

Index

Note: Page numbers in **bold** type refer to figures; those in *italic* refer to tables or boxed material.

A

absolute risk difference (ARD), 26
absolute risk reduction (ARR), 26
absolute risks, 28
accuracy, 42
alternative hypothesis, 65

B

bar charts, 8–9, 10
 bar chart/histogram distinction, 9
Bayes factors, 69
Bayes' Theorem, 68
binary data, 24–35
 common questions, 35–6
binary variable, 2
binary variables, summarizing
 relationship between
 examples
 hay fever and eczema, 30–1
 post-hemorrhagic ventricular
 dilatation (PHVD), 28, 29
 number needed to harm (NNH),
 26–7, 35
 number needed to treat (NNT),
 26–7, 35
odds ratios *see also* odds ratios
 (OR)
 and case–control studies, 31–3
 paired alternatives, 33–5
 relative risks *versus*, 29–31
one variable, 24–5
risks, 29–31 *see also* risks
two variables, 25–9

bin widths, 10
blocked randomization, 42
Bonferroni correction, 92, 149
bootstrap, 56
box-whisker plot, 7, 8, 10

C

case–control study, 31–3, 145–6
categorical variables, 9
 binary, 1, 2 *see also* binary
 variables
 nominal, 1, 2
 ordinal, 1, 2
censored observation, 133, 134,
 136, 141
central limit theorem, 43–4
CHI, 155, 167
chi-squared (χ^2) test, 86–99
 common questions, 98
 comparing proportions, 90–2
 extensions, 98
 formula appreciation, 98–9
 fourfold tables, 89–90
 McNemar's test *see* McNemar's
 test
observed/theoretical distribution
 comparison, 95–6
in OpenEpi, 159–61, **160**, **161**, **162**
procedure, 87–9
reading/reporting, 99
splitting, 92–3
trends, 93–5
Yates' correction, 90

- cluster study, 144
 - coefficient of variation (CV%), 21
 - cohort study, 145
 - computer software *see* software
 - confidence interval (CI), 50–2, 56
 - for an odds ratio, 54–5
 - bootstrap, 56
 - common questions, 56–8
 - for a difference in proportions or percentages, 54
 - mean
 - large sample differences, 52–3
 - small sample, 73–4
 - percentages or proportions, 53–4
 - reading and reporting, 57
 - for a relative risk, 56
 - confirmatory hypotheses, 148
 - continuous variable, 1, 1, 2
 - correlation, 119–25
 - causality, 123
 - coefficient, 119
 - calculation, 121–3
 - common questions, 129–30
 - formula appreciation, 131
 - reading/reporting, 131
 - scatter diagrams, 119–21
 - significance test, 123–4
 - Spearman's rank, 124–5
 - Cox regression, 140
 - cross-sectional study, 146
- D
- data
 - binary, 24–35
 - grouped, standard deviation
 - from, 18, 18–19
 - paired, 13, 97–8, 149
 - transformation, 19, 19–20, 20
 - geometric mean, 20
 - types, 1–3
 - categorical variables, 1, 2
 - cut-off points, 2
 - quantitative variables, 1, 1–2
 - ungrouped, standard deviation
 - from, 16–17, 17
 - data display
 - bar charts, 8–9, 10
 - box-whisker plot, 7, 8
 - common questions, 9–10
 - dynamite plunger plot, 9
 - histograms, 8, 8, 9
 - bar chart/histogram distinction, 9
 - in papers, 10–11
 - stem and leaf plots, 3–4, 4
 - data dredging, 70
 - data summary, 1–11
 - common questions, 9–10
 - data types, 1, 1–3
 - median, 4–5, 5
 - range, 4
 - variation measurement, 5–6
 - see also* data display
 - dependent variable, 120, 121, 123
 - design, 143–6
 - matched parallel, 144
 - parallel group, 144
 - quasi-experimental, 145
 - reading/reporting, 145
 - see also*, study
 - diagnostic tests, 102–9
 - diagnosis/screening, 107–8
 - likelihood ratio, 106
 - in OpenEpi, 161, 162, 163
 - prevalence, 103, 105, 105, 106, 107
 - reading/reporting, 109
 - ROC curve, 107, 108
 - sensitivity, 103–4, 104, 105, 106, 107, 108
 - specificity, 104, 104, 105, 106, 107, 108
 - standard table, 102
 - difference
 - ranked *see* rank score tests
 - t* tests, 72–3
 - distribution
 - normal *see* normal distribution
 - skewed *see* skewed distribution
 - dot plots, 7, 20
 - double blinded study, 144
 - dynamite plunger plot, 9
- E
- empirical normal range, 50
 - equation, regression, 125–9
 - error
 - type I, 60–1
 - type II, 65–6
 - exploratory hypotheses, 148

- F**
 Fisher's exact test, 90, 159
 common questions, 98–9
 mid *P*-value, 90, 98
 reading and reporting, 99
 fourfold tables, chi-squared (χ) test, 89–90
- G**
 Gaussian distribution *see* normal distribution
 geometric mean, 20
- H**
 histograms, 8
 bar chart/histogram distinction, 9
 bin widths, 10
 group requirement in a, 10
 hypothesis
 alternative, 65
 confirmatory, 148
 null, 53, 60–1
 study, 65
- I**
 incidence rate, 25
 incidence rate ratio (IRR), 27
 independent variable, 120, 121
 inference, 68–9
 interquartile range (IQR), 5
 data display in papers, 10
- K**
 Kaplan–Meier survival curve, 133–4
- L**
 least squares estimate, 127
 likelihood ratio of a diagnostic test, 106
 Likert scale, 2
 logged data, 19, 20, **20**, 79
 log transform, 20
 logistic regression, 98
 log rank test, 137–40
- M**
 Mann–Whitney *U* test, 113–16
 comparing means/medians, 116–17
 in OpenStat, 162, 163, **165**
 in R, 163, **165**
 and *t* test, 117
 Mantel–Haenszel method, 94
 matched parallel design, 144
 McNemar's test, 97–8
 in OpenEpi, 162
- mean**
 confidence interval for, 72–4
 definition, 13
 standard error of, 52
 and standard deviation, 13–15
 standard error of, 43–5
 measured variables, 1
 measurement error, 21
 measure of location *see* mean, median
 median
 comparison by non-parametric tests, 116–17
 data summary, 4, 4–5
 midpoint *see* median
 mid *P*-value, 90, 98
 mode, 21
 multiple regression, 129
- N**
 nominal variables, 2, 66
 non-parametric test, 110, 116–17
 non-random samples
 common questions, 46–7
 problems with, 45–6
 reading and reporting
 populations and samples, 47
 normal distribution, 14–15, **15**, **180**
 curve, 14, **15**
 normal range, 50
 null hypothesis
 and alternative hypotheses, 66
 and type I error, 60–1
 number needed to harm (NNH), 26–7, 35
 number needed to treat (NNT), 26–7, 35
- O**
 odds, 25
 odds ratio (OR), 28
 and case–control studies, 31–3
 confidence interval for, 54–5
 in OpenEpi, 157, **158**
 versus relative risks, 29–31, 35

one-sided test, 67–8
 one way analysis of variance, 83
 OpenEpi, 27, **154**, 155, 167
 χ^2 test, 159–61, **160**, **161**, **162**
 diagnostic tests in, 161, 162, **163**
 odds ratio in, 157, **158**
 risks analysis in, 157, **157**
 OpenOffice Calc, 153, 167
 OpenStat, **154**, 155, 167
 comparison of two means, **158**,
 158–9, **159**
 kurtosis, 156
 Mann–Whitney *U* test, 162, 163,
165
 mean, 157
 regression data analysis, 163,
 164, 166, **166**
 skewness, 156
 survival analysis, 166, **166**
 Wilcoxon test, 162, 163, **164**
 ordinal variables, 1, 2
 outliers, 5, 83

P

paired alternatives, 33–5
 paired data, 13, 97–8, 149
 paired samples, rank score tests,
 110–13
 Wilcoxon test, 111–13
 parallel group design, 144
 Pearson's correlation coefficient,
 119–23
 percentages
 confidence interval for a
 difference in, 54
 standard error of difference
 between, 45, 53–4
 pie charts, 11
 populations
 common questions, 46–7
 defined, 39
 parameter, 39–40
 reading and reporting, 47
 samples, 40–1
 positive predictive value (PPV), 105
 power, of study, 65
 PPV *see* positive predictive value (PPV)
 precision, 42
 prevalence, 24
 diagnostic tests, 103, 105, 105,
 106, 107

probability
 Fisher's exact test, 90
P value, 63–5
 statement of, 49–50
 proportional hazards assumption,
 survival analysis, 137
 proportions
 χ^2 comparisons, 90–2
 confidence interval for a
 difference in, 54
 standard error of, 45
 standard error of difference
 between, 53–4
 prospective studies, 143
P-value, 63–5
 common questions, 66–9
 confidence intervals, and
 clinically important
 results, 64–5
 mid *P*-value, 90, 98
 one-sided *P*-value, 64, 69, 90
 reading/reporting, 69–70

Q

quantitative variables
 continuous, 1, 2
 counted, 1, 1, 2
 measured, 1, 1, 2
 quartiles, 5
 quasi-experimental design,
 145

R

randomization, 42–3, 143–4
 random sampling, 40–1, **148**
 stratified, 41
 systematic, 41
 rank score tests, 110
 common questions, 116–17
 paired samples, 110–13
 Wilcoxon test, 111–13
 reading/reporting, 117
 unpaired samples, 113–16
 Mann–Whitney *U* test,
 113–16
 rate, 24
 recall bias, 146
 receiver operating characteristics
 (ROC) curve, 107, **108**
 reference ranges, 49–50
 confidence interval, 56–7

- regression
 - analysis in OpenStat, 163, 164, 166, **166**
 - assumption testing, 128
 - common questions, 129–30
 - equation, 125–9
 - least squares estimate, 127
 - line, 127–9
 - multiple, 129
 - reading/reporting, 131
- relative risk reduction (RRR), 27
- relative risks *versus* odds ratios, 29–31
- risks
 - absolute risk difference (ARD), 26
 - absolute risk reduction (ARR), 26
 - analysis in OpenEpi, 157, **157**
 - ratio or relative risk (RR), 27
- ROC *see* receiver operating characteristics (ROC) curve
- R program, software, 155
 - Wilcoxon test, 163, **164**, **165**
- S
- sample size, 146–8
- samples, 40–6
 - accuracy, 42
 - central limit theorem, 43–4
 - convenience, 45
 - mean
 - confidence interval, 52–3, 72–4
 - differences *see t* test
 - standard error of, 43–5
 - non-random, problems, 45–6
 - paired, 110–13
 - population mean differences *see t* test
 - precision, 42
 - random, 40–1, **148**
 - reading/reporting, 47
 - response rate, 46
 - stratified random, 41
 - systematic random, 41
 - unbiased, 42
 - unpaired, 113–16
 - variation between, 43
- scatter diagrams, 119–21, **122**, **127**
- sensitivity, diagnostic tests, 103–4, 104, 105, 106, 107, **108**
- sequential study, 144
- significance test
 - correlation, 123–4
 - difference in two proportions, 53–4
 - single blind study, 144
- SISA, 155
- skewed distribution, 19
- software
 - CHI, 155
 - OpenEpi *see* OpenEpi
 - OpenOffice Calc, 153
 - OpenStat *see* OpenStat
 - R program, 155, 167
 - SISA, 155
 - Stata, 153
 - Websites, 167
 - Xycoon, 155
- Spearman's rank correlation, 124–5
- specificity, diagnostic tests, 104, 104, 105, 106, 107, **108**
- splitting, χ^2 test, 92–3
- standard deviation, 14
 - calculation of, 17
 - from count data, 18
 - from grouped data, 18–19
 - between subjects and within subjects, 21
 - from ungrouped data, 16–17
- standard error (SE)
 - common questions, 66–9
 - of difference between means, 52
 - of difference between percentages or proportions, 53–4
 - of the mean, 43–5
 - of a proportion or a percentage, 45
 - reading and reporting *P*-values, 69–70
- Stata, 153, 167
- stem and leaf plots, 3–4, 4
- stratified random sampling, 41
- Student's *t* test *see t* test
- study
 - carryover effect, 144
 - case-control, 145–6
 - cluster, 144
 - cohort, 145
 - control, 143, 144
 - cross-sectional, 146
 - double blinded, 144
 - placebo effects, 144
 - prospective, 143
 - randomization, 143–4

Study (contd)

- randomized controlled trial, 143
- reading/reporting, 151
- recall bias, 146
- sample size, 146–8
- sequential, 144
- statistical test selection, 148–51, 149, **150**
- study hypothesis, 65
- summary statistics *see also* interquartile range (IQR); median
 - for binary data from a non-matched study
 - common questions, 35–6
 - reading and displaying summary statistics, 36
- defined, 13
- reading and displaying, 23
- survival analysis
 - censored observation, 133, 134, 136, 141
 - common questions, 141
 - Kaplan–Meier survival curve, 133–4
 - long rank test, 137–40
 - in OpenStat, 166
 - proportional hazards assumption, 137
 - reading/reporting, 141
 - survival curve calculation, 134–6
- systematic random sampling, 41

T

testing

- for a difference in two proportions, 62–3
- for differences of two means, 61–2
- test selection, 148–51, 149, **150**
- trends, χ^2 test, 93–5
- t* test
 - common questions, 83–4
 - confidence interval for the mean from a small sample, 73–4
 - difference between means of paired samples (paired *t* test), 80–2
 - difference between means of two samples, 75–8
 - difference of sample mean from population mean, 74–5

- formula appreciation, 84
- and Mann-Whitney *U* test, 117
- reading and reporting, 84
- significance test of correlation, 123–4
- unequal standard deviations, 79–80
- Tukey's hinges, 6
- two sided tests, 67–8
- type I error
 - and null hypothesis, 60–1
 - relation with type II error, 66
- type II error, 65
 - relation with type I error, 66

U

- unbiased samples, 42
- unpaired samples
 - rank score tests, 113–16
 - Mann-Whitney *U* test, 113–16
 - t* test, 75–78

V

variables

- correlation/causality, 123
- dependent, 120, 121, 123
- independent, 120, 121
- input, 150
- outcome, 150–1
- test selection, 149
- variance, 16–17
- variation, coefficient of (CV%), 21
- variation measurement, 5–6
 - data summary, 5–6
 - interquartile range (IQR), 5
 - outliers, 5
 - quartiles, 5
 - samples, 43

W

- Welch test, 79
- Wilcoxon test, 111–13
 - in OpenStat, 162, 163, **164**
 - in R, 163, **164, 165**

X

- Xycoon software, 155, 167

Y

- Yates' correction, χ^2 test, 90