Introductory Statistics 2nd edition Gould Solutions Manual

SKU: 9780321978271-SOLUTIONS

INSTRUCTOR'S SOLUTIONS MANUAL

JAMES LAPP

Colorado Mesa University

INTRODUCTORY STATISTICS:
EXPLORING THE WORLD THROUGH DATA
SECOND EDITION

Robert Gould

University of California Los Angeles

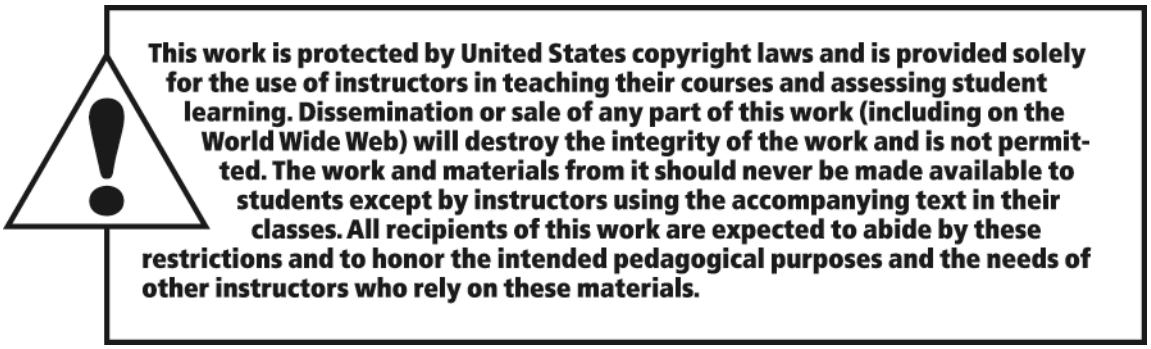
Colleen Ryan

California Lutheran University

PEARSON

Boston Columbus Hoboken Indianapolis New York San Francisco
Amsterdam Cape Town Dubai London Madrid Milan Munich Paris Montreal Toronto
Delhi Mexico City Sao Paulo Sydney Hong Kong Seoul Singapore Taipei Tokyo

Introductory Statistics 2nd edition Gould Solutions Manual



The author and publisher of this book have used their best efforts in preparing this book. These efforts include the development, research, and testing of the theories and programs to determine their effectiveness. The author and publisher make no warranty of any kind, expressed or implied, with regard to these programs or the documentation contained in this book. The author and publisher shall not be liable in any event for incidental or consequential damages in connection with, or arising out of, the furnishing, performance, or use of these programs.

Reproduced by Pearson from electronic files supplied by the author.

Copyright © 2016, 2013 Pearson Education, Inc.
Publishing as Pearson, 75 Arlington Street, Boston, MA 02116.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of the publisher. Printed in the United States of America.

ISBN-13: 978-0-321-97840-0
ISBN-10: 0-321-97840-4

www.pearsonhighered.com

PEARSON

Introductory Statistics 2nd edition Gould Solutions Manual

CONTENTS

Chapter 1: Introduction to Data

Section 1.2: Classifying and Storing Data 1
Section 1.3: Organizing Categorical Data 2
Section 1.4: Collecting Data to Understand Causality....5
Chapter Review Exercises 6

Chapter 2: Picturing Variation with Graphs

Section 2.1: Visualizing Variation in Numerical Data 9
Section 2.2: Summarizing Important Features of a Numerical Distribution 9
Section 2.3: Visualizing Variation in Categorical Variables 13
Section 2.4: Summarizing Categorical Distributions....13
Section 2.5: Interpreting Graphs 14
Chapter Review Exercises 15

Chapter 3: Numerical Summaries of Center and Variation

Section 3.1: Summaries for Symmetric Distributions 17
Section 3.2: What's Unusual? The Empirical Rule and *z*-Scores....22
Section 3.3: Summaries for Skewed Distributions 23
Section 3.4: Comparing Measures of Center 24
Section 3.5: Using Boxplots for Displaying Summaries 25
Chapter Review Exercises 27

Chapter 4: Regression Analysis: Exploring Associations between Variables

Section 4.1: Visualizing Variability with a Scatterplot 35
Section 4.2: Measuring Strength of Association with Correlation 35
Section 4.3: Modeling Linear Trends 37
Section 4.4: Evaluating the Linear Model 41
Chapter Review Exercises 44

Chapter 5: Modeling Variation with Probability

Section 5.1: What Is Randomness?....51
Section 5.2: Finding Theoretical Probabilities....51
Section 5.3: Associations in Categorical Variables 56
Section 5.4: Finding Empirical Probabilities 60
Chapter Review Exercises 61

Introductory Statistics 2nd edition Gould Solutions Manual

Chapter 6: Modeling Random Events: The Normal and Binomial Models

Section 6.1: Probability Distributions Are Models of Random Experiments.....71
Section 6.2: The Normal Model.....72
Section 6.3: The Binomial Model (Optional).....79
Chapter Review Exercises.....81

Chapter 7: Survey Sampling and Inference

Section 7.1: Learning about the World through Surveys.....85
Section 7.2: Measuring the Quality of a Survey.....86
Section 7.3: The Central Limit Theorem for Sample Proportions.....87
Section 7.4: Estimating the Population Proportion with Confidence Intervals.....90
Section 7.5: Comparing Two Population Proportions with Confidence.....94
Chapter Review Exercises.....98

Chapter 8: Hypothesis Testing for Population Proportions

Section 8.1: The Essential Ingredients of Hypothesis Testing.....103
Section 8.2: Hypothesis Testing in Four Steps.....104
Section 8.3: Hypothesis Tests in Detail.....108
Section 8.4: Comparing Proportions from Two Populations.....109
Chapter Review Exercises.....113

Chapter 9: Inferring Population Means

Section 9.1: Sample Means of Random Samples.....123
Section 9.2: The Central Limit Theorem for Sample Means.....124
Section 9.3: Answering Questions about the Mean of a Population.....124
Section 9.4: Hypothesis Testing for Means.....127
Section 9.5: Comparing Two Population Means.....131
Chapter Review Exercises.....139

Chapter 10: Associations between Categorical Variables

Section 10.1: The Basic Ingredients for Testing with Categorical Variables.....149
Section 10.2: The Chi-Square Test for Goodness of Fit.....151
Section 10.3: Chi-Square Tests for Associations between Categorical Variables.....154
Section 10.4: Hypothesis Tests When Sample Sizes Are Small.....161
Chapter Review Exercises.....165

Chapter 11: Multiple Comparisons and Analysis of Variance

Section 11.1: Multiple Comparisons.....173
Section 11.2: The Analysis of Variance.....177
Section 11.3: The ANOVA Test.....178
Section 11.4: Post-Hoc Procedures.....182
Chapter Review Exercises.....186

Introductory Statistics 2nd edition Gould Solutions Manual

Chapter 12: Experimental Design: Controlling Variation

Section 12.1: Variation Out of Control....189
Section 12.2: Controlling Variation in Surveys....194
Section 12.3: Reading Research Papers....195
Chapter Review Exercises....197

Chapter 13: Inference without Normality

Section 13.1: Transforming Data....199
Section 13.2: The Sign Test for Paired Data....201
Section 13.3: Mann-Whitney Test for Two Independent Groups....202
Section 13.4: Randomization Tests....204
Chapter Review Exercises....205

Chapter 14: Inference for Regression

Section 14.1: The Linear Regression Model....209
Section 14.2: Using the Linear Model....210
Section 14.3: Predicting Values and Estimating Means....211
Chapter Review Exercises....213

Introductory Statistics 2nd edition Gould Solutions Manual

Introductory Statistics 2nd edition Gould Solutions Manual

Chapter 1: Introduction to Data

Section 1.2: Classifying and Storing Data

1.1. There are nine variables: "Male", "Age", "Eye Color", "Shoe Size", "Height, Weight", "Number of Siblings", "College Units This Term", and "Handedness".

1.2. There are eleven observations.

1.3. a. Handedness is categorical.

b. Age is numerical.

1.4. a. Shoe size is numerical.

b. Eye color is categorical.

1.5. Answers will vary but could include such things as number of friends on Facebook or foot length. *Don't copy these answers*.

1.6. Answers will vary but could include such things as class standing ("Freshman", "Sophomore", "Junior", or "Senior") or favorite color. *Don't copy these answers*.

1.7. The label would be "Brown Eyes" and there would be eight 1's and three 0's.

1.8. There would be nine 1's and two 0's.

1.9. Male is categorical with two categories. The 1's represent males, and the 0's represent females. If you added the numbers, you would get the number of males, so it makes sense here.

1.10.

Units	Full
16.0	1
13.0	1
5.0	0
15.0	1
19.5	1
11.5	0
9.5	0
8.0	0
13.5	1
12.0	1
14.0	1

1.11 a. The data is stacked.

b. 1 means male and 0 means female.

c.

Female	Male
9.5	9.4
9.5	9.5
9.9	9.5
	9.7

1.12 a. The data is unstacked.

b. Labels for columns will vary.

Age	5 p.m.
31	1
34	1
46	1
47	1
50	1
24	0
18	0
21	0
20	0
20	0

Copyright © 2016 Pearson Education, Inc.

Introductory Statistics 2nd edition Gould Solutions Manual

1.13 a. Stacked and coded

Calories	Sweet
90	1
310	1
500	1
500	1
600	1
90	1
150	0
600	0
500	0
550	0

The second column could be labeled "Salty" with the 1's being 0's and the 0's being 1's.

b. Unstacked

Sweet	Salty
90	150
310	600
500	500
500	550
600	
90	

1.14 a. Stacked and coded

Cost	Male
10	1
15	1
15	1
25	1
12	1
8	0
30	0
15	0
15	0

The second column could labeled "Female" with the 1's being 0's and the 0's being 1's.

b. Unstacked

Male	Female
10	8
15	30
15	15
25	15
12	

Section 1.3: Organizing Categorical Data

1.15 a.

	Men	Women	Total
Yes, Older S	12	55	$12+55=67$
No, Older S	11	39	50
Total	23	$55+39=94$	117

b. $12/23=52.2\%$
c. $11/23=47.8\%$
d. $55/94=58.5\%$
e. $67/117=57.3\%$
f. $55/67=82.1\%$
g. $0.585(600)=351$

1.16 a.

	Men	Women	Total
Work	15	65	$15+65=80$
Not Work	23	28	51
Total	38	$65+28=93$	131

b. $15/38=39.5\%$
c. $23/38=60.5\%$
d. $65/93=69.9\%$
e. $80/131=61.1\%$
f. $65/80=81.25\%$
g. $15/80=18.75\%$
h. $65/93\times 800=559$

Copyright © 2016 Pearson Education, Inc.

Introductory Statistics 2nd edition Gould Solutions Manual

1.17 a. $15/38$, or 39.5%, of the class were male.

b. $0.641(234) = 149.99$, or about 150, men in the class

c. $0.40(x) = 20$

$$x = \frac{20}{0.40}$$

$= 50$ people in the class

1.18 a. $0.35(346) = 121$ male nurses

b. $66/178 = 37.1\%$ female engineers

c. $0.65(x) = 169$

$$x = \frac{169}{0.65}$$

$= 260$ lawyers

1.19 The frequency of women is 7, the proportion is $7/11$, and the percentage is 63.6%.

1.20 The frequency of righties is 9, the proportion is $9/11$, and the percentage is 81.8%.

1.21 The answers follow the guidance on page 34.

a. and b.

	Men	**Women**	**Total**
Right	4	5	9
Left	0	2	2
Total	4	7	11

c. $5/7 = 71.4\%$

d. $5/9 = 55.6\%$

e. $9/11 = 81.8\%$

f. $0.714(70) = 50$

1.22 a. and b.

	Men	**Women**	**Total**
Brown	3	5	8
Blue	1	1	2
Hazel	0	1	1
Total	4	7	11

c. $5/7 = 71.4\%$

d. $5/8 = 62.5\%$

e. $8/11 = 72.7\%$

f. $0.714(60) = 42.84$ or about 43

1.23 $0.202x = 88,547,000$

$$x = \frac{88,547,000}{0.202}$$

$x = 438,351,485$ (final value could be rounded differently)

1.24 $0.055x = 12,608,000$

$$x = \frac{12,608,000}{0.055}$$

$x = 229,236,364$ (final value could be rounded differently)

Copyright © 2016 Pearson Education, Inc.

Introductory Statistics 2nd edition Gould Solutions Manual

1.25 The answers follow the guidance on page 34.

1–3:

State	AIDS/HIV	Rank Cases	Population	Population (thousands)	AIDS/HIV per 1000	Rank Rate
New York	192,753	1	19,421,005	19,421	$\frac{192,753}{19,421} = 9.92$	2
California	160,293	2	37,341,989	37,342	$\frac{160,293}{37,342} = 4.29$	5
Florida	117,612	3	18,900,773	18,901	$\frac{117,612}{18,901} = 6.22$	3
Texas	77,070	4	25,258,418	25,258	$\frac{77,070}{25,258} = 3.05$	6
New Jersey	54,557	5	8,807,501	8,808	$\frac{54,557}{8808} = 6.19$	4
District of Columbia	9257	6	601,723	602	$\frac{9257}{602} = 15.38$	1

4: No, the ranks are not the same. The District of Columbia had the highest rate and had the lowest number of cases. (Also, the rate for Florida puts its rank above California, and the rate for New Jersey puts it above Texas in ranking.)

5: The District of Columbia is the place (among these six regions) where you would be most likely to meet a person diagnosed with AIDS/HIV, and Texas is the place (among these six regions) where you would be least likely to do so.

1.26 a.

State	**Population Density**	**Rank**
Pennsylvania	$\frac{12,448,279}{44,817} = 277.76$	3
Illinois	$\frac{12,901,563}{55,584} = 232.11$	5
Florida	$\frac{18,328,340}{53,927} = 339.87$	2
New York	$\frac{19,490,297}{47,214} = 412.81$	1
Texas	$\frac{24,326,974}{261,797} = 92.92$	6
California	$\frac{36,756,666}{155,959} = 235.68$	4

b. Texas has the lowest population density.

c. New York has the highest population density.

Copyright © 2016 Pearson Education, Inc.

Introductory Statistics 2nd edition Gould Solutions Manual

1.27

Year	Percentage
1990	$\frac{112.6}{191.8} = 58.7\%$
1997	$\frac{116.8}{207.2} = 56.4\%$
2000	$\frac{120.2}{213.8} = 56.2\%$
2007	$\frac{129.9}{235.8} = 55.1\%$

The percentage of married people is decreasing over time (at least with these dates).

1.28

Year	Percentage
2006	$\frac{2426}{4266} = 56.9\%$
2007	$\frac{2424}{4316} = 56.2\%$
2008	$\frac{2473}{4248} = 58.2\%$
2009	$\frac{2437}{4131} = 59.0\%$
2010	$\frac{2452}{4007} = 61.2\%$

The rate of death as a percentage of the rate of birth tends to go up over this time period This is primarily due to the birth rate decreasing.

1.29 We don't know the percentage of female students in the two classes. The larger number of women at 8 a.m. may just result from a larger number of students at 8 a.m., which may be because the class can accommodate more students because perhaps it is in a large lecture hall.

1.30 We don't know the rate of fatalities—that is, the number of fatalities per pedestrian. There may be fewer pedestrians in Hillsborough County, and that may be the source of the difference.

Section 1.4: Collecting Data to Understand Causality

1.31 Observational study

1.32 Observational study

1.33 Controlled experiment

1.34 Controlled experiment

1.35 Controlled experiment

1.36 Observational study

1.37 Observational study

1.38 Controlled experiment

1.39 This was an observational study, and from it you cannot conclude that the tutoring raises the grades. Possible confounders (answers may vary): 1. It may be the more highly motivated who attend the tutoring, and this motivation is what causes the grades to go up. 2. It could be that those with more time attend the tutoring, and it is the increased time studying that causes the grades to go up.

1.40 a. If the doctor decides on the treatment, you could have bias.
b. To remove this bias, randomly assign the patients to the different treatments.
c. If the doctor knows which treatment a patient had, that might influence his opinion about the effectiveness of the treatment.
d. To remove that bias, make the experiment double-blind. Neither the patients nor the doctor evaluating the patients should know whether each patient received medication or talk therapy.

1.41 a. It was a controlled experiment, as you can tell by the random assignment. This tells us that the researchers determined who received which treatment.
b. We can conclude that the early surgery caused the better outcomes, because it was a randomized controlled experiment.

Copyright © 2016 Pearson Education, Inc.

Introductory Statistics 2nd edition Gould Solutions Manual

1.42 This is an observational study, because researchers did not determine who received PCV7 and who did not. You cannot conclude causation from an observational study. We must assume that it is possible that there were confounding variables (such as other advances in medicine) that had a good effect on the rate of pneumonia.

1.43 Answers will vary. However, they should all mention randomly dividing the 100 people into two groups and giving one group the copper bracelets. The other group could be given (as a placebo) bracelets that look like copper but are made of some other material. Then the pain levels after treatment could be compared.

1.44 a. Heavier people might be more likely to choose to eat meat. Also, people who are not prepared to change their diet very much (such as by excluding meat) might also not change other variables that affect weight, such as how much exercise they get.

b. It would be better to randomly assign some of the subjects to eat meat and some of the subjects to consume a vegetarian diet.

1.45 No. This was an observational study, because researchers could not have deliberately exposed people to weed killers. There was no random assignment, and no one would randomly assign a person to be exposed to pesticides. From an observational study, you cannot conclude causation. This is why the report was careful to use the phrase associated with rather than the word caused.

1.46 a. The survival rate for TAC $(473/539$, or $87.8\%)$ was higher than the survival rate for FAC $(426/521$, or $81.8\%)$.

b. Controlled experiment: Yes, we can conclude cause and effect, because this was a controlled experiment with random assignment. The random assignment balances out other variables, so the only difference is the treatment, which must be causing the effect.

1.47 Ask whether the patients were randomly assigned the full or the half dose. Without randomization there could be bias, and we cannot infer causation. With randomization we can infer causation.

1.48 Ask whether there was random assignment to groups. Without random assignment there could be bias, and we cannot infer causation.

1.49 This was an observational study: vitamin C and breast milk. We cannot conclude cause and effect from observational studies.

1.50 This is likely to be from observational studies. It would not be ethical to assign people to overeat. We cannot conclude causation from observational studies because of the possibility of confounding variables.

1.51 a. LD: $\frac{4}{4+46}=\frac{2}{25}=8\%$ tumors; LL: $\frac{14}{14+36}=\frac{7}{25}=28\%$ tumors

b. A controlled experiment. You can tell by the random assignment.

c. Yes, we can conclude cause and effect because it was a controlled experiment, and random assignment will balance out potential confounding variables.

1.52 a. $\frac{43}{43+10}=\frac{43}{53}$, or 81.1%, of the males who were assigned to Scared Straight were rearrested. $\frac{37}{37+18}=\frac{37}{55}$, or 67.3%, of those receiving no treatment were rearrested. So the group from Scared Straight had a higher arrest rate.

b. No, Scared Straight does not cause a lower arrest rate, because the arrest rate was higher.

Chapter Review Exercises

1.53 a. Dating: 81/440, or 18.4%

b. Cohabiting: 103/429, or 24.0%

c. Married: 147/424, or 34.7%

d. No, this was an observational study. Confounding variables may vary. Perhaps married people are likely to be older, and older people are more likely to be obese.

1.54 No, this was an observational study. There is no mention of random assignment. We cannot conclude causation from observational studies because of the possibility of confounding factors.

Copyright © 2016 Pearson Education, Inc.

Introductory Statistics 2nd edition Gould Solutions Manual

1.55 a

	Boy	Girl	Total
Violent	10	11	21
Nonviolent	19	4	23
Total	29	15	44

b. For the boys, 10/29, or 34.5%, were on probation for violent crime. For the girls, 11/15, or 73.3%, were on probation for violent crime.

c. The girls were more likely to be on probation for violent crime.

1.56 For those getting the antivenom, 87.5% got better. For those given the placebo, only 14.3% got better.

	Antivenom	Placebo	Total
Better	7	1	8
Not Better	1	6	7
Total	8	7	15

1.57 Answers will vary. Students should not copy the words they see in these answers. Randomly divide the group in half, using a coin flip for each woman: Heads she gets the vitamin D, and tails she gets the placebo (or vice versa). Make sure that neither the women themselves nor any of the people who come in contact with them know whether they got the treatment or the placebo ("double-blind"). Over a given length of time (such as three years), note which women had broken bones and which did not. Compare the percentage of women with broken bones in the vitamin D group with the percentage of women with broken bones in the placebo group.

1.58 Answers will vary. Students should not copy the words they see here. Randomly divide the group in half, using a coin flip for each person: Heads they get Coumadin, and tails they get aspirin (or vice versa). Make sure that neither the subjects nor any of the people who come in contact with them know which treatment they received ("double-blind"). Over a given length of time (such as three years), note which people had second strokes and which did not. Compare the percentage of people with second strokes in the Coumadin group with the percentage of people with second strokes in the aspirin group. There is no need for a placebo, because we are comparing two treatments. However, it would be acceptable to have three groups, one of which received a placebo.

1.59 a. The treatment variable was Medicaid expansion or not and the response variables were the death rate and the rate of people who reported their health as excellent or very good.

b. This was observational. Researchers did not assign people either to receive or not to receive Medicaid.

c. No, this was an observational study. From an observational study, you cannot conclude causation. It is possible that other variables that differed between the states caused the change.

1.60 a. The treatment variable is whether the person has both forms of HIV infection (HIV-1 and HIV-2) or only one form (HIV-1). The response variable is the time to the development of AIDS.

b. This was an observational study. No one would assign a person to a form of HIV.

c. The median time to development of AIDS was longer for those with both infections.

d. No, you cannot infer causation from an observational study.

1.61 No, we cannot conclude causation. There was no control group for comparison, and the sample size was very small.

1.62 No, it does not show that the exercise works. There is no control group. (Also, the sample size is very small.)

Copyright © 2016 Pearson Education, Inc.

Introductory Statistics 2nd edition Gould Solutions Manual

Introductory Statistics 2nd edition Gould Solutions Manual

Chapter 2: Picturing Variation with Graphs

Section 2.1: Visualizing Variation in Numerical Data and Section 2.2: Summarizing Important Features of a Numerical Distribution

2.1 a. 11 are morbidly obese.

b. $\frac{11}{134} = 0.082$, or about 8%, which is much more than 3%.

2.2 a. 21 have levels above 240.

b. $\frac{21}{93} = 0.226$, or about 23%. That is a bit more than the 18% mentioned.

2.3 New vertical axis labels: $\frac{1}{25} = 0.04$, $\frac{2}{25} = 0.08$, $\frac{3}{25} = 0.12$, $\frac{4}{25} = 0.16$, $\frac{5}{25} = 0.20$, $\frac{6}{25} = 0.24$, $\frac{7}{25} = 0.28$.

2.4 a. $0.04 + 0.13 = 0.17$ and $0.17(24) = 4.08$, or about 4

b. The two modes are 7 and 8.

2.5 a. 1 (or 2) have no TVs

b. 9 TVs

c. Between 25 and 30

d. Around 6

e. $\frac{6}{90} = \frac{1}{15}$, or 0.0667

2.6 a. 18 or 19 hours

b. 8

c. About 5 or 6

d. $\frac{5}{50} = \frac{1}{10}$ or $\frac{6}{50} = \frac{3}{25}$ (or about 0.10 or 0.12)

2.7 a. Both dotplots are right-skewed. The dotplot for the females is also multimodal.

b. The females tend to have more pairs of shoes.

c. The numbers of pairs for the females are more spread out. The males' responses tend to be clustered at about 10 pairs or fewer.

2.8 a. Detroit

b. Seattle

c. Left-skewed

2.9 There will be a lot of people who have no tickets and maybe a few with 1, 2, 3, or more, so the distribution will be right-skewed.

2.10 This should be left-skewed with a lot of people reporting 7 and a few reporting various values less than 7.

2.11 It would be bimodal because men and women tend to have different heights, with men being taller overall, and therefore longer armspans.

2.12 It might be bimodal because private colleges and public colleges tend to differ in amount of tuition.

2.13 About 58 years (between 56 and 60)

2.14 The typical number of sleep hours is around 7 or 7.5 hours.

2.15 Riding the bus shows a larger typical value and also more variation.

2.16 a. Both graphs are bimodal with modes at about 100 and 200 dollars per month.

b. The women tend to spend a bit more.

c. The data for the women have more variation.

2.17 a. The distribution is multimodal with modes at 12 years (high school), 14 years (junior college), 16 years (bachelor's degree), and 18 years (possible master's degree). It is also left-skewed with numbers as low as 0.

b. Estimate: 300 + 50 + 100 + 40 + 50, or about 500 to 600, had 16 or more years.

Copyright © 2016 Pearson Education, Inc.

Introductory Statistics 2nd edition Gould Solutions Manual

2.17 (continued)

c. Between $\frac{500}{2018}$, or about 25%, and $\frac{600}{2018}$, or about 30%, have a bachelor's degree or higher. This is very similar to the 27% given.

2.18 a. The distribution is right-skewed.

b. About 2 or 3

c. Between 80 and 100

d. $\frac{80}{2000} = 4\%$ or $\frac{100}{2000} = 5\%$

2.19 Both graphs go from about 0 to about 20 years of education, but the data for years of formal education for the respondents (compared to their mothers) include more with education above 12 years. For example, the bar at 16 (college bachelor's degree) is higher for the respondents than for the mothers, which shows that the respondents tend to have a bit more education than their mothers. Also, the bar at 12 is taller for the mothers, showing that the mothers were more likely to get only a high school diploma. Furthermore, the bar graph for the mothers includes more people (taller bars) at lower numbers of years, such as 0 and 3 and 6.

2.20 For men the data go from about 0 to about 90, and for women the data go from about 0 to about 80. There are more men who worked more than 40 hours. For example, the bars at 45, 50, 55, and 60 are taller for the men, showing that more men than women worked those numbers of hours.

2.21 1. Most psychology students would be younger, with a few older students: This is histogram C.
2. The number of psychology students should roughly the same for each year: This is histogram B.
3. Most students would eat breakfast every day: This is histogram A.

2.22 1. Most students would do well on an easy test: This is histogram A.
2. The number of hours of television watched would be left-skewed, with fewer people watching many hours of television: This is histogram B.
3. The heights of adults would be unimodal and roughly symmetrical: This is histogram C.

2.23 1. The heights of students would be bimodal and roughly symmetrical: This is histogram B.
2. The number of hours of sleep would be unimodal and roughly symmetrical, with any outliers more likely being fewer hours of sleep: This is histogram A.
3. The number of accidents would be left skewed, with most student being involved in no or a few accidents: This is histogram C.

2.24 1. The SAT scores would be unimodal and roughly symmetrical: This is histogram C.
2. The weights of men and women would be bimodal and roughly symmetrical, but with more variation that SAT scores: This is histogram A.
3. The ages of students would be left skewed, with most student being younger: This is histogram B.

2.25 The answers follow the guidance on page 76.

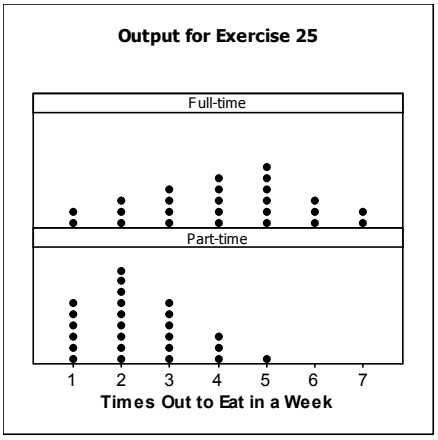
1: See the dotplots. Histograms would also be good for visualizing the distributions. Stemplots would not work with these data sets because all the observed values have only one digit.

2: Full-time is a bit left-skewed, and part-time is a bit right-skewed.

3: Those with full-time jobs tend to go out to eat more than those with part-time jobs.

4: The full-time workers have a distribution that is more spread out; full-time goes from 1 to 7, whereas part-time goes only from 1 to 5.

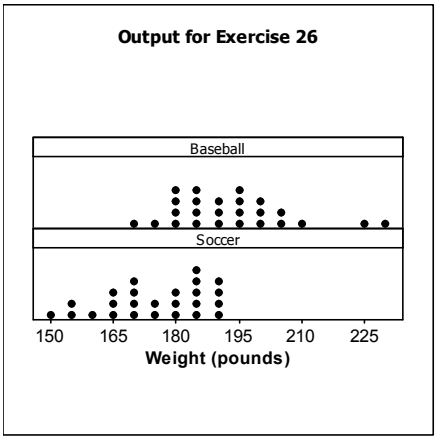
5: There are no outliers—that is, no dots detached from the main group with an empty space between.

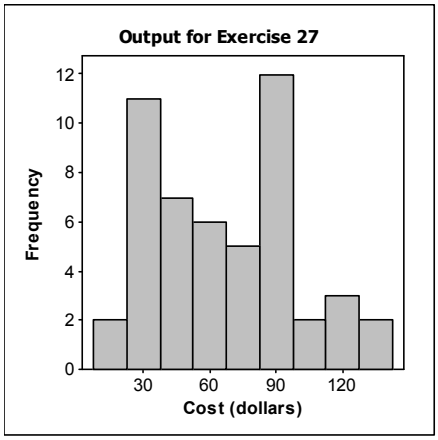


Copyright © 2016 Pearson Education, Inc.

Introductory Statistics 2nd edition Gould Solutions Manual

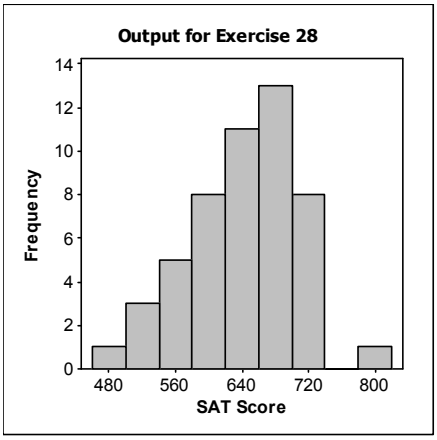
2.26 The figure shows dotplots of both groups. Histograms or stemplots would also be appropriate. The baseball players' weights are right-skewed with outliers at about 225 pounds or more. The soccer players' distribution of weights is left-skewed with no outliers. The baseball players tend to weigh more, and that data set is also more spread out. Both graphs appear bimodal with this grouping.

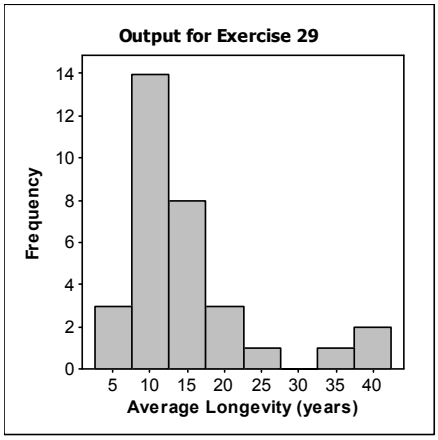




2.27 See histogram. The shape will depend on the binning used. The histogram is bimodal with modes at about \$30 and about \$90.

2.28 See histogram. The shape will depend on the binning used. The 800 score could be an outlier or not, and the graph could appear left-skewed or not.



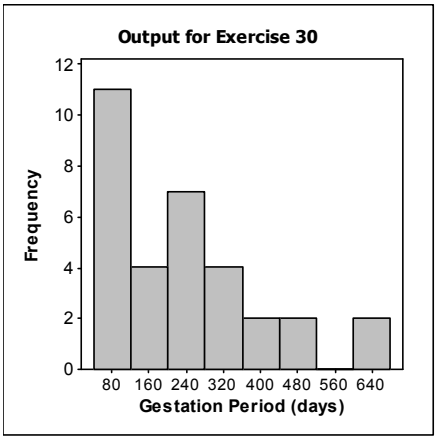


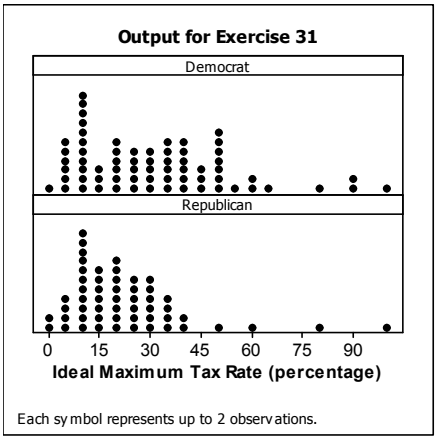
2.29 See histogram. The histogram is right-skewed. The typical value is around 12 (between 10 and 15) years, and there are three outliers: Asian elephant (40 years), African elephant (35 years), and hippo (41 years). Humans (75 years) would be way off to the right; they live much longer than other mammals

Copyright © 2016 Pearson Education, Inc.

Introductory Statistics 2nd edition Gould Solutions Manual

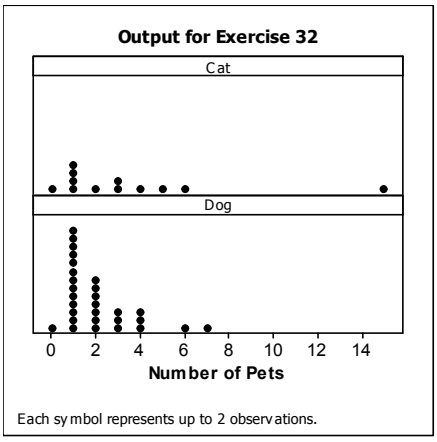
2.30 The histogram is right-skewed and also bimodal (at least with this grouping). The modes are at about 80 days and 240 days. The typical value is about 240 days (between 160 and 320 days). There are two outliers at more than 600 days, the Asian elephant and the African elephant. Humans (266 days) would be near the middle of the graph.

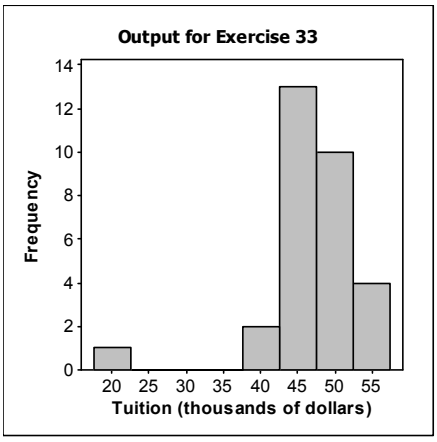




2.31 Both graphs are multimodal and right-skewed. The Democrats have a higher typical value, as shown by the fact that the center is roughly around 35 or 40%, while the center value for the Republicans is closer to 20 to 30%. Also note the much larger proportion of Democrats who think the rate should be 50% or higher. The distribution for the Democrats appears more spread out because the Democrats have more people responding with both lower and higher percentages.

2.32 Both distributions are right-skewed. A large outlier did represent a cat lover, but typically, cat lovers and dog lovers both seem to have about 2 pets, although there are a whole lot of dog lovers with one dog.



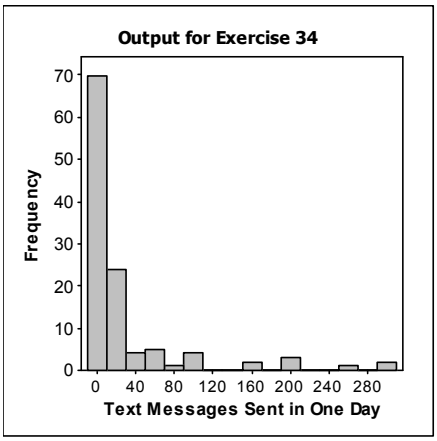


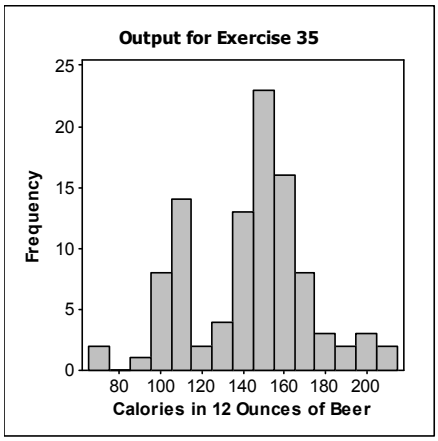
2.33 The distribution appears left-skewed because of the low-end outlier at about $20,000 (Brigham Young University).

Copyright © 2016 Pearson Education, Inc.

Introductory Statistics 2nd edition Gould Solutions Manual

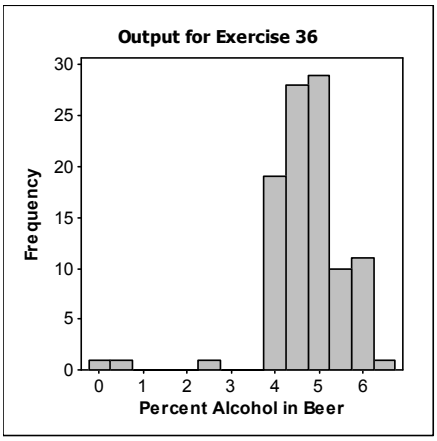
2.34 The histogram is strongly right-skewed, with outliers.





2.35 With this grouping the distribution appears bimodal with modes at about 110 and 150 calories. (With fewer—that is, wider—bins, it may not appear bimodal.) There is a low-end outlier at about 70 calories. There is a bit of left skew.

2.36 The distribution is left-skewed primarily because of the outliers at about 0% alcohol.



Section 2.3: Visualizing Variation in Categorical Variables and Section 2.4: Summarizing Categorical Distributions

2.37 No, the largest category is Wrong to Right, which suggests that changes tend to make the answers more likely to be right.

2.38 a. About 7.5 million
b. About 5 million
c. No, overweight and obesity do not result in the highest rate. That is from high blood pressure.
d. This is a Pareto chart.

2.39 a. 80 to 82%
b. Truth, since almost all observations are in the Top Fifth category.
c. Ideal, since these are almost uniformly spread across the five groups.
d. They underestimate the proportion of wealth held by the top 20%.

2.40 a. Oxnard tends to have more highly educated residents. Note that the bars for Oxnard are taller than the bars for Nyeland Acres for all the categories that show at least one year of college. Also note that the bars for Nyeland Acres are taller for the category with the least education.
b. Nyeland Acres has the least variation, because a substantially greater percentage of residents are in a single category (Less than HS). Oxnard also has residents in more categories, which suggests that it is more variable.

Copyright © 2016 Pearson Education, Inc.

Introductory Statistics 2nd edition Gould Solutions Manual

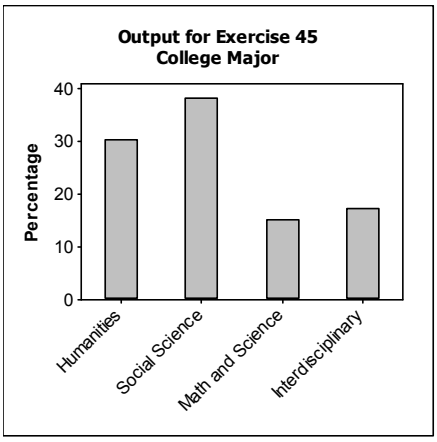
2.41 a. Dem (not strong)
b. Other. It is easier to pick out the second tallest bar in the bar chart. (Answers may vary.)

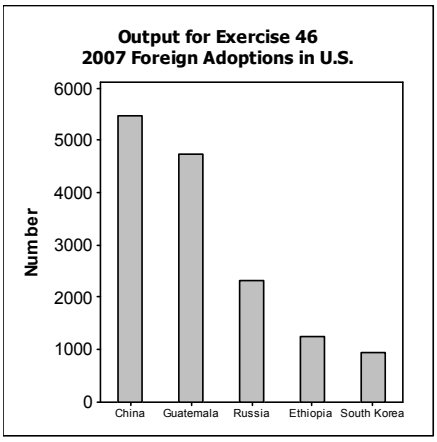
2.42 a. Dem (not strong)
b. Other. It is easier to pick out the second tallest bar in the bar chart.
c. There is no evidence of that. The percentage of men who are Democrats may even be larger than the percentage for women.

2.43 a. The percentage of old people is increasing, the percentage of those 25–64 is decreasing, and the percentage of those 24 and below is relatively constant.
b. The money for Social Security normally comes from those in a working age range (which includes those 25–64), and that group is decreasing in percentage. Also, the group receiving Social Security (those 65 and older) is becoming larger. This suggests that in the future, Social Security might not get enough money from the workers to support the old people.

2.44 a. Midsize
b. The percentage for small cars is going up, at least from 2000 to 2007.
c. The percentage for large cars went down between 1985 and 2000 but went part of the way back up in 2007.

2.45 A Pareto chart or pie chart would also be appropriate. Note that the mode is Social Science and that there is substantial variation. (Of course, individual majors such as chemistry were grouped into Math and Science.)





2.46 This is a Pareto chart, but a bar chart or pie chart would also be appropriate. The mode is China, but there is substantial variation.

Section 2.5: Interpreting Graphs

2.47 This is a histogram, which we can see because the bars touch. The software treated the values of the variable *Garage* as numbers. However, we wish them to be seen as categories. A bar graph or pie chart would be better for displaying the distribution.

2.48 The graph is a histogram (the bars touch), and histograms are used for numerical data. But this data set is categorical, and the numbers (1, 2, and 3) represent categories. A more appropriate graph would be a bar graph or pie graph.

2.49 Hours of sleep is a numerical variable. A histogram or dotplot would better enable us to see the distribution of values. Because there are so many possible numerical values, this pie chart has so many "slices" that it is difficult to tell which is which.

2.50 a. This is a bar chart (or bar graph), as you can see by the separation between bars.
b. These numerical data would be better shown as a pair of histograms (with a common horizontal axis) or a pair of dotplots. Bar graphs are for categorical data.

2.51 Those who still play tended to have practiced more as teenagers, which we can see because the center of the distribution for those who still play is about 2 or 2.5 hours, compared to only about 1 or 1.5 hours for those who do not. The distribution could be displayed as a pair of histograms or a pair of dotplots.

Copyright © 2016 Pearson Education, Inc.

Introductory Statistics 2nd edition Gould Solutions Manual

2.52 a. *Gender* is categorical and *Hours on Cell Phone* is also categorical.
b. Because in this data set both variables are categorical, the bar chart is appropriate.
c. You could make two histograms (or two dotplots) for the data because the time would be numerical. It would be ideal to use a common horizontal axis for easy comparison of the two graphs.
d. The distributions show that the women tend to talk more. (The mode for women is 4–8 hours, and the mode for the men is 0–4 hours.)

Chapter Review Exercises

2.53 TV: Histograms: One for the males and one for the females would be appropriate. Dotplots or stemplots would also work for this numerical data set.

2.54 Jobs: Bar graphs would allow comparison of men and women in one graph. If you chose pie charts, you would need two.

2.55 a. The diseases with higher rates for HRT were heart disease, stroke, pulmonary embolism, and breast cancer. The diseases with lower rates for HRT were endometrial cancer, colorectal cancer, and hip fracture.
b. Comparing the rates makes more sense than comparing just the numbers, in case there were more women in one group than in the other.

2.56 a. South Korea and the United States have the highest rate of access to the Internet.
b. China and Thailand have the highest percentage of music purchased over the Internet.

2.57 The vertical axis does not start at zero and exaggerates the differences. Make a graph for which the vertical axis starts at zero.

2.58 In histograms the bars should generally touch, and these don't touch. Also, we cannot see the top of the range because "More" is a poor label. Change the numbers on the horizontal axis and increase the width of the bins so as to make the bars touch.

2.59 The shapes are roughly bell-shaped and symmetric; the later period is warmer, but the spread is similar. This is consistent with theories on global warming. The difference is 57.9 – 56.7 = 1.2, so the difference is only a bit more than 1 degree Fahrenheit.

2.60 The typical percentage of students with jobs at the top schools is higher than the percentage for the bottom 91 schools. In other words, you are more likely to find a job if you went to a law school in the top half of the rankings. Both histograms are left-skewed. Also, the range for the bottom schools is wider, because it goes down to lower employment rates.

2.61 a. The graph shows that a greater percentage of people survived when lying prone (on their stomachs) than when lying supine (on their backs). This suggests that we should recommend that doctors ask these patients to lie prone.
b. Both variables (*Position* and *Outcome*) are categorical, so a bar chart is appropriate.

2.62 In 2012, more people thought global warming was happening than thought so in 2010.

2.63 The created 10-point dotplot will vary, but the dotplot for this exercise should be right-skewed.

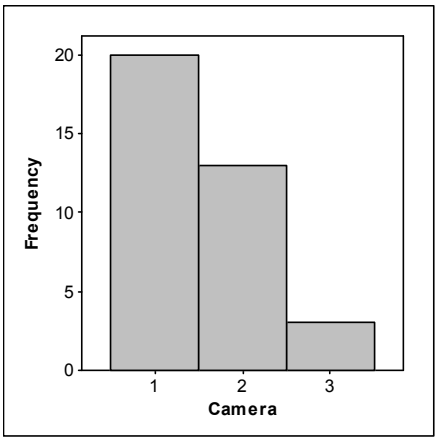
2.64 The created 10-point dotplot will vary, but the dotplot for this exercise should be not be skewed.

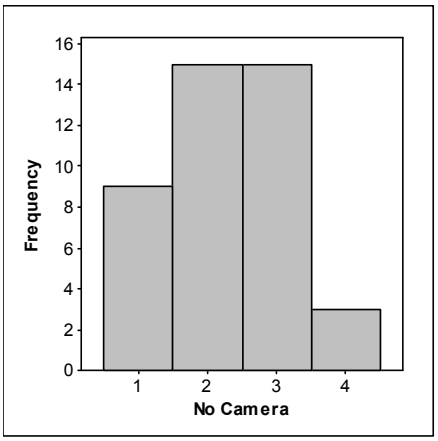
2.65 Graphs will vary. Histograms and dotplots are both appropriate. For the group without a camera the distribution is roughly symmetrical, and for the group with a camera it is right-skewed. Both are unimodal. The number of cars going through a yellow light tends to be less at intersections with cameras. Also, there is more variation in the intersections without cameras.

Copyright © 2016 Pearson Education, Inc.

Introductory Statistics 2nd edition Gould Solutions Manual

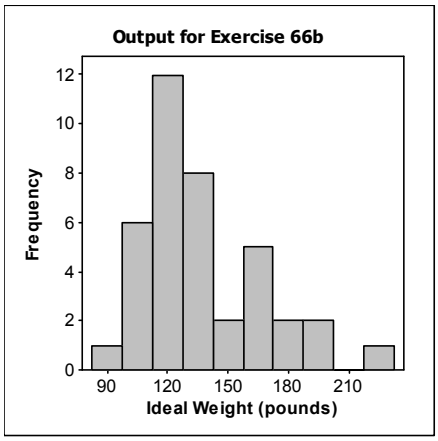
2.65 (continued)

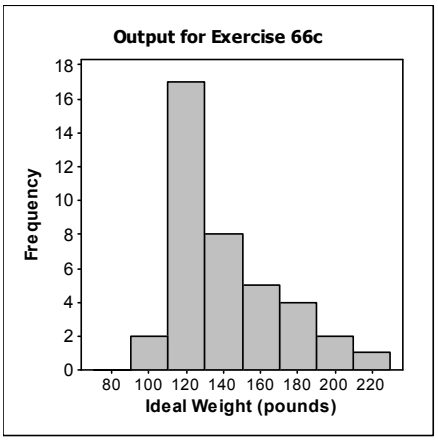




2.66 a. You might expect bimodality because men tend to have ideal weights that are larger than women's ideal weights.

b. and c.





Graphs may vary, depending on technology and the choice of bins for the second histogram. On the two graphs given here, the bin width for the first is 15 pounds and for the second is 20 pounds. The first distribution is bimodal and the second is not.

2.67 Both distributions are right-skewed. The typical speed for the men (a little above 100 mph) is a bit higher than the typical speed for the women (which appears to be closer to 90 mph). The spread for the men is larger primarily because of the outlier of 200 mph for the men.

2.68 Both graphs are relatively symmetric and unimodal. The center for the men is larger than the center for the women, showing that men tend to wear larger shoes than women. The spread is a bit more for women because their sizes range from about 5 to about 10 whereas the men's sizes range from about 8 to about 12. There are no outliers in either group.

2.69 The distribution should be right-skewed.

2.70 Since most of the physician's patients probably do not smoke and a few may be heavy smokers, the distribution should be right-skewed with lots of zeros and a few high numbers.

2.71 a. The tallest bar is Wrong to Right, which suggests that the instruction was correct.

b. For both instructors, the largest group is Wrong to Right, so it appears that changes made tend to raise the grades of the students.

2.72 a. The raw numbers would be affected by how many were in each group, and that might hide the rate. For example, because there are many more old women than old men, that information would hide the rates.

b. The males up to about 64 have a higher rate of visits to the ER. From 65 to 74 the rates are about the same, and for 75 and up the rates are higher for the women.

Copyright © 2016 Pearson Education, Inc.

Introductory Statistics 2nd edition Gould Solutions Manual

Chapter 3: Numerical Summaries of Center and Variation

Answers may vary slightly, especially for quartiles and interquartile ranges, due to type of technology used, or rounding.

Section 3.1: Summaries for Symmetric Distributions

3.1 c

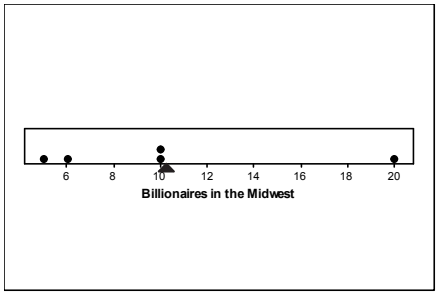
3.2 b

3.3 The typical age of the CEOs is between about 56 and 60 (or any number from 56 to 60). The distribution is symmetric, so the mean should be about in the middle.

3.4 The mean number of televisions is about 3 or 4. It is near the center because the distribution is roughly symmetric.

3.5 a. The mean number of billionaires in the five states is $\bar{x} = \frac{20+10+10+6+5}{5} = \frac{51}{5} = 10.2$.

b.



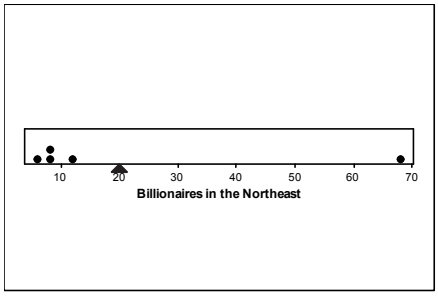
c. $s = \sqrt{\frac{140.8}{4}} = 5.9$

x	$x - \bar{x}$	$(x - \bar{x})^2$
20	9.8	96.04
10	–0.2	0.04
10	–0.2	0.04
6	–4.2	17.64
5	–5.2	27.04
51	0.0	140.80

d. The number farthest from the mean is 20, which is the largest number of billionaires.

3.6 a. The mean number of billionaires in the five states is $\bar{x} = \frac{67+11+7+7+5}{5} = \frac{97}{5} = 19.4$.

b.



Copyright © 2016 Pearson Education, Inc.

Introductory Statistics 2nd edition Gould Solutions Manual