

## Estimating The Difference of Two Population Proportions

In a recent survey on academic dishonesty simple random samples of 200 female and 100 male college students were taken. 24 of females and 26 of the males agreed or strongly agreed with the statement “Under some circumstances academic dishonesty is justified.” Researchers would like to compare the population proportions,  $p_f$  of all female and  $p_m$  of all male college students who agree or strongly agree with this statement.

We know that a point estimate of  $p_f$  is the sample proportion  $\hat{p}_f = 24/200 = 0.12$ , and a point estimate of  $p_m$  is the sample proportion  $\hat{p}_m = 26/100 = 0.26$ . Therefore it makes sense to estimate the difference  $p_f - p_m$  with the difference in estimates  $\hat{p}_f - \hat{p}_m = 0.12 - 0.26 = -0.14$ .

We will compute a 95% approximate score confidence interval for the difference in the proportions  $p_f$  of all female and  $p_m$  of all male college students who agree or strongly agree with the statement.

Since  $z_{0.975} = 1.96$ , the adjusted estimates of  $n_f$  and  $n_m$  are

$$\tilde{n}_f = 200 + 0.5 \cdot 1.96^2 = 201.9208, \quad \tilde{n}_m = 100 + 0.5 \cdot 1.96^2 = 101.9208$$

The adjusted estimates of  $p_f$  and  $p_m$  are then

$$\tilde{p}_f = \frac{24 + 0.25 \cdot 1.96^2}{\tilde{n}_f} = 0.1236,$$

and

$$\tilde{p}_m = \frac{26 + 0.25 \cdot 1.96^2}{\tilde{n}_m} = 0.2645.$$

The approximate score interval for  $p_f - p_m$  is then

$$\begin{aligned} & 0.1236 - 0.2645 \pm \\ & 1.96 \sqrt{\frac{0.1236(1 - 0.1236)}{201.9208} + \frac{0.2645(1 - 0.2645)}{101.9208}} \\ & = (-0.2378, -0.0440) \end{aligned}$$

Based on this interval, we conclude that there is a difference in the true population proportions. With 95% confidence, we estimate that the proportion of males who agree with the statement is between 4.4% and 23.78% higher than the proportion of females.