STAT 515 Lec 15 slides

p values

Karl Gregory

University of South Carolina

These slides are an instructional aid; their sole purpose is to display, during the lecture, definitions, plots, results, etc. which take too much time to write by hand on the blackboard. They are not intended to explain or expound on any material.

Discuss: Two researchers will test H_0 vs H_1. Each gathers data, and:

- The first rejects H_0 based on a test with significance level $\alpha = 0.10$ and
- The second rejects H_0 based on a test with significance level $\alpha = 0.01$.

Whose result is more "significant"?

🤔

At what significance levels would the observed data lead to a rejection of H_0?

This is a way to measure the strength of observed evidence against H_0.

The p value

The smallest significance level α at which the observed data would lead to a rejection of H_0 is called the p *value*.

Interpretion: Probability (under H_0) of observing data that carry as much or more evidence against the null as the data observed.

Once we have the p value, we reject H_0 if p value $< \alpha$.

The p value is also called the *observed significance level*.

[Fly example from Lec 14]

(2)

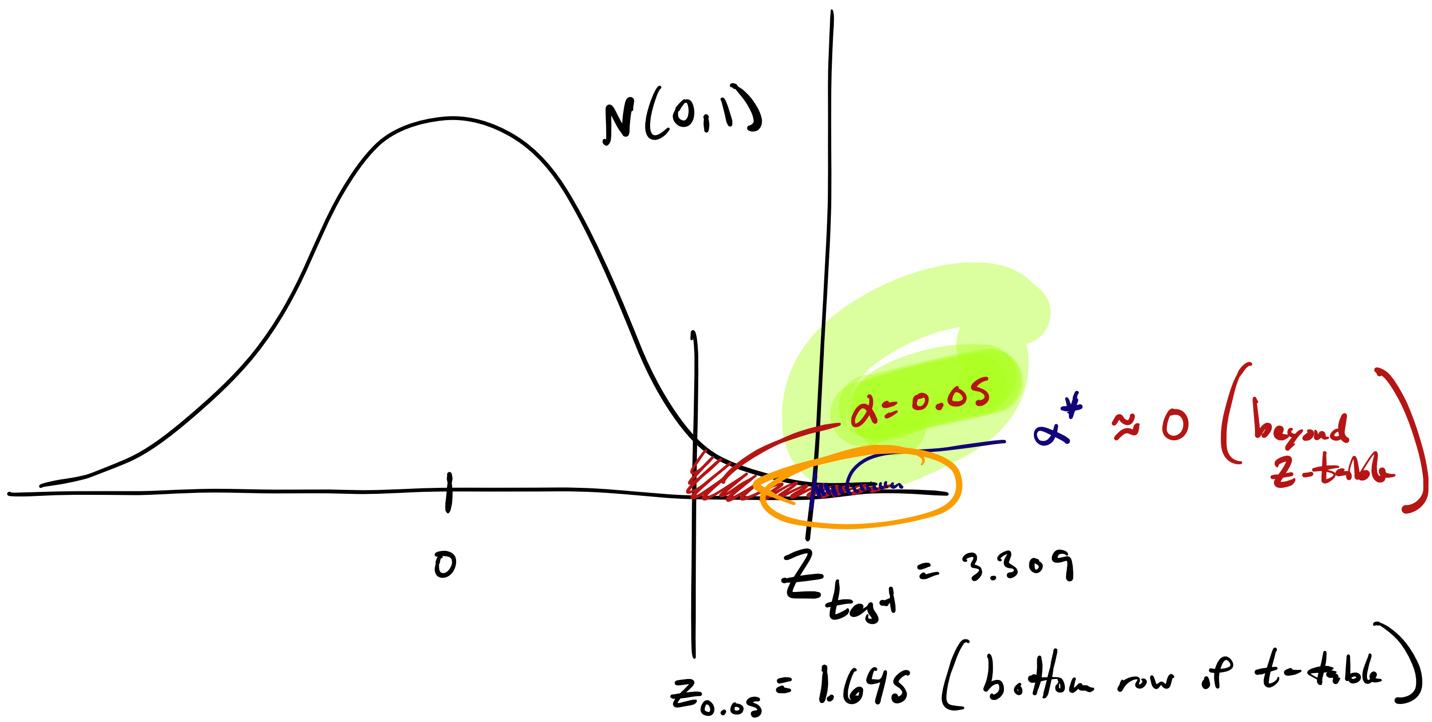
$\alpha = 0.05$ $\quad p_0 = 1/2 \quad \hat{p}_n = 0.574$

$n = 500$

$H_0: p \le 1/2$ vs $H_1: p > 1/2$

$$Z_{test} = \frac{0.574 - 0.5}{\sqrt{\frac{0.5(1-0.5)}{500}}} \qquad Z_{test} = \frac{\hat{p}_n - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

$$= 3.309$$



Find p-value by finding α^* such that

$$z_{\alpha^*} = Z_{test}.$$

In above case — just find area under curve to the right of Z_{test}.

Let $X_1, \dots, X_n \overset{\text{ind}}{\sim} \text{Normal}(\mu, \sigma^2)$, with μ and σ^2 unknown.

Tests about μ when σ is unknown

For some null value μ_0, define the test statistic

$$T_{\text{test}} = \frac{\bar{X}_n - \mu_0}{S_n/\sqrt{n}}.$$

Then the following tests have $P(\text{Type I error}) \leq \alpha$.

$H_0\colon \mu \geq \mu_0$ $H_1\colon \mu < \mu_0$	$H_0\colon \mu = \mu_0$ $H_1\colon \mu \neq \mu_0$	$H_0\colon \mu \leq \mu_0$ $H_1\colon \mu > \mu_0$
Reject H_0 if $T_{\text{test}} < -t_{n-1,\alpha}$	Reject H_0 if $\lvert T_{\text{test}} \rvert > t_{n-1,\alpha/2}$	Reject H_0 if $T_{\text{test}} > t_{n-1,\alpha}$
p val $= P(T < T_{\text{test}})$	p val $= 2 \cdot P(T > \lvert T_{\text{test}} \rvert)$	p val $= P(T > T_{\text{test}})$

For computing the p values, let $T \sim t_{n-1}$. **Draw pictures.**

Exercise: A machine should produce ball bearings with Normally distributed diameters having mean 0.5 inches. Is the mean truly 0.5 inches.? (Ex 6.84 in [1]).

With $n = 5$ you get $\bar{X}_n = 0.499$ and $S_n = 0.001$. Find the p value for testing

1. H_0: $\mu \geq 0.5$ vs H_1: $\mu < 0.5$ μ_0
2. H_0: $\mu \leq 0.5$ vs H_1: $\mu > 0.5$
3. H_0: $\mu = 0.5$ vs H_1: $\mu \neq 0.5$

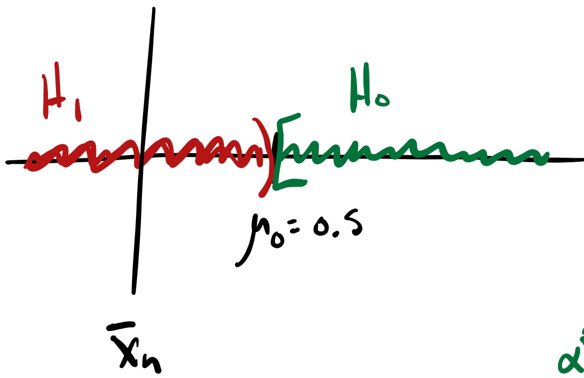
$\boxed{\alpha = 0.05}$

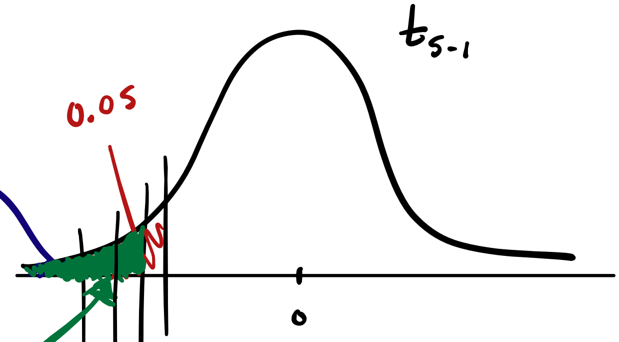
$$T_{test} = \frac{\bar{X}_n - \mu_0}{S_n/\sqrt{n}}$$

$$= \frac{0.499 - 0.5}{0.001/\sqrt{5}}$$

$$= -2.236.$$

①





$-t_{5-1,0.05} = -2.1318$

$\boxed{T_{test} = -2.236}$

$-t_{5-1,0.025} = -2.7764$

$-t_{5-1,0.01} = -3.7469$

α^*, p-value.

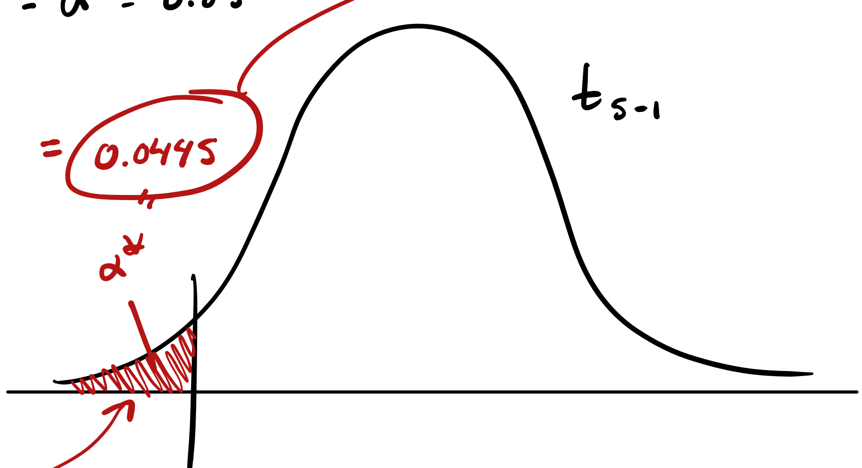
$0.01 \le \alpha^* \le 0.05$

ν	0.100	0.050	0.025	0.010	0.005	0.001	0.0005
			α				
1	3.0777	6.3138	12.7062	31.8205	63.6567	318.3088	636.6192
2	1.8856	2.9200	4.3027	6.9646	9.9248	22.3271	31.5991
3	1.6377	2.3534	3.1824	4.5407	5.8409	10.2145	12.9240
4	1.5332	2.1318	2.7764	3.7469	4.6041	7.1732	8.6103

$0.025 \le \alpha^* \le 0.05$

p-value.

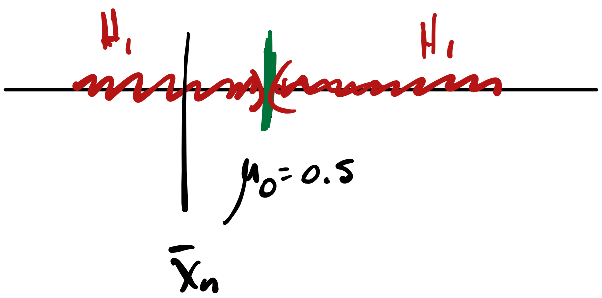
$pt(-2.236, 4) = 0.0445$

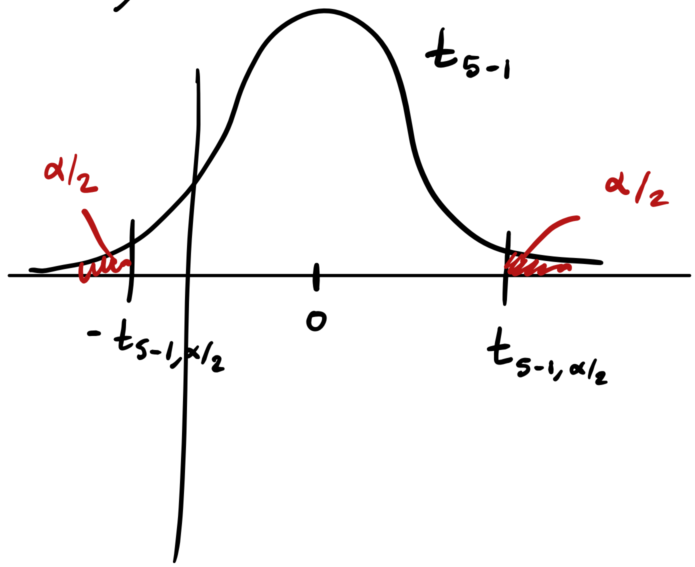


$T_{test} = -2.236$

$pt(T_{test}, 5-1)$, gives area under t_{5-1} pdf to the left of T_{test}.

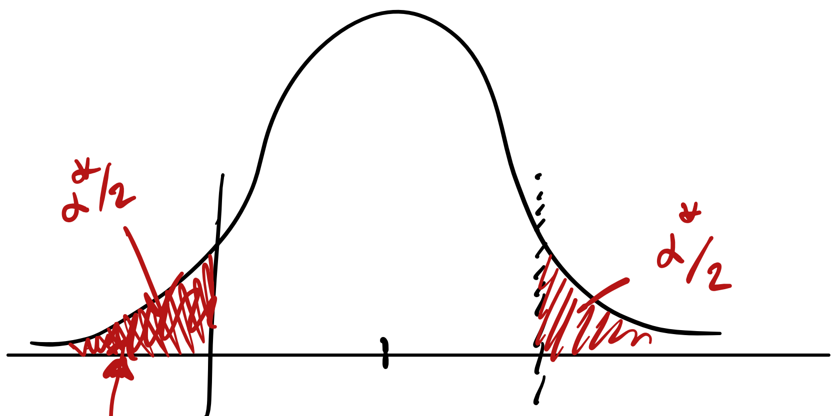
③ $H_0: \mu = 0.5$ vs $H_1: \mu \neq 0.5$





$\alpha = 0.05$: $-t_{5-1,\frac{0.05}{2}} = -2.7764$

$T_{test} = -2.236$



0.0445

$T_{test} = -2.236$

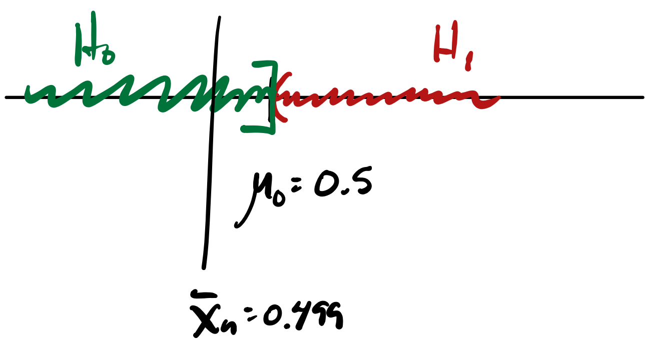
$$\alpha^* = 2 * 0.0445 = \boxed{0.089.}$$

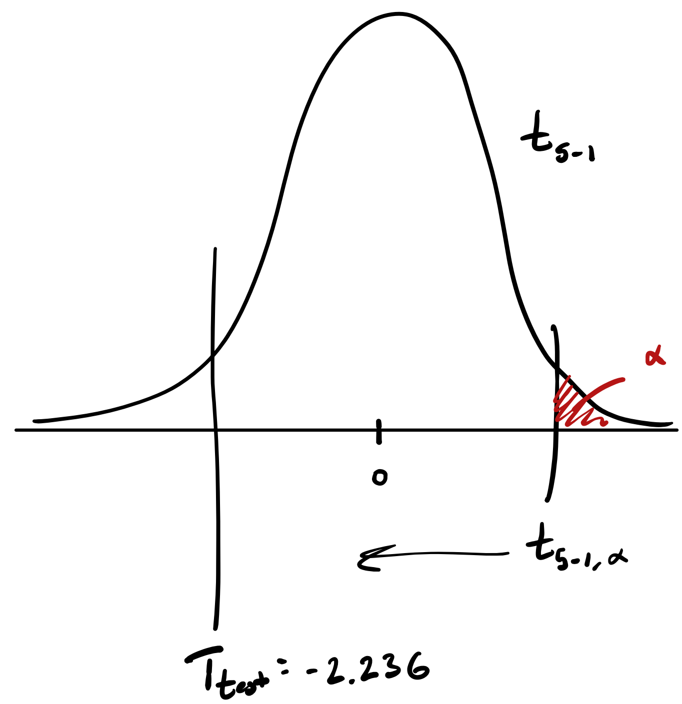
p-value

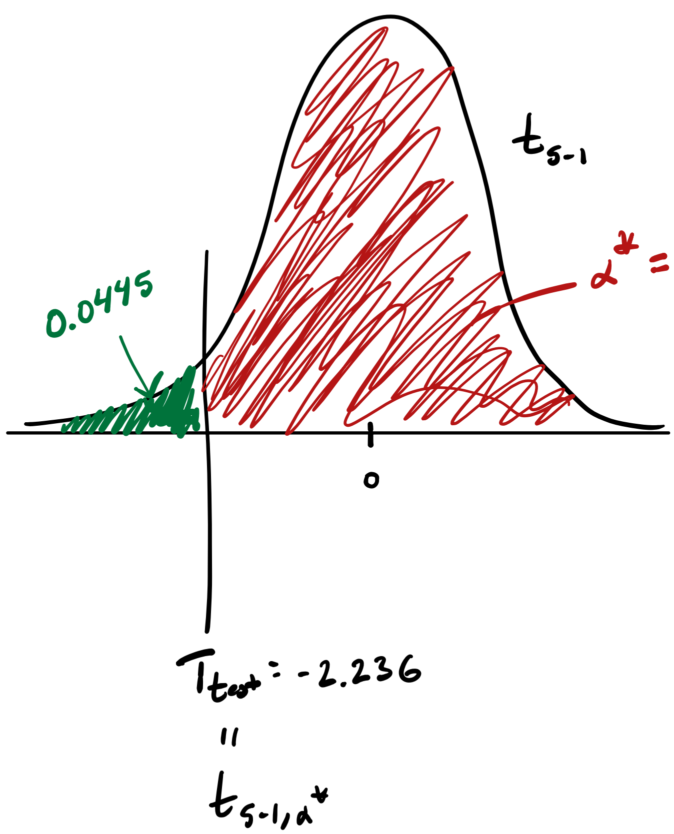
To test H_0 vs H_1 at significance level α, you can just compute the p-value, and reject H_0 if p-value $< \alpha$.

② $H_0: \mu \le 0.5$ vs $H_1: \mu > 0.5$

$\bar{X}_n = 0.499$







$\alpha^* = 1 - 0.0445 = \boxed{0.9555}$

p-value

Large, indicating very weak (possibly no) evidence against H_0.

A smaller p-value indicates stronger evidence against H_0.

```
n <- 5
xbar <- 0.499
sn <- 0.001
mu0 <- 0.5
Ttest <- (xbar - mu0)/(sn/sqrt(n))

pt(Ttest,n-1)
1 - pt(Ttest,n-1)
2*(1-pt(abs(Ttest),n-1))
```

Exercise: Suppose you wish to test whether the LDL (bad cholesterol) level of South Carolinians exceeds the nationwide mean of 150 mg/dl.

With $n = 20$ you get $\bar{X}_n = 162.5$ and $S_n = 27.6$. Find the p value for testing

1. H_0: $\mu \geq 150$ vs H_1: $\mu < 150$
2. H_0: $\mu \leq 150$ vs H_1: $\mu > 150$
3. H_0: $\mu = 150$ vs H_1: $\mu \neq 150$

Assume the LDL levels are Normally distributed.

```
n <- 20
xbar <- 162.5
sn <- 27.6
mu0 <- 150
Ttest <- (xbar - mu0)/(sn/sqrt(n))

pt(Ttest,n-1)
1 - pt(Ttest,n-1)
2*(1-pt(abs(Ttest),n-1))
```

Let $X_1, \dots, X_n$ iid non-Normal with mean μ and unknown variance σ^2.

Large-n tests about μ when data are $\boxed{\text{non-Normal}}$

$\boxed{\text{need } n \geq 30}$

For some null value μ_0, define the test statistic

$$T_{\text{test}} = \frac{\bar{X}_n - \mu_0}{S_n/\sqrt{n}}.$$

Then for large n, the following tests have (approximately) $P(\text{Type I error}) \leq \alpha$.

$H_0\colon \mu \geq \mu_0$ $H_1\colon \mu < \mu_0$	$H_0\colon \mu = \mu_0$ $H_1\colon \mu \neq \mu_0$	$H_0\colon \mu \leq \mu_0$ $H_1\colon \mu > \mu_0$
Reject H_0 if $T_{\text{test}} < -\underline{z_\alpha}$	Reject H_0 if $\lvert T_{\text{test}} \rvert > \underline{z_{\alpha/2}}$	Reject H_0 if $T_{\text{test}} > \underline{z_\alpha}$
p val $= P(Z < T_{\text{test}})$	p val $= 2 \cdot P(Z > \lvert T_{\text{test}} \rvert)$	p val $= P(Z > T_{\text{test}})$

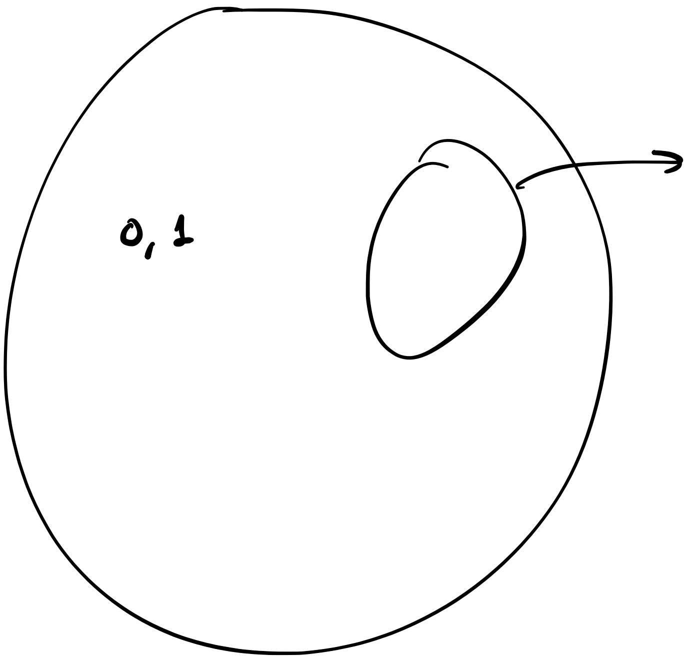
Time allowing:

1. Draw a random sample of size $n = 35$ from the 2009 Boston Marathon `women's finishing times` and compute the p value for testing

$$H_0\colon \mu \leq 4 \text{ versus } H_1\colon \mu > 4$$

2. Repeat this 1000 times and make a histogram of the p values.

population: p unknown "success" probability/proportion



$$X_1, \dots, X_n \overset{ind}{\sim} \text{Bernoulli}(p)$$

$$\hat{p}_n = \bar{X}_n = \frac{1}{n}(X_1 + \dots + X_n)$$

To test hypotheses about p, compute

$$Z_{test} = \frac{\hat{p}_n - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}.$$

Let $X_1, \dots, X_n \stackrel{\text{ind}}{\sim} \text{Bernoulli}(p)$.

Tests about p

For some null value p_0, define the test statistic

$$Z_{\text{test}} = \frac{\hat{p}_n - p_0}{\sqrt{p_0(1-p_0)/n}}.$$

Then the following tests have (approximately) $P(\text{Type I error}) \leq \alpha$.

$H_0\colon p \geq p_0$ $H_1\colon p < p_0$	$H_0\colon p = p_0$ $H_1\colon p \neq p_0$	$H_0\colon p \leq p_0$ $H_1\colon p > p_0$
Reject H_0 if $Z_{\text{test}} < -z_\alpha$	Reject H_0 if $\lvert Z_{\text{test}}\rvert > z_{\alpha/2}$	Reject H_0 if $Z_{\text{test}} > z_\alpha$
p val $= P(Z < Z_{\text{test}})$	p val $= 2 \cdot P(Z > \lvert Z_{\text{test}}\rvert)$	p val $= P(Z > Z_{\text{test}})$

Discuss: Draw pictures of how to get the p values.

$p = ?$

$p_0 = 0.10$

$\hat{p}_n = \frac{70}{527} = 0.133$

$n = 527$

$70 = X_1 + \cdots + X_n$

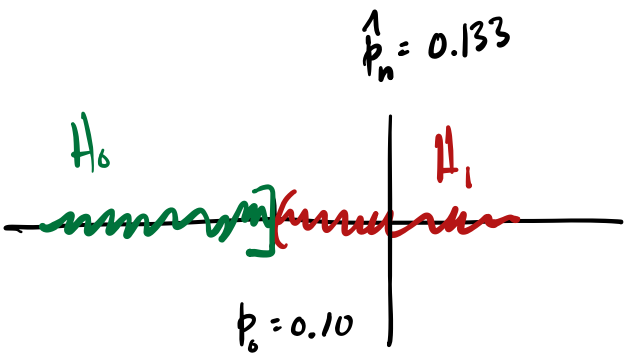
$>$

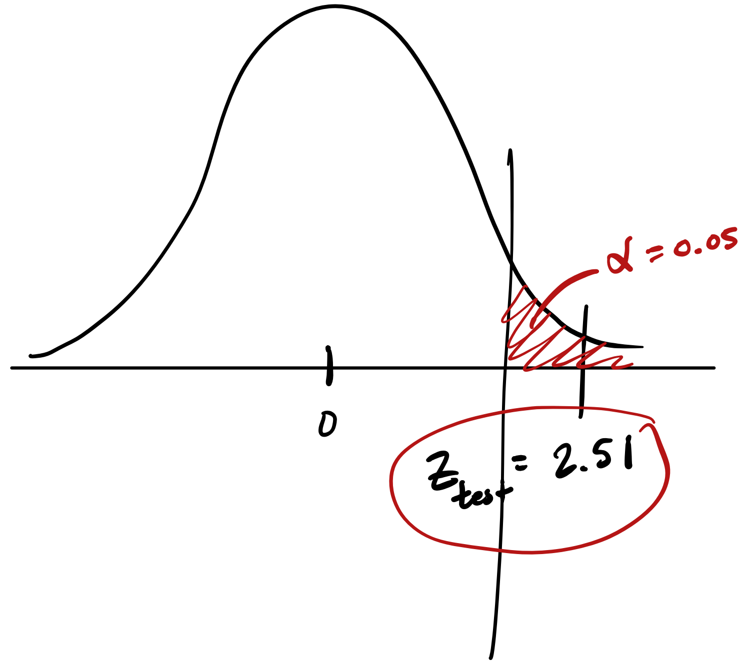
Exercise: The DNR will take action if an invasive fish is concluded to comprise $\boxed{\text{more than}}$ 10% of the fish population in a habitat. In a random sample of 527 fish, 70 were of the invasive species.

1. What are the appropriate null and alternate hypotheses?
2. What is the p value?
3. What would the p value be if the two-sided test were of interest?

(1) $H_0: p \leq 0.10$ vs $H_1: p > 0.10$

(2) $$Z_{test} = \frac{\hat{p}_n - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0.133 - 0.10}{\sqrt{\frac{0.10(1-0.10)}{527}}} = 2.51$$

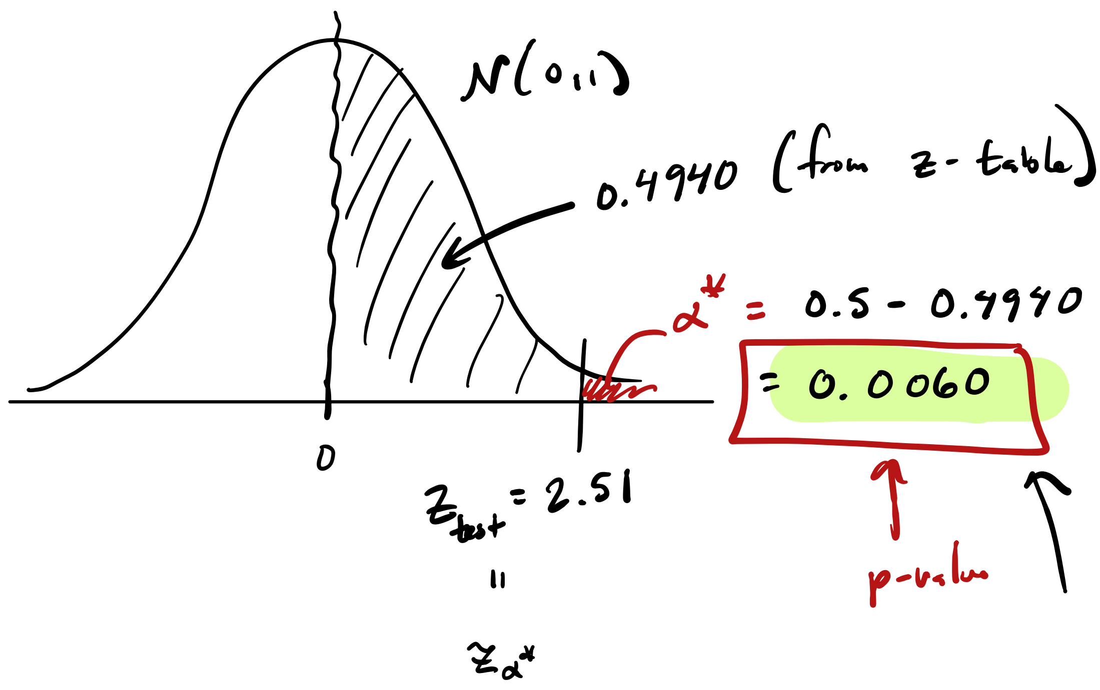




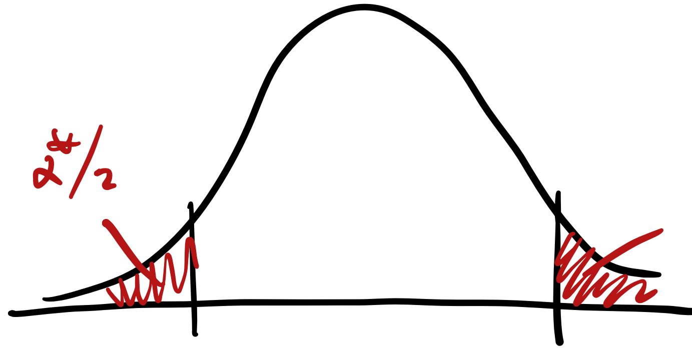
$z_{0.05} = 1.645$

$\alpha = 0.05$

$z_{0.05} = 1.645$



③ $H_0: \; p = 0.10$ vs $H_1: \; p \neq 0.10$



$\alpha^*/2 = 0.006$

double it!

$\Rightarrow \boxed{\alpha^* = 0.012}$

$Z_{test} = 2.51$

```
n <- 527
pn <- 70/527
p0 <- 0.10
Ztest <- (pn - p0)/sqrt(p0*(1-p0)/n)

1 - pnorm(Ztest)
2*(1 - pnorm(Ztest))
```

J.T. McClave and T.T. Sincich.
Statistics.
Pearson Education, 2016.