Week 4: Simulation

SOC6302 Winter 2023

In this lab, we will simulate flipping a coin four times and recording the number of heads that turn up. First, load in the tidyverse package:

```
library(tidyverse)
```

1 Flipping a coin four times

Now we need to define all possible events that can occur in our trial — in this case, we flip a coin and either get a head or a tail:

```
possible_events <- c("H", "T")
```

We can then simulate flipping the coin four times using the sample function:

```
coin_flips <- sample(possible_events,
                     size = 4, # how many times we repeat the trial
                     replace = TRUE) # allows a particular event to occur more than once
coin_flips
```

```
[1] "T" "T" "H" "T"
```

We can then sum up the number of heads we obtained in that particular trial:

```
number_of_heads <- sum(coin_flips=="H")
```

2 Repeating the experiment over and over

Now imagine we repeat the act of tossing a doing four times over and over. We can simulate this in R with the help of a for loop. First, specify the number of trials we would like to do. Let's start with 5:

```r
number_of_trials <- 5
```

Second, we want to define a new data frame in which we can store the results of each trial (i.e. the number of heads). This data frame has two columns, the trial number and the corresponding number of heads. Note the : in the code below means "to", so we are saying the trial number goes from 1 to the specified number of trials.

```r
trial_data <- tibble(trial_number = 1:number_of_trials,
                     number_heads = NA) # just put NA for now because we haven't done the
trial_data
```

```
# A tibble: 5 x 2
  trial_number number_heads
         <int> <lgl>
1            1 NA
2            2 NA
3            3 NA
4            4 NA
5            5 NA
```

Now we can repeat the experiment (tossing a coin four times) over and over in a for loop. A for loop repeats the code within the curly brackets for the number of times specified in the first parentheses. Within the loop, we flip the coin four times, calculate the number of heads, and then store that number in our data frame:

```r
for(i in 1:number_of_trials){ # "for all numbers i in 1 to the number of trials (i.e. 1 to
  this_coin_flip <- sample(possible_events, size = 4, replace = TRUE)
  this_number_of_heads <- sum(this_coin_flip=="H")
  trial_data[i,2] <- this_number_of_heads
}
```

3 Calculate summary statistics

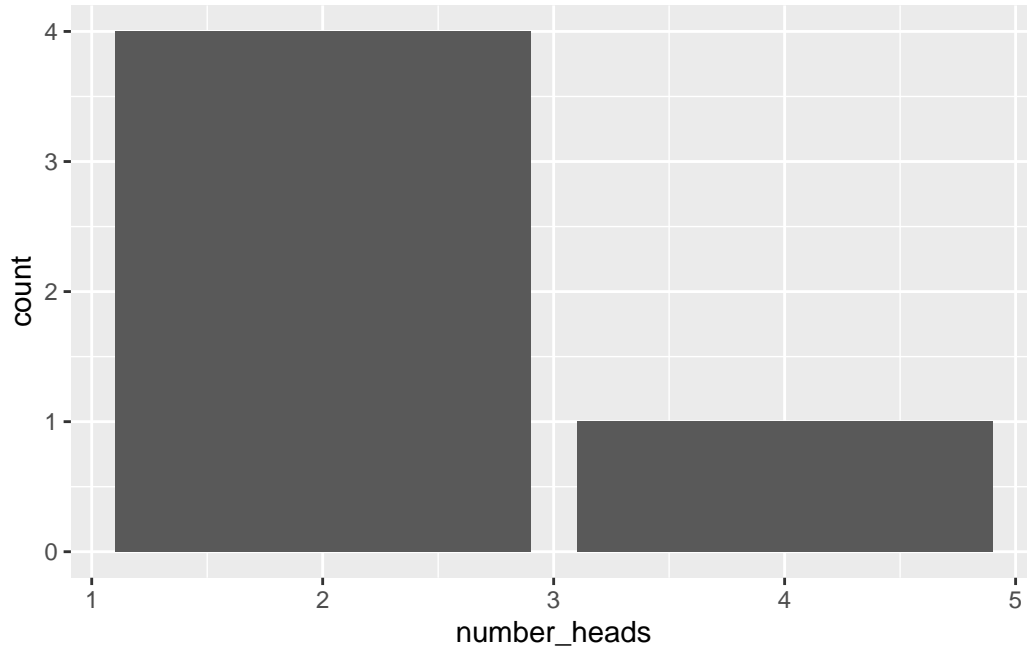Now we can summarize the data across all the trials. In particular, calculate the mean and variance:

```
trial_data |>
  summarize(mean = mean(number_heads),
            variance = var(number_heads))
```

```
# A tibble: 1 x 2
   mean variance
  <dbl>    <dbl>
1   2.4      0.8
```

4 Plot the results

We can also plot the distribution of the number of heads as a bar chart:

```
ggplot(trial_data, aes(number_heads)) +
  geom_bar()
```

5 Exercises

1. Increase the number of trials to be 1000 and run the code again. What do you notice about the mean number of heads? What do you notice about the shape of the bar chart?
2. (ggplot review). Change the color of the `fill` in the bar chart above. Give the chart a title and change the x axis.