13. Tagung der Fachgruppe Methoden & Evaluation der Deutschen Gesellschaft für Psychologie

17.-20. September 2017 in Tübingen

Tagungsband

METHODEN + EVALUATION

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Inhaltsverzeichnis

Grußwort

Liebe Kolleginnen und Kollegen, liebe Tagungsteilnehmerinnen und Tagungsteilnehmer,

wir möchten Sie sehr herzlich zur 13. Tagung der Fachgruppe Methoden und Evaluation der Deutschen Gesellschaft für Psychologie im Namen des gesamten Organisationsteams in Tübingen begrüßen. Wir freuen uns, die Fachgruppentagung 2017 in Tübingen ausrichten zu dürfen und wünschen uns sehr, Ihnen einen schönen Rahmen für fachlich anregende Gespräche und ein kollegiales Beisammensein bieten zu können.

Unter dem Motto „Methoden- und Evaluationsforschung in ihrer Breite“ hatten wir Sie eingeladen, zu dieser Fachgruppentagung beizutragen. Ziel war es, die Diversität der Enwicklungsarbeiten in diesem wichtigen Grundlagenfach aufzuzeigen, welches eine starke Ausstrahlung in empirisch arbeitende, angrenzende Anwendungsfächer (z.B. Empirische Bildungsforschung) hat.

Mit Ihrer Hilfe war es uns möglich, ein abwechslungsreiches wissenschaftliches Tagungsprogramm mit interessanten Symposien, Arbeitsgruppen und Posterbeiträgen zusammenzustellen. Neben den traditionellen Themen, wie Strukturgleichungsmodelle, Item Response Modelle, Kausalität und Mehrebenenmodelle, liegen besondere Schwerpunkte bei der diesjährigen Tagung auf Themen wie der Analyse längsschnittlicher Daten, Modellierung von Antwortzeiten, Evaluationsmaßnahmen, Qualitativen Datenanalyse, Bayesschen Modellen, Statistik Software und Knowledge Space Theory.

Ebenso war es uns wichtig, Keynote-Vortragende sowie Workshop-Veranstaltende zu gewinnen, die nicht nur international überaus visibel sind, sondern über fachliche Grenzen hinweg Impulse geben. So konnten wir mit Xin Yuan Song (Statistik), Michel Regenwetter (Mathematische Psychologie, Politikwissenschaften), Carolin Strobl (Psychologische Methoden) und Kenneth A. Bollen (Psychologische Methoden, Soziologie) beeindruckende Forscherpersönlichkeiten gewinnen, die in Keynotes ihre Entwicklungstätigkeiten in sehr unterschiedlichen Teilgebieten darstellen werden. Die Workshops wurden in gleicher Weise thematisch breit zusammengestellt, um einerseits die hohe Nachfrage nach Themen wie Bayessche Modellierung mit Stan (Shravan Vasishth, Bruno Nicenboim) und Mehrebenenstrukturgleichungsmodellierung mit lavaan (Yves Rosseel, Axel Mayer) zu bedienen und andererseits Impulse zu geben, wie mit den Workshops zur robusten Schätzung mit Anwendungen in der

Psychophysiologie und Biomedizin (Michael Muma) und zum quantitativen Testen von Theorien binärer Entscheidungen (Michel Regenwetter).

Am Rande des wissenschaftlichen Programms wollen wir Ihnen außerdem mit einer Stadtführung durch Tübingen (Montag, 18.09.2017), einem Conference Dinner im Kloster Bebenhausen (Dienstag, 19.09.2017) und einem Ausflug zur Burg Hohenzollern (Mittwoch, 20.09.2017) Gelegenheit bieten, Tübingen und die Umgebung ein wenig kennenzulernen.

Sollten Sie nach Ihrer Anreise am Sonntag Abend (17.09.2017) noch Zeit und Lust haben, würden wir uns freuen, Sie in dem Gasthaus „Tübinger Wurstküche“, Am Lustnauer Tor 8, 72074 Tübingen, ab 19:00 Uhr willkommen zu heißen, um mit Ihnen in ungezwungener Atmosphäre zu essen und zu trinken.

Wir hoffen, dass Sie sich in Tübingen Zuhause fühlen, sich fachlich inspieren lassen und mit Freude lange auf die Tagung zurückblicken werden.

Ihr Organisationsteam

Allgemeines

Anreise

Tübingen ist gut mit dem **Zug** erreichbar. Die meisten Verbindungen werden Sie über Stuttgart führen.

Vom Stuttgarter Flughafen aus können Sie direkt mit dem **Bus 828** nach Tübingen kommen.

Da nur eine begrenzte Anzahl an Parkplätzen am Veranstaltungsort und in der Innenstadt vorhanden ist, wird davon abgeraten, mit dem **Auto** direkt zum Tagungsort anzureisen.

Von Ort zu Ort

Die FGME 2017 wird nahe der Altstadt in den Gebäuden des Instituts für Psychologie (Pre-Conference-Workshops) und der Neuen Aula (eigentliche Tagung) stattfinden. Beide Orte sind sehr gut zu Fuß erreichbar und liegen nur 3 Gehminuten voneinander entfernt. Die Tagungsorte sind bei langsamen Gehen in max. 10 Minuten Fußweg von der Altstadt aus zu erreichen.

Die Altstadt selbst ist eine **autofreie Zone** und kann gut zu Fuß besichtigt werden.

Falls Sie öffentliche Busse während der Tagung nutzen möchten, fahren Sie die Haltestellen *Hölderlinstraße* bzw. *Uni/Neue Aula* an, um das Psychologische Institut bzw. die Neue Aula zu erreichen. Tickets können an Automaten in den Bussen gekauft werden – halten Sie bitte dafür Kleingeld bereit.

Pre-Conference Workshops am 17.09.2017

Eberhard Karls Universität Tübingen
Psychologisches Institut
Gebäude Alte Frauenklinik
Schleichstraße 4
72076 Tübingen
Räume: 4.332, 4.333, 4.326 und 4.329

Fachgruppentagung vom 18.-20.09.2017

Eberhard Karls Universität Tübingen
Neue Aula
Geschwister-Scholl-Platz
72074 Tübingen
Räume: Audimax, HS 2, HS 4, HS 5, HS 6 und Foyer sowie Kleiner und Großer Senat

Verpflegung

Während der Kaffeepausen stehen Ihnen reichlich Getränke, wie Kaffee, Tee, Saft und Wasser sowie Snacks, wie Obst und Gebäck, zur Verfügung.

Restaurants und Imbisse

In der Tübinger Innenstadt gibt es einge Restaurants und vor allem kleinere Imbisse, die auch einen Mittagstisch anbieten. Die folgenden Entfernungen sind jeweils vom Konferenzort (Neue Aula) ausgehend berechnet:

Imbisse sind hier zu finden:

- Hao's Box - Mühlstraße 2 (550m)
- Salam Box - Mühlstraße 14 (450m)
- El Chico - Mühlstraße 14 (450m)
- Salam Imbiss - Am Lustnauer Tor 9 (400m)
- Asia-Imbiss Wok-In - Wilhelmstraße 20 (160m)
- Kalender-Kebab - Gartenstraße 1 (550m)
- Kichererbse - Metzgergasse 2 (500m)

Restaurants sind z.B.:

- Tübinger Wurstküche - Am Lustnauer Tor 8 (400m)
- Zum Alten Fritz - Gartenstraße 13 (750m)
- Unckel Café - Wilmhelmstraße 17 (250m)
- Café Bar Centrale – Doblerstraße 10 (450m)
- China Restaurant San Bao - Neckargasse 22 (550m)
- Gasthausbrauerei Neckarmüller - Gartenstraße 4 (750m)

Am Sonntag Abend nach den Workshops gibt es einen informellen Empfang in der Tübinger Wurstküche ab 19 Uhr.

Internetzugang

Sofern Sie keinen Zugang über den **eduroam**-Verbund haben, können Sie an dem Registrierungstisch einen kostenlosen Zugang für die Dauer der Tagung erhalten.

Abendprogramm

Stadtführung durch Tübingen (Montag, 18.09.2017)

Am ersten Abend der Konferenz wird eine Stadtführung durch das schöne Tübingen angeboten. Das gibt Ihnen die Möglichkeit, etwas von der Stadt zu sehen und mit anderen Tagunsteilnehmerinnen und Tagungsteilnehmern ins Gespräch zu kommen.

Die Stadtführung beginnt um **17:30 Uhr** auf der Neckarbrücke. Wir treffen uns deshalb um **17:15 im Foyer der Neuen Aula** um gemeinsam dorthin zu gehen. Selbstverständlich kann auch jeder direkt um 17:25 zur Neckarbrücke kommen.

Conference Dinner: Gesellschaftabend im Kloster Bebenhausen (Dienstag, 19.09.2017)

Idyllisch im Schönbuch gelegen, hat sich die mittelalterliche Klosteranlage von Bebenhausen fast vollständig erhalten. Bald nach der Gründung übernahm der Zisterzienserorden Bebenhausen. Schenkungen und Erwerbungen machten das neue Kloster reich. Bis zu 80 Mönche und 130 Laienbrüder lebten zeitweise hier. Im 16. Jahrhundert, nach der Einführung der Reformation, wurde Bebenhausen zur evangelischen Klosterschule. Während des Dreißigjährigen Krieges übernahmen nochmals Mönche das Kloster, aber schon 1648 endete die Zeit der Zisterzienser endgültig. Die evangelische Klosterschule bestand fort bis ins 19. Jahrhundert. Im 18. und im 19. Jahrhundert bauten die württembergischen Herrscher einen Teil des Klosters zum Jagdschloss aus.

Das Conference Dinner wird in den prächtigen Räumen des Klosters Bebenhausen stattfinden. Da das Kloster außerhalb Tübingens liegt, wird es **Shuttlebusse** geben, die die Gäste vor dem Dinner zum Kloster bringen und anschließend wieder nach Tübingen fahren.

Die Fahrten werden zu drei verschiedenen Uhrzeiten zum Dinner angeboten. Die Abfahrt der Busse ist am **Stadtgraben 2, an der Bushaltestelle „Tübingen Nonnenhaus“**. Diese erreicht man leicht, wenn man vom Konferenzgebäude die Wilhelmstraße entlang Richtung Innenstadt läuft. Sobald man die Kreuzung erreicht, die die Innenstadt begrenzt, ist die Bushaltestelle auf der rechten Seite. Dort werden Mitglieder des Organisations-Teams stehen, um Ihnen die Suche zu erleichtern.

Die **Abfahrtszeiten** sind **18:00 Uhr**, **18:20 Uhr** und **18:40 Uhr**. Da der Bus dort nicht lange halten kann, versuchen Sie bitte am Besten wenige Minuten vor der Abfahrt an der Bushaltestelle zu sein.

Die **Rückfahrt** nach Tübingen wird einmal um **23:00 Uhr** angeboten und einmal um **24:00 Uhr**. Weiterhin stellen wir dort die Nummern von Taxi-Unternehmen zur Verfügung.

Ausflug zur Burg Hohenzollern (Mittwoch, 20.09.2017)

Nach dem Ende der Konferenz wird für alle Interessierten ein Ausflug zur Burg Hohenzollern angeboten.

Die Burg Hohenzollern, am Rand der Schwäbischen Alb gelegen, ist kein Museum im herkömmlichen Sinne, sondern ein zwar geschichtsträchtiger, aber zugleich ausgesprochen lebendiger Ort, der alljährlich Hunderttausende von Besuchern aus der ganzen Welt anzieht. Neben wesentlichen Teilen der Kunstsammlung, darunter bedeutende Gemälde, kostbares Silber und Porzellan sowie die preußische Königskrone, machen zahlreiche Konzerte, Open-Air-Kino, Ausstellungen sowie einer der schönsten Weihnachtsmärkte Deutschlands die Burg zu einer ganzjährig attraktiven Kultureinrichtung.

Die heutige Burganlage wurde im 19. Jahrhundert von beiden Zweigen des Hauses Hohenzollern gemeinschaftlich wiederhergestellt und befindet sich bis heute in deren Privatbesitz. Sie dient seit nahezu eintausend Jahren als zeitweiliger Wohnsitz der Familie. Nicht zuletzt ist es auch ihre einzigartige Lage, die bereits Kaiser Wilhelm II. zu dem Ausspruch animierte, dass der „Ausblick von der Burg Hohenzollern wahrlich eine Reise wert ist“.

Die **Abfahrt** für den Ausflug zum Schloss Hohenzollern ist am **Stadtgraben 2, an der Bushaltestelle „Tübingen Nonnenhaus“**. Diese erreicht man leicht, wenn man vom Konferenzgebäude die Wilhelmstraße entlang Richtung Innenstadt läuft. Sobald man die Kreuzung erreicht, die die Innenstadt begrenzt, ist die Bushaltestelle auf der rechten Seite. Dort werden Mitglieder des Organisations-Teams stehen, um Ihnen die Suche zu erleichtern.

Die **Abfahrtszeit** ist um **14:30 Uhr**. Da der Bus dort nicht lange halten kann, werden Sie gebeten, spätestens um 14:25 an der Bushaltestelle zu sein.

Die **Rückfahrt** ist am Parkplatz des Schlosses um **18:00 Uhr**, sodass wir circa um 18:45 Uhr wieder in Tübingen sein sollten.

Sehenswürdigkeiten

Die Stadt Tübingen liegt am Neckar im geografischen Mittelpunkt des Bundeslandes Baden-Württemberg im Südwesten Deutschlands. Sie ist zweifellos eine der schönsten Städte Deutschlands mit einer mittelalterlich geprägten Altstadt, einem markanten Marktplatz und der malerischen Neckarfront, die Besucher von nah und fern verzaubert. Mit über 83.000 Einwohnern und über 28.000 Studierenden ist Tübingen außerdem eine junge und weltoffene Stadt, der vor allem im Sommer ein mediterranes Flair nachgesagt wird.

- Tübingen hat eine **sehr schöne Altstadt**. Es lohnt sich ein wenig durch die Gassen zu schlendern und in den vielen kleinen Läden zu stöbern.
- Sehenswert sind zum Beispiel der **Hölderlinturm** und die **Stiftskirche**, deren Kirchturm man besteigen kann.
- Ebenfalls einen Besuch wert ist das **Schloss Hohentübingen**, das auch ein Museum beherbergt, das spannende Einblicke in die Geschichte vom Anbeginn der menschlichen Kultur gibt.
- Im Tübinger Umfeld lohnt sich zum Beispiel ein Ausflug zur **Wurmlinger Kapelle** oder zur Burg Hohenzollern.

Vorbereitung der Beiträge (Talks)

Die Sessions sind derart gestaltet, dass jeder individuelle Beitrag ca. 15 Minuten plus 5 Minuten Diskussion umfassen sollte. In den 90 Minuten Slots steht Ihnen damit insgesamt ein kleiner zeitlicher Puffer zur Verfügung. Bitte halten Sie sich an die vorgegebene Zeit, sodass Sie Zuhörern die Gelegenheit geben können, zwischen den parallelen Sitzungen zu wechseln.

Vorträge können auf Deutsch oder Englisch gehalten werden.

Bitte bringen Sie Ihren Vortrag auf einem USB-Stick **rechtzeitig** vor der Session in den jeweiligen Hörsaal, sodass er auf den bereitgestellten Laptop von einer technisch helfenden Person aufgespielt werden kann. Alle üblichen Formate wie pdf, ppt und doc können auf den Rechnern verwendet werden (Windows 7). Sie können – wenn Sie dies wünschen – auch Ihren eigenen Laptop anschließen. In diesem Fall stellen Sie bitte sicher, dass Sie eventuell notwendige passende VGA- und HDMI-Adapter mitbringen.

Tagungsablauf

Sonntag, 17. September

Pre-Conference Workshops

9:00	Introduction to Bayesian Modeling using Stan *Shravan Vasishth (with Bruno Nicenboim)* **PI 4.326**	Multilevel Structural Equation Modeling with lavaan *Yves Rosseel and Axel Mayer* **PI 4.329**
	Robust Estimation with Applications in Psychophysiological and Biomedical Signal Processing *Michael Muma* **PI 4.332**	Quantitative Testing of Theories of Binary Choice *Michel Regenwetter* **PI 4.333**
19:00	Informeller Empfangsabend in der Tübinger Wurstküche (Am Lustnauer Tor 8, 72074 Tübingen)	

Montag, 18. September

	HS 2	**HS 4**	**HS 5**
9:00	Tagungseröffnung (Audimax)		
9:30	Komplexe Strukturgleichungsmodelle *Chair: Tobias Koch*	Modellierung von Antwortstilen *Chair: Christof Schuster*	Diskrete latente Variablenmodelle *Chair: Jürgen Heller*
	Ein Likelihood-Quotienten-Test zur Beurteilung des Model Fit für nichtlineare Strukturgleichungsmodelle *Büchner, Rebecca; Klein, Andreas; Schermelleh-Engel, Karin*	Vergleich von Ansätzen zur Schätzung von extremem Antwortstil aufgrund des Differential Discrimination Model *Schuster, Christof; Lubbe, Dirk*	Discrete-state modeling of discrete and continuous variables: A generalized processing tree framework *Heck, Daniel W.; Edgar, Erdfelder*
	Identification of Latent Variables *Avello, Danny Andrés; San Martin, Ernesto*	Adjacent Category Logits versus Cumulative Logits: A Note on Scaled Threshold Models for Measuring Extreme Response Style *Lubbe, Dirk; Schuster, Christof*	Accounting for parameter heterogeneity in multinomial processing tree models using model-based recursive partitioning *Wickelmaier, Florian; Zeileis, Achim*
	A multilevel MTMM latent state-trait graded response model for interchangeable and structurally different methods *Holtmann, Jana; Koch, Tobias; Eid, Michael*	Ordinale IRT-Modelle und Mehrprozess-Modelle zur Analyse item- und zeitabhängiger Effekte von Antworttendenzen auf Rating-Daten *Meiser, Thorsten; Henninger, Mirka; Plieninger, Hansjörg*	Persönlichkeitsdiagnostik auf der Grundlage verallgemeinerter Wissensstrukturen *Heller, Jürgen*
	Explaining General and Specific Factors in G-Factor Models Without Bias: Caveats and Recommendations *Koch, Tobias; Holtmann, Jana; Bohn, Johannes; Eid, Michael*	Ein neues Modell für Akquieszenz an der Schnittstelle von Psychometrie und Kognitiver Psychologie *Plieninger, Hansjörg; Heck, Daniel W.*	Random Forest Variable Importance Maße - Ein Überblick *Bollmann, Stella; Debeer, Dries; Deutschmann, Eva; Strobl, Carolin*
11:00	Kaffeepause (Foyer)		

	HS 2	HS 4	HS 5
11:20	Keynote (Audimax)		
	Song, Xin Yuan: Joint Modeling Approach for Analyzing Complex Data with Latent Variable (*Chair: Augustin Kelava*)		
12:20	Mittagspause		
13:45	Nichtlineare SEM (Symposium)	Item Fit und Differential Item Functioning in IRT-Modellen	Anwendungsbeispiele: Antworttendenzen und Kontrollmöglichkeiten
	Chair: Andreas Klein	*Chair: Johannes Hartig*	*Chair: Juliane Wilcke*
	Latent Variable Models with Nonlinear Structures: Problems and Challenges	Schätzung der Itemfit-Statistik RMSD in Stichproben in Relation zu ihrer Definition in der Population	Differential discrimination model of experimental data
	Klein, Andreas	*Hartig, Johannes; Robitzsch, Alexander; Köhler, Carmen*	*Wilcke, Juliane; Glock, Christian; Lubbe, Dirk*
	A lasso estimator for accurate estimation in complex nonlinear SEM	Bias Korrektur der RMSD Itemfit Statistik mithilfe des Bootstrap	Careless Responding als kulturvergleichende Dimension
	Brandt, Holger; Cambria, Jenna; Kelava, Augustin	*Köhler, Carmen; Robitzsch, Alexander; Hartig, Johannes*	*Grau, Ina; Ebbeler, Christine; Banse, Rainer*
	Nonparametric estimation of a latent variable model	Detection of Differential Item Functioning based on penalized Conditional Likelihood	Kontrollmöglichkeiten in Quasi-Experimenten – Die Abschätzung des Bias in einer Evaluationsstudie zum Schulungsprogramm "famoses"
	Kelava, Augustin; Kohler, Michael; Krzyzak, Adam; Schaffland, Tim	*Gürer, Can; Draxler, Clemens*	*Hagemann, Anne; Nußbeck, Fridtjof W.; May, Theodor W.*
	Evaluation of a new nonparametric factor score estimator with a simulation study	Advances in DIF analysis: Evaluating the cluster approach of differential item pair functioning	Wirkt sich eine finanzielle Förderung von Kitas auf den Sprachstand der Kinder aus? Anwendung eines Propensity Score Matching-Verfahrens für die Stichprobenauswahl
	Schaffland, Tim; Noventa, Stefano; Kelava, Augustin	*Schulze, Daniel; Stets, Eric; Pohl, Steffi*	*Bihler, Lilly; Agache, Alexandru; Leyendecker, Birgit*
15:15	Kaffeepause (Foyer)		
15:30	Keynote (Audimax)		
	Regenwetter, Michel: The geometry of probabilistic choice induced by heterogeneous hypothetical constructs and/or error-prone responses (*Chair: Jürgen Heller*)		

Poster-Session (Foyer)

16:30	A Comparison of Unidimensionality and Measurement Precision of the Narcissistic Personality Inventory and the Narcissistic Admiration and Rivalry Questionnaire *Grosz, Michael P.; Emons, Wilco H.M.; Wetzel, Eunike; Leckelt, Marius; Chopik, William J.; Rose, Norman; Back, Mitja D.*	Analysis of longitudinal ordinal data: effect of dependence on the robustness of generalized linear mixed models *Bono, Roser; Blanca, María José; Arnau, Jaume; Alarcón, Rafael*	Comparing Maximum Likelihood and Bayesian Exploratory Factor Analysis – a Simulation Study *Scharf, Florian; Pförtner, Jana; Nestler, Steffen*
	Diffusion model analysis: a graphical user interface to fast-dm *Radev, Stefan; Lerche, Veronika; Mertens, Ulf; Voß, Andreas*	Die Bedeutung der longitudinalen Messinvarianz für klinische Studien am Beispiel der Alzheimer Erkrankung *Kleineidam, Luca; Maier, Wolfgang; Wagner, Michael*	Du bist, was du XING'st - Validierung einer Skala zur Erfassung des Nutzerverhaltens auf dem beruflichen Online-Netzwerk "XING" *Brandenberg, Gabriel; Janker, Christine; Ozimek, Phillip; Förster, Jens*
	Why vulnerable narcissists benefit from Facebook use, but grandiose narcissists do not - An examination of narcissistic Facebook use in the light of self-regulation *Ozimek, Phillip; Hanke, Stephanie; Bierhoff, Hans-Werner*	Vergleich der Struktur kognitiver Fähigkeiten zwischen Lernförderschülern und Jugendlichen der Bildungsgänge Hauptschulabschluss und Mittlerer Schulabschluss *Böhme, Hendryk; Sander, Nicolas; Sengewald, Erik*	Response Surface Analyse von Grad 3: Polynomiale Modelle zum Testen komplexer Übereinstimmungshypothesen *Humberg, Sarah; Schönbrodt, Felix; Nestler, Steffen; Back, Mitja*
	Einführung in die Datenanalyse mit dem R-Paket "dplyr" *Sauer, Sebastian*	Comparing identification constraints in response style IRT models *Henninger, Mirka; Meiser, Thorsten*	Gewinnung normativer Traitschätzer aus mehrdimensionalen Forced-Choice Daten – Eine Simulationsstudie *Frick, Susanne; Wetzel, Eunike*

17:30 Stadtführung (Treffpunkt: Foyer)

19:00 Fachgruppensitzung (Grosser Senat)

Dienstag, 19. September

	HS 2	HS 4	HS 5
8:30	Multilevelmodelle	Analyse bedingter und durchschnittlicher Behandlungseffekte (anaconda): Probleme und Lösungen (Symposium)	Evaluation von Massnahmen zur Förderung mathematischer Kompetenzen im Psychologiestudium (Symposium)
	Chair: Thorsten Meiser	*Chair: Rolf Steyer*	*Chair: Sarah Bebermeier*
	Einfluss der Stichprobengröße auf Parameterschätzungen in Drei-Ebenen-Modellen *Kerkhoff, Denise; Nussbeck, Fridtjof W.*	Metaanalytische Schätzung kausaler Effekte bei binären Outcome-Variablen *Jusepeitis, Adrian; Steyer, Rolf*	Wie kann der Studienerfolg in Statistik erhöht werden? Wer nutzt welche Lernhilfen und was bewirken sie? *Bebermeier, Sarah; Austerschmidt, Kim Laura; Nussbeck, Fridtjof*
	Reliability of empirical Bayes estimates in multilevel models – Implications for the analysis of inter-individual differences in within-person effects *Neubauer, Andreas; Voelkle, Manuel; Voß, Andreas; Mertens, Ulf*	Die Amplifizierung des verfälschenden Einflusses einer unreliablen Kovariate *Sengewald, Marie-Ann; Pohl, Steffi*	Evaluation eines Übungsblatts zur einfachen linearen Regression *Warnken, Aileen Sabine Bernadette; Bebermeier, Sarah*
	The effect of long-memory-type autocorrelations on the estimation of continuous and categorical effects in linear mixed-effects regression models *Wallot, Sebastian; Schlotz, Wolff*	Die Bedeutung stochastischer Kovariaten bei der Schätzung von Behandlungseffekten mittels Poisson-Regression *Kiefer, Christoph; Mayer, Axel*	Nutzung und Nutzen eines mathematischen Vorkurses in der Psychologie *Austerschmidt, Kim Laura; Bebermeier, Sarah; Nußbeck, Fridtjof*
	Ein hierarchisches stochastisches Differentialgleichungsmodell der psychobiologischen Stress-Reaktivität *Miller, Robert*	Problem der Effektschätzung bei latenten nicht normalverteilten Variablen mit dichotomen Indikatoren *Plötner, Jan; Steyer, Rolf*	Langfristige Effekte der Teilnahme an mathematischen Vorkursen auf die Studienzufriedenheit und -bewältigung *Brinkmann, Sven; Austerschmidt, Kim*
10:00	Kaffeepause (Foyer)		

	HS 2	HS 4	HS 5
10:15	Analyse von Längsschnittdaten *Chair: Holger Brandt*	Gemischte Beiträge zu IRT und SEM *Chair:*	Reaktionszeiten als psychometrische Informationsquelle: Informationsakkumulationsmodelle 1 (Symposium) *Chair: Rainer Alexandrowicz*
	Pearl's Causality and Longitudinal Data *Gische, Christian; Voelkle, Manuel*	Ein Vergleich parametrischer und nicht-parametrischer Ansätze zur Prüfung des Rasch-Modells *Debelak, Rudolf*	Zur Speed-Power-Problematik psychologischer Diagnostik: Kann uns das Diffusionsmodell weiterhelfen? *Alexandrowicz, Rainer; Bartosz, Gula; Donker, Saskia; Reich, Lars; Gottstein, Janis*
	Individual Parameter Contributions Regression for Longitudinal Data *Arnold, Manuel; Oberski, Daniel; Voelkle, Manuel*	Dealing with a two-dimensional ability phenomenon when calibrating a test according to the Rasch model *Kubinger, Klaus D.; Yanagida, Takuya; Hagenmüller, Bettina*	Vergleich von Parameterschätzungen mit dem Diffusionsmodell und dem diffIRT-Ansatz *Conci, Anna; Donker, Saskia; Gula, Bartosz; Alexandrowicz, Rainer*
	Schätzung von Wachstumsparametern ohne Verwendung von longitudinalen Informationen: Das Cohort Growth Model *Fischer, Kevin; Klein, Andreas; Reinecke, Jost*	Constraint Interaction Revisited: How to Interpret Factor Loadings under Alternative Scaling Methods *Klößner, Stefan; Klopp, Eric*	Validierung der personen- und itembezogenen Komponenten des boundary-separation Parameters im diffIRT *Reich, Lars; Alexandrowicz, Rainer W.*
	On the Definition of Latent-state-trait Models with Autoregressive Effects: Insights from LST-R theory *Eid, Michael; Jana, Holtmann; Philip, Santangelo; Ulrich, Ebner-Priemer*		Zur Validierung des Vigilanzparameters im diffIRT *Alexandrowicz, Rainer; Gottstein, Janis*
11:45	Kaffeepause und Mini-Imbiss (Foyer)		
12:05	Keynote (Audimax)		

Strobl, Carolin: Model-based recursive partitioning of psychometric models: A data-driven approach for detecting heterogeneity in model parameters (*Chair: Florian Wickelmaier*)

13:05 Mittagspause

	HS 2	HS 4	HS 5
14:25	Bayesian Structural Equation Modeling	Entwicklungen in der Statistik-Software	Reaktionszeiten als psychometrische Informationsquelle: Informationsakkumulationsmodelle 2 (Symposium)
	Chair: Augustin Kelava	*Chair: Timo von Oertzen*	*Chair: Rainer Alexandrowicz*
	Bayesian two-level model for partially ordered repeated responses *Wang, Xiaoqing; Feng, Xiangnan; Song, Xinyuan*	Introducing approxbayes – a software for approximate Bayes factors *Mertens, Ulf; Radev, Stefan; Voß, Andreas*	Modeling responses and response times in rating scales *Ranger, Jochen; Kuhn, Jörg-Tobias*
	Semiparametric latent variable models for multivariate censored data with Bayesian analysis *Ouyang, Ming; Song, Xinyuan*	Ωnyx: Ein grafisches Modellierungswerkzeug für Strukturgleichungsmodelle *Brandmaier, Andreas M.; von Oertzen, Timo*	Erweiterungen des Diffusionsmodells: Adaptiven Schwellen und Lévy-Flight-Modelle *Voß, Andreas*
	Gaussian Process Panel Modeling – A unification of longitudinal modeling approaches *Karch, Julian; Brandmaier, Andreas; Voelkle, Manuel*	Dirichlet Process Clustering in Onyx *von Oertzen, Timo; Glassen, Thomas; Brandmaier, Andreas*	Speed-Accuracy Manipulations: Are Effects in Non-Decision Time attributable to a Lack of Discriminant Validity of the Manipulation or to Trade-Offs in Parameter Estimation? *Lerche, Veronika; Voß, Andreas*
	Emotionales Erleben und tägliche Ereignisse im dynamischen Zusammenspiel: Modellierung durch Gaußsche und Poisson Vektor Autoregressionen *Adolf, Janne; Karch, Julian; Brose, Annette; Schmiedek, Florian; Voelkle, Manuel*	Observation Oriented Modeling (OOM) – Ein alternativer Ansatz für statistische Modellierung *Sauer, Sebastian; Lübke, Karsten*	Validität von Diffusionsmodellparametern in elementaren kognitiven Aufgaben *Schmitz, Florian; Wilhelm, Oliver*
15:55	Kaffeepause (Foyer)		
16:10	Modellierung von Antwortzeiten	Qualitative Datenanalyse	Anwendungsbeispiele in Onlinekontexten
	Chair: Steffi Pohl	*Chair: Philipp Mayring*	*Chair: Michael Eid*
	Using response time models to investigating the mechanism of missing values at the end of a test *Pohl, Steffi; Ulitzsch, Esther; von Davier, Matthias*	Why we cannot prove theories through experiments (or any other method): Underdetermination, the problem of induction, falsificationism, and the question why they are so rarely talked about in psychology *Holtz, Peter*	Optimal number of response categories for assessing job satisfaction: Results from an experimental online study *Kutscher, Tanja; Crayen, Claudia; Michael, Eid*
	A Dynamic Response Time Model for Time-Limited Tests *Ulitzsch, Esther; Pohl, Steffi; von Davier, Matthias*	Stichprobengröße und Stichprobenziehung in nichtrandomisierten Studien *Mayring, Philipp*	Statistik lernen mit shiny-Apps – Was? Wie? Weitere Anregungen? *Hahn, Sonja; Mundt, Fabian; Horvath, Kenneth*
	The role of copula for reaction time modeling *Colonius, Hans*	Inference Statistic: An uncertain tool to explore an unknown territory - a simulation study *Krefeld-Schwalb, Antonia; Zenker, Frank; Witte, Erich*	Development and Evaluation of a Social Brokerage Index for Social Media – A Case Study of Twitter *Rehm, Martin; Cornelissen, Frank; ten Thij, Marijn*

19:00 Gesellschaftsabend (Kloster Bebenhausen) (Abfahrt: Stadtgraben 2 um 18.00, 18.20 und 18.40 Uhr)

Mittwoch, 20. September

	HS 2	HS 4
8:30	Messinvarianz *Chair:*	Gemischte Beiträge *Chair: Rüdiger Mutz*
	IRT-basierter Item-Fit zur Analyse der Messinvarianz *Buchholz, Janine; Hartig, Johannes*	A possible connection between Knowledge Space Theory and Item Response Theory using Information Theory *Noventa, Stephano; Heller, Jürgen; Kelava, Augustin*
	Vom Testfairness-Paradox zu einem einheitlichen psychometrischen Fairness-Begriff *Yousfi, Safir*	On the relationship between maximum entropy and maximum likelihood methods with applications in psychometrics, categorical data analysis, and network analysis *Hosoya, Georg*
	Metric Measurement Invariance of Latent Variables: Foundations, Testing, and Correct Interpretation *Klößner, Stefan; Klopp, Eric*	Vorhersage oder Produktion? – Ein Bayessches Strukturgleichungsmodell für Stochastic Frontier Analysen (B-SEM-SFA) *Mutz, Rüdiger*
9:30	Kaffeepause (Foyer)	
9:45	Vorträge der PreisträgerInnen (Audimax)	
11:00	Kaffeepause (Foyer)	
11:20	Keynote (Audimax) *Bollen, Kenneth A.*: Model implied instrumental variables (MIIVs): A new orientation to structural equation models (SEMs) (*Chair: Michael Eid*)	
12:20	Kaffeepause und Mini-Imbiss (Foyer)	

	HS 2	HS 4
12:40	Modellierung von Missing Data *Chair: Oliver Lüdtke*	IRT: Kalibrierungsdesigns und Instruktionssensitivität *Chair: Andreas Frey*
	Schätzung der prognostischen Validität von Auswahlverfahren mittels multipler Imputation – Illustration des Vorgehens am Beispiel realer Datensätze *Pfaffel, Andreas; Spiel, Christiane*	Kriteriumsorientierte adaptive Hochschulklausuren: Methodische Probleme typischer Klausuren und Möglichkeiten zur Verbesserung *Frey, Andreas; Born, Sebastian; Spoden, Christian; Fink, Aron*
	Imputation fehlender Werte auf Ebene 2: Ein Vergleich verschiedener Verfahren und Implementation mithilfe von Plausible Values *Grund, Simon; Lüdtke, Oliver; Robitzsch, Alexander*	Kriteriumsorientierte adaptive Hochschulklausuren: Auswirkung verschiedener Kalibrierungsdesigns auf die Personenparameterschätzung *Born, Sebastian; Frey, Andreas; Spoden, Christian; Fink, Aron*
	Modellerierung fehlender Werte unter Berücksichtigung entscheidungstheorethischer Erkenntnisse *Buntins, Katja*	Die Rolle von Itemkovarianzstrukturen auf Klassenebene in der Messung von Instruktionssensitivität *Naumann, Alexander; Hartig, Johannes; Hochweber, Jan*
13:40	Tagungsabschluss (Audimax)	
14:30	Ausflug zur Burg Hohenzollern (Lunchpakete) (Abfahrt: Stadtgraben 2 um 14:30 Uhr)	

Workshops

Introduction to Bayesian Modeling using Stan

So 9:00
PI 4.326

Shravan Vasishth (with Bruno Nicenboim)
University of Potsdam

In this one-day workshop, we will give a comprehensive introduction to using Stan for Bayesian data analysis and Bayesian modeling. By the end of the course, participants should be able to:

1. Understand the syntax of Stan and the high-level ideas behind MCMC and HMC.
2. Fit standard models (such as linear models) in Stan.
3. Understand how hierarchical modeling works, and be able to fit complex hierarchical models in different settings.
4. Carry out sensitivity analyses to investigate how posteriors change as a result of prior specification.
5. Visualize and interpret different models.
6. Carry out posterior predictive checks and cross-validation for model evaluation.

We will provide lecture notes and suggested readings for further study. We assume that everyone has a laptop with them and has the R package rstan installed within R. This one-day workshop will involve lectures interspersed with short exercises to be done in class. In order to consolidate understanding, we will assign a project that participants can carry out (this is optional). Students have the option to submit it to the instructor a week later and get feedback.
More information on the workshop are provided here:
`http://www.ling.uni-potsdam.de/~vasishth/courses/IntroStanFGME2017.html`

Multilevel Structural Equation Modeling with lavaan

So 9:00
PI 4.329

Yves Rosseel[1], Axel Mayer[2]
1: Universität Gent; 2: RWTH Aachen

The aim of this workshop is to provide an introduction to the multilevel structural equation modeling (SEM) framework with lavaan. We focus on the application of this framework to analyze multilevel data (for example: student scores, where students are nested in schools). First, we will discuss the relationship between classic (single-level) regression, multilevel regression, and SEM. We will do this both from a theoretical point of view as well as from a software point of view. We will show how and under which conditions (classic, non-multilevel) SEM software can produce identical results as dedicated multilevel (or mixed modeling) software. Second, we will demonstrate how lavaan can be used to analyze multilevel data. We will start from a regression perspective, and gradually proceed from a simple regression analysis, to a two-level regression analysis, towards more complicated models, exploiting the full power of the multilevel SEM framework. Third, we will take a latent-variable (CFA) perspective, and give various examples of multilevel CFA, and multilevel SEM involving latent variables. Fourth, we will introduce moderation and mediation within the multilevel SEM framework and provide some examples. Along the way, we will discuss many practical issues including the role of centering, the treatment of missing and/or non-normal data, and how to deal with categorical data.
lavaan-Link: `http://lavaan.ugent.be/`

Robust Estimation with Applications in Psychophysiological and Biomedical Signal Processing

So 9:00
PI 4.332

Michael Muma
TU Darmstadt

Statistical signal processing is a powerful area of research that has been successfully applied to generations of research problems in order to extract useful information from empirical data. An effective way to incorporate knowledge about the application at hand is to use parametric models. When applying parametric methods to real-world problems, it often happens that the observations do not exactly follow the assumptions that were made to model the problem. In these cases, the nominally excellent performance can drastically degrade.

This seminar introduces the concept of robust estimation in a way that it is accessible to researchers in the area of psychology. Robust statistics formalize the theory of approximate parametric models. On the one hand, they are able to leverage upon a parametric model, but on the other hand, they do not depend critically on the exact fulfillment of the model assumptions.

The basic concepts and foundations of robust statistical theory are introduced by considering the robust estimation of the location and scale parameters of a univariate random variable. In particular, measures such as the influence function, the breakdown point and the trade-off between robustness and statistical efficiency are discussed.

After laying the foundations using simple models the concepts of robust statistics are extended to more challenging situations. Parameter estimation in linear regression is considered because of its importance of modeling many practical problems. Next, the theory of robust estimation of the multivariate location and scatter (covariance) matrix is introduced and an application of robust regularized discriminant analysis for emotion classification is provided in detail. Finally, correlated data streams, which are commonly measured, e.g., in psychophysiology are treated. A focus will lie on robust parameter estimation for ARMA models as well as methods of robust filtering and outlier cleaning.

During the entire tutorial, ample real-life applications from engineering, bio-medicine and psychophysiology will be given along with information to publicly available implementations of the considered algorithms.

QTEST: Quantitative Testing of Theories of Binary Choice

So 9:00
PI 4.333

Michel Regenwetter
University of Illinois at Urbana-Champaign

This half-day workshop is aimed at graduate students, postdocs and faculty interested in learning about distribution-free models of binary choice: How to build them and how to test them. The format will combine a lecture with hand-on exercises. Attendees should bring a laptop with the QTEST software installed. While the workshop does not require prior reading, attendees will benefit from reading any or all of the papers (reprints at `internal.psychology.illinois.edu/reprints/index.php?site_id=38`) listed below.

Key papers:

- Regenwetter, M., Davis-Stober, C.P., Lim, S.H., Cha, Y.-C., Guo, Y., Messner, W., Popova, A., & Zwilling, C. (2014). QTEST: Quantitative Testing of Theories of Binary Choice. *Decision, 1,1*, 2-34.
- Regenwetter, M. Dana, J. & Davis-Stober, C. (2011). Transitivity of Preferences. *Psychological Review, 118*, 684-688.

Also relevant:

- Brown, N. R., Davis-Stober, C.P., & Regenwetter, M. (2015). Commentary: Neural signatures of intransitive preferences. *Frontiers in Human Neuroscience.*
- Davis-Stober, C., Park, S., Brown, N. & Regenwetter, M. (2016). Reported violations of rationality may be aggregation artifacts. *Proceedings of the National Academy of Sciences of the United States of America.*
- Guo, Y. & Regenwetter, M. (2014). Quantitative Tests of the Perceived Relative Argument Model: Comment on Loomes (2010). *Psychological Review, 121*, 696-705.
- Regenwetter, M., Cavagnaro, D., Popova, A., Guo, Y., Zwilling, C., Lim, S.H., & Stevens, J.R. (in press). Heterogeneity and Parsimony in Intertemporal Choice. *Decision.*
- Regenwetter, M. & Davis-Stober, C.P. (2012). Behavioral Variability of Choices versus Structural Inconsistency of Preferences. *Psychological Review, 119*, 408-416.
- Regenwetter, M. & Robinson, M. (in press). The construct-behavior gap in behavioral decision research: A challenge beyond replicability. *Psychological Review.*

Abstracts

Ein Likelihood-Quotienten-Test zur Beurteilung des Model Fit für nichtlineare Strukturgleichungsmodelle

Mo 9:30
HS2

Rebecca Büchner*, Andreas Klein, Karin Schermelleh-Engel
Goethe Universität Frankfurt am Main, Deutschland

Seit einigen Jahren werden, zusätzlich zu linearen Strukturgleichungsmodellen (SEM), auch nichtlineare Beziehungen zwischen latenten, nicht direkt beobachtbaren Variablen (nichtlineare SEM) modelliert. In gewöhnlichen, linearen SEM wird die Gesamtmodellgüte (Model Fit) mit dem χ^2-Test und deskriptiven Gütemaßen, die meist auf dem χ^2-Test basieren, bewertet. Allerdings ist der χ^2-Test für nichtlineare SEM ungeeignet, da das in ihm verwendete saturierte Modell die nichtlinearen Terme nicht berücksichtigt. Daher existierte bisher kein geeigneter globaler Modelltest für nichtlineare SEM. Es wurde nun ein alternatives saturiertes Modell für nichtlineare SEM aus einem Quasi-Maximum-Likelihood-Ansatz (Klein & Muthen, 2006) entwickelt. Das neu entwickelte saturierte Modell lässt sich, äquivalent zum χ^2-Test, in einem Likelihood-Quotienten-Test verwenden. Zu diesem neuen Test wurden Monte-Carlo-Simulationen durchgeführt, in denen der α-Fehler und die Power für verschiedene Modelle geprüft wurden. Es wurden unter anderem die Stichprobengröße, Art und Ausprägung der Nichtlinearität der Terme variiert (quadratische Terme, Interaktionsterme). Die Simulationen zeigten gute bis sehr gute Ergebnisse für α-Fehler und Power.

Identification of Latent Variables

Mo 9:30
HS2

Danny Andrés Avello*[1], Ernesto San Martin[2]

1: Universidad Católica de Chile; 2: Universidad Católica de Chile, Center for Operations Research and Econometrics (CORE)

We studied the identification problem of latent variables in psychometric. By using the decomposition of Lord and Novick (1968) the Classical Test Theory (CTT) is the more general framework to work with (P. De Boeck & M. Wilson, 2004). We worked at this level of generality. Zimmerman (1975) pointed out that Hilbert spaces characteristics are suitable to framing the CTT. We extended his work by using the geometrical properties of Hilbert spaces. A Hilbert space is a complete inner product space. Therefore it can be defined angles (Bouldin, 1972) and norms (length), and so make geometry from linear operators such as expected value among others. To solve the identification problem we used the Adjoint Operator and Minimal Splitting Subspaces concepts (Lindquist et al., 2015). The last one is close related with the Axiom of Local Independence (ALI) as a weak version of him (Boorsboom et al., 2003; Florens & Mouchart, 1985). We generalized the classical definition of reliability into a multidimensional definition, and prove that is equal to a post-reliability (as we could "observe" the latent variable). We showed that under ALI the kernel of the empirical bayes is empty. And finally we prove that under ALI the True Score is equal to the latent variable. This means that is possible to identify the latent variable from the CTT. This is very important because latent variables represents the psychological attributes of interests in a test, and now we know how to identify those attributes from the observed scores.

A multilevel MTMM latent state-trait graded response model for interchangeable and structurally different methods

Mo 9:30
HS2

Jana Holtmann*[1], Tobias Koch[2], Michael Eid[1]

1: Freie Universität Berlin, Deutschland; 2: Leuphana Universität Lüneburg

A multilevel multitrait-multimethod latent state-trait graded response model for the combination of different types of methods (LST-Com GRM) is presented. The model combines CFA-MTMM modeling approaches for interchangeable and structurally different methods with LST modeling approaches for the analysis of ordered categorical observed variables. The model allows researchers to (a) analyze convergent and discriminant validity on the trait- and occasion-specific levels, (b) investigate the stability and occasion-specificity of constructs and method effects over time, (c) scrutinize the generalizability of time-stable and occasion-specific method effects across different methods and constructs. The impact of between- and within-level sample sizes (between: 250, 500, 750; within: 2, 5, 10, 20), the number of measurement occasions (2, 3, 4), the degree of convergent validity (high vs. low), as well as the use of diffuse vs. informative prior specifications on estimation accuracy were investigated in a Monte Carlo simulation study. Results of the simulation study indicate that the LST-Com GRM can be accurately estimated with Bayesian estimation techniques using diffuse priors for models with low convergent validity and at least 500 between- and 5 within-level observations. Estimation accuracy is better for models with more than two measurement occasions. Models with high levels of convergent validity exhibit biased parameter estimates primarily for parameters connected to the method factors. Setting informative priors on loading parameters improved estimation accuracy only slightly, and primarily in conditions with few observations.

Explaining General and Specific Factors in G-Factor Models Without Bias: Caveats and Recommendations

Mo 9:30
HS2

Tobias Koch*[1], Jana Holtmann[2], Johannes Bohn[2], Michael Eid[2]

1: Leuphana Universität Lüneburg, Deutschland; 2: Freie Universität Berlin

Many psychometric models assume general and specific factors, as for example, bi-factor models, latent state-trait models, multimethod, and multilevel confirmatory factor models. To identify potential predictors of general and specific components (e.g., general and domain specific abilities), researchers typically relate untransformed explanatory variables to the latent factors in g-factor models (using a classical mimic approach). This procedure, however, fails if explanatory variables are correlated with both general and specific factors in g-factor models. It is shown that the classical mimic approach leads to model misspecification and parameter bias under such circumstances. We propose two alternative modeling strategies that can be used to circumvent these methodological problems: The multiconstruct bi-factor approach and the residual approach. The basic idea of the two modeling approaches is to transform (manifest and/or latent) explanatory variables in such a way that they can safely be linked to the latent factors in g-factor models. Using real data examples from longitudinal and multimethod research, we illustrate how both modeling approaches can be applied in practice and discuss their advantages and limitations.

Vergleich von Ansätzen zur Schätzung von extremem Antwortstil aufgrund des Differential Discrimination Model

Mo 9:30
HS4

Christof Schuster*, Dirk Lubbe
Universität Gießen, Deutschland

Das Differential Discrimination Model beschreibt Ratingurteile von Personen auf einer stetigen Antwortskala, die auf einem einzelnen Trait laden. Die Itemdiskrimination bzw. Itemladung wird dabei von einer traitunabhängigen, personenspezifischen latenten Variablen, im Folgenden als Personendiskrimination bezeichnet, moderiert. Die Personendiskrimination ist ein nicht-negativer Skalenparameter des Modells. In der Tat lässt sich das Differential Discrimination Model auch als zweidimensionale Faktorenanalyse beschreiben, bei der die Faktoren allerdings multiplikativ anstatt additiv verknüpft sind. Ein extremer Antwortstil lässt sich mit Hilfe des Differential Discrimination Models modellieren, da Personen mit hoher Personendiskrimination zu extremeren Antworten neigen. Wird das Differential Discrimination Model an die Ratingurteile einer Gruppe homogener Items angepasst, lassen sich Faktorwerte der Personendiskrimination als individuelle Schätzwerte für extremeren Antwortstil verwenden. Solche Schätzwerte lassen sich unter sehr allgemeinen Annahmen über die gemeinsame Verteilung der Ratingurteile ermitteln. Allerdings ist dabei nicht ausgeschlossen, dass negative Faktorwerte für die Personendiskrimination resultieren. Obwohl sich im Differential Discrimination Model aufgrund spezifischer Verteilungsannahmen solche negativen Faktorwerte vermeiden lassen, ist die Schätzung mit Hilfe des sog. Marginal Maximum Likelihood Ansatzes rechnerisch sehr aufwändig. Es werden deshalb Wege exploriert und evaluiert, Faktorwerte der Personendiskrimination aufgrund spezifischer Verteilungsannahmen zu ermitteln, ohne solche Annahmen für die Modellanpassung zu verwenden, mit dem Ziel negative Personendiskriminationen als Indikatoren für extremen Antwortstil zu vermeiden.

Adjacent Category Logits versus Cumulative Logits: A Note on Scaled Threshold Models for Measuring Extreme Response Style

Mo 9:30
HS4

Dirk Lubbe*, Christof Schuster
Justus-Liebig Universität Gießen, Deutschland

Extreme response style denotes the tendency of individuals to prefer the extreme categories of a rating scale irrespective of item content. To account for extreme response style in ordered categorical item responses, it has been proposed to model responder-specific sets of category thresholds. Jin and Wang (2014) proposed a parsimonious model achieving this using a responder-specific scaling factor that modifies intervals between thresholds. By individually expanding or contracting these intervals, preferences for selecting either the outer or inner response categories can be modeled. In my talk, I will explain that the utility of this approach depends on the choice of category probability logits. Specifically, I will show that in the present context models using cumulative logits are generally more useful than models based on adjacent category logits.

Ordinale IRT-Modelle und Mehrprozess-Modelle zur Analyse item- und zeitabhängiger Effekte von Antworttendenzen auf Rating-Daten

Mo 9:30
HS4

Thorsten Meiser*, Henninger Mirka, Hansjörg Plieninger
Universität Mannheim, Deutschland

Für die Analyse von Antworttendenzen in Rating-Daten wurden unterschiedliche Modellansätze vorgeschlagen, wie etwa mehrdimensionale IRT-Modelle, IRT-Modelle mit variierenden Schwellenparametern und Mehrprozess-Modelle. Während die ersten beiden Modellklassen auf der Annahme eines ordinalen Antwortprozesses basieren, konzeptualisieren Mehrprozess-Modelle den Antwortprozess als eine Abfolge kategorial unterschiedlicher binärer Entscheidungen. In dem Beitrag werden die unterschiedlichen Modellansätze genutzt, um Prädiktoren für die Wirkung von Antworttendenzen auf beobachtete Rating-Antworten zu analysieren. Hierbei werden als Antworttendenzen die generelle Präferenz mittlerer oder extremer Antwortkategorien bzw. ein Antwortverhalten unabhängig von der Itempolung berücksichtigt. Bislang wurden in der Literatur primär personseitige Prädiktoren für Antwortstile im Sinne von Persönlichkeitsmerkmalen und demographischen Variablen untersucht. In unserem Beitrag stehen hingegen Prädiktoren auf der Seite der Items, wie die Komplexität und Itemposition, sowie dynamische Person-Iteminteraktionen, wie etwa die Veränderung der individuellen Reaktionszeiten über die Items und die itemspezifische Abweichung von der zu erwartenden Bearbeitungszeit einer Person, im Vordergrund. Die modellbasierte Analyse itemspezifischer und dynamischer Prädiktoren für den Einfluss von Antworttendenzen ermöglicht die Prüfung theoriegeleiteter Hypothesen über den Zusammenhang von Antworttendenzen, kognitiven Anforderungen des Antwortprozesses und einer zunehmenden Tendenz von “Satisficing” bzw. schematischen Antworten über den Verlauf eines Fragebogens.

Ein neues Modell für Akquieszenz an der Schnittstelle von Psychometrie und Kognitiver Psychologie

Mo 9:30
HS4

Hansjörg Plieninger*, Daniel W. Heck
Universität Mannheim, Deutschland

Die Beantwortung eines Fragebogenitems wird nicht nur vom Iteminhalt (und dem entsprechenden Konstrukt) beeinflusst, sondern auch von sogenannten Antwortstilen. Diese sind definiert als die inhaltsunabhängige Präferenz für spezifische Antwortkategorien. Vor kurzem schlug Böckenholt (2012) ein Item Response Modell vor, das es erlaubt das inhaltliche Konstrukt von zwei Antwortstilen zu trennen, nämlich der Präferenz für extreme Antworten sowie der Präferenz für den Mittelpunkt einer Antwortskala. Böckenholts Modell ist ein mehrdimensionales, dichotomes Item Response Modell und geht zusätzlich mit einer psychologisch plausiblen Verarbeitungsbaumstruktur einher. Wir schlagen eine Erweiterung des Modells vor um damit zusätzlich Akquieszenz messbar zu machen, also die inhaltsunabhängige Zustimmungstendenz. Das neue Mischverteilungsmodell baut auf Item Response Theorie, Multinomialen Verarbeitungsbaummodellen (aus der Kognitiven Psychologie) sowie hierarchischer Bayesianischer Modellierung auf. Im Detail nimmt das Modell eine Mischverteilungsstruktur von zustimmenden Itemantworten an, die entweder auf das inhaltliche Konstrukt oder Akquieszenz zurückgeführt werden. In einer Simulationsstudie untersuchen wir statistische und inhaltliche Eigenschaften des Modells. Des Weiteren illustrieren wir das Modell anhand eines empirischen Datensatzes aus der Persönlichkeitspsychologie. Insgesamt stellt das Modell eine theoretisch interessante Alternative zu bisherigen Modellierungsansätzen für Akquieszenz dar. Außerdem ist das vorgeschlagene Modell nur ein Beispiel aus einer neuen Modellfamilie an der Schnittstelle von Psychometrie und Kognitiver Psychologie, die einen vielversprechenden Ansatz für zukünftige Entwicklungen und Anwendungen darstellt.

Discrete-state modeling of discrete and continuous variables: A generalized processing tree framework

Mo 9:30
HS5

Daniel W. Heck*, Edgar Erdfelder
Universität Mannheim, Deutschland

Many psychological theories assume that qualitatively distinct cognitive processes may result in identical responses. To test such discrete-state theories, we propose a new class of statistical models that account for discrete and continuous response variables jointly by assuming finite mixture distributions along with assumption of local stochastic independence. These generalized processing tree (GPT) models extend the popular class of multinomial processing tree (MPT) models, which are limited to discrete response categories and cannot account for continuous variables such as response times, eye fixation durations, or other physiological measures. GPT models assume a psychologically meaningful structure on the mixture weights that is constrained by the probabilities of the latent processing paths. Depending on the type and dimensionality of continuous data, the latent components can be modeled by normal, ex-Gaussian, or other continuous distributions with either separate or shared parameters across states. We propose an adapted expectation-maximization (EM) algorithm for maximum-likelihood parameter estimation, and highlight the benefits of jointly modeling discrete and continuous data using a simulation and an empirical example.

Accounting for parameter heterogeneity in multinomial processing tree models using model-based recursive partitioning

Mo 9:30
HS5

Florian Wickelmaier*[1], Achim Zeileis[2]

1: University of Tübingen, Germany; 2: Universität Innsbruck, Austria

In multinomial processing tree (MPT) models, individual differences between the participants in a study can lead to heterogeneity of the model parameters. While subject covariates may explain these differences, it is often unknown in advance how the parameters depend on the available covariates, that is, which variables play a role at all, interact, or have a nonlinear influence, etc. Therefore, we propose a new approach for capturing parameter heterogeneity in MPT models based on the machine learning method MOB for model-based recursive partitioning. This procedure recursively partitions the covariate space, leading to an MPT tree with subgroups that are directly interpretable in terms of effects and interactions of the covariates. The advantages and limitations of MPT trees as a means of analyzing the effects of covariates in MPT model parameters are discussed based on simulation experiments as well as on two empirical applications from memory research.

Persönlichkeitsdiagnostik auf der Grundlage verallgemeinerter Wissensstrukturen

Mo 9:30
HS5

Jürgen Heller*
Eberhard Karls Universität Tübingen, Deutschland

Die Theorie der Wissensstrukturen wurde von Jean-Claude Falmagne und Jean-Paul Doignon (1985) ursprünglich als formaler Rahmen für die Repräsentation und Diagnose von Wissen (etwa in der Schulmathematik) entwickelt. Dabei wird der Wissensstand einer Person mit der Menge der (dichotomen) Aufgaben identifiziert, die von ihr gelöst werden können. Diese mengentheoretische Kennzeichnung vermeidet eine Aggregierung der Antworten über die Aufgaben hinweg (z.B. durch Summenscores) und die damit verbundene Bildung von Äquivalenzklassen von Antwortmustern. Sie erlaubt daher eine differenzierte Beschreibung der Stärken und Schwächen einer Person in einem Wissensbereich und ermöglicht eine detaillierte Analyse der den beobachteten Antwortmustern unterliegenden Struktur. Um diese Vorzüge auch für Daten nutzbar zu machen, wie sie in der psychologischen Diagnostik aus der Anwendung von Persönlichkeitstests resultieren, ist es erforderlich die Theorie der Wissensstrukturen zu verallgemeinern. Der Vortrag stellt eine Verallgemeinerung vor, die sowohl das Antwortformat (polychotom vs. dichotom) wie auch die Art der strukturellen Abhängigkeiten zwischen den Antworten zu verschiedenen Testitems betrifft. Darüber hinaus werden die theoretischen Grundlagen für eine empirische Anwendung des verallgemeinerten Ansatzes bereitgestellt, die anhand von publizierten Datensätzen verschiedener Persönlichkeitstests exemplarisch illustriert wird.

Random Forest Variable Importance Maße - Ein Überblick

Mo 9:30
HS5

Stella Bollmann*, Dries Debeer, Eva Deutschmann, Carolin Strobl
Universität Zürich, Schweiz

Machine Learning Verfahren, wie Bagging und Random Forests, werden auch in der Psychologie immer beliebter. Sie liefern oft sehr gute Vorhersagen, sind aber nicht so eindeutig interpretierbar wie parametrische Modelle. In der psychologischen Forschung ist es jedoch meist nicht nur gewünscht, Vorhersagen zu treffen, sondern auch Erklärungsmodelle zu liefern. Aus diesem Grund wurden sogenannte Variable Importance Maße vorgeschlagen, die einen Einblick in solche "black boxes" gewähren. Damit ist es möglich, relevante Prädiktorvariablen zu identifizieren und nach ihrer Bedeutsamkeit zu sortieren. Dieses Vorgehen ist in einigen Disziplinen, wie z.B. der Genetik, der Ökologie und auch der Linguistik, bereits sehr gebräuchlich, und auch erste Anwendungen in der Psychologie sind vielversprechend. Allerdings verleiten die scheinbar einfachen Variable Importance Maße vielfach zu Überinterpretationen. Das Ziel dieses Vortrags ist es deshalb, einen Überblick über den methodischen Hintergrund und mögliche Anwendungsfelder für Random Forest Variable Importance Maße aufzuzeigen, sowie ihre statistischen Eigenschaften zu illustrieren und kritisch zu diskutieren.

Joint Modeling Approach for Analyzing Complex Data with Latent Variables

Mo 11:20
Audimax

Xinyuan Song*
The Chinese University of Hong Kong

This talk introduces several joint modeling approaches for analyzing complex data with latent variables. Several statistical models, including hidden Markov model, additive hazards model, transformation model, and regularized regression model, are considered to analyze multivariate longitudinal data, time-to-event data, and other non-normal data in the presence of latent variables. The estimating equation method, EM algorithm, and Bayesian methods are used to conduct statistical inference. Applications to real-life studies are presented.

Latent Variable Models with Nonlinear Structures: Problems and Challenges

Mo 13:45
HS2

Andreas Klein*

Goethe Univ. Frankfurt, Deutschland

What part of a nonlinear model structure stays invariant when variables or time points of observation are transformed? How difficult is it to falsify a nonlinear model structure? How can one separate model structure from artefacts? What aspects are important when to judge the size of a nonlinear effect? What do we have to consider when we want to assess model fit? This paper attempts to address some of these issues of nonlinear latent variable modeling from a fairly general methodological perspective.

A lasso estimator for accurate estimation in complex nonlinear SEM

Mo 13:45
HS2

Holger Brandt*[1], Jenna Cambria[2], Augustin Kelava[3]

1: University of Kansas, USA; 2: University of Arkansas, USA; 3: Universitaet Tübingen

Methodological research on latent interaction models within the structural equation modeling framework has mainly focused on small models with a limited number of latent variables. However, in applied research – for example in motivation theories – typically several variables are used to predict an outcome; variables that are often assumed to be correlated. We introduce a Bayesian lasso estimator that overcomes problems of estimation accuracy that particularly occur in situations with high multicollinearity or low reliability. In a simulation study, the lasso is compared to traditional approaches developed to estimate latent interaction effects. Extensions of the lasso approach to semi-parametric nonlinear effects such as splines are discussed.

Nonparametric estimation of a latent variable model

Mo 13:45
HS2

Augustin Kelava*[1], Michael Kohler[2], Adam Krzyzak[3], Tim Fabian Schaffland[1]

1: Universität Tübingen, Deutschland; 2: Technische Universität Darmstadt, Deutschland; 3: Concordia University Montreal, Canada

In this talk we present a new nonparametric latent variable approach. In this approach the model is estimated without specifying the underlying distributions of the latent variables. In a first step we fit a common factor analysis model to the observed variables. The main idea in estimation of the common factor analysis model is to estimate the values of the latent variables in such a way that the corresponding empirical distribution asymptotically satisfies the conditions that characterize the distribution of the latent variables uniquely. In a second step we apply suitable nonparametric regression techniques to analyze the relation between the latent variables in this model. Theoretical results (e.g., concerning consistency of the estimates) are briefly presented.

Evaluation of a new nonparametric factor score estimator with a simulation study

Mo 13:45
HS2

Tim Fabian Schaffland*, Stefano Noventa, Augustin Kelava
Eberhard Karls Universität Tübingen, Deutschland

A new nonparametric approach to estimate factor scores and their relationship – using nonparametric regression – will be investigated. In a first simulation study it will be compared to two different ways to estimate structural equation models, i.e. the Structural Equation Mixture Modeling (SEMM) approach and the Latent Moderated Structural Equations (LMS) approach. We will look at their performance for normal distributions with varying degrees of skewness and kurtosis and different functional relations between the latent variables. In a second simulation study the new approach will be compared to different factor score estimators, e.g. Bartlett's method and the regression method. Here we will look at their performance for different normal and non-normal distributions to compare their biases and their consistency properties.

Schätzung der Itemfit-Statistik RMSD in Stichproben in Relation zu ihrer Definition in der Population

Mo 13:45
HS4

Johannes Hartig*[1], Alexander Robitzsch[2], Carmen Köhler[1]

1: DIPF, Deutschland; 2: IPN, Deutschland

Die Root Mean Squared Deviance (RMSD) Itemfit-Statistik ist in der Software mdltm (von Davier, 2005) implementiert, welche u.a. zur Skalierung von PISA verwendet wird. Die Statistik ist in der Population definiert als die Wurzel der quadrierten Abweichungen zwischen der wahren Item-Response-Funktion (IRF) des Items und der vom Modell postulierten/angepassten IRF. Bei vorliegenden empirischen Daten wird die unbekannte wahre IRF anhand der Stichprobe geschätzt. Hierbei fällt die Schätzung allerdings nicht erwartungstreu aus, wenn die individuelle Posteriorverteilung durch IRFs nicht passender Items im Datensatz verzerrt wird. In der präsentierten Simulationsstudie wird geprüft, unter welchen Bedingungen in der Stichprobe die Schätzung der IRF erwartungstreu ausfällt und sich der RMSD aus der Stichprobe dem RMSD in der Population annähert. Hierzu werden Datensätze mit variierender Stichprobengröße (500, 5000, 100 000), Anzahl an Items (50, 200, 500) und Anzahl an nicht passenden Items (1, 10, 20) simuliert, wobei alle nicht passenden Items dieselben generierenden Parameter und somit denselben Populations-RMSD aufweisen. Die Ergebnisse zeigen, dass erst mit einer sehr hohen Anzahl von passenden Items im Datensatz der Populations-RMSD annähernd approximiert wird; mit steigender Anzahl an nicht passenden Items verschlechtert sich die Approximation.

Bias Korrektur der RMSD Itemfit Statistik mithilfe des Bootstrap

Mo 13:45
HS4

Carmen Köhler*[1], Alexander Robitzsch[2], Johannes Hartig[1]
1: DIPF, Deutschland; 2: IPN, Deutschland

Bei der Skalierung von Daten mithilfe der Item Response Theorie (IRT) ist die Passung der Daten auf das Messmodell eine notwendige Voraussetzung dafür, valide Aussagen anhand des Messmodells treffen zu können. In der Literatur existieren zahlreiche Itemfit Statistiken zur Beurteilung der Passung der beobachteten Itemantworten auf die Annahmen des Messmodells. Die meisten davon zeigen jedoch erhebliche Schwächen auf, wie beispielsweise (1) ein fehlender theoretischer Beweis über die tatsächliche Verteilung der Prüfgröße oder (2) Bias der Statistik aufgrund endlicher Stichproben. Parametrische Bootstrap-Verfahren wurden bislang bei einigen Statistiken erfolgreich eingesetzt, um die Prüfverteilung der Statistik unter der Nullhypothese zu erlangen. Der nichtparametrische Boostrap hingegen liefert die Prüfverteilung der Statistik in der zugrundeliegenden Population und kann somit zur Korrektur des Bias durch endliche Stichproben verwendet werden. In der vorgestellten Studie werden der parametrische sowie der nichtparametrische Bootstrap in Bezug auf die Root Mean Squared Deviation (RMSD) Statistik angewendet, welche u.a. in der PISA Studie verwendet wird. In Simulationsstudien wird die Performanz dreier aktueller Statistiken—Infit/Outfit, S-X^2 und RMSD—sowie die vorgeschlagene Bias korrigierte RMSD Statistik im Hinblick auf Fehler 1. Art und Teststärke unter verschiedenen Bedingungen geprüft. Ergebnisse zeigen, dass der parametrische Bootstrap angemessene Konfidenzintervalle für die RMSD Statistik liefert und die Korrektur durch den nichtparametrischen Bootstrap den positiven Bias des RMSD eliminiert. Infit/Outfit und S-X^2 weisen inkonsistente Fehler 1. Art auf. Außerdem besitzen der RMSD sowie der Bias korrigierte RMSD eine höhere Teststärke als die beiden anderen untersuchten Statistiken.

Detection of Differential Item Functioning based on penalized Conditional Likelihood

Mo 13:45
HS4

Can Gürer*, Clemens Draxler
UMIT – The Health & Life Sciences University, Österreich

Recent developments in detection of Differential Item Functioning (DIF) included approaches like Rasch Trees (Strobl et al., 2015), DIF Lasso (Tutz & Schauberger, 2015) and Item-focussed Trees (Tutz & Berger, 2016) that were able to handle metric covariates inducing DIF in comparison to well established methods, which usually had been restricted to categorical variables. Still each of these methods have downsides which shall be addressed with a new estimation method that mainly aims to combine three central virtues of the three methods: the use of Conditional Likelihood for estimation, the incorporation of genuinely linear influence of covariates on difficulty of items and the possibility to detect different DIF-types: either certain items showing DIF over all covariates, certain covariates inducing DIF over all items, or certain covariates inducing DIF in certain items. Each of the recent methods mentioned lacks in two of these aspects. In this talk we will introduce a method for DIF-detection, which firstly uses the Conditional Likelihood for estimation to tackle the problem of nuisance parameters (Andersen, 1970) combined with an L1-penalization for variable selection, secondly is based on the DIF-model used in DIF Lasso to include genuinely linear effects instead of approximation through step functions, and thirdly leaves the user the option to decide, which of the three DIF-types they'd like to investigate. The method will be described theoretically, the challenges in implementation will be discussed and performance of the approach illustrated presenting first results.

Advances in DIF analysis: Evaluating the cluster approach of differential item pair functioning

Mo 13:45
HS4

Daniel Schulze*, Eric Stets, Steffi Pohl
Freie Universität Berlin, Deutschland

The phenomenon of differential item functioning (DIF, e.g., between two groups) is regularly of interest in the context of item response models. Proofing the mere existence of DIF in an item set is easy but identifying specific items displaying DIF poses a challenge. This challenge arises from the identification issue of any measurement model. Commonly, this issue is solved by making assumptions, such as balancedness of DIF (equal-mean-difficulty approach) or that the majority of items is DIF-free (all-other-iterative-forward algorithm, Kopf, Zeileis, & Strobl; 2015). Bechger & Maris (2015) introduced a new conception in DIF analysis, namely relative DIF of item pairs. Their method allows the identification of item clusters which are invariant in item difficulty compared to the other items within the same cluster. This assumption-free approach comes at the cost of multiple possible solutions (i.e., clusters) for specifying a model in DIF analysis. It reflects the fact that we cannot know which items are DIF free or whether there are DIF free items at all. We have extended the cluster based approach to make it applicable to empirical data and evaluated its performance in a simulation study under a wide range of conditions (cluster size, (un-)balancing of DIF, sample size, missing values) and compare the results to those of other approaches. Although the issue of standardization cannot be circumvented by statistical means, our study explicates the assumptions made in the different approaches and shows under which conditions the different approaches result in unbiased results.

Differential discrimination model of experimental data

Mo 13:45
HS5

Juliane Wilcke*, Christian Glock, Dirk Lubbe
Justus-Liebig-Universität Gießen, Deutschland

Individual differences in the use of continuous scales can be assessed using a recently introduced factor-analytic model, the differential discrimination model (Ferrando, 2014; Lubbe & Schuster, 2016). Our aim was to explore its benefits in a new area of application, namely the analysis of experimental data. A perceptual experiment with 48 trials was performed by 98 participants, who rated the visibility of masked stimuli presented for 33 ms to 117 ms. Using the differential discrimination model, we were able to distinguish ability from response style. The latter showed clear interindividual differences, which were thus responsible for some of the observed variance in the data. We also analysed time effects in the data, such as an increase in the performance despite constant mean visibility ratings. We will discuss the benefits and difficulties using the differential discrimination model on experimental data.

Careless Responding als kulturvergleichende Dimension

Mo 13:45
HS5

Ina Grau*, Christine Ebbeler, Rainer Banse
Universität Bonn, Deutschland

Während Antworttendenzen wie sozial erwünschtes Antworten oder Akquieszenz ein seit Jahrzehnten bekanntes Problem bei der Interpretation von Fragebogenantworten darstellen, wird das Phänomen Careless Responding (CR) erst seit einigen Jahren diskutiert. CR tritt auf, wenn Probanden die Instruktion oder die Items nicht gründlich lesen und äußert sich in eintönigem Antwortverhalten (lange Reihen identischer Antworten) oder im Übersehen umgepolter Items. Vorgestellt werden mehrere Indikatoren zur Quantifizierung von CR anhand eines Fragebogens zur Messung der Big5-Persönlichkeitsmerkmale. Der Fragebogen wurde im Rahmen einer internationalen Ehezufriedenheits-Studie mehr als 8000 Personen aus 34 Ländern vorgelegt. Es wurden CR-Indizes für jede Person und aggregierte CR-Indizes für jedes Land berechnet. Mehrebenenanalysen zeigen, dass sich CR auf individueller Ebene durch Persönlichkeitsmerkmale, Bildung und den Human Development Index des Landes vorhersagen lässt. CR auf Länderebene korreliert sehr hoch mit dem Human Development Index und anderen kulturvergleichenden Dimensionen wie Individualismus und Geschlechterungleichheit (um r = .70). Die Effekte dieser Dimensionen auf Mittelwerte und Korrelationen in der Ehezufriedenheits-Studie verschwinden bei Kontrolle von CR. CR kann daher den Status einer kulturvergleichenden Dimension annehmen und zahlreiche Unterschiede zwischen Ländern erklären.

Kontrollmöglichkeiten in Quasi-Experimenten - Die Abschätzung des Bias in einer Evaluationsstudie zum Schulungsprogramm "famoses"

Mo 13:45
HS5

Anne Hagemann*[1], Fridtjof W. Nußbeck[1], Theodor W. May[1,2]

1: Universität Bielefeld, Deutschland; 2: Gesellschaft für Epilepsieforschung, Bielefeld, Deutschland

In Evaluationsstudien ist eine randomisierte Zuteilung zu Interventions- und Kontrollgruppe nicht immer durchführbar, sei es aus praktischen oder aus ethischen Gründen. Häufig ist in diesem Fall die Selbstselektion der Teilnehmer in die von ihnen bevorzugte Gruppe problematisch, weil die daraus entstehenden strukturellen Unterschiede zwischen den Gruppen die Ergebnisse verzerren können. Mit dem Matching und der statistischen Kontrolle von Kovariaten stehen verschiedene Möglichkeiten zur Reduktion dieses Bias zur Verfügung. Within-Study Comparisons zwischen randomisierter und nichtrandomisierter Zuweisung haben gezeigt, dass insbesondere die Auswahl der Kovariaten eine Rolle bei der Reduktion des Bias spielt. Am Beispiel der quasi-experimentellen Evaluation des Schulungsprogramms "famoses" für Familien von Kindern mit Epilepsie soll daher gezeigt werden, inwieweit auch ohne Vorliegen eines randomisierten Studienzweigs Aussagen zum Ausmaß des Bias in den Studienergebnissen getroffen werden können. Zu diesem Zweck werden die Ergebnisse verschiedener Auswertungsansätze verglichen, die unterschiedliche Einflussfaktoren, beispielweise den Prätest, demografische Variablen oder zeitabhängige Kovariaten, berücksichtigen können. Folgende Analysestrategien werden betrachtet: unabhängige Gruppen ohne statistische Kontrolle von Kovariaten, unabhängige Gruppen mit statistischer Kontrolle von Kovariaten und gematchte Gruppen. Beim Vergleich der Effektstärken der verschiedenen Auswertungsansätze zeigen sich zum Teil deutliche Unterschiede zwischen der Analyse unabhängiger Gruppen ohne Kontrolle von Kovariaten und der Analyse gematchter Gruppen mit Berücksichtigung von Kovariaten. Insbesondere die Nutzung der Prätest-Werte, sowie teilweise auch das Matching können jedoch die Variabilität der Effektstärken verringern. Da verschiedene, aufgrund theoretischer Überlegungen ausgewählte Kovariaten berücksichtigt werden, lässt dies insgesamt auf einen geringen Bias in den Ergebnissen schließen.

Wirkt sich eine finanzielle Förderung von Kitas auf den Sprachstand der Kinder aus? Anwendung eines Propensity Score Matching-Verfahrens für die Stichprobenauswahl

Mo 13:45
HS5

Lilly Bihler*, Alexandru Agache, Birgit Leyendecker
Ruhr-Universität Bochum, Deutschland

In Deutschland werden sehr viele personelle und finanzielle Ressourcen in die sprachliche Frühförderung investiert. Ziel des hier vorgestellten Ansatzes war es zu evaluieren, ob und wie sich zwei Kita-Förderungsprogramme (Treatments) im Land NRW - (1) doppelte und (2) einfache Bezuschussung - auf die Sprachentwicklung der Kinder auswirken. Um den Einfluss von pre-Treatment Variablen zu minimieren wurde für die Stichprobenauswahl ein Propensity Score Matching mit N=8906 Kitas durchgeführt. Dazu wurden 45 potentiell konfundierende Kovariaten aus vier Datenquellen ausgewählt: (1) KiBiz.web (administrative Daten, z.B. Anzahl der Gruppen sowie Anteile von Kindern nach Alter und Migrationshintergrund in der Kita), (2) Microm (sozialräumliche Indikatoren, z.B. Arbeitslosenquote), (3) Kriterien der Jugendämter für die Verteilung der Zuschüsse und (4) Förderung im Bundesprogramm "Frühe Chancen". Anschließend wurden mit dem R-Paket TriMatch drei logistische Modelle geschätzt, die anhand der Kovariaten jeweils zwei Gruppen miteinander verglichen und in einem mehrdimensionalen Raum die Wahrscheinlichkeit berechneten, dass eine Kita nicht bezuschusst bzw. einfach oder doppelt bezuschusst wird. Einrichtungen mit unterschiedlicher Bezuschussungsform, aber ähnlicher Wahrscheinlichkeit wurden zu einem Drilling zusammengefasst. Je nach Modell erklärten die Kovariaten 68-92% der Varianz in den Gruppenunterschieden. Auf diese Weise konnten 292 Kita-Drillinge identifiziert werden, die sich auf den Kovariaten statistisch nicht unterschieden. Es wurde eine Stichprobe von 32 Kita-Drillinge realisiert (n= 2237 Kinder im Querschnitt/ca. = 1000 Kinder Längsschnitt, genestet in 192 Gruppen und 95 Kita-Einrichtungen). In der Diskussion fokussieren wir die Frage, inwieweit nun kausale Inferenzen gezogen und Ergebnisse generalisiert werden können.

The geometry of probabilistic choice induced by heterogeneous hypothetical constructs and/or error-prone responses

Mo 15:30
Audimax

Michel Regenwetter*

University of Illinois at Urbana-Champaign

Heterogeneity of choice behavior has many potential sources: Different people may vary in the underlying hypothetical construct (e.g., preferences in decision making), a given individual may fluctuate in the hypothetical construct or be uncertain about it (e.g., uncertain preferences). Even for a given, fixed, latent state of a construct, overt behavior may vary due to probabilistic errors in responses. It is plausible that much behavior inside and outside the lab combines all of these sources of heterogeneity. This talk reviews a general geometric framework through which we can compare the parameter spaces induced by different types and sources of heterogeneity. This framework makes it possible to diagnose whether heterogeneity in observed behavior is due to heterogeneity in hypothetical constructs or in overt response processes.

A Comparison of Unidimensionality and Measurement Precision of the Narcissistic Personality Inventory and the Narcissistic Admiration and Rivalry Questionnaire

Mo 16:30
Foyer

Michael P. Grosz[1], Wilco H.M. Emons[2], Eunike Wetzel[3], Marius Leckelt[4], William J. Chopik[5], Norman Rose[1], Mitja D. Back[4]

1: University of Tübingen; 2: Tilburg University; 3: University of Konstanz; 4: University of Münster; 5: Michigan State University

We compared the closeness to unidimensionality (CU) and measurement precision (MP) of the subscales of the Narcissistic Personality Inventory (NPI; Raskin & Hall, 1979) to the CU and MP of the subscales of the Narcissistic Admiration and Rivalry Questionnaire (NARQ; Back et al., 2013). In Samples 1 and 2 (N1 = 1,949 and N2 = 695), we compared CU and MP of the NPI with its traditional forced-choice response format to the CU and MP of the NARQ with its Likert-type scale response format. In Sample 3 (N3 = 5,234), we assessed whether CU and MP of the NPI subscales change when the NPI is administered with a Likert-type scale response format. The main indices of CU were the explained common variance based on minimum rank factor analysis (Ten Berge & Socan, 2004) and fit indices of a one-factor confirmatory factor analysis model. MP was assessd via item information curves from item response models. The results indicated that NPI subscales are lower in CU and MP compared with NARQ subscales when the NPI was administered with its traditional forced-choice response format. When the NPI was administered with a Likert-type scale response format, the NPI subscale Leadership/Authority and NPI Grandiose Exhibitionism showed similarly high levels of CU and MP as the two NARQ subscales. While the NPI subscale Entitlement/Exploitativeness had a higher CU than the NARQ subscales it showed considerably lower levels of MP. Limitations and strength of the applied methods to assess CU and MP will be discussed.

Analysis of longitudinal ordinal data: effect of dependence on the robustness of generalized linear mixed models

Mo 16:30
Foyer

Roser Bono[1,2], María José Blanca[3], Jaume Arnau[1], Rafael Alarcón[3]

1: Department of Social Psychology and Quantitative Psychology, Faculty of Psychology, University of Barcelona, Spain; 2: Institute of Neurosciences, University of Barcelona, Spain; 3: Department of Psychobiology and Behavioural Sciences Methodology, Faculty of Psychology, University of Malaga, Spain

Most psychological data collected by tests or questionnaires, measuring subjective perceptions, attitudes or opinions, are ordinal (e.g. Likert-type scales). Longitudinal studies involving ordinal responses are frequently conducted in many fields of health and social sciences. In these cases, when data are observed over time, they are generally correlated. The generalized linear mixed model (GLMM) is a good technique for modelling correlated data with different types of distributions. The GLMM is a hybrid model that combines the linear mixed model (LMM) and the generalized linear model (GLM). In this study, we investigated Type I error in GLMM with the Kenward-Roger degrees of freedom adjustment and multinomial distribution. To this end, we simulated ordinal data with split-plot balanced and unbalanced designs (two groups and three repeated measures), manipulating the number of categories of the outcome, the sample size, the coefficient of sample size variation (null, slight, moderate and severe), the correlation values (low and high) and the pairing of sample size and correlation values. The GLMM was implemented using PROC GLIMMIX in SAS. The results showed that the test is robust with balanced designs under all the conditions analysed. With unbalanced designs, however, the robustness depends on the coefficient of sample size variation, the correlation value and the pairing between them. In sum, GLMM is a good analytical option for correlated ordinal outcomes with balanced groups.

This research was supported by grant PSI2016-78737-P (AEI/FEDER, UE) from the Spanish Ministry of Economy, Industry and Competitiveness.

Comparing Maximum Likelihood and Bayesian Exploratory Factor Analysis – a Simulation Study

Mo 16:30
Foyer

Florian Scharf, Jana Pförtner, Steffen Nestler
Universität Leipzig, Deutschland

The comparison of different estimation methods for exploratory factor analysis (EFA) has a long tradition in methodological research. However, although different approaches for Bayesian Exploratory Factor Analysis (BEFA) have been proposed (e.g., Rowe, 2003), most of the published research focuses on comparisons of more 'traditional' methods such as principal axis factor analysis, principal component analysis, and maximum likelihood factor analysis (MLFA) (e.g., De Winter & Dodou, 2012, De Winter & Dodou, 2016). To fill this gap, we performed a simulation study to examine the performance of MLFA, BEFA using uninformative priors, and BEFA using mildly informative priors. We compared the proportion of improper solutions as well as recovery and stability of the estimated loadings in a baseline condition with perfect simple structure in the population and in conditions in which several distortions from this ideal were implemented (e.g. cross-loadings and correlated residuals). Results showed that MLFA was prone to improper solutions (Heywood Cases) in small samples. Both methods performed equally well in the baseline condition. However, due to its less restrictive assumptions, BEFA showed a superior performance especially when correlated residuals were introduced. Implications of these findings will be discussed.

Diffusion model analysis: a graphical user interface to fast-dm

Mo 16:30
Foyer

Stefan Radev, Veronika Lerche, Ulf Mertens, Andreas Voß
Psychologisches Institut Heidelberg, Deutschland

In the following, we present the latest development of fast-dm, an open-source tool for diffusion model data analysis (Voß, Voß, & Lerche, 2015). The program enables estimation of all parameters of Ratcliff's (1978) diffusion model from the empirical response time distributions of any binary decision task. The current build introduces a brand new graphical user interface (GUI) to the user. The GUI greatly improves the ease of use and flexibility of the program. It also comes with enhanced functionality, e.g., a built-in data visualization tool for producing publication-ready plots. Diffusion model analysis will be performed on exemplary data sets in order to demonstrate the utility of the new software.

Die Bedeutung der longitudinalen Messinvarianz für klinische Studien am Beispiel der Alzheimer Erkrankung

Mo 16:30
Foyer

Luca Kleineidam[1,2], Wolfgang Maier[1,2], Michael Wagner[1,2]

1: Universitätsklinikum Bonn, Deutschland; 2: Deutsches Zentrum für Neurodegenerative Erkrankungen (DZNE), Bonn, Deutschland

Die Entwicklung wirksamer Behandlungen der Alzheimer Erkrankung ist von großer gesundheitspolitischer Bedeutung. Um die Wirksamkeit von Therapien zu beurteilen, werden in klinischen Studien meist Veränderungen in Summenscores kognitiver Tests untersucht. Eine wichtige Voraussetzung für die Interpretierbarkeit dieser Veränderungen ist das Vorliegen longitudinaler Messinvarianz (Horn & McArdle, 1992). Diese Annahme wird jedoch in der Praxis selten geprüft. Um den Einfluss von Verletzungen der Messinvarianz zu ermitteln, wurde der kognitive Abbau von 494 Patienten mit erhöhtem Risiko für eine Alzheimer Demenz mit vier neuropsychologischen Skalen untersucht, die üblicherweise in klinischen Studien verwendet werden. Zunächst wurden die Summenscores der Skalen mittels Latent Growth Curve Models ausgewertet, wobei implizit das Vorliegen von Messinvarianz vorausgesetzt werden muss. Danach erfolgte eine Auswertung mittels Second-Order Latent Growth Curve Models, welche die explizite Prüfung und Berücksichtigung von Verletzungen der longitudinalen Messinvarianz ermöglichen (Widaman, Ferrer & Conger, 2010). Die Ergebnisse zeigen, dass bei allen Skalen partielle Verletzungen der Messinvarianz vorliegen. Je nach Art und Ausprägung der Verletzungen unterscheiden sich die Schätzungen von Mittelwert und Varianz der kognitiven Veränderungen zwischen beiden Methoden erheblich, was die Stichprobenplanung klinischer Studien substantiell beeinflussen kann. Monte Carlo Simulationsstudien belegen, dass Second-Order Latent Growth Curve Models unabhängig von Verletzungen der Messinvarianz eine unverzerrte Schätzungen der kognitiven Veränderung ermöglichen, wohingegen Summenscores nur beim Vorliegen vollständiger strikter Messinvarianz zuverlässige Parameterschätzungen erlauben. Zudem sind die Effektstärken bekannter Prädiktoren der Krankheitsprogression geringer, wenn der kognitive Abbau mittels Summenscores analysiert wird. Die Bedeutung dieser Befunde für klinische Studien im Bereich der Alzheimer Erkrankung und anderer psychischer Störungen wird diskutiert.

Du bist, was du XING'st - Validierung einer Skala zur Erfassung des Nutzerverhaltens auf dem beruflichen Online-Netzwerk "XING"

Mo 16:30
Foyer

Gabriel Brandenberg, Christine Janker, Phillip Ozimek, Jens Förster
Ruhr-Universität Bochum, Deutschland

Soziale Online-Netzwerke (SNS) bieten eine neue Möglichkeit, vielfältiges Sozialverhalten zu erforschen. Dabei blieben berufliche SNS wie das deutsche Berufsnetzwerk „XING" bisher in der Forschung weitestgehend unberücksichtigt, sodass es auch bislang noch keine validierten, testdiagnostischen Verfahren zur Erfassung der Motivation und Nutzung dieses Netzwerkes zu geben scheint. Ziel der Studie ist die Validierung einer Skala zur differenzierten Erfassung des Nutzungsverhaltens auf „XING". Zu den Nutzungsbereichen Identitätsmanagement, Netzwerkbewusstsein, Netzwerkmodifikation, Kontaktmanagement, Kommunikation & Austausch und Karrieremanagement wurden in einem ersten Schritt Items entwickelt. Mittels einer groß angelegten Online-Erhebung wurden sowohl die strukturelle als auch die inhaltliche Validität untersucht. Zur inhaltlichen Validierung wurden sieben konvergente und drei divergente Validierungsmaße im Hinblick auf das Online-Verhalten erhoben. Faktoranalytische Berechnungen ermittelten acht Nutzungsdomänen mit zufriedenstellenden bis sehr guten Reliabilitätswerten. Zudem lassen sich viele erwartete signifikante Zusammenhänge zentraler Persönlichkeits- und Verhaltenskonstrukte, wie beispielsweise zu Materialismus, grandiosem Narzissmus, zur Sozialen Vergleichsorientierung oder zur Facebook-Aktivität finden, wie auch gleichzeitig Nullkorrelationen zu den postulierten divergenten Maßen. Die vorliegende Studie bietet erste Hinweise für eine zufriedenstellende Güte der neuentwickelten Skala zur Erfassung des XING-Nutzungsverhaltens. Im Rahmen des Posters werden differenzierte Ergebnisse vorgestellt und kritisch diskutiert, sowie ein Ausblick für künftige Forschung gegeben.

Why vulnerable narcissists benefit from Facebook use, but grandiose narcissists do not - An examination of narcissistic Facebook use in the light of self-regulation

Mo 16:30
Foyer

Phillip Ozimek, Stephanie Hanke, Hans-Werner Bierhoff
Ruhr-Universität Bochum, Deutschland

Narcissism influences how people communicate with others. Previous theory and research has led to the distinction between grandiose and vulnerable narcissism. This work examines the assumption that grandiose and vulnerable narcissism, although positively correlated, have different consequences for frequency and time expenditure of Facebook use. Respondents were recruited for participation in an online study. They answered demographic questions and completed measures of Facebook use, grandiose narcissism, vulnerable narcissism, and self-esteem. Three studies (Ns = 384, 175, and 289) provided evidence that vulnerable narcissism, not grandiose narcissism, was linked to Facebook use if partial correlations between narcissism and Facebook use were employed which controlled for the second narcissism scale. Implications for understanding the distinction between grandiose and vulnerable narcissism in the prediction of Facebook use are discussed and an important suggestion is that vulnerable narcissism, not grandiose narcissism is the primary determinant of Facebook use: Vulnerable narcissists seem to use Facebook as means to attain narcissistic goals whereas grandiose narcissists seem to utilize different strategies in order to attain self-regulatory goals. Additionally, we used the self-esteem scale for validating the distinction between the two facets of narcissism. Moreover, the inclusion of a second measure of Facebook use permitted the examination of the validity of the Facebook activity questionnaire. These validity checks provided encouraging results.

Vergleich der Struktur kognitiver Fähigkeiten zwischen Lernförderschülern und Jugendlichen der Bildungsgänge Hauptschulabschluss und Mittlerer Schulabschluss

Mo 16:30
Foyer

Hendryk Böhme, Nicolas Sander, Erik Sengewald
Bundesagentur für Arbeit Nürnberg, Deutschland

Die Bundesagentur für Arbeit (BA) unterstützt Jugendliche bei der beruflichen Orientierung, unter anderem im Hinblick auf Ausbildungsberufe. Eine wesentliche Zielgruppe sind dabei Jugendliche aus den Bildungsgängen Hauptschulabschluss und Mittlerer Schulabschluss. Es stellt sich die Frage, inwieweit die in diesem Kontext üblicherweise eingesetzten psychologischen Testverfahren auch bei Lernförderschülern anwendbar sind, also Schülerinnen und Schüler, die sonderpädagogische Förderung (Schwerpunkt „Lernen“) erhalten. Da ein Fokus der psychologischen Testdiagnostik auf der Erfassung kognitiver Fähigkeiten liegt, ist in diesem Zusammenhang unter anderem relevant, ob sich hinsichtlich deren Struktur Unterschiede im Vergleich zu Jugendlichen aus den Bildungsgängen Hauptschulabschluss und Mittlerer Schulabschluss zeigen. Dazu wurden die Daten von mehr als 50.000 Jugendlichen analysiert, die im Zuge der beruflichen Orientierung im Berufspsychologischen Service der BA untersucht wurden und dabei (unter anderem) kognitive Fähigkeitstests bearbeitet haben. Die Teilstichproben der Lernförderschüler sowie der Jugendlichen der Bildungsgänge Hauptschulabschluss und Mittlerer Schulabschluss können als repräsentativ für den jeweiligen Bildungsgang angenommen werden. Sie wurden jeweils als Gruppe eines Drei-Gruppen-Strukturgleichungsmodells spezifiziert, mit entsprechenden Invarianzen in Bezug auf das dreifaktorielle Modell kognitiver Fähigkeiten. Die Ergebnisse sprechen für eine vergleichbare Struktur kognitiver Fähigkeiten zwischen den drei Bildungsgängen, wobei sich – wie zu erwarten – Unterschiede hinsichtlich des Fähigkeitsniveaus zeigen.

Gewinnung normativer Traitschätzer aus mehrdimensionalen Forced-Choice Daten – Eine Simulationsstudie

Mo 16:30
Foyer

Susanne Frick[1], Eunike Wetzel[1,2]

1: Universität Konstanz, Deutschland; 2: Otto-von-Guericke-Universität Magdeburg, Deutschland

Das Thurstonian Item Response Modell (TIRT; Brown & Maydeu-Olivares, 2011) ermöglicht die Gewinnung normativer Traitschätzer aus mehrdimensionalen Forced-Choice (MFC) Daten. Im MFC-Antwortformat müssen Personen Items, die verschiedene Eigenschaften messen, danach in eine Rangreihe bringen, wie gut die Items sie als Person beschreiben. Die Eigenschaften des Fragebogens hängen dabei unter anderem von den Korrelationen der untersuchten Traits und der Zusammenstellung von Items zu Blöcken ab. In bisherigen Simulationen wurden in erster Linie die Itemparameter und der Modellfit untersucht.
Das Ziel dieser Studie war es, die Traitschätzer aus dem TIRT unter verschiedenen realistischen Bedingungen genauer zu untersuchen und mit den Traitschätzern aus der klassischen ipsativen Auswertung und aus Ratingskalendaten zu vergleichen. Hierzu wurden für einen Fragebogen mit fünf Traits MFC-Daten für Blöcke aus drei Items und Ratingskalendaten für eine vierstufige Antwortskala simuliert. Die Daten wurden jeweils klassisch (Mittelwerte) und mit dem TIRT bzw. dem Graded Response Modell ausgewertet. Zusätzlich wurden vier Faktoren komplett gekreuzt: Korrelationen der Traits angelehnt an die Big Five, Itemanzahl pro Trait, Polungen der Items und gesamte Itemanzahl. Pro Bedingung wurden 100 Replikationen durchgeführt.
Zur Bewertung der Traitrecovery wurden Biasmaße für die einzelnen Traits, den Gesamtwert, die Summen und Differenzen je zweier Traits und die Traitkorrelationen berechnet. Zudem wurden die Mahalanobis-Distanz der Profile, die tatsächliche und die empirische Reliabilität berechnet. Diese wurden zwischen den Faktorstufenkombinationen, Datengenerierungs- und Auswertungsmethoden verglichen. Die Ergebnisse werden in Bezug auf die Modelle diskutiert und Empfehlungen für die Verwendung und Entwicklung von MFC-Fragebogen abgeleitet.

Response Surface Analyse von Grad 3: Polynomiale Modelle zum Testen komplexer Übereinstimmungshypothesen

Mo 16:30
Foyer

Sarah Humberg[1], Felix Schönbrodt[2], Steffen Nestler[3], Mitja Back[1]
1: WWU Münster, Deutschland; 2: LMU München; 3: Universität Leipzig

Übereinstimmungshypothesen nehmen eine zentrale Rolle in der psychologischen Forschung ein, z.B. wenn es um Konsequenzen von Person-Umwelt-Passung, um Ähnlichkeits- oder um Genauigkeitseffekte geht. Nur eine Beispielfragestellung lautet: Ist es adaptiv, sich selbst zu kennen? D.h. geht bspw. eine Übereinstimmung von selbsteingeschätzter und tatsächlicher Intelligenz mit höherem Wohlbefinden einher? Um Hypothesen dieser Art zu testen werden Methoden der Response Surface Analyse von Grad 2 (Grad-2-RSA) immer populärer. Innerhalb der Grad-2-RSA würde bspw. Wohlbefinden aus selbsteingeschätzter und tatsächlicher Intelligenz in einer polynomialen Regression von Totalgrad 2 vorhergesagt, und die Regressionskoeffizienten in geeigneter Weise interpretiert. Allerdings stellten Anwender in letzter Zeit vermehrt genauere Hypothesen über solche Übereinstimmungseffekte auf, bei deren Untersuchung die Grad-2-RSA an ihre Grenzen stößt: Profitieren bspw. Personen auf niedrigem Intelligenzlevel mehr von einer genaueren Selbsteinschätzung als Personen auf höheren Leveln; oder hat eine Überschätzung der eigenen Intelligenz fatalere Auswirkungen auf das Wohlbefinden als eine Unterschätzung? Um die Untersuchung von Hypothesen dieser Art zu ermöglichen, stellen wir RSA-Modelle von Totalgrad 3 für die beliebtesten komplexen Übereinstimmungshypothesen vor. Anhand des Beispiels der Adaptivität von Selbstkenntnis beschreiben wir die inhaltliche Interpretation dieser Grad-3-RSA-Modelle und zeigen, wie sie mit Hilfe des R Pakets RSA getestet werden können.

Einführung in die Datenanalyse mit dem R-Paket "dplyr"

Mo 16:30
Foyer

Sebastian Sauer
FOM Hochschule, Deutschland

Innerhalb der R-Landschaft (R Core Team, 2016) hat sich das Paket "dplyr" (Wickham & Francois, 2016) binnen kurzer Zeit zu einem der verbreitesten Pakete entwickelt; es stellt ein innovatives Konzept der Datenanalyse zur Verfügung. dplyr zeichnet sich durch zwei Ideen aus. Die erste Idee ist, dass nur Tabellen ("dataframes" oder "tibbles") verarbeitet werden, keine anderen Datenstrukturen. Diese Tabellen werden von Funktion zu Funktion durchgereicht. Der Fokus auf Tabellen vereinfacht die Analyse, da Spalten nicht einzeln oder mittels Schleifen werden müssen. Die zweite Idee ist, typische Tätigkeiten der Datenanalyse anhand einer Taxonomie zu "grammatikalisieren". Es lassen sich einige Bausteine identifizieren, mit der die typischen Aufgaben der Datenanalyse durchgeführt werden können. Der Vortrag stellt beide Ideen von dplyr vor; dabei wird zuerst die Logik von dplyr erläutert ohne Rückgriff auf die R-Syntax. Danach werden praktische Umsetzung von dplyr anhand publizierter Daten aus psychologischen Studien vorgestellt. Syntax und Daten dazu werden parallel zum Vortrag bereit gestellt, so dass Interaktivität im Vortrag ermöglicht wird.

Comparing identification constraints in response style IRT models

Mo 16:30
Foyer

Mirka Henninger, Thorsten Meiser
Universität Mannheim, Deutschland

Many model variants are proposed in the IRT literature to correct for response styles, such as the tendency towards extreme categories or midscale responses. A general approach is the random threshold model: this model accounts for response tendencies heuristically through person-specific threshold shifts, and constrains these to be unrelated. We use simulations to explore this constraint in two ways. First, we compare the random threshold model to multidimensional IRT models that allow for covariation between trait and response styles by imposing constraints on response style dimensions, as well as to the Partial Credit Model that ignores response style effects. We examine how the models react to varying covariance structures among traits and response styles in the population. Results suggest that the level of covariation between trait and response styles plays a minor role for model fit and trait parameter recovery. However, the random threshold model performed worse than the multidimensional model in recovering true response style parameters. As a second step, we aim to test alternative identification strategies for the random threshold model. With the aid of a simulation, we contrast the zero-covariance constraint to a sum-to-zero constraint of threshold shifts within respondents. The possibility to estimate covariances between traits and response styles in the random threshold model allows exploratory analyses that may inform further psychometric research on response style bias. This comparison of model constraints can be a step towards regarding response styles as an integral part of the psychological process underlying item responses.

Einfluss der Stichprobengröße auf Parameterschätzungen in Drei-Ebenen-Modellen

Di 8:30
HS2

Denise Kerkhoff*, Fridtjof W. Nussbeck
Universität Bielefeld, Deutschland

Daten psychologischer Forschung weisen oftmals eine geschachtelte Struktur auf, bei der Beobachtungseinheiten in höher geordneten Strukturen genested sind. Die Methode der Wahl zur Analyse dieser Datenstrukturen sind sogenannte Mehrebenenmodelle (random effects models). Allerdings ist in Anwendungsfällen oftmals unklar, wie groß eine Stichprobe sein muss und vor allem wie viele Einheiten auf den jeweiligen Ebenen erhoben werden müssen, um eine ausreichend hohe Power zur Hypothesenprüfung und eine ausreichende Stabilität der Parameterschätzung zu erreichen. Zur Beantwortung dieser Frage und zur Ableitung von Daumenregeln wird eine Simulationsstudie durchgeführt. Basierend auf einer Anwendung aus der Bildungspsychologie wird ein Populationsdatensatz erzeugt, in dem Schüler (Lev-2) aus unterschiedlichen Schulklassen (Lev-3) messwiederholt (Lev-1) Daten liefern. Auf den drei Ebenen werden entsprechend der empirischen Anwendung Prädiktoren aufgenommen. Aus dem Populationsdatensatz werden mittels Ziehen mit Zurücklegen Bootstrap-Stichproben gezogen, wobei die Anzahl der Schulklassen (15, 35 oder 55), der Schüler (5, 15 oder 35) und der fehlenden Werte zu den fünf Messzeitpunkten (zufällig, zu den letzten zwei Messzeitpunkten oder zu den letzten drei Messzeitpunkten) variiert werden. Für die simulierten, in diesem Forschungsbereich typischen Effektgrößen legen die Ergebnisse nahe, dass Stichproben mindestens 35 Level-3 Einheiten mit jeweils 15 Level-2 Einheiten enthalten müssen, um feste Effekte unverzerrt zu schätzen, während die zufälligen Effekte eine größere Stichprobe benötigen. Die größte Verbesserung der Schätzergebnisse wird erreicht, wenn die Anzahl der Level-2 Einheiten in der Stichprobe erhöht wird.

Reliability of empirical Bayes estimates in multilevel models – Implications for the analysis of inter-individual differences in within-person effects

Di 8:30
HS2

Andreas Neubauer*[1], Manuel Voelkle[2,3], Andreas Voß[1], Ulf Mertens[1]

1: Universität Heidelberg, Deutschland; 2: Humboldt-Universität zu Berlin, Deutschland; 3: Max Planck Institut für Bildungsforschung, Berlin

Within-person effects are at the center of many psychological theories and models. Recently, inter-individual differences in the within-person effect stress reactivity (the intra-individual coupling of stress and an indicator of well-being) have been studied as a potential predictor for future outcomes such as sleep quality, overall health, and even mortality. However, issues of reliability of the within-person effect estimates have been completely ignored so far. In the present talk, we show that the reliability of within-person effects (operationalized as random slopes) can be estimated by drawing upon an established measure for growth rate reliability. We show in two simulation studies that this reliability estimate slightly overestimates the true reliability, although this bias is small in most studied scenarios. We further illustrate the application of the reliability estimate in an empirical example investigating stress reactivity in a measurement burst design (two bursts separated by two weeks). In this study, reliability of inter-individual differences in stress reactivity reached only around .50; however, this low unreliability cannot account for the low obtained test-retest correlation of stress reactivity estimates across two weeks. From a substantive point of view our findings echo recent claims that stress reactivity estimates might (a) confound reactivity, recovery, and stress anticipation effects, resulting in blurred stress reactivity estimates, and (b) be less stable over time but rather influenced by situational factors. From a methodological perspective, we point out key factors that researchers interested in using inter-individual differences in random slope estimates as person-level characteristic should consider when designing their studies.

The effect of long-memory-type autocorrelations on the estimation of continuous and categorical effects in linear mixed-effects regression models

Di 8:30
HS2

Sebastian Wallot*, Wolff Schlotz
Max-Planck-Institut für empirische Ästhetik, Deutschland

A number of stationary autocorrelation processes such as the first-order autoregressive process (AR1) have been implemented in linear mixed-effects regression models (MRMs) to account for non-independence due to autocorrelated errors in time-series of behavioral and physiological data. Although such models adequately represent local autocorrelations spanning a limited number of lags (short-memory), long-memory processes, i.e. autocorrelations spanning almost the entire time-series, might not be satisfactorily modelled. While such long-memory processes have been observed in time-series response times, eye movements, or (neuro-)physiological signals, their influence on precision of parameter estimation in MRMs is unknown. We investigated the effects of long-memory processes on estimation of linear trend and randomized-trial within-subject effect parameters in a number of different conditions. Data were simulated for five different effect sizes, five different lengths of time-series (64, 128, 256, 512, 1024), one anti-persistent and two persistent types of long-memory processes – a time-series without autocorrelation (iid) was used as a baseline. Particularly persistent types of long-memory increased the standard errors of trend and within-subject parameter estimates. While bias and standard errors for iid-normal cases quickly converged towards zero with increasing length of time-series, parameter estimation of time-series with persistent long-memory was clearly affected even for longer time-series. Our results suggest that long-memory processes adversely affect parameter estimation in MRMs. Therefore, detection procedures and potential remediation for long-memory should be investigated and applied in MRMs to model time-series that include such processes.

Ein hierarchisches stochastisches Differentialgleichungsmodell der psychobiologischen Stress-Reaktivität

Di 8:30
HS2

Robert Miller*[1,2]

1: TU Dresden, Deutschland; 2: Karolinska Institutet, Stockholm, Schweden

Zur validen Messung von akuten Stressreaktionen werden in der Psychobiologie zumeist Zeitreihendaten zu Stresshormonen und selbstberichteter Stimmung erhoben. Die Analyse solcher Daten erfolgt aus pragmatischen Gründen häufig in zwei Stufen: (1) Der Datenaggregation zu Kennwerten (wie Anstiegs- und Abfallsmaßen) auf Probanden-Ebene, und (2) der anschließenden Effektschätzung aus solchen Kennwerten auf Stichproben-Ebene. Dieses Vorgehen hat jedoch den entscheidenden Nachteil, dass sich sowohl Effektstärken und somit auch Teststärken auffallend verringern, wenn Kennwerte durch mehrere latente und unabhängige Teilprozesse gespeist werden. Vor diesem Hintergrund, wurde ein physiologisch informiertes Modell zur Abbildung von Stress-induzierten Veränderungen des Hormons Cortisol entwickelt. Dieses Differentialgleichungsmodell wurde unter Annahme von Stress-unabhängigen stochastischen Fluktuationen der Hormonsekretion auf Daten von 210 Probanden angepasst, welche den Trierer Sozial Stress Test (TSST) absolvierten. Die empirischen Bayes-Schätzer der Prozessparameter des Modells wurden anschließend zur Simulation von künstlichen Zeitreihen verwendet, welche schließlich im Sinne des oben dargestellten zweistufigen Analysenverfahrens zu verschiedenen Kennwerten aggregiert wurden. Die Analyse bestätigt, dass viele der populären Stress-Kennwerte die interindividuelle Varianz von mehreren Teilprozessen akuter Stressreaktionen inkorporieren, sodass selektive Effekte auf einzelne Teilprozesse oft unterschätzt und daher mit einer geringeren Wahrscheinlichkeit detektiert werden können. Die Entwicklung von Bayesianischen Expertensystemen in der jüngsten Vergangenheit eröffnet jedoch die Möglichkeit, dass eine einfachere, theoriegeleitete Modellierung von hierarchischen Zeitreihendaten in der Psychologie die Verwendung suboptimaler ad-hoc Analyseverfahren zu Gunsten robusterer Forschungsbefunde ablösen könnte.

Metaanalytische Schätzung kausaler Effekte bei binären Outcome-Variablen

Di 8:30
HS4

Adrian Jusepeitis*, Rolf Steyer
Friedrich Schiller Universität, Deutschland

Trotz einer Vielzahl theoretischer und praktischer Hürden hat sich die Metaanalyse als Maßstab für Entscheidungen in medizinischer und psychologischer Praxis etabliert. Sie dient der quantitativen Zusammenfassung von Studienergebnissen und letztlich dem Errechnen einer aggregierten Effektgröße. In Szenarien mit binären Treatment- und Outcome-Variablen, wie sie in der Epidemiologie oft vorgefunden werden, wird das Ziel dieser Aggregation herkömmlicherweise in der Schätzung des Erwartungswertes der Studienzentren-spezifischen log odds ratios gesehen, die aus den Studienzentren-spezifischen Daten in logistischen Regressionen als (logistische) Regressionskoeffizienten geschätzt werden. Aus der Perspektive der Kausalitätstheorie ist diese Zielstellung jedoch problematisch, da der so geschätzte Effekt nicht denjenigen Effektgrößen entsprechen, die man auch im randomisierten Experiment schätzen würde. Darüber hinaus entsprechen auch die aus den Erwartungswerten der Studienzentren-spezifischen logistischen Regressionskoeffizienten berechneten Treatment-bedingten Outcome-Wahrscheinlichkeiten nicht den adjustierten Erwartungswerten der Treatment-bedingten Outcome-Wahrscheinlichkeiten, welche man in einem randomisierten Experiment schätzen würde. Es wird daher ein alternatives Verfahren vorgeschlagen, in dem zunächst eben diese Erwartungswerte geschätzt werden, die anschließend in ein (log) odds ratio oder in eine andere gängige Effektgröße übersetzt werden können. Die so berechneten Effektgrößen sind dann genau diejenigen, die man auch im randomisierten Experiment schätzen würde. Es wird gezeigt, unter welchen Bedingungen die Schätzung des Erwartungswertes eines log odds ratio zu gravierenden Fehlern und sogar systematisch zu falschen Vorzeichen bei der Effektschätzung führt.

Die Amplifizierung des verfälschenden Einflusses einer unreliablen Kovariate

Di 8:30
HS4

Marie-Ann Sengewald*[1], Steffi Pohl[2]

1: Universität Jena, Deutschland; 2: Freie Universität Berlin, Deutschland

Bei der Schätzung von adjustierten Effekten in nicht-randomisierten Studien können wichtige Kovariaten, beispielsweise der Vortest des Outcomes, messfehlerbehaftet sein. Falls die Behandlungswahrscheinlichkeit und das Outcome von einer latenten Kovariate abhängen, dann wird durch die Adjustierung für die manifeste, messfehlbehaftete Kovariate ein verfälschter durchschnittlicher Behandlungseffekt geschätzt. Wie bereits in Studien gezeigt wurde, hängt die Verfälschung durch eine manifeste Kovariate von deren Reliabilität ab. Unklar ist jedoch inwiefern weitere Kovariaten im Modell Einfluss auf die Größe der Verfälschung haben. In unserer Arbeit untersuchen wir das Ausmaß der Verfälschung durch die Adjustierung für eine manifeste Kovariate falls (i) nur die latente Kovariate für die Adjustierung relevant ist und (ii) die latente Kovariate und eine zusätzliche Kovariate für die Adjustierung relevant sind. Im Kontext von kontinuierlichen Variablen mit linearen Zusammenhängen kann die Verfälschung analytisch quantifiziert werden. So zeigen wir die Verfälschung bei Verwendung unterschiedlicher Kovariaten in der Adjustierung (keine Kovariate, manifeste Kovariate, manifeste und zusätzliche Kovariate). Die Verfälschung hängt von der Reliabilität der manifesten Kovariate, der Stärke des Zusammenhangs der latenten Kovariate mit der Behandlungsvariablen, und der Zusammenhangsstruktur der zusätzlichen Kovariate mit den Variablen im Modell ab. Wir zeigen unter welchen Bedingungen die zusätzliche Kovariate die Verfälschung teilweise kompensiert und wann die zusätzliche Kovariate die Verfälschung amplifiziert. Limitationen und Implikationen für empirische Anwendungen werden diskutiert.

Die Bedeutung stochastischer Kovariaten bei der Schätzung von Behandlungseffekten mittels Poisson-Regression

Di 8:30
HS4

Christoph Kiefer*, Axel Mayer
RWTH Aachen University, Deutschland

Bei der Analyse von Behandlungseffekten auf eine abhängige Zählvariable (y = 0, 1, 2, 3,...) wird üblicherweise ein Poisson-Regressionsmodell betrachtet. Bei diesem nicht-linearen Modell wird die untere Beschränkung der Zählvariablen berücksichtigt. Zur Schätzung einer Poisson-Regression kann ein Generalisiertes Lineares Modell mit logarithmischer Linkfunktion und Poisson-verteilten Residuen herangezogen werden. Für die Bestimmung des durchschnittlichen Behandlungseffekts (ATE) ergeben sich bei Zählvariablen damit zwei Implikationen: Erstens, anders als im linearen Modell kann der ATE nicht über den bedingten Effekt am Mittelwert der Kovariaten bestimmt werden (wie z.B. beim mean-centering). Tatsächlich muss die Verteilung der Kovariaten berücksichtigt werden, um zu einer unverzerrten Effektschätzung zu gelangen. Zweitens, gängige Verfahren, die den ATE über die Verteilung der bedingten Effekte schätzen, basieren auf der Annahme fixierter Kovariaten. Das heißt, die Kovariaten werden nicht als zufällig betrachtet, wodurch es i.d.R. zu einer Unterschätzung des Standardfehlers für den ATE kommt. Anhand eines neuen Schätzverfahrens, das sowohl die Verteilung der Kovariaten als auch stochastische Kovariaten berücksichtigen kann, wird illustriert inwieweit beide Aspekte zur unverzerrten Schätzung des ATE und deren Standardfehler beitragen.

Problem der Effektschätzung bei latenten nicht normalverteilten Variablen mit dichotomen Indikatoren

Di 8:30
HS4

Jan Plötner*, Rolf Steyer
Friedrich-Schiller-Universität Jena, Deutschland

Um die Wirksamkeit von Behandlungen zu untersuchen, ist oft die Betrachtung von durchschnittlichen Behandlungseffekten (ATE) sinnvoll, wie sie beispielsweise in einem randomisierten Experiment geschätzt werden. Dabei ist der ATE oft auch von einer oder mehreren (qualitativen oder quantitativen) Kovariaten abhängig. Zur Schätzung des ATE können Mehrgruppen-Strukturgleichungsmodelle (SEM) verwenden werden, die es zudem ermöglichen, latente abhängige und unabhängige Variablen zu analysieren. Im Fall von einer latenten abhängigen Variablen, die durch mehrere dichotome Indikatoren gemessen wird, können dabei Modelle der Item-Response-Theorie (IRT), zum Beispiel das Raschmodell, verwendet werden. Bisherige Simulationen unter Verwendung dieser Modelle haben gezeigt, dass die ATEs unter Einbeziehung aller Kovariaten, von denen die abhängige Variable abhängt, zuverlässig geschätzt werden können. Ein Problem entsteht jedoch, wenn nicht alle im o. g. Sinn relevanten Kovariaten berücksichtigt werden. In Simulationsstudien wurde gezeigt, dass es in diesem Fall, trotz stochastischer Unabhängigkeit zwischen der Behandlung und den ignorierten Kovariaten, zu einer Verfälschung der Schätzung des ATE kommen kann. Die Nichtberücksichtigung relevanter Kovariaten kann zu einer schiefen Verteilung der latenten Variable in den Gruppen des Strukturgleichungsmodells führen. Bei Verwendung der gängigen Schätzmethoden, wie der Maximum-Likelihood-Schätzung kommt es zu einer systematischen Verfälschung der Schätzung des ATEs, da diese die Normalverteilung der latenten Variablen in den Gruppen annehmen. In diesem Beitrag werden diese Problematik anhand simulierter Daten und die Implikationen für die Schätzung von Behandlungseffekten dargestellt.

Wie kann der Studienerfolg in Statistik erhöht werden? Wer nutzt welche Lernhilfen und was bewirken sie?

Di 8:30
HS5

Sarah Bebermeier*, Kim Laura Austerschmidt, Fridtjof Nussbeck
Universität Bielefeld, Deutschland

In einer umfassenden längsschnittlichen Evaluationsstudie werden Psychologiestudierende aus vier aufeinanderfolgenden Kohorten zu verschiedenen Zeitpunkten des ersten Studienjahres befragt. Vor Studienbeginn werden fachspezifische Fähigkeiten (in Mathematik), Selbstwirksamkeit und soziodemografische Variablen erfasst, nach dem ersten Semester Fachkompetenz in Statistik und Studienwahlsicherheit und nach dem zweiten Semester Lernhilfennutzung, Fachkompetenz in Statistik und Studienerfolg, operationalisiert über die Note in der Statistikklausur und die Zufriedenheit mit dem Studium. Mediationsanalysen zeigen einen Effekt der fachspezifischen Fähigkeiten (in Mathematik) über die Fachkompetenz in Statistik auf die Note in der Statistikklausur, sowie einen Effekt der Selbstwirksamkeit über die Studienwahlsicherheit auf die Zufriedenheit mit dem Studium. Außerdem zeigt sich, dass Studierende, deren fachspezifische Fähigkeiten in Mathematik vor Studienbeginn gering waren, häufiger Lernhilfen nutzen. Ein positiver Effekt auf die Note in der Statistikklausur zeigt sich anschließend vor allem bei Lernhilfen, in denen die Studierenden aktiv mit den Inhalten arbeiten (vs. diese passiv rezipieren). Dieser Effekt kann zumindest teilweise über eine Erhöhung der Fachkompetenz in Statistik durch die Lernhilfenutzung erklärt werden, ist jedoch stärker bei Studierenden, die von vornherein kompetent(er) sind.

Evaluation eines Übungsblatts zur einfachen linearen Regression

Di 8:30
HS5

Aileen Sabine Bernadette Warnken*, Sarah Bebermeier
Universität Bielefeld, Deutschland

Mathematisch-statistische Kompetenzen spielen eine zentrale Rolle für den Studienerfolg im Fach Psychologie. Daher gibt es zahlreiche Ansätze, den Kompetenzerwerb im Verlauf des Studiums durch Unterstützungsangebote zu fördern. Im Rahmen einer Masterarbeit wird ein solches Unterstützungsangebot für Psychologiestudierende im ersten Studienjahr evaluiert. Mit Hilfe eines quasi-experimentellen Untersuchungsdesigns, in dem Studierende teils randomisiert und teils autoselektiv auf die Untersuchungsbedingung Einzel- vs. Gruppenarbeit zugewiesen wurden, wird untersucht, wie Psychologiestudierende ein Übungsblatt zum Thema einfache lineare Regression bearbeiten, welche Hilfsmittel sie für die Bearbeitung benötigen und einsetzen, wie schnell und erfolgreich sie die Aufgaben lösen, ob eine Bearbeitung in Gruppen- (vs. Einzelarbeit) effektiver ist und wie die Studierenden das Übungsblatt bewerten. Die Effektivität der Bearbeitung wird operationalisiert durch die Bearbeitungsdauer bis zur Lösung der Aufgaben, sowie durch die Anzahl und den Umfang an Hilfsmitteln. Die abhängigen Variablen werden außerdem für das Kompetenzniveau der Studierenden (bzw. der Gruppe) kontrolliert. Es liegen Daten von n = 40 Studierenden, die sich zum ersten Mal auf die Prüfung in Statistik vorbereiten, vor. Die Ergebnisse der Evaluation, sowie Implikationen für die Optimierung und den Einsatz des Angebots werden vorgestellt und diskutiert.

Nutzung und Nutzen eines mathematischen Vorkurses in der Psychologie

Di 8:30
HS5

Kim Laura Austerschmidt*, Sarah Bebermeier, Fridtjof Nußbeck
Universität Bielefeld, Deutschland

Die empirische Ausrichtung des Studiengangs Psychologie und der damit verbundene große Anteil methodisch-statistischer Inhalte wird von vielen Studierenden als Hürde erlebt. Dies kann zum einen an einer mangelnden Passung schulisch gelehrter Inhalte, falschen Erwartungen an das Studium oder einem länger zurückliegenden Schulabschluss liegen. Um Studierende auf die methodische Ausbildung vorzubereiten wurde im WS16/17 ein freiwilliger Vorkurs zur Auffrischung schulischer Mathematikinhalte angeboten. Zu Semesterbeginn fand eine Befragung aller Erstsemester statt. Neben demographischen Daten wurden sie zu ihren mathematik- und studienerfolgsbezogenen Fähigkeiten und Erwartungen befragt. Außerdem beinhaltete die Befragung einen kurzen Test der mathematischen Kompetenzen. Aufgrund der wahrgenommenen Schwierigkeiten mit den Aufgaben sollten Studierende über ihre Teilnahme am Vorkurs entscheiden. Aufgrund weiterer erhobener Daten zeigt sich, dass die anvisierte Zielgruppe erreicht wurde. Im Gegensatz zu Nicht-Teilnehmer/innen geben Teilnehmer/innen seltener an, einen Mathematik-Leistungskurs belegt zu haben und weisen eine signifikant schlechtere Mathematik- sowie Gesamtabiturnote auf. Sie schneiden schlechter im Mathematiktest ab und bewerten diesen als schwieriger, schätzen ihre mathematischen Kompetenzen und Kenntnisse geringer ein und ihre Selbstwirksamkeitserwartung ist geringer. Die Selbsteinschätzung hinsichtlich der Mathematikkenntnisse und -kompetenzen sowie der Studierfähigkeit und Relevanz von Statistik verändert sich bei den Vorkursteilnehmer/innen im Gegensatz zu Nicht-Teilnehmer/innen positiv.

Langfristige Effekte der Teilnahme an mathematischen Vorkursen auf die Studienzufriedenheit und -bewältigung

Di 8:30
HS5

Sven Brinkmann*, Kim Austerschmidt
Universität Bielefeld, Deutschland

In Fächern mit mathematischen Lehrinhalten haben Studierende häufig Probleme diese zu bewältigen und es kommt nicht selten zu einer Studienzeitverlängerung oder sogar zu einem vorzeitigen Abbruch des Studiums. Um die Studierenden bei der Bewältigung der mathematischen Anforderungen zu unterstützen, werden an der Universität Bielefeld nicht nur im Fach Psychologie, sondern auch in anderen Fächern mathematische Vor- bzw. Aufbaukurse angeboten. 160 Studierenden des 1. und 5. Fachsemesters verschiedener Bachelorstudiengänge mit einem solchen Kursangebot wurde ein Online-Fragebogen vorgelegt. Es wurden u.a. demographische Merkmale, die Abiturnote, die Teilnahme am Vor- bzw. Aufbaukurs (ja/nein), die wahrgenommene Relevanz von Mathematik im Studium, die Beurteilung der eigenen mathematischen Kompetenz, die Studienzufriedenheit und der Studienerfolg gemäß den Vorgaben, das Studium in Regelstudienzeit zu absolvieren (nur für Studierende des 5. Fachsemesters) erfasst. Es wurde analysiert, welche Studierenden einen Vor- bzw. Aufbaukurs nutzten und welche langfristigen Effekte die (Nicht)Teilnahme am Vorkurs auf die Studienzufriedenheit und den Studienerfolg hat. Im Rahmen des Vortrags werden die Ergebnisse der Studie vorgestellt und Implikationen im Hinblick auf die Relevanz der Unterstützung von Psychologiestudierenden zu Studienbeginn diskutiert.

Pearl's Causality and Longitudinal Data

Di 10:15
HS2

Christian Gische*, Manuel Voelkle
Humboldt-Universität zu Berlin, Deutschland

During the last two decades a comprehensive theory of causal inference based on directed acyclical graphs (DAGs) has been developed (Pearl, 2000). Furthermore, it is well known that non-recursive structural equation models can be represented as DAGs and thus can be analyzed using Pearl's framework of causal inference. This not only applies to cross-sectional models but also to dynamic longitudinal models, such as vector autoregressive processes, which can be readily formulated as DAGs, as long as they do not contain feedback loops. Despite the crucial role of time for the study of causal effects, however, surprisingly little attention has been put on integrating longitudinal models of change into Pearl's approach to causal inference. In this presentation, we apply Pearl's general causal effect formula to time series models, as frequently employed in the study of human development. Based on previous work, we define a set of assumptions under which cross-lagged coefficients warrant a causal interpretation in line with Pearl's framework of causal inference. The formal and substantive meaning of these assumptions will be demonstrated and evaluated in the context of modelling human behavior and development. Causal effects for different research designs (i.e. unique intervention vs. repeated interventions) will be derived. In practical applications some of the assumptions on which our arguments are build on are likely to be met while others might be violated. We demonstrate the consequences of violating one or more core assumptions. Furthermore we discuss strategies for ensuring valid causal inference under more general circumstances.

Individual Parameter Contributions Regression for Longitudinal Data

Di 10:15
HS2

Manuel Arnold*[1], Daniel Oberski[2], Manuel Voelkle[1]

1: Humboldt University of Berlin, Germany; 2: Utrecht University, Netherlands

Structural equation modelling (SEM) is a broad and flexible framework, frequently applied in the analysis of longitudinal data. In standard SEM, model parameters are assumed to be invariant across individual. In many longitudinal applications, however, interindividual or subgroup differences are the rule rather than the exception. During the last decades, a number of different SEM extensions have been developed to account for heterogeneity. One of those methods is "Individual parameter contributions" (IPC) regression, recently proposed by Oberski (2013) to identify and estimate subgroup differences in structural equation model parameters. IPC regression is conducted in three steps. First, a theory-guided structural equation model is fitted using the complete sample. Second, the contributions of every individual to the model parameters are calculated based on the first-step model. Third, these contributions are regressed on a set of explanatory variables, such as grouping variables, which may have caused heterogeneity. IPC regression stands out due to a comparatively small computational burden and encompasses the use of multiple discrete and continuous explanatory variable to identify heterogeneity in model parameters. This talk aims to illustrate how IPC regression can be used as a data-driven procedure to detect and provide rough estimates of subgroup differences in contemporary longitudinal structural equation models, focussing on cross-lagged autoregressive panel models in discrete and continuous time. It will be shown that IPC regression may underestimate subgroup differences in models strongly affected by heterogeneity. Bias correction approaches as well as the model-specific power of the method to identify subgroup differences will be discussed.

Schätzung von Wachstumsparametern ohne Verwendung von longitudinalen Informationen: Das Cohort Growth Model

Di 10:15
HS2

Kevin Fischer*[1], Andreas Klein[1], Jost Reinecke[2]

1: Goethe Universität Frankfurt, Deutschland; 2: Universität Bielefeld, Deutschland

Die Datengrundlage für die Analyse von Wachstumsprozessen sind in den Sozialwissenschaften üblicherweise Längschnittsdaten. Die Erhebung von Längschnittsdaten verlangt, dass dieselben Personen in Bezug auf ein bestimmtes Merkmal wiederholt in einem Zeitintervall gemessen werden. Erfolgt die Messung an diskreten Punkten im Zeitintervall, werden häufig Längschnittsdaten mit Hilfe eines "Latenten Wachstumsmodells" ausgewertet. Der Vorteil dieser Wachstumsmodelle ist, dass sie im Gegensatz zu Standardverfahren, wie ANOVA oder Regression, neben Veränderungen in den Mittelwerten auch intra- und interindividuelle Entwicklungsprozesse erfassen. Um die Modellidentifikation sicherzustellen, werden Informationen derselben Person benötigt, die aus unterschiedlichen Punkten auf dem Zeitintervall gewonnen wurden. Stehen für die Analyse Daten von unterschiedlichen Personen zur Verfügung, die jedoch aus verschiedenen Geburtskohorten stammen und an nur einem Punkt auf dem Zeitintervall gewonnen wurden (Querschnittsdaten), lassen sich Wachstumsparameter aufgrund von fehlender empirischer Information nicht schätzen. In diesem Fall werden auf Standardverfahren wie Regression und Varianzanalyse zurückgegriffen, in denen Informationen über interindividuelle Unterschiede verloren gehen. Das Cohort Growth Model erlaubt die Schätzung aller Parameter eines latenten Wachstumsmodells basierend ausschließlich auf Querschnittsdaten. In einer Simulationsstudie wurde die Verzerrung in den Parameterschätzungen getestet und die statische Teststärke ermittelt unter Voraussetzungen, welche sich häufig in empirischen Längschnittsstudien finden lassen. Die Ergebnisse zeigen, dass für unverzerrte Parameterschätzungen und ausreichende Teststärke mindestens vier Kohorten erhoben werden sollten und die Stichprobengröße in den jeweiligen Kohorten etwa N = 1000 betragen sollte. Abschließend wird diskutiert unter welche Voraussetzungen das Verfahren in der Empirie angewandt werden kann.

On the Definition of Latent-state-trait Models with Autoregressive Effects: Insights from LST-R theory

Di 10:15
HS2

Michael Eid*[1], Jana Holtmann[1], Philip Santangelo[2], Ulrich Ebner-Priemer[2]
1: Freie Universität Berlin, Deutschland; 2: Karlsruhe Institute of Technology, Deutschland

In particular in applications with short time lags, classical models of latent state-trait (LST) theory assuming that there are no carry-over effects between neighboring occasions of measurement are often not appropriate, and have to be extended by including autoregressive effects. However, the way how autoregressive effects should be defined is still an open question. Referring to a recently published revision of LST theory (LST-R theory), Steyer et al. (2015) state that the trait-state-occasion (TSO) model (Cole et al., 2005), one of the most widely applied LST models with autoregressive effects, is not an LST-R model, implying that proponents of LST-R theory might recommend not to apply this model. In this talk, we show that the TSO model can be defined on the basis of LST-R theory and some of its restrictions can be reasonably relaxed. The model is based on the idea that situational effects can change time-specific dispositions, and it makes full use of the basic idea of LST-R theory that dispositions to react to situational influences are dynamic and malleable. The latent variables of the model have a clear meaning that is explained in detail.

Dealing with a two-dimensional ability phenomenon when calibrating a test according to the Rasch model

Di 10:15
HS4

Klaus D. Kubinger*, Takuya Yanagida, Bettina Hagenmüller

University of Vienna, Faculty of Psychology, Division of Psychollogical Assessment and Applied Psychometrics, Österreich

Our example deals with a Memory test consisting eleven three-letter mindless syllables, which the testee has to learn by heart and then to reproduce exactly according to their sequence. Data included 5241 pupils. Apart from Andersen's Likelihood-Ratio test Rasch's graphical model check disclosed severe Rasch model violation: Both the last items are in relation to the other items more difficult for high scorers than for low scorers; for the items rather in the middle of the presentation sequence the situation is exactly reverse. The hypothesis emerged that a two-dimensional ability phenomenon occurs, that is testees apply two different strategies: One type focus on repeatedly recalling all items from the very beginning, hence the late items become more difficult due to less time left and the memory storage almost full; the other type does not insist as much in recalling former items, hence the late items do not have such a time and storage handicap but may replace former ones in the memory storage. This meets Formann's (1982, 1985) early approach of looking for those latent types for each of which the items match the Rasch model, although the item difficulty parameters' relations differ from type to type; this approach is in the meantime known as the so-called Mixed Rasch model (cf. Rost, 1990). Analyses confirmed our hypothesis. For an enhancement oriented psychological assessment it might be actually useful to determine that a testee belongs to type two: he/she then can be instructed to meet respective memory tasks more systematically.

Constraint Interaction Revisited: How to Interpret Factor Loadings under Alternative Scaling Methods

Di 10:15
HS4

Stefan Klößner*, Eric Klopp
Universität des Saarlandes, Deutschland

For achieving model identification, latent variables must be scaled. In the literature, several methods have been developed for scaling latent variables: the fixed marker method, the fixed factor method, and the effects coding method. When using the fixed marker method, model identification is achieved by setting one loading per latent variable to unity. Under the fixed factor method, all latent variables' variances are set to unity, while the effects coding method requires that the average of all loadings belonging to a latent variable must be unity. Constraint interaction denotes the phenomenon that the results of testing certain hypotheses about latent variables' loadings depend on which scaling method is employed. Examining the reasons behind constraint interaction, we find that the scaling methods interfere with the testing procedures because scaling methods implicitly determine which quantity loadings, or model parameters in general, actually estimate. As a consequence, using different scaling methods leads to different hypotheses that are being tested, none of which typically corresponds to the hypothesis that the researcher actually wants to test. Finally, in order to make researchers aware of the consequences of choosing a scaling method, we also develop new rules on how to interpret loadings and other model parameters when the fixed marker, fixed factor, or effects coding method is used for scaling the latent variables.

Ein Vergleich parametrischer und nicht-parametrischer Ansätze zur Prüfung des Rasch-Modells

Di 10:15
HS4

Rudolf Debelak*
Universität Zürich, Schweiz

In den vergangenen Jahren wurden eine Reihe von Modelltests für das Rasch-Modell vorgestellt. Bei diesen Modelltests lassen sich Tests erster Ordnung, welche die erwarteten und beobachteten Häufigkeiten für einzelne Items vergleichen, von Tests zweiter Ordnung unterscheiden, welche diese Häufigkeiten für Itempaare vergleichen. Ein bekanntes Beispiel für einen Test erster Ordnung basiert auf der LR-Statistik von Andersen (1973), während die M2-Statistik von Maydeu-Olivares & Joe (2005) zu einem Test zweiter Ordnung führt. Neben diesen parametrischen Modelltests wurden von Ponocny (2001) die T10- und T11-Statistiken für nicht-parametrische Modelltests erster und zweiter Ordnung vorgestellt. In einer Reihe von Simulationsstudien werden die LR-, M2-, T10- und T11-Statistiken in Hinsicht auf ihre Typ I-Fehlerrate und ihre Teststärke gegen verschiedene Alternativmodelle (2PL-Modell, Rasch-Modell mit Rateparameter, mehrdimensionales Rasch-Modell, Rasch-Modell mit Lerneffekt) verglichen. Die Ergebnisse zeigen, dass Tests zweiter Ordnung sensitiv gegen alle untersuchten Alternativmodelle sind, während Tests erster Ordnung nur gegen das 2PL-Modell und ein Rasch-Modell mit Rateparameter sensitiv sind. Die Bedeutung dieser Ergebnisse für die praktische Testanalyse wird diskutiert.

Zur Speed-Power-Problematik psychologischer Diagnostik: Kann uns das Diffusionsmodell weiterhelfen?

Di 10:15
HS5

Rainer Alexandrowicz*, Gula Bartosz, Saskia Donker, Lars Reich, Janis Gottstein
Alps-Adria-University Klagenfurt, Österreich

Die Ausrichtung gängiger psychologischer Testverfahren orientiert sich überwiegend an einer Erfassung der "Leistungsfähigkeit" einer getesteten Person hinsichtlich des zu erfassenden psychologischen Konstruktes. Damit zielt die Testung auf die "Power"-Komponente ab. Ein wichtiges psychometrisches Werkzeug zur Modellierung stellt dabei das Rasch-Modell (RM) samt seiner Derivate und Erweiterungen dar. In Kontrast zu dieser differentialpsychologisch orientierten Herangehensweise schlug Ratcliff (1978) das Diffusionsmodell (DM) vor, das einem allgemeinspychologisch-experimentellen Paradigma entspringt. Charakteristisch für dieses Modell ist die simultane Modellierung von sowohl richtig-falsch-Information der Antworten als auch der für jede Antwort benötigte Reaktionszeit (hier wird vor allem auf Reaktionen im Sekundenbereich abgezielt). Allerdings wird im ursprünglichen DM keine Differenzierung von Item- und Personenbezogenen Maßen vorgenommen (eine solche wäre durch entsprechendes Versuchsdesign vorzunehmen). Die wesentlich jüngere Erweiterung des DM zum sogenannten diffIRT-Modell (Tuerlinckx & De Boeck, 2005) nimmt jedoch genau diese Unterscheidung vor, indem zwei zentrale Modellparameter (boundary separation und drift diffusion) in eine personen- und eine itembezogene Komponente unterteilt werden. Hierin könnte ein Ansatzpunkt zur Anwendung im psychometrisch-diagnostischen Kontext liegen. Im Vortrag werden die verschiedenen Modellansätze vorgestellt und Studien präsentiert, in denen Fragen der Parameterschätzung und –vergleichbarkeit der unterschiedlichen Ansätze erörtert werden.

Vergleich von Parameterschätzungen mit dem Diffusionsmodell und dem diffIRT-Ansatz

Di 10:15
HS5

Anna Conci*, Saskia Donker, Bartosz Gula, Rainer Alexandrowicz
Alpen-Adria Universität Klagenfurt, Österreich

Das Ratcliff'sche Diffusionsmodell (DM; Ratcliff, 1978) stellt eine wichtige Erweiterung der Modellierung von Antwortprozessen für dichotome Entscheidungen unter gleichzeitiger Berücksichtigung der Reaktionszeit dar. Es beschreibt die getrennten Reaktionszeiten der beiden Antwortmöglichkeiten unter Annahme der vier Hauptparameter: boundary separation (a), bias (z), non-response time (tER) und drift parameter (nu) (sowie dreier weiterer Streuungsparameter). Molenaar, Tuerlinckx und van der Maas stellten 2015 die Erweiterung diffIRT vor, in der die boundary separation und der drift parameter jeweils durch eine Funktion einer personen- und einer itembezogenen Komponente ersetzt werden. Hierbei wurden zwei unterschiedliche Reparametrisierungen (D-diffusion und Q-diffusion) gewählt. Damit sollte eine differenzierte Modellierung ermöglicht werden, die laut Autoren wahlweise eher für Leistungs- bzw. Persönlichkeitstests geeignet sein sollen. In der vorliegenden Studie wird anhand publizierter Datensätze, auf die bislang nur das DM bzw. andere Auswertungsverfahren angewendet wurden, das DM und das diffIRT (in beiden Varianten) simultan untersucht. Je nach Versuchsdesign sollten sich Effekte primär in den Personen- oder den Itemkomponenten widerspiegeln, wobei auch die Unterscheidung in D- und Q-diffusion eine besondere Rolle einnimmt. Damit soll gezeigt werden, ob das diffIRT generell sowie spezifisch für die beiden diagnostischen Zielsetzungen Leistungs- vs. Persönlichkeitsdiagnostik geeignet ist und eine differenziertere Betrachtung gegenüber dem ursprünglichen DM tatsächlich gegeben ist.

Validierung der personen- und itembezogenen Komponenten des boundary-separation Parameters im diffIRT

Di 10:15
HS5

Lars Reich*, Rainer W. Alexandrowicz
Universität Klagenfurt, Österreich

Das Diffusionsmodell (DM) nach Ratcliff (1978) ermöglichte die simultane Modellierung der Richtig-Falsch-Information und der hierfür benötigten Bearbeitungszeit. Molenaar, Tuerlinckx und van der Maas (2015) stellten eine Erweiterung des DM (diffIRT) unter Berücksichtigung des Grundprinzips der Trennung von item- und personenbezogenen Parametern der Item Response Theorie vor. Durch diese Kombination gelingt eine differenziertere Modellierung des Antwortprozesses. In der vorliegenden Studie soll der itemboundary Parameter einer näheren Betrachtung unterzogen werden mit gleichzeitiger Überprüfung der Differenzierung durch das Q- beziehungsweise D-Diffusionsmodell, welche besonders für Persönlichkeits-, beziehungsweise Leistungstests eignen sollen. Dazu wurde ¡strong¿e¡/strong¿in experimentelles Design mit einer studentischen Stichprobe durchgeführt. Entsprechend der Modellannahme müsste sich Zeitdruck bei der Bearbeitung einer Aufgabe im item-boundary Parameter (a) des DM zeigen. Im diffIRT sollte eine weitere Differenzierung auf Itemebene nachweisbar sein. Die Relevanz unserer Forschungsergebnisse bezieht sich auf Testverfahren, in denen es um Entscheidungsprozesse geht, diese in der kognitiven Psychologie häufig zur Anwendung kommen. Unsere empirischen Ergebnisse geben weiteres darüber Aufschluss, ob das Modell von Molenaar, Tuerlinckx und van der Maas adäquat zur Beschreibung des antwortgenerierenden Prozesses der Daten herangezogen werden kann.

Zur Validierung des Vigilanzparameters im diffIRT

Di 10:15
HS5

Rainer Alexandrowicz*, Janis Gottstein
Alpen Adria Universität Klagenfurt, Österreich

Während das Diffusionsmodell nach Ratcliff (1978) die Reaktionszeiten dichotomer Antworten unter Verwendung der vier Hauptparameter boundary separation (a), bias (z), non-response time (tER) und drift parameter (nu) (sowie dreier weiterer Streuungsparameter) modelliert, nimmt das diffIRT-Modell (Molenaar, Tuerlinckx & van der Maas, 2015) eine Reparametrisierung von a und nu in jeweils eine personen- und eine itembezogene Komponente vor. Darüber hinaus wird zwischen einer D- und einer Q-Reparametrisierung unterschieden, die in einem Fall eher für Leistungstests und im anderen Fall für Persönlichkeitsverfahren geeignet sein sollen. Jede dieser vier Komponenten wird eine spezifische psychologische Bedeutung zugeschrieben: Die Personenkomponente von nu spiegelt die Informationsverarbeitungsgeschwindigkeit wider, die Itemkomponente die reine Itemschwierigkeit; die Personenkomponente von a die Vigilanz und die Itemkomponente den Zeitdruck. Die vorliegende Studie widmet sich der Parametervalidierung des so definierten Vigilanzparameters. In einem Experiment wird ein "klassischer" Vigilanztest durchgeführt und entprechend der traditionellen Richtlinien ausgewertet. Parallel werden die Parameter des diffIRT geschätzt und hinsichtlich der spezifischen Übereinstimmung des Vigilanzparameters untersucht. Damit soll überprüft werden, ob das diffIRT eine mögliche Erfassung der Vigilanz ohne der (vor allem zeit-) aufwendigen Erfassung, wie sie in klassischen Verfahren erforderlich ist, erlaubt.

Model-based recursive partitioning of psychometric models: A data-driven approach for detecting heterogeneity in model parameters

Di 12:05
Audimax

Carolin Strobl*
Universität Zürich

Model-based recursive partitioning is a flexible framework for detecting differences in model parameters between two or more groups of subjects. Its origins lie in machine learning, where its predecessor methods, classification and regression trees, had been introduced around the 1980s as a nonparametric regression technique. Today, after the statistical flaws of the early algorithms have been overcome, their extension to detecting heterogeneity in parametric models makes recursive partitioning methods a valuable addition to the statistical "toolbox" in various areas of application, including econometrics and psychometrics. This talk gives an overview about the rationale and statistical background of model-based recursive partitioning in general and in particular with extensions to psychometric models for paired comparisons as well as item response models. In this context, the data-driven approach of model-based recursive partitioning proves to be particularly suited for detecting violations of homogeneity or invariance, such as differential item functioning, where we usually have no a priori hypotheses about the underlying group structure.

Bayesian two-level mo del for partially ordered repeated responses

Di 14:25
HS2

Xiaoqing* Wang, Xiangnan Feng, Xinyuan Song
The Chinese University of Hong Kong, Hong Kong (VR China)

We consider a multivariate generalized latent variable model to investigate the effects of observable and latent explanatory variables on multiple responses of interest. Various types of correlated responses, such as continuous, count, ordinal, and nominal variables, are considered in the regression. A generalized confirmatory factor analysis model that is capable of managing a variable-type data is proposed to characterize latent variables via correlated observation indicators. In the context of the proposed model, we develop a multivariate version of the Bayesian adaptive least absolute shrinkage and selection operator procedure, which is implemented with a Markov chain Monte Carlo (MCMC) algorithm in a full Bayesian context. The empirical performance of the proposed methodology is by a simulation study. An application of the proposed method to a study of adolescent substance abuse based on the National Longitudinal Survey of Youth is presented.

Semiparametric latent variable models for multivariate censored data with Bayesian analysis

Di 14:25
HS2

Ming Ouyang*, Xinyuan Song
The Chinese University of HongKong, Hong Kong S.A.R. (China)

In this study, we aim to develop a semiparametric latent variable model to analyze multivariate censored data. In the proposed model, time-to-event responses are regressed on latent variables and covariates through an additive structural equation formulated by a series of unspecified smooth functions. The Bayesian approach, along with Markov chain Monte Carlo algorithm and Bayesian P-splines technique, is implemented to estimate the unknown parameters and smooth functions. The empirical performance of the proposed methodology is evaluated by simulation studies. An application to a study of cardiovascular diseases is presented.

Gaussian Process Panel Modeling – A unification of longitudinal modeling approaches

Di 14:25
HS2

Julian Karch*[1], Andreas Brandmaier[1], Manuel Voelkle[2]

1: Max-Planck-Institut für Bildungsforschung, Deutschland; 2: Humboldt-Universität zu Berlin, Deutschland

In machine learning, Bayesian Gaussian process models are known as a flexible modeling technique for time-series data. In psychology, Structural Equation Modeling (SEM) is a popular modeling technique for the analysis of longitudinal panel data. We integrate ideas from both approaches to create a novel longitudinal panel data analytic method that we call Gaussian Process Panel Modeling (GPPM). The main advantage of GPPM is its flexibility, with most popular modeling approaches for longitudinal data being a special case of GPPM. This not only includes SEM and hierarchical linear modeling, but also state-space modeling in its time-discrete and time-continuous variant, as well as generalized additive models. GPPM allows specifying all aforementioned model families in a consistent modeling language. The generality of GPPM enables specifying novel models by either combining approaches from different traditions or by relying on the wealth of models used in Gaussian process time series modeling. As an example, we present the exponential squared model, which implements the generic assumption of smooth process trajectories. We demonstrate the utility of GPPM using an empirical example. In summary, GPPM is a unification as well as an extension of existing panel data modeling techniques.

Emotionales Erleben und tägliche Ereignisse im dynamischen Zusammenspiel: Modellierung durch Gaußsche und Poisson Vektor Autoregressionen.

Di 14:25
HS2

Janne Adolf*[1], Julian Karch[1], Annette Brose[2,1], Florian Schmiedek[3,1], Manuel Voelkle[2,1]

1: Max-Planck-Institut für Bildungsforschung, Berlin, Deutschland; 2: Humboldt-Universität zu Berlin, Berlin, Deutschland; 3: Deutsches Institut für Internationale Pädagogische Forschung, Frankfurt am Main, Deutschland

Mikro-längsschnittliche psychologische Studien versprechen Einblicke in die Dynamiken des alltäglichen Erlebens und Verhaltens und erfreuen sich zunehmender Beliebtheit. Die Analyse der resultierenden Daten erfolgt oft über dynamische Modelle, die zeitliche Zusammenhänge zwischen interessierenden Variablen auf intra-individueller Ebene abbilden und so die beteiligten psychologischen Prozesse annähern. Aktuell wird die Modellierung emotionaler Dynamiken besonders beachtet. Darüber, dass die zugrunde liegenden Prozesse der Emotionsregulation allerdings nur unter Berücksichtigung von kontextuellen Bedingungen verstanden werden können, besteht in der theoretischen Literatur Konsens. In diesem Beitrag erweitern wir deshalb bestehende Modellierungsansätze, in dem wir Veränderungen in kontextuellen Bedingungen, hier speziell stressigen Ereignissen, explizit dynamisch modellieren. Ein solcher Ansatz bildet nicht nur zeitliche Zusammenhänge des Alltagskontexts ab, er erlaubt auch die gleichzeitige Betrachtung emotionaler und kontextueller Prozesse, gibt Aufschluss über deren dynamisches Zusammenspiel und zeichnet damit ein vollständigeres Bild emotionalen Alltagslebens. Für die dynamische Analyse von Ereigniszeitreihen sind außerhalb der Psychologie Poisson Autoregressive Modelle entwickelt worden, die der charakteristischen Verteilungsform von Ereignissen Rechnung tragen. Wir modifizieren diese Modelle zur Analyse durchschnittlicher intra-individueller Dynamiken stressiger Ereignisse, für sich und im Zusammenspiel mit affektivem Erleben. Die Modellschätzung erfolgt simulationsbasiert über Markov-Ketten-Monte-Carlo-Verfahren im Rahmen Bayesscher Statistik. Wir illustrieren die Modelle anhand von mikro-längsschnittlichen Daten der COGITO Studie (Schmiedek, Bauer, Lövdén, Brose, & Lindenberger, 2010).

Introducing approxbayes – a software for approximate Bayes factors

Di 14:25
HS4

Ulf Mertens*, Stefan Radev, Andreas Voß
Universität Heidelberg, Deutschland

The Bayes factor (BF) has become a popular alternative to classical hypothesis testing. BFs allow for the comparison among hypotheses by taking into account the uncertainty in the parameters of the models. One important advantage of the BF is the ability to express support for both alternative and the null hypothesis. Although BFs are favorable in many ways, they are nonetheless difficult to compute for non-standard tests. Approximate Bayesian Computation (ABC) offers a way to approximate BFs for arbitrarily specified models. Due to this flexibility, ABC is a promising tool for BF computation if it cannot be computed using standard software. We introduce our software for ABC model comparison named approxbayes and show how to compare two multinomial processing tree models.

Onyx: Ein grafisches Modellierungswerkzeug für Strukturgleichungsmodelle

Di 14:25
HS4

Andreas M. Brandmaier*[1], Timo von Oertzen[2]

1: Max-Planck-Institut für Bildungsforschung, Deutschland; 2: Universität der Bundeswehr München, Deutschland

Ωnyx: Ein grafisches Modellierungswerkzeug für Strukturgleichungsmodelle In diesem Beitrag geben wir einen Überblick über Ωnyx, ein frei verfügbares, grafisches Werkzeug zur Modellierung und Schätzung von Strukturgleichungsmodellen. Die intuitive grafische Oberfläche erlaubt das schnelle Erstellen von Strukturgleichungsmodellen in Form von Pfaddiagrammen und unterstützt sowohl Multigruppenmodelle als auch Definitionsvariablen. Darüberhinaus verfügt Ωnyx über ein Maximum-Likelihood-Schätzverfahren, dass schwarmbasiert arbeitet, ohne Startwerte auskommt und multiple Minima finden kann. Ωnyx-Modelle können grafisch in vektor- und bitmaporientierte Formate exportiert werden. Darüberhinaus erzeugt Ωnyx aus dem Pfaddiagramm automatisch Modellspezifikationen für gängige Modellierungsprogramme wie Mplus, lavaan, sem, oder OpenMx. Modelle, die in lavaan und OpenMx spezifiert wurden, können von Ωnyx gelesen und dargestellt werden. Likelihood-Quotienten-Tests können direkt in Ωnyx mittels grafischer Testkanten zwischen Modellen ausgeführt werden. Ωnyx ist durch seine Anschaulichkeit ideal für die Lehre geeignet und integriert sich als Bindeglied in die Landschaft existierender Lösungen im wissenschaftlichen Modellierungsalltag.

Dirichlet Process Clustering in Onyx

Di 14:25
HS4

Timo von Oertzen*[1], Thomas Glassen[1], Andreas Brandmaier[2]

1: UniBW München, Deutschland; 2: Max-Planck-Institut für Bildungsforschung, Berlin, Deutschland

Miss-specification in SEM often is created by clusters in the data which each in itself fit the SEM perfectly. Such clusters of participants can be identified either with supervised methods based on covariates, e.g. with SEM-Trees, or with unsupervised clustering methods. Sometimes, the clustering itself has a prior distribution, for example a 'rich gets richer' property in which cluster that already are large have higher probability to grow compared to smaller cluster. Dirichlet Clustering allows to include such priors, also removing the limitation of many clustering methods that the number of clusters must be specified beforehand. In this presentation, we present how Dirichlet Clustering can be used for SEM and demonstrate a new feature in the graphical SEM software Onyx that allows to apply Dirichlet Clustering on data for pre-specified SEM, using some longitudinal model as examples for clustering based on more complex SEM.

Observation Oriented Modeling (OOM) – Ein alternativer Ansatz für statistische Modellierung

Di 14:25
HS4

Sebastian Sauer*, Karsten Lübke
FOM Hochschule, Deutschland

Observation Oriented Modeling (OOM; Grice, 2011) ist ein alternativer Ansatz der Datenanalyse. Zentrales Merkmal von OOM ist, dass Theorien beobachtungsbezogen verstanden werden: Anstatt des durchschnittlichen Verhaltens auf Ebene der Variablen wird das generelle Verhalten der Beobachtungen untersucht. Dafür wird die Übereinstimmung von beobachteten und erwarteten bzw. prognostizierten Werten in einer Kreuztabelle betrachtet. Dabei werden Gedanken der Configural Frequency Analysis (Stemmler, 2014) und Ordinal Pattern Analysis (Thorngate & Ma, 2016) mit elementaren Grundgedanken des Statistical Learnings (Hastie, Tibshirani, & Friedman, 2009) kombiniert. OOM bietet einen "frischen Blick" auf statistische Modellierung, obwohl mit bekannten Ansätzen argumentiert wird. Als Weiterführung des bestehenden Konzepts von OOM stellen wir eine Modellanpassung auf Basis der Moore-Penrose-Inverse vor (Ben-Israel & Greville, 2003), die dem ursprünglichen Ansatz (Grice, 2011) überlegen ist. OOM ist einfach und voraussetzungarm; auf die Annahme von metrischem Niveau wird genauso verzichtet wie auf auf Verteilungs- und Zusammenhangsformen, da nur mit Indikatormatrizen argumentiert wird. Im ersten Teil des Vortrags wird die Idee von OOM vorgestellt; Ansatz, Stärken und Grenzen der Methode werden diskutiert. Im zweiten Teil wird OOM beispielhaft auf eine publizierte Studie angewendet.

Modeling responses and response times in rating scales

Di 14:25
HS5

Jochen Ranger*[1], Jörg-Tobias Kuhn[2]

1: Martin-Luther-Universität Halle-Wittenberg, Deutschland; 2: Westfälische Wilhelms-Universität Münster

In this manuscript, a new model is proposed for the responses and the response times in attitudinal or personality inventories with graded response format. The model is based on the lognormal race model (Heathcote & Love, 2012) and assumes two accumulators that aggregate evidence in favor of and against the statement claimed by an item of the inventory. The accumulator that first reaches a response threshold determines the direction of the response (agreement/disagreement). The strength of the response, which is reflected in the choice of a graded response option, is a function of the difference between the two accumulators when responding. By relating the accumulators to latent traits, the model can be embedded into a latent trait model that accounts for individual differences. The model can be fit to data with marginal maximum likelihood estimation. A test of model fit is described and it is shown how the model can be used for attitudinal and personality assessment. Finally, the application of the model is demonstrated with a real data set.

Erweiterungen des Diffusionsmodells: Adaptiven Schwellen und Lévy-Flight-Modelle

Di 14:25
HS5

Andreas Voß*
Universität Heidelberg, Deutschland

Das Diffusionsmodell ist eines der bekanntesten mathematischen Modelle der kognitiven Psychologie. Typischerweise wird dieses Modell für die Auswertung von Reaktionszeitdaten aus einfachen Wahrnehmungsaufgaben mit sehr kurzen Antwortzeiten angewandt. Erste Versuche, auch komplexere Aufgaben mit längeren Antwortlatenzen mit dem Diffusionsmodell auszuwerten, lieferten erfolgversprechende Ergebnisse. Dennoch muss gerade bei solchen Aufgaben sorgfältig geprüft werden, ob die Voraussetzungen der Diffusionsmodellanalyse gegeben sind: Insbesondere stellt sich die Frage ob Versuchspersonen ihre initialen Entscheidungskriterien aufrechterhalten, wenn nach einigen hundert Millisekunden keine Antwort offensichtlich ist. Eine Anpassung der Antwortkriterien kann in die Diffusionsmodellanalyse durch die Annahme dynamischer Schwellen einbezogen werden. Eine andere Möglichkeit, mit der Versuchspersonen reagieren könnten, wenn ein Stimulus wenig objektive Information bietet, ist eine sprunghafte Veränderung der subjektiven Informationssammlung ("jumping to conclusion"). Mathematisch kann ein solcher Prozess als Lévy-Flight beschrieben werden, also als ein Informationssammlungsprozess bei dem Zufallseinflüsse nicht einer Normalverteilung sondern einer Verteilung mit häufigeren extremen Ereignissen folgen (z.B. Lévy, Cauchy). Im vorliegenden Beitrag wird der Fit von sechs unterschiedliche Varianten solcher Erweiterungen des Diffusionsmodells anhand von Daten einer einfachen und einer komplexeren Klassifikationsaufgabe (jeweils unter Geschwindigkeits- vs. und Genauigkeits-Instruktionen) verglichen.

Speed-Accuracy Manipulations: Are Effects in Non-Decision Time attributable to a Lack of Discriminant Validity of the Manipulation or to Trade-Offs in Parameter Estimation?

Di 14:25
HS5

Veronika Lerche*, Andreas Voß
Heidelberg University, Deutschland

The validity of the parameters of the diffusion model (Ratcliff, 1978) has been tested with several experimental studies. In these studies, speed-accuracy manipulations have been employed to examine the validity of the threshold separation parameter. In the accuracy condition, participants—as expected—have higher threshold separations than in the speed condition. However, also effects in non-decision time have been observed: In the accuracy condition, the non-decision time estimates are higher than in the speed condition. There are mainly two possible reasons for this finding: (1) Trade-offs in parameter estimation lead to the unexpected effect in non-decision time or (2) the speed-accuracy manipulations do not only influence the threshold settings but also non-decisional processes (e.g., the motoric response execution). To disentangle these two possible accounts, we conducted a set of simulation studies that were based on parameter values observed in experimental studies. The simulation studies revealed that the unexpected difference in non-decision time is attributable to a lack of discriminant validity of the speed-accuracy manipulations. These manipulations seem to influence not only the threshold separation but also non-decisional processes.

Validität von Diffusionsmodellparametern in elementaren kognitiven Aufgaben

Di 14:25
HS5

Florian Schmitz*, Oliver Wilhelm
Universität Ulm, Deutschland

Das Anliegen der vorgestellten Analysen war die Untersuchung der psychometrischen Validität von Parameterschätzern des Diffusionsmodells. Dafür wurden Reaktionszeitdaten aus elementaren kognitiven Aufgaben aus zwei multivariaten Studien (N=200; N=120) mit dem Diffusionsmodell analysiert. Als Korrelate wurden Arbeitsgedächtniskapazität, fluide Intelligenz und mit konventionellen Papiertests erfasste Mentale Geschwindigkeit berücksichtigt. Korrespondierende Parameterschätzer der Driftrate, der Nicht-Entscheidungszeit und des Antwortkriteriums waren moderat bis hoch über die Aufgaben korreliert, wobei die Zusammenhänge bei Aufgaben desselben Typs höher ausfielen. Konfirmatorische Modellvergleiche für die Parameterschätzer als beobachtete Indikatoren bestätigten eine gute Passung von Bifaktormodellen mit jeweils einem breiten g-Faktor für jeden Parametertyp sowie methodenspezifischen genesteten Faktoren. In Übereinstimmung mit früheren Ergebnissen zeigte der allgemeine Driftratenfaktor eine klare Validität für kognitive Leistung, während die anderen Parameterfaktoren keine bedeutsamen Relationen mit Leistungs- oder mit Persönlichkeitsvariablen aufwiesen. Zusätzlich wurden die Diffusionsmodellanalysen mit konventionellen reaktionszeit- und fehlerbasierte Scores sowie andere Parametrisierungen von Reaktionszeitverteilungen (bspw. Ex-Gauß-Modell) verglichen. Hierbei zeichnete sich das Diffusionsmodell durch seine theoretische begründeten Parameter und seinen Umgang mit individuellen Unterschieden in der Balance von Geschwindigkeit und Akkuratheit in der Aufgabenbearbeitung aus. Implikationen der Befunde für die Modellierung und Diagnostik von Leistung in elementaren kognitiven Aufgaben werden diskutiert.

Using response time models to investigating the mechanism of missing values at the end of a test

Di 16:10
HS2

Steffi Pohl*[1], Esther Ulitzsch[1], Matthias von Davier[2]

1: Freie Universität Berlin, Deutschland; 2: National Board of Medical Examiners, Philadelphia, USA

In large-scale assessments tests are often presented with a time limit and missing values due to not reaching the end of the test result. These are typically non-ignorable and may have a considerable impact on analyses results. Integrating research on response time modeling with research on modeling missing responses, we recently proposed using response time models to account for missing values due to time limits in the test. In this study, we will use these models to investigate the mechanisms underlying missing values at the end of the test in empirical data. We use competence data of PISA 2015, which have been assessed via CBA. Our results indicate that different mechanisms may be present that result in missing values at the end of the test. We discuss possible approaches that may represent these mechanisms and that may be used to account for them in data analyses.

A Dynamic Response Time Model for Time-Limited Tests

Di 16:10
HS2

Esther Ulitzsch*[1], Steffi Pohl[1], Matthias von Davier[2]

1: Freie Universität Berlin, Deutschland; 2: National Board of Medical Examiners, USA

Recently developed approaches for the joint modeling of responses and response times (van der Linden, 2007) assume stationarity of both ability and speed with an assumed application to data stemming from tests with a generous time limit. These assumptions are violated whenever examinees encounter tight time restrictions and are either running out of time, or perceive to do so. Building on previous research on dynamic processes, we present a dynamic extension for time-limited tests of van der Linden's response time model which would allow for varying speed resulting from examinees monitoring and adjusting their pace. We a) introduce a dynamic response time model for time-limited tests in its most general - but still conceptual - form and b) discuss special cases of the general dynamic response time model. Parameter recovery of the proposed model is investigated within a simulation study. An illustration of the model by means of an application to real data is provided.

The role of copula for reaction time modeling

Di 16:10
HS2

Hans Colonius*
Universität Oldenburg, Deutschland

An n-copula is an n-variate distribution function with univariate margins uniformly distributed on [0, 1]. Sklar's theorem shows that copulas remain invariant under strictly increasing transformations of the underlying random variables. It is possible to construct a wide range of multivariate distributions by choosing the marginal distributions and a suitable copula. This way, it is possible to specify the multivariate dependence structure of a set of random variables without specifying the marginal distributions. Here we demonstrate how the concept of copulas can be drawn upon to develop stochastic models of reaction time with a minimal number of parameters. Several examples from reaction time modeling (redundant signals task, stop sign task, multisensory integration) will be presented.

Why we cannot prove theories through experiments (or any other method): Underdetermination, the problem of induction, falsificationism, and the question why they are so rarely talked about in psychology

Di 16:10
HS4

Peter Holtz*

Leibniz Institut für Wissensmedien Tübingen, Deutschland

There are several fundamental epistemological reasons why we cannot prove theories through experiments or any other research method. For example, even if all available evidence as of now is in line with specific predictions of our theory, we do neither know whether all future observations will support our theory as well, nor whether there is no hitherto unknown theory that explains the phenomena in question in a "better" (e.g., more parsimonious) way than our theory - this is known as the "problem of induction". Furthermore, we cannot in a strict sense test specific hypothesis that are derived from our theory in isolation, because the validity of an experiment (or any other research method) depends on an infinite number of "auxiliary hypotheses" as well; these auxiliary hypothesis range from rather trivial assumptions (e.g., the measurement instrument "worked") to more intricate methodological premises (e.g., the statistics that we used "work" in a mathematical sense and are applicable to the current case) and theoretical conjectures (certain claims of other researchers, on whose work the experiment is based, were correct) - this is known in epistemology as the underdetermination problem or the "Duhem-Quine Thesis". Unfortunately, these and other epistemological fundamentals (such as Popper's critical rationalist approach as one answer to the aforementioned problems) are rarely talked about in texts on research methods in psychology (Holtz & Monnerjahn, 2017), although discussing epistemological principles in more depth could potentially help psychology to move forward in view of the recent "crisis of confidence".

Stichprobengröße und Stichprobenziehung in nicht-randomisierten Studien

Di 16:10
HS4

Philipp Mayring*
Alpen-Adria Universität Klagenfurt, Österreich

Die Stichprobenqualität ist eines der zentralsten Gütekriterien in empirischer Forschung, sie ist der Ansatzpunkt zur Generalisierung der Ergebnisse. Traditionelle quantitative Studien berufen sich hier auf das Modell der Repräsentativität per Zufallsauswahl. Die gängigen inferenzstatistischen Verfahren der Datenauswertung setzen eigentlich randomisierte Stichproben voraus. Allerdings erfüllen in der Praxis nur sehr wenige psychologische Studien diese Voraussetzung zur Gänze (Harden, 2009). So verwenden in psychologischer Forschung viele Studien vermischte Sampling-Strategien. Für qualitativ orientierte Studien haben Onwuegbuzie & Leech (2005) 19 verschiedene Sampling-Strategien in der Literatur gefunden, darunter die verbreiteten purposive sampling und theoretical sampling. Die jeweils spezifischen Schwächen dieser Strategien werden an einer Beispieluntersuchung diskutiert. Die Stichprobengröße stellt ein weiteres fundamentales Problem psychologischer Forschung dar. Nur eine Minderzahl erfüllt hier alle Kriterien (Maxwell, 2004). In qualitativ orientierten Studien werden hier meist nur Faustregeln gehandelt, und von Autor zu Autor sehr verschiedene Angaben gemacht, beispielsweise für Fallanalyse 1 bis 15 Fälle (Onwuegbuzie & Leech, 2007). Andererseits sind hier die Aussageformen bei den Ergebnissen meist vorsichtiger. Entlang eines Datensatzes wird diskutiert, welche Arten von Aussagen (explorativ, deskriptiv, ...) in Abhängigkeit von der Stichprobengröße sinnvoll gemacht werden können und welche Einflussfaktoren hier eine Rolle spielen. Der Beitrag soll zeigen, dass gerade in Studien mit nicht-randomisierten Samples eine äußerst sorgfältige und begründete Stichprobenplanung notwendig ist.

Inference Statistic: An uncertain tool to explore an unknown territory - a simulation study

Di 16:10
HS4

Antonia Krefeld-Schwalb*[1], Frank Zenker[2], Erich Witte[3]

1: University of Geneva, Schweiz; 2: University of Konstanz, Germany; 3: University of Hamburg, Germany

The empirical sciences aim at developing and testing informative theories and corresponding hypotheses. Null Hypothesis Significance Testing NHST has nonetheless remained the main statistical tool, although it broadly serves to draw inferences from data, rather than provide an optimal tool to test hypothesis. In fact, NHST not only provides a bare minimum of relevant information about the corroboration of a focal hypothesis, questionable research practices also further inflate the probability of drawing erroneous inferences (e.g., Type I errors), while under-powered studies additionally increase the probability of Type II errors. To address these shortcomings, we outline a research program strategy consisting of six consecutive steps leading from the discovery of an effect against a random model, to a statistical verification of the alternative hypothesis. Here, NHST (applied under unknown power) marks the first step of this strategy, and will at most lead to preliminary discoveries. Subsequent steps generally demand that data are of better induction quality before they serve to properly test hypotheses. Eventually, the null-hypothesis can be falsified and the alternative hypothesis be verified. As verification/falsification-criteria, we propose the ratio of the likelihood of the respective hypothesis along with the Wald-criterion, the power of a test divided by its level of significance. The final step of the strategy is reached if the ratio of the likelihood of the point-alternative hypothesis and the null-hypothesis exceeds the Wald-criterion. We support our proposal with a simulation study and present a hands-on web-application allowing researchers to quickly gauge the induction quality of data.

Optimal number of response categories for assessing job satisfaction: Results from an experimental online study

Di 14:25
HS5

Tanja Kutscher*, Claudia Crayen, Michael Eid
Freie Universität Berlin, Deutschland

When a trait or an attitude is measured with a rating scale consisting of many response categories, the psychometric quality of the data is often reduced. One reason lies in different response styles (e.g., an extreme response style or ignoring superfluous categories), which participants use when faced with an overwhelmingly large number of response options. In turn, a rating scale with an insufficient number of response categories can also fail to provide reasonable reliability. The objective of this study is to gather evidence for an optimal number of categories. We examine the extent of inappropriate scale usage in three conditions, which differed in the number of response categories used in the assessment of job satisfaction. A sample of American employees and employers (N=6999) filled out a 12-item online questionnaire on job satisfaction, with the number of response categories (4, 6, or 11) randomly assigned. Considering the three-dimensional structure of the job satisfaction measure, we applied a multidimensional extension of the restrictive mixed generalized partial credit model (rmGPCM) within each experimental condition. This resulted in similar configuration of three response-style classes in all conditions. Nevertheless, the proportion of respondents with inappropriate scale usage was lower in conditions with fewer response categories. Based on these results and a model-based reliability index, we conclude that a 6-point rating scale is most adequate for assessing job satisfaction, presumably extending to other aspects of life satisfaction.

Statistik lernen mit shiny-Apps – Was? Wie? Weitere Anregungen?

Di 16:10
HS5

Sonja Hahn*, Fabian Mundt, Kenneth Horvath
Pädagogische Hochschule Karlsruhe, Deutschland

Applets, in denen Sachverhalte aus der Statistik veranschaulicht werden können, gibt es seit vielen Jahren. Das R-Paket shiny (Chang, Cheng, Allaire, Xie, & McPherson, 2016) ermöglicht Lehrenden relativ einfach Apps zu entwickeln, welche für die eigene Veranstaltung oder eine bestimmte Zielgruppe zugeschnitten sind (vgl. Doi, Potter, Wong, Alcaraz, & Chi, 2016). Im vorliegenden Beitrag werden mehrere shiny-Apps vorgestellt, die im Rahmen Projektes Bildungsinitiative L^2 an der Pädagogischen Hochschule Karlsruhe entwickelt wurden, um Studierenden einen interaktiven Zugang zur Statistik zu ermöglichen. Dabei werden didaktische Szenarien, Vorteile, aber auch Grenzen der Nutzung von entsprechenden Apps aufgezeigt. Eine besondere Herausforderung stellen hierbei Visualisierungen komplexer statistischer Konzepte dar (Ellis & Merdian, 2015). Anhand eines kleinen randomisierten Experimentes, in dem Studierende mit einer App zu Konzepten aus der Inferenzstatistik arbeiten, wird aufgezeigt, wie durch eine entsprechende didaktische Einbettung Lernziele besser erreicht werden können. Im Rahmen dieses Beitrags soll auch der Austausch zwischen Lehrenden verschiedener Hochschulen angeregt werden.

Development and Evaluation of a Social Brokerage Index for Social Media – A Case Study of Twitter

Di 16:10
HS5

Martin Rehm*[1], Frank Cornelissen[2], Marijn ten Thij[3]

1: Universität Duisburg-Essen, Deutschland; 2: University of Amsterdam, Niederlande; 3: Vrije Universiteit Amsterdam, Niederlande

Social Network Analysis (SNA) has been acknowledged as a valuable tool to assess and describe (learning) networks in online social media (e.g. Rehm & Notten, 2016). However, a commonly used metric – brokerage – stems from a predominately offline environment, where individuals are in direct contact within a physical location. In an online realm, this distinctive feature is getting blurred and the explanatory power might be diminished (Daly, et al., 2013). Hence, while brokerage is still very relevant and useful to apply to SNA in online settings (e.g. De Nooy, et al., 2011), we believe that this metric can be modified to better represent the "new" circumstances. In order to account for these considerations, we propose a social brokerage index (SBI). In order to demonstrate our proposed metric, we employ a multi-method approach and use Twitter as an underlying network structure. We collected data from two hashtag conversations: #edtech and #edchat. Our metric is tested and contrasted with simulated data from scale-free networks (e.g. Barabási & Albert, 1999) and actor-based modeling (e.g. Maanen & Vecht, 2014). In the context of our actor-based modeling, we also add bilbiometric analyses, such as latent semantic analysis (Deer Wester et al. 1990) and topic modeling (e.g. Blei & Lafferty, 2009), to further inform our decision about which explanatory variables to include. This contribution will highlight our preliminary findings and stipulate how our approach and metric can help to better understand and anticipate people engagement in social media and possibly their attained brokerage positions.

IRT-basierter Item-Fit zur Analyse der Messinvarianz

Mi 8:30
HS2

Janine Buchholz*, Johannes Hartig
Deutsches Institut für Internationale Pädagogische Forschung (DIPF), Deutschland

Ein Ziel internationaler Large-Scale Assessments (LSA) besteht darin, die anhand von Leistungstests und Fragebögen gemessenen Konstrukte zwischen einer großen Anzahl an Ländern zu vergleichen. Solche Vergleiche sind jedoch nur dann zulässig, wenn in allen Ländern tatsächlich das gleiche Konstrukt gemessen wurde (Messinvarianz). Da zur Skalierung in LSAs häufig Modelle der Item-Response-Theorie (IRT) Anwendung finden, betrifft Messinvarianz die Gleichheit von Itemparametern des IRT-Modells. Zur Analyse der Messinvarianz wurde in PISA 2015 (OECD, 2016) ein innovativer Ansatz gewählt: Auf die gemeinsame Skalierung der Daten aus allen Ländern im Rahmen eines Generalized-Partial-Credit-Modells (Muraki, 1992) folgte die Berechnung der RMSD (Root Mean Square Deviance)-Item-Fit Statistik für jedes Item in jedem Land. Sie zeigt die Abweichung zwischen modellimplizierter und beobachteter Item Characteristic Curve und somit die Passung der internationalen Parameter auf die Daten eines einzelnen Landes an. Die Entscheidung über das Vorliegen von Invarianz hängt folglich am Cut-off-Wert auf dieser Statistik. In dieser Simulationsstudie wurden daher polytome Antwort-Daten mit unterschiedlichen Arten (bezogen auf Schwierigkeit bzw. Diskrimination) und Stärken von Non-Invarianz erzeugt und die empirischen Verteilungen der resultierenden RMSD-Statistik untersucht. Es zeigt sich, dass die Statistik stärker auf Zwischen-Gruppen-Unterschiede bezogen auf die Itemschwierigkeit reagiert, insgesamt aber sensitiv für beide Invarianz-Arten ist. Dies bedeutet jedoch auch, dass ein Anschlagen der Statistik keine Auskunft über die tatsächliche Ursache der Zwischen-Gruppen-Unterschiede anzeigen kann. Ähnlich zum Vorgehen in Mehrgruppen-Strukturgleichungsmod (z.B. Meredith, 1993) schlagen wir daher ein schrittweises Vorgehen vor, in dem die Itemparameter nacheinander freigesetzt werden, und präsentieren erste Befunde bezogen auf ein empirisches Datenbeispiel aus dem PISA-2015-Fragebogen.

Vom Testfairness-Paradox zu einem einheitlichen psychometrischen Fairness-Begriff

Mi 8:30
HS2

Safir Yousfi*

Bundesagentur für Arbeit, Deutschland

Der vorherrschende Ansatz zur Überprüfung von Testfairness in Bezug auf Personenmerkmale wie Geschlecht oder Alter beruht auf einem Vergleich der Regressionen des Kriteriums auf die Testscores. Wenn die Regresssion(sgerad)en in den betreffenden Gruppen identisch sind, liegt prädiktive Invarianz vor. In der psychometrischen Literatur wird dagegen Messinvarianz als zentrale Voraussetzung für die Testfairness betont. Nach den einschlägigen Teststandards zeichnet sich ein fairer Test dadurch aus, dass prädiktive Invarianz und Messvarianz gegeben sind. Die Arbeiten von Millsap zeigen jedoch den paradoxen Befund, dass prädiktive Invarianz und Messinvarianz sich in der Regel gegenseitig ausschließen. Während in der psychometrischen Literatur daher gefordert wird, dass man sich zwischen diesen beiden Fairnesskonzepten entscheidet, scheinen Testentwickler und Testanwender die Inkompatibilität von Messinvarianz und prädiktiver Invarianz einfach zu ignorieren. Daher soll hier ein psychometrisches Konzept der Testfairness vorgestellt werden, dass die Grundideen von Messinvarianz und prädiktiver Invarianz aufgreift und in einen einheitlichen formalen Begriff von Testfairness integriert und so das vermeintliche Testfairness-Paradox auflöst. Innerhalb dieses Ansatzes sind Messinvarianz und prädiktive Invarianz Spezialfälle psychometrischer Fairness. Selbst "weichere" Aspekte der Testfairness, die auch auf Werturteilen beruhen, lassen sich vielfach anhand eines solchen psychometrischen Fairnessbegriffs rekonstruieren. Damit zeichnet sich ein übergreifendes Konzept von Testfairness ab, das die Beziehungen zwischen Aspekten der Testfairness transparent macht, die vielfach getrennt diskutiert werden.

Metric Measurement Invariance of Latent Variables: Foundations, Testing, and Correct Interpretation

Mi 8:30
HS2

Stefan Klößner, Eric Klopp*
Universität des Saarlandes, Deutschland

In multi-group and longitudinal studies, it is important to test for metric measurement invariance. Recently, several authors have pointed out that currently used test procedures for measurement invariance (MI) do not fully test for MI and that additional assumptions about the invariance of the referent indicator are needed in order to conclude that actual data satisfy MI. Introducing the new concept of proportional factor loadings (PFL), we show that tests for MI actually only test for PFL, because PFL is empirically indistinguishable from metric MI. More precisely, if the loadings are only proportional over groups or time, the implied distribution of the observed variables is identical to one that results from invariant factor loadings. Thus, it is impossible to differentiate between MI and PFL based on empirical data only. Furthermore, PFL affects tests about the equality of latent variables' variances, leading to wrong conclusions when the data only satisfy PFL, but not MI. We also discuss how the empirical indistinguishability between PFL and MI affects partial MI. We find that it is typically impossible to differentiate invariant from non-invariant indicators. Empirically, one can only detect which indicators form subsets whose loadings are proportional. These findings explain why procedures for detecting invariant indicators perform poorly under certain conditions, a disturbing fact that several recent studies have found in Monte Carlo studies. Finally, we also discuss the referent indicator problem and how different scaling methods potentially affect the results of the above-mentioned tests.

A possible connection between Knowledge Space Theory and Item Response Theory using Information Theory

Mi 8:30
HS4

Stefano Noventa*, Jürgen Heller, Augustin Kelava
Universität Tübingen

Error parameters represent the noise inherent in any framework designed to assess educational and psychological performances and competences, like Item Response Theory (IRT), Knowledge Space Theory (KST) and Cognitive Diagnostic Modeling (CDM). Generally, they are modeled as conditional probabilities (e.g., guessing and slipping) under an assumption of conditional independence given the ability (IRT), the knowledge state (KST), or the set of skills (CDM) that characterize individuals. Information entropy is suggested as a natural framework to model error parameters, establishing a possible connection between these different theories. In particular, a two-steps model is hypothesized to separate guessing and slipping parameters into a first component (informed guess or mistake) due to the effects of individual ability on item mastering, and a second one (blind guess or careless error) due to the effects of pure chance on item solving. Based on requiring maximal mutual information between prior and actual knowledge, general families of logistic models are derived which account for the probabilities of informed guessing (mastering some item with only partial item-specific knowledge) and mistaking (not mastering some item in spite of item-specific knowledge). Results on the dichotomous case may be extended to arbitrary knowledge structures. Finally, a correspondence between generalized families of Rasch models and Logistic Knowledge Structures is obtained in the limit of minimum mutual information, when latent traits and item characteristics are introduced. Some implications on the assumptions of IRT and Rasch models are also drawn.

Vorhersage oder Produktion? – Ein Bayessches Strukturgleichungsmodell für Stochastic Frontier Analysen (B-SEM-SFA)

Mi 8:30
HS4

Rüdiger Mutz*
ETH Zürich, Schweiz

In der Sozialpsychologie oder in der Arbeits- und Organisationspsychologie wird häufig von Einheiten (z.B. Gruppe, Forschungsteam) ausgegangen, die aus einem Input (z.B. Wissen, Forschungsgelder, Personal) einen bestimmten Output (z.B. Artikel) erstellen. Damit wird ein empirischer Produktionszusammenhang zwischen Input und Output hergestellt, der bei einer Vielzahl von Einheiten statistisch über eine Regressionsanalyse, der "Stochastic Frontier Analysis" (SFA), modelliert werden kann. Statt jedoch aus dem Input den durchschnittlich zu erwarten Output vorherzusagen, wie es in der üblichen Regressionsanalyse der Fall ist, wird in der SFA der maximal aufgrund des Inputs zu erwartende Output als Grenze der Produktion oder Produktionsfunktion geschätzt. Die technische Effizienz einer Einheit bezeichnet dabei das Ausmass, in welchem eine Einheit den maximal zu erwartenden Output bei gegebenen Input erreicht. Die klassische SFA geht jedoch von einem messfehlerfreien Output bzw. Input aus. Für die Nutzung in den Sozialwissenschaften, insbesondere der Psychologie ist diese Annahme jedoch unrealistisch. Mittels Linearen Strukturgleichungsmodellen soll in diesem Beitrag ein SFA-Modell vorgestellt werden, das von einem Messmodell von Input und Output ausgeht und die SFA auf der latenten Ebene definiert. Die komplexe Residuenstruktur (technische Ineffizienz, Zufallsfehler, Messfehler) macht es erforderlich, das Modell in einem Bayesschen Rahmen zu definieren (SAS PROC MCMC). Klassische Softwareprogramme für Strukturgleichungsmodelle (z.B. MPLUS, LISREL, Lavaan) eignen sich dafür nicht. Neben der Definition von Produktion, der Ableitung des Modells und einer Simulationsstudie, die das Verhalten des Modells unter unterschiedlichen Bedingungen (Stichprobengrösse, Reliabilität, ...) untersucht, soll an einer Effizienzanalyse von Forschungsprojekten des FWF (Wissenschaftsfonds Österreichs), das Modell beispielhaft dargestellt werden.

On the relationship between maximum entropy and maximum likelihood methods with applications in psychometrics, categorical data analysis, and network analysis.

Mi 8:30
HS4

Georg Hosoya*
Freie Universität Berlin, Deutschland

This theory oriented talk highlights the formal intersection between maximum entropy and maximum likelihood methods. The maximum entropy method provides the researcher with a solution to the primal optimization problem of finding the least informative, theoretical probability distribution given a set of testable information in data. The solution to the dual problem of finding the parameters of that distribution is optimizing the likelihood given the Lagrange multipliers that are introduced in the solution to the primal problem. Although these results seem to be well-known, particularly in the domain of machine learning, their practical implications in psychology are new. To highlight the practical applicability of maximum entropy methods to (psychometric) model formulation, the derivation of a psychometric measurement model from first maximum entropy principles is shown in detail, including the gradient and information functions. In addition, the formal relationships to various other well-known models of psychometrics, categorical data analysis (classification), as well as network analysis, are discussed.

Model implied instrumental variables (MIIVs): A new orientation to structural equation models (SEMs)

Mi 11:20
Audimax

Kenneth A. Bollen*
University of North Carolina at Chapel Hill

It is hardly controversial to say that our models are approximations to reality. Yet when it comes to estimating structural equation models (SEMs), we use estimators that assume true models (e.g., ML) and that readily spread bias through estimated parameters when the model is approximate. This talk presents the Model Implied Instrumental Variable (MIIV) approach to SEMs initially proposed in Bollen (1996). It has greater robustness to structural misspecifications and the conditions for robustness are well understood. In addition, the MIIV-2SLS estimator is asymptotically distribution free. Furthermore, MIIV-2SLS has equation based overidentification tests that can help pinpoint errors in specification. Beyond these features, the MIIV approach has other desirable qualities. It permits new tests of dimensionality and tests of causal vs. reflective indicators. MIIV methods apply to higher order factor analyses, categorical measures, growth curve models, and nonlinear latent variables. Finally, it permits researchers to estimate and test only the latent variable model or any other subset of equations. This presentation will provide an overview of this new orientation to SEMs and illustrate MIIVsem, an R package that implements it.

Schätzung der prognostischen Validität von Auswahlverfahren mittels multipler Imputation – Illustration des Vorgehens am Beispiel realer Datensätze

Mi 12:40
HS2

Andreas Pfaffel*, Christiane Spiel
Universität Wien, Österreich

Die statistische Schätzung der prognostischen Validität von Auswahlverfahren (z.B. im Hochschulbereich) beruht typischerweise auf systematisch zensierten Daten, weil nur von den ausgewählten Personen Werte des Erfolgskriteriums (z.B. Noten, ECTS-Punkte, Studienabschluss) vorliegen. Die Nicht-Beachtung der fehlenden Kriteriumswerte führt zu verzerrten Schätzern der prognostischen Validität der Auswahlverfahren ("range restriction"-Problem). Anhand von Simulationsstudien konnte gezeigt werden, dass eine Schätzung mittels multipler Imputation mit Markov-Ketten auch bei einem sehr hohem Anteil an fehlenden Werten zu weniger verzerrten Schätzern führt und bisherigen Korrekturmethoden überlegen ist (Pfaffel, Kollmayer, Schober & Spiel, 2016; Pfaffel, Schober & Spiel, 2016). Bisher wurde dieses Vorgehen jedoch noch nicht an realen Datensätzen angewandt. Im Vortrag wollen wir deshalb die Anwendung der multiplen Imputation zur Schätzung der prognostischen Validität anhand realer Datensätze von Auswahlverfahren für den Hochschulzugang zeigen. Dabei ergaben sich methodische Probleme, die in den Simulationsstudien nicht berücksichtigt wurden (z.B. durch ungleiche Standardfehler der Messwerte der Erfolgskriterien, Multikollinearität der Prädiktoren, hohe Korrelation oder stark asymmetrische Verteilung der Kriteriumsvariablen). Die Bewältigung dieser Probleme erforderte Annahmen über die Verteilung der fehlenden Werte sowie zusätzliche Spezifikationen des Imputationsmodells. Aktuell fehlen in der Forschung empirische Studien, die die Verteilungsannahmen und die inkrementelle Effektivität komplexerer Modelle belegen. Mit den TeilnehmerInnen wollen wir deshalb einerseits das methodische Vorgehen zur Lösung dieser Probleme diskutieren, sowie andererseits weitere typische Probleme in diesem Bereich für künftige Forschungsvorhaben sammeln.

Imputation fehlender Werte auf Ebene 2: Ein Vergleich verschiedener Verfahren und Implementation mithilfe von Plausible Values

Mi 12:40
HS2

Simon Grund*[1,2], Oliver Lüdtke[1,2], Alexander Robitzsch[1,2]

1: Leibniz-Institut für die Pädagogik der Naturwissenschaften und Mathematik; 2: Zentrum für internationale Bildungsvergleichsstudien

In der empirischen Forschung weisen Daten häufig eine Mehrebenenstruktur auf, in der fehlende Werte sowohl auf Ebene 1 (z.B. Schülerinnen und Schüler) als auch auf Ebene 2 (z.B. Schulen) auftreten können. Die multiple Imputation (MI) kann verwendet werden, fehlende Werte auf Ebene 2 zu behandeln und dabei Informationen von Variablen auf Ebene 1 zu berücksichtigen. Obwohl sich zahlreiche Arbeiten mit fehlenden Werten auf Ebene 1 befasst haben (z.B. Enders, Mistler & Keller, 2016), wirft die Behandlung fehlender Werte auf Ebene 2 weiterhin Fragen auf, z.B. wie Informationen von Variablen auf Ebene 1 berücksichtigt werden können (Resche-Rigon & White, in Druck). Anhand theoretischer Argumente und einer Computersimulation vergleichen wir Ansätze des "joint modeling" (JM) und der "fully conditional specification" (FCS) für fehlende Werte auf Ebene 2 sowie verschiedene Ansätze, Variablen auf Ebene 1 anhand manifester und latenter Gruppenmittelwerte zu berücksichtigen. Die Technik der Plausible Values (Mislevy, 1991) wird genutzt, um die Simulation latenter Gruppenmittelwerte im FCS-Ansatz zu ermöglichen. Wir zeigen, dass (a) ein FCS-Ansatz mit latenten Gruppenmittelwerten vergleichbar ist mit JM und dass (b) die Verwendung manifester Gruppenmittelwerte, obwohl nicht völlig äquivalent mit JM, zu ähnlichen Ergebnisse führt, so lange die Stichproben nicht in extremer Weise unbalanciert sind. Wir skizzieren das computationale Vorgehen zur Simulation von Plausible Values im FCS-Ansatz und illustrieren die verschiedenen Verfahren an einem Beispiel aus PISA 2012.

Modellerierung fehlender Werte unter Berücksichtigung entscheidungstheorethischer Erkenntnisse

Mi 12:40
HS2

Katja Buntins*
Universität Duisburg-Essem, Deutschland

Obwohl das Auslassen von Items in Aufgaben mit Multiple Choice Antwortformat im entscheidungstheorethischen Sinne nicht rational ist (Lord, 1974), kommt das Auslassen von Items in Kompetenztests häufig vor. Daher ist in der Skalierung von Kompetenzwerten zu klären, wie mit fehlenden Werten umgegangen werden soll. In der Forschungspraxis gibt es verschiedene Vorschläge, Kompetenzwerte am besten, d.h. möglichst unverzerrt, zu schätzen (vgl. Robitzsch, 2016; Pohl, Gräfe, Rose, & 2014; Pohl, & Carstensen, 2013). Hierbei wird jeweils eine Implikation gemacht, warum das Item nicht beantwortet wurde. Einige Studien zeigen durch experimentelle Manipulation, dass ein gut funktionierende Möglichkeit zur Vorhersage der Antwortwahrscheinlichkeit auf einem Item, die Annahmen der Prospect Theory sind (vergl. Budescu & BarHillel, 1993; Bar-Hillel, Budescu & Attali, 2005; Espinosa; & Gardeazabal, 2013, Budesco, & Bo, 2015). In Vorstudien konnten wir Evidenz dafür finden, dass sich die Annahmen der Prospect Theory auch im Bereich der Large Scale Assessments anwenden lässt. Hierbei stellt sich allerdings die Frage inwieweit unterschiedliche Wertefunktionen in die Schätzung der Kompetenz mit eingehen können. Mittles neuere Ansätze der Modellierung von fehlenden Werten können auch Drittvariablen in den Modellierungsprozess mit einbezogen werden (Köhler, Pohl, & Carstensen, 2015; Hafez, Moustaki & Kuha, 2015). Eine dieser Drittvariablen könnte die persönliche Risikoaversivität sein. Mögliche Operationalsierungen von Risikoaversivität in Large Scale Assessment werden modelliert. Was die Ergebnisse für die Diskussion über fehlende Werte in Kompetzensklaierungen bedeuten und welchen Mehrwert entscheidungstheorethische Modelle bringen werden im Rahmen des Vortrags kritisch reflektiert.

Kriteriumsorientierte adaptive Hochschulklausuren: Auswirkung verschiedener Kalibrierungsdesigns auf die Personenparameterschätzung

Mi 12:40
HS4

Sebastian Born*[1], Andreas Frey[1,2], Christian Spoden[1], Aron Fink[1]

1: Friedrich-Schiller-Universität Jena, Deutschland; 2: Centre for Educational Measurement (CEMO) at the University of Oslo, Norway

Die Kalibrierung von Aufgabenpools für kriteriumsorientierte adaptive Klausuren stellt aufgrund der üblicherweise an Hochschule vorzufindenden Rahmenbedingungen (z. B. vergleichsweise kleine Kalibrierungsstichproben, begrenzte Ressourcen für die Aufgabenentwicklung) eine besondere Herausforderung dar. Die Methode der Online-Kalibrierung (z. B. Makransky & Glas, 2010) liefert einen Baustein für eine praxistaugliche Lösung. Dabei erfolgt nach einer initialen Kalibrierungsphase die Kalibrierung neuer Items in der operationalen Phase der adaptiven Klausur. Basierend auf dem Beitrag von Frey, Born, Spoden & Fink wird hier nun der Forschungsfrage nachgegangen, welche Auswirkung das verwendete Kalibrierungsdesign auf die Qualität der Personenparameterschätzung hat. Zur Beantwortung der Forschungsfrage wird in einer Simulationsstudie die Online-Kalibrierung des Aufgabenpools auf Basis der Item-Response-Theory über mehrere Testzeitpunkte simuliert. Der Studie liegt ein zweifaktorielles Design mit den Faktoren Länge der initialen Kalibrierungsphase und Größe der Kalibrierungsstichprobe (50, 100, 300) zu Grunde. Der Faktor Länge der initialen Kalibrierungsphase beschreibt dabei die Anzahl der Testzeitpunkte, die für eine initiale Kalibrierung aller Aufgaben notwendig ist. Evaluationskriterien für die Qualität der Personenparameterschätzung sind der Bias, der absoluten Bias und die Messpräzision (MSE). Die Studie ist noch nicht abgeschlossen, wird bis zur Tagung aber vollständig fertig sein. Bezüglich der Ergebnisse wird erwartet, dass die Länge der initialen Kalibrierungsphase bei großen Stichproben nur geringen Einfluss auf die Qualität der Personenparameterschätzung hat. Bei kleinen Stichproben hingegen wird erwartet, dass eine längere initiale Kalibrierungsphase von Vorteil ist.

Kriteriumsorientierte adaptive Hochschulklausuren: Methodische Probleme typischer Klausuren und Möglichkeiten zur Verbesserung

Mi 12:40
HS4

Andreas Frey*[1,2], Sebastian Born[1], Christian Spoden[1], Aron Fink[1]

1: Friedrich-Schiller-Universität Jena, Deutschland; 2: Centre for Educational Measurement (CEMO) at the University of Oslo, Norway

Das Ausmaß mit dem Studierende die Ziele ihres Hochschulstudiums erreichen wird formal am Abschneiden bei Prüfungen festgemacht. Eine zentrale Form solcher Prüfungen sind Klausuren. Wie andere Prüfungen auch, haben Klausuren für die Studierenden eine hohe persönliche Relevanz, da ein erfolgreicher Studienabschluss vom Abschneiden bei Klausuren abhängt. Dieser hohen Bedeutung werden typische Hochschulklausuren aus methodischer Sicht indes nicht gerecht. Vielmehr klafft eine deutliche Lücke zwischen psychometrischem Kenntnisstand und Prüfungspraxis. Problematisch an derzeitigen Hochschulklausuren sind im Wesentlichen vier Aspekte: 1. Lernziele werden nicht angemessen durch die genutzten Klausuraufgaben operationalisiert. 2. Klausuren sind nicht als kriteriumsorientierte Verfahren konzipiert, so dass Klausurergebnisse nicht als Ausmaß des Erreichens von Lernzielen interpretiert werden können. 3. Testzeitpunkte werden nicht statistisch verlinkt, so dass Unabhängigkeit der vergebenen Noten von Kohortenleistungsfähigkeit und Klausurschwierigkeit nicht gewährleistet sind. 4. Die Messpräzision schwankt über den Merkmalsbereich mit typischerweise deutlich niedrigerer Messpräzision in den Randbereichen der Merkmalsverteilung. Im Vortrag wird erörtert wie diese Probleme durch eine Kombination aktueller Methoden und Ansätze aus den Bereichen Psychometrie und Educational Measurement gelöst werden können. Der Fokus liegt dabei auf der psychometrisch anspruchsvollen Frage, wie die Kalibrierung des Aufgabenpools für kriteriumsorientierte, adaptive Hochschulklausuren auf Basis der Item-Response-Theory im laufenden Lehrbetrieb erfolgen kann. Dabei werden die Ziele (a) kontinuierliche Vergrößerung des Aufgabenpools, (b) Optimierung der Itemparameterschätzung, (c) Konstanthaltung der Metrik über Messzeitpunkte, (d) Identifikation defizitärer Items und (e) Berücksichtigung von Itempositionseffekten adressiert. Bei dem Vortrag handelt es sich um den ersten Teil. Im zweiten Teil von Born, Frey, Spoden & Fink wird das Kalibrierungskonzept anhand einer Simulationsstudie illustriert.

Die Rolle von Itemkovarianzstrukturen auf Klassenebene in der Messung von Instruktionssensitivität

Mi 12:40
HS4

Alexander Naumann*[1], Johannes Hartig[1], Jan Hochweber[2]

1: Deutsches Institut für Internationale Pädagogische Forschung (DIPF), Deutschland; 2: Pädagogische Hochschule St. Gallen (PHSG)

Testergebnisse von Schülerinnen und Schülern werden regelmäßig auf einen mehr oder weniger guten Unterricht zurückgeführt (Creemers & Kyriakides, 2008). Gültige Rückschlüsse über Unterricht setzen jedoch voraus, dass die eingesetzten Instrumente potentiell dazu in der Lage sind, Effekte des Unterrichts aufzufangen. Polikoff (2010) bezeichnet diese psychometrische Eigenschaft eines Test oder einzelnen Items als Instruktionssensitivität. Während diverse Verfahren zur Messung der Instruktionssensitivität einzelner Items verfügbar sind, wird die Varianz von Testscores zwischen Lerngruppen innerhalb einer Stichprobe (z.B. Klassen) als ausschlaggebend für die Messung der Instruktionssensitivität eines Tests angesehen (Naumann, Hartig, & Hochweber, 2017). Allerdings ist die Beziehung von Indikatoren der Itemebene zur Testebene nach wie vor wenig untersucht. Unsere Studie zielt darauf ab, diese Beziehung für einen kürzlich vorgeschlagenen Ansatz zur Messung der Itemsensitivität von Naumann und Kollegen (2017) zu prüfen. In deren längsschnittlichen Mehrebenen-IRT (LMLIRT) Modell dienen klassenspezifische Veränderungswerte für Itemschwierigkeiten über Messzeitpunkte als Indikatoren für Instruktionssensitivität. Wir vermuten, dass die Instruktionssensitivität eines Tests ansteigt, je stärker diese klassenspezifischen Veränderungswerte über die einzelnen Items innerhalb eines Tests hinweg kovariieren. Zur Überprüfung unserer Hypothese simulieren wir auf Basis des LMLIRT-Modells Itemantworten und varrieren dabei systematisch die Kovarianz der klassenspezifischen Veränderungswerte über Items hinweg. Die Ergebnisse unterstützen unsere Hypothese. Ein Test scheint umso instruktionssensitiver, je stärker die Instruktionssensitivität der einzelnen Items kovariiert. Einerseits deutet dieser Befund darauf hin, das der LMLIRT-Ansatz konsistent zur Testebene ist. Andererseits ergeben sich daraus Fragen zur Dimensionalität von Veränderungswerten in Längsschnittsmessungen, wenn ein Test unterschiedlich instruktionssensitive Items umfasst, also Lernzuwächse nicht homogen über Items hinweg erscheinen.

Autorenverzeichnis

Impressum

Postanschrift

Eberhard Karls Unversität Tübingen
Prof. Dr. Augustin Kelava
Europastr. 6
72072 Tübingen

Organisationsteam

Prof. Dr. Augustin Kelava
Prof. Dr. Holger Brandt
Prof. Dr. Jürgen Heller
M.Sc. Tim Schaffland

Gestaltung des Tagungsbandes

Prof. Dr. Holger Brandt
Prof. Dr. Augustin Kelava
M.Sc. Tim Schaffland

Dank

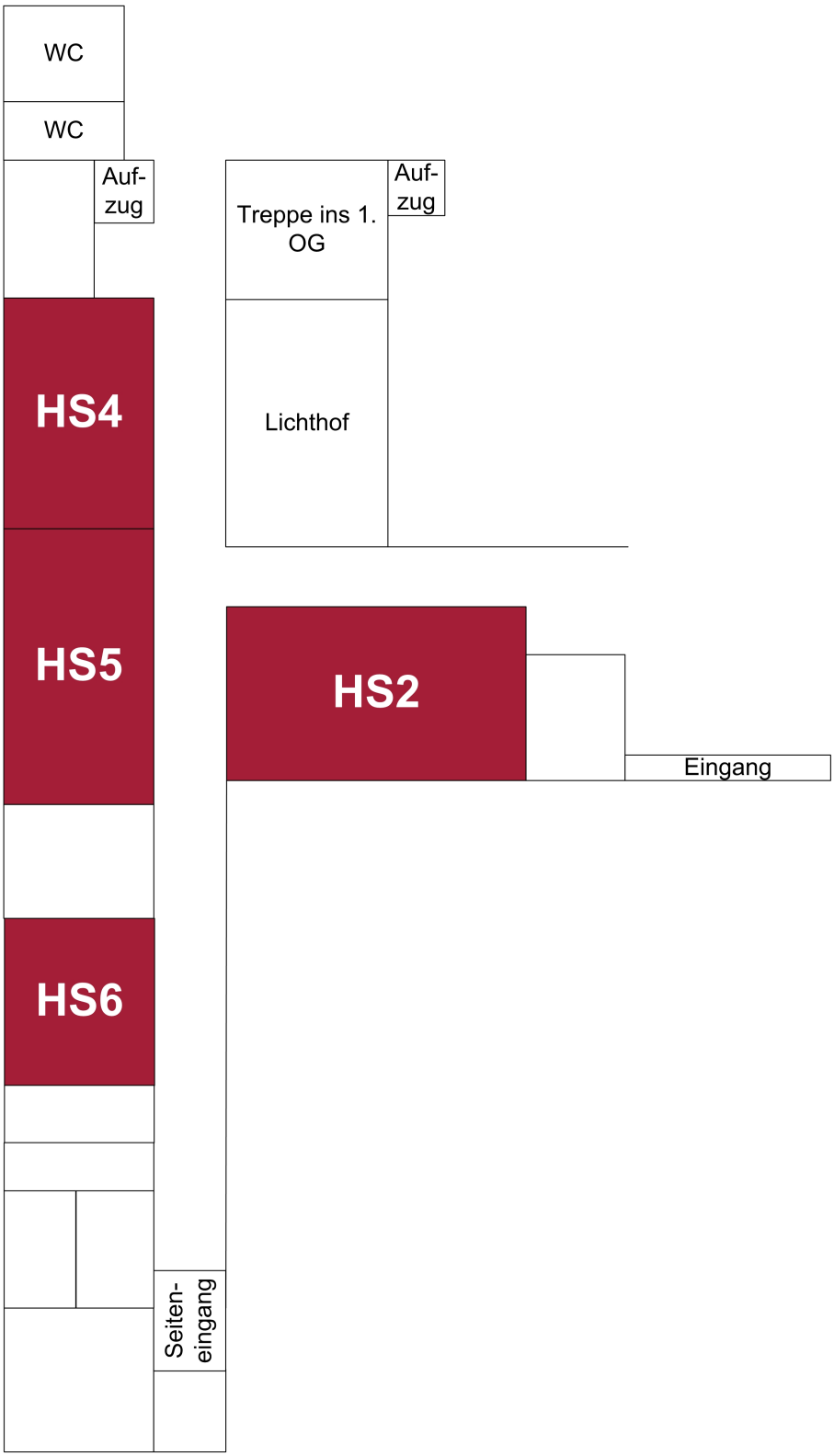
Wir bedanken uns herzlich für die finanzielle Unterstützung durch die Fachgruppe Methoden und Evaluation der DGPs und der Deutschen Forschungsgemeinschaft (DFG) für die Ausrichtung der Tagung.

Dem Springer Verlag, dem Verein zur Förderung qualitativer Forschung – Association for Supporting Qualitative Research ASQ sowie der Fachgruppe Methoden und Evaluation der DGPs danken wir für die Finanzierung der vergebenen wissenschaftlichen Preise. Ebenso gilt unser Dank den Mitgliedern der Kommissionen zur Vergabe der Preise sowie allen Gutachterinnen und Gutachtern der Tagungsbeiträge (Einzelvorträge, Symposien etc.)!

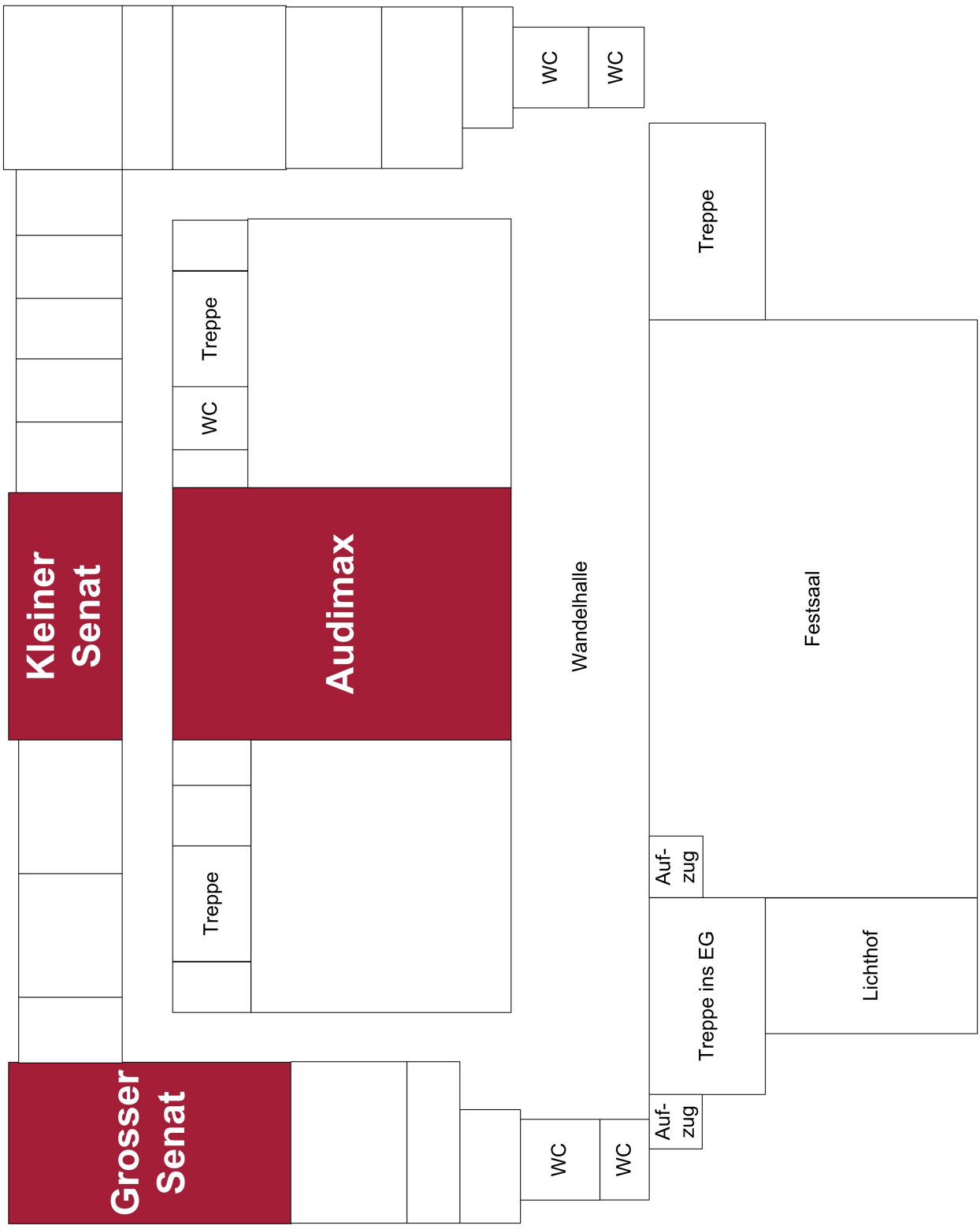
Besonderer Dank gilt den Hilfskräften und Mitarbeitern, die bei der Organisation und Webseitengestaltung unverzichtbare Hilfe geleistet haben.

Raumpläne

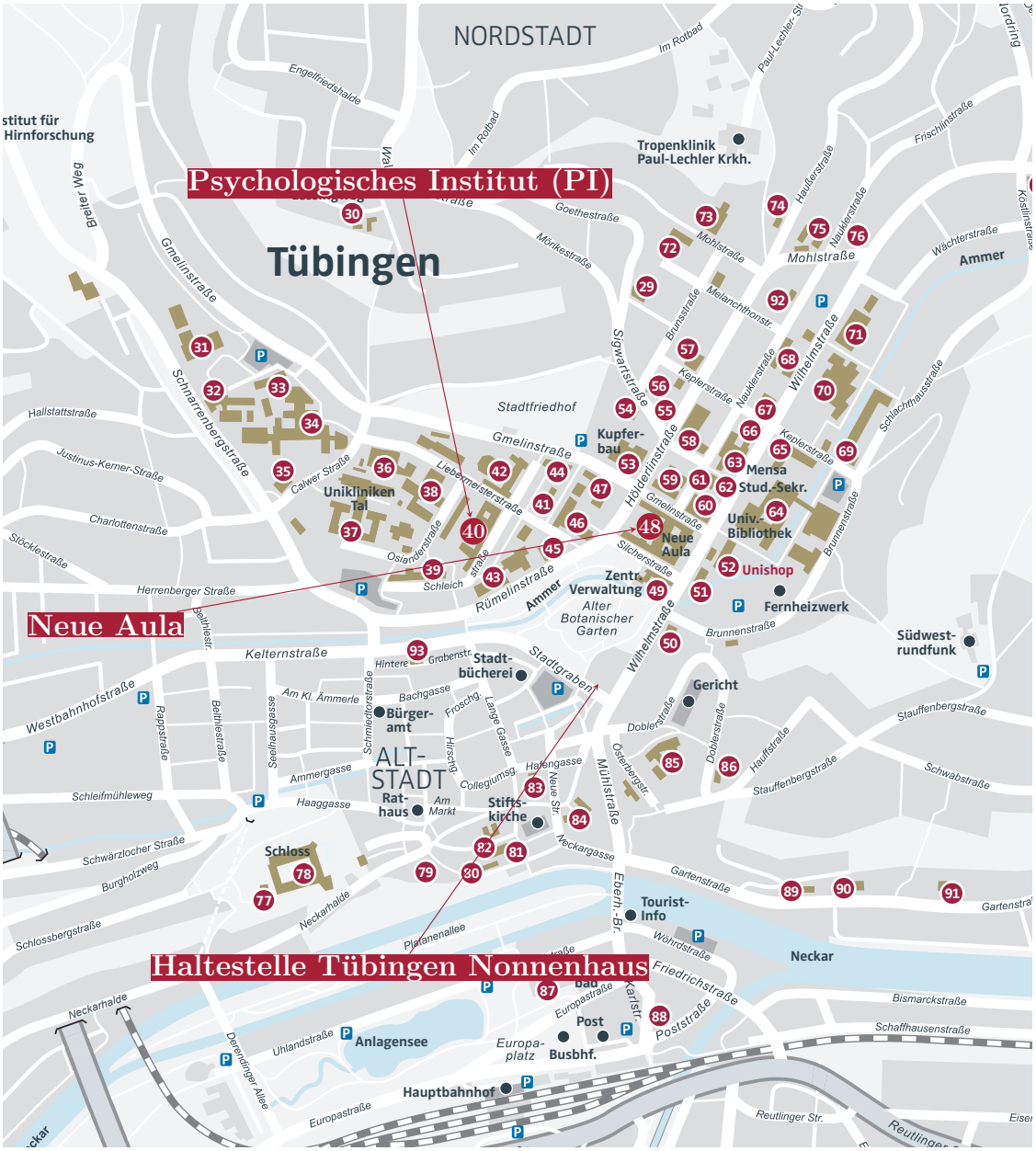
Räume für die Workshops im Psychologischen Institut (PI)



Räume im EG der Neuen Aula



Räume im OG der Neuen Aula



Lageplan der Universität