

Index

Note: Page numbers in *italics* relate to illustrations

- addition rule 290–2
- ambiguous questions 25
- arithmetic mean *see* mean
- average *see* mean
- background images 98, 99
- bar charts 86–8, 86, 87, 88
 - horizontal 86, 86
 - grouped 87, 87
 - scale 86–7, 87
 - stacked 87, 88
 - vertical 86, 86, 87
 - visual perspective 99–100, 99
- bias 28
- biased questions 25
- bimodal data 126
- bivariate data 203
 - correlation coefficient 207
 - scatter diagrams 205, 210
- box plots 168–72, 169, 170
 - identifying outliers 180–1
 - modified 171–2, 171
- cause-and-effect relationship 212, 213
- census 14
- central limit theorem 322
- central tendency *see* measures of central tendency
- certain events 284
- classes 50
 - boundaries 55
 - choosing 56
 - limits 54
 - lower class boundary 55
 - lower class limit 54
 - choosing classes 56
 - median 132, 134
 - mid-point 56
 - number of 56
 - upper class boundary 55
 - upper class limit 54
 - width 55
- classical approach to probability 281–2
- closed questions 24–5
- cluster sampling 19–20
- coefficient of variation 178–80
 - advantages 181
 - disadvantages 181
- combined mean 128, 130
- comparison of sampling methods 22–3
- complement rule 288–9, 288
- compound events 286
- compound questions 26
- conditional probability 292–3
 - multiplication rule 293
- confidence intervals 324, 324
 - interpreting 326
 - point estimate 324
 - for population mean 324–5, 325
- confidence levels 324
 - for population mean 325
- continuous data
 - grouped frequency distributions 50, 53–4
 - mean 128
 - median 132
- continuous random variables 310, 311
 - normal distribution 313–21
 - probability distribution 312–13, 313
- continuous variables 3
 - random *see* continuous random variables
- convenience sampling 17–18
- correlation 203–44
 - cause-and-effect relationship 212
 - coefficient 203, 205–11, 213
 - interpretation 207–8, 209
 - limitations 210–11
 - line of best fit 250–1
 - negative 203
 - positive 203
 - rank 203, 211–12
 - scatter diagrams 204–5
 - units 208–10
- cumulative frequency 48–9
 - grouped 50, 54
- cumulative relative frequency 49–50
 - grouped 50, 54

- data 1
 - bias 28
 - bimodal 126
 - collection 2, 13–42
 - collection techniques 23–7
 - multimodal 126, 135
 - primary 15
 - qualitative *see* qualitative data
 - quantitative *see* quantitative data
 - raw 44
 - representative 15, 16, 17, 18, 19, 20–1, 22, 28
 - sampling methods 15–23
 - secondary 15
- data distributions 43–83
 - choice of distribution 57
 - cumulative frequency 48–9
 - cumulative relative frequency 49–50
 - frequency distributions 44–7
 - grouped frequency distributions 50–4
 - numerical class characteristics 54–7
 - relative frequency 47–8
- data sets 3
- dependent variables 246
 - extrapolation 253
 - interpolation 253
 - interpretation 253–4, 255
 - line of best fit 250
 - linear equations 246–8
 - predicted values 252–3
 - simple linear regression model 248, 254, 257
- descriptive statistics 2
- deviation 172–4
- discrete data
 - grouped frequency distributions 50, 51–2
 - mean 128
 - median 132
- discrete random variables 310
 - probability distribution 311–12
- discrete variables 3
 - random *see* discrete random variables
- dispersion *see* measures of dispersion
- empirical approach to probability 282–3
- equally likely outcomes 281
- equation of regression model 249–50
- events 280, 281
 - certain 284
 - complement 288
 - compound 280
 - impossible 284
 - independent 287
 - conditional probability 292
 - multiplication rule 293–4
 - mutually exclusive 287, 287
 - addition rule 290–2
 - simple 280
 - sum of probabilities 285
 - Venn diagrams 286–7, 287
- Excel 3–4
 - built-in functions 6–10
 - POWER 8
 - SQRT 9
 - SUM 10
 - correlation coefficient 213–15
 - formulae, constructing basic 4–5
 - graphical representation 100–3
 - measures of central tendency 137–42
 - simple linear regression 256–7
 - statistical functions 11–12
- experiments 280, 281
- explanatory variables *see* independent variables
- extrapolation 253
- extreme values *see* outliers
- first quartile *see* lower quartile
- five-number summary 167–8
 - box plots 168–72
- frequency distributions 44–7
 - grouped *see* grouped frequency distributions
 - mean 128, 129, 136–7
 - median 132, 133
 - mode 126–7, 136
 - variance 177
- gradient 247–8, 247
 - interpretation 253–4
 - line of best fit 250, 251, 253
 - rounding errors 255
- graphical representation 85–124
 - background images 98, 99
 - bar charts 86–8, 86, 87, 88
 - histograms 93–5, 94, 95
 - pie charts 88–90, 88, 89, 90
 - presentation guidelines 99
 - scatter diagrams 97–8, 97, 98
 - stem and leaf diagrams 90–3, 91, 92, 93
 - time series plots 95–6, 95, 96
 - visual perspective 99–100, 99, 100
- grouped bar charts 87, 87
- grouped frequency distributions 50–4
 - mean 128, 129–30
 - median 132, 133–4, 137
 - mode 126, 127
 - variance 178–9
- histograms 93–5, 94, 95
 - equal class width 93–4, 94
 - unequal class width 94–5, 95
- impossible events 284
- independent events 287
 - conditional probability 292
 - multiplication rule 293–4
- independent variables 246
 - extrapolation 253
 - interpolation 253

- interpretation 253–4, 255
- line of best fit 250
- linear equations 246–8
- predicted values 252–3
- simple linear regression model 248, 254, 257
- inference *see* statistical inference
- inferential statistics 2
- interpolation 253
- interquartile range (IQR) 165, 167
 - advantages 181
 - box plots 169
 - disadvantages 181
 - five-number summary 167–8
 - limitations 172
 - outliers 171, 180
- intersection of events 289–90, 289
 - addition rule 290
 - equivalence 295
 - multiplication rule 293–4
 - for three events 294–5, 295
- interval estimates 323–4
- interviewees 23
- interviews 23
 - comparison with questionnaires 27
- IQR *see* interquartile range
- leading questions 25
- least squares method 250, 251
 - predicted values 252
- Legendre, Adrien-Marie 250
- line of best fit 249–51, 249, 251
 - predicted values 252, 252
 - rounding errors 255
- linear equations 246–8, 247
 - interpretation 254
- linear regression *see* simple linear regression
- linear relationships 254
 - correlation 212, 213
 - correlation coefficient 203, 207–8, 210–11, 210
 - rank correlation coefficient 211–12
 - scatter diagrams 204, 204, 205, 207, 208, 209, 210–11, 210
 - negative
 - correlation coefficient 207
 - scatter diagrams 204–5, 204
 - positive
 - correlation coefficient 207
 - scatter diagrams 204, 204, 205, 205
 - simple linear regression 248–9, 250–1, 253, 254
- lower class boundary 55
- lower class limit 54
 - choosing classes 56
- lower quartile 165–7
 - box plots 168–9, 169
 - modified 171, 172
 - five-number summary 167–8
 - outliers 171–2, 180
- lurking variables 212, 213
- mean 127–30
 - advantages 134
 - coefficient of variation 179
 - combined 128, 130
 - confidence intervals 326
 - deciding to use 135–6, 136
 - deviation 172–4
 - disadvantages 134
 - frequency distribution 128, 129, 136–7
 - grouped frequency distribution 128, 129–30
 - normal distribution 314–15, 314, 315
 - standard 315–6, 315
 - standardising normally distributed variables 319–20
 - population parameters 321, 324–5, 325
 - raw data 127, 129
 - sample data 179
 - sample statistics 321, 322–3
 - standard deviation 174, 177, 178
 - variance 173, 174–8
 - weighted 128, 130
- measures of central tendency 125–62
 - advantages 134
 - choice of measure 135–6, 136
 - disadvantages 134
 - five-number summary 167–8
 - mean 127–30
 - median 131–4
 - mode 126–7
- measures of dispersion 163–202
 - box plots 168–72, 168–170
 - choice of measure 181
 - coefficient of variation 178–80
 - five-number summary 167–8
 - interquartile range 165–7
 - quartiles 165–7
 - range 164–5
 - standard deviation 172–8
 - variance 173–9
- median 131–4
 - advantages 134
 - box plots 168, 169, 170
 - class 132, 134
 - deciding to use 135, 136
 - disadvantages 134
 - five-number summary 167–8
 - frequency distribution 132, 133
 - grouped frequency distribution 132, 133–4, 137
 - quartiles 165–7
 - raw data 131, 132–3, 137
- median class 132, 134
- Microsoft Excel *see* Excel
- mode 126–7
 - advantages 134
 - deciding to use 135, 136
 - disadvantages 134
 - frequency distribution 126–7, 136
 - grouped frequency distribution 126, 127
 - raw data 126

- modified box plots 171–2, 172
 - identifying outliers 180
- multimodal data 126, 135
- multiple regression models 246
- multiplication rule 293–4
 - conditional probability 292–3
- mutually exclusive events 287, 287
 - addition rule 290, 291, 292
- negative correlation 203
 - correlation coefficient 207, 208, 209
 - rank correlation coefficient 211
 - scatter diagrams 209
- non-response bias 28
- normal curve 313, 313
 - properties 314, 314, 315
 - standard *see* standard normal curve
- normal distribution 313–21, 313
 - parameters 314
 - properties 314–15, 314, 315
 - sampling distribution of $\bar{x}$ 322, 323
 - standard *see* standard normal distribution
 - standardising normally distributed variables 319–21
- numerical class characteristics 54–7
- observations 3
- open questions 24, 25
- outcomes 280, 281
- outliers
 - identifying 180
 - modified box plots 171–2, 180
- overlapping intervals 25–6
- parameters
 - normal distribution 314
 - population 321–5, 326
- Pearson, Karl 205
- Pearson product-moment correlation coefficient (PMCC)
 - see* correlation
- pie charts 88–90, 88, 89, 90
 - angle calculation 88–9
 - comparisons 89–90
 - number of segments 90
 - visual perspective 99, 100, 100
- pilot studies 29
- point estimates 323–4
- population data
 - coefficient of variation 179
 - standard deviation 174, 175
 - variance 174, 175
 - frequency distributions 178
- population parameters 321–5, 326
- populations 14
- positive correlation 203
 - correlation coefficient 207, 208, 209
 - rank correlation coefficient 211
 - scatter diagrams 209
- predicted values 252, 252
 - extrapolation 253
 - interpolation 253
- primary data 15
 - advantages 15
 - disadvantages 15
- probability 279–307
 - addition rule 290–2
 - approaches 281–4
 - classical 281–2
 - empirical 282–3
 - subjective 283–4
 - basic properties 284–5, 285
 - complement rule 288–9, 288
 - conditional 292–3
 - expressing 295
 - independent events 287
 - intersection 289–90
 - model 285–6
 - multiplication rule 293–4
 - mutually exclusive events 287
 - union 289–90
 - Venn diagrams 286–7
- probability density function 312–13
 - normal curve 313
- probability distributions 311–13
 - continuous random variables 312–13, 313
 - discrete random variables 311–12
 - sample statistics 322
- probability model 285–6
- proportional allocation 20, 22
- qualitative data 3
 - frequency distributions 45–6
 - graphical representation
 - bar charts 86
 - pie charts 88
- quantitative data 3
 - frequency distributions 44, 46–7
 - graphical representation
 - histograms 93
 - scatter diagrams 97, 204
 - stem and leaf diagrams 90
- quartiles 165–7
 - lower *see* lower quartile
 - upper *see* upper quartile
- questionnaires 23–7
 - comparison with interviews 27
- questions
 - ambiguous 25
 - biased 28
 - closed 24–5
 - compound 26
 - design of 25–7
 - leading 25
 - open 24, 25
- quota sampling 18

- random variables 310
 - continuous 310, 311
 - normal distribution 313–21
 - probability distribution 312–13, 313
 - discrete 310
 - probability distribution 311–12
- range 164–5
 - advantages 181
 - box plots 169–170
 - disadvantages 181
 - interquartile *see* interquartile range
 - limitations 172
- rank correlation 203, 211–12
- raw data 44
- regression line 249
 - see also* line of best fit
- relative frequency 47–8
 - cumulative 49–50
 - grouped 50, 54
 - grouped 50, 52
 - sum of 58
- relative frequency probability 283
- representative data 15, 16, 17, 18, 19, 20–1, 22, 28
- residuals 249–50, 249
- resistant measures 132
- respondents 24
- response bias 28
- response rate, improving 29
- response variables *see* dependent variables
- rounding errors 255
- sample mean 178–9
- sample space 280, 281
 - complement rule 288
 - probability model 285–6
 - sum of probabilities of all elements 285, 288
- sample standard deviation 174, 177, 178
- sample statistics 321–5, 326
- sample variance 174, 176–7
- samples 14
- sampling distribution 321–2
 - of $\bar{x}$ 322–3
- sampling frame 14
- sampling methods 15–23
 - cluster sampling 19–20
 - comparison 22–3
 - convenience sampling 17–18
 - quota sampling 18
 - simple random sampling 16–17
 - stratified sampling 20–2
 - systematic sampling 18–19
- scatter diagrams 97–8, 97, 98, 204–5, 204, 205
 - correlation coefficient 207, 209, 210–11, 210
 - extreme values 210
 - line of best fit 249–50, 249, 251, 251
 - linear relationships 204, 204, 205, 207, 208, 209, 210–11, 210
 - multiple 98, 98
 - simple linear regression 248, 249, 249
- second quartile *see* median
- secondary data 15
 - advantages 15
 - disadvantages 15
- simple events 280
 - sum of probabilities 285
- simple linear regression 245–78
 - dependent variables 246
 - extrapolation 253
 - independent 246
 - interpolation 253
 - interpretation 253–4, 255
 - line of best fit 249–51
 - linear equations 246–8
 - model 248–9
 - predicted values 252, 254, 255
- simple random sampling 16–17
- Spearman, Charles 211
- Spearman's rank correlation coefficient *see* rank correlation coefficient
- stacked bar charts 17, 18
- standard deviation 174–8
 - advantages 181
 - coefficient of variation 179
 - different units of measurement 178
 - disadvantages 181
 - normal distribution 314–15, 314, 315
 - standard 315–6, 315
 - standardising normally distributed variables 319–20
 - population data 174, 175
 - population parameters 321
 - sample data 173, 176, 178
 - sample statistics 321, 322, 323
- standard error 322
- standard normal curve 315, 316
 - sampling distribution of $\bar{x}$ 323
 - standard normal distribution table 316–9, 316, 317, 318, 320
- standard normal distribution 315–16
 - standardising normally distributed variables 319–21
 - table 316–19
 - confidence interval for population mean 324–5, 325
 - non-exact values 326
 - sampling distribution of $\bar{x}$ 323
 - variations 326
- statistical inference 309–40
 - continuous random variables 310–11
 - discrete 310
 - normal distribution 313–21
 - population parameters 321–5
 - probability distributions 311–13
 - random variables 310
 - sample statistics 321–5
- statistics
 - descriptive 2
 - inferential 2

- stem and leaf diagrams 90–3, 91, 92, 93
 - back-to-back 92, 92
 - decimal numbers 93
 - key 91
 - separating digits 91
 - split stems 92
- strata 20–2
- stratified sampling 20–2
- subjective approach to probability 283–4
- systematic sampling 18–19
- tables
 - standard normal distribution 341–3
- third quartile *see* upper quartile
- time series plots 95–6, 95, 96
 - background images 99
 - comparisons 96
 - key 96
 - scale 96
- union of events 289–90, 289
 - addition rule 290–1
 - equivalence 295, 295
 - for three events 294–5, 295
- upper class boundary 55
- upper class limit 54
- upper quartile 165–7
 - box plots 168, 169, 169
 - modified 171, 172
 - five-number summary 167–8
 - outliers 171, 172, 180
- variables 3
- variance 172–8
 - advantages 181
 - calculations 180–181
 - disadvantages 181
 - population data 174, 175
 - frequency distributions 178
 - sample data 173, 176–8
 - frequency distributions 191
- Venn, John 286
- Venn diagrams 286–7, 287
 - addition rule 290
 - complement rule 288
 - intersection 289
 - equivalence 295
 - for three events 295–6, 295
 - mutually exclusive events 287
 - union 289, 290
 - equivalence 295, 295
 - for three events 295–6, 295
- weighted mean 128, 130
- y-intercept 247–8, 247
 - interpretation 252–4, 255
 - line of best fit 250, 251, 253
 - rounding errors 255
- z-scores 316
 - confidence interval for population mean 324–5, 325
 - normal distribution table 316–19
 - sampling distribution of $\bar{x}$ 323
 - standard normal distribution table
 - non-exact values 326
 - variations 326
- standardising normally distributed variables 319–20