BRIOT_HW_chapter 6

Valerie Briot

April 3, 2016

This document pertains to the homework assigned for CUNY606: Probability and Statistics on Chapter 6. The following exercises are included; 6.6, 6.12, 6.20, 6.28, 6.44, and 6.48.

Exercise 6.6: 2010 Health Care 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 decisions indicates that 46% of 1,012 Amercnas agree with the decisions. At a 95% confidence level, this sample has a 3% margin of error. Based on htis informatioln, 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 decsision of the U.S. Supreme Court on the 2010 healthcare law. FALSE - Actually we are 100% of 46% of Americans in suvey support the the decision.

b) We are 95% confident that between 43% and 49% of Americans support the decsiion. TRUE - Confidence interval allow us to make inference on the wider population. For proportion, it means that the true proportion of the population at large will be between the end point of interval.

c) If we considered many random samples of 1,012 Americans and we calculated the sample proportion of those who support the decsiion of the U.S. Supreme Court, 95% of those examples would be between 43% and 49%. FALSE - Again, the conclusion is about the proportion of the population of American that support the decision. The probability interpretation of the interval is that: 95% of time, the true proporation of the American will fall within interval constructed as above based on the sample data.

d) The margin of error at a 90% confidence level would be higher than 3%.
FALSE - At 90% confidence level, the interval would be shorter not wider hence the margin of error would be smaller.

Exercise 6.12: Legalization of Marijuana, Part I

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

a) Is 48% a sample statistic or a population parameter? Explain. This correspond that a sample statistic since this is the percentage of respondats to the survey that support the legalization of marijuana.

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.

Checking for Condition: 1,259 < 10% of US residents and we can probably assume that survey was conducted using random people so we have indedence of observations and

p=.48, success = 0.48 * 1,259 = 604.32 > 10 q = 1-p , failure = 0.52 * 1,259 = 654.68 > 10

For 95% interval we will use z = 1.96

Confidence interval: 0.48 +- 1.96 * SE, where SE = sqrt(p*(1-p)/n), 0.48 +- 1.96* sqrt(0.48*0.52/1259) = 0.48 +- 1.96* 0.014 = 0.48 +- 0.0276 (0.4524, 0.5076)

c) Yes, we can assume a normal distribution, we will assume that survey was conducted with random selection (see above)

d) No, the only statement we can make is that we are 95% confident that the proportion of Americans that support legalization of marijuana is 48%. Which does not refelct the majority (<50%).

Exercise 6.20: Legalize Marijuana, Part II

At 95% Confidence level, me = 1.96 * SE, with SE = sqrt(p * (1-p)/n), solving for n,

me/1.96 = sqrt(p*(1-p)/n), squaring both side (me/1.96)^2 = (p*(1-p)/n), solving for n n = (p*(1-p))/(me/1.96)^2, substituing values n = (0.48 0.52) * (1.96/0.02)^2 n = 0.2496 * 9604 = 2,397.1584, hence n = 2,398

Exercise 6.28: Sleep deprivation, CA vs OR, PartI

According to a report on sleep deprivation by the Center of Disaease Control and Presentation, 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 Oregons residents. These data are based on simple random samples of 11,545 California and 4.691 Oregon residents. Calculate a 95% Condifence interval for the difference between the proportions of California and Oregornians who are sleep deprived and interpret it in context of the data.

Checking Condition: Each sample were random and represent less that 10% of California and Oregon population, hence each proprotion follow a normal model in addition, we can assume that the 2 samples are independent of each other.

Also, the success and failure conditions are met, we have at least 10 sucess/failure for each sample.

```
p_c <- 0.08
n_c <- 11545

p_o <- 0.088
n_o <- 4691

p_dif <- p_o - p_c
se_c_o <- sqrt((p_c*(1-p_c))/n_c + (p_o*(1-p_o))/n_o)

interval_p_c_low <- p_dif - 1.96 * se_c_o
interval_p_c_high <- p_dif + 1.96 * se_c_o
```

interval (-0.0014981 , 0.0174981)

We are 95% confident that the proportion of sleep deprived residents change between 0 and 1% from California and Oregon.

Exercise 6.44: Barking deer

a) H0: barking deers have no preferred habitats for forage Ha: barking deers have preferred habitats for forage over other habitats

b) Chi-square test

c) Total count = 426

Expected_wood = 0.048 * 426 = 20.5 > 5 Expected_culitvated_grassplot = 0.147 * 426 = 62.6 > 5 Expected_decidious_forests = 0.396 * 426 = 168.7 >5 Expected_other = 0.409 * 426 = 174.2 >5

Condition for Chi-square test met.

d)

```
df <- 4-1

e_wood <- 20.5
e_cul <- 62.6
e_deci <- 168.7
e_oth <- 174.2

o_wood <- 4
o_cul <- 16
o_deci <-67
o_oth <- 345

z_wood <- (o_wood-e_wood)/sqrt(e_wood)
z_cul <- (o_cul-e_cul)/sqrt(e_cul)
z_deci <- (o_deci-e_deci)/sqrt(e_deci)
z_oth <- (o_oth-e_oth)/sqrt(e_oth)

x <- z_wood^2 + z_cul^2 + z_deci^2 + z_oth^2
```

For degree of freedom = 3, p-value for 276.7456709 is less than 0.05, hence we would reject H0 hypothesis in favor of Ha.

Exercise 6.48: Coffee and Depression

a) We would use a two-way-table

b) Ho: There is not relationship between the consumption of caffeinated coffee and risk of depression in women Ha: There is a relationship bewteen the consumption of caffeinated coffee and risk of depression in women

c) w_depression = 2607/50739 = 0.05

w_no_depression = 48132/50739 = 0.95

d) expected_count_1_2 = 2607*6617/50739 = 340

(oberved count - expected count)^2/expected count = (373 - 340)^2/340 = 3.2

e) degree of freedom = (number of row - 1) * (number of columns - 1) = (5-1) * (2-1) = 4

p-value = 0.001 (from table, for df = 4, last value is 18.47)

f) p_value < 0.05, we would reject H0 in favor of Ha

g) Yes, we would agree. we just inferred that there is a relationship between caffeinated coffee consumption and depression hovever we have not determined in which direction the impact is.