Score: ___________ / **68**

PSTAT 5A / **FINAL EXAM** / Spring 2023 Instructor: **Ethan Marzban**

Name:____ **UCSB NetID:**____________________

First then Last *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:_____________

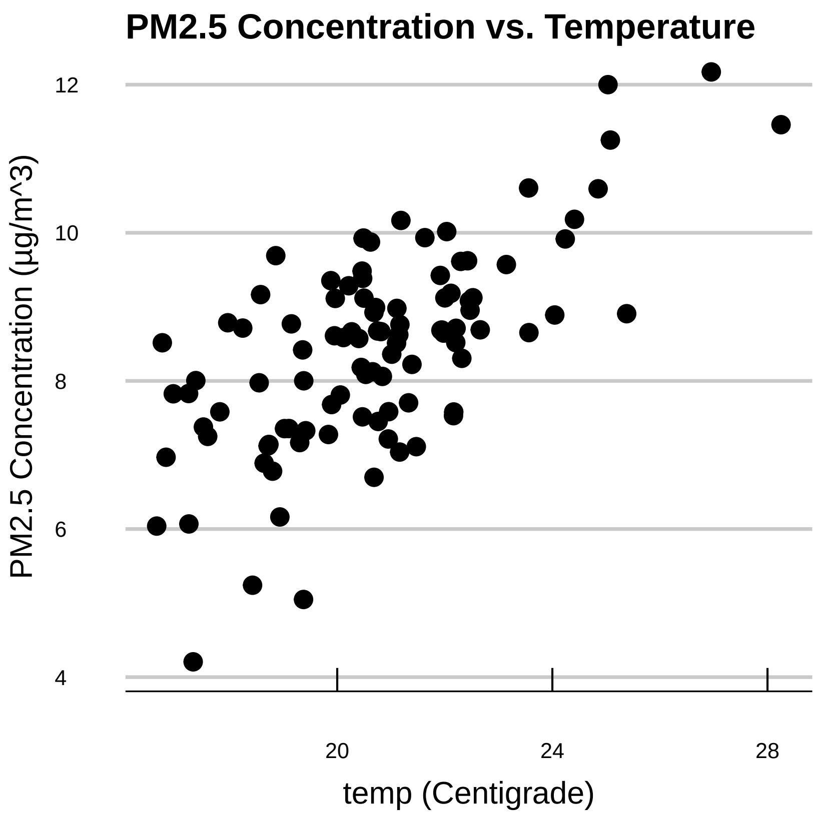
FREE RESPONSE QUESTIONS: VERSION A

Instructions:

- You will have **165 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 $\times$ 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!!!

---

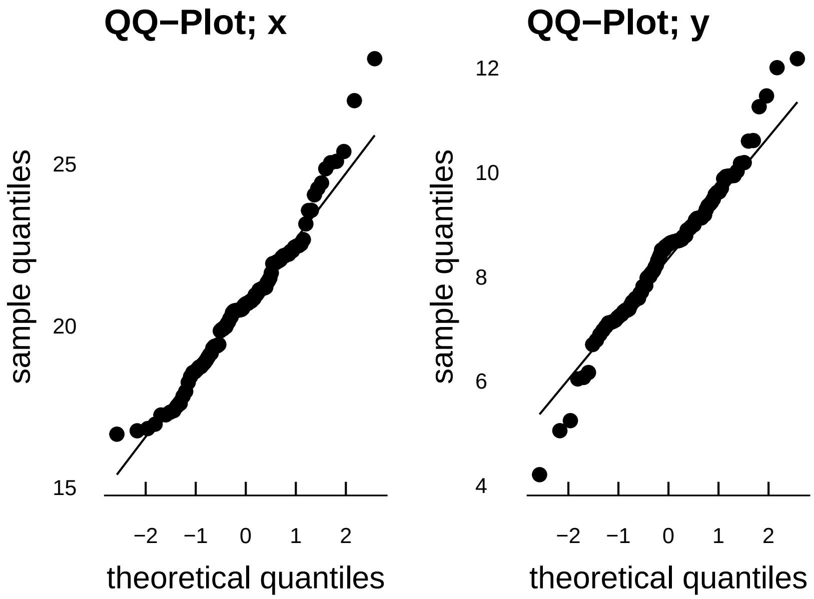
1. A researcher is interested in determining the relationship between temperature and the concentration of pollutants in the air. To do so, she takes temperature readings (in Celsius) and PM2.5 concentration readings (PM2.5 concentration can be viewed as a measure of air pollutants) at her house across 100 days. A scatterplot of her data is displayed below:



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

$$\sum_{i=1}^{100} x_i = 2072.653 \quad \sum_{i=1}^{100}(x_i - \overline{x})^2 = 486.1703$$
$$\sum_{i=1}^{100} y_i = 844.7383 \quad \sum_{i=1}^{100}(y_i - \overline{y})^2 = 180.921$$
$$\sum_{i=1}^{100}(x_i - \overline{x})(y_i - \overline{y}) = 200.9171$$

Finally, below are the QQ-plots of `temperature` and `PM2.5 Concentration`
, respectively:



(a) (3 points) Compute $\text{Cor}(x,\ y)$, the correlation between x (temperature) and y (PM2.5 concentration).

(b) (2 points) Compute $\widehat{\beta_1}$, the slope of the OLS regression line.

(c) (2 points) Compute $\widehat{\beta_0}$, the intercept of the OLS regression line.

(d) (2 points) Provide an interpretation of your value of $\widehat{\beta_1}$. Specifically, what does a one-degree change in temperature correspond to with regards to a change in PM2.5 concentration?

(e) (3 points) It is found that $\text{SD}(\widehat{\beta_1}) = 0.0453$. Construct a 95% confidence interval for β_1, the slope of the true underlying linear relationship between `x` and `y`. Interpret your confidence interval.

(f) (3 points) Suppose on day 101 (i.e. the day right after the researcher stops recording data), the temperature at the researcher's house reaches 23 degrees celsius. What is a good estimate for the concentration of PM2.5 on day 101?

(g) (3 points) Is it dangerous to try and use the OLS regression line to predict the PM2.5 concentration of a day in which the temperature is 2 degrees Centigrade? Why or why not? (There is a specific word/term I'm looking for here.)

(h) (2 points) Do you think it is reasonable to assume the datapoints were independent from each other? Why or why not?

(i) (2 points) Assuming the independence assumption *was* reasonable (which doesn't mean that is the correct answer for the previous part!), do the normality assumptions appear to hold? Why or why not (i.e. what did you look at to answer this question)?

2. At an animal shelter, 55% of all cats are female. Additionally, 40% of all cats at the shelter are tabby cats. Furthermore, 25% of the tabby cats are female.

(a) (2 points) Define events, and translate the information provided in the problem.

(b) (3 points) What is the probability that a randomly selected cat is both tabby and female?

(c) (3 points) What is the probability that a randomly selected cat is tabby or female?

(d) (3 points) What is the probability that a randomly selected cat is tabby, given that it is female?

3. A scientist believes that the average amount of air pollution in City A is the same as the average amount of air pollution in City B. As a metric for measuring air pollution, the scientist decides to use the concentration of PM2.5 particles. She then collects 32 measurements from City A and 32 measurements from City B, and computes the following summaries:

	Sample Mean	**Sample Stnd. Dev.**
City A	8	2.5
City B	7	3

(a) (2 points) Classify this as either an observational study or an experiment. Explain your reasoning.

(b) (2 points) Classify this as either a Longitudinal or Cross-Sectional study. Explain your reasoning.

(c) (2 points) Which of the three sampling procedures discussed in class (simple random, stratified, and clustered) do you think the researcher used when collecting her data? Explain your reasoning.

Parts (d) - (i) refer to the following: Suppose that the scientist now wishes to statistically test her beliefs against a two-sided alternative using a 5% level of significance. Assume all normality and independence assumptions hold. Additionally, let Population 1 refer to City A and Population 2 refer to City B.

(d) (2 points) Define the parameters of interest, μ_1 and μ_2.

(e) (2 points) Write down the null and alternative hypotheses.

(f) (3 points) Compute the value of the test statistic.

(g) (4 points) Assuming the null is correct, what distribution does the test statistic follow? Be sure to include any/all relevant parameter(s).

(h) (3 points) What is the critical value of the test?

(i) (2 points) Now, conduct the test and phrase your conclusions in the context of the problem.

4. At a particular salad bar, a salad is created from a choice of base (lettuce, spring mix, or kale), 5 toppings (to be selected from a list of 12 possible toppings, and it is permitted to order multiple servings of the same topping), and a dressing (Caesar, balsamic vinaigrette, or ranch). A single salad (i.e. base + topping + dressing) is to be selected at random from the total set of salads that can be created.

   (a) (4 points) If Ω denotes the outcome space of this experiment, how many elements are in Ω?

   (b) (1 point) Are we justified in utilizing the classical approach to probability in this problem? Why or why not?

(c) (4 points) Use the Classical Approach to Probability to compute the probability that the selected salad has ranch dressing.

(d) (4 points) Use the Classical Approach to Probability to compute the probability that the selected salad has kale <u>and</u> ranch dressing.