

# Index

- a posteriori test, 131–8, 183
  - after analysis of variance, 131
  - after non-parametric analysis, 257
  - ambiguous result, 136
  - power of, 137
  - Tukey test, 132
- a priori test, 138–40
- accuracy, 28
- alpha ( $\alpha$ ), 60
  - of 0.05, 57, 60
  - other values of, 58–9
- analysis of variance (ANOVA)
  - a posteriori testing, 128
  - assumptions
    - homoscedasticity, 166
    - independence, 171
    - normality, 167
  - fixed and random effects, 128
  - Model I, 129
  - Model II, 129
  - multiple comparisons, 131
  - nested, 185–92
  - risk of Type 1 error, 116
  - single-factor, 115–29
    - a posteriori testing, 128
    - unbalanced design, 128
  - three or more factors, 164
  - two-factor, 142–64
    - a posteriori testing, 155
    - cautions and complications, 154–64
    - fixed and random factors, 160
    - interaction, 144
    - interaction obscuring main effects, 157
    - no interaction, 144
    - simple main effects, 150
    - unbalanced designs, 164
  - two-factor without replication, 178–84
    - randomized blocks, 184
- ANOVA, *see* analysis of variance
- autocorrelation, 301–7
- autoregression, 317–19
- average, *see* mean
- Bayes’ theorem, 61–2
- beta ( $\beta$ ), *see* Type 2 error
- binomial distribution, 56, 81, 245
- box-and-whiskers plot, 167–71
- Box–Ljung statistic, 307
- central limit theorem, 80
- chi-square
  - statistic ( $\chi^2$ ), 62
    - table of critical values, 374
  - test, 60–4, 131–8, 231–3
    - bias with one degree of freedom, 237
    - example to illustrate statistical testing, 60
    - for heterogeneity, 235
    - for one sample, 231
    - worked example, 62
    - Yates’ correction, 237
- cluster analysis, 291–4
- Cochran *Q* test, 245
- confidence interval, 85
- confidence limits, 85
- contingency table, 235
- control treatments, 35
  - for disturbance, 36
  - for time, 36
- correlation, 29, 194–203
  - artifact of closed data sets, 30
  - confused with causality, 30
  - contrasted with regression, 195
  - in sequence analysis, 302
  - linear, 195–202
    - assumptions, 202
    - non-parametric, 266
  - Pearson correlation coefficient ( $r$ ), 195–202
- correlogram, 303
- D* statistic, *see* Kolmogorov–Smirnov one-sample test
- data
  - bivariate, 16, 194
  - continuous, 17
  - discrete, 17

Cambridge University Press

978-0-521-76322-6 - Geostatistics Explained: An Introductory Guide for Earth Scientists

Steve McKillup and Melinda Darby Dyar

Index

[More information](#)**392 Index**

- data (cont.)
  - displaying, 15–27
    - bivariate, 21
    - circular histogram, 348
    - cumulative graph, 20
    - directional, 348
    - histograms, 17
    - line graphs, 19
    - multivariate, 26
    - nominal scale, 21
    - ordinal scale, 21
    - pie diagrams, 25
    - rose diagrams, 26, 348
    - univariate, 17
  - interval scale, 16
  - multivariate, 16, 270
  - nominal scale, 16, 230
  - ordinal scale, 16
  - ratio scale, 16
  - univariate, 16
- degrees of freedom, 90
  - additivity in ANOVA, 127
  - for chi-square test, 232
  - for contingency tables, 236
  - for *F* statistic, 127
  - in regression analysis, 216
- directional data, 346–52
- dissimilarity, 284
- effect size, 107
- eigenvalue, 275, 280
- eigenvector, 275
- empirical survival graph, 329
- error
  - Type 1, 58, 105
  - Type 2, 58, 106
  - uncontrollable, 118
- ethical behavior, 45–50
  - acknowledging previous work, 46
  - approvals and permits, 47
  - fair dealing, 46
  - input of others, 46
  - moral judgements, 47
  - plagiarism, 45
  - pressure from peers or superiors, 49
  - quality control, 50
  - record keeping, 49
  - reporting results, 48
- Euclidian distance, 285
  - for cluster analysis, 291
- exact tests
  - Fisher Exact Test, 238
  - for two independent samples, 252–4
- examples
  - air pollution in a city, 283
  - aluminum oxide in lower crustal xenoliths, 106
  - apatite
    - and lead contamination, 8
    - chlorapatite and fluorapatite binding lead, 186
    - in human teeth, 24
    - in sandstone, 136
  - asbestos and mesothelioma, 58, 259
  - beach erosion/accretion in Massachusetts, 243
  - coal depth in West Virginia, 358
  - Copernicus' model of the solar system, 12
  - Cretaceous/Tertiary boundary, 238
  - dental caries in Hale and Yarvard, 22
  - earthquakes and tilt of volcanoes, 30
  - estuary sediments, 278
  - flood frequency over time, 327
  - forams and their coiling, 232
  - fossil clams in outcrops, 100, 250
  - gold pan performance, 34
  - grain size of granites, 96
  - greenhouse gasses forming minerals, 143
  - groundwater
    - arsenic contamination, 59
    - nitrate contamination, 235
  - hafnium in New England granites, 134
  - hydrogen diffusion in pyroxene, 315
  - impact craters
    - iridium concentration, 264
    - on the Earth and Moon, 41
    - on Venus, 343
  - irradiation
    - of amethyst, 94, 146
    - of quartz, 153
    - of topaz, 5
    - of zircons, 60
  - kimberlite and diamonds, 51
  - limestone
    - and acid rain, 195
    - isotopic analysis of, 19
    - weathering of tombstones, 248
  - magma temperature estimated from SiO<sub>2</sub> content, 204
  - magnetic field strength of the Earth, 298
  - meteorites in Antarctica, 337
  - mine site bioremediation, 35
  - oil shale test cores, 178
  - olivine production and temperature, 41
  - pearl growth, 184
  - potable water from wells, 61
  - quartz crystals and handedness, 231
  - radiometric decay and age of rock, 210–11
  - rainfall prediction at Neostrata coal mine, 317
  - sand mineralogy and location, 230
  - sea level
    - at Port Magmamo, 320
    - from sedimentary deposits, 324
  - sediments
    - heavy metal content of, 278, 283

	Index	393
hydrocarbons in, 282		
in a glacial lake, 32		
in the Holyoke estuary, 31		
snow and school cancellations, 217–19		
stream directions in western Queensland, 350		
student visits to a lecturer's office, 17		
temperature and pressure on feldspar growth, 179		
topaz crystals, heat-treated, 4		
tornadoes in the USA, 21		
tourmalines		
in Maine pegmatites, 117		
magnesium content, 122		
stable isotopes in, 135		
turbidity of well water (NTU), 310		
vermiculite		
and soil improvement, 262		
water content of, 94		
Vienna Standard Mean Ocean Water, 88		
examples, worked		
a posteriori Tukey test, 131–8		
chi-square test		
for goodness of fit, 62, 232		
for heterogeneity, 235		
of directional data, 351		
Fisher Exact Test, 238		
multidimensional scaling, 287–9		
nearest neighbor analysis, 345		
regression		
linear, 217–19		
non-linear, 320–2		
of a sequence, 310–11		
of event frequency over time, 327		
rose diagram, 350–1		
semivariogram, 358		
single-factor analysis of variance, 122–8		
spatial distribution of meteorites, 337–40		
<i>t</i> test		
independent samples, 100		
paired-sample, 98		
single-sample, 95–6		
two-factor analysis of variance, 153–4		
Type 2 error, 106–9		
<i>Z</i> test, 88–9		
experiment		
advantage of repeating, 34		
common sense, 42		
good design versus cost, 44		
manipulative, 28, 34–40		
apparent replication, 38		
control treatments, 35		
need for independent replicates, 34		
pseudoreplication, 37		
mensurative, 28, 29–34		
confounded, 31		
need for independent samples, 32		
need for replicates, 32		
pseudoreplication, 32		
negative outcome, 12		
realism, 41		
unreplicated, 40		
experimental design, 28–44		
orthogonal, 143		
experimental unit, 2		
<i>F</i> statistic, 102, 122		
degrees of freedom for, 127		
in regression testing, 215		
table of critical values, 378		
factorial ( $X!$ ), 337		
false negative, <i>see</i> Type 2 error		
false positive, <i>see</i> Type 1 error		
Fisher Exact Test, 238		
Fisher, R., 57, 116		
Friedman test, 262–4		
<i>G</i> test, 236		
<i>H</i> statistic, <i>see</i> Kruskal–Wallis test		
heteroscedasticity, 166		
Levene test for, 167, 174		
homoscedasticity, 166		
Hurlbert, S, 32, 37		
hypothesis, 8		
alternate, 12		
becoming a theory, 11		
cannot be proven, 11		
null, 12		
predictions, 8		
rejection of, 8		
retained or rejected, 11		
two-tailed contrasted with one-tailed, 91		
interaction, 142		
Kolmogorov–Smirnov one-sample test, 248–50		
Kriging, 360		
Kruskal–Wallis test, 256–7		
lag, 301		
leptokurtic, 80		
Levene test, 167		
line graph, 19		
log-likelihood ratio, <i>see G</i> test		
Mann–Whitney <i>U</i> test, 250–2		
Markov chains, 324		
McNemar test, 244		
mean, 67		
grand mean, 117		
population, 67, 86		
sample, 72		
standard error of (SEM), 73		

394 Index

- median, 82, 167
- meta-analysis, 40
- mode, 83
- Monte Carlo method, 234, 237
- multidimensional scaling, 284–91
- multivariate analyses, 270–95
  - choice of, 295
  - cluster analysis, 291–4
    - dendrogram, 292
    - group average linkage method, 292
    - hierarchical, 292
  - multidimensional scaling, 284–91
    - cautions, 290
    - Euclidian distance, 285
    - stress, 289
  - principal components analysis (PCA), 272–84
    - cautions in use of, 283
    - eigenvalues, 275
    - number of components to plot, 282
    - practical use of, 282
    - principal components, 272
    - redundancy, 272
  - Q-mode, 271, 291
  - R-mode, 271
- non-parametric tests, 66, 227–9
  - correlation, 266
  - independent samples of nominal scale data
    - chi-square test, 131–8, 231–3
    - Fisher Exact Test, 238
    - G test, 236
    - inappropriate use of, 242
    - randomization test, 233–4
  - independent samples of ratio, interval or ordinal data, 248–59
    - analysis of variance on ranks, 257
    - exact test, 252–4
    - Kolmogorov–Smirnov one-sample test, 248–50
    - Kruskall–Wallis test, 256–7
    - Mann–Whitney *U* test, 250–2
    - randomization test, 252, 257
  - related samples of nominal scale data, 243–5
    - Cochran *Q* test, 245
    - McNemar test, 243
  - related samples of ratio, interval or ordinal data, 259–64
    - Friedman test, 262–4
    - Wilcoxon test, 259–62
  - when transformation cannot remedy heteroscedasticity, 264
- normal distribution, 66
  - bivariate, 202
  - leptokurtic, 80
  - platykurtic, 80
  - skew, 80
- nugget (effect), 356
- one-tailed
  - hypotheses, 91
    - appropriateness, 94
    - cautions, 92
  - tests and critical value, 93
- orthogonal, 143
- outliers, 167, 169
- parametric tests, 66
- Pearson correlation coefficient (*r*), 195–202
- planned comparisons, *see* a priori test
- platykurtic, 80, 81
- Poisson distribution, 82, 336
- Popper, K., 8, 13
- population, 1
  - statistics (parameters), 71
- post hoc test, *see* a posteriori test
- power of a test, 109
  - and sample size, 111
  - controllable factors, 110
  - desirable, 110
  - uncontrollable factors, 110
- precision, 28
- principal components analysis (PCA), 272–84
- probability
  - and statistical testing, 51–65
  - essential concepts, 53–5
  - reported value
    - < 0.001, 60
    - < 0.01, 60
    - < 0.05, 60
    - ≥ 0.05 not significant (N.S.), 60
    - close to 0.05, 65
    - of exactly 0.05, 64
  - statistically significant, 57
  - Type 1 error, 58
  - Type 2 error, 58
- probability plot (P-P), 101, 167
- pseudoreplication, 37
  - inappropriate generalization, 34
  - replicates
    - clumped, 38
    - placed alternately, 38
    - segregated, 38
    - sharing a condition, 39
- Q-mode, 284
- quadrat, 335
- r* statistic, 195
- r*<sup>2</sup> statistic, 217
- randomization test
  - concept of, 233

	Index	395
<ul style="list-style-type: none"> <li>for contingency tables, 237</li> <li>for nominal scale data, 233</li> <li>for three or more independent samples, 257</li> </ul>		
randomized blocks, 184		
range, 68, 83		
ranks, 247		
tied, 251		
redundancy, 272		
regional dependence, 352		
regression		
contrasted with correlation, 195		
linear, 204–23		
assumptions, 220		
coefficient of determination ( $r^2$ ), 217		
danger of extrapolation, 220		
equation, 205		
intercept, 208		
predicting $X$ from $Y$ , 219		
predicting $Y$ from $X$ , 219		
significance testing, 211–17		
slope, 205–8		
multiple linear, 223–4		
polynomial, 311–15		
replicates, 9		
residuals, 220		
use in sequence analysis, 309		
$R$ -mode, 271		
rose diagram, 348		
sample, 117		
mean, 72		
random, 2		
representative, 1, 6, 66		
standard deviation ( $s$ ), 71		
statistics, 71		
as estimates of population statistics, 76		
variance ( $s^2$ ), 72		
sampling unit, 1		
replicates, 117		
scientific method, 8, 45		
core theories, 14		
hypothesis, 8		
hypothetico-deductive, 8, 13		
paradigms, 13		
semivariance, 353		
compared to variance, 353		
semivariogram		
application of, 359		
experimental (empirical), 354		
nugget effect, 356		
region of influence, 357		
sill, 357		
theoretical, 355		
sequence analysis, 297–332		
nominal scale data, 323–30		
changes of state, 323–7		
randomness over time, 329		
regression modeling of frequency, 329		
repeated occurrence of an event, 327–30		
sampled at regular intervals, 324		
transitions, 325, 326		
ratio, interval or ordinal data, 298–323		
autocorrelation, 301–7		
autoregression, 317–19		
Box–Ljung statistic, 307		
cautions in use of, 323–30		
correlogram, 303–7		
correlogram of residuals, 309		
cross-correlation, 307		
cyclic pattern, 320–2		
detrending with regression, 309		
lag, 301		
polynomial regression analysis of, 311–15		
preliminary inspection, 298		
regression modeling of, 308–23		
simple linear regression, 309, 310–11		
similarity within a sequence, 307		
serial correlation, <i>see</i> autocorrelation		
significance level		
of 0.01, 59		
of 0.05, 57		
of 0.3, 59		
skew, 80, 167		
negative, 80		
positive, 80		
spatial analysis, 334–62		
direction of objects, 346–52		
testing for an even distribution, 351		
distribution of objects, 335–46		
edge effects, 344		
guard region, 344		
Monte Carlo simulation, 343		
nearest neighbor analysis, 342–6		
scale and sensitivity, 342		
testing for randomness, 335–46		
orientation of objects, 351		
prediction and interpolation in two dimensions, 352–62		
Kriging, 360		
regional dependence, 352		
semivariance, 353		
semivariogram, 354		
theoretical semivariogram, 355		
Spearman's rank correlation, 266		
standard deviation, 67		
population, 70		
sample, 71		
standard error of the mean (SEM), 73		
calculated from ANOVA mean square (error), 132		
estimated from one sample, 75		

396 Index

- standard error of the mean (SEM) (cont.)
  - for a population, 74
  - samples from non-normal distributions, 80
- standard normal distribution, 71
- statistic, 60
- statistical significance, 52
- stress, 289
- Student (W. S. Gossett), 78
- Student's *t*, *see t* statistic
- survival graph, 329
- t* statistic, 86
  - distribution of, 78
  - table of critical values, 376
- t* test
  - assumptions, 100
    - equal variances, 102
    - normality, 101
  - critical values, table of, 90, 376
  - degrees of freedom, 90–1
  - for planned comparison after ANOVA, 138
  - independent samples, 98
  - paired sample, 96
  - single sample, 87
- theory, 11
- time series analysis, *see* sequence analysis
- transformations, 171
  - arc-sine, 172
  - legitimacy, 172
  - logarithmic, 171
  - square root, 171
- Tukey test, *see a posteriori* test
- two-tailed hypotheses, 91
- Type 1 error, 58, 105–6
  - a posteriori* testing, 137
  - concept of risk, 113
  - multiple comparisons, 116
  - trade-off with Type 2 error, 110
- Type 2 error, 58, 106–9
  - concept of risk, 113
  - trade-off with Type 1 error, 110
- variable, 15
  - dependent, 195
  - independent, 195
  - measured, 298
  - sequence, 298
- variance
  - among group, 119
  - for a population, 68
  - for a sample, 72
  - tests for heteroscedasticity, 174
  - total, 119
  - within group (error), 119
- variation
  - among group, 125
  - total, 126, 148
  - within group (error), 124, 148
- Wilcoxon test, 259–62
- X axis, 18
- Y axis, 18
- Yates' correction, *see* chi-square test
- Z statistic, 70, 75, 76, 86, 284
  - test using, 86
  - use in correlation, 198