

## Lecture 19: Confidence Intervals, T-Statistics, and Hypothesis Testing

November 13, 2000

### *Administrative Issues*

- 1) Assignment 5 will be back to you on Wednesday along with an answer sheet.
- 2) Assignment 6 is on the Web—it combines z-statistics, confidence intervals, hypothesis testing, and difference in means tests. It is due on Monday the 20<sup>th</sup>.
- 3) Your next exam is on Wednesday the 22<sup>nd</sup> of November. If you will be gone on Wednesday, you can take it by appointment on Tuesday. You must tell me by Wednesday the 13<sup>th</sup> if you'd like to take this option.
- 4) Change the last problem on Assignment 6 to be 80% and 99.9% for parts A and B, respectively.
- 5) We will cover the remaining material this week and, if time permits, have an in-class review next Monday the 20<sup>th</sup>.
- 6) Question, comments, concerns?

---

### Confidence Intervals and T-Statistics: What's the difference?

Because we do not know the true population standard deviation and variance, we must rely on an estimate of that value. Thus, we have to substitute  $s$ , which is our sample standard error for  $\bar{x}$ , for  $\sigma$ , which is the standard deviation of our population mean  $\mu$ .

When we do not know  $\sigma$ , we substitute the standard error  $s / \sqrt{n}$  of  $\bar{x}$  for its standard deviation  $\sigma / \sqrt{n}$ . The statistic that results does not have a normal distribution. It has a new distribution called a **t distribution**.

### What is a t distribution?

If we draw a simple random sample of size  $n$  from a population that has the normal distribution with mean  $\mu$  and standard deviation  $\sigma$ . The **one-sample t statistic**

$$t = \frac{\bar{x} - \mu}{s / \sqrt{n}}$$

has the **t distribution** with  $n-1$  degrees of freedom. The **t distribution** is more flexible because there is a different **t distribution** for every sample size. We specify a specific **t distribution** by giving its **degrees of freedom**. The degrees of freedom for the one-sample t statistic come from the sample standard deviation  $s$  in the denominator of  $t$ . Recall that  $s$  has  $n-1$  degrees of freedom.

### Using t instead of z for estimating our confidence intervals.

We can use the t-statistic to estimate our confidence intervals. We do this just as we estimated the z-statistics and the confidence intervals. We use the t-statistic because we do not know  $\sigma$ .

Thus, using the t distribution, we draw a simple random sample of size  $n$  from a population having unknown mean  $\mu$ . The confidence interval for  $\mu$  using the t-statistic is:

$$\bar{x} \pm t^* \frac{s}{\sqrt{n-1}}$$

where  $t^*$  is the upper  $(1-C)/2$  critical value for the t ( $n-1$ ) distribution. This interval is exact when the population distribution is normal and is approximately correct for large  $n$  in other cases.

Knowing this, how do we go about calculating our confidence intervals using the t-statistic?

### Calculating Confidence Intervals Using the t-statistic.

Let's say we want to find out a confidence interval such that we are 80% certain that the true population parameter  $\mu$  falls within that confidence interval. We begin by writing out the problem using our t-statistic rather than our z-statistic:

We are given a sample where the mean is 90 and the sample standard deviation is 5. We also know that the sample size is 25. So to calculate the 80% C.I. using the t-statistic, we just need to employ the following formula:

$$\text{80\% confidence interval} = \bar{x} \pm t^* \frac{s}{\sqrt{n-1}}$$

where  $t$  = the value for the t-statistic found in table B

Plugging the values in from above, we obtain:

$$= 90 \pm t^* \frac{5}{\sqrt{24}}$$

$$= 90 \pm t * 5 / 4.899$$

We know that our degrees of freedom,  $n - 1$ , equals 24. So we need to examine Table B in Appendix C to find the appropriate t-value for completing our equation.

Table B is set up to indicate  $\alpha$ , which represents the area in the tails of the  $t$  distribution. Alpha equals one minus the level of confidence.

$$\alpha = 1 - \text{level of confidence}$$

In our case,  $\alpha = 0.2$ . Because the t-distribution has two tails, we know from Table B that the sum of the area in the two tails equals 20% of the total area of the distribution (i.e.  $P(x) = 0.2$ ). So we simply look in the 0.20 column to find the appropriate t-value.

Now that we know where to look in Table B, we need only to find the t-value associated with 24 degrees of freedom. From Table B, we see that 24 degrees of freedom at the 0.10 significance level equals 1.318.

$$t (\text{d.f.} = 24, \alpha = 0.20) = 1.318$$

Plugging this value back into our equation from above yields:

$$\begin{aligned} &= 90 \pm 1.318 * \frac{5}{\sqrt{24}} \\ &= 90 \pm 1.318 * 0.9802 \\ &= 90 \pm 1.292 \\ &= 90 \pm 1.345 \end{aligned}$$

**So our confidence interval is simply: 88.655 to 91.345**

---

**Let's try another example:**

Let's say that we want to use the same numbers as before, but instead we want to find the 95% confidence interval for our population mean. Let's set it up the same as we did earlier:

$$\text{95\% confidence interval} = \bar{x} \pm t * \frac{s}{\sqrt{n - 1}}$$

where  $t =$  the value for the t-statistic found in table B

Plugging the values in from above, we obtain:

$$= 90 + / - t * \frac{5}{\sqrt{24}}$$

$$= 90 + / - t * 5 / 4.899$$

Up to this point, everything is the same as it was above. Now, we need to consider the other relevant information from our problem.

- 1) Degrees of freedom  $(n - 1) = 24$
- 2) Alpha  $(\alpha) = 1 -$  desired confidence level

$$= 1 - 0.95$$

$$= 0.05$$

This tells us that between the two tails of the distribution, there is a 5% probability that our population mean exists in those tails.

- 3) Using Table B, we find the 0.05 level of statistical significance with 24 degrees of freedom. This gives us:

$$t (\text{d.f.} = 24, \alpha = 0.05) = 2.064$$

$$4) = 90 + / - 2.064 * \frac{5}{\sqrt{24}}$$

$$= 90 + / - 2.064 * 1.0206$$

$$= 90 + / - 2.107$$

$$= \mathbf{87.893 \text{ to } 92.107}$$

And the final example:

Let's say that we want to find the 99% confidence interval...

$$\mathbf{99\% \text{ confidence interval}} = \bar{x} + / - t * \frac{s}{\sqrt{n}}$$

$$\sqrt{n - 1}$$

where  $t$  = the value for the t-statistic found in table B

Plugging the values in from above, we obtain:

$$= 90 + / - t * \frac{5}{\sqrt{24}}$$

$$= 90 + / - t * 5 / 4.899$$

Up to this point, everything is the same as it was above. Now, we need to consider the other relevant information from our problem.

- 4) Degrees of freedom  $(n - 1) = 24$
- 5) Alpha  $(\alpha) = 1 - \text{desired confidence level}$
- =  $1 - 0.99$
- =  $0.01$

This tells us that between the two tails of the distribution, there is a 5% probability that our population mean exists in those tails.

- 6) Using Table B, we find the 0.01 level of statistical significance with 24 degrees of freedom. This gives us:

$$t \text{ (d.f. = 24, } \alpha = 0.01) = 2.797$$

$$4) = 90 + / - 2.797 * \frac{5}{\sqrt{24}}$$

$$= 90 + / - 2.797 * 1.0206$$

$$= 90 + / - 2.855$$

$$= \mathbf{87.145 \text{ to } 92.855}$$

When do we use the t-statistic versus the z-statistic?

- 1) When you have the **population mean and standard deviation**, use the z-statistic.
- 2) When you have the **sample mean and standard deviation**, use the t-statistic.