9.4 Comparing Proportions

OBJECTIVE: Test the Difference Between Two Proportions. Find the confidence intervals for $p_1 - p_2$.

Recall from Chapter 7 that the symbol $\widehat{p}$ is the sample proportion used to estimate the population proportion, denoted by p. The formula for the sample proportion $\widehat{p}$ is

$$\widehat{p} = \frac{X}{n}$$

$$\hat{p} \sim N\left(p, \frac{pq}{n}\right)$$

where X = number of units that possess the characteristic of interest
n = sample size

■ Hypothesis test for Two Proportions

Our test statistic uses $\hat{p}_1 - \hat{p}_2$

a) Left-tailed test (or lower tailed test): $\quad \boldsymbol{H_0: p_1 = p_2}$ vs $\boldsymbol{H_1: p_1 < p_2}$
b) Right-tailed test (or upper tailed test): $\quad \boldsymbol{H_0: p_1 = p_2}$ vs $\boldsymbol{H_1: p_1 > p_2}$
c) Two-tailed test (or two-sided test): $\quad \boldsymbol{H_0: p_1 = p_2}$ vs $\boldsymbol{H_1: p_1 \neq p_2}$

Test Value (or Test Statistic)

$$\frac{\bar{X} - \mu}{\sigma}$$

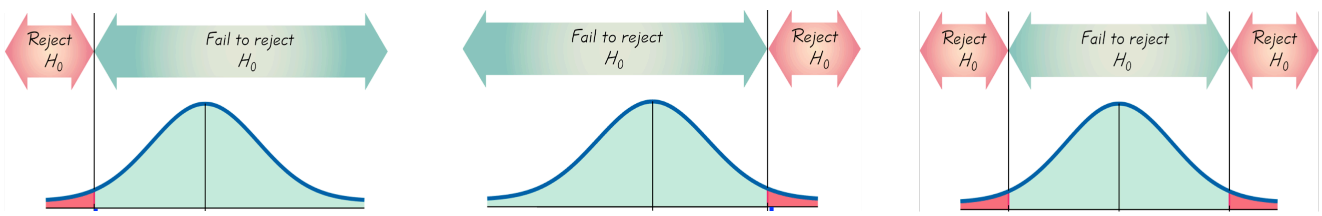
$$\hat{p}_1 - \hat{p}_2 \sim N\left(p_1 - p_2, \left[\text{dependent on pooling}\right]\right)$$

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\bar{p}\,\bar{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

where $p_1 - p_2 = 0$ (assumed in the null hypothesis).

$\hat{p}_1 = \frac{X_1}{n_1}$, $\hat{p}_2 = \frac{X_2}{n_2}$, $\bar{p} = \frac{X_1 + X_2}{n_1 + n_2}$ (**pooled** sample proportion), $\bar{q} = 1 - \bar{p}$

Note: Rejection regions/p-values are the same as before.



<u>*P*-value method:</u>

If *P*-value ≤ α, reject H_0.
If *P*-value > α, fail to reject H_0.

Std Deviation of $\hat{p}_1 - \hat{p}_2$

$$Var(\hat{p}_1 - \hat{p}_2) = Var(\hat{p}_1) + Var(\hat{p}_2)$$

$$= \frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}$$

$$= \bar{p}\bar{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)$$

assuming

H_0 $p_1 = p_2$

$= \bar{p}$ ← contingent on H_0

$$Z^* = \frac{\hat{p}_1 - \hat{p}_2 - E(\hat{p}_1 - \hat{p}_2)}{\sigma(\hat{p}_1 - \hat{p}_2)}$$

$$= \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\bar{p}\bar{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

by H_0,

$p_1 - p_2 = 0$

$\bar{p}$ ← pooled proportion

$$\boxed{\bar{p} = \frac{x_1 + x_2}{n_1 + n_2}}$$

$$\sigma(\hat{p}_1 - \hat{p}_2) = \sqrt{\bar{p}\bar{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

■ **Confidence Interval Estimate of $p_1 - p_2$**

$$(\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2}\underbrace{\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}}_{\mathrm{Var}(\hat{p}_1 - \hat{p}_2)}$$

The margin of error E is given by $E = z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$

Example 1: In the nursing home study mentioned in the chapter-opening Statistics Today, the researchers found that 12 out of 34 randomly selected small nursing homes had a resident vaccination rate of less than 80%, while 17 out of 24 randomly selected large nursing homes had a vaccination rate of less than 80%. At $\alpha = 0.05$, test the claim that there is no difference in the proportions ~~of the small and large nursing homes with~~ a resident vaccination rate of less than 80%.

$H_0: p_1 = p_2 \quad H_a: p_1 \neq p_2$

$\hat{p}_1 = \frac{12}{34} = 0.35$

$x_1 = 12$

$n_1 = 34$

$\hat{p}_2 = \frac{17}{24} = 0.71$

$x_2 = 17$

$n_2 = 24$

$\alpha = 0.05$

$z^* = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\bar{p}\bar{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$

$= \frac{0.35 - 0.71}{\sqrt{0.25\left(\frac{1}{34} + \frac{1}{24}\right)}} = -2.667$ REJECT

$\pm z_{\frac{\alpha}{2}} = \pm$ invNorm $(0.025, 0, 1) = \pm 1.96$

$P = 2P(Z < -2.7006) = 2$ normalcdf $(-10^{99}, -2.7006, 0, 1) = 0.00767$

$\bar{p} = \frac{29}{58} = 0.5 \quad \bar{q} = 0.5$

Example 2: Find the 95% confidence interval for the difference of proportions for the data in Example 1.

$(\hat{p}_1 - \hat{p}_2) \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$

$(-0.596, -0.1128)$

$\not\ni 0$ $\therefore$ Reject

answer you know how to plug in numbers

Example 3: A survey of 200 randomly selected male and female workers (100 in each group) found that 7% of the male workers said that they worked more than 5 days per week while 11% of the female workers said that they worked more than 5 days per week. At $\alpha = 0.01$, can it be concluded that the percentage of males who work more than 5 days per week is less than the percentage of female workers who work more than 5 days per week?

Male:

$n_1 = 100$

$X_1 = 7$

$\hat{p}_1 = 0.07$

Female:

$n_2 = 100$

$x_2 = 11$

$\hat{p}_2 = 0.11$

$\alpha = 0.01$, $H_0: p_1 = p_2$, $H_a: p_1 < p_2$

$Z_{0.01}$ = invNorm(0.01, 0, 1, LEFT) $= -2.32$

2-Prop Z Test:

$Z^* = -0.988$

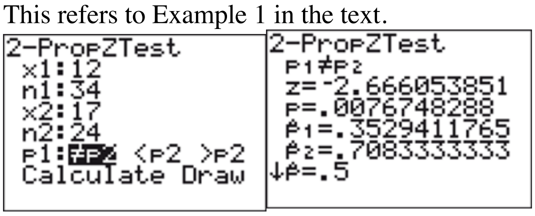
$p = 0.16149 > 0.01 \Rightarrow$ Fail to reject

CI: $(-0.144, 0.06355) \ni 0$

TI-84 Plus Step by Step

■ Hypothesis Test for the Difference Between Two Proportions

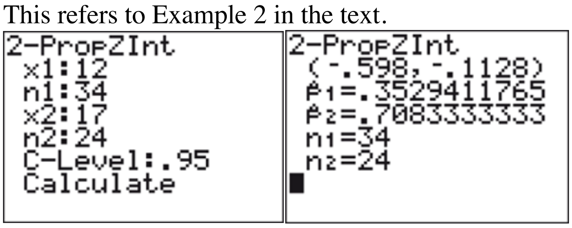
This refers to Example 1 in the text.



1. Press **STAT** and move the cursor to TESTS.
2. Press **6** for 2-PropZTEST.
3. Type in the appropriate values.
4. Move the cursor to the appropriate alternative hypothesis and press **ENTER**.
5. Move the cursor to Calculate and press **ENTER**.

■ Confidence Interval for the Difference Between Two Proportions

This refers to Example 2 in the text.



1. Press **STAT** and move the cursor to TESTS.
2. Press **B (ALPHA APPS)** for 2-PropZInt.
3. Type in the appropriate values.
4. Move the cursor to Calculate and press **ENTER**.

Extra Problems

1. In a sample of 50 men, 44 said that they had less leisure time today than they had 10 years ago. In a random sample of 50 women, 48 women said that they had less leisure time than they had 10 years ago. At $\alpha = 0.10$, is there a difference in the proportions? Find the 90% confidence interval for the difference of the two proportions. Does the confidence interval contain 0?

2. In a random sample of 200 men, 130 said they used seat belts. In a random sample of 300 women, 63 said they used seat belts. Test the claim that men are more safety-conscious than women, at $\alpha = 0.01$.

3. The percentage of males 18 years and older who have never married is 30.4. For females the percentage is 23.6. Looking at the records in a particular populous county, a random sample of 250 men showed that 78 had never married and 58 of 200 women had never married. At the 0.05 level of significance, is the proportion of men greater than the proportion of women?

4. Find the 95% confidence interval for the difference of proportions for the data in Problem 3.