

Score: \_\_\_\_\_ / 15

PSTAT 5A / MIDTERM EXAM 1 / Sum. Sess. A 2023

Instructor: **Ethan Marzban**

Name: \_\_\_\_\_  
*First, then Last*

UCSB NetID: \_\_\_\_\_  
*NOT your Perm Number!*

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

## MULTIPLE CHOICE QUESTIONS VERSION A

---

### 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 **MULTIPLE CHOICE** part of the exam.
    - There is a separate booklet containing Free-Response questions that should have been distributed to you at the same time as this booklet.
  - Fill in the bubble corresponding to your answer **on the provided scantron**; **Absolutely NOTHING** written directly on this exam booklet will be graded. Partial credit will **not** be awarded.
    - Unless explicitly instructed otherwise, mark only one answer per question. If you mark multiple answers for the same question, you will receive 0 points for the question even if one of your choices is correct.
  - The use of calculators is permitted; the use of any other aids (including notes, laptops, phones, etc.) is strictly prohibited. A list of formulae is included with this exam.
  - **PLEASE DO NOT DETACH ANY PAGES FROM THIS EXAM.**
  - Good Luck!!!
-

**Problem 1.** Suppose that, in a particular Jupyter Notebook, the variable  $x$  has previously been defined. When evaluating the expression  $x = x * 7$ , which side of the equality does Python evaluate first? [1pts.]

- A. Left
- B. Right**

**Problem 2.** A set of numbers  $X = \{x_i\}_{i=1}^4$  is such that  $\bar{x} = 3$ . Define  $Y = \{x_1, \dots, x_4, 3, 4\}$ ; that is,  $Y$  consists of the same numbers in  $X$  along with the numbers 3 and 4. What is  $\bar{y}$ , the mean of  $Y$ ? [1pts.]

- A. 1.6667
- B. 3.0000
- C. 3.1667**
- D. 4.0000
- E. None of the above.

**Problem 3.** Let  $E$  be an event with  $\mathbb{P}(E) = 0.1$ . What is  $\mathbb{P}(E \mid \emptyset)$ , where  $\emptyset$  denotes the empty set? [1pts.]

- A. 0
- B. 1
- C.  $\infty$
- D. Undefined**
- E. None of the above.

**Problem 4.** Events  $A$  and  $B$  are such that  $\mathbb{P}(A) = 0.5$ ,  $\mathbb{P}(B) = 0.5$ , and  $\mathbb{P}(A \cap B) = 0.25$ . Select the statement that is correct. [1pts.]

- A.  $A$  and  $B$  are independent, but not mutually exclusive**
- B.  $A$  and  $B$  are mutually exclusive, but not independent
- C.  $A$  and  $B$  are both mutually exclusive and independent
- D.  $A$  and  $B$  are neither mutually exclusive nor independent

**Problem 5.** A *GachoLotto* ticket number consists of 2 Letters ( $A - Z$ ) and 1 digit ( $0 - 9$ ), in any order. For example, A3A is a valid *GachoLotto* ticker number. Assuming repeated digits are allowed, what is the total number of *GachoLotto* tickets that can be created? [1pts.]

- A. 6,500
- B. 6,760
- C. 39,000
- D. 40,560**
- E. None of the above

Problems 6 - 10 refer to the following: Consider the following data matrix:

color	height	weight	grade
red	2.1	3.2	A+
red	3.3	4.1	B-
orange	1.1	2.2	A-
blue	4.4	5.7	C
orange	2.3	2.9	A

- Problem 6.** How many observational units are present in this data matrix? [1pts.]
- A. 4
  - B. 5**
  - C. 6
  - D. 7
  - E. None of the above.
- Problem 7.** What is the correct classification of the `color` variable? [1pts.]
- A. Discrete
  - B. Continuous
  - C. Ordinal
  - D. Nominal**
- Problem 8.** What type of visualization is best suited for the `grade` variable? [1pts.]
- A. Barplot/Bargraph**
  - B. Boxplot
  - C. Histogram
  - D. Scatterplot
  - E. None of the above.
- Problem 9.** What type of visualization is best suited to visualize the relationship between the variables `height` and `grade`? [1pts.]
- A. Histogram
  - B. Scatterplot
  - C. Side-by-side Boxplot**
  - D. Line Graph
  - E. None of the above.
- Problem 10.** Crucially, this data set is missing a description of what the variables represent. What is such a description called? [1pts.]
- A. Data Descriptor
  - B. Data Matirx Supplement
  - C. Data Dictionary**
  - D. Data Science
  - E. None of the above.

**Problems 11 - 15 refer to the following situation:** Katrina would like to write a function called `length_classifier()` that takes in a single list  $x = [x_1, x_2, \dots, x_n]$ . The function is meant to output one of three things:

- If  $x$  has 3 or fewer elements, the function should output the string "short"
- If  $x$  has between 4 and 7 elements (inclusive on both ends), the function should output the string "medium"
- If  $x$  has 8 or more elements, the function should output the string "long"

To that end, she has written the following skeleton code, but it is missing some crucial parts. (Assume this is the **only** code in Katrina's Jupyter Notebook, and that there are **no** other code cells before or after.

```
def length_classifier(x):
    if len(x) Blank 1 3:
        return "short"
    Blank 2 4 Blank 3 len(x) Blank 4 8:
        return "medium"
    else:
        return "long"
```

---

**Problem 11.** What should go in Blank 1?

[1pts.]

- A. <
- B. <=**
- C. >
- D. >=
- E. None of the above.

**Problem 12.** What should go in Blank 2?

[1pts.]

- A. **else**
- B. **else if**
- C. elif**
- D. **e\_if**
- E. None of the above.

**Problem 13.** What should go in Blank 3?

[1pts.]

- A. <
- B. <=**
- C. >
- D. >=
- E. None of the above.

**Problem 14.** What should go in Blank 4?

[1pts.]

- A. <**
- B. <=
- C. >
- D. >=
- E. None of the above.

**Problem 15.** What is missing from the body of Katrina's function (specifically, this is something we mentioned in Lab that should *always* be included with a function)

[1pts.]

- A. An output statement
- B. A return statement
- C. An exception statement
- D. A docstring**
- E. None of the above.

.....