Introduction:

It is well known that the waters near Antarctica are cold. As a result of ocean movements and currents, the temperature of the water can vary depending on a lot of factors. Possibly one of the main factors that cause this difference in temperature to be apparent is water current in the ocean. Similar to wind direction, water currents often carry water that has a certain temperature to it depending on where the currents come from.

Since water currents can be very complex, it can be hard to predict the temperature of the water based on water movements alone. This means that other means of prediction can be very useful. I think that people who study the effects on global warming would benefit from the outcome of this investigation as the study on ocean temperature plays a big part in climate change.

Question:

I wonder if the median water temperature at the Kerguelen Islands is greater than the median water temperature at Macquarie Island?

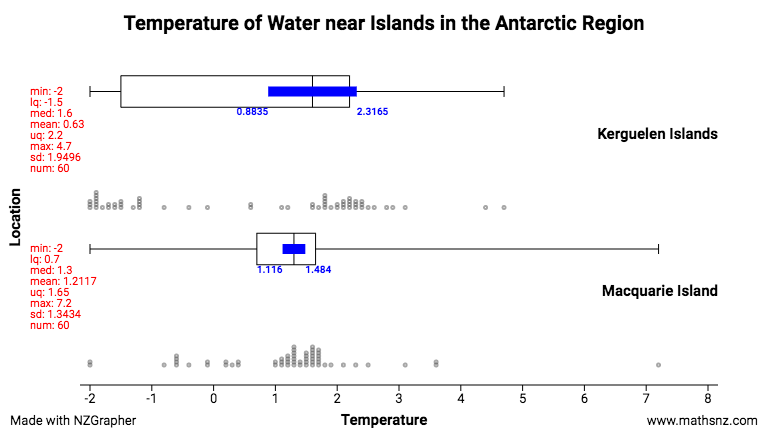
Plan:

I will be taking a sample out of the population: water around Macquarie Island and the Kerguelen Islands in the Antarctic region. My numerical variable that will be looking at is the temperature of the water in degress centegrade (˚C).

I will take a sample of 60 data points for each of the two islands. This equates to 20% of the population being sampled. I have chosen this to minimise the effects of sampling variability so that I can be more confident that my sample is representative of the population.

I will be using a systematic sampling method to generatre my sample. I have chosen this because it is faster for large populations such as this and that it coveres the entire population. There are also no obvious patterns in the dataset which means that there should be no bias in my sample. I will be using a true random number generator (<https://www.random.org/>) to generate my starting point. I will then increment this number by 5 (300/60 = 5) and highlight the corresponding data point to generate my sample. I will repeat this for the second island.

Data:

Below is my sample shown as a dot plot and a box and whisker chart.

Analysis:

From the box and whisker chart, I notice that the middle 50% of the sample from Macquarie Island falls within the middle 50% of the sample from the Kerguelen Islands. As a result, there is no noticable shift in my sample. The lower quartile for the Maquarie Island sample is 0.7˚C and the lower quartile for the Kerugen Islands sample is -1.5˚C. The upper quartile for the Maquarie island sample is 1.65˚C and the upper quartile for the Kerguelen Islands sample is 2.2˚C. This does not macth the direction in my question as there is no noticable shift. This could mean that back in the population, there is likely to be no nociable shift in temperature between the two islands.

From the box and whisker chart, I also notice that the median value for temperature near the Kerguelen Islands (1.6) falls further up the scale than the median temperatue near Maquarie Island (1.3). This is a difference of 0.3 degrees. This difference does match the direction in my question but is rather insignificant. It is possible that due to sampling variability, this difference could change or even reverse. This is an indicator that back in the population, there is likely to be no signnificant difference in median water temperature between the two islands. Given that both the shift and center show no significant features, I can be more confident that this is true.

There are a number of unusual features present in this sample. One unusual feature is a group of extreme values located between -2˚C and -1˚C for the Kergulen islands. This group may be caused by a number of factors. One could be that seasonal effects could be present within the population. It is possible that throughout the year, the water temperature varies depending on the time of year. Another possible reason could be bacause of my sampling method. Earlier I stated that I chose the systematic sampling method to extract data from the population. If any pattern is present within my dataset, it is possible that this may be seen by grouping in my sample.

Conclusion:

Each time you take a sample, it is very likely that the statistics for the sample (median etc.) will be different. This is known as sampling variability. Sampling variability can be defined by how much different samples of the same population differenciate from each other. The main factor that influences sampling variability is sample size.

An effect of sampling variability can be seen by how much the sample medain deviates from the population medain. If we are unable to get the population median, we must use an Informal Confidence Interval to estemate where the population medain is likely to be located.

We can generate an informal confidence interval by using this formula:

Where n is the sample size and IQR is the interquartile range.

This formula takes into account the sample size and the interquartile range to produce an accurate estimate as to where the population median is likely to be located.

Based on the informal confidence intervals (displayed as blue bars on the chart), I am unable to make the claim that the median water temperature at the Kerguelen Islands is greater than the median water temperature at Macquarie Island from the population of information collected in 2017 from tags attached to Sothern Elephant Seals by members of the MEOP scientific community.I say this because the informal confidence intervals overlap, which means that the populaiton medain for both categories could deviate in such a way that the they are inverted.

My informal confidence intervals indicate that the median water temperature at the Kerguelen Islands may be less than the median water temperature at Macquarie Island.

I can also add to this conclusion given that there are a lot of extreme values in my sample. This may be an indicator that my chosen sampling method is not representative of the population and could therefore invalidate any claims that I make.

I can additionally support this conclusion becasue the difference in my sample median is very weak (0.3˚C difference) and is highly likely to be different in another sample.