**Assignment #2**

**Stats 147 Fall 2017 Sec. 2**

Sarah Ruckman

SID: 7194

**Using R**

1. A dog food manufacturer was interested in comparing the sales of its four main varieties of canned dog food: Beef & Rice, Chicken & Rice, Liver & Rice and Turkey & Rice. The data represents the national sales (in millions of dollars) for a random sample of 10 months and has been saved in the .csv file dogfood sales2 f17b.csv.

(i) (2 pts) Read in and print out the .csv file, dogfood sales2 f17b.csv.

**R Code:**

> #Stats 147 Assignment #2

> #Section 002

> #Fall 2017

> #Sarah Ruckman

> #R Question 1

> #Part i

> #Read in file (csv)

> dogfoodsales <-read.csv("C:\\Users\\sarah\\Downloads\\dogfood\_sales2\_f17b.csv", header = TRUE)

> #Print as check

> dogfoodsales

Beef Chicken Liver Turkey

1 1.40 1.70 1.65 1.58

2 1.75 1.85 1.72 1.77

3 1.58 1.36 1.50 1.48

4 1.65 2.05 1.37 1.69

5 1.55 1.80 1.60 1.65

6 1.45 2.10 1.40 1.65

7 1.66 1.95 1.75 1.79

8 1.70 1.65 1.38 1.58

9 1.85 1.80 1.65 1.77

10 1.24 2.00 1.55 1.60

(ii) (4 pts) Generate the following descriptive statistics for each variety. Complete the following table:

**R Code:**

> #R Question 1 part ii

> #Display names using names function

> names(dogfoodsales)

[1] "Beef" "Chicken" "Liver" "Turkey"

> #Use attach function to attach columns

> attach(dogfoodsales)

> #Print as check

> Beef

[1] 1.40 1.75 1.58 1.65 1.55 1.45 1.66 1.70 1.85 1.24

> Chicken

[1] 1.70 1.85 1.36 2.05 1.80 2.10 1.95 1.65 1.80 2.00

> Liver

[1] 1.65 1.72 1.50 1.37 1.60 1.40 1.75 1.38 1.65 1.55

> Turkey

[1] 1.58 1.77 1.48 1.69 1.65 1.65 1.79 1.58 1.77 1.60

> #Use length function to determine sample length

> length(Beef)

[1] 10

> length(Chicken)

[1] 10

> length(Liver)

[1] 10

> length(Turkey)

[1] 10

> #Use mean(x) to determine mean of sample

> mean(Beef)

[1] 1.583

> mean(Chicken)

[1] 1.826

> mean(Liver)

[1] 1.557

> mean(Turkey)

[1] 1.656

> #Use sd(x) to find standard deviation

> sd(Beef)

[1] 0.1804962

> sd(Chicken)

[1] 0.2203129

> sd(Liver)

[1] 0.1403211

> sd(Turkey)

[1] 0.1004656

> #Use max(x) to determine maximum value

> max(Beef)

[1] 1.85

> max(Chicken)

[1] 2.1

> max(Liver)

[1] 1.75

> max(Turkey)

[1] 1.79

> #Use min(x) to determine minimum value

> min(Beef)

[1] 1.24

> min(Chicken)

[1] 1.36

> min(Liver)

[1] 1.37

> min(Turkey)

[1] 1.48

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Variety** | **# of Obs** | **Mean** | **Standard Deviation** | **Max Value** | **Min Value** |
| **Beef** | **10** | **1.583** | **0.1804962** | **1.85** | **1.24** |
| **Chicken** | **10** | **1.826** | **0.2203129** | **2.10** | **1.36** |
| **Liver** | **10** | **1.557** | **0.1403211** | **1.75** | **1.37** |
| **Turkey** | **10** | **1.656** | **0.1004656** | **1.79** | **1.48** |

2. A sales representative for a dog food manufacturing company was interested in determining whether there is a difference in preference among it main varieties of canned dog food: Beef & Rice, Chicken & Rice, Liver & Rice and Turkey & Rice. A random sample of 19 customers was selected and asked to name their preference. The data was recorded and saved in a .csv file named dogfood prefer f17b.csv.

(i) (2 pts) Read in and print out the .csv file, dogfood prefer f17b.csv.

**R Code:**

> #R Question 2 part i

> #Read in dog food prefer csv file using read.csv

> dogfoodprefer <- read.csv("C:\\Users\\sarah\\Downloads\\dogfood\_prefer\_f17b.csv", header =TRUE)

> #Print as check

> dogfoodprefer

Prefer1

1 liver

2 chicken

3 chicken

4 beef

5 turkey

6 liver

7 chicken

8 chicken

9 beef

10 beef

11 chicken

12 chicken

13 beef

14 turkey

15 turkey

16 chicken

17 liver

18 beef

19 turkey

(ii) (4 pts) Create a vertical bar chart for the preference of the dog food.

1. Use the following colors for the bars of the chart: Blue, Pink, Yellow and Green.

**R Code:**

> #R Question 2 Part ii part a

> #create a vertical bar chart with bars being colors: Blue, Pink, Yellow, and Green

> #Create frequency table

> dogfoodptable = table(dogfoodprefer)

> #Print as check

> dogfoodptable

dogfoodprefer

beef chicken liver turkey

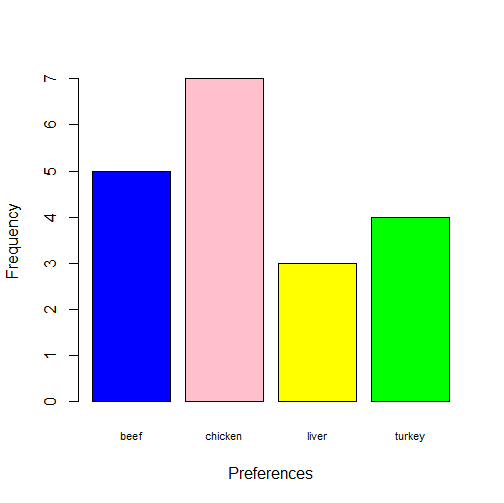
5 7 3 4

> #Create list of names of colors to be used

> colors = c("Blue", "Pink", "Yellow", "Green")

> #Create bar chart with labels shruck with cex.names = 0.7 to read the labels, col to attach colors, and xlab/ylab to make labels

> barplot(dogfoodptable, xlab = "Preferences", ylab = "Frequency", col = colors, cex.names = 0.70)



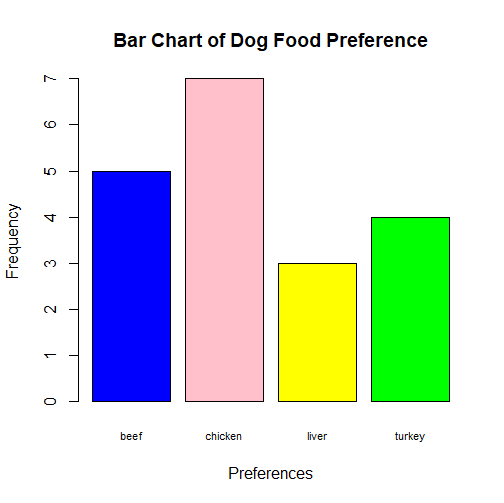
1. Include the following title: Bar Chart of Dog Food Preference

**R Code:**

> #R Question 2 part ii part b

> #Use main function to create title

> barplot(dogfoodptable, xlab = "Preferences", ylab = "Frequency", main = "Bar Chart of Dog Food Preference", col = colors, cex.names = 0.70)



(iii) (1 pt) Identify the variety that has the largest preference. (Be sure to state the number that prefer that variety.)

**See frequency table and code above. Chicken has the largest preference with 7 dogs preferring that variety.**

**Using SAS**

A dog food manufacturer was interested in comparing the sales of its four main varieties of canned dog food: Beef & Rice, Chicken & Rice, Liver & Rice and Turkey & Rice. The data represents the national sales (in millions of dollars) for a random sample of 10 months.

1. Write a SAS program to complete the following:

(i) (3 pts) Read in and print out the data. 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 gunit=pct border cback =white ctext=black colors=(blue green red) ftext=swissb ftitle=swissb htitle=5 htext=2.75 hpos = 10;

**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 \*/

/\* Set up some options for gchart and gplot \*/

goptions reset=global gunit=pct border cback =white ctext=black colors=(blue green red) ftext=swissb ftitle=swissb htitle=**5** htext=**2.75** hpos = **10**;

/\* Create titles \*/

title1 'Statstics 147 Assignment #2';

title2 'Section 002';

title3 'Fall 2017';

title4 'Sarah Ruckman';

title5 'SAS Question 1 Part i';

/\* Read file \*/

/\* Create temporary data set called dfsales \*/

**data** dfsales;

infile "C:\Users\sarah\Downloads\DOGFOOD\_SALES\_F17A.DAT" firstobs = **2** ;

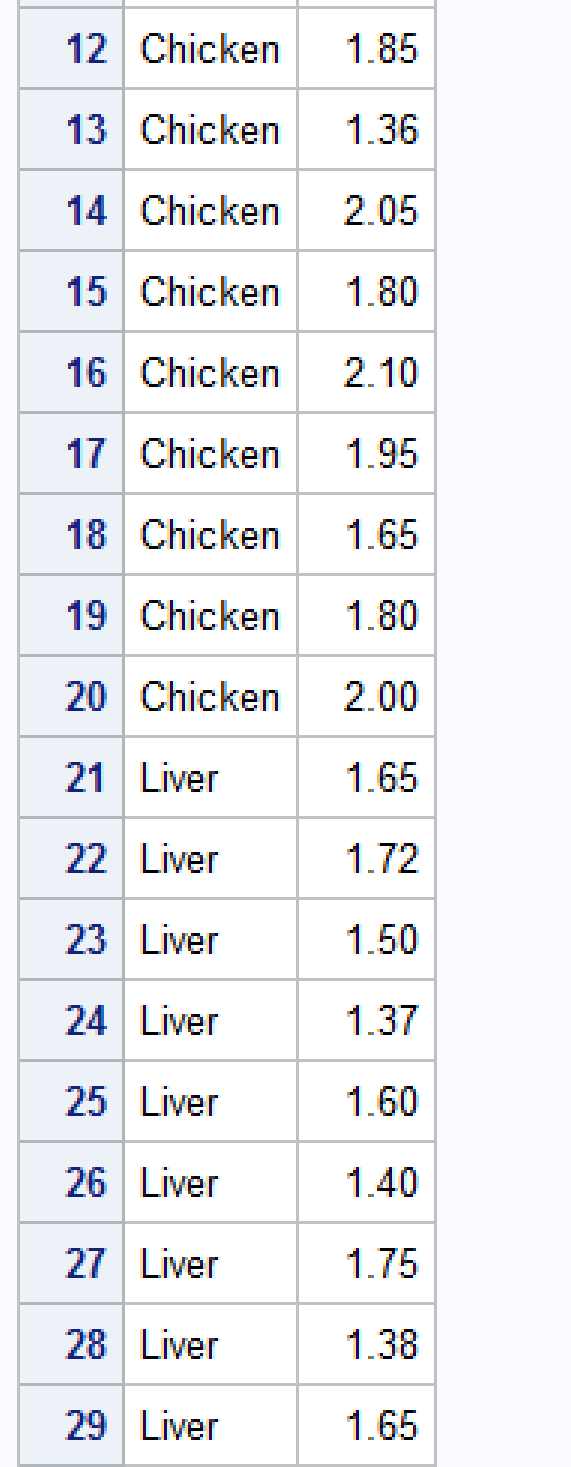
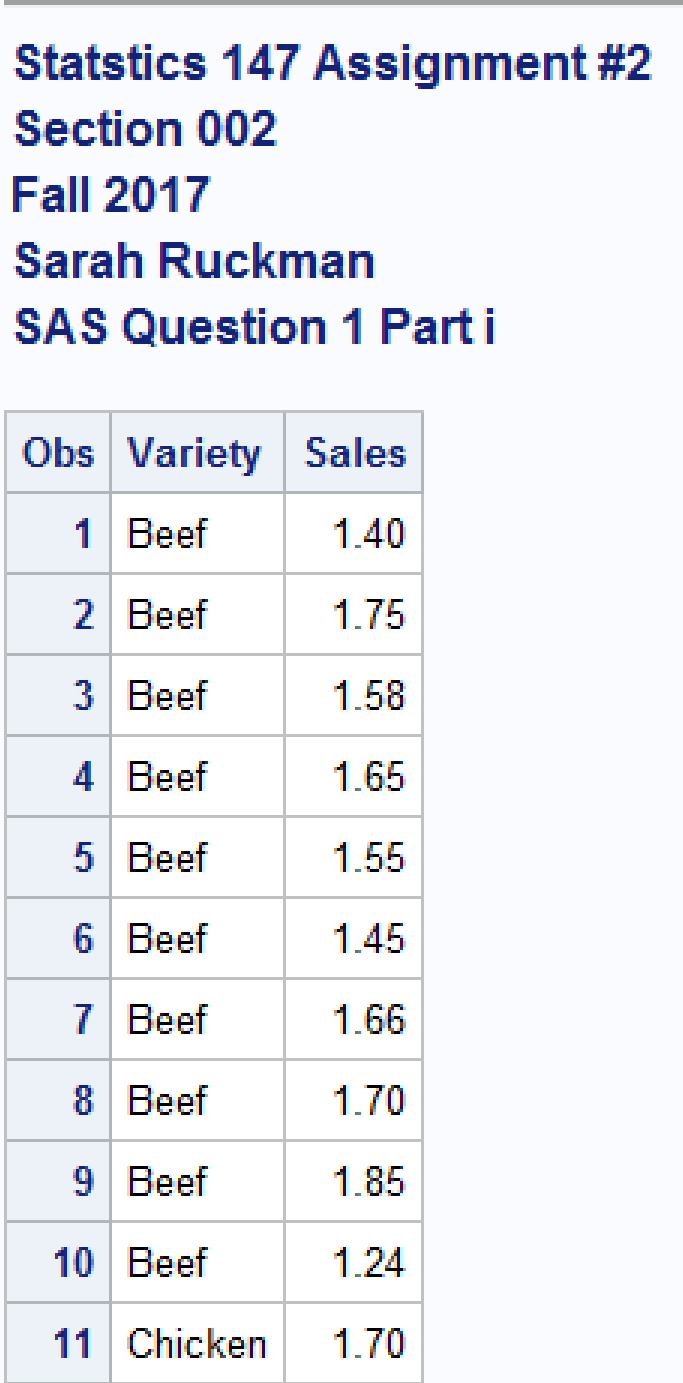
input Variety $ Sales;

/\* Print the data as a check \*/

**proc** **print**;

**run**;

**quit**;



(ii) (4 pts) Modify your program to generate a 3-D vertical bar chart for the sales of dogfood.

1. Be sure to include some coloration, including verylightgrayishblue 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 lightblue, pink, yellow and lightgreen as the colors for the bars. (You do not have to submit your graph in color, but your code must reflect the color scheme.)

(b) Be sure to include the sumvar, outside and subgroup options.

**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 \*/

/\* Set up some options for gchart and gplot \*/

goptions reset=global gunit=pct border cback =white ctext=black colors=(blue green red) ftext=swissb ftitle=swissb htitle=**5** htext=**2.75** hpos = **10**;

/\* Create titles \*/

title1 'Statstics 147 Assignment #2';

title2 'Section 002';

title3 'Fall 2017';

title4 'Sarah Ruckman';

title5 'SAS Question 1 Part ii Part b';

/\* Read file \*/

/\* Create temporary data set called dfsales \*/

**data** dfsales;

infile "C:\Users\sarah\Downloads\DOGFOOD\_SALES\_F17A.DAT" firstobs = **2** ;

input Variety $ Sales;

/\* Print the data as a check \*/

**proc** **print**;

/\* Create 3D vbar chart using proc gchart statement, include coloration: background frame = verylightgrayishblue,

outline = verydarkblue, hexagon shape of bars, text within = verydarkgreen,

bar colors = lightblue, pink, yellow, and lightgreen \*/

**proc** **gchart**;

vbar3d Variety / sumvar = Sales

outside = sum

subgroup = Variety

cframe = verylightgrayishblue

coutline = verydarkblue

shape = hexagon

ctext = verydarkgreen;

pattern1 color = lightblue;

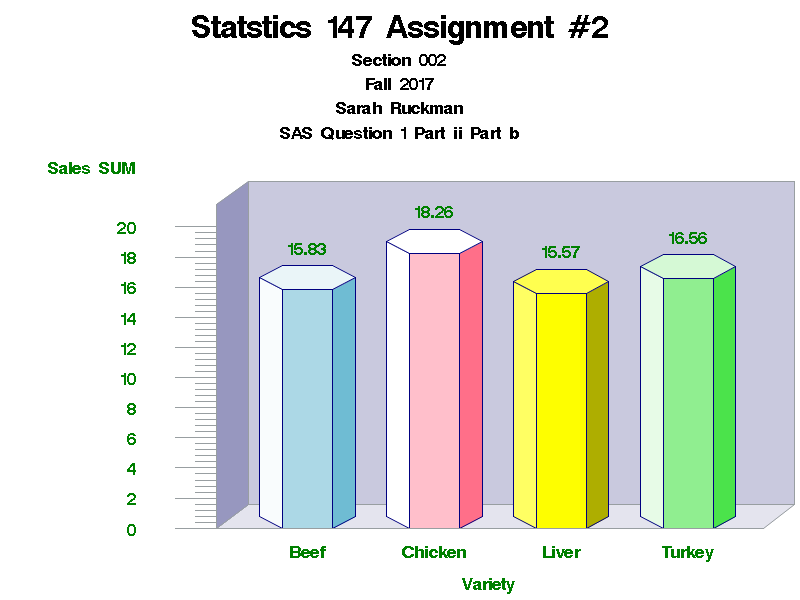
pattern2 color = pink;

pattern3 color = yellow;

pattern4 color = lightgreen;

**run**;

**quit**;

****

(iii) (4 pts) Modify your program to generate the descriptive statistics for the sales for each variety. Identify the mean, median, and standard deviation for each of the varieties.

**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 \*/

/\* Set up some options for gchart and gplot \*/

goptions reset=global gunit=pct border cback =white ctext=black colors=(blue green red) ftext=swissb ftitle=swissb htitle=**5** htext=**2.75** hpos = **10**;

/\* Create titles \*/

title1 'Statstics 147 Assignment #2';

title2 'Section 002';

title3 'Fall 2017';

title4 'Sarah Ruckman';

title5 'SAS Question 1 Part iii';

/\* Read file \*/

/\* Create temporary data set called dfsales \*/

**data** dfsales;

infile "C:\Users\sarah\Downloads\DOGFOOD\_SALES\_F17A.DAT" firstobs = **2** ;

input Variety $ Sales;

/\* Print the data as a check \*/

**proc** **print**;

/\* Create 3D vbar chart using proc gchart statement, include coloration: background frame = verylightgrayishblue,

outline = verydarkblue, hexagon shape of bars, text within = verydarkgreen,

bar colors = lightblue, pink, yellow, and lightgreen \*/

**proc** **gchart**;

vbar3d Variety / sumvar = Sales

outside = sum

subgroup = Variety

cframe = verylightgrayishblue

coutline = verydarkblue

shape = hexagon

ctext = verydarkgreen;

pattern1 color = lightblue;

pattern2 color = pink;

pattern3 color = yellow;

pattern4 color = lightgreen;

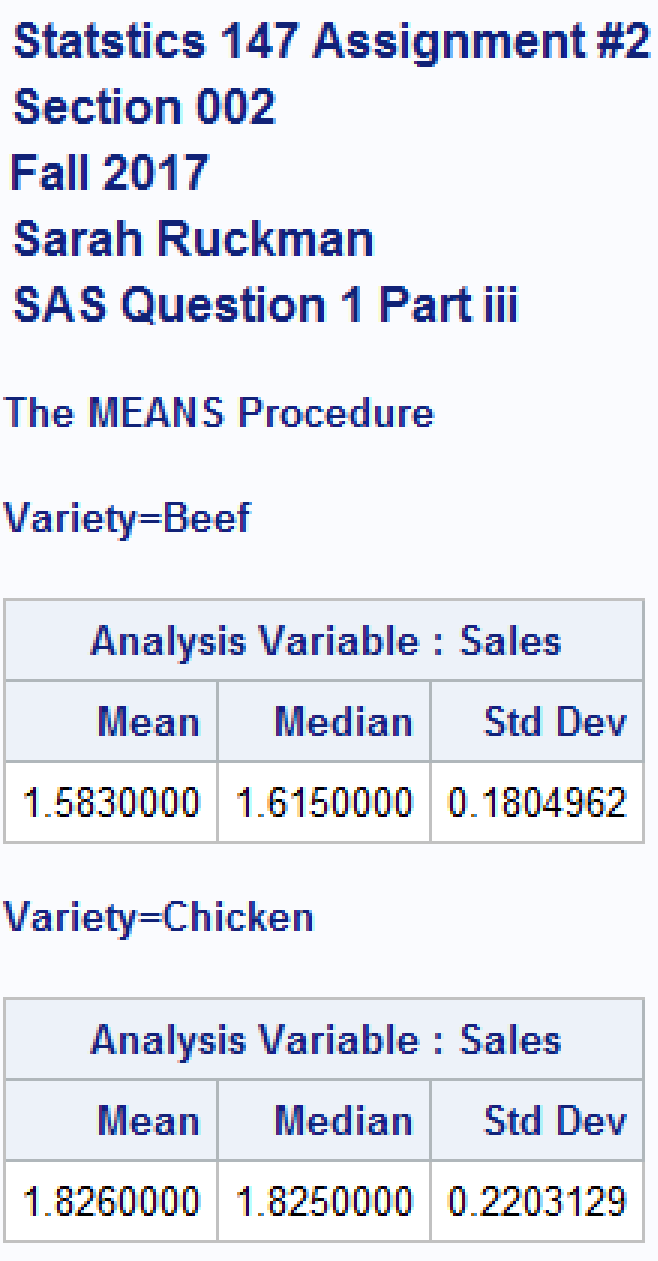
/\* Use proc means statement to specify which statistics to display \*/

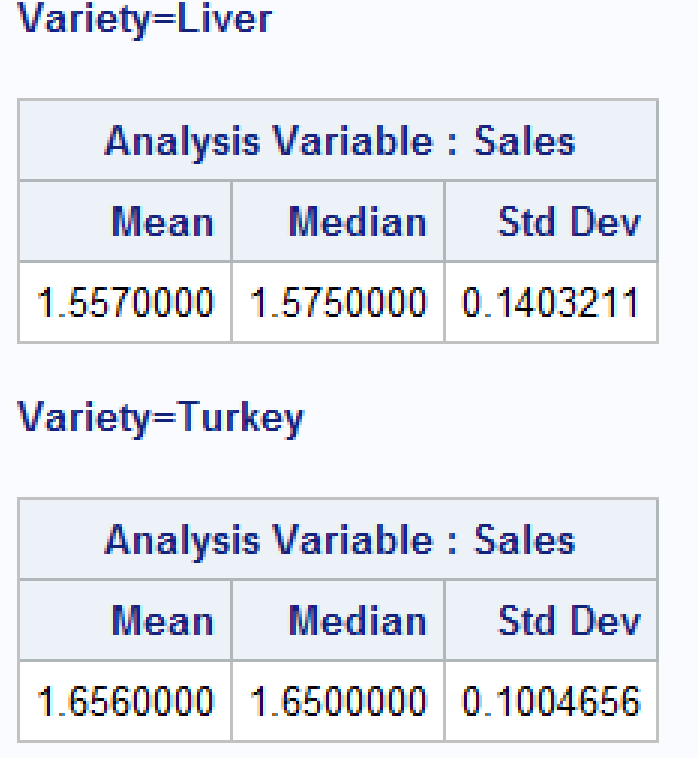
**proc** **means** mean median stddev;

by Variety;

**run**;

**quit**;





|  |  |  |  |
| --- | --- | --- | --- |
| **Variety** | **Mean** | **Median** | **Standard Deviation** |
| **Beef** | **1.583** | **1.615** | **0.1804962** |
| **Chicken** | **1.826** | **1.825** | **0.2203129** |
| **Liver** | **1.557** | **1.575** | **0.1403211** |
| **Turkey** | **1.656** | **1.65** | **0.1004656** |