Ivan Jacob Agaloos Pesigan

February 10, 2024

References

M. W.-L. Cheung: Synthesizing indirect effects in mediation models with meta-analytic methods   **Cheung-2021**

Mike W.-L. Cheung. "Synthesizing indirect effects in mediation models with meta-analytic methods". In: *Alcohol and Alcoholism* 57.1 (June 2021), pp. 5–15. DOI: 10.1093/alcalc/agab044.

Abstract: Aims A mediator is a variable that explains the underlying mechanism between an independent variable and a dependent variable. The indirect effect indicates the effect from the predictor to the outcome variable via the mediator. In contrast, the direct effect represents the predictor's effort on the outcome variable after controlling for the mediator. Methods A single study rarely provides enough evidence to answer research questions in a particular domain. Replications are generally recommended as the gold standard to conduct scientific research. When a sufficient number of studies have been conducted addressing similar research questions, a meta-analysis can be used to synthesize those studies' findings. Results The main objective of this paper is to introduce two frameworks to integrating studies using mediation analysis. The first framework involves calculating standardized indirect effects and direct effects and conducting a multivariate meta-analysis on those effect sizes. The second one uses meta-analytic structural equation modeling to synthesize correlation matrices and fit mediation models on the average correlation matrix. We illustrate these procedures on a real dataset using the R statistical platform. Conclusion This paper closes with some further directions for future studies.

S. F. Cheung et al.: FINDOUT: Using either SPSS commands or graphical user interface to identify influential cases in structural equation modeling in AMOS

Shu Fai Cheung and Ivan Jacob Agaloos Pesigan. "FINDOUT: Using either SPSS commands or graphical user interface to identify influential cases in structural equation modeling in AMOS". In: *Multivariate Behavioral Research* (Jan. 2023), pp. 1–5. DOI: .

Abstract: The results in a structural equation modeling (SEM) analysis can be influenced by just a few observations, called influential cases. Tools have been developed for users of R to identify them. However, similar tools are not available for AMOS, which is also a popular SEM software package. We introduce the FINDOUT toolset, a group of SPSS extension commands, and an AMOS plugin, to identify influential cases and examine how these cases influence the results. The SPSS commands can be used either as syntax commands or as custom dialogs from pull-down menus, and the AMOS plugin can be run from AMOS pull-down menu. We believe these tools can help researchers to examine the robustness of their findings to influential cases.

S. F. Cheung et al.: semlbci: An R package for forming likelihood-based confidence intervals for parameter estimates, correlations, indirect effects, and other derived parameters

Shu Fai Cheung and Ivan Jacob Agaloos Pesigan. "semlbci: An R package for forming likelihood-based confidence intervals for parameter estimates, correlations, indirect effects, and other derived parameters". In: *Structural Equation Modeling: A Multidisciplinary Journal* (May 2023), pp. 1–15. DOI: .

Abstract: There are three common types of confidence interval (CI) in structural equation modeling (SEM): Wald-type CI, bootstrapping CI, and likelihood-based CI (LBCI). LBCI has the following advantages: (1) it has better coverage probabilities and Type I error rate compared to Wald-type CI when the sample size is finite; (2) it correctly tests the null hypothesis of a parameter based on

likelihood ratio chi-square difference test; (3) it is less computationally intensive than bootstrapping CI; and (4) it is invariant to transformations. However, LBCI is not available in many popular SEM software packages. We developed an R package, semlbci, for forming LBCI for parameters in models fitted by lavaan, a popular open-source SEM package, such that researchers have more options in forming CIs for parameters in SEM. The package supports both unstandardized and standardized estimates, derived parameters such as indirect effect, multisample models, and the robust LBCI proposed by Falk.

S. F. Cheung et al.: DIY bootstrapping: Getting the nonparametric bootstrap confidence interval in SPSS for any statistics or function of statistics (when this bootstrapping is appropriate) Cheung-Pesigan-Vong-2022

Abstract: Researchers can generate bootstrap confidence intervals for some statistics in SPSS using the BOOTSTRAP command. However, this command can only be applied to selected procedures, and only to selected statistics in these procedures. We developed an extension command and prepared some sample syntax files based on existing approaches from the Internet to illustrate how researchers can (a) generate a large number of nonparametric bootstrap samples, (b) do desired analysis on all these samples, and (c) form the bootstrap confidence intervals for selected statistics using the OMS commands. We developed these tools to help researchers apply nonparametric bootstrapping to any statistics for which this method is appropriate, including statistics derived from other statistics, such as standardized effect size measures computed from the t test results. We also discussed how researchers can extend the tools for other statistics and scenarios they encounter.

Abstract: Objective: Affect regulation models of drinking state that affect motivates and reinforces drinking. Few studies have been able to elucidate the timing of these associations in natural settings. We tested positive affect (PA) and negative affect (NA) as predictors of drinking behavior, both prior to and during drinking episodes, and whether drinking predicted changes in affect during episodes. Method: Two hundred twentytwo regularly drinking young adults (21–29 years, 84% undergraduates), completed an ecological momentary assessment (EMA) protocol for five consecutive 24-hr periods stretching across 6 days (Wednesday–Monday). Participants provided PA and NA reports three times daily and every half hour during drinking episodes. Alcohol consumption reports were provided each morning and every half hour during drinking episodes. Results: Multi-level models showed that greater pre-drinking PA predicted higher odds of drinking and greater number of drinks consumed. Pre-drinking NA did not predict same day odds of drinking or drinks consumed. Episode-level results revealed different associations for PA and NA with drinking. Current PA did not predict drinks consumed over the next half hour; however, increased drinking was associated with greater increases in PA over the next half hour. Higher NA predicted fewer drinks consumed in the next half hour and higher odds of the end of a drinking episode; however, increased drinking was not associated with changes in NA. Conclusions: PA increased following drinking during episodes. Our results suggest that a focus on PA prior to episodes and a focus on NA during episodes may interrupt processes leading to heavy drinking, and may therefore aid prevention efforts.

Didier et al.: Signal processing and machine learning with transdermal alcohol concentration to predict natural environment alcohol consumption.

Abstract: Wrist-worn alcohol biosensors continuously and discreetly record transdermal alcohol concentration (TAC) and may allow alcohol researchers to monitor alcohol consumption in participants' natural environments. However, the field lacks established methods for signal processing and detecting alcohol events using these devices. We developed software that streamlines analysis of raw data (TAC, temperature, and motion) from a wrist-worn alcohol biosensor (BACtrack Skyn) through a signal processing and machine learning pipeline: biologically implausible skin surface temperature readings (¡ 28C) were screened for potential device removal and TAC artifacts were corrected, features that describe TAC (e.g., rise duration) were calculated and used to train models (random forest and logistic regression) that predict self-reported alcohol consumption, and model performances were measured and summarized in autogenerated reports. The software was tested using 60 Skyn data sets recorded during 30 alcohol drinking episodes and 30 nonalcohol drinking episodes. Participants (N = 36; 13 with alcohol use disorder) wore the Skyn during one alcohol drinking episode and one nonalcohol drinking episode in their natural environment. In terms of distinguishing alcohol from nonalcohol drinking, correcting artifacts in the data resulted in 10% improvement in model accuracy relative to using raw data. Random forest and logistic regression models were both accurate, correctly predicting 97% (58/60; AUC-ROCs = 0.98, 0.96) of episodes. Area under TAC curve, rise duration of TAC curve, and peak TAC were the most important features for predictive accuracy. With promising model performance, this protocol will enhance the efficiency and reliability of TAC sensors for future alcohol monitoring research.

Abstract: Influential psychological theories hypothesize that people consume alcohol in response to the experience of both negative and positive emotions. Despite two decades of daily diary and ecological momentary assessment research, it remains unclear whether people consume more alcohol on days they experience higher negative and positive affects in everyday life. In this preregistered meta-analysis, we synthesized the evidence for these daily associations between affect and alcohol use. We included individual participant data from 69 studies (N = 12,394), which used daily and momentary surveys to assess the affect and the number of alcoholic drinks consumed. Results indicate that people are not more likely to drink on days they experience high negative affect but are more likely to drink and drink heavily on days high in positive affect. People self-reporting a motivational tendency to drink-to-cope and drink-to-enhance consumed more alcohol but not on days they experienced higher negative and positive affects. Results were robust across different operationalizations of affect, study designs, study populations, and individual characteristics. These findings challenge the long-held belief that people drink more alcohol following increase in negative affect. Integrating these findings under different theoretical models and limitations of this field of research, we collectively propose an agenda for future research to explore open questions surrounding affect and alcohol use.

Fridberg et al.: Examining features of transdermal alcohol biosensor readings: A promising approach with implications for research and intervention

Daniel J. Fridberg, Yan Wang, and Eric Porges. "Examining features of transdermal alcohol biosensor readings: A promising approach with implications for research and intervention". In: *Alcoholism: Clinical and Experimental Research* 46.4 (Feb. 2022), pp. 514–516. ISSN: 1530-0277. DOI: 10.1111/acer.14794.

Georgeson et al.: A sensitivity analysis for temporal bias in cross-sectional mediation

A. R. Georgeson, Diana Alvarez-Bartolo, and David P. MacKinnon. "A sensitivity analysis for temporal bias in cross-sectional mediation". In: *Psychological Methods* (Dec. 2023). DOI: 10.1037/met0000628.

Abstract: For over three decades, methodologists have cautioned against the use of cross-sectional mediation analyses because they yield biased parameter estimates. Yet, cross-sectional mediation models persist in practice and sometimes represent the only analytic option. We propose a sensitivity analysis procedure to encourage a more principled use of cross-sectional mediation analysis, drawing inspiration from Gollob and Reichardt (1987, 1991). The procedure is based on the two-wave longitudinal mediation model and uses phantom variables for the baseline data. After a researcher provides ranges of possible values for cross-lagged, autoregressive, and baseline Y and M correlations among the phantom and observed variables, they can use the sensitivity analysis to identify longitudinal conditions in which conclusions from a cross-sectional model would differ most from a longitudinal model. To support the procedure, we first show that differences in sign and effect size of the b-path occur most often when the cross-sectional effect size of the b-path is small and the cross-lagged and the autoregressive correlations are equal or similar in magnitude. We then apply the procedure to cross-sectional analyses from real studies and compare the sensitivity analysis results

to actual results from a longitudinal mediation analysis. While no statistical procedure can replace longitudinal data, these examples demonstrate that the sensitivity analysis can recover the effect that was actually observed in the longitudinal data if provided with the correct input information. Implications of the routine application of sensitivity analysis to temporal bias are discussed. R code for the procedure is provided in the online supplementary materials.

Gunn et al.: Characterising patterns of alcohol use among heavy drinkers: A cluster analysis utilising alcohol biosensor data   Gunn-Steingrimsson-Merrill-etal-2021

Abstract: Introduction: Previous research has predominately relied on person-level or single characteristics of drinking episodes to characterise patterns of drinking that may confer risk. This research often relies on self-report measures. Advancements in wearable alcohol biosensors provide a multifaceted objective measure of drinking. The current study aimed to characterise drinking episodes using data derived from a wearable alcohol biosensor. Methods: Participants ($n = 45$) were adult heavy drinkers who wore the Secure Continuous Remote Alcohol Monitoring (SCRAM) bracelet and reported on their drinking behaviours. Cluster analysis was used to evaluate unique combinations of alcohol episode characteristics. Associations between clusters and self-reported person and event-level factors were also examined in univariable and multivariable models. Results: Results suggested three unique clusters: Cluster 1 (most common, slowest rate of rise to and decline from peak), Cluster 2 (highest peak transdermal alcohol concentration and area under the curve) and Cluster 3 (fastest rate of decline from peak). Univariable analyses distinguished Cluster 1 as having fewer self-reported drinks and fewer episodes that occurred on weekends relative to Cluster 2. The effect for number of drinks remained in multivariable analyses. Discussion and Conclusions: This is the first study to characterise drinking patterns at the event-level using objective data. Results suggest that it is possible to distinguish drinking episodes based on several characteristics

derived from wearable alcohol biosensors. This examination lays the groundwork for future studies to characterise patterns of drinking and their association with consequences of drinking behaviour.

Li et al.: Bayesian forecasting with a regime-switching zero-inflated multilevel poisson regression model: An application to adolescent alcohol use with spatial covariates

Li-Oravecz-Zhou-etal-2022

Abstract: In this paper, we present and evaluate a novel Bayesian regime-switching zero-inflated multilevel Poisson (RS-ZIMLP) regression model for forecasting alcohol use dynamics. The model partitions individuals' data into two phases, known as regimes, with: (1) a zero-inflation regime that is used to accommodate high instances of zeros (non-drinking) and (2) a multilevel Poisson regression regime in which variations in individuals' log-transformed average rates of alcohol use are captured by means of an autoregressive process with exogenous predictors and a person-specific intercept. The times at which individuals are in each regime are unknown, but may be estimated from the data. We assume that the regime indicator follows a first-order Markov process as related to exogenous predictors of interest. The forecast performance of the proposed model was evaluated using a Monte Carlo simulation study and further demonstrated using substance use and spatial covariate data from the Colorado Online Twin Study (CoTwins). Results showed that the proposed model yielded better forecast performance compared to a baseline model which predicted all cases as non-drinking and a reduced ZIMLP model without the RS structure, as indicated by higher AUC (the area under the receiver operating characteristic (ROC) curve) scores, and lower mean absolute errors (MAEs) and root-mean-square errors (RMSEs). The improvements in forecast performance were even more pronounced when we limited the comparisons to participants who showed at least one instance of transition to drinking.

Abstract: The influx of intensive longitudinal data creates a pressing need for complex modeling tools that help enrich our understanding of how individuals change over time. Multilevel vector autoregressive (mlVAR) models allow for simultaneous evaluations of reciprocal linkages between dynamic processes and individual differences, and have gained increased recognition in recent years. High-dimensional and other complex variations of mlVAR models, though often computationally intractable in the frequentist framework, can be readily handled using Markov chain Monte Carlo techniques in a Bayesian framework. However, researchers in social science fields may be unfamiliar with ways to capitalize on recent developments in Bayesian software programs. In this paper, we provide step-by-step illustrations and comparisons of options to fit Bayesian mlVAR models using Stan, JAGS and Mplus, supplemented with a Monte Carlo simulation study. An empirical example is used to demonstrate the utility of mlVAR models in studying intra- and inter-individual variations in affective dynamics.

Abstract: Background: Alcohol-attributable costs to society are captured by cost-of-illness studies, however estimates are often not comparable, e.g. due to the omission of relevant cost components.

In this contribution we (1) summarize the societal costs attributable to alcohol use, and (2) estimate the total costs under the assumption that all cost components are considered. Methods: A systematic review and meta-analyses were conducted for studies reporting costs from alcohol consumption for the years 2000 and later, using the EMBASE and MEDLINE databases. Cost estimates were converted into 2019 international dollars (Int\$) per adult and into percentage of gross domestic product (GDP). For each study, weights were calculated to correct for the exclusion of cost indicators. Results: Of 1708 studies identified, 29 were included, and the mean costs of alcohol use amounted to 817.6 Int\$ per adult (95% confidence interval [CI] 601.8-1033.4), equivalent to 1.5% of the GDP (95% CI 1.2-1.7%). Adjusting for omission of cost components, the economic costs of alcohol consumption were estimated to amount to 1306 Int\$ per adult (95% CI 873-1738), or 2.6% (95% CI 2.0-3.1%) of the GDP. About one-third of costs (38.8%) were incurred through direct costs, while the majority of costs were due to losses in productivity (61.2%). Discussion: The identified cost studies were mainly conducted in high-income settings, with high heterogeneity in the employed methodology. Accounting for some methodological variations, our findings demonstrate that alcohol use continues to incur a high level of cost to many societies.

McNeish et al.: A primer on two-level dynamic structural equation models for intensive longitudinal data in Mplus   McNeish-Hamaker-2020

Abstract: Technological advances have led to an increase in intensive longitudinal data and the statistical literature on modeling such data is rapidly expanding, as are software capabilities. Common methods in this area are related to time-series analysis, a framework that historically has received little exposure in psychology. There is a scarcity of psychology-based resources introducing the basic ideas of time-series analysis, especially for data sets featuring multiple people. We begin with basics of N = 1 time-series analysis and build up to complex dynamic structural equation

models available in the newest release of Mplus Version 8. The goal is to provide readers with a basic conceptual understanding of common models, template code, and result interpretation. We provide short descriptions of some advanced issues, but our main priority is to supply readers with a solid knowledge base so that the more advanced literature on the topic is more readily digestible to a larger group of researchers.

McNeish et al.: Intensive longitudinal mediation in Mplus   McNeish-MacKinnon-2022

Abstract: Much of the existing longitudinal mediation literature focuses on panel data where relatively few repeated measures are collected over a relatively broad timespan. However, technological advances in data collection (e.g., smartphones, wearables) have led to a proliferation of short duration, densely collected longitudinal data in behavioral research. These intensive longitudinal data differ in structure and focus relative to traditionally collected panel data. As a result, existing methodological resources do not necessarily extend to nuances present in the recent influx of intensive longitudinal data and designs. In this tutorial, we first cover potential limitations of traditional longitudinal mediation models to accommodate unique characteristics of intensive longitudinal data. Then, we discuss how recently developed dynamic structural equation models (DSEMs) may be well-suited for mediation modeling with intensive longitudinal data and can overcome some of the limitations associated with traditional approaches. We describe four increasingly complex intensive longitudinal mediation models: (a) stationary models where the indirect effect is constant over time and people, (b) person-specific models where the indirect effect varies across people, (c) dynamic models where the indirect effect varies across time, and (d) cross-classified models where the indirect effect varies across both time and people. We apply each model to a running example featuring a mobile health intervention designed to improve health behavior of individuals with binge eating disorder. In each example, we provide annotated Mplus code and interpretation of the output

to guide empirical researchers through mediation modeling with this increasingly popular type of longitudinal data.

Nüst et al.: The Rockerverse: Packages and applications for containerisation with R

Daniel Nüst et al. "The Rockerverse: Packages and applications for containerisation with R". In: *The R Journal* 12.1 (2020), p. 437. DOI: 10.32614/rj-2020-007.

Abstract: The Rocker Project provides widely used Docker images for R across different application scenarios. This article surveys downstream projects that build upon the Rocker Project images and presents the current state of R packages for managing Docker images and controlling containers. These use cases cover diverse topics such as package development, reproducible research, collaborative work, cloud-based data processing, and production deployment of services. The variety of applications demonstrates the power of the Rocker Project specifically and containerisation in general. Across the diverse ways to use containers, we identified common themes: reproducible environments, scalability and efficiency, and portability across clouds. We conclude that the current growth and diversification of use cases is likely to continue its positive impact, but see the need for consolidating the Rockerverse ecosystem of packages, developing common practices for applications, and exploring alternative containerisation software.

Park et al.: Subgrouping with chain graphical VAR models

Jonathan J. Park, Sy-Miin Chow, et al. "Subgrouping with chain graphical VAR models". In: *Multivariate Behavioral Research* (2023).

Abstract: Recent years have seen the emergence of an "idio-thetic" class of methods to bridge the gap between nomothetic and idiographic inference. These methods describe nomothetic trends in idiographic processes by pooling intraindividual information across individuals to inform group-level

inference or vice versa. The current work introduces a novel "idio-thetic" model: the subgrouped chain graphical vector autoregression (scGVAR). The scGVAR is unique in its ability to identify subgroups of individuals who share common dynamic network structures in both lag(1) and contemporaneous effects. Results from Monte Carlo simulations indicate that the scGVAR shows promise over similar approaches when clusters of individuals differ in their contemporaneous dynamics and in showing increased sensitivity in detecting nuanced group differences while keeping Type-I error rates low. In contrast, a competing approach–the Alternating Least Squares VAR (ALS VAR) performs well when groups were separated by larger distances. Further considerations are provided regarding applications of the ALS VAR and scGVAR on real data and the strengths and limitations of both methods.

Park et al.: On subgrouping continuous processes in discrete time

Park-Fisher-Chow-etal-2023a

Park et al.: Evaluating discrete time methods for subgrouping continuous processes

Park-Fisher-Chow-etal-2023b

Abstract: Rapid developments over the last several decades have brought increased focus and attention to the role of time scales and heterogeneity in the modeling of human processes. To address these emerging questions, subgrouping methods developed in the discrete-time framework—such as the vector autoregression (VAR)—have undergone widespread development to identify shared

nomothetic trends from idiographic modeling results. Given the dependence of VAR-based parameters on the measurement intervals of the data, we sought to clarify the strengths and limitations of these methods in recovering subgroup dynamics under different measurement intervals. Building on the work of Molenaar and collaborators for subgrouping individual time-series by means of the subgrouped chain graphical VAR (scgVAR) and the subgrouping option in the group iterative multiple model estimation (S-GIMME), we present results from a Monte Carlo study aimed at addressing the implications of identifying subgroups using these discrete-time methods when applied to continuous-time data. Results indicate that discrete-time subgrouping methods perform well at recovering true subgroups when the measurement intervals are large enough to capture the full range of a system's dynamics, either via lagged or contemporaneous effects. Further implications and limitations are discussed therein.

Pesigan et al.: SEM-based methods to form confidence intervals for indirect effect: Still applicable given nonnormality, under certain conditions   Pesigan-Cheung-2020

Abstract: A SEM-based approach using likelihood-based confidence interval (LBCI) has been proposed to form confidence intervals for unstandardized and standardized indirect effect in mediation models. However, when used with the maximum likelihood estimation, this approach requires that the variables are multivariate normally distributed. This can affect the LBCIs of unstandardized and standardized effect differently. In the present study, the robustness of this approach when the predictor is not normally distributed but the error terms are conditionally normal, which does not violate the distributional assumption of ordinary least squares (OLS) estimation, is compared to four other approaches: nonparametric bootstrapping, two variants of LBCI, LBCI assuming the predictor is fixed (LBCI-Fixed-X) and LBCI based on ADF estimation (LBCI-ADF), and Monte Carlo. A simulation study was conducted using a simple mediation model and a serial mediation

model, manipulating the distribution of the predictor. The Monte Carlo method performed worst among the methods. LBCI and LBCI-Fixed-X had suboptimal performance when the distributions had high kurtosis and the population indirect effects were medium to large. In some conditions, the problem was severe even when the sample size was large. LBCI-ADF and nonparametric bootstrapping had coverage probabilities close to the nominal value in nearly all conditions, although the coverage probabilities were still suboptimal for the serial mediation model when the sample size was small with respect to the model. Implications of these findings in the context of this special case of nonnormal data were discussed.

Pesigan et al.: Monte Carlo confidence intervals for the indirect effect with missing data   Pesigan-Cheung-2023

Abstract: Missing data is a common occurrence in mediation analysis. As a result, the methods used to construct confidence intervals around the indirect effect should consider missing data. Previous research has demonstrated that, for the indirect effect in data with complete cases, the Monte Carlo method performs as well as nonparametric bootstrap confidence intervals (see MacKinnon et al., Multivariate Behavioral Research, 39(1), 99–128, 2004; Preacher & Selig, Communication Methods and Measures, 6(2), 77–98, 2012; Tofighi & MacKinnon, Structural Equation Modeling: A Multidisciplinary Journal, 23(2), 194–205, 2015). In this manuscript, we propose a simple, fast, and accurate two-step approach for generating confidence intervals for the indirect effect, in the presence of missing data, based on the Monte Carlo method. In the first step, an appropriate method, for example, full-information maximum likelihood or multiple imputation, is used to estimate the parameters and their corresponding sampling variance-covariance matrix in a mediation model. In the second step, the sampling distribution of the indirect effect is simulated using estimates from the first step. A confidence interval is constructed from the resulting sampling distribution.

A simulation study with various conditions is presented. Implications of the results for applied research are discussed.

Pesigan et al.: betaDelta and betaSandwich: Confidence intervals for standardized regression coefficients in R Pesigan-Sun-Cheung-2023

Abstract: The multivariate delta method was used by Yuan and Chan to estimate standard errors and confidence intervals for standardized regression coefficients. Jones and Waller extended the earlier work to situations where data are nonnormal by utilizing Browne's asymptotic distribution-free (ADF) theory. Furthermore, Dudgeon developed standard errors and confidence intervals, employing heteroskedasticity-consistent (HC) estimators, that are robust to nonnormality with better performance in smaller sample sizes compared to Jones and Waller's ADF technique. Despite these advancements, empirical research has been slow to adopt these methodologies. This can be a result of the dearth of user-friendly software programs to put these techniques to use. We present the betaDelta and the betaSandwich packages in the R statistical software environment in this manuscript. Both the normal-theory approach and the ADF approach put forth by Yuan and Chan and Jones and Waller are implemented by the betaDelta package. The HC approach proposed by Dudgeon is implemented by the betaSandwich package. The use of the packages is demonstrated with an empirical example. We think the packages will enable applied researchers to accurately assess the sampling variability of standardized regression coefficients.

Rousselet et al.: The percentile bootstrap: A primer with step-by-step instructions in R Rousselet-Pernet-Wilcox-2021

with step-by-step instructions in R". In: *Advances in Methods and Practices in Psychological Science* 4.1 (Jan. 2021), pp. 1–10. DOI: 10.1177/2515245920911881.

Abstract: The percentile bootstrap is the Swiss Army knife of statistics: It is a nonparametric method based on data-driven simulations. It can be applied to many statistical problems, as a substitute to standard parametric approaches, or in situations for which parametric methods do not exist. In this Tutorial, we cover R code to implement the percentile bootstrap to make inferences about central tendency (e.g., means and trimmed means) and spread in a one-sample example and in an example comparing two independent groups. For each example, we explain how to derive a bootstrap distribution and how to get a confidence interval and a p value from that distribution. We also demonstrate how to run a simulation to assess the behavior of the bootstrap. For some purposes, such as making inferences about the mean, the bootstrap performs poorly. But for other purposes, it is the only known method that works well over a broad range of situations. More broadly, combining the percentile bootstrap with robust estimators (i.e., estimators that are not overly sensitive to outliers) can help users gain a deeper understanding of their data than they would using conventional methods.

Russell et al.: Baseline protective behavioral strategy use predicts more moderate transdermal alcohol concentration dynamics and fewer negative consequences of drinking in young adults' natural settings.   Russell-Smyth-Turrisi-Rodriguez-2023

Michael A. Russell, Joshua M. Smyth, et al. "Baseline protective behavioral strategy use predicts more moderate transdermal alcohol concentration dynamics and fewer negative consequences of drinking in young adults' natural settings." In: *Psychology of Addictive Behaviors* (June 2023). ISSN: 0893-164X. DOI: 10.1037/adb0000941.

Abstract: Objective: Test whether frequent protective behavioral strategies (PBS) users report (a) fewer alcohol-related consequences and (b) less risky alcohol intoxication dynamics (measured via transdermal alcohol concentration [TAC] sensor "features") in daily life. Method: Two hundred

twenty-two frequently heavy-drinking young adults ($M_{\text{age}} = 22.3$ years) wore TAC sensors for 6 consecutive days. TAC features peak (maximum TAC), rise rate (speed of TAC increase), and area under the curve (AUC) were derived for each day. Negative alcohol-related consequences were measured in the morning after each self-reported drinking day. Past-year PBS use was measured at baseline. Results: Young adults reporting more frequent baseline PBS use showed (a) fewer alcohol-related consequences and (b) lower intoxication dynamics on average (less AUC, lower peaks, and slower rise rates). Limiting/stopping and manner of drinking PBS showed the same pattern of findings as the total score. Serious harm reduction PBS predicted fewer negative alcohol-related consequences, but not TAC features. Multilevel path models showed that TAC features peak and rise rate partially explained associations between PBS (total, limiting/stopping, and manner of drinking) and consequences. Independent contributions of PBS subscales were small and nonsignificant, suggesting that total PBS use was a more important predictor of risk/protection than the specific types of PBS used. Conclusions: Young adults using more total PBS may experience fewer alcohol-related consequences during real-world drinking episodes in part through less risky intoxication dynamics (TAC features). Future research measuring PBS at the daily level is needed to formally test TAC features as day-level mechanisms of protection from acute alcohol-related consequences.

Russell et al.: Transdermal sensor features correlate with ecological momentary assessment drinking reports and predict alcohol-related consequences in young adults' natural settings   **Russell-Turrisi-Smyth-2022**

Abstract: Background: Wearable transdermal alcohol concentration (TAC) sensors allow passive monitoring of alcohol concentration in natural settings and measurement of multiple features from

drinking episodes, including peak intoxication level, speed of intoxication (absorption rate) and elimination, and duration. These passively collected features extend commonly used self-reported drink counts and may facilitate the prediction of alcohol-related consequences in natural settings, aiding risk stratification and prevention efforts. Method: A total of 222 young adults aged 21-29 ($M_{\text{age}} = 22.3$, 64 female, 79% non-Hispanic white, 84% undergraduates) who regularly drink heavily participated in a 5-day study that included the ecological momentary assessment (EMA) of alcohol consumption (daily morning reports and participant-initiated episodic EMA sequences) and the wearing of TAC sensors (SCRAM-CAM anklets). The analytic sample contained 218 participants and 1274 days (including 554 self-reported drinking days). Five features—area under the curve (AUC), peak TAC, rise rate (rate of absorption), fall rate (rate of elimination), and duration—were extracted from TAC-positive trajectories for each drinking day. Day- and person-level associations of TAC features with drink counts (morning and episodic EMA) and alcohol-related consequences were tested using multilevel modeling. Results: TAC features were strongly associated with morning drink reports ($r = 0.60.7$) but only moderately associated with episodic EMA drink counts ($r = 0.30.5$) at both day and person levels. Higher peaks, larger AUCs, faster rise rates, and faster fall rates were significantly predictive of day-level alcohol-related consequences after adjusting for both morning and episodic EMA drink counts in separate models. Person means of TAC features added little above daily scores to the prediction of alcohol-related consequences. Conclusions: These results support the utility of TAC sensors in studies of alcohol misuse among young adults in natural settings and outline the specific TAC features that contribute to the day-level prediction of alcohol-related consequences. TAC sensors provide a passive option for obtaining valid and unique information predictive of drinking risk in natural settings.

Ryan et al.: Time to intervene: A continuous-time approach to network analysis and centrality   **Ryan-Hamaker-2021**

Oisin Ryan and Ellen L. Hamaker. "Time to intervene: A continuous-time approach to network

analysis and centrality". In: *Psychometrika* 87.1 (June 2021), pp. 214–252. DOI: .

Abstract: Network analysis of ESM data has become popular in clinical psychology. In this approach, discrete-time (DT) vector auto-regressive (VAR) models define the network structure with centrality measures used to identify intervention targets. However, VAR models suffer from time-interval dependency. Continuous-time (CT) models have been suggested as an alternative but require a conceptual shift, implying that DT-VAR parameters reflect total rather than direct effects. In this paper, we propose and illustrate a CT network approach using CT-VAR models. We define a new network representation and develop centrality measures which inform intervention targeting. This methodology is illustrated with an ESM dataset.

Savalei et al.: Computational options for standard errors and test statistics with incomplete normal and nonnormal data in SEM   Savalei-Rosseel-2021

Abstract: This article provides an overview of different computational options for inference following normal theory maximum likelihood (ML) estimation in structural equation modeling (SEM) with incomplete normal and nonnormal data. Complete data are covered as a special case. These computational options include whether the information matrix is observed or expected, whether the observed information matrix is estimated numerically or using an analytic asymptotic approximation, and whether the information matrix and the outer product matrix of the score vector are evaluated at the saturated or at the structured estimates. A variety of different standard errors and robust test statistics become possible by varying these options. We review the asymptotic properties of these computational variations, and we show how to obtain them using lavaan in R. We hope

that this article will encourage methodologists to study the impact of the available computational options on the performance of standard errors and test statistics in SEM.

Tofighi et al.: Improved inference in mediation analysis: Introducing the model-based constrained optimization procedure   Tofighi-Kelley-2020

Abstract: Mediation analysis is an important approach for investigating causal pathways. One approach used in mediation analysis is the test of an indirect effect, which seeks to measure how the effect of an independent variable impacts an outcome variable through one or more mediators. However, in many situations the proposed tests of indirect effects, including popular confidence interval-based methods, tend to produce poor Type I error rates when mediation does not occur and, more generally, only allow dichotomous decisions of "not significant" or "significant" with regards to the statistical conclusion. To remedy these issues, we propose a new method, a likelihood ratio test (LRT), that uses non-linear constraints in what we term the model-based constrained optimization (MBCO) procedure. The MBCO procedure (a) offers a more robust Type I error rate than existing methods; (b) provides a p-value, which serves as a continuous measure of compatibility of data with the hypothesized null model (not just a dichotomous reject or fail-to-reject decision rule); (c) allows simple and complex hypotheses about mediation (i.e., one or more mediators; different mediational pathways), and (d) allows the mediation model to use observed or latent variables. The MBCO procedure is based on a structural equation modeling framework (even if latent variables are not specified) with specialized fitting routines, namely with the use of non-linear constraints. We advocate using the MBCO procedure to test hypotheses about an indirect effect in addition to reporting a confidence interval to capture uncertainty about the indirect effect because this combination transcends existing methods.

L. Wang et al.: Investigating the impact of the time interval selection on autoregressive mediation modeling: Result interpretations, effect reporting, and temporal designs

Wang-Zhang-2020

Lijuan Wang and Qian Zhang. "Investigating the impact of the time interval selection on autoregressive mediation modeling: Result interpretations, effect reporting, and temporal designs". In: *Psychological Methods* 25.3 (June 2020), pp. 271–291. DOI: [10.1037/met0000235](10.1037/met0000235).

Abstract: This study investigates the impact of the time interval (the time passed between 2 consecutive measurements) selection on autoregressive mediation modeling (AMM). For a widely used autoregressive mediation model, via analytical derivations, we explained why and how the conventionally reported time-specific coefficient estimates (e.g., $\hat{a}\hat{b}$ and $\hat{c}'$) and inference results in AMM provide limited information and can arrive in even misleading conclusions about direct and indirect effects over time. Furthermore, under the stationarity assumption, we proposed an approach to calculate the overall direct and indirect effect estimates over time and the time lag lengths at which they reach maxima, using AMM results. The derivation results revealed that the overall direct and indirect effect curves are asymptotically invariant to the time interval selection, under stationarity. With finite samples and thus sampling errors and potential computing problems, however, our simulation results revealed that the overall indirect effect curves were better recovered when the time interval is selected to be closer to half of the time lag length at which the overall indirect effect reaches its maximum. An R function and an R Shiny app were developed to obtain the overall direct and indirect effect curves over time and facilitate the time interval selection using AMM results. Our findings provide another look at the connections between AMM and continuous time mediation modeling and the connections are discussed.

Zeileis et al.: Various versatile variances: An object-oriented implementation of clustered covariances in R

Zeileis-Koll-Graham-2020

Achim Zeileis, Susanne Köll, and Nathaniel Graham. "Various versatile variances: An object-

Abstract: Clustered covariances or clustered standard errors are very widely used to account for correlated or clustered data, especially in economics, political sciences, and other social sciences. They are employed to adjust the inference following estimation of a standard least-squares regression or generalized linear model estimated by maximum likelihood. Although many publications just refer to "the" clustered standard errors, there is a surprisingly wide variety of clustered covariances, particularly due to different flavors of bias corrections. Furthermore, while the linear regression model is certainly the most important application case, the same strategies can be employed in more general models (e.g., for zero-inflated, censored, or limited responses). In R, functions for covariances in clustered or panel models have been somewhat scattered or available only for certain modeling functions, notably the (generalized) linear regression model. In contrast, an object-oriented approach to "robust"' covariance matrix estimation - applicable beyond lm() and glm() - is available in the sandwich package but has been limited to the case of cross-section or time series data. Starting with sandwich 2.4.0, this shortcoming has been corrected: Based on methods for two generic functions (estfun() and bread()), clustered and panel covariances are provided in vcovCL(), vcovPL(), and vcovPC(). Moreover, clustered bootstrap covariances are provided in vcovBS(), using model update() on bootstrap samples. These are directly applicable to models from packages including MASS, pscl, countreg, and betareg, among many others. Some empirical illustrations are provided as well as an assessment of the methods' performance in a simulation study.