Aalto University
School of Science

TU-L0040 - Advanced Statistical Research Methods
Fall 2019-Spring 2020
6-10 Credits
In English

Note: All course lectures are openly accessible on the course website and the units are available as playlists at https://www.youtube.com/mronkko

Aalto University - Department of Industrial Engineering and Management

1 COURSE STAFF AND CONTACT INFORMATION

Dr. Mikko Rönkkö   Course instructor   mikko.ronkko@aalto.fi

- All course assignments are available and returned through Aalto MyCourses.
- All articles and other reading materials are available through Zotero.
- Except for personal matters, all course communications are done through the course discussion forum at Aalto MyCourses.

2 OVERVIEW OF THE COURSE

This is a blended learning course that contains both online and in-person elements. All in-person elements are optional except for the final seminar days that contain student presentations, so it is possible to complete the course mostly online. While online participation is possible, we follow a schedule, which means that completing the course as independent self-study for credits is not possible.

This is an advanced level course that is designed for student that have already some experience in using quantitative research designs and data analysis techniques in their own research. The goal of the course is to develop a more thorough understanding of how and why certain techniques are used and what principles these techniques are based on. The course focuses on longitudinal and multilevel designs and analyses. The techniques covered include all advanced econometrics and latent variable techniques used in *Academy of Management Journal, Journal of Applied Psychology, Journal of Management, Journal of Operations Management, Journal of Organizational Behavior, The Leadership Quarterly, Organization Science, Personnel Psychology,* and *Strategic Management Journal* as reported in a recent review[1]. The focus of the course will be on how these techniques can be used to empirically support causal claims taking particularly the issues of endogeneity and measurement validity into account.

The course consists of eight units, that each take two to four weeks and contain video lectures, online and in-person discussions, and assignments. The number of credits varies between 6-10 depending on which assignments students choose to complete. The content of each course component is explained later in the course brochure. All data analysis assignments can be completed with Stata or R. Mplus can be used for some of the assignment or their parts.

Credits	Content
6.5	- Pre-exam - Video lectures and online participation - Readings and written assignments for units 2 and 3 - Data-analysis assignment for units 2-9 - Monte Carlo study - Paper presentation - Final exam
+0.5 each	Written assignments for units 4-9
+0.5	Writing a methodological review

Max 10 credits. The number of credits is rounded down.

---

[1] John Antonakis, Nicolas Bastardoz, and Mikko Rönkkö, "On Ignoring the Random-Effects Assumption in Multilevel Models: Review, Critique, and Recommendations," *Organizational Research Methods* forthcoming (2019).

Aalto University - Department of Industrial Engineering and Management

3   PARTICIPATION AND SIGNING UP

This course is targeted to industrial engineering and management doctoral students that have already taken the course TU-L0021 – Statistical Research Methods or an equivalent quantitative research methods course. Faculty or students from other universities are admitted if space permits.

You can sign up for the course at the following URL using the enrolment key SRM2020 (in capital letters):

https://mycourses.aalto.fi/enrol/index.php?id=26732

4   LEARNING OUTCOMES

The main goal of the course is to provide an overview of most commonly used data analysis techniques and research designs that go beyond cross-sectional designs and simple linear models. Instead of just explaining how the methods are used, we focus on why certain methods are used and how and why these methods work. Completing the 6-credit base module will introduce you to the logic of structural equation models, generalized linear models, longitudinal data analysis, missing data, advanced measurement concepts, and endogeneity and causality.

The optional readings and written assignments and data-analysis assignments are more challenging and recommended only for those who plan to use the methods and designs covered in these assignments in their own research.

5   WORKLOAD

TO BE ADDED[2]. If you want to complete all mandatory and optional assignments on the course, you should book about two full work days per week for the duration of the course.

6   COURSE CONTENT

The course consists of a pre-exam, a learning diary, readings and written assignments, data analysis assignments, video lectures, online interactions, seminar sessions, and computer classes. All assignments are distributed and returned through Aalto MyCourses where online interactions also take place.

https://mycourses.aalto.fi/course/view.php?id=26732

The course is structured as ten units that are completed sequentially. At the beginning of a unit, the students receive a set of readings and video materials that they study independently. After this, the content and possible questions and thoughts that it raises are discussed on the course forum. Each unit concludes with a seminar where the materials are further discussed in-person.

The different parts of the course are listed below:

6.1   Pre-exam (mandatory)

The pre-exam is a written exam that you must pass to be able to participate and get credits. The exam materials are

Singleton, R., & Straits, B. C. (2018). *Approaches to social research* (Sixth edition). New York: Oxford University Press.

Wooldridge, J. M. (2013). *Introductory econometrics: A modern approach* (5th ed). Mason, OH: South-Western Cengage Learning. Chapters 1-9.

---

[2] Asko Karjalainen, Katariina Alha, and Suvi Jutila, *Anna Aikaa Ajatella: Suomalaisten Yliopisto-Opintojen Mitoitusjärjestelmä* (Oulun yliopisto, opetuksen kehittämisyksikkö, 2007).

Aalto University - Department of Industrial Engineering and Management

and either of these books depending on whether you want to use Stata or R as your main statistical analysis software for the course:

StataCorp. (2019). *Stata user's guide*. College Station, TX: Stata Press. ([PDF on Stata's website](#))

Wickham, H., & Grolemund, G. (2016). *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. O'Reilly Media, Inc. ([Available online](#))

all reading materials of TU-L0021 Statistical Research Methods.

The exam consists of four questions. The first question is term definitions, in which you need to define 8 terms from the course material. The three remaining questions are essays. The exam questions are chosen randomly from a question pool that can be found on the course website.

6.2 Readings and written assignments (2 mandatory, 6 optional)

The written assignments consist of reading methodological literature and empirical papers and then doing an assignment where you apply the methodological ideas that you just studied to analyze the empirical papers. You will be provided individual feedback on the written assignments and these are discussed in both the online and in-person interactions.

The written assignments 1, and 2 are mandatory. The written assignments 3-8 are optional and can completed for extra credits. All returned written assignment should follow the American Psychological Association (APA) Publication Manual style for article manuscripts. All work will be checked for plagiarism using the TurnItIn service and will be returned to the students for revising if plagiarism problems are found.

6.3 Presentation of a methodological paper (mandatory)

In this assignment, the students pick an article published in *Organizational Research Methods* during the last three years. The article should be chosen based on how useful or thought provoking you found the article to be. The articles that are discussed in detail on the course cannot be picked. Read through the article and prepare a 15-minute presentation of the article that you will give to the other students during the last seminar.

The purpose of this assignment is to familiarize you with the kind of research that *Organizational Research Methods* publishes and how articles presented in this journal should be read and interpreted. This is useful if you want to keep yourself up to date with the latest developments in research methods that are applied in management research.

6.4 Writing a methodological review (optional)

In this assignment, the students are given real articles that have been submitted to *Journal of Operations Management* and were assigned to be methods checked before the actual review. The task is to identify weaknesses from the empirical part and suggest ways that the authors can use to fix these problems. You are given a template that contains explanations of the most common problems present in articles submitted to the journal, but are of course free to give other suggestions as well.

All reviews will be given individual feedback and will be used as actual reviews by the journal. If you want, you can then continue working as a reviewer of the article if you are interested in seeing how the article develops in the review process or just commit a single reviewer statement and then drop out.

The purpose of this assignment is to teach you how to review the methods parts of empirical research and to show the kinds of problems that initial versions of articles submitted to journals often include.

6.5 Data-analysis assignments (8 mandatory)

Because the course is focused on doing quantitative analysis effectively, all data analysis assignments are mandatory. Each assignment contains three parts, tools, analyses, and challenge.

The tools part introduces you to general concepts such as reproducibility, programming your statistical software, data management, and visualization. Reproducibility refers to the being able to reproduce the exact analysis results presented in an article from the original raw data. Reproducibility can be compromised by poor data-analysis practices as well as by manual transferring of results from the statistical software to the research article.

Data management is an important skill of a quantitative researcher for two reasons. First, the raw data are typically in a different format than what the final analysis requires and thus some preparations are required before the actual analysis can even start. Second, the data preparation step is error prone and therefore it is important to know how to do data preparation in a transparent and reproducible way so that possible data preparation errors can be identified. Reproducibility can be increased by structuring all analyses as well-documented analysis files and by using the statistical software to produce the final tables shown in articles.

Data visualization is another important skill of a quantitative researcher. While good statistical software provides you with useful visualization tools (e.g. margins and marginsplot in Stata), there are scenarios where these tools are not applicable. For example, if you want to do a marginal effect or marginal prediction plot based on a latent variable model or a multilevel model, the built-in commands are often inadequate. Moreover, sometimes the impact of the different options and what they mean for the graphics may be difficult to evaluate.

The analyses part contains a series of short data-analysis tasks where the students apply the techniques discussed in the unit to either real or simulated dataset and interpret the results.

The challenge part contains one or two more challenging questions from the question pool of the final exam and can be useful for evaluating one's level of skill at various stages of the course.

Assignments can be done with either Stata or R. Some assignments can be also be completed fully or partly with Mplus. The course assumes that the students have a basic understanding of their statistical software before the start of the course. The MyCourses page contains additional resources for familiarizing yourself with different statistical software.

The answer to each data-analysis assignment must be submitted as two separate files. A commented analysis file (R file or Stata do file) that implements the analyses and a report that is produced using the reproducibility techniques taught on the course and also shows the analysis results and the required interpretations. All returned data analysis assignments will receive individual feedback.

6.6 Monte Carlo assignment (mandatory)

A Monte Carlo study is a study where repeated samples are generated from a known population which are then each analyzed with a technique being studied. This kind of studies are the most common form of evidence presented to support new data analysis techniques. In the Monte Carlo assignment, the task is to pick a rule of thumb or a guideline and present evidence either supporting or refuting the rule. Examples of rules that you might study are:

- Variance inflation factor (VIF) should be less than ten to ensure that the estimates are interpretable.
- Control variables that are uncorrelated with the dependent variable should be left out from a regression model[3].

---

[3] e.g., Herman Aguinis and Robert J. Vandenberg, "An Ounce of Prevention Is Worth a Pound of Cure: Improving Research Quality Before Data Collection," *Annual Review of Organizational Psychology and Organizational Behavior* 1, no. 1 (2014): 574, https://doi.org/10.1146/annurev-orgpsych-031413-091231.

Aalto University - Department of Industrial Engineering and Management

- Approximate fit indices indicating RMSEA < .06 and CFI > .95 indicate a good fitting model that can be causally interpreted.
- Regression analysis requires the error term to be normally distributed.
- Common method variance will always inflate regression estimates.
- Measurement error will always attenuate regression estimates.

The Monte Carlo studies are reported as short presentations (15 minutes) that are given in the last seminar session by the students.

The purpose of this assignment is to familiarize the students with the programming features of their statistical software as well as how methodological evidence is generated and interpreted. Being able to conduct Monte Carlo simulations can also be useful for applied researchers, because a simple Monte Carlo can sometimes be enough to convince your readers and reviewers of the soundness of your empirics (e.g. a power analysis in the case of small-sample research).

6.7 Video lectures (mandatory and optional)

The course follows a flipped classroom design and the lectures are delivered as videos that students watch on their own before online and in-person meetings with the instructor. The video library consists of about 100 videos forming a total of about 20 hours of content.

Each of the videos is assigned to a unit and the completion of watching the videos is tracked so that a student can complete a unit only if they have watched all mandatory videos for that unit. Some videos may contain interactive content (e.g. quizzes) that must be completed successfully to complete the video.

6.8 Online interactions (mandatory)

Each unit contains mandatory online participation in the form of course forum discussion. At the beginning of each unit, the students are assigned the materials for that unit including readings, assignments, and video content. After a few days of familiarizing with the content, students are expected to post a question or a comment about the materials on the course forum. These questions and comments are then discussed online with the course instructor. To pass an online interaction session, a student must either start at least one discussion thread by posting a question or comment or reply to at least one thread started by someone else. Participation is graded.

Detailed instructions for how to participate online are delivered by email when the course forums open.

6.9 Seminar sessions (mandatory and optional)

Each unit concludes with an in-person seminar. These seminars follow a flipped classroom design. Each seminar starts with an overview of the lecture materials that the students have viewed as videos in advance, but the focus is more on discussion and classroom assignments.

The seminar sessions are optional, but participation is highly recommended. The final seminar session includes student presentations and is mandatory.

6.10 Computer exercise sessions (optional and not graded)

The computer exercise sessions are optional and not graded. The course instructor is present to answer questions and give hand-to-hand guidance with the computer assignments. It is unlikely that a data-analysis assignment can be completed fully within the short time of a computer class, so it is recommended that the students start working on the assignments on their own already before the class so that we can use the class to focus on problems and questions.

In addition to working with data, we will do "manual calculations" without statistical software. Most statistical estimation involve minimizing or maximizing an estimation function. For example, least squares estimator minimizes the sum of squares of prediction errors (residuals)

and maximum likelihood estimator maximizes the likelihood of the data given a hypothesized model. During these exercises, we specify the estimation functions in Excel and estimate the models by minimizing or maximizing this function with the Solver tool in Excel. The purpose of these assignments is to make the students understand how the analysis tools work in practice. While you are unlikely to encounter problems with linear regression, more advanced modeling techniques may not always work well or you could get nonsense results. In these scenarios understanding what the analysis software actually does is very important so that you can troubleshoot the analysis.

6.11 Final exam

The course concludes with a final exam. The exam is done using either Stata and R and the answers are submitted as a report similar to the assignment reports. The exam will require the students to carry out and interpret analyses on datasets given by the instructor, troubleshooting problematic analyses, finding problems in analyses and research designs reported by other, and so on. The exam is an open book exam and the students can use their notes, written material, and internet, but they may not communicate with each other or anyone else over the internet.

The exam time is 4 hours.

7   UNITS AND SCHEDULE

The course consists of eight units. Each unit starts with a self-study of the materials, followed by online interactions where the materials are discussed on the course forum, and concludes with an in-person seminar and computer class. During the in-person teaching days, we will discuss theory and principles of quantitative research in the mornings and do hands-on assignments during the computer class in the afternoon. There will be no computer class on the final unit, but we will instead have student presentations and discuss theory and principles for the full day.

Date	Time	Location	Topic
Unit 1: Course introduction and pre-exam **Weeks 45-50**			
5.11.	14:15-16	TU 5 (Aalto)	Introductory lecture
5.11.	14:15	Online	Course forum for unit 1 opens. The task is to write a forum post where you introduce yourself to others on the course.
11.12.	9:00-12:00	TUAS-building	Course pre-exam
Unit 2: Causality, endogeneity, and natural experiments **Weeks 2-4**			
7.1.	8:00	Online	Course forum for unit 2 opens.
13.1.	24:00	Online	Deadline for written assignment 1 (mandatory)
15.1.	9:15-12	TU 5 (Aalto)	Unit 2 seminar
22.1.	13:15-17	TU 4 (Aalto)	Unit 2 computer class
26.1.	24:00	Online	Deadline for data-analysis assignment 1
Unit 3: Structural equation modeling, mediation, and instrumental variables **Weeks 4-6**			

Aalto University - Department of Industrial Engineering and Management

20.1.	8:00	Online	Course forum for unit 3 opens.
27.1.	24:00	Online	Deadline for written assignment 2 (mandatory)
29.1.	9:15-12	TU 5 (Aalto)	Unit 3 seminar
5.2.	13:15-17	TU 4 (Aalto)	Unit 3 computer class
9.2.	24:00	Online	Deadline for data-analysis assignment 2
Unit 4: Generalized linear models **Weeks 6-8**			
3.2.	8:00	Online	Course forum for unit 4 opens.
10.2.	24:00	Online	Deadline for written assignment 3 (optional)
12.2.	9:15-12	TU 5 (Aalto)	Unit 4 seminar
19.2.	13:15-17	TU 4 (Aalto)	Unit 4 computer class
23.3.	24:00	Online	Deadline for data-analysis assignment 3
Unit 5: Moderation, structural equation modeling with latent variables **Week 8-10**			
17.2.	8:00	Online	Course forum for unit 5 opens.
24.2.	24:00	Online	Deadline for written assignment 4 (optional)
26.2.	9:15-12	TU 5 (Aalto)	Unit 5 seminar
4.3.	13:15-17	TU 4 (Aalto)	Unit 5 computer class
8.3.	24:00	Online	Deadline for data-analysis assignment 4
Unit 6: Measurement and measurement validation **Week 10-12**			
2.3.	8:00	Online	Course forum for unit 6 opens.
9.3	24:00	Online	Deadline for written assignment 5 (optional)
11.3.	9:15-12	TU 5 (Aalto)	Unit 6 seminar
18.3.	13:15-17	TU 4 (Aalto)	Unit 6 computer class
22.3.	24:00	Online	Deadline for data-analysis assignment 5
Unit 7: Multilevel models **Weeks 13-15**			
23.3.	8:00	Online	Course forum for unit 7 opens.
29.3.	24:00	Online	Deadline for written assignment 6 (optional)

Aalto University - Department of Industrial Engineering and Management

31.3.	9:15-12	TU 5 (Aalto)	Unit 7 seminar
8.4.	13:15-17	TU 4 (Aalto)	Unit 7 computer class
12.4.	24:00	Online	Deadline for data-analysis assignment 6
Unit 8: Sampling, sample selection, and missing data **Weeks 15-17**			
6.4.	8:00	Online	Course forum for unit 8 opens.
13.4.	24:00	Online	Deadline for written assignment 7 (optional)
15.4.	9:15-12	TU 5 (Aalto)	Unit 8 seminar
22.4.	13:15-17	TU 4 (Aalto)	Unit 8 computer class
26.4.	24:00	Online	Deadline for data-analysis assignment 7
Unit 9: Longitudinal analysis and multilevel SEM **Weeks 17-19**			
20.4.	8:00	Online	Course forum for unit 9 opens.
27.4.	24:00	Online	Deadline for written assignment 8 (optional)
29.4.	9:15-12	TU 5 (Aalto)	Unit 9 seminar
6.5.	13:15-17	TU 4 (Aalto)	Unit 9 computer class
10.5.	24:00	Online	Deadline for data-analysis assignment 8
Unit 10: Student presentations and course conclusion **Weeks 21**			
18.5.	24:00	Online	Deadline for Monte Carlo assignment
20.5.	9:15-17	AS 4 (Aalto)	Unit 10 seminar with student presentations
Final exam **Week 22**			
27.5.	13:15-17	TU 4 (Aalto)	Final exam

The detailed description of the ten units follows.

7.1 Unit 1: Course introduction

The purpose of this unit is to introduce the students to the work practices on the course. The students may not be familiar with working on a blended learning / online course, and this first unit will introduce the course tools (MyCourses/Moodle, Zotero, TurnItIn) and work practices. We will also discuss what the requirements for the written assignments are and go over the course plagiarism policy. We will also discuss the importance of reproducibility.

The unit concludes with the course pre-exam.

7.2 Unit 2: Causality, endogeneity, and natural experiments

Research designs and techniques:

The unit discusses assumptions and principles behind making causal claims. We briefly cover different philosophical approaches for causation after which we focus on the counterfactual model. The average causal (treatment) effect is contrasted with other possible ways to define causal effects.

We also cover natural experimental designs for making causal claims. Difference-in-differences, propensity score matching, and regression discontinuity models are introduced as statistical tools for causal analysis.

The concept of endogeneity and why it presents a dilemma is discussed. We discuss the use of instrumental variables to address endogeneity and the instrument relevance and exclusion criterions.

Computer tools:

Besides the analysis techniques presented during this unit, the computer assignment addresses random number generation and reproducibility. Generating random numbers is a useful skill because it allows one to generate datasets from known populations and then test if a particular analysis technique can recover the correct population values. This is useful for example when you are learning new analysis techniques or are troubleshooting existing analyses (e.g. checking for model identification).

This unit introduces Markdown and StatTag[4] to automate linking analysis results to research reports as well as how to export results in CSV or Excel format so that they can be copy-pasted to research reports as full tables.

Readings:

Hittchcock, C. (2010). Causation. In S. Psillos & M. Curd (Eds.), *The Routledge companion to philosophy of science* (1. publ. in paperback, pp. 317–326). London: Routledge.

Morgan, S. L., & Winship, C. (2007). *Counterfactuals and causal inference methods and principles for social research*. Leiden: Cambridge University Press. (Chapters 1-2)

Wooldridge, J. M. (2013). *Introductory econometrics: a modern approach* (5th ed.). Mason, OH: South Western, Cengage Learning. (Chapter, 15)

Lonati, S., Quiroga, B. F., Zehnder, C., & Antonakis, J. (2018). On doing relevant and rigorous experiments: Review and recommendations. *Journal of Operations Management*, *64*, 19–40. https://doi.org/10.1016/j.jom.2018.10.003

Antonakis, J., Bendahan, S., Jacquart, P., & Lalive, R. (2010). On making causal claims: A review and recommendations. *The Leadership Quarterly*, *21*(6), 1086-1120. doi:10.1016/j.leaqua.2010.10.010

Athey, S., & Imbens, G. W. (2017). The state of applied econometrics: Causality and policy evaluation. *Journal of Economic Perspectives*, *31*(2), 3–32.

Mochon, D., Johnson, K., Schwartz, J., & Ariely, D. (2017). What Are Likes Worth? A Facebook Page Field Experiment. *Journal of Marketing Research (JMR)*, *54*(2), 306–317.

Ketokivi, M., & McIntosh, C. N. (2017). Addressing the endogeneity dilemma in operations management research: Theoretical, empirical, and pragmatic considerations. *Journal of*

---

[4] Leah J Welty et al., *StatTag* (Galter Health Science Library, 2016), https://doi.org/10.18131/G36K76.

Aalto University - Department of Industrial Engineering and Management

Operations Management, *52*, 1–14. https://doi.org/10.1016/j.jom.2017.05.001

Deaton, A., & Cartwright, N. (2018). Understanding and misunderstanding randomized controlled trials. *Social Science & Medicine*, *210*, 2–21. https://doi.org/10.1016/j.socscimed.2017.12.005

Cuddy, A. J. C., Wilmuth, C. A., Yap, A. J., & Carney, D. R. (2015). Preparatory power posing affects nonverbal presence and job interview performance. *Journal of Applied Psychology*, *100*(4), 1286–1295. https://doi.org/10.1037/a0038543

7.3 Unit 3: Structural equation modeling, mediation, and instrumental variables

Research designs and techniques:

This unit explains the basic principles of structural equation modeling with observed variables for independent observations (e.g. a cross-sectional study). We discuss identification, estimation, interpretation, and diagnostics. The focus will be on maximum likelihood (ML) estimation.

Mediation is introduced in more detail including estimation of models with endogenous mediators.

We discuss further instrumental variable techniques including estimation of instrumental variable models with ML SEM, generalized method of moments (GMM), and two-stage least squares. Tests for endogeneity, weak instruments, and instrument exclusion are introduced as is the concept of plausible exogeneity.

Computer tools:

This unit introduces two new tools: matrix algebra and numerical optimization. Understanding the basics of matrix algebra is useful for two reasons. First, many texts on quantitative methods that go beyond introductory level use matrix notation if not in full then at least in part. If you understand how to read matrix equations and what these equations mean, this will make it easier to study these sources. Second, if you are generating datasets yourself, doing that with matrix equations instead of writing a separate (scalar) equation for each variable is much more straightforward. The introduction covers the topics of matrix sums, matrix multiplication, inverse, and determinant.

Every statistical model addressed on the course is estimated by minimizing and objective function. While some techniques, such as least-squares regression, has a closed from solution which can be simply calculated by applying algebra to the data, many techniques do not. In these cases, the estimates are obtained by numerically minimizing or maximizing the objective function. This unit covers the basics of numerical optimization and how gradient vector and Hessian matrix are used in this process and can be used for troubleshooting non-convergent models.

Readings:

Peng, D. et al. (2019). Addressing Endogeneity in Operations Management Research: Recent Developments, Common Problems, and Directions for Future Research. *Journal of Operations Management*.

Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York, NY: Guilford Press. (Chapters 1, 5-8, 153 pages)

7.4 Unit 4: Moderation, transformations, and generalized linear models

Research designs and techniques:

We discuss the use of linear regression to estimate moderation models and how transformations can be used for modeling non-linear effects.

Aalto University - Department of Industrial Engineering and Management

The unit introduced the generalized linear model, which is an extension to linear regression covering most commonly used single dependent variable models as special cases (e.g. logistic regression, Poisson regression, tobit regression, etc.). Maximum likelihood estimation of these models is introduced.

We will focus on the choice of non-linear model as well as their graphical interpretation and common non-graphical interpretations including their misuses.

Computer tools:

The computer tools covered in this unit are about visualization. We will focus particularly on two-way plots and their application for calculating marginal prediction and marginal effects plots. These plots are produced by both the built-in plots (e.g. margins in Stata) as well as without these tools by using generic graphing functions.

Readings:

Wooldridge, J. M. (2013). *Introductory econometrics: a modern approach* (5th ed.). Mason, OH: South Western, Cengage Learning. (Chapters 6-8, 17)

Dawson, J. F. (2014). Moderation in management research: What, why, when, and how. *Journal of Business and Psychology*, *29*(1), 1–19. https://doi.org/10.1007/s10869-013-9308-7

Rönkkö, M., Aalto, E., Tenhunen, H., & Aguirre-Urreta, M. (2019). Eight simple guidelines for improved understanding of transformations and nonlinear effects. *Organizational Research Methods*.

Hekman, D. R., Aquino, K., Owens, B. P., Mitchell, T. R., Schilpzand, P., & Leavitt, K. (2010). An Examination of Whether and How Racial and Gender Biases Influence Customer Satisfaction. *Academy of Management Journal*, 53(2), 238-264. doi:10.5465/AMJ.2010.49388763

Deephouse, D. L. (1999). To be different, or to be the same? It's a question (and theory) of strategic balance. *Strategic Management Journal*, 20(2), 147-166. doi:10.1002/(SICI)1097-0266(199902)20:2<147::AID-SMJ11>3.0.CO;2-Q

Mochon, D., Johnson, K., Schwartz, J., & Ariely, D. (2017). What Are Likes Worth? A Facebook Page Field Experiment. *Journal of Marketing Research (JMR)*, *54*(2), 306–317.

7.5 Unit 5: Structural equation modeling with latent variables, latent variable moderation

Research designs and techniques:

The unit introduces students to structural equation models with both latent and observed variables for cross-sectional designs. Structural equation models are a basic tool in the course and many of the techniques covered can be considered as special cases of this general modeling framework.

Different techniques for analyzing latent interactions are presented. Estimation techniques other than ML are introduced (MIIV, WLS).

Various diagnostics and model fit indices are discussed.

Computer tools:

This unit introduces loops, which allow automating repetitive tasks in statistical software. Loops are used for example to construct and export tables of model fit indices. Such tables are not available for export directly from the software that we use, but must be constructed manually.

We also continue practicing matrix calculations and troubleshoot problematic models by inspecting the output of numerical optimization routines of statistical software.

Readings:

Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York, NY: Guilford Press. (Chapter 5-9, 10, 12)

Ropovik, I. (2015). A cautionary note on testing latent variable models. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01715

Antonakis, J., Bendahan, S., Jacquart, P., & Lalive, R. (2010). On making causal claims: A review and recommendations. *The Leadership Quarterly*, *21*(6), 1086–1120. doi:10.1016/j.leaqua.2010.10.010

Mesquita, L. F., & Lazzarini, S. G. (2008). Horizontal and Vertical Relationships in Developing Economies: Implications for Smes' Access to Global Markets. *Academy of Management Journal*, *51*(2), 359–380.

Yli-Renko, H., Autio, E., & Sapienza, H. J. (2001). Social capital, knowledge acquisition, and knowledge exploitation in young technology-based firms. *Strategic Management Journal*, *22*(6-7), 587–613. doi:10.1002/smj.183

7.6 Unit 6: Measurement and measurement validation

Research designs and techniques:

Measurement related problems, particularly in studies where data are collected with surveys, are some of the most common reasons for rejecting manuscripts from publication. In this unit we will address issues related to reliability, validity, and measurement modeling.

A fairly typical way of addressing reliability in empirical research is to first calculate "Cronbach's alphas" that are then compared against fixed 0.70 cutoff. If all alphas exceed 0.70, reliability is declared acceptable and then ignored for the remainder of the article. This practice is problematic for two main reasons: alpha is rarely the best reliability index and reliability assessment is not a yes or no decision. We will discuss different reliability indices and their assumptions and how reliability statistics and the effect of unreliability can be interpreted and quantified beyond using fixed benchmarks. The concept of bi-factor model is introduced to support the discussion.

The concept of validity is more challenging. We will discuss a couple of different measurement theories and how these relate to current (psychometric) modeling practices. The problematic concepts of formative measurement and multidimensional constructs are discussed.

Common method variance is another issue that is often raised during the review process of survey-based research. We will discuss the different potential causes of this phenomenon, what different statistical approaches are available to detect and possibly control for method variance, and why most of these techniques are actually inadequate for the task.

Finally, we will also discuss what to do when a structural equation model does not fit or fails to converge. Exploratory structural equation models and Bayesian structural equation models are introduced as possible solutions for relaxing the strict covariance implications of traditional linear structural equation models.

Computer tools:

This unit discusses programming and data management. We discuss how data can be merged and reshaped and how repetitive tasks can be automated with programs (Stata) and functions (R).

Readings:

DeVellis, R. F. (2017). *Scale development theory and applications* (4th ed.). Thousand Oaks: Sage. (Chapters 1-4, 6)

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity.

Aalto University - Department of Industrial Engineering and Management

Psychological Review, *111*(4), 1061.

Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York, NY: Guilford Press. (Chapter 8)

Guide, D., & Ketokivi, M. (2015). Notes from the editors: Redefining some methodological criteria for the journal. *Journal of Operations Management*, *37*, v–viii. https://doi.org/10.1016/S0272-6963(15)00056-X

McNeish, D. (2017). Thanks Coefficient Alpha, We'll Take It From Here. *Psychological Methods*. https://doi.org/10.1037/met0000144

Rönkkö, M., & Cho, E. (2019). Discriminant validity: A synthesis of definitions and a test of techniques. *Organizational Research Methods*.

Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of Method Bias in Social Science Research and Recommendations on How to Control It. *Annual Review of Psychology*, *63*(1), 539–569. https://doi.org/10.1146/annurev-psych-120710-100452

7.7   Unit 7: Multilevel models

Research designs and techniques:

Independence of observations is one of the key assumptions of linear regression. However, this assumption is violated when working with longitudinal or clustered data (i.e. members in teams) because of unobserved heterogeneity. We start by introducing the concepts of nested and crossed clustering, the between, within, and contextual effects, and the concepts of random and fixed effects.

The data-analysis techniques covered include GLS random and fixed effects estimators and the Mundlak/correlated random effect/hybrid approach from the econometrics tradition for analyzing these types of data and mixed/random/multilevel models from the modeling tradition. Generalized estimation equations and cluster robust standard errors are presented as an alternative to these techniques. Additionally, we discuss the use of cluster means or cluster mean centering for estimating effects on different levels. Different empirical approaches for model choice (Hausman test and different nested model tests) are presented. The data analysis techniques conclude with models for cross-level interactions and generalized mixed models.

Finally, we will also discuss what to do when a multilevel model fails to converge.

Computer tools:

This unit will teach the students how to run Monte Carlo simulations how to summarize the Monte Carlo results as tables and figures.

Readings:

Luke, D. A. (2004). *Multilevel Modeling*. Sage. (Chapters 1, 2)

Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York, NY: Guilford Press. (Chapter 11)

Hausknecht, J. P., Hiller, N. J., & Vance, R. J. (2008). Work-Unit Absenteeism: Effects of Satisfaction, Commitment, Labor Market Conditions, and Time. *Academy of Management Journal*, *51*(6), 1223–1245. doi:10.5465/AMJ.2008.35733022

McNeish, D., Stapleton, L. M., & Silverman, R. D. (2016). On the Unnecessary Ubiquity of Hierarchical Linear Modeling. *Psychological Methods*. http://dx.doi.org/10.1037/met0000078

Antonakis, J., Bendahan, S., Jacquart, P., & Lalive, R. (2010). On making causal claims: A review and recommendations. *The Leadership Quarterly*, *21*(6), 1086–1120. doi:10.1016/j.leaqua.2010.10.010

Aalto University - Department of Industrial Engineering and Management

Deephouse, D. L. (1999). To be different, or to be the same? It's a question (and theory) of strategic balance. *Strategic Management Journal*, *20*(2), 147–166.

Antonakis, J., Bastardoz, N., & Rönkkö, M. (2019). On ignoring the random-effects assumption in multilevel models: Review, critique, and recommendations. *Organizational Research Methods*, *forthcoming*.

Ketokivi (2019) Avoiding Bias and Fallacy in Survey Research: A Behavioral Multilevel Approach. *Journal of Operations Management*

Bou, J. C., & Satorra, A. (2017). Univariate Versus Multivariate Modeling of Panel Data: Model Specification and Goodness-of-Fit Testing. *Organizational Research Methods*, 109442811771550. https://doi.org/10.1177/1094428117715509

Bliese, P. D., & Ployhart, R. E. (2002). Growth Modeling Using Random Coefficient Models: Model Building, Testing, and Illustrations. *Organizational Research Methods*, *5*(4), 362–387. https://doi.org/10.1177/109442802237116

Aguinis, H., Gottfredson, R. K., & Culpepper, S. A. (2013). Best-Practice Recommendations for Estimating Cross-Level Interaction Effects Using Multilevel Modeling. *Journal of Management*, *39*(6), 1490–1528. https://doi.org/10.1177/0149206313478188

Holcomb, T. R., Combs, J. G., Sirmon, D. G., & Sexton, J. (2010). Modeling Levels and Time in Entrepreneurship Research An Illustration With Growth Strategies and Post-IPO Performance. *Organizational Research Methods*, *13*(2), 348–389. https://doi.org/10.1177/1094428109338401

7.8 Unit 8: Sampling, sample selection, and missing data

Research designs and techniques:

Missing data refers to a phenomenon where part of the data that would be ideally be available for an analysis is in fact not available. Sample selection is a special case of missing data where the missigness depends systematically on the studied phenomenon. A classic example is a study of how strongly womens' income depend on their level of education. Studying this effect is challenging because some women choose to not work and stay at home with the children and this decision depends on the income that they would receive from the job markets.

This unit starts by introducing the three different missing data mechanisms, missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR) and different missing data patterns. After this, we will discuss when and why missing data could become a problem and how a potential missing data or sample selection issue can be diagnosed.

Multiple imputation and maximum likelihood estimation with missing data (FIML) are presented as solutions to the MCAR and MAR scenarios. Selection models, including the widely used Heckman model, are presented as a potential solution to MCAR scenarios, focusing particularly on the assumptions that these techniques make and the fact that sometimes these assumptions may be more untenable than the MAR assumption.

Finally, we will discuss on endogenous selection where data are not missing, but whether a company makes a decision depends on the company's expected returns from that decision.

Computer tools:

Besides the analysis techniques discussed in this unit, there will be no new computer tools. The students will have an opportunity to ask questions and get guidance on their Monte Carlo projects.

Readings:

Wooldridge, J. M. (2013). *Introductory econometrics: a modern approach* (5th ed.). Mason, OH: South Western, Cengage Learning. (Chapters 9.5, 17.1, 17.4, 17.5)

Aalto University - Department of Industrial Engineering and Management

Little, T. D. (Ed.). (2013). *The Oxford handbook of quantitative methods* (Vol. 2). New York: Oxford University Press. (Chapter 27)

Enders, C. K. (2010). *Applied missing data analysis*. New York: Guilford Publications. (Chapters 1-4, 10)

Clougherty, J. A., Duso, T., & Muck, J. (2015). Correcting for Self-selection Based Endogeneity in Management Research: Review, Recommendations and Simulations. *Organizational Research Methods*. https://doi.org/10.1177/1094428115619013

Rogelberg, S. G., & Stanton, J. M. (2007). Introduction Understanding and Dealing With Organizational Survey Nonresponse. *Organizational Research Methods*, *10*(2), 195–209. https://doi.org/10.1177/1094428106294693

Newman, D. A. (2014). Missing Data Five Practical Guidelines. *Organizational Research Methods*, *17*(4), 372–411. https://doi.org/10.1177/1094428114548590

7.9 Unit 9: Longitudinal analysis and multilevel SEMs

Research designs and techniques:

Longitudinal analysis can refer to either studying the effects of time on the dependent variable or studying the effect of the dependent variable itself on its future values. This unit introduces the concepts of lagged dependent variables, and the choice of lags more generally, autocorrelation, dynamic panel models, endogeneity in panel models.

The econometrics techniques cover the use of instrumental variables to address the endogeneity issue in dynamic panels, particularly by using the Arellano-Bond approach. Longitudinal analysis with structural equation models are covered using the latent growth model, cross-lagged model and its variants, and latent change score models.

Stata's generalized structural equation modeling and other approaches to multilevel structural equation modeling, including the MUML and WLSMV estimators as implemented in Mplus and Lavaan are presented. Longitudinal latent class analysis is introduced briefly.

Computer tools:

Besides the analysis techniques discussed in this unit, there will be no new computer tools. The students will have an opportunity to ask questions and get guidance on their Monte Carlo projects.

Readings:

Bliese, P. D., & Ployhart, R. E. (2002). Growth Modeling Using Random Coefficient Models: Model Building, Testing, and Illustrations. *Organizational Research Methods*, *5*(4), 362–387. https://doi.org/10.1177/109442802237116

Ketokivi (2019) Avoiding Bias and Fallacy in Survey Research: A Behavioral Multilevel Approach. *Journal of Operations Management*

Antonakis, J., Bastardoz, N., & Rönkkö, M. (2019). On ignoring the random-effects assumption in multilevel models: Review, critique, and recommendations. *Organizational Research Methods*, *forthcoming*.

Bou, J. C., & Satorra, A. (2017). Univariate Versus Multivariate Modeling of Panel Data: Model Specification and Goodness-of-Fit Testing. *Organizational Research Methods*, 109442811771550. https://doi.org/10.1177/1094428117715509

Aguinis, H., Gottfredson, R. K., & Culpepper, S. A. (2013). Best-Practice Recommendations for Estimating Cross-Level Interaction Effects Using Multilevel Modeling. *Journal of Management*, *39*(6), 1490–1528. https://doi.org/10.1177/0149206313478188

Little, T. D. (Ed.). (2013). *The Oxford handbook of quantitative methods* (Vol. 2). New York:

Aalto University - Department of Industrial Engineering and Management

Oxford University Press. (Chapter 18)

Hamaker, E. L., Kuiper, R. M., & Grasman, R. P. (2015). A critique of the cross-lagged panel model. *Psychological Methods*, *20*(1), 102–116.'

Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York, NY: Guilford Press. (Chapter 11)

Ployhart, R. E., & Vandenberg, R. J. (2010). Longitudinal Research: The Theory, Design, and Analysis of Change. *Journal of Management*, *36*(1), 94–120. https://doi.org/10.1177/0149206309352110'

Meier, L. L., & Spector, P. E. (2013). Reciprocal effects of work stressors and counterproductive work behavior: A five-wave longitudinal study. *Journal of Applied Psychology*, *98*(3), 529-539.

Wooldridge, J. M. (2013). *Introductory econometrics: a modern approach* (5th ed.). Mason, OH: South Western, Cengage Learning. (Chapters 13.2)

7.10 Unit 10: Student presentations and course conclusion

The unit we will address research design and current issues and debates in quantitative management research. We end with a summary of the course units. This unit concludes with a full day seminar.

The seminar has two sets of student presentations. Each student is required to read one recent article from *Organizational Research Methods* and present that in the class. Additionally, the students present their Monte Carlo simulations.

8   GRADING

All submitted work will be graded between 1-5 and your grade will be a weighted average of the parts of the course that you completed.

Course part	Weight	Notes
Pre-exam	4	
Seminar and forum participation	1 each (total 9)	By default, you will receive 2 for posting something on the course forum or being present at a seminar and your grade will increase based on your unit participation
Final exam	15	
Assignments	1 each (total 11)	8 data analysis assignments, 8 written assignments, 1 Monte Carlo presentation, 1 methodological article presentation, 1 methods review.

9   COURSE MATERIAL

The reading materials for the course are distributed through the Zotero reference management system. To get access to the materials:

1. Create a user account at Zotero.org
2. Email your username to the course instructor
3. The course instructor will send you an invitation to a group library, which you need to accept.

After you have accepted the invitation, you can access the material either online with a web browser or by installing the Zotero software on your computer. See the MyCourses page for information.

All seminars will be recorded on video and will be made available for the students.

Aalto University - Department of Industrial Engineering and Management

9.1 Books

To be added.

XX pages total.

9.2 Articles

To be added.

XX pages total.

9.3 Empirical articles used as examples

To be added.

XX pages total.

10 ABOUT THE INSTRUCTOR

The course is taught by Dr. Mikko Rönkkö. He is Associate Professor of Entrepreneurship at Jyväskylä University School of Business and Economics and Docent in Management at Aalto University. His work focuses on the use of quantitative research methods in management and and on software entrepreneurship. He has published articles about quantitative research methods in multiple journals including leading applied research methods journals, (e.g. Organizational Research Methods, Psychological Methods), and leading field journals (e.g. Journal of Operations Management and MIS Quarterly), and has taught quantitative research methods at multiple universities in and outside Finland. He is currently a department editor at Journal of Operations Management and on the editorial boards of Organizational Research Methods and Entrepreneurship Theory and Practice.