Day 14 – Assignment Questions

1. What is the average sales value?

The average sales value is calculated using the mean formula. It gives the central value of sales data.

Formula:
Mean = Sum of all sales / Number of sales

Example in Python:
```
import numpy as np
sales = [200, 300, 250, 400, 500, 600, 150]
mean_sales = np.mean(sales)
print(mean_sales)
```
The mean helps to understand the overall sales performance.

2. Is income distribution skewed?

To check whether income distribution is skewed, we calculate skewness.
If skewness > 0 → Right skewed
If skewness < 0 → Left skewed
If skewness = 0 → Normally distributed

Example:
```
from scipy.stats import skew
income = [20000, 25000, 30000, 50000, 100000]
print(skew(income))
```
If the value is positive and large, the income distribution is right skewed.

3. What is the most common gender?
The most common gender is found using mode.
Example:
```
import pandas as pd
df = pd.DataFrame({"gender": ["Male","Female","Male","Male","Female"]})
print(df["gender"].mode())
```

Mode gives the most frequently occurring value.

4. How much variability exists in delivery time?
Variability is measured using standard deviation or variance.
Standard deviation shows how much delivery time varies from the mean.
Example:

```
delivery_time = [2, 3, 5, 4, 6, 3]
print(np.std(delivery_time))
```

Higher standard deviation means higher variability.

5. What is the spread of sales?
Spread of sales is measured using standard deviation and variance.
It shows how widely sales values are distributed from the average.

6. What is the sales range?
Range = Maximum sales – Minimum sales
Example:
```
print(max(sales) - min(sales))
```
Range gives the simplest measure of dispersion.

7. Is income right skewed?
If the mean is greater than median and skewness is positive, income is right skewed.
This usually happens when a few people earn very high income.

8. Are sales heavy tailed?
Heavy tailed distribution means extreme values are present.
We check kurtosis.

If kurtosis > 3 → Heavy tailed
If kurtosis < 3 → Light tailed

Example:
```
from scipy.stats import kurtosis
print(kurtosis(sales))
```

9. Detect sales outliers
Outliers can be detected using IQR method.
Q1 = 25th percentile
Q3 = 75th percentile
IQR = Q3 – Q1
Lower limit = Q1 – 1.5 * IQR
Upper limit = Q3 + 1.5 * IQR
Values outside this range are outliers.

10. Standardize the sales
Standardization is done using Z-score.

$$Z = (x - mean) / \text{standard deviation}$$

Example:
```
z_scores = (sales - np.mean(sales)) / np.std(sales)
print(z_scores)
```

Standardization helps in machine learning models.

11. What are quartiles of sales?
Quartiles divide data into four equal parts.
Q1 – 25%
Q2 – Median (50%)
Q3 – 75%

Example:
```
print(np.percentile(sales, [25, 50, 75]))
```

12. Compare volatility using coefficient of variation
Coefficient of Variation (CV) = (Standard Deviation / Mean) × 100
It is used to compare variability between datasets.

Example:
```
cv = (np.std(sales) / np.mean(sales)) * 100
print(cv)
```

Higher CV means higher volatility.