Alunos: Álvaro Kothe, Giovanni Piccirilli, Larissa Eleutério, Marcos Andrade
Docente: Viviana Giampaoli
Data: 16/09/2021

Modelos Mistos - Lista 1

Exercício 1

Prove a expressão

$$cov(y, w) = cov_u(E(y|u), E(w|u)) + E_u(cov(y, w)|u)$$

Resposta:

Como

$$cov(y, w) = E[yw] - E(y)E(w),$$

e sabendo que $E(g) = E_y[E(g|y)]$, temos que

$$cov(y, w) = E_u[E(yw|u)] - E_u[E(y|u)]E_u[E(w|u)],$$

tomando que $E(yw|u) = cov(y, w|u) + E(y|u)E(w|u)$, temos que

$$\begin{aligned} cov(y, w) &= E_u[cov(y, w|u) + E(y|u)E(w|u)] - E_u[E(y|u)]E_u[E(w|u)] \\ &= E_u[cov(y, w|u)] + E_u[E(y|u)E(w|u)] - E_u[E(y|u)]E_u[E(w|u)], \end{aligned}$$

como $E_u[E(y|u)E(w|u)] - E_u[E(y|u)]E_u[E(w|u)] = cov_u(E(y|u), E(w|u))$

$$\therefore cov(y, w) = E_u[cov(y, w|u)] + cov_u(E(y|u), E(w|u)).$$

Exercício 1 (seção 2.1)

1. Estimate slope β using the data Family.txt, assuming that the family-specific intercepts are fixed, using the dummy variable technique. [Hint: Estimate the linear regression model with $1 + 18 = 19$ parameters using lm(y~X-1), where y=famdat$Weight and X is the 71×19 and the other 18 columns as dummy variables.]

Resposta:

Como no comando do exercício será ajustado um modelo linear com apenas efeitos fixos, sendo eles, Height e FamilyID, sem intercepto. Logo nosso modelo terá 19 parâmetros no total, sendo 18 interceptos (o efeito espcifico para cada família) e o parâmetro relacionado a Height.

$$y_{ij} = \alpha_i + \beta_1 x_{ij} + \epsilon_{ij}$$

com $i = 1, \ldots, 18$ e $j = 1, \ldots, n_i$. O modelo matricialmente pode ser escrito como

$$\begin{pmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{18n_{18}} \end{pmatrix} = \begin{pmatrix} 1_{n_1} & 0 & \dots & 0 & x_{1.} \\ 0 & 1_{n_2} & \dots & 0 & x_{2.} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 1_{n_{18}} & x_{18.} \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_{18} \\ \beta_1 \end{pmatrix} + \begin{pmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \vdots \\ \epsilon_{18n_{18}} \end{pmatrix}$$

```r
family = read.table("Family.txt", header = TRUE, stringsAsFactors = FALSE)
family$FamilyID = factor(family$FamilyID)
mod1 = lm(Weight ~ Height + FamilyID - 1, data = family)
```

O coeficiente angular para a altura é 5.907, assim, O peso e altura se relacionam como $\widehat{peso}_{ij} = \widehat{familia}_i + 5.907altura_{ij}$, em que o índice $i = 1, 2, \dots, 18$ é a família do j-ésimo indivíduo, com os coeficientes:

```r
summary(mod1)
```

```
##
## Call:
## lm(formula = Weight ~ Height + FamilyID - 1, data = family)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -47.379 -14.714  -0.495  11.200  76.806
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## Height        5.9073     0.8562   6.899 7.16e-09 ***
## FamilyID1  -215.0825    57.6663  -3.730 0.000475 ***
## FamilyID2  -251.8518    62.6789  -4.018 0.000190 ***
## FamilyID3  -227.1979    62.9412  -3.610 0.000689 ***
## FamilyID4  -281.9329    61.9447  -4.551 3.24e-05 ***
## FamilyID5  -263.8219    60.6871  -4.347 6.44e-05 ***
## FamilyID6  -240.7572    59.0119  -4.080 0.000155 ***
## FamilyID7  -252.6809    57.4435  -4.399 5.42e-05 ***
## FamilyID8  -262.8287    62.8887  -4.179 0.000112 ***
## FamilyID9  -238.3499    58.1752  -4.097 0.000147 ***
## FamilyID10 -254.8798    60.1633  -4.236 9.30e-05 ***
## FamilyID11 -248.5835    60.5823  -4.103 0.000144 ***
## FamilyID12 -240.1530    59.9539  -4.006 0.000198 ***
## FamilyID13 -261.5787    62.8887  -4.159 0.000120 ***
## FamilyID14 -209.8782    61.3525  -3.421 0.001223 **
## FamilyID15 -224.1346    57.8615  -3.874 0.000302 ***
## FamilyID16 -256.5140    64.8852  -3.953 0.000234 ***
## FamilyID17 -260.7408    61.4206  -4.245 9.03e-05 ***
## FamilyID18 -238.6131    57.7516  -4.132 0.000131 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 24.85 on 52 degrees of freedom
## Multiple R-squared:  0.9835, Adjusted R-squared:  0.9775
## F-statistic: 163.4 on 19 and 52 DF,  p-value: < 2.2e-16
```

Todas as estimativas dos parâmetros foram significativas, mostrando que realmente existe o efeito da família. O problema é que a variância das nossas estimativas foram altas, que de certa forma pode atrapalhar na inferência do modelo, mostrando uma fragilidade desse modelo. Além disso, não temos como assumir que os efeitos das famílias são fixos, não sabemos se as observações são independentes dentro das famílias. Os modelos mistos nos permite adicionar um efeito aleatório que prediz o efeito da família, além de captar essa variabilidade que existe entre as famílias.

Exercício 2 (seção 2.1)

2. Denote $d = \frac{\sigma_d^2}{\sigma^2}$ as the scaled (or relative) variance of the random intercept. Express the covariance matrix (2.3) in he form $I + d11'$, where $\mathbf{1}$ is the column of vectors of 1s.

Resposta:

Segue abaixo matriz de covariância (2.3) dos elementos do grupo i.

$$V_i = \begin{bmatrix} 1+\frac{\sigma_d^2}{\sigma^2} & \dots & \frac{\sigma_d^2}{\sigma^2} \\ \vdots & \vdots & \vdots \\ \frac{\sigma_d^2}{\sigma^2} & \dots & 1+\frac{\sigma_d^2}{\sigma^2} \end{bmatrix}$$

$$V_i = \begin{bmatrix} 1+d & d & \dots & d \\ d & 1+d & \dots & d \\ \vdots & \vdots & \vdots & \vdots \\ d & d & \dots & 1+d \end{bmatrix} = I + \begin{bmatrix} d & d & \dots & d \\ d & d & \dots & d \\ \vdots & \vdots & \vdots & \vdots \\ d & d & \dots & d \end{bmatrix} = I + d \begin{bmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & \dots & 1 \end{bmatrix} = I + d11'$$

em que I é uma matriz identidade.

Exercício 4 (seção 2.1)

4. Incorporate sex and (possibly Age) into the model using fixed = Weight+Age. Is Age statistically significant? Provide an interpretation for the model.

Resposta:

Como o exercício não define se a solução deve ser usando modelos fixos ou mistos, vamos trazer as duas soluções.

Modelo de efeitos fixos Primeiramente foi ajustado um modelo de efeitos fixos com as variáveis *Sex*, *Weight* e *Age*. Nesse modelo o intercepto também é um parâmetro fixo e único para todas as pessoas, independente da família.

```
## leitura dos dados
family = read.table("Family.txt", header = TRUE, stringsAsFactors = FALSE)
## transformando identificador de familia e sexo em fator
family$FamilyID = factor(family$FamilyID)
family$Sex = factor(family$Sex)

## ajuste do modelo com efeitos fixos
mod2 = lm(Weight ~ Height + Age + Sex, data = family)

summary(mod2)
```

```
##
## Call:
## lm(formula = Weight ~ Height + Age + Sex, data = family)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -61.66 -13.65  -2.80  13.01  88.68
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -65.0745    76.3525  -0.852  0.39709
## Height        2.7425     1.1479   2.389  0.01971 *
## Age           0.6562     0.2030   3.233  0.00190 **
## Sex1         26.0965     9.5576   2.730  0.00807 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 25.48 on 67 degrees of freedom
## Multiple R-squared:  0.473,  Adjusted R-squared:  0.4494
## F-statistic: 20.05 on 3 and 67 DF,  p-value: 2.211e-09
```

Vemos que apenas o interpecto não é significativo a um nível de confiança de 95%. Observando os coeficientes agulares dessas variáveis podemos concluir que:

- **Age**: Com coeficiente positivo, podemos dizer que peso e idade são diretamente proporcionais.
- **Height**: Com coeficiente positivo, podemos dizer que peso e altura são diretamente proporcionais.
- **Sex**: pessoas do sexo 1 tem, em média 26,06 libras a mais de peso do que pessoas do sexo 0.

Modelo de efeitos mistos Ajustando agora um modelo de efeitos mistos que passa a considerar o intercepto como uma variável aleatória com distribuição Normal. Os efeitos aleatórios são preditos para cada uma das famílias (fator aleatório).

Com a alteração o intercepto passa a ser estatísticamente significativo a um nível de 95% e todas as variáveis também são significativas, mantndo a interpretação acima. A diferença entre os modelos é que agora consideramos a correlação dos membros da mesma família, e temos um desvio padrão associado ao fator aleatório de família *14.06325*.

```r
library(nlme)
## ajuste do modelo misto
mod3 = lme (Weight ~Height + Age + Sex, random=~ 1 |FamilyID, data=family)

summary(mod3)
```

```
## Linear mixed-effects model fit by REML
##  Data: family
##        AIC      BIC    logLik
##   651.1885 664.4167 -319.5943
##
## Random effects:
##  Formula: ~1 | FamilyID
##         (Intercept) Residual
## StdDev:    14.06325 21.30134
##
```

```
## Fixed effects: Weight ~ Height + Age + Sex
##                  Value Std.Error DF   t-value p-value
## (Intercept) -136.98742  75.37559 50 -1.817398  0.0752
## Height         3.80575   1.13282 50  3.359528  0.0015
## Age            0.67120   0.17223 50  3.896990  0.0003
## Sex1          20.96406   8.88688 50  2.358990  0.0223
##  Correlation:
##        (Intr) Height Age
## Height -0.994
## Age    -0.247  0.165
## Sex1    0.756 -0.791 -0.075
##
## Standardized Within-Group Residuals:
##         Min          Q1         Med          Q3         Max
## -2.11753193 -0.56529380 -0.07806256  0.46585409  3.58329184
##
## Number of Observations: 71
## Number of Groups: 18
```

A tabela abaixo mostra os valores preditos para os efeitos aleatórios e as estimativas dos parâmetros. Vemos que cada uma das 18 famílias possui um valor predito de intercepto diferente e os demais parâmetros têm a mesma estimativa para todas as familias, pois são fixos.

```r
coef(mod3)
```

```
##    (Intercept)   Height       Age     Sex1
## 1    -117.2095 3.805745 0.6711957 20.96406
## 2    -139.5046 3.805745 0.6711957 20.96406
## 3    -121.8201 3.805745 0.6711957 20.96406
## 4    -158.8300 3.805745 0.6711957 20.96406
## 5    -149.5370 3.805745 0.6711957 20.96406
## 6    -130.8892 3.805745 0.6711957 20.96406
## 7    -141.5041 3.805745 0.6711957 20.96406
## 8    -141.2162 3.805745 0.6711957 20.96406
## 9    -133.0409 3.805745 0.6711957 20.96406
## 10   -141.2518 3.805745 0.6711957 20.96406
## 11   -140.1266 3.805745 0.6711957 20.96406
## 12   -136.7304 3.805745 0.6711957 20.96406
## 13   -144.1791 3.805745 0.6711957 20.96406
## 14   -118.9669 3.805745 0.6711957 20.96406
## 15   -124.8528 3.805745 0.6711957 20.96406
## 16   -145.9798 3.805745 0.6711957 20.96406
## 17   -144.0644 3.805745 0.6711957 20.96406
## 18   -136.0699 3.805745 0.6711957 20.96406
```