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**, novembro 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

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

que pode ser reescrito como

$$\boldsymbol{y} = \boldsymbol{X}\beta + \boldsymbol{e}, \quad (2)$$

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

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

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

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

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

  de dimensão $n \times p$ $(n > p)$ conhecida e assumida de posto completo.
- $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_{p-1})^\top$: vetor $(p \times 1)$ de parâmetros de regressão.
- $\boldsymbol{e} = (e_1, \dots, e_n)^\top$: 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

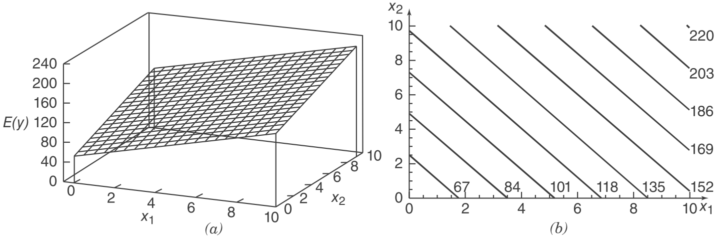
- Este modelo descreve um hiperplano $(p-1)$-dimensional no espaço gerado pelas colunas de $\boldsymbol{X}$ ($\mathbb{C}(\boldsymbol{X})$), i.e., no espaço gerado pelas variáveis explicatvas $x_1, \ldots, x_{p-1}$.
- Para exemplificar, considere o caso particular com $p = 3$ e a seguinte função de regressão

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

- 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.

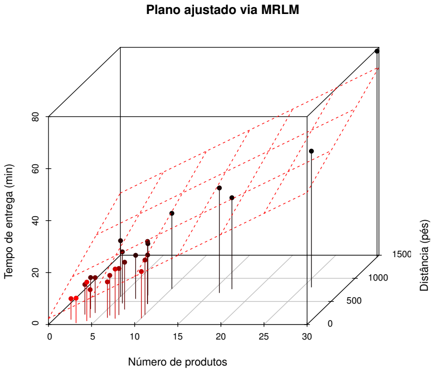
Ilustração

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

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

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

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

implicando que

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

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

$$\boldsymbol{\lambda}^\top\boldsymbol{X} = \boldsymbol{c}^\top. \tag{3}$$

Por outro lado,

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

Propriedades

Lembrando que

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

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

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

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.

<u>**Dica**</u>: 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)

<u>Teorema</u>: 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):

<u>Teorema</u>: 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) $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}$ 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

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

Logo,

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

Propriedades

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

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

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

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

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

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	$p-1$	$\boldsymbol{y}^\top(\boldsymbol{H} - \boldsymbol{J}_n/n)\boldsymbol{y}$	$\mathrm{SQReg}/(p-1) = \mathrm{QMReg}$	$\frac{\mathrm{QMReg}}{\mathrm{QMRes}}$
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

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

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

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

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

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

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

- 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)

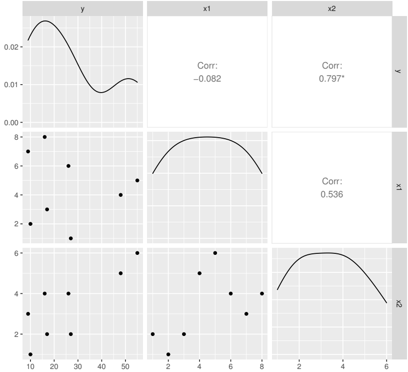
$$y_i = 8 - 5x_{1i} + 12x_{2i}. \tag{7}$$

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.

Testes de hipóteses

No contexto de MRLM podemos realizar diversos testes de hipóteses, em especial se tem interesse em responder perguntas do tipo:

- O modelo como um todo é adequado?
- Dentre todas as possíveis variáveis explicativas, quais devem ser incluídas no modelo?
- Para responder a primeira pergunta, basta utilizarmos o teste de ANOVA global, todavia a resposta para a segunda pergunta não é tão imediata.

Testes de hipóteses

No contexto de MRLM podemos realizar diversos testes de hipóteses, em especial se tem interesse em responder perguntas do tipo:

- O modelo como um todo é adequado?
- Dentre todas as possíveis variáveis explicativas, quais devem ser incluídas no modelo?
- Para responder a primeira pergunta, basta utilizarmos o teste de ANOVA global, todavia a resposta para a segunda pergunta não é tão imediata.

Testes de hipóteses

No contexto de MRLM podemos realizar diversos testes de hipóteses, em especial se tem interesse em responder perguntas do tipo:

- O modelo como um todo é adequado?
- Dentre todas as possíveis variáveis explicativas, quais devem ser incluídas no modelo?
- Para responder a primeira pergunta, basta utilizarmos o teste de ANOVA global, todavia a resposta para a segunda pergunta não é tão imediata.

Testes de hipóteses

No contexto de MRLM podemos realizar diversos testes de hipóteses, em especial se tem interesse em responder perguntas do tipo:

- O modelo como um todo é adequado?
- Dentre todas as possíveis variáveis explicativas, quais devem ser incluídas no modelo?
- Para responder a primeira pergunta, basta utilizarmos o teste de ANOVA global, todavia a resposta para a segunda pergunta não é tão imediata.

Teste para um único parâmetro

Para testar

$$\mathcal{H}_0 : \beta_i = b \text{ (especificado) vs. } \mathcal{H}_1 : \beta_i \neq (>, <) b$$

para $i = 0, 1, \ldots, (p-1)$, usa-se a estatística de teste

$$t_{\text{teste}} = \frac{\widehat{\beta}_i - b}{\widehat{S}(\widehat{\beta}_i)},$$

em que $\widehat{S}^2(\widehat{\beta}_i)$ é o $(i+1)$-ésimo elemento da diagonal principal da matriz $\widehat{\sigma}^2(\boldsymbol{X}^\top \boldsymbol{X})^{-1}$, com $\widehat{\sigma}^2 = \text{QMRes} = \dfrac{\boldsymbol{y}^\top(\boldsymbol{I}_n - \boldsymbol{H})\boldsymbol{y}}{n-p}$. Sob $\mathcal{H}_0$, tem-se que $t_{\text{teste}} \sim t_{(n-p)}$.

IC para um único parâmetro e para uma combinação linear

Para $i = 0, 1, \ldots, p-1$, tem-se que o melhor IC de nível α para β_i é dado por

$$\mathrm{IC}_{1-\alpha}(\beta_i) = \left[\widehat{\beta}_i \pm t_{(1-\alpha/2, n-p)} \widehat{S}(\widehat{\beta}_i)\right].$$

Por outro, tem-se que o melhor IC de nível α para $\boldsymbol{c}^\top \boldsymbol{\beta}$, $\boldsymbol{c} \in \mathbb{R}^p$, é dado por

$$\mathrm{IC}_{1-\alpha}(\boldsymbol{c}^\top \boldsymbol{\beta}) = \left[\boldsymbol{c}^\top \widehat{\boldsymbol{\beta}} \pm t_{(1-\alpha/2, n-p)} \sqrt{\mathrm{QMRes}\, \boldsymbol{c}^\top (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{c}}\right].$$

Os ICs de confiança para um único parâmetro podem ser obtidos como casos particulares do IC de uma combinação linear, especificando de forma conveniente o vetor $\boldsymbol{c}$. 😉

Exercícios (entregar próxima aula)

Exercício 13: Desenvolva um teste de hipóteses para $\mathcal{H}_0 : \boldsymbol{c}^\top \boldsymbol{\beta} = a$ (a especificado) vs. $\mathcal{H}_1 : \boldsymbol{c}^\top \boldsymbol{\beta} \neq a$ em um MRLM sob a suposição de normalidade.

Exercício 14: Obter um intervalo de predição de nível $(1-\alpha)$ para y quando o vetor de variáveis explicativas for igual a $\boldsymbol{x}_i = (1, x_{1i}, x_{2i}, \ldots, x_{(p-1)i})^\top$ em um MRLM sob a suposição de normalidade.

Teste para um subconjunto de coeficientes de regressão

- Considere um MRLM com p coeficientes de regressão, i.e., com a seguinte forma funcional

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

em que $\boldsymbol{y}$ $(n \times 1)$, $\boldsymbol{X}$ $(n \times p)$, $\boldsymbol{\beta}$ $(p \times 1)$ e $\boldsymbol{e}$ $(n \times 1)$ já definidos anteriormente.

- O interesse é avaliar se algum subconjunto $r < p$ de regressores contribuem de forma significativa com o modelo.

- Inicialmente considere a seguinte partição do vetor $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top)^\top$, com $\boldsymbol{\beta}_1$ um vetor $(p - r) \times 1$ e $\boldsymbol{\beta}_2$ um vetor $r \times 1$. O interesse é testar

$$\mathcal{H}_0 : \boldsymbol{\beta}_2 = \boldsymbol{0} \text{ vs. } \mathcal{H}_1 : \boldsymbol{\beta}_2 \neq \boldsymbol{0}$$

- O modelo original consegue ser reescrito como

$$\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{e},$$

em que $\boldsymbol{X}_1$ é uma matriz $(n \times (p - r))$ de especificação, representando as colunas de $\boldsymbol{X}$ associadas aos elementos de $\boldsymbol{\beta}_1$ e $\boldsymbol{X}_2$ é uma matriz $(n \times r)$ de especificação contendo as colunas de $\boldsymbol{X}$ associadas com $\boldsymbol{\beta}_2$. Este modelo é denominado de modelo completo.

Teste para um subconjunto de coeficientes de regressão

- Considere um MRLM com p coeficientes de regressão, i.e., com a seguinte forma funcional

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

em que $\boldsymbol{y}$ $(n \times 1)$, $\boldsymbol{X}$ $(n \times p)$, $\boldsymbol{\beta}$ $(p \times 1)$ e $\boldsymbol{e}$ $(n \times 1)$ já definidos anteriormente.

- O interesse é avaliar se algum subconjunto $r < p$ de regressores contribuem de forma significativa com o modelo.

- Inicialmente considere a seguinte partição do vetor $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top)^\top$, com $\boldsymbol{\beta}_1$ um vetor $(p-r) \times 1$ e $\boldsymbol{\beta}_2$ um vetor $r \times 1$. O interesse é testar

$$\mathcal{H}_0 : \boldsymbol{\beta}_2 = \mathbf{0} \text{ vs. } \mathcal{H}_1 : \boldsymbol{\beta}_2 \neq \mathbf{0}$$

- O modelo original consegue ser reescrito como

$$\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{e},$$

em que $\boldsymbol{X}_1$ é uma matriz $(n \times (p-r))$ de especificação, representando as colunas de $\boldsymbol{X}$ associadas aos elementos de $\boldsymbol{\beta}_1$ e $\boldsymbol{X}_2$ é uma matriz $(n \times r)$ de especificação contendo as colunas de $\boldsymbol{X}$ associadas com $\boldsymbol{\beta}_2$. Este modelo é denominado de modelo completo.

Teste para um subconjunto de coeficientes de regressão

- Considere um MRLM com p coeficientes de regressão, i.e., com a seguinte forma funcional

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

em que $\boldsymbol{y}$ $(n \times 1)$, $\boldsymbol{X}$ $(n \times p)$, $\boldsymbol{\beta}$ $(p \times 1)$ e $\boldsymbol{e}$ $(n \times 1)$ já definidos anteriormente.

- O interesse é avaliar se algum subconjunto $r < p$ de regressores contribuem de forma significativa com o modelo.

- Inicialmente considere a seguinte partição do vetor $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top)^\top$, com $\boldsymbol{\beta}_1$ um vetor $(p - r) \times 1$ e $\boldsymbol{\beta}_2$ um vetor $r \times 1$. O interesse é testar

$$\mathcal{H}_0 : \boldsymbol{\beta}_2 = \mathbf{0} \text{ vs. } \mathcal{H}_1 : \boldsymbol{\beta}_2 \neq \mathbf{0}$$

- O modelo original consegue ser reescrito como

$$\boldsymbol{y} = \boldsymbol{X}_1\beta_1 + \boldsymbol{X}_2\beta_2 + \boldsymbol{e},$$

em que $\boldsymbol{X}_1$ é uma matriz $(n \times (p - r))$ de especificação, representando as colunas de $\boldsymbol{X}$ associadas aos elementos de β_1 e $\boldsymbol{X}_2$ é uma matriz $(n \times r)$ de especificação contendo as colunas de $\boldsymbol{X}$ associadas com β_2. Este modelo é denominado de modelo completo.

Teste para um subconjunto de coeficientes de regressão

- Considere um MRLM com p coeficientes de regressão, i.e., com a seguinte forma funcional

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

em que $\boldsymbol{y}$ $(n \times 1)$, $\boldsymbol{X}$ $(n \times p)$, $\boldsymbol{\beta}$ $(p \times 1)$ e $\boldsymbol{e}$ $(n \times 1)$ já definidos anteriormente.

- O interesse é avaliar se algum subconjunto $r < p$ de regressores contribuem de forma significativa com o modelo.
- Inicialmente considere a seguinte partição do vetor $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top)^\top$, com $\boldsymbol{\beta}_1$ um vetor $(p - r) \times 1$ e $\boldsymbol{\beta}_2$ um vetor $r \times 1$. O interesse é testar

$$\mathcal{H}_0 : \boldsymbol{\beta}_2 = \mathbf{0} \text{ vs. } \mathcal{H}_1 : \boldsymbol{\beta}_2 \neq \mathbf{0}$$

- O modelo original consegue ser reescrito como

$$\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{e},$$

em que $\boldsymbol{X}_1$ é uma matriz $(n \times (p - r))$ de especificação, representando as colunas de $\boldsymbol{X}$ associadas aos elementos de $\boldsymbol{\beta}_1$ e $\boldsymbol{X}_2$ é uma matriz $(n \times r)$ de especificação contendo as colunas de $\boldsymbol{X}$ associadas com $\boldsymbol{\beta}_2$. Este modelo é denominado de modelo completo.

Teste para um subconjunto de coeficientes de regressão

- Para o modelo completo, temos que

$$\begin{aligned} \widehat{\boldsymbol{\beta}} &= (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{y} \\ \mathrm{SQReg} &= \mathrm{SQReg}(\boldsymbol{\beta}) = \widehat{\boldsymbol{\beta}}\boldsymbol{X}^\top \boldsymbol{y} - n\bar{y}_n^2 \text{ com } (p-1) \text{ gl} \\ \widehat{\sigma}^2 &= \mathrm{QMRes} = \frac{\boldsymbol{y}^\top(\boldsymbol{I}_n - \boldsymbol{H})\boldsymbol{y}}{n-p}. \end{aligned}$$

- Para determinar a contribuição de β_2 para o modelo, a ideia é ajustar o modelo reduzido (i.e., o modelo sob a hipótese nula $\mathcal{H}_0 : \boldsymbol{\beta}_2 = \mathbf{0}$)

$$\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{e}^*.$$

- Com base no modelo reduzido temos que o EMQ de $\boldsymbol{\beta}_1$ e a SQReg são dados, respectivamente, por

$$\begin{aligned} \widehat{\boldsymbol{\beta}}_1 &= (\boldsymbol{X}_1^\top \boldsymbol{X}_1)^{-1}\boldsymbol{X}_1^\top \boldsymbol{y} \\ \mathrm{SQReg}(\boldsymbol{\beta}_1) &= \widehat{\boldsymbol{\beta}}_1\boldsymbol{X}_1^\top \boldsymbol{y} - n\bar{y}_n^2 \text{ com } (p-r-1) \text{ gl}. \end{aligned}$$

Teste para um subconjunto de coeficientes de regressão

- Para o modelo completo, temos que

$$\begin{aligned}
\widehat{\boldsymbol{\beta}} &= (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{y} \\
\text{SQReg} &= \text{SQReg}(\boldsymbol{\beta}) = \widehat{\boldsymbol{\beta}}\boldsymbol{X}^\top \boldsymbol{y} - n\bar{y}_n^2 \text{ com } (p-1) \text{ gl} \\
\widehat{\sigma}^2 &= \text{QMRes} = \frac{\boldsymbol{y}^\top(\boldsymbol{I}_n - \boldsymbol{H})\boldsymbol{y}}{n-p}.
\end{aligned}$$

- Para determinar a contribuição de $\boldsymbol{\beta}_2$ para o modelo, a ideia é ajustar o modelo reduzido (i.e., o modelo sob a hipótese nula $\mathcal{H}_0 : \boldsymbol{\beta}_2 = \mathbf{0}$)

$$\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{e}^*.$$

- Com base no modelo reduzido temos que o EMQ de $\boldsymbol{\beta}_1$ e a SQReg são dados, respectivamente, por

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}_1 &= (\boldsymbol{X}_1^\top \boldsymbol{X}_1)^{-1}\boldsymbol{X}_1^\top \boldsymbol{y} \\
\text{SQReg}(\boldsymbol{\beta}_1) &= \widehat{\boldsymbol{\beta}}_1\boldsymbol{X}_1^\top \boldsymbol{y} - n\bar{y}_n^2 \text{ com } (p-r-1) \text{ gl.}
\end{aligned}$$

Teste para um subconjunto de coeficientes de regressão

- Para o modelo completo, temos que

$$\begin{aligned}\widehat{\boldsymbol{\beta}} &= (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{y}\\ \mathrm{SQReg} &= \mathrm{SQReg}(\boldsymbol{\beta}) = \widehat{\boldsymbol{\beta}}\boldsymbol{X}^\top\boldsymbol{y} - n\bar{y}_n^2 \text{ com } (p-1) \text{ gl}\\ \widehat{\sigma}^2 &= \mathrm{QMRes} = \frac{\boldsymbol{y}^\top(\boldsymbol{I}_n - \boldsymbol{H})\boldsymbol{y}}{n-p}.\end{aligned}$$

- Para determinar a contribuição de $\boldsymbol{\beta}_2$ para o modelo, a ideia é ajustar o modelo reduzido (i.e., o modelo sob a hipótese nula $\mathcal{H}_0 : \boldsymbol{\beta}_2 = \mathbf{0}$)

$$\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{e}^*.$$

- Com base no modelo reduzido temos que o EMQ de $\boldsymbol{\beta}_1$ e a SQReg são dados, respectivamente, por

$$\begin{aligned}\widehat{\boldsymbol{\beta}}_1 &= (\boldsymbol{X}_1^\top\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1^\top\boldsymbol{y}\\ \mathrm{SQReg}(\boldsymbol{\beta}_1) &= \widehat{\boldsymbol{\beta}}_1\boldsymbol{X}_1^\top\boldsymbol{y} - n\bar{y}_n^2 \text{ com } (p-r-1) \text{ gl}.\end{aligned}$$

Teste para um subconjunto de coeficientes de regressão

- A soma de quadrados de regressão devido a $\boldsymbol{\beta}_2$, dado que $\boldsymbol{\beta}_1$ já está no modelo, é definida por

$$\mathrm{SQReg}(\boldsymbol{\beta}_2/\boldsymbol{\beta}_1) := \mathrm{SQReg}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) - \mathrm{SQReg}(\boldsymbol{\beta}_1)$$

que possui $(p-1)-(p-r-1) = r$ graus de liberdade.

- Esta soma de quadrados é chamada de soma de quadrados devido a $\boldsymbol{\beta}_2$ pois mede o acréscimo na soma de quadrados de regressão devido a adição de r regressores, digamos $x_{(p-r)+1}, x_{(p-r)+2}, \ldots, x_{(p)}$ em um modelo que já continha $(p-r)$ regressores, digamos $x_1, x_2, \ldots, x_{(p-r)}$.

- Além disso, sob suposição de normalidade, temos que (exercício)

$$\mathrm{SQReg}(\boldsymbol{\beta}_2/\boldsymbol{\beta}_1) \perp \mathrm{SQRes}$$

e que $\frac{\mathrm{SQReg}(\boldsymbol{\beta}_2/\boldsymbol{\beta}_1)}{\sigma^2} \sim \chi^2_{(r,\lambda_1)}$, em que

$$\lambda_1 = \frac{1}{2\sigma^2}\beta_2^\top \boldsymbol{X}_2^\top [\boldsymbol{I}_n - \boldsymbol{X}_1(\boldsymbol{X}_1^\top \boldsymbol{X}_1)^{-1}\boldsymbol{X}_1^\top]\boldsymbol{X}_2\beta_2.$$

Teste para um subconjunto de coeficientes de regressão

- A soma de quadrados de regressão devido a $\boldsymbol{\beta}_2$, dado que $\boldsymbol{\beta}_1$ já está no modelo, é definida por

$$\mathrm{SQReg}(\boldsymbol{\beta}_2/\boldsymbol{\beta}_1) := \mathrm{SQReg}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) - \mathrm{SQReg}(\boldsymbol{\beta}_1)$$

que possui $(p-1)-(p-r-1) = r$ graus de liberdade.

- Esta soma de quadrados é chamada de soma de quadrados devido a $\boldsymbol{\beta}_2$ pois mede o acréscimo na soma de quadrados de regressão devido a adição de r regressores, digamos $x_{(p-r)+1}, x_{(p-r)+2}, \ldots, x_{(p)}$ em um modelo que já continha $(p-r)$ regressores, digamos $x_1, x_2, \ldots, x_{(p-r)}$.

- Além disso, sob suposição de normalidade, temos que (exercício)

$$\mathrm{SQReg}(\boldsymbol{\beta}_2/\boldsymbol{\beta}_1) \perp \mathrm{SQRes}$$

e que $\frac{\mathrm{SQReg}(\boldsymbol{\beta}_2/\boldsymbol{\beta}_1)}{\sigma^2} \sim \chi^2_{(r,\lambda_1)}$, em que

$$\lambda_1 = \frac{1}{2\sigma^2}\beta_2^\top \boldsymbol{X}_2^\top [\boldsymbol{I}_n - \boldsymbol{X}_1(\boldsymbol{X}_1^\top \boldsymbol{X}_1)^{-1}\boldsymbol{X}_1^\top]\boldsymbol{X}_2\beta_2.$$

Teste para um subconjunto de coeficientes de regressão

- A soma de quadrados de regressão devido a $\boldsymbol{\beta}_2$, dado que $\boldsymbol{\beta}_1$ já está no modelo, é definida por

$$\mathrm{SQReg}(\boldsymbol{\beta}_2/\boldsymbol{\beta}_1) := \mathrm{SQReg}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) - \mathrm{SQReg}(\boldsymbol{\beta}_1)$$

que possui $(p-1)-(p-r-1) = r$ graus de liberdade.

- Esta soma de quadrados é chamada de soma de quadrados devido a $\boldsymbol{\beta}_2$ pois mede o acréscimo na soma de quadrados de regressão devido a adição de r regressores, digamos $x_{(p-r)+1}, x_{(p-r)+2}, \ldots, x_{(p)}$ em um modelo que já continha $(p-r)$ regressores, digamos $x_1, x_2, \ldots, x_{(p-r)}$.

- Além disso, sob suposição de normalidade, temos que (exercício)

$$\mathrm{SQReg}(\boldsymbol{\beta}_2/\boldsymbol{\beta}_1) \perp \mathrm{SQRes}$$

e que $\frac{\mathrm{SQReg}(\boldsymbol{\beta}_2/\boldsymbol{\beta}_1)}{\sigma^2} \sim \chi^2_{(r,\lambda_1)}$, em que

$$\lambda_1 = \frac{1}{2\sigma^2}\boldsymbol{\beta}_2^\top \boldsymbol{X}_2^\top [\boldsymbol{I}_n - \boldsymbol{X}_1(\boldsymbol{X}_1^\top \boldsymbol{X}_1)^{-1}\boldsymbol{X}_1^\top]\boldsymbol{X}_2\boldsymbol{\beta}_2.$$

Exercício (entregar próxima aula)

Exercício 15: Considere o MRLM e a suposição de normalidade válida.

i) Reescreva $\mathrm{SQReg}(\boldsymbol{\beta}_1)$ e $\mathrm{SQReg}(\boldsymbol{\beta}_2/\boldsymbol{\beta}_1)$ como formas quadráticas.

ii) Prove que $\mathrm{SQReg}(\boldsymbol{\beta}_2/\boldsymbol{\beta}_1) \perp \mathrm{SQRes}$.

iii) Prove que $\dfrac{\mathrm{SQReg}(\boldsymbol{\beta}_2/\boldsymbol{\beta}_1)}{\sigma^2} \sim \chi^2_{(r,\lambda_1)}$, em que

$$\lambda_1 = \frac{1}{2\sigma^2}\boldsymbol{\beta}_2^\top \boldsymbol{X}_2^\top [\boldsymbol{I}_n - \boldsymbol{X}_1(\boldsymbol{X}_1^\top \boldsymbol{X}_1)^{-1}\boldsymbol{X}_1^\top]\boldsymbol{X}_2\boldsymbol{\beta}_2.$$

Teste para um subconjunto de coeficientes de regressão

- Portanto, podemos utilizar para testar

$$\mathcal{H}_0 : \boldsymbol{\beta}_2 = \mathbf{0} \text{ vs. } \mathcal{H}_1 : \boldsymbol{\beta}_2 \neq \mathbf{0}$$

a estatística

$$F_0^* = \frac{\text{SQReg}(\boldsymbol{\beta}_2/\boldsymbol{\beta}_1)/r}{\text{QMRes}},$$

de forma que sob $\mathcal{H}_0$, $F_0^* \sim \mathcal{F}(r, n-p)$.

- Para obter a contribuição de um subconjunto de variáveis explicativas no software `R`, basta utilizar

```
ANOVA(ajuste,ajuste1)
```

sendo `ajuste` e `ajuste1` os objetos referentes, respectivamente, aos ajustes dos modelos completo e reduzido.

Teste para um subconjunto de coeficientes de regressão

- Portanto, podemos utilizar para testar

$$\mathcal{H}_0 : \boldsymbol{\beta}_2 = \mathbf{0} \text{ vs. } \mathcal{H}_1 : \boldsymbol{\beta}_2 \neq \mathbf{0}$$

a estatística

$$F_0^* = \frac{\mathrm{SQReg}(\boldsymbol{\beta}_2/\boldsymbol{\beta}_1)/r}{\mathrm{QMRes}},$$

de forma que sob $\mathcal{H}_0$, $F_0^* \sim \mathcal{F}(r, n-p)$.

- Para obter a contribuição de um subconjunto de variáveis explicativas no software `R`, basta utilizar

```
ANOVA(ajuste,ajuste1)
```

sendo `ajuste` e `ajuste1` os objetos referentes, respectivamente, aos ajustes dos modelos completo e reduzido.

ANOVA para um subconjunto de coeficientes de regressão

Abaixo, segue a tabela de ANOVA associada ao teste em questão

Causas de Variação	GL	SQ	QM	F
$\text{Reg}(\boldsymbol{y}/\boldsymbol{x}_1)$	$p-1-r$	$\mathrm{SQReg}(\boldsymbol{\beta}_1)$	-	-
Cont. de $\boldsymbol{x}_2$	r	$\mathrm{SQReg}(\boldsymbol{\beta}_2/\boldsymbol{\beta}_1)$	$\mathrm{QMC}(\boldsymbol{x}_2)$	$\dfrac{\mathrm{QMC}(\boldsymbol{x}_2)}{\mathrm{QMRes}}$
Regressão	$p-1$	$\boldsymbol{y}^\top(\boldsymbol{H}-\boldsymbol{J}_n/n)\boldsymbol{y}$	$\mathrm{SQReg}/(p-1)=\mathrm{QMReg}$	-
Resíduo	$n-p$	$\boldsymbol{y}^\top(\boldsymbol{I}_n-\boldsymbol{H})\boldsymbol{y}$	$\mathrm{SQRes}/(n-p)=\mathrm{QMRes}$	-
Total	$n-1$	$\boldsymbol{y}^\top(\boldsymbol{I}_n-\boldsymbol{J}_n/n)\boldsymbol{y}$	$\mathrm{SQT}/(n-1)$	-

De forma que rejeitamos $\mathcal{H}_0 : \boldsymbol{\beta}_2 = \mathbf{0}$ ao nível α se $F_0^* = \dfrac{\mathrm{QMC}(\boldsymbol{x}_2)}{\mathrm{QMRes}} > \mathcal{F}_{1-\alpha}(r, n-p)$.

Hipótese Linear Geral

- Todos os testes mostrados anteriormente podem ser expressos como uma situação particular de uma classe maior de testes, denominada hipótese linear geral, dada por

$$\mathcal{H}_0 : \boldsymbol{C}\boldsymbol{\beta} = \boldsymbol{m} \text{ vs. } \mathcal{H}_1 : \boldsymbol{C}\boldsymbol{\beta} \neq \boldsymbol{m} \tag{8}$$

com $\boldsymbol{C}$ denotando uma matriz e $\boldsymbol{m}$ um vetor, ambos conhecidos e que possuem papéis de destaque para especificar as hipóteses de interesse.

- Para testar $\mathcal{H}_0$ acima, utiliza-se a seguinte estatística de teste

$$Q_0 = \frac{(\boldsymbol{C}\widehat{\boldsymbol{\beta}} - \boldsymbol{m})^\top \left[\boldsymbol{C}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{C}^\top\right]^{-1}(\boldsymbol{C}\widehat{\boldsymbol{\beta}} - \boldsymbol{m})}{k\text{QMRes}},$$

em que $k = \text{posto}(\boldsymbol{C})$.

- Sob a hipótese de normalidade e sob $\mathcal{H}_0$, tem-se que

$$Q_0 \sim \mathcal{F}(k, n-p) \xrightarrow{\mathcal{D}} \chi^2_k, \text{ quando } n \to \infty.$$

- No software R, pode-se usar a função `glht` implementada no pacote `gmodels`.

Hipótese Linear Geral

- Todos os testes mostrados anteriormente podem ser expressos como uma situação particular de uma classe maior de testes, denominada hipótese linear geral, dada por

$$\mathcal{H}_0 : \boldsymbol{C\beta} = \boldsymbol{m} \text{ vs. } \mathcal{H}_1 : \boldsymbol{C\beta} \neq \boldsymbol{m} \tag{8}$$

com $\boldsymbol{C}$ denotando uma matriz e $\boldsymbol{m}$ um vetor, ambos conhecidos e que possuem papéis de destaque para especificar as hipóteses de interesse.

- Para testar $\mathcal{H}_0$ acima, utiliza-se a seguinte estatística de teste

$$Q_0 = \frac{(\boldsymbol{C}\widehat{\boldsymbol{\beta}} - \boldsymbol{m})^\top \left[\boldsymbol{C}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{C}^\top\right]^{-1}(\boldsymbol{C}\widehat{\boldsymbol{\beta}} - \boldsymbol{m})}{k\text{QMRes}},$$

em que $k = \text{posto}(\boldsymbol{C})$.

- Sob a hipótese de normalidade e sob $\mathcal{H}_0$, tem-se que

$$Q_0 \sim \mathcal{F}(k, n-p) \xrightarrow{\mathcal{D}} \chi^2_k, \text{ quando } n \to \infty.$$

- No software R, pode-se usar a função `glht` implementada no pacote `gmodels`.

Hipótese Linear Geral

- Todos os testes mostrados anteriormente podem ser expressos como uma situação particular de uma classe maior de testes, denominada hipótese linear geral, dada por

$$\mathcal{H}_0 : \boldsymbol{C}\boldsymbol{\beta} = \boldsymbol{m} \text{ vs. } \mathcal{H}_1 : \boldsymbol{C}\boldsymbol{\beta} \neq \boldsymbol{m} \tag{8}$$

  com $\boldsymbol{C}$ denotando uma matriz e $\boldsymbol{m}$ um vetor, ambos conhecidos e que possuem papéis de destaque para especificar as hipóteses de interesse.

- Para testar $\mathcal{H}_0$ acima, utiliza-se a seguinte estatística de teste

$$Q_0 = \frac{(\boldsymbol{C}\widehat{\boldsymbol{\beta}} - \boldsymbol{m})^\top \left[\boldsymbol{C}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{C}^\top\right]^{-1}(\boldsymbol{C}\widehat{\boldsymbol{\beta}} - \boldsymbol{m})}{k\text{QMRes}},$$

  em que $k = \text{posto}(\boldsymbol{C})$.

- Sob a hipótese de normalidade e sob $\mathcal{H}_0$, tem-se que

$$Q_0 \sim \mathcal{F}(k, n-p) \xrightarrow{\mathcal{D}} \chi^2_k, \text{ quando } n \to \infty.$$

- No software R, pode-se usar a função `glht` implementada no pacote `gmodels`.

Hipótese Linear Geral

- Todos os testes mostrados anteriormente podem ser expressos como uma situação particular de uma classe maior de testes, denominada hipótese linear geral, dada por

$$\mathcal{H}_0 : \boldsymbol{C}\boldsymbol{\beta} = \boldsymbol{m} \text{ vs. } \mathcal{H}_1 : \boldsymbol{C}\boldsymbol{\beta} \neq \boldsymbol{m} \quad (8)$$

com $\boldsymbol{C}$ denotando uma matriz e $\boldsymbol{m}$ um vetor, ambos conhecidos e que possuem papéis de destaque para especificar as hipóteses de interesse.

- Para testar $\mathcal{H}_0$ acima, utiliza-se a seguinte estatística de teste

$$Q_0 = \frac{(\boldsymbol{C}\widehat{\boldsymbol{\beta}} - \boldsymbol{m})^\top \left[\boldsymbol{C}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{C}^\top\right]^{-1}(\boldsymbol{C}\widehat{\boldsymbol{\beta}} - \boldsymbol{m})}{k\mathrm{QMRes}},$$

em que $k = \mathrm{posto}(\boldsymbol{C})$.

- Sob a hipótese de normalidade e sob $\mathcal{H}_0$, tem-se que

$$Q_0 \sim \mathcal{F}(k, n-p) \xrightarrow{\mathcal{D}} \chi^2_k, \text{ quando } n \to \infty.$$

- No software `R`, pode-se usar a função `glht` implementada no pacote `gmodels`.

Exercícios (entregar próxima aula)

Exercício 16: Mostre que todos os testes já apresentados anteriormente, podem ser expressos no formato da hipótese linear geral.

Exercício 17: Considere $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3, \beta_4)^\top$. Expresse as hipóteses abaixo no formato da hipótese linear geral:

i) $\mathcal{H}_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4$.

ii) $\mathcal{H}_0 : \beta_1 = \beta_2$ e $\beta_3 = \beta_4$.

iii) $\mathcal{H}_0 : \beta_1 = 2\beta_2 - 3\beta 3 + 7\beta_4$.

iv) $\mathcal{H}_0 : \beta_1 = \beta_4$, $\beta_2 = 3$ e $\beta_4 = 5$.

v) $\mathcal{H}_0 : \beta_4 - \beta_3 = \beta_2 - \beta_1$.

Exercício 18: Forneça um exemplo hipotético que a hipótese de interesse não consegue ser expressa no formato da hipótese linear geral.

Exercício 19: Faça um resumo da subseção 3.11 de Montgomery et al. (2012): "Why do regression coefficients have the wrong sign?"

Elipsóides de confiança

- Em um MRLM sob a suposição de normalidade, tem-se que uma região de confiança conjunta para $\boldsymbol{\beta}$ é dada pelo seguinte Elipsóide de Confiança de nível $(1-\alpha)$:

$$\mathrm{EC}_{1-\alpha}(\boldsymbol{\beta}) = \left[\boldsymbol{\beta} \in \mathbb{R}^p; (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})^\top (\boldsymbol{X}^\top \boldsymbol{X})^{-1} (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \leq p \mathrm{QMRes} \mathcal{F}_{1-\alpha}(p, n-p)\right].$$

- No caso de termos dois parâmetros de regressão, β_0 e β_1, então se $\widehat{\beta}_0$ e $\widehat{\beta}_1$ forem não-correlacionados

$$\mathrm{EC}_{1-\alpha}(\boldsymbol{\beta}) = \mathrm{IC}_{1-\alpha}(\beta_0) \times \mathrm{IC}_{1-\alpha}(\beta_1),$$

i.e., a região resultante é retangular.

- A inclinação da elipse é relacionada com $\mathrm{sign}(\mathrm{Cov}(\widehat{\beta}_0, \widehat{\beta}_1))$.
- No software R, pode-se usar o pacote `ellipse` para gerar os elipsóides de confiança em diversas classes de modelos de regressão.

Elipsóides de confiança

- Em um MRLM sob a suposição de normalidade, tem-se que uma região de confiança conjunta para $\boldsymbol{\beta}$ é dada pelo seguinte Elipsóide de Confiança de nível $(1-\alpha)$:

$$\mathrm{EC}_{1-\alpha}(\boldsymbol{\beta}) = \left[\boldsymbol{\beta} \in \mathbb{R}^p ; (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})^\top (\boldsymbol{X}^\top \boldsymbol{X})^{-1} (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \leq p \mathrm{QMRes} \mathcal{F}_{1-\alpha}(p, n-p)\right].$$

- No caso de termos dois parâmetros de regressão, β_0 e β_1, então se $\widehat{\beta}_0$ e $\widehat{\beta}_1$ forem não-correlacionados

$$\mathrm{EC}_{1-\alpha}(\boldsymbol{\beta}) = \mathrm{IC}_{1-\alpha}(\beta_0) \times \mathrm{IC}_{1-\alpha}(\beta_1),$$

i.e., a região resultante é retangular.

- A inclinação da elipse é relacionada com $\mathrm{sign}(\mathrm{Cov}(\widehat{\beta}_0, \widehat{\beta}_1))$.
- No software R, pode-se usar o pacote `ellipse` para gerar os elipsóides de confiança em diversas classes de modelos de regressão.

Elipsóides de confiança

- Em um MRLM sob a suposição de normalidade, tem-se que uma região de confiança conjunta para $\boldsymbol{\beta}$ é dada pelo seguinte Elipsóide de Confiança de nível $(1-\alpha)$:

$$\mathrm{EC}_{1-\alpha}(\boldsymbol{\beta}) = \left[\boldsymbol{\beta} \in \mathbb{R}^p; (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})^\top (\boldsymbol{X}^\top \boldsymbol{X})^{-1} (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \leq p \mathrm{QMRes} \mathcal{F}_{1-\alpha}(p, n-p)\right].$$

- No caso de termos dois parâmetros de regressão, β_0 e β_1, então se $\widehat{\beta}_0$ e $\widehat{\beta}_1$ forem não-correlacionados

$$\mathrm{EC}_{1-\alpha}(\boldsymbol{\beta}) = \mathrm{IC}_{1-\alpha}(\beta_0) \times \mathrm{IC}_{1-\alpha}(\beta_1),$$

  i.e., a região resultante é retangular.

- A inclinação da elipse é relacionada com $\mathrm{sign}(\mathrm{Cov}(\widehat{\beta}_0, \widehat{\beta}_1))$.

- No software R, pode-se usar o pacote `ellipse` para gerar os elipsóides de confiança em diversas classes de modelos de regressão.

Elipsóides de confiança

- Em um MRLM sob a suposição de normalidade, tem-se que uma região de confiança conjunta para $\boldsymbol{\beta}$ é dada pelo seguinte Elipsóide de Confiança de nível $(1-\alpha)$:

$$\mathrm{EC}_{1-\alpha}(\boldsymbol{\beta}) = \left[\boldsymbol{\beta} \in \mathbb{R}^p; (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})^\top (\boldsymbol{X}^\top \boldsymbol{X})^{-1} (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \leq p\mathrm{QMRes}\mathcal{F}_{1-\alpha}(p, n-p)\right].$$

- No caso de termos dois parâmetros de regressão, β_0 e β_1, então se $\widehat{\beta}_0$ e $\widehat{\beta}_1$ forem não-correlacionados

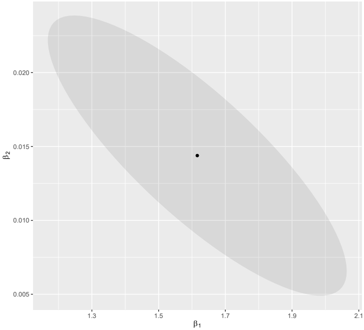
$$\mathrm{EC}_{1-\alpha}(\boldsymbol{\beta}) = \mathrm{IC}_{1-\alpha}(\beta_0) \times \mathrm{IC}_{1-\alpha}(\beta_1),$$

  i.e., a região resultante é retangular.

- A inclinação da elipse é relacionada com $\mathrm{sign}(\mathrm{Cov}(\widehat{\beta}_0, \widehat{\beta}_1))$.
- No software `R`, pode-se usar o pacote `ellipse` para gerar os elipsóides de confiança em diversas classes de modelos de regressão.

Ilustração

Figura: Elipsóide de confiança com 95% de confiança para os parâmetros de regressão do modelo para os dados do Exemplo 3.1- The Delivery Time Data (Montgomery et al., 2012, pag. 74)



Coeficiente de correlação parcial

- Considere um MRLM com duas variáveis explicativas centralizadas

$$y_i = \beta_0^* + \beta_1(x_{1i} - \bar{x}_1) + \beta_2(x_{2i} - \bar{x}_2) + e_i, i = 1, \ldots, n. \quad (9)$$

- Denotando ν_j o j-ésimo resíduo ordinário da regressão linear de x_{1i} contra x_{2i}, e por z_j o j-ésimo resíduo ordinário linear de y_i contra x_{2i}.
- O coeficiente de correlação parcial de y e x_1, denotado por $r_{y1.2}$, é por definição, o coficiente de correlação entre ν e z, i.e.,

$$r_{y1.2} = \frac{S_{z\nu}}{\sqrt{S_{zz}S_{\nu\nu}}}.$$

- Após alguma álgebra (exercício), pode-se mostrar que

$$r_{y1.2} = \frac{r_{y1} - r_{12}r_{y2}}{\sqrt{(1 - r_{12}^2)(1 - r_{y2}^2)}}.$$

Coeficiente de correlação parcial

- Considere um MRLM com duas variáveis explicativas centralizadas

$$y_i = \beta_0^* + \beta_1(x_{1i} - \bar{x}_1) + \beta_2(x_{2i} - \bar{x}_2) + e_i, i = 1, \ldots, n. \quad (9)$$

- Denotando ν_j o j-ésimo resíduo ordinário da regressão linear de x_{1i} contra x_{2i}, e por z_j o j-ésimo resíduo ordinário linear de y_i contra x_{2i}.

- O coeficiente de correlação parcial de y e x_1, denotado por $r_{y1.2}$, é por definição, o coficiente de correlação entre ν e z, i.e.,

$$r_{y1.2} = \frac{S_{z\nu}}{\sqrt{S_{zz}S_{\nu\nu}}}.$$

- Após alguma álgebra (exercício), pode-se mostrar que

$$r_{y1.2} = \frac{r_{y1} - r_{12}r_{y2}}{\sqrt{(1 - r_{12}^2)(1 - r_{y2}^2)}}.$$

Coeficiente de correlação parcial

- Considere um MRLM com duas variáveis explicativas centralizadas

$$y_i = \beta_0^* + \beta_1(x_{1i} - \bar{x}_1) + \beta_2(x_{2i} - \bar{x}_2) + e_i, i = 1, \ldots, n. \tag{9}$$

- Denotando ν_j o j-ésimo resíduo ordinário da regressão linear de x_{1i} contra x_{2i}, e por z_j o j-ésimo resíduo ordinário linear de y_i contra x_{2i}.

- O coeficiente de correlação parcial de y e x_1, denotado por $r_{y1.2}$, é por definição, o coficiente de correlação entre ν e z, i.e.,

$$r_{y1.2} = \frac{S_{z\nu}}{\sqrt{S_{zz}S_{\nu\nu}}}.$$

- Após alguma álgebra (exercício), pode-se mostrar que

$$r_{y1.2} = \frac{r_{y1} - r_{12}r_{y2}}{\sqrt{(1 - r_{12}^2)(1 - r_{y2}^2)}}.$$

Coeficiente de correlação parcial

- Considere um MRLM com duas variáveis explicativas centralizadas

$$y_i = \beta_0^* + \beta_1(x_{1i} - \bar{x}_1) + \beta_2(x_{2i} - \bar{x}_2) + e_i, i = 1, \ldots, n. \tag{9}$$

- Denotando ν_j o j-ésimo resíduo ordinário da regressão linear de x_{1i} contra x_{2i}, e por z_j o j-ésimo resíduo ordinário linear de y_i contra x_{2i}.

- O coeficiente de correlação parcial de y e x_1, denotado por $r_{y1.2}$, é por definição, o coficiente de correlação entre ν e z, i.e.,

$$r_{y1.2} = \frac{S_{z\nu}}{\sqrt{S_{zz}S_{\nu\nu}}}.$$

- Após alguma álgebra (exercício), pode-se mostrar que

$$r_{y1.2} = \frac{r_{y1} - r_{12}r_{y2}}{\sqrt{(1 - r_{12}^2)(1 - r_{y2}^2)}}.$$

Coeficiente de correlação parcial

- **Pergunta**: O que mede/avalia $r_{y1.2}$?
- Qual a relação entre y e x_1, dado x_2, i.e., será que existe relação linear entre y e x_1, dado que existe relação linear entre y e x_2?
- Em outras palavras, $r_{y1.2}$ serve para avaliar a importância da inclusão da variável x_1 no modelo

$$y_i = \beta_0^* + \beta_2(x_{2i} - \bar{x}_2) + e_i, \tag{10}$$

i.e., devemos incluir a variável x_1 neste modelo de regressão?

Coeficiente de correlação parcial

- **Pergunta**: O que mede/avalia $r_{y1.2}$?
- Qual a relação entre y e x_1, dado x_2, i.e., será que existe relação linear entre y e x_1, dado que existe relação linear entre y e x_2?
- Em outras palavras, $r_{y1.2}$ serve para avaliar a importância da inclusão da variável x_1 no modelo

$$y_i = \beta_0^* + \beta_2(x_{2i} - \bar{x}_2) + e_i, \tag{10}$$

i.e., devemos incluir a variável x_1 neste modelo de regressão?

Coeficiente de correlação parcial

- **Pergunta**: O que mede/avalia $r_{y1.2}$?
- Qual a relação entre y e x_1, dado x_2, i.e., será que existe relação linear entre y e x_1, dado que existe relação linear entre y e x_2?
- Em outras palavras, $r_{y1.2}$ serve para avaliar a importância da inclusão da variável x_1 no modelo

$$y_i = \beta_0^* + \beta_2(x_{2i} - \bar{x}_2) + e_i, \tag{10}$$

i.e., devemos incluir a variável x_1 neste modelo de regressão?

Coeficiente de correlação parcial

- O gráfico da variável adicionada usa o mesmo *princípio*. Tal gráfico consiste em plotar o diagrama de dispersão entre os resíduos z e ν e através de sua análise consegue-se verificar a importância da adição de x_1 no modelo (10).
- É possível mostrar que (exercício) o EMQ de β_1 no modelo (9) é dado por

$$\widehat{\beta}_1 = r_{y1.2}\sqrt{\frac{\mathrm{SQReg}(y/x_2)}{\mathrm{SQReg}(x_1/x_2)}},$$

ou seja, o coeficiente de correlação parcial $r_{y1.2}$ tem o mesmo sinal de $\widehat{\beta}_1$, todavia este pode ter sinal diferente do coeficiente de correlação marginal r_{y1}.

Coeficiente de correlação parcial

- O gráfico da variável adicionada usa o mesmo *princípio*. Tal gráfico consiste em plotar o diagrama de dispersão entre os resíduos z e ν e através de sua análise consegue-se verificar a importância da adição de x_1 no modelo (10).

- É possível mostrar que (exercício) o EMQ de β_1 no modelo (9) é dado por

$$\widehat{\beta}_1 = r_{y1.2}\sqrt{\frac{\mathrm{SQReg}(y/x_2)}{\mathrm{SQReg}(x_1/x_2)}},$$

ou seja, o coeficiente de correlação parcial $r_{y1.2}$ tem o mesmo sinal de $\widehat{\beta}_1$, todavia este pode ter sinal diferente do coeficiente de correlação marginal r_{y1}.

Exercícios (entregar próxima aula)

Exercício 20: Considere um MRLM com dois regressores centralizados. Mostre que

$$\begin{aligned} r_{y1.2} &= \frac{r_{y1} - r_{12} r_{y2}}{\sqrt{(1 - r_{12}^2)(1 - r_{y2}^2)}} \text{ e} \\ \widehat{\beta}_1 &= r_{y1.2}\sqrt{\frac{\text{SQReg}(y/x_2)}{\text{SQReg}(x_1/x_2)}}, \end{aligned}$$

Exercício 21: Faça um resumo sobre a técnica associada ao gráfico da variável adicionada e mostre um exemplo prático (com somente dois regressores) em que ele é utilizado para selecionar um MRLM.

Coeficientes de regressão padronizados

- É comum se ter dificuldade em comparar diretamente os coeficientes de regressão estimados, tendo em vista que a magnitude de $\widehat{\beta}_j$ reflete a unidade de medida da variável regressora x_j.
- Para exemplificar, considere o exemplo (Montgomery et al., 2012)

$$\widehat{y}_i = 5 + x_{1i} + 1000x_{2i},$$

em que y é medido em litros (L), x_1 em mililitros (ml) e x_2 em litros (L).

- Embora $\widehat{\beta}_2 >> \widehat{\beta}_1$, o efeito de ambos regressores (x_1 e x_2) em $\widehat{y}$ é o mesmo. Como se chega a esta conclusão? 😇
- Por esta razão, é comum se considerar regressores escalonados de forma a produzir coeficientes de regressão adimensionais, que são comumente denominados de coeficientes de regressão padronizados.

Coeficientes de regressão padronizados

- É comum se ter dificuldade em comparar diretamente os coeficientes de regressão estimados, tendo em vista que a magnitude de $\widehat{\beta}_j$ reflete a unidade de medida da variável regressora x_j.

- Para exemplificar, considere o exemplo (Montgomery et al., 2012)

$$\widehat{y}_i = 5 + x_{1i} + 1000x_{2i},$$

em que y é medido em litros (L), x_1 em mililitros (ml) e x_2 em litros (L).

- Embora $\widehat{\beta}_2 >> \widehat{\beta}_1$, o efeito de ambos regressores (x_1 e x_2) em $\widehat{y}$ é o mesmo. Como se chega a esta conclusão? 😇

- Por esta razão, é comum se considerar regressores escalonados de forma a produzir coeficientes de regressão adimensionais, que são comumente denominados de coeficientes de regressão padronizados.

Coeficientes de regressão padronizados

- É comum se ter dificuldade em comparar diretamente os coeficientes de regressão estimados, tendo em vista que a magnitude de $\widehat{\beta}_j$ reflete a unidade de medida da variável regressora x_j.
- Para exemplificar, considere o exemplo (Montgomery et al., 2012)

$$\widehat{y}_i = 5 + x_{1i} + 1000x_{2i},$$

em que y é medido em litros (L), x_1 em mililitros (ml) e x_2 em litros (L).

- Embora $\widehat{\beta}_2 >> \widehat{\beta}_1$, o efeito de ambos regressores (x_1 e x_2) em $\widehat{y}$ é o mesmo. Como se chega a esta conclusão?
- Por esta razão, é comum se considerar regressores escalonados de forma a produzir coeficientes de regressão adimensionais, que são comumente denominados de coeficientes de regressão padronizados.

Coeficientes de regressão padronizados

- É comum se ter dificuldade em comparar diretamente os coeficientes de regressão estimados, tendo em vista que a magnitude de $\widehat{\beta}_j$ reflete a unidade de medida da variável regressora x_j.
- Para exemplificar, considere o exemplo (Montgomery et al., 2012)

$$\widehat{y}_i = 5 + x_{1i} + 1000x_{2i},$$

em que y é medido em litros (L), x_1 em mililitros (ml) e x_2 em litros (L).
- Embora $\widehat{\beta}_2 >> \widehat{\beta}_1$, o efeito de ambos regressores (x_1 e x_2) em $\widehat{y}$ é o mesmo. Como se chega a esta conclusão?
- Por esta razão, é comum se considerar regressores escalonados de forma a produzir coeficientes de regressão adimensionais, que são comumente denominados de coeficientes de regressão padronizados.

Unit normal scaling- Escala normal unitária

- Considere

$$
\begin{aligned}
z_{ij} &= \frac{x_{ij} - \bar{x}_j}{S_j}, i = 1, 2, \ldots, n, j = 1, \ldots, p-1 \text{ e} \\
y_i^* &= \frac{y_i - \bar{y}_n}{S_y}, i = \ldots, n.
\end{aligned}
$$

- Em que

$$
\begin{aligned}
S_j^2 &= S_{jj} = \frac{\sum_{i=1}^{n}(x_{ij} - \bar{x}_j)^2}{n-1}, j = 1, \ldots, p-1 \text{ e} \\
S_y^2 &= S_{yy} = \frac{\sum_{i=1}^{n}(y_i - \bar{y}_n)^2}{n-1}.
\end{aligned}
$$

- Perceba que tanto a variável resposta como as variáveis explicativas possuem média e variância amostrais iguais a 0 e 1, respectivamente.

- Assim, considera-se o seguinte MRLM

$$
y_i^* = b_1 z_{1i} + b_2 z_{2i} + \ldots + b_{p-1} z_{(p-1)i} + e_i, i = 1, \ldots, n. \tag{11}
$$

Unit normal scaling- Escala normal unitária

- Considere

$$
\begin{aligned}
z_{ij} &= \frac{x_{ij} - \bar{x}_j}{S_j}, i = 1, 2, \ldots, n, j = 1, \ldots, p-1 \text{ e} \\
y_i^* &= \frac{y_i - \bar{y}_n}{S_y}, i = \ldots, n.
\end{aligned}
$$

- Em que

$$
\begin{aligned}
S_j^2 &= S_{jj} = \frac{\sum_{i=1}^{n}(x_{ij} - \bar{x}_j)^2}{n-1}, j = 1, \ldots, p-1 \text{ e} \\
S_y^2 &= S_{yy} = \frac{\sum_{i=1}^{n}(y_i - \bar{y}_n)^2}{n-1}.
\end{aligned}
$$

- Perceba que tanto a variável resposta como as variáveis explicativas possuem média e variância amostrais iguais a 0 e 1, respectivamente.

- Assim, considera-se o seguinte MRLM

$$
y_i^* = b_1 z_{1i} + b_2 z_{2i} + \ldots + b_{p-1} z_{(p-1)i} + e_i, i = 1, \ldots, n. \tag{11}
$$

Unit normal scaling- Escala normal unitária

- Considere

$$\begin{aligned} z_{ij} &= \frac{x_{ij} - \bar{x}_j}{S_j}, i = 1, 2, \ldots, n, j = 1, \ldots, p-1 \text{ e} \\ y_i^* &= \frac{y_i - \bar{y}_n}{S_y}, i = \ldots, n. \end{aligned}$$

- Em que

$$\begin{aligned} S_j^2 &= S_{jj} = \frac{\sum_{i=1}^{n}(x_{ij} - \bar{x}_j)^2}{n-1}, j = 1, \ldots, p-1 \text{ e} \\ S_y^2 &= S_{yy} = \frac{\sum_{i=1}^{n}(y_i - \bar{y}_n)^2}{n-1}. \end{aligned}$$

- Perceba que tanto a variável resposta como as variáveis explicativas possuem média e variância amostrais iguais a 0 e 1, respectivamente.

- Assim, considera-se o seguinte MRLM

$$y_i^* = b_1 z_{1i} + b_2 z_{2i} + \ldots + b_{p-1} z_{(p-1)i} + e_i, i = 1, \ldots, n. \quad (11)$$

Unit normal scaling- Escala normal unitária

- Considere

$$
\begin{aligned}
z_{ij} &= \frac{x_{ij} - \bar{x}_j}{S_j}, i = 1, 2, \ldots, n, j = 1, \ldots, p-1 \text{ e} \\
y_i^* &= \frac{y_i - \bar{y}_n}{S_y}, i = \ldots, n.
\end{aligned}
$$

- Em que

$$
\begin{aligned}
S_j^2 &= S_{jj} = \frac{\sum_{i=1}^n (x_{ij} - \bar{x}_j)^2}{n-1}, j = 1, \ldots, p-1 \text{ e} \\
S_y^2 &= S_{yy} = \frac{\sum_{i=1}^n (y_i - \bar{y}_n)^2}{n-1}.
\end{aligned}
$$

- Perceba que tanto a variável resposta como as variáveis explicativas possuem média e variância amostrais iguais a 0 e 1, respectivamente.
- Assim, considera-se o seguinte MRLM

$$
y_i^* = b_1 z_{1i} + b_2 z_{2i} + \ldots + b_{p-1} z_{(p-1)i} + e_i, i = 1, \ldots, n. \tag{11}
$$

Unit normal scaling- Escala normal unitária

- Dado que foi centralizado (padronizado) tanto as regressoras como a variável resposta, então o intercepto é removido do modelo, tendo em vista que $\widehat{b}_0 = \bar{y}_n^* = 0$.
- O EMQ de $\boldsymbol{b} = (b_1, \ldots, b_{p-1})^\top$ é dado por

$$\widehat{\boldsymbol{b}} = (\boldsymbol{Z}^\top \boldsymbol{Z})^{-1} \boldsymbol{Z}^\top \boldsymbol{y}^*.$$

- Os coeficientes de regressão $\widehat{\boldsymbol{b}}$ são comumente denominados por coeficientes de regressão padronizados. A relação entre os coeficientes originais estimados e os padronizados é dada por

$$\begin{aligned}\widehat{\beta}_j &= \widehat{b}_j \left(\frac{S_{yy}}{S_{jj}}\right)^{1/2}, \ j = 1, \ldots, p-1 \text{ e} \\ \widehat{\beta}_0 &= \bar{y}_n - \sum_{j=1}^{p-1} \widehat{\beta}_j \bar{x}_j.\end{aligned}$$

Unit normal scaling- Escala normal unitária

- Dado que foi centralizado (padronizado) tanto as regressoras como a variável resposta, então o intercepto é removido do modelo, tendo em vista que $\widehat{b}_0 = \bar{y}_n^* = 0$.
- O EMQ de $\boldsymbol{b} = (b_1, \ldots, b_{p-1})^\top$ é dado por

$$\widehat{\boldsymbol{b}} = (\boldsymbol{Z}^\top \boldsymbol{Z})^{-1} \boldsymbol{Z}^\top \boldsymbol{y}^*.$$

- Os coeficientes de regressão $\widehat{\boldsymbol{b}}$ são comumente denominados por coeficientes de regressão padronizados. A relação entre os coeficientes originais estimados e os padronizados é dada por

$$\begin{aligned} \widehat{\beta}_j &= \widehat{b}_j \left(\frac{S_{yy}}{S_{jj}} \right)^{1/2}, \, j = 1, \ldots, p-1 \text{ e} \\ \widehat{\beta}_0 &= \bar{y}_n - \sum_{j=1}^{p-1} \widehat{\beta}_j \bar{x}_j. \end{aligned}$$

Unit normal scaling- Escala normal unitária

- Dado que foi centralizado (padronizado) tanto as regressoras como a variável resposta, então o intercepto é removido do modelo, tendo em vista que $\widehat{b}_0 = \bar{y}_n^* = 0$.
- O EMQ de $\boldsymbol{b} = (b_1, \ldots, b_{p-1})^\top$ é dado por

$$\widehat{\boldsymbol{b}} = (\boldsymbol{Z}^\top \boldsymbol{Z})^{-1} \boldsymbol{Z}^\top \boldsymbol{y}^*.$$

- Os coeficientes de regressão $\widehat{\boldsymbol{b}}$ são comumente denominados por coeficientes de regressão padronizados. A relação entre os coeficientes originais estimados e os padronizados é dada por

$$\begin{aligned}
\widehat{\beta}_j &= \widehat{b}_j \left(\frac{S_{yy}}{S_{jj}}\right)^{1/2}, \; j = 1, \ldots, p-1 \text{ e} \\
\widehat{\beta}_0 &= \bar{y}_n - \sum_{j=1}^{p-1} \widehat{\beta}_j \bar{x}_j.
\end{aligned}$$

Unit length scaling- Escala de comprimento unitário

O outro método de padronização é obtido fazendo

$$y_i^* = \frac{y_i - \bar{y}_n}{\sqrt{(n-1)S_{yy}}} \text{ e } z_{ij} = \frac{x_{ij} - \bar{x}_j}{\sqrt{(n-1)S_{jj}}}$$

gerando o mesmo estimador do vetor $\boldsymbol{b}$. Tal método é denominado por escala de comprimento unitário (*Unit length Scaling*).

Multicolinearidade

- Um sério problema que pode ocorrer no ajuste de um MRLM é o de multicolinearidade ou dependência linear aproximada entre as variáveis explicativas, fazendo com a que a matriz de especificação $\boldsymbol{X}$ não seja de posto completo e por conseguinte, resultando que $\boldsymbol{X}^\top\boldsymbol{X}$ seja singular (ou aproximadamente singular).
- Mesmo que não exista multicolinearidade perfeita, é comum a existência de multicolinearidade elevada, causando alguns problemas, como por exemplo:
- Instabilidade numérica para se obter $(\boldsymbol{X}^\top\boldsymbol{X})^{-1}$ e por conseguinte o EMQ de $\boldsymbol{\beta}$.
- Os valores de $\widehat{\mathrm{Var}}(\widehat{\beta}_i)$ serão superestimados.
- Problemas para selecionar submodelos.
- Falta de robustez ao se fazer pequenas perturbações nos dados.

Multicolinearidade

- Um sério problema que pode ocorrer no ajuste de um MRLM é o de multicolinearidade ou dependência linear aproximada entre as variáveis explicativas, fazendo com a que a matriz de especificação $\boldsymbol{X}$ não seja de posto completo e por conseguinte, resultando que $\boldsymbol{X}^\top\boldsymbol{X}$ seja singular (ou aproximadamente singular).
- Mesmo que não exista multicolinearidade perfeita, é comum a existência de multicolinearidade elevada, causando alguns problemas, como por exemplo:
- Instabilidade numérica para se obter $(\boldsymbol{X}^\top\boldsymbol{X})^{-1}$ e por conseguinte o EMQ de β.
- Os valores de $\widehat{\mathrm{Var}}(\widehat{\beta}_i)$ serão superestimados.
- Problemas para selecionar submodelos.
- Falta de robustez ao se fazer pequenas perturbações nos dados.

Multicolinearidade

- Um sério problema que pode ocorrer no ajuste de um MRLM é o de multicolinearidade ou dependência linear aproximada entre as variáveis explicativas, fazendo com a que a matriz de especificação $\boldsymbol{X}$ não seja de posto completo e por conseguinte, resultando que $\boldsymbol{X}^\top \boldsymbol{X}$ seja singular (ou aproximadamente singular).
- Mesmo que não exista multicolinearidade perfeita, é comum a existência de multicolinearidade elevada, causando alguns problemas, como por exemplo:
- Instabilidade numérica para se obter $(\boldsymbol{X}^\top \boldsymbol{X})^{-1}$ e por conseguinte o EMQ de $\boldsymbol{\beta}$.
- Os valores de $\widehat{\mathrm{Var}}(\widehat{\beta}_i)$ serão superestimados.
- Problemas para selecionar submodelos.
- Falta de robustez ao se fazer pequenas perturbações nos dados.

Multicolinearidade

- Um sério problema que pode ocorrer no ajuste de um MRLM é o de multicolinearidade ou dependência linear aproximada entre as variáveis explicativas, fazendo com a que a matriz de especificação $\boldsymbol{X}$ não seja de posto completo e por conseguinte, resultando que $\boldsymbol{X}^\top\boldsymbol{X}$ seja singular (ou aproximadamente singular).
- Mesmo que não exista multicolinearidade perfeita, é comum a existência de multicolinearidade elevada, causando alguns problemas, como por exemplo:
- Instabilidade numérica para se obter $(\boldsymbol{X}^\top\boldsymbol{X})^{-1}$ e por conseguinte o EMQ de $\boldsymbol{\beta}$.
- Os valores de $\widehat{\mathrm{Var}}(\widehat{\beta}_i)$ serão superestimados.
- Problemas para selecionar submodelos.
- Falta de robustez ao se fazer pequenas perturbações nos dados.

Multicolinearidade

- Um sério problema que pode ocorrer no ajuste de um MRLM é o de multicolinearidade ou dependência linear aproximada entre as variáveis explicativas, fazendo com a que a matriz de especificação $\boldsymbol{X}$ não seja de posto completo e por conseguinte, resultando que $\boldsymbol{X}^\top \boldsymbol{X}$ seja singular (ou aproximadamente singular).
- Mesmo que não exista multicolinearidade perfeita, é comum a existência de multicolinearidade elevada, causando alguns problemas, como por exemplo:
- Instabilidade numérica para se obter $(\boldsymbol{X}^\top \boldsymbol{X})^{-1}$ e por conseguinte o EMQ de $\boldsymbol{\beta}$.
- Os valores de $\widehat{\mathrm{Var}}(\widehat{\beta}_i)$ serão superestimados.
- Problemas para selecionar submodelos.
- Falta de robustez ao se fazer pequenas perturbações nos dados.

Multicolinearidade

- Um sério problema que pode ocorrer no ajuste de um MRLM é o de multicolinearidade ou dependência linear aproximada entre as variáveis explicativas, fazendo com a que a matriz de especificação $\boldsymbol{X}$ não seja de posto completo e por conseguinte, resultando que $\boldsymbol{X}^\top \boldsymbol{X}$ seja singular (ou aproximadamente singular).
- Mesmo que não exista multicolinearidade perfeita, é comum a existência de multicolinearidade elevada, causando alguns problemas, como por exemplo:
- Instabilidade numérica para se obter $(\boldsymbol{X}^\top \boldsymbol{X})^{-1}$ e por conseguinte o EMQ de $\boldsymbol{\beta}$.
- Os valores de $\widehat{\mathrm{Var}}(\widehat{\beta}_i)$ serão superestimados.
- Problemas para selecionar submodelos.
- Falta de robustez ao se fazer pequenas perturbações nos dados.

Multicolinearidade

- Como conseguir identificar/diagnosticar a multicolinearidade?
- Podemos inicialmente fazer uso do gráfico de dispersão multivariado somente das variáveis explicativas.
- Na literatura, sugere-se avaliar o valor de

$$k = \frac{\mathrm{e}_{\max}}{\mathrm{e}_{\min}},$$

em que $\mathrm{e}_{\max}$ e $\mathrm{e}_{\min}$ representam, respectivamente, o maior e menor autovalor da matriz $\boldsymbol{X}^\top\boldsymbol{X}$, com k denominado por número de condição de $\boldsymbol{X}^\top\boldsymbol{X}$.
- Quanto maior for o número de condição, maior será o grau de multicolinearidade. Valores de referência para k são dados em Montgomery et al. (2012, Cap. 9).

Multicolinearidade

- Como conseguir identificar/diagnosticar a multicolinearidade?
- Podemos inicialmente fazer uso do gráfico de dispersão multivariado somente das variáveis explicativas.
- Na literatura, sugere-se avaliar o valor de

$$k = \frac{\mathrm{e}_{\max}}{\mathrm{e}_{\min}},$$

em que $\mathrm{e}_{\max}$ e $\mathrm{e}_{\min}$ representam, respectivamente, o maior e menor autovalor da matriz $\boldsymbol{X}^\top \boldsymbol{X}$, com k denominado por número de condição de $\boldsymbol{X}^\top \boldsymbol{X}$.
- Quanto maior for o número de condição, maior será o grau de multicolinearidade. Valores de referência para k são dados em Montgomery et al. (2012, Cap. 9).

Multicolinearidade

- Como conseguir identificar/diagnosticar a multicolinearidade?
- Podemos inicialmente fazer uso do gráfico de dispersão multivariado somente das variáveis explicativas.
- Na literatura, sugere-se avaliar o valor de

$$k = \frac{\mathrm{e}_{\max}}{\mathrm{e}_{\min}},$$

em que $\mathrm{e}_{\max}$ e $\mathrm{e}_{\min}$ representam, respectivamente, o maior e menor autovalor da matriz $\boldsymbol{X}^\top \boldsymbol{X}$, com k denominado por número de condição de $\boldsymbol{X}^\top \boldsymbol{X}$.

- Quanto maior for o número de condição, maior será o grau de multicolinearidade. Valores de referência para k são dados em Montgomery et al. (2012, Cap. 9).

Multicolinearidade

- Como conseguir identificar/diagnosticar a multicolinearidade?
- Podemos inicialmente fazer uso do gráfico de dispersão multivariado somente das variáveis explicativas.
- Na literatura, sugere-se avaliar o valor de

$$k = \frac{\mathrm{e}_{\max}}{\mathrm{e}_{\min}},$$

em que $\mathrm{e}_{\max}$ e $\mathrm{e}_{\min}$ representam, respectivamente, o maior e menor autovalor da matriz $\boldsymbol{X}^\top\boldsymbol{X}$, com k denominado por número de condição de $\boldsymbol{X}^\top\boldsymbol{X}$.
- Quanto maior for o número de condição, maior será o grau de multicolinearidade. Valores de referência para k são dados em Montgomery et al. (2012, Cap. 9).

Fontes de Multicolinearidade

- Planejamento de coleta, experimentos do tipo mistura, por exemplo.
- Restrições no modelo ou na população.
- Especificação do modelo.
- Modelo super ajustado (*overdefined model*).
- Para detalhes de como diagnosticar multicolinearidade, vejam por exemplo, Montgomery et al. (2012, Cap.11) e Draper and Smith (1998, Cap. 16), por exemplo.
- Uma forma de contornar este problema é usar regressão em cristas (*ridge regression*). Para detalhes, veja por exemplo, Montgomery et al. (2012, Cap.11) e Draper and Smith (1998, Cap. 17).

Fontes de Multicolinearidade

- Planejamento de coleta, experimentos do tipo mistura, por exemplo.
- Restrições no modelo ou na população.
- Especificação do modelo.
- Modelo super ajustado (*overdefined model*).
- Para detalhes de como diagnosticar multicolinearidade, vejam por exemplo, Montgomery et al. (2012, Cap.11) e Draper and Smith (1998, Cap. 16), por exemplo.
- Uma forma de contornar este problema é usar regressão em cristas (*ridge regression*). Para detalhes, veja por exemplo, Montgomery et al. (2012, Cap.11) e Draper and Smith (1998, Cap. 17).

Fontes de Multicolinearidade

- Planejamento de coleta, experimentos do tipo mistura, por exemplo.
- Restrições no modelo ou na população.
- Especificação do modelo.
- Modelo super ajustado (*overdefined model*).
- Para detalhes de como diagnosticar multicolinearidade, vejam por exemplo, Montgomery et al. (2012, Cap.11) e Draper and Smith (1998, Cap. 16), por exemplo.
- Uma forma de contornar este problema é usar regressão em cristas (*ridge regression*). Para detalhes, veja por exemplo, Montgomery et al. (2012, Cap.11) e Draper and Smith (1998, Cap. 17).

Fontes de Multicolinearidade

- Planejamento de coleta, experimentos do tipo mistura, por exemplo.
- Restrições no modelo ou na população.
- Especificação do modelo.
- Modelo super ajustado (*overdefined model*).
- Para detalhes de como diagnosticar multicolinearidade, vejam por exemplo, Montgomery et al. (2012, Cap.11) e Draper and Smith (1998, Cap. 16), por exemplo.
- Uma forma de contornar este problema é usar regressão em cristas (*ridge regression*). Para detalhes, veja por exemplo, Montgomery et al. (2012, Cap.11) e Draper and Smith (1998, Cap. 17).

Fontes de Multicolinearidade

- Planejamento de coleta, experimentos do tipo mistura, por exemplo.
- Restrições no modelo ou na população.
- Especificação do modelo.
- Modelo super ajustado (*overdefined model*).
- Para detalhes de como diagnosticar multicolinearidade, vejam por exemplo, Montgomery et al. (2012, Cap.11) e Draper and Smith (1998, Cap. 16), por exemplo.
- Uma forma de contornar este problema é usar regressão em cristas (*ridge regression*). Para detalhes, veja por exemplo, Montgomery et al. (2012, Cap.11) e Draper and Smith (1998, Cap. 17).

Fontes de Multicolinearidade

- Planejamento de coleta, experimentos do tipo mistura, por exemplo.
- Restrições no modelo ou na população.
- Especificação do modelo.
- Modelo super ajustado (*overdefined model*).
- Para detalhes de como diagnosticar multicolinearidade, vejam por exemplo, Montgomery et al. (2012, Cap.11) e Draper and Smith (1998, Cap. 16), por exemplo.
- Uma forma de contornar este problema é usar regressão em cristas (*ridge regression*). Para detalhes, veja por exemplo, Montgomery et al. (2012, Cap.11) e Draper and Smith (1998, Cap. 17).

Modelos de regressão polinomiais

- A classe de MRLM com forma funcional

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e} \tag{12}$$

é extremamente ampla e é apropriada em situações em que a função de regressão é linear com relação ao vetor de parâmetros $\boldsymbol{\beta}$.

- Esta classe de modelos inclui como caso particular várias classes importantes, como por exemplo, os modelos polinomias. Para exemplificar, consideremos o modelo polinomial univariado de segunda ordem

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + e_i$$

e o modelo polinomial bivariado de segunda ordem

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_{11} x_{1i}^2 + \beta_{22} x_{2i}^2 + \beta_{12} x_{1i} x_{2i} + e_i,$$

que são casos particulares de (12).

- **Pergunta**: Como mostramos que estes modelos são casos particulares de (12)?

Modelos de regressão polinomiais

- A classe de MRLM com forma funcional

$$\boldsymbol{y} = \boldsymbol{X}\beta + \boldsymbol{e} \tag{12}$$

é extremamente ampla e é apropriada em situações em que a função de regressão é linear com relação ao vetor de parâmetros $\boldsymbol{\beta}$.

- Esta classe de modelos inclui como caso particular várias classes importantes, como por exemplo, os modelos polinomias. Para exemplificar, consideremos o modelo polinomial univariado de segunda ordem

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + e_i$$

e o modelo polinomial bivariado de segunda ordem

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_{11} x_{1i}^2 + \beta_{22} x_{2i}^2 + \beta_{12} x_{1i} x_{2i} + e_i,$$

que são casos particulares de (12).

- **Pergunta**: Como mostramos que estes modelos são casos particulares de (12)?

Modelos de regressão polinomiais

- A classe de MRLM com forma funcional

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e} \tag{12}$$

é extremamente ampla e é apropriada em situações em que a função de regressão é linear com relação ao vetor de parâmetros $\boldsymbol{\beta}$.

- Esta classe de modelos inclui como caso particular várias classes importantes, como por exemplo, os modelos polinomias. Para exemplificar, consideremos o modelo polinomial univariado de segunda ordem

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + e_i$$

e o modelo polinomial bivariado de segunda ordem

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_{11} x_{1i}^2 + \beta_{22} x_{2i}^2 + \beta_{12} x_{1i} x_{2i} + e_i,$$

que são casos particulares de (12).

- **Pergunta**: Como mostramos que estes modelos são casos particulares de (12)?

Modelos de regressão polinomiais

- Esses modelos são bastante utilizados quando o comportamento médio é curvilíneo ou quando o comportamento é não linear e não se tem a ideia da forma (ou é extremamente complexa) da função de regressão.
- A última afirmação segue do fato que qualquer função diferenciável consegue ser expandida/aproximada por polinômios através da expansão em série de Taylor.

Modelos de regressão polinomiais

- Esses modelos são bastante utilizados quando o comportamento médio é curvilíneo ou quando o comportamento é não linear e não se tem a ideia da forma (ou é extremamente complexa) da função de regressão.
- A última afirmação segue do fato que qualquer função diferenciável consegue ser expandida/aproximada por polinômios através da expansão em série de Taylor.

Modelos de regressão polinomiais - Princípios básicos

- Quando a função de regressão é dada por

$$\mu(x, \boldsymbol{\beta}) = \sum_{j=0}^{k} \beta_j x_i^j,$$

o modelo é dito ser polinomial univariado de k-ésima ordem.

- Fazendo $x_{ij} := x_i^j, j = 0, \ldots, k$, podemos ajustar o modelo polinomial de k-ésima ordem univariado através da metodologia de MRLM. Todavia, os parâmetros associados não possuem a mesma interpretação.

- Se tomarmos $k = 2$, obtemos o modelo polinomial univariado de segunda ordem

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + e_i,$$

ou simplesmente o modelo polinomial quadrático.

- Neste caso, β_1 representa o **efeito linear**, enquanto que β_2 representa o **efeito quadrático**. O intercepto possui interpretação padrão, caso a amplitude dos dados contenha $x_i = 0$, caso contrário ele não possui interpretação física.

Modelos de regressão polinomiais - Princípios básicos

- Quando a função de regressão é dada por

$$\mu(x, \boldsymbol{\beta}) = \sum_{j=0}^{k} \beta_j x_i^j,$$

o modelo é dito ser polinomial univariado de k-ésima ordem.

- Fazendo $x_{ij} := x_i^j, j = 0, \ldots, k$, podemos ajustar o modelo polinomial de k-ésima ordem univariado através da metodologia de MRLM. Todavia, os parâmetros associados não possuem a mesma interpretação.

- Se tomarmos $k = 2$, obtemos o modelo polinomial univariado de segunda ordem

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + e_i,$$

ou simplesmente o modelo polinomial quadrático.

- Neste caso, β_1 representa o **efeito linear**, enquanto que β_2 representa o **efeito quadrático**. O intercepto possui interpretação padrão, caso a amplitude dos dados contenha $x_i = 0$, caso contrário ele não possui interpretação física.

Modelos de regressão polinomiais - Princípios básicos

- Quando a função de regressão é dada por

$$\mu(x, \boldsymbol{\beta}) = \sum_{j=0}^{k} \beta_j x_i^j,$$

o modelo é dito ser polinomial univariado de k-ésima ordem.

- Fazendo $x_{ij} := x_i^j, j = 0, \ldots, k$, podemos ajustar o modelo polinomial de k-ésima ordem univariado através da metodologia de MRLM. Todavia, os parâmetros associados não possuem a mesma interpretação.

- Se tomarmos $k = 2$, obtemos o modelo polinomial univariado de segunda ordem

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + e_i,$$

ou simplesmente o modelo polinomial quadrático.

- Neste caso, β_1 representa o **efeito linear**, enquanto que β_2 representa o **efeito quadrático**. O intercepto possui interpretação padrão, caso a amplitude dos dados contenha $x_i = 0$, caso contrário ele não possui interpretação física.

Modelos de regressão polinomiais - Princípios básicos

- Quando a função de regressão é dada por

$$\mu(x, \boldsymbol{\beta}) = \sum_{j=0}^{k} \beta_j x_i^j,$$

o modelo é dito ser polinomial univariado de k-ésima ordem.

- Fazendo $x_{ij} := x_i^j, j = 0, \ldots, k$, podemos ajustar o modelo polinomial de k-ésima ordem univariado através da metodologia de MRLM. Todavia, os parâmetros associados não possuem a mesma interpretação.

- Se tomarmos $k = 2$, obtemos o modelo polinomial univariado de segunda ordem

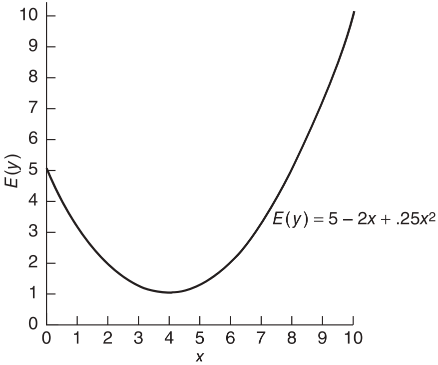
$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + e_i,$$

ou simplesmente o modelo polinomial quadrático.

- Neste caso, β_1 representa o **efeito linear**, enquanto que β_2 representa o **efeito quadrático**. O intercepto possui interpretação padrão, caso a amplitude dos dados contenha $x_i = 0$, caso contrário ele não possui interpretação física.

Ilustração

Figura: Exemplo de um modelo quadrático (Montgomery et al. , 2012).



Considerações

1. **Ordem do modelo**:

- Valores mais altos de k implicam em um modelo mais bem ajustado. Por que?
- Todavia, é preferível que a ordem seja a mais baixa possível, tendo em vista que valores altos de k implicam em uma quantidade maior de parâmetros, prejudicando a interpretabilidade e a parcimônia do modelo. Isso considerando apenas uma variável explicativa, a medida que temos mais variáveis a complexidade aumenta bastante.
- Antigamente, consideravam-se transformações na variável resposta e posteriormente ajustava-se modelos de baixa ordem, mas hoje não há mais necessidade de tal procedimento. Por que?
- Ajustes arbitrários de polinômios de alta ordem são vistos em geral como abuso sério da metodologia de análise de regressão.
- Alguns autores sugerem a utilização dos métodos backward ou forward para selecionar a ordem do modelo, i.e, começando do maior/menor modelo possível até obter todos os coeficientes significativos. Os métodos não necessariamente fornecem o mesmo modelo final e também acabam não levando em consideração o critério da parcimônia.

Considerações

1. **Ordem do modelo**:

- Valores mais altos de k implicam em um modelo mais bem ajustado. Por que?
- Todavia, é preferível que a ordem seja a mais baixa possível, tendo em vista que valores altos de k implicam em uma quantidade maior de parâmetros, prejudicando a interpretabilidade e a parcimônia do modelo. Isso considerando apenas uma variável explicativa, a medida que temos mais variáveis a complexidade aumenta bastante.
- Antigamente, consideravam-se transformações na variável resposta e posteriormente ajustava-se modelos de baixa ordem, mas hoje não há mais necessidade de tal procedimento. Por que?
- Ajustes arbitrários de polinômios de alta ordem são vistos em geral como abuso sério da metodologia de análise de regressão.
- Alguns autores sugerem a utilização dos métodos backward ou forward para selecionar a ordem do modelo, i.e, começando do maior/menor modelo possível até obter todos os coeficientes significativos. Os métodos não necessariamente fornecem o mesmo modelo final e também acabam não levando em consideração o critério da parcimônia.

Considerações

1. **Ordem do modelo**:

- Valores mais altos de k implicam em um modelo mais bem ajustado. Por que?
- Todavia, é preferível que a ordem seja a mais baixa possível, tendo em vista que valores altos de k implicam em uma quantidade maior de parâmetros, prejudicando a interpretabilidade e a parcimônia do modelo. Isso considerando apenas uma variável explicativa, a medida que temos mais variáveis a complexidade aumenta bastante.
- Antigamente, consideravam-se transformações na variável resposta e posteriormente ajustava-se modelos de baixa ordem, mas hoje não há mais necessidade de tal procedimento. Por que?
- Ajustes arbitrários de polinômios de alta ordem são vistos em geral como abuso sério da metodologia de análise de regressão.
- Alguns autores sugerem a utilização dos métodos backward ou forward para selecionar a ordem do modelo, i.e, começando do maior/menor modelo possível até obter todos os coeficientes significativos. Os métodos não necessariamente fornecem o mesmo modelo final e também acabam não levando em consideração o critério da parcimônia.

Considerações

1. **Ordem do modelo**:
   - Valores mais altos de k implicam em um modelo mais bem ajustado. Por que?
   - Todavia, é preferível que a ordem seja a mais baixa possível, tendo em vista que valores altos de k implicam em uma quantidade maior de parâmetros, prejudicando a interpretabilidade e a parcimônia do modelo. Isso considerando apenas uma variável explicativa, a medida que temos mais variáveis a complexidade aumenta bastante.
   - Antigamente, consideravam-se transformações na variável resposta e posteriormente ajustava-se modelos de baixa ordem, mas hoje não há mais necessidade de tal procedimento. Por que?
   - Ajustes arbitrários de polinômios de alta ordem são vistos em geral como abuso sério da metodologia de análise de regressão.
   - Alguns autores sugerem a utilização dos métodos backward ou forward para selecionar a ordem do modelo, i.e, começando do maior/menor modelo possível até obter todos os coeficientes significativos. Os métodos não necessariamente fornecem o mesmo modelo final e também acabam não levando em consideração o critério da parcimônia.

Considerações

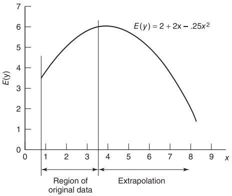
1. **Ordem do modelo**:

- Valores mais altos de k implicam em um modelo mais bem ajustado. Por que?
- Todavia, é preferível que a ordem seja a mais baixa possível, tendo em vista que valores altos de k implicam em uma quantidade maior de parâmetros, prejudicando a interpretabilidade e a parcimônia do modelo. Isso considerando apenas uma variável explicativa, a medida que temos mais variáveis a complexidade aumenta bastante.
- Antigamente, consideravam-se transformações na variável resposta e posteriormente ajustava-se modelos de baixa ordem, mas hoje não há mais necessidade de tal procedimento. Por que?
- Ajustes arbitrários de polinômios de alta ordem são vistos em geral como abuso sério da metodologia de análise de regressão.
- Alguns autores sugerem a utilização dos métodos backward ou forward para selecionar a ordem do modelo, i.e, começando do maior/menor modelo possível até obter todos os coeficientes significativos. Os métodos não necessariamente fornecem o mesmo modelo final e também acabam não levando em consideração o critério da parcimônia.

Considerações

2. Extrapolação:

Figura: Perigo da extrapolação (Montgomery et al. , 2012).

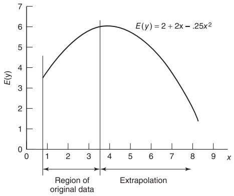


- Assim como todo modelo de regressão, deve-se ter muita cautela a se fazer extrapolações.
- Em geral, os modelos polinomiais podem ser extremamente imprecisos e inadequados para se fazer extrapolações, especialmente em regiões remotas do subespaço $\mathbb{C}(\boldsymbol{X})$.

Considerações

2. Extrapolação:

Figura: Perigo da extrapolação (Montgomery et al. , 2012).

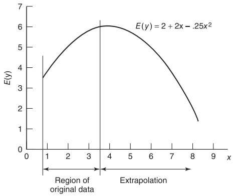


- Assim como todo modelo de regressão, deve-se ter muita cautela a se fazer extrapolações.
- Em geral, os modelos polinomiais podem ser extremamente imprecisos e inadequados para se fazer extrapolações, especialmente em regiões remotas do subespaço $\mathbb{C}(\boldsymbol{X})$.

Considerações

2. Extrapolação:

Figura: Perigo da extrapolação (Montgomery et al. , 2012).



- Assim como todo modelo de regressão, deve-se ter muita cautela a se fazer extrapolações.
- Em geral, os modelos polinomiais podem ser extremamente imprecisos e inadequados para se fazer extrapolações, especialmente em regiões remotas do subespaço $\mathbb{C}(\boldsymbol{X})$.

Considerações

3. **Mal-condicionamento (ill-conditioned):**

- Dependendo da amplitude dos dados, a matriz $\boldsymbol{X}^\top \boldsymbol{X}$ pode apresentar problemas numéricos para ser invertida, gerando por exemplo elevados erros-padrão estimados.
- Esse problema tende a acontecer a medida que se aumenta o grau do polinômio a ser ajustado.
- Uma primeira alternativa para contornar o mal-condicionamento é centralizar as variáveis, i.e., considerar a função de regressão

$$\mu(x_i; \boldsymbol{\beta}) = \sum_{j=0}^{k} \beta_j (x_i - \bar{x})^j,$$

todavia nem sempre uma simples centralização na média pode resolver, implicando que em casos mais gerais, deve-se partir para procedimentos mais sofisticados. Apresentaremos um procedimento em breve que resolver este inconveniente.
- Em geral este problema acontece com mais intensidade quando a variável explicativa assume valores somente em um intervalo limitado.

Considerações

3. **Mal-condicionamento (ill-conditioned):**

- Dependendo da amplitude dos dados, a matriz $\boldsymbol{X}^\top \boldsymbol{X}$ pode apresentar problemas numéricos para ser invertida, gerando por exemplo elevados erros-padrão estimados.

- Esse problema tende a acontecer a medida que se aumenta o grau do polinômio a ser ajustado.

- Uma primeira alternativa para contornar o mal-condicionamento é centralizar as variáveis, i.e., considerar a função de regressão

$$\mu(x_i; \boldsymbol{\beta}) = \sum_{j=0}^{k} \beta_j (x_i - \bar{x})^j,$$

todavia nem sempre uma simples centralização na média pode resolver, implicando que em casos mais gerais, deve-se partir para procedimentos mais sofisticados. Apresentaremos um procedimento em breve que resolver este inconveniente.

- Em geral este problema acontece com mais intensidade quando a variável explicativa assume valores somente em um intervalo limitado.

Considerações

3. **Mal-condicionamento (ill-conditioned):**

- Dependendo da amplitude dos dados, a matriz $\boldsymbol{X}^\top \boldsymbol{X}$ pode apresentar problemas numéricos para ser invertida, gerando por exemplo elevados erros-padrão estimados.
- Esse problema tende a acontecer a medida que se aumenta o grau do polinômio a ser ajustado.
- Uma primeira alternativa para contornar o mal-condicionamento é centralizar as variáveis, i.e., considerar a função de regressão

$$\mu(x_i; \boldsymbol{\beta}) = \sum_{j=0}^{k} \beta_j (x_i - \bar{x})^j,$$

todavia nem sempre uma simples centralização na média pode resolver, implicando que em casos mais gerais, deve-se partir para procedimentos mais sofisticados. Apresentaremos um procedimento em breve que resolver este inconveniente.
- Em geral este problema acontece com mais intensidade quando a variável explicativa assume valores somente em um intervalo limitado.

Considerações

3. **Mal-condicionamento (ill-conditioned):**

- Dependendo da amplitude dos dados, a matriz $\boldsymbol{X}^\top\boldsymbol{X}$ pode apresentar problemas numéricos para ser invertida, gerando por exemplo elevados erros-padrão estimados.
- Esse problema tende a acontecer a medida que se aumenta o grau do polinômio a ser ajustado.
- Uma primeira alternativa para contornar o mal-condicionamento é centralizar as variáveis, i.e., considerar a função de regressão

$$\mu(x_i;\boldsymbol{\beta}) = \sum_{j=0}^{k} \beta_j (x_i - \bar{x})^j,$$

todavia nem sempre uma simples centralização na média pode resolver, implicando que em casos mais gerais, deve-se partir para procedimentos mais sofisticados. Apresentaremos um procedimento em breve que resolver este inconveniente.
- Em geral este problema acontece com mais intensidade quando a variável explicativa assume valores somente em um intervalo limitado.

Considerações

4. **Modelos hierárquicos** :

- O modelo

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{1i}^2 + \beta_3 x_{1i}^3 + e_i$$

é dito ser hierárquico, pois contém todos os termos até terceira ordem, ao contrário do modelo

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_3 x_{1i}^3 + e_i.$$

- Peixoto (1987, 1990, American Statistician) comenta que apenas modelos hierárquicos são invariantes sob transformações lineares e sugere que todos os modelos polinomiais devem ter esta propriedade, surgindo assim o termo modelo hierarquicamente bem formulado.
- Conforme Montgomery et al. (2012) citam: "É certamente atraente ter a forma do modelo preservada sob transformações lineares (como ajustar o modelo em variáveis transformadas e, em seguida, converter para um modelo nas variáveis originais), mas é puramente uma sutileza matemática".

Considerações

4. **Modelos hierárquicos :**

- O modelo

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{1i}^2 + \beta_3 x_{1i}^3 + e_i$$

é dito ser hierárquico, pois contém todos os termos até terceira ordem, ao contrário do modelo

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_3 x_{1i}^3 + e_i.$$

- Peixoto (1987, 1990, American Statistician) comenta que apenas modelos hierárquicos são invariantes sob transformações lineares e sugere que todos os modelos polinomiais devem ter esta propriedade, surgindo assim o termo modelo hierarquicamente bem formulado.

- Conforme Montgomery et al. (2012) citam: "É certamente atraente ter a forma do modelo preservada sob transformações lineares (como ajustar o modelo em variáveis transformadas e, em seguida, converter para um modelo nas variáveis originais), mas é puramente uma sutileza matemática".

Considerações

4. **Modelos hierárquicos :**

- O modelo

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{1i}^2 + \beta_3 x_{1i}^3 + e_i$$

é dito ser hierárquico, pois contém todos os termos até terceira ordem, ao contrário do modelo

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_3 x_{1i}^3 + e_i.$$

- Peixoto (1987, 1990, American Statistician) comenta que apenas modelos hierárquicos são invariantes sob transformações lineares e sugere que todos os modelos polinomiais devem ter esta propriedade, surgindo assim o termo modelo hierarquicamente bem formulado.

- Conforme Montgomery et al. (2012) citam: "É certamente atraente ter a forma do modelo preservada sob transformações lineares (como ajustar o modelo em variáveis transformadas e, em seguida, converter para um modelo nas variáveis originais), mas é puramente uma sutileza matemática".

Exercício (entregar próxima aula)

Exercício 22: Reproduzir o Exemplo 7.1 (The Hardwood Data) de Montgomery et al. (2012) em formato de relatório adicionando a metodologia de diagnóstico.

Observações

- No software R para ajuste de um modelo polinomial de terceira ordem, utiliza-se o comando `lm(y~x+I(x^2)+I(x^3))`.
- Lembrando que o operador `I( )` é o adequado para realizar operações aritméticas. Para detalhes, ver `?formula`, por exemplo.
- É comum em regressão não paramétrica fazer uso de Splines (ajustes polinomiais por regiões- Piecewise Polynomial Fitting), que constitui um procedimento consistente de estimação de curvas/densidades.
- Outros procedimentos bastante utilizados são o ajuste de modelos de regressão usando suavizadores do tipo Kernel, regressão ponderada localmente (Locally Weighted Regression-Loess), todas fazendo bastante uso de modelos polinomiais.

Observações

- No software R para ajuste de um modelo polinomial de terceira ordem, utiliza-se o comando `lm(y~x+I(x^2)+I(x^3))`.
- Lembrando que o operador `I( )` é o adequado para realizar operações aritméticas. Para detalhes, ver `?formula`, por exemplo.
- É comum em regressão não paramétrica fazer uso de Splines (ajustes polinomiais por regiões- Piecewise Polynomial Fitting), que constitui um procedimento consistente de estimação de curvas/densidades.
- Outros procedimentos bastante utilizados são o ajuste de modelos de regressão usando suavizadores do tipo Kernel, regressão ponderada localmente (Locally Weighted Regression-Loess), todas fazendo bastante uso de modelos polinomiais.

Observações

- No software `R` para ajuste de um modelo polinomial de terceira ordem, utiliza-se o comando `lm(y~x+I(x^2)+I(x^3))`.
- Lembrando que o operador `I( )` é o adequado para realizar operações aritméticas. Para detalhes, ver `?formula`, por exemplo.
- É comum em regressão não paramétrica fazer uso de Splines (ajustes polinomiais por regiões- Piecewise Polynomial Fitting), que constitui um procedimento consistente de estimação de curvas/densidades.
- Outros procedimentos bastante utilizados são o ajuste de modelos de regressão usando suavizadores do tipo Kernel, regressão ponderada localmente (Locally Weighted Regression-Loess), todas fazendo bastante uso de modelos polinomiais.

Observações

- No software R para ajuste de um modelo polinomial de terceira ordem, utiliza-se o comando `lm(y~x+I(x^2)+I(x^3))`.
- Lembrando que o operador `I( )` é o adequado para realizar operações aritméticas. Para detalhes, ver `?formula`, por exemplo.
- É comum em regressão não paramétrica fazer uso de Splines (ajustes polinomiais por regiões- Piecewise Polynomial Fitting), que constitui um procedimento consistente de estimação de curvas/densidades.
- Outros procedimentos bastante utilizados são o ajuste de modelos de regressão usando suavizadores do tipo Kernel, regressão ponderada localmente (Locally Weighted Regression-Loess), todas fazendo bastante uso de modelos polinomiais.

Modelos polinomiais multivariados

- Modelos polinomiais multivariados constituem uma simples extensão dos modelos polinomias univariados vistos anteriormente.
- Por exemplo, um modelo polinomial bivariado de segunda ordem é um modelo com a seguinte forma funcional

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_{11} x_{1i}^2 + \beta_{22} x_{2i}^2 + \beta_{12} x_{1i} x_{2i} + e_i.$$

  Este modelo contém dois parâmetros associados aos **efeitos lineares**, β_1 e β_2, dois parâmetros associados aos **efeitos quadráticos**, β_{11} e β_{22}, e um parâmetro associado ao **efeito de interação** β_{12}.
- Ajustes de modelos deste tipo recebem atenção especial na literatura, por exemplo, o modelo

$$\mu(\boldsymbol{x}; \boldsymbol{\beta}) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_{11} x_{1i}^2 + \beta_{22} x_{2i}^2 + \beta_{12} x_{1i} x_{2i},$$

  é chamado de **superfície de resposta**, pois podemos representar a função de regressão em termos de uma superfície bidimensional, por exemplo.

Modelos polinomiais multivariados

- Modelos polinomiais multivariados constituem uma simples extensão dos modelos polinomias univariados vistos anteriormente.
- Por exemplo, um modelo polinomial bivariado de segunda ordem é um modelo com a seguinte forma funcional

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_{11} x_{1i}^2 + \beta_{22} x_{2i}^2 + \beta_{12} x_{1i} x_{2i} + e_i.$$

Este modelo contém dois parâmetros associados aos **efeitos lineares**, β_1 e β_2, dois parâmetros associados aos **efeitos quadráticos**, β_{11} e β_{22}, e um parâmetro associado ao **efeito de interação** β_{12}.

- Ajustes de modelos deste tipo recebem atenção especial na literatura, por exemplo, o modelo

$$\mu(\boldsymbol{x}; \boldsymbol{\beta}) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_{11} x_{1i}^2 + \beta_{22} x_{2i}^2 + \beta_{12} x_{1i} x_{2i},$$

é chamado de **superfície de resposta**, pois podemos representar a função de regressão em termos de uma superfície bidimensional, por exemplo.

Modelos polinomiais multivariados

- Modelos polinomiais multivariados constituem uma simples extensão dos modelos polinomias univariados vistos anteriormente.
- Por exemplo, um modelo polinomial bivariado de segunda ordem é um modelo com a seguinte forma funcional

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_{11} x_{1i}^2 + \beta_{22} x_{2i}^2 + \beta_{12} x_{1i} x_{2i} + e_i.$$

  Este modelo contém dois parâmetros associados aos **efeitos lineares**, β_1 e β_2, dois parâmetros associados aos **efeitos quadráticos**, β_{11} e β_{22}, e um parâmetro associado ao **efeito de interação** β_{12}.
- Ajustes de modelos deste tipo recebem atenção especial na literatura, por exemplo, o modelo

$$\mu(\boldsymbol{x}; \boldsymbol{\beta}) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_{11} x_{1i}^2 + \beta_{22} x_{2i}^2 + \beta_{12} x_{1i} x_{2i},$$

  é chamado de **superfície de resposta**, pois podemos representar a função de regressão em termos de uma superfície bidimensional, por exemplo.

Modelos polinomiais multivariados

- Podemos representar a superfície de resposta bidimensional graficamente desenhando os eixos x_1 e x_2 no plano e visualizando a função de regressão $\mathbb{E}[y|\boldsymbol{x}] = \mu(\boldsymbol{x}, \boldsymbol{\beta})$ no eixo perpendicular.
- Por exemplo, na Figura (7) plotamos a superficie de resposta associada a função de regressão

$$\mu(\boldsymbol{x}; \boldsymbol{\beta}) = 800 + 10x_1 + 7x_2 - 8,5x_1^2 - 5x_2^2 + 4x_1x_2.$$

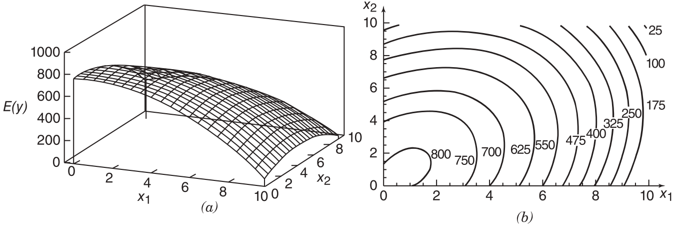
Modelos polinomiais multivariados

- Podemos representar a superfície de resposta bidimensional graficamente desenhando os eixos x_1 e x_2 no plano e visualizando a função de regressão $\mathbb{E}[y|\boldsymbol{x}] = \mu(\boldsymbol{x}, \boldsymbol{\beta})$ no eixo perpendicular.
- Por exemplo, na Figura (7) plotamos a superficie de resposta associada a função de regressão

$$\mu(\boldsymbol{x}; \boldsymbol{\beta}) = 800 + 10x_1 + 7x_2 - 8,5x_1^2 - 5x_2^2 + 4x_1x_2.$$

Ilustração

Figura: (a) Hiperplano de regressão e (b) gráfico de contorno para o modelo $\mu(\boldsymbol{x}_i) = \mu(\boldsymbol{x};\boldsymbol{\beta}) = 800 + 10x_1 + 7x_2 - 8,5x_1^2 - 5x_2^2 + 4x_1x_2$ (Montgomery et al., 2012).



Modelos polinomiais multivariados

- Muitas vezes o interesse é determinar o ponto de máximo/mínima resposta.
- Essa técnica é comumente denominada de metodologia de superfície de resposta e é extremamente utilizada em processos industriais, tecnologia de alimentos, etc.
- No exemplo hipotético apresentado, se tem uma resposta máxima atingida/obtida no par $(0, 0)$.
- Um exemplo hipotético de aplicação, seria determinar as condições de operação que otimizam o processo de resposta.

Modelos polinomiais multivariados

- Muitas vezes o interesse é determinar o ponto de máximo/mínima resposta.
- Essa técnica é comumente denominada de metodologia de superfície de resposta e é extremamente utilizada em processos industriais, tecnologia de alimentos, etc.
- No exemplo hipotético apresentado, se tem uma resposta máxima atingida/obtida no par $(0, 0)$.
- Um exemplo hipotético de aplicação, seria determinar as condições de operação que otimizam o processo de resposta.

Modelos polinomiais multivariados

- Muitas vezes o interesse é determinar o ponto de máximo/mínima resposta.
- Essa técnica é comumente denominada de metodologia de superfície de resposta e é extremamente utilizada em processos industriais, tecnologia de alimentos, etc.
- No exemplo hipotético apresentado, se tem uma resposta máxima atingida/obtida no par $(0, 0)$.
- Um exemplo hipotético de aplicação, seria determinar as condições de operação que otimizam o processo de resposta.

Modelos polinomiais multivariados

- Muitas vezes o interesse é determinar o ponto de máximo/mínima resposta.
- Essa técnica é comumente denominada de metodologia de superfície de resposta e é extremamente utilizada em processos industriais, tecnologia de alimentos, etc.
- No exemplo hipotético apresentado, se tem uma resposta máxima atingida/obtida no par $(0, 0)$.
- Um exemplo hipotético de aplicação, seria determinar as condições de operação que otimizam o processo de resposta.

Outras Extensões

- Para deixar o modelo ainda mais flexível e com possibilidade de modelar sazonalidade, por exemplo, podemos adicionar na classe de modelos polinomiais termos trigonométricos, como por exemplo

$$y_i = \beta_0 + \sum_{j=1}^{k} \beta_j x_i^j + \sum_{j=1}^{r} [\delta_j \sin(jx_i) + \gamma_j \cos(jx_i)] + e_i.$$

  Para detalhes a respeito desta classe de modelos, veja Montgomery et al. (2012), por exemplo.

- Outra extensão extremamente interessante e que vem recebendo atenção especial na literatura é a classe de polinôminos fracionários proposta por Royston and Altman (1994, Applied Statistics), artigo publicado com discussão. Existem vários artigos recentes e inclusive um livro (Royston and Saurbrei, 2008) versando sobre esta classe de modelos.

Outras Extensões

- Para deixar o modelo ainda mais flexível e com possibilidade de modelar sazonalidade, por exemplo, podemos adicionar na classe de modelos polinomiais termos trigonométricos, como por exemplo

$$y_i = \beta_0 + \sum_{j=1}^{k} \beta_j x_i^j + \sum_{j=1}^{r} [\delta_j \sin(jx_i) + \gamma_j \cos(jx_i)] + e_i.$$

Para detalhes a respeito desta classe de modelos, veja Montgomery et al. (2012), por exemplo.

- Outra extensão extremamente interessante e que vem recebendo atenção especial na literatura é a classe de polinôminos fracionários proposta por Royston and Altman (1994, Applied Statistics), artigo publicado com discussão. Existem vários artigos recentes e inclusive um livro (Royston and Saurbrei, 2008) versando sobre esta classe de modelos.

Polinômios Ortogonais

- Nem sempre a simples centralização das variáveis elimina o problema de mal-condicionamento. Isso pode ser resolvido usando polinômios ortogonais para ajustar o modelo.
- Lembre-se que toda função pode ser reescrita como uma expansão em termos de senos e cossenos (série de Fourier), bem como através de uma combinação de polinômios em uma base ortogonal. Existem algumas classes, tais como polinômios associados de Legendre, polinômios de Geronimus, polinômios de Jacob, polinômios de Chebychev, polinômios de Hermite e polinômios de Laguerre, por exemplo.
- A ideia é ajustar

$$y_i = \alpha_0 P_0(x_i) + \alpha_1 P_1(x_i) + \ldots + \alpha_k P_k(x_i) + e_i, \ i = 1, \ldots, n,$$

em que $P_u(x)$ representa um polinômio de grau u em x, definidos tais que

$$\begin{aligned} \sum_{i=1}^{n} P_R(x_i) P_S(x_i) &\equiv 0, \forall R \neq S \in \{0, 1, \ldots, k\}, \\ P_0(x_i) &\equiv 1. \end{aligned}$$

Polinômios Ortogonais

- Nem sempre a simples centralização das variáveis elimina o problema de mal-condicionamento. Isso pode ser resolvido usando polinômios ortogonais para ajustar o modelo.
- Lembre-se que toda função pode ser reescrita como uma expansão em termos de senos e cossenos (série de Fourier), bem como através de uma combinação de polinômios em uma base ortogonal. Existem algumas classes, tais como polinômios associados de Legendre, polinômios de Geronimus, polinômios de Jacob, polinômios de Chebychev, polinômios de Hermite e polinômios de Laguerre, por exemplo.
- A ideia é ajustar

$$y_i = \alpha_0 P_0(x_i) + \alpha_1 P_1(x_i) + \ldots + \alpha_k P_k(x_i) + e_i, \; i = 1, \ldots, n,$$

em que $P_u(x)$ representa um polinômio de grau u em x, definidos tais que

$$\begin{aligned} \sum_{i=1}^{n} P_R(x_i) P_S(x_i) &\equiv 0, \forall R \neq S \in \{0, 1, \ldots, k\}, \\ P_0(x_i) &\equiv 1. \end{aligned}$$

Polinômios Ortogonais

- Nem sempre a simples centralização das variáveis elimina o problema de mal-condicionamento. Isso pode ser resolvido usando polinômios ortogonais para ajustar o modelo.
- Lembre-se que toda função pode ser reescrita como uma expansão em termos de senos e cossenos (série de Fourier), bem como através de uma combinação de polinômios em uma base ortogonal. Existem algumas classes, tais como polinômios associados de Legendre, polinômios de Geronimus, polinômios de Jacob, polinômios de Chebychev, polinômios de Hermite e polinômios de Laguerre, por exemplo.
- A ideia é ajustar

$$y_i = \alpha_0 P_0(x_i) + \alpha_1 P_1(x_i) + \ldots + \alpha_k P_k(x_i) + e_i, \; i = 1, \ldots, n,$$

em que $P_u(x)$ representa um polinômio de grau u em x, definidos tais que

$$\begin{aligned} \sum_{i=1}^{n} P_R(x_i) P_S(x_i) &\equiv 0, \forall R \neq S \in \{0, 1, \ldots, k\}, \\ P_0(x_i) &\equiv 1. \end{aligned}$$

Polinômios Ortogonais

- Então, o modelo pode ser reescrito como

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\alpha} + \boldsymbol{e},$$

em que

- $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \ldots, \alpha_k)^\top$.
- E com matriz de especificação $\boldsymbol{X}$ dada por

$$\boldsymbol{X} = \begin{pmatrix} P_0(x_1) & P_1(x_1) & \ldots & P_K(x_1) \\ P_0(x_2) & P_1(x_2) & \ldots & P_K(x_2) \\ \vdots & \vdots & \vdots & \vdots \\ P_0(x_n) & P_1(x_n) & \ldots & P_K(x_n) \end{pmatrix}.$$

- Isso implica por exemplo que

$$\boldsymbol{X}^\top \boldsymbol{X} = \mathrm{diag}\left(\sum_{i=1}^{n} P_0(x_i)^2, \ldots, \sum_{i=1}^{n} P_k(x_i)^2\right).$$

Polinômios Ortogonais

- Então, o modelo pode ser reescrito como

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\alpha} + \boldsymbol{e},$$

em que

- $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \ldots, \alpha_k)^\top$.
- E com matriz de especificação $\boldsymbol{X}$ dada por

$$\boldsymbol{X} = \begin{pmatrix} P_0(x_1) & P_1(x_1) & \ldots & P_K(x_1) \\ P_0(x_2) & P_1(x_2) & \ldots & P_K(x_2) \\ \vdots & \vdots & \vdots & \vdots \\ P_0(x_n) & P_1(x_n) & \ldots & P_K(x_n) \end{pmatrix}.$$

- Isso implica por exemplo que

$$\boldsymbol{X}^\top \boldsymbol{X} = \mathrm{diag}\left(\sum_{i=1}^{n} P_0(x_i)^2, \ldots, \sum_{i=1}^{n} P_k(x_i)^2\right).$$

Polinômios Ortogonais

- Então, o modelo pode ser reescrito como

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\alpha} + \boldsymbol{e},$$

em que

- $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \ldots, \alpha_k)^\top$.
- E com matriz de especificação $\boldsymbol{X}$ dada por

$$\boldsymbol{X} = \begin{pmatrix} P_0(x_1) & P_1(x_1) & \ldots & P_K(x_1) \\ P_0(x_2) & P_1(x_2) & \ldots & P_K(x_2) \\ \vdots & \vdots & \vdots & \vdots \\ P_0(x_n) & P_1(x_n) & \ldots & P_K(x_n) \end{pmatrix}.$$

- Isso implica por exemplo que

$$\boldsymbol{X}^\top\boldsymbol{X} = \mathrm{diag}\left(\sum_{i=1}^{n} P_0(x_i)^2, \ldots, \sum_{i=1}^{n} P_k(x_i)^2\right).$$

Polinômios Ortogonais

- Então, o modelo pode ser reescrito como

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\alpha} + \boldsymbol{e},$$

em que

- $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \ldots, \alpha_k)^\top$.
- E com matriz de especificação $\boldsymbol{X}$ dada por

$$\boldsymbol{X} = \begin{pmatrix} P_0(x_1) & P_1(x_1) & \ldots & P_K(x_1) \\ P_0(x_2) & P_1(x_2) & \ldots & P_K(x_2) \\ \vdots & \vdots & \vdots & \vdots \\ P_0(x_n) & P_1(x_n) & \ldots & P_K(x_n) \end{pmatrix}.$$

- Isso implica por exemplo que

$$\boldsymbol{X}^\top\boldsymbol{X} = \mathrm{diag}\left(\sum_{i=1}^{n} P_0(x_i)^2, \ldots, \sum_{i=1}^{n} P_k(x_i)^2\right).$$

Polinômios Ortogonais

- O estimador de mínimos quadrados de $\boldsymbol{\alpha}$ é dado por

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

de forma que

$$\widehat{\alpha}_j = \frac{\sum_{i=1}^{n} P_j(x_i) y_i}{\sum_{i=1}^{n} P_j(x_i)^2}, j = 0, 1, \ldots, k. \tag{13}$$

- Dado que P_0 é um polinômio identicamente nulo (de grau zero), então $\widehat{\alpha}_0 = \bar{y}_n$.
- É possível mostrar que (Lista)

$$\mathrm{SQRes} = \mathrm{SQT} - \underbrace{\sum_{j=1}^{k} \widehat{\alpha}_j \left[\sum_{i=1}^{n} P_j(x_i) y_i \right]}_{=\mathrm{SQReg}}. \tag{14}$$

Polinômios Ortogonais

- O estimador de mínimos quadrados de $\boldsymbol{\alpha}$ é dado por

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

de forma que

$$\widehat{\alpha}_j = \frac{\sum_{i=1}^{n} P_j(x_i) y_i}{\sum_{i=1}^{n} P_j(x_i)^2}, j = 0, 1, \ldots, k. \quad (13)$$

- Dado que P_0 é um polinômio identicamente nulo (de grau zero), então $\widehat{\alpha}_0 = \bar{y}_n$.

- É possível mostrar que (Lista)

$$\mathrm{SQRes} = \mathrm{SQT} - \underbrace{\sum_{j=1}^{k} \widehat{\alpha}_j \left[\sum_{i=1}^{n} P_j(x_i) y_i \right]}_{=\mathrm{SQReg}}. \quad (14)$$

Polinômios Ortogonais

- O estimador de mínimos quadrados de $\boldsymbol{\alpha}$ é dado por

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

de forma que

$$\widehat{\alpha}_j = \frac{\sum_{i=1}^{n} P_j(x_i) y_i}{\sum_{i=1}^{n} P_j(x_i)^2}, j = 0, 1, \ldots, k. \tag{13}$$

- Dado que P_0 é um polinômio identicamente nulo (de grau zero), então $\widehat{\alpha}_0 = \bar{y}_n$.
- É possível mostrar que (Lista)

$$\mathrm{SQRes} = \mathrm{SQT} - \underbrace{\sum_{j=1}^{k} \widehat{\alpha}_j \left[\sum_{i=1}^{n} P_j(x_i) y_i \right]}_{=\mathrm{SQReg}}. \tag{14}$$

Exercício (entregar próxima aula)

Exercício 23: Prove as identidades (13) e (14).

Polinômios Ortogonais

- Os polinômios ortogonais $P_j(x_i)$ são facilmente construidos para o caso em que os valores de x_i são equidistantes. Neste caso, os 5 primeiros polinômios ortogonais são dados por (Montogomery at el., 2012):

$$
\begin{aligned}
P_0(x_i) &\equiv 1\\
P_1(x_i) &= \lambda_1\left[\frac{x_i-\bar{x}_n}{d}\right]\\
P_2(x_i) &= \lambda_2\left[\left(\frac{x_i-\bar{x}_n}{d}\right)^2-\left(\frac{n^2-1}{12}\right)\right]\\
P_3(x_i) &= \lambda_3\left[\left(\frac{x_i-\bar{x}_n}{d}\right)^3-\left(\frac{x_i-\bar{x}_n}{d}\right)\left(\frac{3n^2-7}{20}\right)\right]\\
P_4(x_i) &= \lambda_4\left[\left(\frac{x_i-\bar{x}_n}{d}\right)^4-\left(\frac{x_i-\bar{x}_n}{d}\right)^2\left(\frac{3n^2-13}{14}\right)+\left(\frac{3(n^2-1)(n^2-9)}{560}\right)\right].
\end{aligned}
$$

- No caso d representa o incremento (diferença constante entre os valores sucessivos da variável explicativa) e os λ's são constantes escolhidas de forma a fazer com que os polinômios assumam valores inteiros.

Polinômios Ortogonais

- Os polinômios ortogonais $P_j(x_i)$ são facilmente construidos para o caso em que os valores de x_i são equidistantes. Neste caso, os 5 primeiros polinômios ortogonais são dados por (Montogomery at el., 2012):

$$
\begin{aligned}
P_0(x_i) &\equiv 1 \\
P_1(x_i) &= \lambda_1 \left[\frac{x_i - \bar{x}_n}{d}\right] \\
P_2(x_i) &= \lambda_2 \left[\left(\frac{x_i - \bar{x}_n}{d}\right)^2 - \left(\frac{n^2 - 1}{12}\right)\right] \\
P_3(x_i) &= \lambda_3 \left[\left(\frac{x_i - \bar{x}_n}{d}\right)^3 - \left(\frac{x_i - \bar{x}_n}{d}\right)\left(\frac{3n^2 - 7}{20}\right)\right] \\
P_4(x_i) &= \lambda_4 \left[\left(\frac{x_i - \bar{x}_n}{d}\right)^4 - \left(\frac{x_i - \bar{x}_n}{d}\right)^2\left(\frac{3n^2 - 13}{14}\right) + \left(\frac{3(n^2 - 1)(n^2 - 9)}{560}\right)\right].
\end{aligned}
$$

- No caso d representa o incremento (diferença constante entre os valores sucessivos da variável explicativa) e os λ's são constantes escolhidas de forma a fazer com que os polinômios assumam valores inteiros.

Polinômios Ortogonais

- Para valores numéricos e procedimentos para os casos não equidistantes, veja Montgomery et al. (2012), por exemplo.
- No software `R`, pode-se usar a função `poly` implementada no pacote `STATS` para se ajustar um modelo utilizando polinômios ortogonais.
- Os procedimentos inferenciais de testes de hipóteses, ANOVA, IC, etc... seguem o padrão da classe do MRLM.

Polinômios Ortogonais

- Para valores numéricos e procedimentos para os casos não equidistantes, veja Montgomery et al. (2012), por exemplo.
- No software `R`, pode-se usar a função `poly` implementada no pacote `STATS` para se ajustar um modelo utilizando polinômios ortogonais.
- Os procedimentos inferenciais de testes de hipóteses, ANOVA, IC, etc... seguem o padrão da classe do MRLM.

Polinômios Ortogonais

- Para valores numéricos e procedimentos para os casos não equidistantes, veja Montgomery et al. (2012), por exemplo.
- No software `R`, pode-se usar a função `poly` implementada no pacote `STATS` para se ajustar um modelo utilizando polinômios ortogonais.
- Os procedimentos inferenciais de testes de hipóteses, ANOVA, IC, etc... seguem o padrão da classe do MRLM.

Exercício (entregar próxima aula)

Exercício 24: Reproduzir o Exemplo 7.5 (Inventory Data) de Montgomery et al. (2012) em formato de relatório.