8. ANALIZA KORELACJI I REGRESJI

8.1.1. Miary zależności

Macierz kowariancji

```
C = cov(X)
C = cov(X, rodzaj)
C = cov(x, y)
C = cov(x, y, rodzaj)
```

Macierz korelacji

```
R = corrcoef(X)
R = corrcoef(x, y)
```

gdzie:

`X` – macierz której kolumny odpowiadają kolejnym zmiennym a wiersze kolejnym obserwacjom;

`x`, y – wektory zawierające wartości cech dwóch zmiennych;

`rodzaj` – jeśli równy 0 wykorzystywane są wzory z dzieleniem przez (n-1), jeśli równy 1 wykorzystywane wzory z dzieleniem przez n, domyślnie równy 0.

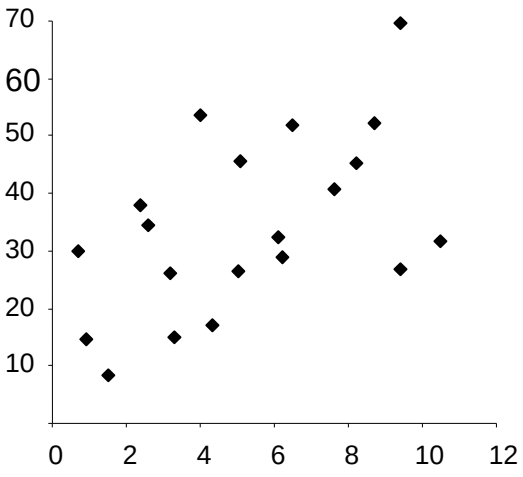
Przykład 1.

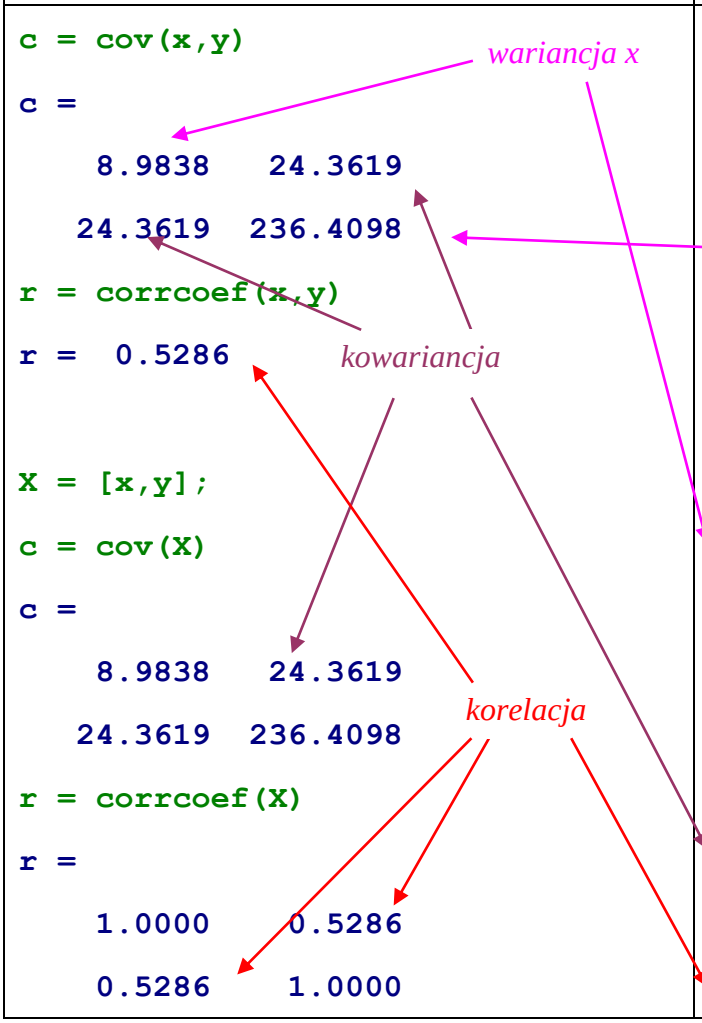
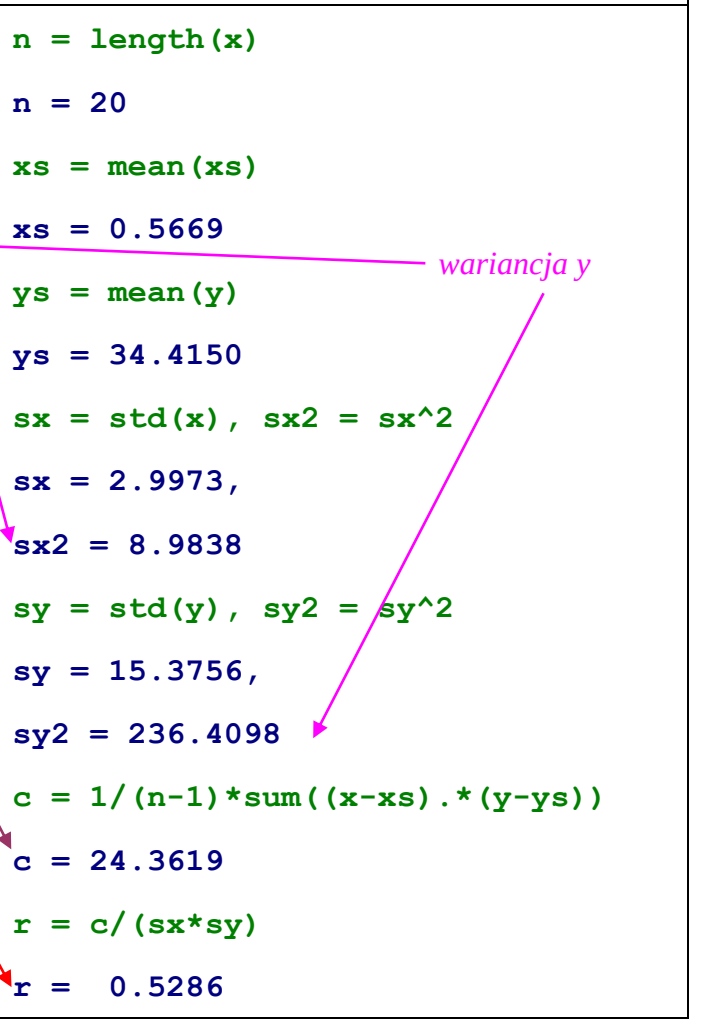
Podczas badania zależności dwóch wymiarów pewnego detalu otrzymano z próby n=20 następujące wyniki:

x_i	3.3	2.4	3.2	10.5	5.1	2.6	1.5	9.4	6.2	0.7	4	7.6	9.4	6.1	4.3	0.9	8.2	6.5	5	8.7
y_i	15.1	38	26	31.6	45.6	34.4	8.5	26.8	28.9	30	53.4	40.6	69.4	32.3	17.1	14.6	45.2	51.9	26.6	52.3

Narysować wykres rozrzutu, wyznaczyć kowariancję i współczynnik korelacji.

```
x=[3.3;2.4;3.2;10.5;5.1;2.6;1.5;9.4;6.2;
0.7;4;7.6;9.4;6.1;4.3;0.9;8.2;6.5;5;8.7];
y=[15.1;38;26;31.6;45.6;34.4;8.5;26.8;
28.9;30;53.4;40.6;69.4;32.3;17.1;14.6;
45.2;51.9;26.6;52.3];

plot(x,y,'kd')
```



wersja 1.	*wersja 2.*
	


```
c = cov(x,y)
c =
    8.9838   24.3619
   24.3619  236.4098
r = corrcoef(x,y)
r =  0.5286

X = [x,y];
c = cov(X)
c =
    8.9838   24.3619
   24.3619  236.4098
r = corrcoef(X)
r =
    1.0000    0.5286
    0.5286    1.0000
```

```
n = length(x)
n = 20
xs = mean(xs)
xs = 0.5669
ys = mean(y)
ys = 34.4150
sx = std(x), sx2 = sx^2
sx = 2.9973,
sx2 = 8.9838
sy = std(y), sy2 = sy^2
sy = 15.3756,
sy2 = 236.4098
c = 1/(n-1)*sum((x-xs).*(y-ys))
c = 24.3619
r = c/(sx*sy)
r =  0.5286
```

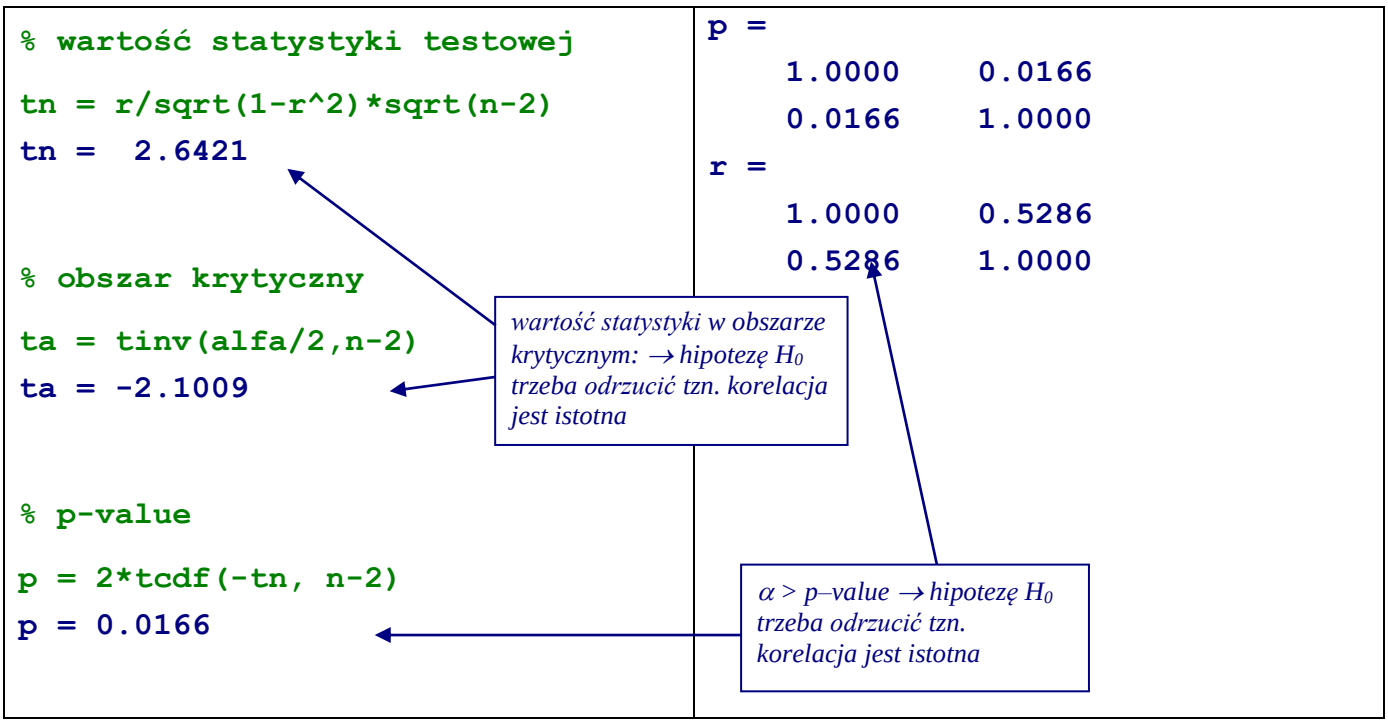
8.1.2. Istotność współczynnika korelacji

Przykład 2.

Na poziomie istotności $\alpha = 0.05$ zbadać istotność współczynnika korelacji z przykładu 1.

wersja 1.	*wersja 2.*

`alfa = 0.05;`
`n = length(x)`
`n = 20`
`r = corrcoef(x,y)`
`r =  0.5286`
	funkcja `corrcoef` *wywołana z dwoma parametrami wyjściowymi zwraca w drugim parametrze wyliczone dla każdego współczynnika korelacji wartość p-value, więc:*
	`[r p]=corrcoef(x,y)`



8.2. Analiza regresji

Funkcja regresji

```
b = regress(y, X)
```

gdzie:

`X` – macierz wejść; `y` – wektor wyjść; `b` – wektor współczynników funkcji regresji; funkcja `regress` może być wywołana również z większą liczbą parametrów wyjściowych w których zwracane są przedziały ufności dla współczynników regresji, wartości wektora błędów (reszt) i przedziały ufności dla elementów tego wektora oraz wartości współczynnika determinacji, statystyki *F* itp.

Macierz wejść

```
X = x2fx(y, x)
X = x2fx(y, x, model)
```

gdzie:

`X` – macierz wejść; `y` – wektor wyjść; `x` – macierz , której kolejne kolumny zawierają wartości kolejnych zmiennych niezależnych poszukiwanej funkcji; `model` – określa postać macierzy wejść, domyślnie wybierany jest `model = 'linear',` dozwolone wartości:

- `'linear'` – uzupełnia model o stałą (macierz jest uzupełniana kolumną jedynek),
- `'interaction'` – uzupełnia model o stałą i iloczyny zmiennych (tzw. interakcje),
- `'quadratic'` – uzupełnia model o stałą, interakcje i kwadraty zmiennych,
- `'purequadratic'` – uzupełnia model o stałą i kwadraty zmiennych,

Statystyki testowe analizy regresji

```
stat = regstats(y, X, model, stats)
```

gdzie:

`X`, `y`, `b`, `model` – jw.; `model` może być także podany w postaci macierzy, np. przekazanie macierzy jednostkowej o wymiarach równych liczbie kolumn macierzy `X` powoduje, że macierz `X` nie jest modyfikowana,

`stats` – tablica komórkowa zawierająca nazwy obliczanych statystyk, wybrane wartości:

- `'rsquare'` – współczynnik determinacji R^2,
- `'adjrsquare'` – skorygowany współczynnik determinacji $\overline{R}^2$,
- `'tstat'` – parametry statystyki t - *Studenta* wykorzystanej do badania istotności współczynników regresji,
- `'fstat'` – parametry statystyki F wykorzystanej do badania istotności funkcji regresji.

Przykład 3.

Podczas badania zależności kosztów produkcji od ilości produkowanych sztuk otrzymano z próby n=4 następujące wyniki:

x_i	1	2	3	4
y_i	3	4	6	8

Zakładając liniową postać funkcji regresji wyznaczyć jej parametry, obliczyć otrzymaną wartość błędu, narysować wykres rozrzutu, wykreślić znalezioną funkcję.

W zadaniu przyjęto założenie:

$$\hat{y} = b_0 + b_1 x .$$

Z porównania założonej postaci funkcji z jej ogólną postacią:

$$\hat{y} = b_0 \varphi_0(x) + b_1 \varphi_1(x),$$

wynika, że funkcje bazowe $\varphi_0(x)$ i $\varphi_1(x)$ wynoszą w tym przypadku:

$$\varphi_0(x) = 1 \text{ i } \varphi_1(x) = x$$

Dla otrzymanych danych pomiarowych macierz wejść X oraz wektor wyjść y wynoszą więc:

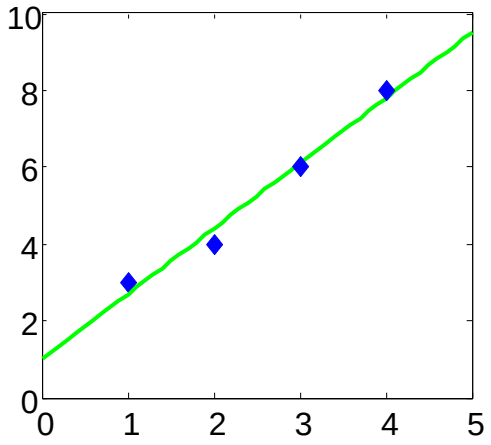
$$X = \begin{pmatrix} \varphi_0(x_1) & \varphi_1(x_1) \\ \varphi_0(x_2) & \varphi_1(x_2) \\ \varphi_0(x_3) & \varphi_1(x_3) \\ \varphi_0(x_4) & \varphi_1(x_4) \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{pmatrix} \qquad y = \begin{pmatrix} 3 \\ 4 \\ 6 \\ 8 \end{pmatrix}.$$

```
x = [1;2;3;4]; y = [3;4;6;8];
```

wersja 1	*wersja 2*

`X = [ones(4,1) x]`
`X =`
`1 1`
`1 2`
`1 3`
`1 4`
 	
`b = inv(X'*X)*X'*y`
`b = 1.0000 1.7000`
 	
`ya = X*b`
`ya = 2.7000 4.4000 6.1000 7.8000`
`sse = (y-ya)'*(y-ya)`
`sse = 0.3000`
`sse = norm(y-ya)^2`
`sse = 0.3000`
	`X = x2fx(x)`
	`X =`
	`1 1`
	`1 2`
	`1 3`
	`1 4`
	
	`b = regress(y, X)`
	`b = 1.0000 1.7000`
	
	`stat=regstats(y,x,'linear','fstat')`
	`stat.fstat`
	`ans =`
	`sse: 0.3000`
	`dfe: 2`
	`dfr: 1`
	`ssr: 14.4500`
	`f: 96.3333`
	`pval: 0.0102`

```
xf = 0:0.1:5;
yf = b(1)+b(2)*xf;
plot(x,y,'bd', xf,yf,'g-')
```



Komentarz.

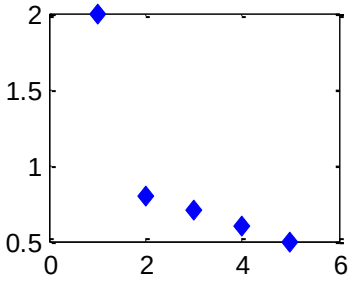
Stosując metodę najmniejszych kwadratów otrzymano współczynniki funkcji regresji: $b = (1, 1.7)^T$, *a w konsekwencji funkcję regresji postaci:* $\hat{y} = 1 + 1.7\,x$. *Wektor przybliżonych wartości zmiennej zależnej wyniósł w tym przypadku* $\hat{y} = (2.7, 4.4, 6.1, 7.8)^T$ *a suma kwadratów błędów* $sse = (3 - 2.6)^2 + \ldots + (8 - 7.8)^2 = 0.3$

Znalezioną funkcję regresji narysowano wyznaczając jej wartości dla punktów $x \in [0, 5]$.

Przykład 4.

Podczas badania zależności zmian pewnego parametru procesu od czasu otrzymano następujące wyniki:

t_i	1	2	3	4	5
z_i	2	0.8	0.7	0.6	0.5



Wykres rozrzutu wskazuje, że zależność parametrów nie jest zależnością liniową. Przyjęto, że zależność ta ma postać:

$$\hat{z} = a_0 e^{a_1/t}.$$

Wyznaczyć parametry funkcji, narysować wykres rozrzutu, wykreślić znalezioną funkcję.

Przyjęta postać funkcji nie może być zapisana bezpośrednio w postaci liniowej kombinacji funkcji bazowych:

$$\hat{z} = b_0\,\varphi_0(t) + b_1\,\varphi_1(t) + \ldots + b_k\,\varphi_k(t).$$

Funkcję tą można jednak przekształcić do takiej postaci:

$$\hat{z} = a_0 e^{a_1/t} \quad \Rightarrow \quad \ln(\hat{z}) = \ln\left(a_0 e^{a_1/t}\right) \quad \Rightarrow \quad \ln(\hat{z}) = \ln(a_0) + \ln\left(e^{a_1/t}\right) \quad \Rightarrow \quad \ln(\hat{z}) = \ln(a_0) + \frac{a_1}{t}.$$

Wprowadzając oznaczenia:

$$\hat{y} = \ln(\hat{z}), \quad b_0 = \ln(a_0), \quad b_1 = a_1$$

można funkcję zapisać w postaci: $\hat{y} = b_0 + b_1 \frac{1}{t}$.

Funkcja ta jest kombinacją liniową funkcji bazowych $\varphi_0(t) = 1$ i $\varphi_1(t) = \frac{1}{t}$. Zastosowanie metody najmniejszych kwadratów wymaga podania macierzy wejść oraz wektora zaobserwowanych wyjść:

$$X = \begin{pmatrix} \varphi_0(t_1) & \varphi_1(t_1) \\ \varphi_0(t_2) & \varphi_1(t_2) \\ \varphi_0(t_3) & \varphi_1(t_3) \\ \varphi_0(t_4) & \varphi_1(t_4) \\ \varphi_0(t_5) & \varphi_1(t_5) \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 0.5 \\ 1 & 0.33 \\ 1 & 0.25 \\ 1 & 0.2 \end{pmatrix} \qquad y = \ln[z] = \begin{pmatrix} 0.69 \\ -0.22 \\ -0.36 \\ -0.51 \\ -0.69 \end{pmatrix}.$$

Po wyznaczeniu parametrów b_i funkcję $\hat{y}(t)$ przekształca się do postaci $\hat{z}(t)$ wykorzystując zależności:

$$\hat{z} = e^{\hat{y}}, \quad a_0 = e^{b_0}, \quad a_1 = b_1.$$

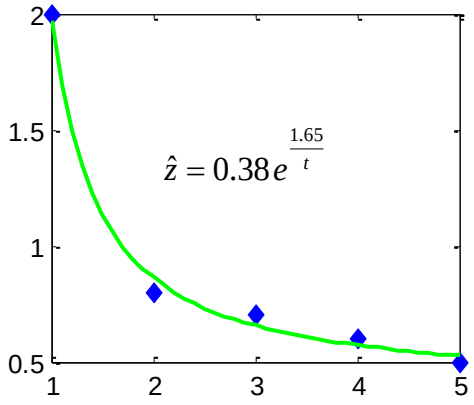
```
t = [1;2;3;4;5]; z = [2; 0.8; 0.7; 0.6; 0.5];
[r,p] = corrcoef(t,z)

r =
    1.0000   -0.8240
   -0.8240    1.0000
p =
    1.0000    0.0862
    0.0862    1.0000
```

p–value > α więc nie można odrzucić hipotezy H_0 o braku istotności korelacji

wersja 1	*wersja 2*

`X = [ones(5,1) 1./t];`
`y = log(z);`
`b = inv(X'*X)*X'*y`
`b = -0.9716  1.6499`
	`X=x2fx(1./t);`
	
	`b = regress(y, X)`
	`b = -0.9716  1.6499`

```
a = [exp(b(1)) b(2)];
a = 0.3785  1.6499
tf = 1:.1:5; zf = a(1)*exp(a(2)./tf);
plot(t,z,'bd',tf,zf,'g-')
```



Przykład 5.

W celu zbadania wpływu dwóch parametrów procesu obróbki na wysokość nierówności wykonano n=10 pomiarów i otrzymano następujące wyniki:

x_{i1}	2.1	1.1	3.1	1.1	2.4	4.4	4.1	1.7	1.1	1.1
x_{i2}	5.8	4.6	2.4	5.6	2.2	2.3	4.1	3.7	3.7	2.1
y_i	24.4	18.4	16.0	22.2	14.0	18.6	24.2	17.7	16.4	11.4

Zakładając następującą postać funkcji regresji:

$$\hat{y}(x) = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_1 x_2 ,$$

wyznaczyć jej parametry, obliczyć otrzymaną wartość błędu, narysować wykres rozrzutu, wykreślić znalezioną funkcję.

Z porównania założonej postaci funkcji z jej ogólną postacią:

$$\hat{y} = b_0\,\varphi_0(x) + b_1\,\varphi_1(x) + b_2\,\varphi_2(x) + b_3\,\varphi_3(x),$$

wynika, że funkcje bazowe $\varphi_0(x)$, $\varphi_1(x)$, $\varphi_2(x)$, $\varphi_3(x)$ wynoszą w tym przypadku:

$$\varphi_0(x) = 1, \quad \varphi_1(x) = x_1, \quad \varphi_2(x) = x_2, \quad \varphi_3(x) = x_1\,x_2\,.$$

Dla otrzymanych danych pomiarowych macierz wejść X oraz wektor wyjść y wynoszą więc:

$$X = \begin{pmatrix} \varphi_0(x_1) & \varphi_1(x_1) & \varphi_2(x_1) & \varphi_3(x_1) \\ \varphi_0(x_2) & \varphi_1(x_2) & \varphi_2(x_2) & \varphi_3(x_2) \\ \vdots & \vdots & \vdots & v \\ \varphi_0(x_{10}) & \varphi_1(x_{10}) & \varphi_2(x_{10}) & \varphi_3(x_{10}) \end{pmatrix} = \begin{pmatrix} 1 & 2.1 & 5.8 & 12.18 \\ 1 & 1.1 & 4.6 & 5.06 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1.1 & 2.1 & 2.31 \end{pmatrix} \qquad y = \begin{pmatrix} 24.4 \\ 18.4 \\ \vdots \\ 11.4 \end{pmatrix}.$$

```
x1=[2.1;1.1;3.1;1.1;2.4;4.4;4.1;1.7;1.1;1.1];
x2=[5.8;4.6;2.4;5.6;2.2;2.3;4.1;3.7;3.7;2.1];
y =[24.4;18.4;16.0;22.2;14.0;18.6;24.2;17.7;16.4;11.4];
```

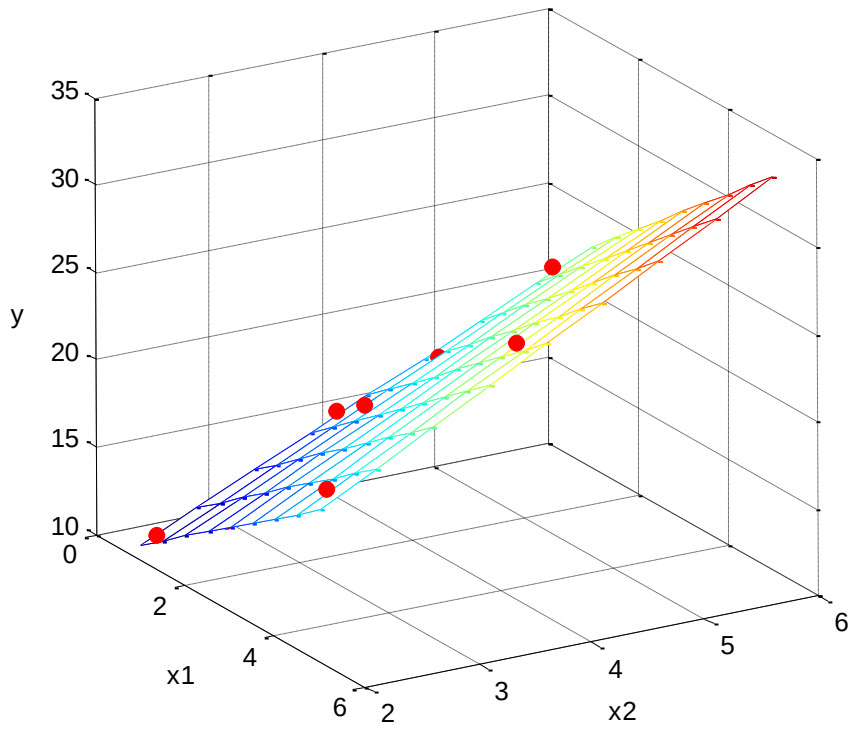
wersja 1

```
X = [ones(10,1) x1 x2 x1.*x2];

b = inv(X'*X)*X'*y
b = 3.2173 1.6963 2.8484  0.1255

ya = X*b;
sse = norm(y-ya)^2
sse = 0.8364
```

wersja 2

```
X=x2fx([x1 x2])

b = regress(y, X)
b = 3.2173 1.6963 2.8484 0.1255

stat=regstats(y,[x1,x2],...
          'interaction','fstat')
stat.fstat
ans =
     sse: 0.8364
     dfe: 6
     dfr: 3
     ssr: 161.8446
       f: 386.9976
    pval: 2.9673e-007
```

```
[x1f x2f]=meshgrid([1:0.1:5],[2:0.1:6]);
yf = b(1) + b(2)*x1f + b(3)*x2f + b(4)*x1f.*x2f;
mesh(x1f, x2f, yf);
hold on;
plot3(x1,x2,y,'ro');
hold off
```



Komentarz.
Znaleziona funkcja regresji postaci: $\hat{y} = 3.2 + 1.7x_1 + 2.8x_2 + 0.1x_1x_2$. *Jest to funkcja dwóch zmiennych, do jej wykreślenia została wykorzystana funkcja* `mesh` *(patrz dodatek: Wykresy 3D). W przykładzie przyjęto, że:* $x_1 \in [1, 5]$ *a* $x_2 \in [2, 6]$. *Wyniki pomiarów* (x_{i1}, x_{i2}, y_i) *to punkty w przestrzeni trójwymiarowej więc do ich wykreślenia wykorzystano funkcję* `plot3` *(a nie* `plot`*).*

Przykład 6.

Dla funkcji regresji z przykładu 5. zweryfikować na poziomie istotności $\alpha = 0.05$ hipotezę o istotności funkcji regresji i istotności jej współczynników. Wyznaczyć współczynnik determinacji.

```
alfa = 0.05;
x1=[2.1;1.1;3.1;1.1;2.4;4.4;4.1;1.7;1.1;1.1];
x2=[5.8;4.6;2.4;5.6;2.2;2.3;4.1;3.7;3.7;2.1];
y =[24.4;18.4;16.0;22.2;14.0;18.6;24.2;17.7;16.4;11.4];
```

wersja 1.	*wersja 2.*

`X = [ones(10,1) x1 x2 x1.*x2];`
`b = inv(X'*X)*X'*y`
`b = 3.2173 1.6963 2.8484 0.1255`
`ya = X*b; ys = mean(y);`
`n = length(y), k=length(b)-1`
`n = 10, k = 3`
`dft = n-1; dfr = k, dfe = n-k-1`
`dfr = 3, dfe = 6`
	`X=x2fx([x1 x2], 'interaction')`
	`b = regress(y, X)`
	`b = 3.2173 1.6963 2.8484 0.1255`
	
	`stat=regstats(y,X,eye(size(X,2)),...`
	`{'fstat', 'tstat', 'rsquare'})`
	`stat =`
	`source: 'regstats'`

```
sst = norm(y-ys)^2;
ssr = norm(ya-ys)^2
ssr = 161.8446
sse = norm(y-ya)^2
sse = 0.8364

R2 = ssr/sst
R2 = 0.9949

Fn = (ssr/dfr)/(sse/dfe)
Fn = 386.9976
Fa = finv(1-alfa, dfr, dfe)
Fa = 4.7571
p = 1-fcdf(Fn, dfr, dfe)
p = 2.9673e-007

C = inv(X'*X); c = diag(C);
tn = b./sqrt(sse/dfe*c)
tn   =   3.8706   5.0534   13.3209
1.2890
ta = -tinv(alfa/2, dfe)
ta = 2.4469
p = 2*tcdf(-abs(tbn), dfe)
p = 0.0083 0.0023 0.0000 0.2449
```

```
    rsquare: 0.9949
      tstat: [1x1 struct]
      fstat: [1x1 struct]

fstat=stat.fstat
fstat =
     sse: 0.8364
     dfe: 6
     dfr: 3
     ssr: 161.8446
       f: 386.9976
    pval: 2.9673e-007

tstat=stat.tstat
tstat =
    beta: [4x1 double]
      se: [4x1 double]
       t: [4x1 double]
    pval: [4x1 double]
     dfe: 6
tn=tstat.t
tn = 3.8706 5.0534 13.3209 1.2890
p=tstat.pval
p = 0.0083 0.0023 0.0000 0.2449
```

Komentarz:

Obydwa przedstawione rozwiązania dały identyczne wyniki. Współczynnik determinacji oceniający jakość dopasowania otrzymanego równania regresji do danych empirycznych wyniósł $R^2 = 0.9949$, co oznacza, że ponad 99% zmienności zmiennej zależnej jest wyjaśniona równaniem regresji.

Test istotności funkcji regresji nakazuje odrzucić hipotezę zerową o braku wpływu zmiennych niezależnych na zmienną zależną (wartość statystyki testowej F_n leży w obszarze krytycznym $F_n = 386.9976 \;>\; F_a = 4.7571$, *graniczny poziom istotności p–value* $p = 2.9673e-007$ jest mniejszy od założonego poziomu istotności $\alpha = 0.05$). Funkcję regresji należy więc uznać za istotną.

Test istotności dla współczynników funkcji regresji wskazuje, że wszystkie współczynniki z wyjątkiem ostatniego związanego z iloczynem zmiennych x_1x_2 są istotne. Wartości statystyk testowych sprawdzających brak istotności współczynników b_0, b_1, b_2, b_3 wynoszą w tym przypadku: $t_{0n} = 3.8706$, $t_{1n} = 5.0534$, $t_{2n} = 13.3209$ i $t_{3n} = 1.2890$, obszar krytyczny w teście jest dwustronny $|t_{in}| > |t_a| = 2.4469$, statystyki t_{0n}, t_{1n}, t_{2n} leżą więc w obszarze krytycznym więc hipotezę o nieistotności współczynników

b_0, b_1, b_2 należy odrzucić. Identyczne wnioski daje analiza wartości *granicznych poziomów istotności* współczynników, które wynoszą odpowiednio: $p_0 = 0.0083$, $p_1 = 0.0023$, $p_2 = 0.0000$ i $p_3 = 0.2449$. Tylko w przypadku współczynnika b_3 założony poziom istotności $\alpha = 0.05$ jest niższy od granicznego poziomu istotności, więc tylko w tym przypadku nie ma podstaw do odrzucenia hipotezy o nieistotności współczynnika.

W wersji 2. rozwiązania funkcja `regress` wykorzystana została do znalezienia współczynników regresji a funkcja `regstats` do obliczenia wartości współczynnika determinacji R^2 i przeprowadzenia testów istotności. Funkcja `regstats` zwraca strukturę (w przykładzie: `stat`), której pola zawierają obliczane przez nią statystyki. Sposób zwracania obliczonych wartości jest uzależniony od rodzaju statystyki, np. wartość współczynnika determinacji jest umieszczana wprost w zwracanej strukturze (pole: `rsquare`), obliczenia wynikające z przeprowadzonych testów istotności są zapisywane w dodatkowych strukturach zapisywanych w strukturze podstawowej (pola: `fstat` i `tstat`). Odczyt wyników zapisywanych w strukturach wewnętrznych wymaga odwołania się do pól, w których struktury te zostały zapisane. W przypadku *testu F* wartości sum: `sse`, `ssr`, liczba ich stopni swobody: `dfe` i `dfr`, wartość statystyki testowej i graniczny poziom istotności są wyświetlane bezpośrednio po wyświetleniu struktury zawierającej informacje o *teście F*. W przypadku *testu t* obliczane wartości są w większości tablicami więc żeby wyświetlić należy odwołać się do pól w których tablice te zostały zapisane.
Z przeprowadzonej analizy wynika, że znalezioną funkcję regresji należy uznać za istotną, jednak tylko trzy spośród czterech współczynników zostały uznane za istotne. Czwarty współczynnik b_3 powinien zostać wyeliminowany z modelu.

Dobór funkcji regresji

Funkcja regresji powinna w możliwie jak największym stopniu wyjaśniać zmienność zmiennej zależnej, jednocześnie jednak powinna mieć możliwie najprostszą strukturę. Dobór modelu funkcji regresji nie jest sprawą prostą, stosowane są różne strategie. W selekcji postępującej konstruowanie modelu rozpoczyna się od jednej zmiennej niezależnej, w kolejnych krokach dodawane są kolejne zmienne. W eliminacji wstecznej poszukiwanie optymalnego modelu rozpoczyna się od modelu maksymalnego a w kolejnych krokach kolejno usuwane są zmienne o najmniejszym wpływie na zmienną zależną. Kombinacją obydwu podejść jest tzw. metoda krokowa.

Przykład 7. Stosując eliminację wsteczną uprościć funkcję regresji z przykładu 6.

Funkcja regresji w przykładzie 6. zawierała współczynnik, który został uznany za nieistotny – współczynnik ten, a właściwie skojarzona z nim interakcja zmiennych x_1 i x_2, zostanie usunięty z modelu.

W przypadku gdyby model zawierał kilka nieistotnych współczynników eliminację należałoby rozpocząć od współczynnika o największym granicznym poziomie istotności. Po wyeliminowaniu tego najmniej istotnego współczynnika należałoby ponownie przeprowadzić testy istotności. Proces eliminacji kończy się gdy w modelu występują wyłączne zmienne, które w istotny sposób wyjaśniają zmienność zmiennej zależnej.

```
X = [ones(10,1) x1 x2];
b = regress(y, X)
b = 2.3533 2.1075 3.0953

stat=regstats(y, X, eye(size(X,2)), {'fstat', 'tstat', 'rsquare'});

R2 = stat.rsquare
R2 = 0.9934

fstat=stat.fstat
fstat =
     sse: 1.0680
     dfe: 7
     dfr: 2
     ssr: 161.6130
       f: 529.6127
    pval: 2.2928e-008

p = stat.tstat.pval
p = 0.0026 0.0000 0.0000
```

Komentarz

Po wyeliminowaniu interakcji zmiennych x_1 i x_2 znaleziona została funkcja regresji postaci: $\hat{y} = 2.4 + 2.1x_1 + 3.1x_2$. Współczynnik determinacji dla znalezionej funkcji wynosi: $R^2 = 0.9934$, ma więc nieznacznie mniejszą wartość w stosunku do obliczonego dla modelu pełnego: $R^2 = 0.9949$. Funkcja regresji jest statystycznie istotna (*graniczny poziom istotności testu F* ma wartość: 2.2928e-008), dla założonego poziomu istotności wszystkie współczynniki funkcji są istotne (*graniczne poziomy istotności testu t* mają wartości: 0.0026, 0.0000, 0.0000).