Score: __________ / **35**

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

FREE RESPONSE QUESTIONS

Instructions:

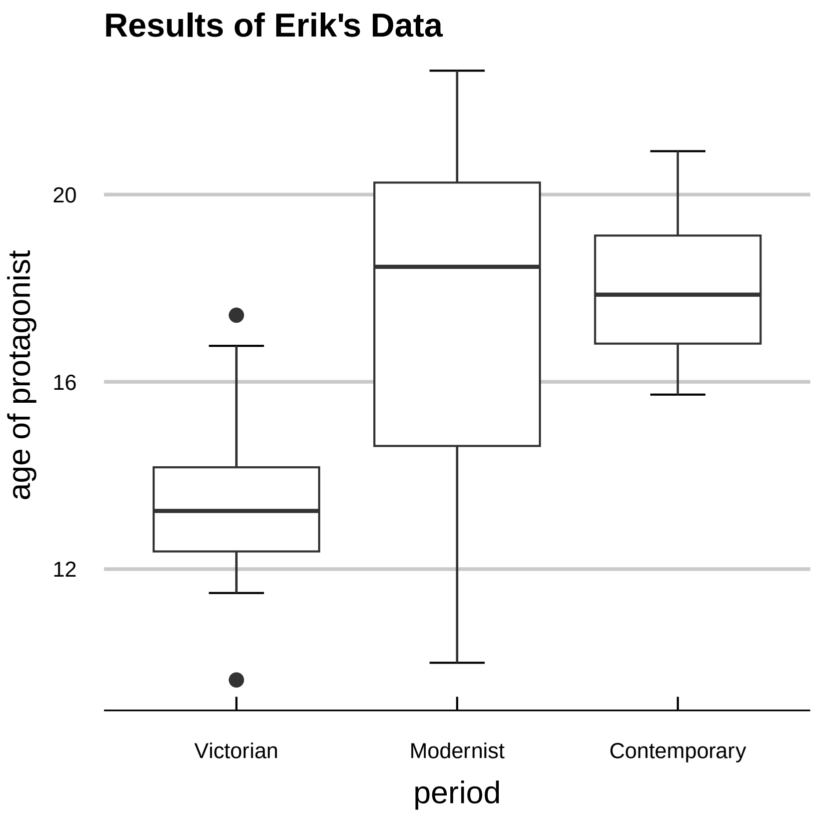
- You will have **75 minutes** to complete the entire exam
  - Do not begin working on the exam until instructed to do so.
  - During the final 10 minutes of the exam, we will ask everyone to remain seated until the exam concludes.
- This exam comes in **TWO PARTS**: this is the **FREE RESPONSE** part of the exam.
  - There is a separate booklet containing Multiple Choice questions that should have been distributed to you at the same time as this booklet.
- Write your answers directly in the space provided on this exam booklet.
  - You do not need to write anything on your scantron for this part of the exam.
- Be sure to show all of your work; correct answers with no supporting work will not receive full credit.
- 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!!!

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.

× ____

Free Response Questions

Problem 1. Erik decides to collect data on the ages of protagonists from pieces of fictional literature across three main historical periods: Victorian (19th century), Modernist (early 20th century), and Contemporary (late 20th century and early 21st century). The results of his study are displayed below, in graphical format:



(a) What is the median age of protagonists from Victorian-era literature included in Erik's data? [2pts.]

> **Solution:** From the boxplot of ages from the Victorian era, we see that the median age was approximately 13.

(b) Approximately what proportion of protagonists in Contemporary-era literature are <u>older</u> than 19 years old? [2pts.]

> **Solution:** From the boxplot of ages from the Contemporary era, we see that 19 is around he 75^{th} percentile of ages, meaning 75% of protagonists were *younger* than 19; i.e. 25% of protagonists were *older*.

(c) Are there any outliers in the data collected in any of the eras? If so, which eras contain outliers? How do you know? [2pts.]

> **Solution:** Outliers appear as points on either side of the reach of the whiskers; we see that only the Victorian era possessses such points, meaning it was the only era containing outliers.

(d) Erik's advisor believes that, over time, the average age of protagonists in literature has increased. Does Erik's data support this claim? Explain briefly. [2pts.]

> **Solution:** Answers may vary. It does seem as though the average age in Modernist and Contemporary literature is noticeably higher than that in Victorian literature, however it is a bit difficult to determine whether there was a significant change in average age between the Modernist and Contemporary eras.

Problem 2. A recent survey in Santa Barbara polled several people about their age, along with whether they prefer to communicate over phone or over email. The results of the survey are displayed below:

```
         Communication_Channel
Age        Email Phone
  18 - 24    30    20
  25 - 34    40    25
  35 - 44    25    30
```

A person is to be selected at random from the pool of people who participated in the survey.

(a) Are we justified in using the Classical Approach to probability? Justify your answer. [1pts.]

> **Solution:** Since the person is to be selected at random, we can use the Classical Approach to probability.

(b) What is the probability that the randomly-selected person was between 35 and 44 years old or preferred Email? [3pts.]

Solution: Let A denote the event "the person was between 35 and 44 years old" and B denote the event "they preferred email." We seek $\mathbb{P}(A \cup B)$, which we compute using the Addition Rule:

$$\begin{aligned}\mathbb{P}(A \cup B) &= \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) \\ &= \frac{25+30}{30+20+40+25+25+30} \\ &\quad + \frac{30+40+25}{30+20+40+25+25+30} \\ &\quad - \frac{25}{30+20+40+25+25+30} \\ &= \boxed{\frac{125}{170} \approx 73.6\%}\end{aligned}$$

(c) It is noted that the randomly-selected person preferred to communicate via Phone. What is the probability that they are between 25 and 34 years old? [3pts.]

Solution: Let C denote the event "the person prefers to communicate over Phone" and D denote the event "the person is between 25 and 34 years old." We seek $\mathbb{P}(D \mid C)$, which, since we are using the Classical Approach to Probability, is computed as

$$\mathbb{P}(D \mid C) = \frac{\#(D \cap C)}{\#(C)} = \frac{25}{20+25+30} = \boxed{\frac{25}{75} = \frac{1}{3} = 33.\overline{3}\%}$$

Problem 3. Two numbers are to be selected from the set $\{-1, 0, 1\}$. Suppose that the selection of numbers is done at random, and that the numbers are replaced after each selection (so the same number could be selected more than once). Additionally, suppose that the order in which the two numbers are selected <u>is</u> important. The two numbers are then recorded.

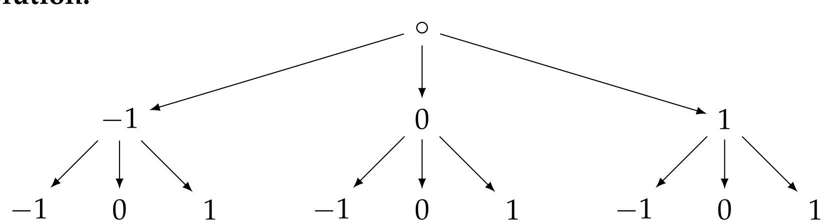
(a) Express the outcome space Ω for this experiment as a table. [2pts.]

Solution: Letting rows denote the first number selected and columns denote the second, we have

	-1	0	1
-1	$(-1, -1)$	$(-1, 0)$	$(-1, 1)$
0	$(0, -1)$	$(0, 0)$	$(0, 1)$
1	$(1, -1)$	$(1, 0)$	$(1, 1)$

(b) Express the outcome space Ω for this experiment as a tree. [2pts.]

Solution:

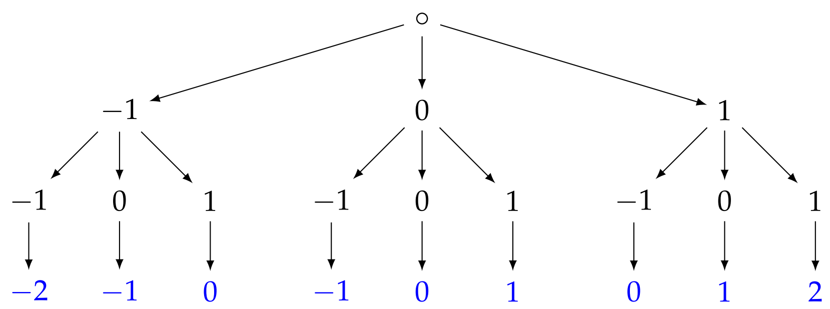


(c) How many outcomes are in Ω? Justify your answer. [2pts.]

Solution: We can find the number of elements in Ω by either summing up the entries in the table in part (a) or by counting the leaves on the tree in part (b). Either way, we find the number of elements in Ω to be 9.

(d) If A denotes the event "the sum of the two numbers is even", compute $\mathbb{P}(A)$ using the Classical Approach to Probability. (Recall that 0 is even, and that negative numbers can be even as well.) For full points, you should list out the elements in A. [4pts.]

Solution: It may be useful to append another row to our table, totaling the sum of the two numbers in each outcome:



This shows us that the sum of the two numbers will only ever be -2, -1, 0, 1, or 2. Of these numbers, only -2, 0, and 2 are even; hence,

$$A = \{\underbrace{(-1,\ -1)}_{\text{sum} = -2},\ \underbrace{(-1,\ 1),\ (0,\ 0),\ (1,\ -1)}_{\text{sum} = 0},\ \underbrace{(1,\ 1)}_{\text{sum} = 2}\}$$

which means A contains 5 elements and so

$$\mathbb{P}(A) = \frac{\#(A)}{\#(\Omega)} = \boxed{\frac{5}{9}}$$

Problem 4. Consider the list of numbers

$$X = \{-2,\ 0,\ 5\}$$

(a) Compute $\overline{x}$, the mean of X. [2pts.]

Solution:

$$\begin{aligned} \overline{x} &= \frac{1}{n}\sum_{i=1}^{n} x_i \\ &= \frac{1}{3}(-2+0+5) = \boxed{1} \end{aligned}$$

(b) Compute s_X^2, the variance of X. [3pts.]

Solution:

$$\begin{aligned} s_X^2 &= \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 \\ &= \frac{1}{3-1}\cdot\left[(-2-1)^2 + (0-1)^2 + (5-1)^2\right] = 13 \end{aligned}$$

(c) Compute s_X, the standard deviation of X. [1pts.]

Solution:

$$s_X = \sqrt{s_X^2} = \sqrt{13} \approx 3.61$$

(d) Compute the **sample skewness** of X, defined as [4pts.]

$$\widehat{\alpha}_3 := \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2}{\left[\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2}\right]^3}$$

Solution: The numerator can be computed as

$$\frac{1}{3}\left[(-2-1)^2 + (0-1)^2 + (5-1)^2\right] = \frac{26}{3}$$

and the denominator is actually just s_X^3, which, by part (b), is $13^{3/2}$. Hence,

$$\widehat{\alpha}_3 = \frac{\left(\frac{26}{3}\right)}{13^{3/2}} = \frac{26}{39\sqrt{13}} \approx 0.1849$$

An Aside: The formula for sample skewness actually has a typo in the way it is stated above: the true definition of sample skewness is actually

$$\widehat{\alpha}_3 := \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^3}{\left[\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2}\right]^3}$$

This does not affect the problem at all, though; as long as you used the formula that was provided to you, you'll be fine.