Modelo de Regressão Linear Múltiplo

Prof. Juvêncio Santos Nobre

Departamento de Estatística e Matemática Aplicada

Universidade Federal do Ceará-Brasil

http://www.dema.ufc.br/∼juvencio

DEMA-UFC

Capital do **Ceará**, outubro de 2022

Conteúdo

Modelo de regressão Linear Múltiplo - MRLM

- Na grande maioria das aplicações, necessitamos de mais de uma variável explicativa para conseguir modelar de forma adequada a variabilidade da variável resposta, surgindo assim os modelos de regressão múltipla (múltiplas variáveis explicativas).
- O modelo de regressão linear múltiplo (MRLM) admite a seguinte forma funcional

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_{p-1} x_{(p-1)i} + e_i, i = 1, \ldots, n, \tag{1}$$

que pode ser reescrito como

$$\boldsymbol{y} = \boldsymbol{X}\beta + \boldsymbol{e}, \tag{2}$$

em que:

Modelo de regressão Linear Múltiplo - MRLM

- Na grande maioria das aplicações, necessitamos de mais de uma variável explicativa para conseguir modelar de forma adequada a variabilidade da variável resposta, surgindo assim os modelos de regressão múltipla (múltiplas variáveis explicativas).
- O modelo de regressão linear múltiplo (MRLM) admite a seguinte forma funcional

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_{p-1} x_{(p-1)i} + e_i, i = 1, \ldots, n, \tag{1}$$

que pode ser reescrito como

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{e}, \tag{2}$$

em que:

MRLM

- $\boldsymbol{y} = (y_1, \ldots, y_n)^\top$: vetor $(n \times 1)$ de variáveis respostas.
- $\boldsymbol{X}$: matriz de especificação do modelo, dada por

$$\boldsymbol{X} = \begin{pmatrix} 1 & x_{11} & x_{21} & \ldots & x_{(p-1)1} \\ 1 & x_{12} & x_{22} & \ldots & x_{(p-1)2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1n} & x_{2n} & \ldots & x_{(p-1)n} \end{pmatrix} = (\mathbf{1}_n, \boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_{p-1}) = (\mathbf{1}_n, \boldsymbol{X}_R)$$

  de dimensão $n \times p$ $(n > p)$ conhecida e assumida de posto completo.
- $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_{p-1})^\top$: vetor $(p \times 1)$ de parâmetros de regressão.
- $\boldsymbol{e} = (e_1, \ldots, e_n)^\top$: vetor $(n \times 1)$ com elementos representando a fonte de variação associada aos elementos da amostra.

MRLM

- $\boldsymbol{y} = (y_1, \ldots, y_n)^\top$: vetor $(n \times 1)$ de variáveis respostas.
- $\boldsymbol{X}$: matriz de especificação do modelo, dada por

$$\boldsymbol{X} = \begin{pmatrix} 1 & x_{11} & x_{21} & \ldots & x_{(p-1)1} \\ 1 & x_{12} & x_{22} & \ldots & x_{(p-1)2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1n} & x_{2n} & \ldots & x_{(p-1)n} \end{pmatrix} = (\mathbf{1}_n, \boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_{p-1}) = (\mathbf{1}_n, \boldsymbol{X}_R)$$

  de dimensão $n \times p$ $(n > p)$ conhecida e assumida de posto completo.
- $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_{p-1})^\top$: vetor $(p \times 1)$ de parâmetros de regressão.
- $\boldsymbol{e} = (e_1, \ldots, e_n)^\top$: vetor $(n \times 1)$ com elementos representando a fonte de variação associada aos elementos da amostra.

MRLM

- $\boldsymbol{y} = (y_1, \ldots, y_n)^\top$: vetor $(n \times 1)$ de variáveis respostas.
- $\boldsymbol{X}$: matriz de especificação do modelo, dada por

$$\boldsymbol{X} = \begin{pmatrix} 1 & x_{11} & x_{21} & \ldots & x_{(p-1)1} \\ 1 & x_{12} & x_{22} & \ldots & x_{(p-1)2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1n} & x_{2n} & \ldots & x_{(p-1)n} \end{pmatrix} = (\mathbf{1}_n, \boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_{p-1}) = (\mathbf{1}_n, \boldsymbol{X}_R)$$

  de dimensão $n \times p$ $(n > p)$ conhecida e assumida de posto completo.
- $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_{p-1})^\top$: vetor $(p \times 1)$ de parâmetros de regressão.
- $\boldsymbol{e} = (e_1, \ldots, e_n)^\top$: vetor $(n \times 1)$ com elementos representando a fonte de variação associada aos elementos da amostra.

MRLM

- $\boldsymbol{y} = (y_1, \ldots, y_n)^\top$: vetor $(n \times 1)$ de variáveis respostas.
- $\boldsymbol{X}$: matriz de especificação do modelo, dada por

$$\boldsymbol{X} = \begin{pmatrix} 1 & x_{11} & x_{21} & \ldots & x_{(p-1)1} \\ 1 & x_{12} & x_{22} & \ldots & x_{(p-1)2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1n} & x_{2n} & \ldots & x_{(p-1)n} \end{pmatrix} = (\mathbf{1}_n, \boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_{p-1}) = (\mathbf{1}_n, \boldsymbol{X}_R)$$

  de dimensão $n \times p$ $(n > p)$ conhecida e assumida de posto completo.
- $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_{p-1})^\top$: vetor $(p \times 1)$ de parâmetros de regressão.
- $\boldsymbol{e} = (e_1, \ldots, e_n)^\top$: vetor $(n \times 1)$ com elementos representando a fonte de variação associada aos elementos da amostra.

MRLM

- Este modelo descreve um hiperplano $(p-1)$-dimensional no espaço gerado pelas colunas de $\boldsymbol{X}$ ($\mathbb{C}(\boldsymbol{X})$), i.e., no espaço gerado pelas variáveis explicatvas $x_1, \ldots, x_{p-1}$.
- Para exemplificar, considere o caso particular com $p = 3$ e a seguinte função de regressão

$$\mu(\boldsymbol{x}, \boldsymbol{\beta}) = \mathbb{E}[y_i | x_1, x_2] = \beta_0 + \beta_1 x_1 + \beta_2 x_2.$$

- Na Figura 1, fixando $\beta_0 = 50, \beta_1 = 10$ e $\beta_2 = 7$ mostramos os gráficos do hiperplano de regressão e o gráfico de contorno associados.

MRLM

- Este modelo descreve um hiperplano $(p-1)$-dimensional no espaço gerado pelas colunas de $\boldsymbol{X}$ ($\mathbb{C}(\boldsymbol{X})$), i.e., no espaço gerado pelas variáveis explicatvas $x_1, \ldots, x_{p-1}$.
- Para exemplificar, considere o caso particular com $p = 3$ e a seguinte função de regressão

$$\mu(\boldsymbol{x}, \boldsymbol{\beta}) = \mathbb{E}[y_i|x_1, x_2] = \beta_0 + \beta_1 x_1 + \beta_2 x_2.$$

- Na Figura 1, fixando $\beta_0 = 50, \beta_1 = 10$ e $\beta_2 = 7$ mostramos os gráficos do hiperplano de regressão e o gráfico de contorno associados.

MRLM

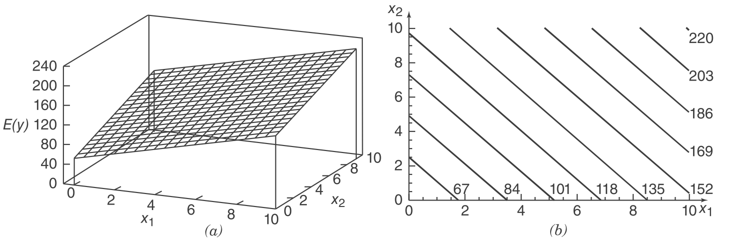
- Este modelo descreve um hiperplano $(p-1)$-dimensional no espaço gerado pelas colunas de $\boldsymbol{X}$ ($\mathbb{C}(\boldsymbol{X})$), i.e., no espaço gerado pelas variáveis explicatvas $x_1, \ldots, x_{p-1}$.
- Para exemplificar, considere o caso particular com $p = 3$ e a seguinte função de regressão

$$\mu(\boldsymbol{x}, \boldsymbol{\beta}) = \mathbb{E}[y_i | x_1, x_2] = \beta_0 + \beta_1 x_1 + \beta_2 x_2.$$

- Na Figura 1, fixando $\beta_0 = 50, \beta_1 = 10$ e $\beta_2 = 7$ mostramos os gráficos do hiperplano de regressão e o gráfico de contorno associados.

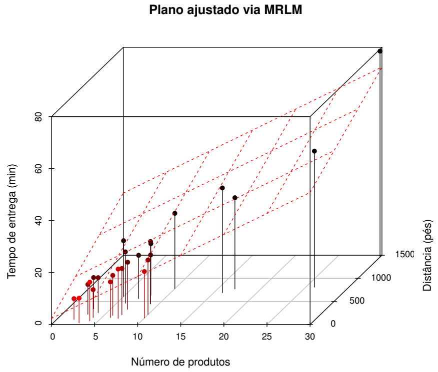
Ilustração

Figura: (a) Hiperplano de regressão e (b) gráfico de contorno para o modelo $\mu(\boldsymbol{x}_i) = 50 + 10x_1 + 7x_2$ (Montgomery et al. , 2012).



Ilustração

Figura: Hiperplano de regressão ajustado para os dados do Exemplo 3.1- The Delivery Time Data (Montgomery et al., 2012, pag. 74)



Interpretação dos parâmetros

- Considere, hipoteticamente, o seguinte MRLM

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + e_i,\ i = 1 \ldots, n.$$

- O parâmetro β_0 é o intercepto do plano de regressão, i.e.,

$$\beta_0 = \mathbb{E}[y_i | x_{1i} = x_{2i} = 0]$$

e **só possui** interpretação quando a amplitude dos dados **incluir** o par $x_1 = x_2 = 0$. Caso contrário, no modelo original, β_0 não possui interpretação física/prática.

- O parâmetro β_1 é dado por

$$\beta_1 = \mathbb{E}[y_i | x_{1i} = x + 1, x_{2i} = y] - \mathbb{E}[y_i | x_{1i} = x, x_{2i} = y], \forall x, y \in \mathbb{R}.$$

- O parâmetro β_2 é dado por

$$\beta_2 = \mathbb{E}[y_i | x_{1i} = x, x_{2i} = y + 1] - \mathbb{E}[y_i | x_{1i} = x, x_{2i} = y], \forall x, y \in \mathbb{R}.$$

- Por essa razão, os parâmetros $\beta_1, \ldots, \beta_{p-1}$ também são denominados de coeficientes de regressão **parcial**.

Interpretação dos parâmetros

- Considere, hipoteticamente, o seguinte MRLM

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + e_i, \; i = 1 \ldots, n.$$

- O parâmetro β_0 é o intercepto do plano de regressão, i.e.,

$$\beta_0 = \mathbb{E}[y_i | x_{1i} = x_{2i} = 0]$$

e **só possui** interpretação quando a amplitude dos dados **incluir** o par $x_1 = x_2 = 0$. Caso contrário, no modelo original, β_0 não possui interpretação física/prática.

- O parâmetro β_1 é dado por

$$\beta_1 = \mathbb{E}[y_i | x_{1i} = x + 1, x_{2i} = y] - \mathbb{E}[y_i | x_{1i} = x, x_{2i} = y], \forall x, y \in \mathbb{R}.$$

- O parâmetro β_2 é dado por

$$\beta_2 = \mathbb{E}[y_i | x_{1i} = x, x_{2i} = y + 1] - \mathbb{E}[y_i | x_{1i} = x, x_{2i} = y], \forall x, y \in \mathbb{R}.$$

- Por essa razão, os parâmetros $\beta_1, \ldots, \beta_{p-1}$ também são denominados de coeficientes de regressão **parcial**.

Interpretação dos parâmetros

- Considere, hipoteticamente, o seguinte MRLM

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + e_i,\ i = 1\ldots, n.$$

- O parâmetro β_0 é o intercepto do plano de regressão, i.e.,

$$\beta_0 = \mathbb{E}[y_i | x_{1i} = x_{2i} = 0]$$

e **só possui** interpretação quando a amplitude dos dados **incluir** o par $x_1 = x_2 = 0$. Caso contrário, no modelo original, β_0 não possui interpretação física/prática.

- O parâmetro β_1 é dado por

$$\beta_1 = \mathbb{E}[y_i | x_{1i} = x + 1, x_{2i} = y] - \mathbb{E}[y_i | x_{1i} = x, x_{2i} = y], \forall x, y \in \mathbb{R}.$$

- O parâmetro β_2 é dado por

$$\beta_2 = \mathbb{E}[y_i | x_{1i} = x, x_{2i} = y + 1] - \mathbb{E}[y_i | x_{1i} = x, x_{2i} = y], \forall x, y \in \mathbb{R}.$$

- Por essa razão, os parâmetros $\beta_1, \ldots, \beta_{p-1}$ também são denominados de coeficientes de regressão **parcial**.

Interpretação dos parâmetros

- Considere, hipoteticamente, o seguinte MRLM

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + e_i,\ i = 1\ldots, n.$$

- O parâmetro β_0 é o intercepto do plano de regressão, i.e.,

$$\beta_0 = \mathbb{E}[y_i | x_{1i} = x_{2i} = 0]$$

e **só possui** interpretação quando a amplitude dos dados **incluir** o par $x_1 = x_2 = 0$. Caso contrário, no modelo original, β_0 não possui interpretação física/prática.

- O parâmetro β_1 é dado por

$$\beta_1 = \mathbb{E}[y_i | x_{1i} = x+1, x_{2i} = y] - \mathbb{E}[y_i | x_{1i} = x, x_{2i} = y], \forall x, y \in \mathbb{R}.$$

- O parâmetro β_2 é dado por

$$\beta_2 = \mathbb{E}[y_i | x_{1i} = x, x_{2i} = y+1] - \mathbb{E}[y_i | x_{1i} = x, x_{2i} = y], \forall x, y \in \mathbb{R}.$$

- Por essa razão, os parâmetros $\beta_1, \ldots, \beta_{p-1}$ também são denominados de coeficientes de regressão **parcial**.

Interpretação dos parâmetros

- Considere, hipoteticamente, o seguinte MRLM

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + e_i, \; i = 1 \ldots, n.$$

- O parâmetro β_0 é o intercepto do plano de regressão, i.e.,

$$\beta_0 = \mathbb{E}[y_i | x_{1i} = x_{2i} = 0]$$

e **só possui** interpretação quando a amplitude dos dados **incluir** o par $x_1 = x_2 = 0$. Caso contrário, no modelo original, β_0 não possui interpretação física/prática.

- O parâmetro β_1 é dado por

$$\beta_1 = \mathbb{E}[y_i | x_{1i} = x + 1, x_{2i} = y] - \mathbb{E}[y_i | x_{1i} = x, x_{2i} = y], \forall x, y \in \mathbb{R}.$$

- O parâmetro β_2 é dado por

$$\beta_2 = \mathbb{E}[y_i | x_{1i} = x, x_{2i} = y + 1] - \mathbb{E}[y_i | x_{1i} = x, x_{2i} = y], \forall x, y \in \mathbb{R}.$$

- Por essa razão, os parâmetros $\beta_1, \ldots, \beta_{p-1}$ também são denominados de coeficientes de regressão **parcial**.

Exercício (fazer agora)

Exercício 1: Considere o seguinte MRLM

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + e_i, \ i = 1 \ldots, n.$$

em que y_i denota o lucro do i-ésimo mês (em milhares de reais), x_{1i} e x_{2i} denotam o capital investido e o gasto em publicidade, em milhares de reais, de uma determinada empresa no i-ésimo mês. Interprete os parâmetros do modelo de regressão.

Suposições

Para ajustar o MRLM, considera-se as seguintes suposições:

i) A função de regressão $\mu(\boldsymbol{x};\boldsymbol{\beta})$ é linear nos parâmetros.

ii) Os valores das variáveis explicativas são conhecidos e fixados, ou de uma forma geral, a matriz de especificação $\boldsymbol{X}$ $(n \times p)$ é conhecida, não estocástica e de posto completo.

iii) $\boldsymbol{e} \sim (\mathbf{0}, \sigma^2\boldsymbol{I}_n)$.

A suposição de homoscesdaticidade e não-correlação por parte das fontes de variação pode ser expressa somente na suposição iii). Para efeito de inferência de segunda ordem, se faz necessário alguma suposição distribucional a respeito da fonte de variação.

Comentários

- A classe de MRLM são sempre utilizados como modelos empíricos (ou aproximados), i.e., a verdadeira relação funcional entre y e $x_1, \ldots, x_{p-1}$ é **desconhecida**, todavia dentro de um certo intervalo/região a relação pode ser bem aproximada por uma função linear.
- Modelos mais complexos também podem ser expressos e analisados como um MRLM, como por exemplo:
- $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{1i} x_{2i} + e_i$ que corresponde a um modelo de duas variáveis explicativas com interação.
- $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_{11} x_{1i}^2 + \beta_{22} x_{2i}^2 + \beta_{12} x_{1i} x_{2i} + e_i$ que corresponde a um modelo de segunda ordem completo com interação.
- Um detalhe pertinente, é que apesar de conseguirem ser expressos na forma (1), as interpretações dos parâmetros se **modificam**.
- Apesar da possibilidade de ajustar vários modelos na classe de MRLM, eles podem possuir comportamentos bem distintos. Para ilustrar isso considere o Exercício 2 para ser entregue na próxima aula.

Comentários

- A classe de MRLM são sempre utilizados como modelos empíricos (ou aproximados), i.e., a verdadeira relação funcional entre y e $x_1, \ldots, x_{p-1}$ é **desconhecida**, todavia dentro de um certo intervalo/região a relação pode ser bem aproximada por uma função linear.
- Modelos mais complexos também podem ser expressos e analisados como um MRLM, como por exemplo:
- $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{1i} x_{2i} + e_i$ que corresponde a um modelo de duas variáveis explicativas com interação.
- $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_{11} x_{1i}^2 + \beta_{22} x_{2i}^2 + \beta_{12} x_{1i} x_{2i} + e_i$ que corresponde a um modelo de segunda ordem completo com interação.
- Um detalhe pertinente, é que apesar de conseguirem ser expressos na forma (1), as interpretações dos parâmetros se **modificam**.
- Apesar da possibilidade de ajustar vários modelos na classe de MRLM, eles podem possuir comportamentos bem distintos. Para ilustrar isso considere o Exercício 2 para ser entregue na próxima aula.

Comentários

- A classe de MRLM são sempre utilizados como modelos empíricos (ou aproximados), i.e., a verdadeira relação funcional entre y e $x_1, \ldots, x_{p-1}$ é **desconhecida**, todavia dentro de um certo intervalo/região a relação pode ser bem aproximada por uma função linear.
- Modelos mais complexos também podem ser expressos e analisados como um MRLM, como por exemplo:
- $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{1i} x_{2i} + e_i$ que corresponde a um modelo de duas variáveis explicativas com interação.
- $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_{11} x_{1i}^2 + \beta_{22} x_{2i}^2 + \beta_{12} x_{1i} x_{2i} + e_i$ que corresponde a um modelo de segunda ordem completo com interação.
- Um detalhe pertinente, é que apesar de conseguirem ser expressos na forma (1), as interpretações dos parâmetros se **modificam**.
- Apesar da possibilidade de ajustar vários modelos na classe de MRLM, eles podem possuir comportamentos bem distintos. Para ilustrar isso considere o Exercício 2 para ser entregue na próxima aula.

Comentários

- A classe de MRLM são sempre utilizados como modelos empíricos (ou aproximados), i.e., a verdadeira relação funcional entre y e $x_1, \ldots, x_{p-1}$ é **desconhecida**, todavia dentro de um certo intervalo/região a relação pode ser bem aproximada por uma função linear.
- Modelos mais complexos também podem ser expressos e analisados como um MRLM, como por exemplo:
- $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{1i} x_{2i} + e_i$ que corresponde a um modelo de duas variáveis explicativas com interação.
- $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_{11} x_{1i}^2 + \beta_{22} x_{2i}^2 + \beta_{12} x_{1i} x_{2i} + e_i$ que corresponde a um modelo de segunda ordem completo com interação.
- Um detalhe pertinente, é que apesar de conseguirem ser expressos na forma (1), as interpretações dos parâmetros se **modificam**. ☹
- Apesar da possibilidade de ajustar vários modelos na classe de MRLM, eles podem possuir comportamentos bem distintos. Para ilustrar isso considere o Exercício 2 para ser entregue na próxima aula.

Comentários

- A classe de MRLM são sempre utilizados como modelos empíricos (ou aproximados), i.e., a verdadeira relação funcional entre y e $x_1, \ldots, x_{p-1}$ é **desconhecida**, todavia dentro de um certo intervalo/região a relação pode ser bem aproximada por uma função linear.
- Modelos mais complexos também podem ser expressos e analisados como um MRLM, como por exemplo:
- $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{1i} x_{2i} + e_i$ que corresponde a um modelo de duas variáveis explicativas com interação.
- $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_{11} x_{1i}^2 + \beta_{22} x_{2i}^2 + \beta_{12} x_{1i} x_{2i} + e_i$ que corresponde a um modelo de segunda ordem completo com interação.
- Um detalhe pertinente, é que apesar de conseguirem ser expressos na forma (1), as interpretações dos parâmetros se **modificam**.
- Apesar da possibilidade de ajustar vários modelos na classe de MRLM, eles podem possuir comportamentos bem distintos. Para ilustrar isso considere o Exercício 2 para ser entregue na próxima aula.

Comentários

- A classe de MRLM são sempre utilizados como modelos empíricos (ou aproximados), i.e., a verdadeira relação funcional entre y e $x_1, \ldots, x_{p-1}$ é **desconhecida**, todavia dentro de um certo intervalo/região a relação pode ser bem aproximada por uma função linear.
- Modelos mais complexos também podem ser expressos e analisados como um MRLM, como por exemplo:
- $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{1i} x_{2i} + e_i$ que corresponde a um modelo de duas variáveis explicativas com interação.
- $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_{11} x_{1i}^2 + \beta_{22} x_{2i}^2 + \beta_{12} x_{1i} x_{2i} + e_i$ que corresponde a um modelo de segunda ordem completo com interação.
- Um detalhe pertinente, é que apesar de conseguirem ser expressos na forma (1), as interpretações dos parâmetros se **modificam**. ☹
- Apesar da possibilidade de ajustar vários modelos na classe de MRLM, eles podem possuir comportamentos bem distintos. Para ilustrar isso considere o Exercício 2 para ser entregue na próxima aula.

Exercícios (entregar próxima aula)

Exercício 2: Para cada uma das funções de regressão abaixo, pede-se para plotar os gráficos dos hiperplanos e respectivas curvas de níveis associadas, além de interpretar os respectivos gráficos.

i) $\mu(\boldsymbol{x}, \boldsymbol{\beta}) = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2$.

ii) $\mu(\boldsymbol{x}, \boldsymbol{\beta}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2$.

iii) $\mu(\boldsymbol{x}, \boldsymbol{\beta}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \beta_4 x_2^2 + \beta_5 x_1 x_2$.

Exercício 3: Mostre que os três modelos acima podem ser expressos como MRLM, i.e., reescreva-os na forma (1) especificando a matriz $\boldsymbol{X}$ e o respectivo vetor de parâmetros $\boldsymbol{\beta}$.

Exercício 4: Considere o seguinte MRLM

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + e_i,\ i = 1, \ldots, n.$$

Considere as suposições adequadas e determine o EMQ de $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)^\top$ sem usar a notação a matricial.

Estimação

Sob as suposições usuais do MRLM, temos que o EMQ é obtido através da minimização da forma quadrática

$$S(\boldsymbol{\beta}) = (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^\top(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}),$$

no qual já mostramos que o respectivo EMQ de $\boldsymbol{\beta}$ é dado por

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{y},$$

desde que $\boldsymbol{X}$ seja de posto completo, i.e., se $\mathrm{rank}(\boldsymbol{X}) = \mathrm{posto}(\boldsymbol{X}) = p < n$ e isso ocorre, se e somente se, as colunas da matriz $\boldsymbol{X}$ forem **linearmente independentes**.

Exercício (Entregar próxima aula)

Exercício 5: Usando a notação matricial, apresente a interpretação geométrica do método de mínimos quadrados.

Propriedades

Considere um MRLM com respectiva matriz de especificação

$$\boldsymbol{X} = \begin{pmatrix} 1 & x_{11} & x_{21} & \ldots & x_{(p-1)1} \\ 1 & x_{12} & x_{22} & \ldots & x_{(p-1)2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1n} & x_{2n} & \ldots & x_{(p-1)n} \end{pmatrix} = (\boldsymbol{1}, \boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_{p-1}) = (\boldsymbol{1}_n, \boldsymbol{X}_R).$$

P1. O sistema de equações normais (equação de estimação resultante do MMQ) é dada por:

$$\boldsymbol{X}^\top \widehat{\boldsymbol{e}} = \boldsymbol{0},$$

implicando que

$$\sum_{j=1}^{n} \widehat{e}_j = 0 \text{ e} \sum_{j=1}^{n} \widehat{e}_j x_{ij} = 0, \ \forall i = 1, \ldots, p-1.$$

Vale ressaltar que $\sum_{j=1}^{n} \widehat{e}_j = 0$ é válida somente se o modelo possui intercepto.

Propriedades

P2. Se o modelo possui intercepto, então

$$\sum_{j=1}^{n} y_j = \sum_{j=1}^{n} \widehat{y}_j.$$

A demonstração segue do fato de que $\widehat{e}_j = y_j - \widehat{y}_j, j = 1, \ldots, n$ e que $\sum_{j=1}^{n} \widehat{e}_j = 0$.

P3. De uma maneira geral, dado que $\widehat{\boldsymbol{\beta}}$ é uma transformação linear de $\boldsymbol{y}$ já mostramos que

$$\widehat{\boldsymbol{\beta}} \sim (\boldsymbol{\beta}, \sigma^2(\boldsymbol{X}^\top \boldsymbol{X})^{-1}).$$

P4. Adicionalmente sob a hipótese de normalidade, tem-se que $\boldsymbol{y} \sim \mathcal{N}_n(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2 \boldsymbol{I}_n)$, implicando

$$\widehat{\boldsymbol{\beta}} \sim \mathcal{N}_p(\boldsymbol{\beta}, \sigma^2(\boldsymbol{X}^\top \boldsymbol{X})^{-1}).$$

Propriedades

P5. (**Teorema de Gauss-Markov**) Considere o MRLM com forma funcional $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}$, tal que $\mathbb{E}[\boldsymbol{e}] = \boldsymbol{0}$, $\boldsymbol{\Sigma}_{\boldsymbol{e}} = \sigma^2 \boldsymbol{I}_n$ e $\boldsymbol{c}^\top \boldsymbol{\beta}$ uma função linear de $\boldsymbol{\beta}$, com $\boldsymbol{c} \neq \boldsymbol{0}$. Então, $\boldsymbol{c}^\top \widehat{\boldsymbol{\beta}} = \boldsymbol{c}^\top (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{y}$ é o BLUE (Best Linear Unbiased Estimator) de $\boldsymbol{c}^\top \boldsymbol{\beta}$.

Dem: Perceba que $\boldsymbol{c}^\top \widehat{\boldsymbol{\beta}} = \boldsymbol{c}^\top (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{y}$ é uma transformação linear de $\boldsymbol{y}$, além de ser um estimador não viesado de $\boldsymbol{c}^\top \boldsymbol{\beta}$. Vamos mostrar que dentre os estimadores lineares não viesados de $\boldsymbol{c}^\top \boldsymbol{\beta}$, $\boldsymbol{c}^\top \widehat{\boldsymbol{\beta}}$ é o que possui a menor variância. 😉

Seja $\boldsymbol{\lambda}^\top \boldsymbol{y}$ um outro estimador linear não viesado de $\boldsymbol{c}^\top \boldsymbol{\beta}$, i.e.,

$$\mathbb{E}[\boldsymbol{\lambda}^\top \boldsymbol{y}] = \boldsymbol{c}^\top \boldsymbol{\beta}, \forall \boldsymbol{\beta} \in \mathbb{R}^p,$$

implicando que $\boldsymbol{\lambda}^\top \boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{c}^\top \boldsymbol{\beta}, \forall \boldsymbol{\beta} \in \mathbb{R}^p$,e isso ocorre, se e somente se

$$\boldsymbol{\lambda}^\top \boldsymbol{X} = \boldsymbol{c}^\top. \qquad (3)$$

Por outro lado,

$$\mathrm{Var}[\boldsymbol{\lambda}^\top \boldsymbol{y}] = \boldsymbol{\lambda}^\top \mathrm{Var}[\boldsymbol{y}] \boldsymbol{\lambda} = \sigma^2 \boldsymbol{\lambda}^\top \boldsymbol{I}_n \boldsymbol{\lambda} = \sigma^2 \boldsymbol{\lambda}^\top \boldsymbol{\lambda}. \qquad (4)$$

Propriedades

Lembrando que

$$\mathrm{Var}(\boldsymbol{c}^\top\widehat{\boldsymbol{\beta}}) = \sigma^2\boldsymbol{c}^\top(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{c}, \tag{5}$$

e usando a identidade (3) em (5) obtemos

$$\mathrm{Var}(\boldsymbol{c}^\top\widehat{\boldsymbol{\beta}}) = \sigma^2\boldsymbol{\lambda}^\top\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{\lambda},$$

de forma que

$$\begin{aligned}
\mathrm{Var}(\boldsymbol{\lambda}^\top\boldsymbol{y}) - \mathrm{Var}(\boldsymbol{c}^\top\widehat{\boldsymbol{\beta}}) &= \sigma^2\boldsymbol{\lambda}^\top\boldsymbol{\lambda} - \sigma^2\boldsymbol{\lambda}^\top\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{\lambda} \\
&= \sigma^2\boldsymbol{\lambda}^\top[\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top]\boldsymbol{\lambda} \\
&= \sigma^2\boldsymbol{\lambda}^\top[\boldsymbol{I}_n - \boldsymbol{H}]\boldsymbol{\lambda} \\
&\geq 0,
\end{aligned}$$

pois a matriz $\boldsymbol{I}_n - \boldsymbol{H}$ é simétrica e idempotente, e por conseguinte não-negativa definida (positiva semidefinida). Portanto,

$$\mathrm{Var}(\boldsymbol{c}^\top\widehat{\boldsymbol{\beta}}) \leq \mathrm{Var}(\boldsymbol{\lambda}^\top\boldsymbol{y}).$$

Propriedades

A igualdade entre $\mathrm{Var}(\boldsymbol{c}^\top\boldsymbol{\beta})$ e $\mathrm{Var}(\boldsymbol{\lambda}^\top\boldsymbol{y})$ é válida, se e somente se,

$$\begin{aligned}
\boldsymbol{I}_n - \boldsymbol{H} = \boldsymbol{0} \\
\Leftrightarrow \boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top = \boldsymbol{0} \\
\Leftrightarrow \boldsymbol{\lambda}^\top[\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top] = \boldsymbol{0}, \forall\boldsymbol{\lambda} \in \mathbb{R}^p \\
\Leftrightarrow \boldsymbol{\lambda}^\top = \boldsymbol{\lambda}^\top\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top, \forall\boldsymbol{\lambda} \in \mathbb{R}^p,
\end{aligned}$$

implicando que

$$\boldsymbol{\lambda}^\top\boldsymbol{y} = \boldsymbol{\lambda}^\top\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{y} = \boldsymbol{\lambda}^\top\boldsymbol{X}\widehat{\boldsymbol{\beta}} = \boldsymbol{c}^\top\widehat{\boldsymbol{\beta}},$$

sendo que esta última segue por (3).

Propriedades

P6. Em geral, a variância σ^2 é desconhecida, logo inferências sobre $\boldsymbol{\beta}$ dependem de um estimador desse parâmetro. Sem a necessidade da suposição de normalidade, pode-se mostrar que

$$\widehat{\sigma}^2 = S^2 := \mathrm{QMRes} = \frac{\mathrm{SQRes}}{n-p} = \frac{\boldsymbol{y}^\top[\boldsymbol{I}_n - \boldsymbol{H}]\boldsymbol{y}}{n-p}$$

é o MINQUE de σ^2, em que $p = \mathrm{posto}(\boldsymbol{X})$. Sob normalidade, tem-se adicionalmente que (será demonstrado logo adiante)

$$\frac{\mathrm{SQRes}}{\sigma^2} \sim \chi^2_{(n-p)}.$$

Resultado

Para provar que o estimador anterior é não viesado, basta usar o resultado do teorema abaixo (associado a um exercício anterior solicitado)

Teorema: Se $\boldsymbol{W} \sim \mathcal{N}_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, então:

i) $\mathbb{E}[\boldsymbol{W}^\top \boldsymbol{A} \boldsymbol{W}] = \mathrm{tr}(\boldsymbol{A}\boldsymbol{\Sigma}) + \boldsymbol{\mu}^\top \boldsymbol{A}\boldsymbol{\mu}$. (resultado geral, i.e., não precisa de normalidade).

ii) $\mathrm{Var}[\boldsymbol{W}^\top \boldsymbol{A} \boldsymbol{W}] = 2\mathrm{tr}(\boldsymbol{A}\boldsymbol{\Sigma})^2 + 4\boldsymbol{\mu}^\top \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}\boldsymbol{\mu}$.

iii) $\mathrm{Cov}(\boldsymbol{W}, \boldsymbol{W}^\top \boldsymbol{A} \boldsymbol{W}) = 2\boldsymbol{\Sigma}\boldsymbol{A}\boldsymbol{\mu}$.

Consistência do QMRes

Usando o teorema anterior, podemos provar facilmente a consistência do QMRes. Para isto, basta perceber que

$$\begin{aligned}
\widehat{\boldsymbol{e}} = \boldsymbol{y} - \widehat{\boldsymbol{y}} &= (\boldsymbol{I}_n - \boldsymbol{H})\boldsymbol{y} \\
&= (\boldsymbol{I}_n - \boldsymbol{H})(\boldsymbol{X}\beta + \boldsymbol{e}) \\
&= (\boldsymbol{I}_n - \boldsymbol{H})\boldsymbol{e},
\end{aligned}$$

desde que $(\boldsymbol{I}_n - \boldsymbol{H})\boldsymbol{X} = \boldsymbol{0}$. Logo, podemos reescrever a SQRes como

$$\begin{aligned}
\text{SQres} = ||\widehat{\boldsymbol{e}}||^2 &= \widehat{\boldsymbol{e}}^\top \widehat{\boldsymbol{e}} \\
&= [(\boldsymbol{I}_n - \boldsymbol{H})\boldsymbol{e}]^\top (\boldsymbol{I}_n - \boldsymbol{H})\boldsymbol{e} \\
&= \boldsymbol{e}^\top (\boldsymbol{I}_n - \boldsymbol{H})\boldsymbol{e} \\
&= \boldsymbol{e}^\top \boldsymbol{e} - \boldsymbol{e}^\top \boldsymbol{H}\boldsymbol{e}.
\end{aligned}$$

Consistência do QMRes

Como assumimos que $\boldsymbol{e} \sim (\mathbf{0}, \sigma^2 \boldsymbol{I}_n)$, então pela lei fraca dos grandes números de Khintchine, assumindo independência no lugar de não correlacionadas), temos que

$$\frac{\boldsymbol{e}^\top \boldsymbol{e}}{n} = \frac{\sum_{i=1}^n e_i^2}{n} \xrightarrow{\mathbb{P}} \sigma^2.$$

Logo (pelo Teorema de Slustky), temos

$$\frac{\boldsymbol{e}^\top \boldsymbol{e}}{n} = \underbrace{\frac{\boldsymbol{e}^\top \boldsymbol{e}}{n-p}}_{\xrightarrow{\mathbb{P}} \sigma^2} \underbrace{\frac{n-p}{n}}_{\to 1} \xrightarrow{\mathbb{P}} \sigma^2.$$

Por outro lado, dado que $\boldsymbol{H}$ é idempotente, temos que

$$\boldsymbol{e}^\top \boldsymbol{H} \boldsymbol{e} = \boldsymbol{e}^\top \boldsymbol{H}^2 \boldsymbol{e} = ||\boldsymbol{H}\boldsymbol{e}||^2 \geq 0.$$

Consistência do QMRes

Logo, $\forall \epsilon > 0$, tem-se que

$$\begin{aligned}
\mathbb{P}\left\{\left|\frac{\boldsymbol{e}^\top \boldsymbol{H}\boldsymbol{e}}{n-p}\right| > \epsilon\right\} &= \mathbb{P}\left\{\boldsymbol{e}^\top \boldsymbol{H}\boldsymbol{e} > \epsilon(n-p)\right\} \\
&\leq \frac{\mathbb{E}\left[\boldsymbol{e}^\top \boldsymbol{H}\boldsymbol{e}\right]}{(n-p)\epsilon} \quad \text{(des. Chebychev)} \\
&= \frac{\mathbb{E}[\mathrm{tr}\{\boldsymbol{e}^\top \boldsymbol{H}\boldsymbol{e}\}]}{(n-p)\epsilon} = \frac{\mathbb{E}[\mathrm{tr}\{\boldsymbol{H}\boldsymbol{e}\boldsymbol{e}^\top\}}{(n-p)\epsilon} \\
&= \frac{\mathrm{tr}\{\mathbb{E}[\boldsymbol{H}\boldsymbol{e}\boldsymbol{e}^\top]\}}{(n-p)\epsilon} = \frac{\mathrm{tr}\{\boldsymbol{H}\mathbb{E}[\boldsymbol{e}\boldsymbol{e}^\top]\}}{(n-p)\epsilon} \\
&= \frac{\mathrm{tr}\{\boldsymbol{H}\sigma^2\boldsymbol{I}\}}{(n-p)\epsilon} = \frac{\sigma^2 p}{n-p} \to 0,
\end{aligned}$$

quando $n \to \infty$. Portanto, $\boldsymbol{e}^\top \boldsymbol{H}\boldsymbol{e}/n-p \xrightarrow{\mathbb{P}} 0$, implicando pelo teorema de Slustky que

$$\mathrm{QMRes} = \frac{\mathrm{SQres}}{n-p} = \frac{\boldsymbol{e}^\top \boldsymbol{e}}{n-p} - \frac{\boldsymbol{e}^\top \boldsymbol{H}\boldsymbol{e}}{n-p} \xrightarrow{\mathbb{P}} \sigma^2.$$

■

Exercícios (Entregar próxima aula)

Exercício 6: Considere o MRLM sob a suposição de normalidade. Obtenha o EMV de $(\boldsymbol{\beta}^\top, \sigma^2)^\top$ e sua distribuição assintótica. Os parâmetros $\boldsymbol{\beta}$ e σ^2 são ortogonais?

Exercício 7: Considere o MRLM sob a suposição de normalidade. Mostre que $\widehat{\boldsymbol{\beta}}$ e SQRes são independentes.

Dica: Sob a suposição de normalidade, temos que

$$\boldsymbol{y} \sim \mathcal{N}_n(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2 \boldsymbol{I}_n).$$

Além disso, como $\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{y} = \boldsymbol{B}\boldsymbol{y}$ (transformação linear) e $\mathrm{SQRes} = \boldsymbol{y}^\top(\boldsymbol{I}_n - \boldsymbol{H})\boldsymbol{y} = \boldsymbol{y}^\top\boldsymbol{A}\boldsymbol{y}$ (forma quadrática), então pode-se usar o teorema do slide seguinte para provar o que se deseja.

Exercício 8: Provar o resultado do exercício anterior usando o teorema de Basu.

Resultado

Para provar que $\widehat{\boldsymbol{\beta}}$ e SQRes podemos usar o resultado do teorema abaixo (associado a um exercício anterior solicitado)

Teorema: Se $\boldsymbol{W} \sim \mathcal{N}_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, então $\boldsymbol{W}^\top \boldsymbol{A}\boldsymbol{W}$ e $\boldsymbol{B}\boldsymbol{W}$ são independentes, se e somente se

$$\boldsymbol{B}\boldsymbol{\Sigma}\boldsymbol{A} = \mathbf{0}.$$

Outro resultado extremamente útil que faremos uso é sobre a independência de formas quadráticas (também pedido anteriormente como exercício):

Teorema: Se $\boldsymbol{W} \sim \mathcal{N}_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, então $\boldsymbol{W}^\top \boldsymbol{A}\boldsymbol{W}$ e $\boldsymbol{W}^\top \boldsymbol{B}\boldsymbol{W}$ são independentes, se e somente se

$$\boldsymbol{B}\boldsymbol{\Sigma}\boldsymbol{A} = \mathbf{0} \Leftrightarrow \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{B} = \mathbf{0}.$$

Decomposição da Soma de Quadrados Total

Considere o MRLM (1) $\boldsymbol{y} = \boldsymbol{X}\beta + \boldsymbol{e}$ com respectiva matriz de especificação $\boldsymbol{X} = (\mathbf{1}_n, \boldsymbol{X}_R)$, em que $\boldsymbol{X}_R$ tem dimensão $n \times (p-1)$ e corresponde a matriz de especificação de um MRLM sem intercepto.

De forma similar ao que já foi visto anteriormente, temos válida (modelo com intercepto) a seguinte decomposição

$$\mathrm{SQT} = \mathrm{SQRes} + \mathrm{SQReg},$$

em que

$$\begin{aligned}
\mathrm{SQT} &= \sum_{i=1}^{n}(y_i - \bar{y}_n)^2 = \boldsymbol{y}^\top \left(\boldsymbol{I}_n - \frac{\boldsymbol{J}_n}{n}\right)\boldsymbol{y}, \\
\mathrm{SQRes} &= \sum_{i=1}^{n}(y_i - \widehat{y}_i)^2 = \boldsymbol{y}^\top(\boldsymbol{I}_n - \boldsymbol{H})\boldsymbol{y}, \\
\mathrm{SQReg} &= \sum_{i=1}^{n}(\widehat{y}_i - \bar{y}_n)^2 = \boldsymbol{y}^\top \left(\boldsymbol{H} - \frac{\boldsymbol{J}_n}{n}\right)\boldsymbol{y}.
\end{aligned}$$

Propriedades

As matrizes núcleo $(\boldsymbol{I}_n - \boldsymbol{J}_n/n)$ e $(\boldsymbol{I}_n - \boldsymbol{H})$ são simétricas e idempotentes, implicando que SQT e SQRes possuem respectivamente, $\mathrm{posto}(\boldsymbol{I}_n - \boldsymbol{J}_n/n) = n-1$ e $\mathrm{posto}(\boldsymbol{I}_n - \boldsymbol{H}) = n-p$, graus de liberdade.

Agora, vamos usar o seguinte teorema já conhecido de todos nós (exercício anterior):

__**Teorema**__: Se $\boldsymbol{W} \sim \mathcal{N}_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, então $\boldsymbol{W}^\top \boldsymbol{A} \boldsymbol{W} \sim \chi^2_{[\mathrm{posto}(\boldsymbol{A}), \frac{1}{2}\boldsymbol{\mu}^\top \boldsymbol{A}\boldsymbol{\mu}]}$ se e somente se $\boldsymbol{A}\boldsymbol{\Sigma}$ é idempotente.

Usando o teorema acima, tem-se diretamente que (exercício para fazer agora)

$$\frac{\mathrm{SQRes}}{\sigma^2} \sim \chi^2_{(n-p)}.$$

Propriedades

Por outro lado,

$$\mathrm{SQReg} = \boldsymbol{y}^\top \left(\boldsymbol{H} - \frac{\boldsymbol{J}_n}{n} \right) \boldsymbol{y},$$

com $\boldsymbol{X} = (\mathbf{1}_n, \boldsymbol{X}_R)$, implicando que

$$\begin{aligned} \boldsymbol{X}^\top \boldsymbol{X} &= \begin{pmatrix} \mathbf{1}_n^\top \\ \boldsymbol{X}_R^\top \end{pmatrix} (\mathbf{1}_n, \boldsymbol{X}_R) \\ &= \left(\begin{array}{c|c} \mathbf{1}_n^\top \mathbf{1}_n & \mathbf{1}_n^\top \boldsymbol{X}_R \\ \hline \boldsymbol{X}_R^\top \mathbf{1}_n & \boldsymbol{X}_R^\top \boldsymbol{X}_R \end{array} \right) \end{aligned}$$

de forma que usando a inversa para matrizes particionadas (exercício) é possível mostrar que

$$\boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \mathbf{1}_n = \mathbf{1}_n \; (\boldsymbol{H}\mathbf{1}_n = \mathbf{1}_n) \text{ e } \boldsymbol{X}_R^\top \boldsymbol{H} = \boldsymbol{X}_R^\top$$

de maneira na qual fica fácil provar que a matriz núcleo $\boldsymbol{H} - \boldsymbol{J}_n/n$ é idempotente.

Propriedades

Portanto, sob a suposição de normalidade, temos que

$$\frac{\mathrm{SQReg}}{\sigma^2} = \frac{\boldsymbol{y}^\top \left(\boldsymbol{H} - \frac{\boldsymbol{J}_n}{n}\right)\boldsymbol{y}}{\sigma^2} \sim \chi^2_{[\mathrm{posto}(\boldsymbol{H}-\boldsymbol{J}_n/n),\lambda]},$$

com $\mathrm{posto}(\boldsymbol{H} - \boldsymbol{J}_n/n) = \mathrm{tr}(\boldsymbol{H} - \boldsymbol{J}_n/n) = p - 1$ pois $\boldsymbol{H} - \boldsymbol{J}_n/n$ é idempotente, e parâmetro de não-centralidade dado por

$$\lambda = \frac{1}{2\sigma^2}\mathbb{E}[\boldsymbol{y}]^\top(\boldsymbol{H} - \boldsymbol{J}_n/n)\mathbb{E}[\boldsymbol{y}].$$

Perceba que

$$\mathbb{E}[\boldsymbol{y}] = \boldsymbol{X}\boldsymbol{\beta} = (\mathbf{1}, \boldsymbol{X}_R)(\beta_0, \boldsymbol{\beta}_R^\top)^\top = \beta_0\mathbf{1}_n + \boldsymbol{X}_R\boldsymbol{\beta}_R.$$

Logo,

$$\begin{aligned}
\mathbb{E}[\boldsymbol{y}]^\top(\boldsymbol{H} - \boldsymbol{J}_n/n)\mathbb{E}[\boldsymbol{y}] &= \boldsymbol{\beta}^\top\boldsymbol{X}^\top(\boldsymbol{H} - \boldsymbol{J}_n/n)\boldsymbol{X}\boldsymbol{\beta} \\
&= (\beta_0, \boldsymbol{\beta}_R^\top)(\mathbf{1}_n, \boldsymbol{X}_R)^\top[\boldsymbol{H} - \boldsymbol{J}_n/n](\mathbf{1}_n, \boldsymbol{X}_R)(\beta_0, \boldsymbol{\beta}_R^\top)^\top
\end{aligned}$$

Propriedades

$$
\begin{aligned}
\mathbb{E}[\boldsymbol{y}]^\top(\boldsymbol{H}-\boldsymbol{J}_n/n)\mathbb{E}[\boldsymbol{y}] &= (\beta_0,\boldsymbol{\beta}_R^\top)(\boldsymbol{1}_n,\boldsymbol{X}_R)^\top[\boldsymbol{H}-\boldsymbol{J}_n/n](\boldsymbol{1}_n,\boldsymbol{X}_R)(\beta_0,\boldsymbol{\beta}_R^\top)^\top \\
&= (\beta_0,\boldsymbol{\beta}_R^\top)\begin{pmatrix} \boldsymbol{1}_n^\top\boldsymbol{H}-\boldsymbol{1}_n\boldsymbol{J}_n/n \\ \boldsymbol{X}_R^\top\boldsymbol{H}-\boldsymbol{X}_R\boldsymbol{J}_n/n \end{pmatrix}(\boldsymbol{1},\boldsymbol{X}_R)(\beta_0,\boldsymbol{\beta}_R^\top)^\top,
\end{aligned}
$$

todavia, dado que $\boldsymbol{1}_n^\top\boldsymbol{H}=\boldsymbol{1}_n$, $\boldsymbol{1}_n^\top\boldsymbol{J}_n=n\boldsymbol{1}_n^\top$ e $\boldsymbol{X}_R^\top\boldsymbol{H}=\boldsymbol{X}_R^\top$, temos que

$$
\begin{aligned}
\mathbb{E}[\boldsymbol{y}]^\top(\boldsymbol{H}-\boldsymbol{J}_n/n)\mathbb{E}[\boldsymbol{y}] &= (\beta_0,\boldsymbol{\beta}_R^\top)\begin{pmatrix} \boldsymbol{0}^\top \\ \boldsymbol{X}_R^\top\boldsymbol{H}-\boldsymbol{X}_R^\top\boldsymbol{J}_n/n \end{pmatrix}(\boldsymbol{1},\boldsymbol{X}_R)(\beta_0,\boldsymbol{\beta}_R^\top)^\top \\
&= (\beta_0,\boldsymbol{\beta}_R^\top)\begin{pmatrix} 0 & \boldsymbol{0}^\top \\ \boldsymbol{0}^\top & \boldsymbol{X}_R^\top\boldsymbol{X}_R-\boldsymbol{X}_R^\top(\boldsymbol{J}_n/n)\boldsymbol{X}_R \end{pmatrix}(\beta_0,\boldsymbol{\beta}_R^\top)^\top \\
&= \boldsymbol{\beta}_R^\top[\boldsymbol{X}_R^\top\boldsymbol{X}_R-\boldsymbol{X}_R^\top(\boldsymbol{J}_n/n)\boldsymbol{X}_R]\boldsymbol{\beta}_R.
\end{aligned}
$$

Propriedades

Se considerarmos as variáveis centralizadas, de forma a obter a matriz de especificação $\boldsymbol{X}_C$ centralizada, então o parâmetro de não-centralidade fica dado por

$$\lambda = \frac{1}{2\sigma^2}\boldsymbol{\beta}_R^\top \boldsymbol{X}_C^\top (\boldsymbol{J}_n/n)\boldsymbol{X}_C\boldsymbol{\beta}_R,$$

implicando que $\mathbb{E}[\mathrm{SQReg}] = (p-1)\sigma^2 + \boldsymbol{\beta}_R^\top \boldsymbol{X}_C^\top (\boldsymbol{J}_n/n)\boldsymbol{X}_C\boldsymbol{\beta}_R$, ou seja, o QMReg é um estimador não viesado para σ^2 se e somente se o parâmetro de não-centralidade for nulo, e isso ocorre, se e somente se, $\boldsymbol{\beta}_R = (\beta_1, \ldots, \beta_{p-1})^\top = \mathbf{0}$.

Propriedades

Além disso, tem-se que

$$\begin{aligned}\mathrm{SQRes} &= \boldsymbol{y}^\top(\boldsymbol{I}_n - \boldsymbol{H})\boldsymbol{y} = \boldsymbol{y}^\top\boldsymbol{y} - \widehat{\boldsymbol{\beta}}^\top\boldsymbol{X}^\top\boldsymbol{y}\\ \mathrm{SQReg} &= \boldsymbol{y}^\top\left(\boldsymbol{H} - \frac{\boldsymbol{J}_n}{n}\right)\boldsymbol{y} = \widehat{\boldsymbol{\beta}}^\top\boldsymbol{X}^\top\boldsymbol{y} - n\bar{y}_n^2.\end{aligned}$$

então, dado que $\boldsymbol{y} \sim \mathcal{N}_n(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2\boldsymbol{I}_n)$ e lembrando que $\boldsymbol{H}\mathbf{1}_n = \mathbf{1}_n \Rightarrow \boldsymbol{H}\boldsymbol{J}_n = \boldsymbol{J}_n$, tem-se

$$(\boldsymbol{I}_n - \boldsymbol{H})\sigma^2\boldsymbol{I}_n\left(\boldsymbol{H} - \frac{\boldsymbol{J}_n}{n}\right) = \sigma^2(\boldsymbol{I}_n - \boldsymbol{H})\left(\boldsymbol{H} - \frac{\boldsymbol{J}_n}{n}\right) = \sigma^2(\boldsymbol{H} - \boldsymbol{J}_n/n - \boldsymbol{H} + \boldsymbol{H}\boldsymbol{J}_n/n) = \mathbf{0},$$

implicando que $\mathrm{SQReg} \perp \mathrm{SQres}$.

ANOVA

Sob $\mathcal{H}_0 : \beta_1 = \ldots = \beta_{p-1} = 0(\boldsymbol{\beta}_R = \mathbf{0})$, tem-se que $\mathrm{SQReg}/\sigma^2 \sim \chi^2_{(p-1)}$.

Logo, De posse dos resultados anteriores, podemos construir o seguinte quadro de Análise de Variância (ANOVA) do MRLM:

Causas de Variação	GL	SQ	QM	F
Regressão	$p-1$	$\boldsymbol{y}^\top(\boldsymbol{H} - \boldsymbol{J}_n/n)\boldsymbol{y}$	$\mathrm{SQReg}/(p-1) = \mathrm{QMReg}$	$\dfrac{\mathrm{QMReg}}{\mathrm{QMRes}}$
Resíduo	$n-p$	$\boldsymbol{y}^\top(\boldsymbol{I}_n - \boldsymbol{H})\boldsymbol{y}$	$\mathrm{SQRes}/(n-p) = \mathrm{QMRes}$	-
Total	$n-1$	$\boldsymbol{y}^\top(\boldsymbol{I}_n - \boldsymbol{J}_n/n)\boldsymbol{y}$	$\mathrm{SQT}/(n-1)$	-

ANOVA

- Sob a hipótese de normalidade e sob a hipótese de não existência de regressão, i.e.,

$$\mathcal{H}_0 : \beta_1 = \beta_2 = \ldots = \beta_{p-1} = 0,$$

tem-se que

$$F_0 := \frac{\text{QMReg}}{\text{QMRes}} \sim \mathcal{F}(p-1, n-p).$$

- Sob a hipótese alternativa $\mathcal{H}_1 : \exists j \in \{1, ..., p-1\}; \beta_j \neq 0$, tem-se que

$$F_0 := \frac{\text{QMReg}}{\text{QMRes}} \sim \mathcal{F}(p-1, n-p, \lambda),$$

de forma que rejeitamos $\mathcal{H}_0$ ao nível α se $F_0 > \mathcal{F}_{(1-\alpha)}(p-1, n-p)$.

ANOVA

- Sob a hipótese de normalidade e sob a hipótese de não existência de regressão, i.e.,

$$\mathcal{H}_0 : \beta_1 = \beta_2 = \ldots = \beta_{p-1} = 0,$$

tem-se que

$$F_0 := \frac{\text{QMReg}}{\text{QMRes}} \sim \mathcal{F}(p-1, n-p).$$

- Sob a hipótese alternativa $\mathcal{H}_1 : \exists j \in \{1, ..., p-1\}; \beta_j \neq 0$, tem-se que

$$F_0 := \frac{\text{QMReg}}{\text{QMRes}} \sim \mathcal{F}(p-1, n-p, \lambda),$$

de forma que rejeitamos $\mathcal{H}_0$ ao nível α se $F_0 > \mathcal{F}_{(1-\alpha)}(p-1, n-p)$.

Exercício (Entregar próxima aula)

Exercício 9: Considere o MRLM de intercepto nulo. Reescreva-o usando a notação matricial, obtenha a distribuição das somas de quadrados correspondentes e a tabela de ANOVA.

Coeficiente de determinação

- Sabemos que **coeficiente de determinação**, definido por

$$R^2 := \frac{\text{SQReg}}{\text{SQT}} = 1 - \frac{\text{SQRes}}{\text{SQT}}, \tag{6}$$

representa a proporção da variabilidade da variável resposta que é explicada pelo modelo de regressão.

- Um dos problemas na utilização do coeficiente de determinação como medida de qualidade do modelo é que ele não leva em consideração o número de parâmetros envolvidos.
- Se aumentarmos o número de variáveis explicativas, o que acontece com o R^2?
- Quanto mais variáveis explicativas forem introduzidas no modelo, mais o coeficiente R^2 se aproximará de 1.
- Para contornar este problema, pode-se considerar o coeficiente de determinação ajustado pelos graus de liberdade, dado por

$$R_a^2 = 1 - \left(\frac{n-1}{n-p}\right)\frac{\text{SQRes}}{\text{SQT}}.$$

Coeficiente de determinação

- Sabemos que **coeficiente de determinação**, definido por

$$R^2 := \frac{\mathrm{SQReg}}{\mathrm{SQT}} = 1 - \frac{\mathrm{SQRes}}{\mathrm{SQT}}, \tag{6}$$

representa a proporção da variabilidade da variável resposta que é explicada pelo modelo de regressão.

- Um dos problemas na utilização do coeficiente de determinação como medida de qualidade do modelo é que ele não leva em consideração o número de parâmetros envolvidos.
- Se aumentarmos o número de variáveis explicativas, o que acontece com o R^2?
- Quanto mais variáveis explicativas forem introduzidas no modelo, mais o coeficiente R^2 se aproximará de 1.
- Para contornar este problema, pode-se considerar o coeficiente de determinação ajustado pelos graus de liberdade, dado por

$$R_a^2 = 1 - \left(\frac{n-1}{n-p}\right)\frac{\mathrm{SQRes}}{\mathrm{SQT}}.$$

Coeficiente de determinação

- Sabemos que **coeficiente de determinação**, definido por

$$R^2 := \frac{\text{SQReg}}{\text{SQT}} = 1 - \frac{\text{SQRes}}{\text{SQT}}, \tag{6}$$

representa a proporção da variabilidade da variável resposta que é explicada pelo modelo de regressão.

- Um dos problemas na utilização do coeficiente de determinação como medida de qualidade do modelo é que ele não leva em consideração o número de parâmetros envolvidos.
- Se aumentarmos o número de variáveis explicativas, o que acontece com o R^2? 😇
- Quanto mais variáveis explicativas forem introduzidas no modelo, mais o coeficiente R^2 se aproximará de 1.
- Para contornar este problema, pode-se considerar o coeficiente de determinação ajustado pelos graus de liberdade, dado por

$$R_a^2 = 1 - \left(\frac{n-1}{n-p}\right)\frac{\text{SQRes}}{\text{SQT}}.$$

Coeficiente de determinação

- Sabemos que **coeficiente de determinação**, definido por
$$R^2 := \frac{\text{SQReg}}{\text{SQT}} = 1 - \frac{\text{SQRes}}{\text{SQT}}, \tag{6}$$
representa a proporção da variabilidade da variável resposta que é explicada pelo modelo de regressão.

- Um dos problemas na utilização do coeficiente de determinação como medida de qualidade do modelo é que ele não leva em consideração o número de parâmetros envolvidos.

- Se aumentarmos o número de variáveis explicativas, o que acontece com o R^2?

- Quanto mais variáveis explicativas forem introduzidas no modelo, mais o coeficiente R^2 se aproximará de 1.

- Para contornar este problema, pode-se considerar o coeficiente de determinação ajustado pelos graus de liberdade, dado por
$$R_a^2 = 1 - \left(\frac{n-1}{n-p}\right) \frac{\text{SQRes}}{\text{SQT}}.$$

Coeficiente de determinação

- Sabemos que **coeficiente de determinação**, definido por
$$R^2 := \frac{\mathrm{SQReg}}{\mathrm{SQT}} = 1 - \frac{\mathrm{SQRes}}{\mathrm{SQT}}, \tag{6}$$
representa a proporção da variabilidade da variável resposta que é explicada pelo modelo de regressão.
- Um dos problemas na utilização do coeficiente de determinação como medida de qualidade do modelo é que ele não leva em consideração o número de parâmetros envolvidos.
- Se aumentarmos o número de variáveis explicativas, o que acontece com o R^2? 😇
- Quanto mais variáveis explicativas forem introduzidas no modelo, mais o coeficiente R^2 se aproximará de 1.
- Para contornar este problema, pode-se considerar o coeficiente de determinação ajustado pelos graus de liberdade, dado por
$$R_a^2 = 1 - \left(\frac{n-1}{n-p}\right)\frac{\mathrm{SQRes}}{\mathrm{SQT}}.$$

Coeficiente de determinação

- O coeficiente R_a^2 pode diminuir quando adicionamos variáveis explicativas ao modelo pois o decréscimo que isso acarreta na soma de quadrados dos resíduos pode ser compensado pela perda de graus de liberdade no denominador $(n - p)$.
- Já vimos que o coeficiente de determinação sozinho pode não ser informativo para avaliar a qualidade de um ajuste.
- Outras medidas de qualidade de ajuste são utilizadas na prática, como por exemplo, `AIC`, `BIC`, `CAIC`, etc... todas baseadas em teoria da informação e usam o princípio da parcimônia do modelo como chave.
- Apesar de seu uso **indiscriminado**, esses critérios só podem/devem ser utilizados se os modelos competidores são encaixados/hierárquicos, por exemplo, possuindo a mesma verossimilhança.

Coeficiente de determinação

- O coeficiente R_a^2 pode diminuir quando adicionamos variáveis explicativas ao modelo pois o decréscimo que isso acarreta na soma de quadrados dos resíduos pode ser compensado pela perda de graus de liberdade no denominador $(n - p)$.
- Já vimos que o coeficiente de determinação sozinho pode não ser informativo para avaliar a qualidade de um ajuste.
- Outras medidas de qualidade de ajuste são utilizadas na prática, como por exemplo, `AIC`, `BIC`, `CAIC`, etc... todas baseadas em teoria da informação e usam o princípio da parcimônia do modelo como chave.
- Apesar de seu uso **indiscriminado**, esses critérios só podem/devem ser utilizados se os modelos competidores são encaixados/hierárquicos, por exemplo, possuindo a mesma verossimilhança.

Coeficiente de determinação

- O coeficiente R_a^2 pode diminuir quando adicionamos variáveis explicativas ao modelo pois o decréscimo que isso acarreta na soma de quadrados dos resíduos pode ser compensado pela perda de graus de liberdade no denominador $(n - p)$.
- Já vimos que o coeficiente de determinação sozinho pode não ser informativo para avaliar a qualidade de um ajuste.
- Outras medidas de qualidade de ajuste são utilizadas na prática, como por exemplo, `AIC`, `BIC`, `CAIC`, etc... todas baseadas em teoria da informação e usam o princípio da parcimônia do modelo como chave.
- Apesar de seu uso **indiscriminado**, esses critérios só podem/devem ser utilizados se os modelos competidores são encaixados/hierárquicos, por exemplo, possuindo a mesma verossimilhança.

Coeficiente de determinação

- O coeficiente R_a^2 pode diminuir quando adicionamos variáveis explicativas ao modelo pois o decréscimo que isso acarreta na soma de quadrados dos resíduos pode ser compensado pela perda de graus de liberdade no denominador $(n - p)$.
- Já vimos que o coeficiente de determinação sozinho pode não ser informativo para avaliar a qualidade de um ajuste.
- Outras medidas de qualidade de ajuste são utilizadas na prática, como por exemplo, `AIC`, `BIC`, `CAIC`, etc... todas baseadas em teoria da informação e usam o princípio da parcimônia do modelo como chave.
- Apesar de seu uso **indiscriminado**, esses critérios só podem/devem ser utilizados se os modelos competidores são encaixados/hierárquicos, por exemplo, possuindo a mesma verossimilhança.

Coeficiente de determinação

- O R_a^2 pode ser utilizado como um critério de seleção de variáveis, embora ele não represente em si um teste de hipóteses. Veremos em breve testes adequados para inclusão/exclusão de variáveis, bem como de forma sucinta métodos de seleção de variáveis. Estes métodos devem ser apresentados e discutidos em detalhes nos seminários finais.
- Os métodos de diagnóstico que apresentamos também podem servir com base para selecionar modelos.

Coeficiente de determinação

- O R_a^2 pode ser utilizado como um critério de seleção de variáveis, embora ele não represente em si um teste de hipóteses. Veremos em breve testes adequados para inclusão/exclusão de variáveis, bem como de forma sucinta métodos de seleção de variáveis. Estes métodos devem ser apresentados e discutidos em detalhes nos seminários finais.
- Os métodos de diagnóstico que apresentamos também podem servir com base para selecionar modelos.

Exercício (entregar próxima aula) - Questão da 2a AP

Exercício 10: Faça um ensaio sobre os critérios de informação `AIC` e `BIC`. Apresente ao menos um exemplo de utilização, constando o respectivo ajuste e implementação (em qualquer software de interesse), no contexto de modelos de regressão lineares.

O perigo do gráfico de dispersão múltiplo

- No contexto de modelo empíricos a análise descritiva é de fundamental importância.
- Por exemplo, quando estamos no caso associado a um MRLS, já vimos que um diagrama de dispersão pode ser bastante útil para especificar a forma funcional a ser utilizada.
- Portanto para o caso mais geral, era de se esperar que examinando pura e simplesmente os diagramas de dispersão de y vs $x_1, x_2, \ldots, x_{p-1}$ já fosse suficiente para se ter uma ideia das relações.
- Note, entretanto, que tais gráficos evidenciam apenas as relações marginais e podem carregar um alto poder de *confundimento*.

O perigo do gráfico de dispersão múltiplo

- No contexto de modelo empíricos a análise descritiva é de fundamental importância.
- Por exemplo, quando estamos no caso associado a um MRLS, já vimos que um diagrama de dispersão pode ser bastante útil para especificar a forma funcional a ser utilizada.
- Portanto para o caso mais geral, era de se esperar que examinando pura e simplesmente os diagramas de dispersão de y vs $x_1, x_2, \ldots, x_{p-1}$ já fosse suficiente para se ter uma ideia das relações.
- Note, entretanto, que tais gráficos evidenciam apenas as relações marginais e podem carregar um alto poder de *confundimento*.

O perigo do gráfico de dispersão múltiplo

- No contexto de modelo empíricos a análise descritiva é de fundamental importância.
- Por exemplo, quando estamos no caso associado a um MRLS, já vimos que um diagrama de dispersão pode ser bastante útil para especificar a forma funcional a ser utilizada.
- Portanto para o caso mais geral, era de se esperar que examinando pura e simplesmente os diagramas de dispersão de y vs $x_1, x_2, \ldots, x_{p-1}$ já fosse suficiente para se ter uma ideia das relações.
- Note, entretanto, que tais gráficos evidenciam apenas as relações marginais e podem carregar um alto poder de *confundimento*.

O perigo do gráfico de dispersão múltiplo

- No contexto de modelo empíricos a análise descritiva é de fundamental importância.
- Por exemplo, quando estamos no caso associado a um MRLS, já vimos que um diagrama de dispersão pode ser bastante útil para especificar a forma funcional a ser utilizada.
- Portanto para o caso mais geral, era de se esperar que examinando pura e simplesmente os diagramas de dispersão de y vs $x_1, x_2, \ldots, x_{p-1}$ já fosse suficiente para se ter uma ideia das relações.
- Note, entretanto, que tais gráficos evidenciam apenas as relações marginais e podem carregar um alto poder de *confundimento*.

O perigo do gráfico de dispersão múltiplo

Daniel and Wood (1980, Fitting Equations to Data , 2nd ed. , Wiley) e Montgomery et al. (2012, pag. 84) mostram a ineficácia do gráfico de dispersão múltiplo através do seguinte exemplo baseado no modelo **matemático** com 2 regressores (variáveis explicativas)

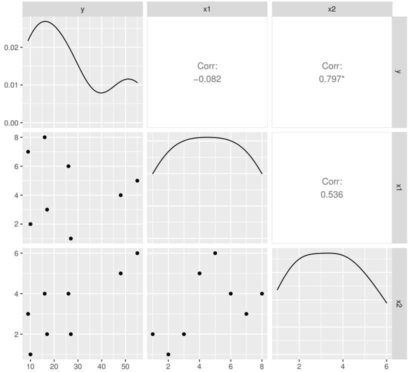
$$y_i = 8 - 5x_{1i} + 12x_{2i}. \tag{7}$$

Com base neste modelo, gerou-se os dados

y	x_1	x_2
10	2	1
17	3	2
48	4	5
27	1	2
55	5	6
26	6	4
9	7	3
16	8	4

O perigo do gráfico de dispersão múltiplo

Figura: Gráfico de dispersão múltiplo para os dados gerados via modelo (7)



O perigo do gráfico de dispersão múltiplo

- O gráfico de dispersão de y vs. x_1 não exibe nenhuma relação aparente entre as variáveis, inclusive não rejeitamos a hipótese de correlação nula ao nível de 5%.
- Analisando o gráfico de dispersão de y vs. x_2 temos indício de relação linear e ajustando um MRLS obtemos uma estimativa igual a $8,1$.
- Ambos os gráficos nos fornecem informações incoerentes com o modelo gerador dos dados.
- Este exemplo ilustra que a construção de diagramas de dispersão de variável resposta vs variáveis explicativas pode ser extremamente ineficaz, mesmo no caso de apenas dois regressores em um modelo sem fonte de variação (matemático).

O perigo do gráfico de dispersão múltiplo

- O gráfico de dispersão de y vs. x_1 não exibe nenhuma relação aparente entre as variáveis, inclusive não rejeitamos a hipótese de correlação nula ao nível de 5%.
- Analisando o gráfico de dispersão de y vs. x_2 temos indício de relação linear e ajustando um MRLS obtemos uma estimativa igual a $8,1$.
- Ambos os gráficos nos fornecem informações incoerentes com o modelo gerador dos dados.
- Este exemplo ilustra que a construção de diagramas de dispersão de variável resposta vs variáveis explicativas pode ser extremamente ineficaz, mesmo no caso de apenas dois regressores em um modelo sem fonte de variação (matemático).

O perigo do gráfico de dispersão múltiplo

- O gráfico de dispersão de y vs. x_1 não exibe nenhuma relação aparente entre as variáveis, inclusive não rejeitamos a hipótese de correlação nula ao nível de 5%.
- Analisando o gráfico de dispersão de y vs. x_2 temos indício de relação linear e ajustando um MRLS obtemos uma estimativa igual a $8,1$.
- Ambos os gráficos nos fornecem informações incoerentes com o modelo gerador dos dados.
- Este exemplo ilustra que a construção de diagramas de dispersão de variável resposta vs variáveis explicativas pode ser extremamente ineficaz, mesmo no caso de apenas dois regressores em um modelo sem fonte de variação (matemático).

O perigo do gráfico de dispersão múltiplo

- O gráfico de dispersão de y vs. x_1 não exibe nenhuma relação aparente entre as variáveis, inclusive não rejeitamos a hipótese de correlação nula ao nível de 5%.
- Analisando o gráfico de dispersão de y vs. x_2 temos indício de relação linear e ajustando um MRLS obtemos uma estimativa igual a $8,1$.
- Ambos os gráficos nos fornecem informações incoerentes com o modelo gerador dos dados.
- Este exemplo ilustra que a construção de diagramas de dispersão de variável resposta vs variáveis explicativas pode ser extremamente ineficaz, mesmo no caso de apenas dois regressores em um modelo sem fonte de variação (matemático).

O perigo do gráfico de dispersão múltiplo

- Em uma situação com mais regressores em um modelo estatístico (i.e. com fonte de variação) o confundimento poderia ser ainda maior.
- Se há apenas um (ou poucos) regressor(es) dominante(s), ou se os regressores operam quase independentemente, o gráfico de dispersão múltiplo (matriz de dispersão) é mais útil.
- No entanto, quando vários regressores importantes são correlacionados (de forma não perfeita, evidentemente), então esses diagramas de dispersão podem ser extremamente confundidos. Para minimizar isto é necessário inicialmente selecionar bem as variáveis explicativas a serem consideradas no modelo.

O perigo do gráfico de dispersão múltiplo

- Em uma situação com mais regressores em um modelo estatístico (i.e. com fonte de variação) o confundimento poderia ser ainda maior.
- Se há apenas um (ou poucos) regressor(es) dominante(s), ou se os regressores operam quase independentemente, o gráfico de dispersão múltiplo (matriz de dispersão) é mais útil.
- No entanto, quando vários regressores importantes são correlacionados (de forma não perfeita, evidentemente), então esses diagramas de dispersão podem ser extremamente confundidos. Para minimizar isto é necessário inicialmente selecionar bem as variáveis explicativas a serem consideradas no modelo.

O perigo do gráfico de dispersão múltiplo

- Em uma situação com mais regressores em um modelo estatístico (i.e. com fonte de variação) o confundimento poderia ser ainda maior.
- Se há apenas um (ou poucos) regressor(es) dominante(s), ou se os regressores operam quase independentemente, o gráfico de dispersão múltiplo (matriz de dispersão) é mais útil.
- No entanto, quando vários regressores importantes são correlacionados (de forma não perfeita, evidentemente), então esses diagramas de dispersão podem ser extremamente confundidos. Para minimizar isto é necessário inicialmente selecionar bem as variáveis explicativas a serem consideradas no modelo.

Uso de variáveis centralizadas

Assim como no MRLS, é comum centralizar as variáveis explicativas em um MRLM, obtendo o seguinte modelo

$$y_i = \beta_0^* + \beta_1(x_{1i} - \bar{x}_1) + \beta_2(x_{2i} - \bar{x}_2) + \ldots + \beta_{p-1}(x_{(p-1)i} - \bar{x}_{p-1}) + e_i,\ i = 1, \ldots, n,$$

ou simplesmente

$$y_i = \beta_0^* + \beta_1 x_{1i}^c + \beta_2 x_{2i}^c + \ldots + \beta_{p-1} x_{(p-1)i}^c + e_i,\ i = 1, \ldots, n,$$

ou em notação matricial

$$\boldsymbol{y} = \boldsymbol{X}_c \boldsymbol{\beta} + \boldsymbol{e}.$$

Uma das vantagens em se utilizar as variáveis centralizadas é tornar o intercepto β_0^* **sempre** interpretável.

Uso de variáveis centralizadas

Ademais,

$$\boldsymbol{X}_c^\top \boldsymbol{X}_c = \begin{pmatrix} n & 0 & 0 & \dots & 0 \\ 0 & S_1^2 & S_{12} & \dots & S_{1(p-1)} \\ 0 & S_{12} & S_2^2 & \dots & S_{2(p-1)} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & S_{1(p-1)} & S_{2(p-1)} & \dots & S_{(p-1)}^2 \end{pmatrix}$$

e $\boldsymbol{X}_c^\top \boldsymbol{y} = (n\bar{y}_n, S_{1y}, \dots, S_{(p-1)y})^\top$, em que $S_i^2 = \sum_{j=1}^n (x_{ij} - \bar{x}_i)^2$, $S_{ik} = \sum_{j=1}^n (x_{ij} - \bar{x}_i)(x_{kj} - \bar{x}_k)$ e $S_{iy} = \sum_{j=1}^n (x_{ij} - \bar{x}_i) y_j$, $i \neq k = 1, \dots, n$.

Temos que

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}_c^\top \boldsymbol{X}_c)^{-1} \boldsymbol{X}_c^\top \boldsymbol{y},$$

de forma que decompondo de forma apropriada, obtemos diretamente que $\widehat{\beta}_0^* = \bar{y}_n$ assim como no MRLS.

Exercício (entregar próxima aula)

Exercício 11: Considere o MRLM com todas as variáveis centralizadas. Mostre que as somas de quadrados podem ser reescritas como

$$\begin{aligned}\text{SQRes} &= \sum_{j=1}^{n} y_j^2 - \sum_{i=1}^{p-1} \widehat{\beta}_i \sum_{j=1}^{n} (x_{ij} - \bar{x}_j) y_j \\ \text{SQReg} &= \sum_{i=1}^{p-1} \widehat{\beta}_i \sum_{j=1}^{n} (x_{ij} - \bar{x}_j) y_j .\end{aligned}$$

Exercício 12: Considere o exemplo 4.8 de Hoffman (2016, pag. 135). Obtenha o EMQ de $\boldsymbol{\beta}$ e a tabela de ANOVA associada ao modelo de forma analítica, i.e., sem uso de software estatístico algum.