Probability and Statistics

Assignment 2

By Himanshu Sardana

Question 1

(a) Suppose there is a chest of coins with 20 gold, 30 silver and 50 bronze coins. You randomly draw 10 coins from the chest. Write an R code that will give us the sample space for the experiment.

```
chest = c(rep("gold", 20), rep("silver", 30), rep("bronze", 50))

# To get a random sample
sample_space = sample(chest, 10)

# To get all possible combinations
sample_space = combn(chest, 10)
```

(b) In a surgical procedure, the chances of success and failure are 90% and 10% respectively. Write an R code that will give us the sample space for the next 10 surgeries.

```
surgery = c(rep("success", 9), "failure")

sample_space = sample(surgery, 10, replace = TRUE)
```

Question 2

A room has n people and each has an equal chance of being born on any of the 365 days of the year. (For simplicity, ignore leap years.) What is the probability that 2 people in the room have the same birthday?

(a) Use an R simulation to estimate this value for various n

```
# Probability that no two people have the same birthday
prob = 1

# Number of people in the room
n = 10

for(i in 1:(n-1)){
  prob = prob * (365 - i)/365
}

# Probability that at least two people have the same birthday
prob = 1 - prob
```

```
# Inbuilt function to calculate the probability
pbirthday(n)
```

(b) Find the smallest value of n for which the probability exceeds 0.5

```
n = 1
prob = 1

while(prob > 0.5){
  n = n + 1
  prob = 1
  for(i in 1:(n-1)){
    prob = prob * (365 - i)/365
  }
  prob = 1 - prob
}

print(n)

# Inbuilt function
qbirthday(0.5)
```

Question 3

Write an R function for computing conditional probability. Call this function to do the following:

Suppose the probability of the weather being cloudy is 40%. Also suppose the probability of rain on a given day is 20% and the probability of clouds on a rainy day is 85%. if its cloudy outside on a given day, what is the probability that it will rain that day?

```
conditional_prob = function(p_a, p_b, p_b_given_a){
  return(p_b_given_a * p_a / p_b)
}

p_a = 0.4
p_b = 0.2

p_b_given_a = 0.85

conditional_prob(p_a, p_b, p_b_given_a)
```

Question 4

The iris dataset is a built-in dataset in R that contains measureents on 4 different attributes (in centimeters) for 150 flowers from 3 different species. Load this dataset and do the following

- Print first few rows of this dataset

```
head(iris)
```

- Find the structure of this dataset

```
str(iris)
```

- Find the range of data regarding the sepal length of the flowers

```
range(iris$Sepal.Length)
```

- Find the mean of the sepal width of the flowers

```
mean(iris$Sepal.Width)
```

- Find the median of the petal length of the flowers

```
median(iris$Petal.Length)
```

- Find the first and third quartile of the petal width of the flowers

```
quantile(iris$Petal.Width, c(0.25, 0.75))

# Interquartile range
IQR(iris$Petal.Width)
```

- Find the standard deviation and variance

```
sd(iris$Sepal.Length)
var(iris$Sepal.Length)
```