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: $\sum_{i=1}^{n} w_i$
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: $\sum_{j \in J_p} b_j \mathbf{G}_j \mathbf{y}_j$

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 k^{th} 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 $\mathbf{u}'\mathbf{W}\mathbf{G}_j\tilde{\mathbf{y}}_j = 0$ 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

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

Nominal: $\mathbf{y}_r^* = \tilde{\mathbf{y}}_r$

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.

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

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

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

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).

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

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):

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

(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

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

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

ANOVA Table

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

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

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 $\mathbf{q}_j = \mathbf{G}_j\mathbf{y}_j \quad j \in J_p$.

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

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

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

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

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

Part correlation

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

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

Importance

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

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

Tolerance

The tolerance for the optimally scaled predictor variables is given by

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

with $r_{p_{jj}}$ the j^{th} 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 ($\mathbf{G}_r\mathbf{y}_r - \left(\mathbf{v} - b_j\mathbf{G}_j\mathbf{y}_j\right)$) 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.