Linear Regression Using Simulated Data in R and SAS

jdt

11/26/2020

Contents

Preface

This document presents an example for generating simulated data for linear regression using R and SAS.

Assignment

Run this example using both R and SAS. If you have difficulty with either, let me know and get some help. Being able to run R will not cut it here!

Change the value for the parameters, n, β_1 and β_0 and observe the effect these changes have on the results. Pay particular attention to the effect on the R-square value.

R

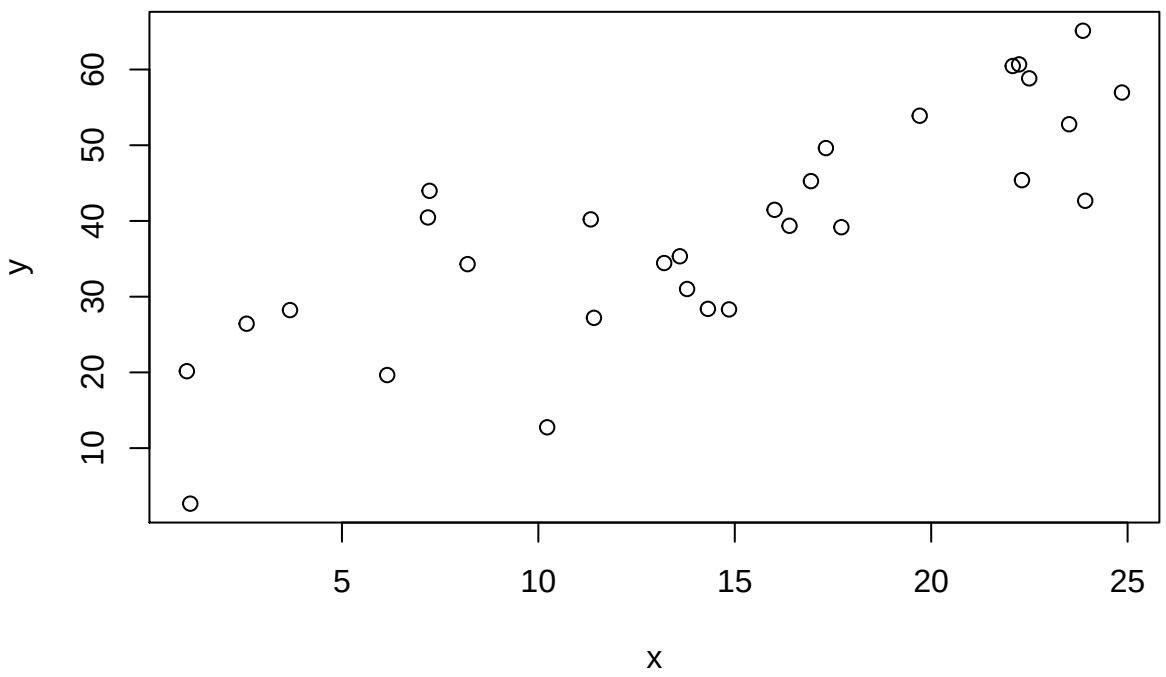
Generate the Simulated Data

```
set.seed(123)
n=30          #sample size
beta_0=10     #true y-intercept
beta_1=2     #true slope
sigma= 9      #true sigma

x=25*runif(n)
y=beta_0 + beta_1*x + sigma*rnorm(n)
```

Graph of data

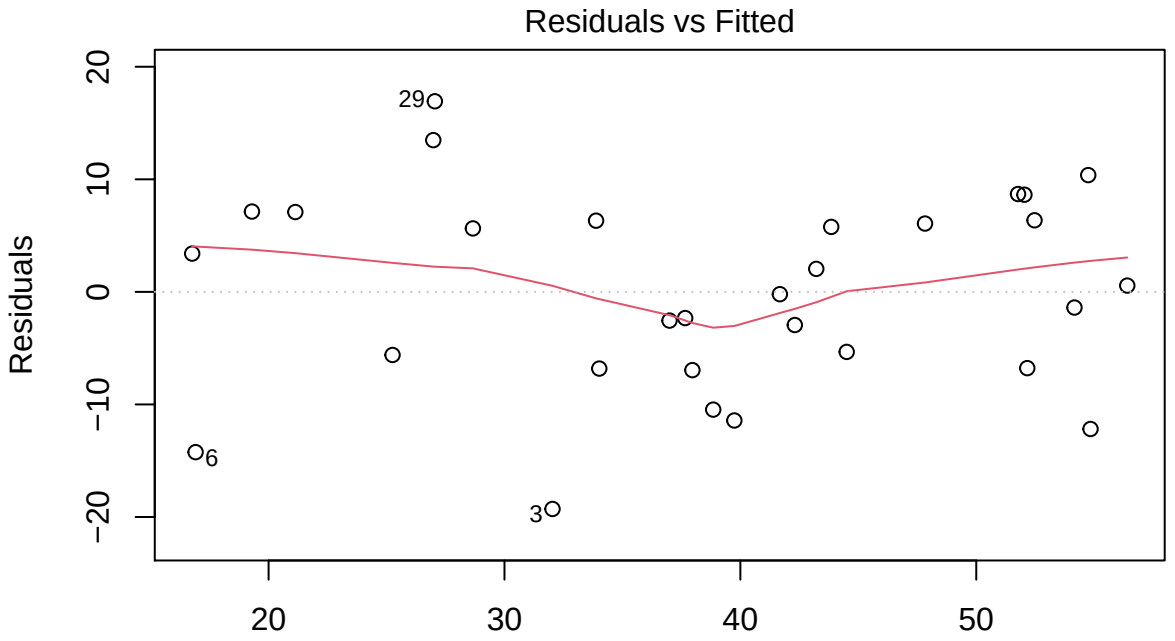
```
plot(y~x)
```

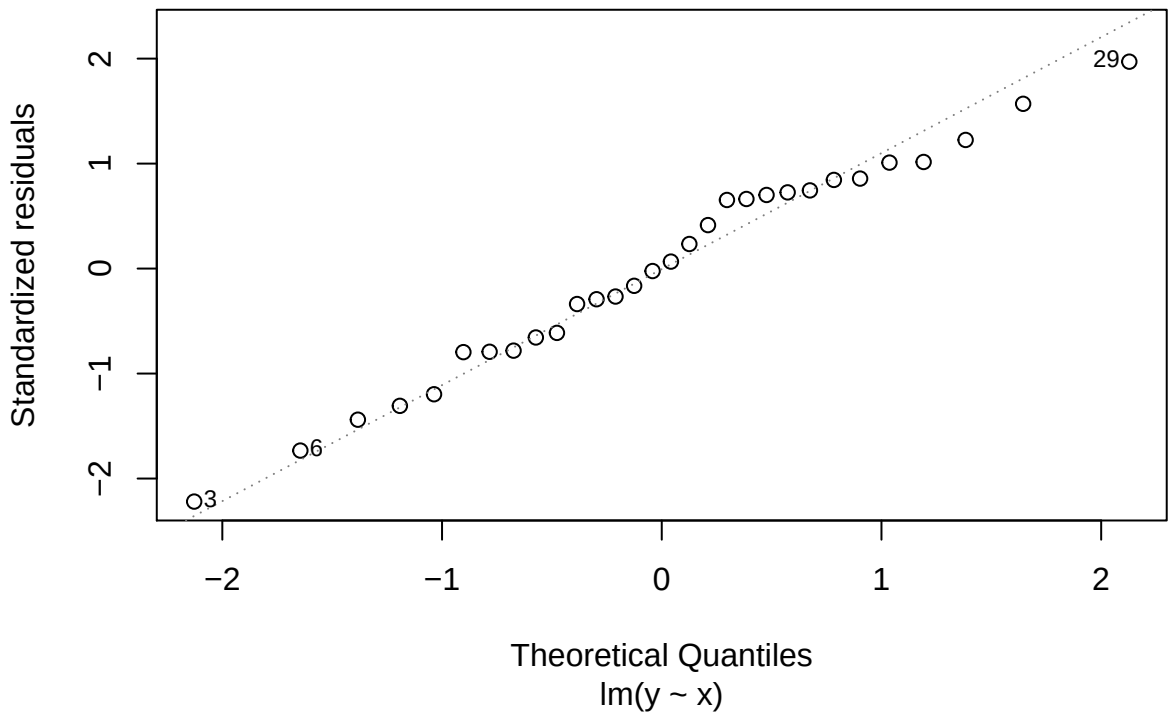


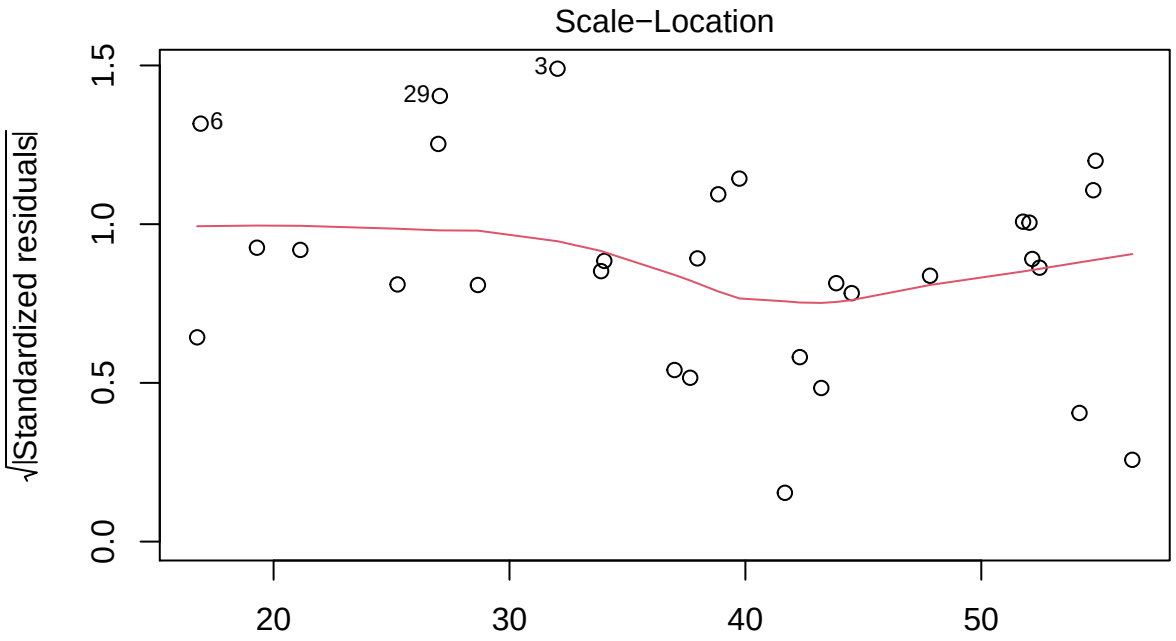

```
result<-lm(y~x)
summary(result)
```

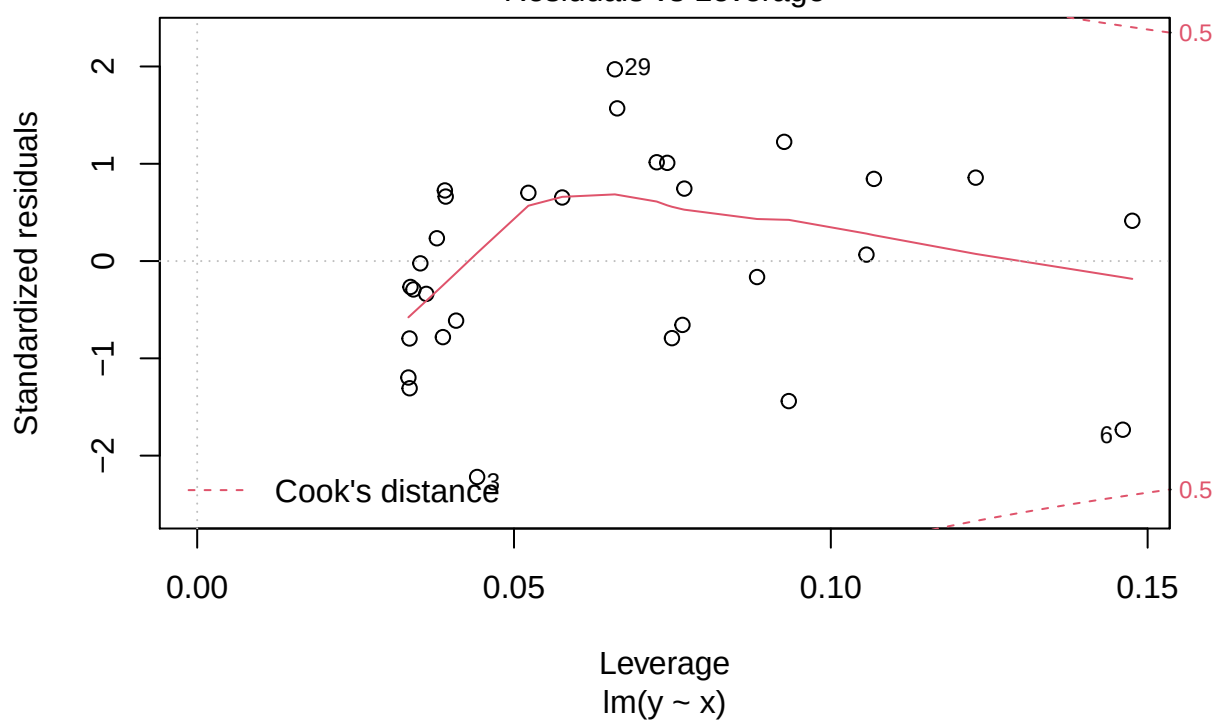
```
##
## Call:
## lm(formula = y ~ x)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -19.2864  -6.4826   0.1758   6.3506  16.9309
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  15.0097     3.6258   4.140 0.000289 ***
## x             1.6652     0.2266   7.349  5.3e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.888 on 28 degrees of freedom
## Multiple R-squared:  0.6586, Adjusted R-squared:  0.6464
## F-statistic: 54.01 on 1 and 28 DF,  p-value: 5.296e-08
```

```
plot(result)
```


Residuals vs Fitted
Residuals
Fitted values
lm(y ~ x)
29
6
3
20
10
0
−10
−20
20
30
40
50


Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(y ~ x)
29
6
3
2
1
0
−1
−2
−2
−1
0
1
2





SAS

SAS Code

```
title1 'Simulated Linear Regression';
```

```
/*************************************
 Simple Linear Regression Models
 *************************************/

%let N = 30;                              /* size of each sample      */
%let beta_0 = 10;                         /* true y-intercept         */
%let beta_1 = 2;                          /* true slope               */
%let sigma=9;                             /* true sigma               */
data Reg1(keep=x y);
call streaminit(1);
do i = 1 to &N;
   x = 10*rand("Uniform");                /* explanatory variable     */
   eps = rand("Normal", 0, &sigma);       /* error term N(0,sigma)    */
   y = &beta_0 + &beta_1*x + eps;
   output;
end;
run;

data reg_out; set Reg1;

proc sgplot data=reg_out;
scatter y=y x=x;
reg y=y x=x;
run;

proc reg data=Reg1 plots=FITPLOT;
   model y = x;
*   ods exclude NObs;
   run;
quit;
```

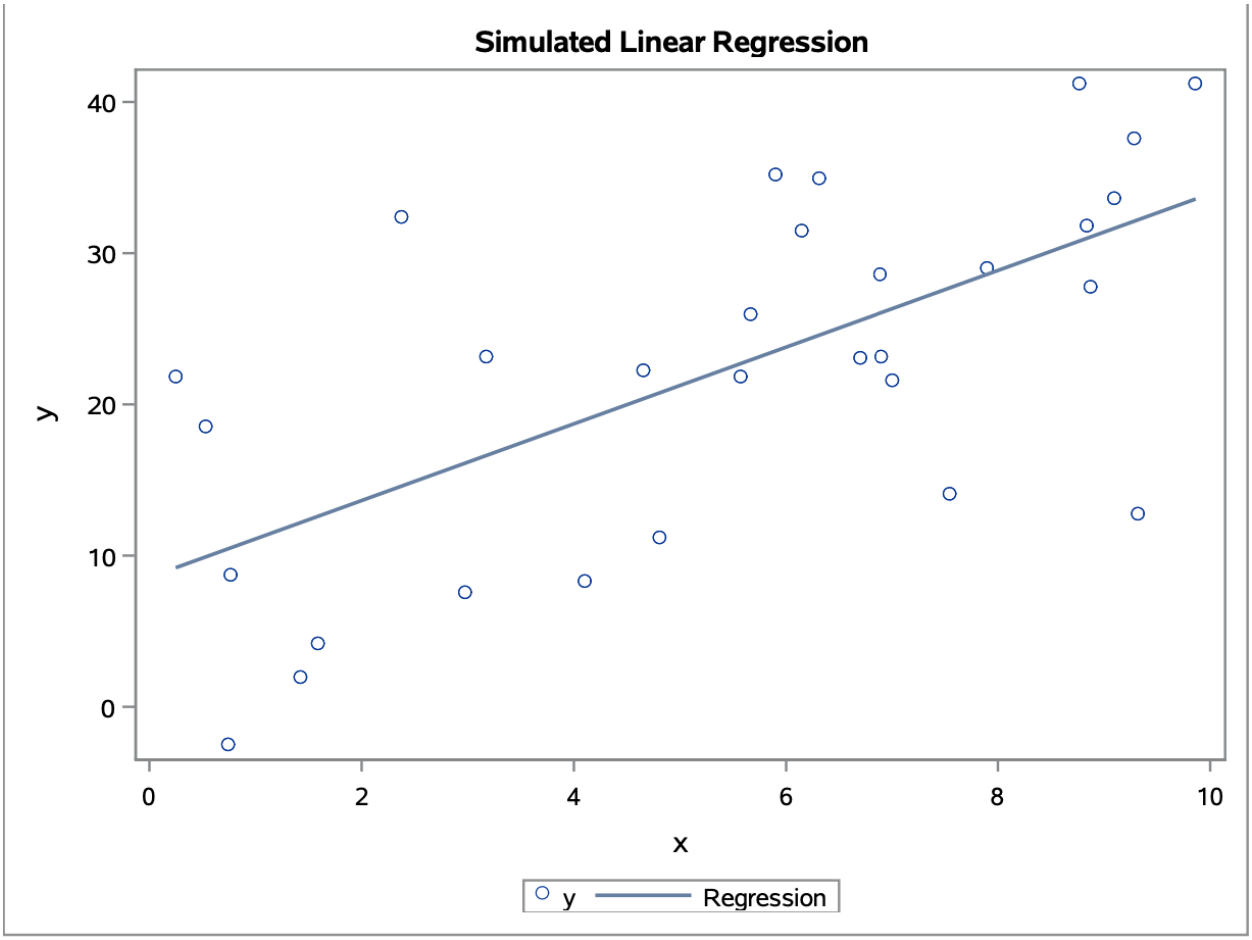
SAS Output

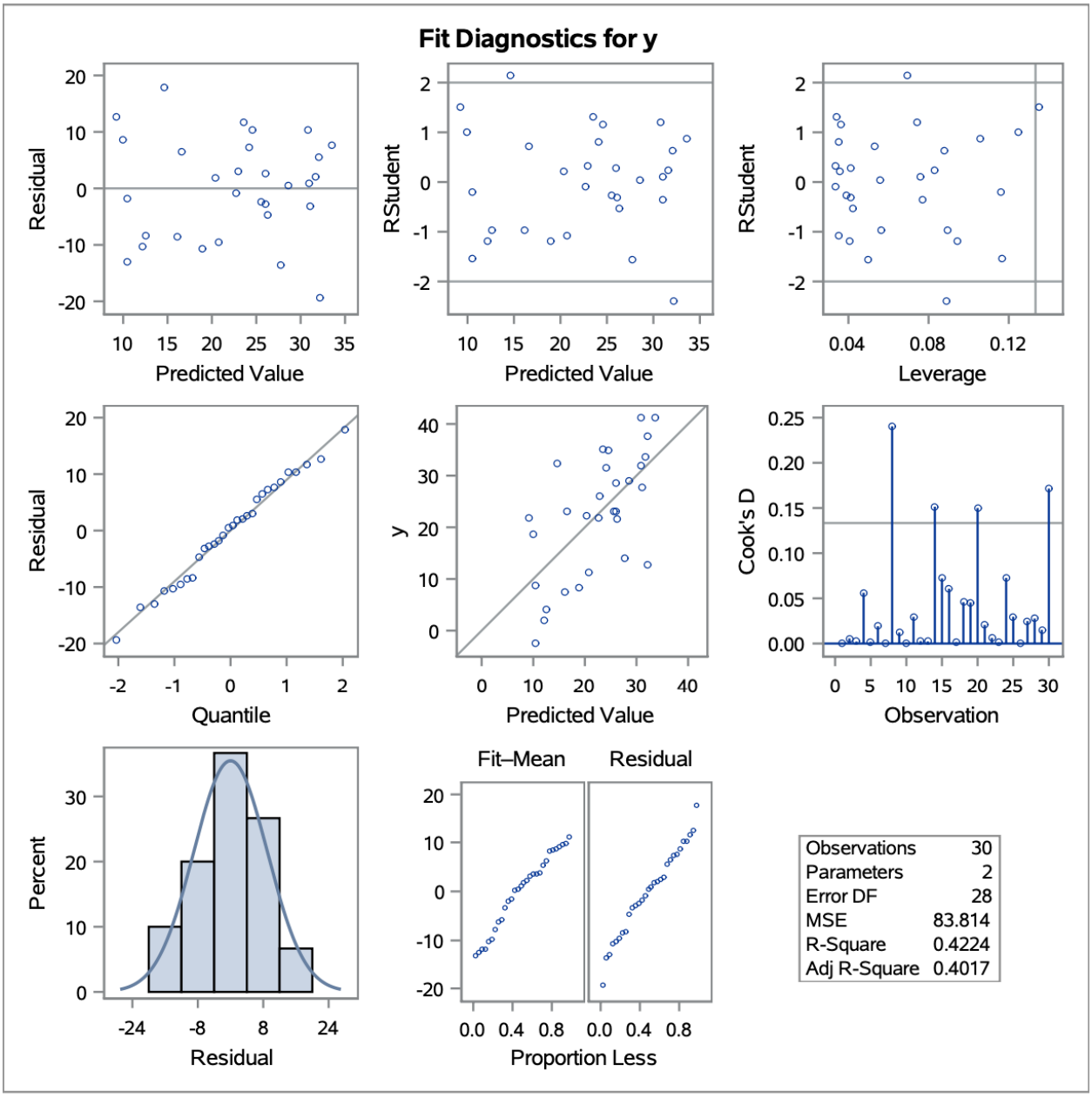
Number of Observations Read	50
Number of Observations Used	50

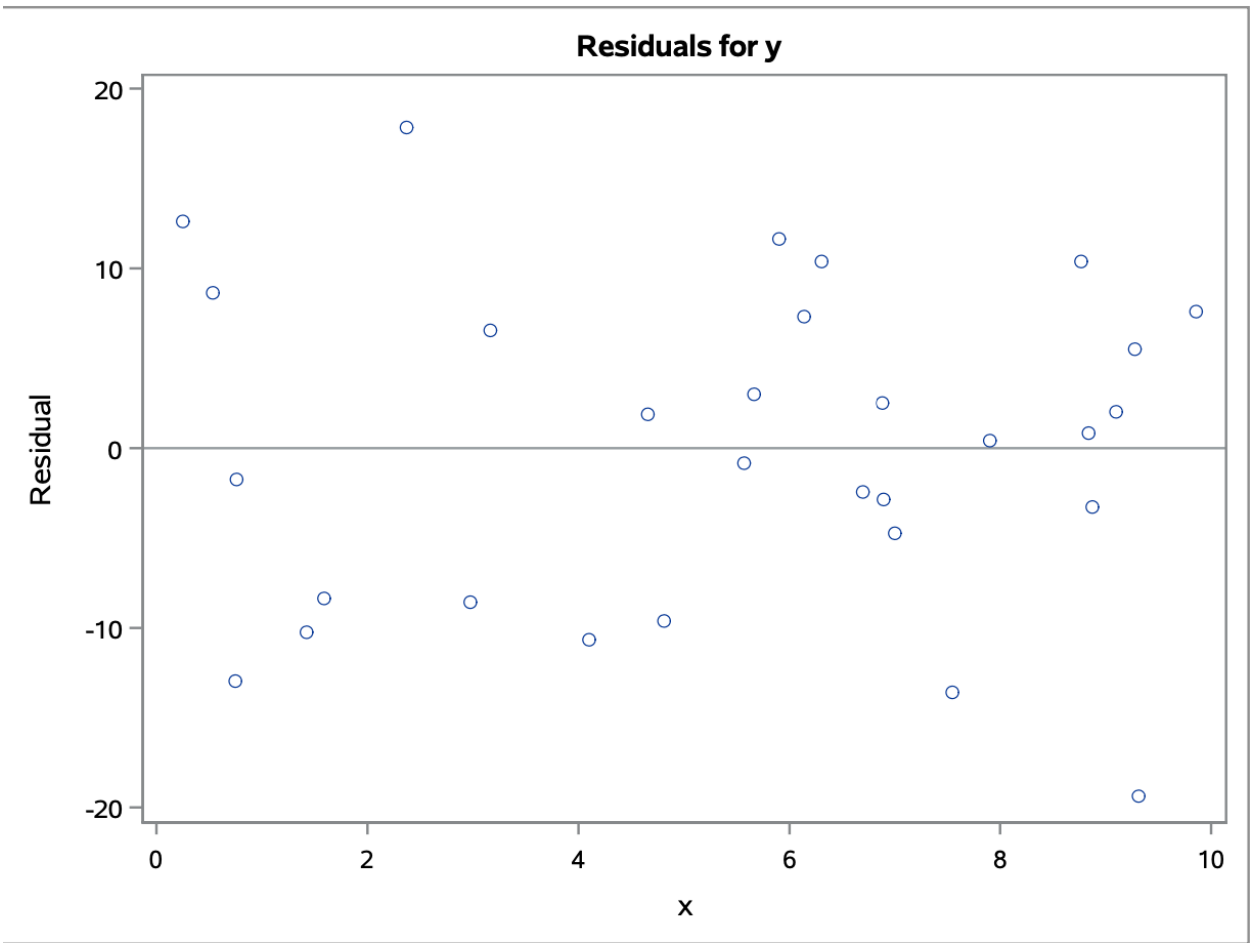
Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	1520.12331	1520.12331	88.32	<.0001
Error	48	826.17436	17.21197		
Corrected Total	49	2346.29767			

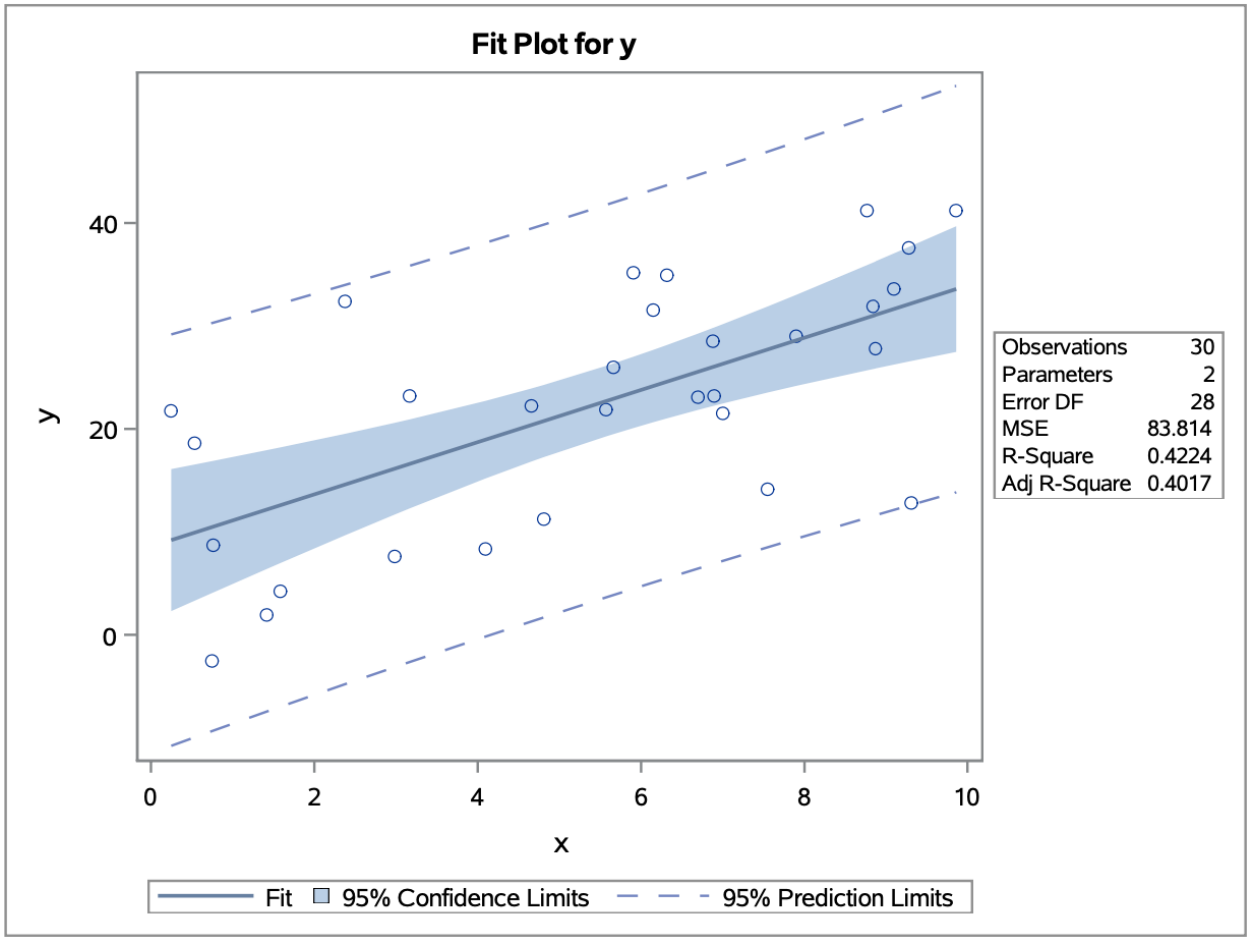
Root MSE	4.14873	R-Square	0.6479
Dependent Mean	−1.51809	Adj R-Sq	0.6405
Coeff Var	−273.28688		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	9.56379	1.31710	7.26	<.0001
x	1	−1.95653	0.20819	−9.40	<.0001




Fit Diagnostics for y
Residual
Predicted Value
RStudent
Predicted Value
RStudent
Leverage
Residual
Quantile
y
Predicted Value
Cook's D
Observation
Percent
Residual
Fit–Mean
Residual
Proportion Less
Observations 30
Parameters 2
Error DF 28
MSE 83.814
R-Square 0.4224
Adj R-Square 0.4017


Residuals for y
Residual
20
10
0
-10
-20
0
2
4
6
8
10
x


Fit Plot for y
y
x
40
20
0
0
2
4
6
8
10
Observations 30
Parameters 2
Error DF 28
MSE 83.814
R-Square 0.4224
Adj R-Square 0.4017
Fit
95% Confidence Limits
95% Prediction Limits