

APPLIED
GENERAL
STATISTICS

SECOND EDITION

CROXTON & COWDEN

Applied General Statistics

Frederick E. Croxton, Ph.D.

Professor of Statistics, Columbia University

and

Dudley J. Cowden, Ph.D.

Professor of Economic Statistics

School of Commerce

University of North Carolina

Second Edition

New York

Prentice-Hall, Inc.

1955

COPYRIGHT, 1939, 1955, by
PRENTICE-HALL, INC.
70 Fifth Avenue, New York

ALL RIGHTS RESERVED. NO PART OF THIS BOOK MAY
BE REPRODUCED IN ANY FORM, BY MIMEOGRAPH OR
ANY OTHER MEANS, WITHOUT PERMISSION IN WRITING
FROM THE PUBLISHERS.

SECOND EDITION

L. C. Cat. Card No. 55-7162

PRINTED IN THE UNITED STATES OF AMERICA

Preface

Although in this edition *Applied General Statistics* has been largely rewritten, the objective is the same as that of the first edition: to describe the more commonly used statistical methods and to illustrate their applications in many fields.

The scope is essentially the same as that of the earlier edition, but the topic of statistical significance has been expanded, while the book as a whole has been shortened some 115 pages. Nearly all of the illustrative examples are new and, as before, represent real rather than hypothetical data. The order of topics has been altered, the fitting of curves to frequency distributions and the discussion of significance tests having been moved to the end of the book. A few of the symbols have been changed, for simplicity and clarity. Each chapter which uses symbols is preceded by a "symbol vocabulary" for that chapter. The forthcoming fourth edition of *Workbook in Applied General Statistics* will conform to this edition in regard to symbols used and the order of topics.

This second edition of *Applied General Statistics* was prepared by Frederick E. Croxton, except that Dudley J. Cowden provided a first draft of three chapters. The third edition of *Practical Business Statistics* will be prepared by Dudley J. Cowden.

I am indebted to Professor Sir Ronald A. Fisher, Cambridge, to Dr. Frank Yates, Rothamsted, and to Messrs. Oliver and Boyd Ltd., Edinburgh, for permission to reprint portions of Tables III and IV from their book *Statistical Tables for Biological, Agricultural, and Medical Research*. I am similarly indebted to Professor Egon S. Pearson and to the Biometrika Trustees for permission to reprint the tables or portions of tables from *Biometrika* and from E. S. Pearson and H. O. Hartley's *Biometrika Tables for Statisticians*, Volume 1, which are shown here in Appendices I, J, M, O, and P, as well as Charts 25.6 and 25.7. Other persons, and organizations, who supplied data or gave permission to reprint material are acknowledged at the appropriate location.

Many people helped with the work incidental to the revision of this textbook. Dr. James D. Paris prepared the first draft of the two chapters on index numbers, and Robert E. Lewis wrote the initial draft of the

description of the method of adjusting a seasonal index for the date of Easter. Portions of the manuscript were read by Alfred J. Kana, Associate in Statistics, Columbia University. Charles H. Wittmann helped with some phases of the research incidental to the preparation of the manuscript. Julius I. Brown, Roy C. Calogeras (now Lecturer in Statistics, Columbia University), and Donald W. Lovejoy performed computations. Keith Galli, Antony Herrey, Marcia Silfin, Mirian Weissman, and Faedon Xydis assisted in the drawing of charts, many of which were lettered by Marie Morisawa. Most of the manuscript was typed by Miss Hsi-lan Wang. To all of these who helped I express my thanks, but my particular thanks are extended to my wife, Rosetta R. Croxton, who was ready to do whatever was needed whether it was computing, drawing charts, lettering, or working with me on the index.

FREDERICK E. CROXTON

Leonia,
New Jersey

Contents

(For a short course in statistical methods, the chapters or sections which are starred in this table of contents may be omitted without destroying the continuity of treatment.)

CHAPTER

PAGE

1. Introduction	1
Statistical Data and Statistical Methods	1
Collection	2
Presentation	3
Analysis	3
Interpretation	6
A Few Improperities	6
Bias	7
Omission of important factor	7
Carelessness	8
Non-sequitur	9
Non-comparable data	9
Confusion of association and causation	9
Insufficient data	10
Unrepresentative data	10
Concealed classification	11
Failure to define units	11
Misleading totals	11
Poorly designed experiment	12
Research Methods	12
2. Statistical Data	15
Collecting Statistical Data	16
Method of collection	16
Outline of procedure	16
1. Planning the study	17
2. Devising the questions and making the schedule	18
3. Selecting the type of sample	25
4. Using the schedules to obtain the information	34
5. Editing the schedules	36
6. Organizing the data	37
7. Presentation and analysis	45
Using Existing Sources	45
Primary versus secondary sources	45

2. Statistical Data (*Continued*)

Suitability of data.....	46
Comparability of data from different sources.....	48

3. Statistical Tables..... 50

Methods of Presentation.....	50
Text presentation.....	50
Tabular presentation.....	51
Semi-tabular presentation.....	51
Graphic presentation.....	52
Leading Considerations.....	53
Types of tables.....	53
Comparisons.....	54
Emphasis.....	56
Arrangement of items in stub and caption.....	56
Details of Table Construction.....	60
Title and identification.....	60
Prefatory notes and footnotes.....	61
Source notes.....	61
Percentages.....	61
Rounding numbers.....	62
Totals.....	63
Units.....	63
Size and shape of table.....	63
Ruling.....	64
Guiding the eye.....	65
Zeros.....	65
Size and style of type.....	65
Statistical Reports.....	65

4. Graphic Presentation I: Curves Using Arithmetic Scales..... 67

The Graphic Method.....	67
Types of Charts.....	68
Plotting a Curve.....	69
Types of Data Shown by Curves.....	71
Time series curves.....	71
Curves of frequency distributions.....	73
Rules for Drawing Curves.....	76
Zero on vertical scale.....	76
Ruling curves.....	80
Coordinates.....	81
Chart proportions.....	81
Lettering.....	84
Title.....	85
Source.....	85
Line Diagrams for Special Purposes.....	85
Net balance charts.....	85
Silhouette charts.....	85
Maximum variation charts.....	85
Range charts.....	87

**4. Graphic Presentation I: Curves Using Arithmetic Scales
(Continued)**

Z-charts.....	87
Varying horizontal-scale charts.....	89
Multiple-axis charts.....	90
Component-part charts.....	90

5. Graphic Presentation II: The Semi-logarithmic or Ratio Chart..... 93

Amount of Change vs. Ratio of Change.....	93
A Grid to Show Ratios of Change.....	98
The logarithmic scale.....	100
Interpretation of curves.....	105
Applications.....	105
Comparing ratios of increase or decrease.....	105
Comparing fluctuations.....	108
Showing ratios.....	109
Interpolation and extrapolation.....	111
Construction of Logarithmic Scales.....	113

6. Graphic Presentation III: Other Types of Charts..... 117

Bases of Comparison.....	117
Bar Charts.....	119
Pictographs.....	123
Component-Part Charts.....	126
Statistical Maps.....	130
Hatched maps.....	131
Dot maps.....	131
Pin maps.....	132

7. Rates, Ratios, and Percentages..... 136

Calculation.....	137
Effect of Changing Base.....	138
Recording Percentages.....	139
Types of Comparisons.....	140
Some Frequently Used Ratios.....	141
Index numbers.....	141
Sex ratio.....	142
Population density.....	142
Ratios per capita.....	142
Death rates.....	143
Birth rates.....	144
Crop yields per acre.....	144
Hog-corn ratio.....	145
Batting averages.....	145
Airline accident ratios.....	146
The 100 per cent statement.....	147
Railroad ratios.....	148
Faulty Use of Percentages.....	149
Confusion in regard to base.....	149
Percentages from small numbers.....	150

7. Rates, Ratios, and Percentages (*Continued*)

Misplaced decimal points.....	150
Arithmetic mistakes.....	151
Improper averaging of percentages.....	151

8. The Frequency Distribution..... 153

Raw Data.....	153
The Array.....	154
The Frequency Distribution.....	156
Selecting the number of classes.....	159
Selecting class limits.....	160
Curves of frequency distributions.....	162
Graphic representation when the class intervals are unequal.....	164
Graphic comparison of frequency distributions.....	166
Cumulative frequency distributions and the ogive.....	168

9. Measures of Central Tendency..... 173

Symbols Used in Chapter 9.....	171
The Arithmetic Mean.....	173
The arithmetic mean from ungrouped data.....	173
Properties of the arithmetic mean.....	174
The arithmetic mean from grouped data: long method... ..	176
The arithmetic mean from grouped data: short methods... ..	179
The arithmetic mean from grouped data having unequal class intervals.....	181
Modified forms of the arithmetic mean.....	182
Averaging percentages.....	183
Averaging averages.....	184
The Median.....	185
The median from ungrouped data.....	185
The median from grouped data.....	186
The quartiles, quintiles, deciles, and percentiles.....	187
The Mode.....	189
The mode from ungrouped data.....	189
The mode from grouped data.....	190
Characteristics of the Mean, Median, and Mode.....	192
Familiarity of the concept.....	192
Algebraic treatment.....	193
Need for classifying data.....	193
Effect of unequal class intervals.....	194
Effect of classes with open end.....	194
Effect of skewness.....	195
Effect of extreme values.....	195
Effect of irregularity of data.....	196
Reliability when based on samples.....	197
Mathematical properties.....	197
Selection of appropriate measure.....	197
Minor Means.....	198
The geometric mean.....	198
The harmonic mean.....	203

CHAPTER	PAGE
10. Dispersion, Skewness, and Kurtosis	212
Symbols Used in Chapter 10	210
Measures of Absolute Dispersion	213
The range	214
The 10-90 percentile range	214
The quartile deviation	215
The average deviation	215
The standard deviation, ungrouped data	215
The standard deviation, grouped data	217
Properties of the standard deviation	219
Measures of Relative Dispersion	222
Skewness	225
Pearsonian measure of skewness	227
Measures of skewness based on quartiles and percentiles	229
Measure of skewness based on the third moment	229
Kurtosis	232
Correction of the Moments for Grouping Error	237
11. The Problem of Time Series	240
Movements in Time Series	240
Secular trend	240
Periodic movements	246
Cyclical movements	249
Irregular variations	251
Other movements	251
A Graphic Review	253
Preliminary Treatment of Data	253
Calendar variation	253
Population changes	256
Price changes	257
Securing comparability	258
12. Analysis of Time Series: Secular Trend I—The Straight Line	261
Symbols Used in Chapter 12	260
Trend Fitted by Inspection	262
Least-Squares Fit of Straight Line	263
The straight line	263
Method of least squares	265
The normal equations	267
Odd number of years	270
Even number of years	273
Adapting Equations to a Monthly Basis	275
Annual totals— X units, one year	276
Annual totals— X units, one-half year	277
Monthly averages— X units, one year	277
Monthly averages— X units, one-half year	277
Selecting the Period for Trend Analysis	278
Selecting the Type of Trend	280

*13. Analysis of Time Series: Secular Trend II—Non-linear Trends	282
Symbols Used in Chapter 13	281
Simple Polynomials	282
Second-degree curve	285
Third-Degree Curve	288
Orthogonal polynomials	289
Use of Logarithms	290
Straight line fitted to logarithms	290
Second-degree curve fitted to logarithms	295
Asymptotic Growth Curves	297
The modified exponential	298
The Gompertz curve	302
The logistic curve	310
Comparison of the Gompertz and logistic curves	316
Selecting a Trend Type	318
14. Analysis of Time Series: Periodic Movements I—Constant Seasonal Patterns	320
An Introductory Illustration	320
Averages of unadjusted data	320
Percentages of simple averages	321
Seasonal Indexes of Monthly Data	324
A seasonal index based on percentages of trend	324
Percentages of centered 12-month moving averages	326
Link relatives	339
Adequacy of the Seasonal Index	339
*15. Analysis of Time Series: Periodic Movements II—Changing Seasonal Patterns	340
Progressive Changes in Seasonal Pattern	340
A moving seasonal	340
Computation of a moving seasonal index	340
Sudden Variations in Seasonal Patterns	352
Adjustment for Easter	352
Sudden changes in entire seasonal pattern	359
Short-time shifts in timing	359
Varying amplitude	360
Further Refinements of Method	361
Continuity of seasonal indexes	361
Combinations of seasonal types	362
Logical basis of methods of construction	362
16. Analysis of Time Series: Cyclical Movements—Adjusting Time Series for Trend, Seasonal, and Irregular Movements	365
Symbols Used in Chapter 16	364
Adjusting Annual Data for Trend	364
Adjustment of Monthly Data	367
Deseasonalizing	367

16. Analysis of Time Series: Cyclical Movements—Adjusting Time Series for Trend, Seasonal, and Irregular Movements (*Continued*)

Adjustment for seasonal and trend	373
Smoothing irregular movements	380
Comparing cyclical movements	384
Other Methods of Estimating Cyclical Movements	388
Direct analysis	388
Harmonic analysis	388
Reference-cycle analysis	388

17. Fundamentals in Index Number Construction 394

Symbols Used in Chapter 17	393
Meaning and Uses of Index Numbers	394
Problems in the Construction of Index Numbers	397
An Illustration of the Behavior of Price Relatives	398
Data for Index Numbers	401
Accuracy	402
Comparability	402
Representativeness	402
Adequacy	403
Selection of Base	404
Aggregative Price Index Numbers	405
Simple aggregates	405
Weighted aggregates	406
Selection of weights	408
Averages of Price Relatives	414
Commodity weights versus group weights	417
Comparison of the four types of price indexes	420
Quantity Index Numbers	421
Aggregative type	421
Averages of relatives	423

18. Index Number Theory and Practice 426

Symbols Used in Chapter 18	425
*Index Number Concepts	426
Mathematical tests	426
Relationship of formula to use	428
The Chain Index	431
*Substituting New Commodities and Changing Weights	432
Descriptions of Indexes	436
Price Indexes	436
The Consumer Price Index	436
United States Bureau of Labor Statistics Index of Wholesale Commodity Prices	438
Indexes of Prices Paid by and Received by Farmers; Parity Ratio	439
Common stock prices	441
Indexes of Physical Volume and of Business Activity	442

18. Index Number Theory and Practice (*Continued*)

Federal Reserve Index of Industrial Production	442
The American Telephone and Telegraph Company Index of Industrial Activity	444
The New York Times Weekly Index of Business Activity	444
Indexes of Business Activity in the Pittsburgh District	445
Indexes of Qualitative Changes or Differences	446
Farm housing indexes for Oklahoma	446
Sources of Current Index Numbers	448

19. Correlation I: Two-variable Linear Correlation 451

Symbols Used in Chapter 19	449
A Simple Explanation	451
Correlation Theory	454
The estimating equation	455
Dependability of estimates	458
The correlation coefficient and explained variation	461
The product-moment formula	465
Practical Methods of Computation	466
Some Cautions	469
Correlation and causation	469
Heterogeneity	470
Errors of measurement	472
Use of averages	472
Non-linear relationship	473
Elimination of relevant data	473
Correlation of Grouped Data	473
Effect of grouping	477
Correlation of Ranked Data	478
Correlation of Data in 2×2 Tables	480

***20. Correlation II: Two-variable Non-linear Correlation 486**

Symbols Used in Chapter 20	483
Polynomials	486
Second-degree curve	486
Third-degree curve	493
Use of Transformations	503
Preliminary examination	503
The $\log Y$, $\log X$ relationship	508
The $\sqrt{Y}$, X relationship	513
Comparison of the three non-linear relationships for diam- eter and volume of trees	516
The $\log Y$, X relationship	518
The $\frac{1}{Y}$, X relationship	519
The Correlation Ratio, η	520

***21. Correlation III: Multiple and Partial Correlation 530**

Symbols Used in Chapter 21	525
Preliminary Explanation	530
Simple correlation	530

***21. Correlation III: Multiple and Partial Correlation**
(Continued)

Multiple correlation.....	531
Partial correlation.....	534
Computation Procedure.....	536
Computation of sums.....	536
Gross measures of relationship.....	538
Two independent variables: multiple correlation.....	541
Two independent variables: partial correlation.....	543
Relationship between $R_{1.23}$ and the measures of gross and partial correlation.....	545
Three independent variables: multiple correlation.....	546
Three independent variables: partial correlation.....	549
Four or more independent variables.....	549
Multiple-partial coefficients.....	551
Another Approach to Multiple and Partial Correlation	
Coefficients.....	552
First-order partial correlation coefficients.....	552
Second-order partial correlation coefficients.....	554
Multiple coefficients.....	555
Coefficients of estimation and standard errors of estimate.....	557
Other Measures of the Individual Importance of the Independent Variables.....	557
Beta coefficients.....	557
Coefficients of separate determination.....	558
Multiple Curvilinear Correlation.....	559
Polynomials.....	559
Transformations.....	559
Graphic method.....	560

22. Correlation IV: Correlation of Time Series..... 562

Symbols Used in Chapter 22.....	561
Annual Data.....	562
Correlation of data unadjusted for trend.....	562
Correlation of percentages of trend.....	563
Correlation of unadjusted data with time as a third variable.....	573
Correlation of amounts of change or percentages of change.....	575
Problems in correlating time series.....	576
Monthly Data.....	578
Synchronous relationships.....	578
Lag and lead.....	579
Procedure for use of lead and lag as an aid in forecasting.....	582

23. Describing a Frequency Distribution by a Fitted Curve. 588

Symbols Used in Chapter 23.....	586
The Normal Curve.....	590
Development of the normal curve.....	590
Explanation of the formula.....	591
Fitting the Normal Curve.....	594
Fitting the normal curve to data of physical ability.....	595

23. Describing a Frequency Distribution by a Fitted Curve (Continued)

The normal curve and collar sizes.....	603
Suitability of the normal curve.....	606
*Binomials.....	607
Fitting a binomial.....	609
*Skewed Curves.....	613
The logarithmic normal curve.....	613
Fitting a logarithmic normal curve.....	616
Fitting a normal curve with adjustment for skewness....	619

24. Statistical Significance I: Arithmetic Means..... 626

Symbols Used in Chapter 24.....	624
How Sample Arithmetic Means Are Distributed.....	626
The arithmetic mean of sample means.....	627
Skewness of sample means.....	627
Kurtosis of sample means.....	629
Sample means and the normal curve.....	631
Dispersion of sample means.....	632
Significance of the Difference between $\bar{X}$ and $\bar{X}_\phi$ When $\bar{X}_\phi$ and σ Are Known.....	635
A difference between $\bar{X}$ and $\bar{X}_\phi$ that is not significant....	635
A difference between $\bar{X}$ and $\bar{X}_\phi$ that is significant.....	637
The value of P and significance.....	638
Probability and everyday occurrences.....	641
Size of sample.....	642
Significance of the Difference Between $\bar{X}$ and $\bar{X}_\phi$ When σ Is Not Known.....	642
A difference between $\bar{X}$ and $\bar{X}_\phi$ that is not significant....	643
A difference between $\bar{X}$ and $\bar{X}_\phi$ that is significant.....	645
Confidence limits of $\bar{X}_\phi$	648
Significance of the Difference Between Two Sample Means.....	651
Independent samples.....	651
Non-independent samples.....	654
Conclusion.....	657

*25. Statistical Significance II: Proportions and the Chi-Square Test..... 661

Symbols Used in Chapter 25.....	659
Part I: Proportions.....	661
Significance of the Difference Between p and π	661
Confidence Limits of π	671
Significance of the Difference Between p_1 and p_2	679
Part 2: The Chi-Square Test.....	681
The 1×2 Table.....	681
The 2×2 Table.....	684
$1 \times R$ Tables, Larger Than 1×2	689
2×3 and Larger Tables.....	691

*26. Statistical Significance III: Variances, Analysis of Variance, Measures of Skewness and Kurtosis, and Correlation Coefficients	699
Symbols Used in Chapter 26	694
Variances	699
Significance of the difference between $\hat{\sigma}^2$ and σ^2	700
Confidence limits of σ^2	701
Significance of the difference between two sample variances	702
Comparison of several values of $\hat{\sigma}^2$	704
Analysis of Variance	706
One criterion of classification	706
Two criteria of classification, one entry in each box	711
Two criteria of classification, more than one entry in a box	714
Measures of Skewness and Kurtosis	720
Skewness	720
Kurtosis	722
Correlation Coefficients	722
Simple correlation	722
Non-linear correlation	726
Multiple correlation	732
Partial correlation	734

APPENDICES

A. Flexible Calendar of Working Days	738
B. Sums of the First Six Powers of the First 50 Natural Numbers	740
C. Sums of the First Six Powers of the First 50 Odd Natural Numbers	742
D. Ordinates of the Normal Curve	744
E. Areas Under the Normal Curve	746
F. Values of $F_2 \left(\frac{x}{s} \right)$	747
G. Areas in One Tail of the Normal Curve at Selected Values of $\frac{x}{s}$ or $\frac{x}{\sigma}$ from the Arithmetic Mean	748
H. Areas in Two Tails of the Normal Curve at Selected Values of $\frac{x}{s}$ or $\frac{x}{\sigma}$ from the Arithmetic Mean	749
I. Values of t	750
J. Values of χ^2	752
K. Values of $\frac{\hat{\sigma}^2}{\sigma^2}$ for Use in Determining Sampling Limits of $\hat{\sigma}^2$	754
L. Values of $\frac{\sigma^2}{\hat{\sigma}^2}$ for Use in Determining Confidence Limits of σ^2	756
M. Values of F	758

APPENDIX

	PAGE
N. Values of L at the 0.05 and 0.01 Points for Specified Values of N_i and k, when $N_1 = N_2 = \dots = N_k = N_i$	763
O. Upper 0.10 and 0.02 Limits of β_1 When Computed from Random Samples from a Normal Population	764
P. Upper and Lower 0.05 and 0.01 Limits of β_2 When Computed from Random Samples from a Normal Population	765
Q. Squares, Square Roots, and Reciprocals, 1-1,000	766
R. Common Logarithms of Numbers	776
S. Demonstrations	792
T. Rounding Numbers	818
U. Selected List of Readily Available Sources of Data	823
A. General	823
B. Commodities—Prices, Production, Consumption, Stocks, Exports, and Imports	824
C. Financial—Money, Banking, Securities, Interest Rates, Taxation, Etc.	826
D. Employment, Wages, and Hours of Labor	826
E. Activities of Individual Concerns	827
F. Miscellaneous	828
Index	831