Lab 5 Final Exam Review

Owen Cheung, Steven Tran

2022-08-17

# Question 1

point\_est = (23.12 - 17.90)  
# 5.22  
SE = sqrt((5.01^2 / 27) + (2.28^2 / 21))  
# 1.084978  
# a)  
# df = 20, so using t table for 99%, the statistic should be 2.528  
upper <- point\_est + (2.528 \* SE)  
lower <- point\_est - (2.528 \* SE)  
# (2.477177, 7.962823)  
# we are 99% confident that the average   
  
# b)  
a = 0.01  
# H0: difference in means is 0, Ha: difference in means is not 0  
t = point\_est - 0 / SE  
# t statistic = 4.81116  
df = min((27-1), (21-1))  
# df = 20  
pval <- pt(q = 4.81116, df = 20)  
# p val = 0.9999469  
# p value is large so we do not reject H0 the data does not provide convincing evidence that suggests the difference in mean sick days between blue collar workers and white collars workers is not 0

# Question 2

n <- 12  
#confidence interval of 95% for mean is (46.2, 56.9)  
  
# a) what is the margin of error for this confidence interval  
ME <- (56.9 - 46.2) / 2  
ME

## [1] 5.35

# margin of error = 5.35  
  
# b) what is the point estimate for that sample  
# finding sample mean  
samp\_mean <- 56.9 - ME  
samp\_mean

## [1] 51.55

# sample mean = 51.55  
  
# c) find the standard deviation for that sample  
# 95% = 2 SDs away from mean  
SD <- (56.9 - 46.2) / 4  
SD

## [1] 2.675

# standard deviation = 2.675

# Question 3

### 3A) Test the hypothesis H0: μ = 65 against H1: μ > 65. Find the test statistic and report your conclusion.  
  
### H0: μ = 65   
  
### Ha: μ > 65  
  
df <- c(60, 62, 63, 63, 63, 66, 66, 66, 66, 67, 67, 68, 68, 68, 69, 70, 71, 71)  
  
mean(df)

## [1] 66.33333

sd(df)

## [1] 3.124853

t <- (66-65) / ((3.12485) / sqrt(18))  
  
pt(1.357, 17, lower.tail = FALSE)

## [1] 0.0962635

### B) Test the hypothesis H0: p = 0.5 against H1: p > 0.5. Find the test statistic and report your conclusion.   
  
  
### H0: p = 0.5   
  
### H1: p > 0.5  
  
Z = (((13/18) - 0.5) / sqrt(0.5^2/18))

### H0: μ = 65

### Ha: μ > 65

### A Since p-val = 0.096 > alpha = 0.05, we fail to reject the null hypothesis that the mean height of all women is equal to 65. We do not have convincing evidence for the alternative hypothesis that the mean height of all women is greater than 65.

### H0: p = 0.5

### H1: p > 0.5

### B Since Pval = 0.0294 < alpha = 0.05, we reject the null hypothesis. We have convincing evidence that the proportion of women taller than 65 inches is greater than 0.5.

1- p(Z > 1.89) = 0.0294

# Question 4

### 4A)   
  
phat <- (127 + 131)/ 1007  
  
q <- 1 - phat  
  
  
## Z <- (0.2 - 0.35) / sqrt((phat\*q)((1/635)+(1/372)))  
  
Zstat <- 5.26  
  
  
  
### B)  
  
  
  
Lower <- 0.15 - (1.96)\*(0.029)  
Upper <- 0.15 + (1.96)\*(0.029)  
  
standard\_error <- sqrt(((0.2\*0.8) / 635) + ((0.35\*0.65) / 372))

H0: p(dem) = p(rep) H0: p(dem) =/ = p(rep)

Since p-val = 1 > 0.05, we fail to reject the null hypothesis. We do not have convincing evidence that there is a difference in the proportion of Democrats and Republicans who said that driving while impaired by marijuana poses a very serious threat.

1. It would not change. The sample statistic is the proportion of democrats subtracted by the proportion of republicans, or vice versa.
2. It would be different, because in hypothesis testing, you are pooling both proportions to create the phat and in confidence interval you re adding both confidence interval.
3. (0.09, 0.21)

We are 95% confident that the difference in democrats and republicans on their opionin that marijuana poses a very serious threat to be between 0.09 and 0.21.

# Question 5

handout1 <- read.csv("data/Handout 1.csv", stringsAsFactors = FALSE)  
head(handout1)

## COMMIT AGE SEX SCHTYPE SCHLEVEL SALARY CLASSSIZE RESOURCES AUTONOMY CLIMATE  
## 1 48 30 1 1 1 42.277 26 5 12 12  
## 2 20 30 1 1 3 40.147 21 5 11 9  
## 3 40 26 1 1 1 41.826 28 7 9 10  
## 4 20 27 1 1 3 37.000 22 7 5 10  
## 5 26 24 1 1 1 39.032 25 7 10 11  
## 6 48 30 1 1 1 40.785 23 5 10 9  
## SUPPORT  
## 1 12  
## 2 11  
## 3 12  
## 4 8  
## 5 9  
## 6 13

grouped <- handout1 %>%  
 group\_by(SEX) %>%  
 summarise\_at(vars(SALARY), list(mean\_salary = mean))  
grouped

## # A tibble: 2 × 2  
## SEX mean\_salary  
## <int> <dbl>  
## 1 1 42.8  
## 2 2 42.8

grouped\_sd <- handout1 %>%  
 group\_by(SEX) %>%  
 summarise\_at(vars(SALARY), list(sd\_salary = sd))  
grouped\_sd

## # A tibble: 2 × 2  
## SEX sd\_salary  
## <int> <dbl>  
## 1 1 2.81  
## 2 2 3.19

# seems like a difference in means question  
# hypothesis  
# H0: no difference, Ha: difference  
  
  
s <- ((80\*7.868) + (68\*10.17)) / (148)  
seed <- sqrt((s/81)+ (s/69))  
  
lower <- (0.044 - 1.96\*seed)  
upper <- (0.044 + 1.96\*seed)

Hypothesis:

Ho: μ1 = μ2

Ho: μ1 =/= μ2

Where μ1 is the population mean for the males and μ2 is the population mean for the females.

We are 95% confident that the mean difference in commitment scores with respect to sex lies between -0.92 and 1.0. Since 0 lies between these intervals, we do not have convincing evidence that there is a difference in commitment scores with respect to sex.