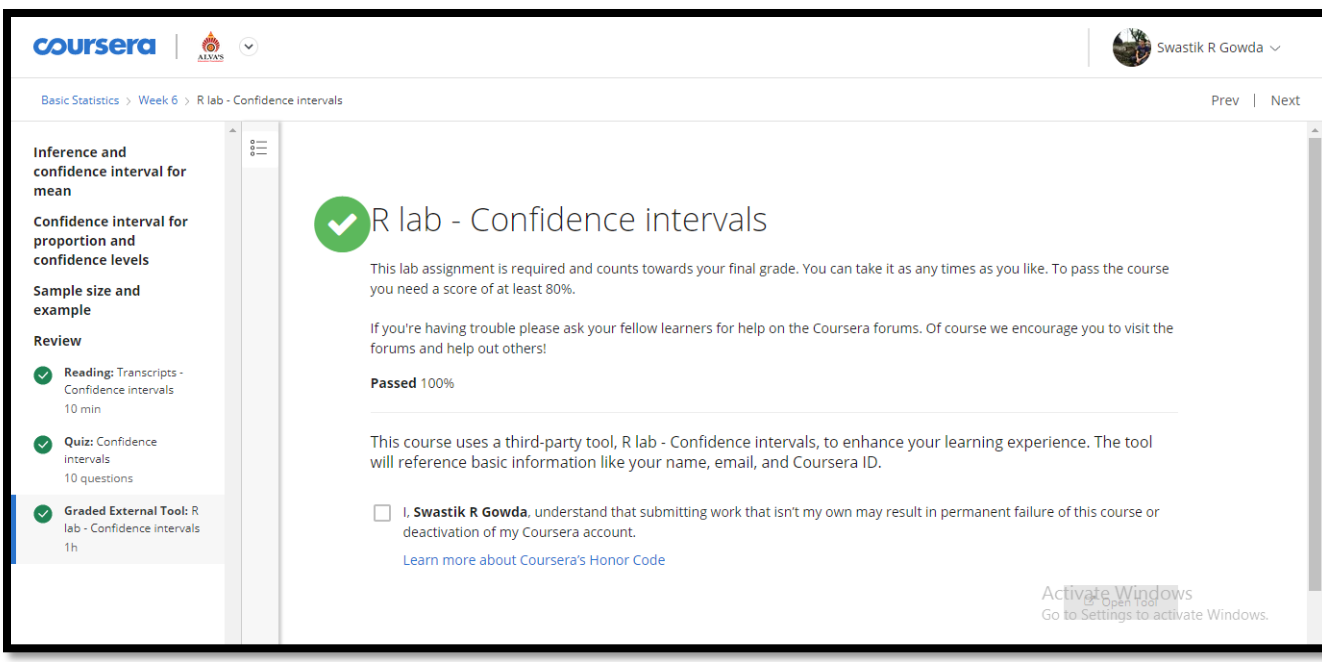
DAILY ASSESSMENT

Date:	**29-July-2020**	**Name:**	**Swastik R Gowda**
Course:	**Coursera – Basic Statistics**	**USN:**	**4AL17EC091**
Topic:	**Week - 6**	**Semester & Section:**	**6th Sem 'B' Sec**
Github Repository:	**swastik-gowda**		

FORENOON SESSION DETAILS

Image of session



Report – Report can be typed or hand written for up to two pages.

Inference and confidence interval for mean

- We can distinguish two types of statistical inference methods.
- We can estimate population parameters and we can test hypotheses about these parameters.
- In the first video of this section we argue that we can estimate the value of a population parameter in two ways: by means of a so-called point estimate (a single number that is our best guess for the population parameter) or by means of an interval estimate (a range of values within which we expect the parameter to fall).
- An interval estimate is a range of numbers, which, most likely, contains the actual population value.
- The probability that the interval contains the population value is what we call the confidence level.
- In the next two videos we'll explain how you can construct a confidence interval yourself, and we'll discuss how such a confidence interval should be interpreted.
- In the first of these two videos we'll show you how to compute a confidence interval if the population standard deviation is known.
- In this case we need the z-distribution.
- However, because this is a rather unlikely scenario, the next video looks at the more realistic situation that the standard deviation in the population is unknown.
- In this case we need the t-distribution.

Confidence interval for proportion:

- In the first video of this section we'll focus on how to construct confidence intervals for population proportions.
- The approach is very similar to constructing confidence intervals for means, although the relevant formulas and assumptions are slightly different.
- The main difference is that instead of the t-distribution we make use of the z-distribution.
- In the second video we'll discuss what it means if we make use of other confidence levels.
- We'll show that a higher confidence level leads to a wider confidence interval. This means that the more confident we are that we draw a correct inference, the larger our margin of error is.
- This implies that we have to compromise between confidence and precision; as one gets better the other gets worse.
- The video concludes with a step-by-step plan for constructing a confidence interval.

Confidence levels:

- The 95% confidence interval, tells us that we can be 95% confident that our point estimate, which could be a mean or a proportion, falls within our confidence interval.
- Or in other words, it tells us that if we would draw an infinite number of samples, similar to our extra sample and for every sample, we would compute a 95% confidence interval with a similar margin of error.
- In 95% of the cases, the population value would fall within this confidence interval.
- This, of course, also means that in 5% of the cases, this method will produce an interval that does not contain the actual population parameter.
- If you would like to reduce the chance of an incorrect inference, you could go for a larger confidence interval, such as, for instance, 99%.
- And what the consequences are, of doing so.
- Imagine you asked a sample of 100 new parents, if the babies like to answer nature's call, during the diaper changing process.
- 17% reported that, this is the case. Our sample, proportion p, thus equals 0.17.
- The formula to compute a 95% confidence interval for a proportion is p plus and minus the Z score for the 95% confidence level, times the standard error, which equals the square root of p, multiplied with 1, minus p divided by n.
- You can look up a z score for a 95% confidence level, in the z table.
- Look at this standard normal distribution here.
- When you have a 0.95 probability that your value falls within z standard errors from the mean that means, that 0.025 probabilities falls in the two tails.
- If you look up the z scores, which are displayed here in the z table, we find values of plus or minus 1.96. You can see that here.
- We can now easily compute the interval, that's 0.17 plus and minus 1.96 times the standard error which is the square root of 0.17, times 0.83, divided by 100.
- This leads to a confidence interval with the end point 0.10 and 0.24. You can now imagine that it is not so difficult to construct intervals with other confidence levels.
- Let's first look at the 99% confidence interval. This is the formula. P plus or minus the Z score for the 99% confidence level, times the standard error. The only difference is, the different Z score.
- Look at this standard normal distribution.
- When you have a 0.99 probability that your value falls within z standard errors from the mean that means, that 0.005 probabilities falls in the two tails. If we look up the z scores, which are indicated here.
- We find values of plus and minus 2.58. You can see that here.
- That is 0.17, plus and minus 2.58, times the standard error. Which was 0.038? This leads to a confidence interval with the endpoints 0.07 and 0.27.
- For the 90% confidence level, we find a z score of 1.645. This leads to a confidence interval of 0.17, plus and minus 1.645, times 0.038.
- That makes a confidence interval with the endpoints of 0.11 and 0.23. I have here displayed confidence intervals graphically.

- You can see that a higher confidence level leads to a wider confidence interval. In other words, the more confident we are that we draw a correct inference, the larger of margin of error.
- That means that we have to compromise between confidence and precision. As one gets better, the other gets worse.
- We never settle for a 100% confidence interval, because the margin of error then is far too large, which means that our conclusions are not very informative.
- In most cases, the 95% confidence interval is used.
- We can also use other confidence intervals when we construct a confidence interval to estimate a population mean.
- Suppose, we've asked a sample of 30 new parents in Amsterdam, how much hours of sleep they've lost after the first child was born.
- The mean is 2.6 hours per night. And the standard deviation is 0.9 hours per night.
- This is the formula we use, to construct a 95% confidence interval. X-bar, plus and minus the t score for the 95% confidence level, times the standard error.
- Which equals the sample standard deviation, divided by the square root of the sample size?
- Now, what is the t score for 95% confidence level? That's dependent on the degrees of freedom, which equals n minus 1.
- That is, 30 minus 1 are 29. In the t table, we should look in the column of the 95% confidence level and in the row of 29 degrees of freedom. That gives a t score of 2.045.
- The confidence interval becomes 2.6 plus and minus 2.045, times 0.9, divided by the square root of 30.
- That gives an interval from 2.26 to 2.94. If we would want to construct an interval with a confidence level of 99%, we simply replace the t score for the 95% level with the t score for the 99% level.
- You can look it up in the table, it's 2.756. The confidence interval is 2.6 plus and minus 2.756, times 0.9, divided by the square root of 30.
- That leads to an interval from 2.15 to 3.05. You can also easily do that for other confidence levels.
- First, decide which confidence level you want to use. For instance, do you settle for the regular 95% level? Or do you want to be more confident and less precise? Or more precise and less confident? Second, decide if you're dealing with a proportion or a mean.
- If you're interested in a proportion, you work with the z distribution, and if you're interested in a mean, you have to use the t distribution.
- So, in the case of a proportion, you look up the relevant Z score and in the case of a mean, you look up the relevant to score.

Sample size and example:

- The first video in this section is about choosing the sample size of a study. We’ll talk about situations in which we are interested in means and about situations in which we are interested in proportions.
- We'll show that the choice of your sample size depends on how precise you would like to be, on which confidence level you would like to use, and on how variable your data are. In an ideal world, you would just go for a super large sample.
- However, in the real world we have to deal with limited time and often we don’t have enough money to draw very large samples.
- Computing how large our sample should be can help us keep costs and time to a minimum.
- In the final video of this module we'll apply what we've learned so far to an example.
- We know that, as long as the sample is large enough, this sampling distribution is normally distributed with a mean that is equal to the population proportion, pi. And a standard deviation that is equal to the square root of pi multiplied with one minus pi divided by n.
- We also know that the probability of finding a sample proportion of less than about two standard deviations of the mean, which is the population proportion, is 0.95.
- More precisely, if we look up the Z score which corresponds to this probability, we'll find a value of 1.96. This means that we have a 95% chance that our sample proportion will fall within 1.96 standard deviations of our population proportion. This is what we call the margin of error.
- The formula with which we can compute a 95% confidence interval looks like this. P plus and minus 1.96 times the standard deviation of the sampling distribution of the sample proportion.
- 1.96 is the Z score that corresponds to the 95% confidence level. So we could also write, p plus and minus the Z score for the 95% confidence level times the standard deviation of the sampling distribution of the sample proportion.