Module 1C Questions:

1. Imagine that we have a population that is skewed to the left. This population has a mean of 112 and a standard deviation of 16. Using a simulation program, Tyler simulated drawing 1000 samples of size 2 from the population. He then plotted the means for each of the samples that he drew. Alex simulated drawing 1000 samples of size 30, and he also plotted the means for each of the samples that he drew.

a) Would you expect the shape of Tyler' s distribution of sample means to differ from the shape of Alex's distribution of sample means? Please explain your answer (i.e., If you do expect the shapes to differ, how will they differ? If you do not expect the shapes to differ, why not?)
b) Is the mean of Tyler's distribution of sample means <,>, or = to the mean of Alex's distribution of sample means and to the mean of the sample?
c) How would you rank, from largest to smallest, the following: the standard deviation of Tyler's distribution of sample means, the standard deviation of Alex's distribution of sample means, and the standard deviation of the sample itself?

2. A researcher is studying the average **Glucose12** level in a large population of mice.
   To estimate this population mean, they repeatedly take random samples of **n = 5 mice** and compute the **sample mean** for each sample.

a) Explain in your own words what a **sampling distribution of the sample mean** is.
b) If the researcher increases the sample size from **n = 5** to **n = 20**, how will the **sampling distribution** of the sample mean change? (Think about its shape and its spread.)
c) What is **sampling error**, and why does it occur even if the researcher samples randomly and measures everything perfectly?
d) Describe a situation where two different random samples of 5 mice from the same population might give noticeably different sample means. What does that tell you about sampling error?

3. You take repeated random samples of **4 mice** from a large colony whose true mean Glucose12 level is **140 mg/dL**.
   You use the **sample mean** to estimate the true mean.

   Scenario A: Your glucose measurement tool sometimes reads **10 mg/dL too high**, but the readings are very consistent (little variation).
   Scenario B: Your tool is **unbiased**, but readings bounce around a lot (high variation).

a) Which scenario shows **high precision but high bias**?
b) Which scenario shows **low precision but low bias**?
c) Which scenario gives a sampling distribution that is *tight but centered away* from the truth?
d) If you're designing an experiment in real life, which scenario is more dangerous for inference, and why?