SZKOŁA GŁÓWNA HANDLOWA W WARSZAWIE
INSTYTUT STATYSTYKI I DEMOGRAFII

STATYSTYKA

PODSTAWOWE WZORY DOZWOLONE NA EGZAMINIE NA STUDIACH LICENCJACKICH

Opracowanie przygotowane przez dr Marię Wieczorek na podstawie:

1. P. Kuszewski, J. Podgórski: Statystyka. Wzory i tablice. SGH, Warszawa, 2008
2. M. Wieczorek: Statystyka. Lubię to! Zbiór zadań. SGH, Warszawa 2016

1. Statystyki opisowe rozkładu empirycznego

Dystrybuanta empiryczna

$$F_n(x) = \begin{cases} 0 & \text{dla } x < x_{min} \\ \sum_{s=1}^{i} w_s & \text{dla } x_i \le x < x_{i+1} \quad i=1,2,\ldots k-1 \\ 1 & \text{dla } x \ge x_{max} \end{cases} \tag{1}$$

Średnia arytmetyczna

$$\bar{x} = \frac{1}{n}\sum_{j=1}^{n} x_j \tag{2}$$

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} x_i n_i = \sum_{i=1}^{k} x_i w_i \tag{3}$$

$$\bar{x} \cong \frac{1}{n}\sum_{i=1}^{k} \dot{x}_i n_i = \sum_{i=1}^{k} \dot{x}_i w_i \tag{4}$$

Mediana

$$me = \begin{cases} x_{\frac{n+1}{2}} & gdy \quad n \quad nieparzyste \\ \dfrac{x_{\frac{n}{2}} + x_{\frac{n+2}{2}}}{2} & gdy \quad n \quad parzyste \end{cases} \tag{5}$$

$$me = x_{0m} + \left[\frac{n}{2} - n(x_{0m})\right]\frac{h_m}{n_m} = x_{0m} + \left[\frac{1}{2} - F_n(x_{0m})\right]\frac{h_m}{w_m} \tag{6}$$

Kwantyl rzędu p

$$k_p = x_{0p} + \left[p - F_n(x_{0p})\right]\frac{h_p}{w_p} \tag{7}$$

Wariancja:

obciążona

$$s^2 = \frac{1}{n}\sum_{j=1}^{n}(x_j - \bar{x})^2 = \frac{1}{n}\left(\sum_{j=1}^{n} x_j^2 - n\bar{x}^2\right) \tag{8}$$

$$s^2 = \frac{1}{n}\sum_{i=1}^{k}(x_i - \bar{x})^2 n_i = \frac{1}{n}\left(\sum_{i=1}^{k} x_i^2 n_i - n\bar{x}^2\right) \tag{9}$$

$$s^2 = \sum_{i=1}^{k}(x_i - \bar{x})^2 w_i = \sum_{i=1}^{k} x_i^2 w_i - \bar{x}^2 \tag{10}$$

nieobciążona

$$s^2 = \frac{1}{n-1}\sum_{j=1}^{n}(x_j - \bar{x})^2 = \frac{1}{n-1}\left(\sum_{j=1}^{n} x_j^2 - n\bar{x}^2\right) \tag{11}$$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{k}(x_i - \bar{x})^2 n_i = \frac{1}{n-1}\left(\sum_{i=1}^{k} x_i^2 n_i - n\bar{x}^2\right) \tag{12}$$

Odchylenie standardowe

$$s = \sqrt{s^2} \tag{13}$$

Współczynnik zmienności

$$V = \frac{s}{\bar{x}}(\cdot 100\%) \tag{14}$$

Rozstęp

$$x_{\max} - x_{\min} \tag{15}$$

Rozstęp ćwiartkowy

$$Q_3 - Q_1 \tag{16}$$

Odchylenie ćwiartkowe

$$Q = \frac{Q_3 - Q_1}{2} \tag{17}$$

Współczynnik zmienności (pozycyjny)

$$V = \frac{Q}{me}(\cdot 100\%) \tag{18}$$

Współczynnik asymetrii

$A = \dfrac{M_3^{'}}{s^3}$ gdzie (trzeci moment centralny): $$M_3^{'} = \frac{1}{n}\sum_{j=1}^{n}(x_j - \bar{x})^3 \tag{19}$$

lub $$M_3^{'} = \frac{1}{n}\sum_{i=1}^{k}(\dot{x}_i - \bar{x})^3 n_i \tag{20}$$

3. Rozkład zmiennej losowej

Funkcja prawdopodobieństwa zmiennej losowej skokowej X

$P(X = x_i) = p_i$ $\qquad$ $i=1,2,\ldots.$ $\qquad$ przy czym $\sum_i p_i = 1$ (1)

Dystrybuanta zmiennej losowej skokowej X

$$F(x) = \sum_{x_i \le x} p_i \qquad (-\infty < x < \infty) \tag{2}$$

Wartość oczekiwana zmiennej losowej skokowej X

$$E(X) = \sum_i x_i p_i \tag{3}$$

Wariancja zmiennej losowej (definicja)

$$D^2(X) = E[X - E(X)]^2 = E(X^2) - [E(X)]^2 \tag{4}$$

Wariancja zmiennej losowej skokowej X (metoda obliczania)

$$D^2(X) = \sum_i [x_i - E(X)]^2 p_i = \sum_i x_i^{\,2} p_i - [E(X)]^2 \tag{5}$$

Standaryzacja zmiennej losowej X

$$U = \frac{X - E(X)}{D(X)} \tag{6}$$

4. Wybrane typy rozkładów

Rozkład zero-jedynkowy (dwupunktowy)

$$\begin{cases} P\,(X = 1) = p & (1) \\ P(X = 0) = 1\text{-}p & (2) \end{cases}$$

przy czym: $E(X) = p\,; \quad D(X) = \sqrt{p(1-p)}$ (3)

Rozkład dwumianowy

liczby sukcesów: $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$ k= 0, 1, 2,….n (4)

przy czym: $E(X) = np\,; \quad D(X) = \sqrt{np(1-p)}$ (5)

częstości sukcesów: $P\left(W = \frac{k}{n}\right) = \binom{n}{k} p^k (1-p)^{n-k}$ k= 0, 1, 2,….n (6)

przy czym: $E(W) = p\,; \quad D(W) = \sqrt{\frac{p(1-p)}{n}}$ (7)

Rozkład normalny N(m,σ)

$$f(x) = \frac{1}{\delta\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}} \qquad (-\infty < x < \infty) \tag{8}$$

przy czym: E(X) = m; D(X) = σ (9)

Rozkład normalny standardowy N(0,1)

$$f(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} \qquad (-\infty < u < \infty) \tag{10}$$

przy czym: E(U)=0; D(U) = 1 (11)

4. Rozkłady statystyk z próby

Założenia	Wielkość próby	Rozkład statystyki
Rozkłady dokładne średniej arytmetycznej z próby		
X: $N(m, \sigma)$ σ – znane	próba dowolna	$\overline{X} : N\left(m, \frac{\sigma}{\sqrt{n}}\right)$ (1)
(X: $N(m, \sigma)$ σ – nieznane	próba dowolna[a)]	(2) $t = \frac{\overline{X} - m}{s}\sqrt{n}$ t; $v = n - 1$ *gdzie* $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$
Rozkłady dokładne różnicy średnich z dwóch prób		
X_1: $N(m_1, \sigma_1)$ X_2: $N(m_2, \sigma_2)$ σ_1, σ_2 – znane	próby dowolne	$\overline{X}_1 - \overline{X}_2 : N\left(m_1 - m_2, \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)$ (3)
X_1: $N(m_1, \sigma)$ X_2: $N(m_2, \sigma)$ σ – nieznane, ale jednakowe	próby dowolne[a)]	(4) $t = \frac{(\overline{X}_1 - \overline{X}_2) - (m_1 - m_2)}{\sqrt{s_p^2(\frac{1}{n_1} + \frac{1}{n_2})}}$ t; $v = n_1 + n_2 - 2$ *gdzie* $s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$
Rozkłady graniczne: średniej, sumy wartości, częstości i liczby sukcesów w rozkładzie dwumianowym, różnicy średnich i częstości		
X ma dowolny rozkład (niekoniecznie normalny) ze średnią m i odchyleniem standardowym σ.	próba duża $n \geq 100$	$\overline{X} : as.N\left(m, \frac{\sigma}{\sqrt{n}}\right)$ (5)
X ma dowolny rozkład (niekoniecznie normalny) ze średnią m i odchyleniem standardowym σ.	próba duża $n \geq 100$	$Z_n = \sum_{i=1}^{n} x_i$; Z_n: $as\ N(nm, \sqrt{n}\sigma)$ (6)

X_1 i X_2 mają dowolne rozkłady (niekoniecznie normalne) z parametrami odpowiednio m_1 i σ_1 oraz m_2 i σ_2.	próby duże	$\overline{X}_1 - \overline{X}_2 : as.N\left(m_1 - m_2, \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)$ (7)
Populacja zero-jedynkowa[b]; X ma rozkład dwumianowy z parametrami n i p. $W = \frac{X}{n}$ $E(W) = p$ $D(W) = \sqrt{\frac{p(1-p)}{n}}$	próba duża $n \geq 100$	$W : as.N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$ (8)
Populacja zero-jedynkowa[b]; X ma rozkład dwumianowy z parametrami n i p (X oznacza liczbę sukcesów w n próbach)	próba duża $n \geq 100$	$X: as\ N\left(np, \sqrt{np(1-p)}\right)$ (9)
Populacje zero-jedynkowe; X_1 i X_2 mają rozkłady dwumianowe z parametrami odpowiednio n_1 i p_1 oraz n_2 i p_2. $W_1 - W_2 = \frac{X_1}{n_1} - \frac{X_2}{n_2}$	próby duże	$W_1 - W_2 : asN\left(p_1 - p_2, \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}\right)$ (10)
Twierdzenia de Moivre'a - Laplace'a oraz Lindeberga - Levy'ego		
Populacja zero-jedynkowa [b]; X ma rozkład dwumianowy z parametrami n i p (X oznacza liczbę sukcesów w n próbach)	próba duża $n \geq 100$	$U_n = \frac{X-np}{\sqrt{np(1-p)}}$; $U_n: as\ N(0,1)$ (11) (Tw. de Moivre'a-Laplace'a)
X ma dowolny rozkład (niekoniecznie normalny) ze średnią m i odchyleniem standardowym σ.	próba duża $n \geq 100$	$U_n = \frac{\sum_{i=1}^{n} x_i - nm}{\sqrt{n}\sigma}$; $U_n: asN(0,1)$ (12) (Tw. Lindeberga-Levy'ego)

[a] Niektórzy autorzy uznają, że już przy próbie n>30 dobrym przybliżeniem rozkładu *t Studenta* jest rozkład normalny, a w związku z tym rozkład *t* stosują dla prób n ≤ 30.

[b] Z populacji o rozkładzie zero-jedynkowym losujemy niezależnie *n* elementów. *X* – liczba elementów wyróżnionych w próbie.

5. Przedziały ufności

Założenia	**Wielkość próby**	**Przedział ufności**	**Współczynnik ufności**
		Przedział ufności dla średniej *m*	
Populacja normalna X: $N(m, \sigma)$ m – nieznane σ – znane	próba dowolna	(1) $$P\left(\bar{X} - u_\alpha \frac{\sigma}{\sqrt{n}} < m < \bar{X} + u_\alpha \frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$	$P(-u_\alpha < U < u_\alpha) = 1 - \alpha$ tzn. $P(\lvert U \rvert \geq u_\alpha) = \alpha$
Populacja normalna X: $N(m, \sigma)$ m, σ – nieznane	próba dowolna[a)]	(2) $$P\left(\bar{X} - t_{\alpha,n-1} \frac{s}{\sqrt{n}} < m < \bar{X} + t_{\alpha,n-1} \frac{s}{\sqrt{n}}\right) = 1-\alpha$$	$P(-t_{\alpha,n-1} < t < t_{\alpha,n-1}) = 1 - \alpha$ tzn. $P(\lvert t \rvert \geq t_{\alpha,n-1}) = \alpha$
Populacja o dowolnym rozkładzie m i σ – nieznane	próba duża $n \geq 100$	(3) $$P\left(\bar{X} - u_\alpha \frac{s}{\sqrt{n}} < m < \bar{X} + u_\alpha \frac{s}{\sqrt{n}}\right) \cong 1-\alpha$$	$P(-u_\alpha < U < u_\alpha) = 1 - \alpha$ tzn. $P(\lvert U \rvert \geq u_\alpha) = \alpha$
		Przedział ufności dla parametru *p* w zero-jedynkowym rozkładzie populacji	
Populacja o rozkładzie zero-jedynkowym p – nieznane	próba duża $n \geq 100$	(4) $$P\left(\hat{p} - u_\alpha \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \hat{p} + u_\alpha \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right) \cong 1-\alpha$$	$P(-u_\alpha < U < u_\alpha) = 1 - \alpha$ tzn. $P(\lvert U \rvert \geq u_\alpha) = \alpha$

[a)] Niektórzy autorzy uznają, że już przy próbie n>30 dobrym przybliżeniem rozkładu *t Studenta* jest rozkład normalny, a w związku z tym rozkład *t* stosują dla prób n ≤ 30.

Minimalna liczebność próby przy szacowaniu *m*

$$n = \frac{u_\alpha^2 \sigma^2}{d^2}$$ d - dopuszczalny błąd szacunku (5)

Minimalna liczebność próby przy szacowaniu *p*

$$n = \frac{u_\alpha^2 p(1-p)}{d^2}$$ d – dopuszczalny błąd szacunku (6)

6. Testowanie hipotez statystycznych

Testy istotności

Hipoteza zerowa	**Założenia**	**Wielkość próby**	**Statystyka testująca i jej rozkład**
H_0: $m = m_0$	populacja normalna X: $N(m, \sigma)$ σ – znane	próba dowolna	(1) $U = \frac{\overline{X} - m_0}{\sigma}\sqrt{n}$ $N(0, 1)$
	populacja normalna X: $N(m, \sigma)$ σ – nieznane	próba dowolna[a)]	(2) $t = \frac{\overline{X} - m_0}{s}\sqrt{n}$ t; $v = n - 1$
	populacja normalna X: $N(m, \sigma)$ σ – nieznane lub dowolny (nieznany) rozkład populacji	próba duża $n \geq 100$	(3) $U = \frac{\overline{X} - m_0}{s}\sqrt{n}$ as. $N(0, 1)$
H_0: $m_1 = m_2$	populacje normalne X_1: $N(m_1, \sigma_1)$ X_2: $N(m_2, \sigma_2)$ σ_1, σ_2 – znane	próby dowolne	(4) $U = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$ $N(0, 1)$
	populacje normalne X_1: $N(m_1, \sigma_1)$ X_2: $N(m_2, \sigma_2)$ σ_1, σ_2 – nieznane, ale jednakowe	próby dowolne[a)]	(5) $t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$ t; $v = n_1 + n_2 - 2$ *gdzie* $s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$
	populacje normalne X_1: $N(m_1, \sigma_1)$ X_2: $N(m_2, \sigma_2)$ σ_1, σ_2 – nieznane lub dowolne (nieznane) rozkłady populacji	próby duże $n_1 + n_2 \geq 100$	(6) $U = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$ as. $N(0, 1)$
H_0: $m_R = m_0$	próby zależne różnice R_i mają rozkład normalny σ_R – nieznane	próby dowolne	(7) $t = \frac{\overline{R} - m_0}{s_R}\sqrt{n}$ t; $v = n - 1$ gdzie: $\overline{R} = \frac{\sum_{i=1}^{n} R_i}{n} = \frac{\sum_{i=1}^{n}(X_{i1} - X_{i2})}{n}$ $s_R^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(R_i - \overline{R}\right)^2$

H_0: $p = p_0$	populacja o rozkładzie zero-jedynkowym	próba duża $n \geq 100$	$$U = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \quad (8)$$ as. $N(0, 1)$ gdzie: $\hat{p} = \frac{X}{n}$
H_0: $p_1 = p_2$	populacje o rozkładzie zero-jedynkowym	próby duże	$$U = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \quad (9)$$ as. $N(0, 1)$ gdzie: $\hat{p}_1 = \frac{X_1}{n_1}$ $\hat{p}_2 = \frac{X_2}{n_2}$ $\hat{p} = \frac{X_1 + X_2}{n_1 + n_2}$

[a] Niektórzy autorzy uznają, że już przy próbie n>30 dobrym przybliżeniem rozkładu *t Studenta* jest rozkład normalny, a w związku z tym rozkład *t* stosują dla prób n ≤ 30.

Test zgodności

$$\chi^2 = \sum_{i=1}^{r} \frac{(n_i - \hat{n}_i)^2}{\hat{n}_i} = \sum_{I=1}^{r} \frac{(n_i - np_i)^2}{np_i} \qquad v = r - k - 1 \qquad (10)$$

7. Analiza wariancji

H_0: $m_1 = m_2 = \ldots = m_r$
H_1: $m_i \neq m_j$ dla co najmniej jednej pary i , j

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{r}\sum_{k=1}^{n_i} y_{ki} = \frac{1}{n}\sum_{i=1}^{r}\bar{y}_i n_i$$ Średnia ogólna (1)

$$\bar{y}_i = \frac{1}{n_i}\sum_{k=1}^{n_i} y_{ki}$$ Średnia w i-tej grupie (2)

$$s_i^2 = \frac{1}{n_i - 1}\sum_{k=1}^{n_i}(y_{ki} - \bar{y}_i)^2$$ Wariancja w i-tej grupie (3)

$$\sum_{i=1}^{r}\sum_{k=1}^{n_i}(y_{ki} - \bar{y})^2 = \sum_{i=1}^{r} n_i(\bar{y}_i - \bar{y})^2 + \sum_{i=1}^{r}\sum_{k=1}^{n_i}(y_{ki} - \bar{y}_i)^2 \quad (4)$$

SKC (SST) = SKM (SSB) + SKW (SSE)

gdzie $SKW = \sum_{i=1}^{r}(n_i - 1)s_i^2$ (5)

$$F = \frac{s_m^2}{s_w^2}$$ $v_1 = r - 1$; $v_2 = n - r$ (6)

gdzie $s_m^2 = \dfrac{SKM}{r-1}$ *oraz* $s_w^2 = \dfrac{SKW}{n-r}$ (7)

Obszar krytyczny: $P(F \geq F_\alpha) = \alpha$ (8)

8. Zmienna losowa dwuwymiarowa

Funkcja prawdopodobieństwa zmiennej losowej (X,Y)

$p_{ij} = P(X = x_i, Y = y_j)$ i,j =1,2, …. przy czym $\sum_i \sum_j p_{ij} = 1$ (1)

Funkcja prawdopodobieństwa rozkładu brzegowego zmiennej losowej X

$p_{i\cdot} = \sum_j p_{ij} = P(X = x_i)$ i=1,2,… przy czym $\sum_i p_{i\cdot} = 1$ (2)

Funkcja prawdopodobieństwa rozkładu brzegowego zmiennej losowej Y

$p_{\cdot j} = \sum_i p_{ij} = P(Y = y_j)$ j=1,2,… przy czym $\sum_j p_{\cdot j} = 1$ (3)

Funkcja prawdopodobieństwa rozkładu warunkowego zmiennej losowej X

$P(X = x_i \mid Y = y_j) = \dfrac{p_{ij}}{p_{\cdot j}}$ i,j = 1,.2,…. (4)

Funkcja prawdopodobieństwa rozkładu warunkowego zmiennej losowej Y

$P(Y = y_j \mid X = x_i) = \dfrac{p_{ij}}{p_{i\cdot}}$ i,j = 1,.2,…. (5)

Warunek niezależności zmiennych losowych skokowych X i Y

$p_{ij} = p_{i\cdot} \cdot p_{\cdot j}$ dla wszystkich par i, j (6)

Kowariancja zmiennych losowych X i Y

$\mathrm{cov}(X,Y) = E\{[X - E(X)] \cdot [Y - E(Y)]\}$ = E(XY) - E(X) E(Y) (7)

Kowariancja zmiennych losowych skokowych X i Y

$$\mathrm{cov}(X,Y) = \sum_i \sum_j [x_i - E(X)] \cdot [y_j - E(Y)] p_{ij} = \sum_i \sum_j x_i y_j p_{ij} - E(X) \cdot E(Y) \quad (8)$$

Współczynnik korelacji liniowej Pearsona zmiennych losowych X i Y

$$\rho = \frac{\mathrm{cov}(X,Y)}{D(X)D(Y)} \quad (9)$$

9. Badanie zależności dwóch cech

Rozkłady brzegowe w próbie

Średnie

$$\bar{x} = \frac{\sum_{i=1}^{k} x_i n_{i\bullet}}{n} \qquad \bar{y} = \frac{\sum_{j=1}^{l} y_j n_{\bullet j}}{n} \tag{1}$$

Wariancje (nieobciążone) (2)

$$s_x^2 = \frac{\sum_{i=1}^{k} (x_i - \bar{x})^2 n_{i\bullet}}{n-1} = \frac{1}{n-1}\left(\sum_{i=1}^{k} x_i^2 n_{i\bullet} - n\bar{x}^2\right) \qquad s_y^2 = \frac{\sum_{j=1}^{l} (y_j - \bar{y})^2 n_{\bullet j}}{n-1} = \frac{1}{n-1}\left(\sum_{j=1}^{l} y_j^2 n_{\bullet j} - n\bar{y}^2\right)$$

Rozkłady warunkowe w próbie

Średnie

$$\bar{x}_j = \frac{\sum_{i=1}^{k} x_i n_{ij}}{n_{\bullet j}} \qquad \bar{y}_i = \frac{\sum_{j=1}^{l} y_j n_{ij}}{n_{i\bullet}} \tag{3}$$

Wariancje (nieobciążone) (4)

$${}_j s_x^2 = \frac{\sum_{i} (x_i - \bar{x}_j)^2 n_{ij}}{n_{\bullet j} - 1} = \frac{1}{n_{\bullet j} - 1}\left(\sum_{i=1}^{k} x_i^2 n_{ij} - n_{\bullet j}\bar{x}_j^2\right) \qquad {}_i s_y^2 = \frac{\sum_{j} (y_j - \bar{y}_i)^2 n_{ij}}{n_{i\bullet} - 1} = \frac{1}{n_{i\bullet} - 1}\left(\sum_{j=1}^{l} y_j^2 n_{ij} - n_{i\bullet}\bar{y}_i^2\right)$$

Średnie ogólne (brzegowe) wyliczone ze średnich warunkowych

$$\bar{x} = \frac{\sum_{j=1}^{l} \bar{x}_j n_{\bullet j}}{n} \qquad \bar{y} = \frac{\sum_{i=1}^{k} \bar{y}_i n_{i\bullet}}{n} \tag{5}$$

Kowariancja (nieobciążona)

dane indywidualne:

$$c_{xy} = \frac{1}{n-1}\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}\right) \tag{6}$$

dane pogrupowane w tablicy korelacyjnej:

$$c_{xy} = \frac{1}{n-1}\sum_{i=1}^{k}\sum_{j=1}^{l}(x_i - \bar{x})(y_j - \bar{y})n_{ij} = \frac{1}{n-1}\left(\sum_{i=1}^{k}\sum_{j=1}^{l} x_i y_j n_{ij} - n\bar{x}\bar{y}\right) \qquad (7)$$

Test niezależności

$$\chi^2 = \sum_{i=1}^{k}\sum_{j=1}^{l}\frac{(n_{ij} - \hat{n}_{ij})^2}{\hat{n}_{ij}} \qquad \text{v=(k-1)(l-1)} \qquad \text{gdzie } \hat{n}_{ij} = \frac{n_{i\cdot} \cdot n_{\cdot j}}{n} \qquad (8) \quad (9)$$

Współczynnik zbieżności VCramera

$$V = \sqrt{\frac{\chi^2}{\mathrm{n(m-1)}}} \qquad \text{gdzie m=min(k,l)} \qquad (10)$$

Współczynnik korelacji liniowej Pearsona

$$r_{yx} = r_{xy} = \frac{c_{xy}}{s_x s_y} \qquad (11)$$

Test na nieskorelowanie liniowe zmiennych

$$t = \frac{r}{\sqrt{1-r^2}}\sqrt{n-2} \qquad \text{v=n-2} \qquad (12)$$

Współczynnik korelacji rang Spearmana

$$r_d = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)} \qquad (13)$$

10.Klasyczny model regresji liniowej z jedną zmienną objaśniającą

$$\mathbf{Y = E(Y|X = x_i)+\varepsilon_i = \alpha x_i + \beta + \varepsilon_i} \qquad i=1,2,\ldots n \tag{1}$$

Założenia:

$E(\varepsilon_i) = 0$
$D^2(\varepsilon_i)=E(\varepsilon_i^2)=\sigma^2$
$cov(\varepsilon_i,\varepsilon_j)=0$ dla $i\neq j$
$\varepsilon_i : N(0,\sigma)$

Funkcja regresji liniowej: $\hat{y}_i = \hat{\alpha}\cdot x_i + \hat{\beta}$ (2)

Parametry strukturalne:

$$\hat{\alpha} = \frac{\sum_{i=1}^{n} x_i y_i - \frac{\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n}}{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{c_{xy}}{s^2{}_x} = r_{xy}\frac{s_y}{s_x} \tag{3}$$

$$\hat{\beta} = \bar{y} - \hat{\alpha}\bar{x} \tag{4}$$

Parametry stochastyczne

a) Wariancja i odchylenie standardowe składnika resztowego

$e_i = y_i - \hat{y}_i$ - reszty
y_i – wartości empiryczne
$\hat{y}_i = \hat{\alpha}x_i + \hat{\beta}$ - wartości teoretyczne

$$s_e^2 = \frac{\sum_{i=1}^{n} e_i^2}{n-2} \qquad S_e = \sqrt{S_e^2} \tag{5}$$

b) Błędy standardowe oszacowań parametrów:

$$s_{\hat{\alpha}} = \sqrt{\frac{s_e^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}} = \sqrt{\frac{s_e^2}{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}} = \sqrt{\frac{s_e^2}{(n-1)s_x^2}} \tag{6}$$

$$s_{\hat{\beta}} = \sqrt{\frac{s_e^2 \sum_{i=1}^{n} x_i^2}{n\sum_{i=1}^{n}(x_i - \bar{x})^2}} = \sqrt{\frac{s_e^2 \sum_{i=1}^{n} x_i^2}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}} = \sqrt{\frac{s_e^2 \sum_{i=1}^{n} x_i^2}{n(n-1)s_x^2}} \tag{7}$$

Podział całkowitej sumy kwadratów odchyleń

$$\sum_{i=1}^{n}(y_i-\bar{y})^2 = \sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2 + \sum_{i=1}^{n}(y_i-\hat{y}_i)^2 \tag{8}$$

Współczynnik determinacji liniowej:

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2} = 1-\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2} = 1-\frac{\sum_{i=1}^{n}e_i^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2} = 1-\frac{(n-2)s_e^2}{(n-1)s_y^2} \tag{9}$$

Przedział ufności dla α

$$P(\hat{\alpha}-t_{\gamma,n-2}s_{\hat{\alpha}} < \alpha < \hat{\alpha}+t_{\gamma,n-2}s_{\hat{\alpha}}) = 1-\gamma \tag{10}$$

Test dla hipotezy H_0: α = 0

$$t = \frac{\hat{\alpha}}{s_{\hat{\alpha}}}, \qquad v = n-2 \tag{11}$$

Prognoza Y dla $X=x^p$

$$\hat{y}^p = \hat{\alpha}\cdot x^p + \hat{\beta} \tag{12}$$

Błąd standardowy prognozy Y dla $X=x^p$

$$s(y_x^p) = s_e\sqrt{1+\frac{1}{n}+\frac{(x^p-\bar{x})^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}} \qquad \textbf{gdzie} \ \sum_{i=1}^{n}(x_i-\bar{x})^2 = (n-1)s_x^2 \tag{13}$$

Przedział ufności dla prognozowanej wartości Y

$$P[\hat{y}_x - t_{\gamma,n-2}\cdot s(y_x^p) < \hat{y} < \hat{y}_x + t_{\gamma,n-2}\cdot s(y_x^p)] = 1-\gamma \tag{14}$$

11. Wyrównywanie szeregów czasowych

Średnia ruchoma zwykła

$$\bar{y}_t = \frac{1}{2q+1}\sum_{r=-q}^{q} y_{t+r} \qquad t=q+1,\ q+2, \ldots,\ n-q; \qquad q=1, 2, \ldots \tag{1}$$

Średnia ruchoma scentrowana

$$\bar{y}_t = \frac{1}{2q}\left[\frac{1}{2}y_{t-q} + \sum_{r=-q+1}^{q-1} y_{t+r} + \frac{1}{2}y_{t+q}\right] \tag{2}$$

12.Indeksy statystyczne

Indeksy indywidualne

t = 1, …, n

Przyrost absolutny łańcuchowy (1) $\Delta y_t = y_t - y_{t-1}$	Przyrost absolutny jednopodstawowy (4) $\Delta y_t = y_t - y_{t*}$
Przyrost względny łańcuchowy (2) $\delta_t = \frac{y_t - y_{t-1}}{y_{t-1}}$	Przyrost względny jednopodstawowy (5) $\delta_t = \frac{y_t - y_{t*}}{y_{t*}}$
Indeks indywidualny łańcuchowy (3) $i_{t/t-1} = \frac{y_t}{y_{t-1}}$	Indeks indywidualny jednopodstawowy (6) $i_{t/t*} = \frac{y_t}{y_{t*}}$

Średnie tempo zmian w okresie (1, n)

r (1, n) = $\bar{i}_g - 1$ (7)

gdzie:

(8) (9) (10)

$$\bar{i}_g = \sqrt[n-1]{\prod_{t=2}^{n} i_{t/t-1}} = \sqrt[n-1]{\frac{y_n}{y_1}} = \sqrt[n-1]{i_{n/1}}$$

Prognoza na podstawie średniego tempa: $y_{n+k} = y_n \cdot (\bar{i}_g)^k$ (11)

${}_j i_p = \frac{p_{j1}}{p_{jo}}$ indywidualny indeks cen j-tego produktu (12)

${}_j i_q = \frac{q_{j1}}{q_{jo}}$ indywidualny indeks ilości j-tego produktu (13)

${}_j i_w = \frac{w_{j1}}{w_{jo}}$ indywidualny indeks wartości j-tego produktu (14)

${}_j i_w = {}_j i_p \cdot {}_j i_q$ (15)

Indeksy agregatowe wartości, ilości i cen

(sumowanie w indeksach agregatowych po *j* składnikach agregatu)

Agregatowy indeks wartości (zmian nominalnych): (16)

$$I_w = \frac{\sum_j w_{j1}}{\sum_j w_{j0}} = \frac{\sum_j p_{j1} \cdot q_{j1}}{\sum_j p_{j0} \cdot q_{j0}}$$

Agregatowy indeks cen:

Ogólna postać standaryzacyjna: (17)

$$I_p = \frac{\sum_j p_{j1} q_{jt=const}}{\sum_j p_{j0} q_{jt=const}}$$

a) według formuły Laspeyresa: (18)

$$_L I_p = \frac{\sum_j p_{j1} q_{j0}}{\sum_j p_{j0} q_{j0}} = \frac{\sum_j {}_j i_p \, p_{jo} q_{jo}}{\sum_j p_{jo} q_{jo}} = \frac{\sum_j {}_j i_p \, w_{jo}}{\sum_j w_{jo}} = \sum_j {}_j i_p \, u_{jo}$$

b) według formuły Paaschego: (19)

$$_P I_p = \frac{\sum_j p_{j1} q_{j1}}{\sum_j p_{j0} q_{j1}} = \frac{\sum_j p_{j1} q_{j1}}{\sum_j \frac{p_{j1} q_{j1}}{_j i_p}} = \frac{\sum_j w_{j1}}{\sum_j \frac{w_{j1}}{_j i_p}} = \frac{1}{\sum_j \frac{u_{j1}}{_j i_p}}$$

c) według formuły Fishera: (20)

$$_F I_p = \sqrt{_L I_p \, _P I_p}$$

Agregatowy indeks ilości (zmian realnych):

Ogólna postać standaryzacyjna: (21)

$$I_q = \frac{\sum_j q_{j1} p_{jt=const}}{\sum_j q_{j0} p_{jt=const}}$$

a) według formuły Laspeyresa: (22)

$$_L I_q = \frac{\sum_j p_{j0} q_{j1}}{\sum_j p_{j0} q_{j0}} = \frac{\sum_j {_j i_q} p_{jo} q_{jo}}{\sum_j p_{jo} q_{jo}} = \frac{\sum_j {_j i_q} w_{jo}}{\sum_j w_{jo}} = \sum_j {_j i_q} u_{jo}$$

b) według formuły Paaschego: (23)

$$_P I_q = \frac{\sum_j p_{j1} q_{j1}}{\sum_j p_{j1} q_{j0}} = \frac{\sum_j p_{j1} q_{j1}}{\sum_j \frac{p_{j1} q_{j1}}{_j i_q}} = \frac{\sum_j w_{j1}}{\sum_j \frac{w_{j1}}{_j i_q}} = \frac{1}{\sum_j \frac{u_{j1}}{_j i_q}}$$

c) według formuły Fishera: (24)

$$_F I_q = \sqrt{_L I_q \, _P I_q}$$

Równość indeksowa (25)

$$\boldsymbol{I_w = {_L I_p} \cdot {_P I_q} = {_P I_p} \cdot {_L I_q} = {_F I_p} \cdot {_F I_q}}$$