HW 3 Report

Yongchao Qiao

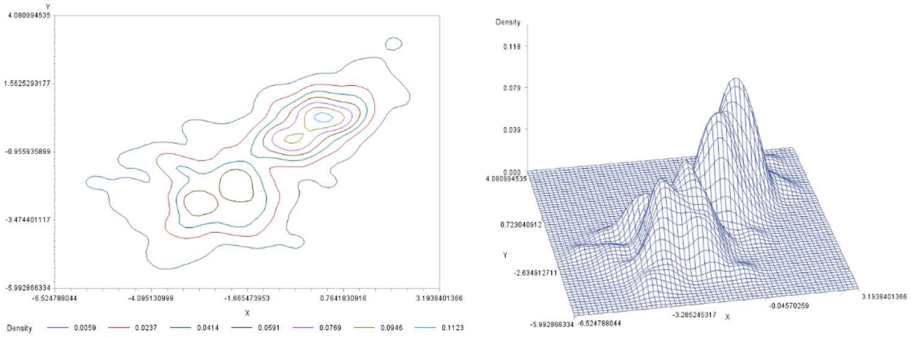
Part A

1. The contour plots and corresponding perspective plots

The contour plots and corresponding perspective plots are shown as below.

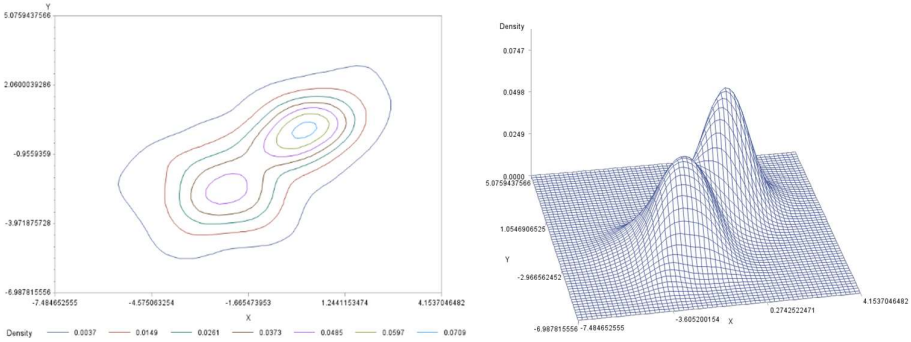
Band width: (0.5, 0.5)

Figure 1: The contour plot and perspective plot for band width (0.5, 0.5)



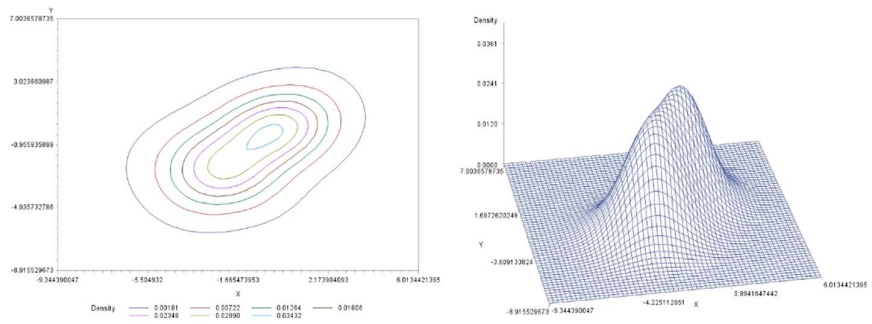
Band width: (1, 1)

Figure 2: The contour plot and perspective plot for band width (1, 1)



Band width: (2, 2)

Figure 3: The contour plot and perspective plot for band width (2, 2)



2. The best density estimation

These plots of different bandwidths tell different stories about the original data. For the plots of the band width as (0.5, 0.5), it shows too much details and it is under-smoothed, while the plots of the band width as (2, 2) shows too little details and it is oversmoothed. Besides, the plots of the band width as (1, 1) shows moderate details and smooth which indicate useful information about the original data. Therefore, the band width as (1, 1) provides the best density estimation about the data.

3. The underlying joint density of (*X*, *Y*)

Based on former plots of different band widths, the band widths as (1, 1) is inferred as the best density estimation after several comparisons. For the plots of band width as (1, 1), it obviously shows two peaks. Since simple normal reference for the bandwidth method and the normal-like shape, I guess the underlying joint density of (X, Y) should follow a mixed bivariate normal distribution.

Part B

1. The relationship between the variables weight and age

Weight vs. Age

$$H_0 : \text{Rho} = 0, \quad H_1 : \text{Rho} \neq 0$$

The result is shown as table 1. The correlation coefficient between Weight and Age is 0.63464 with P value less than 0.05. Thus, it indicates that the null hypothesis that Rho = 0 can be rejected. That is, variable Weight and variable Age have a significant correlation relationship, with Pearson correlation coefficient as 0. 63464.

Table 1: Pearson Correlation Test between Weight and Age|

	Weight	Age
Weight	1.00000	0.63464(<0.0001)
Age	0.63464(<0.0001)	1.00000

Weight vs. Age, by controlling height

$$H_0 : \text{Rho} = 0, \quad H_1 : \text{Rho} \neq 0$$

Table 2: Partial Pearson Correlation Test between Weight and Age|

	Weight	Age
Weight	1.00000	0.27413(<0.0001)
Age	0.27413(<0.0001)	1.00000

The result is shown as table 2. The partial Pearson correlation coefficient between Weight and Age is 0. 27413 with P value less than 0.05. Thus, it indicates that the null hypothesis that

Rho = 0 can be rejected. That is, even if conditional on height, variable Weight and variable Age still have a significant correlation relationship, with partial Pearson correlation coefficient as 0. 27413.

Weight vs. Age, by controlling height, in different sex groups

$$H_0 : Rho = 0, \quad H_1 : Rho \neq 0$$

Table 3: Partial Pearson Correlation Test between Weight and Age, in different sex groups

Sex		Weight	Age
Male	Weight	1.00000	0.31430(0.0004)
	Age	0.31430(0.0004)	1.00000
Female	Weight	1.00000	0.23605(0.0130)
	Age	0.23605(0.0130)	1.00000

The results are shown as table 3. In the male group, the partial Pearson correlation coefficient between Weight and Age is 0.31430 with P value less than 0.05. Thus, it indicates that the null hypothesis that Rho = 0 can be rejected. That is, even if conditional on height, variable Weight and variable Age, in the male group, still have a significant correlation relationship, with partial Pearson correlation coefficient as 0. 31430. In the female group, the partial Pearson correlation coefficient between Weight and Age is 0.23605 with P value less than 0.05. Thus, it indicates that the null hypothesis that Rho = 0 can be rejected. That is, even if conditional on height, variable Weight and variable Age, in the female group, still have a significant correlation relationship, with partial Pearson correlation coefficient as 0. 23605. Therefore, in different sex groups, conditional on height, there is still a correlation between Weight and Age.

2. The simple linear regression and local regression between weight and age

Simple linear regression

Table 3: Analysis of variance

Source	DF	Sum of squares	Mean Square	F Value	Pr > F
Model	1	35924	35924	158.48	<0.0001
Error	235	53270	226.68055		
Corrected Total	236	89194			

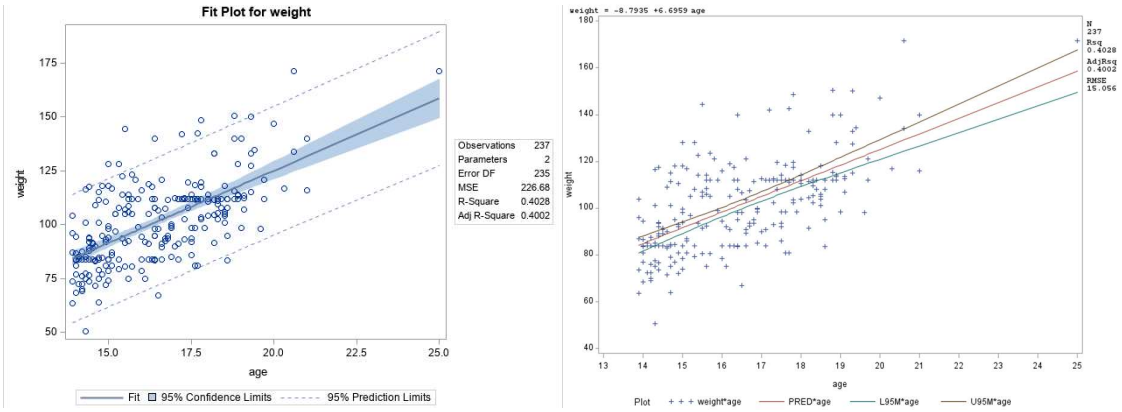
Table 4: Parameter Estimates

Variable	DF	Parameter Estimates	Standard Error	t Value	Pr > |t|
Intercept	1	-8.79351	8.80047	-1.00	0.3187
Age	1	6.69594	0.53189	12.59	<0.0001

Table 5: Statistical indicators

Root MSE	Dependent mean	Coeff Var	R-Square	Adj R-sq
15.05591	101.30802	14.86152	0.4208	0.4002

Figure 4: The scatter plot with 95% confidence interval for simple linear regression

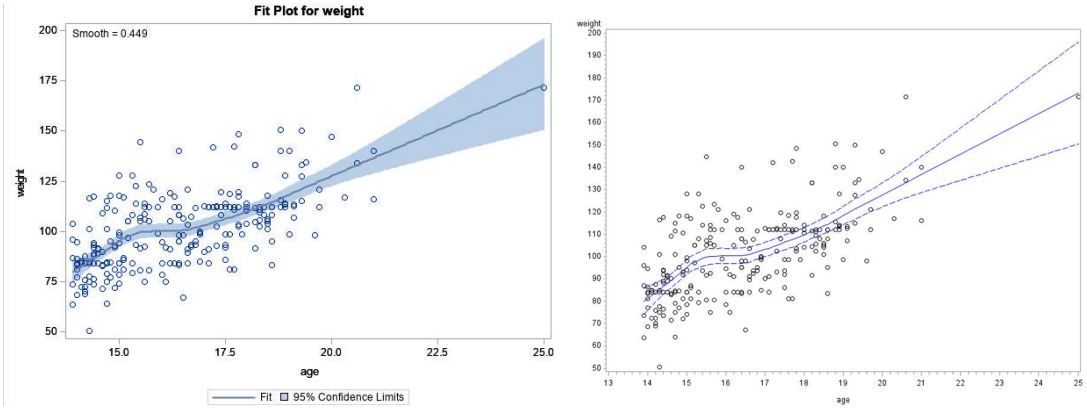


The scatter plots with 95% confidence interval for simple linear regression are shown above. The table 3 shows the p-value of the F test is less than 0.05, which indicates the simple linear regression model is significant to describe the relationship between variable Weight and variable Age. Besides, the table 4 shows the p-value of the t test for the estimated coefficient of Age is less than 0.05, which indicates the variable Age is significantly to explain the change of variable Weight. Then the table 5 shows the R-Square as 0.4208, which is not that big, indicates variable Age can only explain the change of variable Weight in a certain degree.

Local regression

The scatter plot with 95% confidence interval for local regression is shown below.

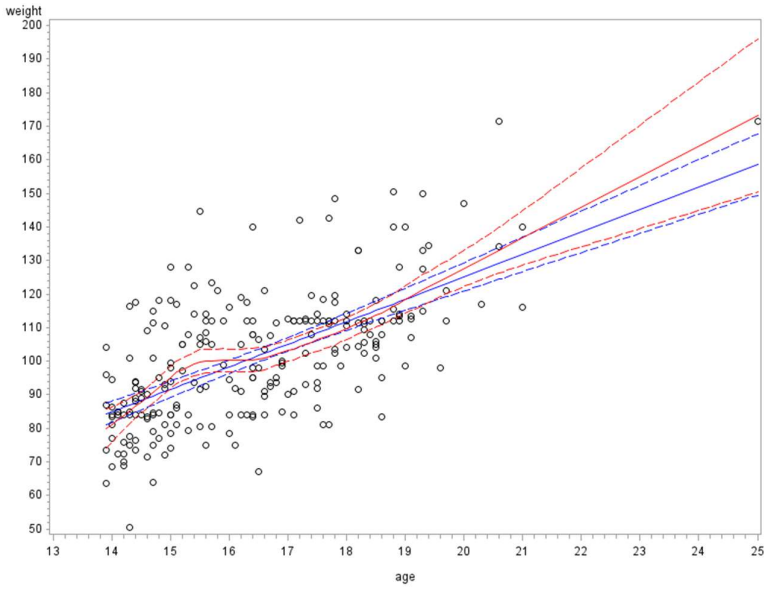
Figure 5: The scatter plot with 95% confidence interval for local regression



Comparison

The scatter plot with 95% confidence interval for simple linear and local regression are shown below. The figure shows that there are many overlapped parts between the 95% confidence intervals of simple linear regression and local regression. So the model fitting based on the simple linear regression and the local regression gives the consistent result.

Figure 6: The scatter plot with 95% confidence interval for simple linear and local regression



Appendix

```
/*Part A*/
proc import datafile = 'E:/GW/Textbook/Data Analysis/HW3/HW3a.csv' /*read the file into
sas*/
dbms = csv /*specify the format of the file*/
out=work.HW3a;      /*specify the saved dataset in sas*/
getnames=yes;      /*get the name of the variables from the original file*/
run;    /*run this procedure*/

/* KDE (0.5,0.5)*/
proc kde data=HW3a out=kdeout0_5 bwm=.5,.5;
   var X Y; /* Kenel density estimation with band width as (0.5, 0.5) */
run;

proc gcontour data=kdeout0_5;
   plot Y*X=density; /*Plot the contour plot with band width as (0.5, 0.5)*/
run;

proc g3d data=kdeout0_5;
  plot Y*X=density /rotate=10 tilt=45; /* Plot the perspective plot with band width as (0.5,
0.5)*/
run;

/* KDE (1,1)*/
proc kde data=HW3a out=kdeout1 bwm=1,1;
   var X Y;/* Kenel density estimation with band width as (1, 1) */
run;

proc gcontour data=kdeout1;
   plot Y*X=density; /*Plot the contour plot with band width as (1, 1)*/
run;

proc g3d data=kdeout1;
  plot Y*X=density /rotate=10 tilt=45; /* Plot the perspective plot with band width as (1, 1)*/
run;

/* KDE (2,2)*/
proc kde data=HW3a out=kdeout2 bwm=2,2;
   var X Y; /* Kenel density estimation with band width as (2, 2) */
run;

proc gcontour data=kdeout2;
   plot Y*X=density; /*Plot the contour plot with band width as (2, 2)*/
run;
```

```
proc g3d data=kdeout2;
  plot Y*X=density /rotate=10 tilt=45; /* Plot the perspective plot with band width as (2, 2)*/
run;

/*Part B*/
proc import datafile = 'E:/GW/Textbook/Data Analysis/HW3/HW3b.csv' /*read the file into
sas*/
dbms = csv /*specify the format of the file*/
out=work.HW3b;      /*specify the saved dataset in sas*/
getnames=yes;      /*get the name of the variables from the original file*/
run;    /*run this procedure*/

/* Weight vs age */
proc corr data=hw3b;
var weight age;    /* calculate the Pearson correlation coefficient and take the Pearson
correlation test */
run;

/* Weight vs age, by controlling height */
proc corr data=hw3b;
var weight age;
partial height; /* calculate the partial Pearson correlation coefficient and take the partial
Pearson correlation test */
run;

/* Weight vs age, by controlling height, in different sex group */
proc corr data=hw3b;
var weight age;
partial height;
by sex;
run; /* calculate the partial Pearson correlation coefficient and take the partial Pearson
correlation test */

proc sort data=hw3b;
by age;
run; /* Sort the data by variable age */

goptions reset=all;
proc reg data=HW3b;
model weight=age; /* Fit the data with simple linear regression model */
output out=regout p=pr uclm=upper lclm=lower; /*Output estimated values */
plot weight*age / conf; /*Plot the scatter plot with 95% confidence interval*/
run;
```

```
proc loess data=HW3b;
model weight=age / clm alpha=0.05; /* Fit the data with local regression model */
ods output Outputstatistics=lofit;/*Output estimated values */
run; /*Plot the scatter plot with 95% confidence interval*/


data both;
   set regout;
   set lofit;
run; /* Combine the two output datasets*/

goptions reset=all;
symbol1 v=circle i=none c=black;
symbol2 v=none i=join c=blue l=1;
symbol3 v=none i=join c=blue l=3;
symbol4 v=none i=join c=blue l=3;
symbol5 v=none i=join c=red l=1;
symbol6 v=none i=join c=red l=3;
symbol7 v=none i=join c=red l=3; /*specify the point type, color and line type */
proc gplot data=both;
   plot (weight pr lower upper pred LowerCL UpperCL)*age / overlay;
run; /*plot the scatter plot with 95% confidence intervals of simple linear regression and
local regression */

proc gplot data=both;
   plot (weight pred LowerCL UpperCL)*age / overlay;
run; /*Plot the scatter plot with 95% confidence interval for local regression*/
```