Lab-4

# Question 1

pnorm(145, mean = 127.1, sd = 11.7, lower.tail = F)

[1] 0.06301894

# Question 2

z <- (145 - 127.1) / 11.7  
z

[1] 1.529915

pnorm(z, lower.tail = F)

[1] 0.06301894

# Question 3

pt(2.3, df = 35, lower.tail = F)

[1] 0.01376397

college <- read.table(file = "CollegeMidwest.txt", header = T)  
head(college)

OnCampus CumGpa  
1 N 2.92  
2 N 3.59  
3 N 3.36  
4 N 2.47  
5 N 3.46  
6 Y 2.98

set.seed(90095)  
  
sample\_index <- sample(2919, size = 40)  
college\_sample <- college[sample\_index, ]  
  
x\_bar <- mean(college\_sample$CumGpa)  
x\_bar

[1] 3.36475

s <- sd(college\_sample$CumGpa)  
t\_stat <- (x\_bar - 3.5) / (s / sqrt(40))  
t\_stat

[1] -1.780155

2 \* pt(t\_stat, df = 40 - 1)

[1] 0.08284292

t.test(college\_sample$CumGpa, mu = 3.5)

One Sample t-test  
  
data: college\_sample$CumGpa  
t = -1.7802, df = 39, p-value = 0.08284  
alternative hypothesis: true mean is not equal to 3.5  
95 percent confidence interval:  
 3.211073 3.518427  
sample estimates:  
mean of x   
 3.36475

# Question 4

t.test(college\_sample$CumGpa,   
 mu = 3.5,  
 conf.level = 0.99)

One Sample t-test  
  
data: college\_sample$CumGpa  
t = -1.7802, df = 39, p-value = 0.08284  
alternative hypothesis: true mean is not equal to 3.5  
99 percent confidence interval:  
 3.159012 3.570488  
sample estimates:  
mean of x   
 3.36475

# Question 5

With the p-value > 0.01, we don’t have sufficient evidence to suggest that the mean cumulative GPA of the students at the College of the Midwest is different from 3.5.

# Question 6

ex\_vec <- c(4,1,2,6,8,5,3,7)  
  
#(a)  
ex\_vec > 4 & ex\_vec < 6

[1] FALSE FALSE FALSE FALSE FALSE TRUE FALSE FALSE

#(b)  
ex\_vec < 4 | ex\_vec > 6

[1] FALSE TRUE TRUE FALSE TRUE FALSE TRUE TRUE

# Question 7

cdc <- read.csv("cdc.csv", header = T)  
head(cdc)

state genhlth physhlth exerany hlthplan smoke100 height weight wtdesire age  
1 22 good 0 0 1 0 70 175 175 77  
2 25 good 30 0 1 1 64 125 115 33  
3 6 good 2 1 1 1 60 105 105 49  
4 6 good 0 1 1 0 66 132 124 42  
5 39 very good 0 0 1 0 61 150 130 55  
6 42 very good 0 1 1 0 64 114 114 55  
 gender  
1 m  
2 f  
3 f  
4 f  
5 f  
6 f

#(a)  
hlth\_notgood <- cdc$genhlth == "poor" | cdc$genhlth == "fair"  
  
tally(~ hlth\_notgood,   
 format = "proportion")

hlth\_notgood  
 TRUE FALSE   
0.1348 0.8652

#(b)  
hlth\_good <- cdc$exerany == "1" &   
 (cdc$genhlth == "good" | cdc$genhlth == "very good" | cdc$genhlth == "excellent")  
  
tally(~ hlth\_good, format = "proportion")

hlth\_good  
 TRUE FALSE   
0.67295 0.32705

#(c)  
hlth\_noex <- cdc$exerany == "0" &   
 (cdc$genhlth == "good" | cdc$genhlth == "very good" | cdc$genhlth == "excellent")  
  
tally(~ hlth\_noex, format = "proportion")

hlth\_noex  
 TRUE FALSE   
0.19225 0.80775

# Question 8

The statistic is of the difference between health and exercise at least once a week and health but do not exercise.

# Question 9

diff\_props <- numeric(1000)  
set.seed(147)  
  
for(i in 1:1000){  
 exerany\_shuffle\_i <- sample(cdc$exerany)  
 cond\_props\_i <- tally(hlth\_notgood ~ exerany\_shuffle\_i, format = "proportion")  
 diff\_props[i] <- cond\_props\_i[2,2] - cond\_props\_i[2, 1]  
}  
  
obs\_table <- tally(hlth\_notgood ~ exerany, data = cdc, format = "proportion")  
obs\_diff <- obs\_table[2, 2] - obs\_table[2, 1]  
  
t.test(diff\_props, mu = obs\_diff, alternative = 'two.side')

One Sample t-test  
  
data: diff\_props  
t = -779.48, df = 999, p-value < 2.2e-16  
alternative hypothesis: true mean is not equal to 0.1464438  
95 percent confidence interval:  
 -0.0002491415 0.0004876040  
sample estimates:  
 mean of x   
0.0001192312

mean(diff\_props) > 0.01 compute the proportion of diff\_props bigger then 0.01.

# Question 10

As the p-value less than 0.05, we reject the null hypothesis.

# Question 11

#(a)  
x <- mean(diff\_props)  
s <- sd(diff\_props)  
t <- (x - obs\_diff) / (s / sqrt(1000))  
t

[1] -779.479

#(b)  
mu <- mean(diff\_props)  
de <- sqrt((obs\_diff \* (1 - obs\_diff))/ 1000)  
z <- (mu - obs\_diff) / (de / sqrt(1000))  
z

[1] -413.8717

# Question 12

2 \* pnorm(z)

[1] 0

The theory-based p-value is smaller than the simulation-based p-value.