

*******Zadanka do poćwiczenia vol. 6** :)******
 ***** **regresja liniowa i inne typy zależności*******

1) Dany jest rozkład zmiennych X oraz Y

Y/X	-1	0	1
0	2	1	2
1	0	1	1
2	0	2	1

- a) Wyznacz regresję średnich zmiennej X względem zmiennej Y
 b) Oblicz wartość odpowiedniego dla tej regresji miernika siły zależności
 c) Wyznacz regresję liniową X względem Y
 d) Wyznacz wartość odpowiedniego dla regresji liniowej miernika siły zależności.

d) Wykorzystując w maksymalnym stopniu obliczenia z poprzednich punktów **wyznacz:**

- Średnią **przewidywania (optymalnego opisu)** Y za pomocą regresji średnich ze względu na X
- Wariancję błędu optymalnego przewidywania X bez wykorzystania informacji o zmiennej Y przy kwadratowej funkcji błędu
- Wariancję błędu przewidywania X za pomocą regresji średnich X|Y
- Średnią kwadratu błędu przewidywania X za pomocą regresji liniowej X|Y
- Sumę kwadratów błędów przewidywania X za pomocą regresji średnich X|Y
- Współczynnik korelacji liniowej pomiędzy standaryzowanymi zmiennymi X oraz Y
- Wariancję przewidywania X za pomocą regresji średnich X|Y
- Wariancję przewidywania X za pomocą regresji liniowej X|Y

(w każdym punkcie podaj wzór lub twierdzenie, do którego się odwołujesz)

2) Poniżej mamy niepełną macierz danych dla sześciu osób. Macierz obejmuje wartości zmiennych X i Y, wartości optymalnego opisu za pomocą regresji liniowej Y względem X, tj \hat{Y}_X i błędu przewidywania przy tym opisie $e_{Y|X}$. Ponadto wiemy, że $E(Y)=2$

	Y	X	\hat{Y}_X	$e_{Y X}$
1		0	1	
2	2	0		1
3		0		0
4		1		0
5		1		1
6		1		-1

- Uzupełnij powyższą tabelkę (o ile uważasz, że to możliwe, a jeśli się nie pomyliłem to jest możliwe!)
- Odtwórz równanie regresji liniowej Y względem X
- Oblicz miernik skorelowania liniowego $\rho^2_{Y,X}$ oraz na tej podstawie określ ile wynosić będzie kwadrat stosunku korelacyjnego (wskazówka: $\rho^2_{Y,X}$ można obliczyć szybko, nie trzeba np. wyznaczać kowariancji)

3) Opisujemy zbiorowość rodzin, w której występują wyłącznie rodziny bezdzietne z jednym dzieckiem i dwójgim dzieci. Wiemy też, że średnie warunkowe zmiennej Z - liczba zbitych bombek choinkowych w okresie świątecznym względem liczby posiadanych dzieci (W) wynoszą odpowiednio $E(Z|W=0)=1,5$ $E(Z|W=1)=3$ $E(Z|W=2)=4,5$ a kwadrat stosunku korelacyjnego Z względem W wynosi 0,4.

- Wyznacz równanie regresji liniowej Z względem W
- Ile wynosi współczynnik korelacji liniowej

4) Zmienne X i Y przyjmują wartości 1, 2, 3 i są niezależne stochastycznie. Zmienna Y ma rozkład symetryczny wokół wartości 2. Wyznacz równanie regresji Y względem X.

5) Mamy trzy zmienne Y - liczba **zaliczonych** sprawdzianów w semestrze
 Z - liczba **nie-zaliczonych** sprawdzianów w semestrze X - liczba zadań wykonanych zadań powtórkowych
 Regresja liniowa zmiennej Y względem zmiennej X przedstawia się wzorem $\hat{Y}_X = -1 + 0,2X$ a współczynnik korelacji Pearsona pomiędzy tymi zmiennymi wynosi 0,6.

- a) Wyznacz wartość współczynnika korelacji Pearsona pomiędzy zmiennymi X i Z
 b) Wiedząc, że w semestrze wszystkich sprawdzianów było aż pięć (ambitny prowadzący, trzeba przyznać!) wyznacz równanie regresji liniowej Z względem X

6) Zmienna X przyjmuje z niezerowymi częstościami wartości 1, 2, i 3. Błąd modalnej tej zmiennej wynosi 0,5. a $P(X=1)=P(X=2)$. Wiadomo też, że $\rho^2_{Y|X}=1$. a parametr regresji $b_{Y|X} = -2$. Wiadomo też, że $E(Y)=3,25$. Wyznacz rozkład łączny obydwu zmiennych.

7) Zmienna X przyjmuje wartości 1, 2, 3 a zmienna Y -1, 0, 1, 2 a $P(Y=-1)=0,3$. Wiemy, że zmienna X jest maksymalnie zależna stochastycznie od Y, ale Y nie jest maksymalnie zależna stochastycznie od X. Mamy zmienną $W=|Y|$ i wiemy, że regresja zmiennej X względem zmiennej W ma postać $X_W = 1+W$ a współczynnik korelacji liniowej jest równy 1. Wiedząc że rozkład zmiennej X jest symetryczny a jej wariancja jest równa 0,4 wyznacz rozkład łączny zmiennych X oraz Y.

Zmienna $Y = X^2$. Czy wynika z tego, że	
Zmienne są maksymalnie skorelowane liniowo	n
Miernik zależności modalnych theta zmiennej Y ze względu na X jest równy 1	t
Miernik zależności median dzeta zmiennej X ze względu na Y jest równy 1	n
Średnie warunkowe Y ze względu na X są sobie równe	n
Zmienna Y jest maksymalnie zależna stochastycznie od zmiennej X	t
Średnia kwadratu błędu przewidywania Y za pomocą regresji średnich ze względu na X jest równy 0	t

Współczynnik korelacji Pearsona pomiędzy zmiennymi X oraz Y jest równy -1. Czy wynika z tego że:	
Miernik zależności średnich $\eta^2_{X Y}$ przyjmuje wartość minimalną	n
Miernik zależności średnich $\eta^2_{Y X}$ przyjmuje wartość 1	t
Miernik zależności median $\zeta_{X Y}$ przyjmuje wartość minimalną	n
Miernik zależności stochastycznej zmiennej X ze względu na Y jest równy 1	t
Jeśli parametr $b_{Y X}$ jest równy 2 to parametr $b_{X Y}$ jest równy 1/2	t

Zmienna $Z=2 \cdot X+1$, a zmienna $W= -3 \cdot Y+1$. Czy wynika z tego, że	
$\rho^2_{Y X} = \rho^2_{W Z}$	t
$C(W,Z) = -6 \cdot C(X,Y)$	t
Wariancja zmiennej Z jest dwa razy większa od wariancji zmiennej W	n
$\rho^2_{Y X} = \rho^2_{Y Z}$	t
$\zeta_{Y W} = 1$	t
$D^2(W) = D^2[E(W X)]$	n

Zmienna Y oznacza liczbę wizyt w teatrze a zmienna X liczbę przeczytanych książek. Wiadomo, że równanie regresji Y względem X przyjmuje postać $\hat{Y}_x = 0,2 \cdot X + 1$ a równanie regresji X względem Y $\hat{X}_y = 0,8 \cdot Y + 2$. Czy wynika z tego, że:	
Kwadrat współczynnika korelacji Pearsona pomiędzy tymi zmiennymi wynosi 0,16	t
Osoba, która przeczytała pięć książek w ciągu roku chodzi do teatru 2 razy w roku.	n
Osoba, która nie czyta książek chodzi do teatru raz w roku	n
Zależność pomiędzy liczbą wizyt w teatrze a liczbą przeczytanych książek za pomocą regresji liniowej jest pozytywna	t
Jeśli $\text{Max}(X)=10$ to $\text{Max}(Y)=3$	n
Średnia błędów opisu zmiennej X za pomocą regresji średnich ze względu na zmienną Y jest równa 0	t

W pewnej zbiorowości równanie regresji liniowej zmiennej Y – liczba godzin poświęconych miesięcznie na oglądanie telewizji ze względu na zmienną X – miesięczne zarobki (w tys. zł), ma postać: $\hat{Y}_x = 100 - 10 \cdot X$. Średnia zarobków w tej zbiorowości wynosi 2 tys zł. Czy wynika z tego, że: że:	
Osoby, które zarabiają 3 tysiące zł oglądają telewizję średnio dłużej niż osoby, które zarabiają 2 tysiące zł.	n
Jeśli osoba A zarabia o tysiąc złotych więcej niż osoba B, to za pomocą regresji liniowej przewidujemy, że w ciągu miesiąca ogląda telewizję 10 godzin krócej niż osoba B	t
Średnia zmiennej Y wynosi 80	t
Osoby bez dochodów oglądają telewizję średnio 100 godzin w miesiącu	n

Mamy dwie zmienne X i Y. Współczynniki regresji liniowej wynoszą odpowiednio $b_{x y}=0,5$ i $b_{y x}=0,5$. czy wynika z tego, że:	
$\rho^2_{x y}=0,25$	t
Współczynniki $\eta^2_{x y}$ oraz $\eta^2_{y x}$ są sobie równe	n
Zestandaryzowany parametr $\beta_{x y} = \rho_{x,y} = 0,5$	t
Współczynnik $\eta^2_{x y}$ jest równy $\rho^2_{x y}$	n

Średnie zmiennej X ze względu na Y są takie same. Czy wynika z tego, że:	
Kowariancja pomiędzy obydwoma zmiennymi jest równa 0	t
$b_{y x} = b_{x y}$	t
$\eta^2_{x y} = 0$	t
$\eta^2_{y x} < 1$	n
Miernik zależności modalnych = 0	n
$E(X \cdot Y) = E(X) \cdot E(Y)$	t

Regresja II rodzaju zarobków Y (w tysiącach złotych) ze względu na płeć X (0-M, 1-K) przedstawia się wzorem: Y przew. = $2100 - 0,1 \cdot X$. Czy oznacza to, że:	
Regresja średnich zarobków ze względu na płeć jest liniowa	t
Regresja modalnych zarobków ze względu na płeć jest liniowa.	t
Przewidywana różnica w zarobkach pomiędzy kobietą a mężczyzną wynosi 100 zł.	t
$\eta^2_{x y} = \eta^2_{y x}$	n
Kowariancja zmiennych X i Y nie może być mniejsza od -0,025	t
Mężczyźni zarabiają średnio 2100 zł	t