Score: _____ / **32**

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:_____

MULTIPLE CHOICE QUESTIONS

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 **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.
- 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!!!

Problem 1. Which of the following is the correct LaTeX syntax used to generate the equation [1pts.]

$$f_X(x) = e^{-\frac{x^2}{2}}$$

A. `$$f_X(x) = e^-x^2/2$$`

B. `$$f_X(x) = e^-\frac{x^2}{2}$$`

C. `$$f_X(x) = e^{-\frac{x^2}{2}}$$`

D. `$$fX(x) = e^{-\frac{x^2}{2}}$$`

E. None of the above

Problem 2. Which of the following values is *not* a valid correlation value? [1pts.]

A. -0.50

B. 0.00

C. 0.75

D. 1.20

E. None of the above

Problem 3. Which of the following values is *not* a valid probability value? [1pts.]

A. -0.5

B. 0.0

C. 0.3

D. 1.0

E. None of the above

Problem 4. Suppose X is a discrete random variable. Which of the following **must** be true? [1pts.]

A. The state space of X consists only of integers.

B. The expected value of X is an element of its state space.

C. The probability that X attains any specific value is zero.

D. The variance of X is less than 1.

E. None of the above.

Problem 5. The time it takes Emily to complete her homework on any given night is uniformly distributed between 10 minutes and 60 minutes. What is the probability that it will take Emily exactly 25 minutes to complete her homework tonight? [1pts.]

A. 0.0000

B. 0.0200

C. 35.0000

D. 208.3333

E. None of the above.

Problem 6. When executing the statement `x = x + 2`, which side of the equality does Python evaluate first? [1pts.]

A. Left

B. Right

Problem 7. Recall that in the `palmerpenguins` dataset, the variable `flipper_length_mm` stores the length (in mm) of the penguins' flippers and `species` stores the species of the penguins. What is the correct type of plot we should use to visualize the relationship between `flipper_length_mm` and `species`? [1pts.]

A. Histogram

B. Side-by-side boxplot

C. Scatterplot

D. Barplot

E. QQ-plot

Problem 8. If $Y \sim \mathcal{N}(2,\ 1.5)$, what is $\mathbb{P}(0.5 \leq Y \leq 1.5)$? [1pts.]

A. 0.0000

B. 0.1587

C. 0.2120

D. 0.3707

E. None of the above.

Problems 9 - 16 refer to the following situtation: A study published by *Gallup* claimed that 48% of Americans regularly drink soda. Suppose we wish to test *Gallup*'s claims against a two-sided alternative, at an $\alpha = 0.05$ level of significance. To that end, we take a representative sample of 75 Americans and note that 52% of these individuals regularly drink soda.

Problem 9. What is the parameter of interest, p? [1pts.]

- **A.** p =the true proportion of Americans that regularly drink soda.
- **B.** p =the true number of Americans that regularly drink soda.
- **C.** p =the proportion of people in a representative sample of 75 Americans that regularly drink soda.
- **D.** None of the above.

Problem 10. What is the random variable of interest, $\widehat{P}$? [1pts.]

- **A.** $\widehat{P}$ =the true proportion of Americans that regularly drink soda.
- **B.** $\widehat{P}$ =the proportion of people in a representative sample of 75 Americans that regularly drink soda.
- **C.** $\widehat{P}$ =the true number of Americans that regularly drink soda.
- **D.** None of the above.

Problem 11. Which of the following are the conditions we need to check to determine the distribution of $\widehat{P}$, under the null? **Select ALL that apply;** incorrect choices incur a deduction of 0.5pts (capped out at zero; i.e. you will never receive negative points for this question.) [2pts.]

- **A.** $n \geq 30$
- **B.** $n \geq 10$
- **C.** $np_0 \geq 10$
- **D.** $n(1-p_0) \geq 10$
- **E.** $n(1-n) \geq 10$

Problem 12. What are the null and alternative hypotheses? [1pts.]

- **A.** $H_0 = 0.48; \quad H_A \neq 0.48$
- **B.** $H_0 = 0.48; \quad H_A < 0.48$
- **C.** $H_0 : \ p = 0.48; \quad H_A : \ p \neq 0.48$
- **D.** $H_0 : \ p = 0.48; \quad H_A : \ p < 0.48$
- **E.** None of the above

Problem 13. What formula do we use to compute the test statistic in this problem? [1pts.]

A. $\text{TS} = \dfrac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}$

B. $\text{TS} = \dfrac{\overline{X} - \mu_0}{s/\sqrt{n}}$

C. $\text{TS} = \dfrac{\widehat{P} - p_0}{\sqrt{\frac{p_0 \cdot (1-p_0)}{n}}}$

D. None of the above

Problem 14. Suppose a **new** dataset (still of 75 Americans) yielded a test statistic of -0.67. What is the $p-$value of this test statistic? [1pts.]

A. 0.2514

B. 0.5028

C. 0.7486

D. 0.9813

E. None of the above

Problem 15. Suppose a **yet another** dataset (still of 75 Americans) yielded a test statistic with a $p-$value of 0.02. Based on this $p-$value, would we reject or fail to reject the null hypothesis? [1pts.]

A. Reject the null

B. Fail to Reject the null

Problem 16. If we reject the null, which of the following would be a fully correct statement of our conclusions in the context of the problem? [1pts.]

A. The true proportion of Americans that regularly drink fast food is not 48%.

B. We reject the null hypothesis that the true proportion of Americans that regularly drink fast food is 48%.

C. At an $\alpha = 0.05$ level of significance, there was sufficient evidence to reject hypothesis that the true proportion of Americans that regularly drink fast food is 48%.

D. At an $\alpha = 0.05$ level of significance, there was sufficient evidence to reject hypothesis that the true proportion of Americans that regularly drink fast food is 48% in favor of the alternative that the true proportion is not 48%.

E. At an $\alpha = 0.05$ level of significance, there was sufficient evidence to reject hypothesis that the true proportion of Americans that regularly drink fast food is 48% in favor of the alternative that the true proportion is less than 48%.

Problems 17 - 19 refer to the following situtation: The temperature at a randomly-selected location in Santa Barbara follows a normal distribution with mean 67°F and standard deviation 10°F. A location is selected at random (within Santa Barbara), and the temperature at this location is recorded.

Problem 17. What is the random variable of interest? [1pts.]

A. $X =$ the temperature.

B. $X =$ the temperature at a randomly selected location within Santa Barbara.

C. $X =$ a randomly selected location in Santa Barbara.

D. $X =$ the average temperature in Santa Barbara.

E. None of the above.

Problem 18. What is the correct notation for the distribution of the random variable of interest? [1pts.]

A. $X \sim \text{Unif}(67,\ 10)$

B. $X \sim \text{Unif}(10,\ 67)$

C. $X \sim \mathcal{N}(67,\ 10)$

D. $X \sim \mathcal{N}(10,\ 67)$

E. None of the above

Problem 19. The temperature at Yuna's house is at the 52^{nd} percentile of temperatures in Santa Barbara. What line of code would correctly output the temperature at Yuna's house? [1pts.]

A. `scipy.stats.norm.cdf(0.52)`

B. `scipy.stats.norm.cdf(0.52, 67, 10)`

C. `scipy.stats.norm.ppf(0.52)`

D. `scipy.stats.norm.ppf(0.52, 67, 10)`

E. None of the above.

Problems 20 - 26 refer to the following situtation: An ANOVA (Analysis of Variance) has been performed on k groups. The resulting ANOVA table is shown below, but has certain entries redacted.

	DF	Sum Sq.	Mean Sq.	$F-$value	$\mathbb{P}(>F)$
Btwn. Grps.	4	36	9	`<blank2>`	0.022
Residuals	100	`<blank1>`	3		

Problem 20. What is k, the number of groups? [1pts.]

A. 3

B. 4

C. 5

D. 6

E. None of the above.

Problem 21. What is n, the total number of observations (aggregated across all groups)? [1pts.]

A. 99

B. 100

C. 104

D. 105

E. None of the above

Problem 22. What is the value of `<blank1>`? [1pts.]

A. 33.3333

B. 36.0000

C. 100.0000

D. 300.0000

E. None of the above

Problem 23. What is the value of `<blank2>`? [1pts.]

A. 33.3333

B. 36.0000

C. 100.0000

D. 300.0000

E. None of the above

Problem 24. Letting $\mu_1, \cdots, \mu_k$ denote the k group means (i.e. the true population mean of each group), what is the null hypothesis being tested? [1pts.]

A. $H_0 : \mu_1 = \mu_2 = \cdots = \mu_k$

B. $H_0 : \mu_1 < \mu_2 < \cdots < \mu_k$

C. $H_0 : \mu_1 > \mu_2 > \cdots > \mu_k$

D. $H_0 : \mu_1 \neq \mu_2 \neq \cdots \neq \mu_k$

E. None of the above.

Problem 25. Again letting $\mu_1, \cdots, \mu_k$ denote the k group means (i.e. the true population mean of each group), what is the alternative hypothesis being tested? [1pts.]

A. $H_A : \mu_1 = \mu_2 = \cdots = \mu_k$

B. $H_A : \mu_1 < mu_2 < \cdots < \mu_k$

C. $H_A : \mu_1 > \mu_2 > \cdots > \mu_k$

D. $H_A : \mu_1 \neq \mu_2 \neq \cdots \neq \mu_k$

E. None of the above.

Problem 26. For which of the following levels of significance would we **reject** the null? **Select ALL that apply;** incorrect choices incur a deduction of 0.5pts (capped out at zero; i.e. you will never receive negative points for this question.) [2pts.]

A. $\alpha = 0.01$

B. $\alpha = 0.05$

C. $\alpha = 0.10$

D. $\alpha = 0.50$

Problems 27 - 30 refer to the following situtation: Yasmina wants to write a function called `my_function()` that takes in two lists `x = [x1, x2, ... , xn]` and `y = [y1, y2, ... , yn]`. The function is meant to output one of two things:

- If `x` and `y` have the same number of elements, the function should return

$$\sum_{i=1}^{n} \sqrt{x_i \cdot y_i}$$

- If `x` and `y` do not have the same number of elements, the function should return the string

```
Inputs must have the same length!
```

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

```
def my_function(x, y):

    if len(x)  __Blank 1__  len(y):

        return "Inputs must have the same length!"

    __Blank 2__:

        return  __Blank 3__  .sqrt(sum(x * y))
```

Problem 27. What should go in Blank 1? [1pts.]

A. `=`

B. `==`

C. `*=`

D. `!=`

E. None of the above.

Problem 28. What should go in Blank 2? [1pts.]

A. **`else`**

B. **`else if`**

C. **`elif`**

D. **`e_if`**

E. None of the above.

Problem 29. What should go in Blank 3? [1pts.]

A. `numpy`

B. `np`

C. `base`

D. `math`

E. None of the above.

Problem 30. What is missing from the body of Yasmina'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.

. . . . . .