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 $Z = \frac{\overline{X} - 65}{\sigma / \sqrt{n}}$ 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?)

Picture:

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

$P(Z > 1.645 \mid \mu = 65) =$

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:

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.

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: $t = \frac{\overline{X} - \mu_0}{s/\sqrt{n}}$**, 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: $H_0: \mu = 65$

$H_a: \mu > 65$

Sample 64 hotels, get $\bar{X}$ = \$67 and *s* = \$10.
Let's set $\alpha = .05$.

Rejection region:

Reject H_0 if *t* is bigger than 1.67.

Conclusion:

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

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

Only very <u>large</u> 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 $t = \frac{\overline{X} - \mu_0}{s/\sqrt{n}}$ **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: $$t = \frac{\overline{X} - \mu_0}{s/\sqrt{n}}$$

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

(where t_α is based on $n - 1$ d.f.)

Rules for two-tailed tests about population mean

H_0: $\mu = \mu_0$

H_a: $\mu \neq \mu_0$

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

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

(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.

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

<u>Assumptions of t-test (and CI) about μ</u>

- **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?**

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

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?

We test:

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 $H_0: p = p_0$

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

Rules for one-tailed tests about population proportion

H_0: p = p_0		**H_0: p = p_0**
H_a: p < p_0	**or**	**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 ≠ p_0

Test statistic:

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

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

Assumptions of test (need large sample):

Need:

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.

Example 1(a): What if we had wanted to test whether the proportion of underage viewers was <u>different from</u> 0.25?

P-values

Recall that the significance level α is the desired P(Type I error) that we specify before the test.

The P-value (or "observed significance level") of a test is the probability of observing as extreme (or more extreme) of a value of the test statistic than we did observe, if H_0 was in fact true.

The P-value gives us an indication of the strength of evidence against H_0 (and for H_a) in the sample.

This is a different (yet equivalent) way to decide whether to reject the null hypothesis:

- **A small p-value (less than α) = strong evidence against the null => Reject H_0**

- **A large p-value (greater than α) = little evidence against the null => Fail to reject H_0**

How do we calculate the P-value? It depends on the alternative hypothesis.

One-tailed tests

Alternative	P-value
H_a: " < "	**Area to the left** of the test statistic value in the appropriate distribution (t or z).
H_a: " > "	**Area to the right** of the test statistic value in the appropriate distribution (t or z).

Two-tailed test

Alternative	P-value
H_a: " ≠ "	**2 times the "tail area"** outside the test statistic value in the appropriate distribution (t or z). **Double** the tail area to get the P-value!

P-values for Previous Examples

Hotel Price Example: H_0: $\mu = 65$ vs. H_a: $\mu > 65$

Test statistic value:

Student height example: H_0: $\mu = 70$ vs. H_a: $\mu \neq 70$

Test statistic value:

Movie theater example: H_0: $p = 0.25$ vs. H_a: $p > 0.25$

Test statistic value:

What if we had done a two-tailed test of H_0: $p = 0.25$ vs. H_a: $p \neq 0.25$ at $\alpha = .01$?