Section 7.3

Estimating *p* in the Binomial Distribution

Estimating p in the Binomial Distribution

Procedure:

HOW TO FIND A CONFIDENCE INTERVAL FOR A PROPORTION p

Requirements

Consider a binomial experiment with n trials, where p represents the population probability of success on a single trial and $q = 1 - p$ represents the population probability of failure. Let r be a random variable that represents the number of successes out of the n binomial trials.

The point estimates for p and q are

$$\hat{p} = \frac{r}{n} \qquad \text{and} \qquad \hat{q} = 1 - \hat{p}$$

The number of trials n should be sufficiently large so that both $n\hat{p} > 5$ and $n\hat{q} > 5$.

Confidence interval for p

$$\hat{p} - E < p < \hat{p} + E$$

where $E \approx z_c\sqrt{\frac{\hat{p}\hat{q}}{n}} = z_c\sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}$

c = confidence level ($0 < c < 1$)

z_c = critical value for confidence level c based on the standard normal distribution (See Table 5(b) of Appendix II for frequently used values.)

Example 6 – *Confidence Interval for p*

Let's return to our flu shot experiment described at the beginning of this section.

Suppose that 800 students were selected at random from a student body of 20,000 and given shots to prevent a certain type of flu.

All 800 students were exposed to the flu, and 600 of them did not get the flu.

Let p represent the probability that the shot will be successful for any single student selected at random from the entire population of 20,000. Let q be the probability that the shot is not successful.

Example 6(a) – *Confidence Interval for p* cont'd

What is the number of trials *n*? What is the value of *r*?

Solution:

Since each of the 800 students receiving the shot may be thought of as a trial, then $n = 800$, and $r = 600$ is the number of successful trials.

Example 6(b) – *Confidence Interval for p* cont' d

What are the point estimates for *p* and *q*?

Solution:

We estimate *p* by the sample point estimate

$$\hat{p} = \frac{r}{n} = \frac{600}{800} = 0.75$$

We estimate *q* by

$$\hat{q} = 1 - \hat{p} = 1 - 0.75 = 0.25$$

Example 6(c) – *Confidence Interval for p* cont'd

Check Requirements Would it seem that the number of trials is large enough to justify a normal approximation to the binomial?

Solution:

Since $n = 800$, $p \approx 0.75$, and $q \approx 0.25$, then

$np \approx (800)(0.75) = 600 > 5$ and $np \approx (800)(0.25) = 200 > 5$

A normal approximation is certainly justified.

Example 6(d) – *Confidence Interval for p* cont'd

Find a 99% confidence interval for *p*.

Solution:

$z_{0.99}$ = 2.58 (Table 5(b) of Appendix II)

$$E \approx z_{0.99}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$\approx 2.58\sqrt{\frac{(0.75)(0.25)}{800}}$$

$$\approx 0.0395$$

Example 6(d) – *Solution*

The 99% confidence interval is then

$$\hat{p} - E < p < \hat{p} + E$$

$$0.75 - 0.0395 < p < 0.75 + 0.0395$$

$$0.71 < p < 0.79$$

Interpretation We are 99% confident that the probability a flu shot will be effective for a student selected at random is between 0.71 and 0.79.

Interpreting Results from a Poll

Interpreting Results from a Poll

Newspapers frequently report the results of an opinion poll. In articles that give more information, a statement about the margin of error accompanies the poll results.

In most polls, the margin of error is given for a *95% confidence interval*.

General interpretation of poll results

1. When a poll states the results of a survey, the proportion reported to respond in the designated manner is $\hat{p}$, the sample estimate of the population proportion.
2. The *margin of error* is the maximal error E of a 95% confidence interval for p.
3. A 95% confidence interval for the population proportion p is

$$\text{poll report } \hat{p} - \text{margin of error } E < p < \text{poll report } \hat{p} + \text{margin of error } E$$

Sample Size for Estimating p

Sample Size for Estimating p

Suppose you want to specify the maximal margin of error in advance for a confidence interval for p at a given confidence level c.

What sample size do you need?

The answer depends on whether or not you have a preliminary estimate for the population probability of success p in a binomial distribution.

Sample Size for Estimating p

Procedure:

HOW TO FIND THE SAMPLE SIZE n FOR ESTIMATING A PROPORTION p

$$n = p(1 - p)\left(\frac{z_c}{E}\right)^2 \text{ if you have a preliminary estimate for } p \quad (21)$$

$$n = \frac{1}{4}\left(\frac{z_c}{E}\right)^2 \text{ if you do } \textit{not} \text{ have a preliminary estimate for } p \quad (22)$$

where E = specified maximal error of estimate

z_c = critical value from the normal distribution for the desired confidence level c. Commonly used value of z_c can be found in Table 5(b) of Appendix II.

If n is not a whole number, increase n to the next higher whole number. Also, if necessary, increase the sample size n to ensure that both $np > 5$ and $nq > 5$. Note that n is the minimal sample size for a specified confidence level and maximal error of estimate.

Example 7 – *Sample Size for Estimating p*

A company is in the business of selling wholesale popcorn to grocery stores. The company buys directly from farmers.

A buyer for the company is examining a large amount of corn from a certain farmer.

Before the purchase is made, the buyer wants to estimate p, the probability that a kernel will pop.

Suppose a random sample of n kernels is taken and r of these kernels pop.

Example 7 – *Sample Size for Estimating p* cont'd

The buyer wants to be 95% sure that the point estimate $\hat{p}$ = *r*/*n* for *p* will be in error either way by less than 0.01.

a. If no preliminary study is made to estimate *p*, how large a sample should the buyer use?

Example 7 – *Solution*

cont'd

In this case, we use Equation (22) with $z_{0.95} = 1.96$ (see Table 7-2) and $E = 0.01$.

Level of Confidence c	Critical Value z_c
0.70, or 70%	1.04
0.75, or 75%	1.15
0.80, or 80%	1.28
0.85, or 85%	1.44
0.90, or 90%	1.645
0.95, or 95%	1.96
0.98, or 98%	2.33
0.99, or 99%	2.58

Some Levels of Confidence and Their Corresponding Critical Values

Table 7-2

Example 7 – *Solution*

cont'd

$$n = \frac{1}{4}\left(\frac{z_c}{E}\right)^2$$

$$= \frac{1}{4}\left(\frac{1.96}{0.01}\right)^2$$

$$= 0.25(38{,}416)$$

$$= 9604$$

The buyer would need a sample of n = 9604 kernels.

Example 7 – *Sample Size for Estimating p* cont'd

(b) A preliminary study showed that p was approximately 0.86. If the buyer uses the results of the preliminary study, how large a sample should he use?

Solution:

In this case, we use Equation (21) with $p \approx 0.86$.

Again, from Table 7-2, $z_{0.95} = 1.96$, and from the problem, $E = 0.01$.

$$n = p(1 - p)\left(\frac{z_c}{E}\right)^2$$

Example 7 – *Solution*

cont'd

$$= (0.86)(0.14)\left(\frac{1.96}{0.01}\right)^2$$

$$= 4625.29$$

The sample size should be at least $n = 4626$. This sample is less than half the sample size necessary without the preliminary study.