Modelo de Regressão Linear Simples

Prof. Juvêncio Santos Nobre

Departamento de Estatística e Matemática Aplicada

Universidade Federal do Ceará-Brasil

http://www.dema.ufc.br/∼juvencio

DEMA-UFC

Capital do **Ceará**, agosto de 2022

Conteúdo

MRLS

- O modelo de regressão linear simples (MRLS) admite a seguinte forma funcional

$$y_i = \beta_0 + \beta_1 x_i + e_i, i = 1, \ldots, n, \quad (1)$$

em que:

- y_i (x_i) denota o valor da variável resposta (explicativa) referente ao i-ésimo elemento da amostra.
- β_0 e β_1 são parâmetros desconhecidos, denominados parâmetros (coeficientes) de regressão.
- e_i representa a fonte de variação associada ao i-ésimo elemento da amostra.

MRLS

- O modelo de regressão linear simples (MRLS) admite a seguinte forma funcional

$$y_i = \beta_0 + \beta_1 x_i + e_i, i = 1, \ldots, n, \tag{1}$$

em que:

  - y_i (x_i) denota o valor da variável resposta (explicativa) referente ao i-ésimo elemento da amostra.
  - β_0 e β_1 são parâmetros desconhecidos, denominados parâmetros (coeficientes) de regressão.
  - e_i representa a fonte de variação associada ao i-ésimo elemento da amostra.

MRLS

- O modelo de regressão linear simples (MRLS) admite a seguinte forma funcional

$$y_i = \beta_0 + \beta_1 x_i + e_i, i = 1, \ldots, n, \quad (1)$$

em que:

  - y_i (x_i) denota o valor da variável resposta (explicativa) referente ao i-ésimo elemento da amostra.
  - β_0 e β_1 são parâmetros desconhecidos, denominados parâmetros (coeficientes) de regressão.
  - e_i representa a fonte de variação associada ao i-ésimo elemento da amostra.

MRLS

- O modelo de regressão linear simples (MRLS) admite a seguinte forma funcional

$$y_i = \beta_0 + \beta_1 x_i + e_i, i = 1, \ldots, n, \tag{1}$$

em que:

  - y_i (x_i) denota o valor da variável resposta (explicativa) referente ao i-ésimo elemento da amostra.
  - β_0 e β_1 são parâmetros desconhecidos, denominados parâmetros (coeficientes) de regressão.
  - e_i representa a fonte de variação associada ao i-ésimo elemento da amostra.

MRLS

- Ao estabelecer o MRLS, pressupomos que:

  i) A função de regressão é linear (nos parâmetros). É comum, apesar de formalmente incorreta, nos textos aparecer a relação entre y_i e x_i é linear nos parâmetros.

  ii) Os valores de x_i são fixos, i.e., x_i não é uma variável aleatória.

  iii) $\mathbb{E}[e_i] = 0$, $\forall i = 1, \ldots, n$. Na verdade, tal suposição deveria ser escrita como (o que acaba implicando a anterior) $\mathbb{E}[e_i|x_i] = 0, \forall i = 1, \ldots, n$.

  iv) Para um dado valor de x_i, a variância da fonte de variação é constante, i.e.,

  $$\mathrm{Var}[e_i] = \mathbb{E}[e_i^2] = \sigma^2, \forall i = 1, \ldots, n \ \ (\textbf{Homoscesdaticidade}).$$

  Na verdade, tal suposição deveria ser escrita como

  $$\mathrm{Var}[y_i|x_i] = \mathrm{Var}[e_i|x_i] = \sigma^2, \forall i = 1, \ldots, n.$$

  v) A fonte de variação associada a uma observação é não-correlacionada com a fonte de variação associada de outra observação, i.e.,

  $$\mathrm{Cov}(e_i, e_j) = \mathbb{E}[e_i e_j] = 0, \forall i \neq j.$$

MRLS

- As suposições iv) e v) podem ser reescritas de sucintamente da seguinte forma

$$\mathrm{Cov}(e_i, e_j) = \sigma^2 \mathbb{1}(i = j), \forall i, j = 1 \ldots, n.$$

- Perceba que no MRLS (1) assume-se essencialmente que a fonte de variação está relacionada somente a variável resposta, i.e, a variável explicativa é medida sem erro, ou seja, com completa exatidão. Isso é razoável no contexto prático?
- Se tivermos uma fonte de variação também associada a variável explicativa x_i, teremos essencialmente um *modelo com erro de medida/erro nas variáveis*.

MRLS

- As suposições iv) e v) podem ser reescritas de sucintamente da seguinte forma

$$\mathrm{Cov}(e_i, e_j) = \sigma^2 \mathbb{1}(i = j), \forall i, j = 1 \ldots, n.$$

- Perceba que no MRLS (1) assume-se essencialmente que a fonte de variação está relacionada somente a variável resposta, i.e, a variável explicativa é medida sem erro, ou seja, com completa exatidão. Isso é razoável no contexto prático? 😵
- Se tivermos uma fonte de variação também associada a variável explicativa x_i, teremos essencialmente um *modelo com erro de medida/erro nas variáveis*. 😎

MRLS

- As suposições iv) e v) podem ser reescritas de sucintamente da seguinte forma

$$\mathrm{Cov}(e_i, e_j) = \sigma^2 \mathbb{1}(i = j), \forall i, j = 1 \ldots, n.$$

- Perceba que no MRLS (1) assume-se essencialmente que a fonte de variação está relacionada somente a variável resposta, i.e, a variável explicativa é medida sem erro, ou seja, com completa exatidão. Isso é razoável no contexto prático?
- Se tivermos uma fonte de variação também associada a variável explicativa x_i, teremos essencialmente um *modelo com erro de medida/erro nas variáveis*.

MRLS

- Para efeito de **inferência de segunda ordem** exata, i.e., construção de IC, testes de hipóteses, é comum considerar também que

$$e_i \sim \mathcal{N}(0, \sigma^2), i = 1, \ldots, n.$$

- Lembrando, que **correlação nula** implica independência sob a suposição de normalidade multivariada, então usando as suposições iv) e v) adicionada com a suposição acima, temos

$$e_i \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2), i = 1, \ldots, n.$$

- Usando o fato que a distribuição normal é **fechada** por transformações lineares, então sob as suposições usuais do MRLS adicionada a suposição de normalidade, tem-se

$$y_i \overset{\text{ind}}{\sim} \mathcal{N}(\beta_0 + \beta_1 x_i, \sigma^2), i = 1, \ldots, n.$$

MRLS

- Para efeito de **inferência de segunda ordem** exata, i.e., construção de IC, testes de hipóteses, é comum considerar também que

$$e_i \sim \mathcal{N}(0, \sigma^2), i = 1, \ldots, n.$$

- Lembrando, que **correlação nula** implica independência sob a suposição de normalidade multivariada, então usando as suposições iv) e v) adicionada com a suposição acima, temos

$$e_i \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2), i = 1, \ldots, n.$$

- Usando o fato que a distribuição normal é **fechada** por transformações lineares, então sob as suposições usuais do MRLS adicionada a suposição de normalidade, tem-se

$$y_i \overset{\text{ind}}{\sim} \mathcal{N}(\beta_0 + \beta_1 x_i, \sigma^2), i = 1, \ldots, n.$$

MRLS

- Para efeito de **inferência de segunda ordem** exata, i.e., construção de IC, testes de hipóteses, é comum considerar também que

$$e_i \sim \mathcal{N}(0, \sigma^2), i = 1, \ldots, n.$$

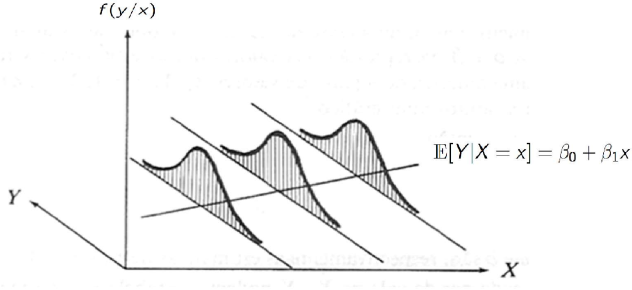
- Lembrando, que **correlação nula** implica independência sob a suposição de normalidade multivariada, então usando as suposições iv) e v) adicionada com a suposição acima, temos

$$e_i \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2), i = 1, \ldots, n.$$

- Usando o fato que a distribuição normal é **fechada** por transformações lineares, então sob as suposições usuais do MRLS adicionada a suposição de normalidade, tem-se

$$y_i \overset{\text{ind}}{\sim} \mathcal{N}(\beta_0 + \beta_1 x_i, \sigma^2), i = 1, \ldots, n.$$

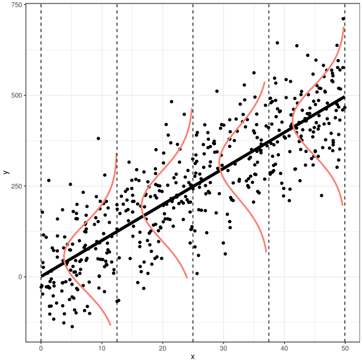
MRLS - Ilustração gráfica



Fonte: Hoffman (2006, Análise de regressão)

MRLS - Ilustração gráfica

Figura: Ilustração gráfica para um exemplo de dados simulados usando o ggplot2.



MRLS - Interpretação dos parâmetros

- Sob as suposições usuais do MRLS, tem-se

$$\mathbb{E}[y_i|X = x] = \beta_0 + \beta_1 x, i = 1, \ldots, n.$$

Logo:

- $\beta_0 = \mathbb{E}[y_i|X = 0]$.
- É válido ressaltar que quando a amplitude amostral não inclui o zero (ou quando não fizer sentido considerar $x = 0$) , então β_0 não possui interpretação prática, sendo necessário centralizar a variável explicativa para tal.
- $\beta_1 = \mathbb{E}[y_i|X = a + 1] - \mathbb{E}[y_i|X = a], \forall a \in \mathbb{R}$, i.e., β_1 representa a variação no valor esperado da variável resposta, quando a variável explicativa é acrescida de uma unidade de medida.

MRLS - Interpretação dos parâmetros

- Sob as suposições usuais do MRLS, tem-se

$$\mathbb{E}[y_i|X = x] = \beta_0 + \beta_1 x, i = 1, \ldots, n.$$

Logo:

- $\beta_0 = \mathbb{E}[y_i|X = 0]$.
- É válido ressaltar que quando a amplitude amostral não inclui o zero (ou quando não fizer sentido considerar $x = 0$) , então β_0 não possui interpretação prática, sendo necessário centralizar a variável explicativa para tal.
- $\beta_1 = \mathbb{E}[y_i|X = a + 1] - \mathbb{E}[y_i|X = a], \forall a \in \mathbb{R}$, i.e., β_1 representa a variação no valor esperado da variável resposta, quando a variável explicativa é acrescida de uma unidade de medida.

MRLS - Interpretação dos parâmetros

- Sob as suposições usuais do MRLS, tem-se

$$\mathbb{E}[y_i|X = x] = \beta_0 + \beta_1 x, i = 1, \ldots, n.$$

Logo:

- $\beta_0 = \mathbb{E}[y_i|X = 0]$.
- É válido ressaltar que quando a amplitude amostral não inclui o zero (ou quando não fizer sentido considerar $x = 0$) , então β_0 não possui interpretação prática, sendo necessário centralizar a variável explicativa para tal.
- $\beta_1 = \mathbb{E}[y_i|X = a + 1] - \mathbb{E}[y_i|X = a], \forall a \in \mathbb{R}$, i.e., β_1 representa a variação no valor esperado da variável resposta, quando a variável explicativa é acrescida de uma unidade de medida.

MRLS - Interpretação dos parâmetros

- Sob as suposições usuais do MRLS, tem-se

$$\mathbb{E}[y_i|X = x] = \beta_0 + \beta_1 x, i = 1, \ldots, n.$$

Logo:

- $\beta_0 = \mathbb{E}[y_i|X = 0]$.
- É válido ressaltar que quando a amplitude amostral não inclui o zero (ou quando não fizer sentido considerar $x = 0$) , então β_0 não possui interpretação prática, sendo necessário centralizar a variável explicativa para tal.
- $\beta_1 = \mathbb{E}[y_i|X = a + 1] - \mathbb{E}[y_i|X = a], \forall a \in \mathbb{R}$, i.e., β_1 representa a variação no valor esperado da variável resposta, quando a variável explicativa é acrescida de uma unidade de medida.

Exemplos - Interpretação dos parâmetros

Exemplo 1: Para os casos abaixo, apresente interpretações práticas dos parâmetros do MRLS:

i) Renda vs. anos estudados (efetivos).

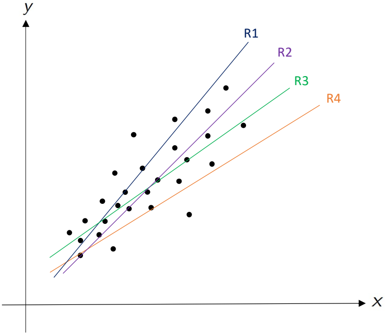
ii) Peso vs altura.

iii) Tempo de processamento vs # de faturas.

iv) Faturamento da empresa vs investimento com propaganda.

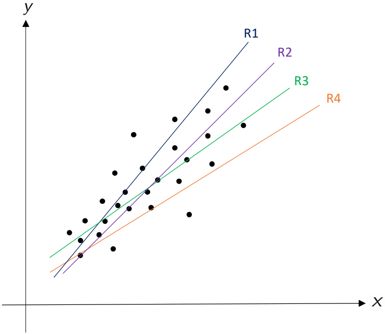
v) Pressão arterial sistólica (mmHg) vs idade (anos).

Método de Mínimos Quadrados (MQ)



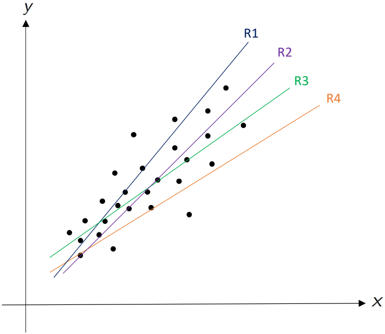
- Dado um conjunto de dados, existem **infinitas** retas candidatas para ajuste.
- Qual delas escolher?

Método de Mínimos Quadrados (MQ)



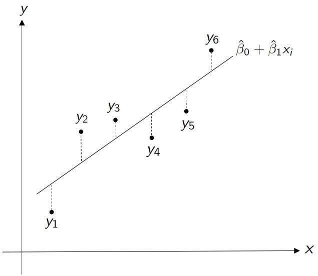
- Dado um conjunto de dados, existem **infinitas** retas candidatas para ajuste.
- Qual delas escolher?

Método de Mínimos Quadrados (MQ)



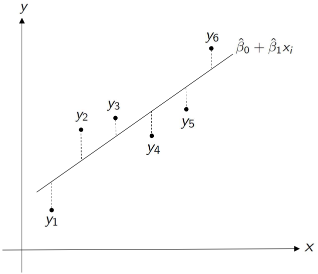
- Dado um conjunto de dados, existem **infinitas** retas candidatas para ajuste.
- Qual delas escolher?

Método de Mínimos Quadrados (MQ) - Ideia



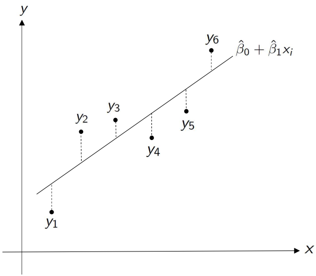
- A melhor reta estimada será aquela que minimiza a distância do valor observado y_i para o valor esperado ajustado $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$, $i = 1, \ldots, n$.
- Infelizmente, não é possível minimizar todas estas distâncias **simultaneamente**, logo, considera-se alguma função conveniente destas distâncias como função objetivo.

Método de Mínimos Quadrados (MQ) - Ideia



- A melhor reta estimada será aquela que minimiza a distância do valor observado y_i para o valor esperado ajustado $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$, $i = 1, \ldots, n$.
- Infelizmente, não é possível minimizar todas estas distâncias **simultaneamente**, logo, considera-se alguma função conveniente destas distâncias como função objetivo.

Método de Mínimos Quadrados (MQ) - Ideia



- A melhor reta estimada será aquela que minimiza a distância do valor observado y_i para o valor esperado ajustado $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$, $i = 1, \ldots, n$.
- Infelizmente, não é possível minimizar todas estas distâncias **simultaneamente**, logo, considera-se alguma função conveniente destas distâncias como função objetivo.

Método de Mínimos Quadrados (MQ) - Ideia

- Podemos considerar, por exemplo, as seguintes funções objetivos:

$$Q_1(\boldsymbol{\beta}) = Q_1(\beta_0, \beta_1) = \sum_{i=1}^{n} |y_i - \beta_0 - \beta_1 x_i| \quad (2)$$

$$Q_2(\boldsymbol{\beta}) = Q_2(\beta_0, \beta_1) = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2. \quad (3)$$

- Acima, temos dois exemplos de funções objetivos de interesse, mas podemos considerar muito mais, basta que seja alguma **norma** (ou **norma** q.c.) com boas propriedades.
- Note que essencialmente temos uma função de perda e o interesse consiste em minimizá-la.
- É possível também utilizar vários outros critérios, como por exemplo, minimizar a diferença máxima, obtendo assim o risco minimax, bem como utilizar utilizar procedimentos paramétricos, tais como EMV, estimadores equivariantes (Pitman, etc...), e outros métodos que fornecem estimadores com propriedades interessantes.

Método de Mínimos Quadrados (MQ) - Ideia

- Podemos considerar, por exemplo, as seguintes funções objetivos:

$$Q_1(\boldsymbol{\beta}) = Q_1(\beta_0, \beta_1) = \sum_{i=1}^{n} |y_i - \beta_0 - \beta_1 x_i| \tag{2}$$

$$Q_2(\boldsymbol{\beta}) = Q_2(\beta_0, \beta_1) = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2. \tag{3}$$

- Acima, temos dois exemplos de funções objetivos de interesse, mas podemos considerar muito mais, basta que seja alguma **norma** (ou **norma** q.c.) com boas propriedades.
- Note que essencialmente temos uma função de perda e o interesse consiste em minimizá-la.
- É possível também utilizar vários outros critérios, como por exemplo, minimizar a diferença máxima, obtendo assim o risco minimax, bem como utilizar utilizar procedimentos paramétricos, tais como EMV, estimadores equivariantes (Pitman, etc...), e outros métodos que fornecem estimadores com propriedades interessantes.

Método de Mínimos Quadrados (MQ) - Ideia

- Podemos considerar, por exemplo, as seguintes funções objetivos:

$$Q_1(\boldsymbol{\beta}) = Q_1(\beta_0, \beta_1) = \sum_{i=1}^{n} |y_i - \beta_0 - \beta_1 x_i| \quad (2)$$

$$Q_2(\boldsymbol{\beta}) = Q_2(\beta_0, \beta_1) = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2. \quad (3)$$

- Acima, temos dois exemplos de funções objetivos de interesse, mas podemos considerar muito mais, basta que seja alguma **norma** (ou **norma** q.c.) com boas propriedades.
- Note que essencialmente temos uma função de perda e o interesse consiste em minimizá-la.
- É possível também utilizar vários outros critérios, como por exemplo, minimizar a diferença máxima, obtendo assim o risco minimax, bem como utilizar utilizar procedimentos paramétricos, tais como EMV, estimadores equivariantes (Pitman, etc...), e outros métodos que fornecem estimadores com propriedades interessantes.

Método de Mínimos Quadrados (MQ) - Ideia

- Podemos considerar, por exemplo, as seguintes funções objetivos:

$$\begin{aligned} Q_1(\boldsymbol{\beta}) &= Q_1(\beta_0, \beta_1) = \sum_{i=1}^{n} |y_i - \beta_0 - \beta_1 x_i| && (2) \\ Q_2(\boldsymbol{\beta}) &= Q_2(\beta_0, \beta_1) = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2. && (3) \end{aligned}$$

- Acima, temos dois exemplos de funções objetivos de interesse, mas podemos considerar muito mais, basta que seja alguma **norma** (ou **norma** q.c.) com boas propriedades.
- Note que essencialmente temos uma função de perda e o interesse consiste em minimizá-la.
- É possível também utilizar vários outros critérios, como por exemplo, minimizar a diferença máxima, obtendo assim o risco minimax, bem como utilizar utilizar procedimentos paramétricos, tais como EMV, estimadores equivariantes (Pitman, etc...), e outros métodos que fornecem estimadores com propriedades interessantes.

Método de Mínimos Quadrados (MQ) - Ideia

- É extremamente comum em alguns textos as funções objetivos (2) e (3) serem apresentadas como

$$
\begin{aligned}
Q_1(\boldsymbol{\beta}) &= \sum_{i=1}^{n} |y_i - \beta_0 - \beta_1 x_i| = \sum_{i=1}^{n} |e_i| \\
Q_2(\boldsymbol{\beta}) &= \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2 = \sum_{i=1}^{n} e_i^2.
\end{aligned}
$$

- Todavia, é válido lembrar que $e_1, \ldots, e_n$ são variáveis **latentes**, i.e., não observadas.
- Portanto, com base no comentário supracitado, é correto expressar as funções objetivos neste formato?

Método de Mínimos Quadrados (MQ) - Ideia

- É extremamente comum em alguns textos as funções objetivos (2) e (3) serem apresentadas como

$$
\begin{aligned}
Q_1(\boldsymbol{\beta}) &= \sum_{i=1}^{n} |y_i - \beta_0 - \beta_1 x_i| = \sum_{i=1}^{n} |e_i| \\
Q_2(\boldsymbol{\beta}) &= \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2 = \sum_{i=1}^{n} e_i^2.
\end{aligned}
$$

- Todavia, é válido lembrar que $e_1, \ldots, e_n$ são variáveis **latentes**, i.e., não observadas.
- Portanto, com base no comentário supracitado, é correto expressar as funções objetivos neste formato?

Método de Mínimos Quadrados (MQ) - Ideia

- É extremamente comum em alguns textos as funções objetivos (2) e (3) serem apresentadas como

$$\begin{aligned} Q_1(\boldsymbol{\beta}) &= \sum_{i=1}^{n} |y_i - \beta_0 - \beta_1 x_i| = \sum_{i=1}^{n} |e_i| \\ Q_2(\boldsymbol{\beta}) &= \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2 = \sum_{i=1}^{n} e_i^2. \end{aligned}$$

- Todavia, é válido lembrar que $e_1, \ldots, e_n$ são variáveis **latentes**, i.e., não observadas. 😵

- Portanto, com base no comentário supracitado, é correto expressar as funções objetivos neste formato? 😐

Método de Mínimos Quadrados (MQ) - Ideia

- O método de estimação $\mathcal{L}_1$, consiste em determinar $\boldsymbol{\beta} = (\beta_0, \beta_1)^\top$ que minimiza (2). Já o método de estimação de Mínimos Quadrados (ordinários)- MQO ou $\mathcal{L}_2$, consiste em determinar $\boldsymbol{\beta} = (\beta_0, \beta_1)^\top$ que minimiza (3).
- Note que o método $\mathcal{L}_1$ é equivalente a obter o estimador de máxima verossimilhança quando considera-se que $e_i \overset{\text{iid}}{\sim} \text{Laplace}(0, \sigma^2)$, implicando que $y_i \overset{\text{ind}}{\sim} \text{Laplace}(\beta_0 + \beta_1 x_i, \sigma^2)$, $i = \ldots, n$.
- Já o método $\mathcal{L}_2$ é equivalente a obter o estimador de máxima verossimilhança quando $e_i \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$, implicando que $y_i \overset{\text{ind}}{\sim} \mathcal{N}(\beta_0 + \beta_1 x_i, \sigma^2)$, $i = \ldots, n$.
- Pelas razões supracitadas, diz-se que o método $\mathcal{L}_1$ é um método **robusto** a presença de valores discrepantes.
- Mas por qual razão o MQO é largamente utilizado frente ao método $\mathcal{L}_1$?

Método de Mínimos Quadrados (MQ) - Ideia

- O método de estimação $\mathcal{L}_1$, consiste em determinar $\boldsymbol{\beta} = (\beta_0, \beta_1)^\top$ que minimiza (2). Já o método de estimação de Mínimos Quadrados (ordinários)- MQO ou $\mathcal{L}_2$, consiste em determinar $\boldsymbol{\beta} = (\beta_0, \beta_1)^\top$ que minimiza (3).
- Note que o método $\mathcal{L}_1$ é equivalente a obter o estimador de máxima verossimilhança quando considera-se que $e_i \overset{\text{iid}}{\sim} \text{Laplace}(0, \sigma^2)$, implicando que $y_i \overset{\text{ind}}{\sim} \text{Laplace}(\beta_0 + \beta_1 x_i, \sigma^2)$, $i = \ldots, n$.
- Já o método $\mathcal{L}_2$ é equivalente a obter o estimador de máxima verossimilhança quando $e_i \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$, implicando que $y_i \overset{\text{ind}}{\sim} \mathcal{N}(\beta_0 + \beta_1 x_i, \sigma^2)$, $i = \ldots, n$.
- Pelas razões supracitadas, diz-se que o método $\mathcal{L}_1$ é um método **robusto** a presença de valores discrepantes.
- Mas por qual razão o MQO é largamente utilizado frente ao método $\mathcal{L}_1$?

Método de Mínimos Quadrados (MQ) - Ideia

- O método de estimação $\mathcal{L}_1$, consiste em determinar $\boldsymbol{\beta} = (\beta_0, \beta_1)^\top$ que minimiza (2). Já o método de estimação de Mínimos Quadrados (ordinários)- MQO ou $\mathcal{L}_2$, consiste em determinar $\boldsymbol{\beta} = (\beta_0, \beta_1)^\top$ que minimiza (3).
- Note que o método $\mathcal{L}_1$ é equivalente a obter o estimador de máxima verossimilhança quando considera-se que $e_i \overset{\text{iid}}{\sim} \text{Laplace}(0, \sigma^2)$, implicando que $y_i \overset{\text{ind}}{\sim} \text{Laplace}(\beta_0 + \beta_1 x_i, \sigma^2)$, $i = \ldots, n$.
- Já o método $\mathcal{L}_2$ é equivalente a obter o estimador de máxima verossimilhança quando $e_i \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$, implicando que $y_i \overset{\text{ind}}{\sim} \mathcal{N}(\beta_0 + \beta_1 x_i, \sigma^2)$, $i = \ldots, n$.
- Pelas razões supracitadas, diz-se que o método $\mathcal{L}_1$ é um método **robusto** a presença de valores discrepantes.
- Mas por qual razão o MQO é largamente utilizado frente ao método $\mathcal{L}_1$?

Método de Mínimos Quadrados (MQ) - Ideia

- O método de estimação $\mathcal{L}_1$, consiste em determinar $\boldsymbol{\beta} = (\beta_0, \beta_1)^\top$ que minimiza (2). Já o método de estimação de Mínimos Quadrados (ordinários)- MQO ou $\mathcal{L}_2$, consiste em determinar $\boldsymbol{\beta} = (\beta_0, \beta_1)^\top$ que minimiza (3).
- Note que o método $\mathcal{L}_1$ é equivalente a obter o estimador de máxima verossimilhança quando considera-se que $e_i \overset{\text{iid}}{\sim} \text{Laplace}(0, \sigma^2)$, implicando que $y_i \overset{\text{ind}}{\sim} \text{Laplace}(\beta_0 + \beta_1 x_i, \sigma^2)$, $i = \ldots, n$.
- Já o método $\mathcal{L}_2$ é equivalente a obter o estimador de máxima verossimilhança quando $e_i \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$, implicando que $y_i \overset{\text{ind}}{\sim} \mathcal{N}(\beta_0 + \beta_1 x_i, \sigma^2)$, $i = \ldots, n$.
- Pelas razões supracitadas, diz-se que o método $\mathcal{L}_1$ é um método **robusto** a presença de valores discrepantes.
- Mas por qual razão o MQO é largamente utilizado frente ao método $\mathcal{L}_1$?

Método de Mínimos Quadrados (MQ) - Ideia

- O método de estimação $\mathcal{L}_1$, consiste em determinar $\boldsymbol{\beta} = (\beta_0, \beta_1)^\top$ que minimiza (2). Já o método de estimação de Mínimos Quadrados (ordinários)- MQO ou $\mathcal{L}_2$, consiste em determinar $\boldsymbol{\beta} = (\beta_0, \beta_1)^\top$ que minimiza (3).
- Note que o método $\mathcal{L}_1$ é equivalente a obter o estimador de máxima verossimilhança quando considera-se que $e_i \overset{\text{iid}}{\sim} \text{Laplace}(0, \sigma^2)$, implicando que $y_i \overset{\text{ind}}{\sim} \text{Laplace}(\beta_0 + \beta_1 x_i, \sigma^2)$, $i = \ldots, n$.
- Já o método $\mathcal{L}_2$ é equivalente a obter o estimador de máxima verossimilhança quando $e_i \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$, implicando que $y_i \overset{\text{ind}}{\sim} \mathcal{N}(\beta_0 + \beta_1 x_i, \sigma^2)$, $i = \ldots, n$.
- Pelas razões supracitadas, diz-se que o método $\mathcal{L}_1$ é um método **robusto** a presença de valores discrepantes.
- Mas por qual razão o MQO é largamente utilizado frente ao método $\mathcal{L}_1$?

Método de Mínimos Quadrados (MQ)

- Como determinar os valores de β_0 e β_1 que minimizam (3)?
- Como a função é diferenciável, vamos tentar encontrar os valores críticos através da equação

$$\left.\frac{\partial}{\partial \boldsymbol{\beta}} Q_2(\boldsymbol{\beta})\right|_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}} = \mathbf{0}.$$

- Ou de forma equivalente, resolver o sistema de equações simultâneas:

$$\begin{cases} \left.\dfrac{\partial}{\partial \beta_0} Q_2(\beta_0, \beta_1)\right|_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}} = 0 \\ \left.\dfrac{\partial}{\partial \beta_1} Q_2(\beta_0, \beta_1)\right|_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}} = 0. \end{cases}$$

Método de Mínimos Quadrados (MQ)

- Como determinar os valores de β_0 e β_1 que minimizam (3)?
- Como a função é diferenciável, vamos tentar encontrar os valores críticos através da equação

$$\frac{\partial}{\partial \boldsymbol{\beta}} Q_2(\boldsymbol{\beta})\bigg|_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}} = \mathbf{0}.$$

- Ou de forma equivalente, resolver o sistema de equações simultâneas:

$$\begin{cases} \dfrac{\partial}{\partial \beta_0} Q_2(\beta_0, \beta_1)\bigg|_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}} = 0 \\ \dfrac{\partial}{\partial \beta_1} Q_2(\beta_0, \beta_1)\bigg|_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}} = 0. \end{cases}$$

Método de Mínimos Quadrados (MQ)

- Como determinar os valores de β_0 e β_1 que minimizam (3)?
- Como a função é diferenciável, vamos tentar encontrar os valores críticos através da equação

$$\frac{\partial}{\partial \boldsymbol{\beta}} Q_2(\boldsymbol{\beta})\bigg|_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}} = \mathbf{0}.$$

- Ou de forma equivalente, resolver o sistema de equações simultâneas:

$$\begin{cases} \dfrac{\partial}{\partial \beta_0} Q_2(\beta_0, \beta_1)\bigg|_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}} = 0 \\ \dfrac{\partial}{\partial \beta_1} Q_2(\beta_0, \beta_1)\bigg|_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}} = 0. \end{cases}$$

Método de Mínimos Quadrados (MQ)

- Para o modelo em questão, tem-se (detalhes no quadro) que

$$
\begin{aligned}
\frac{\partial}{\partial \beta_0} Q_2(\beta_0, \beta_1) &= -2\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i) \\
\frac{\partial}{\partial \beta_1} Q_2(\beta_0, \beta_1) &= -2\sum_{i=1}^{n} x_i(y_i - \beta_0 - \beta_1 x_i).
\end{aligned}
$$

- De forma que o sistema de equações simultâneas que deve ser resolvido é

$$
\begin{cases}
-2\sum_{i=1}^{n}(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i) = 0 \\
-2\sum_{i=1}^{n} x_i(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i) = 0.
\end{cases} \tag{4}
$$

Método de Mínimos Quadrados (MQ)

- Para o modelo em questão, tem-se (detalhes no quadro) que

$$
\begin{aligned}
\frac{\partial}{\partial \beta_0} Q_2(\beta_0, \beta_1) &= -2 \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i) \\
\frac{\partial}{\partial \beta_1} Q_2(\beta_0, \beta_1) &= -2 \sum_{i=1}^{n} x_i (y_i - \beta_0 - \beta_1 x_i).
\end{aligned}
$$

- De forma que o sistema de equações simultâneas que deve ser resolvido é

$$
\begin{cases}
-2 \sum_{i=1}^{n} (y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i) = 0 \\
-2 \sum_{i=1}^{n} x_i (y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i) = 0.
\end{cases} \tag{4}
$$

Sistema de Equações Normais

- Simplificando o sistema (4), detalhes no quadro, temos

$$\begin{cases} n\bar{y}_n - n\widehat{\beta}_0 - n\widehat{\beta}_1\bar{x}_n = 0 \\ \sum_{i=1}^{n} x_i y_i - n\widehat{\beta}_0\bar{x}_n - \widehat{\beta}_1 \sum_{i=1}^{n} x_i^2 = 0. \end{cases}$$

- As equações simultâneas acima, que são equivalentes a (4), são denominadas de **equações normais**. Aqui o termo normal não se refere a distribuição normal e sim ao conceito de ortogonalidade. A razão para isso é que a teoria de mínimos quadrados pode ser desenvolvida por meio de projeções ortogonais.
- **Exercício**: Apresentar a teoria de mínimos quadrados desenvolvida por meio de projeções ortogonais. (Entregar próxima aula). 😇

Sistema de Equações Normais

- Simplificando o sistema (4), detalhes no quadro, temos

$$\begin{cases} n\bar{y}_n - n\widehat{\beta}_0 - n\widehat{\beta}_1\bar{x}_n = 0 \\ \sum_{i=1}^{n} x_i y_i - n\widehat{\beta}_0\bar{x}_n - \widehat{\beta}_1 \sum_{i=1}^{n} x_i^2 = 0. \end{cases}$$

- As equações simultâneas acima, que são equivalentes a (4), são denominadas de **equações normais**. Aqui o termo normal não se refere a distribuição normal e sim ao conceito de ortogonalidade. A razão para isso é que a teoria de mínimos quadrados pode ser desenvolvida por meio de projeções ortogonais.

- **Exercício**: Apresentar a teoria de mínimos quadrados desenvolvida por meio de projeções ortogonais. (Entregar próxima aula).

Sistema de Equações Normais

- Simplificando o sistema (4), detalhes no quadro, temos

$$\begin{cases} n\bar{y}_n - n\widehat{\beta}_0 - n\widehat{\beta}_1\bar{x}_n = 0 \\ \sum_{i=1}^{n} x_i y_i - n\widehat{\beta}_0\bar{x}_n - \widehat{\beta}_1 \sum_{i=1}^{n} x_i^2 = 0. \end{cases}$$

- As equações simultâneas acima, que são equivalentes a (4), são denominadas de **equações normais**. Aqui o termo normal não se refere a distribuição normal e sim ao conceito de ortogonalidade. A razão para isso é que a teoria de mínimos quadrados pode ser desenvolvida por meio de projeções ortogonais.

- **Exercício**: Apresentar a teoria de mínimos quadrados desenvolvida por meio de projeções ortogonais. (Entregar próxima aula).

Sistema de Equações Normais

- Resolvendo a primeira equação de (4), detalhes no quadro, obtemos

$$\widehat{\beta}_0 = \bar{y}_n - \widehat{\beta}_1 \bar{x}_n. \tag{5}$$

- Colocando (5) na segunda equação de (4), para detalhes vide quadro, obtemos

$$\begin{aligned}
\widehat{\beta}_1 &= \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}_n\bar{y}_n}{\sum_{i=1}^{n} x_i^2 - n\bar{x}_n^2} = \frac{\sum_{i=1}^{n}(x_i - \bar{x}_n)(y_i - \bar{y}_n)}{\sum_{i=1}^{n}(x_i - \bar{x}_n)^2} \\
&= \frac{\sum_{i=1}^{n} x_i(y_i - \bar{y}_n)}{\sum_{i=1}^{n}(x_i - \bar{x}_n)^2} = \frac{\sum_{i=1}^{n}(x_i - \bar{x}_n)y_i}{\sum_{i=1}^{n}(x_i - \bar{x}_n)^2} \\
&= \frac{S_{xy}}{S_{xx}},
\end{aligned} \tag{6}$$

em que $S_{xy} := \sum_{i=1}^{n}(x_i - \bar{x}_n)(y_i - \bar{y}_n)$ e $S_{xx} := \sum_{i=1}^{n}(x_i - \bar{x}_n)^2$.

Sistema de Equações Normais

- Resolvendo a primeira equação de (4), detalhes no quadro, obtemos

$$\widehat{\beta}_0 = \bar{y}_n - \widehat{\beta}_1 \bar{x}_n. \tag{5}$$

- Colocando (5) na segunda equação de (4), para detalhes vide quadro, obtemos

$$\begin{aligned}
\widehat{\beta}_1 &= \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}_n\bar{y}_n}{\sum_{i=1}^{n} x_i^2 - n\bar{x}_n^2} = \frac{\sum_{i=1}^{n}(x_i - \bar{x}_n)(y_i - \bar{y}_n)}{\sum_{i=1}^{n}(x_i - \bar{x}_n)^2} \\
&= \frac{\sum_{i=1}^{n} x_i(y_i - \bar{y}_n)}{\sum_{i=1}^{n}(x_i - \bar{x}_n)^2} = \frac{\sum_{i=1}^{n}(x_i - \bar{x}_n)y_i}{\sum_{i=1}^{n}(x_i - \bar{x}_n)^2} \\
&= \frac{S_{xy}}{S_{xx}},
\end{aligned} \tag{6}$$

em que $S_{xy} := \sum_{i=1}^{n}(x_i - \bar{x}_n)(y_i - \bar{y}_n)$ e $S_{xx} := \sum_{i=1}^{n}(x_i - \bar{x}_n)^2$.

Comentário e pergunta

- Note que $\widehat{\beta}_1$ só está definido se existir ao menos dois valores distintos da variável explicativa, i.e., se a variância amostral de $\{x_1, \ldots, x_n\}$ for positiva, ou equivalentemente, se $S_{xx} > 0$. Isto é intuitivo? Faz sentido? Por quê?
- Os pontos críticos $\widehat{\beta}_0 = \overline{Y}_n - \widehat{\beta}_1 \bar{x}_n$ e $\widehat{\beta}_1 = S_{xY}/S_{xx}$, obtidos respectivamente em (5) e (6) são realmente os estimadores de mínimos quadrados (EMQ) de β_0 e β_1, i.e., eles realmente minimizam a função objetivo $Q_2(\beta)$ definida em (3)? Como verificar isso?
- Perceba que no item acima utilizamos Y ao invés de y para evidenciar que é uma **variável aleatória**.

Comentário e pergunta

- Note que $\widehat{\beta}_1$ só está definido se existir ao menos dois valores distintos da variável explicativa, i.e., se a variância amostral de $\{x_1, \ldots, x_n\}$ for positiva, ou equivalentemente, se $S_{xx} > 0$. Isto é intuitivo? Faz sentido? Por quê?
- Os pontos críticos $\widehat{\beta}_0 = \overline{Y}_n - \widehat{\beta}_1\bar{x}_n$ e $\widehat{\beta}_1 = S_{xY}/S_{xx}$, obtidos respectivamente em (5) e (6) são realmente os estimadores de mínimos quadrados (EMQ) de β_0 e β_1, i.e., eles realmente minimizam a função objetivo $Q_2(\boldsymbol{\beta})$ definida em (3)? Como verificar isso?
- Perceba que no item acima utilizamos Y ao invés de y para evidenciar que é uma **variável aleatória**.

Comentário e pergunta

- Note que $\widehat{\beta}_1$ só está definido se existir ao menos dois valores distintos da variável explicativa, i.e., se a variância amostral de $\{x_1, \ldots, x_n\}$ for positiva, ou equivalentemente, se $S_{xx} > 0$. Isto é intuitivo? Faz sentido? Por quê?
- Os pontos críticos $\widehat{\beta}_0 = \overline{Y}_n - \widehat{\beta}_1 \bar{x}_n$ e $\widehat{\beta}_1 = S_{xY}/S_{xx}$, obtidos respectivamente em (5) e (6) são realmente os estimadores de mínimos quadrados (EMQ) de β_0 e β_1, i.e., eles realmente minimizam a função objetivo $Q_2(\boldsymbol{\beta})$ definida em (3)? Como verificar isso?
- Perceba que no item acima utilizamos Y ao invés de y para evidenciar que é uma **variável aleatória**.

Provando que correspondem aos EMQ

- Para provar que (5) e (6) realmente correspondem aos EMQ, i.e., minimizam a função objetivo (3) devemos provar que a matriz Hessiana avaliada nestes pontos

$$\left.\frac{\partial^2 Q_2(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top}\right|_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}}$$

é **positiva definida** (PD).

- A matriz Hessiana é dada por (detalhes no quadro)

$$\begin{aligned}\boldsymbol{H} &= \frac{\partial^2 Q_2(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top} = \begin{pmatrix} \frac{\partial^2 Q_2(\boldsymbol{\beta})}{\partial\beta_0^2} & \frac{\partial^2 Q_2(\boldsymbol{\beta})}{\partial\beta_0\partial\beta_1} \\ \frac{\partial^2 Q_2(\boldsymbol{\beta})}{\partial\beta_1\partial\beta_0} & \frac{\partial^2 Q_2(\boldsymbol{\beta})}{\partial\beta_1^2} \end{pmatrix}. \\ &= 2\begin{pmatrix} n & n\bar{x}_n \\ n\bar{x}_n & \sum_{i=1}^n x_i^2 \end{pmatrix}.\end{aligned}$$

Provando que correspondem aos EMQ

- Para provar que (5) e (6) realmente correspondem aos EMQ, i.e., minimizam a função objetivo (3) devemos provar que a matriz Hessiana avaliada nestes pontos

$$\left.\frac{\partial^2 Q_2(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top}\right|_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}}$$

é **positiva definida** (PD).

- A matriz Hessiana é dada por (detalhes no quadro)

$$\begin{aligned}\boldsymbol{H} &= \frac{\partial^2 Q_2(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top} = \begin{pmatrix} \frac{\partial^2 Q_2(\boldsymbol{\beta})}{\partial\beta_0^2} & \frac{\partial^2 Q_2(\boldsymbol{\beta})}{\partial\beta_0\partial\beta_1} \\ \frac{\partial^2 Q_2(\boldsymbol{\beta})}{\partial\beta_1\partial\beta_0} & \frac{\partial^2 Q_2(\boldsymbol{\beta})}{\partial\beta_1^2} \end{pmatrix}. \\ &= 2\begin{pmatrix} n & n\bar{x}_n \\ n\bar{x}_n & \sum_{i=1}^n x_i^2 \end{pmatrix}.\end{aligned}$$

Provando que correspondem aos EMQ

- Dado que

$$\boldsymbol{H} = \frac{\partial^2 Q_2(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} = 2 \begin{pmatrix} n & n\bar{x}_n \\ n\bar{x}_n & \sum_{i=1}^n x_i^2 \end{pmatrix}.$$

- Como $h_{11} = n > 0$, $h_{22} = \sum_{i=1}^n x_i^2 > 0$ e $|\boldsymbol{H}| = n(\sum_{i=1}^n x_i^2 - n\bar{x}_n^2) = nS_{xx} > 0$, então $\boldsymbol{H}$ é uma matriz **positiva definida**, implicando que $\widehat{\beta}_0$ e $\widehat{\beta}_1$ dados, respectivamente, por (5) e (6) são os valores de β_0 e β_1 que minimizam $Q_2(\boldsymbol{\beta})$, i.e., realmente são os EMQ de β_0 e β_1, respectivamente.

Provando que correspondem aos EMQ

- Dado que

$$\boldsymbol{H} = \frac{\partial^2 Q_2(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} = 2 \begin{pmatrix} n & n\bar{x}_n \\ n\bar{x}_n & \sum_{i=1}^n x_i^2 \end{pmatrix}.$$

- Como $h_{11} = n > 0$, $h_{22} = \sum_{i=1}^n x_i^2 > 0$ e $|\boldsymbol{H}| = n(\sum_{i=1}^n x_i^2 - n\bar{x}_n^2) = nS_{xx} > 0$, então $\boldsymbol{H}$ é uma matriz **positiva definida**, implicando que $\widehat{\beta}_0$ e $\widehat{\beta}_1$ dados, respectivamente, por (5) e (6) são os valores de β_0 e β_1 que minimizam $Q_2(\boldsymbol{\beta})$, i.e., realmente são os EMQ de β_0 e β_1, respectivamente.

Definições

- A reta de regressão ajustada pelo MQ é dada por

$$\widehat{y}_i = \widehat{\beta}_0 + \widehat{\beta}_1 x_i, i, \ldots, n,$$

em que $\widehat{\beta}_0$ e $\widehat{\beta}_1$ representam os EMQ de β_0 e β_1, respectivamente. Note que isso corresponde essencialmente aos valores preditos para a i-ésima observação segundo o MRLS (1).

- Define-se o i-ésimo resíduo **ordinário**, como sendo a diferença entre o valor observado e o valor ajustado para a i-ésima observação, i.e., para $i = 1, \ldots, n$

$$\begin{aligned} \widehat{e}_i &= y_i - \widehat{y}_i \\ &= y_i - (\widehat{\beta}_0 + \widehat{\beta}_1 x_i). \end{aligned}$$

- Veremos posteriormente que os resíduos são primordiais para avaliar a qualidade do ajuste do modelo adotado. ☺

Definições

- A reta de regressão ajustada pelo MQ é dada por

$$\widehat{y}_i = \widehat{\beta}_0 + \widehat{\beta}_1 x_i, i, \ldots, n,$$

em que $\widehat{\beta}_0$ e $\widehat{\beta}_1$ representam os EMQ de β_0 e β_1, respectivamente. Note que isso corresponde essencialmente aos valores preditos para a i-ésima observação segundo o MRLS (1).

- Define-se o i-ésimo resíduo **ordinário**, como sendo a diferença entre o valor observado e o valor ajustado para a i-ésima observação, i.e., para $i = 1, \ldots, n$

$$\begin{aligned} \widehat{e}_i &= y_i - \widehat{y}_i \\ &= y_i - (\widehat{\beta}_0 + \widehat{\beta}_1 x_i). \end{aligned}$$

- Veremos posteriormente que os resíduos são primordiais para avaliar a qualidade do ajuste do modelo adotado. ☺

Definições

- A reta de regressão ajustada pelo MQ é dada por

$$\widehat{y}_i = \widehat{\beta}_0 + \widehat{\beta}_1 x_i, i, \ldots, n,$$

em que $\widehat{\beta}_0$ e $\widehat{\beta}_1$ representam os EMQ de β_0 e β_1, respectivamente. Note que isso corresponde essencialmente aos valores preditos para a i-ésima observação segundo o MRLS (1).

- Define-se o i-ésimo resíduo **ordinário**, como sendo a diferença entre o valor observado e o valor ajustado para a i-ésima observação, i.e., para $i = 1, \ldots, n$

$$\begin{aligned} \widehat{e}_i &= y_i - \widehat{y}_i \\ &= y_i - (\widehat{\beta}_0 + \widehat{\beta}_1 x_i). \end{aligned}$$

- Veremos posteriormente que os resíduos são primordiais para avaliar a qualidade do ajuste do modelo adotado. ☺

Propriedades dos EMQ

Considere o MRLS (1) e $\widehat{\boldsymbol{\beta}} = (\widehat{\beta}_0, \widehat{\beta}_1)^\top$ o EMQ do vetor de coeficientes de regressão. Então,

P1. $\widehat{\beta}_0$ e $\widehat{\beta}_1$ são combinações lineares das observações $y_1, \ldots, y_n$, i.e.,

$$\begin{aligned}\widehat{\beta}_1 &= \frac{S_{xy}}{S_{xx}} = \sum_{i=1}^{n} C_{1i} y_i \\ \widehat{\beta}_0 &= \bar{y}_n - \widehat{\beta}_1 \bar{x}_n = \sum_{i=1}^{n} C_{0i} y_i.\end{aligned}$$

P2. Os EMQ são não viesados, i.e.,

$$\mathbb{E}\left[\widehat{\beta}_0\right] = \beta_0 \text{ e } \mathbb{E}\left[\widehat{\beta}_1\right] = \beta_1,$$

com respectivas variâncias e covariância

$$\begin{aligned}\mathrm{Var}\left[\widehat{\beta}_0\right] &= \sigma^2\left(\frac{1}{n} + \frac{\bar{x}_n^2}{S_{xx}}\right), \ \mathrm{Var}\left[\widehat{\beta}_1\right] = \frac{\sigma^2}{S_{xx}} \text{ e} \\ \mathrm{Cov}\left(\widehat{\beta}_0, \widehat{\beta}_1\right) &= -\frac{\sigma^2 \bar{x}_n}{S_{xx}}.\end{aligned}$$