
The Essence of Statistics for Business

Second Edition

MICHAEL C. FLEMING

*Professor of Economics,
Loughborough University*

and

JOSEPH G. NELLIS

*Professor of International Management Economics,
Cranfield School of Management, Cranfield University*

FINANCIAL TIMES

Prentice Hall

An imprint of **Pearson Education**

Harlow, England · London · New York · Reading, Massachusetts · San Francisco
Toronto · Don Mills, Ontario · Sydney · Tokyo · Singapore · Hong Kong · Seoul
Taipei · Cape Town · Madrid · Mexico City · Amsterdam · Munich · Paris · Milan

Contents

<i>Preface to the first edition</i>	xi
<i>Preface to the second edition</i>	xiii
1 The essence of statistics for business	1
The use of statistics	1
Basic terms and concepts	2
Statistics for managers	3
Schematic overview of the book	4
Statistical computer packages	6
2 Summarizing data using tables and graphs	7
The essence of descriptive tables and graphs in business	7
Tables and frequency distributions	7
Graphical methods of data presentation	10
Use of MINITAB	23
Key learning points	25
Exercises	27
3 Summarizing data numerically	29
The essence of descriptive summary measures in business	29
Measures of location	29
Measures of dispersion	38
Use of MINITAB	47
Key learning points	48
Exercises	50

4 Index numbers	54
The essence of index numbers in business	54
Price index numbers	55
Quantity index numbers	59
Key learning points	63
Exercises	64
5 Probability	67
The essence of probability in business	67
Assigning probabilities	68
The rules of probability	71
Expectations	82
Permutations and combinations	82
Key learning points	85
Exercises	87
6 Probability distributions	89
The essence of probability distributions in business	89
The binomial distribution	90
The Poisson distribution	94
The normal distribution	100
Use of MINITAB	108
Key learning points	111
Exercises	112
7 Using samples to make estimates	115
The essence of statistical inference and estimation in business	115
Sampling distributions	118
Relationship between sample statistics and population parameters	119
Point estimation of the population mean	121
The size and measurement of the standard error of the mean	122
The probability of error	123
Interval estimation of the population mean μ	126
The measurement of the standard error of the mean when σ is unknown	129
Sampling and the t distribution	130
Use of MINITAB	134
Key learning points	135
Exercises	136
8 Tests of statistical hypotheses	138
The essence of statistical inference and testing hypotheses in business	138

The basic principles of hypothesis testing	139
Conducting hypothesis tests	142
Errors in hypothesis testing: Type I and Type II errors	147
Extensions of hypothesis testing	148
Use of MINITAB	155
Key learning points	157
Exercises	158
9 Tests of goodness-of-fit and independence	163
The essence of tests of goodness-of-fit and independence in business	163
Goodness-of-fit tests	164
Tests of independence	171
Use of MINITAB	175
Key learning points	176
Exercises	177
10 Simple regression and correlation analysis	179
The essence of simple regression and correlation in business	179
Simple regression analysis	181
Simple correlation analysis	190
Extensions of simple regression and correlation analysis	194
Use of MINITAB	195
Key learning points	197
Exercises	198
11 Multiple regression and correlation analysis	200
The essence of multiple regression and correlation in business	200
Multiple regression analysis	201
Multiple correlation analysis	204
Pitfalls and limitations of multiple regression analysis	207
Use of MINITAB	208
Key learning points	212
Exercises	213
12 Analyzing time series data	216
The essence of time series analysis in business	216
Measurement of trend (T)	217
Measurement of seasonal variation (S)	220
Measurement of cyclical variation (C)	224
Random variation (R)	226
Key learning points	226
Exercises	227

Appendix A Probability distribution tables	229
A1 Binomial probability distribution	229
A2 Poisson probability distribution	237
A3 Standard normal distribution	242
A4 t distribution	243
A5 χ^2 distribution	244
A6 F distribution	246
Appendix B Computer software for statistical analysis	250
Appendix C Sources of business and economic statistics	252
Appendix D References for further reading	256
Solutions to exercises	258
<i>Index</i>	265

Index

- acceptance region 140-3
- addition model 217, 227
- addition rule 72-4, 85-6
- adjusted multiple coefficient of determination 205, 213
- alternative hypothesis 142-5
- arithmetic mean
 - simple 30-1, 48
 - weighted 30-1, 48
- autocorrelation 208
- average 30-1

- bar charts 11-15, 26
- Bayes' rule 77-9, 86
- bimodal 32
- binomial distribution 112
 - cumulative 94
 - formula 92
 - mean 93
 - normal approximation to 153
 - properties 93, 112
 - tables 229-36
 - variance 93
- binomial experiment 91, 111
- business cycles 224
 - see also* cyclical variation

- causation 185
- census 3
- Central Limit Theorem 123-5, 135
- charts
 - bar 11-15
 - component bar 12

 - cumulative frequency curve 15-16, 18
 - frequency polygon 13-15
 - histogram 11-13, 16
 - line 13
 - logarithmic 20-3
 - of binomial distribution 95
 - of chi-squared distribution 165-6
 - of normal distribution 100-4
 - of Poisson distribution 99
 - of t distribution 132
 - pictogram 18-20
 - pie 18-19
- Chebyshev's inequality 131
- chi-squared (χ^2)
 - distribution 165-6, 169, 176
 - statistic 165-6, 167-8, 171, 176
 - tables 244-5
- chi-squared (χ^2) tests
 - goodness-of-fit 164-71, 176
 - independence 171-5, 177
 - normality of distributions 169-71
- classes for frequency distribution 25
 - boundaries 9-10
 - frequency 8, 26
 - interval 8, 25
 - limits 9, 26
 - open-ended 9
- coefficient of
 - correlation, r 190-1, 198 (see also correlation analysis)
 - determination, r^2 142, 148
 - multiple determination 204-5, 213

- coefficient of - *contd.*
 - variation 46-7, 50
- coefficient, regression 182-3, 201-2
- combinations 84-5, 87
- component bar chart 12-13
- compound formula 37-8, 49
- computer statistical software packages 250-1
- conditional probability 76-9, 86
- confidence
 - intervals 126-8
 - levels 128
 - limits 120
- contingency tables 77, 171-4, 177
- continuous variable 3, 9-10
- control
 - charts 153-5
 - limits 153-5
- correction factors
 - finite population 124
 - Yates 165, 172-3, 177
- correlation
 - coefficient, r 190-1
 - rank 195
- correlation analysis 197
 - coefficient of determination, r^2 192
 - correlation coefficient, r 190-1
 - multiple 204-5
 - significance of r 142-4
 - see also* regression analysis
- critical regions 143
- critical values 143, 158
- cumulative frequency 15-16, 26
- cumulative frequency distribution 26
- cyclical variation 224-5

- deciles 41-2, 49
- decision making, hypothesis testing and 138
- decision tree 82
- degrees of freedom 131, 165, 172, 203
- dependent events 76, 79, 86
- dependent variables 181, 197
- descriptive statistics 2, 29
- de-seasonalizing, in time series 220-4
- determination, coefficient of 192
- deviation
 - about the mean 42-8
 - mean 42-3
 - measures 42
 - standard 43-8

- diagrams
 - scatter 180
 - Venn 72, 74
 - see also* charts
- discrete variable 3, 9, 111
- dispersion, measures of 38-48
- distributions
 - binomial probability 90-4, 112, 229-36
 - chi-squared 165-70, 172, 244-5
 - F 206-7, 246-9
 - normal probability 100-8, 112, 242
 - Poisson probability 94-100, 112, 237-41
 - sampling 116, 118-20
 - skewed 32-3
 - standard normal 102-5, 242
 - t 130-4, 243

- errors in
 - hypothesis testing 147-8
 - in regression 207-8
 - in variables 208
- estimator 43, 121, 129
- events 68, 111
 - dependent 76-7, 79, 86
 - independent 75-6, 79, 86
 - joint 72-3
 - mutually exclusive 71-4
- expected frequencies (in chi-squared tests) 163-4, 168, 169-70, 172, 174
- experiment 91, 111
- explained variation 192
- explanatory variables 181, 201
- extrapolation 188-9

- F distribution 206-7
- F statistic 205-6
- F tests 206, 213
- factorial 84
- finite population correction factor 122, 135
- forecasting 188
- frequency
 - distribution 7, 25
 - distribution, cumulative 15-16
 - observed and expected (in chi-squared tests) 164-5, 168-9
 - polygon 13-15, 26
 - relative 8, 26
- functional form 207

- geometric mean 36-7
 goodness-of-fit tests 164-71, 176
 see also chi-squared tests
 graphs *see* charts
 grouped data 33-6
- heteroscedasticity 208
 histogram 12-13, 26
 hypothesis testing
 alternative hypothesis 142, 157
 correlation 192-4, 198, 205-7
 critical value 143, 158
 decision rules 143-4
 difference between sample means
 149-52
 means 139-47
 multiple regression 157, 200-4
 null hypothesis 142
 one-tailed tests 140-6, 157, 200-4
 proportions 152-3
 quality control 153-5
 regression 185-8, 202-4
 significance level 140, 158
 test statistic 143
 two-tailed tests 142-7, 158
 Type I error 147-8, 158
 Type II error 147-8, 158
 see also statistical inference
- independence, tests of *see* chi-squared tests
 independent events 75-6, 79, 86
 independent variables 181, 197
 index numbers
 price 55-9
 quantity 60-3
 index of output (UK) 61-3
 inference *see* statistical inference
 inferential statistics 2
 interpolation 34, 41, 188-9
 interquartile range 40-1, 49
 interval estimate 135
 interval estimation
 population mean 126-9, 135, 136
 t distribution 133-4, 136
 irregular variation, in time series 226
- joint probability 72-5, 80-1
- Kendall's coefficient of concordance
 195
- Laspeyres index
 price 56-61, 63
 quantity 59-63
 least squares, method of 181-3, 197
 line charts 11, 13, 26
 line graphs 10-11, 26
 linear regression *see* regression analysis
 linear trend 217-18, 227
 location, measures of 29-38
 logarithmic graphs 20-3, 26
- mean
 arithmetic 29-31, 34, 48
 deviation 42-3, 49
 geometric 36-8, 49
 method of least squares 181-3, 197
 MINITAB and
 binomial distribution 109
 chi-squared test of independence
 176
 confidence intervals 134
 cumulative frequency curve 25
 histograms 24
 multiple regression and correlation
 analysis 210
 normal distribution 111
 one-tailed significance tests of
 sample mean 156
 Poisson distribution 110
 quality control charts 157
 reference 250-1
 regression and correlation analysis
 196
 summary measures 47-8
 t distribution 134
 two-tailed significance tests of
 sample mean 155
 MINITAB statistics package 250-1
 mode 32, 35-6, 49
 moving averages 218-19, 221, 227
 multicollinearity 207
 multimodal 32
 multiple coefficient of determination
 204, 212
 multiple correlation analysis 204-7, 212
 multiple correlation coefficient of 205
 multiple regression analysis
 defined 200-1
 explanatory power 204
 goodness-of-fit 204-5
 pitfalls and limitations 207-8

- multiple regression analysis - *contd.*
 - statistical significance 202-4, 205-7
- multiple regression model 201, 212
- multiplication rule 74-7, 86
- multiplicative model 217, 227
- mutually exclusive events 70, 72, 74, 85

- non-parametric tests 164
- normal approximation, of binomial
 - probability distribution 153
- normal curve 100-1
- normal distribution 100-3, 105-8, 112
 - and chi-squared test for goodness-of-fit 169-71
 - as an approximation to the binomial distribution 153
 - see also* standard normal distribution
- null hypothesis *see* hypothesis testing

- ogive 15-16, 26
- operating characteristic curves 148
- outlier 211

- p-value in hypothesis testing 155, 156, 210-11
- Paasche index
 - price 56-9, 63
 - quantity 59-61
- parametric tests 164
- partial regression coefficient 202
- percentiles 41-2, 49
- permutations 83-4, 87
- pictograms 18-20, 26
- pie charts 18-19, 26
- point estimate 135
- point estimation 121
- Poisson distribution 112
 - as limiting form of binomial distribution 94, 97-8
 - cumulative 98-9
 - formula 96
 - mean 98
 - properties 98, 112
 - tables 237-41
 - variance 98
- population 3
- power curves 148
- price indexes 55-9, 63
- price relatives 55-6, 58-9, 64
- probability 67-88
 - joint 72
 - prior 77-8
- probability, determination of
 - classical (theoretical) approach 64, 85
 - relative frequency (empirical) approach 70-1, 85
 - subjective approach 71, 85
- probability distribution 89, 111
- probability distributions
 - binomial 90-4, 112
 - normal 100-8, 112
 - Poisson 94-100, 112
 - see also* chi-squared, *F* and *t* distributions
- probability distribution tables
 - binomial 229-36
 - chi-squared 244-5
 - F* distribution 246-9
 - Poisson 237-41
 - standard normal distribution 242
 - t* distribution 243
- probability, rules of
 - addition 72-4, 85-6
 - conditional 76-8
 - multiplication 74-7, 86
 - see also* Bayes' rule
- proportions 152-3

- quality control 153-5
- quantiles 42
- quantity indexes 59-63, 64
- quantity relatives 60-1
- quartiles 40-1, 49
- quartile deviation 41, 49

- R-bar squared (\bar{R}^2) 205
- random sampling 116
- random variables
 - continuous 90, 111
 - discrete 90, 111, 225-7
- random variation 226
- range 38, 40, 49
 - interdeciles 41
 - interpercentiles 41
 - interquartile range 40-1
 - semi-interquartile range 40-1
- rank correlation 195
- rates of change 20-3, 36-8
- regression analysis *see also* multiple regression analysis
 - and causation 185
 - and extrapolation 188-9
 - and forecasting 188
 - and interpolation 188-9

- best-fitting regression line 181-3
- linear regression 182
- method of least squares 181-3, 197
- regression coefficients 183
- squared residuals 183
- testing significance 185-8, 202-4, 212
 - see also* correlation analysis
- regression coefficients
 - formulae 183, 197
 - interpretation of 185, 202
 - standard error 186
 - testing significance of 185-8, 198, 202-4
- regression equation 182, 197
- rejection region
- relative frequency 8, 26
- residuals 183, 197

- sample 3
- sampling
 - and the t distribution 130-4
 - with replacement 79, 119
 - without replacement 79, 119
- sampling distributions 118-19, 135
 - of means 119-21
- sampling error 116-18
- SAS, reference 251
- scatter diagrams 180, 197
- seasonal
 - adjustment 220-4
 - index 222-3
 - variation 220-4
- semi-interquartile range 40-1, 49
- serial correlation 208
- Shewhart charts 155
- significance
 - level 140
 - test *see* hypothesis testing
- simple regression and correlation 181-95
 - see also* regression analysis
- skewness 32-3
- Spearman's rank correlation 195
- SPSS, reference 251
- standard deviation
 - population 43-6, 50, 122
 - sample 43-6, 50
- standard error
 - of b 186, 198
 - of correlation coefficient, r 193
 - of difference between two means 150, 151-2, 158
 - of the mean 122, 129-30, 135
- standard normal curve 132
- standard normal distribution 102-5, 112, 132
 - tables 242
 - see also* normal distribution
- statistics
 - defined 2-3
 - sources of information 252-5
- Statistical Analysis System (SAS) 251
- statistical estimation
 - and statistical inference 115-36
 - interval estimation 126-9
 - point estimation 121
- statistical inference
 - and hypothesis testing 138-58
 - and statistical estimation 115-36
- Statistical Package for the Social Sciences (SPSS) 251

- t distribution 130-4, 136
 - and interval estimation 133-4
 - compared with standard normal distribution 132
 - tables 243
- t test
 - and correlation coefficient, r 192-4
 - and regression coefficients 185-8, 202-4
- tests of independence *see* chi-squared tests
- tests of hypotheses *see* hypothesis testing
- time series 216, 226
- time series analysis 216-27
 - additive model 217, 227
 - components 216
 - cyclical variation 224-5, 227
 - multiplicative model 217, 227
 - random variation 226, 227
 - seasonal variation 220-4, 227
 - trend 217-20, 226
- trade cycles 224
- tree diagrams 79-81, 86
- trend *see* time series analysis
- trimmed mean 48
- Type I error 147-8, 158
- Type II error 147-8, 158

- unbiased estimator 121
- unexplained variation 192
- unimodal distribution 32
- universe 3

variable

- continuous 3, 9-10
- dependent 181
- discrete 3, 9
- independent 181
- random 90, 111

variance

- pooled samples 151-2
- population 43-6, 49
- sample 43-6, 49

variation

- coefficient of 46-7, 50
- explained 192
- unexplained 192

Venn diagrams 72, 74

warning limits 154

- weighted aggregate price index 56, 63
- weighted aggregate quantity index 59-60, 64

weighted average of price relatives 57-9, 64

weighted average of quantity relatives 60-1, 64

 χ^2 -tests *see* chi-squared

Yates' continuity correction factor 165, 172-3, 177

Z statistic 103

see also standard normal distribution