STATISTICS 147 EXAM I, Part 2: R
Summer 2020; 40 pts; Show all work!

NAME: (Please Print) ⎯⎯⎯⎯ Wesley Chang ⎯⎯⎯⎯

ID: (last 4 #s only) ⎯⎯⎯⎯ 0996 ⎯⎯⎯⎯

General Information:

Complete the following using R.

♣ Do all of your work in an R script. Save your **R script** as **e1Rscript_XX.R** where XX = your initials.

♣ (2 pts) Include the following titles:

```
Statistics 147 Exam I, Part 2 R
Summer 2020
Your Name
Question  X       (where X = question number)
```

♣ You should place `comments` in your R code. (2 pts)

♣ Be sure to include code in your R script to **change the R working directory** to the folder that contains your data files, `dogdiet.dat` and `dogjudge.dat`. This should allow you to read in files using only the file name.

♣ For any plots, place them in an **MS Word/.docx** document. Save it as **ex1Rsu20_XX.docx**, where XX = your initials. Be sure to make the question number very clear.

[♣] When you have *completed* your exam, copy the entire contents of your **Console** and paste it to the end of your **MS Word/.docx** document, **ex1Rsu20_XX.docx**, where XX = your initials.

♣ When you have finished your exam, email the following files to lklei001@ucr.edu:

(i) Your completed exam PDF.

(ii) Your **R script/.R** document.

(iii) Your **MS Word/.docx** document containing your plots and the entirety of the Console from your R session.

1. Carly and Shannon work in Research & Development for a premium dog food company that was interested in marketing a new food for adult dogs that are severely overweight. Carly and Shannon are considering 3 recipes for the new "diet" food and have decided to test them using severely overweight Labrador Retriever adult dogs (of approximately the same age and health condition). Twenty-four severely overweight Labs were selected at random and then each randomly assigned to one of the three recipes of dog food. After an appropriate amount of time, the weight loss, measured in pounds, was recorded.

   The following data has been saved in a file named **dogdiet.dat** which you should have downloaded from Blackboard (iLearn). (Each column represents a diet formula.)

```
file name: dogdiet.dat
F1     F2     F3
4.95   7.70   6.47
4.04   5.81   5.23
3.72   6.61   6.75
2.21   6.07   7.75
2.48   8.04   4.70
3.31   5.96   6.92
3.50   7.30   6.01
2.90   7.46   5.67
```

   NOTE:

   ♠ The data is located in a datafile named **dogdiet.dat**.

   ♠ The headings are included in the data file. The actual data begins on line 3.

   (i) Write the R code to read in and print out the data. (3 pts)

   (ii) Add the appropriate lines of code to make the columns accessible individually. 2 pts)

2. Refer to Question 1.

    (i) Jose and Dung decided they should generate some descriptive statistics. Find the mean, median and standard deviation for **Formula1 (F1)** and complete the following table. (3 pts)

Diet	Mean	Median	Standard Deviation
Formula1 (F1)	3.38875	3.405	0.8818882

    (ii) Create a histogram for the weight loss of the **Formula1**, using breaks from 2 to 5 in increments of 0.5. Be sure to include as many of the following colors (Red, Green, Blue, Purple, Orange, Brown) that are necessary and the following title on your chart: **Histogram of Weight Loss for Formula 1**. Sketch your chart in the space below. (4 pts)

see word document
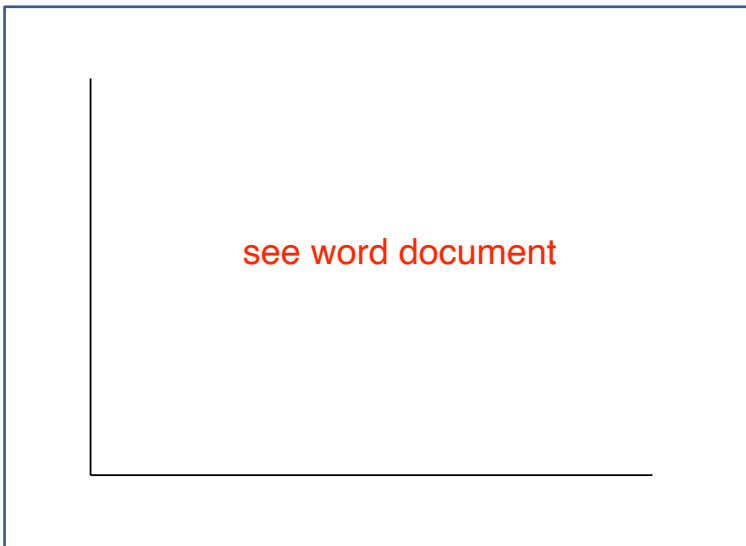
3. *Disc dog* is the more generic name for what is commonly called *Frisbee dog*. In *disc dog* competitions, dogs and their human *flying disc* throwers compete in events such as distance catching and somewhat choreographed freestyle catching. Scores differ among judges, even when the same performance is being evaluated. Khaldoun and Thomas have been selected to be judges in a local *disc dog* competition. The scores, reported by Khaldoun (J1) and Thomas (J2), for 11 competitors are located in a data file named **dogjudge.dat**,

```
file name: dogjudge.dat
dog          J1  J2
Abby         83  84
Cody         79  78
Dexter       77  73
Dusty        94  96
Kali         79  78
Korra        88  90
Lakota       95  90
MaxG         93  87
Mercedes     88  84
Shadow       96  95
MaxS         85  88
```

NOTE: The two lines of headings are included in the data file.

(i) Write the appropriate R code to read in and print out the data. (2 pts)

(ii) Fan and Rene want to see a plot of the data. Create a scatterplot of **J1** vs **J2**, using **J1** as the horizontal axis. Be sure to include the title: **Scatterplot of Judge 1 vs Judge 2**. Use 75 to 100 for the x-axis limits and 70 to 100 for the y-axis limits. Use symbol pch = 16. Sketch your plot in the space below. (3 pts)
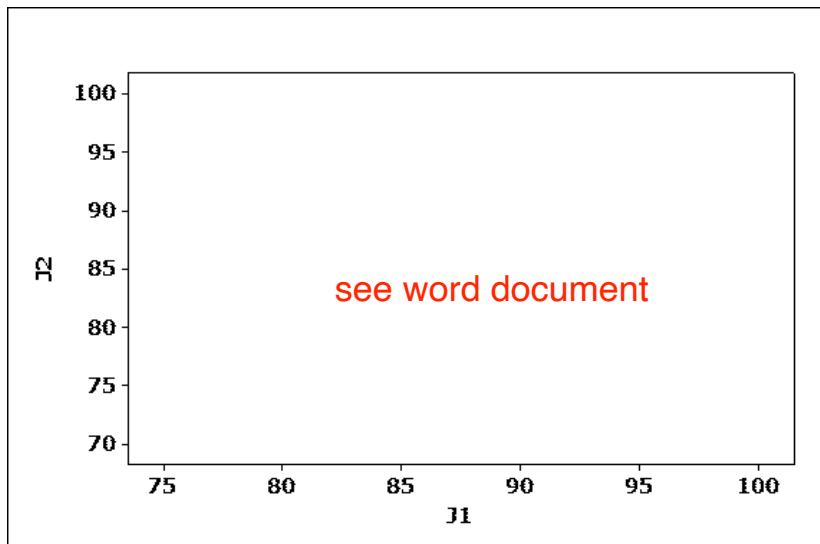
4. Suppose Ernesto and Natascha know the specification limits for a plastic bottle filling machine are 40.0 ± 2.50 ounces. In other words, the filling process is considered to be functioning at an appropriate level (functioning "in control") if the fill in the bottles is between 37.5 ounces and 42.5 ounces. Suppose the filling process is normally distributed with a mean of 40 ounces and a standard deviation of 1.75 ounces. They select a bottle at random from the assembly line. Let X = amount of fill in the bottle.

    (i) Write the appropriate R code to find the probability that a bottle is overfilled (in other words, the amount of fill is more than 42.5 ounces). (2 pts)

ANSWER: 0.07656373

    (ii) Write the appropriate R code to find the probability that the amount of fill is between 37 ounces and 41 ounces. (2 pts)

ANSWER: 0.6729073

    (iii) Write the appropriate R code to find the 96th percentile (i.e., find x such that $P(X \leq x) = 0.96$). (1 pt)

ANSWER: 43.0637

5. Sarah and Ziguo found that, according to a recent survey, 35% of current students prefer the change in class times that begins Fall Quarter 2020. Sarah and Ziguo were interested in examining this claim. A random sample of 24 students was selected.

Let X = # of students that prefer the change in class times. Use this information to complete the following.

    (i) Write the appropriate R code to find the probability exactly 10 of the students prefer the change in class times. (2 pts)

ANSWER: 0.1300175

    (ii) Write the appropriate R code to find the probability between 7 and 10, inclusively, of the students prefer change in class times. (2 pts)

**ANSWER:** 0.4591518

6. Lauren was interested in examining the distribution of colors in a bag of *Skittles* candy. She recorded the following data:

```
blue yellow red blue red green orange blue yellow orange green green yellow yellow
red red red red blue red green blue red yellow orange blue blue
```

(i) Enter and save the data in a variable named **skittles1**. Be sure to print the data as a check! (2 pts)

(ii) Create a frequency table for the colors of *Skittles* in the bag. Complete the following table. (2 pts)

Color	blue	green	orange	red	yellow
Frequency	7	4	3	8	5

(a) (1 pt) Which color/colors appears/appear most often? ___red___

(b)(1 pt) Which color/colors appears/appear least often? ___orange___

(iii) Create a vertical bar chart of the colors of *Skittles*. (4 pts)

(a) Make sure the colors of the bar match the name of the color.

(b) Include the following title: **Distribution of Skittles Colors in the Bag**

see word document
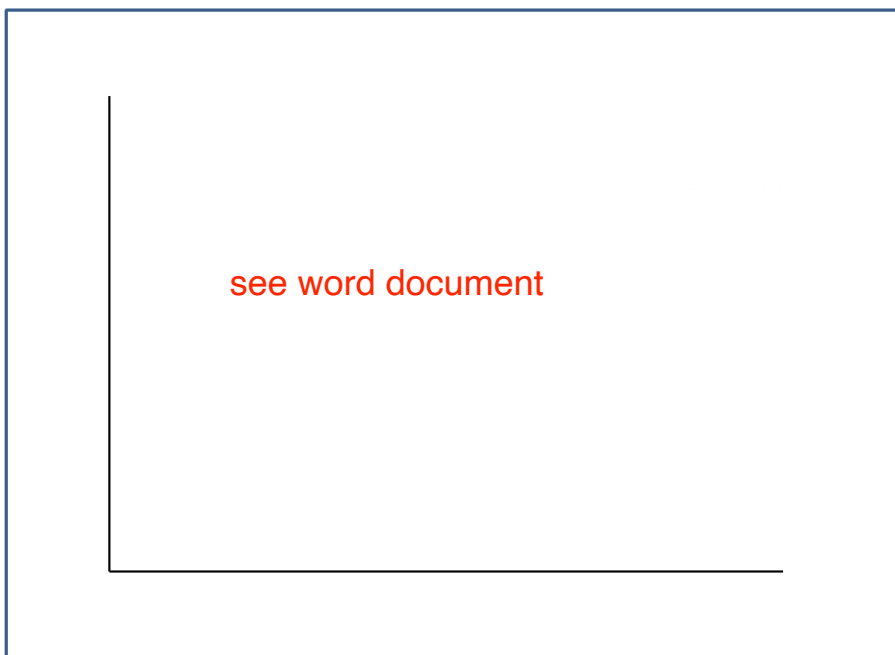
Sketch your chart.

7. Suppose $y = 2x^3 + \sqrt{2x}$ for x = 0.2,0.3,0.4,0.5.

(i) Enter the values of x and write the R code to calculate the values of y. (2 pts)

(ii) Write the R code to combine the two columns together into one table and complete the following table. (2 pts)

	x	y
[1,]	0.2	0.6484555
[2,]	0.3	0.8285967
[3,]	0.4	1.0224272
[4,]	0.5	1.2500000

Turn in your completed exam paper, along with your R script code and the Word document containing your plots and Console output. Have a nice day!!

Luke & Ruihan