Humboldt-Universität zu Berlin, Institut der Informatik

Werkzeuge der empirischen Forschung

Abgabe: 27.05.2019
Blatt 6
Pohl, Oliver
577878
pohloliq

Aufgabe 14a.

Berechnung des Pearson-Korrelationskoeffizienten:

für X_n = k=5

$\bar{\mathsf{X}} = \frac{1}{5} \cdot (1 + 2 + 3 + 4 + 5) = 3$
$\bar{\mathsf{Y}} = \frac{1}{5} \cdot (1 + 2 + 3 + 4 + 5) = 3$

$$\begin{aligned}
S_{xy} &= \frac{1}{4} \cdot \sum_{i=1}^{5} (x_i - 3) \cdot (y_i - 3) \\
&= \frac{1}{4} \cdot ((1-3) \cdot (1-3) + (2-3) \cdot (2-3) \\
&+ (3-3) \cdot (3-3) + (4-3) \cdot (4-3) + (5-3) \cdot (5-3)) \\
&= \frac{1}{4} \cdot (4 + 1 + 0 + 1 + 4) \\
&= \frac{1}{4} \cdot 10 \\
&= \frac{10}{4}
\end{aligned}$$

$$\begin{aligned}
S_x^2 &= \frac{1}{4} \cdot \sum_{i=1}^{5} (x_i - 3) \cdot (x_i - 3) \\
&= \frac{1}{4} \cdot ((1-3) \cdot (1-3) + (2-3) \cdot (2-3) \\
&+ (3-3) \cdot (3-3) + (4-3) \cdot (4-3) + (5-3) \cdot (5-3)) \\
&= \frac{1}{4} \cdot (4 + 1 + 0 + 1 + 4) \\
&= \frac{1}{4} \cdot 10 \\
&= \frac{10}{4}
\end{aligned}$$

$$S_x = \frac{\sqrt{10}}{2}$$

Werkzeuge der empirischen Forschung

Abgabe: 27.05.2019
Blatt 6
Pohl, Oliver
577878
pohloliq

$$
\begin{aligned}
S_y^2 &= \frac{1}{4} \cdot \sum_{i=1}^{5} (y_i - 3) \cdot (y_i - 3) \\
&= \frac{1}{4} \cdot ((1-3) \cdot (1-3) + (2-3) \cdot (2-3) \\
&+ (3-3) \cdot (3-3) + (4-3) \cdot (4-3) + (5-3) \cdot (5-3)) \\
&= \frac{1}{4} \cdot (4+1+0+1+4) \\
&= \frac{1}{4} \cdot 10 \\
&= \frac{10}{4}
\end{aligned}
$$

$$S_y = \frac{\sqrt{10}}{2}$$

$$
\begin{aligned}
r_{xy} &= \frac{S_{xy}}{S_x \cdot S_y} \\
&= \frac{\frac{10}{4}}{\frac{\sqrt{10}}{2} \cdot \frac{\sqrt{10}}{2}} \\
&= \frac{\frac{10}{4}}{\frac{10}{4}} \\
&= 1
\end{aligned}
$$

für X_n = k=10
$\bar{X} = \frac{1}{5} \cdot (1 + 2 + 3 + 4 + 10) = 4$
$\bar{Y} = \frac{1}{5} \cdot (1 + 2 + 3 + 4 + 5) = 3$

Werkzeuge der empirischen Forschung

Abgabe: 27.05.2019
Blatt 6
Pohl, Oliver
577878
pohloliq

$$
\begin{aligned}
S_{xy} &= \frac{1}{4} \cdot \sum_{i=1}^{5} (x_i - 4) \cdot (y_i - 3) \\
&= \frac{1}{4} \cdot ((1-4) \cdot (1-3) + (2-4) \cdot (2-3) \\
&+ (3-4) \cdot (3-3) + (4-4) \cdot (4-3) + (10-4) \cdot (5-3)) \\
&= \frac{1}{4} \cdot (6 + 2 + 0 + 0 + 12) \\
&= \frac{1}{4} \cdot 20 \\
&= \frac{20}{4} \\
&= 5
\end{aligned}
$$

$$
\begin{aligned}
S_x^2 &= \frac{1}{4} \cdot \sum_{i=1}^{5} (x_i - 3) \cdot (x_i - 3) \\
&= \frac{1}{4} \cdot ((1-4) \cdot (1-4) + (2-4) \cdot (2-4) \\
&+ (3-4) \cdot (3-4) + (4-4) \cdot (4-4) + (10-4) \cdot (10-4)) \\
&= \frac{1}{4} \cdot (9 + 4 + 1 + 0 + 36) \\
&= \frac{1}{4} \cdot 50 \\
&= \frac{50}{4}
\end{aligned}
$$

$$
S_x = \frac{\sqrt{50}}{2}
$$

Werkzeuge der empirischen Forschung

Abgabe: 27.05.2019
Blatt 6
Pohl, Oliver
577878
pohloliq

$$\begin{aligned}
S_y^2 &= \frac{1}{4} \cdot \sum_{i=1}^{5} (y_i - 3) \cdot (y_i - 3) \\
&= \frac{1}{4} \cdot ((1-3) \cdot (1-3) + (2-3) \cdot (2-3) \\
&+ (3-3) \cdot (3-3) + (4-3) \cdot (4-3) + (5-3) \cdot (5-3)) \\
&= \frac{1}{4} \cdot (4 + 1 + 0 + 1 + 4) \\
&= \frac{1}{4} \cdot 10 \\
&= \frac{10}{4}
\end{aligned}$$

$$S_y = \frac{\sqrt{10}}{2}$$

$$\begin{aligned}
r_{xy} &= \frac{S_{xy}}{S_x \cdot S_y} \\
&= \frac{5}{\frac{\sqrt{50}}{2} \cdot \frac{\sqrt{10}}{2}} \\
&= 0.89
\end{aligned}$$

für X_n = k=100
$\bar{X} = \frac{1}{5} \cdot (1 + 2 + 3 + 4 + 100) = 22$
$\bar{Y} = \frac{1}{5} \cdot (1 + 2 + 3 + 4 + 5) = 3$

Werkzeuge der empirischen Forschung

Abgabe: 27.05.2019
Blatt 6
Pohl, Oliver
577878
pohloliq

$$
\begin{aligned}
S_{xy} &= \frac{1}{4} \cdot \sum_{i=1}^{5} (x_i - 22) \cdot (y_i - 3) \\
&= \frac{1}{4} \cdot ((1-22) \cdot (1-3) + (2-22) \cdot (2-3) \\
&+ (3-22) \cdot (3-3) + (4-22) \cdot (4-3) + (100-22) \cdot (5-3)) \\
&= \frac{1}{4} \cdot (42 + 20 + 0 - 18 + 156) \\
&= \frac{1}{4} \cdot 200 \\
&= \frac{200}{4} \\
&= 50
\end{aligned}
$$

$$
\begin{aligned}
S_x^2 &= \frac{1}{4} \cdot \sum_{i=1}^{5} (x_i - 22) \cdot (x_i - 22) \\
&= \frac{1}{4} \cdot ((1-22) \cdot (1-22) + (2-22) \cdot (2-22) \\
&+ (3-22) \cdot (3-22) + (4-22) \cdot (4-22) + (100-22) \cdot (100-22)) \\
&= \frac{1}{4} \cdot (441 + 400 + 361 + 324 + 6084) \\
&= \frac{1}{4} \cdot 7610
\end{aligned}
$$

$S_x = 43.62$

Werkzeuge der empirischen Forschung

Abgabe: 27.05.2019
Blatt 6
Pohl, Oliver
577878
pohloliq

$$\begin{aligned}
S_y^2 &= \frac{1}{4} \cdot \sum_{i=1}^{5} (y_i - 3) \cdot (y_i - 3) \\
&= \frac{1}{4} \cdot ((1-3) \cdot (1-3) + (2-3) \cdot (2-3) \\
&+ (3-3) \cdot (3-3) + (4-3) \cdot (4-3) + (5-3) \cdot (5-3)) \\
&= \frac{1}{4} \cdot (4 + 1 + 0 + 1 + 4) \\
&= \frac{1}{4} \cdot 10 \\
&= \frac{10}{4}
\end{aligned}$$

$$S_y = \frac{\sqrt{10}}{2}$$

$$\begin{aligned}
r_{xy} &= \frac{S_{xy}}{S_x \cdot S_y} \\
&= \frac{59}{43.62 \cdot \frac{\sqrt{10}}{2}} \\
&= \frac{50}{69} \\
&= 0.72
\end{aligned}$$

Werkzeuge der empirischen Forschung

Abgabe: 27.05.2019
Blatt 6
Pohl, Oliver
577878
pohloliq

---

für X_n = k=1000
$\bar{\text{X}} = \frac{1}{5} \cdot (1 + 2 + 3 + 4 + 1000) = 202$
$\bar{Y} = \frac{1}{5} \cdot (1 + 2 + 3 + 4 + 5) = 3$

$$\begin{aligned}
S_{xy} &= \frac{1}{4} \cdot \sum_{i=1}^{5} (x_i - 202) \cdot (y_i - 3) \\
&= \frac{1}{4} \cdot ((1 - 202) \cdot (1 - 3) + (2 - 202) \cdot (2 - 3) \\
&+ (3 - 202) \cdot (3 - 3) + (4 - 202) \cdot (4 - 3) + (1000 - 202) \cdot (5 - 3)) \\
&= \frac{1}{4} \cdot (402 + 200 + 0 - 198 + 1596) \\
&= \frac{1}{4} \cdot 200 \\
&= \frac{2000}{4} \\
&= 500
\end{aligned}$$

$$\begin{aligned}
S_x^2 &= \frac{1}{4} \cdot \sum_{i=1}^{5} (x_i - 202) \cdot (x_i - 202) \\
&= \frac{1}{4} \cdot ((1 - 202) \cdot (1 - 202) + (2 - 202) \cdot (2 - 202) \\
&+ (3 - 202) \cdot (3 - 202) + (4 - 202) \cdot (4 - 202) + (1000 - 202) \cdot (1000 - 202)) \\
&= \frac{1}{4} \cdot (40401 + 40000 + 39601 + 39204 + 636804) \\
&= \frac{1}{4} \cdot 796010
\end{aligned}$$

$S_x = 446$

Werkzeuge der empirischen Forschung

Abgabe: 27.05.2019
Blatt 6
Pohl, Oliver
577878
pohloliq

$$
\begin{aligned}
S_y^2 &= \frac{1}{4} \cdot \sum_{i=1}^{5} (y_i - 3) \cdot (y_i - 3) \\
&= \frac{1}{4} \cdot ((1-3) \cdot (1-3) + (2-3) \cdot (2-3) \\
&+ (3-3) \cdot (3-3) + (4-3) \cdot (4-3) + (5-3) \cdot (5-3)) \\
&= \frac{1}{4} \cdot (4 + 1 + 0 + 1 + 4) \\
&= \frac{1}{4} \cdot 10 \\
&= \frac{10}{4}
\end{aligned}
$$

$$
S_y = \frac{\sqrt{10}}{2}
$$

$$
\begin{aligned}
r_{xy} &= \frac{S_{xy}}{S_x \cdot S_y} \\
&= \frac{500}{446 \cdot \frac{\sqrt{10}}{2}} \\
&= \frac{500}{705} \\
&= 0.709
\end{aligned}
$$

Werkzeuge der empirischen Forschung

Abgabe: 27.05.2019
Blatt 6
Pohl, Oliver
577878
pohloliq

Berechnung des Spearman-Rangkorrelationskoeffizienten:

für X_n = k=5,10,100,1000

$$\begin{aligned}
r_s &= 1 - \frac{6 * \sum_{i=1}^{5} (R_i - S_i)^2}{5 \cdot (5^2 - 1)} \\
&= 1 - \frac{6 * \sum_{i=1}^{5} (R_i - S_i)^2}{120} \\
&= 1 - \frac{6 * ((1-1)^2 + (2-2)^2 + (3-3)^2 + (4-4)^2 + (5-5)^2)}{120} \\
&= 1 - \frac{6 * (0)}{120} \\
&= 1 - \frac{0}{120} \\
&= 1 - 0 \\
&= 1
\end{aligned}$$

Berechnung des Kendall-Rangkorrelationskoeffizienten:

für X_n = k=5

$$\begin{aligned}
\tau &= \frac{1}{\binom{5}{2}} \cdot \sum_{i<j} a_{ij} \\
&= \frac{1}{\binom{5}{2}} \cdot sgn((1-2)*(1-2)) + sgn((1-3)*(1-3)) \\
&+ sgn((1-4)*(1-4)) + sgn((1-5)*(1-5)) + sgn((2-3)*(2-3)) \\
&+ sgn((2-4)*(2-4)) + sgn((2-5)*(2-5)) + sgn((3-4)*(3-4)) \\
&+ sgn((3-5)*(3-5)) + sgn((4-5)*(4-5)) \\
&= \frac{1}{\binom{5}{2}} \cdot 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 \\
&= \frac{1}{\binom{5}{2}} \cdot 10 \\
&= \frac{10}{10} \\
&= 1
\end{aligned}$$

Werkzeuge der empirischen Forschung

Abgabe: 27.05.2019
Blatt 6
Pohl, Oliver
577878
pohloliq

---

für X_n = k=10

$$
\begin{aligned}
\tau &= \frac{1}{\binom{5}{2}} \cdot \sum_{i<j} a_{ij} \\
&= \frac{1}{\binom{5}{2}} \cdot sgn((1-2)*(1-2)) \\
&+ sgn((1-3)*(1-3)) + sgn((1-4)*(1-4)) + sgn((1-10)*(1-5)) \\
&+ sgn((2-3)*(2-3)) + sgn((2-4)*(2-4)) + sgn((2-10)*(2-5)) \\
&+ sgn((3-4)*(3-4)) + sgn((3-10)*(3-5)) + sgn((4-10)*(4-5)) \\
&= \frac{1}{\binom{5}{2}} \cdot 1+1+1+1+1+1+1+1+1+1 \\
&= \frac{1}{\binom{5}{2}} \cdot 10 \\
&= \frac{10}{10} \\
&= 1
\end{aligned}
$$

für X_n = k=100

$$
\begin{aligned}
\tau &= \frac{1}{\binom{5}{2}} \cdot \sum_{i<j} a_{ij} \\
&= \frac{1}{\binom{5}{2}} \cdot sgn((1-2)*(1-2)) \\
&+ sgn((1-3)*(1-3)) + sgn((1-4)*(1-4)) + sgn((1-100)*(1-5)) \\
&+ sgn((2-3)*(2-3)) + sgn((2-4)*(2-4)) + sgn((2-100)*(2-5)) \\
&+ sgn((3-4)*(3-4)) + sgn((3-100)*(3-5)) + sgn((4-100)*(4-5)) \\
&= \frac{1}{\binom{5}{2}} \cdot 1+1+1+1+1+1+1+1+1+1 \\
&= \frac{1}{\binom{5}{2}} \cdot 10 \\
&= \frac{10}{10} \\
&= 1
\end{aligned}
$$

Werkzeuge der empirischen Forschung

Abgabe: 27.05.2019
Blatt 6
Pohl, Oliver
577878
pohloliq

---

für $X_n =$ k=1000

$$
\begin{aligned}
\tau &= \frac{1}{\binom{5}{2}} \cdot \sum_{i<j} a_{ij} \\
&= \frac{1}{\binom{5}{2}} \cdot sgn((1-2)*(1-2)) + sgn((1-3)*(1-3)) + sgn((1-4)*(1-4)) \\
&+ sgn((1-1000)*(1-5)) + sgn((2-3)*(2-3)) \\
&+ sgn((2-4)*(2-4)) + sgn((2-1000)*(2-5)) \\
&+ sgn((3-4)*(3-4)) + sgn((3-1000)*(3-5)) + sgn((4-1000)*(4-5)) \\
&= \frac{1}{\binom{5}{2}} \cdot 1+1+1+1+1+1+1+1+1+1 \\
&= \frac{1}{\binom{5}{2}} \cdot 10 \\
&= \frac{10}{10} \\
&= 1
\end{aligned}
$$

Werkzeuge der empirischen Forschung

Abgabe: 27.05.2019
Blatt 6
Pohl, Oliver
577878
pohloliq

Aufgabe 14d.

Grenzwert des Kendall-Korrelationskoeffizient:
$\lim \tau = \pm 1$
-1 $\leq \tau \leq$1