

Class Exercise 2 - Solutions

1. Multiple Choice

For each question below, select **all** the statements that are **correct**. Each question has **at least one correct answer, but not necessarily all options are correct**. You will receive **full credit** if and only if you select all correct answers and **no incorrect answers**. Selecting an incorrect option or missing a correct option may result in **partial credit or no credit**.

Solutions

Incorrect statement(s) are struck out

- A. Suppose the mean of a data set is μ .
- ~~If every data value is multiplied by 10, the new mean is $10 + \mu$.~~
 - If 7 is added to every data value, the new mean is $\mu + 7$.
 - ~~If every data value is divided by 2, the mean remains unchanged.~~
 - If every data value is multiplied by 10 and then 7 is added, the new mean is $10\mu + 7$.
- B. Which of the following statements about the **mean, median, and mode** are true?
- Adding the same constant to every data value increases the mean, median, and mode by that constant.
 - Multiplying every data value by a positive constant multiplies the mean and median by that constant.
 - Adding a new data value that is much larger than the current mean will increase the mean.
 - ~~Adding a new data value equal to 0 will always decrease the mean.~~
- C. Which of the following statements about a **weighted mean** are true?
- If all weights are equal, the weighted mean reduces to the ordinary mean.
 - ~~A value with zero weight affects the weighted mean.~~
 - Repeating a data value multiple times is equivalent to increasing its weight in a weighted mean.
 - Multiplying all weights by the same positive constant does not change the weighted mean.

2. ChatGPT Wedding Vows

Recently, a couple in the Netherlands has had their marriage invalidated in court after their officiant used ChatGPT to create their wedding vows, which neglected to include any legal declarations of marriage.

At a wedding registry office, officiants reviewed a sample of AI-generated wedding vows and recorded the number of legal declarations that were missing from each set of vows. The table below summarizes the results:

Number of Missing Legal Declarations	Number of Vows	
x_i	f_i	$x_i f_i$
0	6	0
1	8	8
2	22	44
3	38	114
4	56	224
	130	390

- (a) Calculate the average number of missing legal declarations per set of vows.

Solution

$$\begin{aligned}\bar{x} &= \frac{\sum x_i f_i}{\sum f_i} = \frac{(0 \times 6) + (1 \times 8) + (2 \times 22) + (3 \times 38) + (4 \times 56)}{130} \\ &= \frac{390}{130} \\ &= 3 \text{ missing legal declarations}\end{aligned}$$

- (b) What is the probability that a randomly selected set of vows is missing 2 or more legal declarations?

Solution

Let X = the number of missing legal declarations

$$P(X \geq 2) = \frac{22 + 38 + 56}{130} = \frac{116}{130} = 0.8923$$

- (c) Would a bar chart or a histogram be more appropriate for representing the number of omissions data? Explain your reasoning and describe the shape of the distribution.

Solution

\therefore The number of omissions is discrete random variable a **bar chart** would be the most appropriate graphical representation of the data. The distributions is unimodal, skewed-left.

The officiants were also asked how long they had been performing wedding ceremonies. Their responses are shown below:

Years of Experience	Number of Officiants		
	f_i	m_i	$f_i m_i$
[0, 5)	30	2.5	75.0
[5, 10)	15	7.5	112.5
[10, 15)	11	12.5	137.5
[15, 20)	13	17.5	227.5
[20, 25)	20	22.5	450.0
[25, 30)	31	27.5	852.5
	120		1855

- (d) Estimate the average number of years of experience for officiants in the study.

Solution

$$\begin{aligned}\bar{x} &= \frac{\sum f_i m_i}{\sum f_i} = \frac{(30 \times 2.5) + (15 \times 7.5) + (11 \times 12.5) + \cdots + (31 \times 27.5)}{120} \\ &= \frac{1855}{120} \\ &= 15.4583 \text{ years of experience}\end{aligned}$$

- (e) Would a bar chart or a histogram be more appropriate for representing the years of experience data? Explain your reasoning and describe the shape of the distribution.

Solution

\therefore The number of years is a continuous random variable a **histogram** would be the most appropriate graphical representation of the data. The distributions is bimodal, symmetric.