

1. In 1997, there were 12,298,000 undergraduate students in U.S. colleges. According to the U.S. Department of Education, there were 127,000 American Indian or Alaskan Native students, 737,000 Asian or Pacific Islander, 1,380,000 non-Hispanic black, 1,108,000 Hispanic, and 8,682,000 non-Hispanic white students. In addition, 265,000 foreign undergraduates were enrolled in U.S. colleges.
- (a) Each number, including the total, is rounded to the nearest thousand. Separate rounding may cause roundoff errors, so that the sum of the counts does not equal the total given. Are roundoff errors present in these data? Explain.

(b) Present the data in a graph.

2. How much oil wells in a given field will ultimately produce is key information in deciding whether to drill more wells. Here are the estimated total amounts of oil recovered from 64 wells in the Devonian Richmond Dolomite area of the Michigan basin, in thousands of barrels.

21.7	53.2	46.4	42.7	50.4	97.7	103.1	51.9	43.4	69.5	156.5	34.6	37.9
12.9	2.5	31.4	79.5	26.9	18.5	14.7	32.9	196	24.9	118.2	82.2	35.1
47.6	54.2	63.1	69.8	57.4	65.6	56.4	49.4	44.9	34.6	92.2	37.0	58.8
21.3	36.6	64.9	14.8	17.6	29.1	61.4	38.6	32.5	12.0	28.3	204.9	44.5
10.3	37.7	33.7	81.1	12.1	20.1	30.5	7.1	10.1	18.0	3.0	2.0	

Construct an appropriate graph of the distribution and describe its shape, center, and spread.

1. Create a set of 5 positive numbers (repeats allowed) that have median 10 and mean 7. What thought process did you use to create your numbers?

2. Of the 50 species of oaks in the United States, 28 grow on the Atlantic Coast and 11 grow in California. We are interested in the distribution of acorn sizes among oak species. Here are data on the volumes of acorns (in cubic centimeters) for these 39 oak species:

Atlantic							California		
1.4	3.4	9.1	1.6	10.5	2.5	0.9	4.1	5.9	17.1
6.8	1.8	0.3	0.9	0.8	2.0	1.1	1.6	2.6	0.4
0.6	1.8	4.8	1.1	3.0	1.1	1.1	2.0	6.0	7.1
3.6	8.1	3.6	1.8	0.4	1.1	1.2	5.5	1.0	

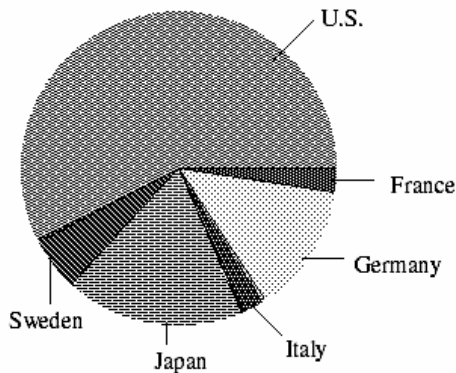
- (a) Make a stemplot of the 39 acorn sizes. Describe the shape, center and spread of the distribution. Choose appropriate numerical summaries.

- (b) Compare the Atlantic and California regional distributions with a graph and numerical summaries. What do you find?

**Directions:** *Work on these sheets. Answer completely, but be concise.*

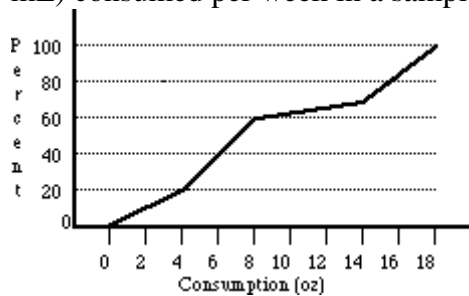
**Part 1: Multiple Choice.** *Circle the letter corresponding to the best answer.*

1. You measure the age, marital status and earned income of an SRS of 1463 women. The number and type of variables you have measured is
  - (a) 1463; all quantitative.
  - (b) four; two categorical and two quantitative.
  - (c) four; one categorical and three quantitative.
  - (d) three; two categorical and one quantitative.
  - (e) three; one categorical and two quantitative.
2. Consumers Union measured the gas mileage in miles per gallon of 38 1978–1979 model automobiles on a special test track. The pie chart below provides information about the country of manufacture of the model cars used by Consumers Union. Based on the pie chart, we may conclude that:
  - (a) Japanese cars get significantly lower gas mileage than cars of other countries. This is because their slice of the pie is at the bottom of the chart.
  - (b) U.S. cars get significantly higher gas mileages than cars from other countries.
  - (c) Swedish cars get gas mileages that are between those of Japanese and U.S. cars.
  - (d) Mercedes, Audi, Porsche, and BMW represent approximately a quarter of the cars tested.
  - (e) More than half of the cars in the study were from the United States.



3. A researcher reports that, on average, the participants in his study lost 10.4 lbs. after two months on his new diet. A friend of yours comments that she tried the diet for two months and lost no weight, so clearly the report was a fraud. Which of the following statements is correct?
  - (a) Your friend must not have followed the diet correctly, since she did not lose weight.
  - (b) Since your friend did not lose weight, the report must not be correct.
  - (c) The report only gives the average. This does not imply that all participants in the study lost 10.4 lbs. or even that all lost weight. Your friend's experience does not necessarily contradict the study results.
  - (d) In order for the study to be correct, we must now add your friend's results to those of the study and recompute the new average.
  - (e) Your friend is an outlier.

4. The following is an ogive on the number of ounces of alcohol (one ounce is about 30 mL) consumed per week in a sample of 150 students.

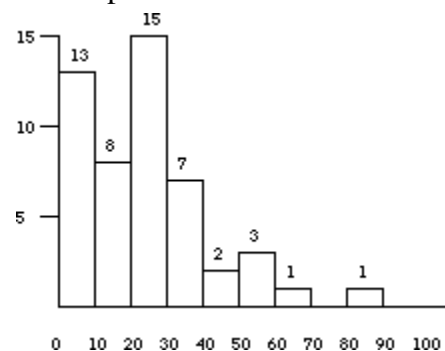


A study wished to classify the students as "light", "moderate", "heavy" and "problem" drinkers by the amount consumed per week. About what percentage of students are moderate drinkers, i.e., consume between 4 and 8 ounces per week?

- (a) 60%
  - (b) 20%
  - (c) 40%
  - (d) 80%
  - (e) 50%
5. Normal" body temperature varies by time of day. A series of readings was taken of the body temperature of a subject. The mean reading was found to be  $36.5^{\circ}\text{C}$  with a standard deviation of  $0.3^{\circ}\text{C}$ . When converted to  $^{\circ}\text{F}$ , the mean and standard deviation are:
- ( $^{\circ}\text{F} = ^{\circ}\text{C}(1.8) + 32$ ).
- (a) 97.7, 32
  - (b) 97.7, 0.30
  - (c) 97.7, 0.54
  - (d) 97.7, 0.97
  - (e) 97.7, 1.80

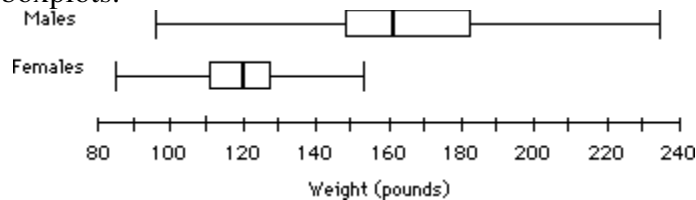
6. The following is a histogram showing the actual frequency of the closing prices on the New York exchange of a particular stock. Based on the frequency histogram for New York Stock exchange, the class that contains the 80th percentile is:

- (a) 20-30
- (b) 10-20
- (c) 40-50
- (d) 50-60
- (e) 30-40



7. Which of the following is likely to have a mean that is smaller than the median?
- The salaries of all National Football League players.
  - The scores of students (out of 100 points) on a very easy exam in which most get nearly perfect scores but a few do very poorly.
  - The prices of homes in a large city.
  - The scores of students (out of 100 points) on a very difficult exam in which most get poor scores but a few do very well.
  - Amounts awarded by civil court juries.
8. There are three children in a room, ages three, four, and five. If a four-year-old child enters the room the
- mean age will stay the same but the variance will increase.
  - mean age will stay the same but the variance will decrease.
  - mean age and variance will stay the same.
  - mean age and variance will increase.
  - mean age and variance will decrease.

9. The weights of the male and female students in a class are summarized in the following boxplots:



Which of the following is NOT correct?

- About 50% of the male students have weights between 150 and 185 lbs.
  - About 25% of female students have weights more than 130 lbs.
  - The median weight of male students is about 162 lbs.
  - The mean weight of female students is about 120 because of symmetry.
  - The male students have less variability than the female students.
10. When testing water for chemical impurities, results are often reported as bdl, i.e., below detection limit. The following are the measurements of the amount of lead in a series of water samples taken from inner city households (ppm).

5, 7, 12, bdl, 10, 8, bdl, 20, 6.

Which of the following is correct?

- The mean lead level in the water is about 10 ppm.
- The mean lead level in the water is about 8 ppm.
- The median lead level in the water is 7 ppm.
- The median lead level in the water is 8 ppm.
- Neither the mean nor the median can be computed because some values are unknown.

## Part 2: Free Response

*Communicate your thinking clearly and completely.*

11. The test grades for a certain class were entered into a Minitab worksheet, and then “Descriptive Statistics” were requested. The results were:

```
MTB > Describe 'Grades'.
```

	N	MEAN	MEDIAN	TRMEAN	STDEV	SEMEAN
Grades	28	74.71	76.00	75.50	12.61	2.38
	MIN	MAX	Q1	Q3		
Grades	35.00	94.00	68.00	84.00		

You happened to see, on a scrap of paper, that the lowest grades were 35, 57, 59, 60, . . . but you don't know what the other individual grades are. Nevertheless, a knowledgeable user of statistics can tell a lot about the dataset simply by studying the set of descriptive statistics above.

a. Write a brief description of what the results tell you about the distribution of grades.

Be sure to address:

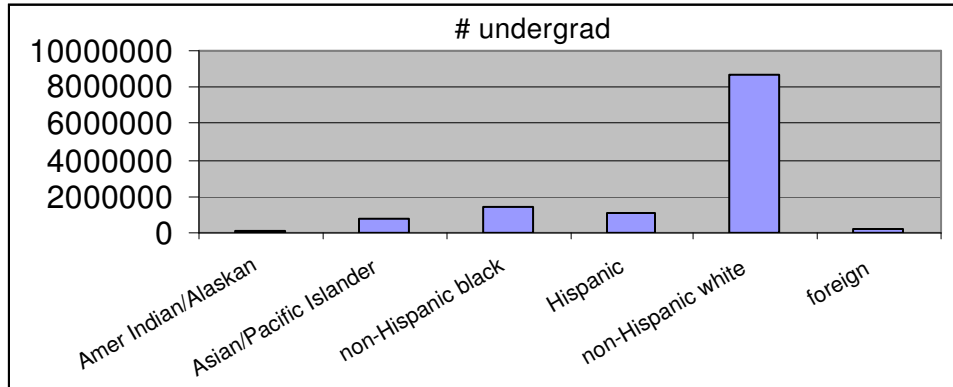
- the general shape of the distribution
- unusual features, including possible outliers
- the middle 50% of the data
- any significance in the difference between the mean and the median

b. Construct a modified boxplot for these data.

# Chapter 1 Solutions

## Quiz 1.1A

(1) (a) Yes. The actual total of the rounded numbers is 12299000. (b) Here is a bar graph of the data.



(2) Shown is a stemplot, with stems (the hundreds digit) split 5 ways; a histogram with classes 0–19, 20–39, etc. would have the same shape. The distribution is strongly right-skewed; the two highest values might be considered outliers. The mean is 48.25 thousand barrels, and the median is 37.80 thousand barrels; the mean is pulled in the direction of the skew.

```

0 0000111111111111
0 222222233333333333
0 44444445555555
0 6666667
0 8899
1 01
1
1 5
1
1 9
2 0
    
```

## Quiz 1.2A

(1) Since the mean is to be 7, the five numbers must add to 35. Also, the third number (in order from smallest to largest) must be 10, since that is the median. Beyond that, there is some freedom in how the numbers are chosen. **Note:** *It is likely that most students will interpret “positive numbers” as meaning positive integers only, which leads to eight possible solutions, shown below.*

```

1 1 10 10 13      1 1 10 11 12      1 2 10 10 12      1 2 10 11 11
1 3 10 10 11      1 4 10 10 10      2 2 10 10 11      2 3 10 10 10
    
```

(2) (a) The distribution is heavily right-skewed and has one high outlier. The median (1.8) and interquartile range (3.7) are the best numerical summaries of this skewed distribution.

Stem-and-leaf of Acorn si N = 39  
Leaf Unit = 1.0

```

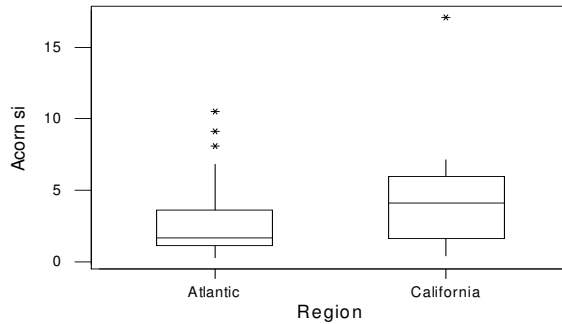
(20) 0 00000001111111111111
      19 0 22223333
      11 0 4455
      7 0 667
      4 0 89
      2 1 0
      1 1
      1 1
      1 1 7
    
```

**Descriptive Statistics: Acorn si**

Variable	N	Mean	Median	TrMean	StDev	SE Mean
Acorn si	39	3.341	1.800	2.914	3.453	0.553

Variable	Minimum	Maximum	Q1	Q3
Acorn si	0.300	17.100	1.100	4.800

(b) Side-by-side boxplots show that the Atlantic oak acorns are smaller in size than those from California. The Atlantic distribution is right-skewed; the California distribution is somewhat symmetric. Both distributions have high outliers. Due to the outliers, it is best to use the median and *IQR* to compare the acorn sizes. California oak acorns are much larger, but also more variable in size than Atlantic oak acorns.



**Descriptive Statistics: Acorn si by Region**

Variable	Region	N	Mean	Median	TrMean	StDev
Acorn si	Atlantic	28	2.750	1.700	2.546	2.732
	Californ	11	4.85	4.10	3.98	4.66

Variable	Region	SE Mean	Minimum	Maximum	Q1	Q3
Acorn si	Atlantic	0.516	0.300	10.500	1.100	3.60
	Californ	1.40	0.40	17.10	1.60	6.00

**Test 1A**

(1) e (2) e (3) c (4) c (5) c (6) e (7) b (8) b (9) e (10) c

(11) (a) The distribution is slightly skewed to the left because the mean is smaller than the median. (Actually, without the outlier (35), the distribution is very symmetric.) The 5-number summary is Min=35, Q1=68, Med=76, Q3=84, and Max=94. The interquartile range, or middle 50%, is  $Q3 - Q1 = 16$ . The test for outliers looks at  $Q1 - 1.5(IQR) = 68 - 24 = 44$ , so 35 is an outlier. At the other end,  $Q3 + 1.5(IQR) = 108$ , so there are no outliers at the high end. (b)

