Scatter Plots

- Types of Variables
  - Quantitative (measurements/counts)
  - Qualitative (groups)

The scatter plot is the basic tool used to investigate relationships between two quantitative variables.

What do I see in these scatter plots?

**Mean January Air Temperatures for 30 New Zealand Locations**

- Temperature (°C)
- Latitude (°S)

There appears to be a linear trend. Trend

There appears to be moderate constant scatter about the trend line.

Negative association.

No outliers or groupings visible.

**% of population who are Internet Users vs GDP per capita for 202 Countries**

- Internet Users (%)
- GDP per capita (thousands of dollars)

There appears to be a non-linear trend.

There appears to be non-constant scatter about the trend line.

Positive association.

One possible outlier (Large GDP, low % Internet Users).

**Average Age New Zealanders are First Married**

- Age
- Year

Two non-linear trends (Male and Female).

Very little scatter about the trend lines

Negative association until about 1970, then a positive association.

Gap in the data collection (Second World War).
What do I look for in scatter plots?

**Trend**

Do you see

- a **linear** trend...
  - straight line
  
or
  - a **non-linear** trend?

Do you see

- a **positive** association...
  - as one variable gets bigger, so does the other
  
or
  - a **negative** association?
  - as one variable gets bigger, the other gets smaller

**Scatter**

Do you see

- a **strong** relationship...
  - little scatter
  
or
  - a **weak** relationship?
  - lots of scatter

Do you see

- **constant** scatter...
  - roughly the same amount of scatter as you look across the plot
  
or
  - **non-constant** scatter?
  - the scatter looks like a "fan" or "funnel"

**Anything unusual**

Do you see

- any **outliers**?
  - unusually far from the trend

Do you see

- any **groupings**?
Rank these relationships from weakest (1) to strongest (4):

How did you make your decisions?

The less scatter there is about the trend line, the stronger the relationship is.
Correlation

- Correlation measures the **strength** of the **linear association** between two **quantitative** variables
- Get the correlation coefficient ($r$) from your calculator or computer
- $r$ has a value between -1 and +1:

<table>
<thead>
<tr>
<th>$r$</th>
<th>Points fall exactly on a straight line</th>
<th>No linear relationship (uncorrelated)</th>
<th>Points fall exactly on a straight line</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1</td>
<td>No linear relationship, but there is a relationship!</td>
<td>No linear relationship, but there is a relationship!</td>
<td>No linear relationship, but there is a relationship!</td>
</tr>
</tbody>
</table>

- The correlation coefficient has no units

What can go wrong?

- Use correlation only if you have two **quantitative** variables
  
  There is an association between gender and weight but there isn’t a correlation between gender and weight!

- Use correlation only if the relationship is **linear**
- Beware of outliers!

**Always** plot the data **before** looking at the correlation!
Tick the plots where it would be OK to use a correlation coefficient to describe the strength of the relationship:

- Distances of Planets from the Sun
- Reaction Times (seconds) for 30 Year 10 Students
- Mean January Air Temperatures for 30 New Zealand Locations
- Average Weekly Income for Employed New Zealanders in 2001
What do I see in this scatter plot?

![Graph of Height and Foot Size for 30 Year 10 Students](image)

- **Appears to be a linear trend, with a possible outlier (tall person with a small foot size.)**
- **Appears to be constant scatter.**
- **Positive association.**

What will happen to the correlation coefficient if the tallest Year 10 student is removed? Tick your answer:

(Remember the correlation coefficient answers the question: “For a linear relationship, how well do the data fall on a straight line?”)

- [ ] It will get smaller
- [ ] It won’t change
- **[✓] It will get bigger**

What do I see in this scatter plot?

![Graph of Life Expectancies and Gestation Period for a sample of non-human Mammals](image)

- **Appears to be a strong linear trend.**
- **Outlier in X (the elephant).**
- **Appears to be constant scatter.**
- **Positive association.**

What will happen to the correlation coefficient if the elephant is removed? Tick your answer:

- **[✓] It will get smaller**
- [ ] It won’t change
- [ ] It will get bigger
Using the information in the plot, can you suggest what needs to be done in a country to increase the life expectancy? Explain.

**Perhaps if you have less people per doctor (i.e., more doctors per person), then the life expectancy will increase.**

Using the information in this plot, can you make another suggestion as to what needs to be done in a country to increase life expectancy?

**It looks like if you decrease the number of people per television (i.e., have more TVs per person), then the life expectancy will increase!**

Can you suggest another variable that is linked to life expectancy and the availability of doctors (and televisions) which explains the association between the life expectancy and the availability of doctors (and televisions)?

Some measure of wealth of a country. E.g., average income per person or GDP.
Causation

Two variables may be strongly associated (as measured by the correlation coefficient for linear associations) but may not have a cause and effect relationship existing between them. The explanation maybe that both the variables are related to a third variable not being measured – a "lurking" or "confounding” variable.

These variables are positively correlated:

- Number of fire trucks vs amount of fire damage
- Teacher’s salaries vs price of alcohol
- Number of storks seen vs population of Oldenburg Germany over a 6 year period
- Number of policemen vs number of crimes

*Only* talk about causation if you have well designed and carefully carried out *experiments*.

Data Sources

http://www.niwa.cri.nz/edu/resources/climate
http://www.cia.gov/cia/publications/factbook
http://www.stats.govt.nz
http://www.censusatschool.org.nz