MATHEMATICS 201-510-LW
Business Statistics
Martin Huard
Fall 2008

XXII – Inferences for Two Proportions
SOLUTIONS

1. A manufacturer has two factories producing Ipods. A random sample of 500 Ipods from the first factory revealed that 44 were defective, whereas a random sample of 600 Ipods from the second factory revealed that 27 were defective.
   a) Construct a 99% confidence interval for the difference between the proportions of defective Ipods at the two factories.

Step 1 Assumptions: $n_1 = 500 > 20$ $\quad n_2 = 600 > 20$

$n_1\hat{p}_1 = 44 > 5$ $\quad n_2\hat{p}_2 = 27 > 5$

$n_1\hat{q}_1 = 456 > 5$ $\quad n_2\hat{q}_2 = 573 > 5$

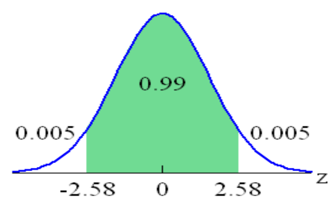
The samples are independent.

Step 2 a) Test statistic: z

b) Level of confidence: $1-\alpha = 0.99$ or $\alpha = 0.01$

Step 3 Point estimate : $\hat{p}_1 - \hat{p}_2 = 0.0880 - 0.0450 = 0.0430$

Step 4 a) $z_{\frac{\alpha}{2}} = z_{0.005} = 2.58$



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

$= 2.58\sqrt{\frac{0.088\cdot 0.912}{500} + \frac{0.045\cdot 0.955}{600}} = 0.0393$

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

$0.0430 - 0.0393 < p_1 - p_2 < 0.0430 - 0.0393$

$0.0037 < p_1 - p_2 < 0.0823$

Step 5 The 99% confidence interval for the difference between the proportions of defective Ipods at the two factories is 0.37% to 8.23%.

b) At the 1% level of significance, can you conclude that the proportion of defective Ipods is higher in the first factory than in the second? Try with both approaches, the classical and the *p*-value.

Classical Approach

Step 1 Assumptions: $n_1 = 500 > 20$ $n_2 = 600 > 20$

$n_1\hat{p}_1 = 44 > 5$ $n_2\hat{p}_2 = 27 > 5$

$n_1\hat{q}_1 = 456 > 5$ $n_2\hat{q}_2 = 573 > 5$

The samples are independent.

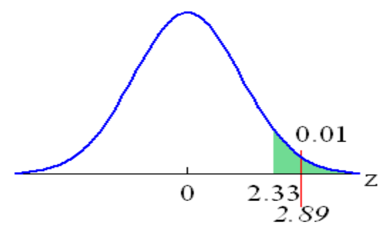
Step 2 $H_0:\ p_1 - p_2 = 0$

$H_A:\ p_1 - p_2 > 0$

Step 3 a) Test statistic: z

b) Right-tailed test with $\alpha = 0.01$

c) $z_\alpha = z_{0.01} = 2.33$



Step 4 a) $\hat{p}_p = \frac{44+27}{500+600} = \frac{71}{1100} = 0.06455$

b) $z = \dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}_p\hat{q}_p\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \dfrac{\frac{44}{500} - \frac{27}{600}}{\sqrt{\frac{71}{1100}\frac{1029}{1100}\left(\frac{1}{500}+\frac{1}{600}\right)}} = 2.89$

Step 5 a) z is in the critical region

b) Reject H_0.

∴ There is sufficient evidence at the 1% level of significance to conclude that the proportion of defective Ipods is higher in the first factory than in the second.

p-value approach

Step 1 Assumptions: $n_1 = 500 > 20$ $n_2 = 600 > 20$

$n_1\hat{p}_1 = 44 > 5$ $n_2\hat{p}_2 = 27 > 5$

$n_1\hat{q}_1 = 456 > 5$ $n_2\hat{q}_2 = 573 > 5$

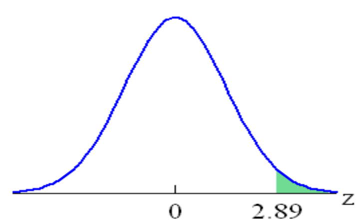
The samples are independent.

Step 2 $H_0:\ p_1 - p_2 = 0$

$H_A:\ p_1 - p_2 > 0$

Step 3 a) Test statistic: z

b) Right-tailed test with $\alpha = 0.01$



Step 4 a) $\hat{p}_p = \frac{44+27}{500+600} = \frac{71}{1100} = 0.06455$

b) $z = \dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}_p\hat{q}_p\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \dfrac{\frac{44}{500} - \frac{27}{600}}{\sqrt{\frac{71}{1100}\frac{1029}{1100}\left(\frac{1}{500}+\frac{1}{600}\right)}} = 2.89$

c) p-value $= P\left(z > 2.89\right) = 1 - 0.9981 = 0.0019$

Step 5 a) p-value $= 0.0019 < \alpha = 0.01$

b) Reject H_0.

∴ There is sufficient evidence at the 1% level of significance to conclude that the proportion of defective Ipods is higher in the first factory than in the second.

2. In a survey of working parents (both parents working), one of the questions asked was "Have you refused a job, promotion, or transfer because it would mean less time with your family?" Two hundred men and 200 women were asked this question. Fifty-eight men and forty-eight women responded "yes".

   a) Construct a 90% confidence interval for the difference in the proportion of men and women who answered "Yes".

Step 1 Assumptions: $n_M = 200 > 20$ $\quad n_M \hat{p}_M = 58 > 5$ $\quad n_M \hat{q}_M = 142 > 5$

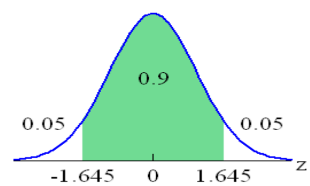
$n_W = 200 > 20$ $\quad n_W \hat{p}_W = 48 > 5$ $\quad n_W \hat{q}_W = 152 > 5$

The samples are independent.

Step 2 a) Test statistic: z

b) Level of confidence: $1 - \alpha = 0.90$ or $\alpha = 0.10$

Step 3 Point estimate : $\hat{p}_M - \hat{p}_W = \frac{58}{200} - \frac{48}{200} = \frac{10}{200} = 0.05$



Step 4 a) $z_{\frac{\alpha}{2}} = z_{0.05} = 1.645$

b) $E = z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}_M \hat{q}_M}{n_M} + \frac{\hat{p}_W \hat{q}_W}{n_W}} = 1.645 \sqrt{\frac{\frac{58}{200}\frac{142}{200}}{200} + \frac{\frac{48}{200}\frac{152}{200}}{200}} = 0.0725$

c) $(\hat{p}_M - \hat{p}_W) - E < p_M - p_W < (\hat{p}_M - \hat{p}_W) + E$

$0.0500 - 0.0725 < p_M - p_W < 0.0500 + 0.0725$

$-0.0225 < p_M - p_W < 0.1225$

Step 5 The 90% confidence interval for the difference in the proportion of men and women who have refused a job, promotion, or transfer because it would mean less time with their family is -2.25% to12.25%.

b) Based on this survey, can we conclude that there is a difference in the proportion of men and women responding "yes" at the 10% level of significance? Try with both approaches, the classical and the *p*-value.

Classical Approach

Step 1 Assumptions: $n_M = 200 > 20$ $\quad n_M\hat{p}_M = 58 > 5$ $\quad n_M\hat{q}_M = 142 > 5$

$n_W = 200 > 20$ $\quad n_W\hat{p}_W = 48 > 5$ $\quad n_W\hat{q}_W = 152 > 5$

The samples are independent.

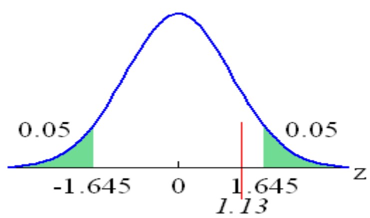
Step 2 $H_0: p_M - p_W = 0$

$H_A: p_M - p_W \neq 0$

Step 3 a) Test statistic: z

b) Two-tailed test with $\alpha = 0.10$

c) $z_{\frac{\alpha}{2}} = z_{0.05} = 1.645$



Step 4 b) $\hat{p}_p = \frac{58+48}{200+200} = \frac{106}{400} = 0.2650$

b) $$z = \frac{\hat{p}_M - \hat{p}_W}{\sqrt{\hat{p}_p\hat{q}_p\left(\frac{1}{n_M} + \frac{1}{n_W}\right)}} = \frac{\frac{58}{200} - \frac{48}{200}}{\sqrt{\frac{106}{400}\frac{294}{400}\left(\frac{1}{200} + \frac{1}{200}\right)}} = 1.13$$

Step 5 a) z is not in the critical region

b) Fail to reject H_0.

∴ There is not sufficient evidence at the 10% level of significance to conclude that there is a difference in the proportion of men and women who have refused a job, promotion, or transfer because it would mean less time with their family.

p-value approach

Step 1 Assumptions: $n_M = 200 > 20$ $\quad n_M\hat{p}_M = 58 > 5$ $\quad n_M\hat{q}_M = 142 > 5$

$n_W = 200 > 20$ $\quad n_W\hat{p}_W = 48 > 5$ $\quad n_W\hat{q}_W = 152 > 5$

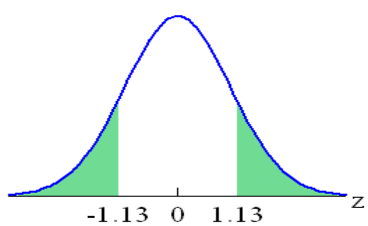
The samples are independent.

Step 2 $H_0: p_M - p_W = 0$

$H_A: p_M - p_W \neq 0$

Step 3 a) Test statistic: z

b) Two-tailed test with $\alpha = 0.10$



Step 4 a) $\hat{p}_p = \frac{58+48}{200+200} = \frac{106}{400} = 0.2650$

b) $$z = \frac{\hat{p}_M - \hat{p}_W}{\sqrt{\hat{p}_p\hat{q}_p\left(\frac{1}{n_M} + \frac{1}{n_W}\right)}} = \frac{\frac{58}{200} - \frac{48}{200}}{\sqrt{\frac{106}{400}\frac{294}{400}\left(\frac{1}{200} + \frac{1}{200}\right)}} = 1.13$$

c) $p\text{-value} = 2P(z < -1.13) = 2 \cdot 0.1292 = 0.2584$

Step 5 a) $p\text{-value} = 0.2584 > \alpha = 0.10$

b) Fail to reject H_0.

∴ There is not sufficient evidence at the 10% level of significance to conclude that there is a difference in the proportion of men and women who have refused a job, promotion, or transfer because it would mean less time with their family.

3. A random sample 76 adults ages 18-24 showed that 11 had donated blood within the past year, while a random sample of 156 adults who were at least 25 years old had 18 people who had donated blood within the past year.
   a) Construct a 95% confidence interval for the difference in the proportion of adults who give blood for the two age groups.

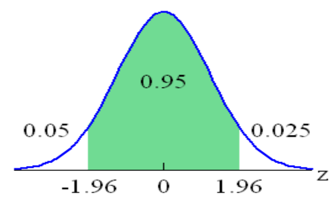
Step 1 Assumptions: $n_{18-24} = 76 > 20$ $n_{18-24}\hat{p}_{18-24} = 11 > 5$ $n_{18-24}\hat{q}_{18-24} = 65 > 5$

$n_{\geq 25} = 156 > 20$ $n_{\geq 25}\hat{p}_{\geq 25} = 18 > 5$ $n_{\geq 25}\hat{q}_{\geq 25} = 138 > 5$

The samples are independent.

Step 2 a) Test statistic: z

b) Level of confidence: $1-\alpha = 0.95$ or $\alpha = 0.05$



Step 3 Point estimate : $\hat{p}_{18-24} - \hat{p}_{\geq 25} = \frac{11}{76} - \frac{18}{156} = 0.0294$

Step 4 a) $z_{\frac{\alpha}{2}} = z_{0.025} = 1.96$

b) $E = z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}_{18-24}\hat{q}_{18-24}}{n_{18-24}} + \frac{\hat{p}_{\geq 25}\hat{q}_{\geq 25}}{n_{\geq 25}}} = 1.96\sqrt{\frac{\frac{11}{76}\frac{65}{76}}{76} + \frac{\frac{18}{156}\frac{138}{156}}{200}} = 0.0937$

c) $(\hat{p}_{18-24} - \hat{p}_{\geq 25}) - E < p_{18-24} - p_{\geq 25} < (\hat{p}_{18-24} - \hat{p}_{\geq 25}) + E$

$0.0294 - 0.0937 < p_{18-24} - p_{\geq 25} < 0.0294 + 0.0937$

$-0.0643 < p_{18-24} - p_{\geq 25} < 0.1230$

Step 5 The 95% confidence interval for the difference in the proportion of adults who give blood for the two age groups is -6.43% to12.30%.

b) Using a 5% level of significance, can you conclude that there is a difference in the proportion of adults who donate blood in the two age groups? Try with both approaches, the classical and the *p*-value.

Classical Approach

Step 1 Assumptions: $n_{18-24} = 76 > 20$ $n_{18-24}\hat{p}_{18-24} = 11 > 5$ $n_{18-24}\hat{q}_{18-24} = 65 > 5$

$n_{\geq 25} = 156 > 20$ $n_{\geq 25}\hat{p}_{\geq 25} = 18 > 5$ $n_{\geq 25}\hat{q}_{\geq 25} = 138 > 5$

The samples are independent.

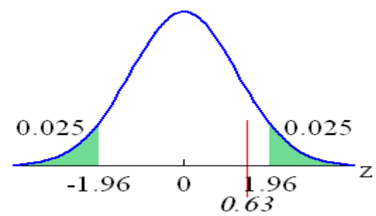
Step 2 $H_0:\ p_{18-24} - p_{\geq 25} = 0$

$H_A:\ p_{18-24} - p_{\geq 25} \neq 0$

Step 3 a) Test statistic: z

b) Two-tailed test with $\alpha = 0.05$

c) $z_{\frac{\alpha}{2}} = z_{0.025} = 1.96$



Step 4 c) $\hat{p}_p = \frac{11+18}{76+156} = \frac{29}{232} = 0.125$

b) $$z = \frac{\hat{p}_{18-24} - \hat{p}_{\geq 25}}{\sqrt{\hat{p}_p\hat{q}_p\left(\frac{1}{n_{18-24}} + \frac{1}{n_{\geq 25}}\right)}} = \frac{\frac{11}{76} - \frac{18}{156}}{\sqrt{\frac{29}{232}\frac{203}{232}\left(\frac{1}{76} + \frac{1}{156}\right)}} = 0.63$$

Step 5 a) z is not in the critical region

b) Fail to reject H_o.

∴ There is not sufficient evidence at the 5% level of significance to conclude that that there is a difference in the proportion of adults who donate blood in the two age groups.

p-value approach

Step 1 Assumptions: $n_{18-24} = 76 > 20$ $n_{18-24}\hat{p}_{18-24} = 11 > 5$ $n_{18-24}\hat{q}_{18-24} = 65 > 5$

$n_{\geq 25} = 156 > 20$ $n_{\geq 25}\hat{p}_{\geq 25} = 18 > 5$ $n_{\geq 25}\hat{q}_{\geq 25} = 138 > 5$

The samples are independent.

Step 2 $H_0:\ p_{18-24} - p_{\geq 25} = 0$

$H_A:\ p_{18-24} - p_{\geq 25} \neq 0$

Step 3 a) Test statistic: z

b) Two-tailed test with $\alpha = 0.05$

Step 4 a) $\hat{p}_p = \frac{11+18}{76+156} = \frac{29}{232} = 0.125$

b) $$z = \frac{\hat{p}_{18-24} - \hat{p}_{\geq 25}}{\sqrt{\hat{p}_p\hat{q}_p\left(\frac{1}{n_{18-24}} + \frac{1}{n_{\geq 25}}\right)}} = \frac{\frac{11}{76} - \frac{18}{156}}{\sqrt{\frac{29}{232}\frac{203}{232}\left(\frac{1}{76} + \frac{1}{156}\right)}} = 0.63$$

c) $p\text{-value} = 2P(z < -0.63) = 2 \cdot 0.2643 = 0.5286$

Step 5 a) $p\text{-value} = 0.5286 > \alpha = 0.10$

b) Fail to reject H_o.

∴ There is not sufficient evidence at the 5% level of significance to conclude that that there is a difference in the proportion of adults who donate blood in the two age groups.

4. In a survey of on commuting habits, a random sample of 1242 homeowners revealed that 756 drove to works, and in a survey of 1106 renters, 467 drove to work.
   a) Construct a 98% confidence interval for the difference in the proportion of homeowners and renters who drive to work.

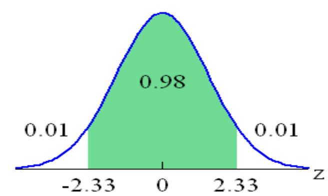
Step 1 Assumptions: $n_H = 1242 > 20$ $\quad n_H\hat{p}_H = 756 > 5$ $\quad n_H\hat{q}_H = 486 > 5$

$n_R = 1106 > 20$ $\quad n_R\hat{p}_R = 467 > 5$ $\quad n_R\hat{q}_R = 639 > 5$

The samples are independent.

Step 2 a) Test statistic: z

b) Level of confidence: $1-\alpha = 0.98$ or $\alpha = 0.02$



Step 3 Point estimate : $\hat{p}_H - \hat{p}_R = \frac{756}{1242} - \frac{467}{1106} = 0.1865$

Step 4 a) $z_{\frac{\alpha}{2}} = z_{0.025} = 2.33$

b) $E = z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}_H\hat{q}_H}{n_H} + \frac{\hat{p}_R\hat{q}_R}{n_R}} = 2.33\sqrt{\frac{\frac{756}{1242}\frac{486}{1242}}{1242} + \frac{\frac{467}{1106}\frac{639}{1106}}{1106}} = 0.0473$

c) $(\hat{p}_H - \hat{p}_R) - E < p_H - p_R < (\hat{p}_H - \hat{p}_R) + E$

$0.1865 - 0.0473 < p_H - p_R < 0.1865 - 0.0473$

$0.1392 < p_H - p_R < 0.2339$

Step 5 The 98% confidence interval for the difference in the proportion of homeowners and renters who drive to work is 13.92% to 23.39%.

b) At the 2% level of significance, test the claim that the proportion of homeowners who drive to work is higher than the proportion of renters? Try with both approaches, the classical and the *p*-value.

Classical Approach

Step 1 Assumptions: $n_H = 1242 > 20$ $\quad n_H\hat{p}_H = 756 > 5$ $\quad n_H\hat{q}_H = 486 > 5$

$n_R = 1106 > 20$ $\quad n_R\hat{p}_R = 467 > 5$ $\quad n_R\hat{q}_R = 639 > 5$

The samples are independent.

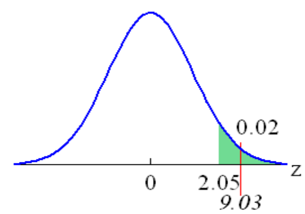
Step 2 $H_0:\ p_H - p_R = 0$

$H_A:\ p_H - p_R > 0$

Step 3 a) Test statistic: z

b) Right-tailed test with $\alpha = 0.02$

c) $z_\alpha = z_{0.02} = 2.05$



Step 4 a) $\hat{p}_p = \frac{756+467}{1242+1106} = \frac{1223}{2348} = 0.5209$

b) $z = \dfrac{\hat{p}_H - \hat{p}_R}{\sqrt{\hat{p}_p\hat{q}_p\left(\frac{1}{n_H}+\frac{1}{n_R}\right)}} = \dfrac{\frac{756}{1242} - \frac{467}{1106}}{\sqrt{\frac{1223}{2348}\frac{1125}{2348}\left(\frac{1}{1242}+\frac{1}{1106}\right)}} = 9.03$

Step 5 a) z is in the critical region

b) Reject H_o.

∴ There is sufficient evidence at the 2% level of significance to conclude that the proportion of homeowners who drive to work is higher than the proportion of renters.

p-value approach

Step 1 Assumptions: $n_H = 1242 > 20$ $\quad n_H\hat{p}_H = 756 > 5$ $\quad n_H\hat{q}_H = 486 > 5$

$n_R = 1106 > 20$ $\quad n_R\hat{p}_R = 467 > 5$ $\quad n_R\hat{q}_R = 639 > 5$

The samples are independent.

Step 2 $H_0:\ p_H - p_R = 0$

$H_A:\ p_H - p_R > 0$

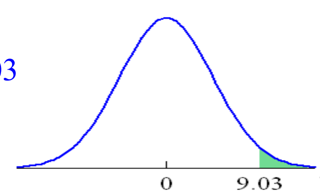
Step 3 a) Test statistic: z

b) Right-tailed test with $\alpha = 0.02$

Step 4 a) $\hat{p}_p = \frac{756+467}{1242+1106} = \frac{1223}{2348} = 0.5209$

b) $z = \dfrac{\hat{p}_H - \hat{p}_R}{\sqrt{\hat{p}_p\hat{q}_p\left(\frac{1}{n_H}+\frac{1}{n_R}\right)}} = \dfrac{\frac{756}{1242} - \frac{467}{1106}}{\sqrt{\frac{1223}{2348}\frac{1125}{2348}\left(\frac{1}{1242}+\frac{1}{1106}\right)}} = 9.03$

c) p-value $= P(z > 9.03) = 0.000$



Step 5 a) p-value $= 0.000 < \alpha = 0.02$

b) Reject H_o.

∴ There is sufficient evidence at the 2% level of significance to conclude that the proportion of homeowners who drive to work is higher than the proportion of renters.

5. A random sample of 2000 American adults showed that 658 had a college degree. Another sample of 1500 Canadians revealed that 542 had a college degree.

   a) Construct a 94% confidence interval for the difference in the proportion of Canadians and Americans who have a college degree.

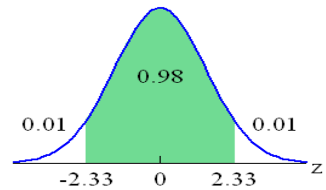
Step 1 Assumptions: $n_A = 2000 > 20$ $\quad n_A\hat{p}_A = 658 > 5$ $\quad n_A\hat{q}_A = 1342 > 5$

$n_C = 1500 > 20$ $\quad n_C\hat{p}_C = 542 > 5$ $\quad n_C\hat{q}_C = 958 > 5$

The samples are independent.

Step 2 a) Test statistic: z

b) Level of confidence: $1-\alpha = 0.94$ or $\alpha = 0.06$



Step 3 Point estimate : $\hat{p}_C - \hat{p}_A = \frac{542}{1500} - \frac{658}{2000} = 0.0323$

Step 4 d) $z_{\frac{\alpha}{2}} = z_{0.03} = 1.88$

e) $E = z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}_A\hat{q}_A}{n_A} + \frac{\hat{p}_C\hat{q}_C}{n_C}} = 1.88\sqrt{\frac{\frac{658}{2000}\frac{1342}{2000}}{2000} + \frac{\frac{542}{1500}\frac{958}{1500}}{1500}} = 0.0306$

f) $(\hat{p}_C - \hat{p}_A) - E < p_C - p_A < (\hat{p}_C - \hat{p}_A) + E$

$0.0323 - 0.0306 < p_C - p_A < 0.0323 + 0.0306$

$0.0017 < p_C - p_A < 0.0629$

Step 5 The 94% confidence interval for the difference in the proportion of Canadians and Americans who have a college degree is 0.17% to 6.29%.

b) Using a 4% level of significance, can we conclude that the proportion of Canadians who have a college degree is higher than the proportion for Americans? Try with both approaches, the classical and the *p*-value.

Classical Approach

Step 1 Assumptions: $n_A = 2000 > 20$ $n_A\hat{p}_A = 658 > 5$ $n_A\hat{q}_A = 1342 > 5$

$n_C = 1500 > 20$ $n_C\hat{p}_C = 542 > 5$ $n_C\hat{q}_C = 958 > 5$

The samples are independent.

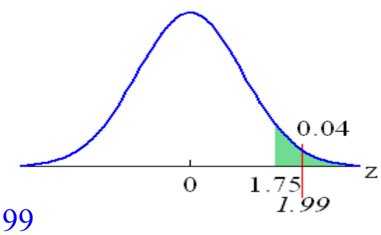
Step 2 $H_0: p_C - p_A = 0$

$H_A: p_C - p_A > 0$

Step 3 a) Test statistic: z

b) Right-tailed test with $\alpha = 0.04$

c) $z_\alpha = z_{0.04} = 1.75$



Step 4 b) $\hat{p}_p = \frac{658+542}{2000+1500} = \frac{1200}{3500} = 0.3429$

b) $$z = \frac{\hat{p}_C - \hat{p}_A}{\sqrt{\hat{p}_p\hat{q}_p\left(\frac{1}{n_C}+\frac{1}{n_A}\right)}} = \frac{\frac{542}{1500} - \frac{658}{2000}}{\sqrt{\frac{1200}{3500}\frac{2300}{3500}\left(\frac{1}{1500}+\frac{1}{2000}\right)}} = 1.99$$

Step 5 a) z is in the critical region

b) Reject H_o.

∴ There is sufficient evidence at the 4% level of significance to conclude that the proportion of Canadians who have a college degree is higher than the proportion for Americans.

p-value approach

Step 1 Assumptions: $n_A = 2000 > 20$ $n_A\hat{p}_A = 658 > 5$ $n_A\hat{q}_A = 1342 > 5$

$n_C = 1500 > 20$ $n_C\hat{p}_C = 542 > 5$ $n_C\hat{q}_C = 958 > 5$

The samples are independent.

Step 2 $H_0: p_C - p_A = 0$

$H_A: p_C - p_A > 0$

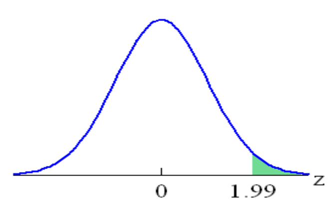
Step 3 a) Test statistic: z

b) Right-tailed test with $\alpha = 0.04$

Step 4 d) $\hat{p}_p = \frac{658+542}{2000+1500} = \frac{1200}{3500} = 0.3429$

e) $$z = \frac{\hat{p}_C - \hat{p}_A}{\sqrt{\hat{p}_p\hat{q}_p\left(\frac{1}{n_C}+\frac{1}{n_A}\right)}} = \frac{\frac{542}{1500} - \frac{658}{2000}}{\sqrt{\frac{1200}{3500}\frac{2300}{3500}\left(\frac{1}{1500}+\frac{1}{2000}\right)}} = 1.99$$

f) $p\text{-value} = P(z > 1.99) = 1 - 0.9767 = 0.0233$



Step 5 a) p-value $= 0.0233 < \alpha = 0.04$

b) Reject H_o.

∴ There is sufficient evidence at the 4% level of significance to conclude that the proportion of Canadians who have a college degree is higher than the proportion for Americans.

6. A researcher conducted a survey of 880 adult drivers. Eight hundred sixty-two of the drivers acknowledged that running red lights is hazardous. Another sample of 240 adult drivers showed that 106 said they do not run red lights.
   a) Construct a 95% confidence interval for the difference in the proportion of drivers who label running red lights as hazardous and the proportion of drivers who do not run red lights.

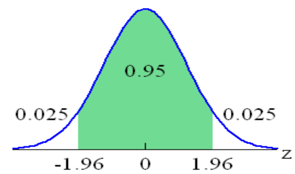
Step 1 Assumptions: $n_{label} = 880 > 20$ $\quad n_{label}\hat{p}_{label} = 862 > 5$ $\quad n_{label}\hat{q}_{label} = 18 > 5$

$n_{run} = 240 > 20$ $\quad n_{run}\hat{p}_{run} = 106 > 5$ $\quad n_{run}\hat{q}_{run} = 134 > 5$

The samples are independent.

Step 2 a) Test statistic: z

b) Level of confidence: $1-\alpha = 0.95$ or $\alpha = 0.05$



Step 3 Point estimate : $\hat{p}_C - \hat{p}_A = \frac{862}{880} - \frac{106}{240} = 0.5379$

Step 4 g) $z_{\frac{\alpha}{2}} = z_{0.025} = 1.96$

h) $E = z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}_{label}\hat{q}_{label}}{n_{label}} + \frac{\hat{p}_{run}\hat{q}_{run}}{n_{run}}} = 1.96\sqrt{\frac{\frac{862}{880}\frac{18}{880}}{880} + \frac{\frac{106}{240}\frac{134}{240}}{240}} = 0.0635$

i) $\left(\hat{p}_{label} - \hat{p}_{run}\right) - E < p_{label} - p_{run} < \left(\hat{p}_{label} - \hat{p}_{run}\right) + E$

$0.5379 - 0.0635 < p_{label} - p_{run} < 0.5379 + 0.0635$

$0.4744 < p_{label} - p_{run} < 0.6014$

Step 5 The 95% confidence interval for the difference in the proportion of Canadians and Americans who have a college degree is 47.44% to 60.14%.

b) At the 5% level of significance, test the claim that the proportion of drivers who label running red lights as hazardous is greater than the proportion of drivers who do not run red lights. Try with both approaches, the classical and the *p*-value.

Classical Approach

Step 1 Assumptions: $n_{label} = 880 > 20$ $\quad n_{label}\hat{p}_{label} = 862 > 5$ $\quad n_{label}\hat{q}_{label} = 18 > 5$

$n_{run} = 240 > 20$ $\quad n_{run}\hat{p}_{run} = 106 > 5$ $\quad n_{run}\hat{q}_{run} = 134 > 5$

The samples are independent.

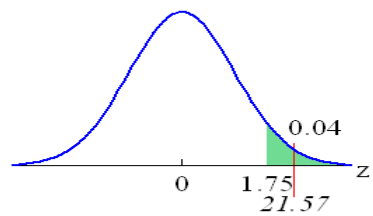
Step 2 $H_0 : p_{label} - p_{run} = 0$

$H_A : p_{label} - p_{run} > 0$

Step 3 a) Test statistic: z

b) Right-tailed test with $\alpha = 0.05$

c) $z_\alpha = z_{0.05} = 1.645$



Step 4 c) $\hat{p}_p = \frac{862+106}{880+240} = \frac{968}{1120} = 0.8643$

b) $$z = \frac{\hat{p}_{label} - \hat{p}_{run}}{\sqrt{\hat{p}_p\hat{q}_p\left(\frac{1}{n_{label}} + \frac{1}{n_{run}}\right)}} = \frac{\frac{862}{880} - \frac{106}{240}}{\sqrt{\frac{968}{1120}\frac{152}{1120}\left(\frac{1}{880} + \frac{1}{240}\right)}} = 21.57$$

Step 5 a) z is in the critical region

b) Reject H_o.

∴ There is sufficient evidence at the 5% level of significance to conclude that the proportion of drivers who label running red lights as hazardous is greater than the proportion of drivers who do not run red lights.

p-value approach

Step 1 Assumptions: $n_{label} = 880 > 20$ $\quad n_{label}\hat{p}_{label} = 862 > 5$ $\quad n_{label}\hat{q}_{label} = 18 > 5$

$n_{run} = 240 > 20$ $\quad n_{run}\hat{p}_{run} = 106 > 5$ $\quad n_{run}\hat{q}_{run} = 134 > 5$

The samples are independent.

Step 2 $H_0 : p_{label} - p_{run} = 0$

$H_A : p_{label} - p_{run} > 0$

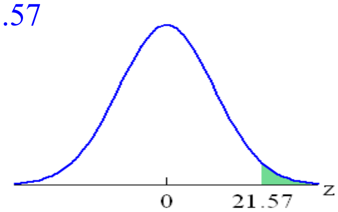
Step 3 a) Test statistic: z

b) Right-tailed test with $\alpha = 0.05$

Step 4 g) $\hat{p}_p = \frac{862+106}{880+240} = \frac{968}{1120} = 0.8643$

h) $$z = \frac{\hat{p}_{label} - \hat{p}_{run}}{\sqrt{\hat{p}_p\hat{q}_p\left(\frac{1}{n_{label}} + \frac{1}{n_{run}}\right)}} = \frac{\frac{862}{880} - \frac{106}{240}}{\sqrt{\frac{968}{1120}\frac{152}{1120}\left(\frac{1}{880} + \frac{1}{240}\right)}} = 21.57$$

i) p-value $= P(z > 21.57) = 0.000$



Step 5 a) p-value $= 0.000 < \alpha = 0.04$

b) Reject H_o.

∴ There is sufficient evidence at the 5% level of significance to conclude that the proportion of Canadians who have a college degree is higher than the proportion for Americans.

7. What is wrong with this statistical project?
Mary wants to know if the proportion of students who have used illicit drugs within the past year is significantly different from the proportion of students who are in favor of legalizing marijuana. For this, she selects 50 students at random and asks each of them if they have used illicit drugs within the past year and, also, if they are in favor of legalizing marijuana. She then applies the hypothesis test introduced in this section to the results she obtains.

The samples are dependent since each student answers both questions.