Statistical Inference for Two Proportions.

Our parameters of interest:

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

e.g., p_1 ... corresponds to the (sub)population who get the treatment;

p_2 ... corresponds to the (sub)population who get the placebo.

Sample sizes are denoted by n_1 and n_2.

Assume that the two samples are independent.

For large n_i, $i=1,2$, we know the approximate dist'n of:

- the count r.v.s:

$$X_i \approx Normal\left(mean = n_i \cdot p_i,\ sd = \sqrt{n_i \cdot p_i(1-p_i)}\right),\ i=1,2$$

and

- the sample proportion r.v.s:

$$\boxed{\hat{P}_i = \frac{X_i}{n_i} \approx Normal\left(mean = p_i,\ sd = \sqrt{\frac{p_i(1-p_i)}{n_i}}\right),\quad i=1,2}$$

We base our statistical inference for $p_1 - p_2$ on $\hat{P}_1 - \hat{P}_2$.

$$\hat{P}_1 - \hat{P}_2 \approx Normal\left(mean = p_1 - p_2,\ sd = ?\right)$$

$$Var[\hat{P}_1 - \hat{P}_2] = Var[\hat{P}_1] + Var[\hat{P}_2] \quad \text{(independence)}$$

$$= \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}$$

$$? = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

Confidence Intervals.

C... confidence level

pt. estimate $\pm$ margin of error

$z^* \cdot \boxed{\text{stderror}}$ $\quad z^* = \text{qnorm}\left(\frac{1+C}{2}\right) = \Phi^{-1}\left(\frac{1+C}{2}\right)$

$$\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

$$\boxed{\hat{p}_1 - \hat{p}_2} \pm z^* \cdot \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

w/ $\hat{p}_1$, $\hat{p}_2$ the observed sample proportions

$\hat{p}_1 = \frac{45}{60} = 0.75$

Problem 17.3. A simple random sample of 60 households in Whoville (p_1) is taken. In the sample, there are 45 households that decorate their houses with lights for the holidays.

A simple random sample of 50 households is also taken from the neighboring Whoburgh (p_2). In the sample, there are 40 households that decorate their houses.

$\hat{p}_2 = \frac{40}{50} = 0.80$

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

$p_1 - p_2$

→ :

pt. estimate ± margin of error

$$z^* \cdot se$$

$z^* = 1.96$

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

$$0.75 - 0.80$$

$$\boxed{-0.05}$$

$$se = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

$$\underbrace{\sqrt{\frac{0.75(0.25)}{60} + \frac{0.8(0.2)}{50}}}_{0.0795}$$

$$p_1 - p_2 = -0.05 \pm 1.96\,(0.0795) = \boxed{-0.05 \pm 0.1558}$$

Hypothesis Testing.

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

Test Statistic: $\hat{P}_1 - \hat{P}_2$

Under the null hypothesis: $\boxed{p_1 = p_2 = p}$

$$\boxed{\hat{P}_1 - \hat{P}_2 \approx \text{Normal}\left(\text{mean} = 0,\ sd = \sqrt{\frac{p(1-p)}{n_1} + \frac{p(1-p)}{n_2}} = \sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}\right)}$$

$$\boxed{\frac{\hat{P}_1 - \hat{P}_2}{\sqrt{\boxed{p(1-p)}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}} \approx N(0,1) \quad \text{under the null hypothesis}$$

Let $\hat{p}_1$ and $\hat{p}_2$ be the observed sample proportions.

Q: What's the form of the observed z-statistic?

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

w/ the estimate $\hat{p}$ based on all the available data