Stata's statistical features empower you to answer a wide range of research questions. From linear and logistic regression to time-series and panel-data analyses, survival models, causal inference, Bayesian analysis, and machine learning, you can fit models, evaluate assumptions, make inferences, and interpret results with confidence. Explore statistical features — Share — copy and redistribute the material in any medium or format for any purpose, even commercially. Adapt — remix, transform, and build upon the material for any purpose, even commercially. The licensor cannot revoke these freedoms as long as you follow the license terms. Attribution — You must give appropriate credit , provide a link to the license, and indicate if changes were made . You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use. ShareAlike — If you remix, transform, or build upon the material, you must distribute your contributions under the same license as the original. No additional restrictions — You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits. You do not have to comply with the license for elements of the material in the public domain or where your use is permitted by an applicable exception or limitation . No warranties are given. The license may not give you all of the permissions necessary for your intended use. For example, other rights such as publicity, privacy, or moral rights may limit how you use the material. The Workflow of Data Analysis Using Stata, by J. Scott Long, is an essential productivity tool for data analysts. Aimed at anyone who analyzes data, this book presents an effective strategy for designing and doing data-analytic projects. In this book, Long presents lessons gained from his experience with numerous academic publications, as a coauthor of the immensely popular Regression Models for Categorical Dependent Variables Using Stata, and as a coauthor of the SPOST routines, which are downloaded over 20,000 times a year. A workflow of data analysis is a process for managing all aspects of data analysis. Planning, documenting, and organizing your work; cleaning the data; creating, renaming, and verifying variables; performing and presenting statistical analyses; producing replicable results; and archiving what you have done are all integral parts of your workflow. Long shows how to design and implement efficient workflows for both one-person projects and team projects. Long guides you toward streamlining your workflow, because a good workflow is essential for replicating your work, and replication is essential for good science. An efficient workflow reduces the time you spend doing data management and lets you produce datasets that are easier to analyze. When you methodically clean your data and carefully choose names and labels for your variables, the time you spend doing statistical and graphical analyses will be more productive and more enjoyable. After introducing workflows and explaining how a better workflow can make it easier to work with data, Long describes planning, organizing, and documenting your work. He then introduces how to write and debug Stata do-files and how to use local and global macros. Long presents conventions that greatly simplify data analysis—conventions for naming, labeling, documenting, and verifying variables. He also covers cleaning, analyzing, and protecting your data. While describing effective workflows, Long also introduces the concepts of basic data management using Stata and writing Stata do-files. Using real-world examples, Stata commands, and Stata scripts, Long illustrates effective techniques for managing your data and analyses. If you analyze data, this book is recommended for you. Comments from readers You have written the book that I had planned to write someday. But I'm glad I didn't—your book is much better. Congratulations, this was greatly needed. Prof. Bill Gardner The Ohio State University I will post the announcement of Workflow on my door with the following note: "I'm glad to help anybody who followed at least 25% of the advice Long provides—and brings me their do-files!" Prof. Alan C. Acock Oregon State University I just wanted to send you a thank you for taking the time to write this book. I feel a little like an obsessed fan because I read it for several hours last night, bought 3 copies for my new research team and am presenting our new organization scheme tomorrow. It turns out that we have just finished a first flurry of data collection and hiring and I've been scratching my head about how to systematize some aspects. It is a perfect time to superimpose a structure. I've used aspects of your plan in my own work (hence my eagerness to adopt) but having this coherent volume is a wonderful and practical resource. I learned a lot from reading this. Thank you! Elizabeth Gifford, Ph.D. Research Scientist Duke University I just received a knock at my door with my new copy of The Workflow of Data Analysis Using Stata. I immediately ripped off the packaging and began perusing it. Just before the knock, I was attempting to write a program to get Stata to save the r(mean) and r(sd) for two variables following a summarize command to be saved for a ttesti command. After looking at your book for about two minutes, I stumbled upon pages 91–92, where it gave me all the information I need. ... I have only had the book about 10 minutes and already it has made my life easier. Thanks much, and I am already looking forward to reading the rest of the book! Claire M. Kamp Dush, Ph.D. The Ohio State University I am a Spanish professor of public economics who is at present enjoying a study-research leave at Melbourne University (Australia). Because of that I have had the time to read your book from cover to cover. I just want to thank you for the incredible work you have done! A book such as this one is a must for anyone trying to make an academic career. Definitely, I will recommend it to my graduate students as soon as I go back to Spain. If I had the chance to reach this book twenty years ago I would have been much more efficient doing my work. Never is it too late! Thanks! Prof. Jose Felix Sanz-Sanz Dept. of Applied Economics Universidad Complutense de Madrid The Workflow of Data Analysis Using Stata This project deals with the principles that guide data analysis and how to implement those principles using Stata. You can order the book from Stata Press. Process of using data analysis Not to be confused with Statistical interference. Part of a series onResearch Research design Ethics Proposal Question Writing Argument Referencing Research strategy Interdisciplinary Multimethodology Qualitative Art-based Quantitative Philosophical schools Antipositivism Constructivism Critical rationalism Empiricism Fallibilism Positivism Postpositivism Pragmatism Realism Critical realism Subtle realism Methodology Action research Art methodology Critical theory Grounded theory Hermeneutics Historiography Human subject research Narrative inquiry Phenomenology Pragmatism Scientific method Methods Analysis Case study Content analysis Descriptive statistics Discourse analysis Ethnography Autoethnography Experiment Field experiment Social experiment Quasi-experiment Field research Historical method Inferential statistics Interviews Mapping Cultural mapping Phenomenography Secondary research Bibliometrics Literature review Meta-analysis Scoping review Systematic review Scientific modelling Simulation Survey Tools and software Argument technology GIS software LIS software Bibliometrics Reference management Science software Qualitative data analysis Simulation Statistics Philosophy portalvte Statistical inference is the process of using data analysis to infer properties of an underlying probability distribution.[1] Inferential statistical analysis infers properties of a population, for example by testing hypotheses and deriving estimates. It is assumed that the observed data set is sampled from a larger population. Inferential statistics can be contrasted with descriptive statistics. Descriptive statistics is solely concerned with properties of the observed data, and it does not rest on the assumption that the data come from a larger population. In machine learning, the term inference is sometimes used instead to mean "make a prediction, by evaluating an already trained model";[2] in this context inferring properties of the model is referred to as training or learning (rather than inference), and using a model for prediction is referred to as inference (instead of prediction); see also predictive inference. Statistical inference makes propositions about a population, using data drawn from the population with some form of sampling. Given a hypothesis about a population, for which we wish to draw inferences, statistical inference consists of (first) selecting a statistical model of the process that generates the data and (second) deducing propositions from the model.[3] Konishi and Kitagawa state "The majority of the problems in statistical inference can be considered to be problems related to statistical modeling".[4] Relatedly, Sir David Cox has said, "How [the] translation from subject-matter problem to statistical model is done is often the most critical part of an analysis".[5] The conclusion of a statistical inference is a statistical proposition.[6] Some common forms of statistical proposition are the following: a point estimate, i.e. a particular value that best approximates some parameter of interest; an interval estimate, e.g. a confidence interval (or set estimate), i.e. an interval constructed using a dataset drawn from a population so that, under repeated sampling of such datasets, such intervals would contain the true parameter value with the probability at the stated confidence level; a credible interval, i.e. a set of values containing, for example, 95% of posterior belief; rejection of a hypothesis;[note 1] clustering or classification of data points into groups. Main articles: Statistical model and Statistical assumptions Any statistical inference requires some assumptions. A statistical model is a set of assumptions concerning the generation of the observed data and similar data. Descriptions of statistical models usually emphasize the role of population quantities of interest, about which we wish to draw inference.[7] Descriptive statistics are typically used as a preliminary step before more formal inferences are drawn.[8] Statisticians distinguish between three levels of modeling assumptions: Fully parametric: The probability distributions describing the data-generation process are assumed to be fully described by a family of probability distributions involving only a finite number of unknown parameters.[7] For example, one may assume that the distribution of population values is truly Normal, with unknown mean and variance, and that datasets are generated by 'simple' random sampling. The family of generalized linear models is a widely used and flexible class of parametric models. Non-parametric: The assumptions made about the process generating the data are much less than in parametric statistics and may be minimal.[9] For example, every continuous probability distribution has a median, which may be estimated using the sample median or the Hodges–Lehmann–Sen estimator, which has good properties when the data arise from simple random sampling. Semi-parametric: This term typically implies assumptions 'in between' fully and non-parametric approaches. For example, one may assume that a population distribution has a finite mean. Furthermore, one may assume that the mean response level in the population depends in a truly linear manner on some covariate (a parametric assumption) but not make any parametric assumption describing the variance around that mean (i.e. about the presence or possible form of any heteroscedasticity). More generally, semi-parametric models can often be separated into 'structural' and 'random variation' components. One component is treated parametrically and the other non-parametrically. The well-known Cox model is a set of semi-parametric assumptions.[citation needed] See also: Statistical model validation The above image shows a histogram assessing the assumption of normality, which can be illustrated through the even spread underneath the bell curve. Whatever level of assumption is made, correctly calibrated inference, in general, requires these assumptions to be correct; i.e. that the data-generating mechanisms really have been correctly specified. Incorrect assumptions of 'simple' random sampling can invalidate statistical inference.[10] More complex semi- and fully parametric assumptions are also cause for concern. For example, incorrectly assuming the Cox model can in some cases lead to faulty conclusions.[11] Incorrect assumptions of Normality in the population also invalidates some forms of regression-based inference.[12] The use of any parametric model is viewed skeptically by most experts in sampling human populations: "most sampling statisticians, when they deal with confidence intervals at all, limit themselves to statements about [estimators] based on very large samples, where the central limit theorem ensures that these [estimators] will have distributions that are nearly normal."[13] In particular, a normal distribution "would be a totally unrealistic and catastrophically unwise assumption to make if we were dealing with any kind of economic population."[13] Here, the central limit theorem states that the distribution of the sample mean "for very large samples" is approximately normally distributed, if the distribution is not heavy-tailed. Main articles: Statistical distance, Asymptotic theory (statistics), and Approximation theory Given the difficulty in specifying exact distributions of sample statistics, many methods have been developed for approximating these. With finite samples, approximation results measure how close a limiting distribution approaches the statistic's sample distribution: For example, with 10,000 independent samples the normal distribution approximates (to two digits of accuracy) the distribution of the sample mean for many population distributions, by the Berry–Esseen theorem.[14] Yet for many practical purposes, the normal approximation provides a good approximation to the sample-mean's distribution when there are 10 (or more) independent samples, according to simulation studies and statisticians' experience.[14] Following Kolmogorov's work in the 1950s, advanced statistics uses approximation theory and functional analysis to quantify the error of approximation. In this approach, the metric geometry of probability distributions is studied; this approach quantifies approximation error with, for example, the Kullback–Leibler divergence, Bregman divergence, and the Hellinger distance.[15][16][17] With indefinitely large samples, limiting results like the central limit theorem describe the sample statistic's limiting distribution if one exists. Limiting results are not statements about finite samples, and indeed are irrelevant to finite samples.[18][19][20] However, the asymptotic theory of limiting distributions is often invoked for work with finite samples. For example, limiting results are often invoked to justify the generalized method of moments and the use of generalized estimating equations, which are popular in econometrics and biostatistics. The magnitude of the difference between the limiting distribution and the true distribution (formally, the 'error' of the approximation) can be assessed using simulation.[21] The heuristic application of limiting results to finite samples is common practice in many applications, especially with low-dimensional models with log-concave likelihoods (such as with one-parameter exponential families). Main article: Randomization See also: Random sample and Random assignment For a given dataset that was produced by a randomization design, the randomization distribution of a statistic (under the null-hypothesis) is defined by evaluating the test statistic for all of the plans that could have been generated by the randomization design. In frequentist inference, the randomization allows inferences to be based on the randomization distribution rather than a subjective model, and this is important especially in survey sampling and design of experiments.[22][23] Statistical inference from randomized studies is also more straightforward than many other situations.[24][25][26] In Bayesian inference, randomization is also of importance: in survey sampling, use of sampling without replacement ensures the exchangeability of the sample with the population; in randomized experiments, randomization warrants a missing at random assumption for covariate information.[27] Objective randomization allows properly inductive procedures.[28][29][30][31][32] Many statisticians prefer randomization-based analysis of data that was generated by well-defined randomization procedures.[33] (However, it is true that in fields of science with developed theoretical knowledge and experimental control, randomized experiments may increase the costs of experimentation without improving the quality of inferences.[34][35]) Similarly, results from randomized experiments are recommended by leading statistical authorities as allowing inferences with greater reliability than do observational studies of the same phenomena.[36] However, a good observational study may be better than a bad randomized experiment. The statistical analysis of a randomized experiment may be based on the randomization scheme stated in the experimental protocol and does not need a subjective model.[37][38] However, at any time, some hypotheses cannot be tested using objective statistical models, which accurately describe randomized experiments or random samples. In some cases, such randomized studies are uneconomical or unethical. It is standard practice to refer to a statistical model, e.g., a linear or logistic models, when analyzing data from randomized experiments.[39] However, the randomization scheme guides the choice of a statistical model. It is not possible to choose an appropriate model without knowing the randomization scheme.[23] Seriously misleading results can be obtained analyzing data from randomized experiments while ignoring the experimental protocol; common mistakes include forgetting the blocking used in an experiment and confusing repeated measurements on the same experimental unit with independent replicates of the treatment applied to different experimental units.[40] Model-free techniques provide a complement to model-based methods, which employ reductionist strategies of reality-simplification. The former combine, evolve, ensemble and train algorithms dynamically adapting to the contextual affinities of a process and learning the intrinsic characteristics of the observations.[41][42] For example, model-free simple linear regression is based either on: a random design, where the pairs of observations (X 1 , Y 1) , (X 2 , Y 2) , ⋯ , (X n , Y n) {\displaystyle (X_{1},Y_{1}),(X_{2},Y_{2}),\cdots ,(X_{n},Y_{n})} are independent and identically distributed (iid), or a deterministic design, where the variables X 1 , X 2 , ⋯ , X n {\displaystyle X_{1},X_{2},\cdots ,X_{n}} are deterministic, but the corresponding response variables Y 1 , Y 2 , ⋯ , Y n {\displaystyle Y_{1},Y_{2},\cdots ,Y_{n}} are random and independent with a common conditional distribution, i.e., P (Y j ≤ y | X j = x) = D x (y) {\displaystyle P\left(Y_{j}\leq y|X_{j}=x\right)=D_{x}(y)} , which is independent of the index j {\displaystyle j} . In either case, the model-free randomization inference for features of the common conditional distribution D x (.) {\displaystyle D_{x}(.)} relies on some regularity conditions, e.g. functional smoothness. For instance, model-free randomization inference for the population feature conditional mean, μ (x) = E (Y | X = x) {\displaystyle \mu (x)=E(Y|X=x)} , can be consistently estimated via local averaging or local polynomial fitting, under the assumption that μ (x) {\displaystyle \mu (x)} is smooth. Also, relying on asymptotic normality or resampling, we can construct confidence intervals for the population feature, in this case, the conditional mean, μ (x) {\displaystyle \mu (x)} .[43] Different schools of statistical inference have become established. These schools—or "paradigms"—are not mutually exclusive, and methods that work well under one paradigm often have attractive interpretations under other paradigms. Bandyopadhyay and Forster describe four paradigms: The classical (or frequentist) paradigm, the Bayesian paradigm, the likelihoodist paradigm, and the Akaikean-Information Criterion-based paradigm.[44] Main article: Frequentist inference This paradigm calibrates the plausibility of propositions by considering (notional) repeated sampling of a population distribution to produce datasets similar to the one at hand. By considering the dataset's characteristics under repeated sampling, the frequentist properties of a statistical proposition can be quantified—although in practice this quantification may be challenging. p-value Confidence interval Null hypothesis significance testing One interpretation of frequentist inference (or classical inference) is that it is applicable only in terms of frequency probability; that is, in terms of repeated sampling from a population. However, the approach of Neyman[45] develops these procedures in terms of pre-experiment probabilities. That is, before undertaking an experiment, one decides on a rule for coming to a conclusion such that the probability of being correct is controlled in a suitable way: such a probability need not have a frequentist or repeated sampling interpretation. In contrast, Bayesian inference works in terms of conditional probabilities (i.e. probabilities conditional on the observed data), compared to the marginal (but conditioned on unknown parameters) probabilities used in the frequentist approach. The frequentist procedures of significance testing and confidence intervals can be constructed without regard to utility functions. However, some elements of frequentist statistics, such as statistical decision theory, do incorporate utility functions.[citation needed] In particular, frequentist developments of optimal inference (such as minimum-variance unbiased estimators, or uniformly most powerful testing) make use of loss functions, which play the role of (negative) utility functions. Loss functions need not be explicitly stated for statistical theorists to prove that a statistical procedure has an optimality property.[46] However, loss-functions are often useful for stating optimality properties: for example, median-unbiased estimators are optimal under absolute value loss functions, in that they minimize expected loss, and least squares estimators are optimal under squared error loss functions, in that they minimize expected loss. While statisticians using frequentist inference must choose for themselves the parameters of interest, and the estimators/test statistic to be used, the absence of obviously explicit utilities and prior distributions has helped frequentist procedures to become widely viewed as 'objective'.[47] See also: Bayesian inference The Bayesian calculus describes degrees of belief using the 'language' of probability; beliefs are positive, integrate into one, and obey probability axioms. Bayesian inference uses the available posterior beliefs as the basis for making statistical propositions.[48] There are several different justifications for using the Bayesian approach. Credible interval for interval estimation Bayes factors for model comparison Many informal Bayesian inferences are based on "intuitively reasonable" summaries of the posterior. For example, the posterior mean, median and mode, highest posterior density intervals, and Bayes Factors can all be motivated in this way. While a user's utility function need not be stated for this sort of inference, these summaries do all depend (to some extent) on stated prior beliefs, and are generally viewed as subjective conclusions. (Methods of prior construction which do not require external input have been proposed but not yet fully developed.) Formally, Bayesian inference is calibrated with reference to an explicitly stated utility, or loss function; the 'Bayes rule' is the one which maximizes expected utility, averaged over the posterior uncertainty. Formal Bayesian inference therefore automatically provides optimal decisions in a decision theoretic sense. Given assumptions, data and utility, Bayesian inference can be made for essentially any problem, although not every statistical inference need have a Bayesian interpretation. Analyses which are not formally Bayesian can be (logically) incoherent; a feature of Bayesian procedures which use proper priors (i.e. those integrable to one) is that they are guaranteed to be coherent. Some advocates of Bayesian inference assert that inference must take place in this decision-theoretic framework, and that Bayesian inference should not conclude with the evaluation and summarization of posterior beliefs. Main article: LikelihoodismLikelihood-based inference is a paradigm used to estimate the parameters of a statistical model based on observed data. Likelihoodism approaches statistics by using the likelihood function, denoted as L (x | θ) {\displaystyle L(x|\theta)} , quantifies the probability of observing the given data x {\displaystyle x} , assuming a specific set of parameter values θ {\displaystyle \theta } . In likelihood-based inference, the goal is to find the set of parameter values that maximizes the likelihood function, or equivalently, maximizes the probability of observing the given data. The process of likelihood-based inference usually involves the following steps: Formulating the statistical model: A statistical model is defined based on the problem at hand, specifying the distributional assumptions and the relationship between the observed data and the unknown parameters. The model can be simple, such as a normal distribution with known variance, or complex, such as a hierarchical model with multiple levels of random effects. Constructing the likelihood function: Given the statistical model, the likelihood function is constructed by evaluating the joint probability density or mass function of the observed data as a function of the unknown parameters. This function represents the probability of observing the data for different values of the parameters. Maximizing the likelihood function: The next step is to find the set of parameter values that maximizes the likelihood function. This can be achieved using optimization techniques such as numerical optimization algorithms. The estimated parameter values, often denoted as y ¯ {\displaystyle {\bar {y}}} , are the maximum likelihood estimates (MLEs). Assessing uncertainty: Once the MLEs are obtained, it is crucial to quantify the uncertainty associated with the parameter estimates. This can be done by calculating standard errors, confidence intervals, or conducting hypothesis tests based on asymptotic theory or simulation techniques such as bootstrapping. Model checking: After obtaining the parameter estimates and assessing their uncertainty, it is important to assess the adequacy of the statistical model. This involves checking the assumptions made in the model and evaluating the fit of the model to the data using goodness-of-fit tests, residual analysis, or graphical diagnostics. Inference and interpretation: Finally, based on the estimated parameters and model assessment, statistical inference can be performed. This involves drawing conclusions about the population parameters, making predictions, or testing hypotheses based on the estimated model. Main article: Akaike information criterion This section needs expansion. You can help by adding to it. (November 2017) The Akaike information criterion (AIC) is an estimator of the relative quality of statistical models for a given set of data. Given a collection of models for the data, AIC estimates the quality of each model, relative to each of the other models. Thus, AIC provides a means for model selection. AIC is founded on information theory: it offers an estimate of the relative information lost when a given model is used to represent the process that generated the data. (In doing so, it deals with the trade-off between the goodness of fit of the model and the simplicity of the model.) Main article: Minimum description length The minimum description length (MDL) principle has been developed from ideas in information theory[49] and the theory of Kolmogorov complexity.[50] The (MDL) principle selects statistical models that maximally compress the data; inference proceeds without assuming counterfactual or non-falsifiable "data-generating mechanisms" or probability models for the data, as might be done in frequentist or Bayesian approaches. However, if a "data generating mechanism" does exist in reality, then according to Shannon's source coding theorem it provides the MDL description of the data, on average and asymptotically.[51] In minimizing description length (or descriptive complexity), MDL estimation is similar to maximum likelihood estimation and maximum a posteriori estimation (using maximum-entropy Bayesian priors). However, MDL avoids assuming that the underlying probability model is known; the MDL principle can also be applied without assumptions that e.g. the data arose from independent sampling.[51][52] The MDL principle has been applied in communication-coding theory in information theory, in linear regression,[52] and in data mining.[50] The evaluation of MDL-based inferential procedures often uses techniques or criteria from computational complexity theory.[53] Main article: Fiducial inference Fiducial inference was an approach to statistical inference based on fiducial probability, also known as a "fiducial distribution". In subsequent work, this approach has been called ill-defined, extremely limited in applicability, and even fallacious.[54][55] However this argument is the same as that which shows[56] that a so-called confidence distribution is not a valid probability distribution and, since this has not invalidated the application of confidence intervals, it does not necessarily invalidate conclusions drawn from fiducial arguments. An attempt was made to reinterpret the early work of Fisher's fiducial argument as a special case of an inference theory using upper and lower probabilities.[57] Developing ideas of Fisher and of Pitman from 1938 to 1939,[58] George A. Barnard developed "structural inference" or "pivotal inference",[59] an approach using invariant probabilities on group families. Barnard reformulated the arguments behind fiducial inference on a restricted class of models on which "fiducial" procedures would be well-defined and useful. Donald A. S. Fraser developed a general theory for structural inference[60] based on group theory and applied this to linear models.[61] The theory formulated by Fraser has close links to decision theory and Bayesian statistics and can provide optimal frequentist decision rules if they exist.[62] The topics below are usually included in the area of statistical inference. Statistical assumptions Statistical decision theory Estimation theory Statistical hypothesis testing Revising opinions in statistics Design of experiments, the analysis of variance, and regression Survey sampling Summarizing statistical data Predictive inference is an approach to statistical inference that emphasizes the prediction of future observations based on past observations. Initially, predictive inference was based on observable parameters and it was the main purpose of studying probability,[citation needed] but it fell out of favor in the 20th century due to a new parametric approach pioneered by Bruno de Finetti. The approach modeled phenomena as a physical system observed with error (e.g., celestial mechanics). De Finetti's idea of exchangeability—that future observations should behave like past observations—came to the attention of the English-speaking world with the 1974 translation from French of his 1937 paper,[63] and has since been propounded by such statisticians as Seymour Geisser.[64] Algorithmic inference Induction (philosophy) Informal inferential reasoning Information field theory Population proportion Philosophy of statistics Prediction interval Predictive analytics Predictive modelling Stylometry ^ According to Peirce, acceptance means that inquiry on this question ceases for the time being. In science, all scientific theories are revisable. ^ Upton, G., Cook, I. (2008) Oxford Dictionary of Statistics, OUP. ISBN 978-0-19-954145-4. ^ "TensorFlow Lite inference". The term inference refers to the process of executing a TensorFlow Lite model on-device in order to make predictions based on input data. ^ Johnson, Richard (12 March 2016). "Statistical Inference". Encyclopedia of Mathematics. Springer: The European Mathematical Society. Retrieved 26 October 2022. ^ Konishi & Kitagawa (2008), p. 75. ^ Cox (2006), p. 197. ^ "Statistical inference - Encyclopedia of Mathematics". www.encyclopediaofmath.org. Retrieved 2019-01-23. ^ a b Cox (2006) page 2 ^ Evans, Michael; et al. (2004). Probability and Statistics: The Science of Uncertainty. Freeman and Company. p. 267. ISBN 9780716747420. ^ van der Vaart, A.W. (1998) Asymptotic Statistics Cambridge University Press. ISBN 0-521-78450-6 (page 341) ^ Kruskal 1988 ^ Freedman, D.A. (2008) "Survival analysis: An Epidemiological hazard?". The American Statistician (2008) 62: 110-119. (Reprinted as Chapter 11 (pages 169–192) of Freedman (2010)). ^ Berk, R. (2003) Regression Analysis: A Constructive Critique (Advanced Quantitative Techniques in the Social Sciences) (v. 11) Sage Publications. ISBN 0-7619-2904-5 ^ a b Brewer, Ken (2002). Combined Survey Sampling Inference: Weighing of Basu's Elephants. Hodder Arnold. p. 6. ISBN 978-0340692295. ^ a b Jörgen Hoffman-Jörgensen's Probability With a View Towards Statistics, Volume I. Page 399 [full citation needed] ^ Le Cam (1986) [page needed] ^ Erik Torgerson (1991) Comparison of Statistical Experiments, volume 36 of Encyclopedia of Mathematics. Cambridge University Press. [full citation needed] ^ Liese, Friedrich & Miescke, Klaus-J. (2008). Statistical Decision Theory: Estimation, Testing, and Selection. Springer. ISBN 978-0-387-73193-3. ^ Kolmogorov (1963, p.369): "The frequency concept, based on the notion of limiting frequency as the number of trials increases to infinity, does not contribute anything to substantiate the applicability of the results of probability theory to real practical problems where we have always to deal with a finite number of trials". ^ "Indeed, limit theorems 'as n {\displaystyle n} tends to infinity' are logically devoid of content about what happens at any particular n {\displaystyle n} . All they can do is suggest certain approaches whose performance must then be checked on the case at hand." — Le Cam (1986) (page xiv) ^ Pfanzagl (1994): "The crucial drawback of asymptotic theory: What we expect from asymptotic theory are results which hold approximately . . . . What asymptotic theory has to offer are limit theorems."(page ix) "What counts for applications are approximations, not limits." (page 188) ^ Pfanzagl (1994) : "By taking a limit theorem as being approximately true for large sample sizes, we commit an error the size of which is unknown. [. . .] Realistic information about the remaining errors may be obtained by simulations." (page ix) ^ Neyman, J.(1934) "On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection", Journal of the Royal Statistical Society, 97 (4), 557–625 JSTOR 2342192 ^ a b Hinkelmann and Kempthorne(2008) [page needed] ^ ASA Guidelines for the first course in statistics for non-statisticians. (available at the ASA website) ^ David A. Freedman et alia's Statistics. ^ Moore et al. (2015). ^ Gelman A. et al. (2013). Bayesian Data Analysis (Chapman & Hall). ^ Peirce (1877-1878) ^ Peirce (1883) ^ Freedman, Pisani & Purves 1978. ^ David A. Freedman Statistical Models. ^ Rao, C.R. (1997) Statistics and Truth: Putting Chance to Work, World Scientific. ISBN 981-02-3111-3 ^ Peirce; Freedman; Moore et al. (2015).[citation needed] ^ Box, G.E.P. and Friends (2006) Improving Almost Anything: Ideas and Essays, Revised Edition, Wiley. ISBN 978-0-471-72755-2 ^ Cox (2006), p. 196. ^ ASA Guidelines for the first course in statistics for non-statisticians. (available at the ASA website) David A. Freedman et alias Statistics. Moore et al. (2015). ^ Neyman, Jerzy. 1923 [1990]. "On the Application of Probability Theory to AgriculturalExperiments. Essay on Principles. Section 9." Statistical Science 5 (4): 465–472. Trans. Dorota M. Dabrowska and Terence P. Speed. ^ Hinkelmann & Kempthorne (2008) [page needed] ^ Dinov, Ivo; Palanimalai, Selvam; Khare, Ashwini; Christou, Nicolas (2018). "Randomization-based statistical inference: A resampling and simulation infrastructure". Teaching Statistics. 40 (2): 64–73. doi:10.1111/test.12156. PMC 6155997. PMID 30270947. ^ Hinkelmann and Kempthorne (2008) Chapter 6. ^ Dinov, Ivo; Palanimalai, Selvam; Khare, Ashwini; Christou, Nicolas (2018). "Randomization-based statistical inference: A resampling and simulation infrastructure". Teaching Statistics. 40 (2): 64–73. doi:10.1111/test.12156. PMC 6155997. PMID 30270947. ^ Tang, Ming; Gao, Chao; Goutman, Stephen; Kalinin, Alexandr; Mukherjee, Bhramar; Guan, Yuanfang; Dinov, Ivo (2019). "Model-Based and Model-Free Techniques for Amyotrophic Lateral Sclerosis Diagnostic Prediction and Patient Clustering". Neuroinformatics. 17 (3): 407–421. doi:10.1007/s12021-018-9406-9. PMC 6527505. PMID 30460455. ^ Politis, D.N. (2019). "Model-free inference in statistics: how and why". IMS Bulletin. 48. ^ Bandyopadhyay & Forster (2011). See the book's Introduction (p.3) and "Section III: Four Paradigms of Statistics". ^ Neyman, J. (1937). "Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability". Philosophical Transactions of the Royal Society of London A. 236 (767): 333–380. Bibcode:1937RSPTA.236..333N. doi:10.1098/rsta.1937.0005. JSTOR 91337. ^ Preface to Pfanzagl. ^ Little, Roderick J. (2006). "Calibrated Bayes: A Bayes/Frequentist Roadmap". The American Statistician. 60 (3): 213–223. doi:10.1198/000313006X117837. ISSN 0003-1305. JSTOR 27643780. S2CID 53505632. ^ Lee, Se Yoon (2021). "Gibbs sampler and coordinate ascent variational inference: A set-theoretical review". Communications in Statistics - Theory and Methods. 51 (6): 1549–1568. arXiv:2008.01006. doi:10.1080/03610926.2021.1921214. S2CID 220935477. ^ Soofi (2000) ^ a b Hansen & Yu (2001) ^ a b Hansen and Yu (2001), page 747. ^ a b Rissanen (1989), page 84 ^ Joseph F. Traub, G. W. Wasilkowski, and H. Wozniakowski. (1988) [page needed] ^ Neyman (1956) ^ Zabell (1992) ^ Cox (2006) page 66 ^ Hampel 2003. ^ Davison, page 12. [full citation needed] ^ Barnard, G.A. (1995) "Pivotal Models and the Fiducial Argument", International Statistical Review, 63 (3), 309–323. JSTOR 1403482 ^ Fraser, D. A. S. (1968). The structure of inference. New York: Wiley. ISBN 0-471-27548-4. OCLC 440926. ^ Fraser, D. A. S. (1979). Inference and linear models. London: McGraw-Hill. ISBN 0-07-021910-9. OCLC 3559629. ^ Taraldsen, Gunnar; Lindqvist, Bo Henry (2013-02-01). "Fiducial theory and optimal inference". The Annals of Statistics. 41 (1). arXiv:1301.1717. doi:10.1214/13-AOS1083. ISSN 0090-5364. S2CID 88520957. ^ De Finetti, Bruno (1937). "La Prévision: ses lois logiques, ses sources subjectives". Annales de l'Institut Henri Poincaré. 7 (1): 1–68. ISSN 0365-320X. Translated in De Finetti, Bruno (1992). "Foresight: Its Logical Laws, Its Subjective Sources". Breakthroughs in Statistics. Springer Series in Statistics. pp. 134–174. doi:10.1007/978-1-4612-0919-5_10. ISBN 978-0-387-94037-3. ^ Geisser, Seymour (1993) Predictive Inference: An Introduction, CRC Press. ISBN 0-412-03471-9 Bandyopadhyay, P. S.; Forster, M. R., eds. (2011), Philosophy of Statistics, Elsevier. Bickel, Peter J.; Doksum, Kjell A. (2001). Mathematical statistics: Basic and selected topics. Vol. 1 (Second (updated printing 2007) ed.). Prentice Hall. ISBN 978-0-13-850363-5. MR 0443141. Cox, D. R. (2006). Principles of Statistical Inference, Cambridge University Press. ISBN 0-521-68567-2. Fisher, R. A. (1955), "Statistical methods and scientific induction", Journal of the Royal Statistical Society, Series B, 17, 69–78. (criticism of statistical theories of Jerzy Neyman and Abraham Wald) Freedman, D. A. (2009). Statistical Models: Theory and practice (revised ed.). Cambridge University Press. pp. xiv+442 pp. ISBN 978-0-521-74385-3. MR 2489600. Freedman, D. A. (2010). Statistical Models and Causal Inferences: A Dialogue with the Social Sciences (Edited by David Collier, Jasjeet Sekhon, and Philip B. Stark), Cambridge University Press. Hampel, Frank R. (February 2003). "The proper fiducial argument". Seminar für Statistik, Eidgenössische Technische Hochschule. 114. doi:10.3929/ethz-a-004526011. Hansen, Mark H.; Yu, Bin (June 2001). "Model Selection and the Principle of Minimum Description Length: Review paper". Journal of the American Statistical Association. 96 (454): 746–774. CiteSeerX 10.1.1.43.6581. doi:10.1198/016214501753168398. JSTOR 2670311. MR 1939352. S2CID 14460386. Archived from the original on 2004-11-16. Hinkelmann, Klaus; Kempthorne, Oscar (2008). Introduction to Experimental Design (Second ed.). Wiley. ISBN 978-0-471-72756-9. Kolmogorov, Andrei N. (1963). "On tables of random numbers". Sankhyā Ser. A. 25: 369–375. MR 0178484. Reprinted as Kolmogorov, Andrei N. (1998). "On tables of random numbers". Theoretical Computer Science. 207 (2): 387–395. doi:10.1016/S0304-3975(98)00075-9. MR 1643414. Konishi S., Kitagawa G. (2008), Information Criteria and Statistical Modeling, Springer. Kruskal, William (December 1988). "Miracles and statistics: the casual assumption of independence (ASA Presidential Address)". Journal of the American Statistical Association. 83 (404): 929–940. doi:10.2307/2290117. JSTOR 2290117. Le Cam, Lucian. (1986) Asymptotic Methods of Statistical Decision Theory, Springer. ISBN 0-387-96307-3 Moore, D. S.; McCabe, G. P.; Craig, B. A. (2015), Introduction to the Practice of Statistics, Eighth Edition, Macmillan. Neyman, Jerzy (1956). "Note on an article by Sir Ronald Fisher". Journal of the Royal Statistical Society, Series B. 18 (2): 288–294. doi:10.1111/j.2517-6161.1956.tb00236.x. JSTOR 2983716. (reply to Fisher 1955) Peirce, C. S. (1877–1878), "Illustrations of the logic of science" (series), Popular Science Monthly, vols. 12–13. Relevant individual papers: (1878 March), "The Doctrine of Chances", Popular Science Monthly, v. 12, March issue, pp. 604–615. Internet Archive Eprint. (1878 April), "The Probability of Induction", Popular Science Monthly, v. 12, pp. 705–718. Internet Archive Eprint. (1878 June), "The Order of Nature", Popular Science Monthly, v. 13, pp. 203–217.Internet Archive Eprint. (1878 August), "Deduction, Induction, and Hypothesis", Popular Science Monthly, v. 13, pp. 470–482. Internet Archive Eprint. Peirce, C. S. (1883), "A Theory of probable inference", Studies in Logic, pp. 126-181, Little, Brown, and Company. (Reprinted 1983, John Benjamins Publishing Company, ISBN 90-272-3271-7) Freedman, D.A; Pisani, R.; Purves, R.A. (1978). Statistics. New York: W. W. Norton & Company. Pfanzagl, Johann; with the assistance of R. Hamböker (1994). Parametric Statistical Theory. Berlin: Walter de Gruyter. ISBN 978-3-11-013863-4. MR 1291393. Rissanen, Jorma (1989). Stochastic Complexity in Statistical Inquiry. Series in Computer Science. Vol. 15. Singapore: World Scientific. ISBN 978-9971-5-0859-3. MR 1082556. Soofi, Ehsan S. (December 2000). "Principal information-theoretic approaches (Vignettes for the Year 2000: Theory and Methods, ed. by George Casella)". Journal of the American Statistical Association. 95 (452): 1349–1353. doi:10.1080/01621459.2000.10474346. JSTOR 2669786. MR 1825292. S2CID 120143121. Traub, Joseph F.; Wasilkowski, G. W.; Wozniakowski, H. (1988). Information-Based Complexity. Academic Press. ISBN 978-0-12-697545-1. Zabell, S. L. (Aug 1992). "R. A. Fisher and Fiducial Argument". Statistical Science. 7 (3): 369–387. doi:10.1214/ss/1177011233. JSTOR 2246073. Casella, G., Berger, R. L. (2002). Statistical Inference. Duxbury Press. ISBN 0-534-24312-6 Freedman, D.A. (1991). "Statistical models and shoe leather". Sociological Methodology. 21: 291–313. doi:10.2307/270939. JSTOR 270939. Held L., Bové D.S. (2014). Applied Statistical Inference—Likelihood and Bayes (Springer). Lenhard, Johannes (2006). "Models and Statistical Inference: the controversy between Fisher and Neyman–Pearson" (PDF). British Journal for the Philosophy of Science. 57: 69–91. doi:10.1093/bjps/axi152. S2CID 14136146. Lindley, D (1958). "Fiducial distribution and Bayes' theorem". Journal of the Royal Statistical Society, Series B. 20: 102–7. doi:10.1111/j.2517-6161.1958.tb00278.x. Rahlf, Thomas (2014). "Statistical Inference", in Claude Diebolt, and Michael Haupert (eds.), "Handbook of Cliometrics (Springer Reference Series)", Berlin/Heidelberg: Springer. Reid, N.; Cox, D. R. (2014). "On Some Principles of Statistical Inference". International Statistical Review. 83 (2): 293–308. doi:10.1111/insr.12067. hdl:10.1111/insr.12067. S2CID 17410547. Sagitov, Serik (2022). "Statistical Inference". Wikibooks. Young, G.A., Smith, R.L. (2005). Essentials of Statistical Inference, CUP. ISBN 0-521-83971-8 Wikimedia Commons has media related to Statistical inference. Wikiversity has learning resources about Statistical inference Statistical Inference – lecture on the MIT OpenCourseWare platform Statistical Inference – lecture by the National Programme on Technology Enhanced Learning An online, Bayesian (MCMC) demo/calculator is available at causaScientia Portal: Mathematics Retrieved from " At StataCorp, we are researchers who build powerful, reliable statistical software designed for researchers, analysts, and data professionals. For over 40 years, Stata has been the trusted choice for statistical analysis, visualization, data manipulation, and reproducible reporting—all in one complete package. From academia to industry, our users rely on Stata's accuracy, reproducibility, and seamless workflow to uncover insights and drive discoveries. Backed by a global user community and expert support, Stata empowers you to focus on your research, not your software. Comprehensive statistical tools for every field Effortless visualization and data management Reproducible workflows and automated reporting Reliable results you can trust Explore how Stata can power your research and analyses: stata.com Stata The all-in-one statistical software package for data science. Analyze, visualize, & automate with Stata. 11,473 followers 1mo Stata 19 is here! Here's a glimpse of some of what we've introduced: Machine learning via H2O: Ensemble decision trees Conditional average treatment effects (CATE) High-dimensional fixed effects (HDFE) New and improved graphics features Do-file Editor enhancements More powerful tables And much more Explore all the new features: stata.com/stata19 11,473 followers 4h Meet our Day 1 speakers for the 2025 Stata Conference. These experts will be sharing practical tips, new tools, and applied research on everything from reproducibility and model building in Mata to health data and visualization. You'll hear from: ↳ James Hardin (University of South Carolina) "A generalized model specification system in Mata" ↳ Sergio Cervantes "Creating pre- and posttest visualizations with Stata, Excel, and Tableau—a dynamic approach" ↳ Jorge Pérez Pérez (Banco de México) "xtevent: Estimation and visualization in the linear panel event-study design" ↳ Tom Stringham (StataCorp) "Control-function linear and probit models in Stata" ↳ William Buchanan (SAG Corporation) "Boosting performance of regular expressions" ↳ Luis Eduardo San Martin (The World Bank) "Introducing repscan: Automated detection of Stata commands linked to common reproducibility failures" ↳ Miguel Pérez Andrade (MultiOn Consulting) "Estimating censored food demand in Mexico with quaidsce" ↳ Sophia Rabe-Hesketh (University of California, Berkeley) "Valid standard errors for misspecified Bayesian models" ↳ Seolah Kim (California State University, Los Angeles) "Estimating correlated random coefficient models with endogeneity" ↳ Rebecca Raciborski (U.S. Department of Veterans Affairs) "cea: A suite of commands for trial-based cost-effectiveness analysis" Connect with the Stata community and explore real-world applications, new commands, and reproducible workflows—all delivered by leading voices in research, policy, and statistics. Join us 31 July–01 August in Nashville. View the full program and register: stata.com/us25 #Stata2025 #StataConference #DataScience #Econometrics #PublicHealth #Statistics #Research 11,473 followers 4d 40 years of Stata. One exciting day in Ottawa. Join us on 3 October for the 2025 Canadian Stata Conference—a gathering of researchers, data professionals, and developers exploring what's new in Stata and how it's being applied in real-world work. Celebrate this milestone year by being part of the program or signing up to be notified when registration opens. Present your research Explore Stata 19's latest features Network with fellow users and developers Ottawa, Canada | 3 October 2025#StataCanada #StataCommunity 11,473 followers 5d New in StataNow: Power and sample-size analysis for logistic regression: Effective study design requires balancing power and resources. Stata's new command helps you determine the optimal sample size when testing one coefficient of interest in a logistic regression model. Use to compute sample size, power, or effect size for logistic regression with: One binary covariate Two binary covariates Mix of binary and continuous covariates And like all Stata power commands, supports tabular output and automated graphs—all customizable. Learn more: 11,473 followers 5d In predictive modeling, obtaining point predictions is only part of the story. Equally important is quantifying the uncertainty around those predictions. Our latest blog post explores how to construct prediction intervals using gradient boosting machine (GBM) and conformal inference in Stata 19's h2oml suite. In this post, we walk through how to: Use Stata's new h2oml commands for conformal inference Use conformal inference to estimate prediction intervals with GBM Visualize those intervals If you do predictive modeling, conformalized prediction intervals provide an additional layer of reliability to your analysis. Read the full blog post: #Stata19 #GBM #ConformalInference #MachineLearning #StataBlog 11,473 followers 6d Take your modeling beyond the linear with Stata 19 + H2O. Some relationships in data aren't linear. They're complex, non-additive, and sometimes unpredictable. With the new suite in Stata 19, you can now use ensemble decision trees—random forest and gradient boosting machine (GBM)—to model those complexities and uncover patterns traditional models may miss. Here's how empowers your workflow: Model classification and regression problems ⚙ Tune hyperparameters to optimize your model Use cross-validation to reduce overfitting Evaluate performance with buil metrics Identify variable importance and understand model drivers Make predictions And you can do all of this without leaving Stata, using familiar syntax or point-and-click. Learn more: 11,473 followers 1w Move beyond average treatment effects with Stata 19. The new command allows researchers to estimate conditional average treatment effects—allowing exploration of heterogeneity in causal effects across individuals and groups. Whether you're comparing policy impacts or analyzing intervention outcomes, cate offers the flexibility and precision your analysis demands. ▶ Watch the video walkthrough: #Stata19 #DataAnalysis #StatisticalSoftware 11,473 followers 1w A conference where you can form connections with, ask questions to, and get answers from the people behind the software. That's what makes the Stata Conference unique. Because it's not just about sessions—it's about community. This July, we invite you to join us in Nashville for the 2025 Stata Conference—our 40th anniversary celebration and a milestone in reproducible statistical research. The 2025 Stata Conference is your opportunity to: Learn from expert-led presentations and real-world research use cases Attend our open panel Q&A with the StataCorp developers—your chance to share feedback and influence future updates Connect with researchers, educators, and data professionals from around the world Participate in our interactive preconference workshop on causal inference and treatment effects Celebrate 40 years of Stata with us at a special reception And don't miss this year's invited speakers: ✔ Ben Jann: Drawing maps in Stata using geoplot ✔ Sophia Rabe-Hesketh: Valid standard errors for misspecified Bayesian models ✔ Jeffrey Wooldridge: Difference in difference for nonbinary treatments in Stata And of course, you'll be surrounded by a community that shares your passion for rigorous, reproducible research. Whether you're looking to sharpen your skills, expand your network, or simply talk shop with fellow Stata users—this is the conference for you. July 31 – August 1, 2025 Nashville, TN Preconference workshop: July 30 ➡ View the full program and register now: #Stata #StataConference #StataCommunity #DataScience 11,473 followers 1w We're getting ready for the 2025 AIR Forum! The Stata team is in full prep mode as we gear up for this year's AIR Forum in Orlando, Florida. We'll be at booths 408 & 410 from 20–23 May, ready to connect with institutional researchers, share what's new in Stata, and showcase tools that support data-driven decision-making in higher education. Stop by to chat with our team Explore how Stata supports institutional research offices Take home some Stata gear See you in Orlando! #Stata19 #AIRForum #InstitutionalResearch 11,473 followers 1w Have a question about Stata? Want to explore a statistical method or dig into best practices? Statalist is the official forum for Stata users—run by users, for users. Whether you're troubleshooting code, exploring statistical techniques, or sharing workflows, Statalist is the place to connect. Ask questions and get answers from experienced users Explore discussions on Stata, statistics, and everything in between Engage with a global community of analysts, researchers, and educators Visit the forum and join the conversation: #Stata #Statalist #StataCommunity Citations are extracted by the CitEc Project, subscribe to its RSS feed for this item. Cited by: Valérie Orozco & Christophe Bontemps & Elise Maigné & Virginie Piguet & Annie Hofstetter & Anne Lacroix & Fabrice Levert & Jean-Marc Rousselle, 2020. "How To Make A Pie: Reproducible Research For Empirical Economics And Econometrics," Journal of Economic Surveys, Wiley Blackwell, vol. 34(5), pages 1134-1169, December. Valérie Orozco & Christophe Bontemps & Élise Maigné & Virginie Piguet & Annie Hofstetter & Anne Marie Lacroix & Fabrice Levert & Jean-Marc Rousselle, 2017. "How to make a pie? Reproducible Research for Empirical Economics & Econometrics," Post-Print hal-01939942, HAL. Orozco, Valérie & Bontemps, Christophe & Maigné, Elise & Piguet, V. & Hofstetter, A. & Lacroix, Anne & Levert, F. & Rousselle, J.M, 2018. "How To Make A Pie: Reproducible Research for Empirical Economics & Econometrics," TSE Working Papers 18-933, Toulouse School of Economics (TSE). Valérie Orozco & Christophe Bontemps & Élise Maigné & Virginie Piguet & Annie Hofstetter & Anne Marie Lacroix & Fabrice Levert & Jean-Marc Rousselle, 2019. "How To Make A Pie: Reproducible Research for Empirical Economics & Econometrics," Working Papers hal-02025843, HAL. Markku Maula & Wouter Stam, 2020. "Enhancing Rigor in Quantitative Entrepreneurship Research," Entrepreneurship Theory and Practice, , vol. 44(6), pages 1059-1090, November. Richard Ball & Norm Medeiros, 2012. "Teaching Integrity in Empirical Research: A Protocol for Documenting Data Management and Analysis," The Journal of Economic Education, Taylor & Francis Journals, vol. 43(2), pages 182-189, April. Fernando Hoces de la Guardia & Sean Grant & Edward Miguel, 2021. "A framework for open policy analysis," Science and Public Policy, Oxford University Press, vol. 48(2), pages 154-163. Hoces de la Guardia, Fernando & Grant, Sean & Miguel, Edward, 2018. "A Framework for Open Policy Analysis," MetaArXiv jnyqh, Center for Open Science. de la Guardia, Fernando Hoces & Grant, Sean & Miguel, Edward, 2021. "A framework for open policy analysis," Department of Economics, Working Paper Series qt05r470xk, Department of Economics, Institute for Business and Economic Research, UC Berkeley. Xiaodan Hu & Pedro Villarreal, 2019. "Public Tuition on the Rise: Estimating the Effects of Louisiana's Performance-Based Funding Policy on Institutional Tuition Levels," Research in Higher Education, Springer;Association for Institutional Research, vol. 60(5), pages 636-669, August. Stefan Friedhoff & Christian Meier zu Verl & Christian Pietsch & Christian Meyer & Johanna Vompras & Stefan Liebig, 2013. "Replicability and Comprehensibility of Social Research and its Technical Implementation," RatSWD Working Papers 219, German Data Forum (RatSWD). David M. Drukker & Hua Peng & Ingmar Prucha & Rafal Raciborski, 2013. "Creating and managing spatial-weighting matrices with the spmat command," Stata Journal, StataCorp LLC, vol. 13(2), pages 242-286, June. All material on this site has been provided by the respective publishers and authors. You can help correct errors and omissions. When requesting a correction, please mention this item's handle: RePEc:tsj:spbook:wdaus. See general information about how to correct material in RePEc. If you have authored this item and are not yet registered with RePEc, we encourage you to do it here. This allows to link your profile to this item. It also allows you to accept potential citations to this item that we are uncertain about. We have no bibliographic references for this item. You can help adding them by using this form . If you know of missing items citing this one, you can help us creating those links by adding the relevant references in the same way as above, for each refering item. If you are a registered author of this item, you may also want to check the "citations" tab in your RePEc Author Service profile, as there may be some citations waiting for confirmation. For technical questions regarding this item, or to correct its authors, title, abstract, bibliographic or download information, contact: Christopher F. Baum or Lisa Gilmore (email available below). General contact details of provider: . Please note that corrections may take a couple of weeks to filter through the various RePEc services.