Ⓐ Variables and Measures

Variables can be classified into **category variables** and **numerical variables**.

A **category variable** (also called a **qualitative variable**) is a property that can be organized into **distinct subgroups**. These subgroups must be defined without overlap so that every value belongs to exactly one subgroup.

Examples of category variables would be the status of a person's home (*freehold*, *mortgaged*, *rented*) or gender (*male*, *female*).

A **numerical variable** (also called a **quantitative variable**) is a property for which the values result from measuring with a measuring device or from counting. Therefore numerical variables are either **continuous (measurement) variables** or **discrete (counting) variables**.

Examples of continuous numerical variables would be *lengths* or *times*.

Examples of discrete numerical variables would be the *number of correct questions in a test* or the *number of days in hospital*.

Deciding on the type of variable can depend on how they are measured. For example, if the *age of a person* is given in a whole number of years it would be a discrete numerical variable; if it was given as a decimal number it would be a continuous numerical variable; if ages were put into groups then it would be a category variable.

Example : Data gathered by a fisheries researcher includes the *breed of fish*, its *length*, its *weight*, the *number of fish in the trap*.

    a) What are the variables for this data?

    b) What units (if this is appropriate) would be used to measure each variable?

    c) Classify each variable as a category variable, continuous numerical variable or discrete numerical variable.

Answer :

a) Variable	b) Unit of Measurement	c) Type of Variable
Breed	*(not appropriate)*	*Category*
Length	*mm*	*Continuous Numerical*
Weight	*g*	*Continuous Numerical*
Number in Trap	*(not appropriate)*	*Discrete Numerical*

1   For each of the following questions state what units you would use to measure each variable (if this is appropriate) and classify them as a category variable, a continuous numerical variable or a discrete numerical variable.

a)   Data from a car dealer's yard consists of the year of manufacture, odometer reading, brand and number of previous owners.

Variable	Unit of Measurement	Type of Variable
……………………………………………………	…………………………..	………………………………………………………………
……………………………………………………	…………………………..	………………………………………………………………
……………………………………………………	…………………………..	………………………………………………………………
……………………………………………………	…………………………..	………………………………………………………………

b)   Data on a donor data base includes people's weights, heights, ages in a whole number of years, eye colours.

Variable	Unit of Measurement	Type of Variable
……………………………………………………	…………………………..	………………………………………………………………
……………………………………………………	…………………………..	………………………………………………………………
……………………………………………………	…………………………..	………………………………………………………………
……………………………………………………	…………………………..	………………………………………………………………

c)   Data from family data base includes the number of children, the work status of the main income earner (full time, part time, retired), distance the main income earner travels to work and the total family income.

Variable	Unit of Measurement	Type of Variable
……………………………………………………	…………………………..	………………………………………………………………
……………………………………………………	…………………………..	………………………………………………………………
……………………………………………………	…………………………..	………………………………………………………………
……………………………………………………	…………………………..	………………………………………………………………

## A The Nature of the Relationship

Now is the time to **analyse** your data. You need to describe *patterns* that you can see in your data.

Start by drawing an oval (or band) that encloses most of the scattered points and comment on the **general trend** (or **nature of the relationship**) by studying this oval.

a) If you see a *cigar shape*, then there will be a **linear relationship**.
   This can be :

   i) a **positive linear relationship**; as one variable increases,
      the other tends to increase.

Positive Linear Relationship
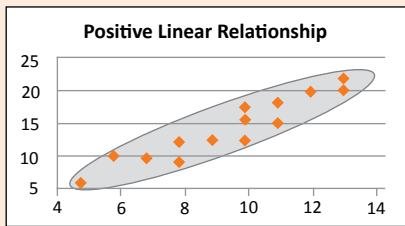
b) If there seems to be a *curved band*, then the relationship
   is **non-linear**.

Non-Linear Relationship
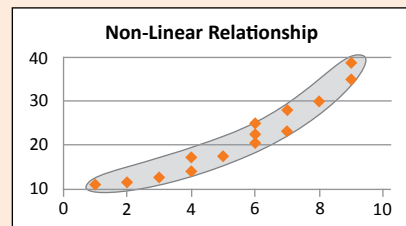
   ii) a **negative linear relationship**; as one variable increases,
       the other tends to decrease.

Negative Linear Relationship
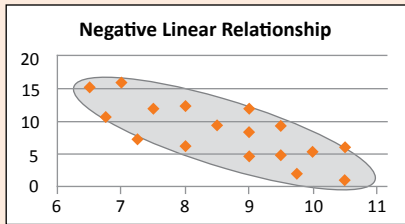
c) If the oval is *almost circular*, then there is **no relationship**.
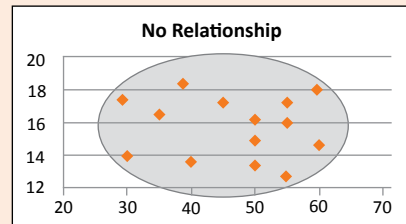
No Relationship

1   Describe the nature of the following relationships :

a)

b)

………………………………………………………………..

………………………………………………………………..

c)

d)

………………………………………………………………..

………………………………………………………………..

15 Analysing Bivariate Data 2

A Features of your Scatter Plot in Context

1   For each of the following discuss features, then draw in an appropriate trend line and discuss any further features, including the strength of the relationship.

Comment on features in the context of the question, refer to the plot to support your answer and give possible reasons for features.

a)

Recovery Time After a Student Runs Different Distances

Recovery Time (sec) vs Distance Run (m)

b)

Shoe Sizes and Heights of Senior Students

Height (cm) vs Shoe Sizes

....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....

## A Making Predictions - continued

For questions d) - f), if it **is sensible**, make and discuss a prediction. If it **is not sensible** to make a prediction, say why this is the case.

d) The time it takes for a drug at a particular concentration to take effect. (See graph page 19, Exercise 1d)

………………………………………………………………..
………………………………………………………………..
………………………………………………………………..
………………………………………………………………..
………………………………………………………………..
………………………………………………………………..
………………………………………………………………..
………………………………………………………………..
………………………………………………………………..

e) How long a student who can run 200 m in a particular time spends on homework. (See graph page 20, Exercise 1e)

………………………………………………………………..
………………………………………………………………..
………………………………………………………………..
………………………………………………………………..
………………………………………………………………..
………………………………………………………………..
………………………………………………………………..
………………………………………………………………..
………………………………………………………………..

f) The weight of a fish with a given length. (See graph page 20, Exercise 1f)

………………………………………………………………..
………………………………………………………………..
………………………………………………………………..
………………………………………………………………..
………………………………………………………………..
………………………………………………………………..
………………………………………………………………..
………………………………………………………………..
………………………………………………………………..
………………………………………………………………..

B Going to Extremes

1 This is the graph for the fertilizer experiment (see page 20). Both the horizontal and vertical axes have been extended.

Crop Yields for Different Applications of Fertiliser

a) The yields for a wide range of application rates need to be predicted. Predict the yield for an application rate of . . .

i) 45 kg/ha. …………………………………………………

ii) 85 kg/ha. …………………………………………………

iii) 110 kg/ha. ………………………………………………

b) Compare these three predictions, discussing how accurate you think they might be. Give reasons for your answer.

………………………………………………………………..
………………………………………………………………..
………………………………………………………………..
………………………………………………………………..
………………………………………………………………..
………………………………………………………………..
………………………………………………………………..
………………………………………………………………..
………………………………………………………………..
………………………………………………………………..
………………………………………………………………..
………………………………………………………………..
………………………………………………………………..
………………………………………………………………..
………………………………………………………………..
………………………………………………………………..
………………………………………………………………..
………………………………………………………………..