

ME720 - Modelos Lineares Generalizados

Profa. Larissa Avila Matos

Atividade

1. Em sala de aula, nós criamos a nossa própria função do `glm` para estimar os parâmetros de um modelo de Poisson com função de ligação `log`. Além disso, calculamos os erros estimados, criamos intervalos de confiança e fizemos testes de hipóteses simples. Usando a mesma função das notas de aula:

```
newton <-function(y,X,init,eps=1e-6,maxiter=50){
  beta <- init
  n<-dim(X)[1]
  out <- matrix(NA, nrow=maxiter+1,ncol=length(t(beta)))
  out[1,] <- t(init)
  i <- 1
  continue <- T
  while (continue) {
    i <- i+1
    beta.o <- beta
    W<-diag(n)
    diag(W)<-exp(X%*%beta.o)
    mu<-exp(X%*%beta.o)
    beta <- beta.o + solve(t(X)%*%W%*%X)%*%t(X)%*%(y-mu)
    if(sum(is.na(beta))>0){stop("NA nas estimativas")}
    out[i,] <- t(beta)
    continue <- (abs(beta-beta.o) > eps) && (i <= maxiter)
  }
  if (i > maxiter) {
    warning("Máximo número de iterações atingido")
  }
  out <- out[!is.na(out[,1]),]
  saida<-list(out=out,est=out[i,],iter=i)
  return(saida)
}
```

(a) Atualize a função para que ela retorne os erros estimados, os ICs e os testes de hipóteses simples.

(b) Acrescente na função o cálculo da função desvio e do AIC.

(c) Gere dados para um modelo poisson com função de ligação `log`, considerando:

- duas variáveis preditoras de tamanho 100, onde $x_{1i} = 1$ e $x_{2i} \sim U(0, 2)$ ($x_i = (1, x_{2i})$); e
- $\beta = (1, 3)$.

Ajuste um modelo Poisson log linear aos seus dados simulados usando a sua função e a função `glm`. Os coeficientes estimados são semelhantes aos verdadeiros que você usou para a geração dos dados?

(d) **Estudo de simulação.**

Repita $N = 1000$ o item (c), para cada amostra (repetição) calcule o viés absoluto V dos parâmetros, ou seja,

$$V_{\beta_j} = |\beta_j - \hat{\beta}_j|,$$

em que $\hat{\beta}_j$ é a estimativa de β_j e β_j é o valor real. Calcule também o AIC e o desvio de cada amostra. Compare os resultados obtidos através da sua função e da função `glm`.

(e) Repita os itens (c) e (d) considerando agora a amostra com tamanho 300. Quais suas conclusões?

2. Análise o conjunto de dados *House Selling Price Data* do livro texto. Escreva um relatório técnico para essa análise.

3. Gere dados para uma regressão logística.

- Considere duas variáveis preditoras de comprimento 100, elas podem ser aleatórias na sua distribuição favorita.
- Considere `beta <- c(-1,2)` e gere a resposta com

```
rbinom(n = 100, size = 1, prob = exp(x%*%beta)/(1+exp(x%*%beta))).
```

(a) Por que esse é o modelo implícito na regressão logística?

(b) Ajuste uma regressão logística aos seus dados simulados usando a função `glm`.

(c) Os coeficientes estimados são semelhantes aos verdadeiros que você usou para a geração dos dados?

(d) **Estudo de simulação:**

1. Gere dados para uma regressão logística, ou seja, gere a resposta com `rbinom(n = n, size = 1, prob = exp(x%*%beta)/(1+exp(x%*%beta)))`, considerando:
 - duas variáveis preditoras de tamanho n , onde $x_{1i} \sim N(10, 2)$ e $x_{2i} \sim N(5, 1)$ ($x_i = (x_{1i}, x_{2i})$);
 - `beta <- c(-1, 2)`;
 - $n = 100$ (tamanho amostral).
2. Ajuste uma regressão logística aos seus dados simulados usando a função `glm`.
3. Repita $N = 1000$ o procedimento acima, para cada amostra (repetição) calcule o viés absoluto V dos parâmetros.
4. Repita os passos 1-3 trocando o tamanho amostral, considere $n = 300$ e $n = 500$.

Qual o comportamento das estimativas?

Nota: Para cada questão explique e justique tudo o que foi feito! Reporte tudo o que você achar necessário. Nos estudos de simulações, reporte os resultados em gráficos e tabelas.