

## Metode regresive –laborator

1. Un lanț de pizerii este interesat să deschidă un nou local și din experiența anterioară au stabilit că apropierea de campusurile universitare constituie un avantaj. Următorul tabel prezintă situația a 10 localuri, fiecare situat în apropierea altui campus: vânzările din ultimul an și numărul de studenți existent în campusul respectiv.

local	vânzări anuale (mii de \$)	număr studenți (mii de studenți)
1	58	2
2	105	6
3	88	8
4	118	8
5	117	12
6	137	16
7	157	20
8	169	20
9	149	22
10	202	26

Calculați în Matlab coeficientul de corelație al numărului de studenți din campus și valoarea vânzărilor pizzeriei de lângă campusul respectiv. Calculați dreapta de regresie, utilizând formulele studiate. Desenați dreapta de regresie și diagrama împrăștierii. Calculați dreapta de regresie folosind metoda celor mai mici pătrate; calculați și reprezentați grafic reziduurile. Care sunt vânzările prognozate ale unui local care se află în vecinătatea unui campus cu 10.000 studenți?

$$\text{dreapta de regresie: } y = a + bx, \text{ unde: } b = \frac{\sum_{i=1}^n x_i \cdot y_i - \frac{1}{n} \cdot \sum_{i=1}^n x_i \cdot \sum_{i=1}^n y_i}{\sum_{i=1}^n x_i^2 - \frac{1}{n} \left( \sum_{i=1}^n x_i \right)^2}, a = \bar{y} - b \cdot \bar{x}$$

```
>> S=[2 6 8 8 12 16 20 20 22 26];V=[58 105 88 118 117 137 157 169 149 202];
>> A=[S' V']
>> corrcoef(A)
ans =
    1.0000    0.9501
    0.9501    1.0000
=> r = 0.9501
>> b=(S*V'-10*mean(S)*mean(V))./(norm(S)^2-10*mean(S)^2)
>> a=mean(V)-b*mean(S)
>> plot(S,a+b*S); hold on
>> plot(S,V,'o'); hold off
>> A=[1 S(1);1 S(2);1 S(3);1 S(4);1 S(5);1 S(6);1 S(7);1 S(8);1 S(9);1 S(10)];
>> a=A\V'
>> plot(S,a(1)+a(2)*S); hold on
>> plot(S,V,'o'); hold off
>> syms S
>> f=a+b*S
>> f10=subs(f,S,10)
```

2. Într-un proces tehnologic dintr-o fabrică s-a constatat că există o viteză (m/min) la care benzii de lucru afectează numărul de defecțiuni descoperite în timpul verificării. Prezentăm un tabel cu diferite viteze ale benzii și numărul defectelor găsite.

viteza (m/min)	nr de defecțiuni
6	21
6	19
12	15
12	16
18	14
24	17

Este corelat numărul de defecțiuni de viteza benzii de lucru? Determinați dreapta de regresie și desenați-o împreună cu diagrama împrăștierii. Calculați dreapta de regresie folosind metoda celor mai mici pătrate și desenați-o împreună cu reziduurile. Care este numărul prognozată de defecțiuni pentru o viteză a benzii de 15m/min?

3. Problema regenerării ierbii (pășunii), după ultimul pășunat este prezentată în următorul tabel:

timpul	9	14	21	28	42	57	63	70	79
recolta	8.93	10.8	18.59	22.33	39.35	56.11	61.73	64.62	67.08

Variabila timp reprezintă timpul scurs de la ultimul pășunat. Nu se cunosc unitățile de măsură.

Calculați dreapta de regresie, utilizând formulele studiate și desenați-o împreună cu diagrama împrăștierii.

Calculați dreapta de regresie folosind metoda celor mai mici pătrate; reprezentați-o grafic împreună cu reziduurile.

Folosiți regresia neliniară, aproximând cu un polinom de gradul al treilea. Care este recolta după 60 de zile de la ultimul pășunat? Desenați polinomul de regresie și diagrama împrăștierii.

regresie neliniară (aproximarea cu un polinom de gradul 3

```
>> A=[1 T(1) T(1)^2 T(1)^3;1 T(2) T(2)^2 T(2)^3;1 T(3) T(3)^2 T(3)^3;1 T(4) T(4)^2
T(4)^3;1 T(5) T(5)^2 T(5)^3;1 T(6) T(6)^2 T(6)^3;1 T(7) T(7)^2 T(7)^3;1 T(8) T(8)^2
T(8)^3;1 T(9) T(9)^2 T(9)^3]; a=A\R'
>> syms t
>> f=a(1)+a(2)*t+a(3)*t^2+a(4)*t^3;
>> f60=subs(f,t,60)
polinomul de regresie și diagrama împrăștierii:
>> plot(T,R,'o');hold on
>> plot(T,a(1)+a(2)*T+a(3)*T.^2+a(4)*T.^3);hold off
```

4. Unor elevi de gimnaziu li s-au determinat două scoruri: un scor de creativitate și un scor de deprinderi practice, cel din urmă fiind de fapt timpul mediu în care sunt îndeplinite câteva activități ce pun în evidență coordonarea mână-ochi. Rezultatele sunt prezentate în tabelul următor:

elev	scor creativitate	scor deprinderi
1	29	4.5
2	35	3.9
3	37	3.9
4	50	6.1
5	69	4.3
6	84	8.8
7	40	2.1
8	65	5.5
9	29	5.7
10	42	3
11	51	7.1
12	45	7.3
13	31	3.3
14	40	5.2

Determinați dreapta de regresie și desenați-o împreună cu diagrama împrăștierii. Calculați dreapta de regresie folosind metoda celor mai mici pătrate și desenați-o împreună cu reziduurile. Care este numărul prognozat de creativitate pentru un scor de deprinderi practice de 5.1?

5. Aproximați seturile de date  $\{(x_i, y_i), 1 \leq i \leq n\}$  din problemele 1-4 printr-o funcție polinomială, cu metoda celor mai mici pătrate., folosind instrucțiunile polifit și polyval. Calculați reziduurile și desenați-le.

### Problema 1

```
>> S=[2 6 8 8 12 16 20 20 22 26];V=[58 105 88 118 117 137 157 169 149 202];
>> p= polyfit(S,V,4)
>> pol = polyval(p,S,V)
>> V1=p(1)+p(2)*S+p(3)*S.^2+p(4)*S.^3+p(5)*S.^4;
>> plot(S,V1,S,V,'O')
```

Este nevoie de noralizarea datelor

```
>> Ss=(S-mean(S))./std(S)
>> Vs=(V-mean(V))./std(V);
>> ps=polyfit(Ss,Vs,4)
>> pol=polyval(ps,S,V)
>> ps=polyfit(Ss,Vs,4)
>> pol=polyval(ps,Ss,Vs)
>> rez=pol-Vs
>> plot(Ss,rez,'*')
```

6. Intr-o mică firmă de IT, se presupune că Y rezultatul vânzărilor depinde de X1 suma cheltuită pentru reclama tradițională la TV și în ziare și 2 suma cheltuită pentru reclama prin internet (sume sunt în zeci de mii de \$:

Y	5.0	7.0	7.0	5.0	9.0	11.0	12.0	13.0	15.0	17.0	18.0	19.0
X1	2.9	2.5	3.0	1.5	4.0	4.5	5.0	5.5	6.0	7.5	8	7
X2	1.0	2.0	2.5	1.5	3.5	4.0	5.0	6.0	6.5	8.0	8.0	7.5

Scrieți matricea corelațiilor. Stabiliți hiperplanul de regresie, utilizând metoda celor mai mici pătrate. Dacă se utilizează 85000 \$ pentru publicitate clasică și 60.000 pentru publicitate prin internet, la cât se prognozează cifra vânzărilor.

```
>> Y=[5 7 7 5 9 11 12 13 15 17 18 19];
>> X1=[2 2.5 3 1.5 4 4.5 5 5.5 6 7.5 8 7];X2=[1 2 2.5 1.5 3.5 4 5 6 6.5 8 8 7.5];
>> A=[X1' X2' Y'];corrcoef(A)
» B=[1 X1(1) X2(1); 1 X1(2) X2(2);1 X1(3) X2(3);1 X1(4) X2(4);1 X1(5) X2(5);1 X1(6)
X2(6);1 X1(7) X2(7);1 X1(8) X2(8);1 X1(9) X2(9);1 X1(10) X2(10);1 X1(11) X2(11);1
X1(12) X2(12)];a=B\Y'
>> [X1,X2]=meshgrid(2:.1:7,1:.1:8);surf(X1,X2, a(1)+a(2)*X1+a(3)*X2)
>> syms X1 X2
>> f=a(1)+a(2)*X1+a(3)*X2;
>>f1=subs(f,[X1,X2],[8.5,6.0])
```

7. Pe baza unor date asupra utilizării unui camion (2004) să construim hiperplanul de regresie, ce va prognoza numărul de ore lucrate săptămânal de un camion. Variabila Y reprezintă numărul de ore lucrate săptămânal, X1 reprezintă cantitatea transportată (în mii de livre), X2 proporția din tonajul camionului ocupată de cantitatea transportată (proportion shipped by truck), X3 greutatea medie a unui transport, x4 numărul săptămânii considerate:

Y	X1	X2	X3	x4
100	5.1	90	20	1
85	3.8	99	22	2
108	5.3	58	19	3
116	7.5	16	15	4
92	4.5	54	20	5
63	3.3	42	26	6
79	5.3	12	25	7
101	5.9	32	21	8
88	4.0	56	24	9
71	4.2	64	29	10
122	6.8	78	10	11

Calculați matricea corelațiilor și comentați. Calculați hiperplanul de regresie folosind metoda celor mai mici pătrate și calculați reziduurile în acest caz.