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

NAME: (Please Print) Wesley Chang **Signature:** Wesley Chang

ID: (last 4 #s only) 0996

General Information:

- ♣ 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 II, 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. (2 pts)

  [♣] 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, **LastName_FirstName_147_Ex2_R.docx**, where the last and first name are your own.

- ♣ 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.
   - ♠ μ_i = true average weight loss for formula i, and σ_i^2 = variance of the weight loss for formula i, $i = 1, 2, 3$

   (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) Natascha and Dung decided they should test for normality of the data for each formula. Use the **Anderson-Darling** test. Complete the following. (2 pts)

Formula	p-value	Reasonable to assume normality? (Answer YES or NO)
F1	0.9363	yes
F2	0.3616	yes
F3	0.9777	yes

   (ii) Sarah and Rene want to determine whether it is reasonable to assume equality (homogeneity) of variances. Perform the appropriate test of hypothesis using $\alpha = 0.05$. (3 pts)

   - ▲ $H_0 : \sigma_1^2 = \sigma_2^2 = \sigma_3^2$
   - ▲ H_a : at least one of the σ_i^2 is different
   - ▲ p-value = 0.9359
   - ▲ Rejection Region: Reject H_0 if p-value $< \alpha$.

▲ **Conclusion:** Since the p-value = 0.9359 is (less than, greater than) [circle your choice] $\alpha = 0.05$, (reject, do not reject) [circle your choice] $H_0 \rightarrow$ it (is, is not) [circle your choice] reasonable to assume equality (homogeneity) of variances.

(iii) Fan decides to perform the appropriate test of hypothesis to determine whether the mean weight loss is significantly different for at least one of the formulas. Use $\alpha = 0.05$. (If the equality of variance assumption is not valid, just state that here!) (3 pts)

★ H_0 : $\mu_1 = \mu_2 = \mu_3$ (all $\alpha_i = 0$) (cannot conclude that at least one of the formulas yields a significantly different mean weight loss)

★ H_a : At least one μ_i is different (At least one $\alpha_i \neq 0$) (can conclude that at least one of the formulas yields a significantly different mean weight loss)

★ p-value = 3.58e-07

★ Rejection Region: Reject H_0 if p-value $< \alpha$.

★ **Conclusion:** Since the p-value = 3.58e-07 is (less than, greater than) [circle your choice] $\alpha = 0.05$, $\rightarrow$(reject, do not reject) [circle your choice] $H_0 \rightarrow$ it (is, is not) [circle your choice] reasonable to assume that at least one of formulas yields a significantly different mean weight loss.

(iv) If appropriate, use **Tukey's** test and the **grouping method** to determine which formulas mean weight loss is/are different. (Note: If you couldn't conclude that at least one of the formulas mean weight loss was significantly different, just state that!)

(a) Complete the following table. (You may round to 4 decimal places.) (4 pts)

Comparison	Diff	lwr	upr	p adj
F1 - F2	3.48000	2.329617	4.6303831	0.0000005
F1 - F3	2.79875	1.648367	3.9491331	0.0000127
F2 - F3	-0.68125	-1.831633	0.4691331	0.314034

(b) What conclusion, if any, may be drawn? (Complete the following table!) (3 pts)

Pair Comparison	p-value	p-value $< \alpha = 0.05$ (Yes or No)	Can conclude a significant difference? (Yes or No)
F1 vs F2	0.0000005	yes	yes
F1 vs F3	0.0000127	yes	yes
F2 vs F3	0.314034	no	no

3. Refer to Question 1.

(i) Zigvo would like to determine whether the true average weight loss for *F1* (Formula 1) is significantly less than 4 pounds. Perform the appropriate test of hypothesis.Use $\alpha = 0.05$. (3 pts)

★ $H_0 : \mu_1 = 4$ ($\mu_{Formula1} = 4$)

★ H_a : mu1 <= 4

★ p-value = 0.04538

★ Rejection Region: Reject H_0 if p-value $< \alpha$.

★ **Conclusion:** Since the p-value = 0.04538 is (less than, greater than) [circle your choice] $\alpha = 0.05$, (reject, do not reject) [circle your choice] $H_0 \rightarrow$ it (is, is not) [circle your choice] reasonable to assume that the true mean true average weight loss for *Formula1* is significantly (greater than, less than, different from) [circle your choice] 4 pounds.

(ii) Find and interpret a 97% confidence interval for the true true mean weight loss of *F1* (Formula 1). (3 pts)

Lower Limit	Upper Limit
2.542361	4.235139

Interpretation:

We can be 97% confident that the true mean weight loss for Formula1 is between 2.5423609 and 4.2351391.

4. Refer to Question 1. Consider ***F1*** (Formula 1) and ***F3*** (Formula 3).

(i) Is it reasonable to assume equality of variances? Complete the following test! (4 pts)

▲ $H_0 : \sigma_1^2 = \sigma_3^2$ $(\sigma^2_{Formula1} = \sigma^2_{Formula3})$

▲ $H_a : \sigma_1^2 \neq \sigma_3^2$ $(\sigma^2_{Formula1} \neq \sigma^2_{Formula3})$

▲ p-value = 0.7763

▲ Rejection Region: Reject H_0 if p-value $< \alpha$.

▲ **Conclusion:** Since the p-value = 0.7763 is (less than, greater than) [circle your choice] $\alpha = 0.05$, (reject, do not reject) [circle your choice] H_0 → it (is, is not) [circle your choice] reasonable to assume equality (homogeneity) of variances.

(ii) Assuming the results of part (i), perform the appropriate test of hypothesis to determine whether the mean weight loss for *Formula1* is significantly less than the mean weight loss for *Formula3*. (4 pts)

★ $H_0 : \mu_1 = \mu_3$ $(\mu_{Formula1} = \mu_{Formula3})$

★ H_a : muF1 <= muF2

★ p-value = 1.671e-05

★ Rejection Region: Reject H_0 if p-value $< \alpha$.

★ **Conclusion:** Since the p-value = 1.671e-05 is (less than, greater than) [circle your choice] $\alpha = 0.05$, (reject, do not reject) [circle your choice] H_0 → it (is, is not) [circle your choice] reasonable to assume that the true mean weight loss for *Formula1* is significantly (greater than, less than, different from) [circle your choice] the true mean weight loss for *Formula3*.

5. *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. Be sure to make the columns accessible. (1 pt)

(ii) Perform the appropriate test of hypothesis to determine whether the average score from Kal (J1) is significantly different from the average score from Thomas (J2). Use $\alpha = 0.05$. (Modify your SAS program code to generate the appropriate information.) (4 pts)

- H_0 : muJ1 = muJ2
- H_a : muJ1 /= muJ2
- p-value = 0.2034
- Rejection Region: Reject H_0 if p-value $< \alpha$.
- **Conclusion:** Since the p-value = 0.2034 is (less than, greater than) [circle your choice] $\alpha = 0.05$, (reject, do not reject) [circle your choice] $H_0 \rightarrow$ it (is, is not) [circle your choice] reasonable to assume that the true average score from Kal (J1) is significantly (greater than, less than, different from) [circle your choice] the true average score from Thomas (J2).

Email your completed exam pdf, along with your R script file and the copy/pasted contents of your console.
Have a nice day!!

Luke & Ruihan

Thanks for being great students!

Have a wonderful summer!

See you in Statistics 157!