AP Statistics: Chapter 12: Inference for Proportions

Our statistical inferences so far has been concerned with making inferences about population **means**. Often, however, we want to answer questions about the **proportion** of some outcome in a population or compare proportions across several populations.

Here's a quick review of proportions:

- $\hat{p}$ is the sample proportion and equals $\frac{\text{number of "successes"}}{\text{sample size}} = \frac{X}{n}$, where X is the random variable.
- The mean of a sample proportion is $\mu_{\hat{p}} = p$, where p is the probability of "success"
- The probability of a "failure" is $q = 1 - p$
- The standard deviation of a sample proportion is $\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}}$

The methods and procedures will be strikingly similar to our calculations for sample means. Of course, there are always assumptions we need to address for our calculations (and predictions) to be accurate:

- The data are a simple random sample from the population of interest
- Normality
- The population is at least 10 times as large as the sample
- The sample size n must be large enough so that $np \geq 10$ AND $nq \geq 10$

These are the assumptions that we must maintain for our inference calculations to hold true. Understand that all of the inference in this chapter models itself directly from all of the models we used on Chapter 10. Confidence intervals will use z^* and the goal of significance tests will be to either **reject** or **fail to reject** the null hypothesis H_0. The key in transitioning to this chapter is remembering how we calculate the standard deviation of a proportion. Below is the comparison of the standard deviation (standard error, SE) of a sample mean versus the standard deviation of a sample proportion.

standard deviation of a sample mean

$$SE = \frac{\sigma}{\sqrt{n}}$$

standard deviation of a sample proportion

$$SE = \sqrt{\frac{pq}{n}}$$

Consequently, the method for finding confidence intervals will we similar.

confidence interval for a sample mean

$$CI = \bar{x} \pm z^* \frac{\sigma}{\sqrt{n}}$$

confidence interval for a sample proportion

$$CI = \hat{p} \pm z^* \sqrt{\frac{pq}{n}}, \ p \approx \hat{p}$$

Oh, so will the test z statistic for hypothesis testing with $H_0 : p = p_o$

test statistic z for a sample mean

$$z = \frac{\bar{x} - \mu}{SE} = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

test statistic z for a sample proportion

$$z = \frac{\hat{p} - p}{SE} = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}, \ p \approx p_o$$

Let's look at an example.

Example 1: In a survey of 2503 men and women aged 18 to 75 years and representative of the nation as a whole, 1927 people said the homeless are not adequately assisted by the government ("Parade Magazine," January 9, 1994). Is the use of the normal approximation justified in this problem? Find a 90% confidence interval.

Answer:
Let's check the assumptions.

We'll assume our sample of people surveyed is from a random sample from the population of ALL men and women aged 18 to 75 years. Check one!

We will assume that the number of men and women aged 18 to 75 years in the United States is 10 times our sample size (Population is at least 25030 people). Check two!

Our sample size is at least 30, so the Central Limit Theorem says our distribution is approximately normal. Check three!

Now we must crunch a few numbers.

First $\hat{p} = \text{proportion saying homeless not adequately assisted} = \dfrac{1927}{2503} = 0.7699$

Check the formulas: $np = (2503)(.7699) = 1927 \geq 10$
$nq = (2503)(1-0.7699) = (2503)(0.2301) = 576 \geq 10$ Check four!

We're now set to use a normal approximation and our interval should be fairly accurate. Let's STATE THE TYPE OF INFERENCE WE'RE USING!!

On-sample confidence interval

NOTE: for a confidence interval, use $\hat{p}$ as an estimate of *p*.

$$CI = \hat{p} \pm z^* \sqrt{\frac{pq}{n}}$$

$$CI = 0.7699 \pm 1.645\sqrt{\frac{(0.7699)(0.2301)}{2503}} = 0.7699 \pm 0.008412$$

We are 90% confident that the true proportion of people aged 18 to 75 who believe the homeless are not adequately assisted is between 75.60% and 78.37%.

CAKE! Right?

Let's try one using a significance test for a proportion.

Example 2: A botanist has produced a new variety of hybrid wheat that is better able to withstand drought than other varieties. The botanist knows that for the parent plants, the probability of the seed germination for the hybrid variety is unknown, but the botanist claims that it is 80%. To test this claim, 400 seeds from the hybrid plant are tested, and it is found that 312 germinated. Does this information indicate that the germination proportion is different from the botanist claim? Use $\alpha = 0.05$.

Answer:
One-proportion *z*-test

NOTE: for a significance test with $H_0 : p = p_o$, use p_o as an estimate of *p*.

We assume our sample of 400 seeds is take from a random sample of all hybrid seed plants.
We assume the population of ally hybrid seeds is at leas 4000
We assume normality by the Central Limit Theorem since our sample of 400 is greater than 30.

$$\hat{p} = \text{proportion of seeds that germinate} = \frac{312}{400} = 0.78$$

Since $np = 400(.78) = 312 \geq 10$ and $nq = 400(.22) = 88 \geq 10$, our inference results will be accurate.

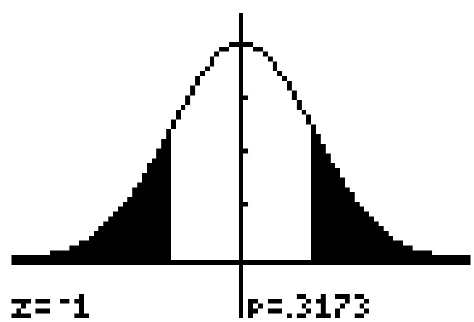
$H_0 : p = 0.8$, the null hypothesis states that the true population proportion of germinating seeds is 80%
$H_a : p \neq 0.8$, the alternative hypothesis states that the true population proportion of germinating seeds is not 80%.

Calculate our *z* test statistic. Remember that *z* is measured in standard deviation units and establishes the cut-off for our acceptance and rejection regions.

$$z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}} = \frac{0.78 - 0.8}{\sqrt{\frac{.8(.2)}{400}}} = -1.00$$

Calculate our *p*-value for the corresponding TWO-SIDED test base on the *z* test statistic. Remember that *p* is measured in probability units and is the area in the tail or tails in the rejection regions.

$$p = 2(.1587) = .3174$$



Interpret the results by comparing the *p*-value to a specified α value.

We fail to reject the null hypothesis since 0.3174 is NOT less than 0.05. There is not sufficient evidence to suggest that the proportion of hybrid seed that germinates is different from 80%. Our results are not statistically significant at the 5% level. In fact, the results we obtained would happen simply by chance approximately 32% of the time if the true population proportion was 80%.

<u>HOMEWORK</u>: Inference for Proportions, Ch 12 worksheet. Mini-quiz to follow.

Technology Toolbox: Inference for a population proportion on the TI-83/84

The TI-83/84 can be used to test a claim about a population proportion and to construct confidence intervals.

<u>Example 3</u>: A coin that is balanced should come up heads half the time in the long run. The French naturalist Count Buffon (1707-1788) tossed a coin 4040 times. He got 2048 heads. The sample proportion of heads is

$$\hat{p} = \frac{X}{n} = \frac{2048}{4040} = 0.5069$$

That's a bit more than one-half. Is this evidence that Buffon's coin was not balanced?

$$H_0 : p = 0.5$$

$$H_a : p \neq 0.5$$

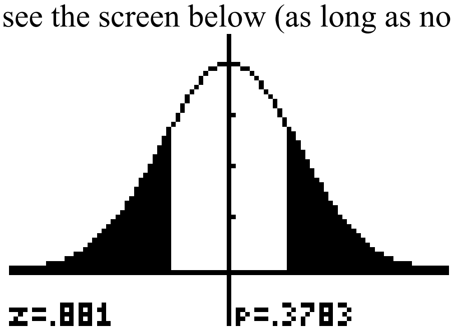
To perform a one-proportion significance test on the TI-83/84

- Press "STAT," then choose "TESTS," then number 5, "1-PropZTest."
- On the "1-PropZTest" screen, enter the values. Specify whether it's a two-sided or the appropriate type of one-sided test. Press "Calculate."

```
1-PropZTest
 p0:.5
 x:2048
 n:4040
 prop≠p0 <p0 >p0
 Calculate Draw
```

```
1-PropZTest
 prop≠.5
 z=.8810434857
 p=.3782942021
 p̂=.5069306931
 n=4040
```

- If you select "Draw," you'll see the screen below (as long as no other graphs are turned on).



Let's construct a 95% confidence interval for the probability p that Buffon's coin gives a head.

<u>To perform a one-proportion confidence interval on the TI-83/84</u>

- Press "STAT," then choose "TESTS" and letter A: "1-PropZInt."
- When the "1-PropZInt" screen appears, enter your data and choose your C-level as a decimal. Press "Calculate."

```
1-PropZInt
 x:2048
 n:4040
 C-Level:.95
 Calculate
```

```
1-PropZInt
 (.49151,.52235)
 p̂=.5069306931
 n=4040
```