Section 3.1.1 – Data Types

VCAA "Dot Points"
Investigating data distributions, including:

• *review of types of data*
Investigating and modelling linear associations, including:

Types of Data

Diagram 1 below provides a visual comparison of the data types or classifications used in statistics:

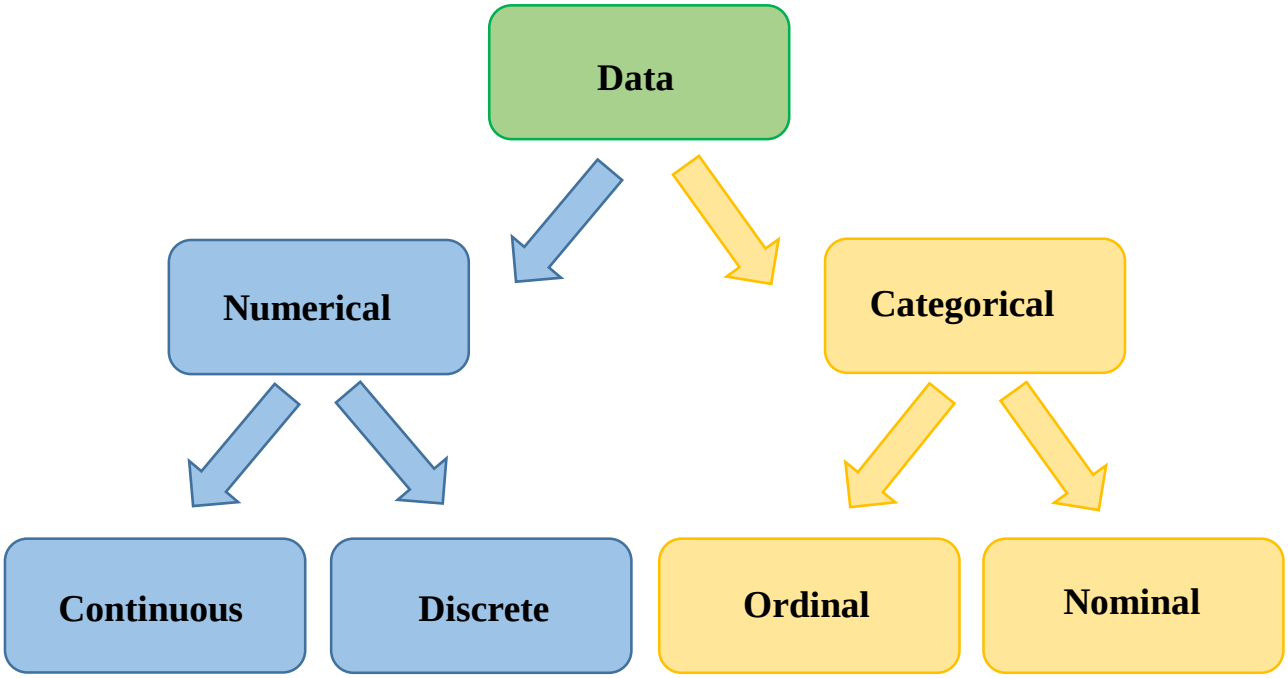


Diagram.1 – *Data Classification table*

Numerical Data is data assigned a **numerical** value.
Categorical Data is data that can be placed into two or more **categories**.

NB: Not all data consisting of numbers is necessarily classified as numerical.

Take for example the typical **numerical rating scale**, where:
1 = Strongly Disagree, 2 = Disagree, 3 = Neutral, 4 = Agree & 5 = Strongly Agree.
In this example these "numbers" represent an **order of categorical data.**

Postcodes too are classified as **categorical data**, rather than numerical data. Despite consisting of numbers, a postcode represent a category, or region.

Region	Postcode
Traralgon	3844
Moe	3825
Morwell	3840
Churchill	3842
Sale	3850

Continuous Numerical Variables

A continuous variable is a variable that has an **infinite number** of possible values, within a range. In other words, **any value is possible** for the variable.

Examples of continuous numerical variables include:

- **A person's weight.** Someone could weigh 100 kilograms, they could weigh 100.10 kilograms or they could weigh 100.1110 kilograms. The number of possibilities for weight are limitless.
- **Time of a race**. The time of a race could be measured to the nearest second, millisecond, microsecond and so forth. There are limitless possibilities depending upon the precission of the timing device.

Discrete Numerical Variables

A discrete variable is a variable that can only take on a **certain number** of values. In other words, they don't have an infinite number of values. If you **can count** a set of items, then it's a discrete variable.

Examples of discrete numerical variables include:

- **Number of coins** in a purse or jar. Discrete because there can only be a certain number of coins (1,2,3,4,5…). In addition, a purse or even a bank is restricted by size so there can only be so many coins.
- The **number of cars** in a parking lot. A parking lot can only hold a certain number of cars.
- **Ages on birthday cards**. Birthday cards only come in years…they don't come in fractions. So there are a finite amount of possibilities (presumably, about one hundred).

Ordinal Categorical Variables

Ordinal variables, can be arranged into **categories that have an order**. Note that the difference between each value is not important.

Examples of ordinal categorical variables include:

- When patients are asked to express the amount of pain they are feeling from 1 to 10.
- Socioeconomic status
- Military rank
- Letter grades for coursework

Nominal Categorical Variables

Nominal variables have **two or more categories** without having any kind of natural order. They are variables with **no numeric value**, such as occupation or political party affiliation. Another way of thinking about nominal variables is that **they are named** (nominal is from Latin "nominalis", meaning pertaining to names).

Nominal variables:

- Cannot be quantified. In other words, you can't perform arithmetic operations on them, like addition or subtraction, or logical operations like "equal to" or "greater than" on them.
- Cannot be assigned any order.

Examples of nominal categorical variables include:

- Gender (Male, Female, Transgender).
- Eye color (Blue, Green, Brown, Hazel).
- Type of house (Bungalow, Duplex, Ranch).
- Type of pet (Dog, Cat, Rodent, Fish, Bird).
- Genotype (AA, Aa, or aa).

Example 1

The back-to-back ordered stemplot below shows the distribution of maximum temperatures (in °Celsius) of two towns, Beachside and Flattown, over 21 days in January.

Beachside		Flattown
9 8 7 5	1	8 9
4 3 2 2 1 1 0 0	2	
9 9 8 7 6 5	2	8 9
3 2	3	3 3 4
8	3	5 5 6 7 7 7 8 8
	4	0 0 1 2
	4	5 6

The variables temperature (°Celsius) and town (Beachside or Flattown) are:

A. both categorical variables.
B. both numerical variables.
C. categorical and numerical variables respectively.
D. numerical and categorical variables respectively.
E. neither categorical nor numerical variables.

D

Back to back stem and leaf plots always display two categories (ie. The towns Beachside & Flattown) and numerical data (ie. Temperature °C) within the stemplot.

Example 2

A survey was taken across several different regions of Australia where the population density for each region was recorded.

The variables:

Region (city, urban, rural)
Population density (number of people per square kilometer)

A. are both categorical.
B. are both numerical.
C. are categorical and numerical respectively.
D. are numerical and categorical respectively.
E. are neither categorical nor numerical.

C

Clearly the variable "*Region*" has a range of different categories and "*Population density*" has a numerical measure not associated with classification.

Exam Styled Questions – Multiple Choice

Use the following information to answer Questions 1-4.

The following table shows the data collected from a random sample of seven drivers drawn from the population of all drivers who used a supermarket car park on one day. The variables in the table are:

- *distance* – the distance that each driver travelled to the supermarket from their home
- *sex* – the sex of the driver (female, male)
- *number of children* – the number of children in the car
- *type of car* – the type of car (sedan, wagon, other)
- *postcode* – the postcode of the driver's home.

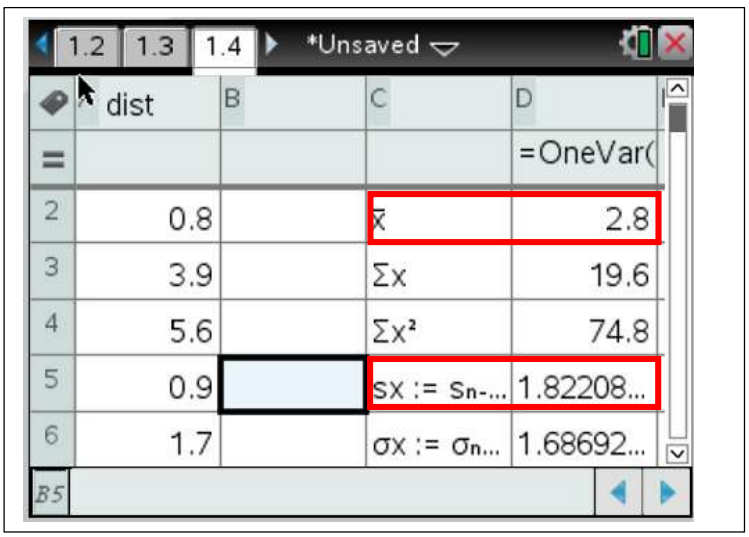
Distance (km)	*Sex* (F = female, M = male)	*Number of children*	*Type of car* (1 = sedan, 2 = wagon, 3 = other)	*Postcode*
4.2	F	2	1	8148
0.8	M	3	2	8147
3.9	F	3	2	8146
5.6	F	1	3	8245
0.9	M	1	3	8148
1.7	F	2	2	8147
2.5	M	2	2	8145

Question 1

(2016 Sample Exam 1 Section A - Qn 3)

The mean, $\bar{x}$, and the standard deviation, s_x, of the variable, distance, for these drivers are closest to

A. $\bar{x} = 2.5$ $s_x = 3.3$
B. $\bar{x} = 2.8$ $s_x = 1.7$
C. $\bar{x} = 2.8$ $s_x = 1.8$
D. $\bar{x} = 2.9$ $s_x = 1.7$
E. $\bar{x} = 3.3$ $s_x = 2.5$



C

Using the TI-nspire

∴ **Option C**

Question 2
(2016 Sample Exam 1 Section A - Qn 4)
The number of discrete numerical variables in this data set is

A. 0
B. 1
C. 2
D. 3
E. 4

B

NB: A discrete numerical variables consists of data that can only take on a certain number of values.
Eg. The number of matches in a match box.

There are four variables that contain numbers:

Distance (km) – Is a continuous numerical variable (a range of possible values are possible)
Number of children – Is a discrete numerical variable (only fixed values that can be counted)
Type of car – Although the variable uses numbers, it is a nominal categorical variable
Postcode – Although the variable uses numbers, it is a nominal categorical variable

∴ **Option B**

Question 3
(2016 Sample Exam 1 Section A - Qn 5)
The number of ordinal variables in this data set is

A. 0
B. 1
C. 2
D. 3
E. 4

A

NB: An ordinal variables consists of data that can be categorised in an order
None of the variables shown in the above table are ordinal

∴ **Option A**

Question 4
(2016 Sample Exam 1 Section A - Qn 6)
The number of female drivers with three children in the car is

A. 0
B. 1
C. 2
D. 3
E. 4

B

From of the table, there are four female drives recorded, only one of which has three children.
∴ **Option B**
