Index

<table>
<thead>
<tr>
<th>Abridged life tables</th>
<th>387–88</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accidents, traffic</td>
<td>27, 85, 458</td>
</tr>
<tr>
<td>Accuracy, test validity</td>
<td>449</td>
</tr>
<tr>
<td>Age, importance of standardising for</td>
<td>119–20</td>
</tr>
<tr>
<td>Aggregate measures, risk of ecological fallacy</td>
<td>90</td>
</tr>
<tr>
<td>Air pollution</td>
<td>10, 25, 46–48, 67, 87–9, 96, 164, 437, 473</td>
</tr>
<tr>
<td>Albumin administration to critically injured patients, meta-analysis</td>
<td>418–22</td>
</tr>
<tr>
<td>Analysis of variance (ANOVA)</td>
<td>504–8</td>
</tr>
<tr>
<td>ANOVA tables</td>
<td>240–1, 245, 246, 506</td>
</tr>
<tr>
<td>Antman, E.M.</td>
<td>396, 397, 409</td>
</tr>
<tr>
<td>Asthma</td>
<td></td>
</tr>
<tr>
<td>and air pollution</td>
<td>10</td>
</tr>
<tr>
<td>paired comparison examples</td>
<td>347–9, 495–501</td>
</tr>
<tr>
<td>trends in</td>
<td>44–6</td>
</tr>
<tr>
<td>Attributable risk (AR)</td>
<td>434–5</td>
</tr>
<tr>
<td>implications for prevention strategies</td>
<td>438–45</td>
</tr>
<tr>
<td>population attributable risk (PAR)</td>
<td>436–8</td>
</tr>
<tr>
<td>Bar charts</td>
<td>181, 182</td>
</tr>
<tr>
<td>Bayesian methods</td>
<td>519</td>
</tr>
<tr>
<td>application and relevance</td>
<td>520–2</td>
</tr>
<tr>
<td>Bayes’ theorem and formula</td>
<td>519–20</td>
</tr>
<tr>
<td>Begg’s test</td>
<td>412, 415</td>
</tr>
<tr>
<td>Beta coefficient</td>
<td>241–2, 291</td>
</tr>
<tr>
<td>Bias</td>
<td>134</td>
</tr>
<tr>
<td>from inaccurate records</td>
<td>271</td>
</tr>
<tr>
<td>interviewer</td>
<td>272</td>
</tr>
<tr>
<td>lead-time</td>
<td>456–7</td>
</tr>
<tr>
<td>non-response</td>
<td>160, 163</td>
</tr>
<tr>
<td>observer</td>
<td>177, 201, 251, 272</td>
</tr>
<tr>
<td>prognostic</td>
<td>456</td>
</tr>
<tr>
<td>publication</td>
<td>399–400, 410–12</td>
</tr>
<tr>
<td>recall</td>
<td>271</td>
</tr>
<tr>
<td>response</td>
<td>144, 162</td>
</tr>
<tr>
<td>screening programmes</td>
<td>446–57</td>
</tr>
<tr>
<td>selection</td>
<td>134, 177, 394</td>
</tr>
<tr>
<td>see also measurement, error</td>
<td></td>
</tr>
<tr>
<td>Binomial distribution</td>
<td>478–9</td>
</tr>
<tr>
<td>Blinding</td>
<td>317, 340–1, 353–7</td>
</tr>
<tr>
<td>Bradford Hill guidelines</td>
<td>226–8</td>
</tr>
<tr>
<td>British Regional Heart Study (BRHS)</td>
<td>194, 196, 199–203</td>
</tr>
<tr>
<td>ascertainment of cancer cases</td>
<td>201</td>
</tr>
<tr>
<td>confounding factors</td>
<td>228–9</td>
</tr>
<tr>
<td>follow-up procedures</td>
<td>202–5</td>
</tr>
<tr>
<td>hypothesis tests</td>
<td>217–22</td>
</tr>
<tr>
<td>measurement instrument</td>
<td>199, 200–1</td>
</tr>
<tr>
<td>multiple regression in</td>
<td>243–4</td>
</tr>
<tr>
<td>passive smoking and CHD</td>
<td>364, 366–9</td>
</tr>
<tr>
<td>relative risk</td>
<td>206–8</td>
</tr>
<tr>
<td>sampling procedure</td>
<td>196–9</td>
</tr>
<tr>
<td>Caffeine and spontaneous abortion, case-control study</td>
<td>259–62</td>
</tr>
<tr>
<td>data collection</td>
<td>260, 263–4</td>
</tr>
<tr>
<td>interpretation of results</td>
<td>295–7</td>
</tr>
<tr>
<td>selection of cases</td>
<td>265–6</td>
</tr>
<tr>
<td>selection of controls</td>
<td>266–7</td>
</tr>
<tr>
<td>study structure</td>
<td>260–2</td>
</tr>
<tr>
<td>Cancer</td>
<td></td>
</tr>
<tr>
<td>breast cancer</td>
<td>19, 89–90</td>
</tr>
<tr>
<td>cervical cancer screening</td>
<td>446–7</td>
</tr>
<tr>
<td>registration of cancers</td>
<td>79</td>
</tr>
<tr>
<td>see also BRHS study; Lung cancer</td>
<td></td>
</tr>
<tr>
<td>Case ascertainment</td>
<td>97, 201–2, 205, 265–6</td>
</tr>
<tr>
<td>Case-control studies</td>
<td>96, 257–305</td>
</tr>
<tr>
<td>analysis approach</td>
<td>262</td>
</tr>
<tr>
<td>applications of design</td>
<td>264–5</td>
</tr>
<tr>
<td>case definition and selection</td>
<td>265–6</td>
</tr>
<tr>
<td>confounding and logistic regression</td>
<td>284–97</td>
</tr>
</tbody>
</table>

Quantitative Methods for Health Research  Nigel Bruce, Daniel Pope and Debbi Stanistreet
© 2008 John Wiley & Sons, Ltd
Case-control studies (Continued)
- control group selection 266–70
- exposure assessment 270–1
- hypothesis tests 279–81
- matched analysis 277–9
- matching to avoid confounding 268–70
- multiple controls 270
- odds ratio (OR) 273–7
- purpose of 259–65
- sample size calculation 281–4
- structure of 260–2
- study design 265–81
- time perspectives 263–4
- unmatched analysis 273–81

Categorical data 180, 208–9
- hypothesis testing for 208–17, 510–17
- ordered 180, 471–2, 478, 487, 491, 499, 504, 510

Causal pathway 228–9, 231

Censoring, survival analysis 366–9

Census data 80–1, 86

Chance variation 25, 97, 116–17, 121–3, 175

Chi-squared distribution 213–14, 476, 478

Chi-squared test 208–17
- chi-squared test for trend 511–13
- Q statistic 413–14

Cholesterol levels
- and coronary heart disease 441–5
- screening tests 448–50

CI, see Confidence interval

Clofibrate controlled trial, implications for prevention 465

Closed questions 165–6

Cluster randomised trials 342
- analysis of 344
- ‘green prescription’ intervention 342–3
- sample size calculation 343–4

Cluster sampling 138–9

Cochrane, Archie 426

Cochrane Collaboration 426–8

Cochrane Controlled Trials Register (CCTR) 428

Cochrane Database of Systematic Reviews 427–28

Cochran’s Q statistic 413

Coefficient of determination 77–8

Cohort effects 457–8, 461–3

Cohort life tables 385

Cohort studies 96, 193
- confounding 226–31
- follow-up 202–5

measurement 199–202
- multiple linear regression 243–8
- obtaining the sample 196–9
- purpose of 193–4
- relative risk 206–8
- result presentation and analysis 205–22
- chi-squared test 208–17
- relative risk 206–8
- t-test 217–22
- z-test 217–22

sample size issues 223–6
- simple linear regression 232–42
- structure of 196
- time perspectives 263–4

Collaborative review groups, Cochrane 427

Communicable disease 40–4

Community-based interventions 344–7

Complex intervention trials, testing 327–30

Compliance with treatment 320, 336, 338

Conditional logistic regression 295

Confidence interval (CI)
- calculating 95 per cent 116–17
- for the odds ratio (OR) 275–6
- for population mean 145–50
- for population proportion 157–60
- for regression coefficients 247

Confidential Enquiries, mortality data 82

Confounding
- Bradford Hill guidelines 226–8
- confounding factors 228–9, 292
- effect of randomisation 315
- methods for dealing with 231

Construct validity 174

Content validity 173

Contingency table 209, 214

Continuous data 179
- calculating frequency distribution of 49–51
- displaying relative frequency distribution of 65–9
- hypothesis testing for 217–22, 490–5
- methods for presenting and summarising 48–69

Continuous probability distributions 476–78

Controls/control group
- case-control comparison 262
- identifying 261, 263
- multiple 270, 278–9
- random allocation of 315–16
- selecting 266–8

Convenience sampling 139
Coronary heart disease (CHD)  
and cholesterol levels 440–5  
and passive smoking 364–9, 373–6, 383–4
Coroner’s inquests 37
Correction factor 148
Correlation coefficient 74–8, 232
Cox regression 379  
applications of 383–5  
hazard function 380  
hazard ratio, calculating 380  
model 380–1  
prediction 383  
proportional hazards assumption 380
Criterion validity 173
Cross-over trials 349–50
Crude death rates 112–13
Cumulative incidence, see Incidence rate
Current life tables 385–8

Data  
describing, displaying and presenting 48–78, 180–3  
summarising 67–8, 180  
types of 179–83  
see also Categorical data; Continuous data;  
Routine data sources
Database of Abstracts of Reviews of Effectiveness (DARE) 428
Data Safety Monitoring Board (DSMB) 326
Deaths, see Mortality
Deduction/deductive reasoning 4
Degrees of freedom 62, 214–15
Demographic data 79–80  
births, registration of 80  
deaths, registration of 31, 80  
national population census 80  
population estimates/projections 81–2
Dependent variable, see Outcome variable
Deprivation indices 86–7
Description of groups 8, 24
Descriptive epidemiology 38–46  
definition of 38  
ecological studies 89  
London smogs of 1950s 87–8  
see also Epidemiology
Diarrhoea reduction in developing countries, systematic review 395, 398–9, 402, 429
Direct standardisation 114–19
Discrete (interval) data 179–80, 181–3  
Discrete probability distributions 478–82
Disease frequency, measuring  
incidence 14–17  
prevalence 13–14
Disease notification 40–4, 85
Disease registers 84–5
Disease surveillance 84
Distribution-free tests 487–8
Distributions  
binary 478–9  
chi-squared 478  
F-distribution 478  
frequency 49–51  
Normal 51, 149, 152–3, 477  
Poisson 479–80  
probability  
continuous 476–7  
discrete 478–80  
relative frequency 65–9  
sampling 145–6  
skewed 485–7  
t-distribution 478
Dummy variables 288

Ecological fallacy 90
Ecological studies 89
Effectiveness and Efficiency (Cochrane) 426
Egger, M. 422, 423, 425
Egger’s test 411, 422, 423, 425
Eligibility (inclusion) criteria, defining 313, 328, 397
Environment, information on 46–8
Epidemiology  
definition of 6–7  
functions of 7  
scientific approach to 2–9  
study designs, overview of 91–2  
see also Descriptive epidemiology
Epi Info (StatCalc)  
output from 224–6  
sample size calculation 282–4
Error, sources of 176–8, 199, 250–1, 271
Ethical issues, intervention studies 318
European Standard Population 121–2, 127–8
Exclusion criteria, defining 313, 328, 352, 357, 399
Experimental research, see Randomised, controlled trials
Explanatory variable 232, 290  
association with outcome variable 235–6  
see also Regression methods
Exponentiation 484–5
Exposure assessment 270–1
  bias in 271–3

F-distribution 478
F-ratio 239, 242
Face validity 173
Factorial design 335–6
Fagan’s nomogram 520–1
Falls prevention trial
  analysis and interpretation 335–50
  factorial design 335–6
  sample size calculation 331–2
Finite population correction factor 148
Fisher’s exact test 513–15
Fixed-effect model, pooled estimate 414
Follow-up
  cohort studies 202–5
  loss to follow-up 320, 366
  randomised, controlled trials 311
Forest plots, meta-analysis 415–17, 421
Frequency distributions
  calculating 49–51
  describing
    location 55
    shape 51–5
    spread 60–2, 63–4
  displaying data 48–9
  for categorical data 180–1
  see also Relative frequency distributions
Frequency matching 269
Funnel plots
  asymmetry, methods for detecting 411–12
  graphical presentation 410–11

General Household Survey (GHS) 85
General practice (GP)
  illness seen in 44–6
  morbidity data 82–3
General Practitioner Research Database (GPRD) 84
Geometric mean 485
Goodness of fit, regression model 237–9, 290–1
Green, J. 5
‘Green prescription’ intervention, cluster randomisation 342–3
Guide table for selecting hypothesis test 516–18
Hard-to-reach people, sampling method for 185
Hazard function 380
Hazard ratio 380, 381
Health event data 82–4
  deaths from specific causes 82
  general practice/primary care 84
  hospital episode statistics (HES) 83
Health information
  demographic 79–81
  describing health events 82–4
  population-based 85–6
  routine collection of 30–7
Health Survey for England (HSE) 85–6
Heterogeneity, testing for 412–14
Histograms
  describing shape of distribution 51–5
  frequency distributions 49
  relative frequency distributions 65–9
Hospital episode statistics (HES) 82–3
Hypothesis testing 500–2
  association between categorical variables 510–15
  chi-squared test 208–17
  chi-squared test for trend 511–13
  Fisher’s exact test 513–15
  McNemar’s test 280–1, 348
  association between two groups 501–4
  Pearson (product moment) correlation 77–8
  Spearman’s rank correlation 502–4
  choosing appropriate test 515–19
  guide table, using 516–18
  multiple significance testing 518–19
  comparing more than two groups 504–10
    analysis of variance (ANOVA) 504–8
    Kruskal–Wallis test 508–10
  comparing two independent groups 490–5
    Mann–Whitney U test 491–5
    t-test 217–22
    z-test 217–22
  comparing two paired (matched) groups 495–501
    paired t-test 495–9
    Wilcoxon signed rank test 499–501
  for heterogeneity in study results 412–14
  robustness of a test 483
  stages of 490
Hypothetico-deductive method 4
I^2 statistic 413
Ideas, development of 11–12
Illness, see Morbidity
IMR (infant mortality rate) 36–8, 386–7
Incidence density 16–17, 205–6
Incidence rate
and attributable risk (AR) 435–8
calculating 15–16, 363
duration of illness and prevalence 17–18
incidence density 16–18, 205–6
and person-time 16–17
and relative risk 206–8
Inclusion (eligibility) criteria 318, 328–9, 399
Independent variable, see Explanatory variable
Index population 120, 123
Indirect standardisation 114–19, 120, 122–3
Individual matching 269
Induction/inductive reasoning 4
Infectious diseases, notification of 40–4
Inference 8, 179, 209, 489, 519
Informed consent 318
Instrument error 172–4, 176–8, 200
Intention to treat approach 324
Internal consistency 170
Interpretative approaches 5–6
Interquartile range (IQR) 60–2
Interval scales 179
Intervention studies/trials 96, 307–61
analysis
checking success of randomisation 318–20
compliance with treatment allocation 320–3
conclusions, drawing 325
effect of intervention 324
by intention to treat 323–4
interim analysis 326
loss to follow-up 320
safety of trial 326
complex interventions, testing 327–30
compliance with treatment 320–3, 336, 338
ethical issues 318
factorial design 335–6
paired data 347–9
purpose of 309–12
sample size calculation 331–4
study design
blinding 317
defining intervention 313–14
eligibility and exclusion criteria 313
outcome assessment 316–17
placebo control 314–15
randomisation 315–16
structure of 310–12
Interviews
advantages/disadvantages of 165
checking success of 170–2
development of questionnaire 168–70
ideal length of 166
interviewer bias 165, 272
Jadad, A.R. 404
Kaplan–Meier survival curves 369–72
Körner data 83
Kruskal-Wallis test 508–10
Kuhn, Thomas 6
‘Ladder of powers’, transformations 485, 486
Lead-time bias 456–7
Least squares regression 234–5
Length-based sampling (prognostic) bias 456
Life expectancy, calculation of 385–7
Life tables 385–8
Likelihood ratio 450–2, 519–20
Linear regression
multiple 243–8
simple 232–42
Line of best fit 233–5, 237
Line charts 181–2, 183
Literature searches 393–5
Location of distribution, measuring 55
Logistic regression 286–8
dummy variables 288
multiple logistic regression 289–95
Log-rank test 372
Log transformation 483–4
London smogs of 1950s 87–9
Longitudinal studies, see Cohort studies
Losses in sampling 144–5
Low back pain and work environment study 389–95
Lung cancer
cohort effects 457–8
and confounding in physical activity study 230–1
RCT survival study 376–9
variations in death rates 36–7
Mann-Whitney U test 491–5
Matching, case-control studies 268–70
simple matched analysis
McNemar’s test 280–1
more than one control per case 278–9
one control per case 277–8
McNemar’s test 280–1, 348
Mean 57, 58, 63–74
case-control studies 267–9
geometric mean 485
Mean (Continued)
  population mean 145–52
  sample mean 143–52
Mean squares (MS) 239, 240, 245, 506
Measurement
  accuracy of 164
  baseline 318–19, 329
  checking success of 170–2
  error
    avoiding/reducing 199, 200–1
    sources of 176–8
    see also Bias
  interviews 164–5
  quality, assessing 172–6
  questionnaire design 165–8
  scales 179
  self-completed questionnaires 164–5
Median 56–7, 58–9, 63–4
Melanoma and sunlight exposure study 425
Meningitis, seasonal and age patterns 42–4
Meta-analysis 393
  historical example 418–22
  methodology 409–21
    hypothesis test for statistical significance 413–14
    pooled estimate calculation 414–15
    publication bias, funnel plots 410–12
    result presentation, forest plots 415–17
    sensitivity analysis 417–18, 425–6
  of observational studies 422–6
Methodological quality, assessing 403–5
Mode 55
Morbidity 39–40
  data, information sources for 40, 81–4
  follow-up of cases 202–5
  seen in general practice 44–6
  see also Survival analysis
Mortality
  Confidential Enquiries 82
  coroner’s inquests 34
  infant mortality rate (IMR) 36–8
  life tables 385–8
  statistics, routine collection of 30–7, 79
  suicide rates 34–6, 458–63
  see also Standardised mortality ratio
Multiple linear regression 243–8
Multiple logistic regression 289–95
Multiple significance testing 518–19
Multistage sampling 139
Multivariate logistic regression analysis 289–95
National Morbidity Studies 40, 44–6
National population census 80
National Survey of Sexual Attitudes and Lifestyles (NATSAL) 129–89
Natural experiments 93
Nelemans, P. J. 425
NHS Economic Evaluation Database (EED) 428
Nominal scales 179
Non-linear relationships 71–3
Non-parametric hypothesis testing 487–8
  Kruskal-Wallis test 508–10
  Mann–Whitney U test 491–5
  Spearman’s rank correlation 502–4
  Wilcoxon signed rank test 491, 495, 499–501
Non-randomised intervention study designs 347
Non-response bias 160, 163
Normal distribution 149–50, 152–3, 489
Null hypothesis 210–13, 218–19, 236, 489
  see also Hypothesis testing
Number needed to treat (NNT) 339
Observational studies
  meta-analysis 422–6
    approach to 422–3
    methods of 423–6
  systematic reviews 422
    stages of 424
  see also Case-control studies; Cohort studies
Observer error (bias) 175, 200–1, 251, 272
Odds ratio (OR) 273–7
  95 per cent CI for 275–6
  calculation of
    more than one control per case 278–9
    simple matched analysis 277–9
    simple unmatched analysis 279–80
  and relative risk (RR) 276
Office of National Statistics (ONS) Longitudinal Study 86
Omnibus Tests of Model Coefficients (OTMC) 290
One-sided tests 333, 524
Open questions 165–6
Ordered categorical data 180, 488, 491, 502, 504, 510–15
OR, see Odds ratio
Outcome assessment 316–17
Outcome variable 232–3
  and explanatory variable 232–3
  see also Regression methods
Outliers 54–5, 58
Over-sampling 134–5, 137, 142
INDEX

P-value 214–15, 219, 221–2, 226, 236–7, 489–90
Paired data, comparison of 347–9, 495–504
Paired t-test 348, 495–9
Parametric hypothesis tests 476, 482
t-test 217–22, 491
z-test 217–22, 491
Passive smoking and CHD study 367–9, 373–6, 383–5
Pearson correlation coefficient 77
Period effects 460, 462
Person-time 16–17
Person, variation of health by 38, 92
Physical activity
and cancer, cohort study 196–206, 230–1, 247–8
GP advice trial 340–1
Pie charts 181
Placebo control 310–11, 314–15, 340–1
Placebo effect 314–15
Place, variation of health by 38, 92
Poisson distribution 479–82
Pollution, air
asthma 10
and health 46
London smogs of 1950s 87–9
routinely available data on 46–8
Pooled estimate, meta-analysis 414–15
Pooled standard deviation 221
Popper, Karl 4, 9
Population attributable risk (PAR) 436–7
Population-based health data 85–6
Population, definition of 133–4
Population estimates and projections 81–2
Population mean
calculating 95 per cent CI 157–60
estimating from random sample 148–9
estimating sample size 155–7
and increasing sample size 147–52
Population proportion
sample size for estimating 158–60
standard error and 95 per cent CI 157–8
Positivism 3–4, 5
Posterior probability 519–22
Power, statistical 223, 249, 262, 270, 282, 422
Precision
of sample estimates and sample size 155–60
of SMR estimate, increasing 113, 117–18, 119, 126
Prediction 235, 236, 237
improving, F-ratio 239, 290
survival analysis 364–79, 380
Predictive validity 173
Predictive value, test validity 450
Prevalence rates, measuring 13–14, 17–18
Prevention
applying both high-risk and population approaches 441–5
concepts of 18–20
high-risk approach 441–2
population (mass) approach 442–3
and safety 444–5
Primary care morbidity data 83–4
Primary prevention 20
Prior probability 519–21
Probability 131–2, 473
Probability density 475
Probability distributions 485–7
continuous 476–7
chi-squared distribution 478
F-distribution 478
normal distribution 477
t-distribution 478
discrete 478–80
binomial distribution 478–9
Poisson distribution 479–80
implications for statistical methods 480–2
Product moment correlation coefficient 77–8
Prognostic bias 456
Proportional hazards regression, see Cox regression
Prospective studies, see Cohort studies
Psychosocial work environment study 289–90
Publication bias
meta-analysis 409–12
systematic reviews 395–409
Q statistic 413
Qualitative versus quantitative research methods 5–6
Quasi-experimental comparison studies 347–9
Questionnaire design
development process 168–70
layout and length of questionnaire 166–7
open and closed questions 165–6
ordering of questions 167
phrasing and structure of questions 167–8
skip questions 168
Quota sampling 140–3
R² value, measuring goodness of fit 237
Radix, figure for current life tables 385
Random-effects model, pooled estimate 414–15
Random error 116, 117, 176, 199, 350
Randomisation
  checking success of 318–20, 336
  purpose and method of 315–16
Randomised, controlled trials (RCTs)
  cluster randomisation 341–4
  Cochrane Collaboration 426–8
  community-based 344–7
  cross-over trials 349–50
  departure from ideal 340–1
  and placebo controls 314–15
  structure of 310–12
  survival analysis 376–9
  see also Systematic reviews
Random sampling 135, 136–7
Random variation 25, 95, 99, 116–17, 175
Rank ordering 487–8
Rates, incidence and prevalence 12–18
Ratio scales 179
Real versus chance association 175, 208–9, 223, 241–2
Reasoning, scientific methods of 2–3
Recall bias 271, 301
Receiver-operator characteristic (ROC) curve 452–5
Regression line 235–8
Regression methods
  Cox 379–85
  least squares 234–5
  linear 232–42
  logistic 284–97
Relative frequency distributions 65–9, 181–2
Relative risk (RR) 206–8, 273, 287
  and attributable risk (AR) 434–8
  hazard ratio 380
Reliability/repeatability 174–5
Representativeness of sample, assessing 161–2
Reproducibility 174–5
Research
  aim(s) and objectives, defining 130–2
  literature searches 400–3
  political context 132–3
  question, formulating well-defined 9–12, 130–2
  scientific approaches to 2–9
  see also Systematic reviews
Residuals 212–13
  residual mean squares (MS) 239, 245
  sum of squared residuals (SSR) 237
Response bias 144, 162
Response rate
determining 160–1
maximising 162–3
Retrospective data collection 263–4
Risk
  change in over time 457–8
  cohort effects 457–8, 461–3
  period effects 458, 460–1
  distribution in a population 438–41
  see also Attributable risk (AR); Relative risk (RR)
Robustness of an hypothesis test 483
ROC curve, see Receiver-operator characteristic (ROC) curve
Routine data sources
demographic data 79–81
deprivation indices 86–7
health (disease) events data 82–4
population-based health data 85–6
secondary data sets 86
strengths and weaknesses of 79
Sample, definition of 133–5
Sample mean
  confidence interval (CI) 153–5
  sampling distribution 145–6
  standard error of 146–52
Sample size
  assumptions 159
  for case-control studies 281–4
  for a cohort study 223–6
  estimating 155, 334–5
  population mean 156–7
  population proportion 158–60
  intervention trials
    categorical data outcomes 331–2
    cluster trial 343–4
    continuous data outcomes 333–4
    type 1 and type 2 errors 223
Sampling distributions 145–7, 149–53
Sampling error 145–60
Sampling frame 143–5
Sampling methods 133–43
  cluster 138–9
  convenience 139
  hard-to-reach people 139–40
  multistage 139
  quota 140–3
  simple random 135, 136–7
  stratified 137–8
  systematic 139
Suicide rates 34–6, 82
  trends in the UK 458–63
‘Summarised’ data 65
Sums of squares (SS) 237–8
Surveys 92, 129–91
  confidence interval (CI) 153–55
  measurement 164–78
  purpose and context 130–3
  response 160–3
  sample size 145–60
  sampling error 145–60
  sampling frame 143–5
  sampling methods 133–43
Survival analysis 363–92
  censoring 366
  Kaplan–Meier survival curves 369–72
  log-rank test 372
  purpose of 365, 366
  in randomised control trial 376–9
Symmetric distribution, see Normal distribution
Systematic error (bias) 134, 176
Systematic reviews 393, 394
  diarrhoea reduction study 395, 398, 399, 402, 405–8
  importance of conducting 395–7
  methodology 397
  assessment of methodological quality 403–5
  extracting data and describing results 405–9
  identifying relevant studies 399–403
  inclusion and exclusion criteria 399
  objectives of review, deciding on 398–9
  of observational studies 422–6
Systematic sampling 139
T-distribution 236, 478
T-test 217–22, 491, 495
T-value 217, 237
Tertiary prevention 20
Throgood, N. 5
Three-year moving averages 459–60
Thrombolytic drugs study 396–7
Time
  period effect 460–3
  survival time 364–5
  trends in food poisoning 41–2
  variation of disease by 34–7
Transformation of data 483–7
  issues in interpretation of results 485–7
  log transformation 483–4
  interpreting results 484–5
  for skewed distributions 483–7
Treatment compliance 320, 336, 338
Two-sample t-test 220–1, 478
Two-sample z-test 218, 491
Two-sided tests 333, 515
Type I and Type II errors 223, 411, 490, 518
Unconditional logistic regression 289–95
Univariate linear regression 244–5
Univariate logistic regression analysis 291–2
Unmatched analysis, case-control studies 273–7, 279–80
Using regression equation 235–6
Validity 172–4
  questionnaire design 170–2
  screening tests 448–55
Variance 62
Variations in health
  by person 38–9
  by place 38
  by time 38
Water quality interventions 395, 398, 416–17
Weighting 142–3
Wilcoxon signed rank test 491, 495, 499–501
Work stress and body weight 244–7
Z-test 217–22, 491