Course Project - Phase 2

Shaun Pritchard

Rasmussen College

STA3215

Fie Wang

April,15 2020

Course Project – Phase 1

**Introduce your scenario and data set.**

Provide a brief overview of the scenario you are given above and the data set that you will be analyzing.

The data set is based on a new infectious disease in which will determine if the age of patients correlates or facilitates for the rise, spread, and possible cause of the new disease. Determining if the increase in patients is related to this new infectious disease. Calculating if it is targeting specific age groups.

The data set consists of 65 patients that have the infectious disease with ages ranging from 35 years of age to 81 years of age for NCLEX Memorial Hospital. Remember this assignment will be completed over the duration of the course.

**Classify the variables in your data set.**

* Which variables are quantitative/qualitative?
  + Quantitative variable would be the age variable
  + Qualitative would be the patient numbers of the sample
* Which variables are discrete/continuous?
  + Patient numbers are discrete
  + Patient age is continuous
* Describe the level of measurement for each variable included in your data set.
  + There are 65 patients being used in this sample training from ages 35 to 81.
* Discuss the importance of the Measures of Center and the Measures of Variation.
  + the measure of center summarizes in variable a representation of all the data from its center. Variation uses this center to find normal distributions as a way to expresses the standard deviation as a percentage of the mean.
* What are the measures of center and why are they important?
  + Mean and median both try to measure the "central tendency" in a data set. The goal of each is to get an idea of a "typical" value in the data set. The mean is commonly used, but sometimes the median is typically preferred. It is resistant to the presence of extreme values
* What are the measures of variation and why are they important?
  + The measures are the range, standard deviation, and variance the range of a set of data values is the difference between the maximum data value and the minimum data value. The standard deviation of a set of sample values, denoted by s, is a measure of how much data values deviate away from the mean. basically, a transformation or velocity from the mean. The variance is a measure of how spread out the set of data is from the average (mean, median).

Calculate the measures of center and measures of variation. Interpret your results in context of the selected topic. Based off the age being they all have the infection and the numbers to identify them are not quantitative.

~~Mean 63~~

~~Median 64~~

~~Mode 69~~

~~Midrange 58~~

~~Range 46~~

~~Variance 86~~

~~Standard Deviation 9.26~~

**Conclusion**

Recap your ideas by summarizing the information presented.

It seems that the age average age range of patients at a median of 64 years old and the mean of 63 years of age have the average infection rate meaning this age range could more susceptible to the infection.it seems the standard deviation is spread apart at 9.26 which means based on the sample of patients that age range of infected patients varies more than those of the median or mean age range of patients at 63-64.

04/13/2020 Re calculated data statistics – Shaun Pritchard

|  |  |  |  |
| --- | --- | --- | --- |
| ***Age range of patients with infections disease sample (Sample Size 65)*** | | | |
|
| Stat | Value | Rounded | Population |
| Mid-Range | 58 |  |  |
| Mean | 62.53846154 | 63 |  |
| Standard Error | 1.148175167 | 1 |  |
| Median | 64 | 64 |  |
| Mode | 69 | 69 |  |
| Standard Deviation | 9.256884133 | 9 |  |
| Sample Variance | 85.68990385 | 86 | 86.37 |
| Kurtosis | -0.106939217 | 0 |  |
| Skewness | -0.600603229 | -1 |  |
| Range | 46 | 46 |  |
| Minimum | 35 | 35 |  |
| Maximum | 81 | 81 |  |
| Sum | 4065 | 4065 |  |
| Count | 65 | 65 |  |
| Confidence Level(95.0%) | 2.293743579 | 2 |  |

I recalculated all terms and corrected the midrange but came out with the same conclusion for the variance? Please advise

Course Project – Phase 2

**Discuss the importance of constructing confidence intervals for the population mean.**

Essentially constructing confidence intervals give us the best point estimate or approximation of a population mean. With this, we can determine whether inference we are looking for is within the range of viability for being true or close to approximation as they pertain to said population means. We can find a range or an interval with a percentage that determines the deviations or approximation of the deviation from this true population.

Confidence intervals help us determine the margin of error which aid in approximation being split evenly into the trials of a normal distribution. The main point in which confidence intervals facilitate is determined the overall viability of statistics for greater proportion using just a small proportion sample from it. This save in computation when dealing with large populations in the real world.

For instance, if we want to determine how many drivers in America use their blinkers. It would make no sense to survey every driver in America 300+ million people to collect data on only to compute the probability in determining the event. We can take a sample from the population of Americans and survey this smaller amount to find a range bound by limits in which we could approximate how many drivers use blinkers.

The only catch is that the approximations for the distribution only apply to the sample population we tested. By doing so we can be confident that the approximation is within the upper and lower bounds of the test we run. It is quite an ingenious way to calculate and determine viability.

**What are confidence intervals?**

Confidence intervals are the intervals or limits in between which a specific chance that specific intervals contain the true mean. a range of values for a parameter that has an associated confidence level.

**What is a point estimate?**

It is a best guess or estimate of some value or parameter. The sample standard deviation, the sample mean, and sample variance are all point estimates. In the case of a confidence interval, it is the sample mean X̄

**What is the best point estimate for the population mean? Explain.**

The sample mean X̄ is the best point estimate for the population mean µ. We can calculate p the point estimate by driving the mean from the sample of the population to best approximate the true value of µ. To find the point estimate we need to calculate the point estimate +- the margin of error using the midpoint formula. we would divide the upper and lower limit values of error after adding them by 2 to get the midpoint or point estimate value **p**

**Why do we need confidence intervals?**

Confidence intervals provide us with an upper and lower limit around our sample mean, and with this interval, we can then be confident to some degree, percentage, and approximation that we have captured the true population mean.

**95% Confidence Interval Experiment**

**Find the best point estimate of the population mean.**

The best point of estimate isX̄ = 62.53

**Construct a 95% confidence interval for the population mean. Assume that your data is normally distributed and σ, the population standard deviation, is unknown.**

In this case we will have to use the T statistic because sigma is unknown.

**Confidence interval formula:** X̄ **+-** t\* α/2 \*  s / √n

*(confidence value)* CI = 95%

*(Sample size)* n = 65

*(Sample Mean)* X̄ = 62.53

*(Compute alpha)* (α): 1 - (confidence level / 100) = 1 - 0.95 = 0.05

*(critical probability)* (p\*) = 1 – α/2 = 1 - 0.05 / 2 = 0.975

*(degrees of Freedom)* DF = (n-1) = 64

*(Sample* σ *)* s = 9.256884133

*(Standard Error)* SEx = s / sqrt( n ) = 9.256884133 / √65 = 1.148561642

*(critical t statistic value)* (t\*) = 1 – α / 2 \* P(DF) = invt(0.025,65) = -1.997729633

*(Margin of Error)* (ME) = Critical Value \* Standard Error = 2.29374

**Please show your work for the construction of this confidence interval and be sure to use the Equation Editor to format your equations.**

Confidence T level formula:

X̄±t\* α/2 \*  s / √n = 62.53 ± 2 \* 9.256884133 / √65 = **2.29**

Upper **=** X̄+t\* α/2 \*  s / √n = 62.53 + 2 \* 9.256884133 / √65 = 62.53 + 2.29 = **64.82**

Lower = X̄ - t\* α/2 \*  s / √n = 62.53 - 2 \* 9.256884133 / √65 = 62.53 - 2.29 = **60.23**

**Write a statement that correctly interprets the confidence interval in context of your selected topic.**

If we continue to repeat the experiment with the same sample size, we will get the same confidence level whereas the sample mean will fall into a range 95% of the time between the range 60.23 < mu < 64.82. With 95% confidence we can be sure the population mean will occur in this range.

**99% Confidence Interval Experiment**

**Find the best point estimate of the population mean.**

The best point of estimate isX̄ = 62.53

**Construct a 99% confidence interval for the population mean. Assume that your data is normally distributed and σ, the population standard deviation, is unknown.**

No population mean **μ**  or standard deviation **σ** is known so we must conduct a 2 tailed test using T critical values to find confidence values for the experiment.

**Confidence interval formula: X̄ +-** t\* α/2 \*  s / √n

*(confidence value)* CI = 99%

*(Sample size)* n = 65

*(Sample Mean)* X̄ = 62.53

*(Compute alpha)* (α): 1 - (confidence level / 100) = 1 - 0.99 = 0.01

*(critical probability)* (p\*) = 1 – α/2 = 1 - 0.01 / 2 = 0.005

*(degrees of Freedom)* DF = (n-1) = 64

*(Sample* σ *)* s = 9.256884133

*(Standard Error)* SEx = s / sqrt( n ) = 9.256884133 / √65 = 1.148561642

*(critical t statistic value)* (t\*) = p\* =1 – α / 2 \* P(DF) = invt (0.005,64) = -2.654854316

*(Margin of Error)* (ME) = Critical Value \* Standard Error = 3.04

**Please show your work for the construction of this confidence interval and be sure to use the Equation Editor to format your equations.**

X̄±t\* α/2 \* s / √n = 62.53 ± -2.655 \* 9.256884133 / √65 = **3.04**

Upper = X̄+t\* α/2 \* s / √n = 62.53 + -2.655 \* 9.256884133 / √65 = 62.53 + 3.04 **= 65.578**

Lower = X̄-t\* α/2 \* s / √n = 62.53 - -2.655 \* 9.256884133 / √65 = 62.53 - 3.04 **= 59.482**

It is 99% Confident the population mean is within the range: 59.48176 < mean < 65.57824

**Write a statement that correctly interprets the confidence interval in context of your selected topic.**

We can conclude that with 99% Confidence the population mean **μ**  is within the range: 59.48176 < mean < 65.57824. This experiment is conclusive to be within the range if we continue to repeat for the sample distribution.

**Compare and contrast your findings for the 95% and 99% confidence interval.**

We can be confident with 95% approximation that the population mean will be within the range 95% of the time between the range 60.23 < mu < 64.82. We can be confident that the mean would fall within the range: 59.48176 < mean < 65.57824 99% giving a higher approximation a higher margin of error at 3.04 as opposed to 2.29 with a 95% confidence level.

**Did you notice any changes in your interval estimate? Explain.**

Yes, I noticed the margin of error was largest for the higher percentage of confidence at 99% as opposed to the 95% confidence level. A meniscal fraction of the deviation became smaller segment in the tails when the percentage value was raised to 99% making the normal distribution wider.

Also, the closer we get to 100% the closer we get to 0 the CI coverage is much wider in relation to the 95% ME upper and lower bounds. It seems 95% get us closer to a smaller margin of error which a is more definitive approximation range.

**What conclusion(s) can be drawn about your interval estimates when the confidence level is increased? Explain.**

We can be confident with 95% approximation that the population mean will be within the range 95% of the time between the range 60.23 < mu < 64.82. We can be confident that the mean would fall within the range: 59.48176 < mean < 65.57824 99% giving a higher approximation a higher margin of error at 3.04 as opposed to 2.29 with a 95% confidence level. Based on the events of patients with infectious disease it would seem more probable that the average age of the patient who contracts the illness would fall within the age between 59 and 65 with only a .01% variance between the confidence level of 95% being patients from ages 60 to 65

# References

Kahn Academy . (2020). *Confidence intervals*. Retrieved from https://www.khanacademy.org/: https://www.khanacademy.org/math/ap-statistics/estimating-confidence-ap/one-sample-t-interval-mean/v/calculating-a-one-sample-t-interval-for-a-mean

Mario F. Triola. (2018). Elementary Statistics. In M. F. Triola, *Elementary Statistics.* Pearson.

References

1