Oefenboek Statistiek

OEFENBOEK STATISTIEK

METHODEN EN TECHNIEKEN

L.W.C. Tavecchio

J. Hoogeweij

Wolters-Noordhoff Groningen

Ontwerp binnenwerk: Studio Wolters-Noordhoff bv Groningen
Ontwerp omslag: Studio Wolters-Noordhoff bv Groningen
Omslagillustratie: Martin Rijpstra, Leeuwarden

0 1 2 3 4 5 / 01 00 99 98 97

© 1997 Wolters-Noordhoff bv Groningen, The Netherlands.

Alle rechten voorbehouden. Niets uit deze uitgave mag worden verveelvoudigd, opgeslagen in een geautomatiseerd gegevensbestand, of openbaar gemaakt, in enige vorm of op enige wijze, hetzij elektronisch, mechanisch, door fotokopieën, opnamen of op enig andere manier, zonder voorafgaande schriftelijke toestemming van de uitgever.

Voorzover het maken van kopieën uit deze uitgave is toegestaan op grond van art. 16b en 17 Auteurswet 1912, dient men de daarvoor verschuldigde vergoedingen te voldoen aan de Stichting Reprorecht, Postbus 882, 1180 AW Amstelveen. Voor het overnemen van een of enkele gedeelte(n) uit deze uitgave in bloemlezingen, readers of andere compilatiewerken dient men zich tot de uitgever te wenden.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of the publisher.

ISBN 90 01 83821 9

Voorwoord

De bedoeling van dit boek is om extra oefenmateriaal te verschaffen bij de door velen als 'lastig' of 'vervelend' ervaren studieonderdelen statistiek en methoden & technieken. Deze studieonderdelen nemen in de curricula van veel (sociaal-wetenschappelijke) opleidingen een belangrijke plaats in, die in betekenis eerder toe- dan afneemt. Daarbij denken we niet alleen aan universitaire, maar ook aan hbo-opleidingen.

Het oefenmateriaal richt zich op de eerste plaats op de statistiek- en m & t-stof in de eerste jaren van de opleiding. Meer geavanceerde onderwerpen zijn buiten beschouwing gelaten.

In **deel I** van dit oefenboek wordt in 12 hoofdstukken aandacht besteed aan de statistiek. Bij elk onderwerp dat in de diverse hoofdstukken aan de orde komt, wordt eerst, de ene keer wat uitgebreider dan de andere, aan de hand van voorbeelden ingegaan op het doel van de betreffende techniek. Deze inleiding gaat vooraf aan de opgaven. De bedoeling van deze inleidingen is **niet** om de bestaande en/of door de studenten gebruikte basisliteratuur te vervangen. Daarvoor is dit boek niet geschreven!

De inleidingen dienen uitsluitend het doel de gebruiker op het spoor te zetten van de betreffende techniek, zeg maar als geheugensteuntje of als opfrisser. Daarnaast vervullen ze de rol de lezer vertrouwd te maken met het taalgebruik van de auteurs, zoals dat in de opgaven en de uitgebreide uitwerkingen daarvan gehanteerd wordt. Evenmin zal men in het boek afleidingen of bewijzen aantreffen. Verondersteld wordt dat de student deze kennis reeds heeft verworven in de over deze onderdelen gevolgde cursus(sen). Het boek is *praktijkgericht* en uitsluitend bedoeld als *hulpmiddel* bij het toepassen en verwerken van de stof.

Het statistiekgedeelte van het boek bestrijkt ongeveer de stof die gangbaar is in de eerste 1 à 2 jaar van de opleiding. Zo wordt begonnen met eenvoudige bewerkingen van ruwe scores en het in beeld brengen daarvan in tabellen en figuren. Daarna komen centrum- en spreidingsmaten aan de orde, percentielpunten en -rangen, de (standaard)normaalverdeling en z-scores. Vervolgens wordt ingegaan op correlatie en regressie en op parametrische en non-parametrische technieken voor het toetsen van hypothesen, inclusief variantie-analyse.

Deel II van het boek behandelt op overeenkomstige wijze een aantal onderwerpen op het gebied van de methoden & technieken. Om de omvang van het boek binnen redelijke perken te houden, is hier eerder sprake van *capita selecta*. De 5 hoofdstukken in dit deel bevatten een relatief beperkte keuze uit de vele onderwerpen die aan bod hadden kunnen komen. Gekozen is voor eenvoudige en complexe proefopzetten, operationaliseren, validiteit en betrouwbaarheid en (bedreigingen van) de interne en externe validiteit van onderzoeksopzetten. Het boek wordt afgesloten met een opgave over een wetenschappelijk artikel, waarin zowel statistische als m & t-onderwerpen aan de orde komen.

De hoofdstukken zijn zo opgezet dat ze, eventueel, elk op zich als een afzonderlijke leereenheid kunnen worden bestudeerd. Op deze wijze kan men zelf

een keuze maken uit die onderwerpen waarvoor (extra) oefenmateriaal wenselijk of noodzakelijk wordt gevonden. Om de gebruiker tegemoet te komen bij het vinden van zijn weg in de grote hoeveelheid onderwerpen die aan bod komt, is aan het begin van het boek een zeer uitgebreide en gedetailleerde inhoudsopgave opgenomen. Hierdoor is de keuze van het gewenste onderwerp eenvoudig en snel te maken.

Een groot deel van de in het boek gepresenteerde opgaven en uitwerkingen is, zij het in soms sterk bewerkte en aangepaste vorm, afkomstig uit syllabi die in de afgelopen tien jaar zijn gebruikt bij het statistiek- en m & t-onderwijs aan eerste- en tweedejaars pedagogiekstudenten van de Leidse Universiteit. De opmerkingen, suggesties en kritische kanttekeningen, gemaakt door vele jaargangen studenten en werkgroepbegeleid(st)ers, hebben zeker bijgedragen aan de 'kwaliteit' en toegankelijkheid van hetgeen uiteindelijk in het voorliggende boek gepresenteerd kon worden. In het bijzonder zijn wij dank verschuldigd aan collega dr. P.M. Kroonenberg, die gedeelten van door hem samengestelde syllabi voor gebruik en bewerking ter beschikking heeft gesteld. Daarnaast danken wij Bart Bosman, die regelmatig op enthousiaste en kundige wijze heeft bijgedragen aan de zo belangrijke vormgeving en lay-out van deze, in meer dan één opzicht, 'lastige' tekst.
Opmerkingen van gebruikers en suggesties ter aanvulling of verbetering van de inhoud van dit boek zien wij met bijzondere belangstelling tegemoet!

Landsmeer/Rotterdam, februari 1993

Louis Tavecchio
Jacqueline Hoogeweij

Bij de tweede, herziene druk

Deze tweede druk is een grondige herziening van de eerste, waarbij een aantal storende fouten en oneffenheden is verwijderd. De op- en aanmerkingen van gebruikers – docenten en studenten – zijn zo goed mogelijk verwerkt. Wij zijn al diegenen die de moeite hebben genomen contact met ons op te nemen, zeer dankbaar. Evenals de vorige keer houden wij ons ten zeerste aanbevolen voor suggesties bij deze nieuwe druk.

Landsmeer/Rotterdam, februari 1997

Louis Tavecchio
Jacqueline Tournier-Hoogeweij

Inhoudsopgave

1 Inleiding statistische begrippen
1.1 Twee soorten statistiek 13
1.2 Metingen 13
1.3 Variabelen 15
1.4 Afronden 17
1.5 Notatie 17
Opgaven 19
Uitwerkingen 21

2 Frequentieverdelingen 25
2.1 Soorten frequentieverdelingen 25
2.1.1 De gewone frequentieverdeling 25
2.1.2 De relatieve frequentieverdeling 25
2.1.3 De cumulatieve frequentieverdeling 25
2.1.4 De relatieve, cumulatieve frequentieverdeling 26
2.1.5 Gegroepeerde frequentieverdelingen 27
2.2 Grafische voorstellingen van frequentieverdelingen 29
2.2.1 Frequentiehistogram 29
2.2.2 Staafdiagram 30
2.2.3 Frequentiepolygoon 31
2.2.4 Cumulatieve frequentiepolygoon/frequentiehistogram 32
2.2.5 Cumulatieve, relatieve frequentiepolygoon/frequentiehistogram 33
2.3 Misleidende plaatjes 34
Opgaven 36
Uitwerkingen 37

3 Kenmerken van verdelingen 40
3.1 Centrale tendentie 40
3.1.1 Het gemiddelde 40
3.1.2 De mediaan (Md) 41
3.1.3 Modus (Mo) 42
3.2 Variabiliteit 43
3.2.1 Het bereik 44
3.2.2 De variantie 45
3.2.3 De standaardafwijking 46
3.3 Kenmerken van een gegroepeerde frequentieverdeling 48
Opgaven 51
Uitwerkingen 53

4 De relatieve positie van gegevens 56
4.1 De percentielberekening 56
4.1.1 Het percentielpunt 56
4.1.2 De percentielrang 58
4.2 De standaardnormaalverdeling 59

4.2.1 De z-transformatie 61
4.2.2 De omgekeerde z-transformatie 65
Opgaven 69
Uitwerkingen 71

5 Regressie en correlatie 76
5.1 Regressie 76
5.1.1 De regressielijn 76
5.2 Correlatie 81
5.2.1 Eigenschappen van de correlatiecoëfficiënt 81
5.2.2 Berekening van de correlatiecoëfficiënt 82
Opgaven 84
Uitwerkingen 86

6 Inleiding in hypothesen toetsen 92
6.1 Populatie en steekproef 92
6.1.1 Kenmerken van populatie en steekproef 93
6.2 Steekproevenverdelingen 93
6.2.1 De empirische steekproevenverdeling 93
6.2.2 De theoretische steekproevenverdeling 93
6.3 Toetsen van hypothesen 98
6.3.1 Waarschijnlijkheid 98
6.3.2 Toetsingsschema 100
6.4 Beslissingsfouten bij het toetsen van hypothesen 107
Opgaven 109
Uitwerkingen 112

7 Toetsen voor het gemiddelde van de populatie op basis van één steekproef en betrouwbaarheidsintervallen 118
7.1 De z-toets 118
7.1.1 Toetskeuze 118
7.1.2 Onderzoeksvraag 119
7.1.3 Hypothesen 119
7.1.4 Aannamen 119
7.1.5 Beslissingsregels 119
7.1.6 Berekening 120
7.1.7 Beslissing 120
7.1.8 Conclusie 121
7.2 De t-toets 121
7.2.1 Toetskeuze 121
7.2.2 Onderzoeksvraag 121
7.2.3 Hypothesen 122
7.2.4 Aannamen 121
7.2.5 Beslissingsregels 122
7.2.6 Berekening 123
7.2.7 Beslissing 123
7.2.8 Conclusie 123
7.3 Tabellen 124

7.4 Betrouwbaarheidsintervallen 124
Opgaven 126
Uitwerkingen 126

8 Vergelijking van de centra van twee onafhankelijke groepen: t-toets voor onafhankelijke groepen en de Mann-Whitney-U-toets 142
8.1 De t-toets voor onafhankelijke groepen 142
8.1.1 Toetskeuze 143
8.1.2 Onderzoeksvraag 143
8.1.3 Hypothesen 143
8.1.4 Aannamen 144
8.1.5 Beslissingsregels 145
8.1.6 Berekening 146
8.1.7 Beslissing 146
8.1.8 Conclusie 146
8.2 De Mann-Whitney-U-toets bij een kleine steekproef (n van beide groepen ≤ 20) 147
8.2.1 Toetskeuze 147
8.2.2 Onderzoeksvraag 147
8.2.3 Hypothesen 148
8.2.4 Aannamen 148
8.2.5 Beslissingsregels 148
8.2.6 Berekening 149
8.2.7 Beslissing 151
8.2.8 Conclusie 151
8.3 Mann-Whitney-U-toets bij een grote steekproef (n van beide groepen > 20) 151
Opgaven 152
Uitwerkingen 154

9 Vergelijking van twee afhankelijke groepen: t-toets voor afhankelijke groepen en de toets van Wilcoxon 162
9.1 De t-toets voor afhankelijke groepen 163
9.1.1 Toetskeuze 163
9.1.2 Onderzoeksvraag 164
9.1.3 Hypothesen 164
9.1.4 Aannamen 164
9.1.5 Beslissingsregels 164
9.1.6 Berekening 165
9.1.7 Beslissing 166
9.1.8 Conclusie 166
9.2 De Wilcoxon-toets bij kleine steekproefomvang (n (= aantal paren scores) ≤ 50) 166
9.2.1 Toetskeuze 167
9.2.2 Onderzoeksvraag 167
9.2.3 Hypothesen 167
9.2.4 Aannamen 168
9.2.5 Beslissingsregels 168

9.2.6 Berekening 169
9.2.7 Beslissing 171
9.2.8 Conclusie 171
9.3 De Wilcoxon-toets bij grotere steekproefomvang (n (= aantal paren scores) > 50) 171
Opgaven 172
Uitwerkingen 174

10 Toetsen voor samenhang: Pearson-, Spearman-, χ^2-toets en vergelijking van twee correlatiecoëfficiënten: z_r-toets 180
10.1 De Pearson-toets voor het toetsen van verband 180
10.1.1 Toetskeuze 181
10.1.2 Onderzoeksvraag 181
10.1.3 Hypothesen 181
10.1.4 Aannamen 183
10.1.5 Beslissingsregels 183
10.1.6 Berekening 183
10.1.7 Beslissing 184
10.1.8 Conclusie 184
10.2 De Spearman-rangordecorrelatietoets 185
10.2.1 Toetskeuze 185
10.2.2 Onderzoeksvraag 186
10.2.3 Hypothesen 186
10.2.4 Aannamen 186
10.2.5 Beslissingsregels 186
10.2.6 Berekening 187
10.2.7 Beslissing 188
10.2.8 Conclusie 188
10.3 De χ^2-toets 188
10.3.1 Toetskeuze 189
10.3.2 Onderzoeksvraag 189
10.3.3 Hypothesen 189
10.3.4 Aannamen 190
10.3.5 Beslissingsregels 190
10.3.6 Berekening 191
10.3.7 Beslissing 192
10.3.8 Conclusie 193
10.3.9 χ^2-toets als verschiltoets 193
10.4 Vergelijken van twee correlatiecoëfficiënten: de z_r-toets
10.4.1 Toetskeuze 193
10.4.2 Onderzoeksvraag 194
10.4.3 Hypothesen 194
10.4.4 Aannamen 194
10.4.5 Beslissingsregels 194
10.4.6 Berekening 195
10.4.7 Beslissing 196
10.4.8 Conclusie 196
Opgaven 197
Uitwerkingen 199

11 Toetsen voor vergelijking van meer dan twee onafhankelijke groepen: de eenwegvariantieanalyse en de toets van Kruskal-Wallis 208
11.1 De eenwegvariantieanalyse 208
11.1.1 Toetskeuze 209
11.1.2 Onderzoeksvraag 209
11.1.3 Hypothesen 209
11.1.4 Aannamen 209
11.1.5 Beslissingsregels 210
11.1.6 Berekening van de geobserveerde waarde met behulp van definitieformules 211
11.1.7 Berekening geobserveerde F-waarde met behulp van rekenformules 214
11.1.8 Beslissing 215
11.1.9 Conclusie 215
11.2 De toets van Kruskall-Wallis 216
11.2.1 Toetskeuze 216
11.2.2 Onderzoeksvraag 216
11.2.3 Hypothesen 216
11.2.4 Aannamen 216
11.2.5 Beslissingsregels 216
11.2.6 Berekening 217
11.2.7 Beslissing 218
11.2.8 Conclusie 218
Opgaven 219
Uitwerkingen 221

12 Tweewegvariantieanalyse 227
12.1 Een voorbeeldanalyse 227
12.1.1 Toetskeuze 228
12.1.2 Onderzoeksvraag 228
12.1.3 Hypothesen 228
12.1.4 Aannamen 229
12.1.5 Beslissingsregels 229
12.1.6 Berekening 230
12.1.7 Beslissing 233
12.1.8 Conclusie 233
12.2 Mogelijke andere effecten bij tweewegvariantieanalyse 233
12.2.1 Hoofdeffect factor A 234
12.2.2 Interactie 234
Opgaven 239
Uitwerkingen 241

13 (Quasi-)experimentele proefopzetten 248
13.1 Voorbeelden van zuivere experimentele proefopzetten 248
13.2 Voorbeelden van quasi-experimenten 249
Opgaven 253
Uitwerkingen 255

14 Complexe experimentele opzetten: Factoriële en gemengde (mixed) proefopzetten en soorten interactie 257
14.1 Voorbeeld 257
14.2 Grafische afbeeldingen van interactie-effecten 259
Opgaven 262
Uitwerkingen 264

15 Operationaliseren, validiteit en betrouwbaarheid 266
15.1 Operationaliseren 266
15.2 Validiteit 267
15.2.1 Soorten validiteit 267
15.3 Betrouwbaarheid 269
15.3.1 Soorten betrouwbaarheid 271
Opgaven 273
Uitwerkingen 274

16 Interne validiteit van onderzoeksopzetten 276
16.1 Bedreigingen van de interne validiteit 276
16.2 Voorbeeld 276
16.2.1 Mogelijke alternatieve verklaringen voor veranderingen tussen voor- en nameting 277
16.2.2 Mogelijke alternatieve verklaringen voor het verschil in verandering tussen experimentele en controlegroep 278
Opgaven 280
Uitwerkingen 281

17 Bedreigingen van de externe validiteit 282
17.1 Soorten externe validiteit 282
17.2 Bedreigingen van de externe validiteit 283
17.3 Het verhogen van de externe validiteit 284
Opgaven 286
Uitwerkingen 289

18 Gemengde opgaven over de totale stof 291
Opgaven 299
Uitwerkingen 301

Geraadpleegde literatuur 307

Appendix I: Statistische tabellen 308

Appendix II: Toetsingschema 325

Over de auteurs 326

1 Inleiding statistische begrippen

Statistiek is de wetenschap die zich bezighoudt met het *verzamelen, bewerken, beschrijven en interpreteren van kwantitatieve informatie*. Kwantitatieve informatie bevat gegevens over de hoeveelheid of grootte waarin een bepaald verschijnsel optreedt.
De vraag rijst nu, waarom het nodig is statistiek toe te passen. Het antwoord op deze vraag ligt op het gebied van onderzoek en het doel daarvan. Over het algemeen is men op zoek naar *algemene uitspraken over bepaalde verschijnselen*.
Een voorbeeld van zo'n algemene uitspraak uit de natuurwetenschappen is, dat onder normale omstandigheden water kookt bij een temperatuur van 100 graden Celsius. Bij het onderzoeken van mensen stuiten we echter op een probleem: zij reageren niet allen op gelijke wijze op behandelingen of situaties, ook al zijn de omstandigheden gelijk. We noemen dit verschijnsel *variabiliteit*. Aangezien de sociale wetenschappen zich bezighouden met onderzoek van mensen, komen we dit verschijnsel vooral hier tegen. De statistiek stelt ons in staat om toch iets over de gegevens te zeggen, ondanks de aanwezigheid van variabiliteit.

1.1 Twee soorten statistiek

Statistiek kunnen we op twee niveaus uitoefenen:
Op het eerste niveau bevindt zich de *descriptieve of beschrijvende statistiek*. Deze bevat een aantal methoden en technieken om gegevens te organiseren, op te sommen en te beschrijven. Op het tweede niveau bevindt zich de *inductieve statistiek*. Deze heeft betrekking op een aantal methoden en technieken die ons in staat stellen gevolgtrekkingen over een grote groep te maken op basis van een kleinere onderzochte groep. Bij deze technieken speelt natuurlijk wel altijd waarschijnlijkheid een rol.
De descriptieve en inductieve statistiek worden door elkaar gebruikt en zijn in onderzoek beide onontbeerlijk. De eerste hoofdstukken zullen betrekking hebben op de descriptieve statistiek.

1.2 Metingen

De gegevens die bij een onderzoek worden verkregen, zijn totstandgekomen door het verrichten van metingen. Bij een *meting* bepalen we de waarde op de meetschaal, die hoort bij de persoon die onderzocht wordt met betrekking tot het kenmerk dat we onderzoeken.
We kunnen bijvoorbeeld iemands lengte meten, de meetschaal is dan een meetlat. Een meetschaal biedt de mogelijkheid tot het toekennen van numerieke waarden (cijfers, getallen) die de mate waarin een bepaald kenmerk aanwezig is, tot uitdrukking brengen. Meetschalen kunnen de volgende kenmerken hebben:

1 *Magnitude*
Als we de gegevens kunnen ordenen, als ze in een verhouding tot elkaar staan, bijvoorbeeld persoon X is langer dan persoon Y.
2 *Gelijke intervallen*
Als er tussen de waarden op de meetschaal een gelijke afstand aanwezig is.
3 *Absoluut nulpunt*
Is aanwezig als er een waarde 0 op de meetschaal is, waarvoor geldt dat als deze waarde aangenomen wordt, het kenmerk dat gemeten wordt niet aanwezig is. Bijvoorbeeld: bij nul centimeter is het kenmerk lengte afwezig.

De soorten meetschalen kunnen we nu als volgt weergeven.

Ratiomeetschaal. Bevat alle bovenstaande kenmerken. Bijvoorbeeld: lengte (in centimeters).

Intervalmeetschaal. Bevat gelijke intervallen en magnitude, maar geen absoluut nulpunt. Bijvoorbeeld: graden Celsius heeft geen absoluut nulpunt, want 0 graden Celsius betekent dat er wel sprake van temperatuur is.

Ordinale meetschaal. Bevat alleen het kenmerk magnitude. Bijvoorbeeld: vier kinderen in volgorde plaatsen van klein naar groot. Je kan dan zeggen: 'dit kind is langer dan dat kind.'

Nominale meetschaal. Bevat geen van de bovenstaande kenmerken. Bijvoorbeeld: sekse, met als categorieën vrouw en man. Of automerken, met als categorieën Renault, Toyota, Citroën enzovoort.
We kunnen een nominale meetschaal herkennen aan de willekeurige volgorde van de categorieën. Daar zit geen logische volgorde in.

In voorbeeld 1.1 tot en met 1.4 zullen we illustreren hoe we de meetschaal kunnen vaststellen.

Voorbeeld 1.1: Ratiomeetschaal
Stel: we zijn benieuwd naar **de verdienste** (uitgedrukt in de hoeveelheid geld) van mensen die werken als oppas voor kinderen. Op welke meetschaal wordt verdienste gemeten?
Deze meetschaal heeft een *absoluut nulpunt*, want als er niets verdiend is, dan is er geen geld verdiend. Dus er is bij de schaalwaarde 0 geen verdienste. Deze meetschaal heeft ook *gelijke intervallen*, want tussen elke waarde op de meetschaal zit telkens evenveel cent. Deze meetschaal bevat ook het kenmerk *magnitude*, want men kan op basis van de waarde op de meetschaal zeggen of iemand meer of minder verdiend heeft dan een ander.
Aangezien alle bovenstaande kenmerken aanwezig zijn, kunnen we concluderen dat verdienste op een **ratiomeetschaal** gemeten kan worden.

Voorbeeld 1.2: Intervalmeetschaal
Stel: we willen weten *wanneer* men begonnen is te werken als oppas voor kleine kinderen. Met welke meetschaal hebben we hier te maken?

Deze meetschaal heeft geen *absoluut nulpunt*: onze jaartelling kent weliswaar het jaar 0, maar dat was een gewoon jaar als alle andere. De waarde '0' betekent in dit geval dus niet de afwezigheid van het kenmerk 'jaar'. De jaartelling kent wel *gelijke intervallen*, tussen elk jaar zit een even grote ruimte (namelijk 12 maanden). De jaartelling heeft ook het kenmerk *magnitude*, want je kan zeggen dat de ene persoon later begonnen is met het werk dan een ander. We zien dat de kenmerken *gelijke intervallen* en *magnitude* wel aanwezig zijn, maar het *absoluut nulpunt* niet; we hebben hier te maken met de **intervalmeetschaal.**

Voorbeeld 1.3: Ordinale meetschaal
Stel: we willen de gezinnen die wel eens gebruikmaken van oppas in een *volgorde van gezinsgrootte* zetten. De gezinnen met de minste kinderen komen dan op een lage schaalwaarde, en de gezinnen met veel kinderen komen dan op een hogere schaalwaarde. In dit geval ordenen we de gezinnen van klein naar groot. Bij ordening hebben we niet te maken met een *absoluut nulpunt*, want het kleinste gezin krijgt geen 0 als schaalwaarde, maar waarschijnlijk een 1. Ook *gelijke intervallen* komen bij ordening niet ter sprake. Het kan gebeuren dat gezin A 1 kind, gezin B 3 kinderen en gezin C 4 kinderen heeft. In dit geval krijgen de gezinnen respectievelijk schaalwaarden 1, 2 en 3, zonder dat er een gelijk aantal kinderen tussen zit. *Magnitude* is bij deze meting echter wel aanwezig, want op grond van de schaalwaarden kunnen we zeggen dat gezin C groter is dan gezin B of A.
Het blijkt dus dat we alleen te maken hebben met het kenmerk magnitude en dus is dit een **ordinale meetschaal**.

Voorbeeld 1.4: Nominale meetschaal
Stel: we willen inventariseren van welke geloofsovertuigingen de gezinnen zijn. We hebben dan een meetschaal met de volgende mogelijkheden:
- rooms-katholiek
- protestants-christelijk
- Nederlands-hervormd
- islamitisch
- andere geloofsovertuiging
- geen geloof

Duidelijk is dat het hier gaat om categorieën die zich onderling niet verhouden tot elkaar. We vergelijken ze ook niet. *Geen van de schaalkenmerken* is hier aanwezig en dus hebben we hier een **nominale meetschaal**.

1.3 Variabelen

Een **variabele** is een bepaald kenmerk dat verschillende waarden aan kan nemen. Lengte bijvoorbeeld is een kenmerk dat per individu verschilt. De tegenhanger van een variabele is een **constante**, een kenmerk dat niet verandert binnen een bepaalde context. Als we de lengte van de Nederlandse student willen bekijken, dan zouden we als constante 1 meter kunnen nemen, iedere student is immers wel 1 meter lang. Wat daar boven op komt aan centimeters is weer variabel per persoon.

Er zijn twee soorten variabelen, die onderscheiden kunnen worden op meetmogelijkheden.
Discrete variabelen zijn variabelen die alleen een begrensde waarde kunnen aannemen tussen twee meetpunten op de meetschaal, bijvoorbeeld: het aantal huisdieren dat iemand heeft; dat kan 1, 2, 3 enzovoort zijn, maar niet 2,5.
Continue variabelen zijn variabelen die theoretisch een oneindig aantal waarden kunnen aannemen tussen twee punten op een meetschaal, bijvoorbeeld: de lengte van persoon X is 1,74 meter. We kunnen de *echte* lengte van deze persoon echter veel secuurder bepalen door naar millimeters, micromillimeters enzovoort te kijken. Zo is het mogelijk om in het oneindige door te gaan met het bepalen van de precieze lengte van persoon X. Omdat dit natuurlijk een onmogelijke taak is om uit te voeren, bepalen we meestal de echte limieten van een continue variabele. Dit gaat als volgt: persoon X is 1,74 meter, de echte lengte van deze persoon ligt waarschijnlijk tussen de echte limieten van deze waarde; namelijk tussen 1,735 en 1,745 meter.
De echte limieten van een continue variabele vindt men door de laatstgemeten meeteenheid (in dit geval zijn dat centimeters) te halveren, op zo'n manier dat er een halve meeteenheid bij de waarde wordt opgeteld en een halve van de waarde wordt afgetrokken. In voorbeeld 1.5 zullen we het bovenstaande illustreren.

Voorbeeld 1.5: Variabelen

Uit de voorbeelden 1.1 tot en met 1.4 kunnen we vier variabelen onderscheiden:

1 Verdienste
2 Jaar waarin men gestart is met het werk
3 Grootte van gezin op basis van het aantal kinderen
4 Geloofsovertuiging van de gezinnen.

In dit voorbeeld zullen we vaststellen of de variabelen continu of discreet zijn.

1 In dit geval is sprake van een *discrete* variabele. Het is namelijk niet mogelijk tot in het oneindige door te gaan met het bepalen van het exact verdiende bedrag. In de praktijk eindigt men met een opgave in centen, bijvoorbeeld *f* 10,75 per uur.
2 Deze variabele is *continu* omdat we in principe tot in het oneindige door kunnen gaan met te bepalen wat het exacte moment was. Iemand zegt bijvoorbeeld 13 mei 1985; het exacte moment kan dan verder bepaald worden in uren, minuten, seconden, milliseconden enzovoort.
3 Dit is een *discrete* variabele omdat op de schaal het aantal personen staat waaruit het gezin bestaat. We kunnen alleen hele aantallen personen hebben en geen halve.
4 Dit is ook een *discrete* variabele omdat men de ene categorie kiest of een andere. Het is niet mogelijk dat iemand als antwoord geeft dat zijn of haar geloofsovertuiging ligt tussen Nederlands hervormd en islamitisch.

In de praktijk wordt iedere variabele, discreet of continu, altijd discreet gemeten.

1.4 Afronden

In het vervolg van dit boek zullen we vaak te maken krijgen met getallen die bestaan uit meerdere decimalen achter de komma. Afronding is dan de oplossing voor deze onhanteerbare getallen. De afrondingsregel waar we ons in dit boek aan zullen houden, is de volgende. Een getal wordt naar boven afgerond als de waarde van de betreffende decimaal 5 of hoger is. Bijvoorbeeld:

1,6 kan worden afgerond naar 2.

1,5677 kan worden afgerond naar 1,568.

Een getal wordt naar beneden afgerond als de waarde van de betreffende decimaal 4 of lager is. Bijvoorbeeld:

200,4 kan worden afgerond naar 200.

3,112 kan worden afgerond naar 3,11.

Het is gebruikelijk om op twee cijfers achter de komma af te ronden.

1.5 Notatie

Een variabele zal in dit boek meestal met een X of een Y worden aangeduid; andere letters kunnen echter ook worden gekozen. Als er een serie getallen moet worden opgeteld, dan wordt dat aangegeven met het $\sum$-teken (sigma). Als we het $\sum$-teken tegenkomen, dan moeten we *per persoon* optellen wat er volgt op het $\sum$-teken. We zullen het gebruik van dit teken illustreren aan de hand van voorbeeld 1.6.

Voorbeeld 1.6: Notatie

Stel we hebben de volgende scores op twee variabelen van 5 studenten:
x (= cijfer statistiek deeltoets 1) en y (= cijfer statistiek deeltoets 2):

Tabel 1.1

Persoon (i)	*Cijfer deeltoets 1 (X)*	*Cijfer deeltoets 2 (Y)*
1	7	6
2	8	5
3	5	4
4	4	6
5	9	8

$\sum x$ = tel alle x-waarden op: dus $7 + 8 + 5 + 4 + 9 = 33$

$\sum y$ = tel alle y-waarden op: dus $6 + 5 + 4 + 6 + 8 = 29$

$\sum (x + y)$ = tel voor alle personen de x-waarde op bij de y-waarde:

$= (7 + 6) + (8 + 5) + (5 + 4) + (4 + 6) + (9 + 8)$

$= 13 + 13 + 9 + 10 + 17 = 62$

$\sum xy$ = vermenigvuldig voor elke persoon de x- en de y-waarden en tel de uitkomsten van elke persoon op:

$= (7 \cdot 6) + (8 \cdot 5) + (5 \cdot 4) + (4 \cdot 6) + (9 \cdot 8)$

$= 42 + 40 + 20 + 24 + 72 = 198$

We kunnen met behulp van het sigmateken ook bepaalde personen nader aanduiden. Dit ziet er als volgt uit:

1 $\sum_{i=1}^{n} x$ = tel alle waarden van x op, begin bij persoon i = 1 tot en met (n), alle personen.

2 $\sum_{i=1}^{3} y$ = tel de waarden van y op, begin met persoon i = 1 tot en met persoon i = 3.

Let goed op het verschil tussen $\sum x^2$ en $(\sum x)^2$. In het eerste geval berekenen we voor iedere persoon eerst de x-waarde in het kwadraat en tellen vervolgens al deze getallen op; in het tweede geval tellen we eerst alle x-waarden op, en kwadrateren deze som.

Opgaven

1.1 Hieronder vindt u enkele vragen uit het Attachment-onderzoek, survey III, vakgroep W.E.P., 1982.

1 Wanneer bent u geboren? in 19..
2a Hoeveel kinderen hebt u? .. kinderen
9b Een vrouw kan alleen echte bevrediging vinden in een rol als echtgenote en moeder 1 2 3 4
1 = zeer mee oneens
2 = min of meer mee oneens
3 = min of meer mee eens
4 = zeer mee eens
11 Heeft uw kind een fopspeen?
- ja
- wel gehad, nu niet meer
- nee, nooit gehad

12 Als uw kind een fopspeen heeft (of heeft gehad), hoe oud was het toen ermee werd begonnen?

0-3 mnd	1
3-6 mnd	2
6-12 mnd	3
12-18 mnd	4
18-24 mnd	5
24-30 mnd	6

34 Welke opleiding heeft u afgerond?
- alleen lager onderwijs
- voortgezet lager onderwijs (LBO, MULO, MAVO)
- voortgezet middelbaar onderwijs (HAVO, VWO, MBO)
- hoger en wetenschappelijk onderwijs

a Op welk niveau zijn deze vragen gemeten?
b Geef van de *onderliggende* variabelen bij de respectievelijke vragen aan of ze discreet of continu zijn.
c Geef van de variabelen aan of ze discreet of continu gemeten zijn.

1.2 Rond direct af op hele getallen.

a 1,3
b 1,45
c 139,51
d −2,2
e 4,05
f 4,50
g −2,5

1.3 *a* Rond af: 5,46 tot een getal van één cijfer achter de komma (tot één decimaal).
b Rond dit nieuwe getal vervolgens af tot nul cijfers achter de komma (geheel getal of integere waarde).
c Rond 5,46 nu direct af tot een geheel getal.
d Bedenk eens de consequenties voor uw tentamencijfer.

1.4 Op de universiteit van Leiden wordt aan verschillende studenten de volgende vragenlijst voorgelegd.

1 Wat is uw lengte?
.....meter
2 In vergelijking tot de meeste bevriende studenten, bent u dan langer of korter? Ik ben meestal
1 = langer
2 = ongeveer even lang
3 = korter
3 Wat is het hoogste tentamencijfer dat u ooit behaalde op de universiteit (afgerond op hele cijfers)?
Juiste cijfer omcirkelen a.u.b.
0 1 2 3 4 5 6 7 8 9 10
4 Heeft u een computer in uw bezit?
1 = ja
2 = nee, maar ik had er wel één
3 = nee
5 Hoeveel computers heeft u in uw bezit?
Ik heb .... computer(s) in mijn bezit.

a Geef van de bovenstaande vragen aan, welke variabelen men beoogt te meten en met welke meetschaal de vraag gemeten wordt.

1.5 Uit een onderzoek naar het aantal stuks fruit (x) en het aantal stuks groente (y) dat in studentenhuizen aanwezig is, zijn de volgende resultaten afkomstig.

Studentenhuis	*X*	*Y*
1	4	6
2	2	1
3	18	12
4	2	3
5	10	12
6	0	1
7	5	4

Bereken:

a $\sum_{i=1}^{n} x$

b $\sum_{i=1}^{n} y$

c $\sum x$

d $\sum y$

e $\sum_{i=1}^{3} x$

f $\sum_{i=2}^{4} y$

g $\sum xy$

h $\sum (x + y)$

i $\sum x + \sum y$

j $\sum x \cdot \sum y$

Uitwerkingen

1.1 *a* **Vraag 1**

Het meetniveau van deze vraag is interval. Er is namelijk geen absoluut nulpunt, men kan in principe het jaar 0 op de meetschaal zetten en dat was een jaar als alle andere.
Gelijke intervallen zijn wel aanwezig op deze meetschaal, tussen elk jaar zitten telkens 12 maanden, enzovoort.
Magnitude is ook aanwezig, de ene persoon kan later in de jaartelling geboren zijn dan een andere persoon.

Vraag 2a

Deze vraag meet het aantal kinderen. Dit is een ratiomeetschaal. Er is wel een absoluut nulpunt, als iemand nul kinderen heeft, is het kenmerk 'kind' niet aanwezig bij die persoon.
Er zijn ook gelijke intervallen, tussen elke waarde op de meetschaal zit steeds een gelijk aantal kinderen.
Ook het kenmerk magnitude is aanwezig, want de ene persoon heeft meer kinderen dan de andere.

Vraag 9b

Deze vraag meet de mening van de ondervraagden over deze stelling. Het betreft hier een ordinaal meetniveau. Er is geen sprake van een absoluut nulpunt, dit is namelijk niet denkbaar bij meningsvragen.
Gelijke intervallen zijn er ook niet, tussen zeer mee oneens en min of meer mee oneens kan onmogelijk evenveel meningsverschil zitten als tussen min of meer mee oneens en min of meer mee eens.
Magnitude is wel een kenmerk van deze meetschaal, want men kan ordening in de mate van mee eens zijn brengen.

Vraag 11

Hier hebben we te maken met een nominale meetschaal. Het betreft hier slechts drie antwoordcategorieën waar men in kan vallen. Deze categorieën staan niet in verhouding tot elkaar. We kunnen het nominale meetniveau ook herkennen aan de willekeurige volgorde van de antwoordcategorieën. Men had ook de volgende volgorde kunnen geven: nee, ja, wel gehad, nu niet meer.

Vraag 12

Deze vraag is op ordinaal meetniveau gemeten. Er is wel sprake van een absoluut nulpunt, maar de gelijke intervallen zijn niet aanwezig; de ene keer zitten er 3 maanden in een interval en de andere keer 6 maanden. Het kenmerk magnitude is wel aanwezig.

Vraag 34

Type afgeronde opleiding is de variabele die in deze vraag aan bod komt. Zoals duidelijk zal zijn, kunnen we hier niet van een absoluut nulpunt spre-

ken. Er is immers niet eens met getallen gewerkt. Ook gelijke intervallen komen niet voor op deze meetschaal, we kunnen niet zeggen dat er tussen elk type opleiding evenveel opleidingsverschil zit. Over het kenmerk magnitude valt te twisten: er zijn mensen die vinden dat de opleidingen in volgorde gezet kunnen worden van hoog naar laag; men kan de schaal echter ook opvatten als een categorieënsysteem van allerlei soorten opleiding. In het eerste geval is er sprake van een ordinale meetschaal; ben je de tweede mening toegedaan, dan is dit een nominale meetschaal.

b/c **Vraag 1**

De variabele jaartelling is een continue variabele; we kunnen immers tot in het oneindige doorgaan met bepalen wanneer het exacte moment van geboorte was. Deze variabele is echter discreet gemeten, omdat we alleen het jaar van geboorte kunnen invullen.

Vraag 2a

De variabele 'kindertal' is een discrete variabele, omdat het gewoonweg niet mogelijk is om een waarde tussen twee meetpunten op de meetschaal aan te nemen: men heeft één of twee kinderen, maar niet anderhalf kind. Deze variabele is ook discreet gemeten, we moeten een geheel kindertal invullen.

Vraag 9b

Deze meningsvraag is een continue variabele, omdat we tot in het oneindige kunnen doorgaan met het precies bepalen van onze mening over de stelling. Er is echter sprake van een discrete meting, omdat we deze precieze mening niet kunnen uitdrukken; er moet gekozen worden tussen 1, 2, 3 of 4.

Vraag 11

De onderliggende variabele 'het al dan niet hebben van een fopspeen' is een discrete variabele, we kunnen niet tot in het oneindige doorgaan met bepalen of een kind een fopspeen heeft. Deze variabele is discreet gemeten.

Vraag 12

Het moment waarop er begonnen werd met de fopspeen is een continue variabele, we kunnen tot in het oneindige doorgaan met de precieze minuut, seconde enzovoort. De vraag wordt discreet gemeten, want we moeten kiezen tussen een aantal afgebakende periodes.

Vraag 34

De onderliggende variabele van deze vraag is type afgeronde opleiding, dit is een discrete variabele, die ook discreet gemeten is.

1.2 *a* 1
b 1
c 140
d −2
e 4
f 5
g −3

1.3 *a* 5,5
b 6
c 5
d Verzoek de desbetreffende docent om 'trapsgewijze' af te ronden.

1.4 **1** Deze vraag beoogt de variabele lengte te meten. De meetschaal van lengte kent een absoluut nulpunt, want iets dat 0 mm of centimeter is, bevat het kenmerk lengte niet.
Er zijn tevens gelijke intervallen tussen elke meeteenheid op de meetschaal. En de meetschaal bevat ook magnitude; je kunt zeggen: 'Jan is langer dan Piet.'
Dit is dus een ratiomeetschaal.

2 Deze vraag beoogt een ordening van lengten te bewerkstelligen. Deze meetschaal bevat geen absoluut nulpunt, want het getal 0 is niet denkbaar op een meetschaal die ordent. Ook bevat deze meetschaal geen gelijke intervallen, het gaat erom of iemand langer of korter is en niet hoeveel langer of korter. Magnitude is het kenmerk dat wel aanwezig is bij deze meetschaal, we ordenen immers mensen van klein naar groot, en dus kunnnen we zeggen: 'Ik ben langer dan jij.'
Dit is dus een ordinale meetschaal.

3 Deze vraag beoogt de variabele 'hoogste tentamencijfer' te meten. Deze meetschaal is wat moeilijker te beoordelen, omdat daar in de praktijk nogal eens over gediscussieerd wordt. Het gaat hierbij vooral om het kenmerk gelijke intervallen. We zullen dit probleem stapsgewijs benaderen.
Het absoluut nulpunt is niet aanwezig, want een 0 is net zo goed een cijfer als een 10 of een 5, het kenmerk cijfer is dus bij 0 wel aanwezig.
Gelijke intervallen zijn wel aanwezig als men de opvatting aanhoudt dat er tussen elk cijfer evenveel verschil in kennis zit, dus dat twee personen met de cijfers 4 en 5 evenveel verschillen wat betreft hun prestatie als twee personen met de cijfers 9 en 10. Anderen zijn niet deze mening toegedaan en vinden dus dat het kenmerk gelijke intervallen niet aanwezig is.
Over het kenmerk magnitude bestaat wel algemeen consensus. We kunnen uit de cijfermeetschaal namelijk afleiden wie er beter gepresteerd heeft.
Van de mening over de voorgaande punten is het dus afhankelijk of het hier gaat om een interval of een ordinale meetschaal.

4 Met deze vraag wil men de variabele 'computerbezit' meten. We hebben hier drie antwoordcategorieën aangeboden gekregen die totaal niet met elkaar in een bepaalde hoeveelheidsrelatie staan; het gaat hier dan ook om een nominale meetschaal.

5 De laatste vraag meet de hoeveelheid computers. Deze meetschaal kent wel een absoluut nulpunt, want als iemand 0 computers heeft, dan heeft diegene geen computer; het kenmerk computer is dan niet aanwezig.
Gelijke intervallen bevat deze meetschaal ook, tussen elke willekeurige twee waarden op de meetschaal zit telkens evenveel ruimte. Ook het kenmerk

magnitude is aanwezig, je kan in termen van meer of minder computers spreken. Er is dus sprake van meting op rationiveau.

1.5 *a* $\sum_{i=1}^{n} x = 4 + 2 + 18 + 2 + 10 + 0 + 5 = 41$

b $\sum_{i=1}^{n} y = 6 + 1 + 12 + 3 + 12 + 1 + 4 = 39$

c $\sum x = 4 + 2 + 18 + 2 + 10 + 0 + 5 = 41$

d $\sum y = 6 + 1 + 12 + 3 + 12 + 1 + 4 = 39$

e $\sum_{i=1}^{3} x = 4 + 2 + 18 = 24$

f $\sum_{i=2}^{4} y = 1 + 12 + 3 = 16$

g $\sum xy = (4 \cdot 6) + (2 \cdot 1) + (18 \cdot 12) + (2 \cdot 3) + (10 \cdot 12) + (0 \cdot 1) + (5 \cdot 4)$
$= 24 + 2 + 216 + 6 + 120 + 0 + 20 = 388$

h $\sum (x + y) = (4 + 6) + (2 + 1) + (18 + 12) + (2 + 3) + (10 + 12) +$
$(0 + 1) + (5 + 4) = 10 + 3 + 30 + 5 + 22 + 1 + 9 = 80$

i $\sum x + \sum y = 41 + 39 = 80$

j $\sum x \cdot \sum y = 41 \cdot 39 = 1599$

2 Frequentieverdelingen

In het vorige hoofdstuk vonden we de basis van onderzoek: het selecteren van de variabele en de bijbehorende meetschaal. Het resultaat van een meting van de variabele bij een groot aantal mensen is een lange rij ruwe, dat wil zeggen onbewerkte scores. Om enig inzicht in deze brij te krijgen, moeten we frequentieverdelingen construeren.
In dit hoofdstuk zullen we alle soorten en maten van de frequentieverdeling de revue laten passeren. We zullen daarbij eerst een overzicht geven van alle mogelijkheden die er zijn bij het maken van zo'n verdeling. Daarna zullen we alle varianten aan bod laten komen in een uitgebreid voorbeeld.

2.1 Soorten frequentieverdelingen

2.1.1 *De gewone frequentieverdeling*

Wil men overzicht en inzicht krijgen in een aantal scores, dan is het handig om een frequentieverdeling te maken. Men zet dan alle scores die voorkomen van hoog naar laag onder elkaar en daarnaast zet men het getal dat correspondeert met het aantal mensen dat die score behaald heeft. Dit noemen we de **frequentie**. In voorbeeld 2.1 zien we in kolom A de x-waarden en in kolom B de frequentie waarmee die x-waarden voorkomen. Daardoor kunnen we in één oogopslag zien hoe vaak elke score voorkomt.

2.1.2 *De relatieve frequentieverdeling*

Een stap verder gaat de relatieve frequentieverdeling. Deze verdeling geeft voor elke score aan welke *proportie* van de mensen deze score behaald heeft. Je komt aan de relatieve frequentie door het aantal (frequentie) dat bij een score hoort, te delen door het totaal aantal scores in het onderzoek. Wil men percentages verkrijgen, dan moet men het verkregen getal vermenigvuldigen met 100. Als we alle relatieve frequenties optellen, komen we op 1 uit, of 100 in het geval van procenten.
In kolom C van voorbeeld 2.1 zien we de relatieve frequenties staan.

2.1.3 *De cumulatieve frequentieverdeling*

Deze verdeling telt de frequenties van de laagste naar de hoogste x-waarden op. We kunnen dan met één blik zien hoeveel mensen een bepaalde score of lager hebben gehaald.
Het echte nut van deze verdeling komen we echter pas tegen in hoofdstuk 3 en 4 van dit boek. We kunnen controleren op de juistheid van de optelling; het totaal aantal x-waarden (N) moet namelijk gelijk zijn aan het totaal van de cumulatieve frequentie bij de hoogste x-waarde.
In kolom D van voorbeeld 2.1 zien we een voorbeeld van de cumulatieve frequenties.

2.1.4 *De relatieve, cumulatieve frequentieverdeling*

Deze verdeling komt op gelijke wijze als de cumulatieve frequentieverdeling tot stand, alleen tellen we hier de relatieve frequenties op in plaats van de gewone frequenties. In kolom E van het voorbeeld (2.1) zien we deze frequentieverdeling afgebeeld.

Voorbeeld 2.1: Frequentieverdelingen

Hieronder staan de scores van een fictief onderzoek naar het aantal uren werk dat verricht is door 20 mensen bij oppascentrale **Slaap kindje slaap**. Elk cijfer stelt een aantal uren voor.

2	3	4	5
5	1	3	3
4	2	6	7
3	5	6	4
6	4	3	3

Om inzicht te verkrijgen in deze scores gaan we de verschillende frequentieverdelingen maken.

Tabel 2.1

Kolom A	*Kolom B*	*Kolom C*	*Kolom D*	*Kolom E*
x-waarden (x)	*frequentie (f)*	*relatieve frequentie (rf)*	*cumulatieve frequentie (cf)*	*cumulatieve relatieve frequentie(crf)*
7	1	0,05	20	1,00
6	3	0,15	19	0,95
5	3	0,15	16	0,80
4	4	0,20	13	0,65
3	6	0,30	9	0,45
2	2	0,10	3	0,15
1	1	0,05	1	0,05
	$\sum f = 20$			

Uit kolom B kunnen we afleiden dat bijvoorbeeld 6 personen 3 uur hebben gewerkt en 3 personen 5 uur.

Kolom C vertelt ons dat 5 procent ($0,05 \times 100$) van de mensen 7 uur heeft gewerkt en 30 procent 3 uur gewerkt heeft.

In kolom D zien we dat bijvoorbeeld 13 mensen 4 uur of minder gewerkt hebben; het getal 13 is immers een optelling van de frequenties van de x-waarden 1, 2, 3 en 4.

Uit kolom E kunnen we aflezen dat bijvoorbeeld 80 procent van de mensen 5 uur of minder heeft gewerkt.

2.1.5 *Gegroepeerde frequentieverdelingen*

In sommige onderzoeken hebben we met zoveel verschillende x-waarden te maken dat een gewone frequentieverdeling geen inzicht kan bieden in de scores. De oplossing voor dit probleem is een frequentieverdeling met *klasseninterval*len.

Meerdere x-waarden vallen dan binnen één klasse en zo wordt het geheel overzichtelijker.

Het doel blijft nog steeds de scores zo goed mogelijk te beschrijven en tot hun recht te laten komen. Daarom moet er goed gelet worden op het kiezen van het aantal klassen. Bij teveel klassen is het nog steeds niet mogelijk om de scores goed samen te vatten, maar te weinig klassen leiden tot een te summiere beschrijving van de scores.

Er moet dus een compromis gevonden worden tussen een goede samenvatting en een adequate beschrijving van de gegevens. Meestal kiest men tussen de 10 en 20 klassen. Het aantal verschillende x-waarden vinden we door van de bovenlimiet van de hoogste score de onderlimiet van de laagste score af te trekken. Het getal dat we dan hebben, delen we door het aantal gekozen klassen, bijvoorbeeld 15. De uitkomst hiervan geeft de klassenbreedte weer. We zullen dit illustreren aan de hand van voorbeeld 2.2.

Voorbeeld 2.2: Bepalen van het aantal klassen

Stel: we hebben een reeks scores met als laagste waarde 8 en als hoogste waarde 99. Het aantal verschillende x-waarden dat binnen deze verdeling kan voorkomen, is dan maximaal: $99{,}5 - 7{,}5 = 92$.

Nu delen we dit aantal door een getal tussen de 10 en de 20 (aantal klassen), bijvoorbeeld 15: $92 : 15 = 6{,}13$

De klassenbreedte is dan 6, dat wil zeggen dat elke klasse bestaat uit 6 verschillende x-waarden.

De nieuwe, representatieve x-waarde voor elke klasse wordt het **klassenmidden**, het midden van de klasse. Handig is wel dat het klassenmidden een heel getal is, dat is namelijk makkelijker rekenen in de volgende hoofdstukken.

Een heel klassenmidden, dat wil zeggen een getal zonder decimalen, krijgen we als de klassenbreedte gelijk is aan een oneven getal. In het bovenstaande voorbeeld hadden we echter een even klassenbreedte. We hadden ook voor een oneven klassenbreedte kunnen kiezen, bijvoorbeeld 5. Dit resulteert in $92 : 5 = 19$ klassen.

Een regel waar men zich aan moet houden bij het bepalen van klassen en klassenbreedte is, dat de afgeronde ondergrens van de onderklasse deelbaar moet zijn door de klassenbreedte, en wel zo dat deze uitkomst een geheel getal moet zijn. Deze regel geldt niet als de ondergrens van de onderste klasse gelijk is aan nul.

De afgeronde grenzen of limieten van een klasseninterval zijn gelijk aan de hoogste en de laagste scores die in de klasse vallen. Om de echte limieten van de klasse te vinden, moet men net als bij de continue variabele een halve meet-

eenheid van de onderste grens aftrekken en een halve meeteenheid bij de bovenste grens optellen.

Voorbeeld 2.3: Limieten van een klasseninterval

Stel: we hebben de klasse 10-14. In deze klasse vallen de x-waarden 10, 11, 12, 13 en 14.

De **afgeronde limieten** van deze klasse zijn 10 en 14, dit zijn namelijk de hoogste en de laagste waarden die binnen de klasse vallen. De **echte limieten** van deze klasse zijn: $10 - 0{,}5 = 9{,}5$ en $14 + 0{,}5 = 14{,}5$.

De **klassenbreedte** kunnen we terugvinden door de laagste echte limiet van de hoogste echte limiet af te trekken, dus: $14{,}5 - 9{,}5 = 5$.

Het **klassenmidden** is gelijk aan de optelling van de klassengrenzen en die uitkomst delen door 2, dus: $10 + 14 = 24 : 2 = 12$ of $9{,}5 + 14{,}5 = 24 : 2 = 12$.

Voorbeeld 2.4: Gegroepeerde frequentieverdeling

Uit een onderzoek naar de leeftijd van 50 mensen die betaald oppaswerk verrichten in Nederland zijn de volgende resultaten afkomstig. De cijfers stellen leeftijd in jaren voor.

37	65	40	46	35	44	30	47	30	28
61	41	38	26	20	32	62	42	45	38
30	40	25	39	55	36	22	34	24	30
20	15	31	21	18	25	36	29	20	31
28	25	17	23	30	29	30	36	59	25

Uit deze gegevens kunnen we zo niets opmaken, en waarschijnlijk brengt een gewone frequentieverdeling ook geen duidelijkheid, want we hebben hier te maken met een tamelijk groot aantal x-waarden die beschreven moeten worden; om precies te zijn is het maximumaantal verschillende x-waarden:
$65{,}5 - 14{,}5 = 51$.

Het lijkt dus handig om hier een gegroepeerde frequentieverdeling te maken. We kiezen een aantal klassen waarin we de 51 x-waarden willen beschrijven, bijvoorbeeld 10. De klassenbreedte waar we op uit komen is dan $51 : 10 = 5{,}1$.
Afgerond op 5 is deze klassenbreedte acceptabel en hanteerbaar, want dit is een oneven getal dat resulteert in klassenmiddens met gehele getallen. Nu beginnen we bij de laagste x-waarde met groeperen.
We moeten daarbij ook letten op de deelbaarheid van dit getal door de klassenbreedte; de leeftijd van 15 jaar, de laagste waarde, is deelbaar door 5. De uitkomst 3 is een geheel getal, dus dat zit goed. In tabel 2.2 zien we de gegroepeerde frequentieverdeling van dit voorbeeld.

Tabel 2.2

Klasse	*Klassen-midden (x)*	*Frequentie (f)*	*Relatieve frequentie (rf)*	*Cumulatieve frequentie (cf)*	*Relatieve cumulatieve frequentie (rcf)*
65-69	67	1	0,02	50	1,00
60-64	62	2	0,04	49	0,98
55-59	57	2	0,04	47	0,94
50-54	52	0	0,00	45	0,90
45-49	47	3	0,06	45	0,90
40-44	42	5	0,10	42	0,84
35-39	37	8	0,16	37	0,74
30-34	32	10	0,20	29	0,58
25-29	27	9	0,18	19	0,38
20-24	22	7	0,14	10	0,20
15-19	17	3	0,06	3	0,06
		n = 50			

2.2 Grafische voorstellingen van frequentieverdelingen

Van elke hierboven behandelde frequentieverdeling is het mogelijk een grafische voorstelling te maken. Deze voorstellingen komen we vaak tegen in de media. In één oogopslag kunnen we de situatie via een grafische voorstelling overzien. We zullen hieronder de verschillende mogelijke afbeeldingen behandelen. Daarbij zullen we van elke mogelijkheid een afbeelding projecteren, uitgaande van voorbeeld 2.4.

2.2.1 *Frequentiehistogram*

Een frequentiehistogram wordt afgebeeld in een assenstelsel met een x- en een y-as.

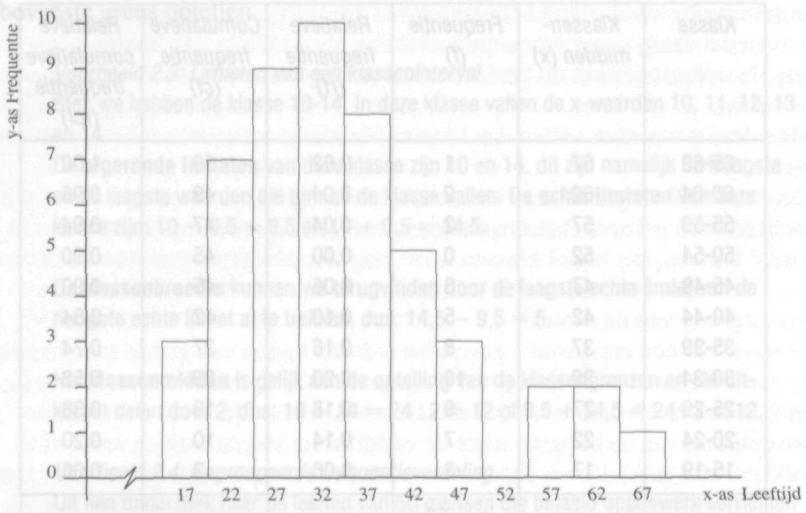
Voor elke x-waarde wordt een staaf getekend waarbij de hoogte van de staaf overeenkomt met de frequentie waarin de x-waarde voorkomt.

Op de x-as komen de scores te staan; in het geval van een gegroepeerde frequentie-verdeling komen de klassenmiddens op de x-as te staan. De staaf van elke x-waarde loopt van de lage echte limiet tot de hoge echte limiet.

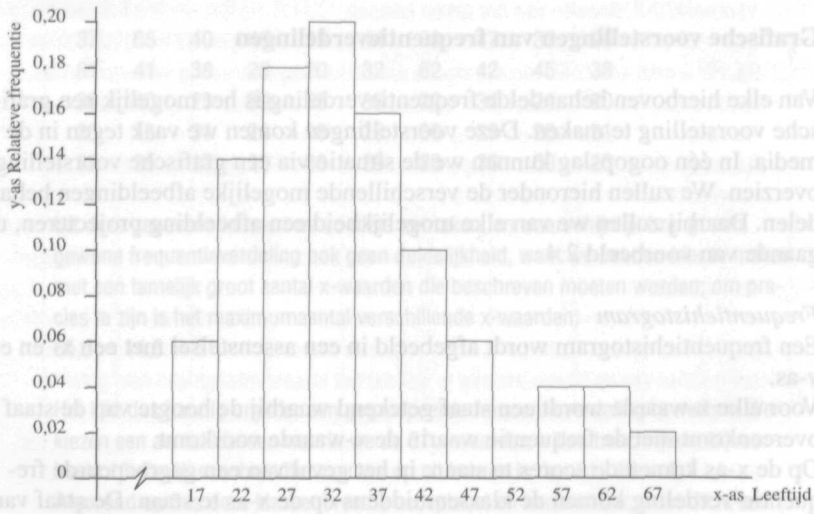
Zodoende sluiten de staven op elkaar aan.

Op de y-as komen de frequenties te staan. Dit kunnen de absolute (figuur 2.1), maar ook de relatieve frequenties zijn (figuur 2.2). Belangrijk is wel dat de y-as gelijkmatig verdeeld wordt, anders zou de afbeelding een scheef beeld van de werkelijkheid geven.

Figuur 2.1 Frequentiehistogram (absolute frequenties)



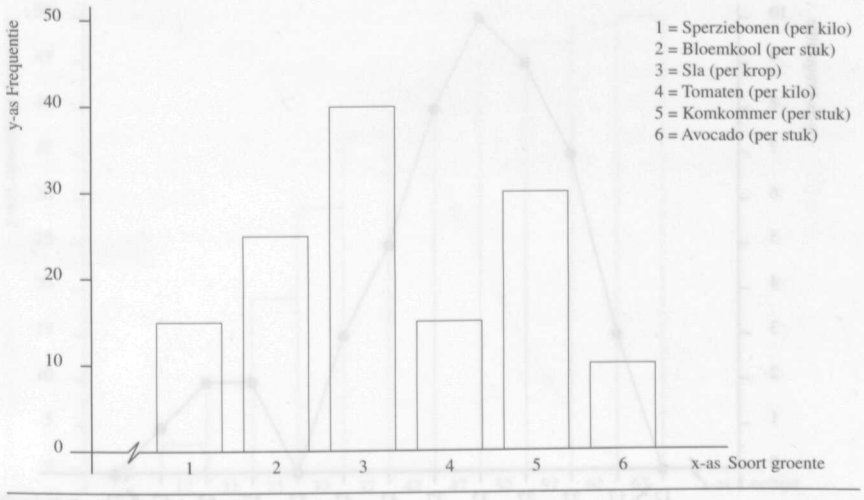
Figuur 2.2 Relatief frequentiehistogram



2.2.2 *Staafdiagram*

Een staafdiagram is vergelijkbaar met een histogram met dit verschil, dat de waarden op de x-as op een nominale meetschaal gemeten zijn. Zodoende sluiten de staven niet op elkaar aan, de categorieën lopen immers niet in elkaar over zoals bij ordinale, interval of ratiomeetschalen. Hieronder staat een voorbeeld van een staafdiagram (figuur 2.3). Op de x-as staan verschillende groenten en op de y-as staat de frequentie waarin deze groenten per dag verkocht worden, in een fictieve groentewinkel te Berlicum.

Figuur 2.3 **Staafdiagram**

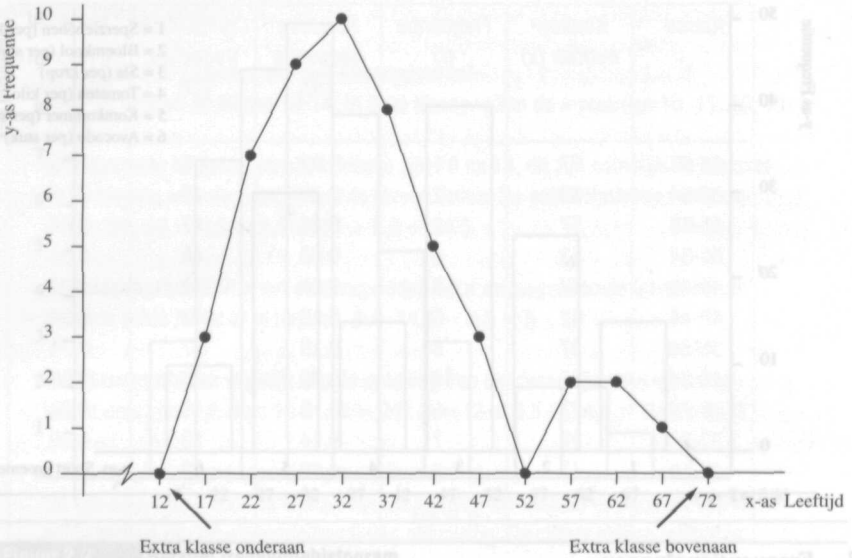


2.2.3 *Frequentiepolygoon*

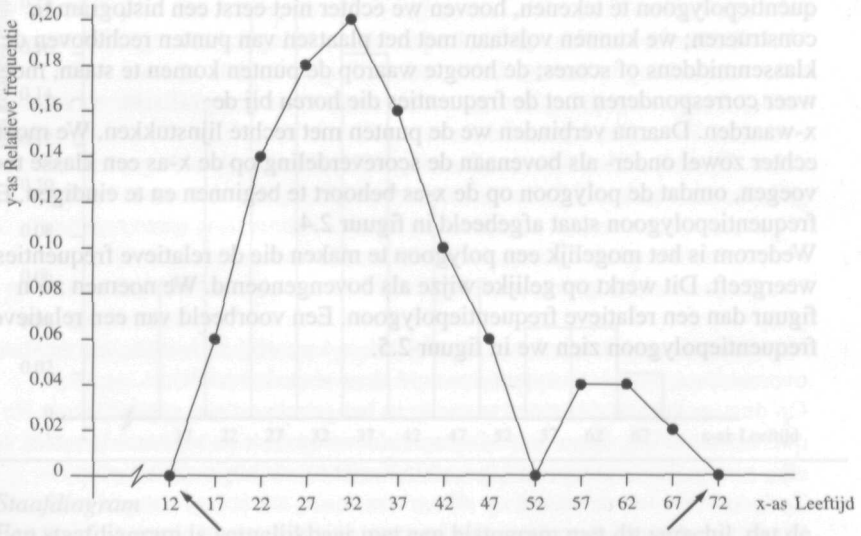
Een frequentiepolygoon is een afbeelding van een lijn die de middens van de bovenkant van de staafjes van een histogram met elkaar verbindt. Om een frequentiepolygoon te tekenen, hoeven we echter niet eerst een histogram te construeren; we kunnen volstaan met het plaatsen van punten rechtboven de klassenmiddens of scores; de hoogte waarop de punten komen te staan, moet weer corresponderen met de frequenties die horen bij de x-waarden. Daarna verbinden we de punten met rechte lijnstukken. We moeten echter zowel onder- als bovenaan de scoreverdeling op de x-as een klasse toevoegen, omdat de polygoon op de x-as behoort te beginnen en te eindigen. Een frequentiepolygoon staat afgebeeld in figuur 2.4.

Wederom is het mogelijk een polygoon te maken die de relatieve frequenties weergeeft. Dit werkt op gelijke wijze als bovengenoemd. We noemen zo'n figuur dan een relatieve frequentiepolygoon. Een voorbeeld van een relatieve frequentiepolygoon zien we in figuur 2.5.

Figuur 2.4 Frequentiepolygoon



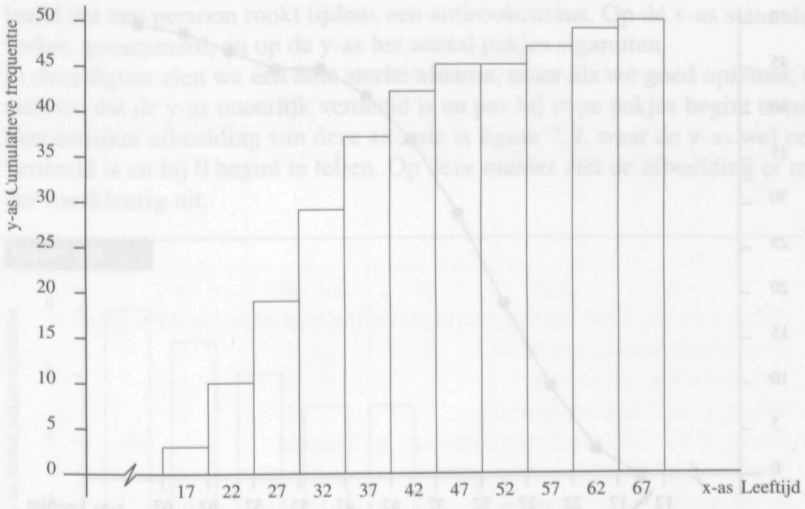
Figuur 2.5 Relatieve frequentiepolygoon



2.2.4 *Cumulatieve frequentiepolygoon/frequentiehistogram*

Bij deze grafische voorstellingen staan op de y-as de cumulatieve frequenties weer-gegeven. Op de x-as staan weer de x-waarden of de klassenmiddens. De hoogte van de staven van het cumulatieve frequentiehistogram komen overeen met de cumulatieve frequenties uit de cumulatieve frequentieverdeling. In figuur 2.6 zien we een cumulatief frequentiehistogram van voorbeeld 2.4.

Figuur 2.6

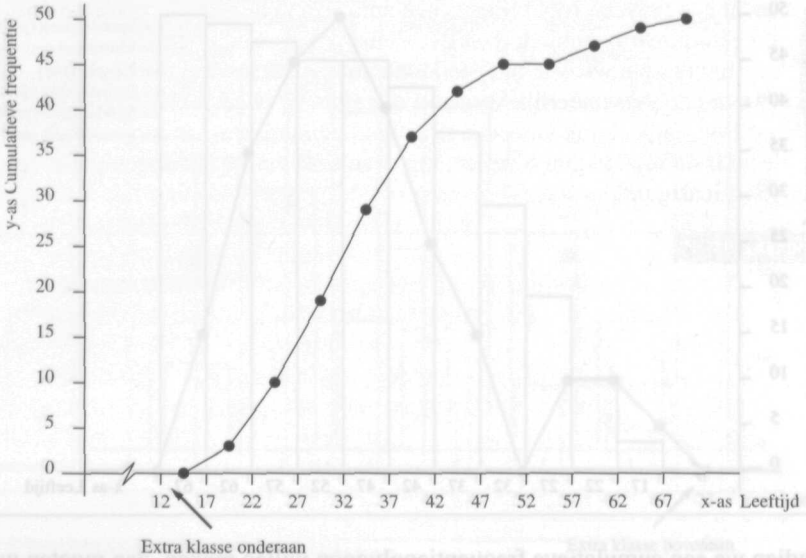


Indien we een cumulatieve frequentiepolygoon willen maken, dan moeten we op een aantal zaken letten. Ten eerste worden de waarnemingspunten niet recht boven het klassenmidden of de x-waarde geplaatst, zoals bij de gewone frequentiepolygoon, maar recht boven de echte bovengrens van het klasseninterval. Ten tweede voegen we alleen aan de onderkant van de scoreverdeling op de x-as een klasse toe; de lijn eindigt immers rechts-boven en kan onmogelijk eindigen op de x-as. In figuur 2.7 staat een cumulatieve frequentiepolygoon.

2.2.5 *Cumulatieve, relatieve frequentiepolygoon/frequentiehistogram*

Deze figuren komen op gelijke wijze tot stand als de cumulatieve frequentiepolygoon c.q. -histogram, met als uitzondering dat de y-as correspondeert met de cumulatieve, relatieve frequenties.

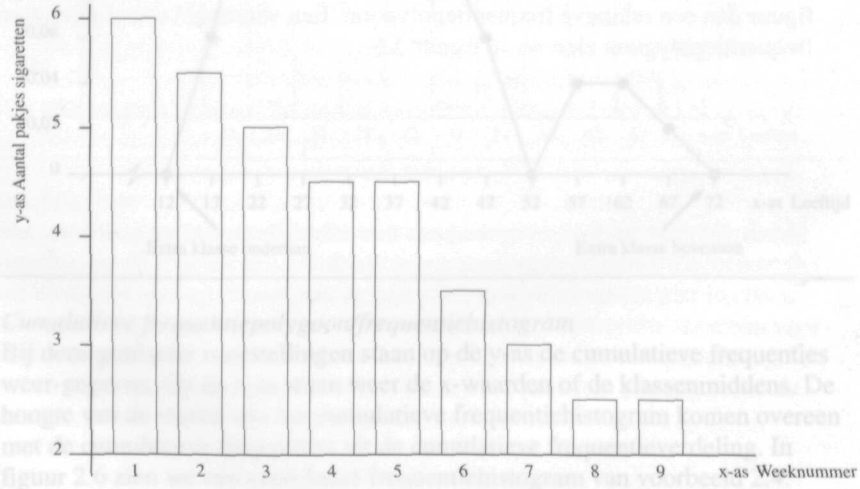
Figuur 2.7 Cumulatieve frequentiepolygoon



2.3 Misleidende plaatjes

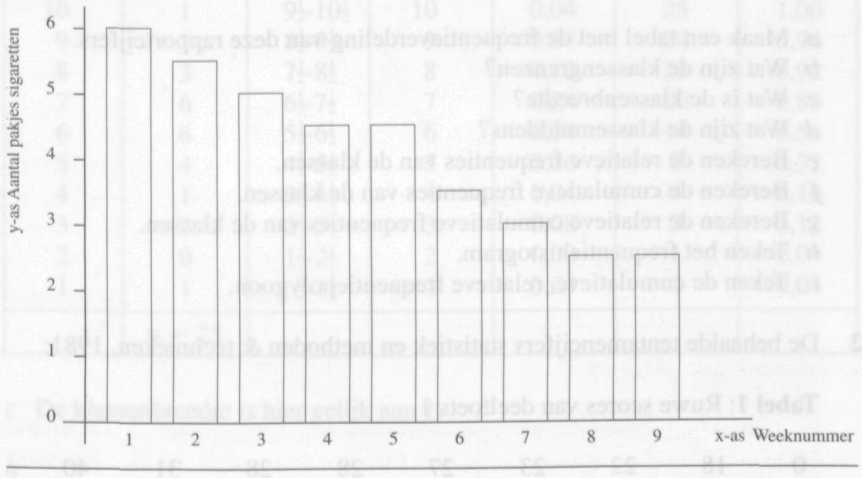
We willen dit hoofdstuk afsluiten met enkele waarschuwende voorbeelden van misleidende grafische afbeeldingen, die we maar al te vaak tegenkomen in de media. Om de harde werkelijkheid te verbloemen, maken de media gebruik van bepaalde technieken om de niet-oplettende lezer of kijker het idee te geven dat de zaken er beter voorstaan dan in werkelijkheid. In figuur 2.8. zien we zo'n plaatje.

Figuur 2.8



In deze fictieve figuur wordt de afname van het aantal pakjes sigaretten afgebeeld dat een persoon rookt tijdens een antirookcursus. Op de x-as staan de weken genummerd, en op de y-as het aantal pakjes sigaretten.
In deze figuur zien we een hele sterke afname, maar als we goed opletten, dan zien we dat de y-as oneerlijk verdeeld is en pas bij twee pakjes begint te tellen. Een eerlijker afbeelding van deze situatie is figuur 2.9, waar de y-as wel eerlijk verdeeld is en bij 0 begint te tellen. Op deze manier ziet de afbeelding er minder rooskleurig uit.

Figuur 2.9



Opgaven

2.1 Hieronder staan de rapportcijfers Grieks van klas B5.

6	8	9	6	5
7	7	3	7	10
6	4	1	6	5
6	7	8	7	6
5	3	8	7	5

a Maak een tabel met de frequentieverdeling van deze rapportcijfers.
b Wat zijn de klassengrenzen?
c Wat is de klassenbreedte?
d Wat zijn de klassenmiddens?
e Bereken de relatieve frequenties van de klassen.
f Bereken de cumulatieve frequenties van de klassen.
g Bereken de relatieve cumulatieve frequenties van de klassen.
h Teken het frequentiehistogram.
i Teken de cumulatieve, relatieve frequentiepolygoon.

2.2 De behaalde tentamencijfers statistiek en methoden & technieken, 1981:

Tabel 1: Ruwe scores van deeltoets 1

0	18	22	23	27	28	28	31	40	40
41	41	41	42	43	44	44	44	44	44
44	44	45	45	45	45	45	45	45	46
32	32	34	34	35	36	37	37	38	38
39	39	39	60	60	60	60	61	61	61
62	62	63	63	63	63	63	63	64	64
66	66	66	66	67	47	47	47	47	47
47	47	48	48	48	49	49	49	50	50
51	51	51	51	51	52	52	52	52	52
52	53	53	53	53	53	53	53	54	54
54	54	54	54	54	54	54	55	55	55
55	55	55	56	56	57	57	57	57	57
58	58	58	58	59	59	59	68	68	68
69	69	69	69	70	70	71	71	72	72
72	72	73	73	74	74	75	75	78	79
81	82	84	84	84					

a Maak hiervan een gegroepeerde frequentieverdeling.
b Bereken de echte klassengrenzen.
c Wat is de klassenbreedte?
d Bereken de klassenmiddens.
e Bereken de cumulatieve frequentieverdeling.
f Teken de frequentiepolygoon.
g Teken de cumulatieve frequentiepolygoon.

Uitwerkingen

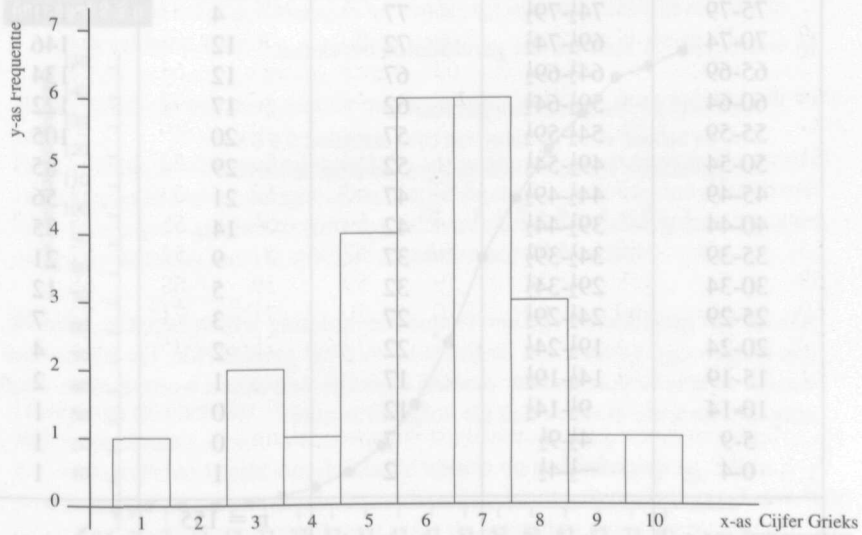
2.1 a, b, d tot en met g

klasse	*frequentie (f)*	*klassen-grenzen*	*klassen-midden (x)*	*relatieve frequentie (rf)*	*cumula-tieve fre-quentie (cf)*	*cumula-tieve rela-tieve fre-quentie (rcf)*
10	1	$9\frac{1}{2}$-$10\frac{1}{2}$	10	0,04	25	1,00
9	1	$8\frac{1}{2}$-$9\frac{1}{2}$	9	0,04	24	0,96
8	3	$7\frac{1}{2}$-$8\frac{1}{2}$	8	0,12	23	0,92
7	6	$6\frac{1}{2}$-$7\frac{1}{2}$	7	0,24	20	0,80
6	6	$5\frac{1}{2}$-$6\frac{1}{2}$	6	0,24	14	0,56
5	4	$4\frac{1}{2}$-$5\frac{1}{2}$	5	0,16	8	0,32
4	1	$3\frac{1}{2}$-$4\frac{1}{2}$	4	0,04	4	0,16
3	2	$2\frac{1}{2}$-$3\frac{1}{2}$	3	0,08	3	0,12
2	0	$1\frac{1}{2}$-$2\frac{1}{2}$	2	0,00	1	0,04
1	1	$\frac{1}{2}$-$1\frac{1}{2}$	1	0,04	1	0,04
	n = 25					

2.1 *c* De klassenbreedte is hier gelijk aan 1.

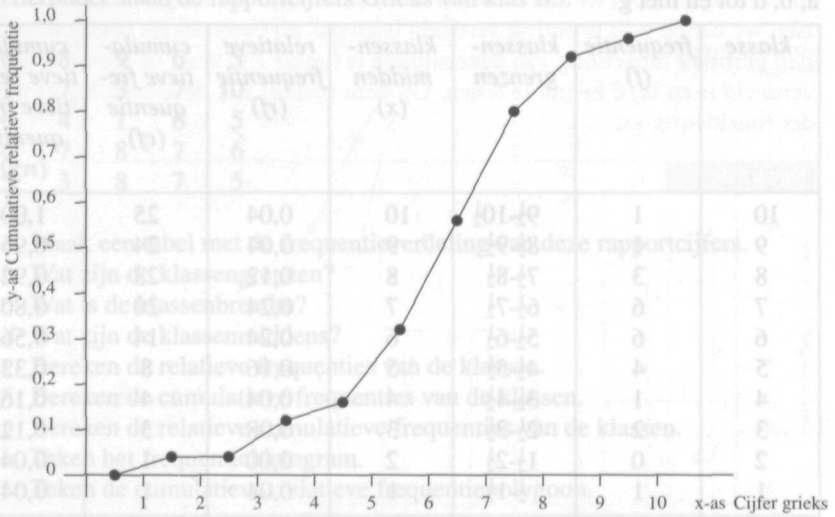
2.1 *h*

Antwoord 2.1 h



2.1 *i*

Antwoord 2.1 i



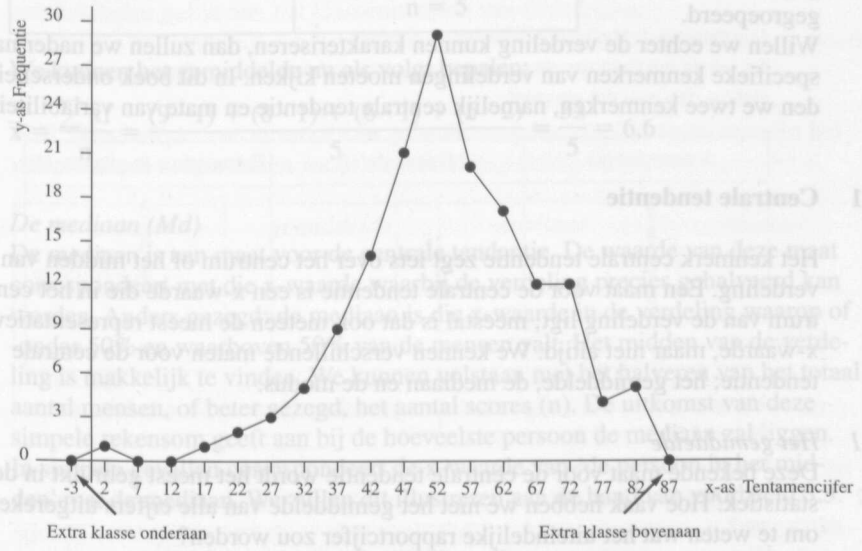
2.2 *a tot en met e*

klasse	*klassengrenzen*	*klassenmidden (x)*	*frequentie (f)*	*cumulatieve frequentie (cf)*
80-84	$79\frac{1}{2}$-$84\frac{1}{2}$	82	5	155
75-79	$74\frac{1}{2}$-$79\frac{1}{2}$	77	4	150
70-74	$69\frac{1}{2}$-$74\frac{1}{2}$	72	12	146
65-69	$64\frac{1}{2}$-$69\frac{1}{2}$	67	12	134
60-64	$59\frac{1}{2}$-$64\frac{1}{2}$	62	17	122
55-59	$54\frac{1}{2}$-$59\frac{1}{2}$	57	20	105
50-54	$49\frac{1}{2}$-$54\frac{1}{2}$	52	29	85
45-49	$44\frac{1}{2}$-$49\frac{1}{2}$	47	21	56
40-44	$39\frac{1}{2}$-$44\frac{1}{2}$	42	14	35
35-39	$34\frac{1}{2}$-$39\frac{1}{2}$	37	9	21
30-34	$29\frac{1}{2}$-$34\frac{1}{2}$	32	5	12
25-29	$24\frac{1}{2}$-$29\frac{1}{2}$	27	3	7
20-24	$19\frac{1}{2}$-$24\frac{1}{2}$	22	2	4
15-19	$14\frac{1}{2}$-$19\frac{1}{2}$	17	1	2
10-14	$9\frac{1}{2}$-$14\frac{1}{2}$	12	0	1
5-9	$4\frac{1}{2}$-$9\frac{1}{2}$	7	0	1
0-4	$-\frac{1}{2}$-$4\frac{1}{2}$	2	1	1
			n = 155	

De klassenbreedte in deze gegroepeerde verdeling is gelijk aan 5, want in elke klasse kunnen 5 verschillende x-waarden voorkomen.

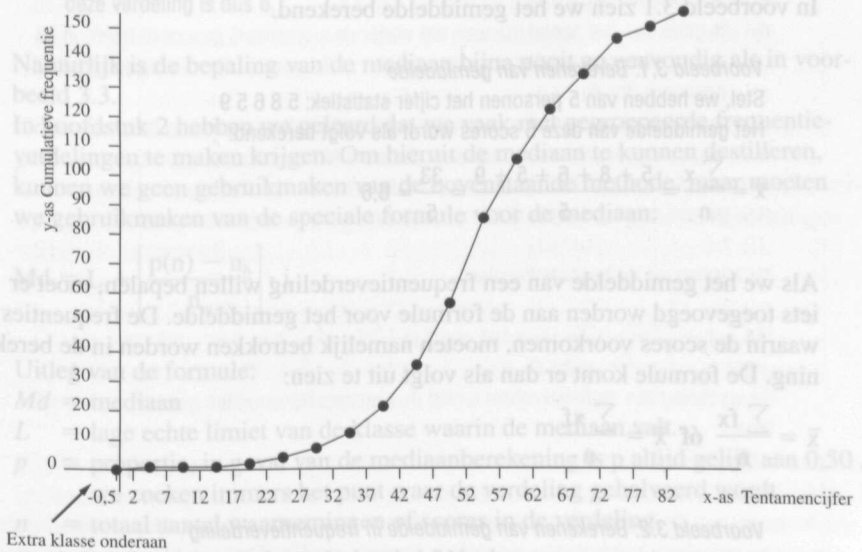
2.2 *f*

Antwoord 2.2 f



2.2 *g*

Antwoord 2.2 g



3 Kenmerken van verdelingen

In voorgaande hoofdstukken hebben we gezien, hoe we onderzoeksgegevens kunnen verzamelen en hoe we inzicht kunnen verkrijgen in grote aantallen gegevens. Dit inzicht verkregen we via frequentieverdelingen, al dan niet gegroepeerd.
Willen we echter de verdeling kunnen karakteriseren, dan zullen we nader naar specifieke kenmerken van verdelingen moeten kijken. In dit boek onderscheiden we twee kenmerken, namelijk centrale tendentie en mate van variabiliteit.

3.1 Centrale tendentie

Het kenmerk centrale tendentie zegt iets over het centrum of het midden van de verdeling. Een maat voor de centrale tendentie is een x-waarde die in het centrum van de verdeling ligt; meestal is dat ook meteen de meest representatieve x-waarde, maar niet altijd. We kennen verschillende maten voor de centrale tendentie: het gemiddelde, de mediaan en de modus.

3.1.1 Het gemiddelde

Deze bekende maat voor de centrale tendentie wordt het meest gebruikt in de statistiek. Hoe vaak hebben we niet het gemiddelde van alle cijfers uitgerekend om te weten wat het uiteindelijke rapportcijfer zou worden?
Het gemiddelde wordt genoteerd als $\bar{x}$, en wordt als volgt bepaald: tel alle x-waarden op en deel ze door het totaal aantal x-waarden.

In formulevorm: $\bar{x} = \frac{\sum x}{n}$

In voorbeeld 3.1 zien we het gemiddelde berekend.

Voorbeeld 3.1: Berekenen van gemiddelde
Stel, we hebben van 5 personen het cijfer statistiek: 5 8 6 5 9
Het gemiddelde van deze 5 scores wordt als volgt berekend:

$$\bar{x} = \frac{\sum x}{n} = \frac{5 + 8 + 6 + 5 + 9}{5} = \frac{33}{5} = 6{,}6$$

Als we het gemiddelde van een frequentieverdeling willen bepalen, moet er iets toegevoegd worden aan de formule voor het gemiddelde. De frequenties waarin de scores voorkomen, moeten namelijk betrokken worden in de berekening. De formule komt er dan als volgt uit te zien:

$$\bar{x} = \frac{\sum fx}{n} \text{ óf } \bar{x} = \frac{\sum xf}{n}$$

Voorbeeld 3.2: Berekenen van gemiddelde in frequentieverdeling
We kunnen de scores van voorbeeld 3.1 als volgt in een frequentieverdeling plaatsen:

x-waarden (x)	*frequentie (f)*
9	1
8	1
6	1
5	2
	n = 5

We kunnen het gemiddelde nu als volgt bepalen:

$$\bar{x} = \frac{\sum xf}{n} = \frac{(9 \cdot 1) + (8 \cdot 1) + (6 \cdot 1) + (5 \cdot 2)}{5} = \frac{33}{5} = 6{,}6$$

3.1.2 *De mediaan (Md)*

De mediaan is een maat voor de centrale tendentie. De waarde van deze maat correspondeert met die x-waarde waarbij de verdeling precies gehalveerd kan worden. Anders gezegd: de mediaan is die x-waarde in de verdeling waarop of -onder 50% en waarboven 50% van de mensen valt. Het midden van de verdeling is makkelijk te vinden. We kunnen volstaan met het halveren van het totaal aantal mensen, of beter gezegd, het aantal scores (n). De uitkomst van deze simpele rekensom geeft aan bij de hoeveelste persoon de mediaan zal liggen. In simpele gevallen correspondeert de x-waarde van 'de persoon in het midden' met de mediaan. We zullen dit illustreren aan de hand van voorbeeld 3.3.

Voorbeeld 3.3: Eenvoudige bepaling van de mediaan

De mediaan van de scoreverdeling van statistiekcijfers uit voorbeeld 3.1 kunnen we als volgt berekenen. Deze cijfers waren: 5 5 6 8 9

↑

We zien dat het cijfer 6 precies in het midden ligt van alle getallen; de mediaan in deze verdeling is dus 6.

Natuurlijk is de bepaling van de mediaan bijna nooit zo eenvoudig als in voorbeeld 3.3.

In hoofdstuk 2 hebben we geleerd dat we vaak met gegroepeerde frequentieverdelingen te maken krijgen. Om hieruit de mediaan te kunnen destilleren, kunnen we geen gebruikmaken van de bovenstaande methode, maar moeten we gebruikmaken van de speciale formule voor de mediaan:

$$Md = L + \left[\frac{p(n) - n_b}{n_w}\right] \cdot i$$

Uitleg van de formule:

Md = mediaan

L = lage echte limiet van de klasse waarin de mediaan valt.

p = proportie, in geval van de mediaanberekening is p altijd gelijk aan 0,50 , we zoeken immers het punt waar de verdeling gehalveerd wordt.

n = totaal aantal waarnemingen of scores in de verdeling

n_b = aantal mensen of scores dat onder de waarde van de mediaan valt (number below).

n_w = aantal mensen dat in de klasse van de mediaan valt (number within)

i = klassenbreedte.

In voorbeeld 3.4 zullen we de toepassing van deze formule toelichten.

Voorbeeld 3.4: Mediaanberekening via formule

Als we de mediaan van de scoreverdeling uit voorbeeld 1 via de formule vast willen stellen, dan gaat dat als volgt:

x-waarden (x)	*Frequentie* (f)	*Cumulatieve frequentie* (cf)
9	1	5
8	1	4
6	1	3
5	2	2
	n = 5	

Allereerst gaan we globaal de plaats van de mediaan bepalen door n te halveren:

p(n) = 0,50(5) = 2,5

Dus bij de 2,5ste persoon ligt de mediaan ongeveer. In de cumulatieve frequentietabel zien we dat 2,5 niet voorkomt, maar wel 2 of 3. Aangezien 2 te laag is, nemen we 3.

Bij deze persoon hoort een score van 6. Deze score vormt het uitgangspunt om de formule in te vullen. We zullen nu eerst alle onbekende tekens in de formule vervangen door getallen.

L is de lage echte limiet van de score 6, in dit geval is dat 5,5.

p(n) is hierboven al berekend en is gelijk aan 2,5.

n_b is gelijk aan het aantal mensen dat onder de x-waarde 6 gescoord heeft, in de cumulatieve frequentiekolom zien we dat er 2 mensen zijn, die lager hebben gescoord dan 6.

n_w is gelijk aan het aantal mensen dat precies 6 heeft gescoord, in dit geval zien we in de gewone frequentiekolom dat slechts één persoon een 6 behaalde.

i is gelijk aan de klassenbreedte, en aangezien we hier te maken hebben met gehele getallen zonder dat ze gegroepeerd zijn, is de klassenbreedte 1.

Nu kunnen we de formule invullen:

$$Md = L + \left[\frac{p(n) - n_b}{n_w}\right] \cdot i \rightarrow Md = 5{,}5 + \left[\frac{2{,}5 - 2}{1}\right] \cdot 1 = 5{,}5 + 0{,}5 \cdot 1 = 6$$

De verdeling van statistiekcijfers wordt dus precies doormidden gedeeld bij het cijfer 6.

3.1.3 *Modus (Mo)*

De modus is de laatste centrummaat die we hier behandelen. Om de modus te kunnen bepalen, hebben we geen berekening nodig. De waarde van de modus

is gelijk aan de x-waarde met de hoogste frequentie. In het geval dat twee x-waarden beide even hoge frequentie hebben, is er sprake van twee modi. We noemen deze verdeling dan ook **bimodaal**. Mochten er meerdere modi voorkomen, dan noemen we een verdeling **multi-modaal**.
In het geval van gegroepeerde frequentieverdelingen moeten we om de modus te kunnen bepalen eerst de klasse met de hoogste frequentie opzoeken. De modus is dan gelijk aan het klassenmidden van deze klasse.
De modus in voorbeeld 3.4 is gelijk aan 5, want 5 komt het meest voor en heeft dus de hoogste frequentie.

N.B. Niet iedere centrummaat kan gebruikt worden op elk meetniveau. In het onderstaande schema zien we een overzicht.

centrummaat → / *meetniveau* ↓	*gemiddelde*	*mediaan*	*modus*
ratio	wel geschikt	wel geschikt	wel geschikt
interval	wel geschikt	wel geschikt	wel geschikt
ordinaal	niet geschikt	wel geschikt	wel geschikt
nominaal	niet geschikt	niet geschikt	wel geschikt

Hoewel het theoretisch niet mogelijk is een gemiddelde te berekenen voor een variabele gemeten op ordinaal meetniveau, wordt het in de praktijk van de sociale wetenschappen maar al te vaak gedaan. Men veronderstelt dan gelijke intervallen tussen de verschillende meetwaarden terwijl deze er niet zijn. Dit gebeurt omdat in de sociale wetenschappen vele variabelen 'slechts' op ordinaal niveau te meten zijn. Als men dan toch gebruik wil maken van 'mooie' statistische maten als het gemiddelde, maakt men van deze valse veronderstelling gebruik.

3.2 Variabiliteit

De mate van variabiliteit geeft aan in hoeverre de scores in een verdeling gespreid zijn. Om een verdeling te karakteriseren, is het niet voldoende te weten wat de waarde van de centrale tendentie is. We kunnen bijvoorbeeld verdelingen hebben die hetzelfde gemiddelde hebben, maar toch verschillen qua variabiliteit (spreiding).

Voorbeeld 3.5: Verdelingen met gelijk gemiddelde
A: 11 12 13 → $\bar{x}$ = 12 → scores liggen dicht bij elkaar, spreiding is klein.
B: 1 12 23 → $\bar{x}$ = 12 → scores liggen veel verder uit elkaar, spreiding is veel groter dan in A.

Uit de informatie over de spreiding van een verdeling kunnen we onder andere informatie inwinnen over de mate van representativiteit van de waarde van de centrale tendentie. Als de scores immers dicht bij bijvoorbeeld het gemiddelde liggen, dan is deze waarde een representatief getal voor de verdeling.

Stel dat de spreiding zeer groot is, dan is bijvoorbeeld het gemiddelde één van de vele getallen in de verdeling zonder representativiteitswaarde. De mate van variabiliteit of spreiding kan in kwantitatieve informatie omgezet worden door middel van de drie spreidingsmaten bereik, variantie en standaardafwijking.

3.2.1 *Het bereik*

Het bereik van een verdeling geeft aan hoeveel verschillende x-waarden de verdeling omvat. Hiertoe bekijken we de hoogste en de laagste waarde van de verdeling, waartussen immers alle verschillende x-waarden liggen. In wezen is het echte bereik van een verdeling te vinden door de onderste echte limiet van de laagste score af te trekken van de bovenste echte limiet van de hoogste waarde van de verdeling. We komen echter tot hetzelfde antwoord als we de laagste waarde aftrekken van de hoogste waarde van de verdeling en daarbij één meeteenheid optellen. We moeten hier dus opletten welke meeteenheid de verdeling volgt. Als de verdeling gemeten is in 1 decimaal achter de komma, dan zullen we in plaats van een geheel getal 0,1 bij de berekening moeten optellen.

Heeft een verdeling een groot bereik, dan betekent dit dat de scores ver uit elkaar liggen en dat deze verdeling dus een grote spreiding bezit. Als een klein aantal x-waarden het resultaat is van de bereikberekening, dan is de spreiding minder breed.

Voorbeeld 3.6: Berekenen van het bereik

In voorbeeld 3.5 kwamen verdelingen A en B voor:

A: 11 12 13

B: 1 12 23

We zullen het bereik van verdeling A op de eerstgenoemde manier berekenen en het bereik van verdeling B op de tweede wijze bepalen.

Het bereik van verdeling A: $13{,}5 - 10{,}5 = 3$

We hebben hier de onderste echte limiet van de laagste waarde afgetrokken van de bovenste echte limiet van de hoogste waarde die in de verdeling voorkomt. We bereiken in verdeling A dus 3 verschillende x-waarden (11, 12 en 13).

Het bereik van verdeling B: $23 - 1 + 1 = 23$

Binnen verdeling B zijn dus 23 verschillende x-waarden mogelijk.

Het bereik blijkt in de praktijk echter nadelen te hebben als spreidingsmaat. Sommige verdelingen met gelijk gemiddelde en gelijk bereik kunnen verschillend zijn wat betreft de spreiding van de x-waarden ten opzichte van het gemiddelde. Dit is zeer belangrijk, want we kunnen de mate van representativiteit van de maat van de centrale tendentie afleiden uit de mate van spreiding. We zullen dit theoretische verhaal illustreren aan de hand van twee verdelingen in voorbeeld 3.7.

Voorbeeld 3.7: Verschillende spreidingen

Cijfers rekentoets klas A: 1, 2, 2, 9, 9, 10

$$\bar{x} = \frac{33}{6} = 5{,}5;\ \text{BEREIK} = 10{,}5 - 0{,}5 = 10$$

Cijfers rekentoets klas B: 1, 5, 5, 6, 6, 10

$$\bar{x} = \frac{33}{6} = 5{,}5;\ \text{BEREIK} = 10{,}5 - 0{,}5 = 10$$

We zien hier, dat zowel gemiddelde als bereik in beide verdelingen gelijk zijn, maar dat de verdelingen toch van elkaar verschillen qua spreiding. Klas A heeft vooral extreme scores en daar kunnen we het gemiddelde niet bepaald als representatieve waarde accepteren, de scores wijken er immers allemaal erg ver van af. In klas B daarentegen zijn er maar enkele extreme waarden en de rest van de scores ligt wel in de buurt van het gemiddelde. In klas B is de spreiding rond het gemiddelde dus kleiner dan in klas A.

3.2.2 *De variantie*

Uit voorbeeld 3.7 is duidelijk geworden, dat het bereik als spreidingsmaat niet echt aan onze eisen voldoet. Een andere maat voor de spreiding is de variantie. Deze maat bepaalt grof gezegd, de *gemiddelde gekwadrateerde afwijking van de scores ten opzichte van het gemiddelde*. Deze maat wordt genoteerd als S^2. Als de variantie groot is, dan betekent dit dat de scores relatief ver van het gemiddelde liggen; dit maakt het gemiddelde iets minder representatief. Is de variantie wat kleiner, dan betekent dit dat de scores dicht bij het gemiddelde liggen; het gemiddelde is dan een goede representatieve maat voor de verdeling. Nu rijst de vraag waarom de variantie de gemiddelde *gekwadrateerde* afstand van het gemiddelde weergeeft en niet de gewone gemiddelde afstand. Het antwoord op deze vraag is niet simpel, maar wel te beantwoorden. De variantie moet gezien worden als een rekenkundige, abstracte maat; in veel formules moeten we de variantie invullen.

De gemiddelde afstand van het gemiddelde komt in paragraaf 3.2.3. aan de orde.

De berekening van de variantie kan op twee manieren gebeuren: met behulp van de definitieformule en met behulp van de rekenformule. In wezen is het altijd het meest verstandig om gebruik te maken van de rekenformule. Toch zullen we in het volgende voorbeeld de variantie eenmalig berekenen met behulp van de definitieformule, zodat duidelijk is wat de variantieberekening eigenlijk inhoudt en dat we werkelijk de gekwadrateerde afwijkingen berekenen (voorbeeld 3.8).

Definitieformule: $S_x^2 = \dfrac{\sum (x_i - \bar{x})^2}{n - 1}$

Rekenformule I: $S_x^2 = \dfrac{n \cdot \sum x^2 - (\sum x)^2}{n(n-1)}$

Rekenformule II: $S_x^2 = \dfrac{n \cdot \sum fx^2 - (\sum fx)^2}{n(n-1)}$

We moeten ook hier, net als bij het gemiddelde, opletten of we te maken hebben met de ruwe scores of met een frequentieverdeling van de scores. Indien we te maken hebben met de ruwe scorelijst, dan kunnen we volstaan met rekenformule I (voorbeeld 3.9). Moeten we de variantie van de frequentieverdeling berekenen, dan zullen we gebruik moeten maken van rekenformule II (voorbeeld 3.10).

3.2.3 *De standaardafwijking*

Deze spreidingsmaat wordt genoteerd als *S*. Aangezien de variantie weergegeven wordt door S^2 en de standaardafwijking door *S*, is de standaardafwijking te berekenen door de wortel te nemen van de variantie, dus $S = \sqrt{S^2}$.
De standaardafwijking of standaarddeviatie is een meer concreet te interpreteren maat dan de variantie: de maat is globaal op te vatten als de gemiddelde afwijking van de scores ten opzichte van het gemiddelde. De definitie lijkt erg op die van de variantie, alleen is het woord 'gekwadrateerd' nu weggelaten. Dit komt weer rekenkundig overeen: om S te berekenen, moeten we immers de wortel uit S^2 nemen, waardoor het kwadraat geëlimineerd wordt. We zullen in de voorbeelden 3.8, 3.9 en 3.10 naast de variantie meteen de standaardafwijking bepalen.
Als we de standaardafwijking direct willen berekenen, dan kunnen we gebruikmaken van de volgende formules:

Definitieformule: $S_x = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n - 1}}$

Rekenformule I: $S_x = \sqrt{\frac{n \cdot \sum x^2 - (\sum x)^2}{n(n-1)}}$

Rekenformule II: $S_x = \sqrt{\frac{n \cdot \sum fx^2 - (\sum fx)^2}{n(n-1)}}$

Voorbeeld 3.8: Variantie en standaardafwijking (definitieformule)
We gaan de variantie van de cijfers in klas A berekenen:

x	$\bar{x}$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
1	5,5	$1 - 5{,}5 = -4{,}5$	$(-4{,}5)^2 = 20{,}25$
2	5,5	$2 - 5{,}5 = -3{,}5$	$(-3{,}5)^2 = 12{,}25$
2	5,5	$2 - 5{,}5 = -3{,}5$	$(-3{,}5)^2 = 12{,}25$
9	5,5	$9 - 5{,}5 = 3{,}5$	$(3{,}5)^2 = 12{,}25$
9	5,5	$9 - 5{,}5 = 3{,}5$	$(3{,}5)^2 = 12{,}25$
10	5,5	$10 - 5{,}5 = 4{,}5$	$(4{,}5)^2 = 20{,}25$
		$\sum (x_i - \bar{x}) = 0$!	$\sum (x_i - \bar{x})^2 = 89{,}50$

$$S_x^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1} = \frac{89{,}50}{6 - 1} = 17{,}9$$

We zien dat we van elke score het gemiddelde hebben afgetrokken en de uitkomsten gekwadrateerd hebben. We kwadrateren omdat als we de gewone afwijkingen ten opzichte van het gemiddelde optellen, we op 0 uitkomen, terwijl er wel degelijk afwijking ten opzichte van het gemiddelde is. De oorzaak daarvan ligt in de mingetallen. Om deze te elimineren, kwadrateren we de afwijkingen. Dit noemen we de *gekwadrateerde afwijkingen van het gemiddelde.* Vervolgens telden we de uitkomsten op; deze uitkomst noemen we de *opgetelde gekwadrateerde afwijkingen van het gemiddelde.*

Daarna deelden we deze uitkomsten weer door het totaal aantal mensen − 1. De uiteindelijke uitkomst is de variantie en dus een soort gemiddelde gekwadrateerde afwijking van het gemiddelde.

Willen we van deze scoreverdeling de standaardafwijking weten, dan nemen we de wortel uit 17,9.

$S_x = \sqrt{17{,}9} = 4{,}23$

De gemiddelde afwijking van de cijfers uit klas A ten opzichte van het gemiddelde is gelijk aan 4,23.

Voorbeeld 3.9: Variantie en standaardafwijking (rekenformule I)

De spreiding van de cijfers van klas B zullen we berekenen door middel van rekenformule I; we gaan dus uit van de ruwe scores.

x	x^2
10	100
6	36
6	36
5	25
5	25
1	1
$\sum x = 33$	$\sum x^2 = 223$

$$S_x^2 = \frac{n \cdot \sum x^2 - (\sum x)^2}{n(n-1)} = \frac{6 \cdot 223 - (33)^2}{6 \cdot (6-1)} = \frac{1338 - 1089}{30} = \frac{249}{30} = 8{,}3$$

We zien dat we via een minder omslachtige wijze dan in voorbeeld 3.8 tot een variantie van 8,3 komen. Ook is hiermee een bevestiging geleverd voor het vermoeden uit voorbeeld 3.7, dat de spreiding in klas A groter was dan de spreiding van klas B.

De standaardafwijking van de cijfers in klas B verkrijgen we door weer de wortel te trekken uit de variantie:

$S_x = \sqrt{8{,}3} = 2{,}88$

De cijfers uit klas B wijken dus gemiddeld met een waarde van 2,88 af van het gemiddelde.

Voorbeeld 3.10: Variantie en standaardafwijking (rekenformule II)

Om het gebruik van rekenformule II te illustreren, zullen we in dit voorbeeld uitgaan van de frequentieverdeling van klas A.

x	f	xf	x^2	fx^2
10	1	10	100	100
9	2	18	81	162
2	2	4	4	8
1	1	1	1	1
		$\sum xf = 33$		$\sum fx^2 = 271$

$$S_x^2 = \frac{n \cdot \sum fx^2 - (\sum fx)^2}{n(n-1)} = \frac{6 \cdot 271 - (33)^2}{6 \cdot (6-1)} = \frac{1626 - 1089}{30} = \frac{537}{30} = 17{,}9$$

$$S_x = \sqrt{17{,}9} = 4{,}23$$

Over het algemeen gebruikt men altijd de rekenformules om tot de variantie en standaardafwijking te komen, omdat ze beter hanteerbaar zijn voor grote groepen scores.

Het is belangrijk te weten dat we de variantie en de standaarddeviatie bepalen op basis van het gemiddelde. We kunnen deze spreidingsmaten dan ook alleen gebruiken in het geval van een interval- of een ratiomeetniveau.

3.3 Kenmerken van een gegroepeerde frequentieverdeling

In bovenstaande voorbeelden van dit hoofdstuk hebben we telkens te maken gehad met niet-gegroepeerde frequentieverdelingen. In de praktijk komen we echter meestal met gegroepeerde frequentieverdelingen in aanraking. Ter illustratie zullen we een gegroepeerde frequentieverdeling karakteriseren op centrale tendentie en variabiliteit. Dit zien we in voorbeeld 3.11.

Voorbeeld 3.11: Karakteriseren van gegroepeerde verdeling

Uit een fictief onderzoek naar het bedrag dat studenten uitgeven aan vakanties per jaar, zijn de volgende resultaten (uitgedrukt in guldens) afkomstig (n = 50):

2250	1568	1608	1998	1050	1460	1185
800	1249	502	0	320	2405	1710
1805	1975	1630	2077	2850	1599	1749
1750	1900	1399	1250	1312	1499	1116
1399	1200	1099	1485	1335	1008	873
498	899	699	750	275	250	120
999	375	99	700	749	575	599
605						

We kunnen van deze bedragen een gegroepeerde frequentieverdeling maken met een klassenbreedte van 500 gulden:

klasse	*x*	*f*	*xf*	*cum.f*
2500-2999	2749,5	1	2749,5 · 1 = 2749,5	50
2000-2499	2249,5	3	2249,5 · 3 = 6748,5	49
1500-1999	1749,5	11	1749,5 · 11 = 19244,5	46
1000-1499	1249,5	15	1249,5 · 15 = 18742,5	35
500-999	749,5	12	749,5 · 12 = 8994	20
0-499	249,5	8	249,5 · 8 = 1996	8
		n = 50	$\sum$ xf = 58475	

Het gemiddelde van deze verdeling vinden we als volgt:

$$\frac{\sum xf}{n} = \frac{58475}{50} = 1169{,}5$$

Gemiddeld geven de studenten dus 1169,5 gulden per jaar uit aan vakanties.

De mediaan van deze verdeling, het punt waar de verdeling precies gehalveerd wordt, kunnen we berekenen met de volgende formule:

$$Md = L + \left[\frac{p(n) - n_b}{n_w}\right] \cdot i$$

Allereerst zoeken we op, waar de mediaan ongeveer zal vallen, dat is bij de helft van het aantal mensen, dus 50 · (0,5) = 25.
Bij de 25ste persoon moet de mediaan liggen. In de cumulatieve kolom gaan we kijken waar deze persoon valt en welke score daarbij hoort. We zien dat die persoon in klasse 1000-1499 valt. Met behulp van dit interval gaan we verder de formule invullen:

$$L = 999{,}5 \rightarrow Md = 999{,}5 + \left[\frac{25 - 20}{15}\right] \cdot 500 = 1166{,}17$$

$n_b = 20$
$n_w = 15$
$i = 500$

Bij het bedrag van 1166,17 wordt de verdeling precies gehalveerd, dus 50% van de mensen heeft meer dan dit bedrag uitgegeven aan hun vakanties en 50% heeft minder dan dit bedrag aan vakanties besteed.
De modus is de meest voorkomende waarde, het bedrag dat het meest aan vakanties is uitgegeven. De klasse 1000-1499 heeft de hoogste frequentie, de modus is dan gelijk aan het klassenmidden van dit interval, dus 1249,5.

Wat we hier zien is, dat het gemiddelde en de mediaan een heel klein beetje van elkaar verschillen; de verdeling is dus niet volledig symmetrisch. Waarschijnlijk is de oorzaak hiervan dat er in verhouding meer hoge dan lage extreme scores zijn. Het gemiddelde trekt daar iets naar toe, terwijl de mediaan zich daar niets van aantrekt. Vandaar dat het gemiddelde hier iets groter is dan de mediaan; deze verdeling is dan scheef naar rechts. Is het gemiddelde kleiner dan de mediaan, dan is de verdeling scheef naar links.

De spreiding van de bovenstaande verdeling zullen we nu uitrekenen met behulp van het bereik, de variantie en de daaruit volgende standaardafwijking.

Als we het bereik van een gegroepeerde frequentieverdeling willen bepalen, hebben we eigelijk de ruwe scores nodig. Daaraan kunnen we immers zien wat de hoogste en laagste scores waren. In dit geval beschikken we over deze ruwe scores en kunnen we het bereik bepalen: hoogste score = 2850, laagste score = 0; bereik = 2850,5 − − 0,5 = 2851.
Indien we de oorspronkelijke ruwe scores niet gegeven krijgen, dan bepalen we het bereik als volgt: hoogste echte limiet van de hoogste waarde in de bovenste klasse min de laagste echte limiet van de laagste waarde in de onderste klasse.

We gaan nu de variantie berekenen.

$$S_x^2 = \frac{n \cdot \sum fx^2 - (\sum fx)^2}{n(n-1)} =$$

Om deze formule volledig te kunnen invullen, moeten we $\sum fx^2$ nog uitrekenen; $\sum fx$ hadden we immers al berekend om het gemiddelde te kunnen vaststellen.

x	f	x^2	fx^2
2749,5	1	$(2749,5)^2 = 7559750,3$	$1 \cdot 7559750,3 = 7559750,3$
2249,5	3	$(2249,5)^2 = 5060250,3$	$3 \cdot 5060250,3 = 15180751$
1749,5	11	$(1749,5)^2 = 3060750,3$	$11 \cdot 3060750,3 = 33668253$
1249,5	15	$(1249,5)^2 = 1561250,3$	$15 \cdot 1561250,3 = 23418755$
749,5	12	$(749,5)^2 = 561750,25$	$12 \cdot 561750,25 = 6741003$
249,5	8	$(249,5)^2 = 62250,25$	$8 \cdot 62250,25 = 498002$
			$\sum fx^2 = 87066514$

$$S_x^2 = \frac{n \cdot \sum fx^2 - (\sum fx)^2}{n(n-1)} = \frac{50 \cdot 87066514 - (58475)^2}{50 \cdot 49} = 381224,52$$

$$S_x = \sqrt{381224,52} = 617,43$$

De studenten wijken dus gemiddeld 617,43 gulden af van het gemiddelde van 1169,5.

Opgaven

3.1 Onder studenten pedagogiek van het hbo is een fictief onderzoek gedaan naar het aantal uren per week dat aan de studie besteed wordt. Totaal deden er 75 mensen mee aan dit onderzoek. Hieronder staan de gegevens in uren.

$10\frac{1}{2}$	22	8	$19\frac{1}{2}$	17	$12\frac{1}{2}$	$10\frac{1}{2}$	9	15
14	13	0	$12\frac{1}{2}$	$10\frac{1}{2}$	24	17	$16\frac{1}{2}$	10
12	15	24	$15\frac{1}{2}$	6	$4\frac{1}{2}$	$3\frac{1}{2}$	31	15
12	12	12	12	12	$13\frac{1}{2}$	$16\frac{1}{2}$	$1\frac{1}{2}$	$15\frac{1}{2}$
14	11	17	7	$16\frac{1}{2}$	4	14	23	$18\frac{1}{2}$
$13\frac{1}{2}$	15	9	13	21	$12\frac{1}{2}$	$6\frac{1}{2}$	7	14
$15\frac{1}{2}$	$15\frac{1}{2}$	$7\frac{1}{2}$	15	28	13	11	$12\frac{1}{2}$	18
19	14	$10\frac{1}{2}$	20	$13\frac{1}{2}$	8	$15\frac{1}{2}$	$9\frac{1}{2}$	15
$16\frac{1}{2}$	16	16						

a Maak hiervan een gegroepeerde frequentieverdeling met een klassenbreedte van 3.
b Bereken het gemiddelde
c Bereken de mediaan
d Bereken de modus
e Bereken het bereik
f Bereken de variantie
g Bereken de standaardafwijking

3.2 Ook onder de studenten die pedagogiek aan de universiteit studeren, hebben we een fictief onderzoek gedaan naar het aantal uren dat thuis per week besteed wordt aan de studie. Totaal onderzochten we weer 75 mensen, hieronder staan de resultaten.

$18\frac{1}{2}$	$16\frac{1}{2}$	$12\frac{1}{2}$	21	15	10	17	12	12
$19\frac{1}{2}$	13	9	10	$9\frac{1}{2}$	15	25	13	16
$10\frac{1}{2}$	$22\frac{1}{2}$	14	$13\frac{1}{2}$	$9\frac{1}{2}$	15	14	11	13
$10\frac{1}{2}$	10	17	10	$12\frac{1}{2}$	20	9	10	19
$10\frac{1}{2}$	$21\frac{1}{2}$	9	$9\frac{1}{2}$	$15\frac{1}{2}$	16	14	12	$13\frac{1}{2}$
$10\frac{1}{2}$	10	11	13	11	$16\frac{1}{2}$	$19\frac{1}{2}$	13	11
$12\frac{1}{2}$	11	$10\frac{1}{2}$	13	$9\frac{1}{2}$	11	16	10	20
14	23	13	12	$10\frac{1}{2}$	$9\frac{1}{2}$	14	$15\frac{1}{2}$	$12\frac{1}{2}$
$19\frac{1}{2}$	19	$16\frac{1}{2}$						

a Maak hiervan een gegroepeerde frequentieverdeling met gelijke klasseninindeling als bij opgave 1.
b Bereken het gemiddelde
c Bereken de mediaan
d Bereken de modus
e Bereken het bereik

f Bereken de variantie
g Bereken de standaardafwijking

3.3 *a* Bekijk de verdelingen uit opgave 1 en 2 goed en vergelijk ze op centrale tendentie (scheef naar links/rechts).
b Vergelijk de verdelingen op variabiliteit (gepiekt/plat)

3.4 *a* Bereken van tabel 2.2 uit hoofdstuk 2 alle maten van de centrale tendentie (deze opgave heeft een beperkte uitwerking).
b Bereken van tabel 2.2 uit hoofdstuk 2 alle maten van de variabiliteit (deze opgave heeft een beperkte uitwerking).
c Geef aan of deze verdeling symmetrisch, scheef naar links of scheef naar rechts is.

Uitwerkingen

3.1 *a*

klasse	*x*	*f*	*cf*	*fx*	x^2	fx^2
30-32	31	1	75	31	961	961
27-29	28	1	74	28	784	784
24-26	25	2	73	50	625	1250
21-23	22	3	71	66	484	1452
18-20	19	5	68	95	361	1805
15-17	16	20	63	320	256	5120
12-14	13	21	43	273	169	3549
9-11	10	10	22	100	100	1000
6-8	7	7	12	49	49	343
3-5	4	3	5	12	16	48
0-2	1	2	2	2	1	2
		n = 75		$\sum$ fx = 1026		$\sum$ fx² = 16314

b Het gemiddelde van deze verdeling vinden we door de volgende formule toe te passen:

$$\frac{\sum fx}{n} = \frac{1026}{75} = 13{,}68$$

Dus gemiddeld besteden de hbo-studenten thuis 13,7 uur aan de studie.

c De mediaan vinden we door de volgende formule te gebruiken.
Eerst zoeken we het punt op waar de mediaan ongeveer zal vallen, daartoe berekenen we $75 \cdot (0{,}5) = 37{,}5$. Dit getal zoeken we op in de cumulatieve kolom (cf). We zien dan dat deze persoon valt in de klasse 12-14; uit deze klasse leiden we de formulewaarden af.

$$Md = L + \left[\frac{p(n) - n_b}{n_w}\right] \cdot i = 11{,}5 + \left[\frac{37{,}5 - 22}{21}\right] \cdot 3 = 13{,}71$$

d De modus is de waarde in de verdeling die het meest voorkomt, dus het aantal uren thuisstudie dat de meeste hbo-ers volmaken. In de frequentietabel zien we dat de hoogste frequentie is gescoord in klasse 12-14; de modus is dan gelijk aan het klassenmidden 13.

e Het bereik van een verdeling kan men vinden door de onderste echte limiet van de laagste score af te trekken van de bovenste limiet van de hoogste score:
dus: $31{,}5 - -0{,}5 = 32$

Dit betekent dat in deze verdeling totaal 32 verschillende x-waarden mogelijk zijn.

f De variantie berekenen we door de volgende formule in te vullen:

$$S_x^2 = \frac{n \cdot \sum fx^2 - (\sum fx)^2}{n(n-1)} = \frac{75 \cdot 16314 - (1026)^2}{75 \cdot 74} = \frac{170874}{5550} = 30{,}79$$

g Om de standaardafwijking te kunnen vinden, moeten we de wortel trekken uit de uitkomst van de variantie, dus:
$S_x = \sqrt{S^2} = \sqrt{30{,}79} = 5{,}55$

3.2 *a*

klasse	*x*	*f*	*cf*	*fx*	x^2	fx^2
27-29	28	0	75	0	784	0
24-26	25	1	75	25	625	625
21-23	22	4	74	88	484	1936
18-20	19	8	70	152	361	2888
15-17	16	13	62	208	256	3328
12-14	13	22	49	286	169	3718
9-11	10	27	27	270	100	2700
6-8	7	0	0	0	49	0
3-5	4	0	0	0	16	0
		n = 75		$\sum fx = 1029$		$\sum fx^2 = 15195$

b $\frac{\sum fx}{n} = \frac{1029}{75} \approx 13{,}7$

Gemiddeld besteden de universtaire studenten pedagogiek dus evenveel tijd aan de studie thuis als de hbo-ers.

c $Md = L + \left[\frac{p(n) - n_b}{n_w}\right] \cdot i = 11{,}5 + \left[\frac{37{,}5 - 27}{22}\right] \cdot 3 = 12{,}93$

De verdeling van thuiswerkuren wordt precies gehalveerd bij 12,93 uur.

d De modus van deze verdeling is gelijk aan het klassenmidden van de klasse met de hoogste frequentie: 10

e Het bereik van deze verdeling is gelijk aan:
25,5 − 8,5 = 17

f De variantie van deze verdeling wordt als volgt bepaald:

$$S_x^2 = \frac{n \cdot \sum fx^2 - (\sum fx)^2}{n(n-1)} = \frac{75 \cdot 15195 - (1029)^2}{75 \cdot 74} = \frac{80784}{5550} = 14{,}56$$

g $S_x = \sqrt{S^2} = \sqrt{14{,}56} = 3{,}82$

3.3 *a* We zien dat het gemiddelde en de mediaan van de hbo-verdeling gelijk zijn aan elkaar. Hier is dus sprake van een symmetrische verdeling. Bij de universiteitsverdeling zien we dat de mediaan kleiner is dan het gemiddelde.

Deze verdeling is dus scheef naar rechts. Dit houdt in dat er in verhouding iets meer hoge extreme scores zijn, waar het gemiddelde naar toe trekt.

b De spreiding in de hbo-verdeling is groter dan de spreiding in de verdeling van de universitaire studenten. We noemen de hbo-verdeling dan platter dan de universiteitsverdeling, of de universiteitsverdeling gepiekter.

3.4 *a* Modus = 32
Mediaan = 32,5
$\bar{x} = 34{,}4$

b bereik = 55
$S_x^2 = 142{,}08$
$S_x = 11{,}92$

c Deze verdeling is scheef naar rechts.

4 De relatieve positie van gegevens

Nu we in staat zijn verdelingen te beschrijven, in kaart te brengen en te karakteriseren op essentiële kenmerken, gaan we de relatieve positie van gegevens vaststellen. Dit is handig, omdat we dan dezelfde scores, voorkomend in verschillende verdelingen, kunnen vergelijken. Bijvoorbeeld: iemand in klas x heeft voor wiskunde een 8 gehaald en zijn vriend uit klas y heeft voor het vak aardrijkskunde ook een 8 gehaald. In eerste instantie zouden we zeggen dat deze twee vrienden even hoog gescoord hebben, maar het maakt natuurlijk wel uit, welke positie het cijfer 8 heeft ten opzichte van de rest van de klas. Als het gemiddelde wiskundecijfer in klas x een 9 is en het gemiddelde voor aardrijkskunde in klas y een 6 is, dan zal iedereen het er over eens zijn, dat het cijfer 8 voor aardrijkskunde een relatief betere prestatie is dan een 8 voor wiskunde.
We kunnen ons ook afvragen hoeveel procent van de leerlingen lager dan een 5 heeft gehaald of zoeken naar het cijfer dat correspondeert met de laagste 25% van de mensen. Kortom: we bepalen de relatieve positie van gegevens.
De twee technieken die in dit hoofdstuk behandeld worden, kunnen beide antwoord geven op de volgende twee vragen:

1 Hoeveel procent van de mensen valt op en onder een bepaald punt(cijfer)?
  We zoeken dan een percentage bij een bepaald punt.
2 Tot en met welk punt (cijfer) valt een bepaald percentage?
  We zoeken dan een punt bij een bepaald percentage.

We kunnen op twee manieren de relatieve plaats van gegevens vaststellen: met behulp van de percentielberekening en met behulp van de standaardnormaalverdeling. Beide technieken drukken de relatieve plaats van gegevens uit in het percentage van de scores of mensen dat op en onder een gegeven score valt binnen de verdeling.

4.1 De percentielberekening

4.1.1 *Het percentielpunt*

In deze berekening wordt antwoord gegeven op de de tweede vraag uit de inleiding:
Tot en met welk punt valt een gegeven percentage?
Deze berekening wordt op gelijke wijze uitgevoerd als de mediaan en kan in principe in elke verdeling worden toegepast. Bij de mediaanberekening zoeken we dát punt, die x-waarde, *waarop of waaronder* 50 procent van alle mensen valt. Zo is het ook mogelijk om dat punt in de verdeling op te zoeken waaronder 30% van de mensen valt, of welk ander percentage dan ook.
De formule voor deze berekening is exact gelijk aan de formule van de mediaan. Voor de p in de formule vullen we dan steeds het percentage in dat gegeven wordt, en daar zoeken we een punt bij.

$$P_{punt} = L + \left[\frac{p(n) - n_b}{n_w}\right] \cdot i$$

We zullen het gebruik van deze formule aan de hand van een voorbeeld illustreren.

Voorbeeld 4.1: Het berekenen van een percentielpunt
In klas x zitten 17 kinderen en voor een proefwerk wiskunde halen zij gemiddeld een 9. We willen weten tot en met welk cijfer de laagste 30% van de kinderen scoort. In klas y haalt men gemiddeld een 6 voor een proefwerk aardrijkskunde, ook hier willen we weten tot en met welk cijfer de laagste 30% van de kinderen scoort (n = 15).

Tabel 4.1

Wiskundecijfers	*f*	*Cumulatieve frequentie*
10	8	17
9	3	9
8 ←	4	→ 6
7	2	2
	n = 17	

Voor de wiskundeklas gaat de berekening als volgt. Allereerst gaan we bepalen in welke klasse het 30%-punt ongeveer zal liggen. We voeren daarom eerst p(n) uit. Voor de p-waarde vullen we 0,30 in, en voor n vullen we 17 in → p(n) = 0,30 · 17 = 5,1.
Dus tot en met de 5,1ste persoon valt 30% van de verdeling.
In de cumulatieve frequentiekolom kijken we, welk cijfer deze persoon heeft gehaald. Dat is het cijfer 8, want tot en met dit cijfer hebben 6 mensen gescoord en daar zit dus ook de 5,1ste persoon in. Nu kunnen we de gegevens voor de formule bepalen en invullen:

$$P_{punt} = L + \left[\frac{p(n) - n_b}{n_w}\right] \cdot i = 7{,}5 + \left[\frac{5{,}1 - 2}{4}\right] \cdot 1 = 7{,}5 + \frac{3{,}1}{4} = 8{,}28$$

Hieruit kunnen we afleiden, dat de onderste 30% van de leerlingen een cijfer van 8,28 of lager hebben gehaald.

Ook voor de aardrijkskundeklas zullen we het percentielpunt behorende bij 30% uitvoeren (zie tabel 4.2).

p(n) = 0,30 · 15 = 4,5

Uit de cumulatieve frequentietabel lezen we af dat het bijbehorende cijfer 5 is. We vullen nu de formule in:

$$P_{punt} = L + \left[\frac{p(n) - n_b}{n_w}\right] \cdot i = 4{,}5 + \left[\frac{4{,}5 - 2}{3}\right] \cdot 1 = 4{,}5 + \frac{2{,}5}{3} = 5{,}33$$

Tabel 4.2

Aardrijkskundecijfers	*f*	*Cumulatieve frequentie*
8	2	15
7	3	13
6	5	10
5 ←	3	→ 5
4	2	2
	n = 15	

Uit deze uitkomst kunnen we afleiden dat de laagste 30% van de cijfers op en onder het cijfer 5,33 ligt.
Als we de percentielpunten van de wiskunde- en de aardrijkskundeklassen vergelijken, dan zien we dat de onderste 30% van de leerlingen in de wiskundeklas relatief hoger scoort dan in de aardrijkskundeklas.

4.1.2 *De percentielrang*

Deze berekening geeft vooral antwoord op de eerste vraag uit de inleiding, namelijk: hoeveel procent van de mensen valt op of onder een bepaald punt (cijfer)? We zoeken dan een percentage, een rang, bij een bepaald punt.
De formule die we nodig hebben om deze vraag te kunnen beantwoorden, ziet er als volgt uit:

$$P_{rang} = \frac{n_w \cdot (X - L) + i \cdot n_b}{n \cdot i}$$

We zullen de onderdelen van de formule hieronder verduidelijken.

n_w = aantal mensen dat in de klasse van X valt
X = de score waarbij we een percentage zoeken
L = de lage ondergrens van de klasse waar de score (X) in valt
i = de klassenbreedte
n_b = het aantal mensen dat lager dan de klasse van X heeft gescoord
n = het totaal aantal mensen in de verdeling

In voorbeeld 4.2 illustreren we het gebruik van deze formule.

Voorbeeld 4.2: Het berekenen van een percentielrang
In dit voorbeeld gaan we weer terug naar de wiskunde- en aardrijkskundeklas uit voorbeeld 1. Stel dat Emile een 8 heeft gescoord voor wiskunde en Thomas een 8 voor aardrijkskunde, wie heeft er relatief beter gescoord?
Deze vraag kunnen we beantwoorden door middel van de percentielrangberekening. Deze berekening geeft namelijk als uitkomst een percentage dat zich op of onder de score 8 bevindt. Hoe hoger dit percentage, hoe beter de score relatief is.
We zullen eerst het percentage bepalen dat zich onder de score van Emile bevindt.
We maken gebruik van tabel 4.1 uit voorbeeld 4.1.

De score waarbij we een percentage zoeken (X) is 8. Nu kunnen we op de gebruikelijke manier de gegevens uit de tabel aflezen.

$n_w = 4$ $\quad i = 1$

$X = 8$ $\quad n_b = 2$

$L = 7{,}5$ $\quad n = 17$

$$P_{rang} = \frac{n_w \cdot (X - L) + i \cdot n_b}{n \cdot i}$$

$$P_{rang} = \frac{4 \cdot (8 - 7{,}5) + 1 \cdot 2}{17 \cdot 1}$$

$$P_{rang} = \frac{4 \cdot 0{,}5 + 2}{17} = \frac{4}{17} = 0{,}2353$$

Willen we deze proportie uitgedrukt zien in een percentage, dan moeten we vermenigvuldigen met 100%, dus $0{,}2353 \cdot 100\% = 23{,}53\%$.
Bij de score 8 van Emile hoort 23,53%, dus 23,53% van de leerlingen heeft een score van 8 of lager gehaald.

Ook bij het cijfer van Thomas gaan we een percentage zoeken. Hierbij maken we gebruik van tabel 4.2 van de aardrijkskundecijfers.

$n_w = 2$ $\quad i = 1$

$X = 8$ $\quad n_b = 13$

$L = 7{,}5$ $\quad n = 15$

$$P_{rang} = \frac{n_w \cdot (X - L) + i \cdot n_b}{n \cdot i}$$

$$P_{rang} = \frac{2 \cdot (8 - 7{,}5) + 1 \cdot 13}{15 \cdot 1}$$

$$P_{rang} = \frac{2 \cdot 0{,}5 + 13}{15} = \frac{14}{15} = 0{,}9333$$

In de aardrijkskundeklas heeft dus 93,33% van de leerlingen een 8 of lager.
Als we de percentages van Emile en Thomas vergelijken, dan zien we dat Thomas relatief een hoger percentage van de leerlingen onder zich heeft dan Emile, relatief gezien heeft hij dus een betere prestatie geleverd.

4.2 De standaardnormaalverdeling

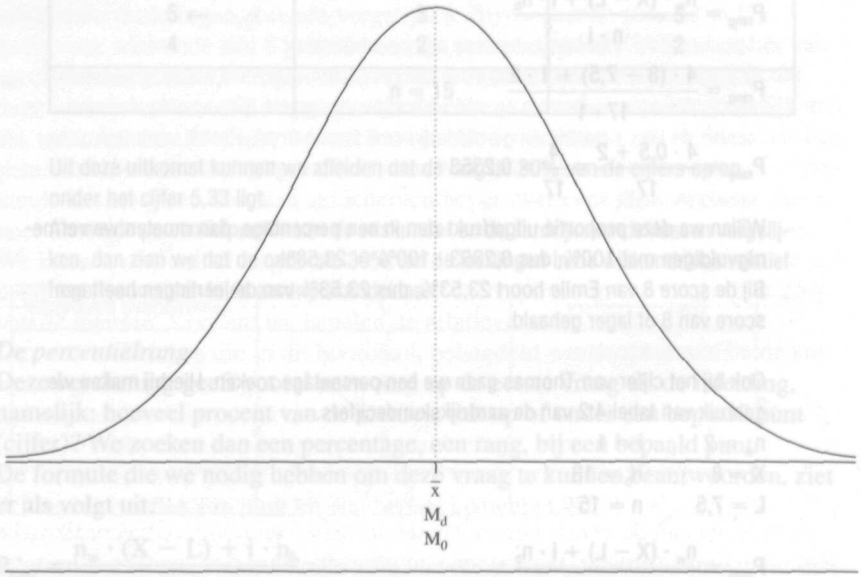
Zoals al in de inleiding van dit hoofdstuk is gezegd, is er nóg een mogelijkheid om de relatieve positie van gegevens te bepalen. Deze methode maakt gebruik van de z-transformatie naar de standaardnormaalverdeling.
De standaardnormale verdeling is een verdeling die als gemiddelde altijd 0 heeft en als standaardafwijking altijd 1. Daarnaast heeft deze verdeling alle kenmerken van een normaalverdeling, namelijk:

1 Modus = mediaan = gemiddelde. Dit wil zeggen dat de drie centrummaten precies dezelfde waarde aannemen.
2 De normaalverdeling is symmetrisch en kan dus gespiegeld worden.

3 De normaalverdeling is klokvormig.

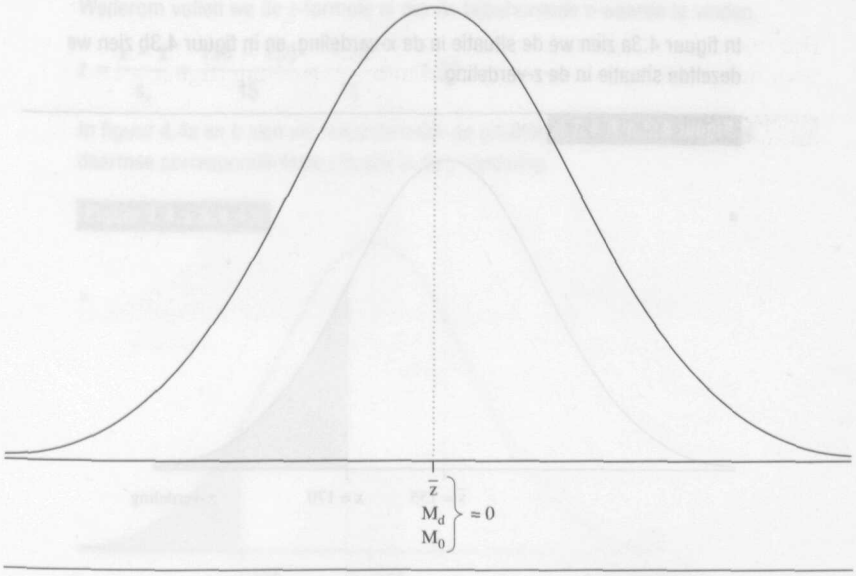
Hieronder staat de afbeelding van een normaalverdeling (figuur 4.1), gevolgd door de afbeelding van een standaardnormaalverdeling (figuur 4.2).

Figuur 4.1 Normaalverdeling



De gedachte achter deze techniek is dat iedere normale verdeling die de drie bovenstaande kenmerken heeft of ze benadert, getransformeerd kan worden naar een soort standaardverdeling, de standaardnormale verdeling dus. Deze verdeling wordt ook wel de z-verdeling genoemd. Van de z-verdeling zijn alle percentages behorend bij verschillende z-scores al bekend en uitgezet in tabel A. We kunnen dus bij elke z-waarde een percentage vinden en bij elk percentage een z-waarde.
Het enige wat ons dus bij deze techniek te doen staat, is onze x-waarde, waarbij we een percentage zoeken, om te zetten in een z-waarde (z-transformatie) en dan vinden we direct een percentage zonder al te veel rekenwerk. Hoe dit in z'n werk gaat, komt aan de orde in subparagraaf 4.2.1.
Als we bij een gegeven percentage een x-waarde zoeken, dan hoeven we slechts de juiste z-waarde bij het percentage te zoeken in de tabel en deze z-waarde om te zetten in een x-waarde via de omgekeerde z-transformatie. Deze techniek komt aan de orde in subparagraaf 4.2.2.

Figuur 4.2 Standaard-normaalverdeling



4.2.1 *De z-transformatie*

Om antwoord te kunnen vinden op vraag 1 uit de inleiding kunnen we, indien we voldoen aan de voorwaarden van de normale verdeling, ook gebruikmaken van de z-transformatie. We zoeken een percentage bij een gegeven x-waarde. Om gebruik van tabel A te kunnen maken, moeten we de gegeven x-waarde eerst transformeren naar een z-waarde. Deze z-transformatie gaat als volgt:

$$z = \frac{x - \bar{x}}{s_x}$$

Voor de x in de formule vullen we de gegeven x-waarde in, verder vullen we de verdelingskenmerken $\bar{x}$ en s_x in, en dan krijgen we een z-waarde waarmee we naar de tabel kunnen gaan om het juiste percentage te vinden.
In voorbeeld 4.3 zullen we het bovenstaande nader verduidelijken.

Voorbeeld 4.3: Gebruik van de z-transformatie

Stel dat we een steekproef nemen van meisjes van 13 jaar en we onderzoeken ze op hun lengte. Totaal onderzoeken we 150 Nederlandse meisjes en hun gemiddelde lengte is 1 meter en 55 centimeter, met een standaarafwijking van 15 centimeter. Stel dat de verdeling redelijk normaal van vorm is.

a Hoeveel procent van de meisjes is langer dan 1,70 meter?

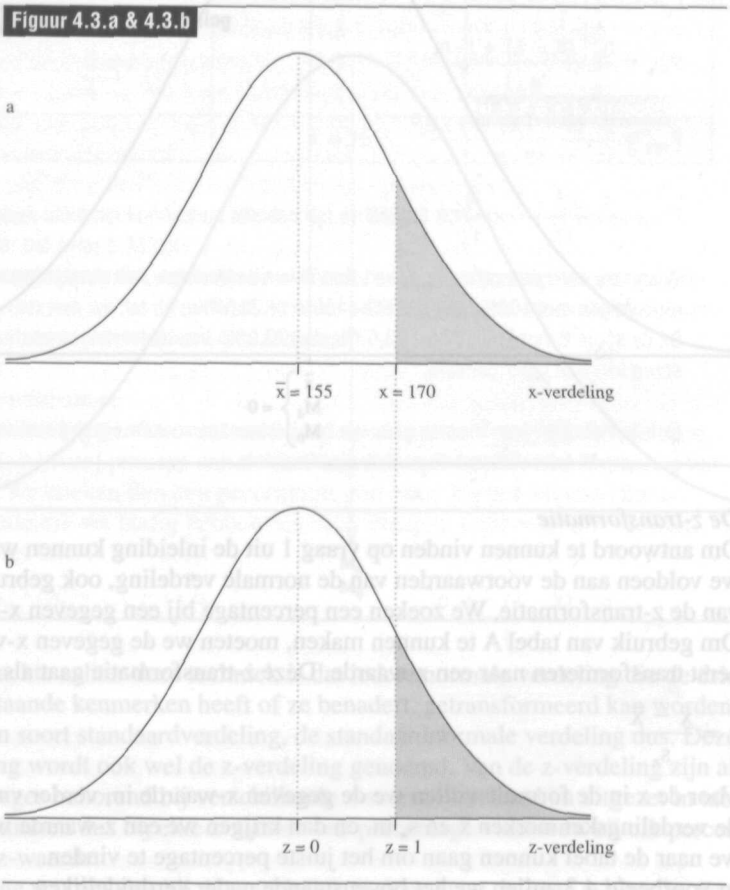
Om deze vraag te kunnen beantwoorden, moeten we de lengte van 1,70 (= de x-waarde) transformeren naar een z-waarde. Let wel op dat alles in meters moet worden uitgedrukt, of in centimeters. We hebben het hier in centimeters gedaan.

$$z = \frac{x - \bar{x}}{s_x} = \frac{170 - 155}{15} = \frac{15}{15} = 1$$

De z-waarde van 170 cm is dus 1.

In figuur 4.3a zien we de situatie in de x-verdeling, en in figuur 4.3b zien we dezelfde situatie in de z-verdeling.

Figuur 4.3.a & 4.3.b



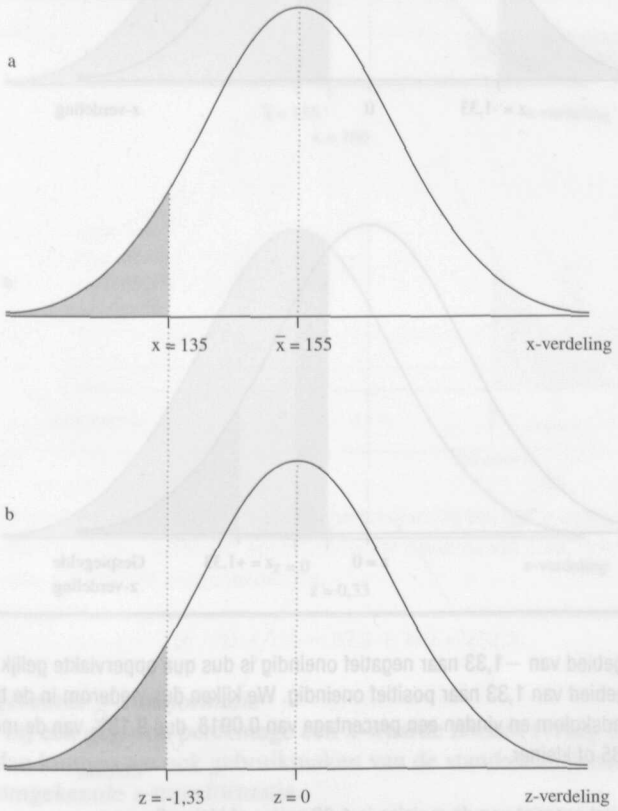
We willen het percentage weten dat in het gearceerde gebied ligt. Hiervoor gaan we naar tabel A. Hier zien we een hele rij z-waarden met daarbij steeds twee gebiedspercentages. De eerste kolom met gebiedspercentages geeft het percentage van de z-waarde naar 0 (= het gemiddelde in de x-verdeling) en de tweede kolom geeft de percentages van de z-waarde naar de staart toe (dit noemen we positief oneindig). Onze z-waarde is gelijk aan 1 en als we naar figuur 4.3b kijken, dan zoeken we het gebiedspercentage van de z-waarde naar positief oneindig (alles wat er boven ligt); we kijken dus naar de percentages in de tweede kolom.
We vinden dan een proportie van 0.1587, dus 15,87% van de meisjes is langer dan 1,70 meter.

b Hoeveel procent van de meisjes is 1,35 meter of kleiner?
Wederom vullen we de z-formule in om de bijbehorende z-waarde te vinden.

$$z = \frac{x - \bar{x}}{s_x} = \frac{135 - 155}{15} = \frac{-20}{15} = -1{,}33$$

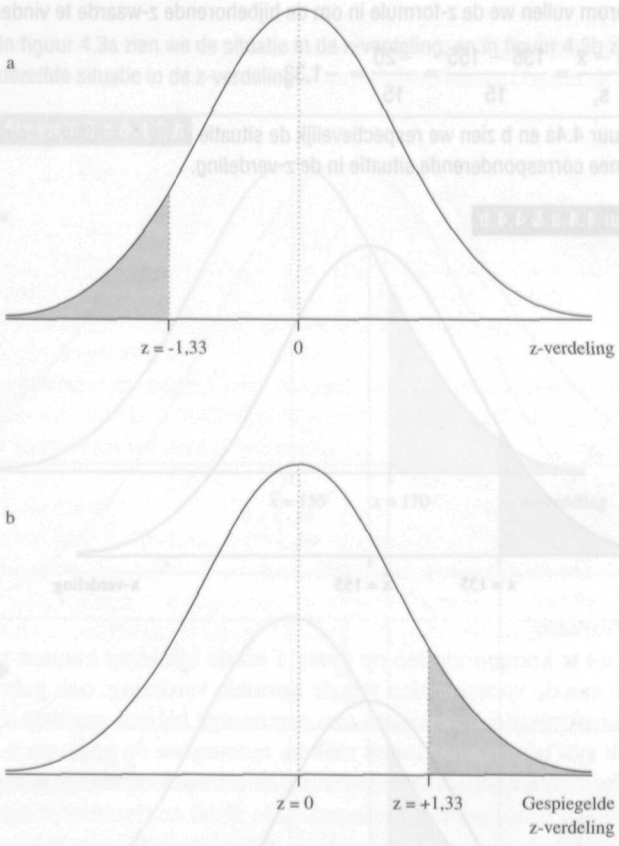
In figuur 4.4a en b zien we respectievelijk de situatie in de x-verdeling en de daarmee corresponderende situatie in de z-verdeling.

Figuur 4.4.a & 4.4.b



Als we nu naar tabel A gaan, dan zien we dat deze verdeling geen informatie geeft over negatieve z-waarden. Maar we weten wel, dat de normale en de standaardnormale verdeling symmetrisch zijn en daarom gespiegeld kunnen worden. Dit moeten we dan gaan doen om aan het percentage van −1,33 en lager te komen. In figuur 4.5a en b zien we eerst de gewone z-situatie en de gespiegelde situatie.

Figuur 4.5.a & 4.5.b



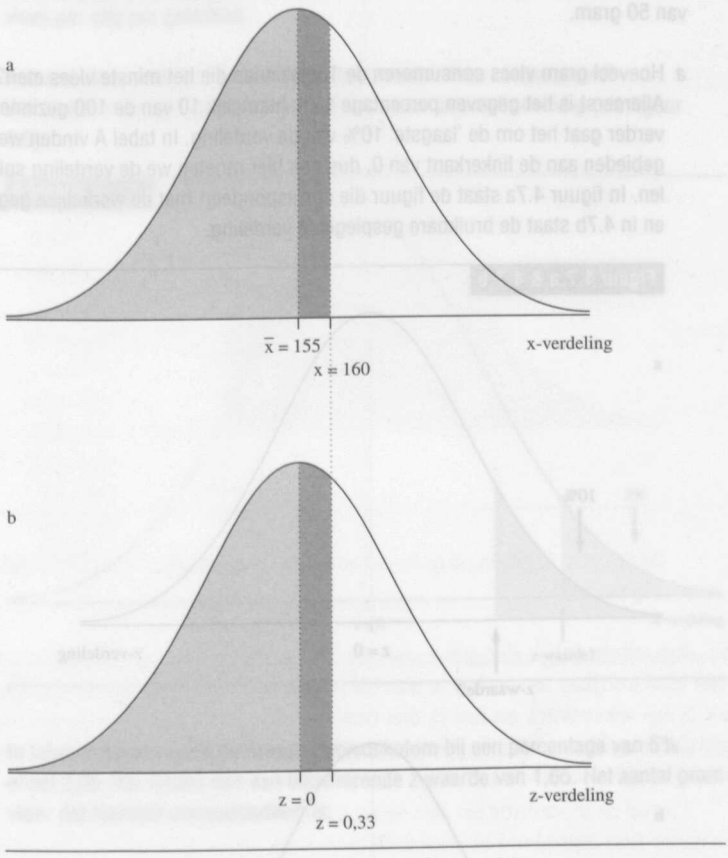
Het gebied van −1,33 naar negatief oneindig is dus qua oppervlakte gelijk aan het gebied van 1,33 naar positief oneindig. We kijken dus wederom in de tweede gebiedskolom en vinden een percentage van 0,0918, dus 9,18% van de meisjes is 1,35 of kleiner.

c Hoeveel procent van de meisjes is 1,60 meter of kleiner?

$$z = \frac{x - \bar{x}}{s_x} = \frac{160 - 155}{15} = \frac{5}{15} = 0{,}33$$

In figuur 4.6a en b staan de situaties weer grafisch weergegeven.
In tabel A zoeken we naar het juiste percentage behorend bij het gearceerde gebied in figuur 4.6b. We zien dan, dat ook hier de tabel niet volledig voorziet in de benodigde informatie: we zoeken het percentage behorend bij het gebied van z naar negatief oneindig. De tabel geeft ons in de eerste gebiedskolom informatie over het gebied van z naar 0; in dit gebied bevindt zich 0,1293 → dus 12,93% van de meisjes. We moeten echter ook het gebied hebben van 0 naar negatief oneindig, dit is de helft van de verdeling, dus in dit gebied bevindt zich 50% van de meisjes. Het totale percentage is dus gelijk aan 62,93%.

Figuur 4.6.a & 4.6.b



4.2.2 *De omgekeerde z-transformatie*

Als we bij een gegeven percentage een x-waarde zoeken (vraag 2 uit de inleiding), dan kunnen we ook gebruikmaken van de standaardnormaalverdeling via de omgekeerde z-transformatie.

We draaien de procedure zoals gevolgd in subparagraaf 4.2.1 om en beginnen bij tabel A om bij het gegeven percentage de bijbehorende z-waarde te zoeken. Als we deze gevonden hebben, dan moeten we van deze z-waarde weer naar de bijbehorende x-waarde; dit doen we via de omgekeerde z-transformatie.

$$x = z \cdot s_x + \bar{x}$$

We zullen het gebruik van deze methode verduidelijken aan de hand van voorbeeld 4.4.

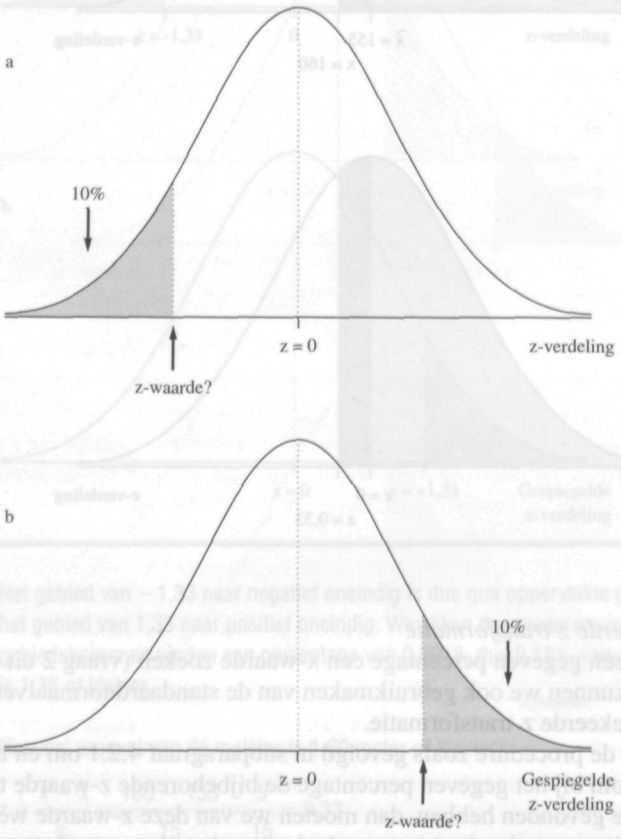
Voorbeeld 4.4: Gebruik van de omgekeerde z-transformatie

Stel dat we een onderzoek hebben gedaan naar de gemiddelde vleesconsumptie per dag van 100 gezinnen. De verdeling bleek normaal verdeeld met een gemiddel-

de vleesconsumptie van 150 gram per gezinslid per dag en een standaardafwijking van 50 gram.

a Hoeveel gram vlees consumeren de 10 gezinnen die het minste vlees eten?
Allereerst is het gegeven percentage 10% (namelijk 10 van de 100 gezinnen); verder gaat het om de 'laagste' 10% van de verdeling. In tabel A vinden we geen gebieden aan de linkerkant van 0, dus ook hier moeten we de verdeling spiegelen. In figuur 4.7a staat de figuur die correspondeert met de werkelijke gegevens en in 4.7b staat de bruikbare gespiegelde verdeling.

Figuur 4.7.a & 4.7.b



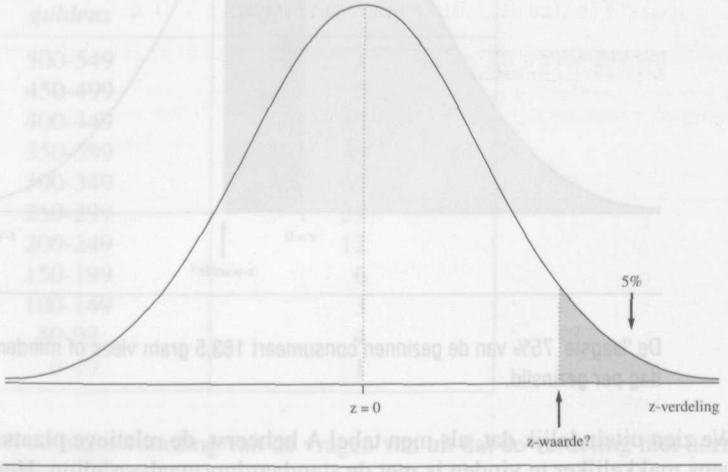
In tabel A gaan we in de tweede gebiedskolom zoeken naar een gebiedswaarde van 10% ofwel 0,10. We vinden dan een z-waarde van 1,28 (10,03% benadert de 10% het dichtst). Vanaf de z-waarde 1,28 naar positief oneindig valt dus 10% van de gezinnen, maar we hebben de z-waarde nodig die naar negatief oneindig loopt, we moeten dus weer terug spiegelen en de z-waarde wordt dan $-1{,}28$.
De bijbehorende vleesconsumptie vinden we door de omgekeerde z-transformatie in te vullen:

$$x = z \cdot s_x + \bar{x} = -1{,}28 \cdot 50 + 150 = -64 + 150 = 86$$

De 10% gezinnen die het minste vlees consumeren, eten 86 gram of minder vlees per dag per gezinslid.

b Hoeveel vlees eten de 5 gezinnen die het meeste vlees consumeren?
We zoeken hier het 5-procentsgebied van z naar positief oneindig (zie figuur 4.8).

Figuur 4.8



In tabel A zoeken we in de tweede gebiedskolom bij een percentage van 5% ofwel 0,05. We vinden dan een bijbehorende z-waarde van 1,65. Het aantal gram vlees dat hiermee correspondeert is:

$x = z \cdot s_x + \bar{x} = 1{,}65 \cdot 50 + 150 = 82{,}5 + 150 = 232{,}5$

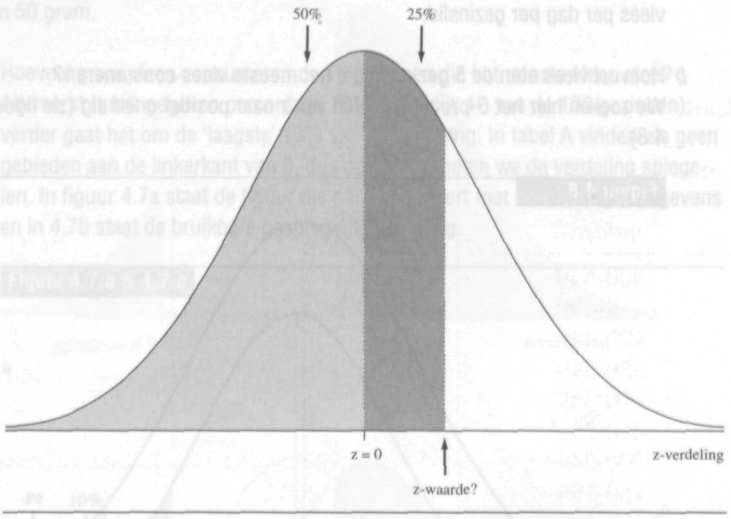
De 5% van de gezinnen die het meeste vlees consumeren, gebruiken dus meer dan 232,5 gram vlees per dag.

c Welke hoeveelheid vlees hoort bij P.75?

P.75 staat voor een proportie van 75%, zie figuur 4.9.
In tabel A gaan we zoeken in de eerste gebiedskolom naar een percentage van 25%. De tabel telt immers de linker 50% niet mee, daarom trekken wij die af van de 75% procent. De dichtstbijzijnde z-waarde is 0,67. Dus het *gehele* gebied dat links van z ligt, is gelijk aan 75%. De vleesconsumptie die hier bij hoort is:

$x = z \cdot s_x + \bar{x} = 0{,}67 \cdot 50 + 150 = 33{,}5 + 150 = 183{,}5$

Figuur 4.9



De 'laagste' 75% van de gezinnen consumeert 183,5 gram vlees of minder per dag per gezinslid.

We zien uiteindelijk dat, als men tabel A beheerst, de relatieve plaats van gegevens makkelijker te vinden is met de standaardnormaalverdeling. Het scheelt een hoop rekenwerk en het is dus ook minder gevoelig voor fouten. Vergeet niet dat we deze methode alleen mogen gebruiken als er sprake is van een (bij benadering) normale verdeling van de scores.

Opgaven

4.1 Uit een fictief onderzoek naar de totale waarde van het boekbezit van 175 studenten zijn de volgende resultaten afkomstig.

Totale waarde van het boekbezit afgerond op guldens	*frequentie*
500-549	1
450-499	5
400-449	10
350-399	39
300-349	65
250-299	29
200-249	12
150-199	6
100-149	3
50-99	4
0-49	1

Ga er bij de beantwoording van de vragen van uit dat de verdeling niet normaal van vorm is.

a Bereken hoeveel procent van de studenten een totale waarde aan boeken van 80 gulden of lager bezit.
b Hoeveel geld aan boeken heeft tweederde van de studenten in bezit?
c Een boekhandelaar beweert dat zeker 25% van de studenten een waarde aan boeken heeft van *meer* dan 375 gulden. Bereken of deze man gelijk heeft.

4.2 Deze opdracht maakt het mogelijk te oefenen met tabel A.
a Hoeveel procent ligt er links van $z = 0$?
b Hoeveel procent ligt er rechts van $z = 0$?
c Hoeveel procent ligt er rechts van $z = 2$?
d Hoeveel procent ligt er links van $z = 1{,}22$?
e Hoeveel procent ligt er tussen $z = 1{,}22$ en $z = 2$?
f Hoeveel procent ligt er links van $z = -1{,}94$?
g Hoeveel procent ligt er tussen $z = -0{,}75$ en $z = 1{,}38$?

4.3 Khadya behaalde voor schoolonderzoek Latijn een 7,5. Het gemiddelde van de klas was 6,6 met een standaardafwijking van 0,8. Hans scoorde op het schoolonderzoek natuurkunde een 8,0. Het gemiddelde van de klas was 6,3 met een standaardafwijking van 1,8. Wie van de twee – Khadya of Hans – was relatief gezien beter ten opzichte van hun klas?
Ga er bij de beantwoording van deze vraag van uit, dat beide verdelingen normaal van vorm zijn.

4.4 Uit een fictief, grootscheeps onderzoek naar het IQ van de Nederlandse bevolking blijkt het gemiddelde IQ 100 te zijn en de standaardafwijking 15. Beantwoord de volgende vragen, ervan uitgaande dat de IQ-scores normaal verdeeld zijn.

a Welk IQ correspondeert met de onderste 95% van de verdeling?

b Welk IQ heeft de laagste 15% van de mensen?

c Hoeveel procent van de mensen heeft een IQ tussen de 110 en de 130?

d Hoeveel procent van de mensen heeft een IQ tussen de 90 en de 135?

Uitwerkingen

4.1 *a* We zoeken hier een percentage bij een gegeven x-waarde van 80 gulden. De verdeling is niet normaal van vorm, dus we berekenen de percentielrang van de score 80.

$$P_{rang} = \frac{n_w \cdot (X - L) + i \cdot n_b}{n \cdot i}$$

Om deze formule in te kunnen vullen, moeten we de cumulatieve frequentietabel maken:

Totale waarde van het boekbezit afgerond op guldens	*frequentie*	*Cumulatieve frequentie*
500-549	1	175
450-499	5	174
400-449	10	169
350-399	39	159
300-349	65	120
250-299	29	55
200-249	12	26
150-199	6	14
100-149	3	8
50-99	4	5
0-49	1	1

$n_w = 4$ $\quad i = 50$
$X = 80$ $\quad n_b = 1$
$L = 49{,}5$ $\quad n = 175$

$$P_{rang} = \frac{n_w \cdot (X - L) + i \cdot n_b}{n \cdot i} = \frac{4 \cdot (80 - 49{,}5) + 50 \cdot 1}{175 \cdot 50}$$

$$= \frac{4 \cdot 30{,}5 + 50}{8750} = \frac{122 + 50}{8750} = \frac{172}{8750} = 0{,}0197 = 1{,}97\%$$

Dus 1,97% van de studenten heeft een boekbezit met een totale waarde van 80 gulden of lager.

b Bij deze opgave zoeken we het bedrag bij een proportie. Tweederde van de studenten is gelijk aan ongeveer 67%, dus een proportie van 0,67.
Aangezien de verdeling niet normaal van vorm is (dat is gegeven), mogen we geen gebruikmaken van de standaardnormaalverdeling; we berekenen dus het percentielpunt bij de proportie P.67.
Eerst gaan we de klasse bepalen waarin deze proportie valt:
$P(n) = 0{,}67 \cdot 175 = 117{,}25$.

Dus bij deze student valt de 67%. In de cumulatieve tabel zoeken we deze persoon op en we kijken in welke klasse deze persoon scoort qua boekbezit: dat is klasse 300-349. We kunnen nu de formule voor het percentielpunt berekenen.

$$P_{punt} = L + \left[\frac{p(n) - n_b}{n_w}\right] \cdot i = 299{,}5 + \left[\frac{117{,}25 - 55}{65}\right] \cdot 50 =$$

$$299{,}5 + \frac{62{,}25}{65} \cdot 50 = 299{,}5 + 0{,}96 \cdot 50 = 299{,}5 + 47{,}88 = 347{,}38$$

Dus het 'laagste' tweederde deel van de studenten bezit boeken met een totale waarde van 347 gulden of minder.

c We hebben in deze opgave een percentage gekregen dat iets zegt over een bepaald bedrag of hoger. We moeten controleren of dit percentage juist is, dus maken we gebruik van de percentielrangberekening. De boekhandelaar heeft het echter over het percentage dat ligt boven het bedrag (dus naar positief oneindig), terwijl de percentielberekening informatie geeft over het percentage dat *op en onder* een bepaalde waarde ligt. We moeten dus ook hier berekenen hoeveel procent er op en onder het bedrag van 375 gulden ligt. Daarna trekken we dit af van de 100% om te weten te komen hoeveel procent van de studenten dan nog een boekbezit van *meer* dan 375 gulden heeft.

$n_w = 39$ $\quad i = 50$
$X = 375$ $\quad n_b = 120$
$L = 349{,}5$ $\quad n = 175$

$$P_{rang} = \frac{n_w \cdot (X - L) + i \cdot n_b}{n \cdot i} = \frac{39 \cdot (375 - 349{,}5) + 50 \cdot 120}{175 \cdot 50}$$

$$= \frac{39 \cdot 25{,}5 + 6000}{8750} = \frac{994{,}5 + 6000}{8750} = \frac{6994{,}5}{8750} = 0{,}7994 = 79{,}94\%$$

Hieruit kunnen we afleiden dat 79,94% van de studenten een totaal boekbezit heeft ter waarde van 375 gulden of minder.
100% − 79,94% = 20,06%
20,06% van de studenten heeft dus een boekbezit met een waarde van *meer* dan 375 gulden.
De boekhandelaar schatte het percentage dus ongeveer 5% te hoog in.

4.2 *a* De standaardnormale verdeling is symmetrisch en in het midden van deze verdeling liggen de drie centrummaten modus, mediaan en gemiddelde (deze waarden zijn in deze verdeling gelijk aan 0). In dit punt wordt de verdeling in tweeën gedeeld, waarbij de helft van de scores boven en de andere helft van de scores onder ligt.
Links van z = 0 ligt dus 50% van de verdeling.

b Rechts van $z = 0$ ligt ook 50% van de verdeling.

c Het gebied rechts van $z = 2$ is het gebied van 2 naar positief oneindig. We zoeken deze waarde op in de tweede gebiedskolom. Daar vinden we een proportie van 0,0228, dus 2,28%.

d Links van $z = 1{,}22 \rightarrow$ Dit gebied moet in twee gedeelten benaderd worden:
1 Het gebied van 1,22 naar 0.
2 Het gebied links van $z = 0$.

Het eerste gebied vinden we in de eerste gebiedskolom, we zoeken namelijk een percentage dat grenst aan 0. We vinden daar een waarde van 0,3888; dus 38,88%.
Het tweede gebied hebben we al gevonden in opdracht *a* en is dus gelijk aan 50%. De twee gebiedswaarden moeten opgeteld worden om aan het gehele gebied te komen, dus 38,88% + 50% = 88,88%.

e Om aan deze gebiedswaarde te komen, moeten we eerst het gebied van $z = 2$ naar $z = 0$ opzoeken in de eerste gebiedskolom; 47,72% valt in dit gebied. Daar hoort echter een stuk *niet* bij, namelijk het gebied van $z = 1{,}22$ naar $z = 0$. Dit gebied zoeken we ook op in de eerste gebiedskolom; het heeft een waarde van 0,3888, dus 38,88%. Deze waarde trekken we af van 47,72%. 47,72% − 38,88% = 8,84% ligt tussen $z = 2$ en $z = 1{,}22$.

f Links van $z = -1{,}94$ wijst op het gebied van $-1{,}94$ naar negatief oneindig. De tabel voorziet niet in informatie over negatieve waarden van z, dus spiegelen we het gebied door de verdeling om te klappen. Het corresponderende gebied is dan het gebied van 1,94 naar positief oneindig. De waarde die hierbij hoort, vinden we in de tweede gebiedskolom en is gelijk aan 0,0262; dus 2,62%.

g Het gebied tussen $z = -0{,}75$ en $z = 1{,}38$ moeten we weer in twee gedeelten benaderen, namelijk:
1 Het gebied van $-0{,}75$ naar $z = 0$.
2 Het gebied van 1,38 naar $z = 0$.

Het eerste gebied moeten we weer vinden door spiegeling, het corresponderende gebied aan de positieve kant loopt van $z = 0{,}75$ naar $z = 0$. Het bijbehorende percentage is gelijk aan 27,34% (eerste gebiedskolom).
Het tweede gebied kunnen we direct opzoeken in de tabel (eerste gebiedskolom) en geeft ons een waarde van 41,62%. We moeten beide gebieden optellen om aan de totale gevraagde gebiedswaarde te komen, en dus is het antwoord: 27,34% + 41,62% = 68,96%.

4.3 Het gaat hier om twee normale verdelingen, dus we mogen beide scores transformeren naar de z-verdeling en dan vergelijken op een 'eerlijke' wijze. Degene die de hoogste z-score heeft, is relatief de 'beste'.

$$\text{z-score van Khadya} = \frac{7{,}5 - 6{,}6}{0{,}8} = 1{,}125$$

Volgens de z-tabel behoort Khadya tot de beste 13%.

$$\text{z-score van Hans} = \frac{8{,}0 - 6{,}3}{1{,}8} = 0{,}944$$

Volgens de z-tabel behoort Hans tot de beste 17%.

Dus, hoewel het gemiddelde in de klas bij Khadya hoger lag én Khadya een lager cijfer dan Hans behaalde, scoorde zij *relatief* beter dan Hans.

4.4 Bij de beantwoording van deze vraag maken we gebruik van de standaardnormaalverdeling omdat de verdeling van IQ-scores normaal van vorm is.

a We moeten bij deze vraag een omgekeerde z-transformatie toepassen omdat we een percentage gegeven krijgen en daarbij een x-waarde moeten zoeken. Allereerst gaan we naar tabel A om te zoeken naar de z-waarde die correspondeert met het percentage. We zoeken dan in de eerste gebiedskolom naar een percentage van 45% (0,45); tabel A telt immers het gebied van z = 0 naar negatief oneindig niet mee. We vinden dan een z-waarde van 1,64 omdat onder deze z-waarde afgerond 45% valt. Om van deze z-waarde weer een x-waarde te maken, vullen we de formule van de omgekeerde z-transformatie in:

$x = z \cdot s_x + \bar{x} = 1{,}64 \cdot 15 + 100 = 24{,}6 + 100 = 124{,}6$

Dus 95% van de mensen heeft een IQ van 124,6 of lager.

b Om de IQ-waarde te vinden die correspondeert met die 15% van de mensen die het laagste IQ hebben, gaan we wederom eerst naar de z-tabel om bij het percentage van 15% de juiste z-waarde te vinden. Aangezien we hier weer geconfronteerd worden met de niet-vermelde negatieve kant van de verdeling, zoeken we in de tweede gebiedskolom naar een percentage van 15. De dichtsbijzijnde z-waarde is 1,04. We moeten deze waarde weer terugspiegelen naar de negatieve kant; we werken nu dus met een z-waarde van −1,04 in de omgekeerde z-transformatie:

$x = z \cdot s_x + \bar{x} = -1{,}04 \cdot 15 + 100 = -15{,}6 + 100 = 84{,}4$

15% van de mensen heeft dus een IQ van 84,4 of lager.

c Bij de beantwoording van opdracht *c* en *d* maken we gebruik van de normale z-transformatie; we hebben immers x-waarden en we zoeken daar een percentage bij.

Het gebied dat we zoeken, moet benaderd worden in twee gedeelten:

1 Het gebied van 130 naar 100.

2 Het gedeelte van gebied 1 dat we niet nodig hebben, namelijk het gebied van 110 naar 100.

ad 1 $z = \frac{x - \bar{x}}{s_x} = \frac{130 - 100}{15} = \frac{30}{15} = 2$

Het corresponderende gebied op de z-verdeling loopt dus van $z = 0$ naar $z = 2$. De bijbehorende gebiedswaarde is gelijk aan 47,72% (eerste gebiedskolom).

ad 2 $z = \frac{x - \bar{x}}{s_x} = \frac{110 - 100}{15} = \frac{10}{15} = 0{,}67$

Op de z-verdeling loopt dit gebied dus van $z = 0$ tot $z = 0{,}67$; dit is het gebied dat we niet willen hebben, maar dat al wel bij het eerste gebied is geteld. We moeten dit percentage van 24,86% (eerste gebiedskolom) aftrekken van het eerste gebied: 47,72% − 24,86% = 22,86%.

d Deze vraag kunnen we op gelijke wijze als in *c* beantwoorden, alleen moeten we nu beide gebieden hebben, dus bij elkaar optellen.
1 Het eerste gebied loopt van 90 naar 100.
2 Dit gebied loopt van 135 naar 100.

ad 1 $z = \frac{x - \bar{x}}{s_x} = \frac{90 - 100}{15} = \frac{-10}{15} = -0{,}67$

Dit gebied bevat 24,86%.

ad 2 $z = \frac{x - \bar{x}}{s_x} = \frac{135 - 100}{15} = \frac{35}{15} = 2{,}33$

Dit gebied bevat 49,01%.
Totaal omvat het gevraagde gebied 24,86% + 49,01% = 73,87%.

5 Regressie en correlatie

De correlatie- en regressietechnieken beschrijven de relatie c.q. het verband tussen variabelen. In dit hoofdstuk beperken we ons tot de relatie tussen twee variabelen. Later in dit boek worden ook technieken behandeld die de relatie tussen meerdere variabelen kunnen beschrijven.

5.1 Regressie

Regressie houdt zich vooral bezig met het voorspellen van de ene variabele (y) door een andere variabele (x).
Om beter inzicht te verkrijgen in de problematiek van regressie en correlatie, wordt in dit gehele hoofdstuk gebruikgemaakt van de volgende voorbeeldsituatie.

variabele 1: werktijd in uren
variabele 2: verdienste

Stel dat we een klein onderzoek doen naar deze variabelen bij vijf verschillende personen; daaruit komen de volgende resultaten:

Tabel 5.1

Persoon	*Werktijd*	*Verdienste*
1	1 uur	25 gulden
2	4 uur	40 gulden
3	5 uur	75 gulden
4	8 uur	100 gulden
5	10 uur	110 gulden

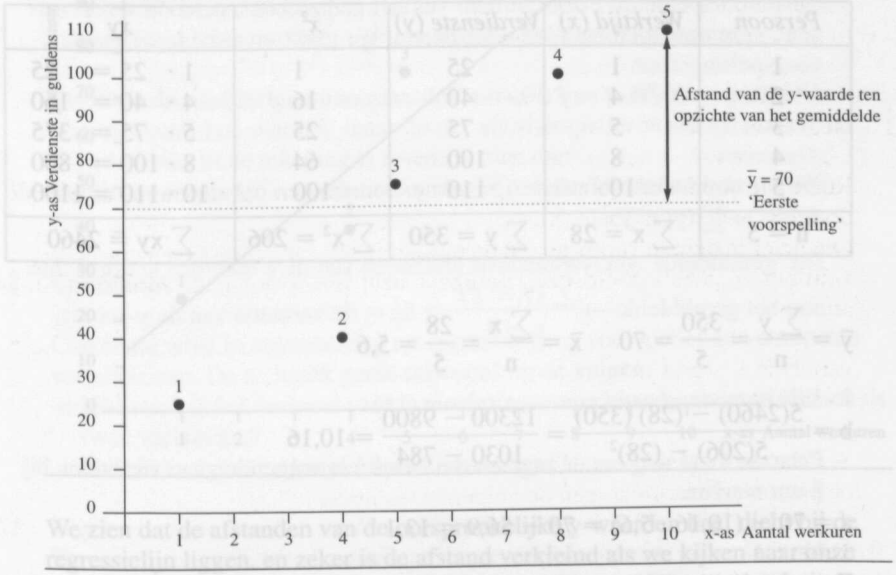
Als we alleen naar de verdienste kijken, dan zien we dat niet iedere persoon hetzelfde verdient, met andere woorden: er bestaat spreiding binnen de verdeling van verdienste. Als men op basis van deze gegevens een voorspelling van verdienste wil maken, dan kunnen we alleen de gemiddelde waarde geven; die is *f* 70,–. Deze voorspelling is niet zo nauwkeurig en we willen dan ook verbetering brengen in de voorspelling. We moeten dan gaan zoeken naar een goede variabele als voorspeller van verdienste, een variabele die de variantie in de verdeling van de geldbedragen voor een deel kan verklaren, bijvoorbeeld de werktijd. Het aantal uren werk heeft immers te maken met het verdiende loon. Om nu nauwkeuriger te voorspellen, moeten we regressie toepassen. Deze procedure gaat als volgt.

5.1.1 *De regressielijn*

Om inzicht te krijgen in een verband kunnen we de variabelen uitzetten in een assenstelsel: de voorspellende (onafhankelijke) variabele (x) op de x-as en de

voorspelde (afhankelijke) variabele (y) op de y-as. Door de waarden van x en y elke keer te combineren, kunnen we een puntenwolk creëren zoals figuur 5.1 laat zien.

Figuur 5.1 **Puntenwolk**



We zien dat de punten ongeveer op één lijn liggen, maar niet precies. Met behulp van regressie kunnen we die lijn berekenen die het best door de puntenwolk heen loopt met een zo klein mogelijke afwijking tot de oorspronkelijke punten. Deze lijn noemen we de **regressielijn.**
Elke lijn loopt anders, de ene lijn is steiler dan de andere of snijdt de y-as op een andere plaats. De ene lijn loopt van boven naar beneden (\) de andere weer van beneden naar boven (/). Deze zaken worden bepaald door de aard van de relatie tussen de twee variabelen. Om achter de aard van een relatie te komen (om op basis van onderzoeksgegevens erachter te komen hoe de lijn loopt), kunnen we de functie van de lijn gaan bepalen:
Elke rechte lijn kent de functie $\mathbf{y = bx + a}$, waarbij **b** de helling van de lijn aangeeft (hoe steil is de lijn) en **a** het snijpunt met de y-as weergeeft. **a** en **b** worden **regressieconstanten** genoemd. De voorspellingslijn, ook wel regressielijn genoemd, ziet er als volgt uit: $\hat{\mathbf{y}} = \mathbf{bx + a}$.
Het dakje boven de **y** geeft aan dat we niet met een oorspronkelijke lijn hebben te maken, maar met een lijn die zo goed mogelijk door de puntenwolk loopt.
We moeten nu de regressieconstanten gaan berekenen om tot een best mogelijke voorspelling van verdienste te komen op basis van werktijd. Om **b** te vinden bestaat er een formule:

$$\mathbf{b} = \frac{\mathbf{n}(\sum \mathbf{xy}) - (\sum \mathbf{x})(\sum \mathbf{y})}{\mathbf{n}\sum \mathbf{x}^2 - (\sum \mathbf{x})^2}$$

Als **b** is gevonden, kunnen we **a** uitrekenen met de volgende formule:

$\mathbf{a = \bar{y} - b\bar{x}}$

We voeren nu de berekeningen uit voor het voorbeeld:

Tabel 5.2

Persoon	*Werktijd (x)*	*Verdienste (y)*	x^2	xy
1	1	25	1	$1 \cdot 25 = 25$
2	4	40	16	$4 \cdot 40 = 160$
3	5	75	25	$5 \cdot 75 = 375$
4	8	100	64	$8 \cdot 100 = 800$
5	10	110	100	$10 \cdot 110 = 1100$
$n = 5$	$\sum x = 28$	$\sum y = 350$	$\sum x^2 = 206$	$\sum xy = 2460$

$$\bar{y} = \frac{\sum y}{n} = \frac{350}{5} = 70 \qquad \bar{x} = \frac{\sum x}{n} = \frac{28}{5} = 5{,}6$$

$$b = \frac{5(2460) - (28)(350)}{5(206) - (28)^2} = \frac{12300 - 9800}{1030 - 784} = 10{,}16$$

$$a = 70 - (10{,}16 \cdot 5{,}6) = 70 - 56{,}9 = 13{,}1$$

De regressielijn die hoort bij onze gegevens, ziet er als volgt uit:

$\hat{y} = 10{,}16x + 13{,}1$

Met deze functie kunnen we gaan voorspellen wat iemand zal gaan verdienen. We moeten dan wel weten hoeveel uur die persoon gewerkt heeft.
Bijvoorbeeld: Eric heeft 3 uur gewerkt, hij wil weten wat hij zal gaan verdienen. We zetten de 3 op de plaats van de x in de functie:

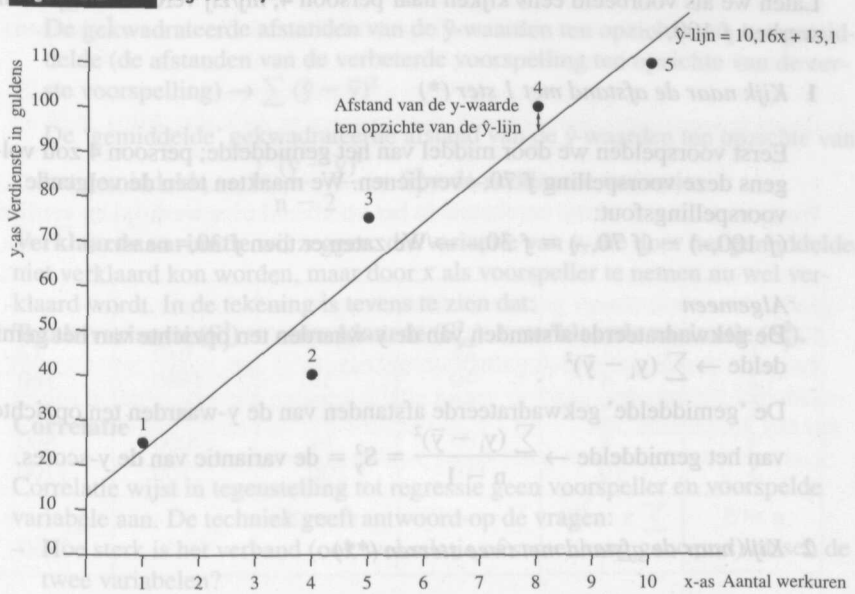
$\hat{y}$ van Eric $= 10{,}16 \cdot 3 + 13{,}1$
$\hat{y}$ van Eric $= 30{,}48 + 13{,}1$
$\hat{y}$ van Eric $= 43{,}58$

Dit betekent dat Eric volgens onze voorspelling bij 3 uur werken *f* 43,58 zal verdienen.

We kunnen de lijn ook gaan tekenen in de puntenwolk, daarvoor moeten we minstens twee punten uitrekenen:

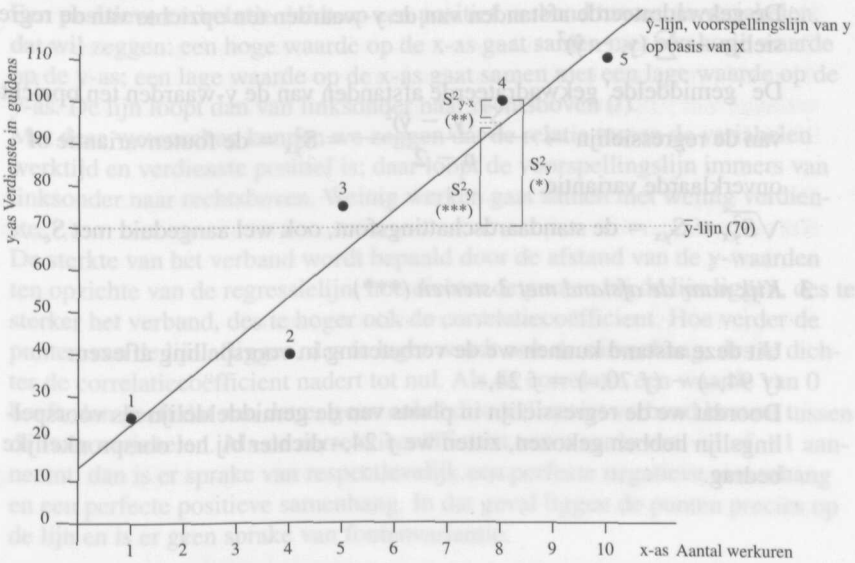
$x = 0 \rightarrow \hat{y} = 10{,}16 \cdot 0 + 13{,}1 \rightarrow \hat{y} = 13{,}1$
dus het punt (0 en 13,1)
$x = 7 \rightarrow \hat{y} = 10{,}16 \cdot 7 + 13{,}1 \rightarrow \hat{y} = 84{,}22$
dus het punt (7 en 84,22)

Figuur 5.2 **Regressielijn**



We zien dat de afstanden van de oorspronkelijke y-waarden heel dicht bij de regressielijn liggen, en zeker is de afstand verkleind als we kijken naar onze 'eerste voorspelling' in figuur 5.1. We hebben dus vooruitgang geboekt in het voorspellen van de y-waarden. Als we beide tekeningen combineren dan krijgen we het volgende:

Figuur 5.3 **Combinatie van puntenwolk en regressielijn**



Laten we als voorbeeld eens kijken naar persoon 4, hij/zij verdiende bij 8 uur werken f 100,–.

1 *Kijk naar de afstand met 1 ster (*)*

Eerst voorspelden we door middel van het gemiddelde; persoon 4 zou volgens deze voorspelling f 70,– verdienen. We maakten toen de volgende voorspellingsfout:
(f 100,–) − (f 70,–) = f 30,– → We zaten er toen f 30,– naast.

Algemeen
De gekwadrateerde afstanden van de y-waarden ten opzichte van het gemiddelde → $\sum (y_i - \bar{y})^2$

De 'gemiddelde' gekwadrateerde afstanden van de y-waarden ten opzichte van het gemiddelde → $\dfrac{\sum (y_i - \bar{y})^2}{n - 1} = S_y^2$ = de variantie van de y-scores.

2 *Kijk naar de afstand met twee sterren (**)*

Persoon 4 zal volgens de regressietechniek het volgende gaan verdienen bij 8 uur werken:
$\hat{y} = 10{,}16 \cdot 8 + 13{,}1 = f\ 94{,}38 \approx f\ 94{,}–$
Nu maken we de volgende voorspellingsfout:
(f 100,–) − (f 94,–) = f 6,– → Nu zitten we nog maar f 6,– naast het werkelijke bedrag. Deze voorspellingsfout is aanzienlijk kleiner dan de eerste fout.

Algemeen
De gekwadrateerde afstanden van de y-waarden ten opzichte van de regressielijn → $\sum (y_i - \hat{y})^2$

De 'gemiddelde' gekwadrateerde afstanden van de y-waarden ten opzichte van de regressielijn → $\dfrac{\sum (y_i - \hat{y})^2}{n - 2} = S_{y.x}^2$ = de foutenvariantie of onverklaarde variantie.

$\sqrt{S_{y.x}^2} = S_{y.x}$ = de standaardschattingsfout, ook wel aangeduid met S_e.

3 *Kijk naar de afstand met 3 sterren (***)*

Uit deze afstand kunnen we de verbetering in voorspelling aflezen:
(f 94,–) − (f 70,–) = f 24,–
Doordat we de regressielijn in plaats van de gemiddeldelijn als voorspellingslijn hebben gekozen, zitten we f 24,– dichter bij het oorspronkelijke bedrag.

Algemeen

De gekwadrateerde afstanden van de $\hat{y}$-waarden ten opzichte van het gemiddelde (de afstanden van de verbeterde voorspelling ten opzichte van de eerste voorspelling) $\rightarrow \sum (\hat{y} - \bar{y})^2$

De 'gemiddelde' gekwadrateerde afstand van de $\hat{y}$-waarden ten opzichte van het gemiddelde $\rightarrow \dfrac{\sum (\hat{y} - \bar{y})^2}{n - 2} = S^2_{\hat{y}}$= de verklaarde variantie.

Verklaarde variantie wil zeggen: die variantie van y, die door het gemiddelde niet verklaard kon worden, maar door x als voorspeller te nemen nu wel verklaard wordt. In de tekening is tevens te zien dat:
Totale variantie (S^2_y) = onverklaarde ($S^2_{y.x}$) + verklaarde variantie ($S^2_{\hat{y}}$).

5.2 Correlatie

Correlatie wijst in tegenstelling tot regressie geen voorspeller en voorspelde variabele aan. De techniek geeft antwoord op de vragen:
- Hoe sterk is het verband (ook wel relatie of samenhang genoemd) tussen de twee variabelen?
- Gaat het om een negatief of positief verband?

5.2.1 *Eigenschappen van de correlatiecoëfficiënt*

De correlatiecoëfficient kan een waarde aannemen tussen -1 en 0 en tussen 0 en $+1$.

Een **negatieve correlatie** duidt op een negatief verband tussen de variabelen, dat wil zeggen: een hoge waarde op de x-as gaat samen met een lage waarde op de y-as; een lage waarde op de x-as gaat samen met een hoge waarde op de y-as. De lijn loopt dan van rechtsonder naar linksboven (\).

Een **positieve correlatie** duidt op een positief verband tussen de variabelen, dat wil zeggen: een hoge waarde op de x-as gaat samen met een hoge waarde op de y-as; een lage waarde op de x-as gaat samen met een lage waarde op de y-as. De lijn loopt dan van linksonder naar rechtsboven (/).

Met deze wetenschap kunnen we zeggen dat de relatie tussen de variabelen werktijd en verdienste positief is; daar loopt de voorspellingslijn immers van linksonder naar rechtsboven. Weinig werken gaat samen met weinig verdienste.

De sterkte van het verband wordt bepaald door de afstand van de y-waarden ten opzichte van de regressielijn: hoe dichter de punten bij de lijn liggen, des te sterker het verband, des te hoger ook de correlatiecoëfficient. Hoe verder de punten van de lijn afliggen, des te lager wordt ook de samenhang, des te dichter de correlatiecoëfficient nadert tot nul. Als de correlatie een waarde van 0 heeft, dan geeft dat aan dat er geen enkel direct (lineair) verband bestaat tussen de twee variabelen. Als de correlatiecoëfficient een waarde van -1 of $+1$ aanneemt, dan is er sprake van respectievelijk een perfecte negatieve samenhang en een perfecte positieve samenhang. In dat geval liggen de punten precies op de lijn en is er geen sprake van foutenvariantie.

5.2.2 *Berekening van de correlatiecoëfficiënt*

De mate van lineaire correlatie wordt gesymboliseerd met de r (van Pearson) en wordt als volgt berekend:

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\Sigma x)^2][n\sum y^2 - (\sum y)^2]}}$$

We gaan nu de correlatie berekenen in het voorbeeld over werktijd en verdienste.

Persoon	x	y	x^2	y^2	xy
1	1	25	1	625	25
2	4	40	16	1600	160
3	5	75	25	5625	375
4	8	100	64	10000	800
5	10	110	100	12100	1100
n = 5	$\sum x = 28$	$\sum y = 350$	$\sum x^2 = 206$	$\sum y^2 =$ 29950	$\sum xy =$ 2460

$$r = \frac{5(2460) - (28)(350)}{\sqrt{[5(206) - (28^2)][5(29950) - (350^2)]}}$$

$$= \frac{12300 - 9800}{\sqrt{[1030 - 784][149750 - 122500]}} = \frac{2500}{\sqrt{[246][27250]}} = \frac{2500}{\sqrt{6703500}}$$

$$= \frac{2500}{2589{,}11} = 0{,}96$$

Het resultaat van de berekening levert een zeer hoge positieve correlatie op. Ondanks het feit dat dit een hoge correlatie is, weten we nog niets concreets. Als we de correlatie kwadrateren, dan krijgen we het percentage verklaarde variantie van y op basis van x! In dit geval: $r^2 = (0{,}96)^2 = 0{,}9216 \times 100\,\% = 92{,}16\%$.

Dit wil zeggen dat 92,16% van de variantie in y verklaard kan worden door x. Het percentage verklaarde variantie kan ook op een andere manier verkregen worden.

$$r^2 = \frac{S^2_{\hat{y}}}{S^2_y} = \frac{\sum(\hat{y} - \bar{y})^2}{\sum(y_i - \bar{y})^2} = \text{proportie verklaarde variantie}$$

$$r^2 = 1 - \frac{S^2_{yx}}{S^2_y} = 1 - \frac{\sum(y_i - \hat{y})^2}{\sum(y_i - \bar{y})^2} = 1 - \text{de proportie onverklaarde variantie}$$

Als de correlatie negatief is, dan is de helling van de regressielijn (b) ook negatief.

b en r verhouden zich tot elkaar als in de volgende formule:

$$r = b \cdot \frac{S_x}{S_y}$$

Hieruit kan worden opgemaakt dat r en b alleen gelijk zijn aan elkaar als de S_x en de S_y gelijk zijn. Ook in het geval van z-scores (standaardnormaalverdeling) geldt, dat $r = b$, want zowel de S_x als de S_y zijn dan gelijk aan 1.
De standaardschattingsfout kan als volgt berekend worden:

$$S_{y.x} = S_y \cdot \sqrt{1 - r^2}$$

Deze waarde symboliseert de gemiddelde afwijking van de oorspronkelijke y-waarden ten opzichte van de regressielijn. Als alle waarnemingspunten op de regressielijn liggen, dan is de standaardschattingsfout gelijk aan 0, dit houdt tevens in dat er dan een perfecte relatie tussen de twee onderzochte variabelen is; de correlatie neemt dan een waarde van $+1$ of -1 aan, afhankelijk van het feit of er een positieve of een negatieve samenhang tussen de variabelen is.

Opgaven

5.1 In een vragenlijst over gehechtheid tussen tweejarige kinderen en hun opvoeders komen de volgende twee vragen voor:

Vraag 12: Hoeveel mensen, behalve de ouders, houden zich regelmatig met uw kind bezig? (meer dan een paar minuten)
Vraag 37: Hoe vaak is uw kind bang? (1 = bijna nooit; 2 = soms; 3 = vaak; 4 = heel vaak)

Hieronder volgen de antwoorden voor negen gezinnen.

Gezin	*Vraag 12*	*Vraag 37*
1	3	3
2	3	2
3	0	3
4	8	1
5	2	2
6	5	1
7	4	1
8	1	3
9	7	niet ingevuld

Beschouw de onderliggende verdelingen van beide variabelen als normaal.

a Is er een reden om een van beide vragen als onafhankelijke en de ander als afhankelijke variabele te beschouwen? Zo ja, waarom?
b Teken een puntenwolk (X-as: vraag 12; Y-as: vraag 37).
c Bereken de regressielijn op grond van de eerste acht gezinnen.
d Teken deze lijn in de onder *b* weergegeven puntenwolk.
e Geef met behulp van het plaatje (puntenwolk + regressielijn) een schatting voor de bangheidsscore van het negende kind. Bereken daarna de exacte schatting.

5.2 *a* Bekijk de puntenwolk uit opdracht 1. Zal de correlatie tussen vraag 12 en 37 negatief of positief zijn? Zal de correlatie hoog, redelijk of laag zijn?
b Bereken de correlatiecoëfficiënt r_{xy} op grond van de eerste acht gezinnen.
c Hoeveel procent van de variantie in de bangheidsscores is te verklaren uit de variantie van de vraag 12-scores?
d Bereken de standaardschattingsfout van de regressielijn.

5.3 *a* Stel dat we *niets* wisten over het verband tussen vraag 12 en 37. Wat zou dan de 'best voorspellende lijn' zijn?
b Vergelijk de echte regressielijn met de lijn uit opgave *a*. Welke van de twee is een meer ideale lijn en waarom?

c Bereken de afstanden van de acht Y-waarden tot de lijn uit opgave *a*. Kwadrateer deze afstanden en tel ze op. Deel deze som door 7. Wat stelt de gevonden waarde voor?

5.4 Er wordt vaak gesuggereerd dat analytisch denkvermogen niets te maken heeft met artistiek vermogen van mensen. Toch is er een 'eigenwijze' onderzoeker die wil kijken of het artistiek vermogen van iemand te voorspellen is uit het analytisch denkvermogen van die persoon. Hieronder staan de gegevens die deze onderzoeker verzameld heeft.

Proefpersoon	*Analytisch denkvermogen*	*Artistiek vermogen*
1	5	6
2	7	8
3	3	1
4	1	4
5	8	3
6	2	2
7	6	7
8	6	5
9	8	7
10	4	0
11	2	3
12	9	2
13	3	1
14	7	6

a Maak een puntenwolk en bereken de voorspellingslijn op basis van deze gegevens.

b Als iemand een analytisch denkvermogen van 9 heeft, wat verwachten we dan dat deze persoon zal scoren op artistiek vermogen?

c Bereken het verband tussen analytisch denkvermogen en artistiek vermogen.

d Hoeveel procent van de totale variantie in artistiek vermogen kan verklaard worden door de variantie in de scores van analytisch denkvermogen?

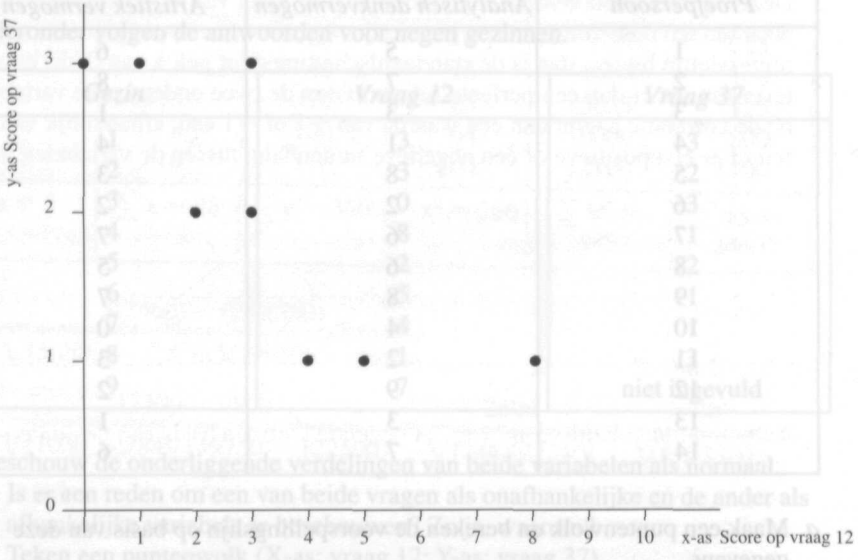
e Bereken de spreiding binnen artistiek vermogen (S_y^2) en de foutenvariantie ($S_{y.x}^2$). Hoeveel procent onverklaarde variantie kunnen we uit deze twee spreidingsmaten afleiden?

Uitwerkingen

5.1 *a* Het meest aannemelijk is angst te voorspellen uit de omgang met mensen, bijvoorbeeld: 'Hoe meer mensen zich met het kind bezighouden, des te minder angstig het kind.'

b

Antwoord 5.1.b



c $\hat{y} = bx + a \rightarrow b = \dfrac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2} = \dfrac{8 \cdot 39 - 26 \cdot 16}{8 \cdot 128 - 676} = \dfrac{-104}{348}$

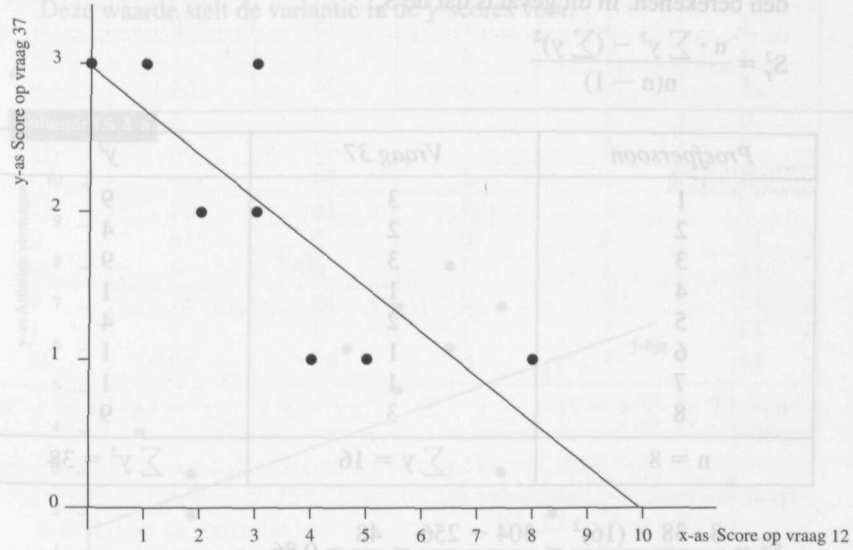
$= -0{,}30$

$a = \bar{y} - b \cdot \bar{x} = \dfrac{16}{8} - (-0{,}30) \cdot \dfrac{26}{8} = 2{,}98$

$\hat{y} = -0{,}30x + 2{,}98$

d

Antwoord 5.1.d



e Schatting 'bangheidsscore' van het negende kind, op basis van de puntenwolk en de regressielijn:
Het negende kind heeft een score van 7 op vraag 12. Trek een loodlijn vanuit 7 naar de regressielijn en vanuit het snijpunt met de regressielijn een loodlijn op de Y-as. We vinden dan een y-waarde van ongeveer 0,85. De exacte schatting kunnen we verkrijgen door de functie van de regressielijn in te vullen. Vul dan voor x 7 in:
$\hat{y} = -0{,}30 \cdot 7 + 2{,}98 = 0{,}88$
We schatten dus voor het negende kind een bangheidsscore van 0,88, berekend op basis van het aantal mensen dat zich met het kind bezighoudt.

5.2 *a* Op het eerste gezicht kunnen we zeggen dat de correlatie negatief is; de regressielijn loopt immers van linksboven naar rechtsonder en we hebben te maken met een vrij hoge correlatie, omdat de punten vrij dicht bij de regressielijn liggen.

b $$r = \frac{n \sum xy - \sum x \cdot \sum y}{\sqrt{[n \sum x^2 - (\sum x)^2] \cdot [n \sum y^2 - (\sum y)^2]}}$$

$$= \frac{312 - 416}{\sqrt{[1024 - 676] \cdot [304 - 256]}} = \frac{-104}{129{,}24} = -0{,}80$$

c $r^2 = (0{,}80)^2 = 0{,}64 \rightarrow 64\%$ van de variantie in de bangheidsscores kan verklaard worden door de variantie in de scores van vraag 37.

d $S_{y.x} = S_y \cdot \sqrt{1 - r^2}$

Om deze formule in te kunnen vullen, moeten we eerst de onbekende waarden berekenen. In dit geval is dat de S_y.

$$S_y^2 = \frac{n \cdot \sum y^2 - (\sum y)^2}{n(n-1)}$$

Proefpersoon	*Vraag 37*	y^2
1	3	9
2	2	4
3	3	9
4	1	1
5	2	4
6	1	1
7	1	1
8	3	9
$n = 8$	$\sum y = 16$	$\sum y^2 = 38$

$$S_y^2 = \frac{8 \cdot 38 - (16)^2}{8(8-1)} = \frac{304 - 256}{56} = \frac{48}{56} = 0{,}86$$

$$S_y = \sqrt{S_y^2} = \sqrt{0{,}86} = 0{,}93$$

Nu kunnen we de formule voor $S_{y.x}$ invullen:

$S_y = 0{,}93$

$r^2 = 0{,}64$

$S_{y.x} = 0{,}93 \cdot \sqrt{1 - 0{,}64}$

$S_{y.x} = 0{,}93 \cdot \sqrt{0{,}36}$

$S_{y.x} = 0{,}93 \cdot 0{,}6$

$S_{y.x} = 0{,}56$

5.3 *a* Dit is de gemiddeldelijn van de y-scores, de lijn die dus evenwijdig aan de x-as loopt en begint op de y-as bij $y = 2$; dat is namelijk het gemiddelde van de y-scores.

b De echte regressielijn lijkt idealer: over het algemeen zijn de afstanden van de punten tot de regressielijn kleiner dan tot de gemiddeldelijn. De regressielijn levert dus een betere voorspelling op, omdat deze dichter bij de oorspronkelijke punten ligt.

c We trekken nu van elke y-score 2 af (y is immers 2)

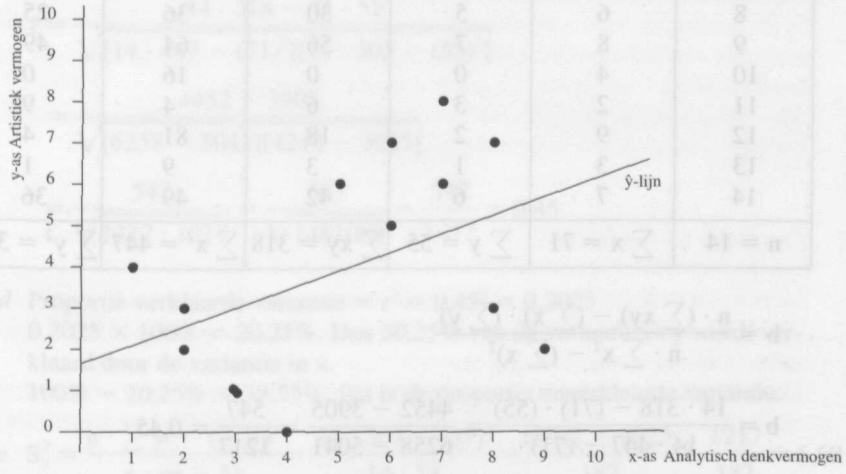
$$\sum (y_i - \bar{y})^2 = (3-2)^2 + (2-2)^2 + (3-2)^2 + (1-2)^2 + (2-2)^2 + (1-2)^2 + (1-2)^2 + (3-2)^2$$
$$= 1^2 + 0^2 + 1^2 + (-1)^2 + 0^2 + (-1)^2 + (-1)^2 + 1^2 = 6$$

$$\frac{\sum (y_i - \bar{y})^2}{n - 1} = \frac{6}{7} = 0{,}86$$

Deze waarde stelt de variantie in de y-scores voor.

5.4 *a*

Antwoord 5.4.a



De onderzoeker blijkt nogal eigenwijs; als we naar de puntenwolk kijken, zien we dat de punten niet duidelijk op een lijn liggen. Toch zullen we de onderzoeker verder helpen.
De regressielijn is de lijn die zo goed mogelijk door de puntenwolk heen loopt; de lijn waar de punten het dichtst bij liggen.

$$\hat{y} = bx + a$$

$$b = \frac{n \cdot (\sum xy) - (\sum x) \cdot (\sum y)}{n \cdot \sum x^2 - (\sum x)^2}$$

$$a = \bar{y} - b \cdot \bar{x}$$

We moeten eerst ***b*** (de helling) uitrekenen om ***a*** (het snijpunt met de y-as te kunnen vinden).

Persoon	x	y	xy	x^2	y^2
1	5	6	30	25	36
2	7	8	56	49	64
3	3	1	3	9	1
4	1	4	4	1	16
5	8	3	24	64	9
6	2	2	4	4	4
7	6	7	42	36	49
8	6	5	30	36	25
9	8	7	56	64	49
10	4	0	0	16	0
11	2	3	6	4	9
12	9	2	18	81	4
13	3	1	3	9	1
14	7	6	42	49	36
$n = 14$	$\sum x = 71$	$\sum y = 55$	$\sum xy = 318$	$\sum x^2 = 447$	$\sum y^2 = 303$

$$b = \frac{n \cdot (\sum xy) - (\sum x) \cdot (\sum y)}{n \cdot \sum x^2 - (\sum x)^2}$$

$$b = \frac{14 \cdot 318 - (71) \cdot (55)}{14 \cdot 447 - (71)^2} = \frac{4452 - 3905}{6258 - 5041} = \frac{547}{1217} = 0{,}45$$

$$a = \bar{y} - b \cdot \bar{x}$$

$$\bar{y} = \frac{\sum y}{n} = \frac{55}{14} = 3{,}93$$

$$\bar{x} = \frac{\sum x}{n} = \frac{71}{14} = 5{,}07$$

$$a = 3{,}93 - 0{,}45 \cdot 5{,}07 = 1{,}65$$

$$\hat{y} = 0{,}45x + 1{,}65$$

Nu moeten we twee punten uitrekenen om de lijn te kunnen tekenen:
$x = 0 \rightarrow \hat{y} = 0{,}45 \cdot 0 + 1{,}65 = 1{,}65$
$x = 7 \rightarrow \hat{y} = 0{,}45 \cdot 7 + 1{,}65 = 4{,}8$

b Analytisch denkvermogen van 9 → invullen voor x in de functie:
$\hat{y} = 0{,}45 \cdot 9 + 1{,}65 = 5{,}7$
We voorspellen voor deze persoon een artistiek vermogen van 5,7.

c Verband → correlatie berekenen. Correlatie geeft de sterkte van het verband aan, hier verwachten we niet zo'n hoge correlatie aangezien de punten ver van de regressielijn afliggen.

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n \sum x^2 - (\Sigma x)^2][n \sum y^2 - (\sum y)^2]}}$$

De waarden om deze formule in te vullen, hebben we al berekend toen we de regressielijn moesten berekenen:

$\sum xy = 318$

$\sum x = 71$

$\sum y = 55$

$\sum x^2 = 447$

$\sum y^2 = 303$

$n = 14$

$$r = \frac{14 \cdot 318 - 71 \cdot 55}{\sqrt{[14 \cdot 447 - (71)^2][14 \cdot 303 - (55)^2]}}$$

$$= \frac{4452 - 3905}{\sqrt{[6258 - 5041][4242 - 3025]}}$$

$$= \frac{547}{\sqrt{1217 \cdot 1217}} = \frac{547}{\sqrt{1481089}} = \frac{547}{1217} = 0{,}45$$

d Proportie verklaarde variantie $= r^2 = 0{,}45^2 = 0{,}2025$

$0{,}2025 \times 100\% = 20{,}25\%$. Dus 20,25% van de variantie in y wordt verklaard door de variantie in x.

$100\% - 20{,}25\% = 79{,}75\%$. Dit is de proportie onverklaarde variantie.

e $S_y^2 = \frac{n \cdot \sum y^2 - (\sum y)^2}{n \cdot (n-1)} = \frac{14 \cdot 303 - (55)^2}{14 \cdot 13} = \frac{4242 - 3025}{182} = \frac{1217}{182} = 6{,}69$

$S_{y.x}^2 = (S_{y.x})^2$

$S_{y.x} = S_y \cdot \sqrt{1 - r^2}$

$S_y = \sqrt{S_y^2} = \sqrt{6{,}69} = 2{,}59$

$S_{y.x} = 2{,}59 \cdot \sqrt{1 - 0{,}2025}$

$S_{y.x} = 2{,}59 \cdot \sqrt{0{,}7975}$

$S_{y.x} = 2{,}59 \cdot 0{,}89 = 2{,}31 \rightarrow S_{y.x} = 2{,}31$

$S_{y.x}^2 = (2{,}31)^2 = 5{,}35$

Proportie onverklaarde variantie $= \frac{S_{y.x}^2}{S_y^2} = \frac{5{,}35}{6{,}69} = 0{,}7997$

$0{,}7997 \times 100\ \% = 79{,}97\%$

Als we even teruggaan naar opgave *d*, dan zien we dat er een gering verschil is in de proportie onverklaarde variantie met deze uitkomst. Dit verschil mogen we wijten aan afrondingsfouten.

6 Inleiding in hypothesen toetsen

In de voorafgaande hoofdstukken hebben we geleerd verdelingen te beschrijven op verscheidene kenmerken; al deze technieken vallen onder de beschrijvende of descriptieve statistiek. Vanaf dit hoofdstuk gaan we ons bezighouden met de *inductieve* statistiek, ofwel de statistiek die ons, door middel van een aantal technieken, in staat stelt om uitspraken te doen over een grote groep op basis van een kleine onderzochte groep.
In dit hoofdstuk zullen we een aantal algemene begrippen die bij de inductieve statistiek van belang zijn, uitleggen, zodat er een goede basis ontstaat om de hoofdstukken 7 tot en met 12 goed te kunnen begrijpen.
Alle toetsen die in de komende hoofdstukken behandeld worden, staan schematisch weergegeven in het toetsingsschema (appendix II).

6.1 Populatie en steekproef

In de korte inleiding van dit hoofdstuk hebben we al gezegd dat inductieve statistiek ons in staat stelt om uitspraken te doen over een grote groep op basis van een kleine onderzochte groep.
De grote groep noemen we de **populatie**, een verzameling van elementen die we willen onderzoeken op een bepaald kenmerk. De populatie is in de meeste gevallen veel te omvangrijk om te onderzoeken op belangrijke kenmerken; denk bijvoorbeeld aan de Nederlandse bevolking, de Nederlandse studenten of alle konijnen van de hele wereld. Indien we de populatie konijnen willen onderzoeken op bijvoorbeeld oorlengte, dan kunnen we onmogelijk alle konijnen van de wereld meten op dit kenmerk. Als we toch iets willen weten van de populatie, moeten we een kleine groep uit die populatie nemen en daarvan de oorlengte onderzoeken. Deze kleine groep die uit een populatie afkomstig is, noemen we een **steekproef**.
Met behulp van de inductieve statistiek kunnen we door middel van de gegevens uit de steekproef iets zeggen over de gehele, niet-onderzochte populatie. Dit proces noemen we (statistische) **inductie**. Bij de inductieve statistiek speelt de waarschijnlijkheid een belangrijke rol, want ondanks het feit dat de steekproef een klein deel van de populatie is, weet je nooit zeker of de situatie in de steekproef vergelijkbaar is met die van de gehele populatie.
Als we weer terug gaan naar de konijnen, dan kunnen zich in de steekproef 'per ongeluk' allemaal Vlaamse reuzen bevinden. De samenstelling van de steekproef wijkt dan af van de samenstelling van de populatie en daardoor zouden we een vertekend beeld van de werkelijkheid kunnen krijgen. Het is dus van groot belang om een goede (dat wil zeggen 'representatieve') steekproef uit de populatie te trekken!
De meest gangbare manier om een steekproef samen te stellen, is de aselecte steekproeftrekking. In een **aselecte steekproef** hebben alle elementen van de populatie een gelijke kans om in de steekproef te komen.

6.1.1 Kenmerken van populatie en steekproef

Zoals we hierboven al gezien hebben, gaan we kenmerken van de populatie schatten op basis van de kenmerken van de steekproef. De kenmerken, zoals we ze in de voorgaande hoofdstukken hebben geleerd, hebben een eigen notatievorm in steekproef en populatie.

De kenmerken van de populatie worden uitgedrukt in Griekse letters en worden **populatieparameters** genoemd. De kenmerken van de steekproef worden uitgedrukt in Romeinse letters en worden **statistische grootheden** genoemd. In tabel 6.1 staat een overzicht van de verschillende notatievormen voor dezelfde kenmerken.

Tabel 6.1

	Gemiddelde	*Standaard-afwijking*	*Variantie*	*Correlatie-coëfficiënt*
Populatie-verdeling	μ_x	σ_x	σ_x^2	ρ
Steekproef-verdeling	$\bar{x}_x$	S_x	S_x^2	r

6.2 Steekproevenverdelingen

6.2.1 De empirische steekproevenverdeling

Zelfs in het geval van aselecte steekproeftrekking bestaat het gevaar dat de samenstelling van de steekproef afwijkt van de populatiesamenstelling. Dit probleem is afhankelijk van toevalsfactoren die een rol spelen bij de aselecte steekproeftrekking. Het is daarom soms beter meerdere aselecte steekproeven te trekken, zodat extreme steekproeven het beeld van de werkelijkheid zo min mogelijk vervormen. Op de x-as zetten we dan alle voorkomende $\bar{x}_x$ (steekproefgemiddelden) en zo krijgen we dus een verdeling van gemiddelden van steekproeven. Gezien het feit dat we hier een steekproef uit alle mogelijke steekproeven nemen, worden de kenmerken van deze verdeling uitgedrukt in Romeinse letters. Het gemiddelde van deze verdeling berekenen we door alle voorkomende $\bar{x}_x$-waarden op te tellen en te delen door het aantal steekproeven dat we hebben getrokken: we vinden dan $\bar{x}_{\bar{x}}$. De spreiding van deze verdeling kunnen we berekenen door de S_x van één steekproef te delen door $\sqrt{n}$, waarbij n de grootte van de steekproef weergeeft. De standaardafwijking van een steekproevenverdeling wordt **standaardmeetfout** genoemd.

De verdeling die hierboven beschreven is, noemen we de **empirische steekproevenverdeling**.

6.2.2 De theoretische steekproevenverdeling

Deze verdeling gaat uit van hetzelfde principe als de empirische steekproevenverdeling, met dit verschil, dat we de de empirische wel kunnen opstellen en de theoretische steekproevenverdeling niet; deze bestaat namelijk uit alle $\bar{x}_x$-waarden afkomstig uit *alle* mogelijke steekproeven. Dit zou nog veel meer

werk zijn dan de gehele populatie onderzoeken op bepaalde kenmerken. Toch gebruiken we deze theoretische verdeling in de inductieve statistiek, maar dan bij het plaatsen van $\bar{x}$ binnen een steekproevenverdeling (via z-scores). Aangezien we bij deze theoretische verdeling werken met de *populatie* van steekproeven, worden bij deze verdeling de kenmerken uitgedrukt in populatieparameters. Het gemiddelde van deze theoretische steekproevenverdeling wordt genoteerd als $\mu_{\bar{x}}$ en is gelijk aan het gemiddelde van de populatie μ_x (we 'onderzoeken' immers alle steekproeven). De standaardmeetfout van deze verdeling wordt als volgt berekend:

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}$$

waarbij we dus de standaardafwijking van de populatieverdeling delen door de wortel uit de omvang van de steekproef.

In tabel 6.2 zien we een overzicht van de tot nu toe behandelde verdelingen, met hun kenmerken. We zien bij elke verdeling ook een z-score, zodat we elke waarde binnen welke verdeling dan ook relatief kunnen maken. In principe gebruiken we de z-score van de empirische steekproevenverdeling nooit, omdat deze verdeling de normale verdeling meestal niet benadert. Om een steekproefgemiddelde relatief te maken, gebruiken we dus meestal de z-score van de theoretische steekproevenverdeling.

Tabel 6.2

	Gemiddelde	*Standaard-afwijking*	*Variantie*	*z-score*
Populatieverdeling *(verdeling van de oorspronkelijke populatie)*	μ_x	σ_x	σ_x^2	$z = \frac{x - \mu_x}{\sigma_x}$
Steekproefverdeling *(verdeling van één steekproef)*	$\bar{x}_x$	S_x	S_x^2	$z = \frac{x - \bar{x}_x}{S_x}$
Empirische steekproevenverdeling *(verdeling van meerdere steekproeven)*	$\bar{x}_{\bar{x}}$	$S_{\bar{x}}$	$S_{\bar{x}}^2$	$z = \frac{\bar{x} - \bar{x}_{\bar{x}}}{S_{\bar{x}}}$
Theoretische steekproevenverdeling *(verdeling van populatie steekproeven)*	$\mu_{\bar{x}}$	$\sigma_{\bar{x}}$	$\sigma_{\bar{x}}^2$	$z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}}$

Aan de hand van het volgende voorbeeld zullen we het een en ander duidelijk maken.

In groep 8 van een basisschool in Rotterdam zitten 25 leerlingen. We beschouwen deze groep als onze populatie. Deze kinderen geven we een biologieles, waarin 10 soorten bloemen worden behandeld. Door middel van een toets na de biologieles wordt gekeken, hoeveel bloemen de kinderen nog herkennen. Hieronder staan de gegevens van deze klas.

2	4	8	7	2
2	3	7	5	5
3	2	4	10	3
1	3	8	5	10
7	9	6	2	0

De kenmerken van deze populatie kunnen we als volgt berekenen:

$$\mu_x = \frac{\sum x}{N} = \frac{118}{25} = 4{,}72$$

$$\sigma_x^2 = \frac{n \cdot \sum x^2 - (\sum x)^2}{n(n-1)} = \frac{25 \cdot 756 - (118)^2}{25(24)} = \frac{18900 - 13924}{600} = \frac{4976}{600}$$

$$= 8{,}29$$

$$\sigma_x = \sqrt{\sigma_x^2} = \sqrt{8{,}29} = 2{,}88$$

Stel dat we iets over de populatie willen zeggen op basis van een steekproef. We trekken aselect (dat wil zeggen op basis van een toevalsmechanisme) een steekproef van 4 kinderen:
kind $1 \rightarrow 2$
kind $5 \rightarrow 7$
kind $13 \rightarrow 4$
kind $20 \rightarrow 2$

De kenmerken van deze steekproef zien er als volgt uit:

$$\bar{x} = \frac{\sum x}{n} = \frac{15}{4} = 3{,}75$$

We zouden op basis van deze verdeling het gemiddeld aantal bloemen dat de kinderen in de populatie onthouden, schatten op 3,75, wat bijna 1 te laag is.

$$S_x^2 = \frac{n \cdot \sum x^2 - (\sum x)^2}{n(n-1)} = \frac{4 \cdot 73 - (15)^2}{4(3)} = \frac{292 - 225}{12} = \frac{67}{12} = 5{,}58$$

$$S_x = \sqrt{S_x^2} = \sqrt{5{,}58} = 2{,}36$$

We zien dat het gemiddelde uit bovenstaande steekproef niet echt overeenkomt met het populatiegemiddelde. Om een minder vertekend beeld van de werkelijkheid te krijgen, voeren we meerdere steekproeven uit:

Steekproef 2
kind 7 → 3
kind 12 → 7
kind 14 → 8
kind 19 → 5

$$\bar{x}_{\text{steekproef 2}} = \frac{\sum x}{n} = \frac{23}{4} = 5{,}75$$

Steekproef 3
kind 16 → 7
kind 25 → 0
kind 3 → 3
kind 22 → 5

$$\bar{x}_{\text{steekproef 3}} = \frac{\sum x}{n} = \frac{15}{4} = 3{,}75$$

Steekproef 4
kind 24 → 10
kind 18 → 10
kind 4 → 1
kind 20 → 2

$$\bar{x}_{\text{steekproef 4}} = \frac{\sum x}{n} = \frac{23}{4} = 5{,}75$$

Steekproef 5
kind 7 → 3
kind 11 → 8
kind 4 → 1
kind 23 → 3

$$\bar{x}_{\text{steekproef 5}} = \frac{\sum x}{n} = \frac{15}{4} = 3{,}75$$

Als we nu alle gevonden gegevens combineren, dan hebben we een empirische steekproevenverdeling met de volgende kenmerken:

$$\mu_{\bar{x}} = \frac{\sum \bar{x}}{n} = \frac{3{,}75 + 5{,}75 + 3{,}75 + 5{,}75 + 3{,}75}{5} = \frac{22{,}75}{5} = 4{,}55$$

Duidelijk is, dat dit gemiddelde dichter bij het populatiegemiddelde ligt, het vormt dus een betere schatting van het populatiegemiddelde.
De spreiding van deze verdeling, uitgedrukt door middel van de standaardmeetfout, is natuurlijk een stuk kleiner dan die van de steekproefverdeling of van de populatieverdeling; we kunnen de standaardmeetfout schatten op basis van de volgende formule:

$$S_{\bar{x}} = \frac{S_x}{\sqrt{n}}$$

We nemen dan de standaardafwijking van één van de steekproeven, bijvoorbeeld van steekproef 1, $S_x = 2{,}36$

$$S_{\bar{x}} = \frac{S_x}{\sqrt{n}} = \frac{2{,}36}{\sqrt{4}} = 1{,}18$$

Gemiddeld schatten we dus dat de gemiddelden in de empirische steekproevenverdeling 1,18 afwijken van $\bar{x}_{\bar{x}}$.

Zoals we aankondigden, hebben we voor iedere verdeling een z-score. In dit voorbeeld zullen we het gebruik daarvan toelichten in de populatieverdeling, steekproefverdeling en de theoretische steekproevenverdeling. Het gebruik van de z-score in de empirische steekproevenverdeling laten we achterwege, omdat deze meestal de normale verdeling niet benadert.

a Gebruik van z-score in de populatieverdeling

Hoeveel procent van alle kinderen in groep 8 kunnen maar 3 of minder bloemen onderscheiden?

$\mu_x = 4{,}72$ De kinderen onthouden dus gemiddeld 4,72 bloemen.
$\sigma_x = 2{,}88$ Men wijkt gemiddeld 2,88 af van het gemiddelde.
$x = 3$ De te standaardiseren x-waarde.

$$z = \frac{x - \mu_x}{\sigma_x} = \frac{3 - 4{,}72}{2{,}88} = \frac{-1{,}72}{2{,}88} = -0{,}60$$

In tabel A zoeken we het gebied van $-0{,}60$ naar negatief oneindig, we spiegelen naar positief, dus we kijken naar het gebied van $+0{,}60$ naar positief oneindig, we kijken dus in de tweede gebiedskolom. We vinden dan een percentage van 27,43%.

b Gebruik van z-score in de steekproefverdeling

Hoeveel procent van de kinderen in de eerste steekproef herkent 3 of minder bloemen?

$\bar{x} = 3{,}75$ De kinderen in deze steekproef kennen dus gemiddeld 3,75 bloemen.
$S_x = 2{,}36$ De kinderen wijken gemiddeld 2,36 af van het gemiddelde van de steekproef.
$x = 3$ De te standaardiseren x-waarde.

$$z = \frac{x - \bar{x}_x}{S_x} = \frac{3 - 3{,}75}{2{,}36} = \frac{-0{,}75}{2{,}36} = -0{,}32$$

In tabel A kijken we in de tweede gebiedskolom naar het gebied van $+0{,}32$ naar positief oneindig (gespiegeld!). Het bijbehorende percentage is 37,45%. We zien dat dit percentage aanzienlijk hoger is dan in de werkelijk-

heid (de populatie); waarschijnlijk komt dat doordat de kinderen in de steekproef relatief weinig bloemen kunnen onthouden.

c *Gebruik van z-score in de theoretische steekproevenverdeling*
Hoe groot is de kans op een steekproefgemiddelde van 3,75 of lager?
We gaan hier niet een ruwe score (X), maar een steekproefgemiddelde standaardiseren. Hiertoe gebruiken we de theoretische steekproevenverdeling en de bijbehorende z-score. Zo kunnen we gaan kijken wat de kans is op een steekproefgemiddelde van 3,75 of lager.

$\mu_{\bar{x}} = \mu_x = 4{,}72$

$\sigma_x = 2{,}88$

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}} = \frac{2{,}88}{\sqrt{4}} = \frac{2{,}88}{2} = 1{,}44$$

$$z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{3{,}75 - 4{,}72}{1{,}44} = \frac{-0{,}97}{1{,}44} = -0{,}67$$

In tabel A kijken we naar het gebied van +0,67 naar positief oneindig (gespiegeld) in de tweede gebiedskolom; we vinden dan het antwoord op de vraag die in C is gesteld. De kans op een steekproefgemiddelde van 3,75 of lager is gelijk aan 25,14%.

6.3 Toetsen van hypothesen

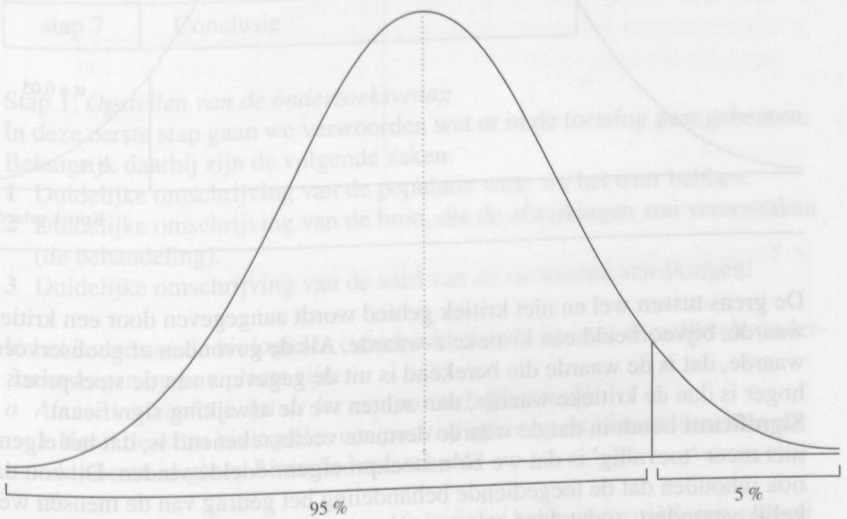
Soms hebben we aanleiding te vermoeden dat een steekproefgemiddelde afwijkt van de populatie waar de steekproef uit getrokken is. Die afwijking kan bijvoorbeeld als oorzaak hebben dat we de subjecten in de steekproef een bepaalde behandeling hebben toegediend, waardoor hun gedrag zou afwijken van dat van de populatie die de behandeling niet ontvangen heeft. Om te kijken of er inderdaad een afwijking bestaat, kunnen we hypothesen gaan toetsen.
Bij het toetsen van hypothesen gaat het erom, dat we *met een bepaalde waarschijnlijkheid uitspraken kunnen doen over de verwachte afwijking van het steekproefgemiddelde.*
We kunnen echter ook een bepaalde overeenkomst verwachten of een verschil tussen twee of meer groepen of andere zaken aan toetsing onderwerpen. Het hierboven gegeven uitgangspunt is slechts één van de vele vragen die we zouden kunnen stellen bij het toetsen van hypothesen.
Waar het in dit hoofdstuk om gaat, is een kader te bieden voor de toetsing van hypothesen, waar we in de rest van de hoofdstukken profijt van zullen hebben.

6.3.1 Waarschijnlijkheid

Om te kunnen vaststellen of een gevonden steekproefgemiddelde afwijkt van een populatie, hanteren we het begrip **waarschijnlijkheid**. We vertalen deze waarschijnlijkheid in een percentage dat overeenkomt met een bepaald deel van de populatieverdeling. Daarbij trekken we een grens tussen 'wel of niet

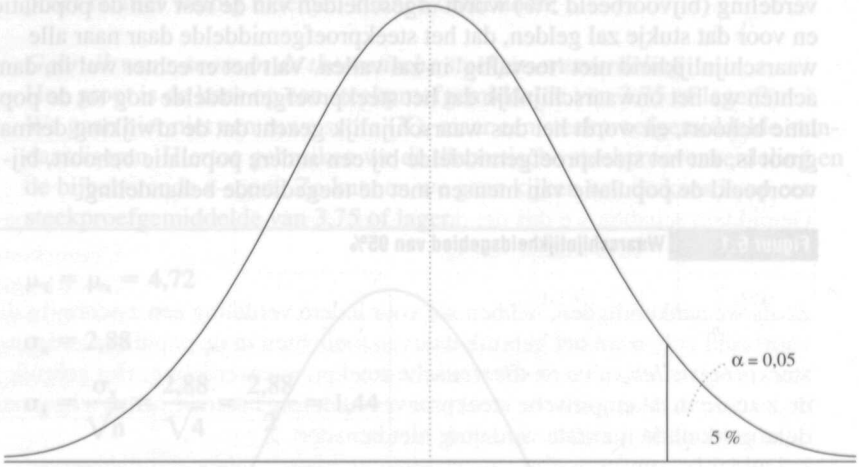
waarschijnlijk' binnen de populatie. Een klein (extreem) deel van de populatieverdeling (bijvoorbeeld 5%) wordt afgescheiden van de rest van de populatie en voor dat stukje zal gelden, dat het steekproefgemiddelde daar naar alle waarschijnlijkheid niet 'toevallig' in zal vallen. Valt het er echter wel in, dan achten we het onwaarschijnlijk dat het steekproefgemiddelde nog tot de populatie behoort, en wordt het dus waarschijnlijk geacht dat de afwijking dermate groot is, dat het steekproefgemiddelde bij een andere populatie behoort, bijvoorbeeld de populatie van mensen met de toegediende behandeling.

Figuur 6.1 **Waarschijnlijkheidsgebied van 95%**



Het gebied waarvoor geldt dat we het onwaarschijnlijk achten dat het steekproefgemiddelde daar 'toevallig' in zal vallen, noemen we het **kritieke gebied**. De grootte van dit kritieke gebied is gelijk aan het percentage dat hoort bij het 'extreme' gedeelte van de populatieverdeling. We noemen dit percentage het **significantieniveau**. Het significantieniveau wordt gesymboliseerd door α. Is het gekozen percentage 5%, dan hanteren we dus een significantieniveau van 5%, en dat wordt genoteerd als $\alpha = 0{,}05$.

Figuur 6.2 Kritiek gebied van 5%



De grens tussen wel en niet kritiek gebied wordt aangegeven door een kritieke waarde, bijvoorbeeld een kritieke z-waarde. Als de gevonden of geobserveerde waarde, dat is de waarde die berekend is uit de gegevens van de steekproef, hoger is dan de kritieke waarde, dan achten we de afwijking significant. **Significant** houdt in dat de waarde dermate veelbetekenend is, dat het eigenlijk niet meer 'toevallig' is dat we zo'n steekproefgemiddelde vinden. Dit zou dan ook inhouden dat de toegediende behandeling het gedrag van de mensen werkelijk verandert, zodat deze mensen niet meer tot de populatie behoren van mensen die de behandeling niet toegediend hebben gekregen. Dit houdt ook in dat we, met een bepaalde waarschijnlijkheid, uitspraken kunnen doen over personen buiten de steekproef, namelijk, dat alle mensen die de behandeling wel toegediend krijgen, zich (waarschijnlijk) anders gedragen dan degenen die haar niet ontvangen.

6.3.2 *Toetsingsschema*

Hoewel we de verschillende toetsen in de volgende hoofdstukken zullen behandelen, geven we hier een algemeen toetsingsschema weer, waar we bij elke toets van uit zullen gaan. Als uitgangspunt kiezen we bijvoorbeeld de biologieles over bloemen; dat noemen we de toegediende behandeling. Voorafgaand aan dit toetsingsschema moeten we dus eerst een keuze voor een bepaalde toets maken. De keuze is afhankelijk van de context waarbinnen het toetsingsprobleem zich bevindt en natuurlijk ook van welke vragen er gesteld worden. In elk volgend hoofdstuk zullen we de context duidelijk aangeven. In de volgende paragrafen wordt elke stap nader toegelicht.

Toetsingsschema bij het toetsen van hypothesen

stap 1	Opstellen van de onderzoeksvraag
stap 2	Opstellen van de hypothesen
stap 3	Opstellen van de aannamen
stap 4	Opstellen van de beslissingsregels
stap 5	Berekening
stap 6	Beslissing
stap 7	Conclusie

Stap 1: *Opstellen van de onderzoeksvraag*
In deze eerste stap gaan we verwoorden wat er in de toetsing gaat gebeuren. Belangrijk daarbij zijn de volgende zaken:

1 Duidelijke omschrijving van de populatie waar we het over hebben.
2 Duidelijke omschrijving van de bron, die de afwijkingen zou veroorzaken (de behandeling).
3 Duidelijke omschrijving van de aard van de verwachte afwijkingen.

In het kader van de biologieles (voorbeeld 1 en 2) zouden de volgende onderzoeksvragen op hun plaats zijn:

a Als we verwachten dat de les een positief effect heeft:

1 Heeft de extra biologieles een *positief effect* op de bloemenkennis van kinderen in groep 8 van de basisschool?

of:

2 Hebben kinderen die een extra biologieles hebben gevolgd, een *grotere* kennis van bloemen dan kinderen die de biologieles niet gevolgd hebben?

b Als we verwachten dat de les effect heeft, positief of negatief:

3 Heeft de biologieles een *effect* op de bloemenkennis van kinderen in groep 8 van de basisschool?

of:

4 Kennen kinderen die de biologieles hebben gevolgd, een *ander* aantal bloemen dan kinderen die de les niet gevolgd hebben?

c Als we verwachten dat de biologieles een negatief effect heeft:

5 Heeft de biologieles een *negatief effect* op de bloemenkennis van kinderen uit groep 8 van de basisschool?

of:

6 Herkennen de kinderen die de biologieles over bloemen gehad hebben, minder bloemen dan de kinderen die de les niet hebben gevolgd?

Natuurlijk zijn er heel wat andere manieren om de onderzoeksvraag te verwoorden. In de volgende hoofdstukken zullen we trachten zo veel mogelijk variatie daarin aan te brengen.

Stap 2: *Opstellen van de hypothesen.*
Bij het toetsen van hypothesen gaan we altijd uit van *twee* hypothesen, de nulhypothese (H_0) en de alternatieve hypothese (H_1).
In de alternatieve hypothese formuleren we altijd wat we verwachten, dat waarvan we hopen dat het waar is. In de nulhypothese komt het tegenovergestelde te staan.
Uiteindelijk willen we dus de H_0 verwerpen en de H_1 aannemen, want dan zou onze verwachting immers worden bevestigd.

Er zijn twee manieren waarop we de hypothesen kunnen opstellen:
1 Hypothesen in termen van de onderzoeksvraag
2 Hypothesen in statistische termen (waarbij altijd populatieparameters gebruikt worden).

Hieronder zullen we de hypothesen opstellen die horen bij de drie typen onderzoeksvragen.

a Als we verwachten dat de les een positief effect heeft:

Hypothesen in termen van de onderzoeksvraag
H_0: De kinderen die de extra biologieles hebben gehad, kennen minder of evenveel bloemen als de kinderen die deze les niet hebben gehad.
H_1: De kinderen die de extra biologieles hebben gehad, kennen meer bloemen dan de kinderen die deze les niet hebben gehad.
Hypothesen in statistische termen
H_0: $\mu_{\text{behandelde groep}} \leq \mu_{\text{onbehandelde groep}}$
H_1: $\mu_{\text{behandelde groep}} > \mu_{\text{onbehandelde groep}}$

In de alternatieve hypothese zeggen we dat het populatiegemiddelde van de behandelde groep hoger ligt dan het populatiegemiddelde van de niet-behandelde groep. We gaan daar dus uit van een andere populatie met een hoger populatiegemiddelde. In de nulhypothese zien we het tegenovergestelde van de alternatieve hypothese staan, datgene wat we niet verwachten.

b Als we verwachten dat de les effect heeft, positief of negatief:

Hypothesen in termen van de onderzoeksvraag
H_0: De kinderen die de extra biologieles hebben gehad, kennen evenveel bloemen als de kinderen die deze les niet hebben gehad.
H_1: De kinderen die de extra biologieles hebben gehad, kennen een ander aantal bloemen dan de kinderen die deze les niet hebben gehad.
Hypothesen in statistische termen
H_0: $\mu_{\text{behandelde groep}} = \mu_{\text{onbehandelde groep}}$
H_1: $\mu_{\text{behandelde groep}} \neq \mu_{\text{onbehandelde groep}}$

Bij deze hypothesen laten we open in welke richting de afwijking gaat, de les kan een positief of negatief effect hebben; we verwachten in ieder geval dat de populatie van de behandelde groep een ander populatiegemiddelde zal hebben dan de niet-behandelde groep.

c Als we verwachten dat de biologieles een negatief effect heeft:

Hypothesen in termen van de onderzoeksvraag
H_0: De kinderen die de extra biologieles hebben gehad, kennen evenveel of meer bloemen dan de kinderen die deze les niet hebben gehad.
H_1: De kinderen die de extra biologieles hebben gehad, kennen minder bloemen dan de kinderen die deze les niet hebben gehad.
Hypothesen in statistische termen
H_0: $\mu_{\text{behandelde groep}} \geq \mu_{\text{onbehandelde groep}}$
H_1: $\mu_{\text{behandelde groep}} < \mu_{\text{onbehandelde groep}}$

Ten slotte nog twee regels waar we ons altijd aan moeten houden bij het toetsen van hypothesen, namelijk:

1 De hypothesen moeten elkaar volledig uitsluiten.
2 Er mag nooit een $=$ in de alternatieve hypothese voorkomen.

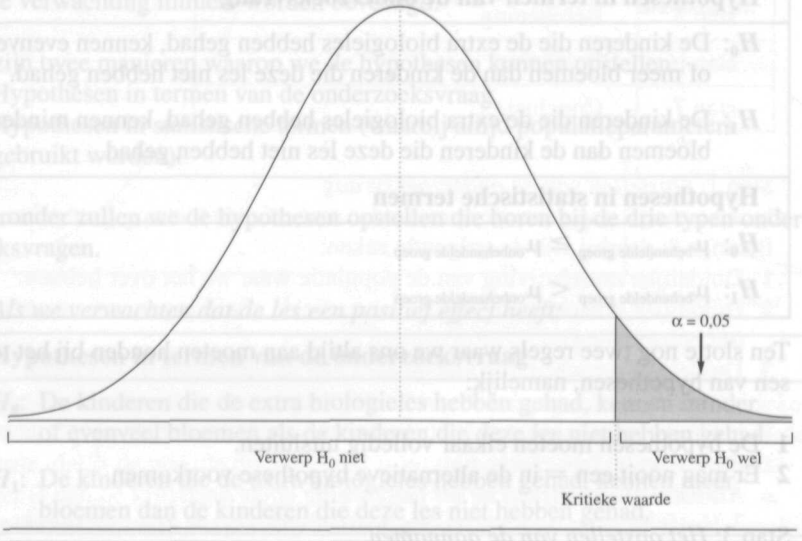
Stap 3: *Het opstellen van de aannamen.*
Elke toets wordt uitgevoerd op basis van een aantal stringente aannamen, bijvoorbeeld, dat de verdeling die we toetsen, normaal van vorm is. Bij elke toets zijn de aannamen weer anders doordat de context waarbinnen de toets wordt uitgevoerd steeds weer anders is. Bij elke toets worden de aannamen genoemd. We kunnen alleen op de resultaten van de toets vertrouwen als er vooraf aan de aannamen voldaan is. Het is daarom van groot belang, voordat men aan de eigenlijke toets begint, te controleren of er wel aan de aannamen van de toets voldaan is.

Stap 4: *Het opstellen van de beslissingsregels*
Deze stap heeft te maken met het afbakenen van het kritieke gebied, het bepalen van een kritieke waarde en het koppelen van regels daaraan, die iets zeggen over *wanneer* we tot *welke* beslissing overgaan. We zullen deze stap in zeer algemene termen behandelen.
Allereerst moeten we de waarden van het kritieke gebied vaststellen, meestal heeft het kritieke gebied een oppervlakte van 5% ($\alpha = 0{,}05$), maar men kan ook kiezen voor 1% ($\alpha = 0{,}01$) of een ander percentage. Wij gaan hier uit van een significantieniveau van 5%, dus een α van 0,05.
Bij elk van de drie onderzoeksvragen zullen we deze stap maken. De beslissingsregels worden in algemene termen opgesteld.

a Als we verwachten dat de les een positief effect heeft:

Gegevens: $\alpha = 0{,}05$
We toetsen een positief effect, we leggen deze α dus in de uiterste staart naar positief oneindig → we toetsen rechtseenzijdig.

Figuur 6.3 **Rechtseenzijdig verwerpingsgebied**



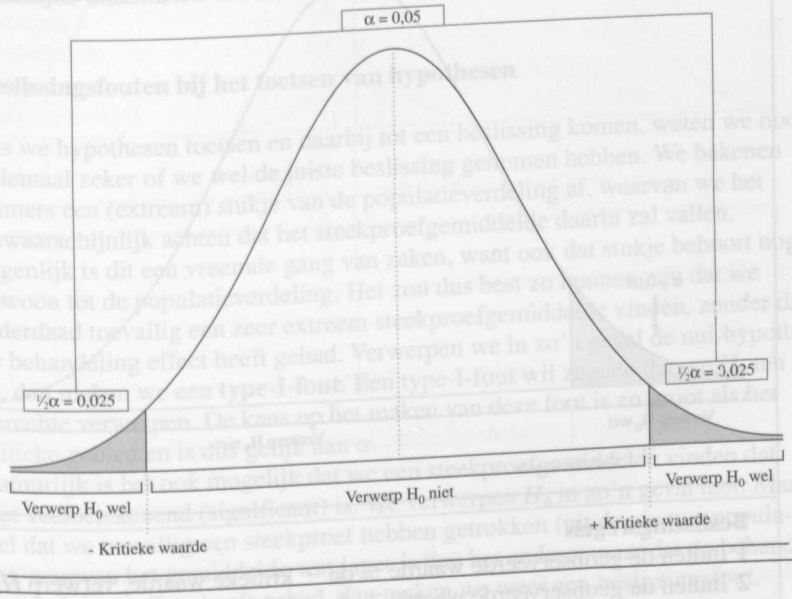
Beslissingsregels
1 Indien de geobserveerde waarde ≥ de kritieke waarde, verwerp $\boldsymbol{H_0}$ wel.
2 Indien de geobserveerde waarde < de kritieke waarde, verwerp $\boldsymbol{H_0}$ niet.

Met de geobserveerde waarde bedoelen we de waarde die straks naar voren komt uit de berekening.

b Als we verwachten dat de les effect heeft, positief of negatief:

Gegevens: $\alpha = 0{,}05$
We toetsen een positief of een negatief effect, we splitsen deze α dus naar twee kanten, de ene helft in de linkerstaart en de andere helft in de rechterstaart. Met andere woorden: we toetsen tweezijdig.

Figuur 6.4 **Tweezijdig verwerpingsgebied**



Beslissingsregels

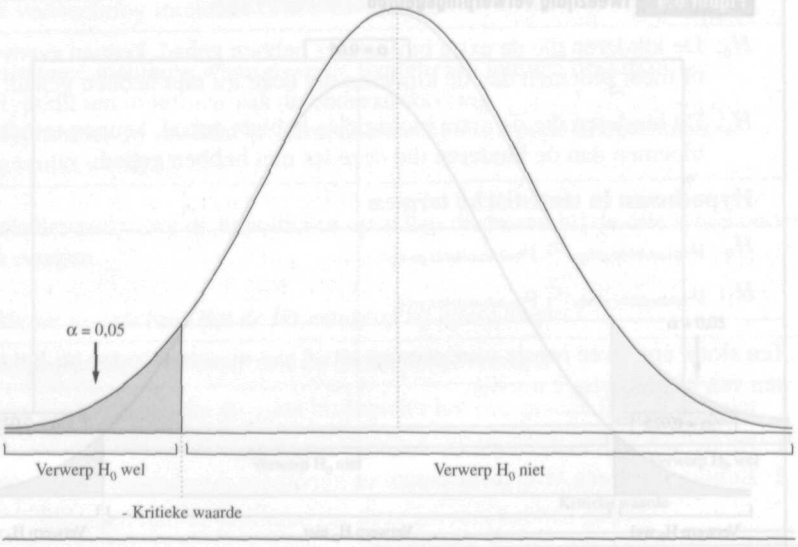
1 Indien de geobserveerde waarde $\geq$ de $+$kritieke waarde, of indien de geobserveerde waarde $\leq$ de $-$kritieke waarde, verwerp $\boldsymbol{H_0}$ wel.

2 Indien de $-$kritieke waarde $<$ de geobserveerde waarde $<$ de $+$kritieke waarde, verwerp $\boldsymbol{H_0}$ niet.

c Als we verwachten dat de biologieles een negatief effect heeft:

Gegevens: $\alpha = 0{,}05$
We toetsen een negatief effect, we leggen deze α dus in de uiterste staart naar negatief oneindig. Met andere woorden: we toetsen **linkseenzijdig**.

Figuur 6.5 **Linkszijdig verwerpingsgebied**



Beslissingsregels
1 Indien de geobserveerde waarde $\leq$ de $-$ kritieke waarde, verwerp H_0.
2 Indien de geobserveerde waarde $>$ de $-$ kritieke waarde, verwerp H_0 niet.

Ten slotte ook nog een regel die bij het opstellen van de beslissingsregels in de gaten gehouden moet worden: de kritieke waarde zelf hoort al bij het kritieke gebied, dus als de geobserveerde waarde gelijk is aan de kritieke waarde, dan verwerpen we de nulhypothese ook.

Stap 5: *De berekening*
Bij elke toets gelden andere berekeningen, doordat de context van elke toets anders is. We kunnen transformeren naar een z-verdeling, maar er is nog een scala aan andere verdelingen waar we naar kunnen transformeren.
Het resultaat van een berekening is een geobserveerde waarde.

Stap 6: *De beslissing*
In deze fase van het toetsen van hypothesen gaan we de geobserveerde en de kritieke waarde met elkaar vergelijken zoals dat in de beslissingsregels is vastgesteld.

Er zijn twee beslissingen mogelijk:

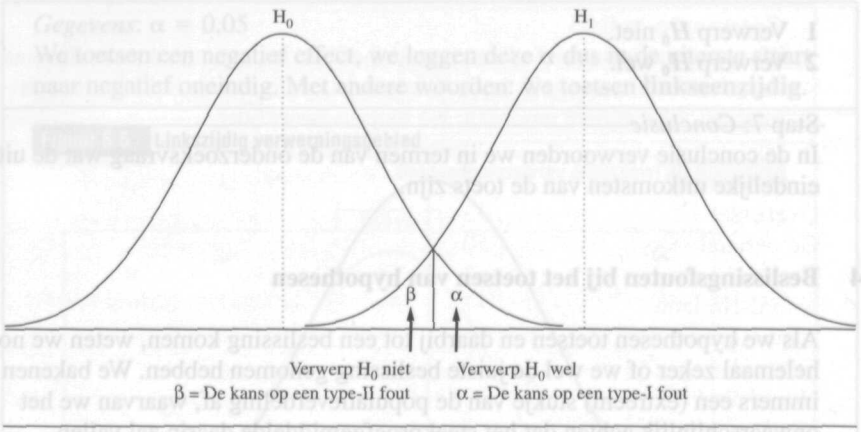
1 Verwerp H_0 niet.
2 Verwerp H_0 wel.

Stap 7: *Conclusie*
In de conclusie verwoorden we in termen van de onderzoeksvraag wat de uiteindelijke uitkomsten van de toets zijn.

6.4 Beslissingsfouten bij het toetsen van hypothesen

Als we hypothesen toetsen en daarbij tot een beslissing komen, weten we nooit helemaal zeker of we wel de juiste beslissing genomen hebben. We bakenen immers een (extreem) stukje van de populatieverdeling af, waarvan we het onwaarschijnlijk achten dat het steekproefgemiddelde daarin zal vallen. Eigenlijk is dit een vreemde gang van zaken, want ook dat stukje behoort nog gewoon tot de populatieverdeling. Het zou dus best zo kunnen zijn dat we inderdaad toevallig een zeer extreem steekproefgemiddelde vinden, zonder dat de behandeling effect heeft gehad. Verwerpen we in zo'n geval de nul-hypothese, dan maken we een **type-I-fout**. Een type-I-fout wil zeggen dat we H_0 ten onrechte verwerpen. De kans op het maken van deze fout is zo groot als het kritieke gebied en is dus gelijk aan α.
Natuurlijk is het ook mogelijk dat we een steekproefgemiddelde vinden dat niet veelbetekenend (significant) is. We verwerpen H_0 in zo'n geval niet. Maar stel dat we toevallig een steekproef hebben getrokken (uit de nieuwe populatie), waarvan het gemiddelde wat lager is dan het andere, maar dat de behandeling toch haar effect heeft gehad, dan maken we weer een beslissingsfout, namelijk een **type-II-fout**. Deze fout houdt in dat we de nulhypothese ten onrechte aanhouden. De kans op het maken van deze fout is zo groot als β. In figuur 6.6 zien we twee populatieverdelingen die elkaar overlappen, de populatie van een nulhypothese en die van een alternatieve hypothese. In deze figuur zijn de fouten aangegeven bij een α van 0,05. De moeilijkheid van deze fouten ligt in het feit dat de beide populaties elkaar overlappen en dat we er nooit achter komen of we één van deze fouten maken. Het is echter van groot belang dat we altijd rekening houden met de *kans* op het maken van deze fouten.
Tevens moet men rekening houden met het evenwicht tussen de kans op een type-I-fout en de kans op een type-II-fout: als we α kleiner maken, bijvoorbeeld 0,01, dan wordt automatisch het gebied van β groter. Daarom heeft men een soort 'standaardwaarde' voor α, namelijk 0,05.

Figuur 6.6 **Gebieden van beslissingsfouten**



Opgaven

6.1 Hieronder is gegeven de oppervlakte in km^2 van de Nederlandse provincies*:

Provincie	*oppervlakte ($\times$ 100 km^2)*
Groningen	23
Friesland	34
Drenthe	26
Overijssel	33
Gelderland	50
Utrecht	13
Noord-Holland	27
Zuid-Holland	29
Zeeland	18
Noord-Brabant	50
Limburg	22

* Sinds Flevoland een provincie is geworden, kloppen de gegevens niet helemaal meer.

a Wat lijkt hier de relevante variabele? (Noem deze verder x.)
b Wat is het meetniveau?
c Wat is de populatie waartoe de variabele behoort?
d Wat is de gemiddelde grootte van de Nederlandse provincie?
e Met welk symbool kunnen we dit gemiddelde weergeven?

6.2 *a* Bedenk een mogelijkheid om vier steekproeven van drie provincies te trekken en voer deze trekking uit. Provincies mogen wel meerdere malen in de vier steekproeven voorkomen, maar streef ernaar *per steekproef drie verschillende provincies* op te nemen.
b Wat is de schatting voor de gemiddelde grootte van de Nederlandse provincies op grond van de eerste steekproef (gebruik daarbij de juiste notatie)?
c Voer *b* uit voor de overige drie steekproeven.
d Vergelijk de vier uitkomsten met de oplossing van opdracht 6.1.*d*.
e Bedenk een oplossing, om juist die steekproef te trekken waarvan het gemiddelde het dichtst bij het gemiddelde van 6.1.*d* ligt.
f Wat is op grond van de vier steekproeven de beste schatting die u kunt maken voor de gemiddelde provinciegrootte? Is dit μ_x, $\mu_{\bar{x}}$, $\bar{x}$ of $\bar{x}_{\bar{x}}$?
g Hoe noemt men de variabele uit *f*? Hoe heet de verdeling?
h Stel: men berekent alle gemiddelden van alle mogelijke steekproeven van drie provincies uit de populatie (dit zijn meer dan 160 mogelijke steekproeven!). Als men nu van deze steekproefgemiddelden het gemiddelde berekent, welke waarde vindt men dan? Is dit μ_x, $\mu_{\bar{x}}$, $\bar{x}$, $\bar{x}_{\bar{x}}$?

6.3 Bekijkt men de oppervlakten van de elf verschillende provincies, dan is te zien dat het verschil tussen Utrecht en Noord-Brabant aanzienlijk is. Welke grootheid geeft een goede maat om de verschillen in oppervlakte tussen de provincies uit te drukken? Is dit een steekproefgrootheid of een populatieparameter?

Neem Zuid-Holland, Noord-Holland en Drenthe.
Wat is S_x^2? Wat is $S_{\bar{x}}^2$?

Neem Utrecht, Gelderland en Drenthe.
Wat is S_x^2? Wat is $S_{\bar{x}}^2$?

Hoe is deze discrepantie tussen de schattingen te verklaren? Hoe kun je ervoor zorgen dat deze schattingen minder ver uiteen liggen?

6.4 Neem de steekproef 'Utrecht, Limburg, Zeeland'. Bereken het gemiddelde oppervlak. Wat is de kans op een zo klein of kleiner $\bar{x}$?
Normaal gesproken acht men die kans te gering om vol te houden dat de steekproef 'Utrecht, Limburg, Zeeland' uit de populatie 'Nederlandse provincies' is getrokken. Men maakt dan een fout. Dankzij deze fout kan men met behulp van toetsende statistiek nooit iets *bewijzen*, doch hooguit (zeer) aannemelijk maken.

6.5 Veel vaktijdschriften zullen geen verslagen van experimenten publiceren, waarvan niet ten minste enkele resultaten significant zijn.

a Als u zelf redacteur van zo'n blad zou zijn, in hoeverre kunt u dan het standpunt van uw collega's onderschrijven?

b Stel dat uit een onderzoek blijkt dat er geen significant verschil tussen twee experimentele groepen te vinden is. Welk voordeel kan het hebben om dit resultaat toch te publiceren?

c Bedenk een voorbeeld waarbij resultaten die op $\alpha = 0{,}10$ significant zijn, meer betekenis kunnen hebben dan andere resultaten die significant zijn op $\alpha = 0{,}001$.

d Bedenk enkele voorbeelden van statistisch zeer significante maar irrelevante verschillen.

6.6 Een onderzoekster vergelijkt het aantal gaatjes in het gebit van kinderen die een jaar lang met Nocarie-tandpasta (= A) en Autoplombum (= B) hebben gepoetst. Ze vindt dat groep B significant minder gaatjes heeft dan groep A ($\alpha = 0{,}001$).

a Stel een nul- en een alternatieve hypothese op.

b Wat betekenen hier type-I- en type-II-fouten?

c Wat is de kans op een type-I-fout?

d Als de onderzoeker $\alpha = 0{,}01$ als significantieniveau had gesteld, zou dan de kans op een type-II-fout gewijzigd zijn? Zo ja, hoe?

e Is het dus slim om alfa altijd zo klein mogelijk te kiezen?

6.7 Hoffmann-La Roche S.A. wil een valiumpreparaat fabriceren. Dit preparaat is bedoeld om epileptische aanvallen te onderdrukken wanneer de geëigende geneesmiddelen om de aanval te voorkomen niet hebben gewerkt of niet zijn ingenomen. Alvorens het middel op de markt te brengen, wordt onderzoek gedaan, waarbij de volgende overwegingen een rol spelen:

1 Het verwachte financiële rendement (Wat brengt het op?).
2 Eventuele neveneffecten op de gezondheid.
3 De effectiviteit van het middel.

Bespreek de consequenties van type-I- en type-II-fouten. Wat kunnen deze fouten in de praktijk betekenen?

Uitwerkingen

6.1 *a* Oppervlakte van de Nederlandse provincies in 100 km^2.

b Ratiomeetniveau.

c Oppervlakte der Nederlandse provincies.

d $\mu_x = \dfrac{325}{11} = 29{,}55\ (\cdot\ 100\ \text{km}^2) = 2955\ \text{km}^2$

e μ_x

6.2 *a* **Bijvoorbeeld**: nummer de provincies als volgt:

Groningen: 01
Friesland: 02
Drenthe: 03
Overijssel: 04
Gelderland: 05
Utrecht: 06
N.Holland: 07
Z.Holland: 08
Zeeland: 09
N.Brabant: 10
Limburg: 11

Doe van elke provincie een briefje in een bak en trek blind steeds (met terugleggen) een briefje uit de bak:

De eerste steekproef is dan: 03 09 02
De tweede steekproef is dan: 04 02 10
De derde steekproef is dan: 06 08 09
De vierde steekproef is dan: 08 06 02

Dus: de provincies Groningen, Gelderland, Noord-Holland en Limburg zijn niet in één van de vier aselecte steekproeven terechtgekomen, de overige zeven wel.
Het kan natuurlijk heel goed dat de steekproef steeds weer anders is, deze getallen zijn ook op aselecte, toevallige wijze totstandgekomen.

b $\bar{x}_{\text{steekproef 1}} = \dfrac{26 + 18 + 34}{3} = 26\ (\cdot\ 100\ \text{km}^2) = 2600\ \text{km}^2$

c $\bar{x}_{\text{steekproef 2}} = \dfrac{33 + 34 + 50}{3} = 39\ (\cdot\ 100\ \text{km}^2) = 3900\ \text{km}^2$

$\bar{x}_{\text{steekproef 3}} = \dfrac{13 + 29 + 18}{3} = 20\ (\cdot\ 100\ \text{km}^2) = 2000\ \text{km}^2$

$$\bar{x}_{\text{steekproef 4}} = \frac{29 + 13 + 34}{3} = 25{,}33\ (\cdot 100\ \text{km}^2) = 2533\ \text{km}^2$$

d We zien dat de steekproefgemiddelden nogal verschillen: steekproef 3 zou de gemiddelde opppervlakte van de Nederlandse provincies veel te laag schatten en steekproef 2 geeft weer een veel te hoge schatting...

e Is niet mogelijk: $\bar{x}$is een *stochast*, een toevalsvariabele; de waarde van $\bar{x}$ zal het populatiegemiddelde wél dichter naderen naarmate de steekproefomvang toeneemt. In dit opzicht is een steekproef van zeven provincies beter dan één van twee of drie provincies.

f $\bar{x}_{\bar{x}} = \dfrac{26 + 39 + 20 + 25{,}33}{4} = 27{,}58\ (\cdot 100\ \text{km}^2) = 2758\ \text{km}^2$

g Gemiddelde provincieoppervlakte (op basis van vier steekproefgemiddelden): een empirische verdeling van steekproefgemiddelden.

h $\mu_{\bar{x}} = \mu_x = 2955\ \text{km}^2$. Dit is dan het gemiddelde van de theoretische steekproevenverdeling.

6.3 Als we de variantie in oppervlakte van alle elf provincies uitrekenen, hebben we de *populatievariantie*, σ_x^2; berekenen we de variantie van de steekproef, dan verkrijgen we de *steekproefvariantie*, S_x^2.

De variantie van de steekproef Zuid-Holland, Noord-Holland, Drenthe:

	x	x^2
Zuid-Holland	29	841
Noord-Holland	27	729
Drenthe	26	676
	$\sum x = 82$	$\sum x^2 = 2246$

$$S_x^2 = \frac{n \cdot \sum x^2 - (\sum x)^2}{n(n-1)} = \frac{3 \cdot 2246 - (82)^2}{3(3-1)} = \frac{6738 - 6724}{6} = \frac{14}{6} = 2{,}33$$

$S_{\bar{x}}^2$ is de variantie van steekproefgemiddelden in een bepaalde *empirische steekproeven verdeling* van het gemiddelde ($\bar{x}_{\bar{x}}$). We kunnen deze waarde berekenen door uit te gaan van de variantie (S_x^2) en de n van *één steekproef*, dus bijvoorbeeld voor de steekproef Zuid-Holland, Noord-Holland en Drenthe krijgt men dan:

$$S_{\bar{x}}^2 = \frac{S_x^2}{n} = \frac{2{,}33}{3} = 0{,}78$$

We schatten dus voor de empirische steekproevenverdeling een variantie van 0,78 (op basis van deze steekproefgegevens).

Bij de steekproef Utrecht, Gelderland, en Drenthe:

	x	x^2
Utrecht	13	169
Gelderland	50	2500
Drenthe	26	676
	$\sum x = 89$	$\sum x^2 = 3345$

$$S_x^2 = \frac{n \cdot \sum x^2 - (\sum x)^2}{n(n-1)} = \frac{3 \cdot 3345 - (89)^2}{3(3-1)} = \frac{10035 - 7921}{6} = \frac{2114}{6}$$

$$= 352{,}33$$

$$S_{\bar{x}}^2 = \frac{S_x^2}{n} = \frac{352{,}33}{3} = 117{,}44$$

De *discrepantie* tussen $S_{\bar{x}}^2(2{,}33)$ van de eerste steekproef en $S_{\bar{x}}^2(352{,}33)$ van de tweede steekproef is *zeer groot*: in het geval van de eerste steekproef lagen de oppervlakten van de drie gekozen provincies dicht bij elkaar (26-29), bij de tweede steekproef zeer ver uiteen (13-50). We zijn het 'slachtoffer' geworden van een vergelijking tussen een 'toevallig' homogene en een 'toevallig' heterogene steekproef. Remedie: men neme méér en vooral ook grotere steekproeven! Op deze wijze verkrijgt men in het algemeen een betere schatting van σ_x (of σ_x^2) op basis van S_x (of S_x^2)!

6.4 Het gemiddelde van de steekproef 'Utrecht, Limburg, Zeeland' =

$$\bar{x} = \frac{13 + 22 + 18}{3} = \frac{53}{3} = 17{,}67\ (\cdot 100\ \text{km}^2) = 1767\ \text{km}^2$$

Om de relatieve plaats van dit gemiddelde te kunnen bepalen, gaan we de z-score van dit steekproefgemiddelde bepalen in de theoretische steekproevenverdeling.

$$Z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}}$$

Om dit in te kunnen vullen, moeten we eerst de populatievariantie uitrekenen:

	x	x^2
Groningen	23	529
Friesland	34	1156
Drenthe	26	676
Overijssel	33	1089
Gelderland	50	2500
Utrecht	13	169
Noord-Holland	27	729
Zuid-Holland	29	841
Zeeland	18	324
Noord-Brabant	50	2500
Limburg	22	484
	$\sum x = 325$	$\sum x^2 = 10997$

$$\sigma_x^2 = \frac{N \cdot \sum x^2 - (\sum x)^2}{N(N-1)} = \frac{11 \cdot 10997 - (325)^2}{11(10)} = \frac{120967 - 105625}{110}$$

$$= \frac{15342}{110} = 139{,}47$$

$$\sigma_x = \sqrt{\sigma_x^2} = \sqrt{139{,}47} = 11{,}81$$

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}} = \frac{11{,}81}{\sqrt{3}} = 6{,}82$$

$$Z = \frac{17{,}67 - 29{,}55}{6{,}82} = \frac{-11{,}88}{6{,}82} = -1{,}74$$

We zoeken de kans op een steekproefgemiddelde van 17,67 en lager. Dit correspondeert in de z-tabel met het gebied van $-1{,}74$ naar negatief oneindig. We kijken in de tweede gebiedskolom van $+1{,}74$ naar positief oneindig, we vinden dan een percentage van 4,09%. De kans op dit steekproefgemiddelde en lager is dus 4,09%.

Normaal gesproken zou men deze kans (van 4%) te gering achten, om vol te houden dat de steekproef 'Utrecht, Limburg, Zeeland' afkomstig is uit de populatie 'Nederlandse provincies' (men neemt immers vaak een standaardwaarde van $\alpha = 5\%$). Toch weten we (in dit geval) dat deze steekproef daar **wel** toe behoort.

6.5 *a* Een tijdschrift vol met verslagen van niet of slechts gedeeltelijk geslaagde experimenten ligt niet goed in de markt; bovendien kan men in zo'n situatie gerede twijfels koesteren of de onderzoeker(s) wel de juiste hypothesen hebben geformuleerd, met andere woorden: of zij hun onderzoek wel op deskundige wijze hebben opgezet en uitgevoerd.

b Het *niet* verwerpen van H_0 (de nulhypothese), c.q. het géén steun (confirmatie) vinden voor de alternatieve hypothese H_1, kan natuurlijk wel degelijk

relevante informatie of een interessant gegeven opleveren. Bovendien kan publicatie collega-onderzoekers uitnodigen of prikkelen om het anders, beter of helemaal niet (verder) te onderzoeken.

c 'Meisjes zijn intelligenter dan jongens ($\alpha = 0{,}10$)
'Jongens zijn langer dan meisjes' ($\alpha = 0{,}001$)
(het zijn maar **voorbeelden**!)

d 'Bij kinderen van 2,5 jaar is de lengte van de grote teen gemiddeld groter dan bij kinderen van 1 jaar' ($\alpha = 0{,}000001$)

6.6 *a* H_0: $\mu_A = \mu_B$
H_1: $\mu_A \neq \mu_B$
De tekst maakt niet duidelijk of de onderzoeker van tevoren reeds verwachtte dat B minder gaatjes zou opleveren. Indien dit wél het geval zou zijn, luiden de hypothesen respectievelijk:
H_0: $\mu_A \leq \mu_B$
H_1: $\mu_A > \mu_B$
N.B.: het gaat om het (gemiddeld) *aantal gaatjes!*

b **Type-I-fout**: het *ten onrechte* verwerpen van de nulhypothese (H_0), dat wil zeggen: aannemen dat er verschil is tussen beide soorten tandpasta (B beter), terwijl dit in werkelijkheid *niet* zo is.
Type-II-fout: het *ten onrechte niet* verwerpen van de nulhypothese.

c $p = 0{,}001$ (één promille, één-duizendste)

d De kans op een type-II-fout zou kleiner geworden zijn.
(bij een eventueel niet verwerpen van H_0)

e Hangt ervan af hoe 'ernstig' een type-II-fout (β) kan zijn: ga je eraan dood, kost het een fortuin?

6.7

	Type-I-fout	Type-II-fout
1 H_0:Geen winst H_1:Wel winst	Verwerpt men ten onrechte de nulhypothese, dan betaalt het bedrijf.	Als men de nulhypothese ten onrechte aanhoudt, dan wordt ons een geneesmiddel onthouden, en het bedrijf loopt de winst mis.
2 H_0:Men wordt er ziek van H_1:Er zijn geen neveneffecten	Verwerpt men de nulhypothese ten onrechte, dan boeten de patiënten in de vorm van gezondheid, maar ongetwijfeld zal het bedrijf schadeclaims aan z'n broek krijgen.	Als men de nulhypothese ten onrechte aanhoudt, dan wordt ons een geneesmiddel onthouden.
3 H_0:Het medicijn werkt niet H_1:Het medicijn werkt wel	Verwerpt men de nulhypothese ten onrechte, dan wordt de patiënt genept, zonder dat het bedrijf daar erg in heeft.	Als men de nulhypothese ten onrechte aanhoudt, dan wordt ons een geneesmiddel onthouden.

Het beste kan men α (type-I-fout) dus klein kiezen (0,01 of 0,001), want de consequenties van een mogelijke β-fout (type-II) zijn niet ernstig, die van een α-fout daarentegen.....!

7 Toetsen voor het gemiddelde van de populatie op basis van één steekproef en betrouwbaarheidsintervallen

In dit hoofdstuk maken we kennis met de eerste uit een reeks toetsen, de z-toets en de t-toets. Door middel van deze toetsen kunnen we nagaan of een steekproefgemiddelde significant verschilt van de populatie waar de steekproef uit getrokken is. We vergelijken hiertoe de *gemiddelde waarden* van steekproef en populatie.
Zowel de z-toets als de t-toets zijn geschikt om deze verschilvraag op te lossen. De z-toets heeft echter de voorkeur, omdat we in deze toets de spreiding van de theoretische steekproevenverdeling schatten met behulp van een populatieparameter. De keuze tussen beide toetsen wordt bepaald door de gegevens waar we wel of geen beschikking over hebben. Beschikken we over de spreidingsgegevens van de populatie (σ_x, σ_x^2), dan kunnen we gebruikmaken van de z-toets. Als we niet over deze gegevens beschikken, dan moeten we 'gedwongen' gebruikmaken van de t-toets. De procedures van deze toetsen worden verduidelijkt aan de hand van voorbeeldsituaties.

7.1 De z-toets

Op een atletiekvereniging wordt een onderzoek gehouden onder de langeafstandslopers. 25 lopers worden onderzocht, en ze lopen gemiddeld op een avond 25 km, terwijl het landelijk gemiddelde op 22 km ligt met een standaardafwijking van 7. De vereniging is van mening dat haar trainingsmethode tot betere loopprestaties leidt. Toets of ze gelijk heeft en neem daarbij een significantieniveau van 5%.
Allereerst moeten we de gegevens die in dit verhaal verweven zijn op een rijtje zetten:

Steekproef:
$n = 25$
$\bar{x} = 25$

Populatie:
$\mu_x = 22$
$\sigma_x = 7$

7.1.1 *Toetskeuze*

We vragen ons nu af of het steekproefgemiddelde van 25 km significant verschilt van de populatie met gemiddelde 22 km. We moeten hier een steekproefgemiddelde gaan plaatsen binnen een theoretische steekproevenverdeling en we beschikken over de spreidingsgevens van de populatie. We kiezen hier dus voor een z-toets.
In de volgende paragrafen volgen we systematisch het toetsingsschema uit hoofdstuk 6.

7.1.2 Onderzoeksvraag

Heeft de trainingsmethode van de vereniging een positief effect op het aantal kilometers dat langeafstandslopers op een avond afleggen? Met andere woorden: leggen de lopers van de vereniging gemiddeld meer kilometers af per avond ten gevolge van de trainingsmethode?

7.1.3 Hypothesen

Hypothesen in termen van de onderzoeksvraag
H_0: Het steekproefgemiddelde $\bar{x} = 25$ is afkomstig uit een populatie met $\mu_x \leq 22$ De lopers van de vereniging lopen evenveel of minder kilometers.
H_1: Het steekproefgemiddelde $\bar{x} = 25$ is afkomstig uit een populatie met $\mu_x > 22$ De lopers van de vereniging lopen meer kilometers.
Hypothesen in statistische termen
H_0: $\mu_x \leq 22$
H_1: $\mu_x > 22$

7.1.4 Aannamen

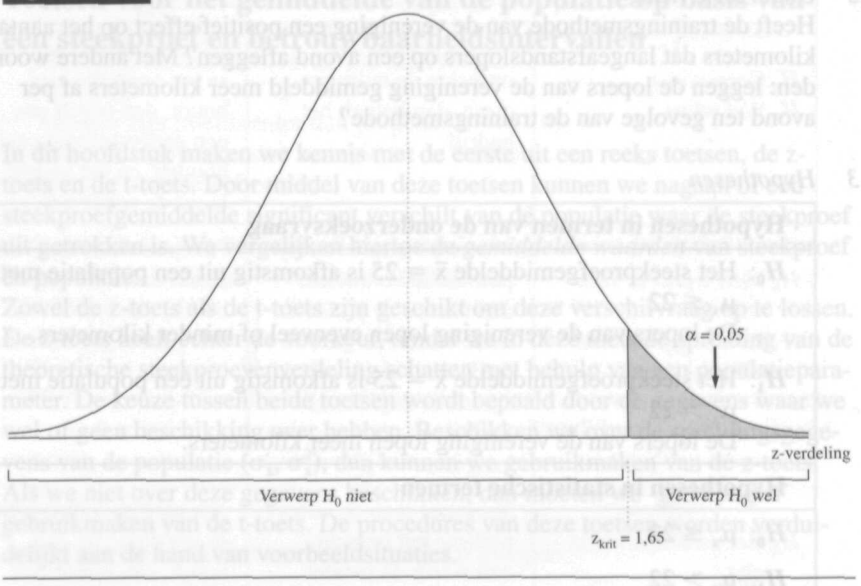
De aannamen van de z-toets zijn altijd gelijk en luiden als volgt:

1 De steekproef is aselect en onafhankelijk totstandgekomen.
2 De populatieverdeling van kilometers is normaal van vorm.

7.1.5 Beslissingsregels

We toetsen hier een positief effect, dus we toetsen rechtseenzijdig met een $\alpha = 0{,}05$. In figuur 7.1 zien we het kritieke gebied aangegeven. We moeten ook een kritieke z-waarde vinden, de grens tussen H_0 wel of niet verwerpen. Deze waarden vinden we in tabel A. We zoeken in de tweede gebiedskolom naar een gebiedswaarde van 0,05, de z-waarde die daarbij hoort is 1,65. $z_{kritiek} = 1{,}65$
In figuur 7.1 staat de situatie schematisch weergegeven.

Figuur 7.1



De beslissingsregels kunnen nu als volgt omschreven worden:

1 Indien $z_{obs} \geq 1{,}65$, verwerp $\boldsymbol{H_0}$.
2 Indien $z_{obs} < 1{,}65$, verwerp $\boldsymbol{H_0}$ niet.

7.1.6 Berekening

We vinden de zogenoemde geobserveerde z-waarde door middel van de z-transformatie binnen de theoretische steekproevenverdeling:

$$z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}}$$

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}} = \frac{7}{\sqrt{25}} = \frac{7}{5} = 1{,}4$$

$$z = \frac{25 - 22}{1{,}4} = 2{,}14$$

$$z_{obs} = 2{,}14$$

7.1.7 Beslissing

We zien dat de geobserveerde z-waarde groter is dan de kritieke z-waarde. De geobserveerde waarde valt dus in het kritieke gebied waarbinnen de nulhypothese wordt verworpen.

$z_{obs} > z_{krit}$, verwerp $\boldsymbol{H_0}$

7.1.8 Conclusie

Er is reden om aan te nemen dat de speciale trainingsmethode van deze vereniging een positief effect heeft op het aantal kilometers dat langeafstandlopers op een avond afleggen.

7.2 De t-toets

Gemiddeld worden er door de Nederlandse bevolking 4,5 kopjes koffie per dag gedronken. Een fabrikant van een bekend koffiemerk introduceert een nieuw soort koffie. Hij neemt een steekproef van 41 mensen, en laat deze mensen een week lang de nieuwe koffie gebruiken. Deze mensen drinken gemiddeld 5 kopjes nieuwe koffie per dag met een variantie van 2. De fabrikant wil toetsen of de nieuwe koffie in een andere hoeveelheid wordt gedronken dan de bekende koffiesoorten. Neem hierbij $\alpha = 0{,}05$.
Allereerst verzamelen we de gegevens uit de tekst:

Steekproef:
$n = 41$
$\bar{x} = 5$
$S_x^2 = 2$

Populatie:
$\mu_x = 4{,}5$

7.2.1 Toetskeuze

De fabrikant wil weten of de nieuwe koffie in een andere hoeveelheid gedronken wordt dan de gewone koffiesoorten. We moeten hier kijken of dit steekproefgemiddelde tot de populatie van 'hoeveelheid gewone koffie' behoort, of dat de nieuwe koffie in een andere populatie thuishoort, met andere hoeveelheden. We hebben alleen de spreidingsgegevens van de steekproef gekregen, niet die van de populatie. We voeren hier dus een t-toets uit. De vraag is of de nieuwe koffie in een *andere* hoeveelheid wordt gedronken; het kan dus meer of minder zijn, daarom kiezen we een tweezijdige toets.

7.2.2 Onderzoeksvraag

Wordt de nieuwe koffie in andere hoeveelheden gedronken dan de gewone, bekende koffiesoorten?

7.2.3 Hypothesen

Hypothesen in termen van de onderzoeksvraag
H_0: Het steekproefgemiddelde $\bar{x} = 5$ is afkomstig uit een populatie met $\mu_x = 4{,}5$. De koffie wordt in dezelfde hoeveelheden gedronken als de gewone soorten koffie.
H_1: Het steekproefgemiddelde $\bar{x} = 5$ is afkomstig uit een populatie met $\mu_x \neq 4{,}5$. De koffie wordt in andere hoeveelheden gedronken dan de gewone soorten koffie.
Hypothesen in statistische termen
H_0: $\mu_x = 4{,}5$
H_1: $\mu_x \neq 4{,}5$

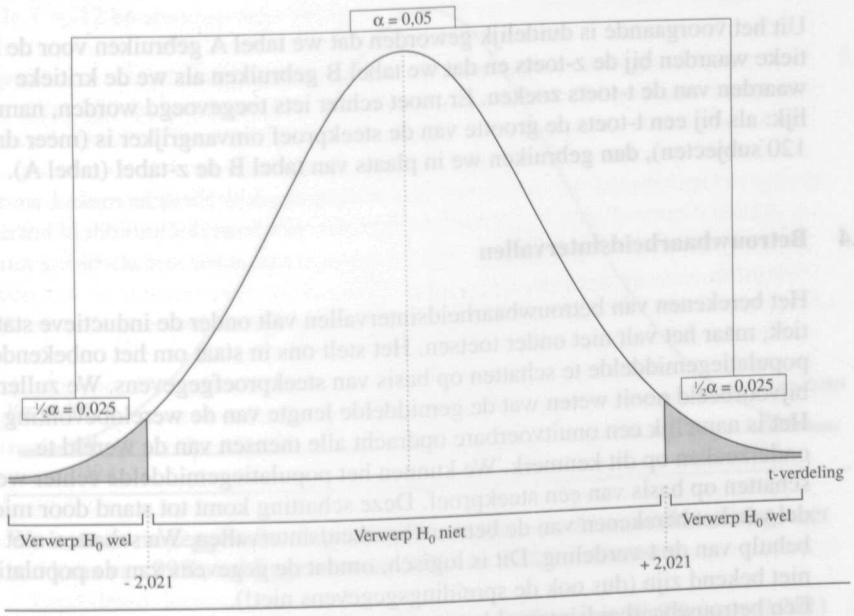
7.2.4 Aannamen

1 De steekproef is aselect en onafhankelijk totstandgekomen.
2 De populatieverdeling van hoeveelheid koffie is normaal van vorm.
We moeten bij deze aannamen een opmerking maken: in de volgende hoofdstukken krijgen we nog enkele malen met een t-toets te maken. Deze toetsen worden onder andere omstandigheden uitgevoerd, dus daar gelden weer andere aannamen voor.

7.2.5 Beslissingsregels

We toetsen hier een effect dat zowel een positieve als een negatieve uitwerking kan hebben. We toetsen dus tweezijdig met een $\alpha = 0{,}05$. In figuur 7.2 zien we het kritieke gebied aangegeven. We moeten ook de kritieke t-waarden vinden, de grenzen tussen H_0 wel of niet verwerpen. Deze waarden vinden we in tabel B.
Tabel B gebruiken we als volgt. We zoeken eerst de juiste α op in de bovenbalken van de tabel, we kijken in de kolom met significantieniveau van 0,05, tweezijdig. Daarnaast moeten we ook de df (degrees of freedom oftewel vrijheidsgraden) bepalen. Het aantal vrijheidsgraden bij deze toets is gelijk aan $n - 1$, dus $41 - 1 = 40$.
$t_{kritiek} = +2{,}021$ en $-2{,}021$
In figuur 7.2 staat de situatie schematisch weergegeven.

Figuur 7.2



De beslissingsregels kunnen nu als volgt omschreven worden:

1 Indien $t_{obs} \geq +2{,}021$ of $t_{obs} \leq -2{,}021$, verwerp $\boldsymbol{H_0}$

2 Indien $-2{,}021 < t_{obs} < +2{,}021$, verwerp $\boldsymbol{H_0}$ niet

7.2.6 *Berekening*

De formule om de geobserveerde t-waarde te vinden, ziet er als volgt uit:

$$t = \frac{\bar{x} - \mu_{\bar{x}}}{S_{\bar{x}}}$$

De formule lijkt veel op die van de z-transformatie, alleen schatten we de spreiding van de theoretische steekproevenverdeling niet op basis van de spreiding in de populatie maar op basis van de spreiding in de steekproefverdeling.

$$S_{\bar{x}} = \frac{\sqrt{S_x^2}}{\sqrt{n}} = \frac{\sqrt{2}}{\sqrt{41}} = \frac{1{,}41}{6{,}40} = 0{,}22$$

$$t = \frac{5 - 4{,}5}{0{,}22} = 2{,}27$$

$$t_{obs} = 2{,}27$$

7.2.7 *Beslissing*

$t_{obs} > t_{krit}$, want $2{,}27 > 2{,}021$, dus verwerp $\boldsymbol{H_0}$

7.2.8 *Conclusie*

Er is reden om aan te nemen, dat de nieuwe soort koffie in andere hoeveelheden gedronken wordt dan de bekende soorten koffie.

7.3 Tabellen

Uit het voorgaande is duidelijk geworden dat we tabel A gebruiken voor de kritieke waarden bij de z-toets en dat we tabel B gebruiken als we de kritieke waarden van de t-toets zoeken. Er moet echter iets toegevoegd worden, namelijk: als bij een t-toets de grootte van de steekproef omvangrijker is (meer dan 120 subjecten), dan gebruiken we in plaats van tabel B de z-tabel (tabel A).

7.4 Betrouwbaarheidsintervallen

Het berekenen van betrouwbaarheidsintervallen valt onder de inductieve statistiek, maar het valt niet onder toetsen. Het stelt ons in staat om het onbekende populatiegemiddelde te schatten op basis van steekproefgegevens. We zullen bijvoorbeeld nooit weten wat de gemiddelde lengte van de wereldbevolking is. Het is namelijk een onuitvoerbare opdracht alle mensen van de wereld te onderzoeken op dit kenmerk. We kunnen het populatiegemiddelde echter wel schatten op basis van een steekproef. Deze schatting komt tot stand door middel van het berekenen van de betrouwbaarheidsintervallen. We schatten met behulp van de t-verdeling. Dit is logisch, omdat de gegevens van de populatie niet bekend zijn (dus ook de spreidingsgegevens niet!).
Een betrouwbaarheidsinterval kent twee grenzen, een onder- en een bovengrens. Hieronder staan de formules van de grenzen van betrouwbaarheidsintervallen:
Voor de ondergrens: $\bar{x} - t_\alpha \cdot (S_{\bar{x}})$
Voor de bovengrens: $\bar{x} + t_\alpha \cdot (S_{\bar{x}})$

In de volgende voorbeeldsituatie wordt het een en ander uiteengezet.

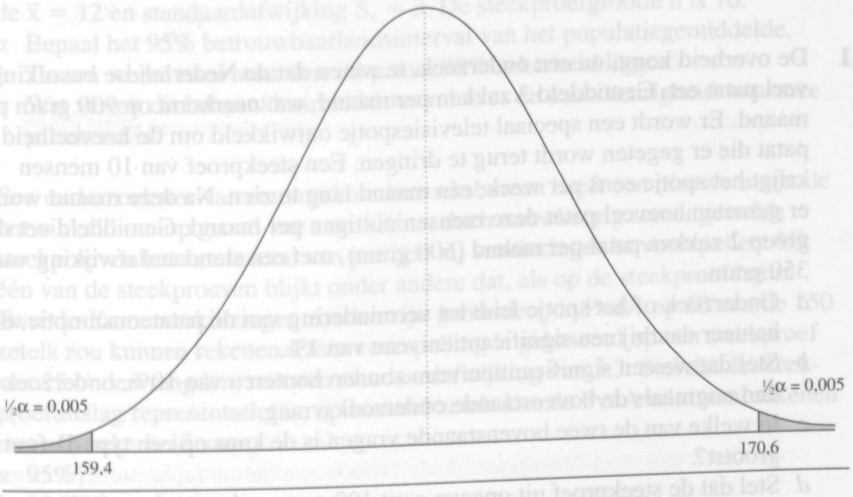
We willen iets te weten komen over de gemiddelde lichaamslengte van de wereldbevolking. Daartoe nemen we een steekproef van 25 mensen. De gemiddelde lengte in de steekproef is 165 centimeter met een standaardafwijking van 10 centimeter. We willen het populatiegemiddelde schatten met een kans van 99% dat het populatiegemiddelde binnen de twee grenzen van het interval ligt. We berekenen dan het 99% betrouwbaarheidsinterval ($\alpha = 0{,}01$).
De benodigde gegevens worden hieronder opgesomd:
$\bar{x} = 165$
t_α = deze waarde zoeken we op in tabel B: tweezijdig, $\alpha = 0{,}01$ en
$df = 24 \rightarrow 2{,}797$

$$S_{\bar{x}} = \frac{S_x}{\sqrt{n}} = \frac{10}{\sqrt{25}} = \frac{10}{5} = 2$$

Nu kunnen we de formules invullen:
Voor de ondergrens: $\bar{x} - t_\alpha \cdot (S_{\bar{x}}) = 165 - 2{,}797 \cdot 2 = 159{,}4$ cm
Voor de bovengrens: $\bar{x} + t_\alpha \cdot (S_{\bar{x}}) = 165 + 2{,}797 \cdot 2 = 170{,}6$ cm

In figuur 7.3 staat de situatie weergegeven:

Figuur 7.3 99% Betrouwbaarheidsinterval



We kunnen nu zeggen dat de gemiddelde lengte van de wereldbevolking, met een kans van 99%, zal liggen tussen 159,4 en 170,6 centimeter.

Op deze wijze kunnen we ook het 95% ($\alpha = 0{,}05$) of het 99,9% ($\alpha = 0{,}001$) betrouwbaarheidsinterval berekenen. We moeten wel rekening houden met het feit dat de steekproef niet representatief zou kunnen zijn, dan kloppen de gegevens niet meer.

Opgaven

7.1 De overheid komt, na een onderzoek, te weten dat de Nederlandse bevolking te veel patat eet. Gemiddeld 3 zakken per maand, wat neerkomt op 900 gram per maand. Er wordt een speciaal televisiespotje ontwikkeld om de hoeveelheid patat die er gegeten wordt terug te dringen. Een steekproef van 10 mensen krijgt het spotje eens per week, een maand lang te zien. Na deze maand wordt er gemeten hoeveel patat deze mensen nuttigen per maand. Gemiddeld eet deze groep 2 zakken patat per maand (600 gram), met een standaardafwijking van 350 gram.

a Onderzoek of het spotje leidt tot vermindering van de patatconsumptie, en hanteer daarbij een significantieniveau van 1%.

b Stel dat we een significantieniveau zouden hanteren van 10%, onderzoek dan nogmaals de bovenstaande onderzoeksvraag.

c In welke van de twee bovenstaande vragen is de kans op een type-II-fout het grootst?

d Stel dat de steekproef uit opgave *a* uit 100 mensen bestond, toets dan nogmaals de onderzoeksvraag. Gebruik ook hier een α van 1%.

e Vergelijk de resultaten van de opgaven *a* en *d*.

7.2 Een restaurantketen in Nederland vindt het gemiddeld aantal gasten per restaurant per avond te laag. Gemiddeld komen er 125 eters op een avond, met een standaardafwijking van 35. De directie van de restaurants wil het gemiddeld aantal gasten vermeerderen door de prijs van alle gerechten met een gulden te verminderen. Zij probeert deze remedie gedurende twee weken uit in 16 restaurants. Gemiddeld komen er 139 bezoekers per avond.

a Toets of de prijsverlaging een positief effect heeft op het aantal restaurantbezoekers. Neem daarbij een α van 0,05.

Niet erg tevreden met dit resultaat, bedenkt de directie een andere manier om het aantal restaurantbezoekers te vergroten. Het idee is om elke dag een voordelig dagmenu te serveren en daarvoor ook te adverteren in de plaatselijke en landelijke dagbladen. In 16 andere restaurants wordt deze aanpak uitgeprobeerd, gemiddeld komen er in deze restaurants nu 145 eters per avond.

b Toets of de aanpak een positief effect heeft op het aantal eters, en neem daarbij weer een α van 0,05.

Natuurlijk is de directie ook geïnteresseerd in de inkomsten van de restaurants die met de nieuwe aanpak werken. Gemiddeld maakten de restaurants een weekomzet van 5000 gulden (per restaurant, per week). De directie vraagt zich af of de nieuwe aanpak überhaupt van invloed is geweest op de weekomzet. De 16 restaurants die het dagmenu introduceerden, maakten gemiddeld een weekomzet van 5100 gulden per restaurant, met een standaardafwijking van 200 gulden.

c Toets of de nieuwe aanpak van invloed is geweest op de omzet van de restaurants, neem daarbij een α van 0,05.

7.3 Gegeven is een steekproef uit een normale verdeling met steekproefgemiddelde $\bar{x} = 12$ en standaardafwijking $S_x = 3$. De steekproefgrootte n is 16.

a Bepaal het 95% betrouwbaarheidsinterval van het populatiegemiddelde.

b Tussen welke twee waarden zal μ met 99% zekerheid liggen?

c Wat vertelt dit betrouwbaarheidsinterval over de nauwkeurigheid waarmee μ bekend is?

7.4 Een onderzoekster van een marktonderzoeksbureau heeft een nieuwe methode ontwikkeld om op grond van zeer kleine, doch nauwkeurig samengestelde steekproeven het zetelaantal van partijen bij verkiezingen te voorspellen. Uit één van de steekproeven blijkt onder andere dat, als op de steekproefdatum Tweede-Kamerverkiezingen zouden zijn gehouden, de PvdA op 60 van de 150 zetels zou kunnen rekenen. Deze voorspelling is gebaseerd op een steekproef van 25 ondervraagden, met een standaardafwijking van 5,3. Stel dat de steekproefuitslag representatief is, op hoeveel zetels kan de PvdA minimaal rekenen met een zekerheid van:

a 95%?

b 99,9%?

c Als de PvdA 99,9% zekerheid wil hebben dat ze voldoende kandidaten op de kieslijst heeft staan om de toegewezen zetels te kunnen bezetten, hoeveel kandidaten dienen er dan minimaal op de kieslijst te staan?

Uitwerkingen

7.1 *a* Allereerst verzamelen we de gegevens uit de tekst:

Steekproef:
$n = 10$
$\bar{x} = 600$
$S_x = 350$

Populatie:
$\mu_x = 900$

Toetskeuze
De overheid wil weten of het televisiespotje het gewenste effect heeft op de patatconsumptie. Men verwacht een vermindering van de consumptie, en we plaatsen een steekproefgemiddelde binnen een theoretische steekproevenverdeling. We hebben geen beschikking over de spreidingsgegevens van de populatie. We voeren een linkseenzijdige t-toets uit.

Onderzoeksvraag
Consumeert men minder patat na het zien van het speciaal ontwikkelde televisiespotje?

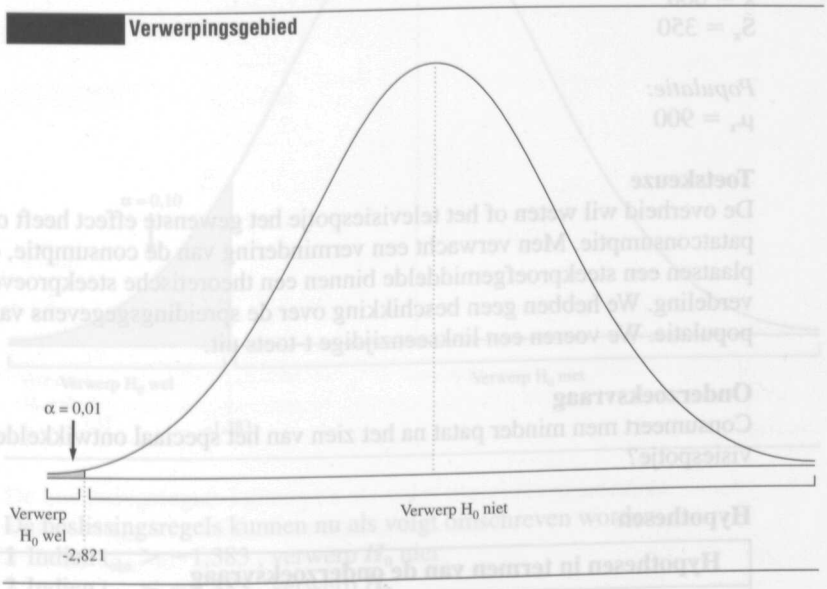
Hypothesen

Hypothesen in termen van de onderzoeksvraag
H_0: Het steekproefgemiddelde $\bar{x} = 600$ is afkomstig uit een populatie met $\mu_x \geq 900$.Men eet meer of evenveel patat na het zien van het spotje.
H_1: Het steekproefgemiddelde $\bar{x} = 600$ is afkomstig uit een populatie met $\mu_x < 900$.Men eet minder patat na het zien van het spotje.
Hypothesen in statistische termen
H_0: $\mu_x \geq 900$
H_1: $\mu_x < 900$

Aannamen
1 De steekproef is aselect en onafhankelijk totstandgekomen.
2 De populatieverdeling is normaal van vorm.

Beslissingsregels
We toetsen hier een vermindering, we toetsen dus linkseenzijdig met een $\alpha = 0{,}01$. We moeten ook de kritieke t-waarde vinden, de grens tussen H_0 wel of niet verwerpen. Deze waarde vinden we in tabel B.

$df = n - 1 = 10 - 1 = 9$
$\alpha = 0{,}01$ eenzijdig
$t_{kritiek} = -2{,}821$
In de volgende figuur zien we het kritieke gebied aangegeven.



De beslissingsregels kunnen nu als volgt omschreven worden:
1 Indien $t_{obs} > -2{,}821$, verwerp $\boldsymbol{H_0}$ niet
2 Indien $t_{obs} \leq -2{,}821$, verwerp $\boldsymbol{H_0}$

Berekening

$$t = \frac{\bar{x} - \mu_{\bar{x}}}{S_{\bar{x}}}$$

$$S_{\bar{x}} = \frac{S_x}{\sqrt{n}} = \frac{350}{\sqrt{10}} = \frac{350}{3{,}16} = 110{,}76$$

$$t = \frac{600 - 900}{110{,}76} = -2{,}71$$

$t_{obs} = -2{,}71$

Beslissing

$t_{obs} > t_{krit}$, want $-2{,}71 > -2{,}821$, dus verwerp $\boldsymbol{H_0}$ niet

Conclusie

Er is geen reden om aan te nemen dat men door het nieuwe televisiespotje minder patat consumeert dan normaal.

b Allereerst verzamelen we de gegevens uit de tekst:

Steekproef:
$n = 10$
$\bar{x} = 600$
$S_x = 350$

Populatie:
$\mu_x = 900$

Toetskeuze
De overheid wil weten of het televisiespotje het gewenste effect heeft op de patatconsumptie. Men verwacht een vermindering van de consumptie, en we plaatsen een steekproefgemiddelde binnen een theoretische steekproevenverdeling. We hebben geen beschikking over de spreidingsgegevens van de populatie. We voeren een linkseenzijdige t-toets uit.

Onderzoeksvraag
Consumeert men minder patat na het zien van het speciaal ontwikkelde televisiespotje?

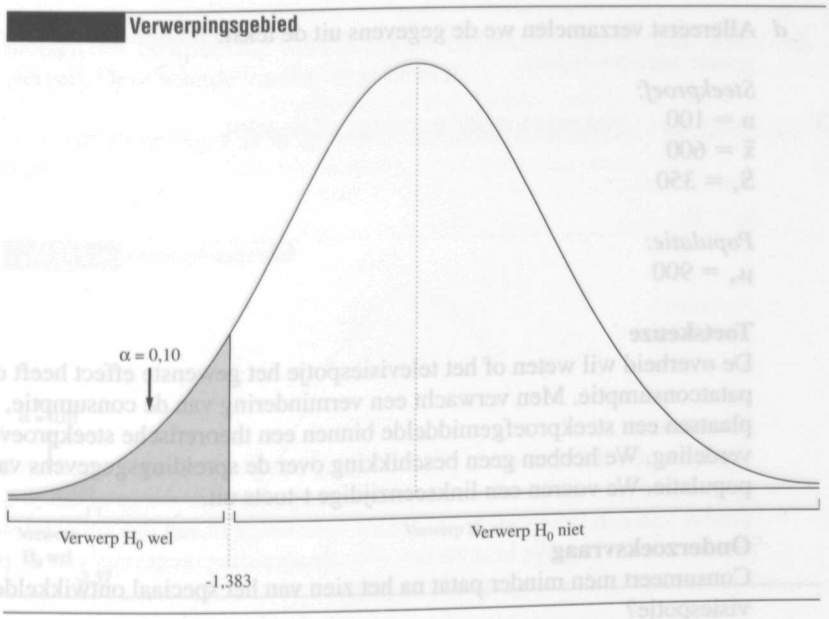
Hypothesen

Hypothesen in termen van de onderzoeksvraag
H_0: Het steekproefgemiddelde $\bar{x} = 600$ is afkomstig uit een populatie met $\mu_x \geq 900$. Men eet meer of evenveel patat na het zien van het spotje.
H_1: Het steekproefgemiddelde $\bar{x} = 600$ is afkomstig uit een populatie met $\mu_x < 900$. Men eet minder patat na het zien van het spotje.
Hypothesen in statistische termen:
H_0: $\mu_x \geq 900$
H_1: $\mu_x < 900$

Aannamen
1 De steekproef is aselect en onafhankelijk totstandgekomen.
2 De populatieverdeling is normaal van vorm.

Beslissingsregels
We toetsen hier een vermindering. We toetsen dus linkseenzijdig met een $\alpha = 0{,}10$. We moeten ook de kritieke t-waarde vinden, de grens tussen H_0 wel of niet verwerpen. Deze waarde vinden we in tabel B.

$df = n - 1 = 10 - 1 = 9$
$\alpha = 0{,}10$ eenzijdig
$t_{kritiek} = -1{,}383$
In de volgende figuur zien we het kritieke gebied aangegeven.



De beslissingsregels kunnen nu als volgt omschreven worden:
1 Indien $t_{obs} > -1{,}383$, verwerp $\boldsymbol{H_0}$ niet
2 Indien $t_{obs} \leq -1{,}383$, verwerp $\boldsymbol{H_0}$

Berekening

$$t = \frac{\bar{x} - \mu_{\bar{x}}}{S_{\bar{x}}}$$

$$S_{\bar{x}} = \frac{S_x}{\sqrt{n}} = \frac{350}{\sqrt{10}} = \frac{350}{3{,}16} = 110{,}76$$

$$t = \frac{600 - 900}{110{,}76} = -2{,}71$$

$$t_{obs} = -2{,}71$$

Beslissing
$t_{obs} < t_{krit}$, want $-2{,}71 < -1.383$, dus verwerp $\boldsymbol{H_0}$

Conclusie
Er is reden om aan te nemen dat men door het nieuwe televisiespotje minder patat consumeert dan normaal.

c De kans op een type-II-fout is de kans dat men ten onrechte de nulhypothese aanhoudt. Deze kans wordt groter in die gevallen waar α kleiner wordt. α geeft immers de grootte van de kans op een type-I-fout weer. In opgave *a* is de α het kleinst, en daar is dus de kans op een type-II-fout het grootst.

d Allereerst verzamelen we de gegevens uit de tekst:

Steekproef:
$n = 100$
$\bar{x} = 600$
$S_x = 350$

Populatie:
$\mu_x = 900$

Toetskeuze
De overheid wil weten of het televisiespotje het gewenste effect heeft op de patatconsumptie. Men verwacht een vermindering van de consumptie, en we plaatsen een steekproefgemiddelde binnen een theoretische steekproevenverdeling. We hebben geen beschikking over de spreidingsgegevens van de populatie. We voeren een linkseenzijdige t-toets uit.

Onderzoeksvraag
Consumeert men minder patat na het zien van het speciaal ontwikkelde televisiespotje?

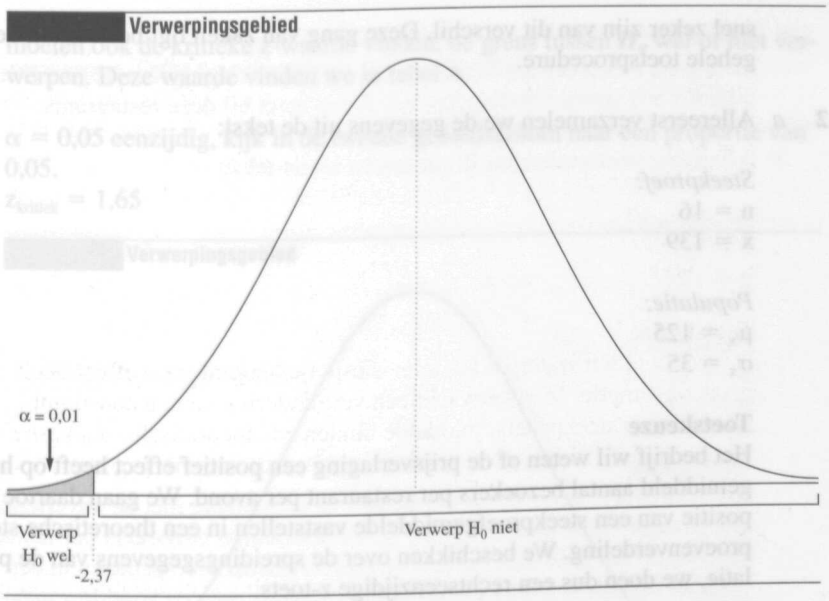
Hypothesen

Hypothesen in termen van de onderzoeksvraag
H_0: Het steekproefgemiddelde $\bar{x} = 600$ is afkomstig uit een populatie met $\mu_x \geq 900$. Men eet meer of evenveel patat na het zien van het spotje.
H_1: Het steekproefgemiddelde $\bar{x} = 600$ is afkomstig uit een populatie met $\mu_x < 900$. Men eet minder patat na het zien van het spotje.
Hypothesen in statistische termen
H_0: $\mu_x \geq 900$
H_1: $\mu_x < 900$

Aannamen
1 De steekproef is aselect en onafhankelijk totstandgekomen.
2 De populatieverdeling is normaal van vorm.

Beslissingsregels
We toetsen hier een vermindering, we toetsen dus linkseenzijdig met een $\alpha = 0{,}01$. We moeten ook de kritieke t-waarde vinden, de grens tussen H_0 wel of niet verwerpen. Deze waarde vinden we in tabel B.

$df = n - 1 = 100 - 1 = 99$
$\alpha = 0{,}01$ eenzijdig
$t_{kritiek} \approx -2{,}37$
In de volgende figuur zien we het kritieke gebied aangegeven.



De beslissingsregels kunnen nu als volgt omschreven worden:
1 Indien $t_{obs} > -2{,}37$, verwerp $\boldsymbol{H_0}$ niet
2 Indien $t_{obs} \leq -2{,}37$, verwerp $\boldsymbol{H_0}$

Berekening

$$t = \frac{\bar{x} - \mu_{\bar{x}}}{S_{\bar{x}}}$$

$$S_{\bar{x}} = \frac{S_x}{\sqrt{n}} = \frac{350}{\sqrt{100}} = \frac{350}{10} = 35$$

$$t = \frac{600 - 900}{35} = -8{,}57$$

$$t_{obs} = -8{,}57$$

Beslissing
$t_{obs} < t_{krit}$, want $-8{,}57 < -2{,}37$, dus verwerp $\boldsymbol{H_0}$

Conclusie
Er is reden om aan te nemen, dat men door het nieuwe televisiespotje minder patat consumeert dan normaal.

e In opdracht *a* verwierpen we de nulhypothese niet bij een α van 1% en bij dezelfde α verwerpen we de nulhypothese in opgave *d* wel. De oorzaak hiervan ligt in de grootte van de steekproef. Hoe groter de steekproef, des te sneller verwerpen we de nulhypothese. Dit is redelijk logisch te verklaren, want als we een verschil bij heel veel mensen constateren, dan kunnen we al

snel zeker zijn van dit verschil. Deze gang van zaken ligt besloten in de gehele toetsprocedure.

7.2 *a* Allereerst verzamelen we de gegevens uit de tekst:

Steekproef:
$n = 16$
$\bar{x} = 139$

Populatie:
$\mu_x = 125$
$\sigma_x = 35$

Toetskeuze
Het bedrijf wil weten of de prijsverlaging een positief effect heeft op het gemiddeld aantal bezoekers per restaurant per avond. We gaan daartoe de positie van een steekproefgemiddelde vaststellen in een theoretische steekproevenverdeling. We beschikken over de spreidingsgegevens van de populatie, we doen dus een rechtseenzijdige z-toets.

Onderzoeksvraag
Hebben de restaurants die de prijsverlaging hebben geïntroduceerd meer bezoekers per avond dan de andere restaurants? Heeft de prijsverlaging een positief effect op het aantal eters?

Hypothesen

Hypothesen in termen van de onderzoeksvraag
H_0: Het steekproefgemiddelde $\bar{x} = 139$ is afkomstig uit een populatie met $\mu_x \leq 125$. Er komen evenveel of minder eters naar de restaurants met prijsverlaging.
H_1: Het steekproefgemiddelde $\bar{x} = 139$ is afkomstig uit een populatie met $\mu_x > 125$. Er komen meer eters naar de restaurants die de prijsverlaging ingevoerd hebben.
Hypothesen in statistische termen
H_0: $\mu_x \leq 125$ H_1: $\mu_x > 125$

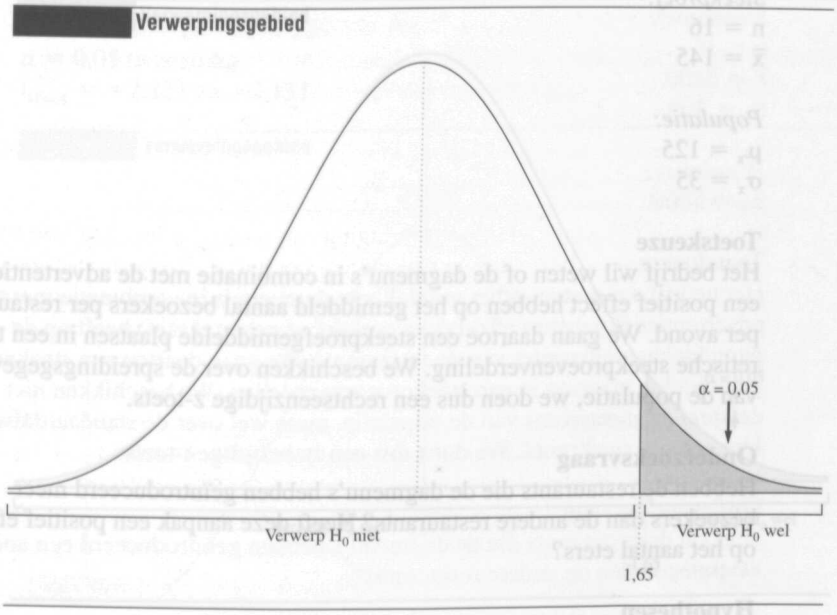
Aannamen
1 De steekproef is aselect en onafhankelijk totstandgekomen.
2 De populatieverdeling van het aantal bezoekers is normaal van vorm.

Beslissingsregels
We toetsen hier een positief effect, we toetsen dus rechtseenzijdig met een $\alpha = 0{,}05$. In de volgende figuur zien we het kritieke gebied aangegeven. We

moeten ook de kritieke z-waarde vinden, de grens tussen H_0 wel of niet verwerpen. Deze waarde vinden we in tabel A.

$\alpha = 0{,}05$ eenzijdig, kijk in de tweede gebiedskolom naar een proportie van 0,05.

$z_{kritiek} = 1{,}65$



De beslissingsregels kunnen nu als volgt omschreven worden:

1 Indien $z_{obs} \geq 1{,}65$, verwerp H_0

2 Indien $z_{obs} < 1{,}65$, verwerp H_0 niet

Berekening

$$z = \frac{\bar{x} - \mu_{\bar{x}}}{S_{\bar{x}}}$$

$$S_{\bar{x}} = \frac{S_x}{\sqrt{n}} = \frac{35}{\sqrt{16}} = \frac{35}{4} = 8{,}75$$

$$z = \frac{139 - 125}{8{,}75} = 1{,}6$$

$z_{obs} = 1{,}6$

Beslissing

$z_{obs} < z_{krit}$, want $1{,}6 < 1{,}65$, dus verwerp H_0 niet

Conclusie
Er is geen reden om aan te nemen dat de prijsverlaging van de gerechten leidt tot een vermeerdering van het aantal eters bij deze restaurants.

b Allereerst verzamelen we de gegevens uit de tekst:

Steekproef:
$n = 16$
$\bar{x} = 145$

Populatie:
$\mu_x = 125$
$\sigma_x = 35$

Toetskeuze
Het bedrijf wil weten of de dagmenu's in combinatie met de advertenties een positief effect hebben op het gemiddeld aantal bezoekers per restaurant per avond. We gaan daartoe een steekproefgemiddelde plaatsen in een theoretische steekproevenverdeling. We beschikken over de spreidingsgegevens van de populatie, we doen dus een rechtseenzijdige z-toets.

Onderzoeksvraag
Hebben de restaurants die de dagmenu's hebben geïntroduceerd meer bezoekers dan de andere restaurants? Heeft deze aanpak een positief effect op het aantal eters?

Hypothesen

Hypothesen in termen van de onderzoeksvraag
H_0: Het steekproefgemiddelde $\bar{x} = 145$ is afkomstig uit een populatie met $\mu_x \leq 125$. Er komen evenveel of minder eters naar de restaurants met de dagmenu's.
H_1: Het steekproefgemiddelde $\bar{x} = 145$ is afkomstig uit een populatie met $\mu_x > 125$. Er komen meer eters naar de restaurants met de dagmenu's.
Hypothesen in statistische termen
H_0: $\mu_x \leq 125$
H_1: $\mu_x > 125$

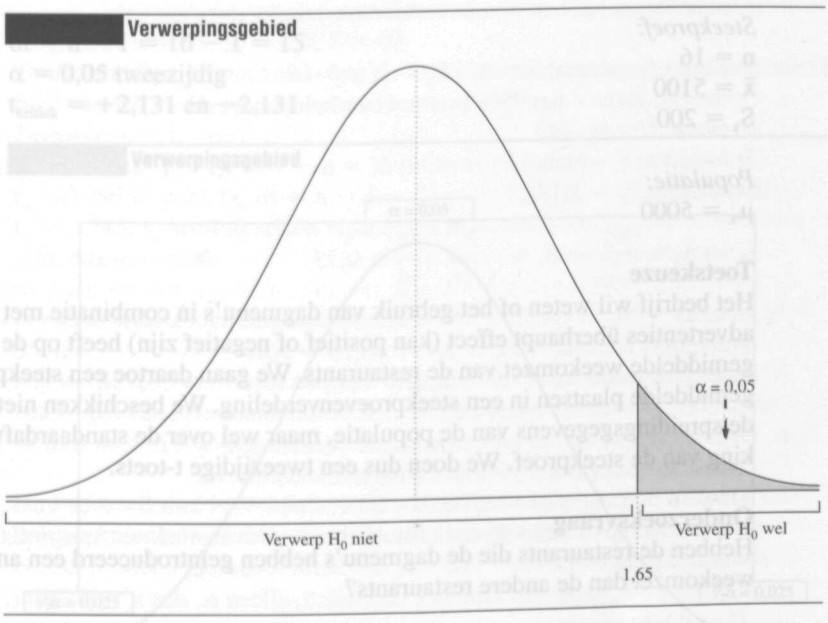
Aannamen
1 De steekproef is aselect en onafhankelijk totstandgekomen.
2 De populatieverdeling van het aantal eters is normaal van vorm.

2 **Beslissingsregels**
We toetsen hier een positief effect, we toetsen dus rechtseenzijdig met een $\alpha = 0{,}05$. In onderstaande figuur zien we het kritieke gebied aangegeven.

We moeten ook de kritieke z-waarde vinden, de grens tussen H_0 wel of niet verwerpen. Deze waarde vinden we in tabel A.

$\alpha = 0{,}05$ eenzijdig, kijk in de tweede gebiedskolom naar een proportie van 0,05.

$z_{kritiek} = 1{,}65$



De beslissingsregels kunnen nu als volgt omschreven worden:

1 Indien $z_{obs} \geq 1{,}65$, verwerp H_0

2 Indien $z_{obs} < 1{,}65$, verwerp H_0 niet

Berekening

$$z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}}$$

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}} = \frac{35}{\sqrt{16}} = \frac{35}{4} = 8{,}75$$

$$z = \frac{145 - 125}{8{,}75} = 2{,}29$$

$z_{obs} = 2{,}29$

Beslissing

$z_{obs} > z_{krit}$, want $2{,}29 > 1{,}65$, dus verwerp H_0

Conclusie
Er is reden om aan te nemen dat de voordelige dagmenu's in combinatie met de advertenties leiden tot een vermeerdering van het aantal eters bij deze restaurants.

c Allereerst verzamelen we weer de gegevens uit de tekst:

Steekproef:
$n = 16$
$\bar{x} = 5100$
$S_x = 200$

Populatie:
$\mu_x = 5000$

Toetskeuze
Het bedrijf wil weten of het gebruik van dagmenu's in combinatie met de advertenties überhaupt effect (kan positief of negatief zijn) heeft op de gemiddelde weekomzet van de restaurants. We gaan daartoe een steekproefgemiddelde plaatsen in een steekproevenverdeling. We beschikken niet over de spreidingsgegevens van de populatie, maar wel over de standaardafwijking van de steekproef. We doen dus een tweezijdige t-toets.

Onderzoeksvraag
Hebben de restaurants die de dagmenu's hebben geïntroduceerd een andere weekomzet dan de andere restaurants?

Hypothesen

Hypothesen in termen van de onderzoeksvraag
H_0: Het steekproefgemiddelde $\bar{x} = 5100$ is afkomstig uit een populatie met $\mu_x = 5000$. De nieuwe aanpak verandert de weekomzet niet.
H_1: Het steekproefgemiddelde $\bar{x} = 5100$ is afkomstig uit een populatie met $\mu_x \neq 5000$. De nieuwe aanpak leidt tot een andere weekomzet.
Hypothesen in statistische termen
H_0: $\mu_x = 5000$
H_1: $\mu_x \neq 5000$

Aannamen
1 De steekproef is aselect en onafhankelijk totstandgekomen.
2 De populatieverdeling is normaal van vorm.

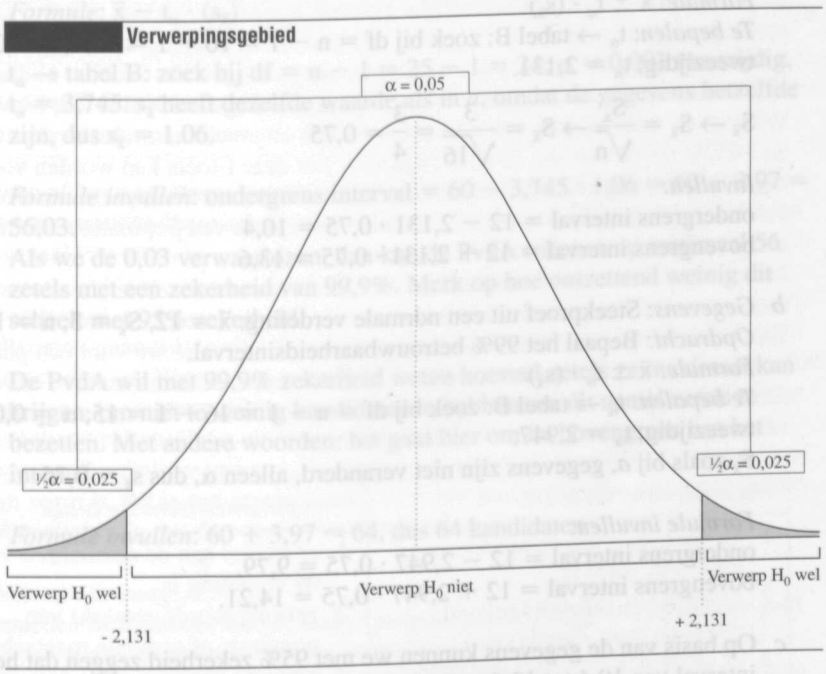
Beslissingsregels

We toetsen hier een effect dat zowel positief als negatief kan zijn. We toetsen dus tweezijdig met een $\alpha = 0{,}05$. In onderstaande figuur zien we het kritieke gebied aangegeven. We moeten ook de kritieke t-waarden vinden, de grenzen tussen H_0 wel of niet verwerpen. Deze waarden vinden we in tabel B.

$df = n - 1 = 16 - 1 = 15$

$\alpha = 0{,}05$ tweezijdig

$t_{kritiek} = +2{,}131$ en $-2{,}131$



De beslissingsregels kunnen nu als volgt omschreven worden:

1 Indien $t_{obs} \geq +2{,}131$ of $t_{obs} \leq -2{,}131$, verwerp H_0

2 Indien $-2{,}131 < t_{obs} < +2{,}131$, verwerp H_0 niet

Berekening

$$t = \frac{\bar{x} - \mu_{\bar{x}}}{S_{\bar{x}}}$$

$$S_{\bar{x}} = \frac{S_x}{\sqrt{n}} = \frac{200}{\sqrt{16}} = \frac{200}{4} = 50$$

$$t = \frac{5100 - 5000}{50} = 2$$

$t_{obs} = 2$

Beslissing

$-t_{krit} < t_{obs} < +t_{krit}$, want $-2{,}131 < 2 < +2{,}131$, dus verwerp $\boldsymbol{H_0}$ niet

Conclusie

Er is geen reden om aan te nemen dat de restaurants die de voordelige dagmenu's in combinatie met de advertenties hebben geïntroduceerd, een andere weekomzet hebben dan de restaurants zonder deze aanpak.

7.3 *a* *Steekproefgegevens:* $\bar{x} = 12$, $S_x = 3$, $n = 16$.
Opdracht: Bepaal het 95% betrouwbaarheidsinterval:
Formule: $\bar{x} \pm t_\alpha \cdot (s_{\bar{x}})$
Te bepalen: $t_\alpha \rightarrow$ tabel B: zoek bij $df = n - 1 = 16 - 1 = 15$; $\alpha = 0{,}05$ tweezijdig; $t_\alpha = 2{,}131$.

$$S_{\bar{x}} \rightarrow S_{\bar{x}} = \frac{S_x}{\sqrt{n}} \rightarrow S_{\bar{x}} = \frac{3}{\sqrt{16}} = \frac{3}{4} = 0{,}75$$

Invullen:
ondergrens interval $= 12 - 2{,}131 \cdot 0{,}75 = 10{,}4$
bovengrens interval $= 12 + 2{,}131 \cdot 0{,}75 = 13{,}6$

b *Gegevens*: Steekproef uit een normale verdeling: $\bar{x} = 12$, $S_x = 3$, $n = 16$.
Opdracht: Bepaal het 99% betrouwbaarheidsinterval.
Formule: $\bar{x} \pm t_\alpha \cdot (s_{\bar{x}})$
Te bepalen: $t_\alpha \rightarrow$ tabel B: zoek bij $df = n - 1 = 16 - 1 = 15$; $\alpha = 0{,}01$ tweezijdig; $t_\alpha = 2{,}947$.
$S_{\bar{x}}$ zoals bij *a*, gegevens zijn niet veranderd, alleen α, dus $s_{\bar{x}} = 0{,}75$.

Formule invullen:
ondergrens interval $= 12 - 2{,}947 \cdot 0{,}75 = 9{,}79$.
bovengrens interval $= 12 + 2{,}947 \cdot 0{,}75 = 14{,}21$.

c Op basis van de gegevens kunnen we met 95% zekerheid zeggen dat het interval van 10,4 tot 13,6 μ (het populatiegemiddelde) bevat, en met 99% zekerheid kunnen we zeggen dat het interval van 9,79 tot 14,21 μ bevat. De nauwkeurigheid waarmee we dat vastgesteld hebben ligt aan α. Zodra deze kleiner wordt, wordt de lengte van het interval groter. Als $\alpha = 0{,}05$, dan is het bijbehorende 95% interval 3,2, en als $\alpha = 0{,}01$ dan is het bijbehorende 99% interval $\rightarrow$ 4,42. Met andere woorden: hoe groter de zekerheid die je wilt hebben dat μ in het interval ligt, hoe groter het interval wordt, oftewel: hoe minder nauwkeurig de waarde van μ kan worden vastgesteld.

7.4 *a* *Gegevens*: $n = 25$, $S_x = 5{,}3$, $\bar{x} = 60$.
Opdracht: Op hoeveel zetels kan de PvdA *minimaal* rekenen, met een zekerheid van 95%?
Formule: $\bar{x} - t_\alpha \cdot (s_{\bar{x}})$
Te bepalen:
$t_\alpha \rightarrow$ tabel B: zoek bij $df = n - 1 = 25 - 1 = 24$; $\alpha = 0{,}05$ tweezijdig; $t_\alpha = 2{,}064$.

$S_{\bar{x}} \rightarrow S_{\bar{x}} = \frac{S_x}{\sqrt{n}} \rightarrow S_{\bar{x}} = \frac{5{,}3}{\sqrt{25}} = \frac{5{,}3}{5} = 1{,}06$

Formule invullen: ondergrens interval = $60 - 2{,}064 \cdot 1{,}06 = 60 - 2{,}19 = 57{,}81$.

De PvdA kan dus minimaal rekenen op 57 zetels.

b *Gegevens*: $n = 25$, $S_x = 5{,}3$, $\bar{x} = 60$

Opdracht: Op hoeveel zetels kan de PvdA minimaal rekenen met een zekerheid van 99,9%?

Formule: $\bar{x} - t_\alpha \cdot (s_{\bar{x}})$

Te bepalen:

$t_\alpha \rightarrow$ tabel B: zoek bij $df = n - 1 = 25 - 1 = 24$; $\alpha = 0{,}001$ tweezijdig, $t_\alpha = 3{,}745$. $s_{\bar{x}}$ heeft dezelfde waarde als in *a*, omdat de gegevens hetzelfde zijn, dus $s_{\bar{x}} = 1.06$.

Formule invullen: ondergrens interval = $60 - 3{,}745 \cdot 1{,}06 = 60 - 3{,}97 = 56{,}03$.

Als we de 0,03 verwaarlozen, dan kan de PvdA rekenen op minimaal 56 zetels met een zekerheid van 99,9%. Merk op hoe ontzettend weinig dit scheelt met 95% zekerheid!

c De PvdA wil met 99,9% zekerheid weten hoeveel zetels ze maximaal kan krijgen, om niet te weinig kandidaten te hebben om dit aantal zetels te bezetten. Met andere woorden: het gaat hier om de bovengrens van het interval.

Formule invullen: $60 + 3{,}97 \approx 64$, dus 64 kandidaten.

8 Vergelijking van de centra van twee onafhankelijke groepen: t-toets voor onafhankelijke groepen en de Mann-Whitney-U-toets

In dit hoofdstuk leren we de centra van twee onafhankelijke groepen te vergelijken en wel zó, dat we kunnen constateren dat het verschil tussen beide verdelingen significant is of niet. We kunnen dit doen aan de hand van twee toetsen, namelijk de t-toets voor onafhankelijke groepen en de Mann-Whitney-U-toets. De t-toets voor onafhankelijke groepen kan alleen gedaan worden als er aan bepaalde stringente voorwaarden voldaan is, dit noemen we een **parametrische toets**: in de hypothesen wordt een uitspraak gedaan over het al dan niet verschillen van parameters, in dit geval de populatiegemiddelden. De Mann-Whitney-U is de variant van deze toets en kan uitgevoerd worden in plaats van de t-toets, als er niet aan alle voorwaarden van deze t-toets kan worden voldaan. Dit noemen we een **niet-parametrische toets**: in de hypothesen wordt een uitspraak gedaan over het al dan niet verschillen van populatieverdelingen als geheel.
In de volgende hoofdstukken zullen we elke keer voor dezelfde situatie twee toetsen aantreffen, een parametrische en een niet-parametrische toets. Hieronder is in een schema weergegeven in welke situatie we voor een parametrische toets kiezen en wanneer we een niet-parametrische moeten uitvoeren. De keuze voor parametrisch of niet-parametrisch zal steeds in de toetskeuze verwerkt worden.

Parametrische toets als:	*Niet-parametrische toets als:*
1 de verdeling van de scores normaal van vorm is	**1** er twijfel aan de normaliteit van de verdeling is
2 er sprake is van een ratio- of interval- meetschaal	**2** gebruik wordt gemaakt van nominale of ordinale meetschaal
3 er gebruik is gemaakt van redelijk grote onderzoeksgroepen	**3** kleinere groepen de onderzoeksgroepen vormen
4 de groepen homogeen zijn	**4** de groepen niet homogeen zijn

De belangrijkste reden om voor een niet-parametrische toets te kiezen, is de twijfel aan de normale verdeling van de scores.

8.1 De t-toets voor onafhankelijke groepen

Deze toets stelt ons in staat het verschil te toetsen tussen twee onafhankelijke steekproefgemiddelden. **Onafhankelijk** wil zeggen dat de groepen waar de steekproefgemiddelden uit afkomstig zijn niets met elkaar te maken hebben. We zullen het gebruik van deze toets weer uitwerken aan de hand van een concrete situatie. We blijven daarbij het toetsingsschema uit hoofdstuk 6 volgen.

De stichting Consument en Geld stelt zich ten doel jongeren minder royaal met hun geld om te leren gaan. Het nieuwe budgetprogramma is een middel om dit doel te bereiken. Om te kijken of jongeren minder geld uitgeven na het zien van het budgetprogramma, laat de stichting een onderzoek uitvoeren.
Aselect worden er twee groepen samengesteld van ieder 8 middelbare scholieren. De eerste groep krijgt tijdens het lesuur maatschappijleer het budgetprogramma te zien, terwijl de andere leerlingen gewoon maatschappijleer volgen. De resultaten staan vermeld in tabel 8.1.

Tabel 8.1

Groep 1: Totale bedrag besteed in een week, in guldens	17	16	12	15	21	13	8	20
Groep 2: Totale bedrag besteed in een week, in guldens	25	17	14	16	22	13	10	19

Toets of het budgetprogramma een verminderend effect heeft gehad op de uitgaven van middelbare scholieren. Ga ervan uit dat de verdeling van bedragen normaal van vorm is.

8.1.1 *Toetskeuze*

Allereerst gaan we de situatie analyseren, zodat we tot een juiste toetskeuze kunnen komen. We moeten nagaan of groep 1 minder geld per week besteedde dan groep 2, we vergelijken dus twee groepen die op het oog niets met elkaar te maken hebben. Het gaat dus om het vergelijken van de gemiddelden van twee onafhankelijke groepen. Verder is ook vermeld dat de verdeling normaal van vorm is, dat is een argument om voor een parametrische variant te kiezen. We nemen hier dus de t-toets voor onafhankelijke groepen.

8.1.2 *Onderzoeksvraag*

Besteden de kinderen die het budgetprogramma hebben gevolgd minder geld per week dan de kinderen die het programma niet hebben gevolgd? Komt de steekproef van de kinderen uit groep 1 uit een populatie met een lager populatiegemiddelde dan de steekproef van de kinderen uit groep 2?

8.1.3 *Hypothesen*

We verwoorden de hypothesen in termen van de onderzoeksvraag op een andere manier dan bij de t-toets uit hoofdstuk 7. Daar vergeleken we een steekproefgemiddelde met een populatiegemiddelde. Nu vergelijken we twee steekproefgemiddelden, die van groep 1 en groep 2. We verwachten dat het gemiddelde van groep 1 lager zal zijn dan dat van groep 2 (vanwege het budgetprogramma). We zeggen dus in de alternatieve hypothese dat de populatieverdeling waar groep 1 uit afkomstig is, naar lagere cijfers tendeert (dus een lager populatiegemiddelde heeft). In de nulhypothese staat het tegenovergestelde van onze verwachting.

Hypothesen in termen van de onderzoeksvraag
H_0: De verdeling van groep 1 tendeert naar gelijke of hogere bedragen dan groep 2. De kinderen uit groep 1 geven evenveel of meer geld uit dan de kinderen in groep 2.
H_1: De verdeling van groep 1 tendeert naar lagere bedragen dan de verdeling van groep 2. De kinderen uit groep 1 geven minder geld uit dan de kinderen uit groep 2.
Hypothesen in statistische termen
H_0: $\mu_{\text{groep 1}} \geq \mu_{\text{groep 2}}$
H_1: $\mu_{\text{groep 1}} < \mu_{\text{groep 2}}$

8.1.4 *Aannamen*

De aannamen van de t-toets voor onafhankelijke groepen luiden als volgt:

1. De subjecten in elke groep zijn aselect en onafhankelijk gekozen.
2. De groepen zijn onafhankelijk.
3. De populatievarianties zijn homogeen.
4. De populatieverdeling $(\bar{x}_1 - \bar{x}_2)$ is normaal van vorm.

We zullen deze aannamen nader verduidelijken:

ad 1: De subjecten in elke groep zijn aselect en onafhankelijk gekozen.
Deze aanname staat voor een eerlijke steekproefprocedure. Wordt hier niet aan voldaan, dan kunnen de resultaten van de toets niet gegeneraliseerd worden naar de populatie middelbare scholieren.

ad 2: De groepen zijn onafhankelijk.
Deze aanname is belangrijk voor de onderlinge verhouding tussen de twee vergelijkingsgroepen. Groepen zijn in principe onafhankelijk als:

a de mensen in groep 1 niet dezelfde mensen zijn als in groep 2. Dit is hier het geval; het zou niet het geval zijn als het ging om de vergelijking van uitgaven in de voortest en de natest.

b de mensen in groep 1 geen nauwe connecties hebben met de mensen in groep 2. Dit zou wel zo zijn als we bijvoorbeeld steeds broer en zus vergelijken op hun uitgaven. In dit geval wordt daar niets over gezegd, dus gaan we ervan uit dat het niet zo is.

c er geen sprake is van 'matching'. Van matching is sprake als de mensen in groep 2 op bepaalde punten, anders dan de afhankelijke variabele 'uitgave', vergelijkbaar zijn. Dit proces noemen we **matchen**: bij iedere persoon uit de ene groep zoeken we een persoon uit de andere groep die bijvoorbeeld even oud is of uit hetzelfde milieu of gebied komt. In dit geval wordt niets vermeld over matchen.

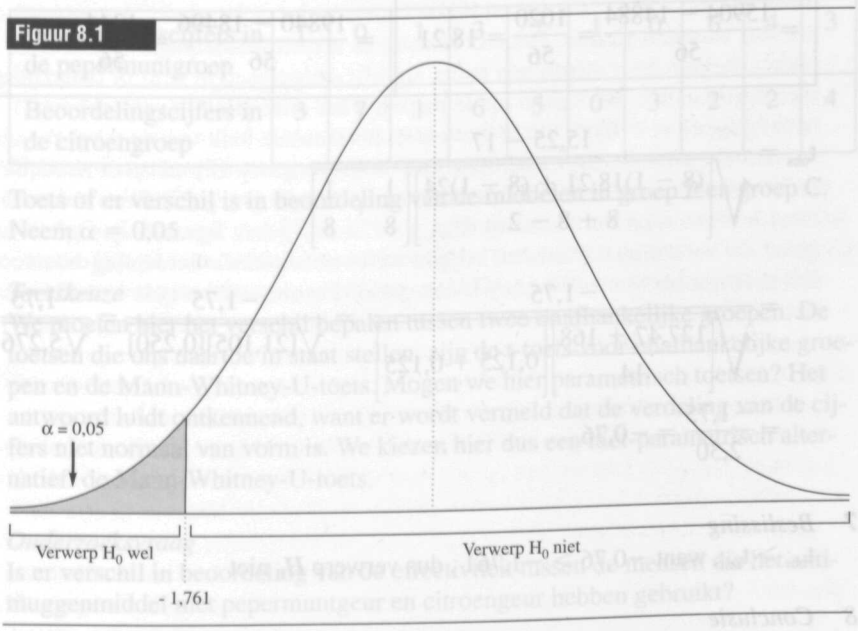
ad 3: De populatievarianties zijn homogeen.
Deze aanname beschrijft de situatie in de populaties. De varianties van de twee steekproeven mogen niet te veel verschillen, zodat de populaties waar de steekproeven uit getrokken zijn niet te veel verschillen qua spreiding.

ad 4: De populatieverdeling $(\bar{x}_1 - \bar{x}_2)$ is normaal van vorm.
Deze aanname geeft aan hoe deze toets in elkaar steekt. We berekenen een verschilscore tussen de gemiddelden van beide groepen (t_{obs}). Deze waarde plaatsen we in een verdeling waar de verschillen tussen de twee steekproefgemiddelden $(\bar{x}_1 - \bar{x}_2)$ zijn uitgezet. In het midden staat een 0, de twee gemiddelden verschillen dan niet. Gaat de t-waarde richting positief oneindig, dan is het gemiddelde van groep 1 hoger dan het gemiddelde van groep 2. Heeft de t-waarde een negatieve waarde, dan is het gemiddelde van groep 2 hoger dan dat van groep 1. De t-toets voor onafhankelijke groepen vertelt ons of het verschil tussen de twee gemiddelden groot genoeg is, of het om een *significant verschil* gaat.

8.1.5 *Beslissingsregels*

Het opstellen van de beslissingsregels bij deze toets verschilt niet veel van de procedure bij de toetsen uit hoofdstuk 7. We verwachten een negatief verschil tussen de gemiddelden van groep 1 en groep 2, we toetsen dus linkseenzijdig en we nemen een α van 0,05. In tabel B zoeken we de kritieke t-waarde. Het aantal vrijheidsgraden bij deze t-toets is gelijk aan:
$(n_{\text{groep 1}} - 1) + (n_{\text{groep 2}} - 1) \rightarrow (8 - 1) + (8 - 1) = 14$
In tabel B vinden we een $t_{\text{kritiek}} = -1{,}761$
In figuur 8.1 zien we de gegevens geprojecteerd.

Figuur 8.1



De beslissingsregels kunnen nu als volgt omschreven worden:
1 Indien $t_{obs} \leq -1{,}761$, verwerp $\boldsymbol{H_0}$
2 Indien $t_{obs} > -1{,}761$, verwerp $\boldsymbol{H_0}$ niet

8.1.6 Berekening

De formule om deze berekening uit te voeren, ziet er als volgt uit:

$$t_{obs} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\left[\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}\right]\left[\frac{1}{n_1} + \frac{1}{n_2}\right]}}$$

Allereerst moeten we de onbekende gegevens berekenen om de formule in te kunnen vullen. In tabel 8.2 berekenen we deze gegevens van beide groepen.

Tabel 8.2

Groep 1	*Groep 2*
$\bar{x}_1 = \frac{\sum x}{n} = \frac{122}{8} = 15{,}25$	$\bar{x}_2 = \frac{\sum x}{n} = \frac{136}{8} = 17$
$S_x^2 = \frac{n \cdot \sum x^2 - (\sum x)^2}{n(n-1)}$	$S_x^2 = \frac{n \cdot \sum x^2 - (\sum x)^2}{n(n-1)}$
$S_1^2 = \frac{8 \cdot 1988 - (122)^2}{8(8-1)}$	$S_2^2 = \frac{8 \cdot 2480 - (136)^2}{8(8-1)}$
$= \frac{15904 - 14884}{56} = \frac{1020}{56} = 18{,}21$	$= \frac{19840 - 18496}{56} = \frac{1344}{56} = 24$

$$t_{obs} = \frac{15{,}25 - 17}{\sqrt{\left[\frac{(8-1)18{,}21 + (8-1)24}{8+8-2}\right]\left[\frac{1}{8} + \frac{1}{8}\right]}}$$

$$= \frac{-1{,}75}{\sqrt{\left[\frac{127{,}47 + 168}{14}\right]\left[0{,}125 + 0{,}125\right]}} = \frac{-1{,}75}{\sqrt{[21{,}105][0{,}250]}} = \frac{-1{,}75}{\sqrt{5{,}276}}$$

$$= \frac{-1{,}75}{2{,}30} = -0{,}76$$

8.1.7 Beslissing

$t_{obs} > t_{krit}$, want $-0{,}76 > -1{,}761$, dus verwerp $\boldsymbol{H_0}$ niet

8.1.8 Conclusie

Er is geen reden om aan te nemen dat het volgen van het budgetprogramma een reducerend effect heeft op de geldbesteding van middelbare scholieren.

8.2 De Mann-Whitney-U-toets bij een kleine steekproef (n van beide groepen ≤ 20)

Als er niet aan de voorwaarden van de parametrische t-toets kan worden voldaan, dan kunnen we hetzelfde toetsingsprobleem benaderen met de niet-parametrische Mann-Whitney-U-toets.
Bij alle niet-parametrische toetsen gebruiken we in de berekening niet de gewone gegevens, maar een rangordening daarvan. Aan de hand van een concrete voorbeeldsituatie zullen we deze toets uitwerken.

In warme zomernachten worden heel wat mensen flink geplaagd door muggen. Voor de farmaceutische industrie is dit reden genoeg te werken aan een doeltreffend middel waardoor de muggen uit de buurt blijven. Uit onderzoek is duidelijk geworden dat muggen niet houden van pepermunt- en citroengeur. Om te kijken welke van de twee geuren het meest afstotend is voor muggen, wordt het volgende experiment gedaan. Aselect kiest men 20 man die in de fabriek werken waar men van plan is een doeltreffend middel op de markt te brengen. 10 mensen gaan een week lang naar bed met de pepermuntlucht (P) en de andere 10 slapen een weekje met een citroengeur (C). Ieder persoon geeft aan het einde van de week een cijfer tussen de 0 en de 10 voor de mate van effectiviteit van het antimuggenmiddel, waarbij het cijfer 0 absoluut niet effectief en het cijfer 10 zeer effectief representeert.
De verdeling van deze beoordelingscijfers bleek niet normaal van vorm. De resultaten van het onderzoekje staan in tabel 8.3.

Tabel 8.3

Beoordelingscijfers in de pepermuntgroep	1	0	1	3	2	1	0	6	4	3
Beoordelingscijfers in de citroengroep	3	7	1	6	5	0	3	2	2	4

Toets of er verschil is in beoordeling van de middelen in groep P en groep C.
Neem $\alpha = 0{,}05$.

8.2.1 *Toetskeuze*

We moeten hier het verschil bepalen tussen twee onafhankelijke groepen. De toetsen die ons daartoe in staat stellen, zijn de t-toets voor onafhankelijke groepen en de Mann-Whitney-U-toets. Mogen we hier parametrisch toetsen? Het antwoord luidt ontkennend, want er wordt vermeld dat de verdeling van de cijfers niet normaal van vorm is. We kiezen hier dus een niet-parametrisch alternatief: de Mann-Whitney-U-toets.

8.2.2 *Onderzoeksvraag*

Is er verschil in beoordeling van de effectiviteit tussen de mensen die het antimuggenmiddel met pepermuntgeur en citroengeur hebben gebruikt?

8.2.3 *Hypothesen*

De hypothesen van een niet-parametrische toets worden nooit in statistische termen verwoord, want de hypothesen van een niet-parametrische toets worden niet in parameters uitgedrukt. De hypothesen in termen van de onderzoeksvraag vormen dus de enige hypothesen.

Hypothesen in termen van de onderzoeksvraag:
H_0: De verdeling van groep P tendeert naar gelijke cijfers als groep C. De pepermuntgroep en de citroengroep geven eenzelfde beoordeling voor het door hen gebruikte middel.
H_1: De verdeling van groep P tendeert naar andere cijfers dan groep C. De pepermuntgroep en de citroengroep geven een andere beoordeling voor het door hen gebruikte middel.

8.2.4 *Aannamen*

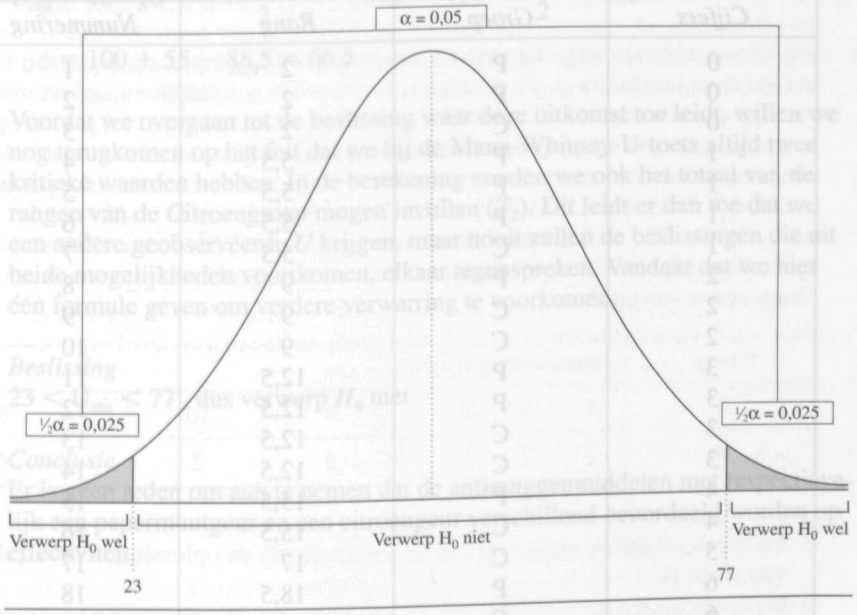
De aannamen van de Mann-Whitney-U-toets luiden als volgt:

1 De groepen zijn aselect en onafhankelijk samengesteld.
2 De groepen zijn onafhankelijk.
3 De afhankelijke variabele (beoordelingscijfer) is continu en de meetschaal is minimaal ordinaal.

8.2.5 *Beslissingsregels*

We voeren hier een tweezijdige toets uit bij een α van 0,05. De kritieke waarden van deze toets noemen we U_{kritiek}, de toetsingsgrootheid bij deze toets is dus U (bij de z-toets is dat z, en bij de t-toets is dat t). In tabel C kunnen we de kritieke waarden vinden. Tabel C bestaat uit verschillende pagina's die zich onderscheiden door steeds een ander significantieniveau in één- of tweezijdige toetsingssituaties. We kijken op de pagina waar onze situatie vermeld staat: tweezijdig en $\alpha = 0{,}05$. Op de bovenste, horizontale balk van de tabel zoeken we naar de juiste steekproefgrootte van de eerste groep (P) en op de verticale balk zoeken we de steekproefgrootte van de tweede groep (C). We vinden dan twee kritieke waarden, namelijk $U_{\text{kritiek}} = 23$ en 77 (zie figuur 8.2). Ook in het geval dat we eenzijdig toetsen, krijgen we twee kritieke waarden bij deze toets. Dit is alleen bij Mann-Whitney-U het geval! De oorzaak hiervan komt bij de berekening nader aan de orde.

Figuur 8.2



De beslissingsregels zien er als volgt uit:

1 Indien $23 < U_{obs} < 77$, verwerp $\boldsymbol{H_0}$ niet

2 Indien $U_{obs} \leq 23$ of $U_{obs} \geq 77$, verwerp $\boldsymbol{H_0}$

8.2.6 *Berekening*

Hieronder staat de formule om de geobserveerde U te berekenen:

$$U_{obs} = n_1 \cdot n_2 + \frac{n_1(n_1 + 1)}{2} - T_1$$

$\boldsymbol{T_1}$ staat voor de som van de rangen in groep 1 en om aan deze waarde te komen, moeten we gaan rangordenen. We ordenen hiertoe vooraf *alle* scores (van groep 1 en van groep 2) van klein naar groot; belangrijk is de vermelding van de groep waar de score uit afkomstig is. In tabel 8.4 zien we deze ordening met benoeming van de groepen respectievelijk in de eerste en tweede kolom. Daarna gaan we de scores rangordenen; deze procedure houdt in dat we de scores op volgorde van grootte een rang toekennen. De score die de laagste waarde heeft, krijgt rangnummer 1. In ons voorbeeld zien we echter dat er drie scores een 'laagste' waarde van 0 hebben. Ze zijn alle goed voor een eerste, tweede en derde rang; het gemiddelde wordt daarvan berekend, en zo komen de drie nullen op een 'gedeelde' tweede plaats. Het daaropvolgende getal is 1. Hier zijn er in totaal vier van, ze staan dus op een gedeelde vierde, vijfde, zesde en zevende plaats, wat neerkomt op een gemiddelde van 5,5. De rangnummers staan in de derde kolom van tabel 8.4. Om fouten te voorkomen bij de rangordening, kunnen we een rangtelling plaatsen naast de getallen (vierde kolom, tabel 8.4), zodat we elke keer kunnen zien bij welk rangnummer we gebleven zijn en tussen welke getallen we moeten middelen bij gelijke scores.

Tabel 8.4

Cijfers	*Groep*	*Rang*	*Nummering*
0	P	2	1
0	P	2	2
0	C	2	3
1	P	5,5	4
1	P	5,5	5
1	P	5,5	6
1	C	5,5	7
2	P	9	8
2	C	9	9
2	C	9	10
3	P	12,5	11
3	P	12,5	12
3	C	12,5	13
3	C	12,5	14
4	P	15,5	15
4	C	15,5	16
5	C	17	17
6	P	18,5	18
6	C	18,5	19
7	C	20	20

Achteraf kunnen we checken of we op de juiste manier gerangordend hebben via de volgende formule:

$$\frac{n \cdot (n + 1)}{2}$$

De uitkomst van deze formule moet gelijk zijn aan de optelling van *alle* toegekende rangen. Ons rangtotaal is gelijk aan de optelling van alle getallen in de derde kolom van tabel 8.4.

We komen dan op een rangtotaal van: $2 + 2 + 2 + 5{,}5 + 5{,}5 + 5{,}5 + 5{,}5 + 9 + 9 + 9 + 12{,}5 + 12{,}5 + 12{,}5 + 12{,}5 + 15{,}5 + 15{,}5 + 17 + 18{,}5 + 18{,}5 + 20 = 210$

De controleformule geeft de volgende uitkomst:

$$\frac{n \cdot (n + 1)}{2} = \frac{20 \cdot (21)}{2} = \frac{420}{2} = 210$$

De uitkomsten van beide berekeningen komen exact overeen, dus is de rangordening op juiste wijze uitgevoerd.

Nu komen we bij het uiteindelijke doel van deze procedure, namelijk de berekening van T_I. In dit geval komt dat neer op de optelling van alle rangen van de scores uit de P(epermunt)groep:

$T_I = 2 + 2 + 5\frac{1}{2} + 5\frac{1}{2} + 5\frac{1}{2} + 9 + 12\frac{1}{2} + 12\frac{1}{2} + 15\frac{1}{2} + 18\frac{1}{2} = 88\frac{1}{2}$

Nu kunnen we de formule invullen:

$$U_{obs} = 10 \cdot 10 + \frac{10\,(10+1)}{2} - 88{,}5 = 100 + \frac{110}{2} - 88{,}5$$

$$= 100 + 55 - 88{,}5 = 66{,}5$$

Voordat we overgaan tot de beslissing waar deze uitkomst toe leidt, willen we nog terugkomen op het feit dat we bij de Mann-Whitney-U-toets altijd twee kritieke waarden hebben. In de berekening zouden we ook het totaal van de rangen van de Citroengroep mogen invullen (T_2). Dit leidt er dan toe dat we een andere geobserveerde U krijgen, maar nooit zullen de beslissingen die uit beide mogelijkheden voortkomen, elkaar tegenspreken. Vandaar dat we hier één formule geven om verdere verwarring te voorkomen.

8.2.7 Beslissing

$23 < U_{obs} < 77$, dus verwerp H_0 niet

8.2.8 Conclusie

Er is geen reden om aan te nemen dat de antimuggenmiddelen met respectievelijk een pepermuntgeur en een citroengeur verschillend beoordeeld worden op effectiviteit.

8.3 Mann-Whitney-U-toets bij een grote steekproef (n van beide groepen > 20)

Als de twee groepen die we willen vergelijken beide een steekproefgrootte van meer dan 20 elementen hebben, dan mogen we in plaats van de hierboven uitgewerkte Mann-Whitney-U-toets een aangepaste versie van de z-toets gebruiken. Het blijft echter een niet-parametrische toetsingssituatie, dus aan de hypothesen en de aannamen verandert er niets in vergelijking met de Mann-Whitney-U. We gebruiken tabel A voor de kritieke waarde(n) van z. Als we tweezijdig toetsen, formuleren we op de gebruikelijke manier de tweezijdige beslissingsregels. Als we eenzijdig toetsen, dan toetsen we afhankelijk van welke U_{obs} we berekend hebben (met behulp van T_1 of T_2) links- of rechtseenzijdig. Omdat de keuze voor het gebruik van T_1 of T_2 willekeurig is, berekenen we eerst de geobserveerde z. Daarna formuleren we de linkseenzijdige beslissingsregels indien deze z-waarde negatief is en de rechtseenzijdige beslissingsregels indien deze z-waarde positief is. De berekening van de geobserveerde z ziet er als volgt uit:

$$z_{obs} = \frac{U_{obs} - n_1 n_2 / 2}{\sqrt{\frac{(n_1)\,(n_2)\,(n_1 + n_2 + 1)}{12}}}$$

Opgaven

8.1 Er is een theorie die zegt dat mensen een taak beter uitvoeren als men erbij vertelt dat deze makkelijk is. 20 proefpersonen worden gelijkelijk en aselect over twee groepen verdeeld. Alle proefpersonen krijgen de opdracht een puzzel op te lossen. Groep A krijgt te horen dat het een makkelijke puzzel is en dat iedereen hem wel zal oplossen. Groep B wordt verteld dat het een moeilijke puzzel is en dat het niet erg is als het niet lukt. De resultaten staan in onderstaande tabel. De getallen stellen een cijfer tussen 1 en 10 voor, dat de kwaliteit van de oplossing aangeeft. Kies een adequate toets voor dit probleem en voer deze uit. Neem een α van 0,05.

Groep	*Kwaliteit oplossing*									
A	4	6	8	7	9	8	10	5	6	9
B	7	7	5	3	5	4	2	1	3	6

a Ga ervan uit dat de verdeling van de kwaliteit van de oplossingen normaal van vorm is.

b Ga ervan uit dat de verdeling van de kwaliteit van de oplossingen niet normaal van vorm is.

8.2 Omdat we nog steeds vaak horen dat meisjes slechter in wiskunde zijn dan jongens, is er op een middelbare school een onderzoekje gedaan om te kijken of deze uitspraak in deze moderne tijd nog hout snijdt. 18 jongens en 12 meisjes zijn aselect gekozen uit de totale populatie brugklassers. Hun wiskundecijfers van het derde rapport staan hieronder afgebeeld.

Rapportcijfer Wiskunde	Frequentie jongens	Frequentie meisjes
3	1	1
4	1	0
5	2	3
6	4	3
7	5	4
8	4	0
9	1	1

a Als we ervan uitgaan dat de verdeling van rapportcijfers Wiskunde normaal verdeeld is, toets dan of meisjes gemiddeld een lager rapportcijfer behalen dan jongens. Neem $\alpha = 0{,}01$.
(deze opgave heeft een beknopte uitwerking)

b Als we ervan uitgaan dat de verdeling van de rapportcijfers niet normaal van vorm is, toets dan nogmaals of meisjes een slechter rapportcijfer hebben dan jongens. Neem $\alpha = 0{,}01$.
(deze opgave heeft een beknopte uitwerking)

c Stel dat dezelfde onderzoeksvraag onderzocht is bij 40 jongens en 38 meisjes. Neem aan dat het totaal van de rangen voor de jongens en meisjes gelijk is aan het totaal van de rangen voor jongens en meisjes uit antwoord van 8.2.b. Toets dan opnieuw of meisjes slechter in wiskunde zijn dan jongens. Neem $\alpha = 0{,}01$.

Uitwerkingen

8.1 Situatie:
Groep A heeft te horen gekregen dat de puzzel makkelijk is. Groep B heeft te horen gekregen dat de puzzel moeilijk is. Het is nu de bedoeling om de groepsgemiddelden te gaan vergelijken. Dat doen we door middel van een toets. We toetsen of het verschil tussen de groepen groot genoeg is (significant), om aan te nemen dat het in de steekproef gevonden verschil naar alle waarschijnlijkheid ook in de populatie te vinden is. Met andere woorden: dat de gevonden verschillen in de steekproef niet toevallig ontstaan zijn door fluctuaties in de steekproef. Nog anders gezegd: kunnen de steekproefresultaten gegeneraliseerd worden naar de populatie?

Gegevens

Groep	*Kwaliteit oplossing*									
A	4	6	8	7	9	8	10	5	6	9
B	7	7	5	3	5	4	2	1	3	6

Onderzoeksvraag
Klopt de theorie dat mensen een opdracht beter uitvoeren als erbij wordt gezegd dat deze makkelijk is?

Toetskeuze
De groepen A en B hebben blijkbaar niets met elkaar te maken, daar wordt niets over vermeld. Dus onafhankelijke groepen. In opgave *a* wordt er gevraagd te toetsen op basis van de aanname van een normale verdeling, we voeren dan een parametrische toets uit. In opgave *b* wordt er gesteld dat de verdeling niet normaal van vorm is, we voeren dan een niet-parametrische toets uit. De keuze valt hier dan op:

a Parametrisch: t-toets voor twee onafhankelijke steekproeven
b Niet-parametrisch: Mann-Whitney-U-toets.

De onderzoeksvraag geeft richting aan het verschil tussen groep A en B, we verwachten dat groep A met betere oplossingen komt dan groep B, dit heeft tot gevolg dat we beide toetsen rechtseenzijdig uitvoeren.

a t-toets voor twee onafhankelijke steekproeven

Hypothesen

Hypothesen in termen van de onderzoeksvraag
H_0: De verdeling van groep A tendeert naar gelijke of lagere cijfers dan groep B. De A-mensen komen met slechtere of even goede oplossingen als de B-mensen.
H_1: De verdeling van groep A tendeert naar hogere cijfers dan de verdeling van groep B. De mensen uit groep A lossen de puzzel beter op dan de B-mensen.
Hypothesen in statistische termen
H_0: $\mu_{\text{groep A}} \leq \mu_{\text{groep B}}$
H_1: $\mu_{\text{groep A}} > \mu_{\text{groep B}}$

Aannamen

De aanamen van de t-toets voor onafhankelijke groepen luiden als volgt:

1 De subjecten in elke groep zijn aselect en onafhankelijk gekozen.
2 De groepen zijn onafhankelijk.
3 De populatievarianties zijn homogeen.
4 De populatieverdeling $(\bar{x}_1 - \bar{x}_2)$ is normaal van vorm.

Beslissingsregels

We verwachten een positief verschil tussen de gemiddelden van groep A en groep B, we toetsen dus rechtseenzijdig en we nemen een α van 0,05. In tabel B zoeken we de kritieke t-waarde. Het aantal vrijheidsgraden bij deze t-toets is gelijk aan:

$(n_{\text{groep A}} - 1) + (n_{\text{groep B}} - 1) \rightarrow (10 - 1) + (10 - 1) = 18$

In tabel B vinden we een $t_{\text{kritiek}} = 1{,}734$.

De beslissingsregels kunnen nu als volgt omschreven worden:

1 Indien $t_{\text{obs}} \geq 1{,}734$, verwerp H_0
2 Indien $t_{\text{obs}} < 1{,}734$, verwerp H_0 niet

Berekening

We moeten nu t_{obs} gaan berekenen om te kijken of het verschil tussen groep A en groep B significant is, met andere woorden de kritieke waarde overschrijdt.

$$t_{\text{obs}} = \frac{\bar{x}_A - \bar{x}_B}{\sqrt{\left[\frac{(n_A - 1)s_A^2 + (n_B - 1)s_B^2}{n_A + n_B - 2}\right]\left[\frac{1}{n_A} + \frac{1}{n_B}\right]}}$$

Om de formule in te kunnen vullen, moeten we eerst de spreiding en het gemiddelde van de verdelingen berekenen. In de onderstaande tabel zijn deze waarden bepaald voor beide groepen.

Groep A:	*Groep B:*
$\bar{x}_A = \frac{\sum x}{n} = \frac{72}{10} = 7{,}2$	$\bar{x}_B = \frac{\sum x}{n} = \frac{43}{10} = 4{,}3$
$S_x^2 = \frac{n \cdot \sum x^2 - (\sum x)^2}{n(n-1)}$	$S_x^2 = \frac{n \cdot \sum x^2 - (\sum x)^2}{n(n-1)}$
$S_A^2 = \frac{10 \cdot 552 - (72)^2}{10(10-1)}$	$S_B^2 = \frac{10 \cdot 223 - (43)^2}{10(10-1)}$
$= \frac{5520 - 5184}{90} = \frac{336}{90} = 3{,}73$	$= \frac{2230 - 1849}{90} = \frac{381}{90} = 4{,}23$

$$t_{obs} = \frac{7{,}2 - 4{,}3}{\sqrt{\left[\frac{(10-1)3{,}73 + (10-1)4{,}23}{10+10-2}\right]\left[\frac{1}{10} + \frac{1}{10}\right]}}$$

$$= \frac{2{,}9}{\sqrt{\left[\frac{33{,}57 + 38{,}07}{18}\right]\left[0{,}1 + 0{,}1\right]}} = \frac{2{,}9}{\sqrt{[3{,}98][0{,}2]}} = \frac{2{,}9}{\sqrt{0{,}80}}$$

$$= \frac{2{,}9}{0{,}89} = 3{,}26$$

Beslissing

$t_{obs} > t_{krit}$, want $3{,}26 > 1{,}734$, dus verwerp $\boldsymbol{H_0}$

Conclusie

Er is reden om aan te nemen dat mensen een taak beter uitvoeren als erbij wordt gezegd dat deze makkelijk is.

b Mann-Whitney-U-toets

Dit is de niet-parametrische toets voor het vergelijken van de centra van twee onafhankelijke steekproeven.

Onderzoeksvraag

Klopt de theorie dat mensen een taak beter uitvoeren als erbij wordt gezegd dat deze makkelijk is?

Hypothesen

Hypothesen in termen van de onderzoeksvraag
H_0: De verdeling van groep A tendeert naar gelijke of lagere cijfers dan groep B. Groep A komt gemiddeld tot gelijke of lagere oplossingscijfers dan groep B.
H_1: De verdeling van groep A tendeert naar hogere cijfers dan groep B. Groep A komt gemiddeld tot hogere oplossingscijfers dan groep B.

Aannamen

1 De groepen zijn aselect en onafhankelijk samengesteld.
2 De groepen zijn onafhankelijk.
3 De afhankelijke variabele (cijfers) is continu en de meetschaal is minimaal ordinaal.

Beslissingsregels

We voeren hier een rechtseenzijdige toets uit bij een α van 0,05. De kritieke waarden van deze toets noemen we $U_{kritiek}$, de toetsingsgrootheid van deze toets is dus U (bij de z-toets is dat z, en bij de t-toets is dat t). In tabel C kunnen we de kritieke waarden vinden. Op de horizontale bovenbalk van de tabel zoeken we naar de juiste steekproefgrootte van de eerste groep (A) en op de verticale balk zoeken we de steekproefgrootte van de tweede groep (B). We vinden dan twee kritieke waarden, namelijk $U_{kritiek} = 27$ en 73. Ondanks het feit dat we eenzijdig toetsen, hebben we twee kritieke waarden bij deze toets.

De beslissingsregels zien er als volgt uit:
1 Indien $27 < U_{obs} < 73$, verwerp H_0 niet
2 Indien $U_{obs} \leq 27$ of $U_{obs} \geq 73$, verwerp H_0

Berekening

Het basisidee bij deze toets is, dat als de verdeling van groep A rechts ligt van die van groep B (d.w.z. de A-personen scoren als groep hoger dan de B-personen), de rangnummers van de A-groep hoger moeten zijn dan die van de B-groep. We moeten dus de sommen van rangnummers bekijken.

$$U_{obs} = (n_A n_B) + \frac{n_A (n_A + 1)}{2} - T_A$$

Alles is bekend behalve T_A, waarbij T_A staat voor de som van de rangen van de A-scores. Om T_A te bepalen moeten we alle scores rangordenen, dus zowel die van groep A als die van groep B. Zet alle scores op een rij, van klein naar groot, zet de groep erbij en ken dan rangen, nummers of gemiddelde rangnummers toe.

Scores	*Groep*	*Rang*
1	B	1
2	B	2
3	B	3,5
3	B	3,5
4	B	5,5
4	A	5,5
5	B	8
5	B	8
5	A	8
6	B	11
6	A	11
6	A	11
7	B	14
7	B	14
7	A	14
8	A	16,5
8	A	16,5
9	A	18,5
9	A	18,5
10	A	20

Omdat er snel fouten sluipen in het rangordenen, is controle gewenst.
Formule voor controle:

$$\frac{n(n+1)}{2} = \frac{20(21)}{2} = 210$$

Dit getal moet hetzelfde zijn als alle opgetelde rangnummers en dat klopt. De som van alle rangen is gelijk aan 210.

Nu kunnen we de onbekende T_A uitrekenen.

$T_A = 5\frac{1}{2} + 8 + 11 + 11 + 14 + 16\frac{1}{2} + 16\frac{1}{2} + 18\frac{1}{2} + 18\frac{1}{2} + 20 = 139\frac{1}{2}$

Formule invullen:

$$U_{obs} = (10 \cdot 10) + \frac{10(11)}{2} - 139{,}5 = 155 - 139{,}5 = 15{,}5$$

Zoals we al in de uitleg van dit hoofdstuk hebben gezien, mogen we in plaats van T_A ook T_B invullen. We zullen dit hier demonstreren:

$T_B = 1 + 2 + 3\frac{1}{2} + 3\frac{1}{2} + 5\frac{1}{2} + 5\frac{1}{2} + 8 + 8 + 11 + 14 + 14 = 70\frac{1}{2}$

$$U_{obs} = (10 \cdot 10) + \frac{10(11)}{2} - 70{,}5 = 155 - 70{,}5 = 84{,}5$$

Beslissing

In beide gevallen valt de geobserveerde in het kritieke gebied, de eerste valt in het linker kritieke gebied:

$U_{obs} < U_{krit}$, want $15{,}5 < 27$, verwerp $\boldsymbol{H_0}$.

De tweede keer vulden we het totaal van de rangen van groep B in, wat resulteerde in een geobserveerde waarde in het rechter kritieke gebied:
$U_{obs} > U_{krit}$, want $84{,}5 > 73$, verwerp H_0.
We kunnen dus concluderen dat de beslissing gelijk blijft, of we nou het rangtotaal van groep A invullen, of dat van groep B. Dit is mogelijk doordat de Mann-Whitney-U ook bij rechtseenzijdige toetsen twee kritieke waarden oplevert.

Conclusie
Er is reden om aan te nemen dat groep A, die te horen kreeg dat de opdracht makkelijk was, beter presteert. Merk op: de parametrische en de niet-parametrische toets leveren hetzelfde resultaat op.

8.2 *a* t-toets voor onafhankelijke groepen

Hypothesen in statistische termen
H_0: $\mu_{meisjes} \geq \mu_{jongens}$
H_1: $\mu_{meisjes} < \mu_{jongens}$

Beslissingsregels
1 Indien $t_{obs} \leq -2{,}467$, verwerp H_0
2 Indien $t_{obs} > -2{,}467$, verwerp H_0 niet

Berekening
$t_{obs} = -0{,}74$

Beslissing
$t_{obs} > t_{krit}$, want $-0{,}74 > -2{,}467$, dus verwerp H_0 niet

Conclusie
Er is geen reden om aan te nemen dat meisjes gemiddeld een lager rapportcijfer voor wiskunde hebben dan jongens. Hoewel hun gemiddelde iets lager is dan dat van de jongens, is dit verschil te klein om te kunnen zeggen dat het systematisch verschilt van het gemiddelde van de jongens. Dit verschil is dus op toeval gebaseerd.

b Mann-Whitney-U-toets

Hypothesen in termen van de onderzoeksvraag
H_0: De verdeling van meisjes tendeert naar gelijke of hogere cijfers dan de verdeling van jongens.
H_1: De verdeling van meisjes tendeert naar lagere cijfers dan de verdeling van jongens.

Beslissingsregels

1 Indien $U_{obs} \leq 53$ of $U_{obs} \geq 163$, verwerp H_0
2 Indien $53 < U_{obs} < 163$, verwerp H_0 niet

Berekening
We berekenen U_{obs}

Totaal van de rangen voor meisjes = 165
Totaal van de rangen voor jongens = 300

Formule invullen:

$$U_{obs} = (12 \cdot 18) + \frac{12(13)}{2} - 165 = 294 - 165 = 129$$

Beslissing
$53 < U_{obs} < 163$, verwerp H_0 niet.

Conclusie
Er is geen reden om aan te nemen dat meisjes een lager rapportcijfer voor wiskunde hebben dan jongens.

c Mann-Whitney-U-toets met de daaraan gekoppelde z-toets (n van beide groepen > 20)

Hypothesen in termen van de onderzoeksvraag
H_0: De verdeling van meisjes tendeert naar gelijke of hogere cijfers dan de verdeling van jongens.
H_1: De verdeling van meisjes tendeert naar lagere cijfers dan de verdeling van jongens.

Berekening
We berekenen eerst U_{obs}

Totaal van de rangen voor meisjes = 165
Totaal van de rangen voor jongens = 300

Formule invullen:

$$U_{obs} = (38 \cdot 40) + \frac{38(39)}{2} - 165 = 2261 - 165 = 2096$$

Nu berekenen we de z_{obs}:

$$z_{obs} = \frac{2096 - 38 \cdot 40 / 2}{\sqrt{\frac{(38) \cdot (40) \cdot (38 + 40 + 1)}{12}}} = 13{,}36$$

Beslissingsregels
Omdat de z_{obs} positief is, formuleren we nu de rechtseenzijdige beslissingsregels:
1 Indien $U_{obs} \geq 2{,}33$, verwerp H_0
2 Indien $z_{obs} < 2{,}33$, verwerp H_0 niet

Beslissing
$13{,}36 > 2{,}33$, verwerp H_0 .

Conclusie
Er is nu wel reden om aan te nemen dat meisjes een lager rapportcijfer voor wiskunde hebben dan jongens.

9 Vergelijking van twee afhankelijke groepen: t-toets voor afhankelijke groepen en de toets van Wilcoxon

In hoofdstuk 8 hebben we geleerd twee onafhankelijke groepen te vergelijken. In dit hoofdstuk gaan we twee afhankelijke groepen met elkaar vergelijken. We noemen twee groepen van observaties **afhankelijk** als:

1 de eerste groep van observaties de voortest vormt van de tweede groep observaties (= natest). Bijvoorbeeld: we vergelijken de statistiekprestaties vóór een bijlessessie (eerste groep observaties) met de statistiekprestaties ná de bijlessessie (tweede groep observaties). Eigenlijk vergelijken we dan steeds dezelfde personen voor en na een behandeling.
2 de mensen in de twee groepen een bepaalde 'relatie' met elkaar hebben. Bijvoorbeeld: we vergelijken steeds broers (eerste groep) met hun zussen (tweede groep). Of vaders (eerste groep) met moeders uit hetzelfde gezin (tweede groep).
3 bij de mensen uit de ene groep, 'vergelijkbare' personen in de andere groep zijn gekozen, zonder dat deze mensen elkaar kennen. Als we bijvoorbeeld de lichaamsconditie van twee groepen willen vergelijken, dan kunnen we beter niet een twintigjarige man (in de ene groep) vergelijken met een tachtigjarige vrouw (in de andere groep). Het is beter om bij de twintigjarige man in de ene groep, iemand in de andere groep te zoeken die een vergelijkbare leeftijd heeft, zodat we geen onmogelijke vergelijking uitvoeren. We noemen deze procedure **matching**, wat inhoudt dat we vergelijkingsgroepen samenstellen waarvan de personen in de ene groep op bepaalde variabelen, anders dan de afhankelijke variabele, 'gelijk' zijn.

Het verschil tussen de toetsen van dit hoofdstuk en die uit hoofdstuk 8, is dat we hier geen rekening hoeven te houden met steekproeffluctuaties. Als mensen in de twee groepen onder dezelfde omstandigheden leven, of vergelijkbaar zijn op bepaalde punten, of zelfs dezelfde personen zijn, dan kan een verschil tussen de twee groepen van observaties eerder toegeschreven worden aan de behandeling (onafhankelijke variabele). Zijn de groepen niet afhankelijk, dan zullen we rekening moeten houden met het feit dat er verschillen zijn tussen de mensen, zonder die ondubbelzinnig toe te kunnen schrijven aan de behandeling. De toetsen uit hoofdstuk 8 zijn zodanig gestructureerd dat er een redelijk groot verschil moet zijn voordat we overgaan tot de verwerping van de nulhypothese. Bij de toetsen uit dit hoofdstuk zullen we, naar verhouding, sneller overgaan tot verwerping van de nulhypothese. Een relatief kleiner verschil is bij deze toetsen eerder significant, omdat we ervan uit mogen gaan dat de verschillen komen door de behandeling.
Ook voor het toetsen van verschil tussen twee afhankelijke groepen hebben we de beschikking over twee varianten: de parametrische t-toets voor afhankelijke groepen en de niet-parametrische toets van Wilcoxon.
We zullen het gebruik van deze toetsen weer illustreren aan de hand van concrete voorbeeldsituaties.

9.1 De t-toets voor afhankelijke groepen

Een docent statistiek heeft het idee dat zijn lessen als saai worden ervaren. Om dit te controleren, kiest hij aselect eerstejaarsstudenten (in 1989) en vraagt hun een cijfer voor zijn colleges te geven. Hij benadrukt het feit dat ze eerlijk hun mening mogen geven, zonder dat ze bang hoeven te zijn voor eventuele represailles. De cijfers liegen er niet om, zoals we kunnen zien in tabel 9.1. De docent geeft de moed niet op en duikt in didactiekboeken, zorgt voor afbeeldingen bij de lesstof en gaat naar een assertiviteitscursus. Hij ontwikkelt op deze manier een, naar zijn mening, aantrekkelijker onderdeel. Hij presenteert dit onderdeel in 1990 aan de nieuwe eerstejaarsstudenten. Aan het eind van de cursus vraagt hij 7 studenten weer eerlijk naar de mening over het onderwijsonderdeel. De docent kiest deze 7 studenten zo, dat ze elk zo veel mogelijk overeenkomen met de studenten die in 1989 hun mening gaven, op de volgende punten:

1. wiskunde-vooropleiding.
2. leeftijd.
3. geslacht.

Tabel 9.1

Waarderingscijfers in 1989	*Waarderingscijfers in 1990*
4	5
3	7
2	4
5	6
4	3
7	9
5	7

Klopt het, dat het onderwijs hoger gewaardeerd wordt na de aangebrachte verbeteringen? Hanteer een significantieniveau van 5%, en ga ervan uit dat de verdeling van waarderingscijfers normaal van vorm is.

9.1.1 *Toetskeuze*

We hebben hier te maken met een toetsingssituatie waarin we het verschil moeten vaststellen tussen twee groepen. De vraag moet nu beantwoord worden of het hier gaat om afhankelijke of onafhankelijke groepen observaties. De eerste groep studenten is op een aselecte wijze totstandgekomen, de tweede groep studenten is echter zo gekozen dat de personen in deze groep op bepaalde punten overeenkomen met de eerste groep studenten. Hier is dus sprake geweest van matching, met als matching-variabelen wiskunde-vooropleiding, leeftijd en geslacht. We hebben hier dus te maken met afhankelijke groepen. Verder is er gegeven dat de verdeling van waarderingscijfers normaal van vorm is, we kunnen dus kiezen voor een parametrische toets, en zo komen we bij de t-toets voor afhankelijke groepen. Deze toets zal rechtseenzijdig uitgevoerd worden,

omdat de docent verwacht dat het onderwijsonderdeel, na de aangebrachte veranderingen, hoger gewaardeerd zal worden.

9.1.2 Onderzoeksvraag

Wordt het statistiekonderwijs van deze docent in 1990 hoger gewaardeerd dan in 1989? Heeft de andere onderwijsaanpak geleid tot een aantrekkelijker onderwijsonderdeel dat hoger gewaardeerd wordt?

9.1.3 Hypothesen

We verwoorden de alternatieve hypothese zo, dat we verwachten dat de waarderingscijfers in 1990 hoger zijn dan in 1989 (rechtseenzijdig). Dit is een keuze, we kunnen ook de verwachting uitspreken dat de waarderingscijfers in 1989 lager zijn dan in 1990 (linkseenzijdig). De beide alternatieve hypothesen komen op hetzelfde neer, alleen de richting waarin we de afwijking verwachten is anders. We moeten goed in de gaten houden hoe we de verwachting uitspreken, en dat we consequent de verwachte richting van de afwijking aanhouden.

Hypothesen in termen van de onderzoeksvraag
H_0: De verdeling van waarderingscijfers in 1990 tendeert naar gelijke of lagere cijfers dan de verdeling van waarderingscijfers in 1989. De veranderingen in het onderwijs leiden tot hetzelfde of een lager waarderingscijfer.
H_1: De verdeling van waarderingscijfers in 1990 tendeert naar hogere cijfers dan de verdeling van waarderingscijfers in 1989. De veranderingen in het onderwijs leiden tot hogere waarderingscijfers.
Hypothesen in statistische termen
H_0: $\mu_{1990} \leq \mu_{1989}$
H_1: $\mu_{1990} > \mu_{1989}$

9.1.4 Aannamen

De aanamen van de t-toets voor afhankelijke groepen luiden als volgt:

1 De subjecten zijn aselect en onafhankelijk gekozen.
2 De scores van de groepen zijn verwant.
3 De populatieverdeling van D_i is normaal van vorm. D_i staat voor de verdeling van de verschilscores tussen de twee steekproeven.

9.1.5 Beslissingsregels

We verwachten een positief verschil tussen de verdeling van 1990 en de verdeling van 1989, we toetsen dus rechtseenzijdig en nemen een α van 0,05. In tabel B zoeken we de kritieke t-waarde. Het aantal vrijheidsgraden bij deze t-toets is gelijk aan: $(n - 1)$, waarbij n het aantal paren scores aangeeft. Er is hier sprake van 7 paren scores $\rightarrow$ $df = (n - 1) = (7 - 1) = 6$

In tabel B vinden we een $t_{kritiek} = 1{,}943$

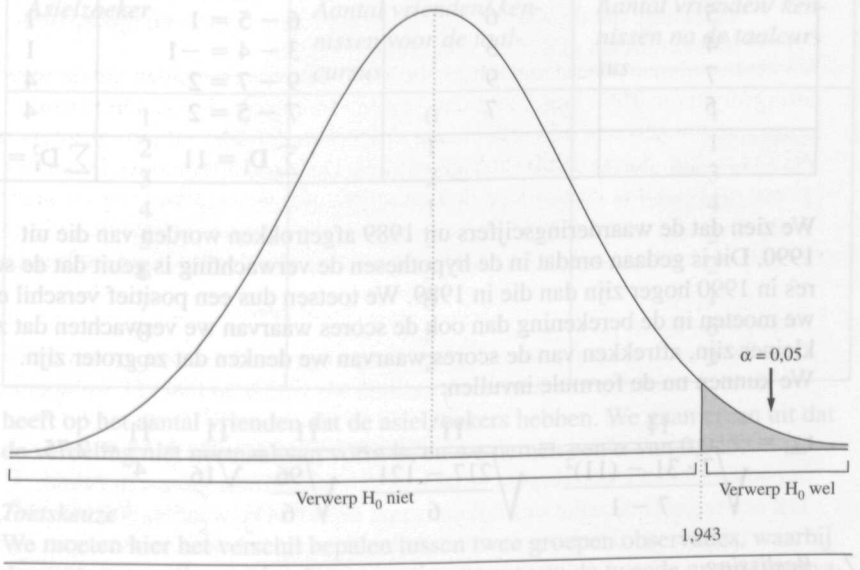
De beslissingsregels kunnen nu als volgt omschreven worden:

1 Indien $t_{obs} \geq 1{,}943$, verwerp $\boldsymbol{H_0}$

2 Indien $t_{obs} < 1{,}943$, verwerp $\boldsymbol{H_0}$ niet

In figuur 9.1 zien we de gegevens geprojecteerd.

Figuur 9.1



9.1.6 Berekening

De formule die bij deze t-toets gebruikt wordt om de geobserveerde t-waarde te vinden, ziet er als volgt uit:

$$t_{obs} = \frac{\sum D_i}{\sqrt{\frac{n\sum D_i^2 - (\sum D_i)^2}{n-1}}}$$

Deze formule gebruiken we als de ruwe scores gegeven zijn. Indien alleen de gemiddelden van de twee groepen gegeven zijn en de standaardafwijking van de verschilscores, dan kunnen we het best gebruikmaken van de volgende formule:

$$t_{obs} = \frac{\overline{D}}{S_D / \sqrt{N}}$$

$\overline{D}$ = het verschil tussen de gegeven gemiddelden

s_D = de standaardafwijking van de verschilscores

Omdat de ruwe scores hier gegeven zijn, berekenen we de eerste formule. Allereerst moeten we de onbekende gegevens berekenen om de formule in te kunnen vullen. In tabel 9.2 berekenen we deze gegevens van beide groepen.

Tabel 9.2

Waarderingscijfers in 1989	*Waarderingscijfers in 1990*	*Verschilscores* $D_i = score_{1990} - score_{1989}$	D_i^2
4	5	$5 - 4 = 1$	1
3	7	$7 - 3 = 4$	16
2	4	$4 - 2 = 2$	4
5	6	$6 - 5 = 1$	1
4	3	$3 - 4 = -1$	1
7	9	$9 - 7 = 2$	4
5	7	$7 - 5 = 2$	4
		$\sum D_i = 11$	$\sum D_i^2 = 31$

We zien dat de waarderingscijfers uit 1989 afgetrokken worden van die uit 1990. Dit is gedaan omdat in de hypothesen de verwachting is geuit dat de scores in 1990 hoger zijn dan die in 1989. We toetsen dus een positief verschil en we moeten in de berekening dan ook de scores waarvan we verwachten dat ze kleiner zijn, aftrekken van de scores waarvan we denken dat ze groter zijn.
We kunnen nu de formule invullen:

$$t_{obs} = \frac{11}{\sqrt{\frac{7 \cdot 31 - (11)^2}{7 - 1}}} = \frac{11}{\sqrt{\frac{217 - 121}{6}}} = \frac{11}{\sqrt{\frac{96}{6}}} = \frac{11}{\sqrt{16}} = \frac{11}{4} = 2{,}75$$

9.1.7 Beslissing

$t_{obs} > t_{krit}$, want $2{,}75 > 1{,}943$, dus verwerp H_0

9.1.8 Conclusie

Er is reden om aan te nemen dat de aangebrachte veranderingen in het statistiekonderwijs tot hogere waarderingscijfers hebben geleid.

9.2 De Wilcoxon-toets bij kleine steekproefomvang (n (= aantal paren scores) ≤ 50)

Nederland kent, net als een groot aantal andere landen in Europa, het probleem van de asielzoekers. Mensen die om politieke redenen niet meer in hun eigen land kunnen of durven wonen, zoeken een nieuwe verblijfplaats. De mensen worden in eerste instantie opgevangen, maar na verloop van tijd zullen ze zich zelf in ons land moeten redden. Dat geeft nogal wat problemen, vooral bij de integratie van deze mensen. Een stichting die zich inzet voor deze mensen, signaleert het probleem dat de asielzoeker vooral moeite heeft met contacten leggen met de Nederlandse bevolking. De stichting wil dit probleem oplossen door de asielzoekers een cursus Nederlandse taal en cultuur aan te bieden. Zij voeren een experiment uit met voor- en nameting. Aselect worden er 9 asielzoekers uitgekozen en men vraagt hun naar het aantal Nederlandse vrienden of kennissen dat ze op dat moment hebben. Daarna geeft men de cursus van

3 maanden, en een maand na deze cursus vraagt men deze mensen weer naar het aantal vrienden en kennissen. De resultaten staan in tabel 9.3.

Tabel 9.3

We gaan toetsen of de cursus Nederlandse taal en cultuur een positief effect

Asielzoeker	*Aantal vrienden/ kennissen voor de taalcursus*	*Aantal vrienden/ kennissen na de taalcursus*
1	0	1
2	1	1
3	2	3
4	1	4
5	2	3
6	3	2
7	0	1
8	5	6
9	2	3

heeft op het aantal vrienden dat de asielzoekers hebben. We gaan ervan uit dat de verdeling niet normaal van vorm is, en we nemen een α van 0,025.

9.2.1 Toetskeuze

We moeten hier het verschil bepalen tussen twee groepen observaties, waarbij de eerste groep observaties de voormeting vormt van de tweede groep observaties. We hebben hier dus te maken met afhankelijke groepen. We gaan de toets van Wilcoxon uitvoeren, omdat er in de vraag is gesteld dat de verdeling van aantal vrienden/kennissen niet normaal van vorm is, wat tot gevolg heeft dat we voor een niet-parametrisch alternatief moeten kiezen. Ten slotte wordt er in de vraag een bepaalde richting van de afwijking verwacht, we toetsen daarom rechtseenzijdig.

9.2.2 Onderzoeksvraag

Heeft de cursus Nederlandse taal en cultuur tot gevolg dat de asielzoekers zich meer vrienden/kennissen kunnen verwerven?

9.2.3 Hypothesen

Ook bij deze niet-parametrische toets gebruiken we geen hypothesen in statistische termen; we verwoorden de hypothesen in termen van de onderzoeksvraag.

Hypothesen in termen van de onderzoeksvraag
H_0: De verdeling van de nameting tendeert naar gelijke of lagere aantallen dan de verdeling van de voormeting. In de nameting heeft men een gelijk aantal of minder vrienden/kennissen dan in de voormeting.
H_1: De verdeling van de nameting tendeert naar hogere aantallen dan de verdeling van de voormeting. In de nameting heeft men meer vrienden/kennissen dan in de voormeting.

9.2.4 Aannamen

De aannamen van de Wilcoxon-toets zijn de volgende:

1 De steekproef is aselect en onafhankelijk getrokken.
2 Er is sprake van gepaarde waarnemingen.
3 Het meetniveau is minimaal ordinaal.

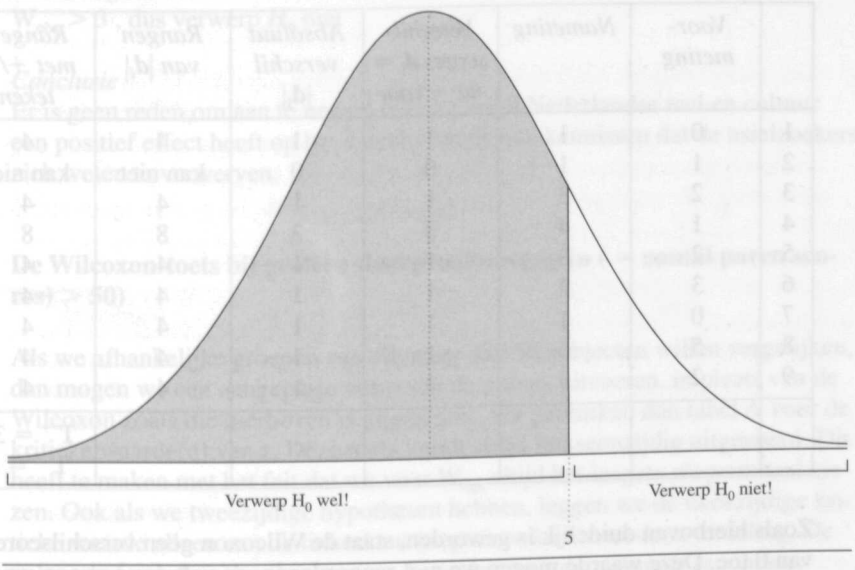
9.2.5 Beslissingsregels

Deze Wilcoxon-toets wordt rechtseenzijdig uitgevoerd, met een α van 2,5%. De kritieke waarden van deze toets kunnen we vinden in tabel D, we noemen deze waarde $W_{kritiek}$.
De beslissingsregels van deze toets, dat wil zeggen in welk gebied we de nulhypothese moeten verwerpen, wijken af van de normale gang van zaken.
Ten eerste geeft de tabel ons telkens maar één kritieke waarde, ook bij een tweezijdige toets. Dit komt doordat de berekening van deze toets uitgaat van absolute waarden. Dit houdt in dat het niet gaat om de negatieve of positieve waarden van de verschilscores, maar om de grootte daarvan. Zo is het getal -4 groter dan $+3$, doordat de absolute waarde van het eerste getal groter is, $4 > 3$. Dit heeft tot gevolg dat we alleen aan de positieve kant van de verdeling toetsen.
Ten tweede ligt het kritieke gebied bij deze toets op een andere plaats dan we gewend zijn: de staart die naar positief oneindig loopt, is het 'veilige gebied' (het gebied waar we de nulhypothese niet verwerpen) en alles wat daaronder ligt, vormt bij deze toets het kritieke gebied.

We zoeken nu de kritieke waarde op in tabel D:
$n = 9$
$\alpha = 0{,}025$, eenzijdig
$W_{kritiek} = 5$

In figuur 9.2 staat de situatie schematisch weergegeven.

Figuur 9.2



De beslissingsregels zien er als volgt uit:

1 Indien $W_{obs} > 5$, verwerp H_0 niet
2 Indien $W_{obs} \leq 5$, verwerp H_0

9.2.6 *Berekening*

Bij deze niet-parametrische toets voor de vergelijking van twee afhankelijke groepen zijn we weer aangewezen op rangordening van de gegevens. We berekenen eerst de verschillen tussen de voor- en de nameting, daarna kennen we rangen toe aan de absolute verschilwaarden. Ten slotte kennen we het teken (+ of −) weer toe aan de rangen. Om tot de geobserveerde W te komen, tellen we vervolgens alle positieve rangen op, dat vormt T^+, daarna tellen we alle negatieve rangen op, wat leidt tot T^-. We nemen dan die totaalwaarde, welke de laagste absolute waarde heeft, dat vormt onze W_{obs}. We voeren deze procedure hierna uit.

Tabel 9.4

	Voor-meting	*Nameting*	*Verschil-scores* d_i = *na − voor*	*Absoluut verschil* $\lvert d_i \rvert$	*Rangen van* $\lvert d_i \rvert$	*Rangen met +/− teken*
1	0	1	1	1	4	4
2	1	1	0	0	kan niet	kan niet
3	2	3	1	1	4	4
4	1	4	3	3	8	8
5	2	3	1	1	4	4
6	3	2	−1	1	4	−4
7	0	1	1	1	4	4
8	5	6	1	1	4	4
9	2	3	1	1	4	4
						$T^+ = 32$ $T^- = 4$

Zoals hierboven duidelijk is geworden, staat de Wilcoxon geen verschilscore van 0 toe. Deze waarde mogen we niet rangordenen, en valt dus gewoon weg. De rangordening begint bij de zeven enen die op de eerste tot en met zevende plaats komen; gemiddeld komen zij op een gedeelde vierde plaats. De drie ten slotte, komt op de achtste plaats.

Het wegvallen van asielzoeker 2 door de verschilscore van nul, heeft consequenties voor de grootte van n in dit experiment. We hebben in plaats van 9 mensen nog maar 8 mensen in ons onderzoek. We moeten daarom onze beslissingsregels herzien, deze zijn immers gebaseerd op n = 9.
In tabel D vinden we nu:
n = 8
α = 0,025, eenzijdig
$W_{kritiek} = 3$

De nieuwe beslissingsregels zien er dan als volgt uit:
1 Indien $W_{obs} > 3$, verwerp H_0 niet
2 Indien $W_{obs} \leq 3$, verwerp H_0

We moeten nu nog de rangordening controleren:

$$\frac{n \cdot (n+1)}{2} = \frac{8 \cdot (8+1)}{2} = \frac{72}{2} = 36$$

Het totaal van onze rangen moet dus gelijk zijn aan 36:
$T^+ + T^- = 32 + 4 = 36$

De uitkomsten van beide berekeningen komen exact overeen, dus is de rangordening op juiste wijze uitgevoerd.
Zoals al eerder vermeld, is onze geobserveerde waarde gelijk aan de laagste rangtotaalwaarde, in ons geval is dat $T^- = 4$, $W_{obs} = 4$.

9.2.7 *Beslissing*
$W_{obs} > 3$, dus verwerp H_0 niet

9.2.8 *Conclusie*
Er is geen reden om aan te nemen dat de cursus Nederlandse taal en cultuur een positief effect heeft op het aantal vrienden en kennissen dat de asielzoekers zich weten te verwerven.

9.3 De Wilcoxon-toets bij grotere steekproefomvang (n (= aantal paren scores) > 50)

Als we afhankelijke groepen van elk meer dan 50 subjecten willen vergelijken, dan mogen we een aangepaste vorm van de z-toets uitvoeren, in plaats van de Wilcoxon zoals die hierboven is uitgewerkt. We gebruiken dan tabel A voor de kritieke waarde(n) van z. Deze toets wordt altijd linkseenzijdig uitgevoerd. Dit heeft te maken met het feit dat we voor W_{obs} altijd het laagste rangentotaal kiezen. Ook als we tweezijdige hypothesen hebben, leggen we de tweezijdige kritieke waarde alleen aan de linkerkant (negatieve kant) van de z-verdeling. De volgende formule gebruiken we om de geobserveerde z-waarde te vinden:

$$z_{obs} = \frac{W_{obs} - n(n + 1) / 4}{\sqrt{\frac{n \cdot (n + 1)\,(2n + 1)}{24}}}$$

We moeten ons realiseren dat deze aangepaste z-toets niet-parametrisch blijft, we mogen alleen de normale verdeling gebruiken omdat er geldt dat een toename van n leidt tot een betere benadering van de normale verdeling.

Opgaven

9.1 Een orthopedagoge heeft een trainingsprogramma ontwikkeld om autistische kinderen onderling betere contacten te laten onderhouden. Alvorens met het programma te beginnen, meet zij bij 10 kinderen het aantal keren dat ze oogcontact hebben met een ander kind. Na afloop van het programma herhaalt zij de meting. De resultaten staan in onderstaande tabel.

Aantal oogcontacten

Kind	*Voormeting*	*Nameting*
a	0	2
b	1	2
c	1	4
d	2	3
e	4	3
f	3	3
g	4	6
h	2	4
i	1	4
j	4	4

a Kies een adequate toets om te kijken of het programma gewerkt heeft. Benader het probleem op twee manieren: parametrisch en niet-parametrisch.
b Voer de parametrische toets uit volgens het toetsingsschema.
c Voer de niet-parametrische toets uit volgens het toetsingsschema.

9.2 De onderzoeker die in opgave 8.2 al onderzoek heeft gedaan naar de verschillen tussen jongens en meisjes als het gaat om hun wiskundecijfers, is bang dat zijn onderzoeksresultaten misschien gemaskeerd worden door steekproeffluctuaties of derde variabelen. Om dit te ondervangen, heeft deze onderzoeker de volgende onderzoeksopzet bedacht. Hij vergelijkt 7 tweelingen, waarvan de één een jongen is en de andere een meisje, op hun rapportcijfer voor wiskunde. Alle geselecteerde tweelingen zitten op moment van het onderzoek in de tweede klas van het middelbaar onderwijs. Omdat de te vergelijken kinderen niet verschillen in leeftijd, sociaal-economische status en opvoedingswijze, denkt de onderzoeker meer kans te hebben om verschillen te vinden tussen de jongens en de meisjes.
Hieronder staan de resultaten.

Tweelingnummer	Wiskundecijfer jongen	Wiskundecijfer meisje
1	5	6
2	7	7
3	9	7
4	7	6
5	8	7
6	8	8
7	7	4

a Toets nu of de jongens uit een tweelingpaar een hoger rapportcijfer voor wiskunde hebben dan de meisjes uit hetzelfde tweelingpaar. Ga er eerst van uit dat de verdeling van rapportcijfers wiskunde normaal van vorm is en neem $\alpha = 0{,}05$. (deze opgave heeft een beknopte uitwerking)

b Toets nu of de jongens uit een tweelingpaar een hoger rapportcijfer voor wiskunde hebben dan de meisjes uit hetzelfde tweelingpaar. Ga er nu van uit dat de verdeling van rapportcijfers wiskunde niet normaal van vorm is en neem $\alpha = 0{,}05$. (deze opgave heeft een beknopte uitwerking)

c Stel dat de onderzoeker meerdere tweelingen gezocht heeft, in totaal 60 tweelingen. Hij vindt na hard zwoegen een $W_{obs} = 700$. Geef de beslissingsregels, berekening en conclusie, als we wederom de hypothese toetsen dat jongens beter in wiskunde zijn dan meisjes. Neem $\alpha = 0{,}05$.

Uitwerkingen

9.1 *a* **Onderzoeksvraag**
Heeft de behandeling succes, met andere woorden: zijn er in de nameting meer oogcontacten dan in de voormeting?

Gegevens:

Kind	*Voormeting*	*Nameting*
a	0	2
b	1	2
c	1	4
d	2	3
e	4	3
f	3	3
g	4	6
h	2	4
i	1	4
j	4	4

Toetskeuze
Ten eerste zien we dat we een verschil tussen groepen gaan bekijken (op basis van hun gemiddelden). De groepen vóór en na hebben in dit geval wel wat met elkaar te maken: ieder kind wordt twee keer gemeten. Er wordt zowel om de parametrische als de niet-parametrische uitvoering gevraagd; respectievelijk t-toets voor verwante steekproeven (gepaarde waarnemingen) en de Wilcoxon-toets.

b t-toets voor twee verwante steekproeven

Hypothesen
De alternatieve hypothese van deze toets kan op twee manieren geformuleerd worden:

1 Uiting van de verwachting dat de cijfers in de voortoets lager zijn dan in de natoets (linkseenzijdig)

2 Uiting van de verwachting dat de cijfers van de natoets hoger zijn dan die van de voortoets (rechtseenzijdig).

Hier is gekozen voor de eerste mogelijkheid, omdat de tweede mogelijkheid al is behandeld in het voorbeeld van dit hoofdstuk.

Hypothesen in termen van de onderzoeksvraag
H_0: De verdeling van de voortoets tendeert naar gelijke of hogere aantallen dan de verdeling van de natoets. De behandeling heeft een negatief of geen effect op het aantal oogcontacten.
H_1: De verdeling van de voortoets tendeert naar lagere aantallen dan de verdeling van de natoets. De behandeling heeft een positief effect op het aantal oogcontacten.
Hypothesen in statistische termen
H_0: $\mu_{voor} \geq \mu_{na}$
H_1: $\mu_{voor} < \mu_{na}$

Aannamen

1 De personen zijn aselect en onafhankelijk getrokken
2 Er zijn gepaarde waarnemingen (verwante steekproeven)
3 De populatieverdeling van de D_i (de verschilscores) is normaal van vorm.

Beslissingsregels

We toetsen hier een negatief effect, we hebben immers de verwachting geuit dat de cijfers van de voortoets kleiner zijn dan die van de natoets. Uit deze berekening (gemiddelde voormeting − gemiddelde nameting) komt dan een negatief getal volgens onze verwachting.

Er zijn n paren van scores:
$n = 10 \rightarrow df = n - 1 = 9$
$\alpha = 0{,}05$ eenzijdig
$t_{krit} = -1{,}833$

De beslissingsregels kunnen nu als volgt omschreven worden:
1 Indien $t_{obs} \leq -1{,}833$, verwerp H_0
2 Indien $t_{obs} > -1{,}833$, verwerp H_0 niet

Berekening

$$t_{obs} = \frac{\sum D_i}{\sqrt{\frac{n \sum D_i^2 - (\sum D_i)^2}{n - 1}}}$$

Allereerst moeten we de onbekende gegevens berekenen om de formule in te kunnen vullen.

Kind	*Voormeting*	*Nameting*	*Verschilscores* $D_i = score_{voor} - score_{na}$	D_i^2
a	0	2	−2	4
b	1	2	−1	1
c	1	4	−3	9
d	2	3	−1	1
e	4	3	1	1
f	3	3	0	0
g	4	6	−2	4
h	2	4	−2	4
i	1	4	−3	9
j	4	4	0	0
			$\sum D_i = -13$	$\sum D_i^2 = 33$

$$t_{obs} = \frac{-13}{\sqrt{\frac{10 \cdot 33 - (-13)^2}{9}}} = \frac{-13}{\sqrt{\frac{330 - 169}{9}}} = \frac{-13}{\sqrt{\frac{161}{9}}} = \frac{-13}{\sqrt{17{,}89}} = \frac{-13}{4{,}23}$$

$$= -3{,}07$$

Beslissing

$t_{obs} < t_{krit}$, want $-3{,}07 < -1{,}833$, verwerp H_0

Conclusie

Er is reden om aan te nemen dat de behandeling positief effect heeft op het aantal oogcontacten bij autistische kinderen.

c Wilcoxon-toets

Omdat de Wilcoxon altijd rechtszijdig getoetst wordt, verwoorden we in deze hypothesen de verwachting dat de cijfers van de natoets groter zullen zijn dan die van de voortoets.

Hypothesen

Hypothesen in termen van de onderzoeksvraag
H_0: De verdeling van de nameting tendeert naar gelijke of lagere aantallen dan de verdeling van de voormeting. Bij de nameting hebben de kinderen evenveel of minder oogcontacten dan bij de voormeting.
H_1: De verdeling van de nameting tendeert naar hogere aantallen dan de verdeling van de voormeting. Bij de nameting hebben de kinderen meer oogcontacten dan bij de voormeting.

Aannamen

1 De steekproef is aselect en onafhankelijk getrokken.
2 Er is sprake van gepaarde waarnemingen.
3 Het meetniveau is minimaal ordinaal.

Beslissingsregels

$n = 10$
$\alpha = 0{,}05$
tabel D éénzijdig
$W_{krit} = 10$
1 Indien $W_{obs} > 10$, verwerp H_0 niet
2 Indien $W_{obs} \leq 10$, verwerp H_0

Berekening

Bij deze toets is de geobserveerde waarde, W_{obs}, een som van rangnummers, en wel óf de som van de rangnummers van de negatieve verschillen (T^-) óf de som van de rangnummers van de positieve verschillen (T^+). Als $T^+ \geq T^-$ dan $W_{obs} = T^-$. Als $T^+ < T^-$ dan $W_{obs} = T^+$.

	Voor-meting	*Nameting*	*Verschil-scores* d_i = na − voor	*Absoluut verschil* $\lvert d_i \rvert$	*Rangen van* $\lvert d_i \rvert$	*Rangen met +/− teken*
a	0	2	2	2	5	5
b	1	2	1	1	2	2
c	1	4	3	3	$7\frac{1}{2}$	$7\frac{1}{2}$
d	2	3	1	1	2	2
e	4	3	−1	1	2	−2
f	3	3	0	0	kan niet	kan niet
g	4	6	2	2	5	5
h	2	4	2	2	5	5
i	1	4	3	3	$7\frac{1}{2}$	$7\frac{1}{2}$
j	4	4	0	0	kan niet	kan niet
						$T^+ = 34$
						$T^- = 2$

Som van rangnummers met negatieve verschillen: $T^- = 2$
Som van rangnummers met positieve verschillen: $T^+ = 34$

We zien dat 2 kinderen uitvallen doordat hun verschilscore nul is. Nul wordt niet gerangordend. Het aantal bruikbare verschillen wordt dus 8 in plaats van 10. Nu moeten we de beslissingsregels en de $W_{kritiek}$ aanpassen.

In tabel D vinden we nu:
$n = 8$
$\alpha = 0{,}05$, eenzijdig
$W_{kritiek} = 5$

De nieuwe beslissingsregels zien er dan als volgt uit:
1 Indien $W_{obs} > 5$, verwerp H_0 niet
2 Indien $W_{obs} \leq 5$, verwerp H_0

We moeten nu nog de rangordening controleren:

$$\frac{n \cdot (n+1)}{2} = \frac{8 \cdot (8+1)}{2} = \frac{72}{2} = 36$$

Het totaal van onze rangen moet dus gelijk zijn aan 36:
$T^+ + T^- = 2 + 34 = 36$

De uitkomsten van beide berekeningen komen exact overeen, dus is de rangordening op juiste wijze uitgevoerd.
Zoals al eerder vermeld, is onze geobserveerde waarde gelijk aan de laagste rangtotaalwaarde, in ons geval is dat $T^- = 2$, $W_{obs} = 2$.

Beslissing
$W_{obs} < 5$, dus verwerp H_0

Conclusie
Er is reden om aan te nemen dat de behandeling een positief effect heeft op het aantal oogcontacten bij autistische kinderen (zelfde resultaat als bij parametrische toets).

9.2 *a* t-toets voor twee verwante steekproeven

Hypothesen in statistische termen
H_0: $\mu_{jongens} \leq \mu_{meisjes}$
H_1: $\mu_{jongens} > \mu_{meisjes}$

Beslissingsregels
1 Indien $t_{obs} \geq 1{,}943$, verwerp H_0
2 Indien $t_{obs} < 1{,}943$, verwerp H_0 niet

Berekening

$$t_{obs} = \frac{6}{\sqrt{\frac{7 \cdot 16 - (6)^2}{6}}} = 1{,}69$$

Beslissing
$t_{obs} < t_{krit}$, want $1{,}69 < 1{,}943$, verwerp H_0 niet

Conclusie
Ook deze onderzoeksopzet geeft ons geen reden om aan te nemen dat jongens beter in wiskunde zouden zijn dan meisjes.

b Wilcoxon-toets

Hypothesen in termen van de onderzoeksvraag

H_0: De verdeling van de jongens tendeert naar gelijke of lagere cijfers dan de verdeling van de meisjes.

H_1: De verdeling van de jongens tendeert naar hogere cijfers dan de verdeling van de meisjes.

Beslissingsregels

1 Indien $W_{obs} > 3$, verwerp H_0 niet

2 Indien $W_{obs} \leq 3$, verwerp H_0

Berekening

Som van rangnummers met negatieve verschillen: $T^- = 2$

Som van rangnummers met positieve verschillen: $T_+ = 13$

$W_{obs} = 2$.

We zien dat 2 tweelingen uitvallen doordat hun verschilscore nul is. Nul wordt niet gerangordend. Het aantal bruikbare verschillen wordt dus 5 in plaats van 7. Nu moeten we de beslissingsregels en de $W_{kritiek}$ aanpassen. De nieuwe beslissingsregels zien er dan als volgt uit:

1 Indien $W_{obs} > 0$, verwerp H_0 niet

2 Indien $W_{obs} \leq 0$, verwerp H_0

Beslissing

$W_{obs} > W_{krit}$, want $2 > 0$, dus verwerp H_0 niet

Conclusie

Ook hier hebben we weer geen reden om aan te nemen dat jongens een beter rapportcijfer voor wiskunde halen dan meisjes.

c **Beslissingsregels**

1 Indien $z_{obs} \leq -1{,}65$, verwerp H_0

2 Indien $z_{obs} > -1{,}65$, verwerp H_0 niet

Berekening

$$z_{obs} = \frac{700 - 60(60 + 1) / 4}{\sqrt{\frac{60 \cdot (60 + 1)\,(2 \cdot 60 + 1)}{24}}} = -1{,}58$$

Beslissing

$z_{obs} > z_{krit}$, want $-1{,}58 > -1{,}65$, verwerp H_0 niet

Conclusie

Ook deze onderzoeksopzet geeft ons geen reden om aan te nemen dat jongens beter in wiskunde zouden zijn dan meisjes.

10 Toetsen voor samenhang: Pearson-, Spearman-, χ^2-toets en vergelijking van twee correlatiecoëfficiënten: z_r-toets

In hoofdstuk 5 hebben we gezien hoe de sterkte van het verband of samenhang tussen twee variabelen in een steekproef kan worden berekend. In dit hoofdstuk introduceren we een drietal toetsen die het ons mogelijk maken vast te stellen, of het verband sterk genoeg is om ook van een verband in de populatie te kunnen spreken (of het verband significant is). Deze drie toetsen zijn de Pearson-, de Spearman- en de χ^2-toets. De eerste, de toets van **Pearson**, passen we toe als de variabelen waartussen we een verband verwachten, op interval- of rationiveau gemeten zijn; dit is een parametrische toets. **Spearman** is de niet-parametrische variant en deze vereist minimaal een ordinaal meetniveau. Spearman moet worden verkozen boven Pearson, als de interval- of ratiogegevens niet-normaal verdeeld zijn. De **χ^2-toets** stelt ons in staat het verband tussen twee nominale variabelen vast te stellen. Dit is een niet-parametrische toets, waar geen parametrische variant van bestaat. De oorzaak daarvan is dat een parametrische toets minimaal een intervalmeetniveau vereist. We zullen elke toets weer aan de hand van concrete voorbeeldsituaties uitwerken.

10.1 De Pearson-toets voor het toetsen van verband

Een reisbureau doet verschillende onderzoeken naar de potentiële klanten. Op een gegeven moment heeft men reden te veronderstellen dat mensen die veel arbeidsuren per week maken, gemiddeld langer op vakantie gaan. Als dit zo is, dan kan het reisbureau zich bezighouden met reclame, gericht op deze groep. Om deze veronderstelling te toetsen, vraagt men aselect aan 15 reisbureaubezoekers hoeveel uur ze gemiddeld werken per week, en hoe lang ze van plan zijn om op vakantie te gaan. De resultaten van dit onderzoek staan in tabel 10.1 weergegeven.

We gaan toetsen of het verband tussen de variabelen gemiddeld aantal arbeidsuren per week en aantal geplande vakantiedagen per jaar significant is. We verwachten een positief verband, namelijk 'hoe meer arbeidsuren, hoe langer de vakantie'. We hanteren daarbij een α van 5%. We mogen ervan uitgaan dat de verdelingen van beide variabelen normaal van vorm zijn.

10.1.1 Toetskeuze

Bij het toetsen van verband hebben we de keuze uit drie toetsen: Pearson-, Spearman- en χ^2-toets. Het gaat hier om gegevens die op rationiveau gemeten zijn, de χ^2-toets valt dus al af. Verder is er vermeld dat er uitgegaan mag worden van een normale verdeling van de beide variabelen, we voeren een parametrische toets uit: dus de Pearson-toets. Verder verwachten we dat het verband positief zal zijn, we toetsen dus rechtseenzijdig.

Tabel 10.1

Klant	*Aantal arbeidsuren per week (x)*	*Aantal geplande vakantiedagen (y)*
1	20	14
2	42	21
3	38	20
4	36	18
5	32	15
6	30	14
7	22	10
8	40	17
9	38	21
10	15	20
11	35	28
12	40	21
13	38	25
14	36	22
15	60	30

10.1.2 Onderzoeksvraag

Bij de correlatietoets zijn drie typen onderzoeksvragen mogelijk die passen bij drie mogelijke verwachtingen:

1. Als we een positieve samenhang verwachten tussen de twee variabelen:
   Bestaat er een positief verband tussen het aantal arbeidsuren per week en het aantal dagen dat men op vakantie gaat?
   (Hoe hoger het aantal arbeidsuren, des te hoger het aantal vakantiedagen.)
2. Als we een negatief verband verwachten:
   Bestaat er een negatief verband tussen het aantal arbeidsuren per week en het aantal dagen dat men op vakantie gaat?
   (Hoe hoger het aantal arbeidsuren, des te lager het aantal dagen dat men op vakantie gaat.)
3. Als we een verband verwachten, positief of negatief:
   Bestaat er een verband tussen de variabelen aantal arbeidsuren per week en aantal dagen dat men op vakantie gaat?

We kiezen hier voor de eerste mogelijkheid, omdat er duidelijk gesteld wordt dat men verwacht dat mensen die meer arbeidsuren per week draaien, meer dagen op vakantie gaan.

10.1.3 Hypothesen

In de statistische hypothesen van een 'verbandtoets' gebruiken we de populatieparameter ρ, dit is het symbool voor correlatie in de populatie. Dit symbool heeft dezelfde kenmerken als de correlatiecoëfficiënt (r), het kan een waarde aannemen tussen de -1 (perfecte negatieve samenhang) en $+1$ (perfecte positieve samenhang). Als ρ een waarde van 0 aanneemt, dan bestaat er geen ver-

band tussen de onderzochte variabelen. Bij een verbandtoets zijn er drie paren van hypothesen mogelijk; elk paar hoort bij de hierboven genoemde onderzoeksvragen.

1 Als we een positieve samenhang verwachten tussen de twee variabelen, rechtseenzijdig:

Hypothesen in termen van de onderzoeksvraag
H_0: Er bestaat een negatieve samenhang of geen samenhang tussen de beide variabelen.
H_1: Er bestaat een positieve samenhang tussen de beide variabelen.
Hypothesen in statistische termen
H_0: $\rho \leq 0{,}00$
H_1: $\rho > 0{,}00$

2 Als we een negatief verband verwachten, linkseenzijdig:

Hypothesen in termen van de onderzoeksvraag
H_0: Er bestaat een positieve samenhang of geen samenhang tussen de beide variabelen.
H_1: Er bestaat een negatieve samenhang tussen de beide variabelen.
Hypothesen in statistische termen
H_0: $\rho \geq 0{,}00$
H_1: $\rho < 0{,}00$

3 Als we een verband verwachten, positief of negatief, tweezijdig:

Hypothesen in termen van de onderzoeksvraag
H_0: Er bestaat geen samenhang tussen de beide variabelen.
H_1: Er bestaat samenhang tussen de beide variabelen.
Hypothesen in statistische termen
H_0: $\rho = 0{,}00$
H_1: $\rho \neq 0{,}00$

In overeenstemming met de keuze van de onderzoeksvraag, kiezen we hier voor het eerste paar hypothesen, we verwachten immers een positieve samenhang.

10.1.4 *Aannamen*

Voor de Pearson-correlatietoets gelden de volgende aannamen:

1 De subjecten zijn aselect en onafhankelijk getrokken.

2 De populatieverdelingen van x en y zijn normaal van vorm.

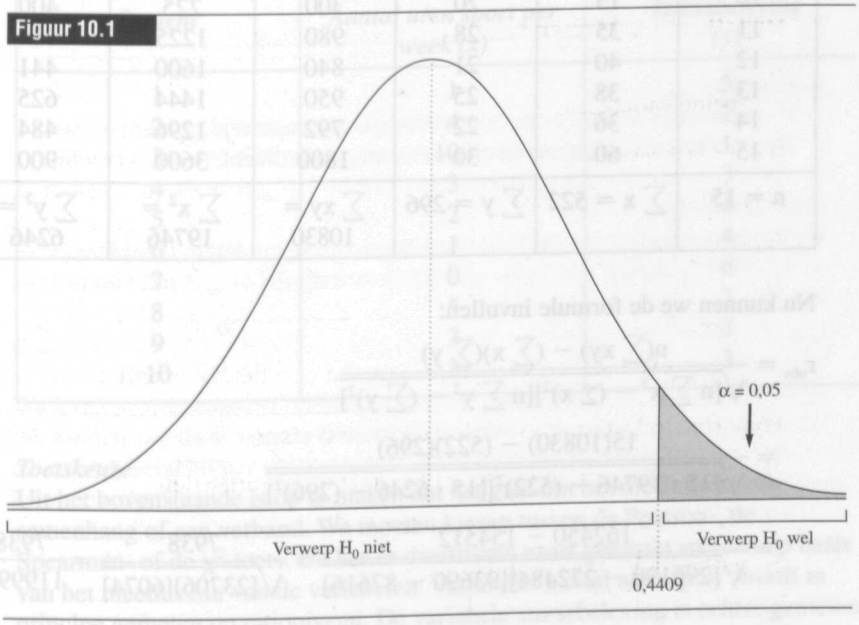
10.1.5 *Beslissingsregels*

Bij deze toets gaan we de correlatiecoëfficiënt die berekend wordt uit de steekproef, plaatsen in een verdeling van allerlei verbanden. In het midden van deze verdeling ligt de waarde 0, wat betekent dat er geen verband is tussen de variabelen. Rechts liggen de verbanden die een positief karakter hebben, links liggen de verbanden die negatief zijn. In tabel E vinden we de kritieke waarde, dat is die correlatiewaarde die de grens aangeeft tussen de nulhypothese wel en niet verwerpen. Het aantal vrijheidsgraden bij deze toets is gelijk aan $(n - 2)$, in ons geval betekent dat, dat we een aantal vrijheidsgraden van $(15 - 2) = 13$ hebben. We toetsen rechtseenzijdig, met $\alpha = 0{,}05 \rightarrow$ de kritieke correlatie-waarde is 0,4409. De toetsingsgrootheid bij deze toets is r.

$r_{krit} = 0{,}4409$

We zien een schematische weergave van de situatie in figuur 10.1.

Figuur 10.1



De beslissingsregels kunnen we als volgt formuleren:

1 Indien $r_{obs} \geq 0{,}4409$, verwerp $\boldsymbol{H_0}$

2 Indien $r_{obs} < 0{,}4409$, verwerp $\boldsymbol{H_0}$ niet

10.1.6 *Berekening*

Bij deze correlatietoets is de berekening gelijk aan de berekening van de correlatiecoëfficiënt uit hoofdstuk 5. Hieronder staat de formule:

$$r_{obs} = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}}$$

Allereerst moeten de onbekende gegevens berekend worden. Dat gebeurt in tabel 10.2.

Tabel 10.2

Klant	x	y	xy	x^2	y^2
1	20	14	280	400	196
2	42	21	882	1764	441
3	38	20	760	1444	400
4	36	18	648	1296	324
5	32	15	480	1024	225
6	30	14	420	900	196
7	22	10	220	484	100
8	40	17	680	1600	289
9	38	21	798	1444	441
10	15	20	300	225	400
11	35	28	980	1225	784
12	40	21	840	1600	441
13	38	25	950	1444	625
14	36	22	792	1296	484
15	60	30	1800	3600	900
n = 15	$\sum x = 522$	$\sum y = 296$	$\sum xy =$ 10830	$\sum x^2 =$ 19746	$\sum y^2 =$ 6246

Nu kunnen we de formule invullen:

$$r_{obs} = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}}$$

$$= \frac{15(10830) - (522)(296)}{\sqrt{[15 \cdot 19746 - (522)^2][15 \cdot 6246 - (296)^2]}}$$

$$= \frac{162450 - 154512}{\sqrt{[296190 - 272484][93690 - 87616]}} = \frac{7938}{\sqrt{[23706][6074]}} = \frac{7938}{11999{,}6}$$

$$= 0{,}66$$

10.1.7 Beslissing

$r_{obs} > r_{krit}$, want $0{,}66 > 0{,}4409$, dus verwerp $\boldsymbol{H_0}$

10.1.8 Conclusie

Er is reden om aan te nemen dat er een positief verband bestaat tussen het aantal arbeidsuren per week en het aantal dagen dat men op vakantie gaat. Het

reisbureau kan zijn reclamecampagne, op basis van deze resultaten, richten op mensen die veel arbeidsuren per week maken; zij boeken immers langere vakanties en dat betekent een grotere winst voor het reisbureau.

10.2 De Spearman-rangordecorrelatietoets

Een socioloog is geïnteresseerd in de samenhang tussen sport en stressbeleving in het dagelijks leven. Hij ontwikkelt zelf een meetschaal om stress te meten. Deze schaal bestaat uit 7 numerieke waarden die lopen van 1 tot en met 7, waarbij 7 een zeer hoge stressbeleving inhoudt en een waarde van 1 zeer lage stress aanduidt. De variabele sporten operationaliseert (operationaliseren is meetbaar maken) hij door de respondenten te vragen naar het aantal uren per week dat men sport. De socioloog wil weten of er verband bestaat tussen beide variabelen; hij legt daarom de desbetreffende vragen voor aan 10 aselect gekozen studenten. De resultaten staan in tabel 10.3. Wij helpen hem een handje, we nemen daarbij een α van 5%.

Tabel 10.3

Student	*Aantal uren sport per week (x)*	*Stressbeleving (y)*
1	1	5
2	4	2
3	10	1
4	3	3
5	2	4
6	1	4
7	0	6
8	5	2
9	3	3
10	2	3

10.2.1 Toetskeuze

Uit het bovenstaande is op te maken dat het gaat om het toetsen van een samenhang of een verband. We moeten kiezen tussen de Pearson-, de Spearman- of de χ^2-toets. De keuze daartussen moet gemaakt worden op basis van het meetniveau van de variabelen. Variabele 'aantal uren sport' wordt in principe gemeten op rationiveau. De variabele stressbeleving is echter gemeten op een ordinale meetschaal. De χ^2-toets valt dus af. Verder wordt vermeld dat de meetschaal door de onderzoeker zelf is ontwikkeld, dat wil zeggen dat we niet zeker zijn van de betrouwbaarheid van dat meetinstrument. We kiezen daarom voor de niet-parametrische variant: de Spearman-toets.
De toets zal tweezijdig uitgevoerd worden, omdat de socioloog geen specifieke richting aan het verwachte verband geeft, hij wil weten of er überhaupt samenhang bestaat.

10.2.2 Onderzoeksvraag

Bestaat er een samenhang tussen het aantal uren dat men per week sport en de dagelijkse stressbeleving?

10.2.3 Hypothesen

De hypothesen van deze toets mogen, ondanks het feit dit een niet-parametrische toets is, geformuleerd worden in statistische termen. We noemen de parameter in deze hypothesen ρ_s, waardoor deze onderscheiden kan worden van de ρ die we gebruiken bij de Pearson.

Hypothesen in termen van de onderzoeksvraag
H_0: Er bestaat geen samenhang tussen aantal uren sport per week en stressbeleving.
H_1: Er bestaat samenhang tussen aantal uren sport per week en stressbeleving.
Hypothesen in statistische termen
H_0: $\rho_s = 0{,}00$
H_1: $\rho_s \neq 0{,}00$

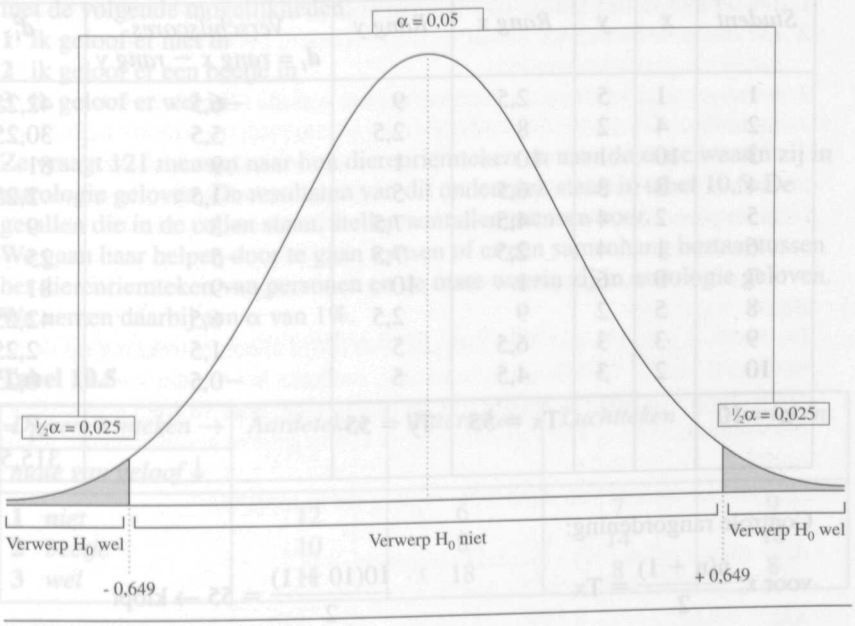
10.2.4 Aannamen

1 De steekproef is aselect en onafhankelijk samengesteld.
2 De meetschaal van de variabelen is minimaal ordinaal.

10.2.5 Beslissingsregels

De kritieke waarden van de Spearman-toets zijn te vinden in tabel F:
$\alpha = 0{,}05$, tweezijdig
$n = 10$
$r_{s\,krit} = +0{,}649$ en $-0{,}649$

Figuur 10.2



De beslissingsregels zien er als volgt uit:

1 Indien $r_{s\,obs} \geq +0{,}649$ of $r_{s\,obs} \leq -0{,}649$, verwerp $\boldsymbol{H_0}$

2 Indien $-0{,}649 < r_{s\,obs} < +0{,}649$, verwerp $\boldsymbol{H_0}$ niet

10.2.6 Berekening

De formule om $\boldsymbol{r_{s\,obs}}$ te berekenen, ziet er als volgt uit:

$$r_{s\,obs} = 1 - \left[\frac{6 \cdot \sum d_i^2}{n^3 - n}\right]$$

We kunnen alle waarden invullen, behalve $\sum d_i^2$. Deze term is nog onbekend. We komen aan deze waarde door de scores van x en y apart te rangordenen, daarna het verschil per rangenpaar te berekenen en ten slotte deze verschilscores te kwadrateren en op te tellen. In tabel 10.4 zien we deze berekeningen uitgevoerd.

Tabel 10.4

Student	*x*	*y*	*Rang x*	*Rang y*	*Verschilscores* d_i = *rang x* − *rang y*	d_i^2
1	1	5	2,5	9	−6,5	42,25
2	4	2	8	2,5	5,5	30,25
3	10	1	10	1	9	81
4	3	3	6,5	5	1,5	2,25
5	2	4	4,5	7,5	−3	9
6	1	4	2,5	7,5	−5	25
7	0	6	1	10	−9	81
8	5	2	9	2,5	6,5	42,25
9	3	3	6,5	5	1,5	2,25
10	2	3	4,5	5	−0,5	0,25
n = 10			Tx = 55	Ty = 55		d_i^2 = 315,50

Controle rangordening:

voor x: $\frac{n(n+1)}{2} = Tx$ $\qquad \frac{10(10+1)}{2} = 55 \rightarrow$ klopt

voor y: $\frac{n(n+1)}{2} = Ty$ $\qquad \frac{10(10+1)}{2} = 55 \rightarrow$ klopt

We kunnen nu de formule invullen:

$$r_{s\,obs} = 1 - \left[\frac{6 \cdot \sum d_i^2}{n^3 - n}\right] = 1 - \left[\frac{6 \cdot 315{,}5}{10^3 - 10}\right] = 1 - \left[\frac{1893}{990}\right] = 1 - 1{,}91 = -0{,}91$$

10.2.7 Beslissing

$r_{s\,obs} < -r_{s\,krit}$, want $-0{,}91 < -0{,}649$, dus verwerp $\boldsymbol{H_0}$

10.2.8 Conclusie

Er is reden om aan te nemen dat er een verband bestaat tussen sporten en stressbeleving.

10.3 De χ^2-toets

Een psychologe, tevens astrologe, doet onderzoek naar de samenhang tussen persoonskenmerken en het al dan niet geloven in astrologie. In het volgende experiment neemt zij het persoonskenmerk dierenriemteken. Dit kenmerk kan in vier categorieën ondergebracht worden, namelijk:
de aardetekens (stier, maagd en steenbok)
de watertekens (kreeft, schorpioen en vissen)
de luchttekens (tweelingen, weegschaal en waterman)
de vuurtekens (ram, leeuw en boogschutter)

De mate waarin mensen geloven in astrologie meet zij op een driepuntsschaal met de volgende mogelijkheden:
1 ik geloof er niet in
2 ik geloof er een beetje in
3 ik geloof er wel in

Ze vraagt 121 mensen naar hun dierenriemteken en naar de mate waarin zij in astrologie geloven. De resultaten van dit onderzoek staan in tabel 10.5. De getallen die in de cellen staan, stellen aantallen mensen voor.
We gaan haar helpen door te gaan toetsen of er een samenhang bestaat tussen het dierenriemteken van personen en de mate waarin zij in astrologie geloven. We nemen daarbij een α van 1%.

Tabel 10.5

Dierenriemteken → / *mate van geloof* ↓	*Aardeteken*	*Waterteken*	*Luchtteken*	*Vuurteken*
1 *niet*	12	6	7	9
2 *beetje*	10	8	14	10
3 *wel*	11	18	8	8

10.3.1 Toetskeuze

Allereerst moeten we een keuze maken tussen een verband- of verschiltoets; er wordt gevraagd een samenhang te toetsen, we kiezen dus één van de verbandtoetsen. Het meetniveau van de gebruikte variabelen bepaalt de toetskeuze. De variabele dierenriemteken is onderverdeeld in 4 categorieën, dit is een nominale variabele. De keuze valt dus op de χ^2-toets. Deze toets wordt altijd tweezijdig uitgevoerd, we kunnen bij deze toets dus geen richting aan de verwachte samenhang geven.

10.3.2 Onderzoeksvraag

Bestaat er een samenhang tussen de variabelen dierenriemteken en mate van geloof in astrologie?

10.3.3 Hypothesen

De hypothesen van deze niet-parametrische toets worden alleen verwoord in termen van de onderzoeksvraag, dus niet in statistische termen.

Hypothesen in termen van de onderzoeksvraag
H_0: Er bestaat geen samenhang tussen de aard van het dierenriemteken en de mate van geloof in astrologie.
H_1: Er bestaat samenhang tussen de aard van het dierenriemteken en de mate van geloof in astrologie.

10.3.4 Aannamen

1 De subjecten zijn aselect en onafhankelijk getrokken.
2 De categorieën moeten elkaar uitsluiten (iedere persoon kan maar in één categorie geplaatst worden).
3 Er mag bij de berekening geen verwachte waarde (E_{jk}) kleiner dan 5 voorkomen.

N.B. Als er sprake is van twee rijen (2 categorieën bij variabele 1) en twee kolommen (twee categorieën bij variabele 2) dan verandert deze aanname in: er mag bij de berekening geen verwachte waarde (E_{jk}) kleiner dan 10 voorkomen.

10.3.5 Beslissingsregels

De beslissingsregels van deze toets hebben *altijd* alleen betrekking op de positieve kant van de verdeling (rechterkant), ondanks het feit dat we altijd tweezijdig toetsen. Dit heeft te maken met de berekening; daar worden alle waarden gekwadrateerd en er kan dus onmogelijk een negatieve geobserveerde waarde ontstaan.

De kritieke waarde van deze toets vinden we in tabel G.

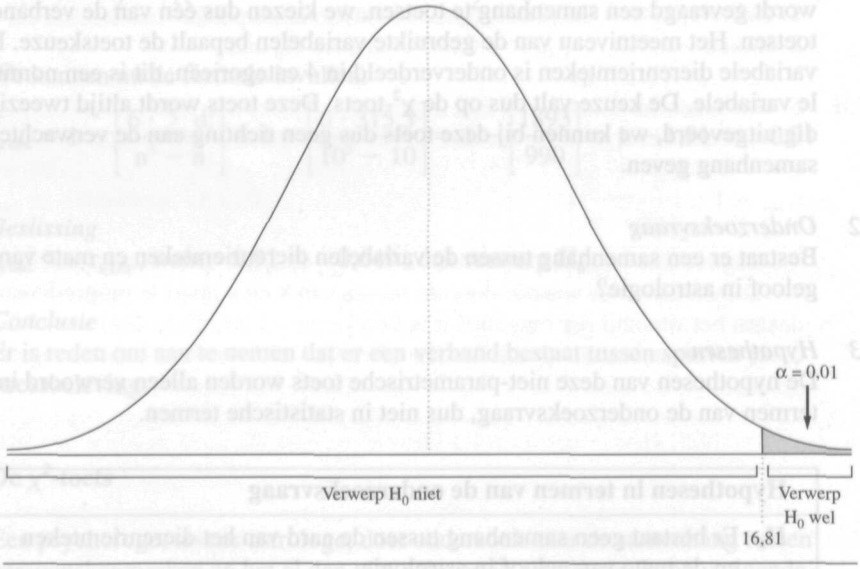
$\alpha = 0{,}01$

df = (aantal rijen -1) $\cdot$ (aantal kolommen -1)

$(3 - 1) \cdot (4 - 1) = 2 \cdot 3 = 6$

$\chi^2_{krit} = 16{,}81$

Figuur 10.3



De beslissingsregels luiden als volgt:

1 Indien $\chi^2_{obs} < 16{,}81$, verwerp H_0 niet
2 Indien $\chi^2_{obs} \geq 16{,}81$, verwerp H_0

10.3.6 Berekening

De berekeningsformule voor deze toets ziet er als volgt uit:

$$\chi^2_{obs} = \sum \frac{(O - E)^2}{E}$$

De ***O*** staat voor de geobserveerde aantallen mensen, de aantallen die in de cellen van tabel 10.5 vermeld staan. De ***E*** staat voor de (volgens de kansrekening) verwachte aantallen mensen in het geval dat er geen samenhang is tussen de twee variabelen. Als het verschil tussen de ***O***- en ***E***-waarden groot genoeg is, dan spreekt men van samenhang.
In tabel 10.6 zien we nogmaals de geobserveerde waarden, met daarbij de rij- en kolomtotalen.

Tabel 10.6

Dierenriemteken → / *mate van geloof ↓*	*Aarde-teken*	*Water-teken*	*Lucht-teken*	*Vuurteken*	*rijtotalen*
1 *niet*	12	6	7	9	34
2 *beetje*	10	8	14	10	42
3 *wel*	11	18	8	8	45
kolomtotalen	33	32	29	27	121

De ***E*** waarden berekenen we per cel, met de volgende formule:

$$\frac{\text{rij totaal} \cdot \text{kolom totaal}}{\text{algeheel totaal}}$$

In tabel 10.7 zijn de ***E***-waarden berekend.

Tabel 10.7

	Aardeteken	*Waterteken*	*Luchtteken*	*Vuurteken*
1	$E = \frac{34 \cdot 33}{121} = 9{,}3$	$E = \frac{34 \cdot 32}{121} = 9$	$E = \frac{34 \cdot 29}{121} = 8{,}1$	$E = \frac{34 \cdot 27}{121} = 7{,}6$
2	$E = \frac{42 \cdot 33}{121} = 11{,}5$	$E = \frac{42 \cdot 32}{121} = 11{,}1$	$E = \frac{42 \cdot 29}{121} = 10{,}1$	$E = \frac{42 \cdot 27}{121} = 9{,}4$
3	$E = \frac{45 \cdot 33}{121} = 12{,}3$	$E = \frac{45 \cdot 32}{121} = 11{,}9$	$E = \frac{45 \cdot 29}{121} = 10{,}8$	$E = \frac{45 \cdot 27}{121} = 10$

We zullen nu de uiteindelijke berekening uitvoeren per rij, per cel.

Rij 1

cel 1: $\frac{(O-E)^2}{E} = \frac{(12-9{,}3)^2}{9{,}3} = 0{,}78$

cel 2: $\frac{(O-E)^2}{E} = \frac{(6-9)^2}{9} = 1$

cel 3: $\frac{(O-E)^2}{E} = \frac{(7-8{,}1)^2}{8{,}1} = 0{,}15$

cel 4: $\frac{(O-E)^2}{E} = \frac{(9-7{,}6)^2}{7{,}6} = 0{,}26$

Rij 2

cel 1: $\frac{(O-E)^2}{E} = \frac{(10-11{,}5)^2}{11{,}5} = 0{,}2$

cel 2: $\frac{(O-E)^2}{E} = \frac{(8-11{,}1)^2}{11{,}1} = 0{,}87$

cel 3: $\frac{(O-E)^2}{E} = \frac{(14-10{,}1)^2}{10{,}1} = 1{,}5$

cel 4: $\frac{(O-E)^2}{E} = \frac{(10-9{,}4)^2}{9{,}4} = 0{,}04$

Rij 3

cel 1: $\frac{(O-E)^2}{E} = \frac{(11-12{,}3)^2}{12{,}3} = 0{,}14$

cel 2: $\frac{(O-E)^2}{E} = \frac{(18-11{,}9)^2}{11{,}9} = 3{,}13$

cel 3: $\frac{(O-E)^2}{E} = \frac{(8-10{,}8)^2}{10{,}8} = 0{,}73$

cel 4: $\frac{(O-E)^2}{E} = \frac{(8-10)^2}{10} = 0{,}4$

$\chi^2_{obs} = 0{,}78 + 1 + 0{,}15 + 0{,}26 + 0{,}2 + 0{,}87 + 1{,}5 + 0{,}04 + 0{,}14 + 3{,}13 + 0{,}73 + 0{,}4 = 9{,}2$

10.3.7 Beslissing

$\chi^2_{obs} < 16{,}81$, dus verwerp $\boldsymbol{H_0}$ niet

10.3.8 Conclusie

We hebben geen reden om aan te nemen dat het persoonskenmerk 'aard van dierenriemteken' samenhangt met het al dan niet geloven in astrologie.

10.3.9 χ^2-toets als verschiltoets

Hoewel we in dit boek de χ^2-toets als een verbandtoets hebben geïntroduceerd, zijn er andere boeken die deze toets vooral als een verschiltoets aanmerken. Eigenlijk berekenen we ook een verschil tussen de geobserveerde waarden en de verwachte ($\boldsymbol{E}$) waarden, die berekend zijn op basis van de randfrequenties (rij- en kolomtotalen). Zo zijn er twee interpretaties van deze toets mogelijk.

10.4 Vergelijken van twee correlatiecoëfficiënten: de z_r-toets

We kunnen naast het vaststellen van een verband of correlatie, ook nagaan of er verschil bestaat tussen twee correlatiecoëfficiënten. De toets die daarvoor gebruikt wordt is de z_r-toets. Dit is een parametrische toets. We observeren een correlatie die in twee onafhankelijke groepen tussen twee (dezelfde) variabelen is bepaald en gaan deze vervolgens vergelijken. We kunnen met behulp van deze toets kijken of het verband in de ene groep sterker is dan in de andere groep (rechtseenzijdig), of het verband in de ene groep zwakker is dan in de andere groep (linkseenzijdig), maar ook of het verband in de ene groep verschilt van dat in de andere groep (tweezijdig).
We zullen het gebruik van deze toets weer illustreren aan de hand van een voorbeeldsituatie.

In het kader van de vrouwenemancipatie wordt er door een tijdschrift onderzoek gedaan naar de hoogte van de positie van de werknemer en de hoogte van het salaris. Men vraagt zich af of het verband tussen deze twee variabelen voor mannen en vrouwen verschilt. Totaal worden er 25 werkende vrouwen en 30 werkende mannen ondervraagd op deze variabelen. Voor vrouwen blijkt dat de correlatie tussen de hoogte van de positie en het salaris gelijk is aan $r = 0{,}64$. Bij de mannen wordt er een correlatie van $r = 0{,}80$ gevonden. We gaan toetsen of er een verschil tussen deze twee correlaties bestaat. We nemen daarbij een α van 5%.

10.4.1 Toetskeuze

We moeten hier twee correlatiecoëfficiënten vergelijken, de correlatie tussen de variabelen positie en salaris bij mannen ($r = 0{,}80$) en de correlatie tussen dezelfde variabelen bij vrouwen ($r = 0{,}64$). De toets die voor dit probleem geschikt is, is de z_r-toets voor de vergelijking van twee correlaties in twee onafhankelijke groepen. We voeren deze toets tweezijdig uit, omdat we toetsen of er verschil is tussen de correlatie bij mannen en vrouwen; dit kan negatief of positief uitvallen.

10.4.2 Onderzoeksvraag

Bestaat er een verschil tussen mannen en vrouwen ten aanzien van de correlatie tussen positie en salaris? Is dat verschil significant, zodat we van een verschil in de populatie kunnen spreken?

10.4.3 Hypothesen

Hypothesen in termen van de onderzoeksvraag
H_0: De correlaties tussen positie en salaris verschillen niet bij vrouwen en mannen.
H_1: De correlaties tussen positie en salaris verschillen wel bij vrouwen en mannen.
Hypothesen in statistische termen
H_0: $\rho_{vrouwen} = \rho_{mannen}$
H_1: $\rho_{vrouwen} \neq \rho_{mannen}$

10.4.4 Aannamen

1 De subjecten in elke groep zijn aselect en onafhankelijk getrokken.
2 De groepen zijn onafhankelijk.
3 De populatieverdelingen van x en y zijn beide normaal van vorm.
4 De steekproeven van beide groepen hebben meer dan 20 subjecten.

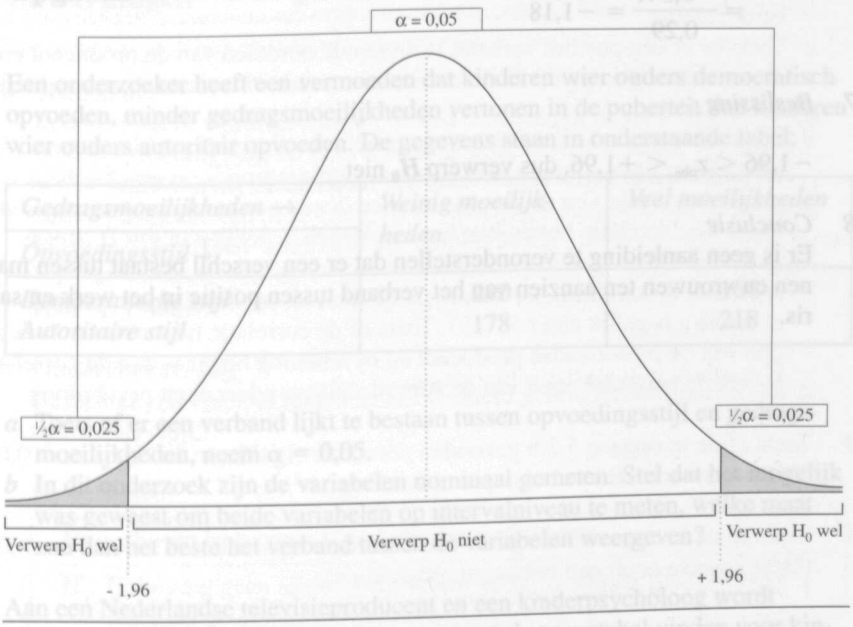
10.4.5 Beslissingsregels

Bij deze toets maken we gebruik van de z-verdeling en we zoeken de kritieke waarden dus op in tabel A.

$\alpha = 0{,}05$, tweezijdig

$z_{krit} = +1{,}96$ en $-1{,}96$

Figuur 10.4



De beslissingsregels voor deze toets zien er als volgt uit:
1 Indien $-1{,}96 < z_{obs} < +1{,}96$, verwerp $\boldsymbol{H_0}$ niet
2 Indien $z_{obs} \leq -1{,}96$ of $z_{obs} \geq +1{,}96$, verwerp $\boldsymbol{H_0}$

10.4.6 Berekening

$$z_{obs} = \frac{z_{r_1} - z_{r_2}}{\sqrt{\left[\frac{1}{n_1 - 3}\right] + \left[\frac{1}{n_2 - 3}\right]}}$$

In de formule hierboven zien we dat er twee onbekende waarden zijn die we moeten vinden om de formule in te kunnen vullen: z_{r1} en z_{r2}; dit zijn de z-waarden die horen bij de correlaties. Deze z-waarden vinden we in tabel H.
In deze tabel zien we allerlei correlaties staan, naast deze correlaties staat iedere keer een z-waarde die behoort bij deze correlatie. De correlatie bij de vrouwen was gelijk aan 0,64, de z-waarde die daarbij hoort is 0,758. Bij de mannen was de correlatie gelijk aan 0,80, de z-waarde die daarbij hoort is 1,099.

$z_{r_1} = 0{,}758$ en $z_{r_2} = 1{,}099$

We kunnen nu de formule invullen:

$$z_{obs} = \frac{0{,}758 - 1{,}099}{\sqrt{\left[\frac{1}{25 - 3}\right] + \left[\frac{1}{30 - 3}\right]}} = \frac{-0{,}341}{\sqrt{0{,}045 + 0{,}037}} = \frac{-0{,}341}{\sqrt{0{,}082}} =$$

$$= \frac{-0{,}341}{0{,}29} = -1{,}18$$

10.4.7 Beslissing

$-1{,}96 < z_{obs} < +1{,}96$, dus verwerp $\boldsymbol{H_0}$ niet

10.4.8 Conclusie

Er is geen aanleiding te veronderstellen dat er een verschil bestaat tussen mannen en vrouwen ten aanzien van het verband tussen positie in het werk en salaris.

Opgaven

10.1 Een onderzoeker heeft een vermoeden dat kinderen wier ouders democratisch opvoeden, minder gedragsmoeilijkheden vertonen in de puberteit dan kinderen wier ouders autoritair opvoeden. De gegevens staan in onderstaande tabel:

Gedragsmoeilijkheden → / *Opvoedingsstijl ↓*	*Weinig moeilijkheden*	*Veel moeilijkheden*
Democratische stijl	290	166
Autoritaire stijl	178	218

a Toets of er een verband lijkt te bestaan tussen opvoedingsstijl en gedragsmoeilijkheden, neem $\alpha = 0{,}05$.

b In dit onderzoek zijn de variabelen nominaal gemeten. Stel dat het mogelijk was geweest om beide variabelen op intervalniveau te meten, welke maat zou dan het beste het verband tussen de variabelen weergeven?

10.2 Aan een Nederlandse televisieproducent en een kinderpsycholoog wordt gevraagd in hoeverre zij 12 televisieprogramma's acceptabel vinden voor kinderen. Zij bekijken de programma's en geven onafhankelijk van elkaar een score van 1 (= zeer acceptabel) tot 10 (= niet acceptabel). De resultaten staan in onderstaande tabel.

Programma	*Beoordeling van producent*	*Beoordeling van psycholoog*
1 Tom & Jerry	3	1
2 Count Down	5	4
3 Knight Rider	9	5
4 Kinderkrant	6	8
5 Studio Sport	4	6
6 A-team	2	6
7 Sesamstraat	1	2
8 Dallas	7	9
9 Dynasty	7	9
10 Fame	8	6
11 Kerkepad	3	1
12 Telekids	5	2

a Wat is een goede maat om het verband tussen de oordelen van de producent en die van de psycholoog te bepalen? Welke parametrische en niet-parametrische toetsen passen bij dit probleem?
Neem bij de beantwoording van de volgende vragen steeds een significantieniveau van 5%.

b Bereken en toets het verband als we ervan uitgaan dat de oordelen normaal verdeeld zijn.

c Toets of er een positief verband is tussen de oordelen van de producent en de psycholoog en ga ervan uit dat we niet weten of wel aan de normaliteitseis is voldaan.

d Uit de Verenigde Staten komen een producent en een psycholoog over. Zij beoordelen 24 in Nederland uitgezonden programma's op geschiktheid voor jeugdige kijkers. Om dit goed te kunnen doen, worden de programma's vertaald. De correlatie tussen hun beoordelingen is gelijk aan $r = 0{,}10$. De Nederlandse producent en de psycholoog beoordeelden deze 24 programma's ook op kwaliteit. De correlatie tussen hun beoordelingen van de programma's is gelijk aan $r = 0{,}57$. Toets of de correlatie tussen de beoordelingen van de Nederlandse producent en psycholoog hoger is dan de correlatie tussen de beoordelingen van de Amerikaanse producent en psycholoog.

10.3 Toets of de in opgave 5.2.b gevonden correlatie significant is. Neem $\alpha = 0{,}01$. (Deze opgave heeft een beknopte uitwerking.)

10.4 Toets of de in opgave 5.4.c gevonden correlatie significant is. Neem $\alpha = 0{,}05$. (Deze opgave heeft een beknopte uitwerking.)

Uitwerkingen

10.1 *a* **Toetskeuze**

In deze opgave wordt gevraagd om vast te stellen of er samenhang bestaat tussen opvoedingsstijl en hoeveelheid gedragsmoeilijkheden. We moeten aan de hand van het meetniveau van deze variabelen bepalen welke toets we zullen gebruiken om deze vraag op te lossen. De variabele opvoedingsstijl heeft de categorieën 'autoritair' en 'democratisch', dit is een nominale variabele, de variabele gedragsmoeilijkheden is ordinaal gemeten. We kiezen voor de χ^2-toets.

Onderzoeksvraag

Bestaat er een verband tussen de hoeveelheid gedragsmoeilijkheden en de opvoedingsstijl?

Hypothesen

Hypothesen in termen van de onderzoeksvraag
H_0: Er bestaat geen samenhang tussen de opvoedingsstijl en de hoeveelheid gedragsmoeilijkheden.
H_1: Er bestaat samenhang tussen de opvoedingsstijl en de hoeveelheid gedragsmoeilijkheden.

Aannamen

1 De subjecten zijn aselect en onafhankelijk getrokken.
2 De categorieën moeten elkaar uitsluiten.
3 Er mag bij de berekening geen verwachte waarde (E_{jk}) kleiner dan 10 voorkomen.

Beslissingsregels

De kritieke waarde van deze toets vinden we in tabel G:

$\alpha = 0{,}05$

df = (aantal rijen $-$ 1) $\cdot$ (aantal kolommen $-$ 1)

$(2 - 1) \cdot (2 - 1) = 1 \cdot 1 = 1$

$\chi^2_{krit} = 3{,}84$

De beslissingsregels luiden als volgt:

1 Indien $\chi^2_{obs} < 3{,}84$, verwerp H_0 niet

2 Indien $\chi^2_{obs} \geq 3{,}84$, verwerp H_0

Berekening

De berekeningsformule voor deze toets ziet er als volgt uit:

$$\chi^2_{obs} = \sum \frac{(O - E)^2}{E}$$

Geobserveerde waarden (O)

Gedragsmoeilijkheden → / *Opvoedingsstijl ↓*	*Weinig moeilijkheden*	*Veel moeilijkheden*	*Rijtotalen*
Democratische stijl	290	166	456
Autoritaire stijl	178	218	396
Kolomtotalen	468	384	852

De ***E***-waarden berekenen we per cel, met de volgende formule:

$$\frac{\text{rij totaal} \cdot \text{kolom totaal}}{\text{algeheel totaal}}$$

Verwachte waarden (E)

Gedragsmoeilijkheden → / *Opvoedingsstijl ↓*	*Weinig moeilijkheden*	*Veel moeilijkheden*
Democratische stijl	$E = \frac{468 \cdot 456}{852} = 250{,}5$	$E = \frac{456 \cdot 384}{852} = 205{,}5$
Autoritaire stijl	$E = \frac{396 \cdot 468}{852} = 217{,}5$	$E = \frac{396 \cdot 384}{852} = 178{,}5$

We zullen nu de uiteindelijke berekening uitvoeren per rij, per cel:

Rij 1

cel 1: $\frac{(O - E)^2}{E} = \frac{(290 - 250{,}5)^2}{250{,}5} = 6{,}23$

cel 2: $\frac{(O - E)^2}{E} = \frac{(166 - 205{,}5)^2}{205{,}5} = 7{,}59$

Rij 2

cel 1: $\frac{(O - E)^2}{E} = \frac{(178 - 217{,}5)^2}{217{,}5} = 7{,}17$

cel 2: $\frac{(O - E)^2}{E} = \frac{(218 - 178{,}5)^2}{178{,}5} = 8{,}74$

$\chi^2_{obs} = 6{,}23 + 7{,}59 + 7{,}17 + 8{,}74 = 29{,}73$

Beslissing

$\chi^2_{obs} > 3{,}84$, dus verwerp H_0

Conclusie
We hebben reden om aan te nemen dat er een samenhang bestaat tussen opvoedingsstijl en hoeveelheid gedragsmoeilijkheden.

b Veronderstel dat de populatieverdelingen van opvoedingsstijl en gedragsmoeilijkheden normaal van vorm zijn en dat beide variabelen op intervalniveau gemeten zijn. Pearsons r is dan de beste maat voor het verband tussen de variabelen.

10.2 *a* Een maat voor overeenstemming, verband of samenhang is correlatie. Het gaat in dit geval niet om het verschil in de hoogte van de waardering tussen de psycholoog en de producent van de programma's, maar of zij overeenstemmen in de relatieve waardering van de programma's. Met andere woorden: of zij het in grote lijnen met elkaar eens zijn over wat de goede en wat de slechte programma's zijn.

Toetskeuze
We gaan kijken naar de mate van verband of samenhang. Hiervoor gebruiken we correlatietoetsen. Het meetniveau is hoger dan nominaal. In de opdracht wordt gevraagd om uit te gaan van 2 situaties:
1 normale verdeling ⇒ parametrisch ⇒ Pearson-correlatiecoëfficiënt
2 niet-normale verdeling ⇒ niet-parametrisch ⇒ Spearman-correlatiecoëfficiënt

b Pearson-correlatietoets

Onderzoeksvraag
Bestaat er een verband tussen de oordelen van de producent en de psycholoog?

Hypothesen

Hypothesen in termen van de onderzoeksvraag
H_0: Er bestaat geen samenhang tussen de oordelen van de producent en de psycholoog.
H_1: Er bestaat samenhang tussen de oordelen van de producent en de psycholoog.
Hypothesen in statistische termen
H_0: $\rho = 0{,}00$
H_1: $\rho \neq 0{,}00$

Aannamen
1 De subjecten zijn aselect en onafhankelijk getrokken.
2 De populatieverdelingen van x en y zijn normaal van vorm.

Beslissingsregels

We toetsen tweezijdig, met $\alpha = 0{,}05$

$df = n - 2 = 12 - 2 = 10$, kijk in tabel E

$r_{krit} = +0{,}5760$ en $-0{,}5760$

De beslissingsregels kunnen we als volgt formuleren:

1 Indien $r_{obs} \geq +0{,}5760$ of indien $r_{obs} \leq -0{,}5760$, verwerp $\boldsymbol{H_0}$

2 Indien $-0{,}5760 < r_{obs} < +0{,}5760$, verwerp $\boldsymbol{H_0}$ niet

Berekening

$$r_{obs} = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}}$$

Allereerst moeten de onbekende gegevens berekend worden.

We noemen de scores van de producent x en de scores van de psycholoog y.

Programma	x	y	xy	x^2	y^2
1	3	1	3	9	1
2	5	4	20	25	16
3	9	5	45	81	25
4	6	8	48	36	64
5	4	6	24	16	36
6	2	6	12	4	36
7	1	2	2	1	4
8	7	9	63	49	81
9	7	9	63	49	81
10	8	6	48	64	36
11	3	1	3	9	1
12	5	2	10	25	4
	$\sum x = 60$	$\sum y = 59$	$\sum xy = 341$	$\sum x^2 = 368$	$\sum y^2 = 385$

Formule invullen:

$$r_{obs} = \frac{(12 \cdot 341) - (60 \cdot 59)}{\sqrt{(12 \cdot 368 - 3600)(12 \cdot 385 - 3481)}}$$

$$= \frac{4092 - 3540}{\sqrt{(4416 - 3600)(4620 - 3481)}} = \frac{552}{\sqrt{816 \cdot 1139}} = \frac{552}{964{,}1} = 0{,}57$$

Beslissing

$r_{obs} < 0{,}5760$, dus verwerp $\boldsymbol{H_0}$ niet

Conclusie

Er is geen reden om aan te nemen dat er een overeenstemming bestaat tussen de oordelen van de producent en die van de psycholoog.

c Spearman-rangordecorrelatietoets

Onderzoeksvraag
Bestaat er een positieve samenhang tussen de beoordelingen van de producent en de psycholoog over de televisieprogramma's?

Hypothesen

Hypothesen in termen van de onderzoeksvraag:
H_0: Er bestaat geen samenhang of een negatieve samenhang tussen de beoordelingen van de producent en de psycholoog.
H_1: Er bestaat een positieve samenhang tussen de beoordelingen van de producent en de psycholoog.
Hypothesen in statistische termen:
H_0: $\rho_s \leq 0{,}00$
H_1: $\rho_s > 0{,}00$

Aannamen
1 De steekproef is aselect en onafhankelijk samengesteld.
2 De meetschaal van de variabelen is minimaal ordinaal.

Beslissingsregels
De kritieke waarden van de Spearman-toets zijn te vinden in tabel F:
$\alpha = 0{,}05$, rechtseenzijdig
$n = 12$
$r_{s\,krit} = +0{,}504$

De beslissingsregels zien er als volgt uit:
1 Indien $r_{s\,obs} \geq +0{,}504$, verwerp H_0
2 Indien $r_{s\,obs} < 0{,}504$, verwerp H_0 niet

Berekening
De formule om $r_{s\,obs}$ te berekenen, ziet er als volgt uit:

$$r_{s\,obs} = 1 - \left[\frac{6 \cdot \sum d_i^2}{n^3 - n}\right]$$

We kunnen alle waarden invullen, behalve $\sum d_i^2$ (is nog onbekend).

	x	y	rang x	rang y	d_i	d_i^2
1	3	1	3,5	1,5	2	4
2	5	4	6,5	5	1,5	2,25
3	9	5	12	6	6	36
4	6	8	8	10	−2	4
5	4	6	5	8	−3	9
6	2	6	2	8	−6	36
7	1	2	1	3,5	−2,5	6,25
8	7	9	9,5	11,5	−2	4
9	7	9	9,5	11,5	−2	4
10	8	6	11	8	3	9
11	3	1	3,5	1,5	2	4
12	5	2	6,5	3,5	3	9
						$\sum d_i^2 = 127,5$

$$r_{s\,obs} = 1 - \left[\frac{6 \cdot 127,5}{1728 - 12}\right] = 1 - 0,45 = 0,55$$

Beslissing

$r_{s\,obs} > 0,504$, dus verwerp H_0

Conclusie

Er is reden om aan te nemen dat er een positief verband bestaat tussen de oordelen van de producent en de psycholoog over de televisieprogramma's.

d **Toetskeuze**

We worden hier geacht twee correlatiecoëfficiënten te vergelijken en wel die van de Nederlandse producent en psycholoog en die van de Amerikaanse producent en psycholoog. Voor de vergelijking van twee correlaties gebruiken we de z_r-toets. Deze zal rechtseenzijdig uitgevoerd worden, omdat we verwachten dat de 'Nederlandse' correlatie hoger zal zijn dan de 'Amerikaanse'.

Onderzoeksvraag

Is de correlatie tussen de oordelen van de Nederlandse producent en psycholoog hoger dan die van de Amerikaanse producent en psycholoog?

Hypothesen

Hypothesen in termen van de onderzoeksvraag:
H_0: De correlatie tussen de oordelen van de Nederlandse producent en psycholoog is gelijk of lager dan die van hun Amerikaanse collega's.
H_1: De correlatie tussen de oordelen van de Nederlandse producent en psycholoog is hoger dan die van hun Amerikaanse collega's.
Hypothesen in statistische termen:
H_0: $\rho_{\text{Nederland}} \leq \rho_{\text{Verenigde Staten}}$
H_1: $\rho_{\text{Nederland}} > \rho_{\text{Verenigde Staten}}$

Aannamen

1 De subjecten in elke groep zijn aselect en onafhankelijk getrokken.
2 De groepen zijn onafhankelijk.
3 De populatieverdelingen van x en y zijn beide normaal van vorm.
4 De steekproeven van beide groepen hebben meer dan 20 subjecten.

Beslissingsregels

Bij deze toets maken we gebruik van de z-verdeling en we zoeken de kritieke waarden dus op in tabel A.

$\alpha = 0{,}05$, eenzijdig

$z_{krit} = +1{,}65$

De beslissingsregels voor deze toets zien er als volgt uit:

1 Indien $z_{obs} < +1{,}65$, verwerp H_0 niet
2 Indien $z_{obs} \geq +1{,}65$, verwerp H_0

Berekening

$$z_{obs} = \frac{z_{r_1} - z_{r_2}}{\sqrt{\left[\frac{1}{n_1 - 3}\right] + \left[\frac{1}{n_2 - 3}\right]}}$$

We vinden de z-waarden in tabel H:

$z_{r_{\text{Nederland}}} = 0{,}648$

$z_{r_{\text{Verenigde Staten}}} = 0{,}100$

We kunnen nu de formule invullen:

$$z_{obs} = \frac{0{,}648 - 0{,}100}{\sqrt{\left[\frac{1}{24 - 3}\right] + \left[\frac{1}{24 - 3}\right]}} = \frac{0{,}548}{\sqrt{0{,}05 + 0{,}05}} = \frac{0{,}548}{\sqrt{0{,}1}} = \frac{0{,}548}{0{,}32} = 1{,}71$$

Beslissing
$z_{obs} > 1{,}65$, dus verwerp H_0

Conclusie
Het is aannemelijk dat de correlatie tussen de beoordelingen over de televisieprogramma's van de producent en de psycholoog uit Nederland hoger is dan de correlatie tussen de beoordelingen over televisieprogramma's van de producent en de psycholoog uit de Verenigde Staten.

10.3 We toetsen op significantie van een Pearson-correlatiewaarde uit opgave 5.2.b. Er is geen richting aangegeven in de vraag, dus we voeren de Pearson-toets tweezijdig uit.

Hypothesen in statistische termen
H_0: $\rho = 0{,}00$
H_1: $\rho \neq 0{,}00$

Beslissingsregels
1 Indien $r_{obs} \geq +0{,}8343$ of indien $r_{obs} \leq -0{,}8343$, verwerp H_0
2 Indien $-0{,}8343 < r_{obs} < +0{,}8343$, verwerp H_0 niet

Berekening
We hoeven de berekening niet meer uit te voeren, want uit het antwoord van 5.2.b kunnen we opmaken dat $r_{obs} = -0{,}80$.

Beslissing
$-0.8343 < r_{obs} < 0{,}8343$, dus verwerp H_0 niet

Conclusie
Er is geen reden om aan te nemen dat de correlatie tussen vraag 12 en vraag 37 significant is.

10.4 We toetsen op significantie van een Pearson-correlatiewaarde uit opgave 5.4.c. Er is geen richting aangegeven in de vraag dus we voeren de Pearson-toets tweezijdig uit.

Hypothesen in statistische termen
H_0: $\rho = 0{,}00$
H_1: $\rho \neq 0{,}00$

Beslissingsregels
1 Indien $r_{obs} \geq +0{,}5324$ of indien $r_{obs} \leq -0{,}5324$, verwerp H_0
2 Indien $-0{,}5324 < r_{obs} < +0{,}5324$, verwerp H_0 niet

Berekening
We hoeven de berekening niet meer uit te voeren, want uit het antwoord van 5.4.c kunnen we opmaken dat $r_{obs} = 0{,}45$.

Beslissing

$-0{,}5324 < r_{obs} < 0{,}5324$, dus verwerp H_0 niet

Conclusie

Er is geen reden om aan te nemen dat er een correlatie bestaat tussen analytisch denkvermogen en artistiek vermogen.

11 Toetsen voor vergelijking van meer dan twee onafhankelijke groepen: de eenwegvariantieanalyse en de toets van Kruskal-Wallis

In dit hoofdstuk behandelen we de parametrische en niet-parametrische toetsen om het verschil tussen meer dan twee onafhankelijke groepen te toetsen. De parametrische toets die we daarvoor gebruiken is de eenwegvariantieanalyse, ook wel bekend als de Oneway ANOVA. De niet-parametrische variant hiervan is de toets van Kruskal-Wallis. We vergelijken de groepen op hun gemiddelde scores op de afhankelijke variabele, net als bij de verschiltoetsen uit hoofdstuk 7 en 8. Uiteindelijk toetsen we of de gevonden groepen uit dezelfde populatie afkomstig zijn of dat ze uit een andere populatie komen. We zullen het gebruik van deze toetsen illustreren aan de hand van concrete voorbeeldsituaties.

11.1 De eenwegvariantieanalyse

In het kader van het beleid van de Nederlandse overheid, gericht op vermindering van de alcoholconsumptie, hebben vele bierbrouwers zich gewaagd aan de ontwikkeling van lekker, schuimend alcoholvrij bier met 'body'. Eén bierbrouwer ontwikkelde zelfs drie soorten: 'maf malt', 'wild taste' en 'hot hop'. Omtrent één van de drie soorten bier (het bier dat men het lekkerst vindt) wil hij een grootscheepse reclamecampagne opzetten. Om erachter te komen welke men het lekkerst vindt, laat hij de drie soorten alcoholvrij bier proeven aan drie groepen van elk zes mensen, die aselect zijn samengesteld. De 'voorproevers' kunnen elk op een 10-puntsschaal aangeven hoe lekker ze het bier vinden, een waarde van 10 houdt in 'heel lekker' en een waarde van 1 houdt in dat men het bier 'smerig' vindt. In tabel 11.1 staan de gegevens uit dit onderzoek.

Tabel 11.1

Maf Malt	*Wild Taste*	*Hot Hop*
5	4	7
3	7	8
5	5	9
4	6	5
4	7	5
3	7	8

We gaan toetsen of er een verschil in smaakbeoordeling is van de drie soorten alcoholvrij bier. We hanteren een significantieniveau van 5%, we gaan ervan uit dat de verdeling van smaakoordelen normaal van vorm is.

11.1.1 *Toetskeuze*

We kiezen hier voor de eenwegvariantieanalyse, de parametrische toets voor vergelijking van gemiddelden van meer dan twee groepen.
We vergelijken drie groepen van smaakoordelen, die van Maf Malt, Wild Taste en Hot Hop. Deze groepen hebben niets met elkaar te maken, ze zijn onafhankelijk. Daarnaast is er vermeld dat de verdeling van smaakoordelen normaal van vorm is, vandaar dat we de parametrische toets kiezen. (Natuurlijk is dit een fictief onderzoek; een werkelijk onderzoek zou gebruikmaken van grotere groepen, maar we nemen dit kleine aantal om de berekeningen overzichtelijk te houden.)

11.1.2 *Onderzoeksvraag*

Is er een verschil in smaakbeoordeling tussen de drie soorten alcoholvrij bier?
De onderzoeksvraag, en dus de gehele variantieanalyse, heeft een tweezijdig karakter; we toetsen alleen of er een verschil is tussen de groepen scores, niet of het één lekkerder wordt gevonden dan het andere, maar weer minder lekker dan de derde soort.

11.1.3 *Hypothesen*

In de statistische hypothesen gebruiken we de populatieparameter μ, omdat we de groepen vergelijken op gemiddelde. We vragen ons af of het verschil tussen de steekproefgemiddelden dermate veelbetekenend is, dat we ook van een verschil in de populatie kunnen spreken. Het aantal malen dat we deze parameter tegenkomen in de hypothesen, is afhankelijk van het aantal groepen dat we vergelijken. In deze hypothesen vergelijken we drie groepen en we verwachten dat de gemiddelden van de populaties waar deze groepen uit afkomstig zijn, van elkaar verschillen. Het is niet vereist dat alle groepen significant van elkaar verschillen: één afwijkend steekproefgemiddelde kan bij deze toets al zorgen voor een significant effect, als deze afwijking tenminste groot genoeg is. Wat we dus aannemelijk willen maken, is dat de groepen in ieder geval niet allemaal uit dezelfde populatie afkomstig zijn. In de nulhypothese beweren we dat dit wel zo is, dus
$\mu_1 = \mu_2 = \mu_3$, en in de alternatieve hypothese uiten we onze verwachting dat de situatie in ieder geval anders is dan in de nulhypothese staat.

Hypothesen in termen van de onderzoeksvraag
H_0: De drie soorten bier worden allemaal even lekker gevonden.
H_1: De drie soorten bier worden qua smaak anders beoordeeld.
Hypothesen in statistische termen
H_0: $\mu_{\text{maf malt}} = \mu_{\text{wild taste}} = \mu_{\text{hot hop}}$
H_1: niet H_0

11.1.4 *Aannamen*

1 De subjecten in elke groep zijn aselect en onafhankelijk getrokken.
2 De groepen zijn onafhankelijk.

3 De populatievarianties zijn homogeen.
4 De populatieverdeling van de scores is normaal van vorm.
5 De uitspraken gelden alleen voor de onderzochte groepen.

11.1.5 *Beslissingsregels*

Bij deze toets gebruiken we altijd maar één kritieke waarde, omdat de variantieanalyse het resultaat is van een ratio(breuk) van *gekwadrateerde* afwijkingen. Deze afwijkingen kunnen nooit een negatieve waarde bezitten, we toetsen dus alleen aan de rechterkant van de verdeling. Om de procedure voor het vinden van de kritieke waarde van de variantie-analyse te begrijpen, moeten we alvast vooruitlopen op de berekening. De formule om het geobserveerde verschil tussen meer dan twee onafhankelijke groepen te berekenen, ziet er als volgt uit:

$$F_{obs} = \frac{MS_{tussen}}{MS_{binnen}}$$

Uit deze formule blijkt dat F de toetsingsgrootheid van de variantieanalyse is.

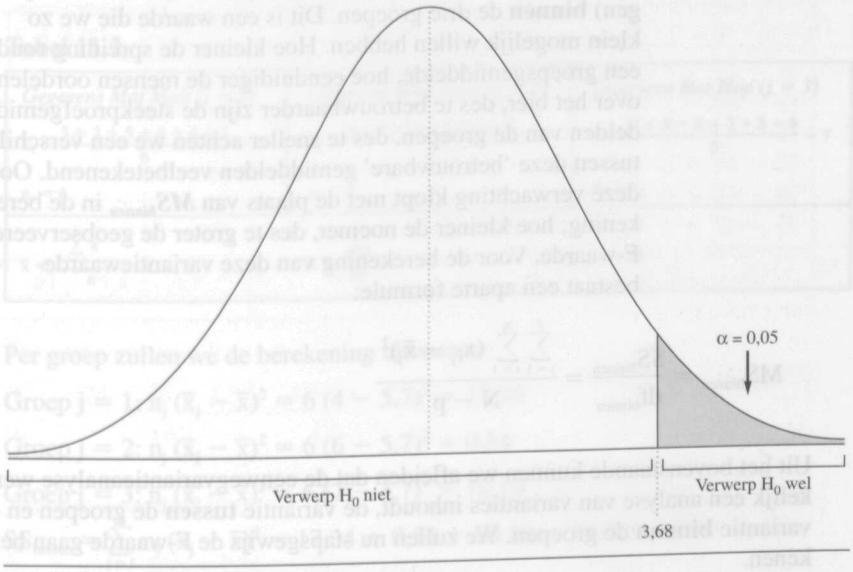
MS_{tussen} staat voor de gemiddelde gekwadrateerde afwijkingen **tussen** de drie groepen,

MS_{binnen} staat voor de gemiddelde gekwadrateerde afwijkingen **binnen** de drie groepen (MS = Mean Square).

Beide waarden worden berekend met een bepaald aantal vrijheidsgraden:
Voor de berekening van de teller van F_{obs}, MS_{tussen}: df = p − 1
Voor de berekening van de noemer van F_{obs}, MS_{binnen}: df = n − p
p = het aantal groepen dat in het experiment vergeleken wordt
n = het totaal aantal mensen/subjecten dat in het experiment gemeten is.

De bepaling van het aantal vrijheidsgraden van de teller en de noemer van de berekening, hebben we nodig om in tabel I de kritieke waarde van F te kunnen vinden.
In tabel I zien we op de horizontale bovenbalk de getallen staan die corresponderen met het aantal vrijheidsgraden van de teller, in de op de volgende bladzijden doorlopende verticale balk staan de getallen die overeenkomen met het aantal vrijheidsgraden van de noemer. Verder zien we overal twee waarden genoemd, de bovenste waarde is de kritieke F-waarde bij een α van 0,05; de onderste, dikgedrukte waarde is de kritieke F-waarde bij een α van 0,01. df van de teller = (p − 1) = (aantal groepen − 1) = (3 − 1) = 2
df van de noemer = (n − p) = (totaal aantal mensen − 3) = (18 − 3) = 15
Wij hanteren een α van 0,05, we nemen dus de bovenste van de twee kritieke waarden: $F_{krit} = 3{,}68$

Figuur 11.1



De beslissingsregels van deze toets worden nu als volgt geformuleerd:

1 Indien $F_{obs} < 3{,}68$, verwerp $\boldsymbol{H_0}$ niet

2 Indien $F_{obs} \geq 3{,}68$, verwerp $\boldsymbol{H_0}$

11.1.6 Berekening van de geobserveerde waarde met behulp van definitieformules

De definitieformule voor de berekening van de geobserveerde F-waarde ziet er als volgt uit:

$$F_{obs} = \frac{MS_{tussen}}{MS_{binnen}}$$

De grootte van F hangt af van twee gemiddelde gekwadrateerde afwijkingen (varianties):

1 $\boldsymbol{MS_{tussen}}$ Staat voor de variantie (gemiddelde gekwadrateerde afwijkingen) **tussen** de drie groepsgemiddelden. Deze variantie (spreiding) moet relatief groot zijn, we willen immers het verschil aantonen tussen de drie groepsgemiddelden. De groepsgemiddelden moeten onderling dus een grote spreiding vertonen. Deze verwachting klopt ook met de plaats van $\boldsymbol{MS_{tussen}}$ in de berekening, een groot getal in de teller leidt tot een grotere waarde van de geobserveerde F, des te sneller verwerpen we de nulhypothese. Voor de berekening van deze waarde bestaat een aparte formule:

$$MS_{tussen} = \frac{SS_{tussen}}{df_{tussen}} = \frac{\sum_{j=1}^{p} n_j (\bar{x}_j - \bar{x})^2}{p - 1}$$

2 MS_{binnen} Staat voor de variantie (gemiddelde gekwadrateerde afwijkingen) **binnen** de drie groepen. Dit is een waarde die we zo klein mogelijk willen hebben. Hoe kleiner de spreiding rond een groepsgemiddelde, hoe eenduidiger de mensen oordelen over het bier, des te betrouwbaarder zijn de steekproefgemiddelden van de groepen, des te sneller achten we een verschil tussen deze 'betrouwbare' gemiddelden veelbetekenend. Ook deze verwachting klopt met de plaats van MS_{binnen} in de berekening; hoe kleiner de noemer, des te groter de geobserveerde F-waarde. Voor de berekening van deze variantiewaarde bestaat een aparte formule:

$$MS_{binnen} = \frac{SS_{binnen}}{df_{binnen}} = \frac{\sum_{j=1}^{p} \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2}{N - p}$$

Uit het bovenstaande kunnen we afleiden dat de eenwegvariantieanalyse werkelijk een analyse van varianties inhoudt, de variantie **tussen** de groepen en de variantie **binnen** de groepen. We zullen nu stapsgewijs de F-waarde gaan berekenen.

Berekening van MS_{tussen}

$$MS_{tussen} = \frac{SS_{tussen}}{df_{tussen}} = \frac{\sum_{j=1}^{p} n_j (\bar{x}_j - \bar{x})^2}{p - 1}$$

Allereerst zullen we alle symbolen en letters die in deze formule voorkomen, verduidelijken.

MS = Mean of Squares, gemiddelde kwadratensom
SS = Sum of Squares, kwadratensom
df = degrees of freedom, aantal vrijheidsgraden waardoor men SS deelt om MS te krijgen
$\sum$ = 'tel op.....'
p = totaal aantal groepen in het experiment
j = de groepsaanduiding
$\sum_{j=1}^{p}$ = 'tel op (...), begin daarmee bij groep j = 1 en doe het voor alle groepen (p).
$\bar{x}_j$ = gemiddelde van groep j
$\bar{x}$ = algemeen gemiddelde, gemiddelde smaakbeoordeling in het hele experiment.
n_j = aantal mensen in groep j
$SS_{tussen} = \sum_{j=1}^{p} n_j (\bar{x}_j - \bar{x})^2$ betekent vrij vertaald: tel op, per groep: het aantal mensen in die groep maal de afwijking tussen het groepsgemiddelde en het algemeen gemiddelde, begin bij groep 1 en doe het voor alle groepen.

In tabel 11.2 zijn de gegevens voor elke groep apart berekend.

Tabel 11.2

Gegevens Maf Malt (j = 1)	*Gegevens Wild Taste (j = 2)*	*Gegevens Hot Hop (j = 3)*
$\bar{x}_1 = \frac{5+3+5+4+4+3}{6} = 4$	$\bar{x}_2 = \frac{4+7+5+6+7+7}{6} = 6$	$\bar{x}_3 = \frac{7+8+9+5+5+8}{6} = 7$
$n_1 = 6$	$n_2 = 6$	$n_3 = 6$
$\bar{x} = \frac{\sum_{i=1}^{n} x}{n} \rightarrow$	$\bar{x} = \frac{102}{18} \rightarrow$	$\bar{x} = 5{,}7$

Per groep zullen we de berekening uitvoeren:

Groep j = 1: $n_j (\bar{x}_j - \bar{x})^2 = 6 (4 - 5{,}7)^2 = 17{,}34$

Groep j = 2: $n_j (\bar{x}_j - \bar{x})^2 = 6 (6 - 5{,}7)^2 = 0{,}54$

Groep j = 3: $n_j (\bar{x}_j - \bar{x})^2 = 6 (7 - 5{,}7)^2 = 10{,}14$

$$SS_{tussen} = \sum_{j=1}^{p} n_j (\bar{x}_j - \bar{x})^2 = 17{,}34 + 0{,}54 + 10{,}14 = 28{,}02$$

$$df_{tussen} = p - 1 = 3 - 1 = 2$$

$$MS_{tussen} = \frac{SS_{tussen}}{df_{tussen}} = \frac{28{,}02}{2} = 14{,}01$$

Berekening van MS_{binnen}

$$MS_{binnen} = \frac{SS_{binnen}}{df_{binnen}} = \frac{\sum_{j=1}^{p} \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2}{N - p}$$

$$SS_{binnen} = \sum_{j=1}^{p} \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2$$

Vrij vertaald houdt de berekening van SS_{binnen} het volgende in: tel op: de score van elke persoon van alle groepen min het groepsgemiddelde, kwadrateer deze waarde.

Tabel 11.3

Maf Malt (j = 1) $\bar{x}_1 = 4$	Wild Taste (j = 2) $\bar{x}_2 = 6$	Hot Hop (j = 3) $\bar{x}_3 = 7$
$(5-4)^2 = 1$	$(4-6)^2 = 4$	$(7-7)^2 = 0$
$(3-4)^2 = 1$	$(7-6)^2 = 1$	$(8-7)^2 = 1$
$(5-4)^2 = 1$	$(5-6)^2 = 1$	$(9-7)^2 = 4$
$(4-4)^2 = 0$	$(6-6)^2 = 0$	$(5-7)^2 = 4$
$(4-4)^2 = 0$	$(7-6)^2 = 1$	$(5-7)^2 = 4$
$(3-4)^2 = 1$	$(7-6)^2 = 1$	$(8-7)^2 = 1$
$\sum (x_{ij} - \bar{x}_j)^2 = 4$	$\sum (x_{ij} - \bar{x}_j)^2 = 8$	$\sum (x_{ij} - \bar{x}_j)^2 = 14$

$$SS_{binnen} = 4 + 8 + 14 = 26$$

$$df_{binnen} = N - p = 18 - 3 = 15$$

$$MS_{binnen} = \frac{SS_{binnen}}{df_{binnen}} = \frac{26}{15} = 1{,}73$$

Berekening van F_{obs}

$$F_{obs} = \frac{MS_{tussen}}{MS_{binnen}} = \frac{14{,}01}{1{,}73} = 8{,}1$$

11.1.7 Berekening geobserveerde F-waarde met behulp van rekenformules

De berekening van de geobserveerde F-waarde met behulp van de definitieformule schept duidelijkheid over wat we nu eigenlijk berekenen; het is echter geen snelle, handige methode. Voor grotere groepen scores adviseren wij dan ook de rekenformules te gebruiken. We passen deze formules hier toe op de gegevens van de bierbrouwer.

Deze berekening vindt plaats door het berekenen van drie tussenberekeningen:

$$\text{I} \quad \frac{T^2}{n}$$

$$\text{II} \quad \sum_{j=1}^{p} \left[\sum_{i=1}^{n_j} x_{ij}^2 \right]$$

$$\text{III} \quad \sum_{j=1}^{p} \left[\frac{T_j^2}{n_j} \right]$$

T = totaal van alle scores

T_j = totaalscore van groep j

In tabel 11.4 worden de benodigde gegevens uitgerekend.

Tabel 11.4

x (j = 1)	x^2 (j = 1)	x (j = 2)	x^2 (j = 2)	x (j = 3)	x^2 (j = 3)
5	25	4	16	7	49
3	9	7	49	8	64
5	25	5	25	9	81
4	16	6	36	5	25
4	16	7	49	5	25
3	9	7	49	8	64
$T_1 = 24$	$\sum x_1^2 = 100$	$T_2 = 36$	$\sum x_2^2 = 224$	$T_3 = 42$	$\sum x_3^2 = 308$

Tussenberekening I:

$$\frac{T^2}{n} = \frac{(24 + 36 + 42)^2}{18} = \frac{10404}{18} = 578$$

Tussenberekening II:

$$\sum_{j=1}^{p}\left[\sum_{i=1}^{n_j} x_{ij}^2\right] = 100 + 224 + 308 = 632$$

Tussenberekening III:

$$\sum_{j=1}^{p}\left[\frac{T_j^2}{n_j}\right] = \frac{24^2}{6} + \frac{36^2}{6} + \frac{42^2}{6} = 96 + 216 + 294 = 606$$

Nu kunnen we alle waarden gaan berekenen:

$$MS_{tussen} = \frac{SS_{tussen}}{df_{tussen}} = \frac{III - I}{p - 1} = \frac{606 - 578}{3 - 1} = 14$$

$$MS_{binnen} = \frac{SS_{binnen}}{df_{binnen}} = \frac{II - III}{n - p} = \frac{632 - 606}{18 - 3} = \frac{26}{15} = 1{,}73$$

$$F_{obs} = \frac{MS_{tussen}}{MS_{binnen}} = \frac{14}{1{,}73} = 8{,}1$$

De afwijkingen tussen de berekening met de definitieformule en die met de rekenformules, zijn te wijten aan afrondingsfouten. De rekenformule is *altijd* nauwkeuriger!

11.1.8 Beslissing

$F_{obs} > 3{,}68$, dus verwerp H_0

11.1.9 Conclusie

Er is reden om aan te nemen dat de drie soorten alcoholvrij bier verschillend zijn beoordeeld op smaak. Als we naar de gemiddelde smaakbeoordelingen kijken, dan wordt 'Hot Hop' het lekkerst gevonden door de voorproevers.

11.2 De toets van Kruskal-Wallis

In tabel 11.5 staan, voor vier leeftijdsgroepen, zogenaamde 'egocentrisme'scores. Deze scores zijn doorgaans nogal scheef verdeeld. Toets de hypothese dat de vier groepen uit dezelfde populatie getrokken zijn. Hanteer een significantieniveau van 5%

Tabel 11.5

Leeftijd	*Scores op egocentrisme*					
I: 21-30	12	8	9	10	10	11
II: 31-40	14	11	7	9	11	12
III: 41-50	9	13	11	12	14	16
IV: 51-60	12	15	14	11	10	13

11.2.1 Toetskeuze

We moeten hier vier groepen scores vergelijken, we gebruiken daarvoor de eenwegvariantieanalyse of de Kruskal-Wallis-toets. Het verschil tussen deze twee toetsen ligt in het feit dat de eerste een parametrische en de tweede een niet-parametrische toets is. Er wordt vermeld dat de scores op egocentrisme doorgaans scheef verdeeld zijn; we nemen daarom de niet-parametrische Kruskal-Wallis-toets.

Deze toets wordt, net als de eenwegvariantieanalyse, altijd tweezijdig uitgevoerd.

11.2.2 Onderzoeksvraag

Bestaat er verschil tussen de vier leeftijdsgroepen op de variabele egocentrisme?

11.2.3 Hypothesen

De hypothesen van deze niet-parametrische toets worden alleen opgesteld in termen van de onderzoeksvraag.

Hypothesen in termen van de onderzoeksvraag
H_0: De populaties waartoe de groepen behoren zijn identiek.
H_1: De populaties waartoe de groepen behoren zijn verschillend.

11.2.4 Aannamen

1 De subjecten zijn aselect en onafhankelijk verdeeld over k onafhankelijke groepen met ieder een $n_j \geq 5$.
2 De afhankelijke variabele is continu en de meetschaal is minimaal ordinaal.

11.2.5 Beslissingsregels

Hoewel deze toets altijd een tweezijdig karakter heeft, worden de beslissingsregels *altijd* alleen over de rechterkant van de verdeling geformuleerd. Dit

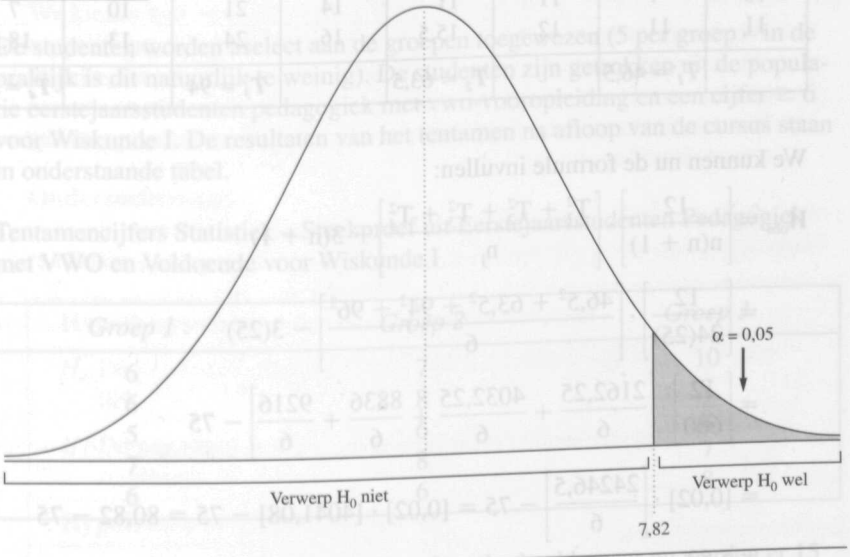
heeft te maken met de berekening, waar gekwadrateerd wordt, zodat de geobserveerde waarde nooit een negatief karakter kan hebben. De toetsingsgrootheid van deze toets noemen we H, de kritieke waarde vinden we in de tabel van X^2, tabel G dus.

$\alpha = 0{,}05$

df = (k − 1) = (aantal groepen − 1) = (4 − 1) = 3

$H_{krit} = 7{,}82$

Figuur 11.2



De beslissingsregels luiden als volgt:

1 Indien $H_{obs} \geq 7{,}82$, verwerp $\boldsymbol{H_0}$

2 Indien $H_{obs} < 7{,}82$, verwerp $\boldsymbol{H_0}$ niet

11.2.6 Berekening

De formule die we hanteren bij deze toets, ziet er als volgt uit:

$$H_{obs} = \left[\frac{12}{n(n+1)}\right] \cdot \left[\frac{\sum T_j^2}{n_j}\right] - 3(n+1)$$

De T-waarden staan bij deze toets voor het totaal van de rangen per groep. We moeten scores dus een rang geven, dat doen we door *alle* scores van klein naar groot te ordenen, en op basis van hun grootte een rang te geven:

Tabel 11.6

Groep I		Groep II		Groep III		Groep IV	
Cijfer	Rang	Cijfer	Rang	Cijfer	Rang	Cijfer	Rang
12	15,5	14	21	9	4	12	15,5
8	2	11	11	13	18,5	15	23
9	4	7	1	11	11	14	21
10	7	9	4	12	15,5	11	11
10	7	11	11	14	21	10	7
11	11	12	15,5	16	24	13	18,5
	$T_1 = 46{,}5$		$T_2 = 63{,}5$		$T_3 = 94$		$T_4 = 96$

We kunnen nu de formule invullen:

$$H_{obs} = \left[\frac{12}{n(n+1)}\right] \cdot \left[\frac{T_1^2 + T_2^2 + T_3^2 + T_4^2}{n_j}\right] - 3(n+1)$$

$$= \left[\frac{12}{24(25)}\right] \cdot \left[\frac{46{,}5^2 + 63{,}5^2 + 94^2 + 96^2}{6}\right] - 3(25)$$

$$= \left[\frac{12}{600}\right] \cdot \left[\frac{2162{,}25}{6} + \frac{4032{,}25}{6} + \frac{8836}{6} + \frac{9216}{6}\right] - 75$$

$$= [0{,}02] \cdot \left[\frac{24246{,}5}{6}\right] - 75 = [0{,}02] \cdot [4041{,}08] - 75 = 80{,}82 - 75$$

$$= 5{,}82$$

11.2.7 Beslissing

$H_{obs} < H_{krit}$, dus verwerp $\boldsymbol{H_0}$ niet

11.2.8 Conclusie

Het is niet aannemelijk dat de mensen in de verschillende leeftijdscategorieën verschillen qua egocentrisme.

Opgaven

11.1 Vijftien studenten pedagogiek doen mee aan een onderwijsexperiment. Met het experiment wil de onderzoeker aantonen dat onderwijsdidactiek in de cursus statistiek verschillen in tentamencijfer 'veroorzaakt'. Er zijn drie groepen:
Groep 1 krijgt een traditionele cursus statistiek
Groep 2 krijgt een cursus met nadruk op de praktische toepassing van statistiek
Groep 3 krijgt een cursus met nadruk op theoretisch inzicht in de statistiek.

De studenten worden aselect aan de groepen toegewezen (5 per groep – in de praktijk is dit natuurlijk te weinig). De studenten zijn getrokken uit de populatie eerstejaarsstudenten pedagogiek met vwo-vooropleiding en een cijfer ≥ 6 voor Wiskunde I. De resultaten van het tentamen na afloop van de cursus staan in onderstaande tabel.

Tentamencijfers Statistiek – Steekproef uit Eerstejaarsstudenten Pedagogiek met VWO en Voldoende voor Wiskunde I

Groep 1	*Groep 2*	*Groep 3*
6	7	10
6	8	7
5	6	8
7	8	7
6	6	8

Hetzelfde experiment werd een half jaar later herhaald, maar nu werden er 15 studenten uit de totale populatie eerstejaars pedagogiekstudenten getrokken. De resultaten zijn te vinden in Tabel 2.

Tabel 2
Tentamencijfers Statistiek – Steekproef uit Alle Eerstejaarsstudenten Pedagogiek

Groep 1	*Groep 2*	*Groep 3*
4	9	9
9	8	8
3	5	8
6	7	7
8	6	8

We gaan ervan uit dat tentamencijfers normaal verdeeld zijn en op intervalniveau gemeten.

a Toets beide experimenten op verschil tussen de groepen ($\alpha = 0{,}05$).
b Verklaar in niet-statistische termen waarom er bij het eerste experiment een sterker effect zal optreden dan bij het tweede. Welk experiment past het beste in de onderzoeksvraag?

11.2 Een onderzoeker is na bestudering van een groot aantal artikelen over de verdelingen van tentamencijfers statistiek tot de conclusie gekomen dat deze cijfers vrijwel altijd scheef naar links verdeeld zijn, en bovendien met name onregelmatig zijn tussen de vijf en de zes.

a Toets nogmaals voor beide experimenten uit de vorige zitting de hypothese dat de drie groepen uit eenzelfde populatie zijn getrokken, rekening houdend met de bovenstaande informatie.
b Wat zou de oorzaak kunnen zijn voor de gerapporteerde onregelmatigheden tussen de vijf en de zes?

Uitwerkingen

11.1 *a* De twee experimenten lijken erg op elkaar, en daarnaast geldt dezelfde toetsingsvraag voor beide experimenten. Alleen de berekeningen, beslissingen en conclusies verschillen van elkaar, verder gelden alle stappen voor beide experimenten.

Toetskeuze
We kiezen hier voor de eenwegvariantieanalyse, de parametrische toets voor vergelijking van gemiddelden van meer dan twee groepen. We vergelijken drie groepen van tentamencijfers. Deze groepen hebben niets met elkaar te maken, ze zijn onafhankelijk. Er is vermeld dat de verdeling van tentamencijfers normaal van vorm is, vandaar dat we de parametrische toets kiezen.

Onderzoeksvraag
Is er een verschil tussen de drie groepen qua tentamencijfer? Heeft de andere onderwijsaanpak effect op de tentamencijfers van pedagogiekstudenten?

Hypothesen in termen van de onderzoeksvraag
H_0: De populaties waar de drie groepen uit afkomstig zijn, zijn identiek.
H_1: De populaties waar de drie groepen uit afkomstig zijn, zijn verschillend.
Hypothesen in statistische termen
H_0: $\mu_1 = \mu_2 = \mu_3$
H_1: niet H_0

Aannamen
1 De subjecten in elke groep zijn aselect en onafhankelijk getrokken.
2 De groepen zijn onafhankelijk.
3 De populatievarianties zijn homogeen.
4 De populatieverdeling van de scores is normaal van vorm.
5 De uitspraken gelden alleen voor de onderzochte groepen.

Beslissingsregels
We gebruiken tabel I om de kritieke waarde van F te vinden:
df van de teller = (p − 1) = (aantal groepen − 1) = (3 − 1) = 2
df van de noemer = (n − p) = (totaal aantal mensen − 3) = (15 − 3) = 12
Wij hanteren een α van 0,05, we nemen dus de bovenste van de twee kritieke waarden: $F_{krit} = 3{,}88$

De beslissingsregels van deze toets worden nu als volgt geformuleerd:
1 Indien $F_{obs} < 3{,}88$, verwerp H_0 niet
2 Indien $F_{obs} \geq 3{,}88$, verwerp H_0

Berekening voor eerste experiment

We gebruiken hier de rekenformules omdat ze sneller tot het antwoord leiden, maar ook nauwkeuriger zijn.

$$\text{I} \quad \frac{T^2}{n}$$

$$\text{II} \quad \sum_{j=1}^{p} \left[\sum_{i=1}^{n_j} x_{ij}^2 \right]$$

$$\text{III} \quad \sum_{j=1}^{p} \left[\frac{T_j^2}{n_j} \right]$$

In onderstaande tabel zijn de benodigde gegevens uitgerekend.

Tabel met gegevens omtrent eerste experiment

$x\ (j = 1)$	$x^2\ (j = 1)$	$x\ (j = 2)$	$x^2\ (j = 2)$	$x\ (j = 3)$	$x^2\ (j = 3)$
6	36	7	49	10	100
6	36	8	64	7	49
5	25	6	36	8	64
7	49	8	64	7	49
6	36	6	36	8	64
$T_1 = 30$	$\sum x_1^2 = 182$	$T_2 = 35$	$\sum x_2^2 = 249$	$T_3 = 40$	$\sum x_3^2 = 326$

Tussenberekening I:

$$\frac{T^2}{n} = \frac{(30 + 35 + 40)^2}{15} = \frac{11025}{15} = 735$$

Tussenberekening II:

$$\sum_{j=1}^{p} \left[\sum_{i=1}^{n_j} x_{ij}^2 \right] = 182 + 249 + 326 = 757$$

Tussenberekening III:

$$\sum_{j=1}^{p} \left[\frac{T_j^2}{n_j} \right] = \frac{30^2}{5} + \frac{35^2}{5} + \frac{40^2}{5} = 180 + 245 + 320 = 745$$

Nu kunnen we alle waarden voor het eerste experiment berekenen:

$$MS_{tussen} = \frac{SS_{tussen}}{df_{tussen}} = \frac{III - I}{p - 1} = \frac{745 - 735}{3 - 1} = 5$$

$$MS_{binnen} = \frac{SS_{binnen}}{df_{binnen}} = \frac{II - III}{n - p} = \frac{757 - 745}{15 - 3} = \frac{12}{12} = 1{,}00$$

$$F_{obs} = \frac{MS_{tussen}}{MS_{binnen}} = \frac{5}{1} = 5$$

Berekening voor tweede experiment
In onderstaande tabel zijn de benodigde gegevens uitgerekend.

Tabel met gegevens omtrent tweede experiment

$x\ (j = 1)$	$x^2\ (j = 1)$	$x\ (j = 2)$	$x^2\ (j = 2)$	$x\ (j = 3)$	$x^2\ (j = 3)$
4	16	9	81	9	81
9	81	8	64	8	64
3	9	5	25	8	64
6	36	7	49	7	49
8	64	6	36	8	64
$T_1 = 30$	$\sum x_1^2 = 206$	$T_2 = 35$	$\sum x_2^2 = 255$	$T_3 = 40$	$\sum x_3^2 = 322$

Tussenberekening I:

$$\frac{T^2}{n} = \frac{(30 + 35 + 40)^2}{15} = \frac{11025}{15} = 735$$

Tussenberekening II:

$$\sum_{j=1}^{p}\left[\sum_{i=1}^{n_j} x_{ij}^2\right] = 206 + 255 + 322 = 783$$

Tussenberekening III:

$$\sum_{j=1}^{p}\left[\frac{T_j^2}{n_j}\right] = \frac{30^2}{5} + \frac{35^2}{5} + \frac{40^2}{5} = 180 + 245 + 320 = 745$$

Nu kunnen we alle waarden voor het tweede experiment berekenen:

$$MS_{tussen} = \frac{SS_{tussen}}{df_{tussen}} = \frac{III - I}{p - 1} = \frac{745 - 735}{3 - 1} = 5$$

$$MS_{binnen} = \frac{SS_{binnen}}{df_{binnen}} = \frac{II - III}{n - p} = \frac{783 - 745}{15 - 3} = \frac{38}{12} = 3{,}17$$

$$F_{obs} = \frac{MS_{tussen}}{MS_{binnen}} = \frac{5}{3{,}17} = 1{,}58$$

Beslissing voor eerste experiment
$F_{obs} > 3{,}88$, dus verwerp H_0

Beslissing voor tweede experiment
$F_{obs} < 3{,}88$, dus verwerp H_0 niet

Conclusie voor eerste experiment
Er is reden om aan te nemen dat deze groepsindeling effect heeft op de tentamencijfers van de groepsleden (met wiskunde-vooropleiding).

Conclusie voor tweede experiment
Er is geen reden om aan te nemen dat deze groepsindeling effect heeft op de gemiddelde tentamencijfers van de groepsleden (alle studenten).

b In het eerste experiment verwerpen we de nulhypothese wel, terwijl dat in het tweede eperiment achterwege blijft. De tussengroepsvarianties zijn in beide experimenten gelijk ($MS_{tussen} = 5$). Het verschil tussen beide experimenten zit in de binnengroepsvariantie. In experiment 1 is er sprake van een homogenere groep door te selecteren op wiskundecijfer ≥ 6, daardoor is hier de binnengroepsvariantie klein. In experiment 2 is er sprake van een heterogene groep; alle studenten zitten daarin, hierdoor is de binnengroepsvariantie veel groter.
De onderzoeksvraag heeft betrekking op alle pedagogiekstudenten, dus experiment 2 past daar beter bij.

11.2 *a* We toetsen hier weer voor beide experimenten tegelijk, maar bij de berekening, de beslissing en de conclusie zullen we onderscheid maken tussen beide experimenten.

Toetskeuze
We moeten hier drie groepen scores vergelijken, we gebruiken daarvoor de eenwegvariantieanalyse of de Kruskal-Wallis-toets. Het verschil tussen deze twee toetsen ligt in het feit dat de eerste een parametrische en de tweede een niet-parametrische toets is. Er wordt vermeld dat de scores niet-normaal verdeeld zijn, we nemen dus de niet-parametrische Kruskal-Wallis.

Onderzoeksvraag
Bestaat er verschil in tentamencijfer tussen de drie verschillende onderwijsgroepjes?

Hypothesen in termen van de onderzoeksvraag
H_0: De populaties waartoe de groepen behoren zijn identiek.
H_1: De populaties waartoe de groepen behoren zijn verschillend.

Aannamen

1 De subjecten zijn aselect en onafhankelijk verdeeld over k onafhankelijke groepen met ieder een $n_j \geq 5$.
2 De afhankelijke variabele is continu en de meetschaal is minimaal ordinaal.

Beslissingsregels

In tabel G zoeken we de kritieke waarde van beide toetsen op; deze is voor beide experimenten gelijk omdat de gegevens dezelfde zijn:

$\alpha = 0{,}05$

$df = (k - 1) = (\text{aantal groepen} - 1) = (3 - 1) = 2$

$H_{krit} = 5{,}99$

De beslissingsregels luiden als volgt:

1 Indien $H_{obs} \geq 5{,}99$, verwerp $\boldsymbol{H_0}$
2 Indien $H_{obs} < 5{,}99$, verwerp $\boldsymbol{H_0}$ niet

11.2.6 Berekening voor het eerste experiment

$$H_{obs} = \left[\frac{12}{n(n+1)}\right] \cdot \left[\frac{\sum T_j^2}{n_j}\right] - 3(n+1)$$

Berekening rangen voor het eerste experiment

Groep I		Groep II		Groep III	
Cijfer	Rang	Cijfer	Rang	Cijfer	Rang
6	4	7	8,5	10	15
6	4	8	12,5	7	8,5
5	1	6	4	8	12,5
7	8,5	8	12,5	7	8,5
6	4	6	4	8	12,5
	$T_1 = 21{,}5$		$T_2 = 41{,}5$		$T_3 = 57$

We kunnen nu de formule invullen:

$$H_{obs} = \left[\frac{12}{n(n+1)}\right] \cdot \left[\frac{\sum T_j^2}{n_j}\right] - 3(n+1)$$

$$= \left[\frac{12}{15(16)}\right] \cdot \left[\frac{21{,}5^2}{5} + \frac{41{,}5^2}{5} + \frac{57^2}{5}\right] - 3(16)$$

$$= \left[\frac{12}{240}\right] \cdot [92{,}45 + 344{,}45 + 649{,}8] - 48$$

$$= [0{,}05] \cdot [1086{,}7] - 48 = 6{,}335$$

Berekening voor het tweede experiment

$$H_{obs} = \left[\frac{12}{n(n+1)}\right] \cdot \left[\frac{\sum T_j^2}{n_j}\right] - 3(n+1)$$

Berekening rangen voor het tweede experiment

Groep I		Groep II		Groep III	
Cijfer	Rang	Cijfer	Rang	Cijfer	Rang
4	2	9	14	9	14
9	14	8	10	8	10
3	1	5	3	8	10
6	4,5	7	6,5	7	6,5
8	10	6	4,5	8	10
	$T_1 = 31{,}5$		$T_2 = 38$		$T_3 = 50{,}5$

We kunnen nu de formule invullen:

$$H_{obs} = \left[\frac{12}{n(n+1)}\right] \cdot \left[\frac{\sum T_j^2}{n_j}\right] - 3(n+1)$$

$$= \left[\frac{12}{15(16)}\right] \cdot \left[\frac{31{,}5^2}{5} + \frac{38^2}{5} + \frac{50{,}5^2}{5}\right] - 3(16)$$

$$= \left[\frac{12}{240}\right] \cdot [198{,}45 + 288{,}8 + 510{,}05] - 48$$

$$= [0{,}05] \cdot [997{,}3] - 48 = 1{,}865$$

Beslissing voor het eerste experiment

$H_{obs} > H_{krit}$, dus verwerp H_0

Beslissing voor het tweede experiment

$H_{obs} < H_{krit}$, dus verwerp H_0 niet

Conclusie voor eerste experiment

Er is reden om aan te nemen dat deze groepsindeling effect heeft op de tentamencijfers van de groepsleden (met wiskundevooropleiding).

Conclusie voor tweede experiment

Er is geen reden om aan te nemen dat deze groepsindeling effect heeft op de gemiddelde tentamencijfers van de groepsleden (alle studenten).

b De verklaring voor rare pieken en dalen bij tentamencijfers tussen 5 en 6 kan waarschijnlijk worden gevonden in het feit dat het daar om de grens gaat tussen wel en niet voldoende.

12 Tweewegvariantieanalyse

In hoofdstuk 11 hebben we gezien dat de eenwegvariantieanalyse ons in staat stelt het verschil tussen meer dan twee groepen te toetsen. We keken naar verschil in scores op de afhankelijke variabele (bijvoorbeeld smaak of tentamencijfer); we waren van mening dat dit verschil veroorzaakt werd door aan- of afwezigheid van een onafhankelijke variabele. Zo keken we naar het verschil in smaakbeoordelingen (afhankelijke variabele) dat veroorzaakt zou zijn doordat de mensen in de drie groepen verschillende soorten alcoholvrij bier te drinken kregen (onafhankelijke variabele). Een ander voorbeeld is dat we naar het verschil in statistiekcijfer kijken, dat veroorzaakt zou zijn door andere onderwijsmethoden; in dit voorbeeld is het cijfer de afhankelijke variabele (datgene wat beïnvloed wordt) en de variabele onderwijsmethode (bestaand uit meerdere methoden) is de onafhankelijke variabele (datgene wat de afhankelijke variabele beïnvloedt).
In dit hoofdstuk toetsen we steeds het verschil op één afhankelijke variabele, dat veroorzaakt wordt (of niet) door *twee* onafhankelijke variabelen.
Om het één en ander te verduidelijken, gebruiken we de volgende fictieve voorbeeldsituatie.

12.1 Een voorbeeldanalyse

Een sportschool beschikt, behalve over een grote sportaccommodatie, ook over een diëtiste, om professionele hulp te bieden aan mensen die wat willen afvallen. De diëtiste begint met de introductie van twee diëten:
dieet K, waarbij men vooral veel koolhydraten moet eten
dieet F, waarbij men vooral veel fruit en groenten moet eten.

Na een aantal weken wil de diëtiste weten welk dieet het beste werkt. Ook wil ze nagaan of de diëten anders werken als men erbij sport. Zij meet de werking van beide variabelen door middel van het aantal kilo's dat men na een maand kwijt is geraakt.
In dit voorbeeld hebben we één afhankelijke variabele, namelijk het aantal kilo's dat men na een maand is afgevallen en twee onafhankelijke variabelen waarvan de diëtiste de invloed op het gewichtsverlies wil bekijken, ten eerste het soort dieet (F of K) en ten tweede het al of niet sporten naast het dieet.
Totaal kiest zij 20 mensen waarvan 10 dieet K en 10 dieet F volgen. In elke dieetgroep sporten 5 mensen wel en 5 mensen niet. In tabel 12.1 staan de resultaten weergegeven in aantal kilo's gewichtsverlies.
We noemen de variabele sporten 'factor A' en de variabele dieet 'factor B'.
Wel en *niet* sporten noemen we niveaus van factor A, *dieet F* en *dieet K* de niveaus van factor B.
We gaan toetsen of de twee onafhankelijke variabelen afzonderlijk effect hebben op gewichtsverlies, en of ze daarnaast wellicht een gecombineerd effect hebben op de afhankelijke variabele. We nemen daarbij een α van 5%.

Tabel 12.1

Soort dieet → / *Sporten* ↓	*Dieet K*	*Dieet F*
Niet	1	2
	3	5
	2	6
	2	4
	1	3
Wel	2	8
	4	4
	5	4
	3	5
	1	3

12.1.1 Toetskeuze

We gaan hier het effect meten van twee onafhankelijke variabelen op één afhankelijke variabele. De toets die hiervoor geschikt is, is de tweewegvariantieanalyse. Deze toets wordt net als de eenwegvariantieanalyse tweezijdig uitgevoerd.

12.1.2 Onderzoeksvraag

Bij de tweewegvariantieanalyse kunnen we door middel van het uitvoeren van één toets drie onderzoeksvragen beantwoorden, namelijk:

1. Is er verschil in gewichtsverlies tussen dieet F en dieet K, ongeacht het feit of men al dan niet sport? Met andere woorden: is er sprake van een hoofdeffect voor factor B?
2. Is er verschil in gewichtsverlies tussen de mensen die wel en niet sporten, ongeacht het feit of men dieet F of K volgt? Met andere woorden: is er sprake van een hoofdeffect voor factor A?
3. Is er een effect van gecombineerd toepassen van het al dan niet sporten en het volgen van één van de diëten? Met andere woorden: is er sprake van interactie tussen de niveaus van beide factoren?

12.1.3 Hypothesen

Doordat we drie onderzoeksvragen hebben, is er bij de tweewegvariantieanalyse ook sprake van drie paren hypothesen. We gebruiken daarbij de volgende symbolen:

α_1 = staat voor het gemiddelde van niveau 1 van factor A. In dit geval staat dat voor het gemiddelde gewichtsverlies bij niet sporten.

α_2 = staat voor het gemiddelde van niveau 2 van factor A. In dit geval staat dat voor het gemiddelde gewichtsverlies bij wel sporten.

β_1 = staat voor het gemiddelde van niveau 1 van factor B. In dit geval staat dat voor het gemiddelde gewichtsverlies bij dieet K.

β_2 = staat voor het gemiddelde van niveau 2 van factor B. In dit geval staat dat voor het gemiddelde gewichtsverlies bij dieet F.

$\alpha_1\beta_1$ = staat voor het celgemiddelde van de combinatie tussen niveau 1 van factor A en niveau 1 van factor B. In dit geval komt dat neer op het gemiddelde gewichtsverlies bij niet sporten en het volgen van dieet K.
$\alpha_2\beta_1$ = staat voor het celgemiddelde van de combinatie tussen niveau 2 van factor A en niveau 1 van factor B. In dit geval komt dat neer op het gemiddelde gewichtsverlies bij wel sporten en het volgen van dieet K.
$\alpha_1\beta_2$ = staat voor het celgemiddelde van de combinatie tussen niveau 1 van factor A en niveau 2 van factor B. In dit geval komt dat neer op het gemiddelde gewichtsverlies bij niet sporten en het volgen van dieet F.
$\alpha_2\beta_2$ = staat voor het celgemiddelde van de combinatie tussen niveau 2 van factor A en niveau 2 van factor B. In dit geval komt dat neer op het gemiddelde gewichtsverlies bij wel sporten en het volgen van dieet F.

Hypothesen voor het toetsen van een hoofdeffect voor factor A
H_0: $\alpha_1 = \alpha_2$
H_1: *niet* H_0
Hypothesen voor het toetsen van een hoofdeffect voor factor B
H_0: $\beta_1 = \beta_2$
H_1: *niet* H_0
Hypothesen voor het toetsen van interactie
H_0: $\alpha_1\,\beta_1 = \alpha_1\,\beta_2 = \alpha_2\,\beta_1 = \alpha_2\,\beta_2$
H_1: *niet* H_0

12.1.4 Aannamen

1 De subjecten zijn aselect en onafhankelijk getrokken.
2 De groepen zijn onafhankelijk.
3 De populatievarianties zijn homogeen.
4 De populatieverdelingen zijn normaal van vorm.
5 De factoren in het onderzoek staan vast.
6 Het aantal waarnemingen in elke groep is gelijk en groter dan 1.

12.1.5 Beslissingsregels

We berekenen straks voor elk paar hypothesen een geobserveerde F-waarde, die ongeveer op gelijke wijze totstandkomt als bij de eenwegvariantieanalyse. Daarbij geldt steeds een ander aantal vrijheidsgraden. Per onderzoeksvraag stellen we het aantal vrijheidsgraden vast, en zoeken we de kritieke waarden op in tabel I.

Beslissingsregels voor het toetsen van een hoofdeffect voor factor A

df van de teller = $p - 1 = 2 - 1 = 1$
(aantal niveaus van factor A $- 1$)

df van de noemer = $n - pq = 20 - 2 \cdot 2 = 16$
(aantal mensen in experiment $-$ aantal niveaus van A $\cdot$ aantal niveaus van B)

tabel I, $\alpha = 0{,}05$ (bovenste waarde)
$F_{kritiek}\ A = 4{,}49$

Beslissingsregels:
1 Indien $F_{obs}\ A \geq 4{,}49$, verwerp $\boldsymbol{H_0}$
2 Indien $F_{obs}\ A < 4{,}49$, verwerp $\boldsymbol{H_0}$ niet

Beslissingsregels voor het toetsen van een hoofdeffect voor factor B
df van de teller = q − 1 = 2 − 1 = 1
(aantal niveaus van factor B − 1)

df van de noemer = n − pq = 20 − 2 · 2 = 16
(aantal mensen in experiment − aantal niveaus van A · aantal niveaus van B)

tabel I, $\alpha = 0{,}05$ (bovenste waarde)
$F_{kritiek}\ B = 4{,}49$

Beslissingsregels:
1 Indien $F_{obs}\ B \geq 4{,}49$, verwerp $\boldsymbol{H_0}$
2 Indien $F_{obs}\ B < 4{,}49$, verwerp $\boldsymbol{H_0}$ niet

Beslissingsregels voor het toetsen van interactie
df van de teller = (p − 1) (q − 1) = 1 · 1 = 1
(aantal niveaus van factor A − 1) · (aantal niveaus van factor B − 1)

df van de noemer = n − pq = 20 − 2 · 2 = 16
(aantal mensen in experiment − aantal niveaus van A · aantal niveaus van B)

tabel I, $\alpha = 0{,}05$ (bovenste waarde)
$F_{kritiek}\ AB = 4{,}49$

Beslissingsregels:
1 Indien $F_{obs}\ AB \geq 4{,}49$, verwerp $\boldsymbol{H_0}$
2 Indien $F_{obs}\ AB < 4{,}49$, verwerp $\boldsymbol{H_0}$ niet

12.1.6 Berekening

De berekening van de geobserveerde F-waarden komt tot stand via verschillende tussenberekeningen:

I $\dfrac{T_{..}^2}{n_{tl}}$

II $\sum_{i=1}^{n}\sum_{j=1}^{p}\sum_{k=1}^{q} x_{ijk}^2$

III $\sum_{j=1}^{p}\dfrac{T_{j.}^2}{n \cdot q}$

$$\text{IV} \quad \sum_{k=1}^{q} \frac{T_{k.}^2}{n \cdot p}$$

$$\text{V} \quad \frac{\sum_{j=1}^{p} \sum_{k=1}^{q} T_{jk}^2}{n}$$

T staat steeds voor het totaal gewichtsverlies.
n staat voor het aantal mensen per cel.
n_{ttl} staat voor het totaal aantal mensen.
q staat voor het totaal aantal kolommen.
k staat voor een bepaalde kolom aanduiding.
p staat voor het totaal aantal rijen.
j staat voor een bepaalde rij aanduiding.

In tabel 12.2 staan de opgetelde gewichtsverliezen. In deze tabel kunnen we van tevoren het één en ander voorzien.
Als we naar de kolomtotalen (totaal gewichtsverlies bij dieet K en F) kijken, zien we een duidelijk verschil tussen deze getallen. Bij dieet F is men meer afgevallen dan bij K, ongeacht het wel of niet sporten naast het dieet. Deze getallen wijzen op een hoofdeffect voor factor B. Als we dan kijken of de mensen die wel of niet sporten anders reageren op dieet F of K, zien we dat de mensen die niet sporten, minder afvallen bij dieet K en dat geldt voor de sporters ook. Dit wijst erop dat er geen interactie-effect is. De verschillen tussen de rijtotalen zijn wat minder groot; waarschijnlijk zal er geen hoofdeffect voor factor A optreden.

Of de hierboven gesignaleerde effecten ook significant zijn, zal toetsing moeten uitwijzen. We gaan daartoe eerst alle tussenberekeningen uitvoeren.

Tabel 12.2

Soort dieet → / *Sporten ↓*	*Dieet K*	*Dieet F*	*Rijtotalen*
Niet	9	20	29
Wel	15	24	39
Kolomtotalen	24	44	68

We moeten hier alle gewichtsverliezen optellen, dit totaal kwadrateren en delen door het totaal aantal mensen.

$$\frac{T^2}{n_{ttl}} = \frac{68^2}{20} = \frac{4624}{20} = 231{,}2$$

Tussenberekening II

We moeten hier voor elke persoon per cel, voor alle rijen en kolommen, het gewichtsverlies kwadrateren, en al deze waarden optellen.

Rij 1: kolom 1: $1^2 + 3^2 + 2^2 + 2^2 + 1^2 = 19$

kolom 2: $2^2 + 5^2 + 6^2 + 4^2 + 3^2 = 90$

Rij 2: kolom 1: $2^2 + 4^2 + 5^2 + 3^2 + 1^2 = 55$

kolom 2: $8^2 + 4^2 + 4^2 + 5^2 + 3^2 = 130$

$$\sum_{i=1}^{n} \sum_{j=1}^{p} \sum_{k=1}^{q} x_{ijk}^2 = 19 + 90 + 55 + 130 = 294$$

Tussenberekening III

We moeten hier elk rijtotaal kwadrateren en delen door het aantal mensen per cel vermenigvuldigd met het aantal kolommen, daarna alle waarden weer optellen.

$$\sum_{j=1}^{p} \frac{T_{j.}^2}{n \cdot q} = \frac{29^2}{5 \cdot 2} + \frac{39^2}{5 \cdot 2} = \frac{841}{10} + \frac{1521}{10} = 84{,}1 + 152{,}1 = 236{,}2$$

Tussenberekening IV

We moeten hier elk kolomtotaal kwadrateren en delen door het aantal mensen per cel vermenigvuldigd met het aantal rijen, daarna alle waarden optellen.

$$\sum_{k=1}^{q} \frac{T_{k}^2}{n \cdot p} = \frac{24^2}{5 \cdot 2} + \frac{44^2}{5 \cdot 2} = \frac{576}{10} + \frac{1936}{10} = 57{,}6 + 193{,}6 = 251{,}2$$

Tussenberekening V

Voor elke cel moeten we het totaal kwadrateren, deze waarden tellen we op, en dat delen we door het aantal mensen per cel.

$$\frac{\sum_{j=1}^{p} \sum_{k=1}^{q} T_{jk}^2}{n} = \frac{9^2 + 20^2 + 15^2 + 24^2}{5} = \frac{81 + 400 + 225 + 576}{5} = \frac{1282}{5} = 256{,}4$$

Berekening van $F_{obs}\,A$

$$F_{obs}\,A = \frac{MS_{tussen}\,A}{MS_{binnen}}$$

$$MS_{tussen}\,A = \frac{SS_{tussen}\,A}{df_{tussen}\,A} = \frac{III - I}{p - 1} = \frac{236{,}2 - 231{,}2}{2 - 1} = 5$$

$$MS_{binnen} = \frac{SS_{binnen}}{df_{binnen}} = \frac{II - V}{n_{ttl} - pq} = \frac{294 - 256{,}4}{20 - 4} = \frac{37{,}6}{16} = 2{,}35$$

$$F_{obs}\,A = \frac{MS_{tussen}\,A}{MS_{binnen}} = \frac{5}{2{,}35} = 2{,}13$$

Berekening van F_{obs} B

$$F_{obs}\,B = \frac{MS_{tussen}\,B}{MS_{binnen}}$$

$$MS_{tussen}\,B = \frac{SS_{tussen}\,B}{df_{tussen}\,B} = \frac{IV - I}{q - 1} = \frac{251{,}2 - 231{,}2}{2 - 1} = \frac{20}{1} = 20$$

$$MS_{binnen} = \frac{SS_{binnen}}{df_{binnen}} = \frac{II - V}{n_{ttl} - pq} = \frac{294 - 256{,}4}{20 - 4} = \frac{37{,}6}{16} = 2{,}35$$

$$F_{obs}\,B = \frac{MS_{tussen}\,B}{MS_{binnen}} = \frac{20}{2{,}35} = 8{,}5$$

Berekening van F_{obs} AB

$$F_{obs}\,AB = \frac{MS_{tussen}\,AB}{MS_{binnen}}$$

$$MS_{tussen}\,AB = \frac{SS_{tussen}\,AB}{df_{tussen}\,AB} = \frac{V + I - III - IV}{(p - 1) \cdot (q - 1)}$$

$$= \frac{256{,}4 + 231{,}2 - 236{,}2 - 251{,}2}{(2 - 1) \cdot (2 - 1)} = \frac{0{,}2}{1} = 0{,}2$$

$$MS_{binnen} = \frac{SS_{binnen}}{df_{binnen}} = \frac{II - V}{n_{ttl} - pq} = \frac{294 - 256{,}4}{20 - 4} = \frac{37{,}6}{16} = 2{,}35$$

$$F_{obs}\,AB = \frac{MS_{tussen}\,AB}{MS_{binnen}} = \frac{0{,}2}{2{,}35} = 0{,}09$$

12.1.7 Beslissing

We zien dat alleen de geobserveerde waarde van het hoofdeffect voor factor B significant is; er is geen hoofdeffect voor factor A en er is geen interactie aanwezig.

12.1.8 Conclusie

Er is reden om aan te nemen dat er verschil in gewichtsverlies bestaat tussen mensen die dieet F of dieet K volgen. Of men daarnaast sport of niet, doet niets af aan dit effect.

12.2 Mogelijke andere effecten bij tweewegvariantieanalyse

In deze paragraaf zullen we een paar zeer overdreven voorbeeldjes presenteren, waaruit we effecten kunnen aflezen. Omdat een hoofdeffect voor factor B al besproken is, behandelen we dit effect hieronder niet. We maken gebruik van het voorbeeld over de diëtiste.

12.2.1 *Hoofdeffect factor A*

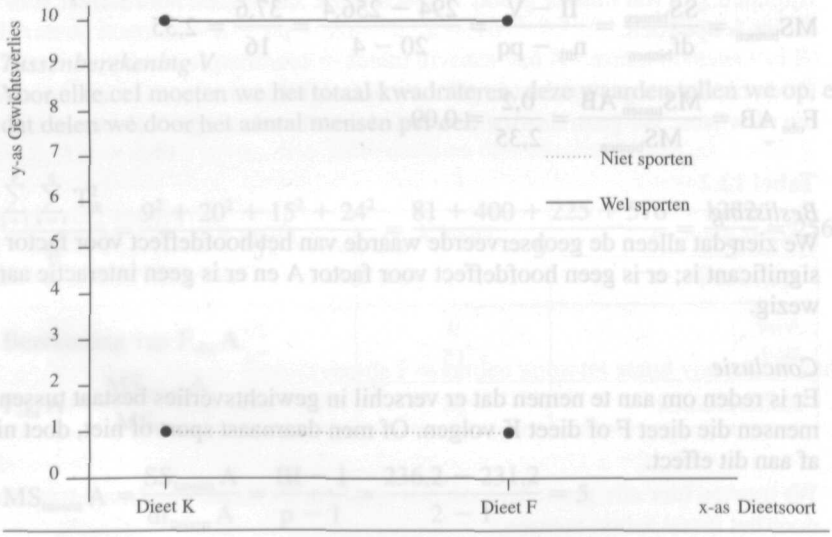
Tabel 12.3

Soort dieet → / *Sporten ↓*	*Dieet K*	*Dieet F*	*Rijtotalen*
Niet	1	1	2
Wel	10	10	20
Kolomtotalen	11	11	22

In tabel 12.3 zien we een situatie waarbij er een duidelijk verschil is tussen de rijtotalen: bij sporten valt men duidelijk veel meer af dan bij niet sporten. Het doet niet terzake of men dieet F of K volgt, want bij beide is het gewichtsverlies hetzelfde. Er is hier dus sprake van een hoofdeffect voor factor A.

In de figuur hieronder staat de situatie weergegeven. Op de y-as is de afhankelijke variabele gepresenteerd, namelijk gewichtsverlies. Op de x-as staat de variabele dieet. De tweede onafhankelijke variabele is in de figuur gebracht door voor elk niveau van factor A een aparte lijn te trekken in de figuur. We noemen dit type figuur een **interactieplaatje**.

Figuur 12.1 **Hoofdeffect factor A**



12.2.2 *Interactie*

Interactie is aanwezig als juist de combinatie van een niveau van de ene factor en een niveau van de andere factor leidt tot verschillen op de afhankelijke variabele. We kunnen drie soorten interactie tegenkomen. Van elke soort laten we weer een overdreven voorbeeld zien.

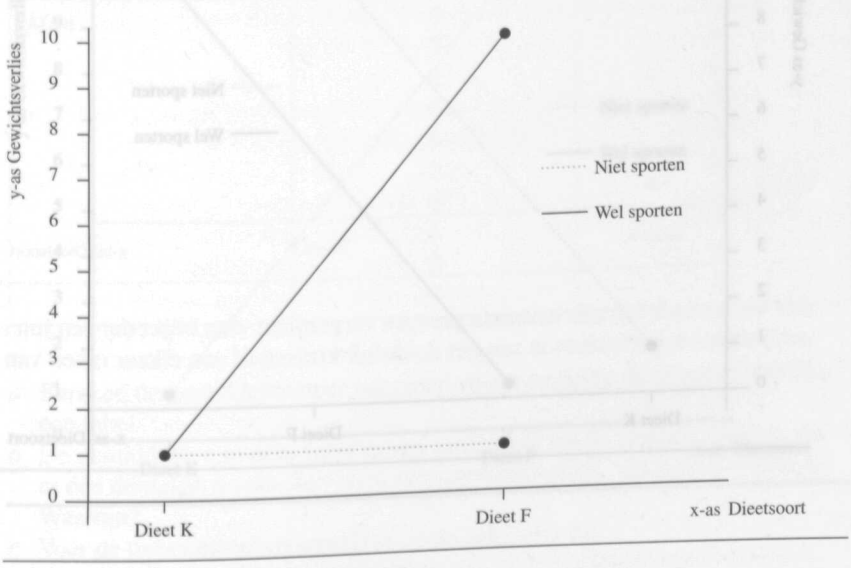
Catalytische interactie

Tabel 12.4

Soort dieet → / *Sporten ↓*	*Dieet K*	*Dieet F*	*Rijtotalen*
Niet	1	1	2
Wel	1	10	11
Kolomtotalen	2	11	13

In dit geval lijkt het of er twee hoofdeffecten zijn, voor factor A en factor B. Deze effecten kunnen bij toetsing significant zijn, maar het probleem bij deze hoofdeffecten is dat ze niet gelden voor elk niveau van de andere factor. Als we kijken naar factor A (sporten), dan zien we (in de rijtotalen) dat wel sporten tot meer gewichtsverlies leidt dan niet sporten. Dit geldt echter niet altijd, want als men dieet K volgt, dan mag het sporten niet baten! Afzonderlijk heeft dus geen enkel niveau van de beide factoren effect op de afhankelijke variabele, slechts een combinatie tussen dieet F en wel sporten leidt tot meer gewichtsverlies. Dit noemen we een **catalytische interactie**. Als men te maken krijgt met zowel significante hoofdeffecten en significante interactie, dan kan men niet zonder meer de hoofdeffecten als zodanig interpreteren. Doordat er sprake van interactie is, gelden ze namelijk niet altijd (niet voor elk niveau van de andere factor).

Figuur 12.2 **Catalytische interactie**



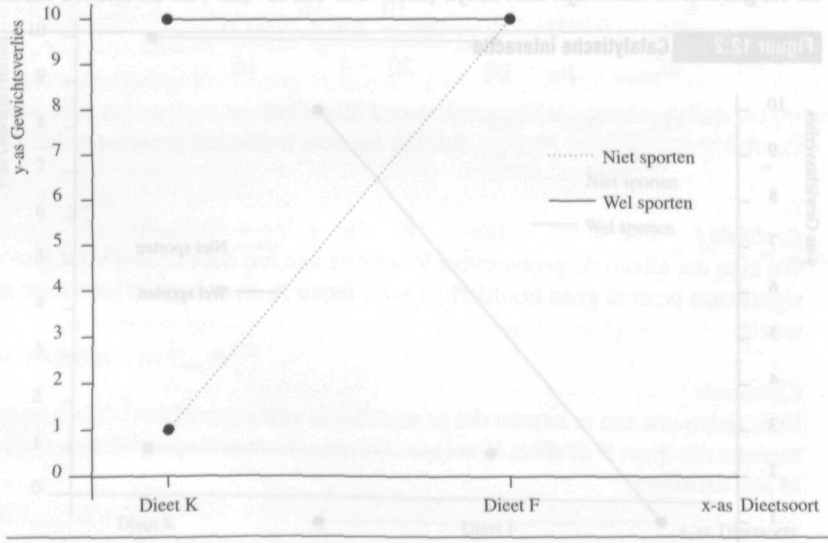
Terminatieve interactie

Tabel 12.5

Soort dieet → / *Sporten ↓*	*Dieet K*	*Dieet F*	*Rijtotalen*
Niet	1	10	11
Wel	10	10	20
Kolomtotalen	11	20	31

In dit geval lijkt het weer of er twee hoofdeffecten zijn, voor factor A en factor B. Als we kijken naar factor B (dieet), dan zien we in de kolomtotalen dat dieet F leidt tot meer gewichtsverlies dan dieet K. Als dit om een echt hoofdeffect zou gaan, dan zou dit effect er bij elk niveau van de andere factor moeten zijn. Bij niveau niet sporten klopt het hoofdeffect, daar is dieet F beter dan dieet K. Bij wel sporten echter maakt het niet uit of men het ene of het andere dieet gebruikt. Het hoofdeffect geldt dus niet onvoorwaardelijk. We zien hier dat de factoren elk hun effect hebben, maar een combinatie van beide leidt niet tot een hoger effect. Dit noemen we een **terminatieve interactie**.

Figuur 12.3 **Terminatieve interactie**



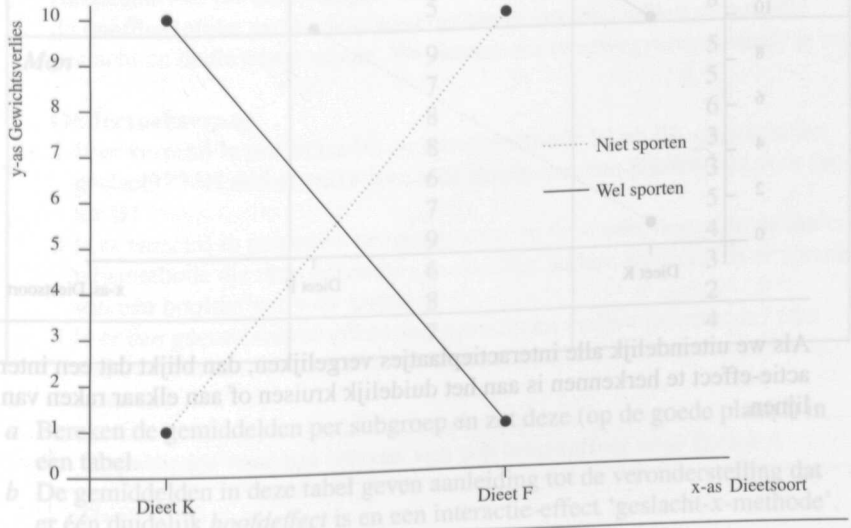
Antagonistische interactie

Tabel 12.6

Soort dieet → / *Sporten ↓*	*Dieet K*	*Dieet F*	*Rijtotalen*
Niet	1	10	11
Wel	10	1	11
Kolomtotalen	11	11	22

In deze tabel lijkt er niets aan de hand te zijn; de rijtotalen verschillen niet van elkaar, er is dus geen hoofdeffect voor factor A. De kolomtotalen verschillen ook niet van elkaar, dus er is ook geen hoofdeffect voor factor B. Als we echter in de cellen kijken, dan zien we wel degelijk verschillende reacties op de verschillende behandelingen. Het volgen van dieet K en daarbij niet sporten leidt tot slechte resultaten, sport men echter wel bij dit dieet, dan boekt men goede resultaten. Dieet F moet men vooral gebruiken als men er niet naast wil sporten, dit leidt tot zeer goede resultaten. Sport men er echter wel bij, dan wacht er een teleurstelling. We noemen dit een **tegengestelde, of antagonistische interactie**.

Figuur 12.4 **Antagonistische interactie**



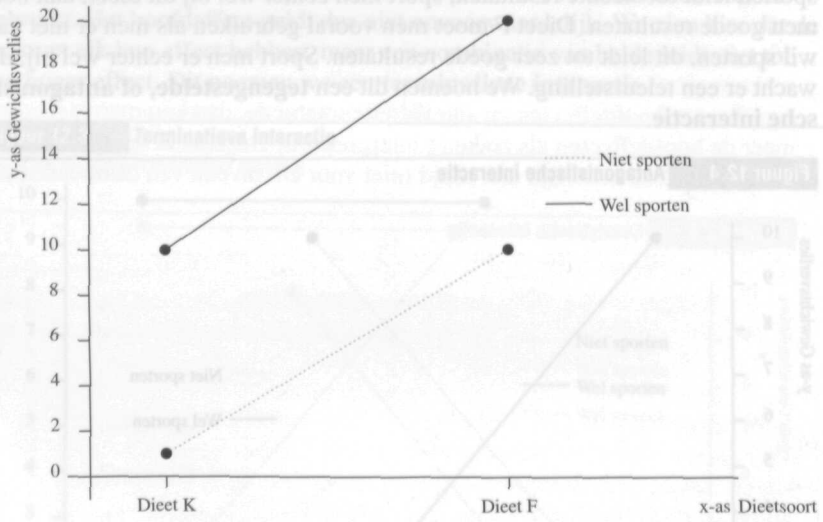
Additieve relatie

Tabel 12.7

Soort dieet → / *Sporten ↓*	*Dieet K*	*Dieet F*	*Rijtotalen*
Niet	1	10	11
Wel	10	20	30
Kolomtotalen	11	30	41

Deze laatste mogelijke uitkomst van de tweewegvariantieanalyse is de additieve relatie. Er is dan sprake van twee hoofdeffecten, en ook van een gecombineerd effect van de beide factoren.

Figuur 12.5 **Additieve relatie**



Als we uiteindelijk alle interactieplaatjes vergelijken, dan blijkt dat een interactie-effect te herkennen is aan het duidelijk kruisen of aan elkaar raken van de lijnen.

Opgaven

12.1 In een onderzoek bij pedagogiekstudenten werd de effectiviteit nagegaan van twee verschillende onderwijsmethoden (onafhankelijke, predictorvariabele). Met name ging het erom of het uitmaakt of studenten alleen colleges krijgen, of colleges en werkgroepen. Afhankelijke variabele was de score op een prestatietest. Een tweede onafhankelijke variabele was het geslacht van de studenten: 20 mannelijke en 20 vrouwelijke studenten werden aselect over de beide onderwijsmethoden verdeeld. In tabel 1 worden de ruwe scores gepresenteerd.

Tabel 1 Ruwe scores van de variabele prestatietest

Onderwijsmethode →	*b1*	*b2*
Geslacht ↓		
Vrouw	7 6 4 6 7 6 5 4 5 5	5 6 7 7 7 6 6 6 7 8
Man	9 7 8 8 6 7 9 6 8 7	5 5 6 3 3 5 4 3 2 4

a Bereken de gemiddelden per subgroep en zet deze (op de goede plaats!) in een tabel.

b De gemiddelden in deze tabel geven aanleiding tot de veronderstelling dat er één duidelijk *hoofdeffect* is en een interactie-effect 'geslacht-x-methode'. Waarom?

c Voer de tweewegvariantieanalyse op de gegevens uit.

d Het zou onverstandig zijn om op basis van deze resultaten zonder meer te stellen dat aangetoond is dat de ene onderwijsmethode beter is dan de andere. Waarom?

e Maak een plaatje van de resultaten van het onderzoek, uitgaande van de subgroepgemiddelden uit tabel 1. Zet de niveaus van de factor onderwijsmethode op de horizontale as.

f In hoeverre vind je in bovenstaand plaatje steun voor het bij vraag d. gegeven antwoord? Waaruit blijkt dat met name?

12.2 *Groepsgemiddelden Prestatietest*

Onderwijsmethode → / *Geslacht↓*	*b1*	*b2*	*Rijgemiddelden*
Vrouw	6,5	8,7	7,6
Man	5,5	7,5	6,5
Kolomgemiddelden	6,0	8,1	7,1

a Stél dat het onderzoek de bovenstaande (gemiddelde) scores op de prestatietest had opgeleverd. Interpreteer de resultaten en vergelijk ze met die uit het eerste onderzoek. Is er zo te zien een *hoofdeffect voor onderwijsmethode*? En voor *geslacht*? Is er een *interactie-effect*? Teken ook een figuur als in vraag 12.1.e.

b Vat de resultaten van dit tweede onderzoek, mede aan de hand van de figuur, overzichtelijk samen.

Uitwerkingen

12.1 *a*

Onderwijsmethode → / *Geslacht↓*	*b1*	*b2*	*Rij-gemiddelden*
Vrouw	5,5	6,5	6,0
Man	7,5	4,0	5,75
Kolomgemiddelden	6,5	5,25	5,875

b Als we naar een hoofdeffect zoeken, moeten we kijken naar de rij- en de kolom-gemiddelden. We zien dan dat de rijgemiddelden niet zo veel verschillen. Dus ♀ en ♂ verschillen nauwelijks van elkaar. Maar voor factor b, onderwijsmethode, zien we dat het kolomgemiddelde van onderwijsmethode b1 hoger is dan bij b2. Opvallend is dat jongens bij methode b2 een zeer laag gemiddelde hebben en bij methode b1 een vrij hoog gemiddelde. Tweewegvariantieanalyse zal uit moeten wijzen of deze in de steekproef gevonden verschillen waarschijnlijk ook zo terug te vinden zijn in de populatie.

c **Toetskeuze**
We gaan hier het effect meten van twee onafhankelijke variabelen op één afhankelijke variabele. De toets die hiervoor geschikt is, is de tweewegvariantieanalyse. De afhankelijke variabele is hier de score op de prestatietest, de onafhankelijke variabelen, waarvan we de invloed willen meten, zijn geslacht en onderwijsmethode. We voeren een tweewegvariantieanalyse uit.

Onderzoeksvraag
1 Is er verschil in prestaties bij onderwijsmethode b1 en b2, ongeacht het geslacht? Met andere woorden: is er sprake van een hoofdeffect voor factor B?
2 Is er verschil in prestaties tussen mannen en vrouwen, ongeacht de onderwijsmethode die men ontvangen heeft? Met andere woorden: is er sprake van een hoofdeffect voor factor A?
3 Is er een gecombineerd effect van geslacht en onderwijsmethode? Met andere woorden: is er sprake van interactie tussen de niveaus van beide factoren?

Hypothesen voor het toetsen van een hoofdeffect voor factor A
H_0: $\alpha_1 = \alpha_2$
H_1: *niet* H_0
Hypothesen voor het toetsen van een hoofdeffect voor factor B
H_0: $\beta_1 = \beta_2$
H_1: *niet* H_0

Hypothesen voor het toetsen van interactie
H_0: $\alpha_1 \beta_1 = \alpha_1 \beta_2 = \alpha_2 \beta_1 = \alpha_2 \beta_2$
H_1: *niet* H_0

Aannamen

1 De subjecten zijn aselect en onafhankelijk getrokken.
2 De groepen zijn onafhankelijk.
3 De populatievarianties zijn homogeen.
4 De populatieverdelingen zijn normaal van vorm.
5 De factoren in het onderzoek staan vast.
6 Het aantal waarnemingen in elke groep is gelijk en groter dan 1.

Beslissingsregels voor toetsen van hoofdeffect voor factor A:

df van de teller = $p - 1 = 2 - 1 = 1$
df van de noemer = $n - pq = 40 - 2 \cdot 2 = 36$

tabel I, $\alpha = 0{,}05$ (bovenste waarde)
$F_{kritiek}$ A = 4,11

Beslissingsregels:

1 Indien F_{obs} A $\geq$ 4,11 , verwerp H_0
2 Indien F_{obs} A $<$ 4,11 , verwerp H_0 niet

Beslissingsregels voor het toetsen van hoofdeffect voor factor B

df van de teller = $q - 1 = 2 - 1 = 1$
df van de noemer = $n - pq = 40 - 2 \cdot 2 = 36$

tabel I, $\alpha = 0{,}05$ (bovenste waarde)
$F_{kritiek}$ B = 4,11

Beslissingsregels:

1 Indien F_{obs} B $\geq$ 4,11 , verwerp H_0
2 Indien F_{obs} B $<$ 4,11 , verwerp H_0 niet

Beslissingsregels voor het toetsen van interactie

df van de teller = $(p - 1)(q - 1) = 1 \cdot 1 = 1$
df van de noemer = $n - pq = 40 - 2 \cdot 2 = 36$

tabel I, $\alpha = 0{,}05$ (bovenste waarde)
$F_{kritiek}$ AB = 4,11

Beslissingsregels:

1 Indien F_{obs} AB $\geq$ 4,11 , verwerp H_0
2 Indien F_{obs} AB $<$ 4,11 , verwerp H_0 niet

Berekening

De berekening van de geobserveerde F-waarden komt tot stand via verschillende tussenberekeningen:

$$\text{I} \quad \frac{T_{..}^2}{n_{ttl}}$$

$$\text{II} \quad \sum_{i=1}^{n} \sum_{j=1}^{p} \sum_{k=1}^{q} x_{ijk}^2$$

$$\text{III} \quad \sum_{j=1}^{p} \frac{T_{j.}^2}{n \cdot q}$$

$$\text{IV} \quad \sum_{k=1}^{q} \frac{T_{k.}^2}{n \cdot p}$$

$$\text{V} \quad \frac{\sum_{j=1}^{p} \sum_{k=1}^{q} T_{jk}^2}{n}$$

Onderwijsmethode → / *Geslacht*↓	*b1*	*b2*	*Rij-gemiddelden*
Vrouw	55	65	120
Man	75	40	115
Kolomgemiddelden	130	105	235

Tussenberekening I

We moeten hier alle prestaties optellen, dit totaal kwadrateren, en delen door het totaal aantal mensen.

$$\frac{T_{..}^2}{n_{ttl}} = \frac{(55 + 65 + 75 + 40)^2}{40} = \frac{235^2}{40} = 1380{,}625$$

Tussenberekening II

We moeten hier voor elke persoon per cel, voor alle rijen en kolommen, de prestaties kwadrateren, en al deze waarden optellen:

II = 1e cel: $7^2 + 6^2 + 4^2 + 6^2 + 7^2 + 6^2 + 5^2 + 4^2 + 5^2 + 5^2 = 313$

2e cel: $9^2 + 7^2 + 8^2 + 8^2 + 6^2 + 7^2 + 9^2 + 6^2 + 8^2 + 7^2 = 573$

3e cel: $5^2 + 6^2 + 7^2 + 7^2 + 7^2 + 6^2 + 6^2 + 6^2 + 7^2 + 8^2 = 429$

4e cel: $5^2 + 5^2 + 6^2 + 3^2 + 3^2 + 5^2 + 4^2 + 3^2 + 2^2 + 4^2 = 174$

1489

$$\sum_{i=1}^{n} \sum_{j=1}^{p} \sum_{k=1}^{q} x_{ijk}^2 = 1489$$

Tussenberekening III

We moeten hier elk rijtotaal kwadrateren en delen door het aantal mensen per cel vermenigvuldigd met het aantal kolommen. Daarna alle waarden weer optellen.

$$\sum_{j=1}^{p} \frac{T_{j.}^2}{n \cdot q} = \frac{(120)^2}{10 \cdot 2} + \frac{(115)^2}{10 \cdot 2} = 720 + 661{,}25 = 1381{,}25$$

Tussenberekening IV

We moeten hier elk kolomtotaal kwadrateren en delen door het aantal mensen per cel vermenigvuldigd met het aantal rijen, daarna alle waarden optellen.

$$\sum_{k=1}^{q} \frac{T_{k.}^2}{n \cdot p} = \frac{(130)^2}{10 \cdot 2} + \frac{(105)^2}{10 \cdot 2} = 845 + 551{,}25 = 1396{,}25$$

Tussenberekening V

Voor elke cel moeten we het totaal kwadrateren, deze waarden tellen we op, en dat delen we door het aantal mensen per cel.

$$\frac{\sum_{j=1}^{p} \sum_{k=1}^{q} T_{jk}^2}{n} = \frac{55^2 + 65^2 + 75^2 + 40^2}{10} = 1447{,}5$$

Berekening van F_{obs} A

$$F_{obs}\,A = \frac{MS_{tussen}\,A}{MS_{binnen}}$$

$$MS_{tussen}\,A = \frac{SS_{tussen}\,A}{df_{tussen}\,A} = \frac{III - I}{p - 1} = \frac{1381{,}25 - 1380{,}625}{2 - 1} = 0{,}625$$

$$MS_{binnen} = \frac{SS_{binnen}}{df_{binnen}} = \frac{II - V}{n_{til} - pq} = \frac{1489 - 1447{,}5}{40 - 4} = \frac{41{,}5}{36} = 1{,}153$$

$$F_{obs}\,A = \frac{MS_{tussen}\,A}{MS_{binnen}} = \frac{0{,}625}{1{,}153} = 0{,}54$$

Berekening van F_{obs} B

$$F_{obs}\,B = \frac{MS_{tussen}\,B}{MS_{binnen}}$$

$$MS_{tussen}\,B = \frac{SS_{tussen}\,B}{df_{tussen}\,B} = \frac{IV - I}{q - 1} = \frac{1396{,}25 - 1380{,}625}{2 - 1} = \frac{15{,}625}{1}$$

$$= 15{,}625$$

$$MS_{binnen} = \frac{SS_{binnen}}{df_{binnen}} = \frac{II - V}{n_{ttl} - pq} = \frac{1489 - 1447{,}5}{40 - 4} = \frac{41{,}5}{36} = 1{,}153$$

$$F_{obs}\ B = \frac{MS_{tussen}\ B}{MS_{binnen}} = \frac{15{,}625}{1{,}153} = 13{,}55$$

Berekening van F_{obs} AB

$$F_{obs}\ AB = \frac{MS_{tussen}\ AB}{MS_{binnen}}$$

$$MS_{tussen}\ AB = \frac{SS_{tussen}\ AB}{df_{tussen}\ AB} = \frac{V + I - III - IV}{(p - 1) \cdot (q - 1)}$$

$$= \frac{1447{,}5 + 1380{,}625 - 1381{,}25 - 1396{,}25}{(2 - 1) \cdot (2 - 1)} = 50{,}625$$

$$MS_{binnen} = \frac{SS_{binnen}}{df_{binnen}} = \frac{II - V}{n_{ttl} - pq} = \frac{1489 - 1447{,}5}{40 - 4} = \frac{41{,}5}{36} = 1{,}153$$

$$F_{obs}\ AB = \frac{MS_{tussen}\ AB}{MS_{binnen}} = \frac{50{,}625}{1{,}153} = 43{,}9$$

Beslissing

$F_{obs\,A} = \frac{0.625}{1.153} = 0.54$; verwerp $\boldsymbol{H_0}$ niet, geen hoofdeffect.

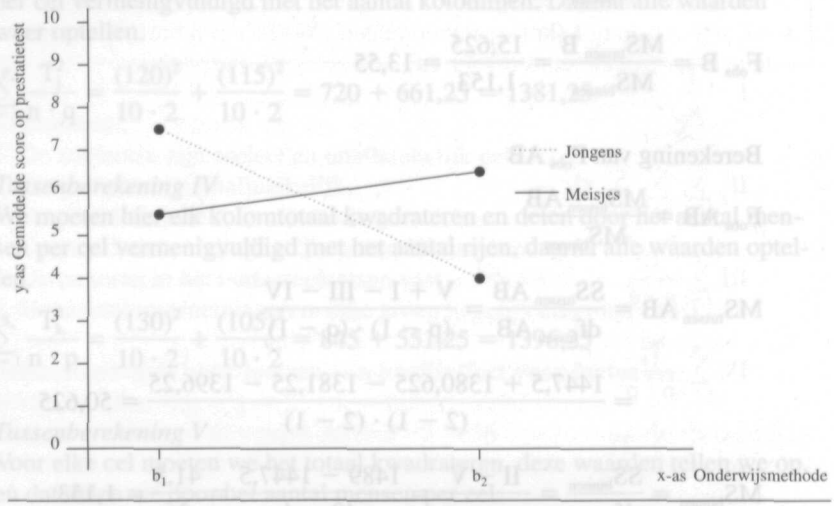
$F_{obs\,B} = \frac{15.625}{1.153} = 13.55$; verwerp $\boldsymbol{H_0}$, hoofdeffect B.

$F_{obs\,AB} = \frac{50.625}{1.153} = 43.90$; verwerp $\boldsymbol{H_0}$, wel interactie!

d Naast het aantonen van een hoofdeffect voor B (onderwijsmethode) zien we dat ook de interactie significant was. Hierdoor moet de interpretatie van het hoofdeffect opgeschort worden omdat eerst de interactie bekeken moet worden. Deze interactie laat zien dat vrouwen baat hebben bij methode b2 en mannen juist bij b1. Het hoofdeffect voor methode kan dus niet zonder meer geïnterpreteerd worden, juist vanwege deze interactie.

e

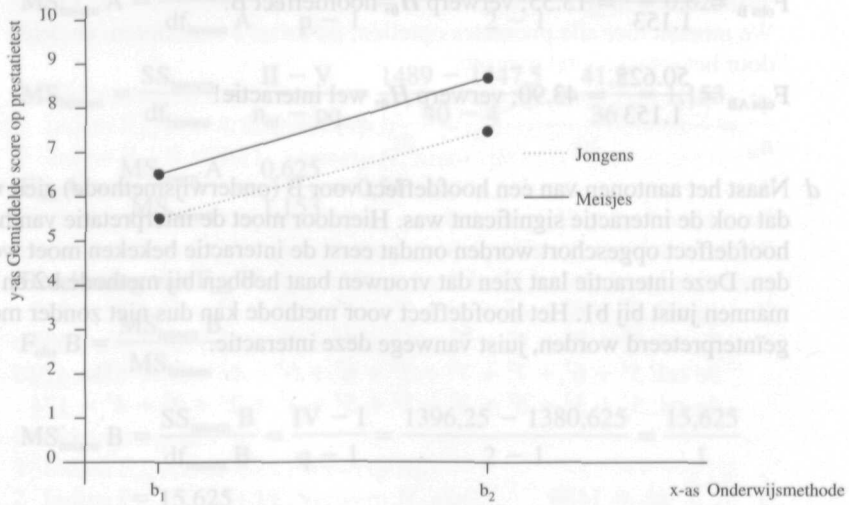
Antwoord 12.1.e Interactieplaatje



f De aanwezigheid van interactie vinden we in het plaatje terug in de kruisende lijnen. Dit is een **antagonistische interactie**, dat wil zeggen: de effecten van de niveaus van de ene factor (B) zijn tegengesteld voor de niveaus van de andere factor.

12.2 Het beste is om eerst een figuur te maken, zodat je in één opslag ziet wat er gaande is.

Antwoord 12.2 Interactieplaatje



We zien nu dat er twee hoofdeffecten zijn, voor geslacht én voor onderwijsmethode. De vrouwen doen het ongeacht de onderwijsmethode beter dan de mannen (rijgemiddelden!). En onderwijsmethode 2 doet het systematisch beter dan onderwijsmethode 1 (kolom-gemiddelden). We behoren het bovenstaande natuurlijk nog wel netjes te toetsen.

13 (Quasi-)experimentele proefopzetten

Bij experimentele proefopzetten worden personen afkomstig uit *dezelfde populatie* aselect toegewezen aan twee (of meer) groepen die elk een verschillende behandeling ondergaan. Als de groepen vervolgens op de criterium (afhankelijke) variabele blijken te verschillen, kunnen deze verschillen worden toegeschreven aan de behandeling. Het experiment is aldus een eenvoudige en logisch dwingende manier om conclusies over causaliteit (oorzaak-gevolgrelaties) te trekken.

Essentiële kenmerken van de experimentele methode zijn:

1 de onafhankelijke variabele wordt *direct* door de onderzoeker gemanipuleerd. Andere veelvuldig gebruikte termen voor de onafhankelijke of gemanipuleerde variabele zijn: predictorvariabele, type behandeling ('treatment') en oorzaak.

2 de toewijzing van personen aan groepen geschiedt *aselect*, dat wil zeggen: iedere persoon heeft een gelijke kans om in de controlegroep (de groep zonder behandeling) of in de experimentele groep (de groep met behandeling) terecht te komen. Het voordeel van aselecte toewijzing is dat verschillen in het begin tussen de groepen (zgn. 'initiële verschillen') en/of verschillen met betrekking tot de gevoeligheid voor de behandeling minder zullen voorkomen.

3 indien goed opgezet en uitgevoerd, is in het experiment sprake van directe *controle* over 'storende' variabelen. Dit betekent onder meer dat *andere* dan de gemanipuleerde (veronderstelde) oorzaakvariabele die eventuele verschillen tussen de groepen zouden kunnen verklaren, worden *uitgeschakeld* (eliminatie van het derde variabeleprobleem).

Een **controlegroep** is nodig om er de resultaten van een experimentele groep tegen af te zetten. De controlegroep vertegenwoordigt het **nulniveau**, dat is het niveau van de experimentele groep als het experiment (de behandeling) er niet geweest was. Zonder controlegroep is er, ook letterlijk, bijna niets onder controle. Vrijwel alle bedreigingen van de interne validiteit (vergelijk hoofdstuk 16) zijn aanwezig.

13.1 Voorbeelden van zuivere experimentele proefopzetten

Zeer bekende (zgn. een- of enkelvoudige) experimentele proefopzetten zijn:

1 Proefopzet met controlegroep, voor- én nameting ('pretest-posttest control group design'), genoteerd als:

(M) R	O_1	X	O_3	Experimentele groep
(M) R	O_2		O_4	Controlegroep

M = 'matchen' (dat al dan niet kan plaatsvinden)
R = aselect toewijzen

X = experimentele behandeling
O_1, O_2 = voormeting
O_3, O_4 = nameting

In deze opzet zijn alle bronnen van interne invaliditeit gecontroleerd; ze vormen dan ook geen bedreiging voor het toeschrijven van verschillen tussen O_3 en O_4 aan X. Bovendien kan de grootte van vele invloeden geschat worden.

2 Proefopzet met controlegroep en alleen nameting ('posttest only control group design'), genoteerd als:

(M)	R	X	O_1
(M)	R		O_2

Alle bronnen van interne invaliditeit zijn gecontroleerd; test-hertesteffect en verval van het meetinstrument ('instrument decay', vgl. hoofdstuk 16) treden zelfs niet op. De grootte van vele invaliditeitsbronnen kan evenwel niet geëvalueerd worden.

13.2 Voorbeelden van quasi-experimenten

Behalve van 'zuivere' experimenten wordt in sociaal-wetenschappelijk onderzoek ten minste even vaak gebruikgemaakt van **quasi-experimentele** proefopzetten. Ook in deze opzetten is er sprake van door de onderzoeker gemanipuleerde variabelen en worden de groepen vergeleken op een criteriumvariabele. Het verschil met 'zuivere' experimenten is dat er *geen aselecte toewijzing* van personen aan groepen plaatsvindt. De proefopzet wordt daardoor aanzienlijk verzwakt:

- Het is (veel) moeilijker om causale relaties vast te stellen.
- De mogelijkheid om conclusies te trekken, hangt vaak af van het patroon van de uitkomsten.

Een belangrijke categorie quasi-experimenten wordt gevormd door de 'proefopzetten met een niet-gelijkwaardige controlegroep' ('nonequivalent control group designs'). Vaak zijn personen ingedeeld in bestaande groepen die niet opgesplitst kunnen worden ten behoeve van een experiment: schoolklassen, bedrijfsafdelingen, clubs en verenigingen en dergelijke. Denk aan een vergelijking van de prestatie tussen de leden van twee judoclubs, waarvan de ene wordt getraind met een nieuw ontwikkelde methode en de andere de standaardmethode volgt. Of denk aan een vergelijking van de gymnastiekprestaties van leerlingen uit twee brugklassen, waarbij de ene klas gedurende een jaar 5 uur lichamelijke opvoeding per week krijgt en de andere klas (het standaardprogramma van) 3 uur per week. Als aan het begin en het eind van het schooljaar metingen plaatsvinden, is sprake van een:

3 Proefopzet met niet-gelijkwaardige groepen, met voor- en nameting ('pretest-posttest nonequivalent control group design'), genoteerd als:

O_1	X	O_3	Experimentele groep
O_2		O_4	Controlegroep

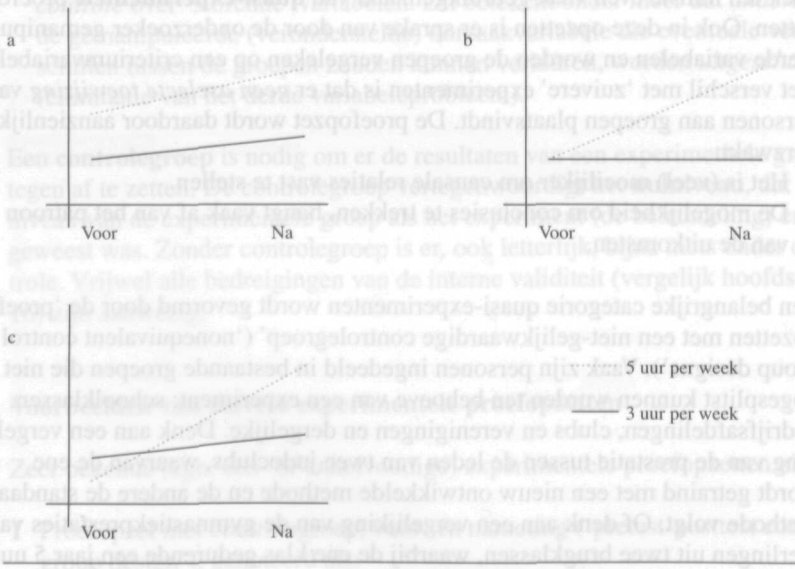
------- = groepen niet gelijkwaardig

Een dergelijke opzet kan gebukt gaan onder (nieuwe) bedreigingen van de interne validiteit door:

- De interactie van selectie en rijping ('maturation'): dit doet zich voor als de te vergelijken groepen (in dit geval de twee brugklassen) verschillend rijpen op een variabele – denk aan de (mate van) fysieke groei en ontwikkeling – die gerelateerd is aan de criteriumvariabele (= de gymnastiekprestaties).
- Lokale externe gebeurtenissen ('local history'): hiervan is sprake als gebeurtenissen gerelateerd aan de criteriumvariabele optreden voor één van de groepen, maar niet voor de andere. Het is bijvoorbeeld denkbaar dat in de ene brugklas door de leerlingen veel meer aan buitenschoolse sportactiviteiten (bijvoorbeeld voetbal, atletiek, turnen) wordt gedaan dan in de andere.

Daarnaast hangt de interpretatie van een quasi-experiment af van (het patroon van) de uitkomsten. Ter illustratie geven we in onderstaande figuren 3 hypothetische resultaten van het gymnastiekonderzoek:

Figuur 13.1



De resultaten in figuur A zijn niet te interpreteren, omdat het verschil tussen de brugklassen op de nameting het gevolg kan zijn van de interactie tussen selectie en rijping.

De resultaten in figuur B zijn waarschijnlijk niet vertekend door een selectie-effect en de resultaten in figuur C vormen de meest overtuigende illustratie van een werkelijk bestaand behandelingseffect.
Toch blijft in alle drie de gevallen de bedreiging van (de invloed van) lokale externe gebeurtenissen ('local history') aanwezig!

Het gymnastiekonderzoek zou ook kunnen worden uitgevoerd zonder voormeting. In dat geval is sprake van een:

4 proefopzet met niet-gelijkwaardige controlegroep en uitsluitend een nameting ('posttest only nonequivalent control group design'), genoteerd als:

X	O	Experimentele groep
	O	Controlegroep

Deze opzet is zeer moeilijk interpreteerbaar: verschillen op de nameting kunnen immers altijd het gevolg zijn van al bestaande verschillen tussen de brugklassen ten gevolge van selectie.

Ten slotte gaan we nog kort in op een andere belangrijke categorie quasi-experimentele opzetten: de zgn. **onderbroken tijdreeksopzetten** ('interrupted time series designs'). In tegenstelling tot quasi-experimenten van het type 3 en 4, waar het gaat om vergelijkingen tussen groepen, ligt hier de nadruk op zgn. *binnenpersoonsvergelijkingen*. Onderbroken tijdreeksopzetten worden gekenmerkt door een onderbreking van de normale gang van zaken door de introductie van een experimentele behandeling. De onderzoeksvraag luidt: 'veroorzaakt de onderbreking de verwachte verandering in de tijdreeks?' Een voorbeeld:

1 eenvoudige onderbroken tijdreeksopzet ('simple interrupted time series design'), bij één of meer personen, genoteerd als:

O O O O	{ X O O O }
A	B

Dit is een A-B-opzet: A = O O O O, de metingen geven de beginsituatie aan (de nullijn) en B = X O O O. Deze metingen geven aan wat voor effect de behandeling heeft op verschillende tijdstippen.
Met deze opzet kan de nullijn (het 'nulniveau') bepaald worden en kan ook het doorwerken van de behandeling worden vastgesteld. Dit soort eenvoudige tijdreeksopzetten is onderhevig aan vele bedreigingen van de interne validiteit (vergelijk hoofdstuk 16), zoals: overdrachtseffecten ('diffusion'), externe gebeurtenissen ('history'), rijping ('maturation'), verval van het meetinstrument ('instrument decay), test-hertesteffect en statistische regressie.

Om daaraan te ontkomen wordt ook wel gebruikgemaakt van een:

2 onderbroken tijdreeksopzet met omkering ('interrupted time series with reversal'), de A-B-A- en A-B-A-B-opzetten, genoteerd als:

O O { X O } O O	A-B-A	opzet: A = { O O } en B = { X O }
O O { X O } O O { X O }	A-B-A-B	opzet: idem, + B aan het eind

Bij de eerste reeks A-metingen wordt de nullijn vastgesteld. Daarna wordt bij B de behandeling ingevoerd. Bij de tweede reeks A-metingen is de behandeling gestopt. Daarna wordt eventueel bij de tweede reeks B-metingen de behandeling heringevoerd.
Bij goede uitkomsten zijn vele bedreigingen van de interne validiteit onder controle en kunnen op verantwoorde wijze causale conclusies worden getrokken.

Van tijdreeksopzetten bestaan vele varianten, waaronder de **complexe onderbroken tijdreeksopzetten** ('complex (multiple treatment) interrupted time series designs') van het type A-B-C-B, waarin B de relevante behandeling is (de veronderstelde oorzakelijke factor) en C een 'concurrerende', als irrelevant beschouwde behandeling die men als verklaring wil uitsluiten.
Bepaalde tijdreeksvarianten voor één persoon worden ook wel aangeduid met de termen 'single subject designs' en 'N = 1 – studies'.

Opgaven

13.1 Uit verschillende onderzoekingen blijkt dat het kijken naar tv-programma's waarin (veel) geweld voorkomt een negatieve invloed heeft op kinderen. Een onderzoekster, die dit een kwalijke zaak vond, heeft een lesprogramma ontworpen. Dit lesprogramma moet kinderen ertoe brengen anders tegen geweld op tv aan te kijken.
Ze wil nu uitzoeken of haar programma effect heeft. Hiertoe heeft ze de medewerking gevraagd van vier groepen 8 van vergelijkbare basisscholen in Almelo. Deze groepen wijst ze aselect toe aan vier condities in een door haar uit te voeren experiment, dat er als volgt uitziet:
Twee groepen krijgen een vragenlijst voorgelegd. Deze vragenlijst moet antwoord geven op de vraag hoe gewelddadig de kinderen de programma's vinden die op dat moment worden uitgezonden. Daarna krijgt één van deze twee groepen, én de derde groep, drie maanden lang één keer in de week les in het 'kritisch bekijken' van geweldprogramma's. Aan het einde van de periode van drie maanden krijgen allevier de groepen de(zelfde) vragenlijst te beantwoorden.
Aan het eind van het experiment blijken de antwoorden op de vragen van groep 1 en 2 bij de eerste afname en de antwoorden van groep 3 en 4 bij de tweede afname niet significant te verschillen. De tweede afname bij groep 1 en 2 levert wel verschillende scores op: de 'behandelde' kinderen staan duidelijk meer gewapend tegenover geweld op tv.
Bestudeer de bovenstaande onderzoeksopzet nauwkeurig en beantwoord de volgende vragen.

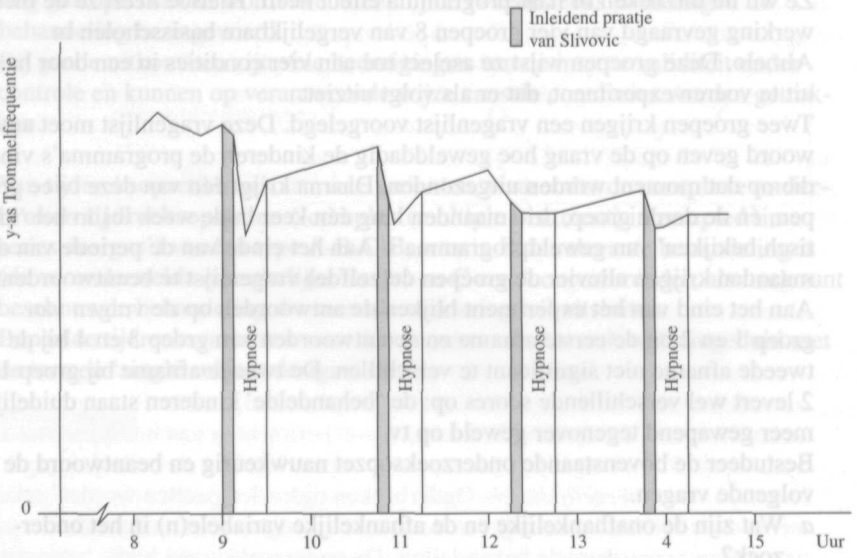
a Wat zijn de onafhankelijke en de afhankelijke variabele(n) in het onderzoek?
b Geef het onderzoeksontwerp schematisch weer.
c Hoe heet dit onderzoeksontwerp? Noem de voor- en nadelen van het gebruik van het ontwerp in dit experiment.
d De onderzoekster is teleurgesteld over het resultaat. Naar haar mening heeft het lesprogramma geen effect gehad. Heeft ze gelijk?
e Geef een verklaring voor de resultaten.

13.2 Een psychiatrische patiënt uit de kliniek van prof. Slivovic heeft zich aangewend de hele dag met zijn vingers te trommelen op alles wat los en vast zit. Slivovic denkt de man te kunnen genezen met behulp van hypnose. De volgende opzet wordt bedacht:

- De patiënt wordt een uur lang stiekem geregistreerd (bandopname van het getrommel).
- Daarna komt Slivovic zijn kamer binnen en begint een geruststellend praatje van 5 minuten.
- Hierna begint Slivovic de hypnose. Dit duurt een half uur. Dan gaat Slivovic weer weg.
- De patiënt wordt een uur lang stiekem geregistreerd.
- Daarna komt Slivovic zijn kamer binnen en ..., enzovoort.

Het resultaat van één dag behandeling is in de grafiek op de volgende bladzijde weergegeven.
De bandopname van het getrommel wordt na ieder kwartier gescoord op frequentie; dit levert de punten in de grafiek op.
Bestudeer bovenstaande onderzoeksopzet en de bijbehorende grafiek nauwkeurig en beantwoord de volgende vragen.

Figuur 13.2



a Hoe heet deze proefopzet?
b Mogelijke bedreigingen van de interne validiteit zoals *externe gebeurtenissen* ('history') en *testgewenning* ('testing') lijken in dit onderzoek geen rol te spelen. Waaruit blijkt dit?
c Hoe zou u de gebruikte opzet noteren?
ABA; ABABA; ABABABABA; AAABBBCCC; ABCABCABC; BACBACBAC.
d Geef twee mogelijke verklaringen voor het verschijnsel dat de patiënt tussen 14:30 en 15:00 minder trommelt dan tussen 8:30 en 9:00.
e Slivovic concludeert dat zijn hypnose effectief is. Heeft hij gelijk?

Uitwerkingen

13.1 *a* De onafhankelijke variabele is het al dan niet volgen van het nieuwe lesprogramma ('kritisch bekijken van geweld op tv').
De afhankelijke variabelen zijn de door de leerling gegeven antwoorden (scores) op de vragenlijst.

b

Onderzoeksopzet					*Betekenis van de tekens*
		T1		**T2**	T1/T2 = meettijdstip 1, resp. 2
1	R	O_1	X	O_2	R = random (aselect)
2	R	O_3		O_4	O = observatie/meting
3	R		X	O_5	X = experimentele behandeling (treatment)
4	R			O_6	

c Dit onderzoeksontwerp staat bekend onder de naam 'Solomon four-group design' ('viergroepenopzet van Solomon'). Voordeel van deze opzet is dat alle kenmerken van het 'zuivere' experiment behouden blijven en dat het effect van het 'ondergaan' van een voormeting kan worden *onderscheiden* van het effect van de gemanipuleerde variabele of 'behandeling' (in dit geval het nieuwe lesprogramma), m.a.w.: de *interactie* tussen voormeting en behandeling kan op deze wijze worden vastgesteld.
Nadeel is dat de opzet nogal omslachtig is, meer werk met zich meebrengt en in de praktijk al gauw leidt tot het gebruik van kleinere groepen.

d en e
Het feit dat de groepen 1 en 2 op de voormeting niet (significant) van elkaar verschillen is gunstig: na aselecte toewijzing van personen 'behoren' deze groepen immers op T1 niet te verschillen op de criteriumvariabele.
Het verschijnsel dat er geen verschil optreedt tussen groep 3 en 4 op de nameting (T2) maar wél tussen groep 1 en 2 (dus O_5 en O_6 verschillen niet, O_2 en O_4 wel) suggereert dat er *interactie* heeft plaatsgevonden tussen voormeting en behandeling.
Met andere woorden: als je kinderen bewust wilt maken van geweld op tv, moet je ze eerst een vragenlijst voorleggen en daarna de lessen geven. Dìe *specifieke combinatie* van voormeting en behandeling heeft blijkbaar het beoogde effect.

13.2 *a* Het gaat hier om een quasi-experiment in de vorm van een 'N = 1 − studie of -opzet' ('single subject design'). Een *quasi*-experiment, omdat er wel sprake is van manipulatie van variabele(n) en vergelijking van prestaties op een criteriumvariabele, maar geen sprake is van aselecte toewijzing (zie Inleiding). Nader gespecificeerd gaat het in dit geval om een binnenpersoonsopzet en wel een *onderbroken tijdreeks met omkering* ('interrupted time series with reversal').

b Externe gebeurtenissen ('history') lijken hier geen rol te spelen omdat het (quasi-) experiment zich op één dag afspeelt. Testeffecten ('testing') spelen hier geen rol omdat het getrommel van de proefpersoon stiekem geregistreerd wordt.

c ABCABCABC

d **1** De patiënt wordt moe
2 De behandeling heeft effect

N.B. De (feitelijke) behandeling bestaat uit twee delen: praatje, hypnose. Uit de grafiek blijkt dat het getrommel het sterkst daalt tijdens het inleidende praatje van de therapeut.

e De hypnose (de veronderstelde oorzakelijke variabele) als zodanig niet, het inleidende praatje wel!

14 Complexe experimentele opzetten: Factoriële en gemengde (mixed) proefopzetten en soorten interactie

Men spreekt van experimentele proefopzetten indien personen afkomstig uit *dezelfde populatie* aselect worden toegewezen aan 2 (of meer) groepen die elk een *verschillende behandeling* ontvangen. Indien de groepen daarna op de criteriumvariabele blijken te verschillen, kunnen deze verschillen worden toegeschreven aan de behandeling.
Ieder verschijnsel wordt echter bijna altijd door meer dan één oorzaak beïnvloed. Proefopzetten waarin het effect van meer dan één oorzaak of factor gemeten wordt, noemt men *complexe* (experimentele) opzetten. Indien *beide* factoren experimenteel of door de onderzoeker gemanipuleerd zijn, spreekt men van factoriële proefopzetten. Denk aan een onderzoek op de basisschool, waarin zowel het effect van een nieuwe rekenmethode (factor A) als het effect van het al dan niet inschakelen van de computer bij het rekenonderwijs (factor B) op de rekenvaardigheid (= criteriumvariabele) wordt nagegaan.
Men spreekt van een *gecombineerde* of *gemengde* opzet ('mixed design') wanneer personen afkomstig uit 2 (of meer) afzonderlijke en niet-overlappende populaties aan de behandeling worden toegewezen.
Een voorbeeld van een gemengde opzet in de meest elementaire vorm is dat personen eerst naar sekse en vervolgens aselect aan een behandeling of geenbehandeling conditie (bijvoorbeeld wel of geen nieuwe rekenmethode) worden toegewezen.
Op deze manier is het bijvoorbeeld mogelijk na te gaan of de eventuele effecten van een behandeling *verschillend* uitvallen voor jongens en meisjes. Men spreekt van een *gemengde* opzet omdat de factor 'sekse' geen experimentele of gemanipuleerde variabele is, maar een *classificerende* (of *correlationele*) variabele. De tweede factor in een gemengde opzet is wel experimenteel (in dit geval het al dan niet volgen van de nieuwe rekenmethode). Met behulp van gemengde opzetten kan tegelijkertijd correlationele en experimentele informatie verkregen worden.
Zowel bij gemengde als bij factoriële opzetten speelt het begrip **interactie** een belangrijke rol. Men spreekt van **interactie** als de invloeden van factoren zodanig combineren of uitwerken dat ze samen een ander resultaat opleveren dan elk voor zich of dan hun optelling.

14.1 Voorbeeld

In het volgende voorbeeld wordt een en ander nader toegelicht:
In een onderzoek bij eerstejaars hbo-studenten werd de effectiviteit nagegaan van *twee verschillende onderwijsmethoden*. Onderwijsmethode b_1 houdt in dat men alleen colleges krijgt; bij methode b_2 krijgt men zowel werkgroep als college. De afhankelijke variabele was de score op een prestatietest. Een tweede onafhankelijke variabele was het *geslacht* van de studenten: 20 mannelijke en

20 vrouwelijke studenten werden aselect over de beide onderwijsmethoden verdeeld. In tabel 14.1 staan de ruwe scores.

Tabel 14.1 Ruwe scores op de variabele 'prestatietest'

vrouwen a_1	vrouwen a_1	mannen a_2	mannen a_2
b_1	b_2	b_1	b_2
7	5	9	5
6	6	7	5
4	7	8	6
6	7	8	3
7	7	6	3
6	6	7	5
5	6	9	4
4	6	6	3
5	7	8	2
5	8	7	4

Voor een eerste, globale indruk van eventuele verschillen tussen de methoden en/of de seksen in deze gemengde proefopzet, vergelijken we de gemiddelden per subgroep, zoals in onderstaande tabel:

Tabel 14.2 Groepsgemiddelden op de prestatietest

		Onderwijsmethode		
		b_1	b_2	Rijgemiddelden
Sekse	Vrouwen a_1	5.5	6.5	6
	Mannen a_2	7.5	4.0	5.75
	Kolomgemiddelden	6.5	5.25	5.875 (totaal gem.)

De tabel laat zien dat de rijgemiddelden (6 en 5.75) nauwelijks van elkaar verschillen. Dit betekent dat er geen hoofdeffect is voor sekse (factor A). Voor factor B (onderwijsmethode) verschillen de niveaus (b_1 en b_2) wel: de kolomgemiddelden zijn respectievelijk 6.5 en 5.25. Uitgaande van dit verschil tussen de kolomgemiddelden, is er aanleiding te veronderstellen dat er wel een hoofdeffect bestaat voor de factor Onderwijsmethode: methode b_1 leidt tot betere prestaties dan methode b_2.

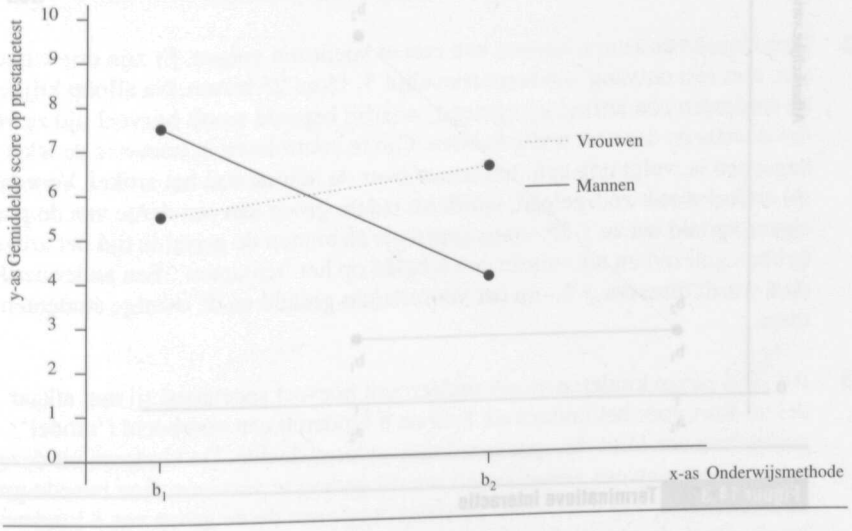
Vergelijking van de celgemiddelden (5.5, 6.5, 7.5 en 4.0) laat echter zien dat er naast een hoofdeffect voor factor B ook een zogenoemd **interactie-effect** aanwezig is. Mannen hebben bij methode b_1 een vrij hoog gemiddelde, maar bij methode b_2 een zeer laag gemiddelde. Zij lijken dus wel baat te hebben bij methode b_1 en relatief slecht te presteren bij methode b_2. De aanwezigheid van een dergelijk interactie-effect *modificeert* dus de betekenis en de interpretatie van het voor factor B gevonden hoofdeffect: het zou immers onverstandig zijn om op basis van deze resultaten zonder meer te stellen dat aangetoond is dat

methode b_1 beter is dan methode b_2. Dat geldt weliswaar voor mannen, maar *niet* voor vrouwen: uit de celgemiddelden blijkt immers dat vrouwen meer baat hebben bij methode b_2!

14.2 Grafische afbeeldingen van interactie-effecten

De aanwezigheid van interactie-effecten kan worden geïllustreerd met behulp van grafische afbeeldingen of 'plaatjes'. Onderstaande figuur bevat zo'n 'plaatje', waarin de hierboven besproken interactie tussen sekse en onderwijsmethode (aangeduid als de interactie sekse × methode) is afgebeeld:

Figuur 14.1



Bij de constructie van zo'n plaatje zet men de afhankelijke variabele (de scores op de prestatietest) op de Y-as en één van de onafhankelijke variabelen op de X-as (hier de factor Onderwijsmethode). De andere onafhankelijke variabele (de factor Sekse) tekent men *in de figuur*, waarbij de twee niveaus (a_1 en a_2, resp. vrouwen en mannen) duidelijk en onderscheidbaar moeten worden aangegeven.

De in de figuur afgebeelde interactie is een voorbeeld van een zgn. **antagonistische** interactie: de niveaus van de ene factor hebben een verschillend (tegengesteld) effect voor elk van de niveaus van de andere factor. Aan de kruisende lijnen in de figuur zien we inderdaad in één oogopslag dat vrouwen baat hebben bij methode b_2 en mannen juist bij b_1.

N.B. In de praktijk zal met behulp van tweewegvariantieanalyse (two-way of 2-factor-ANOVA) moeten worden nagegaan of de in de steekproef gevonden verschillen (waarschijnlijk) ook zo terug te vinden zijn in de populatie. Deze toetsing laten wij hier achterwege. Voor de liefhebbers geven we hier wel de waarden van F_{obs} voor factor A (sekse), factor B (methode) en de interactie AB

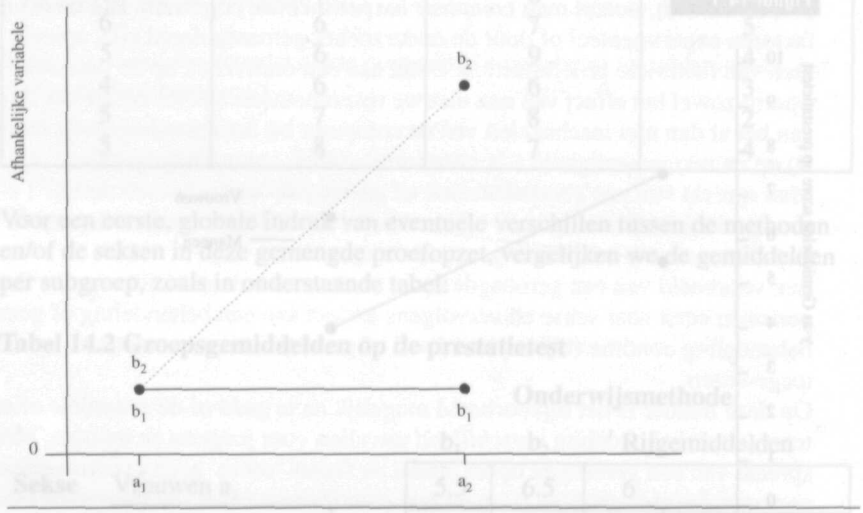
(sekse × methode), nl. resp. 0,54 (niet significant), 13,55 en 43,9 (ga na, zie blz. 245).

Naast antagonistische interactie bestaan er ook andere typerende vormen van interactie. De bekendste daarvan zijn:

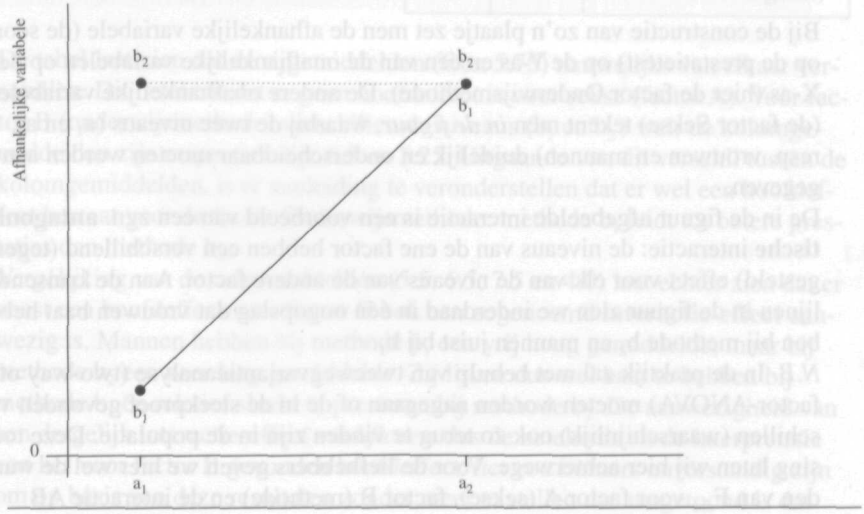
- **catalytische interactie**: twee of meer factoren zijn alleen effectief wanneer ze samen voorkomen
- **terminatieve interactie**: twee factoren hebben elk een effect, maar hun gecombineerd effect is niet groter dan dat van elk apart

Illustraties van beide typen interactie zijn terug te vinden in figuur 14.2 en 14.3.

Figuur 14.2 *Catalytische interactie*



Figuur 14.3 *Terminatieve interactie*



De op de Y-as afgebeelde afhankelijke variabele is het aantal ponden dat men is afgevallen. Factor A is het al dan niet volgen van een strikt afslankingsdieet (a_1 = geen dieet, a_2 = wel dieet) en factor B is het al dan niet drinken van vele glazen water per dag (b_1 = niet drinken, b_2 = wel drinken). De (fictieve) resultaten in de bovenste figuur illustreren een catalytische interactie: alleen de **combinatie** van dieet + veel drinken leidt tot afslanking. Elk van beide afzonderlijk heeft geen effect (zie linksonder in de figuur).
De terminatieve interactie in de onderste figuur duidt daarentegen op (fictieve) resultaten waaruit blijkt dat zowel dieet als veel drinken elk op zich effect hebben, maar het effect wordt **niet** vergroot door beide behandelingen te combineren (zie rechtsboven in de figuur).
N.B. Omdat beide onafhankelijke variabelen ('dieet' en 'veel drinken') door de onderzoeker worden gemanipuleerd, gaat het hier om een **factoriële** proefopzet.

Opgaven

Geef van de onderstaande proefopzetten aan of ze eenvoudig (enkelvoudig) zijn of complex. Geef in het laatste geval tevens aan of er sprake is van een factoriële of een gemengde opzet. Indien het een gecombineerde opzet is, welke zijn dan de *classificerende* (correlationele) variabele(n) en de *experimentele* variabele(n)? Geef ook aan wat de *criterium*variabele is.

14.1 Mannelijke proefpersonen wordt een angsttest afgenomen. Op grond van de testuitslagen worden ze in drie groepen verdeeld. Ieder lid van iedere groep wordt één van de drie mogelijke doseringen van het hormoon adrenaline toegediend. Vervolgens worden dia's vertoond van giftige en afzichtelijke spinnen en slangen. De mate van pupilverwijding wordt gemeten.

14.2 Tweedejaarsstudenten kunnen een cursus snellezen volgen. Er zijn drie cursussen, met een omvang van respectievelijk 5, 15 en 25 lessen. Na afloop krijgen de studenten een artikel voorgelegd, waarbij bepaald wordt hoeveel tijd ze voor het doorlezen daarvan nodig hebben. Om te controleren in hoeverre de tekst begrepen is, volgt nog een 'tentamen' over de inhoud van het artikel. Voordat dit artikel wordt voorgelegd, wordt uit iedere groep aan een derde van de personen verteld dat ze *f* 25,– ontvangen als ze binnen de gestelde tijd het artikel hebben gelezen en ten minste een 6 halen op het 'tentamen'. Een ander derde deel wordt hiervoor *f* 5,– in het vooruitzicht gesteld en de overige studenten niets.

14.3 Bij 32 4-jarige kinderen wordt onderzocht hoeveel speelgoed zij met elkaar delen. Kort voor het onderzoek hebben 8 kinderen een voorbeeld ('model') gezien van een kind dat speelgoed met anderen deelde. De kinderen uit deze groep werden tevens aangespoord om dit gedrag te vertonen. Een tweede groep van 8 kinderen zag alleen het 'modelgedrag', een derde groep van 8 kinderen kreeg alleen de aansporing om dit gedrag te vertonen (in beide gevallen kort voor het onderzoek). De vierde groep van 8 kinderen kreeg geen van beide 'behandelingen'. Observatie tijdens een vrije spelsituatie van één uur gaf de volgende resultaten te zien:

Gemiddeld aantal gedeelde speelgoedvoorwerpen

	Geen voorbeeld	**Wel voorbeeld**
Geen aansporing	2.8	4.5
Wel aansporing	4.8	4.6

a Maak een plaatje van de resultaten van het onderzoek. Zet de niveaus van de factor 'voorbeeld' (model) op de horizontale as.

b Vat de resultaten van het onderzoek, mede aan de hand van de figuur, overzichtelijk samen. Welk type interactie doet zich voor?

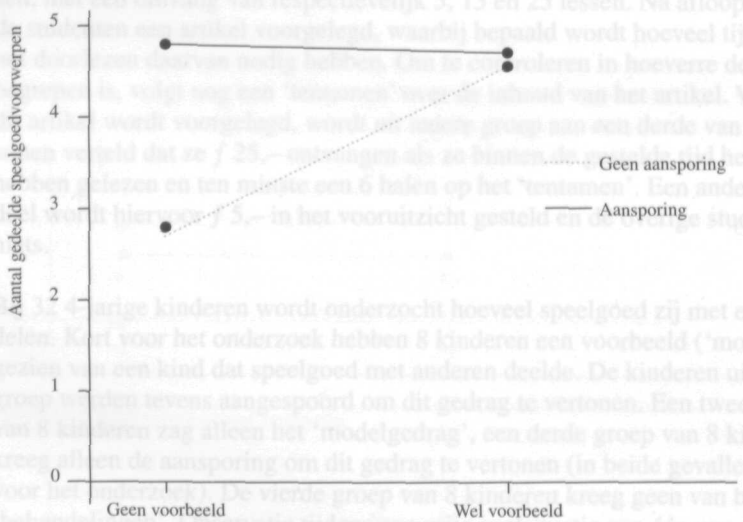
14.4 Een onderzoeker wil bij fobische patiënten het effect van verschillende behandelingen op angstreductie nagaan. Hij vermoedt dat er sprake is van interactie tussen het type fobie en het soort behandeling, zodanig dat personen met agorafobie (vrees voor open ruimten) zullen verschillen van personen met meer specifieke angsten. Daartoe verdeelt hij de personen in twee groepen op basis van type fobie en wijst hij vervolgens leden van elke groep toe aan 4 behandelingscondities: desensitisatie, inzicht, aandacht-placebo en een geen-behandeling controlegroep. Daarna wordt vastgesteld of men de gevreesde situatie beter (met minder angst) durft te benaderen.

Uitwerkingen

14.1 Criteriumvariabele is de pupilverwijding. Er is sprake van een classificerende variabele, want de mannen worden aan de hand van hun angstscore ingedeeld in drie niveaus van angst. De andere onafhankelijke (experimentele) variabele is de dosering adrenaline. Het betreft een gecombineerde opzet: één factor (dosering adrenaline) wordt gemanipuleerd en in het andere geval is er sprake van een classificerende variabele.

14.2 De criteriumvariabelen zijn de 'tijd voor het doorlezen' en het 'tentamen' over de inhoud van het artikel. De twee onafhankelijke variabelen (factoren) zijn beide experimenteel: het aantal lessen snellezen en de (mate van) beloning. Er is hier dus sprake van een complexe opzet en wel een factoriële proefopzet.

14.3 **Antwoord 14.3.a**



b De verschillen tussen de rijgemiddelden (3,65 en 4,7) en de kolomgemiddelden (3,8 en 4,55) zouden kunnen wijzen op de aanwezigheid van hoofdeffecten voor respectievelijk de factoren 'aansporing' en 'voorbeeld'. Dit betekent dat zowel 'het gezien hebben van een voorbeeld' als het 'aangespoord zijn' het aantal gedeelde speelgoedvoorwerpen verhoogt (eigenlijk moet dit worden getoetst met behulp van tweewegvariantieanalyse). De **terminatieve** interactie in de figuur uit a. laat echter zien dat een *combinatie* van de twee 'behandelingen' (voorbeeld + aansporing) *niet* effectiever is dan elk afzonderlijk. Hieruit blijkt, als steeds, dat bij aanwezigheid van interactie de interpretatie van het hoofdeffect nader moet worden bezien! De criteriumvariabele is het aantal gedeelde speelgoedvoorwerpen. Verder is er sprake van een complexe opzet met twee experimentele onafhankelijke variabelen (factoren), m.a.w. een factoriële proefopzet.

14.4 De criteriumvariabele is angstreductie, d.w.z. (de mate van) vermindering van de angst. Soort behandeling (behandelingsconditie) is een gemanipuleerde of experimentele (onafhankelijke) variabele, terwijl de andere onafhankelijke variabele, type fobie, een classificerende (correlationele) variabele is. Er is dus sprake van een complexe opzet met een experimentele en een classificerende variabele, m.a.w.: het gaat hier om een gemengde opzet ('mixed design').

15 Operationaliseren, validiteit en betrouwbaarheid

15.1 Operationaliseren[1]

Operationaliseren speelt een grote rol in elk (sociaal-wetenschappelijk) onderzoek. Door middel van de operationalisatie vertaalt men de *concepten* uit het geformuleerde onderzoeksprobleem in een *empirisch onderzoekbare vorm*. Dergelijke concepten worden ook vaak aangeduid met termen als: *construct*, *begrip-zoals-bedoeld* en dergelijke.

Stel dat het in een onderzoek gaat om het concept (het begrip-zoals-bedoeld, het construct) 'bejaard'. Met deze omschrijving kunnen we geen onderzoek doen: het is te vaag en iedereen heeft een eigen opvatting over wat bejaard is en/of wanneer iemand als zodanig kan worden aangeduid. Als we aangeven dat bejaard een omschrijving is van vijfenzestig jaar of ouder zijn, kunnen we personen op dit kenmerk gaan meten. Van het concept bejaard hebben we nu een **operationele definitie** gegeven; men zegt ook wel dat bij bejaard als begrip-zoals-bedoeld het *begrip-zoals-bepaald* leeftijd van 65 jaar of ouder hoort.

Operationaliseren vindt dus plaats door gebruik te maken van operationele definities. Een zeer bekend voorbeeld is: *intelligentie is wat de intelligentietest meet*. Nauwkeuriger geformuleerd: *iemands intelligentie is het geheel van antwoorden bij het afnemen van de intelligentietest*.

Op deze wijze krijgt het concept intelligentie inderdaad een vertaling in empirisch waarneembare verschijnselen, te weten de antwoorden die worden gegeven op de betreffende test. Dit betekent overigens *niet* dat we niet meer van intelligentie zouden mogen spreken wanneer een *andere* test wordt afgenomen!

In de onderzoekspraktijk wordt juist vaak gewerkt met verschillende operationele definities voor één en hetzelfde concept. Dit heeft als consequentie dat een operationele definitie meestal niet meer precies de volledige betekenisinhoud van het concept zal dekken. Operationele definities zijn slechts *partieel*: conceptinhoud en definitie-inhoud stemmen niet noodzakelijk overeen, met andere woorden: een bepaald concept behoudt meestal een bepaalde *meerbetekenis* boven zijn operationalisatie.

Uiteraard moet geprobeerd worden zo goed mogelijk te operationaliseren. Een voorbeeld kan dit verduidelijken.

Voorbeeld 15.1: Operationaliseren

Bij de operationalisatie van het concept 'sociaal milieu van een gezin' is het in het algemeen onvoldoende om alleen maar te kijken naar *beroep man* (definitie-inhoud of **indicator** 1), maar dient ook rekening gehouden te worden met *beroep vrouw* (indicator 2), *inkomen man* (indicator 3), *inkomen vrouw* (indicator 4), *opleidingsniveau man* (indicator 5), *opleidingsniveau vrouw* (indicator 6) enzovoort. Deze indicatoren zouden ieder afzonderlijk een vrij slechte operationele definitie voor sociaal milieu van een gezin (kunnen) opleveren. Als men alleen met inkomen rekening houdt, dan zouden bijvoorbeeld dominees in een te lage milieuklasse terechtkomen. Wordt alleen op beroep gelet aan de hand van een beroeps-prestigeclassi-

[1] Gedeeltelijk gebaseerd op J.H.L. Oud, 'Sociaal-wetenschappelijk onderzoek' (1975; blz. 30-31).

ficatie, dan komt minstens één bekende Nederlandse kruidenier in een te lage milieuklasse terecht.

Combineert men daarentegen verschillende indicatoren in een **index** om te bepalen tot welke klasse iemand behoort, dan is het resultaat een operationele definitie van hogere kwaliteit.

Bij operationaliseren ligt de nadruk op het empirisch waarneembaar *maken*. De verschijnselen die in de werkelijkheid met het concept corresponderen, zijn namelijk niet op zichzelf of vanzelf empirisch waarneembaar. In dat geval zou een operationele definitie immers niet noodzakelijk zijn. Er moeten juist allerlei handelingen (**operaties**) uitgevoerd worden, eventueel met behulp van instrumenten, om ervoor te zorgen dat de verschijnselen zich in een empirisch waarneembare vorm manifesteren. Zo hanteert een arts een stethoscoop om een toestand van ziekte manifest te maken. Een sociaal-wetenschapper moet het normale gedragspatroon van mensen onderbreken en hun een test, vragenlijst of interview voorleggen, of hij registreert wat hij ziet met behulp van een observatiesysteem. Niet zelden moet hij deze instrumenten (test, vragenlijst, interview, observatiesysteem) van tevoren zelf construeren. Met andere woorden: operationalisering betreft niet alleen het (passief) definiëren van empirisch waarneembare verschijnselen als indicator voor een concept, maar het omvat ook actieve handelingen en daarbij te hanteren meetinstrumenten.
Twee begrippen die steeds weer terugkeren wanneer het gaat om de kwaliteit van een operationalisatie of operationele definitie zijn **validiteit** en **betrouwbaarheid**. Op elk van deze kwaliteitscriteria gaan we in het volgende nader in.

15.2 Validiteit

Een meetresultaat wordt méér valide genoemd naarmate de scores de bedoelde verschillen beter weerspiegelen, met andere woorden: naarmate de afstand tussen begrip-zoals-bedoeld (construct) en begrip-zoals-bepaald (operationalisatie) *kleiner* wordt.
De validiteit van een test heeft betrekking op de mate waarin de test meet wat hij beoogt te meten. Zo kan men zich afvragen of een intelligentietest inderdaad intelligentie (bijvoorbeeld te definiëren als 'probleemoplossend vermogen in nieuwe situaties') meet.
De vraag naar de validiteit van de CITO-toets heeft betrekking op het feit of deze toets inderdaad de 'geschiktheid voor (of: het te verwachten succes op) de middelbare school' meet. In zo'n geval spreekt men van de (mate van) *predictieve validiteit* van de CITO-toets: hoe goed voorspellen de *nu* behaalde CITO-scores *latere* resultaten op de middelbare school?
Validiteit wordt vastgesteld door middel van relaties met andere observaties of metingen bij dezelfde personen.

15.2.1 Soorten validiteit

1 *Inhoudsvaliditeit*: Het meetinstrument moet het bedoelde concept in zijn verschillende aspecten goed weerspiegelen, terwijl er geen overbodige of irrelevante zaken worden gemeten. In hoeverre komen de testinhoud en de

beoogde inhoud overeen? Bijvoorbeeld: vormen de vragen voor een proefwerk wiskunde een goede (representatieve) afspiegeling van de voor dit onderdeel te bestuderen stof?

2 *Begrips- of constructvaliditeit*: Meten we wel het theoretisch begrip dat gemeten moet worden? Men gaat ervan uit dat het begrip onderdeel uitmaakt van of past in een groter theoretisch kader. Op basis daarvan doet men een aantal voorspellingen over hoe het te meten begrip met andere variabelen zal samenhangen. Hoe beter de voorspellingen kloppen, des te sterker is de begripsvaliditeit van het meetinstrument. Bijvoorbeeld, bij de ontwikkeling van een test voor 'democratisch' leerkrachtgedrag verwacht de onderzoeker dat democratische leerkrachten meer inspraak en kritiek van leerlingen accepteren en aanmoedigen dan niet-democratische leerkrachten. Door voor hoog- en laagscorende leerkrachten op de test de tijdens een aantal lessen gevonden frequenties van inspraak en/of kritiek te vergelijken, kan hij de constructvaliditeit van zijn democratietest vaststellen. Daarnaast kan hij nagaan of de scores op de nieuwe test positief correleren met de scores op andere tests waarvan verondersteld mag worden dat ze democratisch leiderschap meten.

3 *Criteriumvaliditeit*: Dit type validiteit geeft aan in hoeverre de testscores samenhang vertonen met de scores van gerelateerde variabelen (criteria). Bijvoorbeeld: is er een (substantieel) verband tussen de resultaten van een bij leerlingen afgenomen taalvaardigheidstoets en de door deze kinderen behaalde cijfers voor dictee, opstel en tekstverklaring?
Criteriumvaliditeit wordt soms nader gespecificeerd in:
- *concurrente validiteit*: Kenmerk van een meetinstrument dat de scores (goed) 'voorspelt' op een criteriumvariabele die *gelijktijdig* (op hetzelfde moment) gemeten wordt (vgl. bovenstaand voorbeeld over taalvaardigheid).
- *predictieve validiteit*: Kenmerk van een meetinstrument dat de scores (goed) voorspelt van een criteriumvariabele die *in de toekomst* gemeten wordt (vgl. het voorbeeld van de CITO-toets).

4 *Convergente en discriminante validiteit*:
- *convergente validiteit*: Uiteenlopende meetinstrumenten voor hetzelfde begrip moeten *positief* met elkaar correleren. Als van een groep kinderen het angstniveau wordt gemeten met behulp van twee methoden, observatie en zelfbeschrijving (via interview en/of vragenlijst), dan is er sprake van convergente validiteit als de resultaten van beide meetmethoden hoog correleren.
- *discriminante validiteit*: Stel dat de observatie niet alleen wordt gebruikt ter meting van het angstniveau maar ook voor het bepalen van de mate van vijandigheid van de kinderen. De correlatie tussen de scores op deze verschillende kenmerken moet dan *lager* zijn dan de correlatie tussen de scores op verschillende meetinstrumenten (observatie en zelfbeschrijving) die hetzelfde begrip meten (bijvoorbeeld angstniveau). Als dat zo is, is er sprake van discriminante validiteit.

De term validiteit wordt óók gebruikt om de kwaliteit en de reikwijdte van de conclusies uit een geheel onderzoeksontwerp aan te duiden. Men maakt daarbij onderscheid tussen:

a *Interne validiteit*: Hier gaat het om de kwaliteit van de (causale) conclusies uit een geheel onderzoeksontwerp, nader te specificeren in:
- *statistische (conclusie)validiteit*: is de conclusie over de aan- of afwezigheid van een statistisch verband tussen twee variabelen juist?
- *causale interpretatievaliditeit*: is het verband dat gevonden is tussen twee variabelen wel een oorzakelijk (causaal) verband?

b *Externe validiteit*: Hier gaat het om de reikwijdte van de onderzoeksresultaten, met andere woorden: in hoeverre zijn de conclusies generaliseerbaar naar: 1 algemene(re) begrippen, 2 populaties en 3 andere omgevingen dan waar het onderzoek gedaan is.
Externe validiteit wordt soms nader gespecificeerd in:
- *Populatievaliditeit*: Betreft de vraag of de verkregen resultaten kunnen worden gegeneraliseerd naar de doelpopulatie.
- *Ecologische validiteit*: Betreft de vraag of de verkregen resultaten gegeneraliseerd kunnen worden naar vergelijkbare situaties (vooral ook *buiten* de test- c.q. onderzoeks-situatie).

Aan de interne validiteit (in het bijzonder de causale interpretatievaliditeit) en de externe validiteit – met name aan de bedreigingen daarvan – wordt in de hoofdstukken 16 en 17 uitgebreid aandacht besteed.

15.3 Betrouwbaarheid

Betrouwbaarheid ('reliability') is de tweede kwaliteitseis waaraan een indicator c.q. meetinstrument moet voldoen. Betrouwbaarheid is de mate waarin herhaalde toepassing van dezelfde indicator (meetinstrument) bij hetzelfde object onder dezelfde omstandigheden overeenkomende resultaten oplevert. Men zegt ook wel dat de betrouwbaarheid (van een meting) aangeeft in hoeverre een bepaalde waarneming een 'ware score' oplevert en vrij is van toevalsfouten: dit blijkt uit een van meting tot meting *stabiel* meetresultaat. Een onbetrouwbare score bevat *meetfouten* en is beïnvloed door irrelevante of toevalsfactoren.
Bronnen van deze toevalsfactoren zijn:
- meetinstrumenten
- onderzoekers
- proefpersonen
- situaties

De invloed van één of meerdere van deze toevalsfactoren verklaart waarom herhaalde metingen aan hetzelfde object verschillende waarden kunnen opleveren en er dus sprake is van onbetrouwbaarheid. Een voorbeeld over het meten van lengte kan dit verduidelijken.

Voorbeeld 15.2: Betrouwbaarheid

Lengte kan worden gemeten door gebruik te maken van een precisie-instrument, zoals dat door artsen wordt gebruikt (*methode 1*). Het gebruik van dit apparaat omvat specifieke en zeer nauwkeurig omschreven procedures (blote voeten, kaarsrecht staan, hielen strak tegen de meetlat enzovoort). Als de lengte van persoon A in een week op 5 verschillende momenten door 5 verschillende mensen, die allen dezelfde instructie krijgen, met behulp van zo'n precisie-instrument wordt gemeten, zullen de meetresultaten nagenoeg identiek zijn. In zo'n geval is sprake van een hoge betrouwbaarheid. Als we de lengte van persoon B, die in werkelijkheid 2 centimeter langer is dan persoon A, op dezelfde manier zouden meten, zal het verschil in lengte tussen beide personen dan ook in de meetresultaten terug te vinden zijn.

Stel dat de lengte van dezelfde twee personen, A en B, nog eens 5 keer door 5 mensen zou worden gemeten, maar nu met behulp van een plastic liniaal van 30 cm (*methode 2*). In zo'n geval is het te verwachten dat de 5 uitkomsten voor beide personen meer variatie vertonen, vanwege het gebrek aan precisie dat gepaard gaat met het gebruik van de plastic liniaal. Methode 2 is duidelijk *minder betrouwbaar* dan methode 1. Als gevolg daarvan kan het *werkelijke verschil in lengte* tussen beide personen (B is 2 cm langer dan A) worden gemaskeerd door de *toevallige verschillen* (toevalsfactoren, meetfouten) die teweeg worden gebracht door de minder betrouwbare meetprocedure (methode 2).

Bovenstaand voorbeeld vormt in feite een illustratie van een veelgebruikte **formele definitie van betrouwbaarheid**. Deze ziet er als volgt uit:

$$\text{betrouwbaarheid} = \frac{\text{systematische variantie}}{\text{systematische variantie} + \text{foutenvariantie}}$$

In de noemer van deze 'formule' staat de totale, werkelijk gemeten variantie. Deze bestaat uit twee delen: een stabiele of systematische component en een deel dat te wijten is aan toevallige fouten. In de teller van de formule staat alleen het systematische deel van de totale gemeten variantie. Uit de formule is direct af te lezen dat met het toenemen van het aandeel van de foutenvariantie de breuk (ratio) kleiner wordt: de meting wordt dus minder betrouwbaar als de foutenvariantie toeneemt. Met andere woorden: *betrouwbaarheid is de proportie systematische variantie ten opzichte van de totale gemeten variantie.*

Terugkerend naar het voorbeeld van de lengtemeting: bij *methode 1* komt het werkelijke verschil van 2 cm tussen persoon A en B óók in de meetresultaten tot uiting, doordat in de totale variantie slechts weinig foutenvariantie voorkomt: de systematische variantie (het werkelijk bestaande verschil) overheerst daardoor ook sterk in de noemer van de formule en de betrouwbaarheid is dus hoog. Bij *methode 2* is het aandeel van de foutenvariantie in de totale variantie groter, waardoor de systematische variantie (het verschil van 2 cm) zich minder goed kan manifesteren: het aandeel van de systematische variantie in de noemer is daardoor geringer dan bij methode 1 en de meting is dus minder betrouwbaar.

Bij het bepalen van kenmerken als lengte en gewicht valt de onbetrouwbaarheid meestal nog wel mee. Het wordt (veel) moeilijker als het gaat om variabe-

len als assertiviteit, beroepskeuze of intelligentie. Uit bovenstaand voorbeeld over lengtemeting blijkt dat de betrouwbaarheid van een meting kan worden vergroot door nauwkeurig te omschrijven hoe en onder welke omstandigheden de meting moet worden uitgevoerd.

15.3.1 *Soorten betrouwbaarheid*

Betrouwbaarheid kan op verschillende manieren worden vastgesteld:

1 *Test-hertestbetrouwbaarheid*

- procedure: geef de test nu en later en bepaal de correlatie tussen de resultaten op beide tijdstippen.
- interpretatie: bij hoge correlatie is sprake van een hoge test-hertestbetrouwbaarheid en is er veel systematische variantie (de door de 'ware scores' verklaarde variantie is groot); bij een lage samenhang tussen de resultaten op beide tijdstippen is er veel foutenvariantie, maar het kan ook zijn dat de eigenschap die gemeten werd, is veranderd (bijvoorbeeld een ontwikkelingskenmerk bij zeer jonge kinderen!).

2 *Inter-itembetrouwbaarheid*:

- procedure: bepaal de correlatie tussen twee 'helften' van de test: (test)halveringsbetrouwbaarheid (bekende Engelse term: 'split-halfreliability') of bepaal de samenhang tussen alle items (vragen, opdrachten) van een test die beoogt één kenmerk/eigenschap te meten.
- interpretatie: als de samenhang tussen de items hoog is, noemen we de test 'intern consistent'; bij een lage samenhang is de test niet intern consistent, maar de vraag is natuurlijk of dat moest (de scores op de rekenitems uit een intelligentietest moeten onderling sterk samenhangen, net als de scores op verbale vaardigheiditems, maar het is goed denkbaar dat de scores op reken- en taalopgaven slechts matig correleren).

3 *Inter-beoordelaarbetrouwbaarheid* ('interrater agreement', 'interobserver reliability'):

- procedure: bepaal de correlatie tussen de scores van twee of meer beoordelaars die hetzelfde verschijnsel observeren of beoordelen.
- interpretatie: als de samenhang hoog is, maakt het niet uit wie er observeert of beoordeelt: de observatoren zijn als het ware onderling 'inwisselbaar'; in het geval van een lage samenhang tussen de scores van de diverse beoordelaars is er een probleem dat moet worden opgelost door extra trainingsbijeenkomsten en/of aanpassing van de categorieën van het gebruikte beoordelings- of observatiesysteem.

Betrouwbaarheid neemt over het algemeen toe als het aantal vragen in de 'steekproef van vragen' toeneemt, mits uit dezelfde (bedoelde) populatie van vragen getrokken is. Je kunt iemands rekenvaardigheid beter meten met tien dan met drie opgaven. In het algemeen geldt dat een redelijk groot aantal items (tussen 10 en 15) tot een goede betrouwbaarheid leidt, maar dat toevoeging van extra items nauwelijks meer effect heeft.

Ter afsluiting staan we kort stil bij de *relatie tussen validiteit en betrouwbaarheid*. Een thermometer die, na te zijn gevallen, defect is geraakt en is blijven staan op 17 °C is uiterst betrouwbaar (herhaalde meting geeft steeds identieke resultaten), maar het instrument bezit voor het concept 'temperatuur' geen enkele validiteit meer. Dit (extreme) voorbeeld maakt duidelijk dat betrouwbaarheid allerminst een garantie vormt voor validiteit. *Toch is betrouwbaarheid van groot belang voor validiteit*. Immers, als een meting onbetrouwbaar is (herhaalde meting aan hetzelfde object geeft zeer verschillende resultaten) en dus veel toevallige meetfouten produceert, meet zij niet meer wat zij bedoelt te meten. Dan is zo'n meting dus weinig valide. Met andere woorden: *betrouwbaarheid is een voorwaarde voor validiteit.*

Opgaven

15.1 Geef voor elk van de volgende concepten aan waarom het nodig is ze nader te omschrijven. Ga daarbij na welke verschillende betekenis(sen) of inhoud(en) ze kunnen hebben als ze niet nader (operationeel) zijn gedefinieerd.

a vrijheid
b inkomen
c prestatiemotivatie
d leeftijd
e bejaard

15.2 Geef van de volgende definities aan of ze conceptueel dan wel operationeel zijn.

a Koorts is een verhoging van de lichaamstemperatuur tot boven de 38 °C.
b Intelligentie is het probleemoplossend vermogen in nieuwe situaties
c Macht is het feitelijk resultaat van beïnvloedingspogingen
d Drankgebruik is het aantal geconsumeerde cl alcohol per uur
e Agressie is het bewust gebruikmaken van of dreigen met fysiek geweld om je doel te bereiken
f Latentietijd is de tijd die verloopt tussen een prikkel (stimulus) en de daaropvolgende waarneembare reactie

15.3 Operationaliseer de volgende begrippen:

a communicatie
b politieke signatuur van kranten
c man/vrouw-rolverdeling in gezin
d inkomen

15.4 *a* Hoe kan men de concurrente validiteit van een intelligentietest vaststellen?
b En hoe de construct- of begripsvaliditeit?

15.5 Een onderzoeker heeft een rekenvaardigheidstoets (25 items) ontwikkeld voor groep 6 van het basisonderwijs.
Beschrijf wat hij moet doen om:

a de test-hertestbetrouwbaarheid van zijn toets te bepalen
b de halveringsbetrouwbaarheid vast te stellen

15.6 *a* Bespreek het onderscheid tussen de begrippen validiteit en betrouwbaarheid.
b Waarom zeggen we dat een meting betrouwbaar kan zijn zonder validiteit, maar dat het omgekeerde (een valide meting die niet betrouwbaar is) niet mogelijk is?

Uitwerkingen

15.1 *a* Vrijheid als een toestand/situatie waarin men *niet* is opgesloten (bijvoorbeeld in een gevangenis) of *niet* in een bezet gebied (bijvoorbeeld tijdens oorlog) woont? Of wordt gedoeld op de toestand/omstandigheid waarin men vrij is alles te doen wat men wil? Of gaat het om de toestand/omstandigheid waarin men alles kan doen wat men wil, zolang dit *niet* ten koste gaat van het recht van iemand anders om alles te doen wat hij wil enzovoort?

b Netto- of bruto-inkomen? Per week, maand of jaar? Met of zonder aftrekbare rente, ziektekostenpremie en dergelijke? Alleen inkomen uit arbeid, of ook uit aandelen, pensioen, uitkeringen enzovoort?

c Gaat het om iemands behoefte aan succes, om een vergroting van de behoefte aan succesrijke ervaringen? Of wordt gedoeld op een eigenschap gekenmerkt door een vasthoudende en competitieve instelling in een moeilijke taaksituatie?

d Vragen we naar iemands geboortejaar en eventueel -maand, of rechtstreeks naar iemands leeftijd? De eerste benadering levert nauwkeuriger informatie op dan de tweede. Bij (zeer) jonge kinderen is het relevant de leeftijd in maanden en eventueel dagen te meten!

e Bejaard als een toestand van teruglopende vitaliteit en afnemende (lichamelijke en/of geestelijke) capaciteiten? Of wordt met het concept niets anders bedoeld dan het overschrijden van een bepaalde leeftijdsgrens en betekent het gewoon '65 jaar en ouder'?

15.2 a, d en f zijn voorbeelden van operationele definities; b, c en e zijn conceptuele definities.

15.3 Het zal duidelijk zijn dat de volgende antwoorden slechts voorbeelden zijn van mogelijke operationalisaties en de daarbij geldende overwegingen.

a Stel dat wordt gekozen voor de methode van observatie van het gedrag (*live* of achteraf via analyse van videobeelden) om vast te stellen 'wie van een bepaalde groep mensen met wie communiceert'. Dan moet vervolgens worden bepaald of *communicatie* alleen maar wordt geoperationaliseerd als 'praten met elkaar' (en voor hoe lang, bijvoorbeeld minstens een halve minuut) of dat ook oogcontact of aanraking (non-verbaal gedrag) wordt meegeteld. Is communicatie wederzijds, of wordt een monoloog ook als communicatie beschouwd, enzovoort.

b Men zou *politieke signatuur* kunnen operationaliseren via een frequentietelling van de aanduiding van bepaalde opstandige bewegingen in termen van 'rebellen', 'guerrilla's', 'opstandelingen' enerzijds, versus de termen 'bevrijdingsleger', 'patriottisch front' en dergelijke anderzijds. Relevante vragen daarbij zijn: wat voor steekproef uit de tijd men neemt (welke periode, welke dag van de week); of alleen de nieuwsberichtgeving of dat ook (of vooral?) het redactioneel commentaar erbij wordt betrokken, enzovoort. Men kan de politieke richting natuurlijk ook vaststellen door een enquête te houden onder journalisten en/of het lezerspubliek.

c In een vragenlijst of interview zou men kunnen vragen naar onder andere de volgende aspecten van rolverdeling:
- kostwinnerschap (wie verdient het (meeste) geld?)
- wie past op/is beschikbaar voor de kinderen?
- wie bepaalt de vakantiekeuze?
- hoe is verdeling van huishoudelijke taken (strijken, koken, (af)wassen, stofzuigen etc.)?

d Stel dat het gaat om het jaarinkomen. Dan moet worden beslist of het alleen gaat over inkomsten uit arbeid, of ook om inkomsten uit vermogen of ondernemingswinst (bij bedrijven). Gaat het om netto- of bruto-inkomen? Of kiest men, zoals vaak gebeurt, voor het belastbaar inkomen, het inkomen waarover belasting verschuldigd is? Dit belastbaar inkomen wordt berekend door het inkomen te verminderen met de kosten op grond van persoonlijke verplichtingen, buitengewone lasten en aftrekbare giften.

15.4 Stel dat de test is afgenomen bij leerlingen van de basisschool.

a Voor het vaststellen van de concurrente validiteit kunnen de testprestaties van de leerlingen worden vergeleken met gelijktijdige, door de leerkracht verrichte prestatiebeoordelingen.

b De begrips- of constructvaliditeit van de test kan worden bepaald door de scores op de test te correleren met schoolprestaties of andere gedragingen, waarvan op theoretische gronden mag worden aangenomen dat ze (nauw) met het construct intelligentie samenhangen.

15.5 *a* De test-hertestbetrouwbaarheid kan worden bepaald door de toets op twee tijdstippen bij dezelfde kinderen af te nemen. Het interval tussen beide meetmomenten moet natuurlijk niet te groot zijn (niet langer dan 5 à 6 weken). Vervolgens worden de op beide momenten behaalde scores gecorreleerd.

b Bij het berekenen van de *split-half*- of halveringsbetrouwbaarheid is sprake van één meettijdstip en worden (bijvoorbeeld) de scores op de oneven items gecorreleerd met de scores op de even items van de toets.

15.6 *a* Validiteit heeft betrekking op de vraag of een test (of meer algemeen: een meetinstrument) al dan niet meet wat hij wordt verondersteld te meten, terwijl betrouwbaarheid betrekking heeft op de vraag of de test (het instrument) datgene wat hij meet op *consistente* wijze meet.

b Een test (instrument) kan betrouwbaar zijn maar niet valide, doordat hij op consistente wijze iets meet waarvoor hij niet ontwikkeld werd en er niet in slaagt datgene te meten waarvoor hij werd ontwikkeld.
Een onbetrouwbare test kan nooit valide zijn, omdat een gebrek aan consistentie (sterk wisselende resultaten bij herhaalde meting aan hetzelfde object) de mogelijkheid uitsluit dat de test meet wat hij wordt verondersteld te meten.
Kortom, betrouwbaarheid is een voorwaarde voor validiteit, al is niet elke betrouwbare meting ook valide.

16 Interne validiteit van onderzoeksopzetten

Nadat is vastgesteld dat twee variabelen onderling samenhangen en aannemelijk is gemaakt dat eventuele bedreigingen van de zogenaamde statistische conclusievaliditeit (zie paragraaf 15.2) in het onderzoek niet plausibel zijn, staat de onderzoeker voor de taak aannemelijk te maken dat de interne validiteit (in engere zin) van het onderzoeksontwerp als geheel niet werkelijk bedreigd is. De daarbij te beantwoorden vraag luidt of er sprake is van een causale relatie tussen de gemanipuleerde (experimentele) variabele (X) en de gemeten criterium- of afhankelijke variabele (Y). Met andere woorden, het gaat om de geldigheid (validiteit) van empirische uitspraken omtrent de vraag of X inderdaad een verandering heeft veroorzaakt in Y (of wordt de verandering in Y veroorzaakt door iets anders?).

16.1 Bedreigingen van de interne validiteit

In een aantal studieboeken op het gebied van methoden en technieken wordt onderscheid gemaakt tussen bedreigingen van de interne validiteit van (I) binnenpersoonsvergelijkingen ('within-subjects comparisons') en (II) tussenpersoonsvergelijkingen ('between-subjects comparisons'). Wij komen daar later op terug.

Tot I worden gerekend:
- Maturation (autonome verandering of 'rijping')
- Testing (testeffecten)
- Instrument decay (verandering van/in het meetinstrument)
- Statistical regression (statistische regressie)
- History (externe gebeurtenissen)

Tot II behoren:
- Selection bias (vertekening ten gevolge van selectie)
- Differential attrition (differentiële uitval)
- Diffusion (overdracht van behandeling)
- Compensatory equalization (compenserende maatregelen)
- Compensatory rivalry (compenserende rivaliteit)
- Resentful demoralization (ontmoediging door boosheid of uit wrok)

16.2 Voorbeeld

Een onderzoekster wil het effect bepalen van een nieuwe taalmethode, die bestaat uit een speciaal ontwikkeld programma om kinderen met behulp van 'videolessen' de beginselen van spelling en grammatica bij te brengen. In onderwijskringen en in voorbeschouwingen in de media is het programma als 'vernieuwend' en 'veelbelovend' gekarakteriseerd.

In overleg met de directie van een grote basisschool besluit zij een onderzoek uit te voeren, dat loopt van het begin van de vijfde groep tot het eind van de zesde. Van de directie krijgt zij 20 namen van kinderen die aan het eind van de vierde groep nog grote problemen met taal hadden. Met deze kinderen start zij aan het begin van de vijfde groep het nieuwe programma. In de controlegroep (deze kinderen volgen de op de school gebruikelijke methode) kiest zij op dezelfde school aselect 20 kinderen van dezelfde leeftijd als die van de leerlingen in de experimentele groep.
Aan het begin van het vijfde en aan het eind van het zesde leerjaar neemt de onderzoekster bij beide groepen een veelgebruikte taalvaardigheidstoets af. Na afloop van het onderzoek vergelijkt ze de vooruitgang in taalvaardigheid van de twee groepen en vindt dat ze *beide gemiddeld dezelfde prestatieverbetering te zien geven.*
We gaan nu systematisch na of en in welke mate er sprake is van bedreigingen van de interne validiteit van (causale) uitspraken die op basis van bovenstaande onderzoeksopzet gedaan kunnen worden. Het gaat er dus om of deze factoren als alternatieve verklaringen (men spreekt ook wel van 'rivaliserende hypothesen') voor de onderzoeksresultaten kunnen worden aangemerkt.

16.2.1 *Mogelijke alternatieve verklaringen voor veranderingen tussen voor- en nameting:*

- Van **maturation** zou in dit geval sprake zijn indien de kinderen in de tijd tussen voor- en natoets veranderen ten gevolge van andere kenmerken dan de behandeling. Denk hierbij aan autonome veranderingen in het kind als ouder, wijzer, zelfstandiger enzovoort worden.
  Voor deze factor geldt dat *in eerste instantie niet* te verwachten is dat deze een *verschillend* effect voor experimentele respectievelijk controlegroep zal hebben. De kinderen hebben immers dezelfde leeftijd en bevinden zich in dezelfde omstandigheden. Anderzijds bevond de experimentele groep zich bij aanvang op een (veel) lager (taal)niveau dan de controlegroep. Het is dan ook niet volledig uit te sluiten dat kinderen in beide groepen zich over een periode van twee jaar anders hebben ontwikkeld.
- Van **testing** zou in dit geval sprake zijn indien de resultaten op de natoets beter zijn doordat de kinderen eenzelfde soort toets al eerder hebben ingevuld. Deze factor vormt weliswaar een potentiële bedreiging in het geval van de aanwezigheid van een voor- en natoets, maar is hier niet waarschijnlijk omdat er (bijna) twee jaar zijn verlopen tussen beide afnames van de toets.
- Van **instrument decay** zou in dit geval sprake zijn indien de taalvaardigheidstoets als instrument in de loop van het onderzoek is veranderd. Deze factor vormt in dit onderzoek geen bedreiging, omdat gewerkt wordt met een veelgebruikte taalvaardigheidstoets: de afnameprocedure ligt vast en is standaard.
- Van **statistical regression** zou in dit geval sprake zijn indien de groepen zijn geselecteerd vanwege hun extreme scores. Deze factor vormt inderdaad een bedreiging, doordat de kinderen uit de experimentele groep gekozen zijn vanwege hun *lage* taalvaardigheid. De kans is aanwezig dat deze leerlingen bij de tweede toetsafname ten gevolge van statistische regressie in de rich-

ting van het gemiddelde zullen opschuiven, waardoor deze verbetering niet ondubbelzinnig is toe te schrijven aan het volgen van de nieuwe taal-methode.
- Van **history** zou in dit geval sprake zijn indien er tussen voor- en natoets externe gebeurtenissen plaatsvinden, afgezien van de behandeling. Op voorhand is deze factor als potentiële bedreiging niet volledig uit te sluiten. Het is bijvoorbeeld denkbaar dat de ouders van de leerlingen uit de experimentele groep de zwakke taalvaardigheid van hun kinderen hebben proberen te verbeteren via ondersteuning door een remedial teacher, veel voorlezen en het geven van extra taaloefeningen. De onderzoekster zal dus moeten nagaan of en in welke mate dit soort 'externe gebeurtenissen' de taalvaardigheidsontwikkeling in de experimentele groep beïnvloed kan hebben.

16.2.2 Mogelijke alternatieve verklaringen voor het verschil in verandering tussen experimentele en controlegroep

- Van **selection bias** zou in dit geval sprake zijn indien er systematisch verschil bestaat tussen de groepen ten gevolge van oorspronkelijke verschillen, te wijten aan niet-aselecte toewijzing van personen. Deze factor vormt hier een bedreiging, omdat de kinderen in de experimentele groep zijn geselecteerd vanwege hun *lage* taalvaardigheid, terwijl de controlegroep wel aselect is samengesteld.
- Van **differential attrition** zou in dit geval sprake zijn indien personen, ten gevolge van het al of niet ondergaan van de behandeling, uit het onderzoek verdwijnen. Deze factor speelt hier geen rol, omdat uit niets blijkt dat er sprake was van uitvallers in één van beide groepen.
- Van **diffusion** zou in dit geval sprake zijn indien er overdracht van behandelingseffecten plaatsvindt van experimentele naar controlegroep. Deze factor vormt zeker een potentiële bedreiging: de leerlingen in beide groepen zitten op dezelfde school, dus de kans dat er gesproken wordt over de nieuwe methode en dat informatie daarover wordt uitgewisseld, is groot.
- Van **compensatory equalization** zou in dit geval sprake zijn indien de personen in de controlegroep een 'vervangende behandeling' krijgen, zodat zij zich niet achtergesteld voelen. Gesteld wordt dat de kinderen in de controlegroep *de op school gebruikelijke taalmethode* volgen. Op grond daarvan lijkt deze factor in *dit* onderzoek geen bedreiging te vormen. In meer algemene zin kan worden opgemerkt dat de schoolleiding in een situatie van gunstige voorpubliciteit in de media over een nieuwe onderwijsmethode best onder druk kan komen te staan (van ouders en/of docenten) ook aan de kinderen van de controlegroep een vorm van 'aangepast' onderwijs te geven waarin leuke elementen (zoals het gebruik van video) gedeeltelijk worden overgenomen. Als dit ook werkelijk gebeurt, zijn dergelijke 'compenserende maatregelen' natuurlijk wel een bedreiging van de interne validiteit.
- Van **compensatory rivalry** zou in dit geval sprake zijn indien de leerlingen in de controlegroep extra hun best doen om te laten zien dat ze 'even goed' zijn als de experimentele groep. Deze factor kan inderdaad een bedreiging vormen, wanneer de leerlingen uit de controlegroep (en hun leerkracht) willen laten zien dat zij ook *zonder* videolessen goede prestaties in taal kunnen leveren. In dat geval zou het experimentele effect niet aangetoond kunnen

worden. De kans daarop is wellicht wat minder groot, omdat de leerlingen uit de controlegroep geen taalachterstand hebben.

– Van **resentful demoralization** zou in dit geval sprake zijn indien de leerlingen in de controlegroep uit boosheid minder goed presteren omdat ze zich achtergesteld voelen. Deze factor zou een rol kunnen spelen wanneer de leerlingen in de controlegroep het missen van de videolessen zo erg vinden dat zij zich bij het taalonderwijs minder gemotiveerd tonen en zich nauwelijks inspannen. In dat geval is er sprake van een bedreiging van de interne validiteit.

In het voorgaande zijn zowel bedreigingen van de interne validiteit van binnenpersoons('within subjects')- als tussenpersoons('between subjects')-vergelijkingen genoemd. Dit omdat er in het onderzoek sprake is van voortoetsnatoetsgegevens *en* van vergelijkingen tussen groepen. In de praktijk is dit regelmatig het geval. Natuurlijk zijn er ook onderzoeksopzetten waarin heel duidelijk sprake is van hetzij een binnenpersoons-, hetzij een tussenpersoonsvergelijking.

De strikte scheiding tussen beide soorten vergelijkingen die in sommige M & T-boeken wordt aangebracht, dient echter voornamelijk een didactisch doel.

Ten slotte is het van groot belang dat men leert inzien dat een potentiële bedreiging van de interne validiteit pas *reëel* is, indien er aanwijzingen zijn of kan worden verwacht dat *de aard en/of de uitwerking ervan in de ene groep verschilt van die in de andere!*

Opgaven

16.1 De schooladviesdienst te Leiden heeft een vijftal remedial teachers aangetrokken om leerlingen met rekenmoeilijkheden te helpen. Samen kunnen deze op jaarbasis 200 kinderen helpen. Om vast te stellen welke kinderen de grootste behoefte aan extra rekenhulp hebben, krijgen alle Leidse kinderen uit groep 4 van de basisschool een rekentoets. De 200 kinderen met de laagste score op deze toets krijgen remediërend rekenonderwijs. Aan het eind van het schooljaar wordt de rekentoets nog eens afgenomen, zowel bij de kinderen die extra hulp hebben gehad als bij de overige vierdegroepers. Vervolgens wordt een vergelijking gemaakt tussen de vooruitgang die door de groep van 200 en de rest van de kinderen is geboekt.
Bestudeer de bovenstaande onderzoeksopzet nauwkeurig en beantwoord de volgende vragen.

a Wat zijn in dit voorbeeld de onafhankelijke en de afhankelijke variabele(n)?

b Welke bedreigingen van de interne validiteit spelen in deze onderzoeksopzet duidelijk een rol en waarom?

c Welke bedreigingen van de interne validiteit spelen hier (zeer waarschijnlijk) geen rol en waarom niet?

d Stel dat de ouders van de kinderen uit de experimentele groep thuis extra gaan oefenen met rekenopgaven. Welke bedreiging(en) van de interne validiteit gaat (gaan) dan een rol spelen?

e Stel dat de rekentoets een pas ontwikkeld instrument is. Welke bedreiging(en) is (zijn) dan aan de orde?

Uitwerkingen

16.1 *a* De onafhankelijke variabele (vaak aangeduid met de Engelse term 'treatment') is het al dan niet ontvangen van remedial teaching op het gebied van rekenvaardigheid. De afhankelijke variabele (de variabele die 'afhangt' van de onafhankelijke variabele) is de score op de rekentoets.

b – **Statistical regression** speelt in dit onderzoek zeker een rol, omdat de groep van 200 kinderen is gekozen vanwege hun (extreem) lage rekenvaardigheid. De kans is dan ook groot dat eventuele vooruitgang op de rekentoets het gevolg is van regressie naar het gemiddelde en niet wordt veroorzaakt door de behandeling.
– **Selection bias** speelt een rol, omdat de kinderen in de experimentele groep zijn geselecteerd op grond van hun lage rekenvaardigheid. Ten gevolge van deze selectie zijn experimentele en controlegroep niet meer vergelijkbaar.

c – **Testing** speelt naar alle waarschijnlijkheid geen rol, omdat er tussen voor- en natoets (bijna) een jaar is verlopen. Het lijkt praktisch uit te sluiten dat kinderen zich na zo'n lange periode de voortoets nog kunnen herinneren.
– **Differential attrition** is naar alle waarschijnlijkheid hier niet aan de orde, omdat (1) er niets over vermeld wordt en (2) de kinderen leerplichtig zijn. Daarnaast is er geen reden om te veronderstellen dat factoren als verhuizing, ziekte en dergelijke zich in één van beide groepen in verschillende mate zullen voordoen.
– **Compensatory equalization** vormt hier geen bedreiging, omdat het aanbieden van extra onderwijs in de vorm van remedial teaching niet als een voorkeursbehandeling wordt beschouwd, waardoor de schoolleiding geen aanleiding heeft om de controlegroep, die geen of minder problemen kent, ter compensatie een vervangende behandeling aan te bieden.

d Een bedreiging die in dit geval met name een rol gaat spelen, is **history**. Eventuele vooruitgang van de experimentele groep op de rekentoets is dan mede toe te schrijven aan de extra training in rekenvaardigheid in het gezin, naast de invloed van remedial teaching.

e Een nieuwe toets kan in de loop van het onderzoek worden veranderd, aangepast of bijgesteld enzovoort, waardoor **instrument decay** een bedreiging zou kunnen gaan vormen.

17 Bedreigingen van de externe validiteit

Bij externe validiteit is de vraag hoe een onderzoek zó kan worden opgezet dat de resultaten kunnen worden gegeneraliseerd:
- vanuit de steekproef naar de populatie;
- vanuit de specifieke behandeling in het onderzoek naar soortgelijke behandelingen daarbuiten;
- vanuit de specifieke omgeving in het onderzoek naar andere omgevingen daarbuiten.

Kortom, bij externe validiteit gaat het om de mate waarin onderzoeksresultaten kunnen worden gegeneraliseerd naar verschillende soorten personen, contexten (omgevingen) en meettijdstippen. Indien een onderzoek op al deze punten niet extern valide is, dan gelden de resultaten van dit onderzoek *slechts* voor die personen, omgeving(en), behandeling en tijdstip zoals die in de steekproef voorkwamen.

17.1 Soorten externe validiteit

De mate waarin onderzoek externe validiteit bezit, kan nader worden toegespitst op twee hoofdcategorieën:

1 *Populatievaliditeit*
Kunnen de verkregen resultaten worden gegeneraliseerd naar de zogenoemde doelpopulatie? Met doelpopulatie bedoelen we die populatie waarin we uiteindelijk geïnteresseerd zijn.

2 *Ecologische validiteit*
Kunnen de verkregen resultaten worden gegeneraliseerd naar vergelijkbare situaties en contexten?

Beantwoording van de vraag naar de mate van populatievaliditeit verloopt meestal in twee stappen:

Stap 1 Kunnen de gegevens uit de steekproef worden gegeneraliseerd naar de experimenteel toegankelijke populatie (de op dat moment toegankelijke populatie)? Dit is een probleem omdat in de praktijk van het sociaal-wetenschappelijk onderzoek nogal vaak gebruik wordt gemaakt van een min of meer toevallige steekproef uit een (op dat moment) toegankelijke populatie. Er is dan géén sprake van een aselecte steekproef uit de doelpopulatie. Dit heeft directe gevolgen voor de (mate van) externe validiteit. We weten nooit zeker of we op deze manier de doelpopulatie in haar geheel bereiken.
N.B. In zo'n geval worden personen vaak wél aselect toegewezen aan groepen, maar dit heeft te maken met de *interne validiteit* van het onderzoek, want dit verandert immers niets aan het eventuele tekort aan externe validiteit!

Stap 2 Kan er vanuit de steekproef (via de toegankelijke populatie) worden gegeneraliseerd naar de uiteindelijke doelpopulatie? Hier is het de vraag of generalisatie naar de populatie als geheel ook betekent dat naar (over) subpo-

pulaties (subgroepen) mag worden gegeneraliseerd. Men zou bijvoorbeeld kunnen vinden dat (1) een nieuwe rekenmethode de prestaties op rekenvaardigheid van de 8-10-jarige kinderen uit de doelpopulatie weliswaar verhoogt, maar (2) dat het patroon voor de subpopulaties jongens en meisjes verschilt.

Bij de beantwoording van de vraag naar de mate van ecologische validiteit gaat het om zaken als:

a Is generalisatie over geografische gebieden geoorloofd? Geldt een onderzoeksresultaat op het platteland ook voor grote stedelijke gebieden of gelden kenmerken van de opvoeder-kindinteractie in Nederland ook voor opvoedingspatronen in Sri Lanka?
Dergelijke vragen kunnen eigenlijk alleen met behulp van zogenaamd **cross-cultureel** onderzoek worden beantwoord. Dit is onderzoek waarin dezelfde probleemstelling in verschillende culturen wordt onderzocht.

b Is er sprake van een interactie tussen behandeling en omgeving? Als van 50 aangeschreven zogenaamde 'zwarte' scholen er slechts 5 willen meewerken aan een onderzoek naar de kwaliteit van het onderwijs, is er duidelijk geen sprake van een representatieve steekproef uit de populatie. In zo'n geval kunnen we de resultaten van het onderzoek natuurlijk niet zonder meer laten gelden voor alle 'zwarte' scholen.

c Is er sprake van temporele validiteit? Gelden de resultaten van vandaag ook morgen, of zouden de resultaten van een onderzoek naar leesvaardigheid uitgevoerd op vrijdagmiddag ook maandagochtend zijn verkregen?

d Kan worden gegeneraliseerd over behandelingen? Is vastgesteld of effecten bij personen die meerdere behandelingen hebben ondergaan zich ook hebben voorgedaan bij personen die elk van deze behandelingen afzonderlijk hebben ontvangen? (Met andere woorden: is er geen sprake van 'multiple treatment interference'?)

e Is generalisatie buiten de specifieke context geoorloofd? Als het onderzoek binnen een zeer specifieke situatie (denk aan laboratoriumonderzoek) is gedaan, gelden de resultaten dan ook daarbuiten?

f Kan worden gegeneraliseerd over onderzoekers? De (persoon van de) onderzoeker is geen neutraal 'stimulusobject'. Geldt wat door de ene onderzoeker is gevonden, ook zonder meer voor resultaten van andere onderzoekers? Kenmerken van onderzoekers die daarbij een rol kunnen spelen zijn sekse, ras, al dan niet 'aardig gevonden worden' en warmte.

17.2 Bedreigingen van de externe validiteit

Bedreigingen van de externe validiteit zijn in essentie voorbeelden van interactie-effecten tussen de 'behandeling' en personen, contexten en tijdstippen. De belangrijkste bedreigingen van de populatievaliditeit zijn:

1 *De interactie tussen selectie en behandeling*
Deze interactie vormt een bedreiging van de externe validiteit omdat het bijna nooit mogelijk is om een aselecte steekproef te trekken uit alle leden van de doelpopulatie (sommige personen zijn minder goed te bereiken dan

andere en/of zijn minder bereid om aan het onderzoek mee te werken). Het verkrijgen van een goede representatieve steekproef uit de doelpopulatie wordt vooral bemoeilijkt door *beperkte financiën* en *zelfselectie*, c.q. vrijwilligerseffecten ('volunteer bias').
Beperkte financiële middelen zullen ertoe leiden dat een onderzoek naar het gedrag van adolescenten in Nederland in de praktijk waarschijnlijk wordt uitgevoerd bij middelbare scholieren afkomstig van enkele scholen in bijvoorbeeld Noord-Holland of Gelderland, in plaats van bij een aselecte steekproef getrokken uit alle adolescenten in Nederland.
Bij een kijk- en luisteronderzoek van de NOS zijn mensen die bereid zijn om daarvoor speciaal naar Hilversum te komen, waarschijnlijk veel 'atypischer' (minder doorsnee) dan mensen die bereid zijn mee te werken aan een huis-aan-huisonderzoek. De kijk- en luistercijfers uit het laatste onderzoek zijn waarschijnlijk veel representatiever dan die uit het eerste, omdat de resultaten uit het eerstgenoemde onderzoek veel sterker onderhevig zijn aan de gevolgen van *zelfselectie*, c.q. de invloed van zogenaamde *vrijwilligerseffecten*.

2 *De interactie tussen context (situatie) en behandeling*
De in een fabriekscontext gevonden causale relatie tussen leiderschapsstijl (democratisch, autoritair) en productiviteit van werknemers hoeft niet te gelden voor de relatie tussen onderwijsstijl van de leerkracht en leerlingprestaties op de basisschool.

3 *De interactie tussen externe gebeurtenissen ('history') en behandeling*
De invloed van antikernenergiepropaganda is waarschijnlijk groter vlak na de ramp in Tsjernobyl dan op een moment dat er al vele jaren geen sprake is (geweest) van rampen met kernenergiecentrales.

17.3 Het verhogen van de externe validiteit

Tot slot noemen we enkele mogelijkheden om de externe validiteit van onderzoek te *verhogen*:

1 Veldonderzoek ('field experiment/field research'). Dit is een type onderzoek waarin een of meer onafhankelijke variabelen worden gemanipuleerd op een manier en in een omgeving die zeer nauw aansluit bij de natuurlijke omgeving/situatie van de onderzochte personen. Ook voor de afhankelijke variabelen geldt dat zoveel mogelijk wordt gestreefd naar het verkrijgen van natuurlijke of 'spontane' reacties.
2 Steekproeven met een grote reikwijdte en een goede representativiteit.
Hierbij zijn een aantal zaken van bijzonder belang:
- de steekproef zorgvuldig vergelijken met de doelpopulatie
- de doelpopulatie nauwkeurig omschrijven
- de beste oplossing is om aselect te trekken, eventueel gelaagd ('gestratificeerd') om alle relevante groepen op te nemen
- vrijwillige deelname aan het onderzoek moet zo gemakkelijk mogelijk worden gemaakt

– verhoging van de representativiteit door trekking uit zo divers mogelijke groeperingen, omgevingen etc. Een nadeel hiervan is echter dat dit ten koste kan gaan van de interne validiteit.

Opgaven

17.1 Bij 8-jarige meisjes werd een onderzoek met herhaalde metingen uitgevoerd, om te bepalen wat het effect is van drie niveaus van beloning op de uitvoering van een eenvoudige puzzel. Uit de gegevens komt naar voren dat gematigde beloning tot de beste resultaten leidt. In hoeverre kunnen deze resultaten gegeneraliseerd worden?

17.2 Een farmaceutische industrie is geïnteresseerd in de effecten van nieuwe antidepressiva. De bedoeling is, deze middelen eerst op ratten uit te proberen en daarna op mannelijke studenten. In hoeverre kunnen de resultaten van elk van de twee steekproeven worden gegeneraliseerd?

17.3 Hieronder staat een fictief voorbeeld van een vragenlijst die huis-aan-huis bezorgd werd. Geef aan waardoor de externe validiteit van dit onderzoek bedreigd kan worden.

Voor u ligt een enquête van Bureau Viewmaster. Deze vragenlijst is samengesteld in opdracht van verschillende bedrijven uit de consumptiesector. Deze bedrijven zijn geïnteresseerd in uw consumptiepatroon. Uw mening over de producten is van groot belang omdat deze fabrikanten er dan voor kunnen zorgen dat er producten in de winkel liggen waar u behoefte aan heeft! Tevens is er een beloning verbonden aan het volledig invullen en insturen van de vragenlijst, namelijk een **boodschappenpakket** met producten van deze fabrikanten!!! Reden genoeg om 5 minuten van uw tijd hieraan te besteden.

Welk wasmiddel gebruikt u?
- ○ All
- ○ Vizir
- ○ Ariël
- ○ Omo
- ○ Huismerk
- ○ Ander merk

Welk afwasmiddel gebruikt u?
- ○ Dreft
- ○ Dubro
- ○ Lux
- ○ Huismerk
- ○ Ander merk

Welke halvarine gebruikt u?
- ○ Blue Band
- ○ Gouda's Glorie
- ○ Wajang
- ○ Lätta
- ○ Bona
- ○ Linera
- ○ Huismerk
- ○ Ander merk

Heeft u een boot? Zo ja, wilt u deze omschrijven?
.....
.....
.....

Rookt u ?
○ Nee ○ Ja

Welk merk rookt u meestal?
- ○ Marlboro
- ○ Peter Stuyvesant
- ○ Roxy
- ○ Camel
- ○ Dunhill
- ○ Caballero
- ○ Ander merk
- ○ Shag

Hoeveel rookt u? ..... per dag.

Welke wasverzachter gebruikt u?
- ○ Silan
- ○ Robijn
- ○ Ander merk
- ○ Ik gebruik geen wasverzachter

Heeft u een computer?
○ Ja ○ Nee

Zo ja, welk merk ?
- ○ Tulip
- ○ Commodore
- ○ Apple
- ○ IBM
- ○ Ander merk

Welk merk chocolade gebruikt u ?
- ○ Baronie
- ○ Verkade
- ○ Droste
- ○ Huismerk
- ○ Ander merk

Indien u katten heeft, welk merk blikvoer gebruikt u ?
- ○ Sheba
- ○ Whiskas
- ○ Tompoes
- ○ Kittekat
- ○ Miauw
- ○ Ander merk

Welke merken huishoudelijke apparaten heeft u?
- ○ Moulinex
- ○ Philips
- ○ Bauknecht
- ○ Indesit
- ○ Miele
- ○ Ander merk

Welk merk pindakaas gebruikt u?
- ○ Calvé
- ○ Gouda's Glorie
- ○ AH huismerk
- ○ Ander huismerk
- ○ Ik gebruik geen pindakaas

Ten slotte nog enkele vragen over u en de samenstelling van uw gezin.

U bent:
- ○ alleenstaand
- ○ samenwonend
- ○ getrouwd, zonder kinderen
- ○ getrouwd, met kinderen
- ○ weduwe/weduwnaar
- ○ gescheiden.

Hoe oud bent u ? .....jaar.

U bent:
○ man
○ vrouw

Uw beroep is .....

Indien u kinderen heeft, hoeveel kinderen heeft u en hoe oud zijn zij?
Ik heb ..... kinderen in de leeftijd van ..... tot ..... jaar.

Indien u een partner heeft:
a. hoe oud is deze? ..... jaar
b. geslacht:
○ man
○ vrouw
c. beroep partner: .....

Hartelijk dank voor het invullen van deze vragenlijst.

Wilt u nog even uw adres en telefoonnummer invullen in verband met het toesturen van uw boodschappenpakket!

naam .....
straat en huisnummer .....
postcode en woonplaats .....
telefoonnummer.....-.....

Uitwerkingen

17.1 De informatie omtrent het onderzoek is zeer summier, zodat eigenlijk allerlei bedreigingen van de populatie- en ecologische validiteit een rol zouden kunnen spelen. Opvallend is dat er niets wordt vermeld over de (aselecte?) keuze van de meisjes. De populatie waarnaar gegeneraliseerd moet worden (de doelpopulatie) is daarmee in feite onbekend.
Verder is het mogelijk dat de herhaalde-metingenopzet betekent dat de kinderen aan méér dan één behandelingsniveau werden blootgesteld (dus: én aan lage, én aan gematigde én aan hoge beloning). Als dat zo is, is het onderzoeksresultaat – een positief effect van gematigde beloning – wellicht alleen generaliseerbaar naar meisjes die meervoudige behandelingen hebben ondergaan en mag niet worden geconcludeerd dat gematigde beloning (één van de drie behandelingsniveaus) afzonderlijk ook effectief is. In deze onderzoeksopzet zou dus sprake kunnen zijn van 'multiple treatment interference'.

17.2 Ten aanzien van de externe validiteit van de resultaten uit de rattensteekproef geldt de bedreiging dat het meestal niet toegestaan is om te generaliseren over soorten. Dat wil zeggen dat resultaten van experimenten bij dieren niet zonder meer gelden voor menselijk gedrag. Eerst dient vastgesteld te worden of het bestudeerde gedrag bij mens en dier op vergelijkbare wijze door antidepressiva zal worden beïnvloed.
De externe validiteit van de resultaten uit de studentensteekproef wordt mogelijk bedreigd door:

1 Beperkte financiële middelen: studenten zijn relatief goedkoper en makkelijk te werven proefpersonen.
2 Zelfselectie en vrijwilligerseffecten: je kunt studenten voor een dergelijk 'vervelend' onderzoek naar het effect van antidepressiva niet aselect kiezen. Dit is ethisch niet verantwoord. Je zult ze moeten informeren over het doel van het onderzoek en bent daarmee aangewezen op vrijwilligers die bijvoorbeeld tegen een forse financiële tegemoetkoming mee willen doen aan zo'n experiment.

17.3 De bedreiging van de externe validiteit van deze onderzoeksaanpak ligt vooral op het gebied van de populatievaliditeit. De vele persoonlijke vragen, de lengte van de vragenlijst en het gratis boodschappenpakket zorgen ervoor dat waarschijnlijk bepaalde groepen juist wel reageren en andere juist niet.
Degenen die voor de belastingdienst vrezen, en goederen bezitten die niet geregistreerd staan bij deze dienst, zullen óf de vragenlijst niet invullen (bedreiging populatievaliditeit) of deze niet naar waarheid invullen.
Daarnaast zullen mensen die het financieel niet al te breed hebben, eerder reageren op de vragenlijst, omdat ze kans maken op een gratis boodschappenpakket. Ook dit is een bedreiging van de populatievaliditeit. Als we ervan uitgaan dat men als doelpopulatie de Nederlandse bevolking heeft, dan zal de experimenteel toegankelijke populatie daar nogal van afwijken. Beter is misschien antwoord op deze vragen te krijgen door middel van interviewers in supermarkten of op straat, zodat de respondenten anoniem kunnen blijven. Het

boodschappenpakket kan dan direct overhandigd worden, zonder dat de respondenten naam, adres en telefoonnummer op hoeven te geven.

18 Gemengde opgaven over de totale stof

De volgende opgaven zijn bedoeld als extra oefenmateriaal. Zij bestrijken de totale stof die in het oefenboek aan de orde is gekomen. De vragen worden gesteld naar aanleiding van de tekst van een artikel van D.B. Baarda, M.P.M. de Goede en G.E.M. van der Meulen (1986). De door ouders geschatte kans op werk voor hun kinderen. *Pedagogische Studiën*, *63*, 74-80.
Het verdient aanbeveling de tekst van het artikel enkele malen zorgvuldig door te lezen en te bestuderen!

De oorspromnkelijke paginering van het artikel staat tussen haakjes vermeld achter de doorlopende paginering.

Vooraf enige toelichting bij enkele in het artikel gebruikte termen:

- In 3.1 op p. 76 (1e kolom) en verderop wordt vermeld dat een variabele als *indicator* wordt gebruikt. Hiermee wordt bedoeld dat die variabele als *operationalisatie* van het (theoretische) begrip wordt gebruikt.
- Op p. 78 (2e kolom) wordt gewag gemaakt van een *stapsgewijze multiple regressie*. Dat komt in dit geval overeen met wat elders als hiërarchische multipele regressie wordt aangeduid.

De door ouders geschatte kans op werk voor hun kinderen

D. B. BAARDA, M. P. M. DE GOEDE EN G. E. M. VAN DER MEULEN
Instituut voor Pedagogische en Andragogische Wetenschappen Rijksuniversiteit Utrecht

Samenvatting

In dit artikel staat de vraag centraal hoe ouders van eind lagere-school- en LTS-leerlingen de kans op werk voor hun kinderen schatten en met welke factoren dit samenhangt. Het blijkt dat de ouders over het algemeen nogal pessimistisch zijn over de verwachte kans op werk voor hun kinderen. Van de ouders van de lagere-schoolleerlingen acht 36% de kans op werk voor hun kinderen klein tot zeer klein, voor de LTS is dat 41%.

De verwachte kans op werk blijkt bij ouders van de lagere-schoolkinderen vooral samen te hangen met de schoolprestaties van de kinderen en in mindere mate met de sociaal-economische status van het gezin en hun ervaring met werkloosheid. Werkloze ouders uit de laagste sociaal-economische milieus taxeren de kans op werk voor hun kinderen het ongunstigst.

1 *Inleiding*

Eind 1983 hebben wij in Pedagogische Studiën (Baarda e.a., 1983) verslag gedaan van een verkennend onderzoek naar de schoolprestaties van kinderen van werkloze vaders. Wij constateerden dat er sprake is van een tendens tot teruggang in leerprestaties van kinderen op het moment dat hun vader werkloos wordt. Dit was niet het geval bij de nameting van de kinderen van werkenden. de achteruitgang is sterker naarmate de vader langer werkloos is.

Het is zinvol om te constateren dat er verschillen bestaan. Veel belangrijker is het om na te gaan waardoor deze verschillen veroorzaakt worden. Daartoe zijn wij een onderzoek begonnen dat zich richt op de vraag welke veranderingen er in een gezin plaats vinden wanneer een vader werkloos wordt (Baarda e.a., 1985). In dit artikel lichten wij er alvast één aspect uit, namelijk de verwachtingen die ouders hebben over de kans op werk voor hun kinderen. Wij vermoeden namelijk dat deze verwachting wel eens van invloed zou kunnen zijn op de leermotivatie van kinderen en als zodanig mogelijkerwijze één van de factoren, die een achteruitgang in leerprestaties bij kinderen van werklozen zou kunnen verklaren. Het is waarschijnlijk dat ouders die met werkloosheid geconfronteerd worden niet alleen pessimistisch zijn over de kans op het verkrijgen van werk voor zichzelf, maar ook de kans op werk voor hun kinderen relatief laag schatten. Deze verwachting zal de kinderen niet ontgaan en mogelijk hun leermotivatie en daarmee ook hun leerprestaties beïnvloeden. Aan de hand van onderzoeksmateriaal dat zowel van ouders van lagere schoolkinderen als van ouders van LTS-leerlingen is verkregen, willen wij nagaan hoe groot ouders de kans op werk voor hun kinderen schatten en welke factoren daarop van invloed zijn. Bij die factoren zullen wij niet alleen letten op ervaringen met werkloosheid, maar ook op de sociaal-economische status van het gezin, geslacht van het kind en schoolresultaten.

De *probleemstelling* van ons onderzoek luidt dan ook: Welke factoren bepalen de verwachting van ouders, over de kans op werk voor hun kinderen?

2 *Theoretische oriëntatie en verwachtingen*

Om bovengenoemde probleemstelling te onderzoeken is geen aparyt onderzoek opgezet. Er is gebruik gemaakt van bestaand onderzoeksmateriaal. Wij zijn daardoor enigszins beperkt in de exploratie van mogelijke samenhangen tussen het centrale concept binnen dit onderzoek, nl. de door ouders geschatte kans op werk voor hun kinderen en enkele achtergrondskenmerken. Het aanwezige materiaal bevat echter een aantal variabelen die voor onze probleemstelling zeer relevant zijn.

Een van die variabelen is de *sociaal-economische status* van het gezin. In de discussie over de

kans op succes in de maatschappij komt het sociaal-economisch milieu regelmatig ter sprake. De kansen van kinderen uit lagere sociaal-economische mileus in school en beroep zouden nog altijd kleiner zijn dan die van kinderen uit de hogere milieus (Peschar, 1979). belangrijk is de vraag of de ouders en kinderen uit de lagere sociaal economische milieus een lagere verwachting van zichzelf hebben en daarom mogelijk minder presteren, of dat de maatschappij gerepresenteerd door bijv. onderwijzers en leerkrachten een lagere verwachting van hen hebben waardoor zij mogelijk anders beoordeeld worden. Uit een onderzoek van Jungbluth (1984) blijkt dat de leerkrachten inderdaad een andere verwachting hebben ten aanzien van kinderen uit lagere milieus. De verwachtingen met betrekking tot het cognitief-intellectueel prestern van leerlingen bleken bij de door hem benaderde leerkrachten onder andere afhankelijk te zijn van het beroepsniveau van de vader. Meijnen (1979) constateert dat ook de verwachtingen van de ouders zelf over de schoolprestaties van hun kinderen statusgebonden zijn. Oders met een hogere sociaal-economische status hebben hogere verwachting dan ouders met een lagere sociaal economische status, hetgeen van invloed zal zij op de prestaties.

Daarom zijn wij ook benieuwd of er sprake is van milieuverschillen, waar het gaat om de verwachte kans op werk voor de kinderen. Wij veronderstellen dat er een positieve samenhang tussen het sociaal-economische milieu van de ouders en de kans op werk die zij voor hun kinderen schatten.

Daarnaast zijn wij geänteresseerd of de *vader al dan niet werk* heeft. Men mag immers verwachten dat mensen die zelf geen werk hebben, de kans opwerk voor anderen - waaronder hun kinderen - ook lager zullen schatten dan anderen. Het is mogelijk dat ouders zich in hun verwachtingen omtrent de kans op werk voor hun kinderen ook laten leiden door de *schoolprestaties* van hun kinderen. De hoogte van de opleiding is maatgevend voor de hoogte van de maatcshappelijke positie. Mensen met een hoge opleiding genieten nog altijd het hoogste beroepspretige en onder hen komt werkloosheid relatief minder vaak voor (Rapportage arbeidsmarkt, 1984). Verondersteld wordt dat naarmate de schoolresultaten van kinderen beter zijn, de verwachte kans op werk groter is.

Niet alleen de hoogte maar ook het *soort opleiding* die kinderen volgen kan mogelijk van invloed zijn op de verwachte kans op werk. Zo is bekend dat in de bouw relatief veel werkloosheid heerst, maar dat er wel veel vraag is naar electrotechnici. Eén van de deelonderzoeken vond plaats op LTS-sen. Op LTS-sen kunnen leerlingen een keize maken uit verschillende richtingen, o.a. bouw en timmeren, mechanische techniek, schilderen, meubelmaken, electrotechniek en motorvoertuigen. Het is interessant na te gaan of er een samenhang bestaat tussen de door ouders geschatte kansop werk voor hun kinderen en de richting die door de kinderen gekozen is.

Niet bij de LTS-leerlingen, dit zijn bijna allemaal jongens, maar wel bij de lagere schoolkinderen zou *het geslacht* van het kind wel eens verschillen op kunnen leveren bij de door de ouders verwachte kans op werk. Nog steeds lijkt werk voor jongens belangrijker dan voor meisjes. Door ouders wordt meer belangstelling getoond voor de beroepskeuze voor hun zoon dan voor hun dochter (Kropman en Collaris, 1974). Wij verwachten dan ook dat ouders de kans op werk voor dochters lager schatten dan voor zonen.

3 *Opzet en uitvoering van het onderzoek*

3.1 *Operationalisaties*

De afhankelijke variabele, de door ouders geschatte kans op werk voor hun kinderen, is geoperationaliseerd in de vraag: 'Hoe groot acht U de kans dat uw kind na het doorlopen van zijn (vervolg)opleiding werk zal kunnen vinden?' De ouders konden kiezen uit 5 antwoordmogelijkheden variërend van zeer groot tot zeer klein.

Sociaal-economisch milieu is geoperationaliseerd met de vraag naar het beroep van de vader. De beroepen zijn aan de hand van de door Van Westerlaak e.a. (1976) gemaakte beschrijvingen, ingedeeld in zes klassen.

Iemand is in het onderzoek als werkloos opgevat, wanneer hij minstens 3 maanden als werkzoende bij het arbeidsbureau staat ingeschreven. Er

is voor een periode van minstens 3 maanden gekozen, omdat de werkloosheid dan geen incidenteel karakter meer heeft. Voor de schoolprestaties zijn de rapportbeoordelingen gebruikt. Als er van een verbale in plaats van numerieke beoordeling sprake was, hebben wij getracht deze omwille van het onderzoek toch in cijfers om te zetten. Rapportbeoordelingen op de verschillende vakken intercorreleren vrij hoog. De gemiddelde pearson-productmoment correlatie tussen de cijfers tussen de verschillende vakken bedraagt namelijk .64. Daarom is besloten om de cijfers voor één vak te gebruiken. Aangezien de rapportcijfers voor rekenen het hoogst met de andere cijfers correleren hebben wij die als indicator genomen. Om niet al te zeer afhankelijk te zijn van toevallige beoordelingselementen, is het geniddelde over de laatste vier rapportcijfers genomen.

Slechts van een beperkt deel van de kinderen zijn evenwel de rapportbeoordelingen bekend. Als vervanging wordt bij de kinderen waarbij deze beoordelingen niet bekend zijn, de verwachte vervolgopleiding na de lagere school als indicator voor de schoolprestaties genomen. Men mag immers aannemen dat de verwachte vervolgopleiding vooral gebaseerd zal zijn op de door het kind geleverde schoolprestaties. Voor de LTS-leerlingen, waarvan de schoolprestaties ook niet bekend zijn zal het niveau van hun opleiding als indicator genomen worden. De LTS kent namelijk drie opleidingsniveaus die verschillen qua moeilijkheidsgraad.

3.2 *De onderzoekspopulatie*

Wij maken gebruik van materiaal dat ontleend is aan twee verschillende onderzoeken. Het eerste is het onderzoek waarover wij in dit blad reeds gerapporteerd hebben (baarda e.a., 1983), dat werd gehouden vijfde- en zesde-klassers van lagere scholen in 's Hertogenbosch en hun ouders. Er werd een aselecte steekproef getrokken uit het bestand van lagere scholen in deze plaats. Uiteindelijk zijn 11 scholen in het onderzoek betrokken. Aan alle ouders (928) van de leerlingen uit de vijfde- en zesde-klas is een brief gestuurd, waarin om toestemming werd gevraagd voor het inzien van de schoolgegevens van hun kind. Deze brief bevatte ook een aantal vragen betreffende de werksituatie van de vader, sociaal-economische status van het gezin, de door hen geschatte kans op werk voor hun kind e.d. (zie par. 2). Voor onze analyse maken wij gebruik van de 583 geretourneerde formulieren. Niet van al deze kinderen zijn de schoolbeoordelingen verzameld. Dit is alleen gedaan voor de 60 kinderen die voor dit gedeelte van het onderzoek in aanmerking kwamen. Dit zijn 30 kinderen van werkloze vaders en 30 qua leeftijd, geslacht, schoolloopbaan, sociaal-economisch milieu en het al dan niet werken van de moeder vergelijkbare kinderen van werkende vaders.

Het andere onderzoek heeft plaatsgevonden bij vierde-klas LTS-leerlingen. Dit is een onderzoek naar de beleving van de toekomst en in het bijzonder toekomstig werk bij kinderen die in hun gezin al of niet te maken hebben gehad met werkloosheid van de vader, een broer of zus. Zes LTS-scholen in en rond Amersfoort werkten mee. Ook aan de ouders van de leerlingen van deze scholen werd een brief geschreven, waarin gevraagd werd om toestemming voor het onderzoek. De brief bevatte verder dezelfde vragen als bij het onderzoek in 's Hertogenbosch. Alleen het responspercentage was hier veel lager; van de 947 aangeschreven ouders reageerden slechts 338 ouders, waarvan 89 te kennen gaven bezwaar te hebben tegen het onderzoek. Er resteerden toen nog 249 ouders. De betrekkelijk geringe respons bij de LTS-leerlingen noopt tot voorzichtigheid bij de interpretatie van de resultaten.

4 *Resultaten*

4.1 *De geschatte kans op werk*

Zowel de ouders van de lagere-schoolkinderen als de ouders van de LTS-leerlingen lijken nogal pessimistisch, waar het gaat om kans op werk voor hun kinderen. Van de ouders met kinderen op de lagere-school zegt 36% en van de ouders met kinderen op de LTS 41% dat de kans op werk voor hun kinderen klein tot zeer klein is (zie Tabel 1). Er is geen sprake van een significant verschil ($t = 1.49/ p > .05$) in de gemiddelde geschatte kans op werk tussen de ouders van de lagere-schoolkinderen (gem. = 2.75) en de ouders van de LTS-leerlingen (gem. = 2.63).

Tabel 1 *De door ouders van lagere-school kinderen en ouders van LTS-leerlingen geschatte kans op werk voor hun kinderen*

	lagere-school		LTS	
	n	%	n	%
zeer groot	13	(2)	13	(6)
groot	78	(14)	20	(9)
niet groot, niet klein	266	(48)	95	(44)
klein	128	(23)	54	(25)
zeer klein	70	(13)	35	(16)
Totaal	555	(100)	217	(100)

4.2 *De samenhang van de geschatte kans op werk met de onafhankelijk variabelen afzonderlijk*

Sociaal-economische status

Bij de lagere is er een duidelijk verband tussen de sociaal-economische status en de geschatte kans op werk ($r = .26/p < .0001$). Ouders met een middelbaar of hoger beroepsniveau schatten de kans op werk voor hun kinderen beduidend hoger; 26% verwacht dat de kans op werk voor hun kind groot tot zeer groot is. Bij de ouders uit de sociaal-economisch laagste klassen (de geschoolde en de ongeschoolde arbeiders) is dit percentage 10. Bij de LTS-leerlingen is er nauwelijks een verband tussen de sociaal-economische status van de ouders en de door hen geschatte kans op werk voor hun kinderen ($r = .10/p = .07$).

Ervaringen met werkloosheid

Mensen die langer werkloos zijn dan een jaar blijken veel pessimistischer dan mensen die korter werkloos zijn (zie Tabel 2). Ook arbeidsongeschikten zijn over het algemeen wat pessimistischer. Mensen die ooit werkloos zijn geweest blijken niet of nauwelijks pessimistischer te zijn dan mensen die nooit werkloos zijn geweest.

Voor de lagere-schoolleerlingen zijn de verschillen significant ($F = 3.82/p = .005$). Dit geldt niet voor de LTS-leerlingen; hoewel werklozen en arbeidsongeschikten de kans op werk voor hun kinderen gemiddeld wat lager schatten, is dit verschil toch niet significant ($F = 1.44/p = .24$). Daarom wordt in de verdere analyses bij de variabele ervaring-met-niet-werken een tweedeling aangehouden, t.w.: 1. de mensen die een jaar of langer werkloos zijn en de arbeidsongeschikten, en 2. de overige respondenten. Voor de LTS-leerlingen behoren alle werkloze vaders tot de eerste groep omdat daar niet specifiek naar de duur is gevraagd. Wel is het zo dat alle mensen die tot de categorie 'werkloos' gerekend worden minstens drie maanden werkloos zijn.

Schoolprestaties, vervolgopleiding, soort opleiding en niveau van de opleiding

Bij de 60 lagere-schoolkinderen, waarvan de rapportbeoordelingen bekend zijn, is de correlatie tussen hun schoolprestaties en de door hun ouders geschatte kans op werk relatief hoog ($r = .40/p = .001$). Dit geldt ook voor het niveau van de vervolgopleiding die de kinderen volgens de verwachting van hun ouders gaan volgen; de correlatie met de geschatte kans op werk is .56 ($p < .001$). Voor deze verbanden geldt: hoe beter de schoolprestaties, des te hoger de vervolgopleiding en des te hoger de geschatte kans op werk. Het verwachte

Tabel 2 *De door ouders geschatte kans op werk voor hun kinderen gerelateerd aan hun ervaringen met werkloosheid (1 = zeer klein - 5 = zeer groot)*

	lagere school			LTS		
	n	gem.	spr.	n	gem.	spr.
werkend	404	2.82	.89	187	2.66	1.02
was werkloos	32	2.75	.84	14	2.50	1.09
is werkloos*				16	2.31	.95
is werkloos < 1 jaar	17	2.76	1.25			
is werkloos > 1 jaar	29	2.28	1.10			
arbeidsongeschikt	26	2.31	1.01	14	2.36	.74

* voor LTS is duur werkloosheid niet bekend

niveau van de vervolgopleiding hangt namelijk sterk samen met de schoolprestaties (r = .55/ p = .001).

Ook voor de totale populatie lagere-schoolkinderen is de correlatie tussen de verwachte vervolgopleiding en de door de ouders geschatte kans op werk relatief hoog (r = .34/p < .0001).

Bij de LTS-leerlingen correleert het opleidingsniveau nauwelijks (r = .07/p = .15) met de verwachte kans op werk. Ook het soort opleiding dat zij volgen (bouw, mechanische techniek e.d.) laat geen verschillen zien in de verwachte kans op werk (F = 1.62/p = .16).

Geslacht

Alleen bij de lagere-schoolleerlingen was het mogelijk te bekijken of er verschil is in de verwachte kans op werk tussen jongens en meisjes. In onze LTS-groep komen namelijk slechts 7 meisjes voor. De ouders van de lagere schoolkinderen blijken echter de kans op werk voor hun dochters niet anders in te schatten dan voor hun zoons (t = .58/p = .57).

4.3 *De samenhang van de geschatte kans op werk met de onafhankelijke variabelen gezamenlijk*

Bij de LTS-leerlingen is er bij geen van de onafhankelijke variabelen sprake van een significant verband met de verwachte kans op werk. Verder analyse heeft hier dan ook geen zin. Bij de lagereschoolkinderen zijn er wel significante verbanden met de sociaal-economische status, de ervaringen met werkloosheid en de schoolprestaties. Het is echter de vraag in hoeverre deze aspecten elkaar overlappen. Daarom hebben wij voor deze drie variabelen het verband met de geschatte kans op werk nog eens berekend, maar nu met de methode van de partiële correlatie. Voor het niveau van de schoolprestaties hebben we niet de schoolprestaties zelf, maar de verwachte vervolgopleiding genomen. Die correleert namelijk hoog met de schoolprestaties en daar hebben wij voor vrijwel alle kinderen de gegevens van.

Tabel 3 laat zien dat vooral de verbanden tussen respectievelijk sociaal-economische status en de revaring met niet-werken met de geschatte kans op werk vrijwel verdwijnen, als er gecorrigeerd wordt voor de andere variabelen. De drie variabelen vertonen namelijk onderlinge samenhang. De productmoment correlatie tussen sociaal-economische status met achtereenvolgens de ervaring met niet-werken en het niveau van de verwachte vervolgopleiding is –.22 en .43. De correlatie tussen de ervaring met niet-werken en het niveau van de verwachte vervolgopleiding is –.17. (het gaat hierbij om de lagere school kinderen, n = 411). Deze correlaties zijn significant op minstens 1%-niveau.

Tabel 3 *Pearson-productmoment correlaties en partiële correlaties van drie onafhankelijke variabelen met de door de ouders geschatte kans op werk voor lagereschoolkinderen (n = 411)*

	correlatie		partiële corr.	
	r	p	r	p
soc.-economische status	.22	.001	.08	.05
ervaring niet-werken	–.14	.003	–.07	.07
niveau verwachte vervolgopl.	.32	< .001	.25	< .003

Uit een stapsgewijze multiple regressieanalyse blijkt dat de verwachte vervolgopleiding de meeste gemeenschappelijke variantie verklaart. Het niveau van de verwachte vervolgopleiding wordt als eerste variabele in de analyse opgenomen; deze verklaart 10% van de variantie van de geschatte kans op werk. De beide andere variabelen (SES en ervaring met niet-werken) voegen daar samen nog slechts één procent aan toe.

5 *Discussie*

Opvallend is het sombere perspectief dat ouders hebben over de kans op werk voor hun kinderen. Ongeveer 40% acht de kans dat hun kinderen werk vinden klein tot zeer klein. Een belangrijke vraag is in hoeverre die sombere verwachting van ouders reëel is. Als wij naar het percentage werklozen onder de schoolverlaters in 1984 kijken, blijkt dat ouders over het algemeen wel erg somber zijn. De door het ministerie van Sociale zaken en

Werkgelegenheid berekende kans op langdurige werkloosheid (meer dan 12 maanden) voor deze schoolverlaters is 18%. Het totale beeld van de werkloosheid en vooral de jeugdwerkloosheid geeft echter ook geen reden tot optimisme. het percentage langdurig werklozen onder schoolverlaters neemt namelijk steeds meer toe. Van de bij het arbeidsbureau geregistreerde schoolverlaters was in 1981 na 6 maanden nog 45% ingeschreven, in 1983 waren dat er al 58% (Schoolverlatersbrief 1984).

Ouders blijken zich voor de kans op werk voor hun kinderen vooral te baseren op de schoolprestaties van hun kinderen. Het is de vraag in hoeverre dit reëel is. Het percentage werklozen in 1983 onder de hoogst opgeleiden (7%) in Nederland is lager dan onder de laagst opgeleiden (21%) (Rapportage arbeidsmarkt 1984, blz. 42). Naast de schoolprestaties speelt ook het sociaal-economische milieu een rol bij de verwachtingen over de kans op werk. In de hogere milieus is men optimistischer. Een (methodisch) probleem is hier, dat er tussen de LTS-leerlingen nauwelijks verschil is wat betreft de sociaal-economische status van het gezin. De LTS-leerlingen vormen zowel qua sociaal-economisch milieu als qua schoolprestatie een homogenere groep dan de lagere-schoolleerlingen, die na de zesde-klas naar diverse vervolgopleidingen overstappen. Daardoor zijn er tussen de LTS-leerlingen nauwelijks verschillen te constateren op de onderzoeksvariabelen. Dat de ouders uit de sociaal-economisch hogere milieus optimistischer zijn, wordt waarschijnlijk mede bepaald door het feit dat in deze milieus werkloosheid relatief minder vaak voorkomt. Verder volgen de kinderen uit de hogere milieus zelf ook relatief vaker hogere opleidingen (Vlaskamp en Van Westerlaak, 1977, blz. 24).

De ongelijkheid wat betreft de toekomstkansen van kinderen uit de sociaal-economisch lagere milieus in vergelijking met kinderen uit hogere milieus lijkt nog steeds niet opgeheven. Het gaat om kansen zoals gepercipieerd door de ouders. Het is waarschijnlijk dat er een samenhang is tussen deze gepercipieerde kansen en de feitelijke mogelijkheden straks. Deze ongelijkheid in kansen wordt waarschijnlijk nog bevorderd door de verwachting van ouders uit sociaal-economisch lagere milieus dat de kans op werk voor hun kinderen kleiner is. Dat zal hun kinderen wellicht minder motiveren om zich extra in te zetten op school, hetgeen hun achterstand nog eens kan versterken.

Ook de ervaringen met al of niet-werken spelen vanzelfsprekend mee in de verwachte kans op werk. Die ervaringen zijn echter wel enigszins milieugebonden. Ouder uit de lagere sociaal-economische milieus schatten de kans op werk voor hun kinderen al relatief laag; wanneer zij werkloos zijn wordt dit nog ongunstiger. Het is van belang na te gaan of dat van invloed is op de leermotivatie van kinderen, en zo ja in welke richting dat gaat. Door werkloosheid van hun vader kunnen kinderen gedemotiveerd raken; het is echter ook mogelijk dat kinderen van werklozen sterker gemotiveerd worden tot het leveren van goede schoolprestaties, juist om te voorkomen dat men later – evenals vader – geen baan krijgt of werkloos wordt. Leermotivatie, ons inziens een centrale variabele, konden wij niet in onze analyses betrekken. Wij maken immers gebruik van bestaand onderzoeksmateriaal, waarin deze variabele niet voorhanden is. Het is duidelijk dat een longitudinaal onderzoek geändiceerd is, om een meer gefundeerde uitspraak te kunnen doen over de veronderstelling 1dat ervaringen met niet-werken' en 'een laag geschatte kans op werk' de leermotivatie en dientengevolge ook de leerprestaties ongunstig beänvloeden. Om de onderzoeksresultaten die hier gepresenteerd zijn adequater te kunnen interpreteren, zal tevens in vervolgonderzoek uitgebreider en diepgaander ingegaan worden op de opinies van ouders over de verwachte kansen op werk voor hun kinderen.

Literatuur

Baarda, D. B., A. P. M. Frowijn, M. P. M. de Goede en M. E. Postma, Schoolprestaties van kinderen van werkloze vaders. *Pedagogische Studiën*, 1983, *60*, 473-484.

Baarda, D. B., A. P. M. Frowijn, M. P. M. de Goede en M. E. Postma, *Kinderen van werklozen; Onderzoeksopzet*. Utrecht: RUU-IPAW, 1985

Jungbluth, P., Herkomst van milieu en sekse van leerlingen in relatie tot leerkrachtverwachtingen; impliciete selectie in het onderwijs. *Pedagogische Studiën*, 1984, *61*, 402-411.

Kropman, J. A. en J. W. M. Collaris, *Van jaar tot jaar*. Nijmegen: Instituut voor Toegepaste Sociologie, 1974.

Meijnen, G. W., *Maatschappelijke achtergronden van intellectuele ontwikkeling*. Groningen: Wolters-Noordhoff, 1979.

Peschar, J. L., *Milieu-school-beroep*. Groningen, 1975.

Rapportage arbeidsmarkt 1984. 's-Gravenhage: Ministerie van Sociale Zaken en Werkgelegenheid, 1984.

Schoolverlatersbrief 1984. 's-gravenhage: Ministerie van Sociale Zaken en Werkgelegenheid, 1984.

Vlaskamp, F. J. M. en J. M. van Westerlaak, Beroepskeuze in het spanningsveld tussen onderwijs en arbeidsmarkt. In: J. A. M. van Wezel (red.). *Arbeidsmarkt in beweging*. 's-Gravenhage: VUGA, 1977.

Westerlaak, J. M. van, J. A. Kropman en J. W. M. Collaris, *Beroepenklapper*. Nijmegen: Instituut voor Toegepaste Sociologie, 1976.

Summary

Baarda, D. B., M. P. M. de Goede, G. E. M. van der Meulen. 'The chance that children will get a job in the future as estimated by their parents'. *Pedagogische Studiën*, 1986, *63*, 74-80.

About 600 parents of Primary school pupils and about 250 parents of Elementary technical school pupils filled in a questionnaire concerning the chance that their children will get a job in the future. In general Parents are rather pessimistic. Of the parents of Primary school pupils 36% estimate the chance for their children to get a job in the future small to very small; for the parents of the Elementary technical schools that percentage is 41. The estimated chance correlates positiveky with the school reports of the children. There is also a slight positive correlation between the estimated chance and the social economic status of the family and a low negative correlation between that estimated chance and being unemployed. In short, unemployed parents of the lowest social economic classes estimate the chance as most unfavourable. The possible implication of these results for the school and the schoolcareers of children are discussed. In the group of Elementary technical schools we did not find significant correlations. However, this category of pupils is rather homogeneous with regard to the schoolreports as well as the social ecenomic status of the family.

Curricula vitae

D. B. Baarda en M. P. M. de Goede zijn als docent verbonden aan het Instituut voor Pedagogische en Andragogische Wetenschappen der Rijksuniversiteit Utrecht.

G. E. M. van der Meulen participeerde in het kader van zijn doctoraalstudie klinische pedagogiek in het onderzoek waarvan hier verslag wordt gedaan.

Correspondentieadres: Heidelberglaan 1, 3584 CS Utrecht.

Manuscript aanvaard 27-8-'85

Opgaven

18.1 Leg uit waarom het onderzoek van Baarda c.s. niet kan worden omschreven als een (quasi-)experiment.

18.2 Om welk type onderzoek gaat het dan wel?

18.3 Op p. 76 (2e kolom, 2e alinea) wordt beschreven hoe de steekproef van lts-leerlingen tot stand is gekomen. Licht aan de hand van de beschrijving toe welke bedreiging(en) van de externe validiteit in dit (deel)onderzoek met name een rol zou(den) kunnen spelen. Licht uw antwoord toe met argumenten.

18.4 In sectie 2 (p. 74/75) worden 5 inhoudelijke hypothesen geopperd. Schrijf deze op en bepaal of ze een- of tweezijdig getoetst moeten worden en geef bij elke hypothese aan met welke toets deze hypothese getoetst moet worden. Beargumenteer de toetskeuze.

18.5 Op p. 77 (2e kolom, 3e alinea) wordt een correlatie van $r = 0{,}40$ gegeven. Toets deze correlatie volgens de geijkte procedure met $\alpha = 0{,}01$ en geef een voor een leek begrijpelijke conclusie van het resultaat, waarin woorden als populatie, steekproef en correlatie niet mogen voorkomen.

18.6 In paragraaf 4.2 (p. 78, paragraaf over *Geslacht*) wordt een t-toets uitgevoerd om één van de gestelde hypothesen te toetsen ($t = 0{,}58$; $p = 0{,}57$). Wat is het aantal vrijheidsgraden bij deze toets? Vanwege het grote aantal vrijheidsgraden mag deze t-waarde in de z-tabel (tabel A) opgezocht worden. Gebruik de t-waarde (is dus gelijk aan een z-waarde) om de bijbehorende p-waarde op te zoeken in de z-tabel. Vergelijk de gevonden waarde met die van de onderzoekers en geef aan waar zij een fout hebben gemaakt.

18.7 In paragraaf 4.2 (p. 77 onderaan) wordt onder het kopje '*Sociaal-economische status*' vermeld dat er bij de lagereschoolkinderen een '*duidelijk verband*' bestaat tussen SES en de afhankelijke variabele. Zou u, gelet op het probleem van statistische en inhoudelijke significantie, net als de onderzoekers concluderen dat in dit geval sprake is van een 'duidelijk verband'?

18.8 Tabel 2 (p. 77 onderaan) is door ons enigszins 'bewerkt' en ziet er nu als volgt uit (zie volgende bladzijde):

a Beschouw deze gegevens als een tweewegvariantieanalyseopzet. Om welke factoren gaat het in dit geval en hoeveel niveaus heeft elke factor?
b Zijn het experimentele of classificerende factoren? Beargumenteer uw antwoord.
c Teken het interactieplaatje behorende bij de tweewegvariantieanalyse. Zorg dat alle benodigde uitleg in het plaatje is terug te vinden.
d Geef een toelichting op de uitkomsten zoals die uit het plaatje te lezen zijn.

Onze Tabel 2: *De door ouders geschatte kans op werk voor hun kinderen, gerelateerd aan hun ervaringen met werkloosheid (1 = zeer klein tot 5 = zeer groot).*

Type school	*Lagere school*			*lts*		
Ervaringen met werkloosheid	n	$\bar{x}$	S_x	n	$\bar{x}$	S_x
1 werkend	404	2,81	0,89	187	2,66	1,02
2 was werkloos	32	2,75	0,84	14	2,50	1,09
3 werkloos (> 3 mnd).	46	2,46	1,15	16	2,31	0,95
4 arbeidsongeschikt	26	2,31	1,01	14	2,36	0,74

(In tabel 2 van het artikel betekent *spr.* spreiding en wordt de standaardafwijking bedoeld.)

18.9 Ga nu terug naar de oorspronkelijke Tabel 2 uit het artikel en **toets** voor de (ouders van) lagere schoolkinderen de hypothese:
Mensen die langer werkloos zijn dan een jaar, zijn pessimistischer dan mensen die korter werkloos zijn

a Formuleer de inhoudelijke nulhypothese.
b Formuleer de formele hypothesen.
c Kies een $\alpha = 0{,}05$ en formuleer de beslissingsregels.
d Bereken de geobserveerde waarde.
e Formuleer de conclusie.

18.10 Beargumenteer waarom (waarschijnlijk) SES en 'ervaring niet werken' zo weinig verklaarde variantie toevoegen, nadat 'niveau verwachte vervolgopleiding' als eerste variabele in de multipele regressieanalyse is opgenomen.

18.11 *a* Schets/construeer voor de lagereschoolkinderen een 'causaal model' met betrekking tot de relatie(s) tussen *'geschatte kans op werk'* en ten minste drie onafhankelijke variabelen. Maak hierbij ook gebruik van p. 74, 75 en 78 van het artikel.
b Beargumenteer uw keuze.
c Geef, voorzover bekend, de grootte van de correlaties bij de pijlen aan.

Uitwerkingen

18.1 Het onderzoek is niet te karakteriseren als een (quasi-)experiment. Er is immers in geen enkel opzicht sprake van manipulatie van variabelen: de onafhankelijke variabelen zijn *classificerende* variabelen.

18.2 Het onderzoek is te omschrijven als (een poging tot) causale analyse vanuit passieve observatie.

18.3 Diverse antwoorden kunnen gegeven worden. Hieronder een aantal belangrijke bedreigingen van de externe validiteit:

a De scholen werkten mee op basis van vrijwilligheid; hierdoor zou een bedreiging van de populatievaliditeit kunnen ontstaan. Welke precies, hangt af van de kenmerken van de scholen die wel en niet meededen. Zijn ze representatief voor dit type scholen?

b Dit vrijwilligersaspect speelt nog sterker bij de zelfselectie van de ouders, daar bleven immers 249 van de 947 (slechts 26 %!) over voor het onderzoek. De mededeling dat 89 bezwaar maakten tegen het onderzoek, versterkt de indruk dat er een sterke zelfselectie heeft plaatsgevonden.

c Men kan de vraag stellen of lts-en uit de omgeving van Amersfoort representatief zijn voor alle lts-en in Nederland. Zonder verdere inhoudelijke argumentatie lijkt het overigens niet nodig zwaar aan die representativiteit te twijfelen.

d Het kan een probleem zijn dat met *bestaand*, dus niet speciaal verzameld materiaal is gewerkt. Doordat dit niet speciaal voor dit onderzoek verzameld is, kunnen er onbedoelde en onverwachte vertekeningen in het materiaal zitten, die te maken hebben met het onderzoek waarvoor het wel verzameld is.

18.4 *a* *Hypothese 1*: Er bestaat een positieve samenhang tussen het sociaal-economisch milieu van de ouders en de kans op werk die zij voor hun kinderen schatten/ verwachten.
– eenzijdige toets, want er wordt in de hypothese een richting aangegeven
– Pearson-correlatietoets of Spearman-rangcorrelatietoets
De bedoeling is het verband vast te stellen tussen twee variabelen, waarbij verondersteld zou kunnen worden dat ze beide normaal verdeeld zijn. SES is een variabele met zes waarden en geschatte kans op werk heeft slechts vijf waarden. Bij dergelijke beoordelingsschaaltjes wordt meestal gedaan alsof ze (bijna) normaal verdeeld zijn (soms ten onrechte) en op intervalniveau gemeten zijn.

b *Hypothese 2*: Mensen (ouders) die zelf geen werk hebben, schatten de kans op werk voor anderen (waaronder hun kinderen) lager dan werkenden.
– eenzijdige toets, want er wordt in de hypothese een richting aangegeven.
– t-toets voor twee onafhankelijke steekproeven (of Mann-Whitney)
Er wordt naar een vergelijking van gemiddelden gevraagd. Toetskeuze hangt af van aanname over meetniveau en al dan niet-normaal verdeeld zijn van de

afhankelijke variabele 'geschatte kans op werk'. Er is geen sprake van verwante steekproeven.

c *Hypothese 3*: Naarmate de schoolresultaten van kinderen beter zijn, wordt de kans op werk groter verondersteld.
– eenzijdige toets, want er wordt in de hypothese een richting aangegeven.
– Pearson-correlatietoets (of Spearman-rangcorrelatietoets)
De bedoeling is het verband vast te stellen tussen twee variabelen.

d *Hypothese 4*: Bestaat er een samenhang tussen de door de ouders geschatte kans op werk voor hun kinderen en de richting (binnen de lts) die door hun kinderen is gekozen?
– tweezijdig (meerzijdig) – geen richting (positief of negatief) vermeld.
– variantieanalyse met 'richting' (binnen lts) als factor.
De formulering van de hypothese suggereert een correlatie, maar lts-richting is een nominale variabele. Er kan dus alleen gekeken worden of de diverse richtingen verschillende gemiddelden hebben op de afhankelijke variabele 'geschatte kans op werk'.

e *Hypothese 5*: Ouders schatten de kans op werk voor dochters lager dan voor zonen.
– eenzijdige toets, want er wordt in de hypothese een richting aangegeven.
– t-toets voor twee (on?)afhankelijke steekproeven.
Of hier sprake is van onafhankelijke of van verwante steekproeven, hangt ervan af hoe het onderzoek opgezet wordt. Het onderzoek is zowel met verwante als met onafhankelijke steekproeven uit te voeren. In het onderzoek van Baarda c.s. worden onafhankelijke steekproeven gebruikt. Er worden geen broertjes en zusjes vergeleken.

18.5 We nemen aan dat hier sprake is van een Pearson-correlatiecoëfficiënt. Bovendien gaan we eenzijdig toetsen (zie hypothese 3 uit de vorige opgave).

Hypothesen
H_0: $\rho \leq 0{,}00$
H_1: $\rho > 0{,}00$.

Beslissingsregels
$\alpha = 0{,}01$
rechtseenzijdig
$n = 60$, $df = 60 - 2 = 58$
dus $r_{krit} = 0{,}2948$ $(n = 60)$.
Interpolatie geeft: $0{,}2948 + (0{,}3218 - 0{,}2948) \times \frac{2}{10} = 0{,}3002$
Dus $r_{krit} = .30$ $(n = 58)$.
Indien $r_{obs} < 0{,}30$, verwerp H_0 niet.
Indien $r_{obs} \geq 0{,}30$, verwerp H_0 wel.

Berekening
Gegeven is dat $r_{obs} = 0{,}40$

Beslissing
0,40 > 0,30, dus H_0 wordt verworpen.

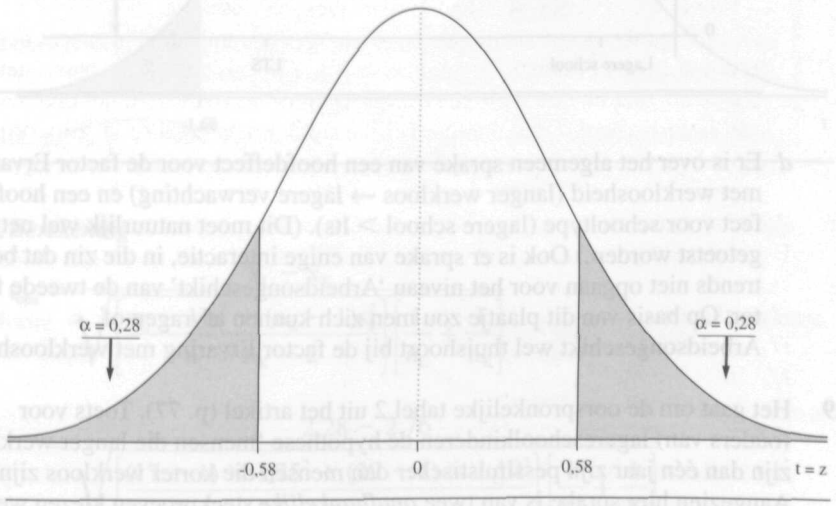
Conclusie
Uit dit onderzoek bij lagereschoolkinderen uit Den Bosch blijkt dat naarmate kinderen het beter doen op school, de ouders ook hogere verwachtingen koesteren over de kans op werk voor hun kinderen.

18.6 *a* Het aantal vrijheidsgraden is: het aantal lagere schoolkinderen − 2. In het artikel komen verschillende aantallen voor, bijvoorbeeld 583 op p. 76 (2e kolom bovenaan), 555 op p. 77 (tabel 1) en 411 op p. 78 (2e kolom bovenaan). Elk van deze getallen zou gebruikt kunnen worden. (Waarom zouden de aantallen trouwens verschillen?).
Stel we kiezen voor 583, dan df = 581. Gezien het grote aantal kinderen mogen we de kritieke t-waarde opzoeken in de z-tabel.

b De door mij gevonden p-waarde bij t = 0,58 is 0,2810.

c Mijn p-waarde wijkt af van die van de onderzoekers, omdat zij waarschijnlijk een tweezijdige toets hebben uitgevoerd, *in tegenstelling* tot hun oorspronkelijke uitgangspunt (vgl. Hypothese 5 uit de vorige opgave). Immers, $2 \times 0{,}2810 \approx 0{,}57$.

Antwoord 18.6.c **t-verdeling met zeer veel vrijheidsgraden**

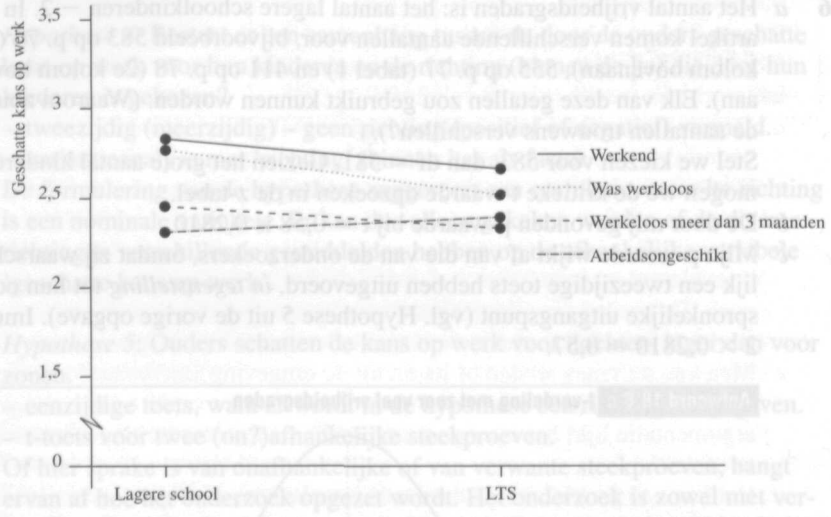


18.7 Ik zou niet tot dezelfde conclusie komen als de onderzoekers, omdat het 'duidelijke verband' betrekking heeft op een correlatie van 0,26. Dit geeft slechts 6,8% verklaarde variantie $(0{,}26)^2$, hetgeen niet bepaald hoog is. Door de grote n is de waarde van de correlatie wel statistisch significant, maar dat wil nog niet zeggen dat ongeveer 7 procent verklaarde variantie ook betekenisvol is.

18.8 *a* factor 1: schooltype (niveaus: lagere school, lts)
factor 2: ervaring met werkloosheid (niveaus: werkend, was werkloos, meer dan 3 maanden werkloos, arbeidsongeschikt).

b Beide factoren zijn *classificerend*, zij worden immers niet door de onderzoekers gemanipuleerd. Het zijn eigenschappen waarmee de kinderen (ouders) het onderzoek binnenkomen.

Antwoord 18.8.c



d Er is over het algemeen sprake van een hoofdeffect voor de factor Ervaring met werkloosheid (langer werkloos → lagere verwachting) en een hoofdeffect voor schooltype (lagere school > lts). (Dit moet natuurlijk wel netjes getoetst worden.) Ook is er sprake van enige interactie, in die zin dat beide trends niet opgaan voor het niveau 'Arbeidsongeschikt' van de tweede factor. Op basis van dit plaatje zou men zich kunnen afvragen of Arbeidsongeschikt wel thuishoort bij de factor Ervaring met werkloosheid.

18.9 Het gaat om de oorspronkelijke tabel 2 uit het artikel (p. 77). Toets voor (ouders van) lagereschoolkinderen de hypothese 'mensen die langer werkloos zijn dan één jaar zijn pessimistischer dan mensen die korter werkloos zijn'. Aangezien hier sprake is van twee *onafhankelijke* steekproeven kiezen we de voor deze situatie geschikte t-toets. Omdat in de hypothese de richting van het verschil wordt aangegeven, wordt de toetsing eenzijdig uitgevoerd.

a *Inhoudelijke nulhypothese:*
Mensen die langer werkloos zijn dan een jaar (L), schatten de kans op werk voor hun kinderen *hoger of gelijk* dan mensen die korter werkloos zijn (K).
Inhoudelijke alternatieve hypothese: Mensen die langer werkloos zijn dan een jaar, (L) schatten de kans op werk voor hun kinderen *lager* dan mensen die korter werkloos zijn (K).

b Formule hypothesen:

H_0: $\mu_L \geq \mu_K$

H_1: $\mu_L < \mu_K \Rightarrow \mu_L - \mu_K < 0$.

c Beslissingsregels:

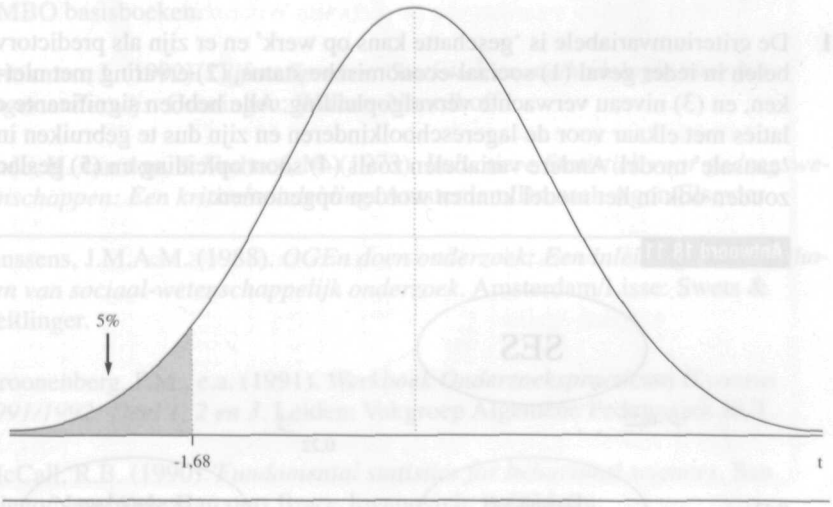
$\alpha = 0{,}05$; $df = n_L + n_K - 2 = 17 + 29 - 2 = 44$; linkseenzijdig; Tabel B;

$t_{krit} \approx 1{,}68$.

Dus: als $t_{obs} \leq -1{,}68$, verwerp H_0;

als $t_{obs} > -1{,}68$, verwerp H_0 niet.

Antwoord 18.9.c



d Berekening

$$t_{obs} = \frac{\bar{x}_L - \bar{x}_K}{\sqrt{\left[\frac{(n_L - 1)s_L^2 + (n_K - 1)s_K^2}{n_L + n_K - 2}\right]\left[\frac{1}{n_L} + \frac{1}{n_K}\right]}}$$

$$= \frac{2{,}28 - 2{,}76}{\sqrt{\left[\frac{(17 - 1) \times 1{,}25^2 + (29 - 1) \times 1{,}10^2}{17 + 29 - 2}\right]\left[\frac{1}{17} + \frac{1}{29}\right]}}$$

$$= \frac{-0{,}48}{\sqrt{\left[\frac{25{,}00 - 33{,}88}{44}\right] \times 0{,}093}} = \frac{-0{,}48}{\sqrt{1{,}338 \times 0{,}093}} = \frac{-0{,}48}{\sqrt{0{,}124}}$$

$$= \frac{-0{,}48}{0{,}35} = -1{,}37$$

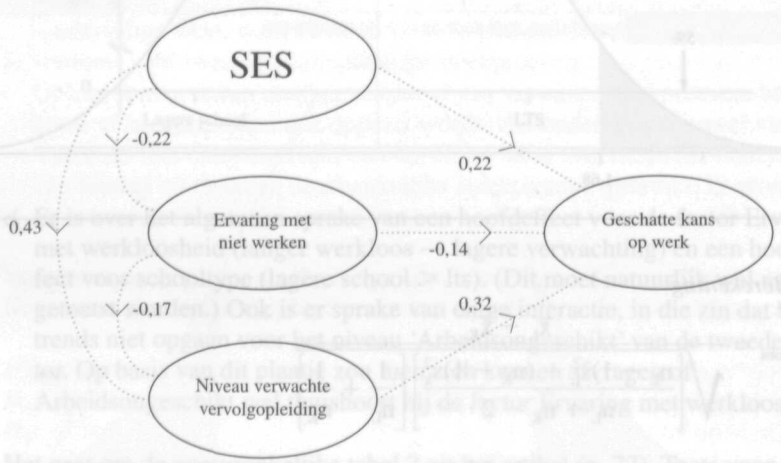
e Conclusie

$t_{obs} = -1{,}37 > t_{krit} = -1{,}68$, dus de nulhypothese wordt niet verworpen. Er is geen aanleiding te veronderstellen dat mensen die langer dan één jaar werkloos zijn, de kans op werk voor hun kinderen lager inschatten dan mensen die korter werkloos zijn.

18.10 SES en 'Ervaring niet werken' voegen (waarschijnlijk) zo weinig verklaarde variantie toe, doordat ze beide correleren met 'niveau verwachte vooropleiding' (respectievelijk 0,43 en −0,17), zodat ze gedeeltelijk dezelfde variantie verklaren. Bovendien correleren SES en 'Ervaring niet-werken' onderling −0,22 (voor de correlaties zie p. 78, 2e kolom bovenaan).

18.11 De criteriumvariabele is 'geschatte kans op werk' en er zijn als predictorvariabelen in ieder geval (1) sociaal-economische status, (2) ervaring met niet-werken, en (3) niveau verwachte vervolgopleiding. Alle hebben significante correlaties met elkaar voor de lagereschoolkinderen en zijn dus te gebruiken in het 'causale' model. Andere variabelen zoals (4) soort opleiding en (5) geslacht, zouden ook in het model kunnen worden opgenomen.

Antwoord 18.11



Het is duidelijk dat alle pijlen naar Geschatte kans op werk (de afhankelijke of criteriumvariabele) moeten lopen. Voor de relaties tussen de predictoren zijn er een aantal mogelijkheden. De hier getrokken pijlen vormen één mogelijkheid die beredeneerd kan worden (ga dit na). Een pijl van Niveau Verwachte Vervolgopleiding naar SES is evenwel niet redelijk. Hoe zou die beïnvloeding moeten werken?

Geraadpleegde literatuur

Bij het schrijven van dit boek zijn diverse bronnen geraadpleegd. De, al dan niet in bewerkte vorm, overgenomen tekstgedeelten en/of ideeën lopen uiteen van enkele regels tot omvangrijker passages. Hieronder volgt een overzicht van boeken en ander schriftelijk materiaal waaruit de schrijvers bewust hebben geput. Hoewel volledigheid wel is nagestreefd, kan deze niet worden gegarandeerd.

Beukelaar, L.J., Donselaar, G.T. van, & Tavecchio, L.W.C. (1980). *Die verdraaide werkelijkheid: Hoe komt de onderzoeker aan zijn gegevens?* Baarn: AMBO basisboeken.

Brinkman, J. (1992). *Cijfers Spreken: Statistiek en methodologie voor het hoger onderwijs*. Groningen: Wolters-Noordhoff.

Ende, H. van den, & Verhoef, M. (1973). *Inductieve Statistiek voor gedragswetenschappen: Een kritische inleiding*. Amsterdam/Brussel: Agon Elsevier.

Janssens, J.M.A.M. (1988). *OGEn doen onderzoek: Een inleiding in de methoden van sociaal-wetenschappelijk onderzoek*. Amsterdam/Lisse: Swets & Zeitlinger.

Kroonenberg, P.M., e.a. (1991). *Werkboek Onderzoekspracticum II cursus 1991/1992: Deel 1, 2 en 3*. Leiden: Vakgroep Algemene Pedagogiek RUL.

McCall, R.B. (1990). *Fundamental statistics for behavioral sciences*. San Diego/New York: Harcourt Brace Jovanovich, Publishers.

McCall, R.B. (1990), *Fundamental statistics for behavioral sciences: Study Guide*. San Diego/New York: Harcourt Brace Jovanovich, Publishers.

Moore, D.S., & McCabe, G.P. (1989). *Introduction to the practice of statistics*. New York: W.H. Freeman and Company.

Neale, J.M., & Liebert, R.M. (1986). *Science and Behavior: An Introduction to Methods of Research*. Englewood Cliffs, NJ: Prentice-Hall.

Oud, J.H.L. (1975), *Sociaal-wetenschappelijk Onderzoek: Inleiding in de methodenleer*. Nijmegen: Berkhout B.V.

Swanborn, P.G. (1987). *Methoden van sociaal-wetenschappelijk onderzoek: Nieuwe editie*. Amsterdam/Meppel: Boom.

Tavecchio, L.W.C., e.a. (1992). *Werkboek Onderzoekspracticum I cursus 1992: Deel A en B*. Leiden: Vakgroep Algemene Pedagogiek RUL.

Wiersma, W. (1975). *Research methods in education: An introduction*. Itasca, Illinois: F.E. Peacock Publishers, Inc.

APPENDIX I

Tabellen

Tabel A Oppervlakte onder de standaard-normaalverdeling 309/310

Tabel B Kritieke waarden van t 311

Tabel C Kritieke waarden van de Mann-Whitney-U toets 312/315

Tabel D Kritieke waarden van de toets van Wilcoxon 316

Tabel E Kritieke waarden van de Pearson-correlatiecoëfficiënt 317

Tabel F Kritieke waarden van de *Spearman-rangorde-correlatietoets* 318

Tabel G Kritieke waarden van de χ^2-toets en de toets van Kruskal-Wallis 319

Tabel H z-waarden behorend bij correlatiewaarden 320

Tabel I Kritieke waarden van de variantieanalyse 321/324

Tabel A
Oppervlakte onder de standaard-normaalverdeling

z	eerste gebiedskolom 0 z	tweede gebiedskolom 0 z	z	eerste gebiedskolom 0 z	tweede gebiedskolom 0 z	z	eerste gebiedskolom 0 z	tweede gebiedskolom 0 z
0.00	.0000	.5000	0.55	.2088	.2912	1.10	.3643	.1357
0.01	.0040	.4960	0.56	.2123	.2877	1.11	.3665	.1335
0.02	.0080	.4920	0.57	.2157	.2843	1.12	.3686	.1314
0.03	.0120	.4880	0.58	.2190	.2810	1.13	.3708	.1292
0.04	.0160	.4840	0.59	.2224	.2776	1.14	.3729	.1271
0.05	.0199	.4801	0.60	.2257	.2743	1.15	.3749	.1251
0.06	.0239	.4761	0.61	.2291	.2709	1.16	.3770	.1230
0.07	.0279	.4721	0.62	.2324	.2676	1.17	.3790	.1210
0.08	.0319	.4681	0.63	.2357	.2643	1.18	.3810	.1190
0.09	.0359	.4641	0.64	.2389	.2611	1.19	.3830	.1170
0.10	.0398	.4602	0.65	.2422	.2578	1.20	.3849	.1151
0.11	.0438	.4562	0.66	.2454	.2546	1.21	.3869	.1131
0.12	.0478	.4522	0.67	.2486	.2514	1.22	.3888	.1112
0.13	.0517	.4483	0.68	.2517	.2483	1.23	.3907	.1093
0.14	.0557	.4443	0.69	.2549	.2451	1.24	.3925	.1075
0.15	.0596	.4404	0.70	.2580	.2420	1.25	.3944	.1056
0.16	.0636	.4364	0.71	.2611	.2389	1.26	.3962	.1038
0.17	.0675	.4325	0.72	.2642	.2358	1.27	.3980	.1020
0.18	.0714	.4286	0.73	.2673	.2327	1.28	.3997	.1003
0.19	.0753	.4247	0.74	.2704	.2296	1.29	.4015	.0985
0.20	.0793	.4207	0.75	.2734	.2266	1.30	.4032	.0968
0.21	.0832	.4168	0.76	.2764	.2236	1.31	.4049	.0951
0.22	.0871	.4129	0.77	.2794	.2206	1.32	.4066	.0934
0.23	.0910	.4090	0.78	.2823	.2177	1.33	.4082	.0918
0.24	.0948	.4052	0.79	.2852	.2148	1.34	.4099	.0901
0.25	.0987	.4013	0.80	.2881	.2119	1.35	.4115	.0885
0.26	.1026	.3974	0.81	.2910	.2090	1.36	.4131	.0869
0.27	.1064	.3936	0.82	.2939	.2061	1.37	.4147	.0853
0.28	.1103	.3897	0.83	.2967	.2033	1.38	.4162	.0838
0.29	.1141	.3859	0.84	.2995	.2005	1.39	.4177	.0823
0.30	.1179	.3821	0.85	.3023	.1977	1.40	.4192	.0808
0.31	.1217	.3783	0.86	.3051	.1949	1.41	.4207	.0793
0.32	.1255	.3745	0.87	.3078	.1922	1.42	.4222	.0778
0.33	.1293	.3707	0.88	.3106	.1894	1.43	.4236	.0764
0.34	.1331	.3669	0.89	.3133	.1867	1.44	.4251	.0749
0.35	.1368	.3632	0.90	.3159	.1841	1.45	.4265	.0735
0.36	.1406	.3594	0.91	.3186	.1814	1.46	.4279	.0721
0.37	.1443	.3557	0.92	.3212	.1788	1.47	.4292	.0708
0.38	.1480	.3520	0.93	.3238	.1762	1.48	.4306	.0694
0.39	.1517	.3483	0.94	.3264	.1736	1.49	.4319	.0681
0.40	.1554	.3446	0.95	.3289	.1711	1.50	.4332	.0668
0.41	.1591	.3409	0.96	.3315	.1685	1.51	.4345	.0655
0.42	.1628	.3372	0.97	.3340	.1660	1.52	.4357	.0643
0.43	.1664	.3336	0.98	.3365	.1635	1.53	.4370	.0630
0.44	.1700	.3300	0.99	.3389	.1611	1.54	.4382	.0618
0.45	.1736	.3264	1.00	.3413	.1587	1.55	.4394	.0606
0.46	.1772	.3228	1.01	.3438	.1562	1.56	.4406	.0594
0.47	.1808	.3192	1.02	.3461	.1539	1.57	.4418	.0582
0.48	.1844	.3156	1.03	.3485	.1515	1.58	.4429	.0571
0.49	.1879	.3121	1.04	.3508	.1492	1.59	.4441	.0559
0.50	.1915	.3085	1.05	.3531	.1469	1.60	.4452	.0548
0.51	.1950	.3050	1.06	.3554	.1446	1.61	.4463	.0537
0.52	.1985	.3015	1.07	.3577	.1423	1.62	.4474	.0526
0.53	.2019	.2981	1.08	.3599	.1401	1.63	.4484	.0516
0.54	.2054	.2946	1.09	.3621	.1379	1.64	.4495	.0505

Tabel A (vervolg)
Oppervlakte onder de standaard-normaalverdeling

z	eerste gebiedskolom	tweede gebiedskolom	z	eerste gebiedskolom	tweede gebiedskolom	z	eerste gebiedskolom	tweede gebiedskolom
	0 z	0 z		0 z	0 z		0 z	0 z
1.65	.4505	.0495	2.22	.4868	.0132	2.79	.4974	.0026
1.66	.4515	.0485	2.23	.4871	.0129	2.80	.4974	.0026
1.67	.4525	.0475	2.24	.4875	.0125	2.81	.4975	.0025
1.68	.4535	.0465	2.25	.4878	.0122	2.82	.4976	.0024
1.69	.4545	.0455	2.26	.4881	.0119	2.83	.4977	.0023
1.70	.4554	.0446	2.27	.4884	.0116	2.84	.4977	.0023
1.71	.4564	.0436	2.28	.4887	.0113	2.85	.4978	.0022
1.72	.4573	.0427	2.29	.4890	.0110	2.86	.4979	.0021
1.73	.4582	.0418	2.30	.4893	.0107	2.87	.4979	.0021
1.74	.4591	.0409	2.31	.4896	.0104	2.88	.4980	.0020
1.75	.4599	.0401	2.32	.4898	.0102	2.89	.4981	.0019
1.76	.4608	.0392	2.33	.4901	.0099	2.90	.4981	.0019
1.77	.4616	.0384	2.34	.4904	.0096	2.91	.4982	.0018
1.78	.4625	.0375	2.35	.4906	.0094	2.92	.4982	.0018
1.79	.4633	.0367	2.36	.4909	.0091	2.93	.4983	.0017
1.80	.4641	.0359	2.37	.4911	.0089	2.94	.4984	.0016
1.81	.4649	.0351	2.38	.4913	.0087	2.95	.4984	.0016
1.82	.4656	.0344	2.39	.4916	.0084	2.96	.4985	.0015
1.83	.4664	.0336	2.40	.4918	.0082	2.97	.4985	.0015
1.84	.4671	.0329	2.41	.4920	.0080	2.98	.4986	.0014
1.85	.4678	.0322	2.42	.4922	.0078	2.99	.4986	.0014
1.86	.4686	.0314	2.43	.4925	.0075	3.00	.4987	.0013
1.87	.4693	.0307	2.44	.4927	.0073	3.01	.4987	.0013
1.88	.4699	.0301	2.45	.4929	.0071	3.02	.4987	.0013
1.89	.4706	.0294	2.46	.4931	.0069	3.03	.4988	.0012
1.90	.4713	.0287	2.47	.4932	.0068	3.04	.4988	.0012
1.91	.4719	.0281	2.48	.4934	.0066	3.05	.4989	.0011
1.92	.4726	.0274	2.49	.4936	.0064	3.06	.4989	.0011
1.93	.4732	.0268	2.50	.4938	.0062	3.07	.4989	.0011
1.94	.4738	.0262	2.51	.4940	.0060	3.08	.4990	.0010
1.95	.4744	.0256	2.52	.4941	.0059	3.09	.4990	.0010
1.96	.4750	.0250	2.53	.4943	.0057	3.10	.4990	.0010
1.97	.4756	.0244	2.54	.4945	.0055	3.11	.4991	.0009
1.98	.4761	.0239	2.55	.4946	.0054	3.12	.4991	.0009
1.99	.4767	.0233	2.56	.4948	.0052	3.13	.4991	.0009
2.00	.4772	.0228	2.57	.4949	.0051	3.14	.4992	.0008
2.01	.4778	.0222	2.58	.4951	.0049	3.15	.4992	.0008
2.02	.4783	.0217	2.59	.4952	.0048	3.16	.4992	.0008
2.03	.4788	.0212	2.60	.4953	.0047	3.17	.4992	.0008
2.04	.4793	.0207	2.61	.4955	.0045	3.18	.4993	.0007
2.05	.4798	.0202	2.62	.4956	.0044	3.19	.4993	.0007
2.06	.4803	.0197	2.63	.4957	.0043	3.20	.4993	.0007
2.07	.4808	.0192	2.64	.4959	.0041	3.21	.4993	.0007
2.08	.4812	.0188	2.65	.4960	.0040	3.22	.4994	.0006
2.09	.4817	.0183	2.66	.4961	.0039	3.23	.4994	.0006
2.10	.4821	.0179	2.67	.4962	.0038	3.24	.4994	.0006
2.11	.4826	.0174	2.68	.4963	.0037	3.25	.4994	.0006
2.12	.4830	.0170	2.69	.4964	.0036	3.30	.4995	.0005
2.13	.4834	.0166	2.70	.4965	.0035	3.35	.4996	.0004
2.14	.4838	.0162	2.71	.4966	.0034	3.40	.4997	.0003
2.15	.4842	.0158	2.72	.4967	.0033	3.45	.4997	.0003
2.16	.4846	.0154	2.73	.4968	.0032	3.50	.4998	.0002
2.17	.4850	.0150	2.74	.4969	.0031	3.60	.4998	.0002
2.18	.4854	.0146	2.75	.4970	.0030	3.70	.4999	.0001
2.19	.4857	.0143	2.76	.4971	.0029	3.80	.4999	.0001
2.20	.4861	.0139	2.77	.4972	.0028	3.90	.49995	.00005
2.21	.4864	.0136	2.78	.4973	.0027	4.00	.49997	.00003

Tabel B

Kritieke waarden van t

	significantieniveau bij een eenzijdige toets					
	.10	.05	.025	.01	.005	.0005
	significantieniveau bij een tweezijdige toets					
df	.20	.10	.05	.02	.01	.001
1	3.078	6.314	12.706	31.821	63.657	636.619
2	1.886	2.920	4.303	6.965	9.925	31.598
3	1.638	2.353	3.182	4.541	5.841	12.941
4	1.533	2.132	2.776	3.747	4.604	8.610
5	1.476	2.015	2.571	3.365	4.032	6.859
6	1.440	1.943	2.447	3.143	3.707	5.959
7	1.415	1.895	2.365	2.998	3.499	5.405
8	1.397	1.860	2.306	2.896	3.355	5.041
9	1.383	1.833	2.262	2.821	3.250	4.781
10	1.372	1.812	2.228	2.764	3.169	4.587
11	1.363	1.796	2.201	2.718	3.106	4.437
12	1.356	1.782	2.179	2.681	3.055	4.318
13	1.350	1.771	2.160	2.650	3.012	4.221
14	1.345	1.761	2.145	2.624	2.977	4.140
15	1.341	1.753	2.131	2.602	2.947	4.073
16	1.337	1.746	2.120	2.583	2.921	4.015
17	1.333	1.740	2.110	2.567	2.898	3.965
18	1.330	1.734	2.101	2.552	2.878	3.922
19	1.328	1.729	2.093	2.539	2.861	3.883
20	1.325	1.725	2.086	2.528	2.845	3.850
21	1.323	1.721	2.080	2.518	2.831	3.819
22	1.321	1.717	2.074	2.508	2.819	3.792
23	1.319	1.714	2.069	2.500	2.807	3.767
24	1.318	1.711	2.064	2.492	2.797	3.745
25	1.316	1.708	2.060	2.485	2.787	3.725
26	1.315	1.706	2.056	2.479	2.779	3.707
27	1.314	1.703	2.052	2.473	2.771	3.690
28	1.313	1.701	2.048	2.467	2.763	3.674
29	1.311	1.699	2.045	2.462	2.756	3.659
30	1.310	1.697	2.042	2.457	2.750	3.646
40	1.303	1.684	2.021	2.423	2.704	3.551
60	1.296	1.671	2.000	2.390	2.660	3.460
120	1.289	1.658	1.980	2.358	2.617	3.373
∞	1.282	1.645	1.960	2.326	2.576	3.291

Tabel C

Kritieke waarden van de Mann-Whitney-U toets

eenzijdig bij $\alpha = 0{,}025$ en tweezijdig bij $\alpha = 0{,}05$

n_B \ n_A	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
2	--	--	--	--	--	--	--	0	0	0	0	1	1	1	1	1	2	2	2	2
								16	18	20	22	23	25	27	29	31	32	34	36	38
3	--	--	--	--	0	1	1	2	2	3	3	4	4	5	5	6	6	7	7	8
					15	17	20	22	25	27	30	32	35	37	40	42	45	47	50	52
4	--	--	--	0	1	2	3	4	4	5	6	7	8	9	10	11	11	12	13	13
				16	19	22	25	28	32	35	38	41	44	47	50	53	57	60	63	67
5	--	--	0	1	2	3	5	6	7	8	9	11	12	13	14	15	17	18	19	20
			15	19	23	27	30	34	38	42	46	49	53	57	61	65	68	72	76	80
6	--	--	1	2	3	5	6	8	10	11	13	14	16	17	19	21	22	24	25	27
			17	22	27	31	36	40	44	49	53	58	62	67	71	75	80	84	89	93
7	--	--	1	3	5	6	8	10	12	14	16	18	20	22	24	26	28	30	32	34
			20	25	30	36	41	46	51	56	61	66	71	76	81	86	91	96	101	106
8	--	0	2	4	6	8	10	13	15	17	19	22	24	26	29	31	34	36	38	41
		16	22	28	34	40	46	51	57	63	69	74	80	86	91	97	102	108	111	119
9	--	0	2	4	7	10	12	15	17	20	23	26	28	31	34	37	39	42	45	48
		18	25	32	38	44	51	57	64	70	76	82	89	95	101	107	114	120	126	132
10	--	0	3	5	8	11	14	17	20	23	26	29	33	36	39	42	45	48	52	55
		20	27	35	42	49	56	63	70	77	84	91	97	104	111	118	125	132	138	145
11	--	0	3	6	9	13	16	19	23	26	30	33	37	40	44	47	51	55	58	62
		22	30	38	46	53	61	69	76	84	91	99	106	114	121	129	136	143	151	158
12	--	1	4	7	11	14	18	22	26	29	33	37	41	45	49	53	57	61	65	69
		23	32	41	49	58	66	74	82	91	99	107	115	123	131	139	147	155	163	171
13	--	1	4	8	12	16	20	24	28	33	37	41	45	50	54	59	63	67	72	76
		25	35	44	53	62	71	80	89	97	106	115	124	132	141	149	158	167	175	184
14	--	1	5	9	13	17	22	26	31	36	40	45	50	55	59	64	67	74	78	83
		27	37	47	51	67	76	86	95	104	114	123	132	141	151	160	171	178	188	197
15	--	1	5	10	14	19	24	29	34	39	44	49	54	59	64	70	75	80	85	90
		29	40	50	61	71	81	91	101	111	121	131	141	151	161	170	180	190	200	210
16	--	1	6	11	15	21	26	31	37	42	47	53	59	64	70	75	81	86	92	98
		31	42	53	65	75	86	97	107	118	129	139	149	160	170	181	191	202	212	222
17	--	2	6	11	17	22	28	34	39	45	51	57	63	67	75	81	87	93	99	105
		32	45	57	68	80	91	102	114	125	136	147	158	171	180	191	202	213	224	235
18	--	2	7	12	18	24	30	36	42	48	55	61	67	74	80	86	93	99	106	112
		34	47	60	72	84	96	108	120	132	143	155	167	178	190	202	213	225	236	248
19	--	2	7	13	19	25	32	38	45	52	58	65	72	78	85	92	99	106	113	119
		36	50	63	76	89	101	114	126	138	151	163	175	188	200	212	224	236	248	261
20	--	2	8	13	20	27	34	41	48	55	62	69	76	83	90	98	105	112	119	127
		38	52	67	80	93	106	119	132	145	158	171	184	197	210	222	235	248	261	273

Tabel C

Kritieke waarden van de Mann-Whitney-U toets

eenzijdig bij $\alpha = 0{,}005$ en tweezijdig bij $\alpha = 0{,}01$

n_B \ n_A	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
2	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	0 38	0 40
3	--	--	--	--	--	--	--	--	0 27	0 30	0 33	1 35	1 38	1 41	2 43	2 46	2 49	2 52	3 54	3 57
4	--	--	--	--	--	0 24	0 28	1 31	1 35	2 38	2 42	3 45	3 49	4 52	5 55	5 59	6 62	6 66	7 69	8 72
5	--	--	--	--	0 25	1 29	1 34	2 38	3 42	4 46	5 50	6 54	7 58	7 63	8 67	9 71	10 75	11 79	12 83	13 87
6	--	--	--	0 24	1 29	2 34	3 39	4 44	5 49	6 54	7 59	9 63	10 68	11 73	12 78	13 83	15 87	16 92	17 97	18 102
7	--	--	--	0 28	1 34	3 39	4 45	6 50	7 56	9 61	10 67	12 72	13 78	15 83	16 89	18 94	19 100	21 105	22 111	24 116
8	--	--	--	1 31	2 38	4 44	6 50	7 57	9 63	11 69	13 75	15 81	17 87	18 94	20 100	22 106	24 112	26 118	28 124	30 130
9	--	--	0 27	1 35	3 42	5 49	7 56	9 63	11 70	13 77	16 83	18 90	20 97	22 104	24 111	27 117	29 124	31 131	33 138	36 144
10	--	--	0 30	2 38	4 46	6 54	9 61	11 69	13 77	16 84	18 92	21 99	24 106	26 114	29 121	31 129	34 136	37 143	39 151	42 158
11	--	--	0 33	2 42	5 50	7 59	10 67	13 75	16 83	18 92	21 100	24 108	27 116	30 124	33 132	36 140	39 148	42 156	45 164	48 172
12	--	--	1 35	3 45	6 54	9 63	12 72	15 81	18 90	21 99	24 108	27 117	31 125	34 134	37 143	41 151	44 160	47 169	51 177	54 186
13	--	--	1 38	3 49	7 58	10 68	13 78	17 87	20 97	24 106	27 116	31 125	34 125	38 144	42 153	45 163	49 172	53 181	56 191	60 200
14	--	--	1 41	4 52	7 63	11 73	15 83	18 94	22 104	26 114	30 124	34 134	38 144	42 154	46 164	50 174	54 184	58 194	63 203	67 213
15	--	--	2 43	5 55	8 67	12 78	16 89	20 100	24 111	29 121	33 132	37 143	42 153	46 164	51 174	55 185	60 195	64 206	69 216	73 227
16	--	--	2 46	5 59	9 71	13 83	18 94	22 106	27 117	31 129	36 140	41 151	45 163	50 174	55 185	60 196	65 207	70 218	74 230	79 241
17	--	--	2 49	6 62	10 75	15 87	19 100	24 112	29 124	34 148	39 148	44 160	49 172	54 184	60 195	65 207	70 219	75 231	81 242	86 254
18	--	--	2 52	6 66	11 79	16 92	21 105	26 118	31 131	37 143	42 156	47 169	53 181	58 194	64 206	70 218	75 231	81 243	87 255	92 268
19	--	0 38	3 54	7 69	12 83	17 97	22 111	28 124	33 138	39 151	45 164	51 177	56 191	63 203	69 216	74 230	81 242	87 255	93 268	99 281
20	--	0 40	3 57	8 72	13 87	18 102	24 116	30 130	36 144	42 158	48 172	54 186	60 200	67 213	73 227	79 241	86 254	92 268	99 281	105 295

Tabel C

Kritieke waarden van de Mann-Whitney-U toets

eenzijdig bij $\alpha = 0{,}05$ en tweezijdig bij $\alpha = 0{,}10$

n_B \ n_A	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	0	0
																			19	20
2	--	--	--	--	0	0	0	1	1	1	1	2	2	2	3	3	3	4	4	4
					10	12	14	15	17	19	21	22	24	26	27	29	31	32	34	36
3	--	--	0	0	1	2	2	3	3	4	5	5	6	7	7	8	9	9	10	11
			9	12	14	16	19	21	24	26	28	31	33	35	38	40	42	45	47	49
4	--	--	0	1	2	3	4	5	6	7	8	9	10	11	12	14	15	16	17	18
			12	15	18	21	24	27	30	33	36	39	42	45	48	50	53	56	59	62
5	--	0	1	2	4	5	6	8	9	11	12	13	15	16	18	19	20	22	23	25
		10	14	18	21	25	29	32	36	39	43	47	50	54	57	61	65	68	72	75
6	--	0	2	3	5	7	8	10	12	14	16	17	19	21	23	25	26	28	30	32
		12	16	21	25	29	34	38	42	46	50	55	59	63	67	71	76	80	84	88
7	--	0	2	4	6	8	11	13	15	17	19	21	24	26	28	30	33	35	37	39
		14	19	24	29	34	38	43	48	53	58	63	67	72	77	82	86	91	96	101
8	--	1	3	5	8	10	13	15	18	20	23	26	28	31	33	36	39	41	44	47
		15	21	27	32	38	43	49	54	60	65	70	76	81	87	92	97	103	108	113
9	--	1	3	6	9	12	15	18	21	24	27	30	33	36	39	42	45	48	51	54
		17	24	30	36	42	48	54	60	66	72	78	84	90	96	102	108	114	120	126
10	--	1	4	7	11	14	17	20	24	27	31	34	37	41	44	48	51	55	58	62
		19	26	33	39	46	53	60	66	73	79	86	93	99	106	112	119	125	132	138
11	--	1	5	8	12	16	19	23	27	31	34	38	42	46	50	54	57	61	65	69
		21	28	36	43	50	58	65	72	79	87	94	101	108	115	122	130	137	144	151
12	--	2	5	9	13	17	21	26	30	34	38	42	47	51	55	60	64	68	72	77
		22	31	39	47	55	63	70	78	86	94	102	109	117	125	132	140	148	156	163
13	--	2	6	10	15	19	24	28	33	37	42	47	51	56	61	65	70	75	80	84
		24	33	42	50	59	67	76	84	93	101	109	118	126	134	143	151	159	167	176
14	--	2	7	11	16	21	26	31	36	41	46	51	56	61	66	71	77	82	87	92
		26	35	45	54	63	72	81	90	99	108	117	126	135	144	153	161	170	179	188
15	--	3	7	12	18	23	28	33	39	44	50	55	61	66	72	77	83	88	94	100
		27	38	48	57	67	77	87	96	106	115	125	134	144	153	163	172	182	191	200
16	--	3	8	14	19	25	30	36	42	48	54	60	65	71	77	83	89	95	101	107
		29	40	50	61	71	82	92	102	112	122	132	143	153	163	173	183	193	203	213
17	--	3	9	15	20	26	33	39	45	51	57	64	70	77	83	89	96	102	109	115
		31	42	53	65	76	86	97	108	119	130	140	151	161	172	183	193	204	214	225
18	--	4	9	16	22	28	35	41	48	55	61	68	75	82	88	95	102	109	116	123
		32	45	56	68	80	91	103	114	123	137	148	159	170	182	193	204	215	226	237
19	0	4	10	17	23	30	37	44	51	58	65	72	80	87	94	101	109	116	123	130
	19	34	47	59	72	84	96	108	120	132	144	156	167	179	191	203	214	226	238	250
20	0	4	11	18	25	32	39	47	54	62	69	77	84	92	100	107	115	123	130	138
	20	36	49	62	75	88	101	113	126	138	151	163	176	188	200	213	225	237	250	262

Tabel C

Kritieke waarden van de Mann-Whitney-U toets

eenzijdig bij $\alpha = 0{,}01$ en tweezijdig bij $\alpha = 0{,}02$

n_B \ n_A	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
2	--	--	--	--	--	--	--	--	--	--	--	--	0 26	0 28	0 30	0 32	0 34	0 36	1 37	1 39
3	--	--	--	--	--	--	0 21	0 24	1 26	1 29	1 32	2 34	2 37	2 40	3 42	3 45	4 47	4 50	4 52	5 55
4	--	--	--	--	0 20	1 23	1 27	2 30	3 33	3 37	4 40	5 43	5 47	6 50	7 53	7 57	8 60	9 63	9 67	10 70
5	--	--	--	0 20	1 24	2 28	3 32	4 36	5 40	6 44	7 48	8 52	9 56	10 60	11 64	12 68	13 72	14 76	15 80	16 84
6	--	--	--	1 23	2 28	3 33	4 38	6 42	7 47	8 52	9 57	11 61	12 66	13 71	15 75	16 80	18 84	19 89	20 94	22 98
7	--	--	0 21	1 27	3 32	4 38	6 43	7 49	9 54	11 59	12 65	14 70	16 75	17 81	19 86	21 91	23 96	24 102	26 107	28 112
8	--	--	0 24	2 30	4 36	6 42	7 49	9 55	11 61	13 67	15 73	17 79	20 84	22 90	24 96	26 102	28 108	30 114	32 120	34 126
9	--	--	1 26	3 33	5 40	7 47	9 54	11 61	14 67	16 74	18 81	21 87	23 94	26 100	28 107	31 113	33 120	36 126	38 133	40 140
10	--	--	1 29	3 37	6 44	8 52	11 59	13 67	16 74	19 81	22 88	24 96	27 103	30 110	33 117	36 124	38 132	41 139	44 146	47 153
11	--	--	1 32	4 40	7 48	9 57	12 65	15 73	18 81	22 88	25 96	28 104	31 112	34 120	37 128	41 135	44 143	47 151	50 159	53 167
12	--	--	2 34	5 43	8 52	11 61	14 70	17 79	21 87	24 96	28 104	31 113	35 121	38 130	42 138	46 146	49 155	53 163	56 172	60 180
13	--	0 26	2 37	5 47	9 56	12 66	16 75	20 84	23 94	27 103	31 112	35 121	39 130	43 139	47 148	51 157	55 166	59 175	63 184	67 193
14	--	0 28	2 40	6 50	10 60	13 71	17 81	22 90	26 100	30 110	34 120	38 130	43 139	47 149	51 159	56 168	60 178	65 187	69 197	73 207
15	--	0 30	3 42	7 53	11 64	15 75	19 86	24 96	28 107	33 117	37 128	42 138	47 148	51 159	56 169	61 179	66 189	70 200	75 210	80 220
16	--	0 32	3 45	7 57	12 68	16 80	21 91	26 102	31 113	36 124	41 135	46 146	51 157	56 168	61 179	66 190	71 201	76 212	82 222	87 233
17	--	0 34	4 47	8 60	13 72	18 84	23 96	28 108	33 120	38 132	44 143	49 155	55 166	60 178	66 189	71 201	77 212	82 224	88 234	93 247
18	--	0 36	4 50	9 63	14 76	19 89	24 102	30 114	36 126	41 139	47 151	53 163	59 175	65 187	70 200	76 212	82 224	88 236	94 248	100 260
19	--	1 37	4 53	9 67	15 80	20 94	26 107	32 120	38 133	44 146	50 159	56 172	63 184	69 197	75 210	82 222	88 235	94 248	101 260	107 273
20	--	1 39	5 55	10 70	16 84	22 98	28 112	34 126	40 140	47 153	53 167	60 180	67 193	73 207	80 220	87 233	93 247	100 260	107 273	114 286

Tabel D

Kritieke waarden van de toets van Wilcoxon

	significantieniveau bij een eenzijdige toets					significantieniveau bij een eenzijdige toets			
	.05	.025	.01	.005		.05	.025	.01	.005
	significantieniveau bij een tweezijdige toets					significantieniveau bij een tweezijdige toets			
N	.10	.05	.02	.01	*N*	.10	.05	.02	.01
5	0	--	--	--	28	130	116	101	91
6	2	0	--	--	29	140	126	110	100
7	3	2	0	--	30	151	137	120	109
8	5	3	1	0	31	163	147	130	118
9	8	5	3	1	32	175	159	140	128
10	10	8	5	3	33	187	170	151	138
11	13	10	7	5	34	200	182	162	148
12	17	13	9	7	35	213	195	173	159
13	21	17	12	9	36	227	208	185	171
14	25	21	15	12	37	241	221	198	182
15	30	25	19	15	38	256	235	211	194
16	35	29	23	19	39	271	249	224	207
17	41	34	27	23	40	286	264	238	220
18	47	40	32	27	41	302	279	252	233
19	53	46	37	32	42	319	294	266	247
20	60	52	43	37	43	336	310	281	261
21	67	58	49	42	44	353	327	296	276
22	75	65	55	48	45	371	343	312	291
23	83	73	62	54	46	389	361	328	307
24	91	81	69	61	47	407	378	345	322
25	100	89	76	68	48	426	396	362	339
26	110	98	84	75	49	446	415	379	355
27	119	107	92	83	50	466	434	397	373

Tabel E

Kritieke waarden van de Pearson-correlatiecoëfficiënt

	significantieniveau bij een eenzijdige toets				
	.05	.025	.01	.005	.0005
	significantieniveau bij een tweezijdige toets				
$df = N-2$	.10	.05	.02	.01	.001
1	.9877	.9969	.9995	.9999	1.0000
2	.9000	.9500	.9800	.9900	.9990
3	.8054	.8783	.9343	.9587	.9912
4	.7293	.8114	.8822	.9172	.9741
5	.6694	.7545	.8329	.8745	.9507
6	.6215	.7067	.7887	.8343	.9249
7	.5822	.6664	.7498	.7977	.8982
8	.5494	.6319	.7155	.7646	.8721
9	.5214	.6021	.6851	.7348	.8471
10	.4973	.5760	.6581	.7079	.8233
11	.4762	.5529	.6339	.6835	.8010
12	.4575	.5324	.6120	.6614	.7800
13	.4409	.5139	.5923	.6411	.7603
14	.4259	.4973	.5742	.6226	.7420
15	.4124	.4821	.5577	.6055	.7246
16	.4000	.4683	.5425	.5897	.7084
17	.3887	.4555	.5285	.5751	.6932
18	.3783	.4438	.5155	.5614	.6787
19	.3687	.4329	.5034	.5487	.6652
20	.3598	.4227	.4921	.5368	.6524
25	.3233	.3809	.4451	.4869	.5974
30	.2960	.3494	.4093	.4487	.5541
35	.2746	.3246	.3810	.4182	.5189
40	.2573	.3044	.3578	.3932	.4896
45	.2428	.2875	.3384	.3721	.4648
50	.2306	.2732	.3218	.3541	.4433
60	.2108	.2500	.2948	.3248	.4078
70	.1954	.2319	.2737	.3017	.3799
80	.1829	.2172	.2565	.2830	.3568
90	.1726	.2050	.2422	.2673	.3375
100	.1638	.1946	.2301	.2540	.3211

Tabel F

Kritieke waarden van de Spearman-rangorde-correlatietoets

	significantieniveau bij een eenzijdige toets			
	.05	.025	.005	.001
	significantieniveau bij een tweezijdige toets			
N	.10	.05	.01	.002
5	.900	1.000		
6	.829	.886	1.000	
7	.715	.786	.929	1.000
8	.620	.715	.881	.953
9	.600	.700	.834	.917
10	.564	.649	.794	.879
11	.537	.619	.764	.855
12	.504	.588	.735	.826
13	.484	.561	.704	.797
14	.464	.539	.680	.772
15	.447	.522	.658	.750
16	.430	.503	.636	.730
17	.415	.488	.618	.711
18	.402	.474	.600	.693
19	.392	.460	.585	.676
20	.381	.447	.570	.661
21	.371	.437	.556	.647
22	.361	.426	.544	.633
23	.353	.417	.532	.620
24	.345	.407	.521	.608
25	.337	.399	.511	.597
26	.331	.391	.501	.587
27	.325	.383	.493	.577
28	.319	.376	.484	.567
29	.312	.369	.475	.558
30	.307	.363	.467	.549

Tabel G

Kritieke waarden van de X^2-toets en de toets van Kruskal-Wallis

	significantieniveau bij een eenzijdige toets					
	.10	.05	.025	.01	.005	.0005
	significantieniveau bij een tweezijdige toets					
df	.20	.10	.05	.02	.01	.001
1	1.64	2.71	3.84	5.41	6.64	10.83
2	3.22	4.60	5.99	7.82	9.21	13.82
3	4.64	6.25	7.82	9.84	11.34	16.27
4	5.99	7.78	9.49	11.67	13.28	18.46
5	7.29	9.24	11.07	13.39	15.09	20.52
6	8.56	10.64	12.59	15.03	16.81	22.46
7	9.80	12.02	14.07	16.62	18.48	24.32
8	11.03	13.36	15.51	18.17	20.09	26.12
9	12.24	14.68	16.92	19.68	21.67	27.88
10	13.44	15.99	18.31	21.16	23.21	29.59
11	14.63	17.28	19.68	22.62	24.72	31.26
12	15.81	18.55	21.03	24.05	26.22	32.91
13	16.98	19.81	22.36	25.47	27.69	34.53
14	18.15	21.06	23.68	26.87	29.14	36.12
15	19.31	22.31	25.00	28.26	30.58	37.70
16	20.46	23.54	26.30	29.63	32.00	39.29
17	21.62	24.77	27.59	31.00	33.41	40.75
18	22.76	25.99	28.87	32.35	34.80	42.31
19	23.90	27.20	30.14	33.69	36.19	43.82
20	25.04	28.41	31.41	35.02	37.57	45.32
21	26.17	29.62	32.67	36.34	38.93	46.80
22	27.30	30.81	33.92	37.66	40.29	48.27
23	28.43	32.01	35.17	38.97	41.64	49.73
24	29.55	33.20	36.42	40.27	42.98	51.18
25	30.68	34.38	37.65	41.57	44.31	52.62
26	31.80	35.56	38.88	42.86	45.64	54.05
27	32.91	36.74	40.11	44.14	46.96	55.48
28	34.03	37.92	41.34	45.42	48.28	56.89
29	35.14	39.09	42.69	46.69	49.59	58.30
30	36.25	40.26	43.77	47.96	50.89	59.70
32	38.47	42.59	46.19	50.49	53.49	62.49
34	40.68	44.90	48.60	53.00	56.06	65.25
36	42.88	47.21	51.00	55.49	58.62	67.99
38	45.08	49.51	53.38	57.97	61.16	70.70
40	47.27	51.81	55.76	60.44	63.69	73.40
44	51.64	56.37	60.48	65.34	68.71	78.75
48	55.99	60.91	65.17	70.20	73.68	84.04
52	60.33	65.42	69.83	75.02	78.62	89.27
56	64.66	69.92	74.47	79.82	83.51	94.46
60	68.97	74.40	79.08	84.58	88.38	99.61

Tabel H

z-waarden behorend bij correlatiewaarden

r	z_r	r	z_r	r	z_r	r	z_r	r	z_r
.000	.000	.200	.203	.400	.424	.600	.693	.800	1.099
.005	.005	.205	.208	.405	.430	.605	.701	.805	1.113
.010	.010	.210	.213	.410	.436	.610	.709	.810	1.127
.015	.015	.215	.218	.415	.442	.615	.717	.815	1.142
.020	.020	.220	.224	.420	.448	.620	.725	.820	1.157
.025	.025	.225	.229	.425	.454	.625	.733	.825	1.172
.030	.030	.230	.234	.430	.460	.630	.741	.830	1.188
.035	.035	.235	.239	.435	.466	.635	.750	.835	1.204
.040	.040	.240	.245	.440	.472	.640	.758	.840	1.221
.045	.045	.245	.250	.445	.478	.645	.767	.845	1.238
.050	.050	.250	.255	.450	.485	.650	.775	.850	1.256
.055	.055	.255	.261	.455	.491	.655	.784	.855	1.274
.060	.060	.260	.266	.460	.497	.660	.793	.860	1.293
.065	.065	.265	.271	.465	.504	.665	.802	.865	1.313
.070	.070	.270	.277	.470	.510	.670	.811	.870	1.333
.075	.075	.275	.282	.475	.517	.675	.820	.875	1.354
.080	.080	.280	.288	.460	.523	.680	.829	.880	1.376
.085	.085	.285	.293	.485	.530	.685	.838	.885	1.398
.090	.090	.290	.299	.490	.536	.690	.848	.890	1.422
.095	.095	.295	.304	.495	.543	.695	.858	.895	1.447
.100	.100	.300	.310	.500	.549	.700	.867	.900	1.472
.105	.105	.305	.315	.505	.556	.705	.877	.905	1.499
.110	.110	.310	.321	.510	.563	.710	.887	.910	1.528
.115	.116	.315	.326	.515	.570	.715	.897	.915	1.557
.120	.121	.320	.332	.520	.576	.720	.908	.920	1.589
.125	.126	.325	.337	.525	.583	.725	.918	.925	1.623
.130	.131	.330	.343	.530	.590	.730	.929	.930	1.658
.135	.136	.335	.348	.535	.597	.735	.940	.935	1.697
.140	.141	.340	.354	.540	.604	.740	.950	.940	1.738
.145	.146	.345	.360	.545	.611	.745	.962	.945	1.783
.150	.151	.350	.365	.550	.618	.750	.973	.950	1.832
.155	.156	.355	.371	.555	.626	.755	.984	.955	1.886
.160	.161	.360	.377	.560	.633	.760	.996	.960	1.946
.165	.167	.365	.383	.565	.640	.765	1.008	.965	2.014
.170	.172	.370	.388	.570	.648	.770	1.020	.970	2.092
.175	.177	.375	.394	.575	.655	.775	1.033	.975	2.185
.180	.182	.380	.400	.580	.662	.780	1.045	.980	2.298
.185	.187	.385	.406	.585	.670	.785	1.058	.985	2.443
.190	.192	.390	.412	.590	.678	.790	1.071	.990	2.647
.195	.198	.395	.418	.595	.685	.795	1.085	.995	2.994

Tabel I

Kritieke waarden (F) van de variantieanalyse

waarden bij α = 0,05 staan boven
dikgedrukte waarden (onder) bij α = 0,01

df van de noemer

df van de teller	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	∞
1	161 4,052	200 4,999	216 5,403	225 5,625	230 5,764	234 5,859	237 5,928	239 5,981	241 6,022	242 6,056	243 6,082	244 6,106	245 6,142	246 6,169	248 6,208	249 6,234	250 6,261	251 6,286	252 6,302	253 6,323	253 6,334	254 6,352	254 6,361	254 6,366
2	18.51 98.49	19.00 99.00	19.16 99.17	19.25 99.25	19.30 99.30	19.33 99.33	19.36 99.36	19.37 99.37	19.38 99.39	19.39 99.40	19.40 99.41	19.41 99.42	19.42 99.43	19.43 99.44	19.44 99.45	19.45 99.46	19.46 99.47	19.47 99.48	19.47 99.48	19.48 99.49	19.49 99.49	19.49 99.49	19.50 99.50	19.50 99.50
3	10.13 34.12	9.55 30.82	9.28 29.46	9.12 28.71	9.01 28.24	8.94 27.91	8.88 27.67	8.84 27.49	8.81 27.34	8.78 27.23	8.76 27.13	8.74 27.05	8.71 26.92	8.69 26.83	8.66 26.69	8.64 26.60	8.62 26.50	8.60 26.41	8.58 26.35	8.57 26.27	8.56 26.23	8.54 26.18	8.54 26.14	8.53 26.12
4	7.71 21.20	6.94 18.00	6.59 16.69	6.39 15.98	6.26 15.52	6.16 15.21	6.09 14.98	6.04 14.80	6.00 14.66	5.96 14.54	5.93 14.45	5.91 14.37	5.87 14.24	5.84 14.15	5.80 14.02	5.77 13.93	5.74 13.83	5.71 13.74	5.70 13.69	5.68 13.61	5.66 13.57	5.65 13.52	5.64 13.48	5.63 13.46
5	6.61 16.26	5.79 13.27	5.41 12.06	5.19 11.39	5.05 10.97	4.95 10.67	4.88 10.45	4.82 10.29	4.78 10.15	4.74 10.05	4.70 9.96	4.68 9.89	4.64 9.77	4.60 9.68	4.56 9.55	4.53 9.47	4.50 9.38	4.46 9.29	4.44 9.24	4.42 9.17	4.40 9.13	4.38 9.07	4.37 9.04	4.36 9.02
6	5.99 13.74	5.14 10.92	4.76 9.78	4.53 9.15	4.39 8.75	4.28 8.47	4.21 8.26	4.15 8.10	4.10 7.98	4.06 7.87	4.03 7.79	4.00 7.72	3.96 7.60	3.92 7.52	3.87 7.39	3.84 7.31	3.81 7.23	3.77 7.14	3.75 7.09	3.72 7.02	3.71 6.99	3.69 6.94	3.68 6.90	3.67 6.88
7	5.59 12.25	4.74 9.55	4.35 8.45	4.12 7.85	3.97 7.46	3.87 7.19	3.79 7.00	3.73 6.84	3.68 6.71	3.63 6.62	3.60 6.54	3.57 6.47	3.52 6.35	3.49 6.27	3.44 6.15	3.41 6.07	3.38 5.98	3.34 5.90	3.32 5.85	3.29 5.78	3.28 5.75	3.25 5.70	3.24 5.67	3.23 5.65
8	5.32 11.26	4.46 8.65	4.07 7.59	3.84 7.01	3.69 6.63	3.58 6.37	3.50 6.19	3.44 6.03	3.39 5.91	3.34 5.82	3.31 5.74	3.28 5.67	3.23 5.56	3.20 5.48	3.15 5.36	3.12 5.28	3.08 5.20	3.05 5.11	3.03 5.06	3.00 5.00	2.98 4.96	2.96 4.91	2.94 4.88	2.93 4.86
9	5.12 10.56	4.26 8.02	3.86 6.99	3.63 6.42	3.48 6.06	3.37 5.80	3.29 5.62	3.23 5.47	3.18 5.35	3.13 5.26	3.10 5.18	3.07 5.11	3.02 5.00	2.98 4.92	2.93 4.80	2.90 4.73	2.86 4.64	2.82 4.56	2.80 4.51	2.77 4.45	2.76 4.41	2.73 4.36	2.72 4.33	2.71 4.31
10	4.96 10.04	4.10 7.56	3.71 6.55	3.48 5.99	3.33 5.64	3.22 5.39	3.14 5.21	3.07 5.06	3.02 4.95	2.97 4.85	2.94 4.78	2.91 4.71	2.86 4.60	2.82 4.52	2.77 4.41	2.74 4.33	2.70 4.25	2.67 4.17	2.64 4.12	2.61 4.05	2.59 4.01	2.56 3.96	2.55 3.93	2.54 3.91
11	4.84 9.65	3.98 7.20	3.59 6.22	3.36 5.67	3.20 5.32	3.09 5.07	3.01 4.88	2.95 4.74	2.90 4.63	2.86 4.54	2.82 4.46	2.79 4.40	2.74 4.29	2.70 4.21	2.65 4.10	2.61 4.02	2.57 3.94	2.53 3.86	2.50 3.80	2.47 3.74	2.45 3.70	2.42 3.66	2.41 3.62	2.40 3.60
12	4.75 9.33	3.88 6.93	3.49 5.95	3.26 5.41	3.11 5.06	3.00 4.82	2.92 4.65	2.85 4.50	2.80 4.39	2.76 4.30	2.72 4.22	2.69 4.16	2.64 4.05	2.60 3.98	2.54 3.86	2.50 3.78	2.46 3.70	2.42 3.61	2.40 3.56	2.36 3.49	2.35 3.46	2.32 3.41	2.31 3.38	2.30 3.36
13	4.67 9.07	3.80 6.70	3.41 5.74	3.18 5.20	3.02 4.86	2.92 4.62	2.84 4.44	2.77 4.30	2.72 4.19	2.67 4.10	2.63 4.02	2.60 3.96	2.55 3.85	2.51 3.78	2.46 3.67	2.42 3.59	2.38 3.51	2.34 3.42	2.32 3.37	2.28 3.30	2.26 3.27	2.24 3.21	2.22 3.18	2.21 3.16

Tabel I (vervolg)

Kritieke waarden (F) van de variantieanalyse

waarden bij α = 0,05 staan boven
dikgedrukte waarden (onder) bij α = 0,01

df van de noemer

df van de noemer	df van de teller 1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	∞	
14	4.60	3.74	3.34	3.11	2.96	2.85	2.77	2.70	2.65	2.60	2.56	2.53	2.48	2.44	2.39	2.35	2.31	2.27	2.24	2.21	2.19	2.16	2.14	2.13	14
	8.86	**6.51**	**5.56**	**5.03**	**4.69**	**4.46**	**4.28**	**4.14**	**4.03**	**3.94**	**3.86**	**3.80**	**3.70**	**3.62**	**3.51**	**3.43**	**3.34**	**3.26**	**3.21**	**3.14**	**3.11**	**3.06**	**3.02**	**3.00**	
15	4.54	3.68	3.29	3.06	2.90	2.79	2.70	2.64	2.59	2.55	2.51	2.48	2.43	2.39	2.33	2.29	2.25	2.21	2.18	2.15	2.12	2.10	2.08	2.07	15
	8.68	**6.36**	**5.42**	**4.89**	**4.56**	**4.32**	**4.14**	**4.00**	**3.89**	**3.80**	**3.73**	**3.67**	**3.56**	**3.48**	**3.36**	**3.29**	**3.20**	**3.12**	**3.07**	**3.00**	**2.97**	**2.92**	**2.89**	**2.87**	
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.45	2.42	2.37	2.33	2.28	2.24	2.20	2.16	2.13	2.09	2.07	2.04	2.02	2.01	16
	8.53	**6.23**	**5.29**	**4.77**	**4.44**	**4.20**	**4.03**	**3.89**	**3.78**	**3.69**	**3.61**	**3.55**	**3.45**	**3.37**	**3.25**	**3.18**	**3.10**	**3.01**	**2.96**	**2.98**	**2.86**	**2.80**	**2.77**	**2.75**	
17	4.45	3.59	3.20	2.96	2.81	2.70	2.62	2.55	2.50	2.45	2.41	2.38	2.33	2.29	2.23	2.19	2.15	2.11	2.08	2.04	2.02	1.99	1.97	1.96	17
	8.40	**6.11**	**5.18**	**4.67**	**4.34**	**4.10**	**3.93**	**3.79**	**3.68**	**3.59**	**3.52**	**3.45**	**3.35**	**3.27**	**3.16**	**3.08**	**3.00**	**2.92**	**2.86**	**2.79**	**2.76**	**2.70**	**2.67**	**2.65**	
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.37	2.34	2.29	2.25	2.19	2.15	2.11	2.07	2.04	2.00	1.98	1.95	1.93	1.92	18
	8.28	**6.01**	**5.09**	**4.58**	**4.25**	**4.01**	**3.85**	**3.71**	**3.60**	**3.51**	**3.44**	**3.37**	**3.27**	**3.19**	**3.07**	**3.00**	**2.91**	**2.83**	**2.78**	**2.71**	**2.68**	**2.62**	**2.59**	**2.57**	
19	4.38	3.52	3.13	2.90	2.74	2.63	2.55	2.48	2.43	2.38	2.34	2.31	2.26	2.21	2.15	2.11	2.07	2.02	2.00	1.96	1.94	1.91	1.90	1.88	19
	8.18	**5.93**	**5.01**	**4.50**	**4.17**	**3.94**	**3.77**	**3.63**	**3.52**	**3.43**	**3.36**	**3.30**	**3.19**	**3.12**	**3.00**	**2.92**	**2.84**	**2.76**	**2.70**	**2.63**	**2.60**	**2.54**	**2.51**	**2.49**	
20	4.35	3.49	3.10	2.87	2.71	2.60	2.52	2.45	2.40	2.35	2.31	2.28	2.23	2.18	2.12	2.08	2.04	1.99	1.96	1.92	1.90	1.87	1.85	1.84	20
	8.10	**5.85**	**4.94**	**4.43**	**4.10**	**3.87**	**3.71**	**3.56**	**3.45**	**3.37**	**3.30**	**3.23**	**3.13**	**3.05**	**2.94**	**2.86**	**2.77**	**2.69**	**2.63**	**2.56**	**2.53**	**2.47**	**2.44**	**2.42**	
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.28	2.25	2.20	2.15	2.09	2.05	2.00	1.96	1.93	1.89	1.87	1.84	1.82	1.81	21
	8.02	**5.78**	**4.87**	**4.37**	**4.04**	**3.81**	**3.65**	**3.51**	**3.40**	**3.31**	**3.24**	**3.17**	**3.07**	**2.99**	**2.88**	**2.80**	**2.72**	**2.63**	**2.58**	**2.51**	**2.47**	**2.42**	**2.38**	**2.36**	
22	4.30	3.44	3.05	2.82	2.66	2.55	2.47	2.40	2.35	2.30	2.26	2.23	2.18	2.13	2.07	2.03	1.98	1.93	1.91	1.87	1.84	1.81	1.80	1.78	22
	7.94	**5.72**	**4.82**	**4.31**	**3.99**	**3.76**	**3.59**	**3.45**	**3.35**	**3.26**	**3.18**	**3.12**	**3.02**	**2.94**	**2.83**	**2.75**	**2.67**	**2.58**	**2.53**	**2.46**	**2.42**	**2.37**	**2.33**	**2.31**	
23	4.28	3.42	3.03	2.80	2.64	2.53	2.45	2.38	2.32	2.28	2.24	2.20	2.14	2.10	2.04	2.00	1.96	1.91	1.88	1.84	1.82	1.79	1.77	1.76	23
	7.88	**5.66**	**4.76**	**4.26**	**3.94**	**3.71**	**3.54**	**3.41**	**3.30**	**3.21**	**3.14**	**3.07**	**2.97**	**2.89**	**2.78**	**2.70**	**2.62**	**2.53**	**2.48**	**2.41**	**2.37**	**2.32**	**2.28**	**2.26**	
24	4.26	3.40	3.01	2.78	2.62	2.51	2.43	2.36	2.30	2.26	2.22	2.18	2.13	2.09	2.02	1.98	1.94	1.89	1.86	1.82	1.80	1.76	1.74	1.73	24
	7.82	**5.61**	**4.72**	**4.22**	**3.90**	**3.67**	**3.50**	**3.36**	**3.25**	**3.17**	**3.09**	**3.03**	**2.93**	**2.85**	**2.74**	**2.66**	**2.58**	**2.49**	**2.44**	**2.36**	**2.33**	**2.27**	**2.23**	**2.21**	
25	4.24	3.38	2.99	2.76	2.60	2.49	2.41	2.34	2.28	2.24	2.20	2.16	2.11	2.06	2.00	1.96	1.92	1.87	1.84	1.80	1.77	1.74	1.72	1.71	25
	7.77	**5.57**	**4.68**	**4.18**	**3.86**	**3.63**	**3.46**	**3.32**	**3.21**	**3.13**	**3.05**	**2.99**	**2.89**	**2.81**	**2.70**	**2.62**	**2.54**	**2.45**	**2.40**	**2.32**	**2.29**	**2.23**	**2.19**	**2.17**	
26	4.22	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.18	2.15	2.10	2.05	1.99	1.95	1.90	1.85	1.82	1.78	1.76	1.72	1.70	1.69	26
	7.72	**5.53**	**4.64**	**4.14**	**3.82**	**3.59**	**3.42**	**3.29**	**3.17**	**3.09**	**3.02**	**2.96**	**2.86**	**2.77**	**2.66**	**2.58**	**2.50**	**2.41**	**2.36**	**2.28**	**2.25**	**2.19**	**2.15**	**2.13**	

Tabel I (vervolg)

Kritieke waarden (F) van de variantieanalyse

waarden bij α = 0,05 staan boven
dikgedrukte waarden (onder) bij α = 0,01

df van de noemer

	df van de teller																								
	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	∞	
27	4.21 7.68	3.35 5.49	2.96 4.60	2.73 4.11	2.57 3.79	2.46 3.56	2.37 3.39	2.30 3.26	2.25 3.14	2.20 3.06	2.16 2.98	2.13 2.93	2.08 2.83	2.03 2.74	1.97 2.63	1.93 2.55	1.88 2.47	1.84 2.38	1.80 2.33	1.76 2.25	1.74 2.21	1.71 2.16	1.68 2.12	1.67 2.10	27
28	4.20 7.64	3.34 5.45	2.95 4.57	2.71 4.07	2.56 3.76	2.44 3.53	2.36 3.36	2.29 3.23	2.24 3.11	2.19 3.03	2.15 2.95	2.12 2.90	2.06 2.80	2.02 2.71	1.96 2.60	1.91 2.52	1.87 2.44	1.81 2.35	1.78 2.30	1.75 2.22	1.72 2.18	1.69 2.13	1.67 2.09	1.65 2.06	28
29	4.18 7.60	3.33 5.42	2.93 4.54	2.70 4.04	2.54 3.73	2.43 3.50	2.35 3.33	2.28 3.20	2.22 3.08	2.18 3.00	2.14 2.92	2.10 2.87	2.05 2.77	2.00 2.68	1.94 2.57	1.90 2.49	1.85 2.41	1.80 2.32	1.77 2.27	1.73 2.19	1.71 2.15	1.68 2.10	1.65 2.06	1.64 2.03	29
30	4.17 7.56	3.32 5.39	2.92 4.51	2.69 4.02	2.53 3.70	2.42 3.47	2.34 3.30	2.27 3.17	2.21 3.06	2.16 2.98	2.12 2.90	2.09 2.84	2.04 2.74	1.99 2.66	1.93 2.55	1.89 2.47	1.84 2.38	1.79 2.29	1.76 2.24	1.72 2.16	1.69 2.13	1.66 2.07	1.64 2.03	1.62 2.01	30
32	4.15 7.50	3.30 5.34	2.90 4.46	2.67 3.97	2.51 3.66	2.40 3.42	2.32 3.25	2.25 3.12	2.19 3.01	2.14 2.94	2.10 2.86	2.07 2.80	2.02 2.70	1.97 2.62	1.91 2.51	1.86 2.42	1.82 2.34	1.76 2.25	1.74 2.20	1.69 2.12	1.67 2.08	1.64 2.02	1.61 1.98	1.59 1.96	32
34	4.13 7.44	3.28 5.29	2.88 4.42	2.65 3.93	2.49 3.61	2.38 3.38	2.30 3.21	2.23 3.08	2.17 2.97	2.12 2.89	2.08 2.82	2.05 2.76	2.00 2.66	1.95 2.58	1.89 2.47	1.84 2.38	1.80 2.30	1.74 2.21	1.71 2.15	1.67 2.08	1.64 2.04	1.61 1.98	1.59 1.94	1.57 1.91	34
36	4.11 7.39	3.26 5.25	2.86 4.38	2.63 3.89	2.48 3.58	2.36 3.35	2.28 3.18	2.21 3.04	2.15 2.94	2.10 2.86	2.06 2.78	2.03 2.72	1.98 2.62	1.93 2.54	1.87 2.43	1.82 2.35	1.78 2.26	1.72 2.17	1.69 2.12	1.65 2.04	1.62 2.00	1.59 1.94	1.56 1.90	1.55 1.87	36
38	4.10 7.35	3.25 5.21	2.85 4.34	2.62 3.86	2.46 3.54	2.35 3.32	2.26 3.15	2.19 3.02	2.14 2.91	2.09 2.82	2.05 2.75	2.02 2.69	1.96 2.59	1.92 2.51	1.85 2.40	1.80 2.32	1.76 2.22	1.71 2.14	1.67 2.08	1.63 2.00	1.60 1.97	1.57 1.90	1.54 1.86	1.53 1.84	38
40	4.08 7.31	3.23 5.18	2.84 4.31	2.61 3.83	2.45 3.51	2.34 3.29	2.25 3.12	2.18 2.99	2.12 2.88	2.07 2.80	2.04 2.73	2.00 2.66	1.95 2.56	1.90 2.49	1.84 2.37	1.79 2.29	1.74 2.20	1.69 2.11	1.66 2.05	1.61 1.97	1.59 1.94	1.55 1.88	1.53 1.84	1.51 1.81	40
42	4.07 7.27	3.22 5.15	2.83 4.29	2.59 3.80	2.44 3.49	2.32 3.26	2.24 3.10	2.17 2.96	2.11 2.86	2.06 2.77	2.02 2.70	1.99 2.64	1.94 2.54	1.89 2.46	1.82 2.35	1.78 2.26	1.73 2.17	1.68 2.08	1.64 2.02	1.60 1.94	1.57 1.91	1.54 1.85	1.51 1.80	1.49 1.78	42
44	4.06 7.24	3.21 5.12	2.82 4.26	2.58 3.78	2.43 3.46	2.31 3.24	2.23 3.07	2.16 2.94	2.10 2.84	2.05 2.75	2.01 2.68	1.98 2.62	1.92 2.52	1.88 2.44	1.81 2.32	1.76 2.24	1.72 2.15	1.66 2.06	1.63 2.00	1.58 1.92	1.56 1.88	1.52 1.82	1.50 1.78	1.48 1.75	44
46	4.05 7.21	3.20 5.10	2.81 4.24	2.57 3.76	2.42 3.44	2.30 3.22	2.22 3.05	2.14 2.92	2.09 2.82	2.04 2.73	2.00 2.66	1.97 2.60	1.91 2.50	1.87 2.42	1.80 2.30	1.75 2.22	1.71 2.13	1.65 2.04	1.62 1.98	1.57 1.90	1.54 1.86	1.51 1.80	1.48 1.76	1.46 1.72	46
48	4.04 7.19	3.19 5.08	2.80 4.22	2.56 3.74	2.41 3.42	2.30 3.20	2.21 3.04	2.14 2.90	2.08 2.80	2.03 2.71	1.99 2.64	1.96 2.58	1.90 2.48	1.86 2.40	1.79 2.28	1.74 2.20	1.70 2.11	1.64 2.02	1.61 1.96	1.56 1.88	1.53 1.84	1.50 1.78	1.47 1.73	1.45 1.70	48

Tabel I (vervolg)

Kritieke waarden (F) van de variantieanalyse

waarden bij α = 0,05 staan boven
dikgedrukte waarden (onder) bij α = 0,01

df van de noemer

df van de teller	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	∞	df van de teller
50	4.03 7.17	3.18 5.06	2.79 4.20	2.56 3.72	2.40 3.41	2.29 3.18	2.20 3.02	2.13 2.88	2.07 2.78	2.02 2.70	1.98 2.62	1.95 2.56	1.90 2.46	1.85 2.39	1.78 2.26	1.74 2.18	1.69 2.10	1.63 2.00	1.60 1.94	1.55 1.86	1.52 1.82	1.48 1.76	1.46 1.71	1.44 1.68	50
55	4.02 7.12	3.17 5.01	2.78 4.16	2.54 3.68	2.38 3.37	2.27 3.15	2.18 2.98	2.11 2.85	2.05 2.75	2.00 2.66	1.97 2.59	1.93 2.53	1.88 2.43	1.83 2.35	1.76 2.23	1.72 2.15	1.67 2.06	1.61 1.96	1.58 1.90	1.52 1.82	1.50 1.78	1.46 1.71	1.43 1.66	1.41 1.64	55
60	4.00 7.08	3.15 4.98	2.76 4.13	2.52 3.65	2.37 3.34	2.25 3.12	2.17 2.95	2.10 2.82	2.04 2.72	1.99 2.63	1.95 2.56	1.92 2.50	1.86 2.40	1.81 2.32	1.75 2.20	1.70 2.12	1.65 2.03	1.59 1.93	1.56 1.87	1.50 1.79	1.48 1.74	1.44 1.68	1.41 1.63	1.39 1.60	60
65	3.99 7.04	3.14 4.95	2.75 4.10	2.51 3.62	2.36 3.31	2.24 3.09	2.15 2.93	2.08 2.79	2.02 2.70	1.98 2.61	1.94 2.54	1.90 2.47	1.85 2.37	1.80 2.30	1.73 2.18	1.68 2.09	1.63 2.00	1.57 1.90	1.54 1.84	1.49 1.76	1.46 1.71	1.42 1.64	1.39 1.60	1.37 1.56	65
70	3.98 7.01	3.13 4.92	2.74 4.08	2.50 3.60	2.35 3.29	2.23 3.07	2.14 2.91	2.07 2.77	2.01 2.67	1.97 2.59	1.93 2.51	1.89 2.45	1.84 2.35	1.79 2.28	1.72 2.15	1.67 2.07	1.62 1.98	1.56 1.88	1.53 1.82	1.47 1.74	1.45 1.69	1.40 1.62	1.37 1.56	1.35 1.53	70
80	3.96 6.96	3.11 4.88	2.72 4.04	2.48 3.56	2.33 3.25	2.21 3.04	2.12 2.87	2.05 2.74	1.99 2.64	1.95 2.55	1.91 2.48	1.88 2.41	1.82 2.32	1.77 2.24	1.70 2.11	1.65 2.03	1.60 1.94	1.54 1.84	1.51 1.78	1.45 1.70	1.42 1.65	1.38 1.57	1.35 1.52	1.32 1.49	80
100	3.94 6.90	3.09 4.82	2.70 3.98	2.46 3.51	2.30 3.20	2.19 2.99	2.10 2.82	2.03 2.69	1.97 2.59	1.92 2.51	1.88 2.43	1.85 2.36	1.79 2.26	1.75 2.19	1.68 2.06	1.63 1.98	1.57 1.89	1.51 1.79	1.48 1.73	1.42 1.64	1.39 1.59	1.34 1.51	1.30 1.46	1.28 1.43	100
125	3.92 6.84	3.07 4.78	2.68 3.94	2.44 3.47	2.29 3.17	2.17 2.95	2.08 2.79	2.01 2.65	1.95 2.56	1.90 2.47	1.86 2.40	1.83 2.33	1.77 2.23	1.72 2.15	1.65 2.03	1.60 1.94	1.55 1.85	1.49 1.75	1.45 1.68	1.39 1.59	1.36 1.54	1.31 1.46	1.27 1.40	1.25 1.37	125
150	3.91 6.81	3.06 4.75	2.67 3.91	2.43 3.44	2.27 3.14	2.16 2.92	2.07 2.76	2.00 2.62	1.94 2.53	1.89 2.44	1.85 2.37	1.82 2.30	1.76 2.20	1.71 2.12	1.64 2.00	1.59 1.91	1.54 1.83	1.47 1.72	1.44 1.66	1.37 1.56	1.34 1.51	1.29 1.43	1.25 1.37	1.22 1.33	150
200	3.89 6.76	3.04 4.71	2.65 3.88	2.41 3.41	2.26 3.11	2.14 2.90	2.05 2.73	1.98 2.60	1.92 2.50	1.87 2.41	1.83 2.34	1.80 2.28	1.74 2.17	1.69 2.09	1.62 1.97	1.57 1.88	1.52 1.79	1.45 1.69	1.42 1.62	1.35 1.53	1.32 1.48	1.26 1.39	1.22 1.33	1.19 1.28	200
400	3.86 6.70	3.02 4.66	2.62 3.83	2.39 3.36	2.23 3.06	2.12 2.85	2.03 2.69	1.96 2.55	1.90 2.46	1.85 2.37	1.81 2.29	1.78 2.23	1.72 2.12	1.67 2.04	1.60 1.92	1.54 1.84	1.49 1.74	1.42 1.64	1.38 1.57	1.32 1.47	1.28 1.42	1.22 1.32	1.16 1.24	1.13 1.19	400
1000	3.85 6.66	3.00 4.62	2.61 3.80	2.38 3.34	2.22 3.04	2.10 2.82	2.02 2.66	1.95 2.53	1.89 2.43	1.84 2.34	1.80 2.26	1.76 2.20	1.70 2.09	1.65 2.01	1.58 1.89	1.53 1.81	1.47 1.71	1.41 1.61	1.36 1.54	1.30 1.44	1.26 1.38	1.19 1.28	1.13 1.19	1.08 1.11	1000
∞	3.84 6.64	2.99 4.60	2.60 3.78	2.37 3.32	2.21 3.02	2.09 2.80	2.01 2.64	1.94 2.51	1.88 2.41	1.83 2.32	1.79 2.24	1.75 2.18	1.69 2.07	1.64 1.99	1.57 1.87	1.52 1.79	1.46 1.69	1.40 1.59	1.35 1.52	1.28 1.41	1.24 1.36	1.17 1.25	1.11 1.15	1.00 1.00	∞

APPENDIX II

Toetsingsschema

	parametrisch	*niet-parametrisch*
Toets van 1 steekproef-gemiddelde in de populatie populatiegegevens bekend	**z-toets** tabel A	
Toets van 1 steekproef-gemiddelde in de populatie populatiegegevens niet bekend	**t-toets voor 1 steekproefgemiddelde** tabel B df = n - 1	
Toets voor verschil tussen twee onafhankelijke steekproef-gemiddelden	**t-toets voor onafhankelijke groepen** tabel B df = n1 + n2 - 2	**Mann-Whitney-U toets** tabel C
Toets voor verschil tussen twee afhankelijke steekproef-gemiddelden	**t-toets voor afhankelijke groepen** tabel B df= n-1	**Wilcoxon toets** tabel D
Toets voor verschil tussen twee correlatie-coëfficiënten	**z_r-toets** tabel A voor kritieke waarden, tabel H voor z-waarden van de correlaties	
Toets voor het vaststellen van samenhang tussen twee variabelen waarvan 1 van beide of allebei nominaal gemeten zijn		**χ^2-toets** tabel G
Toets voor vaststellen van samenhang	**Pearson correlatie toets** tabel E	**Spearman rangorde correlatie toets** tabel F
Toets voor verschil tussen meer dan twee steekproef-gemiddelden	**Eenweg-variantieanalyse** tabel I	**Kruskal-Wallis toets** tabel G
Toets voor verschil tussen groepen op basis van twee onafhankelijke variabelen	**Tweeweg-variantieanalyse** tabel I	

Over de auteurs

Louis Tavecchio (1946, Den Haag) studeerde psychologie aan de Universiteit van Amsterdam (doctoraal 1971) en promoveerde aldaar in 1977 op het proefschrift *'Quantification of teaching behavior in physical education: A methodological study'*. Hij was tot eind 1975 verbonden aan de vakgroep Psychofysiologie van de Universiteit van Amsterdam. Sindsdien is hij werkzaam bij de vakgroep Algemene Pedagogiek van de Rijksuniversiteit Leiden, vanaf 1984 als universitair hoofddocent Empirische Pedagogiek.
Zijn wetenschappelijk onderzoek richtte zich aanvankelijk op evaluatie en effecten van leerkrachtgedrag. Later raakte hij nauw betrokken bij het project 'Vroegkinderlijke Opvoeding', waarin de nadruk ligt op empirisch onderzoek naar gehechtheidsrelaties ('attachment'). Daarnaast verricht hij onderzoek naar seksegebonden sociaal-emotionele en gedragsproblemen bij jongens, en naar de pedagogische aspecten van kinderopvang.
Zijn onderwijstaken bestaan onder andere uit methodologie, methoden & technieken van sociaal-wetenschappelijk onderzoek, en statistiek. Daarnaast geeft hij onderwijs op de gebieden gezinsopvoeding en kinderopvang.

Jacqueline Tournier-Hoogeweij werd geboren in 1969 te Rotterdam. Zij studeerde voorlichtingskunde aan de Rijksuniversiteit Leiden en studeerde in 1992 af op het onderwerp Geloofwaardigheid van het tv-nieuws zoals dat door kinderen wordt ervaren. Deze scriptie kwam tot stand naar aanleiding van een groot onderzoek.
Reeds in haar studietijd was zij actief in statistiek- en m & t-onderwijs. Daarna is zij ruim vier jaar docent geweest aan de Rijksuniversiteit Leiden, met als taak het verzorgen van onderwijs op het gebied van statistiek, methoden en technieken van onderzoek en tevens meer inhoudelijke onderdelen als vroegkinderlijke opvoeding & ontwikkeling en kinderopvang. Zij verwierf in deze periode tweemaal de onderwijsprijs Pedagogiek. Naast onderwijs was zij ook betrokken bij verschillende onderzoeksprojecten op het gebied van kinderopvang en Kind & Media. Momenteel verzorgt zij op freelancebasis cursussen op het gebied van statistiek, methoden en technieken, en wiskunde.