

PSTAT 5A / MIDTERM EXAM 2 / Spring 2023

Instructor: Ethan Marzban

Name: _____
First, then Last

UCSB NetID: _____
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: _____

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 × 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, 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.

× _____

Problem 1. A 97% confidence interval for some population parameter θ has been constructed using a particular dataset. Would we expect the width of a 72% confidence interval for θ , constructed using the same dataset, to be wider than or narrower than the width of the 97% confidence interval? [1pts.]

- Wider
- Narrower**
- The same width

Problem 2. 500 fair six-sided dice are rolled independently of each other, and the outcomes are recorded. What is the expected number of dice (out of these 500) that land on the number 1? [1pts.]

- 0.1667
- 1.0000
- 250.0000
- 83.3333
- None of the above.

Problem 3. Suppose 34% of all residents in a particular country live in a two-story home. In a representative sample of 300 residents, 30% are found to live in a two-story home. **Select the answer choice** that best fills in the blanks in the following sentence: [1pts.]

The value of 34% is a (Blank 1) , and the value of 30% is a (Blank 2) .

- Blank 1:** population parameter; **Blank 2:** population parameter
- Blank 1: population parameter; Blank 2: sample statistic**
- Blank 1:** sample statistic; **Blank 2:** population parameter
- Blank 1:** sample statistic; **Blank 2:** sample statistic

Problem 4. Does the expected value of a random variable have to be an element of the random variable's state space? [1pts.]

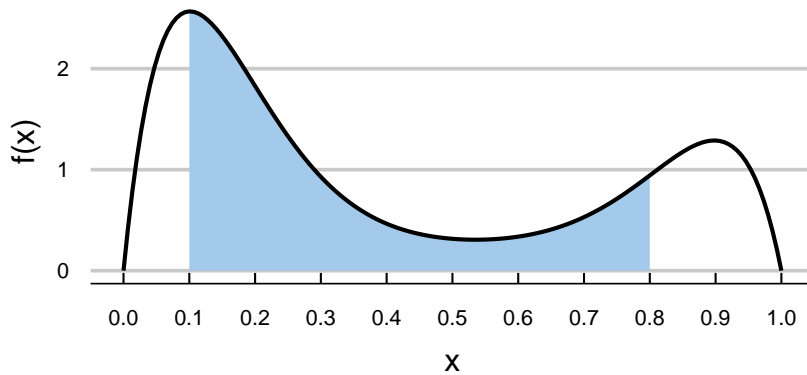
- Yes
- No**

Problem 5. What module contains the function `t.ppf()`? [1pts.]

- numpy
- datascience
- scipy.stats**
- matplotlib.pyplot
- None of the above

Problem 6. Consider a random variable X with the following density curve:

[1pts.]



What is the probability that is shaded in the figure above? (If the figure is hard to read: the region shaded is the region underneath the density curve between $x = 0.1$ and $x = 0.8$.)

- $P(X \leq 0.1) - P(X \leq 0.8)$
- $P(X \leq 0.8) - P(X \leq 0.1)$
- $P(X \geq 0.1) - P(X \leq 0.8)$
- $P(X \leq 0.1) - P(X \geq 0.8)$
- None of the above.

Problem 7. The time it takes Johnny to bike from his apartment to UCSB varies uniformly between 18.3 minutes and 23.5 minutes. What is the probability that Johnny will take exactly 20 minutes to bike from his apartment to UCSB today?

[1pts.]

- 0.0000
- 0.1923
- 0.3269
- 0.6730
- None of the above.

Problem 8. What is the correct syntax to import the `numpy` module with the nickname `np`?

[1pts.]

- `import numpy as np`
- `import numpy with name np`
- `import .numpy.np`
- `import numpy with nickname np`

Problem 9. Karla would like to compute the standard deviation of the list of numbers [1, 4, 3, 6, 10]. To do so, she has imported the `numpy` module (not with any nickname), and is going to run `std([1, 4, 3, 6, 10])`. Will this give her the desired result? As a reminder, the formula we use in this class to compute standard deviation is

[1pts.]

$$\sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

- Yes, it will give her the desired result.
- No, because she needs to pass in the argument `ddof = 1` to her call to `std()`.**
- No, because she has copied down the wrong list of numbers.
- None of the above.

Problem 10. Consider the following code cell:

[1pts.]

```
y = 0

for k in np.arange(0, 10):
    y = k - 1
```

What will be the value of `y` after running the above code cell once?

- 5
- 6
- 7
- 8**
- 9
- 10
- None of the above

Free Response Questions

Problem 11. The duration of a randomly selected flight out of SBA (Santa Barbara Airport) is uniformly distributed between 50 minutes and 200 minutes.

(a) Define the random variable of interest, and call it X .

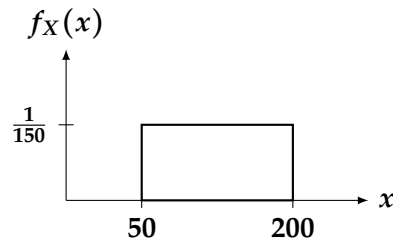
[1pts.]

Solution: Let X denote the duration of a randomly selected flight out of SBA.

(b) Sketch the density curve of X . Make sure you clearly label any points of importance on your graph.

[2pts.]

Solution:



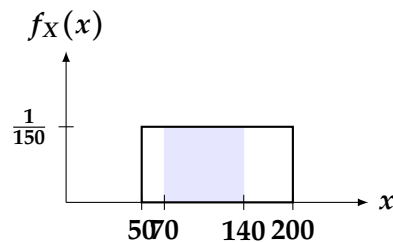
where we obtained the height as

$$\frac{1}{b-a} = \frac{1}{200-50} = \frac{1}{150}$$

(c) What is the probability that a randomly selected flight will have duration between 70 and 140 minutes?

[2pts.]

Solution: As an area, the probability we seek is



This is a rectangle with base $(140 - 70) = 70$ and height $1/150$, meaning the area (and, consequently, the desired probability) is

$$(140 - 70) \cdot \frac{1}{150} = \frac{70}{150} = \frac{7}{15}$$

- (d) A sample of 10 flights is taken (with replacement). What is the probability that this sample contains exactly 4 flights with duration between 70 and 140 minutes?

[4pts.]

Solution: Let Y denote the number of flights, in a sample of 10 flights taken with replacement, that have duration between 70 and 140 minutes. We surmise that Y is binomially distributed- to verify, we check the 4 binomial criteria:

- 1) **Independent trials?** Yes, since sampling is done with replacement.
- 2) **Fixed number of trials?** Yes; $n = 10$.
- 3) **Well-defined notion of success?** Yes; "success" = "flight has duration between 70 and 140 minutes"
- 4) **Fixed probability of success?** Yes; $p = 7/15$, as found in part (c) above.

Hence, we conclude $Y \sim \text{Bin}(10, 7/15)$ and so

$$\mathbb{P}(Y = 4) = \binom{10}{4} \left(\frac{7}{15}\right)^4 \left(\frac{8}{15}\right)^6 = 0.2292$$

Problem 12. An electrician collects a representative sample of 45 *GachoBuzz*-brand batteries and records their average lifetime (in years). The average lifetime of her sample of 45 batteries is 1.92 years, and the standard deviation of lifetimes in her sample is 0.5 years.

- (a) Define the random variable of interest, \bar{X} .

[1pts.]

Solution: Let \bar{X} denote the average lifetime of a sample of 45 *GachoBuzz*-brand batteries.

- (b) What distribution would we use to construct a 95% confidence interval for the true average lifetime of a *GaUCHoBUZZ*-brand battery? Be sure to include any/all relevant parameter(s), and check any/all conditions! [3pts.]

Solution:

- Is the population normally distributed? No.
- Is the sample size at least 30? Yes; $n = 45 \geq 30$.
- Do we have access to σ ? No, we only have access to s , the sample standard deviation.

Based on the answers to these questions, we know we need to use the t -distribution; specifically, with $n - 1 = 45 - 1 = 44$ degrees of freedom. That is, we need to use the t_{44} distribution.

- (c) Construct a 95% confidence interval for the true average lifetime of a *GauchoBUZZ*-brand battery, and interpret your interval. [3pts.]

Solution: Our confidence interval will be of the form

$$\bar{x} \pm t_{44, \alpha} \cdot \frac{s}{\sqrt{45}}$$

where $t_{44, \alpha}$ denotes the appropriate percentile of the t_{44} distribution. Since we are using a 95% confidence level, we need to find the

$$\frac{1 - 0.95}{2} \times 100 = 2.5^{\text{th}}$$

percentile of the t_{44} distribution. From our table, we see that this value is 2.02, meaning our confidence interval is

$$1.92 \pm 2.02 \cdot \frac{0.5}{\sqrt{45}} = 1.92 \pm 0.151 = [1.769, 2.071]$$

The interpretation of this interval is:

We are 95% confident that the true average lifetime of a *GaUCHoBUZZ*-brand battery is between 1.769 and 2.071 years.

Problem 13. Cillian, the owner of *Cillian's Ice Cream*, believes that 77% of his customers pay with credit card. To test his suspicion, he takes a representative sample of 170 customers and finds that 74% of these customers pay with credit card. With this data, he wishes to conduct a two-sided hypothesis test at an $\alpha = 0.05$ level of significance.

- (a) Write down the null and alternative hypotheses for this test. Be sure to use mathematical notation, and define any variables you use. [2pts.]

Solution: Letting p denote the true proportion of customers that pay with credit card, we phrase our hypotheses as

$$\begin{cases} H_0 : p = 0.77 \\ H_A : p \neq 0.77 \end{cases}$$

- (b) Compute the value of the test statistic. [3pts.]

Solution: Recall that if our null hypothesis is $H_0 : p = p_0$, we take

$$TS = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

For this problem, we take $p_0 = 0.77$, $\hat{p} = 0.74$, and $n = 170$. Hence,

$$TS = \frac{0.74 - 0.77}{\sqrt{\frac{0.77 \cdot (1-0.77)}{170}}} = -0.929$$

- (c) What is the distribution of the test statistic, assuming the null hypothesis is correct? Be sure to check any relevant conditions, and be sure to specify any/all necessary parameter(s) in your final answer! [2pts.]

Solution: We must check the success-failure conditions, with p_0 in place of p :

1) $np_0 = (170)(0.77) = 130.9 \geq 10 \checkmark$

2) $n(1 - p_0) = (170)(1 - 0.77) = 39.1 \geq 10 \checkmark$

Since both conditions are met, we conclude

$$TS = \frac{\hat{P} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \stackrel{H_0}{\sim} \mathcal{N}(0, 1)$$

- (d) Find the critical value of the test. [2pts.]

Solution: Because we are using an $\alpha = 0.05$ level of significance, the critical value will be the $(0.05/2) \times 100 = 2.5^{\text{th}}$ percentile of the standard normal distribution, scaled by negative 1. From a lookup table, we see that the critical value is therefore **1.96**.

- (e) Now, conduct the hypothesis test and state your conclusions in the context of the problem. [2pts.]

Solution: Our test rejects H_0 when $|TS|$ is larger than the critical value. By our results from previous parts, we have that $|TS| = |-0.929| = 0.929 \not\geq 1.96$, meaning we fail to reject the null; i.e.

At an $\alpha = 0.05$ level of significance, there is insufficient evidence to reject Cillian's hypothesis that 77% of his customers pay with credit card.

Problem 14. A random variable X has the following probability mass function (p.m.f):

k	-1	0.5	2
$\mathbb{P}(X = k)$	0.1	a	0.5

(a) What is the state space S_X of X ? [1pts.]

Solution: The state space is simply the first row of the PMF table:
 $S_X = \{-1, 0.5, 2\}$.

(b) Is X a discrete random variable or a continuous random variable? Justify your answer. [1pts.]

Solution: Because the state space S_X has jumps, we conclude that X is a discrete random variable.

(c) Find the value of a . [2pts.]

Solution: We know that the probability values in a PMF must sum to 1; that is,

$$0.1 + a + 0.5 = 1 \implies a = 1 - (0.1 + 0.5) = 0.4$$

(d) Compute $\mathbb{P}(X \leq 1)$. [2pts.]

Solution: $\mathbb{P}(X \leq 1) = \mathbb{P}(X = -1) + \mathbb{P}(X = 0) = 0.1 + 0.4 = 0.5$.
Alternatively:

$$\mathbb{P}(X \leq 1) = 1 - \mathbb{P}(X > 1) = 1 - \mathbb{P}(X = 2) = 1 - 0.5 = 0.5$$

- (e) Compute $\mathbb{E}[X]$, the expected value of X . (Make sure you plug in the value for a you found in part (d) above.) [3pts.]

Solution:

$$\begin{aligned}\mathbb{E}[X] &= \sum_{\text{all } k} k \cdot \mathbb{P}(X = k) \\ &= (-1) \cdot \mathbb{P}(X = -1) + (0.5) \cdot \mathbb{P}(X = 0) + (2) \cdot \mathbb{P}(X = 2) \\ &= (-1) \cdot (0.1) + (0.5) \cdot (0.4) + (2) \cdot (0.5) = \mathbf{1.1}\end{aligned}$$

- (f) Compute $\text{Var}(X)$, the variance of X . (Make sure you plug in the value for a you found in part (d) above.) [4pts.]

Solution: If we use the second formula for variance, we first compute

$$\begin{aligned}\sum_{\text{all } k} k^2 \cdot \mathbb{P}(X = k) &= (-1)^2 \cdot \mathbb{P}(X = -1) + (0.5)^2 \cdot \mathbb{P}(X = 0) + (2)^2 \cdot \mathbb{P}(X = 2) \\ &= (-1)^2 \cdot (0.1) + (0.5)^2 \cdot (0.4) + (2)^2 \cdot (0.5) = 2.2\end{aligned}$$

meaning

$$\text{Var}(X) = \sum_{\text{all } k} k^2 \cdot \mathbb{P}(X = k) - (\mathbb{E}[X])^2 = 2.2 - (1.1)^2 = \mathbf{0.99}$$

Alternatively, we could have used the first formula for variance:

$$\begin{aligned}\text{Var}(X) &= \sum_{\text{all } k} (k - \mathbb{E}[X])^2 \cdot \mathbb{P}(X = k) \\ &= (-1 - 1.1)^2 \cdot \mathbb{P}(X = -1) + (0.5 - 1.1)^2 \cdot \mathbb{P}(X = 0) + (2 - 1.1)^2 \cdot \mathbb{P}(X = 2) \\ &= (-1 - 1.1)^2 \cdot (0.1) + (0.5 - 1.1)^2 \cdot (0.4) + (2 - 1.1)^2 \cdot (0.5) = \mathbf{0.99}\end{aligned}$$