VIII. Confidence Intervals: Proportions

Scenario. You will have category or attribute sample data from a population. The observations on the sample will have been taken only once.

Data and parameters. You will be given sample data which estimate unknown population parameters.

	Parameters (unknown)	Sample estimates (known)
"successes"	–	$\boldsymbol{k}$
sample size	–	$\boldsymbol{n}$
proportion	$\boldsymbol{p}$	$\hat{\boldsymbol{p}} = \dfrac{\boldsymbol{k}}{\boldsymbol{n}}$

The "success" category is determined by the focus of the research, not necessarily by a conventional definition of "success."

Research Objective. To place error bounds $\pm$ ERROR about the sample proportion in such a way that the resulting *confidence interval*

$$\hat{\boldsymbol{p}} - \texttt{ERROR} \quad \text{to} \quad \hat{\boldsymbol{p}} + \texttt{ERROR}$$

has a high likelihood of including the true population parameter $\boldsymbol{p}$. The likelihood that the confidence interval

$$\hat{\boldsymbol{p}} - \texttt{ERROR} \quad \text{to} \quad \hat{\boldsymbol{p}} + \texttt{ERROR}$$

contains $\boldsymbol{p}$ is called the *confidence level* of the interval.

Solution Template

Step 1. Make a dictionary of the data given in the problem:

"successes"	$\boldsymbol{k}$
sample size	$\boldsymbol{n}$
sample proportion	$\hat{\boldsymbol{p}} = \dfrac{\boldsymbol{k}}{\boldsymbol{n}}$
confidence level	$\boldsymbol{C}$

The formulae below will require that both $\boldsymbol{n}\hat{\boldsymbol{p}}$ and $\boldsymbol{n}(1-\hat{\boldsymbol{p}})$ be at least five; you should check this before proceeding. Also, *be sure that you have expressed the sample proportion as a proportion and not as a percent* as the formulae require proportions.

Step 2. Find the confidence level in the last column in the table in Appendix B. Then look in the fourth column to find the confidence limits $\pm \boldsymbol{z_c}$.

Step 3. Apply the following formula to find the value of the `ERROR` term:

$$\mathtt{ERROR} = \pm \boldsymbol{z_c}\sqrt{\frac{\hat{\boldsymbol{p}}(1-\hat{\boldsymbol{p}})}{\boldsymbol{n}}}$$

using the dictionary from step one and the confidence limits from step 2. *Make sure that you have expressed the sample proportion as a proportion and not as a percent* when you use this formula.

Step 4. Write down the resulting confidence interval:

$$\hat{\boldsymbol{p}} - \mathtt{ERROR} \quad \text{to} \quad \hat{\boldsymbol{p}} + \mathtt{ERROR}.$$

Step 5. (Optional) Express the confidence interval as percents. Do this only if you are more comfortable thinking in terms of percents than proportions. This will make your confidence limits

$$100\% \times (\hat{\boldsymbol{p}} - \mathtt{ERROR}) \quad \text{to} \quad 100\% \times (\hat{\boldsymbol{p}} + \mathtt{ERROR}).$$

End of Solution Template

Interpretation. You expect that the interval

$$\hat{\boldsymbol{p}} - \texttt{ERROR} \quad \text{to} \quad \hat{\boldsymbol{p}} + \texttt{ERROR}$$

will contain the true value of the parameter $\boldsymbol{p}$. You would be surprised if it were to turn out that the value of $\boldsymbol{p}$ were outside of this interval. The confidence level measures the degree to which you expect the interval will contain the true parameter (or the degree to which you would be surprised if the true parameter were not within the interval).

The researcher is free to select whatever confidence level is appropriate for the problem being studied. However, it is conventional to consider only confidence levels of 90% or higher. Since you believe that the interval contains $\boldsymbol{p}$, the likelihood that this belief is in error is

$$100\% - \text{confidence level.}$$

Thus, for example, an 80% confidence interval results in a 20% chance of error; a chance of error this high is almost always unacceptable.

Assumptions. You must assume that $\boldsymbol{n}\hat{\boldsymbol{p}}$ and $\boldsymbol{n}(1 - \hat{\boldsymbol{p}})$ are both at least five.

Example. An attorney hopes to file a class action suit against a hospital for sexual harassment. In order to develop evidence, the attorney randomly selects 150 nurses in the hospital and conducts a telephone survey. The survey begins by informing the respondents of the purpose of the survey and then asks the question

"Have you been a victim of sexism during work hours at the hospital?"

Of the 150 respondents, 127 answer "yes" to the above question.

Construct an 80% confidence interval for the proportion of nurses in the hospital who would respond "yes" to the above question.

Solution.

Step 1. Our dictionary of values consists of

"successes"	$\boldsymbol{k}$	127
sample size	$\boldsymbol{n}$	150
sample proportion	$\hat{\boldsymbol{p}} = \frac{\boldsymbol{k}}{\boldsymbol{n}}$	$\frac{127}{150} = 0.847$
confidence level	C	80%

Note that we are focusing on "sexism" in our study, so "success" constitutes being a victim of sexism. Since

$$\boldsymbol{n\hat{p}} = 150 \times 0.847 = 127$$

and

$$\boldsymbol{n}(1 - \boldsymbol{\hat{p}}) = 150 \times 0.153 = 22$$

we can proceed with the method.

$\boxed{\textit{Step 2.}}$ The confidence limits are ± 1.281 from the table in Appendix B.

$\boxed{\textit{Step 3.}}$ The error term is

$$\begin{aligned}
\texttt{ERROR} &= \pm \boldsymbol{z_c}\sqrt{\frac{\boldsymbol{\hat{p}}(1-\boldsymbol{\hat{p}})}{\boldsymbol{n}}} \\
&= \pm 1.281\sqrt{\frac{0.847(1-0.847)}{150}} \\
&= \pm 1.281\sqrt{\frac{0.130}{150}} \\
&= \pm 1.281(0.294) \\
&= \pm 0.0377
\end{aligned}$$

$\boxed{\textit{Step 4.}}$ The confidence interval is thus

$$\begin{array}{rcl}
0.847 - 0.377 & \text{to} & 0.847 + 0.377 \\
 & \text{or} & \\
0.809 & \text{to} & 0.885
\end{array}.$$

$\boxed{\textit{Step 5.}}$ Converting back to percents, our confidence interval is 80.9% to 88.5% ▌

Question. Suppose that you were on a jury trying to decide if female employees in the hospital we subject to on-the-job sexual harassment. Would you find the above evidence convincing? List at least four ways to improve the research techniques.

Problems

1. In a sample of size 312 there were 87 successes. Find a 99% confidence interval for the population proportion.

2. In a sample of size 624 there were 174 successes. Find a 99% confidence interval for the population proportion.

3. In a sample of size 1,391 there were 568 successes. Find a 96% confidence interval for the population proportion.

4. In a sample of 1,864 users of extended wear contact lenses, it is found that 112 had experienced an eye infection since using the lenses.

(i) Construct a 99% confidence interval for the proportion of extended wear contact lens users experiencing eye infections.

(ii) Among all contact lens users, only 2% have ever experienced an eye infection. Has researcher gathered convincing evidence that extended wear users are more likely to suffer an eye infection? Has the researcher gathered evidence that users of extended wear contact lenses are more than three times as likely to experience eye infections than other contact lens users?

5. In a sample of 1,783 parolees, it is found that 268 commit another felony during their parole period.

(i) Construct a 95% confidence interval for the proportion of "failed" parolees – those who commit a felony during their parole period.

(ii) Among convicted felons who do not receive parole, it is known that 42% will eventually commit another felony. Is this convincing evidence that parole programs are more effective at reducing recidivism than requiring prisoners to complete their entire sentence?

Video Assignment.

View the following program(s) from the series *Against All Odds*:

Program	*Title*
19	*Confidence Intervals*