DENT SOLUTIONS MANUAL

WITH EXERCISES

for THE

ELEMENTS OF STATISTICS

WITH APPLICATIONS TO ECONOMICS AND THE SOCIAL SCIENCES

JAMES B. RAMSEY

Student's Solutions Manual with Exercises for

The Elements of Statistics

with Applications to Economics and the Social Sciences

James B. Ramsey
New York University

DUXBURY
THOMSON LEARNING
Australia • Canada • Mexico • Singapore • Spain • United Kingdom • United States

Assistant Editor: *Ann Day*
Editorial Assistant: *Nathan Day*
Marketing Manager: *Tom Ziolkowski*

Ancillary Coordinator: *Rita Jaramillo*
Print Buyer: *Christopher Burnham*
Permissions Editor: *Robert Kauser*
Printer: *Webcom Limited*

Printed in Canada.
1 2 3 4 5 6 7 05 04 03 02 01

0-534-38769-1

Asia
Thomson Learning
60 Albert Complex, #15-01
Albert Complex
Singapore 189969

Australia
Nelson Thomson Learning
102 Dodds Street
South Street
South Melbourne, Victoria 3205
Australia

Canada
Nelson Thomson Learning
1120 Birchmount Road
Toronto, Ontario M1K 5G4
Canada

Europe/Middle East/South Africa
Thomson Learning
Berkshire House
168-173 High Holborn
London WC1 V7AA
United Kingdom

Latin America
Thomson Learning
Seneca, 53
Colonia Polanco
11560 Mexico D.F.
Mexico

Spain
Paraninfo Thomson Learning
Calle/Magallanes, 25
28015 Madrid, Spain

June 2001

"Special thanks are due to Joe Newton and Jane Harvill who wrote the labs that appear in many of the exercises. Author and student alike owe Enrique Schroth, Guillermo Felices, and Antonio Guarino grateful thanks for providing and checking the solutions to the exercises."

James B. Ramsey
Department of Economics
New York University
New York, NY 10003
James.Ramsey@NYU.edu
http://www.econ.nyu.edu/user/ramseyj/textbook/main.htm

Contents

Student's Solutions Manual with Exercises for

The Elements of Statistics

with Applications to Economics and the Social Sciences

James B. Ramsey
New York University

DUXBURY
THOMSON LEARNING
Australia • Canada • Mexico • Singapore • Spain • United Kingdom • United States

Chapter 1

Statistics as a Science

1.1 Topics for Consideration

There are no answers provided for the exercises in this chapter.

Chapter 2

Types of Variables, Measurements, and Explanation

2.1 Calculation Practice

2.1 *In the following list of model statements, for each term identify whether they are variables, coefficients, or parameters*

a. $C_t = a + b\,Yh_t$; C_t *is aggregate consumption in year* t, Yh_t *is aggregate household income in year* t.

b. $I_t = \alpha_0 + \alpha_1(Y_t - Y_{t-1}) - \delta K_{t-1} + \gamma r_t$; I_t *is aggregate investment in year* t, Y_t *is net national product in year* t, K_t *is aggregate capital stock in year* t, *and* r_t *is the real interest rate.*

c. $S_{it} = c_i + dA_{it}$; S_{it} *is sales of firm* i *in year* t, A_{it} *is advertizing expenditures in year* t *by firm* i.

d. $M_t = e + fB_t - gB_t^2$; M_t *is the infant mortality rate and* B_t *is the birth rate.*

e. $E = Mc^2$; E *is energy,* M *is mass, and* c *is the velocity of light.*

Answer:

a. Variables: C_t, Yh_t; Parameters: a, b.

b. Variables: I_t, Y_t, K_t, r_t; Parameters: α_0, α_1, δ, γ.

c. Variables: S_{it}, A_{it}; Parameters: c_i, d.

d. Variables: M_t, B_t; Parameters: e, f, g.

e. Variables: E, M; Parameter: c^2.

2.3 *In the following list of variables, specify how each should be measured; that is, identify what type of variable it is, or what type of measurement is applicable. There may be more than one alternative; if so, mention at least two.*

a. Temperature by month at specified points on the earth's surface

b. Number of earthquakes per year

c. Severity of earthquakes as indicated by the Richter scale
d. Scholastic aptitude test scores
e. Volume of a gas
f. Rate of arrival of patrons at a cinema
g. Hostility levels of student subjects
h. Reading speed and comprehension score
i. Train use by commuters (note several possibilities)
j. Preference for alternative mouthwashes
k. Cost of production of all automobiles in the United States by year

Answer: Some of the answers are subject to interpretation, particularly whether something is random vs. deterministic.

a. Random? cardinal, continuous, interval scale, time series
b. Cardinal, discrete, random
c. Random? cardinal, continuous, interval scale
d. Random? cardinal, continuous, interval scale
e. Deterministic, cardinal, continuous, ratio scale
f. Random? cardinal, discrete
g. Random? ordinal?
h. Random? cardinal, ratio scale for speed, interval scale for comprehension
i. Random? time series if recorded over time, index, cardinal, discrete
j. Ordinal
k. Time series, index, cardinal, ratio scale

2.2 Exploring the Tools

2.5 *For the following list of variables give your reasons why, or the extent to which, you think that each is a random or a deterministic variable. What criteria would you use to distinguish between the two categories of variables? What experiments would you perform to check your distinctions?*

a. Extensions of a spring after adding weights
b. Picking cards from a shuffled deck (What role does "shuffling" play and why?)
c. Numbers in the lottery
d. Orbital periods of the planets (that is, the times taken to complete one cycle)
e. Consumption of cigarettes by a specific individual
f. Purchase of cigarettes by a specific individual
g. Cost of production for a specific output by a particular firm
h. Service time at an auto repair facility

Answer: Most of these questions are open to interpretation. For example, extensions of a spring by adding weights may be thought of as deterministic because the degree of extension is proportional to the weights added. On the other hand, it may be thought of as observations on a random variable, because with ten readings you will get ten different results if you measure the extension

sufficiently accurately. If the relatively small measurement error is ignored, the process may be regarded as deterministic, but if we do allow for the errors of measurement, the process has to be regarded as structural.

For questions [b] and [c] the simple answer is that the outcomes are random, but one can speculate that with enough care and precise observation it might be possible to determine the outcome of a coin toss, or of numbers on a lottery. We can speculate, but the reality is that such precise observation is beyond our capabilities. Following are interpretations for the above variables.

a. Spring extension satisfies Hooke's law and is deterministic up to measurement error.

b. Given perfect information about the original orientation of the cards in the deck and the exact nature of the shuffle, one could determine exactly which card will be drawn. However, since it is virtually impossible to ascertain such information, one could claim that it is random. Shuffling is what obscures the information which we might have, shuffling is a "randomization" process.

c. A similar situation to the one above. Given a good description of the exact physical interaction of the lottery system (which is directly observable) one could conceivably predict the lottery, however, it is impossible to find this information in practice, rendering the lottery effectively random.

d. The Orbital periods of the planets follow, up to measurement and some negligible error, well-know laws(e.g. Kepler's Laws). It seems that they are deterministic.

e. There is no specific reason that cigarette consumption of a particular individual should be completely determined. It is difficult to imagine, however, that someone who has smoked two packs a day for the last twenty years would somehow stop smoking in one day. If it is to be modeled as a random variable, then local observations in the time series will be highly correlated.

f. Purchase of cigarettes by a particular individual is functionally related to the consumption of cigarettes by that individual, and hence is a similar type of variable, but "random" variation is likely to be more pronounced.

g. The cost of production for some unit of output is functionally related to the cost and efficiency of various inputs, e.g. wages, capital breakdown, worker productivity, etc., but some of these things are not directly observable, hence I can not ascertain them with an arbitrary degree of certainty. Therefore, we may call the cost of production random.

h. Service time is best modeled as a random variable.

2.8 *If you write the equation:*

$$y = a + bx$$

where x is a continuous deterministic variable and y is a discrete deterministic variable, what is immediately wrong with this formulation? Explain your answer.

Answer: If x is a continuous variable, then $y = a + bx$ is also continuous. We have therefore a contradiction; y cannot be both discrete and continuous. See the Mathematics Appendix for the definition of continuity.

2.3 Applications

2.15 *All relationships in economics must be postulated as structural. Debate this claim.*

Answer: There are always some random components in economic relationships, so they must be postulated as structural. However, the analysis of certain simple technical relationships, for example, in production functions, might usefully be regarded as deterministic. It is unlikely that human decisions or reactions to economic events can ever be regarded as deterministic.

2.16 *The difference between economic laws and physical laws is that the latter are deterministic and the former are structural. Comment.*

Answer: Both are structural in fact; the main difference between the disciplines is that in physical relationships the degree of random variation is much less relative to the size of the variation of the deterministic component.

Chapter 3

How to Describe and Summarize Random Data by Graphical Procedures

3.1 Calculation Practice

3.1 Worked. Objective: Practice calculation of the quartiles. *Find the minimum, maximum, range, interquartile range, and median for the following sets of data.*

a. -7,0,5,6,-1,-2,4

b. 1,2,8,3,6,3

c. 0,1,100,0,90,50,2,80,100,0,10,20

d. -10, 8, 3, -3, 13, 4, -7, 6,5,6,1,9,-2

Answers:

a. Reorder data: -7,-2,-1,0,4,5,6; Min = -7, Max = 6, Range = 13, Q_1 = -2, Q_3 = 5, Inter quartile range = 7, Median = 0

b. Reorder data: 1,2,3,3,6,8; Min = 1, Max = 8, Range = 7, Q_1 = 2, Q_3=6, Interquartile range = 4, Median = 3

c. Reorder data: 0,0,0,1,2,10,20,50,80,90,100,100; Min = 0, Max = 100, Range = 100, Q_1 = 0.5, Q_3 = 85, Interquartile range = 84.5, Median = 15

d. Reorder data: -10,-7,-3,-2,1,3,4,5,6,6,8,9,13; Min = -10, Max = 13, Range = 23, Q_1 = -2.5, Q_3 = 7, Interquartile range = 9.5, Median = 4

3.5 *For practice in constructing histograms, use the following table, which shows a frequency distribution (based on seven "classes") of the weekly wages of 65 employees at BagelBurger. Show*

a. The lower limit of the sixth cell, or interval

b. The upper limit of the fourth cell, or interval

c. The class mark, or cell mark, of the third cell, or interval

d. The relative frequency of the third cell

e. The cell, or interval, having the largest frequency

f. The percentage of employees earning less than \$80 per week

g. The percentage of employees earning less than \$100, but at least \$60 per week

h. If a new employee were to be added to the staff of BagelBurger at the princely sum of \$134.35 per week, what would you do to alter the histogram to include this observation?

Wages(\$)	*No. of Employees*
50.00 – 59.99	8
60.00 – 69.99	10
70.00 – 79.99	16
80.00 – 89.99	14
90.00 – 99.99	10
100.00 – 109.99	5
110.00 – 119.99	2

Answer:

a. 100.00;

b. 89.99;

c. 75.00;

d. Total number of observations is 65. So the relative frequency of the third cell is 16/65, or 0.246;

e. The third cell;

f. (8+10+16)/65 = 0.523, or 52.3%;

g. (10+16+14+10)/65, or (65-8-2)/65 = 0.769, or 76.9%;

h. Two ways. One is to add two cells, "120.00-129.99" and "130.00-139.99." The other is to add just one cell, "120.00 and above."

3.7 *Repeat Exercise 3.6 using two other film studios. What general observations can you make about the choice of number of cells (bars)?*

Answer: The number of cells should be chosen carefully. A reasonable number of observations per cell is needed to have some stability in repeated sampling for the relative frequency, but more cells with narrower widths are needed to get a smoother histogram.

3.2 Exploring the Tools

3.9 *Consider film studio O discussed in this chapter. Suppose that you received the revenue data on this studio piece by piece, starting with the first 18 observations, which are summarized in the following table.*

Fill in the table one line at a time. Examine the extent to which each of the measures that enter a box-and-whisker plot is altered by the addition of new data. Draw conclusions from this experience about the size of the sample of data that you might need to observe and not have the shape measures continue to change substantially.

Observ.	*Min*	*1st Quart*	*Median*	*3rd Quart*	*Max*	*Int.Qt. Range*	*Range*
First 18	4.2	9.09	15.25	25.68	51.6	16.59	47.4
20.30							
2.56							
38.0							
89.4							
20.5							

Construct a box-and-whisker plot by hand for the first 18 observations only, and compare the results with the plot for all the data.

Answer:

Observ.	*Min*	*1st Quart*	*Median*	*3rd Quart*	*Max*	*Int.Qt. Range*	*Range*
First 18	4.2	9.09	15.25	25.68	51.6	16.59	47.4
20.30	4.2	9.56	16.2	25.55	51.6	15.99	47.4
2.56	2.56	8.62	15.25	25.42	51.6	16.80	49.4
38.0	2.56	8.62	16.2	25.8	51.6	17.18	49.4
89.4	2.56	9.09	16.65	26.93	89.4	17.84	86.84
20.5	2.56	9.56	17.1	26.55	89.4	16.99	86.84

One objective in this exercise is to learn how "adding data" refines your view of the data set. For example, if you compare the calculations and box-and-whisker plots for the original data to those for the twenty three observations, you will note many differences in the "shape of the distribution" of the data.

3.11 *If there are five cells (and therefore five relative frequencies), and if you know four of them, you must know the fifth. True? Explain.*

Answer: The sum of all relative frequencies is always one. If you know four relative frequencies out of five, the fifth is equal to one minus the sum of the other four.

3.13 *When discussing the construction of cells, the text states: "If we just pick four cells, we will not have improved matters over our box-and-whisker plot." Why not?*

Answer: When only four cells are picked, the information we get from the histogram is at most range and quartiles, which is just what we learn from box-and-whisker plots.

3.15 *Using the cumulative distributions of heights for men and women aged 18 to 24 contained in file Htmw.xls in folder Xfiles, subfolder Misc. (See Appendix B, Section D, "Data Files"), by hand:*

a. Construct a table showing the relative frequencies.

b. Construct a histogram from these data. What is your choice for the cell marks and the cell boundaries?

c. What differences do you observe between these two distributions?

d. What do the corresponding box-and-whisker plots look like? Comment on the advantages and disadvantages of the two presentations of the data.