

Score: \_\_\_\_\_ / 30

PSTAT 5A / FINAL EXAM / 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 **160 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, as well as a collection of tables, is included with this exam.
  - **PLEASE DO NOT DETACH ANY PAGES FROM THIS EXAM.**
  - Good Luck!!!
-

**Problems 1 - 5 refer to the following situation:** At the *GauchaCinema*, it is found that 60% of people are going to watch *Barbie* and 50% are going to watch *Oppenheimer*. Additionally, of those watching *Barbie* it is found that 50% are going to watch *Oppenheimer* as well. A person is selected at random, and the movie/s they are going to watch is recorded.

**Problem 1.** What is the probability that the randomly-selected person is going to watch both *Barbie* and *Oppenheimer*? [1pts.]

- A. 0.1
- B. 0.3
- C. 0.5
- D. 0.6
- E. None of the above.

**Problem 2.** Given that the person is going to watch *Oppenheimer*, what is the probability that they also watch *Barbie*? [1pts.]

- A. 0.1
- B. 0.3
- C. 0.5
- D. 0.6
- E. None of the above.

**Problem 3.** What is the probability that the randomly-selected person watches *Barbie* but not *Oppenheimer*? Assume that the probability of watching both *Barbie* and *Oppenheimer* is 0.3 (which isn't to say this is the correct answer to Problem 2 above!). [1pts.]

- A. 0.1
- B. 0.3
- C. 0.5
- D. 0.6
- E. None of the above.

**Problem 4.** Let  $B$  denote the event "the person watches *Barbie*" and  $O$  denote the event "the person watches *Oppenheimer*." Are  $B$  and  $O$  independent? [1pts.]

- A. Yes
- B. No
- C. Not enough information to determine.

**Problem 5.** Let  $B$  and  $O$  be defined as in Problem 4 above. Are  $B$  and  $O$  disjoint? [1pts.]

- A. Yes
- B. No
- C. Not enough information to determine.

.....

**Problems 6 - 9 refer to the following situation:** The **geometric mean** of a list of numbers  $\{y_i\}_{i=1}^n$  is defined to be

$$\bar{y}_{\text{geom}} = (y_1 \times y_2 \times \cdots \times y_n)^{\frac{1}{n}}$$

i.e. the geometric mean is computed by first computing the product of the numbers, and then raising the product to the power  $(1/n)$  where  $n$  is the number of observations. João would like to write a Python function called `geom_mean()` that takes in a single input `y = [y1, ..., yn]` and outputs the geometric mean of `y`. To that end, he has written the following code, and has nothing written above it:

```
def geom_mean(y):
    """
    return the geometric mean of y
    """
    n = len(y)
    prod_y = 1
    for k in Blank 1:
        prod_y Blank 2 k
    return (prod_y) Blank 3 (1/n)
```

---

**Problem 6.** What should go in Blank 1?

[1pts.]

- A. `k`
- B. `y`
- C. `geom_mean`
- D. `len`
- E. None of the above

**Problem 7.** What should go in Blank 2?

[1pts.]

- A. `*=`
- B. `+=`
- C. `=*`
- D. `+=`
- E. None of the above

**Problem 8.** What should go in Blank 3?

[1pts.]

- A. ^
- B. ^^
- C. \*
- D. \*\*
- E. None of the above

**Problem 9.** Assuming all blanks are filled in correctly, what would be the output of running `geom_mean(1, 2, 3)`?

[1pts.]

- A. 0.5503
- B. 1.8171
- C. 2.0000
- D. An Error
- E. None of the above

.....  
**Problems 10 - 15 refer to the following situation:** Consider the following data matrix:

grade	sleep	major	fav_color
A+	7.8	PSTAT	Green
B	6.9	PSYCH	Gold
A-	7.0	SOC	Red
B	5.5	PSTAT	Gold
C+	6.7	PSTAT	Purple

We are also provided with the following data dictionary:

- **grade:** letter grade
- **sleep:** amount of sleep (in hours)
- **major:** major
- **fav\_color:** favorite color

**Problem 10.** What is the best type of visualization to visualize the relationship between **sleep** and **fav\_color**? [1pts.]

- A. Histogram
- B. Barplot
- C. Scatterplot
- D. Side-by-side Boxplot
- E. None of the above

**Problem 11.** Which of the variables below is ordinal? (There is only one correct answer choice.) [1pts.]

- A. **grade**
- B. **sleep**
- C. **major**
- D. **fav\_color**

**Problem 12.** Suppose Ayesha wants to model the relationship between **sleep** and **grade**, using **grade** as the response variable and **sleep** as the explanatory variable. Is this a regression problem or a classification problem? [1pts.]

- A. Regression
- B. Classification

**For Problems 13 - 15:** Suppose the above data matrix has been imported into Python as a `datascience` table called `students`. Also assume the `datascience` module has been imported, and that it has been imported without any nickname.

**Problem 13.** What would be the result of running the code [1pts.]

```
students.column(2).item(3)
```

- A. 7.0
- B. 5.5
- C. SOC
- D. PSTAT
- E. None of the above.

**Problem 14.** Which of the answer choices below best describes what the following code is doing:

[1pts.]

```
students.row(students.column(3) == "Gold") [0]
```

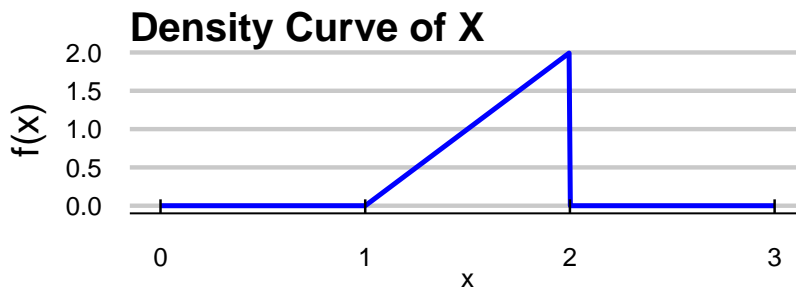
- A. It returns the favorite colors of students whose favorite color was Gold.
- B. It returns the grades of students whose favorite color was Gold.
- C. It returns the number of students whose favorite color was Gold.
- D. It returns an error.
- E. None of the above.

**Problem 15.** What does the output of `len(students.labels)` represent?

[1pts.]

- A. The number of variables
- B. The number of observational units
- C. The number of explanatory variables.
- D. The total number of elements in the table
- E. None of the above.

.....  
**Problems 16 - 18 refer to the following situation:** The random variable  $X$  has the following density curve (if the picture is difficult to read, the density curve is zero up to 1, a straight line from the point  $(1, 0)$  to  $(2, 2)$ , and then zero from 2 onwards):



**Problem 16.** What is the state space of  $X$ ?

[1pts.]

- A.  $S_X = \{0, 1, 2\}$
- B.  $S_X = [0, 2]$
- C.  $S_X = \{1, 2\}$
- D.  $S_X = [1, 2]$
- E. None of the above

**Problem 17.** What is  $P(X = 1.5)$ ? [1pts.]

- A. 0.00
- B. 0.25
- C. 0.50
- D. 0.75
- E. None of the above.

**Problem 18.** What is  $P(X \geq 1.5)$ ? [1pts.]

- A. 0.00
- B. 0.25
- C. 0.50
- D. 0.75
- E. None of the above.

.....

**Problems 19 - 20 refer to the following situation:** Suppose Nitin has imported the `scipy.stats` module with the nickname `sps`, and has also run the following code:

```
a = sps.t.ppf(0.3, 27)
b = sps.t.ppf(0.7, 27)

c = sps.t.cdf(-1.31, 27)
```

---

**Problem 19.** What is the correct relationship between  $a$  and  $b$ ? [1pts.]

- A.  $a = b$
- B.  $a = -b$
- C.  $a = 1 - b$
- D.  $b = 1 - a$
- E. None of the above.

**Problem 20.** What is the value of  $c$ ? [1pts.]

- A.  $-1.31$
- B.  $0.10$
- C.  $0.20$
- D.  $1.31$
- E. None of the above.

.....

**Problems 21 - 23 refer to the following situation:** Karla wants to know whether regular exercise has an effect on overall mental health.

**Problem 21.** Which of the following schemes describes how Karla could conduct an **observational study** to achieve her goal? [1pts.]

- A. Take a sample of 100 volunteers and divide them into two groups. To one group, prescribe regular exercise and to the other prescribe no exercise. Instruct groups to continue for a period of several weeks, and then record mental health at the end of the several weeks.
- B. Take a sample of 100 volunteers, 50 of which already regularly exercise and 50 of which do not regularly exercise. Observe these 100 individuals over a period of a few weeks and then record the mental health of each group at the end of the several weeks.
- C. Take a sample of 100 volunteers that do not regularly exercise, and start by recording the initial mental health of these 100 volunteers. Then, prescribe regular exercise to these volunteers for a period of several weeks, and then record the post-treatment mental health of the volunteers.

**Problem 22.** Suppose Karla has performed her observational study, and found that there is a statistically significant relationship between exercise and mental health; specifically, it seems that more regular exercise is associated with improved mental health. Can Karla then conclude that exercising regularly causes an improvement in mental health? [1pts.]

- A. Yes, Karla is justified in making a causal assertion.
- B. No, because it is not possible to make causal assertions using an observational study.
- C. No, because there may be confounding variables Karla has not controlled for.
- D. Both choices (B) and (C).
- E. None of the above.

**Problem 23.** Suppose Karla has performed her study in the following way: [1pts.]

Take a sample of 100 volunteers that do not regularly exercise, and start by recording the initial mental health of these 100 volunteers. Then, prescribe regular exercise to these volunteers for a period of several weeks, and then record the post-treatment mental health of the volunteers.

Has Karla performed a longitudinal study or a cross-sectional study?

- A. Longitudinal
- B. Cross-Sectional

.....



Problems 24 - 30 are unrelated.

**Problem 24.** Which of the options below gives the correct LaTeX syntax for rendering the following equation (pay attention to the parentheses and exponents!) [1pts.]

$$f_X(x) = \left(\frac{\pi}{x}\right)^{-4}$$

- A. `$$ f_X(x) = \left( \frac{\pi}{x} \right)^{-4} $$`
- B. `$$ f_X(x) = ( \frac{\pi}{x} )^{-4} $$`
- C. `$$ f_X(x) = \left( \frac{\pi}{x} \right)^{-4} $$`
- D. `$$ f_X(x) = ( \frac{\pi}{x} )^{-4} $$`
- E. None of the above.

**Problem 25.** Consider the function  $g()$ , defined as follows: [1pts.]

```
def g(x):
    """
    return negative one times x
    """
    -1 * x
```

---

What will be returned by calling  $g(-1)$ ?

- A.  $-1$
- B.  $1$
- C. An Error
- D. Nothing
- E. None of the above.

**Problem 26.** When running the code  $y = y - 2$ , which side of the equality does Python evaluate first? [1pts.]

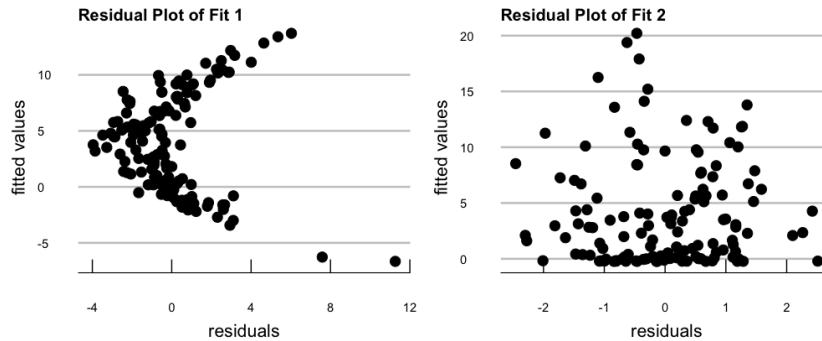
- A. Left
- B. Right

**Problem 27.** Let  $\pi_p$  denote the  $p^{\text{th}}$  percentile of the standard normal distribution for an arbitrary (but fixed) value of  $p$  that is strictly greater than 50%. Which of the following must be true? [1pts.]

- A.  $\pi_p < 0$
- B.  $\pi_p = 0$
- C.  $\pi_p > 0$
- D. None of the above.

**Problem 28.** A variable  $y$  is regressed onto another variable  $x$ . Two different fits are generated, called Fit 1 and Fit 2 respectively; the residual plots are displayed below. Which model is performing “better” (i.e. fitting the data better)?

[1pts.]



- A. Fit 1
- B. Fit 2

**Problem 29. True or False:** The right endpoint of the right whisker on a boxplot will always be the maximum value in the dataset.

[1pts.]

- A. True
- B. False

**Problem 30. True or False:** Variance is a measure of central tendency.

[1pts.]

- A. True
- B. False