Rolling with Marbles

Directions: Follow along with the slides and answer the questions in **BOLDED** font in your journal.

Where we left off

- In the last lab, we looked at how we can use computer simulations to compute estimates of simple probabilities.
- *Such as the probability of drawing a certain color of marble out of a bag.*
- What if we wanted to compute estimates for more complex questions?
- *Such as drawing 2 marbles of the same color from the bag?*
- In this lab, we will explore how to compute these types of estimates.

Start by creating a bag of marbles

- For this lab, we will need a bag of marbles to draw from.
- Use RStudio to create a bag of `marbles` that contains 39 blue marbles, 19 green marbles, and 42 red marbles.
- Be sure to name it: `marbles`.
- **Compute the probability, by hand, of randomly selecting a blue marble.**
- **Estimate the probability of selecting a blue marble by sampling 300 draws from your bag of marbles.** (Refer to Lab 2.3 if you need more guidance.)
- **How far off was your estimated probability from the actual probability?**

Simulating single draws

- From the previous slide, we used the following code to simulate 300 *single* draws from our bag of marbles.
- A *single draw* refers to the action of picking a marble at random, noting its color, and then putting it back.

```
sample(marbles, size = 300,
       replace = TRUE)
```

- We can also use the `sample` command to simulate *multiple draws*.
- A *multiple draw* refers to the action of picking 2 or more marbles from the bag, noting the combination of colors, and then putting the marbles back.

Simulating draws with replacement

- We can simulate multiple draws in two ways:
- We can either pick our first marble, record its color and put it back, and then pick our second marble.
- These are called draws *with replacement* (or *independent* draws):

```
sample(marbles, size = 2, replace = TRUE)
```

- **Run this code and write down the colors of the marbles you randomly chose**.

Simulating draws without replacement

- Or we can pick our first marble, record its color, but leave it out of the bag, and then pick our second marble.
- These are called draws *without replacement* (or *dependent* draws):

```
sample(marbles, size = 2, replace = FALSE)
```

- **Run this code and write down the colors of the marbles you randomly chose**.
- **How does not placing the marbles back into the bag affect the probability of selecting subsequent marbles?**

Drawing multiple marbles with do-loops

- Simulating *single* draws from our bag of marbles was easy.
- It only required us to use the `sample` function.
- Simulating *multiple* draws is a little more complex because we need to use **loops**.
- A **loop** is when we ask the computer to carry out a certain action over-and-over-and-over.
- The type of *loop* we will use is called a *do-loop*.
- You'll see where the name comes from in the next few slides.

Our first do-loop (Step 1)

- Start by writing the code you want to *loop over*.
- *Loop over* is programmer-speak for *do many, many times*.
- In our case, we want to simulate drawing 2 marbles with replacement.
- That is, we want to pick a marble at random, note its color, put it back in the bag, and randomly select a 2nd marble.
- **Write down the code you previously used to simulate drawing 2 marbles with replacement.**

Our first do-loop (Step 2)

- After we have written the code that we want to loop over, we can add in the loop.
- To tell R to `do` something `n` times, we write:

```
do(n) * code-to-loop-over
```

- `do(n) *`, can be translated to mean: *Do the following code over-and-over exactly* ***n*** *times*.
- **Use a `do` loop to simulate 2 draws from your bag of marbles, with replacement, 300 times.**
- **Assign these 300 loops to the name `samples`. Write down the code you used.**

Get ready to compute!

- Type `head(samples)` into the console. What gets printed should look similar to the following:

```
##      V1    V2
## 1   red green
## 2   red   red
## 3 green green
## 4 green   red
## 5   red   red
## 6 green  blue
```

- Notice that our first draw has been given the name `V1` and our second draw `V2`.
- We will need this information to calculate our probabilities.

Using simulations

- We can now use our simulations to estimate compound probabilities.
- Remember, these are *estimates* and not exact probabilities.
- The more loops we use, the more exact our solutions will be (and the longer our code will take to run).
- Actual data scientists sometimes have to wait hours, or even days, for their simulations to run!

Making our estimates

- The first step in making our estimates is to count how often our different outcomes occurred.
- *How often did we draw two reds? A red and then a green? etc.*
- We could use the tally function to compute our estimates directly:

```
tally(V1~V2, data = samples)
```

- But then what would we do if we looked at more than 2 marbles at a time?

Counting our outcomes

- Instead, we use the following code to count the different outcome combinations:

```
tally(~V1:V2, data = samples)
```

- The `V1:V2` is what tells R to create all possible combinations of outcomes.
- If we had drawn 5 marbles instead of just 2, we could write the following to tally up all of the possible outcomes:

```
tally(~V1:V2:V3:V4:V5, data = samples)
```

From counts to estimated probabilities

- Now that we've counted how often each outcome occurred, we can compute our estimated probabilities.
- Start by counting the outcomes we are interested in.
- Then divide by the total number of simulations.
- For example, to estimate the probabilty of drawing 2 red marbles:
- Take the number of simulations that resulted in `red:red`.
- Divide this number by 300.
- **What outcomes would you add to estimate the probability of drawing *at least* 1 blue marble?**

On your own

- Sample 3 independent draws from our bag of marbles and `do` 300 simulations.
- **Write down the code you used to create your 300 simulations.**
- **Write down how often each combination occurred.**
- Use your simulations to estimate the probability of:
- **Drawing exactly 3 blue marbles?**
- **Drawing a red marble first, then a blue marble, and finally a green marble?**
- **Drawing at least 2 green marbles?**