Review for 11th Grade Statistics

1. Finding the favorite flavors of ice cream of students at MMSTC represents the collection of what type of data? Quantitative or Qualitative?_____Qualitative_____

2. The average age of the teachers here represents: Quantitative or Qualitative? Discrete or Continuous   Quantitative and Discrete

_____

3. The average length of hair for the boys at MMSTC: Quantitative or Qualitative?

Discrete or Contiuous?_____Quantitative and Continuos_____

4. Careful watching and taking notes on the number of trees Mr. McMillan runs into as he skies down the slopes is an example _Observational Study_____.

5. Blind-folding Mr. McMillan and then noting the number of trees he runs into as he skies down the slopes is an example of _____Experiment_____

6. When every unit or set of units has an equal chance of being picked, this is called __Simple Random Sample, SRS_____ why is it important?_this helps reduce bias by creating homogeneous groups_____

7. Name four types of sampling and briefly describe each one. (chapter 5)

1. Simple Random Sample (SRS), every individual or set of individuals has an equal chance of being picked.

2. Stratified Random Sample – separates data into strata or categories ten takes a SRS.

3. Multistage Random Sample – chooses by taking SRS from successively smaller categories. (Start with a broad category, Example: Choose 100 school districts across the country randomly, then from those 100 districts randomly choose one high school, from within each of the randomly chosen high schools do an SRS to pick 10 students from each high school.

4. Systematic Random Sampling – Divide sample into equal parts, number each subject successively, then do an SRS in the first group to select an individual then picks next individual from second group by adding same that was picked in first group to the second group and so on and so on.

8. Variables that cannot be controlled are called lurking variables

9. What is the word used to describe an experiment where you can't tell if the results are truly from the treatment or from outside factors. Confounding_____

10. What can we do to help reduce the effects of lurking variables? Conduct a Randomized Comparative Experiment : CRR, Control – lurking variables affect all groups equally, Randomize – creates homogeneous groups to reduce bias, Replicate – reduces variability.

11. If a subject does not know whether he/she is getting the treatment or a placebo the experiment is said to be __blind_____

12. If the subject nor the experimenter knows if the subject is getting the treatment or a placebo the experiment is called__double blind___

13. Define distribution._____the overall pattern that data takes_____
_____

14. A distribution is symmetric if___it is the same (mirror image) on both sides of the median/mean line_____

15. If a distribution is skewed to the right, where is the tail? __to the right____

16. In a right skewed distribution does more data lie to right or the left of the median? Same amount of data on both sides the mean? More data on the left. Mean is pulled by the tail (on the right) where data is more spread out.

17. Q1 - 1.5(IQR) and 1.5(IQR) + Q3 is a method to do what? Finds outliers in any distribution.

18. Points that fall beyond 3 standard deviations on a symmetric distribution are called_____outliers_____

19. Calculate the mean and five number summary of the following list of data

82, 55, 70, 76, 91, 100, 88, 79, 90, 44, 79, 59, 84, 32, 84, 97, 91, 81
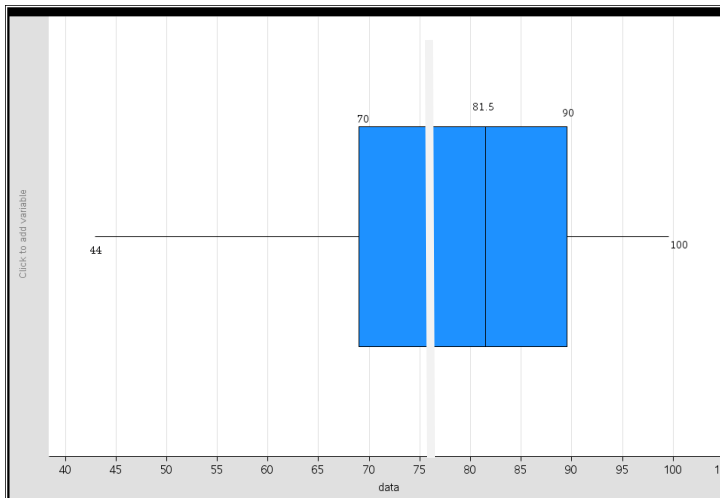
Mean 76.7

20. Circle the statistics that are resistant measures of center and spread: mean, median, standard deviation, inter-quartile range.

median and interquartile range

21. When is the mean equal to the median?_symmetric distribution_____

22. Measure the spread of the data in #19 using the range _68 = 100-32

23. Calculate the standard deviation of the data in #19 write the equation out but use a calc. Define standard deviation $s = \sqrt{\dfrac{\Sigma(x-\bar{x})^2}{n-1}}$ = 18.36

Loosely defined as he average distance that data lie from the mean

24. Define density curve. The area underneath it is always = to 1 whole or 100%. A density curve approximates the shape of the data with a curve instead of a histogram of the data.

25. A single-peaked, bell-shaped density curve can also be called a normal curve.

26. _____68_____% of the data lies within one standard deviation of the mean, _____95_____% lies within two standard deviations of the mean and _____997.___% lies within three standard deviations of the mean of a normal distribution.

27. What is true about the mean of the standard normal distribution? μ = 0

28. Define z-score The number of standard deviations above or below the mean, μ, on a standard normal distribution.

29. Define t-value The number of standard deviations above or below the mean, μ, on a *t* distribution.

30. A __*t*___distribution is single-peaked but is flatter and has thicker tails than the Standard normal distribution

31. A value used to describe a sample is called a statistic.

32. A value used to describe a population is called parameter.

33. The Sampling Distribution of a statistic is the distribution of values taken by the statistic in all possible samples of the same size from the same population. The distribution of x bars or p hats.

34. Define standard error. The standard deviation of the sampling distribution, $s/\sqrt{n}$.

35. The theorem that says that the mean of the sampling distribution is the same as the mean of the population but the standard deviation of the sampling distribution is much narrower. Central Limit Theorem

36. A selected level that deems if results are significant or not. If no level is selected we default to using 0.05. This is called the alpha level.

37. What test would you use to compare a sample mean to a known value and you don't know the population standard deviation. One-sample t Test.

38. What test would be appropriate to see if there is a significant difference in the proportion of 10th graders that love statistics class and the proportion of 11th graders that love statistics class. Two-proportion z Test

39. Confidence intervals are used to find a range that we can be confident will contain the population mean, μ (be general).

40. The probability of getting the statistic (or difference in statistics) you did by chance alone if the null hypothesis is true, is the meaning of the

p-value.

41. How are degrees of freedom estimated for a:

   a. one-sample *t* test? n-1

   b. two-sample *t* test? n-1 of smallest sample

42. Name 5 types of bias and describe each.

   1. Response bias – lie on a survey

   2. Voluntary response – people volunteer their response.

   3.Undercoverage – part of the population is missed

   4. Non-response – people refuse to take survey, hang up on you, close

      the door etc.

   5. Wording of the Question –Question is worded in such a way that elicits

      a certain response.

   6. Convenience Sampling –Only ask people easiest to get to (often they

      have the same view you do

43. What is Blocking? When and Why is it used? An experimental method
    that separates subjects/units into groups with similar characteristics
    because it is believed that this characteristic may have an effect on the
    outcome. So as a form of control you separate them so that you can look
    at one group at a time and if there is a difference in the outcome it can be
    attributed to the treatment and not to the factor (characteristic) by which
    the were separated.

44. When do you use a Matched Pairs *t* Test? How many populations is

    it done on? Use a Matched Pairs t Test when you have one population
and you are comparing the difference of before and after.

45. Are you glad to be back doing statistics? Please circle just one.

   a. Yes   b. YES!   c. Absolutely Yes!

      All of the above!