• A software application (app) lets users enter questions to receive answers in the form of images, texts, or videos. Research indicates that 22 percent of high school students in Country W use the app to help them with their homework at least once per week. Karen is an AP Statistics student in Country W at a high school that has more than 2,000 students. She believes the proportion of all students at her school who use the app to help them with their homework at least once per week is greater than the proportion for her country. To investigate her belief, she took a simple random sample of 130 students from her school and found that 38 of the sampled students use the app to help them with their homework at least once per week.

Is there convincing statistical evidence, at a 0.05 significance level, to support Karen's belief? Justify your answer with the appropriate inference procedure.

Solution:

(A) An appropriate inference procedure is a one-sample z-test for a population proportion.
The null hypothesis is $\mathrm{H}_0 : p = 0.22$, and the alternative hypothesis is $\mathrm{H}_\mathrm{a} : p > 0.22$, where $p =$ the true proportion of students at Karen's high school that use the application at least once per week.

(B) The independent observation condition for performing the one-sample z-test for a population proportion is satisfied. This is because the data were obtained from a random sample of 130 high school students from Karen's high school. Also, the sample of 130 students is less than 10% of the total number of students at this large high school, because $130 < 0.10(2,000)$ and the total number of students in Karen's high school is greater than 2,000. The 10% condition is required as sampling was conducted without replacement from a finite population.
The number of expected successes and expected failures were both more than 10 because $130(0.22) = 28.6$ and $130(0.78) = 101.4$. Thus, the sample size is large enough to support the assumption that the sampling distribution of $\hat{p}$ is approximately normal.

$$\hat{p} = \frac{38}{130} \approx 0.2923$$

Test statistic:

$$z = \frac{0.2923 - 0.22}{\sqrt{\frac{0.22(1-0.22)}{130}}} \approx 1.99$$

$P(z > 1.99) \approx 0.023$

(C) Because this p-value is less than the $\alpha = 0.05$ significance level, the null hypothesis should be rejected. There is convincing statistical evidence that the population proportion of students at Karen's high school that use the application at least once per week is greater than Country W's proportion of 0.22.

• A large exercise center has several thousand members from age 18 to 55 years and several thousand members age 56 and older. The manager of the center is considering offering online fitness classes. The manager is investigating whether members' opinions of taking online fitness classes differ by age. The manager selected a random sample of 170 exercise center members ages 18 to 55 years and a second random sample of 230 exercise center members ages 56 years and older. Each sampled member was asked whether they would be interested in taking online fitness classes.

The manager found that 51 of the 170 sampled members ages 18 to 55 years and that 79 of the 230 sampled members ages 56 years and older said they would be interested in taking online fitness classes.

At a significance level of $\alpha = 0.05$, do the data provide convincing statistical evidence of a difference in the proportion of all exercise center members ages 18 to 55 years who would be interested in taking online fitness classes and the proportion of all exercise center members ages 56 years and older who would be interested in taking online fitness classes? Complete the appropriate inference procedure to justify your response.

Solution:

Section 1: Let p_{younger} represent the proportion of all exercise center members from 18 to 55 years of age who would be interested in taking online fitness classes, and p_{older} represent the proportion of all exercise center members 56 years or older who would be interested in taking online fitness classes.

The null hypothesis is $\text{H}_0 : p_{\text{younger}} = p_{\text{older}}$
The alternative hypothesis is $\text{H}_a : p_{\text{younger}} \neq p_{\text{older}}$.

An appropriate inference procedure is a two-sample z-test for a difference of population proportions.

Section 2: The independent observations condition for performing the two-sample z-test for a difference in population proportions is satisfied because the data were obtained from a random sample of 170 exercise center members ages 18 to 55 years and a second random sample of 230 exercise center members ages 56 years and older.

The 10% condition must be met by both samples because sampling of exercise center members is done without replacement. There are more than $10(170) = 1,700$ adults from 18 to 55 years of age who are members of the exercise center and more than $10(230) = 2,300$ adults ages 56 years and older who are members of the exercise center.

The value of the sample proportions are $\hat{p}_{\text{younger}} = \dfrac{51}{170} = 0.3$ and $\hat{p}_{\text{older}} = \dfrac{79}{230} \approx 0.3435$.

The combined proportion is $\hat{p}_c = \dfrac{170(0.3) + 230(0.3435)}{170 + 230} \approx 0.325$.

The sample size is large enough to support an assumption that the sampling distribution of $\hat{p}_{\text{younger}} - \hat{p}_{\text{older}}$ is approximately normal because $170(0.325) = 55.25$, $170(1 - 0.325) = 114.75$, $230(0.325) = 74.75$, and $230(1 - 0.325) = 155.25$ are all at least 10.

The value of the test statistic is $z = \dfrac{0.3 - 0.3435}{\sqrt{0.3250(1 - 0.3250)\left(\frac{1}{170} + \frac{1}{230}\right)}} \approx -0.918.$

The corresponding p-value is $2 * P(z < -0.918) = 0.359$.

Section 3: Because the p-value of approximately 0.359 is greater than $\alpha = 0.05$, the null hypothesis should not be rejected. The results from this study do not provide convincing statistical evidence that the population proportion of exercise center members from 18 to 55 years of age who would be interested in taking online fitness classes is different from the population proportion of adults ages 56 years and older who would be interested in taking online fitness classes.

•A survey conducted by a national research center asked a random sample of 920 teenagers in the United States how often they use a video streaming service. From the sample, 59% answered that they use a video streaming service every day.

(A) Construct and interpret a 95% confidence interval for the proportion of all teenagers in the United States who would respond that they use a video streaming service every day.

(B) Based on the confidence interval in part (a), do the sample data provide convincing statistical evidence that the proportion of all teenagers in the United States who would respond that they use a video streaming service every day is not 0.5? Justify your answer.

Solutions:

(A) The appropriate procedure is a one-sample z-interval for the proportion of all teenagers in the United States who would respond that they use a video streaming service every day.

This survey selected a random sample of 920 teenagers in the United States, which enables the interval to be generalized to the population of interest. This sample of 920 teenagers is less than 10% of the total number of teenagers in the United States, which is required as sampling was conducted without replacement from a finite population. In addition, there were more than 10 successes and 10 failures as $(920)(0.59) = 542.8$ (or 543) responded that they use a streaming service daily and $(920)(0.41) = 377.2$ (or 377) responded that they did not. Thus, the sample size is large enough to support the assumption that the sampling distribution of $\hat{p}$ is approximately normal.
Therefore, a 95% confidence interval for the population proportion is given by

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.59 \pm 1.96 \sqrt{\frac{(0.59)(0.41)}{920}},$$

which is 0.59 ± 0.032, and the interval is $(0.558, 0.622)$.

We can be 95% confident that the proportion of all teenagers in the United States who would respond that they use a streaming service every day is between 0.558 and 0.622.

(B) The 95% confidence interval of (0.558, 0.622) indicates that any value between 0.558 and 0.622 is a plausible value for the proportion of all teenagers in the United States who use video streaming every day. Because the value 0.5 is not contained in the interval, the sample data provide convincing statistical evidence that the proportion of all teenagers in the United States who would use a streaming service every day is not 0.5.

• The manager of a large company that sells pet supplies online wants to increase sales by encouraging repeat purchases. The manager believes that if past customers are offered \$10 off their next purchase, more than 40 percent of them will place an order. To investigate the belief, 90 customers who placed an order in the past year are selected at random. Each of the selected customers is sent an e-mail with a coupon for \$10 off the next purchase if the order is placed within 30 days. Of those who receive the coupon, 38 place an order.

(A) Is there convincing statistical evidence, at the significance level of $\alpha = 0.05$, that the manager's belief is correct? Complete the appropriate inference procedure to support your answer.

(B) Based on your conclusion from part (a), which of the two errors, Type I or Type II, could have been made? Interpret the consequence of the error in context.

Solutions:

(A) Let p represent the proportion of all customers of the pet supply company who would place an order within 30 days after receiving an e-mail with a coupon for \$10 off the next purchase.

The null hypothesis is $\mathrm{H}_0 : p = 0.40$, and the alternative hypothesis is $\mathrm{H}_\mathrm{a} : p > 0.40$.

An appropriate test is a one-sample z-test for a population proportion.

The independent observations condition for performing the one-sample z-test for a population proportion is satisfied because the data were obtained from a random sample of 90 customers who placed an order in the past year and, because sampling of customers is done without replacement, it is assumed that this large online company has more than $10(90) = 900$ customers.

The sample size is large enough to support an assumption that the sampling distribution of $\hat{p}$ is approximately normal because $(90)(0.4) = 36$ and $(90)(1 - 0.4) = 54$ are both at least 10.

The value of the sample proportion is

$\hat{p} = \dfrac{38}{90} \approx 0.422$ and the value of the test statistic is $z = \dfrac{\frac{38}{90} - 0.40}{\sqrt{\frac{(0.40)(0.60)}{90}}} \approx 0.430$. The corresponding p-value is $P(z > 0.430) \approx 0.333$.

Because the p-value is greater than $\alpha = 0.05$, the null hypothesis should not be rejected. The results from this study do not provide convincing statistical evidence that the manager's belief is correct. That is, there is not convincing statistical evidence that more than 40 percent of all customers of the pet supply company would place an order within 30 days after receiving an e-mail with a coupon for \$10 off the next purchase.

(B) Because the null hypothesis was not rejected in part (a), a Type II error could have been made. A Type II error occurs when the null hypothesisis false and is not rejected. In this case, a Type II error is made by failing to reject the nullhypothesis that 40 percent (or less) of all customersallall of the pet supply company would place an order within 30 days after receiving an e-mail with a coupon for \$10 off the next purchase, when in fact, more than 40 percent would do so. Consequently, the manager may decide not to use the coupon promotion when it actually would result in more than 40 percent of their customers making a purchase within 30 days.

• Tumbleweed, commonly found in the western United States, is the dried structure of certain plants that are blown by the wind. Kochia, a type of plant that turns into tumbleweed at the end of the summer, is a problem for farmers because it takes nutrients away from soil that would otherwise go to more beneficial plants. Scientists are concerned that kochia plants are becoming resistant to the most commonly used herbicide, glyphosate. In 2014, 19.7 percent of 61 randomly selected kochia plants were resistant to glyphosate. In 2017, 38.5 percent of 52 randomly selected kochia plants were resistant to glyphosate. Do the data provide convincing statistical evidence, at the level of alpha equals $\alpha = 0.05$, that there has been an increase in the proportion of all kochia plants that are resistant to glyphosate?

Solution:

Section 1:

Let p_{14} represent the proportion of the population of kochia plants in the western United States that were resistant to glyphosate in 2014. Let p_{17} represent the proportion of the population of kochia plants in the western United States that were resistant to glyphosate in 2017.

The null hypothesis $\mathrm{H_0} : p_{17} - p_{14} = 0$ is to be tested against the alternative hypothesis $\mathrm{H_a} : p_{17} - p_{14} > 0$.

An appropriate inference procedure is a two-sample z-test for a difference in proportions. The formula for the test statistic is:

$$z = \frac{\hat{p}_{17} - \hat{p}_{14}}{\sqrt{\left(\frac{\hat{p}_c(1-\hat{p}_c)}{n_{17}} + \frac{\hat{p}_c(1-\hat{p}_c)}{n_{14}}\right)}}$$

where $\hat{p}_c = \frac{n_{14}\hat{p}_{14} + n_{17}\hat{p}_{17}}{n_{14}+n_{17}}$ is a pooled estimate of the proportion of resistant plants for 2014 and 2017 combined.

Section 2:

The first condition for applying the test is that the data are gathered from independent random samples from the populations of kochia plants in the western United States in 2014 and 2017. The question indicates that a random sample of 61 kochia plants was taken in 2014 and a second random sample of 52 kochia plants was taken in 2017. It is reasonable to assume that the 2017 sample of plants was in no way influenced by the 2014 sample of plants.

The second condition is that the sampling distribution of the test statistic is approximately normal. This condition is satisfied because the expected counts under the null hypothesis are all greater than 10. The pooled estimate of the proportion of resistant plants is $\hat{p}_c = \frac{(61)(0.197)+(52)(0.385)}{61+52} \approx 0.2835$. The estimates of the expected counts are $61(0.2835) \approx 17.29$, $61(1-0.2835) \approx 43.71$, $52(0.2835) \approx 14.74$, $52(1-0.2835) \approx 37.26$, all of which are greater than 10.

Because sampling must have been done without replacement, the independence condition for each sample should be checked. Information on the population sizes of kochia plants is not given for either

2014 or 2017, but it is reasonable to assume that each population has millions of plants. Therefore it is reasonable to assume that the sample sizes are less than 10 percent of the respective population sizes.

Using the pooled estimate of the proportion of resistant plants, $\hat{p}_c \approx 0.2835$, the value of the test statistic is:

$$z = \frac{0.385 - 0.197}{\sqrt{\left(\frac{(0.2835)(0.7165)}{61} + \frac{(0.2835)(0.7165)}{52}\right)}} \approx 2.21$$

The p-value is 0.0135.

Section 3: Because the p-value is less than $\alpha = 0.05$, there is convincing statistical evidence to conclude that the proportion of resistant plants in the 2017 population of kochia plants is greater than the proportion of resistant plants in the 2014 population of kochia plants.

Problems adapted from the College Board released tests.