

R3 – korelacja i regresja – zadania

Korelacja

Zadanie 1. W pewnej grupie gospodarstw domowych badano roczne spożycie na osobę dwóch artykułów spożywczych (w kg) mąki M i cukru C. Otrzymano wyniki:

Mąka:	15	20	25	30	35	50
Cukier:	15	25	30	35	45	50

Obliczyć i przetestować współczynnik korelacji między spożyciem obu artykułów.

Zadanie 2. W pewnym doświadczeniu farmakologicznym bada się korelację pomiędzy wydzielaniem się dwóch substancji A i B. Uzyskano następujące obserwacje (w mg/kg wagi):

A	0,6	0,2	0,3	0,7	0,5	0,8	0,7	0,4	0,6	0,8
B	1,5	1,5	1,1	1,2	1,4	1,1	1,6	1,0	1,9	1,7

Na podstawie powyższej próby oszacować i przetestować współczynnik korelacji między wydzielaniem się obu substancji.

Tablice kontyngencji

Zadanie 3. Podczas badania jakości jabłek oceniano owoce ze względu na uszkodzenia spowodowane przez owocówkę jabłkoweczkę (U – owoce uszkodzone, N – owoce nieuszkodzone) oraz porażenie parchem jabłoniowym (C – owoce czyste, P – owoce z plamami). W wyniku klasyfikacji owoców uzyskano następujące liczebności:

Parch	Owocówka	
	U	N
C	29	194
P	17	68

Czy na poziomie istotności 0,01 można uznać, że badane zmienne są niezależne?

Zadanie 4. Badano istnienie związku między wykształceniem a zarobkami. Wylosowano 950 osób i otrzymano następujące dane:

	podstawowe	średnie	wyższe	ponad wyższe
<500	21	41	93	47
500-1000	33	37	35	53
1000-1500	45	75	27	43
1500-2000	30	48	50	55
>2000	71	47	49	50

Czy powyższe świadczą o istnieniu zależności między wykształceniem i zarobkami?

Zadanie 5. Opinie konsumentów dotyczące dwóch gatunków kawy na podstawie badań sondażowych. Każdy z sześćdziesięciu konsumentów oceniał każdą z dwóch kaw w skali punktowej. Uzyskano następujące wyniki:

Ocena	Kawa Szatanex	Kawa Lurekcja
3	4	6
4	14	16
5	17	21
6	19	13
7	6	4

Czy istnieć zależność między ocenami obu kaw?

Regresja

Zadanie 6. W badaniach nad szybkością oddawania wody przez wędlinę typu serwolotka poszukiwano związku między liczbą dni przechowywania a zawartością wody w wędlinie (w %). Wyniki przedstawia tabela.

Liczba dni przechowywania	1	2	3	4	5	6	7
Zawartość wody w %	49	47,5	47	46,5	44,5	44	45

Wyznaczyć regresję liniową procentowej zawartości wody w wędlinie względem liczby dni przechowywania.

Obliczyć i przetestować współczynnik korelacji między zawartością wody w wędlinie a liczbą dni przechowywania.

$$\text{Odp. } y = 49,29 - 0,77x$$

Zadanie 7. W celu zbadania zależności między wysokością obrotów w tys. zł (Y) w sklepach spożywczych a liczbą sprzedawców X wylosowano 7 sklepów i otrzymano następujące wyniki:

liczba sprzedawców:	1	2	3	4	5	6	6
---------------------	---	---	---	---	---	---	---

obroty (w tys.zł): 1,2 2,2 3 3,5 4 5 6

Oszacować prostą regresji wielkości obrotu względem liczby sprzedawców.

Obliczyć i przetestować współczynnik korelacji między liczbą sprzedawców a obrotami.

$$\text{Odp. } y = 0.43 + 0.81x$$

Zadanie 8. Badano zależność między zawartością tłuszczu w zupach regeneracyjnych (w %) a ich wartością regeneracyjną (w kcal/100 g). W pobranych losowo zupach przygotowanych w tej samej stołówce w oparciu o 7 różnych receptur oznaczono te cechy i uzyskano wyniki:

Zawartość tłuszczu	2,0	2,5	2,9	3,6	4,0	4,5	5,0
Wartość kaloryczna	40	48	50	52	60	65	68

Oszacować równanie regresji wartości kalorycznej względem procentowej zawartości tłuszczu w zupach.

$$\text{Odp. } y = 23,16 + 9,02x$$

Zadanie 9. Wiadomo, że podczas długotrwałego moczenia warzyw przed gotowaniem, następują ubytki witamin rozpuszczalnych w wodzie. Obserwowano straty zawartości witaminy C w moczonych ziemniakach:

Czas moczenia w godzinach	0,5	1	2	5	10
Straty witaminy C w %	5	7	9	11	13

Wyznaczyć i przetestować współczynnik korelacji między badanymi cechami.

Wyznaczyć prostą regresji zależności strat witaminy C względem czasu moczenia warzyw.

$$\text{Odp.: } y = 6,25 + 0,74x$$

Zadanie 10. Badano zależność między czasem (w min/tydzień) przeznaczonym na reklamę telewizyjną margaryny pewnego gatunku a miesięczną ich sprzedażą (w tys. kg). Otrzymano następujące wyniki:

czas reklamy : 5 6 10 2 8 9
wielkość sprzedaży: 2 2,1 2,5 1,6 2,3 2,5 s²

Oszacować prostą regresji wielkości sprzedaży względem czasu reklamy.

$$\text{Odp. } y = 1,39 + 0,12x$$

Zadanie 11. Badano cztery cechy rzepaku ozimego: plon, długość łuszczyń, zawartość tłuszczu i sumę glukozyńlanów. Uzyskano następujące wyniki:

Lp	Plon	Długość łuszczyń	Zawartość tłuszczu	Suma glukozyńlanów
1	23.5	58.28	38.30	13.7
2	19.3	64.24	38.50	10.8
3	15.7	69.28	37.80	16.6
4	21.5	61.00	38.80	12.2
5	27.29	61.12	44.3	12.5
6	24.66	71.00	45.4	9.2
7	33.1	58.2	47.1	5.6
8	33.6	65.1	46.6	5.5
9	38.8	57.6	45.6	6.1
10	32.3	50.4	43.4	6.0
11	47.49	86.3	44.0	6.4
12	38.24	70.2	44.5	10.0
13	27.08	65.2	44.3	11.2
14	42.50	74.2	44.5	13.1
15	30.00	63.8	45.3	8.2
16	26.25	60.6	46.1	8.5
17	19.6	48.80	36.20	14.6
18	14.6	53.08	35.50	13.3
19	14.4	50.60	34.60	14.4
20	20	65.52	35.80	9.0
21	23.62	62.14	44.2	13.7
22	28.53	52.72	45	12.8
23	37.8	60.2	46.2	5.6
24	30.0	58.3	46.6	6.0
25	35.4	67.2	44.6	6.8
26	31.5	59.8	44.0	6.5
27	54.39	87.4	43.8	11.4
28	51.74	60.8	44.0	10.7
29	37.50	70.0	45.5	7.8
30	42.50	61.4	44.6	10.0

Znaleźć równanie regresji wielokrotnej liniowej, określającej zależność plonu od: długości łuszczyń, zawartości tłuszczu i sumy glukozyńlanów.

Odp. $\text{Plon} = -51.0 + 0.47 * \text{DługośćŁuszczyń} + 1.30 * \text{ZawartośćTłuszczu} - 0.35 * \text{SumaGlukozyńlanów}$