CE310 - Modelos de Regressão Linear

Regressão linear múltipla

Cesar Augusto Taconeli

21 de abril, 2025

Introdução

Introdução

- A regressão linear múltipla é uma extensão da regressão linear simples, em que **duas ou mais variáveis explicativas** são incorporadas ao modelo.

Introdução

- A regressão linear múltipla é uma extensão da regressão linear simples, em que **duas ou mais variáveis explicativas** são incorporadas ao modelo.

- Ao considerar conjuntamente o efeito de duas ou mais variáveis explicativas temos condições de avaliar o efeito de uma particular variável explicativa ajustado (controlando) o efeito das demais.

Introdução

- A regressão linear múltipla é uma extensão da regressão linear simples, em que **duas ou mais variáveis explicativas** são incorporadas ao modelo.

- Ao considerar conjuntamente o efeito de duas ou mais variáveis explicativas temos condições de avaliar o efeito de uma particular variável explicativa ajustado (controlando) o efeito das demais.

- A regressão linear múltipla requer maior esforço que a regressão linear simples quanto à especificação do modelo e à avaliação do ajuste.

Definição e propriedades do modelo de regressão linear

Definição e propriedades

- O modelo de regressão linear múltipla é definido da seguinte forma:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_k x_{ik} + \epsilon_i.$$

Definição e propriedades

- O modelo de regressão linear múltipla é definido da seguinte forma:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_k x_{ik} + \epsilon_i.$$

- As seguintes suposições são assumidas:

Definição e propriedades

- O modelo de regressão linear múltipla é definido da seguinte forma:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_k x_{ik} + \epsilon_i.$$

- As seguintes suposições são assumidas:
  - Linearidade: $\mathrm{E}(\epsilon_i) = 0$;

Definição e propriedades

- O modelo de regressão linear múltipla é definido da seguinte forma:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_k x_{ik} + \epsilon_i.$$

- As seguintes suposições são assumidas:
  - Linearidade: $\mathrm{E}(\epsilon_i) = 0$;
  - Variância constante: $\mathrm{Var}(\epsilon_i) = \sigma^2$;

Definição e propriedades

- O modelo de regressão linear múltipla é definido da seguinte forma:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_k x_{ik} + \epsilon_i.$$

- As seguintes suposições são assumidas:
  - Linearidade: $\mathrm{E}(\epsilon_i) = 0$;
  - Variância constante: $\mathrm{Var}(\epsilon_i) = \sigma^2$;
  - Independência: ϵ_i e ϵ_j são independentes para $i \neq j$;

Definição e propriedades

- O modelo de regressão linear múltipla é definido da seguinte forma:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_k x_{ik} + \epsilon_i.$$

- As seguintes suposições são assumidas:
  - Linearidade: $\mathrm{E}(\epsilon_i) = 0$;
  - Variância constante: $\mathrm{Var}(\epsilon_i) = \sigma^2$;
  - Independência: ϵ_i e ϵ_j são independentes para $i \neq j$;
  - x_i é independente de ϵ_i, para todo i;

Definição e propriedades

- O modelo de regressão linear múltipla é definido da seguinte forma:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_k x_{ik} + \epsilon_i.$$

- As seguintes suposições são assumidas:
    - Linearidade: $\mathrm{E}(\epsilon_i) = 0$;
    - Variância constante: $\mathrm{Var}(\epsilon_i) = \sigma^2$;
    - Independência: ϵ_i e ϵ_j são independentes para $i \neq j$;
    - x_i é independente de ϵ_i, para todo i;
    - Normalidade: $\epsilon_i \sim \mathrm{N}(0, \sigma^2)$.

Definição e propriedades

- Como consequências da especificação do modelo, temos:

Definição e propriedades

- Como consequências da especificação do modelo, temos:

1. $\mathrm{E}(y_i|\mathrm{x_i} = (x_{i1}, x_{i2}, ..., x_{ik})') = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_k x_{ik}$;

Definição e propriedades

- Como consequências da especificação do modelo, temos:

1. $\mathrm{E}(y_i|\mathrm{x_i} = (x_{i1}, x_{i2}, ..., x_{ik})') = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_k x_{ik}$;

2. $\mathrm{Var}(y_i|\mathrm{x_i}) = \sigma^2$;

Definição e propriedades

- Como consequências da especificação do modelo, temos:

1. $\mathrm{E}(y_i|\mathrm{x_i} = (x_{i1}, x_{i2}, ..., x_{ik})') = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_k x_{ik}$;

2. $\mathrm{Var}(y_i|\mathrm{x_i}) = \sigma^2$;

3. $y_i|\mathrm{x_i} \sim \mathrm{N}(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_k x_{ik},\ \ \sigma^2)$;

Definição e propriedades

- Como consequências da especificação do modelo, temos:

1. $\mathrm{E}(y_i|\mathrm{x_i} = (x_{i1}, x_{i2}, ..., x_{ik})') = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_k x_{ik}$;
2. $\mathrm{Var}(y_i|\mathrm{x_i}) = \sigma^2$;
3. $y_i|\mathrm{x_i} \sim \mathrm{N}(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_k x_{ik},\ \ \sigma^2)$;
4. Condicional aos respectivos vetores de variáveis explicativas, y_i e y_j são independentes, para todo $i \neq j$.

Definição e propriedades

- Observe que:

$$\frac{\partial \mathrm{E}(y|\mathrm{x})}{\partial x_j} = \beta_j.$$

Definição e propriedades

- Observe que:

$$\frac{\partial \mathrm{E}(y|\mathrm{x})}{\partial x_j} = \beta_j.$$

- Desta forma, β_j representa a alteração esperada (média) na resposta (y) para uma unidade a mais em x_j **quando os valores das demais variáveis** ($x_k \neq x_j$) **são mantidos fixos**.

Definição e propriedades

- Observe que:

$$\frac{\partial \mathrm{E}(y|\mathrm{x})}{\partial x_j} = \beta_j.$$

- Desta forma, β_j representa a alteração esperada (média) na resposta (y) para uma unidade a mais em x_j **quando os valores das demais variáveis** ($x_k \neq x_j$) **são mantidos fixos**.

- Os parâmetros de regressão ($\beta_j' s$) refletem os **efeitos parciais** de cada variável explicativa na resposta.

Definição e propriedades

- Observe que:

$$\frac{\partial \mathrm{E}(y|\mathrm{x})}{\partial x_j} = \beta_j.$$

- Desta forma, β_j representa a alteração esperada (média) na resposta (y) para uma unidade a mais em x_j **quando os valores das demais variáveis** ($x_k \neq x_j$) **são mantidos fixos**.

- Os parâmetros de regressão ($\beta_j's$) refletem os **efeitos parciais** de cada variável explicativa na resposta.

- O intercepto (β_0) é a resposta esperada quando $x_1 = 0$, $x_2 = 0$, $\ldots$, $x_k = 0$, e só tem interpretação válida caso esse ponto pertença ao escopo do problema.

Definição e propriedades

- A interpretação apresentada para os parâmetros $\beta_j's$ somente é válida na ausência de interações (efeitos combinados das covariáveis);

- A interpretação apresentada para os parâmetros $\beta_j's$ somente é válida na ausência de interações (efeitos combinados das covariáveis);

- Para fins de ilustração, considere o seguinte modelo de regressão linear múltipla com termo multiplicativo, correspondente a efeito de interação:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \epsilon.$$

Definição e propriedades

- Nesse caso, por exemplo:

$$\frac{\partial \mathrm{E}(y|\mathrm{x})}{\partial x_1} = \beta_1 + \beta_3 x_2.$$

- Nesse caso, por exemplo:

$$\frac{\partial \mathrm{E}(y|\mathrm{x})}{\partial x_1} = \beta_1 + \beta_3 x_2.$$

- Assim, mantendo x_2 fixa, espera-se uma variação de $\beta_1 + \beta_3 x_2$ em y para cada unidade acrescida em x_1.

Definição e propriedades

- Nesse caso, por exemplo:

$$\frac{\partial \mathrm{E}(y|\mathrm{x})}{\partial x_1} = \beta_1 + \beta_3 x_2.$$

- Assim, mantendo x_2 fixa, espera-se uma variação de $\beta_1 + \beta_3 x_2$ em y para cada unidade acrescida em x_1.

- De forma semelhante, mantendo x_1 fixa, espera-se uma variação de $\beta_2 + \beta_3 x_1$ em y para cada unidade acrescida em x_2.

- Nesse caso, por exemplo:

$$\frac{\partial \mathrm{E}(y|\mathrm{x})}{\partial x_1} = \beta_1 + \beta_3 x_2.$$

- Assim, mantendo x_2 fixa, espera-se uma variação de $\beta_1 + \beta_3 x_2$ em y para cada unidade acrescida em x_1.

- De forma semelhante, mantendo x_1 fixa, espera-se uma variação de $\beta_2 + \beta_3 x_1$ em y para cada unidade acrescida em x_2.

- Assim, a superfície de regressão produzida não é mais plana, pois a taxa de variação de x_1 varia com o valor de x_2 e vice-versa.

Definição e propriedades

- Considere n observações (y_i, x_i), em que $\mathrm{x}_i = (x_{i1}, x_{i2}, ..., x_{ik})'$:

Definição e propriedades

- Considere n observações (y_i, x_i), em que $\mathrm{x}_i = (x_{i1}, x_{i2}, ..., x_{ik})'$:

$$y_1 = \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + ... + \beta_k x_{1k} + \epsilon_1$$

$$y_2 = \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + ... + \beta_k x_{2k} + \epsilon_2$$

$$\vdots$$

$$y_n = \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + ... + \beta_k x_{nk} + \epsilon_n$$

Definição e propriedades

- O modelo de regressão linear múltipla pode ser representado matricialmente por:

$$\mathrm{y} = \mathrm{X}\beta + \epsilon,$$

Definição e propriedades

- O modelo de regressão linear múltipla pode ser representado matricialmente por:

$$\mathrm{y} = \mathrm{X}\beta + \epsilon,$$

em que

$$\mathrm{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \mathrm{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix} \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} \epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

Definição e propriedades

- As suposições e propriedades do modelo de regressão linear múltipla podem ser representados na forma matricial:

Definição e propriedades

- As suposições e propriedades do modelo de regressão linear múltipla podem ser representados na forma matricial:

1. $\mathrm{E}(\epsilon) = 0$;

Definição e propriedades

- As suposições e propriedades do modelo de regressão linear múltipla podem ser representados na forma matricial:

1. $\mathrm{E}(\epsilon) = 0$;
2. $\mathrm{Var}(\epsilon) = \sigma^2\mathrm{I}$;

Definição e propriedades

- As suposições e propriedades do modelo de regressão linear múltipla podem ser representados na forma matricial:

1. $\mathrm{E}(\epsilon) = 0$;
2. $\mathrm{Var}(\epsilon) = \sigma^2 \mathrm{I}$;
3. $\mathrm{E}(\mathrm{y}|\mathrm{X}) = \mathrm{X}\beta$;

Definição e propriedades

- As suposições e propriedades do modelo de regressão linear múltipla podem ser representados na forma matricial:

1. $\mathrm{E}(\epsilon) = 0$;
2. $\mathrm{Var}(\epsilon) = \sigma^2 \mathrm{I}$;
3. $\mathrm{E}(\mathrm{y}|\mathrm{X}) = \mathrm{X}\beta$;
4. $\mathrm{Var}(\mathrm{y}|\mathrm{X}) = \sigma^2 \mathrm{I}$;

Definição e propriedades

- As suposições e propriedades do modelo de regressão linear múltipla podem ser representados na forma matricial:

1. $\mathrm{E}(\epsilon) = 0$;
2. $\mathrm{Var}(\epsilon) = \sigma^2 \mathrm{I}$;
3. $\mathrm{E}(\mathrm{y}|\mathrm{X}) = \mathrm{X}\beta$;
4. $\mathrm{Var}(\mathrm{y}|\mathrm{X}) = \sigma^2 \mathrm{I}$;
5. $\mathrm{y}|\mathrm{X} \sim \mathrm{N}(\mathrm{X}\beta, \sigma^2 \mathrm{I})$, em que I denota a matriz identidade $n \times n$.

Exemplo- Prestígio de profissionais

Exemplo- Prestígio de profissionais

- Nesta aplicação vamos introduzir a análise de regressão linear múltipla por meio de um exemplo com duas variáveis explicativas. Os dados se referem à percepção popular (prestígio) de 45 profissões, em uma pesquisa de opinião conduzida nos EUA. As variáveis são as seguintes:

Exemplo- Prestígio de profissionais

- Nesta aplicação vamos introduzir a análise de regressão linear múltipla por meio de um exemplo com duas variáveis explicativas. Os dados se referem à percepção popular (prestígio) de 45 profissões, em uma pesquisa de opinião conduzida nos EUA. As variáveis são as seguintes:
  - `prestige`: Percentual de respondentes da pesquisa que avaliaram o prestígio de profissionais da área como "bom" ou melhor (variável resposta);

Exemplo- Prestígio de profissionais

- Nesta aplicação vamos introduzir a análise de regressão linear múltipla por meio de um exemplo com duas variáveis explicativas. Os dados se referem à percepção popular (prestígio) de 45 profissões, em uma pesquisa de opinião conduzida nos EUA. As variáveis são as seguintes:
    - `prestige`: Percentual de respondentes da pesquisa que avaliaram o prestígio de profissionais da área como "bom" ou melhor (variável resposta);
    - `income`: Percentual de profissionais da área que ganhavam à época (1950) \$3.500 ou mais;

Exemplo- Prestígio de profissionais

- Nesta aplicação vamos introduzir a análise de regressão linear múltipla por meio de um exemplo com duas variáveis explicativas. Os dados se referem à percepção popular (prestígio) de 45 profissões, em uma pesquisa de opinião conduzida nos EUA. As variáveis são as seguintes:
  - `prestige`: Percentual de respondentes da pesquisa que avaliaram o prestígio de profissionais da área como "bom" ou melhor (variável resposta);
  - `income`: Percentual de profissionais da área que ganhavam à época (1950) $3.500 ou mais;
  - `education`: Percentual de profissionais da área à época que tinham curso superior.

Exemplo- Prestígio de profissionais

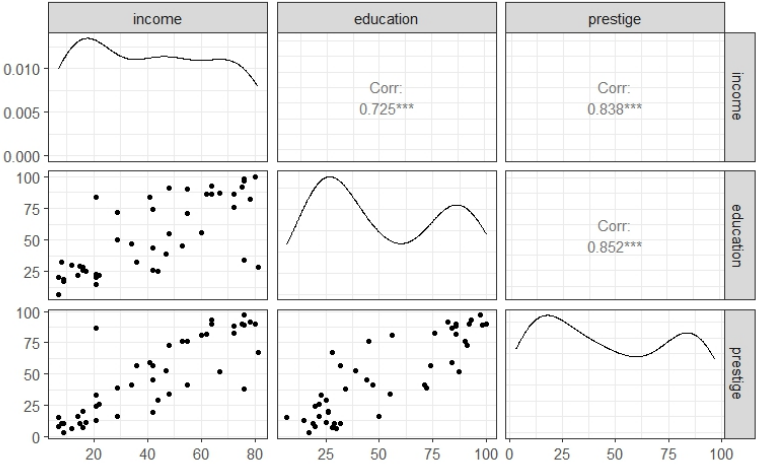

Figura 1: Prestígio de profissionais versus renda e escolaridade

Exemplo- Prestígio de profissionais

- O modelo de regressão linear múltipla definido, inicialmente, tem a seguinte especificação:

$$\texttt{prestige} = \beta_0 + \beta_1 \texttt{income} + \beta_2 \texttt{education} + \epsilon, \quad \epsilon \sim \mathrm{N}(0, \sigma^2),$$

Exemplo- Prestígio de profissionais

- O modelo de regressão linear múltipla definido, inicialmente, tem a seguinte especificação:

$$\texttt{prestige} = \beta_0 + \beta_1\texttt{income} + \beta_2\texttt{education} + \epsilon, \quad \epsilon \sim \mathrm{N}(0, \sigma^2),$$

que pode ser representado de maneira equivalente por:

$$\texttt{prestige}|\boldsymbol{x} \sim \mathrm{N}(\mu_{\boldsymbol{x}}, \sigma^2)$$

$$\mu_{\boldsymbol{x}} = \beta_0 + \beta_1\texttt{income} + \beta_2\texttt{education}$$

Exemplo- Valores de imóveis

Exemplo- Valores de imóveis

- Nesta aplicação, vamos analisar dados de preços de venda de 500 imóveis em determinado município. As variáveis são as seguintes:

Exemplo- Valores de imóveis

- Nesta aplicação, vamos analisar dados de preços de venda de 500 imóveis em determinado município. As variáveis são as seguintes:
  - `valor`: Preço de venda do imóvel (variável resposta, em milhares de dólares);

Exemplo- Valores de imóveis

- Nesta aplicação, vamos analisar dados de preços de venda de 500 imóveis em determinado município. As variáveis são as seguintes:
  - `valor`: Preço de venda do imóvel (variável resposta, em milhares de dólares);
  - `area`: Área do imóvel (m^2);

Exemplo- Valores de imóveis

- Nesta aplicação, vamos analisar dados de preços de venda de 500 imóveis em determinado município. As variáveis são as seguintes:
  - `valor`: Preço de venda do imóvel (variável resposta, em milhares de dólares);
  - `area`: Área do imóvel (m^2);
  - `idade`: Idade do imóvel (anos);

Exemplo- Valores de imóveis

- Nesta aplicação, vamos analisar dados de preços de venda de 500 imóveis em determinado município. As variáveis são as seguintes:
  - `valor`: Preço de venda do imóvel (variável resposta, em milhares de dólares);
  - `area`: Área do imóvel (m^2);
  - `idade`: Idade do imóvel (anos);
  - `distancia`: Distância do imóvel ao marco central do município (em km);

Exemplo- Valores de imóveis

- Nesta aplicação, vamos analisar dados de preços de venda de 500 imóveis em determinado município. As variáveis são as seguintes:
  - `valor`: Preço de venda do imóvel (variável resposta, em milhares de dólares);
  - `area`: Área do imóvel (m^2);
  - `idade`: Idade do imóvel (anos);
  - `distancia:` Distância do imóvel ao marco central do município (em km);
  - `ncomodos`: Número de cômodos;

Exemplo- Valores de imóveis

- Nesta aplicação, vamos analisar dados de preços de venda de 500 imóveis em determinado município. As variáveis são as seguintes:
  - `valor`: Preço de venda do imóvel (variável resposta, em milhares de dólares);
  - `area`: Área do imóvel (m^2);
  - `idade`: Idade do imóvel (anos);
  - `distancia:` Distância do imóvel ao marco central do município (em km);
  - `ncomodos`: Número de cômodos;
  - `pcomerc`: Número de pontos comerciais num raio de um quilômetro.

Exemplo- Valores de imóveis

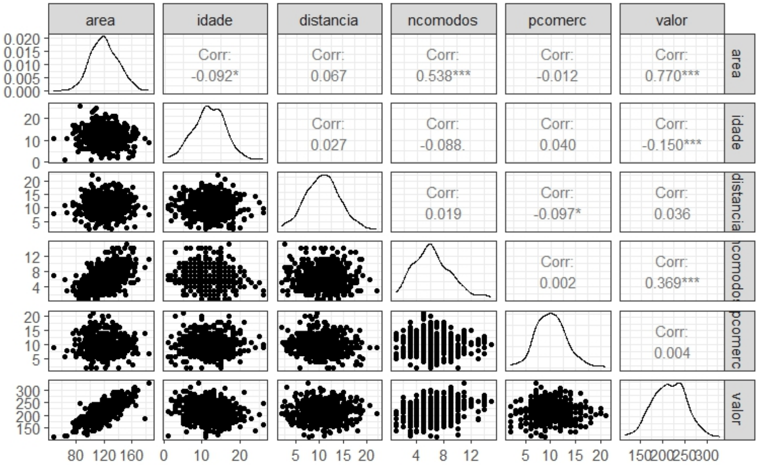

Figura 2: Gráficos de dispersão para os dados de imóveis a venda

Exemplo- Valores de imóveis

- O modelo de regressão linear múltipla definido, inicialmente, tem a seguinte especificação:

$$\texttt{valor} = \beta_0 + \beta_1 \texttt{area} + \beta_2 \texttt{idade} + \beta_3 \texttt{distancia} + \beta_4 \texttt{ncomodos} + \beta_5 \texttt{pcomerc} + \epsilon, \quad \epsilon \sim \mathrm{N}(0, \sigma^2),$$

Exemplo- Valores de imóveis

- O modelo de regressão linear múltipla definido, inicialmente, tem a seguinte especificação:

$$\texttt{valor} = \beta_0 + \beta_1 \texttt{area} + \beta_2 \texttt{idade} + \beta_3 \texttt{distancia} + \beta_4 \texttt{ncomodos} + \beta_5 \texttt{pcomerc} + \epsilon, \quad \epsilon \sim \mathrm{N}(0, \sigma^2),$$

que pode ser representado de maneira equivalente por:

$$\texttt{valor} | \boldsymbol{x} \sim \mathrm{N}(\mu_{\boldsymbol{x}}, \sigma^2)$$

$$\mu_{\boldsymbol{x}} = \beta_0 + \beta_1 \texttt{area} + \beta_2 \texttt{idade} + \beta_3 \texttt{distancia} + \beta_4 \texttt{ncomodos} + \beta_5 \texttt{pcomerc}$$

Estimação

Estimação

- A estimação dos parâmetros do modelo de regressão linear múltipla baseia-se, novamente, no método de mínimos quadrados, mediante determinação de $\beta_0, \beta_1, ..., \beta_k$ que minimizam a soma de quadrados dos erros:

$$S = S(\beta_0, \beta_1, ..., \beta_k) = \sum_{i=1}^{n} \epsilon_i^2 = \sum_{i=1}^{n} (y_i - (\beta_0 + \beta_1 x_{i1} + ... + \beta_k x_{ik}))^2$$

Estimação

- Assim, os estimadores de mínimos quadrados para $\beta_0, \beta_1, ..., \beta_k$ devem satisfazer:

$$\frac{\partial S(\beta)}{\partial \beta} = \begin{bmatrix} \frac{\partial S(\beta)}{\partial \beta_0} \\ \frac{\partial S(\beta)}{\partial \beta_1} \\ \frac{\partial S(\beta)}{\partial \beta_2} \\ \vdots \\ \frac{\partial S(\beta)}{\partial \beta_k} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Estimação

- Derivando parcialmente em relação aos parâmetros de regressão obtemos:

$$\frac{\partial S}{\partial \beta_0}\bigg|_{\hat{\beta}_0,\hat{\beta}_1,...,\hat{\beta}_k} = -2\sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{k}\beta_j x_{ij}\right) = 0$$

e

$$\frac{\partial S}{\partial \beta_j}\bigg|_{\hat{\beta}_0,\hat{\beta}_1,...,\hat{\beta}_k} = -2\sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{k}\beta_j x_{ij}\right) x_{ij} = 0, \quad j = 1, 2, ..., k.$$

Estimação

- Na forma matricial:

$$S(\beta) = \sum_{i=1}^{n} \epsilon_i^2 = \epsilon'\epsilon = (\mathrm{y} - \mathrm{X}\beta)'(\mathrm{y} - \mathrm{X}\beta),$$

Estimação

- Na forma matricial:

$$S(\beta) = \sum_{i=1}^{n} \epsilon_i^2 = \epsilon'\epsilon = (\mathrm{y} - \mathrm{X}\beta)'(\mathrm{y} - \mathrm{X}\beta),$$

de maneira que o vetor $\hat{\beta}$ tal que:

$$\left.\frac{\partial S}{\partial \beta}\right|_{\hat{\beta}} = 0$$

é o estimador de mínimos quadrados de β, dado por:

$$\hat{\beta} = (\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}'\mathrm{y}.$$

Estimação

- Observe que os estimadores de mínimos quadrados somente existem se a matriz $X'X$ tiver inversa;

Estimação

- Observe que os estimadores de mínimos quadrados somente existem se a matriz $X'X$ tiver inversa;

- A condição de existência da inversa de $X'X$ é que as colunas de X sejam linearmente independentes, ou seja, que nenhuma coluna de X seja combinação linear das demais;

Estimação

- Observe que os estimadores de mínimos quadrados somente existem se a matriz $\mathrm{X}'\mathrm{X}$ tiver inversa;

- A condição de existência da inversa de $\mathrm{X}'\mathrm{X}$ é que as colunas de X sejam linearmente independentes, ou seja, que nenhuma coluna de X seja combinação linear das demais;

- O valor ajustado para um vetor $\mathrm{x}' = (1, x_{01}, x_{02}, ..., x_{0k})$ é:

$$\hat{y} = \mathrm{x}'\hat{\beta} = \hat{\beta}_0 + \hat{\beta}_1 x_{01} + ... + \hat{\beta}_k x_{0k}$$

Estimação

- O vetor de valores ajustados para os dados usados no ajuste, $\hat{\mathrm{y}} = (\hat{y}_1, \hat{y}_2, ..., \hat{y}_n)$, é dado por:

$$\hat{\mathrm{y}} = \mathrm{X}\hat{\beta} = \mathrm{X}(\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}'\mathrm{y} = \mathrm{H}\mathrm{y}.$$

Estimação

- O vetor de valores ajustados para os dados usados no ajuste, $\hat{\mathrm{y}} = (\hat{y}_1, \hat{y}_2, ..., \hat{y}_n)$, é dado por:

$$\hat{\mathrm{y}} = \mathrm{X}\hat{\beta} = \mathrm{X}(\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}'\mathrm{y} = \mathrm{H}\mathrm{y}.$$

- A matriz $\mathrm{H} = \mathrm{X}(\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}'$, de dimensão $n \times n$, é chamada matriz chapéu (*hat matrix*) e mapeia o vetor de valores observados no vetor de valores ajustados.

Estimação

- O vetor de valores ajustados para os dados usados no ajuste, $\hat{\mathrm{y}} = (\hat{y}_1, \hat{y}_2, ..., \hat{y}_n)$, é dado por:

$$\hat{\mathrm{y}} = \mathrm{X}\hat{\beta} = \mathrm{X}(\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}'\mathrm{y} = \mathrm{H}\mathrm{y}.$$

- A matriz $\mathrm{H} = \mathrm{X}(\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}'$, de dimensão $n \times n$, é chamada matriz chapéu (*hat matrix*) e mapeia o vetor de valores observados no vetor de valores ajustados.

- O vetor de resíduos $\mathrm{r} = (r_1, r_2, ..., r_n)$ fica definido, em notação matricial, por:

$$\mathrm{r} = \mathrm{y} - \hat{\mathrm{y}},$$

em que $r_i = y_i - \hat{y}_i$, é o resíduo para a i-ésima observação, $i = 1, 2, ..., n$.

Exemplo- Prestígio das profissões

- Vamos determinar as estimativas de mínimos quadrados para os $\beta's$. O vetor resposta $(\boldsymbol{y})$ e a matriz do modelo $(\boldsymbol{X})$ são dados por:

$$\boldsymbol{y} = \begin{bmatrix} 82 \\ 83 \\ 90 \\ \vdots \\ 10 \end{bmatrix} \qquad \boldsymbol{X} = \begin{bmatrix} 1 & 62 & 86 \\ 1 & 72 & 76 \\ 1 & 75 & 92 \\ \vdots & \vdots & \vdots \\ 1 & 8 & 32 \end{bmatrix}$$

Exemplo- Prestígio das profissões

- Desta forma:

$$(\boldsymbol{X'X})^{-1} = \begin{bmatrix} 45 & 1884 & 2365 \\ 1884 & 105148 & 122197 \\ 2365 & 122197 & 163265 \end{bmatrix}^{-1} \qquad (\boldsymbol{X'y}) = \begin{bmatrix} 2146 \\ 118229 \\ 147936 \end{bmatrix}$$

Exemplo- Prestígio das profissões

- Desta forma:

$$(\boldsymbol{X'X})^{-1} = \begin{bmatrix} 45 & 1884 & 2365 \\ 1884 & 105148 & 122197 \\ 2365 & 122197 & 163265 \end{bmatrix}^{-1} \quad (\boldsymbol{X'y}) = \begin{bmatrix} 2146 \\ 118229 \\ 147936 \end{bmatrix}$$

- Logo:

$$\hat{\beta} = (\boldsymbol{X'X})^{-1}(\boldsymbol{X'y}) = \begin{bmatrix} 45 & 1884 & 2365 \\ 1884 & 105148 & 122197 \\ 2365 & 122197 & 163265 \end{bmatrix}^{-1} \times \begin{bmatrix} 2146 \\ 118229 \\ 147936 \end{bmatrix} = \begin{bmatrix} -6.065 \\ 0.599 \\ 0.546 \end{bmatrix}$$

Exemplo- Prestígio das profissões

- De forma que:

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} -6.065 \\ 0.599 \\ 0.546 \end{bmatrix}$$

Exemplo- Prestígio das profissões

- De forma que:

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} -6.065 \\ 0.599 \\ 0.546 \end{bmatrix}$$

- Assim, o modelo ajustado fica dado por:

Exemplo- Prestígio das profissões

- De forma que:

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} -6.065 \\ 0.599 \\ 0.546 \end{bmatrix}$$

- Assim, o modelo ajustado fica dado por:

$$\widehat{\texttt{prestige}} = -6.065 + 0.599 \times \texttt{income} + 0.546 \times \texttt{education}$$

Exemplo- Prestígio das profissões

- Algumas interpretações:

Exemplo- Prestígio das profissões

- Algumas interpretações:
  - Estima-se, em média, um acréscimo de 0.599 pontos percentuais (p.p.) em `prestige` para cada 1 p.p. a mais em `income`, mantendo fixo o percentual em `education`;

Exemplo- Prestígio das profissões

- Algumas interpretações:
  - Estima-se, em média, um acréscimo de 0.599 pontos percentuais (p.p.) em `prestige` para cada 1 p.p. a mais em `income`, mantendo fixo o percentual em `education`;
  - Estima-se, em média, um acréscimo de 0.546 p.p. em `prestige` para cada 1 p.p. a mais em `education`, mantendo fixo o percentual em `income`;

Exemplo- Prestígio das profissões

- Algumas interpretações:
  - Estima-se, em média, um acréscimo de 0.599 pontos percentuais (p.p.) em `prestige` para cada 1 p.p. a mais em `income`, mantendo fixo o percentual em `education`;
  - Estima-se, em média, um acréscimo de 0.546 p.p. em `prestige` para cada 1 p.p. a mais em `education`, mantendo fixo o percentual em `income`;
  - Estima-se, em média, um acréscimo de $10 \times 0.599 = 5.99$ p.p. em `prestige` para cada 10 p.p. a mais em `income`, mantendo fixo o percentual em `education`;

Exemplo- Prestígio das profissões

- Algumas interpretações:
  - Estima-se, em média, um acréscimo de 0.599 pontos percentuais (p.p.) em `prestige` para cada 1 p.p. a mais em `income`, mantendo fixo o percentual em `education`;
  - Estima-se, em média, um acréscimo de 0.546 p.p. em `prestige` para cada 1 p.p. a mais em `education`, mantendo fixo o percentual em `income`;
  - Estima-se, em média, um acréscimo de $10 \times 0.599 = 5.99$ p.p. em `prestige` para cada 10 p.p. a mais em `income`, mantendo fixo o percentual em `education`;
  - Estima-se, em média, um acréscimo de $5 \times 0.546 = 2.73$ p.p. em `prestige` para cada 5 p.p. a mais em `education`, mantendo fixo o percentual em `income`.

Exemplo- Prestígio das profissões

- Vamos calcular os valores ajustados pelo modelo para os profissionais contadores (`accountant`) e pilotos (`pilot`), disponíveis nas duas primeiras linhas da base.

Exemplo- Prestígio das profissões

- Vamos calcular os valores ajustados pelo modelo para os profissionais contadores (`accountant`) e pilotos (`pilot`), disponíveis nas duas primeiras linhas da base.
- Para `accountant`, `income` $= 62$ e `education` $= 86$:

$$\widehat{\texttt{prestige}}_{\text{accountant}} = -6.065 + 0.599 \times 62 + 0.546 \times 86 = 77.99\%$$

Exemplo- Prestígio das profissões

- Vamos calcular os valores ajustados pelo modelo para os profissionais contadores (`accountant`) e pilotos (`pilot`), disponíveis nas duas primeiras linhas da base.
- Para `accountant`, `income` $= 62$ e `education` $= 86$:

$$\widehat{\texttt{prestige}}_{\text{accountant}} = -6.065 + 0.599 \times 62 + 0.546 \times 86 = 77.99\%$$

- Para `pilot`, `income` $= 72$ e `education` $= 76$:

$$\widehat{\texttt{prestige}}_{\text{pilot}} = -6.065 + 0.599 \times 72 + 0.546 \times 76 = 78.53\%$$

Exemplo- Prestígio das profissões

- Os resíduos calculados para os profissionais contadores (`accountant`) e pilotos (`pilot`) são:

Exemplo- Prestígio das profissões

- Os resíduos calculados para os profissionais contadores (`accountant`) e pilotos (`pilot`) são:

$$\text{resíduo}_{\text{accountant}} = \texttt{prestige}_{\text{accountant}} - \widehat{\texttt{prestige}}_{\text{accountant}} = 82 - 77.99 = 4.01\%$$

Exemplo- Prestígio das profissões

- Os resíduos calculados para os profissionais contadores (`accountant`) e pilotos (`pilot`) são:

$$\text{resíduo}_{\text{accountant}} = \texttt{prestige}_{\text{accountant}} - \widehat{\texttt{prestige}}_{\text{accountant}} = 82 - 77.99 = 4.01\%$$

$$\text{resíduo}_{\text{pilot}} = \texttt{prestige}_{\text{pilot}} - \widehat{\texttt{prestige}}_{\text{pilot}} = 83 - 78.53 = 4.47\%$$

Exemplo- Prestígio das profissões

- O vetor de valores ajustados é dado por:

$$\hat{\boldsymbol{y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}} = \begin{bmatrix} 1 & 62 & 86 \\ 1 & 72 & 76 \\ 1 & 75 & 92 \\ \vdots & \vdots & \vdots \\ 1 & 8 & 32 \end{bmatrix} \times \begin{bmatrix} -6.065 \\ 0.599 \\ 0.546 \end{bmatrix} = \begin{bmatrix} 77.99 \\ 78.53 \\ 89.06 \\ \vdots \\ 16.19 \end{bmatrix}$$

Exemplo- Prestígio das profissões

- Já o vetor de resíduos é dado por:

$$\boldsymbol{r} = \boldsymbol{y} - \hat{\boldsymbol{y}} = \begin{bmatrix} 82 \\ 83 \\ 90 \\ \vdots \\ 10 \end{bmatrix} - \begin{bmatrix} 77.99 \\ 78.53 \\ 89.06 \\ \vdots \\ 16.19 \end{bmatrix} = \begin{bmatrix} 4.01 \\ 4.47 \\ 0.94 \\ \vdots \\ -6.19 \end{bmatrix}$$

Exemplo- Valores de imóveis

- O modelo de regressão linear múltipla ajustado tem a seguinte expressão:

$$\widehat{\texttt{valor}} = 48.56 + 1.49 \times \texttt{area} - 0.71 \times \texttt{idade} - 0.14 \times \texttt{distancia} - 0.92 \times \texttt{ncomodos} + 0.18 \times \texttt{pcomerc}$$

Exemplo- Valores de imóveis

- Algumas interpretações:

Exemplo- Valores de imóveis

- Algumas interpretações:
  - Estima-se um aumento médio de 1490 dólares (1.49×1000) no preço de venda para cada m^2 a mais de área, mantendo fixos os valores das demais variáveis;

Exemplo- Valores de imóveis

- Algumas interpretações:
  - Estima-se um aumento médio de 1490 dólares (1.49×1000) no preço de venda para cada m^2 a mais de área, mantendo fixos os valores das demais variáveis;
  - Estima-se uma redução média de 710 dólares (-0.71×1000) no preço de venda para cada ano a mais de idade da construção, mantendo fixos os valores das demais variáveis;

Exemplo- Valores de imóveis

- Algumas interpretações:
  - Estima-se um aumento médio de 1490 dólares (1.49×1000) no preço de venda para cada m^2 a mais de área, mantendo fixos os valores das demais variáveis;
  - Estima-se uma redução média de 710 dólares (-0.71×1000) no preço de venda para cada ano a mais de idade da construção, mantendo fixos os valores das demais variáveis;
  - Estima-se uma redução média de 140 dólares (-0.14×1000) no preço de venda para cada quilômetro a mais de distância ao marco zero, mantendo fixos os valores das demais variáveis;

Exemplo- Valores de imóveis

- Algumas interpretações:
  - Estima-se um aumento médio de 1490 dólares (1.49×1000) no preço de venda para cada m^2 a mais de área, mantendo fixos os valores das demais variáveis;
  - Estima-se uma redução média de 710 dólares (-0.71×1000) no preço de venda para cada ano a mais de idade da construção, mantendo fixos os valores das demais variáveis;
  - Estima-se uma redução média de 140 dólares (-0.14×1000) no preço de venda para cada quilômetro a mais de distância ao marco zero, mantendo fixos os valores das demais variáveis;
  - Estima-se uma redução média de 920 dólares (-0.92×1000) no preço de venda para cada cômodo a mais, mantendo fixos os valores das demais variáveis;

Exemplo- Valores de imóveis

- Algumas interpretações:
  - Estima-se um aumento médio de 1490 dólares (1.49×1000) no preço de venda para cada m^2 a mais de área, mantendo fixos os valores das demais variáveis;
  - Estima-se uma redução média de 710 dólares (-0.71×1000) no preço de venda para cada ano a mais de idade da construção, mantendo fixos os valores das demais variáveis;
  - Estima-se uma redução média de 140 dólares (-0.14×1000) no preço de venda para cada quilômetro a mais de distância ao marco zero, mantendo fixos os valores das demais variáveis;
  - Estima-se uma redução média de 920 dólares (-0.92×1000) no preço de venda para cada cômodo a mais, mantendo fixos os valores das demais variáveis;
  - Estima-se um acréscimo médio de 180 dólares (0.18×1000) no preço de venda para cada ponto comercial a mais no raio de um quilômetro, mantendo fixos os valores das demais variáveis.

Exemplo- Valores de imóveis

- Vamos calcular os valores de venda ajustados pelo modelo para os dois primeiros imóveis da base.

Exemplo- Valores de imóveis

- Vamos calcular os valores de venda ajustados pelo modelo para os dois primeiros imóveis da base.

- Para o primeiro imóvel da base, $\texttt{area} = 100.8$, $\texttt{idade} = 15$, $\texttt{distancia} = 5.8$, $\texttt{ncomodos} = 3$ e $\texttt{pcomerc} = 18$:

$$\widehat{\texttt{valor}}_1 = 48.56 + 1.49 \times 100.8 - 0.71 \times 15 - 0.14 \times 5.8 - 0.92 \times 3 + 0.18 \times 18 = 188.398 (\times 1000)$$

Exemplo- Valores de imóveis

- Vamos calcular os valores de venda ajustados pelo modelo para os dois primeiros imóveis da base.

- Para o primeiro imóvel da base, $\texttt{area} = 100.8$, $\texttt{idade} = 15$, $\texttt{distancia} = 5.8$, $\texttt{ncomodos} = 3$ e $\texttt{pcomerc} = 18$:

$$\widehat{\texttt{valor}}_1 = 48.56 + 1.49 \times 100.8 - 0.71 \times 15 - 0.14 \times 5.8 - 0.92 \times 3 + 0.18 \times 18 = 188.398 (\times 1000)$$

- Para o segundo imóvel da base, $\texttt{area} = 114.4$, $\texttt{idade} = 10$, $\texttt{distancia} = 8.7$, $\texttt{ncomodos} = 6$ e $\texttt{pcomerc} = 13$:

$$\widehat{\texttt{valor}}_2 = 48.56 + 1.49 \times 114.4 - 0.71 \times 10 - 0.14 \times 8.7 - 0.92 \times 6 + 0.18 \times 13 = 208.254 (\times 1000)$$

- Os resíduos calculados para os preços de venda dos dois primeiros imóveis da base são:

Exemplo- Valores de imóveis

- Os resíduos calculados para os preços de venda dos dois primeiros imóveis da base são:

$$\text{resíduo}_1 = \texttt{valor}_1 - \widehat{\texttt{valor}}_1 = 190.232 - 188.398 = 1.834(\times 1000)$$

Exemplo- Valores de imóveis

- Os resíduos calculados para os preços de venda dos dois primeiros imóveis da base são:

$$\text{resíduo}_1 = \texttt{valor}_1 - \widehat{\texttt{valor}}_1 = 190.232 - 188.398 = 1.834(\times 1000)$$

$$\text{resíduo}_2 = \texttt{valor}_2 - \widehat{\texttt{valor}}_2 = 204.368 - 208.254 = -3.886(\times 1000)$$

Estimação

- Propriedades dos estimadores:

Estimação

- Propriedades dos estimadores:

1. $E(\hat{\beta}) = \beta$ ($\hat{\beta}$ é um estimador não viciado de β);

Estimação

- Propriedades dos estimadores:

i. $\mathrm{E}(\hat{\beta}) = \beta$ ($\hat{\beta}$ é um estimador não viciado de β);

ii. $\mathrm{Var}(\hat{\beta}) = \sigma^2(\mathrm{X}'\mathrm{X})^{-1}$;

Estimação

- Propriedades dos estimadores:

i. $\mathrm{E}(\hat{\beta}) = \beta$ ($\hat{\beta}$ é um estimador não viciado de β);

ii. $\mathrm{Var}(\hat{\beta}) = \sigma^2(\mathrm{X}'\mathrm{X})^{-1}$;

iii. $\hat{\beta}$ é o estimador linear não viciado mais eficiente para β (teorema de Gauss Markov);

Estimação

- Propriedades dos estimadores:

i. $\text{E}(\hat{\beta}) = \beta$ ($\hat{\beta}$ é um estimador não viciado de β);

ii. $\text{Var}(\hat{\beta}) = \sigma^2(\text{X}'\text{X})^{-1}$;

iii. $\hat{\beta}$ é o estimador linear não viciado mais eficiente para β (teorema de Gauss Markov);

iv. Sob a suposição de que os erros têm distribuição normal os estimadores de mínimos quadrados equivalem aos de máxima verossimilhança.

Estimação

- Um estimador não viciado para σ^2, baseado na soma de quadrados de resíduos, é dado por:

$$\hat{\sigma}^2 = \mathrm{QM}_{\mathrm{Res}} = \frac{\mathrm{SQ}_{\mathrm{Res}}}{n-p} = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-p},$$

em que $p = k + 1$ é o número de parâmetros do modelo.

Exemplo- Prestígio das profissões

- Vamos calcular a matriz de covariâncias para $\hat{\boldsymbol{\beta}}$. Primeiramente, vamos estimar $\hat{\sigma}^2$:

Exemplo- Prestígio das profissões

- Vamos calcular a matriz de covariâncias para $\hat{\boldsymbol{\beta}}$. Primeiramente, vamos estimar $\hat{\sigma}^2$:

$$\hat{\sigma}^2 = \text{QM}_{\text{Res}} = \frac{\text{SQ}_{\text{Res}}}{n-p} =$$

Exemplo- Prestígio das profissões

- Vamos calcular a matriz de covariâncias para $\hat{\boldsymbol{\beta}}$. Primeiramente, vamos estimar $\hat{\sigma}^2$:

$$\hat{\sigma}^2 = \text{QM}_{\text{Res}} = \frac{\text{SQ}_{\text{Res}}}{n-p} =$$

$$\frac{4.01^2 + 4.47^2 + 0.94^2 + ... + (-6.19)^2}{45-3} = \frac{7507}{42} = 178.7$$

Exemplo- Prestígio das profissões

- Desta forma:

Exemplo- Prestígio das profissões

- Desta forma:

$$\widehat{\text{Var}}(\hat{\boldsymbol{\beta}}) = \hat{\sigma}^2(\boldsymbol{X}'\boldsymbol{X})^{-1} = 178.7 \begin{bmatrix} 45 & 1884 & 2365 \\ 1884 & 105148 & 122197 \\ 2365 & 122197 & 163265 \end{bmatrix}^{-1} =$$

Exemplo- Prestígio das profissões

- Desta forma:

$$\widehat{\mathrm{Var}}(\hat{\boldsymbol{\beta}}) = \hat{\sigma}^2(\boldsymbol{X'X})^{-1} = 178.7 \begin{bmatrix} 45 & 1884 & 2365 \\ 1884 & 105148 & 122197 \\ 2365 & 122197 & 163265 \end{bmatrix}^{-1} =$$

$$\begin{bmatrix} 18.247 & -0.152 & -0.151 \\ -0.152 & 0.014 & -0.008 \\ -0.151 & -0.008 & 0.010 \end{bmatrix}$$

Exemplo- Prestígio das profissões

- Assim, $\widehat{\text{Var}}(\hat{\beta}_0) = 18.247$, $\widehat{\text{Var}}(\hat{\beta}_1) = 0.014$ e $\widehat{\text{Var}}(\hat{\beta}_2) = 0.010$, de forma que os erros padrões são dados por:

Exemplo- Prestígio das profissões

- Assim, $\widehat{\text{Var}}(\hat{\beta}_0) = 18.247$, $\widehat{\text{Var}}(\hat{\beta}_1) = 0.014$ e $\widehat{\text{Var}}(\hat{\beta}_2) = 0.010$, de forma que os erros padrões são dados por:

$$\text{EP}(\hat{\beta}_0) = \sqrt{18.247} = 4.272$$

Exemplo- Prestígio das profissões

- Assim, $\widehat{\text{Var}}(\hat{\beta}_0) = 18.247$, $\widehat{\text{Var}}(\hat{\beta}_1) = 0.014$ e $\widehat{\text{Var}}(\hat{\beta}_2) = 0.010$, de forma que os erros padrões são dados por:

$$\text{EP}(\hat{\beta}_0) = \sqrt{18.247} = 4.272$$

$$\text{EP}(\hat{\beta}_1) = \sqrt{0.014} = 0.118$$

Exemplo- Prestígio das profissões

- Assim, $\widehat{\text{Var}}(\hat{\beta}_0) = 18.247$, $\widehat{\text{Var}}(\hat{\beta}_1) = 0.014$ e $\widehat{\text{Var}}(\hat{\beta}_2) = 0.010$, de forma que os erros padrões são dados por:

$$\text{EP}(\hat{\beta}_0) = \sqrt{18.247} = 4.272$$

$$\text{EP}(\hat{\beta}_1) = \sqrt{0.014} = 0.118$$

$$\text{EP}(\hat{\beta}_2) = \sqrt{0.001} = 0.010$$

Testes de hipóteses e intervalos de confiança

Testes de hipóteses e intervalos de confiança

- Assim como no caso de RLS, também na RLM a inferência sobre os parâmetros do modelo é um ponto importante, que permitirá:

Testes de hipóteses e intervalos de confiança

- Assim como no caso de RLS, também na RLM a inferência sobre os parâmetros do modelo é um ponto importante, que permitirá:

1. Checar a significância do modelo ajustado;

Testes de hipóteses e intervalos de confiança

- Assim como no caso de RLS, também na RLM a inferência sobre os parâmetros do modelo é um ponto importante, que permitirá:

1. Checar a significância do modelo ajustado;
2. Identificar quais variáveis explicativas são relevantes na análise;

Testes de hipóteses e intervalos de confiança

- Assim como no caso de RLS, também na RLM a inferência sobre os parâmetros do modelo é um ponto importante, que permitirá:

1. Checar a significância do modelo ajustado;
2. Identificar quais variáveis explicativas são relevantes na análise;
3. Avaliar o erro de estimativas e das predições geradas pelo modelo ajustado.

Testes de hipóteses e intervalos de confiança

- Assim como no caso de RLS, também na RLM a inferência sobre os parâmetros do modelo é um ponto importante, que permitirá:

1. Checar a significância do modelo ajustado;
2. Identificar quais variáveis explicativas são relevantes na análise;
3. Avaliar o erro de estimativas e das predições geradas pelo modelo ajustado.

- Deste ponto em diante assumiremos todas as suposições especificadas para os erros, inclusive a de normalidade.

Testes de hipóteses e intervalos de confiança

- Primeiramente vamos considerar TH's e IC's para parâmetros individuais do modelo.

Testes de hipóteses e intervalos de confiança

- Primeiramente vamos considerar TH's e IC's para parâmetros individuais do modelo.
- Suponha que se deseja testar a significância de x_j no modelo. Partimos do seguinte par de hipóteses:

$$H_0 : \beta_j = 0 \quad vs \quad H_1 : \beta_j \neq 0.$$

Testes de hipóteses e intervalos de confiança

- Primeiramente vamos considerar TH's e IC's para parâmetros individuais do modelo.
- Suponha que se deseja testar a significância de x_j no modelo. Partimos do seguinte par de hipóteses:

$$H_0 : \beta_j = 0 \quad vs \quad H_1 : \beta_j \neq 0.$$

- A estatística do teste é dada por:

$$t = \frac{\hat{\beta}_j}{\text{ep}(\hat{\beta}_j)},$$

em que $ep(\hat{\beta}_j) = \sqrt{\hat{\sigma}^2 (\mathrm{X}'\mathrm{X})^{-1}_{jj}}$, sendo $(\mathrm{X}'\mathrm{X})^{-1}_{jj}$ o $j-ésimo$ termo da diagonal de $(\mathrm{X}'\mathrm{X})^{-1}$ e $\hat{\sigma}^2 = \mathrm{QM}_{\mathrm{Res}}$.

Testes de hipóteses e intervalos de confiança

- Sob a hipótese nula a estatística t tem distribuição $t - Student$ com $n - p$ graus de liberdade.

Testes de hipóteses e intervalos de confiança

- Sob a hipótese nula a estatística t tem distribuição $t - Student$ com $n - p$ graus de liberdade.

- Assim, a hipótese H_0 deverá ser rejeitada, para um nível de significância α, se $|t| > |t_{n-p,\alpha/2}|$, em que $t_{n-p,\alpha/2}$ é o quantil $\alpha/2$ da distribuição $t - Student$ com $n - p$ graus de liberdade.

Testes de hipóteses e intervalos de confiança

- No `R`, ao aplicar a função `summary()` a um modelo de regressão linear ajustado (objeto da classe `lm`), dispomos das seguintes informações:

Testes de hipóteses e intervalos de confiança

- No `R`, ao aplicar a função `summary()` a um modelo de regressão linear ajustado (objeto da classe `lm`), dispomos das seguintes informações:
  - `Estimate`: estimativas de mínimos quadrados dos parâmetros ($\beta' s$) do modelo;

Testes de hipóteses e intervalos de confiança

- No `R`, ao aplicar a função `summary()` a um modelo de regressão linear ajustado (objeto da classe `lm`), dispomos das seguintes informações:
  - `Estimate`: estimativas de mínimos quadrados dos parâmetros ($\beta's$) do modelo;
  - `Std. Error`: erros padrões (ep) associados a cada uma das estimativas;

Testes de hipóteses e intervalos de confiança

- No `R`, ao aplicar a função `summary()` a um modelo de regressão linear ajustado (objeto da classe `lm`), dispomos das seguintes informações:
    - `Estimate`: estimativas de mínimos quadrados dos parâmetros ($\beta's$) do modelo;
    - `Std. Error`: erros padrões (ep) associados a cada uma das estimativas;
    - `t value`: razão `Estimate\Std. Error`. É a estatística do teste para o par de hipóteses $H_0 : \beta_j = 0$ *vs* $H_1 : \beta_j \neq 0$;

Testes de hipóteses e intervalos de confiança

- No `R`, ao aplicar a função `summary()` a um modelo de regressão linear ajustado (objeto da classe `lm`), dispomos das seguintes informações:
  - `Estimate`: estimativas de mínimos quadrados dos parâmetros ($\beta's$) do modelo;
  - `Std. Error`: erros padrões (ep) associados a cada uma das estimativas;
  - `t value`: razão `Estimate\Std. Error`. É a estatística do teste para o par de hipóteses $H_0 : \beta_j = 0$ *vs* $H_1 : \beta_j \neq 0$;
  - `Pr(>|t|)`: p-valor dos testes.

Testes de hipóteses e intervalos de confiança

- Assim, se desejarmos tomar conclusões ao nível de 5% de significância, comprovamos o efeito de uma particular variável se `Pr(>|t|)<0.05`. Caso contrário, não temos evidências de sua significância.

Testes de hipóteses e intervalos de confiança

- Assim, se desejarmos tomar conclusões ao nível de 5% de significância, comprovamos o efeito de uma particular variável se `Pr(>|t|)<0.05`. Caso contrário, não temos evidências de sua significância.

- Usando a distribuição t_{n-p} como referência, um intervalo de confiança $100(1-\alpha)\%$ para β_j fica definido por:

$$\text{IC}(\beta_j, 100(1-\alpha)\%) = \hat{\beta}_j \pm t_{n-p,\alpha/2}\text{ep}(\hat{\beta}_j).$$

Testes de hipóteses e intervalos de confiança

- Assim, se desejarmos tomar conclusões ao nível de 5% de significância, comprovamos o efeito de uma particular variável se `Pr(>|t|)<0.05`. Caso contrário, não temos evidências de sua significância.

- Usando a distribuição t_{n-p} como referência, um intervalo de confiança $100(1-\alpha)\%$ para β_j fica definido por:

$$\text{IC}(\beta_j, 100(1-\alpha)\%) = \hat{\beta}_j \pm t_{n-p,\alpha/2}\text{ep}(\hat{\beta}_j).$$

- Para qualquer valor β_{j0} pertencente ao intervalo de confiança não se rejeita, ao nível de significância α, que $\beta_j \neq \beta_{j0}$.

Testes de hipóteses e intervalos de confiança

- Assim, se desejarmos tomar conclusões ao nível de 5% de significância, comprovamos o efeito de uma particular variável se `Pr(>|t|)<0.05`. Caso contrário, não temos evidências de sua significância.

- Usando a distribuição t_{n-p} como referência, um intervalo de confiança $100(1-\alpha)\%$ para β_j fica definido por:

$$\text{IC}(\beta_j, 100(1-\alpha)\%) = \hat{\beta}_j \pm t_{n-p,\alpha/2}\text{ep}(\hat{\beta}_j).$$

- Para qualquer valor β_{j0} pertencente ao intervalo de confiança não se rejeita, ao nível de significância α, que $\beta_j \neq \beta_{j0}$.

- No `R`, intervalos de confianças para os $\beta's$ podem ser obtidos usando a função `confint()`.

Exemplo- Valores de imóveis

- Vamos testar o efeito da idade da construção no preço de venda dos imóveis.

Exemplo- Valores de imóveis

- Vamos testar o efeito da idade da construção no preço de venda dos imóveis.

- Hipóteses:

$$H_0 : \beta_{\text{Idade}} = 0 \quad vs \quad H_1 : \beta_{\text{Idade}} \neq 0$$

Exemplo- Valores de imóveis

- Vamos testar o efeito da idade da construção no preço de venda dos imóveis.

- Hipóteses:

$$H_0 : \beta_{\text{Idade}} = 0 \quad vs \quad H_1 : \beta_{\text{Idade}} \neq 0$$

- Estatística do teste:

$$t = \frac{\hat{\beta}_{\text{Idade}} - 0}{ep(\hat{\beta}_{\text{Idade}})} = -\frac{0.71}{0.25} = -2.85$$

Exemplo- Valores de imóveis

- Regra de decisão: Devemos rejeitar H_0, ao nível de significância de 5%, se $|t| > |t_{500-6,2.5\%}| = 1.965$.

Exemplo- Valores de imóveis

- Regra de decisão: Devemos rejeitar H_0, ao nível de significância de 5%, se $|t| > |t_{500-6,2.5\%}| = 1.965$.

- Decisão: Como $|t| = 2.85 > 1.965$, rejeitamos a hipótese nula ao nível de significância de 5%.

Exemplo- Valores de imóveis

- Regra de decisão: Devemos rejeitar H_0, ao nível de significância de 5%, se $|t| > |t_{500-6,2.5\%}| = 1.965$.

- Decisão: Como $|t| = 2.85 > 1.965$, rejeitamos a hipótese nula ao nível de significância de 5%.

- Conclusão: Podemos concluir que há efeito da idade de construção no preço de venda do imóvel.

Exemplo- Valores de imóveis

- Seja $T \sim t_{494}$. O p-valor do teste é dado por:

$$p = 2 \times P(T > 2.85) = 2 \times 0.0022 = 0.0044$$

Exemplo- Valores de imóveis

- Seja $T \sim t_{494}$. O p-valor do teste é dado por:

$$p = 2 \times P(T > 2.85) = 2 \times 0.0022 = 0.0044$$

- Um intervalo de confiança 95% para β_{Idade} é dado por:

$$\text{IC}(\beta_j, 100(1-\alpha)\%) = \hat{\beta}_j \pm t_{n-p,\alpha/2} ep(\hat{\beta}_j) =$$

$$-0.71 \pm 1.965 \times 0.25 = -0.71 \pm 0.49 = (-1.20; -0.22)$$

- Na análise de variância, em regressão linear múltipla, a variação total (corrigida pela média) é novamente decomposta em duas partes: variação explicada pela regressão e variação residual, tal que:

$$\text{SQ}_{\text{Total}} = \text{SQ}_{\text{Reg}} + \text{SQ}_{\text{Res}}.$$

- Usando notação matricial, as somas de quadrados ficam definidas por:

$$\text{SQ}_{\text{Res}} = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \text{y}'\text{y} - \hat{\beta}'\text{X}'\text{y};$$

$$\text{SQ}_{\text{Reg}} = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 = \hat{\beta}'\text{X}'\text{y} - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n};$$

$$\text{SQ}_{\text{Total}} = \sum_{i=1}^{n}(y_i - \bar{y})^2 = \text{y}'\text{y} - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}.$$

Tabela 1: Quadro de análise de variância para o modelo de RLM

Fonte de variação	Graus de liberdade	Soma de quadrados	Quadrados médios	F
Regressão	$p-1$	$\hat{\beta}'\mathrm{X}'\mathrm{y} - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}$	$\mathrm{QM_{Reg}} = \frac{\mathrm{SQ_{Reg}}}{p-1}$	$F = \frac{\mathrm{QM_{Reg}}}{\mathrm{QM_{Res}}}$
Resíduos	$n-p$	$\mathrm{y}'\mathrm{y} - \hat{\beta}'\mathrm{X}'\mathrm{y}$	$\mathrm{QM_{Res}} = \frac{\mathrm{SQ_{Res}}}{n-p}$	
Total	$n-1$	$\mathrm{y}'\mathrm{y} - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}$		

Análise de variância

Tabela 1: Quadro de análise de variância para o modelo de RLM

Fonte de variação	Graus de liberdade	Soma de quadrados	Quadrados médios	F
Regressão	$p-1$	$\hat{\beta}'\mathrm{X}'\mathrm{y} - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}$	$\mathrm{QM}_{\mathrm{Reg}} = \frac{\mathrm{SQ}_{\mathrm{Reg}}}{p-1}$	$F = \frac{\mathrm{QM}_{\mathrm{Reg}}}{\mathrm{QM}_{\mathrm{Res}}}$
Resíduos	$n-p$	$\mathrm{y}'\mathrm{y} - \hat{\beta}'\mathrm{X}'\mathrm{y}$	$\mathrm{QM}_{\mathrm{Res}} = \frac{\mathrm{SQ}_{\mathrm{Res}}}{n-p}$	
Total	$n-1$	$\mathrm{y}'\mathrm{y} - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}$		

- Vale lembrar que n é o tamanho da amostra e $p = k + 1$ o número de parâmetros do modelo.

- Podemos testar a significância do modelo ajustado com base no seguinte par de hipóteses:

$$H_0 : \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad vs \quad H_1 : \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} \neq \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Análise de variância

- Podemos testar a significância do modelo ajustado com base no seguinte par de hipóteses:

$$H_0: \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad vs \quad H_1: \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} \neq \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

- Sob a hipótese nula (não significância do modelo) a estatística F segue distribuição $F-$Snedecor, com $p-1$ e $n-p$ graus de liberdade.

Análise de variância

- Podemos testar a significância do modelo ajustado com base no seguinte par de hipóteses:

$$H_0 : \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad vs \quad H_1 : \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} \neq \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

- Sob a hipótese nula (não significância do modelo) a estatística F segue distribuição $F-$Snedecor, com $p-1$ e $n-p$ graus de liberdade.

- Assim, fixado um nível de significância α, H_0 deve ser rejeitada se o valor da estatística F for maior que o quantil $1-\alpha$ da distribuição $F_{p-1,n-p}$.

- O coeficiente de determinação, como anteriormente, fica definido por:

$$R^2 = 1 - \frac{\text{SQ}_{\text{Res}}}{\text{SQ}_{\text{Total}}} = \frac{\text{SQ}_{\text{Reg}}}{\text{SQ}_{\text{Total}}},$$

e expressa a proporção da variabilidade original dos dados explicada pelo modelo de regressão ajustado.

Análise de variância

- O coeficiente de determinação, como anteriormente, fica definido por:

$$R^2 = 1 - \frac{\text{SQ}_{\text{Res}}}{\text{SQ}_{\text{Total}}} = \frac{\text{SQ}_{\text{Reg}}}{\text{SQ}_{\text{Total}}},$$

e expressa a proporção da variabilidade original dos dados explicada pelo modelo de regressão ajustado.

- Uma propriedade de R^2 que o torna pouco apropriado para a comparação dos ajustes de diferentes modelos é que ele nunca decresce à medida que incluímos novas variáveis ao modelo.

Análise de variância

- Como alternativa ao R^2 podemos considerar o R^2 ajustado, definido por:

$$R^2_{Aj} = 1 - \frac{\text{SQ}_{\text{Res}}/(n-p)}{\text{SQ}_{\text{Total}}/(n-1)}.$$

Análise de variância

- Como alternativa ao R^2 podemos considerar o R^2 ajustado, definido por:

$$R^2_{Aj} = 1 - \frac{\text{SQ}_{\text{Res}}/(n-p)}{\text{SQ}_{\text{Total}}/(n-1)}.$$

- Como $\text{SQ}_{\text{Total}}/(n-1)$ é fixo, então R^2_{Aj} somente aumentará se houver redução do quadrado médio de resíduos.

Análise de variância

- Como alternativa ao R^2 podemos considerar o R^2 ajustado, definido por:

$$R^2_{Aj} = 1 - \frac{\text{SQ}_{\text{Res}}/(n-p)}{\text{SQ}_{\text{Total}}/(n-1)}.$$

- Como $\text{SQ}_{\text{Total}}/(n-1)$ é fixo, então R^2_{Aj} somente aumentará se houver redução do quadrado médio de resíduos.

- Diferentemente de R^2, R^2_{Aj} penaliza a inclusão de variáveis não importantes ao modelo, permitindo comparar adequadamente modelos com diferentes complexidades (números de variáveis).

- O teste da significância da regressão linear baseia-se no par de hipóteses:

$$H_0 : \begin{bmatrix} \beta_{\text{area}} \\ \beta_{\text{idade}} \\ \beta_{\text{distancia}} \\ \beta_{\text{ncomodos}} \\ \beta_{\text{pcomerc}} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad vs \quad H_1 : \begin{bmatrix} \beta_{\text{area}} \\ \beta_{\text{idade}} \\ \beta_{\text{distancia}} \\ \beta_{\text{ncomodos}} \\ \beta_{\text{pcomerc}} \end{bmatrix} \neq \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Exemplo- Valores de imóveis

$$\text{SQ}_{\text{Res}} = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = (190.232 - 188.398)^2 + (204.368 - 208.254)^2 + ... + (264.089 - 250.125)^2 = 262520$$

Exemplo- Valores de imóveis

$$\text{SQ}_{\text{Res}} = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = (190.232 - 188.398)^2 + (204.368 - 208.254)^2 + ... + (264.089 - 250.125)^2 = 262520$$

$$\text{SQ}_{\text{Total}} = \sum_{i=1}^{n}(y_i - \bar{y}_i)^2 = (190.232 - 214.830)^2 + (204.368 - 214.830)^2 + ... + (264.089 - 214.830)^2 = 660888$$

Exemplo- Valores de imóveis

$$\text{SQ}_{\text{Res}} = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = (190.232 - 188.398)^2 + (204.368 - 208.254)^2 + ... + (264.089 - 250.125)^2 = 262520$$

$$\text{SQ}_{\text{Total}} = \sum_{i=1}^{n}(y_i - \bar{y}_i)^2 = (190.232 - 214.830)^2 + (204.368 - 214.830)^2 + ... + (264.089 - 214.830)^2 = 660888$$

$$\text{SQ}_{\text{Reg}} = \sum_{i=1}^{n}(\hat{y}_i - \bar{y}_i)^2 = (188.398 - 214.830)^2 + (208.254 - 214.830)^2 + ... + (250.125 - 214.830)^2 = 398368$$

Exemplo- Valores de imóveis

Tabela 2: Quadro de análise de variância para o modelo de RLM ajustado para os valores de imóveis

Fonte de variação	Graus de liberdade	Soma de quadrados	Quadrados médios	F
Regressão	5	398368	79673.6	149.93
Resíduos	494	262520	531.417	
Total	499	660888		

Exemplo- Valores de imóveis

- O valor da estatística F para o teste da significância da regressão linear é $F = 149.93$.

Exemplo- Valores de imóveis

- O valor da estatística F para o teste da significância da regressão linear é $F = 149.93$.

- Como $F > F_{5,494}(0.95) = 2.232$, então podemos rejeitar a hipótese nula, e concluímos pela significância da regressão linear.

Exemplo- Valores de imóveis

- O coeficiente de determinação do ajuste é:

Exemplo- Valores de imóveis

- O coeficiente de determinação do ajuste é:

$$R^2 = 1 - \frac{\mathrm{SQ_{Res}}}{\mathrm{SQ_{Total}}} = \frac{\mathrm{SQ_{Reg}}}{\mathrm{SQ_{Total}}} = \frac{398368}{660888} = 0.603$$

Exemplo- Valores de imóveis

- O coeficiente de determinação do ajuste é:

$$R^2 = 1 - \frac{\text{SQ}_{\text{Res}}}{\text{SQ}_{\text{Total}}} = \frac{\text{SQ}_{\text{Reg}}}{\text{SQ}_{\text{Total}}} = \frac{398368}{660888} = 0.603$$

- Desta forma, 60.3% da variação dos preços de venda é explicada pela regressão linear.

Exemplo- Valores de imóveis

- O coeficiente de determinação ajustado é dado por:

Exemplo- Valores de imóveis

- O coeficiente de determinação ajustado é dado por:

$$R^2_{Aj} = 1 - \frac{\text{SQ}_{\text{Res}}/(n-p)}{\text{SQ}_{\text{Total}}/(n-1)} = 1 - \frac{262520/494}{660888/499} = 0.598$$

Exemplo- Valores de imóveis

- O coeficiente de determinação ajustado é dado por:

$$R^2_{Aj} = 1 - \frac{\text{SQ}_{\text{Res}}/(n-p)}{\text{SQ}_{\text{Total}}/(n-1)} = 1 - \frac{262520/494}{660888/499} = 0.598$$

- Observe que neste caso R^2_{Aj} é apenas um pouco menor que o valor de R^2.

Estimação da média e predição de novas observações

Estimação da média e predição de novas observações

- Considere interesse em estimar a resposta média em um ponto $\mathrm{x}_0' = (1, x_{01}, x_{02}, ..., x_{0k})$, ou seja, $\mathrm{E}(y|\mathrm{x}_0)$.

Estimação da média e predição de novas observações

- Considere interesse em estimar a resposta média em um ponto $\mathrm{x}_0' = (1, x_{01}, x_{02}, ..., x_{0k})$, ou seja, $\mathrm{E}(y|\mathrm{x}_0)$.

- A estimativa pontual é dada pelo valor ajustado pelo modelo em x_0:

$$\widehat{\mathrm{E}(y|\mathrm{x}_0)} = \hat{y}_0 = \mathrm{x}_0'\hat{\beta}.$$

Estimação da média e predição de novas observações

- Considere interesse em estimar a resposta média em um ponto $\mathrm{x}_0' = (1, x_{01}, x_{02}, ..., x_{0k})$, ou seja, $\mathrm{E}(y|\mathrm{x}_0)$.

- A estimativa pontual é dada pelo valor ajustado pelo modelo em x_0:

$$\mathrm{E}\widehat{(y|\mathrm{x}_0)} = \hat{y}_0 = \mathrm{x}_0'\hat{\beta}.$$

- O estimador apresentado é não viciado para a real resposta média, com variância:

$$\mathrm{Var}(\mathrm{E}\widehat{(y|\mathrm{x}_0)}) = \sigma^2 \mathrm{x}_0'(\mathrm{X}'\mathrm{X})^{-1}\mathrm{x}_0.$$

- Um intervalo de confiança $100(1-\alpha)\%$ para a resposta média em $\mathrm{x}_0' = (1, x_{01}, x_{02}, ..., x_{0k})$ é dado por:

$$\widehat{\mathrm{E}(y|\mathrm{x}_0)} \pm t_{n-p,\alpha/2}\sqrt{\hat{\sigma}^2 \mathrm{x}_0'(\mathrm{X}'\mathrm{X})^{-1}\mathrm{x}_0},$$

em que $\hat{\sigma}^2 = \mathrm{QM}_{\mathrm{Res}}$.

Estimação da média e predição de novas observações

- Um intervalo de confiança $100(1-\alpha)\%$ para a resposta média em $\mathrm{x}_0' = (1, x_{01}, x_{02}, ..., x_{0k})$ é dado por:

$$\widehat{\mathrm{E}(y|\mathrm{x}_0)} \pm t_{n-p,\alpha/2}\sqrt{\hat{\sigma}^2 \mathrm{x}_0'(\mathrm{X}'\mathrm{X})^{-1}\mathrm{x}_0},$$

em que $\hat{\sigma}^2 = \mathrm{QM}_{\mathrm{Res}}$.

- Considere agora que se deseja **predizer** a resposta em um ponto $\mathrm{x}_0' = (1, x_{01}, x_{02}, ..., x_{0k})$.

Estimação da média e predição de novas observações

- Um intervalo de confiança $100(1-\alpha)\%$ para a resposta média em $\mathrm{x}_0' = (1, x_{01}, x_{02}, ..., x_{0k})$ é dado por:

$$\widehat{\mathrm{E}(y|\mathrm{x}_0)} \pm t_{n-p,\alpha/2}\sqrt{\hat{\sigma}^2\mathrm{x}_0'(\mathrm{X}'\mathrm{X})^{-1}\mathrm{x}_0},$$

em que $\hat{\sigma}^2 = \mathrm{QM}_{\mathrm{Res}}$.

- Considere agora que se deseja **predizer** a resposta em um ponto $\mathrm{x}_0' = (1, x_{01}, x_{02}, ..., x_{0k})$.

- A estimativa pontual, novamente, é dada pelo valor ajustado de y em x_0':

$$\hat{y}_0 = \mathrm{x}_0'\hat{\beta}.$$

Estimação da média e predição de novas observações

- A variância de $\hat{y}_0$ fica dada por:

$$\mathrm{Var}(\hat{y}_0) = \sigma^2 \left(1 + \mathrm{x}_0'(\mathrm{X}'\mathrm{X})^{-1}\mathrm{x}_0\right).$$

Estimação da média e predição de novas observações

- A variância de $\hat{y}_0$ fica dada por:

$$\text{Var}(\hat{y}_0) = \sigma^2 \left(1 + \mathrm{x}_0'(\mathrm{X}'\mathrm{X})^{-1}\mathrm{x}_0\right).$$

- Um intervalo de confiança $100(1-\alpha)\%$ para a predição de uma nova observação em x_0 fica dada por:

$$\hat{y}_0 \pm t_{n-p,\alpha/2}\sqrt{\hat{\sigma}^2\left(1 + \mathrm{x}_0'(\mathrm{X}'\mathrm{X})^{-1}\mathrm{x}_0\right)},$$

em que $\hat{\sigma}^2 = \mathrm{QM}_{\mathrm{Res}}$.

Exemplo- Prestígio das profissões

- Vamos começar estimando o percentual esperado (médio) de prestígio para profissões com `income` $= 50$ e `education` $= 75$.

Exemplo- Prestígio das profissões

- Vamos começar estimando o percentual esperado (médio) de prestígio para profissões com `income` $= 50$ e `education` $= 75$.

- Neste caso, $\mathrm{x}_0' = (1 \;\; 50 \;\; 75)$, de forma que:

$$\mathrm{E}(\widehat{\mathrm{prestige}|\mathrm{x}_0}) = \mathrm{x}_0'\hat{\boldsymbol{\beta}} = -6.065 + 0.599 \times 50 + 0.546 \times 75 = 64.81$$

Exemplo- Prestígio das profissões

- Agora, vamos estimar a variância de $\widehat{\mathrm{E}(\text{prestige}|\mathrm{x}_0)}$:

$$\widehat{\mathrm{Var}}(\widehat{\mathrm{E}(y|\mathrm{x}_0)}) = \hat{\sigma}^2 \mathrm{x}_0'(\mathrm{X}'\mathrm{X})^{-1}\mathrm{x}_0 =$$

$$178.7 \begin{bmatrix} 1 & 50 & 75 \end{bmatrix} \begin{bmatrix} 45 & 1884 & 2365 \\ 1884 & 105148 & 122197 \\ 2365 & 122197 & 163265 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 50 \\ 75 \end{bmatrix} = 6.67$$

Exemplo- Prestígio das profissões

- Logo, o intervalo de confiança (95%) para o percentual esperado (médio) de prestígio tem limites:

$$\widehat{\mathrm{E}(y|\mathrm{x}_0)} \pm t_{n-p,\alpha/2}\sqrt{\hat{\sigma}^2\mathrm{x}_0'(\mathrm{X}'\mathrm{X})^{-1}\mathrm{x}_0} =$$

$$64.81 \pm 2.02\sqrt{6.67} = (59.58; 70.04)$$

Exemplo- Prestígio das profissões

- Agora, vamos predizer o percentual de prestígio para uma profissão não considerada no levantamento, para a qual `income` $= 50$ e `education` $= 75$.

Exemplo- Prestígio das profissões

- Agora, vamos predizer o percentual de prestígio para uma profissão não considerada no levantamento, para a qual `income` $= 50$ e `education` $= 75$.

- Neste caso,

$$\widehat{\text{prestige}}|\text{x}_0 = \text{x}_0'\hat{\boldsymbol{\beta}} = -6.065 + 0.599 \times 50 + 0.546 \times 75 = 64.81,$$

como anteriormente.

Exemplo- Prestígio das profissões

- A variância da predição, no entanto, é dada por:

$$\widehat{\mathrm{Var}}(\widehat{\mathrm{E}(y|\mathrm{x}_0)}) = \hat{\sigma}^2(1 + \mathrm{x}_0'(\mathrm{X}'\mathrm{X})^{-1}\mathrm{x}_0) =$$

$$178.7\left\{1 + \begin{bmatrix}1 & 50 & 75\end{bmatrix}\begin{bmatrix}45 & 1884 & 2365\\1884 & 105148 & 122197\\2365 & 122197 & 163265\end{bmatrix}^{-1}\begin{bmatrix}1\\50\\75\end{bmatrix}\right\} = 185.38$$

Exemplo- Prestígio das profissões

- Logo, o intervalo de confiança (95%) para o percentual predito de prestígio para essa específica profissão tem limites:

$$\widehat{\mathrm{E}(y|\mathrm{x}_0)} \pm t_{n-p,\alpha/2}\sqrt{\hat{\sigma}^2(1+\mathrm{x}_0'(\mathrm{X}'\mathrm{X})^{-1}\mathrm{x}_0)} =$$

$$64.81 \pm 2.02\sqrt{185.38} = (37.3; 92.3)$$

Inferência conjunta para os parâmetros do modelo

Inferência conjunta para os parâmetros do modelo

- Em geral os estimadores dos parâmetros do modelo de RLM são correlacionados (a menos que as correspondentes variáveis sejam ortogonais).

Inferência conjunta para os parâmetros do modelo

- Em geral os estimadores dos parâmetros do modelo de RLM são correlacionados (a menos que as correspondentes variáveis sejam ortogonais).

- Avaliar a significância das variáveis explicativas individualmente e conjuntamente, neste caso, são coisas distintas.

Inferência conjunta para os parâmetros do modelo

- Em geral os estimadores dos parâmetros do modelo de RLM são correlacionados (a menos que as correspondentes variáveis sejam ortogonais).

- Avaliar a significância das variáveis explicativas individualmente e conjuntamente, neste caso, são coisas distintas.

- É comum o interesse em analisar a significância conjunta de dois ou mais parâmetros, como no caso de modelos polinomiais e na inclusão de variáveis com múltiplas categorias.

Inferência conjunta para os parâmetros do modelo

- Considere o modelo de regressão linear múltipla:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + \epsilon.$$

Inferência conjunta para os parâmetros do modelo

- Considere o modelo de regressão linear múltipla:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + \epsilon.$$

- O interesse aqui é testar uma hipótese do tipo:

$$H_0 : \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_q \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad vs \quad H_1 : \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_q \end{bmatrix} \neq \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Inferência conjunta para os parâmetros do modelo

- Considere o modelo de regressão linear múltipla:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + \epsilon.$$

- O interesse aqui é testar uma hipótese do tipo:

$$H_0 : \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_q \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad vs \quad H_1 : \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_q \end{bmatrix} \neq \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

- Por simplicidade de notação, vamos considerar que a hipótese nula contemple os $q < p$ primeiros parâmetros do modelo.

Inferência conjunta para os parâmetros do modelo

- O modelo reduzido, induzido pela hipótese nula (com $p-q$ parâmetros), fica definido por:

$$y = \beta_0 + \beta_{q+1}x_{q+1} + ... + \beta_k x_k + \epsilon$$

Inferência conjunta para os parâmetros do modelo

- O modelo reduzido, induzido pela hipótese nula (com $p-q$ parâmetros), fica definido por:

$$y = \beta_0 + \beta_{q+1}x_{q+1} + ... + \beta_k x_k + \epsilon$$

- Uma vez ajustados o modelo original e o modelo reduzido, sejam SQ_{Res} e SQ_{Res0} as respectivas somas de quadrados de resíduos.

Inferência conjunta para os parâmetros do modelo

- Sob a hipótese nula (correspondente ao modelo restrito), a estatística:

$$F = \frac{(SQ_{\text{Res}_0} - \text{SQ}_{\text{Res}})/q}{\text{SQ}_{\text{Res}}/(n-p)}$$

tem distribuição F-Snedecor com q e $n-p$ graus de liberdade.

- Sob a hipótese nula (correspondente ao modelo restrito), a estatística:

$$F = \frac{(SQ_{\text{Res}_0} - \text{SQ}_{\text{Res}})/q}{\text{SQ}_{\text{Res}}/(n-p)}$$

tem distribuição F-Snedecor com q e $n-p$ graus de liberdade.

- Observe que F baseia-se na variação (acréscimo) da soma de quadrados de resíduos resultante da restrição aplicada aos parâmetros do modelo.

Inferência conjunta para os parâmetros do modelo

- A hipótese H_0 deverá ser rejeitada, ao nível de significância α, se F superar o quantil $1-\alpha$ da distribuição F-Snedecor com q e $n-p$ graus de liberdade.

Inferência conjunta para os parâmetros do modelo

- A hipótese H_0 deverá ser rejeitada, ao nível de significância α, se F superar o quantil $1-\alpha$ da distribuição F-Snedecor com q e $n-p$ graus de liberdade.

- **Ao rejeitar** H_0, temos evidências de que a soma de quadrados de resíduos produzida pelo modelo restrito é significativamente maior que a do modelo original, de forma que **o modelo original é preferível**.

Inferência conjunta para os parâmetros do modelo

- A hipótese H_0 deverá ser rejeitada, ao nível de significância α, se F superar o quantil $1-\alpha$ da distribuição F-Snedecor com q e $n-p$ graus de liberdade.

- **Ao rejeitar** H_0, temos evidências de que a soma de quadrados de resíduos produzida pelo modelo restrito é significativamente maior que a do modelo original, de forma que **o modelo original é preferível**.

- Por outro lado, **ao não rejeitar** H_0, não há evidências de que a soma de quadrados de resíduos produzida pelo modelo restrito seja significativamente maior que a do modelo original, de forma que **o modelo reduzido é preferível**.

Exemplo- Valores de imóveis

- Para fins de ilustração vamos testar o par de hipóteses:

$$H_0 : \begin{pmatrix} \beta_{\text{distancia}} \\ \beta_{\text{pcomerc}} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad vs \quad H_1 : \begin{pmatrix} \beta_{\text{distancia}} \\ \beta_{\text{pcomerc}} \end{pmatrix} \neq \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Exemplo- Valores de imóveis

- Para fins de ilustração vamos testar o par de hipóteses:

$$H_0 : \begin{pmatrix} \beta_{\text{distancia}} \\ \beta_{\text{pcomerc}} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad vs \quad H_1 : \begin{pmatrix} \beta_{\text{distancia}} \\ \beta_{\text{pcomerc}} \end{pmatrix} \neq \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

- Para isso, os seguintes modelos foram ajustados:
  - Modelo irrestrito: `ajuste1 <- lm(valor ~ area + idade + distancia + ncomodos + pcomerc, data = dados)`
  - Modelo restrito: `ajuste3 <- lm(valor ~ area + idade + ncomodos, data = dados)`

Exemplo- Valores de imóveis

- Para testar as hipóteses formuladas, vamos comparar os dois modelos ajustados, usando:

```
anova(ajuste3, ajuste1, test = 'F')
```

Exemplo- Valores de imóveis

- Para testar as hipóteses formuladas, vamos comparar os dois modelos ajustados, usando:

```
anova(ajuste3, ajuste1, test = 'F')
```

Tabela 3: Resultado do teste

ajuste	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
3	496	262817.71				
1	494	262520.04	2	297.67	0.28	0.7558

Exemplo- Valores de imóveis

- A soma de quadrados de resíduos do modelo irrestrito (com todas as variáveis) é de 262520.04, e ela aumenta para 262817.71 sob o modelo restrito (sem os efeitos de `distancia` e `pcomerc`).

- A soma de quadrados de resíduos do modelo irrestrito (com todas as variáveis) é de 262520.04, e ela aumenta para 262817.71 sob o modelo restrito (sem os efeitos de `distancia` e `pcomerc`).

- A diferença nas somas de quadrados de resíduos é de 262817.71-262520.04 = 297.67, associada 496-494=2 graus de liberdade (dois parâmetros a mais no modelo irrestrito).

Exemplo- Valores de imóveis

- A soma de quadrados de resíduos do modelo irrestrito (com todas as variáveis) é de 262520.04, e ela aumenta para 262817.71 sob o modelo restrito (sem os efeitos de `distancia` e `pcomerc`).

- A diferença nas somas de quadrados de resíduos é de 262817.71-262520.04 $=$ 297.67, associada 496-494=2 graus de liberdade (dois parâmetros a mais no modelo irrestrito).

- Com base nessas diferenças, o valor calculado da estatística do teste é $F = 0.28$, produzindo $p = 0.7558$.

Exemplo- Valores de imóveis

- Logo, não temos evidência significativa para rejeitar a hipótese H_0, de que os dois parâmetros são conjuntamente iguais a zero.

Exemplo- Valores de imóveis

- Logo, não temos evidência significativa para rejeitar a hipótese H_0, de que os dois parâmetros são conjuntamente iguais a zero.

- Dada a não rejeição de H_0, as variáveis `pcomerc` e `distancia` poderiam ser removidas conjuntamente do modelo.

Exemplo- Valores de imóveis

- Logo, não temos evidência significativa para rejeitar a hipótese H_0, de que os dois parâmetros são conjuntamente iguais a zero.

- Dada a não rejeição de H_0, as variáveis `pcomerc` e `distancia` poderiam ser removidas conjuntamente do modelo.

- Caso contrário, se a hipótese nula fosse rejeitada, `distancia` e `pcomerc` não poderiam ser excluídas conjuntamente do modelo.

Exemplo- Valores de imóveis

- Logo, não temos evidência significativa para rejeitar a hipótese H_0, de que os dois parâmetros são conjuntamente iguais a zero.

- Dada a não rejeição de H_0, as variáveis `pcomerc` e `distancia` poderiam ser removidas conjuntamente do modelo.

- Caso contrário, se a hipótese nula fosse rejeitada, `distancia` e `pcomerc` não poderiam ser excluídas conjuntamente do modelo.

- Frente à rejeição de H_0, poderiamos proceder com os testes dos efeitos de `distancia` e `pcomerc` um de cada vez.

Regressão com variáveis centradas e escalonadas

Regressão com variáveis centradas e escalonadas

- Numa RLM, a comparação das magnitudes dos $\hat{\beta}_j' s$ nem sempre é possível devido ao impacto das diferentes escalas dos $x_j' s$.

Regressão com variáveis centradas e escalonadas

- Numa RLM, a comparação das magnitudes dos $\hat{\beta}_j's$ nem sempre é possível devido ao impacto das diferentes escalas dos $x_j's$.

- Caso seja desejado que tais estimativas sejam comparáveis, pode-se padronizar cada uma das variáveis de forma que as variáveis resultantes tenham mesma escala.

Regressão com variáveis centradas e escalonadas

- Numa RLM, a comparação das magnitudes dos $\hat{\beta}_j's$ nem sempre é possível devido ao impacto das diferentes escalas dos $x_j's$.

- Caso seja desejado que tais estimativas sejam comparáveis, pode-se padronizar cada uma das variáveis de forma que as variáveis resultantes tenham mesma escala.

- Uma allternativa usual de padronização consiste em centrar e escalonar cada uma das variáveis, aplicando:

$$x_{ij}^* = \frac{x_{ij} - \bar{x}_j}{s_j}, \quad i = 1, 2, ..., n; \; j = 1, 2, ..., k,$$

sendo $\bar{x}_j$ e s_j a média e o desvio padrão amostrais de x_j.

Regressão com variáveis centradas

- Inicialmente, vamos considerar apenas as variáveis centradas na média:

$$y_i = \beta_0^* + \beta_1^*(x_{i1} - \bar{x}_1) + \beta_2^*(x_{i2} - \bar{x}_2) + ... + \beta_k^*(x_{ik} - \bar{x}_k) + \epsilon_i, \quad i = 1, 2, ..., n$$

Regressão com variáveis centradas

- Inicialmente, vamos considerar apenas as variáveis centradas na média:

$$y_i = \beta_0^* + \beta_1^*(x_{i1} - \bar{x}_1) + \beta_2^*(x_{i2} - \bar{x}_2) + ... + \beta_k^*(x_{ik} - \bar{x}_k) + \epsilon_i, \quad i = 1, 2, ..., n$$

- Ao centrar as variáveis em suas médias, isso tem o efeito de deslocar a origem dos eixos para o ponto $\bar{\boldsymbol{x}}' = (\bar{x}_1, \bar{x}_2, ..., \bar{x}_k)$.

Regressão com variáveis centradas

- Inicialmente, vamos considerar apenas as variáveis centradas na média:

$$y_i = \beta_0^* + \beta_1^*(x_{i1} - \bar{x}_1) + \beta_2^*(x_{i2} - \bar{x}_2) + ... + \beta_k^*(x_{ik} - \bar{x}_k) + \epsilon_i, \quad i = 1, 2, ..., n$$

- Ao centrar as variáveis em suas médias, isso tem o efeito de deslocar a origem dos eixos para o ponto $\bar{\boldsymbol{x}}' = (\bar{x}_1, \bar{x}_2, ..., \bar{x}_k)$.

- Esse deslocamento não altera o espalhamento dos pontos, de forma que os parâmetros de inclinação são idênticos aos obtidos sem centrar as variáveis ($\beta_1^* = \beta_1$, $\beta_2^* = \beta_2$,..., $\beta_k^* = \beta_k$).

Regressão com variáveis centradas

- Inicialmente, vamos considerar apenas as variáveis centradas na média:

$$y_i = \beta_0^* + \beta_1^*(x_{i1} - \bar{x}_1) + \beta_2^*(x_{i2} - \bar{x}_2) + ... + \beta_k^*(x_{ik} - \bar{x}_k) + \epsilon_i, \quad i = 1, 2, ..., n$$

- Ao centrar as variáveis em suas médias, isso tem o efeito de deslocar a origem dos eixos para o ponto $\bar{\boldsymbol{x}}' = (\bar{x}_1, \bar{x}_2, ..., \bar{x}_k)$.

- Esse deslocamento não altera o espalhamento dos pontos, de forma que os parâmetros de inclinação são idênticos aos obtidos sem centrar as variáveis ($\beta_1^* = \beta_1$, $\beta_2^* = \beta_2, \dots,$ $\beta_k^* = \beta_k$).

- O intercepto do modelo (β_0^*), no entanto, corresponde à resposta esperada quando cada variável explicativa estiver na média, ou seja, $x_1^* = \bar{x}_1$, $x_2^* = \bar{x}_2, \dots,$ $x_k^* = \bar{x}_k$.

Regressão com variáveis centradas

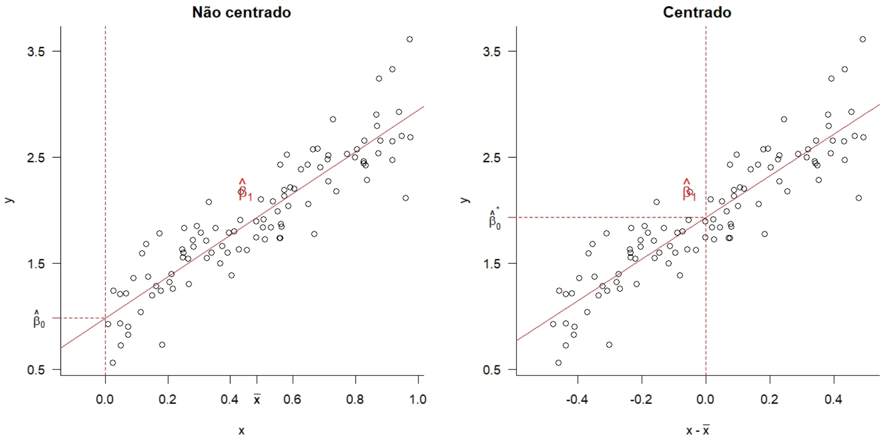

Figura 3: Regressão com variável centrada

Regressão com variáveis centradas e escalonadas

- Usando as variáveis centradas e escalonadas, o modelo de regressão linear múltipla fica definido por:

$$y_i = \beta_0^* + \beta_1^* \left(\frac{x_{i1} - \bar{x}_1}{s_1} \right) + \beta_2^* \left(\frac{x_{i2} - \bar{x}_2}{s_2} \right) + ... + \beta_k^* \left(\frac{x_{ik} - \bar{x}_k}{s_k} \right) + \epsilon_i, \quad i = 1, 2, ..., n$$

Regressão com variáveis centradas e escalonadas

- Usando as variáveis centradas e escalonadas, o modelo de regressão linear múltipla fica definido por:

$$y_i = \beta_0^* + \beta_1^* \left(\frac{x_{i1} - \bar{x}_1}{s_1} \right) + \beta_2^* \left(\frac{x_{i2} - \bar{x}_2}{s_2} \right) + ... + \beta_k^* \left(\frac{x_{ik} - \bar{x}_k}{s_k} \right) + \epsilon_i, \quad i = 1, 2, ..., n$$

- Ao centrar e escalonar as variáveis, o intercepto do modelo novamente é deslocado para o ponto $(x_1 = \bar{x}_1, x_2 = \bar{x}_2, ..., x_k = \bar{x}_k)$.

Regressão com variáveis centradas e escalonadas

- Usando as variáveis centradas e escalonadas, o modelo de regressão linear múltipla fica definido por:

$$y_i = \beta_0^* + \beta_1^* \left(\frac{x_{i1} - \bar{x}_1}{s_1} \right) + \beta_2^* \left(\frac{x_{i2} - \bar{x}_2}{s_2} \right) + ... + \beta_k^* \left(\frac{x_{ik} - \bar{x}_k}{s_k} \right) + \epsilon_i, \quad i = 1, 2, ..., n$$

- Ao centrar e escalonar as variáveis, o intercepto do modelo novamente é deslocado para o ponto $(x_1 = \bar{x}_1, x_2 = \bar{x}_2, ..., x_k = \bar{x}_k)$.

- Novamente, o intercepto do modelo (β_0^*) corresponde à resposta esperada quando cada variável explicativa estiver na média, ou seja, $x_1^* = \bar{x}_1$, $x_2^* = \bar{x}_2, \ldots,$ $x_k^* = \bar{x}_k$.

Regressão com variáveis centradas e escalonadas

- No entanto, desta vez, ao dividir cada variável pelo respectivo desvio padrão, o espalhamento dos pontos (e as inclinações do modelo) ficam alterados.

Regressão com variáveis centradas e escalonadas

- No entanto, desta vez, ao dividir cada variável pelo respectivo desvio padrão, o espalhamento dos pontos (e as inclinações do modelo) ficam alterados.

- Neste caso, β_j^* corresponde à alteração na resposta média para uma alteração de um desvio padrão (s_j) em x_j, mantendo fixos os valores das demais variáveis ($j = 1, 2, ..., k$).

Regressão com variáveis centradas e escalonadas

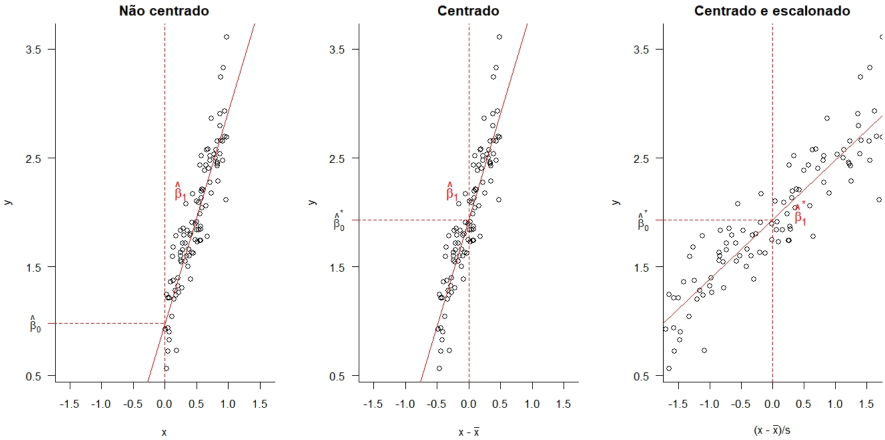

Figura 4: Regressão com variável centrada e escalonada

Exemplo- Prestígio das profissões

- Vamos ajustar e interpretar o modelo de regressão linear múltipla com as variáveis regressoras centradas e escalonadas.

Exemplo- Prestígio das profissões

- Vamos ajustar e interpretar o modelo de regressão linear múltipla com as variáveis regressoras centradas e escalonadas.

- As médias para `income` e `education` são, respectivamente, 41.87 e 52.55 pontos percentuais (p.p.), enquanto os desvios padrões para as duas variáveis são 24.43 e 29.76 p.p.

Exemplo- Prestígio das profissões

- Vamos ajustar e interpretar o modelo de regressão linear múltipla com as variáveis regressoras centradas e escalonadas.

- As médias para `income` e `education` são, respectivamente, 41.87 e 52.55 pontos percentuais (p.p.), enquanto os desvios padrões para as duas variáveis são 24.43 e 29.76 p.p.

- Começando pelo modelo ajustado com as variáveis explicativas apenas centradas (não escalonadas):

$$\widehat{\texttt{prestige}} = 47.69 + 0.599 \times (\texttt{income} - 41.87) + 0.546 \times (\texttt{education} - 52.55)$$

Exemplo- Prestígio das profissões

- Algumas interpretações:

Exemplo- Prestígio das profissões

- Algumas interpretações:
  - Estima-se, em média, percentual de prestígio igual a 47.69% para profissões com `income` e `education` na média, ou seja, `income = 41.87%` e `education = 52.55%`.

Exemplo- Prestígio das profissões

- Algumas interpretações:
  - Estima-se, em média, percentual de prestígio igual a 47.69% para profissões com `income` e `education` na média, ou seja, `income = 41.87%` e `education = 52.55%`.
  - Estima-se, em média, um acréscimo de 0.599 p.p. em prestige para cada 1 p.p. a mais em `income`, mantendo fixo o percentual em `education`;

Exemplo- Prestígio das profissões

- Algumas interpretações:
  - Estima-se, em média, percentual de prestígio igual a 47.69% para profissões com `income` e `education` na média, ou seja, `income = 41.87%` e `education = 52.55%`.
  - Estima-se, em média, um acréscimo de 0.599 p.p. em prestige para cada 1 p.p. a mais em `income`, mantendo fixo o percentual em `education`;
  - Estima-se, em média, um acréscimo de 0.546 p.p. em prestige para cada 1 p.p. a mais em `education`, mantendo fixo o percentual em `income`.

Exemplo- Prestígio das profissões

- O modelo de regressão ajustado para as variáveis centradas e escalonadas é dado por:

$$\widehat{\texttt{prestige}} = 47.69 + 14.63\left(\frac{\texttt{income} - 41.87}{24.43}\right) + 16.24 \times \left(\frac{\texttt{education} - 52.55}{29.76}\right)$$

Exemplo- Prestígio das profissões

- Algumas interpretações:

Exemplo- Prestígio das profissões

- Algumas interpretações:
  - Estima-se, em média, percentual de prestígio igual a 47.69% para profissões com `income` e `education` na média, ou seja, `income = 41.87%` e `education = 52.55%`.

Exemplo- Prestígio das profissões

- Algumas interpretações:
  - Estima-se, em média, percentual de prestígio igual a 47.69% para profissões com `income` e `education` na média, ou seja, `income = 41.87%` e `education = 52.55%`.
  - Estima-se, em média, um acréscimo de 14.63 p.p. em prestige para cada 1 desvio padrão (24.43 p.p.) a mais em `income`, mantendo fixo o percentual em `education`;

Exemplo- Prestígio das profissões

- Algumas interpretações:
  - Estima-se, em média, percentual de prestígio igual a 47.69% para profissões com `income` e `education` na média, ou seja, `income = 41.87%` e `education = 52.55%`.
  - Estima-se, em média, um acréscimo de 14.63 p.p. em prestige para cada 1 desvio padrão (24.43 p.p.) a mais em `income`, mantendo fixo o percentual em `education`;
  - Estima-se, em média, um acréscimo de 16.24 p.p. em prestige para cada 1 desvio padrão (29.76 p.p.) a mais em `education`, mantendo fixo o percentual em `income`.

Ortogonalidade

Ortogonalidade

- A ortogonalidade entre variáveis explicativas é uma propriedade útil em análise de regressão por permitir avaliar o efeito de uma variável independente dos efeitos das demais.

Ortogonalidade

- A ortogonalidade entre variáveis explicativas é uma propriedade útil em análise de regressão por permitir avaliar o efeito de uma variável independente dos efeitos das demais.

- Em geral, ortogonalidade é algo característico de estudos experimentais, em que o desenho do experimento produz variáveis ortogonais, e algo incomum em estudos observacionais.

Ortogonalidade

- A ortogonalidade entre variáveis explicativas é uma propriedade útil em análise de regressão por permitir avaliar o efeito de uma variável independente dos efeitos das demais.

- Em geral, ortogonalidade é algo característico de estudos experimentais, em que o desenho do experimento produz variáveis ortogonais, e algo incomum em estudos observacionais.

- Suponha que a matriz do modelo X possa ser particionada em duas matrizes na forma $\mathrm{X} = [\mathrm{X}_1|\mathrm{X}_2]$ tais que $\mathrm{X'}_1\mathrm{X}_2 = 0$:

$$\mathrm{y} = \mathrm{X}\beta + \epsilon = \mathrm{X}_1\beta_1 + \mathrm{X}_2\beta_2 + \epsilon$$

Ortogonalidade

- Assim:

$$X'X = \begin{pmatrix} X'_1X_1 & X'_1X_2 \\ X'_2X_1 & X'_2X_2 \end{pmatrix} = \begin{pmatrix} X'_1X_1 & 0 \\ 0 & X'_2X_2 \end{pmatrix}$$

Ortogonalidade

- Assim:

$$X'X = \begin{pmatrix} {X'}_1X_1 & {X'}_1X_2 \\ {X'}_2X_1 & {X'}_2X_2 \end{pmatrix} = \begin{pmatrix} {X'}_1X_1 & 0 \\ 0 & {X'}_2X_2 \end{pmatrix}$$

- Os estimadores de mínimos quadrados ficam dados por:

$$\hat{\beta}_1 = (X_1'X_1)^{-1}X_1'y \text{ e } \hat{\beta}_2 = (X_2'X_2)^{-1}X_2'y.$$

Ortogonalidade

- Assim:

$$X'X = \begin{pmatrix} X'_1X_1 & X'_1X_2 \\ X'_2X_1 & X'_2X_2 \end{pmatrix} = \begin{pmatrix} X'_1X_1 & 0 \\ 0 & X'_2X_2 \end{pmatrix}$$

- Os estimadores de mínimos quadrados ficam dados por:

$$\hat{\beta}_1 = (X_1'X_1)^{-1}X_1'y \text{ e } \hat{\beta}_2 = (X_2'X_2)^{-1}X_2'y.$$

- Note que a estimação de β_1 não depende de X_2, da mesma forma que a estimação de β_2 não depende de X_1.

Ortogonalidade

- Assim:

$$X'X = \begin{pmatrix} X'_1X_1 & X'_1X_2 \\ X'_2X_1 & X'_2X_2 \end{pmatrix} = \begin{pmatrix} X'_1X_1 & 0 \\ 0 & X'_2X_2 \end{pmatrix}$$

- Os estimadores de mínimos quadrados ficam dados por:

$$\hat{\beta}_1 = (X_1'X_1)^{-1}X_1'y \text{ e } \hat{\beta}_2 = (X_2'X_2)^{-1}X_2'y.$$

- Note que a estimação de β_1 não depende de X_2, da mesma forma que a estimação de β_2 não depende de X_1.

- Desta forma, o efeito de X_1 será o mesmo, ajustado ou não o efeito de X_2 (e vice-versa).

Exemplo- Odor de produto químico

Exemplo- Odor de produto químico

- Nesta aplicação serão analisados dados de um experimento químico disponíveis na base de dados `odor`, que pode ser acessada na biblioteca `faraway` do `R`.

Exemplo- Odor de produto químico

- Nesta aplicação serão analisados dados de um experimento químico disponíveis na base de dados `odor`, que pode ser acessada na biblioteca `faraway` do `R`.

- As variáveis a serem analisadas são as seguintes:

Exemplo- Odor de produto químico

- Nesta aplicação serão analisados dados de um experimento químico disponíveis na base de dados `odor`, que pode ser acessada na biblioteca `faraway` do `R`.

- As variáveis a serem analisadas são as seguintes:
  - `odor`: escore de odor do produto (variável resposta);

Exemplo- Odor de produto químico

- Nesta aplicação serão analisados dados de um experimento químico disponíveis na base de dados `odor`, que pode ser acessada na biblioteca `faraway` do `R`.

- As variáveis a serem analisadas são as seguintes:
  - `odor`: escore de odor do produto (variável resposta);
  - `temp`: temperatura, codificada como -1, 0 e 1;

Exemplo- Odor de produto químico

- Nesta aplicação serão analisados dados de um experimento químico disponíveis na base de dados `odor`, que pode ser acessada na biblioteca `faraway` do `R`.

- As variáveis a serem analisadas são as seguintes:
    - `odor`: escore de odor do produto (variável resposta);
    - `temp`: temperatura, codificada como -1, 0 e 1;
    - `gas`: razão gás/líquido, codificada como -1, 0 e 1;

Exemplo- Odor de produto químico

- Nesta aplicação serão analisados dados de um experimento químico disponíveis na base de dados `odor`, que pode ser acessada na biblioteca `faraway` do `R`.

- As variáveis a serem analisadas são as seguintes:
  - `odor`: escore de odor do produto (variável resposta);
  - `temp`: temperatura, codificada como -1, 0 e 1;
  - `gas`: razão gás/líquido, codificada como -1, 0 e 1;
  - `pack`: peso da embalagem, codificada como -1, 0 e 1.

Exemplo- Odor de produto químico

- Nesta aplicação serão analisados dados de um experimento químico disponíveis na base de dados `odor`, que pode ser acessada na biblioteca `faraway` do `R`.

- As variáveis a serem analisadas são as seguintes:
  - `odor`: escore de odor do produto (variável resposta);
  - `temp`: temperatura, codificada como -1, 0 e 1;
  - `gas`: razão gás/líquido, codificada como -1, 0 e 1;
  - `pack`: peso da embalagem, codificada como -1, 0 e 1.

- Os códigos R e a discussão dos resultados podem ser encontrados em arquivo disponível na página da disciplina.

Exemplo- Odor de produto químico

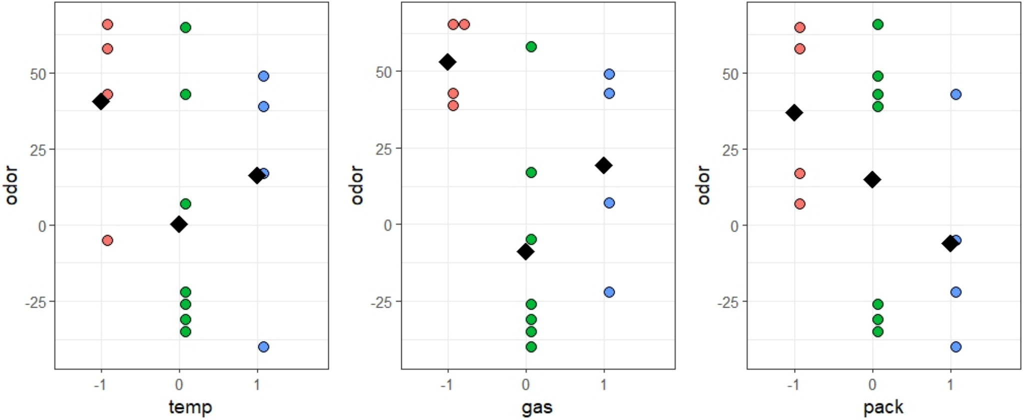

Figura 5: Dados de experimento sobre odor de produto químico

Tópicos adicionais

Inferência conjunta para os parâmetros do modelo

- O teste da significância do modelo de regressão, baseado na hipótese nula:

$$H_0 : \beta_1 = \beta_2 = ... = \beta_k = 0$$

é um caso particular desse teste, em que a estatística F, apresentada no quadro da análise de variância, tem distribuição F-Snedecor com $p-1$ e $n-p$ graus de liberdade sob H_0.

Inferência conjunta para os parâmetros do modelo

- Seja β_q um subconjunto de elementos de β, com os parâmetros que se deseja inferir.

Inferência conjunta para os parâmetros do modelo

- Seja β_q um subconjunto de elementos de β, com os parâmetros que se deseja inferir.

- Adicionalmente, seja $\hat{\beta}_\mathrm{q}$ o vetor de estimadores de mínimos quadrados de β_q.

Inferência conjunta para os parâmetros do modelo

- Seja β_q um subconjunto de elementos de β, com os parâmetros que se deseja inferir.
- Adicionalmente, seja $\hat{\beta}_{\mathrm{q}}$ o vetor de estimadores de mínimos quadrados de β_q.
- Uma região de confiança $100(1-\alpha)\%$ para os componentes de β_q é definido pelo conjunto de todos os vetores $\beta_q^{(0)}$ tais que:

Inferência conjunta para os parâmetros do modelo

- Seja β_q um subconjunto de elementos de β, com os parâmetros que se deseja inferir.

- Adicionalmente, seja $\hat{\beta}_{\text{q}}$ o vetor de estimadores de mínimos quadrados de β_q.

- Uma região de confiança $100(1-\alpha)\%$ para os componentes de β_q é definido pelo conjunto de todos os vetores $\beta_q^{(0)}$ tais que:

$$F_0 = \frac{(\hat{\beta}_q - \beta_q^{(0)})' \mathrm{V}_{qq}^{-1} (\hat{\beta}_q - \beta_q^{(0)})}{q \mathrm{QM}_{\text{Res}}} \leq F_{q,n-p}(1-\alpha)$$

em que $F_{q,n-p}(\alpha)$ é o quantil $1-\alpha$ da distribuição F-Snedecor com q e $n-p$ graus de liberdade.

Inferência para combinações lineares dos parâmetros do modelo

Propriedades das combinações lineares dos estimadores dos parâmetros de regressão

- Seja $\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \hat{\beta}_1, ..., \hat{\beta}_k)$ os estimadores de mínimos quadrados dos parâmetros de um modelo de regressão linear.

Propriedades das combinações lineares dos estimadores dos parâmetros de regressão

- Seja $\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \hat{\beta}_1, ..., \hat{\beta}_k)$ os estimadores de mínimos quadrados dos parâmetros de um modelo de regressão linear.

- Já estudamos anteriormente as seguintes propriedades:

$$\mathrm{E}(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta} \quad ; \quad \mathrm{Var}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{X})^{-1}\sigma^2$$

Propriedades das combinações lineares dos estimadores dos parâmetros de regressão

- Vamos considerar uma combinação linear dos estimadores definida por:

$$\boldsymbol{l}'\hat{\boldsymbol{\beta}} = l_0\hat{\beta}_0 + l_1\hat{\beta}_1 + ... + l_k\hat{\beta}_k$$

Propriedades das combinações lineares dos estimadores dos parâmetros de regressão

- Vamos considerar uma combinação linear dos estimadores definida por:

$$\boldsymbol{l}'\hat{\boldsymbol{\beta}} = l_0\hat{\beta}_0 + l_1\hat{\beta}_1 + ... + l_k\hat{\beta}_k$$

- Neste caso:

$$\mathrm{E}(\boldsymbol{l}'\hat{\boldsymbol{\beta}}) = \boldsymbol{l}'\boldsymbol{\beta} = l_0\beta_0 + l_1\beta_1 + ... + l_k\beta_k$$

e

$$\mathrm{Var}(\boldsymbol{l}'\hat{\boldsymbol{\beta}}) = \sigma^2\boldsymbol{l}'(\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\boldsymbol{l}$$

Propriedades das combinações lineares dos estimadores dos parâmetros de regressão

- Adicionalmente, ao assumir normalidade para os erros, vale a normalidade para $\hat{\boldsymbol{\beta}}$ e para suas combinações lineares, tal que:

$$\boldsymbol{l}'\hat{\boldsymbol{\beta}} \sim \text{Normal}(\boldsymbol{l}'\boldsymbol{\beta}, \sigma^2\boldsymbol{l}'(\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\boldsymbol{l})$$

Propriedades das combinações lineares dos estimadores dos parâmetros de regressão

- Agora, vamos considerar um conjunto de q combinações lineares dos $\hat{\beta}$'s:

Propriedades das combinações lineares dos estimadores dos parâmetros de regressão

- Agora, vamos considerar um conjunto de q combinações lineares dos $\hat{\beta}$'s:

$$L_{10}\hat{\beta}_0 + L_{11}\hat{\beta}_1 + L_{12}\hat{\beta}_2 + ... + L_{1k}\hat{\beta}_k$$

$$L_{20}\hat{\beta}_0 + L_{21}\hat{\beta}_1 + L_{22}\hat{\beta}_2 + ... + L_{2k}\hat{\beta}a_k$$

$$\vdots$$

$$L_{q0}\hat{\beta}_0 + L_{q1}\hat{\beta}_1 + L_{q2}\hat{\beta}_2 + ... + L_{qk}\hat{\beta}_k$$

Propriedades das combinações lineares dos estimadores dos parâmetros de regressão

- Esse conjunto de combinações lineares pode ser expresso matricialmente por $\mathbf{L}\hat{\beta}$ onde:

$$\mathbf{L} = \begin{bmatrix} L_{10} & L_{11} & L_{12} & \cdots & L_{1k} \\ L_{20} & L_{21} & L_{22} & \cdots & L_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ L_{q0} & L_{q1} & L_{q2} & \cdots & L_{qk} \end{bmatrix}$$

Propriedades das combinações lineares dos estimadores dos parâmetros de regressão

- Esse conjunto de combinações lineares pode ser expresso matricialmente por $\mathbf{L}\hat{\beta}$ onde:

$$\mathbf{L} = \begin{bmatrix} L_{10} & L_{11} & L_{12} & \cdots & L_{1k} \\ L_{20} & L_{21} & L_{22} & \cdots & L_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ L_{q0} & L_{q1} & L_{q2} & \cdots & L_{qk} \end{bmatrix}$$

- Desta forma, valem as seguintes propriedades para $\mathbf{L}\hat{\beta}$:

$$\mathrm{E}(\mathbf{L}\hat{\beta}) = \mathbf{L}\beta \quad ; \quad \mathrm{Var}(\mathbf{L}\hat{\beta}) = \mathbf{L}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{L}'\sigma^2$$

Propriedades das combinações lineares dos estimadores dos parâmetros de regressão

- Adicionalmente, ao assumir normalidade para os erros, vale a normalidade para $\hat{\boldsymbol{\beta}}$ e para suas combinações lineares, tal que:

$$\mathbf{L}\hat{\beta} \sim \text{Normal}_q(\mathbf{L}\beta, \mathbf{L}(\mathbf{X}'\mathbf{X})^{\mathbf{-1}}\mathbf{L}'\sigma^2)$$

Inferência para combinações lineares dos parâmetros do modelo

- De forma geral, podemos definir hipóteses lineares na forma:

$$H_0 : \mathrm{L}\beta = \mathrm{c},$$

em que L é uma matriz de constantes de dimensão $q \times p$, de rank linha completo, e c um vetor de constantes de dimensão q (ambos especificados).

Inferência para combinações lineares dos parâmetros do modelo

- De forma geral, podemos definir hipóteses lineares na forma:

$$H_0 : \mathrm{L}\beta = \mathrm{c},$$

em que L é uma matriz de constantes de dimensão $q \times p$, de rank linha completo, e c um vetor de constantes de dimensão q (ambos especificados).

- Neste caso, H_0 compreende q hipóteses lineares sobre os parâmetros do modelo, do tipo:

$$L_{10}\beta_0 + L_{11}\beta_1 + L_{12}\beta_2 + ... + L_{1k}\beta_k = c_1$$

$$L_{20}\beta_0 + L_{21}\beta_1 + L_{22}\beta_2 + ... + L_{2k}\beta_k = c_2$$

$$\vdots$$

$$L_{q0}\beta_0 + L_{q1}\beta_1 + L_{q2}\beta_2 + ... + L_{qk}\beta_k = c_q$$

.

Inferência para combinações lineares dos parâmetros do modelo

- Sob a hipótese H_0, a estatística:

$$F = \frac{(\mathrm{L}\hat{\boldsymbol{\beta}} - \mathrm{c})'[\mathbf{L}(\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{L}']^{-1}(\mathrm{L}\hat{\boldsymbol{\beta}} - \mathrm{c})}{q\mathrm{QM}_{\mathrm{Res}}}$$

tem distribuição F-Snedecor com q e $n - p$ graus de liberdade.

Inferência para combinações lineares dos parâmetros do modelo

- Sob a hipótese H_0, a estatística:

$$F = \frac{(\mathrm{L}\hat{\boldsymbol{\beta}} - \mathrm{c})'[\mathbf{L}(\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{L}']^{-1}(\mathrm{L}\hat{\boldsymbol{\beta}} - \mathrm{c})}{q\mathrm{QM}_{\mathrm{Res}}}$$

tem distribuição F-Snedecor com q e $n - p$ graus de liberdade.

- Assim, a hipótese nula será rejeitada, ao nível de significância α, se o valor calculado da estatística F exceder o quantil $1 - \alpha$ da distribuição F-Snedecor com q e $n - p$ graus de liberdade.

Intervalos de confiança para combinações lineares dos parâmetros

- Seja $\mathrm{l}' = (l_0, l_1, l_2, ..., l_k)$ um vetor de constantes e considere interesse em estimar $\theta = l'\boldsymbol{\beta}$.

- Seja $\mathrm{l}' = (l_0, l_1, l_2, ..., l_k)$ um vetor de constantes e considere interesse em estimar $\theta = l'\boldsymbol{\beta}$.

- A estimativa pontual de $\boldsymbol{l}'\boldsymbol{\beta}$ é dada por $\boldsymbol{l}'\hat{\boldsymbol{\beta}}$.

Intervalos de confiança para combinações lineares dos parâmetros

- Seja $\mathrm{l}' = (l_0, l_1, l_2, ..., l_k)$ um vetor de constantes e considere interesse em estimar $\theta = l'\boldsymbol{\beta}$.

- A estimativa pontual de $\boldsymbol{l}'\boldsymbol{\beta}$ é dada por $\boldsymbol{l}'\hat{\boldsymbol{\beta}}$.

- Um intervalo de confiança $100(1-\alpha)\%$ para $\boldsymbol{l}'\boldsymbol{\beta}$ tem limites:

$$l'\hat{\boldsymbol{\beta}} \pm t_{n-p,\alpha/2}\sqrt{\hat{\sigma}^2\boldsymbol{l}(\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\boldsymbol{l}}$$

Exemplo- Desempenho de alunos

Exemplo- Desempenho de alunos

- Nesta aplicação vamos utilizar dados de desempenho de $n = 100$ alunos no processo seletivo e ao término do primeiro semestre de curso para produzir inferências para combinações lineares dos parâmetros do modelo. As variáveis são as seguintes:

Exemplo- Desempenho de alunos

- Nesta aplicação vamos utilizar dados de desempenho de $n = 100$ alunos no processo seletivo e ao término do primeiro semestre de curso para produzir inferências para combinações lineares dos parâmetros do modelo. As variáveis são as seguintes:
  - `Y`: desempenho no primeiro semestre do curso (variável resposta);

Exemplo- Desempenho de alunos

- Nesta aplicação vamos utilizar dados de desempenho de $n = 100$ alunos no processo seletivo e ao término do primeiro semestre de curso para produzir inferências para combinações lineares dos parâmetros do modelo. As variáveis são as seguintes:
  - `Y`: desempenho no primeiro semestre do curso (variável resposta);
  - `V1`: nota na primeira prova do processo seletivo;

Exemplo- Desempenho de alunos

- Nesta aplicação vamos utilizar dados de desempenho de $n = 100$ alunos no processo seletivo e ao término do primeiro semestre de curso para produzir inferências para combinações lineares dos parâmetros do modelo. As variáveis são as seguintes:
  - `Y`: desempenho no primeiro semestre do curso (variável resposta);
  - `V1`: nota na primeira prova do processo seletivo;
  - `V2`: nota na segunda prova do processo seletivo;

Exemplo- Desempenho de alunos

- Nesta aplicação vamos utilizar dados de desempenho de $n = 100$ alunos no processo seletivo e ao término do primeiro semestre de curso para produzir inferências para combinações lineares dos parâmetros do modelo. As variáveis são as seguintes:
  - `Y`: desempenho no primeiro semestre do curso (variável resposta);
  - `V1`: nota na primeira prova do processo seletivo;
  - `V2`: nota na segunda prova do processo seletivo;
  - `V3`: nota na terceira prova do processo seletivo;

Exemplo- Desempenho de alunos

- Nesta aplicação vamos utilizar dados de desempenho de $n = 100$ alunos no processo seletivo e ao término do primeiro semestre de curso para produzir inferências para combinações lineares dos parâmetros do modelo. As variáveis são as seguintes:
  - `Y`: desempenho no primeiro semestre do curso (variável resposta);
  - `V1`: nota na primeira prova do processo seletivo;
  - `V2`: nota na segunda prova do processo seletivo;
  - `V3`: nota na terceira prova do processo seletivo;
  - `V4`: nota na quarta prova do processo seletivo.

Exemplo- Desempenho de alunos

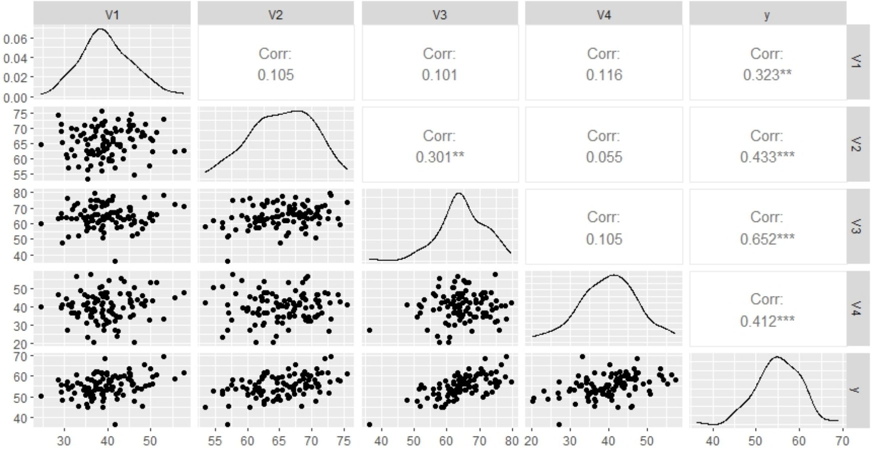

Figura 6: Desempenho dos alunos no processo seletivo e ao término do primeiro semestre

Exemplo- Desempenho de alunos

- O modelo de regressão linear múltipla a ser ajustado é o seguinte:

$$Y_i = \beta_0 + \beta_1 V_{i1} + \beta_2 V_{i2} + \beta_3 V_{i3} + \beta_4 V_{i4} + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma^2)$$

Exemplo- Desempenho de alunos

- O modelo de regressão linear múltipla a ser ajustado é o seguinte:

$$Y_i = \beta_0 + \beta_1 V_{i1} + \beta_2 V_{i2} + \beta_3 V_{i3} + \beta_4 V_{i4} + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma^2)$$

- Com base nas estimativas de mínimos quadrados, o modelo ajustado é o seguinte:

$$\hat{y} = -2.684 + 0.179 V_1 + 0.256 V_2 + 0.389 V_3 + 0.221 V_4$$

Exemplo- Desempenho de alunos

- A matriz de variâncias e covariâncias para os $\hat{\beta}'s$ foi estimada:

$$\widehat{\text{Var}}(\hat{\boldsymbol{\beta}}) = \begin{bmatrix} 25.659991 & -0.074595 & -0.245911 & -0.070114 & -0.049330 \\ -0.074595 & 0.002848 & -0.000287 & -0.000161 & -0.000239 \\ -0.245911 & -0.000287 & 0.004928 & -0.000973 & -0.000049 \\ -0.070114 & -0.000161 & -0.000973 & 0.002277 & -0.000175 \\ -0.049330 & -0.000239 & -0.000049 & -0.000175 & 0.001833 \end{bmatrix}$$

Exemplo- Desempenho de alunos

- Vamos iniciar por uma hipótese simples, de que o efeito médio de V_1 e V_2 em Y é igual ao efeito médio de V_3 e V_4.

Exemplo- Desempenho de alunos

- Vamos iniciar por uma hipótese simples, de que o efeito médio de V_1 e V_2 em Y é igual ao efeito médio de V_3 e V_4.

- Para testar esta conjectura, o seguinte par de hipóteses foi definido:

$$H_0 : \frac{\beta_1 + \beta_2}{2} = \frac{\beta_3 + \beta_4}{2} \quad vs \quad H_1 : \frac{\beta_1 + \beta_2}{2} \neq \frac{\beta_3 + \beta_4}{2}$$

Exemplo- Desempenho de alunos

- Vamos iniciar por uma hipótese simples, de que o efeito médio de V_1 e V_2 em Y é igual ao efeito médio de V_3 e V_4.

- Para testar esta conjectura, o seguinte par de hipóteses foi definido:

$$H_0 : \frac{\beta_1 + \beta_2}{2} = \frac{\beta_3 + \beta_4}{2} \quad vs \quad H_1 : \frac{\beta_1 + \beta_2}{2} \neq \frac{\beta_3 + \beta_4}{2}$$

- Podemos reescrever **a hipótese nula** por:

$$H_0 : \frac{\beta_1 + \beta_2 - \beta_3 - \beta_4}{2} = 0,$$

de tal forma que o contraste a ser testado pode ser definido pelo vetor de coeficientes $l' = (0 \;\; 1/2 \;\; 1/2 \;\; -1/2 \;\; -1/2)$.

Exemplo- Desempenho de alunos

- O valor estimado para o contraste é:

$$l'\hat{\beta} = \begin{bmatrix} 0 & 1/2 & 1/2 & -1/2 & -1/2 \end{bmatrix} \begin{bmatrix} 2.684 \\ 0.179 \\ 0.256 \\ 0.389 \\ 0.221 \end{bmatrix} = -0.0878$$

- Já a variância do contraste é dada por:

$$\widehat{\text{Var}}(l'\hat{\beta}) = l'\widehat{\text{Var}}(\hat{\beta})l = \begin{bmatrix} 0 \\ 1/2 \\ 1/2 \\ -1/2 \\ -1/2 \end{bmatrix}' \begin{bmatrix} 25.659991 & -0.074595 & -0.245911 & -0.070114 & -0.049330 \\ -0.074595 & 0.002848 & -0.000287 & -0.000161 & -0.000239 \\ -0.245911 & -0.000287 & 0.004928 & -0.000973 & -0.000049 \\ -0.070114 & -0.000161 & -0.000973 & 0.002277 & -0.000175 \\ -0.049330 & -0.000239 & -0.000049 & -0.000175 & 0.001833 \end{bmatrix} \begin{bmatrix} 0 \\ 1/2 \\ 1/2 \\ -1/2 \\ -1/2 \end{bmatrix} = 0.0034$$

Exemplo- Desempenho de alunos

- **Estatística do teste:**

$$t = \frac{l'\hat{\boldsymbol{\beta}}}{\sqrt{\widehat{\text{Var}}(l'\hat{\boldsymbol{\beta}})}} = \frac{-0.0878}{\sqrt{0.0034}} = -1.5058$$

Exemplo- Desempenho de alunos

- **Estatística do teste:**

$$t = \frac{l'\hat{\boldsymbol{\beta}}}{\sqrt{\widehat{\text{Var}}(l'\hat{\boldsymbol{\beta}})}} = \frac{-0.0878}{\sqrt{0.0034}} = -1.5058$$

- **Regra de decisão:** devemos rejeitar H_0, ao nível de significância $\alpha = 5\%$, se $|t| > |t_{100-5;2.5\%}| = 1.985$.

Exemplo- Desempenho de alunos

- **Estatística do teste:**

$$t = \frac{l'\hat{\boldsymbol{\beta}}}{\sqrt{\widehat{\text{Var}}(l'\hat{\boldsymbol{\beta}})}} = \frac{-0.0878}{\sqrt{0.0034}} = -1.5058$$

- **Regra de decisão:** devemos rejeitar H_0, ao nível de significância $\alpha = 5\%$, se $|t| > |t_{100-5;2.5\%}| = 1.985$.

- **Decisão:** Como $|t| = 1.5058 < 1.985$, não temos evidências amostrais para rejeitar a hipótese H_0 ao nível de 5% de significância.

Exemplo- Desempenho de alunos

- Um IC(95%) para $l\boldsymbol{\beta}$ tem limites:

Exemplo- Desempenho de alunos

- Um IC(95%) para $l\boldsymbol{\beta}$ tem limites:

$$l'\hat{\boldsymbol{\beta}} \pm t_{100-5;2.5\%}\sqrt{\widehat{\text{Var}}(\hat{\boldsymbol{\beta}})} = -0.0878 \pm 1.985\sqrt{0.0034} =$$

Exemplo- Desempenho de alunos

- Um IC(95%) para $l\boldsymbol{\beta}$ tem limites:

$$l'\hat{\boldsymbol{\beta}} \pm t_{100-5;2.5\%}\sqrt{\widehat{\text{Var}}(\hat{\boldsymbol{\beta}})} = -0.0878 \pm 1.985\sqrt{0.0034} =$$

$$-0.0878 \pm 0.1157 =$$

Exemplo- Desempenho de alunos

- Um IC(95%) para $l\boldsymbol{\beta}$ tem limites:

$$l'\hat{\boldsymbol{\beta}} \pm t_{100-5;2.5\%}\sqrt{\widehat{\text{Var}}(\hat{\boldsymbol{\beta}})} = -0.0878 \pm 1.985\sqrt{0.0034} =$$

$$-0.0878 \pm 0.1157 =$$

$$(-0.2035; 0.02794)$$

Exemplo- Desempenho de alunos

- Vamos testar a hipótese de iguais efeitos de V_2 e V_4 na resposta, e que o efeito de V_3 seja o dobro de V_2.

Exemplo- Desempenho de alunos

- Vamos testar a hipótese de iguais efeitos de V_2 e V_4 na resposta, e que o efeito de V_3 seja o dobro de V_2.

- A **hipótese nula** do teste pode ser formulada da seguinte forma:

$$H_0 : \beta_2 = \beta_4 \text{ e } \beta_3 = 2\beta_2$$

Exemplo- Desempenho de alunos

- Podemos escrever as conjecturas em H_0 de maneira equivalente por:

$$H_0 : \beta_2 - \beta_4 = 0 \text{ e } \beta_3 - 2\beta_2 = 0$$

Exemplo- Desempenho de alunos

- Podemos escrever as conjecturas em H_0 de maneira equivalente por:

$$H_0 : \beta_2 - \beta_4 = 0 \text{ e } \beta_3 - 2\beta_2 = 0$$

- Ou seja, podemos escrever o par de conjecturas por $\boldsymbol{L\beta} = \mathbf{0}$ definindo $\boldsymbol{L}$ por:

$$\boldsymbol{L} = \begin{bmatrix} 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & -2 & 1 & 0 \end{bmatrix}$$

Exemplo- Desempenho de alunos

- De tal forma que:

$$H_0 : \boldsymbol{L\beta} = \begin{bmatrix} 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & -2 & 1 & 0 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Exemplo- Desempenho de alunos

- De tal forma que:

$$H_0 : \boldsymbol{L\beta} = \begin{bmatrix} 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & -2 & 1 & 0 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

- **Estatística do teste:**

$$F = \frac{(\mathrm{L}\hat{\boldsymbol{\beta}} - \mathrm{c})'[\mathrm{L}(\mathrm{X}'\mathrm{X})^{-1}\mathrm{L}']^{-1}(\mathrm{L}\hat{\boldsymbol{\beta}} - \mathrm{c})}{q\mathrm{QM}_{\mathrm{Res}}} = 0.3428$$

Exemplo- Desempenho de alunos

- **Regra de decisão:** a hipótese H_0 deverá ser rejeitada ao nível de 5% de significância se $F > F_{q=2,n-p=95}(0.95) = 3.09$.

Exemplo- Desempenho de alunos

- **Regra de decisão:** a hipótese H_0 deverá ser rejeitada ao nível de 5% de significância se $F > F_{q=2,n-p=95}(0.95) = 3.09$.

- **Decisão:** Como $F = 0.34 < 3.09$, a hipótese nula não pode ser rejeitada.

Exemplo- Desempenho de alunos

- **Regra de decisão:** a hipótese H_0 deverá ser rejeitada ao nível de 5% de significância se $F > F_{q=2,n-p=95}(0.95) = 3.09$.

- **Decisão:** Como $F = 0.34 < 3.09$, a hipótese nula não pode ser rejeitada.

- Estimativas, erros padrões e intervalos de confiança para os contrastes são apresentados nos scripts.

Exemplo- Desempenho de alunos

- Agora, considere o conjunto de conjecturas:

Exemplo- Desempenho de alunos

- Agora, considere o conjunto de conjecturas:
    - Os efeitos de V_2 e V_3 são iguais;

Exemplo- Desempenho de alunos

- Agora, considere o conjunto de conjecturas:
  - Os efeitos de V_2 e V_3 são iguais;
  - O efeito de V_3 é o dobro do efeito de V_1;

Exemplo- Desempenho de alunos

- Agora, considere o conjunto de conjecturas:
    - Os efeitos de V_2 e V_3 são iguais;
    - O efeito de V_3 é o dobro do efeito de V_1;
    - O efeito de V_4 é a média aritmética dos efeitos de V_1, V_2 e V_3.

Exemplo- Desempenho de alunos

- Agora, considere o conjunto de conjecturas:
    - Os efeitos de V_2 e V_3 são iguais;
    - O efeito de V_3 é o dobro do efeito de V_1;
    - O efeito de V_4 é a média aritmética dos efeitos de V_1, V_2 e V_3.

- As conjecturas apresentadas, em função dos parâmetros de regressão, são definidas da seguinte forma:
    - $\beta_2 = \beta_3$ (ou seja, $\beta_2 - \beta_3 = 0$);

Exemplo- Desempenho de alunos

- Agora, considere o conjunto de conjecturas:
  - Os efeitos de V_2 e V_3 são iguais;
  - O efeito de V_3 é o dobro do efeito de V_1;
  - O efeito de V_4 é a média aritmética dos efeitos de V_1, V_2 e V_3.

- As conjecturas apresentadas, em função dos parâmetros de regressão, são definidas da seguinte forma:
  - $\beta_2 = \beta_3$ (ou seja, $\beta_2 - \beta_3 = 0$);
  - $\beta_3 = 2\beta_1$ (ou seja, $-2\beta_1 + \beta_3 = 0$);

Exemplo- Desempenho de alunos

- Agora, considere o conjunto de conjecturas:
  - Os efeitos de V_2 e V_3 são iguais;
  - O efeito de V_3 é o dobro do efeito de V_1;
  - O efeito de V_4 é a média aritmética dos efeitos de V_1, V_2 e V_3.

- As conjecturas apresentadas, em função dos parâmetros de regressão, são definidas da seguinte forma:
  - $\beta_2 = \beta_3$ (ou seja, $\beta_2 - \beta_3 = 0$);
  - $\beta_3 = 2\beta_1$ (ou seja, $-2\beta_1 + \beta_3 = 0$);
  - $\beta_4 = \frac{\beta_1+\beta_2+\beta_3}{3}$ (ou seja, $-\frac{1}{3}\beta_1 + -\frac{1}{3}\beta_2 - \frac{1}{3}\beta_3 + \beta_4 = 0$).

Exemplo- Desempenho de alunos

- Neste caso, a matriz de contrastes fica definida por:

$$\boldsymbol{L} = \begin{bmatrix} 0 & 0 & 1 & -1 & 0 \\ 0 & -2 & 0 & 1 & 0 \\ 0 & -1/3 & -1/3 & -1/3 & 1 \end{bmatrix}$$

Exemplo- Desempenho de alunos

- Neste caso, a matriz de contrastes fica definida por:

$$\boldsymbol{L} = \begin{bmatrix} 0 & 0 & 1 & -1 & 0 \\ 0 & -2 & 0 & 1 & 0 \\ 0 & -1/3 & -1/3 & -1/3 & 1 \end{bmatrix}$$

- Logo, a **hipótese nula** do teste fica dada por:

$$H_0 : \boldsymbol{L\beta} = \begin{bmatrix} 0 & 0 & 1 & -1 & 0 \\ 0 & -2 & 0 & 1 & 0 \\ 0 & -1/3 & -1/3 & -1/3 & 1 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Exemplo- Desempenho de alunos

- **Estatística do teste:**

$$F = \frac{(\mathrm{L}\ \hat{\boldsymbol{\beta}} - \mathrm{c})'[\mathrm{L}(\mathrm{X}'\mathrm{X})^{-1}\mathrm{L}']^{-1}(\mathrm{L}\hat{\boldsymbol{\beta}} - \mathrm{c})}{q\mathrm{QM}_{\mathrm{Res}}} = 1.132$$

Exemplo- Desempenho de alunos

- **Estatística do teste:**

$$F = \frac{(\mathrm{L}\ \hat{\boldsymbol{\beta}} - \mathrm{c})'[\mathrm{L}(\mathrm{X}'\mathrm{X})^{-1}\mathrm{L}']^{-1}(\mathrm{L}\hat{\boldsymbol{\beta}} - \mathrm{c})}{q\mathrm{QM}_{\mathrm{Res}}} = 1.132$$

- **Regra de decisão:** a hipótese H_0 deverá ser rejeitada ao nível de 5% de significância se $F > F_{q=3,n-p=95}(0.95) = 2.70$.

Exemplo- Desempenho de alunos

- **Estatística do teste:**

$$F = \frac{(\mathrm{L}\ \hat{\boldsymbol{\beta}} - \mathrm{c})'[\mathrm{L}(\mathrm{X}'\mathrm{X})^{-1}\mathrm{L}']^{-1}(\mathrm{L}\hat{\boldsymbol{\beta}} - \mathrm{c})}{q\mathrm{QM}_{\mathrm{Res}}} = 1.132$$

- **Regra de decisão:** a hipótese H_0 deverá ser rejeitada ao nível de 5% de significância se $F > F_{q=3,n-p=95}(0.95) = 2.70$.

- **Decisão:** Como $F = 1.132 < 2.70$, a hipótese nula não pode ser rejeitada.

Exemplo- Desempenho de alunos

- **Estatística do teste:**

$$F = \frac{(\mathrm{L}\ \hat{\boldsymbol{\beta}} - \mathrm{c})'[\mathrm{L}(\mathrm{X}'\mathrm{X})^{-1}\mathrm{L}']^{-1}(\mathrm{L}\hat{\boldsymbol{\beta}} - \mathrm{c})}{q\mathrm{QM}_{\mathrm{Res}}} = 1.132$$

- **Regra de decisão:** a hipótese H_0 deverá ser rejeitada ao nível de 5% de significância se $F > F_{q=3,n-p=95}(0.95) = 2.70$.

- **Decisão:** Como $F = 1.132 < 2.70$, a hipótese nula não pode ser rejeitada.

- Estimativas, erros padrões e intervalos de confiança para os contrastes são apresentados nos scripts.