Descriptive Statistics Assignment

Question1.

A bakery tracks the daily sales of muffins (in dozens) over a week: [10, 12, 11, 15, 14, 13, 12]. What is the most representative value of their weekly sales, and why?

Ans. 12 dozen muffins per day.

First sort the data:

[10, 11, 12, 12, 13, 14, 15]

Mean=(10+11+12+12+13+14+15)/7

=87/7

=12.43

Mean=12.4

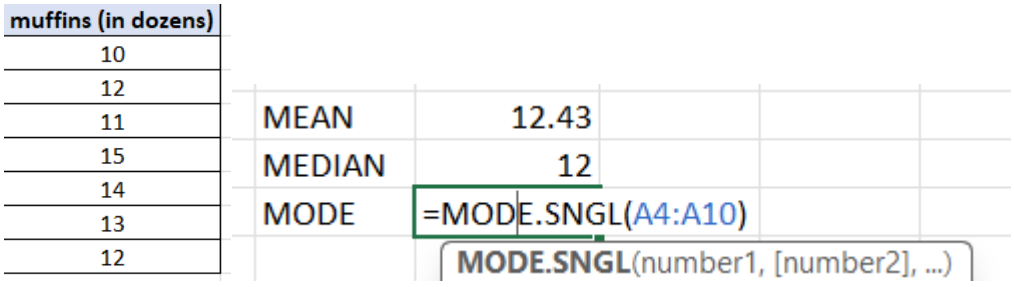
Median= Middle value = 4th item = 12

Median=12 dozens

Mode=12 (most repeated value).

Range=Max-Min=**15**-**10**=**5**

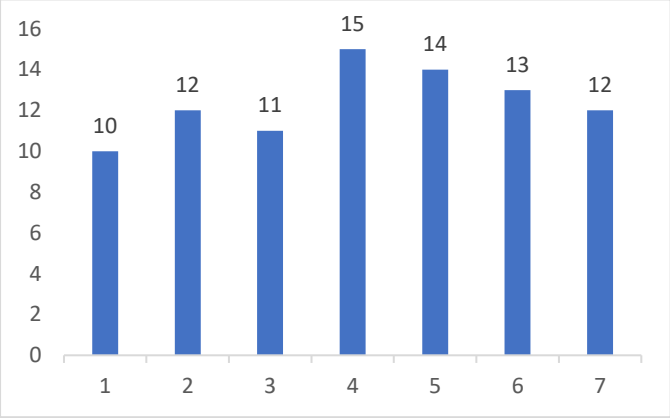
Also in excel,



The **mean (≈12.4)** and **median (12)** are very close, and the **mode (12)** also matches.
This consistency suggests that **12 dozen muffins per day** is the most representative value.
The distribution is balanced (no skew or outliers).

Mean, median, and mode all cluster around 12.

The most representative value is **12 dozen muffins per day**, since it reflects the central tendency and matches both median and mode.

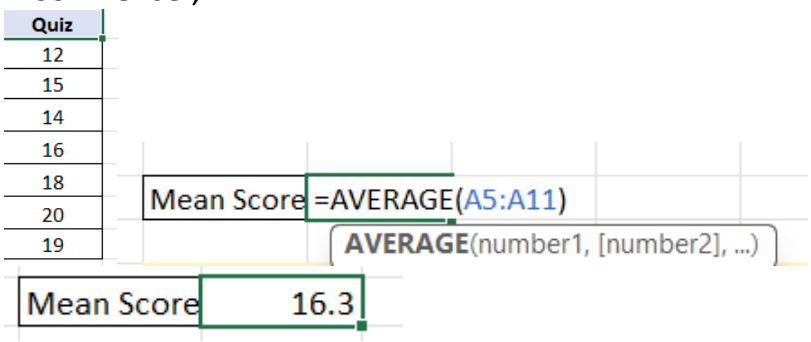


Question2.
A teacher records the marks of her students in a short quiz: [12, 15, 14, 16, 18, 20, 19]. What is the mean score, and what does it tell us about the class's performance?

Ans.

Mean=(12+15+14+16+18+20+19)/7
=114/7
=16.3
Mean score is 16.3

Also in excel,



The quiz scores range from 12 to 20, with the mean sitting at 16.3.
This suggests that most students performed above average, since the mean is closer to the higher end of the scale.
The mean score is 16.3, which indicates that the class performed well overall, with most students achieving marks in the mid-to-high range.

Question3.
A store records the shoe sizes sold in one day: [7, 8, 9, 8, 8, 10, 7, 9]. What is the mode, and why is this information useful for the store manager?

Ans.
The data set is: **[7, 8, 9, 8, 8, 10, 7, 9]**
Size **7** appears **2 times**
Size **8** appears **3 times**
Size **9** appears **2 times**
Size **10** appears **1 time**

The **mode** is **8**, because it occurs most frequently.

Why is this useful for the store manager?

Knowing the mode helps the manager understand **customer demand patterns**. The **mode** tells the manager which shoe size is in **highest demand**, guiding smarter stocking and reducing the risk of running out of the most-sold size.

Question4.
A car dealer notes the prices of used cars: [$8,000, $9,500, $10,200, $11,000, $50,000]. Why is the median a better measure than the mean in this case? Calculate the median.

Ans.

Step 1: Arrange the data
Prices:
$8,000, $9,500, $10,200, $11,000, $50,000
(Already sorted in ascending order)

Step 2: Calculate the **Mean**

Mean = (8000+9500+10200+11000+50000)**/**5

(88700)/5 = 17,740
Mean=$17,740

Step 3: Calculate the **Median**

Since there are **5 values** (odd number), the median is the **middle value**.
Middle value = 3rd item = $10,200
Median = $10,200

Median is a better measure than the mean in this case because
The **mean ($17,740)** is heavily skewed by the **outlier ($50,000)**.
The **median ($10,200)** better represents the "typical" car price because it isn't distorted by one unusually high value.
So, **median ($10,200)** is a better measure because it reflects the typical car price without being distorted by the extreme $50,000 value.

Question5.
A student times how long it takes to finish a puzzle each day:
[25, 30, 27, 35, 40].
What does the range tell us about the variation in the student's puzzle-solving time?

Ans.
Steps:
1 Identify the data
Puzzle times: [25, 30, 27, 35, 40]

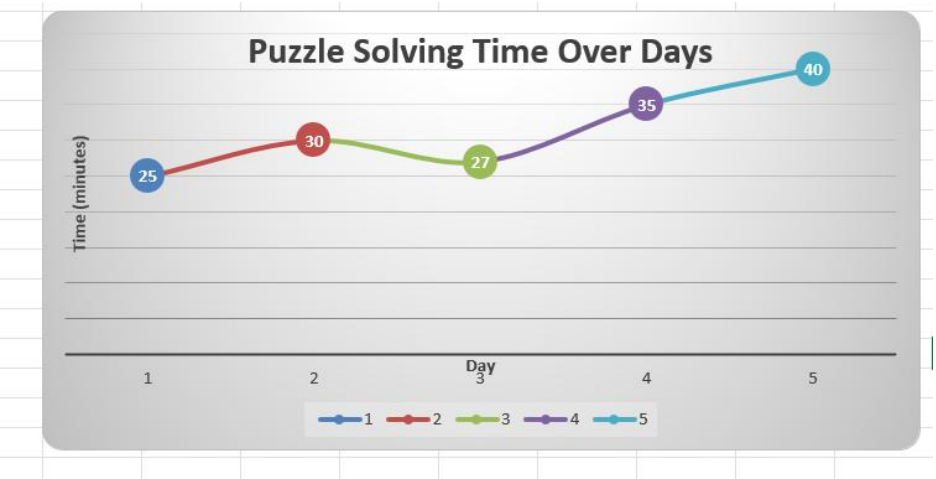
2 Calculate the Range
Range=Max value-Min value
40-25=**15**

The range shows the spread of the data — how much the times vary from the fastest to the slowest day.

A range of 15 minutes means the student's puzzle-solving time can differ by up to 15 minutes depending on the day.

This indicates moderate variation: sometimes the student is quite quick (25 minutes), while other times they take significantly longer (40 minutes).

The range tells us that the student's puzzle-solving time isn't consistent — it can vary by as much as 15 minutes from day to day.



Question6.
A farmer records the weekly weight of harvested apples (kg): [100, 105, 98, 110, 120]. Find the range. How can this help the farmer in planning his packaging?

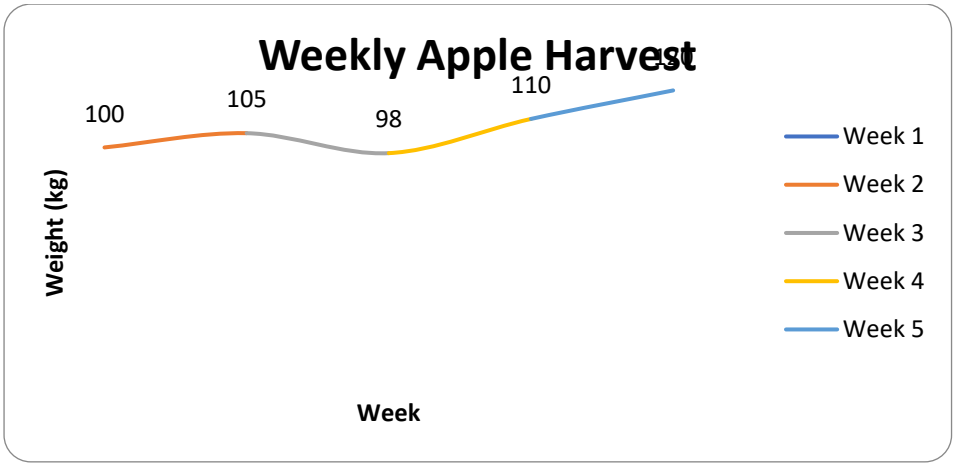
Ans.
Steps:
1 Identify the data
Weekly harvest weights: **[100, 105, 98, 110, 120]**
2 Calculate the Range
Range=Maximum value – Minimum value
120-98**=22**
The range is 22 kg.

The range shows how much the harvest weight fluctuates from week to week.
Packaging planning: Knowing that harvests can vary by up to 22 kg helps the farmer decide how many boxes or crates to keep ready.

Logistics: It ensures transport capacity is flexible enough to handle both smaller (98 kg) and larger (120 kg) weeks.

Sales forecasting: The farmer can estimate minimum and maximum supply, helping negotiate with buyers or markets.

The range (22 kg) tells the farmer how much weekly harvests can vary, helping them plan packaging and logistics to avoid shortages or excess.



Question7.
Two delivery companies track delivery delays (in minutes).
Company A: variance = 6
Company B: variance = 15
Which company is more consistent, and why?

Ans.
Variance measures how spread out the data is around the mean.
A **smaller variance** means the values are closer to the average → more consistent performance.
A **larger variance** means the values fluctuate more → less consistent performance.

Compare the two companies
Company A: variance = 6 smaller spread, delays are more predictable.
Company B: variance = 15 larger spread, delays vary more widely.

Conclusion:
Company A is more consistent because its variance is lower.
This means their delivery delays are more stable and reliable, while Company B's delays are less predictable.

Question8.
A finance student compares the daily price fluctuations of two cryptocurrencies.
Coin A: standard deviation = $30
Coin B: standard deviation = $120
Which coin is riskier to invest in, and why?

Ans.
Standard deviation (SD) is a key statistic measuring how spread out data points are from their average (mean); a low SD means data clusters near the mean, showing stability, while a high SD indicates data is widely dispersed, showing more variability or risk.
A higher standard deviation **= more risk**
A lower standard deviation = **less risk**

Coin A: $30 - smaller fluctuations, more stable.
Coin B: $120 - much larger fluctuations, highly volatile.

Conclusion:
Coin B is riskier to invest in because its standard deviation is much higher. This means its daily price swings are larger, making returns less predictable and exposing investors to greater potential losses (and gains).
The higher the standard deviation, the greater the investment risk. Coin B's $120 deviation signals high volatility, while Coin A's $30 deviation suggests more consistent performance.

Question9.
A family records their monthly electricity usage (in kWh): [400, 420, 390, 450, 410].
Find the mean and standard deviation. What do these values together tell you about the family's energy use pattern?

Ans.
Electricity usage (kWh): **[400, 420, 390, 450, 410]**

Mean=(400+420+390+450+410)/5
=2070/5
=414
The mean usage **= 414 kWh**.

To calculate the Standard Deviation these steps must be follow:

1 Differences from mean:

400-414=-14

420-414=6

390-414=-24

450-414=36

410-414=-4

2 Squares of differences:

$(-14)^2=196$

$6^2=36$

$(-24)^2=576$

$36^2=1296$

$(-4)^2=16$

3 Sum of squares =(196+36+576+1296+16) =**2120**

4 Variance = (2120)/5 = 424

5 Standard deviation = √424 = 20.6 approx.

The **standard deviation ≈ 20.6 kWh**.

Conclusion:

a. The **mean (414 kWh)** shows the family's typical monthly electricity usage.

b. The **standard deviation (≈ 21 kWh)** shows that their usage doesn't vary much — it usually stays within about ±20 kWh of the average.

c. Together, these values suggest the family's energy use is **fairly consistent** month to month, with only small fluctuations.

Question10.

A basketball player's points in 8 games are recorded: [15, 18, 20, 22, 25, 17, 19, 21].

Find the mean, median, mode, range, and standard deviation. What insights can these measures provide about the player's scoring performance?

Ans.

Scores: [15, 18, 20, 22, 25, 17, 19, 21]

First sort the data.

[15,17,18,19,20,21,22,25]

MEAN:

Mean=(15+17+18+19+20+21+22+25)/8

=157/8

=19.62

Mean=19.6 points.

MEDIAN:

8 values → median = average of 4th and 5th values

Median=(19+20)/2

=39/2

=19.5

Median=19.5 points.

MODE:

Each score appears only once. So there is no mode.

RANGE:

Range=MAXIMUM value – MINIMUM value

=25-15=10

Range=10 points

STANDARD DEVIATION:

1 Mean **= 19.625**

2 Differences from mean:

15-19.625=-**4.625**

18-19.625=-**1.625**

20-19.625=**0.375**

22-19.625=**2.375**

25-19.625=**5.375**

17-19.625=-**2.625**

19-19.625=-**0.625**

21-19.625=**1.375**

3 Squares of differences:

21.39, 2.64, 0.14, 5.64, 28.89, 6.89, 0.39, 1.89

4 Sum **= 67.5**

5 Variance **= 67.5/8 = 8.44**

6 Standard deviation **= √8.44=2.9**

Standard deviation ≈ 2.9 points

a. Mean (≈19.6): The player typically scores around 20 points per game.
b. Median (19.5): Confirms the average is representative of the central tendency.
c. No mode: Scoring is varied, no single repeated score.
d. Range (10): The difference between best (25) and worst (15) games is moderate.
e. Standard deviation (≈2.9): Scores are fairly consistent, usually within ±3 points of the average.