

Ćwiczenia 13

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:

ołów (X)	355	190	345	316	269	210	275
cynk (Y)	82	53	93	82	67	46	80

$$\sum_i x_i = 1960; \sum_i x_i^2 = 573092; \sum_i y_i = 503; \sum_i y_i^2 = 37911; \sum_i x_i y_i = 146860$$

- 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) Zprognozować, zgodnie z przyjętym modelem regresji, zawartość cynku w powierzchniowej warstwie, gdy zawartość wapnia wynosi 270 (mg/kg). Odp. 69,374
2. Na terenie byłego województwa konińskiego badano zmniejszenie się emisji pyłu (w t/rok) po zamontowaniu instalacji mokrego odpylania na kominach największych zakładów. Otrzymano dane:

liczba zamontowanych instalacji odpylania (X)	1	2	3	3	4	5
zmniejszenie emisji pyłu (Y)	5,8	6,1	8,4	9,2	9,3	10,4

$$\sum_i x_i = 18; \sum_i x_i^2 = 64; \sum_i y_i = 49,2; \sum_i y_i^2 = 420,7; \sum_i x_i y_i = 160$$

- a) Obliczyć i zinterpretować współczynnik korelacji między zmniejszeniem emisji pyłów a liczbą instalacji mokrego odpylania. Odp. $r = 0,94$
- b) Wyznaczyć równanie regresji liniowej zmniejszenia emisji pyłów względem liczby instalacji mokrego odpylania. Zinterpretować współczynnik regresji. Obliczyć i zinterpretować współczynnik determinacji. Odp. $y = 4,48 + 1,24x; R^2 = 88,36\%$
- c) Zweryfikować hipotezę o istotności regresji liniowej. Przyjąć poziom istotności $\alpha = 0,05$.
- d) Narysować prostą regresji.
- e) Określić przewidywane zmniejszenie emisji pyłów, gdy liczba instalacji wyniesie 3.

Odp. 8,2

3. Dział marketingu pewnej firmy analizował zależność wielkości sprzedaży swych produktów (w tys. sztuk) a liczbą współpracujących z zakładem hurtowni. Otrzymano dane:

liczba hurtowni (X)	1	2	3	4	5	6	7	8	9	10
wielkość sprzedaży (Y)	5,8	6,1	8,4	9,2	9,3	10,4	12,9	14,6	19,1	22,8

$$\sum_i x_i = 55; \sum_i x_i^2 = 385; \sum_i y_i = 118,6; \sum_i y_i^2 = 1684,92; \sum_i x_i y_i = 795,9$$

Ćwiczenia 13

a) Obliczyć i zinterpretować współczynnik korelacji między wielkością sprzedaży produktów a liczbą współpracujących z zakładem hurtowni.
Odp. $r = 0,95$

b) Wyznaczyć równanie regresji liniowej wielkości sprzedaży produktów względem liczby współpracujących z zakładem hurtowni. Zinterpretować współczynnik regresji. Obliczyć i zinterpretować współczynnik determinacji.
Odp. $y = 2,29 + 1,74x$; $R^2 = 90,25\%$

c) Zweryfikować hipotezę o istotności regresji liniowej. Przyjąć poziom istotności $\alpha = 0,05$.
Odp.

$$F_0 = 70,48 > F_{0,05; 1,8} = 5,318; H_0 \text{ odrzucamy}$$

d) Narysować prostą regresji na tle diagramu korelacyjnego.

e) Określić przewidywaną wielkość sprzedaży, gdy liczba hurtowni wyniesie 6.

Odp. 12,73

4. Badano zawartość tlenu rozpuszczonego w wodzie destylowanej (cecha Y w mgO_2/dm^3) w zależności od temperatury (cecha X w $^\circ\text{C}$). Uzyskano dane:

temperatura (X)	5	7	8	11	13	14	16	17	20	21
zawartość tlenu (Y)	12,9	13,6	11,9	11	11,2	11,9	10	11,7	8,8	8,9

$$\sum_i x_i = 132; \sum_i x_i^2 = 2010; \sum_i y_i = 111,9; \sum_i y_i^2 = 1274,57; \sum_i x_i y_i = 1409,9$$

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. $y = 14,5 - 0,25x$; $R^2 = 75,69\%$

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°C .

Odp. 11,5

5. Badano zależność między roczną wielkością wytworzonych odpadów w Polsce w mln ton wg GUS a ilością odpadów wykorzystanych wtórnie w ciągu roku w mln ton. Uzyskano następujące dane:

Dla X (wytworzone odpady):	120,8	122,7	124,6	124,4	133,2
Dla Y (wykorzystane odpady):	65,6	66,9	69,5	80,1	91,7

$$\sum_i x_i = 625,7; \sum_i x_i^2 = 78390,69; \sum_i y_i = 373,8; \sum_i y_i^2 = 28434,12; \sum_i x_i y_i = 46971,69$$

a) Obliczyć współczynnik korelacji między wielkością wykorzystanych odpadów a ilością wytworzonych odpadów oraz obliczyć współczynnik determinacji. Zinterpretować oba wyniki.
Odp. $r = 0,92$; $R^2 = 84,64\%$

b) Oszacować prostą regresji wielkości wykorzystanych odpadów względem wytworzonych odpadów.
 $y = -193,72 + 2,15x$

c) Określić przewidywaną wielkość wykorzystanych odpadów, gdy ilość wytworzonych odpadów wynosi 130 mln ton.
Odp. 85,78

Ćwiczenia 13

współczynnik korelacji Pearsona:

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

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

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

Współczynnik determinacji jest miarą dopasowania prostej regresji do punktów empirycznych

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

Równie regresji liniowej:

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

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