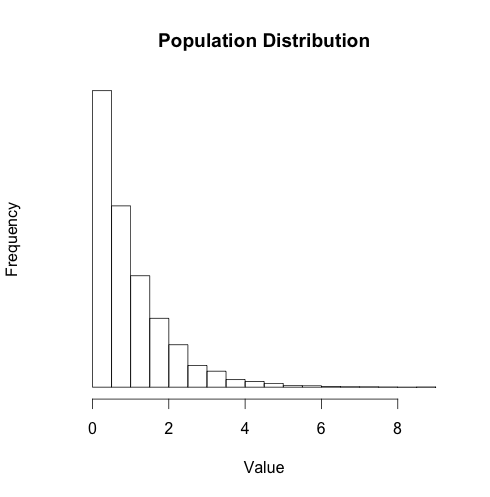
Quiz 1 Name: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Umich Uniqname: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

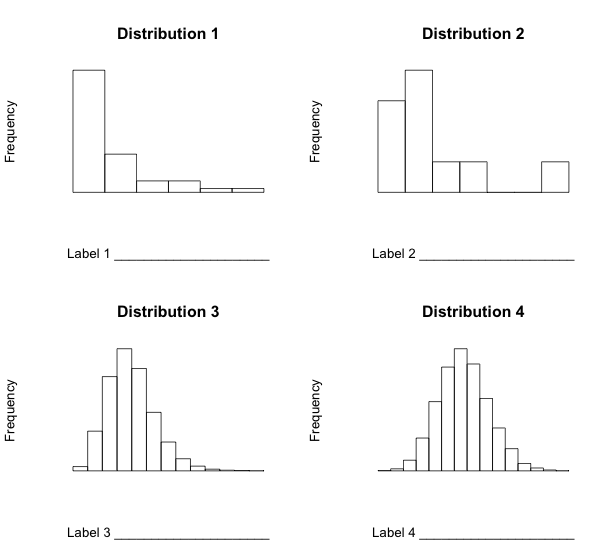
**Multiple Choice/Fill in the blank**

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. Each student’s favorite type of car \_\_\_\_\_\_nominal\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
   2. The maximum daily temperature in Ann Arbor for every day in 2017 \_\_\_\_\_\_\_\_continuous\_\_\_\_\_\_\_\_\_\_
   3. The class size (number of students) of every class at Michigan \_\_\_ \_\_\_\_\_discrete\_\_\_\_\_\_\_\_
   4. The last store that each student purchased something from \_\_nominal\_\_\_\_\_\_\_\_\_
   5. The results of the polls I post every week on Piazza asking how well you understood the lecture material (understood most, understood half, understood less than half) \_\_\_\_\_\_\_\_ordinal\_\_\_\_\_\_\_\_\_\_\_
   6. The length of EAS 538 students’ hair \_\_\_\_continuous\_\_\_\_\_\_
2. Is the mean or median less sensitive to outliers in a dataset (1 point)?
   1. The mean is less sensitive to outliers
   2. The median is less sensitive to outliers
3. 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 compare the total number of students in SEAS, the business school, and the school of public health in 2017 \_\_\_\_bar chart/plot\_\_\_\_\_\_\_\_\_\_\_
   2. You want to plot the distribution of UofM students’ carbon footprints (total amount of carbon emissions per individual) from last year \_\_\_histogram\_\_\_\_\_\_\_\_
   3. You want to see the relationship between the amount of CO2 in the atmosphere and the acidity (pH) of all lakes in Michigan from 1970 to the present \_\_\_\_\_\_\_scatterplot\_\_\_\_\_\_\_\_\_\_\_\_
   4. You want to compare the total number of visitors in 2017 in each of five zoos \_\_\_ bar chart/plot \_\_\_\_\_\_\_\_
   5. You want to see how the median and range of wheat yields compare across the world’s largest producers: China, India, and the US \_\_\_\_boxplot\_\_\_\_\_\_\_\_\_\_
4. 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 left
   2. The minimum parameters you need to describe a normal distribution are its mean and height
   3. The mean is larger than the median
   4. The distribution is bell shaped and symmetric around the mean
   5. The mean = median = mode
5. I am running a study where I am trying to quantify how much money professors across University of Michigan have donated to charity last year (2017). To collect data, I send out an email to all faculty at the University of Michigan asking them the total amount of money they gave to charity last year. Are there any types of bias that I may have based on my sampling strategy? **Please circle all answers that apply** (2 points)**:**
   1. Nonresponse bias
   2. Voluntary response bias
   3. Overcoverage
   4. Social desirability bias
   5. Undercoverage
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.**



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. 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. 7 \_\_t distribution\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
   2. 25 \_\_\_\_ t distribution \_\_\_\_\_\_\_\_\_\_\_\_\_\_
   3. 32 \_\_\_\_ z distribution \_\_\_\_\_\_\_\_\_\_\_\_\_\_
   4. 100 \_\_\_ z distribution \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
3. Which of the following are characteristics of the t distribution. **Please circle all answers that apply** (2 points)**:**
   1. The curve is flatter than the normal distribution at small sample sizes
   2. As you get a sample size of 20 or greater, the t distribution looks very similar to the normal distribution
   3. The shape of the t distribution depends on the standard deviation of your dataset
   4. The shape of the t distribution depends on the degrees of freedom of your dataset
   5. The distribution is bell shaped and symmetric around the mean
4. What happens to the margin of error/confidence interval (does it get bigger, smaller, or stay the same) in each of the following scenarios (3 points):
   1. You decrease your sample size from 1000 to 100 \_ bigger\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
   2. You move from calculating a 99% confidence interval to a 95% confidence interval \_\_smaller\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
   3. You increase the mean of your population dataset from mean = 1 to mean = 2 \_\_\_\_\_ stays the same \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
5. What probability value would you look up in a z-score table if I asked you to calculate the 95% confidence interval of a dataset with N = 100? (2 points)

\_\_0.975, but you can accept 0.025 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. University of Michigan is thinking of holding a two-day festival this Earth Day to highlight all of the great environmental research and outreach faculty, staff, and students are doing at the University. The University estimates that this event will cost $200,000, so they only want to hold the festival if approximately 20% of Ann Arbor’s population will come to the event. Please design a survey to collect this information for the University. Please consider feasibility of data collection and the ideal sampling design that will reduce the multiple types of bias we learned about in this class (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.