Modele mieszane

Model efektów losowych- model II komponentów wariancyjnych

Dotychczas rozważane modele liniowe 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 potraktujemy poziomy czynnika nie jako z góry ustalone, ale jako wylosowane z pewnej hipotetycznej populacji poziomów to taki model nazywamy modelem efektów losowych (**model II**). Może być kilka przyczyn dla których efekty poziomów pewnej zmiennej objaśniającej zdecydujemy się potraktować jako efekty losowe.

- Efektów tych może być tak dużo, że zamiast traktować je jako parametry modelu, przyjmiemy, że są to zmienne losowe z rozkładu, którego parametry będą parametrami modelu.
- Możemy też nie obserwować wszystkich możliwych poziomów zmiennej objaśniającej, a efekt obserwowanych poziomów potraktować jako realizację pewnej zmiennej losowej opisującej efekty w całej populacji.
- Liczba poziomów zmiennej objaśniającej zależy od liczby obserwacji. Jeżeli do badania rekrutujemy więcej osób/obiektów, to możemy obserwować większą liczbę poziomów zmiennej objaśniającej
- W ogólnej sytuacji trudno podać definicję efektów losowego i stałego. Ta sama zmienna może być uwzględniona w jednym modelu przez efekty stałe a w innym przez losowe. Wybór typu efektów często zależy od interpretacji lub kontekstu. Rozróżnienie między efektami stałym a losowymi jest o tyle istotne, że możemy otrzymać różne wyniki w zależności od tego, czy pewną zmienną opiszemy efektami stałymi czy losowymi. Dobrze jest mieć argumenty na obronę dokonanego wyboru.

Podobne jak w klasycznym modelu $\mathbf{Y} = \mathbf{X\beta} + \boldsymbol{\varepsilon} \quad , \quad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$ interesuje nas opisanie zależności pomiędzy zmienną objaśnianą y a zbiorem zmiennych objaśniających. Część efektów zmiennych objaśniających traktowana będzie jako efekty stałe a część jako efekty losowe. Macierz modelu opisująca zmienne odpowiadające efektom stałym oznaczamy symbolem $\mathbf{X}$ a macierz opisująca zmienne odpowiadające efektom losowym symbolem $\mathbf{Z}$. Przyjmując liniową zależność między kolumnami macierzy modelu a zmienną objaśnianą możemy rozważyć model postaci

$$\mathbf{Y} = \mathbf{X\beta} + \mathbf{Zu} + \boldsymbol{\varepsilon},$$

gdzie $\mathbf{Y}$ -obserwowany wektor wartości zmiennej objaśnianej

$\mathbf{X}$ - macierz (n,p) zmiennych objaśniających dla efektów stałych

$\boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}$ wektor nieznanych efektów stałych

$\mathbf{Z}$ - macierz (n,q) zmiennych objaśniających dla efektów losowych

$\mathbf{u} \sim N_q(\mathbf{0}, \sigma^2 D)$ - wektor zmiennych losowych odpowiadający efektom losowym

$\boldsymbol{\varepsilon} \sim N_n(\mathbf{0}, \sigma^2 \mathbf{I})$ - zakłócenie losowe

Równoważnie moglibyśmy określić ten model jako

$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$, gdzie $\mathbf{e} = \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon}$, więc $V(\mathbf{e}) = \sigma^2(\mathbf{I} + \mathbf{Z}\mathbf{D}\mathbf{Z}^{\mathrm{T}})$.

Podsumowując, $\mathbf{Y}$ jest wektorem losowym o rozkładzie brzegowym

$$\mathbf{Y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2(\mathbf{I} + \mathbf{Z}\mathbf{D}\mathbf{Z}^{\mathrm{T}}))$$

oraz o rozkładzie warunkowym

$$\mathbf{Y} \mid \mathbf{u} \sim N(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}, \sigma^2\mathbf{I})).$$

Przykładowe modele mieszane i ich zastosowania Rencher 481

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_1\mathbf{a}_1 + \cdots + \mathbf{Z}_m\mathbf{a}_m + \boldsymbol{\varepsilon} \ ,$$

gdzie jak zwykle w modelach liniowych $E(\boldsymbol{\varepsilon}) = \mathbf{0}$, $\mathrm{cov}(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{I}_n$, macierz $\mathbf{X}$ -(n,p) znana (niekoniecznie pełnego rzędu) $\boldsymbol{\beta}$ -(p,1) wektor nieznanych parametrów. Macierze $\mathbf{Z}_i$ o wymiarach (n,r_i) są pełnego rzędu i opisują przynależność jednostek eksperymentalnych do różnych grup. Nowością jest to że $\mathbf{a}_i$ są (r_i,1) wektorami losowymi o których zakładamy, że $E(\mathbf{a}_i) = \mathbf{0}$, $\mathrm{cov}(\mathbf{a}_i) = \sigma_i^2\mathbf{I}_{r_i}$. Założymy również , że $\mathrm{cov}(\mathbf{a}_i, \mathbf{a}_j) = \mathbf{0}_{(r_i, r_j)}$ $\mathrm{cov}(\mathbf{a}_i, \boldsymbol{\varepsilon}) = \mathbf{0}_{(r_i, n)}$. Założymy także normalność wektorów losowych występujących w modelu, co umożliwi zastosowanie podejścia bayesowskiego do estymacji i testowania hipotez.

- Randomized Block Design (Bloki losowe). RDB to metoda eksperymentalna powszechnie stosowana w rolnictwie, farmacji i eksperymentach przemysłowych, gdzie jednostki eksperymentalne są grupowane w jednorodne bloki na podstawie wspólnych cech (zmiennych zakłócających, które mają potencjalny wpływ na odpowiedź ale nie są one głównym problemem badawczym) a następnie losowe przypisanie zabiegów (primary factor) do jednostek eksperymentalnych w obrębie każdego bloku zwiększając precyzję wnioskowania o zabiegach zmniejszając ryzyko błędu eksperymentalnego poprzez uwzględnienie systematycznych różnic miedzy blokami. (Eksperyment obejmujący 3 zabiegi został przeprowadzony przez losowe

przypisanie zabiegów do jednostek eksperymentalnych w każdym z 4 bloków o rozmiarze 3. Możemy wiec użyć modelu

$$Y_{ij} = \mu + \tau_i + a_j + \varepsilon_{ij}\,;\; i{=}1 \div 3;\; j{=}1 \div 4$$

$$a_j \sim N(0,\sigma_1^2)\,;\; \varepsilon_{ij} \sim N(0,\sigma^2)\,;\; \mathrm{cov}(a_j,\varepsilon_{ij}) = 0$$

Jeżeli przyjmiemy, że obserwacje są sortowane według bloków i zabiegów w obrębie bloków, możemy wyrazić ten model w postaci $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_1\mathbf{a}_1 + \boldsymbol{\varepsilon}$

$$\begin{bmatrix} Y_{11} \\ Y_{21} \\ Y_{31} \\ Y_{12} \\ Y_{22} \\ Y_{32} \\ Y_{13} \\ Y_{23} \\ Y_{33} \\ Y_{14} \\ Y_{24} \\ Y_{34} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \tau_3 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix} + \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{21} \\ \varepsilon_{31} \\ \varepsilon_{12} \\ \varepsilon_{22} \\ \varepsilon_{32} \\ \varepsilon_{13} \\ \varepsilon_{23} \\ \varepsilon_{33} \\ \varepsilon_{14} \\ \varepsilon_{24} \\ \varepsilon_{34} \end{bmatrix}$$

Kolumny w macierzy efektów stałych są zależne(pierwsza jest sumą pozostałych) .Program R zastosuje kodowanie oszczędne przyjmując $\tau_1 = 0$.

$$\begin{bmatrix} Y_{11} \\ Y_{21} \\ Y_{31} \\ Y_{12} \\ Y_{22} \\ Y_{32} \\ Y_{13} \\ Y_{23} \\ Y_{33} \\ Y_{14} \\ Y_{24} \\ Y_{34} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix} + \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{21} \\ \varepsilon_{31} \\ \varepsilon_{12} \\ \varepsilon_{22} \\ \varepsilon_{32} \\ \varepsilon_{13} \\ \varepsilon_{23} \\ \varepsilon_{33} \\ \varepsilon_{14} \\ \varepsilon_{24} \\ \varepsilon_{34} \end{bmatrix}$$

Struktura kowariancyjna wektora obserwacji jest tu następująca

$$\boldsymbol{\Sigma} = \sigma_1^2 \mathbf{Z}_1 \mathbf{Z}_1^T + \sigma^2 \mathbf{I}_{12} = \begin{bmatrix} \boldsymbol{\Sigma}_1 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\Sigma}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{\Sigma}_1 \end{bmatrix}, \text{ gdzie } \boldsymbol{\Sigma}_1 = \begin{bmatrix} \sigma^2 + \sigma_1^2 & \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma^2 + \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 & \sigma^2 + \sigma_1^2 \end{bmatrix}$$

Interpretacja parametrów jest następująca

β_1 - średnia (oczekiwana) wartość odpowiedzi przy pierwszym zabiegu

β_2 - różnica średnich odpowiedzi pomiędzy zabiegiem drugim i pierwszym

β_3 - różnica średnich odpowiedzi pomiędzy zabiegiem trzecim i pierwszym

```
Realizacja modelu w R
set.seed(123)
y1<-c(5,6,7,6,7,8,7,8,9,8,9,10)+rnorm(12)  #generujemy 12 elementowy wektor odpowiedzi  y
x1<-c(1,2,3,1,2,3,1,2,3,1,2,3)
z1=c(1,1,1,2,2,2,3,3,3,4,4,4)  #zmienna jakościowa o 4 poziomach
dane1<-data.frame(y1,x1,z1)  #tworzymy ramkę danych
dane1$x1<-factor(dane1$x1)  #nadajemy kody  dla zmiennej jakościowej
levels(dane1$x1) <- c("T1","T2","T3")
dane1$z1<-factor(dane1$z1)  #nadajemy kody  dla zmiennej jakościowej
levels(dane1$z1) <- c("A","B","C","D")
dane1
library(lme4)
model1<-lmer(y1~x1+(1|z1),data=dane1,REML=FALSE,control = lmerControl(optimizer ="Nelder_Mead"))#estymacja metodą ML
model1@pp$Zt # transponowana macierz Z odczytana ze slotu model0@pp obiekty klasy S4
t(model1@pp$Zt) # macierz Z
image(getME(model1,"Z"))#graficzny obraz macierzy Z dla modelu 1
t(model1@pp$Zt) # macierz Z
t(model1@pp$Zt)%*%model1@pp$Zt
model1@pp$X # macierz X
image(getME(model1,"X"))#graficzny obraz macierzy X dla modelu 1
summary(model1)

#wyznaczymy predykcję efektów losowych
#argument condVar=TRUE dodaje do wyniku informacje o odchyleniach standardowych
library(lattice)
(u=ranef(model1,condVar=TRUE))
dotplot(u)

ranef(model1,drop=TRUE) # predykcje efektów losowych z warunkowymi wariancjami
fixef(model1)
```

Zadanie. Zapisać w postaci macierzowej ten model jeśli dane sortowane według zabiegów i bloków w obrębie zabiegów.

- Subsampling (Podpróbkowanie). Subsampling, to technika redukcji ilości danych przez wybieranie tylko części próbek z większego zbioru, często stosowana w przetwarzaniu sygnałów i obrazów, sieciach neuronowych i statystycznej kontroli jakości kosztem pewnej utraty szczegółów. Wytworzono trzy partie przy użyciu każdego z dwóch procesów. Z każdej partii uzyskano i zmierzono dwie próbki. Ograniczając efekty procesu do sumy zerowej, model jest następujący

$$Y_{ijk} = \mu + \tau_i + a_{ij} + \varepsilon_{ijk} \quad ; i=1,2\ ,\ j=1\div 3,\ k=1,2,\ \tau_2 = -\tau_1$$

$$a_{ij} \sim N(0,\sigma_1^2)\ ;\ \varepsilon_{ijk} \sim N(0,\sigma^2)\ ;\ \mathrm{cov}(a_{ij},\varepsilon_{ijk}) = 0$$

Jeżeli obserwacje są sortowane według procesów, partii w procesach i próbek w partiach, możemy zapisać ten model w postaci $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_1\mathbf{a}_1 + \boldsymbol{\varepsilon}$

$$\begin{bmatrix} Y_{111} \\ Y_{112} \\ Y_{121} \\ Y_{122} \\ Y_{131} \\ Y_{132} \\ Y_{211} \\ Y_{212} \\ Y_{221} \\ Y_{222} \\ Y_{231} \\ Y_{232} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & -1 \\ 1 & -1 \\ 1 & -1 \\ 1 & -1 \\ 1 & -1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} \mu \\ \tau \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} a_{11} \\ a_{12} \\ a_{13} \\ a_{21} \\ a_{22} \\ a_{23} \end{bmatrix}$$

Parametryzacja R (inaczej parametryzuje efekty stałe)

$$\begin{bmatrix} Y_{111} \\ Y_{112} \\ Y_{121} \\ Y_{122} \\ Y_{131} \\ Y_{132} \\ Y_{211} \\ Y_{212} \\ Y_{221} \\ Y_{222} \\ Y_{231} \\ Y_{232} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \beta \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} a_{11} \\ a_{12} \\ a_{13} \\ a_{21} \\ a_{22} \\ a_{23} \end{bmatrix}$$

Struktura kowariancyjna wektora obserwacji jest tu następująca

$$\boldsymbol{\Sigma} = \sigma_1^2 \mathbf{Z}_1 \mathbf{Z}_1^T + \sigma^2 \mathbf{I}_{12} = \begin{bmatrix} \boldsymbol{\Sigma}_1 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_1 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\Sigma}_1 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{\Sigma}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{\Sigma}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{\Sigma}_1 \end{bmatrix}, \text{ gdzie } \boldsymbol{\Sigma}_1 = \begin{bmatrix} \sigma^2 + \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma^2 + \sigma_1^2 \end{bmatrix}$$

```
#Rencher  Subsampling
set.seed(123)
y2<-rnorm(12)  #generujemy 12 elementowy wektor odpowiedzi  y
x2<-rep(c(1,2),each=6)
z2<-rep(1:6,each=2)  #partia
z3<-rep(1:2,6)#próbka
dane2<-data.frame(y2,x2,z2,z3)  #tworzymy ramkę danych
dane2$x2<-factor(dane2$x2)  #nadajemy kody  dla zmiennej jakościowej Proces
levels(dane2$x2) <- c("Proc1","Proc2")
dane2$z2<-factor(dane2$z2)  #nadajemy kody  dla zmiennej jakościowej Partia
levels(dane2$z2) <- c("pa11","pa12","pa13","pa21","pa22","pa23")
dane2$z3<-factor(dane2$z3)  #nadajemy kody  dla zmiennej jakościowej
levels(dane2$z3) <- c("pr1","pr2")
model2<-lmer(y2~x2+(1|z2),data=dane2,REML=FALSE,control = lmerControl(optimizer ="Nelder_Mead"))#estymacja metodą ML
```

```
t(model2@pp$Zt) # macierz Z
image(getME(model2,"Z"))#graficzny obraz macierzy Z dla modelu 2
model2@pp$X # macierz X
t(model2@pp$Zt)%*%model2@pp$Zt
```

- Split-Plot Studies(Badania z podziałem na działki). Split-Plot, to hierarchiczny plan badań czynnikowych gdzie jeden czynnik (whole plot factor) jest przypisany do dużych jednostek eksperymentalnych (whole plots) a drugi czynnik(split-plot factor) jest przypisany do mniejszych jednostek (subplots) w obrębie dużych jednostek. Zwykle whole plot factor jest trudny do kontrolowania (np. metoda nawadniania) a drugi jest znacznie łatwiejszy do kontolowania. Przykład wprowadzający. Agronom chce sprawdzić skuteczność 3 rodzajów nawozów przy 2 metodach nawadniania w uprawie roślin. Wybrał więc (losowo) dwa miejsca (sites) w których można przeprowadzić eksperymenty. Każde miejsce dzieli na dwie duże działki (całe działki- whole plots- main units) aby zastosować na każdej z nich jedną z metod nawadniania. Następnie dzieli każdą działkę na 3 sekcje (dzielone działki-subunits) aby zastosować 3 rodzaje nawozów. Odpowiedzią jest produkcja roślina z poszczególnych działek.

  Przeprowadzono eksperyment czynnikowy 3x 2 (odpowiednio z czynnikami A i B przy użyciu sześciu głównych jednostek(sites), z których każda została podzielona na dwie podjednostki. Poziomy A zostały losowo przypisane do dwóch głównych jednostek, a poziomy B zostały losowo przypisane do podjednostek w obrębie głównych jednostek. Odpowiedni model to

  $$Y_{ijk} = \mu + \tau_i + \delta_j + \theta_{ij} + a_{ik} + \varepsilon_{ijk} \quad ; i=1\div 3\ ,\ j=1,2,\ k=1,2$$

  $$\sum_{i=1}^{3}\tau_i = 0\,, \sum_{j=1}^{2}\delta_j = 0\,, \forall j\ \sum_{i=1}^{3}\theta_{ij} = 0\,,\ \forall i\ \sum_{j=1}^{2}\theta_{ij} = 0$$

  $$a_{ik} \sim N(0,\sigma_1^2)\,;\ \varepsilon_{ijk} \sim N(0,\sigma^2)\,;\ \mathrm{cov}(a_{ik},\varepsilon_{ijk}) = 0$$

  Jeżeli obserwacje są sortowane według poziomów A, jednostek głównych w obrębie poziomów A i poziomów B w obrębie jednostek głównych, możemy wyrazić ten model w postaci

  $$\mathbf{Y} = \mathbf{X\beta} + \mathbf{Z}_1\mathbf{a}_1 + \boldsymbol{\varepsilon}\ ,\text{gdzie}$$

$$\mathbf{Y} = \begin{bmatrix} Y_{111} \\ Y_{121} \\ Y_{112} \\ Y_{122} \\ Y_{211} \\ Y_{221} \\ Y_{212} \\ Y_{222} \\ Y_{311} \\ Y_{321} \\ Y_{312} \\ Y_{322} \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \text{ i } \mathbf{Z}_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Parametryzacja R efektów stałych

$$\begin{bmatrix} Y_{111} \\ Y_{121} \\ Y_{112} \\ Y_{122} \\ Y_{211} \\ Y_{221} \\ Y_{212} \\ Y_{222} \\ Y_{311} \\ Y_{321} \\ Y_{312} \\ Y_{322} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \tau_2 \\ \tau_3 \\ \delta_2 \\ \theta_{22} \\ \theta_{32} \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} a_{11} \\ a_{12} \\ a_{21} \\ a_{22} \\ a_{31} \\ a_{32} \end{bmatrix} + \begin{bmatrix} \varepsilon_{111} \\ \varepsilon_{121} \\ \varepsilon_{112} \\ \varepsilon_{122} \\ \varepsilon_{211} \\ \varepsilon_{221} \\ \varepsilon_{212} \\ \varepsilon_{222} \\ \varepsilon_{311} \\ \varepsilon_{321} \\ \varepsilon_{312} \\ \varepsilon_{322} \end{bmatrix}$$

Struktura kowariancyjna wektora obserwacji jest taka jak w poprzednim przykładzie

$$\mathbf{\Sigma} = \sigma_1^2 \mathbf{Z}_1 \mathbf{Z}_1^T + \sigma^2 \mathbf{I}_{12} = \begin{bmatrix} \mathbf{\Sigma}_1 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}_1 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{\Sigma}_1 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{\Sigma}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{\Sigma}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{\Sigma}_1 \end{bmatrix}, \text{ gdzie } \mathbf{\Sigma}_1 = \begin{bmatrix} \sigma^2 + \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma^2 + \sigma_1^2 \end{bmatrix}$$

```
library(lme4)
#Rencher Split-Plot Studies
set.seed(123)
y3<-rnorm(12)  #generujemy 12 elementowy wektor odpowiedzi  y
x3_1<-rep(c(1:3),each=4)#Main Units
x3_2<-rep(c(1,2),6)#Subunits
#x3_3<-rep(1:6,2)
z3_1<-rep(1:6,each=2)
z3_2<-rep(1:6,2)
dane3<-data.frame(y3,x3_1,x3_2,z3_1,z3_2)  #tworzymy ramkę danych
dane3$x3_1<-factor(dane3$x3_1)  #nadajemy kody  dla zmiennej jakościowej
levels(dane3$x3_1) <- c("MU1","MU2","MU3")
dane3$x3_2<-factor(dane3$x3_2)  #nadajemy kody  dla zmiennej jakościowej Partia
levels(dane3$x3_2) <- c("SU1","SU2")
dane3$z3_1<-factor(dane3$z3_1)  #nadajemy kody  dla zmiennej jakościowej
levels(dane3$z3_1) <- c("pa11","pa12","pa13","pa21","pa22","pa23")
model3<-lmer(y3~x3_1*x3_2+(1|z3_1),data=dane3,REML=FALSE,control = lmerControl(optimizer ="Nelder_Mead"))#estymacja ML
t(model3@pp$Zt) # macierz Z
image(getME(model3,"Z"))#graficzny obraz macierzy Z dla modelu 3
model3@pp$X # macierz X
t(model3@pp$Zt)%*%model3@pp$Zt
```

- One-Way Random Effects. Jednokierunkowe efekty losowe. To plan eksperymentu używany w analizie wariancji lub analizie danych panelowych, gdzie poziomy czynnika grupującego są traktowane jako próba losowa z hipotetycznej populacji poziomów koncentrując się na estymacji komponentów wariancyjnych opisujących zmienność pochodzącą od różnych grup a nie na różnicach średnich pomiędzy tymi grupami.
  Zakład chemiczny wyprodukował dużą liczbę partii. Każda partia została zapakowana w dużą liczbę pojemników. Wybraliśmy trzy partie losowo i losowo wybraliśmy cztery pojemniki z każdej partii, aby zmierzyć y. Model jest następujący:

  $$Y_{ij} = \mu + a_i + \varepsilon_{ij} \; ; i=1 \div 3, j=1 \div 4$$

Jeżeli obserwacje są sortowane według partii i pojemników w partiach, możemy wyrazić ten model w postaci $\mathbf{Y} = \mathbf{X\beta} + \mathbf{Z}_1\mathbf{a}_1 + \boldsymbol{\varepsilon}$,gdzie

$$\mathbf{X} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \qquad \mathbf{Z}_1 = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

Struktura kowariancyjna wektora obserwacji jest następująca

$$\boldsymbol{\Sigma} = \sigma_1^2\mathbf{Z}_1\mathbf{Z}_1^T + \sigma^2\mathbf{I}_{12} = \begin{bmatrix} \boldsymbol{\Sigma}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\Sigma}_1 \end{bmatrix}, \text{ gdzie } \boldsymbol{\Sigma}_1 = \begin{bmatrix} \sigma^2+\sigma_1^2 & \sigma_1^2 & \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma^2+\sigma_1^2 & \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 & \sigma^2+\sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 & \sigma_1^2 & \sigma^2+\sigma_1^2 \end{bmatrix}$$

```
#One-way  Random Effects
y4<-rnorm(12)
z4<-rep(1:3,each=4)
dane4<-data.frame(y4,z4)
dane4$z4<-factor(dane4$z4)
levels(dane4$z4)<-c("P1","P2","P3")
model4<-lmer(y4~(1|z4),data=dane4,REML=FALSE)
model4@pp$X # macierz X
t(model4@pp$Zt)%*%model4@pp$Zt
ranef(model4,drop=TRUE)
fixef(model4)
```

- Independent Random Coefficients (Niezależne współczynniki losowe). To modele w których parametry zmieniają się losowo w różnych grupach lub jednostkach i których zmienność nie jest statystycznie powiązana ze sobą, co oznacza że np. losowe nachylenie dla jednej grupy nie jest zależne od losowego nachylenia dla innej grupy.
  W eksperymencie wykorzystano trzy szczenięta z każdego z czterech miotów myszy. Jedno szczenię z każdego miotu zostało wystawione na jeden z trzech ilościowych poziomów substancji rakotwórczej. Zależność między przyrostem masy ciała (y) a poziomem substancji rakotwórczej jest linią prostą, ale nachylenia i wyrazy wolne różnią się losowo niezależnie między miotami. Trzy poziomy substancji rakotwórczej oznaczono jako **x**. Model jest następujący

$$Y_{ij} = \beta_0 + a_i + \beta_1 x_j + b_{i1} x_j + \varepsilon_{ij} \; ; i=1\div4, j=1\div3$$

gdzie $a_i \sim N(0,\sigma_1^2)$; $b_i \sim N(0,\sigma_2^2)$ $\varepsilon_{ijk} \sim N(0,\sigma^2)$; $\text{cov}(a_{ik},\varepsilon_{ijk}) = 0$

Jeżeli dane zostaną posortowane według miotów i poziomów substancji rakotwórczych w miotach, możemy wyrazić ten model w postaci $\mathbf{Y} = \mathbf{X\beta} + \mathbf{Z}_1\mathbf{a}_1 + \mathbf{Z}_2\mathbf{a}_2 + \boldsymbol{\varepsilon}$ $(\mathbf{a} = \mathbf{a}_1, \mathbf{b} = \mathbf{a}_2)$

$$\mathbf{X}=\begin{bmatrix}\mathbf{j}_3 & \mathbf{x}\\ \mathbf{j}_3 & \mathbf{x}\\ \mathbf{j}_3 & \mathbf{x}\\ \mathbf{j}_3 & \mathbf{x}\end{bmatrix}\quad \mathbf{Z}_1=\begin{bmatrix}\mathbf{j}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3\\ \mathbf{0}_3 & \mathbf{j}_3 & \mathbf{0}_3 & \mathbf{0}_3\\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{j}_3 & \mathbf{0}_3\\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{j}_3\end{bmatrix}\mathbf{Z}_2=\begin{bmatrix}\mathbf{x} & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3\\ \mathbf{0}_3 & \mathbf{x} & \mathbf{0}_3 & \mathbf{0}_3\\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{x} & \mathbf{0}_3\\ \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{0}_3 & \mathbf{x}\end{bmatrix}$$

Struktura kowariancyjna wektora obserwacji jest następująca $\mathbf{\Sigma}=\sigma_1^2\mathbf{Z}_1\mathbf{Z}_1^T+\sigma_2^2\mathbf{Z}_2\mathbf{Z}_2^T+\sigma^2\mathbf{I}_{12}$

$$\mathbf{\Sigma}=\begin{bmatrix}\mathbf{\Sigma}_1 & \mathbf{0} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{\Sigma}_1 & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{\Sigma}_1 & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{\Sigma}_1\end{bmatrix},\ \text{gdzie}\ \mathbf{\Sigma}_1=\sigma_1^2\mathbf{J}_3+\sigma_2^2\mathbf{x}\mathbf{x}^T+\sigma^2\mathbf{I}_3\quad \text{a}\quad \mathbf{J}_3=\begin{bmatrix}1 & 1 & 1\\ 1 & 1 & 1\\ 1 & 1 & 1\end{bmatrix}.$$

```
#Independent Random Coefficients
y5<-2*c(1,2,3,1,2,3,1,2,3,1,2,3)+rnorm(12)
dose<-c(1,2,3,1,2,3,1,2,3,1,2,3)
z1<-rep(1:4,each=3)
dane5<-data.frame(y5,dose,z1)
dane5$z1<-factor(dane5$z1)
levels(dane5$z1)<-c("L1","L2","L3","L4")
model5<-lmer(y5~dose+(dose||z1),data=dane5,REML=FALSE,control = lmerControl(optimizer ="Nelder_Mead"))#estymacja metodą ML
#model5<-lmer(y5~1+dose+(1|z1)+(0+dose|z1),data=dane5,REML=FALSE) #Inaczej
t(model5@pp$Zt) # macierz Z
image(getME(model5,"Z"))#graficzny obraz macierzy Z dla modelu 3
model5@pp$X # macierz X
t(model5@pp$Zt)%*%model5@pp$Zt
ranef(model5,drop=TRUE)
fixef(model5)

#(uncorrelated random intercept and slope)
model5bis<-lmer(y5~dose+(dose|z1),data=dane5,REML=FALSE,control = lmerControl(optimizer ="Nelder_Mead"))#estymacja ML
t(model5bis@pp$Zt) # macierz Z
#model5bis<-lmer(y5~1+dose+(1|z1)+(1+dose|z1),data=dane5,REML=FALSE) #Inaczej
```

- Heterogeneous Variances (Niejednorodne wariancje). Losowo wybrano cztery osoby z każdej z czterech grup. Grupy miały różne średnie i różne wariancje. Zakładamy tutaj, że $\sigma^2=0$. Model jest następujący

$$Y_{ij}=\mu_i+\varepsilon_{ij},\ i=1\div4,\ \ j=1\div4,\ \varepsilon_{ij}\sim N(0,\sigma_i^2).$$

Jeżeli dane są sortowane według grup i jednostek w obrębie grup, możemy wyrazić ten model w postaci $\mathbf{Y}=\mathbf{X}\boldsymbol{\beta}+\mathbf{Z}_1\mathbf{a}_1+\cdots+\mathbf{Z}_4\mathbf{a}_4+\boldsymbol{\varepsilon}$ z

$$\mathbf{X}=\begin{bmatrix}\mathbf{I}_4\\ \mathbf{I}_4\\ \mathbf{I}_4\\ \mathbf{I}_4\end{bmatrix},\ \mathbf{Z}_1=\begin{bmatrix}\mathbf{I}_4\\ \mathbf{0}_4\\ \mathbf{0}_4\\ \mathbf{0}_4\end{bmatrix},\ \mathbf{Z}_2=\begin{bmatrix}\mathbf{0}_4\\ \mathbf{I}_4\\ \mathbf{0}_4\\ \mathbf{0}_4\end{bmatrix},\mathbf{Z}_3=\begin{bmatrix}\mathbf{0}_4\\ \mathbf{0}_4\\ \mathbf{I}_4\\ \mathbf{0}_4\end{bmatrix},\mathbf{Z}_4=\begin{bmatrix}\mathbf{0}_4\\ \mathbf{0}_4\\ \mathbf{0}_4\\ \mathbf{I}_4\end{bmatrix}.$$

Struktura kowariancyjna jest postaci

$$\boldsymbol{\Sigma} = \sigma_1^2 \mathbf{Z}_1 \mathbf{Z}_1^T + \sigma_2^2 \mathbf{Z}_2 \mathbf{Z}_2^T + \sigma_3^2 \mathbf{Z}_3 \mathbf{Z}_3^T + \sigma_4^2 \mathbf{Z}_4 \mathbf{Z}_4^T = \begin{bmatrix} \sigma_1^2 \mathbf{I}_4 & & & \\ & \sigma_2^2 \mathbf{I}_4 & & \\ & & \sigma_3^2 \mathbf{I}_4 & \\ & & & \sigma_4^2 \mathbf{I}_4 \end{bmatrix}$$

Jeżeli znamy macierz **D** a więc także macierz $\mathbf{I} + \mathbf{ZDZ}^{\mathrm{T}}$, to możemy estymować efekty $\boldsymbol{\beta}$ używając metody uogólnionych najmniejszych kwadratów. Jeżeli nie znamy **D**, to możemy je estymować metodą największej wiarygodności. Załóżmy, że macierz **D** jest parametryzowana *g* parametrami. Wektor tych parametrów oznaczmy symbolem $\boldsymbol{\theta}$. Formalnie powinniśmy więc pisać $\mathbf{D}(\boldsymbol{\theta})$. Dla zwiększenia czytelności rezygnujemy z dopisywania parametru. Jeżeli w klasycznym modelu liniowym $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$ wektor parametrów $\boldsymbol{\beta}$ jest liczny (duże *p*) , to aby uniknąć maksymalizacji funkcji wiarygodności po dużej przestrzeni parametrów, rozważa się tzw. *resztowe estymatory największej wiarygodności* (*residual/restricted maximum likelihood* -REML). W pierwszym kroku redukuje się wektor obserwacji $\mathbf{Y}$ do przestrzeni ortogonalnej do kolumn macierzy $\mathbf{X}$ rzutując ortogonalnie $\mathbf{Y}$ na $(\mathrm{Im}(\mathbf{X}))^{\perp}$ a następnie szuka się oceny parametru σ^2 na zredukowanej podprzestrzeni. W ten sposób w procesie estymacji σ^2 pozbywamy się parametrów zakłócających $\boldsymbol{\beta}$. Tak uzyskany estymator dla rozważanego modelu liniowego wyraża się wzorem

$$\hat{\sigma}^2_{REML} = \frac{RSS}{n-p},$$

a więc jest estymatorem nieobciążonym w odróżnieniu od estymatora $\hat{\sigma}^2_{ML} = \frac{RSS}{n}$, który jest tylko asymptotycznie nieobciążony.

Dla modeli mieszanych obie metody estymacji mogą dawać estymatory obciążone, przy czym zazwyczaj metoda REML daje estymatory o mniejszym obciążeniu.

Estymacja parametrów w modelach mieszanych nie jest zagadnieniem prostym i od wielu lat proponuje się nowe metody estymacji. Z grubsza można je podzielić na dwie klasy. W jednej grupie metod wykorzystuje się algebraiczne własności macierzy $\mathbf{ZDZ}^{\mathrm{T}}$ a w drugiej rozwija metody obliczeniowe i optymalizacyjne do numerycznej maksymalizacji funkcji wiarygodności. Wielu polskich statystyków zajmuje się metodami estymacji parametrów w modelach mieszanych szczególnie z pierwszej z wymienionych grup (Gnot S. 1994, 2004, Michalski A. 1996, Caliński T. 2004, Grzędziel M. 2008). Wprowadzeniem do tej grupy problemów jest monografia:

Gnot S. (1992) Estymacja komponentów wariancyjnych w modelach liniowych. WNT. Warszawa.

Metody zaproponowane przez polskich statystyków są bardzo efektywne w sytuacji, gdy możemy kontrolować macierz $\mathbf{Z}$ i mamy możliwość zaplanowania macierzy modelu. W oprogramowaniu ogólnego przeznaczenia popularność zdobyły metody obliczeniowe, w których zarówno macierze $\mathbf{Z}$ i $\mathbf{D}$ mogą być praktycznie dowolnej postaci. Dominują 2 podejścia

- metoda estymacji z użyciem algorytmu Newtona-Rapshona zaimplementowana w programie SAS w procedurze MIXED

- metoda estymacji z wykorzystaniem operacji na macierzach rzadkich (w R pakiety nlme i lme4)

Pakiet mnle jest omówiony w monografii Pinheiro K, Bates D (2009) Mixed –Effects Models in S and S-PLUS. Springer

Nowszą wersją pakietu nelm jest pakiet lme4 (Bates D, Machler M)

https://lme4.r-forge.r-project.org/book/

Bates D, Machler M. Fitting Linear Mixed-Effects Models Using lme4 (plik lme4 Bates.pdf)

Metoda estymacji z użyciem algorytmu Newtona-Rapshona Woofinger R., Tobias R., Sall J., Computing Gaussian Likelihoods and Their Derivatives for General Linear Mixed Models – do wykonania jest następujące zadanie. Należy znaleźć punkt maksymalizujący funkcję wiarygodności (lub resztowej wiarygodności) w przestrzeni $p+q+1$ parametrów $(\boldsymbol{\beta},\boldsymbol{\theta},\sigma^2)$. Wykorzystując procedurę profilowania opiszmy funkcję wiarygodności jako funkcję jedynie parametru $\boldsymbol{\theta}$, użyjemy metody Newtona –Rapshona do znalezienia oceny parametru $\boldsymbol{\theta}$ a następnie użyjemy tej oceny do wyznaczenia ocen $(\boldsymbol{\beta},\sigma^2)$. Metoda optymalizacji Newtona-Rapshona to metoda iteracyjna: w każdym kroku jest wyznaczana lokalna aproksymacja funkcji wiarygodności funkcją kwadratową w celu znalezienia punktu maksymalizującego funkcję wiarygodności. Iteracyjnie wyznacza się nowe oceny

$$\boldsymbol{\theta}^{(i+1)} = \boldsymbol{\theta}^{(i)} - \mathbf{H}^{-1}_{\boldsymbol{\theta}^{(i)}} \mathbf{g}_{\boldsymbol{\theta}^{(i)}},$$

gdzie $\mathbf{g}_{\boldsymbol{\theta}^{(i)}}$ to gradient logarytmu funkcji wiarygodności w punkcie $\boldsymbol{\theta}^{(i)}$ a $\mathbf{H}_{\boldsymbol{\theta}^{(i)}}$ - to hessian (macierz drugiej różniczki) logarytmu funkcji wiarygodności w punkcie $\boldsymbol{\theta}^{(i)}$. Do wad metody Newtona – Rapshona należy zaliczyć brak gwarancji zbieżności nawet do lokalnego maksimum. W pakietach statystycznych stosuje się modyfikacje metody Newtona –Rapshona (np. grzbietowa metoda Newtona –Rapshona z krótszym krokiem , lub tzw. *Fisher scoring* w którym macierz $\mathbf{H}_{\boldsymbol{\theta}^{(i)}}$ zastępuje się macierzą informacji Fishera (wartością oczekiwaną macierzy $\mathbf{H}_{\boldsymbol{\theta}^{(i)}}$)

Rozważmy bliżej zagadnienie maksymalizacji funkcji wiarygodności wyrażonej wzorem

$$L(\boldsymbol{\beta},\boldsymbol{\theta},\sigma^2 \mid \mathbf{Y}) = (2\pi)^{-\frac{n}{2}} \mid \mathbf{V} \mid^{-\frac{1}{2}} \exp(-\tfrac{1}{2}(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})^{\mathrm{T}}\mathbf{V}^{-1}(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta}))$$

gdzie $\mathbf{V}(\sigma^2,\boldsymbol{\theta}) = \sigma^2(\mathbf{I}+\mathbf{Z}^T\mathbf{D}(\boldsymbol{\theta})\mathbf{Z})$ jest macierz kowariancyjną wektora $\mathbf{Y}$ parametryzowaną przez σ^2 i $\boldsymbol{\theta}$ (wektor $\boldsymbol{\theta}$ określa macierz $\mathbf{D}$). Równoważnie możemy minimalizować tzw. dewiancję (*deviance*) czyli minus podwojony logarytm funkcji wiarygodności

$$-2l(\boldsymbol{\beta},\boldsymbol{\theta},\sigma^2 \mid \mathbf{Y}) = n\log(2\pi) + \log \mid \mathbf{V} \mid + (\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})^{\mathrm{T}}\mathbf{V}^{-1}(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta}), \quad (*)$$

Dla ustalonych parametrów $\boldsymbol{\theta}$ estymator największej wiarygodności dla $\boldsymbol{\beta}$ można wyznaczyć metodą uogólnionych ważonych kwadratów jako $\mathbf{b}(\boldsymbol{\theta}) = \left(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{Y}$.

Wstawiając tą wartość do wzoru (*) możemy przedstawić minimalizowaną funkcję jako funkcję wyłącznie parametrów σ^2 i $\boldsymbol{\theta}$. Taki zabieg redukcji parametrów w funkcji wiarygodności nazywamy profilowaniem. Dzięki niemu zmniejszyliśmy wymiar przestrzeni parametrów z $p+q+1$ do $q+1$. Funkcja celu z równania (*) jest teraz równa

$$-2l(\boldsymbol{\theta},\sigma^2 \mid \mathbf{Y}) = n\log(2\pi) + \log|\mathbf{V}| + (\mathbf{Y}-\mathbf{Xb}(\boldsymbol{\theta}))^{\mathrm{T}}\mathbf{V}^{-1}(\mathbf{Y}-\mathbf{Xb}(\boldsymbol{\theta})) \quad (**)$$

W podobny sposób możemy wyznaczyć funkcję resztowej wiarygodności REML

$$-2l_R(\boldsymbol{\theta},\sigma^2 \mid \mathbf{Y}) = (n-p)\log(2\pi) + \log|\mathbf{V}| + (\mathbf{Y}-\mathbf{Xb}(\boldsymbol{\theta}))^{\mathrm{T}}\mathbf{V}^{-1}(\mathbf{Y}-\mathbf{Xb}(\boldsymbol{\theta})) + \log|\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X}| \quad (***)$$

W rozważanym przypadku jeszcze zmniejszyć wymiar przestrzeni parametrów profilując σ^2. Możemy maksymalizować funkcję wiarygodności oddzielnie po σ^2 i oddzielnie po $\boldsymbol{\theta}$. Z macierzy $\mathbf{V} = \sigma^2(\mathbf{I} + \mathbf{Z}^T\mathbf{DZ})$ wyciągamy σ^2, a to co zostanie oznaczmy $\mathbf{V}^*(\boldsymbol{\theta}^*) = \mathbf{I} + \mathbf{Z}^T\mathbf{D}(\boldsymbol{\theta}^*)\mathbf{Z}$. Dla ustalonego $\boldsymbol{\theta}^*$ maksymalizacja po σ^2 prowadzi do ocen

$$\hat{\sigma}^2 = (\mathbf{Y}-\mathbf{Xb})^{\mathrm{T}}(\mathbf{V}^*)^{-1}(\mathbf{Y}-\mathbf{Xb})/n,$$

$$\hat{\sigma}_R^2 = (\mathbf{Y}-\mathbf{Xb})^{\mathrm{T}}(\mathbf{V}^*)^{-1}(\mathbf{Y}-\mathbf{Xb})/(n-p).$$

Wstawiając te oceny do równań (**) i (***) odpowiednio otrzymujemy

$$-2l(\boldsymbol{\theta}^*,\mid \mathbf{Y}) = \log|\mathbf{V}^*| + n\log[(\mathbf{Y}-\mathbf{Xb})^{\mathrm{T}}(\mathbf{V}^*)^{-1}(\mathbf{Y}-\mathbf{Xb})] + n + n\log(2\pi/n).$$

W podobny sposób możemy wyznaczyć funkcję resztowej wiarygodności REML

$$-2l_R(\boldsymbol{\theta},\mid \mathbf{Y}) = n\log(\sigma^2) + \log|\mathbf{V}^*| + n\log[(\mathbf{Y}-\mathbf{Xb})^{\mathrm{T}}(\mathbf{V}^*)^{-1}(\mathbf{Y}-\mathbf{Xb})] + (n-p) + \\ + (n-p)\log(2\pi/(n-p)) + \log|\mathbf{X}^T(\mathbf{V}^*)^{-1}\mathbf{X}|$$

Zauważmy , że $\mathbf{V}$ jest rozmiaru n x n a więc dla dużych n (rzędu milionów czy miliardów obserwacji) nie możemy takiej macierzy zmieścić w pamięci komputera. Okazuje się jednak, że każdy z członów opisujących funkcje $-2l(\boldsymbol{\theta},\mid \mathbf{Y})$ i $-2l_R(\boldsymbol{\theta},\mid \mathbf{Y})$ można wyznaczyć opierając się na macierzy $\mathbf{S} = [\mathbf{Y};\mathbf{X};\mathbf{Z}]^T[\mathbf{Y};\mathbf{X};\mathbf{Z}]$, która jest rozmiaru $(p+q+1)\times(p+q+1)$. Dokładny opis jak to zrobić jest w pracy

Wolfinger R., Tobias R.,Sall J.,(1994) Computing Gaussian Likelihoods and their Derivatives for General Linear Mixed Models. SIAM Journal of Scientific Computing, Vol.15.No.6, pp. 1294-1310.

Jako argument za opisaniem pewnego efektu jako efekt losowy podaliśmy brak zainteresowania oceną wartości tego efektu. Są jednak sytuacje, gdzie pomimo, że efekt będzie traktowany jako losowy będziemy zainteresowani jego oceną. Efekt losowy w ujęciu klasycznym jest inaczej traktowany niż efekt stały lub parametry modelu opisujące wariancję (komponenty wariancyjne). Efekt stały i parametry opisujące wariancję są pewnymi nieznanymi wartościami, a więc ma sens ich ocena czyli estymacja. Efekt losowy jest opisywany przez zmienną losową i nie powinniśmy używać określenia estymacja do jego oceny. Będziemy więc wyznaczać medianę lub średnią (dla rozkładu normalnego to

to samo) rozkładu a posteriori $p(\mathbf{u} \mid \mathbf{Y}, \mathbf{X})$ i te wartości będziemy nazywali predykcjami efektów losowych. Samo sformułowanie „predykcje" pojawiło się dopiero kilkanaście lat temu po tym jak Henderson podał wzory do wyznaczania tych predykcji. W celu odróżnienia predykcji od ocen będziemy jako $\hat{\boldsymbol{\beta}}$ oznaczać ocenę efektów stałych a $\tilde{\mathbf{u}}$ predykcje efektów losowych. Szczegóły wyprowadzenie równań Hendersona w duchu bayesowskim można znaleźć w pracy :

Robinson G.(1999) That BLUP is a good thing:the estimation of random effects. Statistical Science, Vol 6,No.1 pp.15-32.

Dla modelu $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon}$ równania Hendersona można zapisać w postaci:

$$\begin{bmatrix} \mathbf{X}^T\mathbf{X} & \mathbf{X}^T\mathbf{Z} \\ \mathbf{Z}^T\mathbf{X} & \mathbf{Z}^T\mathbf{Z} + \hat{\mathbf{D}}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \tilde{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^T\mathbf{Y} \\ \mathbf{Z}^T\mathbf{Y} \end{bmatrix}.$$

Przy założeniu, że macierz $\mathbf{D}$ jest znana (lub wyestymowana) z powyższych równań można wyznaczyć $\hat{\boldsymbol{\beta}}$ i $\tilde{\mathbf{u}}$. Co więcej można pokazać, że otrzymana ocena $\hat{\boldsymbol{\beta}}$ jest w sensie błędu średniokwadratowego najlepszą w klasie liniowych ocen nieobciążonych BLUE(*best linear unbiased estimation*) a $\tilde{\mathbf{u}}$ jest najlepszą liniową predykcją nieobciążoną BLUP (*best linear unbiased prediction*). W pakiecie R do wyznaczenia predykcji efektów losowych służy funkcja ranef(){lme4}

Uwagi do metody REML estymacji komponentów wariancyjnych.

W modelu $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon}$ wektor ma rozkład $\mathbf{Y} \sim N(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}))$ gdzie $\boldsymbol{\Sigma} = \sigma^2(\mathbf{I} + \mathbf{Z}\mathbf{D}\mathbf{Z}^{\mathrm{T}})$.

Zazwyczaj $\boldsymbol{\Sigma} = \sigma^2\mathbf{I} + \sum_{i=1}^{m} \sigma_i^2 \mathbf{Z}_i\mathbf{Z}_i^{\mathrm{T}} = \sum_{i=0}^{m} \sigma_i^2 \mathbf{Z}_i\mathbf{Z}_i^{\mathrm{T}}$ po przyjęciu $\mathbf{Z}_0 = \mathbf{I}_n$. Idea metody REML (Patterson ,Thompson 1971) polega na estymacji komponentów wariancyjnych 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 wariancyjnych 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 wariancyjnych, 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 $(n-r)\times n$. Ponadto $\mathbf{K}$ jest postaci $\mathbf{K} = \mathbf{C}(\mathbf{I} - \mathbf{H}) = \mathbf{C}(\mathbf{I} - \mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-}\mathbf{X})$ 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 $\mathbf{k}_i^{\mathrm{T}}$ macierzy muszą spełniać równania $\mathbf{k}_i^{\mathrm{T}}\mathbf{X} = \mathbf{0}^{\mathrm{T}}$ lub równoważnie $\mathbf{X}^{\mathrm{T}}\mathbf{k}_i = \mathbf{0}$. 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 $\mathbf{k}_i = (\mathbf{I} - (\mathbf{X}^{\mathrm{T}})^{-}\mathbf{X}^{\mathrm{T}})\mathbf{c}$ dla dowolnego wektora $\mathbf{c}$ wymiaru $n \times 1$.

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

Ale $\operatorname{rank}(\mathbf{X}\mathbf{X}^{-}) = \operatorname{rank}(\mathbf{X}) = r$. Ponadto $\mathbf{I} - \mathbf{X}\mathbf{X}^{-}$ jest idempotentna, więc $\operatorname{rank}(\mathbf{I} - \mathbf{X}\mathbf{X}^{-}) = \operatorname{tr}(\mathbf{I} - \mathbf{X}\mathbf{X}^{-}) = \operatorname{tr}(\mathbf{I}) - \operatorname{tr}(\mathbf{X}\mathbf{X}^{-}) = n - r$. Stąd wynika, że istnieje $n - r$ liniowo niezależnych wektorów $\mathbf{k}_i$, spełniających $\mathbf{X}^{\mathrm{T}}\mathbf{k}_i = \mathbf{0}$ i dlatego maksymalna liczba wierszy macierzy $\mathbf{K}$ jest równa $n - r$. Wobec powyższego $\mathbf{K} = \mathbf{C}(\mathbf{I} - \mathbf{X}\mathbf{X}^{-})$ 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 $\mathbf{X}^{-} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-}\mathbf{X}^{\mathrm{T}}$ (zob. uzupełnienie macierzy) możemy przyjąć $\mathbf{K} = \mathbf{C}(\mathbf{I} - \mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-}\mathbf{X})$.

Dla modelu $\mathbf{Y} \sim N(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma})$ gdzie $\boldsymbol{\Sigma} = \sigma^2\mathbf{I} + \sum_{i=1}^{m}\sigma_i^2\mathbf{Z}_i\mathbf{Z}_i^{\mathrm{T}} = \sum_{i=0}^{m}\sigma_i^2\mathbf{Z}_i\mathbf{Z}_i^{\mathrm{T}}$ i powyżej rozważanej macierzy $\mathbf{K}$ mamy $\mathbf{KY} \sim N_{n-r}(\mathbf{0}, \mathbf{K}\boldsymbol{\Sigma}\mathbf{K}^{\mathrm{T}})$.

Twierdzenie. Układ m+1 równań dla estymacji komponentów $\sigma_0^2, \sigma_1^2, \ldots, \sigma_m^2$ jest dany przez

$$\operatorname{tr}[\mathbf{K}^{\mathrm{T}}(\mathbf{K}\boldsymbol{\Sigma}\mathbf{K}^{\mathrm{T}})^{-1}\mathbf{K}\mathbf{Z}_i\mathbf{Z}_i^{\mathrm{T}}] = \mathbf{Y}^{\mathrm{T}}\mathbf{K}^{\mathrm{T}}(\mathbf{K}\boldsymbol{\Sigma}\mathbf{K}^{\mathrm{T}})^{-1}\mathbf{K}\mathbf{Z}_i\mathbf{Z}_i^{\mathrm{T}}(\mathbf{K}\boldsymbol{\Sigma}\mathbf{K}^{\mathrm{T}})^{-1}\mathbf{K}\mathbf{Y}\ ,\ i=0,\ldots,m.$$

Dowód.

$$\ln L(\sigma_0^2, \sigma_1^2, \ldots, \sigma_m^2) = -\frac{n-r}{2}\ln(2\pi) - \frac{1}{2}\ln|\mathbf{K}\boldsymbol{\Sigma}\mathbf{K}^{\mathrm{T}}| - \frac{1}{2}\mathbf{Y}^{\mathrm{T}}\mathbf{K}^{\mathrm{T}}(\mathbf{K}\boldsymbol{\Sigma}\mathbf{K}^{\mathrm{T}})^{-1}\mathbf{K}\mathbf{Y}$$

$$= -\frac{n-r}{2}\ln(2\pi) - \frac{1}{2}\ln\left|\mathbf{K}\left(\sum_{i=0}^{m}\sigma_i^2\mathbf{Z}_i\mathbf{Z}_i^{\mathrm{T}}\right)\mathbf{K}^{\mathrm{T}}\right| - \frac{1}{2}\mathbf{Y}^{\mathrm{T}}\mathbf{K}^{\mathrm{T}}\left[\mathbf{K}\left(\sum_{i=0}^{m}\sigma_i^2\mathbf{Z}_i\mathbf{Z}_i^{\mathrm{T}}\right)\mathbf{K}^{T}\right]^{-1}\mathbf{K}\mathbf{Y}$$

Stąd korzystając ze wzorów $\frac{\partial \mathbf{A}^{-1}}{\partial x} = -\mathbf{A}^{-1}\frac{\partial \mathbf{A}}{\partial x}\mathbf{A}^{-1}$; $\frac{\partial \log|\mathbf{A}|}{\partial x} = \operatorname{tr}\left(\mathbf{A}^{-1}\frac{\partial \mathbf{A}}{\partial x}\right)$ zob.(Uzupełnienia z macierzy) i pamiętając , że $\boldsymbol{\Sigma} = \sigma^2\mathbf{I} + \sum_{i=1}^{m}\sigma_i^2\mathbf{Z}_i\mathbf{Z}_i^{\mathrm{T}} = \sum_{i=0}^{m}\sigma_i^2\mathbf{Z}_i\mathbf{Z}_i^{\mathrm{T}}$ otrzymujemy

$$\frac{\partial}{\partial \sigma_i^2}\ln L(\sigma_0^2,\sigma_1^2,\ldots,\sigma_m^2) = -\frac{1}{2}\mathrm{tr}\left((\mathbf{K}\Sigma\mathbf{K}^{\mathrm{T}})^{-1}\left[\frac{\partial}{\partial \sigma_i^2}(\mathbf{K}\Sigma\mathbf{K}^{\mathrm{T}})\right]\right)$$

$$+\frac{1}{2}\mathbf{Y}^{\mathrm{T}}\mathbf{K}^{\mathrm{T}}(\mathbf{K}\Sigma\mathbf{K}^{\mathrm{T}})^{-1}\left[\frac{\partial}{\partial \sigma_i^2}(\mathbf{K}\Sigma\mathbf{K}^{\mathrm{T}})\right](\mathbf{K}\Sigma\mathbf{K}^{\mathrm{T}})^{-1}\mathbf{K}\mathbf{Y}$$

$$= -\frac{1}{2}\mathrm{tr}\left((\mathbf{K}\Sigma\mathbf{K}^{\mathrm{T}})^{-1}\mathbf{K}\mathbf{Z}_i\mathbf{Z}_i^{\mathrm{T}}\mathbf{K}^{\mathrm{T}}\right)$$

$$+\frac{1}{2}\mathbf{Y}^{\mathrm{T}}\mathbf{K}^{\mathrm{T}}(\mathbf{K}\Sigma\mathbf{K}^{\mathrm{T}})^{-1}\left[\mathbf{K}\mathbf{Z}_i\mathbf{Z}_i^{\mathrm{T}}\mathbf{K}^{\mathrm{T}}\right](\mathbf{K}\Sigma\mathbf{K}^{\mathrm{T}})^{-1}\mathbf{K}\mathbf{Y} =$$

$$= -\frac{1}{2}\mathrm{tr}\left(\mathbf{K}^{\mathrm{T}}(\mathbf{K}\Sigma\mathbf{K}^{\mathrm{T}})^{-1}\mathbf{K}\mathbf{Z}_i\mathbf{Z}_i^{\mathrm{T}}\right)$$

$$+\frac{1}{2}\mathbf{Y}^{\mathrm{T}}\mathbf{K}^{\mathrm{T}}(\mathbf{K}\Sigma\mathbf{K}^{\mathrm{T}})^{-1}\left[\mathbf{K}\mathbf{Z}_i\mathbf{Z}_i^{\mathrm{T}}\mathbf{K}^{\mathrm{T}}\right](\mathbf{K}\Sigma\mathbf{K}^{\mathrm{T}})^{-1}\mathbf{K}\mathbf{Y}$$

Przyrównując powyższe wyrażenia do 0 otrzymujemy zapowiadany układ.

Uwagi o testach dla efektów stałych i losowych

Testy dla efektów stałych

Testujemy hipotezę $H_0 : \beta_i=0$ wobec alternatywy $H_1 : \beta_i\neq 0$.

Prezentowane niżej testy łatwo rozszerzyć (poza pierwszym) na przypadek testowania hipotezy o zbiorze J efektów β_j.

- Jeżeli liczba obserwacji jest znacznie większa od liczby parametrów $n>>p$ to często wykorzystywany jest **test Walda**. Statystyką testową jest iloraz oceny parametru i błędu standardowego tej oceny

$$T = \frac{\hat{\beta}_i}{se(\hat{\beta}_i)}.$$

  Asymptotycznie statystyka ta ma rozkład normalny. Dla mniejszych n jej rozkład jest przybliżany rozkładem t-Studenta. W modelach mieszanych w przeciwieństwie do modeli liniowych bez efektów losowych rozkład t-Studenta nie jest dokładnym rozkładem statystyki testowej jeżeli nie znamy macierzy $\mathbf{D}$. Potrzebny błąd standardowy $se(\hat{\beta}_i)$ odczytujemy z macierzy

$$\mathbf{V}(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}},\mathbf{u}-\tilde{\mathbf{u}}) = \begin{bmatrix}\mathbf{X}^T\mathbf{X} & \mathbf{X}^T\mathbf{Z}\\ \mathbf{Z}^T\mathbf{X} & \mathbf{Z}^T\mathbf{Z}+\hat{\mathbf{D}}^{-1}\end{bmatrix}^{-1}.$$

  W aktualnej wersji pakietu lme4 statystyka T jest oznaczana jako *t.value* i (świadomie) nie jest podawana p-wartość.

- Jeżeli *n* nie jest bardzo duże (ale i nie jest też bardzo małe) stosujemy **test ilorazu wiarygodności**. Statystyka testowa to różnica logarytmów dwóch funkcji wiarygodności z danym efektem i bez danego efektu. Różnica ta przy prawdziwości hipotezy zerowej ma rozkład χ^2 z liczba stopni swobody odpowiadającą różnicy w licznie parametrów pomiędzy badanymi modelami.Test ilorazu wiarygodności możemy stosować wyłącznie jeżeli jeden model jest zagnieżdżony w drugim. Stwierdzeni to ma następujące konsekwencje. Stosując ML do estymacji modelu z efektem i bez mamy spełniony warunek zagnieżdżenia. Natomiast stosując metodę REML estymacja odbywa się w przestrzeni ortogonalnej do przestrzeni rozpinanej przez efekty stałe. Jeżeli porównywane modele mają różne efekty stałe to estymacja odbywa się w różnych przestrzeniach i warunek zagnieżdżenia nie jest spełniony. W konsekwencji test ilorazu wiarygodności nie powinien być stosowany przy testowaniu istotności efektów stałych metodą estymacji REML.
- Do określenia istotności współczynnika β_i mogą być też wykorzystane kryteria wyboru modelu takie jak **AIC** lub **BIC**. W tym przypadku nie wykonujemy testu na istotność ale sprawdzamy, czy dodanie danej zmiennej do modelu poprawia wybrane kryteriom oceny jakości modelu.. Uwaga. Nie wszystkie własności kryteriów AIC oraz BIC przenoszą się z modeli efektów stałych na modele mieszane.
- Do weryfikowania istotności efektu β_i można stosować test permutacyjny. Jako statystykę testową można użyć funkcji wiarygodności. Wykonując permutacje na kolumnie macierzy X odpowiadającej efektowi β_i można wyznaczyć rozkład statystyki testowej. Minusem jest czas obliczeń.

Testy dla komponentów wariancyjnych

Testujemy hipotezę $H_0: \sigma_j^2 = 0$ wobec alternatywy $H_1: \sigma_j^2 > 0$.

Asymptotyczny rozkład statystyki testowej ilorazu wiarygodności bada się na przestrzeni parametrów, która jest zbiorem otwartym (kłopoty z różniczkowaniem na zbiorach nieotwartych). Jeżeli zbiór parametrów jest otwarty, to każdy konkretny parametr jest punktem wewnętrznym zbioru i nie leży na jego brzegu. W przypadku testowania hipotezy $H_0: \sigma_j^2 = 0$ przestrzenią parametrów jest półprosta nieujemna a interesujący nas parametr, przy prawdziwości H_0, leży na jej brzegu. W takiej sytuacji rozkład asymptotyczny statystyki testowej nie jest rozkładem χ_1^2 ale jest mieszanką 1:1 rozkładów χ_0^2 i χ_1^2 (rozkład $\chi_0^2 = \delta_0$). Dlatego w literaturze sugeruje się aby *p*-wartość wyznaczoną z rozkładu χ_1^2 podzielić przez 2.

- Model z jednym komponentem wariancyjnym, jedna zmienna grupująca

Rozważmy przypadek, w którym mamy do czynienia z jedną zmienną grupującą. Uwzględnimy ją w modelu jako komponent wariancyjny, a poziomy zmiennej grupującej jako efekty losowe. Rozważmy szczególny przypadek modelu mieszanego $\mathbf{Y} = \mathbf{X\beta} + \mathbf{Zu} + \boldsymbol{\varepsilon}$ tzn. $\mathbf{Y} = m\mathbf{1} + \mathbf{Zu} + \boldsymbol{\varepsilon}$,

gdzie $\mathbf{u} = \begin{bmatrix} u_1 \\ \vdots \\ u_q \end{bmatrix} \sim N(\mathbf{0}, \sigma_1^2 \mathbf{I}_{q\times q})$ czyli $\mathbf{D} = \sigma_1^2 \mathbf{I}_{q\times q}$ i $\boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix} \sim N(\mathbf{0}, \sigma_0^2 \mathbf{I}_{n\times n})$. Macierz $\mathbf{Z}$ jest

macierzą eksperymentu kodującą poziomy (q-poziomów) zmiennej objaśniającej z użyciem zmiennych pustych. Dla rozważanego modelu $\mathbf{Y} = m\mathbf{1} + \mathbf{Zu} + \boldsymbol{\varepsilon}$ mamy

$V(\mathbf{Y}) = \sigma_0^2 \mathbf{I} + \sigma_1^2 \mathbf{Z}\mathbf{Z}^T$.

Przyjmijmy że mamy siedem obserwacji. Mamy też zmienną jakościową z, która przyjmuje wartości na trzech poziomach (A,A,A,B,B,C,C). Macierz $\mathbf{Z}$ jest więc postaci

$$\mathbf{Z} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}.$$

Wektor efektów losowych $\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} \sim N(\mathbf{0}, \sigma_1^2 \mathbf{I}_{3\times 3})$, $\mathbf{D} = \sigma_1^2 \mathbf{I}_{3\times 3}$

$$V(\mathbf{Y}) = \begin{bmatrix} \sigma_0^2+\sigma_1^2 & \sigma_1^2 & \sigma_1^2 & 0 & 0 & 0 & 0 \\ \sigma_1^2 & \sigma_0^2+\sigma_1^2 & \sigma_1^2 & 0 & 0 & 0 & 0 \\ \sigma_1^2 & \sigma_1^2 & \sigma_0^2+\sigma_1^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_0^2+\sigma_1^2 & \sigma_1^2 & 0 & 0 \\ 0 & 0 & 0 & \sigma_1^2 & \sigma_0^2+\sigma_1^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_0^2+\sigma_1^2 & \sigma_1^2 \\ 0 & 0 & 0 & 0 & 0 & \sigma_1^2 & \sigma_0^2+\sigma_1^2 \end{bmatrix}.$$

W modelu efektów losowych obserwacje z tej samej klasy są skorelowane

$$\rho(Y_{ij}, Y_{ik}) = \frac{Cov(Y_{ij}, Y_{ik})}{\sqrt{V(Y_{ij})\, V(Y_{ik})}} = \frac{\sigma_1^2}{\sigma_0^2 + \sigma_1^2}$$

Współczynnik korelacji pomiędzy zmiennymi z tej samej klasy (grupy) nosi nazwę współczynnika korelacji wewnątrzklasowej. Wektor $\mathbf{Y}$ ma rozkład $N(m\mathbf{1}, V(\mathbf{Y}))$.

```
#Plik Mixed.R
y<-c(10,10,10,11,11,12,12)+rnorm(7)  #generujemy 7 elementowy wektor odpowiedzi  y
z=c(1,1,1,2,2,3,3)  #zmienna jakościowa o 3 poziomach
dane<-data.frame(y,z)  #tworzymy ramkę danych
dane$z<-factor(dane$z)  #nadajemy kody  dla zmiennej jakościowej
levels(dane$z) <- c("A","B","C")
dane
```

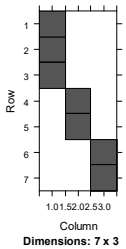
	y	z
1	10.155414	A
2	7.627135	A
3	7.258444	A
4	11.924624	B
5	13.228422	B
6	13.158085	C
7	13.302090	C

```
summary(dane)
```

	y	z
Min. :	7.258	A:3
1st Qu.:	8.891	B:2
Median :	11.925	C:2
Mean :	10.951	
3rd Qu.:	13.193	
Max. :	13.302	

Rozważamy model 0 $\mathbf{Y} = m\mathbf{1} + \mathbf{Zu} + \boldsymbol{\varepsilon}$

```
library(lme4)
model0<-lmer(y~(1|z),data=dane,REML=FALSE)#estymacja metodą ML
model0@pp$Zt # transponowana macierz Z odczytana ze slotu model0@pp obiekty klasy S4
t(model0@pp$Zt) # macierz Z
7 x 3 sparse Matrix of class "dgCMatrix"
  A B C
1 1 . .
2 1 . .
3 1 . .
4 . 1 .
5 . 1 .
6 . . 1
7 . . 1
image(getME(model0,"Z"))#graficzny obraz macierzy Z dla modelu 0
```




```
model0@pp$Xwts # transponowana macierz X
summary(model0)
```

Linear mixed model fit by maximum likelihood ['lmerMod']
Formula: y ~ (1 | z)

```
  Data: dane
    AIC      BIC   logLik deviance df.resid
   34.5     34.4    -14.3     28.5        4
Scaled residuals:
    Min      1Q  Median      3Q     Max
-1.1622 -0.6231  0.1716  0.4932  1.2505

Random effects:
 Groups   Name        Variance Std.Dev.
 z        (Intercept) 4.195    2.048
 Residual             1.442    1.201
Number of obs: 7, groups:  z, 3

Fixed effects:
            Estimate Std. Error t value
(Intercept)   11.333      1.269   8.929
```

Interpretacja wyników estymacji metodą ML

ocena wariancji efektów losowych: $\hat{\sigma}_1^2 = 4.195$

ocena wariancji zakłócenia losowego: $\hat{\sigma}_0^2 = 1.442$

ocena efektu stałego średniej: $\hat{m} = 11.333$

wartość statystyki testu Walda: t=8.929

```
model1<-lmer(y~(1|z),data=dane,REML=TRUE)#estymacja metodą REML
model1@pp$Zt # transponowana macierz Z
t(model1@pp$Zt) # macierz Z
model1@pp$Xwts# transponowana macierz X
summary(model1)
```

```
Linear mixed model fit by REML ['lmerMod']
Formula: y ~ (1 | z)
   Data: dane
REML criterion at convergence: 26

Scaled residuals:
    Min      1Q  Median      3Q     Max
-1.07553 -0.60448  0.09612  0.42932  1.32973

Random effects:
 Groups   Name        Variance Std.Dev.
 z        (Intercept) 6.537    2.557
 Residual             1.451    1.204
Number of obs: 7, groups:  z, 3

Fixed effects:
            Estimate Std. Error t value
(Intercept)   11.350      1.547   7.337
```

Interpretacja wyników estymacji metodą REML

ocena wariancji efektów losowych: $\hat{\sigma}_1^2 = 6.537$

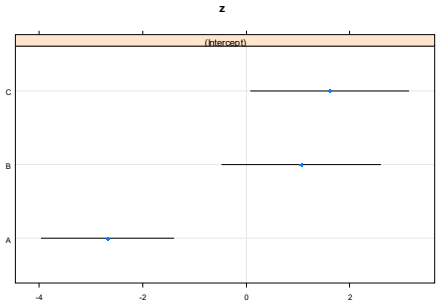
ocena wariancji zakłócenia losowego: $\hat{\sigma}_0^2 = 1.451$

ocena efektu stałego średniej: $\hat{m} = 11.350$

wartość statystyki testu Walda: t=7.337

```
#wyznaczymy predykcję efektów losowych
#argument condVar=TRUE dodaje do wyniku informacje o odchyleniach standardowych
library(lattice)
u=ranef(model0,condVar=TRUE)
```

```
dotplot(u)
```




```
ocenaML<-unlist(ranef(model0))+fixef(model0)# ML
#funkcja unlist() wybiera z listy wektor liczb który można dodać do innego wektora
ocenaML
    z.(Intercept)1          z.(Intercept)2 z.(Intercept)3
          8.653908      12.394254      12.951992
dotplot(ocenaML)
```

Zadanie. Narysować wykres dotplot(ocenaML) z błędami standardowymi

Przypuśćmy, że eksperymentator chce sprawdzić rzetelność pomiaru określonej cechy psychologicznej za pomocą pewnej standaryzowanej skali postaw. W tym celu każdej z pięciu wylosowanych z określonej populacji osób daje do rozwiązania trzy równolegle wersje tej skali, przy czym kolejność w jakiej te wersje występują w badaniu została zrandomizowana dla każdej z osób oddzielnie.

Tabela1. Wyniki fikcyjnego eksperymentu (Brzeziński 297)

Osoby	Wersje skali postaw		
	A1	A2	A3
1	25	26	27
2	21	25	26
3	18	25	26
4	16	20	18
5	12	16	20

II sposób

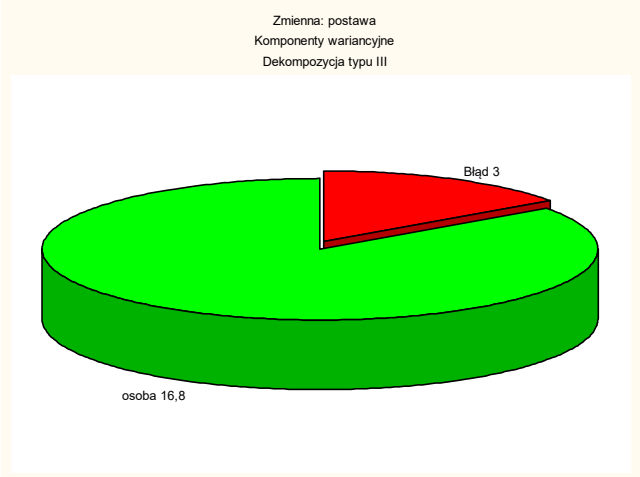
Budujemy bazę danych zestawiając kolumnę **Postawa** $= \begin{bmatrix} A_1 \\ A_2 \\ A_3 \end{bmatrix}$ i dwie kolumny kodów

WERSJA	OSOBA	POSTAWA
1	1	25
1	2	21
1	3	18
1	4	16
1	5	12
2	1	26
2	2	25
2	3	25
2	4	20
2	5	16

3	1	27
3	2	26
3	3	26
3	4	18
3	5	20

Uruchamiamy GLM
ZMIENNE ZALEŻNE : Postawa
PREDYKTORY JAKOŚCIOWE : Wersja , Osoba
Efekty międzygrupowe Wersja + Osoba
Opcje/Czynniki losowe: Osoba

Efekt	Jednowymiarowe testy istotności, wielkości efektów i moce dla postawa (Arkusz1) Model przeparametryzowany Dekompozycja typu III										
	Efekt (S/L)	SS	Stopnie swobody	MS	Łącz. Mn. df dla b	Łącz. Mn. MS dla b	F	p	Eta-kwadrat cząstkowe	Niecentralność	Moc obserwowana (alfa= 0,05)
Wyraz wolny	Stałe	6869,4	1	6869,4	4	53,4	128,6	0,000	0,970	128,6	1,000
wersja	Stałe	70,0	2	35,0	8	3,0	11,7	0,004	0,745	23,3	0,950
osoba	Losowy	213,6	4	53,4	8	3,0	17,8	0,000	0,899	71,2	1,000
Błąd		24,0	8	3,0							

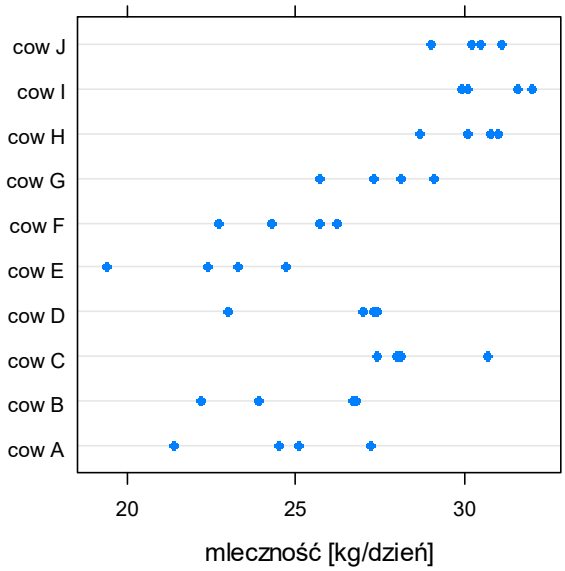



```
postawa<-c(25,21,18,16,12,26,25,25,20,16,27,26,26,18,20)
wersja<-c(1,1,1,1,1,2,2,2,2,2,3,3,3,3,3)
osoba<-c(1,2,3,4,5,1,2,3,4,5,1,2,3,4,5)
dane1<-data.frame(postawa,wersja,osoba)
dane1$wersja<-factor(dane1$wersja)
levels(dane1$wersja) <- c("w1","w2","w3")
dane1$osoba<-factor(dane1$osoba)
levels(dane1$osoba) <- c("o1","o2","o3","o4","o5")

model_a<-lmer(postawa~wersja+(1|osoba),data=dane1,REML=FALSE)#estymacja metodą ML
model_a
```

```
fixef(model_a)
ranef(model_a)
dotplot(ranef(model_a,condVar=TRUE))
model_a@pp$X #macierz X dla modelu 1
t(model_a@pp$Zt) # macierz Z dla modelu 1
image(getME(model_a,"Z"))#graficzny obraz macierzy Z dla modelu_a
summary(model_a)

model_b<-lmer(postawa~wersja+(1|osoba),data=dane1,REML=TRUE)#estymacja metodą REML
model_b
fixef(model_b)
ranef(model_b)
summary(model_b)
#uwaga Komponenty wariancyjne  westymowane metodę REML są takie jak w Statistica

#test ilorazu wiarygodności istotności czynnika wersja
logLik(model_a)  #pełny model
model_aw=update(model_a,.~.-wersja)
logLik(model_aw)
anova(model_a,model_aw)
```

Przykład : mleczność krów. Dla wybranych 10 krów zmierzono ilość mleka wyprodukowanego dzienne przez daną krowę. Dla każdej krowy pomiar powtórzono czterokrotnie

```
 library(PBImisc)
#odczytujemy dane o mleczności krów, w pierwszej kolumnie jest identyfikator krowy,
#w drugiej dzienny udój w kilogramach
data(milk)
summary(milk)
        cow          milk.amount
 cow A  : 4   Min.   :19.40
 cow B  : 4   1st Qu.:24.65
 cow C  : 4   Median :27.30
 cow D  : 4   Mean   :27.02
 cow E  : 4   3rd Qu.:29.95
 cow F  : 4   Max.   :32.00
 (Other):16
#średnie mleczności krów
(średnie=with(milk,tapply(milk.amount,cow,mean)))
   cow A  cow B  cow C  cow D  cow E  cow F  cow G  cow H  cow I  cow J
   24.550 24.900 28.550 26.175 22.450 24.725 27.550 30.150 30.900 30.200
#graficzna prezentacja danych
library(lattice)
dotplot(cow~milk.amount,data=milk,xlab="mleczność [kg/dzień]")
```




```
#model z jednym komponentem wariancyjnym
library(lme4)

(model1<-lmer(milk.amount~(1|cow),data=milk))
        Linear mixed model fit by REML ['lmerMod']
        Formula: milk.amount ~ (1 | cow)
```

```
Data: milk
REML criterion at convergence: 178.9

Scaled residuals:
    Min      1Q  Median      3Q     Max
-1.9981 -0.4136  0.1775  0.6561  1.4021

Random effects:
 Groups   Name        Variance Std.Dev.
 cow      (Intercept) 7.589    2.755
 Residual             2.999    1.732

Number of obs: 40, groups:  cow, 10

Fixed effects:
            Estimate Std. Error t value
(Intercept) 27.0150   0.9132    29.58
```

Otrzymaliśmy oceny:

- efektu stałego średniej 27.015
- wariancji efektów losowych $\hat{\sigma}_1^2 = 7.589$
- wariancji zakłócenia losowego $\hat{\sigma}_0^2 = 2.999$

```
# ten sam model z jednym komponentem wariancyjnym  estymowany ML
model3<-lmer(milk.amount~(1|cow),data=milk,REML=FALSE)
```

Porównajmy uzyskane wyniki z modelem efektów stałych

```
#model ze stałym efektem krowy
summary(model2<-lm(milk.amount~cow,data=milk))
Call:
lm(formula = milk.amount ~ cow, data = milk)

Residuals:
   Min     1Q Median     3Q    Max
-3.175 -1.000  0.150  1.006  2.650
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   24.550      0.866  28.350  < 2e-16 ***
cowcow B       0.350      1.225   0.286  0.77700
cowcow C       4.000      1.225   3.266  0.00273 **
cowcow D       1.625      1.225   1.327  0.19455
cowcow E      -2.100      1.225  -1.715  0.09670 .
cowcow F       0.175      1.225   0.143  0.88733
cowcow G       3.000      1.225   2.450  0.02035 *
cowcow H       5.600      1.225   4.573  7.76e-05 ***
cowcow I       6.350      1.225   5.185  1.38e-05 ***
cowcow J       5.650      1.225   4.614  6.92e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Residual standard error: 1.732 on 30 degrees of freedom
Multiple R-squared:  0.7694,     Adjusted R-squared:  0.7002
```

F-statistic: 11.12 on 9 and 30 DF, p-value: 2.161e-07

W podsumowaniu modelu2 podane są oceny efektów każdej krowy, ponieważ krowa była traktowana jako efekt stały.

Używając funkcji ranef() wyznaczymy predykcje efektów losowych

```
ocenaMR<-unlist(ranef(model1))+fixef(model1) #REML
ocenaMM<-unlist(ranef(model3))+fixef(model3)# ML
ocenaLM<-c(0,model2$coef[-1])+model2$coef[1]#LM
cbind(średnie,ocenaLM,ocenaMR,ocenaMM)
```

	średnie	ocenaLM	ocenaMR	ocenaMM
cow A	24.550	24.550	24.77168	24.79631
cow B	24.900	24.900	25.09020	25.11133
cow C	28.550	28.550	28.41196	28.39662
cow D	26.175	26.175	26.25054	26.25893
cow E	22.450	22.450	22.86053	22.90614
cow F	24.725	24.725	24.93094	24.95382
cow G	27.550	27.550	27.50189	27.49654
cow H	30.150	30.150	29.86807	29.83675
cow I	30.900	30.900	30.55062	30.51181
cow J	30.200	30.200	29.91358	29.88175

Jak widać na powyższym przykładzie oceny mleczności krów w modelu efektów stałych (ocenaLM) są identyczne ze średnimi grupowymi (średnie). Oceny wyznaczone z użyciem modeli mieszanych są inne i co do wartości ściągnięte w kierunku średniej a więc w kierunku 27.015. Oceny ML i REML różnią się nieznacznie.

Przykład : efekt stały genu i jeden komponent wariancyjny

W zbiorze danych poza roczną mlecznością i indeksem krowy mamy informację o numerze laktacji , o genotypie danego osobnika i o ilości tłuszczu w mleku wyprodukowanym przez krowę. Hipoteza badawcza, która nas interesuje, to określenie czy obserwowana mutacja w genie BTN1 wpływa na mleczność krów. Mamy 915 pomiarów mleczności dla osobników o genotypie BTN3A1=1(allel dziki) i 85 pomiarów dla genotypu BTN3A1=2 (allel zmutowany). Nie możemy tu zastosować testu t Studenta dla dwóch prób ani modelu z jednym efektem stałym efektem genu BNT3A1, ponieważ część osobników była mierzona więcej niż 1 raz/przez kilka laktacji/ . Stąd też niektóre pomiary są do siebie bardziej podobne i nie jest spełnione założenie jednorodności wariancji i niezależności zakłócenia losowego.

Budujemy model w którym efektami stałymi będą gen BTN1 i numer laktacji a efekt osobniczy będzie uwzględniony jako efekt losowy.

Plik Milkgene.R

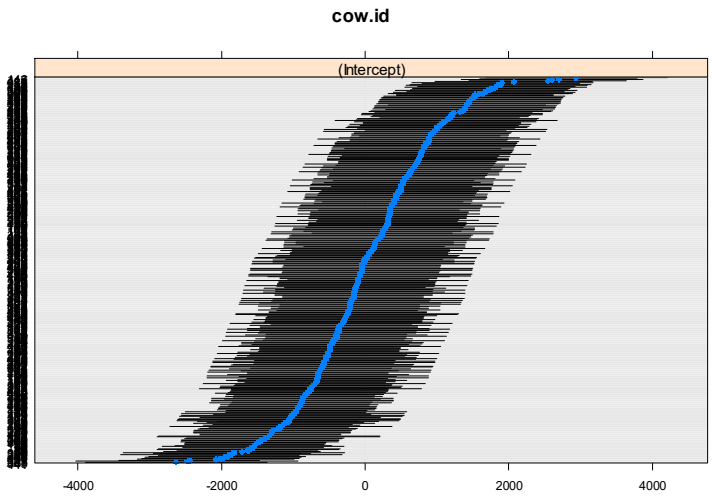
```
library(PBImisc)
data(milkgene)
```

```
library(lattice)
xyplot(milk~fat, data=milkgene)
bwplot(milk~lactation, data=milkgene)

summary(milkgene)

#model z dwoma efektami stałymi jednym komponentem wariancyjnym
library(lme4)
model1<-lmer(milk~btn3a1+lactation+(1|cow.id),data=milkgene)
summary(model1)
```

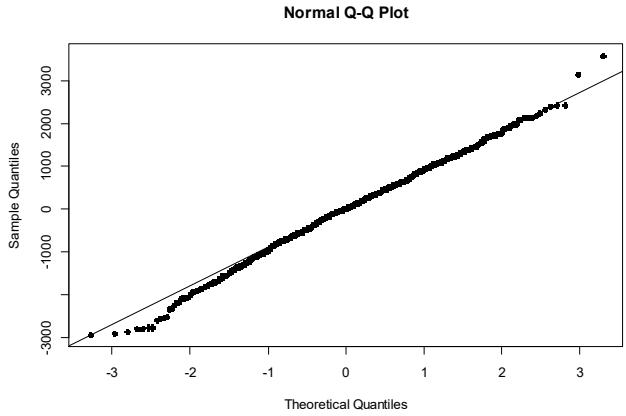
```
Linear mixed model fit by REML ['lmerMod']
Formula: milk ~ btn3a1 + lactation + (1 | cow.id)
   Data: milkgene

REML criterion at convergence: 17303.6

Scaled residuals:
    Min      1Q  Median      3Q     Max
-2.6325 -0.5378  0.0279  0.5505  3.2110

Random effects:
 Groups   Name        Variance Std.Dev.
 cow.id   (Intercept) 1240403  1114
 Residual             1252911  1119
Number of obs: 1000, groups:  cow.id, 409

Fixed effects:
            Estimate Std. Error t value
(Intercept)  6699.45     81.08  82.632
btn3a12      -244.08    235.12  -1.038
lactation2   1307.04     84.63  15.443
lactation3   1800.54    102.10  17.635
lactation4   1669.27    176.45   9.460

Correlation of Fixed Effects:
            (Intr) btn312 lcttn2 lcttn3
btn3a12     -0.269
lactation2 -0.454  0.029
lactation3 -0.374  0.015  0.395
lactation4 -0.214  0.000  0.228  0.234
```

Efekt genu BTN1 został oceniony na -244.08- tzn. o tyle średnio niższa jest mleczność krów o genotypie BTN1=2 od mleczności osobników BTN1=1. Parametr $\sigma^2_{cow.id}$ został oceniony na 1240403 i jest praktycznie równy ocenie parametru σ^2_0. Oznacza to, że zmienność σ^2_0 mleczności dla jednej krowy jest mniej więcej dwukrotnie mniejsza niż zmienność dla różnych krów $\sigma^2_{cow.id} + \sigma^2_0$.

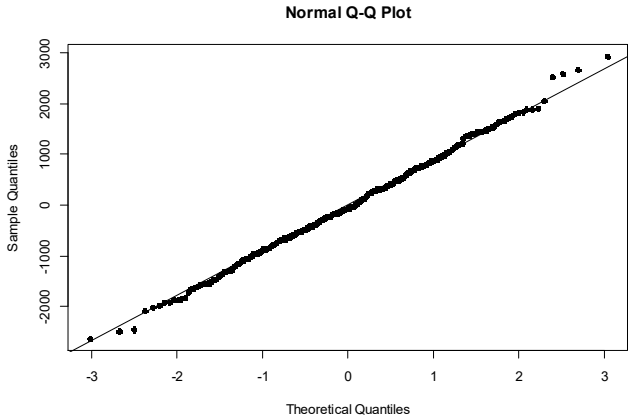
```
#wyznaczymy predykcję efektów losowych
#argument condVar=TRUE dodaje do wyniku informacje o odchyleniach standardowych
u=ranef(model1,condVar=TRUE)
dotplot(u)
```




```
#wyznaczymy reszty z modelu
e=resid(model1)
#wyznaczymy wykres QQplot dla reszt z modelu
qqnorm(e,pch=16)
qqline(e)
```

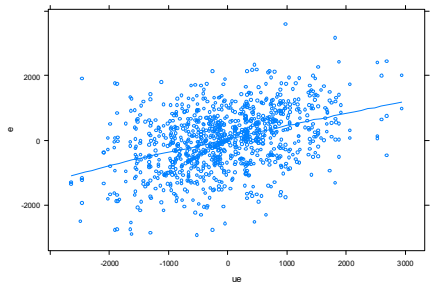


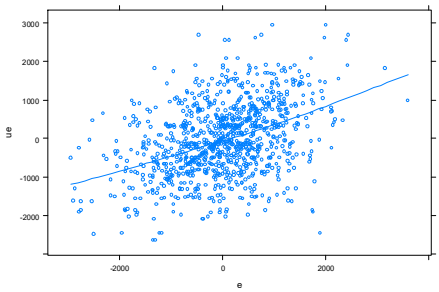

```
#wyznaczymy wykres QQplot dla efektów losowych z modelu
ul=u[[1]][[1]]# wybieramy predykcje efektów losowych bez odchyleń standardowych
qqnorm(ul,pch=16)
qqline(ul)
```



Zależność między predykcjami efektów losowych a ocenami reszt

```
#wektor losowy ul ma inną długość niż epsilon; aby je porównać wyznaczymy  wartość
# odpowiadającą  efektowi u dla każdej obserwacji ze zbioru danych
Z=model.matrix(~factor(cow.id)-1,milkgene)
ue=Z %*% ul# wektor o tej samej długości co wektor epsilon
xyplot(e~ue,type=c("p","smooth"))
xyplot(ue~e,type=c("p","smooth"))
```





Predykcje efektu losowego są nieznacznie skorelowane z wartościami reszt

Testowanie istotności efektów stałych

- Test Walda istotności genu BTN3A1

```
#Testowanie istotności efektów stałych
#istotność  genu
summary(model1)
```

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	6699.45	81.08	82.632
btn3a12	-244.08	235.12	-1.038
lactation2	1307.04	84.63	15.443
lactation3	1800.54	102.10	17.635
lactation4	1669.27	176.45	9.460

```
uu=summary(model1)
wsp<-as.data.frame(uu[10]) #obiekt uu  jest listą która na pozycji 10 zawiera współczynniki
wsp
tse<-wsp[2,3]
#istotność  genu
2*pnorm(tse,lower.tail = TRUE)
```

[1] 0.2991872

Wniosek. Efekt genu nie jest istotny

- Test ilorazu wiarygodności istotności genu BTN3A1 (musimy użyć estymatorów ML a nie REML

```
#Test ilorazu wiarygodności  istotności genu BTN3A1 (musimy użyć estymatorów ML a nie REML)
modelPełny=lmer(milk~btn3a1+lactation+(1|cow.id),data=milkgene,REML=FALSE)
modelBezBTN1=update(modelPełny,.~.-btn3a1)
logLik(modelPełny)
logLik(modelBezBTN1)
anova(modelPełny,modelBezBTN1)
```

Data: milkgene

Models:

modelBezBTN1: milk ~ lactation + (1 | cow.id)

modelPełny: milk ~ btn3a1 + lactation + (1 | cow.id)

	npar	AIC	BIC	logLik	deviance	Chisq	Df	Pr(>Chisq)
modelBezBTN1	6	17373	17403	-8680.7	17361			
modelPełny	7	17374	17409	-8680.2	17360	1.0826	1	0.2981

Wniosek. Efekt genu nie jest istotny

```
#Test permutacyjny istotności genu BTN3A1
N=1000
logs=replicate(N,logLik(lmer(milk~sample(btn3a1)+lactation+(1|cow.id),data=milkgene,REML=FALSE)))
```

```
mean(logs>logLik(modelPełny))
        [1] 0.301
```

Wniosek. Efekt genu nie jest istotny

Testowanie istotności efektów losowych

- Test ilorazu wiarygodności

Aby przeprowadzić test ilorazu wiarygodności wyznaczymy logarytm funkcji wiarygodności dla modelu z komponentem wariancyjnym i bez.

```
#Testowanie istotności efektów losowych
(lL1<-logLik(model1<-lmer(milk~btn3a1+lactation+(1|cow.id),data=milkgene,REML=FALSE)))
(lL2<-logLik(model2<-lm(milk~btn3a1+lactation,data=milkgene)))
(lDelta=as.numeric(lL2-lL1))
pchisq(-2*lDelta,1,lower.tail=FALSE)
pchisq(-2*lDelta,409,lower.tail=FALSE)
```

```
#test permutacyjny
milkgene2=milkgene
N=1000
logs=replicate(N,{
   milkgene2$cow.id=sample(milkgene2$cow.id);
   logLik(lmer(milk~btn3a1+lactation+(1|cow.id),data=milkgene2))})

mean(logs > logLik(model1))
```

Model mieszany z dwoma komponentami wariancyjnymi, dwie zmienne grupujące

Rozważmy zbiór 7 obserwacji i że w tym zbiorze danych mamy dwie zmienne jakościowe z_1 o dwóch poziomach tzn. $z_1=(A,A,A,B,B,B,B)$ i z_2 o trzech poziomach tzn. $z_2=(Q, Q, Q,R, R, S, S)$.
Jeżeli w formule opisującej model w programie R dodamy $(1|\, z_1)+ (1|\, z_2)$ to będzie oznaczało uwzględnienie poziomów zmiennych z_1 i z_2 jako efektów losowych, a więc dodanie dwóch komponentów wariancyjnych.
Macierze Z_1 i Z_2 z modelu mieszanego są postaci

$$Z_1=\begin{bmatrix}1&0\\1&0\\1&0\\0&1\\0&1\\0&1\\0&1\end{bmatrix},\quad Z_2=\begin{bmatrix}1&0&0\\1&0&0\\1&0&0\\0&1&0\\0&1&0\\0&0&1\\0&0&1\end{bmatrix}$$

Macierz kowariancyjna $V(\mathbf{Y})=\sigma_0^2\mathbf{I}+\sigma_1^2\mathbf{Z}_1\mathbf{Z}_1^T+\sigma_2^2\mathbf{Z}_2\mathbf{Z}_2^T$ jest postaci

$$V_1 = \begin{bmatrix} \sigma_1^2 & \sigma_1^2 & \sigma_1^2 & 0 & 0 & 0 & 0 \\ \sigma_1^2 & \sigma_1^2 & \sigma_1^2 & 0 & 0 & 0 & 0 \\ \sigma_1^2 & \sigma_1^2 & \sigma_1^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_1^2 & \sigma_1^2 & \sigma_1^2 & \sigma_1^2 \\ 0 & 0 & 0 & \sigma_1^2 & \sigma_1^2 & \sigma_1^2 & \sigma_1^2 \\ 0 & 0 & 0 & \sigma_1^2 & \sigma_1^2 & \sigma_1^2 & \sigma_1^2 \\ 0 & 0 & 0 & \sigma_1^2 & \sigma_1^2 & \sigma_1^2 & \sigma_1^2 \end{bmatrix}$$

$$V_2 = \begin{bmatrix} \sigma_2^2 & \sigma_2^2 & \sigma_2^2 & 0 & 0 & 0 & 0 \\ \sigma_2^2 & \sigma_2^2 & \sigma_2^2 & 0 & 0 & 0 & 0 \\ \sigma_2^2 & \sigma_2^2 & \sigma_2^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_2^2 & \sigma_2^2 & 0 & 0 \\ 0 & 0 & 0 & \sigma_2^2 & \sigma_2^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_2^2 & \sigma_2^2 \\ 0 & 0 & 0 & 0 & 0 & \sigma_2^2 & \sigma_2^2 \end{bmatrix}$$

$$V = \begin{bmatrix} \sigma_0^2+\sigma_1^2+\sigma_2^2 & \sigma_1^2+\sigma_2^2 & \sigma_1^2+\sigma_2^2 & 0 & 0 & 0 & 0 \\ \sigma_1^2+\sigma_2^2 & \sigma_0^2+\sigma_1^2+\sigma_2^2 & \sigma_1^2+\sigma_2^2 & 0 & 0 & 0 & 0 \\ \sigma_1^2+\sigma_2^2 & \sigma_1^2+\sigma_2^2 & \sigma_0^2+\sigma_1^2+\sigma_2^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_0^2+\sigma_1^2+\sigma_2^2 & \sigma_1^2+\sigma_2^2 & \sigma_1^2 & \sigma_1^2 \\ 0 & 0 & 0 & \sigma_1^2+\sigma_2^2 & \sigma_0^2+\sigma_1^2+\sigma_2^2 & \sigma_1^2 & \sigma_1^2 \\ 0 & 0 & 0 & \sigma_1^2 & \sigma_1^2 & \sigma_0^2+\sigma_1^2+\sigma_2^2 & \sigma_1^2+\sigma_2^2 \\ 0 & 0 & 0 & \sigma_1^2 & \sigma_1^2 & \sigma_1^2+\sigma_2^2 & \sigma_0^2+\sigma_1^2+\sigma_2^2 \end{bmatrix}$$

Hierarchiczny model mieszany

Rozważmy zbiór 8 obserwacji i że w tym zbiorze danych mamy dwie zmienne jakościowe z_1 o dwóch poziomach z_1=(*A*,*A*,*A*,*A*,*B*,*B*,*B*,*B*), z_2=(*Q*, *Q*, *R*, *R*, *Q*, *Q*, *Q*,*R*) przy czym z_2 jest zagnieżdżona w z_1. W programie R w formule definiującej model pojawi się składnik y~ $(1|z_1)+(1|z_1{:}\, z_2)$ (y~(1|z1/z2) inny zapis powyższego modelu)

Macierze Z_1 i Z_2 z modelu mieszanego są postaci

$$Z_1 = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}, \quad Z_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Macierz kowariancyjna $V(\mathbf{Y}) = \sigma_0^2\mathbf{I} + \sigma_1^2\mathbf{Z}_1\mathbf{Z}_1^T + \sigma_2^2\mathbf{Z}_2\mathbf{Z}_2^T$ jest postaci

$$V_1=\begin{bmatrix}\sigma_1^2&\sigma_1^2&\sigma_1^2&\sigma_1^2&0&0&0&0\\ \sigma_1^2&\sigma_1^2&\sigma_1^2&\sigma_1^2&0&0&0&0\\ \sigma_1^2&\sigma_1^2&\sigma_1^2&\sigma_1^2&0&0&0&0\\ \sigma_1^2&\sigma_1^2&\sigma_1^2&\sigma_1^2&0&0&0&0\\ 0&0&0&0&\sigma_1^2&\sigma_1^2&\sigma_1^2&\sigma_1^2\\ 0&0&0&0&\sigma_1^2&\sigma_1^2&\sigma_1^2&\sigma_1^2\\ 0&0&0&0&\sigma_1^2&\sigma_1^2&\sigma_1^2&\sigma_1^2\\ 0&0&0&0&\sigma_1^2&\sigma_1^2&\sigma_1^2&\sigma_1^2\end{bmatrix}\qquad V_2=\begin{bmatrix}\sigma_2^2&\sigma_2^2&0&0&0&0&0&0\\ \sigma_2^2&\sigma_2^2&0&0&0&0&0&0\\ 0&0&\sigma_2^2&\sigma_2^2&0&0&0&0\\ 0&0&\sigma_2^2&\sigma_2^2&0&0&0&0\\ 0&0&0&0&\sigma_2^2&\sigma_2^2&\sigma_2^2&0\\ 0&0&0&0&\sigma_2^2&\sigma_2^2&\sigma_2^2&0\\ 0&0&0&0&\sigma_2^2&\sigma_2^2&\sigma_2^2&0\\ 0&0&0&0&0&0&0&\sigma_2^2\end{bmatrix}$$

Plik Biecek212.R

```
library(lme4)
y<-5+rnorm(8)
z1<-as.factor(c("A","A","A","A","B","B","B","B"))
z2<-as.factor(c("Q","Q","R","R","Q","Q","Q","R"))
dane<-data.frame(y,z1,z2)
model<-lmer(y~ (1|z1)+(1|z1:z2),data=dane,REML=FALSE,
control = lmerControl(optimizer ="Nelder_Mead"))#estymacja metodą ML
#   y~(1|z1/z2) inny zapis powyższego modelu
summary(model)
model@pp$Zt # transponowana macierz Z
t(model@pp$Zt) # macierz Z
image(getME(model,"Z"))#graficzny obraz macierzy Z
```

```
8 x 6 sparse Matrix of class "dgCMatrix"
  A:Q A:R B:Q B:R A B
1   1   .   .   . 1 .
2   1   .   .   . 1 .
3   .   1   .   . 1 .
4   .   1   .   . 1 .
5   .   .   1   . . 1
6   .   .   1   . . 1
7   .   .   1   . . 1
8   .   .   .   1 . 1
```

Widać że macierz $\mathbf{Z}=[\mathbf{Z}_2,\mathbf{Z}_1]$ powstaje przed dołączenie do macierzy $\mathbf{Z}_2$ kolumn macierzy $\mathbf{Z}_1$.

Sprawdzenie w Mathematica 13.3

$$Z=\begin{matrix}1&0&0&0&1&0\\1&0&0&0&1&0\\0&1&0&0&1&0\\0&1&0&0&1&0\\0&0&1&0&0&1\\0&0&1&0&0&1\\0&0&1&0&0&1\\0&0&0&1&0&1\end{matrix}\qquad \mathrm{D}=\begin{matrix}\sigma_2^2&0&0&0&0&0\\0&\sigma_2^2&0&0&0&0\\0&0&\sigma_2^2&0&0&0\\0&0&0&\sigma_2^2&0&0\\0&0&0&0&\sigma_1^2&0\\0&0&0&0&0&\sigma_1^2\end{matrix}$$

$$\mathrm{ZDZ'}=\begin{matrix}\sigma_1^2+\sigma_2^2&\sigma_1^2+\sigma_2^2&\sigma_1^2&\sigma_1^2&0&0&0&0\\ \sigma_1^2+\sigma_2^2&\sigma_1^2+\sigma_2^2&\sigma_1^2&\sigma_1^2&0&0&0&0\\ \sigma_1^2&\sigma_1^2&\sigma_1^2+\sigma_2^2&\sigma_1^2+\sigma_2^2&0&0&0&0\\ \sigma_1^2&\sigma_1^2&\sigma_1^2+\sigma_2^2&\sigma_1^2+\sigma_2^2&0&0&0&0\\ 0&0&0&0&\sigma_1^2+\sigma_2^2&\sigma_1^2+\sigma_2^2&\sigma_1^2+\sigma_2^2&\sigma_1^2\\ 0&0&0&0&\sigma_1^2+\sigma_2^2&\sigma_1^2+\sigma_2^2&\sigma_1^2+\sigma_2^2&\sigma_1^2\\ 0&0&0&0&\sigma_1^2+\sigma_2^2&\sigma_1^2+\sigma_2^2&\sigma_1^2+\sigma_2^2&\sigma_1^2\\ 0&0&0&0&\sigma_1^2&\sigma_1^2&\sigma_1^2&\sigma_1^2+\sigma_2^2\end{matrix}$$