

# Index

NOTE: Page numbers in *italics* refer to Figures

- absolute risk reduction (ARR) 146–7, 147
- absolute risks 146–7, 147, 148
- adjustment
  - confidence intervals 193–4, 195–6
  - confounders 92, 95
  - goodness-of-fit 284–6
- agreement 260–266, 261
- association 260–261
  - Bland–Altman charts 265–6
  - Cohen's kappa 262–4, 262
  - limits 265
  - weighted kappa 263, 264
- analysis of variance (ANOVA) 293
- analytic cross-sectional studies 104–6
- Apgar scores 209
- arithmetic mean 67–8, 68
- assessment bias 122
- association
  - agreement 260–261
  - correlation coefficient 252–4
  - linear 251
  - linear regression 270–271
  - negative 249
  - positive 249
  - see also* scatterplots
- automated variable selection 288–9
  
- backward elimination 289
- backwards selection 288
- bar charts 51–4
  - clustered 35–7, 35–7
  - simple 34, 34–5
  - stacked 37, 38–9
- baseline table 11, 13, 14
- Bayesian probability 141
- beneficial risk factor 156
- bimodal distribution 56, 56
- binary data 224–5, 295
- binary-dependent variables 295–7, 296
  
- binomial distribution 55, 142–3, 170–171
- Bland–Altman charts 265, 265–6, 266
- blinding
  - assessment bias 122
  - double-blind randomised controlled trial 122
  - parallel group trial 122, 122
  - response/placebo bias 121–2
- block randomisation 120–121
- boxplots 42–3, 43–4, 76–8, 77–8
  
- case-control studies
  - versus* cohort design 115–16
  - confidence intervals 196–8, 198
  - confounding 113, 114
  - frequency-matched case-control study 114, 114
  - hypothesis tests 213, 228, 243
  - matched 94, 154
  - nested case-control design 114
  - probability 138
  - risk ratio 155
  - structure 111–12, 112
- case report 102–3
- case series studies 103
- categorical data 11
  - agreement 262
  - charts 32–9, 32–9
  - frequency tables 18–22, 19–22
  - hypothesis test 209, 222
  - linear regression 286
  - nominal 6–7
  - ordered 236, 236
  - ordinal 7–8
- causal relationship 271–3
- chance-corrected proportional agreement statistic 262
- charts 32–48
  - bar charts 34–7, 35–9
  - boxplots 42–3, 43–4, 76–8, 77–8
  - categorical data 32–9, 32–9
  - continuous data 40–42, 41–2, 44–6, 45–6

- charts (*continued*)
  - cumulative data 44–7, 45–6
  - discrete data 39–40, 39–40, 44, 45
  - distribution 51–4
  - histogram 40–42, 41–2
  - metric data 39–42, 39–42
  - nominal data 32–7, 32–7
  - ordinal data 32–7, 32–7
  - pie chart 32–4, 33–4
  - step charts 44, 45
  - time-based data 47–8, 47–8
- chi-squared test
  - hypothesis tests 209, 222, 226–38
  - logistic regression 304
  - survival 319
- clinical databases 326
- clinical trials 119–20
- clustered bar chart 35–7, 35–7
- cluster sampling approach 131–2
- coding design 286–7
- coefficient of determination 281
- Cohen's kappa 262, 262–4
- cohort studies
  - charts 42
  - confounding 110–111
  - features 108
  - principal structure 107
  - prospective cohort study 107–8, 109
  - retrospective cohort study 108, 109
  - shortcomings 109–10
- conditional probability 141
- confidence intervals
  - agreement 262
  - difference between population parameters 173–89
  - hazard ratios 198–200, 199
  - hypothesis tests 173–89, 204, 218–220, 229, 241, 241
  - independent populations 174–81, 186–9
  - linear regression 291
  - logistic regression 302
  - Mann-Whitney rank-sums test 187–8, 189
  - matched populations 189–90, 190
  - mean 164–70, 166, 170
  - median 170–173, 173, 187–90, 189–90
  - Minitab 175–7, 178
  - Normal distribution 165–6, 166
  - odds ratio 196–8, 198
  - proportions 170–171, 172, 186–7, 187
  - ratio of two population parameters 192–200
  - risk ratios 193–6, 195–7
  - single population parameter 164–74
  - SPSS 178, 179
  - standard error 165–70
  - survival 322
  - systematic review 328–9
  - two independent population means 192–3, 192–3
- confounders
  - data analysis stage
    - adjustment 92, 95
    - stratification 94–5
  - design stage
    - matching 94
    - randomisation 95
    - restriction 94
- detection 93
- Down syndrome babies 91–2, 92
- linear regression 290–291
- logistic regression 304
- myocardial infarctions, causes 90–91, 90–91
- negative 92
- odds ratio 158
- positive 92
- risk ratios 152, 151–2, 194
- study design 110–111, 113, 114
- consecutive samples 132
- constant coefficient 273
- contingency table 28–30, 29–30
- chi-squared test 227–8
- logistic regression 304
- risk 149
- continuous metric data 10
  - agreement 264–6
  - charts 40–42, 41–2, 44–6, 45–6
  - frequency tables 22–4, 23, 26–7
  - linear regression 275
  - numeric summary values 66, 70
- controlling for confounders 93
- correlation coefficient 252
- counts 9
- Cox's regression model 322–4
- cross-over randomised controlled trial 123–4, 123–4
- cross-sectional studies
  - analytic 104–6
  - descriptive 103–4
- cross-tabulation 28–30, 29–30
- cumulative data 44–7, 45–6
- cumulative frequency 25–7, 26–8
- data
  - definition 3–4
  - identification algorithm 11
  - metric data 9–10
  - sample data 3–4
- databases 326
- deciles 70
- decision rules 206–207
- dependent variables 282, 295–7
- descriptive cross-sectional studies 103–4
- descriptive statistics
  - charts 32–48
  - definition 17–18
  - distribution 49–58
  - frequency tables 17–30
- design variables 286–7
- deviance coefficient 306
- diagnostics 292–3
- discrete metric data 9, 11
  - charts 39–40, 39–40, 44, 45
  - frequency tables 23–4, 23–4
  - numeric summary values 70
- dispersion measures 60, 73–86
- distribution
  - bimodal 56, 56
  - binomial 55, 170–171

## INDEX

383

- negative skew 50–51, 51  
 Normal distribution 54–5, 54–5  
 numeric summary values 66–8, 70, 71, 82–3, 83, 846  
 outliers 51  
 Poisson 55  
 skew 51–3, 52–3, 56–8, 57  
 symmetric 50, 53, 53  
 transformed data 84, 85, 86  
 uniform 50  
 double-blind randomised controlled trial 122  
 drop-out 125  
 dummy variables 286  
  
 ecological fallacy 117  
 ecological study 116, 116  
 errors  
   drop-out 125  
   hypothesis test 220–221  
   linear regression 276, 277  
   sampling 129–130, 165–6, 166  
 estimates 128, 130–131 *see also* confidence intervals  
 exclusion criteria 326  
 expected values 229–30, 262  
 experimental studies 101  
   clinical trials 119–20  
   randomisation 120–126, 122–4  
  
 false negatives 221  
 false positives 220  
 Fisher's exact test 209, 233, 233–4  
 forest plot 328, 328–9, 329  
 forward elimination 289  
 forward selection 288  
 frequency distribution 19, 19–20  
 frequency-matched case-control study 114, 114  
 frequency matching 94  
 frequency tables 17–30  
   categorical data 18–22, 19–22  
   contingency table 28–30, 29–30  
   continuous data 22–4, 23  
   cross-tabulation 28–30, 29–30  
   cumulative frequency 25–7, 26–8  
   discrete data 23–4, 23–4  
   grouping data 26–7  
   metric data 22–5  
   nominal data 18–20, 19–20  
   open-ended groups 27, 28  
   ordinal data 20–22, 21–2  
   ranking data 14, 14  
   relative frequency 20, 20  
 funnel plots 330–332, 331, 332  
  
 GCS *see* Glasgow Coma Scale  
 generalised linear model 293  
 Glasgow-Blatchford Bleeding Scale (GBS), 340  
 Glasgow Coma Scale (GCS), 7  
 goodness-of-fit 280–281, 284–6, 306, 307  
 grouped data 26–7, 26–7, 40–42, 41–2  
 grouped frequency distributions 26–7, 26–7  
  
 hazard ratios 198–200, 199, 244, 244, 320–321, 323, 324  
 histograms 40–42, 41–2  
  
 confidence intervals 175  
   distribution 51–2, 68, 68, 85  
   numeric summary values 68, 68, 85  
 Hoehn scale scores 22, 22  
 homoskedasticity 278  
 Hosmer-Lemeshow statistic 306  
 hypothesis tests  
   chi-squared test 209, 222  
   confidence intervals 173–89, 204, 218–20, 229, 241, 241  
   *versus*. confidence intervals 218–20  
   decision rules 206–7  
   difference between population parameters 210–213  
   equality of population proportions 214, 227–30  
   errors 220–221  
   Fisher's exact test 209  
   independent population 227  
   independent populations 210–211, 223–5  
   Kruskal-Wallis test 209  
   Mann-Whitney rank sums test 209, 215, 215–18, 217  
   matched-pairs t test 209, 213–14, 214  
   McNemar's test 209  
   mean 208–11  
   median 214–15, 218  
   Minitab 211–12  
   Normal distribution 208, 213, 217  
   null hypothesis 205  
   paired populations 213–14  
   power 221–2  
   procedure 206  
   proportions 227–30  
   *p* values 206–8, 210, 210, 211, 230, 242, 242  
   ratio of two population parameters 204–25, 239–44  
   research questions 204–5  
   rules of thumb 223–5  
   significance level 207, 223–4  
   SPSS 212–13, 218  
   trend 209, 236–8, 238  
   two-sample t test 185, 188, 209–13, 211–12  
   Wilcoxon signed-ranks test 173, 189, 190, 209, 218  
   *see also* null hypothesis  
  
 inclusion criteria 326  
 independent populations  
   difference 187–8, 189  
   hypothesis test 210–211, 223–5, 227  
   ratios 192–3, 192–3  
   Whitney rank sums method 187–8, 189  
 independent variables  
   linear regression 282  
   logistic regression 298, 303–4  
 inferences 128–30  
 Injury Severity Score (ISS) 263, 264  
 intention-to-treat analysis 125–6  
 interquartile range (IQR) 73–6, 74–6, 317  
 inter-rater agreement 260  
  
 journals 329  
  
 Kaplan–Meier curve 315–17, 316–17  
 Kaplan–Meier table 313–15  
 Kendall's rank-order correlation coefficient 258–9

linear regression  
 analysis of variance 293  
 association 270–271  
 assumptions 277–8  
 causal relationship 271–3, 272  
 coding design 286–7  
 confounders 290–291  
 design variables 286–7  
 diagnostics 292–3  
 goodness-of-fit 280–281, 284–6  
 Minitab 279–80  
 model building 273–6, 287–8  
 multiple 282, 282–4  
 nominal independent variables 286–7  
 ordinary least squares 277–8  
 population regression equation 278  
 sample regression equation 275  
 SPSS 278–9, 279  
 statistical significance 278–80  
 variable selection 288–290  
 variation 271–3

location measures 64–71

logistic regression 294–307  
 binary-dependent variables 295–7, 296  
 goodness-of-fit 306  
 maximum likelihood estimation 298–300  
 Minitab 303  
 model building 297–8, 303–6  
 odds ratios 300–301  
 Poisson 306–7  
 regression coefficient 301, 302  
 SPSS 298  
 statistical inference 301–2

log–log plot 324

log-rank test 319–20, 320

Mann-Whitney rank sums test 187–8, 189, 209  
 Minitab output 216–18, 217  
 ranking procedure 215, 215–16  
 SPSS output 218  
 Wilcoxon signed-ranks test 218

Mann-Whitney test 227

manual variable selection 289–90

matched case-control studies 94, 154

matched-pairs *t* test 209, 213–14, 214

matched populations 189–90, 190, 218

maximum likelihood estimation 298–300

McNemar's test 209

mean  
 confidence intervals 164–70, 166, 170  
 hypothesis test 208–11  
 linear regression 275  
 numeric summary values 67–8, 68  
 standard error 165–70  
 statistical inference 129

measurements 8, 10

median  
 confidence intervals 172–3, 173, 189, 190  
 hypothesis test 214–15, 218  
 numeric summary values 66–7  
 survival 316–17

meta-analysis 332–5

charts 39–42, 39–42  
 frequency tables 22–4

metric data 10  
 agreement 264–6  
 continuous 10  
 discrete 9

mode 65, 65–6

model building 273–6, 287–8

mound-shaped distributions 50, 53, 53

multicollinearity 283

multiple linear regression 282, 282–4, 283

n-tiles 70

negative  
 confounding 92  
 outcomes 329  
 skew 50–51, 51

negative predictive value (NPV) 337

nested case-control design 114

nominal categorical data 6–7, 10  
 charts 32–7, 32–7  
 frequency tables 18–20, 19–20  
 linear regression 286–7  
 numeric summary values 64, 70

non-parametric tests 187–8, 208, 209, 218

Normal distribution 54–5, 54–5  
 confidence intervals 165–6, 166  
 linear regression 277  
 probability 144  
 standard deviation 82–3, 83, 84

null hypothesis 205  
 difference between population parameters 204–25  
 ratio of two population parameters 231, 232, 240–241  
 survival 319

number needed to treat 149–50

numeric summary values  
 dispersion measures 60, 73–86  
 distribution 66–8, 70, 71, 82–3, 83, 84  
 interquartile range (IQR) 73–6, 74–6  
 location measures 64–71  
 numbers 60–61  
 outliers 73  
 percentages 60–64  
 percentiles 68–70  
 proportions 60–61  
 quantitation 60  
 range 73  
 skew 67, 71, 86  
 standard deviation 79–84  
 transformed data 84, 85, 86

observational study designs  
 case-control studies 111–16, 112, 114  
 case report 102–3  
 case series studies 103  
 cohort studies 108–11, 109  
 cross-sectional studies 103–6  
 ecological fallacy 117  
 ecological study 116, 116  
 issues 101  
 portfolio 101, 102  
 research questions and methods 100

## INDEX

385

- observed values 231
- odds ratios
  - calculation 157–8, 158
  - confidence intervals 196–8, 198
  - confounding 158
  - hypothesis tests 242–3
  - logistic regression 300–301
  - risk ratio approximating 159
  - stroke and exercise study 156–7, 157
  - systematic review 329
- one-to-one matching 94
- open-ended groups 27, 28
- open trials 122
- ordered categorical data 236, 236
- ordering of data 7–8, 20–22, 21–2
- ordinal categorical data 7–8, 10
  - charts 32–7, 32–7
  - confidence intervals 188, 189
  - frequency tables 20–22, 21–2
  - hypothesis test 208
  - numeric summary values 64, 67, 70, 71
- ordinary least squares (OLS) 277–8
- outliers
  - distribution 51
  - frequency tables 27
  - numeric summary values 73
- p* values
  - hypothesis tests 206–8, 210, 210, 211, 230, 242, 242
  - linear regression 284
  - logistic regression 305
  - survival 323
- Palliative Care Outcome Scale (POS) 263, 264
- parallel design 122
- parametric tests 187, 208, 209, 222
- parsimony 290
- Pearson's correlation coefficient 256, 256
- percentages
  - cumulative frequency 44–7, 45
  - frequency 20, 20
  - numeric summary values 60–64
- percentiles 68–70
- perfect agreement 261
- pie charts 32–4, 33–4
- placebo bias 121–2
- Poisson distribution 55, 143–4
- Poisson regression 306–7
- population
  - correlation coefficient 254
  - difference between parameters 173–89
  - logistic regression 297
  - mean 164–70, 166, 170, 208–11
  - median 172–3, 173, 189, 190, 214–15, 218
  - odds ratios 196–8, 198, 243, 243
  - parameters 128
  - proportions 170–171, 172
  - ratio of two parameters 192–200, 239–44
  - regression equation 278
  - risk ratios 193–6, 195–7, 240, 240–241
  - single parameter 164–74
  - statistical inference 128–130
  - survival 314
- positive
  - confounding 92
  - outcomes 329
  - skew 51–3, 52–3
- positive predictive value (PPV) 337
- power of a test 221–2
- prediction 280
- probability
  - binomial probability distribution 142–3
  - calculation 138, 139
  - case-control studies 138
  - conditional and Bayesian statistics 141
  - definition 138
  - dice rolling experiment 141, 141–2
  - discrete *versus*. continuous 142
  - logistic regression 298, 299
  - Normal distribution 144
  - number needed to treat 149–50
  - odds 156
  - Poisson distribution 55, 143–4
  - proportional frequency 138–9
  - simple probability
    - addition rule 140
    - multiplication rule 139
  - survival 314
- proportional agreement 261
- proportional frequency 138–9
- proportional hazards 324
- proportions
  - confidence intervals 170–171, 172, 186–7, 187
  - hypothesis test 227–30
  - numeric summary values 60–61
  - populations 227
- publication bias 329–30
- quintiles 70
- randomised controlled trial (RCT)
  - hypothesis tests 211, 220
  - study design 120, 123–5
- random number tables 120
- range 73
- ranked data
  - frequency tables 14, 14
  - log-rank test 319–20
  - Mann-Whitney rank-sums test 209, 214–19
  - Spearman's rank correlation coefficient 256–9
  - Wilcoxon signed-rank test 173, 189, 190, 209, 218
- receiver operating characteristic curve (ROC) 339, 340
- reference value 149–51, 150, 151
- regression coefficient 301, 302
- relative frequency 20, 20, 45, 76
- relative risk reduction 148–9
- residuals 276–7
- response bias 121–2
- risk 298
- risk ratios 147–8
  - case-control study 155
  - confidence intervals 193–6, 195–7
  - confounding 152, 151–2
  - formula 149
  - hypothesis tests 240–241

## 386

risk ratios (*continued*)  
 relative risk reduction 148–9  
 survival 322  
 systematic review 331  
 rules of thumb 223–5

sample  
 consecutive 132  
 logistic regression 297  
 mean 165–6, 166  
 percentage 129–30  
 regression equation 275  
 simple random 130  
 statistics 128, 129  
 stratified random 130–131  
 survival 314  
 systematic random 130

sampling  
 errors 129–30, 165  
 inclusion and exclusion criteria 133  
 sample size 132–3  
 sampling error 129–30  
 statistical inference 128–30  
 target population 128–9, 130–131  
 types 129–32

scatterplots  
 limitations 250  
 linear regression 276  
 logistic regression 295–6

sensitivity 337

significance level 207, 223–4

simple bar chart 34, 34–5

simple random sample 130

skew 51–3, 52–3, 67, 71, 86

slope coefficient 273

Spearman's rank correlation coefficient 256–9, 257, 259

specificity 337

spread measures 60, 73–86

stacked bar chart 37, 38–9

standard deviation 79–84, 81, 83–4  
 agreement 266  
 confidence intervals 174, 184

standard error 165–70

statistical inference 128–9, 301–2

step charts 44, 45

stepwise selection 288–9

straight line model *see* linear regression

stratified random sample 130–131

study design  
 blinding 121–2, 122  
 case-control studies 111–16, 112, 114  
 case series studies 103  
 clinical trials 119–20  
 cohort studies 108–11, 109  
 confounders 94–5  
 cross-sectional studies 103–6  
 experimental studies 119–26  
 intention-to-treat analysis 125–6  
 matching 94  
 randomisation 120–126, 122–4  
 randomised controlled trial 120, 123–5  
 study populations 128–130

## INDEX

sum of squares 80

survival  
 censored data 312–13  
 comparison between groups 317–18  
 Cox's regression model 322–4  
 hazard ratio 320–321  
 Kaplan–Meier curve 315–17, 316–17  
 Kaplan–Meier table 313–15  
 log–log plot 324  
 log-rank test 318–20  
 median 316–17  
 probability 314  
 proportional hazards 324  
 single groups 312–13, 313  
 symmetric distribution 50, 53, 53  
 systematic random sample 130

systematic review  
 exclusion criteria 326  
 forest plot 328, 328–9, 329  
 funnel plot 330–332  
 inclusion criteria 326  
 meta-analysis 332–5  
 methods 327  
 publication bias 329–30  
 search strategy 327  
 selection criteria 327

*t* distribution 168, 174, 187–8

target population 128–9

test statistic 236

time series chart 47–8, 47–8

transformed data 84, 85, 86

treatment bias 122

treatment group 122

trend 209, 236–8, 238

two-sample *t* test 209–11, 304  
 Minitab output 211, 211–12  
 SPSS output 212, 212–13  
 type I/II errors 220–221

uniform distributions 50

units 11

univariate logistic regression 304

variables  
 characteristics 11–14  
 definition 3–4  
 selection 289–90  
 types 6, 6  
*see also* categorical; continuous; discrete metric;  
 nominal; ordinal data

variation 271–3

visual analogue scale (VAS) 11

Wald statistic 298

weighted kappa 263, 264

Whitney rank sums method 187–8, 189

Wilcoxon signed-rank test 173, 189–90, 190, 209, 218

Yahr scale scores 22, 22

Yate's correction 233

*z* distribution 298