Exercício 1

Considere um modelo balanceado com intercepto aleatório obtenha a estimativa de máxima verossimilhança da componente d.

Resposta:

O modelo balanceado com intercepto aleatório em notação matricial pode ser excrito como:

$$\boldsymbol{y_i} = \boldsymbol{X\beta} + \boldsymbol{1_n b_i} + \boldsymbol{\varepsilon_i}, \quad i = 1, 2, \ldots, N,$$

onde

- $\boldsymbol{y_i}$ é um vetor $n \times 1$ de variáveis respostas para o i-ésimo sujeito.
- $\boldsymbol{X}$ é uma matriz $n \times m$ de variáveis explicativas.
- $\boldsymbol{\beta}$ é um vetor $m \times 1$ de efeitos fixos.
- $\boldsymbol{\varepsilon_i}$ é um vetor $n \times 1$ de erros independentes, e assume-se que $\boldsymbol{\varepsilon_i} \sim N(0, \sigma^2 I_n)$.
- b_i é o efeito aleatório, e assume-se que $b_i \sim N(0, \sigma^2 d)$.
- $\boldsymbol{1_n}$ é um vetor unitário de dimensão $n \times 1$.
- σ^2 é a variância dentro do sujeito e d é a variância escalada do efeito aleatório.

O modelo também pode ser escrito como

$$\boldsymbol{y_i} \sim N(\boldsymbol{X\beta}, \sigma^2(\boldsymbol{I_n} + d\boldsymbol{1_n 1'_n})), \quad i = 1, 2, \ldots, N.$$

A log-verossimilhança do modelo, ignorando a constante é

$$\ell(\boldsymbol{\theta}) = -\frac{1}{2}\left\{Nn\log(\sigma^2) + \sum_{i=1}^{N}\log|\boldsymbol{I_n} + d\boldsymbol{1_n 1'_n}| + \sigma^{-2}\sum_{i=1}^{N}\boldsymbol{e'_i}(\boldsymbol{I_n} + d\boldsymbol{1_n 1'_n})^{-1}\boldsymbol{e_i}\right\},$$

onde $\boldsymbol{e_i} = \boldsymbol{y_i} - \boldsymbol{X\beta}$.

Utilizando as fórmulas da redução de dimensão, temos que:

$$\begin{aligned}
(\boldsymbol{I_n} + d\boldsymbol{1_n 1'_n})^{-1} &= \boldsymbol{I_n} - \boldsymbol{1_n}\left(\frac{1}{d} + \boldsymbol{1'_n 1_n}\right)^{-1}\boldsymbol{1'_n} \\
&= \boldsymbol{I_n} - \boldsymbol{1_n}\left(\frac{1}{d} + n\right)^{-1}\boldsymbol{1'_n} \\
&= \boldsymbol{I_n} - \frac{d}{1+nd}\boldsymbol{1_n 1'_n}.
\end{aligned}$$

$$|\boldsymbol{I_n} + d\boldsymbol{1_n 1'_n}| = |1 + dn| = 1 + dn.$$

Assim, a log-verossimilhança fica

$$\begin{aligned}
\ell(\boldsymbol{\theta}) &= -\frac{1}{2}\left\{Nn\log(\sigma^2) + \sum_{i=1}^{N}\log(1+dn) + \sigma^{-2}\sum_{i=1}^{N}\boldsymbol{e'_i}\left(\boldsymbol{I_n} - \frac{d}{1+nd}\boldsymbol{1_n 1'_n}\right)\boldsymbol{e_i}\right\} \\
&= -\frac{1}{2}\left\{Nn\log(\sigma^2) + N\log(1+dn) + \sigma^{-2}\sum_{i=1}^{N}\left[\boldsymbol{e'_i e_i} - \frac{d}{1+nd}\boldsymbol{e'_i 1_n 1'_n e_i}\right]\right\} \\
&= -\frac{1}{2}\left\{Nn\log(\sigma^2) + N\log(1+dn) + \sigma^{-2}\sum_{i=1}^{N}\left[\boldsymbol{e'_i e_i} - \frac{d}{1+nd}(\boldsymbol{e'_i 1_n})^2\right]\right\}.
\end{aligned}$$

A derivada parcial da log-verossimilhança em relação a d é:

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial d} = -\frac{1}{2}\left\{\frac{Nn(1+dn)\sigma^2 - \sum_{i=1}^{N}(\boldsymbol{e_i'1_n})^2}{\sigma^2(1+dn)^2}\right\}.$$

A derivada acima é igual a zero quando

$$d = \frac{\sum_{i=1}^{N}(\boldsymbol{e_i'1_n})^2 - Nn\sigma^2}{\sigma^2 Nn^2} = \frac{\sum_{i=1}^{N}(\boldsymbol{e_i'1_n})^2}{\sigma^2 Nn^2} - \frac{1}{n},$$

assim a estimativa de máxima verossimilhança da componente d é:

$$\hat{d}_{ML} = \frac{\sum_{i=1}^{N}(\boldsymbol{\hat{e}_i'1_n})^2}{\hat{\sigma}^2_{ML} Nn^2} - \frac{1}{n},$$

onde $\boldsymbol{\hat{e}_i = y_i - X\hat{\beta}_{ML}}$, e segundo Demidenko (2013), para dados balanceados $\hat{\beta}_{ML} = \hat{\beta}_{OLS} = \boldsymbol{(X'X)^{-1}X'\overline{y}}$, assim, o termo $\boldsymbol{\hat{e}_i'1_n}$ pode ser escrito como

$$\begin{aligned}
\boldsymbol{\hat{e}_i'1_n} &= \boldsymbol{(y_i - X\hat{\beta}_{OLS})'1_n} \\
&= \boldsymbol{(y_i - X(X'X)^{-1}X'\overline{y})'1_n} \\
&= \boldsymbol{y_i'1_n - \overline{y}'X(X'X)^{-1}X'1_n} \\
&= \boldsymbol{y_i'1_n - \overline{y}'1_n} \\
&= n(\overline{y}_i - \overline{\overline{y}}),
\end{aligned}$$

onde $\overline{\overline{y}} = \boldsymbol{\overline{y}'1_n}/n$, $\overline{y}_i = \boldsymbol{y_i'1_n}/n$, e a prova de que $\boldsymbol{X(X'X)^{-1}X'1_n = 1_n}$ se encontra em (Demidenko 2013, p.68).

Portanto, a estimativa de máxima verossimilhança da componente d em um modelo balanceado com intercepto aleatório é

$$\hat{d}_{ML} = \frac{\sum_{i=1}^{N}(\overline{y}_i - \overline{\overline{y}})^2}{\hat{\sigma}^2_{ML} N} - \frac{1}{n},$$

Exercício 2

Mostre que o termo do log-verossimilhança restrita

$$f(N) = -\frac{1}{2}(-m \ln \sigma^2 + \ln |\sum_{i=1}^{N} X_i' V_i X_i|)$$

é de ordem de $\ln N$. Considere: $n_i = n$, X_i e $Z_i = Z$.

Resposta:

Queremos mostrar que $\frac{f(N)}{\ln N} < M$, com $M \in \mathbb{R}$ e $\forall N > N_0$.

Temos que

$$\ln |\sum_{i=1}^{N} X_i' V_i X_i| = \ln |\frac{N}{N} \sum_{i=1}^{N} X_i' V_i X_i| = \ln(N^m |\frac{1}{N} \sum_{i=1}^{N} X_i' V_i X_i|) = \ln |\frac{N}{N} X' V X| + m \ln N$$

Logo,

$$f(N) = -\frac{1}{2}(-m \ln \sigma^2 + \ln |X' V X| + m \ln N)$$

só depende de $\ln N$ quando $\lim_{N \to \infty}$. Logo a seguinte inequação é válida pois existe um M real tal que

$$\frac{-\frac{1}{2}(-m \ln \sigma^2 + \ln |X' V X| + m \ln N)}{\ln N} < M$$

para todo $N > N_0$.

```r
## importando os pacotes
library(glmmTMB)
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --

## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.0.4     v dplyr   1.0.2
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.0

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```r
library(knitr)
library(ggplot2)
library(nlme)
```

```
##
## Attaching package: 'nlme'

## The following object is masked from 'package:dplyr':
##
##     collapse
```

```r
## importando os dados
data(Owls, package = "glmmTMB")
```

Exercício 3

Um estudo com filhotes de coruja foi conduzido com câmeras e microfones para analisar a negociação entre os irmãos definida como segue. Usando a filmagem gravada foram registrados durante intervalos de 30 segundos a cada 15 minutos o número de chamadas feitas por todos os descendentes na ausência dos pais. Para cada visita de um dos pais foi registrado o número de chamadas dos 15 minutos anteriores divido pelo número de filhotes em cada ninho. Os dados estão no arquivo Owls. As variáveis explicativas são o sexo dos pais, tratamento de alimentos, e o tempo de chegada do pai. O tratamento de alimentos foram dois, a metade dos ninhos foram dadas presas extras ("food-satiated"), e na outra metade as presas (remanescentes) foram removidas. ("fooddeprived"). As medições ocorreram em duas noites em cada caso, e o tratamento de alimentos foi trocado na segunda noite. (Fonte Roulin and Bersier, 2007) A negociação entre irmãos (NegPerChick) pode ser transformada por log10(Y + 1)= LogNeg para ser modelada.

a) Selecione um modelo misto que considere adequado para determinar os fatores que podem influenciar a variável de interesse.

b) Descreva o modelo formalmente especificando as matrizes associadas aos efeitos fixos e aleatórios.

c) Realize as interpretações que pertinentes.

item a)

Análise descritiva Ao todo temos 599 observações de 27 ninhos. As características observadas ao longo do experimento em todos os ninhos foram tratamento alimentar, sexo do pai, tempo de chegada do pai e tamanho da ninhada.

A característica de interesse do estudo é o número de negociação entre irmãos.

Como o número de filhotes de cada ninhada é diferente, para conseguirmos comparar as negociações entre ninhadas, dividimos o número de negociações observadas no intervalo de tempo pelo número total de filhotes do ninho. Além disso, conforme instrução do enuncuado, transformamos a variável resposta NegPerChick (negociações por filhote) adicionando 1 e aplicando logaritmo de base 10.

\$ LogNeg = log(NegPerChick + 1) \$

Vemos que cada ninho tem pelo menos 4 observações, desta forma concluimos que os sujeitos não são independentes e portanto, é interessante mensurar a correlação entre as observações do mesmo ninho.

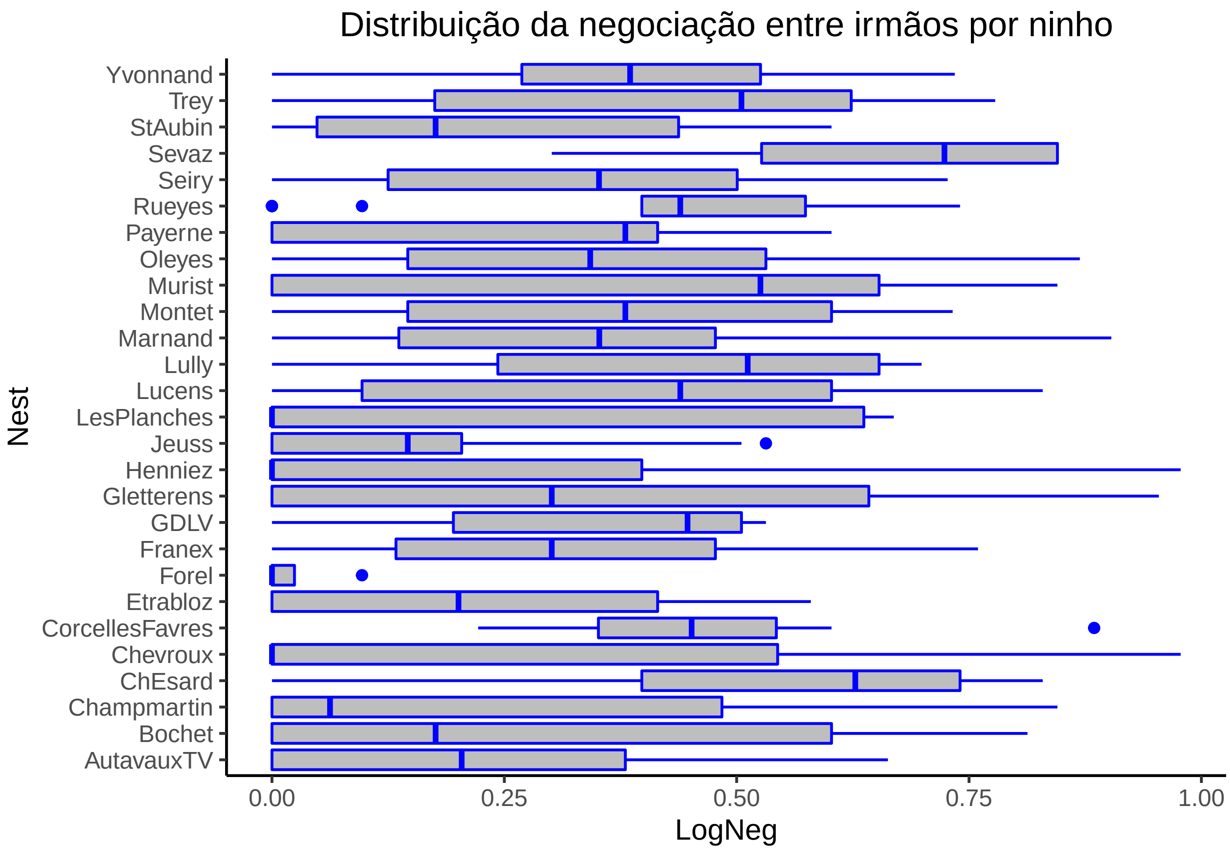
Além disso, os ninhos estudados são uma amostra de todos os ninhos existentes podemos generalizar o modelo para que ele seja capaz de descrever outros ninhos fora da amostra.

Também é importante destacar que ao considerar os ninhos como efeitos fixos, gastaríamos muitos graus de liberdade.

Nest	contagem
Forel	4
Sevaz	4
Chevroux	10
GDLV	10
CorcellesFavres	12
Henniez	13
Gletterens	15
LesPlanches	17
Lully	17
Rueyes	17
Jeuss	19
Trey	19
ChEsard	20
Bochet	23
StAubin	23
Murist	24
Payerne	25
Franex	26
Seiry	26
Marnand	27
AutavauxTV	28
Lucens	29
Champmartin	30
Etrabloz	34
Yvonnand	34
Montet	41
Oleyes	52

Variável resposta

Abaixo fica claro que a distribuição do log das negociações entre irmãos varia entre os ninhos.

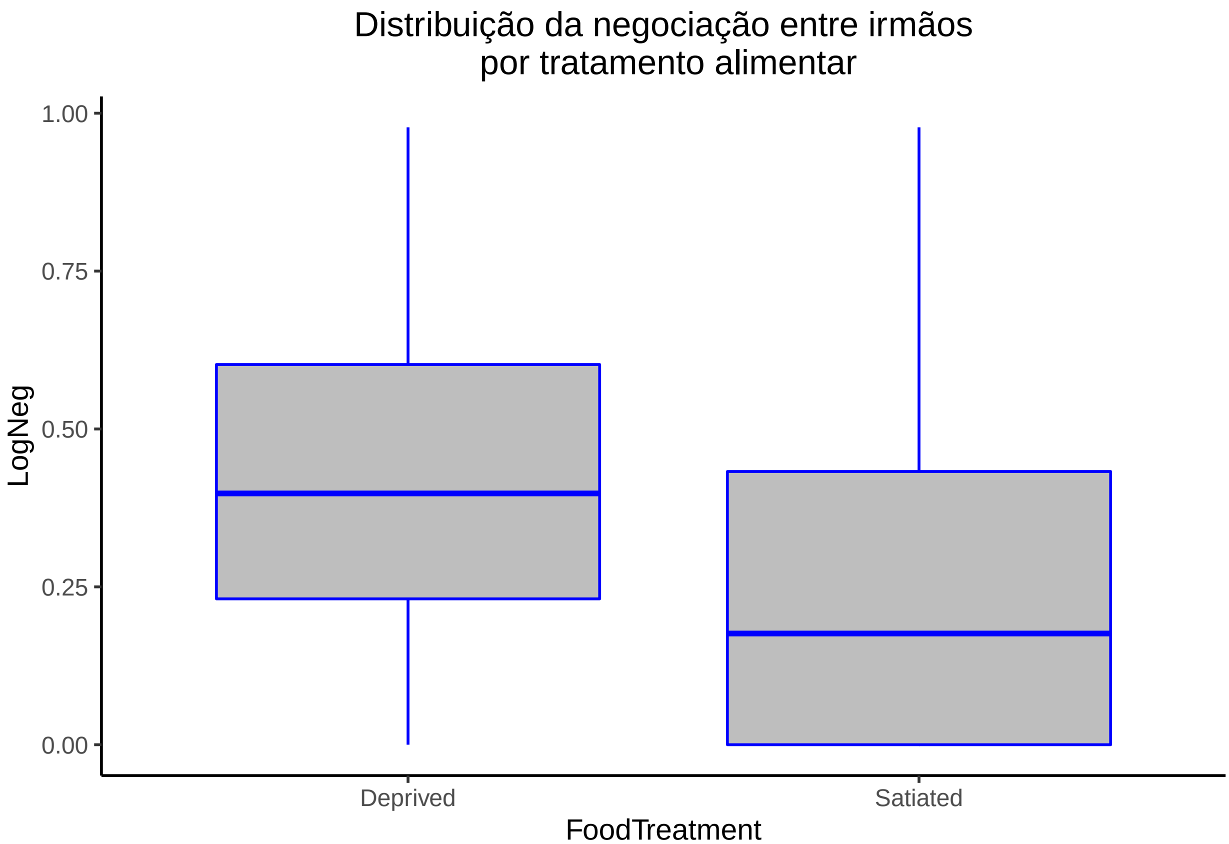


Por todo o exposto vamos construir um modelo de efeitos mistos.

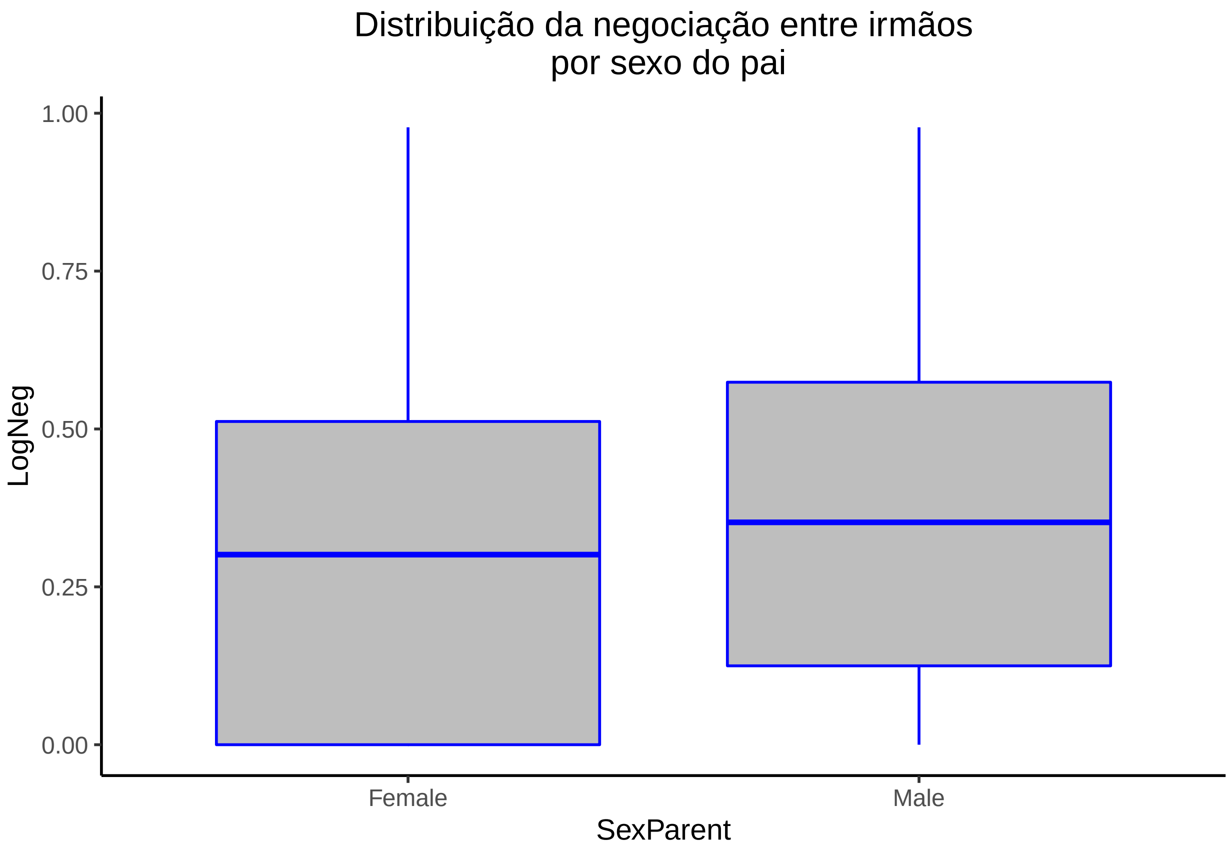
Variáveis explicativas

Quanto às variáveis explicativas do modelo, temos disponíveis: tratamento alimentar, sexo do pai, e tempo de chegada do pai. Os demais dados disponíveis são referentes à resposta e/ou estão sendo usados para calculá-la.

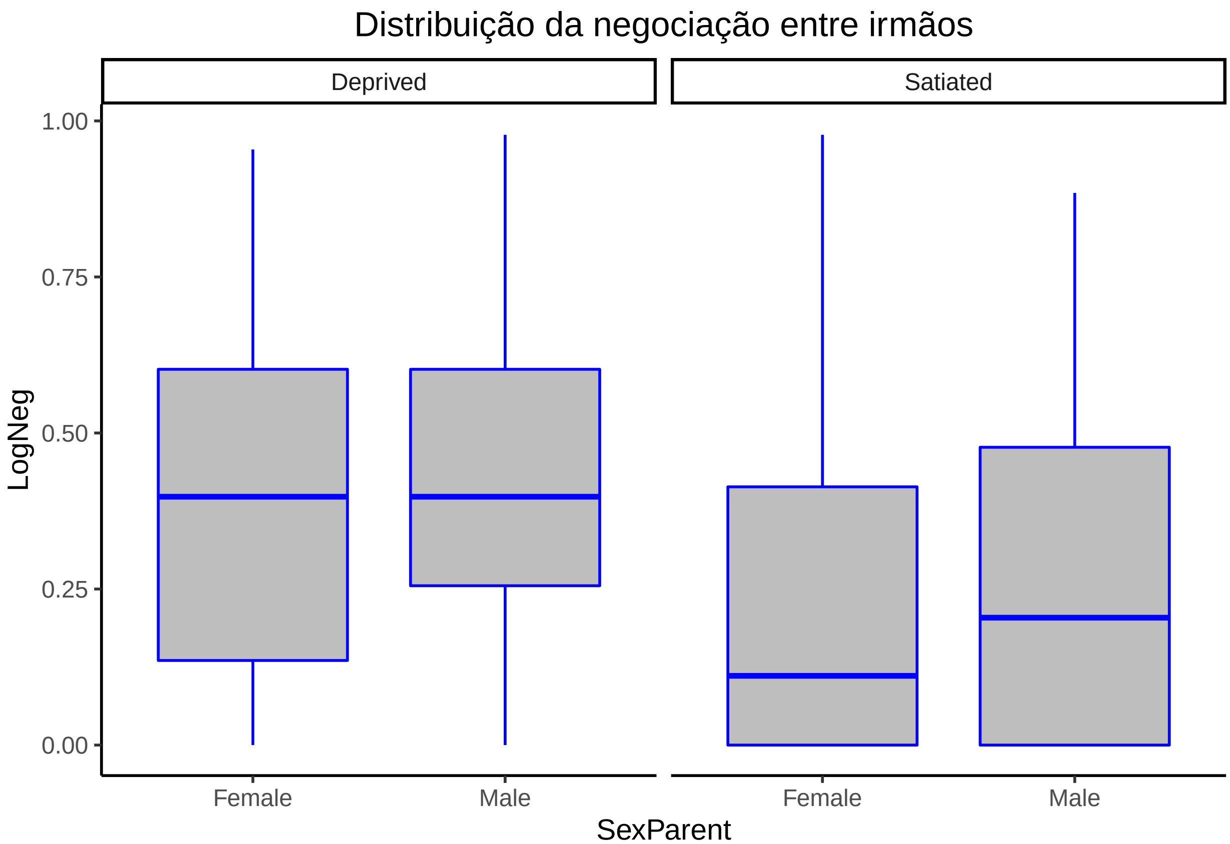
No gráfico de abaixo vemos que a distribuição do log da taxa de negociação por filhote é maior para o tratamento alimentar *Satiated*.



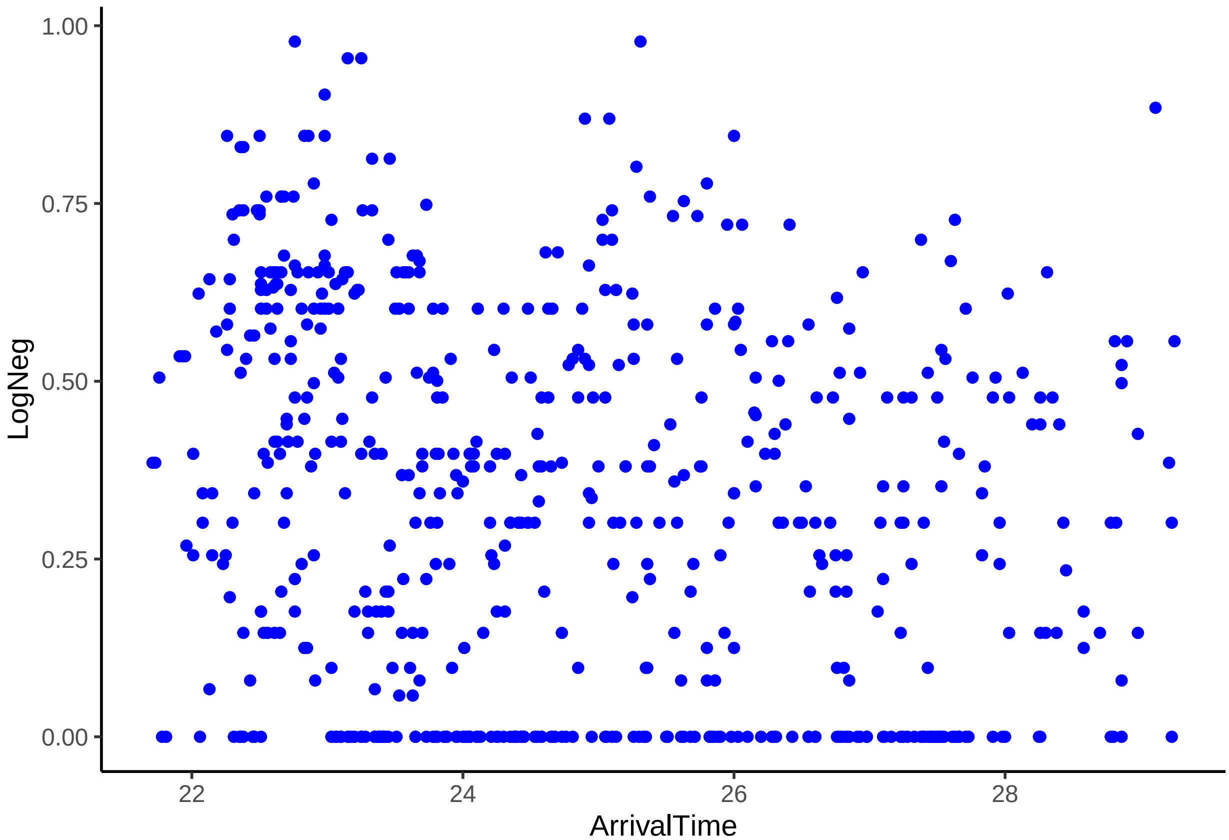
Avaliando agora o sexo dos pais, pela análise gráfica não identificamos diferença na negociação entre irmãos.

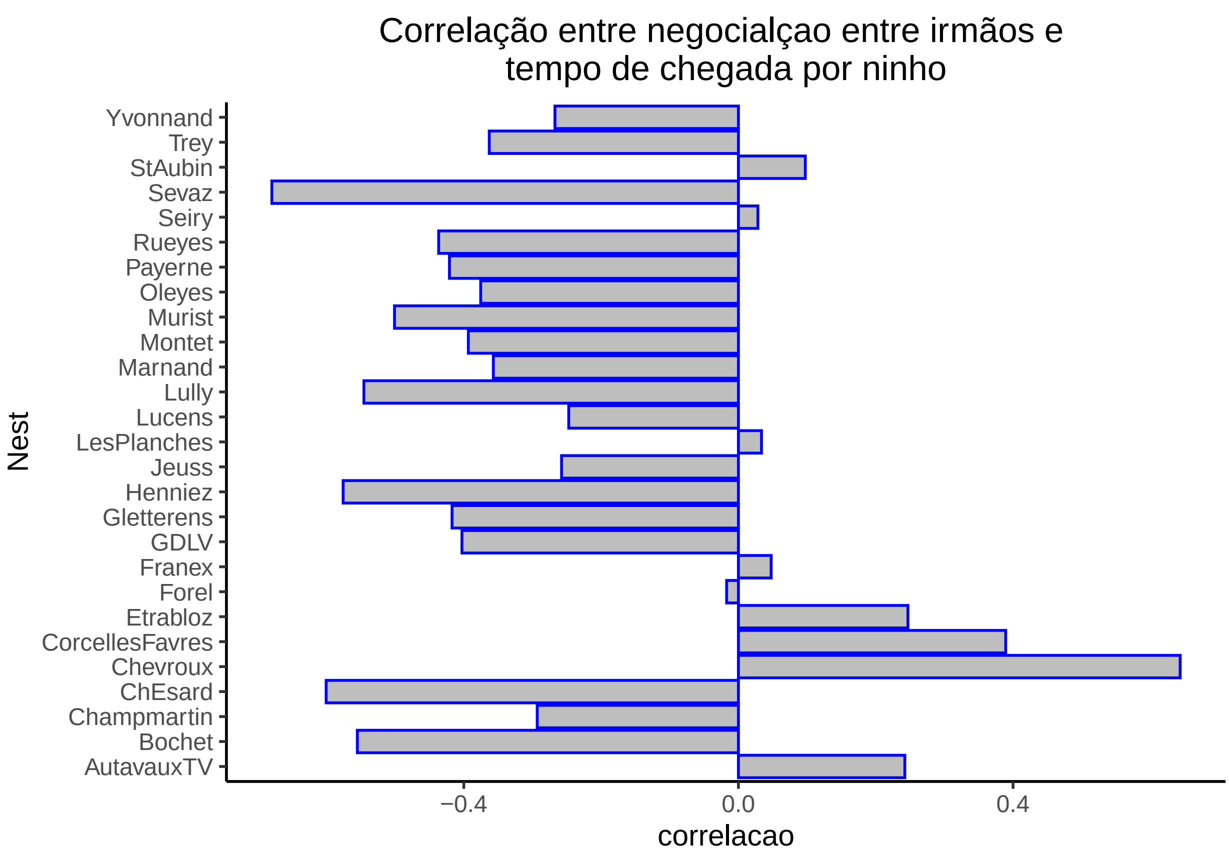


Visualizando as duas variáveis simultaneamente não encontramos evidência de interação entre elas, pois o tratamento alimentar *Deprived* mostra maior negociação entre irmãos para ambos os sexos, assim como o tratamento *Satiated* que mantém próximas as distribuições de negociação para ambos os sexos.

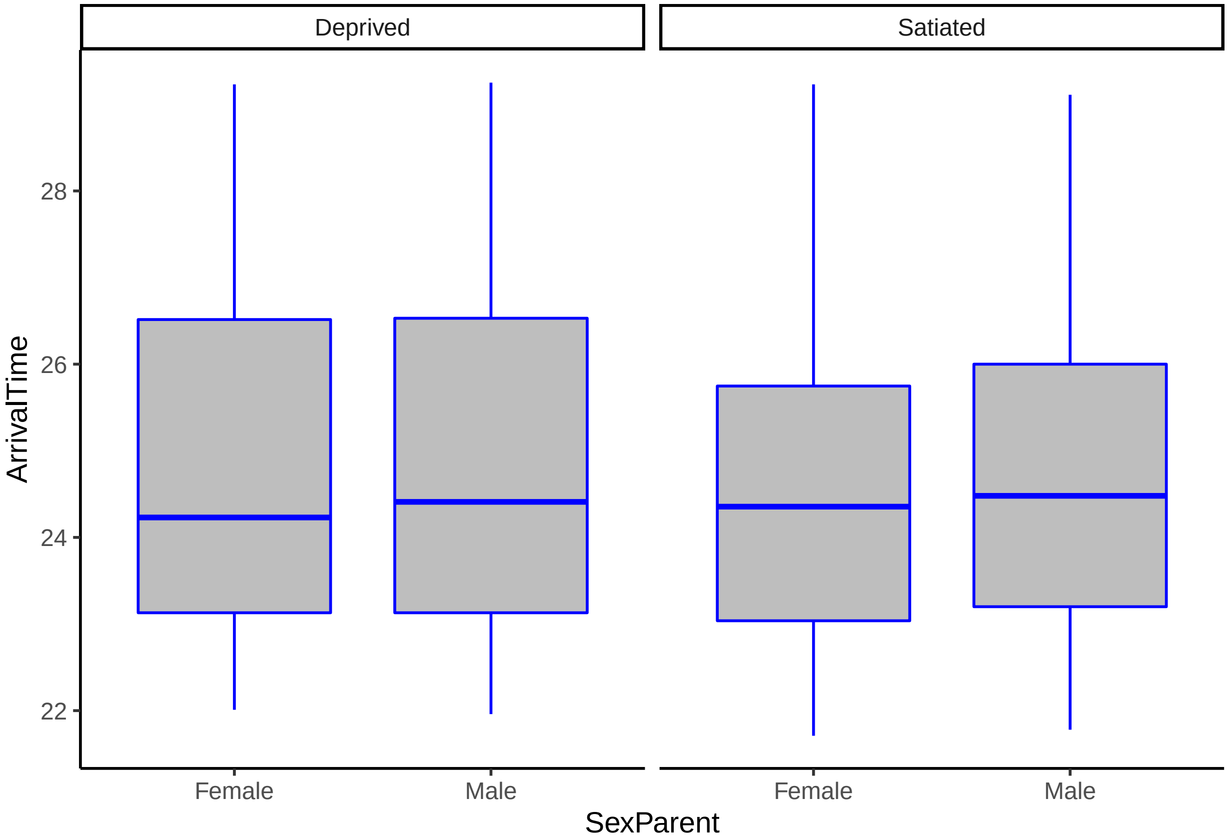


Outra variáves explicativa disponível é o tempo de chegada que aparentemente não tem correlação com a variável resposta. Porém, se avaliarmos a correlação entre essas variáveis para cada um dos ninhos, vemos que o cenário muda.





Procurando interação entre as variáveis, avaliamos o tempo de chegada por tratamento alimentar e sexo dos pais e não encontramos evidências gráficas de interação.



Ajuste do modelo Após ajustar o modelo, vemos que a variável *SexParent* não é sifnificativa.

```
negociacao <- lme(LogNeg~SexParent+FoodTreatment+ArrivalTime, random= ~1| Nest, data = Owls)
summary(negociacao)
```

```
## Linear mixed-effects model fit by REML
##   Data: Owls
##        AIC      BIC    logLik
##   22.03994 48.37131 -5.019969
##
## Random effects:
##  Formula: ~1 | Nest
##         (Intercept)  Residual
## StdDev:  0.09271906 0.2318242
##
## Fixed effects:  LogNeg ~ SexParent + FoodTreatment + ArrivalTime
##                            Value  Std.Error  DF   t-value p-value
## (Intercept)            1.1748045 0.12917288 569  9.094823  0.0000
## SexParentMale          0.0202573 0.02130399 569  0.950870  0.3421
## FoodTreatmentSatiated -0.1735870 0.02001784 569 -8.671614  0.0000
## ArrivalTime           -0.0312510 0.00512116 569 -6.102322  0.0000
##  Correlation:
##                       (Intr) SxPrnM FdTrtS
## SexParentMale         -0.055
```

```
## FoodTreatmentSatiated -0.115  0.068
## ArrivalTime           -0.979 -0.051  0.035
##
## Standardized Within-Group Residuals:
##         Min          Q1         Med          Q3         Max
## -2.23169999 -0.77923134 -0.08570653  0.71095687  3.29621690
##
## Number of Observations: 599
## Number of Groups: 27
```

Outro teste que mostra que a variável sexo dos pais pode ser removida sem prejuízo para o modelo é o teste de verossimilhança, porém, para usá-lo precisamos que o método de estimativa dos parâmetros seja máxima verossimilhança.

```
com_SexParent <- lme(LogNeg~SexParent+FoodTreatment+ArrivalTime, random= ~1| Nest, method = "ML", data =

sem_SexParent <- lme(LogNeg~FoodTreatment+ArrivalTime, random= ~1| Nest, method = "ML", data = Owls)

anova(com_SexParent, sem_SexParent)
```

```
##               Model df       AIC      BIC   logLik   Test   L.Ratio p-value
## com_SexParent     1  6 -4.476920 21.89465 8.238460
## sem_SexParent     2  5 -5.545145 16.43116 7.772572 1 vs 2 0.9317755  0.3344
```

O **modelo final** escolhido para explicar o log da taxa de negociações por filhote é um modelo misto, cujos parâmetros são estimados pelo método da máxima verossimilhança restrita porque as estimativas são mais próximas do valor real.

As variáveis escolhidas para este modelo são tratamento alimentar e tempo de chegada.

```
sem_SexParent <- lme(LogNeg~FoodTreatment+ArrivalTime, random= ~1| Nest, data = Owls)
sem_SexParent
```

```
## Linear mixed-effects model fit by REML
##   Data: Owls
##   Log-restricted-likelihood: -2.536915
##   Fixed: LogNeg ~ FoodTreatment + ArrivalTime
##           (Intercept) FoodTreatmentSatiated           ArrivalTime
##            1.18213859           -0.17507539           -0.03102135
##
## Random effects:
##  Formula: ~1 | Nest
##         (Intercept)  Residual
## StdDev:  0.09468769 0.2316398
##
## Number of Observations: 599
## Number of Groups: 27
```

item b)

item c)