8.4: Comparing Two Independent Population Proportions

Comparing two proportions, like comparing two means, is common. If two estimated proportions are different, it may be due to a difference in the populations or it may be due to chance in the sampling. A hypothesis test can help determine if a difference in the estimated proportions reflects a difference in the two population proportions.

Like the case of differences in sample means, we construct a sampling distribution for differences in sample proportions: $\left(P_A^{\prime}-P_B^{\prime}\right)$ where $P_A^{\prime}=\frac{x_A}{n_A}$ and $P_B^{\prime}=\frac{x_B}{n_B}$ are the sample proportions for the two sets of data in question. X_A and X_B are the number of observations in each sample group of interest, respectively, and n_A and n_B are the respective sample sizes from the two groups. Again we go the Central Limit theorem to find the distribution of this sampling distribution for the differences in sample proportions. And again we find that this sampling distribution, like the ones past, are normally distributed as proved by the Central Limit Theorem.

Figure 8.4.1

Generally, the null hypothesis allows for the test of a difference of a particular value, δ_0, just as we did for the case of differences in means.

$$H_0 : P_A-P_B=\delta_0$$

$$H_1 : P_A-P_B \neq \delta_0$$

$$\textbf{The confidence interval formula is:}$$

$$\left(P_2^{\prime}-P_1^{\prime}\right) \pm z_{\frac{\alpha}{2}} * \sqrt{\frac{P_1^{\prime} *\left(1-P_1^{\prime}\right)}{n_1}+\frac{P_2^{\prime} *\left(1-P_2^{\prime}\right)}{n_2}}$$

Example 8.4.1

A bank has recently acquired a new branch and thus has customers in this new territory. They are interested in the default rate in their new territory. They wish to test the hypothesis that the default rate is different from their current customer base. They sample 200 files in area A, their current customers, and find that 20 have defaulted. In area B, the new customers, another sample of 200 files shows 12 have defaulted on their loans. At a 1% level of significance can we say that the default rates are the same or different?

Answer

This is a test of proportions. We know this because the underlying random variable is binary, default or not default. Further, we know it is a test of differences in proportions because we have two sample groups, the current customer base and the newly acquired customer base. Let A and B be the subscripts for the two customer groups. Then $p_A$and p_B are the two population proportions we wish to test.

Random Variable:

$P_A^{\prime}-P_B^{\prime}$ = difference in the proportions of customers who defaulted in the two groups.

$H_0 : P_A=P_B$

$H_a : P_A \neq P_B$

The words "is different" tell you the test is two-tailed.

Distribution for the test:

Estimated proportion for group A: $P_A^{\prime}=\frac{x_A}{n_A}=\frac{20}{200}=0.1$

Estimated proportion for group B: $P_B^{\prime}=\frac{x_B}{n_B}=\frac{12}{200}=0.06$

The estimated confidence interval for the difference between the two groups is:

$$(P_2' - P_1') \pm z_{\frac{\alpha}{2}} * \sqrt{\frac{P_1' * (1 - P_1')}{n_1} + \frac{P_2' * (1 - P_2')}{n_2}} = [-.03, .11]$$

Make a decision: Since the calculated confidence interval contains 0, we cannot reject H_0.

Conclusion: At a 1% level of significance, from the sample data, there is not sufficient evidence to conclude that there is a difference between the proportions of customers who defaulted in the two groups.

Exercise 8.4.1

Two types of valves are being tested to determine if there is a difference in pressure tolerances. Fifteen out of a random sample of 100 of Valve *A* cracked under 4,500 psi. Six out of a random sample of 100 of Valve *B* cracked under 4,500 psi. Test at a 5% level of significance.