Statystyka matematyczna
Lista 5

1. Zmienna losowa (X, Y) ma rozkład prawdopodobieństwa podany w tabeli

$X \backslash Y$	2	4
-1	0.1	0.06
0	0.3	0.18
1	0.2	0.16

Obliczyć współczynnik korelacji.

2. Wektor losowy (X, Y) ma gęstość wyrażającą się wzorem

$$f(x,y) = \begin{cases} 0.5 & \text{dla} \quad (x,y) \in K, \\ 0 & \text{dla} \quad (x,y) \notin K, \end{cases}$$

gdzie K jest kwadratem o wierzchołkach $(1,0)$, $(0,1)$, $(-1,0)$, $(0,-1)$.

a). Wyznaczyć gęstości brzegowe.

b). Czy X i Y są niezależne?

c). Obliczyć $\mathrm{Cov}(X, Y)$.

3. Gęstość dwuwymiarowej zmiennej losowej (X, Y) dana jest wzorem

$$f(x,y) = \begin{cases} \frac{12}{11}(2x^2 + xy) & \text{dla} \quad x \in (0,1), y \in (0,1), \\ 0 & \text{poza tym.} \end{cases}$$

Obliczyć współczynnik korelacji.

4. Rozpatrzmy model regresji liniowej

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad i = 1, ..., n,$$

gdzie $\epsilon_1, ..., \epsilon_n$ są niezależne oraz $\mathrm{E}(\epsilon_i) = 0$, $\mathrm{Var}(\epsilon_i) = \sigma^2$, $i = 1, ..., n$. Ponadto zakładamy, że istnieją $i, j \in \{1, ..., n\}$ takie, że $x_i \neq x_j$.

a). Wykazać, że estymatory parametrów β_0 i β_1 wyznaczone metodą najmniejszych kwadratów mają postać

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^n (x_i - \overline{x})^2}, \quad \widehat{\beta}_0 = \overline{y} - \widehat{\beta}_1 \overline{x},$$

gdzie $\overline{x} = \sum_{i=1}^n x_i / n$, $\overline{y} = \sum_{i=1}^n y_i / n$.

b). Pokazać, że

$$\widehat{\beta}_1 = r \frac{s_y}{s_x}, \quad \text{gdzie} \quad r = \frac{\sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y})}{(n-1) s_x s_y}.$$

Przypomnijmy, że r oznacza wartość współczynnika korelacji próbkowej a s_x i s_y to odchylenia standardowe wartości $x_1, ..., x_n$ oraz $y_1, ..., y_n$, odpowiednio.

5. Rozpatrzmy model regresji z pominięciem wyrazu stałego

$$Y_i = \beta_1 x_i + \epsilon_i, \quad i = 1, ..., n,$$

gdzie $\epsilon_1, ..., \epsilon_n$ są niezależnymi błędami o rozkładzie normalnym $\mathcal{N}(0, \sigma)$.

a). Wykazać, że estymator współczynnika kierunkowego β_1 wyznaczony metodą najmniejszych kwadratów ma postać

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2}.$$

b). Wyznaczyć estymator parametru σ^2 metodą największej wiarogodności.

6. Rozpatrzmy rezultaty kolokwium (na skali od 0 do 25 punktów) i egzaminu końcowego (na skali od 0 do 50 punktów) ze statystyki matematycznej. W kolokwium i egzaminie brało udział 10 studentów (wyniki w tabeli).

Numer studenta	Kolokwium	Egzamin
1	7	20
2	11	24
3	12	25
4	14	30
5	17	35
6	15	30
7	21	43
8	22	42
9	19	41
10	13	24

a). Sporządzić wykres rozproszenia, odpowiadający tym danym. Czy na podstawie tego rysunku można stwierdzić, że istnieje liniowa zależność między wynikiem egzaminu a wynikiem kolokwium?

b). Obliczyć korelację między wynikiem kolokwium a wynikiem egzaminu końcowego.

c). Za pomocą metody najmniejszych kwadratów wyznaczyć równanie prostej regresji, opisującej zależność wyniku egzaminu od wyniku kolokwium.

7. Rodzicom pięcioletniej Kasi wydaje się, że ich córka rośnie zbyt wolno. W poniższej tabeli podano informacje dotyczące wzrostu dziecka:

Wiek (w miesiącach)	36	48	51	54	57	60
Wzrost (w cm)	86	90	91	93	94	95

a). Sporządzić wykres rozproszenia, odpowiadający tym danym. Czy na podstawie tego rysunku można stwierdzić, że istnieje liniowa zależność między wzrostem a wiekiem Kasi?

b). Za pomocą metody najmniejszych kwadratów wyznaczyć równanie prostej regresji, opisującej zależność wzrostu Kasi od jej wieku.

c). Wykorzystując otrzymany model, podaj prognozę wzrostu Kasi w wieku 40 i 60 miesiący.

d). O ile średnio rośnie Kasia w ciągu miesiąca? Prawidłowo rozwijająca się dziewczynka przyrasta o 6 centymetrów między czwartym a piątym rokiem życia. Czy Kasia rośnie wolniej niż powinna?

8. Zbadano średnią grubość pni pewnego gatunku drzew w zależności od średniej temperatury otoczenia w ciągu dnia. Uzyskano następujące wyniki:

Temperatura (w $^\circ C$)	15	20	25	30
Grubość pni (w cm)	25	26	28	29

Za pomocą metody najmniejszych kwadratów skonstruować prostą regresji zależności grubości pni od temperatury otoczenia. Wykorzystując otrzymany model znaleźć temperaturę, dla której prognozowana grubość pni osiągnie 30cm.

9. Za pomocą modelu regresji liniowej zbadano zależność między temperaturą powierzchni chodnika (x) a jego wygięciem (y). Na podstawie $n = 20$ pomiarów wyznaczono następujące wielkości:

$$\sum_{i=1}^{n} y_i = 12.75, \quad \sum_{i=1}^{n} y_i^2 = 8.86, \quad \sum_{i=1}^{n} x_i = 1478, \quad \sum_{i=1}^{n} x_i^2 = 143,215.8, \quad \sum_{i=1}^{n} x_i y_i = 1083.67.$$

a). Wyznaczyć estymatory parametrów β_0 i β_1.

b). Narysować prostą regresji.

c). Wykorzystać otrzymany model do prognozy wygięcia chodnika przy temperaturze $85^\circ F$.

d). Jakie jest średnie wygięcie chodnika mającego temperaturę $90^\circ F$?

e). Jaka zmiana średnigo wygięcia chodnika odpowiada zmianie temperatury jego powierzchni o $1^\circ F$?