
Applied General Statistics

By

Frederick E. Croxton, Ph. D.

*Associate Professor of Statistics
Columbia University*

and

Dudley J. Cowden, Ph. D.

*Associate Professor of Economics
School of Commerce
University of North Carolina*

New York : 1940
Prentice-Hall, Inc.

Applied General Statistics

COPYRIGHT, 1939, 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.

First Printing.....September 1939
Second Printing.....April 1940
Third Printing.....November 1940

PRINTED IN THE UNITED STATES OF AMERICA

When you cannot measure what you are speaking about, when you cannot express it in numbers, your knowledge is of a meager and unsatisfactory kind; it may be the beginning of knowledge, but you have scarcely in your thoughts advanced to the stage of a *science*, whatever the matter may be.

—*Lord Kelvin.*

* * *

When the facts are gathered or discovered, when they are disentangled and identified, when they are sifted and verified, when they are counted and measured, the real task of the scholar is not ended—it is not even begun, but only prepared for him.

—*MacIver.*

PREFACE

This book is intended for the use of readers who are interested in the understanding of statistical methods, and in their application in various fields, especially the social sciences. Consequently, the illustrative material has been drawn mainly from the fields of economics, sociology, and business, but occasionally from others.

The arrangement of the topics treated in this volume is about the same as in the authors' earlier book, *Practical Business Statistics*. The present text, however, does not stress business applications of statistical methods, but does present a greatly amplified treatment of analytical methods. The extensive discussion of the description and analysis of statistical data and of the making of statistical inferences will, we hope, make it useful to a wide group of teachers emphasizing various aspects of statistics.

Although the treatment of the material in this text is intended to be at an elementary level, probably its scope is so great that it cannot be covered adequately in most introductory courses. Thus, many instructors will deem it advisable to omit certain material in the first course and reserve it for a second, more advanced or more specialized course. Among the chapters which may be completely omitted in a short course without disturbing the continuity of treatment are Chapters XIII, XVI, XVIII, XXIII, and XXIV. Parts of other chapters may, of course, be left out at the discretion of the instructor. In the table of contents, chapters or sections which could well be omitted from an elementary course have been starred.

One problem of the statistician, especially the teacher, is that of selecting symbols which are simple and clear. Some of the symbols used in the present volume differ from those used in the authors'

earlier text. This departure was made in an attempt to arrive at symbols which would be more easily understood and which would, consequently, facilitate the teaching process.

If this volume has any merit, it must be at least partly because of those who first introduced the authors to statistical methods; because of those texts (and other publications) which have preceded this; because of those publishers and individuals who have allowed us to reproduce charts or data of particular value, specific acknowledgment of which is made in the appropriate connection; and because of numerous individuals who have assisted with the task of completing the book. The authors especially extend their thanks to Mr. Morton S. Nagelberg, who assisted in the construction of charts and in connection with certain of the mathematical developments; to Dr. James D. Paris, who assisted and advised concerning a number of charts; to Brant Bonner, Herbert Wolf, and John W. Gunter, for aid in making computations and preparing charts; and to Rosetta R. Croxton, for assistance in the reading of proof.

FREDERICK E. CROXTON
DUDLEY J. COWDEN

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.)

	PAGE
PREFACE	V
CHAPTER	
I. INTRODUCTION	1
Statistical data and statistical methods	1
Collection	2
Presentation	3
Analysis	3
Interpretation	7
A few improprieties	7
Bias	7
Omission of important factor	8
Carelessness	9
Non-sequitur	9
Non-comparable data	9
Confusion of association and causation	10
Insufficient data	10
Unrepresentative data	12
Concealed classification	12
Research methods	13
II. STATISTICAL DATA	15
Collecting statistical data	16
Method of collection	16
Outline of procedure	16
Laying out the general plan	16
Devising questions and making the schedule	18
Selecting the sample	26
Using the schedules to collect the data	33
Editing the schedules	35
Tabulating the data	37
Presentation and analysis	44

CHAPTER	PAGE
II. STATISTICAL DATA (<i>Cont.</i>)	
Statistical sources.....	44
Reliability of data.....	45
Comparability of different sources.....	47
III. STATISTICAL TABLES.....	49
Methods of presentation.....	49
Text presentation.....	49
Tabular presentation.....	50
Semi-tabular presentation.....	51
Graphic presentation.....	52
Leading considerations.....	52
Types of tables.....	52
Comparisons.....	53
Emphasis.....	56
Arrangement of items in stub and caption.....	57
Details of table construction.....	62
Title and identification.....	62
Prefatory note and footnotes.....	62
Source notes.....	62
Percentages.....	63
Rounding numbers.....	63
Totals.....	64
Units.....	64
Size and shape of table.....	64
Ruling.....	65
Guiding the eye.....	66
Zeros.....	66
Size and style of type.....	66
Statistical reports.....	66
IV. GRAPHIC PRESENTATION—SIMPLE CURVES.....	70
The graphic method.....	70
Types of charts.....	71
Plotting a curve.....	72
Types of data shown by curves.....	74
Time series curves.....	74
Curves of frequency distributions.....	77
Rules for drawing curves.....	79
Zero on vertical scale.....	81
Ruling curves.....	85
Coordinates.....	86
Title.....	91
Source.....	91

IV.	GRAPHIC PRESENTATION—SIMPLE CURVES (<i>Cont.</i>)	
	Line diagrams for special purposes.....	91
	Net balance charts.....	91
	Silhouette charts.....	91
	Maximum variation charts.....	92
	Range charts.....	92
	Z charts.....	93
	Varying horizontal scale charts.....	95
	Multiple axis charts.....	95
	Component part charts.....	95
	Frequency distribution and range chart.....	97
V.	GRAPHIC PRESENTATION—THE SEMI-LOGARITHMIC OR RATIO CHART.....	100
	Absolute and relative growth.....	100
	A grid to show rates of change.....	104
	How it is made.....	106
	The logarithmic scale.....	107
	Interpretation of curves.....	109
	Applications.....	109
	Comparing rates of increase or decrease.....	109
	Comparing fluctuations.....	114
	Showing ratios.....	117
	Interpolation and extrapolation.....	118
	Flexible logarithmic scales.....	120
VI.	GRAPHIC PRESENTATION—OTHER TYPES OF CHARTS	124
	Bases of comparison.....	124
	Bar charts.....	126
	Pictorial devices.....	131
	Component part charts.....	133
	Statistical maps.....	137
	Hatched maps.....	137
	Dot maps.....	137
	Pin maps.....	144
VII.	RATIOS AND PERCENTAGES.....	146
	Calculation.....	146
	Effect of changing base.....	148
	Recording percentages.....	149
	Types of comparisons.....	150
	Some frequently used ratios.....	151
	Index numbers.....	151
	Sex ratio.....	151
	Population density.....	152

CHAPTER

PAGE

VII. RATIOS AND PERCENTAGES (*Cont.*)

Persons per family.....	152
Ratios per capita.....	152
Death rates.....	153
Birth rates.....	154
Crop yields per acre.....	155
Hog-corn ratio.....	155
Batting averages.....	156
Airplane accident ratios.....	157
The 100 per cent statement.....	157
Railroad ratios.....	158
Faulty use of percentages.....	159
Confusion in regard to base.....	159
Percentages from small numbers.....	160
Misplaced decimal points.....	161
Arithmetic mistakes.....	161
Improper averaging of percentages.....	161
Unduly large percentages.....	162

VIII. THE FREQUENCY DISTRIBUTION..... 164

Raw data.....	164
The array.....	165
The frequency distribution.....	168
Selecting the number of classes.....	171
Selecting class limits.....	172
Curves of frequency distributions.....	174
Plotting a frequency curve when the class intervals are unequal.....	177
Comparison of frequency distributions.....	180
Cumulative frequency distributions and the ogive.....	184
Comparison of ogives.....	186
The Lorenz curve.....	188
The Pareto curve.....	190

IX. MEASURES OF CENTRAL TENDENCY..... 194

The arithmetic mean.....	194
The arithmetic mean from ungrouped data.....	194
Properties of the arithmetic mean.....	195
The arithmetic mean from grouped data: long method.....	197
The arithmetic mean from grouped data: short methods.....	200
The arithmetic mean from grouped data having unequal class intervals.....	202
Modified forms of the arithmetic mean.....	204

IX. MEASURES OF CENTRAL TENDENCY (*Cont.*)

Averaging percentages	205
Averaging averages	206
The median	207
The median from ungrouped data	207
The median from grouped data	208
The quartiles, quintiles, deciles, and percentiles	210
The mode	212
The mode from ungrouped data	212
The mode from grouped data	212
Characteristics of the mean, median, and mode	215
Familiarity of the concept	215
Algebraic treatment	215
Need for classifying data	216
Effect of unequal class intervals	217
Effect of classes with open end	217
Effect of skewness	217
Effect of extreme values	218
Effect of irregularity of data	219
Reliability when based on samples	220
Mathematical properties	220
Selection of appropriate measure	220
Minor means	221
The geometric mean	221
The harmonic mean	226

X. DISPERSION, SKEWNESS, AND KURTOSIS 234

Measures of absolute dispersion	235
The range	236
The 10-90 percentile range	237
The quartile deviation	237
The average deviation	238
The standard deviation, ungrouped data	240
The standard deviation, grouped data	242
Properties of the standard deviation	243
Comparison of measures of absolute dispersion	246
Measures of relative dispersion	246
Skewness	249
Pearsonian measure of skewness	251
Measures of skewness based on quartiles and per- centiles	253
*Measure of skewness based on the third moment	254
*Kurtosis	258
*Correction of the moments for grouping error	262

XI.	DESCRIBING A FREQUENCY DISTRIBUTION BY A FITTED CURVE.....	265
	The normal curve of error.....	266
	Development of the normal curve.....	266
	Fitting the normal curve.....	271
	Fitting the normal curve to data of physical ability.....	272
	The normal curve fitted to batting averages.....	280
	The normal curve and collar sizes.....	281
	Suitability of the normal curve.....	282
	*Binomials.....	287
	Experimental construction of skewed binomials.....	287
	Fitting a binomial.....	289
	*Skewed curves.....	293
	The logarithmic normal curve.....	293
	Fitting a logarithmic normal curve.....	295
	Fitting a normal curve with adjustment for skewness.....	299
XII.	RELIABILITY AND SIGNIFICANCE OF STATISTICAL MEASURES—ARITHMETIC MEANS.....	305
	Reliability of sample means, large samples.....	305
	The standard error of a sample mean.....	307
	Significance of the deviation of a sample mean from the mean of a known population.....	308
	The null hypothesis.....	310
	Reliability of a sample mean, when X_P and σ_P are unknown.....	311
	Significance of the difference between a sample mean and a hypothetical population mean.....	314
	Significance of the difference between two sample means.....	317
	Procedure when $N_1 \neq N_2$	322
	Reliability of the mean of a stratified sample.....	324
	*Reliability of sample means, small samples.....	325
	The t distribution.....	325
	Reliability of a sample mean when N is small.....	327
	Significance of the difference between two means when $N_1 \neq N_2$ and both N 's are small.....	330
*XIII.	RELIABILITY AND SIGNIFICANCE OF STATISTICAL MEASURES — PERCENTAGES, STANDARD DEVIATIONS, VARIANCES, AND THE CRITERION OF LIKELIHOOD.....	332
	Reliability of sample percentages.....	332
	Standard error of a sample percentage.....	332

*XIII.	RELIABILITY AND SIGNIFICANCE OF STATISTICAL MEASURES — PERCENTAGES, STANDARD DEVIATIONS, VARIANCES, AND THE CRITERION OF LIKELIHOOD (<i>Cont.</i>)	
	Reliability of a sample percentage.....	332
	Reliability of percentages and the χ^2 test.....	333
	Significance of difference between percentages.....	337
	Reliability of measures of dispersion.....	339
	Reliability of a sample σ	339
	Reliability of σ when N is small.....	340
	Significance of difference between two standard deviations when N 's are large and when $N_1 = N_2$	343
	Significance of differences between two σ 's when N 's are small and/or $N_1 \neq N_2$	344
	An application of reliability measures.....	348
	Analysis of variance.....	351
	Criterion of likelihood.....	359
	Comparison of several σ 's.....	359
XIV.	THE PROBLEM OF TIME SERIES.....	363
	The problem stated.....	363
	Characteristics of time series.....	363
	Secular trend.....	364
	Cyclical movements.....	367
	Irregular variations.....	372
	A graphic illustration.....	372
	Another view of time series movements.....	376
	Preliminary treatment of data.....	378
	Calendar variation.....	379
	Population changes.....	382
	Price changes.....	382
	Securing comparability.....	383
XV.	ANALYSIS OF TIME SERIES—SECULAR TREND.....	385
	Objects and method.....	385
	Trend fitted by inspection.....	386
	Moving averages.....	386
	Simple moving averages and annual trend values.....	386
	Obtaining monthly trend values from annual moving averages.....	392
	Moving averages: summary.....	394
	Straight line trend.....	395
	Description.....	395
	Method of selected points.....	397
	Method of least squares.....	399

CHAPTER	PAGE
XV. ANALYSIS OF TIME SERIES—SECULAR TREND (<i>Cont.</i>)	
Normal equations.....	401
Odd number of items.....	404
Even number of items.....	405
Adaptation of equations to monthly data.....	408
Other simple types of trends.....	411
Series of curves.....	411
Related series as trend.....	411
Cyclical averages.....	412
Selecting the type of trend.....	418
Adjustment for trend.....	419
*XVI. OTHER TREND TYPES.....	421
Weighted moving averages.....	421
Simple polynomials.....	426
Second degree curve.....	426
Third degree curve.....	430
Empirical test of data.....	432
Orthogonal polynomials.....	433
Use of logarithms.....	435
Straight line equation.....	435
Second degree curve.....	440
Curves with declining absolute growth.....	440
(1) Modified polynomials.....	440
(2) Straight line to log X	440
(3) Parabolic curve to log Y	440
(4) Modified exponential.....	441
Asymptotic growth curves.....	441
Modified exponential.....	441
Modified exponential fitted to department store sales.....	445
Gompertz curve.....	447
Logistic curve.....	452
Use of arithmetic probability paper.....	458
Objective tests of trends.....	461
XVII. PERIODIC MOVEMENTS.....	464
Averages of unadjusted data.....	464
Percentages of simple averages.....	466
Trend adjustment for averages.....	467
Percentages of trend.....	469
Percentages of 12-month moving average.....	471
A graphic approach to seasonal measurement.....	484
Link relative method.....	486
Comparison of results.....	492

CONTENTS

xv

CHAPTER		PAGE
XVII.	PERIODIC MOVEMENTS (<i>Cont.</i>)	
	Adjustment for seasonal.....	492
	Test of seasonal.....	497
*XVIII.	TYPES OF SEASONAL MOVEMENTS.....	500
	Progressive changes in seasonal pattern.....	500
	Use of moving averages.....	500
	Computation of moving seasonal.....	501
	Sudden variations in seasonal pattern.....	509
	Adjustment for Easter.....	509
	Sudden changes in entire seasonal pattern.....	516
	Short-time shifts in timing.....	516
	Varying amplitude.....	518
	Further refinements of method.....	524
	Continuity of seasonal indexes.....	524
	Combinations of seasonal types.....	525
	Correction by subtraction of seasonal.....	525
	Logical basis of methods of construction.....	527
	Weekly seasonal.....	528
XIX.	CYCLICAL MOVEMENTS.....	540
	Residual method.....	540
	Reducing minor irregularities.....	548
	Comparison of cyclical movements.....	549
	*Direct method.....	552
	*Harmonic analysis.....	554
	Periodogram analysis.....	555
	Fitting a periodic curve.....	559
	*Cyclical averages.....	560
	Specific cycle analysis.....	562
	Reference cycle analysis.....	566
	Comparison of reference and specific cycles.....	568
XX.	FUNDAMENTALS IN INDEX NUMBER CONSTRUCTION	573
	Meaning and uses of index numbers.....	573
	Problems in the construction of index numbers.....	576
	An illustration of the behavior of price relatives....	577
	Data for index numbers.....	582
	Accuracy.....	583
	Comparability.....	583
	Representativeness.....	583
	Adequacy.....	586
	Selection of base.....	586
	Aggregative price index numbers.....	588
	Simple aggregates.....	588

CHAPTER		PAGE
XX.	FUNDAMENTALS IN INDEX NUMBER CONSTRUCTION (<i>Cont.</i>)	
	Weighted aggregates.....	590
	Selection of weights.....	591
	Averages of price relatives.....	597
	Type of average.....	599
	Weighting systems.....	601
	Commodity weights versus group weights.....	603
	Quantity index numbers.....	607
	Aggregative type.....	607
	Averages of relatives.....	609
XXI.	INDEX NUMBER THEORY AND PRACTICE.....	612
	Index number concepts.....	612
	Mathematical tests.....	612
	Relationship of formula to use.....	614
	The chain index.....	616
	Circular test.....	621
	*Substituting new commodities and changing weights.....	623
	Some price indexes.....	627
	Changes in cost of living.....	629
	Geographical variations in cost of living.....	630
	Indexes of physical volume of production and trade..	631
	Indexes of business cycles.....	639
	*Indexes of qualitative changes or differences.....	644
	Measures of adequacy of state school systems....	645
	Sources of current index numbers.....	650
XXII.	SIMPLE CORRELATION.....	651
	A simple explanation.....	651
	Correlation theory.....	654
	The estimating equation.....	655
	Dependability of estimates.....	657
	The correlation coefficient and explained variability	660
	The product-moment formula.....	666
	Practical methods of computation.....	667
	Correlation of grouped data.....	673
	Causation and the correlation coefficient.....	678
	Estimate of correlation in population.....	679
	Reliability of correlation coefficient.....	680
	General measure of reliability.....	680
	The <i>t</i> test.....	681
	Analysis of variance.....	682
	The <i>Z</i> transformation.....	683
	Correlation of ranked data.....	685
	Correlation of qualitative distributions.....	687

CHAPTER	PAGE
*XXIII. NON-LINEAR CORRELATION.....	691
Transforming data to linear form.....	691
Use of logarithms.....	684
Use of reciprocals.....	699
Curves with more than two constants.....	705
Second degree curve.....	705
Third degree curve.....	712
Grouped data.....	721
Use of means.....	727
A simple illustration.....	727
Data grouped on both axes.....	732
Limitations of correlation ratio.....	735
Unreliability of coefficients of curvilinear correlation	736
*XXIV. MULTIPLE AND PARTIAL CORRELATION.....	739
Preliminary explanation.....	739
Simple correlation.....	739
Multiple correlation.....	740
Partial correlation.....	742
Computation procedure.....	743
Computation of product sums.....	743
Computation of gross measures of relationship....	748
Two independent variables: multiple correlation ..	756
Two independent variables: partial correlation....	761
Three independent variables: multiple correlation..	765
Three independent variables: partial correlation..	769
Another approach to multiple and partial correlation	770
Partial coefficients.....	770
σ_s and R	772
Other measures of the individual importance of the	
independent variables.....	773
Estimate of correlation in the population.....	775
Reliability of coefficients.....	775
Standard error of coefficients.....	775
Analysis of variance.....	776
Multiple curvilinear correlation.....	778
Transformation to linear form.....	778
Use of polynomials.....	779
A graphic approach.....	784
Limitations of graphic method.....	788
XXV. CORRELATION OF TIME SERIES AND FORECASTING	791
Correlation of time series.....	791
Preliminary adjustment of data.....	791
Correlation of adjusted cyclical relatives.....	795