Assignment 2

Harshad Kumar Elangovan - 19200349

15/10/2019

The data frame of two dice are created using the below code. These values are then used for calculating the probability of getting the counts of green and red dice.

```
green = c(1:6,1:6,1:6,1:6,1:6,1:6)
red = sort(green,decreasing = FALSE)
dice = data.frame(red,green)
dice$sum = dice$red + dice$green
print(dice)
```

```
##    red green sum
## 1    1     1   2
## 2    1     2   3
## 3    1     3   4
## 4    1     4   5
## 5    1     5   6
## 6    1     6   7
## 7    2     1   3
## 8    2     2   4
## 9    2     3   5
## 10   2     4   6
## 11   2     5   7
## 12   2     6   8
## 13   3     1   4
## 14   3     2   5
## 15   3     3   6
## 16   3     4   7
## 17   3     5   8
## 18   3     6   9
## 19   4     1   5
## 20   4     2   6
## 21   4     3   7
## 22   4     4   8
## 23   4     5   9
## 24   4     6  10
## 25   5     1   6
## 26   5     2   7
## 27   5     3   8
## 28   5     4   9
## 29   5     5  10
## 30   5     6  11
## 31   6     1   7
## 32   6     2   8
## 33   6     3   9
## 34   6     4  10
## 35   6     5  11
## 36   6     6  12
```

Part A

What is the probability of getting a total of 8 when you roll two dice?

```
counts= table(dice$sum)
counts = data.frame(counts)

counts$prob = counts$Freq / 36
print(counts)
```

```
##    Var1 Freq       prob
## 1     2    1 0.02777778
## 2     3    2 0.05555556
## 3     4    3 0.08333333
## 4     5    4 0.11111111
## 5     6    5 0.13888889
## 6     7    6 0.16666667
## 7     8    5 0.13888889
## 8     9    4 0.11111111
## 9    10    3 0.08333333
## 10   11    2 0.05555556
## 11   12    1 0.02777778
```

From the calculation above, we see that the probability of getting a total of 8 is **0.13888889**.

What is the probability of getting an even number when you roll two dice?

From the data frame, we see that there are five values that are even. So taking a sum of the probabilities of those even numbers, we get **0.5** (i.e) 50 percentage of the probability.

Part 2

The sample of 50 values is taken using the dice data frame and relative frequency of these sample is computed.

	sample50	Freq	relfreq
1	2	1	0.02
2	3	6	0.12
3	5	6	0.12
4	6	7	0.14
5	7	13	0.26
6	8	7	0.14
7	9	6	0.12
8	10	2	0.04
9	11	1	0.02
10	12	1	0.02

In your sample of 50, what proportion of the time did you get a total of 8?

0.14 is the proportion of the time I got a total of 8 in the given dataset.

What proportion of the time did you get an even numebr?

The sample produced **0.36** proportion of times on even number.

Part C

In your sample of 500, what proportion of the time did you get a total of 8?

	sample500	Freq	relfreq
1	2	12	0.024
2	3	30	0.060
3	4	36	0.072
4	5	41	0.082
5	6	76	0.152
6	7	81	0.162
7	8	81	0.162
8	9	63	0.126
9	10	43	0.086
10	11	26	0.052
11	12	11	0.022

The sample produced **0.162** proportion of times for a total of 8.

What code did you use to find this?

```
sample500 = sample(dice$sum, 500,replace=TRUE)
prob500_table=table(sample500)
prob500 = data.frame(prob500_table)
prob500$relfreq = prob500$Freq/500
```

Compare the proportion of times you got an 8 in each sample with the true probability of getting an 8. What do you notice?

The true probability of getting an 8 is 0.13888889. For sample of 50 values, I got 0.14 as proportion and for 500 sample values, the proportion is 0.162. From this we see that the proportion increases compared with the actual probability of 8. This is because of increase in count of 8 in sample 50(8 - 7 times in the dataset) and in sample 500(8 - 81 times in the dataset). So, the proportion of times of 8 in each sample increases with increase in sample values : 0.162(sample500)/0.14(sample50)/0.13888889(true probability).