Social Studies 201
Fall 2006

Answers to Problem Set 5

1. **Number of close friends**

   (a) i. Let μ be the true mean number of close friends for Saskatchewan adults aged 15-24. From Table 1, the sample size for the sample of 15-24 year olds is $n = 180$ and the test statistic is $\bar{X}$. Given this large sample size, $\bar{X}$ has a normal distribution with mean μ and standard deviation $\sigma/\sqrt{n}$, where the sample standard deviation $s = 5.04$ is used as an estimate of σ. For 90% confidence, the Z value is 1.645 and the 90% interval estimate is:

$$
\begin{aligned}
\bar{X} \pm Z\frac{\sigma}{\sqrt{n}} &= \bar{X} \pm 1.645\frac{5.04}{\sqrt{180}} \\
&= \bar{X} \pm 1.645\frac{5.04}{13.416} \\
&= \bar{X} \pm (1.645 \times 0.376) \\
&= \hat{X} \pm 0.618 \\
&= 6.38 \pm 0.62
\end{aligned}
$$

   or from 5.76 to 7.00.

   ii. For those aged 75 plus, the method is the same, with μ representing the true mean number of close friends for all Saskatchewan residents aged 75 plus. Again the sample size $n = 134$ is large, so the sample mean has a normal distribution. Also, since n is large, s is used as an estimate of σ, and the 90% interval estimate is

$$
\begin{aligned}
\bar{X} \pm Z\frac{\sigma}{\sqrt{n}} &= \bar{X} \pm 1.645\frac{7.73}{\sqrt{134}} \\
&= \bar{X} \pm 1.645\frac{7.73}{11.576} \\
&= \bar{X} \pm (1.645 \times 0.668)
\end{aligned}
$$

$$
\begin{aligned}
&= \hat{X} \pm 1.098 \\
&= 7.35 \pm 1.10
\end{aligned}
$$

or from 6.25 to 8.45.

(b) For the first of the two age groups, those aged 15-24, let p be the true proportion of all Saskatchewan residents aged 15-24 who have two or less close friends. From Table 1, in the sample the proportion of respondents with two or less close friends is $\hat{p} = (2+28)/180 = 0.167$. Since $n = 180$ is large ($n = 180 > 5/0.167 = 29.9$), the sample proportion $\hat{p}$ is normally distributed with mean p and standard deviation $\sqrt{pq/n}$. For the 95% interval estimate, $Z = \pm 1.96$. By using $p = q = 0.5$, the widest possible interval for any of the proportions is obtained. Alternatively, $\hat{p} = 0.167$ and $\hat{q} = 1 - \hat{p} = 1 - 0.167 = 0.833$ could be used for p and q in the estimate of the standard deviation of $\hat{p}$. Using these latter values, the interval estimate is:

$$
\begin{aligned}
\hat{p} \pm Z\sqrt{\frac{pq}{n}} &= \hat{p} \pm 1.96\sqrt{\frac{0.167 \times 0.833}{180}} \\
&= \hat{p} \pm 1.96\sqrt{\frac{0.1391}{180}} \\
&= \hat{p} \pm 1.96\sqrt{0.000773} \\
&= \hat{p} \pm (1.96 \times 0.0278) \\
&= \hat{p} \pm 0.054 \\
&= 0.167 \pm 0.054
\end{aligned}
$$

or from 0.113 to 0.221. The 95% interval estimate for the proportion of all Saskatchewan residents aged 15-24 with 2 or less close friends is from 0.11 to 0.22.

For those aged 75 plus, the method is exactly the same. The sample size is $n = 134$ and the sample proportion is $\hat{p} = (16 + 22)/134 = 38/134 = 0.284$. Also, $\hat{q} = 1 - \hat{p} = 1 = 0.284 = 0.716$. The 95% interval estimate is

$$
\hat{p} \pm Z\sqrt{\frac{pq}{n}} = \hat{p} \pm 1.96\sqrt{\frac{0.284 \times 0.716}{134}}
$$

$$
\begin{aligned}
&= \hat{p} \pm 1.96\sqrt{\frac{0.20334}{134}} \\
&= \hat{p} \pm 1.96\sqrt{0.001517} \\
&= \hat{p} \pm (1.96 \times 0.0390) \\
&= \hat{p} \pm 0.0764 \\
&= 0.284 \pm 0.076
\end{aligned}
$$

or from 0.208 to 0.360. The 95% interval estimate for the proportion of all Saskatchewan residents aged 75 plus with 2 or less close friends is from 0.21 to 0.36.

(c) i. Let μ be the true mean number of close friends for all those of age 45-54 in Saskatchewan. The hypotheses are:

$$H_0 \ : \ \mu = 6.3 \quad \text{(same as overall mean of 6.3)}$$

$$H_1 \ : \ \mu \neq 6.3 \quad \text{(differs from overall mean 0f 6.3)}$$

This is a two-tailed test with the null hypothesis being that this age group has the same mean number of close friends as the overall mean of 6.3 among all Saskatchewan residents. The alternative hypothesis is that the mean number of close friends for this age group differs from the overall mean of 6.3. Since $n = 173$ is large (over 30), the sample mean has a normal distribution with mean μ and standard deviation $\sigma/\sqrt{n}$, by the Central Limit Theorem. The significance level is $\alpha = 0.05$ and since the test is a two-tailed test the critical Z-value is ± 1.96. H_0 is rejected for all $Z < -1.96$ or $Z > +1.96$. From the data in Table 1, the Z-value is

$$
\begin{aligned}
Z &= \frac{\bar{X} - \mu}{s/\sqrt{n}} \\
&= \frac{6.48 - 6.3}{5.95/\sqrt{173}} \\
&= \frac{0.18}{5.95/13.153} \\
&= \frac{0.18}{0.452} \\
&= -0.398
\end{aligned}
$$

Since this Z-value is greater than -1.96 but less than +1.96, there is not sufficient evidence to reject the null hypothesis at the 0.10 level of significance. The conclusion is that there is not sufficient evidence to conclude that the mean number of close friends for those of aged 45-54 differs from the overall mean of 6.3 close friends. This conclusion is made at 0.05 level of statistical significance.

ii. For the age group 75 plus, the method is the same as in part (i), except that the alternative hypothesis is that the mean for this group exceeds 6.3 close friends. The hypotheses are:

$$H_0 \ : \ \mu = 6.3 \quad \text{(same as overall mean of 6.3)}$$

$$H_1 \ : \ \mu > 6.3 \quad \text{(exceeds overall mean of 6.3)}$$

This is a one-tailed test with the null hypothesis being that this age group has the same mean number of close friends as the overall mean of 6.3 among all Saskatchewan residents. The alternative hypothesis is that the mean number of close friends for this older age group exceeds from the overall mean of 6.3. Since $n = 134$ is large (over 30), the sample mean has a normal distribution with mean μ and standard deviation $\sigma/\sqrt{n}$, by the Central Limit Theorem. The significance level is $\alpha = 0.05$ and since the test is a one-tailed test the critical Z-value is 1.645 H_0 is rejected for all $Z > 1.645$. From the data in Table 1, the Z-value is

$$\begin{aligned} Z &= \frac{\bar{X} - \mu}{s/\sqrt{n}} \\ &= \frac{7.35 - 6.3}{7.73/\sqrt{134}} \\ &= \frac{1.05}{7.73/11.576} \\ &= \frac{1.05}{0.668} \\ &= 1.572 \end{aligned}$$

This Z-value is not greater than 1.645 so, at the 0.05 level of significance, the null hypothesis is not rejected. While sample gives a mean of 7.35, somewhat above 6.3, in this case the difference is not large enough to reject the null hypothesis that the mean number of close friends for those of age 75 plus differs from 6.3.

(d) Let p be the true proportion of all those aged 15-24 with two or less close friends. From part b., the sample size is $n = 180$, $\hat{p} = 0.167$ and $\hat{q} = 0.833$. Since the question asks whether the proportion with two or less close friends is less than one-quarter, this is a one-tailed test with alternative hypothess that $p < 1/4 = 0.25$. The hypotheses are

$$H_0 \ : \ p = 0.25$$
$$H_1 \ : \ p < 0.25$$

The sample size is $n = 180$ and since $\hat{p} = 0.167$ is smaller than $\hat{q} = 0.833$, n divided by the smaller of p or q gives $5/0.167 = 29.9$ and $n = 180 > 29.9$ so this is a large sample size and $\hat{p}$ has a normal distribution with mean p and standard deviation $\sqrt{pq/n}$. At $\alpha = 0.10$ significance, with a one-tailed test, the critical $Z = 1.28$. When the sample yields a Z value to the left of -1.28, the null hypothesis is rejected and the alternative hypothesis is accepted. From the sample data,

$$\begin{aligned} Z &= \frac{\hat{p} - p}{\sqrt{pq/n}} \\ &= \frac{0.167 - 0.25}{\sqrt{(0.25 \times 0.75)/180}} \\ &= \frac{-0.083}{\sqrt{0.001042}} \\ &= \frac{-0.083}{0.0323} \\ &= -2.57 < -1.28 \end{aligned}$$

Since the Z-value is to the left of -1.28, the conclusion is to reject the null hypothesis at the 0.10 level of significance. At the 0.10 level of significance the data provide evidence that the proportion of all 15-24 year olds in Saskatchewan with two or less close friends is less than one-quarter.

(e) From Table 1, the oldest age group has the largest mean number of close friends, about 1 more than for the other two age groups. However, the sample sizes are not large part a. shows that the 90% interval estimates for the mean of the 15-24 and 75 plus age groups overlap. This raises some question about whether the pattern of increasing means in Table 1 would hold if all Saskatchewan residents in each of the three age groups were surveyed. The results from part c. also lead one to question whether the means differ for the three groups. For each of the youngest and oldest residents, the null hypothesis that the mean is 6.3 cannot be rejected.

Part b. gives a result that may seem contradictory to the above, in that the oldest age group has a larger proportion with two or less close friends, as compared with the youngest age group. As well, the two intervals in b. do not overlap all that much. The apparent contradication can be resolved by noting that the oldest age group also has more respondents with more than 20 close friends, as compared with the youngest age group. What seems to be happening here is that older respondents are more varied than are younger respondents in terms of the number of close friends. That is, as compared with the youngest age group, there is a larger proportion of older respondents with few close friends and also with many close friends. Also note that the standard deviation is larger for the oldest age group, demonstrating that the distribution of number of close friends is more varied for the older age group than for the other two age groups.

2. **Sample sizes**

(a) This questions asks whether the sample size necessary to estimate the mean accurate to within $E = 0.1$ units on the five-point scale. The required sample size is

$$n = \left(\frac{Z\sigma}{E}\right)^2$$

where E is the accuracy required. In order to be correct seventeen in twenty times, or 85% of the time, the Z-value is 1.44. Using the larger of the two standard deviations in Table 2, the estimate of σ is 1.29.

If the accuracy of the estimate is to be 0.1 units on the five-point scale, or $E = 0.1$, then the required sample size is

$$\begin{aligned} n &= \left(\frac{Z\sigma}{E}\right)^2 \\ &= \left(\frac{1.44 \times 1.29}{0.1}\right)^2 \\ &= 18.576^2 \\ &= 345.1 \end{aligned}$$

or $n = 346$, rounding up to ensure a large enough sample size.

For an estimate to be accurate to within 0.25 units on the five-point scale, $E = 0.25$. For probability 0.98, $Z = 2.33$. As an estimate of σ, use either of the two sample standard deviations. Using the sample standard deviation from the rural respondents, the required sample size is

$$\begin{aligned} n &= \left(\frac{Z\sigma}{E}\right)^2 \\ &= \left(\frac{2.33 \times 1.19}{0.25}\right)^2 \\ &= 11.091^2 \\ &= 123.0 \end{aligned}$$

or $n = 124$, rounding up to ensure a large enough sample size.

(b) For the three parts of this question, either the formula for the sample size for a proportion or the interval estimate for a proportion can be used. For part (i), the formula for sample size is used. For the other two parts, the formula for an interval estimate is used.

i. For estimating a proportion, the formula for sample size is

$$n = \left(\frac{Z}{E}\right)^2 pq$$

where E is the accuracy of the estimate. The claim is that the accuracy, or margin of error, is 3.9 percentage points or $E = 0.039$. The confidence level is stated to be 95%, so $Z = 1.96$. Using $p = q = 0.5$ gives the largest possible sample size for any given accuracy and confidence level. Using these values in the formula gives

$$\begin{aligned} n &= \left(\frac{Z}{E}\right)^2 pq \\ &= \left(\frac{1.96}{0.039}\right)^2 0.5 \times 0.5 \\ &= 50.256^2 \times 0.25 \\ &= 2525.71 \times 0.25 \\ &= 631.4 \end{aligned}$$

or $n = 632$. This sample size is a little larger than the stated sample size of $n = 615$, but is of a similar order of magnitude.

ii. In this case, the formula for an interval estimate is used. 19 in 20 times is equivalent to $19/20 = 0.95$ or 95%, so the Z-value for this confidence level is 1.96. The interval is:

$$\begin{aligned} \hat{p} \pm Z\sqrt{\frac{pq}{n}} &= \hat{p} \pm 1.96\sqrt{\frac{0.5 \times 0.5}{1,001}} \\ &= \hat{p} \pm 1.96\sqrt{0.00024975} \\ &= \hat{p} \pm (1.96 \times 0.01580) \\ &= \hat{p} \pm 0.03097 \end{aligned}$$

or 0.031, or 3.1%. In this case, the statement seems correct.

iii. Again, using the method of an interval estimate, the interval would be

$$\begin{aligned}
\hat{p} \pm Z\sqrt{\frac{pq}{n}} &= \hat{p} \pm 1.96\sqrt{\frac{0.5 \times 0.5}{602}} \\
&= \hat{p} \pm 1.96\sqrt{0.000415282} \\
&= \hat{p} \pm (1.96 \times 0.020378) \\
&= \hat{p} \pm 0.03994
\end{aligned}$$

or 0.040, or 4.0 percentage points. Again, the interval estimate verifies the statement.

(c) One-third of 615 is 205, so if this many are undecided, the effective sample size is reduced from $n = 615$ to $n = 615 - 205 = 410$. Using the method of interval estimates from parts ii. and iii. of b., the intervals would be:

$$\begin{aligned}
\hat{p} \pm Z\sqrt{\frac{pq}{n}} &= \hat{p} \pm 1.96\sqrt{\frac{0.5 \times 0.5}{410}} \\
&= \hat{p} \pm 1.96\sqrt{0.000609756} \\
&= \hat{p} \pm (1.96 \times 0.024693) \\
&= \hat{p} \pm 0.04840
\end{aligned}$$

or 0.048, or 4.8 percentage points. If the 95% confidence level continues to be used, the margin of error is ± 4.8%. While the result in b. ii. appears to represent an accuracy of ±3.9% with 95% confidence, if the large number of undecided is taken into account, the accuracy of the estimate is considerably reduced, with a wider interval estimate of ±4.8%.

3. **Concern about climate change**

(a) i. Let μ be the true mean number response to the question on climate change, on the 1-5 scale for all Saskatchewan urban respondents. From Table 2, the sample size for the sample of urban residents is $n = 395$ and the test statistic is $\bar{X}$. Given this large sample size, $\bar{X}$ has a normal distribution with mean μ and standard deviation $\sigma/\sqrt{n}$, where the sample standard deviation $s = 1.29$ is used as an estimate of σ. Since no confidence level is stated in the question, you could select whichever level you wish, although the level selected should generally be at 80% or more. For 95% confidence, the Z value is 1.96 and the 95% interval estimate is:

$$\begin{aligned}\bar{X} \pm Z\frac{\sigma}{\sqrt{n}} &= \bar{X} \pm 1.96\frac{1.29}{\sqrt{395}} \\ &= \bar{X} \pm 1.96\frac{1.29}{19.875} \\ &= \bar{X} \pm (1.96 \times 0.0649) \\ &= \hat{X} \pm 0.127 \\ &= 3.68 \pm 0.13\end{aligned}$$

or from 3.55 to 3.81.

ii. For this part, let μ be the true mean number response to the question on climate change, for all Saskatchewan rural residents. From Table 2, the sample size for the sample of rural residents is $n = 220$ and the test statistic is $\bar{X}$. Given this large sample size, $\bar{X}$ has a normal distribution with mean μ and standard deviation $\sigma/\sqrt{n}$, where the sample standard deviation $s = 1.19$ is used as an estimate of σ. Again using the 95% confidence level, the Z value is 1.96 and the 95% interval estimate is:

$$\begin{aligned}\bar{X} \pm Z\frac{\sigma}{\sqrt{n}} &= \bar{X} \pm 1.96\frac{1.19}{\sqrt{220}} \\ &= \bar{X} \pm 1.96\frac{1.19}{14.832}\end{aligned}$$

$$
\begin{aligned}
&= \bar{X} \pm (1.96 \times 0.0802) \\
&= \hat{X} \pm 0.157 \\
&= 3.61 \pm 0.16
\end{aligned}
$$

or from 3.45 to 3.77.

(b) Let p represent the proportion of all Saskatchewan urban residents who consider the issue of climate change to be important or very important. From Table 2, the sample the percentage of urban respondents who consider the issue important or very important is $35.3+24.0 = 59.3\%$. This is the same as a proportion of $\hat{p} = 0.593$. From this, $\hat{q} = 1 - \hat{p} = 1 - 0.593 = 0.407$. The smaller of $\hat{p}$ or $\hat{q}$ is the latter and $n = 395$ is larger than $5/0.407 = 12.3$, so the sample size can be considered large in this sample and the sample proportion $\hat{p}$ is normally distributed with mean p and standard deviation $\sqrt{pq/n}$. Again using the 95% interval estimate, $Z = \pm 1.96$. Using these sample proportions as estimates of p and q, the interval estimate is:

$$
\begin{aligned}
\hat{p} \pm Z\sqrt{\frac{pq}{n}} &= \hat{p} \pm 1.96\sqrt{\frac{0.593 \times 0.407}{395}} \\
&= \hat{p} \pm 1.96\sqrt{\frac{0.2414}{395}} \\
&= \hat{p} \pm 1.96\sqrt{0.00061110} \\
&= \hat{p} \pm (1.96 \times 0.02472) \\
&= \hat{p} \pm 0.048449 \\
&= 0.593 \pm 0.048
\end{aligned}
$$

or from 0.545 to 0.641. The 95% interval estimate for the proportion of all Saskatchewan urban residents aged 15-24 with 2 or less close friends is from 0.545 to 0.641, or from 54.5% to 64.1%.

(c) Let p be the true proportion of rural residents of Saskatchewan who consider the issue of climate change to be important or very important. The question is to test whether this proportion is more than one-half. The hypotheses are

$$H_0 \ : \ p = 0.50$$

$$H_1 \quad : \quad p > 0.50$$

The sample size is $n = 220$ and the test statistic is $\hat{p}$, where $\hat{p} = 0.543$ (sum of 34.0% and 20.3%, expressed as a proportion), $\hat{q} = 1 - \hat{p} = 1 - 0.543 = 0.457$. To ensure this is a large sample size, n divided by the smaller of p or q gives $5/0.457 = 10.9$ and $n = 220 > 10.9$ so this is a large sample size and $\hat{p}$ has a normal distribution with mean p and standard deviation $\sqrt{pq/n}$. Since no significance level is given, pick the $\alpha = 0.05$ significance level, and with a one-tailed test, the critical $Z = 1.645$. When the sample $Z > +1.645$, the null hypothesis is rejected but when $Z < +1.645$, the null hypothesis is not rejected. From the sample data,

$$\begin{aligned} Z &= \frac{\hat{p} - p}{\sqrt{pq/n}} \\ &= \frac{0.543 - 0.50}{\sqrt{(0.50 \times 0.50)/220}} \\ &= \frac{0.043}{0.03371} \\ &= 1.276 < 1.645 \end{aligned}$$

Since the Z-value is less than 1.645, the conclusion is that the null hypothesis cannot be rejected at the 0.05 level of significance. In the sample, there are more than one-half of rural respondents who consider the issue to be important or very important, but using the 0.05 level of significance the sample does not provide sufficient evidence to conclude that among all Saskatchewan rural residents, more than one-half consider the issue to be important or very important.

(d) The samples of urban and rural respondents are very similar in terms of the mean response, with respective means of 3.68 and 3.61. From part a., the two intervals also overlap a lot, so there is not much to distinguish the two groups in terms of the mean response.

From part b., the urban respondents appear to have a stronger expression of the importance of the issue than do the rural respondents. From b., the interval estimate for the urban residents does not cross the 50% point, so this provides fairly strong evidence that more than 50% of all urban residents agree that this issue is important or very important. In contrast, from b., there is insufficient evidence to conclude that more than one-half of rural residents consider the issue important or very important.

It appears that while the mean response for the two groups is very similar, those in urban areas are a little more polarized in their responses, as compared with rural residents. For the rural respondents, responses appear a little more concentrated in the middle (at 2-3), whereas for urban respondents, there are more responses at the extremes, at response 1, 4, or 5. This more diverse set of responses for urban respondents produces a slightly larger standard deviation for the urban than for the rural respondents.

4. **Downsizing and job loss**.

 (a) Chi-square – not required for answering this problem set but the following provides a description of the test and analysis.

 The hypotheses for a chi-square test are as follows:

 Null hyp. H_0: No relationship between age and labour force status

 Alternative hyp. H_1: Relationship between age and labour force status

 The expected values for this test are reported as 'expected count' for each cell. These expected values or counts are obtained under the assumption that the null hypothesis is true, that is, that there is no relationship between the two variables. From the counts and expected counts, the Pearson chi-square value of $\chi^2 = 11.025$ results. The number of degrees of freedom (df) is 6.

 The assumption required for conducting the test is that each of the expected counts exceeds 5. In Table 3, the minimum expected count is 7.6, so this assumption is met.

 If a significance level of $\alpha = 0.05$ is used, the critical chi-square value with 6 degrees of freedom is 12.592. That is, the critical region for rejecting the null hypopthesis is all $\chi^2 > 12.592$. If the chi-square value calculated from the cross-classification table is less than 12.592, the null hypothesis is not rejected.

 From the SPSS printout, $\chi^2 = 11.025 < 12.592$ so the null hypothesis is not rejected. As a result, at the 0.05 level of significance there is insufficient relationship between age and labour force status in Table 3 to reject the null hypothesis. While the expected and actual counts differ in the various cells, these differences are not large enough to reject the null hypothesis.

 While the null hypothesis cannot be rejected at the 0.05 level of significance, it can be rejected at the 0.10 level of significance. For $\alpha = 0.10$ and $df = 6$, the critical region begins at $\chi^2 = 10.645$. From Table 4, the chi-square value from the cross-classification table, 11.025, exceeds this, so at the 0.10 level of significance, the null hypothesis can be rejected and the alternative hypothesis accepted.

From these results, there is some evidence for the author's claim that there is a relation between age and subsequent employment status. However, the relationship is not strong in that the relationship exists only at the 0.10 level of significance, but not at the 0.05 level.

The pattern of relationship, to the extent that it exists, can be seen by comparing the counts and expected counts. For the first row, there is little difference by age, with expected and actual counts about equal for each cell. However, for 'employed elsewhere,' the youngest group has a much larger count than expected while the oldest group has a smaller count than expected. This supports the argument that younger workers were more likely to find jobs elsewhere (43) would be expected (35) if there is no relationship between age and labour force status. In contrast, for older workers, there are fewer employed elsewhere (28) than would be expected if there is no relationship between age and labour force status (35.5). In terms of the status of being unemployed or not in the labour force, the relationship is reversed, with older workers being more likely than expected to experience these labour force statuses, while younger workers are less likely than expected to experience these.

In summary, there is a weak relationship between age and subsequent employment status, perhaps not as strong as the author claims.

(b) The sample proportions are $(3 + 9)/99 = 0.121$ for the youngest age group, $(7 + 11)/100 = 0.18$ for the middle aged group, and $(13 + 14)/100 = 0.27$ for the oldest group. For each group the method is the same – produce a 95% interval estimate. Each sample size is large enough to use the normal distribution for the distribution of $\hat{p}$, so $Z = 1.96$.

For the age group under 37, the 95% interval estimate for the true proportion who were unemployed or not in the labour force is:

$$\hat{p} \pm Z\sqrt{\frac{pq}{n}} \quad = \quad \hat{p} \pm 1.96\sqrt{\frac{0.121 \times 0.879}{99}}$$

$$
\begin{aligned}
&= \hat{p} \pm 1.96\sqrt{\frac{0.1064}{99}} \\
&= \hat{p} \pm 1.96\sqrt{0.0010743} \\
&= \hat{p} \pm (1.96 \times 0.03278) \\
&= \hat{p} \pm 0.0642 \\
&= 0.121 \pm 0.064
\end{aligned}
$$

or from 0.057 to 0.185.

For the 37-44 age group, the 95% interval estimate for the true proportion who were unemployed or not in the labour force is:

$$
\begin{aligned}
\hat{p} \pm Z\sqrt{\frac{pq}{n}} &= \hat{p} \pm 1.96\sqrt{\frac{0.18 \times 0.82}{100}} \\
&= \hat{p} \pm 1.96\sqrt{\frac{0.1476}{100}} \\
&= \hat{p} \pm 1.96\sqrt{0.001476} \\
&= \hat{p} \pm (1.96 \times 0.03842) \\
&= \hat{p} \pm 0.0753 \\
&= 0.180 \pm 0.075
\end{aligned}
$$

or from 0.105 to 0.255.

For those over age 45, the 95% interval estimate for the true proportion who were unemployed or not in the labour force is:

$$
\begin{aligned}
\hat{p} \pm Z\sqrt{\frac{pq}{n}} &= \hat{p} \pm 1.96\sqrt{\frac{0.27 \times 0.73}{100}} \\
&= \hat{p} \pm 1.96\sqrt{\frac{0.1971}{100}} \\
&= \hat{p} \pm 1.96\sqrt{0.001971} \\
&= \hat{p} \pm (1.96 \times 0.04440) \\
&= \hat{p} \pm 0.0870 \\
&= 0.270 \pm 0.087
\end{aligned}
$$

or from 0.183 to 0.357.

From the samples, the pattern of increased proportions without jobs (unemployed or not in the labour force) as the age group is older is fairly clear. However, the interval estimates for each group are fairly wide, with a considerable overlap of the intervals from one group to the next. This leads one to question whether the results are as strong as the author claims. The statement is correct about the sample proportions, but whether the same would hold if all members of the population are surveyed is less clear.