

REGRESJA LINIOWA

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, 2, \dots, n,$$

gdzie ε_i , $i = 1, 2, \dots, n$, są niezależnymi zmiennymi losowymi o tym samym rozkładzie $N(0, \sigma^2)$.

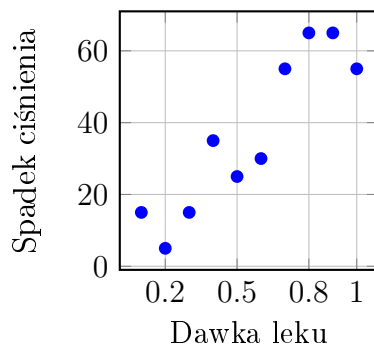
ZADANIE 1

Wskaźnik szczepień i zapadalności na błonnicę w 12 wylosowanych miastach przedstawia tabela. Stosując model regresji liniowej, zbadać istotność regresji. Czy można uznać, że im większy wskaźnik szczepień, tym mniejsza zapadalność? Jeżeli tak, to średnio o ile powinna zmniejszyć się zapadalność na błonnicę jeżeli wskaźnik szczepień zostanie zwiększony o jednostkę? Ile średnio wynosi zapadalność w miastach, w których wskaźnik szczepień jest na poziomie dwóch jednostek? Jaka może być zapadalność w mieście, w którym wskaźnik szczepień wynosi 2.3?

wskaźnik szczepień	3.32	3.42	3.45	3.85	4.25	5.20	5.84	6.25	6.50	6.88	7.60	7.60
wskaźnik zapadalności	3.20	3.00	3.50	2.25	2.50	1.80	1.00	0.79	0.52	1.00	0.30	0.06

ZADANIE 2

W pewnym doświadczeniu farmakologicznym bada się wpływ leku hipotensyjnego na ciśnienie tętnicze krwi zwierząt doświadczalnych. Podano 10 różnej wielkości dawek (w mg/kg wagi ciała) tego leku i otrzymano następujące spadki ciśnienia tętniczego krwi (w mm Hg):



dawka	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
spadek	15	5	15	35	25	30	55	65	65	55

1. Czy zależność spadku ciśnienia od dawki leku jest statystycznie istotna?
2. Jakiego spadku ciśnienia należy oczekiwać przy dawce 0.35?
3. Jaki jest przeciętny spadek ciśnienia przy dawce 0.45?
4. O ile zmienia się średnio spadek ciśnienia przy zwiększaniu dawki o 0.1?

ZADANIE 3

Badano, czy aktywność enzymu zależy od czasu leczenia (w dniach). Zebrano wyniki dla 12 osób:

Czas leczenia	1	2	3	4	5	7	10	14	18	20	24	26
Aktywność enzymu	42	40	35	44	36	35	30	33	22	20	16	18

1. Czy można uznać, że czas leczenia obniża aktywność enzymu?
2. Jakiej aktywności enzymu należy oczekiwać u osoby, która leczyla się 8 dni?

ZADANIE 4

Badano zależność między wzrostem a obwodem klatki piersiowej w populacji osób chorych na choroby reumatyczne kręgosłupa. Otrzymano wyniki:

wzrost	153	158	160	163	166	170	175	178
obwód	74	76	77	78	80	83	85	88

Przeprowadzić analizę regresji.

ZADANIE 5

Wskaźnik szczyptień i zapadalności na błonnicę w 12 wylosowanych miastach przedstawia tabela:

Wskaźnik szczyptień	3.32	3.42	3.45	3.85	4.25	5.20	5.84	6.25	6.50	6.88	7.60	7.60
Wskaźnik zapadalności	3.20	3.00	3.50	2.25	2.50	1.80	1.00	0.79	0.52	1.00	0.30	0.06

1. Stosując model regresji liniowej, zbadać istotność regresji.
2. Czy można uznać, że im większy wskaźnik szczyptień, tym mniejsza zapadalność?
3. O ile powinna zmniejszyć się zapadalność na błonnicę, gdy wskaźnik szczyptień zostanie zwiększony o jednostkę?
4. Ile średnio wynosi zapadalność w miastach, w których wskaźnik szczyptień jest na poziomie dwóch jednostek?
5. Jaka może być zapadalność w mieście, w którym wskaźnik szczyptień wynosi 2.3?

REGRESJA LOGITOWA

ZADANIE 6

Sprawdzano, czy śmiertelność z powodu choroby serca zależy od statusu socjalnego *SOC* (0-niski status, 1-wysoki status). W tym celu przeprowadzono dwunastoletni eksperyment prospektywny. Obserwowano dwustu mężczyzn w wieku co najmniej 60 lat. Jako zmienne kontrolne przyjęto *SMK* – status palenia (0-nie pali, 1-pali) oraz *SBP* – ciśnienie skurczowe serca. Dopasowano dwa modele logitowe:

Model 1

Zmienna	Współczynnik
stała	-1.180
<i>SOC</i>	-0.520
<i>SBP</i>	0.040
<i>SMK</i>	-0.560
<i>SOC</i> × <i>SBP</i>	-0.033
<i>SOC</i> × <i>SMK</i>	0.175

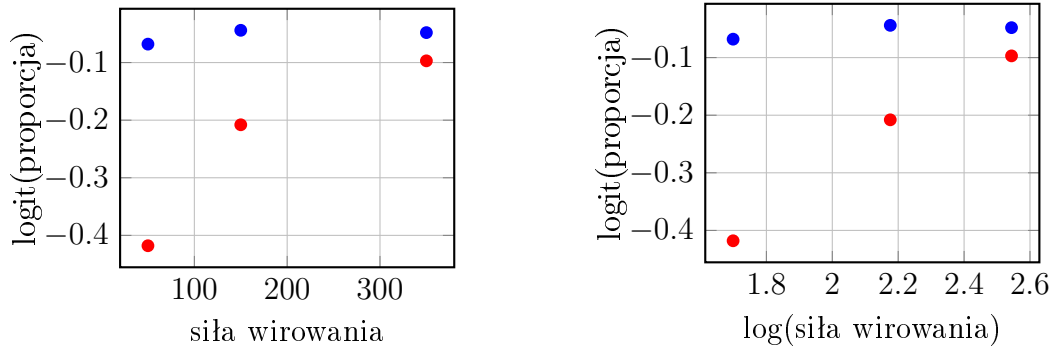
Model 2

Zmienna	Współczynnik
stała	-1.19
<i>SOC</i>	-0.50
<i>SBP</i>	0.01
<i>SMK</i>	-0.42

1. Za pomocą modelu 1. oszacuj ryzyko śmierci osoby o wysokim statusie społecznym, która pali papierosy oraz ma wysokie ciśnienie.
2. Za pomocą modelu 2. wyznacz ryzyko śmierci dla dwóch osób: Osoba 1 (*SOC* = 1, *SMK* = 1, *SBP* = 155), Osoba 2 (*SOC* = 0, *SMK* = 1, *SBP* = 155).
3. W modelu 2. wyznacz iloraz szans dla grupy o wysokim statusie społecznym względem grupy o niskim statusie społecznym. Uwzględnij warianty dla różnych wartości zmiennych kontrolnych.

ZADANIE 7

Obserwujemy liczbę zarodkowych pylników roślin z gatunku *Datura innoxia*, uzyskanych przy różnych warunkach, określonych przez dwa czynniki. Jeden z czynników jest jakościowy na dwóch poziomach. Określa sposób przechowywania pylników (kontrolny/specyficzny). Drugi czynnik jest ilościowy i określa siłę wirowania (40, 150, 350).



$$\text{proporcja} = \frac{\text{liczba pylników zarodkowych}}{\text{liczba wszystkich pylników}}$$

Warunki przechowywania		Siła wirowania		
		40	150	350
kontrolne	y_{1k}	55	52	57
	n_{1k}	102	99	108
specyficzne	y_{2k}	55	50	50
	n_{2k}	76	81	90

Sposób wprowadzenia danych w programie STATISTICA:

Przechowywanie	Wirowanie	Liczba	Zarodkowe
kontrolne	40	55	1
kontrolne	150	52	1
kontrolne	350	57	1
kontrolne	40	47	0
kontrolne	150	47	0
kontrolne	350	51	0
specyficzne	40	55	1
specyficzne	150	50	1
specyficzne	350	50	1
specyficzne	40	21	0
specyficzne	150	31	0
specyficzne	350	40	0

Zauważmy, jak są wprowadzone wyniki dla grupy pylników otrzymanych przy sile wirowania 40 oraz przechowywanych w warunkach określonych, jako kontrolne. Wszystkich pylników jest 102, zarodkowych jest 55, a pozostałych niezarodkowych jest $102-55=47$. Stąd mamy wiersz (dla Zarodkowe=1): (kontrolne, 40, 55, 1) oraz wiersz (dla Zarodkowe=0): (kontrolne, 40, 47, 0).