**Question**

 I wonder if boys aged 14 – 16 tend to have a higher popliteal than girls aged 14 – 16 from the Census in Schools NZ database?

**Plan**

 From the data sample of 92 students aged 14-16 obtained from the Census in Schools NZ database I will select the categorical data of gender and the numerical data of popliteal measurement. These I will process, using INZight, to produce summary statistics, dot plots and box and whisker plots to help with the investigation.

**Analysis**





**Things to Consider**

1. Anything unusual – outliers, clusters, spikes, etc.
2. Shape or distribution – normal, regular, right skewed, etc.
3. Whole spread – range, (max – min)
4. Box spread – interquartile range, IQR for short, (3rd Qu. – 1st Qu.) or lower – upper quartile.
5. Position of boxes
6. Which median is higher
7. Other values in the summary – which is higher ( how different or similar)

**Observations** – I noticed that……..

1. Girls have both low and high outliers. Boys have a low outlier. There are no other unusual features
2. Both displays have a normal distribution
3. Girl’s popliteal measurements are more spread out than boys
4. The boy’s box is a bit wider than the girl’s box
5. The boy’s box is shifted a bit to the right of the girl’s box
6. The median for the boy’s popliteal is bigger than the median for the girl’s popliteal
7. The minimum, 1st quartile and 3rd quartile are all bigger for boys. Only the maximum is bigger for girls.

**Evidence** – Prove it

1. There is a big gap between the lowest girl, the highest girl and the lowest boy and the rest of the data. This is shown on the dot plot.
2. Draw the shape over the dot plot.
3. Boy’s range: 58-21=37 Girl’s range: 67-12=55 Girl’s range is bigger by 55-37=18
4. Boy’s box width or IQR: 47-42=5 Girl’s IQR: 44-40=4 Boy’s IQR is bigger by 5-4=1
5. The boxes part overlap but the median for boys at 45 is higher than the 3rd Qu. for girls at 44. Draw a vertical line on the box plot display to show this.
6. Boy’s median 45, girl’s median 41 so boy’s median is higher by 45-41=4
7. Give the numbers from the summary statistics to back this up

**Meaning** – so what?

1. I think 12cm is probably too small and 67 cm too high, for a girl, for a popliteal measurement, so it might be better to clean these values out and redraw the displays.
2. The data is a good fit to the normal distribution for boys and girls, which is what I would expect for this type of data because most children are quite similar in size at similar ages as they grow up. As the popliteal measurement increases, you would expect to find less and less children and the same going shorter. This pattern is described by a normal distribution.
3. The much bigger range for girls was not expected, but is due to both high and low outliers being included.
4. The middle 50% of boy’s popliteal measurements are a bit more spread out than the middle 50% of girl’s popliteal measurements but not by much as the difference is only 1cm.
5. The highest 50% of boys have a higher popliteal measurement than the lowest 75% of girls.

**Conclusion**

To answer my question, because the boy’s median is higher than the girl’s median, boys tend to have a higher popliteal measurement than girls in this sample.

If another sample was taken, the results might be different.

However, the difference in the position of the boxes is big enough that I would predict that it is quite likely that boys will tend to have a bigger popliteal measurement in the population of 14 – 16 year olds in the Census at Schools NZ database.

**Further**

The sample size was quite small compared to the number of 14 – 16 tear olds in NZ. It would confirm the result better if it was taken from a bigger sample, as a bigger sample should give a more reliable result, as long as it is taken fairly across the whole population.

It would be interesting to see if this result was found for other ages, because boys and girls might have growth spurts at different ages.

It would also be interesting to see if same result for boys and girls was found in different ethnic groups and to compare ethnic groups so see if stereotypes such as Pacific Islanders tend to be big people or Asian people tend to be small hold true.

The popliteal measurement is from bottom of your foot to the bottom of your thigh, by your knee, when sitting; it about ¼ of a person’s height. The lowest girl at 12cm would have a height of about 50cm clearly this is not possible, so a mistake has been made here and it should be cleaned out. The next two girls at about 25cm are also unrealistic, unless dwarves, so is the smallest boy. At the other end, the biggest girl at 67cm would be over 2.5 metres tall, again, not possible so it should be ignored. The biggest boy at 58cm would be over 2.3m tall again extremely unlikely, unless already as tall as an adult professional basketball player, but not impossible. Taking out these unlikely and impossible results would leave both boys and girls with a similar range of around 15cm. This would mean a more realistic spread of heights of about 60cm.

**Note:** in this analysis, observations, evidence and meaning are separated out to help you recognise what each one is. When you write up your analysis it would read better if you put them together so that each observation had evidence and meaning with it.

Eg. Girls have both low and high outliers. Boys have a low outlier. This is shown on the dot plot by the big gap between the lowest girl, the highest girl and the lowest boy and the rest of the data. I think 12cm is probably too small and 67 cm too high, for a girl, for a popliteal measurement, so it might be better to ignore these values. If I had more time I could clean these values out and redraw the displays.

**Suggestion:** once you have dealt with an observation and followed through with evidence and meaning, leave a gap between it and the next one. When more than one type of observation is dealt with at the same time it is more likely that you will muddle them up and end up with a lower grade.

**Remember:** it is the overlap of the two boxes that you use to make the call for the population, not the width of the boxes or IQR.