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

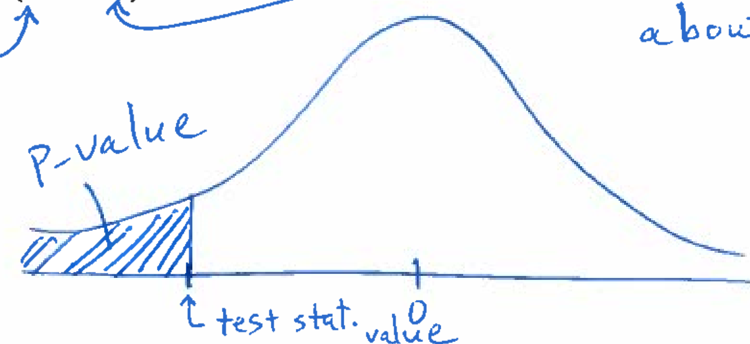
Alternative | **P-value**

H_a: " < "

Area to the left of the test statistic value in the appropriate distribution (t or z).

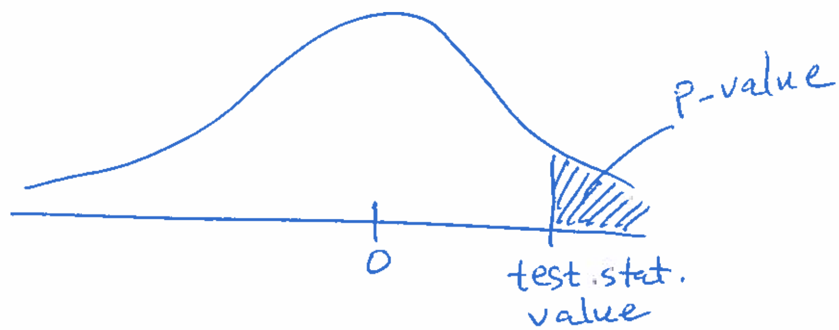
for tests about μ

for tests about p



H_a: " > "

Area to the right of the test statistic value in the appropriate distribution (t or z).

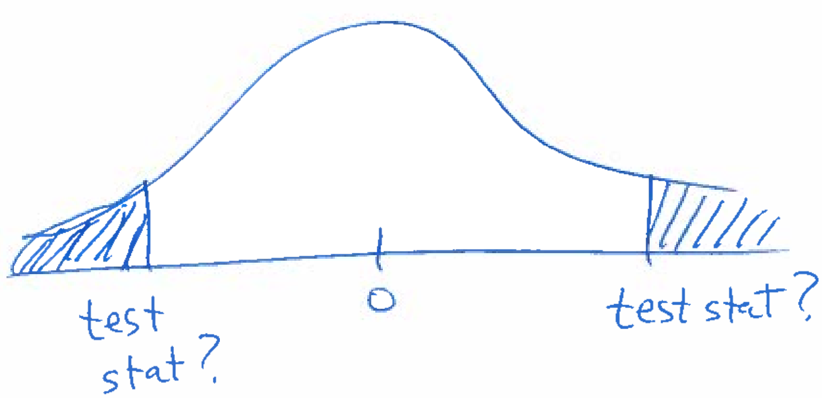


Two-tailed test

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!

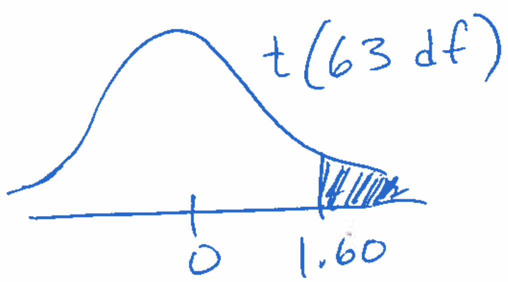


P-values for Previous Examples

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

Test statistic value: $t = 1.60$ (63 d.f.)

P-value = area to the right of 1.60 in t distribution with 63 d.f.



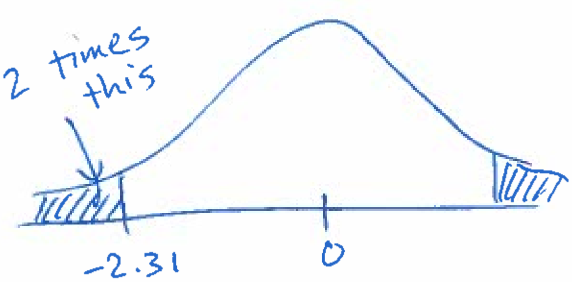
P-value ≈ .06
(definitely between .05 and .10)

Recall $\alpha = .05$. So P-value > α. Therefore the evidence against H_0 is <u>not</u> strong enough to reject H_0.

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

Test statistic value: $t = -2.31$ (25 d.f.)

P-value = Double the tail area outside −2.31 in a t-distribution with 25 d.f.



tail area ≈ .015
(between .025 and .01)

P-value ≈ 2(.015) = .03

– Recall $\alpha = .05 \Rightarrow$ P-value < α, so we have strong enough evidence to reject H_0.

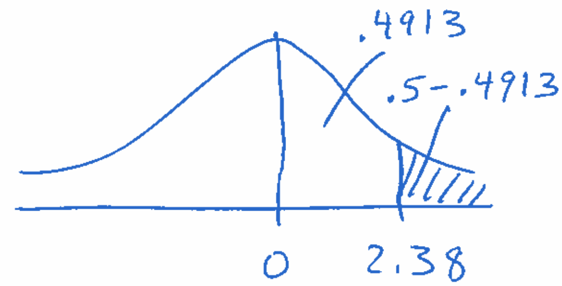
Movie theater example: H_0: $p = 0.25$ vs. H_a: $p > 0.25$

Test statistic value: $z = 2.38$

P-value is area to <u>right</u> of 2.38 in the std. normal (z) distribution:

From z-table (Table II),

Tail area = .0087 = P-value



Using $\alpha = .01$, our P-value $< \alpha$.

We have strong evidence to reject H_0.

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

P-value would have been <u>double</u> the tail area outside 2.38. P-value $= 2(.0087) = .0174$

In that case, P-value would have been $> \alpha$, so we would <u>not</u> have had strong enough evidence to reject H_0 in the 2-tailed test.

Relationship between a CI and a (two-sided) hypothesis test:

- A test of H_0: $\mu = m^*$ vs. H_a: $\mu \neq m^*$ will reject H_0 if and only if a corresponding CI for μ does not contain the number m^*.

Example: A 95% CI for μ is (2.7, 5.5).

(1) At $\alpha = 0.05$, would we reject H_0: $\mu = 3$ in favor of H_a: $\mu \neq 3$?

(2) At $\alpha = 0.05$, would we reject H_0: $\mu = 2$ in favor of H_a: $\mu \neq 2$?

(3) At $\alpha = 0.10$, would we reject H_0: $\mu = 2$ in favor of H_a: $\mu \neq 2$?

(4) At $\alpha = 0.01$, would we reject H_0: $\mu = 3$ in favor of H_a: $\mu \neq 3$?

Power of a Hypothesis Test

• Recall the significance level α is our desired P(Type I error) = P(Reject H_0 | H_0 true)

The other type of error in hypothesis testing:
Type II error = "Fail to reject H_0 | H_0 false"

P(Type II error) = β = P(Fail to reject H_0 | H_0 false)

The power of a test is P(Reject H_0 | H_0 false)

$= 1 - \beta$

• High power is desirable, but we have little control over it (different from α)

<u>Calculating Power</u>: The power of a test about μ depends on several things: α, n, σ, and the true μ.

<u>Example 1</u>: Suppose we test whether the true mean nicotine contents in a population of cigarettes is greater than 1.5 mg, using $\alpha = 0.01$.

H_0: $\mu = 1.5$ $\qquad$ H_a: $\mu > 1.5$

We take a random sample of 36 cigarettes. Suppose we know $\sigma = 0.20$ mg. Our test statistic is

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{\bar{X} - 1.5}{0.20/\sqrt{36}}$$

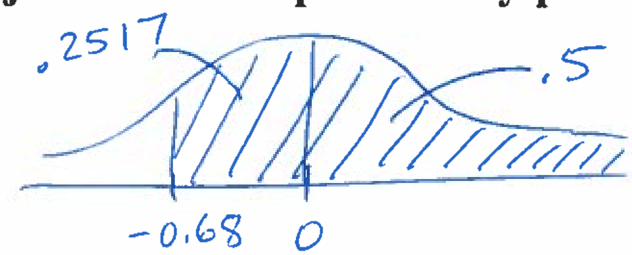
We reject H_0 if: $Z > Z_{.01} = 2.326$

$$\Rightarrow \frac{\bar{X} - 1.5}{0.20/\sqrt{36}} > 2.326 \Rightarrow \bar{X} - 1.5 > 0.0775$$
$$\Rightarrow \bar{X} > 1.5775$$

- **Now, suppose μ is actually 1.6 (implying that H_0 is false). Let's calculate the power of our test if $\mu = 1.6$:**

$$P(\bar{X} > 1.5775 \mid \mu = 1.6) = P\left(\frac{\bar{X} - 1.6}{0.20/\sqrt{36}} > \frac{1.5775 - 1.6}{0.2/\sqrt{36}}\right)$$
$$= P(Z > -0.68)$$

This is just a normal probability problem!



$$P(Z > -0.68) = .7517$$

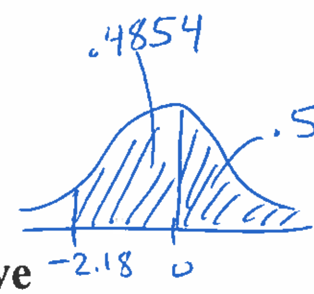
$$\Rightarrow P(\text{Reject } H_0 \mid \mu = 1.6) = .7517$$

$\Rightarrow$ Power when $\mu = 1.6$ is .7517

- **What if the true mean were 1.65?**

Verify: $P(\bar{X} > 1.5775 \mid \mu = 1.65)$

$$= P(Z > -2.18) = .9854$$



- **The farther the true mean is into the "alternative region," the more likely we are to correctly reject H_0.**

<u>Example 2</u>: Testing H_0: $p = 0.9$ vs. H_a: $p < 0.9$ at $\alpha = 0.01$ using a sample of size 225.

Suppose the true p is 0.8. Then our power is: