

Questão D

Guilherme Pazian

15 de junho de 2017

```
dados <- read.table("recrutas.dat")
colnames(dados) <- c("Hab. de nadar", "Local", "Faixa etária", "Sexo", "Nº de infecções")

media <- mean(dados$`Nº de infecções`)
variancia <- var(dados$`Nº de infecções`)
```

Observamos, nos modelos anteriores, que nenhum deles foi adequado para os dados, uma vez que as análises de resíduo indicaram um mal ajuste dos modelos do tipo Poisson. Possivelmente, um dos aspectos à influenciar esse mal ajuste dos modelos Poisson é a característica da variável resposta, a qual apresenta média igual a 1,3868 e variância igual a 5,4688, ou seja, a variância observada na variável resposta é quase quatro vezes o valor observado para a média da variável resposta. Uma das imposições do modelo de Poisson é que a média da variável resposta tem a mesma magnitude do valor de sua respectiva variância, porém isso não é observado nos dados referente à variável resposta, de maneira que temos uma situação de superdispersão, ou seja, a variável resposta apresenta variância maior do que aquela imposta pelo modelo probabilístico.

Dadas as observações feitas quanto ao modelo de Poisson, um modelo de regressão construído a partir de uma distribuição binomial negativa talvez seja mais adequado (veja Azevedo 2017) para os dados em questão.

Temos, portando o modelo:

Seja Y_i o número de infecções de ouvido diagnosticadas pelo i -ésimo indivíduo.

$$Y_i \stackrel{ind.}{\sim} BN(\mu_i, \phi)$$

$$\begin{aligned} \ln(\mu_i) = & \mu + \alpha O_i + \gamma P_i + \beta_1 "20 - 25"_i + \beta_2 "25 - 29"_i + \delta M_i + \\ & (\alpha\gamma)P_i O_i + (\alpha\beta_1)"20 - 25"_i O_i + (\alpha\beta_2)"25 - 29"_i O_i + (\alpha\delta)M_i O_i + (\beta_1\gamma)"20 - 25"_i P_i + \\ & (\beta_2\gamma)"25 - 29"_i P_i + (\delta\gamma)M_i P_i + (\beta_1\delta)M_i "20 - 25"_i + (\beta_2\delta)M_i "25 - 29"_i + \\ & (\alpha\beta_1\gamma)"20 - 25"_i P_i O_i + (\alpha\beta_2\gamma)"25 - 29"_i P_i O_i + (\alpha\delta\gamma)M_i P_i O_i + (\beta_1\delta\gamma)"20 - 25"_i P_i M_i + \\ & (\beta_2\delta\gamma)"25 - 29"_i P_i M_i + (\alpha\beta_1\delta\gamma)"20 - 25"_i P_i O_i M_i + (\alpha\beta_2\delta\gamma)"25 - 29"_i P_i O_i M_i \end{aligned}$$

$O_i = 1$ se o i -ésimo indivíduo nada ocasionalmente e $O_i = 0$ caso contrário.

$P_i = 1$ se o i -ésimo indivíduo nada na piscina e $P_i = 0$ caso contrário.

$"20 - 25"_i = 1$ se o i -ésimo indivíduo pertence à faixa etária 20-25 e $"20 - 25"_i = 0$ caso contrário.

$"25 - 29"_i = 1$ se o i -ésimo indivíduo pertence à faixa etária 25-29 e $"25 - 29"_i = 0$ caso contrário.

$M_i = 1$ se o i -ésimo indivíduo é do sexo masculino e $M_i = 0$ caso contrário.

```
#pacotes caio Usados no exemplo "Número de acidentes" ( que tem a binomial negativa)
library(xtable)
library(plotrix)
library(plyr)
library(lattice)
library(HistogramTools)
library(MASS)
```

```
fit.model <- glm.nb(`Nº de infecções`~ `Hab. de nadar` * `Local` * `Faixa etária` * `Sexo`,link=log,data=
```

```
source("codigos/gerador_tbl_coef.R")
#Gerando a tabela dos coeficientes para o modelo Binomial negativo
#gerador_tbl_coef(fit.model)
```

	Estimativa	Erro Padrão	Estatística Z	P-valor
(Intercept)	0,3567	0,3975	0,8973	0,3696
‘Hab. de nadar‘Occas	0,1133	0,5482	0,2067	0,8362
LocalNonBeach	0,0488	0,8383	0,0582	0,9536
‘Faixa etária‘20-24	-1,0498	0,7136	-1,4712	0,1413
‘Faixa etária‘25-29	-0,3567	0,7594	-0,4697	0,6386
SexoMale	-0,5798	0,5442	-1,0655	0,2866
‘Hab. de nadar‘Occas:LocalNonBeach	0,7340	1,1378	0,6451	0,5189
‘Hab. de nadar‘Occas:‘Faixa etária‘20-24	0,3285	0,9811	0,3348	0,7378
‘Hab. de nadar‘Occas:‘Faixa etária‘25-29	0,7517	0,9802	0,7668	0,4432
LocalNonBeach:‘Faixa etária‘20-24	0,5390	1,1473	0,4698	0,6385
LocalNonBeach:‘Faixa etária‘25-29	-1,4351	1,5810	-0,9077	0,3640
‘Hab. de nadar‘Occas:SexoMale	0,5153	0,7400	0,6964	0,4862
LocalNonBeach:SexoMale	0,5153	0,9541	0,5401	0,5892
‘Faixa etária‘20-24:SexoMale	1,2730	1,0031	1,2690	0,2044
‘Faixa etária‘25-29:SexoMale	-0,5188	1,0078	-0,5148	0,6067
‘Hab. de nadar‘Occas:LocalNonBeach:‘Faixa etária‘20-24	-0,6957	1,5553	-0,4473	0,6546
‘Hab. de nadar‘Occas:LocalNonBeach:‘Faixa etária‘25-29	-0,2127	1,9903	-0,1069	0,9149
‘Hab. de nadar‘Occas:LocalNonBeach:SexoMale	-0,9180	1,2941	-0,7094	0,4781
‘Hab. de nadar‘Occas:‘Faixa etária‘20-24:SexoMale	-2,9030	1,6687	-1,7397	0,0819
‘Hab. de nadar‘Occas:‘Faixa etária‘25-29:SexoMale	-0,3462	1,2922	-0,2679	0,7888
LocalNonBeach:‘Faixa etária‘20-24:SexoMale	-1,1831	1,4431	-0,8199	0,4123
LocalNonBeach:‘Faixa etária‘25-29:SexoMale	1,9696	1,8226	1,0807	0,2798
‘Hab. de nadar‘Occas:LocalNonBeach:‘Faixa etária‘20-24:SexoMale	3,9762	2,1745	1,8286	0,0675
‘Hab. de nadar‘Occas:LocalNonBeach:‘Faixa etária‘25-29:SexoMale	0,6356	2,3180	0,2742	0,7839

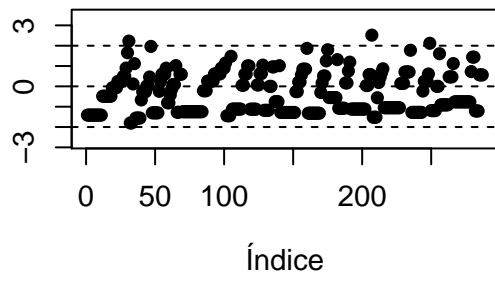
```
source("codigos/diag_nbin.R")
```

```
source("codigos/envel_nbin.R")
```

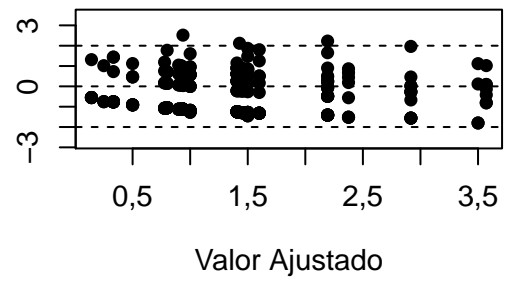
```
desvio <- fit.model$deviance
gl_desvio <- fit.model$df.residual
p_valor_desvio <- 1-pchisq(desvio,gl_desvio)
```

```
diagnbin(fit.model)
```

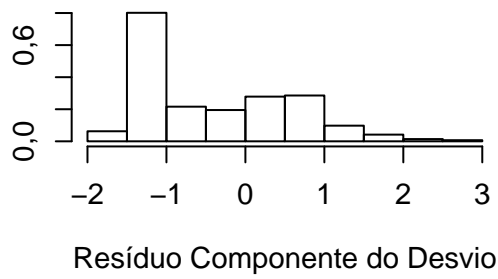
Resíduo Componente do Desvio



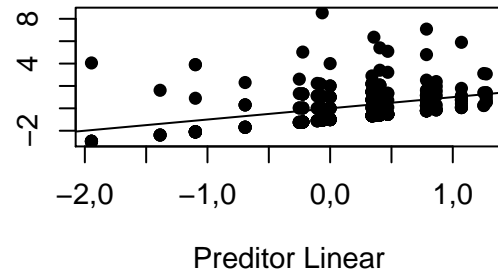
Resíduo Componente do Desvio



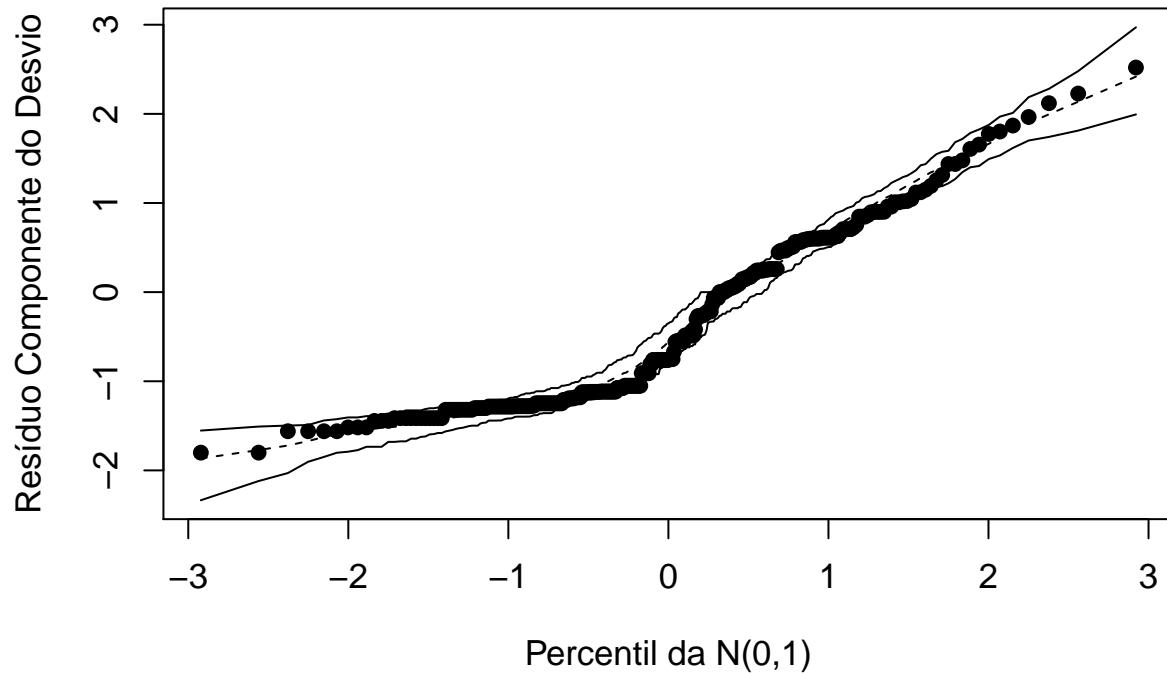
densidade



Variável z



```
ligacaonbin <- "log"
envelnbin(fit.model)
```



Note pela figura XX(diagnostico)XX, que os dois gráficos da parte superior indicam poucos pontos fora dos limites -2 e 2, e estes pontos, embora estejam fora, estão bem próximos dos limites aceitáveis, portanto observamos um comportamento similar ao esperado para um bom ajuste do modelo. Observamos pela figura XX(envelopes)XX que os resíduos estão todos contidos dentro das bandas de confiança do envelope e não se observa nenhuma tendência não esperada no comportamento dos resíduos.

Podemos também, com base no valor da função desvio e a partir da distribuição qui-quadrado com os respectivos graus de liberdade (263 neste caso) testar a hipótese de um bom ajuste do modelo em questão (veja Azevedo 2017), o valor observado para a função desvio foi 269,8345, e o p-valor para o testes foi 0,3728, portanto, à um nível de significância de 5%, não temos evidências para rejeitar a hipótese de um bom ajuste do modelo, ou seja, a função desvio indica um bom ajuste.

Dadas as observações, temos uma boa indicação de bom ajuste do modelo log-linear binomial-negativo.