

7. Macierz planu X podzielona jest na dwie ortogonalne¹ części: $X = [U|Z]$, U o p kolumnach i Z o q kolumnach. Parametry regresji β , odpowiednio są podzielone na dwie o p i q wymiarowe części: $\beta^T = [\gamma^T, \delta^T]$.

a. Oblicz estymatory najmniejszych kwadratów parametrów γ i δ . Zauważ, że $\hat{\gamma}$ i $\hat{\delta}$ zależą tylko od y i (odpowiednio) od U i Z .

b. Oblicz macierz wariancji $\hat{\beta}$ i pokaż, że $\hat{\gamma}$ i $\hat{\delta}$ są nieskorelowane.

c. Oblicz macierz H .

d. Pokaż, że prognozy $\hat{y}$ składają się z dwóch nieskorelowanych części, zależnych od U i Z .

8. Macierz planu X ma *wszystkie* kolumny ortogonalne. Wykonaj punkty a – d zad.7

9. Doświadczenie czynnikowe 3² 2

Badano wpływ temperatury otoczenia na poziomach 10, 20 i 30 °C oraz stężenia czynnika na poziomach 5%, 10% i 15% na parametr y określający zapach (w jednostkach fizycznych)

y	48	50	51	55	57	58	62	63	65
temperatura	10	10	10	20	20	20	30	30	30
stezenie	5	10	15	5	10	15	5	10	15

a. Oblicz estymatory parametrów regresji modelu $y = \beta_0 + \beta_1 \text{temperatura} + \beta_2 \text{stezenie} + \epsilon$.

b. Przeprowadź standaryzację zmiennych:

$$t = \frac{\text{temperatura} - 20}{10}$$

$$s = \frac{\text{stezenie} - 10}{5}$$

Zauważ, że kolumny macierzy planu $X = [1, t, s]$ są ortogonalne. Wykorzystaj wyniki zad. 8 aby wyznaczyć bez użycia programu komputerowego parametry modelu $y = \beta_0 + \beta_1 t + \beta_2 s + \epsilon$.

c. Przelicz współczynniki modelu z punktu b na współczynniki modelu z punktu a. Dlaczego ortogonalizacja zmiennych jest użyteczna?

10. Dane są dwa modele regresji: model M_p z macierzą planu o p kolumnach i jego podmodel M_q z macierzą planu o q kolumnach ($p > q > 1$). Niech $F(M_p, M_q)$ będzie statystyką F Snedecora do oceny różnicy między modelami M_p i M_q , zaś R_p^2 i R_q^2 współczynnikami determinacji modeli M_p i M_q . Pokaż, że

$$F(M_p, M_q) = \frac{n-p}{p-q} \frac{R_p^2 - R_q^2}{1 - R_p^2}.$$

Ile stopni swobody ma F ?

Jak zmodyfikować ten wzór, aby obejmował również przypadek $q = 1$?

¹ Ortogonalność oznacza, że przestrzeń generowana przez U jest ortogonalna do przestrzeni generowanej przez Z .

² W naukach eksperymentalnych, doświadczenie czynnikowe bada wpływ dwóch lub większej liczby czynników, z których każdy jest na skończonej liczbie poziomów. Jeżeli uwzględniamy k czynników, każdy na p poziomach to doświadczenie czynnikowe obserwuje wpływ wszystkich p^k wartości zmiennych objaśniających. (https://en.wikipedia.org/wiki/Factorial_experiment)

11. Dany jest model regresji m z co najmniej dwiema zmiennymi objaśniającymi ($p \geq 2$) Z i U . Niech β_1 i β_2 będą współczynnikami regresji przy tych zmiennych. Model m można więc zapisać jako:

$$y = \beta_0 + \beta_1 Z + \beta_2 U + \dots + \epsilon,$$

gdzie $\dots$ oznacza część regresji z pozostałymi zmiennymi (o ile jeszcze takie istnieją).

a. Rozważamy alternatywny model m_0 , w którym $\beta_2 = \alpha\beta_1$ (α jest zadaną stałą). Pokaż, że w ramach testu anova można porównać model m z jego szczególnym przypadkiem m_0 . Jaka jest różnica stopni swobody modelu m i m_0 ?

Wskazówka. Gdy $\beta_2 = \alpha\beta_1$ model $y = \beta_0 + \beta_1 Z + \beta_2 U + \dots + \epsilon$ sprowadza się do modelu $y = \beta_0 + \beta_1 Z + \alpha\beta_1 U + \dots + \epsilon$ czyli $y = \beta_0 + \beta_1 (Z + \alpha U) + \dots + \epsilon$

b. Rozważamy alternatywny model m_1 , w którym $\beta_1 + \beta_2 = c$ (c jest zadaną stałą). Pokaż, że w ramach testu anova można porównać model $y - cU = \beta_0 + \beta_1 Z + \beta_2 U + \dots + \epsilon$ z jego szczególnym przypadkiem: $y - cU = \beta_0 + \beta_1 (Z - U) + \dots + \epsilon$. Jaka jest różnica stopni swobody tych modeli?

12.

Dane: galapagos.csv

Oblicz współczynniki modelu: `species ~ elevation + adjacent`.

Korzystając z wyników zad. 11 zweryfikuj hipotezy:

a. $30\beta_2 = -7\beta_1$,

b. $\beta_1 + \beta_2 = 0.23$,

c. $\beta_1 = 0.3$, $\beta_2 = -0.07$.

W przypadku c oszacuj z dołu p-value za pomocą nierówności Bonferro-niego: $P(A \cap B) \geq P(A) + P(B) - 1$ (udowodnij ją). $P(A)$ oznacza prawdopodobieństwo zdarzenia A .