15. Hypothesis Tests for Proportions

Conceptually, these are exactly the same as hypothesis test for means. The only differences are that we find sample proprotions $\hat{p}_E$ and $\hat{p}_C$, and then the spreadsheet calculates the test statistic:

$$\text{test statistic} = \frac{\hat{p}_E - \hat{p}_C}{\sqrt{\frac{\hat{p}_E(1-\hat{p}_E)}{n_E} + \frac{\hat{p}_C(1-\hat{p}_C)}{n_C}}}$$

and we test

$$H_0 : \; p_E = p_C \quad \textit{against} \quad H_A : \begin{cases} p_E > p_C & \textit{or} \\ p_E < p_C & \textit{or} \\ p_E \neq p_C \end{cases}$$

15.1. Example.

A researcher surveyed 1,600 randomly chosen females, aged 40-60. In order to participate, the subjects must either be currently married or divorced. The researcher gathered data about whether or not the subjects had cohabited prior to thier first marriage and whether or not that marraige ended in divorce. There were 732 who cohabited prior to marraige, and 345 of this group were divorced. There were 868 who did not cohabit prior to marraige, and 348 of this group were divorced.

Using a signficance level of 5%, does this provide significant evidence that cohabitation prior to marraige is associated with a higher divorce rate?

Solution.

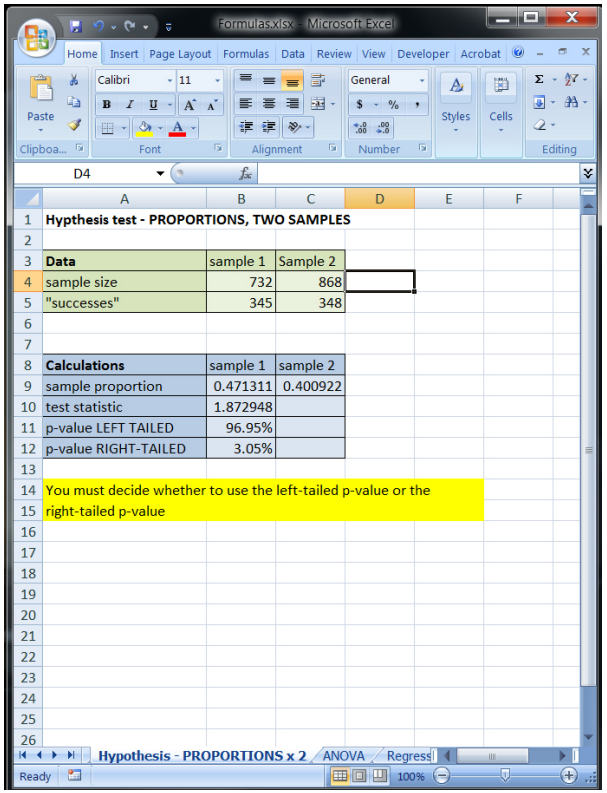
Step 1. First make a dictionary of the information given in the problem; this problem focuses on divorce, so "divorce" constitutes "success."

	Cohabited	Did not Cohabit
n	732	868
Divorced	345	348

Step 2. The treatment in this case is cohabitation, and the experimental outcome is divorce. Since the problem asks whether cohabitation increases the chances of divorce, our hypotheses are:

$$H_0 : p_E = p_C$$
$$H_A : p_E > p_C$$

Step 3. Now enter the list into the spreadsheet FORMULAS.XLS, found in the resources section for this course on LEARN.OU.EDU. Note that you will need to select the tab at the bottom labeled `hypothesis-PROPORTIONS x 2`. You should use `Sample 1` to record the experimental data and `Sample 2` to record the control data.



Step 4. From the alternative hypothesis, this is a right-tailed test. Since the p-value is less than the target significance value, we *reject the null hypothesis* and *accept the alternative.* This means that we believe that cohabitation is associated with higher divorce rates. The

probability that we have made a Type I Error is the p-value, 3.05%.

Solution Template

Step 1. Make a dictionary assigning values to each of the variables:

	Experimental	Control
sample size	n_E	n_C
successes	k_E	k_C
significance level	α	

In order to use the spreadsheet, we must have both

$$np_0 \geq 5$$

and

$$n(1 - p_0) \geq 5$$

This requirement will always be fulfilled in problems and examination questions in this class. There are other techniques one can use (χ-squared tests) when this requirement is violated.

Step 2. Write down the null and alternative hypotheses. The null hypothesis will always be:

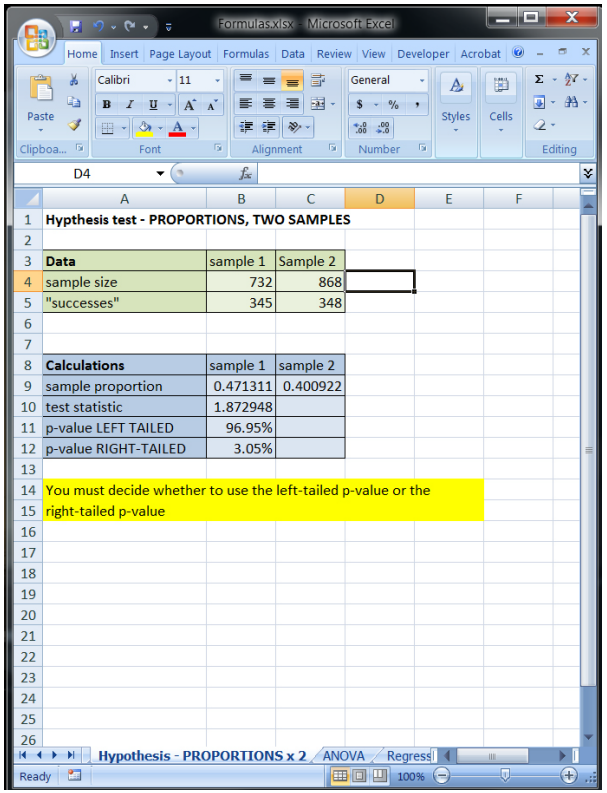
$$H_0 : \quad p_E = p_C$$

while the alternative hypothesis will be one of the following:

$$H_A : \ p_E < p_C \quad \textit{(a left tailed test)}$$
$$H_A : \ p_E > p_C \quad \textit{(a right tailed test)}$$
$$H_A : \ p_E \neq p_C \quad \textit{(a two tailed test)}$$

(The reason for the terms right, left and two tailed tests is the same as in hypothesis testing for means.)

Step 3. Now enter the list into the spreadsheet FORMULAS.XLS, found in the resources section for this course on LEARN.OU.EDU. Note that you will need to select the tab at the bottom labeled `hypothesis-PROPORTIONS x 2`. You should use `Sample 1` to record the experimental data and `Sample 2` to record the control data.



Step 4. Select the appropriate p-value from the spreadsheet using the direction of the inequality in the alternative hypothesis. If the p-value is less than the pre-set significance level, then you *reject the null hypothesis* and *accept the alternative.* Otherwise, you accept the alternative hypothesis.

End of Solution Template

15.2. Example.

A researcher gathers data on 1988 students in a large, urban high school. In this school, 123 students have a history of incarceration in a temporary dentention center, while 1865 have no such history. Among those who have been incarcerated, the researcher determines that 22 have a diagnosis of a personality disorder, while 72 of the non-incarcerated group have a similar diagnosis.

Is this convincing evidence, using a significance level of 5%, that detained youth are at greater risk for a personality disorder than students who do not have a history of incarceration?

Solution.

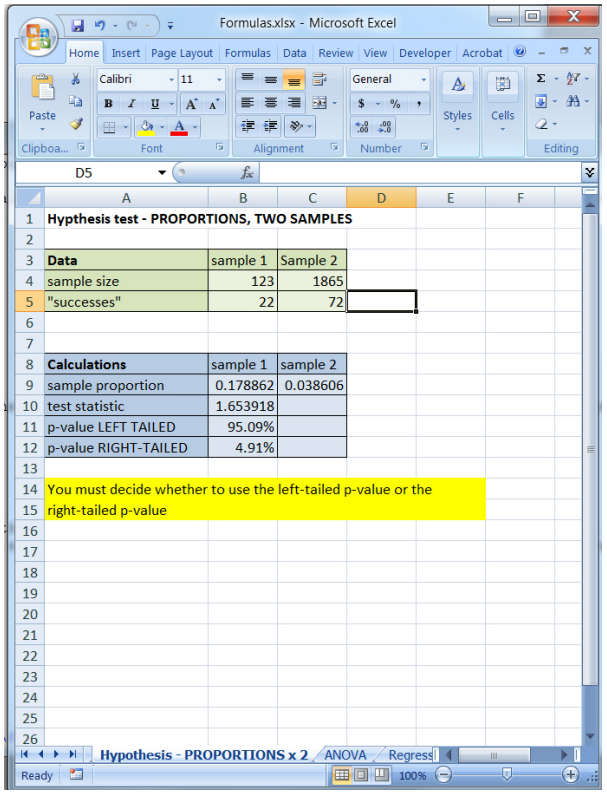
Step 1. First make a list of all the relevant variables.

	Incarcerated	non-Incarcerated
sample size	123	1865
number of “successes”	22	72
α	5%	

Step 2. The treatment in this case is a history of incarceration, and the researcher conjectures that incarcerated youth have a higher incidence of disorder, so:

$$H_0 : p_E = p_C$$
$$H_A : p_E > p_C$$

Step 3. Now enter the list into the spreadsheet FORMULAS.XLS, found in the resources section for this course on LEARN.OU.EDU. Note that you will need to select the tab at the bottom labeled `hypothesis-PROPORTIONS x 2`. You should use `Sample 1` to record the experimental data and `Sample 2` to record the control data.



Step 4. This is a right-tailed test from the direction of the inequality in the alternative hypothesis. Since the p-value of 4.91% is less than the pre-set target of 5%, we reject the null hypothesis and accept the alternative, namely that incarcerated youth are at greater risk for personality

disorders. The chance of a Type I Error is 4.92%.