Chapter 19 — Inference about a proportion.

Ex: Young Adults Living with their Parents

A random sample of $n_1 = 2253$ men (age 19-25)
and $n_2 = 2629$ women (age 19-25)

found that 44% of the men & 35% of the women lived with their parents.

Let $\hat{p}$ be the sample proportion.

Then $$\frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$$ is approximately $N(0,1)$

We get a confidence interval

$$p = \pm z^* \left(\sqrt{\frac{p(1-p)}{n}}\right) + \hat{p}$$

Notice that p occurs on both sides
If the sample contains <u>at least</u> 15 successes and 15 failures, and n is <u>quite large</u>, we use

$$p = \hat{p} \pm z^* \left(\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$$

where p is replaced with $\hat{p}$.

Now we can obtain a confidence interval for the proportion p of young adult males who live with parents.

Recall that $\hat{p} = .44$

So a 95% confidence interval will be

$$.44 \pm 1.96\sqrt{\frac{(.44)(1-.44)}{2253}}$$

$$= .44 \pm .02$$

The above method works well for samples of $n = 1000+$ and 15+ successes & failures.

However we need an alternative for small sample sizes. In fact, this new method is better in the previous case as well.

Plus 4 Method

Add two imaginary successes and two imaginary failures. Then use the previous method. This works well as long as $n \geq 10$

Ex: A study of Spanish Currency found cocaine traces in all 20 of the bills sampled.

A confidence interval for the proportion p of all spanish bills having cocam traces is:

$$p = \tilde{p} \pm z^* \sqrt{\frac{\tilde{p}(1-\tilde{p})}{20+4}}$$

where $\tilde{p} = \frac{22}{24} = \frac{11}{12}$.

For a 99% confidence interval,

we get $p = .917 \pm 2.57 \sqrt{\frac{\frac{11}{12}\left(\frac{1}{12}\right)}{24}}$

$= .917 \pm .145$

$= (.772, 1.062)$

Since $p \leq 1$ we write $(.772, 1)$

Choosing a Sample Size for a survey:

Ex: Gloria is running against Ronald for Mayor in a large city.

You would like to predict who will win the election.

You decide to do a phone survey where you ask registered voters who they plan to vote for.

You want a 95% confidence interval with margin of error 3% for the proportion $\hat{p}$ that will vote for Gloria.

How many people do you need in your sample?

The margin of error is $m = z^* \sqrt{\frac{p(1-p)}{n}}$

If we have an informed guess for the value of p, we can plug that in. Otherwise we err on the side of caution and set $p = \frac{1}{2}$.

Thus $m = z^* \sqrt{\frac{\frac{1}{2}(1-\frac{1}{2})}{n}}$

Solving for n gives:

$$n\, m^2 = (z^*)^2 \frac{1}{4}$$

$$n = \frac{(z^*)^2}{4m^2} = \frac{1}{4}\left(\frac{1.96}{.03}\right)^2 = 1067.11$$

So we will survey 1068 people
(Always round n up for safety)

Ex: Suppose that we have surveyed 1068 people and found that 600 planned to vote for **Gloria.**

Will she win?

One way to procede would be to create a confidence interval.

Instead, lets see what a hypothesis test would look like.

H_0: $p = .5$
H_A: $p > \frac{1}{2}$

Remember that $\dfrac{\hat{p} - p}{\sqrt{\dfrac{p(1-p)}{n}}}$ is approximately $N(0,1)$
(under H_0)

So, the z-test statistic is

$$z = \frac{.562 - .5}{\sqrt{\frac{\frac{1}{2}(1-\frac{1}{2})}{1068}}} = 4.05$$

This gives a P-value that allows us to reject H_0 at any reasonable level.

If H_0 (The two are evenly matched) then you would almost never have 600+ of 1068 sampled vote for Gloria.

Gloria will win (we say this confidently)