Lab: Statistical Inference of Proportions

75694189 16257626 13341225

Chen Wang, Junke Wang, Zhuozhi Xiong

due 11/13 at end of discussion

As A Class

1) Go to: http://www.rossmanchance.com/applets/OneProp/OneProp.htm?candy=1. Set "proportion of orange" to 0.4, and "number of candies" to 50. Select "proportion of orange" and "summary statistics".

2) Draw 1 sample. What sample proportion of orange did you get?

Text:

I got a sample proportion of 0.34

3) Uncheck "Animate". Draw 50 more samples. What is the mean of the sample proportions? What is the standard deviation?

Text:

The mean was 0.382, and the standard deviation was 0.071.

4) By central limit theorem, what should the mean of the sample proportions be? What should the standard error be?

Code (for use as calculator):

```
p <- 0.4
sd <- sqrt(p * (1 - p) / 50)
sd
```

```
## [1] 0.06928203
```

Text: The mean should be 0.4.
The standard error should be 0.06928203.

5) Draw 1 more sample. Conduct a hypothesis test to determine if this sample proportion is greater than 0.4.

Code:

```
p.hat <- 0.56

z <- (p.hat - 0.4) / sqrt(0.4 * (1 - 0.4) / 50)
z
```

```
## [1] 2.309401
```

```
z.crit <- qnorm(0.95)
z.crit
```

```
## [1] 1.644854
```

```
pvalue <- 1 - pnorm(z)
pvalue
```

```
## [1] 0.01046067
```

Text:

$H_0 : p = 0.4$

$H_A : p > 0.4$

Because $Z = 2.31 > Z^* = 1.64$, we have sufficient evidence to conclude that we can reject the null hypothesis, and p is greater than 0.4.

On Your Own, In Groups

1) According to a survey of 1,364 4-year colleges and universities conducted by the Pew Research Center, 53.3% of colleges admitted at least two-thirds of their applicants in 2017.

a) Calculate an 85% confidence interval for the proportion of colleges/universities that admitted at least two-thirds of their applicants in 2017. Interpret it.

Code:

```
p = 0.533
sd = sqrt(p * (1 - p) / 1364)
critic = qnorm((1 - 0.85) / 2)
p - critic * sd
```

```
## [1] 0.5524463
```

```
p + critic * sd
```

```
## [1] 0.5135537
```

Text:

The confidence interval is (0.5135537, 0.5524463).

We have 85% confidence that the proportion of colleges/universities that admitted at least two-thirds of their applicants in 2017 is between 51.3% and 55.2%.

2) In the fall of 2018, Girls Who Code (GWC) surveyed women affiliated with the organization about their experiences applying for internships in the tech industry (read the full report here). Out of 1,017 women, 29% of survey respondents reported having negative experiences during an internship application process.

a) Calculate a 95% confidence interval for the proportion of women affiliated with GWC who had negative internship applications experiences. Interpret it.

Code:

```
p = 0.29
sd = sqrt(p * (1 - p) / 1017)
critic = qnorm((1 - 0.95) / 2)
p - critic * sd
```

```
## [1] 0.3178879
```

```
p + critic * sd
```

```
## [1] 0.2621121
```

Text:

The confidence interval is (0.2621121, 0.3178879).

We have 95% confidence that the proportion of women affiliated with GWC who had negative internship applications experiences is between 26.2% and 31.7%.

b) Conduct a hypothesis test to evaluate whether the proportion is larger than 20%.

Code:

```
p.hat <- 0.29

z <- (p.hat - 0.2) / sqrt(0.2 * (1 - 0.2) / 1017)
z
```

```
## [1] 7.175348
```

```
z.crit <- qnorm(0.95)
z.crit
```

```
## [1] 1.644854
```

```
pvalue <- 1 - pnorm(z)
pvalue
```

```
## [1] 3.606004e-13
```

Text:

$H_0 : p \leq 0.2$

$H_A : p > 0.2$

Because $Z = 7.17 > Z^* = 1.64$, we have sufficient evidence to conclude that we can reject the null hypothesis, and p is greater than 0.2.

c) Do your answers for (a) and (b) agree with each other?

Text:

Yes. In part (a), we have 95% confidence that the proportion is between 26.2% and 31.7% and this is true that the number is greater than 20%, which is indicated in part (b).

Optional Exercises

1) Below is a table illustrating the prevalence of two different biomarkers of Alzheimer's Disease, drawn from a recent paper. Having at least 1 APOE e4 allele is thought it increase your risk for Alzheimer's; having elevated levels of AB protein in the brain is also seen an increased risk.

```
##                   AB Elevated AB Non-Elevated
## 0 APOE Alleles             53              182
## 1-2 APOE Alleles           47               53
```

a) What is the estimated difference in proportions of elevated AB levels between those who have no APOE e4 alleles, and those who have at least 1 APOE e4 allele? What is the standard error?

Code:

```
n1 = 53 + 182
n2 = 47 + 53
p1 = 53 / n1
p2 = 47 / n2
p1 - p2
```

```
## [1] -0.2444681
```

```
sd = sqrt(p1 * (1 - p1) / n1 + p2 * (1 - p2) / n2)
sd
```

```
## [1] 0.0568706
```

Text:

The estimated difference is -0.24.

The standard error is 0.05

b) Conduct a hypothesis test for whether this difference in proportions is equal to 0.

Code:

```
p.hat <- -0.24

z <- (p.hat - 0) / sd
z
```

```
## [1] -4.220107
```

```
z.crit <- qnorm(0.05)
z.crit
```

```
## [1] -1.644854
```

Text:

$H_0 : p_1 - p_2 = 0$

$H_A : p_1 - p_2 \neq 0$

Because $Z = -4.2 < Z^* = -1.6$, we have sufficient evidence to conclude that we can reject the null hypothesis, and $p_1 - p_2 \neq 0$.

2) Below are results from a FiveThirtyEight survey about people's opinions of the Oxford (or serial) comma. Answers on the left correspond to the question "How much, if at all, do you care about the debate over the use of the word 'data' as a singular or plural noun?". Answers on the top correspond to the question "How much, if at all, do you care about the use (or lack thereof) of the serial (or Oxford) comma in grammar?"

```
##
##              Not at all Not much Some A lot
##   Not at all         77       53   46    27
##   Not much           32      143  152    76
##   Some                8       57  178   109
##   A lot               7       11   38    77
```

a) Calculate a 70% confidence interval for the difference in proportions of those who only care "Some" about the Oxford comma debate and "A lot" about the data debate, and those who only care "Some" about the Oxford comma debate and "Not Much" about the data debate. What does this interval tell you?

Code:

Text:

b) Now conduct a hypothesis test for whether the difference in these two proportions is equal to 0. What can we conclude?

Code:

Text: