PSTAT 5A / **MIDTERM EXAM 2** / Sum. Sess. A 2023   Instructor: **Ethan Marzban**

Name:_____   **UCSB NetID:**_____
  First, then Last   *NOT your Perm Number!*

Circle Your Section: Olivier 12:30 - 1:20pm   Mengrui 2 - 2:50pm   Mengrui 3 - 3:50pm

# FREE RESPONSE QUESTIONS

Instructions:

- You will have **75 minutes** to complete the entire exam
  - Do not begin working on the exam until instructed to do so.
  - During the final 10 minutes of the exam, we will ask everyone to remain seated until the exam concludes.

- This exam comes in **TWO PARTS**: this is the **FREE RESPONSE** part of the exam.
  - There is a separate booklet containing Multiple Choice questions that should have been distributed to you at the same time as this booklet.

- Write your answers directly in the space provided on this exam booklet.
  - You do not need to write anything on your scantron for this part of the exam.

- Be sure to show all of your work; correct answers with no supporting work will not receive full credit.

- The use of calculators is permitted; the use of any other aids (including notes, laptops, phones, etc.) is strictly prohibited. A list of formulae, as well as a collection of tables, is included with this exam.

- ▢ **PLEASE DO NOT DETACH ANY PAGES FROM THIS EXAM.**

- Good Luck!!!

Honor Code: In signing my name below, I certify that all work appearing on this exam is entirely my own and not copied from any external source. I further certify that I have not received any unauthorized aid while taking this exam.

    ✕ _____

Free Response Questions

Problem 1. The length of a *GauchoSteel*-brand rod is meant to be 11 feet; due to imperfections in the manufacturing process, however, the length of a randomly-selected *GauchoSteel*-brand rod is actually a random variable X that has the following density curve:

Density Curve of X
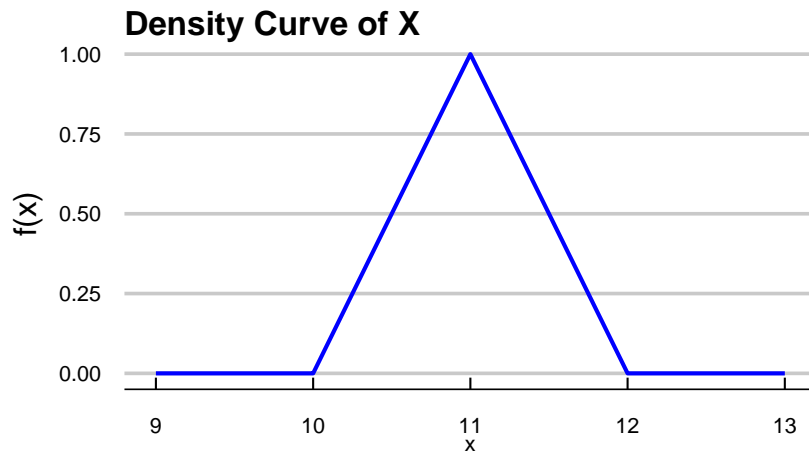
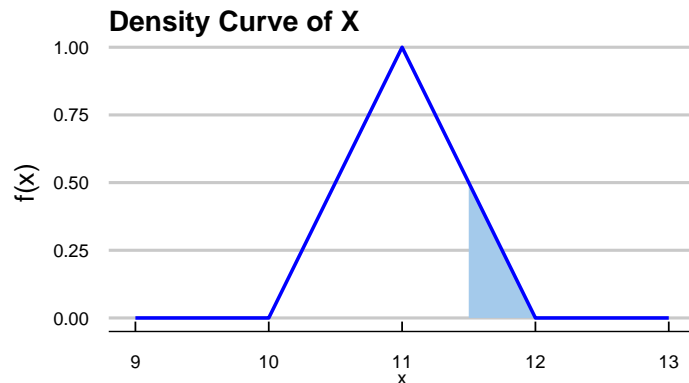(a) What is the probability that a randomly-selected *GauchoSteel*-brand rod is exactly 11 meters in length? [2pts.]

> **Solution:** Let X denote the length of a randomly-selected *GauchoSteel*-brand rod, so that the quantity we seek can be written as $\mathbb{P}(X = 11)$. Since X is continuous, we know that $\mathbb{P}(X = k) = 0$ for any value of k, meaning the desired probability is simply 0 .

(b) What is the probability that a randomly-selected *GauchoSteel*-brand rod is longer than 11.5 meters? [2pts.]

> **Solution:** Letting X be defined as above, we seek $\mathbb{P}(X > 11.5)$. This corresponds to the following area underneath the density curve:
>
> **Density Curve of X**
>
>
>
> This is a triangle with base length $(12 - 11.5) = (0.5)$ and height (0.5),

meaning its area (and, consequently, the desired probability) is simply

$$\frac{1}{2} \cdot (0.5) \cdot (0.5) = \frac{1}{8}$$

(c) A sample of 10 *GauchoSteel*-brand rods is taken with replacement, and the number of rods longer than 11.5 meters is recorded. What is the probability that this sample contains exactly 4 rods that are longer than 11.5 meters? Be sure to define any new random variables clearly and explicitly, and make sure to check any/all relevant conditions! You do **not** need to report your final answer as a decimal.

[4pts.]

> **Solution:** Let Y denote the number of *GauchoSteel*-brand rods, in a sample of 10 rods taken with replacement, that are longer tahn 11.5 meters. We suspect Y to be binomially distributed; to verify this, we check the four Binomial Conditions:
>
> 1) **Independence Across Trials?** Yes, since our sample is taken with replacement.
>
> 2) **Fixed Number of Trials?** Yes; $n = 10$.
>
> 3) **Well-Defined Notion of Success?** Yes; "success" = "given rod is longer than 11.5 meters"
>
> 4) **Fixed Probability of Success?** Yes; $p = 1/8$, as computed in part (a) above.
>
> Since all four conditions are met, we conclude that $Y \sim \text{Bin}(10, \ 1/8)$ and so, using the formula for the probability mass function of the Binomial distribution, we have
>
> $$\mathbb{P}(Y = 4) = \binom{10}{4} \left(\frac{1}{8}\right)^4 \left(1 - \frac{1}{8}\right)^{10-4} \approx 0.023$$

Problem 2. Alayah is interested in performing inference on the true average monthly rent (in thousands of dollars) of a 1-bedroom apartment in Santa Barbara. To that effect, she takes a representative sample of 100 1-bedroom apartments in Santa Barbara, and finds that these 100 apartments have a combined average monthly rent of 2.2 thousand dollars per month. From prior studies, she knows that the standard deviation of <u>all</u> monthly rents of 1-bedroom apartments in Santa Barbara is 0.75 thousand dollars.

(a) Define the parameter of interest.

[1pts.]

> **Solution:** Let μ denote the true average rent (in thousands of dollars) of a 1-bedroom apartment in Santa Barbara.

(b) Define the random variable of interest. [1pts.]

> **Solution:** Let $\overline{X}$ denote the average monthly rent, in thousands of dollars, of a representative sample of 100 1-bedroom apartments.

(c) What distribution should Alayah use when making inferences about the true average monthly rent of a 1-bedroom apartment in Santa Barbara? Be sure to check any/all relevant conditions. [3pts.]

> **Solution:** We go through our flowchart:
>
> - **Normal Population?** No; we are not told that the distribution of *all* monthly rents of 1-bedroom apartments in Santa Barbara are normally distributed.
>
> - **Large Enough Sample?** Yes; $n = 100 \geq 30$ ✓
>
> - **σ or s?** The value of 0.75 is stated to be the standard deviation of *all* rents, meaning it is σ.
>
> Based on the answers to this question, we use the standard normal distribution.
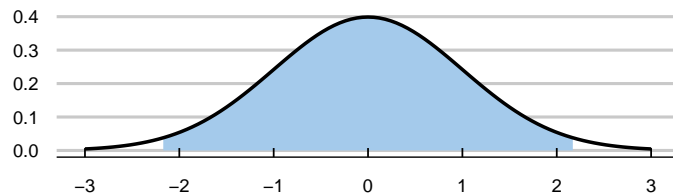
(d) Construct a 97% confidence interval for the true average monthly rent of a 1-bedroom apartment in Santa Barbara. Be sure to interpret your interval in the context of the problem! [3pts.]

> **Solution:** The general form of a confidence interval for μ, assuming we are using the normal distribution and that we have access to σ, is
>
> $$\overline{x} \pm z^* \cdot \frac{\sigma}{\sqrt{n}}$$
>
> Because we seek to construct a 97% confidence interval, we select the confidence coefficient z^* to be the value such that the following blue area is 97%:
>
>
>
> This tells us that the tails, separately, must have area $(1 - 0.97)/2 = 0.015$ meaning we seek either negative one times the 1.5^{th} percentile, or the $(100 - 1.5) = 98.5^{\text{th}}$ percentile. Either way we find $z^* \approx 2.17$, meaning our confidence interval becomes
>
> $$(2.2) \pm (2.17) \cdot \frac{0.75}{\sqrt{100}} \approx [2.0373 \ , \ 2.3628]$$

One interpretation of this interval is as follows:

> We are 97% confident that the true average monthly rent of a 1-bedroom apartment in santa Barbara is between 2.0373 and 2.3628 thousand dollars.

(e) Would a 95% confidence interval for the true average monthly rent of a 1-bedroom apartment in Santa Barbara be wider or narrower than the interval you constructed in part (d) above? Explain briefly; you do not need to construct the interval.

[2pts.]

> **Solution:** We know that, in general, higher confidence levels correspond to wider confidence intervals; conversely, lower confidence levels correspond to narrower confidence intervals. Since 95% is less than 97%, we would expect a 95% confidence interval to be narrower than the 97% confidence interval constructed in part (d) above.

Problem 3. In the field of Psychology, a Reaction Time Test is used to measure the time it takes a given person to respond to a specific stimulus; for example, how long it takes a person to press a button once the button has lit up. Suppose that for a particular stimulus, response times of randomly-selected individuals follow a normal distribution centered at 3 seconds with a standard deviation of 0.5 seconds. A person is selected at random, administered the stimulus, and their reaction time is recorded.

(a) Define the random variable of interest, and call it X.

[1pts.]

> **Solution:** Let X denote the reaction time (in seconds) of a randomly-selected individual.

(b) What is the probability that a randomly-selected person has a reaction time between 2.5 seconds and 3.7 seconds?

[3pts.]

> **Solution:** From the problem statement, we are told that $X \sim \mathcal{N}(3, 0.5)$. We seek $\mathbb{P}(2.5 \leq X \leq 3.7)$, which can be computed as:
>
> $$\mathbb{P}(2.5 \leq X \leq 3.7) = \mathbb{P}(X \leq 3.7) - \mathbb{P}(X \leq 2.5)$$
> $$= \mathbb{P}\left(\frac{X-3}{0.5} \leq \frac{3.7-3}{0.5}\right) - \mathbb{P}\left(\frac{X-3}{0.5} \leq \frac{2.5-3}{0.5}\right)$$
> $$= \mathbb{P}\left(\frac{X-3}{0.5} \leq 1.4\right) - \mathbb{P}\left(\frac{X-3}{0.5} \leq -1\right)$$
> $$= 0.9192 - 0.1587 = 0.7605$$

(c) Can you foresee any potential difficulties in modeling response times using a normal distribution? Specifically, think in terms of state spaces.   [1pts.]

> **Solution:** The state space of a normally-distributed random variable contains negative values, whereas reaction times cannot be negative.

(d) What sort of plot would be best-suited for assessing whether or not a set of reaction times could plausibly have been drawn from a normal distribution?   [1pts.]

> **Solution:** A QQ-plot is best suited for this.

Problem 4. According to the World Bank, only 54.2% of households in Ethiopia live with access to electricity. To test these claims, a sociologist takes a representative sample of 130 Ethiopian households, and observes that 67 of these households live with access to electricity. Suppose that the sociologist wants to test the World Bank's claims against a two-sided alternative, at a 5% level of significance.

(a) Define the parameter of interest, and call it p. [1pts.]

> **Solution:** Let p denote the true proportion of households in Ethiopia that live with access to electricity.

(b) Define the random variable of interest, and call it $\widehat{P}$. [1pts.]

> **Solution:** Let $\widehat{P}$ denote the proportion of Ethiopian households, in a representative sample of 130, that have access to electricity.

(c) State the null and alternative hypotheses in terms of p. [2pts.]

> **Solution:** The null hypothesis is that $p = 0.542$. We are told to adopt a two-sided alternative, meaning we take $H_A : p \neq 0.542$, and so our hypotheses become
> $$\left[\begin{array}{ll} H_0 : & p = 0.542 \\ H_A : & p \neq 0.542 \end{array} \right.$$

(d) Compute the observed value of the test statistic. [2pts.]

> **Solution:**
> $$\text{ts} = \frac{\widehat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{\left(\frac{67}{130}\right) - 0.542}{\sqrt{\frac{0.542*(1-0.542)}{130}}} \approx \boxed{-0.61}$$

(e) Compute the critical value of the test. Be sure to check any/all relevant conditions first! [3pts.]

> **Solution:** We would like to use the normal distribution to find the critical value; before we do so, however, we need to ensure that $\widehat{P}$ is normally distributed under the null, which only occurs when:
>
> **1)** $np_0 = 130 \cdot (0.542) = 70.46 \geq 10$ ✓
>
> **2)** $n(1 - p_0) = 130 \cdot (0.542) = 59.54 \geq 10$ ✓
>
> Since both conditions are satisfied, we can use the normal distribution to find the critical value. We are using a 5% level of significance, which enables us to directly recall that the critical value will be $\boxed{1.96}$.

(f) Now, perform the test and interpret your conclusions in the context of the problem. [2pts.]

> **Solution:** We reject the null only when |TS| exceeds the critical value. Here, $|ts| = |-0.61| = 0.61 < 1.96$, meaning we fail to reject the null:
>
> > At a 5% level of significance, there was insufficient evidence to reject the World Bank's claims that the true proportion of Ethiopian households with access to electricity is 54.2%, in favor of the alternative that the true proportion is *not* 54.2%.