Social Studies 201
Answers to Questions 1-3 of Problem Set 4
March 16, 2001

1. For this questions, success is selection of a Liberal party supporter; the probability of this is $p = 0.52$ and $n = 28$. The probability q of not selecting a Liberal party supporter is $q = 1 - p = 1 - 0.52 = 0.48$. The mean and standard deviation of this binomial are $\mu = np = 28 \times 0.52 = 14.56$ and $\sigma = \sqrt{npq} = \sqrt{28 \times 0.52 \times 0.48} = \sqrt{6.9888} = 2.644$.

   For part (i), the probability of selecting less than 10 Liberal party supporters, the appropriate area under the same normal curve is the area to the left of $X = 9.5$ and this is the area to the left of

$$Z = \frac{X - \mu}{\sigma} = \frac{X - np}{\sqrt{npq}} = \frac{9.5 - 14.56}{2.644} = \frac{-5.16}{2.644} = -1.95.$$

   The area to the left of $Z = -1.95$ is 0.0256 and this is the probability of selecting less than 10 Liberal party supporters in this sample.

   From the MINITAB printout, the exact probability is the sum of the probabilities of 0 through 9 successes. This is $0.0002 + 0.0007 + 0.0025 + 0.0070 + 0.0169 = 0.0273$.

   The approximation provides a low estimate of the exact probability, 0.0256 as opposed to the exact probability of 0.0273. The difference is $0.0256 - 0.0273 = 0.0017$ or 17 out of 10,000. This is not a large difference so the approximation provides a relatively good estimate of the exact probability. Note that $n = 28$, exceeding the minimum acceptable value for n of $5/\min(p, q) = 5/0.48 = 10.4$. The approximation would be closer if n were larger and the reason for a difference between the approximation and the correct value from MINITAB is that $n = 28$ is not too much larger than the minimum of $n = 10.4$.

2. The question askes for the probabilities of various numbers of smokers in a random sample of 8 Saskatchewan adults, 25% of which are smokers. In this problem, success is defined as selecting a smoker and the probability of this is $p = 0.25$. The probability of failure, selecting someone who is not a smoker, is $q = 1 - p = 1 - 0.25 = 0.75$. Since the sample size is 8, for this binomial, $n = 8$.

The mean and standard deviation of this binomial are $\mu = np = 8 \times 0.25 = 2$ and $\sigma = \sqrt{npq} = \sqrt{8 \times 0.25 \times 0.75} = \sqrt{1.5} = 1.225$.

When using the normal approximation to the binomial, the probability of selecting less than 3 smokers is the area under the normal curve which lies to the left of $X = 2.5$. The Z value for $X = 2.5$ is

$$Z = \frac{X - \mu}{\sigma} = \frac{X - np}{\sqrt{npq}} = \frac{2.5 - 2}{1.225} = \frac{0.5}{1.225} = 0.41.$$

The area to the left of $Z = 0.41$ is the 0.5000 of the area to the left of the mean plus the A area of 0.1591 between the mean and $Z = 0.41$. Thus the probability of selecting less than 3 smokers is $0.5000 + 0.1591 = 0.6591$.

Using the normal approximation to the binomial, the probability of selecting 6 or more smokers is the area under the same normal distribution that lies to the right of $X = 5.5$. For this value of X, the appropriate Z is

$$Z = \frac{X - \mu}{\sigma} = \frac{X - np}{\sqrt{npq}} = \frac{5.5 - 2}{1.225} = \frac{3.5}{1.225} = 2.86$$

and the B area to the right of this is 0.0021. The probability of selecting 6 or more smokers in a random sample of size 8 from a population with 25% smokers is approximately 21 in 10,000.

Since there is almost no probability of selecting more than 6 smokers, it is extremely unlikely to select 8 smokers in a sample of size 8 from a population with 25% smokers. There are several reasons why the binomial would not explain the situation described. First, since smoking is not permitted in many buildings, smokers who wish to smoke often congregate outside a building to smoke. As a result, those who gather there to smoke do not represent a random sample of people in the office building, let alone a random sample of adults in Saskatchewan. That is, the probability of selecting a smoker is not 0.25 for those outside the building but may be closer to 1. Second, it is unlikely that there is independence of trials, as required by the binomial. Random selection of people from a large population is generally regarded as leading to independence from one trial to the next. But the individuals gathering

outside the building cannot be regarded as independent – each may be a smoker, and they gather together to smoke, so there is a certain dependence involved. If one goes out to smoke, others may accompany him or her. Third, the question of how closely the normal approximation is to the actual binomial probabilities also emerges here. Since $5/\min(p,q) = 5/0.25 = 20$ and $n = 8$, the small n does not satisfy the conditions for using the normal approximation to the binomial. However, from MINITAB, the exact probability for less than 3 is 0.6786, not too different from the approximation of 0.6591. And the exact probability of more than 6 smokers in 0.0004, not too different from 0.0001. So the while the normal approximation is not real close in this case, it is close enough to provide a rough estimate of the correct probabilities, and the use of the normal approximation is only a minor reason for the problem noted in the question.

3. (a) Let μ be the true mean income for all Saskatchewan adults with a bachelors degree. A sample size of $n = 114$ cases is large, so the sample mean $\bar{X}$ has a normal distribution with mean μ and standard deviation $\sigma/\sqrt{n}$. Note that s can be used as an estimate of σ here because $n = 114$ is large. For 95% confidence, the middle 95% of the normal distribution is between $Z = -1.96$ and $Z = +1.96$. The 95% interval estimate is

$$
\begin{aligned}
\bar{X} \pm Z\frac{\sigma}{\sqrt{n}} &= 35.24 \pm 1.96\,\frac{24.24}{\sqrt{114}} \\
&= 35.24 \pm 1.96\,\frac{24.24}{10.68} \\
&= 35.24 \pm 1.96 \times 2.27 \\
&= 35.24 \pm 4.45
\end{aligned}
$$

or (30.79 , 39.69) or rounded to the nearest hundred dollars, $30,800 to $39,700.

For those with a diploma or certificate from a community college, the method is the same. The 95% interval estimate is

$$
\begin{aligned}
\bar{X} \pm Z\frac{\sigma}{\sqrt{n}} &= 26.47 \pm 1.96\,\frac{19.79}{\sqrt{80}} \\
&= 26.47 \pm 1.96\,\frac{19.79}{8.94}
\end{aligned}
$$

$$
\begin{aligned}
&= 26.47 \pm 1.96 \times 2.21 \\
&= 26.47 \pm 4.34
\end{aligned}
$$

or (22.13 , 30.81) or rounded to the nearest hundred dollars, $22,100 to $30,800.

(b) For those in categories 5 and 8, the method is again the same but the intervals are 90% intervals so that the appropriate Z value for the middle 90% of the normal distribution is $Z = 1.645$. The 90% interval estimate for the true mean income of those with some university is

$$
\begin{aligned}
\bar{X} \pm Z \frac{\sigma}{\sqrt{n}} &= 21.70 \pm 1.645 \frac{19.38}{\sqrt{78}} \\
&= 21.70 \pm 1.645 \frac{19.38}{8.83} \\
&= 21.70 \pm 1.645 \times 2.19 \\
&= 21.70 \pm 3.61
\end{aligned}
$$

or (18.09 , 25.31) or rounded to the nearest hundred dollars, $18,100 to $25,300.

The 90% interval estimate for the true mean income of those with no more than a high school diploma is

$$
\begin{aligned}
\bar{X} \pm Z \frac{\sigma}{\sqrt{n}} &= 20.59 \pm 1.645 \frac{17.36}{\sqrt{145}} \\
&= 20.59 \pm 1.645 \frac{17.36}{12.04} \\
&= 20.59 \pm 1.645 \times 1.44 \\
&= 20.59 \pm 2.37
\end{aligned}
$$

or (18.22 , 22.96) or rounded to the nearest hundred dollars, $18,200 to $23,000.

(c) The true mean could be anywhere within the intervals obtained in (a) and (b). Since these are 95% and 90% confidence intervals, the intervals are very likely to contain the true means, but there is always some chance that these are among the few (5% or 10%) of intervals that do not contain the true mean.

From (a) it appears that the true mean income for Saskatchewan adults with a bachelors degree or with a diploma or certificate is within approximately $4,400-4,500 of the estimated mean from the sample, and this is the case in 95% of the samples. The particular intervals barely overlap, so that a researcher can be relatively certain that the true mean income for those with a diploma or certificate is somewhat less than for Saskatchewan adults with a diploma or certificate. That is, while the intervals just touch each other, the researcher is reasonably confident that the true mean for those with a bachelors degree is between $30,800 and $39,700. For those with a diploma or certificate, the true mean is very likely somewhere between $22,100 and $30,800. So it is very likely that the true mean income for those with a bachelors degree exceeds the true mean income for those with a diploma or certificate.

For those with some university and those with only a high school diploma the situation is quite different. Here the two intervals overlap so much that a researcher could not be very confident that all Saskatchewan adults with only some university have higher incomes than those with a high school diploma only.