STATISTICS FOR BUSINESS

TEACHER: JOSEMARI SARASOLA

SPECIAL COVID-19

CONFIDENCE INTERVALS: MEAN AND PROPORTION

HELLO! As you know, when we apply data to an estimator, we get a numerical value (for example AM, arithmetical mean) that is an estimate for the parameter (population mean). This is a point estimate.

Actually, point estimates are quite poor. They just give a value for the parameter, but we don't know the error underlying in that estimate.

To know about the error in an estimate, we set a confidence interval. For example, a 95% confidence level about the mu population mean would be something like: mu:7+-2, so (5,9); so, the estimate for the mean is 7, with a 2 unit error, with a 95% confidence; in other words, population mean is between 5 and 9 with a 95% confidence.

As we have several sampling distributions about the AM depending on the situation, we will learn only the most usual one: the t interval about the mean (it's related to the t distribution, you know about it from the previous task), that as you know applies when population is normal and sigma unknown.

Take the slides 1-7 from this lesson in our course: https://gizapedia.hirusta.io/tutorial-confidence-intervals/

There you have the general formulas for t intervals about mu. To apply these formulas for the symmetric version, you have plenty of videos in Youtube. For example this one:

https://www.youtube.com/watch?v=UmAJJtEo6cQ

The confidence is called 1-alfa. So in the extremes, we have alfa, on each extrema alfa divided by 2. So for 99% confidence, you have to look into the tables t for 0.995, all the confidence plus the lower extreme!

You should also remark that when n is bigger than 30, instead of t distribution, we must use N(0,1) standard distribution, that is z values, to get the t values for the interval.

Take these data: 4-6-7-3-4-1. Give the 90% interval about mu. Give the 99% interval about mu. Draw a conclusion: the bigger the confidence, the wider or shorter is the interval?

Now let's suppose that instead of 6 data you have 16 data with the same AM and standard deviation. Give the new interval. Is it better or more exact than the previous one? Yes! Because, we have more data, beste information, so we can shorten the interval.

Now suppose that the standard deviation you calculated with 6 data was really the same but divided by 2. Give the new interval. Is it better? Yes! Lesser deviation, lesser uncertainty, shorter interval!

Now, let's learn about asymmetrical intervals. Asymmetrical intervals are interval of type "mu bigger than" or "mu lesser than". You have the formulas in the slides.

View this video: https://www.youtube.com/watch?v=8MeC3EAmnKU

In these cases to look into the tables you must take all 1-alfa, all confidence. For example, if confidence is 99%, take into table t for 0.99, because the remaining alfa is all on the upper side.

Draw the 99% "mu>" type interval with previous data, and the 95% interval of type "mu<".

For each interval draw a graphic showing the t distribution, the central value (AM) and the interval.

Now let's learn how to set confidence intervals about a p population proportion (remark: sample assize must be bigger than 30!). You have the concept and formulas on slides 8-10: p is like mu, an unknown parameter, and we draw inference about it from p hat (sample proportion, slide no 8), jus as like AM for mu.

You can also view this video: https://www.youtube.com/watch?v=3ReWri_jh3M

Now suppose you took a survey and among 200 potential customers and 50 among them said they would buy our product. Set the 90 and 95% intervals for the real proportion of buyers in the overall population. Why is the first one shorter?

Give 90% "p>" and "p<" type proportion intervals.

Suppose now instead of 200 customers, you took the survey from 2000, with the same sample proportion being the same. Why is this shorter?

AND THE NEXT TASK WILL BE THE LAST ONE (AND VERY EASY)!