Qualification of R Linear Regression

Nick Lauerman

March 20, 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

1 Setup

1.1 R

This seriess 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 = 22)
> path="~/R/workspace/qualification/raw data/Linear Regression/"
```

1.2 Computer Information

The degree of accuracy that can be expected from a computer if 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] 1.4901161193847656e-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 imbeded in the file. The file was than 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 The Data Sets

3.1 Norris Data Set

Using the Norris data set `://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

Certifed 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	5436385.54079785
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
-2.352378128659911916 -0.532696716201364384 -0.029629225963878517
                   3Q                  Max
 0.600027773681106935  1.789785852880034112

Coefficients:
                          Estimate              Std. Error            t value
(Intercept) -0.26232307377411717697  0.23281823430115430873   -1.1267300000000
x            1.00211681802045426970  0.00042979684819994022 2331.6057900000001
            Pr(>|t|)
(Intercept)  0.26775
x            < 2e-16 ***
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.050000000000000003 '.' 0.10000000000000001 ' ' 1

Residual standard error: 0.88479639614437944 on 34 degrees of freedom
Multiple R-squared:  0.9999937458837117044,        Adjusted R-squared:  0.9999935619391150388
F-statistic: 5436385.540797759779 on 1 and 34 DF,  p-value: < 2.2204460492503131e-16

> anova(Norris.lm)

Analysis of Variance Table

Response: y
```

```
          Df                 Sum Sq                 Mean Sq            F value
x          1 4255954.1323236934841 4.2559541323236935e+06 5436385.5407999996
Residuals 34      26.6173985294228 7.8286466263010002e-01
              Pr(>F)
x         < 2.22e-16 ***
Residuals
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.050000000000000003 '.' 0.10000000000000001 ' ' 1
```

3.2 Pontius

Using the Pontius data set ://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 Statistic
Regression	2	15.6040343244198	7.80201716220991	185330865.995752
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
-4.4684022556417666e-04 -1.5782706766922605e-04  3.8172932330817434e-05
                      3Q                      Max
 1.0878853383451323e-04  4.2345300751879777e-04

Coefficients:
                           Estimate              Std. Error
(Intercept)  6.7356578947353482e-04  1.0793861203308400e-04
x            7.3205916040100258e-07  1.5781739998166896e-10
I(x^2)      -3.1608187134503184e-15  4.8665284999206759e-17
                          t value   Pr(>|t|)
(Intercept)    6.2402699999999998 2.9705e-07 ***
x           4638.6466899999995803 < 2.22e-16 ***
I(x^2)       -64.9501700000000000 < 2.22e-16 ***
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.050000000000000003 '.' 0.10000000000000001 ' ' 1

Residual standard error: 0.00020517742407619811 on 37 degrees of freedom
Multiple R-squared:  0.9999999001785371266,        Adjusted R-squared:  0.9999998947827823947
F-statistic:  185330865.995727241 on 2 and 37 DF,  p-value: < 2.2204460492503131e-16

> anova(Pontius.lm)
```

```
Analysis of Variance Table

Response: y
          Df                  Sum Sq                 Mean Sq
x          1 1.5603856733899418e+01 1.5603856733899418e+01
I(x^2)     1 1.7759052039473990e-04 1.7759052039473990e-04
Residuals 37 1.5576176879701001e-06 4.2097775350499999e-08
                         F value     Pr(>F)
x         3.7065751346639001e+08 < 2.22e-16 ***
I(x^2)    4.2185250599999999e+03 < 2.22e-16 ***
Residuals
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.050000000000000003 '.' 0.10000000000000001 ' ' 1
```

3.3 NoInt1

Using the NoInt1 data set ://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	15750.2500000000
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
-5.20661157024794718 -2.52066115702480609  0.16528925619836343
                  3Q                   Max
 2.85123966942149032  5.53719008264467316

Coefficients:
              Estimate           Std. Error t value   Pr(>|t|)
x 2.074380165289256173 0.016528925619834767   125.5 < 2.22e-16 ***
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.050000000000000003 '.' 0.10000000000000001 ' ' 1

Residual standard error: 3.5675303400633909 on 10 degrees of freedom
Multiple R-squared:  0.9993654922986627831,        Adjusted R-squared:  0.9993020415285290836
F-statistic:  15750.2499999998945 on 1 and 10 DF,  p-value: < 2.2204460492503131e-16

> anova(NoInt1.lm)

Analysis of Variance Table
```

```
Response: y
          Df                 Sum Sq                Mean Sq  F value     Pr(>F)
x          1 200457.72727272729389 200457.727272727293894 15750.25 < 2.22e-16
Residuals 10    127.27272727272812     12.727272727272812

x         ***
Residuals
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.050000000000000003 '.' 0.10000000000000001 ' ' 1
```

3.4 NoInt2

Using the NoInt2 data set ://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	298.66666666666
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
 0.090909090909088691  0.363636363636364424 -0.363636363636362869

Coefficients:
              Estimate           Std. Error            t value  Pr(>|t|)
x 0.727272727272726960 0.042082731807843207 17.281980000000001 0.0033315 **
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.050000000000000003 '.' 0.10000000000000001 ' ' 1

Residual standard error: 0.36927447293799792 on 2 degrees of freedom
Multiple R-squared:  0.9933481152993347552,        Adjusted R-squared:  0.9900221729490021882
F-statistic: 298.6666666666671404 on 1 and 2 DF,  p-value: 0.003331491769036167

> anova(NoInt2.lm)

Analysis of Variance Table

Response: y
          Df               Sum Sq              Mean Sq            F value
x          1 40.72727272727271242 40.72727272727271242 298.66667000000001
Residuals  2  0.27272727272727232  0.13636363636363616
             Pr(>F)
x         0.0033315 **
Residuals
```

```
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.050000000000000003 '.' 0.10000000000000001 ' ' 1
```

3.5 Filip

Using the Filip data set ://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 Statistic**
Regression	10	0.242391619837339	0.242391619837339E-01	2162.43954511489
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
-0.0099086576296772491 1 -0.0024610257733083457 68  0.0003384703135377156 5
                      3Q                      Max
 0.0020743439598085419 87  0.0071654112010250941 4

Coefficients: (1 not defined because of singularities)
                                     Estimate              Std. Error
(Intercept)               -1.7428044563463959e+02  8.7561162217625807e+01
poly(x, 10, raw = TRUE)1  -3.2688220994861894e+02  1.4804961815707296e+02
poly(x, 10, raw = TRUE)2  -2.6605654051119797e+02  1.0951208451353565e+02
poly(x, 10, raw = TRUE)3  -1.2392161444002367e+02  4.6524714722544466e+01
poly(x, 10, raw = TRUE)4  -3.6381670952047024e+01  1.2514520054377089e+01
poly(x, 10, raw = TRUE)5  -6.9791883807598705e+00  2.2111660989355579e+00
poly(x, 10, raw = TRUE)6  -8.7466017869459656e-01  2.5674490824684437e-01
poly(x, 10, raw = TRUE)7  -6.9060097407069573e-02  1.8900562845353470e-02
poly(x, 10, raw = TRUE)8  -3.1183219015965633e-03  8.0087898469617799e-04
poly(x, 10, raw = TRUE)9  -6.1386708335283072e-05  1.4890516869441546e-05
poly(x, 10, raw = TRUE)10                      NA                    NA
                                  t value   Pr(>|t|)
(Intercept)               -1.9903900000000001 0.05034547 .
poly(x, 10, raw = TRUE)1  -2.2079200000000001 0.03043559 *
poly(x, 10, raw = TRUE)2  -2.4294699999999998 0.01761673 *
```

```
poly(x, 10, raw = TRUE)3  -2.6635700000000000 0.00953389 **
poly(x, 10, raw = TRUE)4  -2.9071600000000002 0.00484493 **
poly(x, 10, raw = TRUE)5  -3.1563400000000001 0.00233320 **
poly(x, 10, raw = TRUE)6  -3.4067300000000000 0.00107912 **
poly(x, 10, raw = TRUE)7  -3.6538599999999999 0.00048725 ***
poly(x, 10, raw = TRUE)8  -3.8936199999999999 0.00021857 ***
poly(x, 10, raw = TRUE)9  -4.1225399999999999  9.907e-05 ***
poly(x, 10, raw = TRUE)10                  NA         NA
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.050000000000000003 '.' 0.10000000000000001 ' ' 1

Residual standard error: 0.0037680121878278122 on 72 degrees of freedom
Multiple R-squared:  0.995764530989105167,        Adjusted R-squared:  0.9952710097362743591
F-statistic: 1895.154690132417954 on 9 and 72 DF,  p-value: < 2.2204460492503131e-16

> anova(Filip.lm)

Analysis of Variance Table

Response: y
                        Df                 Sum Sq                 Mean Sq
poly(x, 10, raw = TRUE)  9 0.2421652212784838054 2.6907246808720423e-02
Residuals               72 0.0010222499410285635 1.4197915847618938e-05
                                    F value     Pr(>F)
poly(x, 10, raw = TRUE) 1895.1546900000001 < 2.22e-16 ***
Residuals
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.050000000000000003 '.' 0.10000000000000001 ' ' 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
-0.0099086576296772491 1 -0.0024610257733083457 68  0.0003384703135377156 5
                       3Q                      Max
 0.0020743439598085419 87  0.0071654112010250941 4

Coefficients: (1 not defined because of singularities)
                           Estimate              Std. Error              t value
(Intercept) -1.7428044563463959e+02  8.7561162217625807e+01 -1.9903900000000001
x           -3.2688220994861894e+02  1.4804961815707296e+02 -2.2079200000000001
I(x^2)      -2.6605654051119797e+02  1.0951208451353565e+02 -2.4294699999999998
I(x^3)      -1.2392161444002367e+02  4.6524714722544466e+01 -2.6635700000000000
I(x^4)      -3.6381670952047024e+01  1.2514520054377089e+01 -2.9071600000000002
I(x^5)      -6.9791883807598705e+00  2.2111660989355579e+00 -3.1563400000000001
I(x^6)      -8.7466017869459656e-01  2.5674490824684437e-01 -3.4067300000000000
I(x^7)      -6.9060097407069573e-02  1.8900562845353470e-02 -3.6538599999999999
I(x^8)      -3.1183219015965633e-03  8.0087898469617799e-04 -3.8936199999999999
I(x^9)      -6.1386708335283072e-05  1.4890516869441546e-05 -4.1225399999999999
I(x^10)                          NA                      NA                   NA
             Pr(>|t|)
(Intercept) 0.05034547 .
x           0.03043559 *
I(x^2)      0.01761673 *
I(x^3)      0.00953389 **
I(x^4)      0.00484493 **
```

```
I(x^5)      0.00233320 **
I(x^6)      0.00107912 **
I(x^7)      0.00048725 ***
I(x^8)      0.00021857 ***
I(x^9)       9.907e-05 ***
I(x^10)             NA
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.050000000000000003 '.' 0.10000000000000001 ' ' 1

Residual standard error: 0.0037680121878278122 on 72 degrees of freedom
Multiple R-squared:  0.9957964530989105167,        Adjusted R-squared:  0.9952710097362743591
F-statistic: 1895.154690132417954 on 9 and 72 DF,  p-value: < 2.2204460492503131e-16

> anova(Filip2.lm)

Analysis of Variance Table

Response: y
          Df                  Sum Sq                 Mean Sq
x          1 2.1288106025947531e-01 2.1288106025947531e-01
I(x^2)     1 7.5340986962444470e-03 7.5340986962444470e-03
I(x^3)     1 6.8374929283144502e-03 6.8374929283144502e-03
I(x^4)     1 9.3592745257209103e-03 9.3592745257209103e-03
I(x^5)     1 3.0458358215425621e-04 3.0458358215425621e-04
I(x^6)     1 3.8053348383291141e-03 3.8053348383291141e-03
I(x^7)     1 4.4441482509089002e-05 4.4441482509089002e-05
I(x^8)     1 1.1576369533475999e-03 1.1576369533475999e-03
I(x^9)     1 2.4129801238862008e-04 2.4129801238862008e-04
Residuals 72 1.0222499410285635e-03 1.4197915847618938e-05
                        F value     Pr(>F)
x         14993.8246199999994133 < 2.22e-16 ***
I(x^2)      530.6482099999999491 < 2.22e-16 ***
I(x^3)      481.5842700000000036 < 2.22e-16 ***
I(x^4)      659.2005900000000338 < 2.22e-16 ***
I(x^5)       21.4527000000000001 1.5707e-05 ***
```

```
I(x^6)      268.0206600000000208 < 2.22e-16 ***
I(x^7)        3.1301399999999999   0.081092 .
I(x^8)       81.5357000000000056 1.8390e-13 ***
I(x^9)       16.9953099999999999 9.9070e-05 ***
Residuals
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.050000000000000003 '.' 0.10000000000000001 ' ' 1
```

3.6 Longley

Using the Longley data set ://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	330.285339234588
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
-410.114621930903809 -157.674719295390105  -28.161984818765305
                  3Q                  Max
 101.550383258075641  455.394094551858075

Coefficients:
                           Estimate              Std. Error
(Intercept) -3.482258634595823e+06  8.904203836073686e+05
x1           1.506187227137476e+01  8.491492577476685e+01
x2          -3.581917929259137e-02  3.349100777224300e-02
x3          -2.020229803816828e+00  4.883996816516961e-01
x4          -1.033226867173590e+00  2.142741631616739e-01
x5          -5.110410565357873e-02  2.260732000693687e-01
x6           1.829151464613554e+03  4.554784991422099e+02
```

```
                       t value   Pr(>|t|)
(Intercept) -3.91080000000000005 0.00356040 **
x1           0.17738000000000001 0.86314083
x2          -1.06952000000000003 0.31268106
x3          -4.13642999999999983 0.00253509 **
x4          -4.82199000000000044 0.00094437 ***
x5          -0.22605000000000000 0.82621180
x6           4.01588999999999974 0.00303680 **
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.050000000000000003 '.' 0.10000000000000001 ' ' 1

Residual standard error: 304.85407356196362 on 9 degrees of freedom
Multiple R-squared:  0.9954790045772956564,        Adjusted R-squared:  0.992465007628826057
F-statistic: 330.2853392345908219 on 6 and 9 DF,  p-value: 4.9840305287246181e-10

> anova(Longley.lm)

Analysis of Variance Table

Response: y
          Df                 Sum Sq                 Mean Sq                F value
x1         1 174397449.77912792563 174397449.779127925634 1876.5326500000001033
x2         1   4787181.04444965813   4787181.044449658133   51.5105099999999965
x3         1   2263971.10981840128   2263971.109818401281   24.3605400000000003
x4         1    876397.16186108603    876397.161861086031    9.4301100000000009
x5         1    348589.39964974835    348589.399649748346    3.7508499999999998
x6         1   1498813.44958734419   1498813.449587344192   16.1273699999999991
Residuals  9    836424.05550590809     92936.006167323125
             Pr(>F)
x1        9.2954e-12 ***
x2        5.2109e-05 ***
x3        0.00080706 ***
x4        0.01333568 *
x5        0.08475523 .
x6        0.00303680 **
```

```
Residuals
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.050000000000000003 '.' 0.10000000000000001 ' ' 1
```