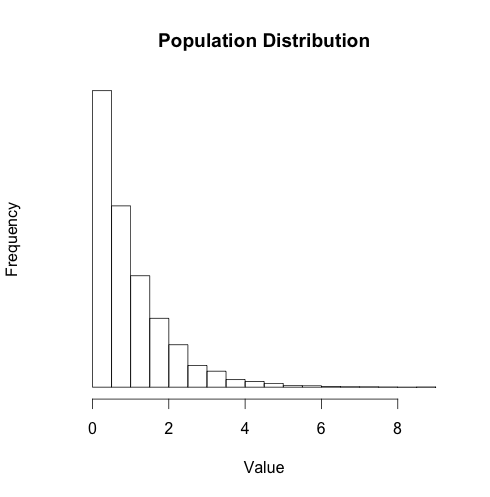
Quiz 1 Name: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Umich Uniqname: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

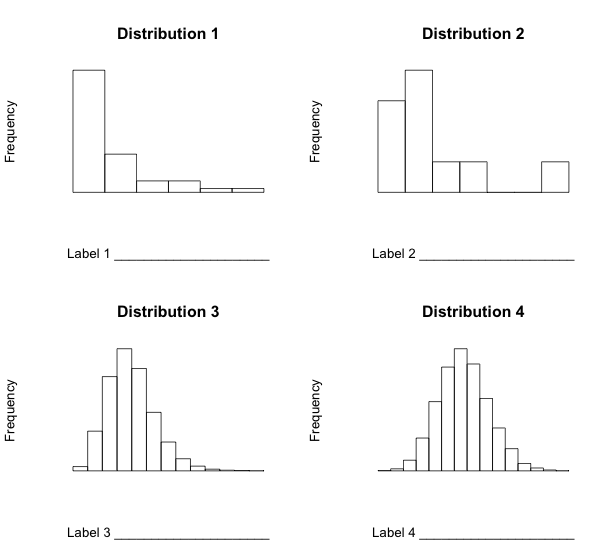
**Multiple Choice/Fill in the blank**

Circle only one choice or fill in only one word for each question unless otherwise noted in bold.

1. Categorize each of the following types of variables as (a) nominal, (b) ordinal, (c) discrete, or (d) continuous. Please write out the full word in the associated blank space (6 points).
   1. Heights of every student in SNRE \_continuous\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
   2. The shoe size of every person in the US \_\_discrete\_\_\_\_\_\_\_\_\_\_\_\_\_\_
   3. All the types of pets sold at the local pet store \_\_\_nominal\_\_\_\_\_\_\_\_\_\_\_\_\_
   4. The amount of water each household in Ann Arbor used last year \_\_continuous\_
   5. The letter grades that everyone has received in this stats class last year ordinal
   6. The brand of phone that everyone owns in Michigan \_nominal\_\_\_\_\_\_\_
2. Which type of graph would you use for each of the scenarios below. The types of graphs you can list as answers are (a) bar chart/plot, (b) histogram, (c) boxplot, (d) scatterplot. Please write out the full word in the associated blank space (5 points).
   1. You want to identify if the number of hours faculty members sleep per night is associated with the number of times they smile in class \_\_scatterplot\_\_\_\_
   2. You want to understand the distribution of a variable \_\_histogram\_\_\_\_\_
   3. You want to compare the total amount of current forest cover between four national parks \_\_bar chart/plot\_\_\_\_\_
   4. You want to visualize the relationship between the number of minutes SNRE students exercise per week and their heart rate \_\_scatterplot\_\_\_\_\_
   5. You want to see how the median and range of heights compare for citizens in Canada, Mexico, and the United States \_boxplot\_\_\_\_\_
3. Which of the following are characteristics of the normal distribution. **Please circle all answers that apply** (2 points)**:**
   1. The distribution is typically skewed to the right
   2. The mean = median = mode
   3. The mean is larger than the median
   4. The minimum parameters you need to describe a normal distribution are its mean, standard deviation, and range
   5. The distribution is bell shaped and symmetric around the mean
4. Is the mean or the median more sensitive to outliers in a dataset (1 point)?
   1. The mean is more sensitive to outliers
   2. The median is more sensitive to outliers
5. I am running a study where I am trying to identify the amount of vegetables that the average undergraduate college student eats per day. To collect data, I send out an email to all undergraduates at the University of Michigan asking them how many times they ate vegetables over the past week. Are there any types of bias that I may have based on my sampling strategy? **Please circle all answers that apply** (2 points)**:**
   1. Social desirability bias
   2. Overcoverage
   3. Undercoverage
   4. Nonresponse bias
   5. None of the above
6. Say you have the following population distribution (4 points):



You then take two sets of samples – one where the sample size is 10 and the other where the sample size is 100. You also calculate the sampling distribution of the mean for each of these sample sizes using simulations in R. Please label the following graphs using (a) Distribution of the sample with N = 10, (b) Distribution of the sample with N = 100, (c) Sampling distribution of the mean with N = 10, and (d) Sampling distribution of the mean with N = 100. **Use each label only once.**



Sampling distribution of the mean N = 100

Sampling distribution of the mean N = 10

Distribution of the sample N = 100

Distribution of the sample N = 10

1. Please circle which of the following statements are true. **Please circle all answers that apply** (2 points)**.**
   1. The standard deviation is a characteristic of a random sampling process and not a characteristic of the population.
   2. The standard error is a characteristic of a random sampling process and not a characteristic of the population.
   3. The standard deviation is a characteristic of the population.
   4. The standard error is a characteristic of the population.
2. Please draw the shape of the distributions of the sampling distribution of the mean when sample size equals 100 given the following population distributions. You do not have to draw every individual bar in the histogram – only the overall shape of the curve (3 points).

Macintosh HD:Users:mehajain:Desktop:graphs.pdf

1. When calculating confidence intervals, would you look up the critical score from either the z distribution or the t distribution for each of the following sample sizes? Please choose either ‘z distribution’ or ‘t distribution’ for each of the blanks (4 points).
   1. 10 \_\_t distribution\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
   2. 20 \_\_\_\_ t distribution \_\_\_\_\_\_\_\_\_\_\_\_\_\_
   3. 50 \_\_\_\_ z distribution \_\_\_\_\_\_\_\_\_\_\_\_\_\_
   4. 100 \_\_\_ z distribution \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
2. Which of the following are characteristics of the t distribution. **Please circle all answers that apply** (2 points)**:**
   1. The distribution is typically skewed to the right
   2. The mean = median = mode
   3. The mean is larger than the median
   4. The shape of the t distribution depends on the standard deviation of your dataset
   5. The shape of the t distribution depends on the degrees of freedom of your dataset
   6. The curve is flatter than the normal distribution at small sample sizes
   7. The distribution is bell shaped and symmetric around the mean
3. What happens to the margin of error (does it get bigger, smaller, or stay the same) in each of the following scenarios (4 points):
   1. You move from calculating a 90% confidence interval to a 95% confidence interval \_\_bigger\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
   2. You increase your sample size from 100 to 1000 \_ smaller \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
   3. You increase the standard deviation of your population dataset from standard deviation = 2 to standard deviation = 5 \_\_\_ bigger\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
   4. You increase the mean of your population dataset from mean = 2 to mean = 5 \_\_\_\_\_ stays the same \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
4. What probability value would you look up in a z-score table if I asked you to calculate the 90th% confidence interval of a dataset with N = 100? (2 points)

\_\_0.95, but you can accept 0.05 too\_\_\_

1. Which of the following is true about a 90th% confidence interval of the mean **Please circle all answers that apply** (2 points)**:**
   1. A range of values so defined that there is a 90% probability that the mean of your sample is the same as the mean of the population
   2. There is a 90% probability that the true population mean is within the confidence interval of one sample
   3. There is a 90% probability that any given confidence interval from a random sample will contain the true population mean.
   4. You must use the z distribution when sample size is < = 30
   5. You need to know the degrees of freedom of your sample when calculating z critical values and using the z-look up table

**Open-ended Questions:**

1. SNRE is interested in building a new research field station at the Matthaei Botanical Gardens that would host research that the faculty are doing and also be open to the public so citizens of Ann Arbor can learn more about the research being conducted by faculty at SNRE. While the research station sounds like a great idea, it will cost SNRE 1 million dollars and SNRE only wants to invest these funds if the majority of Ann Arbor households say that they will visit the research station at least once per year. Design a survey that would help SNRE identify whether a majority of Ann Arbor households would visit the research station. In particular please design a study that reduces selection bias and response bias (6 points).

(1) 2 points for good explanation for **way to pick representative households at random** (e.g. stratified sampling). You can give 1 point instead of 2 if they say something about picking houses at random but don’t think about stratifying their sample (e.g. random sampling). [2 points]

(2) good explanation for how to reduce response bias (e.g. household interviews instead of mail or email surveys). You can give 1 point out of 2 if you feel like they didn’t explain this well… I leave this up to your interpretation. [2 points]

(3) some discussion of trying to get a large enough sample size so that your sample statistics more closely approximate parameters of the population [1 point]

(4) 1 point for feasibility of survey – e.g. they didn’t say I will interview every household in Ann Arbor.