



KURS STATYSTYKA

Lekcja 6
Regresja i linie regresji

ZADANIE DOMOWE



Część 1: TEST

Zaznacz poprawną odpowiedź (tylko jedna jest prawdziwa).

Pytanie 1

Funkcja regresji I rodzaju cechy Y zależnej od cechy X, to:

- a) empiryczna prosta regresji, w której wartościom cechy X odpowiadają średnie warunkowe zmiennej Y
- b) teoretyczna linia prosta regresji, w której wartościom cechy X odpowiadają średnie warunkowe zmiennej Y
- c) funkcja przyporządkowująca wartościom cechy X odpowiadające im średnie warunkowe cechy Y
- d) funkcja przyporządkowująca wartościom cechy X odpowiadające im teoretyczne wartości cechy Y

Pytanie 2

Funkcja regresji II rodzaju...

- a) narysowana na wykresie przedstawia pojedyncze punkty
- b) jest dana wzorem (np. wzorem na prostą)
- c) narysowana na wykresie przedstawia pojedyncze punkty połączone odcinkami
- d) jest wyznaczana po wyznaczeniu funkcji regresji I rodzaju

Pytanie 3

Wzór na oszacowanie liniowej funkcji regresji II rodzaju to:

- a) $\hat{y}_i = a_0 + a_1 x_i$
- b) $\hat{y}_i = a_1 + a_0 x_i$

Pytanie 4

Zaznacz prawidłowe wzory na współczynniki oszacowanie liniowej funkcji regresji II rodzaju (a_0, a_1) :

a) $a_1 = \frac{\sum (x_i - \bar{X})(y_i - \bar{Y})}{\sum (x_i - \bar{X})^2}, a_0 = \bar{Y} - a_1 \bar{X}$

b) $a_1 = \frac{\sum (x_i - \bar{X})(y_i - \bar{Y})}{\sum (x_i - \bar{X})^2}, a_0 = \bar{Y} - a_1 \bar{X}$

c) $a_0 = \frac{\sum (x_i - \bar{X})(y_i - \bar{Y})}{\sum (x_i - \bar{X})^2}, a_1 = \bar{Y} - a_0 \bar{X}$

d) $a_1 = \frac{\sum (x_i - \bar{X})(y_i - \bar{Y})}{\sum (x_i - \bar{X})^2}, a_0 = \bar{Y} - a_0 \bar{X}$

Pytanie 5

Współczynnik korelacji liniowej r_{xy} ...

- a) Jest zawsze dodatni
- b) Musimy określić jego znak biorąc pod uwagę informacje pozamatematyczne
- c) Ma taki sam znak jak współczynniki a_1 i b_1
- d) Może mieć różny znak od współczynnika a_1

Pytanie 6

Mając dany współczynnik korelacji r_{xy} możemy obliczyć a_1 lub b_1 ze wzorów:

a) $a_1 = r_{xy} \frac{S(Y)}{S(X)}, b_1 = r_{xy} \frac{S(X)}{S(Y)},$

b) $a_1 = r_{xy} \frac{S(X)}{S(Y)}, b_1 = r_{xy} \frac{S(Y)}{S(X)},$

c) $a_1 = b_1 = r_{xy} \frac{S(Y)}{S(X)},$

d) $a_1 = r_{xy}^2, b_1 = r_{yx}^2.$

Pytanie 7

Reszty są to:

- a) ilorazy pomiędzy wartościami empirycznymi, a szacowanymi
- b) iloczyny pomiędzy wartościami empirycznymi, a szacowanymi
- c) sumy pomiędzy wartościami empirycznymi, a szacowanymi
- d) różnice pomiędzy wartościami empirycznymi, a szacowanymi

Pytanie 8

Współczynnik zbieżności φ^2 wyznaczamy ze wzoru:

a) $\varphi_{yx}^2 = \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{Y})}$

b) $\varphi_{yx}^2 = \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{X})^2}$

c) $\varphi_{yx}^2 = \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{Y})^2}$

d) $\varphi_{yx}^2 = \frac{\sum (x_i - \hat{y}_i)^2}{\sum (y_i - \bar{Y})^2}$



Pytanie 9

Współczynnik zbieżności ϕ_{xy}^2, ϕ_{yx}^2 wskazuje nam:

- a) stosunek wariancji resztowych dwóch cech,
- b) w ilu promilach zmiana jednej zmiennej nie została wyjaśniona regresją
- c) zależność cech niemierzalnych
- d) w ilu procentach zmiana jednej zmiennej nie została wyjaśniona regresją

Pytanie 10

W przypadku dwóch zmiennych X_1, X_2 mamy do czynienia z płaszczyzną regresji o równaniu:

- a) $x_i = a_0 + a_1 x_{1i} + a_2 x_{2i}$
- b) $\hat{y}_i = a_0 + a_1 x_{1i} + a_2 x_{2i}$
- c) $\hat{y}_i = a_0 - a_1 x_{1i} + a_2 x_{2i}$
- d) $\hat{y}_i = a_1 x_{1i} + a_2 x_{2i}$

Część 2: ZADANIA

I część zadania domowego (regresja zmiennych X i Y)

Zadanie 1

W poniższej tabeli zgrupowano liczbę zgłoszonych skarg przez klientów pewnej sieci komórkowej w ciągu trzech miesięcy wraz z ich średnimi z poprzedniego kwartału, uzyskując wyniki:

Liczba skarg Średnie	1	2	3
<2.0-3.0)		4	8
<3.0-4.0)	3	5	6
<4.0-5.0)	4	2	1

a) Wyznacz funkcję regresji I rodzaju zgłoszonych skarg w zależności od ich średnich z poprzedniego kwartału, a także narysuj empiryczną linię regresji.

b) Wyznacz równanie linii regresji II rodzaju skarg w zależności od średniej i naszkicuj na tym samym wykresie, co empiryczną linię regresji.

Zadanie 2

W poniższej tabeli przedstawiono następujące dane:

X	1	2	3	4	5
Y	100	154	233	324	424

a) Oszacuj funkcje regresji zmiennej Y względem X

b) Oblicz współczynnik korelacji liniowej

Zadanie 3

W poniższej tabeli przedstawiono następujące dane związane z ilością błędów popełnianych przez studentów na kolokwium z matematyki:

X ilość błędów	1	2	3	4	5	6
Y ilość studentów	898	174	143	426	167	152

a) Oszacuj funkcje regresji ilości studentów względem błędów.

b) Oblicz współczynnik korelacji liniowej i zinterpretuj go.

c) Ilu studentów popełniło by 7 błędów?

Zadanie 4

W oparciu o dane z poniższej tabeli oszacuj funkcje regresji i oblicz współczynnik korelacji liniowej:

X	1	2	3	4
Y	36	74	66	81

Zadanie 5

a) Wyznacz współczynnik indeterminacji funkcji regresji zależności zmiennej Y od zmiennej X korzystając z poniższych danych:

n_i	31	16	25	41	25	18	23
X	0-5	5-10	10-15	15-20	20-25	25-30	30-40
Y	5	24	36	11	16	41	65

b) Oblicz wariancję resztową funkcji regresji Y zależnego od X dla danych z poniższej tabeli:

n_i	23	14	27	20	53	52	33
X	0-10	10-20	20-30	30-40	40-50	50-60	60-70
Y	24	21	36	43	37	43	33



Zadanie 6

W oparciu o dane z tabelki, które przedstawiają różne liczby klientów salonu piękności w styczniu i odpowiadające im wielkości zamówień usługi masażu, wyznacz krzywą regresji liczby usług względem liczby klientów:

Liczba dni	11	6	5	4	2	1	2
Liczba klientów	0-5	5-10	10-15	15-20	20-25	25-30	30-40
Liczba usług	1	2	3	1	6	4	5

Zadanie 7

W oparciu o dane z tabelki wyznacz krzywą regresji zmiennej Y zależnej od X:

a)

n_i	21	12	32	19	24	17	42
X	0-5	5-10	10-15	15-20	20-25	25-30	30-40
Y	2	5	1	7	8	6	9

b)

n_i	32	11	42	22	24	42	56
X	0-10	10-20	20-30	30-40	40-50	50-60	60-70
Y	54	12	17	57	86	65	39

Zadanie 8

Dane z poniższej tabelki dotyczą liczby kwiatków na łące, które rozkwitały w maju oraz zależnej od niej ilości pszczoł jakie siadały na kwiatkach:

Liczba dni	2	4	6	3	5	1	10
Liczba kwiatów	0-5	5-10	10-15	15-20	20-25	25-30	30-40
Liczba pszczoł	18	9	10	15	7	26	35

Wyznacz krzywą regresji, wariancję i odpowiedz na pytanie: „W ilu procentach wzrost liczby pszczoł nie został wyjaśniony linią regresji?”



II część zadania domowego (korelacja i regresja wieloraka)

Zadanie 1

Badamy wzajemny wpływ na siebie trzech czynników:

- 1- poziom wody w morzu Czarnym
- 2- ilości opadów w Europie
- 3- ilości ryb w morzu Czarnym

Zmienne związane są ze sobą współczynnikami korelacji całkowitej, które wynoszą odpowiednio: $r_{12} = 0,76$, $r_{13} = 0,15$, $r_{23} = 0,06$. Sprawdź, jak mocno wpływają na siebie poziom wody w morzu Czarnym i ilość opadów w Europie pomijając wpływ na ten związek ilości ryb w morzu Czarnym.

Zadanie 2

Badamy wzajemny wpływ na siebie trzech czynników: pozytywne oceny studentów z kolokwium, samopoczucie studenta oraz pogodę.

Zmienne związane są ze sobą współczynnikami korelacji całkowitej, które wynoszą odpowiednio: $r_{12} = 0,69$, $r_{13} = 0,52$, $r_{23} = 0,33$. Sprawdź, jak mocno wpływają na siebie pozytywne oceny studentów z kolokwium i pogoda pomijając wpływ na ten związek samopoczucia studentów.

Uwaga: zakładamy, że zarówno pogodę jak i samopoczucie studenta można jakoś wyrazić liczbami.

Zadanie 3

Dla zmiennych z zadania 1 związanych ze sobą współczynnikami korelacji całkowitej:

$r_{12} = 0,11$, $r_{13} = 0,31$, $r_{23} = 0,42$ wyznacz współczynnik korelacji wielorakiej między ilością ryb w Morzu Czarnym, a pozostałymi czynnikami.

Zadanie 4

Badamy wzajemny wpływ na siebie trzech czynników: pozytywne oceny studentów z kolokwium, samopoczucie studenta oraz pogodę.

Zmienne związane są ze sobą współczynnikami korelacji całkowitej, które wynoszą odpowiednio: $r_{12} = 0,96$, $r_{13} = 0,1$, $r_{23} = 0,05$. Wyznacz współczynnik korelacji wielorakiej samopoczucia studenta w zależności od pozostałych czynników.

Uwaga: zakładamy, że zarówno pogodę jak i samopoczucie studenta można jakoś wyrazić liczbami.



Zadanie 5

Dane w tabeli przedstawiają zużycie prądu w gospodarstwie domowym w zależności od liczby zamieszkałych w nim osób i grubości kabli:

Zużycie prądu	90	160	280	450	550
Liczba mieszkańców	1	2	3	4	5
Grubość kabli	1,5	2	3	3,5	4,2

Wyznacz liniową funkcję regresji i współczynnik korelacji wielorakiej dla zużycia prądu zależnego od liczby mieszkańców i grubości kabli.

Zadanie 6

Dla danych w tabeli wyznacz liniową funkcję regresji, współczynnik korelacji wielorakiej i wariancję resztową dla Y zależnego od X_1 i X_2 :

Y	15	10	1	-6	4
X_1	1	-1	8	95	2
X_2	3	0	1	6,5	2

KONIEC