Prediction accuracy and stability of regression with optimal scaling transformations

Kooij, A.J. van der

Appendix A
CATREG Algorithm

CATREG

CATREG (*Cat*egorical *reg*ression with optimal scaling using alternating least squares) quantifies categorical variables using optimal scaling, resulting in an optimal linear regression equation for the transformed variables. The variables can be given mixed optimal scaling levels and no distributional assumptions about the variables are made.

Notation

The following notation is used throughout this chapter unless otherwise stated:

n	Number of analysis cases (objects)
n_w	Weighted number of analysis cases: $\displaystyle\sum_{i=1}^{n} w_i$
n_{tot}	Total number of cases (analysis + supplementary)
w_i	Weight of object i ; $w_i = 1$ if cases are unweighted; $w_i = 0$ if object i is supplementary.
$\mathbf{W}$	Diagonal $n_{tot} \times n_{tot}$ matrix, with w_i on the diagonal.
p	Number of predictor variables
m	Total number of variables
r	Index of response variable
J_p	Index set of predictor variables
$\mathbf{H}$	The data matrix (category indicators), of order $n_{tot} \times m$, after discretization, imputation of missings , and listwise deletion, if applicable.

For variable j, $j = 1, \ldots, m$

k_j	Number of categories of variable j (number of distinct values in $\mathbf{h}_j$, thus, including supplementary objects)
$\mathbf{G}_j$	Indicator matrix for variable j, of order $n_{tot} \times k_j$

The elements of $\mathbf{G}_j$ are defined as $i = 1, \ldots, n_{tot}; r = 1, \ldots, k_j$

$$g_{(j)ir} = \begin{cases} 1 & \text{when the } i\text{th object is in the } r\text{th category of variable } j \\ 0 & \text{when the } i\text{th object is not in the } r\text{th category of variable } j \end{cases}$$

$\mathbf{D}_j$	Diagonal $k_j \times k_j$ matrix, containing the weighted univariate marginals; i.e., the weighted column sums of $\mathbf{G}_j$ ($\mathbf{D}_j = \mathbf{G}'_j \mathbf{W} \mathbf{G}_j$)
$\mathbf{f}$	Degrees of freedom for the predictor variables, of order p
$\mathbf{S}_j$	I-spline basis for variable j, of order $k_j \times (s_j + t_j)$ (see Ramsay (1988) for details)
$\mathbf{a}_j$	Spline coefficient vector, of order $s_j + t_j$
d_j	Spline intercept.
s_j	Degree of polynomial
t_j	Number of interior knots

The quantification matrices and parameter vectors are:

$\mathbf{y}_r$	Category quantifications for the response variable, of order k_r
$\mathbf{y}_{j,\ j \in J_p}$	Category quantifications for predictor variable j, of order k_j
$\mathbf{b}$	Regression coefficients for the predictor variables, of order p
$\mathbf{v}$	Accumulated contributions of predictor variables: $\displaystyle\sum_{j \in J_p} b_j \mathbf{G}_j \mathbf{y}_j$

Note: The matrices $\mathbf{W}, \mathbf{G}_j$, and $\mathbf{D}_j$ are exclusively notational devices; they are stored in reduced form, and the program fully profits from their sparseness by replacing matrix multiplications with selective accumulation.

Discretization

Discretization is done on the unweighted data.

Multiplying   First, the orginal variable is standardized. Then the standardized values are multiplied by 10 and rounded, and a value is added such that the lowest value is 1.

Ranking   The original variable is ranked in ascending order, according to the alpanumerical value.

Grouping into a specified number of categories with a normal distribution

First, the original variable is standardized. Then cases are assigned to categories using intervals as defined in Max (1960).

Grouping into a specified number of categories with a unifrom distribution

First the target frequency is computed as n divided by the number of specified categories, rounded. Then the original categories are assigned to grouped categories such that the frequencies of the grouped categories are as close to the target frequency as possible.

Grouping equal intervals of specified size

First the intervals are defined as lowest value + interval size, lowest value + 2*interval size, etc. Then cases with values in the kth interval are assigned to category k.

Imputation of Missing Values

When there are variables with missing values specified to be imputed (with mode or extra category), then first the k_j's for these variables are computed before listwise deletion. Next the category indicator with the highest weighted frequency (mode; the smallest if multiple modes exist), or $k_j + 1$ (extra category) is imputed. Then listwise deletion is applied if applicable. And then the k_j's are adjusted.

If an extra category is imputed for a variable with optimal scaling level Spline Nominal, Spline Ordinal, Ordinal or Numerical, the extra category is not included in the restriction according to the scaling level in the final phase (see step (2) next section).

Objective Function Optimization

Objective Function

The CATREG objective is to find the set of $\mathbf{y}_r$, $\mathbf{b}$, and $\mathbf{y}_j$, $j \in J_p$, so that the function

$$\sigma\left(\mathbf{y}_r; \mathbf{b}; \mathbf{y}_j\right) = \left(\mathbf{G}_r \mathbf{y}_r - \sum_{j \in J_p} b_j \mathbf{G}_j \mathbf{y}_j\right)' \mathbf{W} \left(\mathbf{G}_r \mathbf{y}_r - \sum_{j \in J_p} b_j \mathbf{G}_j \mathbf{y}_j\right)$$

is minimal, under the normalization restriction $\mathbf{y}_r' \mathbf{D}_r \mathbf{y}_r = n_w$ The quantifications of the response variable are also centered; that is, they satisfy $\mathbf{u}' \mathbf{W} \mathbf{G}_r \mathbf{y}_r = \mathbf{0}$ with $\mathbf{u}$ denoting an n-vector with ones.

Optimal Scaling Levels

The following optimal scaling levels are distinguished in CATREG ($j = 1, \ldots, m$):

Nominal   Equality restrictions only.

Spline Nominal $\mathbf{y}_j = d_j + \mathbf{S}_j \mathbf{a}_j$ (equality and spline restrictions).

Spline Ordinal $\mathbf{y}_j = d_j + \mathbf{S}_j \mathbf{a}_j$ (equality and monotonic spline restrictions),

with $\mathbf{a}_j$ restricted to contain nonnegative elements (to garantee monotonic I-splines).

Ordinal $\mathbf{y}_j \in \mathbf{C}_j$ (equality and monotonicity restrictions).

The monotonicity restriction $\mathbf{y}_j \in \mathbf{C}_j$ means that $\mathbf{y}_j$ must be located in the convex cone of all k_j-vectors with nondecreasing elements.

Numerical $\mathbf{y}_j \in \mathbf{L}_j$ (equality and linearity restrictions).

The linearity restriction $\mathbf{y}_j \in L_j$ means that $\mathbf{y}_j$ must be located in the subspace of all k_j-vectors that are a linear transformation of the vector consisting of k_j successive integers.

For each variable, these levels can be chosen independently. The general requirement for all options is that equal category indicators receive equal quantifications. For identification purposes, $\mathbf{y}_j$ is always normalized so that $\mathbf{y}'_j \mathbf{D}_j \mathbf{y}_j = n_w$.

Optimization

Iteration scheme

Optimization is achieved by executing the following iteration scheme:

1. Initialization I or II
2.   Update category quantifications response variable
3.   Update category quantifications and regression coefficients predictor variables
4.   Convergence test: repeat (2)—(3) or continue

Steps (1) through (4) are explained below.

(1) Initialization

I. Random

The initial category quantifications $\tilde{\mathbf{y}}_j$ (for $j = 1, ..., m$) are defined as the k_j category indicators of variable j, normalized such that $\mathbf{u}'\mathbf{WG}_j\tilde{\mathbf{y}}_j = 0$ and $\tilde{\mathbf{y}}_j \mathbf{D}_j \tilde{\mathbf{y}}_j = n_w$, and the initial regression coefficients are the correlations with the response variable.

II. Numerical

In this case, the iteration scheme is executed twice. In the first cycle, (initialized with initialization I) all variables are treated as numerical. The second cycle, with the specified scaling levels, starts with the category quantifications and regression coefficients from the first cycle.

(2) Update category quantifications response variable

With fixed current values $\mathbf{y}_j$, $j \in J_p$, the unconstrained update of $\mathbf{y}_r$ is

$$\tilde{\mathbf{y}}_r = \mathbf{D}_r^{-1} \mathbf{G}_r' \mathbf{W} \mathbf{v}$$

<u>Nominal</u>: $\mathbf{y}_r^* = \tilde{\mathbf{y}}_r$

For the next four optimal scaling levels, if the response variable was imputed with an extra category, $\mathbf{y}_r^*$ is inclusive category k_r in the initial phase, and is exclusive category k_r in the final phase.

<u>Spline nominal and spline ordinal</u>: $\mathbf{y}_r^* = d_r + \mathbf{S}_r \mathbf{a}_r$.

The spline transformation is computed as a weighted regression (with weights the diagonal elements of $\mathbf{D}_r$) of $\tilde{\mathbf{y}}_r$ on the I-spline basis $\mathbf{S}_r$. For the spline ordinal scaling level the elements of $\mathbf{a}_j$ are restricted to be nonnegative, which makes $\mathbf{y}_r^*$ monotonically increasing

<u>Ordinal</u>: $\mathbf{y}_r^* \leftarrow WMON(\tilde{\mathbf{y}}_r)$.

The notation $WMON(\ )$ is used to denote the weighted monotonic regression process, which makes $\mathbf{y}_r^*$ monotonically increasing. The weights used are the diagonal elements of $\mathbf{D}_r$ and the subalgorithm used is the up-and-down-blocks minimum violators algorithm (Kruskal, 1964; Barlow et al., 1972).

<u>Numerical</u>: $\mathbf{y}_r^* \leftarrow WLIN(\tilde{\mathbf{y}}_r)$.

The notation $WLIN(\ )$ is used to denote the weighted linear regression process. The weights used are the diagonal elements of $\mathbf{D}_r$.

Next $\mathbf{y}_r^*$ is normalized (if the response variable was imputed with an extra category, $\mathbf{y}_r^*$ is inclusive category k_r from here on):

$$\mathbf{y}_r^+ = n_w^{1/2} \mathbf{y}_r^* (\mathbf{y}_r^{*\prime} \mathbf{D}_r \mathbf{y}_r^*)^{-1/2}$$

(3) Update category quantifications and regression weights predictor variables; loop across variables j, $j \in J_p$

For updating a predictor variable j, $j \in J_p$, first the contribution of variable j is removed from $\mathbf{v}$: $\mathbf{v}_j = \mathbf{v} - b_j \mathbf{G}_j \mathbf{y}_j$

Then the unconstrained update of $\mathbf{y}_j$ is

$$\tilde{\mathbf{y}}_j = \mathbf{D}_j^{-1}\mathbf{G}_j'\mathbf{W}\left(\mathbf{G}_r\mathbf{y}_r - \mathbf{v}_j\right)$$

Next $\tilde{\mathbf{y}}_j$ is restricted and normalized as in step (2) to obtain $\mathbf{y}_j^+$.

Finally, we update the regression coefficient

$$b_j^+ = n_{\mathrm{w}}^{-1}\tilde{\mathbf{y}}_j'\mathbf{D}_j\mathbf{y}_j^+ \ .$$

(4) Convergence test

The difference between consecutive values of the squared multiple regression coefficient,

$$R^2 = n_{\mathrm{w}}^{-1/2}\left(\mathbf{G}_r\mathbf{y}_r\right)'\mathbf{W}\mathbf{v}\left(\mathbf{v}'\mathbf{W}\mathbf{v}\right)^{-1/2}$$

is compared with the user-specified convergence criterion ε — a small positive number. Steps (2) and (3) are repeated as long as the loss difference exceeds ε.

Diagnostics

Descriptive Statistics

The descriptives tables gives the weighted univariate marginals and the weighted number of missing values (system missing, user defined missing, and values ≤ 0) for each variable.

Fit and error measures

The fit and the error for each iteration are reported in the History table.

Multiple *R* Square

R^2 as computed in step(4) in the last iteration.

Error $\left(1-R^2\right)^{1/2}$

Also, the increase in R^2 for each iteration is reported.

Summary Statistics

Multiple *R*

$$R = \left(R^2\right)^{1/2}$$

Multiple *R* Square

$$R^2$$

Adjusted Multiple *R* Square

$$1-\left(1-R^2\right)\left(n_w-1\right)\left(n_w-1-\mathbf{u'f}\right)$$

with $\mathbf{u}$ a p-vector of ones.

ANOVA Table

	Sum of Squares	df	Mean Sum of Squares
Regression	$n_w R^2$	$\mathbf{u'f}$	$n_w R^2 \left(\mathbf{u'f}\right)^{-1}$
Residual	$n_w \left(1-R^2\right)$	$n_w-1-\mathbf{u'f}$	$n_w \left(1-R^2\right)\left(n_w-1-\mathbf{u'f}\right)^{-1}$

$F = \mathrm{MS}_{reg}/\mathrm{MS}_{res}$

Correlations and Eigenvalues

Before transformation

$\mathbf{R} = n_w^{-1} \mathbf{H}'_c \mathbf{W} \mathbf{H}_c$, with $\mathbf{H}_c$ weighted centered and normalized $\mathbf{H}$ excluding the response variable.

After transformation

$\mathbf{R} = n_w^{-1} \mathbf{Q}' \mathbf{W} \mathbf{Q}$, the columns of $\mathbf{Q}$ are $\mathbf{q}_j = \mathbf{G}_j \mathbf{y}_j$ $j \in J_p$.

Statistics for Predictor Variables $j \in J_p$

Beta

The standardized regression coefficient is

Beta$_j = b_j$

Standard Error Beta

The standard error of Beta$_j$ is estimated by

SE (Beta$_j$) $\left(\left(1 - R^2 \right) \big/ \left(n_w - 1 - \mathbf{u}'\mathbf{f} \right) t_j \right)^{1/2}$

with t_j the tolerance for variable j (see below).

Degrees of Freedom

The degrees of freedom for a variable depend on the optimal scaling level:

numerical: $f_j = 1$;

spline ordinal, spline nominal: $f_j = s_j + t_j$ minus the number of elements equal to zero in $\mathbf{a}_j$.

ordinal, nominal: $f_j =$ the number of distinct values in $\mathbf{y}_j$ minus 1;

F-value

$$F_j = \left(\text{Beta}_j \big/ \text{SE}\left(\text{Beta}_j\right) \right)^2$$

Zero-order correlation

Correlations between the transformed response variable $\mathbf{G}_r \mathbf{y}_r$ and the transformed predictor variables $\mathbf{G}_j \mathbf{y}_j$:

$$r_{rj} = n_w^{-1} \left(\mathbf{G}_r \mathbf{y}_r \right)' \mathbf{W} \mathbf{G}_j \mathbf{y}_j$$

Partial correlation

$$\text{PartialCorr}_j = b_j \left(\left(1 \big/ t_j \right) \left(1 - R^2 \right) + b_j^2 \right)^{-1/2}$$

with t_j the tolerance for variable j (see below).

Part correlation

$$\text{PartCorr}_j = b_j t_j^{1/2}$$

with t_j the tolerance for variable j (see below).

Importance

Pratt's measure of relative importance (Pratt, 1987)

$$\text{Imp}_j = b_j r_{rj} \big/ R^2$$

Tolerance

The tolerance for the optimally scaled predictor variables is given by

$$t_j = r_{p_{jj}}^{-1},$$

with $r_{p_{jj}}$ the jth diagonal element of $\mathbf{R}_p$, where $\mathbf{R}_p$ is the correlation matrix of predictors that have regression coefficients > 0.

The tolerance for the original predictor variables is also reported and is computed in the same way, using the correlation matrix for the original predictor variables, discretized, imputed, and listwise deleted, if applicable.

Quantifications

The quantifications are $\mathbf{y}_j$, $j = 1, \ldots, m$.

Predicted and residual values

There is an option to save the predicted values $\mathbf{v}$ and the residual values $\mathbf{G}_r \mathbf{y}_r - \mathbf{v}$.

Supplementary objects

For supplementary objects predicted and residual values are computed.

The category indicators of supplementary objects are replaced by the quantification of the category. If a category is only used by supplementary objects, the category indicator is replaced by a system-missing value.

Residual Plots

The residual plot for predictor variable j displays two sets of points: unnormalized quantifications ($b_j \mathbf{y}_j$) against category indicators, and residuals when the response variable is predicted from all predictor variables except variable j ($\mathbf{G}_r \mathbf{y}_r - (\mathbf{v} - b_j \mathbf{G}_j \mathbf{y}_j)$) against category indicators.

References

Barlow, R. E., Bartholomew, D. J., Bremner, J. M., and Brunk, H. D. 1972. *Statistical inference under order restrictions*. New York: John Wiley & Sons, Inc.

Gifi, A. 1990. *Nonlinear multivariate analysis*. Leiden: Department of Data Theory.

Kruskal, J. B. 1964. Nonmetric multidimensional scaling: a numerical method. *Psychometrika*, 29: 115–129.

Max, J. (1960), Quantizing for minimum distortion. IRE Transactions on Information Theory, 6, 7-12.

Pratt, J.W. (1987). *Dividing the indivisible: using simle symmetry to partition variance explained.* In T. Pukkila and S. Puntanen (Eds.), Proceedings of the Second International Conference in Statistics (245-260). Tampere, Finland: University of Tampere.

Ramsay, J.O. (1988) Monotone regression Splines in action, Statistical Science, 4, 425-441.

Appendix B
CATREG sections from
SPSS Categories® 11.0

Categorical Regression (CATREG)

Categorical regression quantifies categorical data by assigning numerical values to the categories, resulting in an optimal linear regression equation for the transformed variables. Categorical regression is also known by the acronym CATREG, for *cat*egorical *reg*ression.

Standard linear regression analysis involves minimizing the sum of squared differences between a response (dependent) variable and a weighted combination of predictor (independent) variables. Variables are typically quantitative, with (nominal) categorical data recoded to binary or contrast variables. As a result, categorical variables serve to separate groups of cases, and the technique estimates separate sets of parameters for each group. The estimated coefficients reflect how changes in the predictors affect the response. Prediction of the response is possible for any combination of predictor values.

An alternative approach involves regressing the response on the categorical predictor values themselves. Consequently, one coefficient is estimated for each variable. However, for categorical variables, the category values are arbitrary. Coding the categories in different ways yield different coefficients, making comparisons across analyses of the same variables difficult.

CATREG extends the standard approach by simultaneously scaling nominal, ordinal, and numerical variables. The procedure quantifies categorical variables such that the quantifications reflect characteristics of the original categories. The procedure treats quantified categorical variables in the same way as numerical variables. Using nonlinear transformations allow variables to be analyzed at a variety of levels to find the best-fitting model.

Example. Categorical regression could be used to describe how job satisfaction depends on job category, geographic region, and amount of travel. You might find that high levels of satisfaction correspond to managers and low travel. The resulting regression equation could be used to predict job satisfaction for any combination of the three independent variables.

Statistics and plots. Frequencies, regression coefficients, ANOVA table, iteration history, category quantifications, correlations between untransformed predictors, correlations between transformed predictors, residual plots, and transformation plots.

Data. CATREG operates on category indicator variables. The category indicators should be positive integers. You can use the Discretization dialog box to convert fractional-value variables and string variables into positive integers.

Assumptions. Only one response variable is allowed, but the maximum number of predictor variables is 200. The data must contain at least three valid cases, and the number of valid cases must exceed the number of predictor variables plus one.

Related procedures. CATREG is equivalent to categorical canonical correlation analysis with optimal scaling (OVERALS) with two sets, one of which contains only one variable. Scaling all variables at the numerical level corresponds to standard multiple regression analysis.

To Obtain a Categorical Regression

▶ From the menus choose:

Analyze
  Regression
    Optimal Scaling…

Figure 2.1   Categorical Regression dialog box
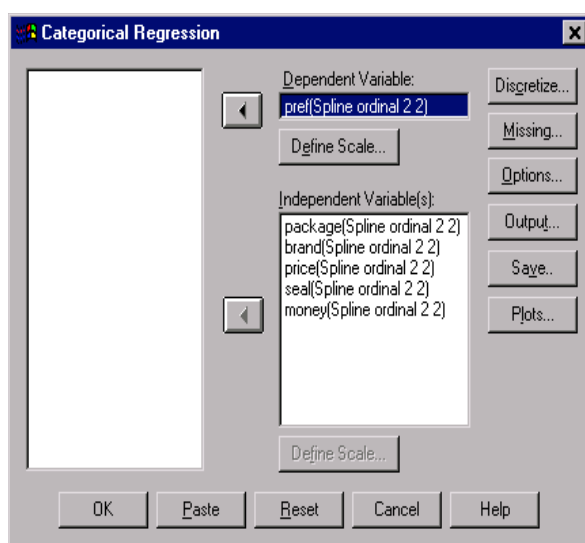
▶ Select the dependent variable and independent variable(s).

▶ Click OK.

Optionally, change the scaling level for each variable.

Define Scale in Categorical Regression

You can set the optimal scaling level for the dependent and independent variables. By default, they are scaled as second-degree monotonic splines (ordinal) with two interior knots. Additionally, you can set the weight for analysis variables.
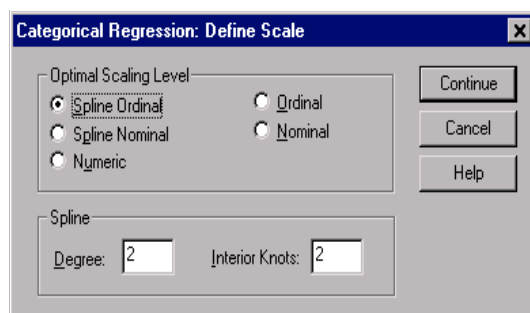
Optimal Scaling Level. You can also select the scaling level for quantifying each variable.

- **Spline Ordinal.** The order of the categories of the observed variable is preserved in the optimally scaled variable. Category points will be on a straight line (vector) through the origin. The resulting transformation is a smooth monotonic piecewise polynomial of the chosen degree. The pieces are specified by the user-specified number and procedure-determined placement of the interior knots.

- **Spline Nominal.** The only information in the observed variable that is preserved in the optimally scaled variable is the grouping of objects in categories. The order of the categories of the observed variable is not preserved. Category points will be on a straight line (vector) through the origin. The resulting transformation is a smooth, possibly nonmonotonic, piecewise polynomial of the chosen degree. The pieces are specified by the user-specified number and procedure-determined placement of the interior knots.

- **Ordinal.** The order of the categories of the observed variable is preserved in the optimally scaled variable. Category points will be on a straight line (vector) through the origin. The resulting transformation fits better than the spline ordinal transformation but is less smooth.

- **Nominal.** The only information in the observed variable that is preserved in the optimally scaled variable is the grouping of objects in categories. The order of the categories of the observed variable is not preserved. Category points will be on a straight line (vector) through the origin. The resulting transformation fits better than the spline nominal transformation but is less smooth.

- **Numeric.** Categories are treated as ordered and equally spaced (interval level). The order of the categories and the equal distances between category numbers of the observed variable are preserved in the optimally scaled variable. Category points will be on a straight line (vector) through the origin. When all variables are at the numeric level, the analysis is analogous to standard principal components analysis.

To Define the Scale in CATREG

▶ Select one or more variables on the variables list in the Categorical Regression dialog box.

▶ Click Define Scale.

**Figure 2.2   Categorical Regression Define Scale dialog box**

▶ Select the optimal scaling level to be used in the analysis.

▶ Click Continue.

Categorical Regression Discretization

The Discretization dialog box allows you to select a method of recoding your variables. Fractional-value variables are grouped into seven categories (or into the number of distinct values of the variable, if this number is less than seven) with an approximately normal distribution, unless specified otherwise. String variables are always converted into positive integers by assigning category indicators according to ascending alphanumeric order. Discretization for string variables applies to these integers. Other variables are left alone by default. The discretized variables are then used in the analysis.

Figure 2.3   Categorical Regression Discretization dialog box
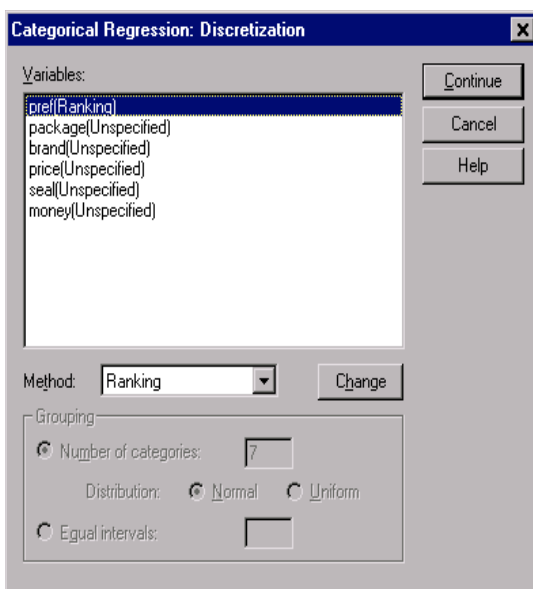
Method. Choose between grouping, ranking, or multiplying.

- **Grouping.** Recode into a specified number of categories or recode by interval.
- **Ranking.** The variable is discretized by ranking the cases.
- **Multiplying.** The current values of the variable are standardized, multiplied by 10, rounded, and have a constant added such that the lowest discretized value is 1.

Grouping. The following options are available when discretizing variables by grouping:

- **Number of categories.** Specify a number of categories and whether the values of the variable should follow an approximately normal or uniform distribution across those categories.
- **Equal intervals.** Variables are recoded into categories defined by these equally sized intervals. You must specify the length of the intervals.

Categorical Regression Missing Values

The Missing Values dialog box allows you to choose the strategy for handling missing values in analysis variables and supplementary variables.

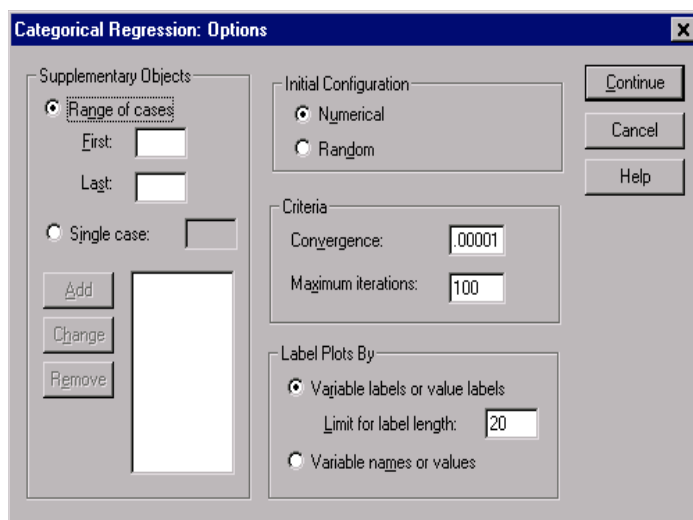**Figure 2.4   Categorical Regression Missing Values dialog box**
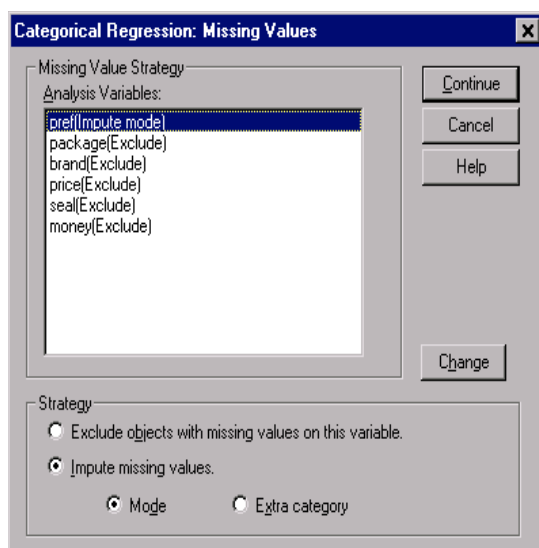
Strategy. Choose to impute missing values (active treatment) or exclude objects with missing values (listwise deletion).

- **Impute missing values.** Objects with missing values on the selected variable have those values imputed. You can choose the method of imputation. Select Mode to re-place missing values with the most frequent category. When there are multiple modes, the one with the smallest category indicator is used. Select Extra category to replace missing values with the same quantification of an extra category. This im-plies that objects with a missing value on this variable are considered to belong to the same (extra) category.

- **Exclude objects with missing values on this variable.** Objects with missing values on the selected variable are excluded from the analysis. This strategy is not available for supplementary variables.

Categorical Regression Options

The Options dialog box allows you to select the initial configuration style, specify iteration and convergence criteria, select supplementary objects, and set the labeling of plots.

Figure 2.5   Categorical Regression Options dialog box

Supplementary Objects. This allows you to specify the objects that you want to treat as supplementary. Simply type the number of a supplementary object and click Add. You cannot weight supplementary objects (specified weights are ignored).

Initial Configuration. If no variables are treated as nominal, select the Numerical configuration. If at least one variable is treated as nominal, select the Random configuration.
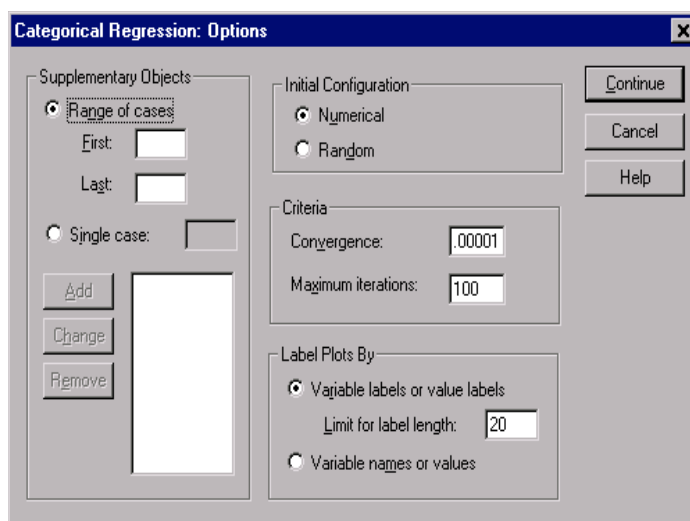
Criteria. You can specify the maximum number of iterations the regression may go through in its computations. You can also select a convergence criterion value. The regression stops iterating if the difference in total fit between the last two iterations is less than the convergence value or if the maximum number of iterations is reached.

Label Plots By. Allows you to specify whether variables and value labels or variable names and values will be used in the plots. You can also specify a maximum length for labels.

Categorical Regression Options

The Options dialog box allows you to select the initial configuration style, specify iteration and convergence criteria, select supplementary objects, and set the labeling of plots.

Figure 2.5   Categorical Regression Options dialog box

Supplementary Objects. This allows you to specify the objects that you want to treat as supplementary. Simply type the number of a supplementary object and click Add. You cannot weight supplementary objects (specified weights are ignored).

Initial Configuration. If no variables are treated as nominal, select the Numerical configuration. If at least one variable is treated as nominal, select the Random configuration.

Criteria. You can specify the maximum number of iterations the regression may go through in its computations. You can also select a convergence criterion value. The regression stops iterating if the difference in total fit between the last two iterations is less than the convergence value or if the maximum number of iterations is reached.

Label Plots By. Allows you to specify whether variables and value labels or variable names and values will be used in the plots. You can also specify a maximum length for labels.

- **ANOVA.** This option includes regression and residual sums of squares, mean squares, and *F*. Two ANOVA tables are displayed: one with degrees of freedom for the regression equal to the number of predictor variables and one with degrees of freedom for the regression taking the optimal scaling into account.

Category Quantifications. Tables showing the transformed values of the selected variables are displayed.

Descriptive Statistics. Tables showing the frequencies, missing values, and modes of the selected variables are displayed.

Categorical Regression Save

The Save dialog box allows you to save results to the working file or an external file.

**Figure 2.7   Categorical Regression Save dialog box**
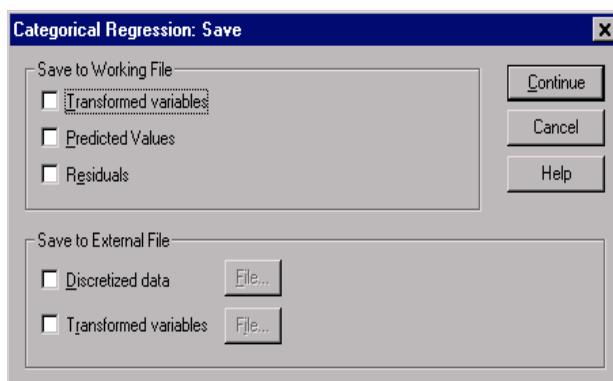
Save to Working File. You can save the transformed values of the variables, model-predicted values, and residuals to the working file.

Save to External File. You can save the discretized data and transformed variables to external files.

Categorical Regression Plots

The Plot dialog box allows you to specify the variables that will produce transformation and residual plots.

Figure 2.8   Categorical Regression Plot dialog box
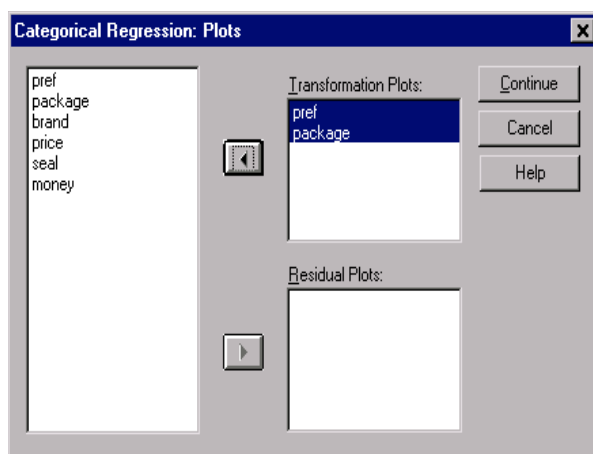
Transformation Plots. For each of these variables, the category quantifications are plotted against the original category values. Empty categories appear on the horizontal axis but do not affect the computations. These categories are identified by breaks in the line connecting the quantifications.

Residual Plots. For each of these variables, residuals (computed for the dependent variable predicted from all predictor variables except the predictor variable in question) are plotted against category indicators and the optimal category quantifications multiplied with beta against category indicators.

CATREG Command Additional Features

You can customize your categorical regression if you paste your selections into a syntax window and edit the resulting CATREG command syntax. SPSS command language also allows you to:

• Specify rootnames for the transformed variables when saving them to the working data file (with the SAVE subcommand).

Categorical Regression Examples

The goal of categorical regression with optimal scaling is to describe the relationship between a response and a set of predictors. By quantifying this relationship, values of the response can be predicted for any combination of predictors.

In this chapter, two examples serve to illustrate the analyses involved in optimal scaling regression. The first example uses a small data set to illustrate the basic concepts. The second example uses a much larger set of variables and observations in a practical example.

Example 1: Carpet Cleaner Data

In a popular example by Green and Wind (1973), a company interested in marketing a new carpet cleaner wants to examine the influence of five factors on consumer preference—package design, brand name, price, a *Good Housekeeping* seal, and a money-back guarantee. There are three factor levels for package design, each one differing in the location of the applicator brush; three brand names (*K2R*, *Glory*, and *Bissell*); three price levels; and two levels (either no or yes) for each of the last two factors. Table 8.1 displays the variables used in the carpet-cleaner study, with their variable labels and values.

Table 8.1   Explanatory variables in the carpet-cleaner study

	Variable label	**Value labels**
package	Package design	A*, B*, C*
brand	Brand name	K2R, Glory, Bissell
price	Price	$1.19, $1.39, $1.59
seal	*Good Housekeeping* seal	No, yes
money	Money-back guarantee	No, yes

Ten consumers rank 22 profiles defined by these factors. The variable *pref* contains the rank of the average rankings for each profile. Low rankings correspond to high preference. This variable reflects an overall measure of preference for each profile. Using categorical regression, you will explore how the five factors in Table 8.1 are related to preference. This data set can be found in *carpet.sav.*

Standard Linear Regression Analysis

To produce standard linear regression output, from the menus choose:

Analyze
  Regression
    Linear...

▸ Dependent: *pref*
▸ Independent(s): *package*, *brand*, *price*, *seal*, *money*

Statistics...
  ☐ Descriptives (deselect)

Save...
  Residuals
  ☑ Standardized

The standard approach for describing the relationships in this problem is linear regression. The most common measure of how well a regression model fits the data is R^2. This statistic represents how much of the variance in the response is explained by the weighted combination of predictors. The closer R^2 is to 1, the better the model fits. Regressing *pref* on the five predictors results in an R^2 of 0.707, indicating that approximately 71% of the variance in the preference rankings is explained by the predictor variables in the linear regression.

Figure 8.1   Model summary for standard linear regression

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.841	.707	.615	3.9981

The standardized coefficients are shown in Figure 8.2. The sign of the coefficient indicates whether the predicted response increases or decreases when the predictor increases, all other predictors being constant. For categorical data, the category coding determines the meaning of an increase in a predictor. For instance, an increase in *money*, *package*, or *seal* will result in a decrease in predicted preference ranking. *money* is coded 1 for *no money-back guarantee* and 2 for *money-back guarantee*. An increase in *money* corresponds to the addition of a money-back guarantee. Thus, adding a money-back guarantee reduces the predicted preference ranking, which corresponds to an increased predicted preference.

Figure 8.2   Regression coefficients

Model		Standardized Coefficients	t	Sig.
		Beta		
1	(Constant)		4.352	.000
	Package design	-.560	-4.015	.001
	Brand name	.056	.407	.689
	Price	.366	2.681	.016
	Good Housekeeping seal	-.330	-2.423	.028
	Money-back guarantee	-.197	-1.447	.167

The value of the coefficient reflects the amount of change in the predicted preference ranking. Using standardized coefficients, interpretations are based on the standard deviations of the variables. Each coefficient indicates the number of standard deviations that the predicted response changes for a one standard deviation change in a predictor, all other predictors remaining constant. For example, a one standard deviation change in *brand* yields an increase in predicted preference of 0.056 standard deviations. The standard deviation of *pref* is 6.44, so *pref* increases by $0.056 \times 6.44 = 0.361$. Changes in *package* yield the greatest changes in predicted preference.

   A regression analysis should always include an examination of the residuals. To produce residual plots, from the menus choose:

Graphs
  Scatter...

Select *Simple*. Click *Define*.

▶ Y Axis: *zre_1*
▶ X Axis: *zpr_1*

Then, recall the Simple Scatterplot dialog box and click *Reset* to clear the previous selections.

▶ Y Axis: *zre_1*
▶ X Axis: *package*

The standardized residuals are plotted against the standardized predicted values in Figure 8.3. No patterns should be present if the model fits well. Here you see a U-shape in which both low and high standardized predicted values have positive residuals. Standardized predicted values near 0 tend to have negative residuals.

Figure 8.3   Residuals versus predicted values

Standardized Residual (y-axis) vs Standardized Predicted Value (x-axis)
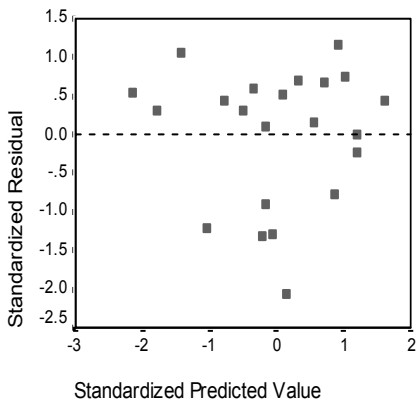
This shape is more pronounced in the plot of the standardized residuals against *package* in Figure 8.4. Every residual for Design B* is negative, whereas all but one of the residuals is positive for the other two designs. Because the regression model fits one parameter for each variable, the relationship cannot be captured by the standard approach.

Figure 8.4   Residuals versus package

Standardized Residual (y-axis) vs Package Design (x-axis)
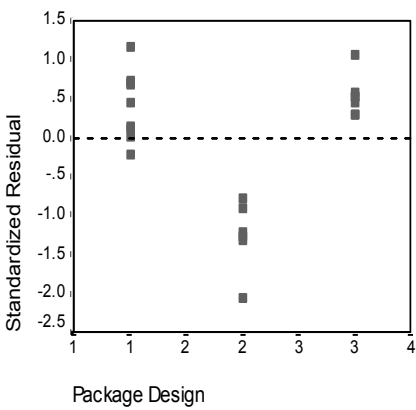
Categorical Regression Analysis

The categorical nature of the variables and the nonlinear relationship between *pref* and *package* suggest that regression on optimal scores may perform better than standard regression. The U-shape of Figure 8.4 indicates that a nominal treatment of *package* should be used. All other predictors will be treated at the numerical scaling level.

The response variable warrants special consideration. You want to predict the values of *pref*. Thus, recovering as many properties of its categories as possible in the quantifications is desirable. Using an ordinal or nominal scaling level ignores the differences between the response categories. However, linearly transforming the response categories preserves category differences. Consequently, scaling the response numerically is generally preferred and will be employed here.

To produce the following categorical regression output, from the menus choose:

Analyze
  Regression
    Optimal Scaling...

▶ Dependent: *pref*
▶ Independent(s): *package*, *brand*, *price*, *seal*, *money*

Select *pref*. Click *Define Scale*.
    Optimal Scaling Level
    ⊙ Numeric

Select *package*. Click *Define Scale*.
    Optimal Scaling Level
    ⊙ Nominal

Select *brand*, *price*, *seal*, and *money*. Click *Define Scale*.
    Optimal Scaling Level
    ⊙ Numeric

Output...
    Display
    ☑ Correlations of original predictors
    ☑ Correlations of transformed predictors
    ☐ Frequencies (deselect)
    ☐ ANOVA table (deselect)

Save...
    Save to Working File
    ☑ Transformed variables
    ☑ Residuals

Plots...
    ▶ Transformation Plots: *package*, *price*

Intercorrelations

The intercorrelations among the predictors are useful for identifying multicollinearity in the regression. Variables that are highly correlated will lead to unstable regression estimates. However, due to their high correlation, omitting one of them from the model only minimally affects prediction. The variance in the response that can be explained by the omitted variable is still explained by the remaining correlated variable. However, zero-order correlations are sensitive to outliers and also cannot identify multicollinearity due to a high correlation between a predictor and a combination of other predictors.

Figure 8.5 and Figure 8.6 show the intercorrelations of the predictors for both the untransformed and transformed predictors. All values are near 0, indicating that multicollinearity between individual variables is not a concern.

Notice that the only correlations that change involve package. Because all other predictors are treated numerically, the differences between the categories and the order of the categories are preserved for these variables. Consequently, the correlations cannot change.

Figure 8.5   Original predictor correlations

	Package design	Brand name	Price	Good Housekeeping seal	Money-back guarantee
Package design	1.000	-.189	-.126	.081	.066
Brand name	-.189	1.000	.065	-.042	-.034
Price	-.126	.065	1.000	.000	.000
Good Housekeeping seal	.081	-.042	.000	1.000	-.039
Money-back guarantee	.066	-.034	.000	-.039	1.000

Figure 8.6   Transformed predictor correlations

	Package design	Brand name	Price	Good Housekeeping seal	Money-back guarantee
Package design	1.000	-.156	-.089	.032	.102
Brand name	-.156	1.000	.065	-.042	-.034
Price	-.089	.065	1.000	.000	.000
Good Housekeeping seal	.032	-.042	.000	1.000	-.039
Money-back guarantee	.102	-.034	.000	-.039	1.000

Model Fit and Coefficients

The Categorical Regression procedure yields an R^2 of 0.948, indicating that almost 95% of the variance in the transformed preference rankings is explained by the regression on the optimally transformed predictors. Transforming the predictors improves the fit over the standard approach.

Figure 8.7   Model summary for categorical regression

Multiple R	R Square	Adjusted R Square
.974	.948	.932

Figure 8.8 shows the standardized regression coefficients. Categorical regression standardizes the variables, so only standardized coefficients are reported. These values are divided by their corresponding standard errors, yielding an F test for each variable. However, the test for each variable is contingent upon the other predictors being in the model. In other words, the test determines if omission of a predictor variable from the model with all other predictors present significantly worsens the predictive capabilities of the model. These values should not be used to omit several variables at one time for a subsequent model. Moreover, alternating least squares optimizes the quantifications, implying that these tests must be interpreted conservatively.

Figure 8.8   Standardized coefficients for transformed predictors

	Standardized Coefficients		
	Beta	Std. Error	F
Package design	-.748	.058	165.495
Brand name	4.530E-02	.058	.614
Price	.371	.057	41.986
Good Housekeeping seal	-.350	.057	37.702
Money-back guarantee	-.159	.057	7.669

The largest coefficient occurs for *package*. A one standard deviation increase in *package* yields a 0.748 standard deviation decrease in predicted preference ranking. However, *package* is treated nominally, so an increase in the quantifications need not correspond to an increase in the original category codes.

Standardized coefficients are often interpreted as reflecting the importance of each predictor. However, regression coefficients cannot fully describe the impact of a predictor or the relationships between the predictors. Alternative statistics must be used in conjunction with the standardized coefficients to fully explore predictor effects.

Correlational Analyses

To interpret the contributions of the predictors to the regression, it is not sufficient to only inspect the regression coefficients. In addition, the correlations, partial correlations, and part correlations should be inspected. Figure 8.9 contains these correlational measures for each variable.

The zero-order correlation is the correlation between the transformed predictor and the transformed response. For this data, the largest correlation occurs for *package*. However, if you can explain some of the variation in either the predictor or the response, you will get a better representation of how well the predictor is doing.

Figure 8.9   Zero-order, part, and partial correlations (transformed variables)

	Correlations		
	Zero-Order	Partial	Part
Package design	-.816	-.955	-.733
Brand name	.206	.192	.045
Price	.441	.851	.369
Good Housekeeping seal	-.370	-.838	-.350
Money-back guarantee	-.223	-.569	-.158

Other variables in the model can confound the performance of a given predictor in predicting the response. The partial correlation coefficient removes the linear effects of other predictors from both the predictor and the response. This measure equals the correlation between the residuals from regressing the predictor on the other predictors and the residuals from regressing the response on the other predictors. The squared partial correlation corresponds to the proportion of the variance explained relative to the residual variance of the response remaining after removing the effects of the other variables. For example, in Figure 8.9, *package* has a partial correlation of –0.955. Removing the effects of the other variables, package explains $(-0.955)^2 = 0.91 = 91\%$ of the variation in the preference rankings. Both *price* and *seal* also explain a large portion of variance if the effects of the other variables are removed.

Figure 8.10 displays the partial correlations for the untransformed variables. All of the partial correlations increase when optimal scores are used. In the standard approach, *package* explained 50% of the variation in *pref* when other variable effects were removed from both. In contrast, *package* explains 91% of the variation if optimal scaling is used. Similar results occur for *price* and *seal*.

Figure 8.10 Zero-order, part, and partial correlations (untransformed variables)

Model		Correlations		
		Zero-order	Partial	Part
1	(Constant)			
	Package design	-.657	-.708	-.544
	Brand name	.206	.101	.055
	Price	.440	.557	.363
	Good Housekeeping seal	-.370	-.518	-.328
	Money-back guarantee	-.223	-.340	-.196

As an alternative to removing the effects of variables from both the response and a predictor, you can remove the effects from just the predictor. The correlation between the response and the residuals from regressing a predictor on the other predictors is the part correlation. Squaring this value yields a measure of the proportion of variance explained relative to the total variance of response. From Figure 8.9, if you remove the effects of *brand*, *seal*, *money*, and *price* from *package*, the remaining part of *package* explains $(-0.733)^2 = 0.54 = 54\%$ of the variation in preference rankings.

Importance

In addition to the regression coefficients and the correlations, Pratt's measure of relative importance (Pratt, 1987) aids in interpreting predictor contributions to the regression. Large individual importances relative to the other importances correspond to predictors that are crucial to the regression. Also, the presence of suppressor variables is signaled by a low importance for a variable that has a coefficient of similar size to the important predictors.

Figure 8.11 displays the importances for the carpet cleaner predictors. In contrast to the regression coefficients, this measure defines the importance of the predictors additively—that is, the importance of a set of predictors is the sum of the individual importances of the predictors. Pratt's measure equals the product of the regression coefficient and the zero-order correlation for a predictor. These products add to R^2, so they are divided by R^2, yielding a sum of one. The set of predictors *package* and *brand*, for example, have an importance of 0.654. The largest importance corresponds to *package*, with *package*, *price*, and *seal* accounting for 95% of the importance for this combination of predictors.

Multicollinearity

Large correlations between predictors will dramatically reduce a regression model's stability. Correlated predictors result in unstable parameter estimates. Tolerance reflects how much the independent variables are linearly related to one another. This measure is the proportion of a variable's variance *not* accounted for by other independent variables in the equation. If the other predictors can explain a large amount of a predictor's variance, that predictor is not needed in the model. A tolerance value near 1 indicates that the variable cannot be predicted very well from the other predictors. In contrast, a variable with a very low tolerance contributes little information to a model, and can cause computational problems. Moreover, large negative values of Pratt's importance measure indicate multicollinearity.

Figure 8.11 shows the tolerance for each predictor. All of these measures are very high. None of the predictors are predicted very well by the other predictors and multicollinearity is not present.

Figure 8.11 Predictor tolerances and importances

		Tolerance	
	Importance	After Transformation	Before Transformation
Package design	.644	.959	.942
Brand name	.010	.971	.961
Price	.172	.989	.982
Good Housekeeping seal	.137	.996	.991
Money-back guarantee	.037	.987	.993

Transformation Plots

Plotting the original category values against their corresponding quantifications can reveal trends that might not be noticed in a list of the quantifications. Such plots are commonly referred to as transformation plots. Attention should be given to categories that receive similar quantifications. These categories affect the predicted response in the same manner. However, the transformation type dictates the basic appearance of the plot.

Variables treated as numerical result in a linear relationship between the quantifications and the original categories, corresponding to a straight line in the transformation plot. The order and the difference between the original categories is preserved in the quantifications.

The order of the quantifications for variables treated as ordinal correspond to the order of the original categories. However, the differences between the categories are not preserved. As a result, the transformation plot is nondecreasing but need not be a straight line. If consecutive categories correspond to similar quantifications, the category distinction may be unnecessary and the categories could be combined. Such categories result in a plateau on the transformation plot. However, this pattern can also result from imposing an ordinal structure on a variable that should be treated as nominal. If a subsequent nominal treatment of the variable reveals the same pattern, combining categories is warranted. Moreover, if the quantifications for a variable treated as ordinal fall along a straight line, a numerical transformation may be more appropriate.

For variables treated as nominal, the order of the categories along the horizontal axis corresponds to the order of the codes used to represent the categories. Interpretations of category order or of the distance between the categories is unfounded. The plot can assume any nonlinear or linear form. If an increasing trend is present, an ordinal treatment should be attempted. If the nominal transformation plot displays a linear trend, a numerical transformation may be more appropriate.

Figure 8.12 displays the transformation plot for *price*, which was treated as numerical. Notice that the order of the categories along the straight line correspond to the order of the original categories. Also, the difference between the quantifications for *$1.19* and *$1.39* (–1.173 and 0) is the same as the difference between the quantifications for *$1.39* and *$1.59* (0 and 1.173). The fact that categories 1 and 3 are the same distance from category 2 is preserved in the quantifications.

Figure 8.12   Transformation plot for price (numerical)

The nominal transformation of *package* yields the transformation plot in Figure 8.13. Notice the distinct nonlinear shape in which the second category has the largest quantification. In terms of the regression, the second category decreases predicted preference ranking, whereas the first and third categories have the opposite effect.

Figure 8.13   Transformation plot for package (nominal)

Package Design
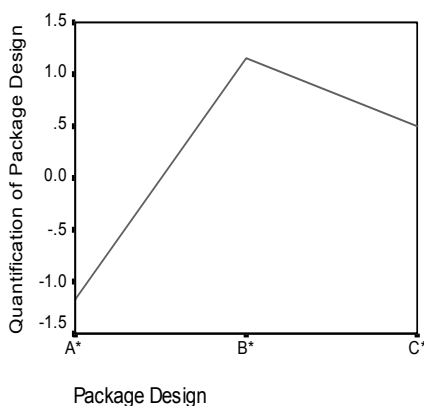
Residual Analysis

Using the transformed data and residuals that you saved to the working file allows you to create a scatterplot like the one in Figure 8.4.

To obtain such a scatterplot, recall the Simple Scatterplot dialog box and click *Reset* to clear your previous selections and restore the default options.

▶ Y Axis: *res_1*
▶ X Axis: *tra2_1*

Figure 8.14 shows the standardized residuals plotted against the optimal scores for *package*. All of the residuals are within two standard deviations of 0. A random scatter of points replaces the U-shape present in Figure 8.4. Predictive abilities are improved by optimally quantifying the categories.

Figure 8.14   Residuals for categorical regression

Quantification of Package Design
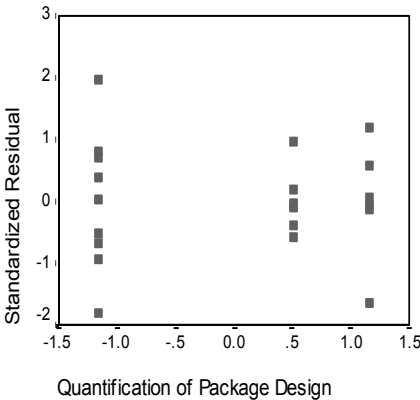
Example 2: Ozone Data

In this example, you will use a larger set of data to illustrate the selection and effects of optimal scaling transformations. The data include 330 observations on six meteorological variables analyzed by Breiman and Friedman (1985), and Hastie and Tibshirani (1990), among others. Table 8.2 describes the original variables. Your categorical regression attempts to predict the ozone concentration from the remaining variables. Previous researchers found nonlinearities among these variables, which hinder standard regression approaches.

Table 8.2   Original variables

Variable	Description
ozon	daily ozone level; categorized into one of 38 categories
ibh	inversion base height
dpg	pressure gradient (mm Hg)
vis	visibility (miles)
temp	temperature (degrees F)
doy	day of the year

This data set can be found in *ozone.sav*.

Categorizing Variables

In many analyses, variables need to be categorized or recoded before a categorical regression can be performed. For example, the Categorical Regression procedure truncates any decimals and treats negative values as missing. If either of these applications is undesirable, the data must be recoded before performing the regression. Moreover, if a variable has more categories than is practically interpretable, you should modify the categories before the analysis to reduce the category range to a more manageable number.

The variable *doy* has a minimum value of 3 and a maximum value of 365. Using this variable in a categorical regression corresponds to using a variable with 365 categories. Similarly, *vis* ranges from 0 to 350. To simplify analyses, divide each variable by 10, add 1, and round the result to the nearest integer. The resulting variables, denoted *ddoy* and *dvis*, have only 38 and 36 categories respectively, and are consequently much easier to interpret.

The variable *ibh* ranges from 111 to 5000. A variable with this many categories results in very complex relationships. However, dividing by 100 and rounding the result to the nearest integer yields categories ranging from 1 to 50 for the variable *dibh*. Using a 50-category variable rather than a 5000-category variable simplifies interpretations significantly.

Categorizing *dpg* differs slightly from categorizing the previous three variables. This variable ranges from –69 to 107. The procedure omits any categories coded with negative numbers from the analysis. To adjust for the negative values, add 70 to all observations to yield a range from 1 to 177. Dividing this range by 10 and adding 1 results in *ddpg*, a variable with categories ranging from 1 to 19.

The temperatures for *temp* range from 25 to 93 on the Fahrenheit scale. Converting to Celsius and rounding yields a range from –4 to 34. Adding 5 eliminates all negative numbers and results in *tempc*, a variable with 39 categories.

To compute the new variables as suggested, from the menus choose:

Transform
  Compute...

Target Variable: *ddoy*
Numeric Expression: RND(doy/10 +1)

Recall the Compute Variable dialog box. Click *Reset* to clear your previous selections.

Target Variable: *dvis*
Numeric Expression: RND(vis/10 +1)

Recall the Compute Variable dialog box. Click *Reset* to clear your previous selections.

Target Variable: *dibh*
Numeric Expression: RND(ibh/100)

Recall the Compute Variable dialog box. Click *Reset* to clear your previous selections.

Target Variable: *ddpg*
Numeric Expression: RND((dpg+70)/10 +1)

Recall the Compute Variable dialog box. Click *Reset* to clear your previous selections.

Target Variable: *tempc*
Numeric Expression: RND((temp-32)/1.8) +5

As described above, different modifications for variables may be required before conducting a categorical regression. The divisors used here are purely subjective. If you desire fewer categories, divide by a larger number. For example, *doy* could have been divided into months of the year or seasons.

Selection of Transformation Type

Each variable can be analyzed at one of three different levels. However, because prediction of the response is the goal, you should scale the response "as is" by employing the numerical optimal scaling level. Consequently, the order and the differences between categories will be preserved in the transformed variable.

To obtain a categorical regression in which the dependent variable is scaled at the numerical level and the independent variables are scaled at the nominal level, from the menus choose:

Analyze
  Regression
    Optimal Scaling...

Dependent: *ozon*
Independent(s): *ddpg*, *ddoy*, *dibh*, *dvis*, *tempc*

Select *ozon*. Click *Define Scale.*
    Optimal Scaling Level
    ⊙ Numerical

Select *ddpg*, *ddoy*, *dibh*, *dvis*, and *tempc*. Click *Define Scale.*
    Optimal Scaling Level
    ⊙ Nominal

Output...
    Display
    ☐ ANOVA table (deselect)

Plots...
    ▶ Transformation Plots: *ddpg*, *ddoy*, *dibh*, *dvis*, *tempc*

Treating all predictors as nominal yields an R^2 of 0.883. This large amount of variance accounted for is not surprising because nominal treatment imposes no restrictions on the quantifications. However, interpreting the results can be quite difficult.

Figure 8.15   Model summary

Multiple R	R Square	Adjusted R Square
.940	.883	.881

Figure 8.16 shows the standardized regression coefficients of the predictors. A common mistake made when interpreting these values involves focusing on the coefficients while neglecting the quantifications. You cannot assert that the large positive value of the *tempc* coefficient implies that as *tempc* increases, predicted *ozon* increases. Similarly, the negative coefficient for *dibh* does not suggest that as *dibh* increases, predicted *ozon* decreases. All interpretations must be relative to the *transformed* variables. As the quantifications for *tempc* increase, or as the quantifications for *dibh* decrease, predicted *ozon* increases. To examine the effects of the original variables, you must relate the categories to the quantifications.

Figure 8.16   Regression coefficients (all predictors nominal)

	Standardized Coefficients		F	Importance
	Beta	Std. Error		
DDOY	-.340	.020	279.077	.110
DVIS	-.199	.019	104.087	.073
DIBH	-.264	.020	175.887	.144
DDPG	.249	.020	152.275	.041
TEMPC	.681	.020	1124.375	.631

Figure 8.17 displays the transformation plot for *ddpg*. The initial categories (1 through 7) receive small quantifications and thus have minimal contributions to the predicted response. Categories 8 through 10 receive somewhat higher, positive values, resulting in a moderate increase in predicted *ozon*. The quantifications decrease up to category 17, where *ddpg* has its greatest decreasing effect on predicted *ozon*. Although the line increases after this category, using an ordinal scaling level for *ddpg* may not significantly reduce the fit, while simplifying the interpretations of the effects. However, the importance measure of 0.04 and the regression coefficient for *ddpg* indicates that this variable is not very useful in the regression.

Figure 8.17   Transformation plot for ddpg (nominal)
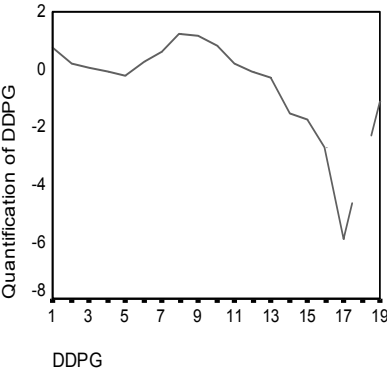
The transformation plots for *dvis* and *dibh* (Figure 8.18 and Figure 8.19) show no apparent pattern. As evidenced by the jagged nature of the plots, moving from low categories to high categories yields fluctuations in the quantifications in both directions. Thus, describing the effects of these variables requires focusing on the individual categories. Imposing ordinal or linear restrictions on the quantifications for either of these variables might significantly reduce the fit.

Figure 8.18 Transformation plot for dvis (nominal)
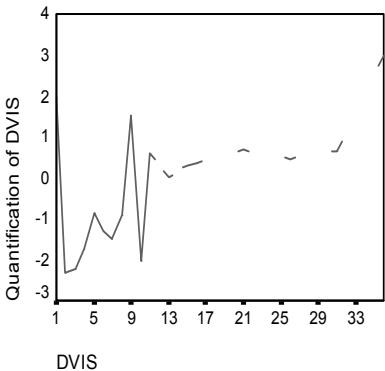
Figure 8.19 Transformation plot for dibh (nominal)
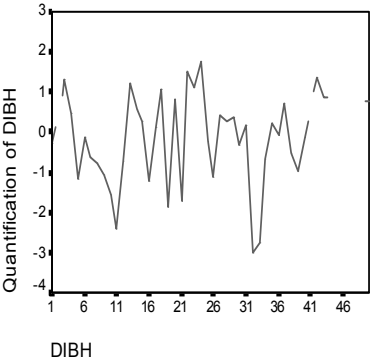
Figure 8.20 shows the transformation plot for *ddoy*. In contrast to Figure 8.18, this plot displays a pattern. The quantifications tend to decrease up to category 21, at which point they tend to increase, yielding a U-shape. Considering the sign of the regression coefficient for *ddoy*, the initial categories (1 through 5) receive quantifications that have a decreasing effect on predicted *ozon*. From category 6 onward, the effect of the quantifications on predicted *ozon* gets more increasing, reaching a maximum around category 21. Beyond category 21, the quantifications tend to decrease the predicted *ozon*. Although the line is quite jagged, the general shape is still identifiable.

Figure 8.20   Transformation plot for ddoy (nominal)
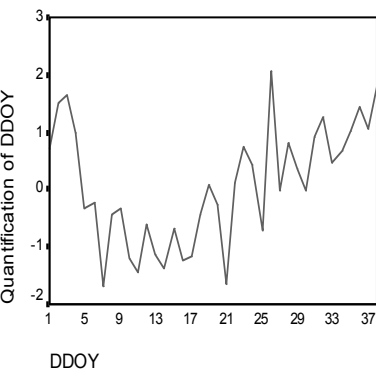
The transformation plot for *tempc* (Figure 8.21) displays an alternative pattern. As the categories increase, the quantifications tend to increase. As a result, as *tempc* increases, predicted *ozon* tends to increase. This pattern suggests scaling *tempc* at the ordinal level.

Figure 8.21   Transformation plot for tempc (nominal)
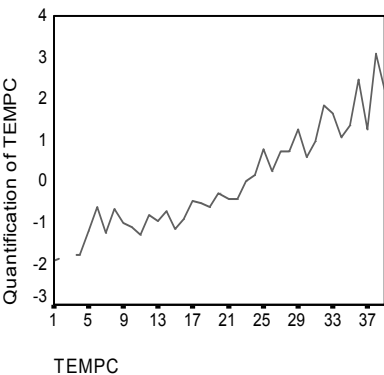
Thus, the transformation plots suggest scaling *tempc* at the ordinal level while keeping all other predictors nominally scaled. To recompute the regression, scaling *tempc* at the ordinal level, recall the Categorical Regression dialog box.

Select *tempc*. Click *Define Range and Scale.*
   Optimal Scaling Level
   ⊙ Ordinal

Options...
   ☑ Save transformed data

Plot...
   ▸ Plot: *tempc*

This model results in an R^2 of 0.873, so the variance accounted for decreases negligibly when the quantifications for *tempc* are restricted to be ordered.

Figure 8.22   Model summary for regression with tempc ordinal

Multiple R	R Square	Adjusted R Square
.934	.873	.871

Figure 8.23 displays the coefficients, correlations, and importances. Comparing the coefficients to those in Figure 8.16, no large changes occur. The importance measures suggest that *tempc* is still much more important to the regression than the other variables. Now, however, as a result of the ordinal scaling level of *tempc* and the positive regression coefficient, you can assert that as *tempc* increases, predicted *ozon* increases.

Figure 8.23   Coefficients and importances

	Standardized Coefficients		F	Importance
	Beta	Std. Error	F	Importance
DVIS	-.197	.020	95.122	.072
DIBH	-.269	.021	164.501	.151
DDPG	.240	.021	128.980	.034
TEMPC	.686	.021	1037.918	.642
DDOY	-.337	.021	253.566	.101

The transformation plot in Figure 8.24 illustrates the ordinal restriction on the quantifications for *tempc*. The jagged line in Figure 8.21 is here replaced by a smooth increasing line. Moreover, no long plateaus are present, indicating that collapsing categories is not needed.

Figure 8.24 Transformation plot for tempc (ordinal)
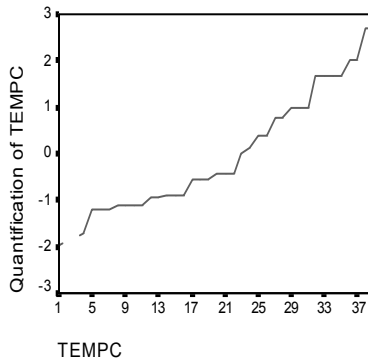
Optimality of the Quantifications

As stated previously, the transformed variables from a categorical regression can be used in a standard linear regression, yielding identical results. However, the quantifications are optimal only for the model that produced them. Using a subset of the predictors in linear regression does not correspond to an optimal scaling regression on the same subset.

For example, the categorical regression that you have computed has an R^2 of 0.873. You have saved the transformed variables, so in order to fit a linear regression using only *tempc*, *dvis*, and *dibh* as predictors, from the menus choose:

Analyze
  Regression
    Linear...

▶ Dependent: *trans1_1*
▶ Independent(s): *trans2_1*, *trans3_1*, *trans5_1*

Statistics...
  ☐ Descriptives (deselect)

  Regression Coefficients
  ☐ Estimates (deselect)

Figure 8.25 Model summary for regression with subset of optimally scaled predictors

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.870	.757	.755	.4962

Using the quantifications for the response, *tempc*, *dvis*, and *dibh* in a standard linear regression results in a fit of 0.757. To compare this to the fit of a categorical regression using just those three predictors, recall the Categorical Regression dialog box:

▸ Independent(s): *tempc*, *dvis*, *dibh*

Options...
  Display
  ☐ Coefficients (deselect)

  ☐ Save transformed data (deselect)
Plot...
  ▸ Plot: (blank)

Figure 8.26 Model summary for categorical regression on three predictors

Multiple R	R Square	Adjusted R Square
.889	.791	.789

The categorical regression analysis has a fit of 0.791, which is better than the fit of 0.757. This demonstrates the property of the scalings that the quantifications obtained in the original regression are only optimal when all five variables are included in the model.

Effects of Transformations

Transforming the variables makes a nonlinear relationship between the original response and the original set of predictors linear for the transformed variables. However, when there are multiple predictors, pairwise relationships are confounded by the other variables in the model.

To focus your analysis on the relationship between *ozon* and *ddoy*, begin by looking at a scatterplot. From the menus choose:

Graphs
  Scatter...

Select *Simple*. Click *Define*.

▸ Y Axis: *ozon*
▸ X Axis: *ddoy*

Figure 8.27 illustrates the relationship between *ozon* and *ddoy*. As *ddoy* increases to approximately 25, *ozon* increases. However, for *ddoy* values greater than 25, *ozon* decreases. This inverted U pattern suggests a quadratic relationship between the two variables. A linear regression cannot capture this relationship.

Figure 8.27   Scatterplot of ozon and ddoy
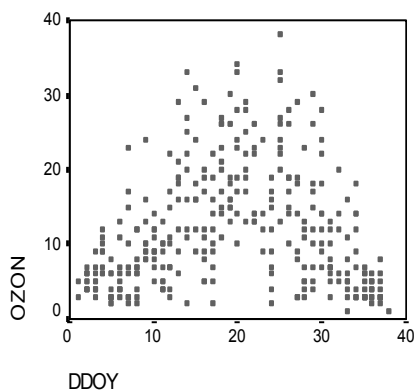
By excluding the other variables from the model, you can focus on the relationship between *ozon* and *ddoy*. However, all interpretations based on the reduced model apply only to the reduced model. Do not generalize the results to the regression involving all predictors.

To obtain a standard linear regression of *ozon* on *ddoy*, recall the Linear Regression dialog box:

▸ Dependent: *ozon*
▸ Independent(s): *ddoy*

Figure 8.28 Model summary for linear regression of ozon on ddoy

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.066	.004	.001	8.0057

The regression of *ozon* on *ddoy* yields an R^2 of 0.004. This fit suggests that *ddoy* has no predictive value for *ozon*. This is not surprising, given the pattern in Figure 8.27. By using optimal scaling, however, you can linearize the quadratic relationship and use the transformed *ddoy* to predict the response.

To obtain a categorical regression of *ozon* on *ddoy*, recall the Categorical Regression dialog box:

▶ Independent(s): *ddoy*

Select *ddoy*. Click *Define Scale*.
   Optimal Scaling
   ⊙ Nominal

Save...
   ☑ Transformed variables

Plots...
   ▶ Transformation Plots: *ddoy*

Figure 8.29 Model summary for categorical regression of ozon on ddoy

Multiple R	R Square	Adjusted R Square
.750	.562	.561

The optimal scaling regression treats *ozon* as numerical and *ddoy* as nominal. This results in an R^2 of 0.562. Although only 56% of the variation in *ozon* is accounted for by the categorical regression, this is a substantial improvement over the original regression. Transforming *ddoy* allows for the prediction of *ozon*.

   Figure 8.30 displays the transformation plot for *ddoy*. The extremes of *ddoy* both receive negative quantifications, whereas the central values have positive quantifications. By applying this transformation, the low and high *ddoy* values have similar effects on predicted *ozon.*

Figure 8.30   Transformation plot for ddoy (nominal)
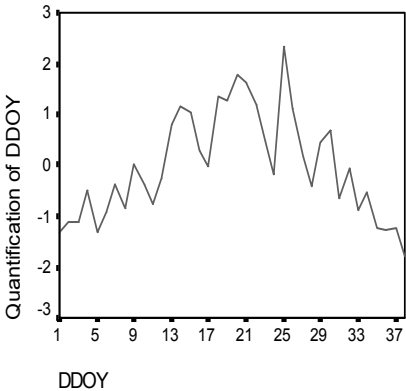
To see a scatterplot of the transformed variables, recall the Simple Scatterplot dialog box, and click *Reset* to clear your previous selections.

▶ Y Axis: *tra1_2*
▶ X Axis: *tra2_2*

Figure 8.31 depicts the relationship between the transformed variables. An increasing trend replaces the inverted U in Figure 8.27. The regression line has a slope of 0.750, indicating that as transformed *ddoy* increases, predicted *ozon* increases. Using optimal scaling linearizes the relationship and allows interpretations that would otherwise go unnoticed.

Figure 8.31   Scatterplot of the transformed variables
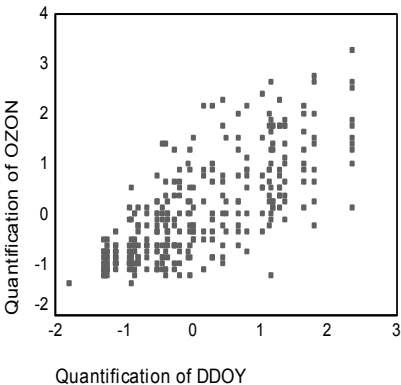
CATREG

```
CATREG [VARIABLES =] varlist

/ANALYSIS
        depvar [([LEVEL={SPORD**}]] [DEGREE={2}] [INKNOT={2}])]]
                                {n}           {n}
                        {SPNOM  }  [DEGREE={2}] [INKNOT={2}]
                                            {n}          {n}
                        {ORDI   }
                        {NOMI   }
                        {NUME   }
     WITH indvarlist [([LEVEL={SPORD**}]] [DEGREE={2}] [INKNOT={2}])]]
                                      {n}           {n}
                              {SPNOM  }  [DEGREE={2}] [INKNOT={2}]
                                                {n}          {n}
                              {ORDI   }
                              {NOMI   }
                              {NUME   }

[/DISCRETIZATION = [varlist [(([{GROUPING  }] [{NCAT*={7}}] [DISTR={NORMAL }])])]]]
                                                  {n}            {UNIFORM}
                                  {EQINTV=d  }
                                  {RANKING   }
                                  {MULTIPLYING}

[/MISSING = [{varlist}({LISTWISE**})]]]
            {ALL**  } {MODEIMPU  }
                      {EXTRACAT  }

[/SUPPLEMENTARY = OBJECT(objlist)]

[/INITIAL = [{NUMERICAL**}]]]
            {RANDOM     }

[/MAXITER = [{100**}]]]
            {n    }

[/CRITITER = [{.00001**}]]]
             {n       }

[/PRINT = [R**] [COEFF**] [DESCRIP**[(varlist)]] [HISTORY] [ANOVA**]
          [CORR] [OCORR] [QUANT[(varlist)]] [NONE]]

[/PLOT = {TRANS(varlist)[(h)]} {RESID(varlist)[(h)]}]

[/SAVE = {TRDATA[({TRA     })]} {PRED[({PRE    })]} {RES[({RES    })]}]
                  {rootname}           {rootname}          {rootname}

[/OUTFILE = {TRDATA('filename')} {DISCRDATA('filename')}] .
```

** Default if subcommand or keyword is omitted.

Overview

CATREG (Categorical regression with optimal scaling using alternating least squares) quantifies categorical variables using optimal scaling, resulting in an optimal linear regression equation for the transformed variables. The variables can be given mixed optimal scaling levels and no distributional assumptions about the variables are made.

Options

Transformation type. You can specify the transformation type (spline ordinal, spline nominal, ordinal, nominal, or numerical) at which you want to analyze each variable.

Discretization. You can use the DISCRETIZATION subcommand to discretize fractional-value variables or to recode categorical variables.

Initial configuration. You can specify the kind of initial configuration through the INITIAL subcommand.

Tuning the algorithm. You can control the values of algorithm-tuning parameters with the MAXITER and CRITITER subcommands.

Missing data. You can specify the treatment of missing data with the MISSING subcommand.

Optional output. You can request optional output through the PRINT subcommand.

Transformation plot per variable. You can request a plot per variable of its quantification against the category numbers.

Residual plot per variable. You can request an overlay plot per variable of the residuals and the weighted quantification, against the category numbers.

Writing external data. You can write the transformed data (category numbers replaced with optimal quantifications) to an outfile for use in further analyses. You can also write the discretized data to an outfile.

Saving variables. You can save the transformed variables, the predicted values, and/or the residuals in the working data file.

Basic Specification

The basic specification is the command CATREG with the VARIABLES and ANALYSIS subcommands.

Syntax Rules

- The VARIABLES and ANALYSIS subcommands must always appear, and the VARIABLES subcommand must be the first subcommand specified. The other subcommands, if specified, can be in any order.
- Variables specified in the ANALYSIS subcommand must be found in the VARIABLES subcommand.
- In the ANALYSIS subcommand, exactly one variable must be specified as a dependent variable and at least one variable must be specified as an independent variable after the keyword WITH.
- The word WITH is reserved as a keyword in the CATREG procedure. Thus, it may not be a variable name in CATREG. Also, the word TO is a reserved word in SPSS.

Operations

- If a subcommand is specified more than once, the last one is executed but with a syntax warning. Note this is true also for the VARIABLES and ANALYSIS subcommands.

Limitations

- If more than one dependent variable is specified in the ANALYSIS subcommand, CATREG is not executed.
- CATREG operates on category indicator variables. The category indicators should be positive integers. You can use the DISCRETIZATION subcommand to convert fractional-value variables and string variables into positive integers. If DISCRETIZATION is not specified, fractional-value variables are automatically converted into positive integers by grouping them into seven categories with a close to normal distribution and string variables are automatically converted into positive integers by ranking.
- In addition to system missing values and user defined missing values, CATREG treats category indicator values less than 1 as missing. If one of the values of a categorical variable has been coded 0 or some negative value and you want to treat it as a valid category, use the COMPUTE command to add a constant to the values of that variable such that the lowest value will be 1. (See the *SPSS Syntax Reference Guide* or the *SPSS Base User's Guide* for more information on COMPUTE). You can also use the RANKING option of the DISCRETIZATION subcommand for this purpose, except for variables you want to treat as numerical, since the characteristic of equal intervals in the data will not be maintained.
- There must be at least three valid cases.
- The number of valid cases must be greater than the number of independent variables plus 1.
- The maximum number of independent variables is 200.
- Split-File has no implications for CATREG.

Example

```
CATREG VARIABLES = TEST1 TEST3 TEST2 TEST4 TEST5 TEST6
                   TEST7 TO TEST9 STATUS01 STATUS02
  /ANALYSIS TEST4 (LEVEL=NUME)
   WITH TEST1 TO TEST2 (LEVEL=SPORD DEGREE=1 INKNOT=3) TEST5 TEST7
   (LEVEL=SPNOM) TEST8 (LEVEL=ORDI) STATUS01 STATUS02 (LEVEL=NOMI)
  /DISCRETIZATION = TEST1(GROUPING NCAT=5 DISTR=UNIFORM)
                    TEST5(GROUPING) TEST7(MULTIPLYING)
  /INITIAL = RANDOM
  /MAXITER = 100
  /CRITITER = .000001
  /MISSING = MODEIMPU
  /PRINT = R COEFF DESCRIP ANOVA QUANT(TEST1 TO TEST2 STATUS01
          STATUS02)
  /PLOT = TRANS (TEST2 TO TEST7 TEST4)
  /SAVE
  /OUTFILE = 'c:\data\qdata.sav'.
```

- VARIABLES defines variables. The keyword TO refers to the order of the variables in the working data file.

- The ANALYSIS subcommand defines variables used in the analysis. It is specified that *TEST4* is the dependent variable, with optimal scaling level numerical and that the variables *TEST1*, *TEST2*, *TEST3*, *TEST5*, *TEST7*, *TEST8*, *STATUS01*, and *STATUS02* are the independent variables to be used in the analysis. (The keyword TO refers to the order of the variables in the VARIABLES subcommand.) The optimal scaling level for *TEST1*, *TEST2*, and *TEST3* is spline ordinal, for *TEST5* and *TEST7* spline nominal, for *TEST8* ordinal, and for *STATUS01* and *STATUS02* nominal. The splines for *TEST1* and *TEST2* have degree 1 and three interior knots, the splines for *TEST5* and *TEST7* have degree 2 and two interior knots (default because unspecified).

- DISCRETIZATION specifies that *TEST5* and *TEST7*, which are fractional-value variables, are discretized: *TEST5* by recoding into seven categories with a normal distribution (default because unspecified) and *TEST7* by "multiplying." *TEST1*, which is a categorical variable, is recoded into five categories with a close-to-uniform distribution.

- Because there are nominal variables, a random initial solution is requested by the INITIAL subcommand.

- MAXITER specifies the maximum number of iterations to be 100. This is the default, so this subcommand could be omitted here.

- CRITITER sets the convergence criterion to a value smaller than the default value.

- To include cases with missing values, the MISSING subcommand specifies that for each variable, missing values are replaced with the most frequent category (the mode).

- PRINT specifies the correlations, the coefficients, the descriptive statistics for all variables, the ANOVA table, the category quantifications for variables *TEST1*, *TEST2*, *TEST3*, *STATUS01*, and *STATUS02*, and the transformed data list of all cases.

- PLOT is used to request quantification plots for the variables *TEST2*, *TEST5*, *TEST7*, and *TEST4*.

- The SAVE subcommand adds the transformed variables to the working data file. The names of these new variables are *TRANS1_1*, ..., *TRANS9_1*.

- The OUTFILE subcommand writes the transformed data to a data file called *qdata.sav* in the directory *c:\data*.

VARIABLES Subcommand

VARIABLES specifies the variables that may be analyzed in the current CATREG procedure.

- The VARIABLES subcommand is required and precedes all other subcommands. The actual keyword VARIABLES can be omitted. (Note that the equals sign is always optional in SPSS syntax.)

- The keyword TO on the VARIABLES subcommand refers to the order of variables in the working data file. (Note that this behavior of TO is different from that in the indvarlist on the ANALYSIS subcommand.)

ANALYSIS Subcommand

ANALYSIS specifies the dependent variable and the independent variables following the keyword WITH.

- All the variables on ANALYSIS must be specified on the VARIABLES subcommand.
- The ANALYSIS subcommand is required and follows the VARIABLES subcommand.
- The first variable list contains exactly one variable as the dependent variable, while the second variable list following WITH contains at least one variable as an independent variable. Each variable may have at most one keyword in parentheses indicating the transformation type of the variable.
- The keyword TO in the independent variable list honors the order of variables on the VARIABLES subcommand.
- Optimal scaling levels are indicated by the keyword LEVEL in parentheses following the variable or variable list.

LEVEL   *Specifies the optimal scaling level.*

LEVEL Keyword

The following keywords are used to indicate the optimal scaling level:

SPORD   *Spline ordinal (monotonic).* This is the default for a variable listed without any optimal scaling level, for example, one without LEVEL in the parentheses after it or with LEVEL without a specification. Categories are treated as ordered. The order of the categories of the observed variable is preserved in the optimally scaled variable. Categories will be on a straight line through the origin. The resulting transformation is a smooth nondecreasing piecewise polynomial of the chosen degree. The pieces are specified by the number and the placement of the interior knots.

SPNOM   *Spline nominal (non-monotonic).* Categories are treated as unordered. Objects in the same category obtain the same quantification. Categories will be on a straight line through the origin. The resulting transformation is a smooth piecewise polynomial of the chosen degree. The pieces are specified by the number and the placement of the interior knots.

ORDI   *Ordinal.* Categories are treated as ordered. The order of the categories of the observed variable is preserved in the optimally scaled variable. Categories will be on a straight line through the origin. The resulting transformation fits better than SPORD transformation, but is less smooth.

NOMI   *Nominal.* Categories are treated as unordered. Objects in the same category obtain the same quantification. Categories will be on a straight line through the origin. The resulting transformation fits better than SPNOM transformation, but is less smooth.

NUME	*Numerical.* Categories are treated as equally spaced (interval level). The order of the categories and the differences between category numbers of the observed variables are preserved in the optimally scaled variable. Categories will be on a straight line through the origin. When all variables are scaled at the numerical level, the CATREG analysis is analogous to standard multiple regression analysis.

SPORD and SPNOM Keywords

The following keywords are used with SPORD and SPNOM :

DEGREE	*The degree of the polynomial.* If DEGREE is not specified the degree is assumed to be 2.
INKNOT	*The number of the interior knots.* If INKNOT is not specified the number of interior knots is assumed to be 2.

DISCRETIZATION Subcommand

DISCRETIZATION specifies fractional-value variables that you want to discretize. Also, you can use DISCRETIZATION for ranking or for two ways of recoding categorical variables.

- A string variable's values are always converted into positive integers by assigning category indicators according to the ascending alphanumeric order. DISCRETIZATION for string variables applies to these integers.

- When the DISCRETIZATION subcommand is omitted, or when the DISCRETIZATION subcommand is used without a varlist, fractional-value variables are converted into positive integers by grouping them into seven categories (or into the number of distinct values of the variable if this number is less than 7) with a close to normal distribution.

- When no specification is given for variables in a varlist following DISCRETIZATION, these variables are grouped into seven categories with a close-to-normal distribution.

- In CATREG, a system-missing value, user-defined missing values, and values less than 1 are considered to be missing values (see next section). However, in discretizing a variable, values less than 1 are considered to be valid values, and are thus included in the discretization process. System-missing values and user-defined missing values are excluded.

GROUPING	*Recode into the specified number of categories.*
RANKING	*Rank cases.* Rank 1 is assigned to the case with the smallest value on the variable.
MULTIPLYING	*Multiplying the standardized values (z-scores) of a fractional-value variable by 10, rounding, and adding a value such that the lowest value is 1.*

GROUPING Keyword

NCAT   *Recode into ncat categories.* When NCAT is not specified, the number of categories is set to 7 (or the number of distinct values of the variable if this number is less than 7). The valid range is from 2 to 36. You may either specify a number of categories or use the keyword DISTR.

EQINTV   *Recode intervals of equal size into categories.* The interval size must be specified (there is no default value). The resulting number of categories depends on the interval size.

DISTR Keyword

DISTR has the following keywords:

NORMAL   *Normal distribution.* This is the default when DISTR is not specified.

UNIFORM   *Uniform distribution.*

MISSING Subcommand

In CATREG, we consider a system missing value, user defined missing values, and values less than 1 as missing values. However, in discretizing a variable (see previous section), values less than 1 are considered as valid values. The MISSING subcommand allows you to indicate how to handle missing values for each variable.

LISTWISE   *Exclude cases with missing values on the specified variable(s).* The cases used in the analysis are cases without missing values on the variable(s) specified. This is the default applied to all variables, when the MISSING subcommand is omitted or is specified without variable names or keywords. Also, any variable which is not included in the subcommand gets this specification.

MODEIMPU   *Impute missing value with mode.* All cases are included and the imputations are treated as valid observations for a given variable. When there are multiple modes, the smallest mode is used.

EXTRACAT   *Impute missing values on a variable with an extra category indicator.* This implies that objects with a missing value are considered to belong to the same (extra) category. This category is treated as nominal, regardless of the optimal scaling level of the variable.

- The ALL keyword may be used to indicate all variables. If it is used, it must be the only variable specification.
- A mode or extra-category imputation is done before listwise deletion.

SUPPLEMENTARY Subcommand

The SUPPLEMENTARY subcommand specifies the objects that you want to treat as supplementary. You cannot weight supplementary objects (specified weights are ignored).

OBJECT   *Supplementary objects.* Objects that you want to treat as supplementary are indicated with an object number list in parentheses following OBJECT. The keyword TO is allowed, for example, OBJECT(1 TO 1 3 5 TO 9).

INITIAL Subcommand

INITIAL specifies the method used to compute the initial value/configuration.

- The specification on INITIAL is keyword NUMERICAL or RANDOM. If INITIAL is not specified, NUMERICAL is the default.

NUMERICAL   *Treat all variables as numerical.* This is usually best to use when there are only numerical and/or ordinal variables.

RANDOM   *Provide a random initial value.* This should be used only when there is at least one nominal variable.

MAXITER Subcommand

MAXITER specifies the maximum number of iterations CATREG can go through in its computations. Note that the output starts from the iteration number 0, which is the initial value before any iteration, when INITIAL = NUMERICAL is in effect.

- If MAXITER is not specified, CATREG will iterate up to 100 times.

- The specification on MAXITER is a positive integer indicating the maximum number of iterations. There is no uniquely predetermined (hard coded) maximum for the value that can be used.

CRITITER Subcommand

CRITITER specifies a convergence criterion value. CATREG stops iterating if the difference in fit between the last two iterations is less than the CRITITER value.

- If CRITITER is not specified, the convergence value is 0.00001.

- The specification on CRITITER is any value less than or equal to 0.1 and greater than or equal to .000001. (Values less than the lower bound might seriously affect performance. Therefore, they are not supported.)

PRINT Subcommand

The PRINT subcommand controls the display of output. The output of the CATREG procedure is always based on the transformed variables. However, the correlations of the original predictor variables can be requested as well by the keyword OCORR. The default keywords are R, COEFF, DESCRIP, and ANOVA. That is, the four keywords are in effect when the PRINT subcommand is omitted or when the PRINT subcommand is given without any keyword. If a keyword is duplicated or it encounters a contradicting keyword, such as /PRINT = R R NONE, then the last one silently becomes effective.

R   *Multiple R.* Includes R^2, adjusted R^2, and adjusted R^2 taking the optimal scaling into account.

COEFF   *Standardized regression coefficients (beta).* This option gives three tables: a Coefficients table that includes betas, standard error of the betas, *t* values, and significance; a Coefficients-Optimal Scaling table, with the standard error of the betas taking the optimal scaling degrees of freedom into account; and a table with the zero-order, part, and partial correlation, Pratt's relative importance measure for the transformed predictors, and the tolerance before and after transformation. If the tolerance for a transformed predictor is lower than the default tolerance value in the SPSS Regression procedure (0.0001), but higher than 10E–12, this is reported in an annotation. If the tolerance is lower than 10E–12, then the COEFF computation for this variable is not done and this is reported in an annotation. Note that the regression model includes the intercept coefficient but that its estimate does not exist because the coefficients are standardized.

DESCRIP(varlist)   *Descriptive statistics (frequencies, missing values, and mode).* The variables in the varlist must be specified on the VARIABLES subcommand, but need not appear on the ANALYSIS subcommand. If DESCRIP is not followed by a varlist, Descriptives tables are displayed for all of the variables in the variable list on the ANALYSIS subcommand.

HISTORY   *History of iterations.* For each iteration, including the starting values for the algorithm, the multiple *R* and the regression error (square root of (1–multiple R^2)) are shown. The increase in multiple *R* is listed from the first iteration.

ANOVA   *Analysis-of-variance tables.* This option includes regression and residual sums of squares, mean squares and *F*. This options gives two ANOVA tables: one with degrees of freedom for the regression equal to the number of predictor variables and one with degrees of freedom for the regression taking the optimal scaling into account.

CORR   *Correlations of the transformed predictors.*

OCORR   *Correlations of the original predictors.*

QUANT(varlist)   *Category quantifications.* Any variable in the ANALYSIS subcommand may be specified in parentheses after QUANT. If QUANT is not followed by a varlist, Quantification tables are displayed for all variables in the variable list on the ANALYSIS subcommand.

NONE   *No* PRINT *output is shown.* This is to suppress the default PRINT output.

- The keyword TO in a variable list can only be used with variables that are in the ANALYSIS subcommand, and TO applies only to the order of the variables in the ANALYSIS subcommand. For variables that are in the VARIABLES subcommand but not in the ANALYSIS subcommand, the keyword TO cannot be used. For example, if /VARIABLES = v1 TO v5 and /ANALYSIS is v2 v1 v4, then /PRINT QUANT(v1 TO v4) will give two quantification plots, one for *v1* and one for *v4*. (/PRINT QUANT(v1 TO v4 v2 v3 v5) will give quantification tables for *v1*, *v2*, *v3*, *v4*, and *v5*.)

PLOT Subcommand

The PLOT subcommand controls the display of plots.

- In this subcommand, if no plot keyword is given, then no plot is created. Further, if the variable list following the plot keyword is empty, then no plot is created, either.

- All the variables to be plotted must be specified in the ANALYSIS subcommand. Further, for the residual plots, the variables must be independent variables.

TRANS(varlist)(l)   *Transformation plots (optimal category quantifications against category indicators).* A list of variables must come from the ANALYSIS variable list and must be given in parentheses following the keyword. Further, the user can specify an optional parameter l in parentheses after the variable list in order to control the global upper boundary of category label lengths in the plot. Note that this boundary is applied uniformly to all transformation plots.

RESID(varlist)(l)   *Residual plots (residuals when the dependent variable is predicted from all predictor variables in the analysis except the predictor variable in varlist, against category indicators, and the optimal category quantifications multiplied with Beta against category indicators).* A list of variables must come from the ANALYSIS variable list's independent variables and must be given in parentheses following the keyword. Further, the user can specify an optional parameter l in parentheses after the variable list in order to control the global upper boundary of category label lengths in the plot. Note that this boundary is applied uniformly to all residual plots.

- The category label length parameter (l) can take any non-negative integer less than or equal to 20. If l = 0, values instead of value labels are displayed to indicate the categories on the *x* axis in the plot. If l is not specified, CATREG assumes that each value label at its full length is displayed as a plot's category label, but currently LINE CHART in GRAPH limit them to 20. Thus, it is equivalent to (l = 20). (Note that the VALUE LABELS command allows up to 60 characters.) If l is an integer larger than 20, then we reset it to 20 and issue a warning saying l must be a non-negative integer less than or equal to 20.

- If a positive value of l is given, but if some or all of the values do not have value labels, then for those values, the values themselves are used as the category labels, and they obey the label length constraint.

- The keyword TO in a variable list can only be used with variables that are in the ANALYSIS subcommand, and TO applies only to the order of the variables in the ANALYSIS subcommand. For variables that are in the VARIABLES subcommand but not in the ANALYSIS subcommand, the keyword TO cannot be used. For example, if /VARIABLES = v1 TO v5 and /ANALYSIS is v2 v1 v4, then /PLOT TRANS(v1 TO v4) will give two transformation plots, one for *v1* and for *v4*. (/PLOT TRANS(v1 TO v4 v2 v3 v5) will give transformation plots for *v1*, *v2*, *v3*, *v4*, and *v5*.)

SAVE Subcommand

The SAVE subcommand is used to add the transformed variables (category indicators replaced with optimal quantifications), the predicted values, and the residuals to the working data file.

Excluded cases are represented by a dot (the sysmis symbol) on every saved variable.

TRDATA   *Transformed variables.*

PRED   *Predicted values.*

RES   *Residuals.*

- A variable rootname can be specified with each of the keywords. Only one rootname can be specified with each keyword, and it can contain up to five characters (if more than one rootname is specified with a keyword, the first rootname is used; if a rootname contains more than five characters, the first five characters are used at most). If a rootname is not specified, the default rootnames (TRA, PRE, and RES) are used.

- CATREG adds two numbers separated by an underscore (_) to the rootname. The formula is *ROOTNAMEk_n* where *k* increments from 1 to identify the source variable names by using the source variables' position numbers in the ANALYSIS subcommand (that is, the dependent variable has the position number 1, and the independent variables have the position numbers 2, 3, ... as they are listed), and *n* increments from 1 to identify the CATREG procedures with the successfully executed SAVE subcommands for a given data file in a continuous SPSS session. For example, with two predictor variables specified on ANALYSIS, the first set of default names for the transformed data, if they do not exist in the data file, would be *TRA1_1*, for the dependent variable, and *TRA2_1*, *TRA3_1* for the predictor variables. The next set of default names, if they do not exist in the data file, would be *TRA1_2*, *TRA2_2*, *TRA3_2*. However, if, for example, *TRA1_2* already exists in the data file, then the default names should be attempted as *TRA1_3*, *TRA2_3*, *TRA3_3*— that is, the last number increments to the next available integer.

- As *k* and/or *n* increase, the rootname is truncated to keep variable names within eight characters. For example, if *TRANS* is specified as rootname, *TRANS1_9* would be followed by *TRAN1_10*. The initial character (*T* in this example) is required. Note that the truncation is done variable-wise, not analysis-wise.

- Variable labels are created automatically. (They are shown in the Procedure Information Table (the Notes table) and can also be displayed in the Data Editor window.)

OUTFILE Subcommand

The OUTFILE subcommand is used to write the discretized data and/or the transformed data (category indicators replaced with optimal quantifications) to an external data file. Excluded cases are represented by a dot (the sysmis symbol) on every saved variable.

DISCRDATA('filename')   *Discretized data.*

TRDATA('filename')   *Transformed variables.*

- Following the keyword, a filename enclosed by single quotation marks should be specified. The filenames should be different for the each of the keywords.

- A working data file, in principle, should not be replaced by this subcommand, and the asterisk (*) file specification is not supported. This strategy also prevents the OUTFILE interference with the SAVE subcommand.

Appendix C Notation

$\lvert \cdot \rvert$	absolute value
$\lVert \cdot \rVert^2$	squared Euclidean norm
$\mathbf{y}$	observed response variable
$\mathbf{X}$	matrix of observed predictor variables
$\mathbf{x}_j$	observed predictor variable j
$\varphi(\mathbf{y})$	transformed response variable
$\varphi_j(\mathbf{x}_j)$	transformed predictor variable j
$\mathbf{G}_r$	indicator matrix for the response variable
$\mathbf{G}_j$	indicator matrix for predictor variable j
$\mathbf{D}_r$	inner product of $\mathbf{G}_r$ (contains marginal frequencies on its main diagonal)
$\mathbf{D}_j$	inner product of $\mathbf{G}_j$
$\mathbf{v}_r$	category quantifications for the response variable
$\mathbf{v}_j$	category quantifications for predictor variable j
$\mathbf{b}$	vector of standardized regression coefficients
β_j	standardized regression coefficient for predictor variable j
a_{k_j}	unstandardized regression coefficient for a dummy variable representing category k of variable j
$\mathbf{W}$	matrix of weights
$\mathbf{w}$	vector of weights
w_j	weight for predictor variable j
$\mathbf{e}$	error vector
$\mathbf{I}$	identity matrix
N	number of cases
P	number of predictor variables
K_j	number of categories for variable j
B	number of bootstrap samples
R^2	squared multiple regression coefficient
$\overline{\mathrm{err}}$	estimate of apparent prediction error
$\widehat{\mathrm{Err}}^{(.632)}$	.632 bootstrap estimate of expected prediction error
$\overline{\mathrm{Err}}^{(1)}$	leave-one-out bootstrap estimate of prediction error
$\widehat{\mathrm{OP}}$	estimate of optimism
C^{-i}	set of indices of bootstrap samples that do not contain observation i
$\lvert C^{-i} \rvert$	number of bootstrap samples that do not contain observation i
λ_2	Ridge penalty parameter
λ_1	Lasso penalty parameter
γ_{A_k}	slope of piece k of regularization paths
A_k	set of indices of the active predictors on piece k of regularization paths

REFERENCES

Altman, D. G. and Bland, J. M. (1994), "Statistics Notes: Diagnostic tests 3: receiver operating characteristic plots," *British Medical Journal*, 309, 188–188.

American Psychiatric Association (1994), *Diagnostic and Statistical Manual of Mental Disorders* (Fourth ed.): Washington, DC: American Psychiatric Association.

Barlow, R. E., Bartholomew, D. J., Brenner, J. M., and Brunk, H. D. (1972), *Statistical inference under order restrictions*: New York: John Wiley and Sons.

Box, G. E. P. and Cox, D. R. (1964), "An Analysis of transformations," *Journal of the Royal Statistical Society Series B*, 26, 211–252.

Box, G. E. P. and Tidwell, P. W. (1962), "Transformation of independent variables," *Technometrics*, 4, 531–550.

Breiman, L. (1995), "Better subset selection using the nonnegative Garotte," *Technometrics*, 37, 373–383.

Breiman, L. and Friedman, J. H. (1985), "Estimating optimal transformation for multiple regression and correlation (with discussion)," *Journal of the American Statistical Association*, 80, 580–619.

Buja, A. (1990), "Remarks on functional canonical variates, alternating least squares methods and ACE," *Annals of Statistics*, 18, 1032–1069.

Buja, A., Hastie, T. J., and Tibshirani, R. J. (1989), "Linear smoothers and additive models (with discussion)," *Annals of Statistics*, 17, 453–555.

Casper, R. C. (1998), "Depression and eating disorders," *Depression and Anxiety*, 8 (suppl. 1), 96–104.

Cassin, S. E. and von Ranson, K. M. (2005), "Personality and eating disorders: A decade in review," *Clinical Psychology Review*, 25, 895–916.

Clausen, L. (2004a), "Review of studies evaluating psychotherapy in bulimia nervosa: The influence of research methods," *Scandinavian Journal of Psychology*, 45, 247–252.

——— (2004b), "Time Course of Symptom Remission in Eating Disorders," *International Journal of Eating Disorders*, 36, 296–306.

Cooper, P. J. and Fairburn, C. G. (1986), "The depressive symptoms of bulimia nervosa," *British Journal of Psychiatry*, 148, 268–274.

De Leeuw, J., Young, F. W., and Takane, Y. (1976), "Additive structure in qualitative data," *Psychometrika*, 41, 471–503.

Derogatis, L. R. (1977), *SCL-90-R, administration, scoring and procedures manual-I for the R(evised) version*: Baltimore: Johns Hopkins University School of Medicine.

Draper, N. R. and Smith, H. (1966), *Applied Regression Analysis*: New York: Wiley.

Efron, B. E. (1983), "Estimating the error rate of a prediction rule: improvements on cross-validation," *Journal of the American Statistical Association*, 78, 316–313.

Efron, B. E., Hastie, T. J., Johnstone, I. M., and Tibshirani, R. J. (2004), "Least Angle Regression," *The Annals of Statistics*, 32, 407–451.

Efron, B. E. and Tibshirani, R. J. (1993), *An introduction to the bootstrap*: New York: Chapman and Hall.

——— (1997), "Improvements on cross-validation: The .632+ bootstap method," *Journal of the American Statistical Association*, 92, 548–560.

Fichter, M. (1991), "Structured Interview for Anorexia and Bulimia Nervosa (SIAB): Development of a new Instrument for the Assessment of Eating Disorders," *International Journal for Eating Disorders*, 10, 571–592.

Fichter, M. and Quadflieg, N. (1997), "Six-year course of bulimia nervosa," *International Journal of Eating Disorders*, 22, 361–384.

——— (2001), "The structured interview for anorexic and bulimic disorders for DSM-IV and ICD-10 (SIAB-EX): reliability and validity," *European psychiatry*, 16, 38–48.

Fichter, M. M., Herpertz, S., Quadflieg, N., and Herpertz-Dahlmann, B. (1998), "Structured Interview for Anorexic and Bulimic disorders for DSM-IV and ICD-10: updated (third) revision," *International Journal of Eating Disorders*, 24, 227–249.

Frank, I. E. and Friedman, J. H. (1993), "A statistical view of some chemometrics regression tools," *Technometrics*, 35, 109–148.

Franke, G. (2002), *Die Symptom-Checkliste von Derogatis (SCL-90-R) - Deutsche Version - Manual* (Second ed.): Göttingen: Hogrefe.

Friedman, J. H. (1984), "A variable span scatterplot smoother," Technical Report 5, Laboratory for Computational Statistics, Stanford University.

Friedman, J. H. and Popescu, B. E. (2004), "Gradient Directed Regularization for Linear Regression and Classification," Technical report, Department of Statistics, available at http://www-stat.stanford.edu/~jhf.

Friedman, J. H. and Silverman, B. W. (1989), "Flexible parsimonious smoothing and additive modelling (with discussion)," *Technometrics*, 31, 3–39.

Friedman, J. H. and Stuetzle, W. (1981), "Projection Pursuit Regression," *Journal of the American Statistical Association*, 76, 817–823.

Fu, W. J. (1998), "Penalized Regression: The Bridge Versus the Lasso," *Journal of Computational and Graphical Statistics*, 7, 397–416.

Garner, D. M. (1991), *Eating Disorder Inventory-2. Professional Manual*: Odessa, Florida: Psychological Assessment Resources.

Gaudart, J., Giusiano, B., and Huiart, L. (2004), "Comparison of the performance of multi-layer perceptron and linear regression for epidemiological data," *Computational Statistics and Data Analysis*, 44, 547–570.

Gifi, A. (1990), *Nonlinear multivariate analysis*: Chichester UK: John Wiley, Original work published 1981.

Golub, G. H., Heath, M., and Wahba, G. (1979), "Generalized Cross-Validation as a Method for Choosing a Good Ridge Parameter," *Technometrics*, 21, 215–223.

Golub, G. H. and Van Loan, C. F. (1983), *Matrix computations*: Johns Hopkins University Press: Baltimore.

Härdle, W. and Hall, P. (1993), "On the backfitting algorithm for additive regression models," *Statistica Neerlandica*, 47, 43–57.

Hastie, T. J. (1989), "Discussion of 'Flexible parsimonious smoothing and additive modelling' by J. Friedman and B. Silverman," *Technometrics*, 31, 3–39.

Hastie, T. J. and Tibshirani, R. J. (1990), *Generalized additive models*: New York: Chapman and Hall.

——— (2000), "Bayesian backfitting," *Statistical Science*, 15, 196–223.

Hastie, T. J., Tibshirani, R. J., and Buja, A. (1994), "Flexible discriminant analysis by optimal scoring," *Journal of the American Statistical Association*, 89, 1255–1270.

Hastie, T. J., Tibshirani, R. J., and Friedman, J. H. (2001), *The Elements of Statistical Learning*: New York: Springer.

Hawkins, D. (1989), "Discussion of 'Flexible parsimonious smoothing and additive modelling' by J. Friedman and B. Silverman," *Technometrics*, 31, 3–39.

Hoerl, A. E. and Kennard, R. (1970a), "Ridge Regression: Applications to nonorthogonal problems," *Technometrics*, 12, 69–82.

——— (1970b), "Ridge Regression: biased estimation for nonorthogonal problems," *Technometrics*, 12, 55–67.

Jacobson, N. S., Follette, W. C., and Revenstorf, D. (1984), "Psychotherapy outcome research: Methods for reporting variability and evaluating clinical significance," *Behavior Therapy*, 15, 336–352.

Jacobson, N. S., Roberts, L. J., Berns, S. B., and McGlinchey, J. B. (1999), "Methods for defining and determining the clinical significance of treatment effects: description, application, and alternatives," *Journal of Consulting and Clinical Psychology*, 67, 300–307.

Jacobson, N. S. and Truax, P. (1991), "Clinical significance: A statistical approach to defining meaningful change in psychotherapy research," *Journal of Consulting and Clinical Psychology*, 59, 12–19.

Judd, C. and McClelland, G. (1989), *Data Analysis: A Model-Comparison Approach*: San Diego, CA: Harcourt Brace Jovanich.

Kim, Y., Kim, J., and Kim, Y. (2006), "Blockwise Sparse Regression," *Statistica Sinica*, 16.

Kruskal, J. B. (1964a), "Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis," *Psychometrika*, 29, 1–28.

—— (1964b), "Nonmetric multidimensional scaling: a numerical method," *Psychometrika*, 29, 115–129.

—— (1965), "Analysis of factorial experiments by estimating monotone transformations of the data," *Journal of the Royal Statistical Society Series B*, 27, 251–263.

Levy, A. B., Dixon, K. N., and Stern, S. L. (1989), "How are depression and bulimia related?" *American Journal of Psychiatry*, 146, 162–169.

Lin, Y. and Zhang, H. (2003), "Component Selection and Smoothing in Smoothing Spline Analysis of Variance Models," Technical Report 1072, Department of Statistics, University of Wisconsin-Madison.

Mammen, E., Linton, O. B., and Nielsen, J. P. (1999), "The existence and asymptotic properties of a backfitting projection algorithm under weak conditions," *The Annals of Statistics*, 27, 1443–1490.

MathSoft, Inc. (1999), *S-PLUS 2000 Professional Release 1*: Seattle WA: MathSoft, Inc.

McCullagh, P. and Nelder, J. A. (1989), *Generalized Linear Models*: London: Chapman and Hall.

Meermann, R., Napierski, C. H., and Schulenkorf, E. M. (1987), "EDI-Münster - Selbstbeurteilungsfragebogen für Eßstörungen," in *Therapie der Magersucht und Bulimia Nervosa. Ein klinischer Leitfaden für den Praktiker*, eds. R. Meermann and W. Vandereycken, Berlin; New York: Walter de Gruyter.

Meulman, J. J., Heiser, W. J., and SPSS Inc. (1999), *SPSS Categories 10.0*: Chicago: SPSS Inc.

—— (2004), *SPSS Categories 13.0*: Chicago: SPSS Inc.

Meulman, J. J. and Van der Kooij, A. J. (2000), *Transformations towards independence through optimal scaling*: Paper presented at the International Conference on Measurement and Multivariate Analysis (ICMMA), Banff, Canada.

Nelder, J. A. and Wedderburn, R. W. M. (1972), "Generalized linear models," *Journal of the Royal Statistical Society Series A*, 135, 370–384.

Nielsen, J. P. and Sperlich, S. (2005), "Smooth backfitting in practice," *Journal of the Royal Statistical Society Series B*, 67, 43–61.

Nollett, C. L. and Button, E. J. (2005), "Questionnaire Measures of Psychopathology in Eating Disorders: Comparisons Between Clinical Groups," *European Eating Disorders Review*, 13, 211–215.

Olmsted, M. P., Kaplan, A. S., and Rockert, W. (2005), "Defining Remission and Relapse in Bulimia Nervosa," *International Journal of Eating Disorders*, 38, 1–6.

Opsomer, J. D. and Ruppert, D. (1997), "Fitting a bivariate additive model by local polynomial regression," *Annals of Statistics*, 25, 186–211.

Osborne, M. R., Presnell, B., and Turlach, B. A. (2000), "On the LASSO and its Dual," *Journal of Computational and Graphical Statistics*, 9, 319–337.

Perkins, S., Lacker, K., and Theiler, J. (2003), "Grafting: Fast, Incremental Feature Selection by Gradient Descent in Function Space," *Journal of Machine Learning Research*, 3, 1333–1356.

Pratt, J. W. (1987), "Dividing the indivisible: using simple symmetry to partition variance explained," in *Proceedings of the Second International Conference in Statistics*, eds. T. Pukkila and S. Puntanen, Tampere, Finland: University of Tampere, pp. 245–260.

Ramsay, J. O. (1988), "Monotone regression splines in action," *Statistical Science*, 4, 425–441.

Ripley, B. D. (1996), *Pattern recognition and neural networks*: Cambridge U.K.: Cambridge University Press.

Rounsaville, B. J., Alarcón, R. D., Andrews, G., Jackson, J. S., Kendell, R. E., and Kendler, K. (2002), "Basic Nomenclature Issues for DSM-V," in *A research agenda for DSM-V*, eds. D. J. Kupfer, M. B. First, and D. A. Regier, Washington, DC: American Psychiatric Association.

SAS/STAT (1989), *User's Guide, Version 6, Vol. 2*: Cary N.C.: SAS Institute Inc.

Seber, G. A. F. and Wild, C. J. (1989), *Nonlinear Regression*: New York: John Wiley and Sons.

SPSS Inc. (1998), *SPSS Categories, 8.0*: Chicago: SPSS Inc.

Stamey, T. A., Kabalin, J. N., McNeal, J. E., Johnstone, I. M., Freiha, F., Redwine, E. A., and Yang, N. (1989), "Prostate specific antigen in the diagnosis and treatment of adenocarnicoma of the prostate, ii: Radical prostatectomy treated patients," *Journal of Urology*, 16, 1076–1083.

Stone, C. J. (1985), "Additive regression and other nonparametric models," *Annals of Statistics*, 13, 689–705.

Tibshirani, R. J. (1996), "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society Series B*, 58, 267–288.

Tobin, D. L. and Griffing, A. S. (1995), "Coping and depression in bulimia nervosa," *International Journal of Eating Disorders*, 18, 359–363.

Van Buuren, S. and Heiser, W. J. (1989), "Clustering n objects into k groups under optimal scaling of variables," *Psychometrika*, 54, 699–706.

Van der Burg, E. and De Leeuw, J. (1983), "Non-linear canonical correlation," *British Journal of Mathematical and Statistical Psychology*, 36, 54–80.

Van der Kooij, A. J. and Meulman, J. J. (1997), "MURALS: Multiple regression and optimal scoring using alternating least squares," in *Softstat '97 Advances in Statistical Software 6*, eds. W. Bandilla and F. Faulbaum, Stuttgart: Lucius & Lucius, pp. 99–106.

—— (1999), "Regression with Optimal Scaling," in *SPSS Categories 10.0*, eds. J. J. Meulman, W. J. Heiser, and SPSS Inc., Chicago: SPSS Inc., pp. 1–8, 77–101.

—— (2004), "Regression with Optimal Scaling," in *SPSS Categories 13.0*, eds. J. J. Meulman, W. J. Heiser, and SPSS Inc., Chicago: SPSS Inc., pp. 1–10, 107–157.

—— (2006a), "Prediction Accuracy of Regression with Optimal Scaling Transformations: The .632 Bootstrap with CATREG," *Manuscript submitted for publication*.

—— (2006b), "Regularization with Ridge Regression, the Lasso, and the Elastic Net for Regression with Optimal Scaling Transformations," *Manuscript submitted for publication*.

Van der Kooij, A. J., Meulman, J. J., and Heiser, W. J. (2006), "Local Minima in Categorical Multiple Regression," *Computational Statistics and Data Analysis*, 50, 446–462.

Vapnik, V. (1996), *The Nature of Statistical Learning Theory*: New York: Springer-Verlag.

Winsberg, S. and Ramsay, J. O. (1980), "Monotonic transformations to additivity using splines," *Biometrica*, 67, 669–674.

Young, F. W. (1981), "Quantitative analysis of qualitative data," *Psychometrika*, 46, 357–388.

Young, F. W., De Leeuw, J., and Takane, Y. (1976), "Regression with qualitative and quantitative variables: An alternating least squares method with optimal scaling features," *Psychometrika*, 41, 505–529.

Yuan, M. and Lin, Y. (2006), "Model Selection and Estimation in Regression with Grouped Variables," *Journal of the Royal Statistical Society Series B*, 68, 49–67.

Zhao, P. and Yu, B. (2004), "Boosted Lasso," Technical Report 678, University of California, Berkeley, available at http://www.stat.berkeley.edu/ users/pengzhao.

Zou, H. and Hastie, T. J. (2005), "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society Series B*, 67, 301–320.

Zou, H., Hastie, T. J., and Tibshirani, R. J. (2004), "On the 'Degrees of Freedom' of the Lasso," *manuscript submitted for publication*, available at http://www-stat.stanford.edu/~hastie.