

## Ćwiczenia 9 i 10 – Analiza korelacji i regresji liniowej

### Zadania

#### Zadanie 1

Badano zanieczyszczenie terenów wokół pewnej elektrowni. W tym celu odsłonięto siedem profili glebowych. W powierzchniowej warstwie badanych profili zawartości ołowiu i cynku (w mg/kg) przedstawiały się następująco:

	1	2	3	4	5	6	7
X (ołów)	355	190	345	316	269	210	275
Y (cynk)	82	53	93	82	67	46	80

$$\sum x_i = 1960, \quad \sum x_i^2 = 573092, \quad \sum y_i = 503, \quad \sum y_i^2 = 37911, \quad \sum x_i y_i = 146860$$

- a) Obliczyć i zinterpretować współczynnik korelacji między cechami X i Y.  
Odp.  $r = 0,92$
- b) Sprawdzić hipotezę o braku korelacji między zawartością ołowiu i cynku w powierzchniowej warstwie badanych profili. Przyjąć poziom istotności 0,05.  
Odp.  $t_0 = 5,25$ ;  $t_{0,05;5} = 2,571$ ;  $H_0$  odrzucamy
- c) Wyznaczyć równanie regresji liniowej zawartości cynku względem zawartości ołowiu w powierzchniowej warstwie badanych profili. Zinterpretować współczynnik regresji.  
Odp.  $y = 2,468 + 0,2478x$
- d) Narysować wykres prostej regresji na tle diagramu korelacyjnego
- e) Sprawdzić, czy regresja liniowa cechy Y względem X jest istotna. Przyjąć poziom istotności 0,05.
- f) Obliczyć i zinterpretować współczynnik determinacji.  
Odp.  $R^2 = 84,64\%$
- g) Zaprognozować, zgodnie z przyjętym modelem regresji, zawartość cynku w powierzchniowej warstwie, gdy zawartość ołowiu wynosi 270 (mg/kg).  
Odp. 69,374

**Zadanie 2**

Badano zawartość tlenu rozpuszczonego w wodzie destylowanej (cecha Y w  $\text{mgO}_2/\text{dm}^3$ ) w zależności od temperatury (cecha X w  $^\circ\text{C}$ ). Uzyskano dane:

	1	2	3	4	5	6	7	8	9	10
X (temperatura)	5	7	8	11	13	14	16	17	20	21
Y (tlen)	12,9	13,6	11,9	11	11,2	11,9	10	11,7	8,8	8,9

$$\sum x_i = 132, \quad \sum x_i^2 = 2010, \quad \sum y_i = 111,9, \quad \sum y_i^2 = 1274,57, \quad \sum x_i y_i = 1409,9$$

- a) Obliczyć współczynnik korelacji między zawartością tlenu w wodzie destylowanej a temperaturą.

Odp.  $r = -0,87$

- b) Sprawdzić hipotezę o braku korelacji między zawartością tlenu rozpuszczonego w wodzie destylowanej a temperaturą. Przyjąć poziom istotności 0,05.

Odp.  $t_0 = 4,99$ ;  $t_{0,05;8} = 2,306$ ;  $H_0$  odrzucamy

- c) Wyznaczyć równanie regresji liniowej zawartości tlenu rozpuszczonego w wodzie destylowanej względem temperatury. Obliczyć i zinterpretować współczynnik determinacji.

Odp.  $R^2 = 75,69\%$

- d) Sprawdzić, czy regresja liniowa cechy Y względem X jest istotna. Przyjąć poziom istotności 0,05.

- e) Określić przewidywaną zawartość tlenu, gdy temperatura wynosi  $12^\circ\text{C}$ .

Odp. 11,5

## Wzory

współczynnik korelacji Pearsona:

$$r_{xy} = \frac{\hat{s}_{xy}}{\sqrt{\hat{s}_x^2 \hat{s}_y^2}} \quad -1 \leq r_{xy} \leq 1$$

$$\hat{s}_{xy} = \frac{1}{n-1} \left( \sum_{i=1}^n x_i y_i - \frac{\sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n} \right)$$

$$\hat{s}_x^2 = \frac{1}{n-1} \left( \sum_{i=1}^n x_i^2 - \frac{\left( \sum_{i=1}^n x_i \right)^2}{n} \right) \quad \hat{s}_y^2 = \frac{1}{n-1} \left( \sum_{i=1}^n y_i^2 - \frac{\left( \sum_{i=1}^n y_i \right)^2}{n} \right)$$

test hipotezy o braku korelacji w populacji $H_0 : \rho = 0,$ $H_1 : \rho \neq 0,$  $t = \frac{r_{xy}}{\sqrt{1 - r_{xy}^2}} \sqrt{n-2}$	Jeżeli  $ t  > t_{\alpha, n-2}$  to hipotezę $H_0$ odrzucamy na rzecz $H_1$ (inaczej: $H_0$ odrzucamy jeśli $t \in (-\infty, -t_{\alpha, n-2}) \cup (t_{\alpha, n-2}, \infty)$ )
---	---

Równanie regresji liniowej:

$$y = \hat{\beta}_0 + \hat{\beta}_1 x \quad \text{gdzie:} \quad \hat{\beta}_1 = \frac{\hat{s}_{xy}}{\hat{s}_x^2} \quad \text{oraz} \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

test hipotezy o braku istotności wsp. regresji $H_0 : \beta_1 = 0,$ $H_1 : \beta_1 \neq 0,$  $t = \frac{\hat{\beta}_1}{\hat{s}_y \sqrt{1 - r_{xy}^2}} \cdot \hat{s}_x \sqrt{n-2}$	Jeżeli  $ t  > t_{\alpha, n-2}$  to hipotezę $H_0$ odrzucamy na rzecz $H_1$ (inaczej: $H_0$ odrzucamy jeśli $t \in (-\infty, -t_{\alpha, n-2}) \cup (t_{\alpha, n-2}, \infty)$ )
---	---

Współczynnik determinacji jest miarą dopasowania prostej regresji do punktów empirycznych

$$R^2 = r_{xy}^2 \cdot 100\%$$