Homework 3

Elikem Asudo Tsatsu Gale-Zoyiku

2023-02-22

load("C:/Users/egale/OneDrive - Ashesi University/Desktop/Statistics with Probability/auto\_premiums.RData")

##Question 1

# Sample size  
n <- 50  
  
# Number of people with each blood type  
type\_o <- 21  
type\_a <- 22  
type\_b <- 5  
type\_ab <- 2  
  
#a. Probability of having type O blood  
prob\_o <- type\_o/n  
prob\_o

## [1] 0.42

#b. Probability of having type A or type B blood  
prob\_a\_b <- (type\_a + type\_b)/n  
prob\_a\_b

## [1] 0.54

#c. Probability of having neither type A nor type O blood  
prob\_not\_a\_o <- 1 - (type\_a + type\_o)/n  
prob\_not\_a\_o

## [1] 0.14

#d. Probability of not having type AB blood  
prob\_not\_ab <- 1 - type\_ab/n  
prob\_not\_ab

## [1] 0.96

#1a.

The probability that a person has type O blood is 0.42.

#1b.

The probability that a person has type A or type B blood is 0.54.

#1c.

The probability that a person has neither type A nor type O blood is 0.14.

#1d.

The probability that a person does not have type AB blood is 0.96.

##Question 2

# Total number of students  
n <- 35  
  
# Number of female students  
n\_female <- 8 + 15   
# there are 8 females from Oxford and 15 from ASU  
  
#a.probability that the student is a female  
prob\_female <- n\_female/n  
prob\_female

## [1] 0.6571429

#b. Probability that the student is from ASU given that they are female  
# there are 15 females from ASU  
n\_female\_asu <- 15   
prob\_asu\_given\_female <- n\_female\_asu/n\_female  
prob\_asu\_given\_female

## [1] 0.6521739

#2a.

The probability that a student selected at random is female is 0.657.

#2b.

The probability that the selected student comes from Arizona State University given that they are female is 0.652.

##Question 3

prob\_2\_defective <- (choose(4, 2) \* choose(20,2)) / choose(24, 4)  
prob\_2\_defective

## [1] 0.107284

prob\_none\_defective <- choose(20, 4) / choose(24, 4)  
prob\_none\_defective

## [1] 0.4559571

prob\_all\_defective <- (choose(4,4)\* choose(20,0)) / choose(24,4)  
prob\_all\_defective

## [1] 9.410879e-05

prob\_at\_least\_1\_defective <- 1 - prob\_none\_defective  
prob\_at\_least\_1\_defective

## [1] 0.5440429

#3a.

The probability that exactly two of the four screws are defective is 0.107.

#3b.

The probability that none of the screws sold are defective is 0.456.

#3c.

The probability that all of the four screws are defective is 0.00094 (practically impossible).

#3d.

The probability that at least one of the four screws sold at random is defective is 0.544.

##Question 4

# Create a frequency table of days stayed  
days\_stayed <- c(3, 4, 5, 6, 7)  
freq <- c(15, 32, 56, 19, 5)  
df <- data.frame(days\_stayed, freq)  
  
# Calculate probabilities  
# a. A patient stayed exactly 5 days.  
prob\_a <- df$freq[df$days\_stayed == 5] / sum(df$freq)  
prob\_a

## [1] 0.4409449

# b. A patient stayed less than 6 days.  
prob\_b <- sum(df$freq[df$days\_stayed < 6]) / sum(df$freq)  
prob\_b

## [1] 0.8110236

# c. A patient stayed at most 4 days.  
prob\_c <- sum(df$freq[df$days\_stayed <= 4]) / sum(df$freq)  
prob\_c

## [1] 0.3700787

# d. A patient stayed at least 5 days.  
prob\_d <- sum(df$freq[df$days\_stayed >= 5]) / sum(df$freq)  
prob\_d

## [1] 0.6299213

#4a.

The probability that a patient stayed exactly 5 days at the hospital is 0.441.

#4b.

The probability that a patient spends less than 6 days at the hospital is 0.811.

#4c.

The probability that a patient stayed at most 4 days at the hospital is 0.370.

#4d.

The probability that a patient stayed at least 5 days at the hospital is 0.630.

##Question 5

# Formula: probability = favorable outcomes / total outcomes  
  
# Define probabilities  
P\_W <- 0.04 # probability that a car needs warranty repair  
P\_US <- 0.60 # probability that car is manufactured by a U.S.-based company  
P\_W\_and\_US <- 0.025 # probability that car needs warranty repair and is manufactured by a U.S.-based company  
P\_not\_W <- 1 - 0.04  
  
#5a.  
#Probability that a car needs warranty repair given that it was manufactured by a US-based company  
  
#P(B|A) = P(A ∩ B) / P(A)  
P\_needing\_repair\_given\_US <- (P\_W\_and\_US) / P\_US  
P\_needing\_repair\_given\_US

## [1] 0.04166667

# Probability that a car needs warranty repair given that it was not manufactured by a US-based company  
  
# P(W and not US)  
P\_W\_and\_not\_US <- P\_W - P\_W\_and\_US  
  
# P(not US)  
P\_not\_US <- 1 - P\_US  
  
# P(W | not US)  
P\_W\_given\_not\_US <- P\_W\_and\_not\_US / P\_not\_US  
  
P\_W\_given\_not\_US

## [1] 0.0375

# Are need for warranty repair and location of the company manufacturing the car statistically independent?  
  
# Formula for independent events: P(A and B) = P(A) \* P(B)  
P\_W\_and\_US == P\_W \* P\_US

## [1] FALSE

#5a.

The probability that the car needs warranty repair given that it is manufactured in the United States is 0.042.

#5b.

The probability that the car needs warranty repair given that it is not manufactured in the United States is 0.038.

#5c.

Need for warranty repair and location of the company manufacturing the car are statistically dependent because the probability that a new car needs warranty repair and was manufactured in the US is not equal to the probability that a new car needs warranty repair and that it was manufactured in the US.

##Question 6

n\_women <- 7  
n\_men <- 5  
n\_choose\_women <- 3  
n\_choose\_men <- 2  
  
n\_possibilities <- choose(n\_women, n\_choose\_women) \* choose(n\_men, n\_choose\_men)  
n\_possibilities

## [1] 350

#b  
n\_digits <- 6  
n\_positions <- 4  
  
n\_possibilities <- n\_digits ^ n\_positions  
n\_possibilities

## [1] 1296

#6a.

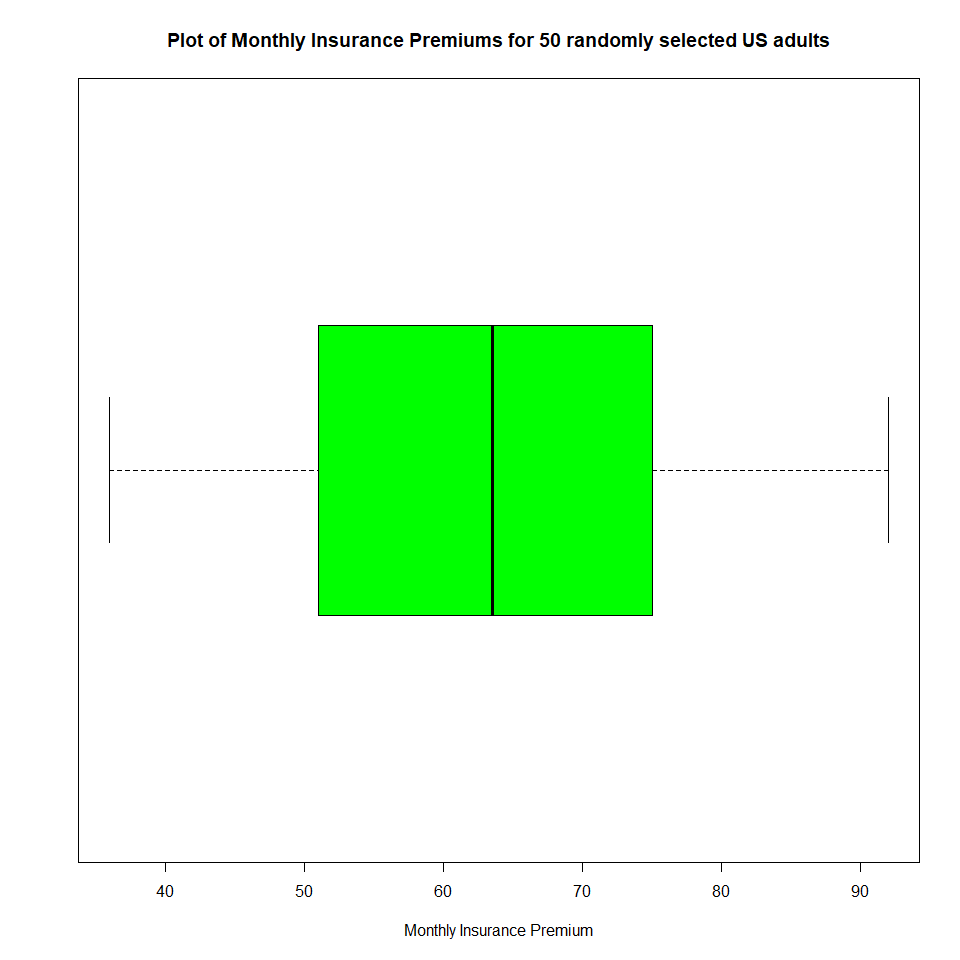
There are 350 possibilities of forming a committee with 3 women and 2 men from the club out of 7 women and 5 men.

#6b.

There are 1296 ways different cards can be made if she uses the digits 1, 2, 3, 4, 5, 6 and repetitions are permitted.

##Question 7

#7a.  
boxplot(data$Premium,main="Plot of Monthly Insurance Premiums for 50 randomly selected US adults",xlab='Monthly Insurance Premium',col='green',horizontal = TRUE)



#7b.

The distribution of auto insurance premiums is roughly symmetric because the median line seems to lie perfectly in between the 1st and 3rd quantile lines. This means that the distribution is evenly spread about the median. I expect the distribution to follow the empirical rule, because from the boxplot, there don’t seem to be outliers in the distribution. This means that the mean will more accurately describe the center of the data. The accuracy of the mean, and the absence of outliers, means that the standard deviation will be a more accurate description of the spread of the distribution. As such, I expect the distribution to follow the empirical rule.

#7c.

#7c.  
summary(data$Premium)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 36.00 51.00 63.50 62.98 74.75 92.00

mean(data$Premium)

## [1] 62.98

sd(data$Premium)

## [1] 15.17852

The mean of the sample distribution of auto premiums is USD 62.98, and the standard deviation is USD 15.18. The mean of 62.98 indicates that the average auto premium in the distribution is approximately USD 63. The standard deviation of 15.17852 indicates that the auto premiums in the distribution are spread out around the mean, with most premiums falling within one standard deviation of the mean, or between approximately USD 48 and USD 79. This provides a measure of variability for the distribution.

#7d.

# Calculate the mean and standard deviation  
mean\_premiums <- mean(data$Premium)  
sd\_premiums <- sd(data$Premium)  
  
# Calculate the range within one standard deviation of the mean  
lower\_bound <- mean\_premiums - sd\_premiums  
upper\_bound <- mean\_premiums + sd\_premiums  
  
# Calculate the probability that a randomly selected insurance premium falls within this range  
#P(lower\_bound < X < upper\_bound)  
prob\_within\_one\_sd <- pnorm(upper\_bound, mean = mean\_premiums, sd = sd\_premiums) - pnorm(lower\_bound, mean = mean\_premiums, sd = sd\_premiums)  
  
prob\_within\_one\_sd

## [1] 0.6826895

The probability that a randomly chosen premium falls within 1 standard deviation of the mean premium is 0.683. This is extremely close to what I expected based on the empirical rule, because for an empirical distribution, 68 percent of all values in the distribution fall within 1 standard deviation of the mean. This implies that the probability that a randomly picked value from a distribution that follows the empirical rule, falls withing 1 standard deviaton of the mean, is 68% or 0.68. This is close to the value that I calculated.

#7e.

# Calculate the mean and standard deviation  
mean\_premiums <- mean(data$Premium)  
sd\_premiums <- sd(data$Premium)  
  
# Calculate the range within one standard deviation of the mean  
lower\_bound <- mean\_premiums - 2\*sd\_premiums  
upper\_bound <- mean\_premiums + 2\*sd\_premiums  
  
# Calculate the probability that a randomly selected insurance premium falls within this range  
#P(lower\_bound < X < upper\_bound)  
prob\_within\_two\_sd <- pnorm(upper\_bound, mean = mean\_premiums, sd = sd\_premiums) - pnorm(lower\_bound, mean = mean\_premiums, sd = sd\_premiums)  
  
prob\_within\_two\_sd

## [1] 0.9544997

The probability that a randomly chosen premium falls within 2 standard deviations of the mean premium is 0.954. This is extremely close to what I expected based on the empirical rule, because for an empirical distribution, 95 percent of all values in the distribution fall within 2 standard deviation of the mean. This implies that the probability that a randomly picked value from a distribution that follows the empirical rule, falls within 2 standard deviations of the mean, is 95% or 0.95. This is close to the value that I calculated for the probability that a randomly chosen premium falls within 2 standard deviations of the mean.