Q SCI 381
Introduction to Probability and Statistics
Spring 2018
Laboratory #3 Homework

Using the "students.csv" file to complete the first two questions below, the data contain gender, hair color, and eye color. Follow the same general procedures you used in the Skills exercise to answer the questions below. **REMEMBER: capitalization, spacing, and parentheses are important when using datasets and functions in R**!

Include all code for each question, plus output from R or graphs, and answer all sub-questions.

To review typed HW, please check my Github page.
https://github.com/UWJiangShao/R_Studio_HW/blob/master/Tiger%20Shao%20R%20studio%20HW3.R

1) Let's work with contingency tables with two variables.
   a. Create a Contingency Table using the gender and hair color data located in the "students.csv" file. Add margins and use the table to answer the following questions.

```
> setwd("C:\\Users\\Jiang Shao\\Desktop\\bees")
> data = read.csv("students.csv")
>
> #Create a Contingency Table using the gender and hair
> #color data located in the “students.csv” file. Add margins
> #and use the table to answer the following questions
>
> conting_hair <- addmargins(table(data$Gender, data$Hair))
> conting_hair

         black Blond brown Red Sum
  female    52    81   143  37 313
  male      56    46   143  34 279
  Sum      108   127   286  71 592
```

b. What is the probability of having brown hair if you are male?

```
>
> #What is the probability of having brown hair if you are male?
> prob_male_brown <- conting_hair[2, 3] / conting_hair[2, 5]
> prob_male_brown
[1] 0.5125448
>
```

c. What is the probability of having red hair if you are female?

```
>
> #What is the probability of having red hair if you are female
> prob_female_red <- conting_hair[1, 4] / conting_hair[1, 5]
> prob_female_red
[1] 0.1182109
```

d. If you have blond hair, what is the probability you are female?

```
>
> #If you have blond hair, what is the probability you are f
emale
> prob_blond_female <- conting_hair[1, 2] / conting_hair[3,
2]
> prob_blond_female
[1] 0.6377953
>
```

2) Let’s work with contingency tables with three variables.

a. Create two Contingency Tables - one for males and one for females with eye color and hair color. Add margins and use the table to answer the following questions. Note: order is important when entering data into the table() function; hair, eye, gender should be the order of variables.

```
> #Create two Contingency Tables - one for males and one for
> #females with eye color and hair color.
> conting_sex <- addmargins(table(data$Eye, data$Hair, data$Gender))
> conting_sex
, ,  = female

        black Blond brown Red Sum
  Blue      9    64    34   7 114
  Brown    36     4    66  16 122
  Green     2     8    14   7  31
  Hazel     5     5    29   7  46
  Sum      52    81   143  37 313

, ,  = male

        black Blond brown Red Sum
  Blue     11    30    50  10 101
  Brown    32     3    53  10  98
  Green     3     8    15   7  33
  Hazel    10     5    25   7  47
  Sum      56    46   143  34 279

, ,  = Sum

        black Blond brown Red Sum
  Blue     20    94    84  17 215
  Brown    68     7   119  26 220
  Green     5    16    29  14  64
  Hazel    15    10    54  14  93
  Sum     108   127   286  71 592
```

b. What are the probabilities of having hazel eyes for males and for females?

```
>
> #What are the probabilities of having hazel eyes for males
> #and for females?
> prob_hazel_male <- conting_sex[4, 5, 2] / conting_sex[5, 5, 2]
> prob_hazel_female <- conting_sex[4, 5, 1] / conting_sex[5, 5, 1]
> prob_hazel_male
[1] 0.1684588
> prob_hazel_female
[1] 0.1469649
>
```

c. What is the probability of having blue eyes if you are blond (for females and for males)?

```
>
> #What is the probability of having blue eyes if you are
> #blond (for females and for males)?
> prob_blue_female <- conting_sex[1, 5, 1] / conting_sex[5, 5, 1]
> prob_blue_male <- conting_sex[1, 5, 2] / conting_sex[5, 5, 2]
> prob_blue_female
[1] 0.3642173
> prob_blue_male
[1] 0.3620072
>
```

d. For males, is having brown eyes independent of hair color? Why or why not? To figure this out, calculate whether the probability of having brown eyes is the same as or different from the conditional probability of having brown eyes given that you have each hair color.

```
> #For males, is having brown eyes independent of hair color?
> #Why or why not? To figure this out, calculate whether the
> #probability of having brown eyes is the same as or different
> #from the conditional probability of having brown eyes given that
> #you have each hair color.
> prob_brown_male <- conting_sex[2, 5, 2] / conting_sex[5, 5, 2]
> prob_brown_male
[1] 0.3512545
> prob_brown_BLAhair <- conting_sex[2, 1, 2] / conting_sex[5, 1, 2]
> prob_brown_BLAhair
[1] 0.5714286
> prob_brown_BLOhair <- conting_sex[2, 2, 2] / conting_sex[5, 2, 2]
> prob_brown_BLOhair
[1] 0.06521739
> prob_brown_BROhair <- conting_sex[2, 3, 2] / conting_sex[5, 3, 2]
> prob_brown_BROhair
[1] 0.3706294
> prob_brown_REDhair <- conting_sex[2, 4, 2] / conting_sex[5, 4, 2]
> prob_brown_REDhair
[1] 0.2941176
```

It looks like if you have brown eyes, you will have low chance of having blond hair. Other color looks pretty random and independent.

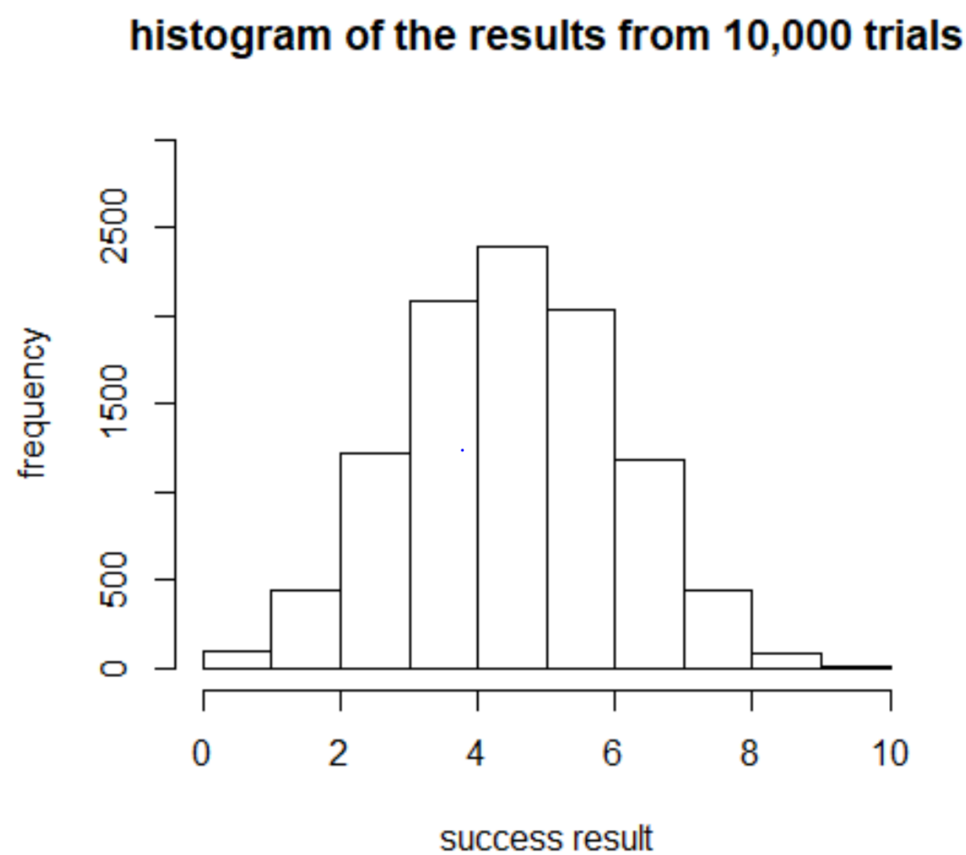
3) We are going to practice with data simulation.

a. Using the rbinom() function, 'simulate' 10,000 trials of flipping a coin 10 times (you should end up with a vector of 10,000 numbers ranging from 0-10).

```
>
> rbinom(10000, 10, 0.5)
   [1] 6 5 2 8 3 3 5 6 6 5 2 4 4 5 4 6 4 4 6 4 4 1 5 5 4 7 4 3 5 5 4 6 5 7
  [35] 6 5 7 8 7 5 6 6 5 7 7 3 3 6 6 6 3 4 3 5 5 3 6 6 7 6 7 6 5 5 6 4 5 5
  [69] 5 6 5 3 6 3 3 4 3 6 4 7 6 2 0 6 3 5 5 3 5 4 5 2 6 2 5 5 5 3 5 3 8 4
 [103] 4 7 7 4 5 4 3 6 5 8 7 7 6 5 2 4 4 3 5 3 7 5 2 3 6 2 4 6 5 3 6 6 6 4
 [137] 7 6 4 6 2 6 5 6 6 6 3 6 5 6 6 5 7 5 4 8 1 5 5 5 7 5 5 8 5 4 5 7 5 5
 [171] 7 5 5 3 5 5 2 8 7 5 8 2 6 7 6 4 5 6 6 4 5 5 6 5 7 4 5 7 7 5 3 4 2 3
 [205] 4 3 4 6 6 7 4 3 6 3 4 3 4 5 3 4 5 4 6 5 6 6 7 4 3 5 4 3 5 7 7 7 8 8
 [239] 4 6 2 5 6 5 5 4 6 7 4 4 7 7 4 5 5 7 4 3 5 6 3 6 6 5 5 6 4 2 6 7 5 6
```

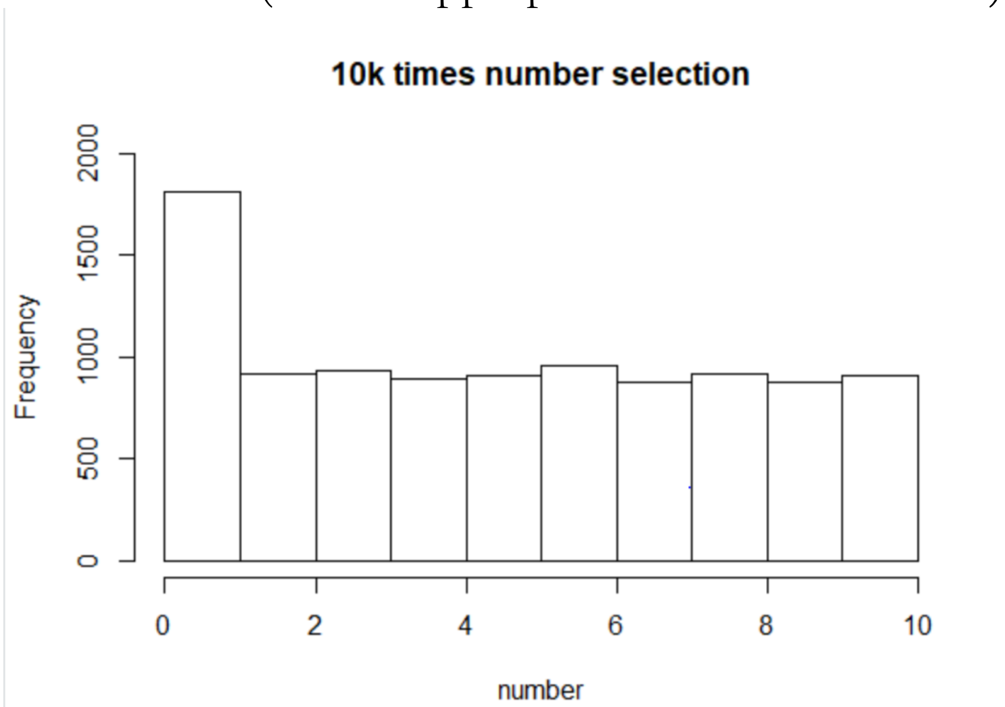
(too many data, not list all)

b. Make a histogram of the results from these 10,000 trials. Add a title and label the axes of your histogram.




```
 [ reached getOption("max.print") -- omitted 9000 entries ]
> hist(f, xlab = "success result", ylab = "frequency", main = "histogram of the results from 10,00
, 10), ylim = c(0, 3000), breaks = 10)
>
> flip10000 <- sample(x <- c(0:10), 10000, replace = T)
> flip10000
   [1]  0  5  4  8  7  5  1  4  3  9  9  3  8  3  6  1  0  4  7  3  5  6  2  1  3  3  0 10
  [29]  9  0 10  0 10  2  2  2  0 10  8  9  5  6  9  3  4  2  6 10  9  4  5  7  6  2  4 10
  [57]  3  0  3  4  9  6  1  4 10  9  6  5  7  1  1  7  6  2  2  4  0  8  4  0  1  3  9  6
  [85]  6  0  5  6  0  5  7  6  0 10  4  0  0 10  2  4  6  3  0  2  4  0  3 10  2 10  9  9
 [113]  1  8  5  8  8  2  5 10  3 10  7 10  9  3  5  6  5  7  9  2  8  9  8  6  5  0  0  9
```

c. Use the sample() function to select a number between 0 and 10 and repeat this 10,000 times. Make a histogram of the results (include appropriate title and axis labels)



> flip10000 <- sample(x <- c(0:10), 10000, replace = T)
> flip10000

```
[ reached getOption("max.print") -- omitted 9000 entries ]
> hist(flip10000, xlab = "number", ylab = "Frequency", xlim = c(0, 10), ylim = c(0, 2000), breaks
es number selection")
>
```

d. Compare the two histograms, what shape should each be and why? Should they look the same or different and how do your actual histograms look (compared to what you think they should look like)?
First one is perfect bell shape. Second one should be right skewed. Since the first one has perfect 0.5 probability, and second one is more random, it is naturally random like this, so they should be different.

4) There are 8 teams in the Pacific Division of the NHL.

a. How many different ranking orders are possible?
b. Of the 8 teams in the Pacific Division, only 3 are guaranteed to move on to the playoffs. How many possible combinations are there?

```
> #4)   There are 8 teams in the Pacific Division of the NHL.
> #How many different ranking orders are possible?
> #Of the 8 teams in the Pacific Division, only 3
> #are guaranteed to move on to the playoffs. How many
> #possible combinations are there?
> factorial(8)
[1] 40320
> factorial(8)/(factorial(5)*factorial(3))
[1] 56
>
```

5) There is a total of 20 fans sitting in a row at the seventh game of the Stanley Cup Finals, 4 are crying because their team lost, 8 are screaming because their team won, 6 are trying to leave so they can beat the traffic, and 2 are asleep because they had bit too much to drink. How many different seating arrangements are possible for this collection of 20 fans?

```
> #How many different seating arrangements are possible
> #for this collection of 20 fans?
> factorial(20)/(factorial(4)*factorial(8)*factorial(2)*factorial(6))
[1] 1745944200
> |
```