Set 30: Inference for two Binomial samples

Example: Motherboards are made by one of two manufacturing processes. 300 motherboards made by the first process and 500 motherboards 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) Before we collected the data, our belief was that the first process makes fewer defective items than the second process.

1. Define the parameters of interest: p_1, p_2 **defined above.**

2. Define null and alternative hypotheses:

$H_0 : p_1 - p_2 = 0$

$H_1 : p_1 - p_2 < 0$

3. Define the test statistic and distribution:

$$Z = \frac{(\hat{p_1} - \hat{p_2}) - (p_1 - p_2)}{\sqrt{\frac{\tilde{p}(1-\tilde{p})}{m} + \frac{\tilde{p}(1-\tilde{p})}{n}}} \approx Normal(0, 1)$$

where $\tilde{p} = \dfrac{15 + 30}{300 + 500}$, **the estimate under** H_0.

4. Find the observed value of the test statistic:

From our data, we find $\hat{p_1} - \hat{p_1} = 15/300 - 30/500 = -0.01$, **and**

Our test statistic has value:

$$Z_{obs} = \frac{-0.01 - 0}{\sqrt{\frac{(45/800)(755/800)}{300} + \frac{(45/800)(755/800)}{500}}} = \frac{-0.01 - 0}{.0168} = -0.594$$

5. and 6. Find the p-value, and draw conclusions:

The p-value is the probability of observing a test statistic value as extreme as the one we've observed in the direction of H_1:

$$p - value = P(Z \leq -0.594) = 0.276$$

There is no evidence against H_0.

(b) **Compute a 93% confidence interval for $p_1 - p_2$.**

Estimate: $\hat{p_1} - \hat{p_2} = -0.01$

Critical value: Using Z, we find 1.81.

Estimated Standard Error:

$$\sqrt{\frac{15/300(285/300)}{300} + \frac{30/500(470/500)}{500}} \approx 0.016466$$

$$-0.01 \pm (1.81)(0.016466) \approx -0.01 \pm 0.0298$$

The interval is $(-0.0398, 0.0198)$.