

## Uogólniona metoda najmniejszych kwadratów

Rozważmy model liniowy  $(\mathbf{Y}, \mathbf{X}\boldsymbol{\beta}, \sigma^2 \boldsymbol{\Sigma})$  tzn.  $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ ,  $E(\boldsymbol{\varepsilon}) = \mathbf{0}$ ,  $\text{cov}(\boldsymbol{\varepsilon}) = \sigma^2 \boldsymbol{\Sigma}$  gdzie  $\boldsymbol{\Sigma}$  jest znaną symetryczną dodatnio określoną macierzą.

Pokażemy, że

- Najlepszym liniowym nieobciążonym estymatorem wektora  $\boldsymbol{\beta}$  jest  $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{Y}$
- Macierz kowariancyjna wektora  $\hat{\boldsymbol{\beta}}$  jest postaci  $\text{Cov}(\hat{\boldsymbol{\beta}}) = \sigma^2 (\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1}$
- Nieobciążonym estymatorem  $\sigma^2$  jest 
$$s^2 = \frac{(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})^T \boldsymbol{\Sigma}^{-1} (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})}{n - k - 1} = \frac{\mathbf{Y}^T (\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1} \mathbf{X} (\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Sigma}^{-1}) \mathbf{Y}}{n - k - 1}, \quad \text{rank } \mathbf{X} = k + 1 = p$$

Ponieważ  $\boldsymbol{\Sigma}$  jest macierzą dodatnio określoną istnieje nieosobliwa macierz  $\mathbf{P}$ , taka że  $\boldsymbol{\Sigma} = \mathbf{P}\mathbf{P}^T$ .

Mnożąc  $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$  przez  $\mathbf{P}^{-1}$  otrzymujemy model  $\mathbf{P}^{-1}\mathbf{Y} = \mathbf{P}^{-1}\mathbf{X}\boldsymbol{\beta} + \mathbf{P}^{-1}\boldsymbol{\varepsilon}$ , przy czym

$$E(\mathbf{P}^{-1}\boldsymbol{\varepsilon}) = \mathbf{P}^{-1}E(\boldsymbol{\varepsilon}) = \mathbf{0},$$

$$\text{cov}(\mathbf{P}^{-1}\boldsymbol{\varepsilon}) = \mathbf{P}^{-1} \text{cov}(\boldsymbol{\varepsilon})(\mathbf{P}^{-1})^T = \mathbf{P}^{-1} \sigma^2 \boldsymbol{\Sigma} (\mathbf{P}^{-1})^T = \sigma^2 \mathbf{P}^{-1} \mathbf{P} \mathbf{P}^T (\mathbf{P}^{-1})^T = \sigma^2 \mathbf{I}.$$

Widać, że założenia twierdzenia Gaussa-Markowa są spełnione dla modelu  $\mathbf{P}^{-1}\mathbf{Y} = \mathbf{P}^{-1}\mathbf{X}\boldsymbol{\beta} + \mathbf{P}^{-1}\boldsymbol{\varepsilon}$ .

Stąd

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= ((\mathbf{P}^{-1}\mathbf{X})^T (\mathbf{P}^{-1}\mathbf{X}))^{-1} (\mathbf{P}^{-1}\mathbf{X})^T \mathbf{P}^{-1}\mathbf{Y} = (\mathbf{X}^T (\mathbf{P}^{-1})^T \mathbf{P}^{-1}\mathbf{X})^{-1} \mathbf{X}^T (\mathbf{P}^{-1})^T \mathbf{P}^{-1}\mathbf{Y} = \\ &= (\mathbf{X}^T (\mathbf{P}\mathbf{P}^T)^{-1} \mathbf{X})^{-1} \mathbf{X}^T (\mathbf{P}\mathbf{P}^T)^{-1} \mathbf{Y} = (\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{Y}. \end{aligned}$$

$$\begin{aligned} \text{Cov}(\hat{\boldsymbol{\beta}}) &= (\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Sigma}^{-1} \text{cov}(\mathbf{Y}) (\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Sigma}^{-1})^T = \\ &= (\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Sigma}^{-1} \sigma^2 \boldsymbol{\Sigma} (\boldsymbol{\Sigma}^{-1})^T \mathbf{X} (\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} = \\ &= \sigma^2 (\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}^T (\boldsymbol{\Sigma}^{-1})^T \mathbf{X} (\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} = \sigma^2 (\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \end{aligned}$$

$$\begin{aligned} \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}} &= \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} - \mathbf{X}(\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} - \mathbf{X}(\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Sigma}^{-1} (\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) = \\ &= \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} - \mathbf{X}\boldsymbol{\beta} - \mathbf{X}(\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\varepsilon} = (\mathbf{I} - \mathbf{X}(\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Sigma}^{-1}) \boldsymbol{\varepsilon} = \mathbf{M}\boldsymbol{\varepsilon} \end{aligned}$$

gdzie  $\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Sigma}^{-1}$

Bezpośrednim rachunkiem sprawdzamy, że  $\mathbf{M}^T \boldsymbol{\Sigma}^{-1} \mathbf{M} = \boldsymbol{\Sigma}^{-1} \mathbf{M}$  i  $\text{tr} \mathbf{M} = n - k - 1 = n - p$

Uwaga (rank  $\mathbf{X} = k + 1 = p$ )

Rzeczywiście

$$\text{tr } \mathbf{M} = \text{tr } \mathbf{I} - \text{tr}(\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X} = n - \text{tr}(\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X} = n - \text{tr } \mathbf{I}_{k+1} = n - (k+1)$$

$$\begin{aligned} E(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})^T \boldsymbol{\Sigma}^{-1} (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}) &= E(\boldsymbol{\varepsilon}^T \mathbf{M}^T \boldsymbol{\Sigma}^{-1} \mathbf{M} \boldsymbol{\varepsilon}) = E(\text{tr}(\boldsymbol{\varepsilon}^T \mathbf{M}^T \boldsymbol{\Sigma}^{-1} \mathbf{M} \boldsymbol{\varepsilon})) = \\ &= E(\text{tr}(\mathbf{M}^T \boldsymbol{\Sigma}^{-1} \mathbf{M} \boldsymbol{\varepsilon} \boldsymbol{\varepsilon}^T)) = \text{tr}(\mathbf{M}^T \boldsymbol{\Sigma}^{-1} \mathbf{M} E(\boldsymbol{\varepsilon} \boldsymbol{\varepsilon}^T)) = \text{tr}(\mathbf{M}^T \boldsymbol{\Sigma}^{-1} \mathbf{M} \sigma^2 \boldsymbol{\Sigma}) = \{\mathbf{M}^T \boldsymbol{\Sigma}^{-1} \mathbf{M} = \boldsymbol{\Sigma}^{-1} \mathbf{M}\}, \\ &= \sigma^2 \text{tr}(\boldsymbol{\Sigma}^{-1} \mathbf{M} \boldsymbol{\Sigma}) = \sigma^2 \text{tr}(\mathbf{M} \boldsymbol{\Sigma} \boldsymbol{\Sigma}^{-1}) = \sigma^2 \text{tr}(\mathbf{M}) \end{aligned}$$

Stąd  $s^2 = \frac{(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})^T \boldsymbol{\Sigma}^{-1} (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})}{n - k - 1}$  jest nieobciążonym estymatorem  $\sigma^2$ .

Rozważmy teraz klasyczny estymator MNK  $\hat{\boldsymbol{\beta}}^* = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}$  parametru  $\boldsymbol{\beta}$  zastosowany do modelu  $(\mathbf{Y}, \mathbf{X}\boldsymbol{\beta}, \sigma^2 \boldsymbol{\Sigma})$ . Oczywiście

$E(\hat{\boldsymbol{\beta}}^*) = E((\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}) = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T E(\mathbf{Y}) = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{X} \boldsymbol{\beta} = \boldsymbol{\beta}$ , czyli klasyczny estymator jest również nieobciążony a jego macierz kowariancyjna jest postaci

$$\text{Cov}(\hat{\boldsymbol{\beta}}^*) = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \sigma^2 \boldsymbol{\Sigma} \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} = \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Sigma} \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1}. \text{ Oczywiście z tw. GM}$$

wynika że macierz  $\text{Cov}(\hat{\boldsymbol{\beta}}^*) - \text{Cov}(\hat{\boldsymbol{\beta}})$  jest nieujemnie określona ( $\text{Cov}(\hat{\boldsymbol{\beta}}) = \sigma^2 (\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1}$ ).

Sprawdźmy to na przykładzie  $\mathbf{X} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 5 \\ 1 & 6 \\ 1 & 7 \end{bmatrix}, \boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 6 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 7 \end{bmatrix}$

$$\text{Cov}(\hat{\boldsymbol{\beta}}) = \sigma^2 \begin{bmatrix} 1.1864 & -0.2966 \\ -0.2966 & 0.1099 \end{bmatrix}, \text{Cov}(\hat{\boldsymbol{\beta}}^*) = \sigma^2 \begin{bmatrix} 1.7143 & -0.4286 \\ -0.4286 & 0.1428 \end{bmatrix}$$

```
X<-matrix(c(rep(1),7),1:7,nrow=7)
Sig<-diag(1:7)
solve(t(X)%*%solve(Sig)%*%X) # macierz kowariancyjna z UMNK
solve(t(X)%*%X)%*%t(X)%*%Sig%*%X%*%solve(t(X)%*%X) # macierz kowariancyjna z MNK
```

## Wybrane struktury kowariancyjne

Symetryczna i dodatnio określona macierz kowariancyjna

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \cdots & \sigma_{1m} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} & \cdots & \sigma_{2m} \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & \vdots \\ \sigma_{m1} & \sigma_{m2} & \cdots & \sigma_{m(m-1)} & \sigma_m^2 \end{bmatrix}$$

jest parametryzowana przez wektor parametrów  $\theta = (\theta_1, \dots, \theta_k)^T$ . Liczba i znaczenie parametrów zależy od rozważanej struktury kowariancyjnej. W wielu strukturach kowariancyjnych używamy dekompozycji

$$\Sigma = \mathbf{D}\mathbf{P}\mathbf{D}, \text{ gdzie}$$

$$\mathbf{D} = \begin{bmatrix} \sigma_1 & 0 & \cdots & \cdots & 0 \\ 0 & \sigma_2 & 0 & \cdots & 0 \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \sigma_m \end{bmatrix} \text{ jest diagonalną macierzą odchyleń standardowych}$$

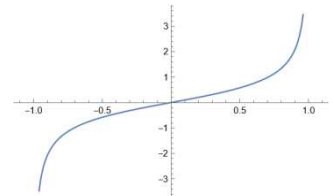
$$\mathbf{P} = \begin{bmatrix} 1 & \rho_{12} & \cdots & \cdots & \rho_{1m} \\ \rho_{21} & 1 & \rho_{23} & \cdots & \rho_{2m} \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & \vdots \\ \rho_{m1} & \rho_{m2} & \cdots & \rho_{m(m-1)} & 1 \end{bmatrix} \text{ jest macierzą korelacyjną z elementami.}$$

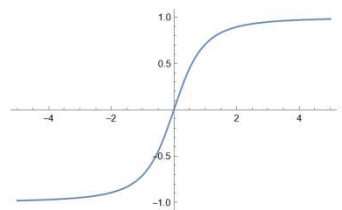
z elementami  $\rho_{ij} \in (-1,1)$ . Jeśli struktura kowariancyjna dopuszcza różne wariancje  $\sigma_i^2$  dla różnych  $i=1, \dots, m$  to nazywamy ją strukturą *heterogeniczną*. Zakładając  $\sigma^2 = \sigma_i^2, i=1, \dots, m$  otrzymamy strukturę *homogeniczną*.

### Transformacja parametrów kowariancyjnych.

Do odchyleń standardowych stosujemy transformację logarytmiczną  $\log(\sigma_i), i=1, \dots, m$  a do parametrów korelacyjnych transformację

$$\theta = f(\rho) = \text{sign}(\rho) \sqrt{\frac{\rho^2}{1-\rho^2}}$$





$$\rho = f^{-1}(\theta) = \frac{\theta}{\sqrt{1+\theta^2}}$$

przekształcającą bijektywnie przedział  $(-1,1)$  na  $R$ . Te transformacje są bardzo ważne gdyż do estymacji parametrów będzie można użyć procedury poszukiwania ekstremów bez ograniczeń.

**Unstructured (us)** Każda macierz kowariancyjna może być reprezentowana przez wysycony (saturated) zestaw  $k=m(m+1)/2$  parametrów

### Compound symmetry (symetria złożona)(cs) i (csh)

Struktury kowariancji symetrii złożonej zakładają stałą korelację między różnymi punktami czasowymi:  $\rho_{ij} = \rho$ ,  $\rho$  - jest tutaj pojedynczym parametrem korelacji.

Założenie stałej wariancji w równaniu przestrzeni stanu daje jednorodną strukturę kowariancji symetrii złożonej (cs) z jedynie dwoma parametrami ( $\theta = (\sigma, \rho)$ ), w przeciwnym razie otrzymujemy niejednorodną strukturę kowariancji(csh) symetrii złożonej z  $k=m+1$  parametrami  $\theta = (\sigma_1, \dots, \sigma_m, \rho)$ .

$$\mathbf{P} = \begin{bmatrix} 1 & \rho & \cdots & \cdots & \rho \\ \rho & 1 & \rho & \cdots & \rho \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & \vdots \\ \rho & \rho & \cdots & \rho & 1 \end{bmatrix}, \quad \mathbf{D} = \begin{bmatrix} \sigma & 0 & \cdots & \cdots & 0 \\ 0 & \sigma & 0 & \cdots & 0 \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \sigma \end{bmatrix} \text{ (cs)}, \quad \mathbf{D} = \begin{bmatrix} \sigma_1 & 0 & \cdots & \cdots & 0 \\ 0 & \sigma_2 & 0 & \cdots & 0 \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \sigma_m \end{bmatrix} \text{ (csh)}$$

### Model efektów losowych- model II komponentów wariancyjnych

Dotychczas rozważane modele ANOVA to tzw. **modele efektów stałych** albo **modele I**. W modelu efektów stałych poziomy czynników są z góry ustalone i interesują nas hipotezy liniowe dotyczące tylko tych rozważanych poziomów. Jeżeli z pewnych względów traktujemy poziomy czynnika nie jako z góry ustalone ale jako wylosowane z pewnej hipotetycznej populacji poziomów, to taki model nazywamy modelem efektów losowych (model II)

Model jednoczynnikowej ANOVA - wariant efektów losowych - z  $k$  poziomami rozważanego czynnika może być tak jak w przypadku modelu efektów stałych zapisany w postaci

$$X_{ij} = m + \alpha_i + \varepsilon_{ij}, \quad i=1, \dots, k; \quad j=1, \dots, n_i$$

gdzie

$m$  oznacza średnią ogólną (w połączonych grupach)

$\alpha_i$  jest obserwacją zmiennej losowej odzwierciedlającą wpływ  $i$  tego poziomu rozważanego czynnika wylosowanego z pewnej populacji poziomów.  
Zakładamy, że

- $\alpha_i$  ma rozkład  $N(0, \sigma_\alpha^2)$
- $\alpha_i \quad i=1, \dots, k$  są niezależne
- $\varepsilon_{ij}$  składnik losowy (niezależne zmienne losowe o rozkładzie  $N(0, \sigma^2)$ )
- zmienne  $\alpha_i$  oraz  $\varepsilon_{ij}$  są niezależne.

Testowana jest hipoteza (**nie jest to hipoteza liniowa !**)

$$H_0: \sigma_\alpha^2 = 0 \quad \text{przeciwko alternatywie} \quad H_1: \sigma_\alpha^2 \neq 0.$$

Dla przypomnienia **model efektów stałych** w tym przypadku jest postaci

$$X_{ij} = m + \alpha_i + \varepsilon_{ij}, \quad i=1, \dots, k; \quad j=1, \dots, n_i$$

$$\varepsilon_{ij} \quad \text{iid } N(0, \sigma^2)$$

$$\sum_i^k \alpha_i = 0$$

## Różnice

- W modelu efektów losowych nie ma sigma ograniczeń, gdyż poziomy są losowe i nie możemy ich kontrolować
- W modelu efektów stałych obserwacje  $X_{ij}$  są niezależne i  $X_{ij} \sim N(m + \alpha_i, \sigma^2)$  a w modelu efektów losowych  $E(X_{ij}) = m$  - wszystkie obserwacje mają wspólną wartość oczekiwaną

$$V(X_{ij}) = \sigma_a^2 + \sigma^2$$

W modelu efektów losowych obserwacje z tej samej klasy są skorelowane

$$\rho(X_{ij}, X_{ik}) = \frac{\text{Cov}(X_{ij}, X_{ik})}{\sqrt{V(X_{ij}) V(X_{ik})}} = \frac{E(X_{ij} - m)(X_{ik} - m)}{\sigma_a^2 + \sigma^2} = \frac{E(\alpha_i + \varepsilon_{ij})(\alpha_i + \varepsilon_{ik})}{\sigma_a^2 + \sigma^2} = \frac{\sigma_a^2}{\sigma_a^2 + \sigma^2}$$

Współczynnik korelacji pomiędzy zmiennymi z tej samej klasy (grupy) nosi nazwę współczynnika korelacji wewnątrzklasowej.

W modelu II efektów losowych wektor obserwacji ma więc następujący rozkład

$$\mathbf{X} \sim N \left( \mathbf{1}m, \begin{bmatrix} \mathbf{V}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{V}_k \end{bmatrix} \right), \text{ gdzie } \mathbf{V}_i = \begin{bmatrix} \sigma_a^2 + \sigma^2 & \cdots & \sigma_a^2 \\ \cdots & \cdots & \cdots \\ \sigma_a^2 & \cdots & \sigma_a^2 + \sigma^2 \end{bmatrix}$$

## Autoregresja (ar1) i (ar1h)

Załóżmy, że reszty w modelu liniowym generowane są przez proces autoregresji  $\varepsilon_t = \rho\varepsilon_{t-1} + u_t$ ,  $t=1, \dots, n$ , gdzie  $u_t$  ( $t=\dots, -2, -1, 0, 1, 2, \dots$ ) są niezależnymi zmiennymi losowymi o jednakowych rozkładach, dla których  $E(u_t) = 0$  i  $V(u_t) = E(u_t^2) = \sigma_u^2$ . Wówczas składniki losowe  $\varepsilon_t$  tworzą próbę pochodzącą ze stacjonarnego procesu stochastycznego o parametrach

$$E(\varepsilon_t) = 0, \quad V(\varepsilon_t) = E(\varepsilon_t^2) = \sigma^2, \quad \text{cov}(\varepsilon_t, \varepsilon_s) = \sigma^2 \rho^{|t-s|} \text{ gdzie } \sigma^2 = \frac{\sigma_u^2}{1-\rho^2}.$$

Czyli

$$\mathbf{P} = \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{n-1} \\ \rho & 1 & \rho & \dots & \rho^{n-2} \\ \rho^2 & \rho & 1 & \dots & \rho^{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \dots & 1 \end{bmatrix}, \quad \mathbf{D} = \begin{bmatrix} \sigma & 0 & \dots & \dots & 0 \\ 0 & \sigma & 0 & \dots & 0 \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & \dots & 0 & \sigma \end{bmatrix} \quad (\text{ar1}), \text{ gdzie } \sigma^2 = \frac{\sigma_u^2}{1-\rho^2}$$

$$\mathbf{D} = \begin{bmatrix} \sigma_1 & 0 & \dots & \dots & 0 \\ 0 & \sigma_2 & 0 & \dots & 0 \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & \dots & 0 & \sigma_m \end{bmatrix} \quad (\text{ar1h}), \dots$$

Opis pozostałych struktur kowariancyjnych dostępnych w R można znaleźć

<https://cran.r-project.org/web/packages/mmrn/vignettes/covariance.html>

i w przeglądowym artykule

<https://intapi.sciendo.com/pdf/10.2478/bile-2022-0010>

W pierwszej z powyższych pozycji można znaleźć informację o algorytmie optymalizacyjnym wykorzystującym procedurę profilowania

**Funkcja `gls()` z biblioteki `{nlme}` w R realizująca UMNK**

This function fits a linear model using generalized least squares. The errors are allowed to be correlated and/or have unequal variances.

**Usage**

```
gls(model, data, correlation, weights, subset, method, na.action,
     control, verbose)
## S3 method for class 'gls'
update(object, model., ..., evaluate = TRUE)
```

**Arguments**

<code>object</code>	an object inheriting from class "gls", representing a generalized least squares fitted linear model.
<code>model</code>	a two-sided linear formula object describing the model, with the response on the left of a <code>~</code> operator and the terms, separated by <code>+</code> operators, on the right.
<code>model.</code>	Changes to the model – see <a href="#">update.formula</a> for details.
<code>data</code>	an optional data frame containing the variables named in <code>model</code> , <code>correlation</code> , <code>weights</code> , and <code>subset</code> . By default the variables are taken from the environment from which <code>gls</code> is called.
<code>correlation</code>	an optional <a href="#">corStruct</a> object describing the within-group correlation structure. See the documentation of <a href="#">corClasses</a> for a description of the available <code>corStruct</code> classes. If a grouping variable is to be used, it must be specified in the form argument to the <code>corStruct</code> constructor. Defaults to <code>NULL</code> , corresponding to uncorrelated errors.
<code>weights</code>	an optional <a href="#">varFunc</a> object or one-sided formula describing the within-group heteroscedasticity structure. If given as a formula, it is used as the argument to <a href="#">varFixed</a> , corresponding to fixed variance weights. See the documentation on <a href="#">varClasses</a> for a description of the available <a href="#">varFunc</a> classes. Defaults to <code>NULL</code> , corresponding to homoscedastic errors.
<code>subset</code>	an optional expression indicating which subset of the rows of data should be used in the fit. This can be a logical vector, or a numeric vector indicating which observation numbers are to be included, or a character vector of the row names to be included. All observations are included by default.
<code>method</code>	a character string. If "REML" the model is fit by maximizing the restricted log-likelihood. If "ML" the log-likelihood is maximized. Defaults to "REML".
<code>na.action</code>	a function that indicates what should happen when the data contain NAs. The default action ( <a href="#">na.fail</a> ) causes <code>gls</code> to print an error message and terminate if there are any incomplete observations.
<code>control</code>	a list of control values for the estimation algorithm to replace the default values returned by the function <a href="#">glsControl</a> . Defaults to an empty list.
<code>verbose</code>	an optional logical value. If <code>TRUE</code> information on the evolution of the iterative algorithm is printed. Default is <code>FALSE</code> .
<code>...</code>	some methods for this generic require additional arguments. None are used in this method.
<code>evaluate</code>	If <code>TRUE</code> evaluate the new call else return the call.

**Możliwe struktury korelacyjne dostępne w użyciu w funkcji `gls()`**

`corClasses {nlme}`

R Documentation

Correlation Structure Classes

**Description**

Standard classes of correlation structures (`corStruct`) available in the `nlme` package.

**Value**

Available standard classes:

<code>corAR1</code>	autoregressive process of order 1.
<code>corARMA</code>	autoregressive moving average process, with arbitrary orders for the autoregressive and moving average components.
<code>corCAR1</code>	continuous autoregressive process (AR(1) process for a continuous time covariate).
<code>corCompSymm</code>	compound symmetry structure corresponding to a constant correlation.

corExp	exponential spatial correlation.
corGaus	Gaussian spatial correlation.
corLin	linear spatial correlation.
corRatio	Rational quadratics spatial correlation.
corSpher	spherical spatial correlation.
corSymm	general correlation matrix, with no additional structure.

Propozycja tematu na zaliczenie.

Przykłady użycia funkcji gls() do estymacji modeli liniowych z różnymi strukturami kowariancyjnymi

## Uwagi do metody REML estymacji komponentów wariacyjnych.

Rozważmy model  $(\mathbf{Y}, \mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma})$  tzn.  $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ . Idea metody REML (Patterson, Thompson 1971) polega na estymacji komponentów wariacyjnych (czyli  $\boldsymbol{\Sigma}$ ) metodą największej wiarygodności zastosowanej do transformowanych danych  $\mathbf{KY}$  a nie oryginalnych danych  $\mathbf{Y}$ , gdzie macierz  $\mathbf{K}$  jest tak dobrana aby rozkład  $\mathbf{KY}$  zależał tylko od komponentów wariacyjnych i nie zależał od  $\boldsymbol{\beta}$ . Aby to zrobić, poszukujemy macierzy  $\mathbf{K}$  takiej że  $\mathbf{KX} = \mathbf{0}$ . Zakładamy że macierz  $\mathbf{K}$  jest pełnego rzędu oraz, że  $\mathbf{KY}$  zawiera tak dużo informacji jak to tylko możliwe o komponentach wariacyjnych, więc  $\mathbf{K}$  musi zawierać maksymalną możliwą liczbę wierszy. Można pokazać

**Twierdzenie.** Macierz  $\mathbf{K}$  pełnego rzędu, taka, że  $\mathbf{KX} = \mathbf{0}$  jest rozmiaru  $(n-r) \times n$ , gdzie  $r = \text{rank}(\mathbf{X})$ . Ponadto  $\mathbf{K}$  jest postaci  $\mathbf{K} = \mathbf{C}(\mathbf{I} - \mathbf{H}) = \mathbf{C}(\mathbf{I} - \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T)$  gdzie  $\mathbf{C}$  jest transformacją pełnego rzędu wierszy macierzy  $\mathbf{I} - \mathbf{H}$  (rzutu ortogonalnego na podprzestrzeń ortogonalną do kolumn macierzy  $\mathbf{X}$ ).

**Dowód.** Wiersze  $\mathbf{k}_i^T$  macierzy muszą spełniać równania  $\mathbf{k}_i^T\mathbf{X} = \mathbf{0}^T$  lub równoważnie  $\mathbf{X}^T\mathbf{k}_i = \mathbf{0}$ . Z poniższego twierdzenia

Jeżeli układ równań liniowych  $\mathbf{Ax} = \mathbf{b}$  jest niesprzeczny, to wszystkie jego rozwiązania dane są wzorem  $\mathbf{x} = \mathbf{A}^{-}\mathbf{b} + (\mathbf{I} - \mathbf{A}^{-}\mathbf{A})\mathbf{h}$  gdzie  $\mathbf{A}^{-}$  jest dowolną uogólnioną odwrotnością a  $\mathbf{h}$  dowolnym wektorem.

zastosowanego dla  $\mathbf{b} = \mathbf{0}$  mamy  $\mathbf{k}_i = (\mathbf{I} - (\mathbf{X}^T)^{-}\mathbf{X}^T)\mathbf{c}$  dla dowolnego wektora  $\mathbf{c}$  wymiaru  $n \times 1$ .

Wobec tego  $\mathbf{k}_i^T = \mathbf{c}^T(\mathbf{I} - (\mathbf{X}^T)^{-}\mathbf{X}^T)^T = \mathbf{c}^T(\mathbf{I} - \mathbf{XX}^{-})$  (bo  $(\mathbf{A}^{-})^T = (\mathbf{A}^T)^{-}$ ).

Ale  $\text{rank}(\mathbf{XX}^{-}) = \text{rank}(\mathbf{X}) = r$ . Ponadto  $\mathbf{I} - \mathbf{XX}^{-}$  jest idempotentna, więc

$\text{rank}(\mathbf{I} - \mathbf{XX}^{-}) = \text{tr}(\mathbf{I} - \mathbf{XX}^{-}) = \text{tr}(\mathbf{I}) - \text{tr}(\mathbf{XX}^{-}) = n - r$ . Stąd wynika, że istnieje  $n - r$  liniowo

niezależnych wektorów  $\mathbf{k}_i$ , spełniających  $\mathbf{X}^T\mathbf{k}_i = \mathbf{0}$  i dlatego maksymalna liczba wierszy macierzy

$\mathbf{K}$  jest równa  $n - r$ . Wobec powyższego  $\mathbf{K} = \mathbf{C}(\mathbf{I} - \mathbf{XX}^{-})$  dla pewnej macierzy  $\mathbf{C}$  wymiaru

$(n-r) \times n$  pełnego rzędu, która specyfikuje liniowo niezależne kombinacje wierszy macierzy

$\mathbf{I} - \mathbf{X}\mathbf{X}^-$ . Przyjmując  $\mathbf{X}^- = (\mathbf{X}^T \mathbf{X})^- \mathbf{X}^T$  (zob. uzupełnienie macierzy) możemy przyjąć

$$\mathbf{K} = \mathbf{C}(\mathbf{I} - \mathbf{X}(\mathbf{X}^T \mathbf{X})^- \mathbf{X}).$$

Dla modelu  $\mathbf{Y} \sim N(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma})$  mamy  $\mathbf{KY} \sim N_{n-r}(\mathbf{0}, \mathbf{K}\boldsymbol{\Sigma}\mathbf{K}^T)$ .

### Rozważmy dwa specjalne przypadki.

Jeżeli próba pochodzi z ekonomicznych szeregów czasowych, to naturalnym założeniem będzie przyjęcie, że składniki losowe  $\varepsilon_1, \dots, \varepsilon_n$  pochodzą ze stacjonarnego procesu stochastycznego. Chociaż wszystkie  $\varepsilon_t$  mają jednakowe rozkłady, to jednak nie są wzajemnie niezależne, bo są skorelowane.

Założmy że w klasycznym modelu regresji zastąpimy założenie

$$E(\boldsymbol{\varepsilon}) = \mathbf{0}, \text{cov}(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{I} \text{ założeniem}$$

$$\varepsilon_t = \rho \varepsilon_{t-1} + u_t, t=1, \dots, n,$$

gdzie  $u_t$  ( $t = \dots, -2, -1, 0, 1, 2, \dots$ ) są niezależnymi zmiennymi losowymi o jednakowych rozkładach, dla których  $E(u_t) = 0$  i  $V(u_t) = E(u_t^2) = \sigma_u^2$ . Wówczas składniki losowe  $\varepsilon_t$  tworzą próbę pochodzącą ze stacjonarnego procesu stochastycznego o parametrach

$$E(\varepsilon_t) = 0, \quad V(\varepsilon_t) = E(\varepsilon_t^2) = \sigma^2, \quad \text{cov}(\varepsilon_t, \varepsilon_s) = \sigma^2 \rho^{|t-s|} \text{ gdzie } \sigma^2 = \frac{\sigma_u^2}{1-\rho^2}.$$

**Uzasadnienie:** zapisując równania  $\varepsilon_t = \rho \varepsilon_{t-1} + u_t, t=1, \dots, n$ , w postaci operatorowej z operatorem  $B$  (przesunięcia w tył tzn.  $B\varepsilon_t = \varepsilon_{t-1}$ ) otrzymujemy

$$(1 - \rho B)\varepsilon_t = u_t, \text{ stąd } \varepsilon_t = (1 - \rho B)^{-1} u_t = \sum_{k=0}^{\infty} \rho^k B^k u_t = \sum_{k=0}^{\infty} \rho^k u_{t-k}$$

$$\text{Stąd otrzymujemy } E(\varepsilon_t) = E\left(\sum_{k=0}^{\infty} \rho^k u_{t-k}\right) = \sum_{k=0}^{\infty} \rho^k E(u_{t-k}) = 0$$

$$\text{cov}(\varepsilon_t, \varepsilon_s) = E\left(\left(\sum_{i=0}^{\infty} \rho^i u_{t-i}\right)\left(\sum_{j=0}^{\infty} \rho^j u_{s-j}\right)\right) = \sum_{i,j=0}^{\infty} \rho^{i+j} E(u_{t-i} u_{s-j})$$

$$\text{Stąd } V(\varepsilon_t) = E(\varepsilon_t^2) = E\left(\left(\sum_{i=0}^{\infty} \rho^i u_{t-i}\right)\left(\sum_{j=0}^{\infty} \rho^j u_{t-j}\right)\right) = \sum_{i,j=0}^{\infty} \rho^{i+j} E(u_{t-i} u_{t-j}) = \sum_{i=0}^{\infty} \rho^{2i} \sigma_u^2 = \frac{\sigma_u^2}{1-\rho^2}$$

$$\text{cov}(\varepsilon_t, \varepsilon_{t+k}) = E\left(\left(\sum_{i=0}^{\infty} \rho^i u_{t-i}\right)\left(\sum_{j=0}^{\infty} \rho^j u_{t+k-j}\right)\right) = \sum_{i,j=0}^{\infty} \rho^{i+j} E(u_{t-i} u_{t+k-j}) = \sum_{i=0}^{\infty} \rho^{2i+k} \sigma_u^2 = \rho^k \frac{\sigma_u^2}{1-\rho^2}$$

Macierz kowariancyjna  $\text{cov}(\boldsymbol{\varepsilon})$  jest więc postaci

$$\text{cov}(\boldsymbol{\varepsilon}) = \sigma^2 \boldsymbol{\Sigma} = \sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{n-1} \\ \rho & 1 & \rho & \dots & \rho^{n-2} \\ \rho^2 & \rho & 1 & \dots & \rho^{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \dots & 1 \end{bmatrix}, \text{ gdzie } \sigma^2 = \frac{\sigma_u^2}{1-\rho^2}.$$

Można pokazać, że  $\boldsymbol{\Sigma}^{-1} = \frac{1}{1-\rho^2} \begin{bmatrix} 1 & -\rho & 0 & \dots & 0 & 0 \\ -\rho & 1+\rho^2 & -\rho & \dots & 0 & 0 \\ 0 & -\rho & 1+\rho^2 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1+\rho^2 & -\rho \\ 0 & 0 & 0 & \dots & -\rho & 1 \end{bmatrix}$

Drugi ważny przypadek to sytuacja, gdzie błędy w modelu liniowym są nieskorelowane ale mają różne wariancje . Jeżeli  $\boldsymbol{\Sigma} = \text{diag}(\frac{1}{w_1}, \dots, \frac{1}{w_n})$  gdzie wagi  $w_1, \dots, w_n$  są znane to , w tym przypadku

$$\mathbf{P} = \text{diag}(\frac{1}{\sqrt{w_1}}, \dots, \frac{1}{\sqrt{w_n}}) \text{ i } \mathbf{P}^{-1} = \text{diag}(\sqrt{w_1}, \dots, \sqrt{w_n}) \text{ i } \mathbf{P}^{-1}\mathbf{Y} = \mathbf{P}^{-1}\mathbf{X}\boldsymbol{\beta} + \mathbf{P}^{-1}\boldsymbol{\varepsilon} .$$

Wektor odpowiedzi oraz kolumny macierzy planu eksperymentu należy poddać procedurze „ważenia” i zastosować klasyczny estymator MNK. W efekcie stosujemy metodę ważonych najmniejszych kwadratów

Błędy są proporcjonalne do predyktora  $V(\varepsilon_i) \propto x_i$  wówczas  $w_i = \frac{1}{x_i}$  .

Jeżeli obserwacje  $Y_i$  są średnimi z  $n_i$  pomiarów, to  $V(Y_i) = V(\varepsilon_i) = \frac{\sigma^2}{n_i}$  i wówczas  $w_i = n_i$