

\*\*\*\*\***Zadanka na miłe zakończenie zajęć**\*\*\*\*\*

\*\*na pewno są tu również zadania innych prowadzących\*\*

1) Wiemy, że zmienna X przyjmuje trzy wartości 0, 1 i 2 a jej rozkład jest symetryczny. Regresja zmiennej Y względem zmiennej X dana jest wzorem  $\hat{Y}_X = 1 + X$ . Wiemy też, że  $\rho^2_{Y|X} = \eta^2_{Y|X} = 0,5$  a wariancja zmiennej Y jest równa 1,2. Oblicz  $P(X=1)$ .

2) Przebadano zbiorowość 100 studentów pewnego wydziału ze względu na następujące informacje: X-posiadanie komputera (0-nie posiada, 1-possada); Y- liczba posiadanych płyt muzycznych CD; Z-liczba posiadanych płyt DVD. Komputer posiada 60 % studentów a wariancja liczby posiadanych płyt CD wynosi 27 . Wiadomo również, że równanie regresji liniowej liczby posiadanych płyt CD ze względu na posiadanie komputera dana jest wzorem  $\hat{Y}_X = 12 + 5X$  a równanie regresji liniowej liczby posiadanych płyt DVD ze względu na posiadanie komputera dana jest wzorem  $Z_X = 2 + 4X$  a miernik siły zależności związany z tą regresją wynosi 0,5. Natomiast współczynnik korelacji liniowej pomiędzy liczbą posiadanych płyt CD a liczbą płyt DVD wynosi 0,8.

Wyznacz o ile uważasz, że to możliwe:

- Średnią i wariancję zmiennej W „łączna liczba posiadanych płyt CD i DVD”
- Regresję liniową zmiennej W względem zmiennej X i związany z tą regresją miernik siły zależności

3)Przebadano zbiorowość 80 studentów dziennych i 100 studentów wieczorowych ze względu na następujące informacje: X- posiadanie komputera (0-nie posiada, 1-possada); Y- liczba posiadanych płyt muzycznych CD. Komputer posiada połowa studentów dziennych i 40 % studentów wieczorowych. Wśród studentów dziennych wariancja liczby posiadanych płyt CD wynosi 27 a wśród studentów wieczorowych 36. Wiemy również, że równania regresji liniowej liczby posiadanych płyt CD ze względu na posiadanie komputera wynosi  $\hat{Y}_X = 10 + 8X$  wśród studentów dziennych i  $\hat{Y}_X = 10 + 10X$  wśród wieczorowych. Wyznacz o ile uważasz, że to możliwe:

- Średnią i wariancję zmiennej „liczba posiadanych płyt CD w połączonej zbiorowości studentów
- Regresję liniową zmiennej Y względem zmiennej X w połączonej zbiorowości studentów i związany z tą regresją miernik siły zależności

4) Przeprowadzono weryfikację hipotezy zerowej o wartości średniej w populacji  $H_0: \mu = 100$ , wykorzystując przy tym informację, że odchylenie standardowe w populacji  $D=100$ . Próba liczyła 400 osób. Przyjęto poziom istotności  $\alpha=0,01$ . Jaka mogłaby być treść prostej hipotezy konkurencyjnej, jeśli założymy, że prawdopodobieństwo błędu II rodzaju powinno być równe prawdopodobieństwu błędu I rodzaju.

5)Przeprowadzono badanie przedwyborcze na prostej próbie losowej, w którym 11% respondentów zadeklarowało, że zamierza głosować na kandydata L, zaś 9% respondentów zadeklarowało chęć głosowania na kandydata R. Na poziomie ufności 0,05 badacz odrzucił hipotezę głoszącą, że w populacji poparcie dla obu kandydatów jest takie samo na rzecz hipotezy, że jest ono różne. Jak duża musiała być próba, na której przeprowadzono badanie?

6) Równanie regresji w postaci standaryzowanej ma postać  $Z_{1|2,3} = 0,5 Z_2 + 0,25 Z_3$  ( $Z_1$  to zmienna  $Y_1$  po standaryzacji itd), a miernik korelacji liniowej wielokrotnej wynosi  $R^2_{1|2,3}=0,5$ . Średnia kwadratu błędu przewidywania  $Y_1$  za pomocą regr. liniowej względem  $Y_2$  i  $Y_3$  wynosi  $E(e^2_{1|2,3})= 2$ , a wariancja błędu przewidywania  $Y_1$  za pomocą regr. liniowej względem  $Y_3$  wynosi 2,5. Ponadto wiemy, że  $E(Y_1)=4$ ,  $E(Y_2)=2,5$ ,  $D^2(Y_2)=0,25$ , że  $E(Y_3)=5$ ,  $D^2(Y_3)=1$ . Wyznacz:

- Równanie w postaci niestandaryzowanej
- Współczynnik korelacji cząstkowej  $\rho^2_{1,2;3}$

7)Dany jest niepełny rozkład łączny zmiennych X i Y

X\Y	-2	-1	0	1
1	2	0	0	?
2	0	1	0	?
3	0	0	2	?
suma	2	1	2	3

- Uzupełnij tabelę w ten sposób, żeby miernik  $\rho^2$  między zmiennymi X oraz  $W=Y^2$  był jak największy
- Uzupełnij tabelę w ten sposób, żeby współczynnik tau-b między zmiennymi X oraz Y był mniejszy od 0
- Uzupełnij tabelę w ten sposób, żeby współczynnik gamma między zmiennymi X oraz Y był równy 1
- Uzupełnij tabelę w ten sposób aby zmaksymalizować wyrażenie  $E[H(X|Y)]$

8)Wiadomo, że  $E(X)=10$ , zaś  $D^2(X)=5$ , oraz  $X_Y = 4+0,8Y$   $\hat{Y}_X = 3,5+0,4X$ . Wyznacz średnią i wariancję Y,  $C(X,Y)$  oraz  $\rho^2_{X,Y}$ .

9) Z populacji, w której  $E(X)=10$  i  $D^2(X)=8$  wylosowano próbę 1600 elementową, a następnie próbę 900 elementową. Jakie jest prawdopodobieństwo, że różnica między średnimi w tych próbach będzie większa niż 0,5 ?

10) Badacz na poziomie istotności 0,02 weryfikuje  $H_0$ , zgodnie z którą  $E(X)=100$  względem  $H_1$  zgodnie z którą  $E(X) \neq 100$ . Wiadomo, że gdyby hipoteza konkurencyjna głosiła, że  $E(X)=110$ , to wynik weryfikacji byłby inny. Próba liczyła 400 osób, zaś średnia w próbie wynosiła 105. (a) Jakie byłyby wyniki weryfikacji przeciwko poszczególnym wersjom  $H_1$ ? (b) Co można powiedzieć [jakie wartości mogła przyjmować] o wariancji  $X$  ?

11) W wyniku badania na liczącej 400 rodzin próbie rodzin z małymi dziećmi stwierdzono, że średni tygodniowy czas poświęcany na zabawę z dzieckiem wynosi dla matek 16 h, przy wariancji  $16 h^2$ , a dla ojców 16,5 godziny, przy wariancji  $25 h^2$ , przy czym ilość czasu poświęcanego przez matkę jest skorelowany z ilością czasu poświęcanego przez ojca na poziomie  $r=0,3$ . Przyjmując poziom ufności  $\alpha=0,05$  zweryfikuj hipotezę głoszącą, że ojcowie i matki poświęcają średnio tyle samo czasu na zabawę z dzieckiem przeciwko hipotezie, że jest inaczej.

12) Mamy podany rozkład łączny częstości dwóch zmiennych  $X$  i  $Y$

$X \backslash Y$	-1	0	1
1	0,1	0,1	
2		0,4	0,2
3		0,1	0,1

a) Jakie parametry korelacji rangowej można obliczyć. Policz jeden z nich

b) Na jakich skalach muszą być mierzone zmienne  $X$  i  $Y$ , aby możliwe było policzenie korelacji rangowej

c) Na jakich skalach muszą być mierzone zmienne  $X$  i  $Y$ , aby możliwe było policzenie współczynnika niepewności

d) Czy na podstawie wartości współczynników obliczonych powyżej jesteśmy w stanie wyznaczyć wartości tych samych współczynników: dla zmiennych

$X$  i  $W$  gdzie  $W = -Y$

$X$  i  $W$  gdzie  $W = 2^Y$

$X$  i  $Z$  gdzie  $Z = Y^2$  ?

Odpowiedź uzasadnij

### Możliwe, że niektóre z poniższych zadań zamieściłem w innych miejscach.

(UWAGA: poniższe odpowiedzi wymagają sprawdzenia)

Zmienna zerowyjatkowa $X$ przyjmuje populacji wartość 1 z prawdopodobieństwem 0,5. Losujemy próby 10-elementowe w sposób prosty niezależny:	
Zmienna „średnia z próby” ma rozkład normalny	n
Wariancja średniej z próby zależy od liczebności próby	t
Wariancja średniej z próby zależy od liczebności populacji	n
Średnia wariancji z próby zależy od liczebności	t

Rozkład łączny częstości zmiennych $X$ oraz $Y$ jest taki sam w zbiorowości kobiet i w zbiorowości mężczyzn. Zbiorowości nie są równoliczne. Czy wynika z tego, że	
Wartości współczynników korelacji rangowej $\tau_A$ w obydwu zbiorowościach są takie same	n
Wartości współczynników korelacji rangowej $\tau_B$ w obydwu zbiorowościach są takie same	t
Wartości współczynników korelacji rangowej $\gamma$ w obydwu zbiorowościach są takie same	t
Zmienna $W = X * Y$ jest niezależna stochastycznie od zmiennej płęć	t

Współczynnik eta kwadrat zmiennej $X$ ze względu na zmienną $Y$ jest równy 0,5. Czy wynika z tego, że	
Współczynnik eta kwadrat zmiennej $Y$ ze względu na zmienną $X$ nie może być większy od 0,5.	n
$\rho^2_{X Y} \leq 0,5$	t
$\rho^2_{Y X} \leq 0,5$	t
Zmienne są zależne stochastycznie	t
$D^2(X) = 2 \cdot E[D^2(X Y)]$	t
Wszystkie mediany zmiennej $X$ mogą być sobie równe	t
Połowa średnich zmiennej $X$ ze względu na zmienną $Y$ jest sobie równa	n

Zmienne $X$ oraz $Y$ przyjmują wyłącznie wartości ujemne. Czy wynika z tego, że:	
Zależność liniowa pomiędzy zmiennymi ma kierunek negatywny	n
Kowariancja pomiędzy zmiennymi jest ujemna	n
Średnie obydwu zmiennych są ujemne	t
Średnia zmiennej $Z$ , która jest funkcją obydwu zmiennych jest mniejsza od 0	n

Wiemy, że $E[D^2(X Y)] = D^2[E(X Y)]$ . Czy wynika z tego, że:	
Miernik zależności średnich $\eta^2_{X Y}$ jest równy $\frac{1}{2}$	t
Miernik zależności średnich $\eta^2_{Y X}$ jest równy $\frac{1}{2}$	n
Miernik skorelowania liniowego $\rho^2_{Y X}$ jest mniejszy niż $\frac{1}{2}$	t
Średnie we wszystkich podzbiórowościach wyróżnionych ze względu na zmienną Y są równe	n
Miernik zależności modalnych jest mniejszy od 1	t

Znając kwadrat współczynnika korelacji wielokrotnej $R^2_{X Y,Z}$ oraz kwadrat współczynnika korelacji $\rho^2_{X Y}$ jesteśmy w stanie wyznaczyć:	
Kwadrat współczynnika korelacji cząstkowej $\rho^2_{X Y,Z}$	n
Kwadrat współczynnika korelacji cząstkowej $\rho^2_{X Z,Y}$	t
Kwadrat współczynnika korelacji cząstkowej $\rho^2_{Y X,Z}$	n
Współczynnik korelacji cząstkowej $\rho_{X Y,Z}$	n
Współczynnik korelacji cząstkowej $\rho_{Y Z,X}$	n

Regresja II rodzaju zarobków Y (w tysiącach złotych) ze względu na płeć X (0-M, 1-K) przedstawia się wzorem: $Y \text{ przew.} = 2100 - 0,1 * X$ . 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 musi być większa od -0,025	t
Mężczyźni zarabiają średnio 2100 zł	t

Wiemy, że kwadrat współczynnika korelacji wielokrotnej $R^2_{X Y,Z} = 0,3$ oraz, że kwadrat współczynnika korelacji $\rho^2_{X Y} = 0,3$ . Czy wynika z tego, że:	
$\rho^2_{X Z,Y} = 0$	t
$\rho^2_{X Y,Z} = 0$	n
$\rho^2_{X Z} = 0$	n
$\rho^2_{Z X,Y} = 0$	t

Czy jest prawdą, że:	
$E(s^2) = D^2(X)$	n
Średnia z próby jest nieobciążonym estymatorem średniej w populacji	t
Efektywny estymator średniej z próby to taki, który ma największą wariancję	n
$E(s^2) / D^2(X) = (n-1)/n$	t

Zmienna losowa X ma rozkład normalny. Czy wynika z tego, że	
Rozkład tej zmiennej jest symetryczny	t
Rozkład tej zmiennej jest jednomodalny	t
Wartość oczekiwana zmiennej jest równa 0 a jej odchylenie standardowe równe jest 1	n
Zmienna „średnia X z próby n-elementowej” musi mieć rozkład normalny	t
Zmienna dotyczy średniej z n-elementowej próby	n

Zmienna losowa X przyjmuje wartości z przedziału $(0;0,5)$ z jednostajnym prawdopodobieństwem, nie przyjmuje natomiast żadnych wartości spoza tego przedziału. Zatem	
Dystrybuanta X w punkcie -2 w ogóle nie jest określona (-2 nie należy do dziedziny dystrybuanty X)	n
Dystrybuanta X w punkcie 0,25 wynosi 0,5	t
Dystrybuanta X w punkcie -1 wynosi 0	t
Funkcja gęstości prawdopodobieństwa X w punkcie 2 może wynosić 1	n
Funkcja gęstości prawdopodobieństwa X w punkcie 0,25 może mieć wartość większą niż 1	t

O każdej próbie losowej można powiedzieć, że:	
Prawdopodobieństwo wylosowania każdego elementu z populacji jest znane	t
Prawdopodobieństwo wylosowania każdego elementu jest takie samo	n
Element populacji $\omega_1$ może zostać wylosowany do danej próby tylko raz	n
Średnia zmiennej z próby nie różni się znacząco od średniej z populacji	n

Badacz nie odrzucił hipotezy zerowej głoszącej, że średnia w populacji wynosi 5 na rzecz hipotezy że ta średnia jest różna od 5 na poziomie istotności równym 0,05. Czy:	
Możliwe jest, podjąłby inną decyzję gdyby taką samą średnią z próby uzyskał z próby o większej liczebności	t
Możliwe jest, podjąłby inną decyzję gdyby przyjął poziom istotności równy 0,03	n
Można określić prawdopodobieństwo beta	n
Bezwarunkowe prawdopodobieństwo popełnienia błędu I rodzaju określa poziom istotności	n