

# modelo poisson

*Guilherme Pazian*

*15 de junho de 2017*

```
dados <- read.table("/home/hugo/Documents/Discretos/discretos_2017/trabalho/TrabalhoME714/Parte02/recre
colnames(dados) <- c("Hab. de nadar", "Local", "Faixa etária", "Sexo", "Nº de infecções")

fit_max <- glm(`Nº de infecções` ~ `Hab. de nadar` * `Local` * `Faixa etária` * `Sexo`, family=poisson(link="log"), data = dados)

fit_min <- glm(`Nº de infecções` ~ 1, family=poisson(link="log"), data = dados)

step_min_to_max <- step(fit_min, scope = list(upper=fit_max), direction = "both")

## Start:  AIC=1198.9
## `Nº de infecções` ~ 1
##
##               Df Deviance    AIC
## + `Hab. de nadar`  1   789.81 1166.2
## + Local           1   800.36 1176.8
## + `Faixa etária`  2   816.03 1194.4
## <none>              824.51 1198.9
## + Sexo            1   824.07 1200.5
##
## Step:  AIC=1166.2
## `Nº de infecções` ~ `Hab. de nadar`
##
##               Df Deviance    AIC
## + Local           1   764.65 1143.0
## + `Faixa etária`  2   781.26 1161.7
## <none>              789.81 1166.2
## + Sexo            1   789.20 1167.6
## - `Hab. de nadar`  1   824.51 1198.9
##
## Step:  AIC=1143.04
## `Nº de infecções` ~ `Hab. de nadar` + Local
##
##               Df Deviance    AIC
## + `Faixa etária`  2   756.07 1138.5
## <none>              764.65 1143.0
## + `Hab. de nadar`:Local  1   763.00 1143.4
## + Sexo            1   764.65 1145.0
## - Local           1   789.81 1166.2
## - `Hab. de nadar`  1   800.36 1176.8
##
## Step:  AIC=1138.46
## `Nº de infecções` ~ `Hab. de nadar` + Local + `Faixa etária`
##
##               Df Deviance    AIC
## + `Hab. de nadar`:`Faixa etária`  2   750.58 1137.0
## + Local:`Faixa etária`           2   751.90 1138.3
## <none>                             756.07 1138.5
```

```

## + `Hab. de nadar`:Local          1  754.37 1138.8
## + Sexo                          1  755.43 1139.8
## - `Faixa etária`                2  764.65 1143.0
## - Local                          1  781.26 1161.7
## - `Hab. de nadar`                1  791.77 1172.2
##
## Step: AIC=1136.97
## `Nº de infecções` ~ `Hab. de nadar` + Local + `Faixa etária` +
##   `Hab. de nadar`:`Faixa etária`
##
##                                Df Deviance    AIC
## + `Hab. de nadar`:Local          1  747.92 1136.3
## + Local:`Faixa etária`           2  746.35 1136.8
## <none>                           750.58 1137.0
## + Sexo                          1  749.97 1138.4
## - `Hab. de nadar`:`Faixa etária` 2  756.07 1138.5
## - Local                          1  776.11 1160.5
##
## Step: AIC=1136.31
## `Nº de infecções` ~ `Hab. de nadar` + Local + `Faixa etária` +
##   `Hab. de nadar`:`Faixa etária` + `Hab. de nadar`:Local
##
##                                Df Deviance    AIC
## + Local:`Faixa etária`           2  743.78 1136.2
## <none>                           747.92 1136.3
## - `Hab. de nadar`:Local          1  750.58 1137.0
## + Sexo                          1  747.36 1137.8
## - `Hab. de nadar`:`Faixa etária` 2  754.37 1138.8
##
## Step: AIC=1136.18
## `Nº de infecções` ~ `Hab. de nadar` + Local + `Faixa etária` +
##   `Hab. de nadar`:`Faixa etária` + `Hab. de nadar`:Local +
##   Local:`Faixa etária`
##
##                                Df Deviance    AIC
## <none>                           743.78 1136.2
## - Local:`Faixa etária`           2  747.92 1136.3
## + `Hab. de nadar`:Local:`Faixa etária` 2  740.21 1136.6
## - `Hab. de nadar`:Local          1  746.35 1136.8
## + Sexo                          1  743.44 1137.8
## - `Hab. de nadar`:`Faixa etária` 2  750.17 1138.6

```

```

step_max_to_min <- step(fit_max,scope = list(lower=fit_min),direction = "both")

```

```

## Start: AIC=1124.11
## `Nº de infecções` ~ `Hab. de nadar` * Local * `Faixa etária` *
##   Sexo
##
##                                Df Deviance    AIC
## <none>                           703.72 1124.1
## - `Hab. de nadar`:Local:`Faixa etária`:Sexo 2  712.26 1128.7

```

```

# Testes do tipo H_0: CB=M vs H_1: CB <> M
# fit.model: saída do modelo ajustado
# m.C & m.M: matrizes de interesse

```

```

testeF.CBM.MLG <- function(fit.model,m.C,m.M)
{
v.beta <- cbind(fit.model$coef)
e.q <- nrow(m.C)
m.cov.beta <- (vcov(fit.model))
e.Q <- t(m.C%*%v.beta-m.M)%*%solve(m.C%*%m.cov.beta%*%t(m.C))%*%(m.C%*%v.beta-m.M)
e.pvalor <- 1-pchisq(e.Q,e.q)
cat("Estatística Q = ",round(e.Q,2),"\\n")
cat("gl = ",e.q,"\\n")
cat("pvalor = ",round(e.pvalor,4),"\\n")
cat("Matriz C :", "\\n")
print(m.C)
cat("Matriz M :", "\\n")
print(m.M)
}

```

```

# fit.model 1: H_0
# fit.model 1: modelo irrestrito

```

```

TRV.MLG <- function(fit.model1,fit.model2)
{
ep <- ncol(model.matrix(fit.model2))
eq <- ep-ncol(model.matrix(fit.model1))
desv1 <- deviance(fit.model1)
desv2 <- deviance(fit.model2)
e.TRV <- (desv1-desv2)
e.pvalor <- 1-pchisq(e.TRV,eq)
cat("Estatística F = ",round(e.TRV,2),"\\n")
cat("g.l = ",eq,"\\n")
cat("pvalor = ",round(e.pvalor,4),"\\n")
}

```

```

# fit.model 1: H_0
# fit.model 1: modelo irrestrito

```

```

AnaDesv.MLG <- function(fit.model1,fit.model2)
{
ep <- ncol(model.matrix(fit.model2))
eq <- ep-ncol(model.matrix(fit.model1))
n <- nrow(model.matrix(fit.model2))
desv1 <- deviance(fit.model1)
desv2 <- deviance(fit.model2)
e.F <- ((desv1-desv2)/eq)/(desv2/(n-ep))
e.pvalor <- 1-pf(e.F,eq,n-ep)
cat("Estatística F = ",round(e.F,2),"\\n")
cat("pvalor = ",round(e.pvalor,4),"\\n")
}

```

```

# Programa extra?do do site: https://www.ime.usp.br/~giapaula/textoregressao.htm
# Cr?ditos: Prof. Dr. Gilberto Alvarenga Paula

```

```

# Adaptado por Caio L. N. Azevedo
# source("D:\\windows\\Unicamp\\Disciplinas\\1_semestre_2016\\ME 720 MLG\\Programas\\diag_pois.r")

diagPoisson<-function(fit.model){
  # fit.model: objeto com o resultado do ajuste do MLG obtido atrav?s da fun??o glm
  X <- model.matrix(fit.model)
  n <- nrow(X)
  p <- ncol(X)
  w <- fit.model$weights
  W <- diag(w)
  H <- solve(t(X)%*%W%*%X)
  H <- sqrt(W)%*%X%*%H%*%t(X)%*%sqrt(W)
  h <- diag(H)
  ts <- resid(fit.model,type="pearson")/sqrt(1-h)
  td <- resid(fit.model,type="deviance")/sqrt(1-h)
  di <- (h/(1-h))*(ts^2)
  par(mfrow=c(2,2))
  a <- max(td)
  b <- min(td)
  plot(td,xlab="Índice", ylab="Resíduo Componente do Desvio",
  ylim=c(b-1,a+1), pch=16,cex.axis=1.1,cex.lab=1.1,cex=1.1,cex.axis=1.1,cex.lab=1.1)
  abline(2,0,lty=2)
  abline(-2,0,lty=2)
  abline(0,0,lty=2)

  # identify(td, n=1)
  # title(sub="(c)")
  fitted = fitted(fit.model)
  plot(fitted ,td,xlab="Valor ajustado (média)", ylab="Resíduo Componente do Desvio",ylim=c(b-1,a+1), pch=
  main="",cex=1.1,cex.axis=1.1,cex.lab=1.1)
  abline(2,0,lty=2)
  abline(-2,0,lty=2)
  abline(0,0,lty=2)

  #
  #hist(td,xlab="Res?duo Componente do Desvio",ylab="densidade",probability=TRUE,main="",cex=1.1,cex.axis=
  #
  eta = predict(fit.model)
  z = eta + resid(fit.model, type="pearson")/sqrt(w)
  plot(predict(fit.model),z,xlab="Preditor Linear",ylab="Variavel z", pch=16,main="",cex=1.1,cex.axis=1.1)
  lines(smooth.spline(predict(fit.model), z, df=2))
  #
  #-----#
}

#-----#
# Para rodar este programa deixe no objeto fit.model a sa?da
# do ajuste da regress?o com erros Poisson e liga??o log.
# Programa extra?do do site: https://www.ime.usp.br/~giapaula/textoregressao.htm
# Cr?ditos: Prof. Dr. Gilberto Alvarenga Paula
# Adaptado por Caio L. N. Azevedo
# source("D:\\windows\\Unicamp\\Disciplinas\\1_semestre_2016\\ME 720 MLG\\Programas\\envel_Bern.r")

```

```

envelPoisson <- function(fit.model,ligacao){
  # fit.model: objeto com o resultado do ajuste do MLG obtido através da função glm
  # ligacao: função de ligação (mesmo nome usado pela função glm (colocar entre aspas))
  #par(mfrow=c(1,1))
  X <- model.matrix(fit.model)
  n <- nrow(X)
  p <- ncol(X)
  w <- fit.model$weights
  W <- diag(w)
  H <- solve(t(X)%*%W%*%X)
  H <- sqrt(W)%*%X%*%H%*%t(X)%*%sqrt(W)
  h <- diag(H)
  td <- resid(fit.model,type="deviance")/sqrt((1-h))
  e <- matrix(0,n,100)
  #
  for(i in 1:100){
    nresp <- rpois(n, fitted(fit.model))
    fit <- glm(nresp ~ X, family=poisson(link=ligacao))
    w <- fit$weights
    W <- diag(w)
    H <- solve(t(X)%*%W%*%X)
    H <- sqrt(W)%*%X%*%H%*%t(X)%*%sqrt(W)
    h <- diag(H)
    e[,i] <- sort(resid(fit,type="deviance")/sqrt(1-h))}
  #
  e1 <- numeric(n)
  e2 <- numeric(n)
  #
  for(i in 1:n){
    eo <- sort(e[i,])
    e1[i] <- (eo[2]+eo[3])/2
    e2[i] <- (eo[97]+eo[98])/2}
  #
  med <- apply(e,1,mean)
  faixa <- range(td,e1,e2)
  #par(pty="s")
  qqnorm(td,xlab="Percentil da N(0,1)",
  ylab="Resíduo Componente do Desvio", ylim=faixa, pch=16,main="",cex=1.1,cex.axis=1.1,cex.lab=1.1)
  #
  par(new=T)
  #
  qqnorm(e1,axes=F,xlab="",ylab="",type="l",ylim=faixa,lty=1,main="")
  par(new=T)
  qqnorm(e2,axes=F,xlab="",ylab="", type="l",ylim=faixa,lty=1,main="")
  par(new=T)
  qqnorm(med,axes=F,xlab="", ylab="", type="l",ylim=faixa,lty=2,main="")
  #-----#
}

```

Pelo stepwise, começando pelo modelo com somente o intercepto em direção ao modelo com todas as interações, chegamos no modelo:

```
fit_step_min_to_max <- glm(`Nº de infecções` ~ `Hab. de nadar` + Local + `Faixa etária` + `Hab. de nadar`
```

```

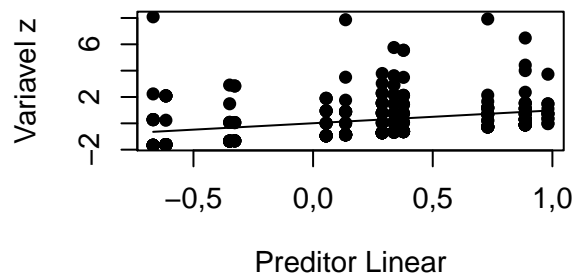
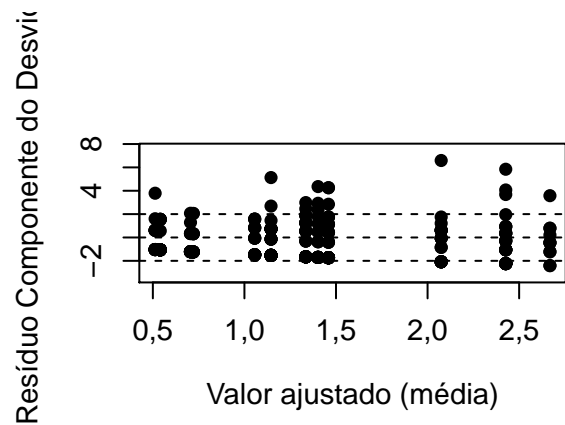
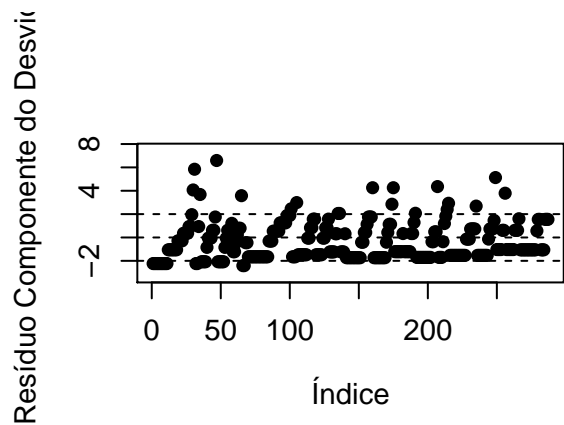
library(xtable)
gerador_tbl_coef <- function(fit.model){
  options(encoding="utf-8")
  tabela <- summary(fit.model)$coefficients
  tabela[,4] <- ifelse(tabela[,4] < 0.0001,"<0,0001",tabela[,4])
  colnames(tabela) <- c("Estimativa","Erro Padrão","Estatística Z", "P-valor")
  options(OutDec= ",")
  return(xtable(tabela,digits = 4))
}
gerador_tbl_coef(fit_step_min_to_max)

## % latex table generated in R 3.4.0 by xtable 1.8-2 package
## % Fri Jun 16 23:09:06 2017
## \begin{table}[ht]
## \centering
## \begin{tabular}{rrrrr}
## \hline
## & Estimativa & Erro Padrão & Estatística Z & P-valor \\
## \hline
## (Intercept) & 0,1356 & 0,1502 & 0,9029 & 0,3666 \\
## `Hab. de nadar`Occas & 0,2421 & 0,1886 & 1,2838 & 0,1992 \\
## LocalNonBeach & 0,1534 & 0,1904 & 0,8055 & 0,4205 \\
## `Faixa etária`20-24 & -0,8041 & 0,2879 & -2,7931 & 0,0052 \\
## `Faixa etária`25-29 & -0,7509 & 0,2762 & -2,7185 & 0,0066 \\
## `Hab. de nadar`Occas:`Faixa etária`20-24 & 0,0776 & 0,2612 & 0,2971 & 0,7664 \\
## `Hab. de nadar`Occas:`Faixa etária`25-29 & 0,7112 & 0,2930 & 2,4272 & 0,0152 \\
## `Hab. de nadar`Occas:LocalNonBeach & 0,3555 & 0,2217 & 1,6038 & 0,1088 \\
## LocalNonBeach:`Faixa etária`20-24 & 0,5696 & 0,2882 & 1,9763 & 0,0481 \\
## LocalNonBeach:`Faixa etária`25-29 & 0,1346 & 0,2620 & 0,5137 & 0,6075 \\
## \hline
## \end{tabular}
## \end{table}

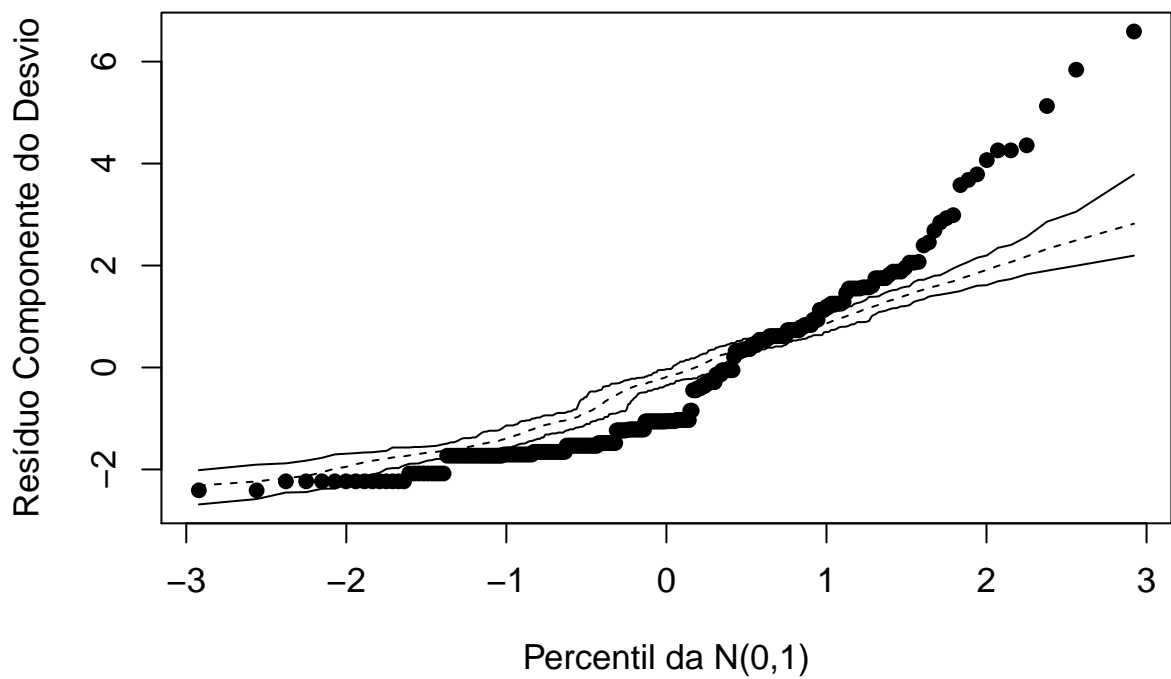
```

	Estimativa	Erro Padrão	Estatística Z	P-valor
(Intercept)	0,1356	0,1502	0,9029	0,3666
`Hab. de nadar`Occas	0,2421	0,1886	1,2838	0,1992
LocalNonBeach	0,1534	0,1904	0,8055	0,4205
`Faixa etária`20-24	-0,8041	0,2879	-2,7931	0,0052
`Faixa etária`25-29	-0,7509	0,2762	-2,7185	0,0066
`Hab. de nadar`Occas:`Faixa etária`20-24	0,0776	0,2612	0,2971	0,7664
`Hab. de nadar`Occas:`Faixa etária`25-29	0,7112	0,2930	2,4272	0,0152
`Hab. de nadar`Occas:LocalNonBeach	0,3555	0,2217	1,6038	0,1088
LocalNonBeach:`Faixa etária`20-24	0,5696	0,2882	1,9763	0,0481
LocalNonBeach:`Faixa etária`25-29	0,1346	0,2620	0,5137	0,6075

```
diagPoisson(fit_step_min_to_max)
```



```
envelPoisson(fit_step_min_to_max, "log")
```



Pelo stepwise, começando pelo modelo com todas as interações em direção ao modelo com somente o intercepto, chegamos no modelo:

```
fit_step_max_to_min <- glm(`Nº de infecções` ~ `Hab. de nadar` * `Local` * `Faixa etária` * `Sexo`, family = poisson)

library(xtable)
gerador_tbl_coef(fit_step_max_to_min)
```

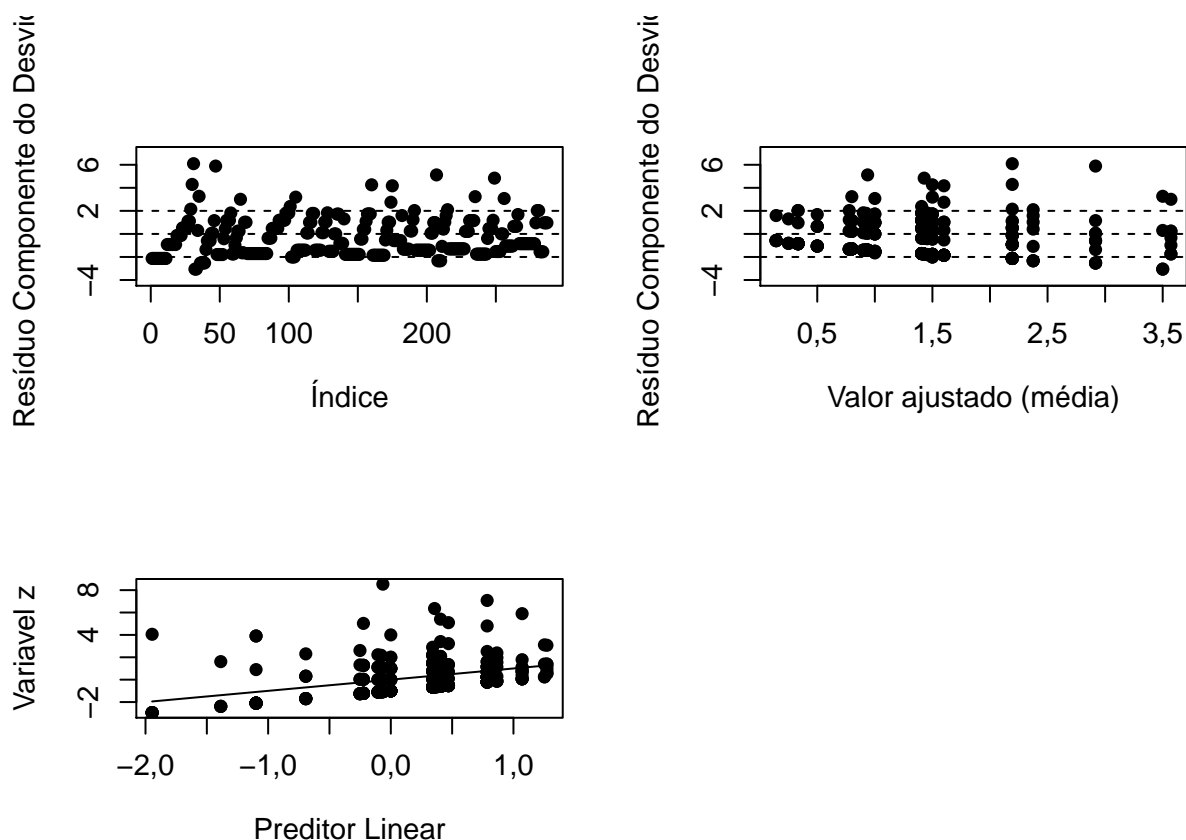
```
## % latex table generated in R 3.4.0 by xtable 1.8-2 package
## % Fri Jun 16 23:09:10 2017
## \begin{table}[ht]
## \centering
## \begin{tabular}{rrrrr}
## \hline
## & Estimativa & Erro Padrão & Estatística Z & P-valor \\
## \hline
## (Intercept) & 0,3567 & 0,2236 & 1,5951 & 0,1107 \\
## `Hab. de nadar`Occas & 0,1133 & 0,3028 & 0,3743 & 0,7082 \\
## LocalNonBeach & 0,0488 & 0,4655 & 0,1048 & 0,9165 \\
## `Faixa etária`20-24 & -1,0498 & 0,5000 & -2,0996 & 0,0358 \\
## `Faixa etária`25-29 & -0,3567 & 0,4655 & -0,7663 & 0,4435 \\
## SexoMale & -0,5798 & 0,3354 & -1,7287 & 0,0839 \\
## `Hab. de nadar`Occas:LocalNonBeach & 0,7340 & 0,5742 & 1,2781 & 0,2012 \\
## `Hab. de nadar`Occas:`Faixa etária`20-24 & 0,3285 & 0,6592 & 0,4983 & 0,6182 \\
## `Hab. de nadar`Occas:`Faixa etária`25-29 & 0,7517 & 0,5576 & 1,3479 & 0,1777 \\
## LocalNonBeach:`Faixa etária`20-24 & 0,5390 & 0,7265 & 0,7419 & 0,4581 \\
## LocalNonBeach:`Faixa etária`25-29 & -1,4351 & 1,1762 & -1,2202 & 0,2224 \\
## `Hab. de nadar`Occas:SexoMale & 0,5153 & 0,4330 & 1,1900 & 0,2341 \\
## LocalNonBeach:SexoMale & 0,5153 & 0,5490 & 0,9386 & 0,3479 \\
## `Faixa etária`20-24:SexoMale & 1,2730 & 0,6748 & 1,8864 & 0,0592 \\
## `Faixa etária`25-29:SexoMale & -0,5188 & 0,6922 & -0,7495 & 0,4536 \\
## `Hab. de nadar`Occas:LocalNonBeach:`Faixa etária`20-24 & -0,6957 & 0,9199 & -0,7563 & 0,4495 \\
## `Hab. de nadar`Occas:LocalNonBeach:`Faixa etária`25-29 & -0,2127 & 1,3413 & -0,1586 & 0,8740 \\
## `Hab. de nadar`Occas:LocalNonBeach:SexoMale & -0,9180 & 0,6801 & -1,3498 & 0,1771 \\
## `Hab. de nadar`Occas:`Faixa etária`20-24:SexoMale & -2,9030 & 1,2935 & -2,2443 & 0,0248 \\
## `Hab. de nadar`Occas:`Faixa etária`25-29:SexoMale & -0,3462 & 0,8206 & -0,4219 & 0,6731 \\
## LocalNonBeach:`Faixa etária`20-24:SexoMale & -1,1831 & 0,9158 & -1,2919 & 0,1964 \\
## LocalNonBeach:`Faixa etária`25-29:SexoMale & 1,9696 & 1,3390 & 1,4709 & 0,1413 \\
## `Hab. de nadar`Occas:LocalNonBeach:`Faixa etária`20-24:SexoMale & 3,9762 & 1,4946 & 2,6604 & 0,0077 \\
## `Hab. de nadar`Occas:LocalNonBeach:`Faixa etária`25-29:SexoMale & 0,6356 & 1,5374 & 0,4134 & 0,6792 \\
## \hline
## \end{tabular}
## \end{table}
```

Que é o mesmo modelo que o modelo completo feito anteriormente.

```
diagPoisson(fit_step_max_to_min)
```



	Estimativa	Erro Padrão	Estatística Z	P-valor
(Intercept)	0,3567	0,2236	1,5951	0,1107
‘Hab. de nadar’Occas	0,1133	0,3028	0,3743	0,7082
LocalNonBeach	0,0488	0,4655	0,1048	0,9165
‘Faixa etária’20-24	-1,0498	0,5000	-2,0996	0,0358
‘Faixa etária’25-29	-0,3567	0,4655	-0,7663	0,4435
SexoMale	-0,5798	0,3354	-1,7287	0,0839
‘Hab. de nadar’Occas:LocalNonBeach	0,7340	0,5742	1,2781	0,2012
‘Hab. de nadar’Occas:‘Faixa etária’20-24	0,3285	0,6592	0,4983	0,6182
‘Hab. de nadar’Occas:‘Faixa etária’25-29	0,7517	0,5576	1,3479	0,1777
LocalNonBeach:‘Faixa etária’20-24	0,5390	0,7265	0,7419	0,4581
LocalNonBeach:‘Faixa etária’25-29	-1,4351	1,1762	-1,2202	0,2224
‘Hab. de nadar’Occas:SexoMale	0,5153	0,4330	1,1900	0,2341
LocalNonBeach:SexoMale	0,5153	0,5490	0,9386	0,3479
‘Faixa etária’20-24:SexoMale	1,2730	0,6748	1,8864	0,0592
‘Faixa etária’25-29:SexoMale	-0,5188	0,6922	-0,7495	0,4536
‘Hab. de nadar’Occas:LocalNonBeach:‘Faixa etária’20-24	-0,6957	0,9199	-0,7563	0,4495
‘Hab. de nadar’Occas:LocalNonBeach:‘Faixa etária’25-29	-0,2127	1,3413	-0,1586	0,8740
‘Hab. de nadar’Occas:LocalNonBeach:SexoMale	-0,9180	0,6801	-1,3498	0,1771
‘Hab. de nadar’Occas:‘Faixa etária’20-24:SexoMale	-2,9030	1,2935	-2,2443	0,0248
‘Hab. de nadar’Occas:‘Faixa etária’25-29:SexoMale	-0,3462	0,8206	-0,4219	0,6731
LocalNonBeach:‘Faixa etária’20-24:SexoMale	-1,1831	0,9158	-1,2919	0,1964
LocalNonBeach:‘Faixa etária’25-29:SexoMale	1,9696	1,3390	1,4709	0,1413
‘Hab. de nadar’Occas:LocalNonBeach:‘Faixa etária’20-24:SexoMale	3,9762	1,4946	2,6604	0,0078
‘Hab. de nadar’Occas:LocalNonBeach:‘Faixa etária’25-29:SexoMale	0,6356	1,5374	0,4134	0,6793



```
envelPoisson(fit_step_max_to_min,"log")
```

