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**, setembro 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

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

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

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

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

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

- 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

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

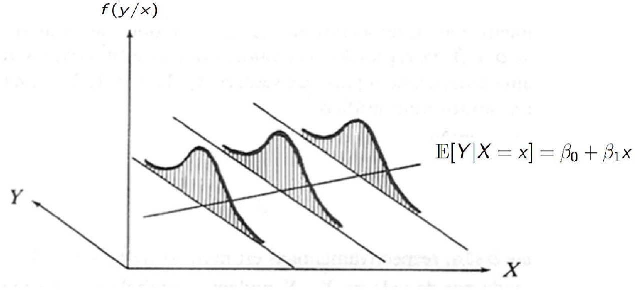
- Lembrando, que **correlação nula** implica independência sob a suposição de normalidade multivariada, então usando as suposições iv) e v) adicionada com a suposição acima, temos

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

- 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

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

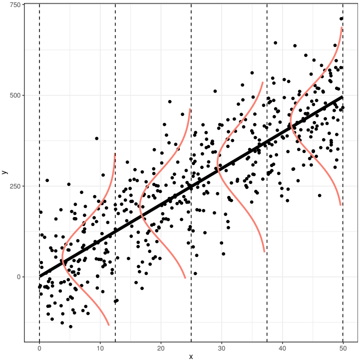
MRLS - Ilustração gráfica



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).

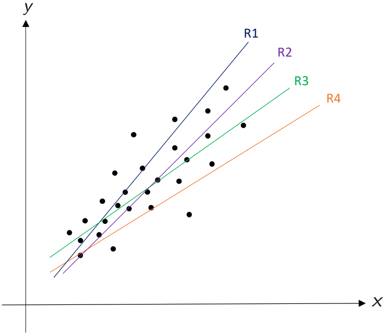
ii) Peso vs altura.

iii) Tempo de processamento vs # de faturas.

iv) Faturamento da empresa vs investimento com propaganda.

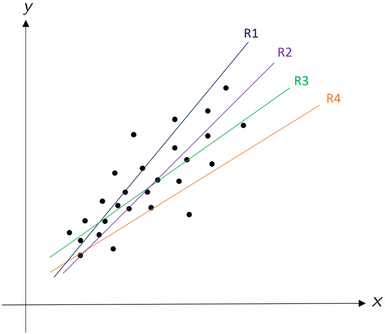
v) Pressão arterial sistólica (mmHg) vs idade (anos).

Método de Mínimos Quadrados (MQ)



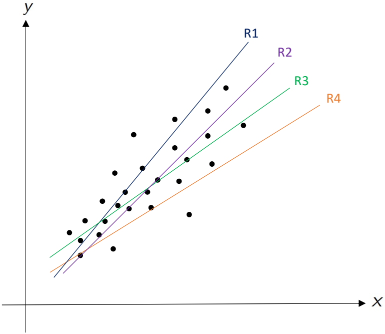
- Dado um conjunto de dados, existem **infinitas** retas candidatas para ajuste.
- Qual delas escolher?

Método de Mínimos Quadrados (MQ)



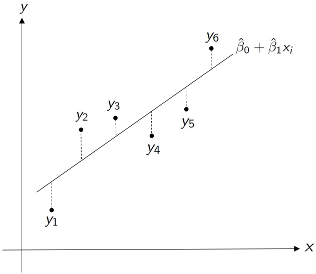
- Dado um conjunto de dados, existem **infinitas** retas candidatas para ajuste.
- Qual delas escolher?

Método de Mínimos Quadrados (MQ)



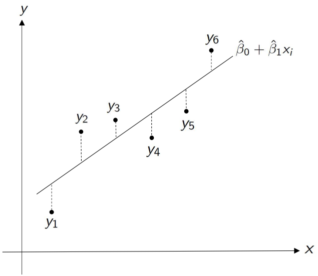
- Dado um conjunto de dados, existem **infinitas** retas candidatas para ajuste.
- Qual delas escolher?

Método de Mínimos Quadrados (MQ) - Ideia



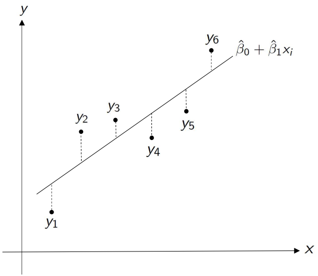
- A melhor reta estimada será aquela que minimiza a distância do valor observado y_i para o valor esperado ajustado $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$, $i = 1, \ldots, n$.
- Infelizmente, não é possível minimizar todas estas distâncias **simultaneamente**, logo, considera-se alguma função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

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

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

- 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

- 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 $e_i \stackrel{\text{iid}}{\sim} \text{Laplace}(0, \sigma^2)$, implicando que $y_i \stackrel{\text{ind}}{\sim} \text{Laplace}(\beta_0 + \beta_1 x_i, \sigma^2)$, $i = \ldots, n$.
- Já o método $\mathcal{L}_2$ é equivalente a obter o estimador de máxima verossimilhança quando $e_i \stackrel{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$, implicando que $y_i \stackrel{\text{ind}}{\sim} \mathcal{N}(\beta_0 + \beta_1 x_i, \sigma^2)$, $i = \ldots, n$.
- 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

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

- 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

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

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

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

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

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

- 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

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

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

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

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

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

Exercícios (entregar próxima aula)

Exercício 2: Considerando adicionalmente que $e_i \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2), i = 1, \ldots, n$, determine o EMV de $\boldsymbol{\beta} = (\beta_0, \beta_1)^\top$, $\widehat{\boldsymbol{\beta}}$ e sua distribuição exata.

Exercício 3: Sob a suposição de normalidade, encontre as estatísticas suficientes minimais considerando o MRLS. Elas são completas?

Exercício 4: Forneça condições suficientes para que os EMQs sejam consistentes.

Propriedades dos EMQ (Cont.)

P3. $\sum_{i=1}^{n}(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i) = \sum_{i=1}^{n}(y_i - \widehat{y}_i) = \sum_{i=1}^{n}\widehat{e}_i = 0.$

P4. Por P3, conclui-se diretamente que

$$\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} \widehat{y}_i.$$

P5. A reta de regressão ajustada ($\widehat{y}_i = \widehat{\beta}_0 + \widehat{\beta}_1 x_i$) sempre passa pelo centróide dos dados, que corresponde ao ponto $(\bar{x}_n, \bar{y}_n)$.

P6. $\sum_{i=1}^{n} x_i \widehat{e}_i = 0$, i.e., a soma dos resíduos ordinários ponderados pelos correspondentes valores da variável explicativa é igual a zero.

P7. $\sum_{i=1}^{n} \widehat{y}_i \widehat{e}_i = 0$, i.e., a soma dos resíduos ordinários ponderados pelos correspondentes valores ajustados é igual a zero.

Propriedades dos EMQ - Teorema de Gauss-Markov

P8. (Teorema de Gauss-Markov) Considere o MRLS com suas pressuposições básicas (exceto a de normalidade). Os EMQs $\widehat{\beta}_0$ e $\widehat{\beta}_1$ são os melhores estimadores lineares não viesados (BLUE-Best Linear Unbiased Estimator) de β_0 e β_1, respectivamente, i.e., dentre todos os estimadores lineares não viesados de β_0 e β_1, $\widehat{\beta}_0$ e $\widehat{\beta}_1$ são os que possuem a menor variância.

Dem: Ver no quadro... 😉

∎

- Por que **teorema de Gauss-Markov**, dado que não são contemporâneos?
- Gauss obteve o resultado sob a suposição de independência e normalidade, enquanto Markov reduziu as suposições de forma a obter da forma apresentada aqui.

Propriedades dos EMQ - Teorema de Gauss-Markov

P8. (Teorema de Gauss-Markov) Considere o MRLS com suas pressuposições básicas (exceto a de normalidade). Os EMQs $\widehat{\beta}_0$ e $\widehat{\beta}_1$ são os melhores estimadores lineares não viesados (BLUE-Best Linear Unbiased Estimator) de β_0 e β_1, respectivamente, i.e., dentre todos os estimadores lineares não viesados de β_0 e β_1, $\widehat{\beta}_0$ e $\widehat{\beta}_1$ são os que possuem a menor variância.

Dem: Ver no quadro... 😉

■

- Por que **teorema de Gauss-Markov**, dado que não são contemporâneos?
- Gauss obteve o resultado sob a suposição de independência e normalidade, enquanto Markov reduziu as suposições de forma a obter da forma apresentada aqui.

Propriedades dos EMQ - Teorema de Gauss-Markov

P8. (Teorema de Gauss-Markov) Considere o MRLS com suas pressuposições básicas (exceto a de normalidade). Os EMQs $\widehat{\beta}_0$ e $\widehat{\beta}_1$ são os melhores estimadores lineares não viesados (BLUE-Best Linear Unbiased Estimator) de β_0 e β_1, respectivamente, i.e., dentre todos os estimadores lineares não viesados de β_0 e β_1, $\widehat{\beta}_0$ e $\widehat{\beta}_1$ são os que possuem a menor variância.

Dem: Ver no quadro... 😉

■

- Por que **teorema de Gauss-Markov**, dado que não são contemporâneos?
- Gauss obteve o resultado sob a suposição de independência e normalidade, enquanto Markov reduziu as suposições de forma a obter da forma apresentada aqui.

Estimação de σ^2

- A forma funcional do MRLS é dada por

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

em que $e_i \sim \mathrm{RB}(0, \sigma^2)$.

- Lembrando que o resíduo é um **preditor** da fonte de variação, então é razoável obter um estimador de σ^2 que seja função do vetor de resíduos ordinários, i.e.,

$$\widehat{\sigma}^2 = \phi(\widehat{\boldsymbol{e}}).$$

- Por outro lado, sob as suposições básicas do MRLS, tem-se que (exercício)

$$\widehat{e}_i = y_i - \widehat{y}_i \sim \left(0, \sigma^2 \left\{1 - \frac{1}{n} - \frac{(x_i - \bar{x}_n)^2}{S_{xx}}\right\}\right).$$

- Como σ^2 representa a variância (segundo momento) da fonte de variação, então um estimador ingênuo (naive estimator) é

$$\sum_{i=1}^{n} \widehat{e}_i^2 = \mathrm{SQRes},$$

que é denominado de Soma de Quadrados dos Resíduos - SQRes.

Estimação de σ^2

- A forma funcional do MRLS é dada por

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

em que $e_i \sim \mathrm{RB}(0, \sigma^2)$.

- Lembrando que o resíduo é um **preditor** da fonte de variação, então é razoável obter um estimador de σ^2 que seja função do vetor de resíduos ordinários, i.e.,

$$\widehat{\sigma}^2 = \phi(\widehat{\boldsymbol{e}}).$$

- Por outro lado, sob as suposições básicas do MRLS, tem-se que (exercício)

$$\widehat{e}_i = y_i - \widehat{y}_i \sim \left(0, \sigma^2\left\{1 - \frac{1}{n} - \frac{(x_i - \bar{x}_n)^2}{S_{xx}}\right\}\right).$$

- Como σ^2 representa a variância (segundo momento) da fonte de variação, então um estimador ingênuo (naive estimator) é

$$\sum_{i=1}^{n} \widehat{e}_i^2 = \mathrm{SQRes},$$

que é denominado de Soma de Quadrados dos Resíduos - SQRes.

Estimação de σ^2

- A forma funcional do MRLS é dada por

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

em que $e_i \sim \mathrm{RB}(0, \sigma^2)$.

- Lembrando que o resíduo é um **preditor** da fonte de variação, então é razoável obter um estimador de σ^2 que seja função do vetor de resíduos ordinários, i.e.,

$$\widehat{\sigma}^2 = \phi(\widehat{\boldsymbol{e}}).$$

- Por outro lado, sob as suposições básicas do MRLS, tem-se que (exercício)

$$\widehat{e}_i = y_i - \widehat{y}_i \sim \left(0, \sigma^2 \left\{1 - \frac{1}{n} - \frac{(x_i - \bar{x}_n)^2}{S_{xx}}\right\}\right).$$

- Como σ^2 representa a variância (segundo momento) da fonte de variação, então um estimador ingênuo (naive estimator) é

$$\sum_{i=1}^{n} \widehat{e}_i^2 = \mathrm{SQRes},$$

que é denominado de Soma de Quadrados dos Resíduos - SQRes.

Estimação de σ^2

- A forma funcional do MRLS é dada por

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

em que $e_i \sim \mathrm{RB}(0, \sigma^2)$.

- Lembrando que o resíduo é um **preditor** da fonte de variação, então é razoável obter um estimador de σ^2 que seja função do vetor de resíduos ordinários, i.e.,

$$\widehat{\sigma}^2 = \phi(\widehat{\boldsymbol{e}}).$$

- Por outro lado, sob as suposições básicas do MRLS, tem-se que (exercício)

$$\widehat{e}_i = y_i - \widehat{y}_i \sim \left(0, \sigma^2 \left\{1 - \frac{1}{n} - \frac{(x_i - \bar{x}_n)^2}{S_{xx}}\right\}\right).$$

- Como σ^2 representa a variância (segundo momento) da fonte de variação, então um estimador ingênuo (naive estimator) é

$$\sum_{i=1}^{n} \widehat{e}_i^2 = \mathrm{SQRes},$$

que é denominado de Soma de Quadrados dos Resíduos - SQRes.

Estimação de σ^2

- Dado que

$$\sum_{i=1}^{n} \mathbb{E}[\widehat{e}_i^2] = \sum_{i=1}^{n} \sigma^2 \left\{ 1 - \frac{1}{n} - \frac{(x_i - \bar{x}_n)^2}{S_{xx}} \right\} = \sigma^2 (n-2).$$

- Então um estimador não viesado para σ^2 é dado por

$$\widehat{\sigma}^2 = \frac{\mathrm{SQRes}}{(n-2)} := \mathrm{QMRes}.$$

- Além de não viesado, o QMRes é consistente e também representa o MINQUE (MInimum Norm Quadratic Unbiased Estimador) de σ^2.
- A teoria do MINQUE foi desenvolvida por C.R. Rao na década de 1970.
  - Rao, C.R. (1970). Estimation of heteroscedastic variances in linear models. *Journal of the American Statistical Association*, 65, 161–172.
  - Rao, C.R. (1971). Estimation of variance and covariance components MINQUE theory. *Journal of Multivariate Analysis*, 1, 257–275.

Estimação de σ^2

- Dado que

$$\begin{aligned}\sum_{i=1}^{n} \mathbb{E}[\widehat{e}_i^2] &= \sum_{i=1}^{n} \sigma^2 \left\{1 - \frac{1}{n} - \frac{(x_i - \bar{x}_n)^2}{S_{xx}}\right\} \\ &= \sigma^2(n-2).\end{aligned}$$

- Então um estimador não viesado para σ^2 é dado por

$$\widehat{\sigma}^2 = \frac{\text{SQRes}}{(n-2)} := \text{QMRes}.$$

- Além de não viesado, o QMRes é consistente e também representa o MINQUE (MInimum Norm Quadratic Unbiased Estimador) de σ^2.

- A teoria do MINQUE foi desenvolvida por C.R. Rao na década de 1970.
  - Rao, C.R. (1970). Estimation of heteroscedastic variances in linear models. *Journal of the American Statistical Association*, 65, 161–172.
  - Rao, C.R. (1971). Estimation of variance and covariance components MINQUE theory. *Journal of Multivariate Analysis*, 1, 257–275.

Estimação de σ^2

- Dado que

$$\sum_{i=1}^{n} \mathbb{E}[\widehat{e}_i^2] = \sum_{i=1}^{n} \sigma^2 \left\{1 - \frac{1}{n} - \frac{(x_i - \bar{x}_n)^2}{S_{xx}}\right\} = \sigma^2(n-2).$$

- Então um estimador não viesado para σ^2 é dado por

$$\widehat{\sigma}^2 = \frac{\mathrm{SQRes}}{(n-2)} := \mathrm{QMRes}.$$

- Além de não viesado, o QMRes é consistente e também representa o MINQUE (MInimum Norm Quadratic Unbiased Estimador) de σ^2.

- A teoria do MINQUE foi desenvolvida por C.R. Rao na década de 1970.
    - Rao, C.R. (1970). Estimation of heteroscedastic variances in linear models. *Journal of the American Statistical Association*, **65**, 161–172.
    - Rao, C.R. (1971). Estimation of variance and covariance components MINQUE theory. *Journal of Multivariate Analysis*, **1**, 257–275.

Estimação de σ^2

- Dado que

$$\begin{aligned}\sum_{i=1}^{n}\mathbb{E}[\widehat{e}_i^2] &= \sum_{i=1}^{n}\sigma^2\left\{1-\frac{1}{n}-\frac{(x_i-\bar{x}_n)^2}{S_{xx}}\right\}\\ &= \sigma^2(n-2).\end{aligned}$$

- Então um estimador não viesado para σ^2 é dado por

$$\widehat{\sigma}^2=\frac{\mathrm{SQRes}}{(n-2)}:=\mathrm{QMRes}.$$

- Além de não viesado, o QMRes é consistente e também representa o MINQUE (MInimum Norm Quadratic Unbiased Estimador) de σ^2.
- A teoria do MINQUE foi desenvolvida por C.R. Rao na década de 1970.
  - Rao, C.R. (1970). Estimation of heteroscedastic variances in linear models. *Journal of the American Statistical Association*, **65**, 161–172.
  - Rao, C.R. (1971). Estimation of variance and covariance components MINQUE theory. *Journal of Multivariate Analysis*, **1**, 257–275.

Estimação de σ^2

- Dado que

$$\begin{aligned}\sum_{i=1}^{n} \mathbb{E}[\widehat{e}_i^2] &= \sum_{i=1}^{n} \sigma^2 \left\{1 - \frac{1}{n} - \frac{(x_i - \bar{x}_n)^2}{S_{xx}}\right\} \\ &= \sigma^2(n-2).\end{aligned}$$

- Então um estimador não viesado para σ^2 é dado por

$$\widehat{\sigma}^2 = \frac{\mathrm{SQRes}}{(n-2)} := \mathrm{QMRes}.$$

- Além de não viesado, o QMRes é consistente e também representa o MINQUE (MInimum Norm Quadratic Unbiased Estimador) de σ^2.
- A teoria do MINQUE foi desenvolvida por C.R. Rao na década de 1970.
  - Rao, C.R. (1970). Estimation of heteroscedastic variances in linear models. *Journal of the American Statistical Association*, **65**, 161–172.
  - Rao, C.R. (1971). Estimation of variance and covariance components MINQUE theory. *Journal of Multivariate Analysis*, **1**, 257–275.

Estimação de σ^2

- Dado que

$$\begin{aligned}\sum_{i=1}^{n} \mathbb{E}[\widehat{e}_i^2] &= \sum_{i=1}^{n} \sigma^2 \left\{1 - \frac{1}{n} - \frac{(x_i - \bar{x}_n)^2}{S_{xx}}\right\} \\ &= \sigma^2(n-2).\end{aligned}$$

- Então um estimador não viesado para σ^2 é dado por

$$\widehat{\sigma}^2 = \frac{\mathrm{SQRes}}{(n-2)} := \mathrm{QMRes}.$$

- Além de não viesado, o QMRes é consistente e também representa o MINQUE (MInimum Norm Quadratic Unbiased Estimador) de σ^2.
- A teoria do MINQUE foi desenvolvida por C.R. Rao na década de 1970.
  - Rao, C.R. (1970). Estimation of heteroscedastic variances in linear models. *Journal of the American Statistical Association*, **65**, 161–172.
  - Rao, C.R. (1971). Estimation of variance and covariance components MINQUE theory. *Journal of Multivariate Analysis*, **1**, 257–275.

Estimação de σ^2

- Em geral, $\widehat{\sigma}$ é denominado por **erro-padrão da regressão**.
- É comum, estimar as variâncias e erros-padrão dos EMQs através de

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

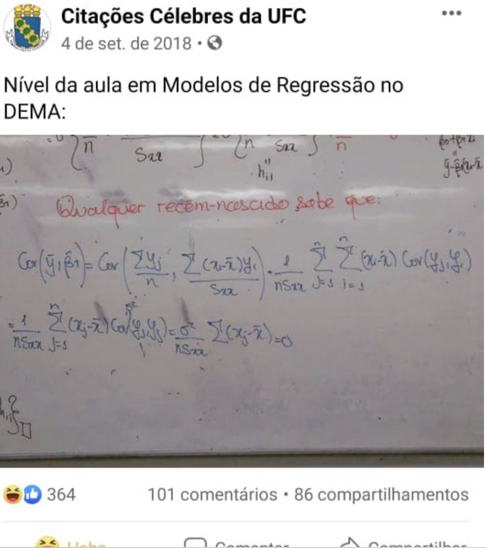
Estimação de σ^2

- Em geral, $\widehat{\sigma}$ é denominado por **erro-padrão da regressão**.
- É comum, estimar as variâncias e erros-padrão dos EMQs através de

$$\widehat{\mathrm{Var}}\left[\widehat{\beta}_0\right] = \widehat{\sigma}^2\left(\frac{1}{n}+\frac{\bar{x}_n^2}{S_{xx}}\right) \text{ e } \widehat{\mathrm{EP}}(\widehat{\beta}_0) = \widehat{\sigma}\sqrt{\left(\frac{1}{n}+\frac{\bar{x}_n^2}{S_{xx}}\right)}$$

$$\widehat{\mathrm{Var}}\left[\widehat{\beta}_1\right] = \frac{\widehat{\sigma}^2}{S_{xx}} \text{ e } \widehat{\mathrm{EP}}(\widehat{\beta}_1) = \frac{\widehat{\sigma}}{\sqrt{S_{xx}}}.$$

De um dia desses... ☺ 😎



Uso da variável centralizada

- Já vimos que algumas situações o parâmetro que representa o intercepto β_0 não possui interpretação **prática**.
- Afim de contornar este inconveniente, é comum centralizar a variável explicativa na forma $x_i - \bar{x}_n$, obtendo o modelo

$$\begin{aligned} y_i &= \beta_0 + \beta_1 x_i + e_i = \beta_0 + \beta_1(x_i + \bar{x}_n - \bar{x}_n) + e_i \\ &= \beta_0 + \beta_1 \bar{x}_n + \beta_1(x_i - \bar{x}_n) + e_i \\ &= \beta_0^* + \beta_1(x_i - \bar{x}_n) + e_i, \end{aligned} \tag{7}$$

perceba que β_0^* sofre apenas uma translação, enquanto β_1 fica inalterado.

Uso da variável centralizada

- Já vimos que algumas situações o parâmetro que representa o intercepto β_0 não possui interpretação **prática**.
- Afim de contornar este inconveniente, é comum centralizar a variável explicativa na forma $x_i - \bar{x}_n$, obtendo o modelo

$$\begin{aligned} y_i &= \beta_0 + \beta_1 x_i + e_i = \beta_0 + \beta_1 (x_i + \bar{x}_n - \bar{x}_n) + e_i \\ &= \beta_0 + \beta_1 \bar{x}_n + \beta_1 (x_i - \bar{x}_n) + e_i \\ &= \beta_0^* + \beta_1 (x_i - \bar{x}_n) + e_i, \end{aligned} \tag{7}$$

perceba que β_0^* sofre apenas uma translação, enquanto β_1 fica inalterado.

Exemplo

Exemplo 1: Considere que se tem interesse em estudar a associação entre idade (x em anos) e a pressão arterial sistólica (y em mmHg), considere o seguinte MRLS

$$y_i = \beta_0^* + \beta_1(x_i - \bar{x}_n) + e_i,$$

em que $\bar{x}_n = 55$.

i) Interprete o parâmetro β_0^* sem usar o jargão estatístico.

ii) Se ao invés do modelo acima, se considerarmos o seguinte MRLS:

$$y_i = \beta_0^{**} + \beta_1(x_i - 50) + e_i.$$

Discuta qual a diferença existente na interpretação entre β_0^{**} do modelo acima e o intercepto do MRLS original. E a interpretação do parâmetro β_1 sofre alguma alteração?

Uso da variável centralizada

- Os valores ajustados e consequentemente o ajuste não se modificam (demonstração abaixo). Agora, neste modelo é sempre possível interpretar o parâmetro relativo ao intercepto, desde que

$$\beta_0^* = \mathbb{E}[y_i | X = \bar{x}_n].$$

- O EMQ de $\boldsymbol{\beta}^* = (\beta_0^*, \beta_1)^\top$ no modelo (7) são dados por

$$\widehat{\beta}_0^* = \overline{Y}_n \text{ e } \widehat{\beta}_1 = S_{xY}/S_{xx}.$$

- Para demonstrar que os valores ajustados não se modificam pelo fato de considerarmos as variáveis centralizadas (pode ser em $\bar{x}_n$ como em um valor $a \in \mathbb{R}$ qualquer), basta perceber que i-ésimo valor ajustado sob o MRLS (7) é dado por

$$\begin{aligned} \widehat{y}_i &= \overline{Y}_n + \widehat{\beta}_1(x_i - \bar{x}_n) = (\overline{Y}_n - \widehat{\beta}_1\bar{x}_n) + \widehat{\beta}_1 x_i \\ &= \widehat{\beta}_0 + \widehat{\beta}_1 x_i. \end{aligned}$$

Uso da variável centralizada

- Os valores ajustados e consequentemente o ajuste não se modificam (demonstração abaixo). Agora, neste modelo é sempre possível interpretar o parâmetro relativo ao intercepto, desde que

$$\beta_0^* = \mathbb{E}[y_i | X = \bar{x}_n].$$

- O EMQ de $\boldsymbol{\beta}^* = (\beta_0^*, \beta_1)^\top$ no modelo (7) são dados por

$$\widehat{\beta}_0^* = \overline{Y}_n \text{ e } \widehat{\beta}_1 = S_{xY}/S_{xx}.$$

- Para demonstrar que os valores ajustados não se modificam pelo fato de considerarmos as variáveis centralizadas (pode ser em $\bar{x}_n$ como em um valor $a \in \mathbb{R}$ qualquer), basta perceber que i-ésimo valor ajustado sob o MRLS (7) é dado por

$$\begin{aligned} \widehat{y}_i &= \overline{Y}_n + \widehat{\beta}_1(x_i - \bar{x}_n) = (\overline{Y}_n - \widehat{\beta}_1 \bar{x}_n) + \widehat{\beta}_1 x_i \\ &= \widehat{\beta}_0 + \widehat{\beta}_1 x_i. \end{aligned}$$

Uso da variável centralizada

- Os valores ajustados e consequentemente o ajuste não se modificam (demonstração abaixo). Agora, neste modelo é sempre possível interpretar o parâmetro relativo ao intercepto, desde que

$$\beta_0^* = \mathbb{E}[y_i | X = \bar{x}_n].$$

- O EMQ de $\boldsymbol{\beta}^* = (\beta_0^*, \beta_1)^\top$ no modelo (7) são dados por

$$\widehat{\beta}_0^* = \overline{Y}_n \text{ e } \widehat{\beta}_1 = S_{xY} / S_{xx}.$$

- Para demonstrar que os valores ajustados não se modificam pelo fato de considerarmos as variáveis centralizadas (pode ser em $\bar{x}_n$ como em um valor $a \in \mathbb{R}$ qualquer), basta perceber que i-ésimo valor ajustado sob o MRLS (7) é dado por

$$\begin{aligned} \widehat{y}_i &= \overline{Y}_n + \widehat{\beta}_1(x_i - \bar{x}_n) = (\overline{Y}_n - \widehat{\beta}_1 \bar{x}_n) + \widehat{\beta}_1 x_i \\ &= \widehat{\beta}_0 + \widehat{\beta}_1 x_i. \end{aligned}$$

Vantagens da centralização

- Possibilidade de interpretar o parâmetro de intercepto β_0^*, desde que centralizado em um valor que pertença a amplitude amostral $(\min\{x_1, \ldots, x_n\}, \ldots, \max\{x_1, \ldots, x_n\})$.
- Centralizando em $\bar{x}_n$, tem-se que $\mathrm{Cov}\left(\widehat{\beta}_0^*, \widehat{\beta}_1\right) = 0$, de forma que sob normalidade implica **independência** dos estimadores. Qual a vantagem disso?
- Na verdade, sob normalidade, ao centralizar a variável explicativa em $\bar{x}_n$, β_0 e β_1 são **ortogonais**.
- Independente de centralização, sob normalidade, temos que $\boldsymbol{\beta}$ e σ são ortogonais.

Vantagens da centralização

- Possibilidade de interpretar o parâmetro de intercepto β_0^*, desde que centralizado em um valor que pertença a amplitude amostral $(\min\{x_1, \ldots, x_n\}, \ldots, \max\{x_1, \ldots, x_n\})$.
- Centralizando em $\bar{x}_n$, tem-se que $\mathrm{Cov}\left(\widehat{\beta}_0^*, \widehat{\beta}_1\right) = 0$, de forma que sob normalidade implica **independência** dos estimadores. Qual a vantagem disso?
- Na verdade, sob normalidade, ao centralizar a variável explicativa em $\bar{x}_n$, β_0 e β_1 são **ortogonais**.
- Independente de centralização, sob normalidade, temos que $\boldsymbol{\beta}$ e σ são ortogonais.

Vantagens da centralização

- Possibilidade de interpretar o parâmetro de intercepto β_0^*, desde que centralizado em um valor que pertença a amplitude amostral $(\min\{x_1, \ldots, x_n\}, \ldots, \max\{x_1, \ldots, x_n\})$.
- Centralizando em $\bar{x}_n$, tem-se que $\mathrm{Cov}\left(\widehat{\beta}_0^*, \widehat{\beta}_1\right) = 0$, de forma que sob normalidade implica **independência** dos estimadores. Qual a vantagem disso?
- Na verdade, sob normalidade, ao centralizar a variável explicativa em $\bar{x}_n$, β_0 e β_1 são **ortogonais**.
- Independente de centralização, sob normalidade, temos que $\boldsymbol{\beta}$ e σ são ortogonais.

Vantagens da centralização

- Possibilidade de interpretar o parâmetro de intercepto β_0^*, desde que centralizado em um valor que pertença a amplitude amostral $(\min\{x_1, \ldots, x_n\}, \ldots, \max\{x_1, \ldots, x_n\})$.
- Centralizando em $\bar{x}_n$, tem-se que $\mathrm{Cov}\left(\widehat{\beta}_0^*, \widehat{\beta}_1\right) = 0$, de forma que sob normalidade implica **independência** dos estimadores. Qual a vantagem disso?
- Na verdade, sob normalidade, ao centralizar a variável explicativa em $\bar{x}_n$, β_0 e β_1 são **ortogonais**.
- Independente de centralização, sob normalidade, temos que $\boldsymbol{\beta}$ e σ são ortogonais.

Decomposição da Soma de Quadrados Total

Considere o desvio de cada observação em torno da média geral $\bar{y}_n$:

$$(y_i - \bar{y}_n),$$

que pode ser reescrito como

$$(y_i - \bar{y}_n) = (y_i - \widehat{y}_i) + (\widehat{y}_i - \bar{y}_n), \forall i = 1, \ldots, n, \tag{8}$$

em que:

- $\widehat{e}_i := y_i - \widehat{y}_i$: Representa a distância entre o valor observado e o valor predito pelo modelo de regressão (resíduo ordinário).
- $\widehat{y}_i - \bar{y}_n$: Representa a distância entre o valor predito pelo método de regressão e o valor predito se as variáveis aleatórias fossem iid, i.e., se não houvesse regressão.

Decomposição da Soma de Quadrados Total

Considere o desvio de cada observação em torno da média geral $\bar{y}_n$:

$$(y_i - \bar{y}_n),$$

que pode ser reescrito como

$$(y_i - \bar{y}_n) = (y_i - \widehat{y}_i) + (\widehat{y}_i - \bar{y}_n), \forall i = 1, \ldots, n, \tag{8}$$

em que:

- $\widehat{e}_i := y_i - \widehat{y}_i$: Representa a distância entre o valor observado e o valor predito pelo modelo de regressão (resíduo ordinário).
- $\widehat{y}_i - \bar{y}_n$: Representa a distância entre o valor predito pelo método de regressão e o valor predito se as variáveis aleatórias fossem iid, i.e., se não houvesse regressão.

Decomposição da Soma de Quadrados Total

Considere o desvio de cada observação em torno da média geral $\bar{y}_n$:

$$(y_i - \bar{y}_n),$$

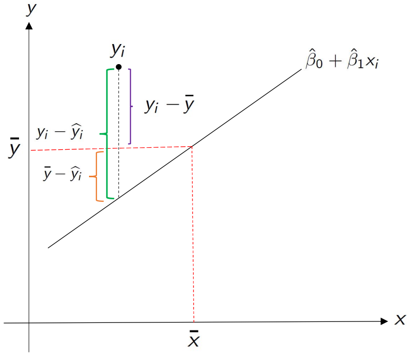
que pode ser reescrito como

$$(y_i - \bar{y}_n) = (y_i - \widehat{y}_i) + (\widehat{y}_i - \bar{y}_n), \forall i = 1, \ldots, n, \tag{8}$$

em que:

- $\widehat{e}_i := y_i - \widehat{y}_i$: Representa a distância entre o valor observado e o valor predito pelo modelo de regressão (resíduo ordinário).
- $\widehat{y}_i - \bar{y}_n$: Representa a distância entre o valor predito pelo método de regressão e o valor predito se as variáveis aleatórias fossem iid, i.e., se não houvesse regressão.

Ilustração gráfica da decomposição



Decomposição da Soma de Quadrados Total

Elevando ao quadrado em (8), obtemos

$$(y_i - \bar{y}_n)^2 = \{\widehat{e}_i + (\widehat{y}_i - \bar{y}_n)\}^2, \forall i = 1, \ldots, n,$$

de forma que

$$\sum_{i=1}^{n}(y_i - \bar{y}_n)^2 = \sum_{i=1}^{n}\widehat{e}_i^2 + \sum_{i=1}^{n}(\widehat{y}_i - \bar{y}_n)^2 + 2\sum_{i=1}^{n}\widehat{e}_i(\widehat{y}_i - \bar{y}_n).$$

Fato: $\sum_{i=1}^{n}\widehat{e}_i(\widehat{y}_i - \bar{y}_n) = 0.$

Dem.: Dado que $\sum_{i=1}^{n}\widehat{e}_i = \sum_{i=1}^{n}x_i\widehat{e}_i = 0$, então

$$\begin{aligned}\sum_{i=1}^{n}\widehat{e}_i(\widehat{y}_i - \bar{y}_n) &= \sum_{i=1}^{n}(\widehat{\beta}_0 + \widehat{\beta}_1 x_i)\widehat{e}_i - \bar{y}\sum_{i=1}^{n}\widehat{e}_i \\ &= \widehat{\beta}_0\sum_{i=1}^{n}\widehat{e}_i + \widehat{\beta}_1\sum_{i=1}^{n}x_i\widehat{e}_i = 0.\end{aligned}$$

QED ■

Decomposição da Soma de Quadrados Total

Logo,

$$\sum_{i=1}^{n}(y_i - \bar{y}_n)^2 = \sum_{i=1}^{n}\widehat{e}_i^2 + \sum_{i=1}^{n}(\widehat{y}_i - \bar{y}_n)^2. \tag{9}$$

Em geral, denotamos:

- $\mathrm{SQT} := \sum_{i=1}^{n}(y_i - \bar{y}_n)^2$: **Soma de Quadrados Total**, que representa a variação total das observações $y_1, \ldots, y_n$ em torno de sua média aritmética.
- $\mathrm{SQRes} := \sum_{i=1}^{n}\widehat{e}_i^2$: **Soma de Quadrados de Resíduos**, que representa a parcela da variabilidade da variável resposta que não é explicada pelo modelo de regressão.
- $\mathrm{SQReg} := \sum_{i=1}^{n}(\widehat{y}_i - \bar{y}_n)^2$: **Soma de Quadrados de Regressão**, que representa a parcela da variabilidade da variável resposta que é explicada pelo modelo de regressão.
- Situação desejável: $\mathrm{SQReg} \approx \mathrm{SQT}$. 😇

Decomposição da Soma de Quadrados Total

Logo,

$$\sum_{i=1}^{n}(y_i - \bar{y}_n)^2 = \sum_{i=1}^{n}\widehat{e}_i^2 + \sum_{i=1}^{n}(\widehat{y}_i - \bar{y}_n)^2. \tag{9}$$

Em geral, denotamos:

- $\mathrm{SQT} := \sum_{i=1}^{n}(y_i - \bar{y}_n)^2$: **Soma de Quadrados Total**, que representa a variação total das observações $y_1, \ldots, y_n$ em torno de sua média aritmética.
- $\mathrm{SQRes} := \sum_{i=1}^{n}\widehat{e}_i^2$: **Soma de Quadrados de Resíduos**, que representa a parcela da variabilidade da variável resposta que não é explicada pelo modelo de regressão.
- $\mathrm{SQReg} := \sum_{i=1}^{n}(\widehat{y}_i - \bar{y}_n)^2$: **Soma de Quadrados de Regressão**, que representa a parcela da variabilidade da variável resposta que é explicada pelo modelo de regressão.
- Situação desejável: $\mathrm{SQReg} \approx \mathrm{SQT}$.

Decomposição da Soma de Quadrados Total

Logo,

$$\sum_{i=1}^{n}(y_i - \bar{y}_n)^2 = \sum_{i=1}^{n}\widehat{e}_i^2 + \sum_{i=1}^{n}(\widehat{y}_i - \bar{y}_n)^2. \tag{9}$$

Em geral, denotamos:

- $\mathrm{SQT} := \sum_{i=1}^{n}(y_i - \bar{y}_n)^2$: **Soma de Quadrados Total**, que representa a variação total das observações $y_1, \ldots, y_n$ em torno de sua média aritmética.
- $\mathrm{SQRes} := \sum_{i=1}^{n}\widehat{e}_i^2$: **Soma de Quadrados de Resíduos**, que representa a parcela da variabilidade da variável resposta que não é explicada pelo modelo de regressão.
- $\mathrm{SQReg} := \sum_{i=1}^{n}(\widehat{y}_i - \bar{y}_n)^2$: **Soma de Quadrados de Regressão**, que representa a parcela da variabilidade da variável resposta que é explicada pelo modelo de regressão.
- Situação desejável: $\mathrm{SQReg} \approx \mathrm{SQT}$.

Decomposição da Soma de Quadrados Total

Logo,

$$\sum_{i=1}^{n}(y_i - \bar{y}_n)^2 = \sum_{i=1}^{n}\widehat{e}_i^2 + \sum_{i=1}^{n}(\widehat{y}_i - \bar{y}_n)^2. \tag{9}$$

Em geral, denotamos:

- $\mathrm{SQT} := \sum_{i=1}^{n}(y_i - \bar{y}_n)^2$: **Soma de Quadrados Total**, que representa a variação total das observações $y_1, \ldots, y_n$ em torno de sua média aritmética.
- $\mathrm{SQRes} := \sum_{i=1}^{n}\widehat{e}_i^2$: **Soma de Quadrados de Resíduos**, que representa a parcela da variabilidade da variável resposta que não é explicada pelo modelo de regressão.
- $\mathrm{SQReg} := \sum_{i=1}^{n}(\widehat{y}_i - \bar{y}_n)^2$: **Soma de Quadrados de Regressão**, que representa a parcela da variabilidade da variável resposta que é explicada pelo modelo de regressão.
- Situação desejável: $\mathrm{SQReg} \approx \mathrm{SQT}$.

Decomposição da Soma de Quadrados Total

Logo,

$$\sum_{i=1}^{n}(y_i - \bar{y}_n)^2 = \sum_{i=1}^{n} \widehat{e}_i^2 + \sum_{i=1}^{n}(\widehat{y}_i - \bar{y}_n)^2. \tag{9}$$

Em geral, denotamos:

- $\mathrm{SQT} := \sum_{i=1}^{n}(y_i - \bar{y}_n)^2$: **Soma de Quadrados Total**, que representa a variação total das observações $y_1, \ldots, y_n$ em torno de sua média aritmética.
- $\mathrm{SQRes} := \sum_{i=1}^{n} \widehat{e}_i^2$: **Soma de Quadrados de Resíduos**, que representa a parcela da variabilidade da variável resposta que não é explicada pelo modelo de regressão.
- $\mathrm{SQReg} := \sum_{i=1}^{n}(\widehat{y}_i - \bar{y}_n)^2$: **Soma de Quadrados de Regressão**, que representa a parcela da variabilidade da variável resposta que é explicada pelo modelo de regressão.
- Situação desejável: $\mathrm{SQReg} \approx \mathrm{SQT}$.

Coeficiente de determinação

O **coeficiente de determinação** é definido por

$$R^2 := \frac{\mathrm{SQReg}}{\mathrm{SQT}} = \frac{\sum_{i=1}^{n}(\widehat{y}_i - \bar{y}_n)^2}{S_{yy}}, \tag{10}$$

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

Note que por (9), tem-se que

$$0 \leq R^2 \leq 1,$$

e quanto mais próximo de **um** maior indicativo de boa qualidade do ajuste. Todavia como veremos posteriormente, o R^2 **sozinho** não garante um bom ajuste, mesmo assumindo valores "altos".

Coeficiente de determinação

- A decomposição (9)

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

  da forma apresentada só é válida em modelos que satisfazem

$$\sum_{i=1}^{n} \widehat{e}_i(\widehat{y}_i - \bar{y}_n) = 0,$$

  em particular isso é válido em modelos de regressão lineares que **possuem** intercepto.

- A decomposição acima é um caso particular da classe de decomposições H (Hoeffding, 1948, "A Class of Statistics with Asymptotically Normal Distribution", Ann. Math. Statist. 19, https://doi.org/10.1214/aoms/1177730196) de Estatísticas U.

Coeficiente de determinação

- A decomposição (9)

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

da forma apresentada só é válida em modelos que satisfazem

$$\sum_{i=1}^{n} \widehat{e}_i(\widehat{y}_i - \bar{y}_n) = 0,$$

em particular isso é válido em modelos de regressão lineares que **possuem** intercepto.

- A decomposição acima é um caso particular da classe de decomposições H (Hoeffding, 1948, "A Class of Statistics with Asymptotically Normal Distribution", Ann. Math. Statist. 19, https://doi.org/10.1214/aoms/1177730196) de Estatísticas *U*.

Coeficiente de determinação

Perceba que

$$\begin{aligned}\mathrm{SQReg} &= \sum_{i=1}^{n}(\widehat{y}_i - \bar{y}_n)^2 = \sum_{i=1}^{n}(\bar{y}_n - \widehat{\beta}_1\bar{x}_n + \widehat{\beta}_1 x_i - \bar{y}_n)^2 \\ &= \widehat{\beta}_1^2\sum_{i=1}^{n}(x_i - \bar{x}_n)^2 = \widehat{\beta}_1^2 S_{xx} = \widehat{\beta}_1 S_{xy},\end{aligned}$$

implicando que

$$\begin{aligned}R^2 := \frac{\mathrm{SQReg}}{\mathrm{SQT}} &= \widehat{\beta}_1^2\frac{S_{xx}}{S_{yy}} = \frac{S_{xy}^2}{S_{xx}^2}\frac{S_{xx}}{S_{yy}} \\ &= \left(\frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}\right)^2 = r_{xy}^2,\end{aligned}$$

i.e., o coeficiente de determinação nada mais é que o quadrado do coeficiente de correlação (amostral) linear de Pearson de y e x. Faz sentido? 😎

Ideia

Ideia: Utilizar a decomposição (9) de SQT para testar se existe regressão, i.e., para testar

$$\mathcal{H}_0 : \beta_1 = 0 \ \text{ versus } \ \mathcal{H}_1 : \beta_1 \neq 0.$$

Exercício - Entregar próxima aula

Exercício: Considere o MRLS

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

com suas pressuposições básicas. Mostre que

i) $\mathbb{E}[\mathrm{SQT}] = (n-1)\sigma^2 + \beta_1^2 S_{xx}$.

ii) $\mathbb{E}[\mathrm{SQReg}] = \beta_1^2 S_{xx} + \sigma^2$.

iii) $\mathbb{E}[\mathrm{SQRes}] = (n-2)\sigma^2$.

Obs.: Logo, pelos resultados acima temos que

a) O $\mathrm{QMRes} := \mathrm{SQRes}/(n-2)$ é um estimador não viesado de σ^2 como já mostramos, na verdade ele é o MINQUE de σ^2, como já discutido em aulas anteriores.

b) Sob $\mathcal{H}_0 : \beta_1 = 0$, o $\mathrm{QMReg} = \mathrm{SQReg}$ também é um estimador não viesado de σ^2.

Distribuição da Soma de Quadrados

Teorema de Cochran

Sob o MRLS adicionada a suposição de normalidade, i.e., $e_i \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$ e sob $\mathcal{H}_0 : \beta_1 = 0$, temos que

i) SQRes e SQReg são independentes.

ii) $\dfrac{\text{SQT}}{\sigma^2} \sim \chi^2_{(n-1)}$.

iii) $\dfrac{\text{SQRes}}{\sigma^2} \sim \chi^2_{(n-2)}$.

iv) $\dfrac{\text{SQReg}}{\sigma^2} \sim \chi^2_{1}$.

Dem.: Lista.

Teorema de Cochran

Dem.: Considerando a suposição de normalidade e sob $\mathcal{H}_0 : \beta_1 = 0$, temos que $Y_1, \ldots, Y_n \overset{\text{iid}}{\sim} \mathcal{N}(\beta_0, \sigma^2)$, de forma que já obtemos diretamente (basta lembrar de Inferência) que

$$\frac{\mathrm{SQT}}{\sigma^2} = \frac{(n-1)S_y^2}{\sigma^2} \sim \chi^2_{(n-1)}.$$

Por outro lado, perceba que $\mathrm{SQReg} = \widehat{\beta}_1^2 S_{xx} = \phi(\widehat{\beta}_1)$, em que $\widehat{\beta}_1 \sim \mathcal{N}(0, \sigma^2/S_{xx})$ e $\mathrm{SQRes} = \sum_{i=1}^{n} \widehat{e}_i^2 = \boldsymbol{\varphi}(\widehat{e}_1, \ldots, \widehat{e}_n)$, em que $\widehat{\boldsymbol{e}} \sim \mathcal{N}_n(\mathbf{0}, \boldsymbol{\Sigma}_e)$. Dado que $\forall i = 1, \ldots, n$

$$\begin{aligned}
\mathrm{Cov}(\widehat{\beta}_1, \widehat{e}_i) &= \mathrm{Cov}(\widehat{\beta}_1, y_i - \widehat{y}_i) = \mathrm{Cov}(\widehat{\beta}_1, y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i) \\
&= \frac{\sigma^2(x_i - \bar{x}_n)}{S_{xx}} + \frac{\sigma^2 \bar{x}_n}{S_{xx}} - \frac{\sigma^2 x_i}{S_{xx}} \\
&= 0.
\end{aligned}$$

E como, $\forall i = 1, \ldots, n$, $\widehat{\beta}_1$ e $\widehat{e}_i$ tem distribuição normal bivariada (funções lineares das observações), então são independentes, pois neste caso, correlação nula, implica independência. Lembrando que qualquer função mensurável de variáveis independentes também são independentes, então SQReg e SQRes são independentes.

Teorema de Cochran

Usando o fato de que sob $\mathcal{H}_0 : \beta_1 = 0$, $\widehat{\beta}_1 \sim \mathcal{N}(0, \sigma^2/S_{xx})$, então segue diretamente que

$$\frac{\mathrm{SQReg}}{\sigma^2} = \frac{\widehat{\beta}_1^2 S_{xx}}{\sigma^2} \sim \chi^2_1.$$

Lembrando que $\dfrac{\mathrm{SQT}}{\sigma^2} = \dfrac{(n-1)S_y^2}{\sigma^2} \sim \chi^2_{(n-1)}$ e que SQReg e SQRes são independentes, então usando funções geradoras de momentos (fazer no quadro), conclui-se que

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

QED ■

ANOVA

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

Causas de Variação	GL	SQ	QM
Regressão	1	$\sum_{i=1}^{n}(\widehat{y}_i - \bar{y}_n)^2$	$\mathrm{SQReg}/1 = \mathrm{QMReg}$
Resíduo	$n-2$	$\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2$	$\mathrm{SQRes}/(n-2) = \mathrm{QMRes}$
Total	$n-1$	$\sum_{i=1}^{n}(y_i - \bar{y}_n)^2$	$\mathrm{SQT}/(n-1)$

Consistência do MINQUE

Teorema

Sob o MRLS, sem necessidade de normalidade, temos que

$$\mathrm{QMRes} \xrightarrow{\mathbb{P}} \sigma^2,$$

i.e., o QMRes é um estimador consistente de σ^2.

Dem.: Lista.

Estatística de teste

Teorema

Sob o MRLS, com a suposição de normalidade, temos que sob $\mathcal{H}_0 : \beta_1 = 0$

$$F_0 = \frac{\text{QMReg}}{\text{QMRes}} \sim \mathcal{F}(1, n-2).$$

Dem.: Segue imediatamente do teorema de Cochran. ■

- Em geral, acrescenta-se uma coluna no quadro de ANOVA com o valor da estatística F_0, além do valor-p associado (que veremos em breve qual o teste relacionado). 😛

Estatística de teste

Teorema

Sob o MRLS, com a suposição de normalidade, temos que sob $\mathcal{H}_0 : \beta_1 = 0$

$$F_0 = \frac{\text{QMReg}}{\text{QMRes}} \sim \mathcal{F}(1, n-2).$$

Dem.: Segue imediatamente do teorema de Cochran. ■

- Em geral, acrescenta-se uma coluna no quadro de ANOVA com o valor da estatística F_0, além do valor-p associado (que veremos em breve qual o teste relacionado).

Estatística de teste

- Sob a hipótese alternativa $\mathcal{H}_1 : \beta_1 \neq 0$, temos $\widehat{\beta}_1 \sim \mathcal{N}(\beta_1, \sigma^2/S_{xx})$.
- Logo, $\text{SQReg} := \sum_{i=1}^{n}(\widehat{y}_i - \bar{y}_n)^2 = \frac{\widehat{\beta}_1^2 S_{xx}}{\sigma^2} \sim \chi^2_{(1, \beta_1^2 S_{xx}/\sigma^2)}$, em que $\chi^2_{(k,\lambda)}$ representa uma variável aleatória com distribuição qui-quadrado não-central com k gl, em que λ é denominado parâmetro de não-centralidade.
- De forma análoga, sob a hipótese alternativa $\mathcal{H}_1 : \beta_1 \neq 0$

$$F_0 = \frac{\text{QMReg}}{\text{QMRes}} \sim \mathcal{F}(1, n-2, \lambda),$$

representando uma variável aleatória com distribuição F de parâmetros $(1, n-2)$ e parâmetro de não-centralidade $\lambda = (\beta_1^2 S_{xx})/\sigma^2$.

Estatística de teste

- Sob a hipótese alternativa $\mathcal{H}_1 : \beta_1 \neq 0$, temos $\widehat{\beta}_1 \sim \mathcal{N}(\beta_1, \sigma^2/S_{xx})$.
- Logo, $\text{SQReg} := \sum_{i=1}^{n}(\widehat{y}_i - \bar{y}_n)^2 = \frac{\widehat{\beta}_1^2 S_{xx}}{\sigma^2} \sim \chi^2_{(1,\beta_1^2 S_{xx}/\sigma^2)}$, em que $\chi^2_{(k,\lambda)}$ representa uma variável aleatória com distribuição qui-quadrado não-central com k gl, em que λ é denominado parâmetro de não-centralidade.
- De forma análoga, sob a hipótese alternativa $\mathcal{H}_1 : \beta_1 \neq 0$

$$F_0 = \frac{\text{QMReg}}{\text{QMRes}} \sim \mathcal{F}(1, n-2, \lambda),$$

representando uma variável aleatória com distribuição F de parâmetros $(1, n-2)$ e parâmetro de não-centralidade $\lambda = (\beta_1^2 S_{xx})/\sigma^2$.

Estatística de teste

- Sob a hipótese alternativa $\mathcal{H}_1 : \beta_1 \neq 0$, temos $\widehat{\beta}_1 \sim \mathcal{N}(\beta_1, \sigma^2/S_{xx})$.
- Logo, $\mathrm{SQReg} := \sum_{i=1}^{n}(\widehat{y}_i - \bar{y}_n)^2 = \dfrac{\widehat{\beta}_1^2 S_{xx}}{\sigma^2} \sim \chi^2_{(1, \beta_1^2 S_{xx}/\sigma^2)}$, em que $\chi^2_{(k,\lambda)}$ representa uma variável aleatória com distribuição qui-quadrado não-central com k gl, em que λ é denominado parâmetro de não-centralidade.
- De forma análoga, sob a hipótese alternativa $\mathcal{H}_1 : \beta_1 \neq 0$

$$F_0 = \frac{\mathrm{QMReg}}{\mathrm{QMRes}} \sim \mathcal{F}(1, n-2, \lambda),$$

representando uma variável aleatória com distribuição F de parâmetros $(1, n-2)$ e parâmetro de não-centralidade $\lambda = (\beta_1^2 S_{xx})/\sigma^2$.

Estatística de teste

Portanto, para testar

$$\mathcal{H}_0 : \beta_1 = 0 \ \text{ versus } \ \mathcal{H}_1 : \beta_1 \neq 0.$$

a um determinado nível de significância $\alpha \in (0, 1)$, pode-se utilizar a estatística F oriunda da ANOVA. Neste caso, rejeita-se $\mathcal{H}_0$ ao nível α se

$$F_0 > \mathcal{F}_{1-\alpha}(1, n-2),$$

desde que sob $\mathcal{H}_1$, tem-se $\mathbb{E}[\mathrm{QMReg}] > \mathbb{E}[\mathrm{QMRes}] = \sigma^2$. O valor-p neste caso é dado por

$$\mathbb{P}[\mathcal{F}(1, n-2) > F_0].$$

Coeficiente de determinação corrigido

Já vimos que $R^2 = r_{xy}^2$ que é uma função decrescente de n. Por quê?

Dado que

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

a ideia é corrigir pelos respectivos graus de liberdade, de forma obter

$$1 - \overline{R}^2 := \frac{\mathrm{QMRes}}{\mathrm{QMTotal}} = \frac{\mathrm{SQRes}/(n-2)}{\mathrm{SQT}/(n-1)} = \frac{n-1}{n-2}(1-R^2),$$

implicando que

$$\begin{aligned}
\overline{R}^2 &= 1 - \frac{n-1}{n-2}(1-R^2) = \frac{(n-2)-(n-1)(1-R^2)}{n-2} \\
&= \frac{(n-2)-[(n-2)(1-R^2)+(1-R^2)]}{n-2} \\
&= 1-(1-R^2) - \frac{(1-R^2)}{n-2} = R^2 - \frac{(1-R^2)}{n-2}.
\end{aligned}$$

O coeficiente de determinação ajustado também possui outras propriedades interessantes como veremos no decorrer do curso.

Exercício - Entregar próxima aula

Exercício: Considere o MRLS

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

com suas pressuposições básicas e considerando também a suposição de normalidade. Mostre que $\widehat{\boldsymbol{\beta}}$ e QMRes são independentes.

Sugestão.: Utilize a mesma ideia usada na demonstração do teorema de Cochran.

Obs.: Nas próximas aulas será apresentado a estrutura de ajustes de MRLS e obtenção do quadro de ANOVA no software `R`.

Distribuição do EMQ

Sob os pressupostos do MRLS, inclusive a de normalidade, temos que

$$\widehat{\boldsymbol{\beta}} = \begin{pmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \end{pmatrix} \sim \mathcal{N}_2 \left[\begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} ; \sigma^2 \begin{pmatrix} \frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} & -\frac{\bar{x}}{S_{xx}} \\ \cdot & \frac{1}{S_{xx}} \end{pmatrix} \right].$$

Dado que $\widehat{\boldsymbol{\beta}}$ e SQRes são independentes, com

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

então,

$$t_0 := \frac{\dfrac{\widehat{\beta}_0 - \beta_0}{\sqrt{\sigma^2 \left(\dfrac{1}{n} + \dfrac{\bar{x}^2}{S_{xx}} \right)}}}{\sqrt{\dfrac{\mathrm{SQRes}}{\sigma^2 (n-2)}}} = \frac{\widehat{\beta}_0 - \beta_0}{\sqrt{\mathrm{QMRes} \left(\dfrac{1}{n} + \dfrac{\bar{x}^2}{S_{xx}} \right)}} \sim t_{(n-2)}. \tag{11}$$

Distribuição do EMQ

Similarmente, temos que

$$t_1 := \frac{\dfrac{\widehat{\beta}_1 - \beta_1}{\sqrt{\dfrac{\sigma^2}{S_{xx}}}}}{\sqrt{\dfrac{\mathrm{SQRes}}{\sigma^2(n-2)}}} = \frac{\widehat{\beta}_1 - \beta_1}{\sqrt{\dfrac{\mathrm{QMRes}}{S_{xx}}}} \sim t_{(n-2)}. \tag{12}$$

- Note que (11) e (12) são quantidades pivotais, para β_0 e β_1, respectivamente.
- Desta forma, podemos utilizar o método da quantidade pivotal para determinar ICs com nível de confiança $(1-\alpha)$ para os parâmetros de interesse.
- Além disso, como as distribuições das quantidades pivotais são **simétricas**, podemos encontrar **facilmente** o **melhor** IC com nível de confiança $(1-\alpha)$, melhor no sentido de possuir o **menor** comprimento, para β_0 e β_1.

Distribuição do EMQ

Similarmente, temos que

$$t_1 := \frac{\dfrac{\widehat{\beta}_1 - \beta_1}{\sqrt{\dfrac{\sigma^2}{S_{xx}}}}}{\sqrt{\dfrac{\mathrm{SQRes}}{\sigma^2(n-2)}}} = \frac{\widehat{\beta}_1 - \beta_1}{\sqrt{\dfrac{\mathrm{QMRes}}{S_{xx}}}} \sim t_{(n-2)}. \tag{12}$$

- Note que (11) e (12) são quantidades pivotais, para β_0 e β_1, respectivamente.
- Desta forma, podemos utilizar o método da quantidade pivotal para determinar ICs com nível de confiança $(1-\alpha)$ para os parâmetros de interesse.
- Além disso, como as distribuições das quantidades pivotais são **simétricas**, podemos encontrar **facilmente** o **melhor** IC com nível de confiança $(1-\alpha)$, melhor no sentido de possuir o **menor** comprimento, para β_0 e β_1.

Distribuição do EMQ

Similarmente, temos que

$$t_1 := \frac{\dfrac{\widehat{\beta}_1 - \beta_1}{\sqrt{\dfrac{\sigma^2}{S_{xx}}}}}{\sqrt{\dfrac{\mathrm{SQRes}}{\sigma^2(n-2)}}} = \frac{\widehat{\beta}_1 - \beta_1}{\sqrt{\dfrac{\mathrm{QMRes}}{S_{xx}}}} \sim t_{(n-2)}. \tag{12}$$

- Note que (11) e (12) são quantidades pivotais, para β_0 e β_1, respectivamente.
- Desta forma, podemos utilizar o método da quantidade pivotal para determinar ICs com nível de confiança $(1-\alpha)$ para os parâmetros de interesse.
- Além disso, como as distribuições das quantidades pivotais são **simétricas**, podemos encontrar **facilmente** o **melhor** IC com nível de confiança $(1-\alpha)$, melhor no sentido de possuir o **menor** comprimento, para β_0 e β_1.

Distribuição do EMQ

Similarmente, temos que

$$t_1 := \frac{\dfrac{\widehat{\beta}_1 - \beta_1}{\sqrt{\dfrac{\sigma^2}{S_{xx}}}}}{\sqrt{\dfrac{\mathrm{SQRes}}{\sigma^2(n-2)}}} = \frac{\widehat{\beta}_1 - \beta_1}{\sqrt{\dfrac{\mathrm{QMRes}}{S_{xx}}}} \sim t_{(n-2)}. \tag{12}$$

- Note que (11) e (12) são quantidades pivotais, para β_0 e β_1, respectivamente.
- Desta forma, podemos utilizar o método da quantidade pivotal para determinar ICs com nível de confiança $(1-\alpha)$ para os parâmetros de interesse.
- Além disso, como as distribuições das quantidades pivotais são **simétricas**, podemos encontrar **facilmente** o **melhor** IC com nível de confiança $(1-\alpha)$, melhor no sentido de possuir o **menor** comprimento, para β_0 e β_1.

Intervalos de confiança ótimos

Portanto, os ICs de comprimento mínimo para β_0 e β_1 com nível $1-\alpha$, são dados respectivamente, por:

$$\mathrm{IC}_{1-\alpha}(\beta_0) = \left[\widehat{\beta}_0 \pm t_{(1-\alpha/2,n-2)}\sqrt{\mathrm{QMRes}\left(\frac{1}{n}+\frac{\bar{x}^2}{S_{xx}}\right)}\right]$$

e

$$\mathrm{IC}_{1-\alpha}(\beta_1) = \left[\widehat{\beta}_1 \pm t_{(1-\alpha/2,n-2)}\sqrt{\frac{\mathrm{QMRes}}{S_{xx}}}\right],$$

em que $t_{(1-\alpha/2,n-2)}$ representa o quantil de ordem $1-\alpha/2$ de uma distribuição $t_{(n-2)}$.

- Se $n \to \infty$ podemos utilizar a distribuição normal como referência, i.e., trocar $t_{(1-\alpha/2,n-2)}$ por $z_{(1-\alpha/2)}$. Por qual razão pode-se fazer isto?
- Se a suposição de normalidade não for atendida, os ICs acima não serão **exatos**, todavia, continuam válidos assintoticamente pelo Teorema Central do Limite de Hájek-Šidak.
- Se a amostra não for grande, podemos utilizar procedimentos de reamostragem, como Jackniffe e Bootstrap, por exemplo, para obter ICs **exatos**.

Intervalos de confiança ótimos

Portanto, os ICs de comprimento mínimo para β_0 e β_1 com nível $1-\alpha$, são dados respectivamente, por:

$$\mathrm{IC}_{1-\alpha}(\beta_0) = \left[\widehat{\beta}_0 \pm t_{(1-\alpha/2,n-2)}\sqrt{\mathrm{QMRes}\left(\frac{1}{n}+\frac{\bar{x}^2}{S_{xx}}\right)}\right]$$

e

$$\mathrm{IC}_{1-\alpha}(\beta_1) = \left[\widehat{\beta}_1 \pm t_{(1-\alpha/2,n-2)}\sqrt{\frac{\mathrm{QMRes}}{S_{xx}}}\right],$$

em que $t_{(1-\alpha/2,n-2)}$ representa o quantil de ordem $1-\alpha/2$ de uma distribuição $t_{(n-2)}$.

- Se $n \to \infty$ podemos utilizar a distribuição normal como referência, i.e., trocar $t_{(1-\alpha/2,n-2)}$ por $z_{(1-\alpha/2)}$. Por qual razão pode-se fazer isto?
- Se a suposição de normalidade não for atendida, os ICs acima não serão **exatos**, todavia, continuam válidos assintoticamente pelo Teorema Central do Limite de Hájek-Šidak.
- Se a amostra não for grande, podemos utilizar procedimentos de reamostragem, como Jackniffe e Bootstrap, por exemplo, para obter ICs **exatos**.

Intervalos de confiança ótimos

Portanto, os ICs de comprimento mínimo para β_0 e β_1 com nível $1-\alpha$, são dados respectivamente, por:

$$\mathrm{IC}_{1-\alpha}(\beta_0) = \left[\widehat{\beta}_0 \pm t_{(1-\alpha/2,n-2)}\sqrt{\mathrm{QMRes}\left(\frac{1}{n}+\frac{\bar{x}^2}{S_{xx}}\right)}\right]$$

e

$$\mathrm{IC}_{1-\alpha}(\beta_1) = \left[\widehat{\beta}_1 \pm t_{(1-\alpha/2,n-2)}\sqrt{\frac{\mathrm{QMRes}}{S_{xx}}}\right],$$

em que $t_{(1-\alpha/2,n-2)}$ representa o quantil de ordem $1-\alpha/2$ de uma distribuição $t_{(n-2)}$.

- Se $n \to \infty$ podemos utilizar a distribuição normal como referência, i.e., trocar $t_{(1-\alpha/2,n-2)}$ por $z_{(1-\alpha/2)}$. Por qual razão pode-se fazer isto?
- Se a suposição de normalidade não for atendida, os ICs acima não serão **exatos**, todavia, continuam válidos assintoticamente pelo Teorema Central do Limite de Hájek-Šidak.
- Se a amostra não for grande, podemos utilizar procedimentos de reamostragem, como Jackniffe e Bootstrap, por exemplo, para obter ICs **exatos**.

Intervalos de confiança ótimos

Portanto, os ICs de comprimento mínimo para β_0 e β_1 com nível $1-\alpha$, são dados respectivamente, por:

$$\mathrm{IC}_{1-\alpha}(\beta_0) = \left[\widehat{\beta}_0 \pm t_{(1-\alpha/2,n-2)}\sqrt{\mathrm{QMRes}\left(\frac{1}{n}+\frac{\bar{x}^2}{S_{xx}}\right)}\right]$$

e

$$\mathrm{IC}_{1-\alpha}(\beta_1) = \left[\widehat{\beta}_1 \pm t_{(1-\alpha/2,n-2)}\sqrt{\frac{\mathrm{QMRes}}{S_{xx}}}\right],$$

em que $t_{(1-\alpha/2,n-2)}$ representa o quantil de ordem $1-\alpha/2$ de uma distribuição $t_{(n-2)}$.

- Se $n \to \infty$ podemos utilizar a distribuição normal como referência, i.e., trocar $t_{(1-\alpha/2,n-2)}$ por $z_{(1-\alpha/2)}$. Por qual razão pode-se fazer isto?
- Se a suposição de normalidade não for atendida, os ICs acima não serão **exatos**, todavia, continuam válidos assintoticamente pelo Teorema Central do Limite de Hájek-Šidak.
- Se a amostra não for grande, podemos utilizar procedimentos de reamostragem, como Jackniffe e Bootstrap, por exemplo, para obter ICs **exatos**.

Exercício - Entregar próxima aula

Exercício: Considere o MRLS considerando válida a suposição de normalidade. Usando o método da quantidade pivotal, determine um IC de nível $1-\alpha$ para σ^2. Discuta a obtenção do IC de comprimento mínimo de nível $1-\alpha$, i.e., o IC **ótimo**. Se a suposição de normalidade não for atendida, como proceder para determinar um IC para σ^2 de forma **exata** ou ao menos assintótica?

Testes de hipóteses marginais

Podemos utilizar as quantidades pivotais (11) e (12) como estatísticas de testes para testar, respectivamente, as hipóteses

$$\mathcal{H}_0 : \beta_0 = b_0 \text{ versus } \mathcal{H}_1 : \beta_0 \neq (>,<) b_0, \; b_0 \text{ especificado}$$

e

$$\mathcal{H}_0 : \beta_1 = b_1 \text{ versus } \mathcal{H}_1 : \beta_1 \neq (>,<) b_1, \; b_1 \text{ especificado.}$$

Tais testes representam na verdade o teste da razão de verossimilhanças, i.e., não caíram do céu.

Teste para o intercepto

- Para testar $\mathcal{H}_0 : \beta_0 = b_0$ vs $\mathcal{H}_1 : \beta_0 \neq b_0$, b_0 especificado, podemos usar a estatística de teste

$$t_0 := \frac{\widehat{\beta}_0 - b_0}{\sqrt{\mathrm{QMRes}\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)}}.$$

- Sob $\mathcal{H}_0 : \beta_0 = b_0$, temos que $t_0 \sim t_{(n-2)}$.
- Rejeita-se $\mathcal{H}_0$ ao nível α se $|t_0| > t_{(1-\alpha/2,n-2)}$.
- O Valor-p associado é dado por

$$\text{Valor-p} := 2\min\{\mathbb{P}[t_{(n-2)} > t_0]; \mathbb{P}[t_{(n-2)} < t_0]\}.$$

Teste para o intercepto

- Para testar $\mathcal{H}_0 : \beta_0 = b_0$ vs $\mathcal{H}_1 : \beta_0 \neq b_0$, b_0 especificado, podemos usar a estatística de teste

$$t_0 := \frac{\widehat{\beta}_0 - b_0}{\sqrt{\mathrm{QMRes}\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)}}.$$

- Sob $\mathcal{H}_0 : \beta_0 = b_0$, temos que $t_0 \sim t_{(n-2)}$.
- Rejeita-se $\mathcal{H}_0$ ao nível α se $|t_0| > t_{(1-\alpha/2,n-2)}$.
- O Valor-p associado é dado por

$$\text{Valor-p} := 2\min\{\mathbb{P}[t_{(n-2)} > t_0]; \mathbb{P}[t_{(n-2)} < t_0]\}.$$

Teste para o intercepto

- Para testar $\mathcal{H}_0 : \beta_0 = b_0$ vs $\mathcal{H}_1 : \beta_0 \neq b_0$, b_0 especificado, podemos usar a estatística de teste

$$t_0 := \frac{\widehat{\beta}_0 - b_0}{\sqrt{\mathrm{QMRes}\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)}}.$$

- Sob $\mathcal{H}_0 : \beta_0 = b_0$, temos que $t_0 \sim t_{(n-2)}$.
- Rejeita-se $\mathcal{H}_0$ ao nível α se $|t_0| > t_{(1-\alpha/2, n-2)}$.
- O Valor-p associado é dado por

$$\text{Valor-p} := 2\min\{\mathbb{P}[t_{(n-2)} > t_0]; \mathbb{P}[t_{(n-2)} < t_0]\}.$$

Teste para o intercepto

- Para testar $\mathcal{H}_0 : \beta_0 = b_0$ vs $\mathcal{H}_1 : \beta_0 \neq b_0$, b_0 especificado, podemos usar a estatística de teste

$$t_0 := \frac{\widehat{\beta}_0 - b_0}{\sqrt{\mathrm{QMRes}\left(\dfrac{1}{n} + \dfrac{\bar{x}^2}{S_{xx}}\right)}}.$$

- Sob $\mathcal{H}_0 : \beta_0 = b_0$, temos que $t_0 \sim t_{(n-2)}$.
- Rejeita-se $\mathcal{H}_0$ ao nível α se $|t_0| > t_{(1-\alpha/2, n-2)}$.
- O Valor-p associado é dado por

$$\text{Valor-p} := 2\min\{\mathbb{P}[t_{(n-2)} > t_0]; \mathbb{P}[t_{(n-2)} < t_0]\}.$$

Teste para inclinação

- Para testar $\mathcal{H}_0 : \beta_1 = b_1$ vs $\mathcal{H}_1 : \beta_1 \neq b_1$, b_1 especificado, podemos usar a estatística de teste

$$t_1 := \frac{\widehat{\beta}_1 - b_1}{\sqrt{\frac{\mathrm{QMRes}}{S_{xx}}}}.$$

- Sob $\mathcal{H}_0 : \beta_1 = b_1$, temos que $t_1 \sim t_{(n-2)}$.
- Rejeita-se $\mathcal{H}_0$ ao nível α se $|t_1| > t_{(1-\alpha/2, n-2)}$.
- O Valor-p associado é dado por

$$\text{Valor-p} := 2\min\{\mathbb{P}[t_{(n-2)} > t_1]; \mathbb{P}[t_{(n-2)} < t_1]\}.$$

Teste para inclinação

- Para testar $\mathcal{H}_0 : \beta_1 = b_1$ vs $\mathcal{H}_1 : \beta_1 \neq b_1$, b_1 especificado, podemos usar a estatística de teste

$$t_1 := \frac{\widehat{\beta}_1 - b_1}{\sqrt{\frac{\mathrm{QMRes}}{S_{xx}}}}.$$

- Sob $\mathcal{H}_0 : \beta_1 = b_1$, temos que $t_1 \sim t_{(n-2)}$.
- Rejeita-se $\mathcal{H}_0$ ao nível α se $|t_1| > t_{(1-\alpha/2, n-2)}$.
- O Valor-p associado é dado por

$$\text{Valor-p} := 2\min\{\mathbb{P}[t_{(n-2)} > t_1]; \mathbb{P}[t_{(n-2)} < t_1]\}.$$

Teste para inclinação

- Para testar $\mathcal{H}_0 : \beta_1 = b_1$ vs $\mathcal{H}_1 : \beta_1 \neq b_1$, b_1 especificado, podemos usar a estatística de teste

$$t_1 := \frac{\widehat{\beta}_1 - b_1}{\sqrt{\dfrac{\mathrm{QMRes}}{S_{xx}}}}.$$

- Sob $\mathcal{H}_0 : \beta_1 = b_1$, temos que $t_1 \sim t_{(n-2)}$.
- Rejeita-se $\mathcal{H}_0$ ao nível α se $|t_1| > t_{(1-\alpha/2, n-2)}$.
- O Valor-p associado é dado por

$$\text{Valor-p} := 2\min\{\mathbb{P}[t_{(n-2)} > t_1]; \mathbb{P}[t_{(n-2)} < t_1]\}.$$

Teste para inclinação

- Para testar $\mathcal{H}_0 : \beta_1 = b_1$ vs $\mathcal{H}_1 : \beta_1 \neq b_1$, b_1 especificado, podemos usar a estatística de teste

$$t_1 := \frac{\widehat{\beta}_1 - b_1}{\sqrt{\dfrac{\mathrm{QMRes}}{S_{xx}}}}.$$

- Sob $\mathcal{H}_0 : \beta_1 = b_1$, temos que $t_1 \sim t_{(n-2)}$.
- Rejeita-se $\mathcal{H}_0$ ao nível α se $|t_1| > t_{(1-\alpha/2, n-2)}$.
- O Valor-p associado é dado por

$$\text{Valor-p} := 2\min\{\mathbb{P}[t_{(n-2)} > t_1]; \mathbb{P}[t_{(n-2)} < t_1]\}.$$

Teste para inclinação - observação importante

- Se fizermos $b_1 = 0$, teremos o teste para significância do modelo, de forma que a estatística de teste fica reduzida a

$$t_1 := \frac{\widehat{\beta}_1}{\sqrt{\frac{\text{QMRes}}{S_{xx}}}}.$$

- Este teste é equivalente ao teste baseado na estatística F_0 oriunda do quadro da ANOVA?
- Note que

$$t_1^2 := \frac{\widehat{\beta}_1^2 S_{xx}}{\text{QMRes}} = \frac{\text{SQReg}}{\text{QMRes}} = F_0.$$

- Além disso, sob $\mathcal{H}_0 : \beta_1 = 0$, $t_1 \sim t_{(n-2)}$, de forma que sob $\mathcal{H}_0 : \beta_1 = 0$

$$t_1^2 \sim t_{(n-2)}^2 \overset{\mathcal{D}}{=} \mathcal{F}(1, n-2),$$

ou seja, os testes são **completamente equivalentes**.

Teste para inclinação - observação importante

- Se fizermos $b_1 = 0$, teremos o teste para significância do modelo, de forma que a estatística de teste fica reduzida a

$$t_1 := \frac{\widehat{\beta}_1}{\sqrt{\frac{\mathrm{QMRes}}{S_{xx}}}}.$$

- Este teste é equivalente ao teste baseado na estatística F_0 oriunda do quadro da ANOVA?

- Note que

$$t_1^2 := \frac{\widehat{\beta}_1^2 S_{xx}}{\mathrm{QMRes}} = \frac{\mathrm{SQReg}}{\mathrm{QMRes}} = F_0.$$

- Além disso, sob $\mathcal{H}_0 : \beta_1 = 0$, $t_1 \sim t_{(n-2)}$, de forma que sob $\mathcal{H}_0 : \beta_1 = 0$

$$t_1^2 \sim t_{(n-2)}^2 \stackrel{\mathcal{D}}{=} \mathcal{F}(1, n-2),$$

ou seja, os testes são **completamente equivalentes**.

Teste para inclinação - observação importante

- Se fizermos $b_1 = 0$, teremos o teste para significância do modelo, de forma que a estatística de teste fica reduzida a

$$t_1 := \frac{\widehat{\beta}_1}{\sqrt{\frac{\text{QMRes}}{S_{xx}}}}.$$

- Este teste é equivalente ao teste baseado na estatística F_0 oriunda do quadro da ANOVA?

- Note que

$$t_1^2 := \frac{\widehat{\beta}_1^2 S_{xx}}{\text{QMRes}} = \frac{\text{SQReg}}{\text{QMRes}} = F_0.$$

- Além disso, sob $\mathcal{H}_0 : \beta_1 = 0$, $t_1 \sim t_{(n-2)}$, de forma que sob $\mathcal{H}_0 : \beta_1 = 0$

$$t_1^2 \sim t_{(n-2)}^2 \stackrel{\mathcal{D}}{=} \mathcal{F}(1, n-2),$$

ou seja, os testes são **completamente equivalentes**.

Teste para inclinação - observação importante

- Se fizermos $b_1 = 0$, teremos o teste para significância do modelo, de forma que a estatística de teste fica reduzida a

$$t_1 := \frac{\widehat{\beta}_1}{\sqrt{\frac{\mathrm{QMRes}}{S_{xx}}}}.$$

- Este teste é equivalente ao teste baseado na estatística F_0 oriunda do quadro da ANOVA?
- Note que

$$t_1^2 := \frac{\widehat{\beta}_1^2 S_{xx}}{\mathrm{QMRes}} = \frac{\mathrm{SQReg}}{\mathrm{QMRes}} = F_0.$$

- Além disso, sob $\mathcal{H}_0 : \beta_1 = 0$, $t_1 \sim t_{(n-2)}$, de forma que sob $\mathcal{H}_0 : \beta_1 = 0$

$$t_1^2 \sim t_{(n-2)}^2 \overset{\mathcal{D}}{=} \mathcal{F}(1, n-2),$$

ou seja, os testes são **completamente equivalentes**.

Observações

- Esses testes são conhecidos como raiz do teste de **Wald**, e são comumente apresentados nos softwares estatísticos mais comuns.
- Se a suposição de normalidade não for satisfeita, mas a amostra for de tamanho **grande**, pode-se continuar utilizando estes testes, mas não como testes exatos e sim assintóticos em que a distribuição de referência sob $\mathcal{H}_0$ será a normal padrão.
- O teste de **Wald** é um teste assintótico muito utilizado, especialmente na área de modelos de regressão, devido a sua simplicidade de obtenção, além de possuir perfomance satisfatória.
- O teste de Wald, em conjunto com o teste escore de Rao e o da razão de verossimilhanças generalizada são os testes assintóticos mais comuns, e são conhecidos como **santíssima trindade**. Em 2002, outro teste entrou nesta classe limitadíssima, o teste Gradiente.

Observações

- Esses testes são conhecidos como raiz do teste de **Wald**, e são comumente apresentados nos softwares estatísticos mais comuns.
- Se a suposição de normalidade não for satisfeita, mas a amostra for de tamanho **grande**, pode-se continuar utilizando estes testes, mas não como testes exatos e sim assintóticos em que a distribuição de referência sob $\mathcal{H}_0$ será a normal padrão.
- O teste de **Wald** é um teste assintótico muito utilizado, especialmente na área de modelos de regressão, devido a sua simplicidade de obtenção, além de possuir perfomance satisfatória.
- O teste de Wald, em conjunto com o teste escore de Rao e o da razão de verossimilhanças generalizada são os testes assintóticos mais comuns, e são conhecidos como **santíssima trindade**. Em 2002, outro teste entrou nesta classe limitadíssima, o teste Gradiente.

Observações

- Esses testes são conhecidos como raiz do teste de **Wald**, e são comumente apresentados nos softwares estatísticos mais comuns.
- Se a suposição de normalidade não for satisfeita, mas a amostra for de tamanho **grande**, pode-se continuar utilizando estes testes, mas não como testes exatos e sim assintóticos em que a distribuição de referência sob $\mathcal{H}_0$ será a normal padrão.
- O teste de **Wald** é um teste assintótico muito utilizado, especialmente na área de modelos de regressão, devido a sua simplicidade de obtenção, além de possuir perfomance satisfatória.
- O teste de Wald, em conjunto com o teste escore de Rao e o da razão de verossimilhanças generalizada são os testes assintóticos mais comuns, e são conhecidos como **santíssima trindade**. Em 2002, outro teste entrou nesta classe limitadíssima, o teste Gradiente.

Observações

- Esses testes são conhecidos como raiz do teste de **Wald**, e são comumente apresentados nos softwares estatísticos mais comuns.
- Se a suposição de normalidade não for satisfeita, mas a amostra for de tamanho **grande**, pode-se continuar utilizando estes testes, mas não como testes exatos e sim assintóticos em que a distribuição de referência sob $\mathcal{H}_0$ será a normal padrão.
- O teste de **Wald** é um teste assintótico muito utilizado, especialmente na área de modelos de regressão, devido a sua simplicidade de obtenção, além de possuir perfomance satisfatória.
- O teste de Wald, em conjunto com o teste escore de Rao e o da razão de verossimilhanças generalizada são os testes assintóticos mais comuns, e são conhecidos como **santíssima trindade**. Em 2002, outro teste entrou nesta classe limitadíssima, o teste Gradiente.

Observações

- Todos os 4 testes são assintoticamente equivalentes sob $\mathcal{H}_0$ e sob hipóteses locais de Pitman, todavia em amostra de tamanho pequeno ou moderado, dependendo do modelo, um deles se torna preferível, seja pelo viés menor, poder local maior ou mesmo simplicidade.
- Quem tiver interesse, pode verificar as monografias Mota (2017, Estatística Gradiente: Conceitos e Aplicações), Santos-Filho (2021, Utilização da Estatística Gradiente e seu refinamento via Bootstrap em Modelos Lineares Simétricos) ou façam a disciplina de Inferência II. 🙂

Observações

- Todos os 4 testes são assintoticamente equivalentes sob $\mathcal{H}_0$ e sob hipóteses locais de Pitman, todavia em amostra de tamanho pequeno ou moderado, dependendo do modelo, um deles se torna preferível, seja pelo viés menor, poder local maior ou mesmo simplicidade.
- Quem tiver interesse, pode verificar as monografias Mota (2017, Estatística Gradiente: Conceitos e Aplicações), Santos-Filho (2021, Utilização da Estatística Gradiente e seu refinamento via Bootstrap em Modelos Lineares Simétricos) ou façam a disciplina de Inferência II. 😕

Exercício - Entregar próxima aula

Exercício: Baseado nas estatísticas de testes apresentadas anteriormente, especifique a região crítica e como obter o valor-p nos seguintes casos:

i) $\mathcal{H}_0 : \beta_0 = b_0$ vs $\mathcal{H}_1 : \beta_0 > b_0$, b_0 especificado.

ii) $\mathcal{H}_0 : \beta_0 = b_0$ vs $\mathcal{H}_1 : \beta_0 < b_0$, b_0 especificado.

iii) $\mathcal{H}_0 : \beta_1 = b_1$ vs $\mathcal{H}_1 : \beta_1 > b_1$, b_1 especificado.

iv) $\mathcal{H}_0 : \beta_1 = b_1$ vs $\mathcal{H}_1 : \beta_1 < b_1$, b_1 especificado.

Valores preditos

Temos que $\forall i = 1, \ldots, n$

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

que representa o valor estimado/predito do valor esperado da variável resposta quando $X = x_i$ segundo o MRLS.

A variância deste respectivo valor predito é dada por

$$\begin{aligned}
\mathrm{Var}[\widehat{y}_i] &= \mathrm{Var}[\widehat{\beta}_0 + \widehat{\beta}_1 x_i] \\
&= \mathrm{Var}[\widehat{\beta}_0] + x_i^2 \mathrm{Var}[\widehat{\beta}_1] + 2x_i \mathrm{Cov}[\widehat{\beta}_0, \widehat{\beta}_1] \\
&= \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right) + \sigma^2 \frac{x_i^2}{S_{xx}} + 2x_i \left(\frac{-\sigma^2 \bar{x}}{S_{xx}} \right) \\
&= \sigma^2 \left\{ \frac{1}{n} + \frac{\bar{x}^2 + x_i^2 - 2x_i \bar{x}}{S_{xx}} \right\} \\
&= \sigma^2 \left\{ \frac{1}{n} + \frac{(x_i - \bar{x})^2}{S_{xx}} \right\} \geq \frac{\sigma^2}{n} = \mathrm{Var}[\overline{y}],
\end{aligned}$$

lembrando que quando $x_i = \bar{x}$ tem-se $\widehat{y}_i = \bar{y}$. Isso faz sentido?