

Modelo linear generalizado

Distribuição binomial

2022

Um modelo binomial com função de ligação canônica é ajustado aos dados do Exemplo 5, pag. 16 do livro Demétrio, C. G. B. (2002), *Modelos Lineares Generalizados em Experimentação Agronômica*, ESALQ (<https://docs.ufpr.br/~niveam/micro%20da%20sala/bom/Apostila%20de%20MLG.pdf>).

O modelo relaciona a probabilidade de que um inseto (pulgão) seja morto com a dose do inseticida (rotenone). Para cada dose foram utilizados m insetos no experimento dos quais y morreram após um certo tempo de exposição ao inseticida. Seis doses diferentes foram testadas. Esse tipo de estudo é conhecido como ensaio dose-resposta.

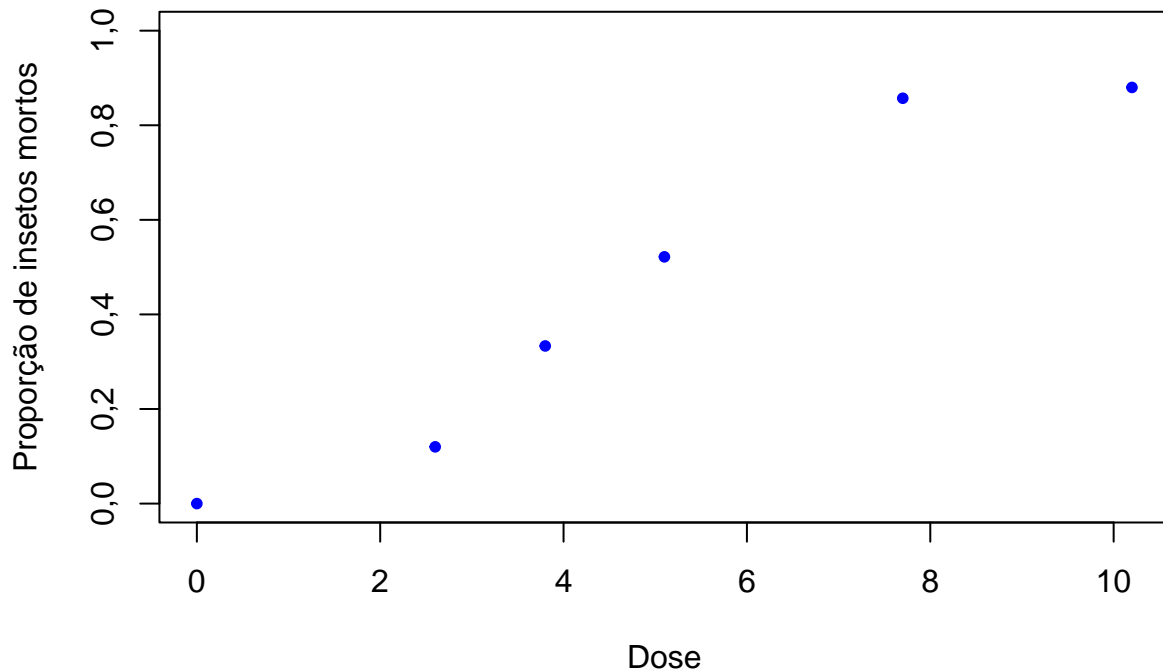
A linguagem R é utilizada no exemplo.

```
# Separador decimal nos resultados: ","  
options(OutDec = ",")
```

```
# Dados  
dose <- c(10.2, 7.7, 5.1, 3.8, 2.6, 0.0)  
m <- c(50, 49, 46, 48, 50, 49)  
y <- c(44, 42, 24, 16, 6, 0)  
dados <- cbind(y, m - y)
```

O objeto dados representa a variável resposta e tem duas colunas: número de insetos mortos e número de insetos sobreviventes.

```
# Gráfico de dispersão  
pmortos <- y / m  
plot(dose, pmortos, pch = 20, ylab = "Proporção de insetos mortos",  
      xlab = "Dose", ylim = c(0, 1), col = "blue")
```



```
# Ligação canônica (default)
```

```
mcan <- glm(dados ~ dose, family = binomial)
summary(mcan)
```

```
##
## Call:
## glm(formula = dados ~ dose, family = binomial)
##
## Deviance Residuals:
##      1       2       3       4       5       6
## -1,9456  0,9145  0,7679  0,7507 -0,8157 -1,9540
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -3,22566    0,36992  -8,720  <2e-16 ***
## dose         0,60513    0,06781   8,923  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0,001 '**' 0,01 '*' 0,05 '.' 0,1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##    Null deviance: 163,745  on 5  degrees of freedom
## Residual deviance:  10,258  on 4  degrees of freedom
## AIC: 33,479
##
## Number of Fisher Scoring iterations: 5
```

O objeto mcan contém diversos resultados do ajuste.

```
names(mcan)
```

```
## [1] "coefficients"      "residuals"        "fitted.values"
## [4] "effects"          "R"                 "rank"
## [7] "qr"               "family"           "linear.predictors"
```

```
## [10] "deviance"          "aic"          "null.deviance"
## [13] "iter"                "weights"      "prior.weights"
## [16] "df.residual"        "df.null"      "y"
## [19] "converged"          "boundary"     "model"
## [22] "call"                "formula"      "terms"
## [25] "data"                "offset"       "control"
## [28] "method"              "contrasts"    "xlevels"
```

Se o modelo binomial faz um bom ajuste aos dados, para n fixo, pode ser provado que a distribuição limite da desviância D_p é χ_{n-p}^2 quando $m_i \rightarrow \infty$ e $m_i\pi_i(1 - \pi_i) \rightarrow \infty$, $i = 1, \dots, n$. No exemplo, temos $n = 6$ e m variando de 46 a 50. O valor de D_p no modelo ajustado é denotado por d_p . Usando a distribuição aproximada e lembrando que quanto maior o valor de D_p , pior o ajuste, calculamos o valor- p como

$$P(D_p \geq d_p) = 1 - P(D_p < d_p) = 1 - P(D_p \leq d_p), \text{ em que } D_p \sim \chi_{n-p}^2. \quad (1)$$

```
pchisq(mcan$deviance, mcan$df.residual, lower.tail = FALSE)
```

```
## [1] 0,03629407
```

Com 1% de significância, não rejeitamos a hipótese de que o modelo faz um bom ajuste aos dados.

Nota 1. Sobreponha a curva ajustada ao gráfico da figura à pag. 1.

```
## Intervalos de confiança assintóticos
# Aproximação pela dist. normal
conf <- 0.95 # coeficiente de confiança
z <- qnorm((1 + conf) / 2) # normal (z = 1,96)
smcan <- summary(mcan)
ep <- smcan$coefficients[, "Std. Error"]
icnorm <- cbind(mcan[["coefficients"]] - z * ep,
               mcan[["coefficients"]] + z * ep)
colnames(icnorm) <- c("Inf.", "Sup.")
print(icnorm, digits = 3)
```

```
##                Inf.    Sup.
## (Intercept) -3,951 -2,501
## dose         0,472  0,738
```

```
# Intervalos de confiança (veros. perfilada)
confint(mcan)
```

```
## Waiting for profiling to be done...
```

```
##                2,5 %    97,5 %
## (Intercept) -3,9952937 -2,540587
## dose         0,4807772  0,747802
```

Razão de chances

Temos $Y \sim \text{binomial}(m, \pi)$ e $E(Y) = \mu = m\pi$. De acordo com o modelo temos que

$$\log\left(\frac{\mu}{m - \mu}\right) = \log\left(\frac{\pi}{1 - \pi}\right) = \beta_1 + \beta_2 \text{ dose}, \quad (2)$$

em que π denota a probabilidade de que um inseto morra e $\pi/(1 - \pi)$ é a chance (*odds*) de que um inseto morra, notando que a chance é uma função monótona crescente em π . O lado esquerdo de (2) é chamado de logito da probabilidade π (função de ligação canônica).

Em seguida comparamos dois valores de doses, $dose_1$ e $dose_2$, sendo que $dose_2 = dose_1 + 1$, com probabilidades π_1 e π_2 , respectivamente. A partir de (2) obtemos

$$\log\left(\frac{\pi_1}{1-\pi_1}\right) = \beta_1 + \beta_2 \text{dose}_1 \quad (3)$$

e

$$\log\left(\frac{\pi_2}{1-\pi_2}\right) = \beta_1 + \beta_2 \text{dose}_2 = \beta_1 + \beta_2 (\text{dose}_1 + 1) = \beta_1 + \beta_2 \text{dose}_1 + \beta_2. \quad (4)$$

Subtraindo (3) de (4) obtemos

$$\log\left(\frac{\pi_2}{1-\pi_2}\right) - \log\left(\frac{\pi_1}{1-\pi_1}\right) = \beta_2, \quad (5)$$

ou seja,

$$\log(RC) = \beta_2 \quad \text{e} \quad RC = e^{\beta_2}, \quad (6)$$

em que

$$RC = \frac{\pi_2}{1-\pi_2} \bigg/ \frac{\pi_1}{1-\pi_1} \quad (7)$$

é a razão de chances (*odds ratio*). Portanto, a chance de um inseto morrer é multiplicada por e^{β_2} quando a dose é aumentada em uma unidade, qualquer que seja a dose. Desta forma, o coeficiente da variável $dose$ (β_2) tem uma interpretação.

Em um modelo com mais de uma covariável ($p > 2$), a interpretação de cada coeficiente β_j , $j \neq 1$, é semelhante tomando-se a covariável x_j acrescida de uma unidade ($x_j + 1$) e mantendo-se constantes os valores de todas as demais covariáveis.

```
# Razão de chances
```

```
coef(mcan)
```

```
## (Intercept)      dose
## -3,2256633    0,6051256
```

```
cat("\n Razão de chances =", exp(coef(mcan)[2]))
```

```
##
```

```
## Razão de chances = 1,831482
```

Nota 2. Apresente um intervalo de confiança para a razão de chances.

Dose letal

É de interesse estimar as doses de inseticida que eliminam 50% e 90% dos insetos, denotadas por DL_{50} e DL_{90} , respectivamente. Segue de (2) que

$$\text{dose} = \frac{\log\left(\frac{\pi}{1-\pi}\right) - \beta_1}{\beta_2}. \quad (8)$$

Da expressão (8) obtemos

$$DL_{50} = \frac{\log\left(\frac{0,5}{1-0,5}\right) - \beta_1}{\beta_2} = -\frac{\beta_1}{\beta_2} \quad (9)$$

e

$$DL_{90} = \frac{\log\left(\frac{0,9}{1-0,9}\right) - \beta_1}{\beta_2} = \frac{2,20 - \beta_1}{\beta_2}. \quad (10)$$

```

# Dose letal
coef(mcan)

## (Intercept)      dose
## -3,2256633    0,6051256

DL <- function(prob) {
  return((log(prob / (1 - prob)) - coef(mcan)[1]) / coef(mcan)[2])
}

cat("\n DL50 e DL90 = ", DL(c(0.5, 0.9)))

##
## DL50 e DL90 = 5,330569 8,961591

```

Nota 3. Apresente um intervalo de confiança para a dose letal.

Nota 4. As funções de ligação logito (regressão logística), probito e complemento log-log são bastante utilizadas no modelo binomial. As expressões são dadas abaixo, sendo que $\eta \in \mathbb{R}$ é o preditor linear, $\pi = \mu/m$ e Φ denota a função distribuição acumulada da distribuição normal(0, 1).

$$\log\left(\frac{\pi}{1-\pi}\right) = \eta, \quad \pi = \frac{e^\eta}{1+e^\eta} : \text{ logito,} \quad (11)$$

$$\Phi^{-1}(\pi) = \eta, \quad \pi = \Phi(\eta) : \text{ probito,} \quad (12)$$

$$\log(-\log(1-\pi)) = \eta, \quad \pi = 1 - \exp(-\exp(\eta)) : \text{ complemento log-log.} \quad (13)$$

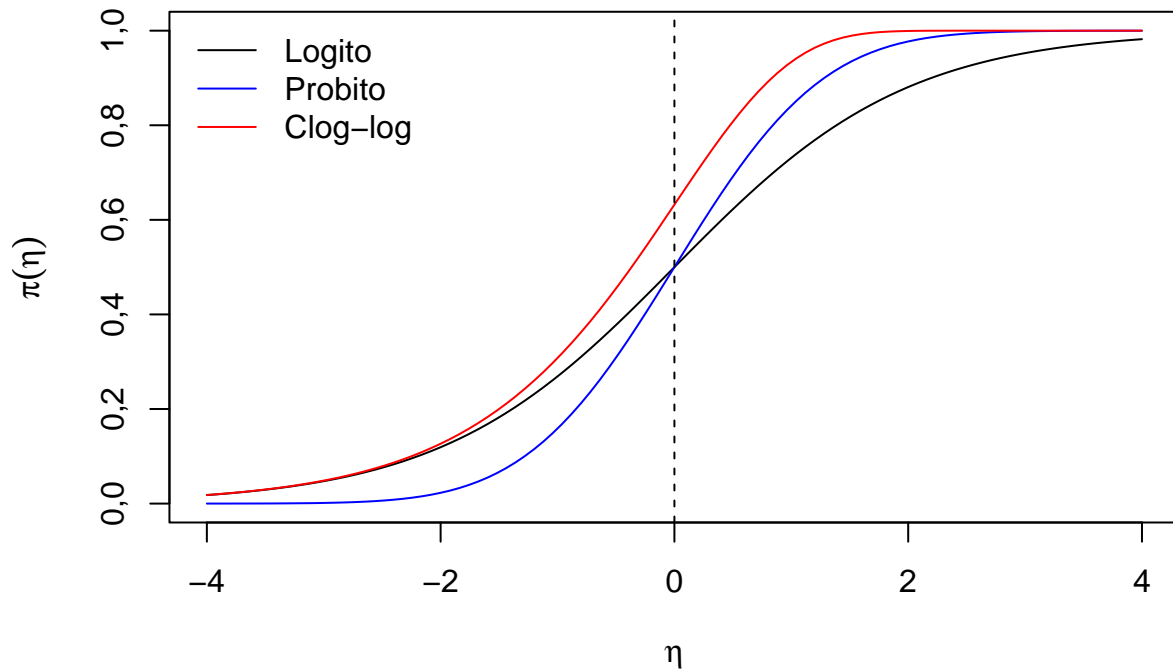
As funções de ligação logito e probito são simétricas, ou seja, $\pi(\eta) = 1 - \pi(-\eta)$ para todo $\eta \in \mathbb{R}$.

```

## Inversa das funções de ligação
# Complemento log-log (Clog-log)
ploglog <- function(x) {
  return(1 - exp(-exp(x)))
}

curve(plogis(x), from = -4, to = 4, ylim = c(0, 1),
      xlab = expression(eta), ylab = expression(pi(eta)))
curve(pnorm(x), add = TRUE, col = "blue")
curve(ploglog(x), add = TRUE, col = "red")
abline(v = 0, lty = 2)
legend("topleft", c("Logito", "Probito", "Clog-log"), col = c("black", "blue", "red"),
      lty = 1, bty = "n")

```



Ajuste modelos com as funções de ligação probito e complemento log-log utilizando `family = binomial(link = "probit")` e `family = binomial(link = "cloglog")`, respectivamente.

Nota 5. Refaça o exemplo em linguagem Python.