Sydney Wilborn
Statistics Paper
November 5, 2019

Quantitative Reasoning Individual Project

The purpose of this paper is to find the estimated average in either high or low temperatures in Celsius daily over several years at Hull Springs Farm and Lancer Park. The data I was given is a simple random sample from Lancer Park over a two year period. Then, I was given the different high and low temperatures of each season (Spring, Summer, Fall, Winter) from the Longwood Environmental Observatory (LEO). From there, I then had to choose a high or low, along with a season to complete my test. However, I chose to use the high temperatures from summer to determine my estimated average of degrees Celsius.

Once I received all of my data (See Figure A), I had to figure out which test I needed to use. I agreed upon myself to use the T-Interval because sigma ($\sigma - \textit{standard deviation}$) is unknown and we are looking for the mean. From then, we need to begin our test by figuring out what our parameter is. The parameter is what is being used to determine what our true mean is. The definition of our parameter (Mu- μ) would be the true average/mean of the high temperatures in the summer at Lancer Park over a two year time period. Next, it asks us to find two different confidence intervals for our data. Given that it can only be between .90 and .99, I chose two intervals of .90 and .95 to test. I cannot exceed beyond or below .90 and .99 because the margin of error has to be below .10, which covers the empty space to fill 100% over the whole bell shaped curve.

Next, we need to figure out what conditions we need to run this T-Interval test. The conditions are not knowing σ, having a simple random sample (SRS), seeing if we have any skews (slight skew = n > 15, strong skew = n > 40), then seeing if we have any outliers. Also, note that 'n' represents the population of the sample, so that is very necessary for running this test. In this case, 'n' is 25 (twenty-five different high temperatures). To figure out if we can run this test, you need to enter all of your data into the calculator. You go to 'STAT', press the first option (1:EDIT…), then enter the data you are given under L1. Once you have entered all of your data, you need to see if you have any slight or strong skews. To find this from entering your data under L1, press the '2nd' button, 'y= (STAT PLOT)', then press the first option that says '1:PLOT1..' that has a graph underneath it. It will then show you a screen that has 'Plot 1', types of graphs, 'Xlist,' and 'Frequency'. You will need to go over to the third option that looks like a bar graph, and hit 'ENTER'. This will determine if you have any slight or strong skews. But also, make sure you have the right Xlist, which would be L1 in this case. If it is in another list, you would then enter whichever list your data is contained in. Once you have picked the bar graph, you hit the 'GRAPH' button which will show you what your graph looks like. But first, to be more accurate, hit 'ZOOM', scroll down to the second option of 'Zoom In' and press 'ENTER'. You will then be able to tell if you have more data on one side than the other (a skew), or a close to normal distribution. In this test, we do not have any big skews. Next, you see if you have any outliers, so you follow the same steps above, but then go to the fourth option that looks like a line, box, another line, then a few dots under the 'Plot1' tab. Once you graph this and you do not have any dots that are not on the line, then you do not have outliers. We also do not have any outliers in our test, so our test is good to run.

Following this, we need to compare the two confidence intervals. On a bell shaped curve, the confidence interval is the area that is in the middle (as it goes up and comes back down), excluding a small section on both sides at the bottom where the curve is close to the horizontal line. To compare the two confidence intervals, you will use the data that you entered earlier in the problem and will be asked to enter more data. To do this, go where you entered your data under L1, press 'STAT', go over to 'TESTS', scroll down to the 8th option named 'TInterval', and press enter. It will then show 'Inpt', so make sure you have 'Data' highlighted because you have entered your data under L1. Make sure you have L1 for 'List', 1 for 'Freq', and have a C-level of .90 (the one we are testing first). Once you have everything entered correctly, go down to 'CALCULATE' and press 'ENTER'. Then, it will give you your z scores of 28.841 and 31.401 ($28.841 < \mu < 31.401$). This means that on the opposites sides of the bell shaped curve being split by x bar (mean of the population), will be at the bottom to represent the area that is outside of the curve (C-level). It will also give you 'x bar', which in this case, equals to 30.1212, 's' that equals to 3.74004492, and 'n' that equals to 25. Finally, you need a conclusion of what our c-level tells us. For this c-level, it would be, "We are 90% confident that the true average/mean of high temperatures at Lancer Park fall between 28.841 and 31.401 (the mean would be 30.1212)." Next, you would follow the same steps as above but enter a confidence level of .95. The calculator will give you z scores of 28.577 and 31.665 ($28.577 < \mu < 31.665$), 'x bar' equals to 30.1212, 's' equals to 3.74004492, and 'n' that equals to 25. Once these numbers are found, the confidence intervals tells us how much margin of error is within the curve. For this conclusion, we would say, "We are 95% confident that the true average/mean of high temperatures at Lancer Park fall between 28.577 and 31.665 (the mean would be 30.1212)."

If the confidence level is .95, there will be a lower margin of error (space outside of c-level), which results in more c-level area. Therefore, if we want to be more confident in our answer, a c-level of .95 is more correct than .90. However, if there is an example that you need more error, the c-level of .90 will benefit more. The confidence level just tells us the likely values of the true population mean, so we trust the system more. The difference between a c-level of .90 and .95 is that the margin of error will differ between the two. This means there is more error on the curve of .90 versus .95. I have chosen these two particular confidence areas because we were not given any specific numbers to use. If we are not given a c-level, we can pick between .90 and .99, nothing more and nothing less.

The purpose of this paper was to find the estimated average of my statistics of high temperatures in the summer at Lancer Park from over a two year period. Once given the data, I had to choose confidence levels to test. I chose .90 and .95. It can only be between .90 and .99 because it leaves room for margin of error. Having a bigger c-level leaves room with a smaller margin of error. However, if your confidence level get smaller, your margin of error gets bigger, so there are more room for mistakes. Confidence levels play a major role in determining how correct and accurate ones information is.

Figure A: High Temperatures at Lancer Park in Celsius

31.76	33.74
31.8	30.63

Sydney Wilborn
Statistics Paper
November 5, 2019

32.58	26.85
34.99	27.67
20.97	34.33
31.57	30.04
30.03	21.16
34.03	27.97
35.98	28.44
29.73	33.92
31.32	
27.85	
28.35	
29.32	
28	