Title: Time series

URL: https://en.wikipedia.org/wiki/Time_series

PageID: 406624

Categories: Category:Machine learning, Category:Mathematical and quantitative methods (economics), Category:Mathematics in medicine, Category:Statistical data types, Category:Time series

Source: Wikipedia (CC BY-SA 4.0).

In mathematics , a time series is a series of data points indexed (or listed or graphed) in time order. Most commonly, a time series is a sequence taken at successive equally spaced points in time. Thus it is a sequence of discrete-time data. Examples of time series are heights of ocean tides , counts of sunspots , and the daily closing value of the Dow Jones Industrial Average .

A time series is very frequently plotted via a run chart (which is a temporal line chart). Time series are used in statistics , signal processing , pattern recognition , econometrics , mathematical finance , weather forecasting , earthquake prediction , electroencephalography , control engineering , astronomy , communications engineering , and largely in any domain of applied science and engineering which involves temporal measurements.

Time series analysis comprises methods for analyzing time series data in order to extract meaningful statistics and other characteristics of the data. Time series forecasting is the use of a model to predict future values based on previously observed values. Generally, time series data is modelled as a stochastic process . While regression analysis is often employed in such a way as to test relationships between one or more different time series, this type of analysis is not usually called "time series analysis", which refers in particular to relationships between different points in time within a single series.

Time series data have a natural temporal ordering. This makes time series analysis distinct from cross-sectional studies , in which there is no natural ordering of the observations (e.g. explaining people's wages by reference to their respective education levels, where the individuals' data could be entered in any order). Time series analysis is also distinct from spatial data analysis where the observations typically relate to geographical locations (e.g. accounting for house prices by the location as well as the intrinsic characteristics of the houses). A stochastic model for a time series will generally reflect the fact that observations close together in time will be more closely related than observations further apart. In addition, time series models will often make use of the natural one-way ordering of time so that values for a given period will be expressed as deriving in some way from past values, rather than from future values (see time reversibility).

Time series analysis can be applied to real-valued , continuous data, discrete numeric data, or discrete symbolic data (i.e. sequences of characters, such as letters and words in the English language [1]).

Methods for analysis

Methods for time series analysis may be divided into two classes: frequency-domain methods and time-domain methods. The former include spectral analysis and wavelet analysis ; the latter include auto-correlation and cross-correlation analysis. In the time domain, correlation and analysis can be made in a filter-like manner using scaled correlation , thereby mitigating the need to operate in the frequency domain.

Additionally, time series analysis techniques may be divided into parametric and non-parametric methods. The parametric approaches assume that the underlying stationary stochastic process has a certain structure which can be described using a small number of parameters (for example, using an autoregressive or moving-average model). In these approaches, the task is to estimate the parameters of the model that describes the stochastic process. By contrast, non-parametric approaches explicitly estimate the covariance or the spectrum of the process without assuming that the process has any particular structure.

Methods of time series analysis may also be divided into linear and non-linear , and univariate and multivariate .

Panel data

A time series is one type of panel data . Panel data is the general class, a multidimensional data set, whereas a time series data set is a one-dimensional panel (as is a cross-sectional dataset). A data set may exhibit characteristics of both panel data and time series data. One way to tell is to ask what makes one data record unique from the other records. If the answer is the time data field, then this is a time series data set candidate. If determining a unique record requires a time data field and an additional identifier which is unrelated to time (e.g. student ID, stock symbol, country code), then it is panel data candidate. If the differentiation lies on the non-time identifier, then the data set is a cross-sectional data set candidate.

Analysis

There are several types of motivation and data analysis available for time series which are appropriate for different purposes.

Motivation

In the context of statistics , econometrics , quantitative finance , seismology , meteorology , and geophysics the primary goal of time series analysis is forecasting . In the context of signal processing , control engineering and communication engineering it is used for signal detection. Other applications are in data mining , pattern recognition and machine learning , where time series analysis can be used for clustering , [2] [3] [4] classification , [5] query by content, [6] anomaly detection as well as forecasting . [7]

Exploratory analysis

A simple way to examine a regular time series is manually with a line chart . The datagraphic shows tuberculosis deaths in the United States, [8] along with the yearly change and the percentage change from year to year. The total number of deaths declined in every year until the mid-1980s, after which there were occasional increases, often proportionately - but not absolutely - quite large.

A study of corporate data analysts found two challenges to exploratory time series analysis: discovering the shape of interesting patterns, and finding an explanation for these patterns. [9] Visual tools that represent time series data as heat map matrices can help overcome these challenges.

Estimation, filtering, and smoothing

This approach may be based on harmonic analysis and filtering of signals in the frequency domain using the Fourier transform , and spectral density estimation . Its development was significantly accelerated during World War II by mathematician Norbert Wiener , electrical engineers Rudolf E. Kálmán , Dennis Gabor and others for filtering signals from noise and predicting signal values at a certain point in time.

An equivalent effect may be achieved in the time domain, as in a Kalman filter ; see filtering and smoothing for more techniques.

Other related techniques include:

Autocorrelation analysis to examine serial dependence

Spectral analysis to examine cyclic behavior which need not be related to seasonality . For example, sunspot activity varies over 11 year cycles. [10] [11] Other common examples include celestial phenomena, weather patterns, neural activity, commodity prices, and economic activity.

Separation into components representing trend, seasonality, slow and fast variation, and cyclical irregularity: see trend estimation and decomposition of time series

Curve fitting

Curve fitting [12] [13] is the process of constructing a curve , or mathematical function , that has the best fit to a series of data points, [14] possibly subject to constraints. [15] [16] Curve fitting can involve either interpolation , [17] [18] where an exact fit to the data is required, or smoothing , [19] [20] in which a "smooth" function is constructed that approximately fits the data. A related topic is regression analysis , [21] [22] which focuses more on questions of statistical inference such as how much uncertainty is present in a curve that is fit to data observed with random errors. Fitted curves can be used as an aid for data visualization, [23] [24] to infer values of a function where no data are available, [25] and to summarize the relationships among two or more variables. [26] Extrapolation refers to the use of a fitted curve beyond the range of the observed data, [27] and is subject to a degree of uncertainty [28] since it may reflect the method used to construct the curve as much as it reflects the observed data.

For processes that are expected to generally grow in magnitude one of the curves in the graphic (and many others) can be fitted by estimating their parameters.

The construction of economic time series involves the estimation of some components for some dates by interpolation between values ("benchmarks") for earlier and later dates. Interpolation is estimation of an unknown quantity between two known quantities (historical data), or drawing conclusions about missing information from the available information ("reading between the lines"). [29] Interpolation is useful where the data surrounding the missing data is available and its trend, seasonality, and longer-term cycles are known. This is often done by using a related series known for all relevant dates. [30] Alternatively polynomial interpolation or spline interpolation is used where piecewise polynomial functions are fitted in time intervals such that they fit smoothly together. A different problem which is closely related to interpolation is the approximation of a complicated function by a simple function (also called regression). The main difference between regression and interpolation is that polynomial regression gives a single polynomial that models the entire data set. Spline interpolation, however, yield a piecewise continuous function composed of many polynomials to model the data set.

Extrapolation is the process of estimating, beyond the original observation range, the value of a variable on the basis of its relationship with another variable. It is similar to interpolation , which produces estimates between known observations, but extrapolation is subject to greater uncertainty and a higher risk of producing meaningless results.

Function approximation

In general, a function approximation problem asks us to select a function among a well-defined class that closely matches ("approximates") a target function in a task-specific way.

One can distinguish two major classes of function approximation problems: First, for known target functions, approximation theory is the branch of numerical analysis that investigates how certain known functions (for example, special functions) can be approximated by a specific class of functions (for example, polynomials or rational functions) that often have desirable properties (inexpensive computation, continuity, integral and limit values, etc.).

Second, the target function, call it g , may be unknown; instead of an explicit formula, only a set of points (a time series) of the form (x , g (x)) is provided. Depending on the structure of the domain and codomain of g , several techniques for approximating g may be applicable. For example, if g is an operation on the real numbers , techniques of interpolation , extrapolation , regression analysis , and curve fitting can be used. If the codomain (range or target set) of g is a finite set, one is dealing with a classification problem instead. A related problem of online time series approximation [31] is to summarize the data in one-pass and construct an approximate representation that can support a variety of time series queries with bounds on worst-case error.

To some extent, the different problems (regression , classification , fitness approximation) have received a unified treatment in statistical learning theory , where they are viewed as supervised learning problems.

Prediction and forecasting

In statistics , prediction is a part of statistical inference . One particular approach to such inference is known as predictive inference , but the prediction can be undertaken within any of the several approaches to statistical inference. Indeed, one description of statistics is that it provides a means of transferring knowledge about a sample of a population to the whole population, and to other related populations, which is not necessarily the same as prediction over time. When information is transferred across time, often to specific points in time, the process is known as forecasting .

Fully formed statistical models for stochastic simulation purposes, so as to generate alternative versions of the time series, representing what might happen over non-specific time-periods in the future

Simple or fully formed statistical models to describe the likely outcome of the time series in the immediate future, given knowledge of the most recent outcomes (forecasting).

Forecasting on time series is usually done using automated statistical software packages and programming languages, such as Julia , Python , R , SAS , SPSS and many others.

Forecasting on large scale data can be done with Apache Spark using the Spark-TS library, a third-party package. [32]

Classification

Assigning time series pattern to a specific category, for example identify a word based on series of hand movements in sign language .

Segmentation

Splitting a time-series into a sequence of segments. It is often the case that a time-series can be represented as a sequence of individual segments, each with its own characteristic properties. For example, the audio signal from a conference call can be partitioned into pieces corresponding to the times during which each person was speaking. In time-series segmentation, the goal is to identify the segment boundary points in the time-series, and to characterize the dynamical properties associated with each segment. One can approach this problem using change-point detection , or by modeling the time-series as a more sophisticated system, such as a Markov jump linear system.

Clustering

Time series data may be clustered, however special care has to be taken when considering subsequence clustering. [33] [34] Time series clustering may be split into

whole time series clustering (multiple time series for which to find a cluster)

subsequence time series clustering (single timeseries, split into chunks using sliding windows)

time point clustering

Subsequence time series clustering

Subsequence time series clustering resulted in unstable (random) clusters induced by the feature extraction using chunking with sliding windows. [35] It was found that the cluster centers (the average of the time series in a cluster - also a time series) follow an arbitrarily shifted sine pattern (regardless of the dataset, even on realizations of a random walk). This means that the found cluster centers are non-descriptive for the dataset because the cluster centers are always nonrepresentative sine waves.

Models

Classical models (AR, ARMA, ARIMA, and well-known variations)

Models for time series data can have many forms and represent different stochastic processes . When modeling variations in the level of a process, three broad classes of practical importance are the autoregressive (AR) models, the integrated (I) models, and the moving-average (MA) models. These three classes depend linearly on previous data points. [36] Combinations of these ideas produce autoregressive moving-average (ARMA) and autoregressive integrated moving-average (ARIMA) models. The autoregressive fractionally integrated moving-average (ARFIMA) model generalizes the former three. Another important generalization is the time-varying autoregressive (TVAR) model, in which the AR coefficients are allowed to change over time, enabling the model to capture evolving or non-stationary dynamics. Extensions of these classes to deal with vector-valued data are available under the heading of multivariate time-series models and sometimes the preceding acronyms are extended by including an initial "V" for "vector", as in VAR for vector autoregression . An additional set of extensions of these models is available for use where the observed time-series is driven by some "forcing" time-series (which may not have a causal effect on the observed series): the distinction from the multivariate case is that the forcing series may be deterministic or under the experimenter's control. For these models, the acronyms are extended with a final "X" for "exogenous".

Time-varying autoregressive (TVAR) models

Time-varying autoregressive (TVAR) models are especially useful for analyzing non-stationary time series in which the underlying dynamics evolve over time in complex ways, not limited to classical forms of variation such as trends or seasonal patterns. Unlike classical AR models with fixed parameters, TVAR models allow the autoregressive coefficients to vary as functions of time, typically represented through basis function expansions whose form and complexity are determined by the user, allowing for highly flexible modeling of time-varying behavior, which enables them to capture specific non-stationary patterns exhibited by the data. This flexibility makes them well suited to modeling structural changes, regime shifts, or gradual evolutions in a system's behavior. TVAR time-series models are widely applied in fields such as signal processing [37] [38] , economics [39] , finance [40] , reliability and condition monitoring [41] [42] [43] , telecommunications [44] [45] , neuroscience [46] , climate sciences [47] , and hydrology [48] , where time series often display dynamics that traditional linear models cannot adequately represent. Estimation of TVAR models typically involves methods such as kernel smoothing [49] , recursive least squares [50] , or Kalman filtering [51] .

Non-linear models

Non-linear dependence of the level of a series on previous data points is of interest, partly because of the possibility of producing a chaotic time series. However, more importantly, empirical investigations can indicate the advantage of using predictions derived from non-linear models, over those from linear models, as for example in nonlinear autoregressive exogenous models . Further references on nonlinear time series analysis: (Kantz and Schreiber), [52] and (Abarbanel) [53]

Among other types of non-linear time series models, there are models to represent the changes of variance over time (heteroskedasticity). These models represent autoregressive conditional heteroskedasticity (ARCH) and the collection comprises a wide variety of representation (GARCH , TARCH, EGARCH, FIGARCH, CGARCH, etc.). Here changes in variability are related to, or predicted by, recent past values of the observed series. This is in contrast to other possible representations of locally varying variability, where the variability might be modelled as being driven by a separate time-varying process, as in a doubly stochastic model .

Other modelling approaches

In a recent work on "model-free" analyses (a term often used to refer to analyses that do not rely on modeling the processes evolution over time with a parametric mathematical expression), wavelet transform based methods (for example locally stationary wavelets and wavelet decomposed neural networks) have gained favor. [54] Multiscale (often referred to as multiresolution) techniques decompose a given time series, attempting to illustrate time dependence at multiple scales. See also Markov switching multifractal (MSMF) techniques for modeling volatility evolution.

A hidden Markov model (HMM) is a statistical Markov model in which the system being modeled is assumed to be a Markov process with unobserved (hidden) states. An HMM can be considered as the simplest dynamic Bayesian network . HMM models are widely used in speech recognition , for translating a time series of spoken words into text.

Many of these models are collected in the python package sktime .

Notation

A number of different notations are in use for time-series analysis. A common notation specifying a time series X that is indexed by the natural numbers is written

Another common notation is

where T is the index set .

Conditions

There are two sets of conditions under which much of the theory is built:

Stationary process

Ergodic process

Ergodicity implies stationarity, but the converse is not necessarily the case. Stationarity is usually classified into strict stationarity and wide-sense or second-order stationarity . Both models and applications can be developed under each of these conditions, although the models in the latter case might be considered as only partly specified.

In addition, time-series analysis can be applied where the series are seasonally stationary or non-stationary. Situations where the amplitudes of frequency components change with time can be dealt with in time-frequency analysis which makes use of a time–frequency representation of a time-series or signal. [55]

Tools

Tools for investigating time-series data include:

Consideration of the autocorrelation function and the spectral density function (also cross-correlation functions and cross-spectral density functions)

Scaled cross- and auto-correlation functions to remove contributions of slow components [56]

Performing a Fourier transform to investigate the series in the frequency domain

Performing a clustering analysis [57]

Discrete, continuous or mixed spectra of time series, depending on whether the time series contains a (generalized) harmonic signal or not

Use of a filter to remove unwanted noise

Principal component analysis (or empirical orthogonal function analysis)

Singular spectrum analysis

"Structural" models: General state space models Unobserved components models

General state space models

Unobserved components models

Machine learning Artificial neural networks Support vector machine Fuzzy logic Gaussian process Genetic programming Gene expression programming Hidden Markov model Multi expression programming

Artificial neural networks

Support vector machine

Fuzzy logic

Gaussian process

Genetic programming

Gene expression programming

Hidden Markov model

Multi expression programming

Queueing theory analysis

Control chart Shewhart individuals control chart CUSUM chart EWMA chart

Shewhart individuals control chart

CUSUM chart

EWMA chart

Detrended fluctuation analysis

Nonlinear mixed-effects modeling

Dynamic time warping [58]

Dynamic Bayesian network

Time-frequency analysis techniques: Fast Fourier transform Continuous wavelet transform Short-time Fourier transform Chirplet transform Fractional Fourier transform

Fast Fourier transform

Continuous wavelet transform

Short-time Fourier transform

Chirplet transform

Fractional Fourier transform

Chaotic analysis Correlation dimension Recurrence plots Recurrence quantification analysis Lyapunov exponents Entropy encoding

Correlation dimension

Recurrence plots

Recurrence quantification analysis

Lyapunov exponents

Entropy encoding

Measures

Time-series metrics or features that can be used for time series classification or regression analysis : [59]

Univariate linear measures Moment (mathematics) Spectral band power Spectral edge frequency Accumulated energy (signal processing) Characteristics of the autocorrelation function Hjorth parameters FFT parameters Autoregressive model parameters Mann–Kendall test

Moment (mathematics)

Spectral band power

Spectral edge frequency

Accumulated energy (signal processing)

Characteristics of the autocorrelation function

Hjorth parameters

FFT parameters

Autoregressive model parameters

Mann–Kendall test

Univariate non-linear measures Measures based on the correlation sum Correlation dimension Correlation integral Correlation density Correlation entropy Approximate entropy [60] Sample entropy Fourier entropy [uk] Wavelet entropy Dispersion entropy Fluctuation dispersion entropy Rényi entropy Higher-order methods Marginal predictability Dynamical similarity index State space dissimilarity measures Lyapunov exponent Permutation methods Local flow

Measures based on the correlation sum

Correlation dimension

Correlation integral

Correlation density

Correlation entropy

Approximate entropy [60]

Sample entropy

Fourier entropy [uk]

Wavelet entropy

Dispersion entropy

Fluctuation dispersion entropy

Rényi entropy

Higher-order methods

Marginal predictability

Dynamical similarity index

State space dissimilarity measures

Lyapunov exponent

Permutation methods

Local flow

Other univariate measures Algorithmic complexity Kolmogorov complexity estimates Hidden Markov model states Rough path signature [61] Surrogate time series and surrogate correction Loss of recurrence (degree of non-stationarity)

Algorithmic complexity

Kolmogorov complexity estimates

Hidden Markov model states

Rough path signature [61]

Surrogate time series and surrogate correction

Loss of recurrence (degree of non-stationarity)

Bivariate linear measures Maximum linear cross-correlation Linear Coherence (signal processing)

Maximum linear cross-correlation

Linear Coherence (signal processing)

Bivariate non-linear measures Non-linear interdependence Dynamical Entrainment (physics) Measures for phase synchronization Measures for phase locking

Non-linear interdependence

Dynamical Entrainment (physics)

Measures for phase synchronization

Measures for phase locking

Similarity measures : [62] Cross-correlation Dynamic time warping [58] Hidden Markov model Edit distance Total correlation Newey–West estimator Prais–Winsten transformation Data as vectors in a metrizable space Minkowski distance Mahalanobis distance Data as time series with envelopes Global standard deviation Local standard deviation Windowed standard deviation Data interpreted as stochastic series Pearson product-moment correlation coefficient Spearman's rank correlation coefficient Data interpreted as a probability distribution function Kolmogorov–Smirnov test Cramér–von Mises criterion

Cross-correlation

Dynamic time warping [58]

Hidden Markov model

Edit distance

Total correlation

Newey–West estimator

Prais–Winsten transformation

Data as vectors in a metrizable space Minkowski distance Mahalanobis distance

Minkowski distance

Mahalanobis distance

Data as time series with envelopes Global standard deviation Local standard deviation Windowed standard deviation

Global standard deviation

Local standard deviation

Windowed standard deviation

Data interpreted as stochastic series Pearson product-moment correlation coefficient Spearman's rank correlation coefficient

Pearson product-moment correlation coefficient

Spearman's rank correlation coefficient

Data interpreted as a probability distribution function Kolmogorov–Smirnov test Cramér–von Mises criterion

Kolmogorov–Smirnov test

Cramér–von Mises criterion

Visualization

Time series can be visualized with two categories of chart: Overlapping Charts and Separated Charts. Overlapping Charts display all-time series on the same layout while Separated Charts presents them on different layouts (but aligned for comparison purpose) [63]

Overlapping charts

Braided graphs

Line charts

Slope graphs

GapChart [fr]

Separated charts

Horizon graphs

Reduced line chart (small multiples)

Silhouette graph

Circular silhouette graph

See also

Anomaly time series

Chirp

Decomposition of time series

Detrended fluctuation analysis

Digital signal processing

Distributed lag

Estimation theory

Forecasting

Frequency spectrum

Hurst exponent

Least-squares spectral analysis

Monte Carlo method

Panel analysis

Random walk

Scaled correlation

Seasonal adjustment

Sequence analysis

Signal processing

Time series database (TSDB)

Trend estimation

Unevenly spaced time series

References

Further reading

De Gooijer, Jan G.; Hyndman, Rob J. (2006). "25 Years of Time Series Forecasting". International Journal of Forecasting . Twenty Five Years of Forecasting. 22 (3): 443– 473. CiteSeerX 10.1.1.154.9227 . doi : 10.1016/j.ijforecast.2006.01.001 . S2CID 14996235 .

Box, George ; Jenkins, Gwilym (1976), Time Series Analysis: forecasting and control, rev. ed. , Oakland, California: Holden-Day

Durbin J. , Koopman S.J. (2001), Time Series Analysis by State Space Methods , Oxford University Press .

Gershenfeld, Neil (2000), The Nature of Mathematical Modeling , Cambridge University Press , ISBN 978-0-521-57095-4 , OCLC 174825352

Hamilton, James (1994), Time Series Analysis , Princeton University Press , ISBN 978-0-691-04289-3

Priestley, M. B. (1981), Spectral Analysis and Time Series , Academic Press . ISBN 978-0-12-564901-8

Shasha, D. (2004), High Performance Discovery in Time Series , Springer , ISBN 978-0-387-00857-8

Shumway R. H., Stoffer D. S. (2017), Time Series Analysis and its Applications: With R Examples (ed. 4) , Springer, ISBN 978-3-319-52451-1

Weigend A. S., Gershenfeld N. A. (Eds.) (1994), Time Series Prediction: Forecasting the Future and Understanding the Past . Proceedings of the NATO Advanced Research Workshop on Comparative Time Series Analysis (Santa Fe, May 1992), Addison-Wesley .

Wiener, Norbert (1949). Extrapolation, Interpolation, and Smoothing of Stationary Time Series: With Engineering Applications . MIT Press . ISBN 9780262257190 . {{ cite book }} : ISBN / Date incompatibility (help)

Woodward, W. A., Gray, H. L. & Elliott, A. C. (2012), Applied Time Series Analysis , CRC Press .

Auffarth, Ben (2021). Machine Learning for Time-Series with Python: Forecast, predict, and detect anomalies with state-of-the-art machine learning methods (1st ed.). Packt Publishing. ISBN 978-1801819626 . Retrieved 5 November 2021 .

External links

Introduction to Time series Analysis (Engineering Statistics Handbook) — A practical guide to Time series analysis.

v

t

e

Outline

Index

Mean Arithmetic Arithmetic-Geometric Contraharmonic Cubic Generalized/power Geometric Harmonic Heronian Heinz Lehmer

Arithmetic

Arithmetic-Geometric

Contraharmonic

Cubic

Generalized/power

Geometric

Harmonic

Heronian

Heinz

Lehmer

Median

Mode

Average absolute deviation

Coefficient of variation

Interquartile range

Percentile

Range

Standard deviation

Variance

Central limit theorem

Moments Kurtosis L-moments Skewness

Kurtosis

L-moments

Skewness

Index of dispersion

Contingency table

Frequency distribution

Grouped data

Partial correlation

Pearson product-moment correlation

Rank correlation Kendall's τ Spearman's ρ

Kendall's τ

Spearman's ρ

Scatter plot

Bar chart

Biplot

Box plot

Control chart

Correlogram

Fan chart

Forest plot

Histogram

Pie chart

Q–Q plot

Radar chart

Run chart

Scatter plot

Stem-and-leaf display

Violin plot

Effect size

Missing data

Optimal design

Population

Replication

Sample size determination

Statistic

Statistical power

Sampling Cluster Stratified

Cluster

Stratified

Opinion poll

Questionnaire

Standard error

Blocking

Factorial experiment

Interaction

Random assignment

Randomized controlled trial

Randomized experiment

Scientific control

Adaptive clinical trial

Stochastic approximation

Up-and-down designs

Cohort study

Cross-sectional study

Natural experiment

Quasi-experiment

Population

Statistic

Probability distribution

Sampling distribution Order statistic

Order statistic

Empirical distribution Density estimation

Density estimation

Statistical model Model specification L p space

Model specification

L p space

Parameter location scale shape

location

scale

shape

Parametric family Likelihood (monotone) Location–scale family Exponential family

Likelihood (monotone)

Location–scale family

Exponential family

Completeness

Sufficiency

Statistical functional Bootstrap U V

Bootstrap

U

V

Optimal decision loss function

loss function

Efficiency

Statistical distance divergence

divergence

Asymptotics

Robustness

Estimating equations Maximum likelihood Method of moments M-estimator Minimum distance

Maximum likelihood

Method of moments

M-estimator

Minimum distance

Unbiased estimators Mean-unbiased minimum-variance Rao–Blackwellization Lehmann–Scheffé theorem Median unbiased

Mean-unbiased minimum-variance Rao–Blackwellization Lehmann–Scheffé theorem

Rao–Blackwellization

Lehmann–Scheffé theorem

Median unbiased

Plug-in

Confidence interval

Pivot

Likelihood interval

Prediction interval

Tolerance interval

Resampling Bootstrap Jackknife

Bootstrap

Jackknife

1- & 2-tails

Power Uniformly most powerful test

Uniformly most powerful test

Permutation test Randomization test

Randomization test

Multiple comparisons

Likelihood-ratio

Score/Lagrange multiplier

Wald

Z -test (normal)

Student's t -test

F -test

Chi-squared

G -test

Kolmogorov–Smirnov

Anderson–Darling

Lilliefors

Jarque–Bera

Normality (Shapiro–Wilk)

Likelihood-ratio test

Model selection Cross validation AIC BIC

Cross validation

AIC

BIC

Sign Sample median

Sample median

Signed rank (Wilcoxon) Hodges–Lehmann estimator

Hodges–Lehmann estimator

Rank sum (Mann–Whitney)

Nonparametric anova 1-way (Kruskal–Wallis) 2-way (Friedman) Ordered alternative (Jonckheere–Terpstra)

1-way (Kruskal–Wallis)

2-way (Friedman)

Ordered alternative (Jonckheere–Terpstra)

Van der Waerden test

Bayesian probability prior posterior

prior

posterior

Credible interval

Bayes factor

Bayesian estimator Maximum posterior estimator

Maximum posterior estimator

Correlation

Regression analysis

Pearson product-moment
Partial correlation
Confounding variable
Coefficient of determination
Errors and residuals
Regression validation
Mixed effects models
Simultaneous equations models
Multivariate adaptive regression splines (MARS)
Simple linear regression
Ordinary least squares
General linear model
Bayesian regression
Nonlinear regression
Nonparametric
Semiparametric
Isotonic
Robust
Homoscedasticity and Heteroscedasticity
Exponential families
Logistic (Bernoulli) / Binomial / Poisson regressions
Analysis of variance (ANOVA, anova)
Analysis of covariance
Multivariate ANOVA
Degrees of freedom
Cohen's kappa
Contingency table
Graphical model
Log-linear model
McNemar's test
Cochran–Mantel–Haenszel statistics
Regression
Manova
Principal components
Canonical correlation
Discriminant analysis
Cluster analysis
Classification

Structural equation model Factor analysis
Factor analysis
Multivariate distributions Elliptical distributions Normal
Elliptical distributions Normal
Normal
Decomposition
Trend
Stationarity
Seasonal adjustment
Exponential smoothing
Cointegration
Structural break
Granger causality
Dickey–Fuller
Johansen
Q-statistic (Ljung–Box)
Durbin–Watson
Breusch–Godfrey
Autocorrelation (ACF) partial (PACF)
partial (PACF)
Cross-correlation (XCF)
ARMA model
ARIMA model (Box–Jenkins)
Autoregressive conditional heteroskedasticity (ARCH)
Vector autoregression (VAR) (Autoregressive model (AR))
Spectral density estimation
Fourier analysis
Least-squares spectral analysis
Wavelet
Whittle likelihood
Kaplan–Meier estimator (product limit)
Proportional hazards models
Accelerated failure time (AFT) model
First hitting time
Nelson–Aalen estimator
Log-rank test
Bioinformatics
Clinical trials / studies

Epidemiology
Medical statistics
Chemometrics
Methods engineering
Probabilistic design
Process / quality control
Reliability
System identification
Actuarial science
Census
Crime statistics
Demography
Econometrics
Jurimetrics
National accounts
Official statistics
Population statistics
Psychometrics
Cartography
Environmental statistics
Geographic information system
Geostatistics
Kriging
Category
Mathematics portal
Commons
WikiProject
Mathematics
United States
France
BnF data
Japan
Israel
Yale LUX