Descriptive Statistics

LLE – Mathematics and Statistics Skills

1 Variable Classification

For the following variables, how would you classify them?

1. A variable recording a person's primary method of transport to work (e.g., Car, Bus, Train, Bicycle).

⊗ **Nominal**　　　○ Ordinal　　　○ Continuous　　　○ Discrete

> **Solution**
> The transport methods are distinct categories with no logical order, making this Nominal data.

2. A variable recording a t-shirt size (e.g., Small, Medium, Large, Extra Large).

○ Nominal　　　⊗ **Ordinal**　　　○ Continuous　　　○ Discrete

> **Solution**
> The sizes have a clear, meaningful order, but the difference in measurement between 'Small' and 'Medium' isn't necessarily the same as between 'Large' and 'Extra Large'. This is Ordinal data.

3. A variable recording the weight of a package in kilograms.

○ Nominal　　　○ Ordinal　　　⊗ **Continuous**　　　○ Discrete

> **Solution**
> Weight is a precise numerical measurement, so continuous.

4. A variable recording the brand of a person's mobile phone (e.g., Apple, Samsung, Google).

⊗ **Nominal** ○ Ordinal ○ Continuous ○ Discrete

> **Solution**
> Phone brands are just labels. There is no inherent ranking or order, so this is Nominal data.

5. A particular company uses the following to rate their financial position 'Thriving', 'OK', 'Breaking even', 'Losing money' or 'Failing'.

○ Nominal ⊗ **Ordinal** ○ Continuous ○ Discrete

> **Solution**
> 'Failing' is clearly worse than 'Breaking even', so there is an order. However, the difference between the levels is not a measurable quantity. This is Ordinal data.

6. A variable that represents a user's shoe size.

○ Nominal ○ Ordinal ○ Continuous ⊗ **Discrete**

> **Solution**
> This is a numerical quantity but can only take fixed values (5, 5.5, 6, …). This is discrete data.

2 Measures of Average

1. Given the data 5, 8, 8, 9, 15, calculate the mean.

Answer: _____ **9**

> **Solution**
> The values add up to 45. There are 5 values. $45 \div 5 = 9$.

2. Given the data 5, 8, 8, 9, 15, calculate the median.

Answer: _____ **8**

3. Given the data 5, 8, 8, 9, 15, find the mode.

Answer: _____ **8**

Solution
The number 8 appears most often.

4. What is another name for the 50th percentile?

○ Range   ○ Mean   ⊗ **Median**

Solution
The 50th percentile is the middle point of the data, which is the Median.

3   Measures of Dispersion

1. Given the data 5, 8, 8, 9, 15, calculate the Range.

Answer: _____ **10**

Solution
The range is the difference between the largest and smallest values.
$15 - 5 = 10$.

Formulas

Variance (s^2): A measure of how disperse the data are.

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

Standard Deviation (s): The square root of the variance.

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

A customer records the number of kilometres they run on 9 gym visits:

$$12, 10, 14, 14, 11, 8, 15, 18, 15$$

2. Calculate the mean distance ($\bar{d}$).

Answer: _____ **13**

> **Solution**
> Sum of values = 117. Count (n) = 9.
> $117 \div 9 = 13$.

3. Complete the table below using the mean ($\bar{d} = 13$).

Distance (d_i)	Difference ($d_i - \bar{d}$)	Squared ($(d_i - \bar{d})^2$)
12	-1	1
10	-3	9
14	1	1
14	1	1
11	-2	4
8	-5	25
15	2	4
18	5	25
15	2	4
SUM	—-	74

Solution

The middle column is the difference from the mean (13). The final column is the square of that difference. The bottom row is the sum of the squared column.

4. Calculate the sample variance (s^2), correct to 2 decimal places.

Answer: _____ **9.25**

Solution

Using the sum from the table (74) and $n = 9$:
$s^2 = \frac{74}{9-1} = \frac{74}{8} = 9.25$

5. Calculate the sample standard deviation (s), correct to 2 decimal places.

Answer: _____ **3.04**

Solution

$s = \sqrt{9.25} \approx 3.04$