Lsn 9 - MA206Y

Clark

Admin

Recall that if certain validity conditions were met we could use

$$z = \frac{\hat{p} - \pi_0}{\sqrt{\pi_0(1-\pi_0)/n}}$$

As a test statistic to test whether our true parameter, $\pi = \pi_0$ vs $\pi \neq \pi_0$. What was the benefit of using this test statistic?

As it turns out, a similar condition applies when we are looking at the population average, μ. What is the difference from π and μ?

When the quantitative variable is symmetric or we have at least 20 observations and the sample distribution is not strongly skewed we can use:

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

As our test statistic. And in this case we know the distribution of $t \sim t_{n-1}$. Below we can see the difference from a t distribution and a normal distribution. Play around with n. What happens as n gets really big?

```
library(tidyverse)
n=5
sim.data.t=data.frame(sim.tstat=rt(1000,n-1))
sim.data.z=data.frame(sim.zstat=rnorm(1000))
ggplot(sim.data.t, aes(x=sim.tstat)) +
  stat_function(fun = dt, args = list(df=n-1), colour = "blue",lwd=2)+
  stat_function(fun=dnorm,colour="red",lwd=2)
```

So, if we want to test hypothesis like $H_0 : \mu = \mu_0$ vs $H_a : \mu \neq \mu_0$ we could form a t-statistic and compare it to a t-distribution. (Note there's a simulation way to do this, but I don't think it's overly insightful in this case).

I want to demonstrate to the BTO that Cadets are getting aren't getting enough sleep. In particular I want to show that the average amount of sleep that Cadets are not getting 6 hours of sleep.

Based on this question, what is the population of interest? What is the parameter? What symbol do we use as a parameter?

Express the null and alternative hypothesis for investigation this question

Ideally, how would we sample Cadets to investigate this question?

I decide to just sample Cadets in my class, what kind of sample is this? Is this potentially a biased study?

How much sleep (to the nearest quarter hour) did you get last night? Combine your answer with those of your classmates.

Load the data into R and create a histogram of your data. Describe the distribution. What is the center of the distribution? What do we call this?

How many students are above the sample mean, how many are below?

What is the **sample standard deviation** for our study? What letter do we use for this?

Are our validity conditions met to use a theory based approach?

Let's assume that they are. Form our **t statistic**. What is the distribution of our t-statistic asusming our null hypothesis is true?

Draw a t distribution. Where does OUR value lie? If the null hypothesis was true would our value be rare?

Just like `pnorm` R has a command called `pt` that integrates for us (HOORAY). Using this, find the p-value for our study.

Let's do the same thing with a built in r function called `t.test`. Run `?t.test` in your console. What inputs do we need to put in?

If we conclude that Cadets are not getting 6 hours of sleep, what is the probability we have committed an error?

In statistics there are two types of errors we need to be concerned with, Type I and Type II. Oftentimes Cadets find it useful to think of Type I and Type II errors as Jurys. Let's look at Table 2.9 and Table 2.10

Oftentimes we want to FIX our Type I error and use this create a decision rule. Note that this is done BEFORE we take our sample. For instance, if we set $\alpha = 0.1$ we are saying:

The notion of Type II error is a bit more complicated. But perhaps a drawing will help:

If we are testing:

$$H_0 : \mu = 6$$
$$H_a : \mu < 6$$

And we set $\alpha = 0.1$ what values of our t-statistic would we reject H_0 for?

What about

$$H_0 : \mu = 6$$
$$H_a : \mu > 6$$

And we set $\alpha = 0.1$ what values of our t-statistic would we reject H_0 for?

If we are testing:

$$H_0 : \mu = 6$$
$$H_a : \mu < 6$$

And we set $\alpha = 0.2$ what values of our t-statistic would we reject H_0 for?

What about

$$H_0 : \mu = 6$$
$$H_a : \mu \neq 6$$

And we set $\alpha = 0.2$ what values of our t-statistic would we reject H_0 for?

Next lesson will will discuss parameter estimation. Recall that there is a difference between $\bar{X}$ and μ. Why can't we say $\bar{X} = \mu$?

While we don't *know* μ from our experiment, perhaps $\bar{X}$ can tell us plausible values of μ. From our study today is it likely that $\mu = 8$? What about $\mu = 5$?