HW_2_Samir_Banjara

Problem 1:

a)

Mean = 92.967 Degrees Fahrenheit

b)

Median = 92 Degrees Fahrenheit

c)

Mode = 88 Degrees Fahrenheit

d)

10% Trimmed mean = 92.75 Degrees Fahrenheit

e)

Considering that the data set posses quantitative data and posses no outliers since the lower bound for outliers is $75.125 = (Q1 - 1.5 * IQR)$ and upper bound is $110.124 = (Q3 + 1.5 * IQR)$ and there are no data point that fall outside these bounds. Since the mean is sensitive to outliers and shows a google overall average when the data set is symmetrically distributed, and the median is robust to outliers and better represents the middle of the data if there is a presence of outliers, and mode which represents the most frequent value in the data which is more useful in the presence of categorical data, and the trimmed mean which removes 10% of the most extreme data and gives a more balanced average making it less sensitive to outliers than the mean but more sensitive than the median. Considering these factors where the data set posses quantitative data without outliers I would recommend that the mean, 92.967 Degrees Fahrenheit, sufficiently describes the center of the data.

Problem 2:

a)

Sample variance of the time until failure : 5.82

$$s^2 = \frac{\Sigma(x_i - \hat{x})^2}{n - 1}$$

b)

Sample standard deviation of the time until failure : 2.41

$$\sqrt{s^2 = \frac{\Sigma(x_i - \hat{x})^2}{n-1}}$$

c)

Range of time until failure : 8

$$\text{Range} = \text{Maximum Data Value} - \text{Minimum Data Value}$$

d)

Outliers in the data set, 2 and 10, significantly affect the variance and range since we have a very small sample size of only 10 observations.

Problem 3:

a)

Portfolio A:

- **Mean:** 150,000
- **Sample Standard Deviation:** 7,905.69
- **Maximum Value:** 160,000
- **Minimum Value:** 140,000
- **Range:** 20,000

Portfolio B:

- **Mean:** 150,000
- **Sample Standard Deviation:** 37,249.16
- **Maximum Value:** 195,000
- **Minimum Value:** 100,000
- **Range:** 95,000

b)

The standard deviation indicates how much data points typically differ from the average. A lower standard deviation suggests that values are generally closer to the mean, indicating less fluctuation and thus lower risk.

Portfolio A has a standard deviation of 7,905.69, while Portfolio B's is much higher at 37,249.16. This means that Portfolio A's values tend to stay nearer to its average of 150,000, showing more stability and lower risk.

In contrast, Portfolio B's larger standard deviation reveals that its values vary more widely from the average, pointing to greater unpredictability and higher risk.

Problem 4:

a)

First Quartile (Q1): 41.25

b)

Second Quartile (Q2): 61.0

c)

Third Quartile (Q3): 83.25

d)

In the marketing study, the quartiles help us understand the distribution of recall test scores. – The **first quartile** (Q1) is the score below which **25% of the test scores fall**. It shows that a quarter of the participants had very low recall. – The second quartile (Q2), or median, represents the middle value of the data. **50% of the participants scored less than** 61.0, and **50% scored more than** 61.0. This gives a central measure of recall ability, meaning half of the group performed above this score, and half performed below it. The second quartile, also known as the median, gives a sense of the overall center of the data. – The third quartile (Q3) is the score below which 75% of the scores fall. In this study, **75% of the participants scored less than** 83.25, meaning only 25% of the participants scored higher than this. It shows that most participants had scores below 83.25, indicating the top 25% of participants had relatively better recall.
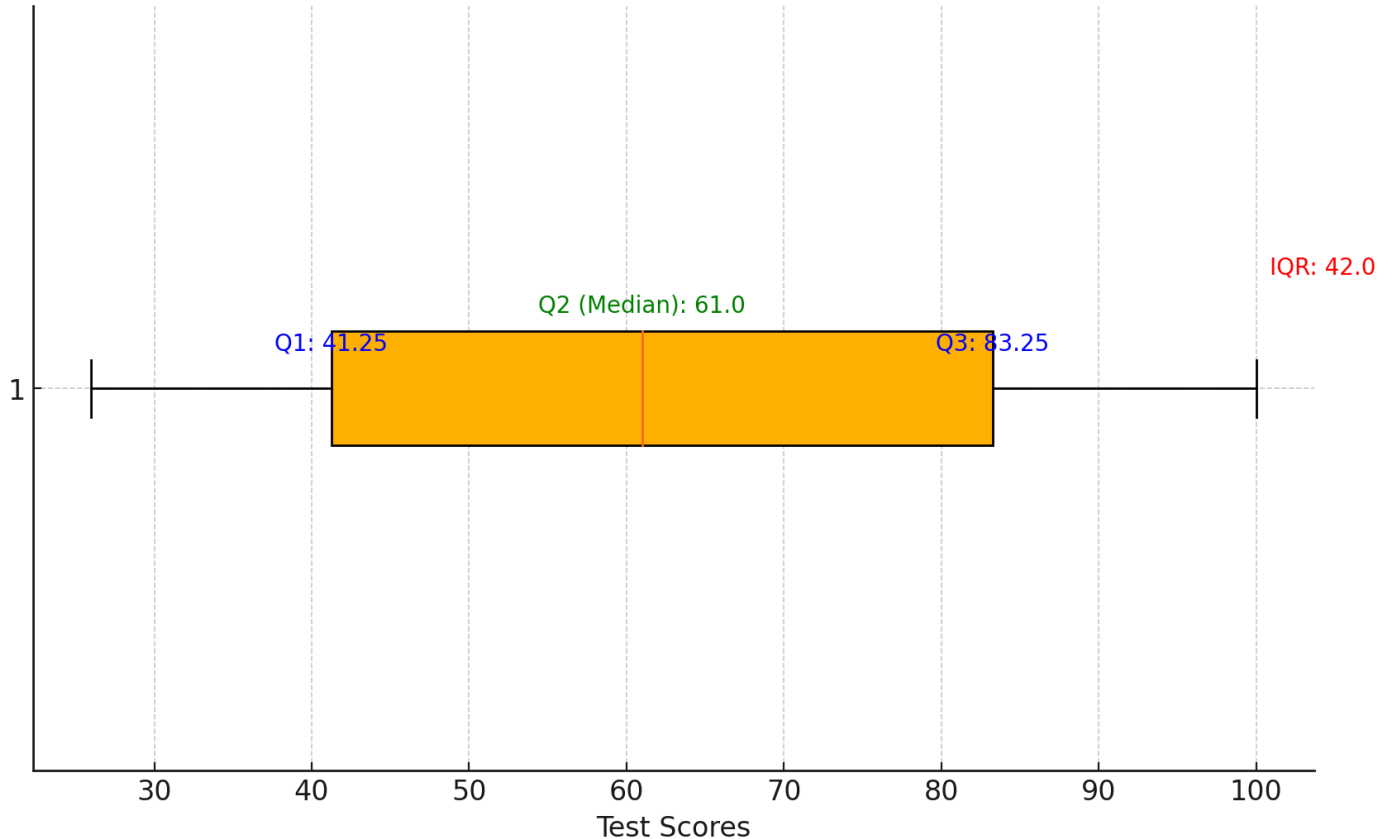
e)

IQR : 42.0 | "Interquartile Range"

f)

Box Plot of Test Scores with Quartiles, IQR, and Outliers

IQR: 42.0

Q2 (Median): 61.0

Q1: 41.25

Q3: 83.25

Test Scores

g)

There are no outliers outside the lower bound and upper bound calculated below,

$$\text{Lower Bound} = Q1 - 1.5 \times IQR$$
$$= 21.75$$

$$\text{Upper Bound} = Q3 + 1.5 \times IQR$$
$$= 146.25$$

Problem 5:

a)

In this scenario, we have two choices to make:

1. **Three levels** (lower level, middle level, upper level).
2. **Aisle seat or not** (yes for aisle, no for non-aisle).

To create the sample space, lets list all possible combinations of these choices:

- For the three levels, we'll call them L_1, L_2, and L_3 (these are lower, middle, and upper levels, respectively).
- For the aisle seat, the options are **Aisle** (A) or **Non-aisle** (N).

The sample space, S, is then:

$$S = \{(L_1, A), (L_1, N),$$
$$(L_2, A), (L_2, N),$$
$$(L_3, A), (L_3, N)\}$$

b)

The bag contains four marbles of different colors: **Green (G)**, **Red (R)**, **Blue (B)**, and **Violet (V)**. We are drawing two marbles without replacement, and the order of the draw does not matter. Since the order does not matter, we only need to consider unique combinations:

The sample space, S, is then:

$$S = \{(G, R), (G, B), (G, V)$$
$$(R, B), (R, V)$$
$$(B, V)\}$$

Problem 6:

a)

Each family member (A, B, C) is assigned at random to one of three stations $(1, 2, 3)$. Since each person has 3 possible station assignments, the total number of outcomes is:

$$3 \times 3 \times 3 = 27 \text{ outcomes}$$

The sample space S consists of all combinations of station assignments for A, B, and C. The outcomes are represented as ordered triples.

$$S = \{(1,1,1), (1,1,2), (1,1,3), (1,2,1), (1,2,2), (1,2,3), (1,3,1), (1,3,2), (1,3,3),$$
$$(2,1,1), (2,1,2), (2,1,3), (2,2,1), (2,2,2), (2,2,3), (2,3,1), (2,3,2), (2,3,3),$$
$$(3,1,1), (3,1,2), (3,1,3), (3,2,1), (3,2,2), (3,2,3), (3,3,1), (3,3,2), (3,3,3)\}$$

b)

$$E = \{(1,1,1), (2,2,2), (3,3,3)\}$$

Problem 7

a)

$$S = \{3, 4, 5, 13, 14, 15, 23, 24, 25, 123, 124, 125, 213, 214, 215\}$$

b)

$$A = \{3, 4, 5\}$$

c)

$$B = \{5, 15, 25, 125, 215\}$$

d)

$$C = \{3, 4, 5, 23, 24, 25\}$$