

HW6

anjali hussan

11/11/2017

6.6 2010 Healthcare Law.

On June 28, 2012 the U.S. Supreme Court upheld the much debated 2010 healthcare law, declaring it constitutional. A Gallup poll released the day after this decision indicates that 46% of 1,012 Americans agree with this decision. At a 95% confidence level, this sample has a 3% margin of error. Based on this information, determine if the following statements are true or false, and explain your reasoning.

a. We are 95% confident that between 43% and 49% of Americans in this sample support the decision of the U.S. Supreme Court on the 2010 healthcare law.

Ans: False, 46% agrees

b. We are 95% confident that between 43% and 49% of Americans support the decision of the U.S. Supreme Court on the 2010 healthcare law.

Ans: True, The confidence interval is 95% with 3% margin of error.

c. If we considered many random samples of 1,012 Americans, and we calculated the sample proportions of those who support the decision of the U.S. Supreme Court, 95% of those sample proportions will be between 43% and 49%.

Ans: False, the confidence interval means that 95% of the population will be in the range.

d. The margin of error at a 90% confidence level would be higher than 3%.

Ans: False, False, Margin of error will also be lower

6.12 Legalization of marijuana, Part I.

The 2010 General Social Survey asked 1,259 US residents: "Do you think the use of marijuana should be made legal, or not?" 48% of the respondents said it should be made legal.

a. Is 48% a sample statistic or a population parameter? Explain.

Ans: Sample statistic. A population parameter should be from a census- not among 1259 US residents

b. Construct a 95% confidence interval for the proportion of US residents who think marijuana should be made legal, and interpret it in the context of the data.

Ans: Assuming that sample is random, but there are more than 10 successes and 10 failures. In this case it meets the criteria for a confidence interval.

```
SE = (0.48*(1-0.48)/1258)^0.5
ME = 1.96*SE
lower<- 0.48-ME
upper <- 0.48+ME
cat("Interval is ", lower, "to", upper)
```

```
## Interval is 0.4523918 to 0.5076082
```

- c. A critic points out that this 95% confidence interval is only accurate if the statistic follows a normal distribution, or if the normal model is a good approximation. Is this true for these data? Explain.

Ans: If sample is not random or biased, then it is not valid. If sample is random and unbiased, then it is valid.

- d. A news piece on this survey's findings states, "Majority of Americans think marijuana should be legalized." Based on your confidence interval, is this news piece's statement justified?

Ans: Considering upper boundary of a confidence interval, we can say that 50% of the Americans think marijuana should be legal. Although it's not a good practice to take upper boundary of a confidence interval.

6.20 Legalize Marijuana, Part II.

As discussed in Exercise 6.12, the 2010 General Social Survey reported a sample where about 48% of US residents thought marijuana should be made legal. If we wanted to limit the margin of error of a 95% confidence interval to 2%, about how many Americans would we need to survey?

Ans:

```
p <- 0.48
ME <- 0.02
Z <- qnorm(0.975)
SE <- ME/Z
n <- (p * (1-p)) / SE^2
cat("We need ", round(n,1), "Americans")
```

```
## We need 2397.1 Americans
```

6.28 Sleep deprivation, CA vs. OR, Part I.

According to a report on sleep deprivation by the Centers for Disease Control and Prevention, the proportion of California residents who reported insufficient rest or sleep during each of the preceding 30 days is 8.0%, while this proportion is 8.8% for Oregon residents. These data are based on simple random samples of 11,545 California and 4,691 Oregon residents. Calculate a 95% confidence interval for the difference between the proportions of Californians and Oregonians who are sleep deprived and interpret it in context of the data.

Ans:

```
California <- 11545
Oregon <- 4691

CalProportion <- 0.08
OregonProportion <- 0.088
PE <- OregonProportion - CalProportion

SE <- sqrt( ((CalProportion * (1 - CalProportion)) / California) +
            ((OregonProportion * (1 - OregonProportion)) / Oregon))
ME <- qnorm(0.975) * SE

lower <- PE - ME
lower
```

```
## [1] -0.001497954
```

```
upper <- PE + ME
upper
```

```
## [1] 0.01749795
```

Confidence interval span is 0, hence, statistically sleep proportions have no difference between people of California and people of Oregon.

6.48 Coffee and Depression.

Researchers conducted a study investigating the relationship between caffeinated coffee consumption and risk of depression in women. They collected data on 50,739 women free of depression symptoms at the start of the study in the year 1996, and these women were followed through 2006. The researchers used questionnaires to collect data on caffeinated coffee consumption, asked each individual about physician-diagnosed depression, and also asked about the use of antidepressants. The table below shows the distribution of incidences of depression by amount of caffeinated coffee consumption.

- What type of test is appropriate for evaluating if there is an association between coffee intake and depression?

Ans: Chi square test.

- Write the hypotheses for the test you identified in part (a).

Ans: Ho: There is no relationship between coffee and depression HA: There is a relationship between coffee and depression

- Calculate the overall proportion of women who do and do not suffer from depression.

Ans:

```
W_with_Depression= 2607/50739
cat("The propotion of Women suffer from Depression: ",W_with_Depression,"\n")
```

```
## The propotion of Women suffer from Depression: 0.05138059
```

```
witout_Depression = 48132/50739
cat("The propotion of Women do not suffer from Depression: ", witout_Depression)
```

```
## The propotion of Women do not suffer from Depression: 0.9486194
```

d. Identify the expected count for the highlighted cell, and calculate the contribution of this cell to the test statistic, i.e. $(\text{Observed} - \text{Expected})^2 / \text{Expected}$.

Ans:

```
expect = 6617 * W_with_Depression
expected_count = (373-expect)^2/expect
cat("Expected count is :", expected_count)
```

```
## Expected count is : 3.205914
```

e. What is the conclusion of the hypothesis test?

Ans:

```
x <- 20.93
diff <- (5-1)*(2-1)
p_value <- 1-pchisq(x, diff)
cat("The p Value is : ", p_value)
```

```
## The p Value is : 0.0003269507
```

The p value is less than .05 Hence, we can reject the null hypothesis which is "There is no relationship between coffee and depression"

g. One of the authors of this study was quoted on the NYTimes as saying it was "too early to recommend that women load up on extra coffee" based on just this study. Do you agree with this statement? Explain your reasoning.

Ans: Yes! I agree with this statement. One test gave results that there is no association between coffee and depression but there could be other factors and affects of having a lot of coffee.

End of this file