Modelos Mistos e Aplicações - Lista 1

Renata Hirota

1. Prove que a estatística F do ANOVA tem distribuição F

Um variável aleatória Y tem distribuição F_{d_1,d_2} se

$$Y = \frac{X_1/d_1}{X_2/d_2}$$

onde $X_1 \sim \chi^2_{d_1}$ e $X_2 \sim \chi^2_{d_2}$ são independentes.

Na ANOVA, a estatística F é definida como

$$\frac{QMA}{QME}, \quad QMA = SQA/(t-1), QME = SQE/(N-t)$$

$SQA = \sum_{i=1}^{t} n(\bar{Y}_{i.} - \bar{Y}_{..})^2$, ou seja, soma de quadrados de uma diferença de normais. Dividido por $(t-1)$, a quantidade de elementos somados, temos uma variável com distribuição qui-quadrado,

$$QMA \sim \chi^2_{(t-1)}$$

O equivalente pode ser dito de SQE e QME, já que SQE é

$$\sum_{i=1}^{t} \sum_{j=1}^{n} (Y_{ij} - \bar{Y}_{i.})^2$$

Logo, $QME \sim \chi^2_{(N-t)}$

Como QME e QMA são independentes, temos que $F \sim F_{t-1,N-t}$

2. MLE para d

Para obter a estimativa de máxima verossimilhança (MLE) para a componente aleatória em um modelo de efeitos mistos, maximizamos a função de verossimilhança condicional.

Dado o modelo

$$Y_{ij} = \beta_0 + b_i + \epsilon_{ij}$$

onde

- Y_{ij} é a observação para o j-ésimo sujeito no i-ésimo nível do tratamento,
- β_0 é o efeito fixo do intercepto,
- b_i é a componente aleatória associada ao i-ésimo tratamento,
- ϵ_{ij} é o erro aleatório associado à observação Y_{ij}.

A função de verossimilhança condicional $L(\theta|Y)$ é definida como a probabilidade conjunta das observações Y_{ij}, dadas as estimativas dos parâmetros θ (que incluem os efeitos fixos e a variância da componente aleatória)

$$L(\theta|Y) = \prod (Y_{11}, Y_{12}, ..., Y_{ij}|\theta)$$
$$ln(L(\theta|Y)) = \sum [ln(f(Y_i j|\theta))]$$

$$\frac{\partial[ln(L(\theta|Y))]}{\partial(\sigma_b^2)} = 0$$

3. Sejam duas variáveis aleatórias G e Y. Prove que Var(G) pode ser escrita como a soma da esperança condicional e variância condicional de Y

Podemos definir a variância de G como

$$Var(G) = E[(G - E[G])^2]$$

Pela esperança condicional, temos

$$Var(G) = E[(E[G|Y] + (G - E[G]))^2]$$
$$Var(G) = E[E[G|Y]^2 + 2(E[G|Y])(G - E[G]) + (G - E[G])^2]$$

$$Var(G) = E[E[G|Y]^2] + 2E[(E[G|Y])(G - E[G])] + E[(G - E[G])^2]$$

$$Var(G) = E[E[G|Y]^2] + 2E[(E[G|Y])(G - E[G])] + E[(G - E[G])^2]$$

O segundo termo é a esperança do produto da esperança condicional $E[G|Y]$ com a diferença $(G - E[G])$, o que é igual a zero, pois a esperança condicional de G é uma função de Y. Portanto, esse termo se reduz a zero:

$$Var(G) = E[E[G|Y]^2] + E[(G - E[G])^2]$$

Então,

$$Var(G) = E[Var(G|Y)] + Var(E[G|Y])$$

5. PRT

```
dados <- nlmeU::prt.fiber |>
  dplyr::filter(fiber.f == "Type 2") |>
  dplyr::inner_join(nlmeU::prt.subjects, "id")
```

Vamos considerar como efeitos aleatórios os participantes, com as seguintes variáveis associadas a eles:

Participante (`id`):

- Sexo (`sex.f`)
- Idade (`age.f`)
- IMC (`bmi`)

Consideraremos como efeito fixo:

- Intensidade do treinamento (`prt.f`)
- Ocasião (`occ.f`)

A variável resposta é a medida de força específica da fibra muscular (`spec.fo`)

Modelo

$$Y = XB + ZU + E$$

- Y é um vetor de observações da variável de resposta `spec.fo`
- X é a matriz de covariáveis fixas, incluindo um vetor de uns para o intercepto e as covariáveis `prt.f`, `age.f`, `sex.f`, `bmi` e `occ.f`.
- B é um vetor de coeficientes correspondentes às covariáveis fixas, incluindo o intercepto.
- Z é a matriz de desenho do efeito aleatório, que contém um vetor de uns para a unidade de agrupamento i
- U é um vetor de efeitos aleatórios u_i associados a cada unidade de agrupamento (`id`), modelado como $u_i \sim N(0, \sigma_u^2)$
- E é um vetor de erros residuais ϵ_{ij} para cada observação, modelado como $\epsilon_{ij} \sim N(0, \sigma^2)$

```r
fit <- nlme::lme(
  spec.fo ~ 1 + prt.f + age.f + sex.f + bmi + occ.f,
  random = ~ 1|id,
  data = dados
)
summary(fit)
```

```
Linear mixed-effects model fit by REML
  Data: dados
       AIC      BIC    logLik
  11187.58 11228.07 -5585.789

Random effects:
 Formula: ~1 | id
        (Intercept) Residual
StdDev:    13.56403 27.49065

Fixed effects:  spec.fo ~ 1 + prt.f + age.f + sex.f + bmi + occ.f
                Value Std.Error   DF   t-value p-value
(Intercept) 149.46213 15.432129 1108  9.685127  0.0000
prt.fLow     -1.61267  3.847214   58 -0.419179  0.6766
age.fOld      6.96845  3.877306   58  1.797240  0.0775
sex.fMale    -4.37663  3.864605   58 -1.132491  0.2621
bmi           0.60780  0.589132   58  1.031689  0.3065
occ.fPos      8.50944  1.661194 1108  5.122487  0.0000
 Correlation:
          (Intr) prt.fL ag.fOl sx.fMl bmi
```

```
prt.fLow  -0.193
age.fOld  -0.242  0.013
sex.fMale -0.028 -0.044 -0.006
bmi       -0.966  0.072  0.106 -0.085
occ.fPos  -0.058  0.003  0.031  0.011 -0.005

Standardized Within-Group Residuals:
        Min          Q1         Med          Q3         Max
-3.03668536 -0.65887912 -0.03786059  0.61449852  3.98073017

Number of Observations: 1172
Number of Groups: 63
```

A variância dos efeitos aleatórios é alta, o que indica que de fato o modelo misto é adequado.

Observamos também que há baixa correlação entre os efeitos fixos, o que descarta uma possível multicolinearidade.

No entanto, não parece haver efeito significante da variável de interesse `prt.f` (intensidade do treinameto) na variável resposta (`spec.fo`), mas sim da ocasião em que o treino ocorre (`occ.f`).