Example 1:

A random sample of 500 people in a certain region of a country, which is about to have a national election, were asked whether they preferred “Candidate A” or “Candidate B”.

From this sample, 320 people responded that they preferred Candidate A.

Example 1, continued:

An independent random sample of 420 people in a second region of that country were asked whether they preferred "Candidate A" or "Candidate B".

From this sample, 302 people responded that they preferred Candidate A.

Construct an approximate 90% confidence interval for the difference between the true proportions for each country.

Let p_1 and p_2 be the true proportions for the first and second regions.

We have

$$\widehat{p}_1 = \frac{320}{500} = \frac{16}{25} \quad \text{and} \quad \widehat{p}_2 = \frac{302}{420} = \frac{151}{210}$$

$$(n_1 = 500,\ n_2 = 420)$$

Both of these satisfy

$$\left(\widehat{p} - 3\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}},\ \widehat{p} + 3\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} \right) \subseteq [0,1]$$

Step One:

An estimator of $p_1 - p_2$ is

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

Step Two: By the CLT, we have

$$\widehat{p}_1 \overset{\text{approx}}{\sim} N\left(p_1, p_1(1-p_1)/n_1\right)$$

$$\widehat{p}_2 \overset{\text{approx}}{\sim} N\left(p_2, p_2(1-p_2)/n_2\right)$$

Find a function of the estimators and the “target” whose distribution is known and “unknown parameter free”.

For simplicity and due to the fact that we have large sample sizes, we replace each

$$\sigma_{\widehat{p}_i} = \frac{p_i(1 - p_i)}{n_i}$$

with the estimator

$$\widehat{\sigma_{\widehat{p}_i}} = \frac{\widehat{p}_i(1 - \widehat{p}_i)}{n_i}$$

We have

$$\hat{p}_1 - \hat{p}_2 \overset{\text{approx}}{\sim} N\left(p_1 - p_2, \frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}\right)$$

Standardize:

$$\frac{p_1 - p_2 - (\hat{p}_1 - \hat{p}_2)}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}} \overset{\text{approx}}{\sim} N(0,1)$$

Step Three:

Put this between appropriate critical values and solve for $p_1 - p_2$ "in the middle.

$$z_{\alpha/2} = z_{0.10/2} = z_{0.05} = 1.645$$

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

Answer: $(-0.12953, -0.02857)$

Answer: $(-0.12953, -0.02857)$

The data suggests that the proportion of people from region 2 that voted for candidate A is larger than the proportion from region 1.

$\widehat{p}_1 = 0.64$ and $\widehat{p}_2 \approx 0.72$

Example 2:

A potato chip manufacturer sells 10 ounce bags of potato chips. The company always overfills the bags slightly so as not to have angry customers.

In addition to overfilling the bags, the manufacturer wants to make sure that the standard deviation of weights is small, so that, even bags on the low fill end will contain at least the amount of product advertised.

For a random sample of 20 bags of chips, the quality control manager finds that the sample variance is 0.51 ounces.

Assuming that fill weights are normally distributed, find a 95% confidence interval for σ, the true standard deviation for all bags.

Step One:

Decide on an estimator.

We will use the sample standard deviation $S = \sqrt{S^2}$.

Step Two:

Look at the distribution of S.

Recall that, for a normal sample,

$$\Rightarrow \quad \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

Step Three:

Put the statistic between appropriate critical values and solve for σ in the middle.

Here are 3 possibilities. (There are more!)

$$\chi^2_{0.975,n-1} < \frac{(n-1)S^2}{\sigma^2} < \chi^2_{0.025,n-1}$$

$$0 < \frac{(n-1)S^2}{\sigma^2} < \chi^2_{0.05,n-1}$$

$$\chi^2_{0.95,n-1} < \frac{(n-1)S^2}{\sigma^2} < \infty$$

$$8.90652 < \frac{(20-1)(0.51)}{\sigma^2} < 32.85233$$

$$\Downarrow$$

$$0.91915 < \frac{1}{\sigma^2} < 3.39033$$

$$\Downarrow$$

$$0.29496 < \sigma^2 < 1.08797$$

$$\Downarrow$$

$$\boxed{0.54310 < \sigma < 1.04306}$$

Alternatively,

$$0 < \frac{(20-1)(0.51)}{\sigma^2} < 30.14353$$

gives

$$\boxed{0.56698 < \sigma < \infty}$$

and

$$10.11701 < \frac{(19-1)(0.51)}{\sigma^2} < \infty$$

gives

$$\boxed{0 < \sigma < 0.97867}$$