Lab 8: Inference for a Single Proportion and Comparing Two Proportions

OBJECTIVES: This lab is designed to expand on the ideas of inference for a single proportion and comparing two proportions.

DATA: The data for the first part of this lab are from a simple random survey conducted by ceiling fan manufacturers in the state of Florida. The data for the second part come from a study about age discrimination in job promotions.

DIRECTIONS: Follow the instructions below, answering all questions. Write out the answers neatly and turn them in to the lab instructor.

1. An association of ceiling fan manufacturers sponsored a sample survey of Florida households to help improve the marketing of ceiling fans. A simple random survey of 1000 households was contacted by telephone and asked several questions in a short survey. One question was "Do you own at least one ceiling fan?" Of the 1000 households, 840 answered "Yes".

   a. What is the sample proportion of Florida households with at least one ceiling fan?

   b. What is the 95% confidence interval for the true population proportion of all Florida households that have at least one ceiling fan? The formula for calculating this is:

   $$\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

   where z^* is the upper (1-C)/2 standard normal critical value, $\hat{p}$ is the proportion of all Florida households that have at least one ceiling fan, and n is the number of people surveyed.

   The value for z^* can be found from the following table. C is simply the confidence level we wish to attain.

C	z^*
90%	1.645
95%	1.960
99%	2.576

   c. What is the interpretation of this confidence interval?

2. Of the 1000 households in the survey, 40% were from Northern Florida, and the other 60% were from Southern Florida. According to the census, 42% of Florida households are in Northern Florida and the remaining 58% are in Southern Florida.

   a. To examine how well the sample represents the state in regard to northern versus southern residence, we perform a significance test of

$$H_0 : p = .42$$
$$H_1 : p \neq .42.$$

b. The test statistic in this case is

$$z = \frac{(\hat{p} - p)}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$$

Calculate the value of the test statistic.

c. Now, we want to figure out the probability that a Z is greater than or equal to the value of the test statistic. This can be found from Table A in the book, which is partially listed below.

z	*.06*	*.07*	*.08*	*.09*
-1.4	.0721	.0708	.0694	.0681
-1.3	.0869	.0853	.0838	.0823
-1.2	.1038	.1020	.1003	.0985
-1.1	.1230	.1210	.1190	.1170

d. From this value, we can calculate the p-value via the following formula

$$\text{p-value} = 2 * (\text{Table Value})$$

Based on this p-value, what can we conclude about the survey, assuming that $\alpha = .05$ is our significance level? In other words, does the proportion of survey households in the north of Florida reasonably reflect the census proportion of Floridians who live in the north portion of the state?

3. Next, we wish to examine two proportions. To do this, we will look at a study about age discrimination in job promotions. A federal agency with offices in several regions is suspected of discriminating against older employees in job promotions. A random sample of 100 individuals was selected from employees across the country who were recently considered for promotions. All the individuals in the sample were classified by age (less than 50 years old, or 50 years old or greater) and by whether or not they were promoted. The data are represented in a 2×2 contingency table below.

Age (years)	*Not Promoted*	*Promoted*
Less than 50	10	40
50 or more	35	15

a. The statisticians wish to determine if the agency has discriminated against older employees in job promotions. In order to determine this, they first looked at a 90% confidence interval for the difference between those under 50 and older than 50 in the proportion of individuals that were promoted. The formula for the confidence interval is

$$(\hat{p}_1 - \hat{p}_2) \pm z^* \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

where $\hat{p}_1$ is the proportion of people under 50 who were promoted, $\hat{p}_2$ is the proportion of people 50 and over who were promoted, n_1 is the total number of people under 50, and n_2 is the total number of people 50 or over in this particular federal agency. Calculate this interval.

b. What is the interpretation of this confidence interval?

c. Although we prefer to compare two proportions by giving a confidence interval for the difference between the two proportions, it is sometimes useful to test the null hypothesis that the two population proportions are the same. So, what we wish to test is whether individuals that are 50 or over in this particular federal agency are just as likely as those that are less than 50 to get promoted. Thus, we wish to test the hypotheses

$$\begin{aligned} H_0 : p_1 &= p_2 \\ H_1 : p_1 &\neq p_2. \end{aligned}$$

d. To test hypotheses of this type, a significance test is generally used. In these types of tests, we do not use an estimate of the standard error like that in the other tests you have been looking at. Instead, we assume that the data come from a single population since we hypothesize that the two proportions are equal. This is typically referred to as 'pooling' the estimate because it combines, or pools, the information from both samples. Thus, the test statistic is

$$z = \frac{\hat{p}_1 - \hat{p}_2}{s_p}$$

where s_p denotes the pooled sample standard error. Note that

$$s_p = \sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1} + \frac{1}{n_2})}$$

where

$$\hat{p} = \frac{\text{total number people promoted}}{\text{total number people}}$$

e. Given all of these formulas, calculate the value of the test statistic z.

f. What is the p-value of this test? Remember that the p-value here is twice the probability that Z is greater than or equal to z. The probability can be found using Table A in your book. Note that if the value of z is extermely large, then the probability that Z is less than or equal to z is 1, so that the p-value is 0.

g. What can we conclude given this p-value? Remember that our significance level, α, is .10.