STATISTICS 147 EXAM II, Part 1: SAS
Summer 2020; 60 pts; Show all work!

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

ID: (last 4 #s only) 0996

General Information:

Complete the following using SAS:

- Save your SAS program in a file named *LastName_FirstName_147_Ex2.sas*, where the last and first name are your own.
- Please include the following titles: (2 pts)

```
Statistics 147 Exam II, Part 1: SAS
Summer 2020
Your name
Question X      (where X = question number)
Part T          (where T = subpart number)
```

- `/*` You should place *meaningful* **comments** in your SAS program file. `*/` (3 pts)
- Be sure to use an `infile` statement to read in your data. DO NOT COPY/PASTE THE DATA DIRECTLY INTO YOUR SAS CODE!
- Use the following options/goptions:

```
/* Set up format of the output */
options nocenter ps = 55 nocenter ls = 78 nodate nonumber formdlim='*';

/* Add line of code to clear all windows except the editor window */
DM log "odsresults; clear; out; clear; log; clear;";
ods graphics off;

/* goptions   formats the plot  */
goptions reset = all colors= (blue,red,green,purple)
       ftitle = swissb ftext=swissb htitle=1.5 htext = 1.0;

/*INCLUDE CODE TO CHANGE THE WORKING DIRECTORY HERE (3 pts)*/
```

- When you have finished your exam, email the following files to lklei001@ucr.edu:
  - (i) Your completed exam PDF.
  - (ii) Your **SAS script/.sas** document.
  - (iii) DO NOT email/submit your SAS output! Luke will run your code for himself.

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 2 lines of headings are included in the data file.

   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 SAS code to read in and print out the data. Use **diet_all** as your temporary SAS dataset name. Use nested do loops! **(NOTE: DO NOT COPY AND PASTE THE DATA INTO YOUR SAS PROGRAM. READ THE DATA IN FROM THE EXTERNAL DATA FILE!)** (3 pts)

   (ii) Using if-then-else structures, name the formulas as follows: (1 pt)

formula	1	2	3
form_name	**Formula1**	**Formula2**	**Formula3**

   (iii) Add the appropriate lines of code to sort the data by form_name. (2 pts)

   (iv) Natascha and Dung decided they should test for normality of the data for each formula. Using the **Shapiro-Wilk** test, complete the following. (4 pts)

form_name	**p-value**	**Reasonable to assume normality? (Answer YES or NO)**
Formula1	0.9433	yes
Formula2	0.3327	yes
Formula3	0.9950	yes

.

(v) Sarah and Rene want to determine whether it is reasonable to assume equality (homogeneity) of variances. Perform the appropriate test of hypothesis using $\alpha = 0.05$. (Modify your SAS program code to generate the appropriate information.) (4 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.

(vi) 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!) (4 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 = <.0001
- ★ Rejection Region: Reject H_0 if p-value $< \alpha$.
- ★ **Conclusion:** Since the p-value = <.0001 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.

(vii) 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 table on the next page. (Modify your SAS program code to generate the appropriate information.) (4 pts)

Tukey's Grouping	Mean	N	Team Formula
A	6.8688	8	Formula2
A			
A	6.1875	8	Formula3
B	3.3888	8	Formula1

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

Pair Comparison	Same Letter (Yes or No)	Can conclude a significant difference? (Yes or No)
Formula1 vs Formula2	no	yes
Formula1 vs Formula3	no	yes
Formula2 vs Formula3	yes	no

2. Refer to Question 1. Create a new temporary SAS data set, named **onlyFormula1**, in which the data is restricted to the **Formula1** labs only. (Be sure to print the data as a check.) (DO NOT COPY THE DATA INTO YOUR SAS PROGRAM!) (2 pts)

(ii) Zigvo would like to determine whether the true average weight loss for *Formula1* is 4 pounds. Perform the appropriate test of hypothesis. Use $\alpha = 0.05$. (Modify your SAS program code to generate the appropriate information.) (6 pts)

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

★ H_a : mu1 /= 4

★ p-value = 0.0908

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

★ Conclusion: **Conclusion:** Since the p-value = 0.0908 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.

(iii) Find and interpret a 97% confidence interval for the true mean weight loss of *Formula1*. (4 pts)

Lower Limit	Upper Limit
2.5423609	4.2351391

Interpretation:

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

3. Refer to Question 1.

(i) Create a new temporary SAS data set, named **bothF1F3** in which the data is restricted to the weight loss of the **Formula1** and **Formula3** labs. (Be sure to print the data as a check.) (DO NOT COPY THE DATA INTO YOUR SAS PROGRAM!) (3 pts)

(ii) Is it reasonable to assume equality of variances? Complete the following test! (Modify your SAS program code to generate the appropriate information.) (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.

(iii) Assuming the results of part (ii), 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*. (5 pts)

- ★ $H_0 : \mu_1 = \mu_3$ $(\mu_{Formula1} = \mu_{Formula3})$
- ★ H_a : mu1 /= mu3
- ★ p-value = <0.0001
- ★ Rejection Region: Reject H_0 if p-value $< \alpha$.
- ★ **Conclusion:** Since the p-value = <0.0001 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*.

4. *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. Kal and Thomas have been selected to be judges in a local *disc dog* competition. The scores, reported by Kal (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 SAS code to read in and print out the data. Use the temporary SAS dataset name of **judging**. (DO NOT use do loops!) **(NOTE: DO NOT COPY AND PASTE THE DATA INTO YOUR SAS PROGRAM. READ THE DATA IN FROM THE EXTERNAL DATA FILE!)** (2 pts)

(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).

Turn in your exam paper with your SAS program attached. Don't forget to email your SAS program file. Prepare for R!

Linda & Lauren