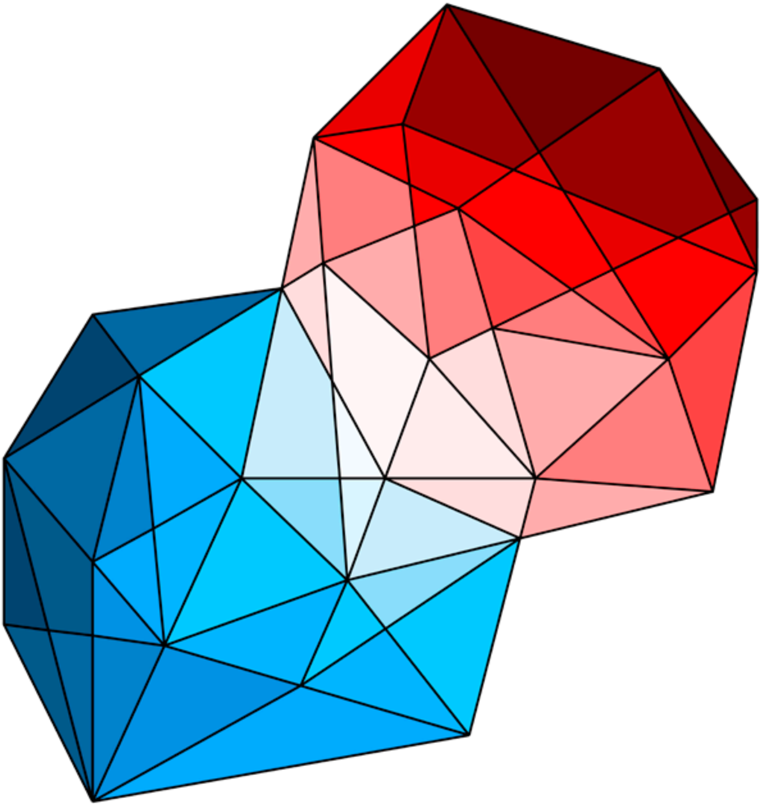
university of groningen

Prof. dr. Marijtje A.J. van Duijn

Statistics in the Social Sciences: The Best of Two Worlds



Inaugural lecture

13 januari 2023

Statistics in the Social Sciences: The Best of Two Worlds

Statistics in the Social Sciences: The Best of Two Worlds

Inaugural lecture by

Prof. dr. M.A.J. van Duijn

13 januari 2023

On acceptance of the post of professor of

Statistics, in particular Models for Social Network Analysis

at the

Faculty of Behavioral and Social Sciences

University of Groningen

university of
groningen

Published by University of Groningen Press
Broerstraat 4
9712 CP Groningen
https://ugp.rug.nl/

Design and layout: LINE UP boek en media bv | Riëtte van Zwol, Mirjam Kroondijk
Illustrations: Marchien Boonstra
Author photo: Hylkia van Meer

DOI: https://doi.org/10.21827/639c2fb106c33

Dear Members of the University Board,
Dear colleagues, friends and students,

0. Introduction

Why do I like being a statistician in the department of Sociology? What is special about statistics in the social sciences? And how can statistics serve as a link between two different worlds to make it the best? In this inaugural lecture, I will try to give an answer to these questions.

A – too - simple answer is that statistics in the social sciences is "applied", or "statistics in context". Both terms are often used. In mathematical statistics to distinguish the theory of statistics from the practice of statistics (Diggle, 2015). And by statistics professors who advocate the teaching of statistics in an applied, contextual, setting, using examples that are close to students' interests (see, e.g., Blatchley, 2019).
By the way: I consider myself an applied statistician and make an effort to teach statistics in context.

A better answer is that statisticians in the social sciences practice in a context with social scientists - often referred to as applied researchers by statisticians. They work with colleagues and students in a consultation setting. To give good advice it is necessary to not only understand the research question but also to familiarize oneself with the context of the scientific problem (see also Heiberger & Holland, 2013; Diggle, 2015).

This approach to statistics in an applied context very much captures the spirit in which I and my statistician-colleagues in Sociology and other departments in the Faculty of Social and Behavioral Sciences teach our students and collaborate with our colleagues.

You might think that students – teaching – and colleagues – collaborating - populate the two worlds of which I perceive the intersection as 'the best'. This is indeed the case, but in this lecture, I would like to explain more precisely what I mean by 'the best' in the practice of statistics. For that I need the idea of statistics connecting two worlds as depicted by Kass (2011) in his article "Statistical Inference: The Big Picture". He sketches two worlds, a real world and a theoretical world, as shown in Figure 1. The theoretical world contains theoretical models and statistical models, whereas the observed data – to support or investigate the theoretical models – live in the real world.

In the following, I will present my view on the role of applied statistics in the social sciences, by investigating the Big Picture in more detail. I will look deeper into the bidirectional relation between theoretical models and statistical models in the theoretical world, and the role of observed data in that relation. Then, I will focus on statistical models for social network data in particular. Finally, I investigate the role of statistics and statisticians in obtaining the best of the two worlds in order to arrive at valid conclusions.

I will only touch upon important themes like teaching, research design, testing or effect sizes, model estimation, or causality. Excellent chapters on these topics and the role of statistics can be found in Panter & Sterba (2011).

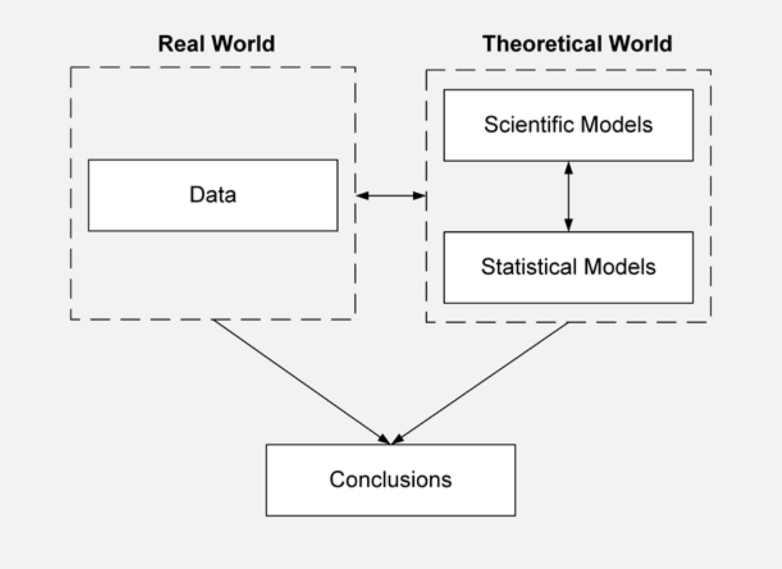


Figure 1. The big picture of statistical inference (based on Figure 1 in Kass, 2011)

1. The Big Picture

Kass uses the Big Picture to make a plea for statistical pragmatism by circumventing the - sometimes difficult and fruitless – discussion about the methods for statistical inference, roughly frequentist vs. Bayesian. Statistical pragmatism is primarily concerned with the assumptions that connect data and models. This approach fits very well with statistical practice, is great for teaching, and highlights the essence of “applied statistics”, finding good, i.e. valid, justifiable, answers to real questions. I view statistical pragmatism as a great, if not ‘the best’ way to achieve statistical thinking. What exactly is statistical thinking? It is a complex concept, that starts with statistical literacy, promoted from an early age in the Anglosaxon countries (see, e.g. Watson, 2000). It may help to develop statistical work ethics and stimulate ethical data science (Utts, 2021).

For now, a simple definition based on Brown & Kass (2009) suffices: understanding the probabilistic nature of statistical models and of data and - thus - their connection. To put it easier: to include ‘error’ in statistical models – because they are never exact – and to separate regularity from variability in data. Understanding the role of assumptions in the statistical models and in data will foster statistical pragmatism, not holding on to absolute rules, not expecting to make final decisions, but being able to make a reasonable assessment of the theoretical model based on the statistical model and the data.

Two important notions to bring into the theoretical and real world are context and control.

Context in the theoretical world is needed because theoretical and statistical models are not universal, not completely specified and therefore not without error. Context in the real world is

needed because the observed data are not ideal, i.e., will not meet the assumptions made in the theoretical world.
Control is a difficult word with various meanings in both worlds. In the Big Picture we can view it as a way to deal with the incompleteness of the models in the theoretical world. In the real, empirical world, we get control of the data by collecting or selecting observations, or by correcting for non-ideal or incomplete observations

I perceive context and control as ways to give substance to statistical thinking, acknowledging error and variability, and deciding how to deal with this.

2. Models

The first topic deals with the theoretical world, in particular the link between theoretical models and statistical models, building models in context.
The arrow between scientific models and statistical models reflects the translation of scientific hypotheses into so-called statistically testable hypotheses. In line with statistical pragmatism and by circumventing the debate about statistical hypothesis testing, I prefer to interpret the arrow as defining a statistical model specification that adequately captures the complete theoretical model including all hypotheses.

A well-known way is to draw a picture - a bit similar to the big picture presented before - to represent a scientific idea, model, or hypotheses. A simple example is represented in Figure 2.

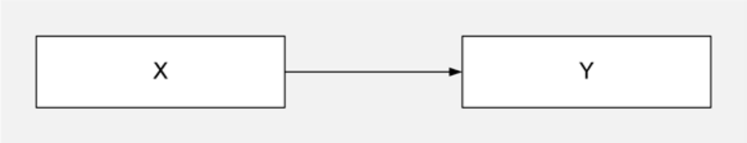


Figure 2. Simple model

Here, X and Y are concepts of interest. For example, X is education and Y is occupational status. In the theoretical model, the arrow from X to Y represents an explanation of the relation between X and Y, or from X to Y, where some form of causality is usually implied. For example, education is positively related to occupational status "because" having a higher education facilitates a higher occupational status. In sociology this is often called the theoretical "mechanism".

The graphical representation is immediately recognizable as a statistical model, often used in regression analysis. The arrow represents the association between X and Y, indicating that for different values of X, different values of Y are expected. In the example, the higher education, the higher occupational status. In teaching we always say that this 'statistical' relation is not causal 'by definition', although we often call X an explanatory variable or a predictor.

Often, the theoretical model will contain multiple explanatory variables, whose association with the outcome of interest need to be investigated. These explanatory variables may also be associated with each other theoretically. An example is given in Figure 3, where two explanatory variables are added to the model, parental education and gender. It shows that parents' education is linked to their child's education and occupational status. Gender is linked to occupational status, reflecting that occupational status is not the same for the genders. The theoretical model also postulates that occupational status depends on gender because of gender differences in the effect of parents' education on education. (I note that this is an example for illustration without a sound theoretical basis.)

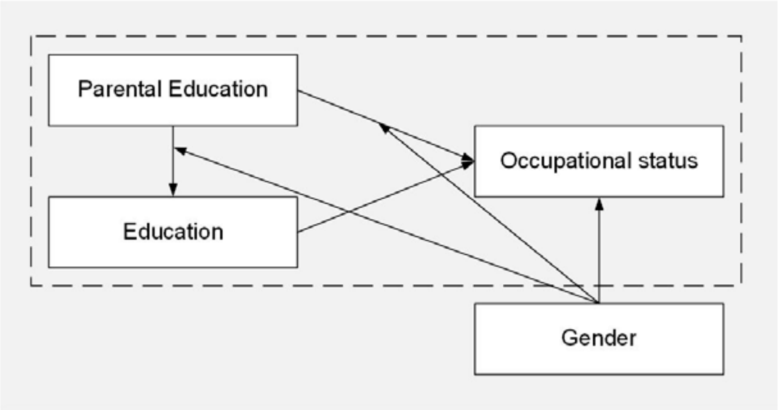


Figure 3. Model with core variables (in dashed area) and a contextual variable (gender)

The link between scientific models and statistical models is relatively straightforward when some form of regression or linear model analysis is used. In the social sciences it is usually assumed that the outcome variable vector Y contains a set of independent identically distributed observations, with normally distributed residuals.

A naïve approach for investigating a theoretical model with multiple explanatory variables, is to estimate the various arrows, that is, the association between pairs of connected variables, ignoring the other variables in the model. A better strategy is to build the complete model in a series of models in growing complexity. Then, the partial association of each X with Y is investigated, that is, conditional on the (partial) associations between the other explanatory variables. One could also say that the 'net' association of X is investigated, that is, association not accounted for by the other variables in the model. It also implies that the model parameters are interpreted under the "ceteris paribus" rule, that is, "all else being equal". It means that all other variables in the model are held constant. This strategy works well if the model is complete, i.e., correctly specified, and thus accounts for the effect of "all" variables of interest.

The order of inclusion of the variables depends on the role of the variable in the model. If it presents a 'known' effect, as for instance the gender effect on occupational status, then it is wise to include the variable in the model from the beginning.

This will alter the partial correlations and provide the relevant interpretation, “ceteris paribus”.
I prefer to call gender a contextual or background variable, and to represent its role by distinguishing a ‘core’ model with the main variables of interest and a periphery containing the contextual variables as depicted in Figure 3. I note that contextual variables are often called ‘control’ variables.

Including “contextual” variables in the model is aimed at obtaining a better estimate of the association between the explanatory variables X and the outcome variable Y, making the “ceteris paribus” assumption concrete. Of course, it does not necessarily make the model correct or complete, but perhaps more complete.

You may wonder why I am concerned about contextual variables. The reason is that we all know that theoretical and statistical models can only approximate reality. One cannot expect theoretical models to represent all factors related to some phenomenon. The researcher will not claim that the theoretical model does so. There will always be “unexplained” differences between human subjects, due to unknown characteristics. This is reflected in the statistical model in the residual term. To find out more about the role of contextual variables, we need to enter the real world.

3. Data

The real world is the domain of the data. Real – empirical – data are collected through experiments, survey research, observations, etc. In my basic understanding of realist ontology and epistemology (based on Maxwell & Delaney, Ch. 1) it works like this. Researchers in the natural sciences may argue that their data are precise except for measurement error and perhaps this also holds for the exactness of their theories and models. Social scientists are more aware of the incompleteness of models as a simplified version of reality and of the imperfection of measurement. This imperfection comes from the difficulties in measuring 'latent' concepts such as occupational status or intelligence, or other variables that we represent by X's and Y's, and the gap between the "human subjects" – people – who participated in the research and the target population of the scientific model. For instance the general population, or a subgroup like women aged over 50. Therefore, social scientists are cautious about making claims about causality.

Ideally, the researcher has complete control over the selection of the subjects whose data are collected. We teach our students that through a – large enough – random sample of some (target) population we can estimate the population statistics. In practice however, the data that are collected in the social sciences do not meet these requirements. The samples are (too) small, not random because of a convenience sample, or not random because of non-response during data collection. Note that I on purpose do not use the term ‘representative’ sample, because it is not immediately clear of what population the sample should be representative, and regarding which characteristics.

The consequences of this non-ideal sample bother me as a statistician because it threatens the performance of the statistical models, affecting the relation with the theoretical models. Therefore, possibly affecting the validity of the statistical and research results as well. It is – and should be - also of concern to the researchers who need to consider the impact on the results and limitations of the conclusions.

The best possible remedy is to carefully inspect the data. In most cases more characteristics of the subjects will be known than included in the theoretical model. This typically concerns demographic information describing the sample, such as gender, age, household situation, geographic region, etc. Statistical insight into the context of the human subjects can be used to discover to what extent the available data are “imbalanced”, or – okay – violate the representativeness of the sample.

Careful data inspection, or exploratory data analysis (Tukey, 1977; see also Tukey, 1962 and Mallows, 2006), provides descriptive statistics such as means, variances and other numerical or graphical distributional information, incomplete observations (missingness) and bivariate correlations.

Descriptive statistics are important ingredients for linking the real world to the theoretical world. They provide the information to identify contextual variables that need to be included in the statistical model and analysis. This is a different way of defining control variables, empirically motivated, and most likely not of primary theoretical interest. Such variables can be viewed as a means to get grip on or "control for" the uncertainty about the influence of the non-ideal or unbalanced data set on the results of the data analysis.

The estimates of the "contributions" (net associations) of the contextual variables are not just nice byproducts of the data analysis. They are essential, because they may lead to a better understanding of the particularities of the "mechanism" under investigation. Perhaps they also lead to a finetuning of the scientific model. In the example I presented, gender may move from context to core, because of new theoretical ideas about the associations between parental education and the gender of the child.

Thus, the real world will identify 'regularity' and 'variability', which the statistical models need to capture adequately in terms of a systematic part and a random part. Exploratory data analysis may convince the statistician and researcher that the statistical analysis they had in mind – usually a ‘normal’ linear model - is appropriate. It may also prepare them for having to use a different model, for instance because of skewness or dependence in the data. This again shows the importance of statistical thinking and statistical pragmatism.

One particular topic that I want to discuss in the area of data requiring special models, is statistical models for social network data, and show that the notion of context is also important for social network analysis. After that I will finish the story about the Big Picture.

4. Statistical models for social network analysis

Just to be sure: social networks are more than social media like Facebook or WhatsApp. Social networks or social network data in the social sciences refer to the relationships (“ties”) between one individual (the “ego”) and a set of other people (the “alters”). So-called complete social network data contain all relationships from and to the members of a group (the “actors”), for instance children in a classroom, or employees in a department.

Statistical modelling of complete social network data is not straightforward. First of all because the observations do not satisfy the usual assumption of independent observations. While in most research data are collected from each individual separately (and thus independently), in social network data the relationships from and to the same individual resemble each other more than relationships involving different individuals. The dependence structure between the ties may well be expected to go beyond pairs of individuals – dyads. For instance, “transitivity”, a friend of my friend is my friend, involves three individuals. This makes the model more complex, with basic ‘structural’ parameters taking care of the dependence and with parameters to represent individual or dyadic characteristics.

The model components and model building steps that I outlined earlier are more difficult for models for social network data. This means that it is also more difficult to translate theoretical models into statistical models. Understanding the behavior of the models, statistically and computationally, in terms of model specification and parameter interpretation is not easy, and a challenging task for both statisticians and social network researchers.

Statistical models for the analysis of social network data have seen a great development, with important contributions by Tom Snijders and his group in the nineties of the last century in Groningen, in close connection with colleagues in Australia and the USA. In the 21st century, this network of statisticians, mathematicians and sociologists has grown in members and locations and is a close-knit community, collaborating on advancing social network analysis. I would like to sketch two topics that are worth studying involving context in a rather different meaning than earlier.

First, context as created by a social network itself. In the early years the work on social network methodology was aimed at estimating a model for just one network, or how it changed over time. Nowadays, it is possible to analyze a 'sample', that is a larger set, of social networks, such as school classrooms, using the same model specification. Through some meta-analytic procedure, or a multilevel analysis, the average associations or mechanisms are estimated, assuming a population of social networks from which the observed networks were drawn.

The assumption of identically distributed social networks is statistically straightforward, but imposes restrictions on the theoretical model, such as same network size, and 'same' (type of) actors. As we saw in the Big Picture, these assumptions need to be investigated in and supported by the real world. How do we decide whether the observed social networks are similar enough to be estimated by the same model? Or, is it possible to account for the differences in the observed data by including contextual variables?

To answer these questions, we first need to understand the behavior of the statistical models for social network data well, in different contexts and conditions. Just like this was done when statisticians first developed the ‘normal’ regression or ANOVA models and evaluated the importance of the distributional assumptions. The straightforward way to study the behavior of statistical models is through simulation studies.

Simulation is a great tool in statistics, because it provides the possibility to use a ‘true’ model, i.e., a model with known parameters, and to generate data from this model. By estimating the parameters from the simulated data, we can investigate whether an estimation method works well, in terms of bias and variance. Simply put, whether the parameters are estimated sufficiently precise. Moreover, the consequences of analyzing data that do not meet the assumptions of the true model, can be studied using ‘corrupted’ data. That is, data generated with a known violation of the assumptions of the true model. If the consequences are relatively small, i.e. the parameters are estimated reasonably precise, then the estimation method is robust against the violation of the model assumption. If the consequences are large, this needs to be taken into account in evaluating the model results.

There are very good examples of this type of work. A recent example is a study of the consequences of missing observations in networks and proposing solutions (e.g., Krause, Huisman, Steglich & Snijders, 2020). To answer the question about the robustness of social network meta-analysis against non-homogeneous samples, a first study shows that structural parameters are less robust than individual and dyadic parameters (Simons, 2021). More work is needed to find out how to identify networks which are responsible for the violation against homogeneity. These may be considered outliers, not belonging to the sample, or perhaps the non-homogeneity can be accommodated for by including contextual variables.

I conclude that developing knowledge about model behavior is essential to judge the link between the real world and the theoretical world, in order to guarantee good results and conclusions.

Second, the context of the individual or actor in the network is a topic of interest. Literally a completely different perspective. It challenges the assumptions of social network research that actors with the same characteristics behave similarly, have the same preferences, etc. It has always surprised me that so many observed friendship dyads are asymmetric. That is, actor i reports a friendship with actor j, but j does not report a friendship to i. This might be due to a difference between individuals in their perception of the concept of friendship, in content - what friends do or feel - or scope - how many friends one can or wants to have. A more methodological explanation is that the measurement of a friendship tie is not precise enough, usually measured by just one dichotomous question. This is an interesting topic that has received attention in the past (see e.g., Marsden, 1990, 2011; Ferligoj & Hlebec, 1999, Hlebec & Ferligoj, 2002), but has not led to scale development.

We can also consider the question whether actors oversee their social environment in such a way that they cannot only report their own friends but also the friendships between others. That is, are individuals sufficiently aware of the position or embeddedness of the other individuals in the group to guide their choices? Knowledge or awareness of the larger network structure, of the absence or presence of mutual relationships, may help individuals to navigate their social context. This is a form of social cognition, a basic need of individuals, for instance to reduce uncertainty about their own position, or whom to trust or turn to for help in situations involving other individuals of the network. A related question is: do people agree in their perceptions of the friendships or social relations present in the group? These are fundamental questions, recognized by many social network researchers but not easy to answer or translate into a statistical model.

Instead of thinking in terms of relationships between two individuals, one needs to think in relationships between three individuals, where "ego" is the perceiver of the relationship. Krackhardt (1987) introduced the cognitive social structure (CSS) as a three-dimensional representation of the perceived social network data. Existing models often either simplify the data to one or a set of 'normal' two-dimensional networks as Krackhardt did, or make too simple and therefore unrealistic dependence assumptions (e.g., van Duijn, 2011; Swartz, Gill, & Muthukumarana, 2015). Other models are difficult to estimate or interpret (e.g., Koskinen, 2002a; 2002b; Sewell, 2020; Sosa & Rodriguez, 2021). Further development of methodology in this area is needed and feasible. I have ongoing work with students searching for viable extensions of existing models and estimation methods.

5. The conclusion

Finally, we get to the conclusion, of this lecture and of the Big Picture. This conclusion is rooted in both the empirical and theoretical world. It relies on the data and on the statistical model in tandem with the theoretical model. The beauty and at the same time essence of the "Big Picture" is that it does not focus on which statistics, models or methods are used. It does not care about Bayesian or frequentist statistics, p-values, confidence intervals, model complexity, etc. These are all concepts that live in the theoretical world. Instead, it focuses on the link between data and theory, to ascertain that the bridge is strong enough to hold the conclusion.

Does this mean that “anything goes” in statistics? Nobody will be surprised that my answer is “of course not”. And it definitely is not what we teach our students.

Good research goes hand in hand with good statistics, and I would even claim that the best research goes with the best statistics. Looking at the left side of the Big Picture, the real, empirical world, it means that the value of data is acknowledged. Data are never perfect but they are to be respected, and cannot be discarded or ignored just when they do not behave according to the model assumptions or do not give the ‘right result’.

We as researchers are responsible for the validity of the conclusions by building the bridge between both worlds. The left side is strengthened by reporting transparently about the data and the statistical analysis, and whatever went wrong or needed adjusting. A solid right side requires careful and nuanced evaluation of the hypothesized theoretical model, acknowledging the uncertainty due to both models and data. This implies that the conclusions based on the statistical analysis are never sure or absolute. Support by the theory or the theoretical mechanism may help to strengthen the conclusion.

With the statement about researchers’ responsibility, I do not want to open Pandora’s box of questionable research practices. What I aim to say is that statistical thinking is needed to arrive at valid conclusions. Involving statisticians in building the bridge is smart. Statisticians are trained in dealing with uncertainty and have a different perspective on the theoretical model and a different relationship with data. Most importantly, statisticians' professional task in consulting or collaborating with colleagues is to make sure that the 'best possible' answer is given to the research question, thus providing a solid base for the conclusion of quantitative research in the social sciences.

Acknowledgments

I am grateful for the support and opportunities I received during my education and long academic career at the University of Groningen. I feel fortunate to have had my professional training in the Department of Statistics and Measurement at the Faculty of Social Sciences, by and with people whom I regard as my academic family, Tom Snijders, Anne Boomsma, Ivo Molenaar, Wendy Post, Herbert Hoijtink, and Jelle van Lenthe. Over the years several family members moved to welcoming neighborhood of the Department of Sociology, joined by Christian Steglich and Mark Huisman. The family has expanded with the many PhD students we supervised.

I specifically thank Rafael Wittek for being my colleague and supportive 'boss' for more than fifteen years.

I am no less grateful for the support I received from the family I grew up in and the family I raised with Hugo Boonstra.

Special thanks go to Wendy Post, Nynke Niezink and Marchien Boonstra for their support in finalizing the inaugural lecture.

Ik heb gezegd.

References

Blatchley B. (2019). Statistics in Context. New York, NY: Oxford University Press.

Brown, E.N. & Kass, R.E. (2009). What Is Statistics? The American Statistician, 63(2), 105-110, https://doi.org/10.1198/tast.2009.0019

Diggle, P. J. (2015). Statistics: a data science for the 21st century. *Journal of the Royal Statistical Society. Series a (Statistics in Society)*, *178*(4), 792–813.

Ferligoj A., & Hlebec, V. (1999). Evaluation of social network measurement instruments. Social Networks, 21(2), 111–130. https://doi.org/10.1016/S0378-8733(99)00007-6.

Gile, K. J., & Handcock, M. S. (2017). Analysis of networks with missing data with application to the national longitudinal study of adolescent health. Journal of the Royal Statistical Society: Series C (Applied Statistics), 66(3), 501–519. https://doi.org/10.1111/rssc.12184.

Heiberger, R. M., & Holland, B. (2015). Statistical analysis and data display: an intermediate course with examples in R (Second, Ser. Springer texts in statistics). Springer. https://doi.org/10.1007/978-1-4939-2122-5.

Hlebec, V., & Ferligoj, A. (2002). Reliability of social network measurement instruments. Field Methods, 14(3), 288–306. https://doi.org/10.1177/15222X014003003

Krackhardt, D. (1987). Cognitive social structures. Social Networks, 9(2), 109–134. https://doi.org/10.1016/0378-8733(87)90009-8.

Krause, R. W., Huisman, M., Steglich, C., & Snijders, T. (2020). Missing data in cross-sectional networks - an extensive comparison of missing data treatment methods. Social Networks, 62, 99–112. https://doi.org/10.1016/j.socnet.2020.02.004.

Mallows, C. (2006). Tukey's paper after 40 years. Technometrics, 48(3), 319–325.

Marsden, P. V. (1990). Network Data and Measurement. *Annual Review of Sociology*, *16*, 435–463. http://www.jstor.org/stable/2083277.

Marsden, P. (2005). Recent Developments in Network Measurement. In P. Carrington, J. Scott, & S. Wasserman (Eds.), *Models and Methods in Social Network Analysis* (Structural Analysis in the Social Sciences, pp. 8-30). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511811395.002

Maxwell, S. E., & Delaney, H. D. (2004). Designing experiments and analyzing data: a model comparison perspective (2nd ed.). Erlbaum.

Panter, A. T., & Sterba, S. K. (2011). *Handbook of ethics in quantitative methodology* (Ser. Multivariate applications series). Routledge.

Simons, J.G. (2021). On the analysis of a sample of exponential random graph model estimates. Research Master's thesis, University of Utrecht.

Sewell, D. (2019). Latent space models for network perception data. *Network Science, 7*(2), 160-179. doi:10.1017/nws.2019.1.

Sosa, J., & Rodríguez, A. (2021). A latent space model for cognitive social structures data. Social Networks, 65, 85-97, https://doi.org/10.1016/j.socnet.2020.12.002.

Swartz, T. B, Gill, P. S, & Muthukumarana, S. (2015). A Bayesian approach for the analysis of triadic data in cognitive social structures. Journal of the Royal Statistical Society, Series C, 64(4), 593–610.

Tukey, J. W. (1977). Exploratory data analysis (Ser. Addison-wesley series in behavioral science : quantitative methods). Addison-Wesley.

Tukey, J. W. (1962). The future of data analysis. *The annals of mathematical statistics*, *33*(1), 1-67.

Utts, J. (2021). Enhancing data science ethics through statistical education and practice. International Statistical Review, 89(1), 1–17. https://doi.org/10.1111/insr.12446

van Duijn, M.A.J. (2011). Modeling three-way social network data: A cross-nested random effects model for gossip in the work place. In: Cerchille, P. and Tarnatola, C. (Eds). CLADAG 2011. Book of Abstracts, (pp. 4). Pavia, Italy, Pavia University Press, ISBN 978-88-906639

Watson, J. M. (2000). Statistics in context. *Mathematics Teacher*, *93*(1), 54–58.

Marijtje van Duijn is Aletta Jacobs Professor of Statistics, in particular models for social network analysis at the Department of Sociology and Inter-University Center for Social Science Theory and Methodology (ICS). She obtained degrees at the University of Groningen, an MSc. in econometrics (1988) and a PhD in statistics in the social sciences in 1993. She received two Fulbright scholarships, in 1993 as a post-doc at the Department of Quantitative Psychology (University of Illinois at Urbana-Champaign), in 2005-2006 as a senior scholar at the Center for Statistics and the Social Sciences (University of Washington, Seattle).

Her research combines statistical modeling and social network analysis with a strong emphasis on applications including software. Next to collaborative work with social and biomedical scientists, she has developed tailor made statistical models for social network data, and written guidelines and tutorials for applying – new or non-standard – models in practice.

University of Groningen Press