Examining Continuous Variables in SPSS (Quiz)

Descriptive statistics for all variables practical questions

The dataset we are using is an excerpt from a cut-down dataset drawn from the Living Costs and Food Survey 2013, available from the UK Data Service: http://doi.org/10.5255/UKDA-SN-7932-2. You will be exploring the characteristics of two variables; household size and total household income (in pounds per week). No conditions are required to use the data; however, respondents are promised that their data will be kept confidential. As a result, high values of both variables are grouped together to prevent households being identified by their large household sizes or unusually high income. This protects respondents, but it also affects the quality of the results produced in this workbook. Users who wish to use better quality data are encouraged to explore the full data from the Living Costs and Food Survey, which is available through the UK Data Service (http://doi.org/10.5255/UKDA-SN-7702-1), for which users need to register and adhere to some conditions of use.

Use the Frequencies window in SPSS to answer the following questions about **income**:

- Question: What is the mean for income?
- Question: What is the median for income and is it bigger than the mean?
- Question: What is the mode for income and how is it calculated?
- Question: What is the standard deviation for income and what does it tell us?
- Question: What is the smallest value for income?
- Question: What is the largest value for income?
- Question: What are the lower and upper limits of the interquartile range for income and how wide is it?

Use the Frequencies window in SPSS to answer the following questions about **hhsize**:

- Question: What is the mean for hhsize?
- Question: What is the median for hhsize and is it bigger than the mean?
- Question: What is the mode for hhsize and how is it calculated?
- Question: What is the standard deviation for hhsize and what does it tell us?
- Question: What is the smallest value for hhsize?
- Question: What is the largest value for hhsize?
- Question: What are the lower and upper limits of the interquartile range for hhsize and how wide is it?

Use the Frequencies window to construct a histogram in SPSS to answer the following questions about **income**:

- Question: What does the histogram show?

Use the Frequencies window to construct a histogram in SPSS to answer the following questions about **hhsize**:

- Question: What should we look for in this histogram?

Use SPSS to construct a boxplot and answer the following questions about **income**:

- Question: From the boxplot what is the median (approximately) for income?
- Question: What is the lower quartile (approximately) for income?
- Question: What is the upper quartile (approximately) for income?

Use SPSS to construct a boxplot and answer the following questions about **hhsize**:

- Question: From the boxplot what is the median (approximately) for hhsize?
- Question: What is the lower quartile (approximately) for hhsize?
- Question: What is the upper quartile (approximately) for hhsize?

Solutions to descriptive statistics for all variables practical questions

The SPSS instructions for the first table of descriptive statistics are as follows:

1. Select **Frequencies** from the **Descriptive Statistics** submenu available from the **Analyze** menu.
2. Copy the **income** variable into the **Variable(s)** box.
3. Click on the **Statistics** button to go to the statistics screen.
4. Here we need to select ALL the summary statistics that we are interested in looking at.
5. Select **Mean, Median** and **Mode** from under **Central Tendency**.
6. **Select Std. deviation, Variance, Range, Minimum** and **Maximum** from under **Dispersion**.
7. **Finally Select Quartiles** from under **Percentile Values.**
8. Click on the **Continue** button to return to the main window.
9. Click on the **OK** button to produce the tables required.

- Question: What is the mean for **income**?
- Question: What is the median for **income** and is it bigger than the mean?
- Question: What is the mode for **income** and how is it calculated?
- Question: What is the standard deviation for **income** and what does it tell us?
- Question: What is the smallest value for **income**?
- Question: What is the largest value for **income**?
- Question: What are the lower and upper limits of the interquartile range for **income** and how wide is it?

Solutions: The first tabular output from SPSS is as follows:

Statistics

income

N	Valid	5144
	Missing	0
Mean		620.4336
Median		563.1500
Mode		1184.99
Std. Deviation		359.15568
Variance		128992.802
Range		1184.99
Minimum		.00
Maximum		1184.99
Percentiles	25	313.9456
	50	563.1500
	75	928.1385

The answers are as follows:

- For the variable **income** the arithmetic mean (or average) value is 620.4336 for the data.
- The median or middle value is 563.1500. This is smaller than the mean so if there is any skew to the distribution it will likely be positive.
- The mode or most frequent value which takes value 1184.99. SPSS calculates this by looking at the frequencies of each possible value so the mode is probably more useful for categorical data. This is an artefact caused by a large number of cases with incomes of £1185 per week or more being grouped together.
- The standard deviation of **income** takes value 359.15568. Typically 95% of observations will lie within 2 standard deviations of the mean i.e. between -97.878 and 1338.745.
- The smallest value observed is .00.
- The largest value observed is 1184.99 (remember, this was set to be the maximum by the creators of the dataset).
- The inter-quartile range runs from the lower quartile to the upper quartile. The lower (25%) quartile takes value 313.9456 meaning that 25% of observations are below this value. Conversely 25% of observations are above the upper (75%) quartile which takes value 928.1385. Therefore the interquartile range is of length 928.1385 - 313.9456 = 614.192886353.

The SPSS instructions for the second table of descriptive statistics are as follows:

1. Select **Frequencies** from the **Descriptive Statistics** submenu available from the **Analyze** menu.
2. Remove the **income** variable from the **Variable(s)** box.
3. Copy the **Household size, number of people in household (recoded) formerly A049r[hhsize]** variable into the **Variable(s)** box.
4. The **Statistics** options will be remembered so do not need adding again.
5. Click on the **OK** button to produce the tables required.

- Question: What is the mean for **hhsize**?

- Question: What is the median for **hhsize** and is it bigger than the mean?

- Question: What is the mode for **hhsize** and how is it calculated?

- Question: What is the standard deviation for **hhsize** and what does it tell us?

- Question: What is the smallest value for **hhsize** ?

- Question: What is the largest value for **hhsize** ?

- Question: What are the lower and upper limits of the interquartile range for **hhsize** and how wide is it?

Solutions: The first tabular output from SPSS is as follows:

Statistics

Household size, number of people in household (recoded) formerly A049r

| N	Valid	5144
	Missing	0
Mean		2.33
Median		2.00
Mode		2
Std. Deviation		1.196
Variance		1.430
Range		4
Minimum		1
Maximum		5
Percentiles	25	1.00
	50	2.00
	75	3.00

The answers are as follows:

- For the variable **hhsize** in the datathe arithmetic mean (or average) value is 2.33.

- The median or middle value is 2.00. This is smaller than the mean so if there is any skew to the distribution it will likely be positive.

- The mode or most frequent value which takes value 2. SPSS calculates this by looking at the frequencies of each possible value so the mode is probably more useful for categorical data.

- The standard deviation of **hhsize** takes value 1.196. Typically 95% of observations will lie within 2 standard deviations of the mean i.e. between -0.063 and 4.72.

- The smallest value observed is 1.

- The largest value observed is 5. Households larger than this will have been included in this group.

- The inter-quartile range runs from the lower quartile to the upper quartile. The lower (25%) quartile takes value 1.00 meaning that 25% of observations are below   this value. Conversely, 25% of observations are above the upper (75%) quartile which takes value 3.00. Therefore, the interquartile range is of length 3.00 - 1.00 = 2.0.

The SPSS instructions for the first histogram are as follows:

1. Select **Frequencies** from the **Descriptive Statistics** submenu available from the **Analyze** menu.
2. Remove the **Household size, number of people in household (recoded) formerly A049r[hhsize]** variable from the **Variable(s)** box.
3. Return the **income** variable into the **Variable(s)** box.
4. Click on the **Charts...** button to bring up the chart options.
5. Click on the **Histogram** Chart type and also the **Show normal curve on histogram** tick box.
6. Click on the **Continue** button to return to the main window.
7. Click on the **OK** button to produce the graph required.

- Question: What does the histogram show?

Solution: The output from SPSS is as follows:
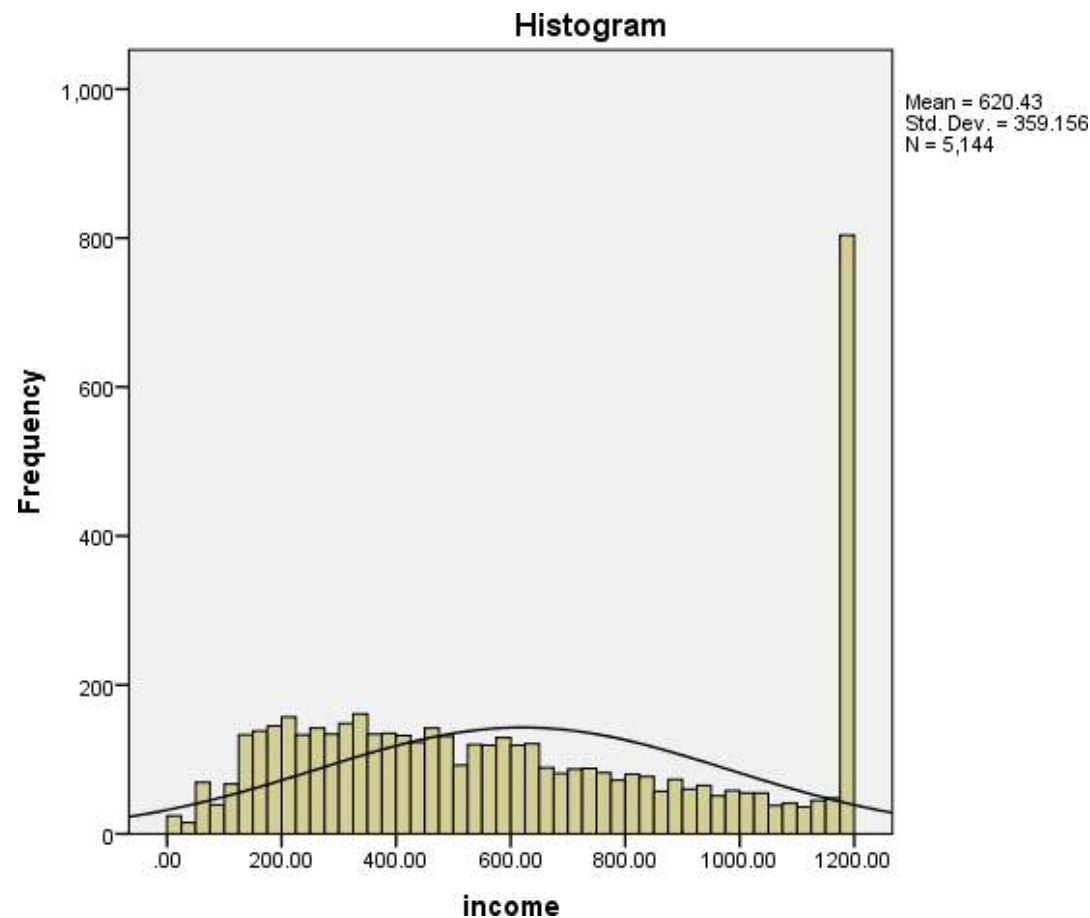
The histogram is somewhat similar to a bar graph but each bar in the histogram represents a range of values and as a result there are no gaps between bars. SPSS chooses the limits for each bar and actually plots frequencies of observations that lie between the limits. We have asked for a normal curve to be superimposed on the plot and this curve is a plot of the normal distribution that has the same mean and standard deviation as the data. If the variable approximately follows a normal distribution then the histogram should roughly follow the curve. If the data is skewed then this will not be the case.

The histogram shows that there is an unusual peak which is accounted for by high values being grouped. Additionally, the distribution is skewed.

The SPSS instructions for the second histogram are as follows:

1. Select **Frequencies** from the **Descriptive Statistics** submenu available from the **Analyze** menu.
2. Remove the **income** variable from the **Variable(s)** box.
3. Copy the **Household size, number of people in household (recoded)formerly A049r[hhsize]** variable into the **Variable(s)** box.
4. The **Charts** options will be remembered so do not need adding again.
5. Click on the **OK** button to produce the graph required.

- Question: What should we look for in this histogram?

Solution: The output from SPSS is as follows:

Histogram

Mean = 2.33
Std. Dev. = 1.196
N = 5,144

Household size, number of people in household (recoded)formerly A049r
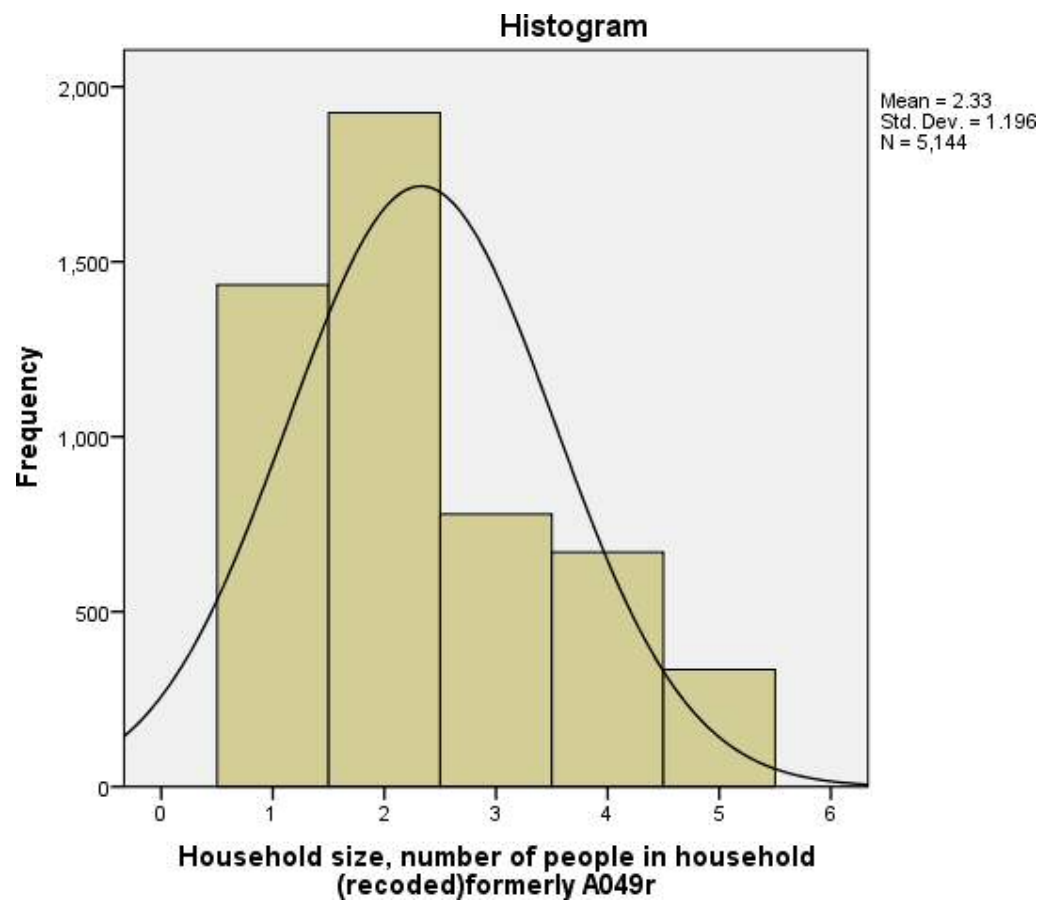
We can look at the shape of the histogram and check for unusual observations and compare the graph with the plot of the normal distribution.

The SPSS instructions for the first boxplot are as follows:

1. Select **Boxplot** from the **Legacy Dialogs** submenu available from the **Graphs** menu.
2. We want to choose **Simple** and **Summaries of separate variables** from the options here.
3. Next click on **Define** to set up the box plot.
4. Copy the **income** variable into the **Boxes Represent:** box.
5. Ignore the rest of the options and click on the **OK** button to produce the graph required.

- Question: From the boxplot what is the median (approximately) for **income?**
- Question: What is the lower quartile (approximately) for **income?**
- Question: What is the upper quartile (approximately) for **income?**

Solution: The output from SPSS is as follows:
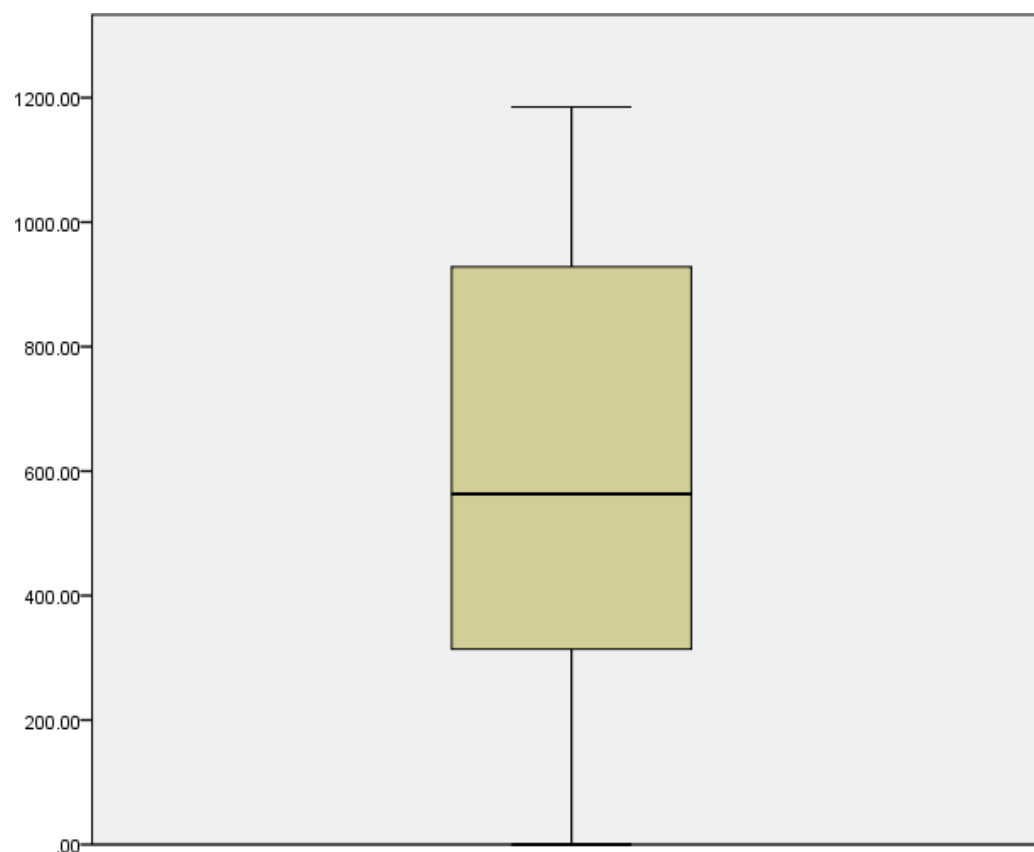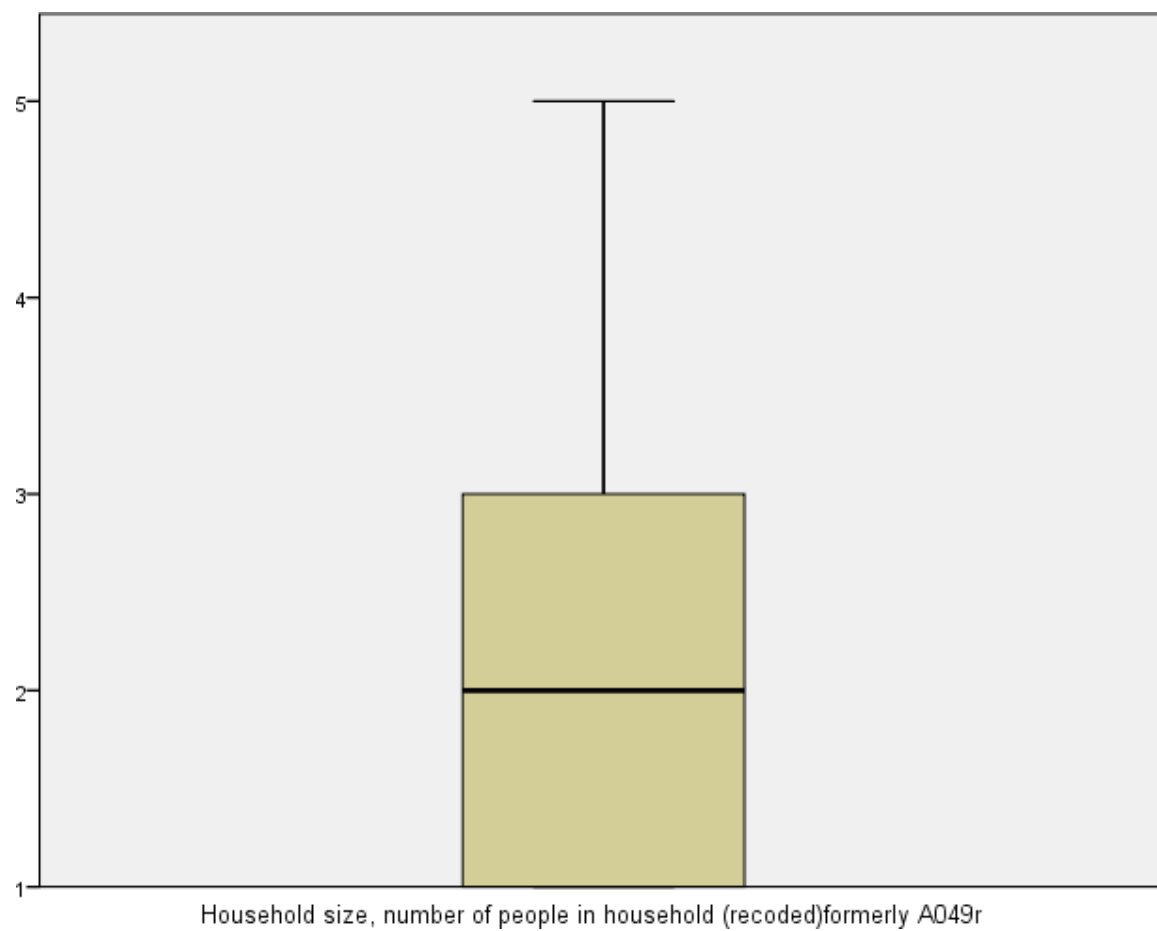

The answers are as follows:

- The median which takes value 563.1500 is represented by a vertical line in the middle of the box.
- The lower quartile which takes value 313.9456 is represented by the bottom of the box.
- The upper quartile which takes value 928.1385 is represented by the top of the box.

The SPSS instructions for the second boxplot are as follows:

1. Select **Boxplot** from the **Legacy Dialogs** submenu available from the **Graphs** menu.
2. Keep the choices of **Simple** and **Summaries of separate variables** and click on **Define** to set up the box plot.
3. Remove the **income** variable into the **Boxes Represent:** box.
4. Copy the **Household size, number of people in household (recoded) formerly A049r[hhsize]** variable into the **Boxes Represent:** box.
5. Ignore the rest of the options and click on the **OK** button to produce the graph required.

- Question: From the boxplot what is the median (approximately) for **hhsize?**

- Question: What is the lower quartile (approximately) for **hhsize?**

- Question: What is the upper quartile (approximately) for **hhsize?**

The output from SPSS is as follows:

Household size, number of people in household (recoded)formerly A049r

The answers are as follows:

- The median which takes value 2.00 is represented by a vertical line in the middle of the box.

- The lower quartile which takes value 1.00 is represented by the bottom of the box.

- The upper quartile which takes value 3.00 is represented by the top of the box.

Copyright and citation

Jo Wathan, Vanessa Higgins, Mark Elliot, William Browne, Chris Charlton, Ana Morales Gomez and Jennifer Buckley (2019) Quantitative methods e-books: Teaching Resources, UK Data Service, NCRM, Centre for Multi-Level Modelling.