

Uogólnione modele liniowe

Laboratorium nr 12

12.1 (Modele mieszane) Przykład niniejszy pochodzi z książki Przemysława Biecka „Analiza danych z programem R. Modele liniowe z efektami stałymi, losowymi i mieszanymi”, Wydawnictwo Naukowe PWN, 2011. Dane dotyczą mleczności krów z pewnej farmy. Aby ocenić mleczność krów, dla 10 wybranych krów mierzone było mleko wyprodukowanego dziennie – dla każdej krowy pomiar powtórzono pięciokrotnie (patrz zbiór `milk`, dostępny w pakiecie `PBImisc`, zawierającym dane do wspomnianej książki).

- (a) Wczytać dane ze zbioru `milk` (w pierwszej kolumnie jest identyfikator krowy, w drugiej zaś dzienny udój w kilogramach). Wyliczyć średnie mleczności poszczególnych krów.
- (b) Wykorzystać funkcję `dotplot` z pakietu `lattice` do graficznego przedstawienia zależności `cow` od `milk.amount`.
- (c) Dopasować do danych model mieszany

$$y_{\text{milk.amount}} = \mu + Z_{\text{cow}} u_{\text{cow}} + \varepsilon,$$

gdzie wektor u_{cow} opisuje efekty mleczności krów, a ε – dzienną zmienność mleczności. Można skorzystać z funkcji `lmer` z pakietu `lme4`:

```
library(lme4)
model1<-lmer(milk.amount~(1|cow), data=milk)
```

To wydaje się być równoważne (sprawdzić) napisaniu

```
model2 <- lmer(milk.amount~1+(1|cow),data=milk)
```

– efekt stały to tylko intercept, reprezentowany przez pierwszą 1 w formule modelu, efekt losowy jest reprezentowany przez składnik $(1|cow)$ – to oznacza, że dane zgrupowane są po `cow` i – to wskazuje 1 – efekt losowy jest stały w każdej grupie wyznaczonej przez `cow`. Odczytać oceny efektów stałych i oceny wariancji efektów losowych. Zauważyć, że dla efektów stałych wyznaczone są także wartości statystyki testu Walda (wartość $t = 29,59$ wskazuje, że efekt średniej faktycznie różni się od zera).

- (d) Dopasować do danych model liniowy z jedną zmienną jakościową (czyli model jednokierunkowej analizy wariancji). Sprawdzić, czy oceny (predykcje) efektów osobniczych krów będą takie same w obu modelach. Do wyznaczania predykcji efektów losowych użyć funkcji `ranef()` z biblioteki `lme4` (podobnie, przy wyznaczaniu oceny efektów stałych można użyć funkcji `fixef()`, także z `lme4`).

12.2 Nieprzetworzone bale wełny zawierają różne ilości nieczystości (np. tłuszczu), które muszą być usunięte przed dalszym przetwarzaniem. Czystość wełny wyraża się jako procent czystej wełny w balu wełny. Zbiór `bales.txt` zawiera informację dotyczącą czystości w siedmiu balach wełny (kolejne kolumny danych) pochodzącej z Urugwaju. Kolejne wiersze oznaczają cztery próbki (pobierane ze wszystkich siedmiu bali). Niech μ_i oznacza średnią czystość wełny w i -tym balu. (Z pojedynczej dostawy wybrano losowo siedem bali, a następnie, także losowo, wybrano po cztery próbki z każdej bali.)

- (a) Narysować wykresy skrzynkowe i wykresy typu *stripchart* oraz `dotplot` czystości wełny w siedmiu grupach.
- (b) Przetestować hipotezy

$$H_0 : y_{i,j} = \mu + \epsilon_{i,j}, \quad H_1 : y_{i,j} = \mu + \alpha_i + \epsilon_{i,j}$$

Czy można twierdzić, że średnie w siedmiu grupach są równe?

- (c) Zaproponować i dopasować model, który uwzględni fakt, że badane bale wełny są losowymi reprezentantami populacji wszystkich bali wełny.