STAR511: HW 7

Megan Sears

# Q1

## [1] FALSE

## [1] TRUE

Checking for large sample normal approximation can be done by multiplying n by and n*(1-). If both of those are greater than or equal to 5, then normal approximation is valid. In the first statement above, n multiplied by is not greater than or equal to 5. The second statement, n*(1-), is greater than 5. However, they both need to be true for the approximate to be valid.

# Q2

##   
## Exact binomial test  
##   
## data: 4 and 65  
## number of successes = 4, number of trials = 65, p-value = 1  
## alternative hypothesis: true probability of success is not equal to 0.06153846  
## 90 percent confidence interval:  
## 0.02129115 0.13531197  
## sample estimates:  
## probability of success   
## 0.06153846

# Q3

##   
## 1-sample proportions test with continuity correction  
##   
## data: 124 out of 215, null probability 0.5  
## X-squared = 4.7628, df = 1, p-value = 0.02908  
## alternative hypothesis: true p is not equal to 0.5  
## 95 percent confidence interval:  
## 0.5075921 0.6431080  
## sample estimates:  
## p   
## 0.5767442

Our data suggest that the true population proportion of voters who support Jones for mayor is somewhere between 51 and 64 percent.

# Q4

##   
## 1-sample proportions test with continuity correction  
##   
## data: 124 out of 215, null probability 0.5  
## X-squared = 4.7628, df = 1, p-value = 0.01454  
## alternative hypothesis: true p is greater than 0.5  
## 95 percent confidence interval:  
## 0.5183674 1.0000000  
## sample estimates:  
## p   
## 0.5767442

We will reject the null hypothesis that less than half of the voters support Jones. We have evidence that more than half of registered voters support candidate Jones based on the confidence intervals being 52 to 100%. Additionally, the p-value is less than 0.05, which provides evidence that more than half off the voters support Jones.

# Q5

## [1] 267

The minimum sample size for no previous information about is 267.

# Q6

## [1] 225

The minimum sample size required for conjectures value of 0.3 is 225.

# Q7 \*\* two side or one side?

##   
## Two-sample comparison of proportions power calculation   
##   
## n = 41.66374  
## p1 = 0.4  
## p2 = 0.1  
## sig.level = 0.05  
## power = 0.9  
## alternative = two.sided  
##   
## NOTE: n is number in \*each\* group

The sample size required per group to achieve 90% is 42.

# Q8 \*\*

## [1] 0.5466667

## [1] 0.4533333

55% preferred candidate A before the debate. 45% preferred candidate A after the debate.

# Q9

## [,1] [,2]  
## [1,] 28 13  
## [2,] 6 28

##   
## McNemar's Chi-squared test with continuity correction  
##   
## data: prefer  
## McNemar's chi-squared = 1.8947, df = 1, p-value = 0.1687

We cannot conclude there is a difference in proportions that voters who changed their preference after debated changed in one candidate or the othe based on the p-value being greater than alpha. \*\*\*

# Q10

This is not correct because the lung cancer patients and healthy controls are not a random sample of the population. It is a case-control study where lung cancer patients were selected and healthy people were selected as a control group.

# Q11

## [,1] [,2]  
## [1,] 328 141  
## [2,] 101 98

## $data  
## Outcome  
## Predictor Disease1 Disease2 Total  
## Exposed1 328 141 469  
## Exposed2 101 98 199  
## Total 429 239 668  
##   
## $measure  
## odds ratio with 95% C.I.  
## Predictor estimate lower upper  
## Exposed1 1.000000 NA NA  
## Exposed2 2.257145 1.60518 3.173915  
##   
## $p.value  
## two-sided  
## Predictor midp.exact fisher.exact chi.square  
## Exposed1 NA NA NA  
## Exposed2 3.052348e-06 3.938413e-06 2.243712e-06  
##   
## $correction  
## [1] FALSE  
##   
## attr(,"method")  
## [1] "Unconditional MLE & normal approximation (Wald) CI"

# Q12 \*\*

The bird owning group has highers odds of getting lung cancer.

# Q13 \*\*

Based on the 95% CI it does not include 1, which indicates bird ownership

# Q14 \*\*

The chi-square test p-value is less than alpha, which indicates the odds ratio of 2.26 is statistically significant.

# Appendix

#Retain (and do not edit) this code chunk!!!  
library(knitr)  
knitr::opts\_chunk$set(echo = FALSE)  
knitr::opts\_chunk$set(message = FALSE)  
  
library(tidyverse)  
library(here)  
  
#Q1  
#n\*pi hat greater than or equal to 5 AND  
#n\*(1-pi hat) greater than or equal t o5  
  
n <- 65  
pi\_hat <- 4/65  
  
n\*pi\_hat >= 5  
  
n\*(1-pi\_hat) >= 5  
  
#Q2  
binom.test(4, 65, p = pi\_hat, conf.level = 0.9)  
  
#Q3  
prop.test(124, 215, correct = T)  
  
#Q4  
prop.test(124, 215, alternative = 'greater', correct = T)  
  
#Q5  
  
me <- 0.06  
z <- 1.96  
pi\_hat <- 0.5  
  
n <- (z^2 \* pi\_hat \* (1-pi\_hat)) / me^2  
  
n <- ceiling(n)  
n  
  
#Q6  
me <- 0.06  
z <- 1.96  
pi\_hat <- 0.3  
  
n <- (z^2 \* pi\_hat \* (1-pi\_hat)) / me^2  
  
n <- ceiling(n)  
n  
  
#Q7  
power.prop.test(power = 0.9, p1 = 0.4, p2 = 0.1, sig.level = 0.05, alternative = 'two.sided')  
  
  
#Q8  
41/75  
  
34/75  
  
#Q9  
prefer <- matrix(c(28, 13, 6, 28), byrow = T, nrow = 2)  
prefer  
  
mcnemar.test(prefer)  
  
library('epitools')  
  
#Q11  
birds <- matrix(c(328, 141, 101, 98), byrow = T, nrow = 2)  
birds  
  
oddsratio(birds, method='wald')