STATA GLOSSARY AND INDEX

RELEASE 14

A Stata Press Publication
StataCorp LLC
College Station, Texas

Published by Stata Press, 4905 Lakeway Drive, College Station, Texas 77845
Typeset in TEX

ISBN-10: 1-59718-156-0
ISBN-13: 978-1-59718-156-3

The suggested citation for this software is

StataCorp. 2015. *Stata: Release 14*. Statistical Software. College Station, TX: StataCorp LLC.

Contents

Combined subject table of contents

This is the complete contents for all manuals. Every estimation command has a postestimation entry; however, not all postestimation entries are listed here.

Getting started

Data manipulation and management

Basic data commands

Creating and dropping variables

Functions and expressions

Strings

Dates and times

Loading, saving, importing, and exporting data

Combining data

Reshaping datasets

Labeling, display formats, and notes

Changing and renaming variables

Examining data

File manipulation

Miscellaneous data commands

Multiple imputation

Utilities

Basic utilities

Error messages

Stored results

Internet

Data types and memory

Advanced utilities

Graphics

Common graphs

Distributional graphs

Item response theory graphs

Multivariate graphs

Quality control

Regression diagnostic plots

ROC analysis

Smoothing and densities

Survival-analysis graphs

Time-series graphs

More statistical graphs

Editing

Graph utilities

Graph schemes

Graph concepts

Statistics

ANOVA and related

Basic statistics

Bayesian analysis

Binary outcomes

Categorical outcomes

Censored and truncated regression models

Cluster analysis

Correspondence analysis

Count outcomes

Discriminant analysis

Do-it-yourself generalized method of moments

Do-it-yourself maximum likelihood estimation

Endogenous covariates

Epidemiology and related

Also see *Treatment effects*

Estimation related

Exact statistics

Factor analysis and principal components

Fractional outcomes

Generalized linear models

Indicator and categorical variables

Item response theory

Linear regression and related

Logistic and probit regression

Longitudinal data/panel data

Mixed models

Multidimensional scaling and biplots

Multilevel mixed-effects models

Multiple imputation

Multivariate analysis of variance and related techniques

Nonlinear regression

Nonparametric statistics

Ordinal outcomes

Other statistics

Pharmacokinetic statistics

Power and sample size

Quality control

ROC analysis

Rotation

Sample selection models

Simulation/resampling

Standard postestimation tests, tables, and other analyses

Structural equation modeling

Survey data

Survival analysis

Also see *Power and sample size*

Time series, multivariate

Time series, univariate

Transforms and normality tests

Treatment effects

Matrix commands

Basics

Programming

Other

Mata

Programming

Basics

Program control

Parsing and program arguments

Console output

Commonly used programming commands

Debugging

Advanced programming commands

Special-interest programming commands

Projects

File formats

Mata

Interface features

Acronym glossary

2SIV	two-step instrumental variables
2SLS	two-stage least squares
3SLS	three-stage least squares
ADF	asymptotic distribution free
AF	attributable fraction for the population
AFE	attributable fraction among the exposed
AFT	accelerated failure time
AIC	Akaike information criterion
AIDS	almost-ideal demand system
AIPW	augmented inverse-probability weights
ANCOVA	analysis of covariance
ANOVA	analysis of variance
APE	average partial effects
AR	autoregressive
AR(1)	first-order autoregressive
ARCH	autoregressive conditional heteroskedasticity
ARFIMA	autoregressive fractionally integrated moving average
ARIMA	autoregressive integrated moving average
ARMA	autoregressive moving average
ARMAX	autoregressive moving-average exogenous
ASCII	American Standard Code for Information Interchange
ASE	asymptotic standard error
ASL	achieved significance level
ATE	average treatment effect
ATET	average treatment effect on the treated
AUC	area under the time-versus-concentration curve
BC	bias corrected
BCa	bias-corrected and accelerated
BCC	boundary characteristic curve
BE	between effects
BFGS	Broyden–Fletcher–Goldfarb–Shanno
BHHH	Berndt–Hall–Hall–Hausman
BIC	Bayesian information criterion
BLOB	binary large object
BLUP	best linear unbiased prediction
BRR	balanced repeated replication

CA	correspondence analysis
CCC	category characteristic curve
CCI	conservative confidence interval
CCT	controlled clinical trial
CD	coefficient of determination
CDC	Centers for Disease Control and Prevention
CDF	cumulative distribution function
CES	constant elasticity of substitution
CFA	confirmatory factor analysis
CFI	comparative fit index
CI	conditional independence
CI	confidence interval
CIF	cumulative incidence function
CMI	conditional mean independence
CMLE	conditional maximum likelihood estimates
ct	count time
cusum	cumulative sum
c.v.	coefficient of variation

DA	data augmentation
DDF	denominator degrees of freedom
DDFs	multiple denominator degrees of freedom
DEFF	design effect
DEFT	design effect (standard deviation metric)
DF	dynamic factor
df / d.f.	degree(s) of freedom
d.f.	distribution function
DFAR	dynamic factors with vector autoregressive errors
DFP	Davidon–Fletcher–Powell
DPD	dynamic panel data

EBCDIC	extended binary coded decimal interchange code
EGARCH	exponential GARCH
EGLS	estimated generalized least squares
EIM	expected information matrix
EM	expectation maximization
EPS	Encapsulated PostScript
ESS	error sum of squares
ESS	effective sample size

FCS	fully conditional specification
FD	first-differenced estimator
FDA	Food and Drug Administration
FE	fixed effects
FEVD	forecast-error variance decomposition
FGLS	feasible generalized least squares
FGNLS	feasible generalized nonlinear least squares
FIML	full information maximum likelihood
FIVE estimator	full-information instrumental-variables efficient estimator
flong	full long
flongsep	full long and separate
FMI	fraction of missing information
FP	fractional polynomial
FPC	finite population correction
GARCH	generalized autoregressive conditional heteroskedasticity
GEE	generalized estimating equations
GEV	generalized extreme value
GHK	Geweke–Hajivassiliou–Keane
GHQ	Gauss–Hermite quadrature
GLIM	generalized linear interactive modeling
GLLAMM	generalized linear latent and mixed models
GLM	generalized linear models
GLS	generalized least squares
GMM	generalized method of moments
GPCM	generalized partial credit model
GRM	graded response model
GSEM	generalized structural equation modeling/model
GUI	graphical user interface
HAC	heteroskedasticity- and autocorrelation-consistent
HR	hazard ratio
HRF	human readable form

IC	information criteria
ICC	item characteristic curve
ICD-9	International Classification of Diseases, Ninth Revision
ICD-10	International Classification of Diseases, Tenth Revision
ICU	International Components for Unicode
IIA	independence of irrelevant alternatives
i.i.d.	independent and identically distributed
IIF	item information function
IPW	inverse-probability weighting
IPWRA	inverse-probability-weighted regression adjustment
IQR	interquartile range
IR	incidence rate
IRF	impulse–response function
IRLS	iterated, reweighted least squares
IRR	incidence-rate ratio
IRT	item response theory
IV	instrumental variables
JAR	Java Archive file
JCA	joint correspondence analysis
JRE	Java Runtime Environment
LAPACK	linear algebra package
LAV	least absolute value
LDA	linear discriminant analysis
LIML	limited-information maximum likelihood
LM	Lagrange multiplier
LOO	leave one out
LOWESS	locally weighted scatterplot smoothing
LR	likelihood ratio
LSB	least-significant byte
MA	moving average
MAD	median absolute deviation
MANCOVA	multivariate analysis of covariance
MANOVA	multivariate analysis of variance
MAR	missing at random
MCA	multiple correspondence analysis
MCAGHQ	mode-curvature adaptive Gauss–Hermite quadrature
MCAR	missing completely at random
MCE	Monte Carlo error
MCMC	Markov chain Monte Carlo
MCSE	MCMC standard errors
MDES	minimum detectable effect size
MDS	multidimensional scaling
ME	multiple equation

MEFF	misspecification effect
MEFT	misspecification effect (standard deviation metric)
MFP	multivariable fractional polynomial
MI / mi	multiple imputation
midp	mid-p-value
MIMIC	multiple indicators and multiple causes
MINQUE	minimum norm quadratic unbiased estimation
MIVQUE	minimum variance quadratic unbiased estimation
ML	maximum likelihood
MLE	maximum likelihood estimate
MLMV	maximum likelihood with missing values
mlong	marginal long
MM	method of moments
MNAR	missing not at random
MNP	multinomial probit
MPL	modified profile likelihood
MS	mean square
MSAR	Markov-switching autoregression
MSB	most-significant byte
MSDR	Markov-switching dynamic regression
MSE	mean squared error
MSL	maximum simulated likelihood
MSS	model sum of squares
MUE	median unbiased estimates
MVAGHQ	mean–variance adaptive Gauss–Hermite quadrature
MVN	multivariate normal
MVREG	multivariate regression
NARCH	nonlinear ARCH
NHANES	National Health and Nutrition Examination Survey
NLS	nonlinear least squares
NPARCH	nonlinear power ARCH
NR	Newton–Raphson
NRM	nominal response model
ODBC	Open DataBase Connectivity
OIM	observed information matrix
OIRF	orthogonalized impulse–response function
OLE	Object Linking and Embedding (Microsoft product)
OLS	ordinary least squares
OPG	outer product of the gradient
OR	odds ratio

PA	population averaged
PARCH	power ARCH
PCA	principal component analysis
PCM	partial credit model
PCSE	panel-corrected standard error
p.d.f.	probability density function
PF	prevented fraction for the population
PFE	prevented fraction among the exposed
PH	proportional hazards
pk	pharmacokinetic data
p.m.f.	probability mass function
PMM	predictive mean matching
PNG	Portable Network Graphics
POM	potential-outcome means
PSS	power and sample size
PSU	primary sampling unit
QDA	quadratic discriminant analysis
QML	quasimaximum likelihood
RA	regression adjustment
rc	return code
RCT	randomized controlled trial
RE	random effects
REML	restricted (or residual) maximum likelihood
RESET	regression specification-error test
RMSE	root mean squared error
RMSEA	root mean squared error of approximation
RNG	random-number generator
ROC	receiver operating characteristic
ROP	rank-ordered probit
ROT	rule of thumb
RR	relative risk
RRR	relative-risk ratio
RSM	rating scale model
RSS	residual sum of squares
RUM	random utility maximization
RVI	relative variance increase

SAARCH	simple asymmetric ARCH
SARIMA	seasonal ARIMA
s.d.	standard deviation
SE / s.e.	standard error
SEM	structural equation modeling/model
SF	static factor
SFAR	static factors with vector autoregressive errors
SIF	Stata internal form
SIR	standardized incidence ratio
SJ	Stata Journal
SMCL	Stata Markup and Control Language
SMR	standardized mortality/morbidity ratio
SMSA	standard metropolitan statistical area
SOR	standardized odds ratio
SQL	Structured Query Language
SRD	standardized rate difference
SRMR	standardized root mean squared residual
SRR	standardized risk ratio
SRS	simple random sample/sampling
SRSWR	SRS with replacement
SSC	Statistical Software Components
SSCP	sum of squares and cross products
SSD	summary statistics data
SSU	secondary sampling unit
st	survival time
STB	Stata Technical Bulletin
STS	structural time series
SUR	seemingly unrelated regression
SURE	seemingly unrelated regression estimation
SUTVA	stable unit treatment value assumption
SVAR	structural vector autoregressive model
SVD	singular value decomposition
TAR	target acceptance rate
TARCH	threshold ARCH
TCC	test characteristic curve
TDT	transmission/disequilibrium test
TIF	test information function
TIFF	tagged image file format
TLI	Tucker–Lewis index
TSS	total sum of squares

UCA	Unicode Collation Algorithm
UCM	unobserved-components model
UI	user interface
UTF-8	Universal character set + Transformation Format—8-bit
VAR	vector autoregressive model
VAR(1)	first-order vector autoregressive
VARMA	vector autoregressive moving average
VARMA(1,1)	first-order vector autoregressive moving average
VCE	variance–covariance estimate
VECM	vector error-correction model
VIF	variance inflation factor
WLC	worst linear combination
WLF	worst linear function
WLS	weighted least squares
WNLS	weighted nonlinear least squares
wrt	with respect to
XML	Extensible Markup Language
ZINB	zero-inflated negative binomial
ZIP	zero-inflated Poisson
ZTNB	zero-truncated negative binomial
ZTP	zero-truncated Poisson

Glossary

1PL. See *one-parameter logistic model*.

1 : M matched case–control study. See *matched study*.

2PL. See *two-parameter logistic model*.

2 × 2 contingency table. A 2 × 2 contingency table is used to describe the association between a binary independent variable and a binary response variable of interest.

2 × 2 × K contingency table. See *stratified 2 × 2 tables*.

3PL. See *three-parameter logistic model*.

100% sample. See *census*.

a posteriori. In the context of Bayesian analysis, we use a posteriori to mean "after the sample is observed". For example, a posteriori information is any information obtained after the data sample is observed. See *posterior distribution, posterior*.

a priori. In the context of Bayesian analysis, we use a priori to mean "before the sample is observed". For example, a priori information is any information obtained before the data sample is observed. In a Bayesian model, a priori information about model parameters is specified by prior distributions.

ability. See *latent trait*.

accelerated failure-time model. A model in which everyone has, in a sense, the same survivor function, $S(\tau)$, and an individual's τ_j is a function of his or her characteristics and of time, such as $\tau_j = t * \exp(\beta_0 + \beta_1 x_{1j} + \beta_2 x_{2j})$.

acceptance rate. In the context of the MH algorithm, acceptance rate is the fraction of the proposed samples that is accepted. The optimal acceptance rate depends on the properties of the target distribution and is not known in general. If the target distribution is normal, however, the optimal acceptance rate is known to be 0.44 for univariate distributions and 0.234 for multivariate distributions.

acceptance region. In hypothesis testing, an acceptance region is a set of sample values for which the null hypothesis cannot be rejected or can be accepted. It is the complement of the rejection region.

accrual period or **recruitment period** or **accrual**. The accrual period (or recruitment period) is the period during which subjects are being enrolled (recruited) into a study. Also see *follow-up period*.

actual alpha, **actual significance level**. This is an attained or observed significance level.

adaptation. In the context of the MH algorithm, adaptation refers to the process of tuning or adapting the proposal distribution to optimize the MCMC sampling. Typically, adaptation is performed periodically during the MCMC sampling. The `bayesmh` command performs adaptation every # of iterations as specified in option `adaptation(every(#))` for a maximum of `adaptation(maxiter())` iterations. In a continuous-adaptation regimes, the adaptation lasts during the entire process of the MCMC sampling. See [BAYES] **bayesmh**.

adaptation period. Adaptation period includes all MH adaptive iterations. It equals the length of the adaptation interval, as specified by `adaptation(every())`, times the maximum number of adaptations, `adaptation(maxiter())`.

adaptive iteration. In the adaptive MH algorithm, adaptive iterations are iterations during which adaptation is performed.

add factor. An add factor is a quantity added to an endogenous variable in a forecast model. Add factors can be used to incorporate outside information into a model, and they can be used to produce forecasts under alternative scenarios.

ADF, method(adf). ADF stands for asymptotic distribution free and is a method used to obtain fitted parameters for standard linear SEMs. ADF is used by `sem` when option `method(adf)` is specified. Other available methods are ML, QML, and MLMV.

administrative censoring. Administrative censoring is the right-censoring that occurs when the study observation period ends. All subjects complete the course of the study and are known to have experienced one of two outcomes at the end of the study: survival or failure. This type of censoring should not be confused with *withdrawal* and *loss to follow-up*. Also see *censored, censoring, left-censoring, and right-censoring*.

AFT, accelerated failure time. See *accelerated failure-time model*.

agglomerative hierarchical clustering methods. Agglomerative hierarchical clustering methods are bottom-up methods for hierarchical clustering. Each observation begins in a separate group. The closest pair of groups is agglomerated or merged in each iteration until all the data are in one cluster. This process creates a hierarchy of clusters. Contrast to *divisive hierarchical clustering methods*.

AIPW estimator. See *augmented inverse-probability-weighted estimator*.

Akaike information criterion, AIC. Akaike information criterion (AIC) is an information-based model-selection criterion. It is given by the formula $-2 \times \text{log likelihood} + 2k$, where k is the number of parameters. AIC favors simpler models by penalizing for the number of model parameters. It does not, however, account for the sample size. As a result, the AIC penalization diminishes as the sample size increases, as does its ability to guard against overparameterization.

allocation ratio. This ratio n_2/n_1 represents the number of subjects in the comparison, experimental group relative to the number of subjects in the reference, control group. Also see [PSS] **unbalanced designs**.

alpha. Alpha, α, denotes the significance level.

alternative hypothesis. In hypothesis testing, the alternative hypothesis represents the counterpoint to which the null hypothesis is compared. When the parameter being tested is a scalar, the alternative hypothesis can be either one sided or two sided.

alternative value, alternative parameter. This value of the parameter of interest under the alternative hypothesis is fixed by the investigator in a power and sample-size analysis. For example, alternative mean value and alternative mean refer to a value of the mean parameter under the alternative hypothesis.

analysis of variance, ANOVA. This is a class of statistical models that studies differences between means from multiple populations by partitioning the variance of the continuous outcome into independent sources of variation due to effects of interest and random variation. The test statistic is then formed as a ratio of the expected variation due to the effects of interest to the expected random variation. Also see *one-way ANOVA*, *two-way ANOVA*, *one-way repeated-measures ANOVA*, and *two-way repeated-measures ANOVA*.

analysis time. Analysis time is like time, except that 0 has a special meaning: $t = 0$ is the time of onset of risk, the time when failure first became possible.

Analysis time is usually not what is recorded in a dataset. A dataset of patients might record calendar time. Calendar time must then be mapped to analysis time.

The letter t is reserved for time in analysis-time units. The term *time* is used for time measured in other units.

The *origin* is the *time* corresponding to $t = 0$, which can vary subject to subject. Thus $t =$ *time* − *origin*.

anchoring, **anchor variable**. A variable is said to be the anchor of a latent variable if the path coefficient between the latent variable and the anchor variable is constrained to be 1. `sem` and `gsem` use anchoring as a way of normalizing latent variables and thus identifying the model.

ANOVA denominator degrees of freedom (DDF) method. This method uses the traditional ANOVA for computing DDF. According to this method, the DDF for a test of a fixed effect of a given variable depends on whether that variable is also included in any of the random-effects equations. For traditional ANOVA models with balanced designs, this method provides exact sampling distributions of the test statistics. For more complex mixed-effects models or with unbalanced data, this method typically leads to poor approximations of the actual sampling distributions of the test statistics.

anti-image correlation matrix or **anti-image covariance matrix**. The image of a variable is defined as that part which is predictable by regressing each variable on all the other variables; hence, the anti-image is the part of the variable that cannot be predicted. The anti-image correlation matrix $\mathbf{A}$ is a matrix of the negatives of the partial correlations among variables. Partial correlations represent the degree to which the factors explain each other in the results. The diagonal of the anti-image correlation matrix is the Kaiser–Meyer–Olkin measure of sampling adequacy for the individual variables. Variables with small values should be eliminated from the analysis. The anti-image covariance matrix $\mathbf{C}$ contains the negatives of the partial covariances and has one minus the squared multiple correlations in the principal diagonal. Most of the off-diagonal elements should be small in both anti-image matrices in a good factor model. Both anti-image matrices can be calculated from the inverse of the correlation matrix $\mathbf{R}$ via

$$\mathbf{A} = \{\text{diag}(\mathbf{R})\}^{-1}\mathbf{R}\{\text{diag}(\mathbf{R})\}^{-1}$$
$$\mathbf{C} = \{\text{diag}(\mathbf{R})\}^{-1/2}\mathbf{R}\{\text{diag}(\mathbf{R})\}^{-1/2}$$

Also see *Kaiser–Meyer–Olkin measure of sampling adequacy*.

approximation denominator degrees of freedom (DDF) methods. The Kenward–Roger and Satterthwaite DDF methods are referred to as approximation methods because they approximate the sampling distributions of test statistics using t and F distributions with the DDF specific to the method for complicated mixed-effects models and for simple mixed models with unbalanced data. Also see *exact denominator degrees of freedom (DDF) methods*.

arbitrary missing pattern. Any missing-value pattern. Some imputation methods are suitable only when the pattern of missing values is special, such as a monotone-missing pattern. An imputation method suitable for use with an arbitrary missing pattern may be used regardless of the pattern.

ARCH model. An autoregressive conditional heteroskedasticity (ARCH) model is a regression model in which the conditional variance is modeled as an autoregressive (AR) process. The ARCH(m) model is

$$y_t = \mathbf{x}_t\boldsymbol{\beta} + \epsilon_t$$
$$E(\epsilon_t^2|\epsilon_{t-1}^2, \epsilon_{t-2}^2, \ldots) = \alpha_0 + \alpha_1\epsilon_{t-1}^2 + \cdots + \alpha_m\epsilon_{t-m}^2$$

where ϵ_t is a white-noise error term. The equation for y_t represents the conditional mean of the process, and the equation for $E(\epsilon_t^2|\epsilon_{t-1}^2, \epsilon_{t-2}^2, \ldots)$ specifies the conditional variance as an autoregressive function of its past realizations. Although the conditional variance changes over time, the unconditional variance is time invariant because y_t is a stationary process. Modeling the conditional variance as an AR process raises the implied unconditional variance, making this model particularly appealing to researchers modeling fat-tailed data, such as financial data.

Arellano–Bond estimator. The Arellano–Bond estimator is a generalized method of moments (GMM) estimator for linear dynamic panel-data models that uses lagged levels of the endogenous variables as well as first differences of the exogenous variables as instruments. The Arellano–Bond estimator removes the panel-specific heterogeneity by first-differencing the regression equation.

ARFIMA model. An autoregressive fractionally integrated moving-average (ARFIMA) model is a time-series model suitable for use with long-memory processes. ARFIMA models generalize autoregressive integrated moving-average (ARIMA) models by allowing the differencing parameter to be a real number in $(-0.5, 0.5)$ instead of requiring it to be an integer.

arguments. The values a function receives are called the function's arguments. For instance, in `lud(`A`, `L`, `U`)`, A, L, and U are the arguments.

ARIMA model. An autoregressive integrated moving-average (ARIMA) model is a time-series model suitable for use with integrated processes. In an ARIMA(p, d, q) model, the data is differenced d times to obtain a stationary series, and then an ARMA(p, q) model is fit to this differenced data. ARIMA models that include exogenous explanatory variables are known as ARMAX models.

ARMA model. An autoregressive moving-average (ARMA) model is a time-series model in which the current period's realization is the sum of an autoregressive (AR) process and a moving-average (MA) process. An ARMA(p, q) model includes p AR terms and q MA terms. ARMA models with just a few lags are often able to fit data as well as pure AR or MA models with many more lags.

ARMAX model. An ARMAX model is a time-series model in which the current period's realization is an ARMA process plus a linear function of a set of exogenous variables. Equivalently, an ARMAX model is a linear regression model in which the error term is specified to follow an ARMA process.

array. An array is any indexed object that holds other objects as elements. Vectors are examples of 1-dimensional arrays. Vector $\mathbf{v}$ is an array, and $\mathbf{v}[1]$ is its first element. Matrices are 2-dimensional arrays. Matrix $\mathbf{X}$ is an array, and $\mathbf{X}[1, 1]$ is its first element. In theory, one can have 3-dimensional, 4-dimensional, and higher arrays, although Mata does not directly provide them. See [M-2] **subscripts** for more information on arrays in Mata.

Arrays are usually indexed by sequential integers, but in associative arrays, the indices are strings that have no natural ordering. Associative arrays can be 1-dimensional, 2-dimensional, or higher. If A were an associative array, then $A[\text{“first”}]$ might be one of its elements. See [M-5] **asarray()** for associative arrays in Mata.

ASCII. ASCII stands for American Standard Code for Information Interchange. It is a way of representing text and the characters that form text in computers. It can be divided into two sections: plain, or lower, ASCII, which includes numbers, punctuation, plain letters without diacritical marks, whitespace characters such as space and tab, and some control characters such as carriage return; and extended ASCII, which includes letters with diacritical marks as well as other special characters.

Before Stata 14, datasets, do-files, ado-files, and other Stata files were encoded using ASCII.

at risk. A subject is at risk from the instant the first failure event becomes possible and usually stays that way until failure, but a subject can have periods of being at risk and not at risk.

ATE. See *average treatment effect*.

ATET. See *average treatment effect on the treated*.

attributable fraction. An attributable fraction is the reduction in the risk of a disease or other condition of interest when a particular risk factor is removed.

augmented inverse-probability-weighted estimator. An augmented inverse-probability-weighted (AIPW) estimator is an inverse-probability-weighted estimator that includes an augmentation term that corrects the estimator when the treatment model is misspecified. When the treatment is correctly specified, the augmentation term vanishes as the sample size becomes large. An AIPW estimator uses both an outcome model and a treatment model and is a doubly robust estimator.

augmented regression. Regression performed on the augmented data, the data with a few extra observations with small weights. The data are augmented in a way that prevents perfect prediction, which may arise during estimation of categorical data. See *The issue of perfect prediction during imputation of categorical data* under *Remarks and examples* of [MI] **mi impute**.

autocorrelation function. The autocorrelation function (ACF) expresses the correlation between periods t and $t-k$ of a time series as function of the time t and the lag k. For a stationary time series, the ACF does not depend on t and is symmetric about $k=0$, meaning that the correlation between periods t and $t-k$ is equal to the correlation between periods t and $t+k$.

autoregressive process. An autoregressive process is a time-series model in which the current value of a variable is a linear function of its own past values and a white-noise error term. A first-order autoregressive process, denoted as an AR(1) process, is $y_t = \rho y_{t-1} + \epsilon_t$. An AR($p$) model contains p lagged values of the dependent variable.

An autoregressive processes can be extended to panel data. An AR(1) process in this is $y_{it} = \rho y_{i,t-1} + \epsilon_{it}$, where i denotes panels, t denotes time, and ϵ_{it} is white noise. In some applications, the parameter ρ is written as ρ_i and is allowed to differ across panels.

average treatment effect. The average treatment effect is the average among all individuals in a population.

average treatment effect on the treated. The average treatment effect on the treated is the average among those individuals who actually get the treatment.

average-linkage clustering. Average-linkage clustering is a hierarchical clustering method that uses the average proximity of observations between groups as the proximity measure between the two groups.

balanced data. A longitudinal or panel dataset is said to be balanced if each panel has the same number of observations. See also *weakly balanced* and *strongly balanced*.

balanced design. A balanced design represents an experiment in which the numbers of treated and untreated subjects are equal. For many types of two-sample hypothesis tests, the power of the test is maximized with balanced designs.

balanced repeated replication. Balanced repeated replication (BRR) is a method of variance estimation for designs with two PSUs in every stratum. The BRR variance estimator tends to give more reasonable variance estimates for this design than does the linearized variance estimator, which can result in large values and undesirably wide confidence intervals. The BRR variance estimator is described in [SVY] **variance estimation**.

band-pass filter. Time-series filters are designed to pass or block stochastic cycles at specified frequencies. Band-pass filters, such as those implemented in `tsfilter bk` and `tsfilter cf`, pass through stochastic cycles in the specified range of frequencies and block all other stochastic cycles.

baseline. In survival analysis, baseline is the state at which the covariates, usually denoted by the row vector $\mathbf{x}$, are zero. For example, if the only measured covariate is systolic blood pressure, the baseline survivor function would be the survivor function for someone with zero systolic blood pressure. This may seem ridiculous, but covariates are usually centered so that the mathematical definition of baseline (covariate is zero) translates into something meaningful (mean systolic blood pressure).

baseline model. A baseline model is a covariance model—a model of fitted means and covariances of observed variables without any other paths—with most of the covariances constrained to 0. That is, a baseline model is a model of fitted means and variances but typically not all the covariances. Also see *saturated model*. Baseline models apply only to standard linear SEMs.

batch means. Batch means are means obtained from batches of sample values of equal size. Batch means provide an alternative method for estimating MCMC standard errors (MCSE). The batch size is usually chosen to minimize the correlation between different batches of means.

Bayes factor. Bayes factor is given by the ratio of the marginal likelihoods of two models, M_1 and M_2. It is a widely used criterion for Bayesian model comparison. Bayes factor is used in calculating the posterior odds ratio of model M_1 versus M_2,

$$\frac{P(M_1|\mathbf{y})}{P(M_2|\mathbf{y})} = \frac{P(\mathbf{y}|M_1)}{P(\mathbf{y}|M_2)}\frac{P(M_1)}{P(M_2)}$$

where $P(M_i|\mathbf{y})$ is a posterior probability of model M_i, and $P(M_i)$ is a prior probability of model M_i. When the two models are equally likely, that is, when $P(M_1) = P(M_2)$, the Bayes factor equals the posterior odds ratio of the two models.

Bayes's rule. The Bayes's rule is a formal method for relating conditional probability statements. For two (random) events X and Y, the Bayes's rule states that

$$P(X|Y) \propto P(Y|X)P(X)$$

that is, the probability of X conditional on Y is proportional to the probability of X and the probability of Y conditional on X. In Bayesian analysis, the Bayes's rule is used for combining prior information about model parameters and evidence from the observed data to form the posterior distribution.

Bayes's theorem. Bayes's theorem states that the probability of an event, A, conditional on another event, B, is generally different from the probability of B conditional on A, although the two are related. Bayes's theorem is that

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

where $P(A)$ is the marginal probability of A, and $P(A|B)$ is the conditional probability of A given B, and likewise for $P(B)$ and $P(B|A)$.

Bayesian analysis. Bayesian analysis is a statistical methodology that considers model parameters to be random quantities and estimates their posterior distribution by combining prior knowledge about parameters with the evidence from the observed data sample. Prior knowledge about parameters is described by prior distributions and evidence from the observed data is incorporated through a likelihood model. Using the Bayes's rule, the prior distribution and the likelihood model are combined to form the posterior distribution of model parameters. The posterior distribution is then used for parameter inference, hypothesis testing, and prediction.

Bayesian hypothesis testing. Bayesian hypothesis testing computes probabilities of hypotheses conditional on the observed data. In contrast to the frequentist hypothesis testing, the Bayesian hypothesis testing computes the actual probability of a hypothesis H by using the Bayes's rule,

$$P(H|\mathbf{y}) \propto P(\mathbf{y}|H)P(H)$$

where $\mathbf{y}$ is the observed data, $P(\mathbf{y}|H)$ is the marginal likelihood of $\mathbf{y}$ given H, and $P(H)$ is the prior probability of H. Two different hypotheses, H_1 and H_2, can be compared by simply comparing $P(H_1|\mathbf{y})$ to $P(H_2|\mathbf{y})$.

Bayesian information criterion, BIC. The Bayesian information criterion (BIC), also known as Schwarz criterion, is an information based criterion used for model selection in classical statistics. It is given by the formula $-0.5 \times \log$ likelihood $+ k \times \ln n$, where k is the number of parameters and n is the sample size. BIC favors simpler, in terms of complexity, models and it is more conservative than AIC.

BCC. See *boundary characteristic curve*.

Bentler's invariant pattern simplicity rotation. Bentler's (1977) rotation maximizes the invariant pattern simplicity. It is an oblique rotation that minimizes the criterion function

$$c(\mathbf{\Lambda}) = -\log[|(\mathbf{\Lambda}^2)'\mathbf{\Lambda}^2|] + \log[|\text{diag}\{(\mathbf{\Lambda}^2)'\mathbf{\Lambda}^2\}|]$$

See *Crawford–Ferguson rotation* for a definition of $\mathbf{\Lambda}$. Also see *oblique rotation*.

Bentler–Weeks formulation. The Bentler and Weeks (1980) formulation of standard linear SEMs places the results in a series of matrices organized around how results are calculated. See [SEM] **estat framework**.

beta. Beta, β, denotes the probability of committing a type II error, namely, failing to reject the null hypothesis even though it is false.

between estimator. The between estimator is a panel-data estimator that obtains its estimates by running OLS on the panel-level means of the variables. This estimator uses only the between-panel variation in the data to identify the parameters, ignoring any within-panel variation. For it to be consistent, the between estimator requires that the panel-level means of the regressors be uncorrelated with the panel-specific heterogeneity terms.

between matrix and **within matrix**. The between and within matrices are SSCP matrices that measure the spread between groups and within groups, respectively. These matrices are used in multivariate analysis of variance and related hypothesis tests: Wilks's lambda, Roy's largest root, Lawley–Hotelling trace, and Pillai's trace.

Here we have k independent random samples of size n. The between matrix $\mathbf{H}$ is given by

$$\mathbf{H} = n\sum_{i=1}^{k}(\overline{\mathbf{y}}_{i\bullet} - \overline{\mathbf{y}}_{\bullet\bullet})(\overline{\mathbf{y}}_{i\bullet} - \overline{\mathbf{y}}_{\bullet\bullet})' = \sum_{i=1}^{k}\frac{1}{n}\mathbf{y}_{i\bullet}\mathbf{y}_{i\bullet}' - \frac{1}{kn}\mathbf{y}_{\bullet\bullet}\mathbf{y}_{\bullet\bullet}'$$

The within matrix $\mathbf{E}$ is defined as

$$\mathbf{E} = \sum_{i=1}^{k}\sum_{j=1}^{n}(\mathbf{y}_{ij} - \overline{\mathbf{y}}_{i\bullet})(\mathbf{y}_{ij} - \mathbf{y}_{i\bullet})' = \sum_{i=1}^{k}\sum_{j=1}^{n}\mathbf{y}_{ij}\mathbf{y}_{ij}' - \sum_{i=1}^{k}\frac{1}{n}\mathbf{y}_{i\bullet}\mathbf{y}_{i\bullet}'$$

Also see *SSCP matrix*.

between–within denominator degrees of freedom (DDF) method. See *repeated denominator degrees of freedom (DDF) method*.

between-subjects design. This is an experiment that has only between-subjects factors. See [PSS] **power oneway** and [PSS] **power twoway**.

between-subjects factor. This is a factor for which each subject receives only one of the levels.

binary 0. Binary 0, also known as the null character, is traditionally used to indicate the end of a string, such as an ASCII or UTF-8 string.

Binary 0 is obtained by using `char(0)` and is sometimes displayed as `\0`. See [U] **12.4.10 strL variables and binary strings** for more information.

binary item. A binary item is an item that is scored as either 0 or 1.

binary operator. A binary operator is an operator applied to two arguments. In `2-3`, the minus sign is a binary operator, as opposed to the minus sign in `-9`, which is a unary operator.

binary string. A binary string is, technically speaking, any string that does not contain text. In Stata, however, a string is only marked as binary if it contains binary 0, or if it contains the contents of a file read in using the `fileread()` function, or if it is the result of a string expression containing a string that has already been marked as binary.

In Stata, `strL` variables, string scalars, and Mata strings can store binary strings.

See [U] **12.4.10 strL variables and binary strings** for more information.

binomial test. A binomial test is a test for which the exact sampling distribution of the test statistic is binomial; see [R] **bitest**. Also see [PSS] **power oneproportion**.

biplot. A biplot is a scatterplot which represents both observations and variables simultaneously. There are many different biplots; variables in biplots are usually represented by arrows and observations are usually represented by points.

biquartimax rotation or **biquartimin rotation**. Biquartimax rotation and biquartimin rotation are synonyms. They put equal weight on the varimax and quartimax criteria, simplifying the columns and rows of the matrix. This is an oblique rotation equivalent to an oblimin rotation with $\gamma = 0.5$. Also see *varimax rotation*, *quartimax rotation*, and *oblimin rotation*.

bisection method. This method finds a root x of a function $f(x)$ such that $f(x) = 0$ by repeatedly subdividing an interval on which $f(x)$ is defined until the change in successive root estimates is within the requested tolerance and function $f(\cdot)$ evaluated at the current estimate is sufficiently close to zero.

BLOB. BLOB is database jargon for binary large object. In Stata, BLOBs can be stored in `strL`s. Thus `strL`s can contain BLOBs such as Word documents, JPEG images, or anything else. See *strL*.

blocking. In the context of the MH algorithm, blocking refers to the process of separating model parameters into different subsets or blocks to be sampled independently of each other. MH algorithm generates proposals and applies the acceptance–rejection rule sequentially for each block. It is recommended that correlated parameters are kept in one block. Separating less-correlated or independent model parameters in different blocks may improve the mixing of the MH algorithm.

BLUPs. BLUPs are best linear unbiased predictions of either random effects or linear combinations of random effects. In linear models containing random effects, these effects are not estimated directly but instead are integrated out of the estimation. Once the fixed effects and variance components have been estimated, you can use these estimates to predict group-specific random effects. These predictions are called BLUPs because they are unbiased and have minimal mean squared errors among all linear functions of the response.

bootstrap. The bootstrap is a method of variance estimation. The bootstrap variance estimator for survey data is described in [SVY] **variance estimation**.

bootstrap, **vce(bootstrap)**. The bootstrap is a replication method for obtaining variance estimates. Consider an estimation method E for estimating θ. Let $\widehat{\theta}$ be the result of applying E to dataset D containing N observations. The bootstrap is a way of obtaining variance estimates for $\widehat{\theta}$ from repeated estimates $\widehat{\theta}_1, \widehat{\theta}_2, \ldots,$ where each $\widehat{\theta}_i$ is the result of applying E to a dataset of size N drawn with replacement from D. See [SEM] **sem option method( )** and [R] **bootstrap**.

vce(bootstrap) is allowed with sem but not gsem. You can obtain bootstrap results by prefixing the gsem command with bootstrap:, but remember to specify bootstrap's cluster() and idcluster() options if you are fitting a multilevel model. See [SEM] **intro 9**.

boundary characteristic curve. A boundary characteristic curve (BCC) expresses the probability of transitioning across a given boundary threshold that separates the ordered item categories into two groups as a function of the latent trait.

boundary kernel. A boundary kernel is a special kernel used to smooth hazard functions in the boundaries of the data range. Boundary kernels are applied when the epan2, biweight, or rectangle kernel() is specified with stcurve, hazard or sts graph, hazard.

boundary solution or **Heywood solution**. See *Heywood case*.

broad type. Two matrices are said to be of the same broad type if the elements in each are numeric, are string, or are pointers. Mata provides two numeric types, real and complex. The term *broad type* is used to mask the distinction within numeric and is often used when discussing operators or functions. One might say, "The comma operator can be used to join the rows of two matrices of the same broad type," and the implication of that is that one could join a real to a complex. The result would be complex. Also see *type, eltype, and orgtype*.

BRR. See *balanced repeated replication*.

Builder. The Builder is Stata's graphical interface for building sem and gsem models. The Builder is also known as the SEM Builder. See [SEM] **intro 2**, [SEM] **Builder**, and [SEM] **Builder, generalized**.

burn-between period. The number of iterations between two draws of an MCMC sequence such that these draws may be regarded as independent.

burn-in period. The burn-in period is the number of iterations it takes for an MCMC sequence to reach stationarity.

byte. Formally, a byte is eight binary digits (bits), the units used to record computer data. Each byte can also be considered as representing a value from 0 through 255. Do not confuse this with Stata's byte variable storage type, which allows values from −127 to 100 to be stored. With regard to strings, all strings are composed of individual characters that are encoded using either one byte or several bytes to represent each character.

For example, in UTF-8, the encoding system used by Stata, byte value 97 encodes "a". Byte values 195 and 161 in sequence encode "á".

CA. See *correspondence analysis*.

calibration. The procedure of estimating parameters of an IRT model.

canonical correlation analysis. Canonical correlation analysis attempts to describe the relationships between two sets of variables by finding linear combinations of each so that the correlation between the linear combinations is maximized.

canonical discriminant analysis. Canonical linear discriminant analysis is LDA where describing how groups are separated is of primary interest. Also see *linear discriminant analysis*.

canonical link. Corresponding to each family of distributions in a generalized linear model is a canonical link function for which there is a sufficient statistic with the same dimension as the number of parameters in the linear predictor. The use of canonical link functions provides the GLM with desirable statistical properties, especially when the sample size is small.

canonical loadings. The canonical loadings are coefficients of canonical linear discriminant functions. Also see *canonical discriminant analysis* and *loading*.

canonical variate set. The canonical variate set is a linear combination or weighted sum of variables obtained from canonical correlation analysis. Two sets of variables are analyzed in canonical correlation analysis. The first canonical variate of the first variable set is the linear combination in standardized form that has maximal correlation with the first canonical variate from the second variable set. The subsequent canonical variates are uncorrelated to the previous and have maximal correlation under that constraint.

case–control studies. In case–control studies, cases meeting a fixed criterion are matched to noncases ex post to study differences in possible covariates. Relative sample sizes are usually fixed at 1:1 or 1:2 but sometimes vary once the survey is complete. In any case, sample sizes do not reflect the distribution in the underlying population.

casewise deletion. See *listwise deletion*.

categorical item. A categorical item is an item that is either ordinal or nominal.

category boundary curve. See *boundary characteristic curve*.

category boundary location. See *difficulty*.

category characteristic curve. A category characteristic curve (CCC) expresses the probability of a response in a given item category as a function of the latent trait.

category response function. See *category characteristic curve*.

cause-specific hazard. In a competing-risks analysis, the cause-specific hazard is the hazard function that generates the events of a given type. For example, if heart attack and stroke are competing events, then the cause-specific hazard for heart attacks describes the biological mechanism behind heart attacks independently of that for strokes. Cause-specific hazards can be modeled using Cox regression, treating the other events as censored.

CCC. See *category characteristic curve*.

c-conformability. Matrix, vector, or scalar A is said to be c-conformable with matrix, vector, or scalar B if they have the same number of rows and columns (they are *p-conformable*), or if they have the same number of rows and one is a vector, or if they have the same number of columns and one is a vector, or if one or the other is a scalar. c stands for colon; c-conformable matrices are suitable for being used with Mata's :*op* operators. A and B are c-conformable if and only if

A	B
$r \times c$	$r \times c$
$r \times 1$	$r \times c$
$1 \times c$	$r \times c$
1×1	$r \times c$
$r \times c$	$r \times 1$
$r \times c$	$1 \times c$
$r \times c$	1×1

The idea behind c-conformability is generalized elementwise operation. Consider C=A:*B. If A and B have the same number of rows and have the same number of columns, then $||C_{ij}|| = ||A_{ij}*B_{ij}||$. Now say that A is a column vector and B is a matrix. Then $||C_{ij}|| = ||A_i*B_{ij}||$: each element of A is applied to the entire row of B. If A is a row vector, each column of A is applied to the entire column of B. If A is a scalar, A is applied to every element of B. And then all the rules repeat, with the roles of A and B interchanged. See [M-2] **op_colon** for a complete definition.

CCT. See *controlled clinical trial study*.

cell means. These are means of the outcome of interest within cells formed by the cross-classification of the two factors. See [PSS] **power twoway** and [PSS] **power repeated**.

cell-means model. A cell-means model is an ANOVA model formulated in terms of cell means.

censored, **censoring**, **left-censoring**, and **right-censoring**. An observation is left-censored when the exact time of failure is not known; it is merely known that the failure occurred before t_l. Suppose that the event of interest is becoming employed. If a subject is already employed when first interviewed, his outcome is left-censored.

An observation is right-censored when the time of failure is not known; it is merely known that the failure occurred after t_r. If a patient survives until the end of a study, the patient's time of death is right-censored.

In common usage, *censored* without a modifier means right-censoring.

Also see *truncation, left-truncation, and right-truncation*.

census. When a census of the population is conducted, every individual in the population participates in the survey. Because of the time, cost, and other constraints, the data collected in a census are typically limited to items that can be quickly and easily determined, usually through a questionnaire.

centered data. Centered data has zero mean. You can center data $\mathbf{x}$ by taking $\mathbf{x} - \overline{\mathbf{x}}$.

central posterior interval. See *equal-tailed credible interval*.

centroid-linkage clustering. Centroid-linkage clustering is a hierarchical clustering method that computes the proximity between two groups as the proximity between the group means.

CFA, **CFA models**. CFA stands for confirmatory factor analysis. It is a way of analyzing measurement models. CFA models is a synonym for measurement models.

chained equations. See *fully conditional specification*.

chi-squared test, χ^2 **test**. This test for which either an asymptotic sampling distribution or a sampling distribution of a test statistic is χ^2. See [PSS] **power onevariance** and [PSS] **power twoproportions**.

Cholesky ordering. Cholesky ordering is a method used to orthogonalize the error term in a VAR or VECM to impose a recursive structure on the dynamic model, so that the resulting impulse–response functions can be given a causal interpretation. The method is so named because it uses the Cholesky decomposition of the error-covariance matrix.

CI. CI is an abbreviation for confidence interval.

CI assumption. See *conditional-independence assumption*.

CIF. See *cumulative incidence function*.

class programming. See *object-oriented programming*.

classical scaling. Classical scaling is a method of performing MDS via an eigen decomposition. This is contrasted to modern MDS, which is achieved via the minimization of a loss function. Also see *multidimensional scaling* and *modern scaling*.

classification. Classification is the act of allocating or classifying observations to groups as part of discriminant analysis. In some sources, classification is synonymous with cluster analysis.

classification function. Classification functions can be obtained after LDA or QDA. They are functions based on Mahalanobis distance for classifying observations to the groups. See *discriminant function* for an alternative. Also see *linear discriminant analysis* and *quadratic discriminant analysis*.

classification table. A classification table, also known as a confusion matrix, gives the count of observations from each group that are classified into each of the groups as part of a discriminant analysis. The element at (i, j) gives the number of observations that belong to the ith group but were classified into the jth group. High counts are expected on the diagonal of the table where observations are correctly classified, and small values are expected off the diagonal. The columns of the matrix are categories of the predicted classification; the rows represent the actual group membership.

clinical trial. A clinical trials is an experiment testing a medical treatment or procedure on human subjects.

clinically meaningful difference, **clinically meaningful effect**, **clinically significant difference**. Clinically meaningful difference represents the magnitude of an effect of interest that is of clinical importance. What is meant by "clinically meaningful" may vary from study to study. In clinical trials, for example, if no prior knowledge is available about the performance of the considered clinical procedure, a standardized effect size (adjusted for standard deviation) between 0.25 and 0.5 may be considered of clinical importance.

cluster. A cluster is a collection of individuals that are sampled as a group. Although the cost in time and money can be greatly decreased, cluster sampling usually results in larger variance estimates when compared with designs in which individuals are sampled independently.

cluster analysis. Cluster analysis is a method for determining natural groupings or clusters of observations.

cluster tree. See *dendrogram*.

clustered, **vce(cluster clustvar)**. Clustered is the name we use for the generalized Huber/White/sandwich estimator of the VCE, which is the `robust` technique generalized to relax the assumption that errors are independent across observations to be that they are independent across clusters of observations. Within cluster, errors may be correlated.

Clustered standard errors are reported when `sem` or `gsem` option `vce(cluster` *clustvar*`)` is specified. The other available techniques are OIM, OPG, robust, bootstrap, and jackknife. Also available for `sem` only is EIM.

clustering. See *cluster analysis*.

Cochran–Armitage test. The Cochran–Armitage test is a test for a linear trend in a probability of response in a $J \times 2$ contingency table. The test statistic has an asymptotic χ^2 distribution under the null hypothesis. See [PSS] **power trend**.

Cochran–Mantel–Haenszel test. See *Mantel–Haenszel test*.

Cochrane–Orcutt estimator. This estimation is a linear regression estimator that can be used when the error term exhibits first-order autocorrelation. An initial estimate of the autocorrelation parameter ρ is obtained from OLS residuals, and then OLS is performed on the transformed data $\widetilde{y}_t = y_t - \rho y_{t-1}$ and $\widetilde{\mathbf{x}}_t = \mathbf{x}_t - \rho\mathbf{x}_{t-1}$.

code pages. A code page maps extended ASCII values to a set of characters, typically for a specific language or set of languages. For example, the most commonly used code page is Windows-1252, which maps extended ASCII values to characters used in Western European languages. Code pages are essentially encodings for extended ASCII characters.

code point. A code point is the numerical value or position that represents a single character in a text system such as ASCII or Unicode. The original ASCII encoding system contains only 128 code points and thus can represent only 128 characters. Historically, the 128 additional bytes of extended ASCII have been encoded in many different and inconsistent ways to provide additional sets of 128 code points. The formal Unicode specification has 1,114,112 possible code points, of which roughly 250,000 have been assigned to actual characters. Stata uses UTF-8 encoding for Unicode. Note that the UTF-8–encoded version of a code point does not have the same numeric value as the code point itself.

coefficient of determination. The coefficient of determination is the fraction (or percentage) of variation (variance) explained by an equation of a model. The coefficient of determination is thus like R^2 in linear regression.

cohort studies. In cohort studies, a group that is well defined is monitored over time to track the transition of noncases to cases. Cohort studies differ from incidence studies in that they can be retrospective as well as prospective.

cointegrating vector. A cointegrating vector specifies a stationary linear combination of nonstationary variables. Specifically, if each of the variables $x_1, x_2, \dots, x_k$ is integrated of order one and there exists a set of parameters $\beta_1, \beta_2, \dots, \beta_k$ such that $z_t = \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k$ is a stationary process, the variables $x_1, x_2, \dots, x_k$ are said to be cointegrated, and the vector $\boldsymbol{\beta}$ is known as a cointegrating vector.

colon operators. Colon operators are operators preceded by a colon, and the colon indicates that the operator is to be performed elementwise. A:*B indicates element-by-element multiplication, whereas A*B indicates matrix multiplication. Colons may be placed in front of any operator. Usually one thinks of elementwise as meaning $c_{ij} = a_{ij}$ *<op>* b_{ij}, but in Mata, elementwise is also generalized to include c-conformability. See [M-2] **op_colon**.

column stripes. See *row and column stripes*.

column-major order. Matrices are stored as vectors. Column-major order specifies that the vector form of a matrix is created by stacking the columns. For instance,

```
: A
        1   2
     +---------+
  1  |  1   4  |
  2  |  2   5  |
  3  |  3   6  |
     +---------+
```

is stored as

```
        1   2   3   4   5   6
     +-------------------------+
  1  |  1   2   3   4   5   6  |
     +-------------------------+
```

in column-major order. The LAPACK functions use column-major order. Mata uses row-major order. See *row-major order*.

colvector. See *vector, colvector, and rowvector*.

command language. Stata's `sem` and `gsem` command provide a way to specify SEMs. The alternative is to use the Builder to draw path diagrams; see [SEM] **intro 2**, [SEM] **Builder**, and [SEM] **Builder, generalized**.

common factors. Common factors are found by factor analysis. They linearly reconstruct the original variables. In factor analysis, reconstruction is defined in terms of prediction of the correlation matrix of the original variables.

common odds ratio. A measure of association in stratified 2×2 tables. It can be viewed as a weighted aggregate of stratum-specific odds ratios.

communality. Communality is the proportion of a variable's variance explained by the common factors in factor analysis. It is also "1 − uniqueness". Also see *uniqueness*.

comparison value. See *alternative value*.

competing risks. Competing risks models are survival-data models in which the failures are generated by more than one underlying process. For example, death may be caused by either heart attack or stroke. There are various methods for dealing with competing risks. One direct way is to duplicate failures for one competing risk as censored observations for the other risk and stratify on the risk type. Another is to directly model the cumulative incidence of the event of interest in the presence of competing risks. The former method uses `stcox` and the latter, `stcrreg`.

complementary log-log regression. Complementary log-log regression is a term for generalized linear response functions that are family Bernoulli, link cloglog. It is used for binary outcome data. Complementary log-log regression is also known in Stata circles as cloglog regression or just cloglog. See *generalized linear response functions*.

complete and incomplete observations. An observation in the $m = 0$ data is said to be complete if no imputed variable in the observation contains soft missing (`.`). Observations that are not complete are said to be incomplete.

complete data. Data that do not contain any missing values.

complete degrees of freedom. The degrees of freedom that would have been used for inference if the data were complete.

complete DF. See *complete degrees of freedom*.

complete-cases analysis. See *listwise deletion*.

completed data. See *imputed data*.

complete-data analysis. The analysis or estimation performed on the complete data, the data for which all values are observed. This term does not refer to analysis or estimation performed on the subset of complete observations. Do not confuse this with completed-data analysis.

completed-data analysis. The analysis or estimation performed on the made-to-be completed (imputed) data. This term does not refer to analysis or estimation performed on the subset of complete observations.

complete-linkage clustering. Complete-linkage clustering is a hierarchical clustering method that uses the farthest pair of observations between two groups to determine the proximity of the two groups.

complex. A matrix is said to be complex if its elements are complex numbers. Complex is one of two numeric types in Stata, the other being real. Complex is generally used to describe how a matrix is stored and not the kind of numbers that happen to be in it: complex matrix Z might happen to contain real numbers. Also see *type, eltype, and orgtype*.

component scores. Component scores are calculated after PCA. Component scores are the coordinates of the original variables in the space of principal components.

compound symmetry. A covariance matrix has a compound-symmetry structure if all the variances are equal and all the covariances are equal. This is a special case of the sphericity assumption.

Comrey's tandem 1 and 2 rotations. Comrey (1967) describes two rotations, the first (tandem 1) to judge which "small" factors should be dropped, the second (tandem 2) for "polishing".

Tandem principle 1 minimizes the criterion

$$c(\mathbf{\Lambda}) = \left\langle \mathbf{\Lambda}^2, (\mathbf{\Lambda}\mathbf{\Lambda}')^2\mathbf{\Lambda}^2 \right\rangle$$

Tandem principle 2 minimizes the criterion

$$c(\mathbf{\Lambda}) = \left\langle \mathbf{\Lambda}^2, \{\mathbf{1}\mathbf{1}' - (\mathbf{\Lambda}\mathbf{\Lambda}')^2\}\mathbf{\Lambda}^2 \right\rangle$$

See *Crawford–Ferguson rotation* for a definition of $\mathbf{\Lambda}$.

concordant pairs. In a 2×2 contingency table, a concordant pair is a pair of observations that are both either successes or failures. Also see *discordant pairs* and *Introduction* under *Remarks and examples* in [PSS] **power pairedproportions**.

condition number. The condition number associated with a numerical problem is a measure of that quantity's amenability to digital computation. A problem with a low condition number is said to be well conditioned, whereas a problem with a high condition number is said to be ill conditioned.

Sometimes reciprocals of condition numbers are reported and yet authors will still refer to them sloppily as condition numbers. Reciprocal condition numbers are often scaled between 0 and 1, with values near `epsilon(1)` indicating problems.

conditional conjugacy. See *semiconjugate prior*.

conditional fixed-effects model. In general, including panel-specific dummies to control for fixed effects in nonlinear models results in inconsistent estimates. For some nonlinear models, the fixed-effect term can be removed from the likelihood function by conditioning on a sufficient statistic. For example, the conditional fixed-effect logit model conditions on the number of positive outcomes within each panel.

conditional hazard function. In the context of mixed- and random-effects survival models, the conditional hazard function is the hazard function computed conditionally on the random effects. Even within the same covariate pattern, the conditional hazard function varies among individuals who belong to different random-effects clusters.

conditional hazard ratio. In the context of mixed- and random effects survival models, the conditional hazard ratio is the ratio of two conditional hazard functions evaluated at different values of the covariates. Unless stated differently, the denominator corresponds to the conditional hazard function at baseline, that is, with all the covariates set to zero.

conditional imputation. Imputation performed using a conditional sample, a restricted part of the sample. Missing values outside the conditional sample are replaced with a conditional constant, the constant value of the imputed variable in the nonmissing observations outside the conditional sample. See *Conditional imputation* under *Remarks and examples* of [MI] **mi impute**.

conditional independence. The assumption that responses are not correlated after controlling for the latent trait.

conditional mean. The conditional mean expresses the average of one variable as a function of some other variables. More formally, the mean of y conditional on $\mathbf{x}$ is the mean of y for given values of $\mathbf{x}$; in other words, it is $E(y|\mathbf{x})$.

A conditional mean is also known as a regression or as a conditional expectation.

conditional normality assumption. See *normality assumption, joint and conditional*.

conditional overdispersion. In a negative binomial mixed-effects model, conditional overdispersion is overdispersion conditional on random effects. Also see *overdispersion*.

conditional variance. Although the conditional variance is simply the variance of a conditional distribution, in time-series analysis the conditional variance is often modeled as an autoregressive process, giving rise to ARCH models.

conditional-independence assumption. The conditional-independence assumption requires that the common variables that affect treatment assignment and treatment-specific outcomes be observable. The dependence between treatment assignment and treatment-specific outcomes can be removed by conditioning on these observable variables.

This assumption is also known as a selection-on-observables assumption because its central tenet is the observability of the common variables that generate the dependence.

configuration. The configuration in MDS is a representation in a low-dimensional (usually 2-dimensional) space with distances in the low-dimensional space approximating the dissimilarities or disparities in high-dimensional space. Also see *multidimensional scaling*, *dissimilarity*, and *disparity*.

configuration plot. A configuration plot after MDS is a (usually 2-dimensional) plot of labeled points showing the low-dimensional approximation to the dissimilarities or disparities in high-dimensional space. Also see *multidimensional scaling*, *dissimilarity*, and *disparity*.

conformability. Conformability refers to row-and-column matching between two or more matrices. For instance, to multiply $A*B$, A must have the same number of columns as B has rows. If that is not true, then the matrices are said to be nonconformable (for multiplication).

Three kinds of conformability are often mentioned in the Mata documentation: *p-conformability*, *c-conformability*, and *r-conformability*.

confounding. In the analysis of epidemiological tables, factor or interaction effects are said to be confounded when the effect of one factor is combined with that of another. For example, the effect of alcohol consumption on esophageal cancer may be confounded with the effects of age, smoking, or both. In the presence of confounding, it is often useful to stratify on the confounded factors that are not of primary interest, in the above example, age and smoking.

confusion matrix. A confusion matrix is a synonym for a classification table after discriminant analysis. See *classification table*.

conjugate. If $z = a + b\mathrm{i}$, the conjugate of z is $\mathrm{conj}(z) = a - b\mathrm{i}$. The conjugate is obtained by reversing the sign of the imaginary part. The conjugate of a real number is the number itself.

conjugate prior. A prior distribution is conjugate for a family of likelihood distributions if the prior and posterior distributions belong to the same family of distributions. For example, the gamma distribution is a conjugate prior for the Poisson likelihood. Conjugacy may provide an efficient way of sampling from posterior distributions and is used in Gibbs sampling.

conjugate transpose. See *transpose*.

constraints. See *parameter constraints*.

containment denominator degrees of freedom (DDF) method. See *ANOVA denominator degrees of freedom (DDF) method*.

continuous parameters. Continuous parameters are parameters with continuous prior distributions.

contrast or **contrasts**. In ANOVA, a contrast in k population means is defined as a linear combination

$$\delta = c_1\mu_1 + c_2\mu_2 + \cdots + c_k\mu_k$$

where the coefficients satisfy

$$\sum_{i=1}^{k} c_i = 0$$

In the multivariate setting (MANOVA), a contrast in k population mean vectors is defined as

$$\boldsymbol{\delta} = c_1\boldsymbol{\mu}_1 + c_2\boldsymbol{\mu}_2 + \cdots c_k\boldsymbol{\mu}_k$$

where the coefficients again satisfy

$$\sum_{i=1}^{k} c_i = 0$$

The univariate hypothesis $\delta = 0$ may be tested with `contrast` (or `test`) after ANOVA. The multivariate hypothesis $\boldsymbol{\delta} = 0$ may be tested with `manovatest` after MANOVA.

control group. A control group comprises subjects that are randomly assigned to a group where they receive no treatment or receives a standard treatment. In hypothesis testing, this is usually a reference group. Also see *experimental group*.

controlled clinical trial study. This is an experimental study in which treatments are assigned to two or more groups of subjects without the randomization.

correlated uniqueness model. A correlated uniqueness model is a kind of measurement model in which the errors of the measurements has a structured correlation. See [SEM] **intro 5**.

correlation structure. A correlation structure is a set of assumptions imposed on the within-panel variance–covariance matrix of the errors in a panel-data model. See [XT] **xtgee** for examples of different correlation structures.

correlogram. A correlogram is a table or graph showing the sample autocorrelations or partial autocorrelations of a time series.

correspondence analysis. Correspondence analysis (CA) gives a geometric representation of the rows and columns of a two-way frequency table. The geometric representation is helpful in understanding the similarities between the categories of variables and associations between variables. CA is calculated by singular value decomposition. Also see *singular value decomposition*.

correspondence analysis projection. A correspondence analysis projection is a line plot of the row and column coordinates after CA. The goal of this graph is to show the ordering of row and column categories on each principal dimension of the analysis. Each principal dimension is represented by a vertical line; markers are plotted on the lines where the row and column categories project onto the dimensions. Also see *correspondence analysis*.

costs. Costs in discriminant analysis are the cost of misclassifying observations.

counterfactual. A counterfactual is an outcome a subject would have obtained had that subject received a different level of treatment. In the binary-treatment case, the counterfactual outcome for a person who received treatment is the outcome that person would have obtained had the person instead not received treatment; similarly, the counterfactual outcome for a person who did not receive treatment is the outcome that person would have obtained had the person received treatment.

Also see *potential outcome*.

count-time data. See *ct data*.

covariance stationarity. A process is covariance stationary if the mean of the process is finite and independent of t, the unconditional variance of the process is finite and independent of t, and the covariance between periods t and $t - s$ is finite and depends on $t - s$ but not on t or s themselves. Covariance-stationary processes are also known as weakly stationary processes.

covariance structure. In a mixed-effects model, covariance structure refers to the variance–covariance structure of the random effects.

covariates. Covariates are the explanatory variables that appear in a model. For instance, if survival time were to be explained by age, sex, and treatment, then those variables would be the covariates. Also see *time-varying covariates*.

covarimin rotation. Covarimin rotation is an orthogonal rotation equivalent to varimax. Also see *varimax rotation*.

Crawford–Ferguson rotation. Crawford–Ferguson (1970) rotation is a general oblique rotation with several interesting special cases.

Special cases of the Crawford–Ferguson rotation include

κ	Special case
0	quartimax / quartimin
$1/p$	varimax / covarimin
$f/(2p)$	equamax
$(f-1)/(p+f-2)$	parsimax
1	factor parsimony

p = number of rows of $\mathbf{A}$.
f = number of columns of $\mathbf{A}$.

Where $\mathbf{A}$ is the matrix to be rotated, $\mathbf{T}$ is the rotation and $\boldsymbol{\Lambda} = \mathbf{AT}$. The Crawford–Ferguson rotation is achieved by minimizing the criterion

$$c(\boldsymbol{\Lambda}) = \frac{1-\kappa}{4}\left\langle \boldsymbol{\Lambda}^2, \boldsymbol{\Lambda}^2(\mathbf{11}' - \mathbf{I})\right\rangle + \frac{\kappa}{4}\left\langle \boldsymbol{\Lambda}^2, (\mathbf{11}' - \mathbf{I})\boldsymbol{\Lambda}^2\right\rangle$$

Also see *oblique rotation*.

credible interval. In Bayesian analysis, the credible interval of a scalar model parameter is an interval from the domain of the marginal posterior distribution of that parameter. Two types of credible intervals are typically used in practice: equal-tailed credible intervals and HPD credible intervals.

credible level. The credible level is a probability level between 0% and 100% used for calculating credible intervals in Bayesian analysis. For example, a 95% credible interval for a scalar parameter is an interval the parameter belongs to with the probability of 95%.

critical region. See *rejection region*.

critical value. In hypothesis testing, a critical value is a boundary of the rejection region.

cross-correlation function. The cross-correlation function expresses the correlation between one series at time t and another series at time $t-k$ as a function of the time t and lag k. If both series are stationary, the function does not depend on t. The function is not symmetric about $k=0$: $\rho_{12}(k) \neq \rho_{12}(-k)$.

crossed variables or **stacked variables**. In CA and MCA crossed categorical variables may be formed from the interactions of two or more existing categorical variables. Variables that contain these interactions are called crossed or stacked variables.

crossed-effects model. A crossed-effects model is a mixed-effects model in which the levels of random effects are not nested. A simple crossed-effects model for cross-sectional time-series data would contain a random effect to control for panel-specific variation and a second random effect to control for time-specific random variation. Rather than being nested within panel, in this model a random effect due to a given time is the same for all panels.

crossed-random effects. See *crossed-effects model*.

crossing variables or **stacking variables**. In CA and MCA, crossing or stacking variables are the existing categorical variables whose interactions make up a crossed or stacked variable.

cross-sectional or **prevalence studies**. Cross-sectional studies sample distributions of healthy and diseased subjects in the population at one point in time.

cross-sectional data. Cross-sectional data refers to data collected over a set of individuals, such as households, firms, or countries sampled from a population at a given point in time.

cross-sectional study. This type of observational study measures various population characteristics at one point in time or over a short period of time. For example, a study of the prevalence of breast cancer in the population is a cross-sectional study.

cross-sectional time-series data. Cross-sectional time-series data is another name for panel data. The term *cross-sectional time-series data* is sometimes reserved for datasets in which a relatively small number of panels were observed over many periods. See also *panel data*.

crude estimate. A crude estimate has not been adjusted for the effects of other variables. Disregarding a stratification variable, for example, yields a crude estimate.

ct data. ct stands for count time. ct data are an aggregate organized like a life table. Each observation records a time, the number known to fail at that time, the number censored, and the number of new entries. See [ST] **ctset**.

cumulative hazard. See *hazard, cumulative hazard, and hazard ratio*.

cumulative incidence estimator. In a competing-risks analysis, the cumulative incidence estimator estimates the cumulative incidence function (CIF). Assume for now that you have one event of interest (type 1) and one competing event (type 2). The cumulative incidence estimator for type 1 failures is then obtained by

$$\widehat{\mathrm{CIF}}_1(t) = \sum_{j:t_j \le t} \widehat{h}_1(t_j)\widehat{S}(t_{j-1})$$

with

$$\widehat{S}(t) = \prod_{j:t_j \le t} \left\{1 - \widehat{h}_1(t_j) - \widehat{h}_2(t_j)\right\}$$

The t_j index the times at which events (of any type) occur, and $\widehat{h}_1(t_j)$ and $\widehat{h}_2(t_j)$ are the cause-specific hazard contributions for type 1 and type 2, respectively. $\widehat{S}(t)$ estimates the probability that you are event free at time t.

The above generalizes to multiple competing events in the obvious way.

cumulative incidence function. In a competing-risks analysis, the cumulative incidence function, or CIF, is the probability that you will observe the event of primary interest before a given time. Formally,

$$\mathrm{CIF}(t) = P(T \le t \text{ and event type of interest})$$

for time-to-failure, T.

cumulative subhazard. See *subhazard, cumulative subhazard, and subhazard ratio*.

curse of dimensionality. The curse of dimensionality is a term coined by Richard Bellman (1961) to describe the problem caused by the exponential increase in size associated with adding extra dimensions to a mathematical space. On the unit interval, 10 evenly spaced points suffice to sample with no more distance than 0.1 between them; however a unit square requires 100 points, and a unit cube requires 1000 points. Many multivariate statistical procedures suffer from the curse of dimensionality. Adding variables to an analysis without adding sufficient observations can lead to imprecision.

curved path. See *path*.

cusum plot, CUSUM plot. The cusum (CUSUM) plot of an MCMC sample is a plot of cumulative sums of the differences between sample values and their overall mean against the iteration number. Cusum plots are useful graphical summaries for detecting early drifts in MCMC samples.

cyclical component. A cyclical component is a part of a time series that is a periodic function of time. Deterministic functions of time are deterministic cyclical components, and random functions of time are stochastic cyclical components. For example, fixed seasonal effects are deterministic cyclical components and random seasonal effects are stochastic seasonal components.

Random coefficients on time inside of periodic functions form an especially useful class of stochastic cyclical components; see [TS] **ucm**.

DA. See *data augmentation*.

data augmentation. An MCMC method used for the imputation of missing data.

data matrix. A dataset containing n observations on k variables in often stored in an $n \times k$ matrix. An observation refers to a row of that matrix; a variable refers to a column. When the rows are observations and the columns are variables, the matrix is called a data matrix.

declarations. Declarations state the *eltype* and *orgtype* of functions, arguments, and variables. In

```
real matrix myfunc(real vector A, complex scalar B)
{
     real scalar i
     ...
}
```

the `real matrix` is a function declaration, the `real vector` and `complex scalar` are argument declarations, and `real scalar i` is a variable declaration. The `real matrix` states the function returns a real matrix. The `real vector` and `complex scalar` state the kind of arguments `myfunc()` expects and requires. The `real scalar i` helps Mata to produce more efficient compiled code.

Declarations are optional, so the above could just as well have read

```
function myfunc(A, B)
{
     ...
}
```

When you omit the function declaration, you must substitute the word `function`.

When you omit the other declarations, `transmorphic matrix` is assumed, which is fancy jargon for a matrix that can hold anything. The advantages of explicit declarations are that they reduce the chances you make a mistake either in coding or in using the function, and they assist Mata in producing more efficient code. Working interactively, most people omit the declarations.

See [M-2] **declarations** for more information.

defective matrix. An $n \times n$ matrix is defective if it does not have n linearly independent eigenvectors.

DEFF and DEFT. DEFF and DEFT are design effects. Design effects compare the sample-to-sample variability from a given survey dataset with a hypothetical SRS design with the same number of individuals sampled from the population.

DEFF is the ratio of two variance estimates. The design-based variance is in the numerator; the hypothetical SRS variance is in the denominator.

DEFT is the ratio of two standard-error estimates. The design-based standard error is in the numerator; the hypothetical SRS with-replacement standard error is in the denominator. If the given survey design is sampled with replacement, DEFT is the square root of DEFF.

degree-of-freedom adjustment. In estimates of variances and covariances, a finite-population degree-of-freedom adjustment is sometimes applied to make the estimates unbiased.

Let's write an estimated variance as $\widehat{\sigma}_{ii}$ and write the "standard" formula for the variance as $\widehat{\sigma}_{ii} = S_{ii}/N$. If $\widehat{\sigma}_{ii}$ is the variance of observable variable x_i, it can readily be proven that S_{ii}/N is a biased estimate of the variances in samples of size N and that $S_{ii}/(N-1)$ is an unbiased estimate. It is usual to calculate variances using $S_{ii}/(N-1)$, which is to say, the "standard" formula has a multiplicative degree-of-freedom adjustment of $N/(N-1)$ applied to it.

If $\widehat{\sigma}_{ii}$ is the variance of estimated parameter β_i, a similar finite-population degree-of-freedom adjustment can sometimes be derived that will make the estimate unbiased. For instance, if β_i is a coefficient from a linear regression, an unbiased estimate of the variance of regression coefficient β_i is $S_{ii}/(N-p-1)$, where p is the total number of regression coefficients estimated excluding the intercept. In other cases, no such adjustment can be derived. Such estimators have no derivable finite-sample properties and one is left only with the assurances provided by its provable asymptotic properties. In such cases, the variance of coefficient β_i is calculated as S_{ii}/N, which can be derived on theoretical grounds. SEM is an example of such an estimator.

SEM is a remarkably flexible estimator and can reproduce results that can sometimes be obtained by other estimators. SEM might produce asymptotically equivalent results, or it might produce identical results depending on the estimator. Linear regression is an example in which `sem` and `gsem` produce the same results as `regress`. The reported standard errors, however, will not look identical because the linear-regression estimates have the finite-population degree-of-freedom adjustment applied to them and the SEM estimates do not. To see the equivalence, you must undo the adjustment on the reported linear regression standard errors by multiplying them by $\sqrt{\{(N-p-1)/N\}}$.

delta. Delta, δ, in the context of power and sample-size calculations, denotes the effect size.

delta method. See *linearization*.

dendrogram or **cluster tree**. A dendrogram or cluster tree graphically presents information about how observations are grouped together at various levels of (dis)similarity in hierarchical cluster analysis. At the bottom of the dendrogram, each observation is considered its own cluster. Vertical lines extend up for each observation, and at various (dis)similarity values, these lines are connected to the lines from other observations with a horizontal line. The observations continue to combine until, at the top of the dendrogram, all observations are grouped together. Also see *hierarchical clustering*.

dereference. Dereferencing is an action performed on pointers. Pointers contain memory addresses, such as 0x2a1228. One assumes something of interest is stored at 0x2a1228, say, a real scalar equal to 2. When one accesses that 2 via the pointer by coding `*`p, one is said to be dereferencing the pointer. Unary `*` is the dereferencing operator.

design effects. See *DEFF* and *DEFT*.

deterministic trend. A deterministic trend is a deterministic function of time that specifies the long-run tendency of a time series.

deviance information criterion, DIC. The deviance information criterion (DIC) is an information based criterion used for Bayesian model selection. It is an analog of AIC and is given by the formula $D(\overline{\theta}) + 2 \times p_D$, where $D(\overline{\theta})$ is the deviance at the sample mean and p_D is the effective complexity, a quantity equivalent to the number of parameters in the model. Models with smaller DIC are preferred.

DFBETA. A DFBETA measures the change in the regressor's coefficient because of deletion of that subject. Also see *partial DFBETA*.

diagonal matrix. A matrix is diagonal if its off-diagonal elements are zero; A is diagonal if `A[i,j]==0` for `i!=j`. Usually, diagonal matrices are also *square*. Some definitions require that a diagonal matrix also be a square matrix.

diagonal of a matrix. The diagonal of a matrix is the set of elements `A[i,j]`.

dichotomous item. See *binary item*.

difference operator. The difference operator Δ denotes the change in the value of a variable from period $t-1$ to period t. Formally, $\Delta y_t = y_t - y_{t-1}$, and $\Delta^2 y_t = \Delta(y_t - y_{t-1}) = (y_t - y_{t-1}) - (y_{t-1} - y_{t-2}) = y_t - 2y_{t-1} + y_{t-2}$.

difficulty. A level of the latent trait needed to pass an item or an item category.

dilation. A dilation stretches or shrinks distances in Procrustes rotation.

dimension. A dimension is a parameter or measurement required to define a characteristic of an object or observation. Dimensions are the variables in the dataset. Weight, height, age, blood pressure, and drug dose are examples of dimensions in health data. Number of employees, gross income, net income, tax, and year are examples of dimensions in data about companies.

diminishing adaptation. Diminishing adaptation of the adaptive algorithm is the type of adaptation in which the amount of adaptation decreases with the size of the MCMC chain.

direct, **indirect**, and **total effects**. Consider the following system of equations:

$$
\begin{aligned}
y_1 &= b_{10} + b_{11}y_2 + b_{12}x_1 + b_{13}x_3 + e_1 \\
y_2 &= b_{20} + b_{21}y_3 + b_{22}x_1 + b_{23}x_4 + e_2 \\
y_3 &= b_{30} + \phantom{b_{21}y_3 +{}} b_{32}x_1 + b_{33}x_5 + e_3
\end{aligned}
$$

The total effect of x_1 on y_1 is $b_{12} + b_{11}b_{22} + b_{11}b_{21}b_{32}$. It measures the full change in y_1 based on allowing x_1 to vary throughout the system.

The direct effect of x_1 on y_1 is b_{12}. It measures the change in y_1 caused by a change in x_1 holding other endogenous variables—namely, y_2 and y_3—constant.

The indirect effect of x_1 on y_1 is obtained by subtracting the total and direct effect and is thus $b_{11}b_{22} + b_{11}b_{21}b_{32}$.

direct standardization. Direct standardization is an estimation method that allows comparing rates that come from different frequency distributions.

Estimated rates (means, proportions, and ratios) are adjusted according to the frequency distribution from a standard population. The standard population is partitioned into categories called standard strata. The stratum frequencies for the standard population are called standard weights. The standardizing frequency distribution typically comes from census data, and the standard strata are most commonly identified by demographic information such as age, sex, and ethnicity.

directional test. See *one-sided test*.

disambiguation: characters, and bytes, and display columns. A character is simply the letter or symbol that you want to represent—the letter "a", the punctuation mark ".", or a Chinese logogram. A byte or sequence of bytes is how that character is stored in the computer. And, a display column is the space required to display one typical character in the fixed-width display used by Stata's Results window and Viewer. Some characters are too wide for one display column. Each character is displayed in one or two display columns.

For plain ASCII characters, the number of characters always equals the number of bytes and equals the number of display columns.

For UTF-8 characters that are not plain ASCII, there are usually two bytes per character but there are sometimes three or even four bytes per character, such as for Chinese, Japanese, and Korean (CJK) characters. Characters that are too wide to fit in one display column (such as CJK characters) are displayed in two display columns.

In general, for Unicode characters, the relationship between the number of characters and the number of bytes and the relationship between the number of characters and the number of display columns is more ambiguous. All characters can be stored in four or fewer bytes and are displayed in Stata using two or fewer display columns.

See [U] **12.4.2.1 Unicode string functions** and [U] **12.4.2.2 Displaying Unicode characters** to learn how to deal with the distinction between characters, bytes, and display columns in your code.

discordant pairs. In a 2×2 contingency table, discordant pairs are the success-failure or failure-success pairs of observations. Also see *concordant pairs* and *Introduction* under *Remarks and examples* in [PSS] **power pairedproportions**.

discordant proportion. This is a proportion of discordant pairs or discordant sets. Also see *Introduction* under *Remarks and examples* in [PSS] **power pairedproportions** as well as *Introduction* under *Remarks and examples* in [PSS] **power mcc**.

discordant sets. In a matched study with multiple controls matched to a case, discordant sets are the sets in which there is any success–failure or failure–success match between the case and any matched control. Also see *Introduction* under *Remarks and examples* in [PSS] **power mcc**.

discrete parameters. Discrete parameters are parameters with discrete prior distributions.

discriminant analysis. Discriminant analysis is used to describe the differences between groups and to exploit those differences when allocating (classifying) observations of unknown group membership. Discriminant analysis is also called classification in many references.

discriminant function. Discriminant functions are formed from the eigenvectors from Fisher's approach to LDA. See *linear discriminant analysis*. See *classification function* for an alternative.

discriminating variables. Discriminating variables in a discriminant analysis are analyzed to determine differences between groups where group membership is known. These differences between groups are then exploited when classifying observations to the groups.

discrimination. A measure of how well an item can distinguish between contiguous latent trait levels near the inflection point of an item characteristic curve.

disparity. Disparities are transformed dissimilarities, that is, dissimilarity values transformed by some function. The class of functions to transform dissimilarities to disparities may either be (1) a class of metric, or known functions such as linear functions or power functions that can be parameterized by real scalars or (2) a class of more general (nonmetric) functions, such as any monotonic function. Disparities are used in MDS. Also see *dissimilarity*, *multidimensional scaling*, *metric scaling*, and *nonmetric scaling*.

display column. A display column is the space required to display one typical character in the fixed-width display used by Stata's Results window and Viewer. Some characters are too wide for one display column. Each character is displayed in one or two display columns.

All plain ASCII characters (for example, "M" and "9") and many UTF-8 characters that are not plain ASCII (for example, "é") require the same space when using a fixed-width font. That is to say, they all require a single display column.

Characters from non-Latin alphabets, such as Chinese, Cyrillic, Japanese, and Korean, may require two display columns.

See [U] 12.4.2.2 Displaying Unicode characters for more information.

dissimilarity, **dissimilarity matrix**, and **dissimilarity measure**. Dissimilarity or a dissimilarity measure is a quantification of the difference between two things, such as observations or variables or groups of observations or a method for quantifying that difference. A dissimilarity matrix is a matrix containing dissimilarity measurements. Euclidean distance is one example of a dissimilarity measure. Contrast to *similarity*. Also see *proximity* and *Euclidean distance*.

disturbance term. The disturbance term encompasses any shocks that occur to the dependent variable that cannot be explained by the conditional (or deterministic) portion of the model.

divisive hierarchical clustering methods. Divisive hierarchical clustering methods are top-down methods for hierarchical clustering. All the data begin as a part of one large cluster; with each iteration, a cluster is broken into two to create two new clusters. At the first iteration there are two clusters, then three, and so on. Divisive methods are very computationally expensive. Contrast to *agglomerative hierarchical clustering methods*.

doubly robust estimator. A doubly robust estimator only needs one of two auxiliary models to be correctly specified to estimate a parameter of interest.

Doubly robust estimators for treatment effects are consistent when either the outcome model or the treatment model is correctly specified.

drift. Drift is the constant term in a unit-root process. In

$$y_t = \alpha + y_{t-1} + \epsilon_t$$

α is the drift when ϵ_t is a stationary, zero-mean process.

dropout. Dropout is the withdrawal of subjects before the end of a study and leads to incomplete or missing data.

dyadic operator. Synonym for binary operator.

dynamic forecast. A dynamic forecast uses forecast values wherever lagged values of the endogenous variables appear in the model, allowing one to forecast multiple periods into the future.

dynamic model. A dynamic model is one in which prior values of the dependent variable or disturbance term affect the current value of the dependent variable.

dynamic-multiplier function. A dynamic-multiplier function measures the effect of a shock to an exogenous variable on an endogenous variable. The kth dynamic-multiplier function of variable i on variable j measures the effect on variable j in period $t + k$ in response to a one-unit shock to variable i in period t, holding everything else constant.

EB. See *empirical Bayes*.

EE estimator. See *estimating-equation estimator*.

effect size. The effect size is the size of the clinically significant difference between the treatments being compared, often expressed as the hazard ratio (or the log of the hazard ratio) in survival analysis.

effective sample size, ESS. Effective sample size (ESS) is the MCMC sample size T adjusted for the autocorrelation in the sample. It represents the number of independent observations in an MCMC sample. ESS is used instead of T in calculating MCSE. Small ESS relative to T indicates high autocorrelation and consequently poor mixing of the chain.

effect-size curve. The effect-size curve is a graph of the estimated effect size or target parameter as a function of some other study parameter such as the sample size. The effect size or target parameter is plotted on the y axis, and the sample size or other parameter is plotted on the x axis.

effect-size determination. This pertains to the computation of an effect size or a target parameter given power, sample size, and other study parameters.

efficiency. In the context of MCMC, efficiency is a term used for assessing the mixing quality of an MCMC procedure. Efficient MCMC algorithms are able to explore posterior domains in less time (using fewer iterations). Efficiency is typically quantified by the sample autocorrelation and effective sample size. An MCMC procedure that generates samples with low autocorrelation and consequently high ESS is more efficient.

eigenvalues and eigenvectors. A scalar, λ, is said to be an eigenvalue of square matrix $\mathbf{A}$: $n \times n$ if there is a nonzero column vector $\mathbf{x}$: $n \times 1$ (called an eigenvector) such that

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x} \tag{1}$$

Equation (1) can also be written

$$(\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = 0$$

where $\mathbf{I}$ is the $n \times n$ identity matrix. A nontrivial solution to this system of n linear homogeneous equations exists if and only if

$$\det(\mathbf{A} - \lambda\mathbf{I}) = 0 \tag{2}$$

This nth-degree polynomial in λ is called the characteristic polynomial or characteristic equation of $\mathbf{A}$, and the eigenvalues λ are its roots, also known as the characteristic roots.

The eigenvector defined by (1) is also known as the right eigenvector, because matrix $\mathbf{A}$ is postmultiplied by eigenvector $\mathbf{x}$. See [M-5] **eigensystem( )** and *left eigenvectors*.

EIM, vce(eim). EIM stands for expected information matrix, defined as the inverse of the negative of the expected value of the matrix of second derivatives, usually of the log-likelihood function. The EIM is an estimate of the VCE. EIM standard errors are reported when `sem` option `vce(eim)` is specified. EIM is available only with `sem`. The other available techniques are OIM, OPG, robust, clustered, bootstrap, and jackknife.

eltype. See *type, eltype, and orgtype*.

EM. See *expectation-maximization algorithm*.

empirical Bayes. In generalized linear mixed-effects models, empirical Bayes refers to the method of prediction of the random effects after the model parameters have been estimated. The empirical Bayes method uses Bayesian principles to obtain the posterior distribution of the random effects, but instead of assuming a prior distribution for the model parameters, the parameters are treated as given.

In IRT models, empirical Bayes refers to the method of prediction of the latent trait after the model parameters have been estimated. The empirical Bayes method uses Bayesian principles to obtain the posterior distribution of the latent trait. However, instead of assuming a prior distribution for the model parameters, one treats the parameters as given.

empirical Bayes mean. See *posterior mean*.

empirical Bayes mode. See *posterior mode*.

encodings. An encoding is a way of representing a character as a byte or series of bytes. Examples of encoding systems are ASCII and UTF-8. Stata uses UTF-8 encoding.

For more information, see [U] **12.4.2.3 Encodings**.

endogenous variable. An endogenous variable is a regressor that is correlated with the unobservable error term. Equivalently, an endogenous variable is one whose values are determined by the equilibrium or outcome of a structural model.

In the context of structural equation modeling and path diagrams, a variable, either observed or latent, is endogenous if any paths point to it.

Also see *exogenous variable*.

epsilon(1), etc.. epsilon(1) refers to the unit roundoff error associated with a computer, also informally called machine precision. It is the smallest amount by which a number may differ from 1. For IEEE double-precision variables, epsilon(1) is approximately 2.22045e–16.

`epsilon(`*x*`)` is the smallest amount by which a real number can differ from *x*, or an approximation thereof; see [M-5] **epsilon( )**.

equal-allocation design. See *balanced design*.

equal-tailed credible interval. An equal-tailed credible interval is a credible interval defined in such a way that both tails of the marginal posterior distribution have the same probability. A $\{100 \times (1-\alpha)\}\%$ equal-tailed credible interval is defined by the $\alpha/2$th and $\{(1-\alpha)/2\}$th quantiles of the marginal posterior distribution.

equamax rotation. Equamax rotation is an orthogonal rotation whose criterion is a weighted sum of the varimax and quartimax criteria. Equamax reflects a concern for simple structure within the rows and columns of the matrix. It is equivalent to oblimin with $\gamma = p/2$, or to the Crawford–Ferguson family with $\kappa = f/2p$, where p is the number of rows of the matrix to be rotated, and f is the number of columns. Also see *orthogonal rotation*, *varimax rotation*, *quartimax rotation*, *oblimin rotation*, and *Crawford–Ferguson rotation*.

error, **error variable**. The error is random disturbance e in a linear equation:

$$y = b_0 + b_1x_1 + b_2x_2 + \cdots + e$$

An error variable is an unobserved exogenous variable in path diagrams corresponding to e. Mathematically, error variables are just another example of latent exogenous variables, but in `sem` and `gsem`, error variables are considered to be in a class by themselves. All (Gaussian) endogenous variables—observed and latent—have a corresponding error variable. Error variables automatically and inalterably have their path coefficients fixed to be 1. Error variables have a fixed naming convention in the software. If a variable is the error for (observed or latent) endogenous variable `y`, then the residual variable's name is `e.y`.

In `sem` and `gsem`, error variables are uncorrelated with each other unless explicitly indicated otherwise. That indication is made in path diagrams by drawing a curved path between the error variables and is indicated in command notation by including `cov(e.`*name1*`*e.`*name2*`)` among the options specified on the `sem` command. In `gsem`, errors for family Gaussian, link log responses are not allowed to be correlated.

error-components model. The error-components model is another name for the random-effects model. See also *random-effects model*.

estimating-equation estimator. An estimating-equation (EE) estimator calculates parameters estimates by solving a system of equations. Each equation in this system is the sample average of a function that has mean zero.

These estimators are also known as M estimators or Z estimators in the statistics literature and as generalized method of moments (GMM) estimators in the econometrics literature.

estimation method. There are a variety of ways that one can solve for the parameters of an SEM. Different methods make different assumptions about the data-generation process, and so it is important that you choose a method appropriate for your model and data; see [SEM] **intro 4**.

Euclidean distance. The Euclidean distance between two observations is the distance one would measure with a ruler. The distance between vector $\mathbf{P} = (P_1, P_2, \ldots, P_n)$ and $\mathbf{Q} = (Q_1, Q_2, \ldots, Q_n)$ is given by

$$D(\mathbf{P},\mathbf{Q}) = \sqrt{(P_1 - Q_1)^2 + (P_2 - Q_2)^2 + \cdots + (P_n - Q_n)^2} = \sqrt{\sum_{i=1}^{n}(P_i - Q_i)^2}$$

event. An event is something that happens at an instant in time, such as being exposed to an environmental hazard, being diagnosed as myopic, or becoming employed.

The failure event is of special interest in survival analysis, but there are other equally important events, such as the exposure event, from which analysis time is defined.

In st data, events occur at the end of the recorded time span.

event of interest. In a competing-risks analysis, the event of interest is the event that is the focus of the analysis, that for which the cumulative incidence in the presence of competing risks is estimated.

exact denominator degrees of freedom (DDF) methods. Residual, repeated, and ANOVA DDF methods are referred to as exact methods because they provide exact t and F sampling distributions of test statistics for special classes of mixed-effects models—linear regression, repeated-measures designs, and traditional ANOVA models—with balanced data. Also see *approximation denominator degrees of freedom (DDF) methods*.

exact test. An exact test is one for which the probability of observing the data under the null hypothesis is calculated directly, often by enumeration. Exact tests do not rely on any asymptotic approximations and are therefore widely used with small datasets. See [PSS] **power oneproportion** and [PSS] **power twoproportions**.

exogenous variable. An exogenous variable is a regressor that is not correlated with any of the unobservable error terms in the model. Equivalently, an exogenous variable is one whose values change independently of the other variables in a structural model.

In the context of structural equation modeling and path diagrams, a variable, either observed or latent, is exogenous if paths only originate from it, or, equivalently, no paths point to it.

Also see *endogenous variable*.

exp. *exp* is used in syntax diagrams to mean “any valid expression may appear here”; see [M-2] **exp**.

expectation-maximization algorithm. In the context of MI, an iterative procedure for obtaining maximum likelihood or posterior-mode estimates in the presence of missing data.

experimental group. An experimental group is a group of subjects that receives a treatment or procedure of interest defined in a controlled experiment. In hypothesis testing, this is usually a comparison group. Also see *control group*.

experimental study. In an experimental study, as opposed to an observational study, the assignment of subjects to treatments is controlled by investigators. For example, a study that compares a new treatment with a standard treatment by assigning each treatment to a group of subjects is an experimental study.

exponential smoothing. Exponential smoothing is a method of smoothing a time series in which the smoothed value at period t is equal to a fraction α of the series value at time t plus a fraction $1-\alpha$ of the previous period's smoothed value. The fraction α is known as the smoothing parameter.

exponential test. The exponential test is the parametric test comparing the hazard rates, λ_1 and λ_2, (or log hazards) from two independent exponential (constant only) regression models with the null hypothesis H_0: $\lambda_2 - \lambda_1 = 0$ [or H_0: $\ln(\lambda_2) - \ln(\lambda_1) = \ln(\lambda_2/\lambda_1) = 0$].

exposure odds ratio. An odds ratio of exposure in cases relative to controls in a case–control study.

extended ASCII. Extended ASCII, also known as higher ASCII, is the byte values 128 to 255, which were not defined as part of the original ASCII specification. Various code pages have been defined over the years to map the extended ASCII byte values to many characters not supported in the original ASCII specification, such as Latin letters with diacritical marks, such as "á" and "Á"; non-Latin alphabets, such as Chinese, Cyrillic, Japanese, and Korean; punctuation marks used in non-English languages, such as "«", complex mathematical symbols such as "±", and more.

Although extended ASCII characters are stored in a single byte in ASCII encoding, UTF-8 stores the same characters in two to four bytes. Because each code page maps the extended ASCII values differently, another distinguishing feature of extended ASCII characters is that their meaning can change across fonts and operating systems.

external variable. See *global variable*.

f test. An f test is a test for which a sampling distribution of a test statistic is an F distribution. See [PSS] **power twovariances**.

factor. A factor is an unobserved random variable that is thought to explain variability among observed random variables.

factor analysis. Factor analysis is a statistical technique used to explain variability among observed random variables in terms of fewer unobserved random variables called factors. The observed variables are then linear combinations of the factors plus error terms.

If the correlation matrix of the observed variables is $\mathbf{R}$, then $\mathbf{R}$ is decomposed by factor analysis as

$$\mathbf{R} = \mathbf{\Lambda}\mathbf{\Phi}\mathbf{\Lambda}' + \mathbf{\Psi}$$

$\mathbf{\Lambda}$ is the loading matrix, and $\mathbf{\Psi}$ contains the specific variances, for example, the variance specific to the variable not explained by the factors. The default unrotated form assumes uncorrelated common factors, $\mathbf{\Phi} = \mathbf{I}$.

factor loading plot. A factor loading plot produces a scatter plot of the factor loadings after factor analysis.

factor loadings. Factor loadings are the regression coefficients which multiply the factors to produce the observed variables in the factor analysis.

factor parsimony. Factor parsimony is an oblique rotation, which maximizes the column simplicity of the matrix. It is equivalent to a Crawford–Ferguson rotation with $\kappa = 1$. Also see *oblique rotation* and *Crawford–Ferguson rotation*.

factor scores. Factor scores are computed after factor analysis. Factor scores are the coordinates of the original variables, $\mathbf{x}$, in the space of the factors. The two types of scoring are regression scoring (Thomson 1951) and Bartlett (1937, 1938) scoring.

Using the symbols defined in *factor analysis*, the formula for regression scoring is

$$\widehat{\mathbf{f}} = \mathbf{\Lambda}'\mathbf{R}^{-1}\mathbf{x}$$

In the case of oblique rotation the formula becomes

$$\widehat{\mathbf{f}} = \boldsymbol{\Phi}\boldsymbol{\Lambda}'\mathbf{R}^{-1}\mathbf{x}$$

The formula for Bartlett scoring is

$$\widehat{\mathbf{f}} = \boldsymbol{\Gamma}^{-1}\boldsymbol{\Lambda}'\boldsymbol{\Psi}^{-1}\mathbf{x}$$

where

$$\boldsymbol{\Gamma} = \boldsymbol{\Lambda}'\boldsymbol{\Psi}^{-1}\boldsymbol{\Lambda}$$

Also see *factor analysis*.

failure event. Survival analysis is really time-to-failure analysis, and the failure event is the event under analysis. The failure event can be death, heart attack, myopia, or finding employment. Many authors—including Stata—write as if the failure event can occur only once per subject, but when we do, we are being sloppy. Survival analysis encompasses repeated failures, and all of Stata's survival analysis features can be used with repeated-failure data.

family distribution. See *generalized linear response functions*.

FCS. See *fully conditional specification*.

feasible initial value. An initial-value vector is feasible if it corresponds to a state with a positive posterior probability.

fictional data. Fictional data are data that have no basis in reality even though they might look real; they are data that are made up for use in examples.

finite population correction. Finite population correction (FPC) is an adjustment applied to the variance of a point estimator because of sampling without replacement, resulting in variance estimates that are smaller than the variance estimates from comparable with-replacement sampling designs.

first- and second-order latent variables. If a latent variable is measured by other latent variables only, the latent variable that does the measuring is called first-order latent variable, and the latent variable being measured is called the second-order latent variable.

first-, second-, and higher-level (latent) variables. Consider a multilevel model of patients within doctors within hospitals. First-level variables are variables that vary at the observational (patient) level. Second-level variables vary across doctors but are constant within doctors. Third-level variables vary across hospitals but are constant within hospitals. This jargon is used whether variables are latent or not.

Fisher–Irwin's exact test. See *Fisher's exact test*.

Fisher's exact test. Fisher's exact test is an exact small-sample test of independence between rows and columns in a 2×2 contingency table. Conditional on the marginal totals, the test statistic has a hypergeometric distribution under the null hypothesis. See [PSS] **power twoproportions** and [R] **tabulate twoway**.

Fisher's z test. This is a z test comparing one or two correlations. See [PSS] **power onecorrelation** and [PSS] **power twocorrelations**. Also see *Fisher's z transformation*.

Fisher's z transformation. Fisher's z transformation applies an inverse hyperbolic tangent transformation to the sample correlation coefficient. This transformation is useful for testing hypothesis concerning Pearson's correlation coefficient. The exact sampling distribution of the correlation coefficient is complicated, while the transformed statistic is approximately standard normal.

fixed effects. In the context of multilevel mixed-effects models, fixed effects represent effects that are constant for all groups at any level of nesting. In the ANOVA literature, fixed effects represent the levels of a factor for which the inference is restricted to only the specific levels observed in the study. See also *fixed-effects model* in [XT] **Glossary**.

fixed-effects model. The fixed-effects model is a model for panel data in which the panel-specific errors are treated as fixed parameters. These parameters are panel-specific intercepts and therefore allow the conditional mean of the dependent variable to vary across panels. The linear fixed-effects estimator is consistent, even if the regressors are correlated with the fixed effects. See also *random-effects model*.

flong data. See *style*.

flongsep data. See *style*.

FMI. See *fraction of missing information*.

follow-up period or **follow-up**. The (minimum) follow-up period is the period after the last subject entered the study until the end of the study. The follow-up defines the phase of a study during which subjects are under observation and no new subjects enter the study. If T is the total duration of a study, and R is the accrual period of the study, then follow-up period f is equal to $T - R$. Also see *accrual period*.

follow-up study. See *cohort study*.

forecast-error variance decomposition. Forecast-error variance decompositions measure the fraction of the error in forecasting variable i after h periods that is attributable to the orthogonalized shocks to variable j.

forward operator. The forward operator F denotes the value of a variable at time $t + 1$. Formally, $Fy_t = y_{t+1}$, and $F^2y_t = Fy_{t+1} = y_{t+2}$.

FPC. See *finite population correction*.

fraction of missing information. The ratio of information lost due to the missing data to the total information that would be present if there were no missing data.

An equal FMI test is a test under the assumption that FMIs are equal across parameters.

An unrestricted FMI test is a test without the equal FMI assumption.

fractional polynomial. A polynomial that may include logarithms, noninteger powers, and repeated powers.

Each time a power repeats in a fractional polynomial of x, it is multiplied by another $\ln(x)$.

We write a fractional polynomial in x as

$$x^{(p_1,p_2,\dots,p_m)\prime}\boldsymbol{\beta}$$

A fractional polynomial in x with powers $(-1, 0, 0.5, 3, 3)$ and coefficients $\boldsymbol{\beta}$ has the following form:

$$x^{(-1,0,0.5,3,3)\prime}\boldsymbol{\beta} = \beta_0 + \beta_1 x^{-1} + \beta_2 \ln(x) + \beta_3 x^{.5} + \beta_4 x^3 + \beta_5 x^3 \ln(x)$$

The notation $x^{(-2,3)}$, for example, means the variable x^{-2} and the variable x^3.

frailty. In survival analysis, it is often assumed that subjects are alike—homogeneous—except for their observed differences. The probability that subject j fails at time t may be a function of j's covariates and random chance. Subjects j and k, if they have equal covariate values, are equally likely to fail.

Frailty relaxes that assumption. The probability that subject j fails at time t becomes a function of j's covariates and j's unobserved frailty value, ν_j. Frailty ν is assumed to be a random variable. Parametric survival models can be fit even in the presence of such heterogeneity.

Shared frailty refers to the case in which groups of subjects share the same frailty value. For instance, subjects 1 and 2 may share frailty value ν because they are genetically related. Both parametric and semiparametric models can be fit under the shared-frailty assumption.

frequency-domain analysis. Frequency-domain analysis is analysis of time-series data by considering its frequency properties. The spectral density and distribution functions are key components of frequency-domain analysis, so it is often called spectral analysis. In Stata, the `cumsp` and `pergram` commands are used to analyze the sample spectral distribution and density functions, respectively. `psdensity` estimates the spectral density or the spectral distribution function after estimating the parameters of a parametric model using `arfima`, `arima`, or `ucm`.

frequentist analysis. Frequentist analysis is a form of statistical analysis where model parameters are considered to be unknown but fixed constants and the observed data are viewed as a repeatable random sample. Inference is based on the sampling distribution of the data.

full conditionals. A full conditional is the probability distribution of a random variate conditioned on all other random variates in a joint probability model. Full conditional distributions are used in Gibbs sampling.

full Gibbs sampling. See *Gibbs sampling, Gibbs sampler*.

full joint and conditional normality assumption. See *normality assumption, joint and conditional*.

fully conditional specification. Consider imputation variables $X_1, X_2, \ldots, X_p$. Fully conditional specification of the prediction equation for X_j includes all variables except X_j; that is, variables $\mathbf{X}_{-j} = (X_1, X_2, \ldots, X_{j-1}, X_{j+1}, \ldots, X_p)$.

function. The words *program* and *function* are used interchangeably. The programs that you write in Mata are in fact functions. Functions receive arguments and optionally return results.

Examples of functions that are included with Mata are `sqrt()`, `ttail()`, and `substr()`. Such functions are often referred to as the built-in functions or the library functions. Built-in functions refer to functions implemented in the C code that implements Mata, and library functions refer to functions written in the Mata programming language, but many users use the words interchangeably because how functions are implemented is of little importance. If you have a choice between using a built-in function and a library function, however, the built-in function will usually execute more quickly and the library function will be easier to use. Mostly, however, features are implemented one way or the other and you have no choice.

Also see *underscore functions*.

For a list of the functions that Mata provides, see [M-4] **intro**.

future history. Future history is information recorded after a subject is no longer at risk. The word *history* is often dropped, and the term becomes simply *future*. Perhaps the failure event is cardiac infarction, and you want to know whether the subject died soon in the *future*, in which case you might exclude the subject from analysis.

Also see *past history*.

gain (of a linear filter). The gain of a linear filter scales the spectral density of the unfiltered series into the spectral density of the filtered series for each frequency. Specifically, at each frequency, multiplying the spectral density of the unfiltered series by the square of the gain of a linear filter yields the spectral density of the filtered series. If the gain at a particular frequency is 1, the filtered and unfiltered spectral densities are the same at that frequency and the corresponding stochastic cycles are passed through perfectly. If the gain at a particular frequency is 0, the filter removes all the corresponding stochastic cycles from the unfiltered series.

gamma regression. Gamma regression is a term for generalized linear response functions that are family gamma, link log. It is used for continuous, nonnegative, positively skewed data. Gamma regression is also known as log-gamma regression. See *generalized linear response functions*.

gaps. Gaps refers to gaps in observation between entry time and exit time; see *under observation*.

GARCH model. A generalized autoregressive conditional heteroskedasticity (GARCH) model is a regression model in which the conditional variance is modeled as an ARMA process. The GARCH(m,k) model is

$$
\begin{aligned}
y_t &= \mathbf{x}_t\boldsymbol{\beta} + \epsilon_t \\
\sigma_t^2 &= \gamma_0 + \gamma_1\epsilon_{t-1}^2 + \cdots + \gamma_m\epsilon_{t-m}^2 + \delta_1\sigma_{t-1}^2 + \cdots + \delta_k\sigma_{t-k}^2
\end{aligned}
$$

where the equation for y_t represents the conditional mean of the process and σ_t represents the conditional variance. See [TS] **arch** or Hamilton (1994, chap. 21) for details on how the conditional variance equation can be viewed as an ARMA process. GARCH models are often used because the ARMA specification often allows the conditional variance to be modeled with fewer parameters than are required by a pure ARCH model. Many extensions to the basic GARCH model exist; see [TS] **arch** for those that are implemented in Stata. See also *ARCH model*.

Gauss–Hermite quadrature. In the context of generalized linear mixed models, Gauss–Hermite quadrature is a method of approximating the integral used in the calculation of the log likelihood. The quadrature locations and weights for individual clusters are fixed during the optimization process.

In the context of IRT models, Gauss–Hermite quadrature (GHQ) is a method of approximating the integral used in the calculation of the log likelihood. The quadrature locations and weights for individuals are fixed during the optimization process.

Gaussian regression. Gaussian regression is another term for linear regression. It is most often used when referring to generalized linear response functions. In that framework, Gaussian regression is family Gaussian, link identity. See *generalized linear response functions*.

generalized eigenvalues. A scalar, λ, is said to be a generalized eigenvalue of a pair of $n \times n$ square numeric matrices $\mathbf{A}$, $\mathbf{B}$ if there is a nonzero column vector $\mathbf{x}$: $n \times 1$ (called a generalized eigenvector) such that

$$\mathbf{Ax} = \lambda\mathbf{Bx} \tag{1}$$

Equation (1) can also be written

$$(\mathbf{A} - \lambda\mathbf{B})\mathbf{x} = 0$$

A nontrivial solution to this system of n linear homogeneous equations exists if and only if

$$\det(\mathbf{A} - \lambda\mathbf{B}) = 0 \tag{2}$$

In practice, the generalized eigenvalue problem for the matrix pair $(\mathbf{A},\mathbf{B})$ is usually formulated as finding a pair of scalars (w, b) and a nonzero column vector $\mathbf{x}$ such that

$$w\mathbf{Ax} = b\mathbf{Bx}$$

The scalar w/b is a generalized eigenvalue if b is not zero.

Infinity is a generalized eigenvalue if b is zero or numerically close to zero. This situation may arise if $\mathbf{B}$ is singular.

The Mata functions that compute generalized eigenvalues return them in two complex vectors, $\mathbf{w}$ and $\mathbf{b}$ of length n. If $\mathbf{b}[i] = 0$, the ith generalized eigenvalue is infinite, otherwise the ith generalized eigenvalue is $\mathbf{w}[i]/\mathbf{b}[i]$.

generalized estimating equations (GEE). The method of generalized estimating equations is used to fit population-averaged panel-data models. GEE extends the GLM method by allowing the user to specify a variety of different within-panel correlation structures.

generalized least-squares estimator. A generalized least-squares (GLS) estimator is used to estimate the parameters of a regression function when the error term is heteroskedastic or autocorrelated. In the linear case, GLS is sometimes described as "OLS on transformed data" because the GLS estimator can be implemented by applying an appropriate transformation to the dataset and then using OLS.

generalized linear mixed-effects model. A generalized linear mixed-effects model is an extension of a generalized linear model allowing for the inclusion of random deviations (effects).

generalized linear model. The generalized linear model is an estimation framework in which the user specifies a distributional family for the dependent variable and a link function that relates the dependent variable to a linear combination of the regressors. The distribution must be a member of the exponential family of distributions. The generalized linear model encompasses many common models, including linear, probit, and Poisson regression.

generalized linear response functions. Generalized linear response functions include linear functions and include functions such as probit, logit, multinomial logit, ordered probit, ordered logit, Poisson, and more.

These generalized linear functions are described by a link function $g(\cdot)$ and statistical distribution F. The link function $g(\cdot)$ specifies how the response variable y_i is related to a linear equation of the explanatory variables, $\mathbf{x}_i\boldsymbol{\beta}$, and the family F specifies the distribution of y_i:

$$g\{E(y_i)\} = \mathbf{x}_i\boldsymbol{\beta}, \qquad y_i \sim F$$

If we specify that $g(\cdot)$ is the identity function and F is the Gaussian (normal) distribution, then we have linear regression. If we specify that $g(\cdot)$ is the logit function and F the Bernoulli distribution, then we have logit (logistic) regression.

In this generalized linear structure, the family may be Gaussian, gamma, Bernoulli, binomial, Poisson, negative binomial, ordinal, or multinomial. The link function may be the identity, log, logit, probit, or complementary log-log.

`gsem` fits models with generalized linear response functions.

generalized method of moments. Generalized method of moments (GMM) is a method used to obtain fitted parameters. In this documentation, GMM is referred to as ADF, which stands for asymptotic distribution free and is available for use with `sem`. Other available methods for use with `sem` are ML, QML, ADF, and MLMV.

The SEM moment conditions are cast in terms of second moments, not the first moments used in many other applications associated with GMM.

generalized partial credit model. The generalized partial credit model (GPCM) is an IRT model for ordinal responses. The categories within each item vary in their difficulty and share the same discrimination parameter.

generalized SEM. Generalized SEM is a term we have coined to mean SEM optionally allowing generalized linear response functions or multilevel models. `gsem` fits generalized SEMs.

GHQ. See *Gauss–Hermite quadrature*.

Gibbs sampling, Gibbs sampler. Gibbs sampling is an MCMC method, according to which each random variable from a joint probability model is sampled according to its full conditional distribution.

GLM. See *generalized linear model*.

GLME model. See *generalized linear mixed-effects model*.

GLMM. Generalized linear mixed model. See *generalized linear mixed-effects model*.

global variable. Global variables, also known as external variables and as global external variables, refer to variables that are common across programs and which programs may access without the variable being passed as an argument.

The variables you create interactively are global variables. Even so, programs cannot access those variables without engaging in another step, and global variables can be created without your creating them interactively.

To access (and create if necessary) global external variables, you declare the variable in the body of your program:

```
function myfunction(...)
{
     external real scalar globalvar
     ...
}
```

See *Linking to external globals* in [M-2] **declarations**.

There are other ways of creating and accessing global variables, but the declaration method is recommended. The alternatives are `crexternal()`, `findexternal()`, and `rmexternal()` documented in [M-5] **findexternal()** and `valofexternal()` documented in [M-5] **valofexternal()**.

GMM. See *generalized method of moments*.

goodness-of-fit statistic. A goodness-of-fit statistic is a value designed to measure how well the model reproduces some aspect of the data the model is intended to fit. SEM reproduces the first- and second-order moments of the data, with an emphasis on the second-order moments, and thus goodness-of-fit statistics appropriate for use after `sem` compare the predicted covariance matrix (and mean vector) with the matrix (and vector) observed in the data.

GPCM. See *generalized partial credit model*.

graded response model. The graded response model (GRM) is an extension of the two-parameter logistic model to ordinal responses. The categories within each item vary in their difficulty and share the same discrimination parameter.

Granger causality. The variable x is said to Granger-cause variable y if, given the past values of y, past values of x are useful for predicting y.

Greenhouse–Geisser correction. See *nonsphericity correction*.

GRM. See *graded response model*.

gsem. `gsem` is the Stata command that fits generalized SEMs. Also see *sem*.

guessing. The guessing parameter incorporates the impact of chance on an observed response. The parameter lifts the lower asymptote of the item characteristic curve above zero.

GUI. See *Builder*.

Hadamard matrix. A Hadamard matrix is a square matrix with r rows and columns that has the property

$$H_r' H_r = r I_r$$

where I_r is the identity matrix of order r. Generating a Hadamard matrix with order $r = 2^p$ is easily accomplished. Start with a Hadamard matrix of order 2 (H_2), and build your H_r by repeatedly applying Kronecker products with H_2.

hard missing and soft missing. A hard missing value is a value of `.a`, `.b`, ..., `.z` in $m = 0$ in an imputed variable. Hard missing values are not replaced in $m > 0$.

A soft missing value is a value of `.` in $m = 0$ in an imputed variable. If an imputed variable contains soft missing, then that value is eligible to be imputed, and perhaps is imputed, in $m > 0$.

Although you can use the terms hard missing and soft missing for passive, regular, and unregistered variables, it has no special significance in terms of how the missing values are treated.

hashing, hash functions, and hash tables. Hashing refers to a technique for quickly finding information corresponding to an identifier. The identifier might be a name, a Social Security number, fingerprints, or anything else on which the information is said to be indexed. The hash function returns a many-to-one mapping of identifiers onto a dense subrange of the integers. Those integers, called hashes, are then used to index a hash table. The selected element of the hash table specifies a list containing identifiers and information. The list is then searched for the particular identifier desired. The advantage is that rather than searching a single large list, one need only search one of K smaller lists. For this to be fast, the hash function must be quick to compute and produce roughly equal frequencies of hashes over the range of identifiers likely to be observed.

hazard, **cumulative hazard**, and **hazard ratio**. The hazard or hazard rate at time t, $h(t)$, is the instantaneous rate of failure at time t conditional on survival until time t. Hazard rates can exceed 1. Say that the hazard rate were 3. If an individual faced a constant hazard of 3 over a unit interval and if the failure event could be repeated, the individual would be expected to experience three failures during the time span.

The cumulative hazard, $H(t)$, is the integral of the hazard function $h(t)$, from 0 (the onset of risk) to t. It is the total number of failures that would be expected to occur up until time t, if the failure event could be repeated. The relationship between the cumulative hazard function, $H(t)$, and the survivor function, $S(t)$, is

$$S(t) = \exp\{-H(t)\}$$

$$H(t) = -\ln\{S(t)\}$$

The hazard ratio is the ratio of the hazard function evaluated at two different values of the covariates: $h(t\,|\,\mathbf{x})/h(t\,|\,\mathbf{x}_0)$. The hazard ratio is often called the relative hazard, especially when $h(t\,|\,\mathbf{x}_0)$ is the baseline hazard function.

hazard contributions. Hazard contributions are the increments of the estimated cumulative hazard function obtained through either a nonparametric or semiparametric analysis. For these analysis types, the estimated cumulative hazard is a step function that increases every time a failure occurs. The hazard contribution for that time is the magnitude of that increase.

Because the time between failures usually varies from failure to failure, hazard contributions do not directly estimate the hazard. However, one can use the hazard contributions to formulate an estimate of the hazard function based on the method of smoothing.

Hermitian matrix. Matrix A is Hermitian if it is equal to its conjugate transpose; $A = A'$; see *transpose*. This means that each off-diagonal element a_{ij} must equal the conjugate of a_{ji}, and that the diagonal elements must be real. The following matrix is Hermitian:

$$\begin{bmatrix} 2 & 4+5i \\ 4-5i & 6 \end{bmatrix}$$

The definition $A = A'$ is the same as the definition for a symmetric matrix, although usually the word *symmetric* is reserved for real matrices and Hermitian, for complex matrices. In this manual, we use the word *symmetric* for both; see *symmetric matrices*.

Hessenberg decomposition. The Hessenberg decomposition of a matrix, $\mathbf{A}$, can be written as

$$\mathbf{Q}'\mathbf{A}\mathbf{Q} = \mathbf{H}$$

where $\mathbf{H}$ is in upper Hessenberg form and $\mathbf{Q}$ is orthogonal if $\mathbf{A}$ is real or unitary if $\mathbf{A}$ is complex. See [M-5] **hessenbergd()**.

Hessenberg form. A matrix, $\mathbf{A}$, is in upper Hessenberg form if all entries below the first subdiagonal are zero: $A_{ij} = 0$ for all $i > j+1$.

A matrix, $\mathbf{A}$, is in lower Hessenberg form if all entries above the first superdiagonal are zero: $A_{ij} = 0$ for all $j > i+1$.

Heywood case or **Heywood solution**. A Heywood case can appear in factor analysis output; this indicates that a boundary solution, called a Heywood solution, was produced. The geometric assumptions underlying the likelihood-ratio test are violated, though the test may be useful if interpreted cautiously.

hierarchical clustering and **hierarchical clustering methods**. In hierarchical clustering, the data is placed into clusters via iterative steps. Contrast to *partition clustering*. Also see *agglomerative hierarchical clustering methods* and *divisive hierarchical clustering methods*.

hierarchical model. A hierarchical model is one in which successively more narrowly defined groups are nested within larger groups. For example, in a hierarchical model, patients may be nested within doctors who are in turn nested within the hospital at which they practice.

higher ASCII. See *extended ASCII*.

highest posterior density credible interval, HPD credible interval. The highest posterior density (HPD) credible interval is a type of a credible interval with the highest marginal posterior density. An HPD interval has the shortest width among all other credible intervals. For some multimodal marginal distributions, HPD may not exists. See *highest posterior density region, HPD region*.

highest posterior density region, HPD region. The highest posterior density (HPD) region for model parameters has the highest marginal posterior probability among all domain regions. Unlike an HPD credible interval, an HPD region always exist.

high-pass filter. Time-series filters are designed to pass or block stochastic cycles at specified frequencies. High-pass filters, such as those implemented in `tsfilter bw` and `tsfilter hp`, pass through stochastic cycles above the cutoff frequency and block all other stochastic cycles.

Holt–Winters smoothing. A set of methods for smoothing time-series data that assume that the value of a time series at time t can be approximated as the sum of a mean term that drifts over time, as well as a time trend whose strength also drifts over time. Variations of the basic method allow for seasonal patterns in data, as well.

Hotelling's T-squared generalized means test. Hotelling's T-squared generalized means test is a multivariate test that reduces to a standard t test if only one variable is specified. It tests whether one set of means is zero or if two sets of means are equal.

hybrid MH sampling, hybrid MH sampler. A hybrid MH sampler is an MCMC method in which some blocks of parameters are updated using the MH algorithms and other blocks are updated using Gibbs sampling.

hybrid model. A hybrid IRT model is a model that performs a single calibration of an instrument consisting of different response formats.

hyperparameter. In Bayesian analysis, hyperparameter is a parameter of a prior distribution, in contrast to a model parameter.

hyperprior. In Bayesian analysis, hyperprior is a prior distribution of hyperparameters. See *hyperparameter*.

hypothesis. A hypothesis is a statement about a population parameter of interest.

hypothesis testing, hypothesis test. This method of inference evaluates the validity of a hypothesis based on a sample from the population. See *Hypothesis testing* under *Remarks and examples* in [PSS] **intro**.

hypothesized value. See *null value*.

ICC. See *item characteristic curve*.

ID variable. An ID variable identifies groups; equal values of an ID variable indicate that the observations are for the same group. For instance, a stratification ID variable would indicate the strata to which each observation belongs.

When an ID variable is referred to without modification, it means subjects, and usually this occurs in multiple-record st data. In multiple-record data, each physical observation in the dataset represents a time span, and the ID variable ties the separate observations together:

idvar	*t0*	*t*
1	0	5
1	5	7

ID variables are usually numbered 1, 2, ..., but that is not required. An ID variable might be numbered 1, 3, 7, 22, ..., or −5, −4, ..., or even 1, 1.1, 1.2, ...

identification. Identification refers to the conceptual constraints on parameters of a model that are required for the model's remaining parameters to have a unique solution. A model is said to be unidentified if these constraints are not supplied. These constraints are of two types: substantive constraints and normalization constraints.

Normalization constraints deal with the problem that one scale works as well as another for each latent variable in the model. One can think, for instance, of propensity to write software as being measured on a scale of 0 to 1, 1 to 100, or any other scale. The normalization constraints are the constraints necessary to choose one particular scale. The normalization constraints are provided automatically by `sem` and `gsem` by anchoring with unit loadings.

Substantive constraints are the constraints you specify about your model so that it has substantive content. Usually, these constraints are zero constraints implied by the paths omitted, but they can include explicit parameter constraints as well. It is easy to write a model that is not identified for substantive reasons; See [SEM] **intro 4**.

idiosyncratic error term. In longitudinal or panel-data models, the idiosyncratic error term refers to the observation-specific zero-mean random-error term. It is analogous to the random-error term of cross-sectional regression analysis.

ignorable missing-data mechanism. The missing-data mechanism is said to be ignorable if missing data are missing at random and the parameters of the data model and the parameters of the missing-data mechanism are distinct; that is, the joint distribution of the model and the missing-data parameters can be factorized into two independent marginal distributions of model parameters and of missing-data parameters.

i.i.d. sampling assumption. See *independent and identically distributed sampling assumption*.

IIF. See *item information function*.

improper prior. A prior is said to be improper if it does not integrate to a finite number. Uniform distributions over unbounded intervals are improper. Improper priors may still yield proper posterior distributions. When using improper priors, however, one has to make sure that the resulting posterior distribution is proper for Bayesian inference to be invalid.

impulse–response function. An impulse–response function (IRF) measures the effect of a shock to an endogenous variable on itself or another endogenous variable. The kth impulse–response function of variable i on variable j measures the effect on variable j in period $t + k$ in response to a one-unit shock to variable i in period t, holding everything else constant.

imputed, passive, and regular variables. An imputed variable is a variable that has missing values and for which you have or will have imputations.

A passive variable is a varying variable that is a function of imputed variables or of other passive variables. A passive variable will have missing values in $m = 0$ and varying values for observations in $m > 0$.

A regular variable is a variable that is neither imputed nor passive and that has the same values, whether missing or not, in all m.

Imputed, passive, and regular variables can be registered using the `mi register` command; see [MI] **mi set**. You are required to register imputed variables, and we recommend that you register passive variables. Regular variables can also be registered. See *registered and unregistered variables*.

imputed data. Data in which all missing values are imputed.

incidence and **incidence rate**. Incidence is the number of new failures (for example, number of new cases of a disease) that occur during a specified period in a population at risk (for example, of the disease).

Incidence rate is incidence divided by the sum of the length of time each individual was exposed to the risk.

Do not confuse incidence with prevalence. Prevalence is the fraction of a population that has the disease. Incidence refers to the rate at which people contract a disease, whereas prevalence is the total number actually sick at a given time.

incidence studies, **longitudinal studies**, and **follow-up studies**. Whichever word is used, these studies monitor a population for a time to track the transition of noncases into cases. Incidence studies are prospective. Also see *cohort studies*.

incomplete observations. See *complete and incomplete observations*.

independent a posteriori. Parameters are considered independent a posteriori if their marginal posterior distributions are independent; that is, their joint posterior distribution is the product of their individual marginal posterior distributions.

independent a priori. Parameters are considered independent a priori if their prior distributions are independent; that is, their joint prior distribution is the product of their individual marginal prior distributions.

independent and identically distributed. A series of observations is independent and identically distributed (i.i.d.) if each observation is an independent realization from the same underlying distribution. In some contexts, the definition is relaxed to mean only that the observations are independent and have identical means and variances; see Davidson and MacKinnon (1993, 42).

independent and identically distributed sampling assumption. The independent and identically distributed (i.i.d.) sampling assumption specifies that each observation is unrelated to (independent of) all the other observations and that each observation is a draw from the same (identical) distribution.

indicator variables, indicators. The term "indicator variable" has two meanings. An indicator variable is a 0/1 variable that contains whether something is true. The other usage is as a synonym for measurement variables.

indirect effects. See *direct, indirect, and total effects*.

individual-level treatment effect. An individual-level treatment effect is the difference in an individual's outcome that would occur because this individual is given one treatment instead of another. In other words, an individual-level treatment effect is the difference between two potential outcomes for an individual.

For example, the blood pressure an individual would obtain after taking a pill minus the blood pressure an individual would obtain had that person not taken the pill is the individual-level treatment effect of the pill on blood pressure.

ineligible missing value. An ineligible missing value is a missing value in a to-be-imputed variable that is due to inability to calculate a result rather than an underlying value being unobserved. For instance, assume that variable `income` had some missing values and so you wish to impute it. Because `income` is skewed, you decide to impute the log of income, and you begin by typing

```
. generate lnincome = log(income)
```

If `income` contained any zero values, the corresponding missing values in `lnincome` would be ineligible missing values. To ensure that values are subsequently imputed correctly, it is of vital importance that any ineligible missing values be recorded as hard missing. You would do that by typing

```
. replace lnincome = .a if lnincome==. & income!=.
```

As an aside, if after imputing `lnincome` using `mi impute` (see [MI] **mi impute**), you wanted to fill in `income`, `income` surprisingly would be a passive variable because `lnincome` is the imputed variable and `income` would be derived from it. You would type

```
. mi register passive income
. mi passive: replace income = cond(lnincome==.a, 0, exp(lnincome))
```

In general, you should avoid using transformations that produce ineligible missing values to avoid the loss of information contained in other variables in the corresponding observations. For example, in the above, for zero values of `income` we could have assigned the log of income, `lnincome`, to be the smallest value that can be stored as `double`, because the logarithm of zero is negative infinity:

```
. generate lnincome = cond(income==0, mindouble(), log(income))
```

This way, all observations for which `income==0` will be used in the imputation model for `lnincome`.

inertia. In CA, the inertia is related to the definition in applied mathematics of "moment of inertia", which is the integral of the mass times the squared distance to the centroid. Inertia is defined as the total Pearson chi-squared for the two-way table divided by the total number of observations or the sum of the squared singular values found in the singular value decomposition.

$$\text{total inertia} = \frac{1}{n}\chi^2 = \sum_k \lambda_k^2$$

In MCA, the inertia is defined analogously. In the case of the indicator or Burt matrix approach, it is given by the formula

$$\text{total inertia} = \left(\frac{q}{q-1}\right)\sum \phi_t^2 - \frac{(J-q)}{q^2}$$

where q is the number of active variables, J is the number of categories and ϕ_t is the tth (unadjusted) eigenvalue of the eigen decomposition. In JCA the total inertia of the modified Burt matrix is defined as the sum of the inertias of the off-diagonal blocks. Also see *correspondence analysis* and *multiple correspondence analysis*.

information. Precision with which an item or an instrument measures the latent trait; also see *item information function* and *test information function*.

informative prior. An informative prior is a prior distribution that has substantial influence on the posterior distribution.

initial values. See *starting values*.

instance and realization. Instance and realization are synonyms for variable, as in Mata variable. For instance, consider a real scalar variable *X*. One can equally well say that *X* is an instance of a real scalar or a realization of a real scalar. Authors represent a variable this way when they wish to emphasize that *X* is not representative of all real scalars but is just one of many real scalars. Instance is often used with structures and classes when the writer wishes to emphasize the difference between the values contained in the variable and the definition of the structure or the class. It is confusing to say that *V* is a class *C*, even though it is commonly said, because the reader might confuse the definition of *C* with the specific values contained in *V*. Thus careful authors say that *V* is an instance of class *C*.

instrument. A collection of items, usually called a test, a survey, or a questionnaire.

instrumental variables. Instrumental variables are exogenous variables that are correlated with one or more of the endogenous variables in a structural model. The term *instrumental variable* is often reserved for those exogenous variables that are not included as regressors in the model.

instrumental-variables (IV) estimator. An instrumental variables estimator uses instrumental variables to produce consistent parameter estimates in models that contain endogenous variables. IV estimators can also be used to control for measurement error.

integrated process. A nonstationary process is integrated of order d, written I(d), if the process must be differenced d times to produce a stationary series. An I(1) process y_t is one in which Δy_t is stationary.

interaction effects. Interaction effects measure the dependence of the effects of one factor on the levels of the other factor. Mathematically, they can be defined as the differences among treatment means that are left after main effects are removed from these differences.

intercept. An intercept for the equation of endogenous variable y, observed or latent, is the path coefficient from `_cons` to y. `_cons` is Stata-speak for the built-in variable containing 1 in all observations. In SEM-speak, `_cons` is an observed exogenous variable.

interval data. Interval data are data in which the true value of the dependent variable is not observed. Instead, all that is known is that the value lies within a given interval.

interval hypothesis testing. Interval hypothesis testing performs interval hypothesis tests for model parameters and functions of model parameters.

interval test. In Bayesian analysis, an interval test applied to a scalar model parameter calculates the marginal posterior probability for the parameter to belong to the specified interval.

intraclass correlation. In the context of mixed-effects models, intraclass correlation refers to the correlation for pairs of responses at each nested level of the model.

invariance. When an IRT model fits the data exactly in the population, then the estimated item parameters should be the same, within sampling error, regardless of what sample the data were derived from, and the estimated person latent traits should be the same regardless of what items they are based on.

inverse-probability-weighted estimators. Inverse-probability-weighted (IPW) estimators use weighted averages of the observed outcome variable to estimate the potential-outcome means. The weights are the reciprocals of the treatment probabilities estimated by a treatment model.

inverse-probability-weighted regression-adjustment estimators. Inverse-probability-weighted regression-adjustment (IPWRA) estimators use the reciprocals of the estimated treatment probability as weights to estimate missing-data-corrected regression coefficients that are subsequently used to compute the potential-outcome means.

IPW estimators. See *inverse-probability-weighted estimators*.

IPWRA estimators. See *inverse-probability-weighted regression-adjustment estimators*.

IRT. See *item response theory*.

istmt. An *istmt* is an interactive statement, a statement typed at Mata's colon prompt.

item. An item is a single question or task on a test or an instrument.

item characteristic curve. An item characteristic curve (ICC) expresses the probability for a given response to a binary item as a function of the latent trait.

item information function. An item information function (IIF) indicates the precision of an item along the latent trait continuum.

item location. Location of an item on the difficulty scale.

item response function. See *item characteristic curve*.

item response theory. Item response theory (IRT) is a theoretical framework organized around the concept of the latent trait. IRT encompasses a set of models and associated statistical procedures that relate observed responses on an instrument to a person's level of the latent trait.

iterated principal-factor method. The iterated principal-factor method is a method for performing factor analysis in which the communalities $\widehat{h}_i^2$ are estimated iteratively from the loadings in $\widehat{\boldsymbol{\Lambda}}$ using

$$\widehat{h}_i^2 = \sum_{j=1}^{m} \widehat{\lambda}_{ij}^2$$

Also see *factor analysis* and *communality*.

J × 2 contingency table. A $J \times 2$ contingency table is used to describe the association between an ordinal independent variable with J levels and a binary response variable of interest.

J(r, c, value). `J()` is the function that returns an $r \times c$ matrix with all elements set to *value*; see [M-5] **J()**. Also, `J()` is often used in the documentation to describe the various types of *void* matrices; see *void matrix*. Thus the documentation might say that such-and-such returns `J(0, 0, .)` under certain conditions. That is another way of saying that such-and-such returns a 0×0 real matrix.

When r or c is 0, there are no elements to be filled in with *value*, but even so, *value* is used to determine the type of the matrix. Thus `J(0, 0, 1i)` refers to a 0×0 complex matrix, `J(0, 0, "")` refers to a 0×0 string matrix, and `J(0, 0, NULL)` refers to a 0×0 *pointer* matrix.

In the documentation, `J()` is used for more than describing 0×0 matrices. Sometimes, the matrices being described are $r \times 0$ or are $0 \times c$. Say that a function `example(`X`)` is supposed to return a column vector; perhaps it returns the last column of X. Now say that X is 0×0. Function `example()` still should return a column vector, and so it returns a 0×1 matrix. This would be documented by noting that `example()` returns `J(0, 1, .)` when X is 0×0.

jackknife. The jackknife is a data-dependent way to estimate the variance of a statistic, such as a mean, ratio, or regression coefficient. Unlike BRR, the jackknife can be applied to practically any survey design. The jackknife variance estimator is described in [SVY] **variance estimation**.

jackknife, vce(jackknife). The jackknife is a replication method for obtaining variance estimates. Consider an estimation method E for estimating θ. Let $\widehat{\theta}$ be the result of applying E to dataset D containing N observations. The jackknife is a way of obtaining variance estimates for $\widehat{\theta}$ from repeated estimates $\widehat{\theta}_1, \widehat{\theta}_2, \ldots, \widehat{\theta}_N$, where each $\widehat{\theta}_i$ is the result of applying E to D with observation i removed. See [SEM] **sem option method( )** and [R] **jackknife**.

`vce(jackknife)` is allowed with `sem` but not `gsem`. You can obtain jackknife results by prefixing the `gsem` command with `jackknife:`, but remember to specify `jackknife`'s `cluster()` and `idcluster()` options if you are fitting a multilevel model. See [SEM] **intro 9**.

jackknifed standard error. See *Monte Carlo error*.

JCA. An acronym for joint correspondence analysis; see *multiple correspondence analysis*.

Jeffreys prior. The Jeffreys prior of a vector of model parameters $\boldsymbol{\theta}$ is proportional to the square root of the determinant of its Fisher information matrix $I(\boldsymbol{\theta})$. Jeffreys priors are locally uniform and, by definition, agree with the likelihood function. Jeffreys priors are considered noninformative priors that have minimal impact on the posterior distribution.

joint correspondence analysis. See *multiple correspondence analysis*.

joint normality assumption. See *normality assumption, joint and conditional*.

Kaiser–Meyer–Olkin measure of sampling adequacy. The Kaiser–Meyer–Olkin (KMO) measure of sampling adequacy takes values between 0 and 1, with small values meaning that the variables have too little in common to warrant a factor analysis or PCA. Historically, the following labels have been given to values of KMO (Kaiser 1974):

0.00 to 0.49	unacceptable
0.50 to 0.59	miserable
0.60 to 0.69	mediocre
0.70 to 0.79	middling
0.80 to 0.89	meritorious
0.90 to 1.00	marvelous

Kalman filter. The Kalman filter is a recursive procedure for predicting the state vector in a state-space model.

Kaplan–Meier product-limit estimate. This is an estimate of the survivor function, which is the product of conditional survival to each time at which an event occurs. The simple form of the calculation, which requires tallying the number at risk and the number who die and at each time, makes accounting for censoring easy. The resulting estimate is a step function with jumps at the event times.

Kenward–Roger denominator degrees of freedom (DDF) method. This method implements the Kenward and Roger (1997) method, which is designed to approximate unknown sampling distributions of test statistics for complex linear mixed-effects models. This method is supported only with restricted maximum-likelihood estimation.

kmeans. Kmeans is a method for performing partition cluster analysis. The user specifies the number of clusters, k, to create using an iterative process. Each observation is assigned to the group whose mean is closest, and then based on that categorization, new group means are determined. These steps continue until no observations change groups. The algorithm begins with k seed values, which act as the k group means. There are many ways to specify the beginning seed values. Also see *partition clustering*.

kmedians. Kmedians is a variation of kmeans. The same process is performed, except that medians instead of means are computed to represent the group centers at each step. Also see *kmeans* and *partition clustering*.

KMO. See *Kaiser–Meyer–Olkin measure of sampling adequacy*.

KNN. See *kth nearest neighbor*.

Kruskal stress. The Kruskal stress measure (Kruskal 1964; Cox and Cox 2001, 63) used in MDS is given by

$$\text{Kruskal}(\widehat{\mathbf{D}}, \mathbf{E}) = \left\{ \frac{\sum (E_{ij} - \widehat{D}_{ij})^2}{\sum E_{ij}^2} \right\}^{1/2}$$

where D_{ij} is the dissimilarity between objects i and j, $1 \leq i, j \leq n$, and $\widehat{D}_{ij}$ is the disparity, that is, the transformed dissimilarity, and E_{ij} is the Euclidean distance between rows i and j of the matching configuration. Kruskal stress is an example of a loss function in modern MDS. After classical MDS, `estat stress` gives the Kruskal stress. Also see *classical scaling*, *multidimensional scaling*, and *stress*.

kth nearest neighbor. kth-nearest-neighbor (KNN) discriminant analysis is a nonparametric discrimination method based on the k nearest neighbors of each observation. Both continuous and binary data can be handled through the different similarity and dissimilarity measures. KNN analysis can distinguish irregular-shaped groups, including groups with multiple modes. Also see *discriminant analysis* and *nonparametric methods*.

lag operator. The lag operator L denotes the value of a variable at time $t-1$. Formally, $Ly_t = y_{t-1}$, and $L^2 y_t = Ly_{t-1} = y_{t-2}$.

Lagrange multiplier test. Synonym for *score test*.

LAPACK LAPACK stands for Linear Algebra PACKage and forms the basis for many of Mata's linear algebra capabilities; see [M-1] **LAPACK**.

Laplacian approximation. Laplacian approximation is a technique used to approximate definite integrals without resorting to quadrature methods. In the context of mixed-effects models, Laplacian approximation is as a rule faster than quadrature methods at the cost of producing biased parameter estimates of variance components.

latent growth model. A latent growth model is a kind of measurement model in which the observed values are collected over time and are allowed to follow a trend. See [SEM] **intro 5**.

latent space. Number of latent traits that are measured by an instrument. All IRT models described in this manual assume a unidimensional latent space or, in other words, that a single latent trait explains the response pattern.

latent trait. A variable or construct that cannot be directly observed.

latent variable. A variable is latent if it is not observed. A variable is latent if it is not in your dataset but you wish it were. You wish you had a variable recording the propensity to commit violent crime, or socioeconomic status, or happiness, or true ability, or even income accurately recorded. Latent variables are sometimes described as imagined variables.

In the software, latent variables are usually indicated by having at least their first letter capitalized.

Also see *first- and second-order latent variables*, *first-, second-, and higher-level (latent) variables*, and *observed variables*.

Lawley–Hotelling trace. The Lawley–Hotelling trace is a test statistic for the hypothesis test H_0 : $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \cdots = \boldsymbol{\mu}_k$ based on the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_s$ of $\mathbf{E}^{-1}\mathbf{H}$. It is defined as

$$U^{(s)} = \mathrm{trace}(\mathbf{E}^{-1}\mathbf{H}) = \sum_{i=1}^{s} \lambda_i$$

where $\mathbf{H}$ is the between matrix and $\mathbf{E}$ is the within matrix, see *between matrix*.

LDA. See *linear discriminant analysis*.

leave one out. In discriminant analysis, classification of an observation while leaving it out of the estimation sample is done to check the robustness of the analysis; thus the phrase "leave one out" (LOO). Also see *discriminant analysis*.

left eigenvectors. A vector $\mathbf{x}$: $n \times 1$ is said to be a left eigenvector of square matrix $\mathbf{A}$: $n \times n$ if there is a nonzero scalar, λ, such that

$$\mathbf{x}\mathbf{A} = \lambda\mathbf{x}$$

left-censoring. See *censored, censoring, left-censoring, and right-censoring*.

left-truncation. See *truncation, left-truncation, and right-truncation*.

life table. Also known as a mortality table or actuarial table, a life table is a table that shows for each analysis time the fraction that survive to that time. In mortality tables, analysis time is often age.

likelihood displacement value. A likelihood displacement value is an influence measure of the effect of deleting a subject on the overall coefficient vector. Also see *partial likelihood displacement value*.

likelihood-ratio test. The likelihood-ratio (LR) test is one of the three classical testing procedures used to compare the fit of two models, one of which, the constrained model, is nested within the full (unconstrained) model. Under the null hypothesis, the constrained model fits the data as well as the full model. The LR test requires one to determine the maximal value of the log-likelihood function for both the constrained and the full models. See [PSS] **power twoproportions** and [R] **lrtest**.

linear discriminant analysis. Linear discriminant analysis (LDA) is a parametric form of discriminant analysis. In Fisher's (1936) approach to LDA, linear combinations of the discriminating variables provide maximal separation between the groups. The Mahalanobis (1936) formulation of LDA assumes that the observations come from multivariate normal distributions with equal covariance matrices. Also see *discriminant analysis* and *parametric methods*.

linear filter. A linear filter is a sequence of weights used to compute a weighted average of a time series at each time period. More formally, a linear filter $\alpha(L)$ is

$$\alpha(L) = \alpha_0 + \alpha_1 L + \alpha_2 L^2 + \cdots = \sum_{\tau=0}^{\infty} \alpha_\tau L^\tau$$

where L is the lag operator. Applying the linear filter $\alpha(L)$ to the time series x_t yields a sequence of weighted averages of x_t:

$$\alpha(L)x_t = \sum_{\tau=0}^{\infty} \alpha_\tau L^\tau x_{t-\tau}$$

linear mixed model. See *linear mixed-effects model*.

linear mixed-effects model. A linear mixed-effects model is an extension of a linear model allowing for the inclusion of random deviations (effects).

linear regression. Linear regression is a kind of SEM in which there is a single equation. See [SEM] **intro 5**.

linearization. Linearization is short for Taylor linearization. Also known as the delta method or the Huber/White/robust sandwich variance estimator, linearization is a method for deriving an approximation to the variance of a point estimator, such as a ratio or regression coefficient. The linearized variance estimator is described in [SVY] **variance estimation**.

link function. See *generalized linear response functions*.

linkage. In cluster analysis, the linkage refers to the measure of proximity between groups or clusters.

listwise deletion, **casewise deletion**. Omitting from analysis observations containing missing values.

LMAX value. An LMAX value is an influence measure of the effect of deleting a subject on the overall coefficient vector and is based on an eigensystem analysis of efficient score residuals. Also see *partial LMAX value*.

LME model. See *linear mixed-effects model*.

loading. A loading is a coefficient or weight in a linear transformation. Loadings play an important role in many multivariate techniques, including factor analysis, PCA, MANOVA, LDA, and canonical correlations. In some settings, the loadings are of primary interest and are examined for interpretability. For many multivariate techniques, loadings are based on an eigenanalysis of a correlation or covariance matrix. Also see *eigenvalues and eigenvector*.

loading plot. A loading plot is a scatter plot of the loadings after LDA, factor analysis or PCA.

local independence. See *conditional independence*.

locale. A locale is a code that identifies a community with a certain set of rules for how their language should be written. A locale can refer to something as general as an entire language (for example, "en" for English) or something as specific as a language in a particular country (for example, "en_HK" for English in Hong Kong).

A locale specifies a set of rules that govern how the language should be written. Stata uses locales to determine how certain language-specific operations are carried out. For more information, see [U] **12.4.2.4 Locales in Unicode**.

logistic discriminant analysis. Logistic discriminant analysis is a form of discriminant analysis based on the assumption that the likelihood ratios of the groups have an exponential form. Multinomial logistic regression provides the basis for logistic discriminant analysis. Because multinomial logistic regression can handle binary and continuous regressors, logistic discriminant analysis is also appropriate for binary and continuous discriminating variables. Also see *discriminant analysis*.

logit regression. Logit regression is a term for generalized linear response functions that are family Bernoulli, link logit. It is used for binary outcome data. Logit regression is also known as logistic regression and also simply as logit. See *generalized linear response functions*.

longitudinal data. Longitudinal data is another term for panel data. See also *panel data*.

long-memory process. A long-memory process is a stationary process whose autocorrelations decay at a slower rate than a short-memory process. ARFIMA models are typically used to represent long-memory processes, and ARMA models are typically used to represent short-memory processes.

LOO. See *leave one out*.

loss. Modern MDS is performed by minimizing a loss function, also called a loss criterion. The loss quantifies the difference between the disparities and the Euclidean distances.

Loss functions include Kruskal's stress and its square, both normalized with either disparities or distances, the strain criterion which is equivalent to classical metric scaling when the disparities equal the dissimilarities, and the Sammon (1969) mapping criterion which is the sum of the scaled, squared differences between the distances and the disparities, normalized by the sum of the disparities.

Also see *multidimensional scaling*, *Kruskal stress*, *classical scaling*, and *disparity*.

loss to follow-up. Subjects are lost to follow-up if they do not complete the course of the study for reasons unrelated to the event of interest. For example, loss to follow-up occurs if subjects move to a different area or decide to no longer participate in a study. Loss to follow-up should not be confused with administrative censoring. If subjects are lost to follow-up, the information about the outcome these subjects would have experienced at the end of the study, had they completed the study, is unavailable. Also see *withdrawal*, *administrative censoring*, and *follow-up period or follow-up*.

lower ASCII. See *plain ASCII*.

lower asymptote. See *guessing*.

lower one-sided test, **lower one-tailed test**. A lower one-sided test is a one-sided test of a scalar parameter in which the alternative hypothesis is lower one sided, meaning that the alternative hypothesis states that the parameter is less than the value conjectured under the null hypothesis. Also see *One-sided test versus two-sided test* under *Remarks and examples* in [PSS] **intro**.

lval. *lval* stands for left-hand-side value and is defined as the property of being able to appear on the left-hand side of an equal-assignment operator. Matrices are *lvals* in Mata, and thus

```
X = ...
```

is valid. Functions are not *lvals*; thus, you cannot code

```
substr(mystr,1,3) = "abc"
```

lvals would be easy to describe except that *pointers* can also be lvals. Few people ever use pointers. See [M-2] **op_assignment** for a complete definition.

M, **m**. M is the number of imputations. m refers to a particular imputation, $m = 1, 2, \dots, M$. In `mi`, $m = 0$ is used to refer to the original data, the data containing the missing values. Thus `mi` data in effect contain $M + 1$ datasets, corresponding to $m = 0$, $m = 1$, ..., and $m = M$.

machine precision. See *epsilon(1), etc.*

Mahalanobis distance. The Mahalanobis distance measure is a scale-invariant way of measuring distance. It takes into account the correlations of the dataset.

Mahalanobis transformation. The Mahalanobis transformation takes a Cholesky factorization of the inverse of the covariance matrix $\mathbf{S}^{-1}$ in the formula for Mahalanobis distance and uses it to transform the data. If we have the Cholesky factorization $\mathbf{S}^{-1} = \mathbf{L}'\mathbf{L}$, then the Mahalanobis transformation of $\mathbf{x}$ is $\mathbf{z} = \mathbf{L}\mathbf{x}$, and $\mathbf{z}'\mathbf{z} = D^2_M(\mathbf{x})$.

main effects. These are average, additive effects that are associated with each level of each factor. For example, the main effect of level j of a factor is the difference between the mean of all observations on the outcome of interest at level j and the grand mean.

MANCOVA. MANCOVA is multivariate analysis of covariance. See *multivariate analysis of variance*.

manifest variables. Synonym for *observed variables*.

MANOVA. *multivariate analysis of variance*.

Mantel–Haenszel test. The Mantel–Haenszel test evaluates whether the overall degree of association in stratified 2×2 tables is significant assuming that the exposure effect is the same across strata. See [PSS] **power cmh**.

MAR. See *missing at random*.

marginal distribution. In Bayesian context, a distribution of the data after integrating out parameters from the joint distribution of the parameters and the data.

marginal homogeneity. Marginal homogeneity refers to the equality of one or more row marginal proportions with the corresponding column proportions. Also see *Introduction* under *Remarks and examples* in [PSS] **power pairedproportions**.

marginal likelihood. In the context of Bayesian model comparison, a marginalized over model parameters $\boldsymbol{\theta}$ likelihood of data $\mathbf{y}$ for a given model M, $P(\mathbf{y}|M) = m(\mathbf{y}) = \int P(\mathbf{y}|\boldsymbol{\theta}, M)P(\boldsymbol{\theta}|M)d\boldsymbol{\theta}$. Also see *Bayes factor*.

marginal posterior distribution. In Bayesian context, a marginal posterior distribution is a distribution resulting from integrating out all but one parameter from the joint posterior distribution.

marginal proportion. This represents a ratio of the number of observations in a row or column of a contingency table relative to the total number of observations. Also see *Introduction* under *Remarks and examples* in [PSS] **power pairedproportions**.

Markov chain. Markov chain is a random process that generates sequences of random vectors (or states) and satisfies the Markov property: the next state depends only on the current state and not on any of the previous states. MCMC is the most common methodology for simulating Markov chains.

mass. In CA and MCA, the mass is the marginal probability. The sum of the mass over the active row or column categories equals 1.

.mata file. By convention, we store the Mata source code for function *function*`()` in file *function*`.mata`; see [M-1] **source**.

matched case–control study. Also known as a retrospective study, a matched case–control study is a study in which persons with positive outcomes are each matched with one or more persons with negative outcomes but with similar characteristics.

matched study. In a matched study, an observation from one group is matched to one or more observations from another group with respect to one or more characteristics of interest. When multiple matches occur, the study design is $1 : M$, where M is the number of matches. Also see *paired data*, also known as $1 : 1$ matched data.

matching coefficient. The matching similarity coefficient is used to compare two binary variables. If a is the number of observations that both have value 1, and d is the number of observations that both have value 0, and b, c are the number of $(1,0)$ and $(0,1)$ observations, respectively, then the matching coefficient is given by

$$\frac{a+d}{a+b+c+d}$$

Also see *similarity measure*.

matching configuration. In MDS, the matching configuration is the low dimensional configuration whose distances approximate the high-dimensional dissimilarities or disparities. Also see *multidimensional scaling*, *dissimilarity*, and *disparity*.

matching configuration plot. After MDS, this is a scatter plot of the matching configuration.

matching estimator. An estimator that compares differences between the outcomes of similar—that is, matched—individuals. Each individual that receives a treatment is matched to a similar individual that does not get the treatment, and the difference in their outcomes is used to estimate the individual-level treatment effect. Likewise, each individual that does not receive a treatment is matched to a similar individual that does get the treatment, and the difference in their outcomes is used to estimate the individual-level treatment effect.

matrix. The most general organization of data, containing r rows and c columns. Vectors, column vectors, row vectors, and scalars are special cases of matrices.

matrix model parameter. A matrix model parameter is any model parameter that is a matrix. Matrix elements, however, are viewed as scalar model parameters.

Matrix model parameters are defined and referred to within the `bayesmh` command as `{`*param*`,matrix}` or `{`*eqname*`:`*param*`,matrix}` with the equation name *eqname*. For example, `{Sigma, matrix}` and `{Scale:Omega, matrix}` are matrix model parameters. Individual matrix elements cannot be referred to within the `bayesmh` command, but they can be referred within postestimation commands accepting parameters. For example, to refer to the individual elements of the defined above, say, 2×2 matrices, use `{Sigma_1_1}`, `{Sigma_2_1}`, `{Sigma_1_2}`, `{Sigma_2_2}` and `{Scale:Omega_1_1}`, `{Scale:Omega_2_1}`, `{Scale:Omega_1_2}`, `{Scale:Omega_2_2}`, respectively. See [BAYES] **bayesmh**.

matrix parameter. See *matrix model parameter*.

maximum likelihood factor method. The maximum likelihood factor method is a method for performing factor analysis that assumes multivariate normal observations. It maximizes the determinant of the partial correlation matrix; thus, this solution is also meaningful as a descriptive method for nonnormal data. Also see *factor analysis*.

MCA. See *multiple correspondence analysis*.

MCAGHQ. See *mode-curvature adaptive Gauss–Hermite quadrature*.

MCAR. See *missing completely at random*.

MCE. See *Monte Carlo error*.

MCMC, Markov chain Monte Carlo. MCMC is a class of simulation-based methods for generating samples from probability distributions. Any MCMC algorithm simulates a Markov chain with a target distribution as its stationary or equilibrium distribution. The precision of MCMC algorithms increases with the number of iterations. The lack of a stopping rule and convergence rule, however, makes it difficult to determine for how long to run MCMC. The time needed to converge to the target distribution within a prespecified error is referred to as mixing time. Better MCMC algorithms have faster mixing times. Some of the popular MCMC algorithms are random-walk Metropolis, Metropolis–Hastings, and Gibbs sampling.

MCMC sample. An MCMC sample is obtained from MCMC sampling. An MCMC sample approximates a target distribution and is used for summarizing this distribution.

MCMC sample size. MCMC sample size is the size of the MCMC sample. It is specified in `bayesmh`'s option `mcmcsize()`; see [BAYES] **bayesmh**.

MCMC sampling, MCMC sampler. MCMC sampling is an MCMC algorithm that generates samples from a target probability distribution.

MCMC standard error, MCSE MCSE is the standard error of the posterior mean estimate. It is defined as the standard deviation divided by the square root of ESS. MCSEs are analogs of standard errors in frequentist statistics and measure the accuracy of the simulated MCMC sample.

McNemar's test. McNemar's test is a test used to compare two dependent binary populations. The null hypothesis is formulated in the context of a 2×2 contingency table as a hypothesis of marginal homogeneity. See [PSS] **power pairedproportions** and the `mcc` command in [R] **epitab**.

MDES. See *minimum detectable effect size*.

MDS. See *multidimensional scaling*.

MDS configuration plot. See *configuration plot*.

mean contrasts. See *contrasts*.

mean–variance adaptive Gauss–Hermite quadrature. In the context of generalized linear mixed models, mean–variance adaptive Gauss–Hermite quadrature is a method of approximating the integral used in the calculation of the log likelihood. The quadrature locations and weights for individual clusters are updated during the optimization process by using the posterior mean and the posterior standard deviation.

In the context of IRT models, mean–variance adaptive Gauss–Hermite quadrature (MVAGHQ) is a method of approximating the integral used in the calculation of the log likelihood. The quadrature locations and weights for individuals are updated during the optimization process by using the posterior mean and the posterior standard deviation.

measure. A measure is a quantity representing the proximity between objects or method for determining the proximity between objects. Also see *proximity*.

measure, **measurement**, **x a measurement of X**, **x measures X**. See *measurement variables*.

measurement models, **measurement component**. A measurement model is a particular kind of model that deals with the problem of translating observed values to values suitable for modeling. Measurement models are often combined with structural models and then the measurement model part is referred to as the measurement component. See [SEM] **intro 5**.

measurement variables, **measure**, **measurement**, **x a measurement of X**, **x measures X**. Observed variable x is a measurement of latent variable X if there is a path connecting $x \leftarrow X$. Measurement variables are modeled by measurement models. Measurement variables are also called indicator variables.

median-linkage clustering. Median-linkage clustering is a hierarchical clustering method that uses the distance between the medians of two groups to determine the similarity or dissimilarity of the two groups. Also see *cluster analysis* and *agglomerative hierarchical clustering methods*.

MEFF and MEFT. MEFF and MEFT are misspecification effects. Misspecification effects compare the variance estimate from a given survey dataset with the variance from a misspecified model. In Stata, the misspecified model is fit without weighting, clustering, or stratification.

MEFF is the ratio of two variance estimates. The design-based variance is in the numerator; the misspecified variance is in the denominator.

MEFT is the ratio of two standard-error estimates. The design-based standard error is in the numerator; the misspecified standard error is in the denominator. MEFT is the square root of MEFF.

method. Method is just an English word and should be read in context. Nonetheless, method is used here usually to refer to the method used to solve for the fitted parameters of an SEM. Those methods are ML, QML, MLMV, and ADF. Also see *technique*.

metric scaling. Metric scaling is a type of MDS, in which the dissimilarities are transformed to disparities via a class of known functions. This is contrasted to *nonmetric scaling*. Also see *multidimensional scaling*.

Metropolis–Hastings (MH) sampling, MH sampler. A Metropolis–Hastings (MH) sampler is an MCMC method for simulating probability distributions. According to this method, at each step of the Markov chain, a new proposal state is generated from the current state according to a prespecified proposal distribution. Based on the current and new state, an acceptance probability is calculated and then used to accept or reject the proposed state. Important characteristics of MH sampling is the acceptance rate and mixing time. The MH algorithm is very general and can be applied to an arbitrary target distribution. However, its efficiency is limited, in terms of mixing time, and decreases as the dimension of the target distribution increases. Gibbs sampling, when available, can provide much more efficient sampling than MH sampling.

mi data. Any data that have been `mi set` (see [MI] **mi set**), whether directly by `mi set` or indirectly by `mi import` (see [MI] **mi import**). The `mi` data might have no imputations (have $M = 0$) and no imputed variables, at least yet, or they might have $M > 0$ and no imputed variables, or vice versa. An `mi` dataset might have $M > 0$ and imputed variables, but the missing values have not yet been replaced with imputed values. Or `mi` data might have $M > 0$ and imputed variables and the missing values of the imputed variables filled in with imputed values.

MIMIC. See *multiple indicators and multiple causes*.

minimum detectable effect size. The minimum detectable effect size is the smallest effect size that can be detected by hypothesis testing for a given power and sample size.

minimum detectable value. The minimum detectable value represents the smallest amount or concentration of a substance that can be reliably measured.

minimum entropy rotation. The minimum entropy rotation is an orthogonal rotation achieved by minimizing the deviation from uniformity (entropy). The minimum entropy criterion (Jennrich 2004) is

$$c(\mathbf{\Lambda}) = -\frac{1}{2}\left\langle \mathbf{\Lambda}^2, \log \mathbf{\Lambda}^2 \right\rangle$$

See *Crawford–Ferguson rotation* for a definition of $\mathbf{\Lambda}$. Also see *orthogonal rotation*.

misclassification rate. The misclassification rate calculated after discriminant analysis is, in its simplest form, the fraction of observations incorrectly classified. See *discriminant analysis*.

missing at random. Missing data are said to be missing at random (MAR) if the probability that data are missing does not depend on unobserved data but may depend on observed data. Under MAR, the missing-data values do not contain any additional information given observed data about the missing-data mechanism. Thus the process that causes missing data can be ignored.

missing completely at random. Missing data are said to be missing completely at random (MCAR) if the probability that data are missing does not depend on observed or unobserved data. Under MCAR, the missing data values are a simple random sample of all data values, so any analysis that discards the missing values remains consistent, albeit perhaps inefficient.

missing not at random. Missing data are missing not at random (MNAR) if the probability that data are missing depends on unobserved data. Under MNAR, a missing-data mechanism (the process that causes missing data) must be modeled to obtain valid results.

misspecification effects. See *MEFF* and *MEFT*.

mixed design. A mixed design is an experiment that has at least one between-subjects factor and one within-subject factor. See [PSS] **power repeated**.

mixed model. See *mixed-effects model*.

mixed-effects model. A mixed-effects model contains both fixed and random effects. The fixed effects are estimated directly, whereas the random effects are summarized according to their (co)variances. Mixed-effects models are used primarily to perform estimation and inference on the regression coefficients in the presence of complicated within-subject correlation structures induced by multiple levels of grouping.

mixing of Markov chain. Mixing refers to the rate at which a Markov chain traverses the parameter space. It is a property of the Markov chain that is different from convergence. Poor mixing indicates a slow rate at which the chain explores the stationary distribution and will require more iterations to provide inference at a given precision. Poor (slow) mixing is typically a result of high correlation between model parameters or of weakly-defined model specifications.

ML, **method(ml)**. ML stands for maximum likelihood. It is a method to obtain fitted parameters. ML is the default method used by `sem` and `gsem`. Other available methods for `sem` are QML, MLMV, and ADF. Also available for `gsem` is QML.

.mlib library. The object code of functions can be collected and stored in a library. Most Mata functions, in fact, are located in the official libraries provided with Stata. You can create your own libraries. See [M-3] **mata mlib**.

MLMV, **method(mlmv)**. MLMV stands for maximum likelihood with missing values. It is an ML method used to obtain fitted parameters in the presence of missing values. MLMV is the method used by `sem` when the `method(mlmv)` option is specified; `method(mlmv)` is not available with `gsem`. Other available methods for use with `sem` are ML, QML, and ADF. These methods omit from the calculation observations that contain missing values.

mlong data. See *style*.

MNAR. See *missing not at random*.

.mo file. The object code of a function can be stored in a `.mo` file, where it can be later reused. See [M-1] **how** and [M-3] **mata mosave**.

mode-curvature adaptive Gauss–Hermite quadrature. In the context of generalized linear mixed models, mode-curvature adaptive Gauss–Hermite quadrature is a method of approximating the integral used in the calculation of the log likelihood. The quadrature locations and weights for individual clusters are updated during the optimization process by using the posterior mode and the standard deviation of the normal density that approximates the log posterior at the mode.

In the context of IRT models, mode-curvature adaptive Gauss–Hermite quadrature (MCAGHQ) is a method of approximating the integral used in the calculation of the log likelihood. The quadrature locations and weights for individuals are updated during the optimization process by using the posterior mode and the standard deviation of the normal density that approximates the log posterior at the mode.

model hypothesis testing. Model hypothesis testing tests hypotheses about models by computing model posterior probabilities.

model parameter. A model parameter refers to any (random) parameter in a Bayesian model. Model parameters can be scalars or matrices. Examples of model parameters as defined in `bayesmh` are `{mu}`, `{scale:s}`, `{Sigma,matrix}`, and `{Scale:Omega,matrix}`. See [BAYES] **bayesmh** and, specifically, *Declaring model parameters* and *Referring to model parameters* in that entry. Also see *Different ways of specifying model parameters* in [BAYES] **bayesmh postestimation**.

model posterior probability. Model posterior probability is probability of a model M computed conditional on the observed data $\mathbf{y}$,

$$P(M|\mathbf{y}) = P(M)P(\mathbf{y}|M) = P(M)m(\mathbf{y})$$

where $P(M)$ is the prior probability of a model M and $m(\mathbf{y})$ is the marginal likelihood under model M.

modern scaling. Modern scaling is a form of MDS that is achieved via the minimization of a loss function that compares the disparities (transformed dissimilarities) in the higher-dimensional space and the distances in the lower-dimensional space. Contrast to *classical scaling*. Also see *dissimilarity*, *disparity*, *multidimensional scaling*, and *loss*.

modification indices. Modification indices are score tests for adding paths where none appear. The paths can be for either coefficients or covariances.

moments (of a distribution). The moments of a distribution are the expected values of powers of a random variable or centralized (demeaned) powers of a random variable. The first moments are the expected or observed means, and the second moments are the expected or observed variances and covariances.

monadic operator. Synonym for unary operator.

monotone-missing pattern, **monotone missingness**. A special pattern of missing values in which if the variables are ordered from least to most missing, then all observations of a variable contain missing in the observations in which the prior variable contains missing.

Monte Carlo error. Within the multiple-imputation context, a Monte Carlo error is defined as the standard deviation of the multiple-imputation results across repeated runs of the same imputation procedure using the same data. The Monte Carlo error is useful for evaluating the statistical reproducibility of multiple-imputation results. See *Example 6: Monte Carlo error estimates* under *Remarks and examples* of [MI] **mi estimate**.

moving-average process. A moving-average process is a time-series process in which the current value of a variable is modeled as a weighted average of current and past realizations of a white-noise process and, optionally, a time-invariant constant. By convention, the weight on the current realization of the white-noise process is equal to one, and the weights on the past realizations are known as the moving-average (MA) coefficients. A first-order moving-average process, denoted as an MA(1) process, is $y_t = \theta\epsilon_{t-1} + \epsilon_t$.

multiarm trial. A multiarm trial is a trial comparing survivor functions of more than two groups.

multidimensional scaling. Multidimensional scaling (MDS) is a dimension-reduction and visualization technique. Dissimilarities (for instance, Euclidean distances) between observations in a high-dimensional space are represented in a lower-dimensional space which is typically two dimensions so that the Euclidean distance in the lower-dimensional space approximates in some sense the dissimilarities in the higher-dimensional space. Often the higher-dimensional dissimilarities are first transformed to disparities, and the disparities are then approximated by the distances in the lower-dimensional space. Also see *dissimilarity*, *disparity*, *classical scaling*, *loss*, *modern scaling*, *metric scaling*, and *nonmetric scaling*.

multilevel models. Multilevel models are models that include unobserved effects (latent variables) for different groups in the data. For instance, in a dataset of students, groups of students might share the same teacher. If the teacher's identity is recorded in the data, then one can introduce a latent variable that is constant within teacher and that varies across teachers. This is called a two-level model.

If teachers could in turn be grouped into schools, and school identities were recorded in the data, then one can introduce another latent variable that is constant within school and varies across schools. This is called a three-level (nested-effects) model.

In the above example, observations (students) are said to be nested within teacher nested within school. Sometimes there is no such subsequent nesting structure. Consider workers nested within occupation and industry. The same occupations appear in various industries and the same industries appear within various occupations. We can still introduce latent variables at the occupation and industry level. In such cases, the model is called a crossed-effects model.

The latent variables that we have discussed are also known as random effects. Any coefficients on observed variables in the model are known as the fixed portion of the model. Models that contain fixed and random portions are known as mixed-effects models.

multinomial logit regression. Multinomial logit regression is a term for generalized linear response functions that are family multinomial, link logit. It is used for categorical-outcome data when the outcomes cannot be ordered. Multinomial logit regression is also known as multinomial logistic regression and as mlogit in Stata circles. See *generalized linear response functions*.

multiple correlation. The multiple correlation is the correlation between endogenous variable y and its linear prediction.

multiple correspondence analysis. Multiple correspondence analysis (MCA) and joint correspondence analysis (JCA) are methods for analyzing observations on categorical variables. MCA and JCA analyze a multiway table and are usually viewed as an extension of CA. Also see *correspondence analysis*.

multiple indicators and multiple causes. Multiple indicators and multiple causes is a kind of structural model in which observed causes determine a latent variable, which in turn determines multiple indicators. See [SEM] **intro 4**.

multiple-record st data. See *st data*.

multivalued treatment effect. A multivalued treatment refers to a treatment that has more than two values. For example, a person could have taken a 20 mg dose of a drug, a 40 mg dose of the drug, or not taken the drug at all.

multivariate analysis of covariance. See *multivariate analysis of variance*.

multivariate analysis of variance. Multivariate analysis of variance (MANOVA) is used to test hypotheses about means. Four multivariate statistics are commonly computed in MANOVA: Wilks's lambda, Pillai's trace, Lawley–Hotelling trace, and Roy's largest root. Also see *Wilks's lambda*, *Pillai's trace*, *Lawley–Hotelling trace*, and *Roy's largest root*.

multivariate GARCH models. Multivariate GARCH models are multivariate time-series models in which the conditional covariance matrix of the errors depends on its own past and its past shocks. The acute trade-off between parsimony and flexibility has given rise to a plethora of models; see [TS] **mgarch**.

multivariate regression. A multivariate regression is a linear regression model in which the regressand is vector valued. Equivalently, a mutivariate regression is a linear regression model in which multiple left-hand-side variables are regressed on the same set of explanatory variables simultaneously, allowing the disturbance terms to be contemporaneously correlated. Multivariate regression is a special case of *seemingly unrelated regression* in which all equations share the same set of explanatory variables.

MVAGHQ. See *mean–variance adaptive Gauss–Hermite quadrature*.

NaN. NaN stands for Not a Number and is a special computer floating-point code used for results that cannot be calculated. Mata (and Stata) do not use NaNs. When NaNs arise, they are converted into . (missing value).

nearest neighbor. See *kth nearest neighbor*.

nearest-neighbor matching. Nearest-neighbor matching uses the distance between observed variables to find similar individuals.

negative binomial regression. Negative binomial regression is a term for generalized linear response functions that are family negative binomial, link log. It is used for count data that are overdispersed relative to Poisson. Negative binomial regression is also known as nbreg in Stata circles. See *generalized linear response functions*.

negative binomial regression model. The negative binomial regression model is for applications in which the dependent variable represents the number of times an event occurs. The negative binomial regression model is an alternative to the Poisson model for use when the dependent variable is overdispersed, meaning that the variance of the dependent variable is greater than its mean.

negative effect size. In power and sample-size analysis, we obtain a negative effect size when the postulated value of the parameter under the alternative hypothesis is less than the hypothesized value of the parameter under the null hypothesis. Also see *positive effect size*.

nested random effects. In the context of mixed-effects models, nested random effects refer to the nested grouping factors for the random effects. For example, we may have data on students who are nested in classes that are nested in schools.

nested-effects models. See *multilevel models*.

Newey–West covariance matrix. The Newey–West covariance matrix is a member of the class of heteroskedasticity- and autocorrelation-consistent (HAC) covariance matrix estimators used with time-series data that produces covariance estimates that are robust to both arbitrary heteroskedasticity and autocorrelation up to a prespecified lag.

nominal alpha, **nominal significance level**. This is a desired or requested significance level.

nominal item. A nominal item is an item scored in categories that have no natural ordering.

nominal response model. The nominal response model (NRM) is an IRT model for nominal responses. The categories within each item vary in their difficulty and discrimination.

noncentrality parameter. In power and sample-size analysis, a noncentrality parameter is the expected value of the test statistic under the alternative hypothesis.

nondirectional test. See *two-sided test*.

noninformative prior. A noninformative prior is a prior with negligible influence on the posterior distribution. See, for example, *Jeffreys prior*.

nonmetric scaling. Nonmetric scaling is a type of modern MDS in which the dissimilarities may be transformed to disparities via any monotonic function as opposed to a class of known functions. Contrast to *metric scaling*. Also see *multidimensional scaling*, *dissimilarity*, *disparity*, and *modern scaling*.

nonparametric methods. Nonparametric statistical methods, such as KNN discriminant analysis, do not assume the population fits any parameterized distribution.

nonrecursive (structural) model (system), recursive (structural) model (system). A structural model (system) is said to be nonrecursive if there are paths in both directions between one or more pairs of endogenous variables. A system is recursive if it is a system—it has endogenous variables that appear with paths from them—and it is not nonrecursive.

A nonrecursive model may be unstable. Consider, for instance,

$$\begin{aligned} y_1 &= 2y_2 + 1x_1 + e_1 \\ y_2 &= 3y_1 - 2x_2 + e_2 \end{aligned}$$

This model is unstable. To see this, without loss of generality, treat $x_1 + e_1$ and $2x_2 + e_2$ as if they were both 0. Consider $y_1 = 1$ and $y_2 = 1$. Those values result in new values $y_1 = 2$ and $y_2 = 3$, and those result in new values $y_1 = 6$ and $y_2 = 6$, and those result in new values, ... Continue in this manner, and you reach infinity for both endogenous variables. In the jargon of the mathematics used to check for this property, the eigenvalues of the coefficient matrix lie outside the unit circle.

On the other hand, consider these values:

$$\begin{aligned} y_1 =& 0.5y_2 + 1x_1 + e_1 \\ y_2 =& 1.0y_1 - 2x_2 + e_2 \end{aligned}$$

These results are stable in that the resulting values converge to $y_1 = 0$ and $y_2 = 0$. In the jargon of the mathematics used to check for this property, the eigenvalues of the coefficients matrix lie inside the unit circle. Finally, consider the values

$$\begin{aligned} y_1 =& 0.5y_2 + 1x_1 + e_1 \\ y_2 =& 2.0y_1 - 2x_2 + e_2 \end{aligned}$$

Start with $y_1 = 1$ and $y_2 = 1$ and that yields new values $y_1 = 0.5$ and $y_2 = 2$ and that yields new values $y_1 = 1$ and $y_2 = 1$, and that yields $y_1 = 0.5$ and $y_2 = 2$, and it will oscillate forever. In the jargon of the mathematics used to check for this property, the eigenvalues of the coefficients matrix lie on the unit circle. These coefficients are also considered to be unstable.

nonsphericity correction. This is a correction used for the degrees of freedom of a regular F test in a repeated-measures ANOVA to compensate for the lack of sphericity of the repeated-measures covariance matrix.

norm. A norm is a real-valued function $f(x)$ satisfying

$$\begin{array}{lcll} f(0) & = & 0 & \\ f(x) & > & 0 & \text{for all } x \neq 0 \\ f(cx) & = & |c|f(x) & \\ f(x+y) & \leq & f(x) + f(y) & \end{array}$$

The word *norm* applied to a vector x usually refers to its Euclidean norm, $p = 2$ norm, or length: the square root of the sum of its squared elements. The are other norms, the popular ones being $p = 1$ (the sum of the absolute values of its elements) and $p =$ infinity (the maximum element). Norms can also be generalized to deal with matrices. See [M-5] **norm()**.

normality assumption, joint and conditional. The derivation of the standard, linear SEM estimator usually assumes the full joint normality of the observed and latent variables. However, full joint normality can replace the assumption of normality conditional on the values of the exogenous variables, and all that is lost is one goodness-of-fit test (the test reported by `sem` on the output) and the justification for use of optional method MLMV for dealing with missing values. This substitution of assumptions is important for researchers who cannot reasonably assume normality of the observed variables. This includes any researcher including, say, variables age and age-squared in his or her model.

Meanwhile, the generalized SEM makes only the conditional normality assumption.

Be aware that even though the full joint normality assumption is not required for the standard linear SEM, `sem` calculates the log-likelihood value under that assumption. This is irrelevant except that log-likelihood values reported by `sem` cannot be compared with log-likelihood values reported by `gsem`, which makes the lesser assumption.

See [SEM] **intro 4**.

normalization. Normalization presents information in a standard form for interpretation. In CA the row and column coordinates can be normalized in different ways depending on how one wishes to interpret the data. Normalization is also used in rotation, MDS, and MCA.

normalization constraints. See *identification*.

normalized residuals. See *standardized residuals*.

NRM. See *nominal response model*.

NULL. A special value for a *pointer* that means "points to nothing". If you list the contents of a pointer variable that contains `NULL`, the address will show as `0x0`. See *pointer*.

null hypothesis. In hypothesis testing, the null hypothesis typically represents the conjecture that one is attempting to disprove. Often the null hypothesis is that a treatment has no effect or that a statistic is equal across populations.

null value, null parameter. This value of the parameter of interest under the null hypothesis is fixed by the investigator in a power and sample-size analysis. For example, null mean value and null mean refer to the value of the mean parameter under the null hypothesis.

null-terminator. See *binary 0*.

numeric. A matrix is said to be numeric if its elements are real or complex; see *type, eltype, and orgtype*.

object code. Object code refers to the binary code that Mata produces from the source code you type as input. See [M-1] **how**.

objective prior. See *noninformative prior*.

object-oriented programming. Object-oriented programming is a programming concept that treats programming elements as objects and concentrates on actions affecting those objects rather than merely on lists of instructions. Object-oriented programming uses classes to describe objects. Classes are much like structures with a primary difference being that classes can contain functions (known as methods) as well as variables. Unlike structures, however, classes may inherit variables and functions from other classes, which in theory makes object-oriented programs easier to extend and modify than non–object-oriented programs.

oblimax rotation. Oblimax rotation is a method for oblique rotation which maximizes the number of high and low loadings. When restricted to orthogonal rotation, oblimax is equivalent to quartimax rotation. Oblimax minimizes the oblimax criterion

$$c(\boldsymbol{\Lambda}) = -\log(\langle \boldsymbol{\Lambda}^2, \boldsymbol{\Lambda}^2 \rangle) + 2\log(\langle \boldsymbol{\Lambda}, \boldsymbol{\Lambda} \rangle)$$

See *Crawford–Ferguson rotation* for a definition of $\boldsymbol{\Lambda}$. Also see *oblique rotation*, *orthogonal rotation*, and *quartimax rotation*.

oblimin rotation. Oblimin rotation is a general method for oblique rotation, achieved by minimizing the oblimin criterion

$$c(\boldsymbol{\Lambda}) = \frac{1}{4} \left\langle \boldsymbol{\Lambda}^2, \{\mathbf{I} - (\gamma/p)\mathbf{1}\mathbf{1}'\} \boldsymbol{\Lambda}^2 (\mathbf{1}\mathbf{1}' - \mathbf{I}) \right\rangle$$

Oblimin has several interesting special cases:

γ	Special case
0	quartimax / quartimin
1/2	biquartimax / biquartimin
1	varimax / covarimin
$p/2$	equamax

p = number of rows of $\mathbf{A}$.

See *Crawford–Ferguson rotation* for a definition of $\boldsymbol{\Lambda}$ and $\mathbf{A}$. Also see *oblique rotation*.

oblique rotation or **oblique transformation**. An oblique rotation maintains the norms of the rows of the matrix but not their inner products. In geometric terms, this maintains the lengths of vectors, but not the angles between them. In contrast, in orthogonal rotation, both are preserved.

observational data. In observational data, treatment assignment is not controlled by those who collected the data; thus some common variables affect treatment assignment and treatment-specific outcomes.

observational study. In an observational study, as opposed to an experimental study, the assignment of subjects to treatments happens naturally and is thus beyond the control of investigators. Investigators can only observe subjects and measure their characteristics. For example, a study that evaluates the effect of exposure of children to household pesticides is an observational study.

observations and variables. A dataset containing n observations on k variables in often stored in an $n \times k$ matrix. An observation refers to a row of that matrix; a variable refers to a column.

observed level of significance. See *p-value*.

observed variables. A variable is observed if it is a variable in your dataset. In this documentation, we often refer to observed variables by using `x1`, `x2`, ..., `y1`, `y2`, and so on; in reality, observed variables have names such as `mpg`, `weight`, `testscore`, etc.

In the software, observed variables are usually indicated by having names that are all lowercase.

Also see *latent variable*.

odds and **odds ratio**. The odds in favor of an event are $o = p/(1-p)$, where p is the probability of the event. Thus if $p = 0.2$, the odds are 0.25, and if $p = 0.8$, the odds are 4.

The log of the odds is $\ln(o) = \text{logit}(p) = \ln\{p/(1-p)\}$, and logistic-regression models, for instance, fit $\ln(o)$ as a linear function of the covariates.

The odds ratio is a ratio of two odds: o_1/o_0. The individual odds that appear in the ratio are usually for an experimental group and a control group, or two different demographic groups.

offset variable and **exposure variable**. An offset variable is a variable that is to appear on the right-hand side of a model with coefficient 1:

$$y_j = \text{offset}_j + b_0 + b_1 x_j + \cdots$$

In the above, b_0 and b_1 are to be estimated. The offset is not constant. Offset variables are often included to account for the amount of exposure. Consider a model where the number of events observed over a period is the length of the period multiplied by the number of events expected in a unit of time:

$$n_j = T_j e(X_j)$$

When we take logs, this becomes

$$\log(n_j) = \log(T_j) + \log\{e(X_j)\}$$

$\ln(T_j)$ is an offset variable in this model.

When the log of a variable is an offset variable, the variable is said to be an exposure variable. In the above, T_j is an exposure variable.

OIM, vce(oim). OIM stands for observed information matrix, defined as the inverse of the negative of the matrix of second derivatives, usually of the log likelihood function. The OIM is an estimate of the VCE. OIM is the default VCE that `sem` and `gsem` report. The other available techniques are EIM, OPG, robust, clustered, bootstrap, and jackknife.

one-at-a-time MCMC sampling. A one-at-a-time MCMC sample is an MCMC sampling procedure in which random variables are sampled individually, one at a time. For example, in Gibbs sampling, individual variates are sampled one at a time, conditionally on the most recent values of the rest of the variates.

one-level model. A one-level model has no multilevel structure and no random effects. Linear regression is a one-level model.

one-parameter logistic model. The one-parameter logistic (1PL) model is an IRT model for binary responses where items vary in their difficulty but share the same discrimination parameter.

one-sample test. A one-sample test compares a parameter of interest from one sample with a reference value. For example, a one-sample mean test compares a mean of the sample with a reference value.

one-sided test, one-tailed test. A one-sided test is a hypothesis test of a scalar parameter in which the alternative hypothesis is one sided, meaning that the alternative hypothesis states that the parameter is either less than or greater than the value conjectured under the null hypothesis but not both. Also see *One-sided test versus two-sided test* under *Remarks and examples* in [PSS] **intro**.

one-step-ahead forecast. See *static forecast*.

one-way ANOVA, one-way analysis of variance. A one-way ANOVA model has a single factor. Also see [PSS] **power oneway**.

one-way repeated-measures ANOVA. A one-way repeated-measures ANOVA model has a single within-subject factor. Also see [PSS] **power repeated**.

operating characteristic curve. See *category characteristic curve*.

operator. An operator is `+`, `-`, and the like. Most operators are binary (or dyadic), such as `+` in *A*`+`*B* and `*` in *C*`*`*D*. Binary operators also include logical operators such as `&` and `|` ("and" and "or") in *E*`&`*F* and *G*`|`*H*. Other operators are unary (or monadic), such as `!` (not) in `!`*J*, or both unary and binary, such as `-` in `-`*K* and in *L*`-`*M*. When we say "operator" without specifying which, we mean binary operator. Thus colon operators are in fact colon binary operators. See [M-2] **exp**.

OPG, vce(opg). OPG stands for outer product of the gradients, defined as the cross product of the observation-level first derivatives, usually of the log likelihood function. The OPG is an estimate of the VCE. The other available techniques are OIM, EIM, robust, clustered, bootstrap, and jackknife.

optimization. Mata compiles the code that you write. After compilation, Mata performs an *optimization* step, the purpose of which is to make the compiled code execute more quickly. You can turn off the optimization step—see [M-3] **mata set**—but doing so is not recommended.

ordered complementary log-log regression. Ordered complementary log-log regression is a term for generalized linear response functions that are family ordinal, link cloglog. It is used for ordinal-outcome data. Ordered complementary log-log regression is also known as ocloglog in Stata circles. See *generalized linear response functions*.

ordered logit regression. Ordered logit regression is a term for generalized linear response functions that are family ordinal, link logit. It is used for ordinal outcome data. Ordered logit regression is also known as ordered logistic regression, as just ordered logit, and as ologit in Stata circles. See *generalized linear response functions*.

ordered probit regression. Ordered probit regression is a term for generalized linear response functions that are family ordinal, link probit. It is used for ordinal-outcome data. Ordered probit regression is also known as just ordered probit and known as oprobit in Stata circles. See *generalized linear response functions*.

ordinal item. An ordinal item is an item scored on a scale where a higher score indicates a "higher" outcome.

ordination. Ordination is the ordering of a set of data points with respect to one or more axes. MDS is a form of ordination.

orgtype. See *type, eltype, and orgtype*.

original data. Original data are the data as originally collected, with missing values in place. In `mi` data, the original data are stored in $m = 0$. The original data can be extracted from `mi` data by using `mi extract`; see [MI] **mi extract**.

orthogonal matrix and unitary matrix. A is orthogonal if A is *square* and $A'A$==I. The word orthogonal is usually reserved for real matrices; if the matrix is complex, it is said to be unitary (and then transpose means conjugate-transpose). We use the word orthogonal for both real and complex matrices.

If A is orthogonal, then $\det(A) = \pm 1$.

orthogonal rotation or **orthogonal transformation**. Orthogonal rotation maintains both the norms of the rows of the matrix and also inner products of the rows of the matrix. In geometric terms, this maintains both the lengths of vectors and the angles between them. In contrast, oblique rotation maintains only the norms, that is, the lengths of vectors.

orthogonalized impulse–response function. An orthogonalized impulse–response function (OIRF) measures the effect of an orthogonalized shock to an endogenous variable on itself or another endogenous variable. An orthogonalized shock is one that affects one variable at time t but no other variables. See [TS] **irf create** for a discussion of the difference between IRFs and OIRFs.

outcome model. An outcome model is a model used to predict the outcome as a function of covariates and parameters.

overdispersion. In count-data models, overdispersion occurs when there is more variation in the data than would be expected if the process were Poisson.

overidentifying restrictions. The order condition for model identification requires that the number of exogenous variables excluded from the model be at least as great as the number of endogenous regressors. When the number of excluded exogenous variables exceeds the number of endogenous regressors, the model is overidentified, and the validity of the instruments can then be checked via a test of overidentifying restrictions.

overlap assumption. The overlap assumption requires that each individual have a positive probability of each possible treatment level.

paired data. Paired data consist of pairs of observations that share some characteristics of interest. For example, measurements on twins, pretest and posttest measurements, before and after measurements, repeated measurements on the same individual. Paired data are correlated and thus must be analyzed by using a paired test.

paired observations. See *paired data*.

paired test. A paired test is used to test whether the parameters of interest of two paired populations are equal. The test takes into account the dependence between measurements. For this reason, paired tests are usually more powerful than their two-sample counterparts. For example, a paired-means or paired-difference test is used to test whether the means of two paired (correlated) populations are equal.

panel data. Panel data are data in which the same units were observed over multiple periods. The units, called panels, are often firms, households, or patients who were observed at several points in time. In a typical panel dataset, the number of panels is large, and the number of observations per panel is relatively small.

panel-corrected standard errors (PCSEs). The term *panel-corrected standard errors* refers to a class of estimators for the variance–covariance matrix of the OLS estimator when there are relatively few panels with many observations per panel. PCSEs account for heteroskedasticity, autocorrelation, or cross-sectional correlation.

parameter constraints. Parameter constraints are restrictions placed on the parameters of the model. These constraints are typically in the form of 0 constraints and equality constraints. A 0 constraint is implied, for instance, when no path is drawn connecting x with y. An equality constraint is specified when one path coefficient is forced to be equal to another or one covariance is forced to be equal to another.

Also see *identification*.

parameters, ancillary parameters. The parameters are the to-be-estimated coefficients of a model. These include all path coefficients, means, variances, and covariances. In mathematical notation, the theoretical parameters are often written as $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\alpha}$ is the vector of intercepts, $\boldsymbol{\beta}$ is the vector of path coefficients, $\boldsymbol{\mu}$ is the vector of means, and $\boldsymbol{\Sigma}$ is the matrix of variances and covariances. The resulting parameters estimates are written as $\widehat{\boldsymbol{\theta}}$.

Ancillary parameters are extra parameters beyond the ones just described that concern the distribution. These include the scale parameter of gamma regression, the dispersion parameter for negative binomial regression, and the cutpoints for ordered probit, logit, and cloglog regression, and the like. These parameters are also included in θ.

parametric methods. Parametric statistical methods, such as LDA and QDA, assume the population fits a parameterized distribution. For example, for LDA we assume the groups are multivariate normal with equal covariance matrices.

parsimax rotation. Parsimax rotation is an orthogonal rotation that balances complexity between the rows and the columns. It is equivalent to the Crawford–Ferguson family with $\kappa = (f-1)/(p+f-2)$, where p is the number of rows of the original matrix, and f is the number of columns. See *orthogonal rotation* and *Crawford–Ferguson rotation*.

partial autocorrelation function. The partial autocorrelation function (PACF) expresses the correlation between periods t and $t-k$ of a time series as a function of the time t and lag k, after controlling for the effects of intervening lags. For a stationary time series, the PACF does not depend on t. The PACF is not symmetric about $k=0$: the partial autocorrelation between y_t and y_{t-k} is not equal to the partial autocorrelation between y_t and y_{t+k}.

partial credit model. The partial credit model (PCM) is an IRT model for ordinal responses. The categories across all items vary in their difficulty and share the same discrimination parameter.

partial DFBETA. A partial DFBETA measures the change in the regressor's coefficient because of deletion of that individual record. In single-record data, the partial DFBETA is equal to the DFBETA. Also see *DFBETA*.

partial likelihood displacement value. A partial likelihood displacement value is an influence measure of the effect of deleting an individual record on the coefficient vector. For single-record data, the partial likelihood displacement value is equal to the likelihood displacement value. Also see *likelihood displacement value*.

partial LMAX value. A partial LMAX value is an influence measure of the effect of deleting an individual record on the overall coefficient vector and is based on an eigensystem analysis of efficient score residuals. In single-record data, the partial LMAX value is equal to the LMAX value. Also see *LMAX value*.

partially specified target rotation. Partially specified target rotation minimizes the criterion

$$c(\mathbf{\Lambda}) = \|\mathbf{W} \otimes (\mathbf{\Lambda} - \mathbf{H})\|^2$$

for a given target matrix $\mathbf{H}$ and a nonnegative weighting matrix $\mathbf{W}$ (usually zero–one valued). See *Crawford–Ferguson rotation* for a definition of $\mathbf{\Lambda}$.

partition clustering and **partition cluster-analysis methods**. Partition clustering methods break the observations into a distinct number of nonoverlapping groups. This is accomplished in one step, unlike hierarchical cluster-analysis methods, in which an iterative procedure is used. Consequently, this method is quicker and will allow larger datasets than the hierarchical clustering methods. Contrast to *hierarchical clustering*. Also see *kmeans* and *kmedians*.

passive variable. See *imputed, passive, and regular variables*.

past history. Past history is information recorded about a subject before the subject was both *at risk* and *under observation*. Consider a dataset that contains information on subjects from birth to death and an analysis in which subjects became at risk once diagnosed with a particular kind of cancer. The past history on the subject would then refer to records before the subjects were diagnosed.

The word *history* is often dropped, and the term becomes simply *past*. For instance, we might want to know whether a subject smoked in the past.

Also see *future history*.

path. A path, typically shown as an arrow drawn from one variable to another, states that the first variable determines the second variable, at least partially. If $x \rightarrow y$, or equivalently $y \leftarrow x$, then $y_j = \alpha + \cdots + \beta x_j + \cdots + e.y_j$, where β is said to be the $x \rightarrow y$ path coefficient. The ellipses are included to account for paths to y from other variables. α is said to be the intercept and is automatically added when the first path to y is specified.

A curved path is a curved line connecting two variables, and it specifies that the two variables are allowed to be correlated. If there is no curved path between variables, the variables are usually assumed to be uncorrelated. We say usually because correlation is assumed among observed exogenous variables and, in the command language, assumed among latent exogenous variables, and if some of the correlations are not desired, they must be suppressed. Many authors refer to covariances rather than correlations. Strictly speaking, the curved path denotes a nonzero covariance. A correlation is often called a standardized covariance.

A curved path can connect a variable to itself and in that case, indicates a variance. In path diagrams in this manual, we typically do not show such variance paths even though variances are assumed.

path coefficient. The path coefficient is associated with a path; see *path*. Also see *intercept*.

path diagram. A path diagram is a graphical representation that shows the relationships among a set of variables using *paths*. See [SEM] **intro 2** for a description of path diagrams.

path notation. Path notation is a syntax defined by the authors of Stata's `sem` and `gsem` commands for entering path diagrams in a command language. Models to be fit may be specified in path notation or they may be drawn using path diagrams into the Builder.

PCA. See *principal component analysis*.

PCM. See *partial credit model*.

p-conformability. Matrix, vector, or scalar A is said to be p-conformable with matrix, vector, or scalar B if `rows(`A`)==rows(`B`)` and `cols(`A`)==cols(`B`)`. p stands for plus; p-conformability is one of the properties necessary to be able to add matrices together. p-conformability, however, does not imply that the matrices are of the same type. Thus `(1,2,3)` is p-conformable with `(4,5,6)` and with `("this","that","what")` but not with `(4\5\6)`.

Pearson's correlation. Pearson's correlation ρ, also known as the product-moment correlation, measures the degree of association between two variables. Pearson's correlation equals the variables' covariance divided by their respective standard deviations, and ranges between -1 and 1. Zero indicates no correlation between the two variables.

penalized log-likelihood function. This is a log-likelihood function that contains an added term, usually referred to as a roughness penalty, that reduces its value when the model overfits the data. In Cox models with frailty, such functions are used to prevent the variance of the frailty from growing too large, which would allow the individual frailty values to perfectly fit the data.

periodogram. A periodogram is a graph of the spectral density function of a time series as a function of frequency. The `pergram` command first standardizes the amplitude of the density by the sample variance of the time series, and then plots the logarithm of that standardized density. Peaks in the periodogram represent cyclical behavior in the data.

permutation matrix and permutation vector. A *permutation matrix* is an $n \times n$ matrix that is a row (or column) permutation of the identity matrix. If P is a permutation matrix, then $P*A$ permutes the rows of A and $A*P$ permutes the columns of A. Permutation matrices also have the property that $P^{-1} = P'$.

A *permutation vector* is a $1 \times n$ or $n \times 1$ vector that contains a permutation of the integers 1, 2, ..., n. Permutation vectors can be used with subscripting to reorder the rows or columns of a matrix. Permutation vectors are a memory-conserving way of recording permutation matrices; see [M-1] **permutation**.

person location. Location of a person on the latent trait scale.

phase function. The phase function of a linear filter specifies how the filter changes the relative importance of the random components at different frequencies in the frequency domain.

Pillai's trace. Pillai's trace is a test statistic for the hypothesis test $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \cdots = \boldsymbol{\mu}_k$ based on the eigenvalues $\lambda_1, \ldots, \lambda_s$ of $\mathbf{E}^{-1}\mathbf{H}$. It is defined as

$$V^{(s)} = \text{trace}[(\mathbf{E}+\mathbf{H})^{-1}\mathbf{H}] = \sum_{i=1}^{s} \frac{\lambda_i}{1+\lambda_i}$$

where $\mathbf{H}$ is the between matrix and $\mathbf{E}$ is the within matrix. See *between matrix*.

plain ASCII. We use plain ASCII as a nontechnical term to refer to what computer programmers call lower ASCII. These are the plain Latin letters "a" to "z" and "A" to "Z"; numbers "0" through "9"; many punctuation marks, such as "!"; simple mathematical symbols, such as "+"; and whitespace and control characters such as space (" "), tab, and carriage return.

Each plain ASCII characters is stored as a single byte with a value between 0 and 127. Another distinguishing feature is that the byte values used to encode plain ASCII characters are the same across different operating systems and are common between ASCII and UTF-8.

Also see *ASCII* and *encodings*.

point estimate. A point estimate is another name for a statistic, such as a mean or regression coefficient.

pointer. A matrix is said to be a pointer matrix if its elements are pointers.

A pointer is the address of a *variable*. Say that variable X contains a matrix. Another variable p might contain 137,799,016 and, if 137,799,016 were the address at which X were stored, then p would be said to point to X. Addresses are seldom written in base 10, and so rather than saying p contains 137,799,016, we would be more likely to say that p contains 0x836a568, which is the way we write numbers in base 16. Regardless of how we write addresses, however, p contains a number and that number corresponds to the address of another variable.

In our program, if we refer to p, we are referring to p's contents, the number 0x836a568. The monadic operator * is defined as "refer to the address" or "dereference": $*p$ means X. We could code `Y = *p` or `Y = X`, and either way, we would obtain the same result. In our program, we could refer to X`[`i`,`j`]` or `(*`p`)[`i`,`j`]`, and either way, we would obtain the i, j element of X.

The monadic operator `&` is how we put addresses into p. To load p with the address of X, we code p `= &`X.

The special address 0 (zero, written in hexadecimal as 0×0), also known as `NULL`, is how we record that a pointer variable points to nothing. A pointer variable contains `NULL` or it contains a valid address of another variable.

See [M-2] **pointers** for a complete description of pointers and their use.

Poisson regression. Poisson regression is a term for generalized linear response functions that are family Poisson, link log. It is used for count data. See *generalized linear response functions*.

Poisson regression model. The Poisson regression model is used when the dependent variable represents the number of times an event occurs. In the Poisson model, the variance of the dependent variable is equal to the conditional mean.

polytomous item. See *categorical item*.

POMs. See *potential-outcome means*.

pooled estimator. A pooled estimator ignores the longitudinal or panel aspect of a dataset and treats the observations as if they were cross-sectional.

population-averaged model. A population-averaged model is used for panel data in which the parameters measure the effects of the regressors on the outcome for the average individual in the population. The panel-specific errors are treated as uncorrelated random variables drawn from a population with zero mean and constant variance, and the parameters measure the effects of the regressors on the dependent variable after integrating over the distribution of the random effects.

portmanteau statistic. The portmanteau, or Q, statistic is used to test for white noise and is calculated using the first m autocorrelations of the series, where m is chosen by the user. Under the null hypothesis that the series is a white-noise process, the portmanteau statistic has a χ^2 distribution with m degrees of freedom.

positive effect size. In power and sample-size analysis, we obtain a positive effect size when the postulated value of the parameter under the alternative hypothesis is greater than the hypothesized value of the parameter under the null hypothesis. Also see *negative effect size*.

posterior distribution, posterior. A posterior distribution is a probability distribution of model parameters conditional on observed data. The posterior distribution is determined by the likelihood of the parameters and their prior distribution. For a parameter vector $\boldsymbol{\theta}$ and data $\mathbf{y}$, the posterior distribution is given by

$$P(\boldsymbol{\theta}|\mathbf{y}) = \frac{P(\boldsymbol{\theta})P(\mathbf{y}|\boldsymbol{\theta})}{P(\mathbf{y})}$$

where $P(\boldsymbol{\theta})$ is the prior distribution, $P(\mathbf{y}|\boldsymbol{\theta})$ is the model likelihood, and $P(\mathbf{y})$ is the marginal distribution for $\mathbf{y}$. Bayesian inference is based on a posterior distribution.

posterior independence. See *independent a posteriori*.

posterior interval. See *credible interval*.

posterior mean. In generalized linear mixed-effects models, posterior mean refers to the predictions of random effects based on the mean of the posterior distribution.

In IRT models, posterior mean refers to the predictions of the latent trait based on the mean of the posterior distribution.

posterior mode. In generalized linear mixed-effects models, posterior mode refers to the predictions of random effects based on the mode of the posterior distribution.

In IRT models, posterior mode refers to the predictions of the latent trait based on the mode of the posterior distribution.

posterior odds. Posterior odds for $\boldsymbol{\theta}_1$ compared with $\boldsymbol{\theta}_2$ is the ratio of posterior density evaluated at $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ under a given model,

$$\frac{p(\boldsymbol{\theta}_1|\mathbf{y})}{p(\boldsymbol{\theta}_2|\mathbf{y})} = \frac{p(\boldsymbol{\theta}_1)}{p(\boldsymbol{\theta}_2)} \frac{p(\mathbf{y}|\boldsymbol{\theta}_1)}{p(\mathbf{y}|\boldsymbol{\theta}_2)}$$

In other words, posterior odds are prior odds times the likelihood ratio.

posterior predictive distribution. A posterior predictive distribution is a distribution of unobserved (future) data conditional on the currently observed data. Posterior predictive distribution is derived by marginalizing the likelihood function with respect to the posterior distribution of model parameters.

posterior probabilities. After discriminant analysis, the posterior probabilities are the probabilities of a given observation being assigned to each of the groups based on the prior probabilities, the training data, and the particular discriminant model. Contrast to *prior probabilities*.

poststratification. Poststratification is a method for adjusting sampling weights, usually to account for underrepresented groups in the population. This usually results in decreased bias because of nonresponse and underrepresented groups in the population. Poststratification also tends to result in smaller variance estimates.

The population is partitioned into categories, called poststrata. The sampling weights are adjusted so that the sum of the weights within each poststratum is equal to the respective poststratum size. The poststratum size is the number of individuals in the population that are in the poststratum. The frequency distribution of the poststrata typically comes from census data, and the poststrata are most commonly identified by demographic information such as age, sex, and ethnicity.

postulated value. See *alternative value*.

potential outcome. The potential outcome is the outcome an individual would obtain if given a specific treatment.

For example, an individual has one potential blood pressure after taking a pill and another potential blood pressure had that person not taken the pill.

potential-outcome means. The potential-outcome means refers to the means of the potential outcomes for a specific treatment level.

The mean blood pressure if everyone takes a pill and the mean blood pressure if no one takes a pill are two examples.

The average treatment effect is the difference between potential-outcome mean for the treated and the potential-outcome mean for the not treated.

power. The power of a test is the probability of correctly rejecting the null hypothesis when it is false. It is often denoted as $1-\beta$ in statistical literature, where β is the type II error probability. Commonly used values for power are 80% and 90%. Also see *type I error* and *type II error*. **power and sample-size analysis**. Power and sample-size analysis investigates the optimal allocation of study resources to increase the likelihood of the successful achievement of a study objective. See [PSS] **intro**.

power curve. A power curve is a graph of the estimated power as a function of some other study parameter such as the sample size. The power is plotted on the y axis, and the sample size or other parameter is plotted on the x axis. See [PSS] **power, graph**.

power determination. This pertains to the computation of a power given sample size, effect size, and other study parameters.

power function. The power functions is a function of the population parameter θ, defined as the probability that the observed sample belongs to the rejection region of a test for given θ. See *Hypothesis testing* under *Remarks and examples* in [PSS] **intro**.

power graph. See *power curve*.

pragma. "(Pragmatic information) A standardised form of comment which has meaning to a compiler. It may use a special syntax or a specific form within the normal comment syntax. A pragma usually conveys non-essential information, often intended to help the compiler to optimise the program." See *The Free On-line Dictionary of Computing*, http://www.foldoc.org/, Editor Denis Howe. For Mata, see [M-2] **pragma**.

Prais–Winsten estimator. A Prais–Winsten estimator is a linear regression estimator that is used when the error term exhibits first-order autocorrelation; see also *Cochrane–Orcutt estimator*. Here the first observation in the dataset is transformed as $\widetilde{y}_1 = \sqrt{1-\rho^2}\ y_1$ and $\widetilde{\mathbf{x}}_1 = \sqrt{1-\rho^2}\ \mathbf{x}_1$, so that the first observation is not lost. The Prais–Winsten estimator is a generalized least-squares estimator.

predetermined variable. A predetermined variable is a regressor in which its contemporaneous and future values are not correlated with the unobservable error term but past values are correlated with the error term.

predictive margins. Predictive margins provide a way of exploring the response surface of a fitted model in any response metric of interest—means, linear predictions, probabilities, marginal effects, risk differences, and so on. Predictive margins are estimates of responses (or outcomes) for the groups represented by the levels of a factor variable, controlling for the differing covariate distributions across the groups. They are the survey-data and nonlinear response analogue to what are often called estimated marginal means or least-squares means for linear models.

Because these margins are population-weighted averages over the estimation sample or subsamples, and because they take account of the sampling distribution of the covariates, they can be used to make inferences about treatment effects for the population.

prevented fraction. A prevented fraction is the reduction in the risk of a disease or other condition of interest caused by including a protective risk factor or public-health intervention.

prewhiten. To prewhiten is to apply a transformation to a time series so that it becomes white noise.

primary sampling unit. Primary sampling unit (PSU) is a cluster that was sampled in the first sampling stage; see *cluster*.

priming values. Priming values are the initial, preestimation values used to begin a recursive process.

principal component analysis. Principal component analysis (PCA) is a statistical technique used for data reduction. The leading eigenvectors from the eigen decomposition of the correlation or the covariance matrix of the variables describe a series of uncorrelated linear combinations of the variables that contain most of the variance. In addition to data reduction, the eigenvectors from a PCA are often inspected to learn more about the underlying structure of the data.

principal factor method. The principal factor method is a method for factor analysis in which the factor loadings, sometimes called factor patterns, are computed using the squared multiple correlations as estimates of the communality. Also see *factor analysis* and *communality*.

prior distribution, prior. In Bayesian statistics, prior distributions are probability distributions of model parameters formed based on some a priori knowledge about parameters. Prior distributions are independent of the observed data.

prior independence. See *independent a priori*.

prior odds. Prior odds for $\boldsymbol{\theta}_1$ compared with $\boldsymbol{\theta}_2$ is the ratio of prior density evaluated at $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ under a given model, $p(\boldsymbol{\theta}_1)/p(\boldsymbol{\theta}_2)$. Also see *posterior odds*.

prior probabilities Prior probabilities in discriminant analysis are the probabilities of an observation belonging to a group before the discriminant analysis is performed. Prior probabilities are often based on the prevalence of the groups in the population as a whole. Contrast to *posterior probabilities*.

probability of a type I error. This is the probability of committing a type I error of incorrectly rejecting the null hypothesis. Also see *significance level*.

probability of a type II error. This is the probability of committing a type II error of incorrectly accepting the null hypothesis. Common values for the probability of a type II error are 0.1 and 0.2 or, equivalently, 10% and 20%. Also see *beta* and *power*.

probability weight. Probability weight is another term for sampling weight.

probit regression. Probit regression is a term for generalized linear response functions that are family Bernoulli, link probit. It is used for binary outcome data. Probit regression is also known simply as probit. See *generalized linear response functions*.

Procrustes rotation. A Procrustes rotation is an orthogonal or oblique transformation, that is, a restricted Procrustes transformation without translation or dilation (uniform scaling).

Procrustes transformation. The goal of Procrustes transformation is to transform the source matrix $\mathbf{X}$ to be as close as possible to the target $\mathbf{Y}$. The permitted transformations are any combination of dilation (uniform scaling), rotation and reflection (that is, orthogonal or oblique transformations), and translation. Closeness is measured by residual sum of squares. In some cases, unrestricted Procrustes transformation is desired; this allows the data to be transformed not just by orthogonal or oblique rotations, but by all conformable regular matrices $\mathbf{A}$. Unrestricted Procrustes transformation is equivalent to a multivariate regression.

The name comes from Procrustes of Greek mythology; Procrustes invited guests to try his iron bed. If the guest was too tall for the bed, Procrustes would amputate the guest's feet, and if the guest was too short, he would stretch the guest out on a rack.

Also see *orthogonal rotation*, *oblique rotation*, *dilation*, and *multivariate regression*.

production function. A production function describes the maximum amount of a good that can be produced, given specified levels of the inputs.

promax power rotation. Promax power rotation is an oblique rotation. It does not fit in the minimizing-a-criterion framework that is at the core of most other rotations. The promax method (Hendrickson and White 1964) was proposed before computing power became widely available. The promax rotation consists of three steps:

1. Perform an orthogonal rotation.
2. Raise the elements of the rotated matrix to some power, preserving the sign of the elements. Typically the power is in the range $2 \leq \text{power} \leq 4$. This operation is meant to distinguish clearly between small and large values.
3. The matrix from step two is used as the target for an oblique Procrustean rotation from the original matrix.

propensity score. The propensity score is the probability that an individual receives a treatment.

propensity-score matching. Propensity-score matching uses the distance between estimated propensity scores to find similar individuals.

proportional hazards model. This is a model in which, between individuals, the ratio of the instantaneous failure rates (the hazards) is constant over time.

proposal distribution. In the context of the MH algorithm, a proposal distribution is used for defining the transition steps of the Markov chain. In the standard random-walk Metropolis algorithm, the proposal distribution is a multivariate normal distribution with zero mean and adaptable covariance matrix.

prospective study. In a prospective study, the population or cohort is classified according to specific risk factors, such that the outcome of interest, typically various manifestations of a disease, can be observed over time and tied in to the initial classification. Also see *retrospective study*.

Also known as a prospective longitudinal study, a prospective study is a study based on observations over the same subjects for a given period.

proximity, **proximity matrix**, and **proximity measure**. Proximity or a proximity measure means the nearness or farness of two things, such as observations or variables or groups of observations or a method for quantifying the nearness or farness between two things. A proximity is measured by a similarity or dissimilarity. A proximity matrix is a matrix of proximities. Also see *similarity* and *dissimilarity*.

pseudoconvergence. A Markov chain may appear to converge when in fact it did not. We refer to this phenomenon as pseudoconvergence. Pseudoconvergence is typically caused by multimodality of the stationary distribution, in which case the chain may fail to traverse the weakly connected regions of the distribution space. A common way to detect pseudoconvergence is to run multiple chains using different starting values and to verify that all of the chain converge to the same target distribution.

pseudolikelihood. A pseudolikelihood is a weighted likelihood that is used for point estimation. Pseudolikelihoods are not true likelihoods because they do not represent the distribution function for the sample data from a survey. The sampling distribution is instead determined by the survey design.

PSS analysis. See *power and sample-size analysis*.

PSS Control Panel. The PSS Control Panel is a point-and-click graphical user interface for power and sample-size analysis. See [PSS] **GUI**.

PSU. See *primary sampling unit*.

p-value. P-value is a probability of obtaining a test statistic as extreme or more extreme as the one observed in a sample assuming the null hypothesis is true.

QDA. See *quadratic discriminant analysis*.

QML, method(ml) vce(robust). QML stands for quasimaximum likelihood. It is a method used to obtain fitted parameters, and a technique used to obtain the corresponding VCE. QML is used by `sem` and `gsem` when options `method(ml)` and `vce(robust)` are specified. Other available methods are ML, MLMV, and ADF. Other available techniques are OIM, EIM, OPG, clustered, bootstrap, and jackknife.

QR decomposition. QR decomposition is an orthogonal-triangular decomposition of an augmented data matrix that speeds up the calculation of the log likelihood; see *Methods and formulas* in [ME] **mixed** for more details.

quadratic discriminant analysis. Quadratic discriminant analysis (QDA) is a parametric form of discriminant analysis and is a generalization of LDA. Like LDA, QDA assumes that the observations come from a multivariate normal distribution, but unlike LDA, the groups are not assumed to have equal covariance matrices. Also see *discriminant analysis*, *linear discriminant analysis*, and *parametric methods*.

quadrature. Quadrature is generic method for performing numerical integration. `gsem` uses quadrature in any model including latent variables (excluding error variables). `sem`, being limited to linear models, does not need to perform quadrature.

quartimax rotation. Quartimax rotation maximizes the variance of the squared loadings within the rows of the matrix. It is an orthogonal rotation that is equivalent to minimizing the criterion

$$c(\mathbf{\Lambda}) = \sum_i \sum_r \lambda_{ir}^4 = -\frac{1}{4}\left\langle \mathbf{\Lambda}^2, \mathbf{\Lambda}^2 \right\rangle$$

See *Crawford–Ferguson rotation* for a definition of $\mathbf{\Lambda}$.

quartimin rotation. Quartimin rotation is an oblique rotation that is equivalent to quartimax rotation when quartimin is restricted to orthogonal rotations. Quartimin is equivalent to oblimin rotation with $\gamma = 0$. Also see *quartimax rotation*, *oblique rotation*, *orthogonal rotation*, and *oblimin rotation*.

random coefficient. In the context of mixed-effects models, a random coefficient is a counterpart to a slope in the fixed-effects equation. You can think of a random coefficient as a randomly varying slope at a specific level of nesting.

random effects. In the context of mixed-effects models, random effects represent effects that may vary from group to group at any level of nesting. In the ANOVA literature, random effects represent the levels of a factor for which the inference can be generalized to the underlying population represented by the levels observed in the study. See also *random-effects model* in [XT] **Glossary**. For a information related to Bayesian analysis, see *random-effects parameters*.

random intercept. In the context of mixed-effects models, a random intercept is a counterpart to the intercept in the fixed-effects equation. You can think of a random intercept as a randomly varying intercept at a specific level of nesting.

random walk. A random walk is a time-series process in which the current period's realization is equal to the previous period's realization plus a white-noise error term: $y_t = y_{t-1} + \epsilon_t$. A *random walk with drift* also contains a nonzero time-invariant constant: $y_t = \delta + y_{t-1} + \epsilon_t$. The constant term δ is known as the drift parameter. An important property of random-walk processes is that the best predictor of the value at time $t+1$ is the value at time t plus the value of the drift parameter.

random-coefficients model. A random-coefficients model is a panel-data model in which group-specific heterogeneity is introduced by assuming that each group has its own parameter vector, which is drawn from a population common to all panels.

random-effects linear form. A linear form representing a random-effects variable that can be used in substitutable expressions.

random-effects model. A random-effects model for panel data treats the panel-specific errors as uncorrelated random variables drawn from a population with zero mean and constant variance. The regressors must be uncorrelated with the random effects for the estimates to be consistent. See also *fixed-effects model*.

random-effects parameters. Parameters associated with a random-effects variable. Random-effects parameters are assumed to be conditionally independent across levels of the random-effects variable given all other model parameters. Often, random-effects parameters are assumed to be normally distributed with a zero mean and an unknown variance–covariance matrix.

random-effects variable. A variable identifying the group structure for the random effects at a specific level of hierarchy.

randomized controlled trial. In this experimental study, treatments are randomly assigned to two or more groups of subjects.

rank. Terms in common use are rank, row rank, and column rank. The row rank of a matrix A: $m \times n$ is the number of rows of A that are linearly independent. The column rank is defined similarly, as the number of columns that are linearly independent. The terms *row rank* and *column rank*, however, are used merely for emphasis; the ranks are equal and the result is simply called the rank of A.

For a square matrix A (where m==n), the matrix is invertible if and only if `rank(`A`)==`n. One often hears that A is of full rank in this case and rank deficient in the other. See [M-5] **rank()**.

rating scale model. The rating scale model (RSM) is an IRT model for ordinal responses. The categories within each item vary in their difficulty; however, the distances between adjacent difficulty parameters are constrained to be the same across the items. The categories across all items share the same discrimination parameter.

r-conformability. A set of two or more matrices, vectors, or scalars A, B, ..., are said to be r-conformable if each is *c-conformable* with a matrix of `max(rows(`A`), rows(`B`), ...)` rows and `max(cols(`A`), cols(`B`), ...)` columns.

r-conformability is a more relaxed form of *c-conformability* in that, if two matrices are c-conformable, they are r-conformable, but not vice versa. For instance, A: 1×3 and B: 3×1 are r-conformable but not c-conformable. Also, c-conformability is defined with respect to a pair of matrices only; r-conformability can be applied to a set of matrices.

r-conformability is often required of the arguments for functions that would otherwise naturally be expected to require scalars. See *R-conformability* in [M-5] **normal()** for an example. **RCT**. See *randomized controlled trial*.

real. A matrix is said to be a real matrix if its elements are all reals and it is stored in a `real` matrix. Real is one of the two numeric types in Mata, the other being complex. Also see *type, eltype, and orgtype*.

recursive regression analysis. A recursive regression analysis involves performing a regression at time t by using all available observations from some starting time t_0 through time t, performing another regression at time $t+1$ by using all observations from time t_0 through time $t+1$, and so on. Unlike a rolling regression analysis, the first period used for all regressions is held fixed.

reference prior. See *noninformative prior*.

reference value. See *null value*.

reflection. A reflection is an orientation reversing orthogonal transformation, that is, a transformation that involves negating coordinates in one or more dimensions. A reflection is a Procrustes transformation.

registered and unregistered variables. Variables in `mi` data can be registered as imputed, passive, or regular by using the `mi register` command; see [MI] **mi set**.

You are required to register imputed variables.

You should register passive variables; if your data are style wide, you are required to register them. The `mi passive` command (see [MI] **mi passive**) makes creating passive variables easy, and it automatically registers them for you.

Whether you register regular variables is up to you. Registering them is safer in all styles except wide, where it does not matter. By definition, regular variables should be the same across m. In the long styles, you can unintentionally create variables that vary. If the variable is registered, `mi` will detect and fix your mistakes.

Super-varying variables, which rarely occur and can be stored only in flong and flongsep data, should never be registered.

The registration status of variables is listed by the `mi describe` command; see [MI] **mi describe**.

regressand. The regressand is the variable that is being explained or predicted in a regression model. Synonyms include dependent variable, left-hand-side variable, and endogenous variable.

regression. A regression is a model in which an endogenous variable is written as a function of other variables, parameters to be estimated, and a random disturbance.

regression-adjustment estimators. Regression-adjustment estimators use means of predicted outcomes for each treatment level to estimate each potential-outcome mean.

regressor. Regressors are variables in a regression model used to predict the regressand. Synonyms include independent variable, right-hand-side variable, explanatory variable, predictor variable, and exogenous variable.

regular variable. See *imputed, passive, and regular variables*.

rejection region. In hypothesis testing, a rejection region is a set of sample values for which the null hypothesis can be rejected.

relative efficiency. Ratio of variance of a parameter given estimation with finite M to the variance if M were infinite.

relative risk. See *risk ratio*.

relative variance increase. The increase in variance of a parameter estimate due to nonresponse.

reliability. Reliability is the proportion of the variance of a variable not due to measurement error. A variable without measure error has reliability 1.

REML. See *restricted maximum likelihood*.

repeated denominator degrees of freedom (DDF) method. This method uses the repeated-measures ANOVA for computing DDF. It is used with balanced repeated-measures designs with spherical correlation error structures. It partitions the residual degrees of freedom into the between-subject degrees of freedom and the within-subject degrees of freedom. The repeated method is supported only with two-level models. For more complex mixed-effects models or with unbalanced data, this method typically leads to poor approximations of the actual sampling distributions of the test statistics.

repeated measures. Repeated measures data have repeated measurements for the subjects over some dimension, such as time—for example test scores at the start, midway, and end of the class. The repeated observations are typically not independent. Repeated-measures ANOVA is one approach for analyzing repeated measures data, and MANOVA is another. Also see *sphericity*.

replicate-weight variable. A replicate-weight variable contains sampling weight values that were adjusted for resampling the data; see [SVY] **variance estimation** for more details.

resampling. Resampling refers to the process of sampling from the dataset. In the delete-one jackknife, the dataset is resampled by dropping one PSU and producing a replicate of the point estimates. In the BRR method, the dataset is resampled by dropping combinations of one PSU from each stratum. The resulting replicates of the point estimates are used to estimate their variances and covariances.

residual. In this manual, we reserve the word “residual” for the difference between the observed and fitted moments of an SEM model. We use the word error for the disturbance associated with a (Gaussian) linear equation; see *error*. Also see *standardized residuals*.

residual denominator degrees of freedom (DDF) method. This method uses the residual degrees of freedom, $n - \text{rank}(X)$, as the DDF for all tests of fixed effects. For a linear model without random effects with independent and identically distributed errors, the distributions of the test statistics for fixed effects are t or F distributions with the residual DDF. For other mixed-effects models, this method typically leads to poor approximations of the actual sampling distributions of the test statistics.

restricted maximum likelihood. Restricted maximum likelihood is a method of fitting linear mixed-effects models that involves transforming out the fixed effects to focus solely on variance–component estimation.

retrospective study. In a retrospective study, a group with a disease of interest is compared with a group without the disease, and information is gathered in a retrospective way about the exposure in each group to various risk factors that might be associated with the disease. Also see *prospective study*.

right-censoring. See *censored, censoring, left-censoring, and right-censoring*.

right-truncation. See *truncation, left-truncation, and right-truncation*.

risk difference. A risk difference is defined as the probability of an event occurring when a risk factor is increased by one unit minus the probability of the event occurring without the increase in the risk factor.

When the risk factor is binary, the risk difference is the probability of the outcome when the risk factor is present minus the probability when the risk factor is not present.

When one compares two populations, a risk difference is defined as a difference between the probabilities of an event in the two groups. It is typically a difference between the probability in the comparison group or experimental group and the probability in the reference group or control group.

risk factor. This is a variable associated with an increased or decreased risk of failure.

risk pool. At a particular point in time, this is the subjects at risk of failure.

risk ratio. In a log-linear model, this is the ratio of probability of survival associated with a one-unit increase in a risk factor relative to that calculated without such an increase, that is, $R(x+1)/R(x)$. Given the exponential form of the model, $R(x+1)/R(x)$ is constant and is given by the exponentiated coefficient.

robust, vce(robust). Robust is the name we use here for the Huber/White/sandwich estimator of the VCE. This technique requires fewer assumptions than most other techniques. In particular, it merely assumes that the errors are independently distributed across observations and thus allows the errors to be heteroskedastic. Robust standard errors are reported when the `sem` (`gsem`) option `vce(robust)` is specified. The other available techniques are OIM, EIM, OPG, clustered, bootstrap, and jackknife.

robust standard errors. Robust standard errors, also known as Huber/White or Taylor linearization standard errors, are based on the sandwich estimator of variance. Robust standard errors can be interpreted as representing the sample-to-sample variability of the parameter estimates, even when the model is misspecified. See also *semirobust standard errors*.

rolling regression analysis. A rolling, or moving window, regression analysis involves performing regressions for each period by using the most recent m periods' data, where m is known as the window size. At time t the regression is fit using observations for times $t-19$ through time t; at time $t+1$ the regression is fit using the observations for time $t-18$ through $t+1$; and so on.

rotation. A rotation is an orientation preserving orthogonal transformation. A rotation is a Procrustes transformation.

row and column stripes. Stripes refer to the labels associated with the rows and columns of a Stata matrix; see *Stata matrix*.

row-major order. Matrices are stored as vectors. Row-major order specifies that the vector form of a matrix is created by stacking the rows. For instance,

```
: A
        1   2   3
    +-------------+
  1 |   1   2   3 |
  2 |   4   5   6 |
    +-------------+
```

is stored as

```
        1   2   3   4   5   6
    +-------------------------+
  1 |   1   2   3   4   5   6 |
    +-------------------------+
```

in row-major order. Mata uses row-major order. The LAPACK functions use column-major order. See *column-major order*.

rowvector. See *vector, colvector, and rowvector*.

Roy's largest root. Roy's largest root test is a test statistic for the hypothesis test $H_0 : \boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_k$ based on the largest eigenvalue of $\mathbf{E}^{-1}\mathbf{H}$. It is defined as

$$\theta = \frac{\lambda_1}{1+\lambda_1}$$

Here $\mathbf{H}$ is the between matrix, and $\mathbf{E}$ is the within matrix. See *between matrix*.

RSM. See *rating scale model*.

RVI. See *relative variance increase*.

Sammon mapping criterion. The Sammon (1969) mapping criterion is a loss criterion used with MDS; it is the sum of the scaled, squared differences between the distances and the disparities, normalized by the sum of the disparities. Also see *multidimensional scaling*, *modern scaling*, and *loss*.

sample. A sample is the collection of individuals in the population that were chosen as part of the survey. Sample is also used to refer to the data, typically in the form of answered questions, collected from the sampled individuals.

sample size. This is the number of subjects in a sample. See [PSS] **intro** to learn more about the relationship between sample size and the power of a test.

sample-size curve. A sample-size curve is a graph of the estimated sample size as a function of some other study parameter such as power. The sample size is plotted on the y axis, and the power or other parameter is plotted on the x axis.

sample-size determination. This pertains to the computation of a sample size given power, effect size, and other study parameters.

sampling stage. Complex survey data are typically collected using multiple stages of clustered sampling. In the first stage, the PSUs are independently selected within each stratum. In the second stage, smaller sampling units are selected within the PSUs. In later stages, smaller and smaller sampling units are selected within the clusters from the previous stage.

sampling unit. A sampling unit is an individual or collection of individuals from the population that can be selected in a specific stage of a given survey design. Examples of sampling units include city blocks, high schools, hospitals, and houses.

sampling weight. Given a survey design, the sampling weight for an individual is the reciprocal of the probability of being sampled. The probability for being sampled is derived from stratification and clustering in the survey design. A sampling weight is typically considered to be the number of individuals in the population represented by the sampled individual.

sampling with and without replacement. Sampling units may be chosen more than once in designs that use sampling with replacement. Sampling units may be chosen at most once in designs that use sampling without replacement. Variance estimates from with-replacement designs tend to be larger than those from corresponding without-replacement designs.

Satterthwaite denominator degrees of freedom (DDF) method. This method implements a generalization of the Satterthwaite (1946) approximation of the unknown sampling distributions of test statistics for complex linear mixed-effects models. This method is supported only with restricted maximum-likelihood estimation.

Satterthwaite's t test. Satterthwaite's t test is a modification of the two-sample t test to account for unequal variances in the two populations. See *Methods and formulas* in [PSS] **power twomeans** for details.

saturated model. A saturated model is a full covariance model—a model of fitted means and covariances of observed variables without any restrictions on the values. Also see *baseline model*. Saturated models apply only to standard linear SEMs.

scalar. A special case of a *matrix* with one row and one column. A scalar may be substituted anywhere a matrix, vector, column vector, or row vector is required, but not vice versa.

scalar model parameter. A scalar model parameter is any model parameter that is a scalar. For example, {mean} and {hape:alpha} are scalar parameters, as declared by the bayesmh command. Elements of matrix model parameters are viewed as scalar model parameters. For example, for a 2×2 matrix parameter {Sigma,matrix}, individual elements {Sigma_1_1}, {Sigma_2_1}, {Sigma_1_2}, and {Sigma_2_2} are scalar parameters. If a matrix parameter contains a label, the label should be included in the specification of individual elements as well. See [BAYES] **bayesmh**.

scalar parameter. See *scalar model parameter*.

Schur decomposition. The Schur decomposition of a matrix, $\mathbf{A}$, can be written as

$$\mathbf{Q}'\mathbf{A}\mathbf{Q} = \mathbf{T}$$

where $\mathbf{T}$ is in Schur form and $\mathbf{Q}$, the matrix of Schur vectors, is orthogonal if $\mathbf{A}$ is real or unitary if $\mathbf{A}$ is complex. See [M-5] **schurd()**.

Schur form. There are two Schur forms: real Schur form and complex Schur form.

A real matrix is in Schur form if it is block upper triangular with 1×1 and 2×2 diagonal blocks. Each 2×2 diagonal block has equal diagonal elements and opposite sign off-diagonal elements. The real eigenvalues are on the diagonal and complex eigenvalues can be obtained from the 2×24 diagonal blocks.

A complex square matrix is in Schur form if it is upper triangular with the eigenvalues on the diagonal.

score. A score for an observation after factor analysis, PCA, or LDA is derived from a column of the loading matrix and is obtained as the linear combination of that observation's data by using the coefficients found in the loading.

score plot. A score plot produces scatterplots of the score variables after factor analysis, PCA, or LDA.

score test, **Lagrange multiplier test**. A score test is a test based on first derivatives of a likelihood function. Score tests are especially convenient for testing whether constraints on parameters should be relaxed or parameters should be added to a model. Also see *Wald test*.

scores. Scores has two unrelated meanings. First, scores are the observation-by-observation first-derivatives of the (quasi) log-likelihood function. When we use the word "scores", this is what we mean. Second, in the factor-analysis literature, scores (usually in the context of factor scores) refers to the expected value of a latent variable conditional on all the observed variables. We refer to this simply as the predicted value of the latent variable.

scree plot. A scree plot is a plot of eigenvalues or singular values ordered from greatest to least after an eigen decomposition or singular value decomposition. Scree plots help determine the number of factors or components in an eigen analysis. Scree is the accumulation of loose stones or rocky debris lying on a slope or at the base of a hill or cliff; this plot is called a scree plot because it looks like a scree slope. The goal is to determine the point where the mountain gives way to the fallen rock.

SDR. See *successive difference replication*.

seasonal difference operator. The period-s seasonal difference operator Δ_s denotes the difference in the value of a variable at time t and time $t-s$. Formally, $\Delta_s y_t = y_t - y_{t-s}$, and $\Delta_s^2 y_t = \Delta_s(y_t - y_{t-s}) = (y_t - y_{t-s}) - (y_{t-s} - y_{t-2s}) = y_t - 2y_{t-s} + y_{t-2s}$.

secondary sampling unit. Secondary sampling unit (SSU) is a cluster that was sampled from within a PSU in the second sampling stage. SSU is also used as a generic term unit to indicate any sampling unit that is not from the first sampling stage.

second-level latent variable. See *first-, second-, and higher-order latent variables*.

second-order latent variable. See *first- and second-order latent variables*.

seemingly unrelated regression. Seemingly unrelated regression is a kind of structural model in which each member of a set of observed endogenous variables is a function of a set of observed exogenous variables and a unique random disturbance term. The disturbances are correlated and the sets of exogenous variables may overlap. If the sets of exogenous variables are identical, this is referred to as multivariate regression.

selection-on-observables. See *conditional-independence assumption*.

SEM. SEM stands for structural equation modeling and for structural equation model. We use SEM in capital letters when writing about theoretical or conceptual issues as opposed to issues of the particular implementation of SEM in Stata with the `sem` or `gsem` commands.

sem. `sem` is the Stata command that fits standard linear SEMs. Also see *gsem*.

semiconjugate prior. A prior distribution is semiconjugate for a family of likelihood distributions if the prior and (full) conditional posterior distributions belong to the same family of distributions. For semiconjugacy to hold, parameters must typically be independent a priori; that is, their joint prior distribution must be the product of the individual marginal prior distributions. For example, the normal prior distribution for a mean parameter of a normal data distribution with an unknown variance (which is assumed to be independent of the mean a priori) is a semiconjugate prior. Semiconjugacy may provide an efficient way of sampling from posterior distributions and is used in Gibbs sampling.

semiparametric model. This is a model that is not fully parameterized. The Cox proportional hazards model is such a model:

$$h(t) = h_0(t)\exp(\beta_1 x_1 + \cdots + \beta_k x_k)$$

In the Cox model, $h_o(t)$ is left unparameterized and not even estimated. Meanwhile, the relative effects of covariates are parameterized as $\exp(\beta_1 x_1 + \cdots + \beta_k x_k)$.

semirobust standard errors. Semirobust standard errors are closely related to robust standard errors and can be interpreted as representing the sample-to-sample variability of the parameter estimates, even when the model is misspecified, as long as the mean structure of the model is specified correctly. See also *robust standard errors*.

sensitivity analysis. Sensitivity analysis investigates the effect of varying study parameters on power, sample size, and other components of a study. The true values of study parameters are usually unknown, and power and sample-size analysis uses best guesses for these values. It is therefore important to evaluate the sensitivity of the computed power or sample size in response to changes in study parameters. See [PSS] **power, table** and [PSS] **power, graph** for details.

sequential limit theory. The sequential limit theory is a method of determining asymptotic properties of a panel-data statistic in which one index, say, N, the number of panels, is held fixed, while T, the number of time periods, goes to infinity, providing an intermediate limit. Then one obtains a final limit by studying the behavior of this intermediate limit as the other index (N here) goes to infinity.

serial correlation. Serial correlation refers to regression errors that are correlated over time. If a regression model does not contained lagged dependent variables as regressors, the OLS estimates are consistent in the presence of mild serial correlation, but the covariance matrix is incorrect. When the model includes lagged dependent variables and the residuals are serially correlated, the OLS estimates are biased and inconsistent. See, for example, Davidson and MacKinnon (1993, chap. 10) for more information.

serial correlation tests. Because OLS estimates are at least inefficient and potentially biased in the presence of serial correlation, econometricians have developed many tests to detect it. Popular ones include the Durbin–Watson (1950, 1951, 1971) test, the Breusch–Pagan (1980) test, and Durbin's (1970) alternative test. See [R] **regress postestimation time series**.

shape parameter. A shape parameter governs the shape of a probability distribution. One example is the parameter p of the Weibull model.

Shepard diagram. A Shepard diagram after MDS is a 2-dimensional plot of high-dimensional dissimilarities or disparities versus the resulting low-dimensional distances. Also see *multidimensional scaling*.

sign test. A sign test is used to test the null hypothesis that the median of a distribution is equal to some reference value. A sign test is carried out as a test of binomial proportion with a reference value of 0.5. See [PSS] **power oneproportion** and [R] **bitest**.

significance level. In hypothesis testing, the significance level α is an upper bound for a probability of a type I error. See [PSS] **intro** to learn more about the relationship between significance level and the power of a test.

similarity, **similarity matrix**, and **similarity measure**. A similarity or a similarity measure is a quantification of how alike two things are, such as observations or variables or groups of observations, or a method for quantifying that alikeness. A similarity matrix is a matrix containing similarity measurements. The matching coefficient is one example of a similarity measure. Contrast to *dissimilarity*. Also see *proximity* and *matching coefficient*.

simple random sample. In a simple random sample (SRS), individuals are independently sampled—each with the same probability of being chosen.

single-linkage clustering. Single-linkage clustering is a hierarchical clustering method that computes the proximity between two groups as the proximity between the closest pair of observations between the two groups.

single-record st data. See *st data*.

singleton-group data. A singleton is a frailty group that contains only 1 observation. A dataset containing only singletons is known as singleton-group data.

singular value decomposition. A singular value decomposition (SVD) is a factorization of a rectangular matrix. It says that if $\mathbf{M}$ is an $m \times n$ matrix, there exists a factorization of the form

$$\mathbf{M} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^*$$

where $\mathbf{U}$ is an $m \times m$ unitary matrix, $\boldsymbol{\Sigma}$ is an $m \times n$ matrix with nonnegative numbers on the diagonal and zeros off the diagonal, and $\mathbf{V}^*$ is the conjugate transpose of $\mathbf{V}$, an $n \times n$ unitary matrix. If $\mathbf{M}$ is a real matrix, then so is $\mathbf{V}$, and $\mathbf{V}^* = \mathbf{V}'$.

size of test. See *significance level*.

slope. See *discrimination*.

smooth treatment-effects estimator. A smooth treatment-effects estimator is a smooth function of the data so that standard methods approximate the distribution of the estimator. The RA, IPW, AIPW, and IPWRA estimators are all smooth treatment-effects estimators while the nearest-neighbor matching estimator and the propensity-score matching estimator are not.

smoothing. Smoothing a time series refers to the process of extracting an overall trend in the data. The motivation behind smoothing is the belief that a time series exhibits a trend component as well as an irregular component and that the analyst is interested only in the trend component. Some smoothers also account for seasonal or other cyclical patterns.

SMR. See *standardized mortality (morbidity) ratio*.

snapshot data. Snapshot data are those in which each record contains the values of a set of variables for a subject at an instant in time. The name arises because each observation is like a snapshot of the subject.

In snapshot datasets, one usually has a group of observations (snapshots) for each subject.

Snapshot data must be converted to st data before they can be analyzed. This requires making assumptions about what happened between the snapshots. See [ST] **snapspan**.

source code. Source code refers to the human-readable code that you type into Mata to define a function. Source code is compiled into object code, which is binary. See [M-1] **how**.

spectral analysis. See *frequency-domain analysis*.

spectral density function. The spectral density function is the derivative of the spectral distribution function. Intuitively, the spectral density function $f(\omega)$ indicates the amount of variance in a time series that is attributable to sinusoidal components with frequency ω. See also *spectral distribution function*. The spectral density function is sometimes called the *spectrum*.

spectral distribution function. The (normalized) spectral distribution function $F(\omega)$ of a process describes the proportion of variance that can be explained by sinusoids with frequencies in the range $(0, \omega)$, where $0 \leq \omega \leq \pi$. The spectral distribution and density functions used in frequency-domain analysis are closely related to the autocorrelation function used in time-domain analysis; see Chatfield (2004, chap. 6) and Wei (2006, chap. 12).

spectrum. See *spectral density function*.

spell data. Spell data are survival data in which each record represents a fixed period, consisting of a begin time, an end time, possibly a censoring/failure indicator, and other measurements (covariates) taken during that specific period.

sphericity. Sphericity is the state or condition of being a sphere. In repeated measures ANOVA, sphericity concerns the equality of variance in the difference between successive levels of the repeated measure. The multivariate alternative to ANOVA, called MANOVA, does not require the assumption of sphericity. Also see *repeated measures*.

square matrix. A matrix is square if it has the same number of rows and columns. A 3×3 matrix is square; a 3×4 matrix is not.

SRS. See *simple random sample*.

SSCP matrix. SSCP is an acronym for the sums of squares and cross products. Also see *between matrix*.

SSD, **ssd**. See *summary statistics data*.

SSU. See *secondary sampling unit*.

st data. st stands for survival time. In survival-time data, each observation represents a span of survival, recorded in variables $t0$ and t. For instance, if in an observation $t0$ were 3 and t were 5, the span would be $(t0, t\,]$, meaning from just after $t0$ up to and including t.

Sometimes variable $t0$ is not recorded; $t0$ is then assumed to be 0. In such a dataset, an observation that had $t = 5$ would record the span $(0, 5\,]$.

Each observation also includes a variable d, called the failure variable, which contains 0 or nonzero (typically, 1). The failure variable records what happened at the end of the span: 0, the subject was still alive (had not yet failed) or 1, the subject died (failed).

Sometimes variable d is not recorded; d is then assumed to be 1. In such a dataset, all time-span observations would be assumed to end in failure.

Finally, each observation in an st dataset can record the entire history of a subject or each can record a part of the history. In the latter case, groups of observations record the full history. One observation might record the period $(0, 5\,]$ and the next, $(5, 8\,]$. In such cases, there is a variable ID that records the subject for which the observation records a time span. Such data are called multiple-record st data. When each observation records the entire history of a subject, the data are called single-record st data. In the single-record case, the ID variable is optional.

See [ST] **stset**.

stacked variables. See *crossed variables*.

stacking variables. See *crossing variables*.

standard linear SEM. An SEM without multilevel effects in which all response variables are given by a linear equation. Standard linear SEM is what most people mean when they refer to just SEM. Standard linear SEMs are fit by `sem`, although they can also be fit by `gsem`; see *generalized SEM*.

standard strata. See *direct standardization*.

standard weights. See *direct standardization*.

standardized coefficient. In a linear equation $y = \ldots\ bx + \ldots$, the standardized coefficient β is $(\widehat{\sigma}_y/\widehat{\sigma}_x)b$. Standardized coefficients are scaled to units of standard deviation change in y for a standard deviation change in x.

standardized covariance. A standardized covariance between y and x is equal to the correlation of y and x, which is to say, it is equal to $\sigma_{xy}/\sigma_x\sigma_y$. The covariance is equal to the correlation when variables are standardized to have variance 1.

standardized data. Standardized data has a mean of zero and a standard deviation of one. You can standardize data $\mathbf{x}$ by taking $(\mathbf{x} - \overline{\mathbf{x}})/\sigma$, where σ is the standard deviation of the data.

standardized mortality (morbidity) ratio. Standardized mortality (morbidity) ratio (SMR) is the observed number of deaths divided by the expected number of deaths. It is calculated using indirect standardization: you take the population of the group of interest—say, by age, sex, and other factors—and calculate the expected number of deaths in each cell (expected being defined as the number of deaths that would have been observed if those in the cell had the same mortality as some other population). You then take the ratio to compare the observed with the expected number of deaths. For instance,

Age	(1) Population of group	(2) Deaths per 100,000 in general pop.	(1)×(2) Expected # of deaths	(4) Observed deaths
25–34	95,965	105.2	100.9	92
34–44	78,280	203.6	159.4	180
44–54	52,393	428.9	224.7	242
55–64	28,914	964.6	278.9	312
Total			763.9	826

$$\text{SMR} = 826/763.9 = 1.08$$

standardized residuals, **normalized residuals**. Standardized residuals are residuals adjusted so that they follow a standard normal distribution. The difficulty is that the adjustment is not always possible. Normalized residuals are residuals adjusted according to a different formula that roughly follow a standard normal distribution. Normalized residuals can always be calculated.

starting values. The estimation methods provided by `sem` and `gsem` are iterative. The starting values are values for each of the parameters to be estimated that are used to initialize the estimation process. The `sem` software provides starting values automatically, but in some cases, these are not good enough and you must (1) diagnose the problem and (2) provide better starting values. See [SEM] **intro 12**.

Stata matrix. Stata itself, separate from Mata, has matrix capabilities. Stata matrices are separate from those of Mata, although Stata matrices can be gotten from and put into Mata matrices; see [M-5] **st_matrix( )**. Stata matrices are described in [P] **matrix** and **[U] 14 Matrix expressions**.

Stata matrices are exclusively numeric and contain real elements only. Stata matrices also differ from Mata matrices in that, in addition to the matrix itself, a Stata matrix has text labels on the rows and columns. These labels are called row stripes and column stripes. One can think of rows and columns as having names. The purpose of these names is discussed in **[U] 14.2 Row and column names**. Mata matrices have no such labels. Thus three steps are required to get or to put all the information recorded in a Stata matrix: 1) getting or putting the matrix itself; 2) getting or putting the row stripe from or into a string matrix; and 3) getting or putting the column stripe from or into a string matrix. These steps are discussed in [M-5] **st_matrix( )**.

state-space model. A state-space model describes the relationship between an observed time series and an unobservable state vector that represents the "state" of the world. The measurement equation expresses the observed series as a function of the state vector, and the transition equation describes how the unobserved state vector evolves over time. By defining the parameters of the measurement and transition equations appropriately, one can write a wide variety of time-series models in the state-space form.

static forecast. A static forecast uses actual values wherever lagged values of the endogenous variables appear in the model. As a result, static forecasts perform at least as well as dynamic forecasts, but static forecasts cannot produce forecasts into the future if lags of the endogenous variables appear in the model.

Because actual values will be missing beyond the last historical time period in the dataset, static forecasts can only forecast one period into the future (assuming only first lags appear in the model); for that reason, they are often called one-step-ahead forecasts.

stationary distribution. Stationary distribution of a stochastic process is a joint distribution that does not change over time. In the context of MCMC, stationary distribution is the target probability distribution to which the Markov chain converges. When MCMC is used for simulating a Bayesian model, the stationary distribution is the target joint posterior distribution of model parameters.

steady-state equilibrium. The steady-state equilibrium is the predicted value of a variable in a dynamic model, ignoring the effects of past shocks, or, equivalently, the value of a variable, assuming that the effects of past shocks have fully died out and no longer affect the variable of interest.

stochastic equation. A stochastic equation, in contrast to an identity, is an equation in a forecast model that includes a random component, most often in the form of an additive error term. Stochastic equations include parameters that must be estimated from historical data.

stochastic trend. A stochastic trend is a nonstationary random process. Unit-root process and random coefficients on time are two common stochastic trends. See [TS] **ucm** for examples and discussions of more commonly applied stochastic trends.

stopping rules. Stopping rules for hierarchical cluster analysis are used to determine the number of clusters. A stopping-rule value (also called an index) is computed for each cluster solution, that is, at each level of the hierarchy in hierarchical cluster analysis. Also see *hierarchical clustering*.

str1, **str2**, ..., **str2045**. See *strL*.

stratification. The population is partitioned into well-defined groups of individuals, called strata. In the first sampling stage, PSUs are independently sampled from within each stratum. In later sampling stages, SSUs are independently sampled from within each stratum for that stage.

Survey designs that use stratification typically result in smaller variance estimates than do similar designs that do not use stratification. Stratification is most effective in decreasing variability when sampling units are more similar within the strata than between them.

stratified 2 × 2 tables. Describe the association between a binary independent variable and a binary response variable of interest. The analysis is stratified by a nominal (categorical) variable with K levels.

stratified model. A stratified survival model constrains regression coefficients to be equal across levels of the stratification variable, while allowing other features of the model to vary across strata.

stratified test. A stratified test is performed separately for each stratum. The stratum-specific results are then combined into an overall test statistic.

stress. See *Kruskal stress* and *loss*.

strict stationarity. A process is strictly stationary if the joint distribution of $y_1, \dots, y_k$ is the same as the joint distribution of $y_{1+\tau}, \dots, y_{k+\tau}$ for all k and τ. Intuitively, shifting the origin of the series by τ units has no effect on the joint distributions.

string. A matrix is said to be a string matrix if its elements are strings (text); see *type, eltype, and orgtype*. In Mata, a string may be text or binary and may be up to 2,147,483,647 characters (bytes) long.

strL. `strL` is a storage type for string variables. The full list of string storage types is `str1`, `str2`, . . . , `str2045`, and `strL`.

`str1`, `str2`, . . . , `str2045` are fixed-length storage types. If variable `mystr` is `str8`, then 8 bytes are allocated in each observation to store `mystr`'s value. If you have 2,000 observations, then 16,000 bytes in total are allocated.

Distinguish between storage length and string length. If `myvar` is `str8`, that does not mean the strings are 8 characters long in every observation. The maximum length of strings is 8 characters. Individual observations may have strings of length 0, 1, . . . , 8. Even so, every string requires 8 bytes of storage.

You need not concern yourself with the storage length because string variables are automatically promoted. If `myvar` is `str8`, and you changed the contents of `myvar` in the third observation to "Longer than 8", then `myvar` would automatically become `str13`.

If you changed the contents of `myvar` in the third observation to a string longer than 2,045 characters, `myvar` would become `strL`.

`strL` variables are not necessarily longer than 2,045 characters; they can be longer or shorter than 2,045 characters. The real difference is that `strL` variables are stored as varying length. Pretend that `myothervar` is a `strL` and its third observation contains "this". The total memory consumed by the observation would be $64+4+1 = 69$ bytes. There would be 64 bytes of tracking information, 4 bytes for the contents (there are 4 characters), and 1 more byte to terminate the string. If the fifth observation contained a 2,000,000-character string, then $64 + 2{,}000{,}000 + 1 = 2{,}000{,}069$ bytes would be used to store it.

Another difference between `str1`, `str2`, . . . , `str2045`, and `strL`s is that the `str#` storage types can store only ASCII strings. `strL` can store ASCII or binary strings. Thus a `strL` variable could contain, for instance, the contents of a Word document or a JPEG image or anything else.

`strL` is pronounce *sturl*.

strongly balanced. A longitudinal or panel dataset is said to be strongly balanced if each panel has the same number of observations, and the observations for different panels were all made at the same times.

structural equation model. Different authors use the term "structural equation model" in different ways, but all would agree that an SEM sometimes carries the connotation of being a structural model with a measurement component, that is, combined with a measurement model.

structural model. A structural model is one that describes the relationship among a set of variables, based on underlying theoretical considerations. In particular, the parameters of a structural model are posited to quantify an actual causal relationship among the variables rather than a mere description of the variables' correlations.

Structural models often have multiple equations and dependencies between endogenous variables, although that is not a requirement.

Structural models can be viewed in a structural equation modeling (SEM) framework and can thus be fitted by `sem` and `gsem`, though these commands are not limited to fitting just structural models. See [SEM] **intro 5** and *structural equation model*.

Structural models are also used in econometric forecasting applications. See [TS] **forecast** for information about forecasting from structural models based on time-series data.

structure (programming version). A structure is an *eltype*, indicating a set of variables tied together under one name. `struct mystruct` might be

```
        struct mystruct {
                real scalar     n1, n2
                real matrix     X
  }
```

If variable `a` was declared a `struct mystruct scalar`, then the scalar `a` would contain three pieces: two real scalars and one real matrix. The pieces would be referred to as `a.n1`, `a.n2`, and `a.X`. If variable `b` were also declared a `struct mystruct scalar`, it too would contain three pieces, `b.n1`, `b.n2`, and `b.X`. The advantage of structures is that they can be referred to as a whole. You can code `a.n1=b.n1` to copy one piece, or you can code `a=b` if you wanted to copy all three pieces. In all ways, `a` and `b` are variables. You may pass `a` to a subroutine, for instance, which amounts to passing all three values.

Structures variables are usually scalar, but they are not limited to being so. If `A` were a `struct mystruct matrix`, then each element of `A` would contain three pieces, and one could refer, for instance, to `A[2,3].n1`, `A[2,3].n2`, and `A[2,3].X`, and even to `A[2,3].X[3,2]`.

See [M-2] **struct**.

structure (statistics version). Structure, as in factor structure, is the correlations between the variables and the common factors after factor analysis. Structure matrices are available after factor analysis and LDA. Also see *factor analysis* and *linear discriminant analysis*.

structured (correlation or covariance). See *unstructured and structured (correlation or covariance)*.

style. Style refers to the format in which the `mi` data are stored. There are four styles: flongsep, flong, mlong, and wide. You can ignore styles, except for making an original selection, because all `mi` commands work regardless of style. You will be able to work more efficiently, however, if you understand the details of the style you are using; see [MI] **styles**. Some tasks are easier in one style than another. You can switch between styles by using the `mi convert` command; see [MI] **mi convert**.

The flongsep style is best avoided unless your data are too big to fit into one of the other styles. In flongsep style, a separate `.dta` set is created for $m = 0$, for $m = 1$, $\ldots$, and for $m = M$. Flongsep is best avoided because `mi` commands work more slowly with it.

In all the other styles, the $M + 1$ datasets are stored in one `.dta` file. The other styles are both more convenient and more efficient.

The most easily described of these `.dta` styles is flong; however, flong is also best avoided because mlong style is every bit as convenient as flong, and mlong is memorywise more efficient. In flong, each observation in the original data is repeated M times in the `.dta` dataset, once for $m = 1$, again for $m = 2$, and so on. Variable `_mi_m` records m and takes on values 0, 1, 2, $\ldots$, M. Within each value of m, variable `_mi_id` takes on values 1, 2, $\ldots$, N and thus connects imputed with original observations.

The mlong style is recommended. It is efficient and easy to use. Mlong is much like flong except that complete observations are not repeated.

Equally recommended is the wide style. In wide, each imputed and passive variable has an additional M variables associated with it, one for the variable's value in $m = 1$, another for its value in $m = 2$, and so on. If an imputed or passive variable is named *vn*, then the values of *vn* in $m = 1$ are stored in variable `_1_`*vn*; the values for $m = 2$, in `_2_`*vn*; and so on.

What makes mlong and wide so convenient? In mlong, there is a one-to-one correspondence of your idea of a variable and Stata's idea of a variable—variable *vn* refers to *vn* for all values of m. In wide, there is a one-to-one correspondence of your idea of an observation and Stata's idea—physical observation 5 is observation 5 in all datasets.

Choose the style that matches the problem at hand. If you want to create new variables or modify existing ones, choose mlong. If you want to drop observations or create new ones, choose wide. You can switch styles with the `mi convert` command; see [MI] **mi convert**.

For instance, if you want to create new variable `ageXexp` equal to `age*exp` and your data are mlong, you can just type `generate ageXexp = age*exp`, and that will work even if `age` and `exp` are imputed, passive, or a mix. Theoretically, the right way to do that is to type `mi passive: generate agexExp = age*exp`, but concerning variables, if your data are mlong, you can work the usual Stata way.

If you want to drop observation 20 or drop `if sex==2`, if your data are wide, you can just type `drop in 20` or `drop if sex==2`. Here the "right" way to do the problem is to type the `drop` command and then remember to type `mi update` so that `mi` can perform whatever machinations are required to carry out the change throughout $m > 0$; however, in the wide form, there are no machinations required.

subhazard, **cumulative subhazard**, and **subhazard ratio**. In a competing-risks analysis, the hazard of the subdistribution (or subhazard for short) for the event of interest (type 1) is defined formally as

$$\overline{h}_1(t) =$$

$$\lim_{\delta\to 0}\left\{\frac{P(t < T \leq t+\delta \text{ and event type 1 })|\ T > t \text{ or } (T \leq t \text{ and not event type 1})}{\delta}\right\}$$

Less formally, think of this hazard as that which generates failure events of interest while keeping subjects who experience competing events "at risk" so that they can be adequately counted as not having any chance of failing.

The cumulative subhazard $\overline{H}_1(t)$ is the integral of the subhazard function $\overline{h}_1(t)$, from 0 (the onset of risk) to t. The cumulative subhazard plays a very important role in competing-risks analysis. The cumulative incidence function (CIF) is a direct function of the cumulative subhazard:

$$\mathrm{CIF}_1(t) = 1 - \exp\{-\overline{H}_1(t)\}$$

The subhazard ratio is the ratio of the subhazard function evaluated at two different values of the covariates: $\overline{h}_1(t|\mathbf{x})/\overline{h}_1(t|\mathbf{x}_0)$. The subhazard ratio is often called the relative subhazard, especially when $\overline{h}_1(t|\mathbf{x}_0)$ is the baseline subhazard function.

subjective prior. See *informative prior*.

subpopulation estimation. Subpopulation estimation focuses on computing point and variance estimates for part of the population. The variance estimates measure the sample-to-sample variability, assuming that the same survey design is used to select individuals for observation from the population. This approach results in a different variance than measuring the sample-to-sample variability by restricting the samples to individuals within the subpopulation; see [SVY] **subpopulation estimation**.

subsampling the chain. See *thinning*.

subscripts. Subscripts are how you refer to an element or even a submatrix of a matrix.

Mata provides two kinds of subscripts, known as list subscripts and range subscripts.

In list subscripts, *A*`[2,3]` refers to the (2,3) element of *A*. *A*`[(2\3), (4,6)]` refers to the submatrix made up of the second and third rows, fourth and sixth columns, of *A*.

In range subscripts, *A*`[|2,3|]` also refers to the (2,3) element of *A*. *A*`[|2,3\4,6|]` refers to the submatrix beginning at the (2,3) element and ending at the (4,6) element.

See [M-2] **subscripts** for more information.

substantive constraints. See *identification*.

successive difference replication. Successive difference replication (SDR) is a method of variance typically applied to systematic samples, where the observed sampling units are somehow ordered. The SDR variance estimator is described in [SVY] **variance estimation**.

summary statistics data. Data are sometimes available only in summary statistics form, as (1) means and covariances, (2) means, standard deviations or variances, and correlations, (3) covariances, (4) standard deviations or variances and correlations, or (5) correlations. SEM can be used to fit models using such data in place of the underlying raw data. The `ssd` command creates datasets containing summary statistics.

super-varying variables. See *varying and super-varying variables*.

supplementary rows or columns or **supplementary variables**. Supplementary rows or columns can be included in CA, and supplementary variables can be included in MCA. They do not affect the CA or MCA solution, but they are included in plots and tables with statistics of the corresponding row or column points. Also see *correspondence analysis* and *multiple correspondence analysis*.

survey data. Survey data consist of information about individuals that were sampled from a population according to a survey design. Survey data distinguishes itself from other forms of data by the complex nature under which individuals are selected from the population.

In survey data analysis, the sample is used to draw inferences about the population. Furthermore, the variance estimates measure the sample-to-sample variability that results from the survey design applied to the fixed population. This approach differs from standard statistical analysis, in which the sample is used to draw inferences about a physical process and the variance measures the sample-to-sample variability that results from independently collecting the same number of observations from the same process.

survey design. A survey design describes how to sample individuals from the population. Survey designs typically include stratification and cluster sampling at one or more stages.

survival-time data. See *st data*.

survivor function. Also known as the survivorship function and the survival function, the survivor function, $S(t)$, is 1) the probability of surviving beyond time t, or equivalently, 2) the probability that there is no failure event prior to t, 3) the proportion of the population surviving to time t, or equivalently, 4) the reverse cumulative distribution function of T, the time to the failure event: $S(t) = \Pr(T > t)$. Also see *hazard*.

SVAR. A structural vector autoregressive (SVAR) model is a type of VAR in which short- or long-run constraints are placed on the resulting impulse–response functions. The constraints are usually motivated by economic theory and therefore allow causal interpretations of the IRFs to be made.

SVD. See *singular value decomposition*.

symmetric matrices. Matrix A is symmetric if $A = A'$. The word *symmetric* is usually reserved for real matrices, and in that case, a symmetric matrix is a square matrix with a_{ij}==a_{ji}.

Matrix A is said to be Hermitian if $A = A'$, where the transpose operator is understood to mean the conjugate-transpose operator; see *Hermitian matrix*. In Mata, the $'$ operator is the conjugate-transpose operator, and thus, in this manual, we will use the word *symmetric* both to refer to real, symmetric matrices and to refer to complex, Hermitian matrices.

Sometimes, you will see us follow the word *symmetric* with a parenthesized Hermitian, as in, "the resulting matrix is symmetric (Hermitian)". That is done only for emphasis.

The inverse of a symmetric (Hermitian) matrix is symmetric (Hermitian).

symmetriconly. Symmetriconly is a word we have coined to refer to a square matrix whose corresponding off-diagonal elements are equal to each other, whether the matrix is real or complex. Symmetriconly matrices have no mathematical significance, but sometimes, in data-processing and memory-management routines, it is useful to be able to distinguish such matrices.

symmetry. In a 2×2 contingency table, symmetry refers to the equality of the off-diagonal elements. For a 2×2 table, a test of marginal homogeneity reduces to a test of symmetry.

t test. A t test is a test for which the sampling distribution of the test statistic is a Student's t distribution.

A one-sample t test is used to test whether the mean of a population is equal to a specified value when the variance must also be estimated. The test statistic follows Student's t distribution with $N - 1$ degrees of freedom, where N is the sample size.

A two-sample t test is used to test whether the means of two populations are equal when the variances of the populations must also be estimated. When the two populations' variances are unequal, a modification to the standard two-sample t test is used; see *Satterthwaite's t test*.

target parameter. In power and sample-size analysis, the target parameter is the parameter of interest or the parameter in the study about which hypothesis tests are conducted.

target rotation. Target rotation minimizes the criterion

$$c(\mathbf{\Lambda}) = \frac{1}{2}||\mathbf{\Lambda} - \mathbf{H}||^2$$

for a given target matrix $\mathbf{H}$.

See *Crawford–Ferguson rotation* for a definition of $\mathbf{\Lambda}$.

taxonomy. Taxonomy is the study of the general principles of scientific classification. It also denotes classification, especially the classification of plants and animals according to their natural relationships. Cluster analysis is a tool used in creating a taxonomy and is synonymous with numerical taxonomy. Also see *cluster analysis*.

Taylor linearization. See *linearization*.

TCC. See *test characteristic curve*.

technique. Technique is just an English word and should be read in context. Nonetheless, technique is usually used here to refer to the technique used to calculate the estimated VCE. Those techniques are OIM, EIM, OPG, robust, clustered, bootstrap, and jackknife.

Technique is also used to refer to the available techniques used with `ml`, Stata's optimizer and likelihood maximizer, to find the solution.

test characteristic curve. A test characteristic curve (TCC) is the sum of item characteristic curves and represents the expected score on the instrument.

test information function A test information function (TIF) is the sum of item information functions and indicates the precision of the entire instrument along the latent trait continuum.

test statistic. In hypothesis testing, a test statistic is a function of the sample that does not depend on any unknown parameters.

tetrachoric correlation. A tetrachoric correlation estimates the correlation coefficients of binary variables by assuming a latent bivariate normal distribution for each pair of variables, with a threshold model for manifest variables.

thinning. Thinning is a way of reducing autocorrelation in the MCMC sample by subsampling the MCMC chain every prespecified number of iterations determined by the thinning interval. For example, the thinning interval of 1 corresponds to using the entire MCMC sample; the thinning interval of 2 corresponds to using every other sample value; and the thinning interval of 3 corresponds to using values from iterations 1, 4, 7, 10, and so on. Thinning should be applied with caution when used to reduce autocorrelation because it may not always be the most appropriate way of improving the precision of estimates.

thrashing. Subjects are said to thrash when they are censored and immediately reenter with different covariates.

three-level model. A three-level mixed-effects model has one level of observations and two levels of grouping. Suppose that you have a dataset consisting of patients overseen by doctors at hospitals, and each doctor practices at one hospital. Then a three-level model would contain a set of random effects to control for hospital-specific variation, a second set of random effects to control for doctor-specific random variation within a hospital, and a random-error term to control for patients' random variation.

three-parameter logistic model. The three-parameter logistic (3PL) model is an IRT model for binary responses where items vary in their difficulty and discrimination and can share or have their own guessing parameter.

ties. After discriminant analysis, ties in classification occur when two or more posterior probabilities are equal for an observation. They are most common with KNN discriminant analysis.

TIF. See *test information function*.

time-domain analysis. Time-domain analysis is analysis of data viewed as a sequence of observations observed over time. The autocorrelation function, linear regression, ARCH models, and ARIMA models are common tools used in time-domain analysis.

time-series–operated variable. Time-series–operated variables are a Stata concept. The term refers to *op*.*varname* combinations such as `L.gnp` to mean the lagged value of variable `gnp`. Mata's [M-5] **st_data()** function works with time-series–operated variables just as it works with other variables, but many other Stata-interface functions do not allow *op*.*varname* combinations. In those cases, you must use [M-5] **st_tsrevar()**.

time-varying covariates. Time-varying covariates appear in a survival model whose values vary over time. The values of the covariates vary, not the effect. For instance, in a proportional hazards model, the log hazard at time t might be $b \times \text{age}_t + c \times \text{treatment}_t$. Variable age might be time varying, meaning that as the subject ages, the value of age changes, which correspondingly causes the hazard to change. The effect b, however, remains constant.

Time-varying variables are either continuously varying or discretely varying.

In the continuously varying case, the value of the variable x at time t is $x_t = x_o + f(t)$, where $f()$ is some function and often is the identity function, so that $x_t = x_o + t$.

In the discretely varying case, the value of x changes at certain times and often in no particular pattern:

idvar	*t0*	*t*	*bp*
1	0	5	150
1	5	7	130
1	7	9	135

In the above data, the value of *bp* is 150 over the period $(0,5\,]$, then 130 over $(5,7\,]$, and 135 over $(7,9\,]$.

titlecase, **title-cased string**, and **Unicode title-cased string**. In grammar, titlecase refers to the capitalization of the key words in a phrase. In Stata, titlecase refers to (a) the capitalization of the first letter of each word in a string and (b) the capitalization of each letter after a nonletter character. There is no judgment of the word's importance in the string or whether the letter after a nonletter character is part of the same word. For example, "it's" in titlecase is "It'S".

A title-cased string is any string to which the above rules have been applied. For example, if we used the `strproper()` function with the book title *Zen and the Art of Motorcycle Maintenance*, Stata would return the title-cased string `Zen And The Art Of Motorcycle Maintenance`.

A Unicode title-cased string is a string that has had Unicode title-casing rules applied to Unicode words. This is almost, but not exactly, like capitalizing the first letter of each Unicode word. Like capitalization, title-casing letters is locale-dependent, which means that the same letter might have different titlecase forms in different locales. For example, in some locales, capital letters at the beginning of words are not supposed to have accents on them, even if that capital letter by itself would have an accent.

If you do not have characters beyond plain ASCII and your locale is English, there is no distinction in results. For example, `ustrtitle()` with an English locale locale also would return the title-cased string `Zen And The Art Of Motorcyle Maintenance`.

Use the `ustrtitle()` function to apply the appropriate capitalization rules for your language (locale).

total characteristic curve. See *test characteristic curve*.

total effects. See *direct, indirect, and total effects*.

total inertia or **total principal inertia**. The total (principal) inertia in CA and MCA is the sum of the principal inertias. In CA, total inertia is the Pearson χ^2/n. In CA, the principal inertias are the singular values; in MCA the principal inertias are the eigenvalues. Also see *correspondence analysis* and *multiple correspondence analysis*.

total information function. See *test information function*.

traceback log. When a function fails—either because of a programming error or because it was used incorrectly—it produces a traceback log:

```
: myfunction(2,3)
               solve():  3200  conformability error
               mysub():     -  function returned error
          myfunction():     -  function returned error
              <istmt>:     -  function returned error
r(3200);
```

The log says that `solve()` detected the problem—arguments are not conformable—and that `solve()` was called by `mysub()` was called by `myfunction()` was called by what you typed at the keyboard. See [M-2] **errors** for more information.

transmorphic. Transmorphic is an *eltype*. A scalar, vector, or matrix can be transmorphic, which indicates that its elements may be real, complex, string, pointer, or even a structure. The elements are all the same type; you are just not saying which they are. Variables that are not declared are assumed to be transmorphic, or a variable can be explicitly declared to be `transmorphic`. Transmorphic is just fancy jargon for saying that the elements of the scalar, vector, or matrix can be anything and that, from one instant to the next, the scalar, vector, or matrix might change from holding elements of one type to elements of another.

See [M-2] **declarations**.

transpose. The transpose operator is written different ways in different books, including $'$, superscript *, superscript T, and superscript H. Here we use the $'$ notation: A' means the transpose of A, A with its rows and columns interchanged.

In complex analysis, the transpose operator, however it is written, is usually defined to mean the conjugate transpose; that is, one interchanges the rows and columns of the matrix and then one takes the conjugate of each element, or one does it in the opposite order—it makes no difference. Conjugation simply means reversing the sign of the imaginary part of a complex number: the conjugate of `1+2i` is `1-2i`. The conjugate of a real is the number itself; the conjugate of 2 is 2.

In Mata, $'$ is defined to mean conjugate transpose. Since the conjugate of a real is the number itself, A' is regular transposition when A is real. Similarly, we have defined $'$ so that it performs regular transposition for string and pointer matrices. For complex matrices, however, $'$ also performs conjugation.

If you have a complex matrix and simply want to transpose it without taking the conjugate of its elements, see [M-5] **transposeonly()**. Or code `conj(`*A*`')`. The extra `conj()` will undo the undesired conjugation performed by the transpose operator.

Usually, however, you want transposition and conjugation to go hand in hand. Most mathematical formulas, generalized to complex values, work that way.

treatment model. A treatment model is a model used to predict treatment-assignment probabilities as a function of covariates and parameters.

trend. The trend specifies the long-run behavior in a time series. The trend can be deterministic or stochastic. Many economic, biological, health, and social time series have long-run tendencies to increase or decrease. Before the 1980s, most time-series analysis specified the long-run tendencies as deterministic functions of time. Since the 1980s, the stochastic trends implied by unit-root processes have become a standard part of the toolkit.

triangular matrix. A triangular matrix is a matrix with all elements equal to zero above the diagonal or all elements equal to zero below the diagonal.

A matrix A is *lower triangular* if all elements are zero above the diagonal, that is, if A`[`i`,`j`]==0`, $j > i$.

A matrix A is *upper triangular* if all elements are zero below the diagonal, that is, if A`[`i`,`j`]==0`, $j < i$.

A *diagonal matrix* is both lower and upper triangular. That is worth mentioning because any function suitable for use with triangular matrices is suitable for use with diagonal matrices.

A triangular matrix is usually *square*.

The inverse of a triangular matrix is a triangular matrix. The determinant of a triangular matrix is the product of the diagonal elements. The eigenvalues of a triangular matrix are the diagonal elements.

truncation, left-truncation, and right-truncation. In survival analysis, truncation occurs when subjects are observed only if their failure times fall within a certain observational period of a study. Censoring, on the other hand, occurs when subjects are observed for the whole duration of a study, but the exact times of their failures are not known; it is known only that their failures occurred within a certain time span.

Left-truncation occurs when subjects come under observation only if their failure times exceed some time t_l. It is only because they did not fail before t_l that we even knew about their existence. Left-truncation differs from left-censoring in that, in the censored case, we know that the subject failed before time t_l, but we just do not know exactly when.

Imagine a study of patient survival after surgery, where patients cannot enter the sample until they have had a post-surgical test. The patients' survival times will be left-truncated. This is a "delayed entry" problem, one common type of left-truncation.

Right-truncation occurs when subjects come under observation only if their failure times do not exceed some time t_r. Right-truncated data typically occur in registries. For example, a cancer registry includes only subjects who developed a cancer by a certain time, and thus survival data from this registry will be right-truncated.

two-independent-samples test. See *two-sample test*.

two-level model. A two-level mixed-effects model has one level of observations and one level of grouping. Suppose that you have a panel dataset consisting of patients at hospitals; a two-level model would contain a set of random effects at the hospital level (the second level) to control for hospital-specific random variation and a random-error term at the observation level (the first level) to control for within-hospital variation.

two-parameter logistic model. The two-parameter logistic (2PL) model is an IRT model for binary responses where items vary in their difficulty and discrimination.

two-sample paired test. See *paired test*.

two-sample test. A two-sample test is used to test whether the parameters of interest of the two independent populations are equal. For example, two-sample means test, two-sample variances, two-sample proportions test, two-sample correlations test.

two-sided test, two-tailed test. A two-sided test is a hypothesis test of a parameter in which the alternative hypothesis is the complement of the null hypothesis. In the context of a test of a scalar parameter, the alternative hypothesis states that the parameter is less than or greater than the value conjectured under the null hypothesis.

two-way ANOVA, **two-way analysis of variance**. A two-way ANOVA model contains two factors. Also see [PSS] **power twoway**.

two-way repeated-measures ANOVA, **two-factor** ANOVA. This is a repeated-measures ANOVA model with one within-subject factor and one between-subjects factor. The model can be additive (contain only main effects of the factors) or can contain main effects and an interaction between the two factors. Also see [PSS] **power repeated**.

type, eltype, and orgtype. The *type* of a matrix (or vector or scalar) is formally defined as the matrix's *eltype* and *orgtype*, listed one after the other—such as `real vector`—but it can also mean just one or the other—such as the *eltype* `real` or the *orgtype* `vector`.

eltype refers to the type of the elements. The *eltypes* are

`real`	numbers such as 1, 2, 3.4
`complex`	numbers such as 1+2i, 3+0i
`string`	strings such as `"bill"`
`pointer`	pointers such as &*varname*
`struct`	structures
`numeric`	meaning `real` or `complex`
`transmorphic`	meaning any of the above

orgtype refers to the organizational type. *orgtype* specifies how the elements are organized. The *orgtypes* are

`matrix`	two-dimensional arrays
`vector`	one-dimensional arrays
`colvector`	one-dimensional column arrays
`rowvector`	one-dimensional row arrays
`scalar`	single items

The fully specified type is the element and organization types combined, as in `real vector`.

type I error or **false-positive result**. The type I error of a test is the error of rejecting the null hypothesis when it is true. The probability of committing a type I error, significance level of a test, is often denoted as α in statistical literature. One traditionally used value for α is 5%. Also see *type II error* and *power*.

type I error probability. See *probability of a type I error*.

type I study. A type I study is a study in which all subjects fail (or experience an event) by the end of the study; that is, no censoring of subjects occurs.

type II error or **false-negative result**. The type II error of a test is the error of not rejecting the null hypothesis when it is false. The probability of committing a type II error is often denoted as β in statistical literature. Commonly used values for β are 20% or 10%. Also see *type I error* and *power*.

type II error probability. See *probability of a type II error*.

type II study. A type II study is a study in which there are subjects who do not fail (or do not experience an event) by the end of the study. These subjects are known to be censored.

unary operator. A unary operator is an operator applied to one argument. In −2, the minus sign is a unary operator. In `!(`*a*`==`*b* `|` *a*`==`*c*`)`, `!` is a unary operator.

unbalanced data. A longitudinal or panel dataset is said to be unbalanced if each panel does not have the same number of observations. See also *weakly balanced* and *strongly balanced*.

unbalanced design. An unbalanced design indicates an experiment in which the numbers of treated and untreated subjects differ. Also see [PSS] **unbalanced designs**.

unconfoundedness. See *conditional-independence assumption*.

under observation. A subject is under observation when failure events, should they occur, would be observed (and so recorded in the dataset). Being under observation does not mean that a subject is necessarily at risk. Subjects usually come under observation before they are at risk. The statistical concern is with periods when subjects are at risk but not under observation, even when the subject is (later) known not to have failed during the hiatus.

In such cases, since failure events would not have been observed, the subject necessarily had to survive the observational hiatus, and that leads to bias in statistical results unless the hiatus is accounted for properly.

Entry time and exit time record when a subject first and last comes under observation, between which there may be observational gaps, but usually there are not. There is only one entry time and one exit time for each subject. Often, entry time corresponds to analysis time $t = 0$, or before, and exit time corresponds to the time of failure.

Delayed entry means that the entry time occurred after $t = 0$.

underscore functions. *Functions* whose names start with an underscore are called underscore functions, and when an underscore function exists, usually a function without the underscore prefix also exists. In those cases, the function is usually implemented in terms of the underscore function, and the underscore function is harder to use but is faster or provides greater control. Usually, the difference is in the handling of errors.

For instance, function `fopen()` opens a file. If the file does not exist, execution of your program is aborted. Function `_fopen()` does the same thing, but if the file cannot be opened, it returns a special value indicating failure, and it is the responsibility of your program to check the indicator and to take the appropriate action. This can be useful when the file might not exist, and if it does not, you wish to take a different action. Usually, however, if the file does not exist, you will wish to abort, and use of `fopen()` will allow you to write less code.

unequal-allocation design. See *unbalanced design*.

Unicode. Unicode is a standard for encoding and dealing with text written in almost any conceivable living or dead language. Unicode specifies a set of encoding systems that are designed to hold (and, unlike extended ASCII, to keep separate) characters used in different languages. The Unicode standard defines not only the characters and encodings for them, but also rules on how to perform various operations on words in a given language (locale), such as capitalization and ordering. The most common Unicode encodings are mUTF-8, UTF-16, and UTF-32. Stata uses UTF-8.

Unicode character. Technically, a Unicode character is any character with a Unicode encoding. Colloquially, we use the term to refer to any character other than the plain ASCII characters.

Unicode normalization. Unicode normalization allows us to use a common representation and therefore compare Unicode strings that appear the same when displayed but could have more than one way of being encoded. This rarely arises in practice, but because it is possible in theory, Stata provides the `ustrnormalize()` function for converting between different normalized forms of the same string.

For example, suppose we wish to search for "ñ" (the lowercase n with a tilde over it from the Spanish alphabet). This letter may have been encoded with the single code point U+00F1. However, the sequence U+006E (the Latin lowercase "n") followed by U+0303 (the tilde) is defined by Unicode to be equivalent to U+00F1. This type of visual identicalness is called canonical equivalence. The one-code-point form is known as the canonical composited form, and the multiple-code-point form is known as the canonical decomposed form. Normalization modifies one or the other string to the opposite canonical equivalent form so that the underlying byte sequences match. If we had strings in a mixture of forms, we would want to use this normalization when sorting or when searching for strings or substrings.

Another form of Unicode normalization allows characters that appear somewhat different to be given the same meaning or interpretation. For example, when sorting or indexing, we may want the code point U+FB00 (the typographic ligature "ff") to match the sequence of two Latin "f" letters encoded as U+0066 U+0066. This is called compatible equivalence.

Unicode title-cased string. See *titlecase, title-cased string, and Unicode title-cased string*.

unidimensionality. See *latent space*.

uniqueness. In factor analysis, the uniqueness is the percentage of a variable's variance that is not explained by the common factors. It is also "1 − communality". Also see *communality*.

unitary matrix. See *orthogonal matrix*.

unit-root process. A unit-root process is one that is integrated of order one, meaning that the process is nonstationary but that first-differencing the process produces a stationary series. The simplest example of a unit-root process is the random walk. See Hamilton (1994, chap. 15) for a discussion of when general ARMA processes may contain a unit root.

unit-root tests. Whether a process has a unit root has both important statistical and economic ramifications, so a variety of tests have been developed to test for them. Among the earliest tests proposed is the one by Dickey and Fuller (1979), though most researchers now use an improved variant called the augmented Dickey–Fuller test instead of the original version. Other common unit-root tests implemented in Stata include the DF–GLS test of Elliott, Rothenberg, and Stock (1996) and the Phillips–Perron (1988) test. See [TS] **dfuller**, [TS] **dfgls**, and [TS] **pperron**.

Variants of unit-root tests suitable for panel data have also been developed; see [XT] **xtunitroot**.

unregistered variables. See *registered and unregistered variables*.

unrestricted transformation. An unrestricted transformation is a Procrustes transformation that allows the data to be transformed, not just by orthogonal and oblique rotations, but by all conformable regular matrices. This is equivalent to a multivariate regression. Also see *Procrustes transformation* and *multivariate regression*.

unstandardized coefficient. A coefficient that is not standardized. If $\texttt{mpg} = -0.006 \times \texttt{weight} + 39.44028$, then -0.006 is an unstandardized coefficient and, as a matter of fact, is measured in mpg-per-pound units.

unstructured and structured (correlation or covariance). A set of variables, typically error variables, is said to have an unstructured correlation or covariance if the covariance matrix has no particular pattern imposed by theory. If a pattern is imposed, the correlation or covariance is said to be structured.

upper one-sided test, **upper one-tailed test**. An upper one-sided test is a one-sided test of a scalar parameter in which the alternative hypothesis is upper one sided, meaning that the alternative hypothesis states that the parameter is greater than the value conjectured under the null hypothesis. Also see *One-sided test versus two-sided test* under *Remarks and examples* in [PSS] **intro**.

UTF-8. UTF-8 stands for Universal character set + Transformation Format—8-bit. It is a type of Unicode encoding system that was designed for backward compatibility with ASCII and is used by Stata 14.

vague prior. See *noninformative prior*.

valid initial state. See *feasible initial value*.

vanishing adaptation. See *diminishing adaptation*.

VAR. A vector autoregressive (VAR) model is a multivariate regression technique in which each dependent variable is regressed on lags of itself and on lags of all the other dependent variables in the model. Occasionally, exogenous variables are also included in the model.

variable. In a program, the entities that store values (a, b, c, $\ldots$, x, y, z) are called variables. Variables are given names of 1 to 32 characters long. To be terribly formal about it: a variable is a container; it contains a matrix, vector, or scalar and is referred to by its variable name or by another variable containing a *pointer* to it.

Also, *variable* is sometimes used to refer to columns of data matrices; see *data matrix*.

variance components. In a mixed-effects model, the variance components refer to the variances and covariances of the various random effects.

variance–covariance matrix of the estimator. The estimator is the formula used to solve for the fitted parameters, sometimes called the fitted coefficients. The VCE is the estimated variance–covariance matrix of the parameters. The diagonal elements of the VCE are the variances of the parameters or equivalent, the square root of those elements are the reported standard errors of the parameters.

variance estimation. Variance estimation refers to the collection of methods used to measure the amount of sample-to-sample variation of point estimates; see [SVY] **variance estimation**.

varimax rotation. Varimax rotation maximizes the variance of the squared loadings within the columns of the matrix. It is an orthogonal rotation equivalent to oblimin with $\gamma = 1$ or to the Crawford–Ferguson family with $\kappa = 1/p$, where p is the number of rows of the matrix to be rotated. Also see *orthogonal rotation*, *oblimin rotation*, and *Crawford–Ferguson rotation*.

varying and super-varying variables. A variable is said to be varying if its values in the incomplete observations differ across m. Imputed and passive variables are varying. Regular variables are nonvarying. Unregistered variables can be either.

Imputed variables are supposed to vary because their incomplete values are filled in with different imputed values, although an imputed variable can be temporarily nonvarying if you have not imputed its values yet. Similarly, passive variables should vary because they are or will be filled in based on values of varying imputed variables.

VCE. See *variance–covariance matrix of the estimator*.

VECM. A vector error-correction model (VECM) is a type of VAR that is used with variables that are cointegrated. Although first-differencing variables that are integrated of order one makes them stationary, fitting a VAR to such first-differenced variables results in misspecification error if the variables are cointegrated. See *The multivariate VECM specification* in [TS] **vec intro** for more on this point.

vector, colvector, and rowvector. A special case of a matrix with either one row or one column. A vector may be substituted anywhere a matrix is required. A matrix, however, may not be substituted for a vector.

A `colvector` is a vector with one column.

A `rowvector` is a vector with one row.

A `vector` is either a `rowvector` or `colvector`, without saying which.

view. A view is a special type of matrix that appears to be an ordinary matrix, but in fact the values in the matrix are the values of certain or all variables and observations in the Stata dataset that is currently in memory. Its values are not just equal to the dataset's values; they are the dataset's values: if an element of the matrix is changed, the corresponding variable and observation in the Stata dataset also changes. Views are obtained by `st_view()` and are efficient; see [M-5] **st_view()**.

void function. A function is said to be void if it returns nothing. For instance, the function [M-5] **printf( )** is a void function; it prints results, but it does not return anything in the sense that, say, [M-5] **sqrt( )** does. It would not make any sense to code `x = printf("hi there")`, but coding `x = sqrt(2)` is perfectly logical.

void matrix. A matrix is said to be void if it is 0×0, $r \times 0$, or $0 \times c$; see [M-2] **void**.

Wald test. A Wald test is a statistical test based on the estimated variance–covariance matrix of the parameters. Wald tests are especially convenient for testing possible constraints to be placed on the estimated parameters of a model. Also see *score test*.

Ward's linkage clustering. Ward's-linkage clustering is a hierarchical clustering method that joins the two groups resulting in the minimum increase in the error sum of squares.

weakly balanced. A longitudinal or panel dataset is said to be weakly balanced if each panel has the same number of observations but the observations for different panels were not all made at the same times.

weighted least squares. Weighted least squares (WLS) is a method used to obtain fitted parameters. In this documentation, WLS is referred to as ADF, which stands for asymptotic distribution free. Other available methods are ML, QML, and MLMV. ADF is, in fact, a specific kind of the more generic WLS.

weighted-average linkage clustering. Weighted-average linkage clustering is a hierarchical clustering method that uses the weighted average similarity or dissimilarity of the two groups as the measure between the two groups.

weighted-regression-adjustment estimator. Weighted-regression-adjustment estimators use means of predicted outcomes for each treatment level to estimate each potential-outcome mean. The weights are used to estimate censoring-adjusted regression coefficients.

white noise. A variable u_t represents a white-noise process if the mean of u_t is zero, the variance of u_t is σ^2, and the covariance between u_t and u_s is zero for all $s \neq t$.

wide data. See *style*.

Wilks's lambda. Wilks's lambda is a test statistic for the hypothesis test $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \cdots = \boldsymbol{\mu}_k$ based on the eigenvalues $\lambda_1, \ldots, \lambda_s$ of $\mathbf{E}^{-1}\mathbf{H}$. It is defined as

$$\Lambda = \frac{|\mathbf{E}|}{|\mathbf{E}+\mathbf{H}|} = \prod_{i=1}^{s} \frac{1}{1+\lambda_i}$$

where $\mathbf{H}$ is the between matrix and $\mathbf{E}$ is the within matrix. See *between matrix*.

Wishart distribution. The Wishart distribution is a family of probability distributions for nonnegative-definite matrix-valued random variables ("random matrices"). These distributions are of great importance in the estimation of covariance matrices in multivariate statistics.

withdrawal. Withdrawal is the process under which subjects withdraw from a study for reasons unrelated to the event of interest. For example, withdrawal occurs if subjects move to a different area or decide to no longer participate in a study. Withdrawal should not be confused with administrative censoring. If subjects withdraw from the study, the information about the outcome those subjects would have experienced at the end of the study, had they completed the study, is unavailable. Also see *loss to follow-up* and *administrative censoring*.

within estimator. The within estimator is a panel-data estimator that removes the panel-specific heterogeneity by subtracting the panel-level means from each variable and then performing ordinary least squares on the demeaned data. The within estimator is used in fitting the linear fixed-effects model.

within matrix. See *between matrix*.

within-subject design. This is an experiment that has at least one within-subject factor. See [PSS] **power repeated**.

within-subject factor. This is a factor for which each subject receives several or all the levels.

WLF. See *worst linear function*.

WLS. See *weighted least squares*.

worst linear function. A linear combination of all parameters being estimated by an iterative procedure that is thought to converge slowly.

Yule–Walker equations. The Yule–Walker equations are a set of difference equations that describe the relationship among the autocovariances and autocorrelations of an autoregressive moving-average (ARMA) process.

z test. A z test is a test for which a potentially asymptotic sampling distribution of the test statistic is a normal distribution. For example, a one-sample z test of means is used to test whether the mean of a population is equal to a specified value when the variance is assumed to be known. The distribution of its test statistic is normal. See [PSS] **power onemean**, [PSS] **power twomeans**, and [PSS] **power pairedmeans**.

Zellner's g-prior. Zellner's g-prior is a form of a weakly informative prior for the regression coefficients in a linear model. It accounts for the correlation between the predictor variables and controls the impact of the prior of the regression coefficients on the posterior with parameter g. For example, $g = 1$ means that prior weight is 50% and $g \to \infty$ means diffuse prior.

References

Bartlett, M. S. 1937. The statistical conception of mental factors. *British Journal of Psychology* 28: 97–104.

——. 1938. Methods of estimating mental factors. *Nature, London* 141: 609–610.

Bellman, R. E. 1961. *Adaptive Control Processes*. Princeton, NJ: Princeton University Press.

Bentler, P. M. 1977. Factor simplicity index and transformations. *Psychometrika* 42: 277–295.

Bentler, P. M., and D. G. Weeks. 1980. Linear structural equations with latent variables. *Psychometrika* 45: 289–308.

Breusch, T. S., and A. R. Pagan. 1980. The Lagrange multiplier test and its applications to model specification in econometrics. *Review of Economic Studies* 47: 239–253.

Chatfield, C. 2004. *The Analysis of Time Series: An Introduction*. 6th ed. Boca Raton, FL: Chapman & Hall/CRC.

Comrey, A. L. 1967. Tandem criteria for analytic rotation in factor analysis. *Psychometrika* 32: 277–295.

Cox, T. F., and M. A. A. Cox. 2001. *Multidimensional Scaling*. 2nd ed. Boca Raton, FL: Chapman & Hall/CRC.

Crawford, C. B., and G. A. Ferguson. 1970. A general rotation criterion and its use in orthogonal rotation. *Psychometrika* 35: 321–332.

Davidson, R., and J. G. MacKinnon. 1993. *Estimation and Inference in Econometrics*. New York: Oxford University Press.

Dickey, D. A., and W. A. Fuller. 1979. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association* 74: 427–431.

Durbin, J. 1970. Testing for serial correlation in least-squares regressions when some of the regressors are lagged dependent variables. *Econometrica* 38: 410–421.

Durbin, J., and G. S. Watson. 1950. Testing for serial correlation in least squares regression. I. *Biometrika* 37: 409–428.

——. 1951. Testing for serial correlation in least squares regression. II. *Biometrika* 38: 159–177.

——. 1971. Testing for serial correlation in least squares regression. III. *Biometrika* 58: 1–19.

Elliott, G. R., T. J. Rothenberg, and J. H. Stock. 1996. Efficient tests for an autoregressive unit root. *Econometrica* 64: 813–836.

Fisher, R. A. 1936. The use of multiple measurements in taxonomic problems. *Annals of Eugenics* 7: 179–188.

Hamilton, J. D. 1994. *Time Series Analysis*. Princeton: Princeton University Press.

Hendrickson, A. E., and P. O. White. 1964. Promax: A quick method for rotation to oblique simple structure. *British Journal of Statistical Psychology* 17: 65–70.

Jennrich, R. I. 2004. Rotation to simple loadings using component loss functions: The orthogonal case. *Psychometrika* 69: 257–273.

Kaiser, H. F. 1974. An index of factor simplicity. *Psychometrika* 39: 31–36.

Kenward, M. G., and J. H. Roger. 1997. Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics* 53: 983–997.

Kruskal, J. B. 1964. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* 29: 1–27.

Mahalanobis, P. C. 1936. On the generalized distance in statistics. *National Institute of Science of India* 12: 49–55.

Phillips, P. C. B., and P. Perron. 1988. Testing for a unit root in time series regression. *Biometrika* 75: 335–346.

Sammon, J. W., Jr. 1969. A nonlinear mapping for data structure analysis. *IEEE Transactions on Computers* 18: 401–409.

Satterthwaite, F. E. 1946. An approximate distribution of estimates of variance components. *Biometrics Bulletin* 2: 110–114.

Thomson, G. H. 1951. *The Factorial Analysis of Human Ability*. London: University of London Press.

Wei, W. W. S. 2006. *Time Series Analysis: Univariate and Multivariate Methods*. 2nd ed. Boston: Pearson.

Vignette index

Author index

A

B

C

D

E

F

G

H

I

J

L

M

N

O

P

Q

R

S

T

U

V

W

X

Y

Z

Subject index

Symbols

A

C

D

F

G

H

J

K

L

O

P

Q

R

T

U

W

X

Y

Z