**Assignment #4**

**Stats 147 Fall 2017 Sec. 2**

Sarah Ruckman

SID: 7194

**Using SAS:**

1. A statistician, with Icelandic heritage1, was interested in comparing the average eruption heights of three world famous geysers: Old Faithful (in US), Grand Geysir2 (in Iceland), and Strokkur (in Iceland). The statistician obtained independent samples of 10 observations from each of the geysers and recorded tin a data file named geysers f17.dat.
2. Read in and print out the data using nested Do loops. (Be sure to give your columns the appropriate full names (i.e., Old Faithful, Grand Geysir and Strokkur)!) (4 pts)

**SAS Code:**

options ls = **70** ps = **55** nocenter formdlim = '\*';

/\* ls = linesize, ps = pagesize, nocenter = justifies output, formdlim = overrides the internal page breaks

and replaces them with the designated symbol \*/

/\* Create titles \*/

title1 'Statstics 147 Assignment #4';

title2 'Fall 2017';

title3 'Sarah Ruckman';

title4 'Section 002';

title5 'SAS Question 1';

/\*Read infile using do loops do rows and then columns and there are 2 lines of headers\*/

/\*Create temporary SAS dataset called geysers\*/

**data** geysers;

infile "C:\Users\sarah\Downloads\geysers\_f17.dat" firstobs = **3**;

do rows = **3** to **12**;

do col = **1** to **3**;

/\*Create if then else statements to label the columns and end with an else statement\*/

if col = **1** then name = 'OF';

else if col = **2** then name = 'GG';

else name = 'S ';

/\*input the values and add @@ to denote that the values are on the same line\*/

input values @@;

/\*Output the results\*/

output;

/\*Close both loops\*/

end;

end;

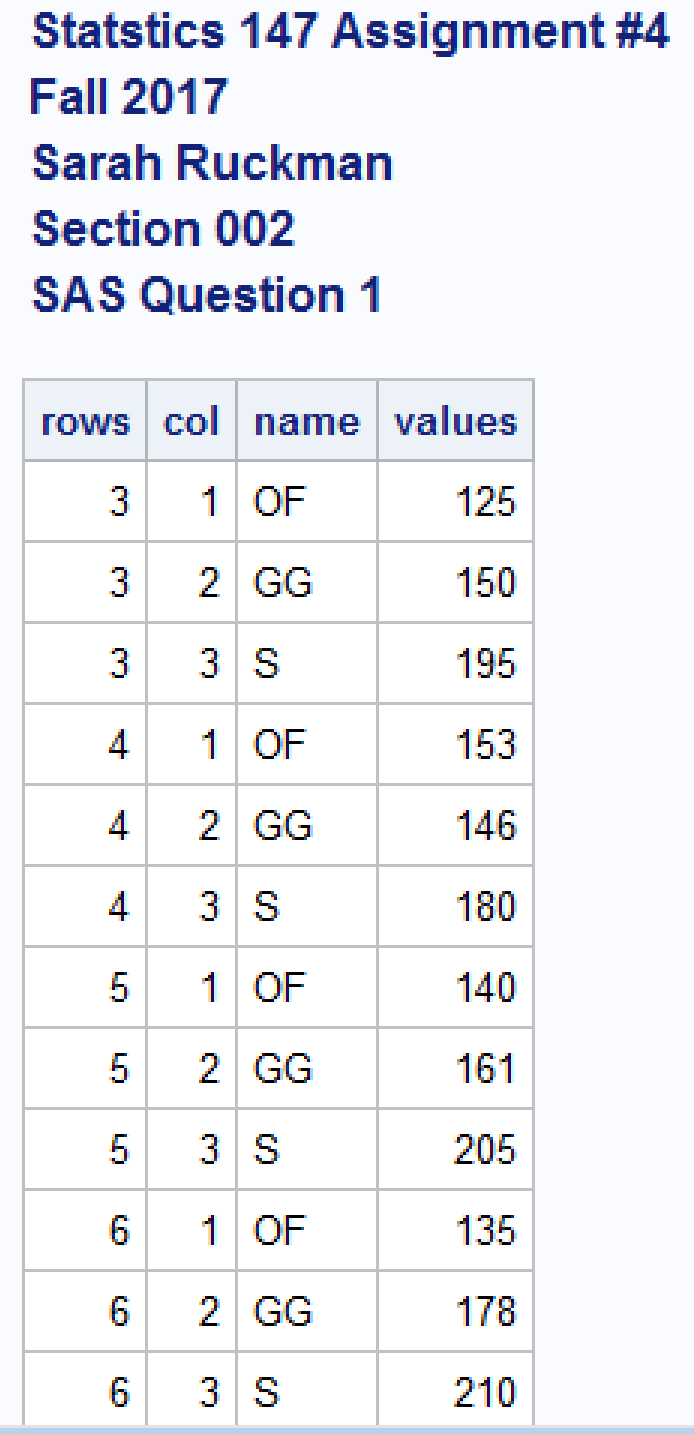
/\*Print as check without observations\*/

**proc** **print** noobs;

**run**;

**quit**;

**Output:**

1. Sort the geysers by the names. (2 pts)

**SAS Code:**

options ls = **70** ps = **55** nocenter formdlim = '\*';

/\* ls = linesize, ps = pagesize, nocenter = justifies output, formdlim = overrides the internal page breaks

and replaces them with the designated symbol \*/

/\* Create titles \*/

title1 'Statstics 147 Assignment #4';

title2 'Fall 2017';

title3 'Sarah Ruckman';

title4 'Section 002';

title5 'SAS Question 1';

/\*Read infile using do loops do rows and then columns and there are 2 lines of headers\*/

/\*Create temporary SAS dataset called geysers\*/

**data** geysers;

infile "C:\Users\sarah\Downloads\geysers\_f17.dat" firstobs = **3**;

do rows = **3** to **12**;

do col = **1** to **3**;

/\*Create if then else statements to label the columns and end with an else statement\*/

if col = **1** then name = 'OF';

else if col = **2** then name = 'GG';

else name = 'S ';

/\*input the values and add @@ to denote that the values are on the same line\*/

input values @@;

/\*Output the results\*/

output;

/\*Close both loops\*/

end;

end;

/\*Sort the data by column using proc sort\*/

**proc** **sort**;

by name;

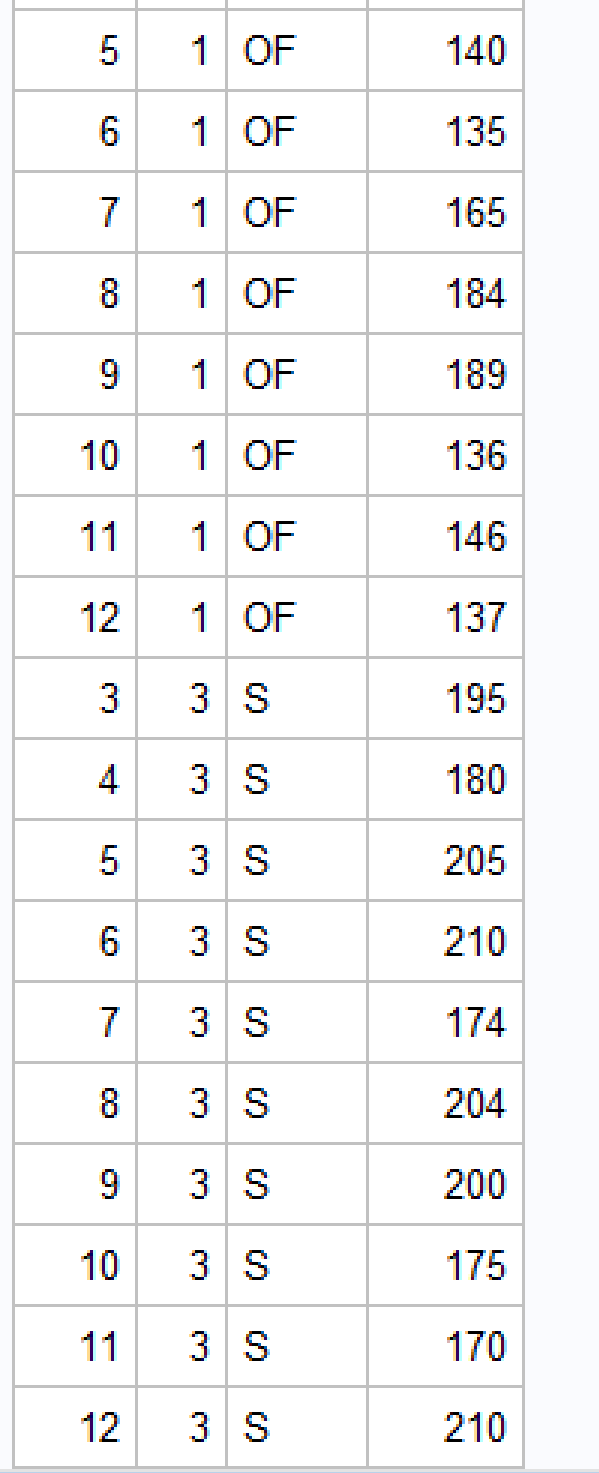
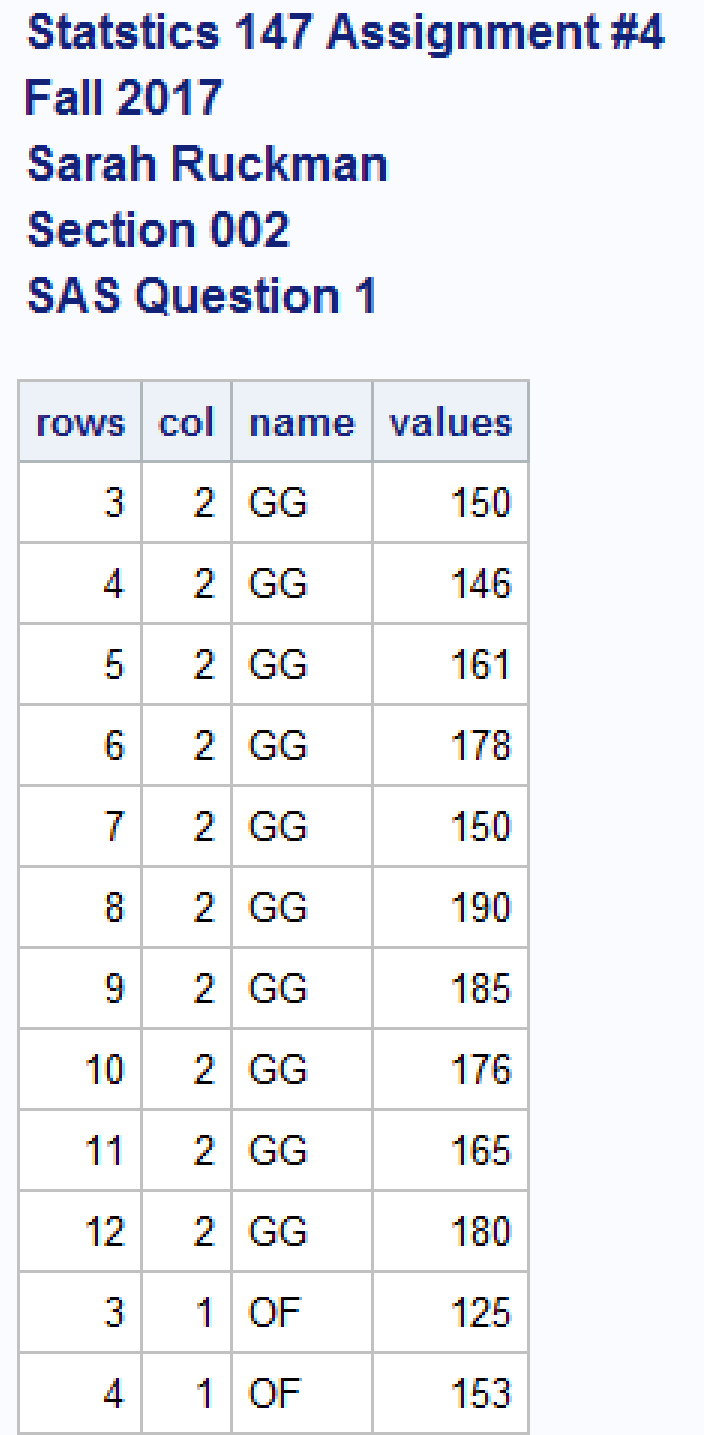
/\*Print as check without observations\*/

**proc** **print** noobs;

**run**;

**quit**;

**Output:**



1. Generate the mean, standard deviation and variance for each of the geysers.

**SAS Code:**

options ls = **70** ps = **55** nocenter formdlim = '\*';

/\* ls = linesize, ps = pagesize, nocenter = justifies output, formdlim = overrides the internal page breaks

and replaces them with the designated symbol \*/

/\* Create titles \*/

title1 'Statstics 147 Assignment #4';

title2 'Fall 2017';

title3 'Sarah Ruckman';

title4 'Section 002';

title5 'SAS Question 1';

/\*Read infile using do loops do rows and then columns and there are 2 lines of headers\*/

/\*Create temporary SAS dataset called geysers\*/

**data** geysers;

infile "C:\Users\sarah\Downloads\geysers\_f17.dat" firstobs = **3**;

do rows = **3** to **12**;

do col = **1** to **3**;

/\*Create if then else statements to label the columns and end with an else statement\*/

if col = **1** then name = 'OF';

else if col = **2** then name = 'GG';

else name = 'S ';

/\*input the values and add @@ to denote that the values are on the same line\*/

input values @@;

/\*Output the results\*/

output;

/\*Close both loops\*/

end;

end;

/\*Sort the data by column using proc sort\*/

**proc** **sort**;

by name;

/\*Print as check without observations\*/

**proc** **print** noobs;

/\*Use proc means to find the mean stddev and var for each geyser\*/

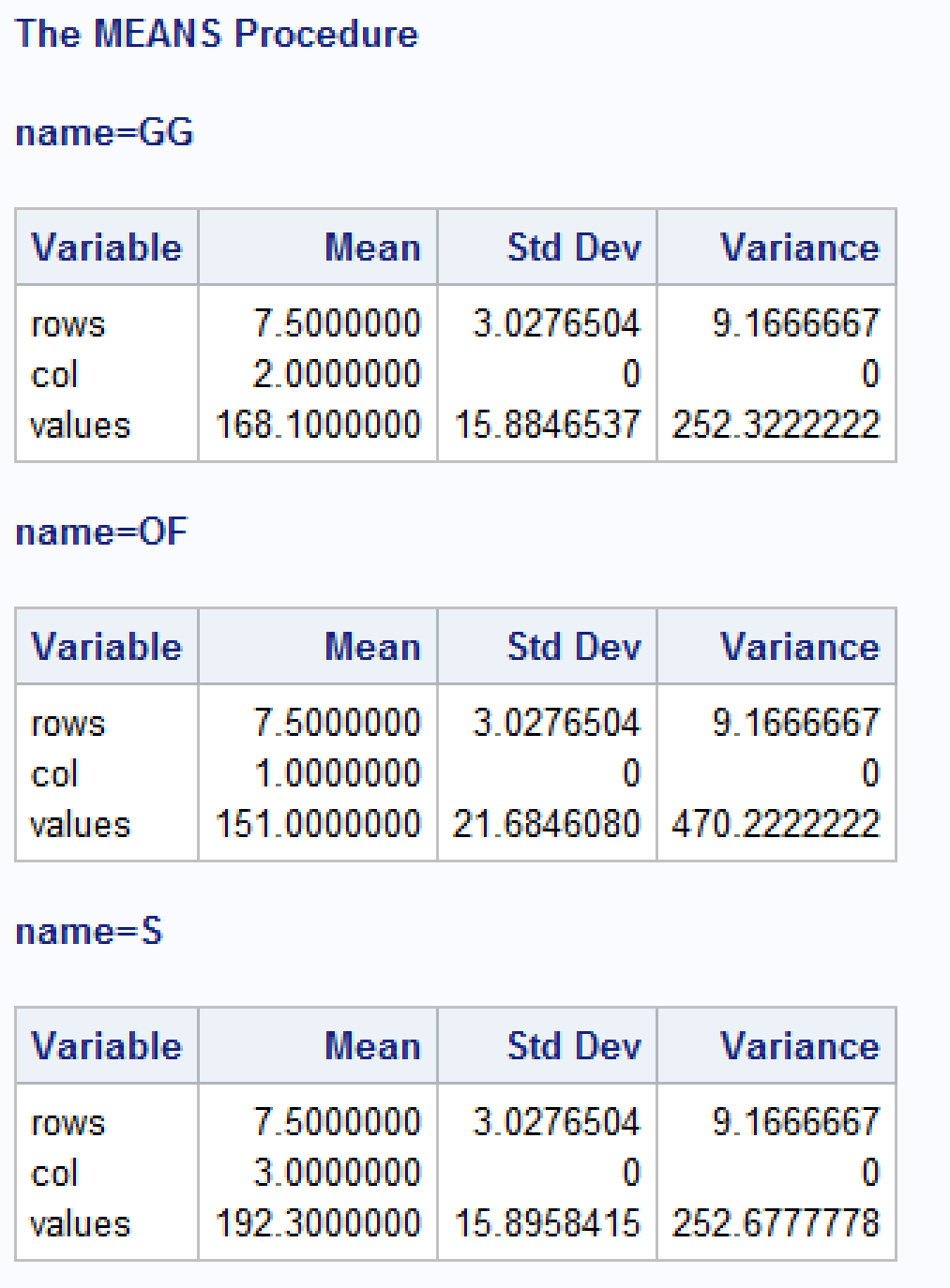
**proc** **means** mean stddev var;

by name;

**run**;

**quit**;

**Output:**



|  |  |  |  |
| --- | --- | --- | --- |
| **Name** | **Mean** | **Stand Deviation** | **Variance** |
| **GG** | 168.1000 | 15.8846537 | 252.3222 |
| **OF** | 151.0000 | 21.6846080 | 470.2222 |
| **S** | 192.3000 | 15.8958415 | 252.6777 |

1. Create a new SAS dataset, called onlyGG, and bring in the data using the SET command. Use the appropriate IF structure to restrict the data to the Grand Geysir geyser. Be sure to print the data! (2 pts)

**SAS Code:**

options ls = 70 ps = 55 nocenter formdlim = '\*';

/\* ls = linesize, ps = pagesize, nocenter = justifies output, formdlim = overrides the internal page breaks

and replaces them with the designated symbol \*/

/\* Create titles \*/

title1 'Statstics 147 Assignment #4';

title2 'Fall 2017';

title3 'Sarah Ruckman';

title4 'Section 002';

title5 'SAS Question 1';

/\*Read infile using do loops do rows and then columns and there are 2 lines of headers\*/

/\*Create temporary SAS dataset called geysers\*/

data geysers;

infile "C:\Users\sarah\Downloads\geysers\_f17.dat" firstobs = 3;

do rows = 3 to 12;

do col = 1 to 3;

/\*Create if then else statements to label the columns and end with an else statement\*/

if col = 1 then name = 'OF';

else if col = 2 then name = 'GG';

else name = 'S ';

/\*input the values and add @@ to denote that the values are on the same line\*/

input values @@;

/\*Output the results\*/

output;

/\*Close both loops\*/

end;

end;

/\*Sort the data by column using proc sort\*/

proc sort;

by name;

/\*Print as check without observations\*/

proc print noobs;

/\*Use proc means to find the mean stddev and var for each geyser\*/

proc means mean stddev var;

by name;

/\*Use the set command and an if statement to create a data set with only GG\*/

/\*Create a temporary SAS dataset\*/

data onlyGG;

set geysers;

if col =2;

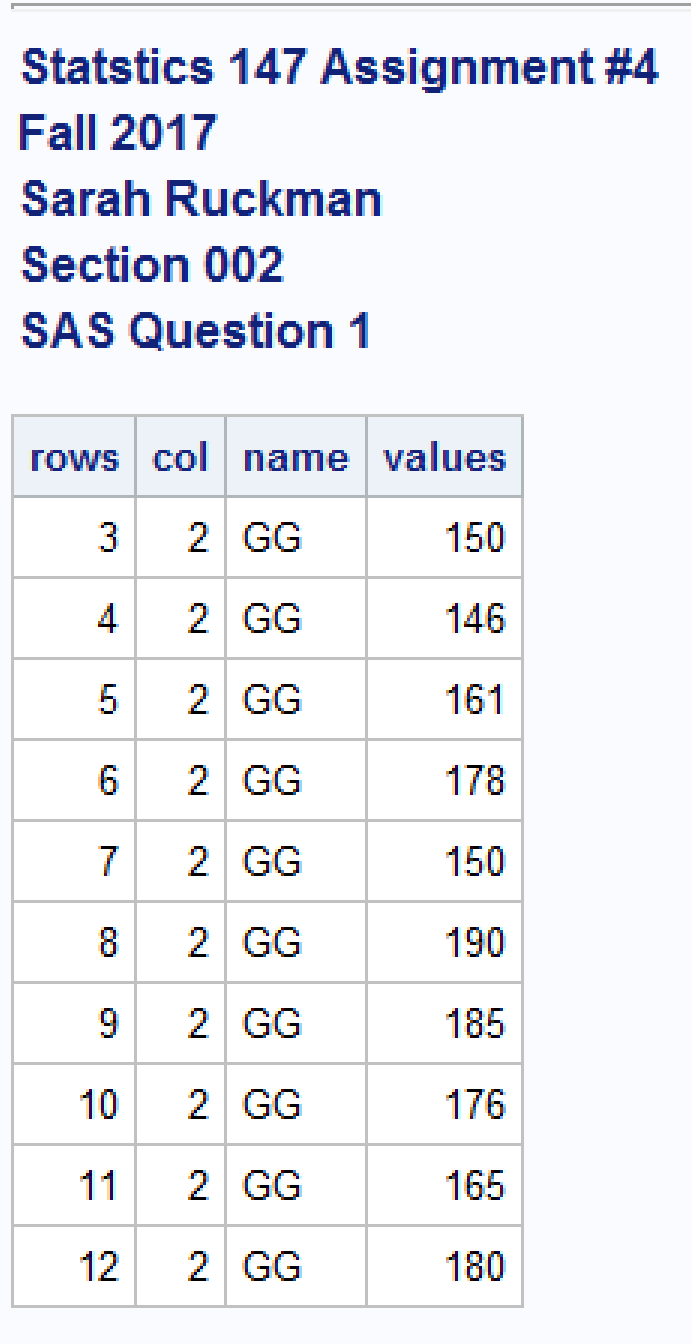
/\*Print the results\*/

proc print noobs;

run;

quit;

**Output:**



1. Create a new SAS dataset, called bothGGS and bring in the data using the SET command. Use the appropriate IF structure to restrict the data to the Grand Geysir and Strokkur geysers. Be sure to print the data! (2 pts)

**SAS Code:**

options ls = **70** ps = **55** nocenter formdlim = '\*';

/\* ls = linesize, ps = pagesize, nocenter = justifies output, formdlim = overrides the internal page breaks

and replaces them with the designated symbol \*/

/\* Create titles \*/

title1 'Statstics 147 Assignment #4';

title2 'Fall 2017';

title3 'Sarah Ruckman';

title4 'Section 002';

title5 'SAS Question 1';

/\*Read infile using do loops do rows and then columns and there are 2 lines of headers\*/

/\*Create temporary SAS dataset called geysers\*/

**data** geysers;

infile "C:\Users\sarah\Downloads\geysers\_f17.dat" firstobs = **3**;

do rows = **3** to **12**;

do col = **1** to **3**;

/\*Create if then else statements to label the columns and end with an else statement\*/

if col = **1** then name = 'OF';

else if col = **2** then name = 'GG';

else name = 'S ';

/\*input the values and add @@ to denote that the values are on the same line\*/

input values @@;

/\*Output the results\*/

output;

/\*Close both loops\*/

end;

end;

/\*Sort the data by column using proc sort\*/

**proc** **sort**;

by name;

/\*Print as check without observations\*/

**proc** **print** noobs;

/\*Use proc means to find the mean stddev and var for each geyser\*/

**proc** **means** mean stddev var;

by name;

/\*Use the set command and an if statement to create a data set with only GG\*/

/\*Create a temporary SAS dataset\*/

**data** onlyGG;

set geysers;

if col =**2**;

/\*Print the results\*/

**proc** **print** noobs;

/\*Use the set command and an if statement to create a data set with GG and S using a not=1\*/

/\*Create temporary SAS dataset\*/

**data** bothGGS;

set geysers;

if col not= **1**;

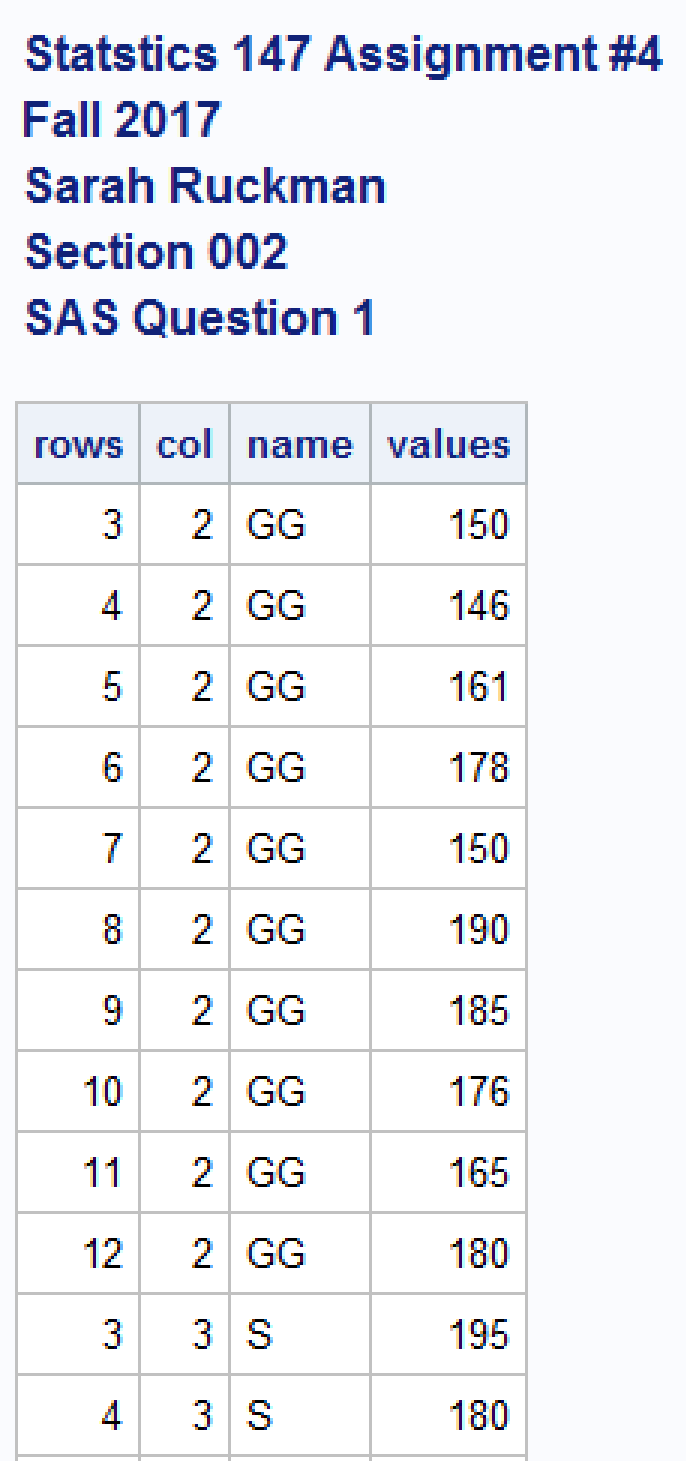
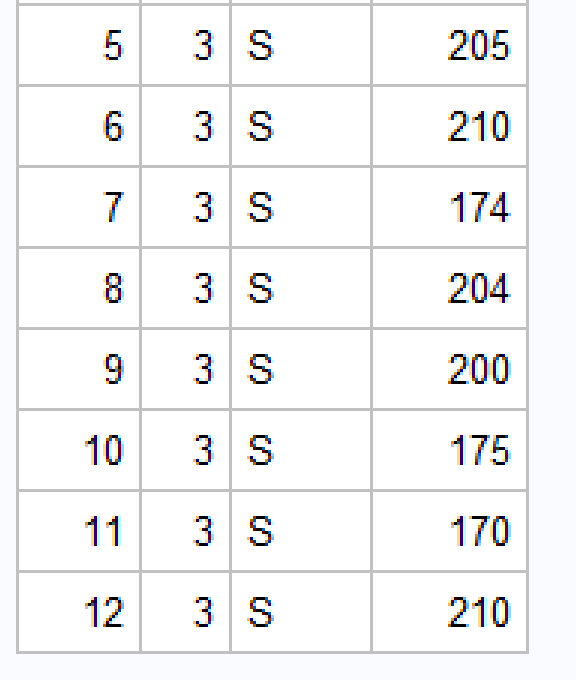
/\*Print the results\*/

**proc** **print** noobs;

**run**;

**quit**;

**Output:**

1. The Great Outdoor Games, sponsored by ESPN, feature the Big Air Dogs event. In this event, also known as dock diving, the dogs race down a dock and fly off the end into the water, with long distance being the goal. During the competition, each dog is allowed 2 or 3 jumps. Linda and Lauren, believing that the dogs dive distance is significantly different on the second dive than on the first dive, attended a recent competition and recorded the dive distance for the first two dives for the top 10 dogs. (NOTE: The data represents the distance of each dive, measured in feet, for each dog.) and record them in a data file named dogdive f17.dat.
2. Read in the data using an infile statement. (Do not use DO loops! Be sure to print your data as a check!) (3 pts)

**SAS Code:**

options ls = 70 ps = 55 nocenter formdlim = '\*';

/\* ls = linesize, ps = pagesize, nocenter = justifies output, formdlim = overrides the internal page breaks

and replaces them with the designated symbol \*/

/\* Create titles \*/

title1 'Statstics 147 Assignment #4';

title2 'Fall 2017';

title3 'Sarah Ruckman';

title4 'Section 002';

title5 'SAS Question 1';

/\*Read infile using do loops do rows and then columns and there are 2 lines of headers\*/

/\*Create temporary SAS dataset called geysers\*/

data geysers;

infile "C:\Users\sarah\Downloads\geysers\_f17.dat" firstobs = 3;

do rows = 3 to 12;

do col = 1 to 3;

/\*Create if then else statements to label the columns and end with an else statement\*/

if col = 1 then name = 'OF';

else if col = 2 then name = 'GG';

else name = 'S ';

/\*input the values and add @@ to denote that the values are on the same line\*/

input values @@;

/\*Output the results\*/

output;

/\*Close both loops\*/

end;

end;

/\*Sort the data by column using proc sort\*/

proc sort;

by name;

/\*Print as check without observations\*/

proc print noobs;

/\*Use proc means to find the mean stddev and var for each geyser\*/

proc means mean stddev var;

by name;

/\*Use the set command and an if statement to create a data set with only GG\*/

/\*Create a temporary SAS dataset\*/

data onlyGG;

set geysers;

if col =2;

/\*Print the results\*/

proc print noobs;

/\*Use the set command and an if statement to create a data set with GG and S using a not=1\*/

/\*Create temporary SAS dataset\*/

data bothGGS;

set geysers;

if col not= 1;

/\*Print the results\*/

proc print noobs;

/\*Create new dataset with infile statement for question 2\*/

data dogdive;

infile "C:\Users\sarah\Downloads\dogdive\_f17.dat" firstobs = 3;

/\*input the variable feet\*/

input Dog$ Dive1 Dive2;

/\*Change title 5\*/

title5 'SAS Question 2';

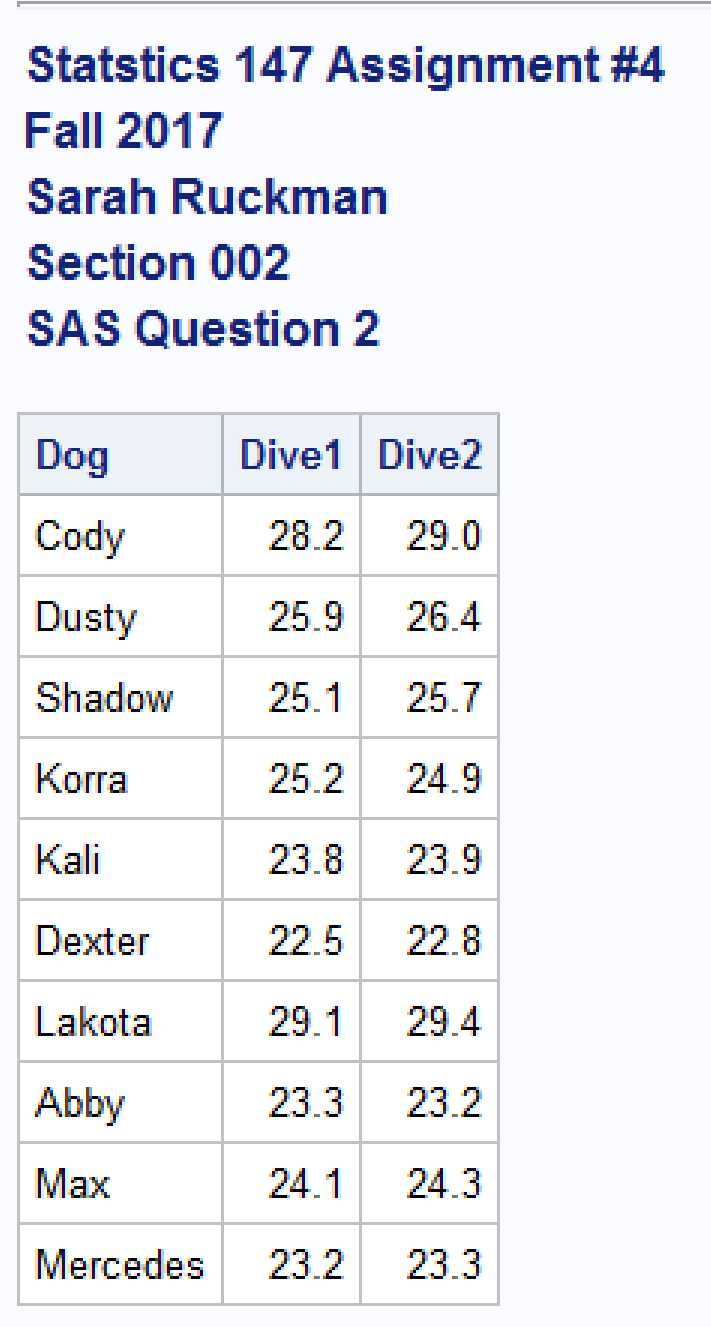
/\*Print the results without observation numbers\*/

proc print noobs;

run;

quit;

**Output:**



1. Generate the mean, median and standard deviation for the variable Dive1. (1 pt)

**SAS Code:**

options ls = **70** ps = **55** nocenter formdlim = '\*';

/\* ls = linesize, ps = pagesize, nocenter = justifies output, formdlim = overrides the internal page breaks

and replaces them with the designated symbol \*/

/\* Create titles \*/

title1 'Statstics 147 Assignment #4';

title2 'Fall 2017';

title3 'Sarah Ruckman';

title4 'Section 002';

title5 'SAS Question 1';

/\*Read infile using do loops do rows and then columns and there are 2 lines of headers\*/

/\*Create temporary SAS dataset called geysers\*/

**data** geysers;

infile "C:\Users\sarah\Downloads\geysers\_f17.dat" firstobs = **3**;

do rows = **3** to **12**;

do col = **1** to **3**;

/\*Create if then else statements to label the columns and end with an else statement\*/

if col = **1** then name = 'OF';

else if col = **2** then name = 'GG';

else name = 'S ';

/\*input the values and add @@ to denote that the values are on the same line\*/

input values @@;

/\*Output the results\*/

output;

/\*Close both loops\*/

end;

end;

/\*Sort the data by column using proc sort\*/

**proc** **sort**;

by name;

/\*Print as check without observations\*/

**proc** **print** noobs;

/\*Use proc means to find the mean stddev and var for each geyser\*/

**proc** **means** mean stddev var;

by name;

/\*Use the set command and an if statement to create a data set with only GG\*/

/\*Create a temporary SAS dataset\*/

**data** onlyGG;

set geysers;

if col =**2**;

/\*Print the results\*/

**proc** **print** noobs;

/\*Use the set command and an if statement to create a data set with GG and S using a not=1\*/

/\*Create temporary SAS dataset\*/

**data** bothGGS;

set geysers;

if col not= **1**;

/\*Print the results\*/

**proc** **print** noobs;

/\*Create new dataset with infile statement for question 2\*/

**data** dogdive;

infile "C:\Users\sarah\Downloads\dogdive\_f17.dat" firstobs = **3**;

/\*input the variable feet\*/

input Dog$ Dive1 Dive2;

/\*Change title 5\*/

title5 'SAS Question 2';

/\*Print the results without observation numbers\*/

**proc** **print** noobs;

/\*Use proc means statement to find the mean median and stddev\*/

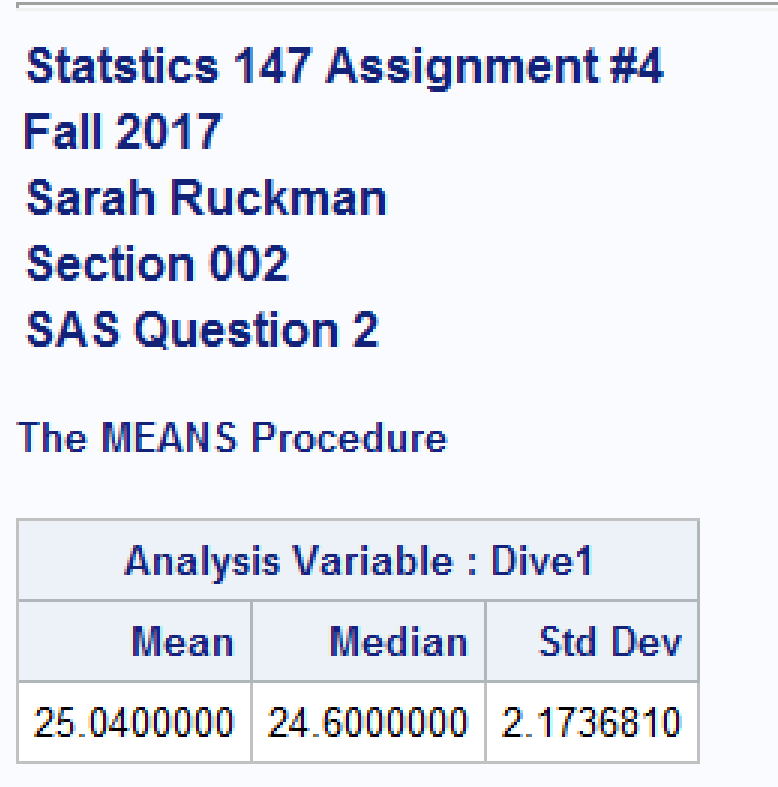
**proc** **means** mean median stddev;

var Dive1;

**run**;

**quit**;

**Output:**



**Using R:**

1. Refer to SAS Question 1.
   1. Write the R code to read in and print out the data. (2 pts)

**R Code:**

> #Statistics 147 Assignment 4

> #Fall 2017

> #Sarah Ruckman

> #Section 002

> #R Question 1

> #Use a read.table function to read in the data

> geysers<-read.table("C:\\Users\\sarah\\Downloads\\geysers\_f17.dat",header=TRUE,skip=1)

> #Print as check

> geysers

OF GG S

1 125 150 195

2 153 146 180

3 140 161 205

4 135 178 210

5 165 150 174

6 184 190 204

7 189 185 200

8 136 176 175

9 146 165 170

10 137 180 210

* 1. Add the appropriate lines of code to make the columns accessible individually and obtain the column names. (1 pt)

**R Code:**

> #Attach the columns to be used separately

> attach(geysers)

> #Use names function to see names

> names(geysers)

[1] "OF" "GG" "S"

> #Print as check

> OF

[1] 125 153 140 135 165 184 189 136 146 137

> GG

[1] 150 146 161 178 150 190 185 176 165 180

> S

[1] 195 180 205 210 174 204 200 175 170 210

1. Refer to SAS Question 2.
   1. Write the R code to read in and print out the data. (2 pts)

**R Code:**

> #R Question 2

> #Read in and print out the data using an read.table statement and skip function

> dogdive<-read.table("C:\\Users\\sarah\\Downloads\\dogdive\_f17.dat",header=TRUE,

skip=1)

> #Print as check

> dogdive

Dog Dive1 Dive2

1 Cody 28.2 29.0

2 Dusty 25.9 26.4

3 Shadow 25.1 25.7

4 Korra 25.2 24.9

5 Kali 23.8 23.9

6 Dexter 22.5 22.8

7 Lakota 29.1 29.4

8 Abby 23.3 23.2

9 Max 24.1 24.3

10 Mercedes 23.2 23.3

* 1. Add the appropriate lines of code to make the columns accessible individually and obtain the column names. (1 pt)

**R Code:**

> #Use attach function to separate the columns

> attach(dogdive)

> #Use names function to see the names

> names(dogdive)

[1] "Dog" "Dive1" "Dive2"

> #Print as check

> Dog

[1] Cody Dusty Shadow Korra Kali Dexter Lakota Abby

[9] Max Mercedes

Levels: Abby Cody Dexter Dusty Kali Korra Lakota Max Mercedes Shadow

> Dive1

[1] 28.2 25.9 25.1 25.2 23.8 22.5 29.1 23.3 24.1 23.2

> Dive2

[1] 29.0 26.4 25.7 24.9 23.9 22.8 29.4 23.2 24.3 23.3