Lsn 11 - MA206Y

Clark

Admin

Let's go back to the boards and write down the number of hours (rounded to the nearest quarter hour) that you got last night.

What are:

$$n =$$
$$\bar{x} =$$
$$s =$$

Suppose we repeated this with MAJ Jonas' class next door. Will $\bar{X}$ be the same or different? What about μ?

What, in words, is μ for this study?

Similar to forming a CI for π we can form a confidence for μ. If our validity conditions are met (see page 183) we can use a theory based test (again I'm skipping simulation based here because I don't think it adds much to our discussion). The general idea is again we are forming:

$$\text{statistic} \pm \text{multiplier} \times \text{SD of Statistic}$$

For forming a CI for for π we used:

Here we need to note that the natural statistic we are using to estimate μ is $\bar{x}$ and the SD of $\bar{x}$ is $\frac{s}{\sqrt{n}}$. Here's an important point:

For a 95% CI the multiplier will be approximately 2. Use this to come up with a 95% CI for μ for our sleep study.

Similar to the `prop.test` command, we can use `t.test` in this instance to come up with any size Confidence Interval. Here we need to put in all of our X values, not $\bar{X}$ and still use `alternative="two.sided"` . For example, if we wanted a 95% CI and our data we observed was `x=c(5,6,4,3,5)` meaning $\bar{x} = 4.6$, we could do

```r
x=c(5,6,4,3,5)
t.test(x,alternative="two.sided",conf.level=0.95)
```

```
##
##  One Sample t-test
##
## data:  x
## t = 9.0213, df = 4, p-value = 0.0008362
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  3.184285 6.015715
## sample estimates:
## mean of x
##       4.6
```

A really important note here is that the `p-value` output is probably meaningless if we're using t.test in this manner. That is because the default test R is doing is:

So unless that's the hypothesis we are testing, we don't really need this p value. So, let's use this to build an 85% CI, but first, does our intuition tell us this should be wider or narrower than our 95% CI?

What about a 99% CI? Should that be wider or narrower?

Confirm or deny your intuition.

Interpret your 85% CI.

Let's look at the formula.

$$\bar{x} \pm \text{multiplier} \times \frac{s}{\sqrt{n}}$$

When we are adjusting our confidence, what in this formula is being adjusted?

What would happen to the width of this CI if n increases?

What would happen to the width of this CI if s increases?

Let's go back to our data, what does it mean if s is decrease? Does this make sense that we would be more certain of μ?

What would it mean if n is increasing? What happesn to our CI as $n \to \infty$?

Cautions when Conducting Inference

Explain the Bradley Effect.

This is an example of a systemic or nonrandom error. Note that above we said as $n \to \infty$ we KNEW the truth. If we have systemic errors, will sampling more people tell us the truth? Why or why not?

I think the most important point in section 3.5 is given on page 201. That is the difference between significant and important. What does it mean if something is statistically significant?

If we wanted to test $H_0 : \mu = 6$ hours of sleep vs $H_a : \mu \neq 6$ and we sampled 1000 Cadets and got $\bar{x} = 6.1$ and $s = .1$ our 95% CI would be:

```
multiplier=qt(.975,999)
s=.1
n=1000
6.1-multiplier*s/sqrt(n)
```

```
## [1] 6.093795
```

```
6.1+multiplier*s/sqrt(n)
```

```
## [1] 6.106205
```

As the CI doesn't cover 6 we would say that it is not a plausible value, so we would therefore reject at an $\alpha = 0.05$ that $\mu = 6$. Why?

But does this really mean anything for our researcher? If I know that Cadets get 6.1 hours sleep instead of 6 hours sleep, then yeah, technically $\mu \neq 6$, but so what? Does that change the BTOs opinion that Cadets are getting 6 hours of sleep?

A survey of 47,000 American households found that 32.4 % own a cat.

Is .324 a parameter or a statistic? What symbol do we use to represent this value?

What, in words, is the parameter of interest here?

Conduct a theory based significance test or whether the sample data provide strong evidence that less than one-third of all American households own a pet cat. Specify the null and althernative hypotheses, calcualte the z statistic and find the p value (it's fine to use `prop.test` here if you can). Is this statistically significant at $\alpha = 0.001$ level?

Determine a 99.9% CI for the proportion of all American households that have a pet cat.

Which statement is true:

- The sample data provide very strong evidence that less than one-third of all American households own a cat
- The sample data provide strong evidence that much less than one-third of all American households own a cat