Section 3.1.12 Scatterplots

VCAA "Dot Points"
Investigating data distributions, including:

- *scatterplots and their use in identifying and qualitatively describing the association between two numerical variables in terms of direction (positive/negative), form (linear/non-linear) and strength (strong/moderate/weak)*
- *construct scatterplots and use them to identify and describe associations between two numerical variables*

Scatterplots

A **scatter plot** is a graph of plotted points that show the **relationship** between **two sets of data**. Scatterplots are used to determine relationships/associations between two **numerical variables**.

Example 1
Consider the following data, which shows the daily sales of ice creams and the noon temperature for that day.

Temp (°C)	Ice Cream Sales ($)
14.2	215
16.4	325
11.9	185
15.2	332
18.5	406
22.1	522
19.4	412
25.1	614
23.4	544
18.1	421
22.6	445
17.2	408
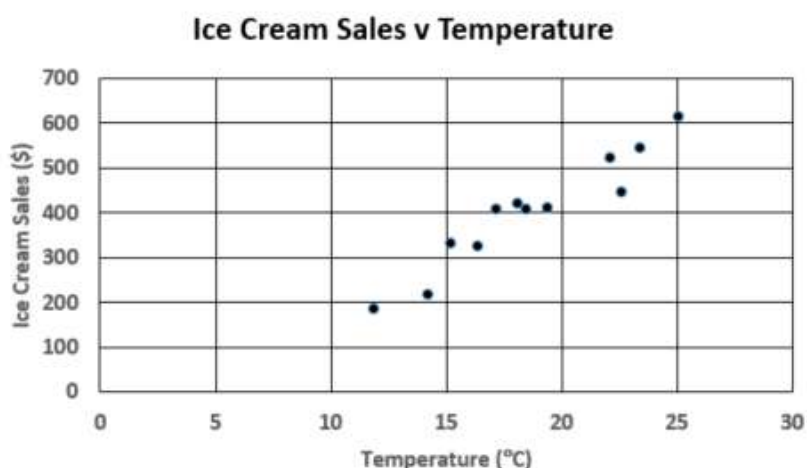
In this example, the ice cream sales **respond** to the temperature of the day, not the other way around. Alternatively, the daily temperature can be used to **explain** ice cream sales.

Accordingly;
- Ice Cream Sales ($) is the **response variable**, &
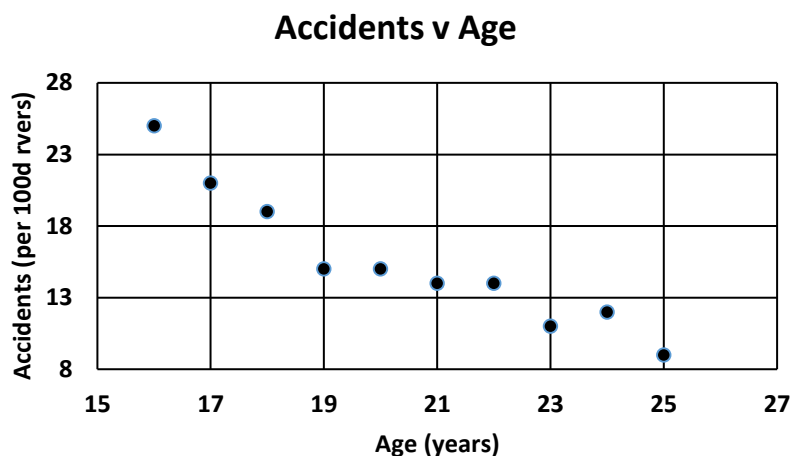- Temperature (°C) is the **explanatory variable**

Recall
When graphing this data upon a scatter plot;
- the **response variable** is plotted upon the **vertical (y) axis**, &
- the **explanatory variable** is plotted upon the **horizontal (x) axis**

Example 2

Consider the following data, which shows the accidents per 100 drivers and the age of the drivers in a given year.

Age (Years)	Accidents (per 100 drivers)
16	25
17	21
18	19
19	15
20	15
21	14
22	14
23	11
24	12
25	9

Accidents v Age

In this example, the number of accidents **respond** to the driver age.
Alternatively, the driver's age can be used to **explain** number of car accidents.

Accordingly;
- Accidents (per 100 drivers) is the **response variable**, &
- Age (years) is the **explanatory variable**

Recall
When graphing this data upon a scatter plot;
- the **response variable** is plotted upon the **vertical (y) axis**, &
- the **explanatory variable** is plotted upon the **horizontal (x) axis**

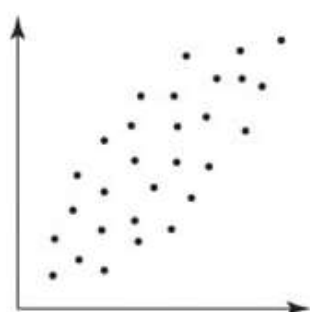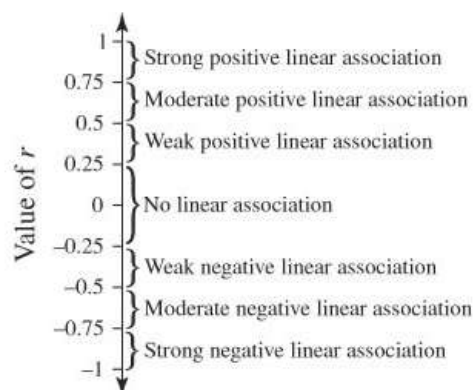When describing the **relationship** between two variables displayed on a scatterplot, we need to comment on:

- the **direction** — whether it is positive or negative
- the **form** — whether it is linear or non-linear
- the **strength** — whether it is strong, moderate or weak
- possible **outliers**
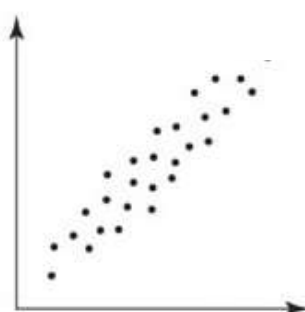
Pearson product-moment correlation coefficient (r)

The **Pearson product-moment correlation coefficient (r)** is a measure of the linear correlation (dependence) between two variables X and Y.

It ranges between values of +1 and −1 inclusive.
- 1 is a perfect positive correlation,
- 0 is no correlation, and
- −1 is a perfect negative correlation.

Value of *r*

1 — } Strong positive linear association
0.75 — } Moderate positive linear association
0.5 — } Weak positive linear association
0.25 —
0 — } No linear association
−0.25 — } Weak negative linear association
−0.5 — } Moderate negative linear association
−0.75 — } Strong negative linear association
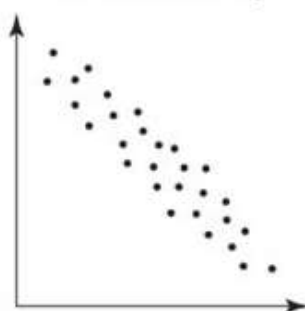−1 —

Weak, positive
linear relationship

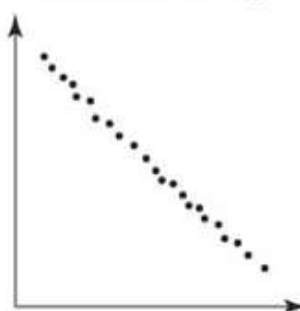Moderate, positive
linear relationship
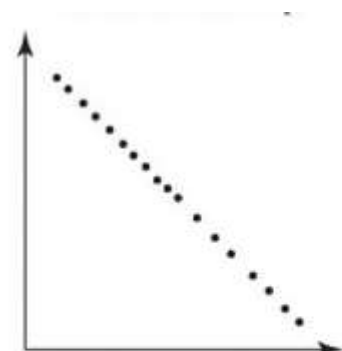
Strong, positive
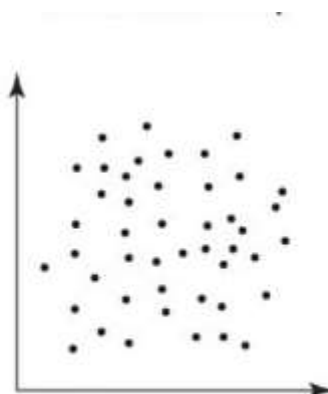linear relationship
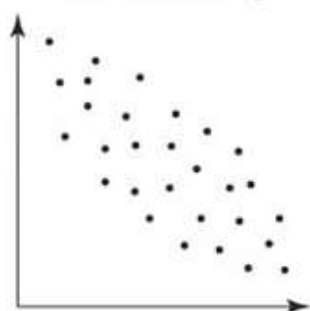
Weak, negative
linear relationship

Moderate, negative
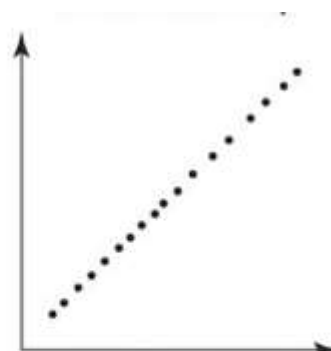linear relationship

Strong, negative
linear relationship

Perfect, negative
linear relationship

No relationship
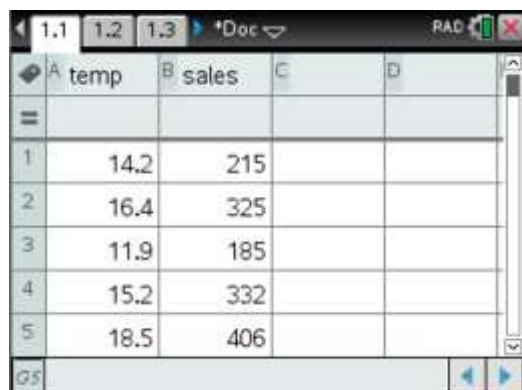
Perfect, positive
linear relationship

NB: A **positive correlation** means that as the X variable increases, so too does the Y variable
A **negative correlation** means that as the X variable increases, the Y variable decreases
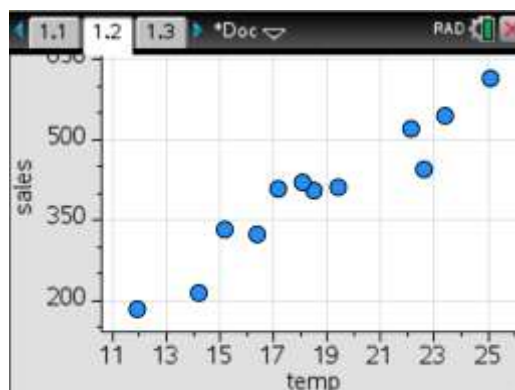
Task 1

Use your TI-Nspire CAS calculator to:
1. Construct a "List & Spreadsheet"
2. Construct a "Data & Statistics"

For the Example.1 data relating to temperature and ice cream sales.

Question:

Classify the association between the variables ice cream sales and daily temperature.
By inspection (i.e. visual appearance) there appears to be a strong, positive, linear association.
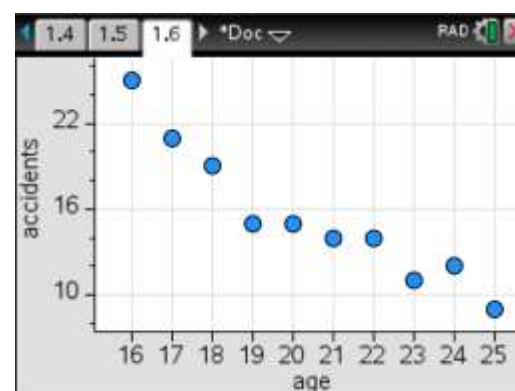

Task 2

Use your TI-Nspire CAS calculator to:
1. Construct a "List & Spreadsheet"
2. Construct a "Data & Statistics"

For the Example.2 data relating to driver age and driver accidents.

Question:

Classify the association between the driver age sales and driver accidents
By inspection (i.e. visual appearance) there appears to be a strong, negative, linear association.