

# Statistics

*A Tutorial Workbook*



**Robert Pisani**

# Statistics

*A Tutorial Workbook*

ROBERT PISANI

UNIVERSITY OF CALIFORNIA · BERKELEY



W · W · NORTON & COMPANY

NEW YORK · LONDON

Copyright © 1985 by W. W. Norton & Company, Inc.

All rights reserved.

Published simultaneously in Canada by Penguin Books Canada Ltd.,  
2801 John Street, Markham, Ontario L3R 1B4.

Printed in the United States of America.

W. W. Norton & Company, Inc.,  
500 Fifth Avenue, New York, N. Y. 10110

W. W. Norton & Company, Ltd.,  
37 Great Russell Street, London WC1B 3NU

ISBN 0-393-95457-9

3 4 5 6 7 8 9 0

# Preface

This workbook is a companion volume to the text

Statistics by D. Freedman, R. Pisani and R. Purves  
(W. W. Norton & Co., New York, 1978)

It is meant to be used in parallel with that text, which is called “FPP” or “the main text” throughout this book. The workbook treats chapter-by-chapter the same topics found in the main text—in some cases it covers additional related material or ideas only suggested in the main text, with the object of deepening the student’s understanding. With the exception of the sections on conditional probability in Chapters 14 and 15 of the workbook, none of the additional material should be omitted. Conditional probability may be skipped if the instructor wishes, although the treatment in this workbook makes it especially easy and transparent.

Each chapter is organized as follows. There is a series of instructional sections which outline the major activities of the chapter. Each of these sections contains some worked examples, followed by a number of similar problems which the student should solve using the technique discussed in the section. The answers to all of these problems are found at the end of the chapter. After these sections on major activities comes a longer problem section, called “Exercises”. The method of solution is not indicated in this section—if the student has read and worked all the problems in the first series of sections, he or she should be able to recognize the problem type and apply the appropriate solution method. Solutions to odd-numbered Exercises are found at the end of the chapter; solutions to the even-numbered Exercises are found in the Instructor’s Manual, so that these problems may be used for homeworks, quizzes and exams.

## Acknowledgments

Many thanks to Shel Silver and David Douglas for helping me launch this project, to countless students who have provided questions that needed to be asked, to Dan Coster for thoughtful proofreading and editorial assistance, to Doug Cooper for many helpful suggestions, and mostly to Yvette Malamud, my editor, who labored long, hard and faithfully and without whom this book would never have made it to press.

# To The Student

## How to Use this workbook

First, read the Preface. It tells you how the workbook is organized. You should use this workbook chapter by chapter in parallel with the main book. It will help you better understand the material. It will also give you many problems on which you can test your understanding.

After reading a chapter in the main book, turn your attention to the same chapter in the workbook. Work through this workbook chapter. You should have no trouble with the problems immediately following the examples. After you have mastered those, try each of the Exercises. Then re-read the chapter in the main book.

## Mathematical Preparations

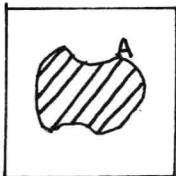
before you begin work in elementary statistics, you should make certain that you have the necessary quantitative skills. You will not need any calculus or algebra—neither FPP nor the workbook use anything more than simple arithmetic and the most rudimentary word-formulas. What you will need, however, is a good feeling for numbers and for the operations of ordinary arithmetic.

The following diagnostic quiz will help you decide if you are adequately prepared. In each case, circle the choice which *most closely* answers the question. Although you may not be able to determine the exact answer, only one of the multiple choice answers will be reasonable. You should be able to answer all these questions without calculating. Give yourself 20 minutes to answer all the questions.

1.  $5/16$  is closest to: a)  $1/2$ ; b)  $1/3$ ; c)  $1/4$ ; d)  $1/5$ .
2. The cube of  $1/3$  is: a)  $3/9$ ; b)  $3/27$ ; c)  $1/81$ ; d)  $1/27$ .
3. Of the following sequences of fractions, what set is arranged in increasing order?
  - a)  $7/12$ ,  $5/6$ ,  $2/3$ ,  $3/4$
  - b)  $7/12$ ,  $7/11$ ,  $8/11$ ,  $4/5$
  - c)  $7/12$ ,  $8/11$ ,  $8/12$ ,  $9/12$
  - d)  $7/12$ ,  $1/2$ ,  $7/15$ ,  $1/3$
4.  $13/38$  is approximately: a) 5%; b) 15%; c) 25%; d) 35%; e) 45%; f) 55%; g) 65%; h) 75%; i) 85%; j) 95%.

## To The Student

5. When 0.3 is divided by  $1/4$ , the answer is approximately: a) 0.75; b) 0.075; c) 1.2; d) 0.12.
6. Of the following, the value closest to that of  $(42.10 \times 0.0003)/0.002$  is: a) 0.063; b) 0.63; c) 6.3; d) 63.0.
7. The number  $(1.20672 \times 2.00012)/0.0502698$  is:  
a) nearly but not quite 50  
b) somewhat more than 50  
c) slightly less than 100  
d) more than 100
8. 72.2376% of 416.9327 is approximately: a) 200; b) 300; c) 400; d) 500.
9. A is older than B. With the passage of time, the ratio of the age of A to the age of B: a) remains the same; b) increases; c) decreases.
10. 300 is what percent of 1500? a) 2%; b) 5%; c) 10%; d) 15%; e) 20%; f) 25%; g) 30%; h) 50%; i) 75%
11. 0.125 written as a percent is: a)  $1/8\%$ ; b) 0.125%; c) 12.5%; d) 125%.



A is the shaded region in the above square.

12. What is the approximate ratio of the area of A to the area of the square?  
a) 0.1; b) 0.3; c) 0.7; d) 0.9.

13. Of the following, which best describes the proportion of A that is in the left half of the square? a)  $23/71$ ; b)  $1/71$ ; c)  $70/71$ ; d)  $37/71$

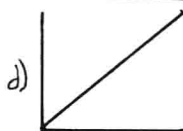
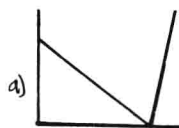
A car is driven at constant speed from San Francisco to Los Angeles by way of Palo Alto. Here are four factors of interest to the driver.

14. The distance from San Francisco after one hour  
15. The time required to go 100 miles  
16. The distance from Palo Alto after one hour  
17. The distance from Los Angeles after one hour

Following are six graphs. In each, the horizontal axis represents the speed of the car. In four of the graphs, the vertical axis is one of the factors listed above. In questions 14 through 17, match the factor with the

## To The Student

proper graph (a, b, c, d, e, or f).



18. Of the following, the pair that is *not* a set of equivalents is:  
a) 0.021%, 0.00021; b) 1/4%, 0.0025; c) 1.5%, 3/200; d) 225%, 0.225
19. One-fourth percent of 360 is: a) 0.09; b) 0.9; c) 9.0; d) 90.
20. The relationships between 0.01% and 0.1 is: a) 1 to 10; b) 1 to 100; c) 1 to 1,000; d) 1 to 10,000.
21. Vodka is 40% alcohol. If one quart of vodka is mixed with 0.6 quarts of orange juice, what percentage of the mixture is alcohol?  
a) 15%; b) 20%; c) 34%; d) 25%; e) 6¾%; f) cannot be determined from the data given
22. Four quarts of a certain mixture of alcohol and water is at 50% strength. To it is added a quart of water. The alcohol strength of the new mixture is: a) 12.5%; b) 20%; c) 25%; d) 40%.
23. The square root of 100,000 is about: a) 30; b) 300; c) 1,000; d) 3,000; e) cannot tell.
24. The closest approximation to the square root of 2/5 is: a) 0.65; b) 1/5; c) 0.16; d) 0.4.

Now grade yourself—the answers are: 1b; 2d; 3b; 4d; 5c; 6c; 7a; 8b; 9c; 10e; 11c; 12b; 13d; 14d; 15e; 16b; 17c; 18d; 19b; 20c; 21d; 22d; 23b; 24a.

If you scored less than 16, you need more preparation. I recommend that you obtain a copy of Mathematics for Statistics by W. L. Bashaw (John Wiley & Co.). Chapters 1 through 6, 18, 19, 21 and 22 will give you the information and practice you need. You should master all of this material *before* you begin the main book or the workbook.

## To the Student

### **A Note on the Approach**

FPP is unusual in its approach to statistics. It attempts to give you a real understanding of the ideas of the subject. It avoids formulas and focuses on reasoning and intuition. On the face of it, the book looks a lot less imposing than most mathematics textbooks. However, there is a price to pay—in the absence of formulas you must learn to think. In some ways, this book is harder than a book filled with formulas and equations. So don't be deceived: although many pages are mostly pictures and prose, they may require a lot of your attention.

Many of the problems in Chapters 3 through 12 emphasize a geometrical approach to the material. It is important to have a good grip on this—if you do, many relationships that otherwise might puzzle you will become clear. I encourage you not to neglect these problems. Work with your pencil and draw all the pictures suggested in the workbook—the reward will be a solid understanding which will carry you beyond memorized procedures.



# Contents

<b>Preface</b>	<b>viii</b>
<b>To the Student</b>	<b>ix</b>
<b>Chapter 1. Controlled Experiments</b>	<b>1</b>
I. Exercises	1
II. Solutions	1
<b>Chapter 2. Observational Studies</b>	<b>2</b>
I. Exercises	2
II. Solutions	3
<b>Chapter 3. The Histogram</b>	<b>4</b>
I. Reading Histograms	4
II. Drawing Histograms	7
III. Exercises	14
IV. Solutions	18
<b>Chapter 4. The Average and the Standard Deviation</b>	<b>22</b>
I. Computing the Average	22
II. Finding the Sum from the Average	23
III. Locating the Average on a Histogram	24
IV. Computing the Median	25
V. Locating the Median on a Histogram	26
VI. The Average and the Median	28
VII. Computing the RMS	29
VIII. Estimating the RMS	30
IX. Computing the SD	30
X. Estimating the SD from a List	31
XI. Estimating the SD from a Histogram	32
XII. Exercises	33
XIII. Solutions	40
<b>Chapter 5. The Normal Approximation for Data</b>	<b>44</b>
I. Thinking in Standard Units	44
II. To Convert a Number to Standard Units	45
III. Finding Areas under the Normal Curve	46
IV. The Normal Approximation for Data	51
V. Percentiles	53

VI. Finding Percentiles on the Normal Curve	53
VII. Interquartile Range	55
VIII. Interquartile Range for the Normal Curve	56
IX. Exercises	56
X. Solutions	63
<b>Chapter 6. Measurement Error</b>	<b>67</b>
<b>Chapter 7. Graphing</b>	<b>68</b>
I. Reading Points off a Graph	68
II. Slope and Intercept	69
III. Solutions	70
<b>Chapter 8. Correlation</b>	<b>71</b>
I. Choosing the Dependent Variable	71
II. Locating the Average and SD on a Scatter Diagram	72
III. Estimating the Correlation Coefficient from a Scatter Diagram	73
IV. Making Judgements about the Correlation Coefficient	74
V. Computing the Correlation Coefficient	75
VI. Exercises	75
VII. Solutions	84
<b>Chapter 9. More about Correlation</b>	<b>88</b>
I. Exercises	88
II. Solutions	88
<b>Chapter 10. Regression</b>	<b>89</b>
I. Computing the Regression Estimate	89
II. The Regression Fallacy	92
III. Exercises	94
IV. Solutions	96
<b>Chapter 11. The RMS Error for Regression</b>	<b>98</b>
I. Computing the RMS Error	98
II. Visualizing the RMS Error	99
III. Computing the New Average and the New SD	100
IV. Approximating the New Average and the New SD Geometrically	102
V. Computing Percentages in Vertical Strips	102
VI. Approximating Percentages in Vertical Strips Geometrically	104
VII. Exercises	105
VIII. Solutions	108

<b>Chapter 12. The Regression Line</b>	<b>109</b>
I. Computing Slope	109
II. Interpreting the Slope	110
III. Computing the Intercept	110
IV. How to Interpret the Intercept	111
V. Exercises	112
VI. Solutions	115
<b>Chapter 13. What Are the Chances?</b>	<b>116</b>
I. Long Run	117
II. Estimation of Chances	118
III. Incompatible Events	119
IV. Addition Rule	119
V. Independent Events	121
VI. Multiplication Rule	122
VII. Multiplication Rule Without Replacement	123
VIII. Rule of Opposites	125
IX. Exercises	126
X. Solutions	130
<b>Chapter 14. More about Chance</b>	<b>135</b>
I. Outcome Lists	135
II. Exercises	138
III. Conditional Probability	139
IV. Solutions	141
<b>Chapter 15. The Binomial Coefficients</b>	<b>145</b>
I. Counting	145
II. How to Recognize a Binomial Experiment	147
III. Simple Binomial	148
IV. Multiple Binomial	150
V. Sign Test	151
VI. Exercises	152
VII. Solutions	154
<b>Chapter 16. The Law of Averages</b>	<b>156</b>
I. Long Run/Short Run — The Law of Averages Simplified	156
II. Box Models	157
III. Solutions	160
<b>Chapter 17. The Expected Value and Standard Error</b>	<b>161</b>
I. Computing the Expected Value (EV)	161
II. Computing the Standard Error (SE)	162

III. Interpreting the SE	163
IV. Normal Curve	164
V. Computing the SD of a 0–1 Box	166
VI. Exercises	166
VII. Solutions	169
<b>Chapter 18. The Normal Approximation for Probability Histograms</b>	<b>172</b>
I. The Continuity Correction	172
II. Normal Approximation, Ranges	173
III. Normal Approximation, Exact Numbers	174
IV. Exercises	174
V. Solutions	176
<b>Chapter 19. Sample Surveys</b>	<b>178</b>
<b>Chapter 20. Chance Errors in Sampling</b>	<b>180</b>
I. Applying the Square Root Law to Percentages	180
II. Computing the SE of a Percentage	181
III. The Correction Factor When Drawing Without Replacement	183
IV. To Compute the SE When Drawing Without Replacement	184
V. Exercises	185
VI. Solutions	187
<b>Chapter 21. The Accuracy of Percentages</b>	<b>190</b>
I. An Estimate of the SE When the Contents of the Box Are Unknown	190
II. To Construct an Approximate X-% Confidence Interval	191
III. Exercises	192
IV. Solutions	193
<b>Chapter 22. Measuring Employment and Unemployment</b>	<b>195</b>
<b>Chapter 23. The Accuracy of Averages</b>	<b>196</b>
I. Computing the EV of the Average of Draws	196
II. Computing the SE of the Average of Draws	197
III. Confidence Intervals for the Average of the Box	198
IV. Exercises	199
V. Solutions	200
<b>Chapter 24. A Model for Measurement Error</b>	<b>201</b>
I. Exercises	201
II. Solutions	203

<b>Chapter 25. Chance Models in Genetics</b>	<b>204</b>
I. Exercises	204
II. Solutions	205
<b>Chapter 26. &amp; 27. Tests of Significance</b>	<b>206</b>
I. Exercises	206
II. Solutions	208
<b>Chapter 28. The Chi-Square Test</b>	<b>209</b>
I. Exercises	209
II. Solutions	212
A Short Normal Table	213

# 1

## Controlled Experiments

### I. Exercises

**1.1** An experimental program to improve reading skills in school children is set up with an aftertest to determine its effect. The children are to be divided into two groups: one to receive instruction in the program and the other to receive only their ordinary reading instruction. Describe a method for deciding how to allocate the children to the “treatment” group (the experimental program) and the control group (with ordinary reading instruction) which will not introduce any bias.

**1.2** It is conjectured that a food supplement improves the intelligence of rats as measured by their ability to run a maze. From a cage containing twenty test rats, ten are chosen to belong to the control group and ten are chosen to belong to the treatment group. The ten for the control group are chosen by the lab assistant who pulls out the ten rats nearest the cage door. What kind of bias might be introduced by this selection process?

### II. Solutions

**1.1** The children should be assigned to treatment and control groups at random. You could for instance put all their names in a box and draw out half the names at random—then assign them to the treatment group and those whose names weren’t drawn to the control group.

# 2

## Observational Studies

### I. Exercises

2.1 An advertisement states that according to a study conducted by an automobile association 90% of all Volvo trucks sold in the last ten years are still on the road. Does this necessarily mean that Volvo trucks last a long time or could there be another explanation?

2.2 In 1979 surveys showed that in Alcoholics Anonymous there are now half as many members under the age of 30 as there were three years ago. This may be due to a decrease in alcoholism among those under 30. Can you think of another contributing factor?

2.3 An English survey of 3000 medical records showed that smokers are more inclined to get depressed than non-smokers. Does this necessarily imply that smoking causes depression?

2.4 Hospital records show that twice as many men as women fall out of hospital beds. Does this indicate that men are clumsier than women or is there another possible explanation?

2.5 Studies show that in the period from 1850 to 1900 in the United States the average marriage lasted only 12 years. Does this show that the divorce rate was high in that period?

2.6 In an investigation of the incidence of heart disease in men it is noticed that the fathers of one-child families have had on the average somewhat fewer heart attacks than the fathers of five-child families. It is concluded that having big families increases one's chance of attack. Do these facts support the conclusion?

2.7 The usual method of catching tuna with nets accidentally results in the capture (and death) of dolphins. In response to reports that the dolphin is in danger of extinction a new kind of net for catching tuna is introduced. It is tried for one year and the percentage of dolphin out of the total tonnage caught is compared with the average percentage caught over the preceding ten-year period using the old method. A large decrease in the percentage of captured dolphins is observed. This is for-

## Solutions

warded as evidence that the new method helps to protect dolphins. Suggest an alternate interpretation.

2.8 Columnist Ann Landers was asked whether having children was worth the problems involved. She asked her readers, "If you had to do it over again would you have children"? A few weeks later her column was headlined, "*Seventy Percent of Parents Say Kids Not Worth It*", because 70% of the parents who wrote said they would not have children if they could make the choice again. Is this a valid conclusion? Comment.

### II. Solutions

2.1 It could be that most trucks were sold this year or in recent years. Or Volvo drivers could use their trucks very little. There are many other possible explanations.

2.3 No—it may be that people with a tendency towards depression also tend to smoke.

2.5 No—the shortened marriages were mostly due to the death of one of the partners.

2.7 It may be that the number of dolphins in the sea has already been reduced by the previous years' catches.



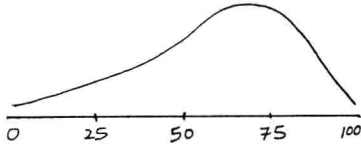
# 3

## The Histogram

### 1. Reading Histograms

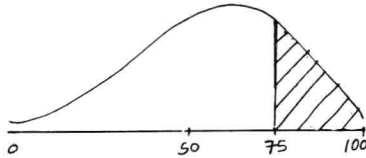
To learn how much of a data set lies between two values, shade in the area above these values on the histogram and estimate the proportion of the histogram that is shaded.

*Example A:* Below is a rough histogram of the test scores in a class.



What percentage scored above 75?

*Solution:* Shading in the area above 75 gives the following picture:



The shaded region appears to be about 30% of the histogram, area-wise. So the answer is: about 30%.

### Problems

3.1 Estimate the percentage scoring below 25.

3.2 Estimate the percentage scoring between 50 and 75.