Examining Continuous Variables in SPSS (Quiz)

The development of this E-Book has been supported by the British Academy.

Descriptives statistics for all variables practical questions

The 2015 version of PISA focused on science, and produced separate scales to measure understanding of different content areas. Here you will explore the distributions of two more test score variables, measuring scientific knowledge of *physical* systems (SCI_PHYS) and of *living* systems (SCI_LIVING). Benchmark scores are the same as those described in the practical section for overall proficiency scores.

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

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

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

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

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

- Question: What does the histogram show?

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

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

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

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

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

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

The development of this E-Book has been supported by the British Academy.

Solutions to Descriptives statistics for all variables practical questions

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

- Select **Frequencies** from the **Descriptive Statistics** submenu available from the **Analyze** menu.
- Copy the **Physical systems sub-score[SCI_PHYS]** variable into the **Variable(s)** box.
- Click on the **Statistics** button to go to the statistics screen.
- Here we need to select ALL the summary statistics that we are interested in looking at.
- Select **Mean, Median** and **Mode** from under **Central Tendency**.
- Select **Std. deviation, Variance, Range, Minimum** and **Maximum** from under **Dispersion**.
- Finally Select **Quartiles** from under **Percentile Values.**
- Click on the **Continue** button to return to the main window.
- Click on the **OK** button to produce the tables required.

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

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

Statistics

Physical systems sub-score

N	Valid	5194
	Missing	0
Mean		520.3541
Median		524.2115
Mode		519.52
Std. Deviation		106.85597
Variance		11418.198
Range		711.08
Minimum		165.58
Maximum		876.66
Percentiles	25	444.3868
	50	524.2115
	75	596.6088

The answers are as follows:

- For the variable **SCI_PHYS** the arithmetic mean (or average) value is 520.3541.
- The median or middle value is 524.2115. This is larger than the mean so if there is any skew to the distribution it will likely be negative.
- The mode or most frequent value which takes value 519.52. 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 **SCI_PHYS** takes value 106.85597. Typically 95% of observations will lie within 2 standard deviations of the mean i.e. between 306.642 and 734.066.
- The smallest value observed is 165.58.
- The largest value observed is 876.66.
- The inter-quartile range runs from the lower quartile to the upper quartile. The lower (25%) quartile takes value 444.3868 meaning that 25% of observations are below this value. Conversely 25% of observations are above the upper (75%) quartile which takes value 596.6088. Therefore the interquartile range is of length 596.6088 - 444.3868 = 152.222.

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

- Select **Frequencies** from the **Descriptive Statistics** submenu available from the **Analyze** menu.
- Remove the **Physical systems sub-score[SCI_PHYS]** variable from the **Variable(s)** box.
- Copy the **Living systems sub-score[SCI_LIVING]** variable into the **Variable(s)** box.
- The **Statistics** options will be remembered so do not need adding again.
- Click on the **OK** button to produce the tables required.

- Question: What is the mean for **SCI_LIVING**?
- Question: What is the median for **SCI_LIVING** and is it bigger than the mean?
- Question: What is the mode for **SCI_LIVING** and how is it calculated?
- Question: What is the standard deviation for **SCI_LIVING** and what does it tell us?

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

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

Statistics

Living systems sub-score

N	Valid	5194
	Missing	0
Mean		523.0700
Median		527.7535
Mode		386.18[a]
Std. Deviation		106.28832
Variance		11297.208
Range		685.93
Minimum		121.92
Maximum		807.85
Percentiles	25	448.3650
	50	527.7535
	75	599.3275

a. Multiple modes exist. The smallest value is shown

The answers are as follows:

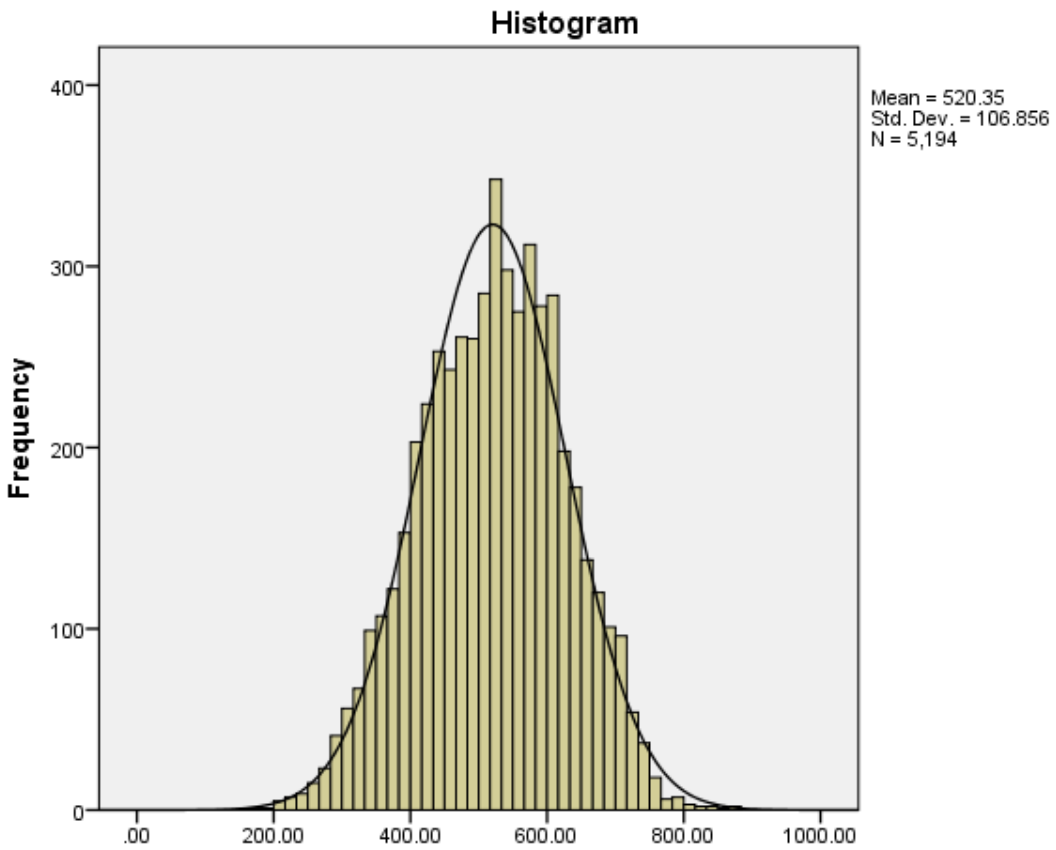
- For the variable **SCI_LIVING** the arithmetic mean (or average) value is 523.0700.
- The median or middle value is 527.7535. This is larger than the mean so if there is any skew to the distribution it will likely be negative.
- The mode or most frequent value which takes value 386.18. SPSS calculates this by looking at the frequencies of each possible value so the mode is probably more useful for categorical data.
- The standard deviation of **SCI_LIVING** takes value 106.28832. Typically 95% of observations will lie within 2 standard deviations of the mean i.e. between 310.493 and 735.647.
- The smallest value observed is 121.92.
- The largest value observed is 807.85.
- The inter-quartile range runs from the lower quartile to the upper quartile. The lower (25%) quartile takes value 448.3650 meaning that 25% of observations are below this value. Conversely 25% of observations are above the upper (75%) quartile which takes value 599.3275. Therefore the interquartile range is of length 599.3275 - 448.3650 = 150.9625.

The SPSS instructions for the first histogram are as follows:

- Select **Frequencies** from the **Descriptive Statistics** submenu available from the **Analyze** menu.
- Remove the **Living systems sub-score[SCI_LIVING]** variable from the **Variable(s)** box.
- Return the **Physical systems sub-score[SCI_PHYS]** variable into the **Variable(s)** box.
- Click on the **Charts...** button to bring up the chart options.
- Click on the **Histogram** Chart type and also the **Show normal curve on histogram** tick box.
- Click on the **Continue** button to return to the main window.
- Click on the **OK** button to produce the graph required.

- Question: What does the histogram show?

Solution: The output from SPSS is as follows:

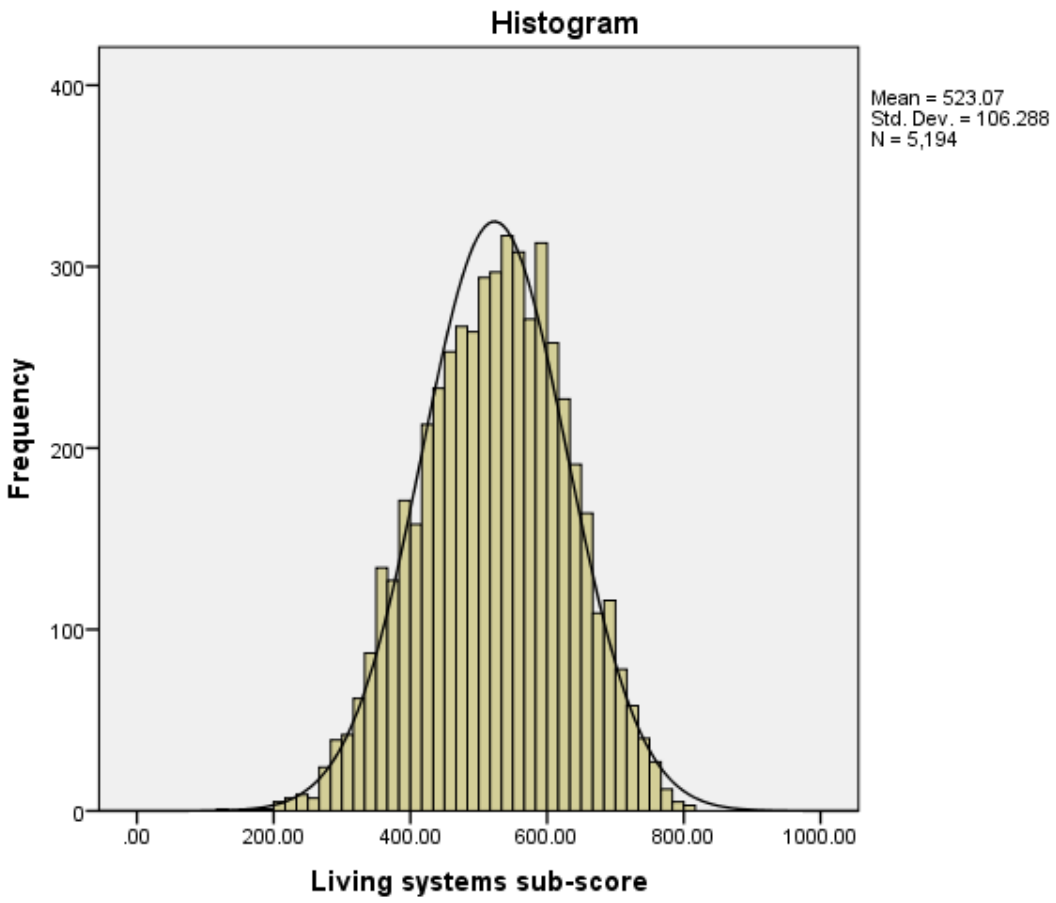


The histogram is somewhat similar to a bar graph but each bar in the histogram represents a range of values and as a result there are no gaps between bars. SPSS chooses the limits for each bar and actually plots frequencies of observations that lie between the limits. We have asked for a normal curve 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 SPSS instructions for the second histogram are as follows:

- Select **Frequencies** from the **Descriptive Statistics** submenu available from the **Analyze** menu.
- Remove the **Physical systems sub-score[SCI_PHYS]** variable from the **Variable(s)** box.
- Copy the **Living systems sub-score[SCI_LIVING]** variable into the **Variable(s)** box.
- The **Charts** options will be remembered so do not need adding again.
- 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:



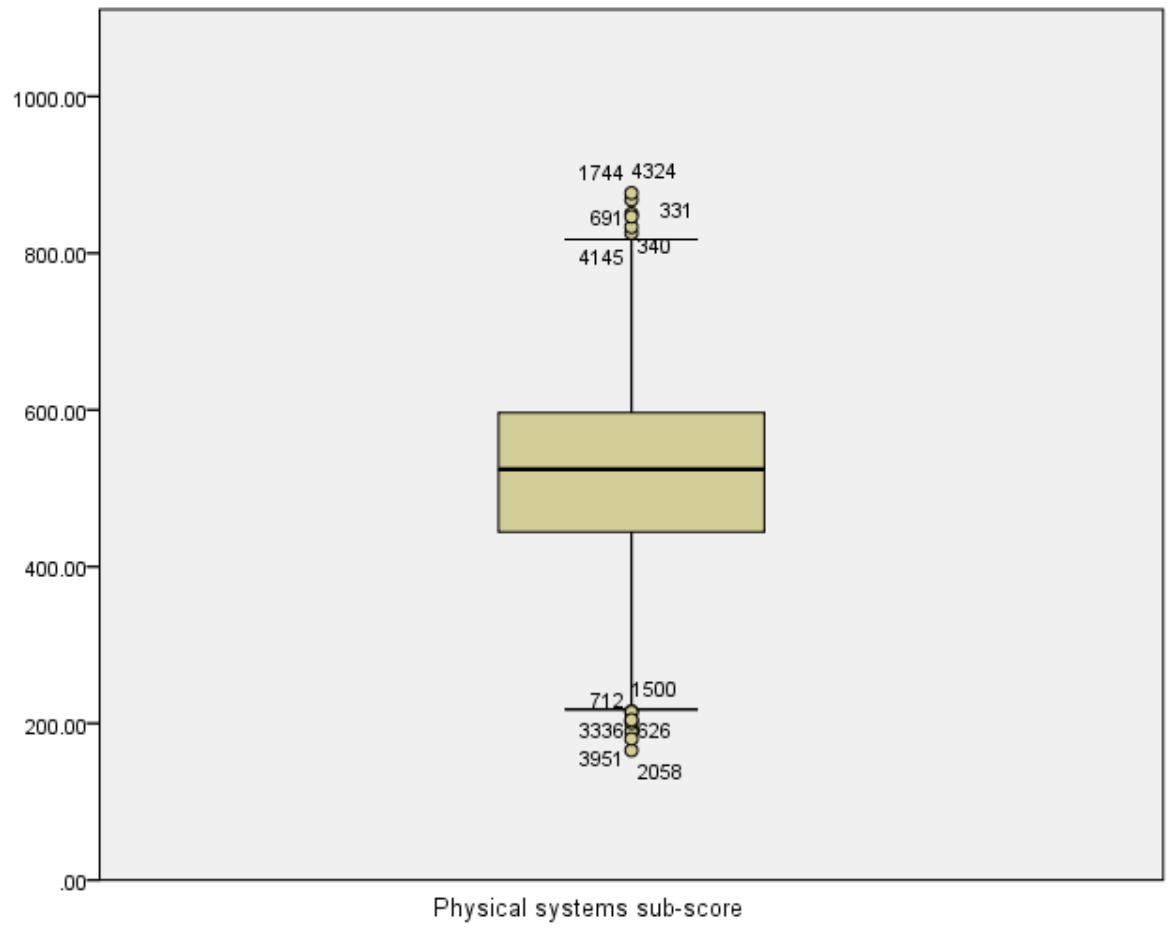
We can look at the shape of the histogram and check for unusual observations and compare the graph with the plot of the normal distribution.
The SPSS instructions for the first boxplot are as follows:

- Select **Boxplot** from the **Legacy Dialogs** submenu available from the **Graphs** menu.
- We want to choose **Simple** and **Summaries of separate variables** from the options here.
- Next click on **Define** to set up the box plot.
- Copy the **Physical systems sub-score[SCI_PHYS]** variable into the **Boxes Represent:** box.

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

Solution: The output from SPSS is as follows:



The answers are as follows:

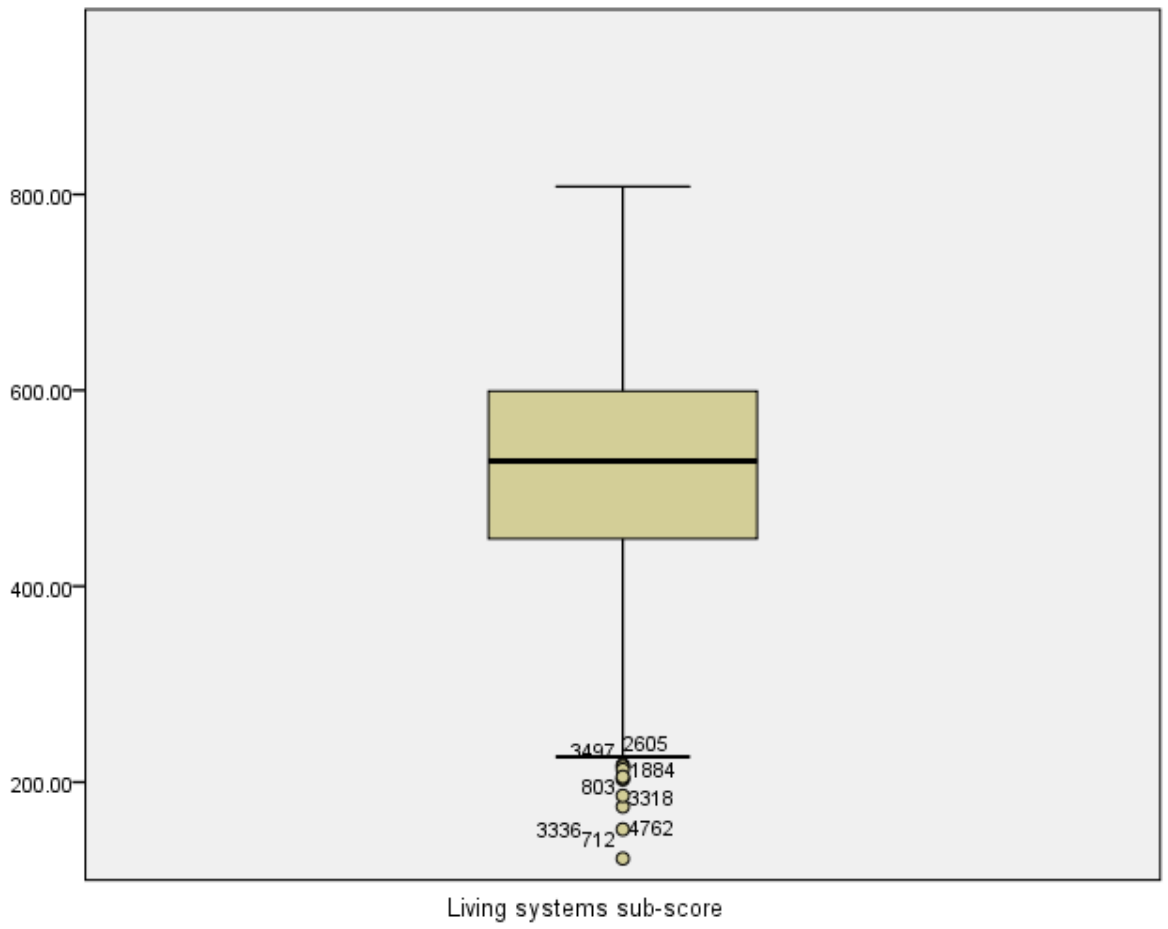
- The median which takes value 524.2115 is represented by a vertical line in the middle of the box.
- The lower quartile which takes value 444.3868 is represented by the bottom of the box.
- The upper quartile which takes value 596.6088 is represented by the top of the box.

The SPSS instructions for the second boxplot are as follows:

- Select **Boxplot** from the **Legacy Dialogs** submenu available from the **Graphs** menu.
- Keep the choices of **Simple** and **Summaries of separate variables** and click on **Define** to set up the box plot.
- Remove the **Physical systems sub-score[SCI_PHYS]** variable into the **Boxes Represent:** box.
- Copy the **Living systems sub-score[SCI_LIVING]** variable into the **Boxes Represent:** box.
- 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 **SCI_LIVING**?
- Question: What is the lower quartile (approximately) for **SCI_LIVING**?
- Question: What is the upper quartile (approximately) for **SCI_LIVING**?

The output from SPSS is as follows:



The answers are as follows:

- The median which takes value 527.7535 is represented by a vertical line in the middle of the box.
- The lower quartile which takes value 448.3650 is represented by the bottom of the box.
- The upper quartile which takes value 599.3275 is represented by the top of the box.

Mean levels of performance were very similar but slightly higher on the physical systems sub-score than on the living systems sub-score. The distributions were also similar and show that considerable variation exists in both dimensions of science understanding among students in England.