Zadania z Analizy danych

zestaw 4

Zadanie 1

Znajdź funkcję generującą momenty rozkładu Poissona.

Wskazówka:

W przypadku rozkładów dyskretnych jakim jest rozkład Poissona funkcję generującą momenty wyznaczamy ze wzoru $\varphi(t) = \sum_{k=0}^{\infty} e^{kt} P(X = k)$. Następnie skorzystaj z tożsamości: $\sum_{k=0}^{\infty} e^{tk} \frac{\lambda^k}{k!} = e^{\lambda e^t}$.

$$\varphi(t) = \sum_{k=0}^{\infty} e^{tk} \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} e^{\lambda e^t} = e^{\lambda(e^t - 1)}$$

Zadanie 2

Korzystając z wyniku poprzedniego zadania oraz własności funkcji generującej momenty, znajdź wartości oczekiwane $E(k)$ oraz $E(k^2)$ rozkładu Poissona $\frac{e^{-\lambda}\lambda^k}{k!}$. Następnie, korzystając z obu wyników wylicz wariancję rozkładu Poissona.

$$\varphi'(t) = \frac{d\varphi(t)}{dt} = \lambda e^t e^{\lambda(e^t - 1)},$$

$$\varphi''(t) = \frac{d^2\varphi(t)}{dt^2} = \lambda e^t \cdot e^{\lambda(e^t-1)} + \lambda e^t \lambda e^t e^{\lambda(e^t-1)} = \lambda e^t e^{\lambda(e^t-1)}(1 + \lambda e^t)$$

$$\varphi'(0) = \hat{k} = \lambda e^0 e^{\lambda(e^0-1)} = \lambda.$$

$$\varphi''(0) = \widehat{k^2} = (1 + \lambda e^0)\lambda e^0 e^{\lambda(e^0-1)} = (1 + \lambda)\lambda$$

$$\sigma^2(k) = \widehat{k^2} - \left(\hat{k}\right)^2 = (1 + \lambda)\lambda - \lambda^2 = \lambda.$$

Zadanie 3

Znajdź funkcję generującą momenty rozkładu Gaussa.

$$\varphi(t) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \begin{pmatrix} u = \frac{x-\mu}{\sqrt{2}\sigma} \\ dx = \sqrt{2}\sigma du \end{pmatrix} = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{t(u\sqrt{2}\sigma+\mu)} e^{-u^2} \sqrt{2}\sigma du =$$

$$\frac{1}{\sqrt{\pi}} e^{t\mu} \int_{-\infty}^{\infty} e^{-u^2+ut\sqrt{2}\sigma} du = \frac{1}{\sqrt{\pi}} e^{t\mu} \int_{-\infty}^{\infty} e^{-u^2+ut\sqrt{2}\sigma-\frac{\sigma^2t^2}{2}+\frac{\sigma^2t^2}{2}} du = \frac{1}{\sqrt{\pi}} e^{t\mu} \int_{-\infty}^{\infty} e^{-\left(u-\frac{\sigma t}{\sqrt{2}}\right)^2+\frac{\sigma^2t^2}{2}} du =$$

$$\frac{1}{\sqrt{\pi}} e^{t\mu+\frac{\sigma^2t^2}{2}} \int_{-\infty}^{\infty} e^{-\left(u-\frac{\sigma t}{\sqrt{2}}\right)^2} du = \frac{1}{\sqrt{\pi}} e^{t\mu+\frac{\sigma^2t^2}{2}} \int_{-\infty}^{\infty} e^{-w^2} dw = \frac{1}{\sqrt{\pi}} e^{t\mu+\frac{\sigma^2t^2}{2}} \sqrt{\pi}.$$

$$\varphi(t) = e^{\frac{2t\mu+\sigma^2t^2}{2}}.$$

Zadanie 4

Korzystając z wyniku poprzedniego zadania oraz własności funkcji generującej momenty, znajdź wartości oczekiwane $E(X)$ oraz $E(X^2)$ rozkładu Gaussa. Następnie, korzystając z obu wyników wylicz wariancję rozkładu Gaussa.

$$\varphi'(t) = \frac{d\varphi(t)}{dt} = (\mu + \sigma^2 t)e^{\frac{2t\mu+\sigma^2 t^2}{2}}.$$

$$\varphi''(t) = \frac{d^2\varphi(t)}{dt^2} = \sigma^2 e^{\frac{2t\mu+\sigma^2 t^2}{2}} + (\mu + \sigma^2 t)(\mu + \sigma^2 t)e^{\frac{2t\mu+\sigma^2 t^2}{2}} =$$

$$e^{\frac{2t\mu+\sigma^2 t^2}{2}}(\sigma^2 + (\mu + \sigma^2 t)^2)$$

$$\varphi'(0) = \hat{x} = \mu.$$

$$\varphi''(0) = \widehat{x^2} = \sigma^2 + \mu^2.$$

$$\sigma^2(X) = \widehat{X^2} - \left(\hat{X}\right)^2 = \sigma^2 + \mu^2 - \mu^2 = \sigma^2.$$

Zadania 5

Przedstaw graficznie dane zebrane w poniższej tabeli. Jakiego współczynnika korelacji między zmiennymi x, y spodziewasz się? Wylicz kowariancję oraz współczynnik korelacji tych zmiennych.

x	0	0,5	1	1,5	2	2,5	3	3,5	4	4,5
y	3,14	4,42	6,21	7,64	8,01	10,18	12,81	13,51	15,46	16,22

Wskazówka

Jako estymatora kowariancji użyj

$$T\big(cov(X,Y)\big) = \frac{\sum_{i=1}^{n}(x_i - \hat{x})(y_i - \hat{y})}{n-1}.$$

Jako estymatora współczynnika korelacji użyj

$$T\big(\rho(X,Y)\big) = \frac{T\big(cov(X,Y)\big)}{\sigma(X)\sigma(Y)}$$

lub, co jest równoważne

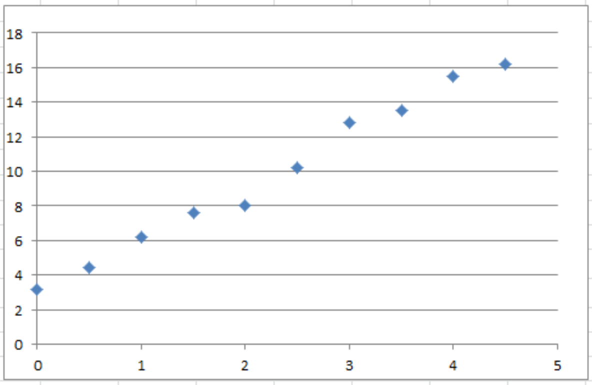
$$T\big(\rho(X,Y)\big) = \frac{\sum_{i=1}^{n}(x_i - \hat{x})(y_i - \hat{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \hat{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \hat{y})^2}}$$

Otrzymane wyniki porównaj z wartościami zwracanymi przez funkcje EXCELA
=KOWARIANCJA.PRÓBKI(Tablica_X;Tablica_Y)
=WSP.KORELACJI(Tablica_X;Tablica_Y)

l.p.	x	y	(xi-x_sr)*(yi-y_sr)	(xi-x_sr)^2	(yi-y_sr)^2
1	0	3,14	14,895	5,0625	43,8244
2	0,5	4,42	9,345	3,0625	28,5156
3	1	6,21	4,4375	1,5625	12,6025
4	1,5	7,64	1,59	0,5625	4,4944
5	2	8,01	0,4375	0,0625	3,0625
6	2,5	10,18	0,105	0,0625	0,1764
7	3	12,81	2,2875	0,5625	9,3025
8	3,5	13,51	4,6875	1,5625	14,0625
9	4	15,46	9,975	3,0625	32,49
10	4,5	16,22	14,535	5,0625	41,7316
			62,295	20,625	190,2624

$x_{śr}$	2,25
$y_{śr}$	9,76
σ(x)	1,5138252
σ(y)	4,5978546

suma((xi-xśr)(yi-yśr))/(n-1)	6,921667		
=KOWARIANCJA.PRÓBKI(B3:B12;C3:C12)	6,921667		
cov(X,Y)/(σ(X)σ(Y))	0,994443	0,994443	<--- suma((xi-xśr)(yi-yśr))/Pierwiastek((xi-xśr)^2(yi-yśr)^2)
=WSP.KORELACJI(B3:B12;C3:C12)	0,994443		



Zadanie 6

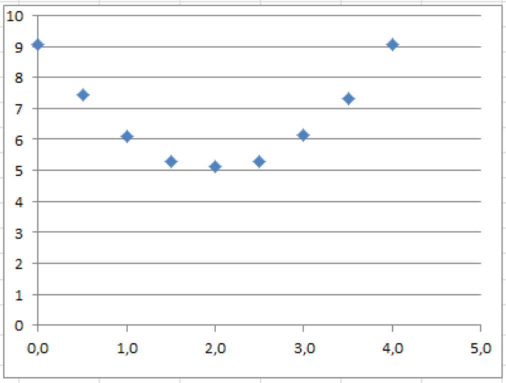
Przedstaw graficznie dane zebrane w poniższej tabeli. Jakiego współczynnika korelacji między zmiennymi x, y spodziewasz się? Wylicz ten współczynnik. Jak uzasadnisz otrzymany wynik?

x	0	0,5	1	1,5	2	2,5	3	3,5	4
y	9,08	7,43	6,08	5,29	5,12	5,27	6,13	7,33	9,05

l.p.	x	y	(xi-x_sr)*(yi-y_sr)	(xi-x_sr)^2	(yi-y_sr)^2
1	0,0	9,08	-4,653	4,00	5,413
2	0,5	7,43	-1,015	2,25	0,458
3	1,0	6,08	0,673	1,00	0,453
4	1,5	5,29	0,732	0,25	2,141
5	2,0	5,12	0,000	0,00	2,668
6	2,5	5,27	-0,742	0,25	2,200
7	3,0	6,13	-0,623	1,00	0,389
8	3,5	7,33	0,865	2,25	0,333
9	4,0	9,05	4,593	4,00	5,275
			-0,170	15,00	19,330

$x_{śr}$	2,0
$y_{śr}$	6,75333
σ(X)	1,36931
σ(Y)	1,55442

suma((xi-xśr)(yi-yśr))/(n-1)	-0,021250		
=KOWARIANCJA.PRÓBKI(B3:B11;C3:C11)	-0,021250		
cov(X,Y)/(σ(X)σ(Y))	-0,009984	-0,009984	<--- suma((xi-xśr)(yi-yśr))/Pierwiastek((xi-xśr)^2(yi-yśr)^2)
=WSP.KORELACJI(B3:B11;C3:C11)	-0,009984		



Zadanie 7

Gęstość prawdopodobieństwa dla dwu zmiennych losowych X i Y jest dana wzorem

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y} e^{-\frac{1}{2}\frac{x^2}{\sigma_x^2}-\frac{1}{2}\frac{y^2}{\sigma_y^2}}.$$

Jest to dwuwymiarowy rozkład Gaussa z $\hat{x} = 0$ i $\hat{y} = 0$.

a) Znajdź brzegowe gęstości prawdopodobieństwa $f(x)$ i $f(y)$.
b) Pokaż, że zmienne X, Y są niezależne

$$g(y) = \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma_x\sigma_y} e^{-\frac{1}{2}\frac{x^2}{\sigma_x^2}-\frac{1}{2}\frac{y^2}{\sigma_y^2}} dx = \frac{1}{2\pi\sigma_x\sigma_y} e^{-\frac{1}{2}\frac{y^2}{\sigma_y^2}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\frac{x^2}{\sigma_x^2}} dx = \frac{1}{2\pi\sigma_x\sigma_y} e^{-\frac{1}{2}\frac{y^2}{\sigma_y^2}} \cdot \sqrt{2}\sigma_x\sqrt{\pi}$$

$$= \frac{1}{\sqrt{2\pi}\sigma_y} e^{-\frac{1}{2}\frac{y^2}{\sigma_y^2}}.$$

$$h(x) = \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma_x\sigma_y} e^{-\frac{1}{2}\frac{x^2}{\sigma_x^2}-\frac{1}{2}\frac{y^2}{\sigma_y^2}} dy = \frac{1}{2\pi\sigma_x\sigma_y} e^{-\frac{1}{2}\frac{x^2}{\sigma_x^2}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\frac{y^2}{\sigma_y^2}} dy = \frac{1}{2\pi\sigma_x\sigma_y} e^{-\frac{1}{2}\frac{x^2}{\sigma_x^2}} \cdot \sqrt{2}\sigma_y\sqrt{\pi}$$

$$= \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{1}{2}\frac{x^2}{\sigma_x^2}}.$$

$$f(x,y) = g(x)h(x).$$

Zadanie 8

Dwie zmienne losowe zdefiniowane są jako $X = \sin Z$, $Y = \cos Z$, gdzie zmienna Z jest zmienną losową o rozkładzie jednostajnym na przedziale $[0, \pi]$. Jak widać zmienne X, Y są zależne, gdyż $x^2 + y^2 = \sin^2 z + \cos^2 z = 1$. Pokaż, że pomimo tego, kowariancja jest równa zeru.

Wskazówka:

Zamiast liczyć kowariancję z definicji skorzystaj z zależności: $cov(X,Y) = E(X \cdot Y) - E(X) \cdot E(Y)$.

$$E(X) = \frac{1}{\pi}\int_0^{\pi} \sin Z\, dz = -\frac{1}{\pi}\cos Z\Big|_0^{\pi} = \frac{1}{\pi}(-\cos\pi + \cos 0) = \frac{1}{\pi}(1+1) = \frac{2}{\pi}.$$

$$E(Y) = \frac{1}{\pi}\int_0^{\pi} \cos Z\, dz = \frac{1}{\pi}\sin Z|_0^{\pi} = \frac{1}{\pi}(\sin\pi - \sin 0) = \frac{1}{\pi}(0-0) = 0.$$

$$E(X \cdot Y) = \frac{1}{\pi}\int_0^{\pi} \sin Z \cos Z\, dz = \frac{1}{2\pi}\sin^2 Z|_0^{\pi} = \frac{1}{2\pi}(\sin^2\pi - \sin^2 0) = \frac{1}{2\pi}(0-0) = 0.$$

$$cov(X,Y) = 0 - \frac{2}{\pi} \cdot 0 = 0.$$