

Index

Applications

- ANCOVA, 158
- approval, 456
- binomial regression, 381, 392
- cattle feed, 458
- coins, 347
- crime, 153
- differential invalidation, 392
- geography, 243
- GLM, 319, 322
- heteroskedasticity, 246
- imputation, 463
- lifetime, 291
- logistic regression, 381, 392
- median regression, 267
- nominal regression, 447
- occupation, 447
- ordinal regression, 456, 458, 463
- physics, 246
- Poisson regression, 426
- property crime, 267
- South Sudan, 209
- terrorism, 426
- time series, 241
- voting, 166, 195, 209, 250, 319
- wealth, 200, 322

R data

- cattleData, 458
- coinflips, 348
- cows, 166, 167, 195, 319
- crime, 153, 266, 418
- fakepoisson, 410

- gdpcap, 322
- gdp, 200
- gssocc, 447, 456
- ocanada, 381
- rur2013parl, 251
- sri2010pres, 392
- summary.aov, 175
- suvr, 463
- terrorism, 426
- xsd2011referendum, 209

R functions

- (drop), 211
- AIC, 364
- BIC, 365
- I, 430
- ROC, 357
- &&, 133, 134
- accuracy, 352, 355
- anova, 388
- aov, 160, 388
- attach, 153
 - no, 210
- autocor.test, 241
- barplot, 587
- binom.test, 118
- bptest, 131, 132, 154
- cbind, 383
- col, 242
- confint, 133, 134, 157, 165, 247, 248
- cor.test, 168, 458
- c, 140, 247, 248

`data.frame`, **178**, 197, 201, 323, 418, 460, 464
`exp`, 410
`fligner.test`, 161
`for-loop`, 117, 119, 133, 134, 154, 587
`function`, 587
`glm.nb`, 424, 428
`glm.nb,offset`, 428
`glm`, 319, **348**, 383, 418, 428
`glm,family`, 320, 322
`glm,link`, 320, 322
`glm,offset`, 428
`glm,quasipoisson`, 422
`head`, 112
`hetero.test`, 160
`hist`, 110, 121
`ks.test`, 118
`legend`, 461
`length`, 122
`library`, 125, 266
`lines`, 213
`lm`, 116, 117, 121, 131, 136, 140, 153, 162, 201, 211, 231
`lm,weights`, 248, 252
`logistic`, 195, 197
`logit.inv`, 195
`logit`, 195
`make.link`, 320, 322, 323, 344
`matrix`, 355
`mean`, 136, 184
`multinom`, 450, 458
`ordered`, **463**
`overlay`, 110, 153, 159
`plot`, 122, 131, 154
`points`, 157
`polr`, 456, 463
`predict`, 136, 157, 165, 178, 197, 201, 212, 323, 349, 418, 450, 457, 460, 464
`predict,se.fit`, **212**
`qchisq`, 420
`qqline`, 108
`qqnorm`, 108
`quantile`, 203
`rcauchy`, 120
`rchisq`, 120
`read.csv`, **153**, 167, 266, 458
`residuals`, 122, 153, 159, 241, 388
`rexp`, 109, 119
`rnorm`, 108, 116, 117, 121, 122, 125, 131, 133, 134, 587
`row`, 242
`rq`, 265, 266, **266**, 267, 269, 270
`rt`, 114
`runif`, 125
`runs.test`, 125, 154, 159, 388, 590
`sample`, 154, 355
`seq`, 122, 131, 133, 134, 140
`set.base`, 162, 452
`set.seed`, 112, 117, 122, 125, 136, 140, 410
`shapiroTest`, 111, 113, 115, 153, 159
`sort`, 410
`source`, **107**, 125, 153, 167
`subset`, 418, 424
`summary.aov`, 158, 161
`summaryHCE`, 217

summary, 116, 117, 121, 153, 160, 319
 tail, 113
 vif, 141, 176
 which, 211
 R packages
 Epi, 357
 MASS, 424, 456, 463
 RFS, 111, 125, 195, 217, 352, 452, 590, **603**
 car, 141, 176
 lawstat, 125, 153
 lmtest, 131, 153
 nnet, 450, 458
 quantreg, 265
 accuracy, 351, 355, 452, 457
 AUC, 356
 maximum, 353
 rate, 351
 relative, 352
 additive model, 140, 159, 176
 AIC, 320, 321, **364**, 387, 412, 452, 464
 alpha-testing, 116, 119, 120
 ANCOVA, 158
 ANOVA, 64, 160, 574
 effects model, 69, 70
 means model, 68
 table, 158
 toy example, 28
 ANOVA table, 173
 appropriate test, **117**
 AR(1), 242
 AUC, 356
 autocorrelation, 241, 242
 back-transform, 197, 206, 349
 base category, 162, 452
 bias, 34, 51, 81, 125, 127, 299, 544
 bias-variance trade-off, 426, **433**
 BIC, **365**, 387, 412
 bijection, 295
 Binomial test, 118
 boolean, 133
 bounds, **192**
 one, 200
 two, 192, **193**
 others, **206**
 Breusch-Pagan test, **129**, 154, 161
 canonical link, 312, 314, 317, 342, 416
 CDF, 344, 554
 center of gravity, 31
 centering, 53
 Central Limit Theorem, *see* CLT
 classical linear model, *see* CLM
 CLM, 152, 190, 232, 240, 294, 308, 309
 CLT, 115, 120, 561, **591**, 600
 coefficient of variation, 167
 conditional distribution, **311**, 341, 379, 382, 416
 confidence band, 100, 212, 355
 confidence interval, 86, 94, 95, 99, 135, 157, 165, 266, 344
 confusion matrix, 355
 correlation, 168, 169, 543, 555
 population, 555
 sample, 543
 symmetric, 543
 covariance, 52, 82, 83, 503, 504, 541, 555
 population, 555

- sample, 541
 - symmetric, 542
- covariance matrix, 309
- coverage, **133**, 136
- cumulative distribution
 - Cauchy, 573
- cumulative distribution function, *see*
 - CDF
- data matrix, *see* design matrix
- degrees of freedom, 496
- design matrix, 54, 107
- deviance, 538
- differential invalidation, **209**, 250, 392
- directional hypotheses, 166
- electoral forensics, 209, 250, 392
- estimator, **48**
- exponential class, **313**, 379
- exposure, *see* offset
- extrapolation, 27
- factor analysis, 143
- Fligner-Killeen test, 161
- function
 - absolute value, 262
 - beta, 575
 - exponential, 202
 - gamma, 569, 572, 575
 - identity, 422
 - logarithm, 200, 413
 - logistic, **194**, 197, 337
 - logit, **193**, **314**, 337
- functional form, 122, 128, 154, 159, 315
- Gaussian process, *see* normality
- generalized least squares, *see* GLS
- generalized linear model, *see* GLM
- geographic regression, 243
- GLM, **309**, 310
 - vs* CLM, 309
 - assumptions, 315
 - components, 310, 332, 339, 341, 348, 413
 - distribution, **311**
 - exponential class, **313**
 - linear predictor, **310**
 - link function, **312**, 318
- GLS, **240**
- grammar of formulas, **171**
- graphics, 155, 163, **179**, 182, **460**, 465
 - R philosophy, 212
- hat matrix, **57**, **237**
- heteroskedasticity, 133, 217, 309, 335, 409
 - adjustment, 216
 - bulge, **134**
 - effects, 125, 133, 135
 - funnel, **134**
 - modeling, 230
 - none, *see* homoskedasticity
 - trumpet, **133**, 136, 414
- heteroskedasticity test, 160, 161
- Hildebrand, **557**
- Hildebrand Rule, 155, 447, 580
- homoskedasticity, 78, **129**, 154, 160, 229, 340
- Huber-White, 216
- iid, 78, 80, 229, 566
- independent, 80, 230, 295
- independent and identically distributed, *see* iid

influential point, 267, 419
 inner product, 483
 interaction model, 158, 171, 173
 interaction term, 431
 interpolation, 27, 155
 interpretation, 135, 182, 209, 254, 320, 452, 461
 of logarithm, 202
 of logit, 196

 Kingdom of Ruritania, *see* Ruritania
 Kolmogorov-Smirnov test, 118, 119–121
 kurtosis, 560, 563, 570, 592

 latent variable, 336
 likelihood, 285, 287, 291
 log-likelihood, 289, 291
 likelihood ratio test, 365, 387
 line of “best” fit, 18, 282
 linear, 78
 linear predictor, 310, 337, 341, 350, 379, 381
 link function, 312, 320, 341, 359, 380, 382, 416, 446
 canonical, 312, 342, 379, 416
 cauchit, 344
 complementary log-log, 344, 359
 identity, 317, 422
 log-log, 344, 361
 logarithm, 415
 logit, 314, 337, 341, 344, 347, 446, 450, 457
 probit, 344
 logistic regression, 332

 machine epsilon, 138
 matrix, 476

 addition, 479
 additive identity, 479
 additive inverse, 479
 adjacency, 244
 associative, 480, 481, 492
 commensurate, 479, 482
 commutative, 480, 481, 489
 confusion, 355
 covariance, 229, 240, 309, 501, 504
 determinant, 484
 diagonalization, 494
 dimension, 476
 distributive, 481
 eigenvalue, 495
 eigenvector, 495
 hat, 237
 idempotent, 57, 496
 inner product, 483
 inverse, 67, 73, 484
 invertible, 484
 multiplication, 482
 multiplicative identity, 484
 multiplicative inverse, 484, 499
 orthogonal, 57, 62, 497
 positive definite, 240, 497, 500
 projection, 58, 497
 rank, 67, 484, 496, 501
 representation, 477
 sample covariance, 504
 sample mean, 502
 sample variance, 503
 singular, 138, 484
 square, 476
 symmetric, 57, 244
 symmetrize, 493

trace, 493
 transpose, 73, 493, 499
 maximum likelihood estimation, *see* MLE
 mean, 313, 413, 502, 537
 distribution of, 576
 linear, 537
 population, 545
 sample, **537**
 mean squared error, *see* MSE
 median regression, 18, **266**, 267
 MGF, 592–597
 minimum-width interval, 95
 vs central, 95, 96, **97**
 MLE, 18, 285, 287, 288, 294, 312, 446
 bias, 292
 function of, 292
 of β_0 , 294
 of β_1 , 297
 of σ^2 , 299
 uniqueness, 292
 model selection, 171, 174, 195, **362**, 429
 AIC, **364**
 BIC, **365**
 likelihood ratio test, **365**
 model specification, 128
 model stability, 446
 moment, 592
 central, 592
 raw, 592
 standardized, 592
 moment generating function, *see* MGF
 Monte Carlo, 181, 197, 203, 355
 MQLE, 420, **422**, 429
 MSE, 36, 84, 95, 238
 multicollinearity, 49, 64, **138**
 approximate, 138
 CS, 138
 fix, **142**
 indications, 140
 logic, 139
 super, 138, 139
 test, 141
 multiple comparisons, 581
 Bonferroni adjustment, **582**
 nominal regression, 449
 norm, 265
 L_1 , 265
 L_2 , 265
 Normality, 108
 normality, 78, 81, 153, 159
 nuisance parameter, **314**
 null model, 38, 452, 464
 Occam's Razor, 158
 odds ratio, 196, 199
 offset variable, **427**
 OLS, 19, 190, 229
 assumptions, 34, **50**, 77, 152, 176, 190, 229
 b_0 , 22, 82, 127
 b_1 , 22, 24, 32, 33, 35, 82, 125
 derivation, matrix, 46–48
 derivation, scalar, 20–22
 matrix, 229
 matrix model, **46**
 scalar, 229
 scalar model, **20**
 toy example, 25, 53
 OLS assumptions, 122

ordinary least squares, *see* OLS
 overdispersion, 384, 415, **420**, 428
 adjustment, 421
 causes, 420
 effects, 420
 test, 420

 p-value, 117
 parameter, 536, **545**, 556
 point estimate, 183
 Poisson regression, 410
 PRE, 386
 definition, **38**, **63**, 352
 R^2 , 38, 141, 170
 \bar{R}^2 , 39
 pseudo- R^2 , 40, 63, 319, 351
 preamble, 167
 precision, 88
 prediction interval, 89, 99, **99**, 157, 165
 principal component analysis, 143
 probability distribution
 Bernoulli, 339, **556**, 594
 Binomial, **561**, 596
 binomial, 283, 284, 377
 bivariate normal, 301
 Cauchy, 120, 573
 Chi-square (χ^2), 536, 569, 578
 chi-square (χ^2), 84, 91, 95, 120
 degenerate, 598
 exponential, 119, 290
 F, 574
 gamma, 423
 Gaussian, 316, 566
 multinomial, 442
 multivariate, **300**
 multivariate normal, 301
 Negative Binomial, 413, 423
 Normal, 566, 576, 597
 Poisson, 286, 288, **413**, 423, **563**
 Snedecor's F, *see* F
 Student's t, *see* t
 t, 91, 94, 95, 114, 571
 uniform, 116, 583
 proportional reduction in error, *see* PRE

 Q-Q plot, 108
 QMLE, 312
 quantile function, 344
 quantile regression, 18, 269
 quasi maximum likelihood estimation,
 see MQLE
 quasi-likelihood, 385, 420

 random variable, 181, **534**
 regression table, 154, 162, 179
 rejection rate, 121, 135
 representative sample, 447
 residuals, 122, 240
 residuals plot, 122, 124, 160
 ROC curve, 355
 Rule of Thumb, 142, 170, 453, 580
 runs test, 123, 159, **586**
 Ruritania, 2
 currency, 3
 drugs, 5
 economics, 5
 geography, 3
 government, 3
 kraj, 4, 243
 Rudolph II, 3
 stát, 4

Státní Univerzita v Ruritánii, 463
 US embassy, 6
 US relations, 6
Valné Shromáždění, 287
Vlajka (1954), 3

sample space, **556**
 sensitivity, 355
 Shapiro-Wilk test, 113, 153, 159
 skew, 109, **557**, 563, 570, 580, 592
 specificity, 355
 standard form, 313, 316, 417
 statistics experiment, 108, 116, 119,
 120, 133, 135, 140, 203, 204,
 355, 410
 S_{xx} , 24, 82, 88
 systematic error, 122

test statistic, **91**, 94, 135
 testing systematic error, 122
 threshold, 337, 353
 time series, 242
 Type I error, **116**, 121, 135, 357, 581
 Type II error, 116, 132, 135, 357

variability, 167
 variable type
 binary, 332
 count, 409
 dichotomous, 332
 nominal, **442**
 ordinal, **454**
 variance, 314, 503, 504, **535**, 538, 552
 distribution of, 578
 population, 552
 quadratic, 539
 sample, 538
 variance inflation factor, *see* VIF
 VGLM, 398
 VIF, 141, 170, 176
 voting, 236

weighted least squares, *see* WLS
 WLS, **230**
 estimator, 235
 matrix, 230
 Working-Hotelling, 100
 bands, 213