

MLG gama em R

Este exemplo refere-se a um experimento em que a resistência (em horas) de um determinado tipo de vidro foi avaliada segundo quatro valores de voltagem (200, 250, 300 e 350, em kilovolts) e duas temperaturas (170 e 180, em graus Celsius). O principal interesse consiste em comparar as resistências médias em relação aos níveis de voltagem e temperatura. Uma descrição do problema encontra-se no livro do Prof. G. A. Paula, p. 175 (http://www.ime.usp.br/~giapaula/texto_2013.pdf). Os dados podem ser obtidos na página <http://www.ime.usp.br/~giapaula/vidros.dat>.

Os dados são lidos com os comandos abaixo. As variáveis `volt` (voltagem) e `temp` (temperatura) são transformadas em qualitativas (`factor`).

```
## Dados
dados <- . . .
colnames(dados) <- c("resist", "volt", "temp")

dados$volt <- factor(dados$volt, labels = c("200kV", "250kV", "300kV",
"350kV"))
dados$temp <- factor(dados$temp, labels = c("170C", "180C"))
```

Algumas estatísticas descritivas (média, desvio padrão e coeficiente de variação) da variável resposta são calculadas para cada nível de voltagem e temperatura.

```
# Estatísticas descritivas
by(dados$resist, dados[, c("volt", "temp")], function(x) c(length(x),
mean(x), sd(x), sd(x) / mean(x)))
```

```
volt: 200kV
temp: 170C
[1] 4.0000000 885.0000000 311.1944730 0.3516322
-----
volt: 250kV
temp: 170C
[1] 4.0000000 814.0000000 229.6490075 0.2821241
-----
volt: 300kV
temp: 170C
[1] 4.0000000 424.2500000 147.8769195 0.3485608
-----
volt: 350kV
temp: 170C
[1] 4.0000000 362.7500000 155.9174461 0.4298207
-----
volt: 200kV
temp: 180C
[1] 4.000000e+00 1.044000e+03 5.760787e+01 5.517995e-02
-----
volt: 250kV
temp: 180C
[1] 4.0000000 364.7500000 121.7439252 0.3337736
```

```
-----
volt: 300kV
temp: 180C
[1] 4.0000000 317.0000000 57.6599225 0.1818925
-----
```

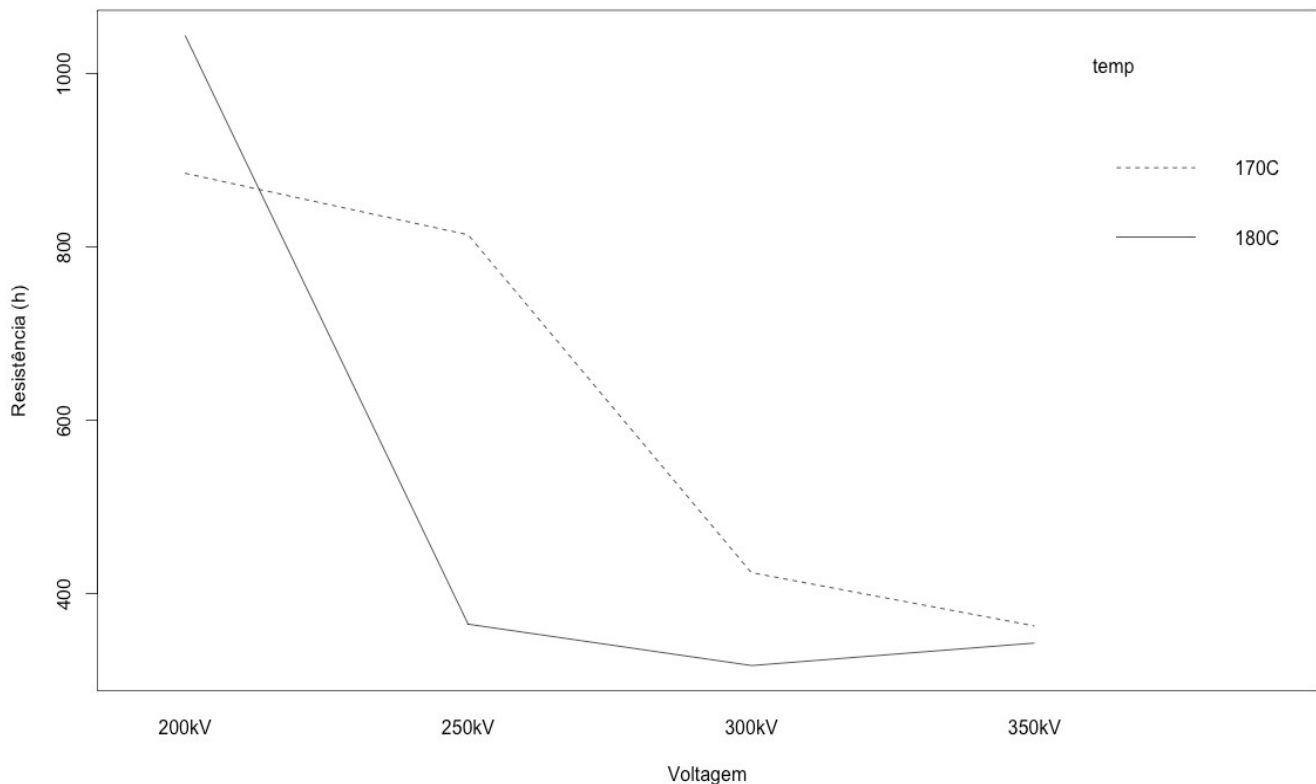
```
volt: 350kV
temp: 180C
[1] 4.0000000 343.0000000 118.0621305 0.3442045
```

Como a variável resposta é positiva, podemos iniciar com o ajuste de um modelo gama. Pelos resultados acima, o coeficiente de variação tem estimativas de 0,055 a 0,43, sendo que dos oito grupos de observações, em apenas um a estimativa é próxima ao mínimo. Ressalte-se que o número de observações em cada grupo é apenas 4. Prosseguimos com o modelo gama, em que o parâmetro de dispersão representa a raiz quadrada do coeficiente de variação. Dizemos que o modelo gama é um modelo para dados com coeficiente de variação constante.

Nota 1. Prove que no modelo gama, o parâmetro de dispersão é igual à raiz quadrada do coeficiente de variação.

O gráfico de interação abaixo indica uma redução da resistência média (por que “média”?) à medida que aumenta o nível de voltagem. Além disso, há indicação de interação entre voltagem e temperatura (por quê?).

```
# Gráfico de interação
with(dados, interaction.plot(volt, temp, resist, xlab = "Voltagem",
                             ylab = "Resistência (h)"))
```



Nota 2. Você recomendaria um gráfico de caixas (*boxplot*) para representar estes dados?

Nota 3. Represente os dados em um gráfico de pontos.

Será utilizada a função de ligação identidade. O primeiro modelo ajustado inclui os efeitos principais e a interação *voltagem × temperatura*.

```
## Modelos
m1 <- glm(resist ~ volt * temp, family = Gamma(link = "identity"), data = dados)
summary(m1)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	885.0	137.8	6.424	1.21e-06	***
volt250kV	-71.0	187.2	-0.379	0.70780	
volt300kV	-460.8	152.8	-3.016	0.00598	**
volt350kV	-522.2	148.9	-3.508	0.00181	**
temp180C	159.0	213.1	0.746	0.46275	
volt250kV:temp180C	-608.2	254.3	-2.392	0.02496	*
volt300kV:temp180C	-266.2	228.4	-1.165	0.25529	
volt350kV:temp180C	-178.8	226.8	-0.788	0.43831	

Nota 4. Levando em conta a função de ligação adotada, os sinais das estimativas estão compatíveis com o gráfico da figura, p. 1?

Nota 5. Interprete as estimativas acima.

Apenas uma interação apresenta coeficiente significativo a um nível de 5% (mas não a 1%). Ajustamos um modelo sem a interação *voltagem × temperatura*.

```
m2 <- update(m1, . ~ . -volt:temp)
summary(m2)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	1039.94	122.50	8.489	4.21e-09	***
volt250kV	-426.49	135.61	-3.145	0.00402	**
volt300kV	-608.81	126.21	-4.824	4.89e-05	***
volt350kV	-612.89	126.04	-4.863	4.40e-05	***
temp180C	-117.77	56.43	-2.087	0.04644	*

Os coeficientes são significativos a um nível de 5%, sendo que no caso da variável temperatura a situação é limítrofe. Em seguida os dois modelos são comparados.

```
anova(m2, m1, test = "F")
Analysis of Deviance Table
Model 1: resist ~ volt + temp
Model 2: resist ~ volt * temp
  Resid. Df Resid. Dev Df Deviance      F Pr(>F)
1      27      3.4306
2      24      2.4277  3    1.0029 3.4487 0.03246 *
```

A diferença não é significativa a um nível de 1%. Por simplicidade, adotamos o modelo contendo apenas os efeitos principais, no qual todos os níveis de voltagem apresentam diferença negativa, decrescente e significativa em relação ao nível de referência (200 kV) sobre a resistência média.

Nota 6. Procure simplificar ainda mais, excluindo a variável temperatura.

Os gráficos de envelopes abaixo não apontam afastamentos sérios das suposições inerentes ao modelo.

```
## Envelope m2
# Número de simulações
M <- 100

X <- model.matrix(m2)
n <- nrow(X)
phi <- sum((resid(m2, type = "response") / (fitted(m2))) ^2) /
      (n - ncol(X))
W <- diag(m2$weights)
h <- diag(sqrt(W) %*% X %*% solve(t(X) %*% W %*% X) %*% t(X) %*% sqrt(W))
eD <- resid(m2, type = "deviance") / sqrt(phi * (1 - h))
eDo <- sort(eD)

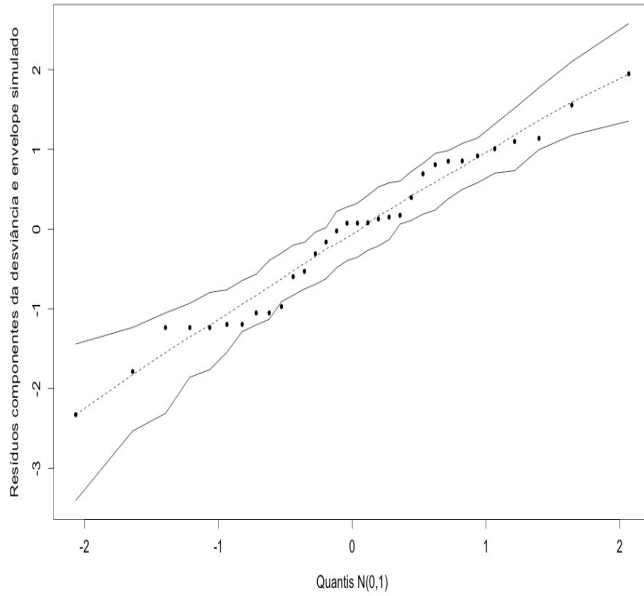
# Simulações
meD <- matrix(0, n, M)
for (i in 1:M) {
  simy <- simulate(m2, nsim = 1)
  m2s <- glm(simy[, 1] ~ X, family = Gamma(link = "identity"))
  phis <- sum((resid(m2s, type = "response") / (fitted(m2s))) ^2) /
    (n - ncol(X))
  W <- diag(m2s$weights)
  h <- diag(sqrt(W) %*% X %*% solve(t(X) %*% W %*% X) %*% t(X) %*% sqrt(W))
  eDs <- resid(m2s, type = "deviance") / sqrt(phis * (1 - h))
  meD[, i] <- eDs
}
meD <- apply(meD, 2, sort)
Z <- qnorm((1:n - 3/8) / (n + 1/4))
eDm <- apply(meD, 1, mean)
eDmin <- apply(meD, 1, min)
eDmax <- apply(meD, 1, max)
meD <- cbind(Z, eDo, eDmin, eDm, eDmax)

# Envelope
par(mai = c(1.2, 1.2, 0.5, 0.1))
plot(meD[, 1], meD[, 2], pch = 20, ylim = range(meD[, -1]),
     cex.axis = 1.2, cex.lab = 1.2, xlab = "Quantis N(0,1)",
     ylab = "Resíduos componentes da desviância e envelope simulado")
lines(meD[, 1], meD[, 3])
lines(meD[, 1], meD[, 4], lty = 2)
lines(meD[, 1], meD[, 5])
```

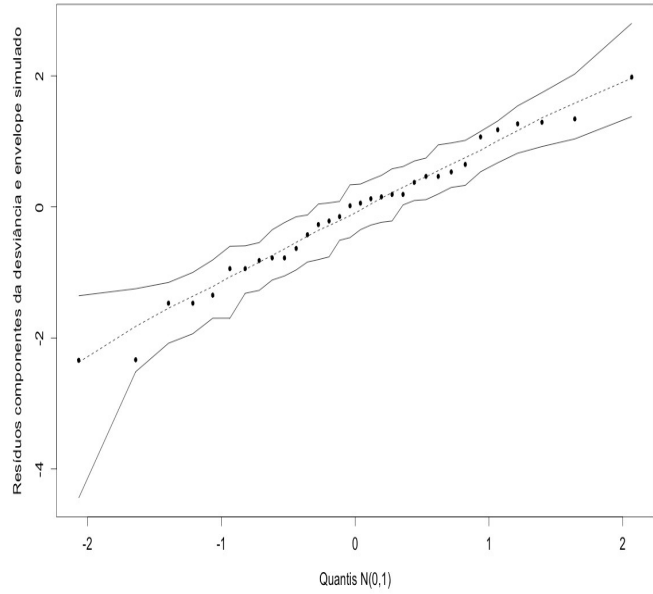
Nota 7. Justifique o estimador do parâmetro de dispersão destacado acima em amarelo. Compare a estimativa obtida com o resultado da função `summary`.

Nota 8. Apresente outra estimativa do parâmetro de dispersão.

Modelo com efeitos principais e interação



Modelo com efeitos principais



Nota 9. Experimente outras funções de ligação.

Nota 10. Verifique os modelos ajustados distribuição normal e normal inversa para a variável resposta.

Nota 11. Procure identificar pontos de alavancagem e, se existirem, avaliar a influência deles sobre as inferências.

Nota 12. Procure reproduzir os resultados utilizando o sistema SAS.