Solution Section 4.1 – Inferences about Two Population Portions

Exercise

A Student surveyed her friends and found that among 20 males, 4 smoke and among 30 female, 6 smoke. Give two reasons why these results should not be used for a hypothesis test of the claim that the proportions of male smokers and female smokers are equal.

Solution

There are two requirements for using the methods of this section, and each of them is violated.

i. The samples should be 2 sample random samples that are independent. These samples are convenience samples, not simple random samples. These samples are likely not independent. Since she surveyed her friends, she may well have males and females that are dating each other (or least that associate with each other) – and people tend to associate with those that have similar behaviors.

ii. The number of successes for each sample should be at least 5, and the number of failures for each sample be at least 5. This is not true for the males, for which $x = 4$.

Using $\hat{p} = \frac{x}{n}$ to estimate p and $\hat{q} = 1 - \frac{x}{n} = \frac{n-x}{n}$ to estimate q.

$$n\hat{p} \geq 5 \qquad\qquad n\hat{q} \geq 5$$

$$n\left(\frac{x}{n}\right) \geq 5 \qquad\qquad n\left(\frac{n-x}{n}\right) \geq 5$$

$$x \geq 5 \qquad\qquad (n-x) \geq 5$$

These inequalities state that the number of successes must be greater than 5, and the number of failures must be greater than 5.

Exercise

In clinical trials of the drug Zocor, some subjects were treated with Zocor and other were given a placebo. The 95% confidence interval estimate of the difference between the proportions of subjects who experienced headaches is $-0.0518 < p_1 - p_2 < 0.0194$. Write a statement interpreting that confidence interval.

Solution

We have 95% confidence that the limits of –0.0518 and 0.01094 contain the true difference between the population proportions of subjects who experience headaches. Repeating the trials many times would result in confidence limits that would include the true difference between the population proportions 95% of the time. Since the interval includes the value 0, there is no significant difference between the two population proportions.

Exercise

Among 8834 malfunctioning pacemakers, in 15.8% the malfunctions were due to batteries. Find the number of successes *x*.

Solution

$$x = (0.158)(8834) \ \underline{\approx 1396}$$

Exercise

Among 129 subjects who took Chantix as an aid to stop smoking, 12.4% experienced nausea. Find the number of successes *x*.

Solution

$$x = (0.124)(129) \ \underline{\approx 16}$$

Exercise

Among 610 adults selected randomly from among the residents of one town, 26.1% said that they have favor stronger gun-control laws. Find the number of successes *x*.

Solution

$$x = (610)(0.261) \ \underline{\approx 159}$$

Exercise

A computer manufacturer randomly selects 2,410 of its computers for quality assurance and finds that 3.13% of these computer are found defective. Find the number of successes *x*.

Solution

$$x = (2,410)(0.0313) \ \underline{\approx 67}$$

Exercise

Assume that you plan to use a significance level of $\alpha = 0.05$ to test the claim that $p_1 = p_2$. Use the given sample sizes and number of successes to find the pooled estimate $\overline{p}$

a) $n_1 = 288, \quad n_2 = 252, \quad x_1 = 75, \quad x_2 = 70$

b) $n_1 = 100, \quad n_2 = 100, \quad \hat{p}_1 = 0.2, \quad \hat{p}_2 = 0.18$

Solution

a) $\hat{p}_1 = \dfrac{x_1}{n_1} = \dfrac{75}{288} = 0.26 \qquad \hat{p}_2 = \dfrac{x_2}{n_2} = \dfrac{70}{252} = 0.278$

$$\bar{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{75+70}{288+252} \underline{= 0.269|}$$

b) $x_1 = n_1 \hat{p}_1 = (100)(0.2) = 20 \qquad x_2 = n_2 \hat{p}_2 = (100)(0.18) = 18$

$$\bar{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{20+18}{100+100} \underline{= 0.19|}$$

Exercise

The numbers of online applications from simple random samples of college applications for 2003 and for the current year are given below.

	2003	Current Year
Number of application in sample	36	27
Number of online applications in sample	13	14

Assume that you plan to use a significance level of $\alpha = 0.05$ to test the claim that $p_1 = p_2$. Find

a) The pooled estimate $\bar{p}$

b) The *x* test statistic

c) The critical *z* values

d) The *P*–value

Assume 95% confidence interval

e) The margin of error E

f) The 95% confidence interval.

Solution

$\hat{p}_1 = \dfrac{x_1}{n_1} = \dfrac{13}{36} = 0.361 \qquad \hat{p}_2 = \dfrac{x_2}{n_2} = \dfrac{14}{27} = 0.519$

$\hat{p}_1 - \hat{p}_2 = 0.361 - 0.519 = -0.157$

a) $\bar{p} = \dfrac{x_1 + x_2}{n_1 + n_2} = \dfrac{13+14}{36+27} = \dfrac{27}{63} \underline{= 0.429|}$

b)
$$z_{\hat{p}_1 - \hat{p}_2} = \frac{\left(\hat{p}_1 - \hat{p}_2\right) - \left(p_1 - p_2\right)}{\sqrt{\dfrac{\bar{p}\bar{q}}{n_1} + \dfrac{\bar{p}\bar{q}}{n_2}}}$$

$$= \frac{-0.157 - 0}{\sqrt{\dfrac{(0.429)(0.571)}{36} + \dfrac{(0.429)(0.571)}{27}}}$$

$\underline{= -1.25|}$

Confidence Level	Critical Value
0.90	1.645
0.95	1.96
0.99	2.575

c) For $\alpha = 0.05$, the critical values are

$$z = \pm z_{\alpha/2} = \pm z_{0.025} \underline{= \pm 1.96}$$

d) $P-value = 2 \cdot P(z < -1.25)$

z	.00	.01	.02	.03	.04	.05
−1.2	.1151	.1131	.1112	.1093	.1075	.1056

$$= 2(0.1056)$$

$$\underline{= 0.2112}$$

e) $E = z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$

$$= 1.96\sqrt{\frac{(0.361)(0.639)}{36} + \frac{(0.519)(0.481)}{27}}$$

$$\underline{= 0.2452}$$

f) $(\hat{p}_1 - \hat{p}_2) - E < (p_1 - p_2) < (\hat{p}_1 - \hat{p}_2) + E$

$$-0.1574 - 0.2452 < p_1 - p_2 < -0.1574 + 0.2452$$

$$\underline{-0.4026 < p_1 - p_2 < 0.0878}$$

Exercise

Chantix is a drug used as an aid to stop smoking. The numbers of subjects experiencing insomnia for each of two treatment groups in a clinical trial of the drug Chantix are given below:

	Chantix Treatment	Placebo
Number in group	129	805
Number experiencing insomnia	19	13

Assume that you plan to use a significance level of $\alpha = 0.05$ to test the claim that $p_1 = p_2$. Find

a) The pooled estimate $\bar{p}$

b) The *x* test statistic

c) The critical *z* values

d) The *P*–value

Assume 95% confidence interval

e) The margin of error E

f) The 95% confidence interval.

Solution

$$\hat{p}_1 = \frac{x_1}{n_1} = \frac{19}{129} = 0.147 \qquad \hat{p}_2 = \frac{x_2}{n_2} = \frac{13}{805} = 0.016$$

$$\hat{p}_1 - \hat{p}_2 = 0.147 - 0.016 = 0.131$$

a) $\bar{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{19+13}{129+805} = \frac{32}{934} \underline{= 0.0343}$

b) $z_{\hat{p}_1-\hat{p}_2} = \dfrac{(\hat{p}_1-\hat{p}_2)-(p_1-p_2)}{\sqrt{\dfrac{\overline{pq}}{n_1}+\dfrac{\overline{pq}}{n_2}}}$

$$= \frac{0.131-0}{\sqrt{\frac{(0.0343)(0.9657)}{129}+\frac{(0.0343)(0.9657)}{805}}}$$

$\underline{=7.60}|$

c) For $\alpha=0.05$, the critical values are

$\frac{\alpha}{2}=\frac{1-0.05}{2}=0.025 \quad \Rightarrow A=1-0.025=0.975$

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767

$z=\pm z_{\alpha/2}=\pm z_{0.025}$

$\underline{=\pm 1.96}|$

d) $P-value=2\cdot P(z>7.60)$

$=2(1-0.9999)$

$\underline{=0.0002}|$

3.50 and up	.9999

e) $E=z_{\alpha/2}\sqrt{\dfrac{\hat{P}_1\hat{q}_1}{n_1}+\dfrac{\hat{P}_2\hat{q}_2}{n_2}}$

$$=1.96\sqrt{\frac{(0.147)(0.853)}{129}+\frac{(0.147)(0.853)}{805}}$$

$\underline{=0.0618}|$

f) $(\hat{p}_1-\hat{p}_2)-E<(p_1-p_2)<(\hat{p}_1-\hat{p}_2)+E$

$0.1311-0.0618<p_1-p_2<0.1311+0.0618$

$\underline{0.0694<p_1-p_2<0.1929}|$

Exercise

In a 1993 survey of 560 college students, 171 said that they used illegal drugs during the previous year. In a recent survey of 720 college students, 263 said that they used illegal drugs during the previous year. Use a 0.05 significance level to test the claim that the proportion of college students using illegal drugs in 1993 was less than it is now.

Solution

$$\hat{p}_1 = \frac{x_1}{n_1} = \frac{171}{560} = 0.305 \qquad \hat{p}_2 = \frac{x_2}{n_2} = \frac{263}{720} = 0.365$$

$$\hat{p}_1 - \hat{p}_2 = 0.305 - 0.365 = -0.60$$

$$\overline{p} = \frac{x_1 + x_2}{n_1 + n_2}$$

$$= \frac{170 + 263}{560 + 720}$$

$$\underline{= 0.339}$$

$(170 + 263)/(560 + 720)$

Original Claim: $p_1 - p_2 < 0$

$$H_0 : p_1 - p_2 = 0$$

$$H_1 : p_1 - p_2 < 0$$

$\alpha = 0.05$, the critical value

z score	Area
1.645	0.9500
2.575	0.9950

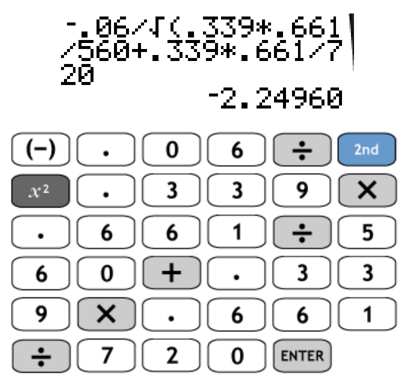
$$z = -z_\alpha = -z_{0.05}$$

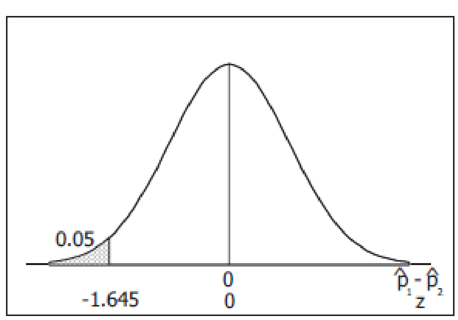
$$\underline{= -1.645}$$

$$z_{\hat{p}_1 - \hat{p}_2} = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{\overline{p}\overline{q}}{n_1} + \frac{\overline{p}\overline{q}}{n_2}}}$$

$$= \frac{-0.060 - 0}{\sqrt{\frac{(0.339)(0.661)}{560} + \frac{(0.339)(0.661)}{720}}}$$

$$\underline{= -2.25}$$





$P-value = P(z < -2.52)$

$\underline{= 0.0122}$

z	.00	.01	.02	.03	.04	.05
-2.2	.0139	.0136	.0132	.0129	.0125	.0122

Conclusion:

Reject H_0; there is sufficient evidence to conclude that $p_1 - p_2 < 0$. There is sufficient evidence to support the claim that the proportion of college students using illegal drugs in 1993 was less than it is now.

Exercise

In a 1993 survey of 560 college students, 171 said that they used illegal drugs during the previous year. In a recent survey of 720 college students, 263 said that they used illegal drugs during the previous year. Construct the confidence interval corresponding to the hypothesis test conducted with a 0.05 significance level. What conclusion does the confidence interval suggest?

Solution

$$\hat{p}_1 = \frac{x_1}{n_1} = \frac{171}{560} = 0.305 \qquad \hat{p}_2 = \frac{x_2}{n_2} = \frac{263}{720} = 0.365$$

$$\hat{p}_1 - \hat{p}_2 = 0.305 - 0.365 = -0.60$$

$$\left(\hat{p}_1 - \hat{p}_2\right) \pm z_{\alpha/2}\sigma_{\hat{p}_1 - \hat{p}_2}$$

$$\left(\hat{p}_1 - \hat{p}_2\right) \pm z_{\alpha/2}\sqrt{\frac{\overline{p}_1\overline{q}_1}{n_1} + \frac{\overline{p}_2\overline{q}_2}{n_2}}$$

$$-0.0599 \pm 1.645\sqrt{\frac{(0.305)(.695)}{560} + \frac{(0.365)(.935)}{720}}$$

$$-0.0599 \pm 0.0480$$

$$-0.0599 - 0.0480 < p_1 - p_2 < -0.0599 + 0.0480$$

$$-0.1079 < p_1 - p_2 < -0.0119$$

Since the confidence interval does not include the value 0, it suggests that the two population proportions are not equal and that the proportion of college students using illegal drugs in 1993 was less than it is now.

Exercise

A simple random sample of front-seat occupants involved in car crashes is obtained. Among 2823 occupants not wearing seat belts, 31 were killed. Among 7765 occupants wearing seat belts, 16 were killed. Construct a 90% confidence interval estimate of the difference between the fatality rates for those not wearing seat belts and those wearing seat belts. What does the result suggest about the effectiveness of seat belts?

Solution

$$\hat{p}_1 = \frac{x_1}{n_1} = \frac{31}{2823} = 0.01098 \qquad \hat{p}_2 = \frac{x_2}{n_2} = \frac{16}{7765} = 0.00206$$

$$\hat{p}_1 - \hat{p}_2 = 0.01098 - 0.00206 = 0.00892$$

$$\left(\hat{p}_1 - \hat{p}_2\right) \pm z_{\alpha/2}\sigma_{\hat{p}_1 - \hat{p}_2}$$

$$\left(\hat{p}_1 - \hat{p}_2\right) \pm z_{\alpha/2}\sqrt{\frac{\bar{p}_1\bar{q}_1}{n_1} + \frac{\bar{p}_2\bar{q}_2}{n_2}}$$

$$-0.00892 \pm 1.645\sqrt{\frac{(.01098)(.98902)}{2823} + \frac{(.00206)(.99794)}{7765}}$$

$$-0.00892 \pm 0.00334$$

$$-0.00892 - 0.00334 < p_1 - p_2 < -0.00892 + 0.00334$$

$$0.00558 < p_1 - p_2 < 0.01226$$

Since the confidence interval does not include the value 0, it suggests that the two population proportions are not equal and that seat belts are effective because the proportion of non-users who killed is greater than the proportion of users who are killed.

Exercise

A Pew Research Center poll asked randomly selected subjects if they agreed with the statement that "It is morally wrong fir married people to have an affair" Among the 386 women surveyed, 347 agrees with the statement. Among the 359 men surveyed, 305 agreed with the statement.

a) Use a 0.05 significance level to test the claim that the percentage of women who agree is difference from the percentage of men who agree. Does there appear to be a difference in the way women and men feel about this issue?

b) Construct the confidence interval corresponding to the hypothesis test conducted with a 0.05 significance level. What conclusion does the confidence interval suggest?

Solution

a) $\hat{p}_1 = \frac{x_1}{n_1} = \frac{347}{386} = 0.899 \qquad \hat{p}_2 = \frac{x_2}{n_2} = \frac{305}{359} = 0.850$

$$\hat{p}_1 - \hat{p}_2 = 0.899 - 0.85 = 0.049$$

$$\bar{p} = \frac{437 + 305}{386 + 359} = 0.875$$

$$\text{Original Claim: } p_1 - p_2 \neq 0$$

$$H_0 : p_1 - p_2 = 0$$

$$H_1 : p_1 - p_2 \neq 0$$

$$\alpha = 0.05,\ \frac{\alpha}{2} = \frac{1 - 0.05}{2} = 0.025 \quad \Rightarrow A = 1 - 0.025 = 0.975$$

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767

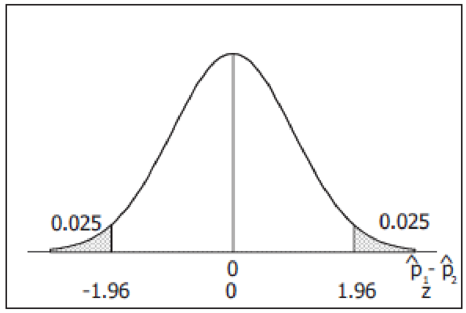
$$z = \pm z_{\alpha/2} = \pm z_{0.025}$$

$$= \pm 1.96$$

$$z_{\hat{p}_1 - \hat{p}_2} = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{\overline{p}\overline{q}}{n_1} + \frac{\overline{p}\overline{q}}{n_2}}}$$

$$= \frac{0.049 - 0}{\sqrt{\frac{(.875)(.125)}{386} + \frac{(.875)(.125)}{359}}}$$

$$= 2.04$$



$$P-value = 2 \cdot P(z > 2.04)$$

$$= 2 \cdot (1 - 0.9793)$$

$$= 0.0414$$

z	.00	.01	.02	.03	.04	.05
2.0	.9772	.9778	.9783	.9788	.9793	

Conclusion:

Reject H_0; there is sufficient evidence to conclude that $p_1 - p_2 \neq 0$ (in fact, that $p_1 - p_2 > 0$. There is sufficient evidence to support the claim that the percentage of women who agree is different from the percentage of men who agree. Yes; there does appear to be a difference in the way that women and men feel about the issue.

b) $(\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2}\sigma_{\hat{p}_1 - \hat{p}_2}$

$$(\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2}\sqrt{\frac{\overline{p}_1\overline{q}_1}{n_1} + \frac{\overline{p}_2\overline{q}_2}{n_2}}$$

$$0.04938 \pm 1.96\sqrt{\frac{(.899)(.101)}{386} + \frac{(.85)(.15)}{359}}$$

$$0.04938 \pm 0.04766$$

$$0.04938 - 0.04766 < p_1 - p_2 < 0.04938 + 0.04766$$

$$0.00172 < p_1 - p_2 < 0.09704$$

Since the confidence interval does not include the value 0, it suggests that the two population proportions are not equal and the percentage of women who agree is different from the percentage of men who agree. Since the interval includes only positive values, conclude that the percentage of women who agree is greater than the percentage of men who agree.

Exercise

Tax returns include an option of designating $3 for presidential election campaigns, and it does not cost the taxpayer anything to make that designation. In a simple random sample of 250 tax returns from 1976, 27.6% of the returns designated the $3 for the campaign. In a simple random sample of 300 recent tax returns, 7.3% of the returns designated the $3 for the campaign. Use a 0.05 significance level to test the claim that the percentage of returns designating the $3 for the campaign was greater in 1973 than it is now.

Solution

$$x_1 = (.276)(250) = 69 \qquad x_2 = (.073)(300) = 22$$

$$\hat{p}_1 = \frac{x_1}{n_1} = \frac{69}{250} = 0.276 \qquad \hat{p}_2 = \frac{x_2}{n_2} = \frac{22}{300} = 0.073$$

$$\hat{p}_1 - \hat{p}_2 = 0.276 - 0.073 = 0.203$$

$$\overline{p} = \frac{69+22}{250+300} = 0.165$$

Original Claim: $p_1 - p_2 > 0$

$$H_0 : p_1 - p_2 = 0$$

$$H_1 : p_1 - p_2 > 0$$

$\alpha = 0.01$

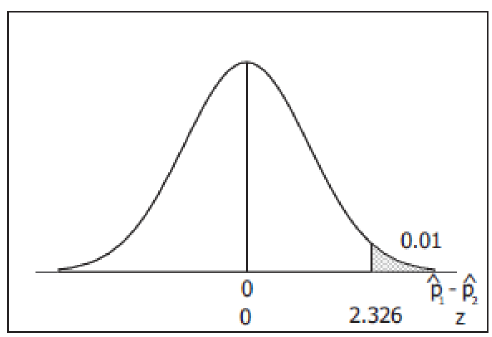
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
$\boxed{2.3}$	.9893	.9896	.9898	$\boxed{.9901}$	.9904					

$$z = z_\alpha = z_{0.01} = \underline{2.326}$$

$$z_{\hat{p}_1 - \hat{p}_2} = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{\overline{p}\overline{q}}{n_1} + \frac{\overline{p}\overline{q}}{n_2}}}$$

$$= \frac{0.203 - 0}{\sqrt{\frac{(.165)(.835)}{250} + \frac{(.165)(.835)}{300}}}$$

$$\underline{= 6.37}$$



3.50 and up	.9999

$$P - value = P(z > 6.37)$$

$$= 1 - 0.9999$$

$$\underline{= 0.0001}$$

Conclusion:

Reject H_0; there is sufficient evidence to conclude that $p_1 - p_2 > 0$. There is sufficient evidence to support the claim that the percentage of returns designated funds for campaigns was greater on 1976 than it is now.

Exercise

In an experiment, 16% of 734 subjects treated with Viagra experienced headaches. In the same experiment, 4% of 725 subjects given a placebo experienced headaches.

a) Use a 0.01 significance level to test the claim that the proportion of headaches is greater for those treated with Viagra. Do headaches appear to be a concern for those who take Viagra?

b) Construct the confidence interval corresponding to the hypothesis test conducted with a 0.01 significance level. What conclusion does the confidence interval suggest?

Solution

a) $\hat{p}_1 = \dfrac{x_1}{n_1} = \dfrac{x_1}{734} = 0.16 \qquad \hat{p}_2 = \dfrac{x_2}{n_2} = \dfrac{x_2}{725} = 0.04$

$$\hat{p}_1 - \hat{p}_2 = 0.16 - 0.04 = 0.12$$

$$\overline{p} = \frac{(.16)(734) + (.04)(725)}{734 + 725} = 0.100$$

Original Claim: $p_1 - p_2 > 0$

$$H_0 : p_1 - p_2 = 0$$

$$H_1 : p_1 - p_2 > 0$$

$$\alpha = 0.01$$

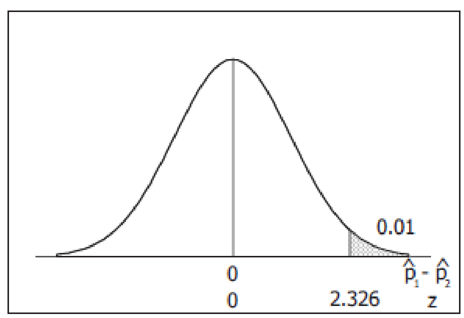
z	.00	.01	.02	.03	.04	.05
2.3	.9893	.9896	.9898	.9901	.9904	

$$z = z_\alpha = z_{0.01} = \underline{2.326}$$

$$z_{\hat{p}_1 - \hat{p}_2} = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{\overline{p}\overline{q}}{n_1} + \frac{\overline{p}\overline{q}}{n_2}}}$$

$$= \frac{0.12 - 0}{\sqrt{\frac{(0.1)(0.9)}{734} + \frac{(0.1)(0.9)}{725}}}$$

$$\underline{= 7.63}$$



$$P - value = P(z > 7.63)$$

$$= 1 - 0.9999$$

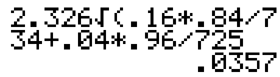
$$\underline{= 0.0001}$$

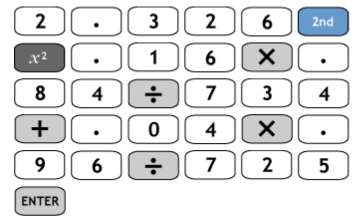
3.50 and up	.9999

Conclusion:

Reject H_0; there is sufficient evidence to conclude that $p_1 - p_2 > 0$. There is sufficient evidence to support the claim that the proportion of persons experiencing headaches is greater for those treated with Viagra. Yes; headaches do appear to be a concern for those who take Viagra.

b) $\left(\hat{p}_1 - \hat{p}_2\right) \pm z_{\alpha/2}\sqrt{\dfrac{\overline{p}_1\overline{q}_1}{n_1} + \dfrac{\overline{p}_2\overline{q}_2}{n_2}}$





$0.12 \pm 2.326\sqrt{\dfrac{(.16)(.84)}{734} + \dfrac{(.04)(.96)}{725}}$

0.12 ± 0.0357

$0.12 - 0.0357 < p_1 - p_2 < 0.12 + 0.0357$

$0.0843 < p_1 - p_2 < 0.1557$

Since the confidence interval does not include the value 0, there is a significant difference the two proportions. Since the confidence interval includes only positive values, the proportion of persons experiencing headaches is greater for those treated with Viagra.

Exercise

Two different simple random samples are drawn from two different populations. The first sample consists of 20 people with 10 having a common attribute. The second sample consists of 2000 people with 1404 of them having the same common attribute. Compare the results from a hypothesis test of $p_1 = p_2$ (with a 0.05 significance level) and a 95% confidence interval estimate of $p_1 - p_2$.

Solution

$\hat{p}_1 = \dfrac{x_1}{n_1} = \dfrac{10}{20} = 0.5$ $\qquad$ $\hat{p}_2 = \dfrac{x_2}{n_2} = \dfrac{1404}{2000} = 0.702$

$\hat{p}_1 - \hat{p}_2 = 0.5 - 0.702 = -0.202$

$\overline{p} = \dfrac{10 + 1404}{20 + 2000} = 0.70$

Original Claim: $p_1 - p_2 \neq 0$

$H_0 : p_1 - p_2 = 0$

$H_1 : p_1 - p_2 \neq 0$

$\alpha = 0.05, \; \dfrac{\alpha}{2} = \dfrac{1 - 0.05}{2} = 0.025 \quad \Rightarrow A = 1 - 0.025 = 0.975$

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767

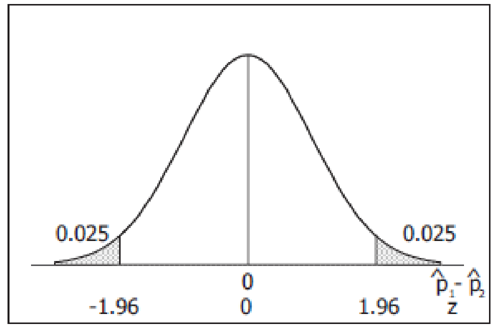
$z = \pm z_{\alpha/2} = \pm z_{0.025}$

$= \pm 1.96$

$$z_{\hat{p}_1-\hat{p}_2} = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{\overline{pq}}{n_1} + \frac{\overline{pq}}{n_2}}}$$

$$= \frac{-0.202 - 0}{\sqrt{\frac{(0.7)(0.3)}{20} + \frac{(0.7)(0.3)}{2000}}}$$

$$= -1.9615$$



$$P-value = 2 \cdot P(z < -1.96)$$

$$= 2 \cdot (.025)$$

$$\approx 0.05$$

z	.00	.01	.02	.03	.04	.05	.06
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250

Conclusion:

Reject H_0; there is sufficient evidence to conclude that $p_1 - p_2 = 0$ and conclude that $p_1 - p_2 \le 0$ (in fact, that $p_1 - p_2 < 0$).

The confidence interval is:

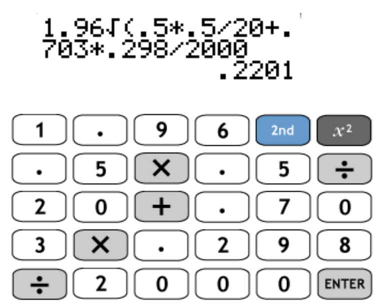
$$(\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2}\sqrt{\frac{\overline{p}_1\overline{q}_1}{n_1} + \frac{\overline{p}_2\overline{q}_2}{n_2}}$$

$$-0.202 \pm 1.96\sqrt{\frac{(0.5)(0.5)}{20} + \frac{(0.702)(0.298)}{2000}}$$

$$-0.202 \pm 0.220$$

$$-0.202 - 0.220 < p_1 - p_2 < -0.202 + 0.220$$

$$-0.422 < p_1 - p_2 < 0.018$$



Since the confidence interval includes the value 0, p_1 *and* p_2 could have the same values and one should not reject the claim that $p_1 - p_2 = 0$.

The test of hypothesis and the confidence interval lead to different conclusions. In this instance, they are not equivalent.

Exercise

A report on the nightly news broadcast stated that 11 out of 142 households with pet dogs were burglarized and 21 out of 217 without pet dogs were burglarized. Find the *z* test statistic for the hypothesis test.

Assume that you plan to use a significance level of $\alpha = 0.05$ to test the claim that $p_1 = p_2$.

Solution

$$\hat{p}_1 = \frac{x_1}{n_1} = \frac{11}{142} = 0.0775 \qquad \hat{p}_2 = \frac{x_2}{n_2} = \frac{21}{217} = 0.0968$$

$$\hat{p}_1 - \hat{p}_2 = .0775 - .0968 = -0.0193$$

$$\bar{p} = \frac{11+21}{142+217} = 0.089$$

$$z = \frac{\left(\hat{p}_1 - \hat{p}_2\right) - \left(p_1 - p_2\right)}{\sqrt{\frac{\bar{p}\bar{q}}{n_1} + \frac{\bar{p}\bar{q}}{n_2}}}$$

$$= \frac{-.0193 - 0}{\sqrt{\frac{(.089)(0.911)}{142} + \frac{(.089)(0.911)}{217}}}$$

$$\underline{= -0.62|}$$

Exercise

Assume that the samples are independent and that they have been randomly selected. Construct a 90% confidence interval for the difference between population proportions $p_1 = p_2$

$n_1 = 39, \quad n_2 = 50, \quad x_1 = 13, \quad x_2 = 28$

Solution

$$\hat{p}_1 = \frac{x_1}{n_1} = \frac{13}{39} = 0.3333 \qquad \hat{p}_2 = \frac{x_2}{n_2} = \frac{28}{50} = 0.56$$

$$\hat{p}_1 - \hat{p}_2 = 0.33 - 0.56 = -0.23$$

$$\bar{p} = \frac{13+28}{39+50} = 0.461$$

$$A = 0.9 \quad \Rightarrow \quad z = 1.645$$

$$\left(\hat{p}_1 - \hat{p}_2\right) \pm z_{\alpha/2}\sqrt{\frac{\bar{p}_1\bar{q}_1}{n_1} + \frac{\bar{p}_2\bar{q}_2}{n_2}} = -0.23 \pm 1.645\sqrt{\frac{(0.461)(0.539)}{39} + \frac{(0.461)(0.539)}{50}}$$

$$-0.23 \pm 0.175$$

$$-0.23 - 0.175 < p_1 - p_2 < -0.23 + 0.175$$

$$\underline{-0.405 < p_1 - p_2 < -0.055|}$$

Exercise

The sample size needed to estimate the difference between two population proportions ti within a margin of error E with a confidence level of $1-\alpha$ can be found as follows:

$$E = z_{\alpha/2}\sqrt{\frac{p_1q_1}{n_1}+\frac{p_2q_2}{n_2}}.$$

In this expression, replace n_1 and n_2 by n (assuming both samples have the same size) and replace each of p_1, q_1, p_2 *and* q_2 by 0.5 (because their values are not known). Then solve for n.

Use this approach to find the size pf each sample of you want to estimate the difference between the proportions of men and women who plan to vote in the next presidential election. Assume that you want 99% confidence that your error is no more than 0.05.

Solution

$$A = 99\% = 0.99 \quad \Rightarrow z = 2.575$$

$$E = z_{\alpha/2}\sqrt{2\frac{p_1q_1}{n}} \qquad p_1 = q_1 = p_2 = q_2 = 0.5$$

$$0.05 = 2.575\sqrt{2\frac{(0.5)^2}{n}}$$

$$\frac{0.05}{2.575} = \sqrt{\frac{2(0.5)^2}{n}}$$

$$\left(\frac{0.05}{2.575}\right)^2 = \frac{2(0.5)^2}{n}$$

$$n = 2(0.5)^2\left(\frac{2.575}{0.05}\right)^2 \quad \approx 1,327$$