

Rachunek prawdopodobieństwa kombinatoryka

$$W_n^k = n^k \quad V_n^k = \frac{n!}{(n-k)!}$$

$$C_n^k = \binom{n}{k} = \frac{n!}{k!(n-k)!}$$

prawdopodobieństwo warunkowe

$$P(A/W) = \frac{P(A \cap W)}{P(W)}$$

prawdopodobieństwo całkowite

$$P(A) = P(A/A_1) \cdot P(A_1) + P(A/A_2) \cdot P(A_2) + \dots + P(A/A_n) \cdot P(A_n)$$

wzór Bayes'a na wielkość prawdopodobieństwa a'posteriori

$$P(A_k/A) = \frac{P(A_k) \cdot P(A/A_k)}{\sum_{i=1}^n P(A_i) \cdot P(A/A_i)}$$

Zmienne losowe

wartość oczekiwana

$$EX = \sum_{i=1}^k x_i \cdot p_i$$

wariancja

$$D^2X = \sum_i (x_i - EX)^2 p_i = \sum_i x_i^2 p_i - (EX)^2$$

kowariancja

$$CXY = \sum_{i,j} (x_i - EX) \cdot (y_j - EY) \cdot p_{ij} = EXY - EX \cdot EY$$

wartość oczekiwana i wariancja sumy zmiennych Z : Z = aX + bY , W = cX + dY

$$EZ = a \cdot EX + b \cdot EY$$

$$D^2Z = a^2 D^2X + b^2 D^2Y + 2ab CXY$$

$$CXZ = aD^2X + bCXY$$

$$CZW = acD^2X + bcCXY + adCXY + bdD^2Y$$

Rozkłady zmiennych losowych - przykłady

rozkład Bernoulli'ego (dwumianowy)

$$X \sim B(p) \quad P(X=k) = P_{n,p}(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

$$EX = np$$

$$D^2X = np(1-p)$$

				1						
			1	1						
		1	2	1						
	1	3	3	1						
1	4	6	4	1						
1	5	10	10	5	1					
1	6	15	20	15	6	1				
1	7	21	35	35	21	7	1			
1	8	28	56	70	56	28	8	1		
1	9	36	84	126	126	84	36	9	1	
1	10	45	120	210	252	210	120	45	10	1

rozkład Poisson'a

$$X \sim P(\lambda) \quad P_\lambda(k) = \frac{\lambda^k}{k!} e^{-\lambda} \quad \lambda = np = EX = D^2X$$

rozkład normalny i standaryzacja

$$X \sim N(\mu; \sigma) \quad U = \frac{X-\mu}{\sigma}$$

Próba

średnia arytmetyczna

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i \quad \bar{x} = \frac{1}{N} \sum_{j=1}^k x_j \cdot n_j = \sum_{j=1}^k x_j \cdot w_j$$

wariancja

$$S_x^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2 = \frac{1}{N-1} \left(\sum_{i=1}^N x_i^2 - \frac{1}{N} \left(\sum_{i=1}^N x_i \right)^2 \right)$$

$$S_x^2 = \frac{1}{N-1} \sum_{j=1}^k (x_j - \bar{x})^2 \cdot n_j = \frac{1}{N-1} \left(\sum_{j=1}^k x_j^2 \cdot n_j - \frac{1}{N} \left(\sum_{j=1}^k x_j \cdot n_j \right)^2 \right)$$

odchylenie standardowe

$$S_x = \sqrt{S_x^2}$$

wskaznik zmienności

$$V_x = \frac{S_x}{\bar{x}} \cdot 100\%$$

współczynnik skośności

$$A'' = \frac{\bar{x} - M_o}{S}$$

kowariancja

$$\text{cov}_{xy} = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x}) \cdot (y_i - \bar{y}) = \frac{1}{N-1} \left(\sum_{i=1}^N x_i \cdot y_i - \frac{\sum_{i=1}^N x_i \cdot \sum_{i=1}^N y_i}{N} \right)$$

współczynnik korelacji Pearsona

$$r = \frac{\text{cov}_{xy}}{S_x \cdot S_y}$$

współczynnik korelacji Spearmana

$$r_s = 1 - \frac{6 \cdot \sum_{i=1}^n d_i^2}{N \cdot (N^2 - 1)}$$

Przedziały ufności

$$P\left(\bar{x} - t_{\alpha, \nu} \frac{S}{\sqrt{N}} \leq \mu \leq \bar{x} + t_{\alpha, \nu} \frac{S}{\sqrt{N}}\right) = 1 - \alpha \quad \nu = N - 1$$

$$P\left(\bar{x} - u_\alpha \frac{\sigma}{\sqrt{N}} < \mu < \bar{x} + u_\alpha \frac{\sigma}{\sqrt{N}}\right) = 1 - \alpha$$

$$P\left(\bar{x} - u_\alpha \frac{S}{\sqrt{N}} < EX < \bar{x} + u_\alpha \frac{S}{\sqrt{N}}\right) = 1 - \alpha$$

$$P\left(w - u_\alpha \sqrt{\frac{w(1-w)}{N}} < p < w + u_\alpha \sqrt{\frac{w(1-w)}{N}}\right) = 1 - \alpha$$

Regresja

współczynniki regresji

$$b_{xy} = \frac{\text{COV}_{xy}}{S_y^2} \quad b_{yx} = \frac{\text{COV}_{xy}}{S_x^2}$$

stałe regresji

$$a_{xy} = \bar{x} - b_{xy} \cdot \bar{y} \quad a_{yx} = \bar{y} - b_{yx} \cdot \bar{x}$$

reszty

$$e_i = y_i - \hat{y}_i$$

równość wariancyjna:

$$SST = SSR + SSE$$

$$SST = \sum y_i^2 - \frac{(\sum y_i)^2}{N} \quad SSR = \sum \hat{y}_i^2 - \frac{(\sum \hat{y}_i)^2}{N} \quad SSE = \sum e_i^2$$

współczynnik determinacji funkcji regresji

$$R^2 = \frac{SSR}{SST}$$

istotność funkcji regresji

$$F_{emp} = \frac{SSR}{SSE} \cdot \frac{N-k}{k-1} = \frac{R^2}{1-R^2} \cdot \frac{N-k}{k-1}$$

$$v_1 = k-1 \quad v_2 = N-k$$

wielkość błędu predykcji dla wybranej wartości x_0

$$S_{\hat{y}_0} = S_e \cdot \sqrt{1 + \frac{1}{N} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^N (x_i - \bar{x})^2}} \quad S_e = \sqrt{\frac{SSE}{N-k}}$$

Ocena niezależności cech – test niezależności

H_0 : brak zależności między cechami

$$\chi^2_{emp} = \sum_{i,j} \frac{(n_{ij} - p_{ij} \cdot N)^2}{p_{ij} \cdot N} \quad v = (k-1)(l-1)$$

współczynnik zbieżności Cramera:

$$V_C = \sqrt{\frac{\chi^2_{emp}}{N \cdot (\min(k-1)(l-1))}}$$

Weryfikacja hipotez

$H_0 : \mu = \mu_0 / EX = EX_0$

$$t_{emp} = \frac{\bar{x} - \mu_0}{S} \sqrt{N} \quad v = N-1$$

$$u_{emp} = \frac{\bar{x} - \mu_0}{\sigma} \sqrt{N} \quad u_{emp} = \frac{\bar{x} - EX_0}{S} \sqrt{N}$$

$H_0 : \mu_1 = \mu_2 / EX_1 = EX_2$

$$t_{emp} = \frac{\bar{x}_1 - \bar{x}_2}{S_{\bar{x}_1 - \bar{x}_2}} \quad v = N_1 + N_2 - 2 \quad u_{emp} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{S_1^2}{N_1} + \frac{S_2^2}{N_2}}}$$

$H_0 : \sigma^2 = \sigma_0^2$

$$\chi^2_{emp} = \frac{(N-1)S^2}{\sigma_0^2} \quad v = N-1$$

$H_0 : \sigma_1^2 = \sigma_2^2$

$$F_{emp} = \frac{S_1^2}{S_2^2} \quad F_{emp} \geq 1 \quad v_1 = N_1 - 1 \quad v_2 = N_2 - 1$$

$H_0 : p = p_0$

$$u_{emp} = \frac{w - p_0}{\sqrt{\frac{p_0(1-p_0)}{N}}}$$

$H_0 : p_1 = p_2$

$$u_{emp} = \frac{w_1 - w_2}{\sqrt{\frac{\bar{w}(1-\bar{w})}{N}}} \quad \bar{w} = \frac{n_1 + n_2}{N_1 + N_2} \quad \bar{N} = \frac{N_1 \cdot N_2}{N_1 + N_2}$$

$H_0 : \rho_{XY} = 0$

$$t_{emp} = \frac{r}{S_r} \quad S_r = \sqrt{\frac{1-r^2}{N-2}} \quad v = N-2$$

H_0 : rozkład cechy jest zgodny z teoretycznym

$$\chi^2_{emp} = \sum_i \frac{(n_i - p_i \cdot N)^2}{p_i \cdot N} \quad v = k-1-m$$

$$W_{emp} = \frac{\left[\sum_{i=1}^{\lfloor \frac{N}{2} \rfloor} \mathbf{a}_{N,i} \cdot (x_{(N-i+1)} - x_i) \right]^2}{\sum_{i=1}^N (x_i - \bar{x})^2}$$