Assignment 8

Natalie Schmer

40 points total, 2 points per problem part unless otherwise noted.

## Q1 Contrasts

{  
 library(tidyverse)  
 library(emmeans)  
}

## Warning: package 'tibble' was built under R version 3.6.2

InData <- read.csv("/Users/natalieschmer/Desktop/GitHub/stats\_511/data/WeightLoss.csv")  
  
WtLoss <- InData %>%  
 gather(key = "Trt", value = "Loss") %>%  
 mutate(Trt = as\_factor(Trt)) %>%  
 mutate(Trt = fct\_relevel(Trt, "S"))  
  
str(WtLoss)

## 'data.frame': 50 obs. of 2 variables:  
## $ Trt : Factor w/ 5 levels "S","A1","A2",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ Loss: num 8.7 9.3 8.2 8.3 9 9.4 9.2 12.2 8.5 9.9 ...

#Summary stats   
SumStats <- WtLoss %>%   
 group\_by(Trt) %>%   
 summarise(n = n(),  
 mean = mean(Loss),  
 sd = sd(Loss),  
 se = sd/sqrt(n))  
  
SumStats

## # A tibble: 5 x 5  
## Trt n mean sd se  
## <fct> <int> <dbl> <dbl> <dbl>  
## 1 S 10 9.27 1.16 0.366  
## 2 A1 10 12.0 0.829 0.262  
## 3 A2 10 11.0 1.12 0.355  
## 4 A3 10 10.3 1.03 0.325  
## 5 A4 10 12.2 0.756 0.239

Use the following additional information about the Trts (called agents in the book): S = Standard A1 = Drug therapy with exercise and with counseling A2 = Drug therapy with exercise but no counseling A3 = Drug therapy no exercise but with counseling A4 = Drug therapy no exercise and no counseling

a\_1\_fit <- lm(Loss ~ Trt, data = WtLoss)   
anova(a\_1\_fit)

## Analysis of Variance Table  
##   
## Response: Loss  
## Df Sum Sq Mean Sq F value Pr(>F)   
## Trt 4 61.618 15.4045 15.681 4.164e-08 \*\*\*  
## Residuals 45 44.207 0.9824   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

a\_1\_emmout <- emmeans(a\_1\_fit, "Trt")  
a\_1\_emmout

## Trt emmean SE df lower.CL upper.CL  
## S 9.27 0.313 45 8.64 9.9  
## A1 12.05 0.313 45 11.42 12.7  
## A2 11.02 0.313 45 10.39 11.7  
## A3 10.27 0.313 45 9.64 10.9  
## A4 12.24 0.313 45 11.61 12.9  
##   
## Confidence level used: 0.95

1. Compare the standard agent mean versus the average of means for the four other agents.

contrast(a\_1\_emmout, list(  
 SvA = c(1, -0.25, -0.25, -0.25, -0.25))  
)

## contrast estimate SE df t.ratio p.value  
## SvA -2.12 0.35 45 -6.064 <.0001

**The standard agent mean is the lowest as comapred to the average of means for the four other agents**

1. Compare the mean for the agents with exercise against those without exercise. (Ignore the standard.)

#Exercise: A1 and A2, no exercise: A3, A4  
contrast(a\_1\_emmout, list(  
 A1\_2vA3\_4 = c(0, 0.5, 0.5, -0.5, -0.5))  
)

## contrast estimate SE df t.ratio p.value  
## A1\_2vA3\_4 0.28 0.313 45 0.893 0.3764

1. Compare the mean for the agents with counseling against those without counseling. (Ignore the standard.)

#Counseling: A1 and A2, no counseling: A3, A4  
contrast(a\_1\_emmout, list(  
 A1\_3vA2\_4 = c(0, 0.5, -0.5, 0.5, -0.5))  
)

## contrast estimate SE df t.ratio p.value  
## A1\_3vA2\_4 -0.47 0.313 45 -1.500 0.1407

1. Compare the mean for the agents with counseling versus the standard.

#Counseling: A1 and A2,   
contrast(a\_1\_emmout, list(  
 SvsA1\_3 = c(1, -0.5, 0, -0.5, 0))  
)

## contrast estimate SE df t.ratio p.value  
## SvsA1\_3 -1.89 0.384 45 -4.924 <.0001

## Q2 Binomial Distribution

{  
 n2 <- 22  
 pi2 <- 0.7  
}  
  
#A: Mean and standard deviation of Y.  
meanY <- n2\*pi2  
sdY <- sqrt(n2\*pi2\*(1-pi2))  
  
#B. P(Y ≤ 15)   
pbinom(15, size = 22, prob = 0.7)

## [1] 0.5058237

# C. P(Y < 15)   
pbinom(14, size = 22, prob = 0.7)

## [1] 0.3287493

# D. P(Y = 15)  
dbinom(15, size = 22, prob = 0.7)

## [1] 0.1770744

# E. P(15 ≤ Y < 18)  
pbinom(17, size = 22, prob = 0.7)- pbinom(15, size = 22, prob = 0.7)

## [1] 0.3296275

# F. P(Y ≥ 18)  
1 - pbinom(17, size = 22, prob = 0.7)

## [1] 0.1645488

# G. The normal approximation to P(Y ≥ 18) without continuity correction.  
(18- meanY)/sqrt(22\*.7\*.3)

## [1] 1.209629

1 - pnorm(1.21)

## [1] 0.1131394

# H. The normal approximation to P(Y ≥ 18) with continuity correction.  
(17.5- meanY)/sqrt(22\*.7\*.3)

## [1] 0.9770084

1 - pnorm(0.9770)

## [1] 0.1642846

## Q3 Chronic Pain

1. Give an estimate for the proportion of persons suffering from chronic pain that are over 50 years of age.

prop <- 424/800

1. Give a 95% confidence interval on the proportion of persons suffering from chronic pain that are over 50 years of age.

lower <- prop- 1.96\*sqrt(((0.53)\*(1-0.53))/800)  
upper <- prop + 1.96\*sqrt(((0.53)\*(1-0.53))/800)  
print(c(lower, upper))

## [1] 0.4954142 0.5645858

C. Using the data in the survey, is there substantial evidence (α = 0.05) that more than half of persons suffering from chronic pain are over 50 years of age? Give the Z statistic, p-value and conclusion **(4 pts)** H0: 0.5 vs HA: > 0.5  
Test Statistic:

Z <- (0.53 - 0.5)/sqrt((0.5\*(1-0.5))/800)  
Zalpha <- 1.645   
pvalue <- 1-pnorm(Zalpha)

**Since Z > Zalpha and p = 0.049 which is very close to 0.05, we fail to reject the null hypothesis and conclude that there is substantial evidence that more than half the persons suffering from chronic pain are over 50.**

## Q4 Defective Items

1. Give an estimate for the proportion of defective items.

prop4 <- 4/50

1. Using R, calculate a 90% confidence interval for the true proportion of defective items using the normal approximation. NOTES: (1) Use correct = TRUE (default). (2) The R CI will not match a hand calculation for this problem because R uses a different formula.

prop.test(4, 50, p = 0.01, alternative = "greater", conf.level = 0.9, correct = F)

## Warning in prop.test(4, 50, p = 0.01, alternative = "greater", conf.level =  
## 0.9, : Chi-squared approximation may be incorrect

##   
## 1-sample proportions test without continuity correction  
##   
## data: 4 out of 50, null probability 0.01  
## X-squared = 24.747, df = 1, p-value = 3.268e-07  
## alternative hypothesis: true p is greater than 0.01  
## 90 percent confidence interval:  
## 0.04316652 1.00000000  
## sample estimates:  
## p   
## 0.08

C. Using R, calculate a 90% confidence interval for the true proportion of defective items using the exact binomial method. Exact 90% CI: ()

binom.test(4, 50, p = 0.01, alternative = "greater", conf.level = 0.9)

##   
## Exact binomial test  
##   
## data: 4 and 50  
## number of successes = 4, number of trials = 50, p-value = 0.001596  
## alternative hypothesis: true probability of success is greater than 0.01  
## 90 percent confidence interval:  
## 0.03534768 1.00000000  
## sample estimates:  
## probability of success   
## 0.08

1. Is the sample size large enough for the normal approximation to be valid? Justify your response using the criteria discussed in the notes (CH10 slide 19).

nD <- 50  
piD <- 0.08  
  
3\*sqrt(piD\*(1-piD)/nD) #ok

## [1] 0.1151

1-prop4 #not ok

## [1] 0.92

**Since the second part of the sample size check was larger than pi hat, sample size is not large enough for the normal approximation to be valid**