MATH 3339
Statistics for the Sciences
Live Lecture Help

James West
jdwest@uh.edu

University of Houston

Session 10

Office Hours: see schedule in the "Office Hours" channel on Teams
Course webpage: www.casa.uh.edu

Email policy

When you email me you **MUST** include the following

- MATH 3339 Section 20024 and a description of your issue in the **Subject Line**
- Your name and ID# in the **Body**
- Complete sentences, punctuation, and paragraph breaks
- Email messages to the class will be sent to your Exchange account (user@cougarnet.uh.edu)

Using R and RStudio

1. Download R from https://cran.r-project.org/
2. Download RStudio from https://www.rstudio.com/

Outline

1. Updates and Announcements

2. Recap

3. Student submitted questions

Updates and Announcements

- Test 2 scheduling opens Thursday, 11/04, at 12 AM.

The confidence interval

The $1 - \alpha$ confidence interval for μ, given that we know the population standard deviation is:

$$\bar{x} \pm z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$$

Margin of Error

The margin of error is

$$m = \text{critical value} \times \text{standard error}$$

For means (given the population standard deviation is known), the margin of error is:

$$m = z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$$

What is the margin of error for the mean monthly cell phone bill?

T distribution

- Used for the inference of the population mean. When population standard deviation σ is unknown.
- The distribution of the population is basically bell-shape.
- Formula for t:

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

- Use t-table, or qt(probability,df) in R.
- Degrees of freedom: $df = n - 1$.

Critical value when σ unknown

- When σ is **unknown** we use t-distribution.
- With degrees of freedom, $df = n - 1$.
- The critical value is $t_{\alpha/2}$ where the area between $-t_{\alpha/2}$ and $+t_{\alpha/2}$ under the T-curve is the confidence level $C = 1 - \alpha$.
- $t_{\alpha/2}$ is found in T-table using the row according to the degrees of freedom and the column according to the confidence level at the bottom of the table.
- In R use qt($(1 + C)/2$, df).

Confidence Interval for μ Recap

- Z-confidence interval, given the population standard deviation, σ is **known**

$$\bar{x} \pm z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$$

- T-confidence interval, given that the population standard deviation, σ is **unknown**

$$\bar{x} \pm t_{\alpha/2,n-1}\left(\frac{s}{\sqrt{n}}\right)$$

Choosing Sample Size

You can have both a high confidence while at the same time a small margin of error by taking enough observations.

- Sample size for confidence intervals of means.

$$n > \left(\frac{z_{\alpha/2}\sigma}{m}\right)^2$$

Confidence Interval for Proportions, p

We have three methods for finding the confidence interval for propotions:

M1: The $1-\alpha$ confidence interval for proportions is:

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

M2: Use $p = \frac{1}{2}$

$0 \le p(1-p) \le \left(\frac{1}{2}\right)^2 = \frac{1}{4}$

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{1}{4n}} = \hat{p} \pm \frac{z_{\alpha/2}}{2\sqrt{n}}$$

M3: Solve the inequality below for p:

$$-z_{\alpha/2} \le \frac{\sqrt{n}\,(\hat{p}-p)}{\sqrt{p(1-p)}} \le z_{\alpha/2}$$

prop.test in R

Choosing Sample Size

- We use the margin of error to determine the sample size and solve for n.
- For the confidence interval for the mean it is:

$$n > \left(\frac{z_{\alpha/2}\sigma}{m}\right)^2$$

- For the confidence interval for proportion it is:

$$n > p^*(1-p^*)\left(\frac{z_{\alpha/2}}{m}\right)^2$$

Where $p^* = 0.5$ if not given a prior proportion.

Confidence Interval for σ^2

A $100(1-\alpha)\%$ confidence interval for the variance σ^2 of a normal population has lower limit

$$LCL = \boxed{\frac{(n-1)s^2}{\chi^2_{\alpha/2,n-1}}}$$

qchisq($\frac{\alpha}{2}$, n-1, lower.tail=F)

and upper limit

$$UCL = \boxed{\frac{(n-1)s^2}{\chi^2_{1-\alpha/2,n-1}}}$$

qchisq($\frac{\alpha}{2}$, n-1, lower.tail=T)

A confidence interval for σ has lower and upper limits that are the square roots of the corresponding limits in the interval for σ^2, where $\alpha/2$ is the area in the upper tail of the chi-squre distribution.

Null Hypothesis of significant tests

- The statement that is assumed to be true. We assume “no effect” or “no difference” for the parameter tested.
- Abbreviate the null hypothesis by H_0.
- From mean body temperature example, $H_0 : \mu = 98.6°$F.
- For a significant test of the mean, the null hypothesis is always equal to some value of what we assume the mean to be.
- The null hypothesis is always $H_0 : \mu = \mu_0$, where μ_0 is some value that is assumed to be the true mean.

Possible values for the Alternative Hypothesis

There are three possible ways that we would want to test against the null hypothesis.

1. Test to prove that the mean is really lower than what is assumed. This is called a **left-tailed test**. $H_a : \mu < \mu_0$
2. Test to prove that the mean is greater than what is assumed. This is called a **right-tailed test**. $H_a : \mu > \mu_0$
3. Test to prove that the mean is not equal (either higher or lower) than what is assumed. This is called a **two-tailed test**. $H_a : \mu \neq \mu_0$

Decision

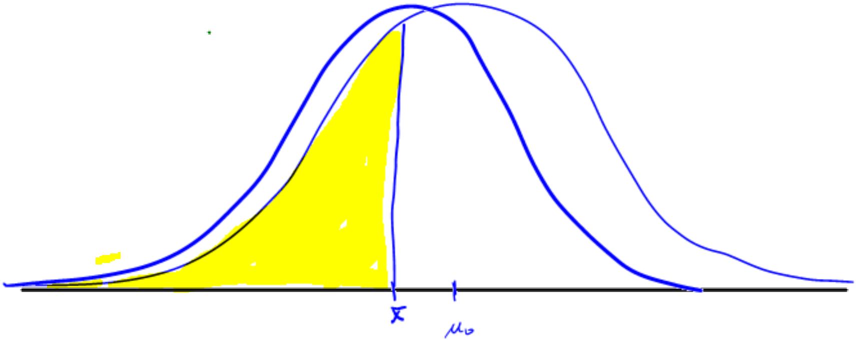
- Since there are only two hypotheses, there are only two possible decisions.
- **Reject** the null hypothesis in favor of the alternative hypothesis. (RH_0)
- **Fail to** reject the null hypothesis. ($FTRH_0$)
- We will **never** say that we accept the null hypothesis.

Type I and Type II Errors

- If we reject H_0 when it is true this is a **Type I error**.
- If we do not reject H_0 when it is false, this is a **Type II error**.
- In the example of the mean body temperature:
  - Type I error: We would conclude that the mean body temperature is less than 98.6 degrees, when in fact it truly is 98.6.
  - Type II error: We would conclude that the mean body temperature is not less than 98.6 degrees, when in fact the mean body temperature is less than 98.6.

Rejection Region

- A **rejection region** is the set of values for which the test statistic leads to a rejection of the null hypothesis.
- The critical value is the boundary of the rejection region, based on the alternative hypothesis and the level of significance, α.



$\bar{x}$
μ_0

Using R and RStudio

1. Download R from https://cran.r-project.org/
2. Download RStudio from https://www.rstudio.com/