

Applied General Statistics

Frederick E. Croxton, Ph.D.

Professor Emeritus of Statistics, Columbia University

Dudley J. Cowden, Ph.D.

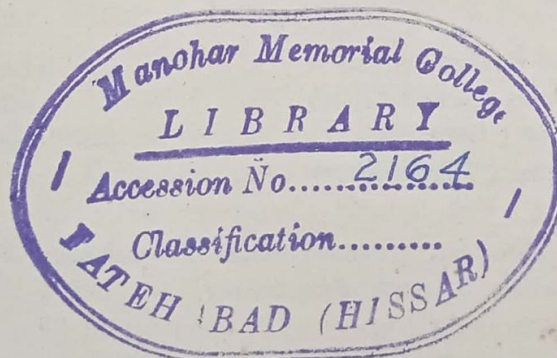
*Professor of Economic Statistics
School of Business Administration
University of North Carolina*

and

Sidney Klein, Ph.D.

*Associate Professor of Economics
Rutgers—The State University*

Third Edition



PRENTICE-HALL OF INDIA PRIVATE LIMITED
New Delhi, 1971

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.....	6
Interpretation.....	6
A Few Improprieties.....	6
Bias.....	7
Omission of important factor.....	8
Carelessness.....	8
Non-sequitur.....	8
Non-comparable data.....	9
Confusion of association and causation.....	9
Insufficient data.....	10
Unrepresentative data.....	10
Concealed classification.....	10
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.....	23
4. Using the schedules to obtain the information.....	31
5. Editing the schedules.....	32
6. Organizing the data.....	34
7. Presentation and analysis.....	40
Using Existing Sources.....	40
Primary versus secondary sources.....	40

2. Statistical Data (Continued)		
Suitability of data	41	
Comparability of data from different sources	43	
3. Statistical Tables		45
Methods of Presentation	45	
Text presentation	46	
Tabular presentation	46	
Semi-tabular presentation	47	
Graphic presentation	48	
Leading Considerations	48	
Types of tables	49	
Comparisons	50	
Emphasis	51	
Arrangement of items in stub and caption	53	
Details of Table Construction	53	
Titles and identification	54	
Prefatory notes and footnotes	54	
Source notes	55	
Percentages	55	
Rounding numbers	56	
Totals	56	
Units	56	
Size and shape of table	57	
Ruling	58	
Guiding the eye	58	
Zeros	58	
Size and style of type	58	
Statistical Reports		

4. Graphic Presentation I: Curves Using Arithmetic Scales 60

The Graphic Method	60
Types of Charts	61
Plotting a Curve	62
Types of Data Shown by Curves	64
Time series curves	64
Curves of frequency distributions	65
Rules for Drawing Curves	68
Zero on vertical scale	68
Ruling curves	71
Coordinates	72
Chart proportions	73
Lettering	73
Title	76
Source	76
Line Diagrams for Special Purposes	76
Net balance charts	77
Silhouette charts	77
Range charts	77

4. Graphic Presentation I: Curves Using Arithmetic Scales (Continued)		
Z-charts	77	
Varying horizontal-scale charts	80	
Multiple-axis charts	80	
Component-part charts	82	
Frequency distribution of range chart	82	
5. Graphic Presentation II: The Semi-logarithmic or Ratio Chart		84
Amount of Change vs. Ratio of Change	84	
A Grid to Show Ratios of Change	89	
The logarithmic scale	90	
Interpretation of curves	93	
Applications	96	
Comparing ratios of increase or decrease	96	
Comparing fluctuations	98	
Showing ratios	99	
Interpolation and extrapolation	101	
Construction of Logarithmic Scales	103	
6. Graphic Presentation III: Other Types of Charts		105
Bases of Comparison	105	
Bar Charts	107	
Pictographs	111	
Component-Part Charts	112	
Statistical Maps	117	
Hatched maps	118	
Dot maps	118	
Pin maps	119	
7. Rates, Ratios, and Percentages		121
Calculation	122	
Effect of Changing Base	123	
Recording Percentages	124	
Types of Comparisons	125	
Some Frequently Used Ratios	126	
Index numbers	126	
Sex ratio	127	
Population density	127	
Ratios per capita	127	
Death rates	129	
Birth rates	129	
Crop yields per acre	129	
Hog-corn ratio	129	
Batting averages	130	
Airline accident ratios	131	
The 100 per cent statement	131	
Railroad ratios	132	

7. Rates, Ratios, and Percentages (Continued)

Faulty Use of Percentages	133
Confusion in regard to base	133
Percentages from small numbers	134
Misplaced decimal points	134
Arithmetic mistakes	135
Improper averaging of percentages	135

8. The Frequency Distribution 136

Raw Data	136
The Array	138
The Frequency Distribution	139
Selecting the number of classes	141
Selecting class limits	143
Curves of frequency distributions	145
Graphic representation when the class intervals are unequal	148
Graphic comparison of frequency distributions	149
Cumulative frequency distributions and the ogive	150

9. Measures of Central Tendency 153

The Arithmetic Mean	153
The arithmetic mean from ungrouped data	153
Properties of the arithmetic mean	154
The arithmetic mean from grouped data: long method	156
The arithmetic mean from grouped data: short methods	159
The arithmetic mean from grouped data having unequal class intervals	161
Modified forms of the arithmetic mean	162
Averaging percentages	163
Averaging averages	164
The Median	165
The median from ungrouped data	165
The median from grouped data	166
The quartiles, quintiles, deciles, and percentiles	167
The Mode	169
The mode from ungrouped data	169
The mode from grouped data	170
Characteristics of the Mean, Median, and Mode	171
Familiarity of the concept	172
Algebraic treatment	172
Need for classifying data	173
Effect of unequal class intervals	174
Effect of classes with open end	174
Effect of skewness	174
Effect of extreme values	175
Effect of irregularity of data	176
Reliability when based on samples	176
Mathematical properties	176
Selection of appropriate measure	176

9. Measures of Central Tendency (Continued)

Minor Means	177
The geometric mean	178
The harmonic mean	182

10. Dispersion, Skewness, and Kurtosis 189

Measures of Absolute Dispersion	190
The range	191
The 10-90 percentile range	191
The quartile deviation	192
The average deviation	192
The standard deviation, ungrouped data	194
The standard deviation, grouped data	195
Properties of the standard deviation	198
Measures of Relative Dispersion	201
Skewness	202
Pearsonian measure of skewness	205
Measures of skewness based on quartiles and percentiles	205
Measure of skewness based on the third moment	208
Kurtosis	212
Correction of the Moments for Grouping Error	212

11. The Problem of Time Series 214

Movements in Time Series	214
Secular trend	214
Periodic movements	218
Cyclical movements	221
Irregular variations	223
Other movements	223
A Graphic Review	223
Preliminary Treatment of Data	225
Calendar variation	225
Population changes	226
Price changes	226
Securing comparability	228

12. Analysis of Time Series: Secular Trend I—The Straight Line 229

Trend Fitted by Inspection	230
Least-Squares Fit of Straight Line	231
The straight line	231
Method of least squares	233
The normal equations	235
Odd number of years	238
Even number of years	241
Adapting Equations to a Monthly Basis	243
Annual totals— X units, one year	243
Annual totals— X units, one-half year	244

CHAPTER PAGE
12. Analysis of Time Series: Secular Trend I—The Straight Line (Continued)

Monthly averages— X units, one year	244
Monthly averages— X units, one-half year	245
Selecting the Period for Trend Analysis	245
Selecting the Type of Trend	247

*13. Analysis of Time Series: Secular Trend II—Non-linear Trends 249

Simple Polynomials	249
Second-degree curve	252
Third-Degree Curve	255
Use of Logarithms	256
Straight line fitted to logarithms	256
Second-degree curve fitted to logarithms	260
Asymptotic Growth Curves	262
The modified exponential	263
The Gompertz curve	267
The logistic curve	274
Comparison of the Gompertz and logistic curves	280
Selecting a Trend Type	282

14. Analysis of Time Series: Periodic Movements I—Constant Seasonal Patterns 285

An Introductory Illustration	285
Averages of unadjusted data	285
Percentages of simple averages	286
Seasonal Indexes of Monthly Data	289
A seasonal index based on percentages of trend	289
Percentages of centered 12-month moving averages	290
Link relatives	303
Adequacy of the Seasonal Index	303

*15. Analysis of Time Series: Periodic Movements II—Changing Seasonal Patterns 304

Progressive Changes in Seasonal Pattern	304
A moving seasonal	304
Computation of a moving seasonal index	304
Sudden Variations in Seasonal Patterns	314
Adjustment for Easter	314
Sudden changes in entire seasonal pattern	314
Short-time shifts in timing	314
Varying amplitude	315
Further Refinements of Method	316
Continuity of seasonal indexes	316
Combinations of seasonal types	317
Logical basis of methods of construction	317

CHAPTER PAGE
16. Analysis of Time Series: Cyclical Movements—Adjusting Time Series for Trend, Seasonal, and Irregular Movements 319

Adjusting Annual Data for Trend	319
Adjustment of Monthly Data	321
Deseasonalizing	321
Adjustment for seasonal and trend	326
Smoothing irregular movements	329
Comparing cyclical movements	336
Other Methods of Estimating Cyclical Movements	340
Direct analysis	340
Harmonic analysis	340
Reference-cycle analysis	340

17. Fundamentals in Index Number Construction 343

Meaning and Uses of Index Numbers	343
Problems in the Construction of Index Numbers	345
An Illustration of the Behavior of Price Relatives	346
Data for Index Numbers	348
Accuracy	349
Comparability	349
Representativeness	350
Adequacy	351
Selection of Base	352
Aggregative Price Index Numbers	352
Simple aggregates	353
Weighted aggregates	354
Selection of weights	355
Averages of Price Relatives	361
Commodity weights versus group weights	364
Comparison of the four types of price indexes	366
Quantity Index Numbers	367
Aggregative type	368
Averages of relatives	368

18. Index Number Theory and Practice 371

*Index Number Concepts	371
Mathematical tests	371
Relationship of formula to use	373
The Chain Index	375
*Substituting New Commodities and Changing Weights	376
Descriptions of Indexes	381
Price Indexes	381
The Consumer Price Index	381
United States Bureau of Labor Statistics Index of Wholesale Commodity Prices	382
Indexes of Prices Paid by and Received by Farmers; Parity Ratio	382

18. Index Number Theory and Practice (Continued) 384

- Common stock prices 386
- Indexes of Physical Volume and of Business Activity 386
- Federal Reserve Index of Industrial Production 387
- Other indexes of physical volume and of business activity 387
- Indexes of Qualitative Changes or Differences 387

19. Correlation I: Two-variable Linear Correlation 389

- A Simple Explanation 389
- Correlation Theory 393
- The estimating equation 393
- Dependability of estimates 395
- The correlation coefficient and explained variation 398
- The product-moment formula 402
- Practical Methods of Computation 404
- Some Cautions 406
- Correlation and causation 406
- Heterogeneity 407
- Errors of measurement 408
- Use of averages 409
- Non-linear relationship 409
- Elimination of relevant data 410
- Correlation of Grouped Data 413
- Effect of grouping 414
- Correlation of Ranked Data 414
- Correlation of Data in 2×2 Tables 416

***20. Correlation II: Two-variable Non-linear Correlation** 419

- Polynomials 419
- Second-degree curve 426
- Third-degree curve 431
- Use of Transformations 431
- Preliminary examination 435
- The log Y , log X relationship 441
- The \sqrt{Y} , X relationship 444
- Comparison of the three non-linear relationships for diameter and volume of trees 446
- The log Y , X relationship 447
- The $\frac{1}{Y}$, X relationship 448
- The Correlation Ratio, η 448

***21. Correlation III: Multiple and Partial Correlation** 453

- Preliminary Explanation 453
- Simple correlation 453
- Multiple correlation 454
- Partial correlation 457
- Computation Procedure 458

CHAPTER

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

- The logarithmic normal curve 527
- Fitting a logarithmic normal curve 528
- Fitting a normal curve with adjustment for skewness 533

24. Statistical Significance I: Arithmetic Means 537

- How Sample Arithmetic Means Are Distributed 537
- The arithmetic mean of sample means 537
- Skewness of sample means 538
- Kurtosis of sample means 540
- Sample means and the normal curve 541
- Dispersion of sample means 543
- Significance of the Difference Between \bar{X} and \bar{X}_ϕ When \bar{X}_ϕ and σ Are Known 544
- A difference between \bar{X} and \bar{X}_ϕ that is not significant 544
- A difference between \bar{X} and \bar{X}_ϕ that is significant 547
- The value of P and significance 547
- Probability and everyday occurrences 550
- Size of sample 551
- Significance of the Difference Between \bar{X} and \bar{X}_ϕ When σ Is Not Known 551
- A difference between \bar{X} and \bar{X}_ϕ that is not significant 553
- A difference between \bar{X} and \bar{X}_ϕ that is significant 555
- Confidence limits of \bar{X}_ϕ 555
- Significance of the Difference Between Two Sample Means 560
- Independent samples 560
- Confidence limits of $\bar{X}_{\phi_1} - \bar{X}_{\phi_2}$ 562
- Non-independent samples 563
- Conclusion 565

***25. Statistical Significance II: Proportions and the Chi-Square Test** 567

- Part I: Proportions 567
- Significance of the Difference Between p and π 567
- Confidence Limits of π 578
- Significance of the Difference Between p_1 and p_2 585
- Part 2: The Chi-Square Test 586
- The 1×2 Table 586
- The 2×2 Table 589
- $1 \times R$ Tables, Larger Than 1×2 594
- 2×3 and Larger Tables 596

***26. Statistical Significance III: Variances, Analysis of Variance, Measures of Skewness and Kurtosis, and Correlation Coefficients** 599

- Variances 599
- Significance of the difference between $\hat{\sigma}^2$ and σ^2 600

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

Computation of sums	458
Gross measures of relationship	461
Two independent variables: multiple correlation	464
Two independent variables: partial correlation	466
Relationship between $R_{1,22}$ and the measures of gross and partial correlation	468
Three independent variables: multiple correlation	469
Three independent variables: partial correlation	472
Four or more independent variables	472
Multiple-partial coefficients	472
Another Approach to Multiple and Partial Correlation Coefficients	473
First-order partial correlation coefficients	473
Second-order partial correlation coefficients	475
Multiple coefficients	476
Coefficients of estimation and standard errors of estimate	477
Other Measures of the Individual Importance of the Independent Variables	478
Multiple Curvilinear Correlation	478
Polynomials	478
Transformations	478
Graphic method	479

22. Correlation IV: Correlation of Time Series 480

Annual Data	481
Correlation of data unadjusted for trend	481
Correlation of percentages of trend	482
Correlation of unadjusted data with time as a third variable	491
Correlation of amounts of change or percentages of change	492
Problems in correlating time series	494
Monthly Data	496
Synchronous relationships	496
Lag and lead	496
Procedure for use of lead and lag as an aid in forecasting	500

23. Describing a Frequency Distribution by a Fitted Curve 503

The Normal Curve	504
Development of the normal curve	504
Explanation of the formula	505
Fitting the Normal Curve	509
Fitting the normal curve to data of physical ability	509
The normal curve and collar sizes	518
Suitability of the normal curve	520
*Binomials	521
Experimental construction of skewed binomials	521
Fitting a binomial	523

***26. Statistical Significance III: Variances, Analysis of Variance, Measures of Skewness and Kurtosis, and Correlation Coefficients (Continued)**

Confidence limits of σ^2	601
Significance of the difference between two sample variances	602
Comparison of several values of σ^2	604
Analysis of Variance	605
One criterion of classification	605
Two criteria of classification, one entry in each box	610
Two criteria of classification, more than one entry in a box	614
Interrelationships Between $\frac{x}{s}$, t , χ^2 , and F	619
Measures of Skewness and Kurtosis	620
Skewness	620
Kurtosis	621
Correlation Coefficients	621
Simple correlation	625
Non-linear correlation	631
Multiple correlation	631
Partial correlation	633

APPENDICES

A. Symbols Used in Each Chapter	638
B. Sums of the First Six Powers of the First 50 Natural Numbers	660
C. Sums of the First Six Powers of the First 50 Odd Natural Numbers	662
D. Ordinates of the Normal Curve	664
E. Areas Under the Normal Curve	666
F. Values of $F^2 \left(\frac{x}{s} \right)$	667
G. Areas in One Tail of the Normal Curve at Selected Values of $\frac{x}{s}$ or $\frac{x}{\sigma}$ from the Arithmetic Mean	668
H. Areas in Two Tails of the Normal Curve at Selected Values of $\frac{x}{s}$ or $\frac{x}{\sigma}$ from the Arithmetic Mean	669
I. Values of t	670
J. Values of χ^2	672
K. Values of $\frac{\hat{\sigma}^2}{\sigma^2}$ for Use in Determining Sampling Limits of $\hat{\sigma}^2$	674
L. Values of $\frac{\sigma^2}{\hat{\sigma}^2}$ for Use in Determining Confidence Limits of σ^2	676
M. Values of F	678

	PAGE
APPENDIX	
N. Values of L at the 0.05 and 0.01 Points for Specified Values of N_i and k , when $N_1 = N_2 = \dots = N_k = N_i$	683
O. Upper 0.10 and 0.02 Limits of β_1 When Computed from Random Samples from a Normal Population.....	684
P. Upper and Lower 0.05 and 0.01 Limits of β_2 When Computed from Random Samples from a Normal Population.....	685
Q. Squares, Square Roots, and Reciprocals, 1-1,000.....	686
R. Common Logarithms of Numbers.....	696
S. Demonstrations.....	712
T. Rounding Numbers.....	738
Index.....	743