

Making History Count

A Primer in Quantitative Methods for Historians

Making History Count introduces the main quantitative methods used in historical research. The emphasis is on intuitive understanding and application of the concepts, rather than formal statistics; no knowledge of mathematics beyond simple arithmetic is required. The techniques are illustrated by applications in social, political, demographic, and economic history.

Students will learn to read and evaluate the application of the quantitative methods used in many books and articles, and to assess the historical conclusions drawn from them. They will see how quantitative techniques can open up new aspects of an enquiry, and supplement and strengthen other methods of research. This textbook will also encourage students to recognize the benefits of using quantitative methods in their own research projects.

Making History Count is written by two senior economic historians with considerable teaching experience on both sides of the Atlantic. It is designed for use as the basic text for taught undergraduate and graduate courses, and is also suitable for students working on their own. It is clearly illustrated with numerous tables, graphs, and diagrams, leading the student through the various key topics; the whole is supported by four specific historical data sets, which can be downloaded from the Cambridge University Press website at <http://uk.cambridge.org/resources/0521806631>.

CHARLES FEINSTEIN is Emeritus Professor of Economic History and Emeritus Fellow of All Souls College, Oxford. His publications include *National Income, Expenditure and Output of the United Kingdom, 1855–1965* (1972), *British Economic Growth, 1856–1973* (with R. C. O. Matthews and J. Odling-Smee, 1982), and *The European Economy between the Wars* (with Peter Temin and Gianni Toniolo, 1997).

MARK THOMAS is Associate Professor of History at the University of Virginia. His publications include *Capitalism in Context: Essays on Economic Development and Cultural Change* (ed., with John A. James, 1994), *Disintegration of the World Economy between the Wars* (ed., 1996), and *The Economic Future in Historical Perspective* (ed., with Paul David, 2002).

Cambridge University Press

0521806631 - Making History Count: A Primer in Quantitative Methods for Historians

Charles H. Feinstein and Mark Thomas

Frontmatter

[More information](#)

Making History Count

*A Primer in Quantitative Methods for
Historians*

CHARLES H. FEINSTEIN
AND MARK THOMAS



CAMBRIDGE
UNIVERSITY PRESS

Cambridge University Press

0521806631 - Making History Count: A Primer in Quantitative Methods for Historians

Charles H. Feinstein and Mark Thomas

Frontmatter

[More information](#)

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS

The Edinburgh Building, Cambridge CB2 2RU, UK

40 West 20th Street, New York NY 10011-4211, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

Ruiz de Alarcón 13, 28014 Madrid, Spain

Dock House, The Waterfront, Cape Town 8001, South Africa

<http://www.cambridge.org>

© Charles H. Feinstein and Mark Thomas 2002

This book is in copyright. Subject to statutory exception and to the provisions of relevant collective licensing agreements, no reproduction of any part may take place without the written permission of Cambridge University Press.

First published 2002

Printed in the United Kingdom at the University Press, Cambridge

Typeface Minion 10.5/14 *System* QuarkXpress™ [SE]

A catalogue record for this book is available from the British Library

Library of Congress Cataloguing in Publication data

Feinstein, C. H.

Making history count : a primer in quantitative methods for historians /

Charles H. Feinstein and Mark Thomas.

p. cm.

Includes bibliographical references and index.

ISBN 0-521-80663-1 (hc) – ISBN 0-521-00137-4 (pbk.)

1. History—Statistical methods. 2. History—Methodology. 3. Historical models.

I. Thomas, Mark, 1954- II. Title.

D16.17 .F45 2001

902'.1—dc21 2001035660

ISBN 0 521 80663 1 hardback

ISBN 0 521 00137 4 paperback

Contents

<i>List of figures</i>	<i>page</i> xiii
<i>List of tables</i>	xvii
<i>List of panels</i>	xix
<i>Preface</i>	xxi
Part I Elementary statistical analysis	
Chapter 1 Introduction	3
1.1 Aims of the book	3
1.2 The four case studies and the data sets	6
1.3 Types of measurement	7
1.3.1 Cases, variables, and values	8
1.3.2 Cross-section and time-series variables	8
1.3.3 Levels of measurement	9
1.3.4 Populations and samples	10
1.3.5 Dummy variables	11
1.4 Simple notation	12
1.5 The equation of a straight line	13
1.6 Powers and logarithms	14
1.6.1 Powers	14
1.6.2 Logarithms	16
1.6.3 Logs as proportions and ratios	17
1.7 Index numbers	19
1.8 Trends and fluctuations	21
1.8.1 Measuring the trend by a moving average	22
1.8.2 Measuring the fluctuations by the ratio to the trend	25
Chapter 2 Descriptive statistics	33
2.1 Presentation of numerical data	33
2.1.1 Frequency distributions	33

2.1.2	Bar charts and histograms	36
2.1.3	Frequency curves	40
2.2	Measures of central tendency	42
2.2.1	The mean, the median, and the mode	43
2.2.2	Weighted averages and standardized rates	44
2.2.3	Percentiles, deciles, and quartiles	46
2.3	Measures of dispersion	47
2.3.1	The range and the quartile deviation	47
2.3.2	The mean deviation, the variance, and the standard deviation	47
2.3.3	The coefficient of variation	50
2.4	Measures of central tendency and dispersion with grouped data	51
2.4.1	The median and mean with grouped data	51
2.4.2	The variance and standard deviation with grouped data	52
2.5	The shape of the distribution	53
2.5.1	Symmetrical and skewed distributions	53
2.5.2	Measures of skewness	55
2.6	The normal distribution	56
2.6.1	Properties of the normal distribution	56
2.6.2	The standard normal distribution	61
2.7	Exercises for chapter 2	66
Chapter 3	Correlation	71
3.1	The concept of correlation	71
3.1.1	Correlation is not causation	72
3.1.2	Scatter diagrams and correlation	72
3.1.3	Outliers	74
3.2	The correlation coefficient	76
3.2.1	Measuring the strength of the association	76
3.2.2	Derivation of the correlation coefficient, r	77
3.2.3	Interpreting a correlation coefficient	82
3.2.4	An illustration of the use of correlation coefficients	83
3.3	Spearman's rank correlation coefficient	86
3.4	Exercises for chapter 3	90
Chapter 4	Simple linear regression	93
4.1	The concept of regression	93
4.1.1	Explanatory and dependent variables	93
4.1.2	The questions addressed by regression	94
4.2	Fitting the regression line	95
4.2.1	How do we define a line?	95
4.2.2	What criterion should be adopted for fitting the line?	96
4.2.3	How do we find this best fitting line?	98

CONTENTS

vii

4.2.4	The regression of Y on X versus the regression of X on Y	99
4.2.5	Fitting a linear trend to a time series	100
4.3	Measuring the goodness of fit	102
4.3.1	Explained variations and unexplained variations (residuals)	102
4.3.2	The coefficient of determination, r^2	105
4.3.3	Some intuitive implications	106
4.4	Introducing the Ballantine	110
4.5	Exercises for chapter 4	112

Part II Samples and inductive statistics

Chapter 5	Standard errors and confidence intervals	117
5.1	Introduction	117
5.2	Sampling distributions	120
5.2.1	Sampling distributions and point estimates of the population mean	120
5.3	Four fundamental statistical propositions	122
5.3.1	The Central Limit Theorem	122
5.4	Theoretical probability distributions	125
5.4.1	Empirical and theoretical frequency distributions	125
5.4.2	The Z- and <i>t</i> -distributions	131
5.5	Confidence intervals for a sample mean	136
5.5.1	Interval estimates and margins of error	137
5.5.2	Confidence interval for an estimate of a population mean	137
5.6	Confidence intervals for other sample statistics	140
5.6.1	Sample proportions	141
5.6.2	Correlation coefficients	141
5.6.3	Regression coefficients	143
5.7	Exercises for chapter 5	145
Chapter 6	Hypothesis testing	149
6.1	Testing hypotheses	149
6.1.1	A simple illustration	150
6.1.2	The five stages of hypothesis testing	151
6.2	The null hypothesis	151
6.2.1	Other assumptions of the model	152
6.3	How confident do we want to be about the results?	153
6.3.1	Two types of error and the null hypothesis	153
6.3.2	Significance levels and the critical region	155
6.3.3	One- and two-tailed tests	158
6.3.4	Statistical versus historical significance	160

6.3.5	Conventional significance levels and prob-values	160
6.3.6	The critical region and the four cases of theft	162
6.4	Test statistics and their sampling distributions	163
6.4.1	The basic ideas	163
6.4.2	Z- and <i>t</i> -tests	164
6.5	Making a decision	167
6.6	Interpreting the results	170
6.7	Three illustrations of the use of <i>t</i> -tests	170
6.7.1	A <i>t</i> -test for a difference of means	170
6.7.2	A <i>t</i> -test for a correlation coefficient when the null hypothesis is zero correlation	173
6.7.3	A <i>t</i> -test for a regression coefficient	175
6.8	Exercises for chapter 6	181
Chapter 7	Non-parametric tests	185
7.1	Introduction	185
7.2	Non-parametric tests for two independent samples	187
7.2.1	The Wald–Wolfowitz runs test	188
7.2.2	The Mann–Whitney <i>U</i> -test	193
7.2.3	The Kolmogorov–Smirnov test	198
7.2.4	Criteria for choosing between alternative tests	201
7.3	The χ^2 -test	202
7.3.1	Contingency tables or cross-tabulations	203
7.3.2	The χ^2 -test for a relationship between two variables	204
7.4	The one-sample runs test of randomness	210
7.5	One-sample non-parametric tests of goodness of fit	216
7.5.1	The one-sample χ^2 -test	216
7.5.2	The one-sample Kolmogorov–Smirnov test	216
7.6	Non-parametric measures of strength of association	218
7.6.1	Four measures of association: ϕ , Cramer's V, the contingency coefficient C, and Goodman and Kruskal's tau	218
7.7	Exercises for chapter 7	224
Part III	Multiple linear regression	
Chapter 8	Multiple relationships	231
8.1	The inclusion of additional explanatory variables	231
8.2	Multiple regression	232
8.2.1	The coefficient of multiple determination, R^2	232
8.2.2	Partial regression coefficients	234
8.2.3	Controlling for the influence of a second explanatory variable	235

CONTENTS

ix

8.2.4	Understanding the impact of additional explanatory variables	238
8.2.5	The Ballantine for multiple regression	240
8.2.6	The least squares criterion in multiple regression	242
8.2.7	Standardized beta coefficients	243
8.2.8	The adjusted coefficient of multiple determination, \bar{R}^2	245
8.2.9	Interpreting the intercept term	246
8.3	Partial and multiple correlation	248
8.3.1	The interpretation and order of partial correlation coefficients	248
8.3.2	Removing the influence of the control variable	252
8.3.3	Calculation of partial correlation coefficients	253
8.3.4	The coefficient of multiple correlation, R	254
8.4	Exercises for chapter 8	255
Chapter 9	The classical linear regression model	258
9.1	Historical research and models of relationships between variables	258
9.1.1	Quantitative methods and historical judgement	260
9.2	Deviations of the dependent variable from the estimated regression line	261
9.2.1	Errors of measurement and specification	261
9.2.2	Stochastic (random) factors and the error term, e	262
9.2.3	Sampling and the regression line	263
9.2.4	Assumptions about the error term	264
9.3	A significance test for the regression as a whole	268
9.3.1	The F -test and the F -distribution	269
9.3.2	F -tests of the statistical significance of the regression as a whole	269
9.3.3	The relationship of F and R^2	271
9.4	The standard error of the estimate, SEE	272
9.4.1	Confidence interval for a prediction of the mean value of a dependent variable	273
9.4.2	Confidence interval for a prediction of a specific value of a dependent variable	276
9.5	Exercises for chapter 9	277
Chapter 10	Dummy variables and lagged values	280
10.1	Dummy variables	280
10.1.1	Dummy variables with two or more categories	280
10.1.2	Dummy variables as a change in the intercept	283
10.1.3	Additional dummy variables	285

	10.1.4	Interaction between two dummy variables	286
	10.1.5	Interaction between dummy and numerical variables	287
10.2		Lagged variables	289
	10.2.1	Lagged dependent variables	291
	10.2.2	Distributed lag models	292
10.3		Exercises for chapter 10	295
Chapter 11		Violating the assumptions of the classical linear regression model	300
11.1		The assumptions of the classical linear regression model	300
	11.1.1	The model is correctly specified	301
	11.1.2	The error term is appropriately specified	301
	11.1.3	The variables are correctly measured and appropriately specified	301
11.2		Problems of model specification	302
	11.2.1	Non-linear models	302
	11.2.2	Omitted and redundant variables	303
	11.2.3	Unstable parameter values	305
11.3		Problems of error specification	308
	11.3.1	Non-zero errors	308
	11.3.2	Heteroscedasticity	309
	11.3.3	Autocorrelation	311
	11.3.4	Outliers	316
11.4		Problems of variable specification	319
	11.4.1	Errors in variables	319
	11.4.2	Multicollinearity	321
	11.4.3	Simultaneity	324
11.5		Exercises for chapter 11	326
Part IV		Further topics in regression analysis	
Chapter 12		Non-linear models and functional forms	333
12.1		Different forms of non-linear relationships	334
12.2		Estimation of non-linear regression models	341
	12.2.1	Two groups of non-linear models	342
	12.2.2	Estimating the double logarithmic and semi-logarithmic models	345
12.3		Interpretation of non-linear relationships	347
	12.3.1	Interpretation of the regression coefficients for logarithmic models	347
	12.3.2	Elasticities	349
	12.3.3	Long-run multipliers	351

CONTENTS

xi

12.4	Fitting a trend to a time series	351
12.4.1	Fitting a log-linear trend	351
12.4.2	Fitting non-linear trends	352
12.5	Modelling a relationship	359
12.5.1	Choice of functional form	360
12.5.2	Some general principles	362
12.5.3	Implementing the principles: some examples from economic history	363
12.5.4	Implementing the principles: some examples from demographic, social, and political history	367
12.6	Model specification	372
12.7	Exercises for chapter 12	380
Chapter 13	Logit, probit, and tobit models	384
13.1	Limited dependent variables	385
13.1.1	Three cases of limited dependent variables	385
13.1.2	Why linear models are not appropriate for limited dependent variables	387
13.2	Estimating the logit model: coefficients, absolute effects, and elasticities	390
13.2.1	The logit regression model	390
13.2.2	Interpreting logit coefficients	391
13.2.3	Absolute changes: marginal effects and impact effects	399
13.2.4	Calculating the marginal and impact effects	402
13.2.5	Proportionate effects: logit elasticities	406
13.2.6	Non-linearity and the measurement of change	412
13.2.7	Goodness of fit	413
13.3	Estimating the logit model: an illustration	415
13.3.1	An illustration of the logit model	415
13.3.2	Estimating the logit model using grouped data	418
13.4	Estimating the probit model	419
13.4.1	The probit model	419
13.4.2	Interpreting the results from a probit model	420
13.5	Multi-response models	422
13.6	Censored dependent variables	423
13.6.1	A censored continuous dependent variable	423
13.6.2	The tobit model	424
13.6.3	Two-part censored models	427
13.7	Exercises for chapter 13	432

Part V Specifying and interpreting models: four case studies

Chapter 14	Case studies 1 and 2: unemployment in Britain and emigration from Ireland	437
14.1	Inter-war benefits and unemployment in Great Britain	438
14.1.1	The Benjamin and Kochin model	438
14.1.2	Parameter stability and outliers in the Benjamin and Kochin model	440
14.2	Emigration from Ireland	445
14.2.1	The time-series model of Irish emigration	446
14.2.2	The panel model of Irish emigration	451
14.2.3	The impact of the explanatory variables on changes in emigration over time	458
Chapter 15	Case studies 3 and 4: the Old Poor Law in England and leaving home in the United States, 1850–1860	463
15.1	The effects of the Old Poor Law	463
15.1.1	Relief payments, wages, and unemployment	464
15.1.2	The Old Poor Law and the birth rate	476
15.2	Leaving home in the United States, 1850–1860	480
15.2.1	Influences on leaving home	482
15.2.2	The logit model	483
Appendix to §15.1.1	Derivation of reduced-form equations from a simultaneous-equations model	491
Appendix A	The four data sets	496
A.1	The iniquitous effects of the Old Poor Law	496
A.2	What explains the scale and timing of the massive emigration from Ireland after the famine?	497
A.3	To what extent were generous unemployment benefits responsible for the high level of unemployment in inter-war Britain?	501
A.4	What influenced children's decision to leave home in the United States in the mid-nineteenth century?	502
Appendix B	Index numbers	507
B.1	Price and quantity relatives	507
B.2	Two types of index number	508
B.3	Laspeyres index numbers with fixed weights	510
B.4	The index number problem and chained indices	513
B.5	Paasche index numbers with changing (given year) weights	520
	<i>Bibliography</i>	529
	<i>Index of subjects</i>	539
	<i>Index of names</i>	545

Figures

1.1	Illustration of SPSS print-out of regression results	page 4
1.2	Graph of a straight line	
	(a) With positive slope	15
	(b) With negative slope	15
1.3	Miles of railway track added in the United States, 1831–1913, with fitted trend	21
1.4	Miles of railway track added in the United States, 1831–1913	
	(a) Original series with 11-year moving average	26
	(b) De-trended series	26
	(c) Percentage deviation from trend	26
2.1	Bar chart of number of workhouses in selected counties in 1831	36
2.2	Histograms with different class intervals for <i>per capita</i> relief payments in 311 parishes in 1831	
	(a) With 8 equal class intervals, each 6 shillings wide	38
	(b) With 23 equal class intervals, each 2 shillings wide	38
	(c) With 6 class intervals of unequal width	39
2.3	<i>Per capita</i> relief payments in 311 parishes in 1831	
	(a) Histogram and frequency polygon	41
	(b) Cumulative frequency distribution	41
2.4	Symmetrical and skewed frequency curves	
	(a) The normal curve	54
	(b) Smooth curve skewed to the right (positive skewness)	54
	(c) Smooth curve skewed to the left (negative skewness)	54
2.5	A family of normal curves	
	(a) Different means but the same standard deviation	57
	(b) The same mean but different standard deviations	57
	(c) Different means and standard deviations	57
2.6	Histogram of <i>per capita</i> relief payments in 311 parishes in 1831 with normal curve superimposed	58

2.7	Areas under the normal curve	
	(a) 1.645 standard deviations either side of the mean covers 90 per cent of the area	60
	(b) 1.96 standard deviations either side of the mean covers 95 per cent of the area	60
	(c) 2.58 standard deviations either side of the mean covers 99 per cent of the area	60
3.1	Scatter diagrams showing different strengths and directions of relationships between X and Y	73
3.2	Scatter diagram of relief and unemployment, 24 Kent parishes in 1831	75
3.3	Deviations from the mean	78
4.1	Deviations from the regression line	97
4.2	Least squares deviations from the regression line	98
4.3	Scatter diagram of relief and unemployment with regression line, 24 Kent parishes in 1831	101
4.4	Net national product, United Kingdom, 1920–1938 and linear trend	103
4.5	Explained and unexplained deviations	104
4.6	The relationship between the correlation coefficient, r , the regression coefficient, b , and the standard deviations of X and Y	
	(a) When s_x is low relative to s_y	109
	(b) When s_x is high relative to s_y	109
4.7	Ballantine for a simple regression with one explanatory variable	110
5.1	Distribution of heads for 2, 4, 10, and 20 tosses of a coin	131
5.2	The t -distribution for selected degrees of freedom	134
6.1	The critical region	
	(a) A two-tailed test at the 5 per cent level of significance	157
	(b) A one-tailed test at the 5 per cent level of significance	157
6.2	A one-tailed hypothesis test at the 5 per cent level of significance	
	(a) Rejection of the null hypothesis	169
	(b) Failure to reject the null hypothesis	169
7.1	The chi-square distribution for selected degrees of freedom	208
8.1	Scatter diagrams for hypothetical data on WELFARE and HOMELESS	
	(a) Scatter diagram and regression line for all observations	236
	(b) Scatter diagram identifying the effect of TEMP on the relationship between WELFARE and HOMELESS	236
8.2	Ballantine for a regression with two explanatory variables	527
8.3	Ballantine for a regression with two explanatory variables	
	(a) With two independent explanatory variables	527
	(b) With two explanatory variables which are not independent	527

LIST OF FIGURES

xv

8.4	Least squares deviation from the regression plane with two explanatory variables	
	(a) Three-dimensional scatter diagram for a dependent variable and two explanatory variables	244
	(b) Three-dimensional diagram with regression plane	244
9.1	The population of Y for different values of X	
	(a) Hypothetical populations	265
	(b) The form of the population of Y <i>assumed</i> in simple linear regression	265
9.2	The true (population) regression line and the estimated (sample) regression line	267
9.3	The <i>F</i> -distribution for selected degrees of freedom	272
9.4	95 per cent confidence interval for a prediction of the mean value of UNEMP	275
10.1	An intercept dummy variable	285
10.2	A slope and an intercept dummy variable	289
11.1	The marginal cost of production of a statistical textbook	303
11.2	Omitted and redundant variables	
	(a) Ballantine to show the effect of a missing variable	527
	(b) Ballantine to show the effect of a redundant variable	527
11.3	Illustration of the use of dummy variables to allow for structural changes	
	(a) Introduction of an intercept dummy variable	306
	(b) Introduction of a slope dummy variable	306
	(c) Introduction of both an intercept and a slope dummy variable	306
11.4	Heteroscedastic error terms	310
11.5	Typical patterns of autocorrelated errors	313
11.6	Autocorrelation	315
11.7	Outliers	
	(a) An outlier as a leverage point	318
	(b) A rogue outlier	318
12.1	A non-linear relationship between age and height	334
12.2	Geometric curves: $Y = aX^b$	336
12.3	Exponential curves: $Y = ab^X$	337
12.4	Reciprocal curves: $Y = a + b/X$	338
12.5	Quadratic curves: $Y = a + b_1X + b_2X^2$	
	(a) Rising and falling without turning points	339
	(b) Maximum and minimum turning points	339
12.6	Logistic curves: $Y = \frac{h}{g + ab^x}$	340

12.7	Double logarithmic transformation of a geometric model	
	(a) The original curve: $Y = aX^b$	344
	(b) The double logarithmic transformation: $\log Y = \log a + b(\log X)$	344
12.8	Semi-logarithmic transformation of an exponential model	
	(a) The original curve: $Y = ab^X$	346
	(b) The semi-logarithmic transformation: $\log Y = \log a + X(\log b)$	346
13.1	Pattern of automobile ownership, Detroit, 1919	388
13.2	Cumulative distribution function of the standardized normal distribution	389
13.3	Maximum-likelihood function	393
13.4	Logit curve of automobile ownership, Detroit, 1919	401
13.5	Flow chart for deriving logit results for the sample as a whole	410
13.6	OLS model with censored observations	
	(a) Censored from above	425
	(b) Censored from below	425
15.1	Departures of children from home by age and sex, United States, 1850–1860	481

Tables

1.1	Calculation of 11-year moving average and deviations from trend, miles of railway track added in the United States, 1831–1850	<i>page 24</i>
1.2	Calculation of a seasonally adjusted series, bankruptcy in England, 1780–1784	29
2.1	<i>Per capita</i> expenditure on relief in 24 Kent parishes in 1831 (shillings)	34
2.2	Frequency, relative frequency, and cumulative frequency of <i>per capita</i> relief payments in Kent in 1831	34
2.3	Frequency of <i>per capita</i> relief payments in 311 parishes in 1831	37
2.4	Population and migration for four Irish counties in 1881 and 1911	45
2.5	Calculation of the sample variance for two sets of farm incomes	49
2.6	Heights of a sample of 40 male students and the standardized distribution	62
2.7	Areas under the standard normal curve for selected values of <i>Z</i>	64
3.1	Calculation of the coefficient of correlation for a sample of four (imaginary) values of <i>X</i> and <i>Y</i>	81
3.2	Coefficients of correlation with the British business cycle, 1854–1913	85
3.3	Fifth-quintile male occupations rank ordered by completed fertility and female age at marriage	88
4.1	Calculation of the regression coefficient for four values of <i>X</i> and <i>Y</i>	100
4.2	Calculation of the coefficient of determination for four values of <i>X</i> and <i>Y</i>	106
5.1	Distribution of <i>t</i> for selected degrees of freedom	135
6.1	Calculation of the regression coefficient and <i>t</i> -ratio for the regression of UNEMP on BWRATIO	178
7.1	Rankings of two independent samples of students	189
7.2	The Wald–Wolfowitz runs test	192
7.3	The Mann–Whitney <i>U</i> -test and the Wilcoxon rank sums test	197
7.4	The Kolmogorov–Smirnov test	200
7.5	Cross-tabulation of patients admitted to the York Retreat, by superintendent’s regime and by treatment outcomes	205
7.6	χ^2 -test for statistical significance: calculations using data for York Retreat, classified by superintendent’s regime and treatment outcomes	210

Cambridge University Press

0521806631 - Making History Count: A Primer in Quantitative Methods for Historians

Charles H. Feinstein and Mark Thomas

Frontmatter

[More information](#)

xviii

LIST OF TABLES

7.7	Cross-tabulation of 379 parishes in England by seasonal type in 1661–1700 and 1701–1740	221
8.1	Notation for simple and multiple correlation and regression	233
8.2	Average monthly data on welfare expenditure, homelessness, temperature bands, and food costs	235
8.3	Correlation coefficients and partial correlation coefficients for five variables	
	(a) Zero-order correlation coefficients	250
	(b) Third-order partial correlation coefficients	250
8.4	Zero-order and partial correlations between church attendance, population density, and church density in England in 1851	251
12.1	Absolute and proportionate regression coefficients	349
12.2	Calculation of change in fitted trend, miles of railway track added in the United States, 1881, 1891, and 1901	359
13.1	Deriving the findings from a logit regression	411
13.2	Detroit automobile ownership, 1919: the full logit model	416
13.3	Detroit automobile ownership, 1919: censored models	428
14.1	Determinants of total Irish emigration rates, time-series data, 1877–1913	447
14.2	Determinants of Irish county emigration rates, panel data, 1881–1911	453
14.3	Decomposition of changes in Irish emigration rates, 1881–1911	459
14.4	Restricted model of Irish county emigration rates, panel data, 1881–1911	460
15.1	The agricultural labour market and the Old Poor Law, simultaneous-equations and reduced-form models, cross-section data, c. 1831	472
15.2	Determinants of birth rates, cross-section data, c. 1831	478
15.3	Impact of child allowances and other factors on the increase in the English birth rate, 1781–1821	480
15.4	The age at leaving home in the United States, 1850–1860, explaining the probability of departure, boys and girls aged 5–11 and 12–18	486
A.1	Data set for investigation of relief payments in England in 1831	498
A.2	Data set for investigation of the birth rate in southern England, c. 1826–1830	499
A.3	Data set for investigation of emigration from Ireland, 1877–1913	500
A.4	Data set for investigation of unemployment in inter-war Britain, 1920–1938	503
A.5	Data set for investigation of decisions to leave home in the United States, 1850–1860	504
B.1	Bituminous coal output in the United States, 1900–1913, original data and two quantity relatives	508
B.2	Illustrative data for construction of price and quantity index numbers	510
B.3	Price and quantity indices for the data in table B.2	511
B.4	Chained index numbers with changes in the reference year	519

Panels

1.1	The log of a variable represents proportional changes in that variable	<i>page 18</i>
1.2	Decomposition of a time series into trend, seasonal factors, and residual fluctuations	28
5.1	Sampling with and without replacement	126
5.2	Generating the normal theoretical probability curve	130
5.3	The concept of degrees of freedom	133
6.1	Calculation of the standard error of the difference between two sample means	172
7.1	The chi-square distribution	207
7.2	The concept of degrees of freedom for a cross-tabulation	211
8.1	The degrees of freedom in a regression equation	246
9.1	Assumptions of the classical linear regression model	266
9.2	The theoretical <i>F</i> -distribution	270
10.1	Koyck's transformation	294
12.1	Measurement of the compound rate of growth	354
12.2	Decomposition of a time series by means of a log-linear trend and seasonal dummy variables	356
13.1	Expected values and probability theory	386
13.2	Maximum-likelihood models	392
14.1	Pooled or panel data	454
15.1	Simultaneous-equations models	468
15.2	Reduced-form models	469
15.3	Estimation by two-stage least squares	470
B.1	Laspeyres index numbers	514
B.2	Paasche index numbers	522

Preface

We would like to express our deep appreciation to Daniel Benjamin, Levis Kochin, Timothy Hatton, Jeffrey Williamson, George Boyer, and Richard Steckel for agreeing to our use of their work for the case studies, which are a central feature of the approach we have adopted for this book. We are particularly grateful to Boyer, Hatton, and Steckel for providing us with the unpublished data that we have used throughout the text and exercises to illustrate quantitative procedures, and for allowing us to reproduce their data sets on our web site. We have also received very helpful comments from a number of our colleagues, notably James Foreman-Peck, Jane Humphries, Sharon Murphy, and Richard Smith. We owe a special debt of gratitude to Joachim Voth, who read the entire text with great care and sent us many thoughtful criticisms and suggestions. None of these scholars has any responsibility for any errors of procedure or interpretation that remain.

We particularly wish to acknowledge the contributions of generations of students who took our courses in quantitative methods in Oxford and Charlottesville. They showed that intuitive appreciation of the importance and limitations of quantification in history could be developed by undergraduates and graduates from a wide range of intellectual and disciplinary backgrounds. The approach adopted in the present text was originally designed primarily for students who had no prior knowledge of statistics, and who were in many cases initially hostile to, or intimidated by, quantitative procedures. It was, above all, their positive response, their success in applying these methods in their own work, and their helpful comments that motivated us to develop the text for a larger readership.

Our final debt is to our families, for their encouragement, forbearance, and general support during the long period over which this book has evolved.