§9.1 Comparing Two Proportions

Summary Statistics

n_1 and n_2 are the Sample Sizes

x_1 and x_2 are the Number of Successes

$\hat{p}_1$ and $\hat{p}_2 = \frac{x_2}{n_2}$ is the proportion of Successes

PE: $\hat{p}_1 - \hat{p}_2$

H_0: $p_1 = p_2$ or $p_1 - p_2 = 0$

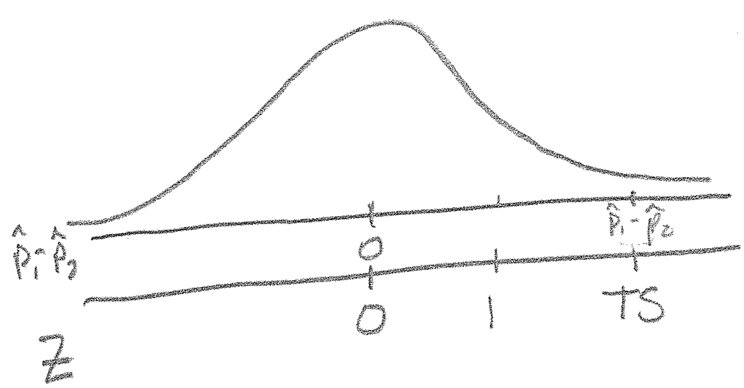
H_1: $p_1 \neq p_2$, $p_1 > p_2$, $p_1 < p_2$

Sampling distribution $\{\hat{p}_1 - \hat{p}_2\}$ ← set of sample differences

Follows a Normal Distribution

$\mu_{\hat{p}_1 - \hat{p}_2} = 0$

$\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}$



Use Calculator

2 prop Z Interval

2 prop Z Test

Test Statistic $\boxed{\text{2 prop Z Test}}$

$$TS: \; Z = \frac{\hat{p}_1 - \hat{p}_2 \quad - \quad 0}{\sqrt{\frac{\bar{p}\bar{q}}{n_1} + \frac{\bar{p}\bar{q}}{n_2}}}$$

← you don't have to use

$$\bar{p} = \frac{x_1 + x_2}{n_1 + n_2} =$$ weighed Average of Two proportions

Confidence Interval for the difference between the population proportions p_1 and p_2

$$(\hat{p}_1 - \hat{p}_2) - E < p_1 - p_2 < (\hat{p}_1 - \hat{p}_2) + E$$

$$E = Z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}$$

STAT Tests

use $\boxed{\text{2PROP Z Interval}}$

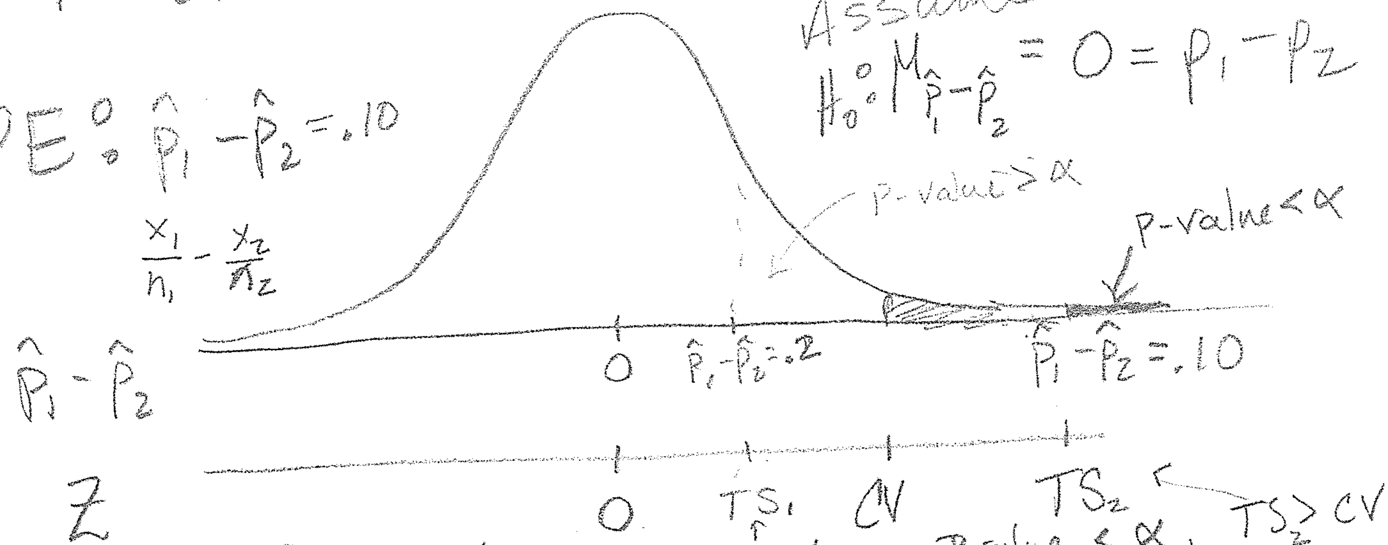
Since Sample Error is different in these two formulas, the Conclusion based on CI May Not be the Same as those based on the HT ← p-value & Tradition give same Answer

The Distribution - Z - or Normal

Assume
$H_0: \mu_{\hat{p}_1 - \hat{p}_2} = 0 = p_1 - p_2$

PE: $\hat{p}_1 - \hat{p}_2 = .10$

$\frac{x_1}{n_1} - \frac{x_2}{n_2}$



① p-value > α, $TS_1 < CV$
② there is Not a significant Difference

Pvalue < α, $TS_2 > CV$
there is a significant Difference

Confidence Interval uses

$$\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}} = \text{Standard Error}$$

Hypothesis Test uses $\bar{p} = \frac{x_1 + x_2}{n_1 + n_2}$ = Weighted Average of the $\hat{p}_1$ & $\hat{p}_2$

$\bar{q} = 1 - \bar{p}$

$$SE = \sqrt{\frac{\bar{p}\,\bar{q}}{n_1} + \frac{\bar{p}\,\bar{q}}{n_2}}$$

Traditional & p-value Method use same SE & K
So give same Conclusion

Since they use a different SE Conclusions May Not be the same. Best to Use HT.

Ex All SRJC students get free bus passes.

Before Bus Pass $n_1 = 250$ $x_1 = 57$ took bus

After Bus Pass $n_2 = 250$ $x_2 = 87$ too buss

Test the claim at $\alpha = .05$ level that α

the proportion riding bus after free passes

is <u>higher than</u> proportion before free pass.

a) Make a confidence Interval for $p_1 - p_2$

B: 2 Prop Z Interval is 90% $(-.186, -.054)$

$-.186 < p_1 - p_2 < -.054$

$x_1 = 57$

$n_1 = 250$

$x_2 = 87$

$n_2 = 25$

PE: $\hat{p}_1 - \hat{p}_2 = \frac{57}{250} - \frac{87}{250} = .228 - .348 = \boxed{-.12 = \hat{p}_1 - p_2}$

CL $= .9 = 1 - 2\alpha$ when One tailed or $1 - \alpha$ if 2-Tailed

↑ use the same cv as the HT CV: $z = (1 - \alpha/2, 0, 1)$

$H_0: p_1 = p_2$, $H_1: p_1 < p_2$

↑ Lt

b) Hypothosis test @ $\alpha = .05$

CV: $Z = \text{inv norm}(\alpha, 0, 1)$ ← Lt

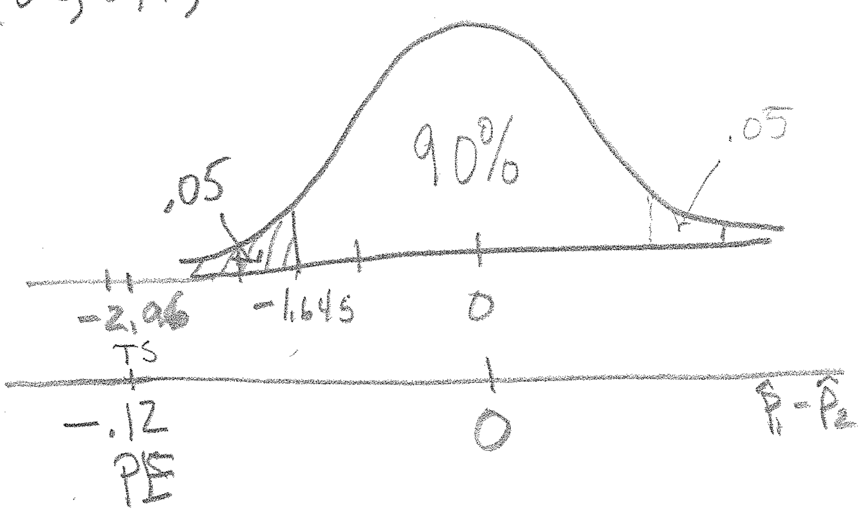
$Z^* = -1.645$

TS: $Z = -2.96$

P-Value $= .0015 < \alpha$

$p_1 < p_2$

$p_1 - p_2 < 0$



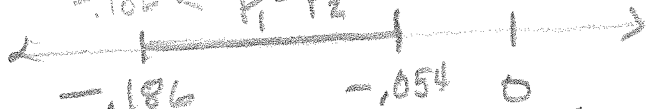
Reject H_0 & Suppor H_1

We Support that the proportion riding the Bus Increased.

Meaning of Confidence Interval

We are 90% that the difference in ridership Before − After is between −.186 and −.054.

$-.186 < p_1 - p_2 < -.054$



the proportion riding the bus increase between 5.4% and 18.6% After free bus passes.

When 0 is Not in the Confidence Interval there is a Significant Difference in the prop riding the bus.

Example All SRJC Students get free Bus passes.

Before Bus $n_1 = 250$ $x_1 = 57$ took Bus
After Bus for Free $n_2 = 250$ $x_2 = 87$ Taking Bus
} Summary Statistics

Question At the .05 level of significance test the claim that the proportion of all SRJC students taking the bus is higher after free bus passes were given.
Proportion befor is less than proportion after.

Requirements $n_1 p_1 > 5$ $n_1 q_1 > 5$ $n_2 p_2 > 5$ $n_2 q_2 > 5$
And SRS - Data is Not Biased.

Confidence Intervals 95%

1-Prop Z interval
1 Before (.176, .280)
2 After (.289, .407)
3 Before-After (-.199, -.041)

Inequality
$.176 < p_1 < .280$

$\boxed{-.199 < p_1 - p_2 < -.041}$

3 Sentences
1 We are 95% confident that the True prop. of all SRJC student that take the Bus is Between .176 and .280

(3) We are 95% confident that the true Popula
difference in proportions who ride Bus
is between −.20 and −.04. (Both Neg

Ridership increased between 4% & 20%
More Students are taking the Bus

$-.20 < p_1 - p_2 < -.04$

$p_1 - p_2 \neq 0$ because zero is Not in interval

$H_0: p_1 = p_2$ Reject

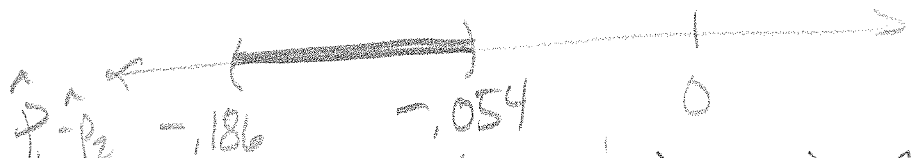
$H_1: p_1 < p_2$ Support

b) Make a 90% Confidence Interval

$CL = 1 - 2\alpha$ if it is one Tailed

$1 - 2(.05) = 90\%$

2-Prop Z Interval $(-.186, -.054)$



Zero is Not in interval $\Rightarrow P_1 - P_2 \neq 0 \Rightarrow$ there is a Significant Difference

c) Meaning of Confidence Interval?

the prop using bus before free Passes was between 18.6% ~~and~~ 5.4% lower than it is with free passes.

Note if $\hat{p}_1$ & $\hat{p}_2$ are switched then

CI $(.054, .186)$

Ridership increased between 5.4% and 18.6%

Hypothesis Test $\alpha = .05$

Claim: $p_1 < p_2$ prop Before is less than prop. After
p_1 p_2
Pop. of all SRJC Students

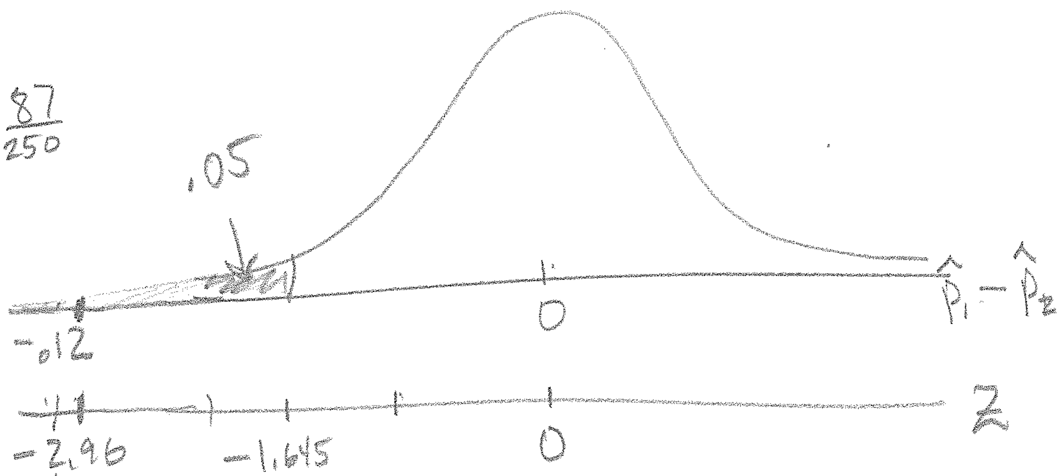
a) ① $H_0: p_1 = p_2$
$H_1: p_1 < p_2$ Same
Left Tail

② CV: $Z^* = \text{invn}(\alpha, 0, 1) = \text{invn}(.05, 0, 1) = -1.645$

b) PE: $\hat{p}_1 - \hat{p}_2 = \frac{57}{250} - \frac{87}{250}$

$= .228 - .348$

4) $\boxed{PE = -.12}$



TS: Z = use 2prop Z Test
Highlight Alt hypothosis $H_1: p_1 < p_2$

③ $\boxed{TS: Z = -2.96}$
p-value $= .0015 < .05 = \alpha$

c) Reject $H_0: p_1 = p_2$ Support $H_1: p_1 < p_2$

d) Words: We Reject that the same proportion are riding the Buss

Conclusion: We Support that the proportion riding bus Before free passes is less than proportion after free passes

Meaning of P-value

It is very unlikely p-value = .0015
that we would see a Sample difference
of .12 if the proportion
riding the bus had Not changed.

Use the traditional method to test the given hypothesis. Assume that the samples are independent and that they have been randomly selected

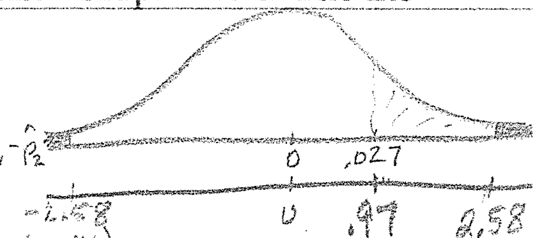
1) A marketing survey involves product recognition in New York and California. Of 558 New Yorkers surveyed, 193 knew the product while 196 out of 614 Californians knew the product. At the 0.01 significance level, test the claim that the recognition rates are the same in both states.

a) (5 Points) State the null and alternate hypothesis. Graph and shade the critical region. Find the critical value.

NY $\hat{p}_1 = \frac{193}{558} = .346$ $H_0: p_1 = p_2$

CA $\hat{p}_2 = \frac{196}{614} = .319$ $H_1: p_1 \neq p_2$

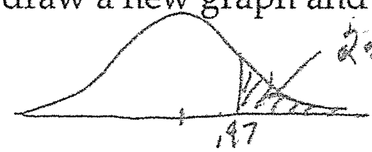
$\alpha = .01$ CV: $z = \text{invnorm}(1 - \alpha/2) = \pm 2.58$



b) (5 Points) Find the point estimate for the difference between the sample proportions and the test statistic for this difference. Label this on your graph. State the initial conclusion. 2-Prop Z Test

PE: $\hat{p}_1 - \hat{p}_2 = .346 - .319 = .027$ TS: $z = .97$

c) (5 points) Find the p–value, draw a new graph and label this area. Explain the meaning of this p–value.



P-value = .333 $> \alpha$
$< \alpha$

It would be usual to see a sample of difference of .027, when there is No Difference in recognition rates.

d) (5 Points) Clearly state your final conclusion in language addressing the original claim.

Initial fail to Reject Ho

Final ~~we failed to~~ Reject that recognition

TINSE to rates are the same.

TINSE ~~to~~ can't Support that they are different

Construct the indicated confidence interval for the difference between population proportions p1 – p2.
Requirements: Assume that the samples are independent and that they have been randomly selected.

2) A marketing survey involves product recognition in New York and California. Of 558 New Yorkers surveyed, 193 knew the product while ~~196~~ 96 out of 614 Californians knew the product. Construct a 99% confidence interval for the difference between the two population proportions.

a) (2 Points) What point estimate of the difference in the population proportions, **p1 – p2**, does this survey give? $\hat{p}_1 - \hat{p}_2 = \frac{193}{558} - \frac{96}{614} =$

b) (3 Points) What is the margin of error? $.346 - .156 = .190$

$$E = z_{\alpha/2} \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}} = 2.58 \sqrt{\frac{.346 \cdot .654}{558} + \frac{96/614 \cdot 518/614}{614}} = \boxed{.0645}$$

c) (5 Points) Find the confidence interval. $\boxed{(.125, .254)}$

☆ 2PropZInt

$$E = \frac{.254 - .125}{2} = .0645 \quad PE \pm E = \hat{p}_1 - \hat{p}_2 \pm E$$

$$\boxed{.190} \pm .0645$$

e) (5 Points) Interpret the meaning of this confidence interval.

$.125 < p_1 - p_2 < .254$, NY'ers recognize the product at a rate between 12.5% and 25.4% more often than CA's

3) When making decisions about a population means or proportions, either a hypothesis test or a confidence interval can be used. Compare these techniques.

Use Same CV; even for one tailed Test
Same answers if Same CV are used
Almost Always. Test uses a different Standard Error than CI, so Not Always the same.

§8.3 #22 Claim: $\mu = 60$ $H_0: \mu = 60$ $H_1: \mu \neq 60$

$\boxed{PE: \bar{x} = 62.7}$ $s = 19.5$ $n = 15$

CV: $t^* = \text{invT}(1 - \alpha/2, df) = \text{invT}(1 - .05/2, 14) = \pm 2.145$

$\boxed{CV: t^* = \pm 2.145}$

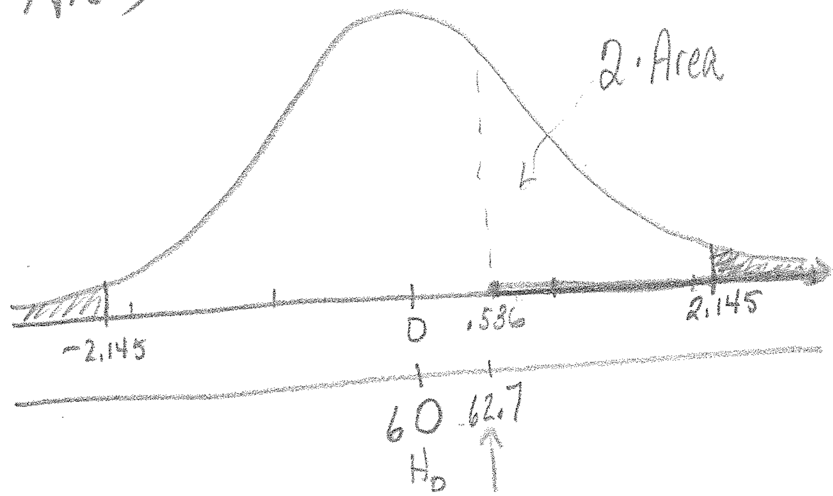
TS: $t = \dfrac{\bar{x} - \mu}{s/\sqrt{n}} = \dfrac{(62.7 - 60)}{(19.5/\sqrt{15})} = .536$

$\boxed{TS: t = .536}$

P-value = 2·tcdf (LB, UB, df)

= 2tcdf(.536, 9999, 14)

= .6004 > .05 = α



.604

fail to Reject H_0

TINSE TO Reject that the mean minute guess is 60 seconds.

the s = 19 seconds is very large so their is alot of variability in their guessing.