I. Two proportions

The two proportion is dealing with two different probability, namely P_1 and P_2. We would test a claim that about the population in regards to P_1 and P_2. Also, we may find a difference of P_1 and P_2's confidence interval.

Since we have two kinds of samples, in which we want to see the difference between them.

$$\text{Goal: } H_0 : P_1 = P_2 \quad \leftarrow P_1 \text{ vs } P_2$$
$$H_1 : P_1 < P_2 \quad \leftarrow \text{test}$$

$$\text{eg Money: } H_0 : P_1 = P_2 \quad \leftarrow P_1 \text{ is Honda, } P_2 \text{ is Toyota}$$
$$H_1 : P_1 < P_2 \quad \leftarrow \text{test with reality}$$

Now, for the two samples

$P_1 :=$ population 1 $\longleftrightarrow$ $P_2 :=$ population 2

$n_1 =$ size of sample 1 $\longleftrightarrow$ $n_2 :=$ size of sample 2

$x_1 =$ 'success' of n_1 $\longleftrightarrow$ $x_2 :=$ 'success' of n_2

<u>With</u>:

$$\underset{\text{success}}{\hat{P}_1} = \frac{x_1}{n_1}, \quad \underset{\text{failure}}{\hat{q}_1} = 1 - \hat{P}_1 \quad \longleftrightarrow \quad \underset{\text{success}}{\hat{P}_2} = \frac{x_2}{n_2}, \quad \underset{\text{failure}}{\hat{q}_2} = 1 - \hat{P}_2$$

Then, we have the new <u>pooled sample</u>, $\bar{P}$ is called "p-pooled"

Then, we have the new pooled sample, $\bar{p}$ is called "p-pooled"

'success' → $\bar{p} = \dfrac{x_1 + x_2}{n_1 + n_2}$

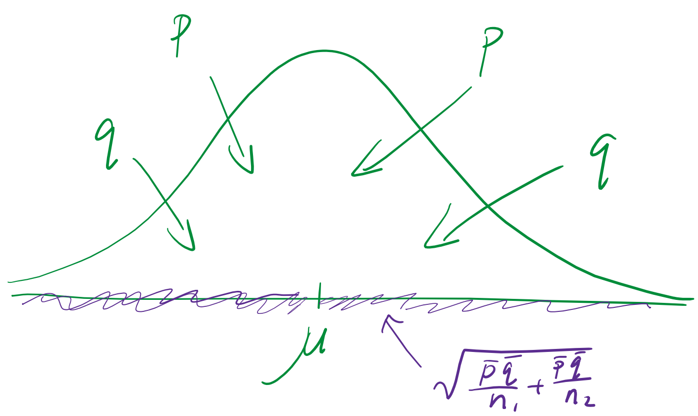
'failure' → $\bar{q} = 1 - \bar{p}$

and, the test statistics is given by

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{\bar{p}\bar{q}}{n_1} + \frac{\bar{p}\bar{q}}{n_2}}}$$

10 Hondas, 8 Toyotas

$$\frac{\vdots \quad \vdots}{10 + 8}$$



Confidence interval:

For the two samples $\hat{p}_1$, $\hat{p}_2$, and the population proportion $p_1 - p_2$,

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

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

eg $0.5 < p_1 - p_2 < 0.6$

eg $-0.2 < p_1 - p_2 < 0.15 \leftarrow$ 0 is there: no difference

Eg Do people having different spending habits depending on the type of money they have?

89 undergraduates were randomly assigned to two groups and were given a choice of keeping the money or buying gum or mints. The claim is that "money in large denominations is less likely to be spent relative to an equivalent amount in many smaller denominations". We test the claim at the 0.05 significance level.

Below are the sample data and summary statistics:

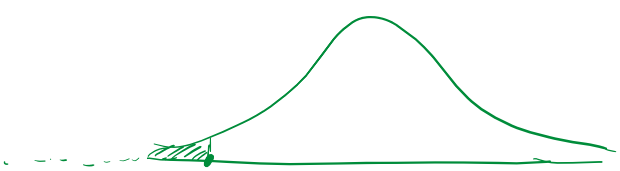
	Group 1	Group 2
	Subjects Given \$1 Bill	Subjects Given 4 Quarters
Spent the money	$x_1 = 12$	$x_2 = 27$
Subjects in group	$n_1 = 46$	$n_2 = 43$
	for p_1	for p_2

S:

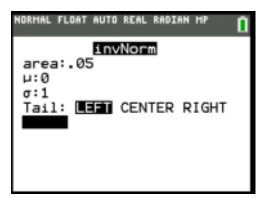
$H_0: p_1 = p_2$

$H_1: p_1 < p_2$

$\alpha = 0.05$, with $<$



−1.64



$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{\bar{p}\bar{q}}{n_1} + \frac{\bar{p}\bar{q}}{n_2}}}$$

$\hat{p}_1 = \frac{12}{46}$, $\hat{p}_2 = \frac{27}{43}$, $\bar{p} = \frac{12+27}{46+43} = \frac{39}{89}$, $\bar{q} = 1 - \bar{p} = 1 - \frac{39}{89} = \frac{50}{89}$

then, $$z = \frac{\left(\frac{12}{46} - \frac{27}{43}\right) - 0}{\sqrt{\frac{\frac{39}{89}\cdot\frac{50}{89}}{46} + \frac{\frac{39}{89}\cdot\frac{50}{89}}{43}}}$$

$H_0: p_1 = p_2$ ✓

$\boxed{p_1 = p_2}$

$\sim$ [illegible]

46 43

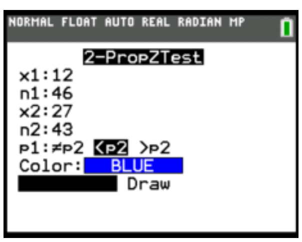
≈ -3.49

Since $z = -3.49 < C.V. = -1.64$.

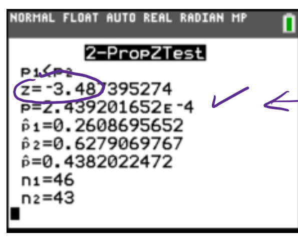
H_1 rejects H_0.

Thus, there is sufficient enough evidence to support the claim that people with money of quarters are more likely to spend than people with money of \$1 bills.

TI-84: Stat → TESTS → 6: 2-PropZTest ...



→



✓ ← $p = 0.0002 < \alpha = 0.05$

Eg Chantix is a drug used as an aid to stop smoking, but giving side effects of insomnia. The numbers of subjects experiencing insomnia for each of two treatment groups in a trial of the drug is given:

	Chantix Treatment	Placebo
Number in group	129	805
Number experiencing insomnia	19	13

Assume α = 0.05. Test the claim that $p_1 = p_2$. Find T.S., C.V., p-value, E and a 95% Confidence Interval.

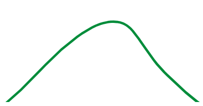
oppose to = is ≠

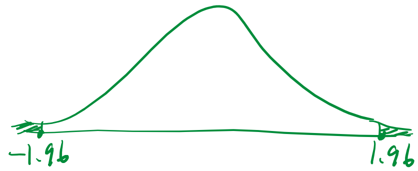
S:

$H_0: p_1 = p_2$

$H_1: p_1 \neq p_2$

$\alpha = 1 - 95\%$, with ≠, then $\frac{\alpha}{2} = \frac{0.05}{2} = 0.025$





then $\frac{\alpha}{2} = \frac{0.05}{2} = 0.025$

T.S.: TI-84: $z = 7.60$, $p = 0$

$$z = 7.60 > C.V. = 1.96 \quad (p = 0 < \alpha = 0.025)$$

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

$$= 1.96\sqrt{\frac{\frac{19}{129}\cdot\frac{110}{129}}{129} + \frac{\frac{13}{805}\cdot\frac{792}{805}}{805}}$$

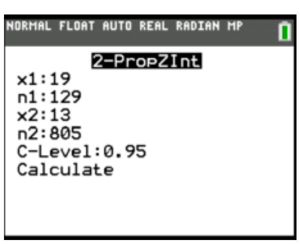
$$\approx 0.06$$

and $\hat{p}_1 - \hat{p}_2 = \frac{19}{129} - \frac{13}{805} \approx 0.131$

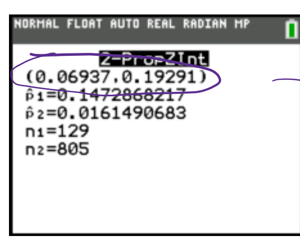
$$0.131 - 0.06 < p_1 - p_2 < 0.131 + 0.06$$

$$\boxed{0.07 < p_1 - p_2 < 0.19}$$

TI-84: Stat → TESTS → B: 2-PropZInt...



→



$$\boxed{0.07 < p_1 - p_2 < 0.19}$$

Eg A poll asked the subjects "Is there solid evidence that the earth is getting warmer?" 69% of $\overset{n_1}{731}$ male $\overset{x_1}{\text{answered}}$ "yes," and 70% of $\overset{n_2\ x_2}{770}$ female said "yes."

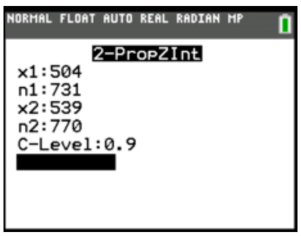
Eg A poll asked the subjects "Is there solid evidence that the earth is getting warmer?" 69% of 731 male answered "yes," and 70% of 770 female said "yes." Construct a 90% confidence interval to estimate the of the difference between the proportion of "yes" responses from males and females.

S: We do not have x_1 and x_2 for TI-84 ← TI-84 takes whole numbers only for x.

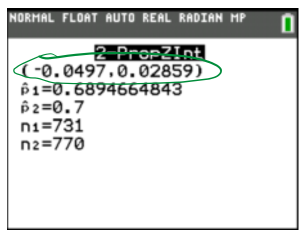
$$x_1 = 69\% \cdot 731 = 504.39 \approx 504$$

$$x_2 = 70\% \cdot 770 = 539$$

90% C.I.:



→



$$\boxed{-0.05 < p_1 - p_2 < 0.03}$$

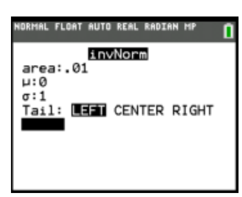
The confidence interval contains 0. Therefore, this is no difference at all.

Eg A poll asked the subjects "Is there solid evidence that the earth is getting warmer?" 69% of 731 male answered "yes," and 70% of 770 female said "yes." Now with a 0.01 significance level test the claim the percentage of male who answer yes is less than the percentage of female who answered yes.

S: $H_0: p_1 = p_2$

$H_1: p_1 < p_2$

$\alpha = 0.01$ with $<$:

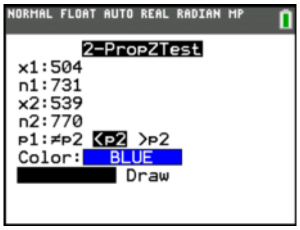


from above ↓

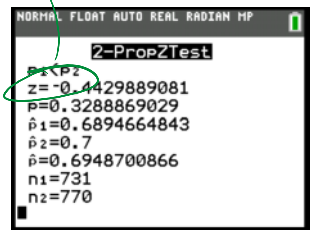
$x_1 = 504$, $n_1 = 731$

$x_1 = 504, \quad n_1 = 731$

$x_2 = 539, \quad n_2 = 770$



$\rightarrow$



since $z = -0.44 > \text{c.v.} = -2.33$ $\quad (p = 0.33 > \alpha = 0.01)$

H_1 fails to reject H_0.

Thus, there is not enough evidence to support the claim that male who answered "Yes" is less that female who answered "Yes".