Describing Data
****

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 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.
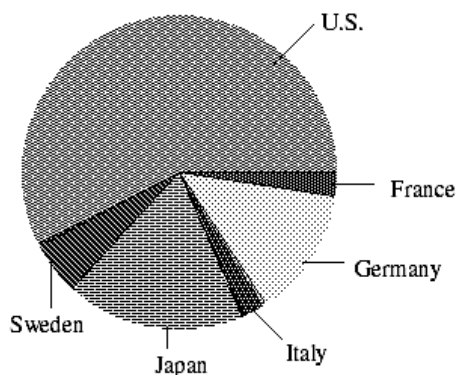
3. A researcher reports that, on average, the participants in his study lost 10.4 pounds after two months on his new diet. A friend of yours comments that she tried the diet for two months and lost no weight, so clearly the report was a fraud. Which of the following statements is correct?
(a) Your friend 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 only gives 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 re-compute 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.
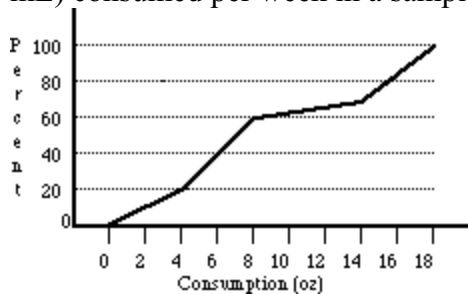

A study wished to classify the students as "light", "moderate", "heavy" and "problem" drinkers by the amount consumed per week. About what percentage of students are moderate drinkers, that is consume between 4 and 8 ounces per week?
(a) 60%
(b) 20%
(c) 40%
(d) 80%
(e) 50%

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. 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.

7. 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.

8. The weights of the male and female students in a class are summarized in the following boxplots:
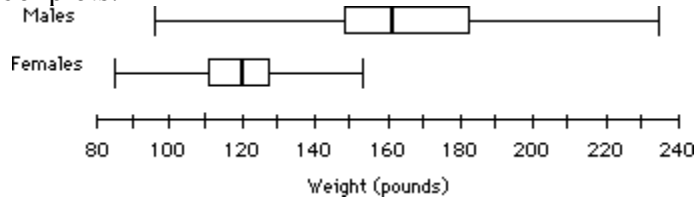
   Which of the following is NOT correct?
   (a) About 50% of the male students have weights between 150 and 185 pounds.
   (b) About 25% of female students have weights more than 130 pounds.
   (c) The median weight of male students is about 162 pounds.
   (d) The mean weight of female students is about 120 pounds because of symmetry.
   (e) The male students have less variability than the female students.

9. 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 (ppm).

   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.

Part 2: Free Response

Communicate your thinking clearly and completely.

10. 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

Normal Calculations
****

1.   A company produces packets of soap powder labeled "Giant Size 32 Ounces." The actual weight of soap powder in a box has a normal distribution with a mean of 33 oz. and a standard deviation of 0.8 oz. What proportion of packets are underweight (i.e., weigh less than 32 oz.)?
(a) 0.159
(b) 0.212
(c) 0.106
(d) 0.841
(e) 0.115

2.   For the density curve shown to the right, what percent of the observations lie above 1.5?
(a) 20%
(b) 25%
(c) 50%
(d) 75%
(e) 80%
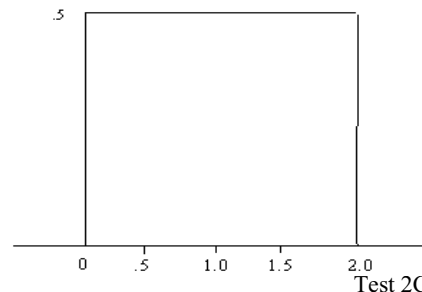
3. For the above density curve, what percent of the observations lie between 0.5 and 1.2?
   (a) 25%
   (b) 35%
   (c) 50%
   (d) 68%
   (e) 70%

4. If the heights of 99.7% of American men are between 5′0″ and 7′0″, what is your estimate of the standard deviation of the height of American men?
   (a) 1"
   (b) 3"
   (c) 4"
   (d) 6"
   (e) 12"

5. The figure below is the density curve of a distribution:
   Five of the seven points marked on the density curve make up the five-number summary for this distribution. Which two points are *not* part of the five-number summary?
   (a) B and E.
   (b) C and F.
   (c) C and E.
   (d) B and F.
   (e) A and G.
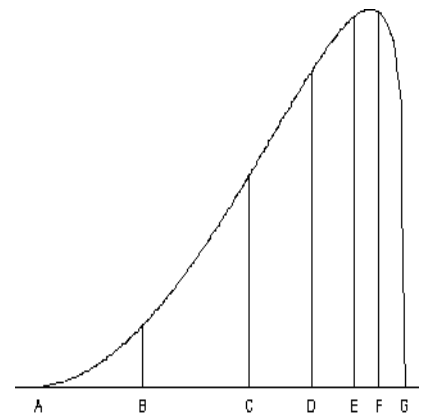
6. Suppose that the distribution of math SAT scores from your state this year is normally distributed with mean 480 and standard deviation 100 for males, and mean 440 and standard deviation 120 for females. If someone who scores 780 or higher on math SAT can be considered a genius, what is the proportion of geniuses among the male SAT takers?

   (a) 28%   (b) 14% (b) 3%   (d) 1.4% (e) 0.14%

7. The average yearly snowfall in Chillyville is normally distributed with a mean of 55 inches. If the snowfall in Chillyville exceeds 60 inches in 15% of the years, what is the standard deviation?
   (a) 4.83 inches   (b) 5.18 inches   (c) 6.04 inches   (d) 8.93 inches
   (e) The standard deviation cannot be computed from the given information.

8. The following graph is a normal probability plot for the amount of rainfall in acre-feet obtained from 26 randomly selected clouds that were seeded with silver oxide:
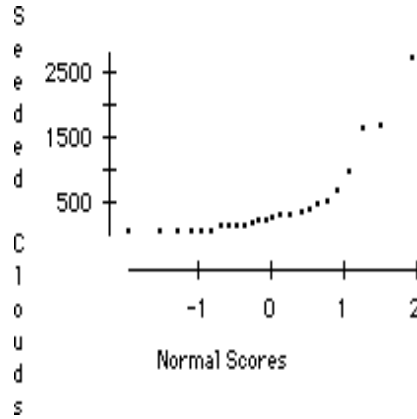
(a) The data appear to show exponential growth; that is, the amount of rainfall increases exponentially as the amount of silver oxide increases.
(b) The pattern suggests that the measurement is not normally distributed.
(c) A least squares regression line should be fitted to the rainfall variable.
(d) It can be expected that the histogram of rainfall amount will look like the normal curve.
(e) The shape of the curve suggests that rainfall is caused by seeding the clouds with silver oxide.

9. The five-number summary of the distribution of scores on a statistics exam is

        0   26   31   36   50

316 students took the exam. The histogram of all 316 test scores was approximately normal. Thus the variance of test scores must be about
(a) 5
(b) 8
(c) 19
(d) 64
(e) 55

10. If the median of a set of data is equal to the mean, then
(a) The data are normally distributed.
(b) The data are approximately distributed.
(c) The distribution is skewed.
(d) The distribution is symmetric.
(e) One can't say anything about the shape of the distribution with any certainty.

Linear Regression and Correlation
****

1. A study found correlation $r = 0.61$ between the sex of a worker and his or her income. You conclude that
   (a) Women earn more than men on the average.
   (b) Women earn less than men on average.
   (c) An arithmetic mistake was made; this is not a possible value of r.
   (d) This is nonsense because r makes no sense here.

2. A copy machine dealer has data on the number x of copy machines at each of 89 customer locations and the number y of service calls in a month at each location. Summary calculations give $\bar{x} = 8.4$, $s_x = 2.1$, $\bar{y} = 14.2$, $s_y = 3.8$, and r = 0.86. What is the slope of the least squares regression line of number of service calls on number of copiers?
   (a) 0.86
   (b) 1.56
   (c) 0.48
   (d) None of these
   (e) Can't tell from the information given

3. In the setting of the previous problem, about what percent of the variation in the number of service calls is explained by the linear relation between number of service calls and number of machines?
   (a) 86%
   (b) 93%
   (c) 74%
   (d) None of these
   (e) Can't tell from the information given

4. If dataset A of (x,y) data has correlation coefficient $r = 0.65$, and a second dataset B has correlation $r = -0.65$, then
   (a) The points in A exhibit a stronger linear association than B.
   (b) The points in B exhibit a stronger linear association than A.
   (c) Neither A nor B has a stronger linear association.
   (d) You can't tell which dataset has a stronger linear association without seeing the data or seeing the scatterplots.

5. There is a linear relationship between the number of chirps made by the striped ground cricket and the air temperature. A least squares fit of some data collected by a biologist gives the model $\hat{y} = 25.2 + 3.3x$, $9 < x < 25$, where x is the number of chirps per minute and $\hat{y}$ is the estimated temperature in degrees Fahrenheit. What is the estimated increase in temperature that corresponds to an increase in 5 chirps per minute?

(a) 3.3°F   (b) 16.5°F   (c) 25.2°F   (d) 28.5°F
(e) 41.7°F

6. The equation of the least squares regression line for the points on the scatterplot below is $\hat{y} = 1.3 + 0.73x$. What is the residual for the point (4,7)?
(a) 2.78
(b) 3.00
(c) 4.00
(d) 4.22
(e) 7.00

7. Linear regression usually employs the method of least squares. Which of the following is the quantity that is minimized by the least squares process?
(a) $\hat{y}_i$
(b) $x_i \quad \bar{x}_i$
(c) $\Sigma (y_i \quad \hat{y}_i)^2$
(d) $(\bar{x}_i, \bar{y}_i)$
(e) $\Sigma (x_i \quad \bar{x}_i)^2$

8. A set of data relates the amount of annual salary raise and the performance rating. The least squares regression equation is $\hat{y} = 1{,}400 + 2{,}000x$ where y is the estimated raise and x is the performance rating. Which of the following statements is *not* correct?
(a) For each increase of one point in performance rating, the raise will increase on average by $2,000.
(b) This equation produces predicted raises with an average error of 0.
(c) A rating of 0 will yield a predicted raise of $1,400.
(d) The correlation for the data is positive.
(e) All of the above are true.

9. Which of the following would *not* be a correct interpretation of a correlation of $r = -.30$?
(a) The variables are inversely related.
(b) The coefficient of determination is 0.09.
(c) 30% of the variation between the variables is linear.
(d) There exists a weak relationship between the variables.
(e) All of the above statements are correct.

10. The following are resistant:
(a) Least squares regression line
(b) Correlation coefficient
(c) Both the least square line and the correlation coefficient
(d) Neither the least square line nor the correlation coefficient
(e) It depends