

Statistics 147 Assignment #1

Wesley Chang

0996

Write-up

3.1 Using R

1. Using R, complete the following.

(i) **(2 pts)** Read in and print out the .csv file, **wheat1.csv**.

```
# read and print wheat1.csv  
# csv has one line of header and data values are separated by a ","  
wheat1 <- read.table("wheat1.csv", header=TRUE, sep=",")  
wheat1
```

```
> wheat1 <- read.table("wheat1.csv", header=TRUE, sep=",")  
> wheat1  
  HardRed SoftRed  
1    29.8    49.0  
2    29.5    53.4  
3    38.2    54.2  
4    39.9    59.0  
5    36.0    56.6  
6    39.0    57.9  
7    36.7    55.8  
8    31.1    43.2  
9    26.9    55.6  
10   36.6    54.2  
11   37.8    59.9  
12   35.4    63.2  
13   37.2    50.0  
14   39.9    60.9  
15   38.1    56.1  
16   28.5    54.3  
17   36.9    62.1
```

(ii) **(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!)

```
# make columns individually accessible variables and print column names  
attach(wheat1)  
names(wheat1)  
  
# ii  
  
# sum of yields/acre for seventeen acres of SoftRed wheat  
sum_SoftRed <- sum(SoftRed)  
sum_SoftRed
```

```
> # ii  
>  
> # sum of yields/acre for seventeen acres of SoftRed wheat  
> sum_SoftRed <- sum(SoftRed)  
> sum_SoftRed  
[1] 945.4
```

(iii) **(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!)

```
# iii  
  
# sum of yields/acre for seventeen acres of HardRed wheat  
sum_HardRed <- sum(HardRed)  
sum_HardRed
```

```
> sum_HardRed <- sum(HardRed)  
> sum_HardRed  
[1] 597.5
```

(iv) **(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!)

```
# iv
```

```
# sum of all thirty-four acres of wheat  
total_all <- sum_HardRed + sum_SoftRed  
total_all
```

```
> total_all <- sum_HardRed + sum_SoftRed  
> total_all  
[1] 1542.9
```

(v) **(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	17	35.147	36.700	4.234	17.923	39.900	26.900
SoftRed	17	55.612	55.800	5.024	25.241	63.200	43.200

```
# v

# generate descriptive statistics obs, mean, median, std dev, var, max, min

# use pastecs library
library("pastecs")

# generate a table of descriptive statistics using the stat.desc function
stat.desc(HardRed)
stat.desc(SoftRed)
```

```
> stat.desc(HardRed)
  nbr.val  nbr.null  nbr.na    min    max    range    sum
17.0000000  0.0000000  0.0000000 26.9000000 39.9000000 13.0000000 597.5000000
  median    mean  SE.mean CI.mean.0.95    var  std.dev  coef.var
36.7000000 35.1470588  1.0267781   2.1766724 17.9226471  4.2335147   0.1204515
> stat.desc(SoftRed)
  nbr.val  nbr.null  nbr.na    min    max    range    sum
17.0000000  0.0000000  0.0000000 43.2000000 63.2000000 20.0000000 945.4000000
  median    mean  SE.mean CI.mean.0.95    var  std.dev  coef.var
55.8000000 55.61176471  1.21851170   2.58312941 25.24110294  5.02405244   0.09034154
```

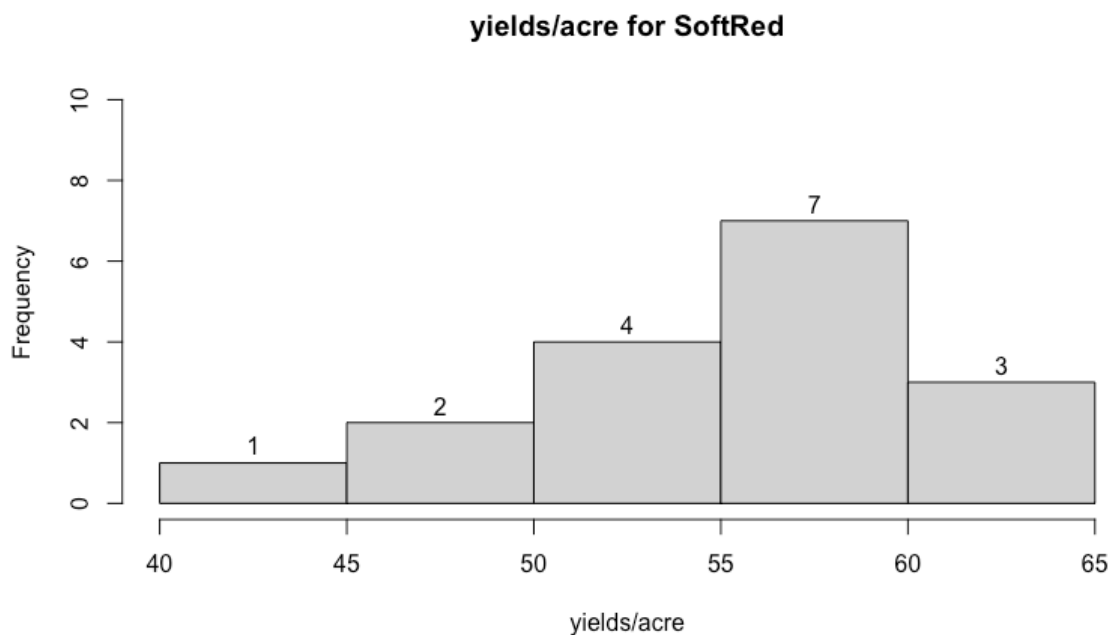
(vi) **(4 pts)** For **SoftRed**, create a histogram for the yield using breaks from 40 to 65 in increments of 5. Make sure the frequencies are listed above each bar in your histogram.

```
# vi

# create a histogram for SoftRed, using breaks from 40 to 65 in increments of 5
# make sure frequencies are listed above each bar in the histogram

# define the breaks for the interval, in increments of 5
brks1 <- c(40,45,50,55,60,65)
brks1

# create a histogram for SoftRed, with title and breaks as defined above
hist(SoftRed,
     main="yields/acre for SoftRed",
     breaks = brks1,
     ylim = c(0,10),
     xlab = "yields/acre",
     labels = TRUE
)
```



- (i) **(1 pt)** Identify interval/class that has the largest number of observations. (Be sure to state the frequency for the class.)
- The class with the largest number of observations is 55-60, with 7 occurrences**

3.2 Using SAS

1. Write a SAS program to complete the following:

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

Statistics 147 Assignment #1
Summer 2020
Wesley Chang
SAS Question 1

Obs	HardRed	SoftRed
1	29.8	49.0
2	29.5	53.4
3	38.2	54.2
4	39.9	59.0
5	36.0	56.6
6	39.0	57.9
7	36.7	55.8
8	31.1	43.2
9	26.9	55.6
10	36.6	54.2
11	37.8	59.9
12	35.4	63.2
13	37.2	50.0
14	39.9	60.9
15	38.1	56.1
16	28.5	54.3
17	36.9	62.1

```
title5 'Subpart (i)';  
/* i */  
/* read data into wheat from file, listing variable names HardRed and SoftRed */  
data wheat;  
    infile 'C:\Users\wchan061_ucr\Downloads\wheat2.dat' firstobs=2;  
    input HardRed SoftRed;  
run;  
  
/* print wheat */  
proc print data = wheat;  
run;
```

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

Statistics 147 Assignment #1

Summer 2020

Wesley Chang

Question 1

Subpart (ii)

Obs	HardRed
1	26.9
2	28.5
3	29.5
4	29.8
5	31.1
6	35.4
7	36.0
8	36.6
9	36.7
10	36.9
11	37.2
12	37.8
13	38.1
14	38.2
15	39.0
16	39.9
17	39.9

```
/* ii */  
/* use proc to sort data by HardRed, then print */  
title5 'Subpart (ii)';  
proc sort;  
    by HardRed;  
proc print;  
    var HardRed;  
run;
```

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

The largest yield is 39.9

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

The smallest yield is 26.9

(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.)

Statistics 147 Assignment #1

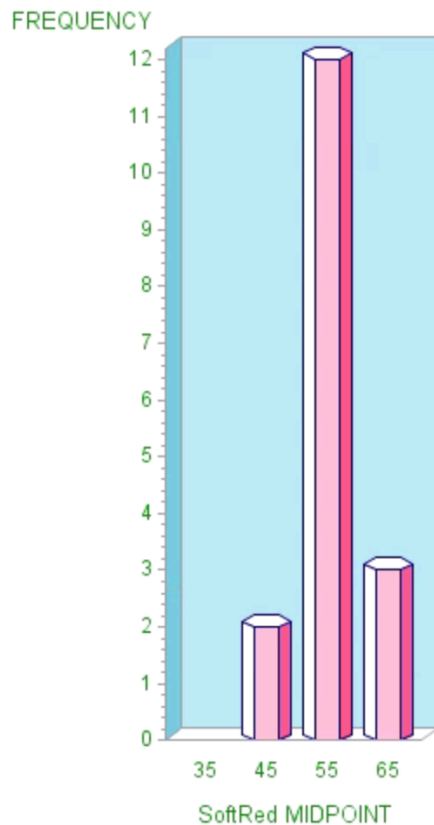
Summer 2020

Wesley Chang

Question 1

Subpart (iii)

Bar Chart for SoftRed



```
title5 'Subpart (iii)';
proc gchart data = wheat;
  title6 'Bar Chart for SoftRed';
  vbar3d SoftRed / midpoints = 35 to 70 by 10
    cfr = lightblue
    coutline = verydarkblue
    shape = hexagon
    ctext = verydarkgreen;
  pattern color = pink;
run;
```

(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.

Statistics 147 Assignment #1
Summer 2020
Wesley Chang
Question 1
Subpart (iv)
SoftRed Summary Statistics

The UNIVARIATE Procedure
Variable: SoftRed

Moments			
N	17	Sum Weights	17
Mean	55.6117647	Sum Observations	945.4
Std Deviation	5.02405244	Variance	25.2411029
Skewness	-0.7863135	Kurtosis	1.06541148
Uncorrected SS	52979.22	Corrected SS	403.857647
Coeff Variation	9.03415396	Std Error Mean	1.2185117

Basic Statistical Measures			
Location		Variability	
Mean	55.61176	Std Deviation	5.02405
Median	55.80000	Variance	25.24110
Mode	54.20000	Range	20.00000
		Interquartile Range	4.80000

Statistics 147 Assignment #1
Summer 2020
Wesley Chang
Question 1
Subpart (iv)
HardRed Summary Statistics

The UNIVARIATE Procedure
Variable: HardRed

Moments			
N	17	Sum Weights	17
Mean	35.1470588	Sum Observations	597.5
Std Deviation	4.23351474	Variance	17.9226471
Skewness	-0.8285625	Kurtosis	-0.7731609
Uncorrected SS	21287.13	Corrected SS	286.762353
Coeff Variation	12.0451465	Std Error Mean	1.02677814

Basic Statistical Measures			
Location		Variability	
Mean	35.14706	Std Deviation	4.23351
Median	36.70000	Variance	17.92265
Mode	39.90000	Range	13.00000

```

title5 'Subpart (iv)';
/* generate descriptive statistics for the yield in each variety */
/* proc univariate */

/* for SoftRed */
proc univariate data = wheat;
    title6 'SoftRed Summary Statistics';
    var SoftRed;
run;

/* for HardRed */
proc univariate data = wheat;
    title6 'HardRed Summary Statistics';
    var HardRed;
run;

```

SoftRed: Mean = 55.612, Median = 55.800, Standard Deviation = 5.024

HardRed: Mean = 35.147, Median = 36.700, Standard Deviation = 4.234

(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.

Statistics 147 Assignment #1

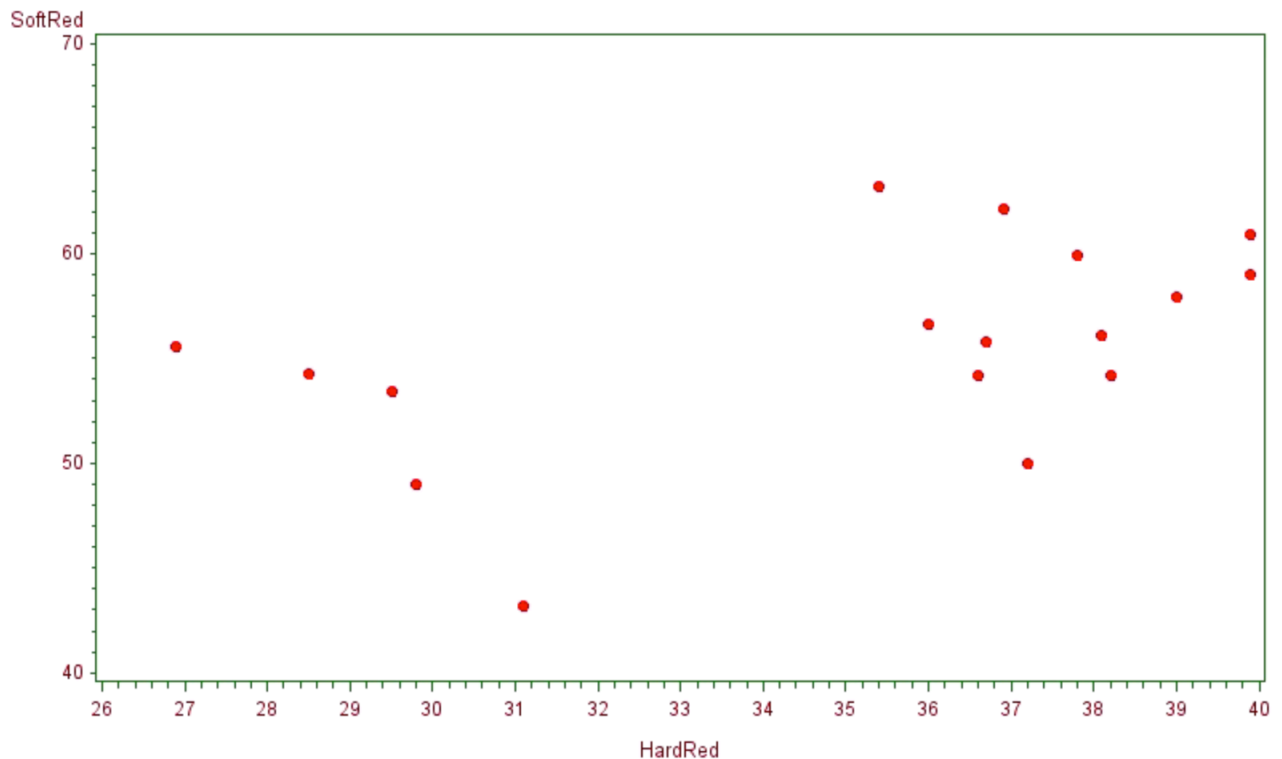
Summer 2020

Wesley Chang

Question 1

Subpart (v)

Plot of HardRed vs SoftRed



```

title5 'Subpart (v)';
/* plot HardRed vs SoftRed, with SoftRed on vertical axis
   use proc gplot */

/* adjust format of the plot symbols */
symbol1 color = red
        value = dot
        height = 1;

```

```

/* generate high resolution plot using proc gplot */
proc gplot data = wheat;
    title6 'Plot of HardRed vs SoftRed';
    plot SoftRed*HardRed /
        caxis = darkgreen|
        ctext = darkred;
run;

```

Code for the R part:

```
# Statistics 147 Assignment #1
# Summer 2020
# Wesley Chang

# R Question 3.1

# change working directory to assignment folder for easy access
setwd("/Users/wes_chang/Library/Mobile Documents/com~apple~CloudDocs/Summer
2020 (UCR)/STAT 147 (Session A)/Assignments/1/")

# i

# read and print wheat1.csv
# csv has one line of header and data values are separated by a ","
wheat1 <- read.table("wheat1.csv", header=TRUE, sep=",")
wheat1

# make columns individually accessible variables and print column names
attach(wheat1)
names(wheat1)

# ii

# sum of yields/acre for seventeen acres of SoftRed wheat
sum_SoftRed <- sum(SoftRed)
sum_SoftRed

# iii

# sum of yields/acre for seventeen acres of HardRed wheat
sum_HardRed <- sum(HardRed)
sum_HardRed

# iv

# sum of all thirty-four acres of wheat
total_all <- sum_HardRed + sum_SoftRed
total_all

# v

# generate descriptive statistics obs, mean, median, std dev, var, max, min

# use pastecs library
library("pastecs")

# generate a table of descriptive statistics using the stat.desc function
stat.desc(HardRed)
```

```

stat.desc(SoftRed)

# vi

# create a histogram for SoftRed, using breaks from 40 to 65
# in increments of 5
# make sure frequencies are listed above each bar in the histogram

# define the breaks for the interval, in increments of 5
brks1 <- c(40,45,50,55,60,65)
brks1

# create a histogram for SoftRed, with title and breaks as defined above
hist(SoftRed,
     main="yields/acre for SoftRed",
     breaks = brks1,
     ylim = c(0,10),
     xlab = "yields/acre",
     labels = TRUE
  )

```

Code for the SAS part:

```
DM log "odsresults; clear; out; clear; log; clear;";
ods graphics off;

title1 'Statistics 147 Assignment #1';
    title2 'Summer 2020';
    title3 'Wesley Chang';

title4 'Question 1';

title5 'Subpart (i)';
/* i */
/* read data into wheat from file, listing variable names HardRed and SoftRed */
data wheat;
    infile 'C:\Users\wchan061_ucr\Downloads\wheat2.dat' firstobs=2;
    input HardRed SoftRed;
run;

/* print wheat */
proc print data = wheat;
run;

/* ii */
/* use proc to sort data by HardRed, then print */
title5 'Subpart (ii)';
proc sort;
    by HardRed;
proc print;
    var HardRed;
run;

/* iii */
/* generate bar chart for SoftRed using midpoints at 35 to 70, in increments of
10
    add coloration (lightblue as color of graph/chart background frame,
verydarkblue
    as the color of the outline, hexagon as the shape of the bars, verydarkgreen
as
    the color of the text within the chart/graph and pink as the color for the
bars */

title5 'Subpart (iii)';
proc gchart data = wheat;
    title6 'Bar Chart for SoftRed';
    vbar3d SoftRed / midpoints = 35 to 70 by 10 /* vertical 3d bar chart */
        cfr = lightblue /* frame */
        coutline = verydarkblue /* outline */
        shape = hexagon /* shape */
        ctext = verydarkgreen; /* text */
    pattern color = pink; /* color for bars */
run;
```

```

title5 'Subpart (iv)';
/* generate descriptive statistics for the yield in each variety */
/* proc univariate */

/* for SoftRed */
proc univariate data = wheat;
    title6 'SoftRed Summary Statistics';
    var SoftRed;
run;

/* for HardRed */
proc univariate data = wheat;
    title6 'HardRed Summary Statistics';
    var HardRed;
run;

title5 'Subpart (v)';
/* plot HardRed vs SoftRed, with SoftRed on vertical axis
   use proc gplot */

/* adjust format of the plot symbols */
symbol1 color = red
        value = dot
        height = 1;

/* generate high resolution plot using proc gplot */
proc gplot data = wheat;
    title6 'Plot of HardRed vs SoftRed';
    plot SoftRed*HardRed /
        caxis = darkgreen
        ctext = darkred;
run;

```