Math 140 Final Review Sheet

1. Vocabulary: Give the term described by each of the following definitions.

_____ a) The collection of all people or objects to be studied.

_____ b) When a sample does not represent the population.

_____ c) When every person or object in a population has a chance to be included in the sample data.

_____ d) A number calculated from sample data in order to better understand the characteristics of the sample.

_____ e) Collecting data from a sub-group or small part of the population.

_____ f) Selecting individual people or objects randomly.

_____ g) Collecting data from every person or object in the population.

_____ h) Selecting groups of people to collect data from.

_____ i) Comparing two or more groups.

_____ j) Collecting data in a way that is easiest for you.

_____ k) Putting a survey out into the world and allowing anyone to fill it out.

_____ l) Taking data from every 50^{th} person on a list.

_____ m) Not collecting the data with randomization or with a census.

_____ n) Phrasing questions in order to persuade people to give a certain answer.

_____ o) Not representing groups from your population or falsifying data.

_____ p) When people will not answer truthfully about a topic.

_____ q) When people selected refuse to take part in the study.

_____ r) Just collecting and analyzing data without trying to control confounding variables.

_____ s) When only the person receiving the treatment does not know if it is real or a placebo.

_____ t) The confounding variable of the human brains capacity to alter physical effects just because the person believes something is true.

_____ u) When both the person receiving the treatment and the person giving the treatment does not know if it is real or a placebo.

_____ v) A group that we can use to compare our treatment group to in order to determine if there was a significant effect.

_____ w) A fake medicine or fake treatment.

_____ x) Scientific process that compares two or more similar groups in order to prove cause and effect.

_____ y) Other variables that might influence our response other than the explanatory variable. These need to be controlled in order to prove cause and effect.

_____ z) Randomly putting people or objects into two groups.

_____ aa) The most accurate measure of typical spread (typical variability) for a skewed data set.

_____ ab) The best average (center) for skewed data sets.

_____ ac) Typical values in a normal data set are between which two statistics.

_____ ad) The best average (center) for a normal (bell shaped) data set.

_____ ae) Typical values in a skewed data set are between which two statistics.

_____ af) The most accurate measure of typical spread (typical variability) for a normal (bell shaped) data set.

_____ ag) Typical distance from the mean.

_____ ah) The center or average when the data values are put in order.

_____ ai) A marker that about 75% of the data values are less than.

_____ aj) The measure of center or average that balances the distances.

_____ ak) A measure of typical spread that measures the difference between the 1^{st} and 3^{rd} quartiles.

_____ al) The smallest number in a quantitative data set.

_____ am) The number that appears most often in a data set.

_____ an) The square of the standard deviation.

_____ ao) The largest number in a quantitative data set.

_____ ap) A marker that about 25% of data values are less than.

_____ aq) A measure of spread that measures the distance between the largest and smallest numbers.

_____ ar) Random sample values will usually come out different and will usually be different than the population value.

_____ as) Collecting lots of random samples from one original random sample with replacement. Then calculating a statistic from each random sample and putting all the statistics on the same graph.

_____ at) Two numbers we think a population value will be in between.

_____ au) When someone takes a sample statistic and tells people it is the population parameter.

_____ av) The standard deviation of a sampling distribution.

_____ aw) 95% of confidence intervals contain the population value and 5% of confidence intervals do not contain the population value.

_____ ax) Collecting lots of random samples from a population. Then calculating a statistic from each random sample and putting all the statistics on the same graph.

_____ ay) 90% of confidence intervals contain the population value and 10% of confidence intervals do not contain the population value.

_____ az) How far off a sample statistic could be from the population parameter.

_____ ba) 99% of confidence intervals contain the population value and 1% of confidence intervals do not contain the population value.

_____ bb) Collecting lots of random samples from one original random sample with replacement.

_____ bc) The probability of getting the sample data or more extreme because of sampling variability if the null hypothesis was true.

_____ bd) A number you compare your P-value to in order to tell if your P-value is low or high. This is also the probability of making a type 1 error.

_____ be) A statement about whether you reject of support with the claim and whether or not you have evidence.

_____ bf) When sample data indicates that we should fail to reject the null hypothesis, but this is not correct in the population.

_____ bg) The probability of making a type 2 error.

_____ bh) A statement about the population that involves equality.

_____ bi) The number of standard errors that a sample mean $(\bar{x})$ is above or below the population mean (μ) in the null hypothesis.

_____ bj) A statement about the population that does <u>not</u> involve equality.

_____ bk) The number of standard errors that a sample proportion $(\hat{p})$ is above or below the population proportion (π) in the null hypothesis.

_____ bl) When sample data indicates that we should support the alternative hypothesis, but this is not correct in the population.

_____ bm) A number to compare your test statistic to in order to determine if the sample data significantly disagrees with the null hypothesis.

_____ bn) The number of standard errors that the sample proportion for group 1 is above or below the sample proportion for group 2.

_____ bo) The sum of the average of the squares of the differences between the observed sample counts and the expected counts from the null hypothesis.

_____ bp) The number of standard errors that the sample mean for group 1 is above or below the sample mean for group 2.

_____ bq) The ratio of the variance between the groups to the variance within the groups.

_____ br) The number of standard errors that the sample slope is above or below zero.

Chapter 1 Questions

2. For each of the following data collecting methods, determine if the method has a significant amount of sampling bias or minimal amount of sampling bias. ("Significant" or "Minimal")

_____ a) Stratified Sample (data was random)

_____ b) Voluntary Response Sample

_____ c) Convenience Sample

_____ d) Simple Random Sample

_____ e) Cluster Sample (data was not random)

3. Explain how experimental design is used to show cause and effect.

4. Classify the following variables as Categorical or Quantitative.

_____ a) The various types of wood carried at Home Depot.

_____ b) The number of houses flying the American flag in various cities on the fourth of July.

_____ c) The temperature in degrees Celsius in Palmdale, CA over the last six months.

_____ d) People's favorite type of beer.

_____ e) The cost of various sweatshirts in dollars at multiple stores.

5. Convert each of the following decimal proportions into a percentage. After converting, round your percentages to the tenths place.

a) 0.024872 = _____

b) 0.00583 = _____

c) 0.854226 = _____

6. Convert each of the following percentages into a decimal proportion. Do <u>not</u> round the answer.

a) 6.3% = _____

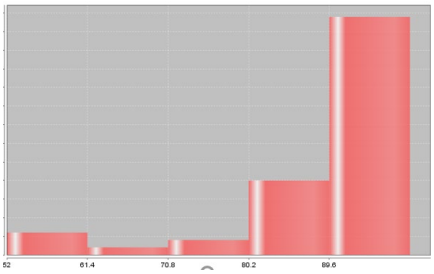
b) 0.00041% = _____

c) 52.7% = _____

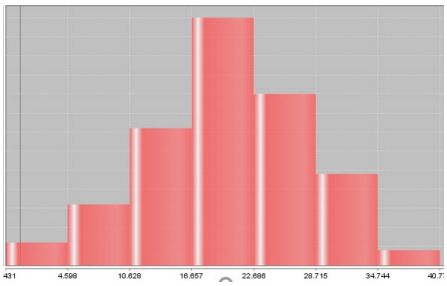
7. Describe the process of analyzing a categorical data set. Include directions on how to find the proportions, percentages and percent of increase.

8. Give the shape of each of the following graphs.

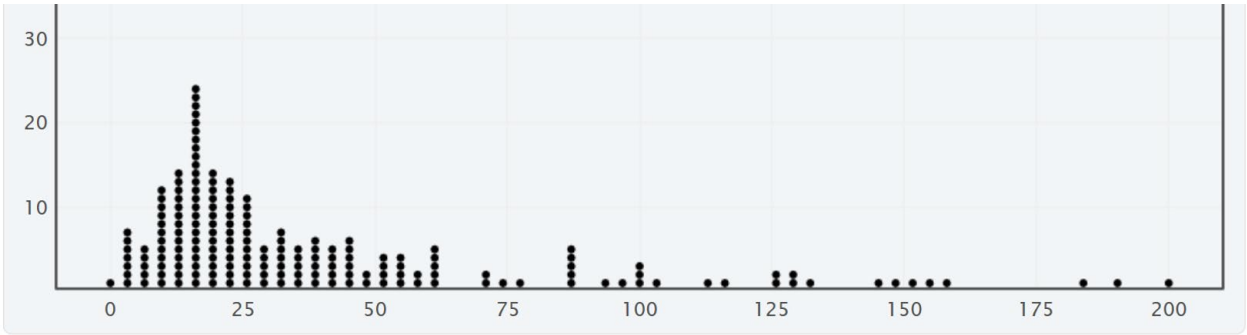
a)



b)



c)



9. Describe how to analyze a normal quantitative data set. Include the shape, center, average, spread, typical values and outliers. What percent of normal quantitative data is considered typical? What percent of normal quantitative data is considered unusually high and unusually low?

10. Describe how to analyze a skewed quantitative data set. Include the shape, center, average, spread, typical values and outliers. What percent of skewed quantitative data is considered typical?

11. Describe the process of finding the mean, variance and standard deviation of a quantitative data set. When are these statistics accurate?

12. Describe the process of finding the 1^{st} quartile, median, 3^{rd} quartile and interquartile range (IQR). When are these statistics accurate?

<u>Chapter 2 Questions</u>

13. What is the difference between a statistic and a parameter? Fill out the following chart to identify the following letters used in statistics.

Letter	Statistic or Parameter?	Standard Deviation, Mean, Proportion,or Frequency?
$\hat{p}$		
N		
μ		
s		
n		
σ		
π		
$\bar{x}$		

14. Write down the three assumptions for a one-population mean confidence interval.

15. Write down the three assumptions for a two-population mean confidence interval from separate groups.

16. Write down the three assumptions for a two-population mean confidence interval from matched pairs.

17. Write down the four assumptions for a one-population proportion confidence interval.

18. Write down the four assumptions for a two-population proportion confidence interval.

19. Describe how a sampling distribution is created and what it shows us about the principle of sampling variability. How can we use the sampling distribution to estimate a population parameter and calculate standard error. Why is it important that sampling distributions for sample means or sample proportions be normal (bell shaped)?

20. State and explain the Central Limit Theorem. What does it tell us about sampling distributions for sample means and sample proportions?

21. What is a confidence interval? How can we use the sample mean, the T score critical value, and the standard error to calculate a confidence interval estimate of the population mean? How can we use the sample proportion, the Z score critical value, and the standard error to calculate a confidence interval estimate of the population proportion? Why is it important that the sample data meet the assumptions for the confidence interval?

22. Explain the process of bootstrapping. How can we use bootstrapping to find the upper and lower limit of the confidence interval without a formula?

23. Explain how two population confidence intervals work. How can we use them to determine if the population mean or proportion from group 1 is significantly higher or lower than the population mean or proportion from group 2? How will we know if they are not significantly different?

24. What happens to the critical value, standard error, margin of error and confidence interval when the sample size increases? What about when the sample size decreases?

25. What happens to the critical value, standard error, margin of error and confidence interval when the confidence level increases? What about when the confidence level decreases?

26. How are Z-scores and T-scores similar? How does a Z-score critical value compare to a T-score critical value when the sample size is very small? What about when the sample size is very large? When should we use a Z-score critical value and when should we use a T-score critical value?

27. What does the margin of error tell us? How is the margin of error used to create a confidence interval? What is the standard sentence we use to explain a one-population confidence interval? Give an example. What is the standard sentence we use to explain a two-population confidence interval? Give an example.

Chapters 3&4 Questions

28. Describe the process for writing the null and alternative hypothesis. How do we know which one is the claim? How do we know if it is a right-tailed, left-tailed or two-tailed test?

29. For each of the following tests, give the type of data, an example null and alternative hypothesis, the test statistic used, and the traditional assumptions required.

a) One-population proportion hypothesis test.

b) One-population mean hypothesis test.

c) Two-population proportion hypothesis test.

d) Two-population mean hypothesis test when the data is in separate groups.

e) Two-population mean hypothesis test when the data is matched pair.

f) Goodness of Fit test.

g) ANOVA test.

h) Categorical Association test.

i) Correlation test.

30. When examining relationships between variables how can we know if we should use ANOVA, Correlation or the Categorical Association Test?

31. How can we use the test statistic and the critical value to determine if the sample data significantly disagrees with the null hypothesis or not.

32. Fill out the following table for interpreting test statistics, tails and significance.

Type	*Test Statistic*	*Critical Value*	*Does the Test Statistic fall in a tail determined by the critical value or not?*	*Does the sample data significantly disagree with H_0 or not?*	
Right-tailed	χ^2 = +6.327	*+9.833*			Are the observed counts significantly different than the expected counts?
Two-tailed	Z = +2.46	*± 1.96*			Is the sample proportion for group 1 significantly higher than group 2?
Left-tailed	*T = −1.377*	*−2.643*			Is the sample mean for group 1 significantly lower than group 2?
Right-tailed	*F = +4.612*	*+3.182*			Is the variance between significantly higher than the variance within?

33. State the definition of P-value. How do we use the P-value and significance level to determine if the sample data could have occurred by sampling variability? How do we use the P-value and significance level to determine if the sample data significantly disagrees with the null hypothesis or not? How do we use the P-value and significance level to determine if we should reject the null hypothesis or fail to reject the null hypothesis?

34. Fill out the following table for interpreting P-value, significance levels, sampling variability, the null hypothesis, and evidence.

P-value	*P-value %*	*Significance Level (α)*	*Significance Level %*	*Is the P-value lower or higher than the sig level?*	*If the null hypothesis was true, could this sample data occur by sampling variability or is it unlikely?*	*Reject H_0 or Fail to reject H_0?*	*Significant Evidence or not?*
0.00248		0.01					
3.4×10^{-5}		0.10					
0.447		0.05					

35. Give the standard sentence for writing a conclusion to a hypothesis test. Explain the various options for a conclusion and how to know which to use.

36. Write the correct conclusion for each of the following addressing evidence and the claim.

a) High P-value and the Claim was H_0
b) Low P-value and the Claim was H_0
c) High P-value and the Claim was H_A
d) Low P-value and the Claim was H_A

37. Explain how to use randomized simulation (randomization) to perform a hypothesis test. How do we use the simulation to calculate the tail for determining significance? How do we use the simulation to understand sampling variability if the null hypothesis is true? How do we use the simulation to estimate the P-value? What are the assumptions needed for using randomized simulation (randomization)?

38. Describe a type 1 and type 2 error. What is the probability of a type 1 error? What is the probability of a type 2 error? Why do these errors sometime occur? What can we do to limit the chances of these error occurring?