

## HOMEWORK 15 Due:

April 2

1. What's the goal when using a confidence interval?

The goal of using a confidence interval is estimate the unknown population parameter. More precisely, to give an interval for which any value in the interval is a *plausible* value of the parameter, while values outside the interval are *implausible*.

2. Explain what we mean when we say that we are 95% confident that an interval estimate contains  $\mu$ .

It means that if many samples are collected and the confidence interval computed for each, in the long run about 95% of these intervals will contain the true population parameter. About 5% of these intervals will miss the population parameter, but, of course, we don't know which ones will miss it, and which ones will capture the unknown parameter.

3. When estimating a population parameter, are you more likely to be correct if you use a point estimate or an interval estimate? Explain.

A *Point Estimate* gives only one plausible estimate for the parameter value in question. But due to *sampling variability*, we cannot be sure how close a point estimate is to the true population parameter. So we are more likely to be correct if instead of just reporting a point estimate, we can make an interval estimate that specifies a range of plausible values for the parameter.

4. Given the same data, which level of confidence would produce the widest confidence interval?

- a. 90%
- b. 99%
- c. 95%

5. What is the probable effect on the width of the confidence interval when the sample size is increased?

- a. The width increases
- b. The width decreases
- c. No effect

6. What value is in the exact middle of the confidence interval, a statistic or a parameter?

The statistic. That's our point estimate.

7. You work for a consumer advocate agency and want to find the mean repair cost of a certain brand of washing machine. As part of your study, you randomly select 40 repairs, determine their costs, and find the mean to be \$100.00 with a standard deviation of \$17.50.

a. Construct a 90% confidence interval for the mean repair cost. (Include all three steps: check conditions, calculate the CI, and interpret it in context.)

**Conditions:** random sample (checked),  $n > 30$  (checked)

Using the t-interval (the z-interval is a good approximation, too, since  $n > 30$ ):

$$\bar{x} = 100 \quad s = 17.50 \quad t^* = 1.685 \quad (\text{or } z^* = 1.645)$$

$$\bar{x} \pm t^* \frac{s}{\sqrt{n}} = (95.34, 104.66)$$

We are 90% confident that the mean repair cost of ALL the washing machines of that certain brand is between \$95.34 and \$104.66.

b. Construct a 95% confidence interval. (You don't need to check the conditions again, but calculate the CI and interpret it in context).

Using the t-interval (the z-interval is a good approximation, too, since  $n > 30$ ):

$$\bar{x} = 100 \quad s = 17.50 \quad t^* = 2.023 \quad (\text{or } z^* = 1.96)$$

$$\bar{x} \pm t^* \frac{s}{\sqrt{n}} = (94.40, 105.60)$$

We are 90% confident that the mean repair cost of ALL the washing machines of that certain brand is between \$94.40 and \$105.60.

c. Construct a 99% confidence interval. (Again interpret it in context).

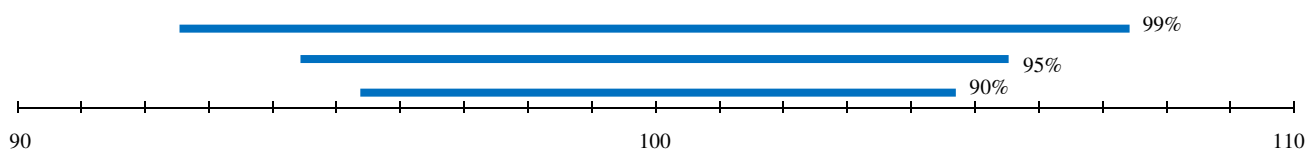
Using the t-interval (the z-interval is a good approximation, too, since  $n > 30$ ):

$$\bar{x} = 100 \quad s = 17.50 \quad t^* = 2.708 \quad (\text{or } z^* = 2.576)$$

$$\bar{x} \pm t^* \frac{s}{\sqrt{n}} = (92.51, 107.49)$$

We are 90% confident that the mean repair cost of ALL the washing machines of that certain brand is between \$92.51 and \$107.49.

d. Draw the intervals from the previous parts on a number line. Draw all three on the same number line with the point estimate in the middle.



8. Based on a sample of 100 employees at a large firm, a 95% confidence interval is calculated for the mean age of all of the firm's employees. The interval is (34.5 years, 47.2 years).

a. What was the sample mean?

The sample mean is the point estimate of the population parameter, and it's always in the middle of the CI. Thus, the sample mean in this case is  $(34.5 + 47.2)/2 = 40.85$  years

b. Find the margin of error.

The margin of error is half of the interval, that is  $(47.2 - 34.5)/2 = 6.35$  years

9. Let's say we want to estimate the proportion of ALL CSUN students who work full time. Again, using our results from question #5 from the class questionnaire, answer the following questions.

a. Clearly define the population, the sample, the parameter (both in symbol and words) and the statistic (both in symbol and words).

Population: ALL CSUN students

Sample: the students in the class

Parameter: the proportion of ALL CSUN students who work full time,  $p$

Statistic: the proportion of the student in our class who work full time,  $\hat{p}$

b. Find the 95% confidence interval for the proportion of ALL CSUN students who work full time. Interpret your results.

**12pm class:** 5 students out of 41 work full time. That is,  $\hat{p} = \frac{5}{41} = 0.122$ .

$x = 5, n = 41, z^* = 1.96$

Using  $\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$  (or your calculator)

We get: (0.022, 0.222)

That is, based on our sample, we are 95% confident that the proportion of ALL CSUN students who work full time is between 2.2% and 22.2%.

**2pm class:** 5 students out of 42 work full time. That is,  $\hat{p} = \frac{5}{42} = 0.119$ .

$x = 5, n = 42, z^* = 1.96$

Using  $\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$  (or your calculator)

We get: (0.021, 0.217)

That is, based on our sample, we are 95% confident that the proportion of ALL CSUN students who work full time is between 2.1% and 21.7%.

c. Are the technical conditions satisfied?

Conditions: it has to be a random sample of students from CSUN. This one, technically, is not satisfied. Your class is not a SRS from all students at CSUN. We have no reason, though, to think that you are not a representative sample.

The other conditions:  $n\hat{p} = 5$  which is not greater than 10. Thus, this condition is not satisfied either.  $n(1 - \hat{p})$  greater than 10, so this one is satisfied.

Since the conditions are not satisfied, the CI we calculated is not a reliable interval.

In this case, it's better to use the +2,+4 rule (given on pg.118 in your workbook):

$$\hat{p} = \frac{x+2}{n+4} = \frac{5+2}{41+4} = \frac{7}{45} \text{ for the 12pm class, and}$$

$$\hat{p} = \frac{x+2}{n+4} = \frac{5+2}{42+4} = \frac{7}{46} \text{ for the 2pm class.}$$

Using this statistic, the 95% confident intervals are:

(0.050, 0.261) for the 12pm class, and (0.048, 0.256) for the 2pm class.

This interval really has a 95% capture rate, so this is a more precise interval.

10. You wish to estimate, with 95% confidence, the proportion of computers that need repairs or have problems by the time the product is three years old. Your estimate must be accurate within 3% of the true proportion.

a. If no preliminary estimate is available, find the minimum sample size required.

If no preliminary estimate is available, we need to use the conservative approach,  $\hat{p} = 0.5$ .

The margin of error is 3% = 0.03.

The minimum sample size required to have our estimate within 3% of the true proportion is:

$$n = \left(\frac{z^*}{m}\right)^2 \hat{p}(1 - \hat{p}) = \left(\frac{1.96}{0.03}\right)^2 0.5(1 - 0.5) = 1067.111$$

But we need to round it up, so we need at least 1068 people in our sample.

b. Now suppose a prior study involving less than 100 computers found that 19% of these computers needed repairs or had problems by the time the product was three years old.

Find the minimum sample size needed.

Now we have an idea about  $\hat{p}$ , that is  $\hat{p} = 19\% = 0.19$

The margin of error we want is still  $3\% = 0.03$ .

The minimum sample size required to have our estimate within  $3\%$  of the true proportion is:

$$n = \left(\frac{z^*}{m}\right)^2 \hat{p}(1 - \hat{p}) = \left(\frac{1.96}{0.03}\right)^2 0.19(1 - 0.19) = 656.9136$$

But we need to round it up, so we need at least 657 people in our sample.