Lab-6.R

daniellesenechal

2019-10-20

#############################################################################################  
## Danielle Senechal  
## CSC 315, Lab #6   
#############################################################################################  
  
library(ggplot2)  
library(gtools)  
library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(cowplot)

##   
## \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

## Note: As of version 1.0.0, cowplot does not change the

## default ggplot2 theme anymore. To recover the previous

## behavior, execute:  
## theme\_set(theme\_cowplot())

## \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

# Question 1a ###############################################################################  
p.hat <- 498/935  
p <- 0.50  
n <- 935  
  
mu <- p  
mu #expected

## [1] 0.5

sd <- sqrt(p\*(1-p)/n)   
sd #expected

## [1] 0.01635175

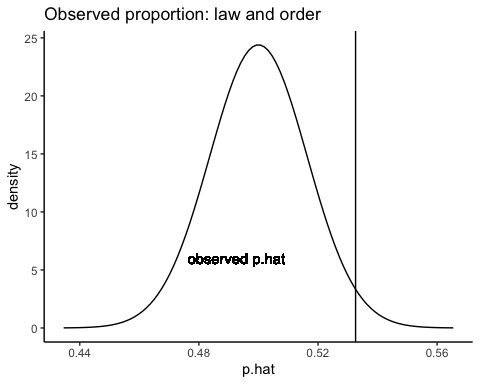
mu.actual <- p.hat  
mu.actual #actual

## [1] 0.5326203

sd.actual <- sqrt(p.hat\*(1-p.hat)/n)   
sd.actual #actual

## [1] 0.01631691

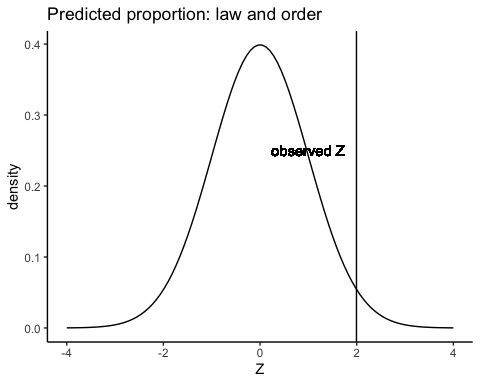
# Question 1b ###############################################################################  
df.phat <- data.frame(x = seq(mu-4\*sd, mu+4\*sd, length.out = 100)) %>%  
 mutate(y = dnorm(x, mean = mu, sd = sd))  
  
plot.phat <- ggplot(df.phat) + geom\_line(aes(x, y)) +  
 theme\_classic() + ggtitle("Observed proportion: law and order") +  
 labs(x = "p.hat", y = "density") +  
 geom\_vline(xintercept = p.hat) +   
 geom\_text(aes(x=p.hat-.04, y=6), label = "observed p.hat")  
plot.phat



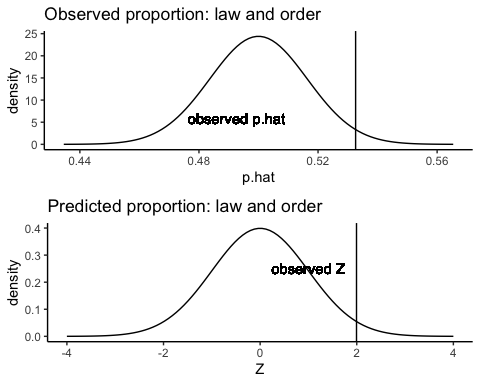
# Question 1c ###############################################################################  
Z = (p.hat - mu) / (sd)  
Z

## [1] 1.994913

x = seq(-4,4, length.out = 100)  
  
df.z <- data.frame(x = seq(-4, 4, length.out = 100)) %>%  
 mutate(y = dnorm(x))  
  
plot.z <- ggplot(df.z) + geom\_line(aes(x, y)) +  
 theme\_classic() + ggtitle("Predicted proportion: law and order") +  
 labs(x = "Z", y = "density") +  
 geom\_vline(xintercept = Z) +  
 geom\_text(aes(x=Z-1, y=.25), label = "observed Z")  
plot.z



plot\_grid(plot.phat, plot.z, nrow = 2)



p.value = 2\*pnorm(-abs(Z))  
p.value

## [1] 0.04605233

# Question 1d ###############################################################################  
p2 = prop.test(498, 935, p = 0.50, correct = FALSE)  
  
p2$statistic

## X-squared   
## 3.979679

sqrt(p2$statistic)

## X-squared   
## 1.994913

p2$p.value

## [1] 0.04605233

# Question 1e ###############################################################################  
  
# Because p-value of 0.04605233 is less than 0.05, we can reject the null hypothesis in favor   
# of the alternative. This means there is sufficient evidence that American adults   
# wish to prioritize “law and order”.  
  
# Question 1f ###############################################################################  
  
# We conclude that there is sufficient evidence that American adults   
# wish to prioritize “law and order”, when in reality American adults do not have a preference   
# over which is the bigger priority.  
  
# Question 2a ###############################################################################  
  
# H0: p = 1/6  
# HA: p > 1/6  
  
# Null hypothesis (H0): This person has a 1/6 chance at correctly predicting the outcome of a roll  
# of a standard die.  
  
# Alternative hypothesis (HA): This person has a greater than 1/6 chance at correctly predicting   
# the outcome of a roll of a standard die, and is a psychic.   
  
# Question 2b ###############################################################################  
  
psychic = prop.test(12, 50, p = 1/6 , correct = FALSE)  
  
psychic$statistic

## X-squared   
## 1.936

sqrt(psychic$statistic)

## X-squared   
## 1.391402

# Question 2c ###############################################################################  
  
psychic$p.value

## [1] 0.1641035

# Question 2d ###############################################################################  
  
# Because p-value of 0.1641035 is greater than 0.05, we fail to reject the null hypothesis.  
# This means there is not sufficient enough evidence that this person can predict the number   
# rolled on a die, and she is not a psychic.  
  
# Question 2e ###############################################################################  
  
# We conclude that there is not sufficient enough evidence that this person can predict the number   
# rolled on a die, when in reality this person is a psychic and can predict what will be rolled.  
  
# Question 3a ###############################################################################  
  
# H0: p = 0.36  
# HA: p ≠ 0.36  
  
# Null hypothesis (H0): There is 36% chance that adults make all/most of their purchases with cash,   
# and the proportion has not changed in the past 5 years.  
  
# Alternative hypothesis (HA): There is a greater or less than 36% chance that adults make   
#all/most of their purchases with cash and the proportion has changed in the past 5 years.  
  
# Question 3b ###############################################################################  
  
cashmoney = prop.test(246, 1024, p = 0.36, correct = FALSE)  
  
cashmoney$statistic

## X-squared   
## 63.75024

sqrt(cashmoney$statistic)

## X-squared   
## 7.984375

# Question 3c ###############################################################################  
  
cashmoney$p.value

## [1] 1.412361e-15

# Question 3d ###############################################################################  
  
# Because p-value of 1.412361e-15 is less than 0.05, we can reject the null hypothesis in favor   
# of the alternative. This means there is sufficient evidence that there is a greater or less   
# than 36% chance that adults make all/most of their purchases with cash,  
# and the proportion has changed in the past 5 years.  
  
# Question 4a ###############################################################################  
  
z.foura = 3.32  
p.value.foura = 2\*pnorm(-abs(z.foura))  
p.value.foura

## [1] 0.0009001745

# With a p-value of 0.0009001745, we reject the null hypothesis in favor of the alternative.  
  
# Question 4b ###############################################################################  
  
z.fourb = -1.3  
p.value.fourb = 2\*pnorm(-abs(z.fourb))  
p.value.fourb

## [1] 0.193601

# With a p-value of 0.193601, we fail to reject the null hypothesis.  
  
# Question 4c ###############################################################################  
  
z.fourc = -2.02  
p.value.fourc = 2\*pnorm(-abs(z.fourc))  
p.value.fourc

## [1] 0.04338339

# With a p-value of 0.04338339, we reject the null hypothesis in favor of the alternative.  
  
# Question 5a ###############################################################################  
  
# H0: p.aspirin – p.placebo = 0  
# HA: p.aspirin – p.placebo ≠ 0  
  
# Null hypothesis (H0): A daily dose of aspirin does not reduce the risk of dying from cancer.  
  
# Alternative hypothesis (HA): A daily dose of aspirin can reduce the risk of   
# dying from cancer.  
  
# Question 5b ###############################################################################  
  
p.hat.one <- 327/14035 #aspirin  
p.hat.two <- 347/11535 #placebo  
n.one <- 14035 #aspirin  
n.two <- 11535 # placebo  
p.five <- ((n.one \* p.hat.one) + (n.two \* p.hat.two)) / (n.one + n.two)  
p.five

## [1] 0.02635901

# The estimate of the common population proportion is 0.02635901.   
  
# Question 5c ###############################################################################  
  
sd.five <- sqrt((p.five \* (1 - p.five)) \* ((1 / n.one) + (1 / n.two)))  
sd.five

## [1] 0.002013327

# The standard deviation of the difference between the two proportions is 0.002013327  
  
# Question 5d ###############################################################################  
  
x.five <- c(347, 327)  
n.five <- c(11353, 14025)  
  
res.five <- prop.test(x.five, n.five, correct = TRUE)  
res.five

##   
## 2-sample test for equality of proportions with continuity  
## correction  
##   
## data: x.five out of n.five  
## X-squared = 12.474, df = 1, p-value = 0.0004126  
## alternative hypothesis: two.sided  
## 95 percent confidence interval:  
## 0.003136651 0.011361550  
## sample estimates:  
## prop 1 prop 2   
## 0.03056461 0.02331551

# The sample proportions are 0.03056461 for placebo, and 0.02331551 for aspirin.   
  
# Question 5e ###############################################################################  
  
z.five <- sqrt(res.five$statistic)  
z.five

## X-squared   
## 3.531874

# The z test statistic is 3.531874.  
  
# Question 5f ###############################################################################  
  
res.five$p.value

## [1] 0.000412626

# The p-value is 0.000412626.  
  
# Question 5g ###############################################################################  
  
# Because p-value of 0.000412626 is less than 0.05, we can reject the null hypothesis in favor   
# of the alternative. This means there is sufficient evidence that a daily dose of aspirin   
# can reduce the risk of dying from cancer.  
  
# Question 5h ###############################################################################  
  
# We conclude that there is sufficient evidence that a daily dose of aspirin could reduce   
# the risk of dying from cancer, when in reality a daily dose of aspirin does not cause any   
# change, meaning it does not reduce or increase the risk of dying of cancer (False positive).  
  
# Question 5i ###############################################################################  
  
# We conclude that there is not sufficient enough evidence that a daily dose of aspirin   
# reduces the risk of dying of cancer, when in reality there is sufficient enough   
# evidence that a daily dose of aspirin reduces the risk of dying of cancer.