

# Qualification of R Linear Regression

Nick Lauerman

February 26, 2018

## Abstract

Evaulation of various NIST data sets for linear regression to see how to model them in R and to use them for the perpose of qualifing R for general statistical use.

## Contents

<b>1 Setup</b>	<b>1</b>	<b>4 Pontius</b>	<b>5</b>
1.1 R . . . . .	1		
1.2 Computer Information .	2	<b>5 NoInt1</b>	<b>7</b>
1.3 R Information . . . . .	2		
1.4 Data Cleaning . . . . .	2	<b>6 NoInt2</b>	<b>8</b>
<b>2 Background Information</b>	<b>2</b>	<b>7 Filip</b>	<b>9</b>
<b>3 Norris Data Set</b>	<b>4</b>	<b>8 Longley</b>	<b>13</b>

## 1 Setup

### 1.1 R

This series of tests is being run on R with only the “base” packages or libraries installed. The following commands are issued prior to runing hte tests for the reason stated

**options(digits = 15)** This is used to specify that 15 digits are to be displayed on numbers.

**path** Sets the path to the directory were the data sets are stored.

```
> options(digits = 15)
> path="~/R/workspace/qualification/raw data/Linear Regression/"
```

## 1.2 Computer Information

The degree of accuracy that can be expected from a computer is a function of several factors including the processor used. R provides a method to determine numeric tolerance based on David Goldberg (1991), “What Every Computer Scientist Should Know About Floating-Point Arithmetic”, ACM Computing Surveys, 23/1, 5-48, also available via <http://www.validlab.com/goldberg/paper.pdf>.

This value can be treated as the error value for the computer and for accuracy beyond requires careful consideration.

```
> .Machine$double.eps ^ 0.5
```

```
[1] 1.49011611938477e-08
```

## 1.3 R Information

```
> version
```

```
platform      x86_64-w64-mingw32
arch           x86_64
os             mingw32
system        x86_64, mingw32
status
major          3
minor          4.3
year           2017
month          11
day            30
svn rev        73796
language       R
version.string  R version 3.4.3 (2017-11-30)
nickname       Kite-Eating Tree
```

## 1.4 Data Cleaning

all data sets downloaded from NIST as a DAT (ASCII Format) file were cleaned up to remove header information that was imbedded in the file. The file was then saved as a TXT file without any additional changes. This clean up was done to simplify the loading of the data into R.

## 2 Background Information

Even with the availability of reliable code for linear least squares fitting, problems persist. Failure to use the best algorithms and to implement them most effectively is often the cause. Therefore, we provide datasets with certified values for key statistics for testing

linear least squares code. Both generated and "real-world" data are included. Generated datasets challenge specific computations and include the Wampler data developed at NIST (formerly NBS) in the early 1970's. Real-world data include the challenging Longley data, as well as more benign datasets from our statistical consulting work at NIST.

Datasets are ordered by level of difficulty (lower, average, and higher). Strictly speaking the level of difficulty of a dataset depends on the algorithm used. These levels are intended to provide rough guidance for the user. Datasets of lower level of difficulty should pose few problems for most code. Discrepancies here may indicate a failure to use correct options for the code. Two datasets are included for fitting a line through the origin. We have encountered codes that produce negative R-squared and incorrect F-statistics for these datasets. Therefore, we assign them an "average" level of difficulty. Finally, several datasets of higher level of difficulty are provided. These datasets are multicollinear. They include the Longley data and several NIST datasets developed by Wampler.

Producing correct results on all datasets of higher difficulty does not imply that your software will pass all datasets of average or even lower difficulty. Similarly, producing correct results for all datasets in this collection does not imply that your software will do the same for your particular dataset. It will, however, provide some degree of assurance, in the sense that your package provides correct results for datasets known to yield incorrect results for some software.

Certified values are provided for the parameter estimates, their standard deviations, the residual standard deviation, R-squared, and the standard ANOVA table for linear regression. Certified values are quoted to 16 significant digits and are accurate up to the last digit, due to possible truncation errors. For more information on certification methodology, see the description provided for each dataset.

If your code fails to produce correct results for a dataset of higher level of difficulty, one possible remedy is to center the data and rerun the code. Centering the data, i.e., subtracting the mean for each predictor variable, reduces the degree of multicollinearity. The code may produce correct results for the centered data. You can judge this by comparing predicted values from the fit of centered data with those from the certified fit.

We plan to update this collection of datasets, and welcome your feedback on specific datasets to include, and on other ways to improve this web service.

### 3 Norris Data Set

Using the Norris data set (<http://www.itl.nist.gov/div898/strd/lls/data/Norris.shtml>)

#### Data Set Description

These data are from a NIST study involving calibration of ozone monitors. The response variable (y) is the customer's measurement of ozone concentration and the predictor variable (x) is NIST's measurement of ozone concentration.

#### About the data set

Data Set Properties

**Level of Difficulty** Lower

**Model Class** Linear

**Number of Parameters** 2

**Number of observations** 36

**Predictor variable(s)** 1

**Response variable** 1

#### Certified Regression Statistics

Parameter	Estimate	Standard Deviation of Estimate
Intercept	-0.262323073774029	0.232818234301152
X	1.00211681802045	0.429796848199937E-03

**Residual Standard Deviation** 0.884796396144373

**R-Squared** 0.999993745883712

#### Certified Analysis of Variance Table

Source of Variation	Degrees of Freedom	Sums of Squares	Mean Squares	F Statistic
Regression	1	4255954.13232369	4255954.13232369	5436385.54
Residual	34	26.6173985294224	0.782864662630069	

```
> Norris <- read.table(file=paste0(path,"Norris.txt"), header=TRUE)
> Norris.lm <- lm(y~x, data=Norris)
> summary(Norris.lm)
```

Call:

```
lm(formula = y ~ x, data = Norris)
```

Residuals:

Min	1Q	Median	3Q
-2.3523781286599	-0.5326967162014	-0.0296292259639	0.6000277736811
Max			

1.7897858528800

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-0.2623230737741	0.2328182343012	-1.12673	0.26775
x	1.0021168180205	0.0004297968482	2331.60579	< 2e-16 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.884796396144 on 34 degrees of freedom

Multiple R-squared: 0.999993745884, Adjusted R-squared: 0.999993561939

F-statistic: 5436385.5408 on 1 and 34 DF, p-value: < 2.220446049e-16

> anova(Norris.lm)

Analysis of Variance Table

Response: y

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
x	1	4255954.132324	4255954.132324	5436385.5408	< 2.22e-16 ***
Residuals	34	26.617399	0.782865		

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## 4 Pontius

Using the Pontius data set (<http://www.itl.nist.gov/div898/strd/lls/data/Pontius.shtml>)

### Data Set Description

These data are from a NIST study involving calibration of load cells. The response variable (y) is the deflection and the predictor variable (x) is load.

### About the data set

Data Set Properties

Level of Difficulty Lower

Model Class Quadratic

Number of Parameters 3

Number of observations 40

Predictor variable(s) 1

Response variable 1

Parameter	Estimate	Standard Deviation of Estimate			
Intercept	0.673565789473684E-03	0.107938612033077E-03			
X	0.732059160401003E-06	0.157817399981659E-09			
X 2	-0.316081871345029E-14	0.486652849992036E-16			
Residual Standard Deviation	0.205177424076185E-03				
R-Squared	0.99999990017853				
Source of Variation	Degrees of Freedom	Sums of Squares	Mean Squares	F	P
Regression	2	15.6040343244198	7.80201716220991	18	
Residual	37	0.155761768796992E-05	0.420977753505385E-07		

```
> Pontius <- read.table(file=paste0(path,"Pontius.txt"), header=TRUE)
> Pontius.lm <- lm(y~x + I(x^2), data=Pontius)
> summary(Pontius.lm)
```

```
Call:
lm(formula = y ~ x + I(x^2), data = Pontius)
```

```
Residuals:
      Min       1Q   Median       3Q      Max
-4.46840225564e-04 -1.57827067669e-04  3.81729323308e-05  1.08788533835e-04
 4.23453007519e-04
```

```
Coefficients:
              Estimate      Std. Error  t value Pr(>|t|)
(Intercept) 6.73565789474e-04  1.07938612033e-04   6.24027 2.9705e-07 ***
x           7.32059160401e-07  1.57817399982e-10 4638.64669 < 2.22e-16 ***
I(x^2)      -3.16081871345e-15  4.86652849992e-17 -64.95017 < 2.22e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 0.000205177424076 on 37 degrees of freedom
Multiple R-squared:  0.999999900179,    Adjusted R-squared:  0.999999894783
F-statistic: 185330865.996 on 2 and 37 DF,  p-value: < 2.220446049e-16
```

```
> anova(Pontius.lm)
```

#### Analysis of Variance Table

```
Response: y
      Df    Sum Sq   Mean Sq    F value    Pr(>F)
x       1 15.603856733899 15.603856733899 370657513.46639 < 2.22e-16 ***
I(x^2)   1  0.000177590520  0.000177590520   4218.52506 < 2.22e-16 ***
Residuals 37  0.000001557618  0.000000042098
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

## 5 NoInt1

Using the NoInt1 data set (<http://www.itl.nist.gov/div898/strd/lls/data/NoInt1.shtml>)

### Data Set Description

This dataset is constructed to test the ability of the software to recognize a statistically significant slope for a line fit through the origin (large positive value of the F statistic)

### About the data set

Data Set Properties

Level of Difficulty Average

Model Class linear

Number of Parameters 1

Number of observations 11

Predictor variable(s) 1

Response variable 1

Parameter	Estimate	Standard Deviation of Estimate			
X	2.07438016528926	0.165289256198347E-01			
Residual Standard Deviation		3.56753034006338			
R-Squared		0.999365492298663			
Source of Variation	Degrees of Freedom	Sums of Squares	Mean Squares	F Statistic	
Regression	1	200457.727272727	200457.727272727	15750.25000	
Residual	10	127.272727272727	12.7272727272727		

```
> NoInt1 <- read.table(file=paste0(path,"NoInt1.txt"), header=TRUE)
> NoInt1.lm <- lm(y~x + 0, data=NoInt1)
> summary(NoInt1.lm)
```

Call:

```
lm(formula = y ~ x + 0, data = NoInt1)
```

Residuals:

	Min	1Q	Median	3Q	Max
-5.206611570248	-2.520661157025	0.165289256198	2.851239669421	5.537190082645	

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
x	2.0743801652893	0.0165289256198	125.5	< 2.22e-16 ***
---				

```

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.56753034006 on 10 degrees of freedom
Multiple R-squared:  0.999365492299,    Adjusted R-squared:  0.999302041529
F-statistic:      15750.25 on 1 and 10 DF,  p-value: < 2.220446049e-16

> anova(NoInt1.lm)

Analysis of Variance Table

Response: y
      Df      Sum Sq      Mean Sq  F value    Pr(>F)
x         1 200457.72727273 200457.72727273 15750.25 < 2.22e-16 ***
Residuals 10    127.27272727    12.72727273
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

## 6 NoInt2

Using the NoInt2 data set (<http://www.itl.nist.gov/div898/strd/lls/data/NoInt2.shtml>)

### Data Set Description

#### About the data set

Data Set Properties

Level of Difficulty Average

Model Class Linear

Number of Parameters 1

Number of observations 3

Predictor variable(s) 1

Response variable 1

Parameter	Estimate	Standard Deviation of Estimate
X	0.727272727272727	

Residual Standard Deviation	0.369274472937998
R-Squared	0.993348115299335

Source of Variation	Degrees of Freedom	Sums of Squares	Mean Squares	F Statistic
Regression	1	40.7272727272727	0.7272727272727	298.666666666667
Residual	2	0.272727272727273	0.136363636363636	



```

> NoInt2 <- read.table(file=paste0(path,"NoInt2.txt"), header=TRUE)
> NoInt2.lm <- lm(y~x + 0, data=NoInt2)
> summary(NoInt2.lm)

Call:
lm(formula = y ~ x + 0, data = NoInt2)

Residuals:
    1          2          3
0.0909090909091  0.3636363636364 -0.3636363636364

Coefficients:
            Estimate      Std. Error  t value Pr(>|t|)
x 0.7272727272727 0.0420827318078 17.28198 0.0033315 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.369274472938 on 2 degrees of freedom
Multiple R-squared:  0.993348115299,    Adjusted R-squared:  0.990022172949
F-statistic: 298.666666667 on 1 and 2 DF,  p-value: 0.00333149176904

> anova(NoInt2.lm)

Analysis of Variance Table

Response: y
      Df      Sum Sq      Mean Sq    F value    Pr(>F)
x       1 40.727272727 40.727272727 298.66667 0.0033315 **
Residuals  2  0.272727273  0.136363636
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

## 7 Filip

Using the Filip data set (<http://www.itl.nist.gov/div898/strd/lls/data/Filip.shtml>)

### Data Set Description

None supplied

### About the data set

Data Set Properties

**Level of Difficulty** Higher

**Model Class** Polynomial

Number of Parameters 11

Number of observations 82

Predictor variable(s) 1

Response variable 1

Parameter	Estimate	Standard Deviation of Estimate
Intercept	-1467.48961422980	298.084530995537
X	-2772.17959193342	559.779865474950
X 2	-2316.37108160893	466.477572127796
X 3	-1127.97394098372	227.204274477751
X 4	-354.478233703349	71.6478660875927
X 5	-75.1242017393757	15.2897178747400
X 6	-10.8753180355343	2.23691159816033
X 7	-1.06221498588947	0.221624321934227
X 8	-0.670191154593408E-01	0.142363763154724E-01
X 9	-0.246781078275479E-02	0.535617408889821E-03
X 10	-0.402962525080404E-04	0.896632837373868E-05

Residual Standard Deviation 0.334801051324544E-02

R-Squared 0.996727416185620

Source of Variation	Degrees of Freedom	Sums of Squares	Mean Squares	F
Regression	10	0.242391619837339	0.242391619837339E-01	21
Residual	71	0.795851382172941E-03	0.112091743968020E-04	

```
> Filip <- read.table(file=paste0(path,"Filip.txt"), header=TRUE)
> Filip.lm <- lm(y ~ poly(x,10,raw = TRUE), data=Filip)
> summary(Filip.lm)
```

Call:

```
lm(formula = y ~ poly(x, 10, raw = TRUE), data = Filip)
```

Residuals:

Min	1Q	Median	3Q
-0.009908657629677	-0.002461025773308	0.000338470313538	0.002074343959809
Max			
0.007165411201025			

Coefficients: (1 not defined because of singularities)

	Estimate	Std. Error	t value
(Intercept)	-1.74280445635e+02	8.75611622176e+01	-1.99039
poly(x, 10, raw = TRUE)1	-3.26882209949e+02	1.48049618157e+02	-2.20792
poly(x, 10, raw = TRUE)2	-2.66056540511e+02	1.09512084514e+02	-2.42947
poly(x, 10, raw = TRUE)3	-1.23921614440e+02	4.65247147225e+01	-2.66357
poly(x, 10, raw = TRUE)4	-3.63816709520e+01	1.25145200544e+01	-2.90716
poly(x, 10, raw = TRUE)5	-6.97918838076e+00	2.21116609894e+00	-3.15634

```

poly(x, 10, raw = TRUE)6 -8.74660178695e-01 2.56744908247e-01 -3.40673
poly(x, 10, raw = TRUE)7 -6.90600974071e-02 1.89005628454e-02 -3.65386
poly(x, 10, raw = TRUE)8 -3.11832190160e-03 8.00878984696e-04 -3.89362
poly(x, 10, raw = TRUE)9 -6.13867083353e-05 1.48905168694e-05 -4.12254
poly(x, 10, raw = TRUE)10 NA NA NA
      Pr(>|t|)
(Intercept) 0.05034547 .
poly(x, 10, raw = TRUE)1 0.03043559 *
poly(x, 10, raw = TRUE)2 0.01761673 *
poly(x, 10, raw = TRUE)3 0.00953389 **
poly(x, 10, raw = TRUE)4 0.00484493 **
poly(x, 10, raw = TRUE)5 0.00233320 **
poly(x, 10, raw = TRUE)6 0.00107912 **
poly(x, 10, raw = TRUE)7 0.00048725 ***
poly(x, 10, raw = TRUE)8 0.00021857 ***
poly(x, 10, raw = TRUE)9 9.907e-05 ***
poly(x, 10, raw = TRUE)10 NA
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.00376801218783 on 72 degrees of freedom
Multiple R-squared:  0.995796453099,    Adjusted R-squared:  0.995271009736
F-statistic: 1895.15469013 on 9 and 72 DF,  p-value: < 2.220446049e-16

> anova(Filip.lm)

Analysis of Variance Table

Response: y
      Df Sum Sq Mean Sq F value
poly(x, 10, raw = TRUE) 9 0.24216522127848 0.026907246808720 1895.15469
Residuals              72 0.00102224994103 0.000014197915848
      Pr(>F)
poly(x, 10, raw = TRUE) < 2.22e-16 ***
Residuals
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> Filip2.lm <- lm(y~x +
+               I(x^2) +
+               I(x^3) +
+               I(x^4) +
+               I(x^5) +
+               I(x^6) +
+               I(x^7) +
+               I(x^8) +
+               I(x^9) +

```

```

+               I(x^10)
+               , data=Filip)
> summary(Filip2.lm)

Call:
lm(formula = y ~ x + I(x^2) + I(x^3) + I(x^4) + I(x^5) + I(x^6) +
    I(x^7) + I(x^8) + I(x^9) + I(x^10), data = Filip)

Residuals:
      Min       1Q   Median       3Q      Max
-0.009908657629677 -0.002461025773308  0.000338470313538  0.002074343959809
 0.007165411201025

Coefficients: (1 not defined because of singularities)
              Estimate      Std. Error  t value  Pr(>|t|)
(Intercept) -1.74280445635e+02  8.75611622176e+01 -1.99039  0.05034547 .
x            -3.26882209949e+02  1.48049618157e+02 -2.20792  0.03043559 *
I(x^2)       -2.66056540511e+02  1.09512084514e+02 -2.42947  0.01761673 *
I(x^3)       -1.23921614440e+02  4.65247147225e+01 -2.66357  0.00953389 **
I(x^4)       -3.63816709520e+01  1.25145200544e+01 -2.90716  0.00484493 **
I(x^5)       -6.97918838076e+00  2.21116609894e+00 -3.15634  0.00233320 **
I(x^6)       -8.74660178695e-01  2.56744908247e-01 -3.40673  0.00107912 **
I(x^7)       -6.90600974071e-02  1.89005628454e-02 -3.65386  0.00048725 ***
I(x^8)       -3.11832190160e-03  8.00878984696e-04 -3.89362  0.00021857 ***
I(x^9)       -6.13867083353e-05  1.48905168694e-05 -4.12254  9.907e-05 ***
I(x^10)      NA                  NA          NA      NA
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.00376801218783 on 72 degrees of freedom
Multiple R-squared:  0.995796453099,    Adjusted R-squared:  0.995271009736
F-statistic: 1895.15469013 on 9 and 72 DF,  p-value: < 2.220446049e-16

> anova(Filip2.lm)

Analysis of Variance Table

Response: y
      Df Sum Sq Mean Sq F value Pr(>F)
x      1 0.21288106025948 0.21288106025948 14993.82462 < 2.22e-16 ***
I(x^2) 1 0.00753409869624 0.00753409869624  530.64821 < 2.22e-16 ***
I(x^3) 1 0.00683749292831 0.00683749292831  481.58427 < 2.22e-16 ***
I(x^4) 1 0.00935927452572 0.00935927452572  659.20059 < 2.22e-16 ***
I(x^5) 1 0.00030458358215 0.00030458358215   21.45270 1.5707e-05 ***
I(x^6) 1 0.00380533483833 0.00380533483833  268.02066 < 2.22e-16 ***
I(x^7) 1 0.00004444148251 0.00004444148251    3.13014 0.081092 .

```

```

I(x^8)      1 0.00115763695335 0.00115763695335      81.53570 1.8390e-13 ***
I(x^9)      1 0.00024129801239 0.00024129801239      16.99531 9.9070e-05 ***
Residuals 72 0.00102224994103 0.00001419791585
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

## 8 Longley

Using the Longley data set (<http://www.itl.nist.gov/div898/strd/lls/data/Longley.shtml>)

### Data Set Description

This classic dataset of labor statistics was one of the first used to test the accuracy of least squares computations. The response variable (y) is the Total Derived Employment and the predictor variables are GNP Implicit Price Deflator with Year 1954 = 100 (x1), Gross National Product (x2), Unemployment (x3), Size of Armed Forces (x4), Non-Institutional Population Age 14 & Over (x5), and Year (x6).

### About the data set

Data Set Properties

**Level of Difficulty** Higher

**Model Class** Multilinear

**Number of Parameters** 7

**Number of observations** 16

**Predictor variable(s)** 6

**Response variable** 1

Parameter	Estimate	Standard Deviation of Estimate
Intercept	-3482258.63459582	890420.383607373
X1	15.0618722713733	84.9149257747669
X2	-0.358191792925910E-01	0.334910077722432E-01
X3	-2.02022980381683	0.488399681651699
X4	-1.03322686717359	0.214274163161675
X5	-0.511041056535807E-01	0.226073200069370
X6	1829.15146461355	455.478499142212

**Residual Standard Deviation** 304.854073561965

**R-Squared** 0.995479004577296

Source of Variation	Degrees of Freedom	Sums of Squares	Mean Squares	F Statistic
Regression	6	184172401.944494	30695400.3240823	330.2853392
Residual	9	836424.055505915	92936.0061673238	

```

> Longley <- read.table(file=paste0(path,"Longley.txt"), header=TRUE)
> Longley.lm <- lm(y ~ .,data=Longley)
> summary(Longley.lm)

Call:
lm(formula = y ~ ., data = Longley)

Residuals:
    Min       1Q   Median       3Q      Max
-410.1146219309 -157.6747192954  -28.1619848188  101.5503832581  455.3940945519

Coefficients:
              Estimate      Std. Error  t value Pr(>|t|)
(Intercept) -3.48225863460e+06  8.90420383607e+05  -3.91080  0.00356040 **
x1             1.50618722714e+01  8.49149257748e+01   0.17738  0.86314083
x2            -3.58191792926e-02  3.34910077722e-02  -1.06952  0.31268106
x3            -2.02022980382e+00  4.88399681652e-01  -4.13643  0.00253509 **
x4            -1.03322686717e+00  2.14274163162e-01  -4.82199  0.00094437 ***
x5            -5.11041056536e-02  2.26073200069e-01  -0.22605  0.82621180
x6             1.82915146461e+03  4.55478499142e+02   4.01589  0.00303680 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 304.854073562 on 9 degrees of freedom
Multiple R-squared:  0.995479004577,    Adjusted R-squared:  0.992465007629
F-statistic: 330.285339235 on 6 and 9 DF,  p-value: 4.98403052872e-10

> anova(Longley.lm)

Analysis of Variance Table

Response: y
      Df    Sum Sq    Mean Sq    F value    Pr(>F)
x1      1 174397449.77913 174397449.77913 1876.53265 9.2954e-12 ***
x2      1  4787181.04445  4787181.04445  51.51051 5.2109e-05 ***
x3      1  2263971.10982  2263971.10982  24.36054 0.00080706 ***
x4      1    876397.16186    876397.16186   9.43011 0.01333568 *
x5      1    348589.39965    348589.39965   3.75085 0.08475523 .
x6      1  1498813.44959  1498813.44959  16.12737 0.00303680 **
Residuals 9    836424.05551    92936.00617
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```