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 2:

- Interpret diagrams for single-variable data.

- Interpret measures of central tendency and variation, extending to standard deviation.

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

- Select or critique data presentation techniques in the context of a statistical problem.

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

The problem

To investigate the hypothesis:
Male and female waist girth measurements are the same.

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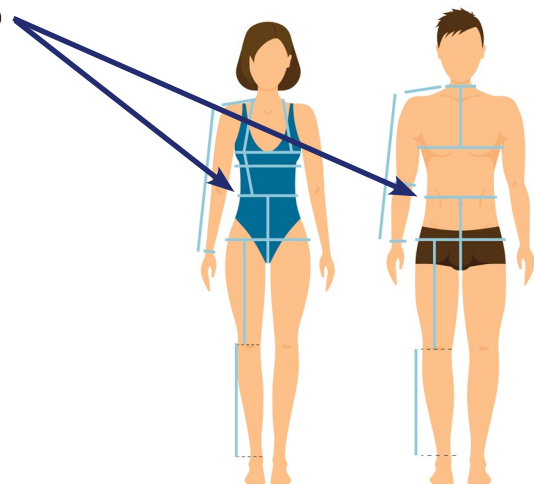
The worksheet named **Information** has the definitions for the variables and coding for gender.

Waist girth is the narrowest part of torso below the rib cage.

The average of contracted and relaxed position was recorded.

From **Section 1**:

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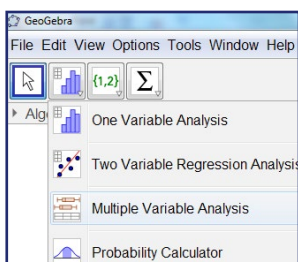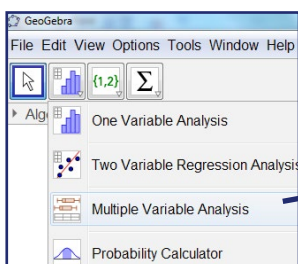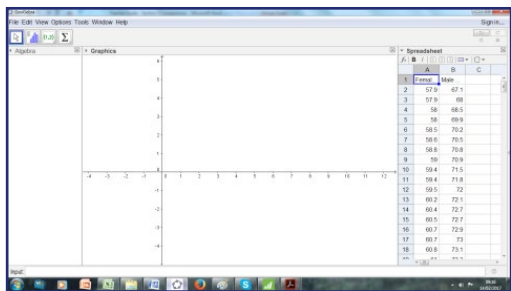- The target population is American adults who are physically active.

Process

Distributions of variables can be compared using box plots or histograms. Whichever graphs are plotted it is important that the same scales are used.

It is tricky to plot box plots in Excel (see '**Box plots in Excel.docx**' for instructions on how to do this).

Plot box plots of '**Waist girth**' for males and females in **GeoGebra**..

To plot a box plot in GeoGebra

- **Open GeoGebra** select **View** then **Spreadsheet**

- Open the Excel spreadsheet '**Adult Measurements.xlsx**'

- In the worksheet **MaleFemaleWaist** select columns **A** and **B**

- Right click **Copy**

- In **GeoGebra** select **A1**, right click **Paste**

In **GeoGebra**:
- Select columns **A** and **B**
- Click on the histogram icon
- Select **Multiple Variable Analysis**

- Click on the ⚙ icon, select **Use Header as Title**, then **Analyse**

- Click and drag the borders of the graph window to make the box plots bigger
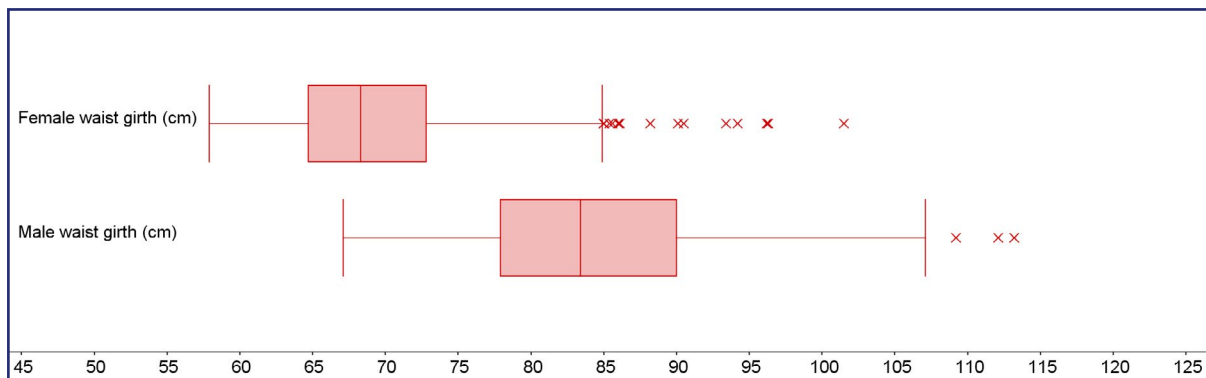
Right click on the graph and you can either:
- **Copy to Graphics View** (this is within GeoGebra)
- **Copy to Clipboard**
- **Export as Picture.**

Below is the box plot copied to Clipboard and pasted into Word. A title was added in Word.
(Note: Clicking on the Σ icon will give summary statistics for the two variables.)

GeoGebra shows outliers as small crosses.

Waist girth for a sample of 260 US females and 247 US males

Comment on the outliers.
There do not seem to be any lower outliers for either variable.
There are many upper outliers for both variables.

Use the formulae in the table to calculate the lower and upper limits for outliers.

Outlier limit	Female waist girth (cm)	Male waist girth (cm)
Lower limit = $Q_1 - 1.5 \times IQR$	*52.55*	*59.75*
Upper limit = $Q_3 + 1.5 \times IQR$	*84.95*	*108.15*

Do the calculated limits agree with the outliers on the box plots?
Since the minimum waist girth for females is 57.9 cm, there are no lower outliers. However the maximum waist girth is 101.5 cm so there will be upper outliers. The situation is similar for males.

To decide if these outliers are errors, anomalies or accurate data, in the worksheet **Male** sort the data and inspect upper outliers for '**Male waist girth**'.

Similarly, in the worksheet **Female** sort the data and inspect upper outliers for '**Female waist girth**'.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Male shoulder girth (cm)	Male waist girth (cm)	Male navel girth (cm)	Male hip girth (cm)	Male thigh girth (cm)	Male bicep girth (cm)	Male forearm girth (cm)	Male knee girth (cm)	Male calf girth (cm)	Male ankle girth (cm)	Male wrist girth (cm)	Male age (years)	Male weight (kg)	Male height (cm)	Gender
2	127.0	113.2	102.9	107.9	57.7	37.4	28.9	37.8	37.4	24.1	17.1	28	102.5	177.8	Male
3	119.7	112.1	105.9	106.3	56.9	35.7	27.8	37.3	36.6	22.7	16.4	46	92.7	179.1	Male
4	110.6	109.2	104.4	101.7	56.4	34.0	29.2	39.3	38.5	24.3	17.4	55	94.1	185.4	Male
5	127.0	107.1	107.2	108.3	67.0	39.6	30.6	40.0	39.6	24.6	17.7	33	101.6	192.0	Male
6	134.8	105.2	105.0	115.5	69.9	39.4	32.1	42.2	47.7	27.0	19.2	27	116.4	177.8	Male
7	118.8	105.2	103.4	108.1	60.5	38.0	30.2	36.9	37.7	21.6	18.3	42	85.9	166.4	Male

Upper outliers for '**Male waist girth**'

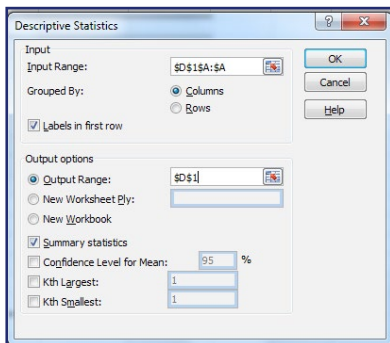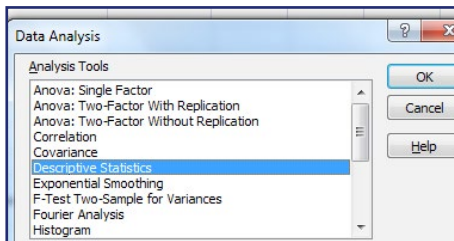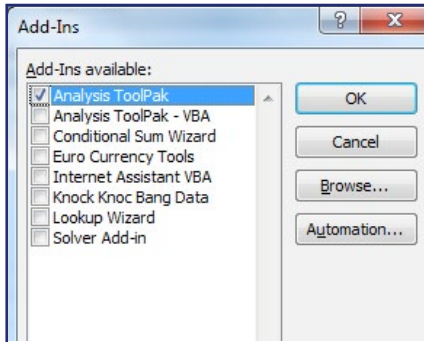|	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Female shoulder girth (cm)	Female waist girth (cm)	Female navel girth (cm)	Female hip girth (cm)	Female thigh girth (cm)	Female bicep girth (cm)	Female forearm girth (cm)	Female knee girth (cm)	Female calf girth (cm)	Female ankle girth (cm)	Female wrist girth (cm)	Female age (years)	Female weight (kg)	Female weight (cm)	Gender
2	129.5	101.5	105.5	114.0	70.0	40.3	30.8	45.6	45.0	25.8	18.2	34	105.2	172.7	Female
3	110.8	96.3	104.2	102.6	57.5	34.3	26.5	38.3	41.1	22.9	16.4	41	80.0	161.0	Female
4	127.1	96.2	121.1	128.3	72.3	35.9	30.6	49.0	45.4	24.1	17.8	43	104.1	165.1	Female
5	115.6	94.2	110.5	103.5	57.8	35.7	25.6	34.8	35.5	19.5	15.0	37	81.8	162.6	Female
6	111.0	93.4	111.1	109.8	67.7	35.7	28.6	41.5	38.6	22.3	17.3	41	84.5	162.6	Female
7	115.5	90.5	103.5	108.1	61.1	33.6	26.6	37.2	35.8	22.6	16.3	40	75.2	160.2	Female
8	112.1	90.1	90.1	95.2	53.8	32.1	25.8	34.7	37.2	21.5	15.8	32	67.2	156.5	Female
9	115.2	88.2	106.5	107.9	63.1	30.3	25.2	39.2	38.4	23.1	16.4	26	86.3	165.1	Female
10	106.4	86.1	107.4	112.0	69.4	31.7	25.5	39.4	38.7	23.7	16.3	33	80.5	160.0	Female
11	104.0	86.0	107.1	112.1	64.4	32.3	26.4	35.7	37.0	21.4	15.8	40	76.8	157.5	Female
12	99.5	85.6	111.7	112.1	64.3	31.3	25.0	35.5	38.0	21.0	14.6	36	76.4	167.6	Female
13	102.1	85.4	105.2	109.5	68.3	29.7	24.5	40.3	37.0	22.4	15.3	24	73.6	176.5	Female
14	114.0	85.0	92.5	106.4	65.6	31.6	26.6	39.6	38.9	23.3	15.9	22	83.0	176.5	Female
15	117.1	85.0	91.2	111.0	68.0	34.4	28.3	40.1	40.0	24.2	17.6	29	87.8	176.5	Female
16	109.8	84.9	98.0	102.4	59.9	32.0	25.7	37.6	33.7	21.6	15.5	40	70.7	166.4	Female
17	108.9	83.6	108.6	108.3	63.0	31.9	26.0	38.1	36.4	24.5	16.0	29	81.8	182.9	Female
18	112.2	83.6	94.2	104.1	61.5	31.0	25.4	37.2	35.2	22.3	16.1	31	73.2	170.0	Female
19	110.7	83.4	100.5	107.5	67.4	34.2	28.3	40.9	41.2	23.4	15.6	26	80.9	165.1	Female

Upper outliers for '**Female waist girth**'

State, with a reason, whether you are removing any of the outliers.
*No data are removed from either the male or female samples as the entries do not seem to be anomalies or errors. The female in row 2 has maximum or near maximum measurements for each variable except for '**Age**'. This female's data are not deleted as the measurements seem consistent.*

Summary statistics table in Excel

To find summary statistics in Excel

- First make sure the **Data Analysis Toolpak** is in Excel

- Press at the same time **Alt T I** (i.e. Alt and capital T and capital I)

- The **Add-Ins** menu will appear.

- Tick **Analysis ToolPak** then **OK**.

- **Data Analysis** will appear in the **Data** tab.

- Select the worksheet **MaleFemaleWaist**

- Select column **A** (Female waist girth (cm)), then **Data Analysis** (in the **Data** tab)

- Then **Descriptive Statistics, OK**

- **Input Range** column **A** (cells containing the data)

- **Labels in first row** ticked (Excel recognises the first row as headers)

- **Output Range D1** (where you want the output)

- **Summary Statistics** ticked, **OK**

Similarly for the Male waist girth:

- Select column **B** (Male waist girth (cm)), then **Data Analysis** (in the **Data** tab)
- Then **Descriptive Statistics, OK**
- **Input Range** column **B**
- **Labels in first row** ticked
- **Output Range F1**
- **Summary Statistics** ticked, **OK**

Express all values, except for count, to 2 decimal place values:

- Select **E3:E14** in the **Home** tab click on the [icon] until the values are expressed to 2 decimal place values. Do the same for **G3:G14**.

Complete the table

	Female waist girth (cm)	Male waist girth (cm)
Mean	69.80	84.53
Standard Error	0.47	0.56
Median	68.30	83.40
Mode	65.00	75.00
Standard Deviation	7.59	8.78
Sample Variance	57.57	77.13
Kurtosis	2.03	0.21
Skewness	1.25	0.64
Range	43.60	46.10
Minimum	57.90	67.10
Maximum	101.50	113.20
Sum	18148.90	20879.70
Count	260	247

Note: the lower and upper quartiles are not included in this table. These are calculated using the Quartile function, =QUARTILE(range,1) for lower quartile and =QUARTILE(range,3) for upper quartile. These values, and more, can be calculated using Excel functions. See Section 1 for details.

Select the appropriate statistics from the table.
Not all the statistics in the table are needed. In this case only the means, medians, standard deviations, minimums and maximums are needed. It is important for students to select the appropriate statistics.

Report

Back to the problem.

Are male and female waist girth measurements the same?
The box plots, medians and means suggest that the waist girths for males, on average, are larger than those for females. However, there is more variability in the waist girths for males than females.

Note: this conclusion must not state 'the box plots clearly show that the waist girth for males is larger than for females' as a sample is being analysed. Also the expression 'on average' is really important.

Extension

The problem
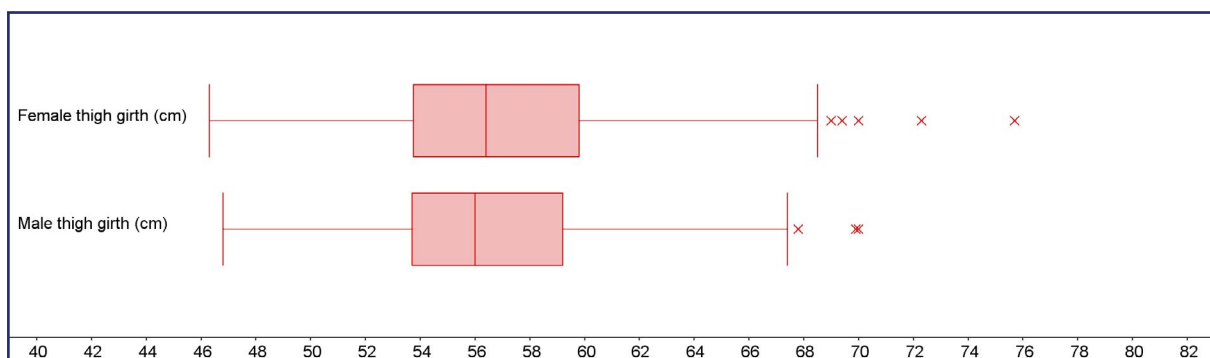Investigate if male and female thigh girths are the same.

Data collection
The data are in the Excel spreadsheet 'Adult Measurements.xlsx'. Worksheet 'MaleFemaleThigh'.

Process

Plot the appropriate graphs to compare the distributions of 'Thigh girth' for males and females.

Thigh girth for a sample of 260 US females and 247 US males

Calculate the appropriate summary statistics.

	Female thigh girth (cm)	Male thigh girth (cm)
Mean	*57.20*	*56.50*
Median	*56.40*	*56.00*
Standard Deviation	*4.64*	*4.25*
Range	*29.40*	*23.20*
Minimum	*46.30*	*46.80*
Maximum	*75.70*	*70.00*
Sum	*14870.90*	*13955.00*
Count	*260*	*247*

or

Statistic	Female (cm)	Male (cm)
Mean	*57.20*	*56.50*
Standard deviation	*4.64*	*4.25*
Minimum	*46.30*	*46.80*
Q1	*53.78*	*53.70*
Q2	*56.40*	*56.00*
Q3	*59.80*	*59.20*
Maximum	*75.70*	*70.00*
IQR	*6.02*	*5.5*

Check for outliers.

Outlier limit	Female thigh girth (cm)	Male thigh girth (cm)
Lower limit = $Q_1 − 1.5 × IQR$	*44.75*	*45.45*
Upper limit = $Q_3 + 1.5 × IQR$	*68.83*	*67.45*

Since the minimum thigh girth for females is 46.3 cm, there are no lower outliers. However the maximum thigh girth is 75.70 cm so there will be upper outliers. The situation is similar for males.

Inspect any outliers in the data set.
No data are removed as for the outliers all the measurement for the other variables seem consistent.

Report
The box plots (or histograms) and medians (or means) suggest that, on average, there is no difference between the thigh girths for male and females. Furthermore, the two distributions seem of similar shape.

Note: this conclusion must not state 'the box plots (or histograms) and medians show that there is no difference between the thigh girths for male and females' .