

Elementy statystyki matematycznej – laboratoria

Współzależność zjawisk. Eksploracja danych jakościowych (2 godziny)

Program ćwiczeń obejmuje następująca zadania:

1. Do danych

x	0	1	2	3	4	5	6
y	1000	993	1072	1405	2232	3865	6688

dopasować równanie regresji

$$y = ax^4 + bx^3 + cx^2 + dx + e.$$

2. Tablica

czas t	2	4	6	8	10
populacja N	2500	6000	15000	35000	90000

przedstawia rozmiar populacji bakterii (N) w różnych chwilach czasu (t). Zobrazować te dane na diagramie korelacyjnym, dodając linię trendu wykładniczego wraz z jego równaniem. Ten sam model wykładniczy znaleźć za pomocą funkcji `regexp()`. Jak wykorzystać do tego samego celu funkcję `reglinp()`?

3. Do danych

x	1	2	3	4	5	6	7	8	9	10
y	0.50	0.52	0.86	1.55	2.68	2.76	4.14	5.23	6.18	7.25

dopasować liniową funkcję regresji, a następnie sporządzić wykres residuów. Zauważyć, że residua układają się wokół pewnej charakterystycznej nieliniowej linii trendu. Na tej podstawie skorygować postać funkcji regresji.

4. Poniższy przykład dotyczy kinetyki reakcji chemicznych. W pewnym eksperymencie, w celu określenia energii aktywacji ΔE , zmierzono stałą szybkości reakcji k dla różnych temperatur T . Zmienne są związane zależnością $k = A \exp(-\Delta E/RT)$, gdzie: A – nieznaną stałą, R – stała gazowa ($R = 8.314 \text{ J} \cdot \text{K}^{-1} \cdot \text{mol}^{-1}$). Poniższa tabela zawiera dane eksperymentalne:

t	0	10	20	30	40	50	60
k	2.46E-5	1.08E-4	4.75E-4	1.63E-3	5.76E-3	1.85E-2	5.48E-2

Określić wartość ΔE z zastosowaniem

- (a) funkcji `nachylenie()` i `odciet()`,
 (b) wykresu XY z równaniem linii trendu;
 (c) funkcji `reglinp()`.
5. W biologii, koncepcja izometrii (stałego kształtu) głosi, że związek między pewną zmienną morfologiczną lub fizjologiczną, Y , oraz pewną podstawową zmienną wymiaru, X , ma postać $Y = aX^b$, gdzie: b – wykładnik skalujący. Pewien doświadczalnik zebrał dane, w której zmienną wymiaru jest długość ryby. Zmierzono przy tym dwie zmienne morfologiczne – długość i powierzchnia płetwy piersiowej – dla kilku ryb. Rezultaty przedstawia tabela:

Długość ryby (cm)	Długość płetwy (cm)	Powierzchnia płetwy (cm ²)
33.8	4.98	6.65
23.1	3.45	2.73
13.4	1.91	1.51
11.3	1.78	0.87
4.85	0.68	0.192
3.92	0.55	0.126

Spodziewane wartości wykładnika skalującego dla długości płetwy i powierzchni płetwy wynoszą odpowiednio 1 i 2. Sprawdzić tę hipotezę.

6. Producent drażetek wielosmakowych dokonał badania zawartości czterech torebek jednego ze swoich produktów. Skład torebek przedstawia tabela:

torebka \ smak	poziomka	malina	porzeczka	wiśnia	truskawka
Nr 1	85	87	84	92	89
Nr 2	108	85	88	71	86
Nr 3	95	99	61	84	80
Nr 4	103	80	82	71	98

Określić rozkłady brzegowe i warunkowe. Przedstawić je graficznie. Sprawdzić czy rozważane zmienne są niezależne.

7. *Paradoks Simpsona*. W roku 1973, badania nad rekrutacją na studia magisterskie na uniwersytecie kalifornijskim w Berkeley dały następujące rezultaty:

kierunek \ płeć	mężczyźni		kobiety	
	liczba kandydatów	liczba przyjętych	liczba kandydatów	liczba przyjętych
Kierunek 1	800	520	120	102
Kierunek 2	550	341	32	23
Kierunek 3	400	160	410	148
Kierunek 4	350	126	347	129
Kierunek 5	200	48	387	105

Z danych tych wynika, że ogółem, na wszystkich kierunkach razem, przyjęto 51,96% spośród wszystkich mężczyzn oraz 39,12% spośród wszystkich kobiet, co sugeruje dyskryminację kobiet. Wyznaczając rozkłady brzegowe i warunkowe pokazać, że sytuacja miała się wręcz odwrotnie!