COB 191 Name_____

Questions 1 - 18 deal with the following Real Estate Scenario.

Stan Dalone is a real estate agent in Statville, and has recently joined a real estate firm that has 600 residential properties listed for sale in Statville. 40% of these properties are in the relatively wealthy Northtown district, 40% are in the moderately wealthy Midtown district, and 20% of the listings are in the relatively poor Southtown district. Curious about what the properties might be like, he decides to take a random sample of 60 of the 600 properties. His sample is made of 24 Northtown properties, 24 Midtown properties, and 12 Southtown properties. Within each district, properties are chosen by simple random sample from among all the properties listed with the real estate firm. He visits each of these 60 properties (and only these 60), and for each records the following information:

Number of bedrooms in the house (e.g., 3)
Primary construction material (brick, wood, etc.)
Asking price of the house (e.g., $125,500)
Condition of house (excellent, good, fair, poor)
Elevation of house foundation (compared to elevation of the ground at City Hall)
(e.g., -40.2 feet)

1. Which of these quantities, if any, should be considered at categorical variables?

    a) number of bedrooms
    b) construction materials
    c) asking price
    d) all of these
    e) none of these

2. Which of these quantities, if any, should best be considered as discrete numeric variables?

    a) number of bedrooms
    b) construction materials
    c) asking price
    d) all of these
    e) none of these

3. Which of these quantities could reasonably be considered to be measured at the interval, but not the ratio, level of measurement?

    a) number of bedrooms
    b) construction materials
    c) asking price
    d) condition of house
    e) elevation of house

4. The method of sampling employed by Stan is best termed

    a) simple random sampling
    b) stratified sampling
    c) systematic sampling
    d) cluster sampling
    e) nonprobabilistic sampling

5. Stan's sampling technique gives him a random sample drawn from the sampled population. What is the sampled population?

    a) all of the residential properties in Statville.
    b) all of the residential properties listed for sale in Statville.
    c) all of the residential properties listed for sale with Stan's firm in Statville.
    d) all of the residential properties that have already been sold in Statville.
    e) the 60 residential properties that Stan goes to visit.

6. *The fraction of the visited houses that are made of brick* is an example of a

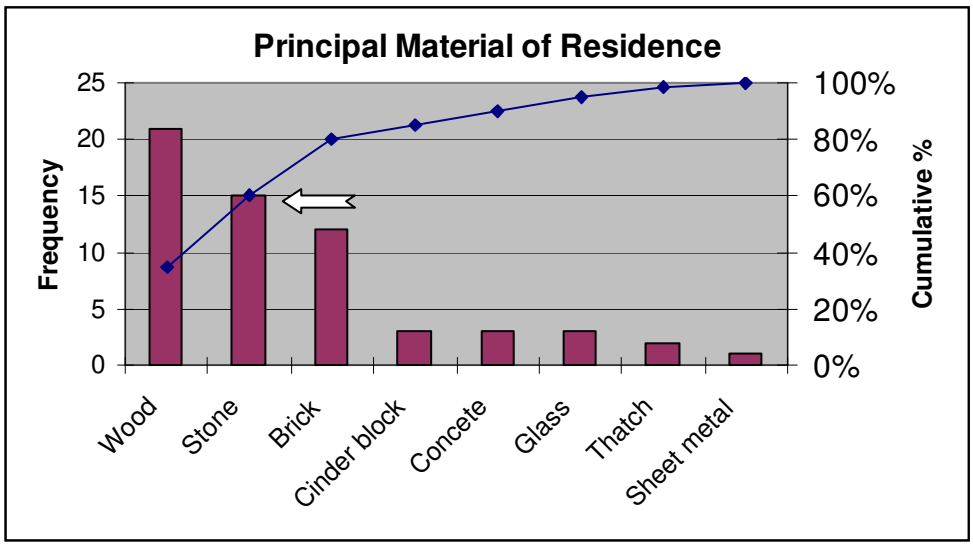
   a) factoid b) mean c) parameter d) range e) statistic

7. Using the results of Stan's survey to draw conclusions about all houses in the US would not be appropriate. No house outside of Statville could possibly be included in his sample. Using Stan's results to draw conclusions about all US houses would be an example of

   a) coverage (or frame) error.
   b) nonresponse error (or nonresponse bias).
   c) Sampling error.
   d) measurement error.
   e) overestimation.

8. 24 of the houses in the sample were ranked as being in "good" condition. What is the relative frequency of houses in the sample that were in "good" condition?

   a) 24 to 36 b) 24 to 60 c) 2/3 d) 0.4 e) 24



9. The Pareto diagram above summarizes Stan's findings at to construction materials. According to this diagram, what is the *relative frequency* of brick houses in the sample? (Ignore the white arrow for this question. It's used in Question 10.)

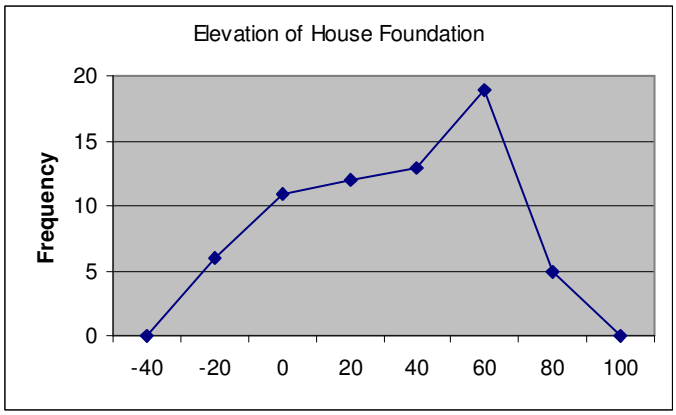
   a) 0.12 b) 0.2 c) 0.48 d) 0.8 e) 12

10. The white arrow in the graph above is pointing to the second point on the **line**. What information does this point convey?

    a) 15 houses in the sample are made of stone
    b) 15 houses in the sample are *not* made of stone
    c) 15 houses in the sample are made of stone or wood
    d) 60% of the houses in the sample are made of stone
    e) 60% of the houses in the sample are made of stone or wood

11. The asking prices in the sample of 60 houses ran from $55,000 to $255,000, with a median value of $120,000. Stan wishes to summarize the relative frequencies of the 60 asking prices in a histogram with six categories. What would be the most sensible class width for Stan to use?

a) 10 b) 3,333.33 c) 20,000 d) 35,000 e) 200,000



12. The frequency polygon above shows the distribution of house foundation elevations for the 60 houses included in Stan's sample. Focus on the third dot, the one directly over the number "0". What information does this dot convey?

a) 11 houses have an elevation of 0.
b) 11 houses have an elevation of 0 or less.
c) No houses have an elevation of 11 feet.
d) 11 houses have an elevation between -10 feet and +10 feet.
e) 5 houses have elevations between -20 feet and 0 feet.

Questions 13 and 15 use the frequency distribution below. The number of bedrooms in the 60 houses sampled are distributed as follows:

# of bedrooms	frequency
1	4
2	11
3	35
4	7
6	3

13. What is the **range** of the number of bedrooms in the sample?

a) 5 b) 32 c) 35 d)1 to 6 e) 3 to 35

14. What is the value of **Q1**, the first quartile, for this sample?

a) 2 b) 2.5 c) 3 d) 11 e) 35

15. The mean number of bedrooms in the sample is 2.95, and the median number of bedrooms is 3. If I wanted to compute the **variance** of this sample, I would have to perform a fairly complicated calculation. Which of the steps listed below would be a part of the computation of the variance of the sample?

a) subtracting 2.95 from 1, and squaring the result
b) dividing a sum by 59
c) taking a square root
d) you would do both a and b, but not c
e) you would do both a and c, but not b

16. Stan makes a histogram of the asking prices of the 60 houses, and discovers that they are strongly skewed to the right. What does this mean?

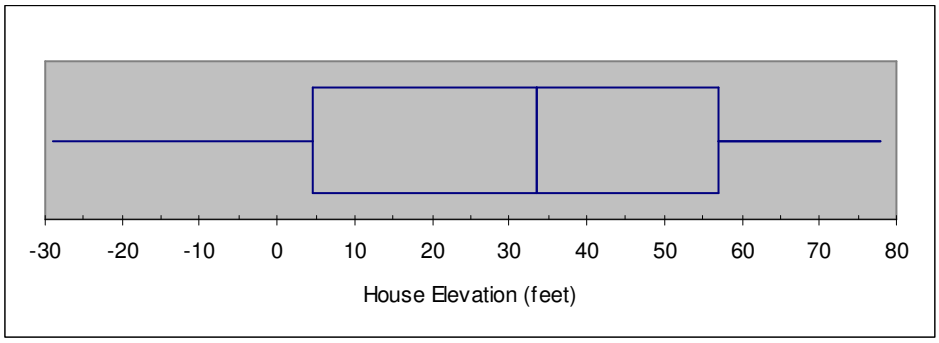
a) There are a lot more than 60 houses for sale by Stan's company in Statsville.
b) Among the 60 houses, many tend to clump around the lower prices, while relatively few have very high prices.
c) Among the 60 houses, many tend to clump around the median price, while the number falling much higher than the median is about the same as the number falling much lower than the median.
d) Among the 60 houses, many tend to clump around the higher prices, while relative few have very low prices.
e) Almost all of the houses have prices far away from the median price.

17. Stan made a scatterplot with *asking price* on the horizontal axis and *number of bedrooms* on the vertical axis. He then computed the correlation coefficient, *r*, for this scatter plot. His value of *r* is one of the five values below. Which value of *r* is most likely?

a) -1.5 b) -0.3 c) 0 d) +0.55 e) +0.99

18. Why can Stan not make a box and whisker plot of the data that he has collected on *house condition*?

a) The data is discrete.
b) The data is categorical.
c) The data constitutes a sample, not a population.
d) There are only four different possible conditions, and a box and whisker plot needs at least five values.
e) There is an even number of observations in the sample.



19. The box and whisker plot above shows the distribution of house elevations for the 60 houses in Stan's sample. The approximate interquartile range for the elevation of the houses in Stan's sample

a) is about 26
b) is about 33

c) is about 53
d) is about 57
e) is about 103

Questions 20 - 23 deal with the scenario below.

When assessing student performance on an 8 question assignment, the 8 questions were found to have the **levels of difficulty** shown below. Difficulty is measured on a scale from 0 to 25, with higher numbers representing more difficult problems.

Question	1	2	3	4	5	6	7	8
Difficulty	8	10	22	1	6	3	5	10

20. What is the mean difficulty level of problems on this assignment?

a) 7 b) 8 c) 8.125 d) 10 e) 10.5

21. What is the median difficulty level of problems on this assignment?

a) 7 b) 8 c) 8.125 d) 10 e) 10.5

22. What is the mode of difficulty levels of problems on this assignment?

a) 7 b) 8 c) 8.125 d) 10 e) doesn't exist

23. If the difficulty of question 3 were changed from 22 to 12, which measures of central tendency would change their values?

a) mean only b) median only c) mode only d) mean and mode only
e) median and mode only

24. StreetArt encourages citizens to send them $1 or $5 cash donations that will be used to support art programs for city children. Today StreetArt has received 100 contribution envelopes. To get an idea of how much money that they have received today, StreetArt opens up the first 16 envelopes. Four of them contain a $1 bill, and the remaining 12 contain a $5 bill. It's easy to determine, then, that the mean contribution in these 16 envelopes is $4. As part of StreetArt's analysis, they wish to compute the variance of the value of these 16 contributions.

The value of the variance for the collection of 16 envelopes is

a) 1/3 b) 3 c) 3.2 d) 10 e) 48

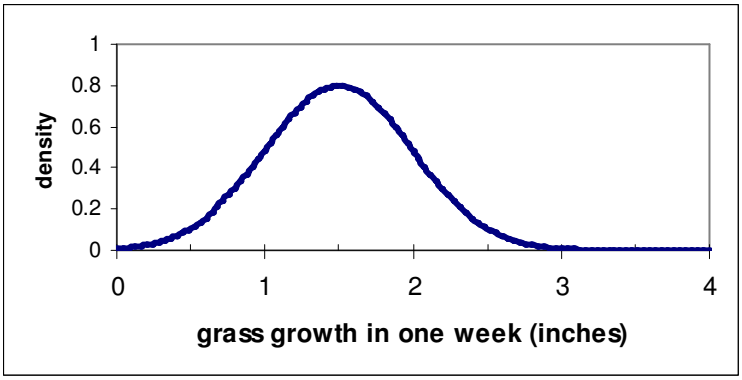
25. I am interested in studying the high daily temperature in Harrisonburg during the year 2003. I obtain records of the high temperatures for each of the 365 days in the year. I enter these numbers into Excel in cells A1 through A365. I now want to compute the mean daily high temperature for 2002 and the standard deviation in daily high temperature for 2003. Which Excel commands would give me the desired values?

a) =AVERAGE(A1:A365) and = STDEV(A1:A365)
b) =AVERAGE(A1:A365) and =STDEVP(A1:A365)

c) =MEAN(A1:A365) and = STDEV(A1:A365)
d) =MEAN(A1:A365) and = STDEVP(A1:A365)
e) =AVE(A1:A365) and =STANDARD(A1:A365)

Questions 26 – 28 deal with Grass Growing Scenario.

The number of inches that a blade of grass will grow during one particular week in summer is normally distributed, as shown in the graph below.



26. Based on this graph, 1.5 inches is

a) the mean number of inches grown
b) the median number of inches grown
c) the modal number of inches grown
d) all of these
e) none of these

27. Based on this graph, the standard deviation grass growth per week is about

a) 0.5” b) 1” c) 1.5” d) 3” e) 4”

28. During a different week in summer, the growth rates were still normally distributed and had the same standard deviation, but the mean growth during the week was 2”. How would this change the graph above?

a) the “hump” will become narrower
b) the “hump” will become wider
c) the “hump” will shift to the left
d) the “hump” will shift to the right
e) the “hump” will become taller

End of Growing Grass Scenario

29. The time required for customers to complete their lunch in a certain restaurant is essentially normally distributed (bell shaped) with a mean time of 35 minutes. About 95% of all customers finish their lunch in between 25 and 45 minutes. Knowing this, the standard deviation of the time required for a customer to complete his or her lunch is about

a) 5 minutes b) 10 minutes c) 15 minutes d) 20 minutes e) 25 minutes

30. Which of the following values is a possible value for a population variance? (As usual, ft means “feet”.)

a) -10 ft b) -10 ft^2 c) -10 ft^3 d) 10 ft e) 10 ft^2

31. The Mars Rovers *Spirit* and *Opportunity* are identical, sophisticated craft, but Mars is a hostile environment. Part of the logic of sending two craft was that, even if one failed, the other would probably continue to work. (Part was also to increase the area explored, since the two craft landed far

away from each other on the Martian surface.) Suppose that NASA computed that either craft (*Spirit* or *Opportunity*) has a 4% chance of failing. How likely is it that at least one of the craft will fail?

a) 0.0016 b) 0.0384 c) 0.0768 d) 0.0784 e) 0.0800

Questions 32-34 deal with Brenda's preference in soft drinks. For any brand of soda, let **S** = *the soda contains sugar* and **L** = *Brenda likes the soda*. Take your time on #32.

32. The claim that **S** and **L** are <u>mutually exclusive</u> is equivalent to the statement that

a) Brenda likes all sodas that contain sugar.
b) Brenda likes all sugar-free sodas.
c) Brenda dislikes all sugar-free sodas.
d) Brenda dislikes any sodas that contain sugar.
e) The presence or absence of sugar in a soda does not change the probability that Brenda likes the soda.

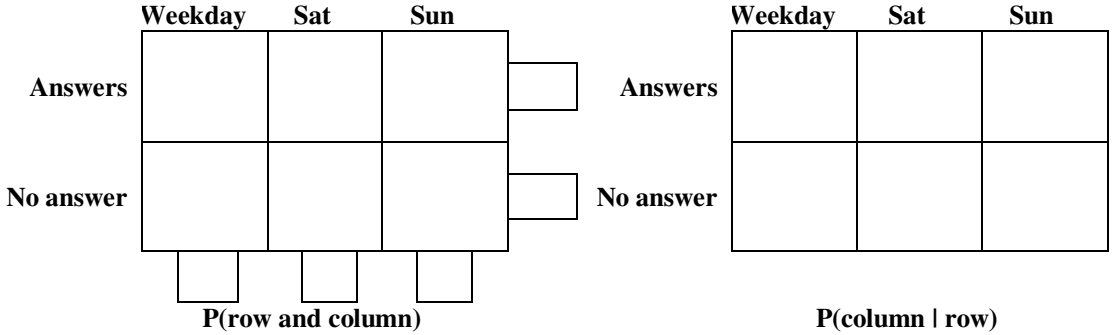
33. The claim that **S** and **L** are <u>independent</u> is equivalent to the statement that

a) Brenda likes all sodas that contain sugar.
b) Brenda likes all sugar-free sodas.
c) Brenda dislikes all sugar-free sodas.
d) Brenda dislikes any sodas that contain sugar.
e) The presence or absence of sugar in a soda does not change the probability that Brenda likes the soda.

34. The claim that **S** and **L** are a <u>completely exhaustive</u> set of events is equivalent to the statement that

a) Brenda likes all sodas that contain sugar.
b) Brenda likes all sugar-free sodas.
c) Brenda dislikes all sugar-free sodas.
d) Brenda dislikes any sodas that contain sugar.
e) The presence or absence of sugar in a soda does not change the probability that Brenda likes the soda.

Questions 35-40 deal with the scenario below. (You will probably wish to fill in part of the table below to hold the information provided or needed.)



Weekday	Sat	Sun

Answers			
No Answer			

P(row | column)

Phone calls were made to a large number of American households at 12 noon on various days of the week. 60% of all calls were made on weekdays (Monday through Friday) and 30% of the calls were made on Saturday. All of the Sunday calls were answered, and half of the weekday calls were answered. Additionally, half of all calls were answered.

35. What fraction of the phone calls was made on a Sunday?

a) 0.1 b) 0.2 c) 0.3 d) 1/3 e) 0.4

36. What fraction of the Saturday calls was answered?

a) 0.1 b) 0.2 c) 0.3 d) 1/3 e) 0.4

37. Suppose that a particular phone call went unanswered. What is the probability that the call was made on Saturday?

a) 0.1 b) 0.2 c) 0.3 d) 1/3 e) 0.4

38. Suppose that we select a phone call at random from all calls made. How likely is it that the call was made on Saturday and was answered?

a) 0.1 b) 0.15 c) 0.3 d) 1/3 e) 0.4

39. Which pair of events listed below are <u>mutually exclusive</u>, given the information in the problem?

a) *Calling on Sunday* and *getting an answer*
b) *Calling on Sunday* and *getting no answer*
c) *Calling on Saturday* and *getting an answer*
d) *Calling on Saturday* and *getting no answer*
e) *Calling on a weekday* and *getting no answer*

40. Which pair of events listed below are <u>independent</u>, given the information in the problem? (Hint: remember the relationships implied by independence.)

a) *Calling on Sunday* and *getting an answer*
b) *Calling on Sunday* and *getting no answer*
c) *Calling on Saturday* and *getting an answer*
d) *Calling on Saturday* and *calling on Sunday*
e) *Calling on a weekday* and *getting no answer*