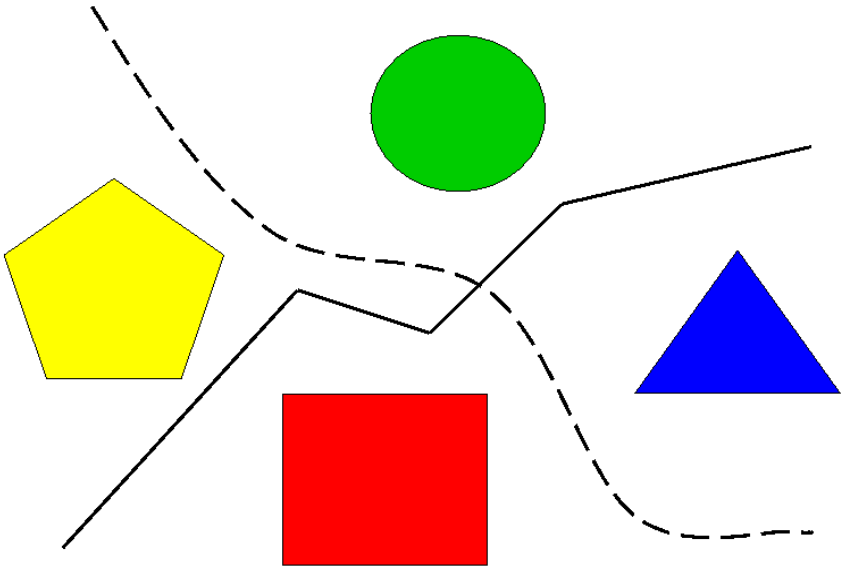
technische universität
dortmund

Fakultät Statistik

8. Herbstkolloquium
des Graduiertenkollegs

"Statistische Modellbildung"



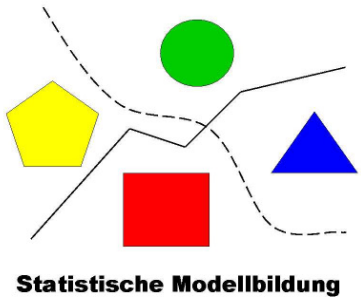
Statistische Modellbildung

Zu diesem Kolloquium wird eingeladen

Freitag / Samstag, 25./26. November 2011

Veranstaltungsort:

TRYP Hotel Dortmund
Emil-Figge Straße 41, 44227 Dortmund, T: 0231 9705-0



8. Herbstkolloquium des Graduiertenkollegs "Statistische Modellbildung"

Freitag, 25. November 2011

Vortragsprogramm I

14:45 **Begrüßung**
Joachim Kunert

15:00 **Olaf Schoffer**
University Cancer Center Dresden, Germany
Which Impact has the demographic change on the structure of patients in German hospitals?

15:45 **Gero Szepannek**
ORCLUS Subspace Clustering using R

16:30 **Jens-Peter Kreiß**
TU Braunschweig, Germany
Bootstrapping Stationary and Locally Stationary Processes

17:15 **Pieter M. Kroonenberg**
Leiden University, The Netherlands
Three-way correspondence analysis illustrated with an ecological example from the sea bed

Diskussion zu den Projektbereichen

19:00 **Posterausstellung (Moderation - Dirk Husmeier):**
Präsentation der Dissertationsprojekte im Kolleg, Diskussion in Arbeitsgruppen

Samstag, 26. November 2011

Vortragsprogramm II

9:00 **Pi-Wen Tsai**
National Taiwan Normal University, Taiwan
Q_B-optimal saturated two-level main effects designs

9:45 **Hugo Maruri-Aguilar**
Queen Mary University of London, U.K.
Sequential screening with elementary effects

10:30 ***Pause***

Vortragsprogramm III

11:00 **Heiko Grossmann**
Queen Mary University of London, U.K.
AutomaticANOVA - a package for automatically carrying out the analysis of variance

11:45 **Nancy Flournoy**
University of Missouri, Columbia, MO, U.S.A.
Dose-Finding Experiments in Clinical Trials

12:30 ***Pause***

13:15 ***Gruppenfoto***

Vortragsprogramm IV

13:30 **Tina Müller**
Bayer Pharma AG, Germany
Analyzing SCORE data – Small Counts of Rough Endpoints

14:15 **Holger Schwender**
TU Dortmund University, Germany
Testing SNPs and Gene-Environment Interactions in Case-Parent Trio Studies Based on Exact Analytic Parameter Estimation

15:00 ***Kaffeetrinken, Abschlussbesprechung und Diskussion***

ABSTRACTS

Dose-Finding Experiments in Clinical Trials

Nancy Flournoy
University of Missouri, Columbia, MO, U.S.A.

Consider two situations. In one, toxicity increases with dose. In the second, one considers efficacy in addition, and assumes efficacy increases with dose – so except at the extremes, the probability of efficacy without toxicity (success) will increase up to a point at which the toxicity is great enough to cause it to turn down. In the first case, typically, one seeks to identify a dose with a prescribed toxicity rate; in the second case, one seeks to identify the dose that maximizes the probability of success. These goals can be posed in terms of estimation or dose selection, as a finite number of doses are typically permitted. Two common classes of procedures that differ in many fundamental ways are described: up-and-down designs and best intention designs.

AutomaticANOVA - a package for automatically carrying out the analysis of variance

Heiko Grossmann
Queen Mary University of London, U.K.

Abstract: Bailey's theory for the analysis of variance of orthogonal designs covers a wide range of models with complicated crossing and nesting structures, where block factors have random and treatment factors have fixed effects.
This talk shows how the theory can be implemented in a user-friendly software package which automatically deduces the model from the design. In particular, no model formula needs to be specified for the analysis.
As time does not allow me to cover the theory, the presentation will focus on examples and a demonstration of the AutomaticANOVA package, but pointers to the relevant literature will be given.

Bootstrapping Stationary and Locally Stationary Processes

Jens-Peter Kreiß
TU Braunschweig, Germany

We propose a nonparametric method to bootstrap for stationary and locally stationary processes which combines a time domain wild bootstrap approach with a nonparametric frequency domain approach. The method generates pseudo-time series which mimic (asymptotically) correctly the (local) second and to the necessary extent the fourth order moment structure of the underlying process. Thus it can be applied to approximate the distribution of statistics of interest that are based on observations of the (locally) stationary process. We prove a bootstrap central limit theorem for a general class of statistics that can be expressed as functionals of the periodogram and accordingly the preperiodogam.

Three-way correspondence analysis illustrated with an ecological example from the sea bed

Pieter M. Kroonenberg
Leiden University, The Netherlands

Three-way correspondence analysis will be introduced as an extension of standard two-way correspondence analysis. The latter technique uses a generalisation of the singular value decomposition for creating coordinates from proportions. It will be demonstrated that three-way correspondence analysis uses a generalisation of a type of three-way singular value decomposition in the same way (see e.g. Kroonenberg, 2008).
The technique will be illustrated with the data from an experiment which was conducted at the Norwegian Institute for Water Research using sediment collected from Bjørnhordenbukta, a small sheltered bay in Oslofjørd. Ninety-eight units of homogenized sediment were subjected to one of seven levels of organic enrichment, combined with one of seven different frequencies of physical disturbance, each replicated once (Widdicombe & Austin, 2001). The effect on the biodiversity of the different levels of the factors and their interaction was examined via graphical displays resulting from three-way correspondence analysis using the program suite 3WayPack (Kroonenberg & De Roo, 2010).

Kroonenberg, P. M. (2008). Applied multiway data analysis. Hoboken, NJ: Wiley.

Kroonenberg, P. M., De Roo, Y. (2010). 3WayPack. A program suite for three-way analysis. Leiden: The Three-Mode Company.

Widdicombe, S. & Austin, M. C. (2001). The interaction between physical disturbance and organic enrichment: An important element in structuring benthic communities. Limnology & Oceanography, 46, 1720-1733.

Sequential screening with elementary effects

Hugo Maruri-Aguilar
Queen Mary University of London, U.K.

Keywords: Elementary effects, space filling, sequential
The Elementary Effects (EE) method is a simple but effective screening strategy for computer models. Starting from a number of initial points, the method creates random trajectories to then estimate factor effects. In turn, those estimates are used for factor screening. Recent research advances have enhanced the performance of the elementary effects method and the projections of the resulting design.

The presentation concentrates on a new proposal which turns the elementary effects method into a sequential design strategy. After describing the methodology, some examples are given and compared against the traditional EE method.

Analyzing SCORE data - Small Counts of Rough Endpoints

Tina Müller
Bayer Pharma AG, Berlin

In many environments, experiments or studies with small sample sizes are either economically desired or practically unavoidable. The need for suitable statistical analysis methods which also give interpretable results is obvious.
But how much improvement can we gain by using suitable exact methods which might have considerable disadvantages in comparison to asymptotic strategies?

In the given example, two groups of small sample size are compared based on what we call score data: Experts assess the outcome of a specific experiment with ordered score values. The differentiation between different score values lies within the scope of the expert.
As the scores are not normally distributed, rank-based methods are used.
For the group comparison, a suitable effect measure is “Probability that the score value of a randomly drawn observation from group 1 is smaller than from group 2”, which is closely related to the Wilcoxon test statistics. The test decision can be based on an exact or an asymptotic test statistics. Additionally, several confidence intervals for the effect measure can be constructed.
The exact test is preferred, the exact confidence interval, however, is not always available. The question of practical interest is if and how often the significance statements from the different tests and confidence intervals contradict each other.
We will show results of a real-world data set as well as results of a simulation study that is based on the given original data.

Which impact has the demographic change on the structure of patients in German hospitals?

Olaf Schoffer[1,2], Andreas Werblow[3], Mike Kühne[3]
[1]Cancer Epidemiology, University Cancer Center Dresden;
2Research Data Centre, Statistical Office of the Free State of Saxony; [3]Health Economics Centre, TU Dresden

Background: The demographic change has an impact on various areas, not least on the German health care system. Thus, there is substantial interest in forecasting the future development, particularly in the area of inpatient care. But many predictions represent only a static projection of the current use of health services based on the predicted population development. This approach appears to be insufficient. It takes into account neither exogenous influences, nor the dynamics of health care utilization and patient mobility.
Methods: We are using the official hospital statistics from 2000 to 2008. On the one hand, diagnosis groups are determined, which are particularly characteristic of certain age groups of patients. On the other hand, all considered hospitals are examined for significant streams of patient mobility. Thus, each institution is assigned to a specific hospital catchment area. This leads to an accurate linkage between the population and the utilization of inpatient care services in a model (for whole patients, but also specific to certain diagnoses). This is based on a micro-economic production model of a hospital (Cobb Douglas production function). Supplemented by exogenous factors, this model is estimated using panel regression. Furthermore, it allows quantification and verification of the assumed relationships.
Results: According to the estimated model components of the production function (medical and non-medical staff, number of beds) as well as demographic factors affect the number of cases significantly in the expected direction. Important are also exogenous effects, such as urbanization, mobility of patients within the catchment area, competition and outsourcing. In particular, the average income in the catchment area of hospitals plays an important role. Thus, a 1% increase in deflated disposable income would lead to an average reduction of 0.24% of the total number of cases. Apart from the development of the total number of cases, the diagnosis specific considerations show shifts between the numbers of cases of individual diagnostic groups. For example, the total number of cases decreases at the expected increase in the age group 65-79. In the same circumstance the number of cases in the diagnostic group "ischaemic heart diseases" (ICD I20-I25) increases and the number of cases in the diagnostic group "malignant neoplasms" (ICD C00-C97) remains almost unaffected.
Conclusions & Perspectives: The demographic change affects both the number of inpatient cases and the patient structure regarding illness and age - in particular with consideration of the changing supply structures. Therefore, there is a great significance of diagnosis specific analysis in investigations of the dependency structure of the inpatient health care utilization. The consideration of the relevant exogenous influences is important as well. In further research also such diagnosis groups with increasing magnitude in the future should be considered. In addition, the interaction between production factors and demographic variables should be considered in detail.

Testing SNPs and Gene-Environment Interactions in Case-Parent Trio Studies Based on Exact Analytic Parameter Estimation

Holger Schwender
TU Dortmund University, Dortmund, Germany

Alternatively to population-based case-control studies, case-parent trio designs considering children affected by a disease and their parents are frequently used to test for association between single nucleotide polymorphisms (SNPs) and the disease. One of the most prominent and powerful procedures for detecting SNPs associated with disease in such studies is the genotypic transmission/disequilibrium test (gTDT), which is equivalent to a Wald test based on a conditional logistic regression model. Usually, the parameters of such a model need to be estimated numerically by an iterative approach. In particular in genome-wide association studies, this procedure can thus be time-consuming, as this model needs to be fitted for hundreds of thousands or millions of SNPs.

In my presentation, I will show that for all typically considered genetic modes of inheritance (i.e. the additive, the dominant, and the recessive mode) closed-form solutions for the parameter estimates in the conditional logistic regression models, and thus, for the test statistics of the gTDT exist. Using these analytic solutions instead of the conventionally used approach for iteratively maximizing the conditional likelihood reduces the computing time from several hours to a few minutes, as a genome-wide application of the gTDT will show. These closed-form solutions can also be adapted to gTDTs for testing whether interactions between SNPs and binary environmental factors are associated with disease. Since the genetic mode of inheritance underlying an association between a SNP and a disease is typically unknown, I will finally present a MAX test based on the gTDT statistics for an additive, a dominant, and a recessive effect as well as a procedure that makes it feasible to compute genome-wide permutation-based p-values for the individual gTDTs and the MAX test.

ORCLUS Subspace Clustering using R

Gero Szepannek

The challenge of clustering data often grows with an increasing dimension. A brief introduction of the idea of subspace clustering will be given in terms of one of its most common representatives: the 'arbitrarily ORiented projected CLUSter generation'. The algorithm is presented from a practical point of view: its implementation in R, its parameterization as well as its performance under several data conditions.

Q_B-optimal saturated two-level main effects designs

Pi-Wen Tsai
National Taiwan Normal University, Taiwan

In this talk we study saturated two-level main effects designs which are commonly used for screen experiments, and we provide a general framework that incorporates experimenters' prior beliefs into the selection of saturated two-level main effects designs. We show that under the sets of priors with more weights on models of one or two active factors, p-efficient designs is recommended; when models with more parameters are of interest, D-optimal designs would be better. Additionally, we present a new class of designs which can be found between these two designs under different sets of priors. The way in which the choice of designs depends on experimenters' prior beliefs will be demonstrated for the cases when N=2 mod 4. The construction of the new class of designs using the conference matrix is presented.

This is a joint work with Professor Steven G. Gilmour.