Rachunek prawdopodobieństwa

kombinatoryka

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

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

prawdopodobieństwo warunkowe

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

prawdopodobieństwo całkowite

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

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

$$D^2X = \sum (x_i - EX)^2 p_i = \sum x^2\, p_i - (EX)^2$$

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

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

$$D^2Z = a^2D^2X + b^2D^2Y + 2abCXY$$

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

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

Rozkłady zmiennych losowych - przykłady

rozkład Bernoulli'ego (dwumianowy)

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

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

rozkład Poisson'a

$X \sim P(\lambda)$ $\quad P_\lambda(k) = \frac{\lambda^k}{k!} 2{,}718^{-\lambda}$ $\qquad \lambda = np = EX = D^2X$

rozkład normalny i standaryzacja

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

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

Próba – statystyka opisowa

średnia arytmetyczna

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

wariancja

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

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

kowariancja

$$cov_{xy} = \frac{1}{N-1}\sum (x_i - \bar{x})\cdot(y_i - \bar{y}) = \frac{1}{N-1}\left(\sum x_i\cdot y_i - \frac{\sum x_i \cdot \sum y_i}{N}\right)$$

odchylenie standardowe **wskaźnik zmienności**

$$S_x = \sqrt{S_x^2} \qquad V_x = \frac{S_x}{\bar{x}}\cdot 100\%$$

kwantyl rzędu q obliczany na postawie szeregu rozdzielczego

$$x_q = x_{0q} + (q - F_{N(q-1)})\frac{x_{1q} - x_{0q}}{w_q}$$

wartość modalna obliczana na podstawie szeregu rozdzielczego

$$M_o = x_{0M_o} + \frac{n_{M_o} - n_{M_o-1}}{2n_{M_o} - n_{M_o-1} - n_{M_o+1}}\cdot\left(x_{1M_o} - x_{1M_o}\right)$$

współczynnik asymetrii (skośności)

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

Przedziały ufności

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

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

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

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

Analiza dynamiki

$$i_{t/t_*} = \frac{y_t}{y_{t_*}}\cdot 100\% \qquad \bar{r} = \sqrt[k-1]{\prod_{a=2}^{k} i_{a/a-1}} = \sqrt[k-1]{i_{k/1}}$$

Regresja prosta $x = b_{xy} \cdot y + a_{xy}$ lub $y = b_{yx} \cdot x + a_{yx}$

współczynniki regresji

$$b_{xy} = \frac{cov_{xy}}{S_y^2} \qquad b_{yx} = \frac{cov_{xy}}{S_x^2}$$

stałe regresji

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

reszty

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

równość wariancyjna:

SST = SSR + SSE

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

współczynnik determinacji (współczynnik dopasowania) funkcji regresji

$$R^2 = \frac{SSR}{SST} = b_{xy} \cdot b_{yx} = r_{xy}^2$$

istotność funkcji regresji

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

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

wielkość błędu predykcji dla wybranej wartości x_0

$$\boldsymbol{S_{\hat{y}_0}} = \boldsymbol{S_e} \cdot \sqrt{\mathbf{1} + \frac{\mathbf{1}}{\boldsymbol{N}} + \frac{(\boldsymbol{x_0} - \bar{\boldsymbol{x}})^2}{\sum(\boldsymbol{x_i} - \bar{\boldsymbol{x}})^2}} \qquad S_e = \sqrt{\frac{SSE}{N-k}}$$

Ocena niezależności cech – test niezależności

H_0 : <u>brak</u> zależności między cechami

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

współczynnik zbieżności Cramera:

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

Weryfikacja hipotez statystycznych

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

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

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

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

$$t_{emp} = \frac{\bar{x}_1 - \bar{x}_2}{S_{\bar{x}_1 - \bar{x}_2}} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{(N_1-1) \cdot S_1^2 + (N_2-1) \cdot S_2^2}{N_1 + N_2 - 2} \cdot \left(\frac{1}{N_1} + \frac{1}{N_2}\right)}} \qquad \nu = N_1 + N_2 - 2$$

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

$H_0 : \sigma^2 = \sigma^2_0$

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

$H_0 : \sigma^2_1 = \sigma^2_2$

$$F_{emp} = \frac{S_1^2}{S_2^2} \qquad F_{emp} \geq 1 \qquad \nu_1 = N_1 - 1 \qquad \nu_2 = N_2 - 1$$

$H_0 : p = p_0$

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

$H_0 : p_1 = p_2$

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

współczynnik korelacji Pearsona

$$r = \frac{cov_{xy}}{S_x \cdot S_y}$$

współczynnik korelacji Spearmana

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

$H_0 : \rho_{XY} = 0$

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

Ocena zgodności z rozkładem teoretycznym – test na zgodność

H_0 : rozkład cechy jest zgodny z teoretycznym

$$\chi^2_{emp} = \sum_{i}^{k} \frac{(n_i - p_i \cdot N)^2}{p_i \cdot N} \qquad \nu = k-1-m$$