1. Texas A&M wants to conduct a survey to what students think about a change in student fees. They randomly select a UIN and then sample every 250th UIN after that until they reach 200 students. This is a ___ sample.
   a) Cluster
   b) Stratified
   c) Systematic
   d) Simple Random
   e) Convenience

2. The Political Science Department wants to know if students like the format of POLS 206. They want to know specifically how Political Science majors feel about the course. They plan on sampling 200 students. A university census shows 2% of undergraduates are Political Science majors. What sampling method is the best for their needs and why?
   a) Convenience sample of students in the Allen building, because this sample would be easy and provide an unbiased sample.
   b) Simple Random, because every student would have an equal chance of being selected.
   c) Cluster of a single class of 100, because it will ensure that political science students are sampled.
   d) Stratified by major, because it will ensure that political science students are sampled.

   [Use the following information to answer the next 5 questions.]
   Consider the following scenario: A marriage counselor is interested in the proportion of clients she counsels who stay married. For the next five questions.

3. What is the population of interest?
   a) a group of clients of this marriage counselor
   b) all the clients of this counselor
   c) the proportion of all her clients who stay married
   d) the number of couples who stay married

4. What would the sample be?
   a) a group of clients of this marriage counselor
   b) all the clients of this counselor
   c) the proportion of all her clients who stay married
   d) the number of couples who stay married
5. The parameter of interest is?
   a) a group of clients of this marriage counselor
   b) the proportion of the sample of her clients who stay married
   c) the proportion of all her clients who stay married
   d) the number of couples who stay married

6. The statistic of interest is?
   a) a group of clients of this marriage counselor
   b) the proportion of the sample of her clients who stay married
   c) the proportion of all her clients who stay married
   d) the number of couples who stay married

7. The variable of interest is?
   a) a group of clients of this marriage counselor
   b) all the clients of this counselor
   c) the proportion of all her clients who stay married
   d) the number of couples who stay married

---

[Use the following information to answer the next 2 questions.]
A realtor is interested in what factors affect prices of various houses. A realtor looks at many houses that have been put on sale, views the price of the offered house, their square footage, the number of rooms in the house, and the number of bathrooms in the house.

8. Which of the following best describes this situation?
   a) Its an Experiment, we are examining the effect of different explanatory variables (square footage, number of rooms, and number of bathrooms) on housing (a response variable), which only happens in experiments.
   b) Its an Observational Study, the realtor is looking at past transactions and not directly interacting with any of these houses or transactions, so it must be an Observational Study.
   c) Its an Experiment, the realtor is probably doing this with the intention of establishing some causation (she can price houses differently based on their traits), which can only be done in experiments, thus, this is an experiment.
   d) Its an Observational Study, the realtor is trying her hardest to price houses, but this isn’t an easy task. This is an observational study due to the difficulty of the task she’s facing.

9. After looking at the data she gathered, determines that the price of the house increases because of increases in the number of rooms. Is this a justifiable conclusion?
   a) No, experimental studies can never be used as evidence of causation
   b) Yes, experimental studies can always be used as evidence of causation
   c) Possibly, experimental studies can be used as evidence of causation when the data is properly gathered.
   d) No, Making this conclusion is unethical as it will lead to higher prices for some houses.
10. A group of scientists are interested in comparing the effects of caffeine on learning to a group of school children. They give each student in one class a cup of caffeinated coffee, and each student in another class a placebo (decaffeinated coffee). The two classes grades on an assignment are then reported. The grades on the assignment come out to be much larger for the caffeinated group, and they conclude that caffeine is responsible for the grade difference. Comment on the validity of this experiment.

a) The experiment is valid, the scientists have made sure the only variable being changed is the caffeine content, and the change in grade must be because of this.

b) The experiment is valid, none of the students know what theyre taking, so they cant manipulate the results to their advantage, and thus the results cant be falsified.

c) The experiment is valid, numbers dont lie.

d) The experiment is not valid, the groups are not randomly assigned, lurking variables can be present.

11. Which of the following are Quantitative Data?

I. Phone Numbers
II. Amount of Money in a person’s wallet
III. Number of Credit Hours

a) I only
b) II only
c) III only
d) II & III
e) I, II, & III

12. Which of the following cases is a Stem Plot not a good choice of graph?

a) An elementary school teacher wants to visualize the data of how her 20 students did on a test.
b) A professor wants to visualize the data of the scores of her 200 students on a test.
c) A researcher wants to visualize the results of the growth of her 13 plants.
d) A golf player wants to visualize his past 10 scores.

13. What is the difference between a bar graph and a histogram?

a) A bar graph is for qualitative data, and a histogram is for quantitative data.
b) A bar graph is for quantitative data, and a histogram is for qualitative data.
c) Bar graphs and histograms are the same thing.
d) Histograms can only be used for continuous quantitative data, but bar graphs are for discrete quantitative data.
In a survey of 20-year-olds in China, Germany, and the United States, people were asked the number of foreign countries they had visited in their lifetime. The box plot below displays the results.

14. Have more Americans or more Germans surveyed been to over eight foreign countries? Why?
   a) More Americans, the distribution is skewed to the right meaning that most of them have traveled to more than 8 countries
   b) More Germans, the median is located at 8 meaning that 50% of sample has been to at least 8 countries
   c) More Americans, the distribution is skewed to the left meaning that most of them have traveled to more than 8 countries
   d) We cannot answer the question using this data as the maximum for both groups are the same

15. A study was conducted to determine approximately how long it takes for skin to burn when standing in direct sunlight without sunscreen. Four people burned in 5-9 minutes, twelve people burned in 10-14 minutes, eighteen people burned in 15-19 minutes, thirteen burned in 21-24 minutes, and seven burned in 25-29 minutes. Using this information, what percentage of people got a sunburn between 10 and 19 minutes?
   a) 55.56
   b) 31.28
   c) 62.96
   d) 52.43

16. Which of the following is true about relative frequency?
   a) In the set of all outcomes, it is the number of times a data value occurs.
   b) Summing over the entire column of relative frequencies always equals the sample size.
   c) The last entry of the cumulative relative frequency column (not the total) is 1, indicating that one hundred percent of the data has been accumulated.
   d) It is the proportion of times an observation occurs.

17. Suppose a group of college students was asked how many cups of coffee they drink per week. The data are the following: 1, 1, 2, 3, 4, 4, 4, 5, 7, 8, 9. Based on the data, what is the 37th percentile of the number of cups of coffee consumed per week?
   a) 4.44
   b) 4.5
   c) 3
   d) 3.5
18. Assume that the 25th percentile of a set of data is 13.5 and the 75th percentile is 18.2. Which data value is considered an outlier?
   a) 24.78  
   b) 6.54  
   c) 6.42  
   d) 20.17  

19. A twelve year-old girl who is 411 tall is in the 72nd percentile of height for her age group, according to her physician. Which of the following is the incorrect interpretation of this?
   a) The girl is taller than 72% of other girls her age.  
   b) The girl is 72% taller than other twelve-year old girls.  
   c) If we sort out all twelve year-old girls who have a height less than 411, they will comprise approximately 72% of the population.  
   d) The girl is shorter than 28% of twelve year-old girls.  

   [Use the following information to answer the next 2 questions.]

Consider the distribution of the following histograms

20. How would you describe the distribution of the histograms above?
   a) Symmetric  
   b) Skewed Left  
   c) Skewed right  
   d) Trimodal
21. Which of the following explains the two clear modes in both of the histogram above?  
   a) For both of the histograms, one of the modes corresponds to infant deaths, the other corre-
      sponds to death by diseases at old age.
   b) The typical death age of male and females are very different. Both of the modes for males
      and females are different by a very large amount as a result.
   c) There is no interpretation to these modes, any narrative can be told and spun to fit the mode.
   d) I don’t see two modes in the above histograms. This question is loaded.

22. Based on the histogram above, which number do you feel best describes the location for the
      age at which the typical person dies?  
   a) The mean  
   b) The median  
   c) The standard deviation  
   d) The IQR

   [Use the following information to answer the next 2 questions.]

23. Compute the mean.
   a) 1  
   b) 4  
   c) 2.67  
   d) 3

24. Compute the standard deviation.
   a) 1  
   b) 1.53  
   c) 4  
   d) 1.25

25. Find the median of the data set: 31, 37, 37, 37, 38, 40, 40 , 41, 41.
   a) 38  
   b) 37.5  
   c) 37  
   d) 39

26. Find the mean of 53, 64, 67, 68, 69.
   a) 67  
   b) 64.2  
   c) 63  
   d) 66

27. What is the standard deviation of 0,0,0,-5,5? This is from a sample.
   a) 3.54  
   b) 3.16  
   c) 5  
   d) 4.08
28. What can be concluded from the boxplot right?

![Boxplot Image]

a) About 50% of the data values are larger than 88.
b) About 50% of the data values are either larger than 94 or smaller than 80.
c) About 50% of the data values are between 80 and 94
d) Two of the above are true

e) All of the above are true.

29. A study was undertaken to determine if atmospheric pressure has an impact on plant heights. To study this, 20 Spartina alterniflora plants were planted at sea level while another 20 Spartina alterniflora plants were grown in a special chamber aboard the International Space Station, and the heights of the plants were recorded. The goal was to see if the plant heights were affected by the atmospheric pressure exerted on the plant. What are the variables, and what kind are they?

a) Explanatory: Location (categorical) Response: Height of plant (numerical)
b) Explanatory: Location (categorical) Response: Height of plant (categorical)
c) Explanatory: Height of plant (numerical) Response: Location (categorical)
d) Explanatory: Species of plant (categorical) Response: Height of plant (numerical)
e) Explanatory: Species of plant (categorical) Response: Location (categorical)

30. Suppose you have the following 2 data sets.

DATA SET A takes the values: 1, 2, 3, 3, 5, 7, 8, 9, 17, 17, 22, 25
DATA SET B takes the values: 6, 7, 8, 8, 10, 12, 13, 14, 22, 22, 27, 30

Which of the following will be the same for both DATA SET A and DATA SET B?

a) IQR and range will be the same
b) Median and IQR will be the same
c) Only the median is the same
d) Only the IQR is the same
e) None of the center and variability statistics are the same
31. In the following figure, the relationship between x and y can be described as

- a) Strong linear relationship
- b) Weak linear relationship
- c) Weak relationship
- d) No relationship
- e) None of the above are a good description of the relationship

![Relationship Graph](image)

32. A study found that there is a large positive correlation between self-esteem scores and performance in high school students. Which of the following can you conclude? Assume there are no outliers and the data indicates a linear relationship exists between self-esteem scores and performance.

- a) Raising self-esteem improves performance in schools
- b) Better performance in schools causes higher self-esteem
- c) Raising self-esteem causes poorer performance in schools
- d) Lurking variables such as parental support may affect both self-esteem and performance in school.
- e) Two of the above are equally good answers

33. Which one of the following statements is FALSE?

- a) The only way the standard deviation can be 0 is when all the observations have the same value.
- b) If you interchange the explanatory variable and the response variable, the correlation coefficient remains the same.
- c) If the correlation coefficient between two variables is 0, that means that there is no possible relationship between the two variables.
- d) The correlation coefficient has no units
- e) None of the above
34. Multiple drugs are often used in treating high blood pressure. Suppose 10% of people taking blood pressure medicine A experience negative side effects and 20% of people taking medicine B experience negative side effects. If the side effects of the two drugs occur independently, what is the probability a person taking both drugs experiences at least one side effect?
   a) 0.28  
   b) 0.72  
   c) 0.30  
   d) 0.02  
   e) 0.98

35. A recent article in an educational research journal reports a positive correlation of +0.75 between math PSAT score and math aptitude. It also reported a negative correlation of −0.75 between math PSAT score and score on a math anxiety test. Which of the following interpretations is most appropriate?
   a) Both relationships are equally strong  
   b) A correlation of +0.75 indicates a stronger relationship than a correlation of −0.75.  
   c) A correlation of −0.75 indicates a stronger relationship than a correlation of +0.75.  
   d) People with high math anxiety have low math aptitude and people with low math anxiety have high math aptitude.  
   e) We cant tell from the above information which relationship is stronger.

36. Stem cell transplants therapy for adults with non-Hodgkin lymphoma is a very new form of treatment for the disease. A study of 40 patients with non-Hodgkin lymphoma who underwent stem cell transplants was conducted to see if there was a difference in outcome between those non-Hodgkin patients who underwent stem cell transplants with their own cells compared to those non-Hodgkin patients who underwent stem cell transplants using someone elses stem cells. The study found that patients who used their own stem cells had fewer side effects. Identify the population of interest in this study.
   a) All people with cancer  
   b) All people with lymphoma  
   c) The 40 patients with non-Hodgkin lymphoma  
   d) All people with non-Hodgkin lymphoma who undergo stem cell transplants  
   e) All people with non-Hodgkin lymphoma who undergo stem cell transplants using their own stem cells.

37. A researcher is reporting characteristic of the subjects she used in a recent study. Two of the variables are hair color and age. Which of these are appropriate choices to summarize these data?
   a) Bar charts for both hair color and age.  
   b) Histograms for both hair color and age.  
   c) A bar chart for hair color and a histogram for age.  
   d) A histogram for hair color and a bar chart for age.  
   e) Either bar chart or histograms are good choices for both hair color and age.