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

1. Which of the following statements is NOT true?
   (a) In a symmetric distribution, the mean and the median are equal.
   (b) The first quartile is equivalent to the twenty-fifth percentile.
   (c) In a symmetric distribution, the median is halfway between the first and third quartiles.
   (d) The median is always greater than the mean.
   (e) The range is the difference between the largest and the smallest observation in the data set.

2. Consumers’ Union measured the gas mileage in miles per gallon of 38 automobiles from the same model year 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 mileage 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 pounds 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 gives only the average. This does not imply that all participants in the study lost 10.4 pounds 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 of the number of ounces of alcohol (one ounce is about 30 milliliters) consumed per week in a sample of 150 college students.

A study wished to classify the students as “light,” “moderate,” “heavy,” and “problem” drinkers by the amount consumed per week. About what percent of students are moderate drinkers, that is, 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°C with a standard deviation of 0.3°C. When converted to °F, the mean and standard deviation are (°F = °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 of a particular stock on the 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?
   (a) The salaries of all National Football League players.
   (b) 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.
   (c) The prices of homes in a large city.
   (d) 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.
   (e) 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
   (a) mean age will stay the same but the variance will increase.
   (b) mean age will stay the same but the variance will decrease.
   (c) mean age and variance will stay the same.
   (d) mean age and variance will increase.
   (e) 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?
   (a) About 50% of the male students have weights between 150 and 185 pounds.
   (b) About 25% of female students have weights more than 130 pounds.
   (c) The median weight of male students is about 162 pounds.
   (d) The mean weight of female students is about 120 pounds because of symmetry.
   (e) The male students have less variability than the female students.

10. When testing water for chemical impurities, results are often reported as bdl, that is, below detection limit. The following are the measurements of the amount of lead in a series of water samples taken from inner-city households (in parts per million):
    5, 7, 12, bdl, 10, 8, bdl, 20, 6

    Which of the following is correct?
    (a) The mean lead level in the water is about 10 ppm.
    (b) The mean lead level in the water is about 8 ppm.
    (c) The median lead level in the water is 7 ppm.
    (d) The median lead level in the water is 8 ppm.
    (e) Neither the mean nor the median can be computed because some values are unknown.
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 data set simply by studying the set of descriptive statistics above.

(a) Construct a modified boxplot for these data.

(b) 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
12. The University of Miami Hurricanes has been among the more successful teams in college football. The weights in pounds and positions of the players on the 2005 team were recorded. The positions are quarterback (QB), running back (RB), offensive line (OL), wide receiver (WR), tight end (TE), kicker/punter (KP), defensive back (DB), linebacker (LB), and defensive line (DL). Here are side-by-side boxplots of the weights.

(a) Briefly compare the weight distributions. Which position has the heaviest players overall? Which has the lightest?

(b) Are any individual players outliers within their position?

13. Give an example of a small data set for which the mean is greater than the third quartile. Indicate the mean and the third quartile.

I pledge that I have neither given nor received aid on this test. _________________________________________
Part 1: Multiple Choice.  Circle the letter corresponding to the best answer.

1. Mr. Yates picked up a dozen items in the grocery store with a mean cost of $3.25. Then he added an apple pie for $6.50. The new mean for all 13 items is
   (a) $3.00
   (b) $3.50
   (c) $3.75
   (d) $4.88
   (e) None of the above

Use the following to answer Question 2:

2. Which of the following bar graphs is equivalent to the pie chart?
   (a) ![Graph A]
   (b) ![Graph B]
   (c) ![Graph C]
   (d) ![Graph D]
   (e) None of these.
3. Consider the following ogive of the scores of students in an introductory statistics course:

A grade of C or C+ is assigned to a student who scores between 55 and 70. The percentage of students who obtained a grade of C or C+ is

(a) 25%
(b) 30%
(c) 20%
(d) 50%
(e) 15%

4. For the following histogram, what is the proper ordering of the mean and median? Note that the graph is NOT numerically precise—only the relative positions are important.

(a) I is the mean and II is the median.
(b) II is the median and III is the mean.
(c) I is the median and II is the mean.
(d) II is the mean and III is the median.
(e) I is the mean and III is the median.

5. A researcher wishes to calculate the average height of patients suffering from a particular disease. From patient records, the mean was computed as 156 cm, and standard deviation as 5 cm. Further investigation reveals that the scale was misaligned, and that all readings are 2 cm too large, for example, a patient whose height is really 180 cm was measured as 182 cm. Furthermore, the researcher would like to work with statistics based on meters. The correct mean and standard deviation are:

(a) 1.56m, 0.05m
(b) 1.54m, 0.05m
(c) 1.56m, 0.03m
(d) 1.58m, 0.05m
(e) 1.58m, 0.07m
6. A medical researcher collects health data on many women in each of several countries. One of the variables measured for each woman in the study is her weight in pounds. The following list gives the five-number summary for the weights of women in one of the countries.

Country A: 100, 110, 120, 160, 200

About what percent of Country A women weigh between 110 and 200 pounds?
(a) 50%
(b) 65%
(c) 75%
(d) 85%
(e) 95%

7. The median age of five people in a meeting is 30 years. One of the people, whose age is 50 years, leaves the room. The median age of the remaining four people in the room is

(a) 40 years.
(b) 30 years.
(c) 25 years.
(d) less than 30 years.
(e) Cannot be determined from the information given.

8. The time plot below gives the number of burglaries committed each month for a city in Ohio. The plot is for the three-year period January 1987 to December 1989.

Which of the following is a true statement?
(a) The number of burglaries in each month of 1988 was lower than the number of burglaries in each month of 1989.
(b) The median number of burglaries for a month in 1988 was a little over 25.
(c) The total number of burglaries in 1989 was higher than in 1988.
(d) More burglaries seem to be committed in June, July, and August during 1987, 1988, and 1989.
(e) None of the above.
9. Here is a summary graph of complex carbohydrates (in grams) for each of three fiber groups in a set of data related to cereals.

Which of the following is NOT correct?
(a) The low-fiber group is more variable than the medium-fiber group because the central box is larger.
(b) About 25% of low-fiber cereals have less than 12 g of complex carbohydrates per serving.
(c) About 50% of medium-fiber cereals have more than 15 g of complex carbohydrates per serving.
(d) The average amount of complex carbohydrates per serving for the high-fiber group appears to be much smaller than for the other two groups.
(e) About 25% of the medium-fiber cereals have less than 10 g of complex carbohydrates.

10. Earthquake intensities are measured using a device called a seismograph, which is designed to be most sensitive to earthquakes with intensities between 4.0 and 9.0 on the open-ended Richter scale. Measurements of nine earthquakes gave the following readings:

4.5  L  5.5  H  8.7  8.9  6.0  H  5.2

where L indicates that the earthquake had an intensity below 4.0 and H indicates that the earthquake had an intensity above 9.0. The median earthquake intensity of the sample is
(a) Cannot be computed since all of the values are not known
(b) 8.70
(c) 5.75
(d) 6.00
(e) 6.47
11. We all “know” that the body temperature of a healthy person is 98.6°F. In reality, the actual body temperature of individuals varies. Here is a back-to-back stemplot of the body temperatures of 130 healthy individuals (65 males and 65 females).

(a) Here are boxplots, produced by Minitab, for these distributions. Label both boxplots with the five-number summary values.

(b) Determine whether the 3 points graphed by the + symbol are indeed outliers by our defined criteria.

(c) Write a few sentences comparing the body temperatures of adult males and females.
12. The following data represent scores of 50 students on a calculus test.

```
72  72  93  70  59  78  74  65  73  80
57  67  72  57  83  76  74  56  68  67
74  76  79  72  61  72  73  76  67  49
71  53  67  65  100  83  69  61  72  68
65  51  75  68  75  66  77  61  64  74
```

(a) Construct a relative frequency histogram for this data set.

(b) Describe the shape, center, and spread of the distribution of test scores.

(c) Would the mean and standard deviation be appropriate measures of center and spread for these test scores? Explain.

---

*I pledge that I have neither given nor received aid on this test.*

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

1. Here are the IQ test scores of 10 randomly chosen fifth-grade students:
   \[145 \quad 139 \quad 126 \quad 122 \quad 125 \quad 130 \quad 96 \quad 110 \quad 118 \quad 118\]
   To make a stemplot of these scores, you would use as stems
   (a) 0 and 1.
   (b) 09, 10, 11, 12, 13, and 14.
   (c) 96, 110, 118, 122, 125, 126, 130, 139, and 145.
   (d) 0, 2, 3, 5, 6, 8, 9.
   (e) None of the above is a correct answer.

2. For the IQ test scores from the previous question, what kind of plot is appropriate?
   (a) a stemplot but not a boxplot
   (b) a boxplot but not a histogram
   (c) a bar graph but not a pie chart
   (d) a histogram but not a dotplot
   (e) None of the above is a correct answer.

3. Rainwater was collected in water collectors at 30 different sites near an industrial complex and the amount of acidity (pH level) was measured. The data ranged from pH 2.6 to pH 6.3. The following stemplot of the data was constructed.
   \[2| 679\]
   \[3| 237789\]
   \[4| 1222446899\]
   \[5| 0556788\]
   \[6| 0233\]
   Which of the following boxplots is correct?

4. A scientist is weighing each of 30 fish. She obtains a mean of 30 g and a standard deviation of 2 g. After completing the weighing, she finds that the scale was misaligned and always under reported every weight by 2 g that is, a fish that really weighed 26 g was reported to weigh 24 grams. What are the mean and standard deviation after correcting for the error in the scale?
   (a) 28 g, 2 g  (b) 30 g, 4 g  (c) 32 g, 2 g  (d) 32 g, 4 g  (e) 28 g, 4 g
5. If a distribution is skewed to the right,
   (a) the mean is less than the median.
   (b) the mean and median are equal.
   (c) the mean is greater than the median.
   (d) It’s impossible to tell without seeing the data.
   (e) None of the above is a correct answer.

6. The population of the United States is aging, though less rapidly than in other developed countries. Here is a stemplot of the percents of residents aged 65 and older in the 50 states, according to the 2000 census:

   There are two outliers: Alaska has the lowest percent of older residents, and Florida has the highest. What is the percent for Florida?

   | 5 | 7
   | 6 |
   | 7 |
   | 8 | 5
   | 9 | 679
   | 10 | 6
   | 11 | 0223677
   | 12 | 0011113445789
   | 13 | 00012233345568
   | 14 | 034579
   | 15 | 36
   | 16 |
   | 17 | 6

7. Ignoring the outliers, the shape of the distribution in the previous question is
   (a) strongly skewed to the right.
   (b) slightly skewed to the right.
   (c) roughly symmetric.
   (d) slightly skewed to the left.
   (e) strongly skewed to the left.

8. The center of the distribution in Question 6 is close to
   (a) 12.7%.
   (b) 13.5%.
   (c) 13.8%.
   (d) 12.
   (e) It’s impossible to tell.

9. To make a boxplot of a distribution, you must know
   (a) all of the individual observations.
   (b) the mean and the standard deviation.
   (c) the quartiles.
   (d) the five-number summary.
   (e) the individual observations, the mean, and the IQR.

10. What are all the values that a standard deviation can possibly take?
    (a) \(0 \leq s\)
    (b) \(0 \leq s \leq 1\)
    (c) \(-1 \leq s \leq 1\)
    (d) \(s \leq 0\)
    (e) any real number
Part 2: Free Response

Communicate your thinking clearly and completely.

11. Here are data on the percent of people in several age groups who attended a movie in the past 12 months:

<table>
<thead>
<tr>
<th>Age group</th>
<th>Movie attendance</th>
</tr>
</thead>
<tbody>
<tr>
<td>18 to 24 years</td>
<td>83%</td>
</tr>
<tr>
<td>25 to 34 years</td>
<td>73%</td>
</tr>
<tr>
<td>35 to 44 years</td>
<td>68%</td>
</tr>
<tr>
<td>45 to 54 years</td>
<td>60%</td>
</tr>
<tr>
<td>55 to 64 years</td>
<td>47%</td>
</tr>
<tr>
<td>65 to 74 years</td>
<td>32%</td>
</tr>
<tr>
<td>75 years and over</td>
<td>20%</td>
</tr>
</tbody>
</table>

(a) Display these data in a bar graph in the space above. What is the main feature of the data?

(b) Would it be correct to make a pie chart of these data? Why?

(c) A movie studio wants to know what percent of the total audience for movies is 18 to 24 years old. Explain why these data do not answer this question.

12. Here is a histogram of the number of unprovoked attacks by alligators on people in Florida over a 33-year period. The classes are “1 ≤ attacks < 3,” “3 ≤ attacks < 5,” and so on.

<table>
<thead>
<tr>
<th>Attacks</th>
<th>N =</th>
<th>33</th>
<th>N* =</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Midpoint</td>
<td>Count</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(a) What is the overall shape of the distribution?

(b) What is the approximate median of the yearly counts of alligator attacks?
(c) Here is a time plot for the alligator data. Connect the numbers from 1 to 8 four times, then 1.

What overall pattern does this plot show?

(d) Why is the typical number of attacks from 1972 to 2004 not very useful in, say, 2006?

13. The states differ greatly in the kinds of severe weather that afflict them. The table below shows the average property damage caused by tornadoes per year over the period from 1950 to 1999 in each of the 50 states and Puerto Rico. (To adjust for the changing buying power of the dollar over time, all damages were restated in 1999 dollars.)

<table>
<thead>
<tr>
<th>State</th>
<th>($)millions</th>
<th>State</th>
<th>($)millions</th>
<th>State</th>
<th>($)millions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alabama</td>
<td>51.88</td>
<td>Louisiana</td>
<td>27.75</td>
<td>Ohio</td>
<td>44.36</td>
</tr>
<tr>
<td>Alaska</td>
<td>0.00</td>
<td>Maine</td>
<td>0.53</td>
<td>Oklahoma</td>
<td>81.94</td>
</tr>
<tr>
<td>Arizona</td>
<td>3.47</td>
<td>Maryland</td>
<td>2.33</td>
<td>Oregon</td>
<td>5.52</td>
</tr>
<tr>
<td>Arkansas</td>
<td>40.96</td>
<td>Mass.</td>
<td>4.42</td>
<td>Penn.</td>
<td>17.11</td>
</tr>
<tr>
<td>California</td>
<td>3.68</td>
<td>Michigan</td>
<td>29.88</td>
<td>Puerto Rico</td>
<td>0.05</td>
</tr>
<tr>
<td>Colorado</td>
<td>4.62</td>
<td>Minnesota</td>
<td>84.84</td>
<td>Rhode Island</td>
<td>0.05</td>
</tr>
<tr>
<td>Connecticut</td>
<td>2.26</td>
<td>Mississippi</td>
<td>43.62</td>
<td>S Carolina</td>
<td>17.19</td>
</tr>
<tr>
<td>Delaware</td>
<td>0.27</td>
<td>Missouri</td>
<td>68.93</td>
<td>S Dakota</td>
<td>10.64</td>
</tr>
<tr>
<td>Florida</td>
<td>37.32</td>
<td>Montana</td>
<td>2.27</td>
<td>Tennessee</td>
<td>23.47</td>
</tr>
<tr>
<td>Georgia</td>
<td>51.68</td>
<td>Nebraska</td>
<td>30.26</td>
<td>Texas</td>
<td>88.60</td>
</tr>
<tr>
<td>Hawaii</td>
<td>0.34</td>
<td>Nevada</td>
<td>0.10</td>
<td>Utah</td>
<td>3.57</td>
</tr>
<tr>
<td>Idaho</td>
<td>0.26</td>
<td>N Hampshire</td>
<td>0.66</td>
<td>Vermont</td>
<td>0.24</td>
</tr>
<tr>
<td>Illinois</td>
<td>62.94</td>
<td>New Jersey</td>
<td>2.94</td>
<td>Virginia</td>
<td>7.42</td>
</tr>
<tr>
<td>Indiana</td>
<td>53.13</td>
<td>New Mexico</td>
<td>1.49</td>
<td>Washington</td>
<td>2.37</td>
</tr>
<tr>
<td>Iowa</td>
<td>49.51</td>
<td>New York</td>
<td>15.73</td>
<td>W Virginia</td>
<td>2.14</td>
</tr>
<tr>
<td>Kansas</td>
<td>49.28</td>
<td>N Carolina</td>
<td>14.90</td>
<td>Wisconsin</td>
<td>31.33</td>
</tr>
<tr>
<td>Kentucky</td>
<td>24.84</td>
<td>N Dakota</td>
<td>14.69</td>
<td>Wyoming</td>
<td>1.78</td>
</tr>
</tbody>
</table>
(a) Here is a histogram of the data, with classes “0 ≤ damage < 10,” “10 ≤ damage < 20,” and so on.

![Histogram](image)

Describe the shape, center, and spread of the distribution. Which states may be outliers? (To understand the outliers, note that most tornadoes in largely rural states such as Kansas cause little property damage. Damage to crops is not counted as property damage.)

(b) Here is the “default” histogram that the calculator makes for these data. How does this calculator histogram compare with the graph in (a)?

(c) What are the top five states for tornado damage? The bottom five? (Include Puerto Rico, though it is not a state.)

(d) Give the five-number summary. Explain why you can see from these five numbers that the distribution is strongly skewed to the right.

(e) The histogram suggests that a few states are outliers. Show that there are no suspected outliers according to the $1.5 \times IQR$ rule. (You see once again that a rule is not a substitute for plotting your data.)

(f) Find the mean property damage. Explain why the mean and median differ so greatly for this distribution.

I pledge that I have neither given nor received aid on this test. ________________________________