Midterm 1: Practice Problems

PSTAT 5A: Spring 2023, with Ethan P. Marzban

! Disclaimer

- This is not meant to be comprehensive, nor is this meant to directly emulate the format of the actual exam.
  - Rather, these problems are designed to *supplement* your already-existing notes, lecture examples, homeworks, labs, and quiz.

Multiple Choice Questions

1. Given two events A and B with $\mathbb{P}(A) = 0.5$, $\mathbb{P}(B) = 0.7$, is it possible for A and B to be disjoint?
   a. Yes
   b. No
   c. Not enough information to determine

   Solution: The answer is (b); if the events A and B were disjoint then $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) = 0.5 + 0.7 = 1.2$ which is a contradiction as probabilities must be less than 1.

2. In what module is the function `make_array()` found?
   a. `numpy`
   b. `math`
   c. `datascience`
   d. `python_arrays`
   e. None of the above

   Solution: The answer is (c).

3. What do we call the verbal description of a function included in the definition of a function, often enclosed in three quotation marks?
   a. functional description
   b. docstring
   c. user guide
   d. vignette
   e. None of the above

Solution: The answer is (b).

4. Which of the following expressions will, when run, return a value of `True`?
   a. `1 > 4`
   b. `2 + 2`
   c. `'a' < 'b'`
   d. `'b' < 'b'`
   e. None of the above

Solution: The answer is (c).

5. Masha records the weight (in grams) of 294 different baby goats. What is the correct classification of the variable containing the weights of these goats?
   a. discrete
   b. continuous
   c. nominal
   d. ordinal
   e. None of the above

Solution: The answer is (b).

6. Morgan collects 78 pieces of fruit and records both the type of fruit (`apple`, `orange`, etc.), along with its weight (in grams). What is the correct type of visualization Morgan should generate if they want to determine the relationship between fruit type and fruit weight?
   a. Boxplot
   b. Histogram
   c. Bargraph
   d. Side-by-side Boxplot
   e. Scatterplot
   f. None of the above

Solution: The answer is (d).

7. Given two events E and F, which of the following statements must be true?
   a. $\mathbb{P}(E \cap F) = \mathbb{P}(E \mid F) \cdot \mathbb{P}(F)$
   b. $\mathbb{P}(E \cap F) = \mathbb{P}(E) \cdot \mathbb{P}(F)$
   c. $\mathbb{P}(E \cup F) = \mathbb{P}(E) + \mathbb{P}(F)$
   d. $\mathbb{P}(E \cup F) = 1 - \mathbb{P}(E^C \cup F^C)$

e. None of the above.

Solution: The answer is (a), which is the Multiplication Rule. Choice (b) is correct only when the events are independent; choice (c) is only correct when the events are mutually exclusive, and choice (d) is seldom correct.

8. If E and F are two events such that $\mathbb{P}(E) = 0.3$, $\mathbb{P}(F) = 0.2$, and $\mathbb{P}(E \cap F) = 0.1$, which of the statements is true?
   a. E and F are disjoint, but not independent
   b. E and F are independent, but not disjoint
   c. E and F are independent and disjoint
   d. E and F are neither independent nor disjoint

Solution: The answer is (d).

9. How do we add multi-line comments to a Jupyter notebook code cell?
   a. Using a hashtag (#)
   b. Using three qutoation marks (""")
   c. Using a percent sign (%)
   d. Using a backslash (\)
   e. None of the above

Solution: The answer is (b), as was discussed in Lab01 and on Quiz01.

Free Response Questions

Problem 1: Political Opinions

A group of political scientists have surveyed several individuals in downtown Santa Barbara and asked participants their political affiliation (Democrat, Republican, Independent) and whether or not they support an upcoming Bill. The results of the survey are included in the contingency table below:

```
         Affiliation
Support   Democrat Independent Republican
  Oppose        10           3          7
  Support       10           5         15
```

a. If a person from the survey is selected at random, what is the probability that they are a Democrat?

Solution: Let D denote the event "a randomly selected person is a Democrat." By the classical approach to probability, $\mathbb{P}(D)$ is simply the number of Democrats divided by the total number of surveyed people:

$$\mathbb{P}(D) = \frac{10 + 10}{50} = \frac{2}{5} = \boxed{40\%}$$

b. If a person from the survey is selected at random, what is the probability that they oppose the bill?

Solution: Let O denote the event "a randomly selected person opposes the bill." By the classical approach to probability, $\mathbb{P}(O)$ is simply the number of people who oppose the bill divided by the total number of surveyed people:

$$\mathbb{P}(O) = \frac{10 + 3 + 7}{50} = \frac{2}{5} = \boxed{40\%}$$

c. If a person from the survey is selected at random, what is the probability that they are an Independent that supports the bill?

Solution: Let I denote the event "a randomly selected person is an Independent" and S denote "a randomly selected person supports the bill". We therefore compute

$$\mathbb{P}(I \cap D) = \frac{5}{50} = \boxed{10\%}$$

d. If a person from the survey is selected at random, what is the probability that they are either an Independent or they oppose the bill (or both)?

Solution: We seek $\mathbb{P}(I \cup O)$; by the Addition Rule, this is computed as

$$\mathbb{P}(I \cup O) = \mathbb{P}(I) + \mathbb{P}(O) - \mathbb{P}(I \cap O) = \frac{3+5}{50} + \frac{10+3+7}{50} - \frac{3}{50} = \boxed{50\%}$$

e. A person is selected at random, and it is noted that they oppose the bill. What is the probability that this individual is a Republican?

Solution: Let R denote "a randomly selected person is a republican". We seek $\mathbb{P}(R \mid O)$, which is computed as

$$\mathbb{P}(R \mid O) = \frac{\#(R \cap O)}{\#(O)} = \frac{7}{10+3+7} = \frac{7}{20} = \boxed{35\%}$$

f. Are 'being a Democrat' and 'supporting the bill' independent events? Justify your answer **mathematically**.

Solution: There are several ways to solve this problem. Recall that to check for independence, checking any of the following conditions will suffice:

- $\mathbb{P}(E \cap F) = \mathbb{P}(E) \cdot \mathbb{P}(F)$
- $\mathbb{P}(E \mid F) = \mathbb{P}(E)$
- $\mathbb{P}(F \mid E) = \mathbb{P}(F)$

I shall check the first condition. As such, we compute:

$$\begin{aligned}
\mathbb{P}(D) &= \frac{20}{50} = 40\% \\
\mathbb{P}(S) &= \frac{30}{50} = 60\% \\
\mathbb{P}(D \cap S) &= \frac{10}{50} = 20\% \\
&\neq (0.4) \cdot (0.6) = \mathbb{P}(D) \cdot \mathbb{P}(S)
\end{aligned}$$

Therefore, the events are not independent since $\mathbb{P}(D \cap S) \neq \mathbb{P}(D) \cdot \mathbb{P}(S)$.

Problem 2: Descriptive Statistics

Consider the set of numbers $X = \{-3, -1, 0, 4, 8\}$.

a. Compute $\overline{x}$, the mean of X.

Solution:

$$\overline{X} = \frac{1}{5}[(-3) + (-1) + (0) + (4) + (8)] = \boxed{1.6}$$

b. Compute median(X), the median of X.

Solution:

$$\cancel{-3} \quad \cancel{-1} \quad 0 \quad \cancel{4} \quad \cancel{8} \quad \Longrightarrow \quad \boxed{\text{median}(X) = 0}$$

c. Compute range(X), the range of X.

Solution:

$$\text{range}(X) = \max\{X\} - \min\{X\} = 8 - (-3) = \boxed{11}$$

d. Compute s_X, the standard deviation of X.

Solution: We first compute the variance:

$$\begin{aligned} s_X^2 &= \frac{1}{5-1}\left[(-3-1.6)^2 + (-1-1.6)^2 + (0-1.6)^2 + (4-1.6)^2 + (8-1.6)^2\right] \\ &= \frac{1}{4}(77.2) = 19.3 \quad \Longrightarrow \quad s_b = \boxed{\sqrt{32.1} \approx 4.393} \end{aligned}$$

Problem 3: Transformations

Consider a set $X = \{x_i\}_{i=1}^n$, and define $Y = \{x_i + a\}_{i=1}^n$ for some fixed constant a.

a. Show that $\overline{y} = \overline{x} + a$. (For extra practice, show this from scratch as opposed to citing previously-derived results).

Solution:

$$\begin{aligned} \overline{y} &= \frac{1}{n}(y_1 + y_2 + \cdots + y_n) \\ &= \frac{1}{n}[(x_1 + a) + (x_2 + a) + \cdots + (x_n + a)] \\ &= \frac{1}{n}\left[(x_1 + x_2 + \cdots + x_n) + \underbrace{(a + a + \cdots + a)}_{n \text{ terms}}\right] \\ &= \frac{1}{n}[(x_1 + x_2 + \cdots + x_n) + n \cdot a] \\ &= \underbrace{\frac{1}{n}(x_1 + x_2 + \cdots + x_n)}_{=\overline{x}} + \frac{1}{\cancel{n}} \cdot \cancel{n} \cdot a = \overline{x} + a \end{aligned}$$

b. How does s_Y^2 compare with s_X^2?

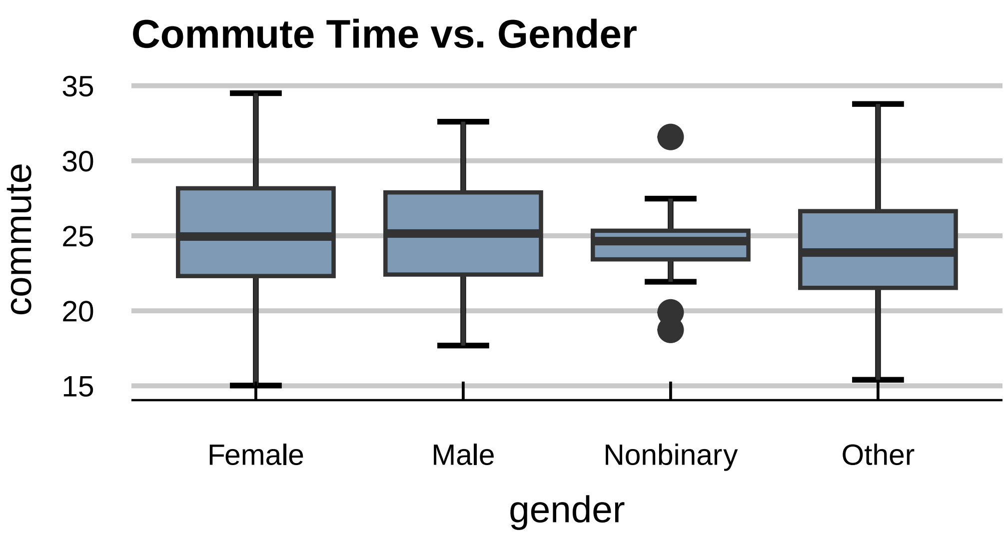
Solution:

$$\begin{aligned} s_Y^2 &= \frac{1}{n-1}\left[(y_1-\overline{y})^2+(y_2-\overline{y})^2+\cdots+(y_n-\overline{y})^2\right] \\ &= \frac{1}{n-1}\left[([x_1+a]-[\overline{x}+a])^2+([x_2+a]-[\overline{x}+a])^2+\cdots+([x_n+a]-[\overline{x}+a])^2\right] \\ &= \frac{1}{n-1}\left[(x_1+\cancel{a}-\overline{x}-\cancel{a})^2+(x_2+\cancel{a}-\overline{x}-\cancel{a})^2+\cdots+(x_n+\cancel{a}-\overline{x}-\cancel{a})^2\right] \\ &= \frac{1}{n-1}\left[(x_1-\overline{x})^2+(x_2-\overline{x})^2+\cdots+(x_n-\overline{x})^2\right] \\ &= s_X^2 \end{aligned}$$

Hence, the variance of X and Y are the same. Intuitively, this makes sense- if we take a set of numbers and shift every element by the same amount, the overall spread does not change!

Problem 4: Interpreting Plots

A sample of 135 UCSB students was taken and their gender (Male, Female, Nonbinary, Other) along with their commute time to campus (in minutes). The results of this study are summarized in the following graph:



a. Provide the five-number summary of commute times among Female respondents.

Solution:

min	Q_1	median	Q_3	max
15	22	25	27	35

b. Provide the range of commute times among Nonbinary respondents.

Solution:

$$32 - 17 = 15$$

c. Does there appear to a difference in commute times across genders?

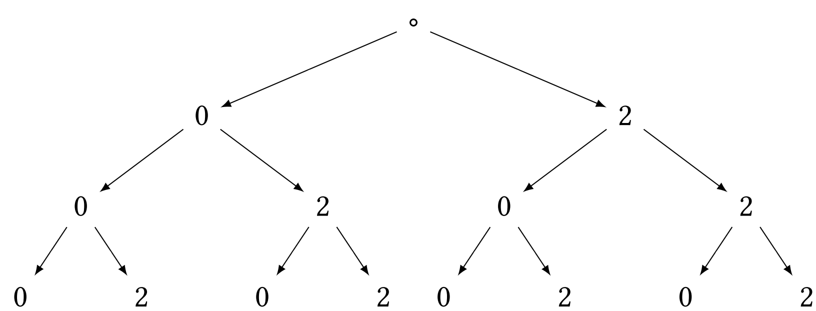
Solution: No, there does not as the commute times all appear to be centered around 25 minutes.

Problem 5: Picking Numbers

Suppose we pick three numbers from the set $\{0, 2\}$, replacing the number after each draw, and record the selected numbers.

a. Write down the outcome space of this experiment.

Solution:

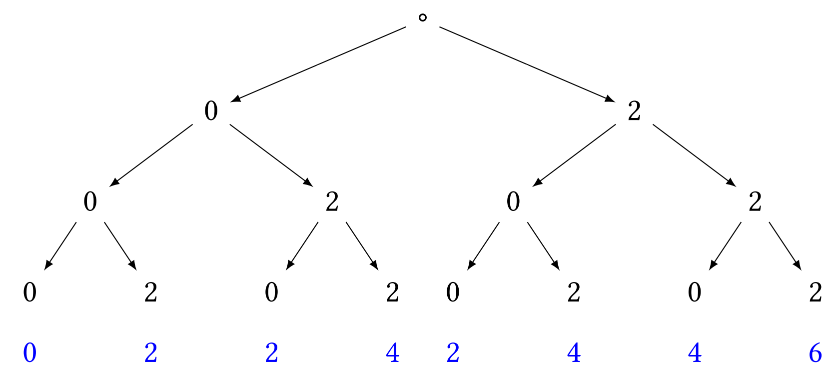


b. How many elements are in Ω?

Solution: There are 8 outcomes in Ω, as can be seen by either the tree above or using a slot diagram.

c. Let A be the event "the sum of the three numbers is 2". Write down the mathematical formulation of A; i.e. identify the outcomes that comprise A.

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



d. Are we justified in using the Classical Approach to probability? Why or why not.

> **Solution:** We are **not** justified in using the Classical Approach, since it is not stated whether the numbers are selected at random or not.

Problem 6: Events

Let E and F be two events with $\mathbb{P}(E) = 0.6,\ \mathbb{P}(F) = 0.5$, and $\mathbb{P}(E \cap F) = 0.12$.

a. What is the probability that either E or F occur?

> **Solution:** $\mathbb{P}(E \cup F) = \mathbb{P}(E) + \mathbb{P}(F) - \mathbb{P}(E \cap F) = 0.6 + 0.5 - 0.12 = 0.98$

b. What is the probability that neither E nor F occur?

> **Solution:** $\mathbb{P}(E^{\mathsf{C}} \cap F^{\mathsf{C}}) = 1 - \mathbb{P}(E \cup F) = 1 - 0.98 = 0.02$

c. What is $\mathbb{P}(E \mid F)$?

> **Solution:** $\mathbb{P}(E \mid F) = \dfrac{\mathbb{P}(E \cap F)}{\mathbb{P}(F)} = \dfrac{0.12}{0.5} = 24\%$

d. What is $\mathbb{P}(F \mid E)$?

> **Solution:** $\mathbb{P}(F \mid E) = \dfrac{\mathbb{P}(E \cap E)}{\mathbb{P}(E)} = \dfrac{0.12}{0.6} = 20\%$