

SalesPeople

Kenneth Manuel (160419041), Jehuda Rivaldo (160419133)

4/25/2021

```
library("readxl")
library("corrplot")
library("ggplot2")
library("ggfortify")
```

Data Preparation

```
data <- read_excel('Salespeople-data.xlsx')
head(data)
```

```
## # A tibble: 6 x 7
##   Salegrow saleproft Newsale createst Mechtest absttest mathtest
##   <dbl>      <dbl>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>
## 1      93        96      97.8        9      12        9      20
## 2     88.8      91.8      96.8        7      10       10      15
## 3      95       100.        99        8      12        9      26
## 4     101.      104.      107.       13      14       12      29
## 5     102       108.      103        10      15       12      32
## 6     95.8      97.5      99.3        10      14       11      21
```

```
summary(data)
```

```
##           Salegrow           saleproft           Newsale           createst
##   Min.   : 81.50   Min.   : 87.3   Min.   : 94.30   Min.   : 1.00
##   1st Qu.: 94.42   1st Qu.: 99.5   1st Qu.: 99.08   1st Qu.: 8.25
##   Median :100.65   Median :106.2   Median :103.15   Median :10.00
##   Mean   : 98.96   Mean   :106.6   Mean   :102.81   Mean   :11.22
##   3rd Qu.:105.05   3rd Qu.:114.8   3rd Qu.:106.45   3rd Qu.:14.00
##   Max.    :110.80   Max.    :122.3   Max.    :115.30   Max.    :18.00
##           Mechtest           absttest           mathtest
##   Min.   : 5.00   Min.   : 5.00   Min.   : 9.00
##   1st Qu.:12.00   1st Qu.: 9.00   1st Qu.:21.50
##   Median :15.00   Median :11.00   Median :31.50
##   Mean   :14.18   Mean   :10.56   Mean   :29.76
##   3rd Qu.:17.00   3rd Qu.:12.00   3rd Qu.:37.00
##   Max.    :20.00   Max.    :15.00   Max.    :51.00
```

Gunakan variabel independen

```
data_ip = data[4:7]
head(data_ip)
```

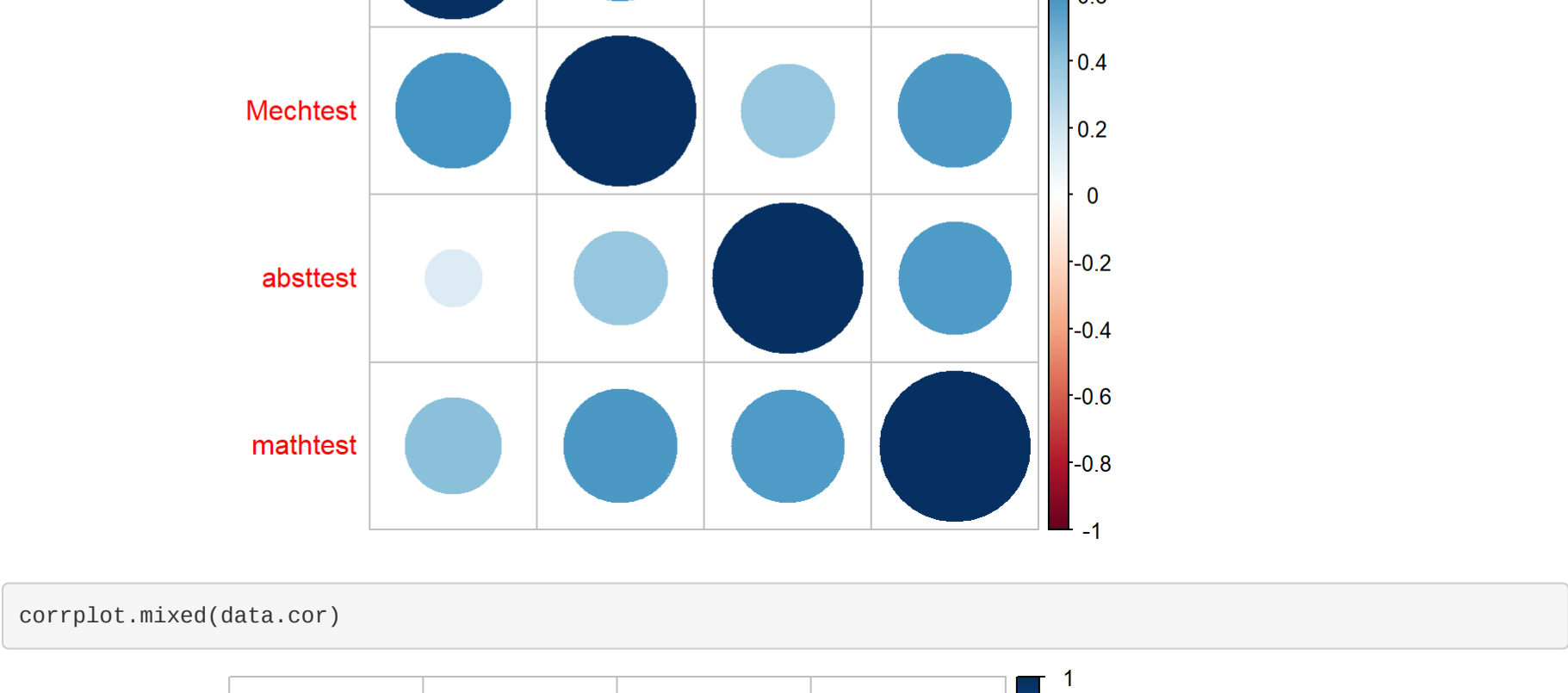
```
## # A tibble: 6 x 4
##   createst Mechtest absttest mathtest
##   <dbl>      <dbl>    <dbl>    <dbl>
## 1      9      12        9      20
## 2      7      10       10      15
## 3      8      12        9      26
## 4     13      14       12      29
## 5     10      15       12      32
## 6     10      14       11      21
```

Melihat *correlation plot* sepertinya semua variabel independen agak berkorelasi kecuali abstract test dan creative test.

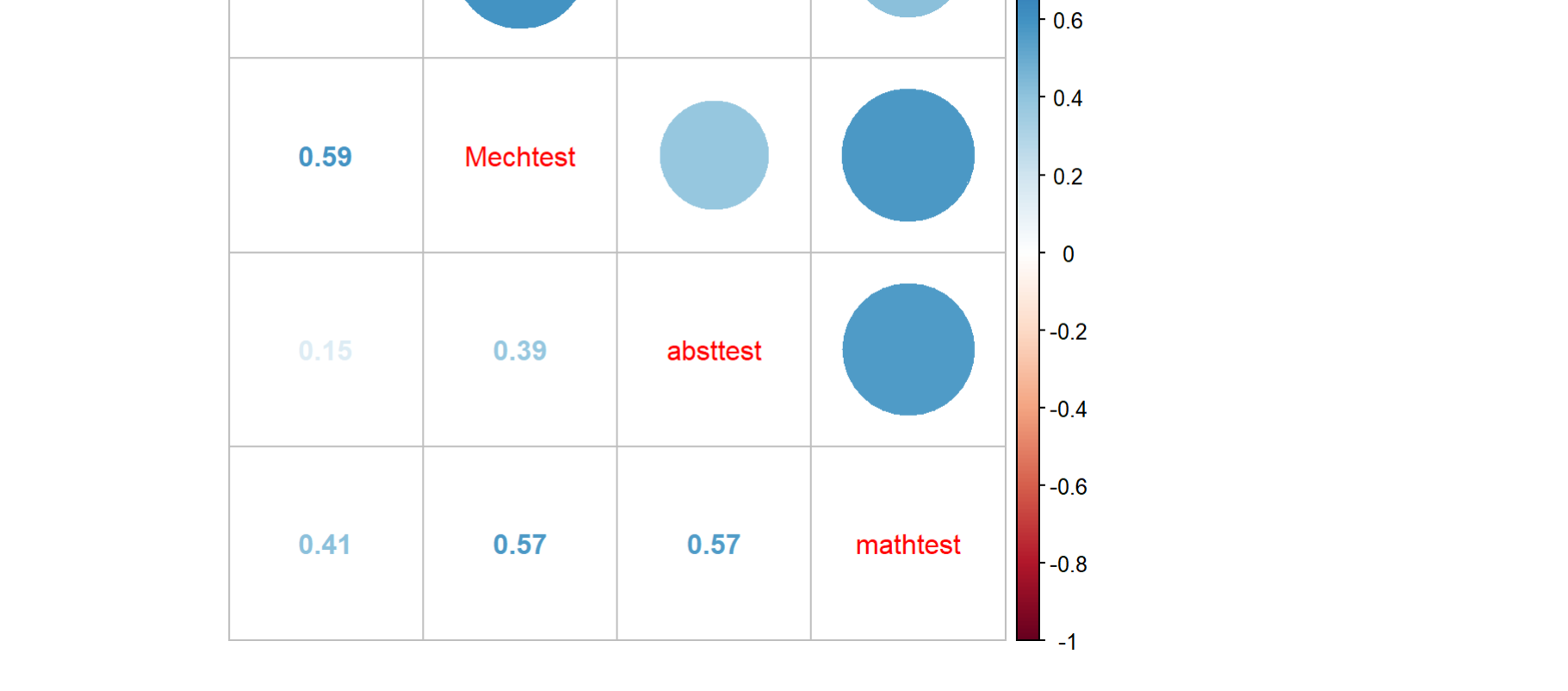
```
# Matrix correlation
data.cor = cor(data_ip[,])
data.cor
```

```
##           createst Mechtest absttest mathtest
## createst 1.0000000 0.5907360 0.1469074 0.4126395
## Mechtest 0.5907360 1.0000000 0.3859502 0.5745533
## absttest 0.1469074 0.3859502 1.0000000 0.5663721
## mathtest 0.4126395 0.5745533 0.5663721 1.0000000
```

```
corrplot(data.cor)
```



```
corrplot.mixed(data.cor)
```



Principal Component Analysis

Melihat visualisasi dan kesimpulan yang lebih ringkas dengan PCA

```
fit <- prcomp(data_ip[,], scale=TRUE)
fit
```

```
## Standard deviations (1, .., p=4):
## [1] 1.5368630 0.9583733 0.6036777 0.5959413
##
## Rotation (n x k) = (4 x 4):
##           PC1      PC2      PC3      PC4
## createst 0.4526830 0.6434339 -0.2055992 0.5820652
## Mechtest 0.5514235 0.2466951 -0.1987056 -0.7717446
## absttest 0.4339023 -0.6850154 -0.5382775 0.2296521
## mathtest 0.5502173 -0.2364071 0.7927814 0.1134470
```

Dari masing-masing **Principal Component** dapat kita lihat bahwa 59% data dijelaskan dalam PC1, 23% data dijelaskan dalam PC2, 9% dijelaskan dalam PC3, dan 8% dijelaskan dalam PC4.

```
summary(fit)
```

```
## Importance of components:
##           PC1      PC2      PC3      PC4
## Standard deviation 1.5369 0.9584 0.6036 0.5959
## Proportion of Variance 0.5905 0.2296 0.0911 0.0887
## Cumulative Proportion 0.5905 0.8201 0.9112 1.0000
```

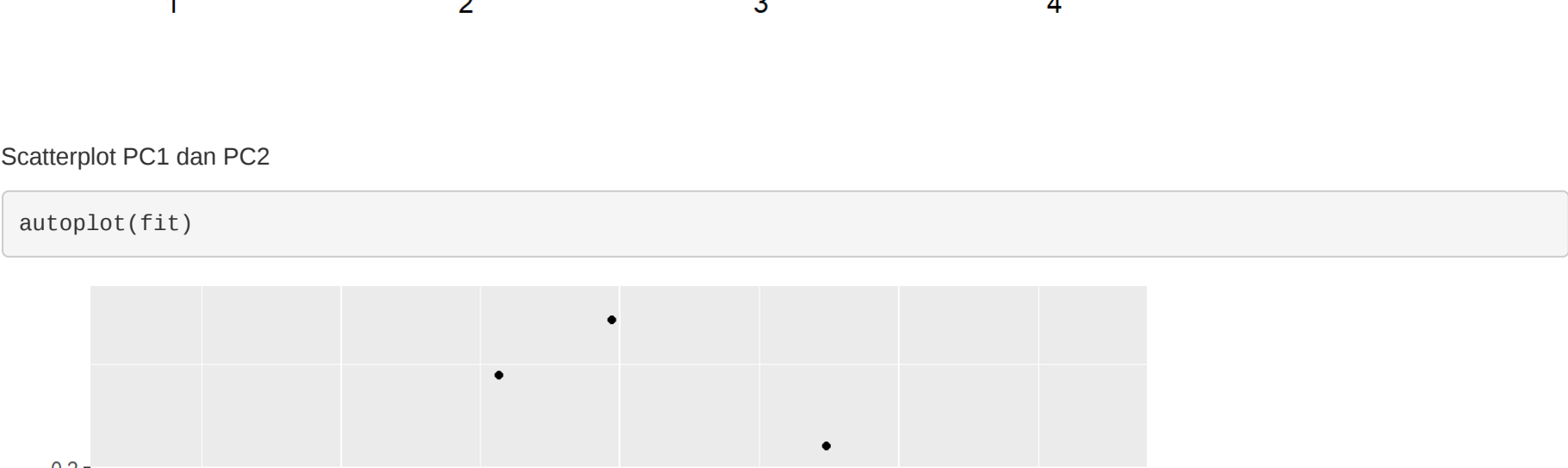
Elbow pada screeplot terletak pada PC2 PC1 dan PC2 akan digunakan untuk analisis ini. (82% data dijelaskan)

```
plot(fit, type="l")
```



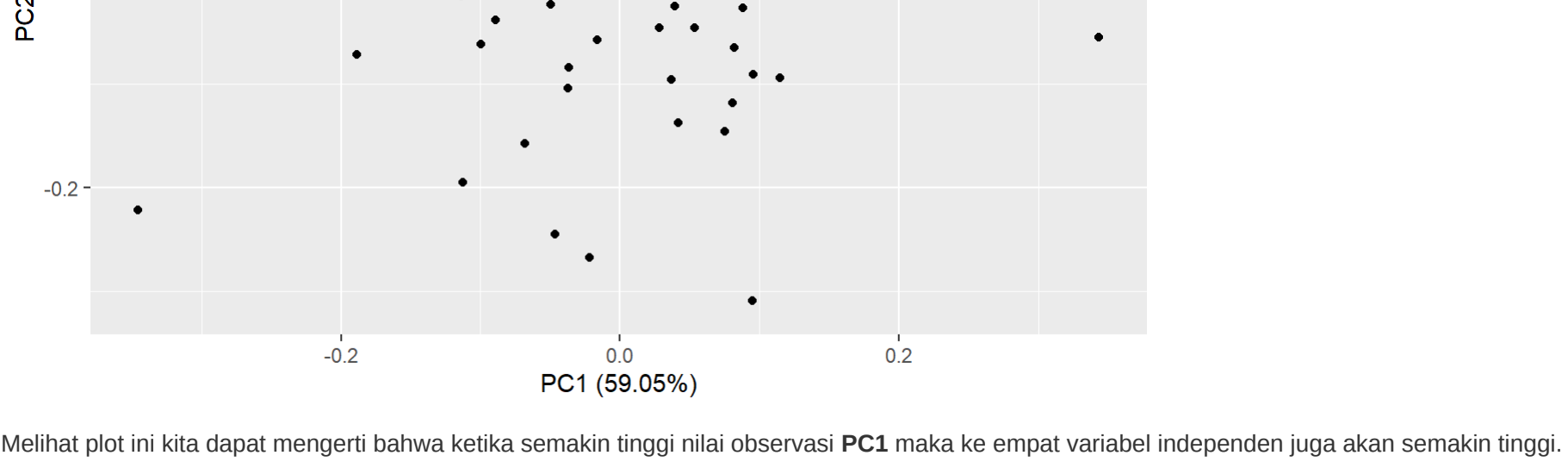
Scatterplot PC1 dan PC2

```
autoplot(fit)
```



Melihat plot ini kita dapat mengerti bahwa ketika semakin tinggi nilai observasi **PC1** maka ke empat variabel independen juga akan semakin tinggi. Untuk observasi nilai **PC2**, semakin tinggi nilai PC2 maka nilai *Creativity Test* dan *Mechanical Test* semakin tinggi, sedangkan semakin kecil nilai PC2 maka nilai *Math Test* dan *Abstract Test* akan semakin rendah.

```
biplot(fit)
```



Model Regresi

Koefisien regresi **salesgrow**

```
lm(data$Salegrow~data$createst+data$Mechtest+data$absttest+data$mathtest)
```

```
##
## Call:
## lm(formula = data$Salegrow ~ data$createst + data$Mechtest +
##   data$absttest + data$mathtest)
##
## Coefficients:
##   (Intercept) data$createst data$Mechtest data$absttest data$mathtest
##      69.8770      0.3292      0.3591      0.6708      0.4439
```

- Nilai uji statistik F bernilai 177.4 dengan p-value < 0.05 tersebut menunjukkan bahwa keempat variabel independen sangat berpengaruh terhadap **salesgrow**.
- Nilai R-squared 0.9351 menunjukkan bahwa keempat variabel independen menjelaskan 93% variasi variabel **salesgrow**

```
summary(lm(data$Salegrow~data$createst+data$Mechtest+data$absttest+data$mathtest))
```

```
##
## Call:
## lm(formula = data$Salegrow ~ data$createst + data$Mechtest +
##   data$absttest + data$mathtest)
##
## Residuals:
##   Min       1Q   Median       3Q      Max
## -4.8845 -0.9712  0.4010  1.1489  5.8399
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  69.87696   1.55502   44.936 < 2e-16 ***
## data$createst  0.32921   0.08522   3.863 0.000356 ***
## data$Mechtest  0.35906   0.11039   3.253 0.002172 ***
## data$absttest  0.67077   0.15371   4.364 7.38e-05 ***
## data$mathtest  0.44390   0.03518  12.616 2.22e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.859 on 45 degrees of freedom
## Multiple R-squared:  0.9404, Adjusted R-squared:  0.9351
## F-statistic: 177.4 on 4 and 45 DF, p-value: < 2.2e-16
```

Koefisien regresi **saleproft**

```
lm(data$Saleproft~data$createst+data$Mechtest+data$absttest+data$mathtest)
```

```
##
## Call:
## lm(formula = data$Saleproft ~ data$createst + data$Mechtest +
##   data$absttest + data$mathtest)
##
## Coefficients:
##   (Intercept) data$createst data$Mechtest data$absttest data$mathtest
##      75.4198      0.1195      0.8674      -0.5778      0.7951
```

- Nilai uji statistik F bernilai 321.9 dengan p-value < 0.05 tersebut menunjukkan bahwa keempat variabel independen sangat berpengaruh terhadap **saleproft**.
- Nilai R-squared 0.9632 menunjukkan bahwa keempat variabel independen menjelaskan 96% variasi variabel **saleproft**

```
summary(lm(data$Saleproft~data$createst+data$Mechtest+data$absttest+data$mathtest))
```

```
##
## Call:
## lm(formula = data$Saleproft ~ data$createst + data$Mechtest +
##   data$absttest + data$mathtest)
##
## Residuals:
##   Min       1Q   Median       3Q      Max
## -4.8038 -1.3943 -0.2561  1.0269  3.9511
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  75.41979   1.62433   46.431 < 2e-16 ***
## data$createst  0.11954   0.08902   1.343 0.186046
## data$Mechtest  0.86739   0.11531   7.522 1.73e-09 ***
## data$absttest -0.57782   0.16056  -3.599 0.000792 ***
## data$mathtest  0.79513   0.03675  21.634 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.941 on 45 degrees of freedom
## Multiple R-squared:  0.9662, Adjusted R-squared:  0.9632
## F-statistic: 321.9 on 4 and 45 DF, p-value: < 2.2e-16
```

Koefisien regresi **newsale**

```
lm(data$Newsale~data$createst+data$Mechtest+data$absttest+data$mathtest)
```

```
##
## Call:
## lm(formula = data$Newsale ~ data$createst + data$Mechtest + data$absttest +
##   data$mathtest)
##
## Coefficients:
##   (Intercept) data$createst data$Mechtest data$absttest data$mathtest
##      83.70818      0.56792      -0.08071      0.65839      0.23258
```

- Nilai uji statistik F bernilai 321.9 dengan p-value < 0.05 tersebut menunjukkan bahwa keempat variabel independen sangat berpengaruh terhadap **newsale**.
- Nilai R-squared 0.9255 menunjukkan bahwa keempat variabel independen menjelaskan 92% variasi variabel **newsale**

```
summary(lm(data$Newsale~data$createst+data$Mechtest+data$absttest+data$mathtest))
```

```
##
## Call:
## lm(formula = data$Newsale ~ data$createst + data$Mechtest + data$absttest +
##   data$mathtest)
##
## Residuals:
##   Min       1Q   Median       3Q      Max
## -4.5193 -0.6151 -0.1532  0.7213  3.0132
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  83.70818   1.07632  77.773 < 2e-16 ***
## data$createst  0.56792   0.05899   9.628 1.68e-12 ***
## data$Mechtest -0.08071   0.07641  -1.056  0.296
## data$absttest  0.65839   0.10639   6.188 1.64e-07 ***
## data$mathtest  0.23258   0.02435  9.550 2.15e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.286 on 45 degrees of freedom
## Multiple R-squared:  0.9316, Adjusted R-squared:  0.9255
## F-statistic: 153.1 on 4 and 45 DF, p-value: < 2.2e-16
```