

ePub^{WU} Institutional Repository

Torsten Hothorn and Achim Zeileis
Generalized Maximally Selected Statistics

Paper

Original Citation:

Hothorn, Torsten and Zeileis, Achim

(2007)

Generalized Maximally Selected Statistics.

Research Report Series / Department of Statistics and Mathematics, 52. Department of Statistics and Mathematics, WU Vienna University of Economics and Business, Vienna.

This version is available at: <https://epub.wu.ac.at/1252/>

Available in ePub^{WU}: April 2007

ePub^{WU}, the institutional repository of the WU Vienna University of Economics and Business, is provided by the University Library and the IT-Services. The aim is to enable open access to the scholarly output of the WU.

Generalized Maximally Selected Statistics



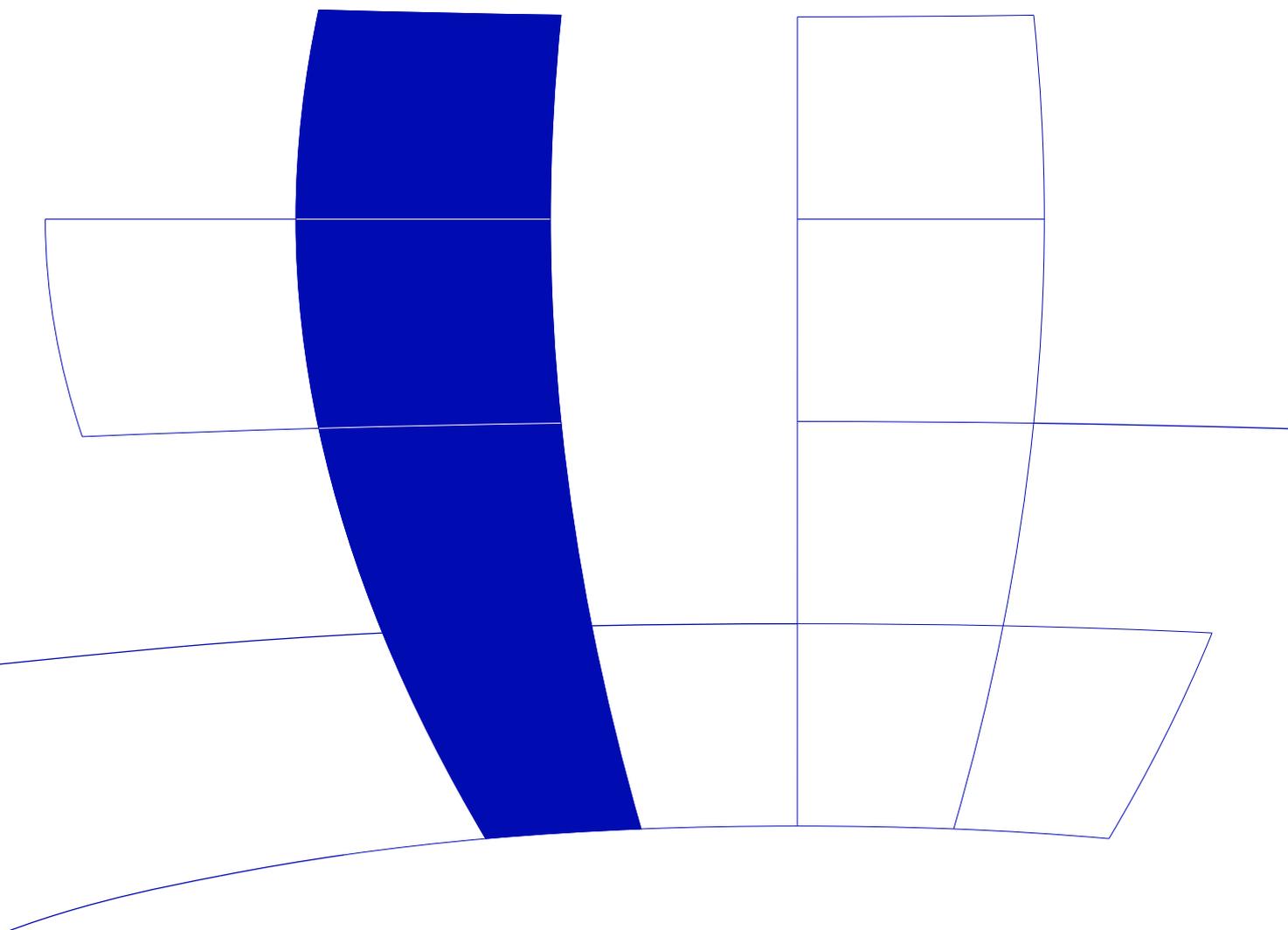
Torsten Hothorn, Achim Zeileis

Department of Statistics and Mathematics
Wirtschaftsuniversität Wien

Research Report Series

Report 52
April 2007

<http://statmath.wu-wien.ac.at/>



Generalized Maximally Selected Statistics

Torsten Hothorn

Ludwig-Maximilians-Universität München

Achim Zeileis

Wirtschaftsuniversität Wien

Abstract

Maximally selected statistics for the estimation of simple cutpoint models are embedded into a generalized conceptual framework based on conditional inference procedures. This powerful framework contains most of the published procedures in this area as special cases, such as maximally selected χ^2 and rank statistics, but also allows for direct construction of new test procedures for less standard test problems. As an application, a novel maximally selected rank statistic is derived from this framework for a censored response partitioned with respect to two ordered categorical covariates and potential interactions. This new test is employed to search for a high-risk group of rectal cancer patients treated with a neo-adjuvant chemoradiotherapy. Moreover, a new efficient algorithm for the evaluation of the asymptotic distribution for a large class of maximally selected statistics is given enabling the fast evaluation of a large number of cutpoints.

Keywords: asymptotic distribution, changepoint, conditional inference, maximally selected statistics.

1. Introduction

Dichotomization of variables measured at higher scale levels prior to model building is bad practice (Royston *et al.* 2006, among many others). It will result in loss of power and sophisticated regression models that adapt itself to the complexity of the regression problem at hand are widely available. However, simple regression models capturing step-shaped relationships between two variables are valuable for the implementation of scientific results into the real world: a one-parameter ‘good–poor’ or ‘high–low’ decision rule is attractive to practitioners because of its simplicity.

Such rules of thumb are frequently used to investigate new predictor variables for patient survival in oncology. Galon *et al.* (2006) estimate cutpoints for various characteristics of immune cells within colorectal tumor samples, such as type, density or location, with respect to their ability to differentiate between patients with good and poor prognosis. Buccisano *et al.* (2006) follow a similar approach, obtaining a threshold for residual leukemic cells in acute myeloid leukemia patients. The ability for expression levels of HER2 and co-amplified genes to predict breast cancer survival is investigated by Vinatzer *et al.* (2005) utilizing maximally selected log-rank statistics. Beyond applications in oncology, the identification of ecological thresholds is of increasing interest (see Huggett 2005), e.g., the estimation of cutpoints for habitat factors discriminating between ecosystems with low and high abundance of certain indicator species (Müller and Hothorn 2004).

Two questions arise from a statistical point of view. In a first step, we have to make sure that there is some relevant association between response and covariate and in a second step we want to estimate the ‘best’ cutpoint in order to approximate this relationship by a simple model. It is convenient to deal with both problems separately. The first problem needs to be addressed by a formal hypothesis test for the null hypothesis of independence between covariate (to be dichotomized) and response variable. A test with power against shift alternatives, i.e., departures from the null hypothesis where the distribution of the response variable varies between two groups of observations, is of special interest. Once we are able to reject the null hypothesis, we are

interested in the alternative which lead to the rejection, i.e., want to estimate a cutpoint or partition.

The first procedure of this kind, utilizing the maximum over multiple χ^2 statistics for 2×2 tables, was described by Miller and Siegmund (1982). Lausen and Schumacher (1992) derived an approximation for the asymptotical distribution of maximally selected rank statistics, extending the area of application to continuous and censored response variables. Betensky and Rabinowitz (1999) propose a maximally selected χ^2 test for nominal response variables measured at $k > 2$ levels and ordered categorical data (maximally selected Cochran-Armitage test). Finally, Rabinowitz and Betensky (2000) suggested a maximally selected McNemar's test. Lausen *et al.* (2004) extended maximally selected rank statistics to more than one covariate.

Based on the ideas underlying these established techniques, we suggest a new generalized class of maximally selected statistics that contains the tests above as special cases but also allows for direct construction of new test procedures for less standard test problems. For evaluating the distribution of the test statistics, a conditional inference approach is adopted by embedding the tests into the theory of permutation tests of Strasser and Weber (1999). This permits efficient computation of the complete correlation structure of the statistics to be maximized. For statistics derived from cutpoints, the correlations have a special product form which we exploit for evaluation of the conditional asymptotic distribution: A linear-time algorithm is described which enables the fast assessment of a large number of cutpoints and and improves significantly upon approximations for the asymptotic distribution currently in use.

Since no assumptions on the scale level of neither the covariate to be dichotomized nor the response variable are necessary, the class of generalized maximally selected statistics can not only be used to implement already published procedures but also extend the methodology to new areas of application. We do so by constructing a maximally selected log-rank statistic for a censored response partitioned with respect to two ordered categorical covariates and potential interactions. This new test is employed to search for a high-risk group determined by the T and N-category of rectal cancer patients.

2. Binary partitions and two-sample statistics

We are provided with independent and identically distributed observations $(\mathbf{Y}_i, \mathbf{X}_i)$ for $i = 1, \dots, n$ and are interested in testing the null hypothesis of independence of the response variable $\mathbf{Y} \in \mathcal{Y}$ and and covariate(s) $\mathbf{X} \in \mathcal{X}$

$$H_0 : D(\mathbf{Y}|\mathbf{X}) = D(\mathbf{Y})$$

against shift alternatives. That is, departures from the null hypothesis where the distribution of the response variable varies between two groups of observations are of special interest. Such binary partitions are defined in advance by p sets A_1, \dots, A_p partitioning the observations into two groups based on the \mathbf{X} measurements only. For ordered covariates \mathbf{X} , these sets are typically constructed via cutpoints, i.e., $A_j = \{\mathbf{X} | \mathbf{X} \leq \xi_j\}$. Here,

$$g_j(\mathbf{X}) = I(\mathbf{X} \in A_j)$$

denotes the indicator function partitioning the observations into two groups. Only partitions satisfying a size constraint $\sum_i g_j(\mathbf{X}_i) \in (n\varepsilon, n - n\varepsilon)$ for some fixed $\varepsilon \in (0, 0.5)$ are considered.

The two-sample problem associated with the j th binary partition can be tested using a linear statistic

$$\mathbf{T}_j = \text{vec} \left(\sum_{i=1}^n g_j(\mathbf{X}_i) h(\mathbf{Y}_i)^\top \right) \in \mathbb{R}^{q \times 1}$$

where $h : \mathcal{Y} \rightarrow \mathbb{R}^{q \times 1}$ is an *influence function* applied to the responses. The function $h(\mathbf{Y}_i) = h(\mathbf{Y}_i, (\mathbf{Y}_1, \dots, \mathbf{Y}_n))$ may depend on the full vector of responses $(\mathbf{Y}_1, \dots, \mathbf{Y}_n)$, however only in

a permutation symmetric way, i.e., the value of the function must not depend on the order in which $\mathbf{Y}_1, \dots, \mathbf{Y}_n$ appear. For example, with h being a rank transformation for a continuous response \mathbf{Y} , the linear statistic \mathbf{T}_j is the sum of the ranks for observations from A_j , i.e., equals the Wilcoxon-Mann-Whitney statistic. A joint linear statistic for all binary partitions is

$$\mathbf{T} = (\mathbf{T}_1, \dots, \mathbf{T}_p) = \text{vec} \left(\sum_{i=1}^n g(\mathbf{X}_i) h(\mathbf{Y}_i)^\top \right) \in \mathbb{R}^{pq \times 1}$$

including all p two-sample partitions, as defined by $g(\mathbf{X}) = (g_1(\mathbf{X}), \dots, g_p(\mathbf{X}))$, simultaneously for testing H_0 .

3. Standardization and estimation

To assess the partitions/cutpoints on a common scale, the corresponding statistics \mathbf{T}_j are typically standardized using some location and scale measure. Consequently, inference can be based on the maximally selected absolute standardized statistics and the best separating partition is the one for which the maximum is attained.

For obtaining valid estimates of the mean and covariance of \mathbf{T} , either a parametric model needs to be specified or non-parametric techniques can be employed, such as permutation or re-sampling approaches. Here, we adopt the latter and utilize the permutation test framework established by [Strasser and Weber \(1999\)](#). Thus, \mathbf{T} is standardized via its conditional expectation $\mu = \mathbb{E}(\mathbf{T}|S) \in \mathbb{R}^{pq \times 1}$ and covariance $\Sigma = \mathbb{V}(\mathbf{T}|S) \in \mathbb{R}^{pq \times pq}$, derived under H_0 by conditioning on all possible permutations S of the responses $\mathbf{Y}_1, \dots, \mathbf{Y}_n$. Closed-form expressions for μ and Σ are available, see [Strasser and Weber \(1999\)](#) or [Hothorn et al. \(2006\)](#).

The key step for the construction of a maximally selected statistic based on the multivariate linear statistic \mathbf{T} is its standardization utilizing the conditional expectation μ and the diagonal elements of the covariance matrix Σ : the maximum of the absolute values of the standardized linear statistic is used as test statistic

$$T_{\max} = \max \frac{|\mathbf{T} - \mu|}{\sqrt{\text{diag}(\Sigma)}}.$$

When the test statistic is large enough to indicate a deviation from the null hypothesis we are interested in determining the partition with largest standardized statistic: the best separating partition A_{j^*} is the one for which the maximum is attained, i.e., for which the absolute value of the standardized statistic \mathbf{T}_{j^*} equals T_{\max} .

4. Inference

For testing H_0 the conditional distribution of T_{\max} given all permutations of the responses is used as reference distribution. Ideally, we want to compute the exact conditional distribution but this is only possible in special small sample situations ([Boulesteix 2006a,b](#)). Conditional Monte-Carlo methods can be used to approximate the exact conditional distribution rather easily (evaluate the test statistic T_{\max} for randomly shuffled responses \mathbf{Y} and compute the proportion of random permutations where T_{\max} exceeds the test statistic for the permutation $1, \dots, n$).

Moreover, the exact conditional distribution can be approximated by its limiting distribution. For $n \rightarrow \infty$ the distribution of the multivariate linear statistic \mathbf{T} tends to a multivariate normal distribution with mean μ and covariance matrix Σ ([Strasser and Weber 1999](#), Theorem 3). Thus, in order to approximate $\mathbb{P}(T_{\max} > c)$ we have to evaluate the probability $\mathbb{P}(\max(|Z_1|, \dots, |Z_{pq}|) > c)$ for standard normal random variables Z_1, \dots, Z_{pq} with correlation matrix $\mathbf{R} = \text{cor}(\Sigma)$ and some $c > 0$. The computation of this probability is possible using Quasi-Monte-Carlo methods ([Genz 1992](#)) for moderate dimensions ($pq < 100$, say) but remains infeasible for higher dimensions.

However, for the most important case of statistics maximally selected over cutpoints induced by an ordered covariate \mathbf{X} and an ordered or binary response \mathbf{Y} , the distribution of the maximum of standard normal variables can be evaluated numerically by an algorithm with computing time being linear in the number of cutpoints considered as will be shown in the following.

Let $A_j = (-\infty, \xi_j]$ with $\xi_j < \xi_k$ for $1 \leq j < k \leq p$ denote the partitioning sets and let $q = 1$ (i.e., ordered or binary response variable). Then, the correlation between \mathbf{T}_j and \mathbf{T}_k is given by

$$\rho_{j,k} = \frac{\Sigma_{j,k}}{\sqrt{\Sigma_{j,j}\Sigma_{k,k}}} = \sqrt{\frac{(n - \sum_i g_k(\mathbf{X}_i)) \sum_i g_j(\mathbf{X}_i)}{(n - \sum_i g_j(\mathbf{X}_i)) \sum_i g_k(\mathbf{X}_i)}}.$$

It follows that the correlation matrix is completely determined by the subdiagonal elements $\rho_{j,j-1}, j = 2, \dots, p$ and it holds that

$$\rho_{1,k} = \prod_{j=2}^k \rho_{j,j-1}.$$

With $\mathbf{v} = (\rho_{1,1}, \dots, \rho_{1,p})$ the lower triangular part of the correlation matrix \mathbf{R} can be written as $\mathbf{v}(1/\mathbf{v})^\top$ and it follows (from Meurant 1992, Section 2.1) that the inverse \mathbf{R}^{-1} of the correlation matrix is a tridiagonal symmetric band matrix:

$$\mathbf{R}^{-1} = \begin{pmatrix} r_{1,1} & r_{1,2} & 0 & 0 & \dots & 0 \\ r_{1,2} & r_{2,2} & r_{2,3} & 0 & \dots & 0 \\ 0 & r_{2,3} & r_{3,3} & r_{3,4} & \dots & 0 \\ 0 & 0 & r_{3,4} & r_{4,4} & \ddots & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & r_{p-1,p-1} & r_{p-1,p} \\ 0 & 0 & 0 & 0 & r_{p-1,p} & r_{p,p} \end{pmatrix}.$$

We have to evaluate the probability that any of $|Z_1|, \dots, |Z_p|$ exceeds $c > 0$, i.e., the probability

$$\mathbb{P}(T_{\max} > c) = 1 - \frac{1}{\sqrt{|\mathbf{R}|}(2\pi)^p} \int_{-c}^c \exp\left(-\frac{1}{2} \mathbf{z}^\top \mathbf{R}^{-1} \mathbf{z}\right) d\mathbf{z}.$$

Because of the special band structure of \mathbf{R}^{-1} the quadratic form $\mathbf{z}^\top \mathbf{R}^{-1} \mathbf{z}$ simplifies to

$$\mathbf{z}^\top \mathbf{R}^{-1} \mathbf{z} = r_{1,1}z_1^2 + 2r_{2,1}z_1z_2 + r_{2,2}z_2^2 + \dots + 2r_{p,p-1}z_pz_{p-1} + r_{p,p}z_p^2.$$

and one can make use of this structure when evaluating the multivariate normal distribution numerically (Genz and Kahaner 1986). With $\phi(z) = \exp(-z/2)$ we have

$$\int_{-c}^c \phi(\mathbf{z}^\top \mathbf{R}^{-1} \mathbf{z}) d\mathbf{z} = \int_{-c}^c \phi(r_{1,1}z_1^2) \int_{-c}^c \phi(2r_{2,1}z_1z_2 + r_{2,2}z_2^2) \int_{-c}^c \dots \int_{-c}^c \phi(2r_{p,p-1}z_pz_{p-1} + r_{p,p}z_p^2) d\mathbf{z}$$

and with recursively defined functions f_j ($j = 2, \dots, p+1$) with

$$f_j(z) = \int_{-c}^c \phi(2r_{j,j-1}z\tilde{z} + r_{j,j}\tilde{z}^2) f_{j+1}(\tilde{z}) d\tilde{z} \quad \forall j = 2, \dots, p$$

and $f_{p+1}(z) \equiv 1$, the above integral can be re-formulated recursively:

$$\mathbb{P}(T_{\max} > c) = 1 - \frac{1}{\sqrt{|\mathbf{R}|}(2\pi)^p} \int_{-c}^c \phi(\mathbf{z}^\top \mathbf{R}^{-1} \mathbf{z}) d\mathbf{z} = 1 - \frac{1}{\sqrt{|\mathbf{R}|}(2\pi)^p} \int_{-c}^c \phi(r_{1,1}z^2) f_2(z) dz.$$

This integral can be evaluated numerically in $O(p)$ starting with f_p utilizing the techniques described by [Miwa et al. \(2000\)](#): For a two-dimensional grid of $z \in [-c, c]$ and $\tilde{z} \in [-c, c]$ values, the function f_j is evaluated and aggregated over \tilde{z} only, yielding values of $f_j(z)$ for a grid of z values. These values are then re-used when computing f_{j-1} .

5. Applications and illustration

Maximally selected statistics as described in Section 2 can be applied to covariates \mathbf{X} and responses \mathbf{Y} measured at arbitrary scales; appropriate influence functions h for nominal, ordered, numeric, censored and multivariate response variables are given in the sequel (further details can be found in [Hothorn et al. 2006](#)), followed by a description of how to partition the covariate space for nominal, ordered and multivariate covariates and the derivation of a novel maximally selected statistic.

For nominal responses, h is typically a straightforward dummy coding of the k factor levels (where it is sufficient to code for the first $k - 1$ levels only). For ordinal responses, scores are usually attached to each of the k levels and h is the vector of those numeric scores. Many possible influence functions are available for discrete or continuous covariates: The identity transformation is the most natural choice; square root, log, or rank transformations are possible as well. When a numeric response is censored, the log-rank transformation or Savage scores can be applied. For a multivariate response, possibly consisting of variables with different scale levels, the influence function h is a combination of influence functions appropriate for any of the univariate response variables as suggested above. In the presence of a grouping of the observations into independent blocks, only permutations within blocks are eligible and the conditional expectation and covariance matrix need to be computed separately for each block. Therefore, it is easily possible to take a block randomization scheme in a randomized clinical trial into account.

The most important situation of a univariate and at least ordinaly measured covariate \mathbf{X} leads to partitions, and thus functions g_j , induced by cutpoints defined by the realizations $\mathbf{X}_1, \dots, \mathbf{X}_n$. More specifically, $A_j = (-\infty, \xi_j]$, where ξ_j is the j th element of the increasingly sorted unique realizations of \mathbf{X} (the sample size constraints apply, of course). Thus, having identified the best separating partition A_{j^*} , the estimated cutpoint is ξ_{j^*} . For nominal covariates, all 2^{k-1} binary partitions of the k levels that meet the sample size constraints are taken into account. For multiple covariates, we simply look at all binary partitions induced by all covariates simultaneously.

This flexible framework can now be utilized to implement the conditional versions of already published maximally selected statistics and to develop new maximally selected statistics for other applications. For binary responses and ordered covariates, the statistic T_{\max} is equivalent to a maximally selected χ^2 statistic ([Miller and Siegmund 1982](#); [Kozioł 1991](#)). For nominal responses with $k > 2$ levels, the T_{\max} statistic is the maximum over the maximum of standardized $k \times 2$ tables, an alternative to maximally selected χ^2 statistics for larger tables ([Betensky and Rabinowitz 1999](#)). For equidistant scores $1, \dots, k$ attached to an ordered response, the statistic T_{\max} is equivalent to a maximally selected Cochran-Armitage statistic ([Betensky and Rabinowitz 1999](#)), however, no restrictions to the choice of scores apply here. For numeric or censored responses, T_{\max} corresponds to maximally selected rank statistics ([Lausen and Schumacher 1992](#)). In the extreme case of each observation being a block in repeated measurements, our test statistic T_{\max} for binary responses corresponds to maximally selected McNemar's statistics ([Rabinowitz and Betensky 2000](#)). The special case of maximally selected rank statistics for multiple covariates has first been studied by [Lausen et al. \(2004\)](#).

Going beyond these established techniques is also easily possible in less standard situations: New tests can be constructed by choosing an influence function $h(\cdot)$ determined by the scale level of the response, and selecting a set of potential partitions $g(\cdot)$ determined by the available covariates. As an illustration, we attempt to identify a high- and low-risk groups of rectal cancer patients—treated with a neo-adjuvant chemoradiotherapy regime—by differentiating them with respect to their combination of pathological T and N category. Survival times of $n = 349$ patients from the preoperative arm of the CAO/ARO/AIO-94 trial ([Sauer et al. 2004](#)) are under test (48 patients

T category	N category			Total
	N0	N1	N2 + N3	
ypT0	36	4	0	40
ypT1	16	5	0	21
ypT2	90	14	7	111
ypT3	107	33	29	169
ypT4	3	1	4	8
Total	252	57	40	349

Table 1: Pathological T and N category of 349 rectal cancer patients treated with a preoperative chemoradiotherapy.

with distant metastases found at surgery were excluded from this analysis). Therefore, we propose a new maximally selected statistic for a censored response and two ordered covariates with potential interactions. Log-rank scores are used as influence function h for the censored response and the potential partitions g are constructed from all combinations of the five T and three N categories. As both categories are ordered, only those partitions are used which are ordered in T given N and vice versa. This yields 194 potential partitions, 187 of which meet the sample size constraints. The maximum of the absolute values of the corresponding 187 standardized statistics is 8.69 (p -value smaller than 0.0001). The partition chosen by the algorithm identifies all patients with N category N2 or N3 as being under high risk and almost all patients from N0 and N1 as being under low risk. As an exception, a single patient with ypT4 and N1 is assigned to the high risk group as well—whether or not this decision is sensible or results from random variation cannot be judged based on one observation alone. Figure 1 depicts Kaplan-Meier estimates of the survival times in the two risk groups.

6. Discussion

Maximally selected statistics for the estimation of simple cutpoint models have been in use since many years. Many researchers appreciate a model that is easy to communicate and easy to implement in practical situations. Of course, the tradeoff between between simplicity and accuracy has to be carefully investigated.

The new class of generalized maximally selected statistics based on the conditional inference framework of [Strasser and Weber \(1999\)](#) allows for a unified treatment of different kinds of maximally selected statistics. Test procedures from this framework can be adapted to new test problems by specifying an influence function h , suitable for the scale level of the response, and setting up a set of potential partitions g determined from the available covariates. As the number of potential partitions can become large, efficient algorithms are required for evaluating the distribution of the maximum statistic. For partitions based on cutpoints, we provide such an algorithm that computes the asymptotic distribution in linear time by exploiting the special product structure of the correlation matrix and utilizing numerical integration techniques.

The implementation of (known and newly designed) maximally selected statistics only requires the specification of the binary partitions, via a function g , and a problem-specific influence function h . Linear statistics \mathbf{T} and the test statistic T_{\max} can be computed in the R system for statistical computing ([R Development Core Team 2007](#)) utilizing the function `maxstat_test()` from the add-on package `coin` ([Hothorn et al. 2006, 2007](#)). The distribution of T_{\max} can be approximated by its asymptotic distribution or by Monte-Carlo methods (also in the presence of a grouping of the observations into independent blocks) readily available from the same package.

In summary, a unified treatment of maximally selected statistics for nominal, ordered, discrete and continuous numeric, censored and multivariate response variables as well as nominal, ordered and multivariate covariates to be dichotomized is now possible both conceptually and practically.

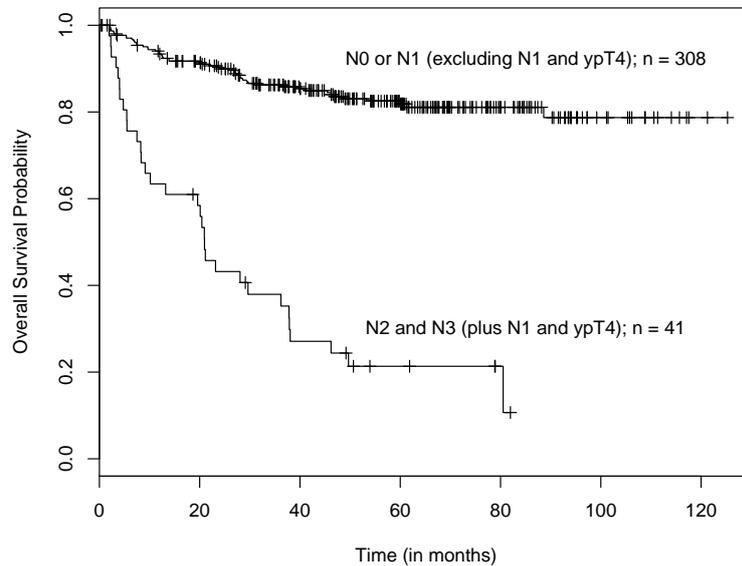


Figure 1: Survival times of rectum cancer patients in two risk groups identified by a maximally selected log-rank statistic.

Acknowledgments

We would like to thank A. Genz for pointing out the special product structure of the covariance matrix. C. Rödel kindly gave us permission to use the CAO/ARO/AIO-94 trial data. The work of T. Hothorn was supported by Deutsche Forschungsgemeinschaft (DFG) under grant HO 3242/1-3.

References

- Betensky RA, Rabinowitz D (1999). “Maximally Selected χ^2 Statistics for $k \times 2$ Tables.” *Biometrics*, **55**, 317–320.
- Boulesteix AL (2006a). “Maximally Selected Chi-square Statistics and Binary Splits of Nominal Variables.” *Biometrical Journal*, **48**, 838–848.
- Boulesteix AL (2006b). “Maximally Selected Chi-square Statistics for Ordinal Variables.” *Biometrical Journal*, **48**, 451–462.
- Buccisano F, Maurillo L, Gattei V, Del Poeta G, Del Principe MI, Cox MC, Panetta P, Consalvo MI, Mazzone C, Neri B, Ottaviani L, Fraboni D, Tamburini A, Lo-Coco F, Amadori S, Venditti A (2006). “The Kinetics of Reduction of Minimal Residual Disease Impacts on Duration of Response and Survival of Patients with Acute Myeloid Leukemia.” *Leukemia*, **20**, 1783–1789.
- Galon J, Costes A, Sanchez-Cabo F, Kirilovsky A, Mlecnik B, Lagorce-Pages C, Tosolini M, Camus M, Berger A, Wind P, Zinzindohoue F, Bruneval P, Cugnenc PH, Trajanoski Z, Fridman WH, Pages F (2006). “Type, Density, and Location of Immune Cells Within Human Colorectal Tumors Predict Clinical Outcome.” *Science*, **313**, 1960–1964.

- Genz A (1992). “Numerical Computation of Multivariate Normal Probabilities.” *Journal of Computational and Graphical Statistics*, **1**, 141–149.
- Genz A, Kahaner DK (1986). “The Numerical Evaluation of Certain Multivariate Normal Integrals.” *Journal of Computational and Applied Mathematics*, **16**, 255–258.
- Hothorn T, Hornik K, van de Wiel MA, Zeileis A (2006). “A Lego System for Conditional Inference.” *The American Statistician*, **60**(3), 257–263.
- Hothorn T, Hornik K, van de Wiel MA, Zeileis A (2007). **coin**: *Conditional Inference Procedures in a Permutation Test Framework*. R package version 0.5-2, URL <http://CRAN.R-project.org/>.
- Huggett AJ (2005). “The Concept and Utility of ‘Ecological Thresholds’ in Biodiversity Conservation.” *Biological Conservation*, **124**, 301–310.
- Koziol JA (1991). “On Maximally Selected Chi-square Statistics.” *Biometrics*, **47**, 1557–1561.
- Lausen B, Hothorn T, Bretz F, Schumacher M (2004). “Assessment of Optimal Selected Prognostic Factors.” *Biometrical Journal*, **46**(3), 364–374.
- Lausen B, Schumacher M (1992). “Maximally Selected Rank Statistics.” *Biometrics*, **48**, 73–85.
- Meurant G (1992). “A Review on the Inverse of Symmetric Tridiagonal and Block Tridiagonal Matrices.” *SIAM Journal on Matrix Analysis and Applications*, **13**(3), 707–728.
- Miller R, Siegmund D (1982). “Maximally Selected Chi Square Statistics.” *Biometrics*, **38**, 1011–1016.
- Miwa T, Hayter AJ, Liu W (2000). “Calculations of Level Probabilities for Normal Random Variables with Unequal Variances with Applications to Bartholomew’s Test in Unbalanced One-Way Models.” *Computational Statistics & Data Analysis*, **32**, 17–32.
- Müller J, Hothorn T (2004). “Maximally Selected Two-Sample Statistics as a New Tool for the Identification and Assessment of Habitat Factors with an Application to Breeding-Bird Communities in Oak Forests.” *European Journal of Forest Research*, **123**, 219–228.
- Rabinowitz D, Betensky RA (2000). “Approximating the Distribution of Maximally Selected McNemar’s Statistics.” *Biometrics*, **56**, 897–902.
- R Development Core Team (2007). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.
- Royston P, Altman DG, Sauerbrei W (2006). “Dichotomizing Continuous Predictors in Multiple Regression: A Bad Idea.” *Statistics in Medicine*, **25**, 127–141.
- Sauer R, Becker H, Hohenberger W, Rödel C, Wittekind C, Fietkau R, Martus P, Tschmelitsch J, Hager E, Hess CF, Karstens JH, Liersch T, Schmidberger H, Raab R (2004). “Preoperative Versus Postoperative Chemoradiotherapy for Rectal Cancer.” *New England Journal of Medicine*, **351**, 1731–1740.
- Strasser H, Weber C (1999). “On the Asymptotic Theory of Permutation Statistics.” *Mathematical Methods of Statistics*, **8**, 220–250. Preprint available from http://epub.wu-wien.ac.at/dyn/openURL?id=oai:epub.wu-wien.ac.at:epub-wu-01_94c.
- Vinatzer U, Dampier B, Streubel B, Pacher M, Seewald MJ, Stratowa C, Kaserer K, Schreiber M (2005). “Expression of HER2 and the Coamplified Genes GRB7 and MLN64 in Human Breast Cancer: Quantitative Real-Time Reverse Transcription-PCR as a Diagnostic Alternative to Immunohistochemistry and Fluorescence In Situ Hybridization.” *Clinical Cancer Research*, **11**, 8348–8357.