Assessment 01 - Introduction to Descrete probability

Gabriele Mineo - Harvard Data Science Professional

Probability of cyan - generalized

In the edX exercises for this section, we calculated some probabilities by hand. Now we'll calculate those probabilities using R.

One ball will be drawn at random from a box containing: 3 cyan balls, 5 magenta balls, and 7 yellow balls.

What is the probability that the ball will be cyan?

Instructions

- Define a variable **p** as the probability of choosing a cyan ball from the box.
- Print the value of **p**.

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

# Assign a variable `p` as the probability of choosing a cyan ball from the box
p <- cyan / (cyan + magenta + yellow)

# Print the variable `p` to the console
print(p)
```

```
## [1] 0.2
```

Probability of not cyan - generalized

We defined the variable **p** as the probability of choosing a cyan ball from a box containing: 3 cyan balls, 5 magenta balls, and 7 yellow balls.

What is the probability that the ball you draw from the box will NOT be cyan?

Instructions

- Using the probability of choosing a cyan ball, **p**, calculate the probability of choosing any other ball.

```
# `p` is defined as the probability of choosing a cyan ball from a box containing: 3 cyan balls, 5 mage
# Using variable `p`, calculate the probability of choosing any ball that is not cyan from the box
1 - p
```

```
## [1] 0.8
```

Sampling without replacement - generalized

nstead of taking just one draw, consider taking two draws. You take the second draw without returning the first draw to the box. We call this sampling without replacement.

What is the probability that the first draw is cyan and that the second draw is not cyan?

Instructions

- Calculate the **conditional probability** `p_2` of choosing a ball that is not cyan after one cyan ball has been removed from the box.
- Calculate the **joint probability** of **both** choosing a cyan ball on the first draw and a ball that is not cyan on the second draw using `p_1` and `p_2`.

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

# The variable `p_1` is the probability of choosing a cyan ball from the box on the first draw.
p_1 <- cyan / (cyan + magenta + yellow)

# Assign a variable `p_2` as the probability of not choosing a cyan ball on the second draw without rep
p_2 <- (magenta + yellow) / (cyan + magenta + yellow - 1)
# Calculate the probability that the first draw is cyan and the second draw is not cyan.
p_1 * p_2
```

```
## [1] 0.1714286
```

Sampling with replacement - generalized

Now repeat the experiment, but this time, after taking the first draw and recording the color, return it back to the box and shake the box. We call this sampling with replacement.

What is the probability that the first draw is cyan and that the second draw is not cyan?

Instructions

- Calculate the probability `p_2` of choosing a ball that is not cyan on the second draw, with replacement.
- Next, use `p_1` and `p_2` to calculate the probability of choosing a cyan ball on the first draw and a ball that is not cyan on the second draw (after replacing the first ball).

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

# The variable 'p_1' is the probability of choosing a cyan ball from the box on the first draw.
p_1 <- cyan / (cyan + magenta + yellow)

# Assign a variable 'p_2' as the probability of not choosing a cyan ball on the second draw with replac
p_2 <- (magenta + yellow) / (cyan + magenta + yellow)
# Calculate the probability that the first draw is cyan and the second draw is not cyan.
p_1 * p_2
```

```
## [1] 0.16
```