Module 3: Basic Probability Concepts

Empirical Probabilities: Dice Example

We can start by simulating rolling two dice and finding the sum of the rolls:

```
# define the sample space of a fair, 6 sided die
S_die <- c(1,2,3,4,5,6)

# "roll" the first die
(die1 <- sample(S_die, 1))
```

```
[1] 1
```

```
# "roll" the second die
(die2 <- sample(S_die, 1))
```

```
[1] 1
```

```
# find the sum of the two rolls
die1 + die2
```

```
[1] 2
```

Simulation with 10 trials

Now, let's simulate rolling the pair of dice 10 times. After each roll, we will find the sum and record it.

```r
# first, we will initialize a vector to store the sum results from our rolls
sum10 <- vector(mode = "numeric", length = 10)

for(i in 1:10){

  # roll the dice
  die1 <- sample(S_die, 1)
  die2 <- sample(S_die, 1)

  # find and record the sum
  sum10[i] <- die1 + die2
}
```

Since there are few enough trials, we can few the raw data easily:

```r
sum10
```

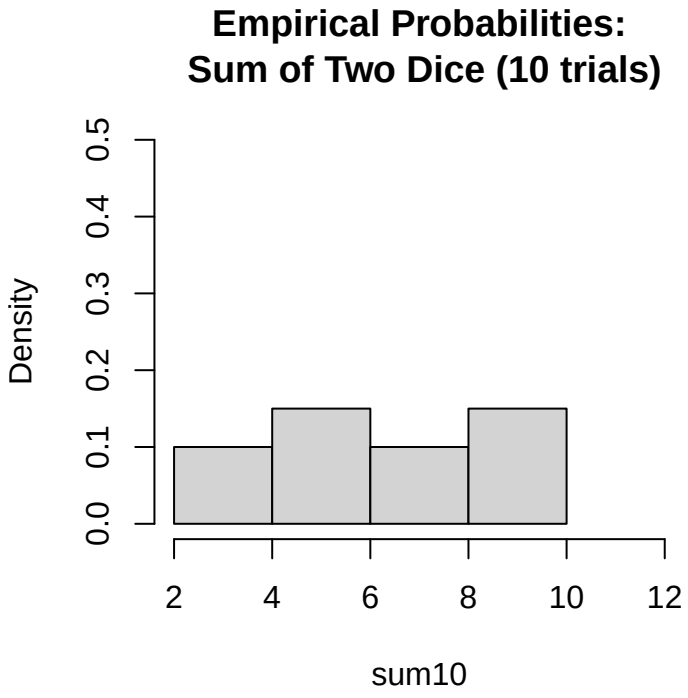
```
 [1]  7  9  7  2  6  6  4 10 10  5
```

We are more interested in the summary of our sums:

```r
# frequency table of our 10 trials
table(sum10)
```

```
sum10
 2  4  5  6  7  9 10
 1  1  1  2  2  1  2
```

```r
# histogram of our 10 trials
hist(sum10,
     freq = FALSE, # to show densities instead of frequencies
     breaks = 4,
     xlim = c(2, 12),
     ylim = c(0,0.5),
     main = "Empirical Probabilities:\n Sum of Two Dice (10 trials)")
```



Simulation with 100 trials

This code will be very similar, but now we will run the simulation 100 times:

```
# first, we will initialize a vector to store the sum results from our rolls
sum100 <- vector(mode = "numeric", length = 100)

for(i in 1:100){

  # roll the dice
  die1 <- sample(S_die, 1)
  die2 <- sample(S_die, 1)

  # find and record the sum
  sum100[i] <- die1 + die2
}
```

To view a summary of our data from the 100 trials:

```
# frequency table of our 10 trials
table(sum100)
```

```
sum100
 2  3  4  5  6  7  8  9 10 11 12
 5  3 12 14 11 14 19 10  7  4  1
```

```r
# histogram of our 10 trials
hist(sum100,
     freq = FALSE, # to show densities instead of frequencies
     breaks = 8,
     xlim = c(2, 12),
     ylim = c(0,0.5),
     main = "Empirical Probabilities:\n Sum of Two Dice (100 trials)")
```