## Metoda najmniejszych kwadratów

**8.1.** Przypuśćmy, że obserwacje $X_1, X_2, \ldots, X_n$ mogą być przedstawione w postaci

$$X_i = \beta_0 + \beta_1 a_i + \varepsilon_i \qquad (i = 1, 2, \ldots, n),$$

gdzie $a_1, a_2, \ldots, a_n$ są znanymi wartościami pewnej zmiennej towarzyszącej, a $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ są nieskorelowanymi błędami o wspólnej wariancji $\sigma^2$. Sprawdzić, że oba parametry $\beta_0$ i $\beta_1$ są estymowalne wtedy i tylko wtedy, gdy nie wszystkie $a_i$ są sobie równe. Przedyskutować intuicyjną treść tego wyniku sporządzając odpowiedni wykres punktów $(a_i, X_i)$. Pokazać, że gdy nie wszystkie $a_i$ są sobie równe, EMNK $\hat{\beta}_0$ i $\hat{\beta}_1$ mają postać

$$\hat{\beta}_1 = \frac{\sum(a_i - \bar{a})X_i}{\sum(a_i - \bar{a})^2} \qquad \hat{\beta}_0 = \bar{X} - \bar{a}\hat{\beta}_1.$$

Wykazać, na podstawie pierwszego z tych wyrażeń, że $D^2\hat{\beta}_1 = \sigma^2 / \sum(a_i - \bar{a})^2$. Pokazać, że kowariancja zmiennych $\bar{X}$ i $\hat{\beta}_1$ jest równa zeru i wyprowadzić następujące wzory

(a) $$Cov(\hat{\beta}_0, \hat{\beta}_1) = -\bar{a}D^2\hat{\beta}_1,$$

(b) $$D^2\hat{\beta}_0 = \sigma^2\left[\frac{1}{n} + \frac{\bar{a}^2}{\sum(a_i - \bar{a})^2}\right].$$

Sprawdzić te wyniki zapisując model w postaci macierzowej

$$\boldsymbol{X} = \boldsymbol{A}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

i korzystając z ogólnych wzorów.

**8.2.** Wyniki obserwacji $X_1, X_2, \ldots, X_n$ mogą być przedstawione w postaci

$$X_i = \beta_0 + \beta_1 a_i + \beta_2 a_i^2 + \varepsilon_i \qquad (i = 1, 2, \ldots, n),$$

gdzie $a_1, a_2, \ldots, a_n$ są znanymi wartościami pewnej zmiennej towarzyszącej, oraz $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ są wielkościami nieskorelowanymi i mającymi jednakową wariancję. Pokazać, że parametr $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)$ jest identyfikowalny wtedy i tylko wtedy, gdy wśród $a_1, a_2, \ldots, a_n$ znajdują się co najmniej trzy różne wartości.

**8.3.** Model $X_i = \beta_0 + \beta_1 a_i + \varepsilon_i$ $(i = 1, 2, \ldots, n)$ może być zapisany w postaci

$$X_i = \beta_0 + \beta_1\bar{a} + \beta_1(a_i - \bar{a}) + \varepsilon_i = \alpha + \beta_1(a_i - \bar{a}) + \varepsilon_i.$$

Pokazać, że wybór parametrów $\alpha$ i $\beta_1$ zamiast $\beta_0$ i $\beta_1$ ułatwia rachunki związane z wyznaczeniem EMNK. Sprawdzić, że w ogólnym przypadku model

$$\boldsymbol{X} = \boldsymbol{A}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

może być zawsze, po odpowiedniej zmianie parametrów, przedstawiony w postaci

$$\boldsymbol{X} = \boldsymbol{B}\boldsymbol{\gamma} + \boldsymbol{\varepsilon},$$

w której $\boldsymbol{B}$ jest macierzą o ortogonalnych kolumnach; estymator parametru $\boldsymbol{\gamma}$ otrzymuje się za pomocą łatwych rachunków.

**8.4.** Obserwacje $X_{ij}$ $(i = 1, 2, \ldots, r, j = 1, 2, \ldots, n)$ mają postać

$$X_{ij} = \mu + \tau_i + \varepsilon_{ij},$$

gdzie $\varepsilon_{ij}$ są nieskorelowanymi błędami o wspólnej wariancji. Pokazać, że $\mu, \tau_1, \tau_2, \ldots, \tau_r$ są nieidentyfikowalne, ale że stają się identyfikowalne po wprowadzeniu warunku $\tau_1 + \tau_2 + \cdots + \tau_r = 0$. Pokazać, że EMNK mają przy tym warunku postać

$$\hat{\mu} = X_{\cdot\cdot} = \frac{1}{rn}\sum_{i,j} X_{ij} \qquad \hat{\tau}_i = X_{i\cdot} - X_{\cdot\cdot}$$

gdzie

$$X_{i\cdot} = \frac{1}{n}\sum_{j} X_{ij}.$$

**8.5.** Niech $X_1, X_2, X_3, X_4$ będą wynikami lotniczych pomiarów kątów $\theta_1, \theta_2, \theta_3, \theta_4$ pewnego czworokąta na powierzchni ziemi. Założyć, że obserwacje obciążone są błędami, które są niezależne, mają wartość oczekiwaną zero i taką samą wariancję $\sigma^2$ i przy tych założeniach wyznaczyć EMNK wielkości $\theta$. Wyznaczyć nieobciążony estymator wariancji $\sigma^2$.
Przypuśćmy, że wiadomo, iż dany czworokąt jest równoległobokiem takim, że $\theta_1 = \theta_3$ oraz $\theta_2 = \theta_4$. Jaką postać mają wtedy EMNK kątów i jak można oszacować wariancję $\sigma^2$.

**8.6.** Pewien produkt chemiczny można wytwarzać bez użycia katalizatora, ale przypuszcza się, że w obecności katalizatorów wydajność procesu będzie większa. Dla zbadania tego zagadnienia przeprowadzono pięć doświadczeń według następującego schematu:

| Doświadczenie | Warunki | Wydajność |
|---|---|---|
| 1 | bez katalizatora | $Y_1$ |
| 2 | katalizator A w ilości $a_1$ | $Y_2$ |
| 3 | katalizator A w ilości $2a_1$ | $Y_3$ |
| 4 | katalizator B w ilości $a_2$ | $Y_4$ |
| 5 | katalizator B w ilości $2a_2$ | $Y_5$ |

Zakładając liniową regresję między wydajnością procesu i ilością każdego z katalizatorów skonstruować EMNK „poziomu bezkatalizatorowego” i obu współczynników regresji. Wyznaczyć, przy zwykłych założeniach o błędzie, macierz kowariancji tych estymatorów. Wywnioskować, że przy danym $a_1 + a_2$ estymator różnicy współczynników regresji ma najmniejszą wariancję, gdy $a_1 = a_2$.

**8.7.** Pewien deterministyczny proces $y_0, y_1, \ldots, y_n$ przebiega tak, że

$$y_{i+1} = ay_i \qquad (i = 0, 1, \ldots, n-1),$$

przy czym $a$ jest znaną stałą. Wielkości $y_i$ nie mogą być obserwowane bezbłędnie, lecz ich obserwacje $X_0, X_1, \ldots, X_n$ mają postać

$$X_i = y_i + \varepsilon_i \qquad (i = 0, 1, \ldots, n-1),$$

gdzie $\varepsilon_i$ są nieskorelowanymi błędami o jednakowej wariancji. Wyznaczyć EMNK $y_0, y_1, \ldots, y_n$. Jak można by szacować $y_0, y_1, \ldots, y_n$, gdyby stała $a$ nie była znana? W każdym z powyższych przypadków skonstruować estymator wariancji błędów.