**Statistics 147 Assignment #1**

**Wesley Chang**

**0996**

# Write-up

# 3.1 Using R

1. Using **R**, complete the following.

1. **(2 pts)** Read in and print out the **.csv** file, **wheat1.csv**.

A picture containing drawing

Description automatically generated

A screenshot of a cell phone

Description automatically generated

1. **(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!)

A screenshot of a cell phone

Description automatically generated

A picture containing drawing

Description automatically generated

1. **(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!)

A picture containing drawing

Description automatically generated

A picture containing drawing

Description automatically generated

1. **(2 pts)** Find the sum of the yields/acre for the thirty-four acres of wheat. Call this variable **totalall**. (Be sure to print out the value of the sum!)



A close up of a logo

Description automatically generated

1. **(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 |

A screenshot of a cell phone

Description automatically generated

A close up of a sign

Description automatically generated

1. **(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.

A screenshot of a cell phone

Description automatically generated

![A picture containing screenshot

Description automatically generated]()

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

1. **(4 pts)** Read in and print out the data.

A picture containing cellphone

Description automatically generated A screenshot of a cell phone

Description automatically generated

1. **(2 pts)** Sort the data by the variable HardRed. (Be sure to print the sorted data!)

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

(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

1. **(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.)

A screenshot of a cell phone

Description automatically generated  
A screenshot of a cell phone

Description automatically generated

1. (**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.

A screenshot of a cell phone

Description automatically generated A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

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

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

1. **(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 changec 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.

A screenshot of a cell phone

Description automatically generated

A screenshot of a social media post

Description automatically generated

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 invidiually 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;