

Score: _____ / 0

PSTAT 5A / **FINAL EXAM** / Spring 2023

Instructor: **Ethan Marzban**

Name: _____
First, then Last

UCSB NetID: _____
NOT your Perm Number!

Circle the section you **attend**:

Yuan 10 - 10:50am Jason 11 - 11:50am Nickolas 12 - 12:50pm Nickolas 1 - 1:50pm

Your Seat Number: _____

SAMPLE FREE RESPONSE QUESTIONS

Instructions:

- You will have **180 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.
 - You are allowed the use of two **8.5 × 11-inch** sheets, front and back, of notes. You are also permitted the use of **calculators**; the use of any and all other electronic devices (laptops, cell phones, etc.) is prohibited.
 - **PLEASE DO NOT DETACH ANY PAGES FROM THIS EXAM.**
 - Good Luck!!!
-

- Leonard believes that PSTAT students are just as good at bowling as Math students. To test this, he organizes a bowling match in which a group of 50 PSTAT students and a group of 50 Math students independently and simultaneously each played a game of bowling. The average number of points was used as a metric of assessing how good each team is at bowling: as such, Leonard collected the following information:

	Sample Mean	Sample Stnd. Dev.
Math	61	10.0
PSTAT	65	13.5

- Classify this as either an observational study or an experiment. Explain your reasoning.
- Classify this as either a Longitudinal or Cross-Sectional study. Explain your reasoning.

Parts (c) - (h) refer to the following: Suppose that Leonard now wishes to statistically test his claims against a two-sided alternative using a 5% level of significance. Assume all normality and independence assumptions hold. Additionally, let Population 1 refer to Math students and Population 2 refer to PSTAT students.

- Define the parameters of interest, μ_1 and μ_2 .
 - Write down the null and alternative hypotheses.
 - Compute the value of the test statistic.
 - Assuming the null is correct, what distribution does the test statistic follow? Be sure to include any/all relevant parameter(s).
 - What is the critical value of the test?
 - Now, conduct the test and phrase your conclusions in the context of the problem.
- Consider a random variable X with the following probability mass function:

k	-3	2	0	3
$P(X = k)$	0.25	0.25	a	0.25

- What must be the value of a ?
- What is $P(X \leq 1)$?
- Compute $E[X]$, the expected value of X .
- Compute $SD(X)$, the standard deviation of X .
- If $F_X(x)$ denotes the cumulative distribution function of X , what is $F_X(0)$?

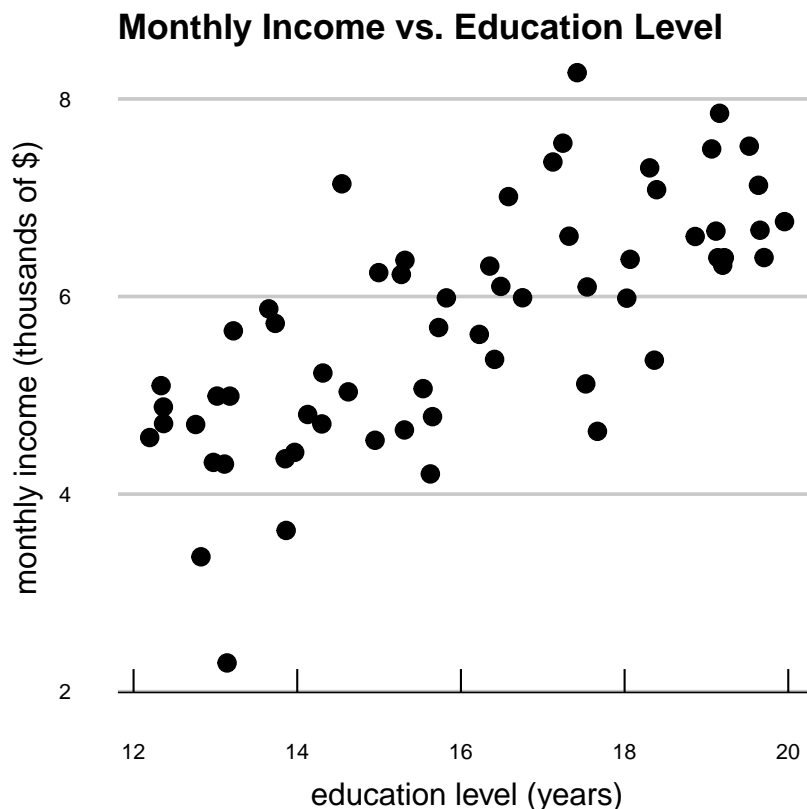
3. Recent scientific studies have revealed that there are aliens living among us (sus). Experts believe the true proportion of aliens on Earth is around 26%. To test this claim, a representative sample of 100 people is taken; it is found that 24% of these people are aliens.

- (a) What is the population in this problem?
- (b) Define the parameter of interest, p .
- (c) Define the random variable of interest, \hat{P} .

For parts (d) - (h): Assume, wherever relevant, that we are conducting a two-sided test at a 5% level of significance.

- (d) What are the null and alternative hypotheses?
 - (e) Assuming the null is correct, what is the distribution of the test statistic? Be sure to check any/all relevant conditions.
 - (f) Compute the value of the test statistic.
 - (g) Compute the p -value.
 - (h) Conduct the relevant hypothesis test, and phrase your conclusions in the context of the problem.
4. The time it takes Juan to commute to school from his apartment is normally distributed with a mean of 20 minutes and a standard deviation of 4 minutes.
- (a) What is the probability that Juan will commute to school in under 15 minutes on a randomly selected day?
 - (b) What is the probability that Juan will commute to school in over 30 minutes on a randomly selected day?
 - (c) What is the probability that Juan will commute to school in between 10 and 25 minutes on a randomly selected day?
 - (d) Let X denote the time in minutes it takes Juan to commute to school on any given day, and let Y denote the time in hours it takes Juan to commute to school on any given day. What is the distribution of Y ? be sure to include any/all relevant parameter(s)!
5. In a very old town, it is found that each building has a 15% chance of containing asbestos, independently of all other buildings. Suppose a sample of 15 buildings is taken (with replacement), and the number of buildings in this sample that contain asbestos is recorded.
- (a) Define the random variable of interest, and call it X .
 - (b) What is the distribution of X ? Be sure to check any/all relevant conditions.
 - (c) What is the probability that exactly 3 buildings in the sample of 15 contain asbestos?
 - (d) What is the standard deviation of the number of buildings (in the sample of 15) that contain asbestos?

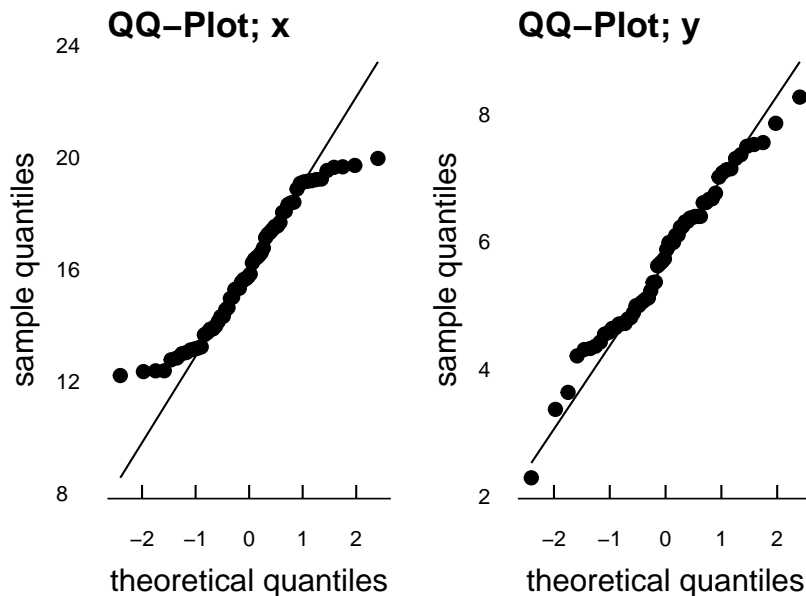
6. Leah is interested in determining whether students who listen to music while studying perform better (academically) than those who do not. To do so, she seeks out 50 people who regularly listen to music while studying and 50 do not. She then collects the average GPA from each group to use as a metric of "performance in school".
- (a) Explain why this is an observational study, and not an experiment.
 - (b) Briefly explain how Leah might restructure her study to conduct an experiment as opposed to an observational study.
 - (c) Is this a longitudinal or cross-sectional study? Explain your reasoning.
 - (d) Suppose that Leah is now interested in seeing whether the results of her study (i.e. whether listening to music while studying affects overall performance) varies between majors. What type of sampling procedure do you think Leah should carry out? Explain your reasoning.
7. Tadhg would like to model the relationship between income and education level (as measured using years of education). He collects a sample of 100 people and records their education level (i.e. years of education) and average monthly income, and produces the following scatterplot from his data:



Additionally, the following numerical summaries of his data are provided:

$$\begin{aligned}\sum_{i=1}^{62} x_i &= 992.7295 & \sum_{i=1}^{62} (x_i - \bar{x})^2 &= 343.1438 \\ \sum_{i=1}^{62} y_i &= 354.8923 & \sum_{i=1}^{62} (y_i - \bar{y})^2 &= 87.11993 \\ \sum_{i=1}^{62} (x_i - \bar{x})(y_i - \bar{y}) &= 122.4954\end{aligned}$$

Finally, below are the QQ-plots of education level (x) and monthly income (y), respectively:



- Compute $\text{Cor}(x, y)$, the correlation between x (education level) and y (monthly income).
- Compute $\hat{\beta}_0$, the intercept of the OLS regression line.
- Compute $\hat{\beta}_1$, the slope of the OLS regression line.
- Provide an interpretation of your value of $\hat{\beta}_1$. Specifically, what does a one-year change in education level correspond to with regards to a change in monthly income?
- It is known that $\text{Var}(\hat{\beta}_1) = 0.002914$. Construct a 95% confidence interval for β_1 , the slope of the true underlying linear relationship between x and y . Interpret your confidence interval.
- What is the predicted monthly income (in thousands of dollars) of someone with 15.25 years of education?
- Is it dangerous to try and use the OLS regression line to predict the monthly income (in thousands of dollars) of someone with 27 years of education? (There is a specific word/term I'm looking for here.)
- Does x appear to be normally distributed? What about y ? Why or why not (i.e. what *specifically* did you look at to answer this question)?