STATS 260 Class 22

Gavin Jaeger-Freeborn

1. Set 30 Comparing Two Population Proportions

In this section we will consider scenarios where we take samples of 2 independent scenarios. Here we compare the two population proportions p_1 and p_2.

To compair them we use $p_1 - p_2$

$p_1 - p_2 \neq 0$	different
$p_1 - p_2 > 0$	larger
$p_1 - p_2 < 0$	smaller
$p_1 - p_2 = 0.1$	requires a reason to test this

To estimate p_1 and p_2 we use $\hat{p}_1$ and $\hat{p}_2$

Where $\hat{p}_1$ and $\hat{p}_2$ are 2 **sample proportions**.

$$\hat{p}_n = \frac{x_n}{\hat{n}_n}$$

2. Confidence Interval

$$\text{CI: estimated} \pm (\text{c. v})(\text{ese})$$

2.1. Option 1 for if $p_1 - p_2 \neq 0$ AKA unpooled

$$\hat{p}_1 - \hat{p}_2 \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

2.2. Option 2 for if $p_1 - p_2 = 0$
AKA pooled

$$\hat{p} \pm z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$$

3. Test Statistic

$$\text{test statistic} = \frac{\text{estimate} - \text{parameter value}}{\text{ese}}$$

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} \sim N(0, 1)$$

Here the **estimated standard error** is since we dont have $\hat{p}_1$ or $\hat{p}_2$

$$\sqrt{\frac{\hat{p}_1(1-ph1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

3.1. Option 1 for if $p_1 - p_2 \neq 0$
AKA unpooled

$$\therefore\ Z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{\hat{p}_1(1-ph1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}} \sim N(0, 1)$$

3.2. Option 2 for if $p_1 - p_2 = 0$
AKA pooled

We assume that $p_1 = p_2 = p$ by combining them to form a single random sample.

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

$$\therefore \text{ese} = \sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

$$\hat{p} \equiv \text{pooled sample}$$

NOTE: $V(\hat{p}_1 - \hat{p}_2) = V(\hat{p}_1) + V(\hat{p}_2)$

NOTE: $\hat{p}_{1or2} \approx N\left(p_n, \sqrt{\frac{p_{1or2}(1-p_{1or2})}{n_1}}\right)$

Example

Motherboards are made by one of two manufacturing processes. 300 motherboards made by the first process and 500 mother boards made by the second process are sampled at random. From the first process, 15 have flaws. From those made by the second process, 30 have flaws. Let p_1 , p_2 denote the proportion of motherboards made by process one, two (respectively) which are defective.

(a) What is the estimate for $p_1 - p_2$?

(b) What is the unpooled estimated standard error of $\hat{p}_1 - \hat{p}_2$?

(c) Test the research hypothesis that the first process makes a smaller proportion of defective items than the second process, using the e.s.e. from part (b).

(d) Test the same hypotheses in (c), this time using the pooled estimated standard error.

(e) Create a 93% confidence interval for $p_1 - p_2$.

(f) What does the confidence interval tell you about $p_1 - p_2$?

(g) Suppose we wish to use these data as a pilot study to estimate the sample size we would need in the future to create a 95% confidence interval with a margin of error of 0.01. What sample size is needed
(assuming that $n_1 = n_2$).