M358K: November 9th, 2020.

Statistical Inference for Two Proportions [Practice].

p_i, $i=1,2$... the population proportion for (sub)population i

$$H_0: p_1 = p_2 \quad \text{vs.} \quad H_a: \begin{cases} p_1 < p_2 \\ p_1 \neq p_2 \\ p_1 > p_2 \end{cases}$$

Two independent samples from the two (sub)populations of sizes n_1 and n_2, resp.

For sample i, $i=1,2$: x_i... the count of successes

$\Rightarrow$ the sample proportion: $\hat{p}_i = \frac{x_i}{n_i}$

Under the null hypothesis, we POOL the two samples to provide an overall point estimate for the population proportion of the whole population:

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

Under the null hypothesis, the z-statistic for our observed difference in sample proportions:

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

To calculate the p-value, we just behave analogously to the previous situations.

UNIVERSITY OF TEXAS AT AUSTIN

Problem Set # 14

Difference in two proportions.

Problem 14.1. A simple random sample of 200 students is selected from a large university. In this sample, there are 35 minority students. A simple random sample of 80 students is selected from the community college in the same town. In this sample, there are 28 minority students. What is the standard error of the difference in sample proportions of minority students?

Problem 14.2. Suppose that, in our usual notation, $\hat{p}_1 = 0.5$, $\hat{p}_2 = 0.2$, $n_1 = 20$ and $n_2 = 30$. What is the $p-$value for testing

$$H_0 : p_1 = p_2 \quad vs. \quad \boxed{H_a : p_1 \neq p_2.}$$

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{\hat{p}_1 \cdot n_1 + \hat{p}_2 \cdot n_2}{n_1 + n_2} = \frac{0.5(20) + 0.2(30)}{20 + 30}$$

$$\hat{p} = \frac{10 + 6}{50} = 0.32$$

z-statistic: $z = \dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \dfrac{0.5 - 0.2}{\sqrt{0.32(0.68)\left(\frac{1}{20} + \frac{1}{30}\right)}} \cong 2.23$

p-value: $\mathbb{P}[Z \geq 2.23] + \mathbb{P}[Z \leq -2.23] =$

$= 2 \cdot \mathbb{P}[Z \leq -2.23] =$

$= 2(0.01287) = 0.0257 4$ ■

#1: p_1 $\hat{p}_1 = 45/60 = 0.75$

Problem 14.3. A simple random sample of 60 households in Whoville is taken. In the sample, there are 45 households that decorate their houses with lights for the holidays.

#2: p_2 $\hat{p}_2 = 40/50 = 0.80$

A simple random sample of 50 households is also taken from the neighboring Whoburgh. In the sample, there are 40 households that decorate their houses.

(i) What is a 95% confidence interval for the difference in population proportions of households that decorate their houses with lights for the holidays?

$p_1 - p_2$

$$p_1 - p_2 = \boxed{\text{pt. estimate}} \pm z^* (\text{std. error})$$

$$\text{pt. estimate} = \hat{p}_1 - \hat{p}_2 = 0.75 - 0.80 = \boxed{-0.05}$$

$$z^* = 1.96$$

$$\text{std. error} = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} = \sqrt{\frac{0.75(0.25)}{60} + \frac{0.8(0.2)}{50}} = 0.07953$$

$$\boxed{p_1 - p_2 = -0.05 \pm 1.96(0.07953) = -0.05 \pm 0.1558}$$

(ii) If you want to test the hypothesis whether one of the two cities has more festive inhabitants, i.e., whether one of the two cities has a higher proportion of decorated domiciles or not, what p–value would you obtain?

$$H_0: p_1 = p_2 \quad \text{vs.} \quad H_a: p_1 \neq p_2$$

The POOLED observed sample proportion:

$$\hat{p} = \frac{45+40}{60+50} \doteq 0.77$$

The observed value of the z statistic:

$$z = \frac{-0.05}{\sqrt{0.77(0.23)\left(\frac{1}{60} + \frac{1}{50}\right)}} = -0.62$$

The p-value:

$$\mathbb{P}[Z \leq -0.62] + \mathbb{P}[Z \geq 0.62] = $$
$$= 2 \cdot \mathbb{P}[Z \leq -0.62] = 2(0.2676)$$
$$= 0.5352$$