30th IOPS Summer Conference, 18-19 June 2015

Dept. Methodology & Statistics, Social and Behavioral Sciences, Utrecht University,
Bestuursgebouw, Heidelberglaan 8, 3584 CS Utrecht
Zaal Van Lier & Eggink

Program prior to conference - Thursday 18 June 2015

10.00 - 13.00	**IOPS Board meeting (lunch incl)** *Sjoerd Groenmangebouw, Padualaan 14, de Uithof, 3584 CH Utrecht. Ruimte B1.13*
11.30 - 12.00	**IOPS PhD student meeting.** *Bestuursgebouw, Heidelberglaan 8, 3584 CS Utrecht Zaal Van Lier & Eggink*
12.00 - 13.00	**Arrange your own lunch**

Program Thursday 18 June 2015

Bestuursgebouw, Heidelberglaan 8, 3584 CS Utrecht. Zaal Van Lier & Eggink

13.00 - 13.10	**Registration**
13.10 - 13.30	**Official opening** by **Rob Meijer** and welcome by **Herbert Hoijtink**
13.30 – 13:55	Presentation **Noémi Schuurman:** *Multilevel autoregressive modeling with measurement error.* Discussant: Kirsten Bulteel
13.55 - 14.20	Presentation **Xin Gu:** *Controlling for error probabilities when using default Bayes factors.* Discussant: Florian Boeing-Messing
14.20 - 14.45	Presentation **Dereje Gudicha:** *Power Analysis for the Likelihood Ratio Test in Latent Markov Models: Short-cutting the bootstrap p-value based method* Discussant: Erwin Nagelkerke
14.45 – 15.30	**Break / Poster Session**
	Posters **Jedelyn Cabrieto:** *Comparing the performance of non-parametric change point detection methods for capturing response concordance*
	Jolien Cremers: *Bayesian Longitudinal Modelling of Circular Data: Application and Interpretation*
	Laura Dekkers *Decision Making in a Sequential Context :A Drift Diffusion Model Study*

Dino Dittrich
Bayesian Analysis of the Network Autocorrelation Model

Paulette Flore
Publication bias in practice: The case of Stereotype Threat

Frank Bais
Intercoder Reliability: Coding Surveys on their Item Characteristics for Constructing Questionnaire Profiles

15.30 – 16.30 Invited speaker
Eric-Jan Wagenmakers:
JASP: A Fresh Way to do Bayesian Hypothesis Testing
Discussant: Jesper Tijmstra

16.30 - 16.55 Presentation
Joke Heylen:
Two-mode K-Spectral Centroid analysis for studying multivariate dynamical processes
Discussant: Camelia Minica

16.55 - 17.30 **Plenary meeting IOPS staff and students**

Presentation IOPS Best Paper Award 2014

17.30 **Conference dinner**
the Basket, Genevelaan 8, de Uithof, 3584 CC Utrecht
Offered to you by IOPS Utrecht

-

Program Friday 19 June 2015

Bestuursgebouw, Heidelberglaan 8, 3584 CS Utrecht. Zaal Van Lier & Eggink

09.00 – 09.30 **Registration**

09.30 – 09.55 Presentation
Maria Bolsinova
Testing conditional independence and modeling conditional dependence between response time and accuracy
Discussant: Abe Hofman

09.55 – 10.20 Presentation
Ruslan Jabrayilov
Comparison of classical and modern testing methods in change assessment
Discussant: Paulette Flore

10.20– 10.45 Presentation
Pieter Oosterwijk
Reliability estimation and coefficient alpha revisited
Discussant: Jurian Meijering

10.45 – 11.30 **Break / poster session**

Posters
Robert Hillen
A Critical Assessment of Taxometrics

Jurian Meijering
The Delphi method: methodological issues and its application to the development of rankings

Camelia Minica
Family-based genetic association analysis: methods and applications to addiction phenotypes

Kees Mulder
Extending Bayesian analysis of circular data to comparison of multiple groups

Aniek Sies
Comparing four methods for estimating tree-based treatment regimes

Mariëlle Zondervan-Zwijnenburg
Development and evaluation of a belief elicitation procedure

11.30 – 11.55 Presentation
Marije Fagginger Auer
Exploring relations between instruction, strategies and achievement in mathematics: latent variable modeling of large-scale assessment data and experiments
Discussant: Michèle Nuijten

11.55 – 12.20 Presentation
Xinru Li
Meta---Cart: Integrating Classification and Regression Trees into Meta- analysis
Discussant: Robbie van Aert

12.20 – 13.10 Invited speaker
Daniel Oberski
Model fit evaluation by sensitivity analysis
Discussant: Dylan Molenaar

13.10 – 13.15 **Presentation IOPS Best Poster Award 2014**

Closing conference by **Rob Meijer**

---

presentation June 18 13.30-13.55

"Multilevel autoregressive modeling with measurement error"

Noémi Schuurman

There is an increasing interest in psychology in studying within-person processes that unfold over time. Further, the development of personal devices such as smart phones has made it possible for researchers to collect intensive longitudinal data, consisting of many repeated measures for many persons.
Multilevel, multivariate autoregressive modeling provides a rich framework for analyzing such data, because the hierarchical structure allows for modeling both person-specific dynamics, and individual differences in these dynamics.
However, the vast majority of applications of autoregressive time series modeling in psychology - both single subject and multilevel studies - do not take measurement error into account. Given that measurement error seems omnipresent in psychological data, this is potentially quite problematic. We discuss the difference between dynamic errors and measurement errors, consequences of disregarding measurement error, and ways to incorporate measurement error in the (multilevel) autoregressive model, illustrated with an empirical data set

presentation June 18 13.55-14.20

Controlling for error probabilities when using default Bayes factors

Xin Gu:

This presentation introduces an alternative way of using default Bayes factors for the one sample *t* test, which takes into account the frequentist error probabilities. Default Bayes factors have the advantage that they can be used when prior information is unavailable. We investigate the type I and type II error probabilities of the Fractional Bayes factor (FBF) (O'Hagan, 1995) in which the prior is not a translation of the substantive beliefs but a proportion of likelihood from the data. It is shown that in most typical situations the FBF results in unequal error probabilities. This implies that the FBF has a tendency to either select the null or alternative hypothesis. This is an undesirable property of default Bayes factor because there is no reason for incorrectly selecting the null or alternative more often than the other. We show how the FBF can be tuned such that the error probabilities are equal under certain conditions.

Power Analysis for the Likelihood Ratio Test in Latent Markov Models: Short-cutting the bootstrap p-value based method

Dereje W. Gudicha, Verena D. Schmittmann, Fetene B. Tekle, and Jeroen K. Vermunt
Department of Methodology and Statistics, Tilburg University, The Netherlands

In recent years, the latent Markov (LM) model has proven useful to identify distinct unobserved states and transitions between these states over time in longitudinally observed responses. The bootstrap likelihood ratio (BLR) test is becoming a gold standard for testing the number of states, yet little is known about power analysis methods for this test. This paper presents a short-cut to a p-value based power computation for the BLR test. The p-value based power computation involves computing the power as the proportion of the bootstrap p-value (PBP) for which the null hypothesis is rejected. This requires to perform the full bootstrap for multiple samples of the model under the alternative hypothesis. Power computation using the short-cut method involves the following simple steps: obtain the parameter estimates of the model under the null hypothesis, construct the empirical distributions of the likelihood ratio under the null and alternative hypotheses via Monte Carlo simulations, and use these empirical distributions to compute the power. The advantage of this short-cut method is that it is computationally cheaper and is simple to apply for sample size determina- tion.

Keywords: Latent Markov; Number of States; Likelihood Ratio; Bootstrap;
Power Analysis; sample size.

Comparing the performance of non-parametric change point detection methods for capturing response concordance

Jedelyn Cabrieto, Francis Tuerlinckx and Eva Ceulemans.
KU Leuven, Belgium

Response concordance is a key concept in the behavioral sciences. It can be defined as the occurrence of changes in response patterning (change in mean) and/or response synchronization (change in covariation) in a multivariate time series. Revealing response concordance can be viewed as a change point detection problem, where the number of change points is unknown a priori. To solve this problem, DeCon was recently developed, which detects change points by combining a moving windows approach and robust PCA. Yet, in the literature, several other methods have been proposed that employ other non-parametric tools: E-divisive (Matteson et al., 2014), Multirank (Lung-Yut-Fong et al., 2012) and KCP (Arlot et al., 2012). The relative performance of all these methods for capturing response concordance is still unknown, however. Therefore, we compare E-divisive, Multirank, KCP and Decon, through extensive simulations. Specifically, we use the simulation settings of Bulteel et al. implying changes in mean and in correlation structure and those of Matteson et al. implying different numbers of (noise) variables. KCP emerged as the best method in almost all settings. However, in case of two or more noise variables, only DeCon performed adequately.

References

Bulteel, K., Ceulemans, E., Thompson, R., Waugh, C., Gotlib, I., Tuerlinckx, F. and Kuppens, P. (2014): DeCon: A tool to detect emotional concordance in multivariate time series data of emotional responding. *Biological Psychology, 98, 29–42.*

Matteson, D. and James, N. (2014): A Nonparametric Approach for Multiple Change Point Analysis of Multivariate Data. *Journal of the American Statistical Association, 109:505, 334-345.*

Keywords

Response Concordance, Patterning, Synchronization, Change Point Detection

Bayesian Longitudinal Modelling of Circular Data: Application and Interpretation.

Cremers, J[1*], Klugkist, I[1]
[1] Department of Methodology and Statistics, Utrecht University, the Netherlands
* Presenting author

One of the disciplines in which circular data may be encountered is the educational sciences, where they measure teachers' positions on the interpersonal circumplex. However, in the literature this type of data is not yet analyzed using circular statistics even though theoretically it is assumed that the data are circular. This results in a discrepancy between the questions that researchers are interested in and the questions that can be answered with the (linear) methods that are currently employed by the researchers who have longitudinal circular data.

Using an example dataset in which secondary school teachers' position on the interpersonal circumplex was measured during the first 16 weeks of the schoolyear, a Bayesian longitudinal model for circular data is presented. This model was originally developed by Nuñez-Antonio and Gutiérrez-Peña (2014) and is a mixed-effects model based on a projected normal distribution. This model assumes that the circular outcome variable in the data has an underlying linear bivariate normal distribution which was projected on the unit circle to produce a circular variable. Due to this nature of the projected normal distribution, the parameters of this model are given on two linear components. Ideally however, we would like to interpret these parameters on a circular scale. The present research attempts to provide both graphical and quantitative ways of interpreting circular fixed and random effects.

References

Nuñez-Antonio, G., Gutiérrez-Peña, E. (2014). A bayesian model for longitudinal data based on the projected normal distribution. *Computational Statistics and Data Analysis*, *71*, 506-519.

Decision Making in a Sequential Context:A Drift Diffusion Model Study

Laura M.S.Dekkers[1], Noor Seijdel[1], Wouter D. Weeda[2],[3], Brenda R. J. Jansen[1], Hilde M. Huizenga[1]
[1]Department of Developmental Psychology, University Amsterdam
[2]Department of Clinical Neuropsychology, Free University Amsterdam
[3]Department Educational Psychology, Free University Amsterdam

An important controversy in the decision making literature is whether people adopt integrative or sequential decision strategies. When using *integrative* strategies, people would integrate all option attributes into a subjective value, and choose the option that yields the highest subjective value. When using *sequential* strategies, people would assess the merit of each option by first comparing them on the most important attribute, choosing the option that is favourable on this attribute; in case differences between options on the most important attribute are not considered large enough, people proceed by comparing options on the second most important attribute. Thus far these decision strategies have been studied by employing paradigms in which information on all attributes is simultaneously provided. In daily life, though, option attributes are often encountered sequentially. Therefore, we modelled people's decision strategies in the more ecologically valid situation where option attributes are presented sequentially. We thereby aimed at studying the processes underlying people's decision strategies in this sequential decision context. Twenty subjects performed on a Sequential Task that required them to choose between two 'gambling machines' that could differ on two attributes. The task provided a sequential decision context, in that information on gain probability was presented first, followed by information on gain value. Reaction time distributions of correct and erroneous choices after presentation of the second attribute were modelled using the Drift Diffusion Model (DDM). Results revealed that people's choices were biased by the firstly presented attribute, reflected by adjustments in the DDM's starting point parameter. In addition, information on the firstly and secondly presented attribute was traded-off against each other. Specifically, as reflected in lower DDM's drift rates, speed of evidence accumulation was slowed in case information on the firstly and secondly presented attribute was conflicting. We conclude that in a sequential context, people thus seem to adopt a combined sequential-integrative decision strategy, such that they are biased by information on the attribute that was encountered first, yet their final choices are a trade-off of information on all attributes.

poster June 18 14.45-15.30

Bayesian Analysis of the Network Autocorrelation Model

Dino Dittrich

The Network Autocorrelation Model (NAM) has been extensively used by researchers interested in the effects of social network ties in diverse areas, such as political science, criminology and organizational studies. The most common inferential method in the model has been classical maximum likelihood (ML) estimation. This approach however has known problems such as negative bias of the network autocorrelation parameter and poor coverage of confidence intervals. To address the issues of the classical ML approach we develop new Bayesian techniques for the NAM. A key ingredient of a Bayesian approach is the choice of the prior distribution. The prior reflects the information we have about the model parameters before observing data. First, we derive two versions of Jeffreys prior, the Jeffreys-rule and Independence Jeffreys priors, which has not yet been developed for the NAM. This prior can be used for Bayesian analysis of the NAM when prior information is weak or completely unavailable. Second, we propose an informative prior for the network autocorrelation parameter based on an extensive literature review on the NAM. Moreover, new computational techniques are proposed to efficiently get posterior estimates and credibility intervals (the Bayesian equivalence of a classical confidence interval). Simulation results suggest that our Bayesian approaches are favorable to maximum likelihood ones with respect to bias and coverage of the network autocorrelation parameter.

poster June 18 14.45-15.30

Publication bias in practice: The case of Stereotype Threat.

Paulette Flore

The stereotype threat effect on girls' quantitative test performance has been an established theory within social psychology, and investigated by means of numerous small experiments. These experiments usually show that girls who are exposed to a gender related stereotype threat on average perform worse on a mathematical test compared to women who are not exposed to such a threat. However, critics suggest (Stoet & Geary, 2012; Ganley et al., 2013) that these effects of stereotype threat might be distorted by publication bias. With a meta-analysis we studied to which extent stereotype threat influenced quantitative test performance of girls and whether this mean effect was distorted by publication bias.

poster June 18 14.45-15.30

Intercoder Reliability: Coding Surveys on their Item Characteristics for Constructing Questionnaire Profiles

Frank Bais

In multi-mode questionnaire design, usually some consideration is given to mode-specific measurement error. Despite this consideration, however, these measurement effects are frequently unexpectedly large. Measurement effects are determined by the interplay between characteristics of the questionnaire and characteristics of the respondents. For predicting measurement effects, we will first investigate the utility of so-called questionnaire profiles, in which item characteristics of questionnaires are summarized. As a first research question, we ask whether questionnaires can be coded reliably on item characteristics that are suggested in the literature as influential in evoking measurement effects. We constructed a typology of item characteristics from the literature and applied it to a wide range of national surveys. Results show that the characteristics difficult language usage, sensitive information, potential presumption of a filter question, emotional charge, and centrality could not be coded reliably. We will give possible explanations for these results and make suggestions for coping with low intercoder reliability.

invited speaker June 18 15.30-16.30

JASP: A Fresh Way to do Bayesian Hypothesis Testing

Eric-Jan Wagenmakers

Bayesian hypothesis testing presents an attractive alternative to p-value hypothesis testing. The most prominent advantages of Bayesian hypothesis testing include (1) ability to quantify evidence in favor of the null hypothesis; (2) ability to quantify evidence in favor of the alternative hypothesis; and (3) ability to monitor and update evidence as the data come in. Despite these practical advantages, Bayesian hypothesis testing is still relatively rare. An important impediment to the widespread use of Bayesian tests is arguably the lack of user-friendly software for the run-of-the-mill statistical problems that confront psychologists for almost every experiment: the t-test, ANOVA, correlation, regression, and contingency tables. In this presentation I introduce JASP (http://jasp-stats.org), an open-source, cross platform, user-friendly graphical software package that allows the user to carry out both classical and Bayesian hypothesis tests for standard statistical problems. Not only is JASP a fresh and innovative statistical software in its own right, JASP enables easy Bayesian analysis via Morey and Rouder's powerful BayesFactor (http://bayesfactorpcl.r-forge.r-project.org/) software, but without users having to know R.

presentation June 18 16.30-17.30

Two-mode K-Spectral Centroid analysis for studying multivariate dynamical processes.

Joke Heylen

Researchers that study dynamic processes, often collect multivariate time profiles, mapping the evolution of a set of variables over time, for multiple subjects. For instance, many clinical studies focus on the differential effect of an intervention on different symptoms, by repeatedly measuring symptom severity. To parsimoniously describe the huge information in such data and to pursue an insightful overview on how time profiles vary as a function of both subjects and variables, we propose two-mode K-Spectral Centroid (2M-KSC) analysis. This method, that combines the key ideas of multi-mode partitioning and one-mode K-Spectral Centroid analysis, simultaneously reduces subjects to subject clusters and variables to variable clusters. This clustering is based on the shape of the time profiles under study, implying that time profiles that correspond to a specific combination of a person cluster and variable cluster are modeled with one specific reference profile, reflecting the typical evolution over time. Furthermore, each time profile receives an amplitude score, indicating its overall intensity relative to its corresponding reference profile. We apply the new 2M-KSC method to time profiles reflecting the intensity of depression symptoms during citalopram treatment.

presentation June 19 9.30-9.55

Testing conditional independence and modeling conditional dependence between response time and accuracy

Maria Bolsinova

presentation June 19 9.55-10.20

Comparison of classical and modern testing methods in change assessment

Ruslan Jabrayilov

In clinical psychology, change assessment is common for patients undergoing therapy. In this talk I will discuss research on possible differences between the classical test theory (CTT) and item response theory (IRT) with respect to change assessment. Besides testing method (CTT and IRT), I also talk about factors such as test length, test type and the change assessment method that can influence change assessment and patient classification (i.e. no change, recovery etc.).

presentation June 19 10.20-10.45

Reliability estimation and coefficient alpha revisited.

Pieter Oosterwijk

The relation between psychometricians and psychologists on the topic of test score reliability estimation has been strenuous (for psychometricians). For decades psychometricians tried to increase the use of superior reliability estimation methods and tried to stop the convolution between Cronbach's coefficient alpha (alpha) and internal consistency. Efforts to accomplish these aims have been largely unsuccessful and thus the problems remain. The main focus of this presentation is on the estimation of reliability using alpha and several other reliability estimation methods and only shortly addresses alpha as internal consistency measure. Some theoretical and empirical aspects of reliability estimation methods will be addressed in this presentation, which is structured as follows. First, reliability and alpha are defined using matrix algebra in order provide insight into what both entail. Second, results of empirical research investigating how interpreting alpha without interpreting sampling fluctuation can lead to errors will be provided. Third, characteristics like the size of sampling fluctuation and the underestimation of reliability of several reliability estimation methods will be addressed. To conclude, some references to useful literature will be given. The references will include some to authors with different view on how to best estimate reliability.

poster June 19 10.45-11.30

A Critical Assessment of Taxometrics

Robbert Hillen

Whether psychological attributes can be best represented by dimensions or categories has been a longstanding debate in psychology, and is referred to as the classification problem (Acton & Zodda, 2005; Kendell, 1975; Meehl, 1995). To solve the classification problem, Meehl introduced the statistical framework known as taxometrics (Meehl, 1965, 1968, 1973). Taxometrics has increased in popularity in the last two decades, particularly in the fields of psychopathology and personality psychology. The performance of taxometrics in detecting latent categories has been frequently studied under highly idealized data conditions (Ruscio et al., 2010; Walters et al., 2010; Walters & Ruscio, 2009, 2010). Our goal was to study how well taxometrics performed under data with measurement properties of typical of clinical scales. We investigated whether the taxometric curves for the MAXCOV, MAMBAC and L-Mode procedures were biased for this type of data. Furthermore, we added a simulation study to explore the degree to which the comparison curve fit index (CCFI, Ruscio et al., 2007), a popular taxometric fit statistic, can distinguish categories from dimensions in data that are characteristic of scales in clinical psychology. Our results indicated that the taxometric curves often did not indicate a categorical latent structure or the base rate was underestimated. The simulation study indicated that the CCFI was an inaccurate method for detecting latent categories and that it is biased to a dimensional inference for typical clinical data. We discuss the properties of the measurement model and the population model associated with this bias.

The Delphi method: methodological issues and its application to the development of rankings

Jurian Meijering

The Delphi method is a structured data-collection technique aimed at acquiring a certain level of agreement among experts on a given topic. The method seems to provide opportunities for developing expert-based and indicator-based rankings. However, research into its application to the development of rankings seems to be non-existent. Furthermore, the Delphi method has several unresolved methodological issues. Therefore, this PhD project aims to improve the methodology of the Delphi method and investigate its application to the development of rankings by conducting four studies. In the first study, simulations were performed to find out how various agreement indices behave within and across the rounds of a Delphi study. In the second study, the Delphi method was applied to develop an expert-based ranking. Additionally, an experiment within the study investigated the effect of two types of controlled opinion feedback on the drop-out rate, experts' degree of opinion change, and the level of agreement among experts. In the third study, content analysis and interviews were conducted to identify and evaluate the methodological characteristics of six indicator-based European green city rankings. In a subsequent and final study the Delphi method will be applied to find out which dimensions and indicators experts prioritize as most relevant for defining and measuring the sustainability of European cities. An experiment within this study will investigate the effect of feeding back experts' own initial ratings on the degree in which they change their opinion and reach an agreement. Finally, different data-analysis techniques will be applied to model experts' degree of conformity.

poster June 19 10.45-11.30

Family-based genetic association analysis: methods and

Camelia C. Minica
Department of Biological Psychology, Vrije Universiteit Amsterdam

My project aims to identify genes implicated in substance use phenotypes. This involves testing the statistical association between measured genetic variants scattered across the genome and the measured behavior, in a regression model (**G**enome-**W**ide **A**ssociation **S**tudies). Toward this aim I use measured genetic variants and substance use behavior measured longitudinally in family members at the Netherlands Twin Register. Information was missing regarding real-world factors affecting the power to detect genetic association when such clustered/longitudinal data are used in GWAS. I have addressed this issue first and this resulted in several papers (http://cameliaminica.nl/publications.php) embedding useful recommendations regarding optimal use of such data in GWAS. Namely, I have studied the power advantages conferred by the use of multivariate data in association studies, modeling alternatives of family-based imputed genotypes and efficient modeling of the familial covariance matrix in family-based GWAS. My current research concerns the use of the linear mixed models in rare variant association studies. In addition, I am involved in two meta-analyses projects aimed at detecting genetic variants implicated in cannabis use phenotypes.

Next, my project will address more complex questions about the role a genetic variant has in substance (ab)use. It may predict the probability of being initialized. Among the initiated, it may influence the amount used. Further, it may predict the switching to escalating use. This dynamic development of substance use can be captured in process mixture models that accommodate initiation, substance use regimes (light/moderate/heavy) and switching between regimes. Such an approach is novel in the context of GWAS and methodologically challenging. The challenge is in the inclusion of many genetic variants (>6000K) in complex, developmentally realistic, mixture models. Results of this project will provide valuable theoretical insights regarding the contribution of genetic variants which are common in the population (i.e., >5% frequency) as well as of rare variants (i.e., <1-5% frequency) to individual differences in addiction phenotypes.

Extending Bayesian analysis of circular data to comparison of multiple groups

Kees Mulder

Circular data are data measured in angles and occur in a variety of scientific disciplines.
Bayesian methods promise to allow for flexible analysis of circular data, for which few methods are available. Three existing MCMC methods (Gibbs, Metropolis-Hastings, and Rejection) for a single group of circular data were extended to be used in a between-subjects design, providing a novel procedure to compare groups of circular data. Investigating the performance of the methods by simulation study, all methods were found to overestimate the concentration parameter of the posterior, while coverage was reasonable. The rejection sampler performed best. In future research, the MCMC method may be extended to include covariates, or a within-subjects design.

Comparing four methods for estimating tree-based treatment regimes

Aniek Sies, Iven Van Mechelen,
KU Leuven, Leuven, Belgium

Background
When multiple treatment alternatives are available for a certain disease, an important challenge is to find an optimal treatment regime, which specifies for each patient the most effective treatment alternative (given his or her pattern of pretreatment characteristics). Here we focus on tree-based treatment regimes, which link a preferred treatment alternative to the leaves of a tree; as such they provide an insightful representation of the decision structure underlying the regime. Recently, several methods have been proposed that can be used for the estimation of tree-based treatment regimes. Up to now, however, only partial information is available concerning their absolute and relative performance.

Methods
Our paper addresses this issue by an extensive simulation study to evaluate four of these methods, namely Interaction Trees (IT), Model-based Recursive Partitioning, a classification-based approach developed by Zhang et al., and Qualitative Interaction Trees (QUINT). The main evaluation criterion was the expected potential outcome if the entire population would be subjected to the treatment regime resulting from each method under study. In addition, we also evaluated the Type 1 and Type 2 error probabilities of each method.

Results & conclusion

The expected outcome was highest for the approach of Zhang et al., and lowest for IT and QUINT. This appeared to be associated with the method of Zhang et al. being more liberal in nature (higher Type 1 and lower Type 2 risk), with the impact of an increased Type 1 risk on the expected outcome being negligible.

poster June 19 10.45-11.30

Development and evaluation of a belief elicitation procedure

Mariëlle Zondervan-Zwijnenburg
Universiteit Utrecht

The purpose of this study was to develop and evaluate a procedure to elicit expert beliefs about correlations. Theory, interviews with key informants, and a pilot study all contributed to the development of the elicitation procedure. The final procedure was established by eliciting behavioral scientists' beliefs about the correlation between cognitive potential and academic performance for youth enrolled in special education because of severe behavioral problems. Validity measures showed satisfactory results. Reliability measures indicated that a face-to-face group process is important for consistent responses. It was concluded that the elicitation procedure is feasible and sufficiently valid for future use. The poster further illustrates how the result of the elicitation procedure can be used in a Bayesian analysis.

presentation June 19 11.30-11.55

Exploring relations between instruction, strategies and achievement in mathematics: latent variable modeling of large-scale assessment data and experiments

Marije Fagginger Auer

National large-scale assessments have demonstrated a decrease in Dutch primary school students' achievement on multidigit multiplication and division problems (e.g., 56×23 and 544÷34) in the past two decades. This achievement decrease appears to be related to changes in students' use of mathematical strategies: the use of relatively accurate strategies for which calculations are written down decreased, while the use of inaccurate mental strategies increased. This leads to the question of what factors influence students' strategy use and achievement, and the role of instruction in this. This issue was investigated using two approaches: secondary analyses of large-scale assessment data using latent variable models and experiments in schools. With the large-scale assessment data, relations between teachers' reports on their instruction and students' strategy use and achievement were investigated using multilevel latent class analysis and LASSO penalized explanatory IRT. With two experiments in schools, it was investigated whether requiring students to write down calculations actually improved their performance, and whether additional instruction in writing down calculations positively affected students' strategy use and achievement.

presentation June 19 11.55-12.20

Meta-Cart: Integrating Classification and Regression Trees into Meta-analysis

Xinru Li, Elise Dusseldorp and Jacqueline Meulman
Mathematical Institute, Leiden University, The Netherlands

Meta-analysis is an important tool to synthesize results from multiple studies in a systematic way. Interaction effects play a central role in assessing conditions under which the relationship between study features and effect size (the outcome variable) changes in strength and/or direction. Within the framework of meta-analysis, when several study features are available, meta-regression lacks sufficient power to detect interactions between them. To overcome this shortcoming, a new approach named "meta-CART" (Dusseldorp et al. 2014) introduced Classification and Regression Trees (CART) in the field of meta-analytic data to identify interactions. The current implementation of meta-CART has its shortcomings: when applying CART, the sample sizes of studies are not taken into account, and the effect size is dichotomized around the median value. In our presentation, we will overcome these shortcomings by 1) weighting the study effect sizes by their accuracy, and 2) using the numerical values of the outcome variable instead of dichotomization. The new methodology will be compared to the current meta-CART in terms of Type I error, Power, and recovery performance in a Monte Carlo simulation study. Our initial results are promising, and an extensive simulation study for different population effect sizes and heterogeneity magnitudes will be presented.

invited speaker June 19 12.20-13.15

Model fit evaluation by sensitivity analysis

Daniel Oberski,
Dept. of Methodology & Statistics, Tilburg University, The Netherlands

Latent variable models involve restrictions that can also be seen as "misspecifications": restrictions with a model-based meaning. Examples include zero cross-loadings in factor analysis, zero local dependencies in latent class modeling, and “measurement invariance” or “differential item functioning” in IRT. If incorrect, such misspecifications can potentially disturb the main purpose of latent variable modeling. This possible disturbance makes model fit evaluation essential, because conclusions are unlikely to be affected when the model fits the data.

In practice, however, the model rarely fits the data. Which should we then stop doing: the modeling or the evaluation of the modeling? Both choices are bad. Abandoning modeling will needlessly throw away information when the misspecifications are irrelevant to the conclusions at hand. Abandoning evaluation, meanwhile, is disastrous when the misspecifications *are* relevant to the conclusions.

I therefore proposed a third option recently. When the model does not fit the data according to a null hypothesis test, I suggest evaluating whether the conclusions could be substantively affected by the misspecification.

To do this, I defined a measure based on the likelihood of the restricted model that approximates the change in the parameters of interest if the misspecification were freed, the *EPC-interest*. The main idea is to examine the EPC-interest and free those misspecifications that are “important” while ignoring those that are not. The measure is implemented in the *lavaan* software for structural equation modeling and the Latent Gold software for latent class analysis.

References

Preprints of the papers can be found at http://daob.nl/publications

Oberski, DL. (2013). “Evaluating Sensitivity of Parameters of Interest to Measurement Invariance in Latent Variable Models”. *Political Analysis*. (http://dx.doi.org/10.1093/pan/mpt014)

Oberski, DL & Vermunt, JK (2013). "A Model-Based Approach to Goodness-of-Fit Evaluation in Item Response Theory", *Measurement: Interdisciplinary Research & Perspectives*, vol. 11, pp. 117-122. (http://dx.doi.org/10.1080/15366367.2013.835195)
