Part 1. (3 points) Sampling Distribution and The Central Limit Theorem Simulations

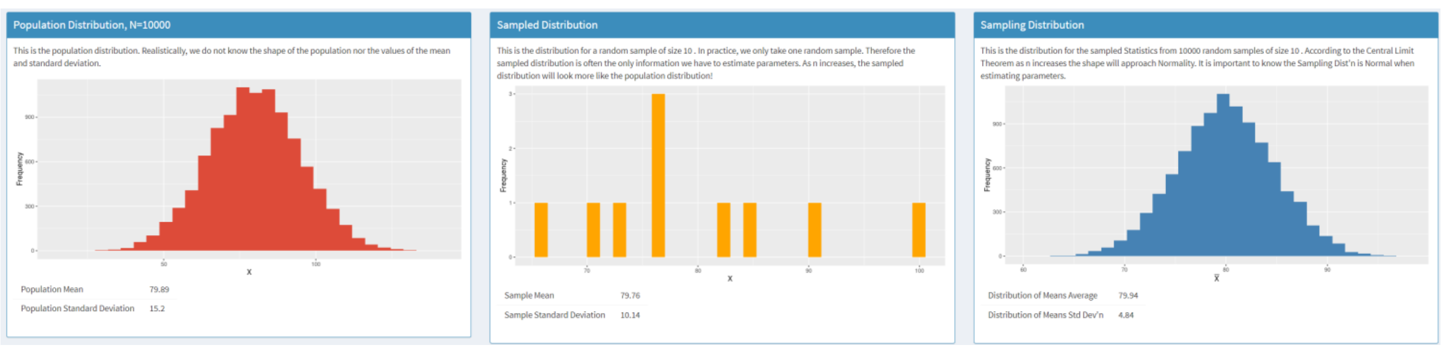
See the Central Limit Theorem in Action! Simulation is an important tool to investigate theoretical ideas. In this exercises we will:

- Randomly generate populations of different shapes
- Take random samples of different sizes
- Visualize the sampling distribution of a mean from repeated sampling.
- Identify the affect of sample size and the shape of the population on the sampled and sampling distribution.

The goal of this simulation is to visualize and validate the central limit theorem.

Getting Started: Go to the Simulation in Lesson 24 in the Week 4 Module in Canvas.

a. Simulate a normal distribution with a mean of $\mu = \mathbf{80}$ and standard deviation of $\sigma = \mathbf{15}$. Take a sample of $n = \mathbf{20}$.
   - (0.5 points) Take a screen shot of the Population, Sampled and Sampling Distribution. Paste it.



   - (0.5 points) Is the mean of the sampling distribution $\mu_{\bar{x}}$ approximately the same as the mean for the population μ_x?
     Yes, they are about the same.
   - (0.5 points) Is the standard deviation of the sampling distribution $\sigma_{\bar{x}}$ approximately $\frac{\sigma_x}{\sqrt{n}}$?
     Yes, $\frac{15.2}{\sqrt{10}}$ = 4.84.
   - (0.5 points) Is the sampling distribution approximately normal?
     Yes, this can be seen by the bell-shaped graph.

b. (1 point) Simulate different population shapes and manipulate sample sizes. You can change population shapes by selecting one of the other three options under the *Population Distribution* menu. Which population shapes need larger sample sizes in order for the sampling distribution to follow a Normal Distribution?
   A skewed population shape would need a bigger sample size to be observed as normal.

Part 2. (5 points) Confidence Interval Simulations

Confidence Intervals at Work. The goal of a confidence interval is to estimate an unknown parameter.
A confidence interval is comprised of an estimate from a sample, the standard error of the statistic and a level of confidence. We choose a confidence level based on how precise we need our estimate to be and how willing we are to risk not obtaining the parameter at all.

The definition of a 95% confidence interval states: *Out of all possible samples of size n taken from the population, the confidence intervals calculated based on those samples will contain the true parameter value 95% of the time.*

This means when we construct a 95% confidence interval, 5% of all intervals will not contain the true parameter. Therefore, we assume a 5% risk we might get an interval that does not contain the true parameter. We hope we get one of the "good" intervals. In practice, we will not know. The simulation repeatedly samples from a population, calculates a confidence interval for each sample and indicates how many confidence intervals obtain the true mean.

The goal of this simulation is to visualize and validate the definition of a confidence interval.

Getting Started: Go to the Simulation in Lesson 25 in the Week 4 Module in Canvas.

1. Start with a 90% confidence interval and the population for standard deviation.
2. Change **Sample Size to 15** and **"# of Simulations" to 1**.
3. This means you are just taking 1 sample of n = 15. This is most similar to what we do "in the real world". We only take one sample to estimate a parameter form that sample.

    a. **(1 point)** Does your 90% confidence interval contain the true mean?

    Yes, it does at the edge.

    b. **(1 point)** Increase **"# of Simulations" to 1000**. Theoretically, 90% of the sample means we obtain should result in an interval that contains the true parameter. Does this seem to be the case?
    Yes, 89.6% were within the range.

    c. **(1 point) What type of sample will fail to capture the true parameter?**
    Smaller sample sizes will fail to capture the true parameter.

- Decrease **"# of Simulations" to 100.** The intervals that don't contain the true mean are indicated in red. You can hover over a sample mean (dot in center of interval) to see it's value and the interval's margin of error.
    - Is there a common feature from the intervals that do not contain the true mean?
      They're all outliers, they are fairly far away from containing the true mean and aren't that close.
    - Where are their sample means with respect to the sample means of the intervals that do contain the parameter?
      The sample means are around ±10 compared to the ones that contain the parameter.
    - Consider the placement of the sample mean in the sampling distribution.
- *Optional:* Perform the previous steps using confidence levels 95% and 99%.

d. (1 point) How does sample size affect your confidence intervals?

- Continue with a 90% Confidence Level and "# of Simulations" at 100.
- Choose a smaller sample size between 2 and 10 observe the width of your intervals.
- Increase the sample size to something between 30 and 100 observe the width of your intervals.
- Increase your sample size to 1000 observe the width of your intervals.
- Comment on how the different sample sizes affect the performance of the confidence intervals.
  The smaller the sample size was, the wider the intervals were. At 1000 sample size, the interval was as small as 1.27. That means the more samples there are, the more confident it is.

e. (1 point) How does the confidence *level* affect your confidence *intervals*?

- Continue with a 90% Confidence Level, "# of Simulations" at 100 and a moderate sample size between 30 and 100. Observe the width of your intervals.
- Increase the confidence level to 95% observe your intervals.
- Increase the confidence level to 99% observe your intervals.
- Comment on how the different confidence levels affect the performance of the confidence intervals.
  Increasing the confidence level widens the confidence intervals, allowing more sample means to contain the true mean.

Part 3: (3 points) Solving a basic hypothesis test for a mean.

Given the following information: $\bar{x} = \mathbf{207}$, $\sigma = \mathbf{66}$, $n = \mathbf{134}$, $\alpha = \mathbf{0.10}$

$$H_0: \mu = \mathbf{200}$$

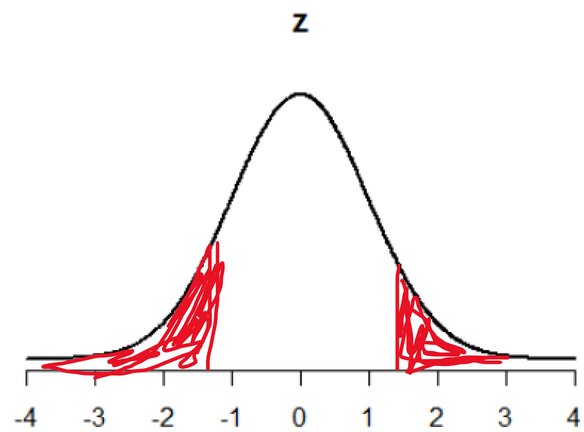
$$H_a: \mu \neq \mathbf{200}$$

a. (1 point) Calculate the z test statistic and p-value.

   Z = 1.23

   P = 0.22

b. (1 point) Draw the area under the the curve that corresponds to the p-value
   - Hint: There are many ways you can do this. In word, use the Draw feature.
   - You can draw it by hand on paper, take a picture and paste it below.



c. (1 point) Make a conclusion.
   - State whether we should reject or fail to reject the null hypothesis based on the significance level

     Since the p-value (0.22) is greater than the significance level (0.10), we fail to reject the null hypothesis.
   - Make a statement in terms of the alternative hypothesis.

     we do not have enough evidence to conclude that the true population mean is different from 200. The sample data is not different from the null hypothesis at the significance level.

Part 4. (12 points) A winery bottles 1000's of bottles of wine per season. The winery has a machine that automatically dispenses the amount of wine per bottle. Each season the wine maker randomly samples **25 bottles** of wine to ensure the amount of wine per bottle is **750 ml.** If there is evidence that the amount is **different than** (less or more than) 750 ml the winery will need to evaluate the machine and perhaps rebottle or consider selling the wine at a discount. The ***sample*** yields a mean of $\bar{x} = $ **745.4 ml.** Assume the population is Normal, with a standard deviation of $\boldsymbol{\sigma} = $ **9.9 ml**. Use a significance level of 0.05

State: Is there sufficient evidence that the average fill of the wine bottles is different than 750 milliliters?

Plan:

a. (2 points) State the null and alternative hypotheses to answer the question of interest.
The null hypothesis states that there is no difference between the mean amount of wine dispensed and the standard amount, while the alternative hypothesis states that there is a difference.

b. (2 points) Check conditions for inference. List the conditions and state whether they are met.

Randomness of sample: States as random (met)
Distribution normality: States as normal (met)
Independence: We can assume that the observations are independent due to the sample and population size. (met)
Standard deviation: known and given as 9.9. (met)

Solve:

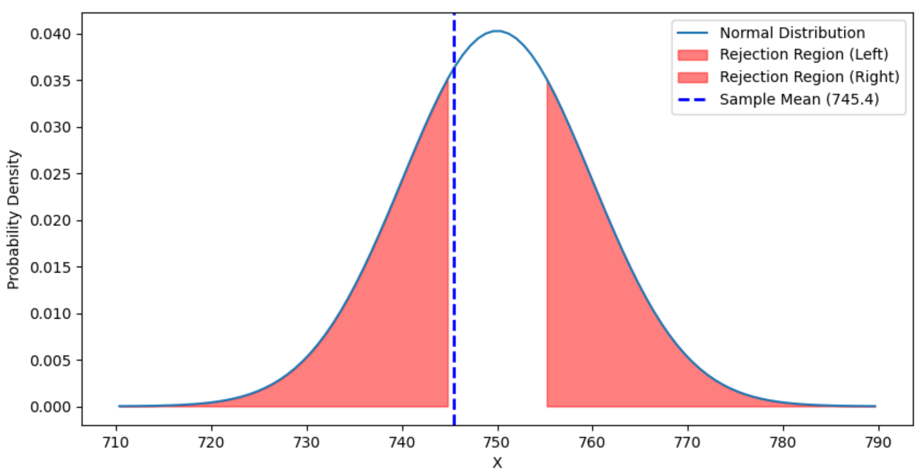
c. (1 point) Calculate the test statistic. **Show work**.

$$Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

$$Z = \frac{745.4 - 750}{\frac{9.9}{\sqrt{25}}}$$

$$Z = -2.32$$

d. (2 points) Draw the distribution for the test statistic and shade the region corresponding to the p-value (your sketch does not have to be perfect, but it should be clear what distribution the test statistic follows). What is the p-value for the test?



The p value is around 0.0203.

e. (1 point) Calculate a 95% confidence interval for μ. **Show work.**

$CI = \bar{x} \pm Z_{\frac{\alpha}{2}} * \frac{\sigma}{\sqrt{n}}$ where $Z_{\frac{\alpha}{2}}$ is the z value cutoff area. If we simply plugin in our values, we get a critical Z of 1.96 with a margin of error of 3.88. This means the confidence interval is 741.52 – 749.28.

Conclude:

f. Write a four-part conclusion describing the results.

- (1 point) Provide a statement in terms of the alternative hypothesis.
- (1 point) State whether (or not) to reject the null.
- (1 point) Give an interpretation of the point and interval estimate.
- (1 points) Include context.

The alternative hypothesis for the test conducted is that the true mean volume of wine per bottle is different from 750 ml. Given that the p value is 0.0203, which is less than the significance level o0f 0.05, we reject the null hypothesis. The point estimate for the mean volume of wine per bottle is 745.4 ml. Since we are rejecting the null hypothesis, we infer that this sample provides enough evidence that the true mean differs from 750 ml. The test suggests that the winery's machines are not dispensing the correct volume of wine per bottle. This could potentially affect product quality and customer satisfaction and thus, the winery should investigate and solve this problem.

Gradescope Page Matching (2 points)

When you upload your PDF file to Gradescope, you will need to match each question on this assignment to the correct pages. Video instructions for doing this are available in the Start Here module on Canvas on the page "Submitting Assignments in Gradescope". Failure to follow these instructions will result in a 2-point deduction on your assignment grade. Match this page to outline item "Gradescope Page Matching".