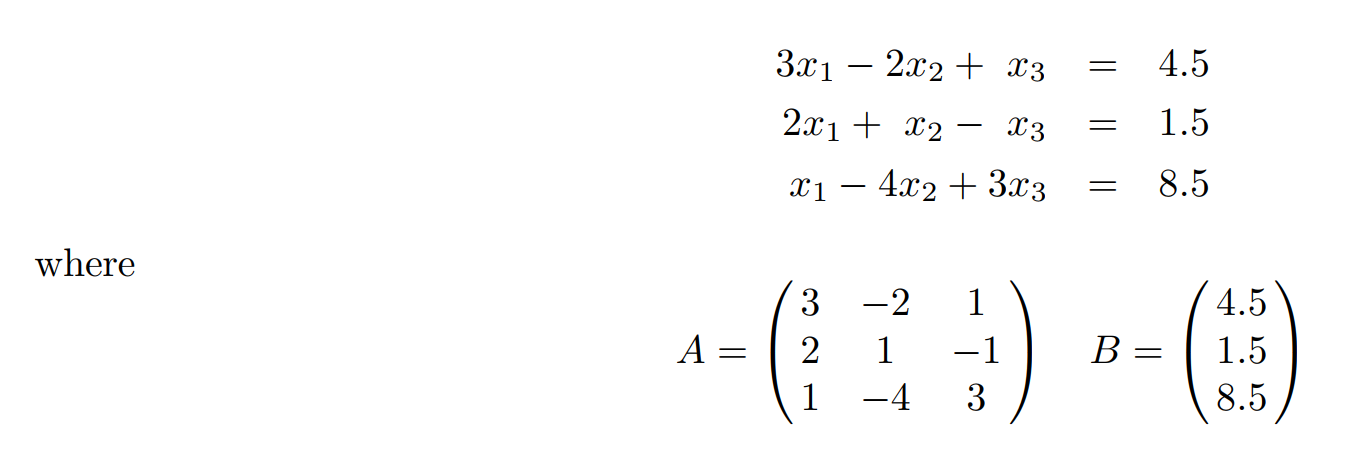
**Assignment #3**

**Stats 157 Winter 2018**

Sarah Ruckman

SID: 7194

1. Consider the following system of equations: (Ax = B)



Using SAS, complete the following.

1. Find the trace of matrix A. (1 pt)

**SAS Code:**

/\*Set up options and turn off extra graphics\*/

options nocenter nodate nonumber ps=55 ls=78;

ods graphics off;

/\*goptions formats the plot

cback color of the plot background

colors colors to use

ftitle font of plot title

htitle height of the title

htext height of the text on the plot \*/

goptions reset = all colors=(blue,red,green,purple) ftitle = swissb ftext=swissb htitle=3;

/\*Create new SAS temporary dataset\*/

data mat1;

/\*Set up titles\*/

title1 'Statistics 157 Winter 2018';

title2 'Assignment #3';

title3 'Sarah Ruckman';

title4 'Question 1 Part i';

/\*Invoke proc iml to complete matrix calculations\*/

proc iml;

/\*initialize matricies\*/

A = {3 -2 1, 2 1 -1, 1 -4 3};

B = {4.5, 1.5, 8.5};

/\*Use trace(A) to find the trace of matrix A\*/

TraceA = trace(A);

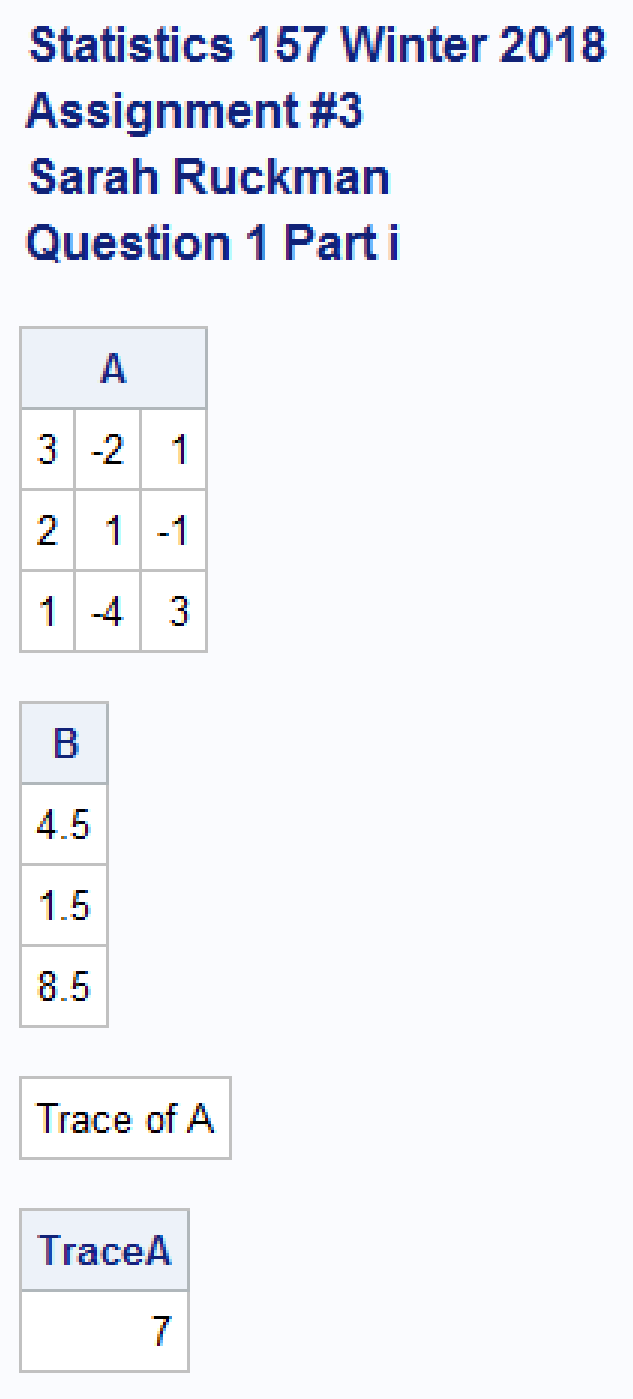
/\*Print the results\*/

print ,,, A, B, 'Trace of A',, TraceA;

run;

quit;

**Output:**



**The trace of A = 7.**

1. Find the product A\*B. (1 pt)

**SAS Code:**

/\*Find the product of A and B by creating a new variable PROD1\*/

PROD1 = A\*B;

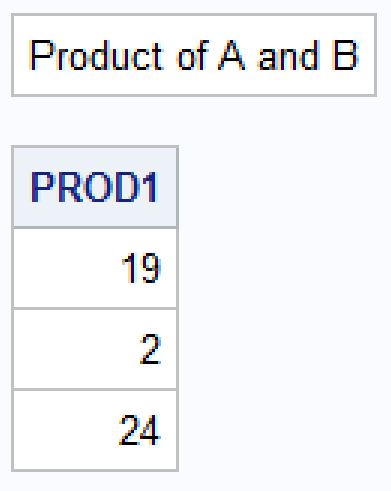
/\*Print the results\*/

print ,,,'Product of A and B',, PROD1;

run;

quit;

**Output:**



**The product of A and B is**

1. Find the determinant of matrix A. (2 pts)

**SAS Code:**

/\*Find the determinat of A\*/

/\*Revise title4\*/

title4 'Question 1 Part iii';

DET1 = det(A);

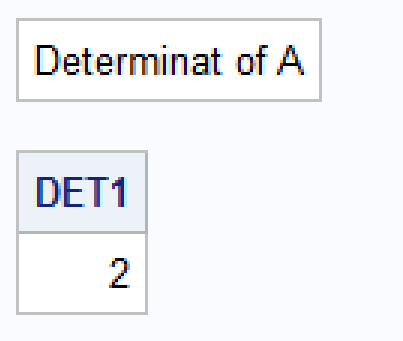
/\*Print the results\*/

print ,,, 'Determinat of A', DET1;

run;

quit;

**Output:**



**The determinant of A is 2.**

1. Find the inverse of A, if it exists. (If the inverse does not exist, make sure your SAS program prints why it does not exist!) (3 pts)

**SAS Code:**

/\*Find the inverse of matrix A\*/

/\*Revise title4\*/

title4 'Question 1 Part iv';

/\*Use do loops with if then else to get the answer\*/

if DET1 = **0** then

do;

print ,,, 'Since the determinant = 0, the matrix A is singular and does not have an inverse';

end;

else

do;

AINV = inv(A);

/\*Check Product\*/

PROD2 = A\*AINV;

/\*Print the results\*/

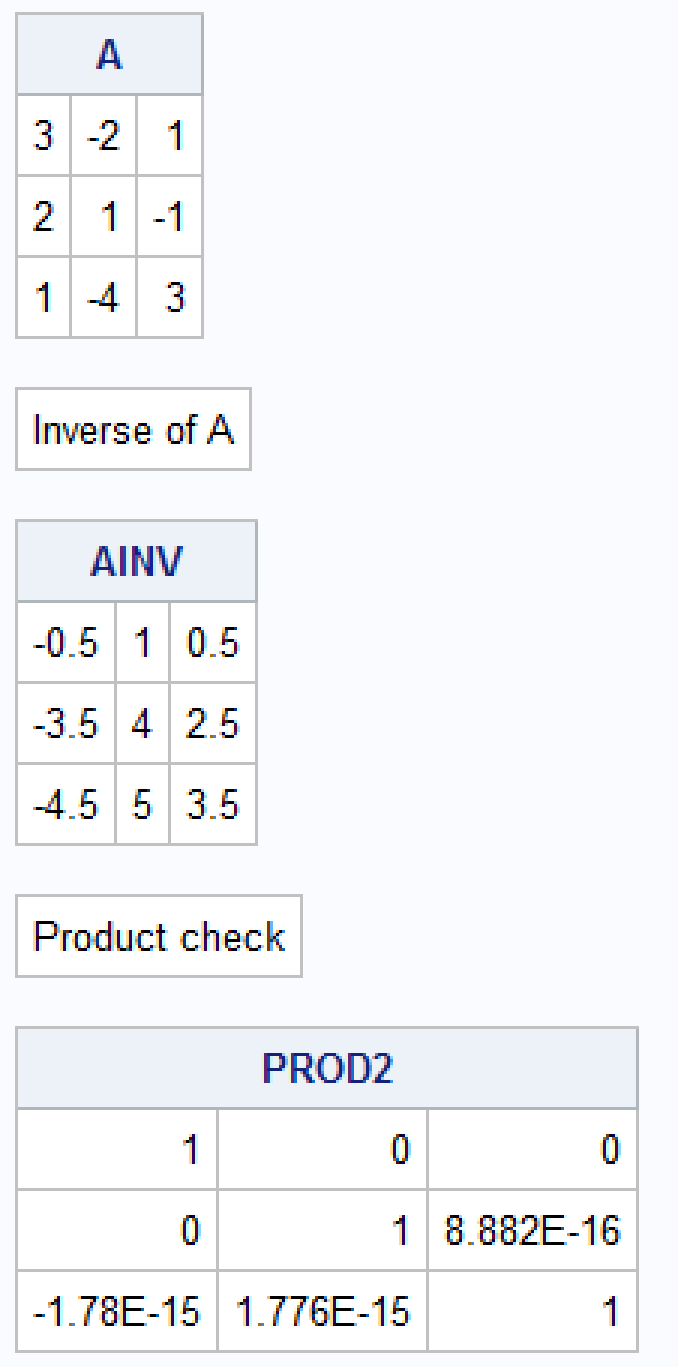
print ,,, A, 'Inverse of A', AINV, 'Product check', PROD2;

end;

**run**;

**quit**;

**Output:**



**The inverse of A is**

1. Find the solution to the system of equations, if it exists. (If the solution does not exist, make sure your SAS program prints why it does not exist!) (3 pts)

**SAS Code:**

/\*Find the solution to the system of equations using AINV\*B\*/

/\*Revise title4\*/

title4 'Question 1 part v';

SOLN = AINV\*B;

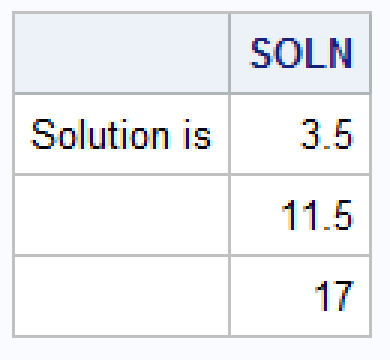
/\*Print the results\*/

print ,,, 'Solution is' SOLN;

**run**;

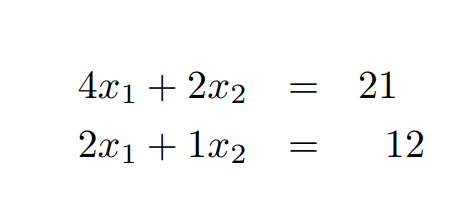
**quit**;

**Output:**



**The solution is**

1. Consider the following system of equations: (Ax = B)



1. Find the determinant of A. (1 pt)

**SAS Code:**

/\*Create new SAS temporary dataset\*/

data q2;

/\*Revise title4\*/

title4 'Question 2 part i';

/\*Invoke proc iml to complete matrix calculations\*/

proc iml;

/\*Create matricies\*/

A = {4 2, 2 1};

B = {21, 12};

/\*Use det(A) to find the determinant of A\*/

DETA = det(A);

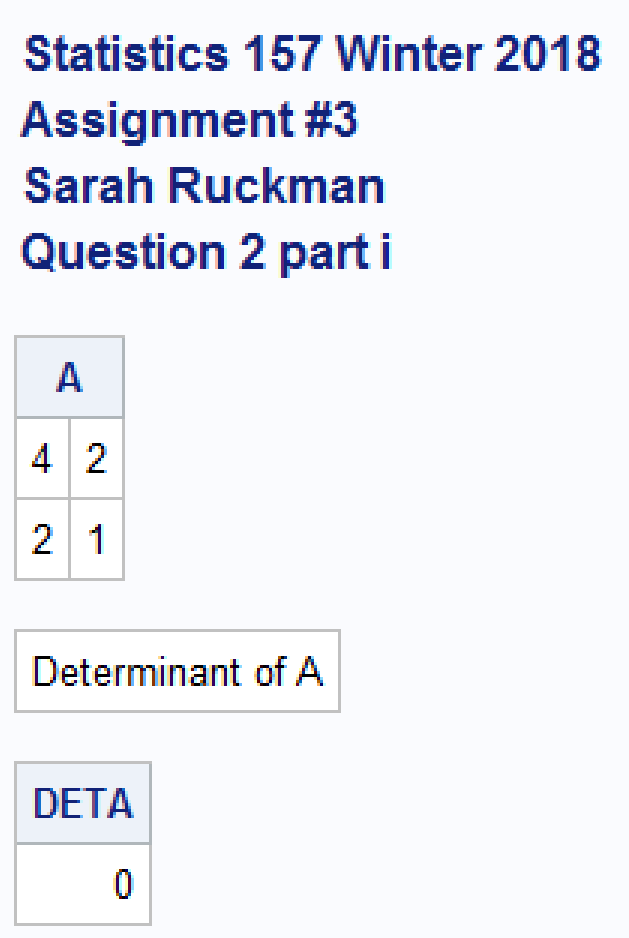
/\*Print the results\*/

print ,,,A, 'Determinant of A', DETA;

run;

quit;

**Output:**



**The determinant of A is 0.**

1. Find the inverse of A, if it exists. (If the inverse does not exist, make sure your SAS program prints why it does not exist!) (2 pts)

**SAS Code:**

/\*Find the inverse of matrix A\*/

/\*Revise title4\*/

title4 'Question 2 Part ii';

/\*Use do loops with if then else to get the answer\*/

if DETA = **0** then

do;

print ,,, 'Since the determinant = 0, the matrix A is singular and does not have an inverse';

end;

else

do;

AINV = inv(A);

/\*Check Product\*/

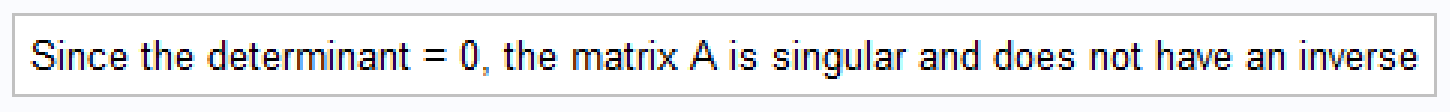
PROD2 = A\*AINV;

/\*Print the results\*/

print ,,, A, 'Inverse of A', AINV, 'Product check', PROD2;

end;

**Output:**



**There is no inverse.**

1. Find the solution to the system of equations, if it exists. (If the solution does not exist, make sure your SAS program prints why it does not exist!) (2 pts)

**SAS Code:**

/\*Find the solution to the system of equations using AINV\*B\*/

/\*Revise title4\*/

title4 'Question 2 part iii';

/\*Use do loops with if then else to get the answer\*/

if DETA = **0** then

do;

print,'Since the determinant = 0, the matrix A is singular and does not have an inverse nor a solution to the system of equations';

end;

else

do;

INVA = inv(A);

SOLN1 = INVA\*B;

/\*Print the results\*/

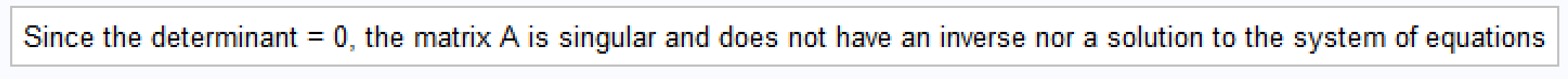
print ,,, 'Inverse of A', AINV, 'The solution is' SOLN1;

end;

**run**;

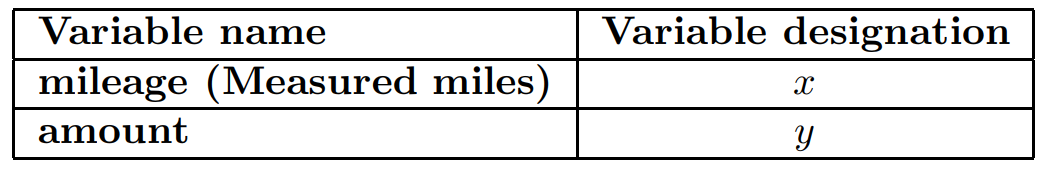
**quit**;

**Output:**



**There is no solution.**

1. Anyone that has owned a vehicle knows repair bills can be expensive. Luke was interested in developing a model that would represent the relationship between mileage of a vehicle and the amount of a repair bill. They obtained a random sample of 15 repair bills and recorded the data in a file named repair1 w18.dat. The variable names and variable designations are as follows:



NOTE: You may use the variable name or the variable designation in your program file!

1. Write a SAS program to read in and print out the data using a temporary SAS dataset named assign3q3 (i.e., data assign3q3;). (3 pts)

**SAS Code:**

/\*Create a new SAS temporary dataset called assign3q3\*/

**data** assign3q3;

/\*Revise title4\*/

title4 'Question 3 part i';

/\*Read in the data using an infile statement it starts on line 3\*/

infile 'C:\Users\sarah\Downloads\REPAIR1\_w18.dat' firstobs = **3**;

/\*Input the variables x (Mileage) and y (Amount)\*/

input x y @@;

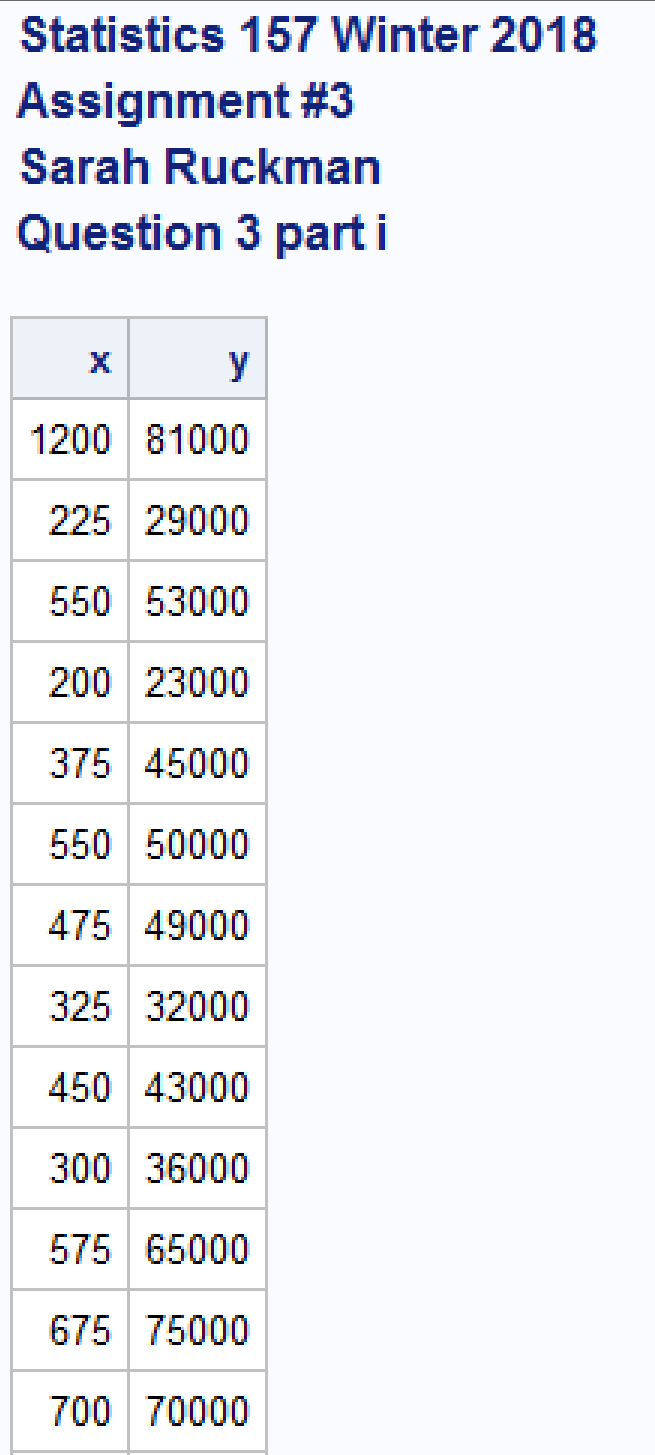
/\*Print as check\*/

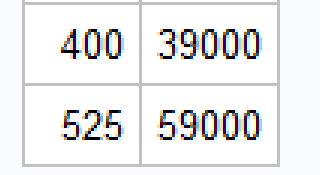
**proc** **print** noobs;

**run**;

**quit**;

**Output:**





1. Create a high-resolution plot (using proc gplot) of y versus x. (2 pts)

**SAS Code:**

/\*Create a high resolution plot of y versus x\*/

/\*Revise title4\*/

title4 'Question 3 Part ii';

/\*Use a symbol statement to set up the format of the plt symbols

value symbol of the data points

height height of the symbol of the data points

cv color of the symbol\*/

symbol1 value = star height = **3** cv = red;

/\*Use proc gplot to generate the high resolution plot

plot vertical(y) vs horizontal(x)

caxis color of the axes

ctext color of the text on the plot\*/

**proc** **gplot**;

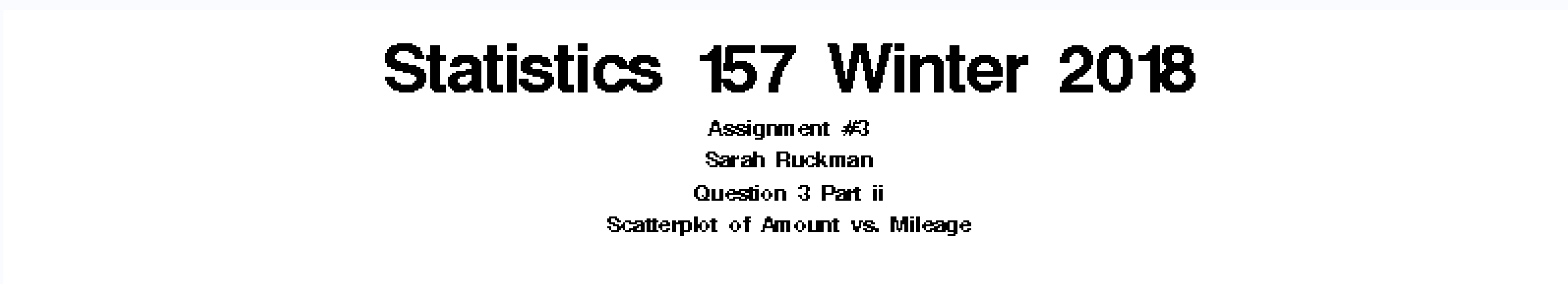
title5 'Scatterplot of Amount vs. Mileage';

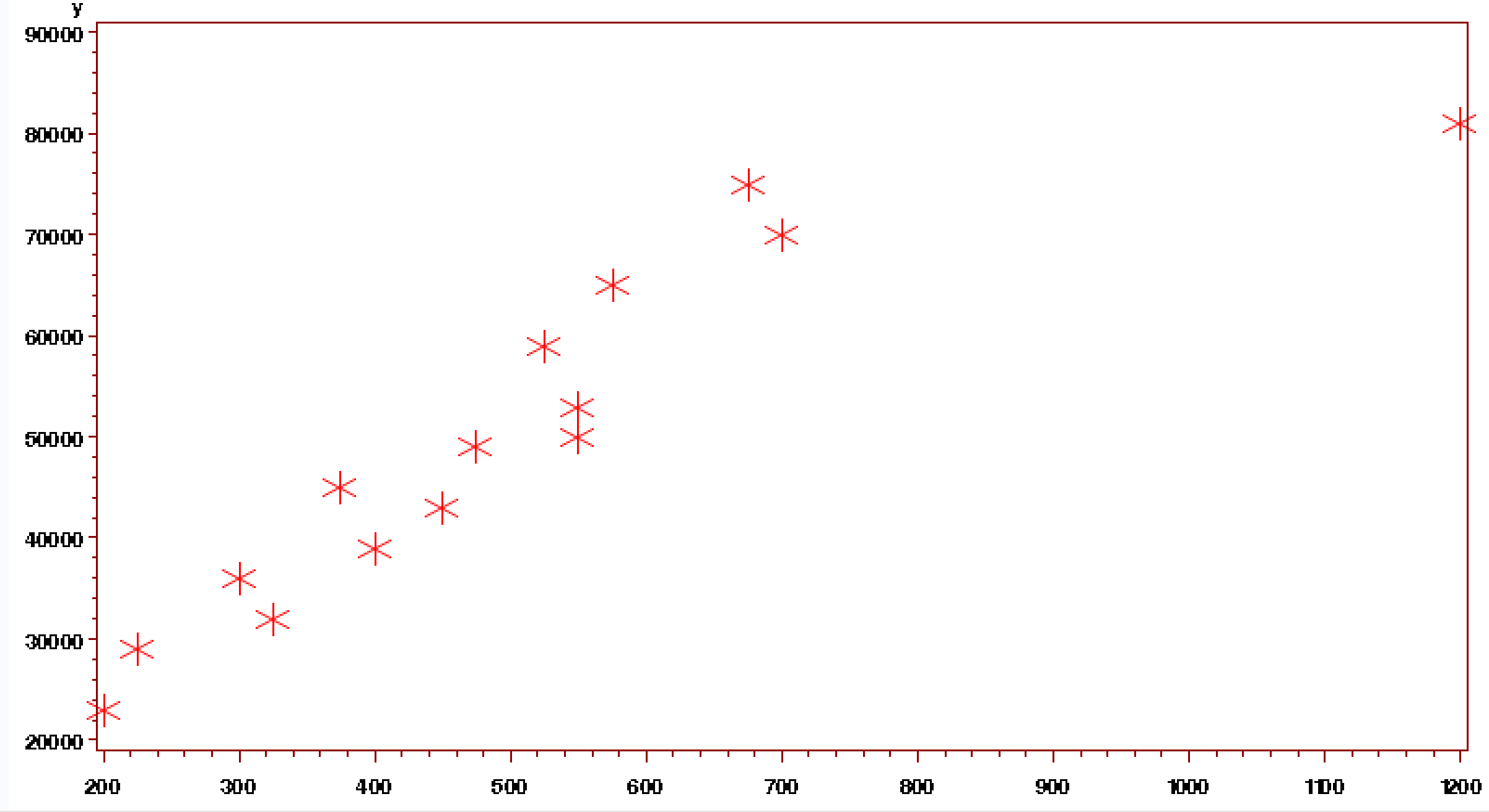
plot y\*x / caxis = darkred ctext=black;

**run**;

**quit**;

**Output:**





1. Find and interpret the correlation between mileage and amount of repair bill. (2)

**SAS Code:**

/\*Find the correlation coefficient using proc corr

Use proc corr to generate correlation between x and y

Use nosimple to supress the printing of the descriptive stats

Use noprob to suppress printing of the stats for testing the correlation = 0\*/

**proc** **corr** nosimple noprob;

/\*Revise title 4 and 5\*/

title4 'Question 3 Part iii';

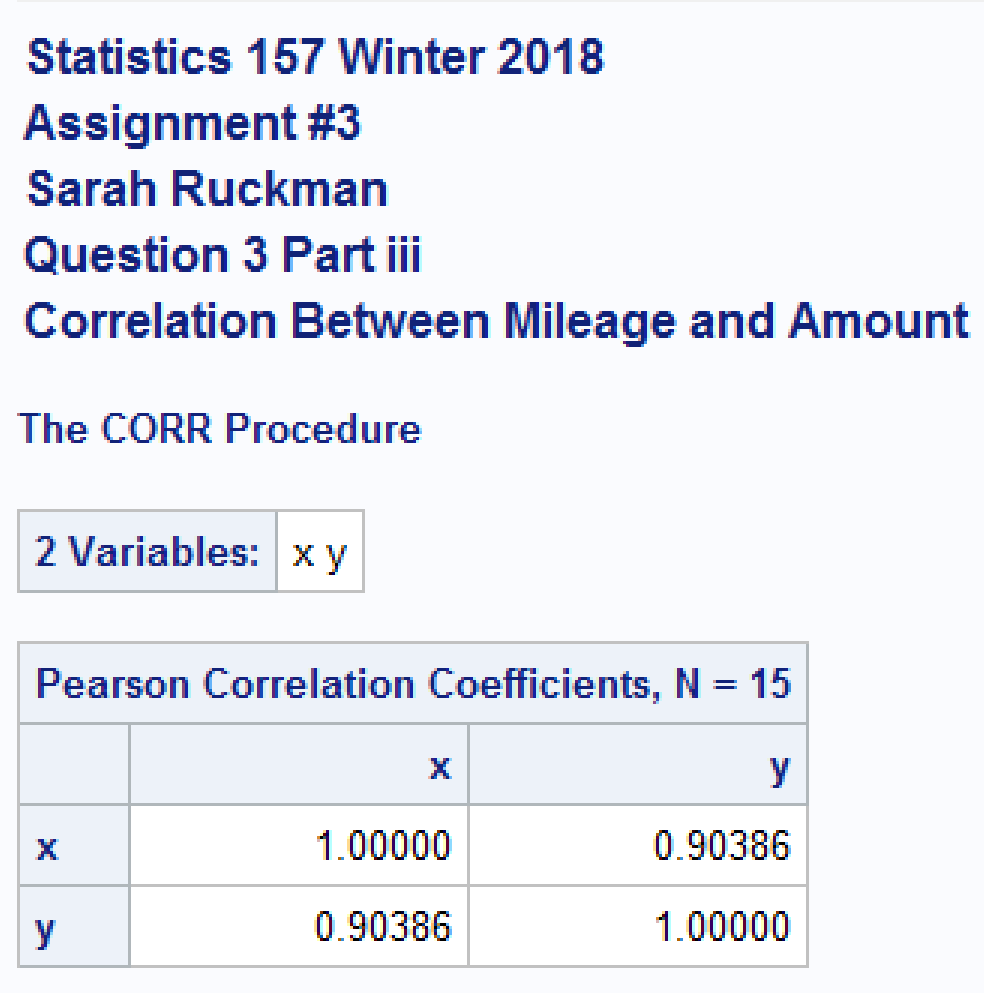
title5 'Correlation Between Mileage and Amount';

var x y;

**run**;

**quit**;

**Output:**



**The correlation coefficient is 0.90386. There seems to be a strong positive linear relationship between amount and mileage because the correlation coefficient = 0.90386 is greater than 0.8.**

1. Find the estimated regression equation, ˆy = βˆ 0 + βˆ 1x. (2 pts)

**SAS Code:**

/\*Use proc reg to compute ANOVA table for linear regression

model dependent\_variable = independent\_variable\*/

**proc** **reg**;

/\*Revise title 4 and 5\*/

title4 'Question 3 Part iv';

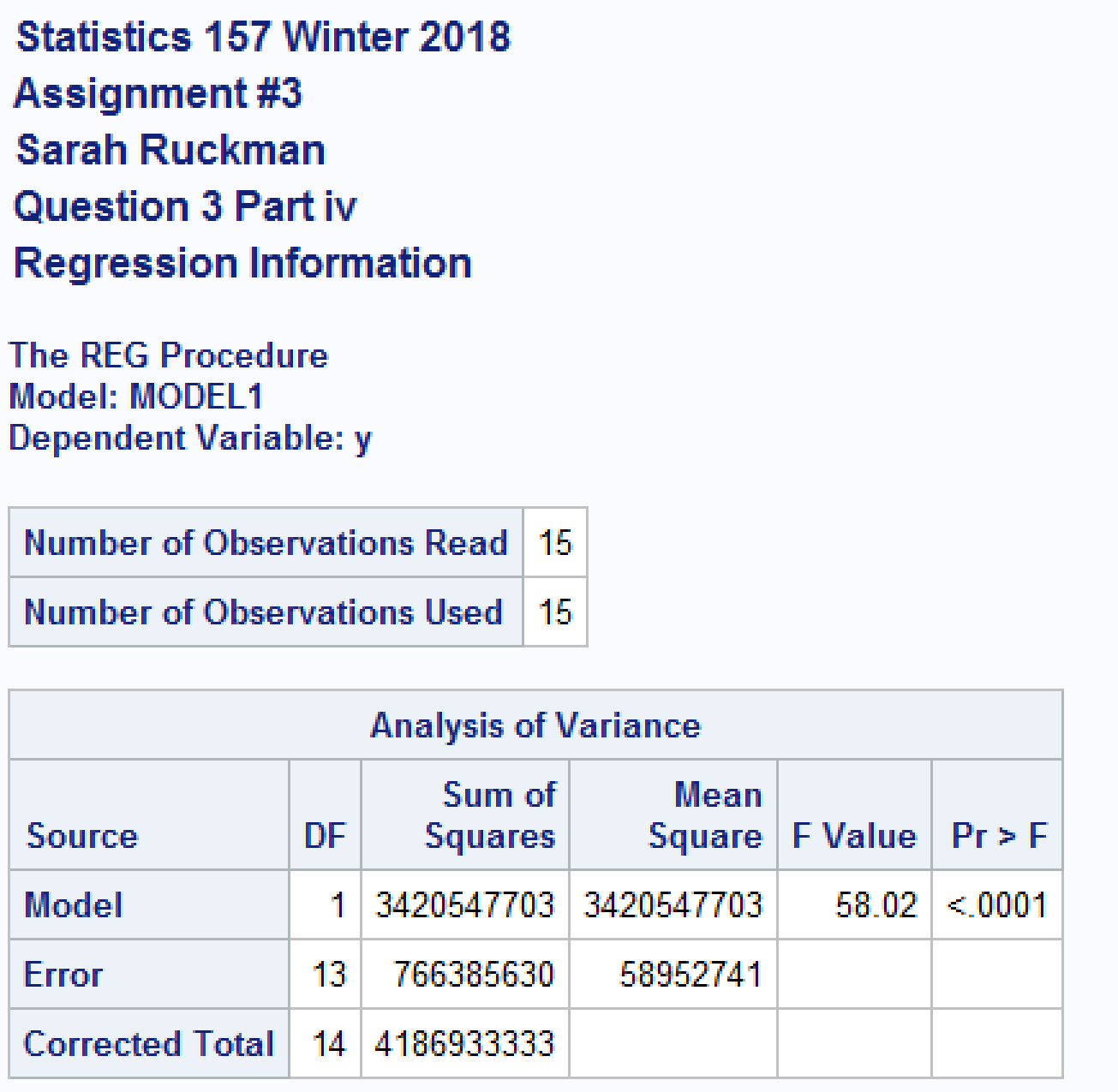
title5 'Regression Information';

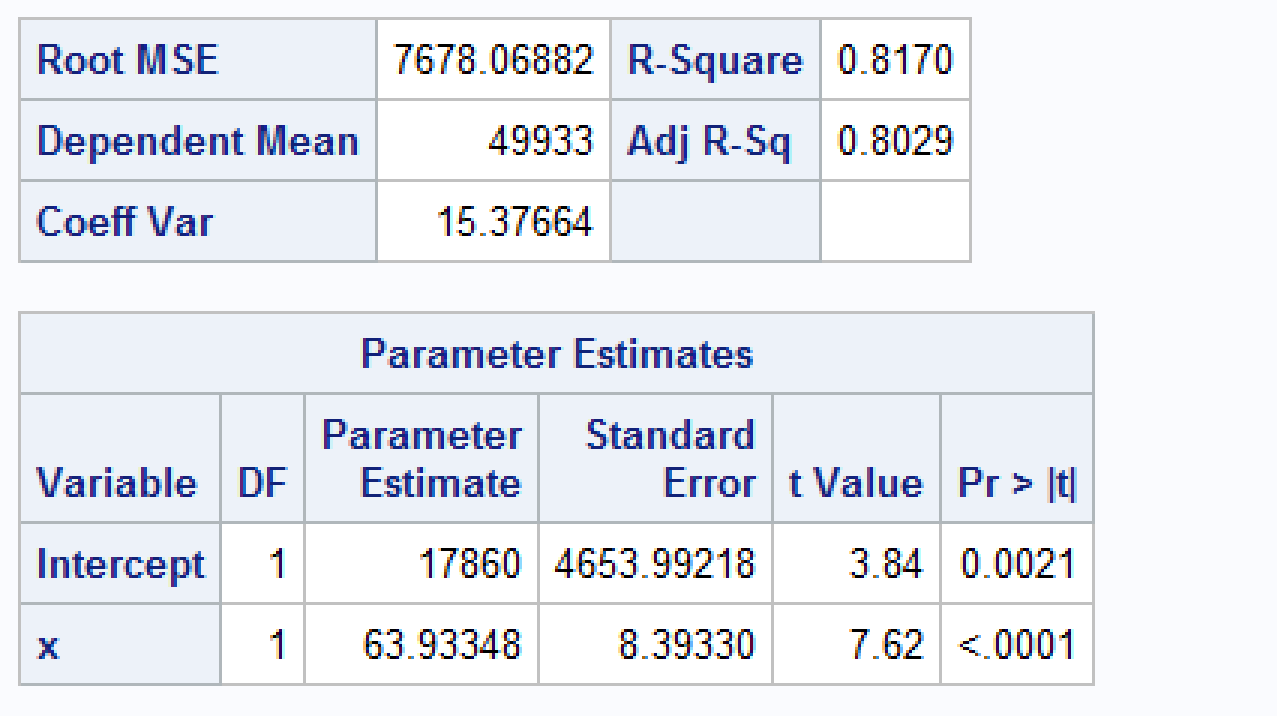
model y = x;

**run**;

**quit**;

**Output:**





**The regression equation is ŷ = 17860 + 63.93348x**

1. Find and interpret the coefficient of determination for the model in part (iv). (2pt)

**SAS Code:**

**No new code needed.**

**Output:**

**See above**

**The coefficient of determination, R2 = 0.8170. 81.70% of the variability in amount of the repair bill can be explained by the mileage on the car according to the model.**

1. Generate the predicted values, residuals and student (studentized) residuals for the model in part (iv). Do any of the data points appear to be potential outliers? If so, which one(s) and why. If not, indicate why none of the data points appear to be outliers. (2 pts)

**SAS Code:**

/\*Use proc reg to compute ANOVA table for linear regression. Use the P and R options to generate the predicted residual

values Output the data so that a residual plot may be generated\*/

**proc** **reg**;

/\*Revise title 4 and 5\*/

title4 'Question 3 Part vi';

title5 ' ';

model y = x / P R;

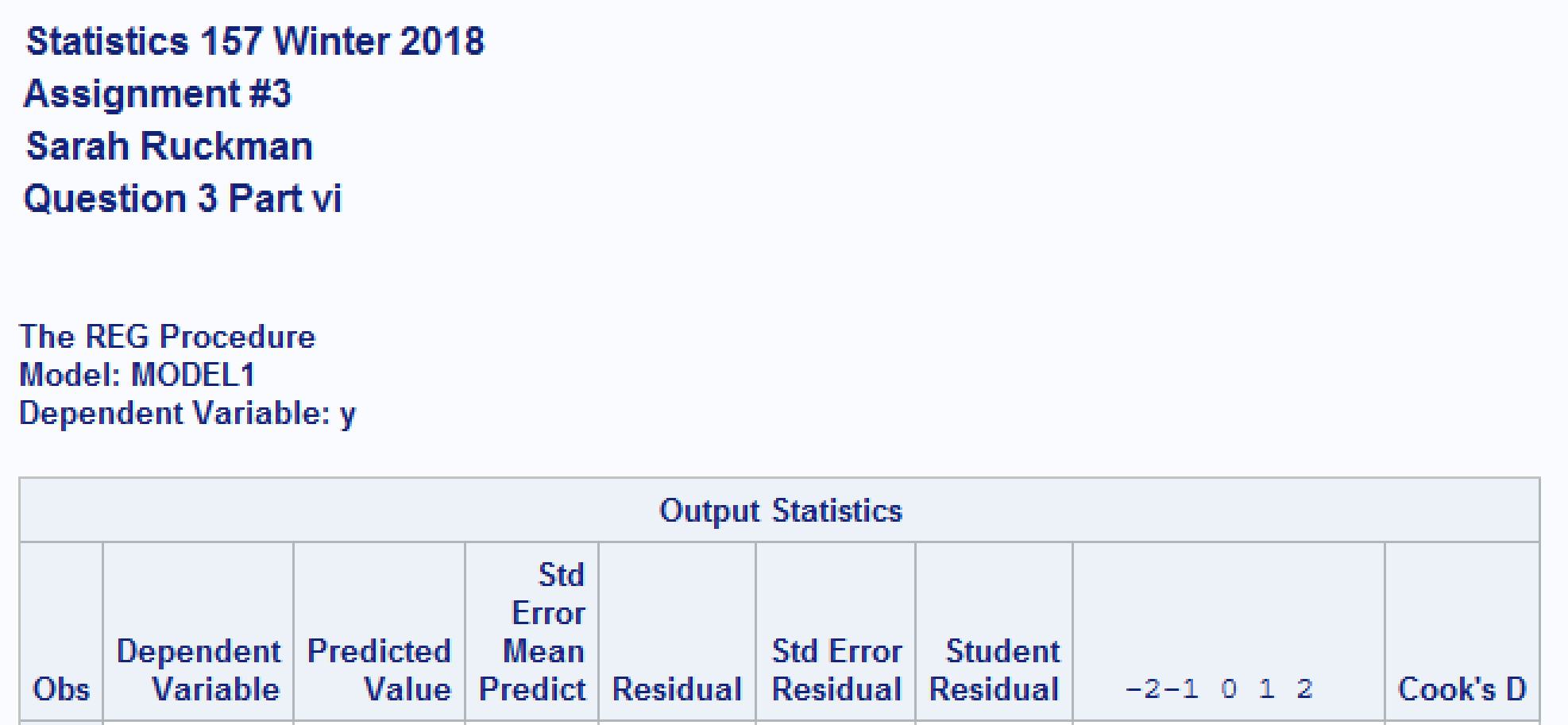
/\*Output the data to a SAS data set named q3\*/

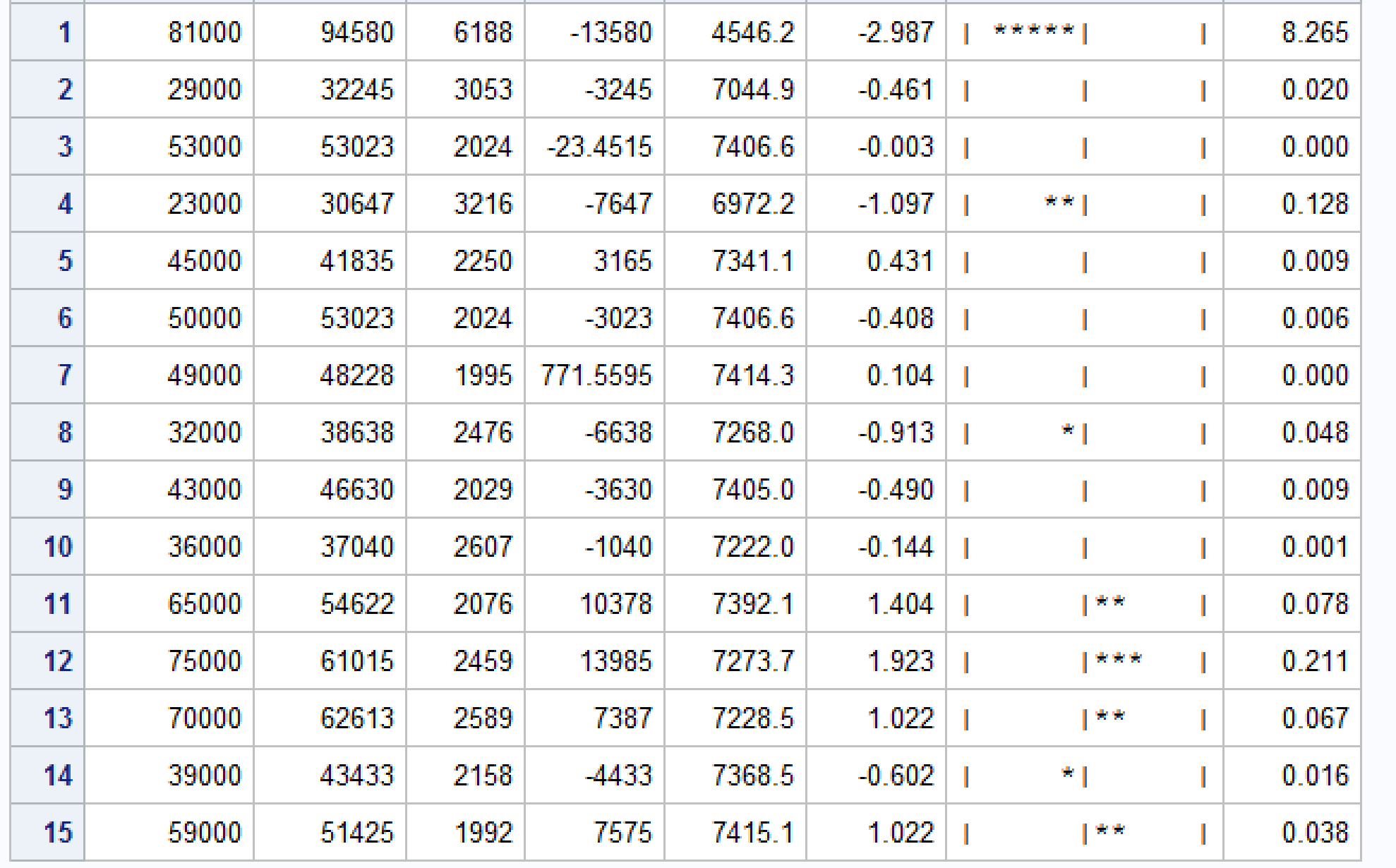
output out = q3 P = pred R = resid Student = stdres;

**run**;

**quit**;

**Output:**







**There is one suspect outlier, it is observation number 1 and has a studentized residual value of -2.987 because the residual value lies between -2 and -3.**

1. Generate a high-resolution residual plot of student (studentized) residuals versus predicted values. What conclusions, if any, can be drawn from the residual plot?

**SAS Code:**

/\*Create a residual plot using proc gplot\*/

**proc** **gplot** data = q3;

/\*Revise title 4 and 5\*/

title4 'Question 3 Part vii';

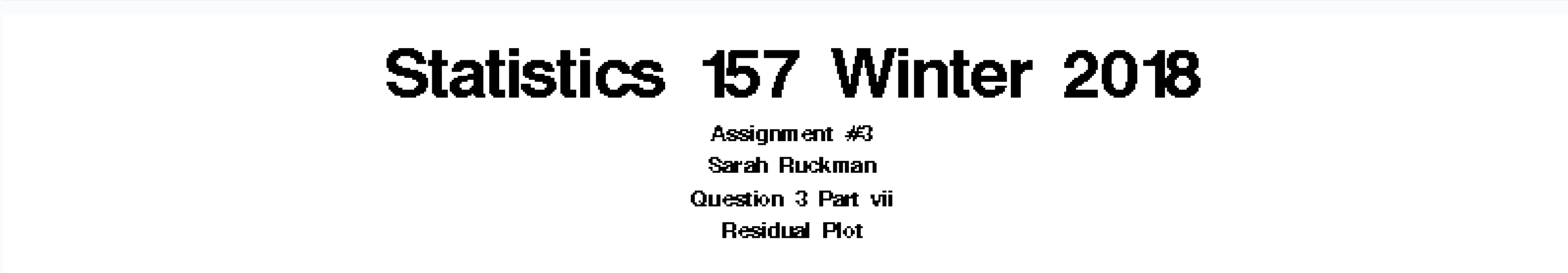
title5 'Residual Plot';

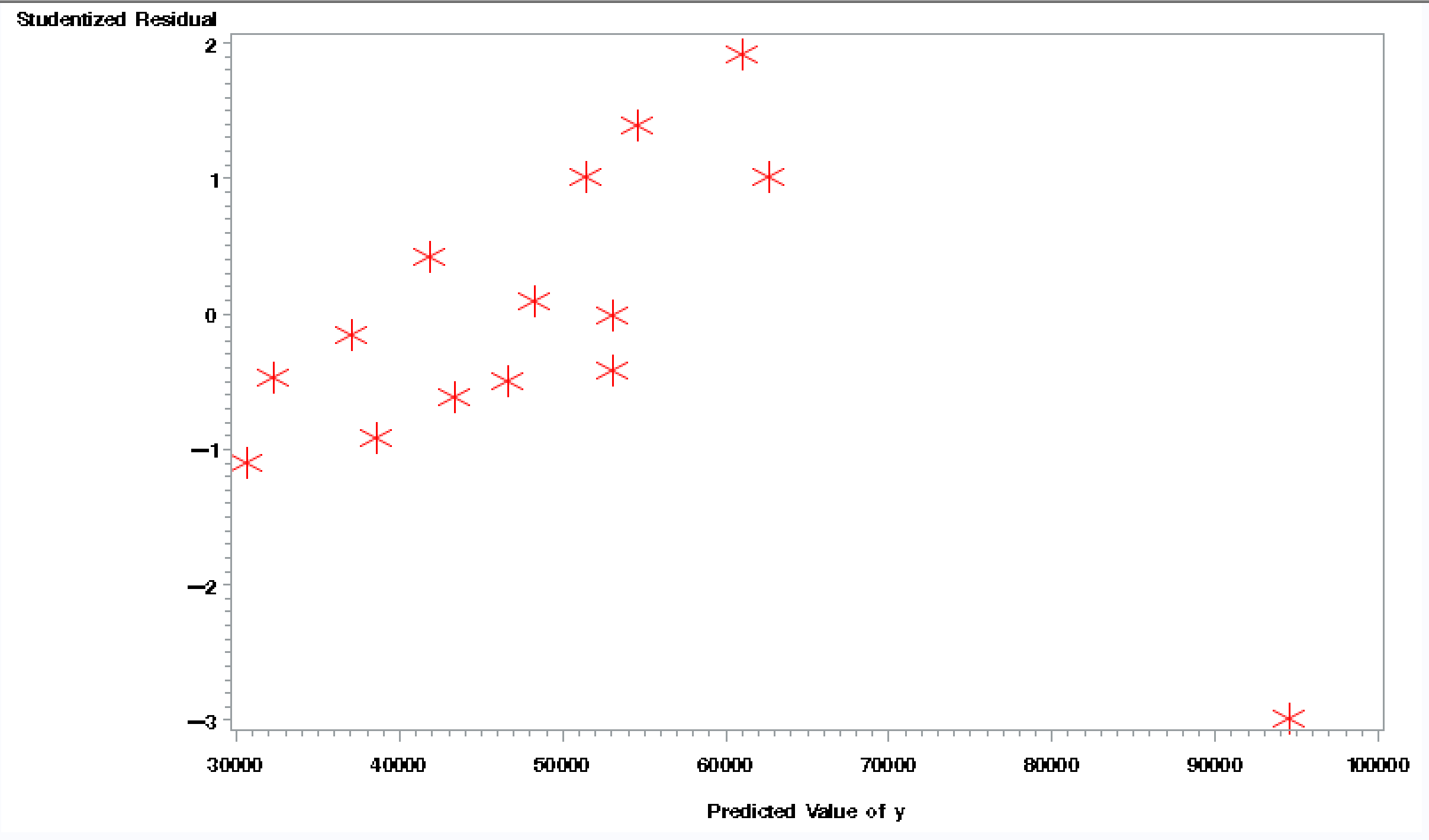
plot stdres\*pred;

**run**;

**quit**;

**Output:**





**There is one suspect outlier that is circled above because it is between -2 and -3. The plot however, seems to be linear with an upward trajectory.**

1. Refer to Question 3. Remove any outliers and complete the following. (NOTE: If you didn’t find any outliers, just state that.)

**There was a suspect outlier and the following was then done.**

1. Find and interpret the correlation between mileage and amount of repair bill. (2)

**SAS Code:**

/\*Create a new SAS temporary dataset called assign3q4\*/

**data** assign3q4;

/\*Revise title4\*/

title4 'Question 4 part i';

/\*Read in the data using an infile statement it starts on line 3, but I need to remove the first obs as it is a suspect outlier\*/

infile 'C:\Users\sarah\Downloads\REPAIR1\_w18.dat' firstobs = **4**;

/\*Input the variables x (Mileage) and y (Amount)\*/

input x y @@;

/\*Print as check\*/

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

/\*Find the correlation coefficient using proc corr

Use proc corr to generate correlation between x and y

Use nosimple to supress the printing of the descriptive stats

Use noprob to suppress printing of the stats for testing the correlation = 0\*/

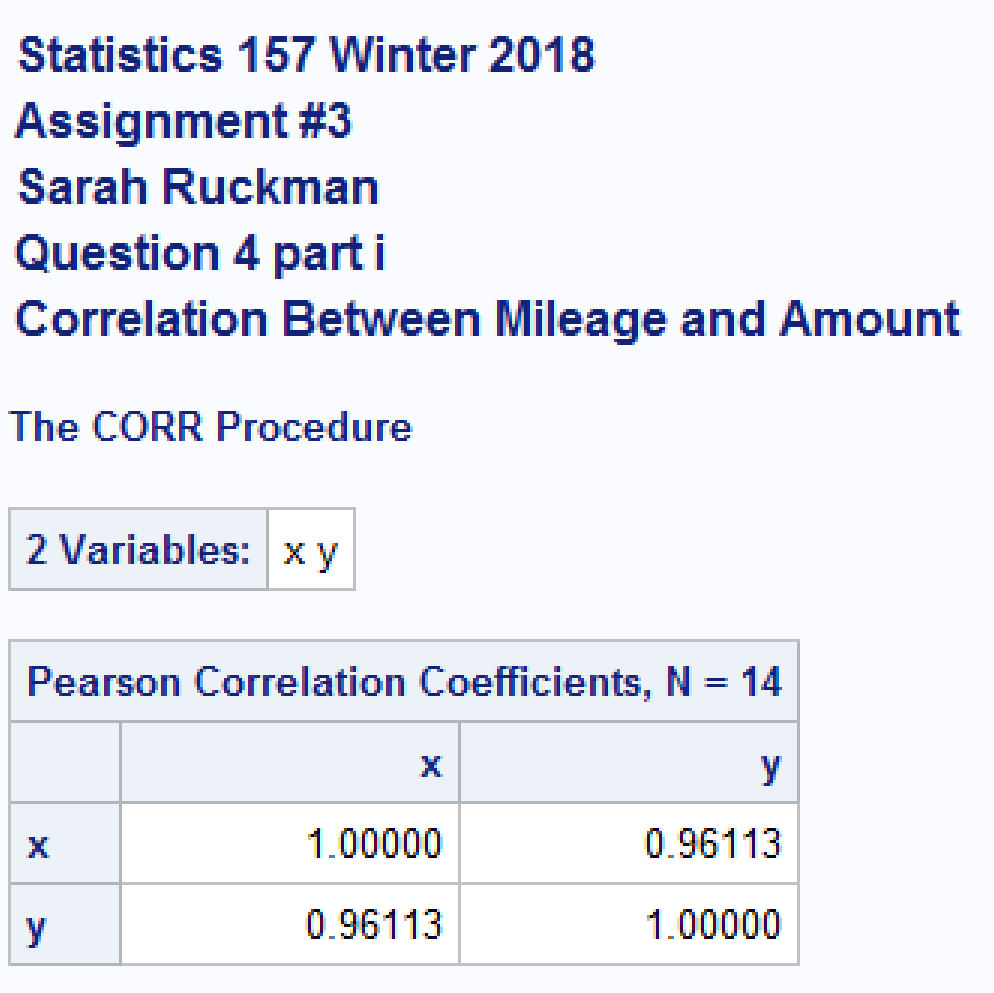
**proc** **corr** nosimple noprob;

/\*Revise title 4 and 5\*/

title5 'Correlation Between Mileage and Amount';

var x y;

**Output:**



**The correlation coefficient is 0.96113. There seems to be a strong positive linear relationship between amount and mileage because the correlation coefficient = 0.96113 is greater than 0.8.**

1. Find the estimated regression equation, ˆy = βˆ 0 + βˆ 1x. (2 pts)

**SAS Code:**

/\*Use proc reg to compute ANOVA table for linear regression

model dependent\_variable = independent\_variable\*/

**proc** **reg**;

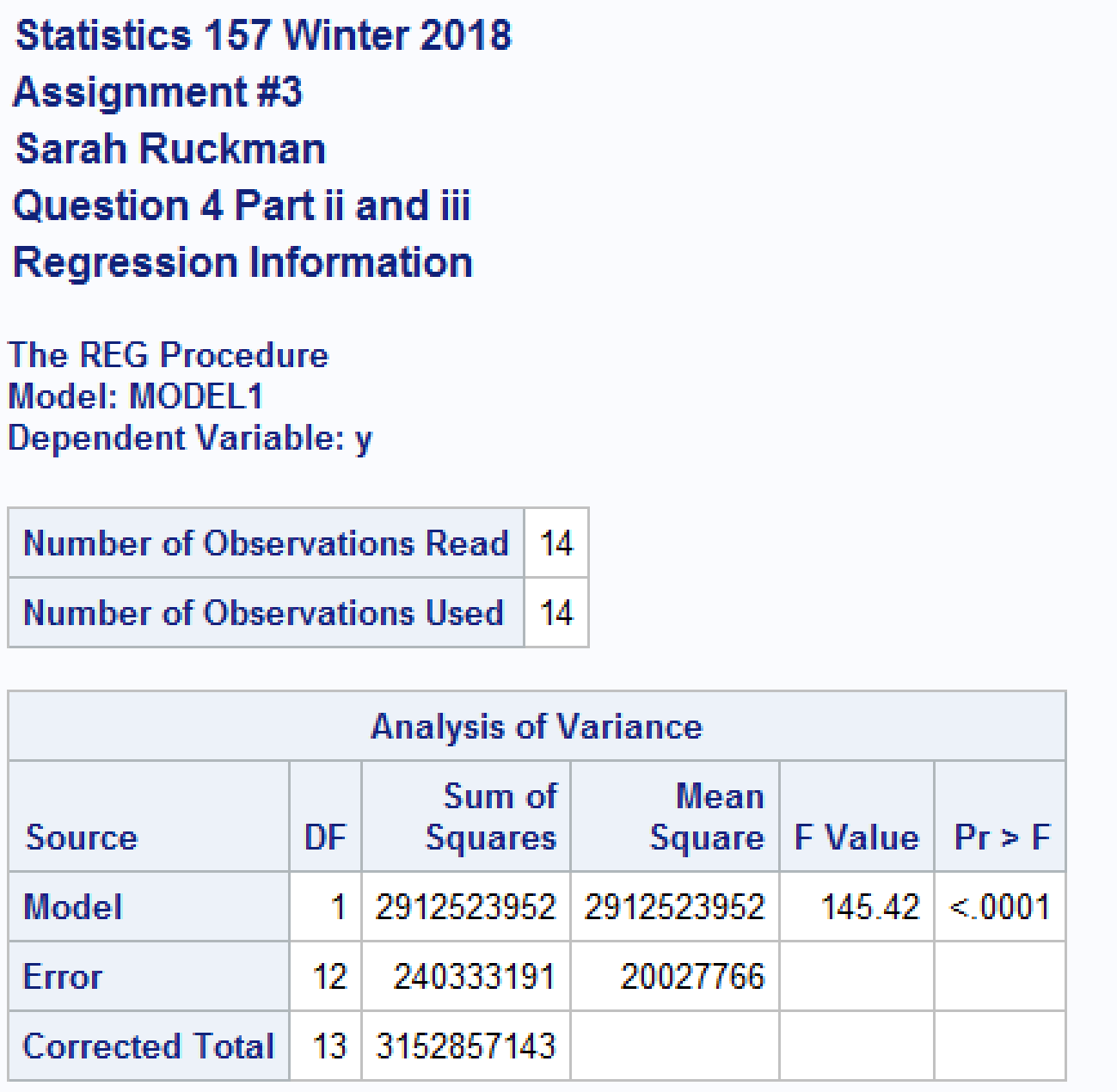
/\*Revise title 4 and 5\*/

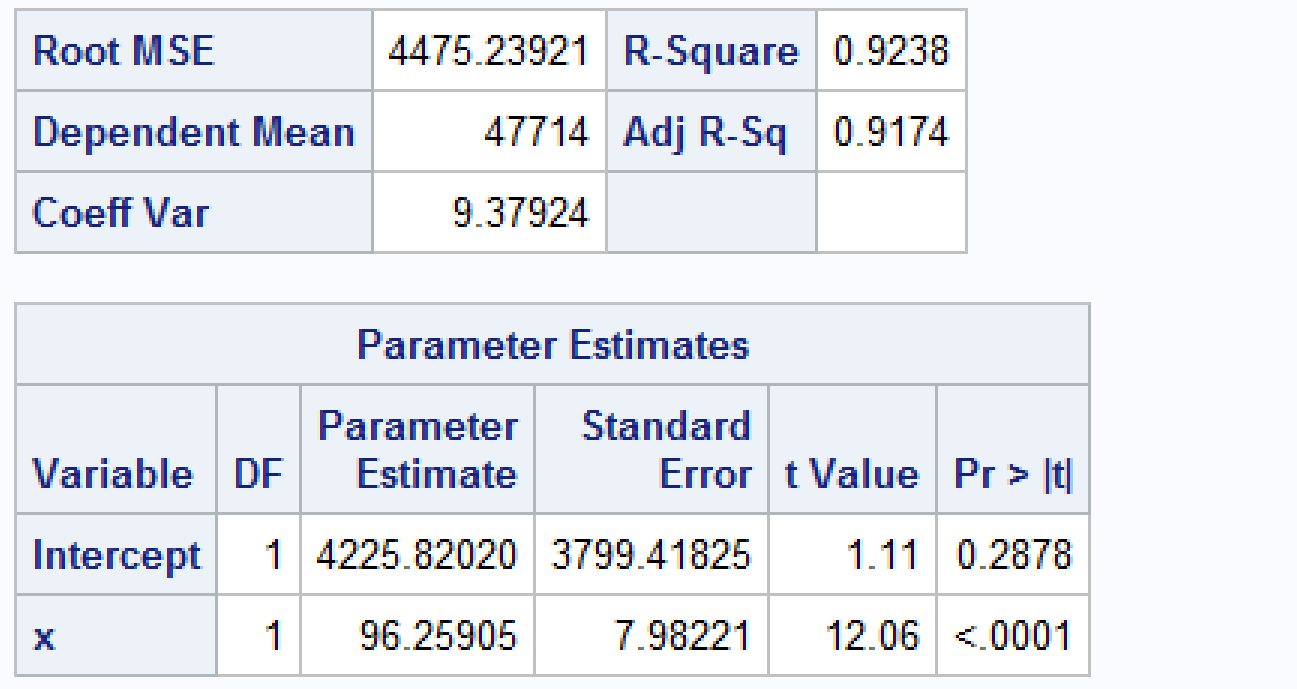
title4 'Question 4 Part ii';

title5 'Regression Information';

model y = x;

**Output:**





**The regression equation is ŷ = 4225.82020 + 96.25905x**

1. Find and interpret the coefficient of determination for the model in part (ii). (2)

**SAS Code:**

**See above**

**Output:**

**See above**

**The coefficient of determination, R2 = 0.9238. 92.38% of the variability in amount of the repair bill can be explained by the mileage on the car according to the model.**

1. Generate the predicted values, residuals and student (studentized) residuals for the model in part (ii). Do any of the data points appear to be potential outliers? If so, which one(s) and why. If not, indicate why none of the data points appear to be outliers. (1 pt)

**SAS Code:**

/\*Use proc reg to compute ANOVA table for linear regression. Use the P and R options to generate the predicted residual

values Output the data so that a residual plot may be generated\*/

**proc** **reg**;

/\*Revise title 4 and 5\*/

title4 'Question 4 Part iv';

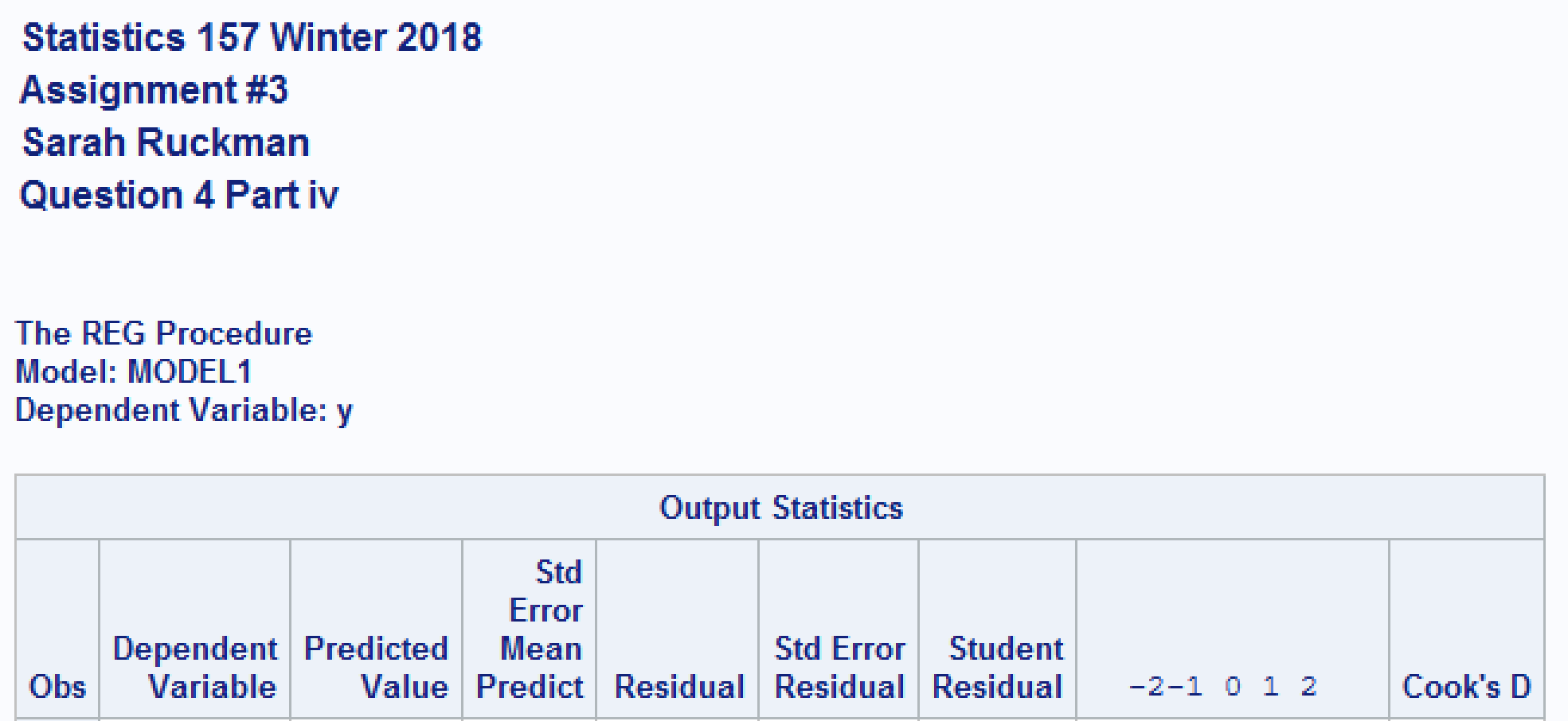
title5 ' ';

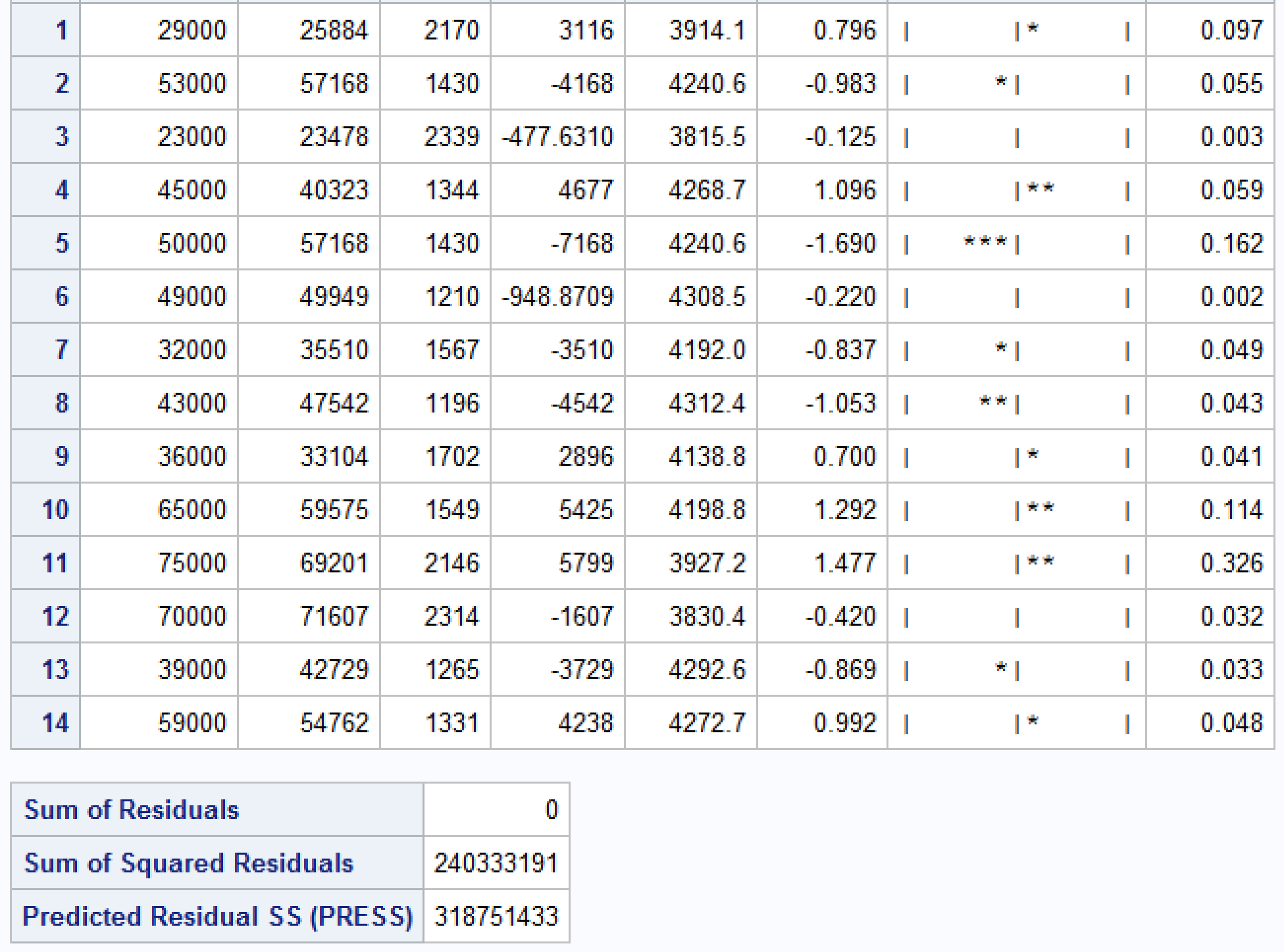
model y = x / P R;

/\*Output the data to a SAS data set named q3\*/

output out = q4 P = pred1 R = resid1 Student = stdres1;

**Output:**





**There are no outliers nor suspect outliers. All studentized residuals lie between -2 and 2 indicating that they are not outliers/suspects.**

1. Generate a high-resolution residual plot of student (studentized) residuals versus predicted values. What conclusions, if any, can be drawn from the residual plot? (1 pt)

**SAS Code:**

/\*Create a residual plot using proc gplot\*/

**proc** **gplot** data = q4;

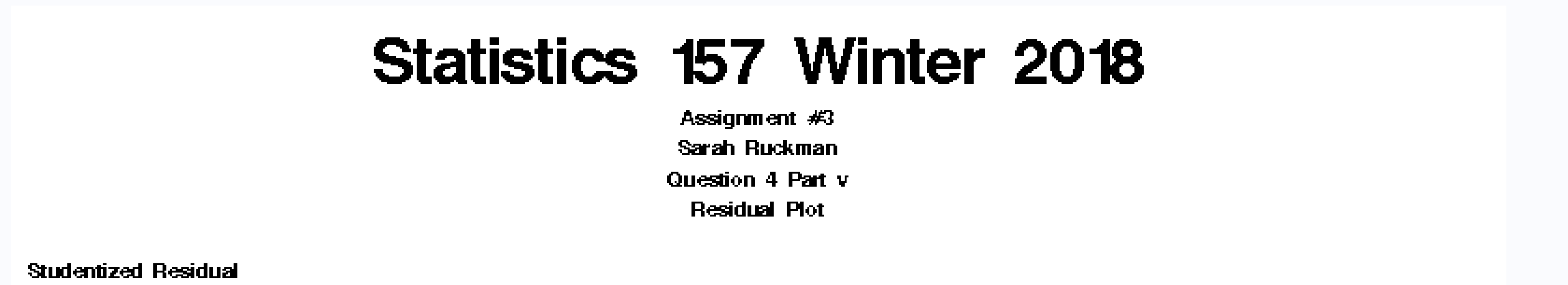
/\*Revise title 4 and 5\*/

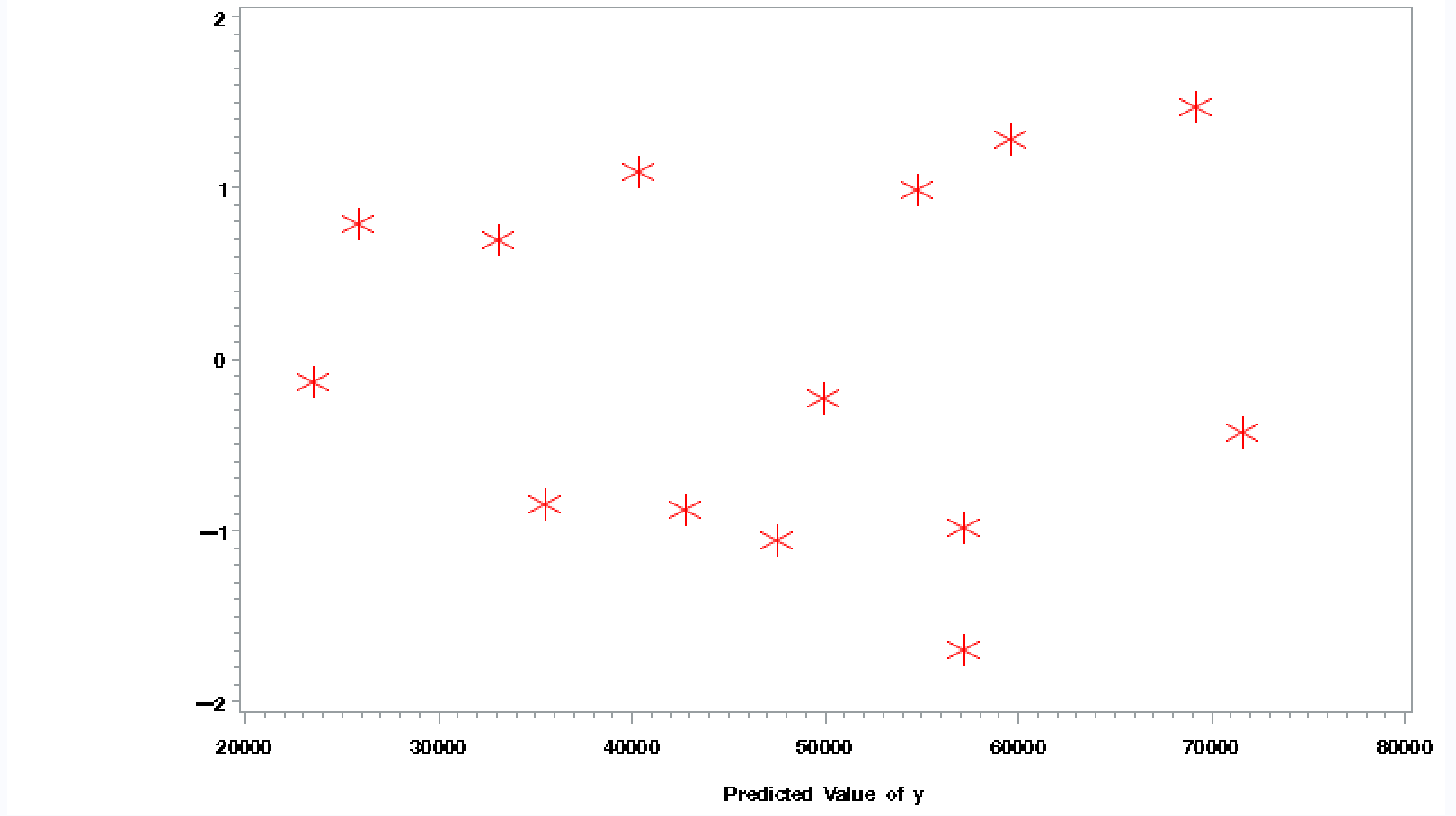
title4 'Question 4 Part v';

title5 'Residual Plot';

plot stdres1\*pred1;

**Output:**





**The residual plot seems random and does not have a pattern.**

1. Use the Shapiro-Wilks test to test normality of the student residuals. (3 pts)

**SAS Code:**

/\*Test for normality of the student residuals using proc univariate with the normal option

Use ods select TestsForNormality to only print the normality test information\*/

**proc** **univariate** data = q4 normal;

ods select TestsForNormality;

var stdres1;

/\*Revise title 4 and 5\*/

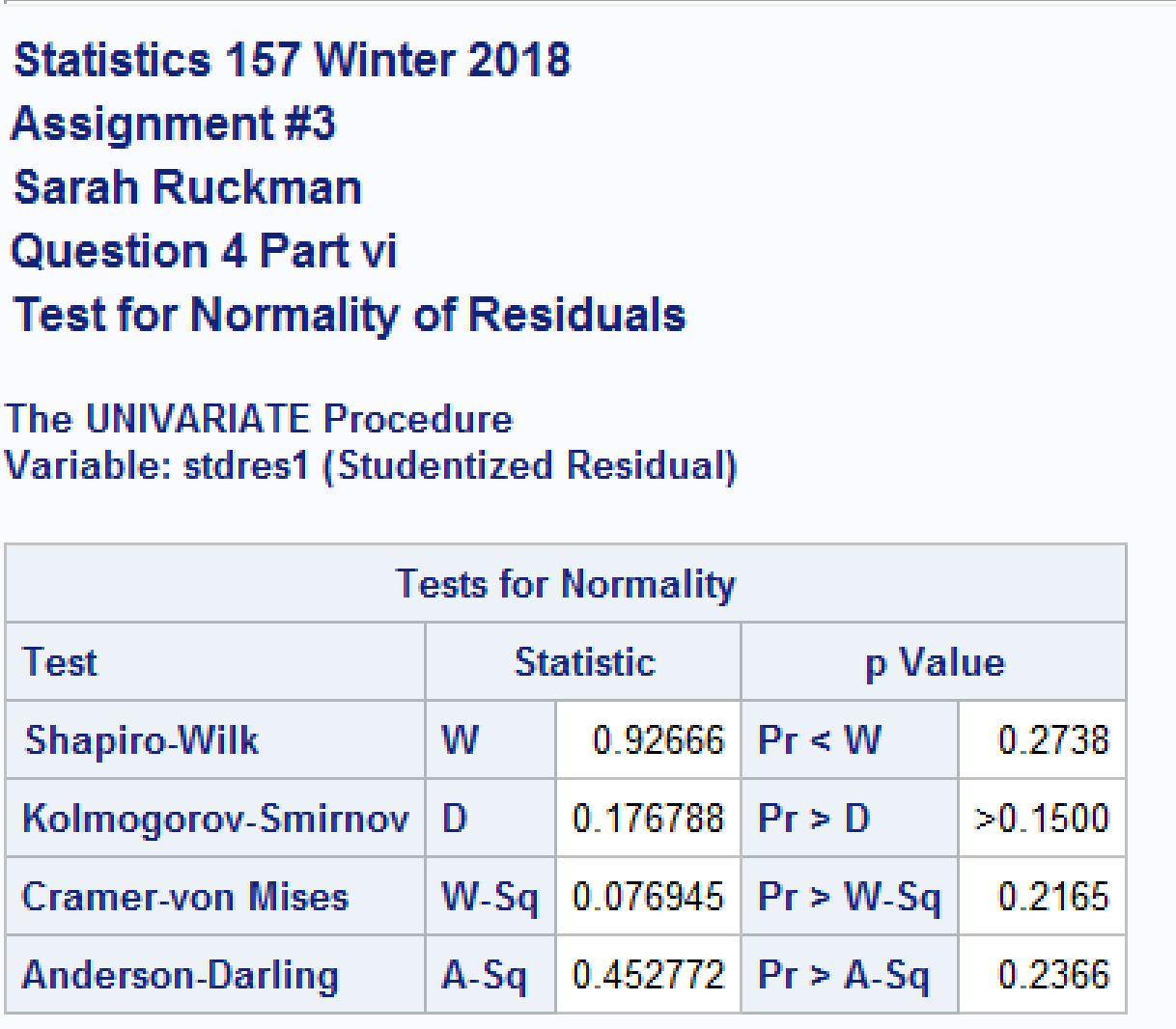
title4 'Question 4 Part vi';

title5 'Test for Normality of Residuals';

**run**;

**quit**;

**Output:**



**H0: Errors are normally distributed.**

**Ha: Errors are not normally distributed.**

**TS = Shapiro-Wilks = 0.92666 with p-value = 0.2738**

**RR: Reject H0 if p-value < α = 0.05**

**Conclusion:**

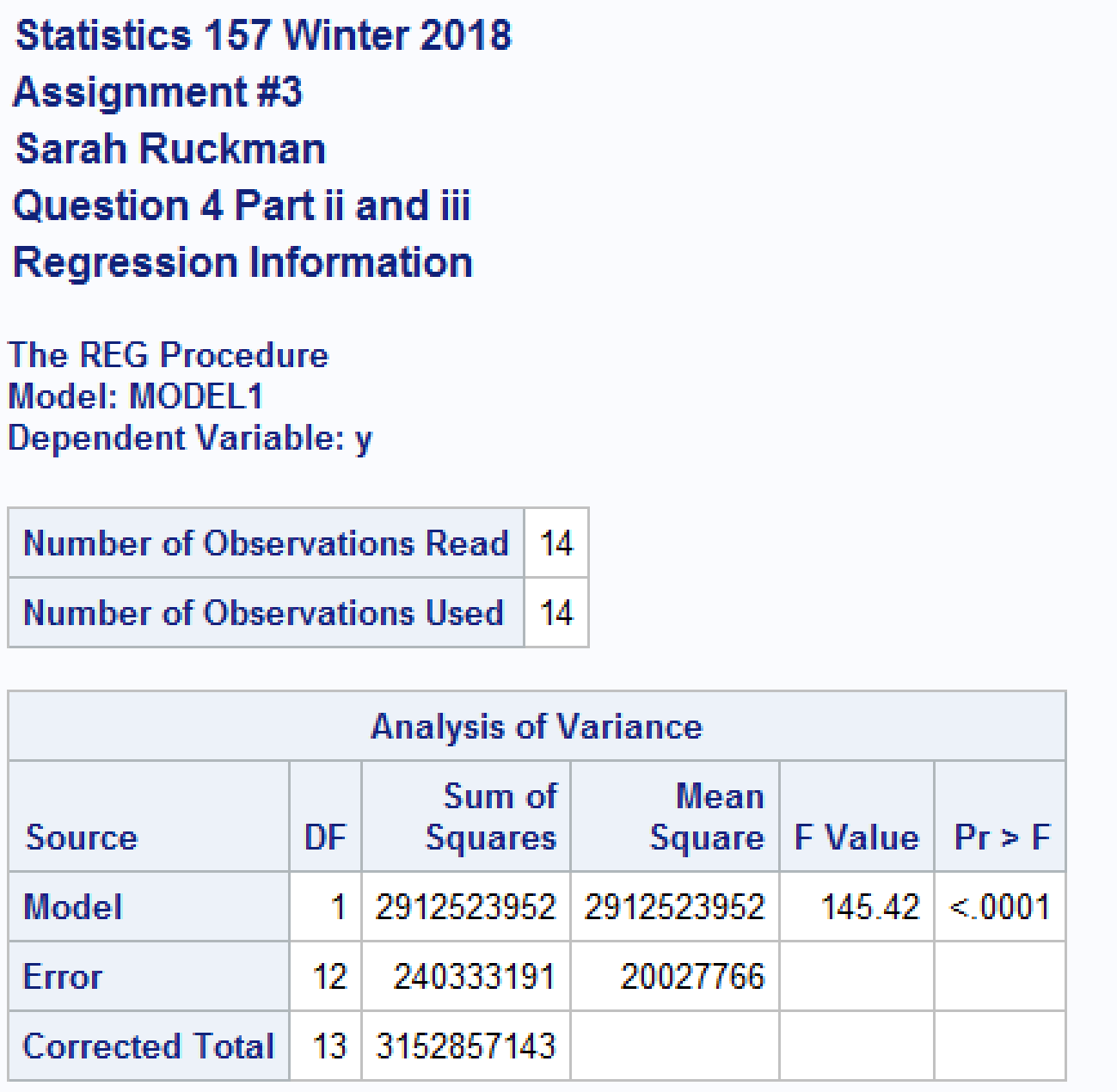
**Since the p-value = 0.2738 is greater than α = 0.05, We do not reject H0 and it is reasonable to assume the errors are normally distributed.**

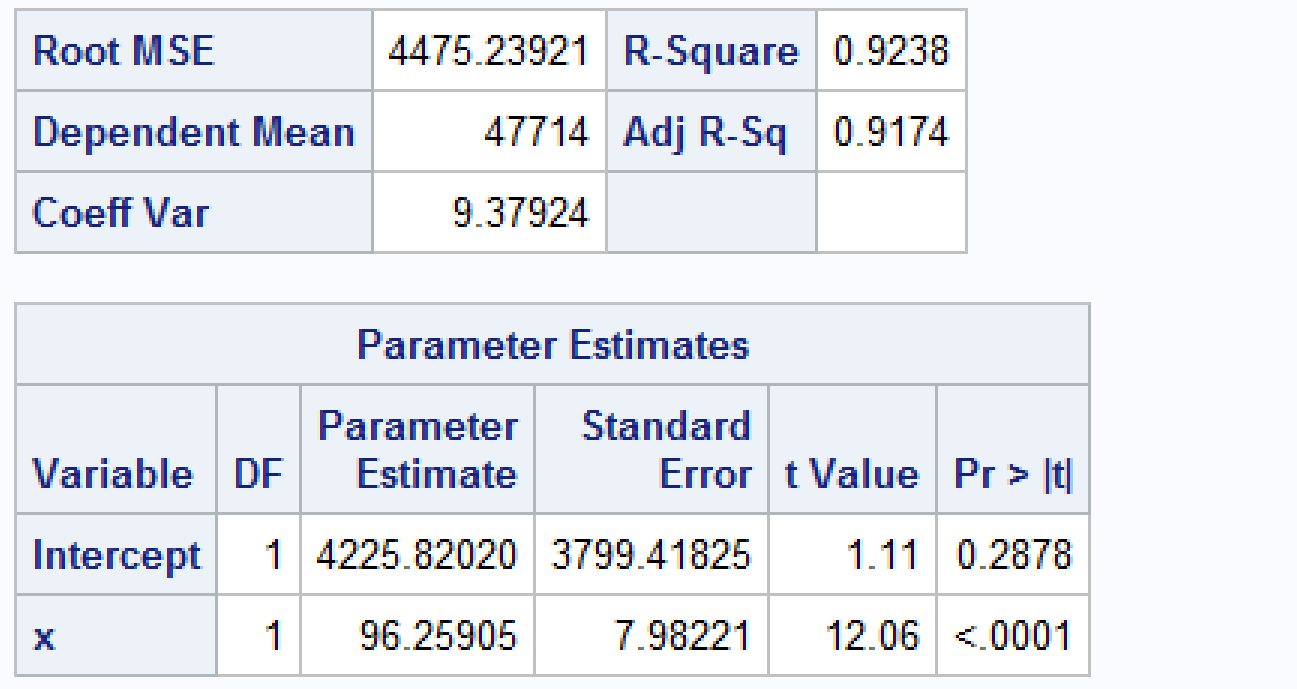
1. Test β1 = 0. (2 pts)

**SAS Code:**

**Please see question 4 part ii**

**Output:**





**H0: β1 = 0**

**Ha: β1 ≠ 0**

**p-value = <0.0001**

**RR: Reject H0­ if p-value < α = 0.05**

**Conclusion:**

**Since the p-value = <0.0001 is less than α = 0.05, We reject H0 and it is reasonable to assume the slope (β1) is significantly different from 0.**

**Complete SAS Code:**

/\*Set up options and turn off extra graphics\*/

options nocenter nodate nonumber ps=**55** ls=**78**;

ods graphics off;

/\*goptions formats the plot

cback color of the plot background

colors colors to use

ftitle font of plot title

htitle height of the title

htext height of the text on the plot \*/

goptions reset = all colors=(blue,red,green,purple) ftitle = swissb ftext=swissb htitle=**3**;

/\*Create new SAS temporary dataset\*/

**data** mat1;

/\*Set up titles\*/

title1 'Statistics 157 Winter 2018';

title2 'Assignment #3';

title3 'Sarah Ruckman';

title4 'Question 1 Part i';

/\*Invoke proc iml to complete matrix calculations\*/

**proc** **iml**;

/\*initialize matricies\*/

A = {**3** -**2** **1**, **2** **1** -**1**, **1** -**4** **3**};

B = {**4.5**, **1.5**, **8.5**};

/\*Use trace(A) to find the trace of matrix A\*/

TraceA = trace(A);

/\*Print the results\*/

print ,,, A, B, 'Trace of A',, TraceA;

/\*Find the product of A and B by creating a new variable PROD1\*/

PROD1 = A\*B;

/\*Revise title4\*/

title4 'Question 1 Part ii';

/\*Print the results\*/

print ,,,'Product of A and B',, PROD1;

/\*Find the determinat of A\*/

/\*Revise title4\*/

title4 'Question 1 Part iii';

DET1 = det(A);

/\*Print the results\*/

print ,,, 'Determinat of A', DET1;

/\*Find the inverse of matrix A\*/

/\*Revise title4\*/

title4 'Question 1 Part iv';

/\*Use do loops with if then else to get the answer\*/

if DET1 = **0** then

do;

print ,,, 'Since the determinant = 0, the matrix A is singular and does not have an inverse';

end;

else

do;

AINV = inv(A);

/\*Check Product\*/

PROD2 = A\*AINV;

/\*Print the results\*/

print ,,, A, 'Inverse of A', AINV, 'Product check', PROD2;

end;

/\*Find the solution to the system of equations using AINV\*B\*/

/\*Revise title4\*/

title4 'Question 1 part v';

SOLN = AINV\*B;

/\*Print the results\*/

print ,,, 'Solution is' SOLN;

/\*Create new SAS temporary dataset\*/

**data** q2;

/\*Revise title4\*/

title4 'Question 2 part i';

/\*Invoke proc iml to complete matrix calculations\*/

**proc** **iml**;

/\*Create matricies\*/

A = {**4** **2**, **2** **1**};

B = {**21**, **12**};

/\*Use det(A) to find the determinant of A\*/

DETA = det(A);

/\*Print the results\*/

print ,,,A, 'Determinant of A', DETA;

/\*Find the inverse of matrix A\*/

/\*Revise title4\*/

title4 'Question 2 Part ii';

/\*Use do loops with if then else to get the answer\*/

if DETA = **0** then

do;

print ,,, 'Since the determinant = 0, the matrix A is singular and does not have an inverse';

end;

else

do;

AINV = inv(A);

/\*Check Product\*/

PROD2 = A\*AINV;

/\*Print the results\*/

print ,,, A, 'Inverse of A', AINV, 'Product check', PROD2;

end;

/\*Find the solution to the system of equations using AINV\*B\*/

/\*Revise title4\*/

title4 'Question 2 part iii';

/\*Use do loops with if then else to get the answer\*/

if DETA = **0** then

do;

print,'Since the determinant = 0, the matrix A is singular and does not have an inverse nor a solution to the system of equations';

end;

else

do;

INVA = inv(A);

SOLN1 = INVA\*B;

/\*Print the results\*/

print ,,, 'Inverse of A', AINV, 'The solution is' SOLN1;

end;

/\*Create a new SAS temporary dataset called assign3q3\*/

**data** assign3q3;

/\*Revise title4\*/

title4 'Question 3 part i';

/\*Read in the data using an infile statement it starts on line 3\*/

infile 'C:\Users\sarah\Downloads\REPAIR1\_w18.dat' firstobs = **3**;

/\*Input the variables x (Mileage) and y (Amount)\*/

input x y @@;

/\*Print as check\*/

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

/\*Create a high resolution plot of y versus x\*/

/\*Revise title4\*/

title4 'Question 3 Part ii';

/\*Use a symbol statement to set up the format of the plt symbols

value symbol of the data points

height height of the symbol of the data points

cv color of the symbol\*/

symbol1 value = star height = **3** cv = red;

/\*Use proc gplot to generate the high resolution plot

plot vertical(y) vs horizontal(x)

caxis color of the axes

ctext color of the text on the plot\*/

**proc** **gplot**;

title5 'Scatterplot of Amount vs. Mileage';

plot y\*x / caxis = darkred ctext=black;

/\*Find the correlation coefficient using proc corr

Use proc corr to generate correlation between x and y

Use nosimple to supress the printing of the descriptive stats

Use noprob to suppress printing of the stats for testing the correlation = 0\*/

**proc** **corr** nosimple noprob;

/\*Revise title 4 and 5\*/

title4 'Question 3 Part iii';

title5 'Correlation Between Mileage and Amount';

var x y;

/\*Use proc reg to compute ANOVA table for linear regression

model dependent\_variable = independent\_variable\*/

**proc** **reg**;

/\*Revise title 4 and 5\*/

title4 'Question 3 Part iv';

title5 'Regression Information';

model y = x;

/\*Use proc reg to compute ANOVA table for linear regression. Use the P and R options to generate the predicted residual

values Output the data so that a residual plot may be generated\*/

**proc** **reg**;

/\*Revise title 4 and 5\*/

title4 'Question 3 Part vi';

title5 ' ';

model y = x / P R;

/\*Output the data to a SAS data set named q3\*/

output out = q3 P = pred R = resid Student = stdres;

/\*Create a residual plot using proc gplot\*/

**proc** **gplot** data = q3;

/\*Revise title 4 and 5\*/

title4 'Question 3 Part vii';

title5 'Residual Plot';

plot stdres\*pred;

/\*Create a new SAS temporary dataset called assign3q4\*/

**data** assign3q4;

/\*Revise title4\*/

title4 'Question 4 part i';

/\*Read in the data using an infile statement it starts on line 3, but I need to remove the first obs as it is a suspect outlier\*/

infile 'C:\Users\sarah\Downloads\REPAIR1\_w18.dat' firstobs = **4**;

/\*Input the variables x (Mileage) and y (Amount)\*/

input x y @@;

/\*Print as check\*/

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

/\*Find the correlation coefficient using proc corr

Use proc corr to generate correlation between x and y

Use nosimple to supress the printing of the descriptive stats

Use noprob to suppress printing of the stats for testing the correlation = 0\*/

**proc** **corr** nosimple noprob;

/\*Revise title 4 and 5\*/

title5 'Correlation Between Mileage and Amount';

var x y;

/\*Use proc reg to compute ANOVA table for linear regression

model dependent\_variable = independent\_variable\*/

**proc** **reg**;

/\*Revise title 4 and 5\*/

title4 'Question 4 Part ii and iii';

title5 'Regression Information';

model y = x;

/\*Use proc reg to compute ANOVA table for linear regression. Use the P and R options to generate the predicted residual

values Output the data so that a residual plot may be generated\*/

**proc** **reg**;

/\*Revise title 4 and 5\*/

title4 'Question 4 Part iv';

title5 ' ';

model y = x / P R;

/\*Output the data to a SAS data set named q3\*/

output out = q4 P = pred1 R = resid1 Student = stdres1;

/\*Create a residual plot using proc gplot\*/

**proc** **gplot** data = q4;

/\*Revise title 4 and 5\*/

title4 'Question 4 Part v';

title5 'Residual Plot';

plot stdres1\*pred1;

/\*Test for normality of the student residuals using proc univariate with the normal option

Use ods select TestsForNormality to only print the normality test information\*/

**proc** **univariate** data = q4 normal;

ods select TestsForNormality;

var stdres1;

/\*Revise title 4 and 5\*/

title4 'Question 4 Part vi';

title5 'Test for Normality of Residuals';

**run**;

**quit**;