Tutorial 5:

The data given below is the total fat, in grams per serving, for a sample of 20 chicken sandwiches from fast-food chains.

7 8 4 5 16 20 20 24 19 30 23 30 25 19 29 29 30 30 40 56

a. Compute the mean, median, first quartile, and third quartile.

b. Compute the variance, standard deviation, range, interquartile range, Are there any outliers? Explain.

c. Are the data skewed? If so, how?

d. Based on the results of (a) through (c), what conclusions can you reach concerning the total fat of chicken sandwiches?

Solution:

Part (a) - Compute the Mean, Median, First Quartile, and Third Quartile

Step 1: Sort the Data in Ascending Order

4, 5, 7, 8, 16, 19, 19, 20, 20, 23, 24, 25, 29, 29, 30, 30, 30, 30, 40, 56

Step 2: Compute the Mean

The **mean** is the sum of all values divided by the number of values:

$$\text{Mean} = \frac{\sum X}{n} = \frac{4+5+7+8+16+19+19+20+20+23+24+25+29+29+30+30+30+30+40+56}{20}$$

$$\text{Mean} = \frac{420}{20} = 21$$

Step 3: Compute the Median

The **median** is the middle value of the sorted data. Since there are 20 data points, the median is the average of the 10th and 11th values:

$$\text{Median} = \frac{23+24}{2} = \frac{47}{2} = 23.5$$

Step 4: Compute the First Quartile (Q1)

The first quartile is the median of the lower half of the data (first 10 values): 4, 5, 7, 8, 16, 19, 19, 20, 20, 23

The median of this subset is the average of the 5th and 6th values:

$$Q1 = \frac{16+19}{2} = \frac{35}{2} = 17.5$$

Step 5: Compute the Third Quartile (Q3)

The third quartile is the median of the upper half of the data (last 10 values): 24, 25, 29, 29, 30, 30, 30, 30, 40, 56

The median of this subset is the average of the 5th and 6th values:

$$Q3 = \frac{30+30}{2} = 30$$

Part (b) - Compute the Variance, Standard Deviation, Range, Interquartile Range, and Outliers

Step 1: Compute the Range

The **range** is the difference between the maximum and minimum values:

$$\text{Range} = 56 - 4 = 52$$

Step 2: Compute the Variance and Standard Deviation

The **variance** is the average of the squared differences from the mean. First, compute the squared differences from the mean (21):

$$\text{Squared differences} = (4-21)^2, (5-21)^2, (7-21)^2, (8-21)^2, (16-21)^2, (19-21)^2, (19-21)^2, (20-21)^2, (20-21)^2, (23-21)^2, (24-21)^2, (25-21)^2, (29-21)^2, (29-21)^2, (30-21)^2, (30-21)^2, (30-21)^2, (30-21)^2, (40-21)^2, (56-21)^2$$

$$= 289, 256, 196, 169, 25, 4, 4, 1, 1, 4, 9, 16, 64, 64, 81, 81, 81, 81, 361, 1225$$

Sum of squared differences:

$$289 + 256 + 196 + 169 + 25 + 4 + 4 + 1 + 1 + 4 + 9 + 16 + 64 + 64 + 81 + 81 + 81 + 81 + 361 + 1225 = 3911$$

Variance:

$$\text{Variance} = \frac{3911}{20} = 195.55$$

Standard deviation:

$$\text{Standard deviation} = \sqrt{195.55} \approx 13.98$$

Step 3: Compute the Interquartile Range (IQR)

The **IQR** is the difference between the third quartile and the first quartile:

$$IQR = Q3 - Q1 = 30 - 17.5 = 12.5$$

Step 4: Identify Outliers

Outliers are typically defined as values outside the range:

$$\text{Lower bound} = Q1 - 1.5 \times IQR = 17.5 - 1.5 \times 12.5 = 17.5 - 18.75 = -1.25$$

$$\text{Upper bound} = Q3 + 1.5 \times IQR = 30 + 1.5 \times 12.5 = 30 + 18.75 = 48.75$$

Any data points below -1.25 or above 48.75 are considered outliers.

The only value above 48.75 is 56, so **56 is an outlier.**

Part (c) - Are the Data Skewed?

The **skewness** can be determined by comparing the mean and median:

- **Mean (21)** is less than **Median (23.5)**, indicating a **left-skewed** distribution (negative skew).

Part (d) - Conclusions

- The **mean** (21) and **median** (23.5) are fairly close, but the left-skew suggests that most of the data are clustered towards the higher fat values, with a few lower fat values pulling the mean down.
- The **interquartile range** (12.5) is moderate, showing that the middle 50% of the data are reasonably spread out.
- **56** is an **outlier**, indicating that one sandwich contains a significantly higher amount of fat compared to the others.
- The **skewness** suggests that the fat content distribution is not symmetrical, with more values concentrated towards the higher end.

2.Find the Variance and Standard Deviation of the Following Numbers: 1, 3, 5, 5, 6, 7, 9, 10.

SOL:-

Step 1: Compute the Mean

The **mean** is the sum of all values divided by the number of values:

$$\text{Mean} = \frac{1+3+5+5+6+7+9+10}{8} = \frac{46}{8} = 5.75$$

Step 2: Compute the Squared Differences from the Mean

Next, calculate the squared difference of each value from the mean (5.75):

$$(1-5.75)^2 = (-4.75)^2 = 22.5625$$

$$(3-5.75)^2 = (-2.75)^2 = 7.5625$$

$$(5-5.75)^2 = (-0.75)^2 = 0.5625$$

$$(5-5.75)^2 = (-0.75)^2 = 0.5625$$

$$(6-5.75)^2 = (0.25)^2 = 0.0625$$

$$(7-5.75)^2 = (1.25)^2 = 1.5625$$

$$(9-5.75)^2 = (3.25)^2 = 10.5625$$

$$(10-5.75)^2 = (4.25)^2 = 18.0625$$

Step 3: Compute the Variance

The **variance** is the average of the squared differences:

$$\text{Variance} = \frac{22.5625 + 7.5625 + 0.5625 + 0.5625 + 0.0625 + 1.5625 + 10.5625 + 18.0625}{8} = \frac{61.5}{8} = 7.6875$$

Step 4: Compute the Standard Deviation

The **standard deviation** is the square root of the variance:

$$\text{Standard Deviation} = \sqrt{7.6875} \approx 2.77$$

3. For the following list, n = 19. Find the median.

24, 25, 28, 31, 33, 33, 36, 42, 42, 48, 51, 57, 57, 68, 75, 79, 79, 79, 85

Solution:

Given Data:

24, 25, 28, 31, 33, 33, 36, 42, 42, 48, 51, 57, 57, 68, 75, 79, 79, 79, 85

- The data is already sorted.
- The number of data points is **n = 19** (odd number of values).

To find the **median** in an odd set of numbers, the median is the middle number. The middle number is located at position $\frac{n+1}{2}$.

$$\text{Position of median} = \frac{19+1}{2} = 10$$

The 10th value is **48.**

So, the **median** is **48.**

4. Five people play golf and at one hole their scores are 3, 4, 4, 5, 7 For these scores,

find (a) the mean (b) the median (c) the mode (d) the range.

Solution:

Given Scores:
3, 4, 4, 5, 7

(a) Mean

The **mean** is the sum of all values divided by the number of values:

$$\text{Mean} = \frac{3+4+4+5+7}{5} = \frac{23}{5} = 4.6$$

(b) Median

The **median** is the middle value when the data is ordered. The ordered data is: 3, 4, 4, 5, 7

The middle value is the 3rd value:

$$\text{Median} = 4$$

(c) Mode

The **mode** is the value that appears most frequently. The value 4 appears twice, while the others appear once:

$$\text{Mode} = 4$$

(d) Range

The **range** is the difference between the maximum and minimum values:

$$\text{Range} = 7 - 3 = 4$$

5. The following data represent the battery life (in shots) for three-pixel digital cameras: 300 180 85 170 380 460 260 35 380 120 110 240 List the Five-point summary.

Solution:

Step 1: Sort the Data in Ascending Order

35, 85, 110, 120, 170, 180, 240, 260, 300, 380, 380, 460

Step 2: Find the Five-Point Summary

The five-point summary consists of:

1. **Minimum:** The smallest value.
2. **First Quartile (Q1)**: The median of the lower half of the data.
3. **Median (Q2)**: The middle value of the data.
4. **Third Quartile (Q3)**: The median of the upper half of the data.
5. **Maximum**: The largest value.

1. Minimum:

The minimum value is the first value in the sorted data:

$$\text{Minimum} = 35$$

2. Median (Q2):

There are 12 values, so the median is the average of the 6th and 7th values in the sorted list:

$$\text{Median} = \frac{180 + 240}{2} = 210$$

3. First Quartile (Q1):

The first quartile is the median of the lower half of the data (values before the median):

$$\text{Lower half} = 35, 85, 110, 120, 170, 180$$

The median of this group is the average of the 3rd and 4th values:

$$Q1 = \frac{110 + 120}{2} = 115$$

4. Third Quartile (Q3):

The third quartile is the median of the upper half of the data (values after the median):

$$\text{Upper half} = 240, 260, 300, 380, 380, 460$$

The median of this group is the average of the 3rd and 4th values:

$$Q3 = \frac{300 + 380}{2} = 340$$

5. Maximum:

The maximum value is the last value in the sorted data:

$$\text{Maximum} = 460$$

VIVA:-

1. What is the coefficient of variation (CV), and when is it useful?

2. How is the standard deviation calculated, and what is its significance?

3. Define variance and how it relates to the spread of data points.

4. When is the median a better measure of central tendency compared to the mean?\

ANSWERS:

1. **Coefficient of Variation (CV):**
   CV is the ratio of the standard deviation to the mean, expressed as a percentage. It measures relative variability and is useful when comparing data with different units or scales.
2. **Standard Deviation:**
   Standard deviation is the square root of the variance. It measures the average distance of data points from the mean, indicating the spread or dispersion of data.
3. **Variance:**
   Variance is the average of the squared differences from the mean. It quantifies how spread out the data points are from the mean.
4. **Median vs. Mean:**
   The median is better when data is skewed or has outliers, as it is less sensitive to extreme values.