2020-11-1

yincy

11/1/2020

library(mosaic)

NCbirths <- read.csv(file = "births.csv")

### Question 1

Answer.  
Variable Habit indicates that the mother is smoked or not.  
Variable weight indicates the weight of the baby.

### Question 2

Answer.  
Variable Feduc, an numerical variable, might affect the weight of the baby weight, because well educated women known more how scientifically take care of herself during pregnancy.

Variable Premie, an categorical variable, will affect the weight of the baby, because the earlier birth of the baby the lower weight he/she will be.

### Question 3

Answer.

df <- subset(NCbirths[, c("Habit", "weight", "Feduc", "Premie")], subset = NCbirths$Habit == "NonSmoker")

### Question 4

head(df)

Habit weight Feduc Premie  
1 NonSmoker 124 13 No  
2 NonSmoker 177 9 No  
3 NonSmoker 107 12 No  
4 NonSmoker 144 12 No  
5 NonSmoker 117 10 No  
6 NonSmoker 98 14 No

### Question 5

Answer.  
(a)

mean(~ weight, data = NCbirths)

[1] 116.0591

mean(~ weight, data = NCbirths[NCbirths$Habit == "Smoker", ])

[1] 108.4225

mean(~ weight, data = NCbirths[NCbirths$Habit == "NonSmoker", ])

[1] 116.8416

### Question 6

Answer.  
(a)

sd(~ weight, data = NCbirths)

[1] 20.40667

sd(~ weight, data = NCbirths[NCbirths$Habit == "Smoker", ])

[1] 20.03352

sd(~ weight, data = NCbirths[NCbirths$Habit == "NonSmoker", ])

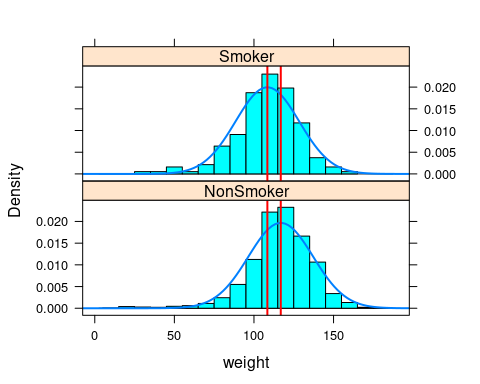
[1] 20.29014

Births <- NCbirths[NCbirths$Habit != '', ]

### Question 7

Answer.

histogram(~ weight | Habit, data = Births,   
 layout = c(1, 2),   
 width = 10,   
 density = T,   
 fit = "normal",   
 v = mean(weight ~ Habit, data = Births),   
 stripes = "vertical",   
 gcol = "red")



It seems that the baby’s weight of Non-smoker mother is a little more weight than smoker mother’s.

### Question 8

Answer.

set.seed(49)  
  
rflip(n = 20, prob = 0.4)

Flipping 20 coins [ Prob(Heads) = 0.4 ] ...  
  
T T T H T T T T T H T T T H H T H T T T  
  
Number of Heads: 5 [Proportion Heads: 0.25]

### Question 9

Answer.

set.seed(49)  
  
sim\_toss <- do(1000) \* rflip(20, 0.4)  
head(sim\_toss)

n heads tails prop  
1 20 9 11 0.45  
2 20 12 8 0.60  
3 20 4 16 0.20  
4 20 11 9 0.55  
5 20 4 16 0.20  
6 20 10 10 0.50

### Question 10

Answer.

set.seed(19)  
  
sample(1:1998, 25, replace = F)

[1] 1533 1002 1718 387 1413 338 697 707 1581 479 15 801 637 1174 620  
[16] 1963 1476 340 1384 1688 688 1391 1029 236 85

### Question 11

Answer.

set.seed(49)  
  
mean(sample(c(1, 0), 20, prob = c(0.4, 0.6), replace = T))

[1] 0.45

### Question 12

Answer.

hyp\_null <- 75 / 100 # the null probability  
hyp\_alt <- 8 / 10 # the alternative probability

### Question 13

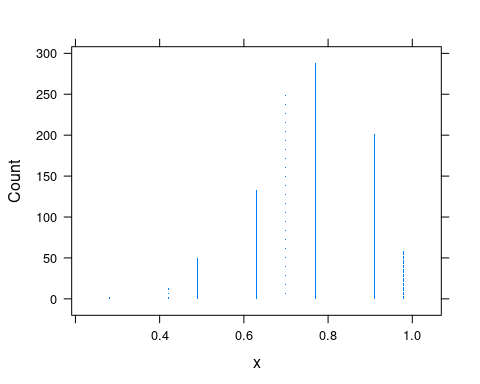
Answer.

set.seed(32)  
n <- 10  
N <- 1000  
  
sim\_shoot\_prob <- numeric(length = N)  
  
for(i in 1:N){  
 shooting = sample(c(0, 1), size = 10, replace = T, prob = c(1 - hyp\_null, hyp\_null))  
 sim\_shoot\_prob[i] <- mean(shooting)  
}

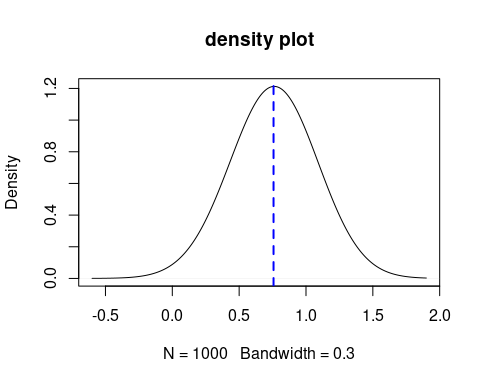
### Question 14

Answer.

dotPlot(sim\_shoot\_prob)



plot(density(sim\_shoot\_prob, kernel = "gaussian", bw = 0.3),   
 main = "density plot",   
 xlab = NULL)  
abline(v = mean(sim\_shoot\_prob), lty = 2, col = "blue", lwd = 2)



### Question 15

Answer.

mean(sim\_shoot\_prob > hyp\_alt)

[1] 0.26

### Question 16

Answer.

z\_score <- (sim\_shoot\_prob - mean(sim\_shoot\_prob)) / sd(sim\_shoot\_prob)

### Question 17

Answer.

z <- (hyp\_alt - hyp\_null) / (sqrt((hyp\_null \* (1- hyp\_null))/length(sim\_shoot\_prob)))  
pnorm(z, mean(z\_score), sd(z\_score), lower.tail = F)

[1] 0.0001303648

the player’s free throws is indeed higher than the average of NBA player.

### Question 18

hyp\_alt <- 10 / 10 # alternative probability  
z <- (hyp\_alt - hyp\_null) / (sqrt((hyp\_null) \* (1 - hyp\_null)) / length(sim\_shoot\_prob))  
pnorm(z, mean(z\_score), sd(z\_score), lower.tail = F)

[1] 0

* the player’s free throws rate is higher than the average of the NBA player.
* limitation: the test player should make multiple group of shooting and use the average shooting rate.