Confidence Intervals for Population Proportions – Example

Evan L. Ray

November 6, 2017

Teenage Drivers

An insurance company checks police records on 582 accidents that occurred in the US, selected at random, and notes that teenagers were at the wheel in 91 of them.

(a) What is the population parameter, p? Describe it in a sentence.

The population parameter is the proportion of all accidents in the US that involved teenage drivers.

(b) What is the sample statistic, $\widehat{p}$? Describe it in a sentence and calculate its value.

The sample statistic is the proportion of accidents in this sample that involved teenage drivers.

$\hat{p} = \frac{91}{582} = 0.156$

(c) What is the standard error of $\widehat{p}$? Describe it in a sentence and calculate its value.

The standard error of $\hat{p}$ is the estimated standard deviation of sample proportions that you would obtain across all possible samples of size 582.

$$\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0.156(1-0.156)}{582}} = 0.015$$

(d) What is the sampling distribution of $\widehat{p}$? Describe the meaning of a sampling distribution in a sentence or two, and write down the sampling distribution of $\widehat{p}$ for this particular example.

The sampling distribution of $\hat{p}$ is the distribution of all values of the sample proportion that you could get from all possible samples of size 582.

Since there were $91 \geq 10$ "successes" and $491 \geq 10$ "failures" in this data set we can approximate the sampling distribution with a normal model:

$$\hat{p} \sim \text{Normal}\left(p, \sqrt{\frac{p(1-p)}{n}}\right).$$

Note that we should also really check that there are 2 outcomes (teen driver or not), each accident has the same probability of involving a teen driver, and the accidents in our sample are independent from each other.

(e) **Calculate a 90% confidence interval for $\hat{p}$. You may use the following output from R:**

```
qnorm(0.9, mean = 0, sd = 1)

## [1] 1.281552

qnorm(0.95, mean = 0, sd = 1)

## [1] 1.644854

qnorm(0.975, mean = 0, sd = 1)

## [1] 1.959964

pnorm(0.9, mean = 0, sd = 1)

## [1] 0.8159399

pnorm(0.95, mean = 0, sd = 1)

## [1] 0.8289439

pnorm(0.975, mean = 0, sd = 1)

## [1] 0.8352199
```

Our confidence interval has the form

$$(\hat{p} - z^* \cdot SE(\hat{p}),\ \hat{p} + z^* \cdot SE(\hat{p})).$$

We found $\hat{p} = 0.156$ in part (b) and $SE(\hat{p}) = 0.015$ in part (c).

For a 90% CI, z^* is the 95th percentile of a standard normal distribution. This is the output from qnorm

$$(0.156 - 1.645 \cdot 0.015,\ 0.156 + 1.645 \cdot 0.015)$$

$$= (0.131,\ 0.181)$$

(f) **Interpret the interval you obtained in part (e) in the context of this example.**

We are 90% confident that the proportion of all US accidents involving teenage drivers is between 0.131 and 0.181. If we were to take many different samples of 582 accidents and use this procedure to calculate a different confidence interval based on the data in each of those samples, about 90% of those confidence intervals would contain the actual proportion of all US accidents that involve teenage drivers.

(g) **Would a 95% confidence interval for $\hat{p}$ be wider or narrower than the 90% interval you obtained in part (e)?**

A 95% confidence interval would be wider than the 90% interval from part (e).