STAT 515 -- Chapter 8: Hypothesis Tests

- CIs are possibly the most useful forms of inference because they give a range of "reasonable" values for a parameter.
- But sometimes we want to know whether one particular value for a parameter is "reasonable."
- In this case, a popular form of inference is the hypothesis test.

We use data to test a claim (about a parameter) called the null hypothesis.

Example 1: We claim the proportion of USC students who travel home for Christmas is 0.95.

Example 2: We claim the mean nightly hotel price for hotels in SC is no more than $65.

- Null hypothesis (denoted H_0) often represents "status quo", "previous belief" or "no effect".
- Alternative hypothesis (denoted H_a) is usually what we seek evidence for.

We will reject H_0 and conclude H_a if the data provide convincing evidence that H_a is true.

Evidence in the data is measured by a test statistic.

A test statistic measures how far away the corresponding sample statistic is from the parameter value(s) specified by H_0.

If the sample statistic is extremely far from the value(s) in H_0, we say the test statistic falls in the "rejection region" and we reject H_0 in favor of H_a.

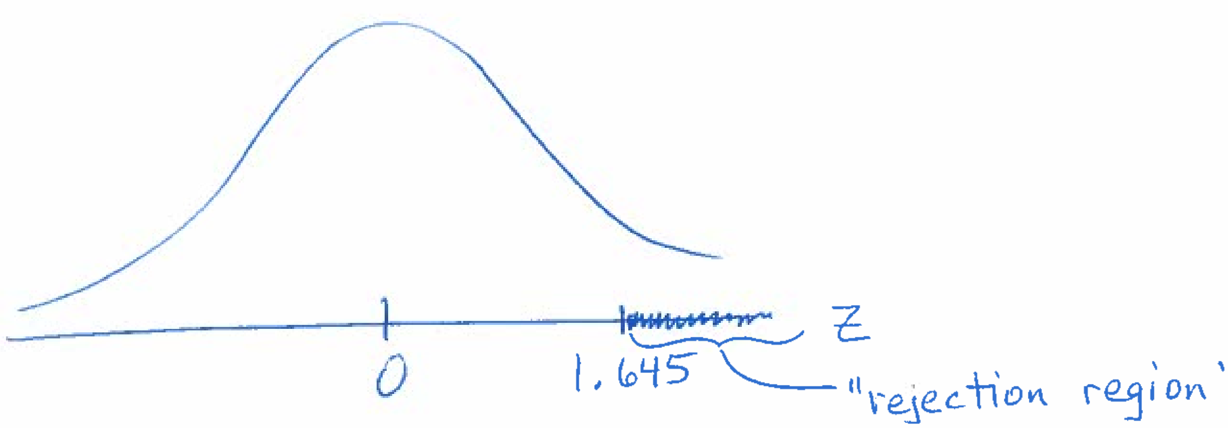
Example 2: We assumed the mean nightly hotel price in SC is no more than \$65, but we seek evidence that the mean price is actually greater than \$65. We randomly sample 64 hotels and calculate the sample mean price $\overline{X}$. Let $Z = \dfrac{\overline{X} - 65}{\sigma / \sqrt{n}}$ be our "test statistic" here.

Note: If this Z value is much bigger than zero, then we have evidence against H_0: $\mu \leq 65$ and in favor of H_a: $\mu > 65$.

Suppose we'll reject H_0 if $Z > 1.645$.

If μ really is 65, then Z has a standard normal distribution. (Why?) $\overline{X} \mathrel{\dot\sim} N\left(65, \frac{\sigma}{\sqrt{64}}\right)$ by the CLT in this case.

Picture:



If we reject H_0 whenever $Z > 1.645$, what is the probability we reject H_0 when H_0 really is true?

$P(Z > 1.645 \mid \mu = 65) =$ $\boxed{.05}$

This is the probability of making a Type I error (rejecting H_0 when it is actually true).

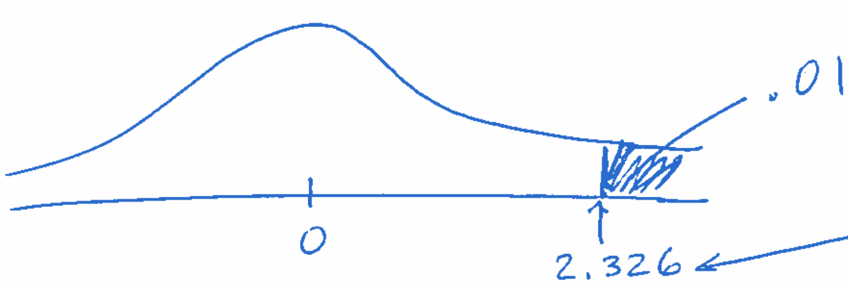
P(Type I error) = "level of significance" of the test (denoted α).

We don't want to make a Type I error very often, so we choose α to be small: Common choices of α: .01, .05, .10.

The α we choose will determine our rejection region (determines how strong the sample evidence must be to reject H_0).

In the previous example, if we choose $\alpha = .05$, then $Z > 1.645$ is our rejection region.

If we had chosen $\alpha = .01$,



found in bottom row, Table III

Rejection region would have been: Reject H_0 if $Z > 2.326$

Hypothesis Tests of the Population Mean

In practice, we don't know σ, so we don't use the *Z*-statistic for our tests about μ.

Use the t-statistic: $t = \dfrac{\overline{X} - \mu_0}{s/\sqrt{n}}$ **, where μ_0 is the value in the null hypothesis.**

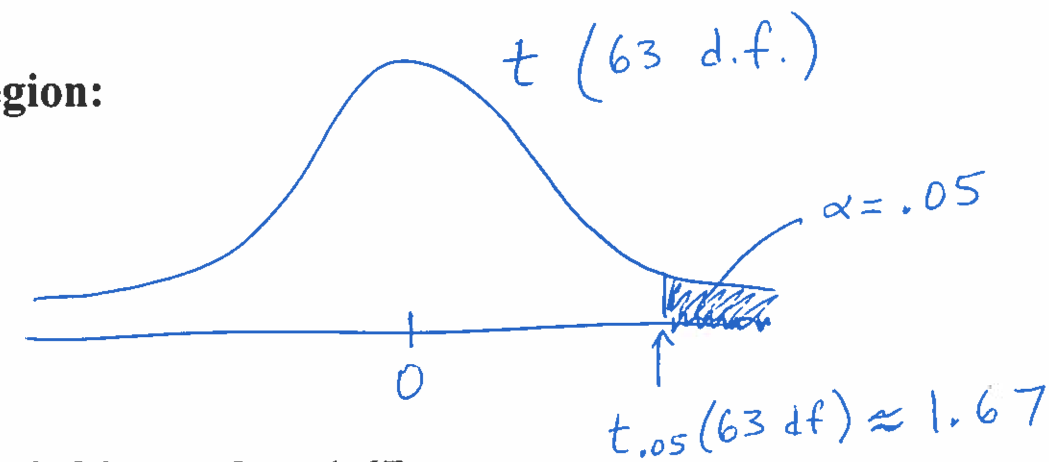
This has a t-distribution (with *n* – 1 d.f.) if H_0 is true (if μ really equals μ_0).

Example 2: Hotel prices: $H_0: \mu \overset{(\le)}{=} 65$ $\longrightarrow$ $t = \dfrac{\overline{X} - 65}{s/\sqrt{n}}$

$H_a: \mu > 65$

Sample 64 hotels, get $\overline{X} = {}^{\$}67$ and $s = {}^{\$}10$.
Let's set α = .05.

Rejection region:



Reject H_0 if t is bigger than 1.67.
$\underbrace{\text{if } t \text{ is bigger than } 1.67}_{\text{rejection region}}$

Conclusion: $t = \dfrac{67-65}{10/\sqrt{64}} = \dfrac{2}{1.25} = 1.60$

t = 1.60 < 1.67, so we do not have strong ~~[illegible]~~ enough evidence to reject H_0.

We never accept H_0; we simply "fail to reject" H_0.

This example is a one-tailed test, since the rejection region was in one tail of the t-distribution.

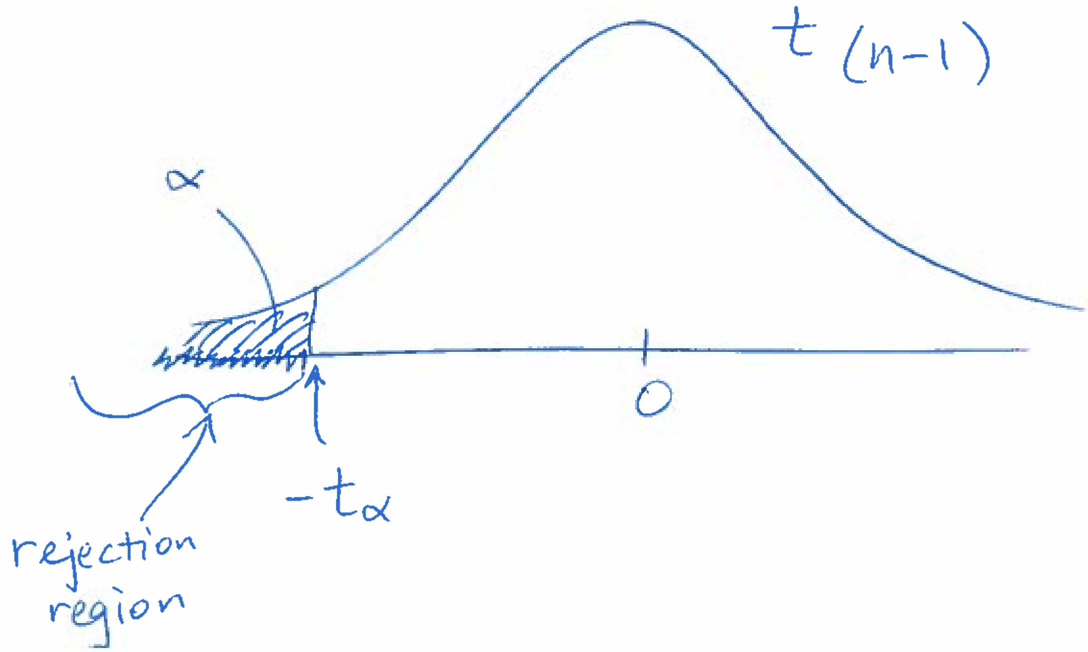
Only very large values of t provided evidence against H_0 and for H_a.

Suppose we had sought evidence that the mean price was less than $72. The hypotheses would have been:

$$H_0: \mu = 72$$
$$H_a: \mu < 72$$

Now very small values of $t = \dfrac{\overline{X} - \mu_0}{s/\sqrt{n}}$ would be evidence against H_0 and for H_a.

Rejection region would be in left tail:



Rules for one-tailed tests about population mean

$H_0: \mu = \mu_0$		$H_0: \mu = \mu_0$
$H_a: \mu < \mu_0$	or	$H_a: \mu > \mu_0$

Test statistic: $$t = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}$$

test statistic

Rejection Region: $t < -t_\alpha$ $\qquad$ $t > t_\alpha$

(where t_α is based on $n - 1$ d.f.)

t-table value (from Table III)

Rules for two-tailed tests about population mean

$H_0: \mu = \mu_0$

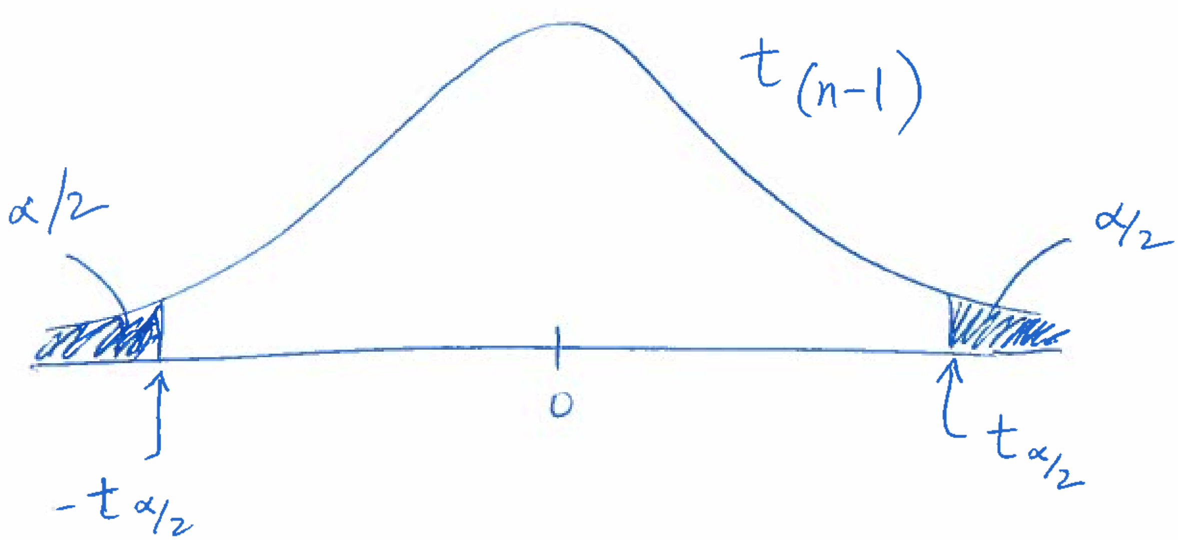
$H_a: \mu \neq \mu_0$

Test statistic: $$t = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}$$

Rejection Region: $\boxed{t < -t_{\alpha/2} \text{ or } t > t_{\alpha/2}}$ **(both tails)**

Reject H_0 if $|t| > t_{\alpha/2}$

(where $t_{\alpha/2}$ is based on $n - 1$ d.f.)



Example: We want to test (using $\alpha = .05$) whether or not the true mean height of male USC students is 70 inches.

$H_0: \mu = 70$

$H_a: \mu \neq 70$

Sample 26 male USC students. Sample data: $\bar{X} = 68.5$ inches, $s = 3.3$ inches.

$\frac{\alpha}{2} = .025 \qquad t_{.025}(25\ df) = 2.06$ (Table III)

Reject Ho if $t < -2.06$ or $t > 2.06$.

$$t = \frac{68.5 - 70}{3.3/\sqrt{26}} = \frac{-1.5}{0.6472} = -2.31$$

$t = -2.31 < -2.06$, so we reject Ho. We conclude the population mean height of male USC students is not 70 inches.

Assumptions of t-test (and CI) about μ

- **We assume the data come from a population that is approximately normal.**
- **If this is not true, our conclusions from the hypothesis test may not be accurate (and our true level of confidence for the CI may not be what we specify).**
- **How to check this assumption?**

Q-Q plot, histogram

- **The t-procedures are robust: If the data are "close" to normal, the t-test and t CIs will be quite reliable.**

- If sample size is large, t-test and t CIs will generally be reliable (CLT)

Hypothesis Tests about a Population Proportion

We often wish to test whether a population proportion *p* equals a specified value.

Example 1: We suspect a theater is letting underage viewers into R-rated movies. Question: Is the proportion of R-rated movie viewers at this theater greater than 0.25? who are underage

We test: $H_0: p = 0.25$

$H_a: p > 0.25$

Recall: The sample proportion $\hat{p}$ is approximately $N\left(p, \sqrt{\frac{pq}{n}}\right)$ for large *n*, so our test statistic for testing H_0: $p = p_0$

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}}$$

has a standard normal distribution when H_0 is true (when *p* really is p_0).

Note $q_0 = 1 - p_0$

Rules for one-tailed tests about population proportion

$H_0: p = p_0$ or $H_0: p = p_0$
$H_a: p < p_0$ $H_a: p > p_0$

Test statistic: $$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}}$$

Rejection Region: $z < -z_\alpha$ $z > z_\alpha$

Rules for two-tailed tests about population proportion

$H_0: p = p_0$
$H_a: p \neq p_0$

Test statistic: $$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}}$$

$|z| > z_{\alpha/2}$

Rejection Region: $z < -z_{\alpha/2}$ or $z > z_{\alpha/2}$ (both tails)

Assumptions of test (need large sample):

Need: $p_0 - 3\sqrt{\frac{p_0 q_0}{n}} \geq 0$

and $p_0 + 3\sqrt{\frac{p_0 q_0}{n}} \leq 1$

(similar to the check for CI about p)

Alternatively, if $np_0 \geq 15$ and $nq_0 \geq 15$, the test is valid.

Example 1:
Test H_0: $p = 0.25$ vs. H_a: $p > 0.25$ using $\alpha = .01$.

We randomly select 60 viewers of R-rated movies, and 23 of those are underage.

$p_0 = 0.25 \Rightarrow$ $\quad .25 - 3\sqrt{\frac{(.25)(.75)}{60}} = .082 \geq 0$ ✓

$\quad .25 + 3\sqrt{\frac{(.25)(.75)}{60}} = .418 \leq 1$ ✓

$\hat{p} = \frac{23}{60} = .383$

Rejection Region:
Reject H_0 if $z > z_{.01}$
$\Rightarrow$ if $z > 2.326$
bottom row, t-table

$$z = \frac{.383 - .25}{\sqrt{\frac{(.25)(.75)}{60}}} = \frac{.133}{.056} = \boxed{2.38}$$

Since $z = 2.38 > 2.326$, we reject H_0 and conclude the population proportion of underage viewers is greater than 0.25.

Example 1(a): What if we had wanted to test whether the proportion of underage viewers was <u>different from</u> 0.25? $H_0: p = 0.25$ vs. $H_a: p \neq 0.25$.

For the same data, $z = 2.38$, still.

Note $z_{\alpha/2} = z_{.005} = 2.576$ (bottom row, t-table)

Reject H_0 if $z < -2.576$ <u>or</u> $z > 2.576$.

Here $2.38 \not> 2.576$, so we would have failed to reject H_0. There would <u>not</u> have been sufficient evidence to conclude that the true proportion of underage viewers is different from 0.25.