

Subject Index

- a prior distribution, 91–93
- a prior probability, 96
- Acceptance region, 111–112
- Accuracy vs. precision, 219
- Agronomy, 38, 74
- Algorithms, 192, 193
- Allocation (of treatment), see Treatment allocation
- Aly's statistic, 69–70
- Analysis of variance, 71, 73–74, 113
- Angiograms, 103
- Animals, 7
- Antibodies, 28
- ARIMA, 189
- Arithmetic vs. geometric mean, 106
- Aspirin, 40, 74
- Association,
 - spurious, 143
 - versus causation, 154, 160
- Assumptions, 4, 36, 57, 68, 99, 111, 147, 181
- Astronomy, 8, 83, 114
- Asymptotic, 57, 165
 - approximation, 68
 - relative efficiency (ARE), 49–50
- Audit, 157
- Authors, 119
- Autocorrelation, 115
- Autoregressive process, see time series
- Axis
 - label, 130, 141
 - range, 130
- Bacteria, 106
- Balanced design, 74
- Bar chart, 126, 135
- Baseline, 65
- Bayes
 - analysis, 98
 - factor, 95
 - Theorem, 89–90
- Behrens-Fisher problem, 61
- Bias
 - estimation, 53, 114, 179, 206
 - observer, 41
 - sample, 7, 39–40, 97, 113–115, 158
 - sources, 121, 166, 181
 - systematic error, 168
- Bias-corrected and accelerated, 54, 69
- Blinding, 41, 102
- Blocks, 38, 66, 80
- Blood, 9, 30, 83
- Blood pressure, 28
- Bonferroni correction, 81
- Bootstrap, 55, 79, 90, , 205
 - limitations, 88
 - nonparametric, 52
 - parametric, 55, 109
 - percentile, primitive, 68
 - sample, 33, 52, 64, 177
 - test, 57–63
 - tilted, 89
- Box and whiskers plot, 107–108
- Boyle's Law, 5, 148

Common Errors in Statistics (and How to Avoid Them), 2e, by Phillip I. Good and James W. Hardin.
Copyright © 2006 John Wiley & Sons, Inc.

Cancer, 19, 23, 94, 97, 158
 Caption, 140
 CART, 187
 Case controls, 42
 Categories, 28, 80
 Cause and effect, 149, 153, 185, 187, 190
 Censored data, 104, 114
 Census, 5, 158
 Central limit theorem, 108
 Chi-square
 statistic, 110
 statistic vs. distribution, 79, 111
 test, 21
 Classification and regression tree (CART), 183
 Clinical
 chemistry, 167
 significance, 61
 trials, 9, 24, 34–35, 40, 90, 94, 98, 103, 143
 Clinician, 38
 Cluster, 6, 63
 Cold fusion, 8
 Computer, 20
 output, 113, 123
 simulation, see Simulation
 Confidence interval, 31, 53, 69, 83, 111–112, 120, 124
 Confounded effects, 4, 79, 97, 112, 120, 147, 151, 177
 Contingency table, 18, 79, 99, 110
 Contour plot, 133, 135
 Contrast, 73
 Controls, 38–40
 positive, 40
 Correlation, 51, 153
 spurious, 159, 176
 vs. slope, 167
 Corticosteroids, 7
 Cost, 29
 Cost-benefit analysis, 23
 Covariances, 64
 Covariates, 66
 Criminology, 152
 Criteria, 117
 Crossovers, 103
 Cross-products, 175
 Cross-validation, 205
 Curve-fitting, 160
 Cutoff value, 70, 110
 Cuts, 114
 Data collection, 28, 47, 113, 117
 Data mining, 182
 Deaths, 104
 Decimal places, 106
 Decision tree, see CART
 Decision theory, 24, 47, 106, 117
 Deduction, 21, 83
 Delete-one, 158, 191
 Density, 133
 Descriptive statistics, 120
 Deterministic vs. stochastic, 5, 219
 Discrimination, 121, 154, 143
 Disease process, 17
 Dispersion, 105–106
 Distribution, 31, 49, 156, 220
 a prior, 94
 cumulative, 220
 empirical, 220
 exponential, 86
 F, 64
 function, 9
 heavy-tailed, 49
 mixture, 54, 58, 74, 109
 multiparameter, 54
 multivariate normal, 62
 nonsymmetric, 49, 54, 106
 normal, 32, 106, 109
 Poisson, 106
 sampling, 68
 symmetric, 49
 uniform, 108
 Weibull, 74
 Diurnal rhythm, 73
 Dropouts, 35, 36
 Drugs, 23, 37, 103
 Ecological fallacy, 170
 Effects, 179
 Elections, 171
 Eligibility, 103
 Emissions, 24
 Endpoints, 112
 Epidemiology, 97, 188
 Equivalence, 61, 75
 Error
 common, 80, 112, 120
 frequency, 158

- probabilities, 16
- terms, 72, 156, 161
- Estimate, 4,
 - consistent, 48
 - efficient, 48
 - interval vs. point, 51
 - least-squares, 51
 - mini-max, 50
 - minimum loss, 50
 - minimum variance, 51, 156
 - optimal, 49
 - plug-in, 51
 - semiparametric, 48
 - unbiased, 51, 156
- Exact, 74
- Experimental design, 69
 - unbalanced, 58
- Experimental unit, 37
- Exploratory analysis, 177
- Extrapolate, 9

- Factors, 7, 24
- Factor analysis, 178
- False dimension, 127
- False negative, see Type I error
- False positive, see Type II error
- F-distribution, 68
- Fisher's exact test, 20
- Forecast, 112
- Found data, 16, 115
- F-ratio, 21, 71–74, 111
- Fraud, 8, 21, 59
- Frequency plot, 10

- GEE, 63, 66, 180
- Geometric mean, 106
- Generalized Linear Models (GLM), 65, 178
- Goodness of fit, 159, 190
- Grammar, 143
- Graphics, 104, 120, 143
- Grid line, 127, 135, 143
- Ground water, 181
- Group, 77
- Group randomized trials, 65, 115, 179
- Growth, 106

- Heterogeneity, 98–99
- Hierarchical models, 98
- Histogram, 107

- HLM, 180
- Hodges-Lehmann estimator, 49
- Hypertension, 14
- Hypothesis, 4, 27, 161,
 - alternative, 18–21, 71–72, 94, 220
 - null, 16, 64, 72, 90, 94, 220
 - ordered, 21
 - post-hoc, 7
 - primary, 18, 70, 86
- Hypothesis testing, 17

- Immunology, 151, 153
- Income, 28, 170
- Independence vs. correlation, 73
- Indifference region, 116
- Induce, 83
- Induction, 21
- Interaction, 74–78
- Interpolation, 124
- Interquartile range, 53
- Interval estimate, 55, 89
- Intraclass correlation, 65

- Jackknife, 191, 205
- Jonckheere-Terpstra statistic, 72

- Kepler's Law, 5
- k-fold resampling, 190
- Kinetic molecular theory, 149
- Kruskal's gamma, 110
- k-sample analysis, 71, 99

- Labels, 141
- Lattice, 77
- Least-absolute deviation, 49
- Least-squares, 49
- Legal applications, 6, 90–93, 121, 154, 158, 170–1
- Legend, 137
- Likelihood, 18
- Likert scale, 60, 73
- Linear vs. nonlinear, 147
- Link function, 178
- Litter, 37
- Location parameter, 32
- Logical assertion, 14
- Logistic regression, 181, 212
- Losses
 - absolute deviation, 22, 48, 50, 72, 157, 160, 161, 164

- jump, 48
- monotone, 48
- square deviation, 48, 50, 161
- step function, 48

- Mail, 37
- Main effect, 74, 78
- Malmquist bias, 114
- Mann-Whitney test, 72
- Marginals, 19, 79
- Maximum, 52, 109
- Maximum likelihood, 47, 51
- Maximum tolerable dose, 15
- Mean absolute deviation, 192
- Mean, arithmetic, 16, 32, 50, 55, 63, 104–7
- Measurements, 28, 29, 48, 114, 105, 117, 148
- Median, 33, 49–52, 89, 105–109
- Medical device, 39, 164
- Medicine, 15, 205–206
- Meta-analysis, 97, 90, 114
- Minimum, 52, 109
- Minimum effective dose, 25
- Missing data, 4, 78, 82, 103, 124
- Mitosis, 38
- Model,
 - choosing, 3, 147, 150
 - nonlinear, 149
 - nonunique, 151
 - range of application, 191
 - general linear, see GLM
 - parametric vs. non-parametric, 220
 - physical, 160, 181
- Monotone function, 70, 78, 148
- MRPP, 157
- Multiple
 - end points, 34
 - tests, 74
- Multivariate analysis, 161, 175

- Negative findings, 106
- Neural network, 182
- Neurology, 114
- Newton's Law, 22, 83
- Neyman-Pearson theory, 17
- Nominative label, 132
- Nonsignificant results, 97, 116, 121
- Nonresponders, 36, 82
- Normal alternatives, 60
- Normal distribution, 80, 107
- Normal scores, 60
- Nuisance parameters, 180

- Objectives, 4, 13
- O'Brien's test, 62
- Observational studies, 97
- Observations
 - dependent, 63,
 - exchangeable, 57
 - identically-distributed, 37
 - independent, 51
 - independent, identically distributed, 37, 57, 111
 - multivariate, 63
 - transformed, 72
- Odds, 95
- One-sided vs. two-sided, 19, 31, 34
- Ordinal data, 132, 219
- Outliers, 88, 157, 164
- Overfitting, 192

- Paired observations, 61
- Parameters, 51, 55, 58, 62, 82, 98, 111, 149
 - location, 32, 49
 - nuisance, 180
 - scale, 32
 - shift, 50
- Paternity, 91
- Patterns, 7, 24
- Percentages, 122
- Percentiles, 51, 53, 89, 109
- Permutation test, 16, 21, 23, 58–63, 66–71, 76, 79–80, 87, 99
- Permutations, 78, 157
- Perspective plot, 133, 135
- Phase III trials, 25
- Physician, 39
- Pie chart, 135–6, 143
- Pitman correlation, 72
- Pivotal quantity, 55, 64, 88
- Placebo, 40
- Plotting region, 129, 139
- Plotting symbol, 137, 141
- Poker, 9–10
- Polar coordinates, 135
- Polyline, 131, 137
- Polynomial, 150
- Population, 3–5, 27, 36–38, 83

Population statistics, 5
 Poverty, 152
 Power, 18, 31–36, 60, 62, 70, 74, 81, 94
 post-hoc, 115
 related to sample size, 31, 101
 Precision, 52, 105, 107, 155, 158
 Prediction, 159, 191–193, 205
 Pricing, 114
 ProcARIMA, 157
 Proc Means, 59
 Proc Mixed, 65
 Protocol, 8, 15, 97, 117
 Proxy variable, 153
 p-value, 10–11, 60, 81–83, 110–112, 116

 Radioimmune assay, 148
 Rain, 188
 Random number, 7, 108
 Randomizing, 40–42, 60, 94
 Ranks, 58
 Rates, 122
 Ratio, 53
 Raw data, 90
 Recruitment, 30, 41
 Recruits, 6, 15, 39
 Redheads, 14, 17, 38
 Redshift, 115
 Regression
 coefficients, 155–7
 Deming (EIV), 165
 LAD, 163
 linear vs. nonlinear, 149, 163
 methods, 163
 multivariable,
 OLS, 164, 166
 quantile, 165
 scope, 147
 stepwise, 176
 stratified, 171
 Regulatory agency, 17, 40
 Rejection region, 18
 Relationship, 148
 Relativity, 22, 83
 Repeated measures, 63, 184
 Report, 28
 Resampling, 188
 Residuals, 76, 105, 160, 162
 Robust, 58, 80, 90

 R-squared, 95
 Rugplot, 109

 Sales, 38
 Sample, 6
 representative, 4, 7, 9, 88
 splitting, 189
 size, 31–36, 41, 58, 104, 109, 105, 119, 167
 universe, 121
 Scale alternative, 62
 Scale parameter, 32
 Scatterplot, 138, 143
 Scope, 147
 Seismology, 114
 Self-selection, 7
 Semiparametric models, 87
 Sex, 14, 19, 154
 Shading, 142
 Shift alternative, 62
 Significance
 practical, 16, 61, 105
 statistical, 155
 Significance level, 18, 31–33, 57–59, 68, 70, 80, 110–111
 Significance level vs. p-value, 221
 Silicon implants, 40
 Simpson’s paradox, 121, 154, 162
 Simulations, 70, 74, 176, 196, 207
 Slope, 150, 152
 Smirnov test, 62
 Sociology, 110
 Software, 32, 34, 156
 Stacked bar chart, 138
 Standard error, 105–106, 193
 Stationarity, 158
 Statistic, 88
 Stein’s paradox, 56
 Stepwise regression, 176
 Stochastic, 148
 Strata, 7, 41, 154, 143
 Studentize, 54
 Subgroups, 15, 38, 80, 97, 121
 Subjective data, 73
 Sufficient statistic, 55
 Surgery, 9
 Surrogates, 153, 158, 189
 Surveys, 7, 28–31, 37–38, 78, 102, 120
 Survival, 19

Tables, 105
 Testing hypotheses, 156
 Tests
 analysis of variance, 71, 74, 76
 bootstrap, 58
 chi-square, 21
 Fisher's exact, 20
 for equality of variances, 67
 for equivalence, 75
 for independence, 79
 F-test, 21, 58, 68, 71
 inferior, 80
 k-sample, 72
 locally most powerful, 63
 Mann-Whitney, 72
 most powerful, 58
 multiple, 81, 112, 176
 one- vs. two-tailed, 21, 60
 permutation, 16, 21, 23, 57-63,
 66-70, 76, 79-80, 87, 99
 t-test, 58, 73
 Texture, 142
 Time series, 57, 114
 Time-to-event data, 74
 Title, 140
 Toxicology, 37
 Transformations, 159
 Treatment allocation, 41, 101, 119
 t-test, 58-60, 64, 73, 81, 111, 123
 Type I and II errors, 23, 29, 32, 58,
 81, 94, 221
 Type I error, 8, 68
 Type II error vs. power, 222
 Unbalanced vs balanced design,
 74
 U-statistic, 58
 Vaccine, 17
 Validation, 3, 160, 186, 188
 Variable
 categorical, 131
 continuous, 131
 confounding, 39
 exchangeable, 64
 explanatory, 176
 indicator, 159
 predictor vs. dependent, 163
 selection of, 10, 175
 Variance, 51, 55, 61, 67-68, 115
 estimator,
 inflation factor, 65
 Variation, 4, 5, 18, 24, 58, 105
 Viewpoint, 133, 135
 Virus, 36
 Voting, electronic, 113
 War, 81
 Weak exchangeability, 76, 99
 Weather, 17, 153
 Wilcoxon test, 62
 Withdrawals, 103