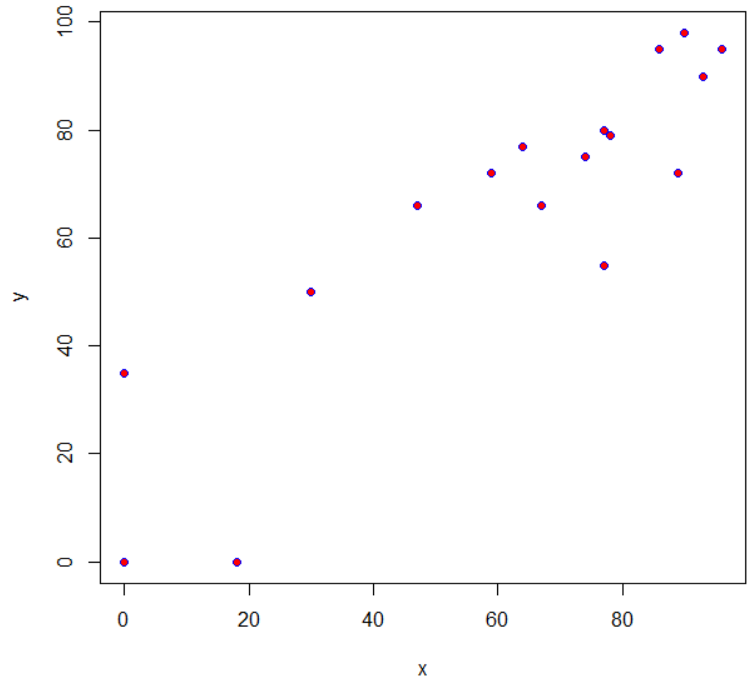
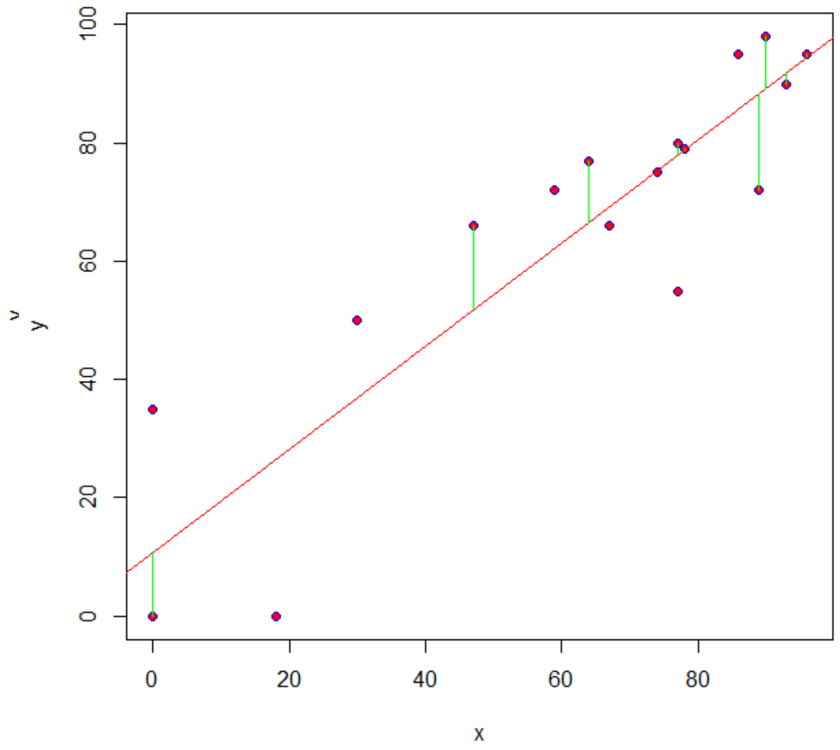
Άσκηση 1

Θέλουμε να δούμε εάν επηρέασε η επίλυση των ασκήσεων, που ανατέθηκαν στους φοιτητές κατά τη διάρκεια της εκπαιδευτικής διαδικασίας, την επίδοσή τους στην πρόοδο. Οι βαθμοί στις ασκήσεις x, και ο βαθμός στην πρόοδο y (με άριστα το 100) δίδονται για 18 φοιτητές.

y	x	y	x	y	x
95	96	72	89	35	0
80	77	66	47	50	30
0	0	98	90	72	59
0	0	90	93	55	77
79	78	0	18	75	74
77	64	95	86	66	67




```
> x<-scan()
1: 96 77 0 0 78 64 89 47 90 93 18 86 0 30 59 77 74 67
19:
Read 18 items
> y<-scan()
1: 95 80 0 0 79 77 72 66 98 90 0 95 35 50 72 55 75 66
19:
Read 18 items
> cbind(y,x,y*x,x*x)
       y  x
 [1,] 95 96 9120 9216
 [2,] 80 77 6160 5929
 [3,]  0  0    0    0
 [4,]  0  0    0    0
 [5,] 79 78 6162 6084
 [6,] 77 64 4928 4096
 [7,] 72 89 6408 7921
 [8,] 66 47 3102 2209
 [9,] 98 90 8820 8100
[10,] 90 93 8370 8649
[11,]  0 18    0  324
[12,] 95 86 8170 7396
[13,] 35  0    0    0
[14,] 50 30 1500  900
[15,] 72 59 4248 3481
[16,] 55 77 4235 5929
[17,] 75 74 5550 5476
[18,] 66 67 4422 4489
>
> cbind(sum(y),sum(x),sum(y*x),sum(x*x))
     [,1] [,2]  [,3]  [,4]
[1,] 1105 1045 81195 80199
```




```
> yx.lm<-lm(y~x)
>
> abline(yx.lm,col="red")
> yx.lm

Call:
lm(formula = y ~ x)

Coefficients:
(Intercept)            x
    10.7269       0.8726
```

```
> cbind(x,y,yx.lm$fitted.values,yx.lm$residuals)
    x  y
1  96 95 94.50098   0.4990229
2  77 80 77.92069   2.0793072
3   0  0 10.72691 -10.7269091
4   0  0 10.72691 -10.7269091
5  78 79 78.79334   0.2066607
6  64 77 66.57629  10.4237122
7  89 72 88.39245 -16.3924513
8  47 66 51.74130  14.2587034
9  90 98 89.26510   8.7349022
10 93 90 91.88304  -1.8830375
11 18  0 26.43455 -26.4345469
12 86 95 85.77451   9.2254883
13  0 35 10.72691  24.2730909
14 30 50 36.90631  13.0936946
15 59 72 62.21306   9.7869449
16 77 55 77.92069 -22.9206928
17 74 75 75.30275  -0.3027532
18 67 66 69.19423  -3.1942274
> sum(yx.lm$residuals)
[1] 1.160183e-14
```

$$\widehat{\beta_1} = \frac{\sum_{i=1}^{n} X_i Y_i - \frac{\sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{n}}{\sum_{i=1}^{n} X_i^2 - \frac{(\sum_{i=1}^{n} X_i)^2}{n}} \equiv \frac{S_{XY}}{S_{XX}}$$

$$\widehat{\beta_0} = \frac{\sum_{i=1}^{n} Y_i}{n} - \widehat{\beta_1} \frac{\sum_{i=1}^{n} X_i}{n} \equiv \overline{Y} - \widehat{\beta_1} \overline{X}$$

```
> predict(yx.lm,list(x=c(45)))
     1
49.996
```

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

```
> cbind(x,y,yx.lm$fitted.values,yx.lm$residuals)
    x  y
1  96 95 94.50098   0.4990229
2  77 80 77.92069   2.0793072
3   0  0 10.72691 -10.7269091
4   0  0 10.72691 -10.7269091
5  78 79 78.79334   0.2066607
6  64 77 66.57629  10.4237122
7  89 72 88.39245 -16.3924513
8  47 66 51.74130  14.2587034
9  90 98 89.26510   8.7349022
10 93 90 91.88304  -1.8830375
11 18  0 26.43455 -26.4345469
12 86 95 85.77451   9.2254883
13  0 35 10.72691  24.2730909
14 30 50 36.90631  13.0936946
15 59 72 62.21306   9.7869449
16 77 55 77.92069 -22.9206928
17 74 75 75.30275  -0.3027532
18 67 66 69.19423  -3.1942274
> sum(yx.lm$residuals)
[1] 1.160183e-14
```

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

$$\hat{\varepsilon}_i = y_i - \hat{y}_i = y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i),$$

εκτιμώμενα σφάλματα κατάλοιπα

Sum of Squares of Errors

$$SSE = \sum_{i=1}^{\nu} \hat{\varepsilon}_i^2 = \sum_{i=1}^{\nu} (y_i - \hat{y}_i)^2$$

Total Sum of Squares

$$SSTO = \sum_{i=1}^{\nu} (y_i - \bar{y})^2$$

Regression Sum of Squares

$$SSR = \sum_{i=1}^{\nu} (\hat{y}_i - \bar{y})^2$$

Το συνολικό άθροισμα τετραγώνων γράφεται στη μορφή

$$SSTO = SSR + SSE$$

$$(y_i - \bar{y})^2 = [(y_i - \hat{y}_i) + (\hat{y}_i - \bar{y})]^2 = (y_i - \hat{y}_i)^2 + (\hat{y}_i - \bar{y})^2 + 2(y_i - \hat{y}_i)(\hat{y}_i - \bar{y})$$

μπορούμε να αναλύσουμε το άθροισμα $SSTO$ σε τρία επιμέρους αθροίσματα, ως εξής

$$SSTO = \sum_{i=1}^{\nu} (y_i - \bar{y})^2 = \sum_{i=1}^{\nu} (y_i - \hat{y}_i)^2 + \sum_{i=1}^{\nu} (\hat{y}_i - \bar{y})^2 + 2\sum_{i=1}^{\nu} (y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) .$$

Σύμφωνα με τον τύπο έχουμε

$$y_i - \hat{y}_i = y_i - [\bar{y} + \hat{\beta}_1 (x_i - \bar{x})] = (y_i - \bar{y}) - \hat{\beta}_1 (x_i - \bar{x}),$$
$$\hat{y}_i - \bar{y} = [\bar{y} + \hat{\beta}_1 (x_i - \bar{x})] - \bar{y} = \hat{\beta}_1 (x_i - \bar{x})$$

όποτε το 3ο άθροισμα του δεξιού μέλους της θα είναι ίσο με

$$\sum_{i=1}^{\nu} (y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) = \sum_{i=1}^{\nu} [(y_i - \bar{y}) - \hat{\beta}_1 (x_i - \bar{x})] \hat{\beta}_1 (x_i - \bar{x}) =$$
$$= \hat{\beta}_1 \sum_{i=1}^{\nu} (x_i - \bar{x})(y_i - \bar{y}) - \hat{\beta}_1^2 \sum_{i=1}^{\nu} (x_i - \bar{x})^2 =$$
$$= \hat{\beta}_1 \left(\sum_{i=1}^{\nu} (x_i - \bar{x})(y_i - \bar{y}) - \hat{\beta}_1 \sum_{i=1}^{\nu} (x_i - \bar{x})^2 \right) = 0$$

$$SSR = \sum_{i=1}^{\nu} (\hat{y}_i - \bar{y})^2$$

$$SSE = \sum_{i=1}^{\nu} \hat{\varepsilon}_i^2 = \sum_{i=1}^{\nu} (y_i - \hat{y}_i)^2$$

```
> anova(yx.lm)
Analysis of Variance Table

Response: y
          Df  Sum Sq Mean Sq F value    Pr(>F)
x          1 14873.0   14873  77.483 1.571e-07 ***
Residuals 16  3071.2     192
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

$$R^2 = \frac{SSR}{SSTO} = 1 - \frac{SSE}{SSTO}$$

λέγεται **συντελεστής προσδιορισμού**

```
> summary(yx.lm)

Call:
lm(formula = y ~ x)

Residuals:
     Min       1Q   Median       3Q      Max
-26.4345  -8.8437   0.3528   9.6466  24.2731

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 10.72691    6.61733   1.621    0.125
x            0.87265    0.09914   8.802 1.57e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 13.85 on 16 degrees of freedom
Multiple R-squared:  0.8288,    Adjusted R-squared:  0.8181
F-statistic: 77.48 on 1 and 16 DF,  p-value: 1.571e-07
```

Ιδιότητες των εκτιμητών

Πρόταση

Για το απλό γραμμικό μοντέλο οι εκτιμητές ελαχίστων τετραγώνων είναι γραμμικοί συνδυασμοί των παρατηρήσεων Y_i. Δηλαδή,

$$\widehat{\beta_1} = \sum_{i=1}^{n} \kappa_i Y_i \quad \text{και} \quad \widehat{\beta_0} = \sum_{i=1}^{n} \lambda_i Y_i,$$

$$\text{όπου} \quad \kappa_i = \frac{X_i - \overline{X}}{\sum_{j=1}^{n}(X_j - \overline{X})^2} \equiv \frac{X_i - \overline{X}}{S_{XX}} \quad \text{και} \quad \lambda_i = \frac{1}{n} - \kappa_i \overline{X} \quad i = 1, \ldots, n.$$

Για τα κ_i και λ_i ισχύουν τα εξής:

$$\sum_{i=1}^{n} \kappa_i = 0, \quad \sum_{i=1}^{n} \kappa_i^2 = \frac{1}{S_{XX}}, \quad \sum_{i=1}^{n} \kappa_i X_i = 1,$$

$$\sum_{i=1}^{n} \lambda_i = 1, \quad \sum_{i=1}^{n} \lambda_i^2 = \frac{1}{n} + \frac{\overline{X}^2}{S_{XX}}, \quad \sum_{i=1}^{n} \lambda_i X_i = 0.$$