Lab 3 : Probability- Part 1

Estimate the probability of getting X heads over four independent flips

1. Load tydiverse
2. The sample() function in R samples elements from a vector x. First, check the help page (?sample). Then, define a vector c("HEADS", "TAILS") and take four samples from it: sample(c("HEADS", "TAILS"), 4, TRUE)
3. If you define heads as one and tails as zero, you can write the same command as follows: sample(0:1, 4, TRUE)
4. Count the number of heads by simply summing up the values in the vector: **sample**(0:1, 4, TRUE) **%>%** sum() #NOTE: The pipe symbol %>%is not part of base R, but is part of an add-on package
5. Run this function 5 times (you can use the up arrow in the console to avoid having to type it over and over). How many heads do you get?
6. Repeat the experiment a whole bunch more times via: replicate(20, sample(0:1, 4, TRUE) %>% sum())
7. Repeat it 50 times: heads50 <- replicate(50, sample(0:1, 4, TRUE) %>% sum())
8. Estimate the probability of each of the outcomes (0, 1, 2, 3, 4 heads) by counting them up and dividing through by the number of experiments. Do this by putting the experiments in a data_frame() and then using count():

```
dat50 <- data_frame(x = heads50) %>%
group_by(x) %>%
summarise(n = n(), p = n / 50)
```

What we have done here is created a variable x in a new data_frame object that contains the results from our 50 experiments, grouped it by the values in x (0, 1, 2, 3, or 4), and counted the number of times each value was observed in **summarise**(n = n()), and then calculated the probability by dividing the newly created variable n by the number of experiments (50)

9. Plot a histogram of the outcomes using ggplot2: **ggplot**(dat50, aes(x, p)) + **geom_bar**(stat = "identity")
10. Repeat the Monte Carlo simulation, but with 10,000 experiments instead of just 50:
11. Again, put the vector in a data frame and calculate counts and probabilities of each outcome using group_by() and summarise():

```
heads10k <- replicate(10000, sample(0:1, 4, TRUE) %>% sum())
dat10k <- data_frame(x = heads10k) %>%
group_by(x) %>%
summarise(n = n(), p = n / 10000)
```

12. Plot a histogram of the new outcomes using ggplot2.
13. What is the probability of getting *two or more heads* in four throws? The outcomes meeting the criterion are 2, 3, or 4 heads:

```
dat10k %>%
filter(x >= 2) %>%
summarise(p2 = sum(p))
```

Note : You can add probabilities for various outcomes together as long as the outcomes are *mutually exclusive*