



Universiteit
Leiden
The Netherlands

Advances in Survival Analysis and Optimal Scaling Methods

Willems, S.J.W.

Citation

Willems, S. J. W. (2020, March 19). *Advances in Survival Analysis and Optimal Scaling Methods*. Retrieved from <https://hdl.handle.net/1887/87058>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/87058>

Note: To cite this publication please use the final published version (if applicable).

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/87058> holds various files of this Leiden University dissertation.

Author: Willems, S.J.W.

Title: Advances in Survival Analysis and Optimal Scaling Methods

Issue Date: 2020-03-19

BIBLIOGRAPHY

- Ajzen, I. *The theory of planned behaviour*. Organisational behaviour and human decision processes, 50:179–211, 1991. – Cited on page 55.
- Andersen, P. K. and Gill, R. D. *Cox's Regression Model for Counting Processes: A Large Sample Study*. The Annals of Statistics, 10(4):1100–1120, 1982. – Cited on page 33.
- Bender, R., Augustin, T., and Blettner, M. *Generating survival times to simulate Cox proportional hazards models*. Statistics in Medicine, 24(11):1713–1723, 2005. – Cited on page 38.
- Boršič, D. and Kavkler, A. *Modeling unemployment duration in Slovenia using Cox regression models*. Transition Studies Review, 16(1):145–156, 2009. – Cited on page 45.
- Boyle, C. M. *Difference Between Patients' and Doctors' Interpretation of Some Common Medical Terms*. British Medical Journal, 2(5704):286–289, 1970. – Cited on page 140.
- Breslow, N. E. *Contribution to the Discussion of Paper by D.R. Cox*. Journal of the Royal Statistical Society. Series B (Methodological), 34:216–217, 1972. – Cited on pages 81 and 99.
- Brouwer, S., Bakker, R. H., and Schellekens, J. M. H. *Predictors for re-employment success in newly unemployed: A prospective cohort study*. Journal of Vocational Behaviour, 89:32–38, 2015. – Cited on pages 46, 47, 51, and 53.
- Bryant, G. D. and Norman, G. R. *Expressions of probability: words and numbers*. New England Journal of Medicine, 302(7):411, 1980. – Cited on pages 136 and 140.
- Budescu, D. V. and Wallsten, T. S. *Consistency in interpretation of probabilistic phrases*. Organizational Behavior and Human Decision Processes, 36(3):391–405, 1985. – Cited on page 136.

- Castro, C. M., Wilson, C., Wang, F., and Schillinger, D. *Babel babble: Physicians' use of unclarified medical jargon with patients*. American Journal of Health Behavior, 31(Suppl 1):S85–95, 2007. – Cited on page 140.
- Chang, I. M., Gelman, R., and Pagano, M. *Corrected group prognostic curves and summary statistics*. Journal of Chronic Diseases, 35(8):669 – 674, 1982. ISSN 0021-9681. – Cited on page 30.
- Cox, D. R. *Regression Models and Life Tables (with Discussion)*. Journal of the Royal Statistical Society B, 34:187–220, 1972. – Cited on pages 8 and 53.
- Davidson, R. A. and Chrisman, H. H. *Translations of uncertainty expressions in Canadian accounting and auditing standards*. Journal of International Accounting, Auditing and Taxation, 3(2):187 – 203, 1994. – Cited on page 137.
- Doupnik, T. S. and Richter, M. *Interpretation of uncertainty expressions: a cross-national study*. Accounting, Organizations and Society, 28(1):15 – 35, 2003. – Cited on page 137.
- Druzdzel, M. J. *Verbal Uncertainty Expressions: Literature Review*. Technical Report, 1989. – Cited on pages 10, 136, 138, and 140.
- Eekhof, J. A. H., Mol, S. S. L., and Pielage, J. C. *Is doorgaans vaker dan dikwijls; of hoe vaak is soms?* Nederlands Tijdschrift voor Geneeskunde, 136(1):41, 1992. – Cited on pages 137, 141, 142, and 151.
- Erev, I. and Cohen, B. L. *Verbal versus Numerical Probabilities: Efficiency, Biases, and the Preference Paradox*. Organizational Behavior and Human Decision Processes, 45:1–18, 1990. – Cited on page 138.
- European Food Safety Authority, Hart, A., Maxim, L., Siegrist, M., da Cruz, C., Merten, C., Lahaniatis, M., and Smith, A. *Outcome of the Public Consultation on the draft Guidance on Communication of Uncertainty in Scientific Assessments*. EFSA Journal, 17(1):1–73, 2019. – Cited on page 136.
- Fox, J. and Carvalho, M. S. *The RcmdrPlugin.survival Package: Extending the R Commander Interface to Survival Analysis*. Journal of Statistical Software, 49(7):1–32, 2012. – Cited on page 34.
- Friedman, J. H. and Stuetzle, W. *Projection Pursuit Regression*. Journal of the American Statistical Association, 76(376):817–823, 1981. – Cited on pages 102 and 105.
- Gifi, A. *Nonlinear Multivariate Analysis*. John Wiley & Sons, Ltd., 1990. Original work published 1981. – Cited on pages 9, 69, 102, and 105.

- Google. *Google Translate*. <https://translate.google.com>, June 2018. – Cited on page 141.
- Graf, E., Schmoor, C., Sauerbrei, W., and Schumacher, M. *Assessment and comparison of prognostic classification schemes for survival data*. *Statistics in Medicine*, 18:2529 – 2545, 1999. – Cited on page 65.
- Green, P. J. *Iteratively Reweighted Least Squares for Maximum Likelihood Estimation, and some Robust and Resistant Alternatives*. *Journal of the Royal Statistical Society: Series B*, 46(2):149–192, 1984. – Cited on page 99.
- Harrell, F. E., Lee, K. L., and Mark, D. B. *Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors*. *Statistics in Medicine*, 15:361–387, 1996. – Cited on page 65.
- Harris, A. J. L., Corner, A., Xu, J., and Du, X. *Lost in translation? Interpretations of the probability phrases used by the Intergovernmental Panel on Climate Change in China and the UK*. *Climatic Change*, 121(2):415–425, 2013. – Cited on page 137.
- Hartmann, A., Van der Kooij, A. J., and Zeeck, A. *Exploring nonlinear relations: Models of clinical decision making by regression with optimal scaling*. *Psychotherapy Research*, 19(4-5):482–492, 2009. – Cited on page 115.
- Hastie, T. J. and Tibshirani, R. J. *Generalized Additive Models*. *Monographs on Statistics & Applied Probability*. Chapman & Hall/CRC, 1990. – Cited on pages 102, 105, and 131.
- Jenkins, S. C., Harris, A. J. L., and Lark, R. M. *Understanding ‘Unlikely (20% Likelihood)’ or ‘20% Likelihood (Unlikely)’ Outcomes: The Robustness of the Extremity Effect*. *Journal of Behavioural Decision Making*, 31, 2018. – Cited on page 152.
- Jolliffe, I. T. *Principal component analysis*. New York, NY: Springer-Verlag, 2002. – Cited on page 52.
- Juanchich, M. and Sirota, M. *Do people really prefer verbal probabilities?* *Psychological Research*, Jun 2019. – Cited on page 138.
- Kaplan, E. L. and Meier, P. *Nonparametric Estimation from Incomplete Observations*. *Journal of the American Statistical Association*, 53:457–481, 1958. – Cited on pages 15 and 35.

- Kavkler, A., Dănaćică, D. E., Babucea, A. G., Bićanić, I., Böhm, B., Tevdovski, D., Toševska, K., and Boršič, D. *Cox regression models for unemployment duration in Romania, Austria, Slovenia, Croatia, and Macedonia*. Romanian Journal of Economic Forecasting, 10(2):81–104, 2009. – Cited on page 45.
- Kim, Y., Kim, J., and Kim, Y. *Blockwise Sparse Regression*. Statistica Sinica, 16(2):375–390, 2006. – Cited on page 132.
- Klein, J. P. and Moeschberger, M. L. *Survival Analysis: Techniques for Censored and Truncated Data*. Statistics for Biology and Health. Springer, second edition edition, 2003. – Cited on pages 53 and 71.
- Kleinbaum, D. G. and Klein, M. *Survival Analysis: A Self-Learning Text*. Statistics for Biology and Health. Springer, third edition edition, 2012. – Cited on page 16.
- Kruskal, J. B. *Nonmetric multidimensional scaling: a numerical method*. Psychometrika, 29:115–129, 1964. – Cited on pages 74, 78, 103, and 106.
- Lichtenstein, S. and Newman, J. R. *Empirical scaling of common verbal phrases associated with numerical probabilities*. Psychonomic Science, 9(10):563–564, 1967. – Cited on pages 10, 138, and 139.
- Lim, T. S., Loh, W. Y., and Shih, Y. S. *A Comparison of Prediction Accuracy, Complexity, and Training Time of Thirty-Three Old and New Classification Algorithms*. Machine Learning, 40(3):203–228, Sep 2000. – Cited on page 120.
- Linting, M. and van der Kooij, A. *Nonlinear Principal Components Analysis With CATPCA: A Tutorial*. Journal of Personality Assessment, 94(1):12–25, 2012. – Cited on pages 52, 54, and 55.
- Linting, M., Meulman, J. J., Groenen, P. J., and van der Kooij, A. J. *Nonlinear principal components analysis: introduction and application*. Psychological Methods, 12(3):336–358, September 2007. – Cited on pages 44, 52, 70, and 95.
- MacLeod, A. and Pietravallo, S. *Communicating risk: variability of interpreting qualitative terms*. EPPO Bulletin, 47(1):57–68, 2017. – Cited on page 140.
- McCullagh, P. and Nelder, J. A. *Generalized Linear Models*. Monographs on Statistics & Applied Probability. Chapman & Hall/CRC, 1989. – Cited on pages 9 and 102.
- Meulman, J. J., Van der Kooij, A. J., and Heiser, W. J. *Principal components analysis with nonlinear optimal scaling transformations for ordinal and nominal data*. In Kaplan, D., editor, *Handbook of Quantitative Methodology for the*

- Social Sciences.*, pages 49–70. Sage, London, 2004. – Cited on pages 10, 44, 52, 70, and 95.
- Meulman, J. J., Van der Kooij, A. J., and Duisters, K. L. W. *ROS Regression: Integrating Regularization with Optimal Scaling Regression*. Statistical Science, 34:361–390, 2019. – Cited on pages 9, 70, 75, 104, and 132.
- Montgomery, S. A. and Åsberg, M. *A new depression scale designed to be sensitive to change*. The British Journal of Psychiatry, 134(4):382–9, 1979. – Cited on page 22.
- Mooney, C. Z. *Monte Carlo Simulation (Quantitative Applications in the Social Sciences)*. SAGE Publications, April 1997. ISBN 0803959435. – Cited on page 38.
- Mosteller, F. and Youtz, C. *Quantifying Probabilistic Expressions*. Statistical Science, 5(1):2–12, 1990. – Cited on pages 139, 140, and 146.
- Pander Maat, H. and Klaassen, R. *De invloed van de formulering van informatie over bijwerkingen op interpretaties door patiënten en op gemelde bijwerkingen*. Nederlands Tijdschrift voor Geneeskunde, 140(8):424–427, 1996. – Cited on pages 137, 139, 140, and 141.
- Putter, H., Fiocco, M., and Geskus, R. B. *Tutorial in biostatistics: competing risks and multi-state models*. Statistics in Medicine, 26(11):2389–2430, 2007. – Cited on page 53.
- Qualtrics. *Qualtrics*. <https://www.qualtrics.com>, 2005-2018. – Cited on page 143.
- R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2018. URL <https://www.R-project.org/>. – Cited on pages 11, 18, and 114.
- Ramsay, J. O. *Monotone Regression Splines in Action*. Statistical Science, 3(4): 425–442, November 1988. – Cited on pages 75, 103, and 106.
- Reagan, R. T., Mosteller, F., and Youtz, C. *Quantitative meanings of verbal probability expressions*. Journal of Applied Psychology, 74(3):433–442, 1989. – Cited on pages 10, 137, 138, and 139.
- Renooij, S. and Witteman, C. *Talking probabilities: communicating probabilistic information with words and numbers*. International Journal of Approximate Reasoning, 22(3):169–194, 1999. – Cited on pages 137 and 141.



- Robins, J. M. *Information Recovery and Bias Adjustment in Proportional Hazards Regression Analysis of Randomized Trials Using Surrogate Markers*. Proceedings of the Biopharmaceutical Section, American Statistical Association, pages 24–33, 1993. – Cited on pages 14 and 16.
- Robins, J. M. and Finkelstein, D. M. *Correcting for Noncompliance and Dependent Censoring in an AIDS Clinical Trial with Inverse Probability of Censoring Weighted (IPCW) Log-Rank Tests*. Biometrics, 56:779–788, 2000. – Cited on pages 14, 16, and 18.
- Robins, J. M. and Rotnitzky, A. *Recovery of Information and Adjustment for Dependent Censoring Using Surrogate Markers*. In Jewell, N. P., Dietz, K., and Farewell, V. T., editors, *AIDS Epidemiology*, pages 297–331. Birkhäuser Boston, 1992. ISBN 978-1-4757-1231-5. – Cited on pages 14 and 16.
- Simon, N., Friedman, J., Hastie, T., and Tibshirani, R. *Regularization Paths for Cox’s Proportional Hazards Model via Coordinate Descent*. Journal of Statistical Software, 39(5):1–13, 2011. – Cited on pages 80, 81, 97, 99, and 109.
- SPSS Inc. *SPSS Statistics for Windows, Version 17.0*. Chicago, 2008. – Cited on page 52.
- Stheeman, S. E., Mileman, P. A., van ’t Hof, M. A., and van der Stelt, P. F. *Blind chance? An investigation into the perceived probabilities of phrases used in oral radiology for expressing chance*. Dento maxillo facial radiology, 22:135–139, 1993. – Cited on pages 10, 139, and 140.
- Takane, Y., Young, F. W., and de Leeuw, J. *The principal components of mixed measurement level multivariate data: An alternating least squares method with optimal scaling features*. Psychometrika, 49:279–281, 1978. – Cited on page 10.
- Theil, M. *The role of translations of verbal into numerical probability expressions in risk management: a meta-analysis*. Journal of Risk Research, 5(2):177–186, 2002. – Cited on pages 136, 140, 141, 151, and 152.
- Therneau, T. M. *A Package for Survival Analysis in S*, 2013. URL <http://CRAN.R-project.org/package=survival>. R package version 2.37-4. – Cited on pages 14 and 18.
- Therneau, T. M. and Grambsch, P. M. *Modeling Survival Data: Extending the Cox Model*. Springer, New York, 2000. – Cited on pages 14 and 18.
- Timmermans, D. *The Roles of Experience and Domain of Expertise in Using Numerical and Verbal Probability Terms in Medical Decisions*. Medical Decision Making, 14:146–156, 1994. – Cited on pages 137, 141, and 142.

- Timmermans, D. R. M. and Mileman, P. A. *Lost for words: using verbal terms to express probabilities in oral radiology*. *Dento maxillo facial radiology*, 22(4): 171–172, 1993. – Cited on page 139.
- Tutkun, N. A. and Karasoy, D. *Applying frailty models to analyse the duration of unemployment in Turkey*. *International Journal of Statistics & Economics*, 17(2):44–57, 2016. – Cited on page 45.
- Twitter Inc. *Twitter*. <https://twitter.com>, 2018. – Cited on page 144.
- Van Dale Uitgevers. *Van Dale*. <https://www.vandale.nl>, June 2018. – Cited on page 141.
- Van der Kooij, A. J. *Prediction Accuracy and Stability of Regression with Optimal Scaling Transformations*. PhD thesis, Leiden University, 2007. – Cited on page 104.
- Van der Kooij, A. J. and Meulman, J. J. *Regression with Optimal Scaling*. In Meulman, J., Heiser, W., and Inc., S., editors, *SPSS Categories 10.0.*, pages 1–8, 77–101, 239–246. Chicago: SPSS Inc, 1999. – Cited on pages 9, 52, and 102.
- van Swieten, J. C., Koudstaal, P. J., Visser, M. C., Schouten, H. J., and van Gijn, J. *Interobserver agreement for the assessment of handicap in stroke patients*. *Stroke*, 19:604–607, 1988. – Cited on page 68.
- Venhuizen, G. J., Hut, R., Albers, C., Stoof, C. R., and Smeets, I. *Flooded by jargon: how the interpretation of water-related terms differs between hydrology experts and the general audience*. *Hydrology and Earth System Sciences*, 2019. Accepted for publication. – Cited on page 140.
- Visschers, V. H. M., Meertens, R. M., Passchier, W. W. F., and de Vries, N. N. K. *Probability Information in Risk Communication: A Review of the Research Literature*. *Risk Analysis*, 29(2):267–287, 2009. – Cited on page 136.
- Vroom, V. H. *Work and motivation*. John Wiley & Sons, New York, 1964. – Cited on page 55.
- Wallsten, T. S., Fillenbaum, S., and Cox, J. A. *Base rate effects on the interpretations of probability and frequency expressions*. *Journal of Memory and Language*, 25(5):571–587, 1986. – Cited on pages 11 and 140.
- Wanberg, C. R., Hough, L. M., and Song, Z. *Predictive validity of a multidisciplinary model of re-employment success*. *Journal of Applied Psychology*, 85: 1100–1120, 2002. – Cited on page 45.

- Weber, E. U. and Hilton, D. J. *Contextual Effects in the Interpretations of Probability Words: Perceived Base Rate and Severity of Events*. Journal of Experimental Psychology: Human Perception and Performance, 16(4):781–789, 1990. – Cited on pages 139 and 140.
- Wijnhoven, M. A. and Havinga, H. *The Work Profiler: A digital instrument for selection and diagnosis of the unemployed*. Local Economy, 29(6-7):740–749, 2014. – Cited on page 46.
- Willems, S. J. W., Fiocco, M., and Meulman, J. J. *Optimal scaling for survival analysis with ordinal data*. Computational Statistics & Data Analysis, 115:155 – 171, 2017. – Cited on page 109.
- Winsberg, S. and Ramsay, J. O. *Monotonic transformations to additivity using splines*. Biometrika, 67(3):669–674, 1980. ISSN 0006-3444. – Cited on page 102.
- Wintle, B. C., Fraser, H., Wills, B. C., Nicholson, A. E., and Fidler, F. *Verbal probabilities: Very likely to be somewhat more confusing than numbers*. PLoS ONE, 14(4):1–18, 04 2019. – Cited on page 152.
- Yang, Y. and Zou, H. *A cocktail algorithm for solving the elastic net penalized Cox’s regression in high dimensions*. Statistics and Its Interface, 6:167–173, 2013. – Cited on page 85.
- Young, F. W. *Quantitative analysis of qualitative data*. Psychometrika, 46(4): 357–388, 1981. – Cited on page 102.
- Young, F. W., de Leeuw, J., and Takane, Y. *Regression with qualitative and quantitative variables: An alternating least squares method with optimal scaling features*. Psychometrika, 41:505–528, December 1976. – Cited on pages 9, 102, and 105.
- Yuan, M. and Lin, Y. *Model Selection and Estimation in Regression With Grouped Variables*. Journal of the Royal Statistical Society: Series B, 68:49–67, 02 2006. – Cited on page 132.
- Zeeck, A., Hartmann, A., Küchenhoff, J., Weiss, H., Sammet, I., Gaus, E., Semm, E., Harms, D., Eisenberg, A., Rahm, R., and von Wietersheim, J. *Differenzielle Indikationsstellung stationärer und tagesklinischer Psychotherapie: die DINSTAP-Studie[1]*. Psychotherapie, Psychosomatik, Medizinische Psychologie, 59(9/10):354–363, 2009. – Cited on page 115.

INDEX

- (G)AM: (Generalized) Additive Model
 - application to breast cancer recurrence data, 128–130
 - application to contraceptive method choice data, 124–127
 - Method, 104, 107
- Censoring, 14
 - dependent, 14–16
 - correct for, *see* IPCW
 - informative, 14
 - noninformative, 14
 - time, 15, 53, 71, 80
- Cox' proportional hazards model
 - for censoring mechanism, 16, 18, 69
 - for survival mechanism, 71, 81, 86
 - IRLS estimation, 80–82
 - method, 53
 - with optimal scaling transformations, *see* Cox-OS
 - with ties, 81
 - with weighted objects, 81
- Cox-OS: Cox model with optimal scaling transformations
 - application to simulated data, 87–92
 - method, 70, 82–85
- Data types
 - Categorical data
 - nominal, 68–69
 - ordinal, 68–69
 - Numeric/continuous data, 68, 69
 - see* Scaling Levels, 1
- Datasets
 - breast cancer recurrence data, 127
 - contraceptive method choice data, 122
 - DINSTAP data, 117
 - ROM data, 22–24
 - simulated survival data with an ordinal predictor, 86–87
 - simulated survival data with dependent censoring, 26, 38–40
 - toy survival data, 18
 - UWV data, 46–48
- Dummy coding, 71–72, 117
- Factor analysis
 - application to UWV data, 47, 51
 - method, 44, 62–64
- GLM-OS: GLM with optimal scaling transformations
 - application to breast cancer recurrence data, 127–130
 - application to contraceptive method choice data, 122–127
 - application to DINSTAP data, 116–122
 - logistic regression, 115–116
 - method, 106, 109–113
- GLM: Generalized Linear Model
 - method, 104, 110–111

- Hazard
 rate, 15–16, 18, 26, 71
 ratio, 53–54, 71
 method, 45, 64–65, 135–136
 OS, *see* GLM-OS, logistic regression
- Integer coding, 72–73
- Interpretation of verbal probability phrases
 (a)symmetry, 138–141, 148, 153
 communication mode preference paradox, 140
 context dependence, 141–142, 144, 150, 153–154
 difference between sub-populations, 142–143, 150–152, 154, 158–159
 other languages, 139–140, 153
 probability scales, 138–139, 141
 variability, 138, 147–148, 153
- IPCW: Inverse Probability Censoring
 Weighting
 application to ROM data, 23–25
 application to simulated data, 26–28
 application to toy data, 18–21, 33, 36, 37
 assumption, 17–18
 implementation in R, 32–37
 method, 14, 16–21
 weights, 16–17, 20
- Kaplan-Meier Estimator, *see* Product-Limit Estimator
- Linear Regression
 method, 75, 104
 with OS, *see* Optimal scaling regression
- Logistic regression
 application to breast cancer recurrence data, 128–130
 application to contraceptive method choice data, 124–127
 application to UWV data, 46, 51, 53–54
- Optimal scaling
 Cox, *see* Cox-OS
 GLM, *see* GLM-OS
 logistic regression, *see* GLM-OS, logistic regression
 PCA, *see* OS-PCA
 quantifications, 45, 69, 73, 75–76, 105
 regression, 69–70, 74–80, 105–109
- OS-PCA: Optimal scaling principal components analysis
 application to UWV data, 52–56, 58–59
 method, 44–45, 52, 62–64, 70
- PCA: Principal Components Analysis
 method, 52
 nonlinear, *see* OS-PCA
- Product-Limit Estimator
 for censoring mechanism, 19
 method, 15, 20
- Scaling levels
 monotone spline function, 74–75, 105, 108, 112
 nominal step function, 69, 74, 78, 105, 108, 112, 117
 nonmonotone spline function, 105, 108, 112
 numeric, 105
 ordinal step function, 74–75, 78, 105, 108, 112
- Survival analysis, 45, 53, 64–65
 application to UWV data, 54, 56–59
 censoring, *see* Censoring
 competing risks, 53
 hazard rate/ratio, *see* Hazard