Michael Dang – 16257750

COMP-SCI 5565

Linear Regression Lab

Sep 24, 2023

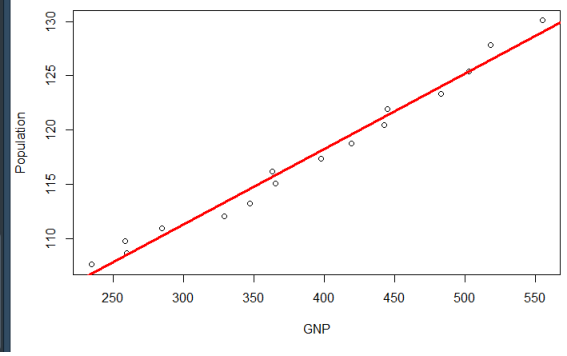
1. Simple Linear Regression

```
> head(longley)
     GNP.deflator     GNP Unemployed Armed.Forces Population Year Employed
1947         83.0 234.289      235.6        159.0    107.608 1947   60.323
1948         88.5 259.426      232.5        145.6    108.632 1948   61.122
1949         88.2 258.054      368.2        161.6    109.773 1949   60.171
1950         89.5 284.599      335.1        165.0    110.929 1950   61.187
1951         96.2 328.975      209.9        309.9    112.075 1951   63.221
1952         98.1 346.999      193.2        359.4    113.270 1952   63.639
>
```

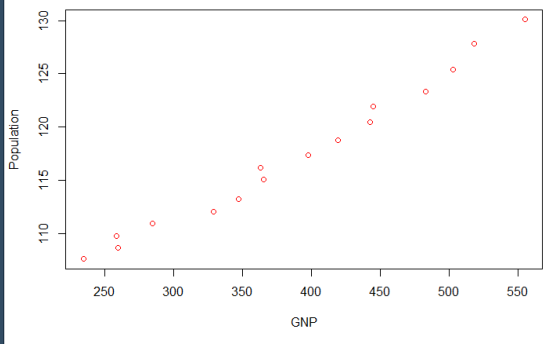
```
> lm.fit <- lm(Population ~ GNP)
> lm.fit

Call:
lm(formula = Population ~ GNP)

Coefficients:
(Intercept)          GNP
   90.53288      0.06936

> summary(lm.fit)

Call:
lm(formula = Population ~ GNP)

Residuals:
     Min       1Q   Median       3Q      Max
-1.33105 -0.78643  0.03967  0.69821  1.37817

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 90.532884   0.995173   90.97  < 2e-16 ***
GNP          0.069361   0.002491   27.84 1.17e-13 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.959 on 14 degrees of freedom
Multiple R-squared:  0.9823,    Adjusted R-squared:  0.981
F-statistic: 775.2 on 1 and 14 DF,  p-value: 1.168e-13
```

```
> names(lm.fit)
 [1] "coefficients"  "residuals"     "effects"       "rank"          "fitted.values" "assign"        "qr"
 [8] "df.residual"   "xlevels"       "call"          "terms"         "model"
> coef(lm.fit)
(Intercept)         GNP
90.53288417  0.06936091
> confint(lm.fit)
                  2.5 %      97.5 %
(Intercept) 88.39845055 92.66731780
GNP          0.06401767  0.07470414
> predict(lm.fit, data.frame(GNP = (c(300.000, 350.000, 400.000))), interval = 'confidence')
       fit      lwr      upr
1 111.3412 110.6454 112.0369
2 114.8092 114.2569 115.3615
3 118.2772 117.7588 118.7957
> predict(lm.fit, data.frame(GNP = (c(300.000, 350.000, 400.000))), interval = 'prediction')
       fit      lwr      upr
1 111.3412 109.1698 113.5125
2 114.8092 112.6794 116.9390
3 118.2772 116.1560 120.3985
>
```

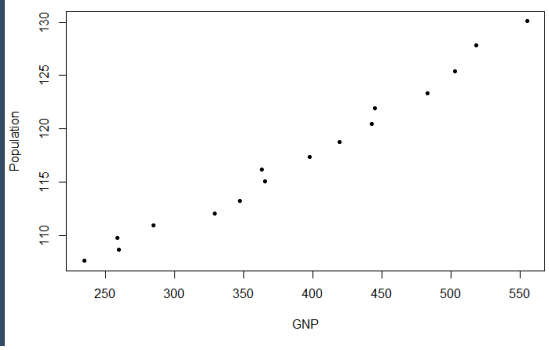
```
1 111.3412 110.6454 112.0369
2 114.8092 114.2569 115.3615
3 118.2772 117.7588 118.7957
> predict(lm.fit, data.frame(GNP = (c(300.000, 350.000, 400.000))), interval = 'prediction')
       fit      lwr      upr
1 111.3412 109.1698 113.5125
2 114.8092 112.6794 116.9390
3 118.2772 116.1560 120.3985
> plot(GNP, Population)
> abline(lm.fit)
> abline(lm.fit, lwd = 3)
> abline(lm.fit, lwd = 3, color = 'red')
Warning message:
In int_abline(a = a, b = b, h = h, v = v, untf = untf, ...) :
  "color" is not a graphical parameter
> abline(lm.fit, lwd = 3, col = 'red')
Warning messages:
1: In doTryCatch(return(expr), name, parentenv, handler) :
  "color" is not a graphical parameter
2: In doTryCatch(return(expr), name, parentenv, handler) :
  "color" is not a graphical parameter
>
```



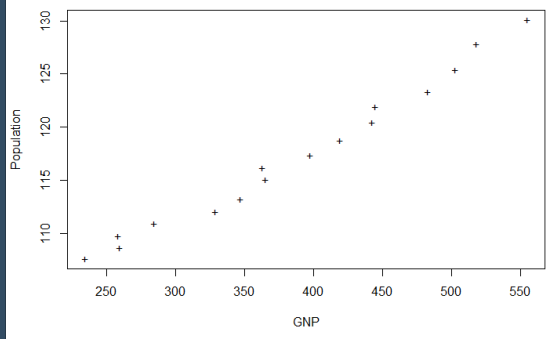

```
> plot(GNP, Population, col = 'red')
Warning messages:
1: In doTryCatch(return(expr), name, parentenv, handler) :
  "color" is not a graphical parameter
2: In doTryCatch(return(expr), name, parentenv, handler) :
  "color" is not a graphical parameter
>
```



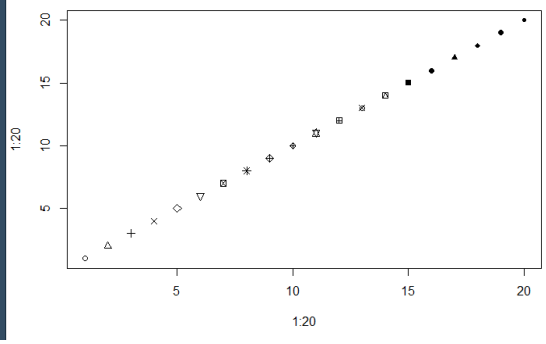

```
> plot(GNP, Population, pch = 20)
>
```



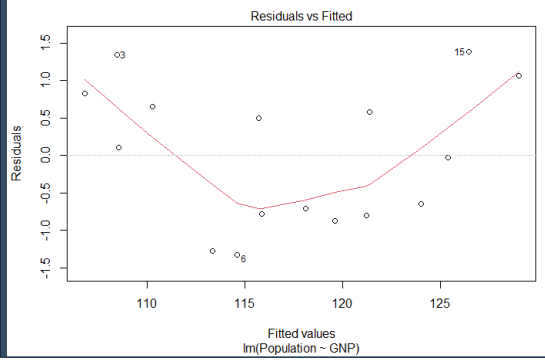
```
> plot(GNP, Population, pch = '+')
> 
```



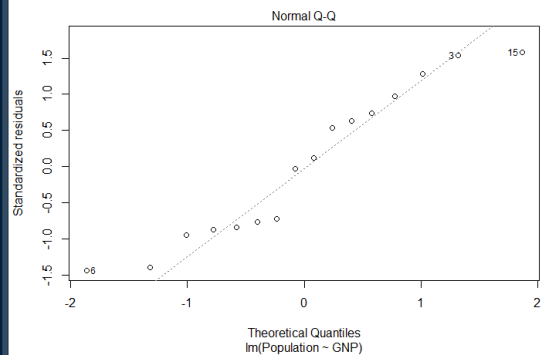

```
> plot(1:20, 1:20, pch = 1:20)
> 
```



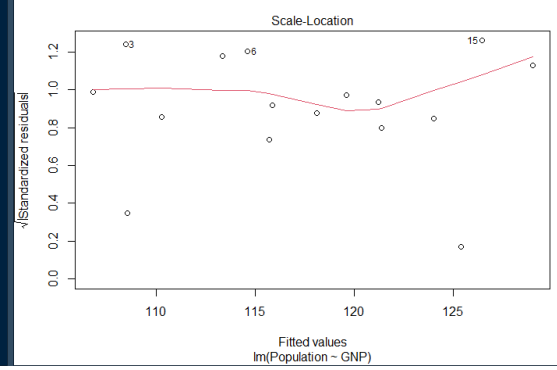

```
> par(nfrow = c(2, 2))
Warning message:
In par(nfrow = c(2, 2)) : "nfrow" is not a graphical parameter
> plot(lm.fit)
Hit <Return> to see next plot: 1
Hit <Return> to see next plot: 2
Hit <Return> to see next plot: 3
Hit <Return> to see next plot: 4
> 
```



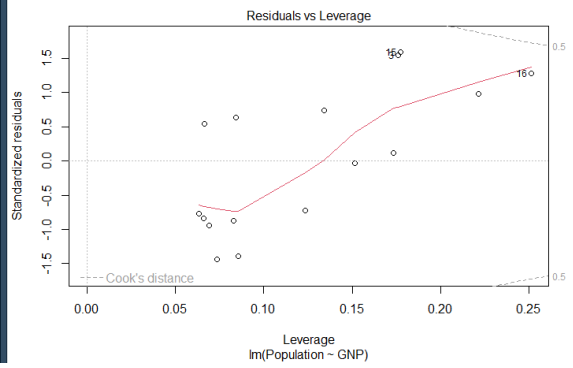

```
> par(nfrow = c(2, 2))
Warning message:
In par(nfrow = c(2, 2)) : "nfrow" is not a graphical parameter
> plot(lm.fit)
Hit <Return> to see next plot: 1
Hit <Return> to see next plot: 2
Hit <Return> to see next plot: 3
Hit <Return> to see next plot: 4
> 
```



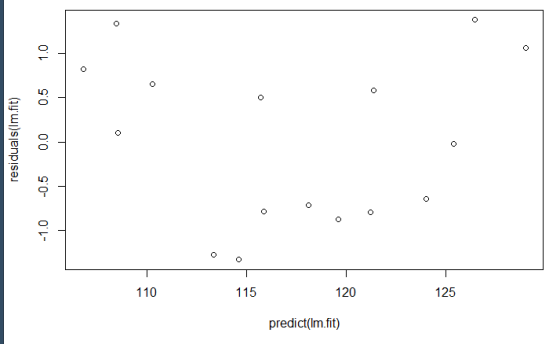

```
> par(nfrow = c(2, 2))
Warning message:
In par(nfrow = c(2, 2)) : "nfrow" is not a graphical parameter
> plot(lm.fit)
Hit <Return> to see next plot: 1
Hit <Return> to see next plot: 2
Hit <Return> to see next plot: 3
Hit <Return> to see next plot: 4
> 
```



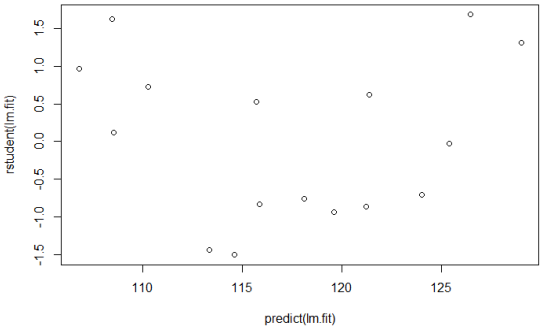
```
> par(nfrow = c(2, 2))
Warning message:
In par(nfrow = c(2, 2)) : "nfrow" is not a graphical parameter
> plot(lm.fit)
Hit <Return> to see next plot: 1
Hit <Return> to see next plot: 2
Hit <Return> to see next plot: 3
Hit <Return> to see next plot: 4
>
```



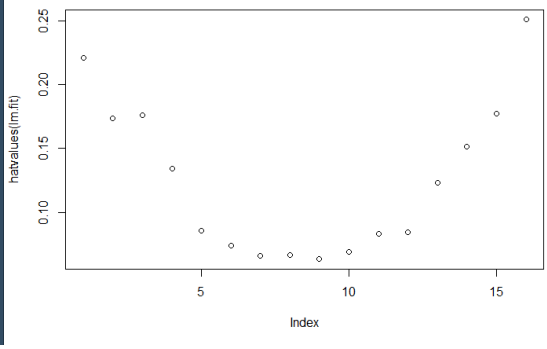

```
> plot(predict(lm.fit), residuals(lm.fit))
>
```




```
> plot(predict(lm.fit), rstudent(lm.fit))
>
```




```
> plot(hatvalues(lm.fit))
> which.max(hatvalues(lm.fit))
16
16
>
```



2. Multiple Linear Regression

```
> lm.fit <- lm(GNP ~ Population + Employed, data = longley)
> summary(lm.fit)

Call:
lm(formula = GNP ~ Population + Employed, data = longley)

Residuals:
     Min       1Q   Median       3Q      Max
-11.6892  -4.5857  -0.6947   3.8580  14.2413

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -1372.0954    36.1406 -37.965 1.05e-14 ***
Population      8.5561     0.9837   8.698 8.85e-07 ***
Employed       11.5606     1.9484   5.933 4.96e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7.385 on 13 degrees of freedom
Multiple R-squared:  0.9952,    Adjusted R-squared:  0.9945
F-statistic:  1352 on 2 and 13 DF,  p-value: 8.297e-16

>
```

```
> lm.fit <- lm(GNP ~ ., data = longley)
> summary(lm.fit)

Call:
lm(formula = GNP ~ ., data = longley)

Residuals:
    Min      1Q  Median      3Q     Max
-3.9488 -1.3209  0.2327  0.7264  5.3511

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  -3.021e+04  9.310e+03  -3.245  0.01007 *
GNP.deflator  1.508e+00  6.190e-01   2.437  0.03757 *
Unemployed   -1.860e-01  4.733e-02  -3.929  0.00347 **
Armed.Forces -5.913e-02  3.252e-02  -1.818  0.10240
Population    4.829e+00  1.388e+00   3.480  0.00694 **
Year          1.543e+01  4.944e+00   3.121  0.01230 *
Employed     -3.148e+00  2.944e+00  -1.070  0.31268
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.858 on 9 degrees of freedom
Multiple R-squared:  0.9995,    Adjusted R-squared:  0.9992
F-statistic:  3022 on 6 and 9 DF,  p-value: 2.409e-14

>
```

```
> library(abind)
> library(car)
> vif(lm.fit)
GNP.deflator   Unemployed Armed.Forces   Population         Year     Employed
   81.946226    35.924858     9.406108   171.158675  1017.609561   196.247880
> lm.fit1 <- lm(GNP ~ . - Armed.Forces, data = longley)
> summary(lm.fit1)

Call:
lm(formula = GNP ~ . - Armed.Forces, data = longley)

Residuals:
    Min      1Q  Median      3Q     Max
-4.5346 -1.4636 -0.4553  1.4473  4.3074

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.651e+04  6.065e+03  -2.723 0.021459 *
GNP.deflator 1.665e+00  6.799e-01   2.449 0.034290 *
Unemployed  -1.085e-01  2.286e-02  -4.745 0.000786 ***
Population   6.442e+00  1.184e+00   5.439 0.000285 ***
Year         8.147e+00  3.214e+00   2.535 0.029614 *
Employed     1.337e+00  1.782e+00   0.750 0.470450
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.17 on 10 degrees of freedom
Multiple R-squared:  0.9993,    Adjusted R-squared:  0.999
F-statistic:  2947 on 5 and 10 DF,  p-value: 1.683e-15

> |
```

```
> library(abind)
> library(car)
> vif(lm.fit)
GNP.deflator   Unemployed Armed.Forces   Population         Year     Employed
   81.946226    35.924858     9.406108   171.158675  1017.609561   196.247880
> lm.fit1 <- lm(GNP ~ . - Armed.Forces, data = longley)
> summary(lm.fit1)

Call:
lm(formula = GNP ~ . - Armed.Forces, data = longley)

Residuals:
    Min      1Q  Median      3Q     Max
-4.5346 -1.4636 -0.4553  1.4473  4.3074

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.651e+04  6.065e+03  -2.723 0.021459 *
GNP.deflator 1.665e+00  6.799e-01   2.449 0.034290 *
Unemployed  -1.085e-01  2.286e-02  -4.745 0.000786 ***
Population   6.442e+00  1.184e+00   5.439 0.000285 ***
Year         8.147e+00  3.214e+00   2.535 0.029614 *
Employed     1.337e+00  1.782e+00   0.750 0.470450
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.17 on 10 degrees of freedom
Multiple R-squared:  0.9993,    Adjusted R-squared:  0.999
F-statistic:  2947 on 5 and 10 DF,  p-value: 1.683e-15

> lm.fit1 <- update(lm.fit, ~ . - Armed.Forces)
> |
```

3. Qualitative Predictors

```
> head(iris)
  Sepal.Length Sepal.Width Petal.Length Petal.Width Species
1          5.1         3.5          1.4         0.2  setosa
2          4.9         3.0          1.4         0.2  setosa
3          4.7         3.2          1.3         0.2  setosa
4          4.6         3.1          1.5         0.2  setosa
5          5.0         3.6          1.4         0.2  setosa
6          5.4         3.9          1.7         0.4  setosa
> lm.fit <- lm(Sepal.Length ~ . + Sepal.Width:Petal.Width, data = iris)
> summary(lm.fit)

Call:
lm(formula = Sepal.Length ~ . + Sepal.Width:Petal.Width, data = iris)

Residuals:
     Min       1Q   Median       3Q      Max
-0.79855 -0.24256  0.02035  0.19395  0.71661

Coefficients:
                       Estimate Std. Error t value Pr(>|t|)
(Intercept)             1.65191    0.43393   3.807 0.000208 ***
Sepal.Width             0.64533    0.12848   5.023  1.5e-06 ***
Petal.Length            0.83663    0.06835  12.240  < 2e-16 ***
Petal.Width             0.22029    0.37469   0.588 0.557510
Speciesversicolor      -0.76953    0.24079  -3.196 0.001716 **
Speciesvirginica       -1.10951    0.33662  -3.296 0.001237 **
Sepal.Width:Petal.Width -0.15891    0.10184  -1.560 0.120889
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3053 on 143 degrees of freedom
Multiple R-squared:  0.8695,    Adjusted R-squared:  0.8641
F-statistic: 158.8 on 6 and 143 DF,  p-value: < 2.2e-16

>
```

```
> attach(iris)
> contrasts(Species)
           versicolor virginica
setosa              0         0
versicolor          1         0
virginica           0         1
>
```

4. Writing Function