

MLG para respostas positivas (assimétricas)

Prof. Caio Azevedo

MLG para repostas positivas (assimétricas): modelo gama

- Suponha que $Y \sim \text{gama}(r, k)$, $\mathcal{E}(Y) = rk$ e $\mathcal{V}(Y) = rk^2$, então:

$$f_Y(y) = \frac{1}{\Gamma(r)k^r} y^{r-1} \exp\left\{-\frac{y}{k}\right\} \mathbb{1}_{(0,\infty)}(y)$$

- Se fizermos $k = \frac{\mu}{\phi}$ e $\mathcal{E}(Y) = \mu$, teremos a parametrização adotada no curso, ou seja:

$$f_Y(y) = \frac{1}{\Gamma(\phi)} \left(\frac{\phi y}{\mu}\right)^{\phi} \exp\left\{-\frac{\phi y}{\mu}\right\} y^{-1} \mathbb{1}_{(0,\infty)}(y)$$

MLG para repostas positivas (assimétricas): modelo gama

$$Y_i \stackrel{ind.}{\sim} \text{gama}(\mu_i, \phi) \quad , \quad \mathcal{E}(Y_i) = \mu_i, \mathcal{V}(Y_i) = \frac{\mu_i^2}{\phi}, i = 1, \dots, n$$

$$g(\mu_i) = \mathbf{X}_i' \boldsymbol{\beta} = \sum_{j=1}^p X_{ji} \beta_j; \mathbf{X}_i = (X_{1i}, \dots, X_{pi})'$$

em geral $g(\cdot) = \ln(\cdot)$ função de ligação log (ou logaritmica). Nesse caso,

$$\ln(\mu_i) = \eta_i \rightarrow \mu_i = \exp(\eta_i)$$

- $CV(Y_i) = \frac{\mu_i / \sqrt{\phi}}{\mu_i} = \frac{1}{\sqrt{\phi}}$ (o coeficiente de variação do modelo gama é o mesmo para todas as observações).
- (Coeficiente de assimetria de Pearson) $CA(Y_i) = \frac{2}{\sqrt{\phi}}$. Assim, quando $\phi \rightarrow \infty$, a distribuição de Y_i tende a apresentar um comportamento simétrico.

Estimação por MV sob a função de ligação log

- Lembremos que, nesse caso, a função de ligação log não corresponde à função de ligação canônica.
- $S(\beta) = \phi \mathbf{X}' \mathbf{W}^{1/2} \mathbf{V}^{-1/2} (\mathbf{y} - \boldsymbol{\mu})$, em que $\mathbf{W} = \text{diag}(\omega_1, \dots, \omega_n)$, $\mathbf{V} = \text{diag}(V_1, \dots, V_n)$. Temos que $V_i = \mu_i^2 = \exp(2\eta_i)$ (ligação log). Além disso, $\omega_i = \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2 / V_i = \frac{\exp(2\eta_i)}{\mu_i} = 1$ (função de ligação log), $i = 1, 2, \dots, n$. Assim $\mathbf{W} = \mathbf{I}_n$.

Estimação por MV sob a função de ligação log

- $S(\phi) = \sum_{i=1}^n \{y_i \theta_i - b(\theta_i)\} + \sum_{i=1}^n c'(y_i, \phi)$. Como $c(y_i, \phi) = (\phi - 1) \ln(y_i) + \phi \ln(\phi) - \ln \Gamma(\phi)$, temos que $c'(y_i, \phi) = \ln(y_i) + \ln(\phi) + 1 - \Psi(\phi)$, em que $\Gamma'(\phi) = \frac{d\Gamma(\phi)}{d\phi}$ e $\Psi(\phi) = \frac{\Gamma'(\phi)}{\Gamma(\phi)}$ é a função digama (implementada no R).

Estimação por MV sob a função de ligação log

■ $l_{11}(\beta, \phi) = \phi \mathbf{X}' \mathbf{W} \mathbf{X} = \phi \mathbf{X}' \mathbf{X}$ e $l_{22}(\beta, \phi) = -\mathcal{E} \left(\sum_{i=1}^n c''(Y_i, \phi) \right)$.

Como $c''(y_i, \phi) = \frac{1}{\phi} - \Psi'(\phi)$, em que

$$\Psi'(\phi) = \frac{1}{(\Gamma(\phi))^2} \left(\Gamma''(\phi)\Gamma(\phi) - (\Gamma'(\phi))^2 \right) \text{ (função trigama,}$$

implementada no R) e $\Gamma''(\phi) = \frac{d^2\Gamma(\phi)}{d\phi^2}$, então

$$l_{22}(\beta, \phi) = -n \left(\frac{1}{\phi} - \Psi'(\phi) \right).$$

Estimação por MV sob a função de ligação log

- Dessa forma, devemos utilizar o processo iterativo (algoritmo Escore de Fisher), apresentado anteriormente (slides 16 e 17 de http://www.ime.unicamp.br/~cnaber/aula_Intro_MLG_Parte2_MLG_1S_2016.pdf), para obtermos estimativas para β e ϕ .
- Alternativamente, a etapa do algoritmo Escore de Fisher, para obtermos estimativas para o parâmetro ϕ , pode ser substituída pela utilização do estimador do tipo método dos momentos (slide 36 de http://www.ime.unicamp.br/~cnaber/aula_Intro_MLG_MLG_1S_2016.pdf).

Desvio e RCD

- As formas do desvio e do RCD para o modelo gama já foram vistas anteriormente, respectivamente: slide 25 (http://www.ime.unicamp.br/~cnaber/aula_Intro_MLG_Parte2_MLG_1S_2016.pdf) e slide 6 (http://www.ime.unicamp.br/~cnaber/aula_Ana_Res_MLG_MLG_1S_2016.pdf).

Utilizando o programa R

- função *glm* (disponível quando na instalação básica do R).
- Sua sintaxe se assemelha bastante com a da função *lm* (na forma de definir a estrutura de regressão)

(com intercepto)

```
result <- glm(Y ~ X1 + X2 + ... + Xp,
```

```
family = Gamma(link = "log"));
```

```
summary(result)
```

(sem intercepto)

```
result <- glm(Y ~ -1 + X1 + X2 + ... + Xp,
```

```
family = Gamma(link = "log"));
```

```
summary(result)
```

- A estimativa default do R para o parâmetro ϕ é a estimativa obtida pelo método dos momentos.
- Uma outra estimativa para ϕ é $\tilde{\phi} = \frac{n - p}{D(\mathbf{y}; \tilde{\boldsymbol{\mu}})}$, a qual não é consistente, em que $D(., .)$ é o desvio do modelo.
- Para se estimar ϕ no modelo gama, por MV, devemos utilizar a função *gamma.shape* do pacote *MASS* (não é um pacote default), através do comando *gamma.shape(result)*.
- Note que a estimativa de ϕ pode afetar os valores de: $\widetilde{\text{Cov}}(\hat{\boldsymbol{\beta}})$, RCD, AIC, BIC, desvio escalonado e da estatística Q (teste $\mathbf{C}\boldsymbol{\beta} = \mathbf{M}$).

Resumo dos comandos básicos

Distribuição (family)	Função de ligação (link)
binomial	logit (logito), probit (probito), cauchit (cauchito), log , cloglog (complemento log-log)
gaussian (normal)	identity, log, inverse
Gamma (gama)	inverse, identity, log
poisson	log, identity, sqrt
inverse.gaussian (inversa gaussiana)	$1/\mu^2$, inverse, identity, log

Distribuição (family)	Função de ligação (link)
identity	$g(\mu) = \mu$
log	$g(\mu) = \ln(\mu)$
inverse	$g(\mu) = \frac{1}{\mu}$
sqrt	$g(\mu) = \sqrt{\mu}$
logit	$g(\mu) = \ln\left(\frac{\mu}{1-\mu}\right)$
probit	$g(\mu) = \Phi^{-1}(\mu)$; $\Phi^{-1}(\cdot)$ inversa da cdf da $N(0, 1)$
cauchito	$g(\mu) = F^{-1}(\mu)$; $F^{-1}(\cdot)$ inversa da cdf da Cauchy(0, 1)
cloglog	$g(\mu) = \ln(-\ln(1 - \mu))$

Outros comandos

Quantidade	Comando
$\tilde{\beta}$	<code>coef(result)</code>
$\widetilde{Cov}(\hat{\beta})$	<code>vcov(result)</code>
$\frac{1}{\tilde{\phi}}$	<code>(summary(result))\$dispersion</code>

Para o modelo gama, em relação à estimativa de MV do parâmetro ϕ , devemos fazer `resultphi < - gamma.shape(result)`, `resultphi$alpha` (estimativa) e `resultphi$SE` (estimativa do erro-padrão)

Outros comandos

- Resultados sob a estimativa pelo MM de ϕ .

summary(fit.model)

- Resultados sob a estimativa por MV de ϕ (*).

summary(fit.model, dispersion = gamma.dispersion(fit.model))

(aceita outros valores para $1/\phi$, bastando digitar “dispersion = 1/phi”, em que phi é uma estimativa de interesse).

- Em princípio, vamos considerar a estimativa de MV de ϕ .

Outros comandos

- Para ver os resultados disponibilizados pela função *glm* faça *names(fit.model)* ou *names(summary(fit.model))*.
- O comando (*) é útil para obter valores corrigidos para $\widetilde{\text{Cov}}(\hat{\beta})$. O restante (veja página 10) deve ser calculado utilizando-se comandos específicos juntamente com a estimativa de interesse de ϕ .

Desvio e RCD

- Desvio não escalonado: $\text{deviance}(\text{result})$. Desvio escalonado: $f_i \cdot \text{deviance}(\text{result})$.

- RCD:

```
X = model.matrix(result); V = fitted(result)
```

```
V = diag(V); w = result$weights
```

```
W = diag(w); H = solve(t(X) %*% W %*% X)
```

```
H = sqrt(W) %*% X %*% H %*% t(X) %*% sqrt(W); h = diag(H)
```

```
rd = resid(result, type = deviance)
```

```
td = rd * sqrt(fi / (1 - h))
```

em que “ f_i ” é uma estimativa de ϕ .

Análise do desvio

- Suponha que $\beta_{(p \times 1)} = (\beta_1', \beta_2')'$ de tal forma que $\beta_{1(q \times 1)}$ e que desejamos testar $H_0 : \beta_1 = \mathbf{0}$ vs $H_1 : \beta_1 \neq \mathbf{0}$.
- Defina $Q_{AD} = \frac{(D(\mathbf{y}; \hat{\mu}^{(0)}) - D(\mathbf{y}; \hat{\mu})) / q}{D(\mathbf{y}; \hat{\mu}) / (n - p)}$, em que $D(\mathbf{y}; \hat{\mu}^{(0)})$ e $D(\mathbf{y}; \hat{\mu})$ representam, respectivamente, o desvio associado ao modelo ajustado sob H_0 e ao modelo ajustado sem a restrição.

Análise do desvio

- Sob H_0 e para n suficientemente grande, temos que $Q_{AD} \approx F_{(q, n-p)}$.
Note que para se utilizar tal estatística não é necessário estimar ϕ .

- Assim, rejeita-se H_0 se p -valor $\leq \alpha$, em que

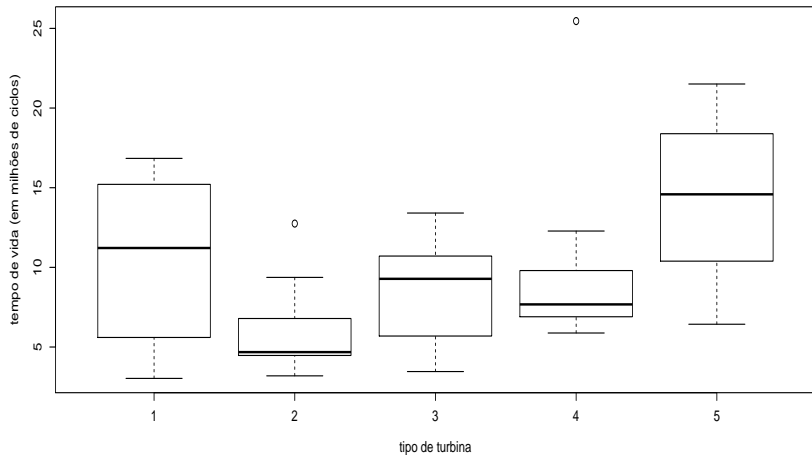
p -valor $\approx P(X \geq q_{AD} | H_0)$, $X \sim F_{(q, n-p)}$ e

$q_{AD} = \frac{\left(D(\mathbf{y}; \tilde{\boldsymbol{\mu}}^{(0)}) - D(\mathbf{y}; \tilde{\boldsymbol{\mu}}) \right) / q}{D(\mathbf{y}; \tilde{\boldsymbol{\mu}}) / (n-p)}$ ($(\tilde{\cdot})$ representa a respectiva estimativa).

Voltando ao Exemplo 1 (potência de turbinas de aviões)

- Vamos considerar os 5 tipos de turbinas analisados no experimento, doravante tipos 1, 2, 3, 4 e 5.
- $n_i = 10, \forall i$ (tamanho amostral por grupo).
- Y_{ij} : tempo de vida (em milhões de ciclos) da j -ésima turbina do i -ésimo tipo.
- Quanto maior o número médio de ciclos, melhor o desempenho da turbina.

Análise descritiva



Análise descritiva

Tipo de turbina	Média	DP	Var.	CV(%)	CA	Mín.	Máx.
1	10,69	4,82	23,23	45,07	-0,20	3,03	16,84
2	6,05	2,92	8,50	48,18	1,20	3,19	12,75
3	8,64	3,29	10,83	38,10	-0,08	3,46	13,41
4	9,80	5,81	33,71	59,26	1,89	5,88	25,46
5	14,71	4,86	23,65	33,07	-0,13	6,43	21,51

Comentários

- A rigor, duas médias populacionais (quase) nunca são iguais.
- Algumas vezes, tem-se o interesse em testar se a diferença em módulo entre elas é maior que um certo valor positivo:

$$H_0 : |\mu_1 - \mu_2| < \Delta \text{ vs } H_1 : |\mu_1 - \mu_2| > \Delta.$$

- No exemplo em questão é preferível, em princípio, afirmar que duas turbinas são equivalentes do que concluir que uma seja melhor do que a outra.

Modelo normal

$$Y_{ij} \stackrel{ind.}{\sim} N(\mu_i, \sigma^2),$$

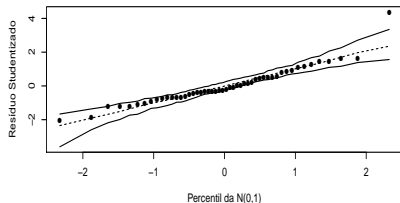
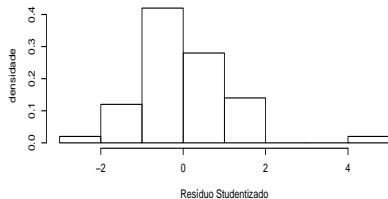
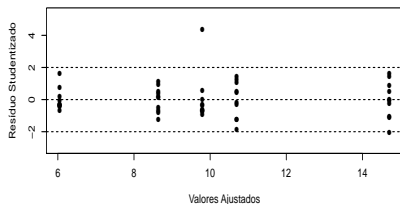
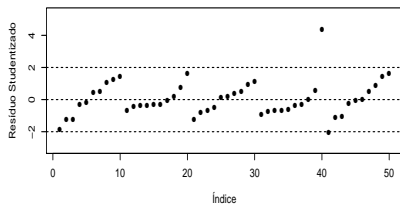
$$i = 1, \dots, 5 \text{ (tipo de turbina)}$$

$$j = 1, 2, \dots, 10 \text{ (turbina - unidade experimental)}$$

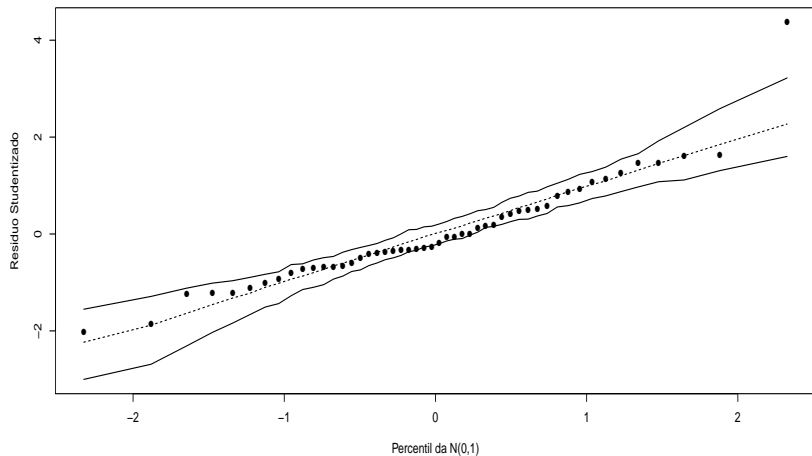
$$\mu_i = \alpha + \beta_i, \beta_1 = 0$$

- $\beta_i = \mu_i - \mu_1, i = 2, \dots, 5$: incremento (aditivo) da média do tipo de turbina i com relação ao tipo de turbina 1 (referência).
- Alguns resultados do ajuste - AIC = 298,37; BIC = 309,84.

Gráficos de diagnóstico: Modelo normal



Envelope para os resíduos: Modelo normal



Modelo gama 1

$$Y_{ij} \stackrel{ind.}{\sim} \text{gama}(\mu_i, \phi),$$

$$i = 1, \dots, 5 \text{ (tipo de turbina)}$$

$$j = 1, 2, \dots, 10 \text{ (turbina - unidade experimental)}$$

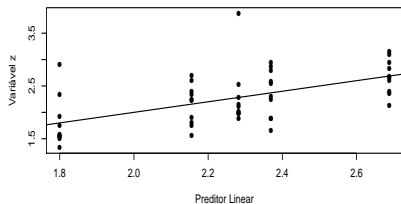
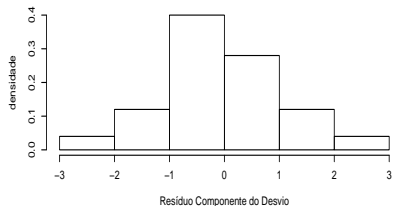
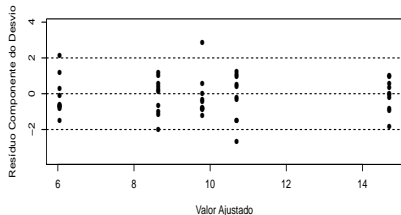
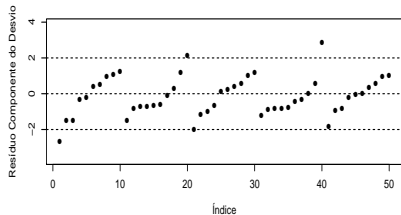
$$\ln \mu_i = \alpha + \beta_i, \beta_1 = 0$$

$$\mathcal{E}(Y_{ij}) = \mu_i \quad ; \quad \mathcal{V}(Y_{ij}) = \frac{\mu_i^2}{\phi}$$

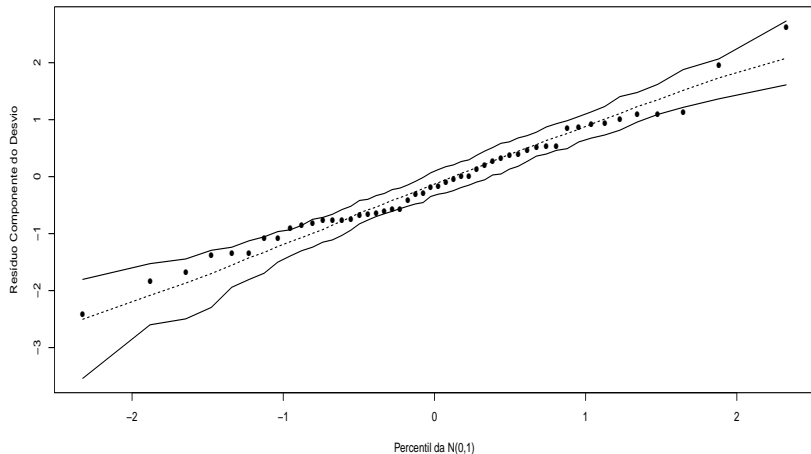
Modelo gama 1

- $e^{\beta_i} = \mu_i / \mu_1, i = 2, \dots, 5$: incremento (multiplicativo) da média do tipo de turbina i com relação ao tipo de turbina 1 (referência).
- Alguns resultados do ajuste: AIC = 285,89; BIC = 297,36.
 $D(\mathbf{y}; \tilde{\boldsymbol{\mu}}) = 51,43(p < 0,2365)$ (considerando-se a aproximação pela $\chi^2_{(45)}$ adequada) o que indica que o modelo se ajustou de modo satisfatório aos dados.

Gráficos de diagnóstico: Modelo gama 1



Envelope para os resíduos: Modelo gama 1



Modelo gama 2

$$Y_{ij} \stackrel{ind.}{\sim} \text{gama}(\mu_i, \phi),$$

$$i = 1, \dots, 5 \text{ (tipo de turbina)}$$

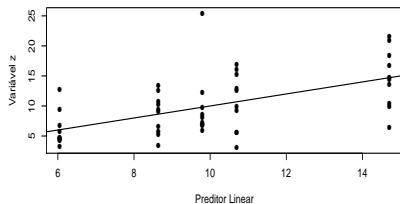
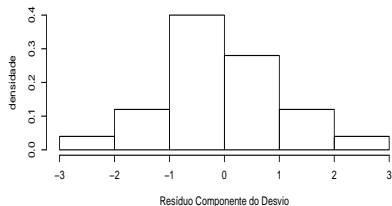
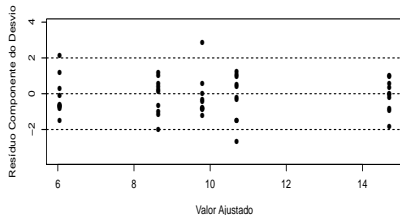
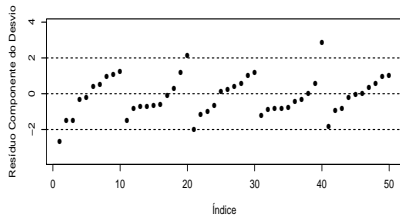
$$j = 1, 2, \dots, 10 \text{ (turbina - unidade experimental)}$$

$$\mu_i = \alpha + \beta_i, \beta_1 = 0$$

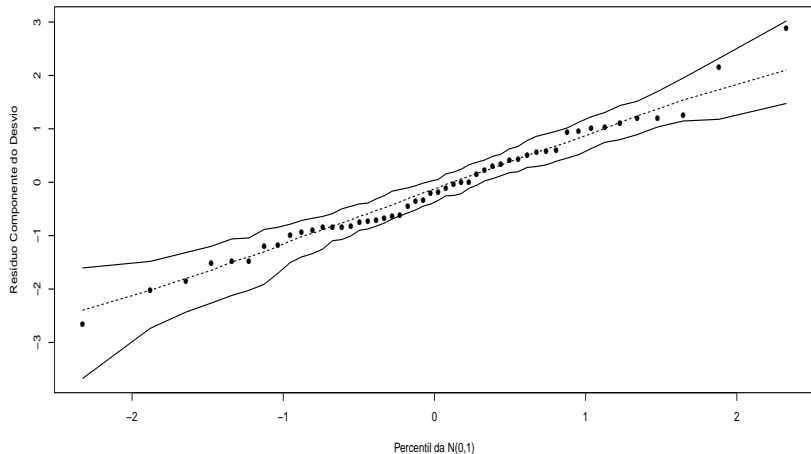
$$\mathcal{E}(Y_{ij}) = \mu_i \quad ; \quad \mathcal{V}(Y_{ij}) = \frac{\mu_i^2}{\phi}$$

- AIC = 285,89; BIC = 297,36. $D(\mathbf{y}; \tilde{\boldsymbol{\mu}}) = 51,43(p = 0,2365)$ (considerando-se a aproximação pela $\chi^2_{(45)}$ adequada), o que indica que o modelo se ajustou de modo satisfatório aos dados.

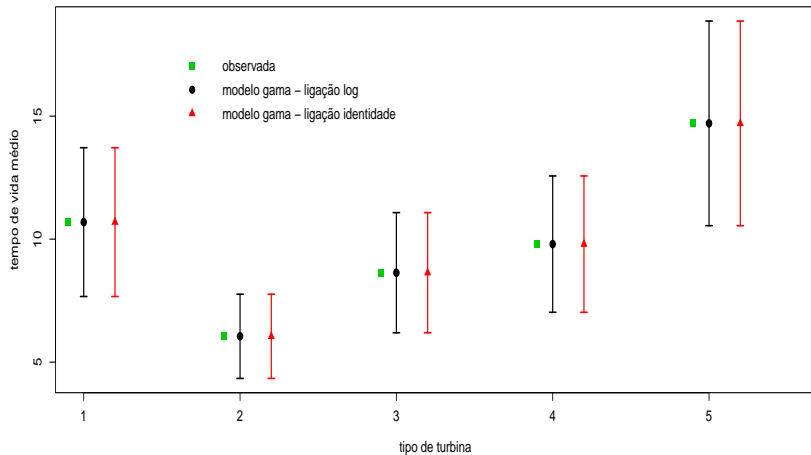
Gráficos de diagnóstico: Modelo gama 2



Envelope para os resíduos: Modelo gama 2



Médias previstas pelos modelos gama



Conclusões

- Os intervalos de confiança para as médias (figura anterior), no caso do modelo gama 2, têm de ser calculados através do método delta, através da exponenciação dos limites do IC (calculados para $\ln(\mu_i)$), ou algum procedimento numérico.
- Os modelos gama se ajustaram aos dados melhor do que o modelo normal.
- O modelo gama 2 apresentou uma ajuste levemente melhor do que o modelo gama 1.

Conclusões

- Escolheremos o modelo com função de ligação identidade, também pela facilidade de interpretação dos parâmetros.
- O resultado anterior (equivalência na predição das médias) é esperado dado que os modelos gama diferem apenas em termos da função de ligação e tem-se apenas fatores (um fator).

Estimativas dos parâmetros: Modelo gama 2

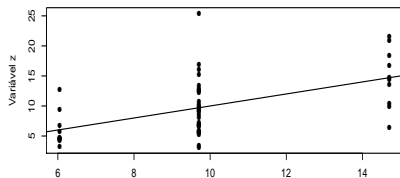
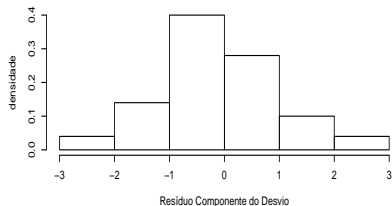
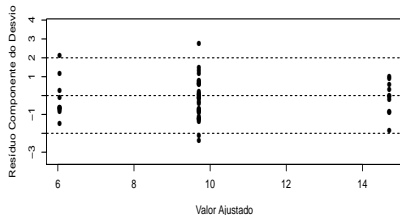
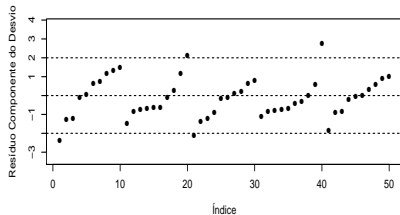
Parâmetro	Estimativa	EP	IC(95%)	Estat. Z	p-valor
α	10,69	1,40	[7,94 ; 13,44]	7,62	<0,0001
β_2	-4,64	1,61	[-7,80 ; -1,48]	-2,88	0,0040
β_3	-2,06	1,80	[-5,59 ; 1,48]	-1,14	0,2542
β_4	-0,89	1,90	[-4,63 ; 2,84]	-0,47	0,6383
β_5	4,01	2,39	[-0,66 ; 8,69]	1,68	0,0927
ϕ	5,80	1,13	[3,59 ; 8,02]	-	-

Modelo gama é preferível ao modelo exponencial. Aparentemente os tipos de turbina 1, 3 e 4 são equivalentes entre si, o tipo 2 é o pior e o tipo 5 o melhor.

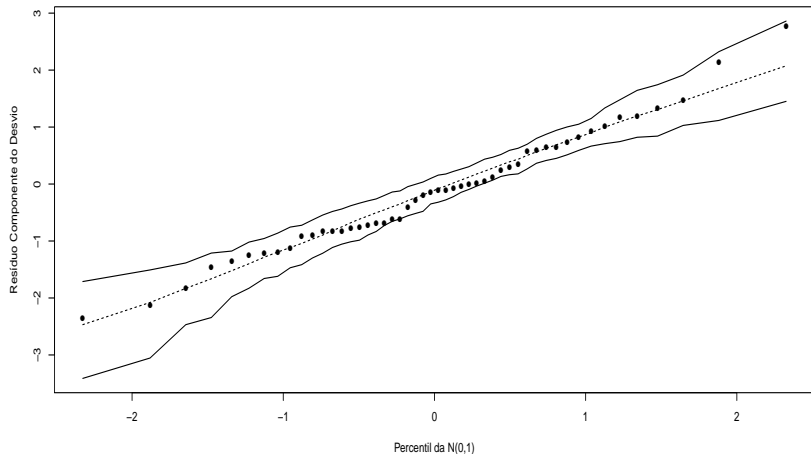
Modelo gama 3

- Ajustaremos um modelo reduzido (modelo gama 2 com $\beta_3 = \beta_4 = 0$, doravante modelo gama 3).
- $AIC = 283,21$; $BIC = 290,85$. $D(\mathbf{y}; \tilde{\boldsymbol{\mu}}) = 51,47(p = 0,3032)$ (considerando-se a aproximação pela $\chi^2_{(47)}$ adequada), o que indica que o modelo se ajustou de modo satisfatório aos dados.
- Os resultados acima indicam que o modelo 3 se ajustou melhor que o modelo 2.

Gráficos de diagnóstico: Modelo gama 3



Envelope para os resíduos: Modelo gama 3

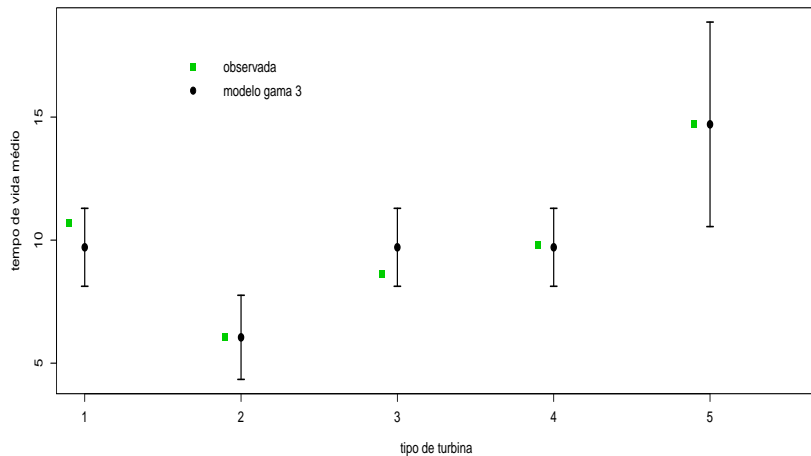


Estimativas dos parâmetros: Modelo gama 3

Parâmetro	Estimativa	EP	IC(95%)	Estat. Z	p-valor
α	9,71	0,74	[8,25 ; 11,17]	13,03	< 0,0001
β_2	-3,66	1,10	[-5,81 ; -1,51]	-3,34	0,0008
β_5	5,00	2,09	[0,90 ; 9,10]	2,39	0,0169
ϕ	5,66	1,10	[3,50 ; 7,82]	-	-

Escolheremos o modelo 3 como modelo final, fazendo a ressalva ao pesquisador que seria interessante testar outras turbinas dos tipos 1 e 5, para termos mais informação a respeito da possível diferença entre elas. Caso não fosse possível, em princípio, ajustariamos um outro modelo reduzido (modelo 3 com $\beta_5 = 0$) e o compararíamos com o modelo 3.

Médias previstas pelo modelo gama 3

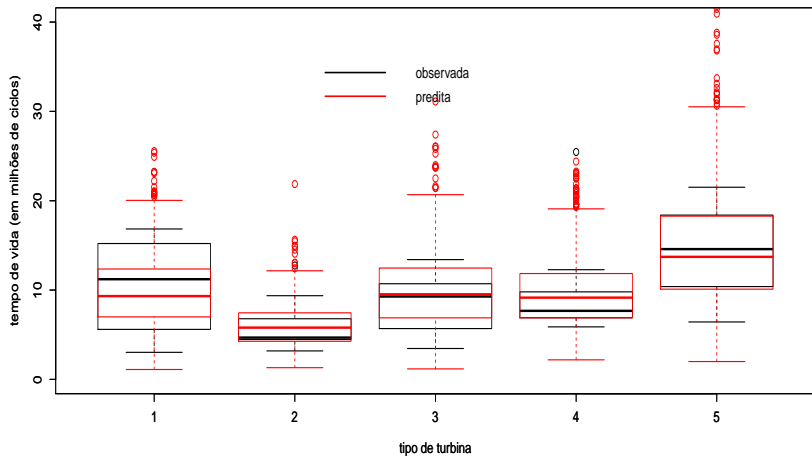


Médias previstas pelo modelo gama 3

Tipo de turbina	Estimativa	EP	IC(95%)
1	9,71	0,81	[8,13 ; 11,29]
2	6,05	0,87	[4,34 ; 7,76]
3	9,71	0,81	[8,13 ; 11,29]
4	9,71	0,81	[8,13 ; 11,29]
5	14,71	2,12	[10,55 ; 18,86]

O melhor tipo de turbina é o 5, o pior é o tipo 2, enquanto que os outros três tipos possuem desempenho semelhante entre si, e intermediário em relação aos dois primeiros tipos.

Distribuições observadas e previstas pelo modelo gama 3



Médias previstas pelo modelo gama 3

- Alguns modelos alternativos: normal inverso (MLG), normal assimétrico, t de Student assimétrico, Birnbaum-Saunders, Weibull, log-normal.
- Exercício: repetir a análise considerando o modelo normal inverso.
- Aparentemente, para o modelo normal inverso, a estimativa default do R, para o parâmetro ϕ também equivale à estimativa pelo método dos momentos.