

MAT02034 - Métodos bayesianos para análise de dados

Introdução à regressão linear bayesiana

Rodrigo Citton P. dos Reis
citton.padilha@ufrgs.br

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DEPARTAMENTO DE ESTATÍSTICA

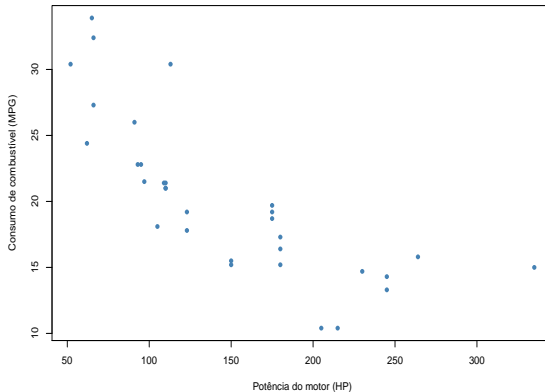
Porto Alegre, 2022

Revisão da Regressão Linear

Revisão da Regressão Linear

- ▶ Lembre-se de que, na análise de regressão, temos duas ou mais variáveis que podem ser medidas nos mesmos indivíduos (**unidades**).
- ▶ Desejamos usar uma ou mais delas, as **variáveis preditoras** (também chamadas de **variáveis independentes** ou **covariáveis**), para **explicar** ou **prever** uma **variável resposta** (também chamada de **variável desfecho** ou **variável dependente**).
 - ▶ Como definimos qual variável é a resposta e quais são preditores depende de nossa questão de pesquisa.

Revisão da Regressão Linear



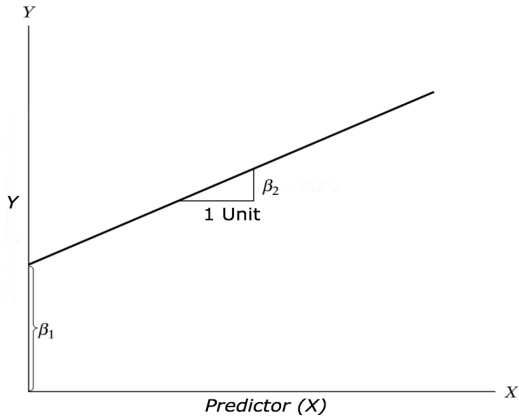
Revisão da Regressão Linear

- ▶ Na regressão linear, a variável resposta é quantitativa.
- ▶ Na regressão linear simples, há apenas uma variável preditora e a relação entre a variável resposta e o preditor é aproximadamente linear.
- ▶ Normalmente, a notação Y é usada para a variável resposta e X para um preditor, de modo que y_i e x_i denotam os valores observados da resposta e do preditor para o i -ésimo indivíduo em um conjunto de dados.
- ▶ A equação de regressão populacional com uma covariável é

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i,$$

em que β_0 é o **intercepto** (geralmente definido como o valor esperado de Y quando $X = 0$) e β_1 é a **inclinação** (a diferença esperada entre dois valores de Y cujos valores de X correspondentes diferem em uma unidade).

Revisão da Regressão Linear



Centralizando a Covariável

- **Centralizar** significa simplesmente calcular a média amostral da covariável e **subtrair essa média de cada valor de covariável**. Em símbolos, antes da centralização, o modelo a ser ajustado é

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i,$$

mas com a centralização, torna-se

$$Y_i = \beta_0^* + \beta_1 (X_i - \bar{X}) + \epsilon_i.$$

- A interpretação e o valor da inclinação β_1 permanecem inalterados pela centralização.

Centralizando a Covariável

- ▶ No entanto, a centralização dá ao intercepto um significado diferente e geralmente um valor muito diferente. Podemos ver isso reorganizando o lado direito da expressão anterior.

$$Y_i = (\beta_0^* - \beta_1 \bar{X}) + \beta_1 X_i + \epsilon_i.$$

- ▶ O intercepto no modelo centrado é o valor esperado da variável resposta quando a covariável centrada é 0.
 - ▶ Isto é, quando a covariável na escala original tem o valor típico \bar{X} em vez do valor irracional de 0. Assim, este tem uma interpretação significativa.
- ▶ Uma **vantagem adicional da centralização** de covariáveis para a inferência bayesiana é que a centralização pode melhorar a convergência de **amostradores MCMC** para ajustar modelos complicados de regressão bayesiana.

Introdução à regressão linear simples bayesiana

Introdução à RLS bayesiana

- ▶ A regressão linear simples bayesiana está intimamente relacionada aos modelos normais (para estimação de uma média populacional).
- ▶ A diferença é que ao invés de uma única média populacional, cada observação i tem sua própria média que depende dos coeficientes de regressão e do valor i da covariável.
- ▶ A seguir, assumiremos que a covariável foi centralizada, pois isso simplifica alguns cálculos.

Introdução à RLS bayesiana

- ▶ A verossimilhança pode ser escrita distributivamente da seguinte forma:

$$y_i | x_i, \beta_0, \beta_1, \sigma^2 \stackrel{\text{ind.}}{\sim} N(\beta_0 + \beta_1(x_i - \bar{x}), \sigma^2), \quad i = 1, \dots, n.$$

- ▶ O parâmetro de maior interesse na regressão linear simples bayesiana geralmente é a inclinação, β_1 . Assim, precisamos encontrar a densidade *a posteriori* conjunta de todos os três parâmetros de regressão e então integrar β_0 e σ^2 para obter a densidade marginal *a posteriori* de β_1 .

Distribuição *a priori* não informativa padrão

- ▶ Consideraremos primeiro a **distribuição *a priori* não informativa padrão** que produz inferência bayesiana análoga aos resultados frequentistas.
- ▶ Como nos modelos com verossimilhança normal em que tanto a média quanto a variância são desconhecidas, a distribuição *a priori* não informativa padrão na regressão bayesiana é o produto de distribuição *a priori* **impróprias** independentes nos parâmetros relacionados à média e no parâmetro de variância.
- ▶ Os parâmetros relacionados às médias são os coeficientes de regressão β_0 e β_1 .

Distribuição *a priori* não informativa padrão

- ▶ Multiplicando distribuições *a priori flat* (proporcionais a uma constante ao longo de toda a reta real) em ambos os coeficientes vezes uma distribuição *a priori* **gama inversa** com ambos os parâmetros tendendo a 0 para σ^2

$$p(\beta_0, \beta_1, \sigma^2) \propto 1/\sigma^2, -\infty < \beta_0, \beta_1 < \infty, 0 < \sigma^2 < \infty.$$

- ▶ Aproximamos esta distribuição *a priori* com distribuições *a priori* **normais vagas** (no **OpenBUGS** “`dflat()`”) em β_0 e β_1 e distribuição *a priori* gama muito vaga no parâmetro de precisão (o **OpenBUGS** utiliza o parâmetro de precisão na parametrização da distribuição normal).

Distribuição *a priori* não informativa padrão

As três estatísticas suficientes simplificam a derivação das distribuições *a posteriori* conjuntas e marginais dos parâmetros de regressão.

$$\begin{aligned}\hat{\beta}_0 &= \bar{y}, \\ \hat{\beta}_1 &= \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_i (x_i - \bar{x})^2}, \\ SSR &= \sum_i \left[y_i - \hat{\beta}_0 - \hat{\beta}_1(x_i - \bar{x}) \right]^2.\end{aligned}$$

- Lembre-se também de que a variância amostral na regressão é $s^2 = \frac{SSR}{(n-2)}$.

Distribuição *a priori* não informativa padrão

- ▶ Com a distribuição *a priori* não informativa padrão e uma covariável centralizada, a densidade *a posteriori* conjunta é

$$\begin{aligned} p(\beta_0, \beta_1, \sigma^2 | y) &\propto \frac{1}{\sigma^2} \frac{1}{(\sigma^2)^{\left(\frac{n}{2}\right)}} \exp \left[\frac{-\sum_i (y_i - \beta_0 - \beta_1(x_i - \bar{x}))^2}{2\sigma^2} \right] \\ &= \frac{1}{\sigma^2} \frac{1}{(\sigma^2)^{\left(\frac{n+2}{2}\right)}} \exp \left[-\frac{SSR + n(\beta_0 - \hat{\beta}_0)^2 - \sum_i (x_i - \bar{x})^2 (\beta_1 - \hat{\beta}_1)^2}{2\sigma^2} \right]. \end{aligned}$$

Distribuição *a priori* não informativa padrão

- ▶ Se integrarmos a distribuição *a posteriori* com respeito a β_1 , obtemos

$$p(\beta_0, \sigma^2 | y) \propto \frac{1}{(\sigma^2)^{\left(\frac{n+1}{2}\right)}} \exp \left[-\frac{SSR + n(\beta_0 - \hat{\beta}_0)^2}{2\sigma^2} \right].$$

- ▶ Para obter a densidade marginal *a posteriori* de β_0 , devemos integrar a expressão anterior com respeito a σ^2 que, após alguma álgebra, produz

$$\beta_0 | y \sim t \left(\hat{\beta}_0, \frac{s^2}{n}, n - 2 \right),$$

uma distribuição com média $\widehat{\beta}_0$, parâmetro de escala s^2/n e graus de liberdade $n - 2$.

Distribuição *a priori* não informativa padrão

- Um par semelhante de integrações leva à densidade marginal *a posteriori* que é nossa principal preocupação:

$$\beta_1|y \sim t\left(\hat{\beta}_1, \frac{s^2}{\sum_i (x_i - \bar{x})^2}, n-2\right).$$

- Com base nessas distribuições *t*, os intervalos de credibilidade bayesianos para β_0 e β_1 terão exatamente os mesmos limites que os intervalos de confiança *t* frequentistas para os mesmos parâmetros.
- E finalmente,

$$\sigma^2|y \sim Gl\left(\frac{n-2}{2}, \frac{SSR}{2}\right).$$

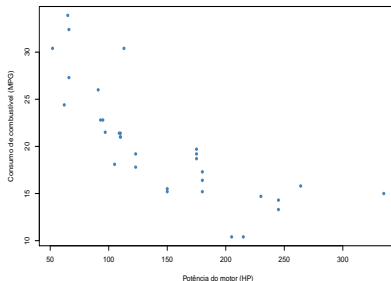
Distribuição *a priori* não informativa padrão

Verificando se a densidade *a posteriori* é própria

- ▶ Lembre-se de que sempre que um estatístico usar uma distribuição *a priori* imprópria, ele deve verificar se os dados fornecem informações suficientes para tornar a densidade *a posteriori* própria.
- ▶ No caso da regressão linear simples e da distribuição *a priori* conjunta imprópria padrão, os requisitos de dados são que o **tamanho da amostra n** seja estritamente maior que dois, e que nem todos os **valores das covariáveis** sejam iguais.

Exemplo

- ▶ **(mtcars)** Dados extraídos da revista *Motor Trend US* de 1974 e abrangem o consumo de combustível e 10 aspectos do *design* e desempenho do automóvel para 32 modelos (de 1973 a 1974).
- ▶ Vamos nos concentrar na relação **Consumo de combustível (MPG)** e **Potência do motor (HP)**.



Exemplo

- ▶ O modelo linear é assumido

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i,$$

em que Y representa a variável resposta “Consumo de combustível (MPG)” e X representa a variável explicativa “Potência do motor (HP)”.

- ▶ Vamos utilizar o **OpenBUGS** para obter uma amostra da distribuição *a posteriori* conjunta de β_0 , β_1 e σ^2 (variância do termo de erro aleatório; ou, de forma equivalente, σ).

Exemplo

```
model
{
  for (i in 1:N) {
    mu[i] <- beta0 + beta1 * xcent[i] # media de y
    y[i] ~ dnorm( mu[i], tausq ) # dist de y (verossimilhança)
  }
  beta0 ~ dflat() # dist. a priori beta0
  beta1 ~ dflat() # dist. a priori beta1
  tausq ~ dgamma( 0.001, 0.001 ) # dist. a priori tau (precisao)
  sigma <- 1/sqrt(tausq) # desvio padrao da regressao
}
```

Exemplo

- ▶ Vamos “chamar” o **OpenBUGS** do R via pacote R2OpenBUGS.
- ▶ Para isso, é preciso salvar o script do modelo em um arquivo `.txt` (`rlb.txt`).
- ▶ **Observação:**
 - ▶ O R2OpenBUGS funciona como uma interface para o **OpenBUGS** no R.
 - ▶ Quem gera a amostra da distribuição *a posteriori* do modelo é o **OpenBUGS**.
 - ▶ Logo este precisa estar instalado no computador).

Exemplo

```
# install.packages("R2OpenBUGS")
library(R2OpenBUGS)

# Objeto de dados
N <- nrow(mtcars)
y <- mtcars$mpg
# centralizando a covariável
xcent <- mtcars$hp - mean(mtcars$hp)
data <- list("N", "y", "xcent")

# Valores iniciais
inits <- function(){
  list(beta0 = rnorm(1, 0, 100),
        beta1 = rnorm(1, 0, 100),
        tausq = 5)
}

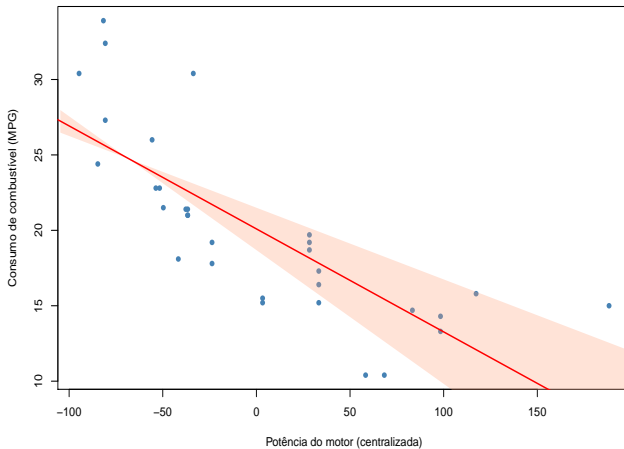
# Rodando o modelo/gerando as cadeias
rlb.sim <- bugs(data = data, inits = inits,
               model.file = here::here("material_de_aula",
                                       "OpenBugsExemplos", "rlb.txt"),
               parameters = c("beta0", "beta1", "sigma"),
               n.chains = 3, n.iter = 10000)
```

Exemplo

```
print(rlb.sim)
```

```
## Inference for Bugs model at "C:/Users/rodri/OneDrive/Documentos/UFRGS/Disiplinas/metodo
## Current: 3 chains, each with 10000 iterations (first 5000 discarded)
## Cumulative: n.sims = 15000 iterations saved
##           mean  sd  2.5%   25%   50%   75%  97.5%  Rhat  n.eff
## beta0      20.1 0.7  18.7  19.6  20.1  20.6  21.5    1   9000
## beta1      -0.1 0.0  -0.1  -0.1  -0.1  -0.1   0.0    1   6800
## sigma       4.0 0.5   3.1   3.6   3.9   4.3   5.1    1  15000
## deviance  178.4 2.6 175.5 176.5 177.7 179.5 185.0    1 15000
##
## For each parameter, n.eff is a crude measure of effective sample size,
## and Rhat is the potential scale reduction factor (at convergence, Rhat=1).
##
## DIC info (using the rule,  $pD = \bar{D} - D_{\text{hat}}$ )
##  $pD = 3.1$  and  $DIC = 181.5$ 
## DIC is an estimate of expected predictive error (lower deviance is better).
```


Exemplo



Exemplo

```
rlb.sim <- bugs(data = data, inits = inits,  
               model.file = here::here("material_de_aula",  
                                       "OpenBugsExemplos", "rlb.txt"),  
               parameters = c("beta0", "beta1", "sigma"),  
               n.chains = 3, n.iter = 10000,  
               n.burnin = 1000,  
               n.thin = 10,  
               codaPkg = TRUE)
```

Exemplo

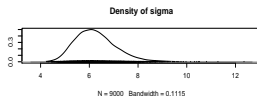
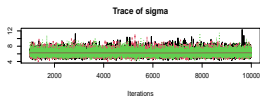
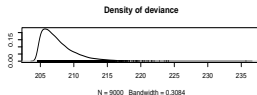
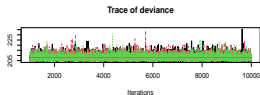
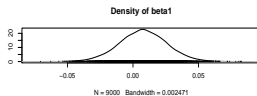
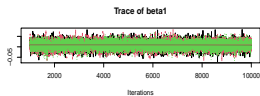
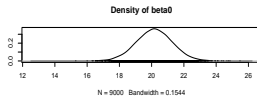
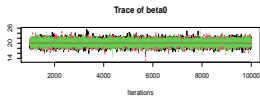
```
library(coda)

codaobject <- read.bugs(rlb.sim)

## Abstracting beta0 ... 9000 valid values
## Abstracting beta1 ... 9000 valid values
## Abstracting deviance ... 9000 valid values
## Abstracting sigma ... 9000 valid values
## Abstracting beta0 ... 9000 valid values
## Abstracting beta1 ... 9000 valid values
## Abstracting deviance ... 9000 valid values
## Abstracting sigma ... 9000 valid values
## Abstracting beta0 ... 9000 valid values
## Abstracting beta1 ... 9000 valid values
## Abstracting deviance ... 9000 valid values
## Abstracting sigma ... 9000 valid values

plot(codaobject)
```

Exemplo



Densidades *a priori* informativas para coeficientes de regressão e variância

- ▶ E se tivermos informações *a priori* que gostaríamos de incluir em nosso modelo bayesiano?
- ▶ O procedimento mais simples (e provavelmente o mais comumente usado) é assumir independência *a priori* entre β_0 , β_1 e σ^2 e colocar distribuições *a priori* **normais próprias independentes** em β_0 e β_1 e uma distribuição *a priori* **gama inversa própria** na variância σ^2 (ou equivalentemente, uma distribuição *a priori* **gama** para o parâmetro de precisão $[1/\sigma^2]$).
- ▶ O produto dessas três densidades *a priori* não é uma distribuição *a priori* conjugada, porque a densidade *a posteriori* resultante não será fatorada em três densidades independentes das mesmas famílias.
 - ▶ Um método de aproximação (por exemplo, MCMC) será necessário para avaliar a distribuição *a posteriori* resultante.

Para casa

- ▶ Revisar a aula de hoje (passos de instalação do OpenBUGS, rodar o exemplo).
- ▶ Implemente no R modelo do exemplo da aula de hoje. Compare com os resultados encontrados no OpenBUGS.
- ▶ Utilize o OpenBUGS para rodar o modelo do exemplo dos dados de taxas de scram em usinas nucleares (aulas 08 e 09). Compare com os resultados encontrados pela a implementação em R feita em aula.
- ▶ Trazer as dúvidas para o Fórum Geral do Moodle e para a próxima aula.

Próxima aula

- ▶ Modelos lineares.

Por hoje é só!

Bons estudos!

