

An insurance company evaluates many variables about a person before deciding on an appropriate rate for automobile insurance. For the first four questions, use the following answer set and select the type of data represented by each variable.

- | | | | |
|----|-----------------------|----|-------------------------|
| a. | qualitative | c. | quantitative-continuous |
| b. | quantitative-discrete | d. | qualitative-numerical |

1. How long an applicant has been licensed to drive
2. Number of insurance claims an applicant has made in the last three years
3. Distance driven per year
4. Existence of air bags
5. If you wish to make inferences from a sample to a population, which of the following types of samples can you use?
 - a. a judgment sample
 - b. a quota sample
 - c. a chunk sample
 - d. a probabilistic sample
6. Which of the following statements is **TRUE?**
 - a. Complete populations are often inaccessible.
 - b. As a population becomes large, it is usually better to obtain statistical information from the entire population.
 - c. We use samples because we want absolute accuracy.
 - d. We always have plenty of time to do a study.
7. Which of the following is most likely a parameter as opposed to a statistic?
 - a. the mean score of the first five students who turn in this exam.
 - b. the proportion of females registered to vote in Monroe County.
 - c. the mean height of people randomly selected from a data base.
 - d. the proportion of trucks stopped on US-37 yesterday that were cited for bad brakes.

8. The evening host of a dinner dance reached into a bowl, mixed all the tickets around, and selected the ticket to award the grand door prize. What sampling method was used?
- a. simple random sample
 - b. systematic sample
 - c. stratified sample
 - d. cluster sample

Refer to the following table of data to answer the next three questions. At a meeting of information systems officers for regional offices of a national company, a survey was taken to determine the number of employees the officers supervise in the operation of their departments, where X is the number of employees supervised by each information systems officer.

X	f
1	7
2	5
3	11
4	8
5	9

9. How many regional offices are represented in the survey results?
- a. 127
 - b. 40
 - c. 15
 - d. 5
10. What is the total number of employees supervised by those surveyed?
- a. 127
 - b. 40
 - c. 15
 - d. 5
11. Estimate the mean number of employees supervised by an information systems officer.
- a. 1.5
 - b. 3
 - c. 3.175
 - d. 8

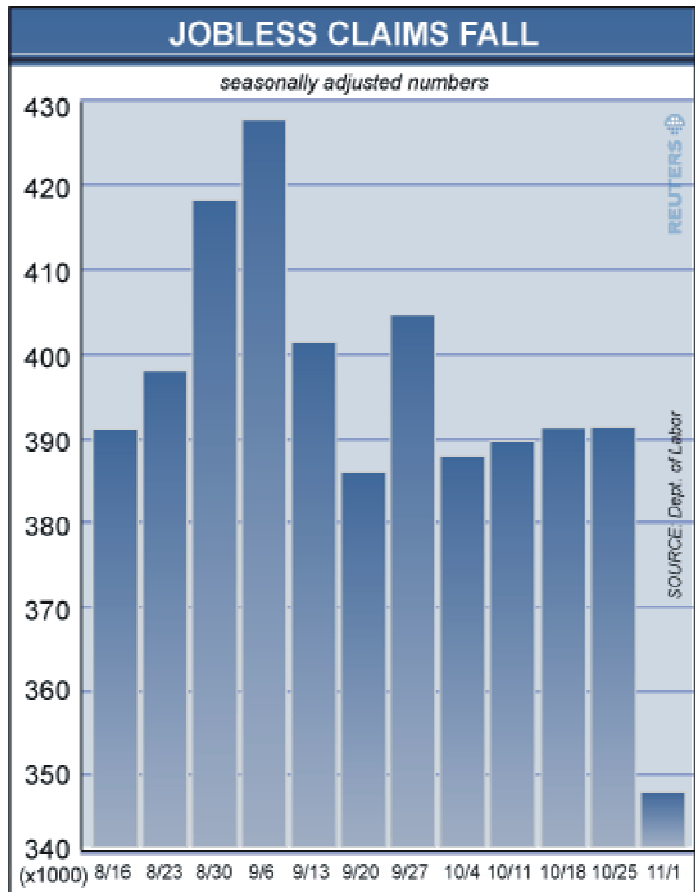
Refer to the following table to answer the next three questions. The following are the duration in minutes of a sample of long-distance phone calls made within the continental United States reported by one long-distance carrier.

Time (in minutes)	Relative Frequency
0 up to 5	0.37
5 up to 10	0.22
10 up to 15	0.15
15 up to 20	0.10
20 up to 25	0.07
25 up to 30	0.07
30 or more	0.02

12. If 1000 calls were randomly sampled, how many calls lasted under 10 minutes?
- a. 220 b. 370 c. 410 d. 590
13. If 10 calls lasted 30 minutes or more, what percent of calls lasted less than 5 minutes?
- a. 185 b. .185 c. 0.63 d. 37
14. Referring to the table above and using your knowledge of measures of center and their relationship to one another, in what order would you expect to find the mean, median and mode for this data set?
- a. mean < median < mode
b. mode < median < mean
c. median < mode < mean
d. median < mean < mode
15. Which of the following statements is TRUE?
- a. The sum of relative frequencies in a distribution always equals one.
b. The sum of cumulative frequencies in a distribution always equals one.
c. The sum of cumulative percentages in a distribution always equals one.
d. None, that is, they are all false.

16. The most appropriate type of chart for determining the number of observations at or below a specific value is:
- ogive
 - time-series
 - pie chart
 - histogram
17. For which of the following data sets would you wish to construct a Pareto Diagram?
- Political Party Affiliation for a field of three parties, all approximately equal in proportion of affiliations.
 - The ages of 25 salespersons with a range of 10 years.
 - The number of students selecting one of 23 majors, where 18 major fields attract 23% of the students.
 - The oral temperature in degrees Fahrenheit of 30 students presenting themselves at the health center last Friday.
18. Of the following characteristics of graphical excellence, which is/are ***VIOLATED*** in the following graph? A) Shows the data; B) Directs viewer to the substance of the data; and C) Does not distort the data?

- | |
|--|
| <ol style="list-style-type: none"> A only B only C only A and B only A and C only |
|--|



19. The mean score on an exam for a class of 30 students was 75. The 20 males students in the class averaged 70. The female students in the class averaged:
- a. 75 b. 80 c. 85 d. 90
20. Based on the following data values {7, 5, 6, 4, 7, 8, 12}, which of these statements is TRUE?
- a. mean = median = mode
b. mean = median < mode
c. mean = mode > median
d. median = mode < mean
21. Which of the following statements about the mean is not always correct?
- a. The sum of the deviations from the mean is zero.
b. Half of the observations are on either side of the mean.
c. The mean is a measure of the center of a distribution.
d. The value of the mean times the number of observations equals the sum of the observations.
22. In a histogram, if the distribution is left skewed, the proportion of the total area which must be to the right of the mean is
- a. less than 50%.
b. exactly 50%.
c. more than 50%.
d. Impossible to determine.
23. A sample of 20 observations has a standard deviation of 3. The sum of the squared deviations from the sample mean is
- a. 171 b. 60 c. 57 d. 20

The following are Descriptive Statistics output as calculated by Excel for two data series from the E370 survey completed for Fall Semester, 2000. The same number of students responded to each question. Refer to this output for the next **SEVEN** questions. Note: the questions continue onto the next page.

<i>Cumulative GPA</i>	
Mean	3.10
Median	3.10
Mode	3.00
Standard Deviation	0.51
Sample Variance	
Skewness	-1.00
Range	4
Minimum	0
Maximum	4
Sum	698.28
Count	

<i>Hours of Exercise per Week</i>	
Mean	6.11
Median	5.00
Mode	5.00
Standard Deviation	
Sample Variance	21.20
Skewness	1.41
Range	
Minimum	0
Maximum	25
Sum	
Count	

24. Of the two variables, *Cumulative GPA* and *Hours of Exercise per Week*, which has the larger relative dispersion?
- Cumulative GPA
 - Hours of Exercise
 - Their relative dispersion is virtually identical.
 - Insufficient data
25. The data set used represents the population of E370 students responding to the survey, however Excel always assumes data are from samples. How would the standard deviation of *Cumulative GPA* differ if Excel correctly assumed a population?
- The population standard deviation would be smaller than that reported.
 - The population standard deviation would be larger than that reported.
 - The population standard deviation would be the same as that reported.
 - The standard deviation depends on the specific data, so there is no way to tell.
26. What is the sample variance of *Cumulative GPA*?
- 0.26
 - 0.71
 - 2.04
 - The sample size is required to make this calculation.

27. Refer to the Skewness Coefficient for **Cumulative GPA** as calculated by Excel. Such coefficients typically range between -3 and +3; the closer the coefficient is to either extreme, the more skewed the distribution; a value of 0 indicates a symmetric distribution. Does this coefficient make sense? Why or why not?
- Yes, the mode is smaller than the mean and median, so the distribution must be negatively skewed.
 - Yes, the mean and median are the same, so the distribution must be left-skewed.
 - No, the distribution is almost perfectly symmetric and should have a skewness coefficient of 0.
 - No, the distribution is right-skewed and should have a coefficient of **positive** 1.
28. Refer to the **Cumulative GPA** output. Using the Empirical rule, what is the range over which approximately 68% of the observations lie?
- 2.59 to 3.61
 - 3.10 to 3.61
 - 1.02
 - 0.51
29. Refer to the **Hours of Exercise per Week** output. Which summary more closely conforms to the calculated statistics?
- A very symmetric distribution with relatively little dispersion implies that IU students are very similar in their taste for exercise.
 - At IU most students report getting 5 hours of exercise per week, and 50% of students report getting 5 hours or less per week.
 - Approximately 2/3rds of IU students report getting 5 to 10.5 hours of exercise per week.
 - b and c
30. Refer to the **Hours of Exercise per Week** output. You would like to create a frequency distribution for the data from which these statistics come. You have decided that you would like to have four classes, the first labeled "0 up to 5 hours". Which of the following is most likely?
- Your classes are mutually exclusive and collectively exhaustive.
 - Your classes are mutually exclusive but not collectively exhaustive.
 - Your classes are collectively exhaustive but not mutually exclusive.
 - Your classes are neither mutually exclusive or collectively exhaustive.

ANSWER KEY											
1	C	6	A	11	C	16	A	21	B	26	A
2	B	7	B	12	D	17	C	22	C	27	C
3	C	8	A	13	D	18	C	23	A	28	C
4	A	9	B	14	B	19	C	24	B	29	B
5	D	10	A	15	A	20	A	25	A	30	B