Лабораторна робота №13. Побудова багатофакторної регресійної моделі

Мета: вивчення методу найменших квадратів (МНК) для практичного розв'язання задачі побудови багатофакторної регресійної моделі, придбання навичок використання цього методу для розв'язання задачі множинної лінійної регресії із застосуванням комп'ютера.

Теоретична частина

За темою лабораторної роботи студент повинен: *знати* загальне формулювання задачі множинної регресії; *вміти* розв'язувати задачу побудови багатофакторної лінійної регресійної моделі методом найменших квадратів [9].

Загальна задача множинної регресії полягає в знаходженні за даними експерименту залежності (у загальному випадку нелінійної) деякої змінної y від декількох факторів (змінних) $x_1, x_2, \ldots, x_m$ $(x \in R^m)$. У випадку розв'язання задачі лінійної множинної регресії структура моделі розглядається (у загальному випадку) у вигляді залежності:

$$y = b_0 + b_1\phi_1(x) + \ldots + b_{k-1}\phi_{k-1}(x) + e, \qquad (4.1)$$

де $\phi_j(x), j = \overline{1,k-1}$ – деякі задані функції від $x \in R^m$ (узагальнені регресори);

$b_j, j = \overline{0,k-1}$ – деякі коефіцієнти (параметри регресійної моделі);

e – помилка моделі (враховуючи і випадкові помилки випробувань).

Зазначимо, що залежність (3.1) має вид лінійної комбінації функцій $\phi_j(x), j = \overline{1,k-1}$. При цьому структура моделі лінійно залежить від параметрів $b_j, j = \overline{0,k-1}$.

Модель (4.1) є частковим випадком моделі (3.1), а саме тут:

$$G(x,a) = a_1 + a_2\phi_1(x) + \ldots + a_k\phi_{k-1}(x) + e, \qquad (4.2)$$

де $a \in R^k$, $a_j = b_{j+1}$, $j = 1,\ldots,k$.

Тоді для даних експерименту $(x^i, y_i), i = \overline{1,n}$, $(x^i \in R^m)$, оцінки параметрів моделі за методом МНК мають бути знайдені з умови мінімуму по a функції:

$$\Phi(a) \equiv \sum_{i=1}^{n} \left[(y_i - G(x^i; a) \right]^2 . \qquad (4.3)$$

Уведемо позначення:

$$Y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \quad F = \begin{pmatrix} 1 & \phi_1(x^1) & \dots & \phi_{k-1}(x^1) \\ 1 & \phi_1(x^2) & \dots & \phi_{k-1}(x^2) \\ \dots & \dots & \dots & \dots \\ 1 & \phi_1(x^n) & \dots & \phi_{k-1}(x^n) \end{pmatrix}.$$

Тоді згідно з (4.2), (4.3) функція $\Phi(a)$ може бути записана в матричному вигляді:

$$\Phi(a) = \|Y - Fa\|^2 .$$

Відповідно до необхідної умови мінімуму 1-го порядку для функції $\Phi(a)$ оцінки параметрів a можуть бути знайдені з рівняння:

$$\Phi'(a) = 2F^T (Y - Fa) = 0 ,$$

звідки (при умові, що $rank(F) = k$, тобто матриця $(F^T F)$ не вироджена):

$$a^* = (F^T F)^{-1} F^T Y . \qquad (4.4)$$

Таким чином, параметри b_j, $j = \overline{0, k-1}$, моделі (4.1) дорівнюють $b_{j-1} = a_j^*$, $j = \overline{1,k}$. При цьому вектор залишків $r = Y - Fa^*$.

Практична частина

Приклад 1

Побудувати двофакторну лінійну регресійну модель виду:

$$y = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_1 x_2 + b_4 x_1^2 + b_5 x_2^2 \qquad (4.5)$$

при даних експерименту (табл. 4.1).

Таблиця 4.1

Дані експерименту

№ експерименту	x_1	x_2	y
1	-1	-1	4,7
2	1	-1	8,6
3	-1	1	11,2
4	1	1	20,9
5	0	0	7,35
6	0	1	17,0
7	0	-1	7,4
8	1	0	10,2
9	-1	0	3,4

Розв'язання.

У цьому випадку, відповідно (4.1), $k = 6$,

$$\phi_0(x) = 1, \phi_1(x) = x_1, \phi_2(x) = x_2, \phi_3(x) = x_1 x_2, \phi_4(x) = x_1^2, \phi_5(x) = x_2^2$$

.

Введемо відправні дані задачі:

D:\ХНЭУ\МС\Методичка_Моя\New\Здано_в_печать\от ИП\R\Exaple 4_1.R - R Editor

```
m = 2 # кількість факторів
k = 6 # кількість функцій моделі регресії
n = 9 # кількість експериментів

X=matrix(c(-1, 1, -1, 1, 0, 0, 0, 1, -1, -1, -1, 1, 1, 0, 1, -1, 0, 0), nrow=n, ncol=m)
Y=c(4.8, 8.7, 11.2, 20.9, 7.3, 17.0, 7.4, 10.2, 3.4)

FunFi = function(x){
      Fi = c(1:k)
      Fi[1] = 1
      Fi[2] = x[1]
      Fi[3] = x[2]
      Fi[4] = x[1]*x[2]
      Fi[5] = x[1]*x[1]
      Fi[6] = x[2]*x[2]
      return (Fi)
}
```

Визначимо та обчислимо матрицю F:

D:\ХНЭУ\МС\Методичка_Моя\New\3д...

```
F = matrix(0, nrow=n, ncol=k)
for(i in 1:n){
      Fi = FunFi(X[i,])
      for(j in 1:k){
            F[i,j] = Fi[j]
      }
}
```

R Console

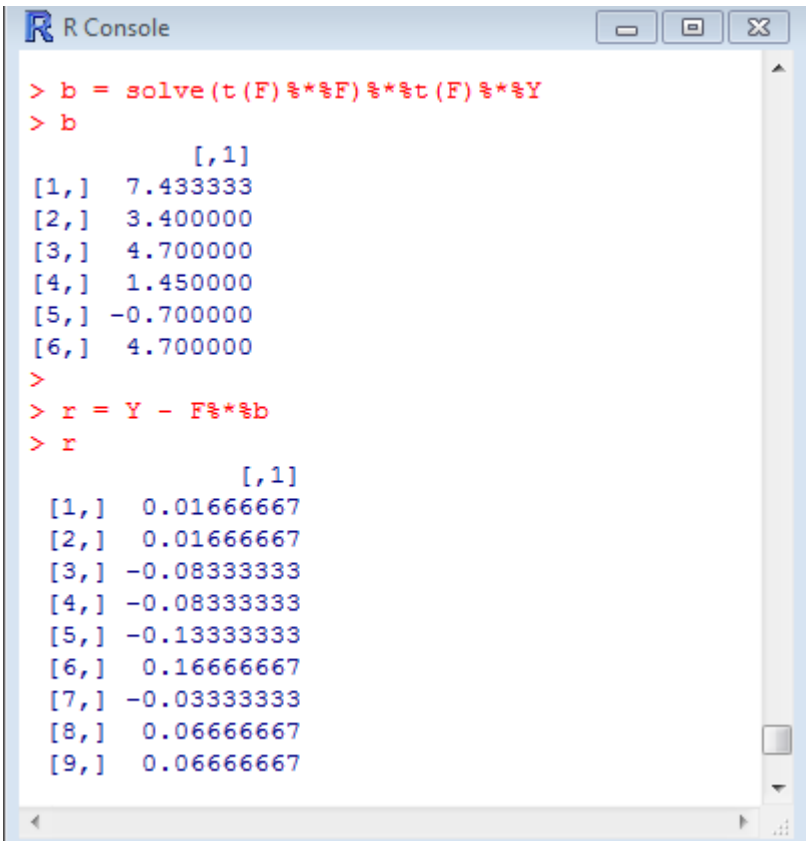
```
> for(i in 1:n){
+ Fi = FunFi(X[i,])
+ for(j in 1:k){
+ F[i,j] = Fi[j]
+ }
+ }
> F
      [,1] [,2] [,3] [,4] [,5] [,6]
 [1,]    1   -1   -1    1    1    1
 [2,]    1    1   -1   -1    1    1
 [3,]    1   -1    1   -1    1    1
 [4,]    1    1    1    1    1    1
 [5,]    1    0    0    0    0    0
 [6,]    1    0    1    0    0    1
 [7,]    1    0   -1    0    0    1
 [8,]    1    1    0    0    1    0
 [9,]    1   -1    0    0    1    0
>
```

та знайдемо оцінки параметрів моделі і залишки:

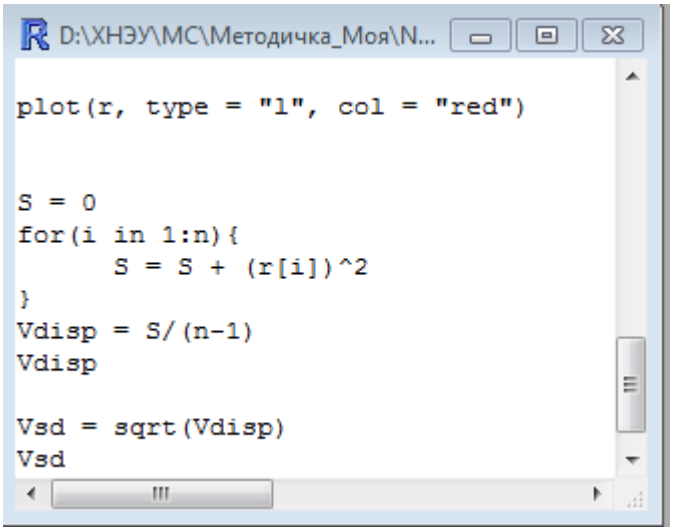
D:\ХНЭУ\МС\Методичка_Моя\New\3д...

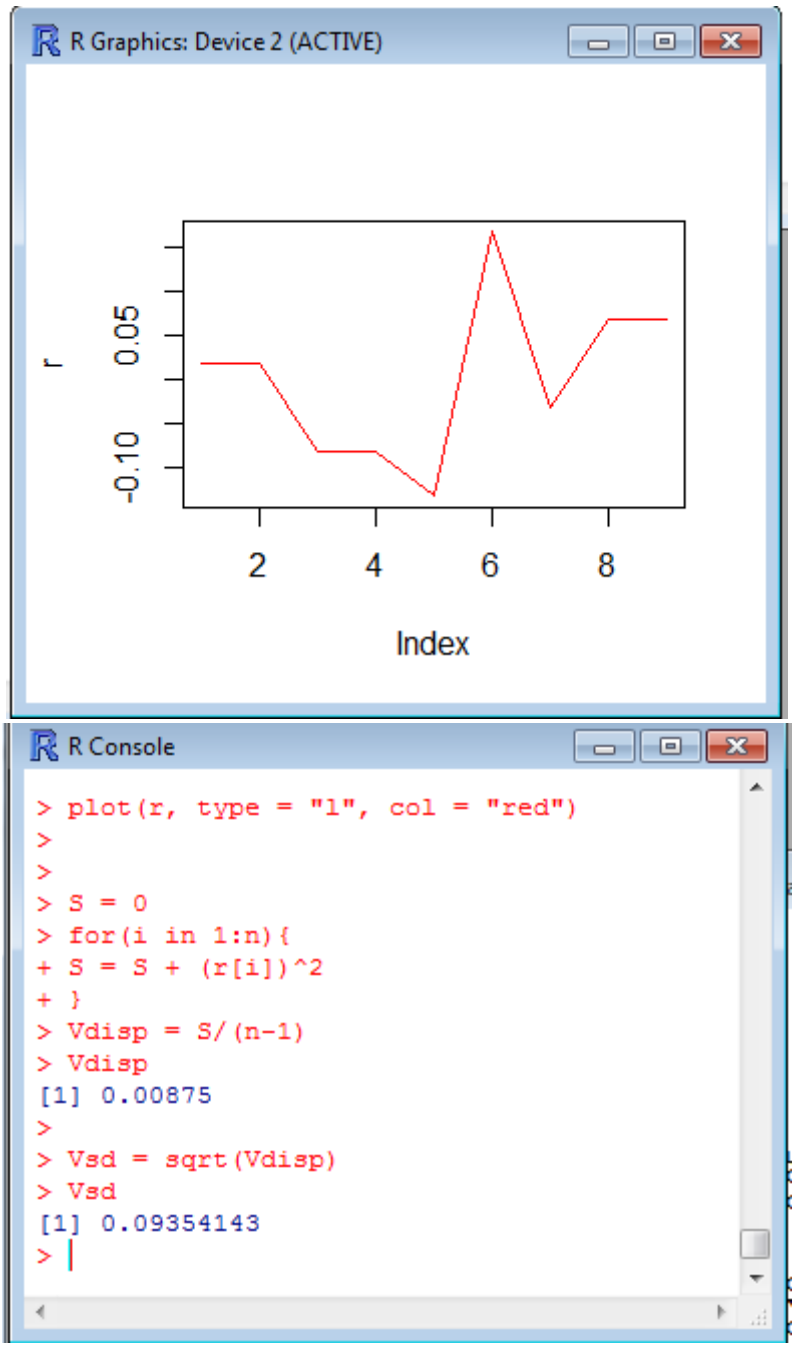
```
b = solve(t(F)%*%F)%*%t(F)%*%Y
b

r = Y - F%*%b
r
```



Для перевірки правильності отриманого розв'язку, а також наочної інтерпретації результатів моделювання, побудуємо графік залишків та знайдемо оцінки дисперсії та середньоквадратичного відхилення залишків:





З вигляду графіка залишків і значення оцінки середньоквадратичного відхилення залишків *Vsd* можна зробити висновок про те, що лінійна регресійна модель (4.5) досить добре описує залежність вихідної змінної y від факторів x_1 й x_2, хоча деяка залежність у залишках і спостерігається.

Приклад 2 [15].

Треба знайти функцію, що описує залежність показника міцності сталі від %-го складу вуглецю в сталі та температури відпуску сталі протягом години при наступних даних експерименту:

$$X := \begin{pmatrix} 6 & 133 \\ 10 & 122 \\ 23 & 112 \\ 24 & 103 \\ 15 & 116 \\ 24 & 99 \\ 57 & 83 \\ 54 & 69 \\ 65 & 76 \\ 66 & 74 \\ 68 & 61 \\ 62 & 38 \\ 83 & 40 \\ 100 & 47 \\ 116 & 31 \end{pmatrix} \qquad Y := \begin{pmatrix} 555 \\ 499 \\ 588 \\ 559 \\ 608 \\ 507 \\ 603 \\ 653 \\ 661 \\ 678 \\ 661 \\ 708 \\ 724 \\ 703 \\ 749 \end{pmatrix}$$

У першому стовпці матриці X знаходиться %-й склад вуглеця в сталі, помножений на 100, а в другому – температура відпуску сталі протягом години (в F^0), помножена на 10. У векторі Y знаходяться відповідні значення показника міцності.

Розв'язання

Застосуємо також модель (4.5):

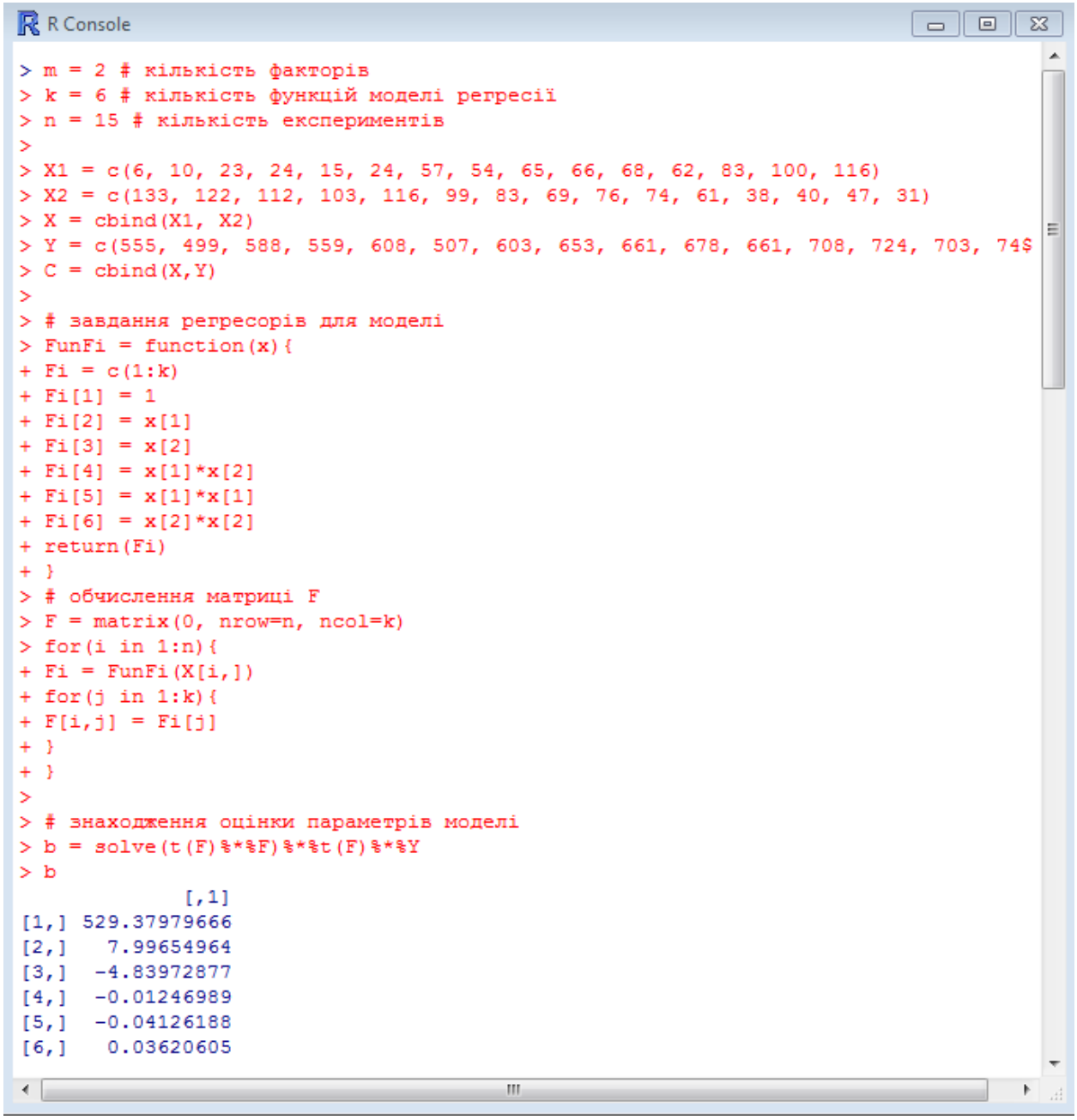
F:\Victor\ХНЭУ\M C\Методичка_Моя\New\Здано_в_печать\Лаб-ки\от ИГ\R\Exaple 4_2.R - R Editor

```
m = 2 # кількість факторів
k = 6 # кількість функцій моделі регресії
n = 15 # кількість експериментів

X1 = c(6, 10, 23, 24, 15, 24, 57, 54, 65, 66, 68, 62, 83, 100, 116)
X2 = c(133, 122, 112, 103, 116, 99, 83, 69, 76, 74, 61, 38, 40, 47, 31)
X = cbind(X1, X2)
Y = c(555, 499, 588, 559, 608, 507, 603, 653, 661, 678, 661, 708, 724, 703, 749)
C = cbind(X,Y)

# завдання регресорів для моделі
FunFi = function(x){
      Fi = c(1:k)
      Fi[1] = 1
      Fi[2] = x[1]
      Fi[3] = x[2]
      Fi[4] = x[1]*x[2]
      Fi[5] = x[1]*x[1]
      Fi[6] = x[2]*x[2]
      return(Fi)
}
# обчислення матриці F
F = matrix(0, nrow=n, ncol=k)
for(i in 1:n){
      Fi = FunFi(X[i,])
      for(j in 1:k){
            F[i,j] = Fi[j]
      }
}

# знаходження оцінки параметрів моделі
b = solve(t(F)%*%F)%*%t(F)%*%Y
b
```

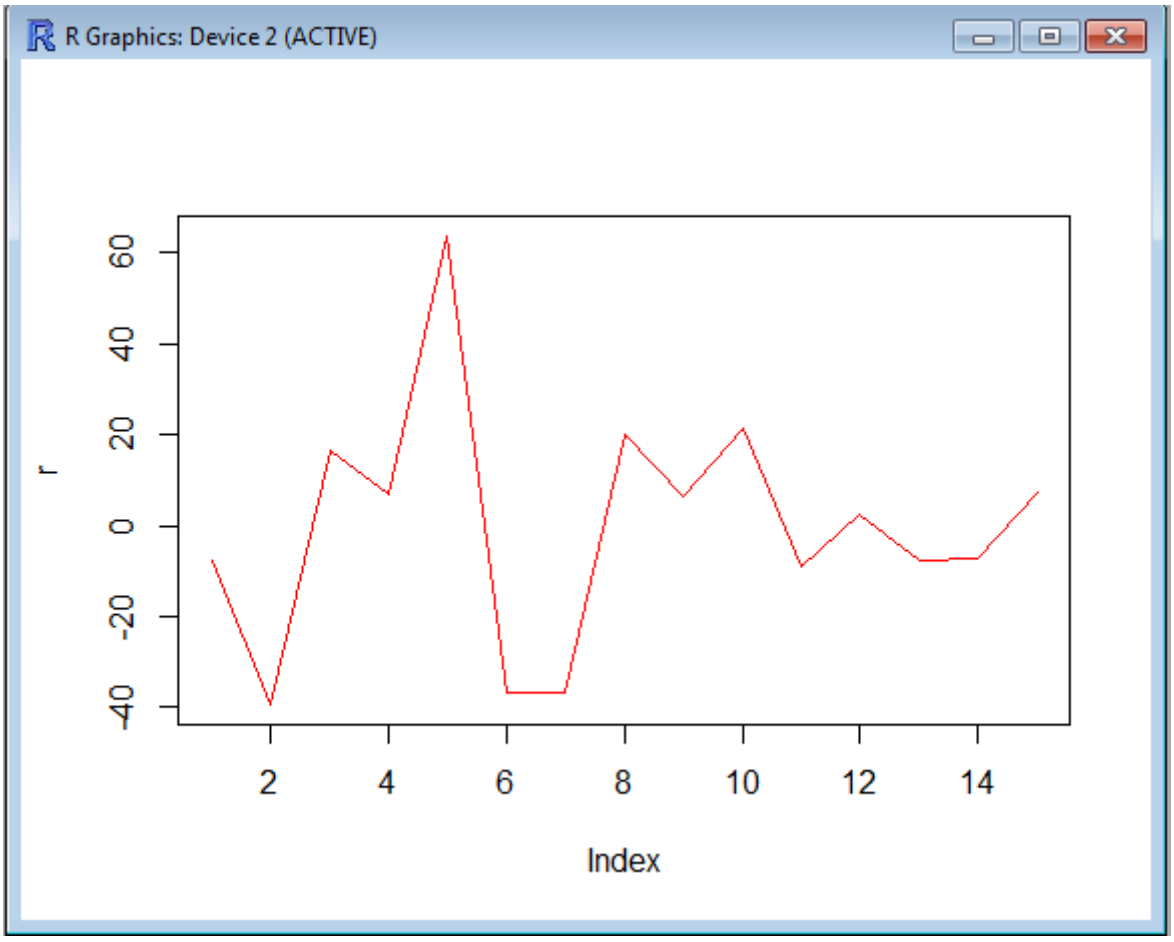


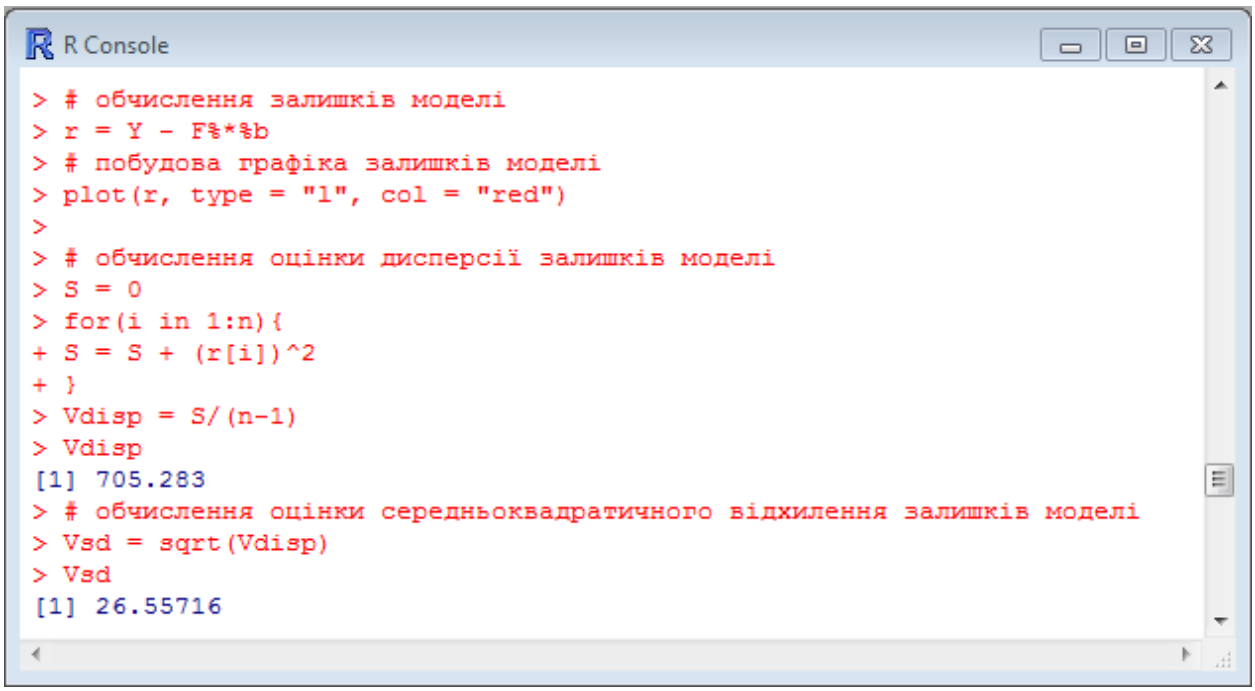
Для перевірки правильності отриманого розв'язку, а також наочної інтерпретації результатів моделювання, побудуємо графік залишків та знайдемо оцінки дисперсії та середньоквадратичного відхилення залишків:

F:\Victor\ХНЭУ\М С\Методичка_Моя\New\Здано_в_печать\Лаб-ки\от ИГ\R\Exaple ...

```
# обчислення залишків моделі
r = Y - F%*%b
# побудова графіка залишків моделі
plot(r, type = "l", col = "red")

# обчислення оцінки дисперсії залишків моделі
S = 0
for(i in 1:n){
      S = S + (r[i])^2
}
Vdisp = S/(n-1)
Vdisp
# обчислення оцінки середньоквадратичного відхилення залишків моделі
Vsd = sqrt(Vdisp)
Vsd
```





Для ілюстрації залежності показника міцності сталі від %-го складу вуглецю в сталі та температури відпуску сталі протягом години побудуємо графік:

F:\Victor\ХНЭУ\М С\Методичка_Моя\New\Здано_в_печать\Лаб-ки\от ИГ\R\Exaple 4_2.R - ...

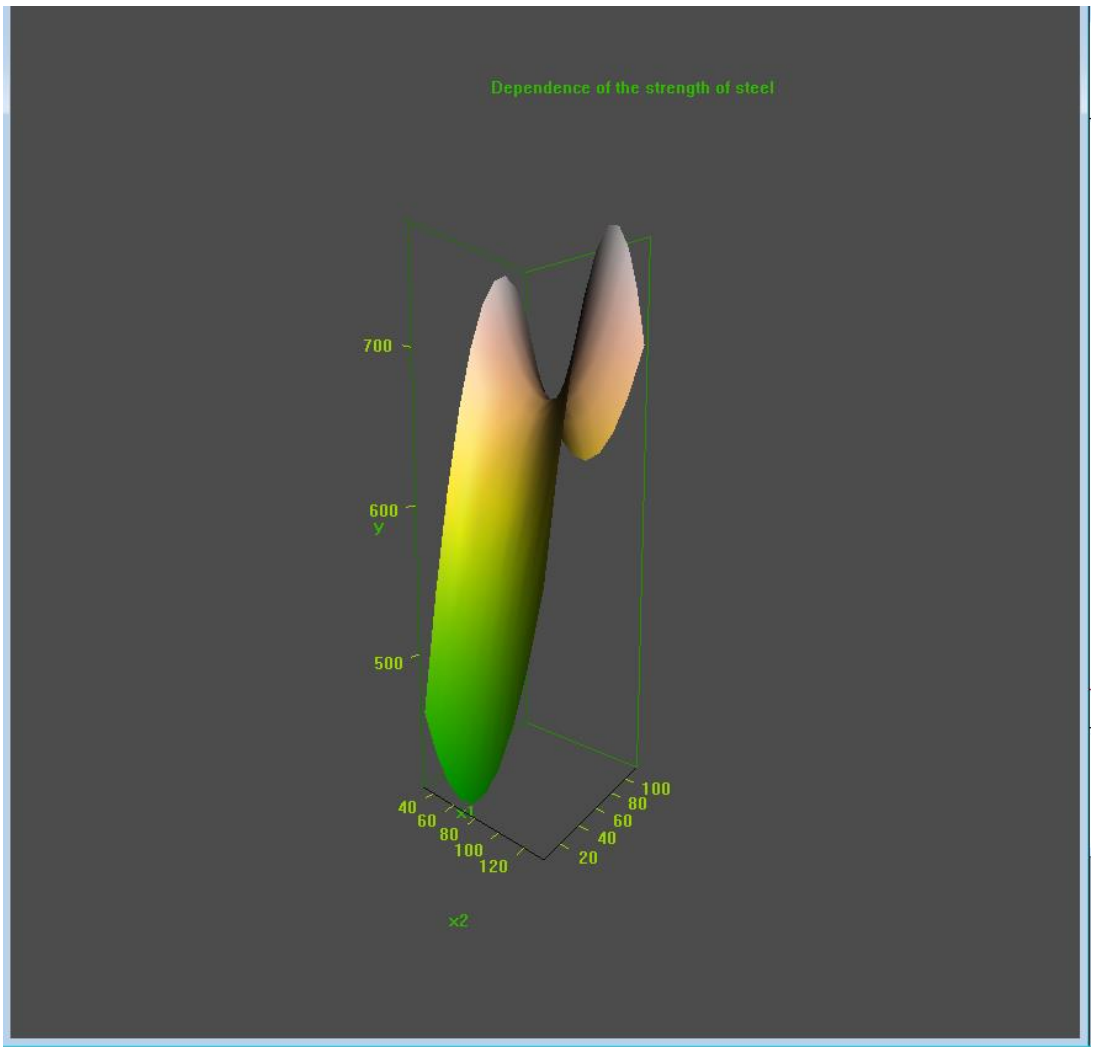
```
# завдання функції багатофакторної лінійної регресійної моделі
FunModel = function(x, a){
      Fi = FunFi(x)
      S = 0
      for(j in 1:k){
            S = S + a[j]*Fi[j]
      }
      return(S)
}

# завдання точок сітки для побудови графіка
N1 = 10
N2 = 10
x1 = c(1:N1)
x2 = c(1:N2)
y = matrix(0, nrow=N1, ncol=N2)
h1 = (max(X1) - min(X1))/(N1-1)
h2 = (max(X2) - min(X2))/(N2-1)
minX1 = min(X1)
minX2 = min(X2)
for(i in 1:N1){
      x1[i] = minX1 + h1*(i-1)
}
for(i in 1:N2){
      x2[i] = minX2 + h2*(i-1)
}
# обчислення значень функції багатофакторної регресійної моделі на сітці
xv = c(1:2)
for(i in 1:N1){
      xv[1] = x1[i]
      for(j in 1:N2){
            xv[2] = x2[j]
            y[i,j] = FunModel(xv, b)
      }
}
```

F:\Victor\ХНЭУ\М С\Методичка_Моя\New\Здано_в_печать\Лаб-ки\от ИГ\R\Exaple 4_2.R - R Editor

```
# побудова графіка залежності показника міцності сталі від %-го складу
# вуглецю в сталі та температури відпуску сталі протягом години
library(rgl)
ylim = range(y)
ylen = ylim[2] - ylim[1] + 1
colorlut = terrain.colors(ylen) # height color lookup table
col = colorlut[ y-ylim[1]+1 ]   # assign colors to heights for each point
#rgl.viewpoint(zoom = 0.5)

rgl.surface(x1, x2, y, coords=1:3, color=col, back="fill")
axes3d()                  # вивід осей на графік
title3d('Dependence of the strength of steel','','x1','y','x2') # вивід назви осей
```



Варіанти завдань

Побудувати двофакторну лінійну регресійну модель виду:

$$y = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_1 x_2 + b_4 x_1^2 + b_5 x_2^2$$

при даних експерименту ($n = 9$, фактор x_1 приймає значення у 1-му рядку таблиці, фактор x_2 приймає значення у 1-му стовпці таблиці, значення змінної y перебувають на перетині).

Варіант 1			
2ф/1ф	0	1	2
20	14	19,1	21,5
10	4	7,9	9,5
0	5,8	8,6	8,7

Варіант 2			
2ф/1ф	0	10	20
0	13	18,1	20,5
1	3	6,9	8
2	4,8	7,6	7,7

Варіант 3			
2ф/1ф	0	10	20
1	4,8	7,4	8,6
2	3	7,2	10
3	11	17	21,5

Варіант 4			
2ф/1ф	1	2	3
20	4,8	7,4	8,6
10	3,5	7	10
0	11,1	17	21,5

Варіант 5			
2ф/1ф	0	10	20
1	6,5	8,5	9,8
2	4,4	9,2	13
3	9,2	17	24

Варіант 6			
2ф/1ф	0	1	2
0	13,1	18,1	20,5
10	3	6,9	8,1
20	4,8	7,6	7,7

Варіант 7			
2ф/1ф	1	2	3
0	13	18,1	20,2
10	3,5	6,9	8
20	4,8	7,6	7,7

Варіант 8			
2ф/1ф	0	1	2
0	6,8	9,3	10,5
10	5,5	9,2	12
20	13,1	19	23,4

Варіант 9			
2ф/1ф	1	2	3
20	5,9	8,4	9,6
10	4,5	8,2	11
0	12,1	18	22,5

Варіант 10			
2ф/1ф	0	10	20
1	6,5	8,5	9,8
2	4,4	9,2	13
3	9,2	16,6	24

Варіант 11			
2ф/1ф	0	10	20
0	13	18,2	20,5
1	3	6,9	8,1
2	4,8	7,6	7,7

Варіант 12			
2ф/1ф	1	2	3
0	13	18,1	20,4
10	3,1	6,9	8
20	4,8	7,6	7,7

Варіант 13			
2ф/1ф	0	1	2
20	14	19,1	21,5
10	4	7,9	9,1
0	5,8	8,6	8,8

Варіант 14			
2ф/1ф	0	10	20
0	13	18,1	20,5
1	3	6,9	8,1
2	4,9	7,6	7,7

Варіант 15			
2ф/1ф	0	10	20
1	4,8	7,3	8,6
2	3,5	7,2	10
3	11	17	21,5