

Index

- actuarial & financial terms and concepts
 - a posteriori premium, 464
 - a priori premium, 464
 - adverse selection, 12, 111, 199
 - bankruptcy, 383
 - bonus-malus, 464
 - capital asset pricing model, CAPM, 46
 - capitation, 326
 - case estimates, 469
 - closed claim, 4, 17, 121, 346, 468
 - conditional tail expectation, CTE, 28
 - credibility, 301, 453
 - credit scoring, 308, 394
 - demand, 11, 70, 316, 327, 417
 - elasticity, 78, 312, 443
 - experience rating, 452
 - financial leverage, 158
 - general insurers, 348
 - health provider
 - fee for service, FFS, 128, 326, 517
 - health maintenance organization, HMO, 128, 326, 517
 - point of service, 326
 - preferred provider organization, 326
 - incremental, cumulative payments, 469
 - incurred but not reported, 468
 - initial public offering, IPO, 59
 - inpatient admissions, 315
 - insurance company branch office, 97
 - malpractice insurance, 346
 - manual rate, 454
 - merit rating, 452
 - outpatient events, 315
 - pricing, 12
 - ratemaking, 348, 366
 - additive plan, 366
 - multiplicative plan, 366
 - redlining, 192
 - reserve, 12, 467
 - risk classification, 452
 - run-off, 472
 - securities market line, 47
 - solvency testing, 12
 - stock liquidity, 157
 - valuation date, 468
 - value-at-risk, VaR, 28
 - workers compensation, 461
- analysis of variance, ANOVA, table, 34, 83
- best linear unbiased predictors, *BLUP*, 402
- bootstrap, 410
 - loss reserves, 412
 - parametric, 411
 - replication, 411
 - resampling pairs, 411
 - resampling residuals, 411
 - sample, 411
- categorical variable, 107
 - binary, 56, 305
 - dummy, 108
 - factor, 108
 - multicategory, 318
 - nominal, 318
 - ordinal, 325
 - polychotomous, 318
 - polytomous, 318
 - reference level, 111
 - unordered, 108
- censoring, 200, 383, 418, 469
 - fixed, 385
 - interval, 385
 - left-, 385
 - random, 385
 - right-, 385
- chain ladder, 412, 469
- claims triangle, 468
- collinearity, 165, 169
 - tolerance, 167
 - variance inflation factor, 166
- confidence interval, 39, 88

- correlation coefficients
 - autocorrelations, 252
 - multiple, 83
 - ordinary, 25
 - partial, 91, 187
 - partial autocorrelation, 265
 - Pearson, 25
- datasets
 - Anscombe's data, 58
 - auto industry claim triangle, 479
 - automobile injury insurance claims, 17, 331
 - automobile insurance claims, 16, 121, 124, 135
 - automobile UK collision, 366
 - capital asset pricing model, 47
 - CEO compensation, 504
 - Euro exchange rates, 249
 - Galton heights, 2
 - general liability reinsurance, 478, 479
 - Hong Kong exchange rates, 231, 265, 268
 - Hong Kong horse racing, 332
 - hospital costs, 16, 133
 - initial public offerings, 59
 - insurance company expenses, 17, 103, 136, 204
 - insurance redlining, 219
 - labor force participation rates, 240, 242, 247, 285
 - Massachusetts bodily injury claims, 4
 - medical care payment triangle, 479
 - medical price inflation, 273
 - Medicare hospital costs, 291
 - MEPS health expenditures, 14, 315, 360, 370, 421
 - national life expectancies, 18, 61, 105, 181
 - nursing home utilization, 15, 56, 59, 134, 407, 438, 444, 447
 - outliers and high leverage points, 43, 164
 - prescription drugs, 280, 282
 - refrigerator prices, 89
 - risk managers cost effectiveness, 210, 511
 - Singapore automobile data, 343
 - Singapore property damage, 469
 - Singapore third party injury, 471
 - Standard and Poor's daily returns, 270, 286, 287
 - Standard and Poor's quarterly index, 244
 - stock market liquidity, 157, 167
 - Swedish automobile claims, 491
 - term life insurance, 70, 93, 108, 117, 181
 - TIPS – inflation bond returns, 251, 255, 258
 - Wisconsin hospital costs, 128, 517
 - Wisconsin lottery sales, 23, 102, 136
 - workers compensation, 461
- density estimation, 406
 - Epanechnikov, 406
 - kernel, 372, 406, 444
 - Silverman's procedure, 407
- dependent variable, 10
 - endogenous, 10
 - explained, 10
 - left-hand side, 10
 - outcome of interest, 10
 - regressand, 10
 - response, 10
- diagnostic checking
 - data criticism, 148, 169
 - model criticism, 148, 169
 - residual analysis, 41, 153, 445
- dispersion
 - equidispersion, 352
 - overdispersion, 352, 368
 - underdispersion, 352
- distributions
 - linear exponential family, 364
 - chi-square, 8, 177, 268, 314, 344, 548, 556
 - exponential, 388
 - extreme value, 310, 391
 - F -, 115, 164, 558
 - gamma, 437
 - generalized beta of the second kind
 - distribution, $GB2$, 442
 - generalized extreme value, GEV , 448
 - generalized gamma, 442
 - generalized Pareto distribution, 449
 - inverse gaussian, 438
 - location-scale, 391
 - log location-scale, 391
 - log-normal, 391
 - logistic, 309, 328, 391
 - multivariate normal, 69
 - negative binomial, 353
 - normal, 3, 437, 554
 - Pareto, 438, 449
 - Poisson, 343, 474
 - posterior, 404
 - prior, 404
 - sampling, 547
 - t -, 38, 86, 548, 557
 - Tweedie, 375
 - Weibull, 391
- estimable function, 145
- estimator, 548
 - consistency, 548
 - interval, *see* confidence interval
 - maximum likelihood, 312, 339, 549

- point, 548
- unbiased, 79, 548
- examples, *see* datasets
 - automobile insurance, 355
 - California automobile accidents, 345
 - choice of health insurance, 323
 - credit scores, 308, 394
 - data snooping in stepwise regression, 152
 - dental expenditures, 206
 - divorce rates, 290
 - health plan choice, 326
 - historical heights, 207
 - job security, 314
 - large medical claims, 448
 - Lee-Carter mortality rate forecasts, 262
 - life insurance company expenses, 96
 - Literary Digest poll, 198
 - long-term stock returns, 232
 - medical malpractice insurance, 346
 - race, redlining and automobile insurance prices, 192, 203
 - Rand health insurance experiment, 11, 111, 358
 - Spanish third party automobile liability insurance, 357
 - success in actuarial studies, 335
 - summarizing simulations, 28
 - suppressor variables, 170
 - time until bankruptcy, 383
 - warranty automobile claims, 396
- explanatory variable, 2
 - binary, 92
 - categorical, 107
 - combining categorical and continuous variables, 126
 - covariate, 126
 - endogenous, 201
 - exogenous, 10
 - factor, 107, 126
 - independent variable, 10
 - interaction, 95, 167
 - omitted, 197, 201
 - predictor, 10
 - quadratic, 26, 105
 - regressor, 10
 - right-hand side, 10
 - suppressor, 151, 169
 - transformed, 94, 179
- exposure, 344
- goodness of fit statistics, 81
 - Akaike's information criterion, *AIC*, 342, 348, 369
 - Bayesian information criterion, *BIC*, 342, 369
- coefficient of determination adjusted for degrees of freedom, R_a^2 , 84
- coefficient of determination, R^2 , 33, 83
- deviance statistic, 370
- log-likelihood, 313
- max-scaled R^2 , 314
- multiple correlation coefficient, R , 83
- Pearson chi-square, 344, 347, 369
- pseudo- R^2 , 314
- scaled deviance statistic, 370
- graphs, *see* plots
- heterogeneity, 292
- heteroscedasticity, 42, 154, 175, 179, 209, 307, 434
- heteroscedasticity-consistent standard error, 307, 420
- homoscedasticity, 30, 42, 176, 236
- hypothesis test, 549
 - F -test, 116, 341
 - t -test, 37, 85
 - alternative hypothesis, 549
 - critical region, 549
 - extra sum of squares principle, 116
 - general linear hypothesis, 113
 - null hypothesis, 549
 - one-sided alternative, 549
 - p -value, 550
 - rejection region, 549
 - test statistics
 - Lagrange multiplier, 340
 - likelihood ratio, 313, 340, 348
 - Rao, 340
 - Wald, 340
 - two-sided alternative, 549
 - Type I error, 549
 - Type II error, 549
- independence of irrelevant alternatives, 322
- inverse Mills ratio, 419
- least squares
 - generalized, 401
 - intercept, 28
 - iterated reweighted, 367, 381
 - least absolute deviations, 446
 - method, 27, 70
 - principle, 548
 - recursive calculation, 141
 - regression plane, 74
 - slope, 28
 - weighted, 178, 400
- leverage, 42, 151, 154, 161, 169, 183, 185
 - Cook's distance, 163

- likelihood inference
 - accelerated failure time model, 391
 - censoring, 387, 420
 - Fisher scoring algorithm, 340
 - generalized estimating equations, 380
 - hypothesis test, *see* hypothesis test
 - information matrix, 313, 330, 331, 338, 346, 368, 380
 - Kullback-Leibler distance, 341
 - likelihood, 312
 - log-likelihood, 312, 367
 - maximum likelihood estimation, 549
 - maximum likelihood estimator, 312, 339, 367, 380
 - model-based standard errors, 381
 - Newton-Raphson algorithm, 340, 381
 - proportional hazards model, 393
 - quasi-likelihood estimator, 352
 - robust standard error, 352, 381
 - score equations, 313
 - score function, 338, 346, 367
- link function, 307, 345, 362
 - canonical, 365
 - logarithmic, 345
- log odds, 311
- logit function, 310
- matrix algebra
 - eigenvalue, 179, 553
 - generalized inverse, 145, 553
 - Hessian, 381, 553
 - matrix inverse, 66, 552
 - orthogonal matrices, 170, 553
 - variance-covariance matrix, 69, 346
- maximum likelihood estimation, 548
- model assumptions, 9
 - error representation, 30, 78
 - observables representation, 29, 78
- model validation, 172
 - cross-validation, 174
 - data snooping, 172
 - leave-one-out cross-validation, 174
 - model development subsample, 172
 - out-of-sample, 172
 - predicted residual sum of squares, *PRESS*, 175
 - sum of squared prediction errors, *SSPE*, 173
 - testing subsample, 172
 - training subsample, 172
 - validation subsample, 172
- multicollinearity, *see* collinearity
- normal equations, 28, 76, 144, 367
- odds, 310
- odds ratio, 311
- offset, 345
- omitted variable, 290
- plots
 - R* chart, 242
 - added variable, 88, 157
 - aspect ratio, 513
 - box, 5
 - box and whiskers, 5
 - chartjunk, 506, 514
 - components
 - legend, 485
 - plotting symbols, 485
 - scale lines and labels, 485
 - tick marks, 485
 - title and caption, 485
 - control chart, 241, 258
 - data density, 519
 - half scatterplot matrix, 72
 - histogram, 5
 - letter, 93
 - multiple time series, 292
 - partial regression, 88
 - quantile-quantile, *qq*, 6
 - scatter, 25
 - scatter plot with symbols, 292
 - scatterplot matrix, 72
 - scatterplot smoothing, 406
 - lowess, 217, 409
 - Nadaraya-Watson estimator, 409
 - time series, 229
 - trellis, 292
 - Xbar chart, 241
- principal components, 171, 263
- product-limit estimator, 389
- recurrent events, 395
- regression function, 76
- regression model
 - accelerated failure time, *AFT*, 390
 - analysis of covariance, 127
 - basic linear, 29
 - Bayesian, 403
 - broken stick, 100, 409
 - conditional logit, 321
 - count
 - heterogeneity, 356
 - hurdle, 355
 - latent class, 357
 - negative binomial GLM, 476
 - overdispersed Poisson, 474
 - Poisson, 343
 - zero-inflated, 354

- cumulative logit, 326
- cumulative probit, 327
- fat-tailed, 433
- frequency-severity, 417
- general linear model, 131, 144
- generalized additive, *GAM*, 409
- generalized linear model, *GLM*, 364, 437
- generalized logit, 319
- linear probability, 306
- logistic, 307
- mixed linear, 399, 461
- mixed logit, 322
- multinomial logit, 321
- nested logit, 325
- normal linear hierarchical model, 403
- one-way random effects, 399, 459
- overdispersed Poisson, 412
- piecewise linear, 100
- probit, 307
- proportional hazards, *PH*, 392
- quantile regression, 445
- semiparametric, 410
- simple linear, 29
- tobit, 418
- Tweedie *GLM*, 375
- two factor additive, 130
- two factor interaction, 131
- two-part, 417
- residual, 33, 153
 - analysis, *see* diagnostic checking
 - Anscombe, 374
 - Cox-Snell, 374
 - deviance, 374
 - outlier, 42, 151, 155
 - Pearson, 347, 374
 - standardized, 154
 - studentized, 154
- response function, 98
- sampling, 198
 - frame, 198
 - Heckman procedure, 208
 - ignorable case, 204, 206
 - impute, 204
 - limited sampling regions, 200
 - missing at random, 204
 - missing at random, *MAR*, 206
 - missing completely at random, *MCAR*, 205
 - missing data, 203
 - non-ignorable case, 207
- shrinkage, 459
- significance
 - causal effect, 192
 - statistical, 190
 - substantive, 190
- statistic, 8, 547
- stochastic process, 397
- symbols
 - D_i , Cook's distance, 164
 - E_i , exposure, 344
 - Error MS*, error mean square, 83
 - Error SS*, error sum of squares, 33, 53, 83
 - H_0 , null hypothesis, 38
 - H_a , alternative hypothesis, 38
 - LRT*, likelihood ratio test statistic, 313
 - MSE*, error mean square, 83
 - PRESS*, predicted residual sum of squares, 175
 - R , multiple correlation coefficient, 83
 - R^2 , coefficient of determination, 33, 53, 83
 - R_a^2 , coefficient of determination adjusted for degrees of freedom, 53, 84
 - Regression SS*, regression sum of squares, 33, 53, 83
 - Regression MS*, regression mean square, 83
 - SSPE*, sum of squared prediction errors, 173
 - Total SS*, total sum of squares, 32, 53, 82
 - U , utility, 310
 - V , underlying value, 310
 - VIF*, variance inflation factor, 166
 - $\Phi(\cdot)$, standard normal distribution function, 18
 - Pr, probability operator, 18
 - \bar{y} , sample mean, 4
 - β_0 , (population) regression intercept, 29
 - β_1 , (population) regression coefficient associated with x_1 , 29
 - β_j , regression coefficient associated with x_j , 76
 - $\boldsymbol{\beta}$, vector of regression coefficients, 79, 113
 - $\boldsymbol{\epsilon}$, vector of disturbance terms, 64
 - χ_k^2 , chi-square random variable with k degrees of freedom, 351, 556
 - η , systematic component, 362
 - \mathbf{X} , matrix of explanatory variables, 66, 74
 - \mathbf{b} , vector of regression coefficients, 76
 - \mathbf{b}_{MLE} , maximum likelihood estimator of $\boldsymbol{\beta}$, 313
 - \mathbf{y} , vector of dependent variables, 64, 74
 - $\pi(\cdot)$, probability function, 307
 - μ , population mean, 9
 - π_i , probability of a 1 for subject i , 306
 - σ , population standard deviation, 9
 - σ^2 , population variance, 9
 - ε_i , "error," or disturbance term, 30
 - \hat{y} , fitted value of y , 28

symbols (*cont.*)

- b_0 , least squares intercept, 28, 53
 - b_0, b_1, \dots, b_k , least squares regression coefficients, 74
 - b_1 , least squares slope, 28, 53
 - c , number of levels in a categorical variable, 108
 - e_i , residual, 33
 - h_{ii} , i th leverage, 162
 - k , number of explanatory variables, 71
 - n , sample size, 4, 53
 - p -value, probability value, 38, 53, 550
 - s , residual standard deviation, 34, 53
 - s^2 , mean square error, 34, 53
 - s_x , sample standard deviation of $\{x_1, \dots, x_n\}$, 25
 - s_y , sample standard deviation, 4, 25
 - $se(b)$, standard error of b , 36, 53
 - $se(pred)$, standard error of a prediction, 40
 - $t(b)$, t -ratio for b , 37, 53
 - $t_{n-2, 1-\alpha/2}$, a $1-\alpha/2$ percentile from the t -distribution with $n - 2$ degrees of freedom, 37
 - $v(\cdot)$, variance function, 363
 - x , observed variable, typically an explanatory variable, 23
 - y , observed variable, typically the outcome of interest, 4
 - y^* , unobserved latent variable, 309
 - B, backshift operator, 261
 - E, expectation operator, 9
 - Var, variance operator, 9
- table, 483
- theorems
- central limit, 9, 18, 37
 - Cramer-Rao, 340
 - Edgeworth approximation, 20
 - Gauss-Markov, 81, 176, 401
 - linearity of normal random variables, 547
- time series models
- autoregressive changing heteroscedasticity model of order p , $ARCH(p)$, 286
 - autoregressive integrated moving average (ARIMA) model, 260, 262
 - autoregressive model of order p , $AR(p)$, 260
 - autoregressive model of order one, $AR(1)$, 254
 - autoregressive moving average (ARMA) model, 262
 - causal, 228
 - fixed seasonal effects, 278
 - generalized $ARCH$ model of order p , $GARCH(p)$, 287
 - linear trend in time, 230
 - longitudinal
 - basic fixed effects, 293
 - basic random effects, 299
 - extended fixed effects, 296
 - extended random effects, 301
 - least squares dummy variable, 295
 - one-way fixed effects, 296
 - random effects, 299
 - two-way fixed effects, 296
 - variable coefficients, 297
 - moving average model of order q , $MA(q)$, 261
 - quadratic trend in time, 231
 - random walk, 237
 - seasonal autoregressive, 282
 - seasonal exponential smoothing, 283
 - white noise, 236
- time series statistics
- augmented Dickey-Fuller, 285
 - Box-Ljung chi-square, 268
 - Box-Pierce chi-square, 268
 - Dickey-Fuller, 285
 - Durbin-Watson, 270
 - exponential smoothed estimate, 276
 - lag k autocorrelation, 253
 - moving average estimate, 273
 - running average estimate, 273
- time series terms and concepts
- ψ -coefficient representation, 264
 - backshift operator B, 261
 - chain rule of forecasting, 259, 263
 - double smoothing, 274
 - filter, 237, 243
 - forecast, 229
 - innovation uncertainty, 237
 - irreducible, 237
 - longitudinal data, 227
 - meandering process, 256
 - seasonal adjustment, 234
 - seasonal component, 233
 - smoothed series, 259
 - stationary, 236
 - strong stationarity, 236
 - weak stationarity, 236
 - stochastic process, 227
 - time series, 227
 - unit root tests, 284
- transformations, 94, 434
- approximate normality, 374
 - Box-Cox family, 7, 435
 - Burbidge-Magee family, 436
 - John-Draper family, 436
 - logarithmic, 7, 77

Cambridge University Press

978-0-521-76011-9 - Regression Modeling with Actuarial and Financial Applications

Edward W. Frees

Index

[More information](#)*Index*

565

- power, 7
- rescaling, 435
- signed-log, 436
- variance-stabilizing, 374
- Yeo-Johnson family, 436
- truncated, 200, 383, 448
 - left-, 386
 - right-, 386
- utility function, 310
- variable selection, 196
 - backwards stepwise regression, 150
 - best regressions, 152
 - forwards stepwise regression, 150
 - stepwise regression, 150
- variance components estimation, 401