Chapter 19

Estimating with Confidence:

Suppose I want to know what proportion of teenagers typically goes to the movies on a Friday night.

Suppose I take an SRS of 25 teenagers and calculate the sample proportion to be $\hat{p} = 0.40$.

The sample proportion $\hat{p}$ is an unbiased estimator of the unknown population proportion p, so I would estimate the population proportion to be approximately 40%. However, using a different sample would have given a different sample proportion, so I must consider the amount of variation in the sampling model for $\hat{p}$.

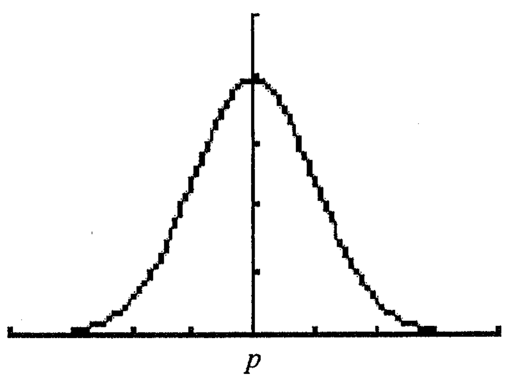
Based on one sample, it would NOT be correct to conclude that 40% of all teenagers typically go to the movies on a Friday night.

But don't despair!… based on my one sample, I can come up with an interval that **may** contain the true proportion of teenagers who typically go to the movies on a Friday night.

Not only will I tell you what that interval is, but I will also tell you how confident I am that the true proportion falls somewhere in that interval.

Remember…

- The sampling model for $\hat{p}$ is approx normal. assuming $np \geq 10$ and $nq \geq 10$.
- The mean of the sampling model is p.
- The standard deviation of the sampling model is $\sqrt{\frac{pq}{n}}$ assuming the population size is at least 10 times larger than the sample size $(N \geq 10n)$.



Since we don't know p, we cannot calculate the standard deviation of the sampling model. We can, however, use $\hat{p}$ to estimate the value of p and calculate the standard error $\sqrt{\frac{\hat{p}\hat{q}}{n}}$ instead.

So the standard error for the sampling model for the proportion of teenagers who typically go to the movies on a Friday night is:

$$\sqrt{\frac{(.4)(.6)}{25}} = .0979$$

According to the 68-95-99.7 Rule, 95% of all possible samples of size 25 will produce a statistic $\hat{p}$ that is within 2 standard errors of the mean of our sampling model.

This means that, in our example, 95% of the $\hat{p}$'s will be between $p - .196$ and $p + .196$.

So the distance between the actual p value and the statistic $\hat{p}$ will usually (95% of the time) be less than or equal to .196.

Therefore, in 95% of our samples, the interval between $\hat{p} - \cancel{0.11}$ .196 and $\hat{p} + \cancel{0.11}$ .196 will contain the parameter p.

We say that the margin of error is ~~0.11~~ .196

For our sample of 25 teenagers, $\hat{p} = 0.40$. Because the margin of error is ~~0.11~~ .196, then we are 95% confident that the true population proportion lies somewhere in the interval $.40 \pm .196$, or [.21, .59].

The interval [~~0.29, 0.51~~ .21, .59] is called a 95% confidence interval because we are 95% Confident that the true proportion of teenagers who typically go to the movies on a Friday night is between about 21% and 59%

CAUTION!!

This does **NOT** mean the probability that p is between ~~0.29~~ and ~~0.51~~ is 95%. Confidence does not mean the same thing as probability.

We CANNOT calculate the probability that p is within a given interval without using a Normal model, and we CANNOT draw a Normal model because we don't know the center, p.

If you assume that $p = \hat{p}$, then $P(x_1 \le p \le x_2)$ is either 0 or 1.

Okay… maybe you're not satisfied with the interval we constructed. Is it too wide? Would you prefer a more precise conclusion? One way of changing the length of the interval is to change the Confidence level.

So how do we construct 90% confidence intervals? 99% confidence intervals? C% confidence intervals?

Since the sampling model of the sample proportion $\hat{p}$ is approx normal, we can use normal calculations to construct confidence intervals.

- For a 95% confidence interval, we want the interval corresponding to the middle 95% of the normal curve.
- For a 90% confidence interval, we want the interval corresponding to the middle 90% of the normal curve.
- And so on…

If we are using the standard normal curve, we want to find the interval using z-values

Suppose we want to find a 90% confidence interval for a standard normal curve. If the middle 90% lies within our interval, then the remaining 10% lies outside our interval. Because the curve is symmetric, there is 5% below the interval and 5% above the interval. Find the z-values with area 5% below and 5% above.

These *z*-**values** are denoted $\pm z^*$. Because they come from the standard normal curve, they are centered at mean 0.

z^* is called the upper p critical value with probability p lying to its right under the standard normal curve.

To find the upper critical p value, we find the complement of C and divide it in half, or find:

$$\frac{1-C}{2}$$

For a 95% confidence interval, we want the *z*-**values** with upper p critical value ~~2.5%~~ **1.96**.

invNorm (.025) =

For a 99% confidence interval, we want the *z*-**values** with upper p critical value ~~0.5%~~ **2.57**.

invNorm (.005) =

Remember that **z-values** tell us how many standard deviations we are above or below the mean.

To construct a 95% confidence interval, we want to find the values 1.96 standard deviations below the mean and 1.96 standard deviations above the mean, or:

$$p \pm 1.96\sqrt{\frac{pq}{n}}$$

Using our sample data, this is $\hat{p} \pm 1.96\sqrt{\frac{\hat{p}\hat{q}}{n}}$, assuming the population is at least 10 times as large as the sample size, n.

In general, to construct a level C confidence interval using our sample data, we want to find:

$$\hat{p} \pm z^*\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

The margin of error is $z^*\sqrt{\frac{\hat{p}\hat{q}}{n}}$. Note that the margin of error is a positive number. **<u>It is not an interval</u>**.

We would like high confidence and a small margin of error.

A higher confidence level means a higher percentage of all samples produce a statistic close to the true value of the parameter. Therefore we want a high level of confidence.

A smaller margin of error allows us to get closer to the true value of the parameter (length of the interval is small), so we want a small margin of error.

So how do we reduce the margin of error?

- Lower the confidence level (by decreasing the value of z^*)
- Lower the standard deviation
- Increase the sample size. To cut the margin of error in half, increase the sample size by four times the previous size.

You can have high confidence and a small margin of error if you choose the right sample size.

To determine the sample size n that will yield a confidence interval for a population proportion with a specified margin of error m, set the expression for the margin of error to be equal to m and solve for n. Always round n up to the next greatest integer.

$$m = z^* \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

CAUTION!!

These methods only apply to certain situations. In order to construct a level C confidence interval using the formula $\hat{p} \pm z^* \sqrt{\frac{\hat{p}\hat{q}}{n}}$, for example, the data must come from a random sample. Also, we want to eliminate (if possible) any outliers.

The margin of error only covers random sampling errors. Things like undercoverage, nonresponse, and poor sampling designs can cause additional errors.

If you are asked to find a C.I. you must PANIC!

P

A

N

I

C