

STATISTICS

SECOND EDITION

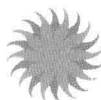
NORMA GILBERT

STATISTICS

SECOND EDITION

NORMA GILBERT

Department of Mathematics
Drew University



SAUNDERS COLLEGE PUBLISHING

Philadelphia New York Chicago
San Francisco Montreal Toronto
London Sydney Tokyo Mexico City
Rio de Janeiro Madrid

Address orders to:
383 Madison Avenue
New York, NY 10017

Address editorial correspondence to:
West Washington Square
Philadelphia, PA 19105

This book was set in Times Roman by York Graphics Services, Inc.
The editors were Jim Porterfield and Carol Field.
The art director and cover designer was Nancy E.J. Grossman.
The text design was done by North 7 Atelier, Ltd.
The production manager was Tom O'Connor.
Von Hoffmann was the printer.

Front cover art: Watercolor by Richard Lutzke.

Library of Congress Catalog Card No.: 80-53939

STATISTICS

ISBN 0-03-058091-1

© 1981 by CBS College Publishing. Copyright 1976 by W. B. Saunders Company.

Printed in the United States of America.

All rights reserved.

1234 032 987654321

CBS COLLEGE PUBLISHING
Saunders College Publishing
Holt, Rinehart and Winston
The Dryden Press

PREFACE

This book is intended for students who need to understand how statistical decisions are made but who have little mathematical background (a year or two of high school algebra is sufficient). So many statistics texts aimed at this group have appeared that some explanation—other than the pleasure I’ve had in writing it—is needed of why another appears. Here are several reasons:

First, many formulas appear in any statistics book. I have tried to explain where a formula comes from, why it should be used, and when it should not be used. Explanations are often intuitive and informal, but a student should get a grasp of why a decision is reached rather than just learn what is reached.

Second, in some texts the notation gets pretty involved. In this one it doesn’t. (The exception is Chapter 15; by then a student is well accustomed to statistics. Some instructors will prefer to omit this chapter.)

Third, the book is suited to self-paced (“Keller Plan”) study as well as to the usual lecture approach: An introductory section at the beginning of each chapter explains its scope and where emphasis should be put. About 250 examples are worked out completely. Answers are given in detail to about 600 of the exercises in the belief that it does no good to know that the probability of an event, for example, is .40 if the student doesn’t know how .40 is derived.

Last, some texts avoid probability, but an understanding of statistics without it is impossible. Others dive into probability immediately. Many students have trouble with this topic, however, so I have begun with descriptive statistics and delayed probability until Chapter 5; it is then used in every succeeding chapter.

Chapters 1 through 8, and parts of 9 and 10 should ordinarily be taken in sequence; material from the remainder of Chapters 9 and 10 and Chapters 11 through 17 can be selected according to the interests and speed of the students and the length of the course.

The second edition differs from the first in that (1) hundreds of new exercises have been added; many of these are statistical problems from real life; (2) the concepts behind confidence intervals and hypothesis tests are emphasized again and again; (3) notation is simplified; and (4) the errors in arithmetic have been corrected.

I am especially indebted to Dr. Peter Nemenyi who suggested many of the new and interesting problems. Special thanks are due to others who read and criticized the first edition: Katharine J. Kharas, Karen Rappaport, Donald L. Evans, Stephen B. Vardeman, Neil W. Henry, Martin D. Fraser, Michael Jacobs, and Fred Morgan.

NORMA M. GILBERT

CONTENTS

CHAPTER 1

WHAT MATHEMATICAL BACKGROUND IS NEEDED?	1
1.1 Introduction	1
1.2 The Order of Operations	2
1.3 Σ Notation	3
1.4 Exercises	4
1.5 Formulas	5
1.6 Exercises	7
1.7 Greek Letters	9
1.8 Vocabulary and Symbols	10
1.9 Review Exercises	10

CHAPTER 2

INTRODUCTION TO STATISTICS	13
2.1 What Is Statistics?	13
2.2 Populations and Samples	14
2.3 Exercises	15
2.4 Formulating a Statistical Experiment	18
2.5 Random Samples	19
2.6 Other Methods of Sampling	22
2.7 Exercises	23
2.8 Vocabulary	26
2.9 Review Exercises	26

CHAPTER 3

PRESENTING DATA IN A TABLE OR GRAPH	29
3.1 Types of Data	29

3.2	Presentation of Data in a Table	30
3.3	How to Group Scores	32
3.4	Exercises	35
3.5	Graphical Presentation of Data	38
3.6	Exercises	41
3.7	Vocabulary and Symbols	48
3.8	Review Exercises	49

CHAPTER 4

MEASURES OF CENTER, LOCATION, AND VARIABILITY 54

4.1	Introduction	54
4.2	The Mode	55
4.3	The Median	56
4.4	The Mean	57
4.5	Exercises	61
4.6	Quartiles, Percentiles, and Percentile Rank	65
4.7	Exercises	66
4.8	Introduction to Variability	68
4.9	The Range	68
4.10	Variance and Standard Deviation	69
4.11	Interpretation of the Standard Deviation	75
4.12	Exercises	76
4.13	Standard Scores, or Scores in Standard Deviation Units	78
4.14	Exercises	81
4.15	Vocabulary and Symbols	83
4.16	Review Exercises	84

CHAPTER 5

PROBABILITY 88

5.1	Why Study Probability?	88
5.2	Introduction to Probability	89
5.3	Exercises	95
5.4	Mutually Exclusive Events	98
5.5	Axioms of Probability	100
5.6	$Pr(A \text{ or } B)$ If A and B Are Not Mutually Exclusive	101
5.7	Exercises	103
5.8	Independent and Dependent Events, $Pr(A \text{ and } B)$	105
5.9	Exercises	109
5.10	Vocabulary and Symbols	111
5.11	Review	111

CHAPTER 6

DISCRETE PROBABILITY DISTRIBUTIONS 117

6.1	Random Variables	117
6.2	Discrete Probability Distributions	119
6.3	Exercises	122
6.4	Assumptions for a Binomial Distribution	124
6.5	$Pr(X \text{ Successes in } n \text{ Trials})$	124
6.6	The Mean and Standard Deviation of the Binomial Distribution	127
6.7	Applications of the Binomial Distribution	127
6.8	Exercises	128
6.9	Vocabulary and Symbols	130
6.10	Review Exercises	131

CHAPTER 7

NORMAL DISTRIBUTIONS 134

7.1	Histograms of Binomial Distributions	134
7.2	Normal Curves	136
7.3	The Standard Normal Curve	136
7.4	Areas Under the Normal Curve	138
7.5	Percentages of Scores Within 1, 2, and 3 Standard Deviations of the Mean	141
7.6	Exercises	142
7.7	Normal Distributions in General	143
7.8	Area Under Any Normal Curve	144
7.9	Exercises	147
7.10	Normal Approximation to the Binomial Distribution	150
7.11	Exercises	151
7.12	Vocabulary and Symbols	152
7.13	Review Exercises	152

CHAPTER 8

CONFIDENCE INTERVALS FOR MEANS 155

8.1	The Sampling Distribution of Means	155
8.2	Uses of the Sampling Distribution of Means	158
8.3	Exercises	160
8.4	Statistical Inference	162
8.5	Estimates of the Population Mean If σ Is Known	163
8.6	Exercises	168
8.7	Interval Estimates of μ When σ Is Unknown and $n \geq 30$	170
8.8	Exercises	172

8.9	Vocabulary and Notation	173
8.10	Review Exercises	174

CHAPTER 9

INTERVAL ESTIMATES OF PROPORTIONS, DIFFERENCE OF MEANS, AND DIFFERENCE OF PROPORTIONS; SAMPLE SIZE	177
---	-----

9.1	The Sampling Distribution of Proportions	178
9.2	Interval Estimates of Proportions	180
9.3	Exercises	183
9.4	The Sampling Distribution of Differences of Means	186
9.5	Use of the Sampling Distribution of Differences of Means	187
9.6	Sampling Distribution of Difference of Proportions	190
9.7	Summary of Formulas for Statistical Inference	194
9.8	Exercises	194
9.9	Size of a Sample	197
9.10	Sample Size for Estimating a Population Mean	198
9.11	Sample Size for Proportions	200
9.12	Exercises	201
9.13	Vocabulary and Notation	202
9.14	Review Exercises	202

CHAPTER 10

HYPOTHESIS TESTING	207
--------------------	-----

10.1	Introduction	207
10.2	Choosing the Hypothesis To Be Tested (H_0)	208
10.3	Choosing the Alternative Hypothesis (H_1)	210
10.4	Exercises	211
10.5	Choosing a Rule for Making a Decision	212
10.6	Exercises	218
10.7	One-Tail Tests	221
10.8	Exercises	224
10.9	Type I and Type II Errors	227
10.10	Warnings	231
10.11	Exercises	232
10.12	Vocabulary and Notation	234
10.13	Review Exercises	235

CHAPTER 11

STUDENT'S t DISTRIBUTIONS	239
-----------------------------	-----

11.1	The Need for t Distributions	239
------	--------------------------------	-----

11.2	Characteristics of t Distributions	242
11.3	When Is a t Distribution Used?	243
11.4	Degrees of Freedom	243
11.5	Table of t Distributions	244
11.6	Confidence Intervals for, and Tests of Hypotheses About, the Mean	245
11.7	Exercises	247
11.8	Differences of Means for Independent Samples	250
11.9	Exercises	252
11.10	Differences Between Means of Dependent Samples	255
11.11	Exercises	257
11.12	Vocabulary and Symbols	259
11.13	Exercises	259

CHAPTER 12

χ^2 (CHI SQUARE)	265
-----------------------	-----

12.1	Introduction	265
12.2	χ^2 Distributions	267
12.3	Testing “Goodness of Fit” with χ^2	269
12.4	Exercises	271
12.5	Related and Independent Variables; Contingency Tables	274
12.6	Exercises	280
12.7	Vocabulary and Symbols	284
12.8	Review Exercises	284

CHAPTER 13

ANALYSIS OF VARIANCE	288
----------------------	-----

13.1	F Distributions	288
13.2	How to Use an F Table	289
13.3	Use of F Distributions in Testing Hypotheses About Variance	291
13.4	Exercises	293
13.5	Introduction to Analysis of Variance	295
13.6	Where Do the Computational Formulas Come From?	302
13.7	Exercises	303
13.8	Which Means Are Different?	307
13.9	Exercises	309
13.10	Vocabulary and Symbols	310
13.11	Review Exercises	310

CHAPTER 14

LINEAR REGRESSION AND CORRELATION 314

14.1	Introduction	314
14.2	Graph of a Straight Line	315
14.3	The Slope of a Line	317
14.4	The Equation of a Line Through Two Given Points	320
14.5	Exercises	322
14.6	The Regression Equation	325
14.7	Exercises	330
14.8	Pearson Correlation Coefficient: Rough Estimates	335
14.9	Error, Explained, and Total Variations	337
14.10	Computational Formulas for r	341
14.11	Some Warnings	343
14.12	Exercises	346
14.13	The Relation Between Regression and Correlation	350
14.14	Vocabulary and Symbols	351
14.15	Review Exercises	351

CHAPTER 15

LINEAR REGRESSION AND CORRELATION: STATISTICAL INFERENCE 357

15.1	Inferences About Regression: Assumptions About the Population	357
15.2	Testing $H_0: B = 0$	358
15.3	Confidence and Prediction Intervals for the Regression Equation	360
15.4	Exercises	366
15.5	The Sampling Distribution of r	369
15.6	Testing $H_0: \rho = 0$	370
15.7	Testing $H_0: \rho = A, A \neq 0$	372
15.8	Exercises	374
15.9	Vocabulary and Symbols	376
15.10	Review Exercises	376

CHAPTER 16

NONPARAMETRIC TESTS 380

16.1	Introduction	380
16.2	The Sign Test	381
16.3	The Mann-Whitney Test	383
16.4	Exercises	385
16.5	Spearman Rank Correlation Coefficient	389

16.6	Exercises	392
16.7	Vocabulary and Symbols	395
16.8	Review Exercises	396

APPENDIX A

SYMBOLS	399
---------------	-----

APPENDIX B

FORMULAS	402
----------------	-----

APPENDIX C

TABLES	407
--------------	-----

1	Squares and Square Roots	407
2	Binomial Probabilities	414
3	Random Numbers	419
4	Areas Under the Normal Curve	420
5	Student's t Distributions	421
6	χ^2 Distributions	422
7	F Distributions	423
a	$\alpha = .05$	423
b	$\alpha = .025$	425
c	$\alpha = .01$	427
8	Transformation of r to Z and ρ to μ_z Values	429
9	Critical Values of U in the Mann-Whitney Test	430

INDEX	431
-------------	-----

1

WHAT MATHEMATICAL BACKGROUND IS NEEDED?

Most of this chapter should be familiar stuff. Summation notation (Section 1.3, pages 3–4) may be new to you. You will use the remaining parts of this chapter so frequently, however, that you will be foolish indeed if you let anything slip by without understanding it. Even students who are good at mathematics are sometimes confused about whether multiplication or addition is carried out first [$4 + 2(8) = 4 + 16$ or $6(8)?$] and how parentheses are used to change the usual order.

1.1 INTRODUCTION

Many students approach statistics with apprehension, having resolved after junior year in high school never to take another mathematics course. But be of good cheer for two reasons. The background in mathematics which you need is remarkably slight. And in your high school courses you may have learned how to solve quadratic equations, for example, without understanding *why* you were learning to do this. Now those of you who are concerned about using mathematics are studying statistics because a knowledge of this material is necessary background for your study of psychology, economics, zoology, or sociology. You will find an almost immediate application of the mathematics you learn to a subject that interests you, and this delightful situation—in contrast to your past experience—will give most of you a new and splendid ability to cope with mathematics.

Try always to gain some perception of the general arguments that produce results rather than accept them blindly. An idiot can be trained (or a computer can be programmed) to plug numbers into formulas; you must learn when and why a particular formula is used. Occasionally an explanation, in order to be mathematically proper, would require that you know calculus; in these cases the explanation is purely intuitive or is omitted. Fair warning will be given when such occasions arise.

A special notation will be used to warn you of traps that have been set for you: an unusual twist, a problem that cannot be solved, or some special reason to be wary.



1.2 THE ORDER OF OPERATIONS

The word “operations” here refers to squaring, multiplication or division, and addition or subtraction. The rules agreed on by mathematicians are that these are carried out in the order given below when more than one operation is to be performed:

- (a) First square.
- (b) Then multiply or divide.
- (c) Then add or subtract.

Example 1

$$2(3) + 4 = ?$$

$$2^2(3) + 1 = ?$$

$$3 - \frac{12}{2^2} = ?$$

$$2(3) + 4 = 6 + 4 = 10.$$

$$2^2(3) + 1 = 4(3) + 1 = 12 + 1 = 13.$$

$$3 - \frac{12}{2^2} = 3 - \frac{12}{4} = 3 - 3 = 0.$$

But suppose you don't want to do things in that order? What if, for example, 2 and 3 are first to be added and then the result squared? Parentheses (), square brackets [], or curly brackets { } are used, and a new rule is added:

If there are parentheses or brackets, simplify what is inside them first before continuing.

Example 2

$$6(4 + 1)^2 = ?$$

$$[2^2(3) - 2]^2(4 - 1) = ?$$

$$2(3^2 + 4 \div 2) - 3(2^2 - 4)^2 = ?$$

$$6(4 + 1)^2 = 6(5)^2 = 6(25) = 150.$$

$$[2^2(3) - 2]^2(4 - 1) = [4(3) - 2]^2(3) = [12 - 2]^2(3) = 10^2(3) = 100(3) = 300.$$

$$\begin{aligned} 2(3^2 + 4 \div 2) - 3(2^2 - 4)^2 &= 2(9 + 4 \div 2) - 3(4 - 4)^2 \\ &= 2(9 + 2) - 3(0)^2 \\ &= 2(11) - 0 = 22. \end{aligned}$$

1.3 Σ NOTATION

You will soon need to learn seven Greek letters. The first of these is the Greek (capital) S, called **sigma**, and written Σ . We shall use Σ as an abbreviation for the word “sum” or for the phrase “the sum of _____.”

Notation: Σ = Sum.

If X values are 1, 2, 3, 5, then $\Sigma X = 1 + 2 + 3 + 5 = 11$. Add all the X values.

If the weights W of five students are 110, 100, 160, 200, 190 pounds, then $\Sigma W = 110 + 100 + 160 + 200 + 190 = 760$ pounds.

Remember the rule for order of operations when Σ (standing for addition) is combined with exponents, multiplication, or parentheses.

Example 1 If Y values are 4, 3, 1, 6, 2, find (a) ΣY , (b) ΣY^2 , (c) $(\Sigma Y)^2$.

$$(a) \Sigma Y = 4 + 3 + 1 + 6 + 2 = 16.$$

$$(b) \Sigma Y^2 = 4^2 + 3^2 + 1^2 + 6^2 + 2^2 = 16 + 9 + 1 + 36 + 4 = 66.$$

$$(c) (\Sigma Y)^2 = 16^2 = 256.$$

There are three rules for working with summations:

Rule 1. $\Sigma(X + Y) = \Sigma X + \Sigma Y$. (This rule only makes sense if the number of X values is the same as the number of Y values.)

Rule 2. If k is a constant, $\Sigma kX = k\Sigma X$.

Rule 3. If A is a constant, $\Sigma A = nA$, where n is the number of values that are being added.

Example 2 X values: 4, 1, 6, 3; corresponding Y values: 2, 3, 7, -1 . (a) $\Sigma(X + Y) = ?$

$$(b) \Sigma 2X = ? \quad (c) \Sigma 5 = ? \quad (d) \Sigma(X - Y) = ?$$

$$(a) \Sigma(X + Y) = \Sigma X + \Sigma Y = (4 + 1 + 6 + 3) + (2 + 3 + 7 - 1) = 14 + 11 = 25.$$

(Rule 1)

$$(b) \Sigma 2X = 8 + 2 + 12 + 6 = 28, \text{ but also} \\ = 2\Sigma X = 2(4 + 1 + 6 + 3) = 2(14) = 28.$$

(Rule 2, which says a constant can be factored out)

(c) $\Sigma 5 = 4(5) = 20$. (Rule 3) It is assumed that the number of values being added (here, 4) is known from the context of the problem. A more sophisticated summation notation, using subscripts, can be adopted, but we shall not need it.

$$\begin{aligned}
 \text{(d) } \Sigma(X - Y) &= \Sigma[X + (-1)Y] \\
 &= \Sigma X + \Sigma(-1)Y && \text{(Rule 1)} \\
 &= \Sigma X + (-1)\Sigma Y && \text{(Rule 2)} \\
 &= \Sigma X - \Sigma Y \\
 &= (4 + 1 + 6 + 3) - (2 + 3 + 7 - 1) \\
 &= 14 - 11 \\
 &= 3.
 \end{aligned}$$

1.4 EXERCISES

1. $4.38 + .12 + .02 =$

2. $6.21(.01) =$

3. $5(0) =$

4. $\frac{27(75)}{25(9)} =$

5. $\frac{25 + 75}{25 + 50} =$

6. Add: $-10, 5, -3, 2, 3, -5, 2, 10$.

7. Add: $-10, -2, 5, 3, 4, 12, -8, -6$.

8. $3(5^2) + 5 =$

9. $3(5^2 + 5) =$

10. $4 - \frac{10^2}{20} =$

11. $3(1 - 2)^2 - 4^2(2^2 - 3) =$

12. $2^2(2^2 + 1)^2 + 3(1 + 3)^2 =$

13. X values: 2, 5, 3. Find (a) ΣX , (b) ΣX^2 , (c) $(\Sigma X)^2$, (d) $\Sigma 3X$, (e) $\Sigma 2X^2$, (f) $\Sigma 4$.

14. Y values: 4, 1, 1, 3. Find (a) ΣY , (b) $\Sigma 4Y$, (c) $(\Sigma Y)^2$, (d) ΣY^2 , (e) $\Sigma 3$.

15. Z values: 2, -1 , 0, 6, 3. Find (a) ΣZ , (b) ΣZ^2 , (c) $\Sigma 5$, (d) $(\Sigma Z)^2$, (e) $\Sigma(-Z)^2$.

16. $\frac{f}{3} \quad \frac{X}{-2}$ Find (a) ΣfX , (b) ΣfX^2 , (c) $(\Sigma fX)^2$.
 $\begin{array}{cc} 4 & 5 \\ 2 & 3 \end{array}$

17. X : 2, -1 , 0, 3. Find $\Sigma X^2 - \frac{(\Sigma X)^2}{4}$.

18. $\frac{f}{2} \quad \frac{Y}{0}$ Find (a) ΣfY , (b) $(\Sigma fY)^2$, (c) ΣfY^2 .
 $\begin{array}{cc} 9 & 1 \\ 3 & 4 \\ 1 & 8 \end{array}$

19. X values are 3, 5, 8, 2, 0, 3, 0; corresponding Y values are 2, 3, 1, -1, 0, 2, 4. Find (a) ΣX , (b) ΣY , (c) $\Sigma(X + Y)$, (e) $(\Sigma X)^2$.

ANSWERS

1. 4.52 2. .0621 3. 0
4. $3(3) = 9$ ("Cancel" first; not $\frac{2025}{225} = 9$.)
5. $\frac{100}{75} = \frac{4}{3}$ (Don't try to cancel the initial 25's.)
6. Cancel -10 and 10, -5 and 5, -3 and 3. $2 + 2 = 4$.
7. Cancel -10, -2, and 12; 5, 3, and -8. $4 - 6 = -2$; OR add the positive numbers and negative numbers separately: $24 - 26 = -2$.
8. $3(25) + 5 = 75 + 5 = 80$.
9. $3(25 + 5) = 3(30) = 90$.
10. $4 - \frac{100}{20} = 4 - 5 = -1$.
11. $3(-1)^2 - 16(4 - 3) = 3(1) - 16(1) = -13$.
12. $4(4 + 1)^2 + 3(4)^2 = 4(5)^2 + 3(16) = 4(25) + 48 = 100 + 48 = 148$.
13. (a) 10, (b) $4 + 25 + 9 = 38$, (c) $10^2 = 100$, (d) $3(10) = 30$, (e) $2(38) = 76$, (f) $3(4) = 12$.
14. (a) 9, (b) $4(9) = 36$, (c) $9^2 = 81$, (d) $16 + 1 + 1 + 9 = 27$, (e) $4(3) = 12$.
15. (a) 10, (b) $4 + 1 + 0 + 36 + 9 = 50$, (c) $5(5) = 25$, (d) $10^2 = 100$, (e) 50.
16. (a) (Add a ΣfX column) $6 + 20 + 6 = 32$, (b) (Add a ΣfX^2 column) $12 + 100 + 18 = 130$, (c) $32^2 = 1024$.
17. $14 - \frac{4^2}{4} = 10$.
18. (a) $0 + 9 + 12 + 8 = 29$, (b) $29^2 = 841$, (c) $0 + 9 + 48 + 64 = 121$.
19. (a) 21, (b) 11, (c) $21 + 11 = 32$, (d) $21 - 11 = 10$, (e) $21^2 = 441$.

1.5 FORMULAS

A number of formulas will be developed in this course; they simply give shorthand directions for carrying out operations. Study the following examples, and learn how to translate from English into mathematical language and vice versa.

Example 1

Translate into English: $\Sigma(X - 4)$.

Subtract 4 from each X value, then add the new set.

Example 2

Write a formula that gives directions to subtract 4 from each X value, square each number in the new set, and then add the squares.

$$\Sigma(X - 4)^2$$