Social Studies 201
Fall 2003

Answers to Problem Set 5

1. **Use of media**

   (a) i. Let μ be the true mean minutes of media use daily by all Saskatchewan respondents aged 15-24. The sample mean is $\bar{X}$ and, with large sample size $n = 80$, the distribution of $\bar{X}$ is a normal distribution with mean μ and standard deviation $s/\sqrt{n}$. For the 96% interval estimate, $Z = 2.05$ or 2.06 – at this value of Z, there is $96/2 = 48\%$ of the area in each half of the centre of the normal distribution. The 96% interval estimate is

$$\begin{aligned}
\bar{X} \pm Z\frac{\sigma}{\sqrt{n}} &= 168 \pm 2.06\ \frac{128}{\sqrt{80}} \\
&= 168 \pm 2.06\ \frac{128}{8.944} \\
&= 168 \pm 2.06 \times 14.311 \\
&= 168 \pm 29.480
\end{aligned}$$

   or, to one decimal, (138.5 , 197.5).

   ii. For the 55-64 year olds, the method is the same. The 96% interval estimate for μ, the true mean minutes of media use daily, for this age group is

$$\begin{aligned}
\bar{X} \pm Z\frac{\sigma}{\sqrt{n}} &= 202 \pm 2.06\ \frac{160}{\sqrt{88}} \\
&= 202 \pm 2.06\ \frac{160}{9.381} \\
&= 202 \pm 2.06 \times 17.056 \\
&= 202 \pm 35.135
\end{aligned}$$

   or (166.9 , 237.1).

   While the mean number of minutes using media daily is 34 minutes greater for 55-64 year olds than for 15-24 year olds, as indicated by the respective standard deviations, the variation in use is also large for each group. The upper end of the 15-24 interval is 197 or 198 minutes daily while the lower end of the 55-64 interval is 167 minutes so the two interval estimates overlap rather considerably. As a result, there is no great assurance that the true mean is greater for the 55-64 year olds than for the 15-24 year olds.

(b) For all Saskatchewan adults, the method is agin the same. For 99% confidence level, there is $99/2 = 49.5\%$ or 0.4950 of the area on each side of centre – for this A area of the normal distribution, $Z = 0.2.575$. The 99% interval estimate for μ, the true mean minutes of media use daily, for all Saskatchewan adults is

$$\begin{aligned} \bar{X} \pm Z\frac{\sigma}{\sqrt{n}} &= 207 \pm 2.575 \, \frac{162}{\sqrt{631}} \\ &= 207 \pm 2.575 \, \frac{162}{25.120} \\ &= 207 \pm 2.575 \times 6.449 \\ &= 207 \pm 16.606 \end{aligned}$$

or (190.4 , 223.6). The 99% interval estimate for the true mean number of minutes of media use daily by all Saskatchewan adults if between 190 and 224 minutes.

This interval is narrower than either of the intervals in part a. There are three factors that affect interval width – confidence level or Z, standard deviation of the sample or population, and the sample size. In the case of the this interval estimate, the sample size is large with $n = 631$, as opposed to the smaller sample sizes of 80 and 88 for the other two age groups. The standard deviation is larger for the total population and the Z value is also larger, both factors that indicate a wider interval for the total population than for each of the two age groups. But the effect of the larger sample size in increasing the denominator of $Z(s/\sqrt{n})$, and hence reducing the size of this expression, dominates over these other two factors to produce a narrower interval for the total population.

2. **Stress levels**

(a) Let μ be the true mean stress levels for all undergraduates on the relation with parents issue. The sample mean $\bar{X}$ is the point estimate of μ and since the sample size for sample 1 is $n = 20$, the t-distribution must be used. The sample mean has a t-distribution with mean μ, standard deviation $s/\sqrt{n}$, and $n - 1 = 20 - 1 = 19$ degrees of freedom. For 19 d.f. and 98% confidence level, the t-value is 2.540. The 98% interval estimate is

$$\begin{aligned}\bar{X} \pm Z\frac{\sigma}{\sqrt{n}} &= 1.45 \pm 2.540\ \frac{0.76}{\sqrt{20}} \\ &= 1.45 \pm 2.540\ \frac{0.76}{4.472} \\ &= 1.45 \pm 2.540 \times 0.170 \\ &= 1.45 \pm 0.432\end{aligned}$$

or (1.018 , 1.882). On the relations with parents issue, the 98% interval estimate for the mean stress level for all undergraduates is from 1.0 to 1.9, a wide interval for a variable measured on a three-point scale.

(b) For each group, let μ be the true mean stress level for undergraduates on the academic performance issue. The question is to test whether this mean exceeds 2, so this is a one-tailed test in the positive direction and the hypotheses are:

$$\begin{aligned}H_0 &: \mu = 2 \\ H_1 &: \mu > 2\end{aligned}$$

For sample 1, the sample size is $n = 21$ and the test statistic is $\bar{X}$. Given the sample size of less than 30, $\bar{X}$ has a t-distribution with mean μ, standard deviation $s/\sqrt{n}$, and $d = n - 1 = 21 - 1 = 20$ degrees of freedom. At $\alpha = 0.05$ significance, the critical $t = 1.725$ and H_0 is rejected for all $t > +1.725$.

$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}}$$

$$\begin{aligned} &= \frac{2.24 - 2}{0.77/\sqrt{21}} \\ &= \frac{0.24}{0.168} \\ &= 1.428 \end{aligned}$$

Since this t-value is less than 1.725, the the null hypothesis cannot be rejected at the 0.05 level of significance. At the 0.05 level of significance the data do not provide evidence that the mean stress level of undergraduates exceeds the neutral response of 2.

For sample 2, the sample size is $n = 697$ 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 is used as an estimate of σ. For a normal distribution, a one tailed test at the $\alpha = 0.05$ significance level has a critical value of $Z = 1.645$. and H_0 is rejected for all $Z > +1.645$.

$$\begin{aligned} Z &= \frac{\bar{X} - \mu}{s/\sqrt{n}} \\ &= \frac{2.36 - 2}{0.66/\sqrt{697}} \\ &= \frac{0.36}{0.0250} \\ &= 14.400 \end{aligned}$$

Since this Z-value is much greater than 1.645, the the null hypothesis is rejected at the 0.05 level of significance. At the 0.05 level of significance sample 2 provides very strong evidence that the mean stress level of undergraduates exceeds the neutral response of 2.

The results of these two hypotheses tests differ primarily because of the difference in sample size. For the smaller sample size in sample 1, the numerical difference of 0.24 between 2 and 2.24 is actually less than the numerical difference of 0.36 between 2.36 and 2 for sample 2. But the smaller sample size of sample 1 produces a smaller t-value than the corresponding Z-value for sample 2. While the effect of the sample size is the dominating factor in comparison of these two tests, also note that the Z-value required to reject H_0 in the case of sample 2 is less than the corresponding t-value for sample 1. Together these two factors result in the null hypothesis that $\mu = 2$ being rejected for sample 2 but not for sample 1.

(c) For sample size,

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

Z=2.575 for 99% confidence, $E = 0.1$, and s can be used as an estimate for σ. Since the two questions have different standard deviations, in order to ensure that a large enough sample size is obtained, use the larger of the two standard deviations. Using these values,

$$n = \left(\frac{2.575 \times 0.77}{0.1}\right)^2 = 19.8275^2 = 393.1.$$

A random sample of $n = 394$ undergraduates would provide the required precision.

3. **Importance of issues**

   (a) Let p be the true proportion of adults in the 18-24 age group who identify spiritual issues as of increased importance. Since $n = 107$ is large ($n = 107 > 5/0.23 = 21.7$), the sample proportion $\hat{p}$ is normally distributed with mean p and standard deviation $\sqrt{pq/n}$. Since no confidence level is stated, select the 95% interval estimate and for this confidence level, $Z = \pm 1.96$. By using $p = q = 0.5$, the widest possible interval for any of the proportions is obtained. The interval estimates are:

$$\begin{aligned}
\hat{p} \pm Z\sqrt{\frac{pq}{n}} &= \hat{p} \pm 1.96\sqrt{\frac{0.5 \times 0.5}{107}} \\
&= \hat{p} \pm 1.96\sqrt{0.0023364} \\
&= \hat{p} \pm (1.96 \times 0.0483) \\
&= \hat{p} \pm 0.0947 \\
&= 0.23 \pm 0.0947
\end{aligned}$$

   or from 0.14 to 0.32. The 95% interval estimate for the 18-24 age group is from 14% to 32%.

   For the 35-44 age group, the method is the same and the sample size is even larger so $\hat{p}$ is normally distributed and $z = \pm 1.96$ for 95% confidence. The interval is

$$\begin{aligned}
\hat{p} \pm Z\sqrt{\frac{pq}{n}} &= \hat{p} \pm 1.96\sqrt{\frac{0.5 \times 0.5}{249}} \\
&= \hat{p} \pm 1.96\sqrt{0.001004} \\
&= \hat{p} \pm (1.96 \times 0.0317) \\
&= \hat{p} \pm 0.0621 \\
&= 0.31 \pm 0.0621
\end{aligned}$$

   or from 0.25 to 0.37. The 95% interval estimate for the 35-44 age group is from 25% to 37%.

   (b) Let p be the true proportion of 18-24 year olds who identify sex as of increased importance. The question is to test whether this proportion exceeds one-half, so this is a one-tailed test in the positive direction. As usual, the null hypothesis is an equality so the hypotheses are:

$$\begin{aligned} H_0 &: p = 0.5 \\ H_1 &: p > 0.5 \end{aligned}$$

The sample size is $n = 107$ and the test statistic is $\hat{p}$. Given that $\hat{p} = 0.61$, $\hat{q} = 1 - \hat{p} = 1 - 0.61 = 0.39$. To ensure this is a large sample size, n divided by the smaller of p or q gives $5/0.39 = 12.8$ and $n = 107 = 13$ so this is a large sample size and $\hat{p}$ has a normal distribution with mean p and standard deviation $\sqrt{pq/n}$. At $\alpha = 0.05$ significance, with a one-tailed test, the critical $Z = 1.645$ and H_0 is rejected for all $Z > +1.645$.

$$\begin{aligned} Z &= \frac{\hat{p} - p}{\sqrt{pq/n}} \\ &= \frac{0.61 - 0.50}{\sqrt{(0.5 \times 0.5)/107}} \\ &= \frac{0.11}{0.0483} \\ &= 2.276 \end{aligned}$$

Since this Z-value is greater than 1.645, the the null hypothesis is rejected at the 0.05 level of significance. At the 0.05 level of significance the data provide evidence that the proportion of those aged 18-24 who identify sex as of increased importance exceeds one-half.

(c) For estimation of a proportion,

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

For 18 in 20 times, this is 90% confidence and $Z = 1.645$. From the question the precision required is ± 5 percentage points so

$E = 0.05$, and the largest possible sample size for this Z and E are given by $p = q = 0.5$. The required sample size is

$$n = \left(\frac{Z}{E}\right)^2 pq = \left(\frac{1.645}{0.05}\right)^2 \times 0.5 \times 0.5 = 32.9^2 \times 0.25 = 270.60$$

The required sample size is $n = 271$.

(d) The percentages in the table support Bibby's contentions – sex and intellectual growth are reported as much less important for the older group, while spiritual issues are more important, all as compared to the younger age group.

The hypothesis test of part b also supports the contention of Bibby. The evidence is that for 18-24 year olds, the proportion who view sex as of increased importance exceeds one-half, while it is much less than one-half for the age 35-44 group. While no test was conducted for the latter group, the 25% who viewed sex as of increased importance would certainly be evidence that this was less than one-half.

For spiritual issues though, it is not clear that Bibby's argument is correct. The percentage of the younger age group that rated spiritual issues as of increased importance was 23%, as opposed to 31% for the older group. But the two 95% interval estimates overlap considerably – for the lower age group the interval is 14 to 32 per cent, while for the older age group it is 25 to 37 per cent. Since these intervals overlap so much, and the true percentage for each group could be in the overlapping region, it is not clear that older people rate spiritual issues more highly than do younger people.

Finally, the question Bibby appears to have asked may bias the results somewhat. The question is not which issues are important but which issues have become more important. Since there is no indication of what the base importance of these issues is, it is not too clear exactly what it means to say issues are of increased importance.

In conclusion, it seems that Bibby may be correct about physical and intellectual issues but the conclusion concerning spiritual issues is questionable.