Assessment-HarvadX-Combinations-and-Permutations

NOTEBOOK BY ISABELLA HEDER

Assessment HarvardX: Combinations and Permutations

Data Science: Probability Course on EDX

EXERCISES:

Exercise 1:

Two teams, say the Celtics and the Cavs, are playing a seven game series. The Cavs have a 60% chance of winning each game. What is the probability that the Celtics win at least one game? Remember that the Celtics must win one of the first four games, or the series will be over!

1. **Instructions:**
   - Calculate the probability that the Cavs will win the first four games of the series.
   - Calculate the probability that the Celtics win at least one game in the first four games of the series.

–> Use p_cavs_win4 as the probability that the Cavs will win the first four games of the series.

```
p_cavs_win4 <-  0.6^4

1 - p_cavs_win4
```

```
## [1] 0.8704
```

The solution was based in the **Multiplication Rule** for **independent** events

P(A, B and C) = P(A).P(B).P(C)

The Cavs had 60% chance of winning EACH game, no matter the outcome of the other games.

–> For the second part:

If the Cavs lose one game = the Celtics win one.

That's why I used:

1 (meaning 100%) - probability that the cavs will win the first four = probability that the celtics will win at least one game

Exercise 2:

Create a Monte Carlo simulation to confirm your answer to the previous problem by estimating how frequently the Celtics win at least 1 of 4 games. Use B $\leftarrow$ 10000 simulations. The provided sample code simulates a single series of four random games, simulated_games.

1. **Instructions:**

   - Use the replicate function for B $\leftarrow$ 10000 simulations of a four game series. The results of replicate should be stored in a variable named celtic_wins.
   - Use the set.seed function to make sure your answer matches the expected result after random sampling;
   - Create an object called celtic_wins that replicates two steps for B iterations: (1) generating a random four-game series named simulated_games, then (2) determining whether the simulated series contains at least one win for the Celtics.

```r
B <- 10000

set.seed(1)

celtic_wins <- replicate(B, {
  simulated_games <- sample(c("lose", "win"), 4, replace = TRUE, prob = c(0.6, 0.4))
  any(simulated_games == "win")})

mean(celtic_wins)
```

```
## [1] 0.8757
```

The solution was based in **Monte Carlo Simulations** for Celtics winning a game, using the *same values as the previous exercise.*

B = number of times the exercise wanted the Monte Carlo simulation to run

Celtic_wins:

Using **replicate** function to repeat the simulation for the number of games.

Using **sample** function to randomly select outcomes (win or lose) for each game.

- 4 games, each with 40% chance of winning and 60% chance of losing.

Replace = TRUE; means that every time the simulation selects a value, it goes back to the mass, not taking it away from the total.

Using **any( )** to check if at least one game received the "win" value.

–

Using **mean( )** to calculate the proportion of simulations where the Celtics won at least one game.

–

Exercise 3:

Say you've drawn 5 balls from a box that has 3 cyan balls, 5 magenta balls, and 7 yellow balls, with replacement, and all have been yellow. What is the probability that the next one is yellow?

1. **Instructions:**

- Assign the variable p_yellow as the probability of choosing a yellow ball on the first draw.
- Using the variable p_yellow, calculate the probability of choosing a yellow ball on the sixth draw.

```
box <- rep(c('cyan', 'magenta', 'yellow'), times = c(3, 5, 7))
p_yellow <- mean(box == 'yellow')
print(p_yellow)
```

```
## [1] 0.4666667
```

```
print(p_yellow)
```

```
## [1] 0.4666667
```

For this exercise I decided to **create a vector** using the amount of balls by color.

To calculate the probability of choosing a yellow ball I used the mean( ) function, to calculate the proportion of yellow balls over the total.

It would be the same as: 7/15

To calculate the probability of drawing a yellow ball on the sixth draw, knowing that they've drawn 5 balls from the box and all have been yellow, I used the **same proportion!**

Because this exercise was done **WITH replacement**, meaning, every time a ball was drawn, it would immediately go back to the box.

Exercise 4:

If you roll a 6-sided die once, what is the probability of not seeing a 6? If you roll a 6-sided die six times, what is the probability of not seeing a 6 on any of those rolls?

1. **Instructions:**

- Assign the variable p_no6 as the probability of not seeing a 6 on a single roll.
- Then, calculate the probability of not seeing a 6 on six rolls using p_no6. Do not assign it to a variable.

```
p_no6 <- 5/6
p_no6
```

```
## [1] 0.8333333
```

```
p_no6^6
```

```
## [1] 0.334898
```

For the final exercise:

I was asked to calculate the probability of not seeing a 6 after rolling a 6-sided die one time only.

Since the die has 6 sides: only one of them has the value "6"

–> So not seeing a 6 means I would see any of the other numbers (1, 2, 3, 4 and 5) –> that would mean 5/6

Not seeing a 6 on 6 rolls:

I used the **multiplication rule** for **independent** events, because each roll does not change because of the other outcomes.

P(A, B, C, D, E and F) = P(A) . P(B) . P(C) . P(D) . P(E) . P(F)