BLUE VERSION **Score:** ________ / **50**

PSTAT 5A / **MIDTERM EXAM 1** / Spring 2023 Instructor: **Ethan Marzban**

Name:____ **UCSB NetID:**____________________

First, then Last *NOT your Perm Number!*

Circle the section you <u>attend</u>:

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

Your Seat Number:____________

Person Sitting to your Left:____

Person Sitting to your Right:____

Instructions:

- You will have **65 minutes** to complete this exam.
- You are allowed the use of a single **8.5 $\times$ 11-inch** sheet, 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, airpods/headphones, etc.) is prohibited.
- For Multiple Choice Questions: fill in the bubble corresponding to your answer directly on the exam. Partial credit will **not** be awarded.
- For Free Response Questions: be sure to include **all** of your work! Correct answers with no supporting work will **not** receive full points.
- **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.

$\times$ ____

Problem 1. If the variable X contains measurements on the duration (in minutes) of 100 different flights from SBA to EWR, what is the correct classification of X? [1pts.]

- ◯ discrete
- √ **continuous**
- ◯ nominal
- ◯ ordinal

Problem 2. Suppose a password for a particular website must be 5 characters long, consisting of exactly 2 digits (0 through 9), 2 letters (A through Z), and 1 special character (!, @, #, $, %), in that order. What is the total number of passwords that can be constructed using this scheme? [1pts.]

- ◯ 1,300
- ◯ 73,125
- ◯ 292,500
- √ 338,000
- ◯ None of the above.

Problem 3. Jana has run the following code: [1pts.]

```
def f(x, y):
    """return the sum of x and y"""
    x + y
```

What will be the output of running `f(1, 2)`?

- ◯ `3`
- √ **Nothing**
- ◯ An Error
- ◯ `[1, 2]`
- ◯ None of the above.

Problem 4. Events A and B are such that $\mathbb{P}(A) = 0.3$, $\mathbb{P}(B) = 0.8$, and $\mathbb{P}(A \cap B) = 0.24$. Select the statement that is correct. [1pts.]

- √ **A and B are independent, but not disjoint**
- ◯ A and B are disjoint, but not independent
- ◯ A and B are both disjoint and independent
- ◯ A and B are neither disjoint nor independent

Problem 5. **True or False:** If $\{x_i\}_{i=1}^n$ is a set of numbers with mean $\overline{x}$, then the mean of the set $\{ax_i\}_{i=1}^n$ for a fixed constant a is simply $a \cdot \overline{x}$. [1pts.]

√ **True**

◯ False

◯ Not Enough Information to Determine

Problem 6. Guadalupe would like to visualize the relationship between a person's favorite color and their height. Which type of graph should she use? [1pts.]

◯ A bargraph

◯ A histogram

√ **A side-by-side boxplot**

◯ A scatterplot

◯ None of the above

Problem 7. In what module is the function `make_array()` found? [1pts.]

√ `datascience`

◯ `numpy`

◯ `python_arrays`

◯ None of the above

Problem 8. In a variable re-assignment statement in Python, which side of the equality does Python evaluate first? [1pts.]

√ **Right**

◯ Left

Problem 9. In order for $\mathbb{P}(A \mid B)$ to be defined for two events A and B, which of the following conditions must be true? **Select only ONE answer choice**. [1pts.]

◯ $\mathbb{P}(A) \neq 0$

√ $\mathbb{P}(B) \neq 0$

◯ $\mathbb{P}(A \cap B) \neq 0$

◯ $\mathbb{P}(A \cup B) \neq 0$

◯ None of the above.

Problem 10. Which of the following is **not** a measure of spread? [1pts.]

◯ Interquartile Range

◯ Standard Deviation

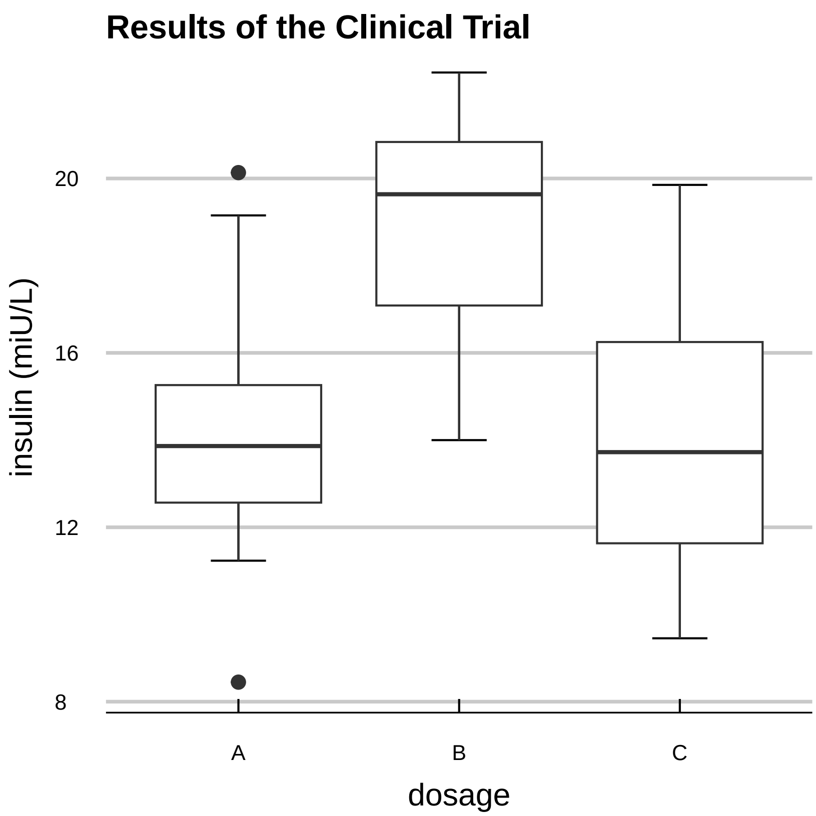
√ **50th Percentile**

◯ Range

◯ None of the above

Free Response Questions

Problem 11. In a clinical trial, subjects were administered one of three different dosages of a particular drug. 3 hours later, the insulin count (in miU/Liter) of each subject was taken and recorded. The results of the trial are displayed below:



(a) Provide the 5-number summary for the insulin levels of subjects who were administered dosage A. Round your numbers to the nearest decimal place. [3pts.]

Solution:

min	Q_1	median	Q_3	max
8.5	12.5	14.0	15.3	20.1

(These are, of course, only approximate.)

(b) Approximately what percent of subjects who were administered dosage C had insulin levels lower than 16.1 miU/L? [2pts.]

Solution: It appears that 16.1 is the third quartile of insulin measurements of individuals administered dosage C. As such, by definition of the third quartile, this means that approximately 75% of subjects administered dosage C has insulin levels lower than 16.1.

(c) Does there appear to be a difference in insulin levels across dosages? Explain in one or two brief sentences. [3pts.]

Solution: **Answers may vary**. Based on the relative positions of the boxplots, it seems that there was no significant difference in average insulin levels between individuals administered dosages A and C, whereas individuals administered dosage B appear to have (on average) higher insulin levels.

Problem 12. Consider the set of numbers

$$B = \{-2,\ -1.5,\ 0,\ 8\}$$

(a) Compute $\overline{b}$, the mean of B. [3pts.]

Solution:

$$\overline{b} = \frac{1}{4}[(-2) + (-1.5) + (0) + (8)] = \frac{4.5}{4} = 1.125$$

(b) Compute the standard deviation of B. [4pts.]

Solution:

$$\begin{aligned} s_b^2 &= \frac{1}{5-1}\left[(-2-1.125)^2 + (-1.5-1.125)^2 + (0-1.125)^2 + (8-1.125)^2\right] \\ &= \frac{1}{4}(65.1875) \approx 16.297 \\ s_b &= \sqrt{s_b^2} = \sqrt{16.297} \approx 4.037 \end{aligned}$$

(c) Compute the median of B. [2pts.]

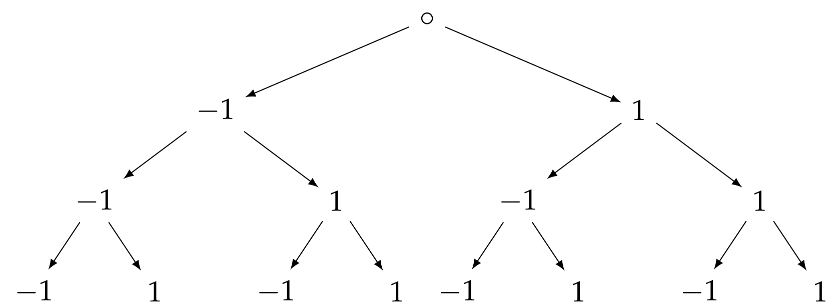
Solution:

$$B = \{\cancel{-2},\ -1.5,\ 0,\ \cancel{8}\} \implies \text{median}(B) = \frac{-1.5+0}{2} = -0.75$$

Problem 13. Three numbers are to be selected at random from the set $\{-1, 1\}$. Assume we replace the numbers after each draw, and assume that the order in which the numbers are selected <u>is</u> important.

(a) Write down the outcome space Ω for this experiment. [3pts.]

Solution: Using a Tree will be easiest:



Alternatively, we could have listed out all of the elements explicitly:

$$\Omega = \{(-1,-1,-1),\ (-1,-1,\ 1),\ (-1,\ 1,-1),\ (1,-1,-1),\\ (-1,\ 1,\ 1),\ (1,-1,\ 1),\ (1,\ 1,-1),\ (1,\ 1,\ 1)\}$$

(b) How many elements are in Ω? [2pts.]

Solution: From part (a), we see there are 8 outcomes in Ω. We could have also found this using a slot diagram with three slots (one for each number):

$$\underline{2} \times \underline{2} \times \underline{2} = 8$$

(c) Are we justified in using the Classical Approach to probability in this problem? Why or why not? [1pts.]

Solution: Yes, since the numbers are stated to have been selected "at random". In the Yellow version, it was not stated whether the numbers were selected "at random" so the answer would have been No; however, I decided to award everyone the full point on this problem (across versions) so long as they wrote *something*.

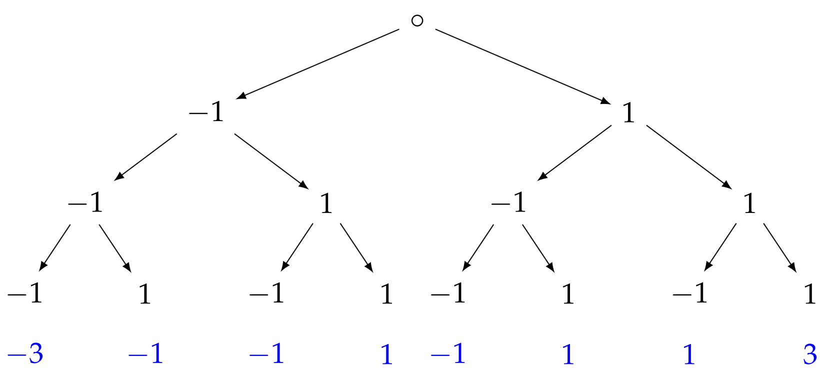
(d) Let A denote the event "the first number selected was greater than the second number selected." Write down the mathematical formulation of A; i.e. identify the outcomes that are contained in A. [3pts.]

Solution: If the first number selected was -1, then the second number must have been 1 if it is to be greater than the first number selected. If first number selected was 1, then there are no possibilities for the second number; as such,

$$A = \{(-1,\ 1, -1),\ (-1,\ 1,\ 1)\}$$

(e) Let E denote the event "the sum of the three numbers selected is 1". Compute $\mathbb{P}(E)$ using the classical approach to probability. [4pts.]

Solution: We should figure out what sum each of the 8 outcomes in Ω correspond to:



We can now see that there are 3 outcomes in which the sum of the three numbers is 1:

$$E = \{(-1,\ 1,\ 1),\ (1, -1,\ 1),\ (1,\ 1,\ -1)\}$$

and so, by the Classical Approach to Probability,

$$\mathbb{P}(E) = \frac{3}{8} = 37.5\%$$

Please Note: For full credit, you needed to have justified your answer for the numerator somehow, either by writing the mathematical formulation of E or by making some sort of explicit counting argument. If you just jumped straight to $\#(E) = 3$, you did not receive full credit.

Problem 14. It is known that 5% of people in the town of *Gauchoville* are affected by a particular disease. There is a test for this disease, however it is imperfect- specifically, it has a 25% false positive rate and a 10% false negative rate.

(a) Define appropriate notation (i.e. define relevant events), and translate the information provided into the problem to be in terms of the events you define. [2pts.]

> **Solution:** Let D denote the event "a randomly selected person has the disease", and $+$ denote "a person tests positive". From the problem statement, we therefore have
>
> $$\mathbb{P}(D) = 0.05; \quad \mathbb{P}(+ \mid D^{\complement}) = 0.25; \quad \mathbb{P}(- \mid D) = 0.1$$

(b) What is the probability that a randomly selected person will both have the disease <u>and</u> test positive? [2pts.]

> **Solution:** We seek $\mathbb{P}(D \cap +)$. By the Multiplication Rule,
>
> $$\mathbb{P}(D \cap +) = \mathbb{P}(+ \mid D) \cdot \mathbb{P}(D) = (1 - 0.1) \cdot (0.05) = \boxed{0.045 = 4.5\%}$$

(c) What is the probability that a randomly selected person will test positive? [3pts.]

> **Solution:** We seek $\mathbb{P}(+)$, which can be computed using the Law of Total Probability:
>
> $$\begin{aligned} \mathbb{P}(+) &= \mathbb{P}(+ \mid D) \cdot \mathbb{P}(D) + \mathbb{P}(+ \mid D^{\complement}) \cdot \mathbb{P}(D^{\complement}) \\ &= (1 - 0.1) \cdot (0.05) + (0.25) \cdot (1 - 0.05) \\ &= (0.9) \cdot (0.05) + (0.25) \cdot (0.95) = \boxed{0.2825 = 28.25\%} \end{aligned}$$

(d) Suppose Fatima has tested herself for the disease, and her test returned a positive result. What is the probability that she actually has the disease? [3pts.]

> **Solution:** We seek $\mathbb{P}(D \mid +)$, which can be computed using Bayes' Rule:
>
> $$\mathbb{P}(D \mid +) = \frac{\mathbb{P}(+ \mid D) \cdot \mathbb{P}(D)}{\mathbb{P}(+)} = \frac{0.045}{0.2825} = \boxed{\frac{18}{113} \approx 15.9\%}$$

You may use this page for scratch work, if necessary. Keep in mind that NOTHING on this page will be graded.

You may use this page for scratch work, if necessary. Keep in mind that NOTHING on this page will be graded.