

A. Colin Cameron

Analysis of Economics Data: An Introduction to Econometrics

<http://cameron.econ.ucdavis.edu/aed/>

Version 1.1 February 2022

Index

- Acemoglu, A., 301, 365, 501
- added variable plot, 369
- adjusted R-squared, 208
 - and overall F test, 234
 - compared to t test, 224
- Akaike information criterion, 208
- analysis of variance table, 108, 210
- Angrist, J. D., 416, 501
- ANOVA, 327
- Aron-Dine, A., 292, 501
- assumptions
 - for causal methods, 391
 - for regression, 126, 133, 146, 218, 358, 464
 - for sample mean, 41
 - relaxing for regression, 137, 158, 222
- asymptotic properties, 43, 48, 136, 139, 222, 250
- asymptotically normal, 43, 136, 222, 242, 467
- autocorrelated errors, 173, 247, 364
- autocorrelation, 100, 403, 411
 - tests for, 404, 411
 - using a statistical package, 404
- autoregressive distributed lag model, 405, 414
- autoregressive error model, 405
- autoregressive model, 404, 414
- average absolute deviation, 11
- average marginal effect, 335, 398, 400
- bar chart, 19
- basis point change, 28
- Bayesian information criterion, 208, 408
- Bayesian methods, 265
 - posterior distribution, 265
 - prior information, 265
 - uninformative prior, 266
- before-after comparison, 295
- Bernoulli distribution, 81, 454
- bootstrap standard errors, 249
 - pairs bootstrap, 249
- box-and-whisker plot, 12
- Breusch-Godfrey test, 404
- Calonico, S., 298, 501
- Cameron, A. C., 416, 501
- Card, D. E., 389, 501
- case study
 - automobile efficiency, 288–291
 - capital asset pricing model, 170–173
 - cleaning data, 303–308
 - Cobb-Douglas production function, 280–284
 - gains from political incumbency, 298–300
 - health and access to health, 295–298
 - health expenditures across countries, 170
 - health insurance experiment, 291–295
 - health outcomes across countries, 165–168
 - health spending across countries, 168
 - institutions and country growth, 301–303
 - output and unemployment, 173–176
 - Phillips curve, 284–288
 - school academic performance, 275–280
- categorical data, 2, 96, 315, 397
- categorical variables, *see* indicator variables
- Cattaneo, M., 298, 501
- causal methods, 391–396
 - applications, 291–303
 - difference-in-differences, 295–298, 393
 - experiments, 291–295, 393
 - instrumental variables, 301–303, 387–391
 - inverse probability weighting, 396
 - local average treatment effect, 395
 - matching, 396

- propensity score, 396
- regression adjustment, 393
- regression discontinuity design, 298–300, 394
- causation, 112–113, 206
 - and correlation, 112
- Census example, 45–46, 131–132
- central limit theorem, 43, 61, 136, 146, 220, 458
- charts, 15–22
- Chebyshev's inequality, 12
- chi-squared distribution, 232, 451
 - compared to F distribution, 232
 - tables of critical values, 500
- cluster-robust standard errors, 245–246, 364, 378, 380, 467
 - degrees of freedom, 246
 - using a statistical package, 250–251
- clustered data, 378–380
 - using a statistical package, 383
- clustered errors, 245, 364
- Cobb, C. W., 280, 501
- Cochrane-Orcutt estimator, 405
- coefficient of variation, 11
- coin toss example, 36, 40, 50, 454
- cointegrated series, 407
- collinear variables, *see* multicollinearity
- column chart, 19
- complier, 395
- component plus residual plot, 369
- compounding, 187
 - continuous compounding, 191
- conditional mean, 125, 126, 218
 - definition, 461
 - prediction, 252–258
- conditional probability, 460
- conditional variance, 461
- confidence interval
 - best confidence intervals, 261
 - critical value, 66, 148, 223
 - derivation, 66
 - for actual value forecast, 254
 - for conditional mean, 254
 - for proportions data, 81
 - for slope parameter, 148–150, 222–223
 - for the mean, 65–69
 - generalizations, 79
 - interpretation, 68, 149
 - margin of error, 80, 82
 - two standard error interval, 68, 149
 - what level of confidence, 67, 148
- consistent estimator, 48, 139, 221, 358
- convergence in probability, 458
- correlation, 97–100, 200, 201, 462
 - and causation, 112
 - and covariance, 98
 - and R-squared, 107, 207
 - and regression, 111
 - and tests of statistical significance, 152
 - autocorrelation, 100, 403, 411
 - definition, 97
 - strength of correlation, 99
- covariance, 98, 462
- covariate, 101, 201
- critical value, 64, 72, 151, 224, 230
 - F distribution, 228
 - tables, 493–500
 - using a statistical package, 451
- cross tabulation, 95
- cross-section data, 3, 377–402
 - fixed effects, 379, 394
 - random effects, 378
- cumulative distribution function, 454
- Cunningham, S., 416, 502
- data analytics, 264
- data cleaning, 303–308
- data collection methods, 3
- data science, 264
- data transformation, 22–28, 307, 333–353, 403
- data types, 2–3
- dataset in main text
 - API99, 275
 - AUTOSMPG, 288
 - CAPM, 170
 - CENSUSAGEMEANS, 46
 - CENSUSREGRESSIONS, 132

- COBBDOUGLAS, 280
- COINTOSSMEANS, 40
- DEMOCRACY, 365
- EARNINGS, 7, 59, 185
- EARNINGS_COMPLETE, 315, 333, 399
- EARNINGSMALE, 74
- FISHING, 22
- GASPRICE, 73
- GDPEMPLOY, 174
- GENERATEDDATA, 129
- GENERATEDREGRESSIONS, 131
- HEALTH2009, 165
- HEALTHACCESS, 296
- HEALTHCATEGORIES, 19
- HEALTHINSEXP, 292
- HOUSE, 94, 145, 197, 223, 244
- INCUMBENCY, 298
- INSTITUTIONS, 301
- INTERESTRATES, 409
- MONTHLYHOMESALES, 25
- NBA, 384
- PHILLIPS, 284
- PSID raw data, 303
- REALGDPPC, 19, 26, 75, 249
- RETURNSTOSCHOOLING, 389
- SP500INDEX, 188
- default standard errors, 158, 241, 465
- degrees of freedom, 39, 63, 105, 206, 227, 242
 - cluster-robust standard errors, 246
- dependent variable, 101, 201
- descriptive statistics, *see* summary statistics
- DFBETA, 370
- DFITS, 370
- diagnostics, 367–371
 - influential observation, 368, 370
 - outlying observation, 367
 - plots, 368
- Dickey-Fuller test
 - in first differences, 413
- difference in means, *see* indicator variables
- difference in means tests, 317–318
 - using a statistical package, 318
- difference-in-differences, 295–298, 393
- disturbance term, *see* error term
- Douglas, P. H., 280, 501
- Duan, N., 292, 502
- dummy variable, 315
- dummy variable trap, 323
- dummy variables, *see* indicator variables
- economic significance
 - versus statistical significance, 151, 225
- Einav, L., 292, 501
- elasticities, 181–187, 342–344, 400
- endogenous regressor, 362, 387
- error sum of squares, 107, 207
- error term, 126, 218
 - autocorrelated, 247, 468
 - clustered, 245, 467
 - compared to residual, 127, 371
 - correlated with regressor, 362
 - definition, 126, 218
 - heteroskedastic, 159, 243, 466
 - homoskedastic, 134, 219, 465
- estimate
 - defined, 47
- estimator
 - best linear unbiased, 139, 221
 - best unbiased, 139, 222
 - consistent, 48, 221, 361, 362
 - defined, 47
 - efficient, 47
 - minimum variance, 47, 139, 221, 260, 473
 - unbiased, 47, 138, 221
- Eviews, *see also* statistical packages
 - essentials, 445–449
 - robust standard errors, 251
- Excel, 449–451
- exogenous regressor, 362, 388
- expected value, 36, 454
- experimental data, 3
- experiments, 291–295, 393
- explained sum of squares, 106, 207
- explanatory variable, 101, 201
- exponential function, 190

- continuous compounding, 191
- exponential growth
 - linearizing, 188
- F distribution, 227–228, 451
 - compared to chi-squared distribution, 232
 - critical value, 228
 - degrees of freedom, 227
 - inverse probabilities, 228
 - mean, 227
 - probabilities, 228
 - tables of critical values, 493
- F statistic, 229
- F tests, 229–234
 - and adjusted R-squared, 234
 - compared to t test, 231
 - complete model, 229
 - computation in a statistical package, 231
 - of exclusion restrictions, 230, 234
 - of overall significance, 230, 233
 - of statistical significance, 231
 - of subsets of regressors, 230, 234
 - reduced model, 229
 - restricted model, 229
 - under assumptions 1-4, 232–234
 - unrestricted model, 229
- Farrell, M., 298, 501
- feasible generalized least squares, 260
- finite distributed lag model, 405, 414
- Finkelstein, A., 292, 501
- fitted value, 101, 201
- fixed effects estimator, 379, 382, 386, 394
 - using a statistical package, 383
- forecast, 254
 - confidence interval, 254
 - with time-series data, 408
- Gauss-Markov theorem, 221
- generalized least squares, 260
- Gini coefficient, 11
- Google sheets, 449–451
- Greene, W. H., 416, 502
- Gretl, *see also* statistical packages
 - essentials, 440–445
 - robust standard errors, 251
- grouped data, *see* clustered data
- growth rate, 28
 - rule of 72, 187
- HAC-robust standard errors, 247–249, 402, 468
 - using a statistical package, 250–251
- Hansen, B. E., 416, 502
- Hastie, T., 416, 502
- heteroskedastic errors, 159, 243–245, 363
 - and estimator properties, 466
- heteroskedastic-robust standard errors, 159–160, 243–245, 466
 - using a statistical package, 250–251
- histogram, 16–18
 - smoothed histogram, 17
 - using a statistical package, 18
- homoskedastic errors, 134, 219
 - and estimator properties, 463, 466
- hypothesis tests
 - best tests, 262
 - critical value, 72, 155, 230
 - examples for the mean, 73–76
 - F tests, 227–234
 - for proportions data, 81
 - for slope parameter, 158, 223–225
 - for the mean, 69, 80
 - generalizations, 80
 - joint hypothesis tests, 227–234
 - most powerful test, 263
 - multiple testing, 232
 - of autocorrelation, 404
 - of exclusion restrictions, 230
 - of overall significance, 230
 - of statistical significance, 150, 224, 231
 - of subsets of regressors, 230
 - on interacted indicator variables, 320
 - on sets of indicator variables, 324
 - one-sided, 76–79, 156–158
 - p-value, 71, 155, 230
 - rejection using critical region, 72, 155, 224
 - rejection using p-values, 71, 155, 224

- relationship with confidence interval, 73, 155
 - significance level, 70, 153, 263
 - test power, 263
 - test size, 70, 153, 263
 - two-sided, 76, 153–156
 - type I error, 69, 153, 262
 - type II error, 262
 - which significance level, 72, 151
- impulse response function, 415
- inconsistent estimator, 358–363
- independence, *see* statistical independence
- independent variable, 101, 201
- indicator variables, 315–329
 - and additional controls, 319–320
 - ANOVA, 327
 - as dependent variable, 322, 397–399
 - base category, 323
 - creating in statistical package, 327
 - difference in means, 104, 316, 393
 - difference in several means, 327
 - dummy variable trap, 323
 - hypothesis tests, 324
 - interacted, 320, 322
 - mutually exclusive, 323
 - omitted category, 323
 - reference category, 323
 - sets of indicator variables, 322–328
 - single indicator variable, 315–318
- inequality measures, 11
- influential observation, 368
 - DFBETA, 370
 - DFITS, 370
- information criteria, 208
- instrument, 387
- instrumental variables, 301–303, 387–391, 470
 - Anderson-Rubin test, 390
 - complier, 395
 - local average treatment effect, 395
 - two-stage least squares estimator, 389
 - using a statistical package, 391
 - weak instruments, 390
- interacted variables, 340–342
 - with indicator variables, 322
- intercept coefficient, 101, 102, 201
- intercept-only regression, 104, 248
- interquartile range, 11
- inverse probability weighting, 396
- irrelevant variables, 359
- James, G., 416, 502
- jitter, 97
- Johnson, S., 301, 365, 501
- joint hypothesis tests, *see* F tests
- Keeler, E. B., 292, 502
- kernel density estimate, 17
 - using a statistical package, 18
- Kling, J. R., 389, 502
- Knittel, C., 288, 502
- Krueger, A., 352, 502
- kurtosis statistic, 15
- law of large numbers, 48, 458
- least squares regression, *see* regression
- Leibowitz, A., 292, 502
- likelihood function, 473
- line chart, 18
- linear probability model, 399
- linear-log model, 183
- Ljung-Box test
 - of autocorrelation, 404
- local average treatment effect, 395
- local constant regression, 115
- local linear regression, 116
- local projection method, 415
- loess regression, 116
- log-likelihood function, 473
- log-linear model, 183, 342
 - prediction, 345
- log-log model, 183, 280, 343
 - prediction, 344
- logit regression, 397–399, 474
 - marginal effects, 398
- longitudinal data, 3, *see* panel data
- lower quartile, 10

- machine learning, 264
- Manning, W. G., 292, 502
- margin of error, 80, 82
- marginal effect at a representative value, 335
- marginal effect at the mean, 335
- marginal effects, 333–337
 - average marginal effect, 335
 - calculus method, 334
 - computing in a statistical package, 336, 402
 - computing in a statistical package, 402
 - finite difference method, 334
 - for logit regression, 398
 - for Poisson regression, 400
 - for probit regression, 398
- Marquis, M. S., 292, 502
- matching estimators, 396
- matrix algebra for regression, 471–474
- maximum likelihood estimator, 401, 473
- mean of a random variable
 - conditional mean, 461
 - definition, 36, 454
 - estimator for, 47–48
- mean of sample, *see* sample mean
- measurement error bias, 363
- median of sample, 9
- mid-range, 9
- missing data, 305
 - codes in statistical packages, 305
- misspecification tests, 361
- mode of sample, 9
- model sum of squares, 106, 207
- moving average, 24
- multicollinearity, 210, 356–357
- Nadaraya-Watson estimator, 115
- natural logarithm, 23, 179–196
 - and elasticities, 181–187, 342–344
 - and semi-elasticities, 181–187, 342–344
 - approximating, 187
 - approximating percentage change, 180
 - approximating proportionate change, 180
 - Cobb-Douglas example, 280–284
 - defined, 179
 - linearizing exponential growth, 188
 - regression models, 181–187, 342–348
 - rule of 72, 187
- Newey, W. K., 469, 502
- Newhouse, J. P., 292, 502
- Nollenberger, N., 313, 502
- nominal data, 26
- nonlinear least squares estimator, 401
- nonlinear regression models, 397–402
- nonparametric regression, 115–117
 - using a statistical package, 116
- nonrepresentative sample, 48, 258, 272
- nonstationary time series, 406–408
- normal distribution
 - fraction within k standard deviations, 12
 - in large samples, 43, 61, 136, 147, 220, 458
 - needed for exact t distribution, 62, 147, 220
 - standard normal distribution, 455
 - table of critical values, 493
- numerical data, 2
- observational data, 2
- omitted variables, 360
- ordinary least squares regression, *see* regression
- outlying observation, 11, 12, 111, 367–371
- p-value, 71, 155, 224, 230
 - using a statistical package, 451
- panel data, 3, 380–387
 - dynamic models, 383
 - example, 384–387
 - fixed effects, 382, 386, 394
 - long panel, 380
 - pooled OLS, 385
 - random effects, 381, 386
 - short panel, 380
 - special considerations, 380–384
 - using a statistical package, 383
 - within variation, 381
- panel-robust standard errors, 381
- parameters
 - defined, 47, 454
 - of linear regression model, 126, 218

- partial regression plot, 369
- per capita data, 27
- percentage change, 28
- percentage point change, 28
- percentiles, 9
- Phillips, A. W., 284, 502
- pie chart, 22
- Pischke, J.-S., 416, 501
- plots, *see* charts
 - diagnostic, 368
- Poisson regression, 399
 - marginal effects, 400
- pooled data, 3
- population mean, 47, 243
- population model for regression, 126, 218, 464
- power of test, 263
- predicted value, 101, 201
- prediction, 110, 252–258
 - average versus actual, 252, 257
 - bivariate regression, 256
 - bivariate regression example, 257
 - for log-linear model, 345
 - for log-log model, 344
 - multiple regression example, 255
 - of actual value, 254
 - of average outcome, 253
 - of conditional mean, 253
 - using a statistical package, 255
 - using big data, 264
- probability density function, 455
- probability mass function, 453
- probit regression, 397–399
 - marginal effects, 398
- propensity score, 396
- proportions data, 81
 - using a statistical package, 81
- quadratic model, 337–340
- quartiles, 9
- quasi-maximum likelihood estimator, 473
- R, *see also* statistical packages
 - essentials, 436
 - robust standard errors, 251
- R-squared, 105–108, 207–208
 - adjusted R-squared, 208
 - alternative definition, 107, 207
 - and correlation, 107, 207
 - definition, 105, 207
 - interpretation, 108
 - with transformed dependent variable, 345
- random effects estimator, 379, 381, 386
 - using a statistical package, 383
- random number generator, 50
- random sample
 - definition, 38
 - generation in a statistical package, 50–52
 - nonrepresentative, 48, 258
 - simple random sample, 41
- random variables, 35–39, 453–462
 - conditional mean, 126, 461
 - conditional probability, 460
 - conditional variance, 461
 - continuous, 455
 - correlation, 462
 - covariance, 462
 - cumulative probability distribution function, 454
 - discrete, 453
 - expected value, 454
 - joint probability, 459
 - linear transformation of, 455
 - mean, 36, 454
 - probability density function, 455
 - probability mass function, 453
 - standard deviation, 37, 454
 - standardized, 455
 - statistical independence, 457, 460
 - sums of, 457
 - variance, 37, 454
- random walk process, 406
- range of sample, 11
- real data, 26
- regression
 - a brief history, 266

- adjusted R-squared, 208
- and correlation, 111
- and difference in means, 104, 318
- and sample mean, 104, 243, 248
- assumptions, 126, 133, 146, 218, 358, 464
- best linear unbiased estimator, 139, 221
- causal methods, *see* causal methods
- Census example, 131–132
- commonly-used models, 400
- computation by hand, 113
- computer output, 108–110, 209–210
- conditional mean, 125, 126, 218
- conditional mean of the error, 127
- confidence interval, 148–150, 222–223
- consistent estimator, 139, 221
- cross-section data, 377–402
- degrees of freedom, 105, 206, 227
- derivations of estimator properties, 463–470
- diagnostics, *see* diagnostics
- estimated coefficients, 102, 133, 203, 472
- experiment example, 128–131
- F tests, 229
- hypothesis tests, 150, 158, 223–234
- in first differences, 407, 413
- inadequate variation in regressor, 210
- incorrect model, 359
- indicator variables, 315–329
- inestimable models, 210, 356
- influential observation, 368, 370
- information criteria, 208
- instrumental variables, 301–303, 470
- intercept-only, 104, 248
- interpretation of slope coefficient, 103, 204, 206
- least squares, 102, 202
- linear-log model, 183
- log-linear model, 183, 342
- log-log model, 183, 343
- marginal effects, 333–337
- matrix algebra, 471–474
- mean of slope coefficient, 134, 219, 464
- minimum variance estimator, 138, 260
- model fit, 105–108, 206, 209
- multicollinearity, 210, 356–357
- nonparametric, 115–117
- nonrepresentative sample, 258
- normal distribution in large samples, 136, 147, 220
- not identified, 211
- ordinary least squares, 102, 202
- outlying observation, 11, 111, 367
- panel data, 380–387
- parameters, 126, 218
- partial effect, 204
- perfectly collinear regressors, 210, 323, 356
- population line, 126, 218
- population model, 126, 218
- prediction, *see* prediction
- R-squared, 105–108, 207–208
- regression line, 101, 201
- regressor orthogonal to residual, 202
- relaxing assumptions, 137, 222
- residual, 101, 202
- residual sum of squares, 102, 202
- robust standard errors, *see* robust standard errors
- spurious regression, 407
- standard error of slope coefficient, 135, 220
- standard error of the regression, 105, 135, 206, 219
- standardized coefficients, 347
- t statistic, 146–147, 220
- tests of statistical significance, 150, 224
- time-series data, 402–415
- total effect, 204
- transformed variables, 333–353
- unbiased estimator, 138, 221, 465
- variance of slope coefficient, 134, 219, 465
- weighted least squares, 259
- regression adjustment, 393
- regression analysis, 4
- regression discontinuity design, 298–300, 394
- regression sum of squares, 106, 207
- regressor, 101, 201

- correlated with error term, 362
- repeated cross-section data, 3
- RESET test, 361
- residual, 101, 202
 - compared to error term, 127, 371
 - standard error of the residual, 105, 206
- residual sum of squares, 102, 107, 202, 207
- residual versus regressor plot, 369
- right-skewed data, 13
- Robinson, J. A., 301, 365, 501
- robust standard errors, 158, 160, 241–252, 466–469
 - bootstrap, 249
 - cluster-robust, 245–246, 364, 378, 380, 467
 - for sample mean, 243, 248
 - HAC-robust for time series, 247–249, 364, 402, 468
 - heteroskedastic-robust, 160, 243–245, 363, 466
 - panel-robust, 381
 - using a statistical package, 250–251
 - when to use, 252
- Rodríguez-Planas, N., 313, 502
- rolling average, 24
- root mean squared error of the residual, 105, 135, 206
- rule of 72, 187
- sample correlation coefficient, *see* correlation
- sample covariance, 98
- sample mean, 8
 - and intercept-only regression, 104, 243, 248
 - as outcome of random variable, 39
 - assumptions, 41
 - Census example, 45–46
 - coin toss example, 40
 - mean, 41, 458
 - normal distribution in large samples, 43
 - robust standard error, 243, 248
 - standard deviation, 42, 458
 - statistical properties, 41–45, 457
 - variance, 42, 458
- sample selection bias, 363
- sample standard deviation, 10, 39
- sample variance, 10, 39
- scatterplot, 96, 199
- seasonal adjustment, 25
- seed for random number generator, 50
- semi-elasticities, 181–187, 342–344, 400
- sets of indicator variables, *see* indicator variables
- Sevilla, A., 313, 502
- simple random sample
 - computer generation, 50–52
 - definition, 41
- simultaneous equations bias, 362
- skewness, 13
- skewness statistic, 13
- slope coefficient, 101, 102, 201
 - consistent estimator, 139, 221
 - derivation of mean, 358, 464
 - derivation of variance, 359, 465
 - economic significance, 151, 225
 - interpretation, 103, 204, 206
 - mean, 134, 219
 - minimum variance estimator, 138
 - standard error, 135
 - statistical significance, 150–153, 224
 - unbiased estimator, 138, 221
 - variance, 134, 219
- smoothed histogram, 17
- spatial map, 21
- spurious regression, 407, 415
- standard deviation of a random variable
 - definition, 37, 454
 - estimator for, 44
- standard deviation of sample
 - definition, 10
 - interpretation, 11
- standard error
 - default, 158
 - of slope coefficient, 135, 220
 - of the regression, 105, 135, 206
 - of the residual, 105, 206
 - of the sample mean, 44
 - robust, 160, 241–252, 466–469
- standard normal distribution, 451, 455

- standardized random variable, 455
- standardized regression coefficients, 347
- standardized score, 23
- Stata, *see also* statistical packages
 - essentials, 434–436
 - robust standard errors, 251
- stationary process, 403
- statistical independence, 457, 460
- statistical learning, 264
- statistical packages, 429–452
 - autocorrelations, 404
 - computer precision, 433
 - computing marginal effects, 336
 - critical values, 451
 - difference in means tests, 318
 - Eviews, 445–449
 - Excel, 449–451
 - fixed effects estimator, 383
 - generating a random sample, 50–52
 - Gretl, 440–445
 - grouped data, 383
 - histogram, 18
 - hypothesis tests, 231
 - indicator variables creation, 327
 - instrumental variables estimation, 391
 - intercept-only regression, 104
 - kernel density estimate, 18
 - local linear regression, 116
 - loess regression, 116
 - missing data codes, 305
 - multicollinearity, 210
 - nonlinear model estimators, 402
 - nonparametric regression, 116
 - p-values, 451
 - prediction, 255
 - proportions data, 81
 - R, 436–440
 - random effects estimator, 383
 - regression output, 108–110, 209–210
 - robust standard errors, 250–251
 - standardized regression coefficients, 347
 - Stata, 434–436
 - time series data, 408
 - statistical significance, 150–153, 224
 - based on correlation, 152
 - of subsets of regressors, 230
 - overall significance, 230
 - versus economic significance, 151
 - Stock, J. H., 416, 502
 - Student’s t distribution, *see* t distribution
 - summary statistics, 7–15
 - survey data, 49, 272
 - symmetric distribution, 13
- t distribution, 62–65, 147, 451
 - critical value, 64
 - exact if data are normally distributed, 62
 - exact if errors are normally distributed, 147
 - inverse probabilities, 64
 - mean and variance, 63
 - probabilities for, 63
 - table of critical values, 493
- t statistic
 - for slope parameter, 146–147, 220, 242
 - for the mean, 62
- tables of critical values, 493–500
 - chi-squared distribution, 500
 - F distribution, 493
 - normal distribution, 493
 - t distribution, 493
- tabulation, 22, 95
- Tanaka, S., 295, 502
- tests, *see* hypothesis tests
- Tibshirani, R., 416, 502
- time-series data, 3, 402–415
 - autocorrelation, 100, 403
 - autoregressive error model, 405
 - autoregressive model, 404
 - differenced, 402
 - finite distributed lag model, 405
 - forecasting, 408
 - impulse response function, 415
 - lagged, 402
 - nonstationary process, 406–408
 - random walk process, 406

- regression, 402–415
 - spurious regression, 407
 - stationary process, 403
 - transformations for, 24–28, 403
 - using a statistical package, 408
- Titunik, R., 298, 501
- total sum of squares, 105, 207
- transformed variables in regression, 353, 333 –
–353
- treatment-control comparison, 295
- Trivedi, P. K., 416, 501
- two-stage least squares estimator, 389
- two-way scatterplot, 96, 199

- unbiased estimator
 - definition, 47
- unit root process, 406, 412
- upper quartile, 10

- variance of a random variable
 - conditional variance, 461
 - definition, 37, 454
 - estimator for, 43
- variance of sample, 10

- Watson, M. K., 416, 502
- weighted least squares, 259, 272
- weighted mean, 49
- West, K. D., 469, 502
- white noise process, 403
- White, H. A., 467, 502
- Witten, D., 416, 502
- Wooldridge, J. M., 416, 502

- Yared, P., 365, 501

- z-score, 24, 111, 347