The data are from the paper:

Exploring Relationships in Body Dimensions

Grete Heinz and Louis J. Peterson
San José State University

Roger W. Johnson and Carter J. Kerk
South Dakota School of Mines and Technology

Topics from GCE AS and A Level Mathematics covered in Sections 3:

- Interpret scatter diagrams and regression lines for bivariate data, including recognition of scatter diagrams which include distinct sections of the population.

- Understand informal interpretation of correlation.

- Recognise and interpret possible outliers in data sets and statistical diagrams.

- Be able to clean data (including dealing with missing data, errors and outliers).

- Use samples to make informal inferences about the population.

The problem

Some people can be sensitive about having their waist measured.
Investigate if there is a relationship between waist girth and wrist girth.
Could this relationship be used to predict someone's waist girth based on their wrist girth?

Data collection

The data are in the Excel spreadsheet '**Adult Measurements.xlsx**'.

The worksheet named **Data** has15 variables in the columns with column headers and data for 507 respondents in the rows.

The worksheet named **Information** has the definitions for the variables and coding for gender.

Waist girth is the narrowest part of torso below the rib cage. The average of contracted and relaxed position was recorded.

Wrist girth is the average of right and left minimum girths

From *Section 1*:

- There are no missing values.
- The measurements can be considered accurate.
- The target population is American adults who are physically active.
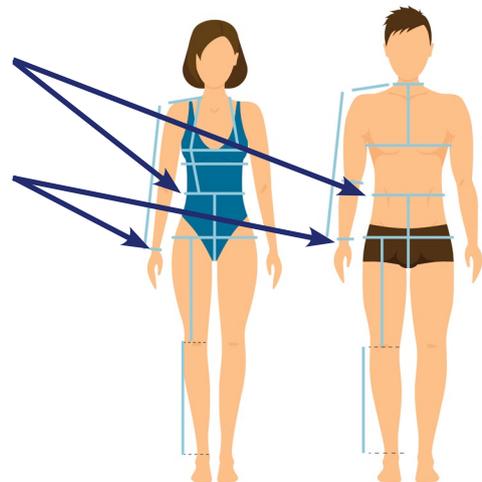
Some definitions

Bivariate data
Bivariate data are data that have two variables i.e. where each member of the sample requires the values of two variables.

Independent variable (Explanatory variable)
In bivariate data, if one of the variables is controlled (or explains the other variable), it is known as the independent (or explanatory variable). Independent variables are the variables that the experimenter changes to test the dependent variable. For example, when taking a temperature every minute, time is the independent variable.

Dependent variable (Response variable)
A dependent variable is a variable whose value depends on the value of another variable. In the example above, temperature is the dependent variable.
The dependent variable is usually plotted on the vertical axis.
Note: if you are using a regression model to predict a value, the variable for the value you wish to predict should be the y variable and plotted on the vertical axis of a scatter diagram.

Linear Correlation
For bivariate data, if all the points in a scatter diagram seem to lie close to a straight line of best fit, there is a linear correlation between the two variables.

For bivariate variables x and y, there is:
- a positive linear correlation if y tends to increase as x increases;

- a negative linear correlation if y tends to increase as x decreases;

- no linear correlation if there is no linear relationship between x and y.

Pearson's product−moment correlation coefficient
This provides a standardised measure of linear correlation. Its value lies between -1 and +1.

- A value of +1 means perfect positive correlation – in this case, all the points on the scatter diagram would lie in a straight line with a positive gradient.

- A value of -1 means perfect negative correlation – in this case, all the points on the scatter diagram would lie in a straight line with a negative gradient.

- A value close to 0 means there is no correlation.

Process

Open the Excel workbook '**Adult Measurements.xlsx**'.

Select the worksheet **WaistWristGirth**.

To plot a scatter diagram in Excel

When plotting a scatter diagram, Excel will always plot the variable in left hand column on the x-axis and the variable in the right hand column on the y-axis.

In this case the dependent variable '**Waist girth**', should be plotted on the y-axis.
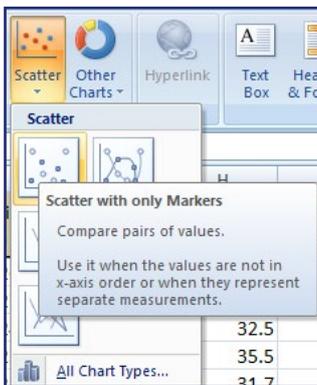
- Select columns **A** and **B**

- Select the **Insert** tab, then select **Scatter** and select **Scatter with only Markers.**

The default graph needs formatting.
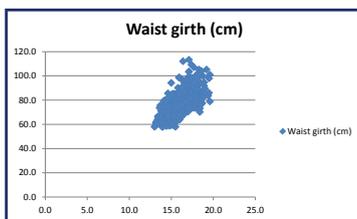
Delete the legend:
- Click on the **legend** and **delete**.

Add a title:
- Click on the title and type **Waist girth versus Wrist girth for a sample of 507 adults** then **Enter**.

Add a vertical axis title:
- Click on the chart then select the **Layout** tab. Select **Axis Titles**, then **Primary Vertical Axis Title**, then **Vertical Title** and type **Waist girth (cm)** and **Enter**.

Change the alignment of the text in the vertical title:
- Right click on the **vertical title**, select **Format Axis Title**, then **Alignment,** select one of the options from the drop down list next to **Text direction** and **Close**.
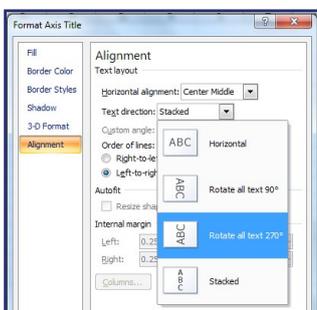
Add a horizontal axis title:
- In the **Layout** tab select **Axis Titles,** then **Primary Horizontal Axis Title,** then **Title Below Axis** and type **Wrist girth (cm)** and **Enter**.
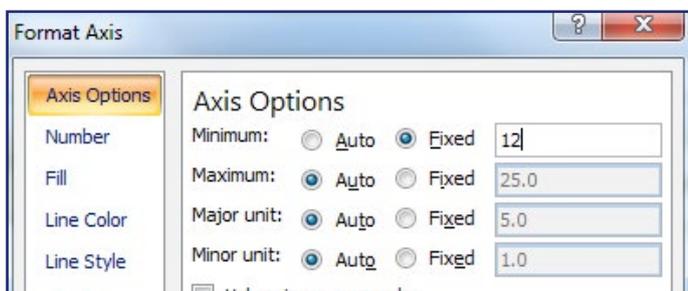
Add a chart border:
- Right click on the **chart**, select **Format Plot Area**, then **Border Color,** select **Solid line**, select the colour required from the drop down box next to **Color** and **Close**.

Add major and minor gridlines:
- Right click on the *x*-axis then **Add Major Gridlines**

- Similarly, add minor gridlines to the *x*-axis

- Add major and minor gridlines to the *y*-axis.

Adjust the scale on the *x*-axis:

Format Axis			?	X
Axis Options	Axis Options			
Number	Minimum:	○ Auto ● Fixed	12	
Fill	Maximum:	● Auto ○ Fixed	25.0	
Line Color	Major unit:	● Auto ○ Fixed	5.0	
Line Style	Minor unit:	● Auto ○ Fixed	1.0	

- Right click on the *x*-axis

- Select **Format Axis**

- In **Axis Options** click on **Fixed** next to **Minimum** and type in **12**

- **Save your work**

Waist Girth vs Wrist Girth for a sample of 507 US adults
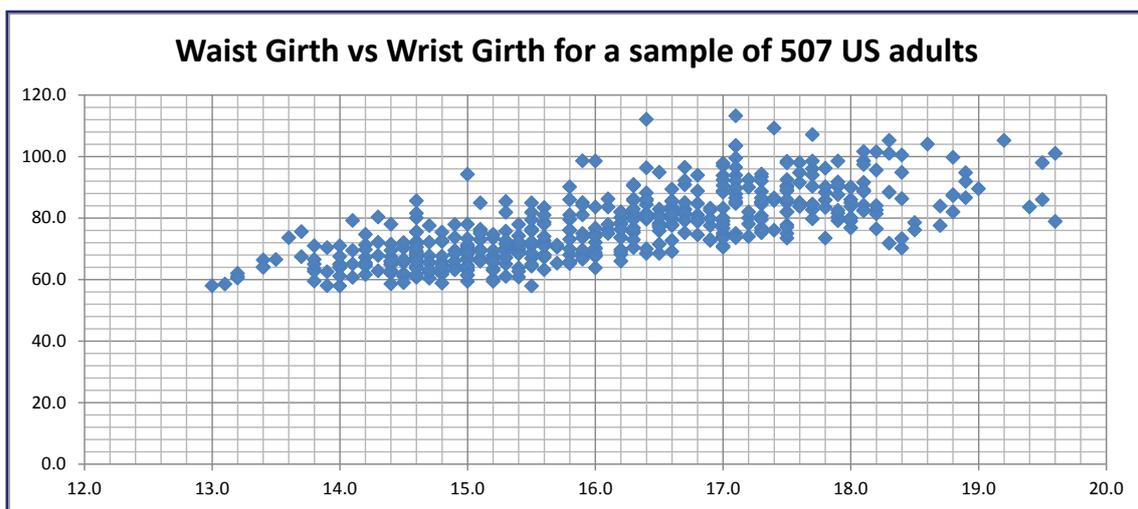
**1. Comment on the correlation between the two variables.**

*The scatter diagram suggests there is a (strong) positive correlation between **waist girth** and **wrist girth** for American adults who are physically active. A large wrist girth suggests a large waist girth and vice versa.*

Students need to comment on the correlation in context

To add a line of regression in Excel (This is called a trend line in Excel)
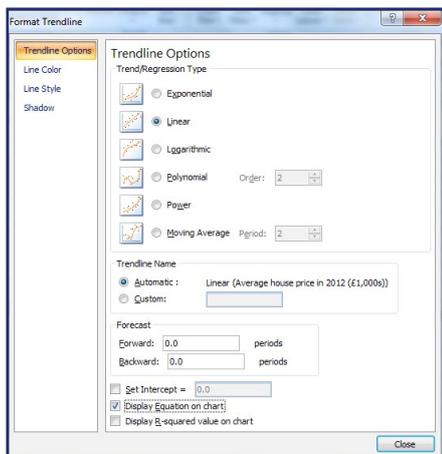
- Right click on any **data point**

- Select **Add Trendline**

- Select **Linear** then **Display Equation on Chart** and **Close**

- Drag the equation for the line of regression (trend line) so it is clearly visible on the chart

- Reduce the number of decimal place values to one and format the text.

Waist Girth vs Wrist Girth for a sample of 507 US adults

$y = 5.8x - 16.6$
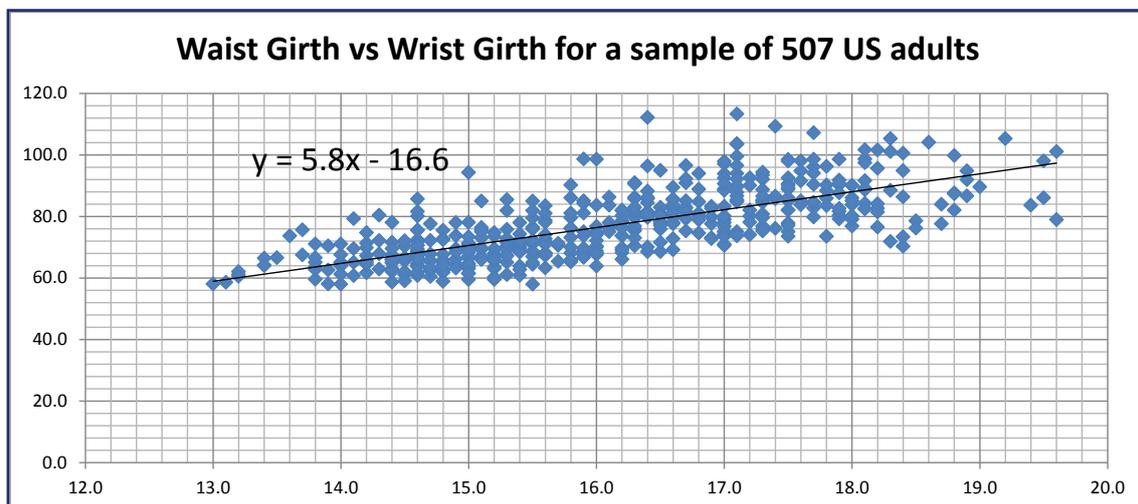
2. Write the line of regression using the names of the variables.

Waist girth = 5.8 x Wrist girth − 16.6

3. Interpret the gradient of the line of regression for 'Waist girth' against 'Wrist girth'.

The regression model suggests that on average the waist girth for American adults who are physically active is approximately 6 times their wrist girth.

4. State, with a reason, if the intercept should be set to zero.

The intercept is the waist girth measurement when the wrist girth measurement is zero. It seems reasonable to set the intercept to zero.

To set the intercept to zero
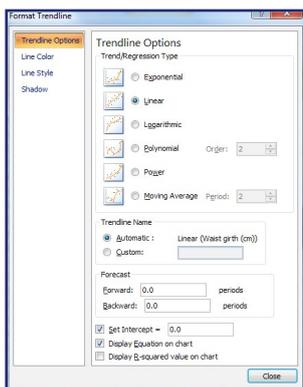
- Right click on the **trend line**

- Select **Format trendline**

- Tick **Set Intercept = 0.0** then **Close**.

- **Save your work**

---

5. Does this regression model seem to fit the data?

From the scatter diagram the regression model with the intercept of -16.6 seems to fit the data better. This is an example of where the model should not be used to predict values outside the range of the data. The relationship between waist girth and wrist girth may be different for children who, of course, will be smaller.

6. Use the regression model (line of regression) to predict the waist girth of an American adult who is physically active, with a wrist girth of 15 cm.

Waist girth = 5.8 x 15 − 16.6 = 70.4 cm

7. Comment on the accuracy of the predicted waist girth in question 6.
Not very accurate as there is a lot of scatter at a wrist girth of 15 cm, (minimum and maximum measurements of waist girth at 15 cm wrist girth are 59.4 cm and 94.2 cm repectively).

8. The scatter diagram shows the relationship between *waist girth* and *wrist girth* for a sample of 507 American adults who are physically active, split by gender.
**Is the relationship between *waist girth* and *wrist girth* stronger for males than females?**
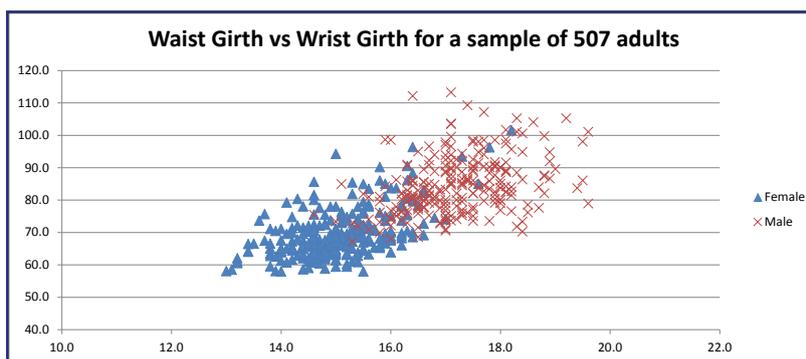
There seems to be a difference between the ranges for the waist girth and wrist girth for males and females. However the gradients within each of the ranges seem to be similar. The scatter seems to be greater for males with larger wrist girths.

To plot the scatter diagram 'Waist girth' vs 'Wrist girth' for a sample of 507 American adults who are physically active, split by gender
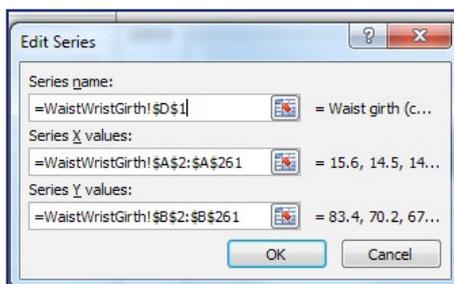
In the worksheet **WaistWristGirth**:
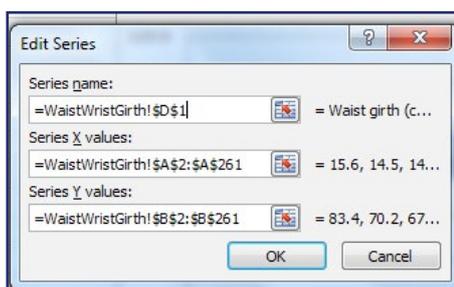
- Sort the data according to column **C** (Gender)
- Plot a scatter diagram for females only cells **A1:B261**
- Move the chart to the top of the worksheet, either drag or **Cut** and **Paste**.

Add a label for females:
- Right click on the **chart,** select **Select Data**
- Select **Edit,** under **Series name** select **D1** and **OK**.

Add males' data and label:
- Right click on the **chart**, select **Select Data** and **Add**
- Under **Series name** select **E1**
- Under **Series X values** select **A262:A508**
- Under **Series Y values** select **B262:B508**
- Delete =\{y\}
- Click **OK**
- Format the chart.

Extension

The problem

1. Investigate the relationship between **wrist girth** and **ankle girth** for the sample of physically active American adults.
2. Identify any outliers and decide whether these should remain in the data set.
3. Comment on the correlation between the two variables.
4. Find the equation of regression for **wrist girth** against **ankle girth**.
5. Discuss if the regression model could be used to predict the wrist girth given a person's ankle girth of 28 cm.
6. Is the regression model the same for males and females?

Process

Data are in '**Adult Measurements.xlsx',** worksheet **WaistAnkleGirth**.

Plot the appropriate graph to investigate the relationship between the two variables.

1. Investigate the relationship between *wrist girth* and a*nkle girth* for the sample of physically active American adults.

Wrist girth vs Ankle girth for a sample of 507 US adults

$y = 0.56x + 3.72$
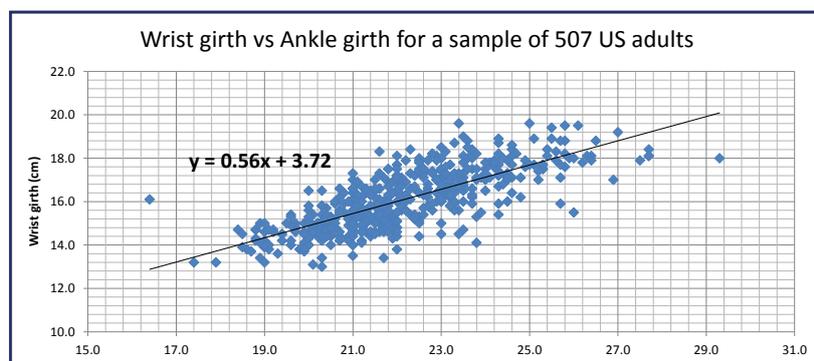
2. Identify any outliers and decide whether these should remain in the data set.
There seems to be one outlier. On inspection this could be an error or an inaccurate result. It is difficult to tell. Removing this does not affect the regression model very much
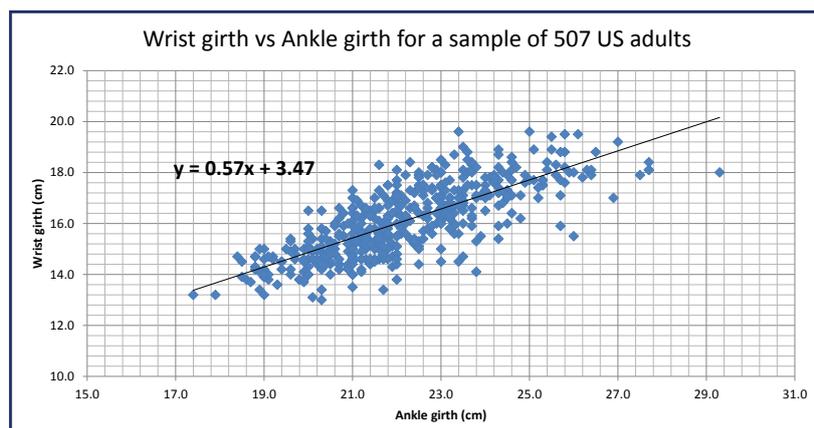
Wrist girth vs Ankle girth for a sample of 507 US adults

$y = 0.57x + 3.47$

3. Comment on the correlation between the two variables

*The scatter diagram suggests there is a (strong) positive correlation between **wrist girth** and **ankle girth** for American adults who are physically active. A large ankle girth suggests a large wrist girth and vice versa.*

4. Find the equation of regression for *wrist girth* against *ankle girth*.

The regression model with the outlier:
 ***Wrist girth** = 0.6 × **Ankle girth** + 3.7*

The regression model without the outlier:
 ***Wrist girth** = 0.6 × **Ankle girth** + 3.5*

## 5. interpret the gradient of the equation of regression

*The regression model suggests that **on average** the wrist girth for American adults who are physically active is approximately 0.6 times their ankle girth.*

6. Discuss if the regression model could be used predict ankle girth given a person's wrist girth of 28 cm

The prediction would not be very reliable as there are very few points plotted near to a wrist girth of 28 cm. This value is very nearly outside the range of the data.

7. Plot the scatter diagram '*Waist girth*' against '*Ankle girth*' for the sample of 507 American adults who are physically active, split by gender.

Comment on the relationships
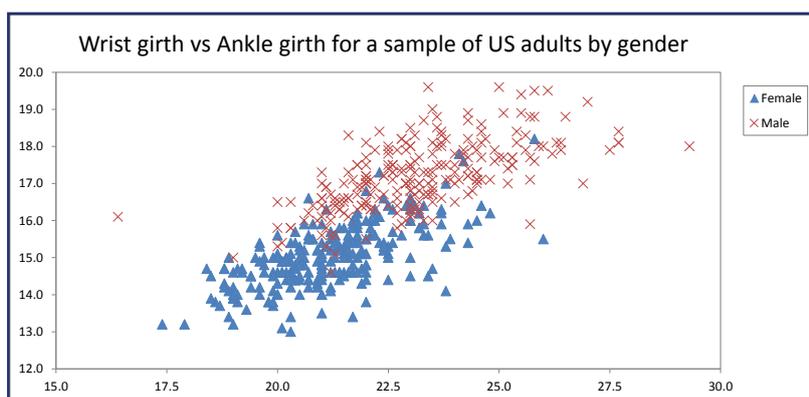
There seems to be a difference between the ranges for the ankle girth and wrist girth for males and females. However the gradient within each of the ranges seems to be similar.