Homework Assignment 3

David Abeiku Saah

2023-02-26

library(readr)  
  
load("auto\_premiums.RData")

### Question 1

n <- 50;  
O <- 21;  
A <- 22;  
B <- 5;  
AB <- 2;  
  
p\_A <- A/n;  
p\_B <- B/n;  
p\_AB <- AB/n  
  
# a.   
p\_O <- O/n;p\_O

## [1] 0.42

# b.  
p\_A\_or\_B <- p\_A + p\_B;p\_A\_or\_B

## [1] 0.54

# c.  
p\_A\_nor\_O <- 1 - (p\_A + p\_O);p\_A\_nor\_O

## [1] 0.14

# d.  
p\_no\_AB <- 1 - p\_AB;p\_no\_AB

## [1] 0.96

1. The probability that a person has type O blood is 0.42.
2. The probability that a person has type A or type B blood is 0.54.
3. The probability that a person has neither type A nor type O blood is 0.14.
4. The probability that a person does not have type AB blood is 0.96.

### Question 2

students\_OU <- 15;  
students\_ASU <- 20;  
n <- students\_ASU + students\_OU;  
f\_students\_OU <- 8;  
m\_students\_ASU <- 5;  
f\_students\_ASU <- students\_ASU - m\_students\_ASU;  
  
# a.   
f\_students <- f\_students\_OU + f\_students\_ASU;  
p\_f\_student <- f\_students / n;round(p\_f\_student, 4)

## [1] 0.6571

# b.  
p\_f\_students\_ASU <- (f\_students\_ASU / n);  
pp = p\_f\_students\_ASU / p\_f\_student;round(pp, 4)

## [1] 0.6522

1. The probability that the student is a female is 0.6571.
2. If the selected student is female, the probability that the student comes from Arizona State University is 0.6522.

### Question 3

n <- 24;  
d <- 4;  
p\_d = d / n;  
  
# a.  
round(dbinom(2, 4, p\_d), 4)

## [1] 0.1157

# b.  
round(dbinom(0, 4, p\_d), 4)

## [1] 0.4823

# c.  
round(dbinom(4, 4, p\_d), 4)

## [1] 8e-04

# d.   
round(1 - dbinom(0, 4, p\_d), 4)

## [1] 0.5177

1. The probability that exactly 2 are defective is 0.1157
2. The probability that none is defective is 0.4823.
3. The probability that all are defective is 0.0008.
4. The probability that at least one is defective is 0.5177.

### Question 4

days\_3 <- 15;  
days\_4 <- 32;  
days\_5 <- 56;  
days\_6 <- 19;  
days\_7 <- 5;  
total <- 127;  
  
# a.   
round(days\_5 / total, 4)

## [1] 0.4409

# b.  
round(1 - ((days\_6 + days\_7) / total), 4)

## [1] 0.811

# c.   
round((days\_3 + days\_4) / total, 4)

## [1] 0.3701

# d.   
round((days\_5 + days\_6 + days\_7) / total, 4)

## [1] 0.6299

1. The probability that a patient stayed exactly 5 days is 0.4409.
2. The probability that a patient stayed less than 6 days is 0.811.
3. The probability that a patient stayed at most 4 days is 0.3701.
4. The probability that a patient stayed at least 5 days is 0.6299.

### Question 5

p\_need\_w <- 0.04;  
p\_us <- 0.60;  
p\_need\_w\_and\_us <- 0.025;  
  
# a.   
round(p\_need\_w\_and\_us / p\_us, 4)

## [1] 0.0417

# b.   
p\_no\_us = 1 - p\_us;  
p\_no\_us\_and\_need\_w = p\_need\_w - p\_need\_w\_and\_us;  
round(p\_no\_us\_and\_need\_w / p\_no\_us, 4)

## [1] 0.0375

# c.   
# Test for independence: is P(need\_w\_and\_us) == P(need\_w) \* P(us)  
p\_need\_w\_and\_us == p\_need\_w \* p\_us

## [1] FALSE

1. If a company based in the United States manufactured a particular car, the probability that the car needs warranty repair is 0.0417.
2. If a company based in the United States did not manufacture a particular car, the probability that the car needs warranty repair is 0.375.
3. From question a and b, knowing whether or not a particular car is manufactured from the United states affects the probability (the probabilities are different). Also, from a mathematical standpoint, the location of the company manufacturing the car and warranty repair can be determined by comparing P(needing warranty and location of manufacturing company), and the product of P(needing warranty) and P(location of manufacturing company). From our comparison, the values are not the same. Hence, the two events are not statistically independent.

### Question 6

men <- 5;  
women <- 7;  
total <- men + women;  
  
# a.  
# To calculate combinations, combn() function is used. However, it generates all possible sequences. To count the number of sequences, ncol() function is used.  
# 7C3 \* 5C2  
ncol(combn(7, 3)) \* ncol(combn(5, 2))

## [1] 350

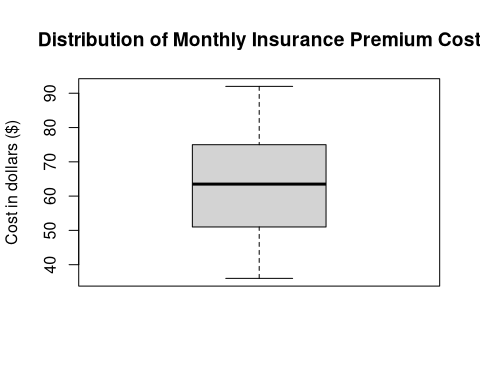
# b.   
# The number of digits available are 6, and there are 4 slots.  
possible\_ways <- 6 \* 6 \* 6 \* 6;possible\_ways

## [1] 1296

1. There are 350 possibilities.
2. The manager can make 1296 different cards.

### Question 7a

boxplot(main ="Distribution of Monthly Insurance Premium Cost", data$Premium, ylab = "Cost in dollars ($)")



### Question 7b

summary(data$Premium)

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

The distribution of auto insurance premiums is not symmetric because the mean (62.98) and the median (63.50) are not the same. Since the median is greater than the mean, the distribution is left-skewed.

I do not expect the distribution to follow the empirical rule because the rule holds only for normal distributions (that is, when mean = median). However, the auto insurance premiums distribution is left-skewed, which does not satisfy the criteria for the empirical rule.

### Question 7c

mean(data$Premium)

## [1] 62.98

sd(data$Premium)

## [1] 15.17852

Based on the sample data, a US adult pays 62.98 dollars for monthly automobile insurance premium on average. However, this cost varies greatly for different adults since the standard deviation of the distribution is 15.17 dollars.

### Question 7d

pos\_sd1 = mean(data$Premium) + sd(data$Premium);  
neg\_sd1 = mean(data$Premium) - sd(data$Premium);  
  
bet\_1\_sd = data[(data$Premium < pos\_sd1) & (data$Premium > neg\_sd1), ];  
prob\_sd1 = length(bet\_1\_sd$Premium)/length(data$Premium);prob\_sd1

## [1] 0.62

The probability that an insurance premium chosen at random is within one standard deviation of the mean is 0.62. Based on the empirical formula, the expected probability is 0.68. 0.62 (actual probability) is close, but varies from the expected empirical probability (0.68) value by 0.06.

### Question 7e

pos\_sd2 = mean(data$Premium) + (2 \* sd(data$Premium));  
neg\_sd2 = mean(data$Premium) - (2 \* sd(data$Premium));  
  
bet\_2\_sd = data[(data$Premium < pos\_sd2) & (data$Premium > neg\_sd2), ];  
prob\_sd2 = length(bet\_2\_sd$Premium)/length(data$Premium);prob\_sd2

## [1] 1

The probability that an insurance premium chosen at random is within two standard deviations of the mean is 1. This means it is certain that an insurance premium chosen at random falls within two standard deviations of the mean. However, based on the empirical formula, the expected probability is 0.95, which is 0.5 less than the actual probability calculated. The probabilities are close, but not the same.