Assessment 02 - Independence and Multiplication Rule

Gabriele Mineo - Harvard Data Science Professional

Independence

Imagine you draw two balls from a box containing colored balls. You either replace the first ball before you draw the second or you leave the first ball out of the box when you draw the second ball.

Under which situation are the two draws independent of one another?

Remember that two events A and B are independent if $\Pr(A \text{ and } B) = \Pr(A)P(B)$.

Possible Answers

- You don't replace the first ball before drawing the next.
- You do replace the first ball before drawing the next. [X]
- Neither situation describes independent events.
- Both situations describe independent events.

Sampling with replacement

Say you've drawn 5 balls from the 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?

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.

```
cyan <- 3
magenta <- 5
yellow <- 7

# Assign the variable 'p_yellow' as the probability that a yellow ball is drawn from the box.
p_yellow <- yellow / (cyan + magenta + yellow)
# Using the variable 'p_yellow', calculate the probability of drawing a yellow ball on the sixth draw.
p_yellow
```

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

Rolling a die

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 roll?

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`.

```
# Assign the variable 'p_no6' as the probability of not seeing a 6 on a single roll.
p_no6 <- 5/6
```

```r
# Calculate the probability of not seeing a 6 on six rolls.
p_no6**6
```

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

Probability the Celtics win a game

Two teams, say the Celtics and the Cavs, are playing a seven game series. The Cavs are a better team and 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!

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.

```r
# Assign the variable `p_cavs_win4` as the probability that the Cavs will win the first four games of t
p_cavs_win4 <- 0.6^4

# Using the variable `p_cavs_win4`, calculate the probability that the Celtics win at least one game in
1 - p_cavs_win4
```

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

Monte Carlo simulation for Celtics winning a game

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 <- 10000` simulations.

The provided sample code simulates a single series of four random games, `simulated_games`.

Instructions

- Use the `replicate` function for `B <- 10000` simulations of a four game series.
- Within each simulation, replicate the sample code to simulate a four-game series named `simulated_games`.
- Then, use the `any` function to indicate whether the four-game series contains at least one win for the Celtics.
- Use the `mean` function to find the proportion of simulations that contain at least one win for the Celtics out of four games.

```r
# This line of sample code simulates four random games where the Celtics either lose or win. Each game
simulated_games <- sample(c("lose","win"), 4, replace = TRUE, prob = c(0.6, 0.4))

# The variable 'B' specifies the number of times we want the simulation to run. Let's run the Monte Car
B <- 10000

# Use the `set.seed` function to make sure your answer matches the expected result after random samplin
set.seed(1)

# Create an object called `celtic_wins` that first replicates the sample code generating the variable c
celtic_wins <- replicate(B, {
  simulated_games <- sample(c("lose","win"), 4, replace = TRUE, prob = c(0.6, 0.4))
  any(simulated_games=="win")
```

```r
})

# Calculate the frequency out of B iterations that the Celtics won at least one game. Print your answer
mean(celtic_wins)
```

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