MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

1) School administrators collect data on students attending the school. Which of the following variables is quantitative?

A) whether the student is in AP classes  
B) class (freshman, soph., junior, senior)  
C) whether the student has taken the SAT  
D) grade point average

2) A professor has kept records on grades that students have earned in his class. If he wants to examine the percentage of students earning the grades A, B, C, D, and F during the most recent term, which kind of plot could he make?

A) dotplot  
B) timeplot  
C) pie chart  
D) boxplot  
E) histogram

3) Which is true of the data shown in the histogram?

I. The distribution is approximately symmetric.
II. The mean and median are approximately equal.
III. The median and IQR to summarize these data summarize the data better than the mean and standard deviation.

A) I only  
B) III only  
C) I and III  
D) I, II, and III  
E) I and II
4) Two sections of a class took the same quiz. Section A had 15 students who had a mean score of 80, and Section B had 20 students who had a mean score of 90. Overall, what was the approximate mean score for all of the students on the quiz?

   A) 85.0  
   B) 84.3  
   C) none of these  
   D) 85.7  
   E) It cannot be determined.

5) It is permissible to violate the area principle if
   A) we need a flashy display to make a point.
   B) we do not care about being true to the data.
   C) the percentages do not add up to 100%.
   D) None of the above.

6) Suppose that a Normal model described student scores in a history class. Parker has a standardized score (z-score) of +2.5. This means that Parker
   A) is 2.5 points above average for the class.
   B) is 2.5 standard deviations above average for the class.
   C) has a score that is 2.5 times the average for the class.
   D) has a standard deviation of 2.5.
   E) None of the above.

7) The advantage of making a stem-and-leaf display instead of a dotplot is that a stem-and-leaf display
   A) preserves the individual data values.
   B) shows the shape of the distribution better than a dotplot.
   C) A stem-and-leaf display has no advantages over a dotplot.
   D) satisfies the area principle.

8) The five-number summary of credit hours for 24 students in a statistics class is:
<table>
<thead>
<tr>
<th>Min</th>
<th>Q1</th>
<th>Median</th>
<th>Q3</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>13.0</td>
<td>15.0</td>
<td>16.5</td>
<td>18.0</td>
<td>22.0</td>
</tr>
</tbody>
</table>

   From this information, we know that
   A) there are no outliers in the data.
   B) there is at least one low outlier in the data.
   C) there are both low and high outliers in the data.
   D) there is at least one high outlier in the data.
   E) None of the above.

9) The SPCA collect data on breed, age, weight, number of days housed, and veterinary costs for the dogs they house. The SPCA has kept these data records for the past 20 years. If they want to show the trend in the number of dogs they have housed, what kind of plot should they make?

   A) pie chart  
   B) boxplot  
   C) timeplot  
   D) bar graph  
   E) histogram
10) Which of the following data summaries are changed by adding a constant to each data value?
   I. the mean
   II. the median
   III. the standard deviation
   A) I and III
   B) I and II
   C) III only
   D) I only
   E) I, II, and III

We collect data from 50 male students on eye color, head circumference, hours of homework last week, number of cigarettes smoked daily, and number of TV sets at home.

11) Which variable is categorical?
   A) hours of homework last week
   B) eye color
   C) head circumference
   D) number of TV sets at home
   E) number of cigarettes smoked daily

12) A correlation of zero between two quantitative variables means that
   A) there is no association between the two variables.
   B) we have done something wrong in our calculation of $r$.
   C) re-expressing the data will guarantee a linear association between the two variables.
   D) there is no linear association between the two variables.
   E) None of the above.

13) Which of the following is not a goal of re-expressing data?
   A) Make the distribution of a variable more symmetric.
   B) Make the scatter in a scatterplot spread out evenly rather than following a fan shape.
   C) Make the spread of several groups more alike.
   D) Make the form of a scatterplot more nearly linear.
   E) All of the above are goals of re-expressing data.

14) The correlation between $X$ and $Y$ is $r = 0.35$. If we double each $X$ value, decrease each $Y$ by 0.20, and interchange the variables (put $X$ on the $Y$-axis and vice versa), the new correlation
   A) is 0.50
   B) is 0.70
   C) is 0.90
   D) is 0.35
   E) cannot be determined.
SHORT ANSWER. Write the word or phrase that best completes each statement or answers the question.

15) Costs for standard veterinary services at a local animal hospital follow a Normal model with a mean of $80 and a standard deviation of $20. Draw and clearly label this model.

![Normal distribution graph]

Has the percentage of young girls drinking milk changed over time? The following table is consistent with the results from "Beverage Choices of Young Females: Changes and Impact on Nutrient Intakes" (Shanthy A. Bowman, *Journal of the American Dietetic Association*, 102(9), pp. 1234–1239):

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>354</td>
<td>502</td>
<td>366</td>
<td></td>
<td>1222</td>
</tr>
<tr>
<td>No</td>
<td>226</td>
<td>335</td>
<td>366</td>
<td></td>
<td>927</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>580</strong></td>
<td><strong>837</strong></td>
<td><strong>732</strong></td>
<td></td>
<td><strong>2149</strong></td>
</tr>
</tbody>
</table>

16) Find the following:

a. What percent of the young girls reported that they drink milk?
   Soln: ... \[ \frac{580}{2149} \times 100 \approx 27.2\% \]

b. What percent of the young girls were in the 1989–1991 survey?
   Soln: ... \[ \frac{837}{2149} \times 100 \approx 39.0\% \]

c. What percent of the young girls who reported that they drink milk were in the 1989–1991 survey?
   Soln: ... \[ \frac{580}{2149} \times 100 \approx 27.2\% \]

d. What percent of the young girls in 1989–1991 reported that they drink milk?
   Soln: ... \[ \frac{580}{837} \approx 69.0\% \]
Suppose that the student taking 22 credit hours in the data set in the previous was actually taking 28 credit hours instead of 22 (so we would replace the 22 in the data set with 28). Indicate whether changing the number of credit hours for that student would make each of the following summary statistics increase, decrease, or stay about the same:

a. mean
   increase
b. median
   stay same
c. range
   increase
d. IQR
   stay same
e. standard deviation
   increase

(B): Suppose the class took a 40-point quiz. Results show a mean score of 30, median 32, IQR 8, SD 6, min 12 and Q1 27. (Suppose YOU got a 35). What happens to each of the statistics if . . .

I decide to count the quiz as 100 points; I’ll double each score and add 20 points. (Now YOUR score is 90).

Now fill out the table by examining the effects of the transformations.

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Original (X)</th>
<th>2*X + 20</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>30</td>
<td>80</td>
</tr>
<tr>
<td>Median</td>
<td>32</td>
<td>84</td>
</tr>
<tr>
<td>IQR</td>
<td>8</td>
<td>16</td>
</tr>
<tr>
<td>SD</td>
<td>6</td>
<td>12</td>
</tr>
<tr>
<td>Minimum</td>
<td>12</td>
<td>44</td>
</tr>
<tr>
<td>Q1</td>
<td>27</td>
<td>74</td>
</tr>
<tr>
<td>Your score</td>
<td>35</td>
<td>90</td>
</tr>
</tbody>
</table>
The boxplots show the age of people involved in accidents according to their role in the accident.

Answer the followings:

a. Which role involved the youngest person, and what is the age?
   Soln: ... Passenger, less than 1 year.

b. Which role had the lowest median age, and what is the age?
   Soln: ... Passenger, ≥ 21 years

c. Which role had smallest range of age, and what is it?
   Soln: ... Cyclist, ≥ 40 years

d. Which role had the largest IQR of age, and what is it?
   Soln: ... Pedestrian, ≥ 44 years

e. Which role generally involves the oldest people?
   Soln: ... Pedestrian
19). The times that the city bus arrives at a certain apartment complex each morning are approximately normally distributed, with an average time of 7:30 AM and a standard deviation of 10 minutes. {Hint: you may convert 7:30am = 7*60+30 = 450 min...and convert back later}

a) About what percentage of the time will the bus arrive between 7:25 and 7:35?

\[
\frac{Z}{7:25} = \frac{-0.5}{0.5} \rightarrow 30.85\% \\
Z_{7:35} = 0.5 \rightarrow 69.15\% \\
\text{Subtracting} \rightarrow \boxed{38.3\%}
\]

b) About what percentage of the time will the bus arrive between 7:15 and 7:45?

\[
\frac{Z}{7:15} = \frac{-1.5}{0.5} \rightarrow 6.68\% \\
Z_{7:45} = 1.5 \rightarrow 93.32\% \\
\text{Subtracting} \rightarrow \boxed{86.64\%}
\]

c) If you want to have a 95% certainty that you will catch the bus (that is, you arrive before the bus does), what time should you arrive at the stop?

\[
X = Z \cdot s_x + \bar{x} = -16.4 + 7:30 \\
\text{7:13 am to 7:14 am have way between}
\]

d) 90% of the time, the bus will arrive no later than ______.

\[
X = 12.8 + 7:30 \rightarrow 7:43 \text{ am}
\]

20): Find the Standard Deviation of \{1, 2, 3, 4, 5, 6, 7, 8, 9\}:

\[
s = \sqrt{\frac{\sum(y_i - \bar{y})^2}{n-1}}
\]

\[
S_{\text{dev}}: \text{Here, } \bar{y} = 5 \\
(1-5)^2 = 16, (2-5)^2 = 9, \ldots; (9-5)^2 = 16 \text{ add all result = 60} \\
\text{Divide by } n-1 = 8 \rightarrow 60/8 = 7.5 \\
\text{Take sq. root} \rightarrow \sqrt{7.5} \approx 2.74
\]
(BONUS, 10 Points; show details to get full credit)

A company's manufacturing process uses 500 gallons of water at a time. A "scrubbing" machine then removes most of a chemical pollutant before pumping the water into a nearby lake. Legally the treated water should contain no more than 80 parts per million of the chemical, but the machine isn't perfect and it is costly to operate. Since there's a fine if the discharged water exceeds the legal maximum, the company sets the machine to attain an average of 75 ppm for the batches of water treated. They believe the machine's output can be described by a Normal model with standard deviation 4.2 ppm.

a. What percent of the batches of water discharged exceed the 80 ppm standard?

\[
Z_0 = \frac{80 - \mu}{\sigma} = \frac{80 - 75}{4.2} = 1.19 
\Rightarrow 88.2\%
\]

\[
\therefore \text{Percent exceed} = 100\% - 88.2\% = 11.8\%
\]

b. The company's lawyers insist that they have not more than 2% of the water over the limit. To what mean value should the company set the scrubbing machine? Assume the standard deviation does not change.

\[
Z = \frac{X - \mu}{\sigma} = \frac{X - 75}{4.2} = 2.05
\]

\[
X = \mu + Z \cdot \sigma = 71.37
\]

c. Because achieving a mean that low would raise the costs too much, they decide to leave the mean set at 75 ppm and try to reduce the standard deviation to achieve the "only 2% over" goal. Find the new standard deviation needed.

\[
\sigma_0 = \frac{X - \mu}{Z} = \frac{80 - 75}{2.05} = \frac{80 - 75}{2.05} = 2.44
\]

d. Explain what achieving a smaller standard deviation means in this context.

The scrubber must be more consistent in its performance from batch to batch.

Hurray!... You are Done!!