|  |  |  |
| --- | --- | --- |
| **Nama :**  **NIM :** ….... | c9824-logouniversitastrisakti  **Praktikum Data Analitik** | **Modul 4A**  **Nama Dosen:** **Syandra Sari, S.Kom, M.Kom** |
| **Hari/Tanggal** :  .., .. Oktober 2021 | **Nama Aslab :**   1. **Tasya Aulia (065001800022)** 2. **Shabrina Irsyadyahwati (065001800023)** |

**Praktikum 4**

**ANALISIS REGRESI SEDERHANA DAN BERGANDA**

**DESKRIPSI MODUL** : Melakukan pemodelan linier sederhana dan berganda.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| No | Elemen Kompetensi | Indikator Kinerja | Jml  Jam | hlm |
| 1 | Mampu melakukan pemodelan linier sederhana dan berganda. | Dapat melakukan pemodelan linier sederhana dan berganda. | 2 |  |

**TEORI SINGKAT**

Dalam praktikum ini akan dipelajari dan dipraktekkan bagaimana melakukan pemodelan linier sederhana dan berganda dengan menggunakan library sebagai berikut :

> library(olsrr)

> library(car)

> library(lmtest)

> library(ggpubr)

#### Regresi Linier Sederhana

Pada regresi linier akan dibicarakan masalah pendugaan atau peramalan sebuah variabel dependen Y dengan sebuah variabel independen X yang telah diketahui nilainya. Model persamaan linier yang digunakan di sini adalah : .

Nilai a dan b dapat diperoleh dari rumus :



dan



**LAB SETUP**

Untuk dapat menjalankan praktikum ini maka yang harus disiapkan adalah :

1. Aplikasi RStudio

2. Xampp

**ELEMEN KOMPETENSI I**

Deskripsi : Dapat melakukan pemodelan linier sederhana dan berganda.

Kompetensi Dasar : Mampu melakukan pemodelan linier sederhana dan berganda.

**PRAKTIKUM**

Dalam praktikum ini akan dipelajari dan dipraktekkan bagaimana melakukan pemodelan linier sederhana dan berganda menggunakan beberapa library dengan menggunakan data houseprices, directmarketing, dan mtcars.

**Data latihan : Houseprices.csv**

|  |
| --- |
| > library(RMySQL)  > con = dbConnect(MySQL(), user = 'root', password = '', dbname =  + 'houseprices', host = 'localhost')  > myQuery <- "select \* from tabel1;"  > df <- dbGetQuery(con, myQuery)  > View(df)  > relasi = lm(df$Price ~ df$SqFt)  > relasi  **OUTPUT** |

**Data latihan : DirectMarketing.csv**

|  |
| --- |
| > library(RMySQL)  > con = dbConnect(MySQL(), user = 'root', password = '', dbname =  + 'directmarketing', host = 'localhost')  > myQuery <- "select \* from tabel1;"  > DirectMarketing <- dbGetQuery(con, myQuery)  > View(DirectMarketin)  > regresi = lm (DirectMarketing$AmountSpent ~ DirectMarketing$Salary)  > regresi  **OUTPUT**    > summary(regresi)  **OUTPUT** |

**Data latihan : mtcars.csv**

|  |
| --- |
| >data(mtcars) #memanggil data mtcars  >View(mtcars) #melihat data mtcars  > plot(mpg ~ wt, data=mtcars) # membuat plot regresi  **OUTPUT** |

|  |
| --- |
| > model <- lm(mpg ~ wt, data=mtcars) #membuat model regresi  > abline(model) # membuat garis pada plot regresi  **OUTPUT** |
| > summary(model) #melihat model regresi  **OUTPUT** |

|  |  |
| --- | --- |
|  |  |
|  | |

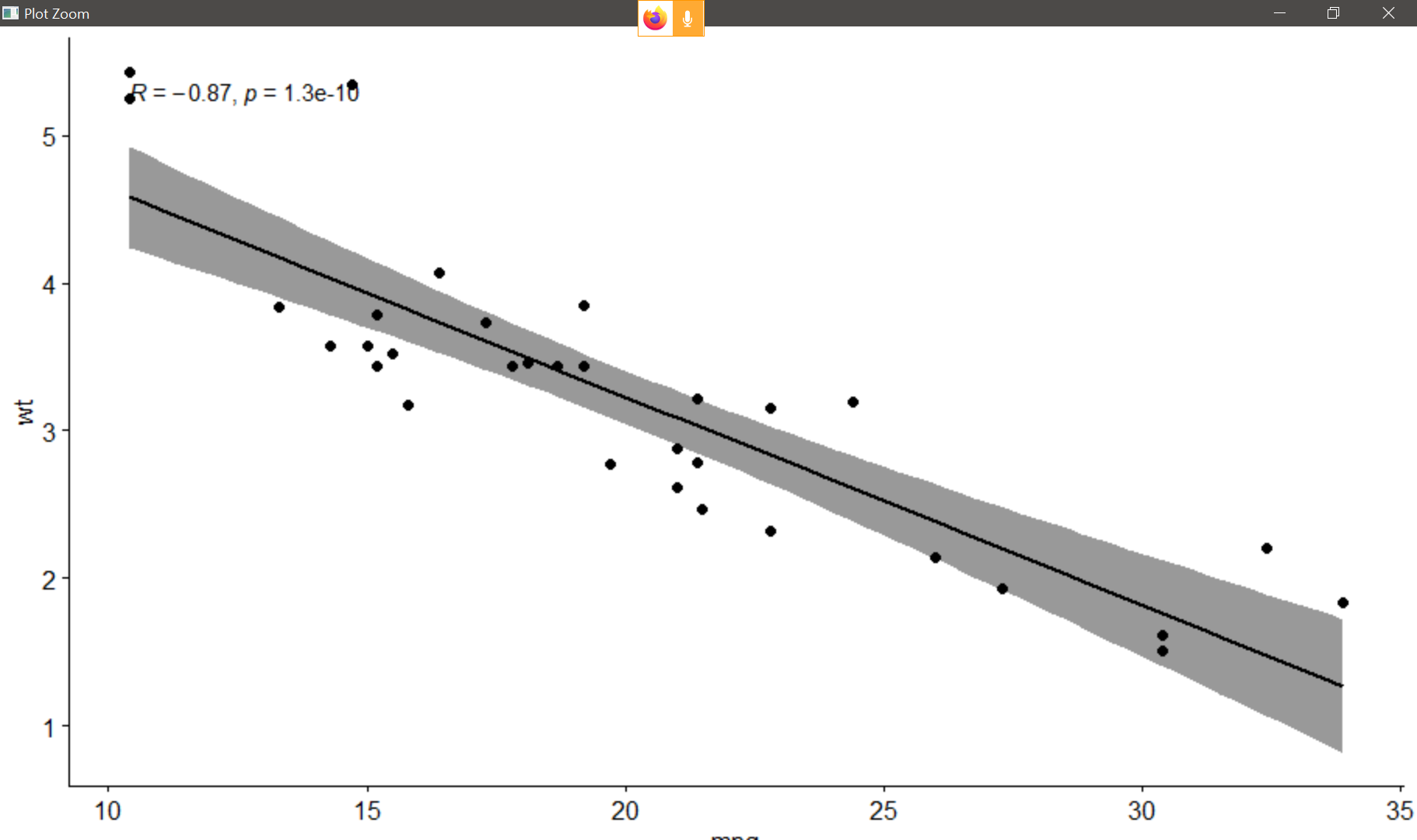
|  |
| --- |
| > predict(model, newdata=data.frame(wt=6)) #misalkan nilai wt adalah 6 maka hasil prediksinya adalah  **OUTPUT**  > predict(model, newdata = data.frame(wt=6))  1  5.218297 |

> ggscatter(mtcars, x = "mpg", y = "wt",

add = "reg.line", conf.int = TRUE,

cor.coef = TRUE, cor.method = "pearson")

**OUTPUT**



#### Regresi Linier Berganda

Jika variabel dependen-nya dihubungkan dengan lebih dari satu variabel independen, maka persamaan yang dihasilkan adalah persamaanregresi linier berganda (*multiple linier regression*). Dalam hal ini kita membatasi pada kasus dua peubah bebas X1 dan X2 saja. Dengan hanya dua peubah bebas, persamaan regresi contohnya menjadi :



Salah satu ukuran kebaikan model adalah dengan melihat koefisien determinasi R2 yang menyatakan proporsi keragaman variabel Y yang dapat dijelaskan oleh variabel X. Namun penggunaan yang lebih baik adalah dengan menggunakan nilai **R-Sq(adj)**, yang merupakan nilai estimasi yang tidak bias (*unbiased estimate*) dari populasi.

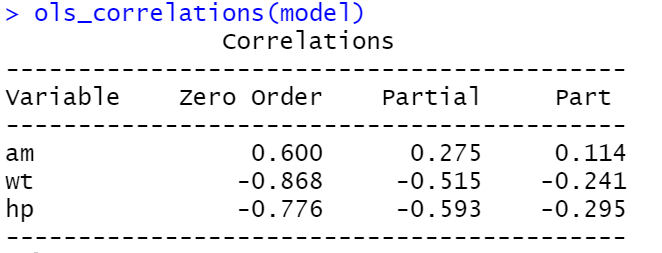
|  |
| --- |
| > data("mtcars")  > model=lm(mpg ~ am + wt + hp, data = mtcars)  > summary(model)  **OUTPUT** |

**Pada output diatas tedapat variabel yang tidak signifikan…maka variabel tersebut jangan langsung dibuang. Lakukanlah explorasi data.**

|  |
| --- |
| > hist(mtcars$am)  > plot(mpg~am,data=mtcars) |
| **OUTPUT**    Melihat Korelasi Antar Variabel  > dataku=mtcars[,c(1,4,6,9)]  > cor(dataku,method = "pearson")  **OUTPUT** |

> ols\_correlations(model)

**OUTPUT**



**Zero Order**

Pearson correlation coefficient between the dependent variable and the independent variables.

**Part**

Unique contribution of independent variables. How much R2R2 will decrease if that variable is removed from the model?

**Partial**

How much of the variance in **Y**, which is not estimated by the other independent variables in the model, is estimated by the specific variable?

**ELEMEN KOMPETENSI II**

Deskripsi : Dapat melakukan pemodelan linier sederhana dan berganda.

Kompetensi Dasar : Mampu melakukan pemodelan linier sederhana dan berganda.

**TUGAS**

Dalam praktikum ini akan dipelajari dan dipraktekkan bagaimana melakukan pemodelan linier sederhana dan berganda menggunakan beberapa library dengan menggunakan data advertising.

*Data is about advertising data sales (in thousands of units) for a particular product advertising budgets (in thousands of dollars) for TV, radio, and newspaper media. On the basis of this data, suggest a marketing plan for next year that will result in high product sales using regression analysis.*

https://rstudio-pubs static.s3.amazonaws.com/249959\_d71491a56f8242909331dfee0e25b813.html

### Linear Regression

> library(MASS)

> library(ISLR)

> library(RMySQL)

> con = dbConnect(MySQL(), user = 'root', password = '', dbname =

+ 'Advertising', host = 'localhost')

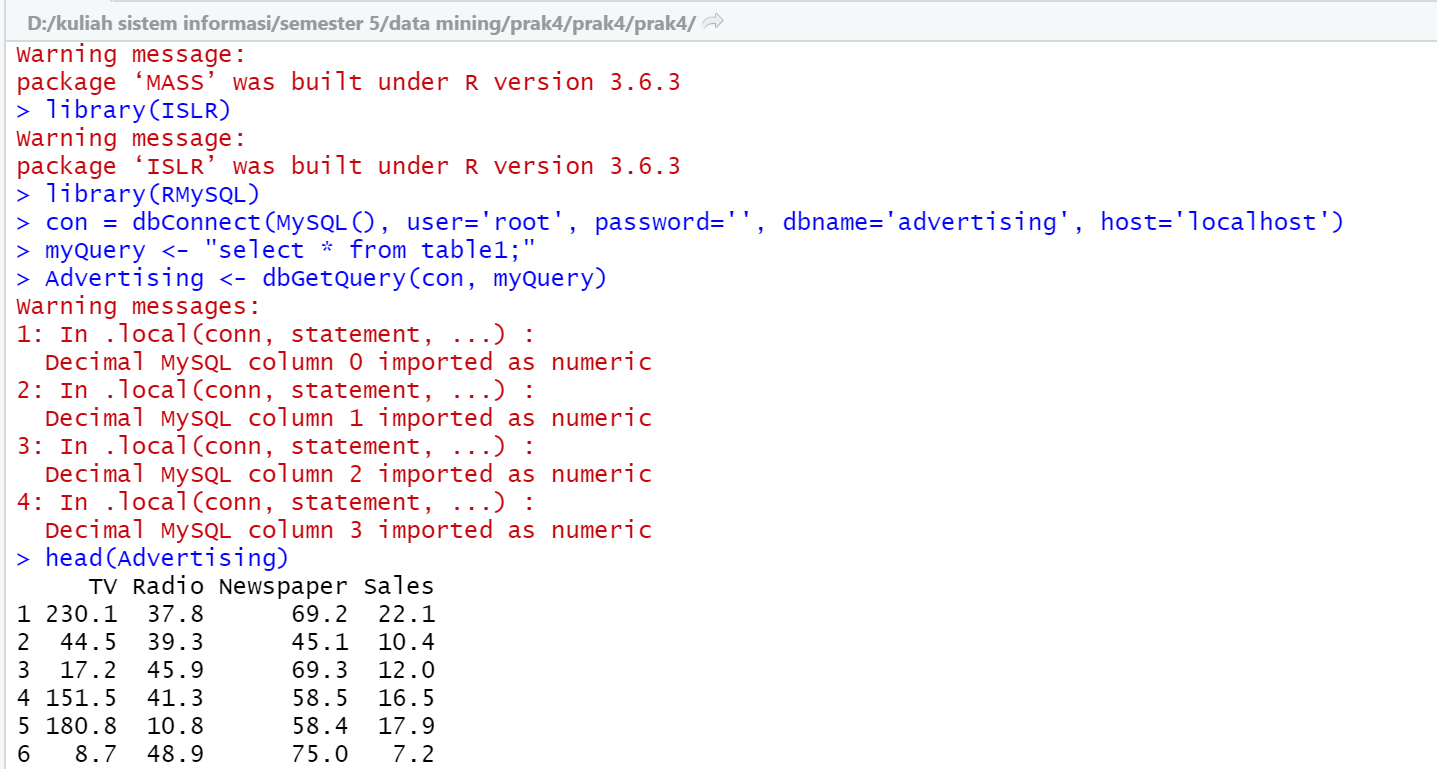
>

> myQuery <- "select \* from tabel1;"

> Advertising <- dbGetQuery(con, myQuery)

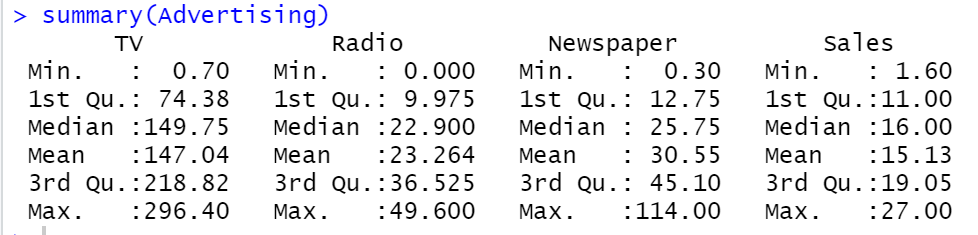
> head(Advertising)

**OUTPUT**



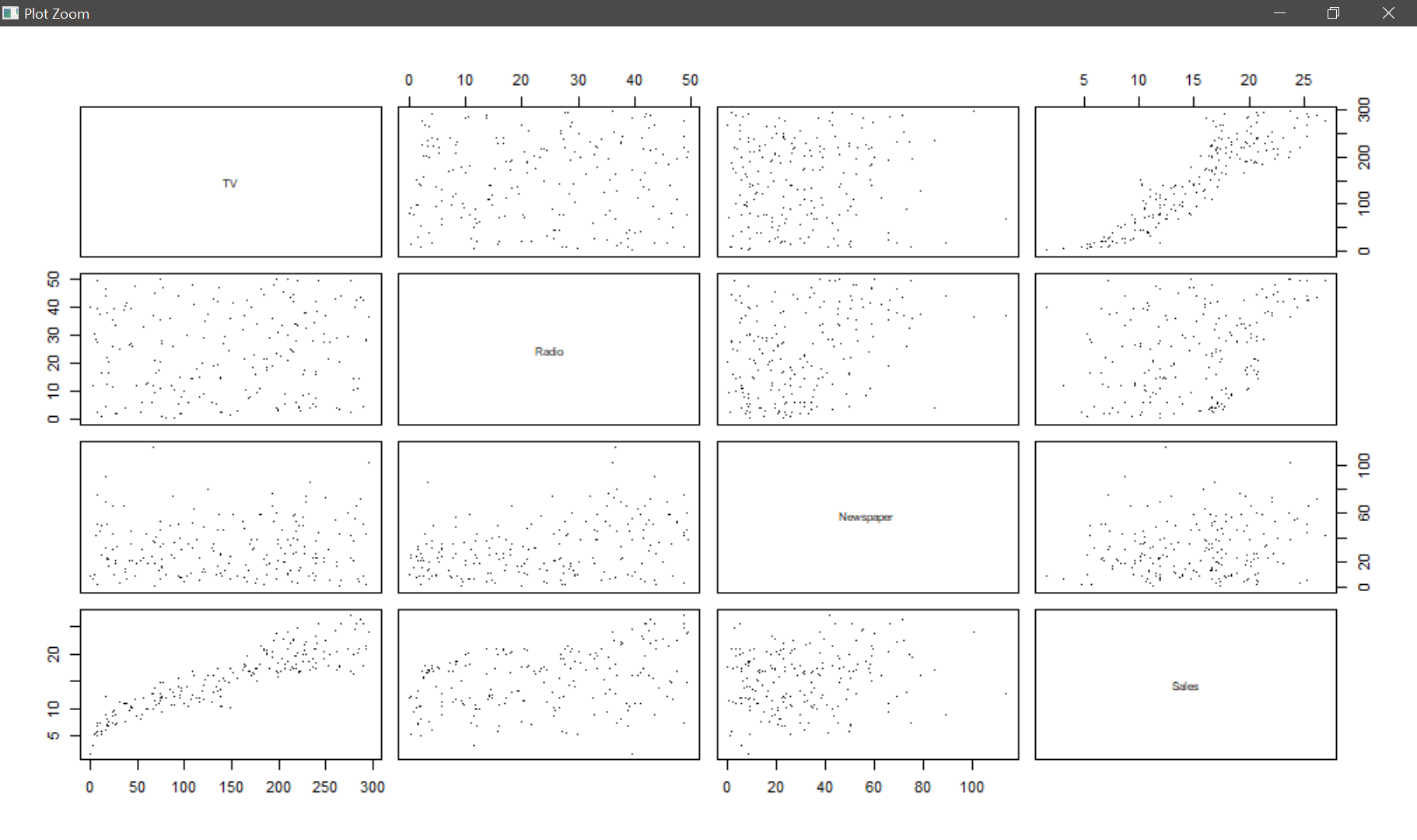
> summary(Advertising)

**OUTPUT**



> pairs(Advertising, pch=".")

**OUTPUT**

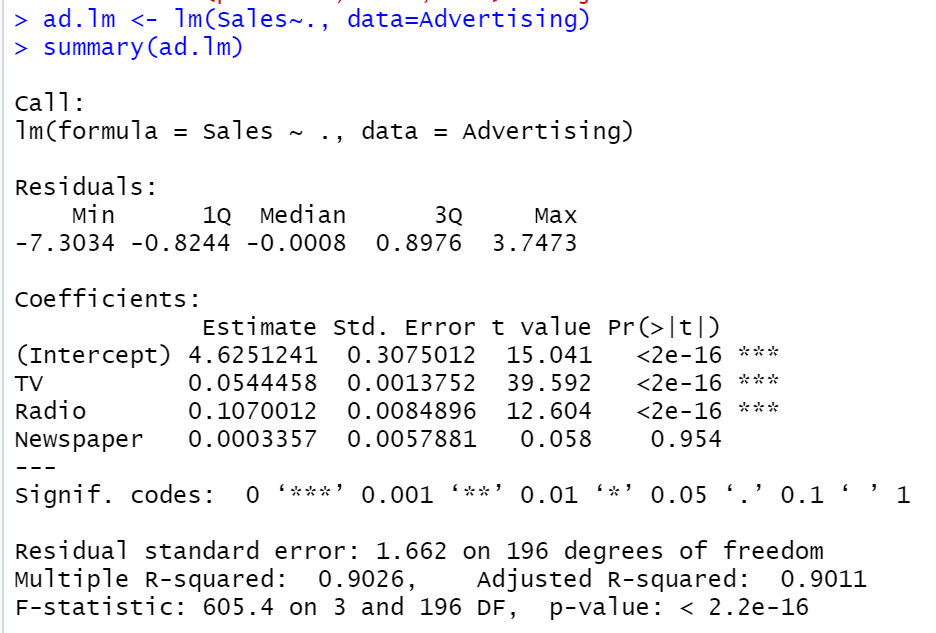


### Is there a relationship between advertising sales and budget?

> ad.lm <- lm(sales~., data=Advertising)

> summary(ad.lm)

**OUTPUT**



### How strong is the relationship?

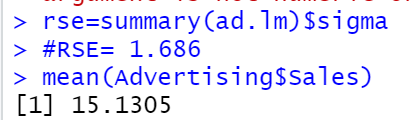
> rse=summary(ad.lm)$sigma

>

> #RSE= 1.686

> mean(Advertising$sales)

**OUTPUT**



> rse/mean(Advertising$sales)

**OUTPUT**

[1] 0.1098242

> rsq=summary(ad.lm)$r.sq

> rsq #0.8972106

**OUTPUT**

[1] 0.9025913

> # rsq is calculated by the following formuala

> yhat=ad.lm$fitted.values #predicted

> y=Advertising$Sales #observed

> rsq=1-sum((y-yhat)^2)/sum((y-mean(y))^2) #orginal formula

Warning message:

In mean.default(y) : argument is not numeric or logical: returning NA

>

>

> #Other way to get R2

> var(yhat)/var(y) #other formula

Error in var(y) : 'x' is NULL

> # rsq is calculated by the following formuala

> yhat=ad.lm$fitted.values #predicted

> y=Advertising$sales #observed

> rsq=1-sum((y-yhat)^2)/sum((y-mean(y))^2) #orginal formula

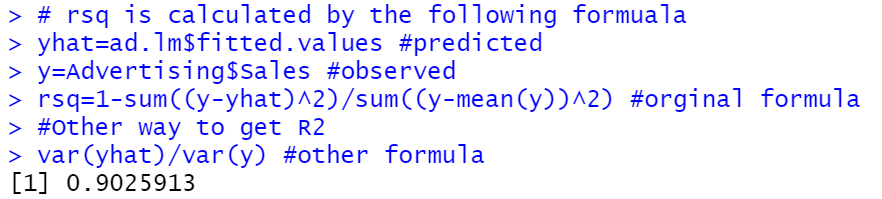
>

>

> #Other way to get R2

> var(yhat)/var(y) #other formula

**OUTPUT**



> 1-sum((y-yhat)^2)/sum((y-mean(y))^2) #orginal formula

**OUTPUT**

…

> cor(yhat,y)^2 #alternate formula

**OUTPUT**

…

> Coef1=summary(ad.lm)$coefficients #Coefficient matrix

> Coef1

**OUTPUT**

Estimate Std. Error t value Pr(>|t|)

(Intercept) 4.6251240788 0.307501165 15.04099695 1.682677e-34

TV 0.0544457803 0.001375188 39.59152448 1.892945e-95

Radio 0.1070012282 0.008489563 12.60385655 4.602097e-27

Newspaper 0.0003356579 0.005788056 0.05799148 9.538145e-01

How large is the effect of each medium on sales?

> lolim=Coef1[,1] - 1.96\*Coef1[,2]

> uplim=Coef1[,1] + 1.96\*Coef1[,2]

> cbind(lolim,uplim)

**OUTPUT**

lolim uplim

(Intercept) 4.02242180 5.22782636

TV 0.05175041 0.05714115

Radio 0.09036169 0.12364077

Newspaper -0.01100893 0.01168025

> confint(ad.lm)

**OUTPUT**

2.5 % 97.5 %

(Intercept) 4.01868836 5.23155980

TV 0.05173372 0.05715784

Radio 0.09025861 0.12374384

Newspaper -0.01107921 0.01175052

> require(car)

Loading required package: car

Loading required package: carData

Attaching package: ‘car’

The following object is masked from ‘package:arules’:

recode

Warning messages:

1: package ‘car’ was built under R version 3.4.4

2: package ‘carData’ was built under R version 3.4.4

> vif(ad.lm)

**output**

TV Radio Newspaper

1.004611 1.144952 1.145187

### How accurately can we predict future sales?

> #for the average response f(X)

> predict(ad.lm, newdata=data.frame(TV=149,radio=22,newspaper=25),

+ interval="confidence")

**OUTPUT**

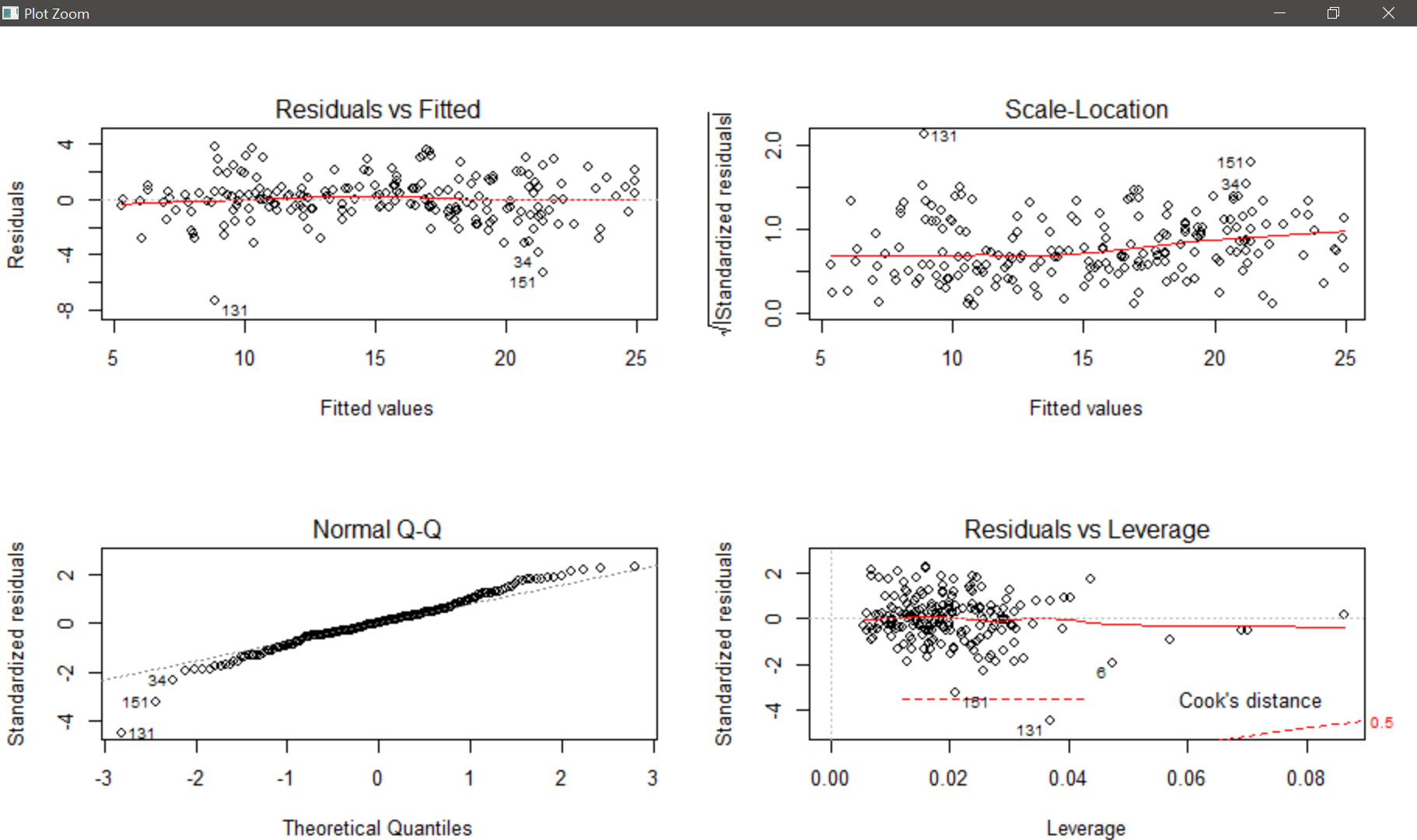
fit lwr upr

1 15.09996 14.86063 15.3393

#### Is the relationship linear?

> plot(ad.lm) #diagnostic plot

**OUTPUT**

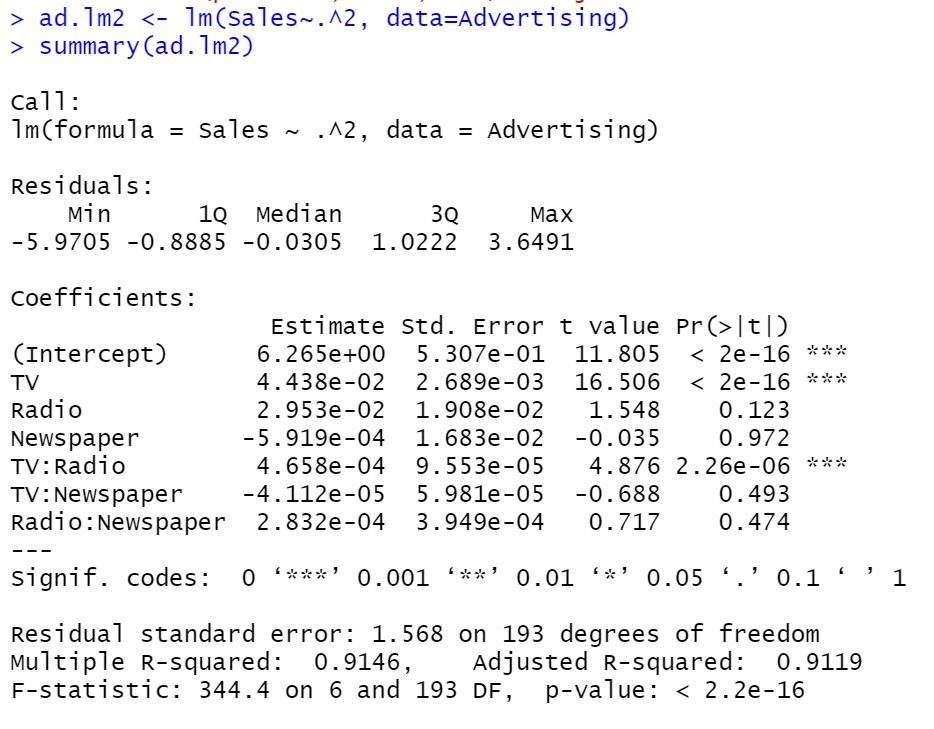


##### Is there synergy among the advertising media?

> ad.lm2 <- lm(sales~.^2, data=Advertising)

> summary(ad.lm2)

**OUTPUT**



# the Advertising data may not be additive.

> summary(ad.lm2)$r.sq;summary(ad.lm)$r.sq

**OUTPUT**

[1] 0.914582

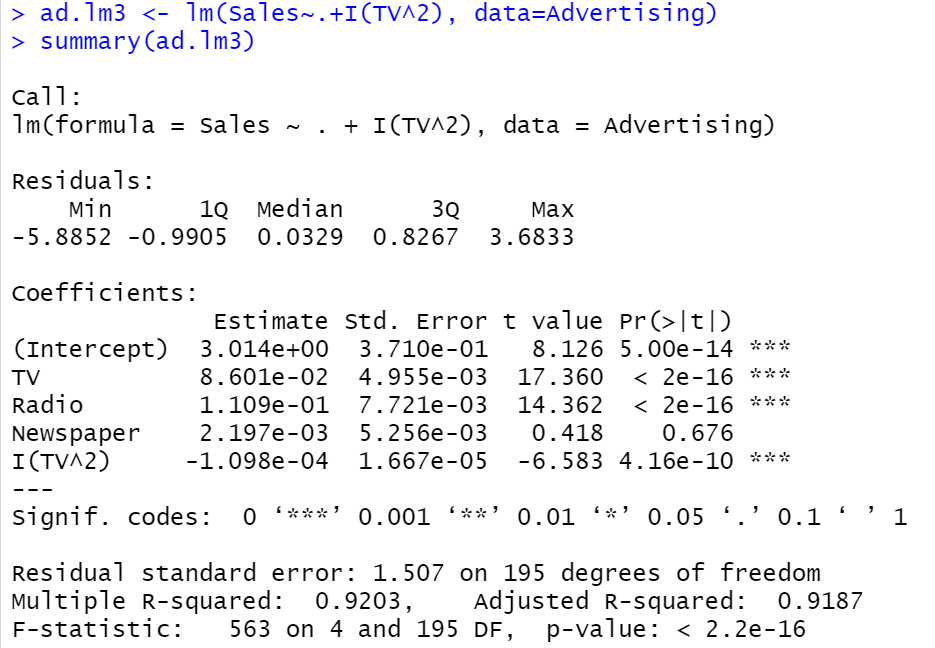
[1] 0.9025913

### Non-linear Transformations of the Predictors

> ad.lm3 <- lm(sales~.+I(TV^2), data=Advertising)

> summary(ad.lm3)

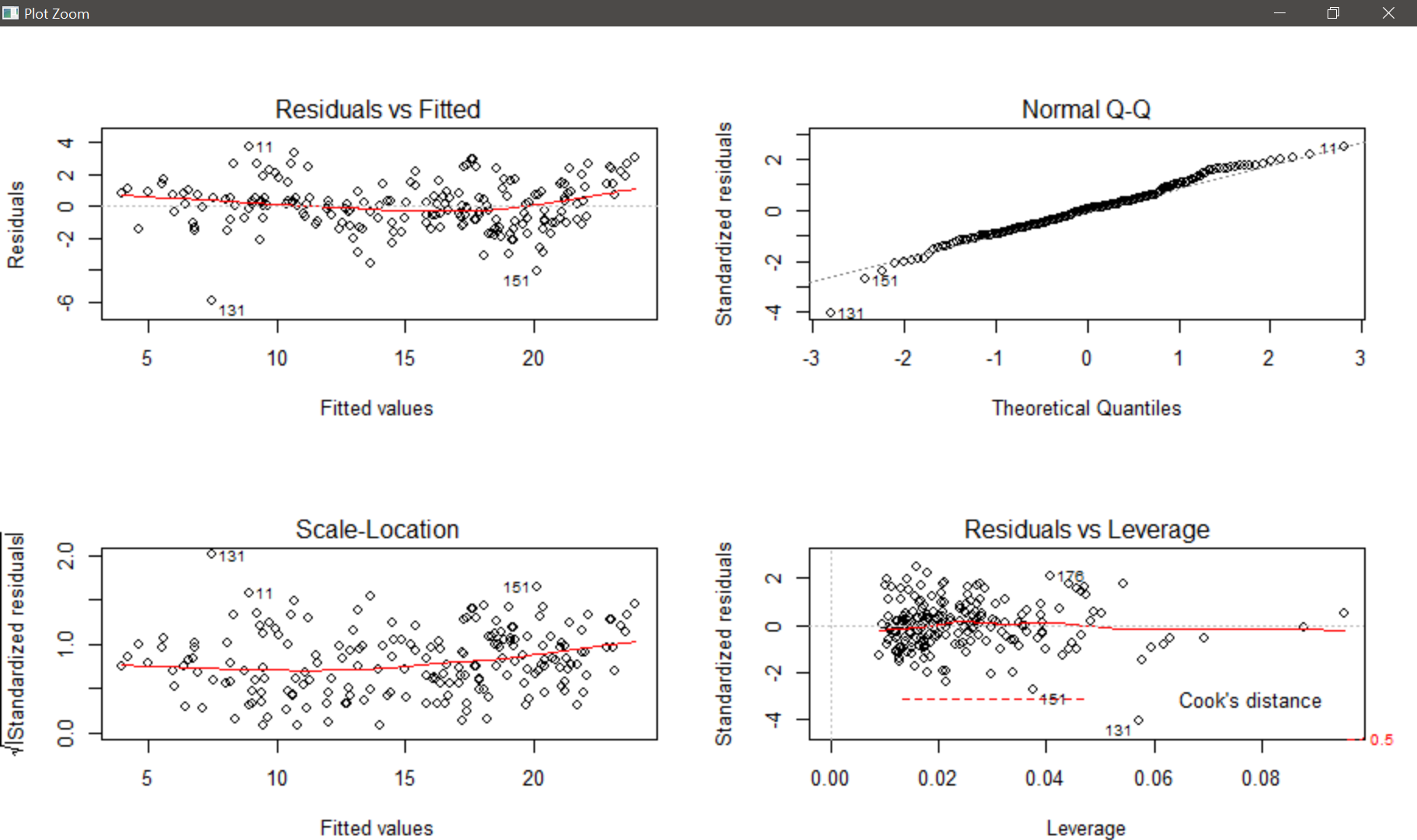
**OUTPUT**



> par(mfrow=c(2,2))

> plot(ad.lm3)

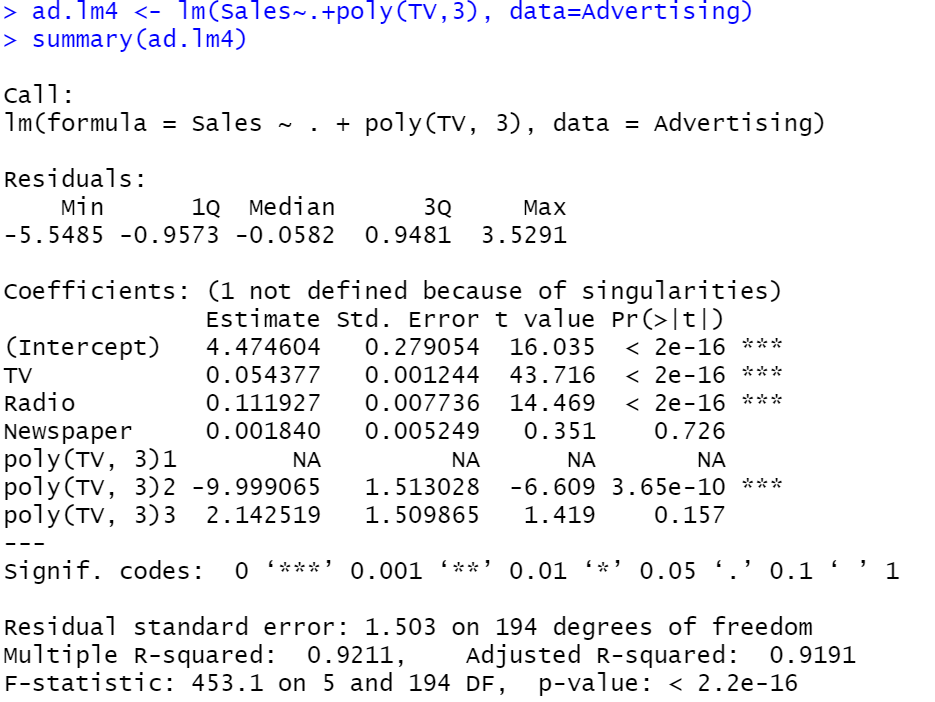
**OUTPUT**



> ad.lm4 <- lm(sales~.+poly(TV,3), data=Advertising)

> summary(ad.lm4)

**OUTPUT**



> anova(ad.lm,ad.lm4)

**OUTPUT**

Analysis of Variance Table

Model 1: Sales ~ TV + Radio + Newspaper

Model 2: Sales ~ TV + Radio + Newspaper + poly(TV, 3)

Res.Df RSS Df Sum of Sq F Pr(>F)

1 196 541.20

2 194 438.24 2 102.96 22.79 1.289e-09 \*\*\*

---

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

> anova(ad.lm3,ad.lm4)

**OUTPUT**

Analysis of Variance Table

Model 1: Sales ~ TV + Radio + Newspaper + I(TV^2)

Model 2: Sales ~ TV + Radio + Newspaper + poly(TV, 3)

Res.Df RSS Df Sum of Sq F Pr(>F)

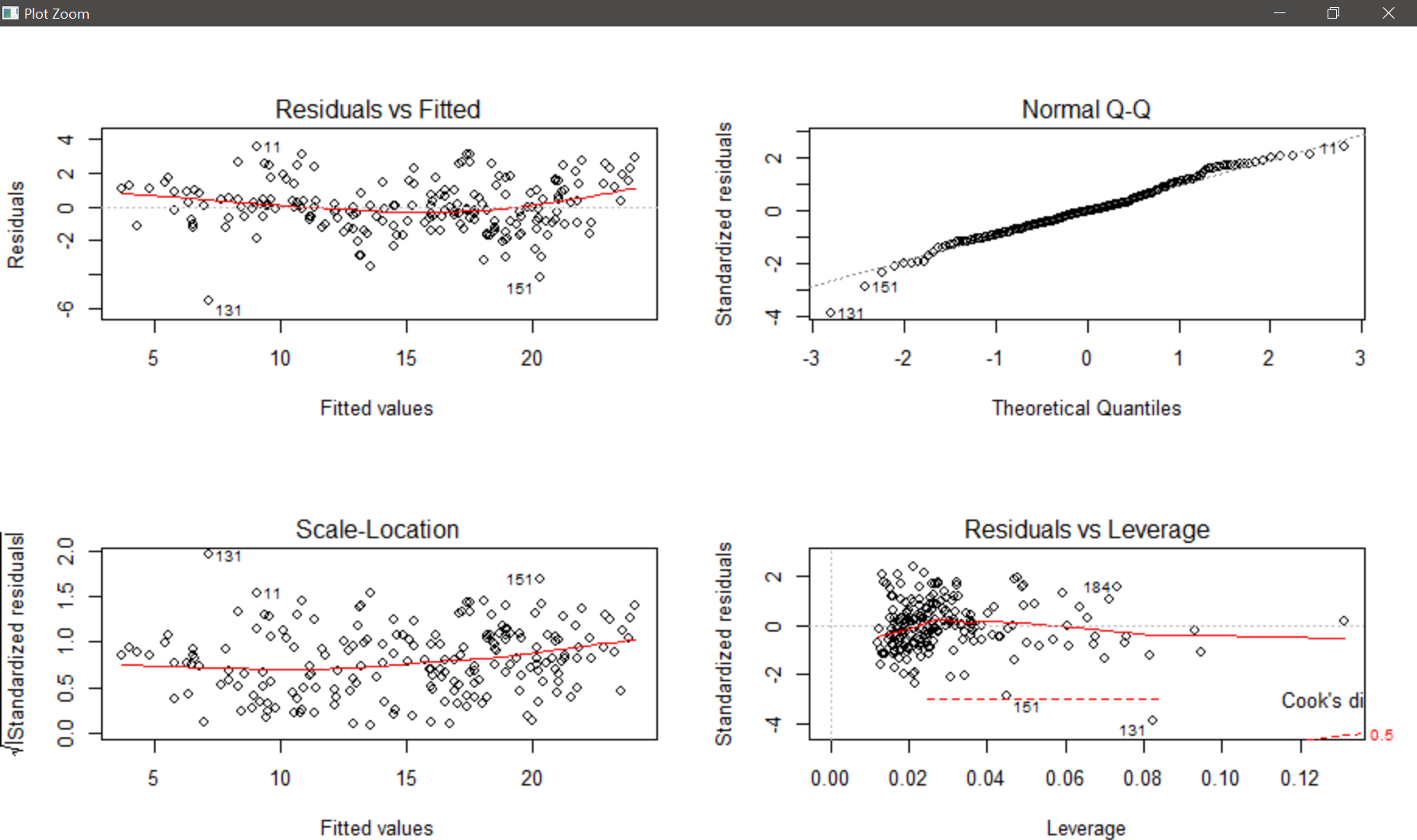
1 195 442.79

2 194 438.24 1 4.5486 2.0136 0.1575

> par(mfrow=c(2,2))

> plot(ad.lm4)

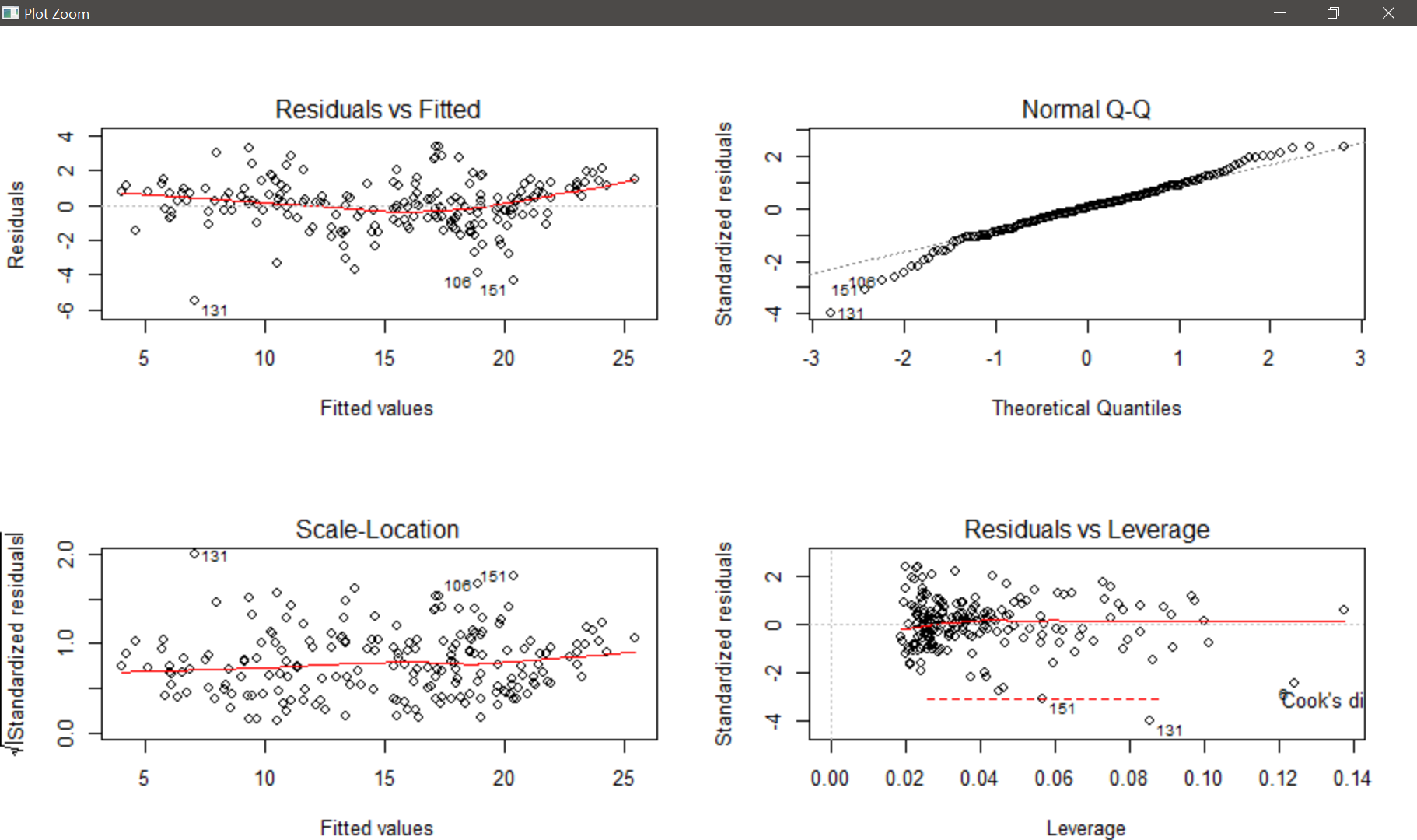
**OUTPUT**



> ad.lm5 <- lm(sales~.+poly(TV,3)+poly(radio,3), data=Advertising)

> plot(ad.lm5)

**OUTPUT**



> anova(ad.lm4,ad.lm5)

**OUTPUT**

Analysis of Variance Table

Model 1: Sales ~ TV + Radio + Newspaper + poly(TV, 3)

Model 2: Sales ~ TV + Radio + Newspaper + poly(TV, 3) + poly(Radio, 3)

Res.Df RSS Df Sum of Sq F Pr(>F)

1 194 438.24

2 192 398.33 2 39.91 9.6185 0.0001045 \*\*\*

---

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

**Output:**

1. **Cek List**

|  |  |  |  |
| --- | --- | --- | --- |
| **No** | **Elemen Kompetensi** | **Penyelesaian** | |
| **Selesai** | **Tidak** |
| **1** | **Elemen Kompetensi I**  Dapat melakukan pemodelan linier sederhana dan berganda menggunakan beberapa library dengan menggunakan data houseprices, directmarketing, dan mtcars. |  |  |
| **2** | **Elemen Kompetensi II**  Dapat melakukan pemodelan linier sederhana dan berganda menggunakan beberapa library dengan menggunakan data advertising. |  |  |

1. **Form Umpan Balik**

|  |  |  |
| --- | --- | --- |
| **Elemen Kompetensi** | **Waktu Pengerjaan** | **Kriteria** |
| **Elemen Kompetensi I**  Dapat melakukan pemodelan linier sederhana dan berganda menggunakan beberapa library dengan menggunakan data houseprices, directmarketing, dan mtcars. |  |  |
| **Elemen Kompetensi II**  Dapat melakukan pemodelan linier sederhana dan berganda menggunakan beberapa library dengan menggunakan data advertising. |  |  |

Kriteria

1.Sangat Menarik

2.Cukup Menarik

3.Kurang Menarik

4.Sangat Kurang Menarik