Teoria Zaufania (CREDIBILITY)

Wojciech Niemiro

Uniwersytet Warszawski i UMK Toruń

LV Szkoła Matematyki Poglądowej
Wola Ducka, styczeń 2017

Plan

1. Przykład
    - Dane
    - Cel analizy
    - Model statystyczny
    - Model „credibility”

2. Teoria zaufania
    - Model Bühlmanna-Strauba
    - Mieszany model liniowy

Plan

1. Przykład
   - Dane
   - Cel analizy
   - Model statystyczny
   - Model „credibility”
2. Teoria zaufania
   - Model Bühlmanna-Strauba
   - Mieszany model liniowy

Dane syntetyczne

klient	czas	liczba strat
1	5	5
2	4	4
3	1	0
4	1	1
5	2	1
6	1	0
7	1	1
8	1	0
9	3	7
10	2	3
11	1	1
12	4	0
13	1	0
14	2	2
15	3	1
razem	32	26

Dane syntetyczne

klient	czas	liczba strat	estymator „indywidualny”
1	5	5	1
2	4	4	1
3	1	0	0
4	1	1	1
5	2	1	0.5
6	1	0	0
7	1	1	1
8	1	0	0
9	3	7	2.33
10	2	3	1.5
11	1	1	1
12	4	0	0
13	1	0	0
14	2	2	1
15	3	1	0.33
razem	32	26	

Dane syntetyczne

klient	czas	liczba strat	estymator „indywidualny”	estymator „kolektywny”
1	5	5	1	0.8125
2	4	4	1	0.8125
3	1	0	0	0.8125
4	1	1	1	0.8125
5	2	1	0.5	0.8125
6	1	0	0	0.8125
7	1	1	1	0.8125
8	1	0	0	0.8125
9	3	7	2.33	0.8125
10	2	3	1.5	0.8125
11	1	1	1	0.8125
12	4	0	0	0.8125
13	1	0	0	0.8125
14	2	2	1	0.8125
15	3	1	0.33	0.8125
razem	32	26		

Cel analizy statystycznej

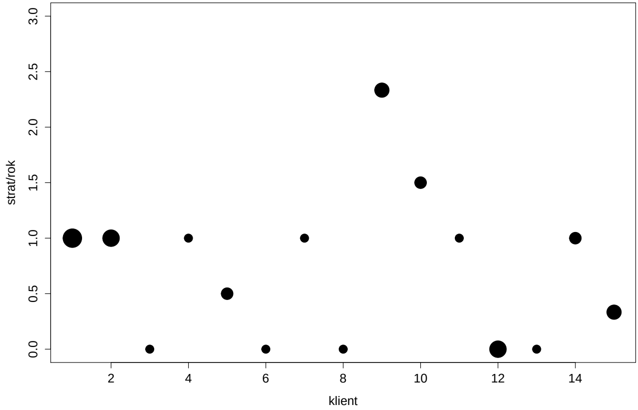
- Możliwie dokładnie przewidzieć przyszłe zachowanie poszczególnych klientów.

- Wyestymować przewidywaną liczbę szkód na rok dla poszczególnych klientów.

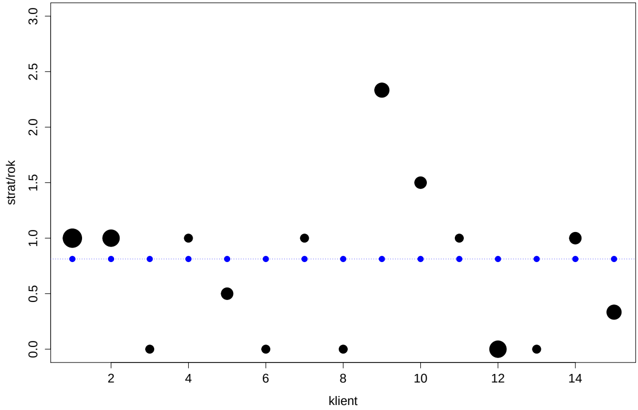
Cel analizy statystycznej

- Możliwie dokładnie przewidzieć przyszłe zachowanie poszczególnych klientów.
- Wyestymować przewidywaną liczbę szkód na rok dla poszczególnych klientów.

Estymatory Indywidualne



Estymatory Indywidualne i Kolektywny



Model statystyczny

- t_i – czas ubezpieczenia (lata) dla i-tego klienta.
- Y_i – liczba szkód zgłoszonych przez i-tego klienta.

$$Y_i \sim \text{Poisson}(t_i\theta_i),$$

$$\mathbf{E}(Y_i|\theta_i) = t_i\theta_i.$$

- θ_i – średnia liczba szkód na rok dla i-tego klienta.
- $\hat{\theta}_i = \frac{Y_i}{t_i}$ – estymator indywidualny dla i-tego klienta.
- $\hat{\theta} = \frac{\sum_i Y_i}{\sum_i t_i}$ – estymator kolektywny (przy założeniu, że $\theta_1 = \cdots = \theta_k$).

Model statystyczny

- t_i – czas ubezpieczenia (lata) dla i-tego klienta.
- Y_i – liczba szkód zgłoszonych przez i-tego klienta.

$$Y_i \sim \text{Poisson}(t_i\theta_i),$$

$$\mathbf{E}(Y_i|\theta_i) = t_i\theta_i.$$

- θ_i – średnia liczba szkód na rok dla i-tego klienta.
- $\hat{\theta}_i = \frac{Y_i}{t_i}$ – estymator indywidualny dla i-tego klienta.
- $\hat{\theta} = \frac{\sum_i Y_i}{\sum_i t_i}$ – estymator kolektywny (przy założeniu, że $\theta_1 = \cdots = \theta_k$).

Model bayesowski

- t_i – czas ubezpieczenia (lata) dla i-tego klienta.
- Y_i – liczba szkód zgłoszonych przez i-tego klienta.
- $Y_i \sim \text{Poisson}(t_i\theta_i)$.

$$p(y_i|\theta_i) = \mathbf{P}(Y_i = y_i|\theta_i) = \mathrm{e}^{-t_i\theta_i}\frac{(t_i\theta_i)^{y_i}}{y_i!}$$

Rozkład *a priori* (losowy wybór klienta z populacji):

- $\theta_i \sim \text{Exponential}(\lambda)$.

$$\pi(\theta_i) = \frac{1}{\lambda}\mathrm{e}^{-\lambda\theta_i}.$$

Rozkład *a posteriori* (wzór Bayesa):

$$\pi(\theta_i|y_i) = \frac{p(y_i|\theta_i)\pi(\theta_i)}{p(y_i)} = \frac{(t_i+\lambda)^{y_i+1}}{y_i!}\theta_i^{y_i}\mathrm{e}^{-(t_i+\lambda)\theta_i}.$$

Rozkład Gamma.

Model bayesowski

- t_i – czas ubezpieczenia (lata) dla i-tego klienta.
- Y_i – liczba szkód zgłoszonych przez i-tego klienta.
- $Y_i \sim \text{Poisson}(t_i\theta_i)$.

$$p(y_i|\theta_i) = \mathbf{P}(Y_i = y_i|\theta_i) = \mathrm{e}^{-t_i\theta_i}\frac{(t_i\theta_i)^{y_i}}{y_i!}$$

Rozkład *a priori* (losowy wybór klienta z populacji):

- $\theta_i \sim \text{Exponential}(\lambda)$.

$$\pi(\theta_i) = \frac{1}{\lambda}\mathrm{e}^{-\lambda\theta_i}.$$

Rozkład *a posteriori* (wzór Bayesa):

$$\pi(\theta_i|y_i) = \frac{p(y_i|\theta_i)\pi(\theta_i)}{p(y_i)} = \frac{(t_i+\lambda)^{y_i+1}}{y_i!}\theta_i^{y_i}\mathrm{e}^{-(t_i+\lambda)\theta_i}.$$

Rozkład Gamma.

Model bayesowski

- t_i – czas ubezpieczenia (lata) dla i-tego klienta.
- Y_i – liczba szkód zgłoszonych przez i-tego klienta.
- $Y_i \sim \mathrm{Poisson}(t_i\theta_i)$.

$$p(y_i|\theta_i) = \mathbf{P}(Y_i = y_i|\theta_i) = \mathrm{e}^{-t_i\theta_i}\frac{(t_i\theta_i)^{y_i}}{y_i!}$$

Rozkład *a priori* (losowy wybór klienta z populacji):

- $\theta_i \sim \mathrm{Exponential}(\lambda)$.

$$\pi(\theta_i) = \frac{1}{\lambda}\mathrm{e}^{-\lambda\theta_i}.$$

Rozkład *a posteriori* (wzór Bayesa):

$$\pi(\theta_i|y_i) = \frac{p(y_i|\theta_i)\pi(\theta_i)}{p(y_i)} = \frac{(t_i+\lambda)^{y_i+1}}{y_i!}\theta_i^{y_i}\mathrm{e}^{-(t_i+\lambda)\theta_i}.$$

Rozkład Gamma.

Model bayesowski

Estymator bayesowski:

$$\hat{\theta}_i = \mathbf{E}(\theta_i | Y_i) = z_i \frac{Y_i}{t_i} + (1 - z_i) m,$$

gdzie

- $\frac{Y_i}{t_i}$ - średnia z danych (estymator indywidualny).
- m - średnia w całej populacji (estymator *a priori*, $\mathbf{E} Y_i = t_i m$).
- $z_i = \frac{t_i}{t_i + \lambda}$ - *współczynnik zaufania*.

Empiryczne podejście bayesowskie

Estymator bayesowski dla *i*-tego klienta

$$\hat{\theta}_i = z_i \frac{Y_i}{t_i} + (1 - z_i)m,$$

- zależy od danych indywidualnych Y_i,
- zależy od rozkładu a priori (poprzez m i λ).

Estymujemy rozkład a priori na podstawie wszystkich danych $Y_1, \ldots, Y_k$.

Empiryczny estymator bayesowski:

$$\hat{\theta}_i = \hat{z}_i \frac{Y_i}{t_i} + (1 - \hat{z}_i)\hat{m}.$$

Empiryczne podejście bayesowskie

Estymator bayesowski dla *i*-tego klienta

$$\hat{\theta}_i = z_i \frac{Y_i}{t_i} + (1 - z_i) m,$$

- zależy od danych indywidualnych Y_i,
- zależy od rozkładu a priori (poprzez m i λ).

Estymujemy rozkład a priori na podstawie wszystkich danych $Y_1, \ldots, Y_k$.

Empiryczny estymator bayesowski:

$$\hat{\theta}_i = \hat{z}_i \frac{Y_i}{t_i} + (1 - \hat{z}_i) \hat{m}.$$

Empiryczne podejście bayesowskie

Estymujemy rozkład a priori na podstawie wszystkich danych $Y_1, \ldots, Y_k$**.**

Rozkład **brzegowy** Y_i:

$$\mathbf{P}(Y_i = y_i) = p(y_i) = \int \mathbf{P}(Y_i = y_i | \theta_i) \mathrm{d}\theta_i.$$

W naszym przykładzie:

$$p(y_i) = \left(\frac{t_i}{t_i + \lambda}\right)^{y_i} \frac{\lambda}{t_i + \lambda}.$$

Rozkład geometryczny.

Estymator największej wiarygodności, $p(y_1) \cdots p(y_k) \to \max_\lambda$: rozwiązanie równania

$$\frac{k}{\lambda} = \sum_{i=1}^{k} \frac{Y_i + 1}{t_i + \lambda}.$$

Empiryczne podejście bayesowskie

Estymujemy rozkład a priori na podstawie wszystkich danych $Y_1, \ldots, Y_k$**.**

Rozkład **brzegowy** Y_i:

$$\mathbf{P}(Y_i = y_i) = p(y_i) = \int \mathbf{P}(Y_i = y_i|\theta_i)\mathrm{d}\theta_i.$$

W naszym przykładzie:

$$p(y_i) = \left(\frac{t_i}{t_i + \lambda}\right)^{y_i} \frac{\lambda}{t_i + \lambda}.$$

Rozkład geometryczny.

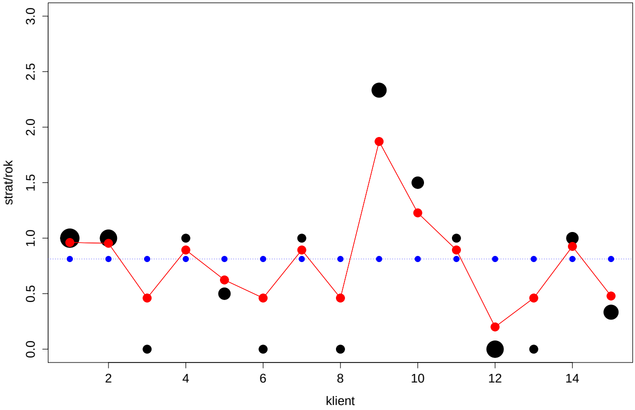
Estymator największej wiarygodności, $p(y_1) \cdots p(y_k) \to \max_\lambda$: rozwiązanie równania

$$\frac{k}{\lambda} = \sum_{i=1}^{k} \frac{Y_i + 1}{t_i + \lambda}.$$

Estymatory „teorii zaufania”

klient	czas	liczba strat	estymator „indywidualny”	estymator „kolektywny”	estymator „credibility”
1	5	5	1	0.8125	0.9611
2	4	4	1	0.8125	0.9537
3	1	0	0	0.8125	0.4608
4	1	1	1	0.8125	0.8937
5	2	1	0.5	0.8125	0.6237
6	1	0	0	0.8125	0.4608
7	1	1	1	0.8125	0.8937
8	1	0	0	0.8125	0.4608
9	3	7	2.33	0.8125	1.8710
10	2	3	1.5	0.8125	1.2279
11	1	1	1	0.8125	0.8937
12	4	0	0	0.8125	0.2005
13	1	0	0	0.8125	0.4608
14	2	2	1	0.8125	0.9258
15	3	1	0.33	0.8125	0.4790

Estymatory „CREDIBILITY”



Model teorii zaufania (Bühlmanna-Strauba)

Dane: Y_{ij} - szkody i-tego klienta w j-tym roku, $i = 1, \ldots, k$, $j = 1, \ldots, n_i$. Parametry (nieobserwowane modelowe zmienne losowe): θ_i.

$$
\begin{array}{lllllllll}
\theta_1; & Y_{11}, & \ldots & Y_{1j}, & \ldots & Y_{1n_1}, & & & \\
\vdots & \vdots & \ddots & \vdots & \ddots & & & & \\
\theta_i; & Y_{i1}, & \ldots & Y_{ij}, & \ldots & \ldots & \ldots & & Y_{in_i}, \\
\vdots & \vdots & \ddots & \vdots & \ddots & & & & \\
\theta_k; & Y_{k1}, & \ldots & Y_{kj}, & \ldots & \ldots & Y_{kn_k}. & &
\end{array}
$$

- $\theta_1, \ldots, \theta_k \sim_{\text{iid}} \pi(\cdot)$;
- $Y_{i1}, \ldots, Y_{in_i} \sim_{\text{iid}} p(\cdot|\theta_i)$.

Łączny rozkład prawdopodobieństwa:

$$p((\theta_i), (y_{ij})) = \prod_{i=1}^{k} \pi(\theta_i) \prod_{j=1}^{n_i} p(y_{ij}|\theta_i).$$

Model Bühlmanna-Strauba

Cel: estymacja/predykcja $\mu(\theta_i)$, gdzie

$$\mu(\theta) = \int y p(y|\theta) \mathrm{d}\theta,$$

czyli

$$\mu(\theta_i) = \mathbf{E}(Y_{ij}|\theta_i) = \mathbf{E}(Y_{i,\text{nowy } j}|\theta_i).$$

Model Bühlmanna-Strauba = mieszany model liniowy

$$Y_{ij} = m + \underbrace{(\mu(\theta_i) - m)}_{=\alpha_i \text{ efekt losowy}} + \underbrace{(Y_{ij} - \mu(\theta_i))}_{=\varepsilon_{ij} \text{ błąd losowy}}.$$

Mieszany model liniowy:

$$Y_{ij} = m + \alpha_i + \varepsilon_{ij},$$

α_i, ε_{ij} nieskorelowane, $\mathbf{E}\alpha_i = 0$, $\mathbf{E}\varepsilon_{ij} = 0$, $\mathbf{Var}\alpha_i = v^2$, $\mathbf{Var}\varepsilon_{ij} = s^2$.

Cel: estymacja/predykcja $\mu_i = \mu(\theta_i) = m + \alpha_i$.

Mieszany model liniowy: BLUP i EBLUP

Mieszany model liniowy:

$$Y_{ij} = m + \alpha_i + \varepsilon_{ij},$$

α_i, ε_{ij} nieskorelowane, $\mathbf{E}\alpha_i = 0$, $\mathbf{E}\varepsilon_{ij} = 0$, $\mathbf{Var}\alpha_i = v^2$, $\mathbf{Var}\varepsilon_{ij} = s^2$.

Cel: estymacja/predykcja $\mu_i = m + \alpha_i$.

BLUP (*Best Linear Unbiased Predictor*):

$$\hat{\mu}_i = z_i \bar{Y}_i + (1 - z_i)\bar{Y},$$

gdzie $\bar{Y}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} Y_{ij}$, $z_i = \frac{n_i v^2}{n_i v^2 + s^2}$, $\bar{Y} = \frac{\sum_i z_i \bar{Y}_i}{\sum_i z_i}$.

EBLUP (*Empirical* BLUP): $z_i = \frac{n_i v^2}{n_i v^2 + s^2} \longrightarrow \hat{z}_i = \frac{n_i \hat{v}^2}{n_i \hat{v}^2 + \hat{s}^2}$
(estymujemy komponenty wariancyjne v^2 i s^2; Searle, Casella & McCulloch).

Mieszany model liniowy: BLUP i EBLUP

Mieszany model liniowy:

$$Y_{ij} = m + \alpha_i + \varepsilon_{ij},$$

α_i, ε_{ij} nieskorelowane, $\mathbf{E}\alpha_i = 0$, $\mathbf{E}\varepsilon_{ij} = 0$, $\mathbf{Var}\alpha_i = v^2$, $\mathbf{Var}\varepsilon_{ij} = s^2$.

Cel: estymacja/predykcja $\mu_i = m + \alpha_i$.

BLUP (*Best Linear Unbiased Predictor*):

$$\hat{\mu}_i = z_i \bar{Y}_i + (1 - z_i)\bar{Y},$$

gdzie $\bar{Y}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} Y_{ij}$, $z_i = \frac{n_i v^2}{n_i v^2 + s^2}$, $\bar{Y} = \frac{\sum_i z_i \bar{Y}_i}{\sum_i z_i}$.

EBLUP (*Empirical* BLUP): $z_i = \frac{n_i v^2}{n_i v^2 + s^2} \longrightarrow \hat{z}_i = \frac{n_i \hat{v}^2}{n_i \hat{v}^2 + \hat{s}^2}$

(estymujemy komponenty wariancyjne v^2 i s^2; Searle, Casella & McCulloch).

Mieszany model liniowy: BLUP i EBLUP

Mieszany model liniowy:

$$Y_{ij} = m + \alpha_i + \varepsilon_{ij},$$

α_i, ε_{ij} nieskorelowane, $\mathbf{E}\alpha_i = 0$, $\mathbf{E}\varepsilon_{ij} = 0$, $\mathbf{Var}\alpha_i = v^2$, $\mathbf{Var}\varepsilon_{ij} = s^2$.

Cel: estymacja/predykcja $\mu_i = m + \alpha_i$.

BLUP (*Best Linear Unbiased Predictor*):

$$\hat{\mu}_i = z_i \bar{Y}_i + (1 - z_i)\bar{Y},$$

gdzie $\bar{Y}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} Y_{ij}$, $z_i = \frac{n_i v^2}{n_i v^2 + s^2}$, $\bar{Y} = \frac{\sum_i z_i \bar{Y}_i}{\sum_i z_i}$.

EBLUP (*Empirical* BLUP): $z_i = \frac{n_i v^2}{n_i v^2 + s^2} \longrightarrow \hat{z}_i = \frac{n_i \hat{v}^2}{n_i \hat{v}^2 + \hat{s}^2}$

(estymujemy komponenty wariancyjne v^2 i s^2; Searle, Casella & McCulloch).

Mieszany model liniowy: BLUP i EBLUP

Mieszany model liniowy:

$$Y_{ij} = m + \alpha_i + \varepsilon_{ij},$$

α_i, ε_{ij} nieskorelowane, $\mathbf{E}\alpha_i = 0$, $\mathbf{E}\varepsilon_{ij} = 0$, $\mathbf{Var}\alpha_i = v^2$, $\mathbf{Var}\varepsilon_{ij} = s^2$.

Cel: estymacja/predykcja $\mu_i = m + \alpha_i$.

BLUP (*Best Linear Unbiased Predictor*):

$$\hat{\mu}_i = z_i \bar{Y}_i + (1 - z_i)\bar{Y},$$

gdzie $\bar{Y}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} Y_{ij}$, $z_i = \frac{n_i v^2}{n_i v^2 + s^2}$, $\bar{Y} = \frac{\sum_i z_i \bar{Y}_i}{\sum_i z_i}$.

EBLUP (*Empirical* BLUP): $z_i = \frac{n_i v^2}{n_i v^2 + s^2} \longrightarrow \hat{z}_i = \frac{n_i \hat{v}^2}{n_i \hat{v}^2 + \hat{s}^2}$
(estymujemy komponenty wariancyjne v^2 i s^2; Searle, Casella & McCulloch).

Mieszany model liniowy, różne warianty:

Możemy włączyć do modelu zmienne objaśniające:

$$Y_{ij} = m + \alpha_i + \varepsilon_{ij} + \sum_{l=1}^{d} x_{ijl}.$$

Możemy zbudować model hierarchczny

$$Y_{ijl} = m + \alpha_i + \beta_{il} + \varepsilon_{ilj}.$$

Możemy uwzględnić wpływ 2 czynników losowych (krzyżowy):

$$Y_{ijl} = m + \alpha_i + \beta_l + \varepsilon_{ilj}.$$

Itp. itd.

Mieszany model liniowy, różne warianty:

Możemy włączyć do modelu zmienne objaśniające:

$$Y_{ij} = m + \alpha_i + \varepsilon_{ij} + \sum_{l=1}^{d} x_{ijl}.$$

Możemy zbudować model hierarchczny

$$Y_{ijl} = m + \alpha_i + \beta_{il} + \varepsilon_{ilj}.$$

Możemy uwzględnić wpływ 2 czynników losowych (krzyżowy):

$$Y_{ijl} = m + \alpha_i + \beta_l + \varepsilon_{ilj}.$$

Itp. itd.

Mieszany model liniowy, różne warianty:

Możemy włączyć do modelu zmienne objaśniające:

$$Y_{ij} = m + \alpha_i + \varepsilon_{ij} + \sum_{l=1}^{d} x_{ijl}.$$

Możemy zbudować model hierarchczny

$$Y_{ijl} = m + \alpha_i + \beta_{il} + \varepsilon_{ilj}.$$

Możemy uwzględnić wpływ 2 czynników losowych (krzyżowy):

$$Y_{ijl} = m + \alpha_i + \beta_l + \varepsilon_{ilj}.$$

Itp. itd.

Mieszany model liniowy, różne warianty:

Możemy włączyć do modelu zmienne objaśniające:

$$Y_{ij} = m + \alpha_i + \varepsilon_{ij} + \sum_{l=1}^{d} x_{ijl}.$$

Możemy zbudować model hierarchczny

$$Y_{ijl} = m + \alpha_i + \beta_{il} + \varepsilon_{ilj}.$$

Możemy uwzględnić wpływ 2 czynników losowych (krzyżowy):

$$Y_{ijl} = m + \alpha_i + \beta_l + \varepsilon_{ilj}.$$

Itp. itd.

Mieszany model liniowy

Trochę ogólniej, w języku macierzowym:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{\alpha} + \boldsymbol{\varepsilon},$$

$$\underset{n\times 1}{\boldsymbol{y}} = \underset{n\times p}{\boldsymbol{X}} \cdot \underset{p\times 1}{\boldsymbol{\beta}} + \underset{n\times k}{\boldsymbol{Z}} \cdot \underset{k\times 1}{\boldsymbol{\alpha}} + \underset{n\times 1}{\boldsymbol{\varepsilon}} .$$

$\boldsymbol{X}$ i $\boldsymbol{Z}$ – znane macierze planu,

$\boldsymbol{\beta}$ – nielosowy wektor „efektów stałych”,

$\boldsymbol{\alpha}$ – wektor „efektów losowych”,

$\boldsymbol{\varepsilon}$ – wektor „błędów losowych”,

$\mathbf{E}(\boldsymbol{\varepsilon}) = \mathbf{0}$, $\mathbf{E}(\boldsymbol{\alpha}) = \mathbf{0}$, $\mathbf{VAR}(\boldsymbol{\varepsilon}) = \boldsymbol{S}$, $\mathbf{VAR}(\boldsymbol{\alpha}) = \boldsymbol{V}$, $\mathbf{COV}(\boldsymbol{\alpha}, \boldsymbol{\varepsilon}) = \mathbf{0}$.

Mieszany model liniowy: BLUE i BLUP

$$\boldsymbol{y} = \boldsymbol{X}\beta + \boldsymbol{Z}\alpha + \varepsilon.$$

Cel: estymacja β i predykcja α.

BLUE i BLUP:

$$\hat{\beta} = (\boldsymbol{X}^\top(\boldsymbol{S} + \boldsymbol{Z}\boldsymbol{V}\boldsymbol{Z}^\top)^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^\top(\boldsymbol{S} + \boldsymbol{Z}\boldsymbol{V}\boldsymbol{Z}^\top)^{-1}\boldsymbol{y},$$

$$\hat{\alpha} = \boldsymbol{V}\boldsymbol{Z}^\top(\boldsymbol{S} + \boldsymbol{Z}\boldsymbol{V}\boldsymbol{Z}^\top)^{-1}(\boldsymbol{y} - \boldsymbol{X}\hat{\beta}).$$

- $\mathrm{VAR}(\boldsymbol{y}) = \boldsymbol{S} + \boldsymbol{Z}\boldsymbol{V}\boldsymbol{Z}^\top$,
- $\mathrm{COV}(\alpha, \boldsymbol{y}) = \boldsymbol{V}\boldsymbol{Z}^\top$.

Mieszany model liniowy: BLUE i BLUP

$$\boldsymbol{y} = \boldsymbol{X}\beta + \boldsymbol{Z}\alpha + \varepsilon.$$

Cel: estymacja β i predykcja α.

BLUE i BLUP:

$$\hat{\beta} = (\boldsymbol{X}^\top(\boldsymbol{S} + \boldsymbol{Z}\boldsymbol{V}\boldsymbol{Z}^\top)^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^\top(\boldsymbol{S} + \boldsymbol{Z}\boldsymbol{V}\boldsymbol{Z}^\top)^{-1}\boldsymbol{y},$$

$$\hat{\alpha} = \boldsymbol{V}\boldsymbol{Z}^\top(\boldsymbol{S} + \boldsymbol{Z}\boldsymbol{V}\boldsymbol{Z}^\top)^{-1}(\boldsymbol{y} - \boldsymbol{X}\hat{\beta}).$$

- $\mathrm{VAR}(\boldsymbol{y}) = \boldsymbol{S} + \boldsymbol{Z}\boldsymbol{V}\boldsymbol{Z}^\top$,
- $\mathrm{COV}(\alpha, \boldsymbol{y}) = \boldsymbol{V}\boldsymbol{Z}^\top$.

Mieszany model liniowy: BLUE i BLUP

$$\boldsymbol{y} = \boldsymbol{X}\beta + \boldsymbol{Z}\alpha + \varepsilon.$$

Cel: estymacja β i predykcja α.

BLUE i BLUP:

$$\hat{\beta} = (\boldsymbol{X}^\top(\boldsymbol{S} + \boldsymbol{Z}\boldsymbol{V}\boldsymbol{Z}^\top)^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^\top(\boldsymbol{S} + \boldsymbol{Z}\boldsymbol{V}\boldsymbol{Z}^\top)^{-1}\boldsymbol{y},$$

$$\hat{\alpha} = \boldsymbol{V}\boldsymbol{Z}^\top(\boldsymbol{S} + \boldsymbol{Z}\boldsymbol{V}\boldsymbol{Z}^\top)^{-1}(\boldsymbol{y} - \boldsymbol{X}\hat{\beta}).$$

- $\mathrm{VAR}(\boldsymbol{y}) = \boldsymbol{S} + \boldsymbol{Z}\boldsymbol{V}\boldsymbol{Z}^\top$,
- $\mathrm{COV}(\alpha, \boldsymbol{y}) = \boldsymbol{V}\boldsymbol{Z}^\top$.

Mieszany model liniowy: EBLUP i EBLUP

$$\boldsymbol{y} = \boldsymbol{X}\beta + \boldsymbol{Z}\alpha + \varepsilon.$$

Cel: estymacja β i predykcja α.

EBLUE i EBLUP:

$$\hat{\beta} = (\boldsymbol{X}^\top(\hat{\boldsymbol{S}} + \boldsymbol{Z}\hat{\boldsymbol{V}}\boldsymbol{Z}^\top)^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^\top(\hat{\boldsymbol{S}} + \boldsymbol{Z}\hat{\boldsymbol{V}}\boldsymbol{Z}^\top)^{-1}\boldsymbol{y},$$

$$\hat{\alpha} = \hat{\boldsymbol{V}}\boldsymbol{Z}^\top(\hat{\boldsymbol{S}} + \boldsymbol{Z}\hat{\boldsymbol{V}}\boldsymbol{Z}^\top)^{-1}(\boldsymbol{y} - \boldsymbol{X}\hat{\beta}).$$

EBLUE i EBLUP: estymujemy macierze kowariancji $\boldsymbol{V}$ i $\boldsymbol{S}$; Searle, Casella & McCulloch).

Mieszany model liniowy: EBLUP i EBLUP

$$\boldsymbol{y} = \boldsymbol{X}\beta + \boldsymbol{Z}\alpha + \varepsilon.$$

Cel: estymacja β i predykcja α.

EBLUE i EBLUP:

$$\hat{\beta} = (\boldsymbol{X}^\top(\hat{\boldsymbol{S}} + \boldsymbol{Z}\hat{\boldsymbol{V}}\boldsymbol{Z}^\top)^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^\top(\hat{\boldsymbol{S}} + \boldsymbol{Z}\hat{\boldsymbol{V}}\boldsymbol{Z}^\top)^{-1}\boldsymbol{y},$$

$$\hat{\alpha} = \hat{\boldsymbol{V}}\boldsymbol{Z}^\top(\hat{\boldsymbol{S}} + \boldsymbol{Z}\hat{\boldsymbol{V}}\boldsymbol{Z}^\top)^{-1}(\boldsymbol{y} - \boldsymbol{X}\hat{\beta}).$$

EBLUE i EBLUP: estymujemy macierze kowariancji $\boldsymbol{V}$ i $\boldsymbol{S}$; Searle, Casella & McCulloch).

Mieszany model liniowy: EBLUP i EBLUP

$$\boldsymbol{y} = \boldsymbol{X}\beta + \boldsymbol{Z}\alpha + \varepsilon.$$

Cel: estymacja β i predykcja α.

EBLUE i EBLUP:

$$\hat{\beta} = (\boldsymbol{X}^\top(\hat{\boldsymbol{S}} + \boldsymbol{Z}\hat{\boldsymbol{V}}\boldsymbol{Z}^\top)^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^\top(\hat{\boldsymbol{S}} + \boldsymbol{Z}\hat{\boldsymbol{V}}\boldsymbol{Z}^\top)^{-1}\boldsymbol{y},$$

$$\hat{\alpha} = \hat{\boldsymbol{V}}\boldsymbol{Z}^\top(\hat{\boldsymbol{S}} + \boldsymbol{Z}\hat{\boldsymbol{V}}\boldsymbol{Z}^\top)^{-1}(\boldsymbol{y} - \boldsymbol{X}\hat{\beta}).$$

EBLUE i EBLUP: estymujemy macierze kowariancji $\boldsymbol{V}$ i $\boldsymbol{S}$; Searle, Casella & McCulloch).

Literatura

- P. Biecek. „Analiza danych z programem R. Modele liniowe z efektami stałymi, losowymi i mieszanymi”, Wydawnictwo Naukowe PWN, 2013.
- H. Bühlmann and A. Gisler. *A Course in Credibility Theory and its Applications*, Published by Springer 2005.
- W. Niemiro. Liniowe modele klasyfikacji z efektami losowymi w ubezpieczeniowej teorii wiarogodności. *Roczniki Kolegium Analiz Ekonomicznych SGH*, 8 (2000), 102–114.
- G. K. Robinson. That BLUP is a Good Thing: The Estimation of Random Effects. *Statistical Science*, Vol. 6, No. 1 (Feb., 1991), pp. 15-32.
- S.R. Searle, G. Casella & C.E. McCulloch, *Variance Components*, Wiley 2006.