Math107

Inference for a Single Proportion
(Sections 6.1–6.3)

Central Limit Theorem

The sampling distribution of a sample proportion is approximately normal and centered at p, if n is sufficiently large.

$$\widehat{p} \sim \mathcal{N}\left(p,\ \sqrt{\frac{p(1-p)}{n}}\right)$$

Success/Failure condition: n is “large enough” if np ≥ 10 and n(1-p)≥10

Confidence Intervals

Confidence Interval Formula

General formula:

$$\text{sample statistic} \pm \mathrm{z}^* \times \mathrm{SE}$$

Formula for a proportion:

$$\widehat{\mathbf{p}} \pm \mathbf{z}^* \sqrt{\frac{\widehat{\mathbf{p}}(\mathbf{1} - \widehat{\mathbf{p}})}{\mathbf{n}}}$$

Example

In a random sample of 500 movie goers in January 2013, 320 of them said they are more likely to wait and watch a new movie in the comfort of their own home.

a) Does the Central Limit Theorem for a sample proportion apply here? That is, are the conditions required to approximate the sampling distribution of a sample proportion using the normal distribution met?

b) Calculate and interpret a 95% confidence interval for the proportion of movie goers who are more likely to watch a new movie from home.

“64% are More Likely to Watch a Movie at Home”, rasmussenreports.com, January 20, 2013.

Margin of Error

$$\mathbf{ME} = \mathbf{z}^{*}\sqrt{\frac{\mathbf{p}(\mathbf{1}-\mathbf{p})}{\mathbf{n}}}$$

You can choose your sample size in advance based on your desired margin of error!

$$\mathbf{n} = \left(\frac{\mathbf{z}^{*}}{\mathbf{ME}}\right)^{\mathbf{2}}\mathbf{p}(\mathbf{1}-\mathbf{p})$$

If you don’t have an educated guess for p, use 0.5 to be conservative

Example

Suppose now that you wanted to obtain a new sample to update the interval estimate for the proportion of movie goers who are more likely to watch a new movie from home for 2014.

a) What was the margin of error for the CI calculated from the 2013 sample?

b) What sample size is needed if you want a margin of error of 0.02 (i.e., within ±2%)? (Use the sample proportion from the original sample.)

c) What sample size is needed if we want a margin of error of 0.02, and if you use the conservative estimate of p=0.5?

Hypothesis Tests

Hypothesis Test for p

1. State the hypotheses
2. Check the conditions necessary to use the normal distribution
3. Compute the test statistic
4. Compute the p-value
5. Make a decision and state its implications in the context of the problem

Test Statistic Formula

We want to use the sampling distribution of the sample proportion assuming the the null hypothesis is true

$$\mathbf{H_0 : p = p_0}$$

$$\mathbf{SE = \sqrt{\frac{p(1-p)}{n}}}$$

What should we use for p?

$$\mathbf{SE = \sqrt{\frac{p_0(1-p_0)}{n}}}$$

Cohen v. Brown University

- In 1991, a suit was filed against Brown University after Brown terminated funding for it's women's gymnastics and volleyball teams and its men's water polo and golf teams.
- The suit charged that Brown was violating Title IX of the Education Amendments of 1972, the federal law that prohibits sex discrimination by all educational institutions receiving federal funds. This requires men and women to have equivalent opportunities for participation.
- A main component of the plaintiff's case was that while 51% of the undergraduate student body was women, only 38% of the 897 students engaged in intercollegiate athletics were women.

Cohen v. Brown University

a) State the null and alternative hypothesis about the proportion of athletes that will be female at Brown University.

b) We can assess the strength of evidence against the null hypothesis by treating the 897 current athletes as a random sample from the process of athlete determination at Brown. If we initially give the university the benefit of the doubt and assume the null hypothesis is true, does the Central Limit Theorem for a sample proportion apply here? Explain.

c) If p=0.51 was hypothesized, calculate the test statistic.

d) Use the Central Limit Theorem to approximate the p-value.

e) Based on this p-value, what conclusion do you draw about whether this discrepancy could have arisen by chance? In other words, does your analysis suggest that the proportion of women involved in intercollegiate athletics is significantly lower than the proportion of women students at the university? Explain.

Using R

News on Twitter

- On HW 9 you used the bootstrap to get a CI for the proportion of U.S. adult Twitter users who get at least some of their news on Twitter.
- Now, we can use the normal-based approach.

Confidence intervals

```
# Load the mosaic package
library(mosaic)

# Load the data
twitternews <- read.csv("data/twitternews.csv")

# Constructing a 95% confidence interval
confint(prop.test(~ new.on.twitter == "yes",
                  data = twitternews,
                  conf.level = 0.95,
                  correct = FALSE))
```

Hypothesis tests

```
# Testing whether a majority get their news on Twitter

result <- prop.test(~ new.on.twitter == "yes",
                    data = twitternews,
                    p = 0.5,
                    alternative = "greater",
                    correct = FALSE)
result

# You can easily grab the p-value
pval(result)
```