

## MAE5911/IME: Fundamentos de Estatística e Machine Learning. Prof.: Alexandre Galvão Patriota

**Questão 01:** Defina formalmente o valor-p (**aproximado**) para testar uma hipótese geral e discuta os problemas de interpretação usuais ao utilizar a sua versão condicional à hipótese nula. Apresente um exemplo numérico e simulações de Monte Carlo, usando um modelo de regressão Poisson, para ilustrar a sua aplicação.

**Questão 02:** Apresente um Teorema da Aproximação Universal que generalize o Teorema de Funahashi, discuta as diferenças entre os resultados por meio de exemplos de redes neurais.

**Questão 03:** Proponha uma função de estimação robusta para estimar os parâmetros de uma regressão Binomial,  $Y|X = x \sim \text{Bin}(m, \mu_\theta(x))$ , em que  $\mu_\theta(\cdot)$  é uma função com imagem em  $[0, 1]$  e  $\theta$  é o vetor de parâmetros. Apresente um exemplo numérico e uma simulação de Monte Carlo, como feito em sala, para ilustrar os resultados. Perturbe a distribuição dos dados e mostre que a sua proposta é de fato robusta contra essa perturbação.

### Resposta para o Questão 01

O conceito do p-valor, tal como proposto originalmente por *Fisher* (1925), é uma medida **condicional sob  $H_0$** , definida por

$$\text{valor-p}(H_0, z_n) = \sup_{\theta \in \Theta_0} P_\theta^{(n)}(T_{H_0}(Z_n) \geq T_{H_0}(z_n)),$$

isto é, a probabilidade de se observar uma estatística tão extrema quanto a obtida, assumindo  $H_0$  verdadeira. O valor-p foi concebido por Fisher como um *instrumento de contraste* entre o dado e hipótese, mas não como uma medida de verdade. Seu objetivo era expressar o grau de incompatibilidade empírica entre o fenômeno observado e as consequências lógicas de  $H_0$ .

Posteriormente, a escola *Neyman–Pearson* reinterpretou esse conceito dentro de uma estrutura de decisão. Nessa transição, o valor-p perdeu seu caráter exploratório e passou a ser tratado como um critério binário, externo e supostamente não condicionado a  $H_0$ , capaz de expressar a decisão: “rejeita” ou “não rejeita”  $H_0$ . Esta leitura gerou um equívoco lógico: o valor-p, que é definido dentro do universo em que  $H_0$  é verdadeira, passou a ser interpretado como se pudesse mensurar a probabilidade de  $H_0$  ser verdadeira.

O **valor-p assintótico**, que é o valor usualmente utilizado, na ausência da distribuição exata de  $T_{H_0}$ , é definido por:

$$p^{(a)}(H_0, z_n) = \sup_{\theta \in \Theta_0} P_\theta(W_\theta \geq T_{H_0}(z_n)),$$

em que  $W_\theta$  representa a distribuição-limite de  $T_{H_0}$  quando  $n \rightarrow \infty$ .

Confirmando essa natureza epistêmica, o valor-p assintótico nem sempre converge para o valor-p exato, isto porque, estando definido condicionalmente a  $H_0$ , pode não ser sensível a falhas estruturais do modelo, como a ausência de independência entre as observações, a presença de hipóteses condicionais aninhadas, ou a violação das condições de regularidade que garantem a validade assintótica — como homocedasticidade.

Por isso, o valor-p deve ser entendido apenas como originalmente proposto — uma **medida de discrepância entre o real e o teórico** — que opera em um único sentido: como evidência contra a hipótese nula, mas nunca para confirmá-la.

Em termos das  $\sigma$ -álgebras envolvidas, a construção inferencial pode ser representada como uma sequência de aplicações mensuráveis, onde cada espaço de probabilidade possui a sua própria  $\sigma$ -álgebra, refletindo níveis distintos de abstração:

$$(\Omega_n, \mathcal{F}_n, \mathbb{P}_n) \xrightarrow{Z_n} (\mathcal{Z}_n, \mathcal{B}_{\mathcal{Z}_n}, \mathbb{P}_{\mathcal{Z}_n}) \xrightarrow{T} (\mathcal{T}_n, \mathcal{B}_{\mathcal{T}_n}, \mathbb{P}_{\mathcal{T}_n}).$$

- $\mathcal{F}_n$  — descreve o universo empírico de incertezas, isto é, os eventos do mundo observável;

- $\mathcal{B}_{Z_n}$  — corresponde ao espaço amostral modelado, onde as variáveis aleatórias  $Z_n$  são definidas sob  $\mathbb{P}_{Z_n}$ ;
- $\mathcal{B}_{T_n}$  — é a  $\sigma$ -álgebra induzida pela estatística  $T$ , onde vivem as distribuições teóricas dos testes e, em particular, o valor-p.

O valor-p pertence ao universo lógico em que a hipótese nula  $H_0$  é assumida como verdadeira. Ele não pertence à  $\sigma$ -álgebra empírica  $\mathcal{F}_n$ , e portanto não pode ser interpretado como uma probabilidade sobre o mundo real ou sobre a veracidade de  $H_0$ . Tratá-lo dessa forma constitui um erro de violação da  $\sigma$ -álgebra — um deslize epistemológico em que se tenta extrair informação empírica de uma construção em sua definição epistêmica condicional e autoreferente.

## Implementações

O exemplo numérico apresentado na Listing 1 tem o objetivo de ilustrar o comportamento do valor- $p$  sob diferentes hipóteses nulas, mantendo fixos os dados observados. O experimento baseia-se em um modelo de regressão de Poisson definido por

$$Y_i \sim \text{Poisson}(\lambda_i), \quad \text{com} \quad \log(\lambda_i) = \theta_0 + \theta x_i,$$

onde  $\theta_0 = 1$  e o valor verdadeiro do parâmetro é  $\theta = 0,20$ . Uma única amostra de tamanho  $n = 100$  foi gerada segundo esse modelo, representando o conjunto de dados “observados” em um mundo empírico fixo.

A partir dessa amostra, foram testadas três hipóteses nulas distintas sobre o mesmo conjunto de dados:

$$H_0^{(A)} = 0,20, \quad H_0^{(B)} = 0,28, \quad H_0^{(C)} = 0,60.$$

Cada hipótese foi avaliada por meio da estatística de Wald,

$$Z = \frac{\hat{\theta} - \theta_{H_0}}{\text{se}(\hat{\theta})},$$

e o valor- $p$  correspondente foi calculado de forma:

$$p = P(Z \leq z_{\text{obs}}),$$

correspondendo à hipótese alternativa  $H_1 : \theta < \theta_{H_0}$ .

Os resultados obtidos encontram-se na Tabela 1. Observa-se que, quando a hipótese nula coincide com o valor real do parâmetro ( $\theta$ ), o valor- $p$  é elevado e a hipótese não é rejeitada. À medida que o valor hipotético se afasta de  $\theta$ , o valor- $p$  decresce rapidamente, levando à rejeição de  $H_0$  nos casos de maior discrepância ( $H_0^{(C)}$ ). Essa simulação evidencia o papel do valor- $p$  como medida de compatibilidade entre modelo e dado: ele quantifica o grau de discrepância entre a hipótese estatística e a estrutura empírica observada.

Tabela 1: Resultados do cálculo do valor- $p$  para diferentes hipóteses nulas sob o modelo de regressão de Poisson.

Cenário	Hipótese nula	Valor- $p$	Decisão	$\hat{\theta}$	$\theta$
Cenário A ( $H_0^{(A)}$ exato)	$\theta_{H_0} = 0,20$	0,547473	Não rejeita $H_0$	0,211726	0,20
Cenário B ( $H_0^{(B)}$ com pequeno desvio)	$\theta_{H_0} = 0,28$	0,243689	Não rejeita $H_0$	0,211726	0,20
Cenário C ( $H_0^{(C)}$ com grande desvio)	$\theta_{H_0} = 0,60$	0,000039	Rejeita $H_0$	0,211726	0,20

O experimento de Monte Carlo apresentado na Listing ?? tem como objetivo avaliar, de forma empírica, o comportamento assintótico do valor- $p$  quando múltiplas amostras são geradas a partir de um mesmo modelo de regressão de Poisson. Considera-se o modelo

$$Y_i \sim \text{Poisson}(\lambda_i), \quad \text{com} \quad \log(\lambda_i) = \theta_0 + \theta x_i,$$

onde o intercepto  $\theta_0 = 1$  e o valor verdadeiro do parâmetro é  $\theta = 0,20$ . Em cada replicação, é gerada uma amostra de tamanho  $n = 100$  com covariáveis  $x_i \sim \text{Unif}(0, 2)$  e respostas  $Y_i$  conforme o modelo acima. O processo é repetido  $M = 10,000$  vezes, mantendo  $\theta$  fixo.

Para cada amostra, ajusta-se um modelo de regressão de Poisson e testam-se três hipóteses nulas distintas sobre o mesmo conjunto de dados:

$$H_0^{(A)} = 0,20, \quad H_0^{(B)} = 0,28, \quad H_0^{(C)} = 0,60.$$

Cada hipótese é avaliada pela estatística de Wald,

$$Z = \frac{\hat{\theta} - \theta_{H_0}}{\text{se}(\hat{\theta})},$$

e o valor- $p$  correspondente é calculado da seguinte forma, considerando o teste contra a alternativa  $H_1 : \theta < \theta_{H_0}$ :

$$p = P(Z \leq z_{\text{obs}}).$$

A partir das  $M$  replicações, calcula-se a proporção de rejeição de cada hipótese, a média e a mediana dos valores- $p$ , bem como a média dos estimadores  $\hat{\theta}$ . Os resultados são apresentados na Tabela 2.

Observa-se que, quando  $H_0^{(A)}$  coincide com o valor verdadeiro do parâmetro, a proporção de rejeição aproxima-se do nível  $\alpha = 0,05$ , e os valores- $p$  seguem aproximadamente uma distribuição uniforme. À medida que as hipóteses nulas se deslocam para valores superiores ao verdadeiro — como em  $H_0^{(B)}$  e  $H_0^{(C)}$  — o teste unicaudal à esquerda ganha poder, e os valores- $p$  tornam-se progressivamente menores. Nos casos de grande discrepância ( $H_0^{(C)}$ ), a rejeição de  $H_0$  é praticamente certa, refletindo a coerência do teste de Wald com o comportamento teórico esperado sob o modelo de Poisson.

Tabela 2: Resultados da simulação de Monte Carlo para o teste de Wald em regressão de Poisson. Cada linha apresenta os resultados obtidos para uma hipótese nula distinta, com  $M = 10,000$  replicações e  $n = 100$ .

Cenário	H0	% de rejeição	Média( $p$ )	Mediana( $p$ )	$p_{0,05}$	$p_{0,95}$	$\bar{\hat{\theta}} / \theta$
Cenário A	$H_0^{(A)} = 0,20$	0,053700	0,494656	0,488848	0,045770	0,950321	0,1983 / 0,2
Cenário B	$H_0^{(B)} = 0,28$	0,204100	0,282815	0,204148	0,006543	0,804670	0,1983 / 0,2
Cenário C	$H_0^{(C)} = 0,60$	0,989600	0,002594	0,000030	0,000000	0,010802	0,1983 / 0,2

```

1 set.seed(111)
2
3 n      = 100
4 beta0  = 1.0
5 theta  = 0.20      # valor verdadeiro
6 alpha  = 0.05
7
8 # Covariável e única amostra observada
9 x      = runif(n, 0, 2)
10 lambda = exp(beta0 + theta * x)
11 y      = rpois(n, lambda)
12
13 # Ajuste único do GLM Poisson
14 modelo = glm(y ~ x, family = poisson)
15
16 # Função: testa H0
17 testa_glm_theta = function(mod, h0, alpha = 0.05) {
18   theta_hat = coef(mod)[2]
19   se_hat    = sqrt(vcov(mod)[2, 2])
20   Z         = (theta_hat - h0) / se_hat
21
22   # valor-p unicaudal (contra H1: theta < h0)
23   p_valor   = pnorm(Z)
24
25   decisao   = ifelse(p_valor < alpha, "Rejeita H0", "Não rejeita H0")
26
27   data.frame(
28     hipotese_nula = paste0("H0: theta = ", format(h0, nsmall = 2)),
29     valor_p       = p_valor,

```

```

30     decisao      = decisao,
31     theta_hat    = theta_hat,
32     theta        = theta
33   )
34 }
35
36 # Hipóteses nulas a testar no mesmo conjunto
37 HO_A = 0.20   # H0 exato
38 HO_B = 0.28   # H0 com pouco desvio
39 HO_C = 0.60   # H0 com muito desvio
40
41 resA = testa_glm_theta(modelo, HO_A, alpha)
42 resB = testa_glm_theta(modelo, HO_B, alpha)
43 resC = testa_glm_theta(modelo, HO_C, alpha)
44
45 # Consolida com rbind
46 resumo_glm = rbind(
47   "Cenário A (H0 exato)"      = resA,
48   "Cenário B (H0 pouco desvio)" = resB,
49   "Cenário C (H0 muito desvio)" = resC
50 )
51
52 # Formatação
53 options(scipen = 999)
54
55 resumo_fmt      = resumo_glm
56 resumo_fmt$valor_p = format(round(resumo_fmt$valor_p, 6), nsmall = 6)
57 resumo_fmt$theta_hat = format(round(resumo_fmt$theta_hat, 6), nsmall = 6)
58 resumo_fmt$theta = format(round(resumo_fmt$theta, 6), nsmall = 6)
59
60 # Tabela final (analítica)
61 noquote(resumo_fmt)

```

Listing 1: Exemplo numérico — Regressão de Poisson ( $n = 100$ ,  $\theta = 0,20$ ).

```

1  set.seed(111)
2
3  # -----
4  # Parâmetros do experimento
5  # -----
6  n      = 200      # tamanho da amostra
7  beta0  = 1.0      # intercepto fixo
8  theta  = 0.20     # (verdadeiro) = coef. de x, usado para gerar os dados
9  alpha  = 0.05     # nível
10 M      = 10000    # repetições de Monte Carlo
11
12 # hipóteses nulas a testar (todas no mesmo conjunto de dados por replicação)
13 HO_A = 0.20   # H0 exato
14 HO_B = 0.28   # H0 com pouco desvio
15 HO_C = 0.60   # H0 com grande desvio
16 HOs  = c(HO_A, HO_B, HO_C)
17
18 # -----
19 # Design fixo: x é gerado 1 vez e mantido fixo
20 # -----
21 x = runif(n, 0, 2)
22
23 # -----
24 # Função: calcula p-valor unicaudal (esquerda) e decisão para b0
25 # -----
26 p_unicaudal_wald = function(mod, b0, alpha) {
27   theta_hat = coef(mod)[2]
28   se_hat    = sqrt(vcov(mod)[2, 2])
29   Z         = (theta_hat - b0) / se_hat
30   # unicaudal à esquerda: H1: theta < b0
31   p_valor   = pnorm(Z)
32   decisao   = ifelse(p_valor < alpha, "Rejeita H0", "Não rejeita H0")
33   list(p = p_valor, decisao = decisao, theta_hat = theta_hat)
34 }

```

```

35 # -----
36 # Uma replicação: gera  $y \sim \text{Poisson}(\exp(\beta_0 + \theta * x))$ , ajusta GLM
37 # e testa  $H_0_A$ ,  $H_0_B$ ,  $H_0_C$  na mesma amostra
38 # -----
39
40 uma_replicacao = function() {
41   lambda = exp(beta0 + theta * x)
42   y      = rpois(n, lambda)
43   mod    = glm(y ~ x, family = poisson)
44   lapply(H0s, function(b0) p_unicaudal_wald(mod, b0, alpha))
45 }
46
47 # -----
48 # Roda M replicações e consolida resultados por cenário
49 # -----
50 # estruturas de acumulação
51 p_list_A = numeric(M); dec_A = character(M); thA = numeric(M)
52 p_list_B = numeric(M); dec_B = character(M); thB = numeric(M)
53 p_list_C = numeric(M); dec_C = character(M); thC = numeric(M)
54
55 for (m in 1:M) {
56   out = uma_replicacao()
57   # A
58   p_list_A[m] = out[[1]]$p
59   dec_A[m]    = out[[1]]$decisao
60   thA[m]      = out[[1]]$theta_hat
61   # B
62   p_list_B[m] = out[[2]]$p
63   dec_B[m]    = out[[2]]$decisao
64   thB[m]      = out[[2]]$theta_hat
65   # C
66   p_list_C[m] = out[[3]]$p
67   dec_C[m]    = out[[3]]$decisao
68   thC[m]      = out[[3]]$theta_hat
69 }
70
71 # -----
72 # Sumários: proporção de rejeição (potência/erro tipo I),
73 # estatísticas de p e média de theta_hat
74 # -----
75 resumo_mc = data.frame(
76   hipotese_nula = c(
77     paste0("H0_A: theta = ", format(H0_A, nsmall = 2)),
78     paste0("H0_B: theta = ", format(H0_B, nsmall = 2)),
79     paste0("H0_C: theta = ", format(H0_C, nsmall = 2))
80   ),
81   prop_rejeicao = c(
82     mean(dec_A == "Rejeita H0"),
83     mean(dec_B == "Rejeita H0"),
84     mean(dec_C == "Rejeita H0")
85   ),
86   media_p = c(mean(p_list_A), mean(p_list_B), mean(p_list_C)),
87   mediana_p = c(median(p_list_A), median(p_list_B), median(p_list_C)),
88   q05_p = c(quantile(p_list_A, 0.05), quantile(p_list_B, 0.05), quantile(p_list_C, 0.05)),
89   q95_p = c(quantile(p_list_A, 0.95), quantile(p_list_B, 0.95), quantile(p_list_C, 0.95)),
90   theta_hat_medio = c(mean(thA), mean(thB), mean(thC)),
91   theta_real = theta
92 )
93
94 # formatação
95 options(scipen = 999)
96 resumo_mc$prop_rejeicao = format(round(resumo_mc$prop_rejeicao, 6), nsmall = 6)
97 resumo_mc$media_p      = format(round(resumo_mc$media_p, 6), nsmall = 6)
98 resumo_mc$mediana_p    = format(round(resumo_mc$mediana_p, 6), nsmall = 6)
99 resumo_mc$q05_p        = format(round(resumo_mc$q05_p, 6), nsmall = 6)
100 resumo_mc$q95_p        = format(round(resumo_mc$q95_p, 6), nsmall = 6)
101 resumo_mc$theta_hat_medio = format(round(resumo_mc$theta_hat_medio, 6), nsmall = 6)
102 resumo_mc$theta_real    = format(round(resumo_mc$theta_real, 6), nsmall = 6)

```

```

103 # Impressao da tabela
104 rownames(resumo_mc) = c(
105   "Cenário A (H0_A exato)",
106   "Cenário B (H0_B pouco desvio)",
107   "Cenário C (H0_C muito desvio)"
108 )
109 )
110
111 noquote(resumo_mc)

```

Listing 2: Simulação de Monte Carlo do teste de Wald sob regressão de Poisson ( $M = 10,000$ ,  $n = 100$ ,  $\theta = 0,20$ ).

## Resposta para o Questão 02

## Resposta para o Questão 03

# 1 Regressão Binomial Robusta: Formulação e Evidências

## 1.1 Modelo e Construção do Experimento

Considerou-se o modelo binomial

$$Y_i | X_i \sim \text{Bin}(m, \mu_\theta(X_i)), \quad \text{com} \quad \mu_\theta(x) = \frac{1}{1 + e^{-(\theta_0 + \theta_1 x)}},$$

onde  $\theta = (\theta_0, \theta_1)$  é o vetor de parâmetros a ser estimado. Os parâmetros verdadeiros foram fixados em  $\theta_0 = -0.5$  e  $\theta_1 = 1.2$ , com  $n = 200$  observações e  $m = 10$  tentativas por ponto.

A variável explicativa foi gerada como  $X_i \sim U(-2, 2)$  e as respostas  $Y_i$  a partir do modelo binomial acima. Para introduzir contaminação, selecionaram-se 10% das observações com  $x_i < 0.5$  e aplicou-se a seguinte modificação adversária:

$$x_i \leftarrow x_i + 3, \quad y_i \leftarrow \begin{cases} 0, & \text{com probabilidade } 0.5, \\ m, & \text{com probabilidade } 0.5. \end{cases}$$

Esse procedimento gera simultaneamente *outliers de alavancagem* (no eixo  $x$ ) e *outliers de resposta* (no eixo  $y$ ), de modo análogo à estratégia demonstrada em sala.

## 1.2 Função de Estimação Robusta

O estimador clássico de máxima verossimilhança (MLE) foi obtido pela regressão logística usual:

$$\hat{\theta}_{\text{MLE}} = \arg \max_{\theta} \sum_{i=1}^n [y_i \log \mu_\theta(x_i) + (m - y_i) \log(1 - \mu_\theta(x_i))].$$

Para tornar a estimação robusta, definiu-se uma **função de estimação ponderada** baseada no resíduo  $r_i = y_i - m\mu_\theta(x_i)$ :

$$U_i(\theta) = \begin{pmatrix} w_i(y_i - m\mu_\theta(x_i)) \\ w_i x_i(y_i - m\mu_\theta(x_i)) \end{pmatrix}, \quad w_i = \exp(-(|r_i| - c)_+), \quad c = p \cdot \text{IQR}(r),$$

em que  $(a)_+ = \max(0, a)$  e  $p = 1.5$  controla o ponto de corte do intervalo interquartil. O estimador robusto  $\hat{\theta}_R$  é então obtido pela minimização do quadrado da soma das funções de estimação:

$$\hat{\theta}_R = \arg \min_{\theta} \left\| \sum_{i=1}^n U_i(\theta) \right\|^2.$$

Essa construção faz com que observações cujos resíduos ultrapassem o limite  $c = p \text{IQR}(r)$  recebam peso reduzido, atenuando sua influência no ajuste.

### 1.3 Simulação e Resultados

Foram realizadas  $R = 500$  repetições do experimento (Monte Carlo), calculando-se para cada método a média, o viés e o erro quadrático médio (RMSE) das estimativas. A Tabela 3 apresenta os resultados obtidos.

Tabela 3: Resumo da simulação Monte Carlo ( $R = 500$ ): comparação entre MLE e estimador robusto.

Estimador	Média	Viés	RMSE
MLE $\beta_0$	-0.561	-0.061	0.087
MLE $\beta_1$	0.845	-0.355	0.370
Robusto $\beta_0$	-0.484	0.016	0.068
Robusto $\beta_1$	1.281	0.081	0.104

### 1.4 Evidências Gráficas

A Figura 1 mostra as curvas ajustadas pelo MLE e pelo método robusto. Os pontos destacados em vermelho correspondem às observações atenuadas pelo corte baseado no IQR.

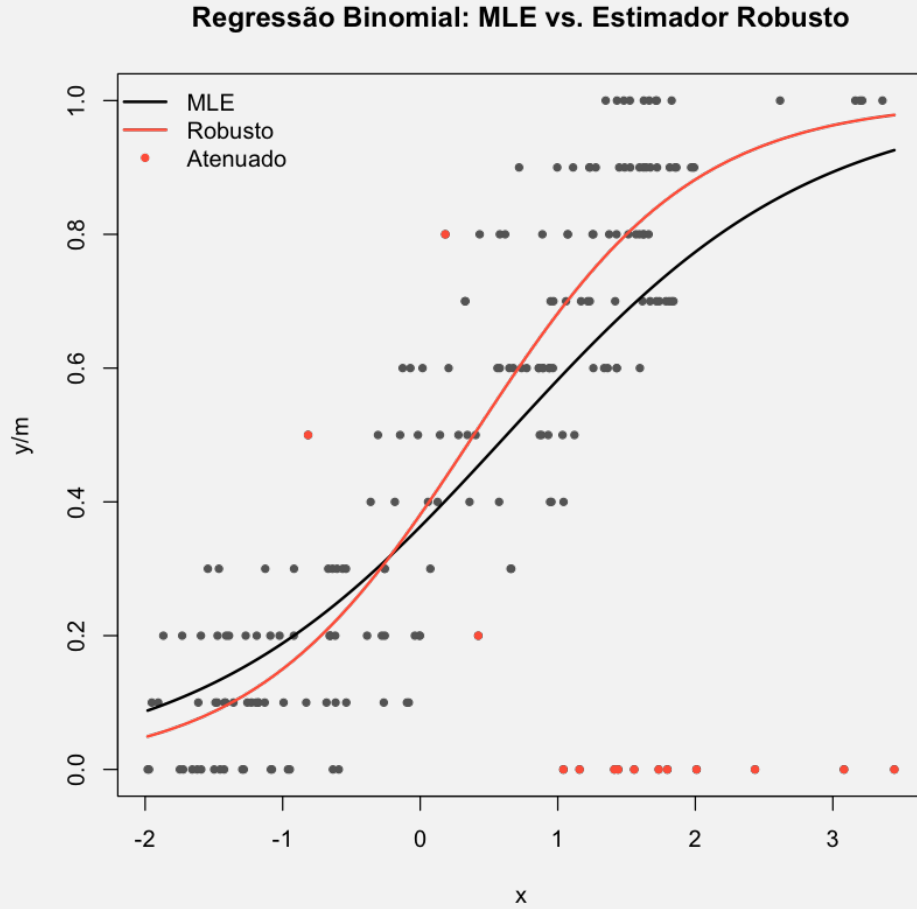


Figura 1: Curvas ajustadas sob contaminação (pontos atenuados em vermelho). Parâmetros:  $n = 200$ ,  $m = 10$ ,  $\theta = (-0.5, 1.2)$ , contaminação 10%,  $p = 1.5$ .

A Figura 2 apresenta as distribuições empíricas (histogramas) das estimativas sob  $R = 500$  repetições. As curvas

vermelhas representam a densidade normal ajustada a cada conjunto de estimativas, e as linhas tracejadas indicam os valores verdadeiros dos parâmetros.

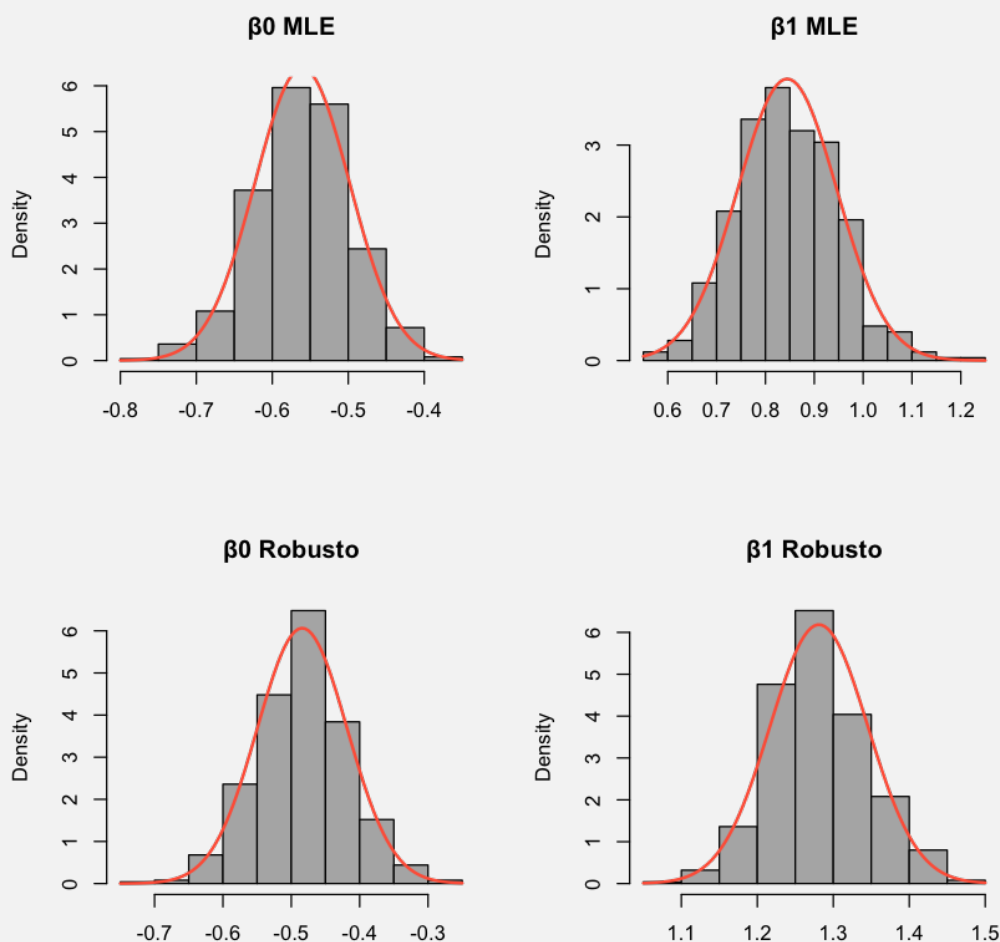


Figura 2: Distribuição das estimativas sob contaminação: comparação entre MLE e estimador robusto.

## 1.5 Conclusão

Os resultados mostram que a contaminação de apenas 10% das observações foi suficiente para produzir um viés expressivo no estimador clássico de máxima verossimilhança, especialmente no coeficiente  $\beta_1$  (viés médio de  $-0.355$ ). O estimador robusto, por sua vez, manteve os parâmetros próximos dos valores verdadeiros e apresentou menores valores de RMSE, indicando maior estabilidade e menor sensibilidade aos outliers.

As curvas ajustadas confirmam que o método robusto atenua as observações contaminadas, sem comprometer o ajuste das demais, e os histogramas de Monte Carlo mostram distribuições aproximadamente normais centradas em torno dos valores verdadeiros. Essas evidências demonstram que o tratamento robusto via ponderação  $w_i = \exp(-(|r_i| - c)_+)$  aproxima as estimativas da distribuição original do modelo, recuperando a forma assintótica esperada mesmo em presença de contaminação.

Os códigos deste estudo estão disponibilizados em <http://academic-codex.github.io/MAE5911-Estatistica-e-Machine-Learning>.

