

Statistics 147 Assignment #1

Summer 2019; 45 pts

DUE DATE: Monday, July 1, 2019 by 1:10 pm

1 GENERAL INSTRUCTIONS:

♣ **NO LATE PAPERS WILL BE ACCEPTED FOR CREDIT!**

♣ Your write-up should be neat, well-organized and concise but complete. Your write-up should include a cover page which includes your name, the last 4 digits of your student ID. Your write-up must be typed. Use a word processor such as Word or Google Docs, etc. **No hand-written papers will be accepted!**

♣ For your **SAS** session, name your program file *hwk1su20.sas*.

♠ (3 pts) Include the following titles:

```
title1 'Statistics 147 Assignment #1';
title2 'Summer 2020';
title3 'Your Name';
title4 'SAS Question YY' /* replace the YY with the actual question # */
```

♠ (3 pts) Be sure to turn in a complete copy of just your SAS code at the end!

♠ (3 pts) Be sure to include comments/documentation in your SAS program.

♣ For your **R** session,

♠ (2 pt) include the following titles (as comments)

```
> # Statistics 147 Assignment #1
> # Summer 2020
> # Your Name
> # R Question ZZ (where ZZ = question number)
```

♠ (3 pts) Be sure to include comments in your R code.

♠ (3 pts) Be sure to include a copy of your R code at the end!

2 The Data

The following data represents the wheat yields in bushels per acre (1 bushel = 60 pounds) for two varieties of wheat: *hard red winter wheat* (used to make breads) and *soft red winter wheat* (used to make cakes, cookies, snack foods, crackers and pastries):

| HardRed | SoftRed |
|---------|---------|
| 29.8 | 49.0 |
| 29.5 | 53.4 |
| 38.2 | 54.2 |
| 39.9 | 59.0 |
| 36.0 | 56.6 |
| 39.0 | 57.9 |
| 36.7 | 55.8 |
| 31.1 | 43.2 |
| 26.9 | 55.6 |
| 36.6 | 54.2 |
| 37.8 | 59.9 |
| 35.4 | 63.2 |
| 37.2 | 50.0 |
| 39.9 | 60.9 |
| 38.1 | 56.1 |
| 28.5 | 54.3 |
| 36.9 | 62.1 |

NOTE: This data has been saved in two formats:

- A **.csv** file named **wheat1.csv** (The headings are included in this data file! The actual data starts on Line 2 of the data file.)
- A text file named **wheat2.dat** (The headings are included in this data file! The actual data starts on Line 2 of the data file.)

These files have been uploaded to Blackboard under *Data Files*.

3 The Questions

3.1 Using R

- Using **R**, complete the following.
 - (2 pts) Read in and print out the **.csv** file, **wheat1.csv**.
 - (2 pts) Find the sum of the yields/acre for the seventeen acres of SoftRed wheat. Call this variable **sum_SoftRed**. (Be sure to print out the value of the sum!)
 - (2 pts) Find the sum of the yields/acre for the seventeen acres of HardRed wheat. Call this variable **sum_HardRed**. (Be sure to print out the value of the sum!)
 - (2 pts) Find the sum of the yields/acre for the thirty-four acres of wheat. Call this variable **total_all**. (Be sure to print out the value of the sum!)
 - (4 pts) Generate the following descriptive statistics for each type of wheat. Complete the following table:

| State | # of Obs | Mean | Median | Standard Deviation | Variance | Max Value | Min Value |
|---------|----------|------|--------|--------------------|----------|-----------|-----------|
| HardRed | | | | | | | |
| SoftRed | | | | | | | |

- (4 pts) For **SoftRed**, create a histogram for the yield using breaks from 40 to 65 in increments of 5. Make sure the frequencies are listed above each bar in your histogram.

(i) **(1 pt)** Identify interval/class that has the largest number of observations. (Be sure to state the frequency for the class.)

3.2 Using SAS

1. Write a SAS program to complete the following:

(i) **(4 pts)** Read in and print out the data.

(ii) **(2 pts)** Sort the data by the variable **HardRed**. (Be sure to print the sorted data!)

(a) **(0.5 pts)** What is the largest yield/acre for **HardRed** wheat?

(b) **(0.5 pts)** What is the smallest yield/acre for **HardRed** wheat?

If you don't already have the `goptions` statement in your program, add the following right after the options statement:

```
/* Set up some options for gchart and gplot */
goptions reset = global colors=(red,blue,green,yellow,pink,purple);
```

(iii) **(3 pts)** Modify your program to generate a 3-D vertical bar chart for **SoftRed** using midpoints beginning at 35 and ending at 70, in increments of 10. Be sure to include some coloration, including lightblue as color of the graph/chart background frame, verydarkblue as the color of the outline, hexagon as the shape of the bars, verydarkgreen as the color of text within the chart/graph and pink as the colors for the bars. (You do not have to submit your graph in color, but your code must reflect the color scheme.)

(iv) **(3 pts)** Modify your program to generate the descriptive statistics for the yield for each variety. Identify the mean, median, and standard deviation for each of the varieties.

(v) **(3 pts)** Modify your program to generate a high resolution plot of the yield of **HardRed** versus that of **SoftRed**. Use **SoftRed** as the **vertical** axis! (Use **proc gplot** so you have flexibility in the presentation of your plot. You can use the coloration options with **proc gplot** just like you did with **proc gchart**.)

NOTE: You may change the symbol that represents the data points on your graph by placing the following code **right before** your **proc gplot** statement:

```
symbol1 color=red
        value=dot
        height=1;
```

Feel free to change the color, height, etc. Some values you can use with `symbol1` are circle, dot, plus, block, square, diamond, triangle, hash, and star.

Have fun!