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/$\sim$juvencio

DEMA-UFC

Capital do **Ceará**, setembro de 2022

Conteúdo

MRLS

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

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

em que:

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

MRLS

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

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

em que:

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

MRLS

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

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

em que:

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

MRLS

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

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

em que:

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

MRLS

- Ao estabelecer o MRLS, pressupomos que:
  i) A função de regressão é linear (nos parâmetros). É comum, apesar de formalmente incorreta, nos textos aparecer a relação entre y_i e x_i é linear nos parâmetros.
  ii) Os valores de x_i são fixos, i.e., x_i não é uma variável aleatória.
  iii) $\mathbb{E}[e_i] = 0$, $\forall i = 1, \ldots, n$. Na verdade, tal suposição deveria ser escrita como (o que acaba implicando a anterior) $\mathbb{E}[e_i|x_i] = 0, \forall i = 1, \ldots, n$.
  iv) Para um dado valor de x_i, a variância da fonte de variação é constante, i.e.,
  $$\mathrm{Var}[e_i] = \mathbb{E}[e_i^2] = \sigma^2, \forall i = 1, \ldots, n \ \ (\textbf{Homoscesdaticidade}).$$
  Na verdade, tal suposição deveria ser escrita como
  $$\mathrm{Var}[y_i|x_i] = \mathrm{Var}[e_i|x_i] = \sigma^2, \forall i = 1, \ldots, n.$$
  v) A fonte de variação associada a uma observação é não-correlacionada com a fonte de variação associada de outra observação, i.e.,
  $$\mathrm{Cov}(e_i, e_j) = \mathbb{E}[e_i e_j] = 0, \forall i \neq j.$$

MRLS

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

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

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

MRLS

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

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

- Perceba que no MRLS (1) assume-se essencialmente que a fonte de variação está relacionada somente a variável resposta, i.e, a variável explicativa é medida sem erro, ou seja, com completa exatidão. Isso é razoável no contexto prático? 😵

- Se tivermos uma fonte de variação também associada a variável explicativa x_i, teremos essencialmente um *modelo com erro de medida/erro nas variáveis*. 😎

MRLS

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

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

- Perceba que no MRLS (1) assume-se essencialmente que a fonte de variação está relacionada somente a variável resposta, i.e, a variável explicativa é medida sem erro, ou seja, com completa exatidão. Isso é razoável no contexto prático? 😵

- Se tivermos uma fonte de variação também associada a variável explicativa x_i, teremos essencialmente um *modelo com erro de medida/erro nas variáveis*. 😎

MRLS

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

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

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

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

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

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

MRLS

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

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

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

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

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

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

MRLS

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

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

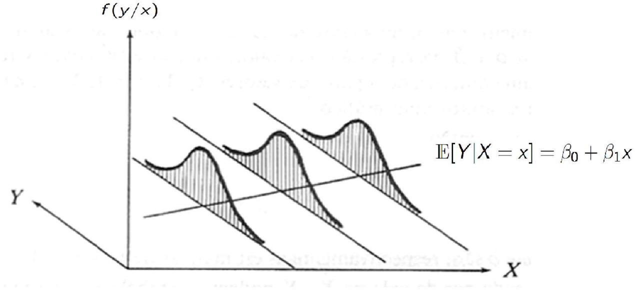
- Lembrando, que **correlação nula** implica independência sob a suposição de normalidade multivariada, então usando as suposições iv) e v) adicionada com a suposição acima, temos

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

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

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

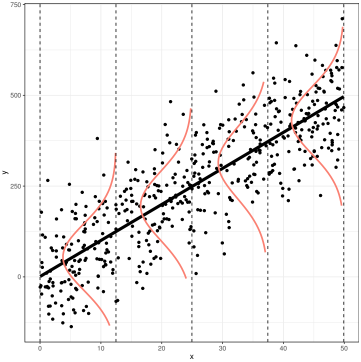
MRLS - Ilustração gráfica



Fonte: Hoffman (2006, Análise de regressão)

MRLS - Ilustração gráfica

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



MRLS - Interpretação dos parâmetros

- Sob as suposições usuais do MRLS, tem-se

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

Logo:

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

MRLS - Interpretação dos parâmetros

- Sob as suposições usuais do MRLS, tem-se

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

Logo:

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

MRLS - Interpretação dos parâmetros

- Sob as suposições usuais do MRLS, tem-se

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

Logo:

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

MRLS - Interpretação dos parâmetros

- Sob as suposições usuais do MRLS, tem-se

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

Logo:

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

Exemplos - Interpretação dos parâmetros

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

i) Renda vs. anos estudados (efetivos).

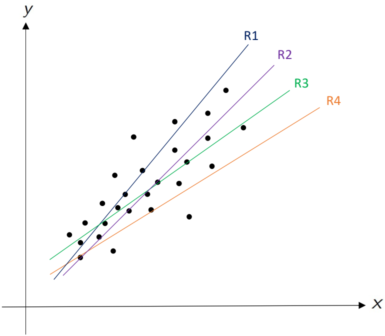
ii) Peso vs altura.

iii) Tempo de processamento vs # de faturas.

iv) Faturamento da empresa vs investimento com propaganda.

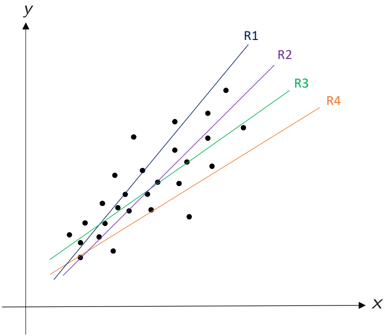
v) Pressão arterial sistólica (mmHg) vs idade (anos).

Método de Mínimos Quadrados (MQ)



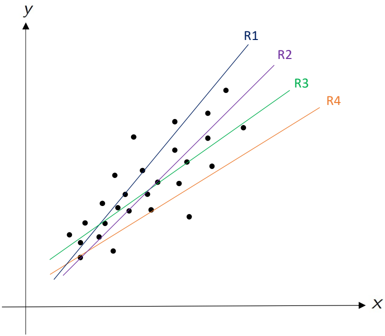
- Dado um conjunto de dados, existem **infinitas** retas candidatas para ajuste.
- Qual delas escolher?

Método de Mínimos Quadrados (MQ)



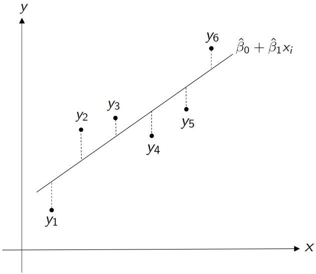
- Dado um conjunto de dados, existem **infinitas** retas candidatas para ajuste.
- Qual delas escolher?

Método de Mínimos Quadrados (MQ)



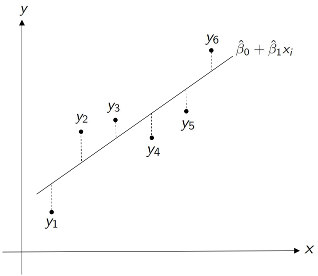
- Dado um conjunto de dados, existem **infinitas** retas candidatas para ajuste.
- Qual delas escolher?

Método de Mínimos Quadrados (MQ) - Ideia



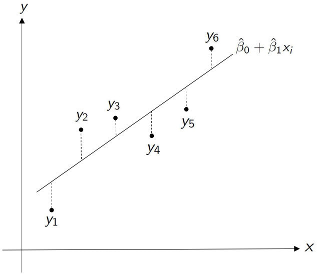
- A melhor reta estimada será aquela que minimiza a distância do valor observado y_i para o valor esperado ajustado $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$, $i = 1, \ldots, n$.
- Infelizmente, não é possível minimizar todas estas distâncias **simultaneamente**, logo, considera-se alguma função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

Método de Mínimos Quadrados (MQ) - Ideia



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

Método de Mínimos Quadrados (MQ) - Ideia



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

Método de Mínimos Quadrados (MQ) - Ideia

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

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

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

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

Método de Mínimos Quadrados (MQ) - Ideia

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

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

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

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

Método de Mínimos Quadrados (MQ) - Ideia

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

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

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

- 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

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

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

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

Método de Mínimos Quadrados (MQ) - Ideia

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

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

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

Método de Mínimos Quadrados (MQ) - Ideia

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

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

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

Método de Mínimos Quadrados (MQ) - Ideia

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

Método de Mínimos Quadrados (MQ) - Ideia

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

Método de Mínimos Quadrados (MQ) - Ideia

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

Método de Mínimos Quadrados (MQ) - Ideia

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

Método de Mínimos Quadrados (MQ) - Ideia

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

Método de Mínimos Quadrados (MQ)

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

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

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

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

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)

- 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

$$\begin{aligned}\frac{\partial}{\partial\beta_0}Q_2(\beta_0,\beta_1) &= -2\sum_{i=1}^{n}(y_i-\beta_0-\beta_1x_i)\\ \frac{\partial}{\partial\beta_1}Q_2(\beta_0,\beta_1) &= -2\sum_{i=1}^{n}x_i(y_i-\beta_0-\beta_1x_i).\end{aligned}$$

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

$$\begin{cases} -2\sum_{i=1}^{n}(y_i-\widehat{\beta}_0-\widehat{\beta}_1x_i)=0\\ -2\sum_{i=1}^{n}x_i(y_i-\widehat{\beta}_0-\widehat{\beta}_1x_i)=0.\end{cases} \tag{4}$$

Método de Mínimos Quadrados (MQ)

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

$$\begin{aligned}\frac{\partial}{\partial\beta_0}Q_2(\beta_0,\beta_1) &= -2\sum_{i=1}^{n}(y_i-\beta_0-\beta_1x_i)\\ \frac{\partial}{\partial\beta_1}Q_2(\beta_0,\beta_1) &= -2\sum_{i=1}^{n}x_i(y_i-\beta_0-\beta_1x_i).\end{aligned}$$

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

$$\begin{cases}-2\sum_{i=1}^{n}(y_i-\widehat{\beta}_0-\widehat{\beta}_1x_i)=0\\ -2\sum_{i=1}^{n}x_i(y_i-\widehat{\beta}_0-\widehat{\beta}_1x_i)=0.\end{cases} \tag{4}$$

Sistema de Equações Normais

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

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

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

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

Sistema de Equações Normais

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

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

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

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

Sistema de Equações Normais

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

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

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

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

Sistema de Equações Normais

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

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

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

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

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

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

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 $i = 1, \dots, n$

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

- 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

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

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

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

- 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

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

- 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

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

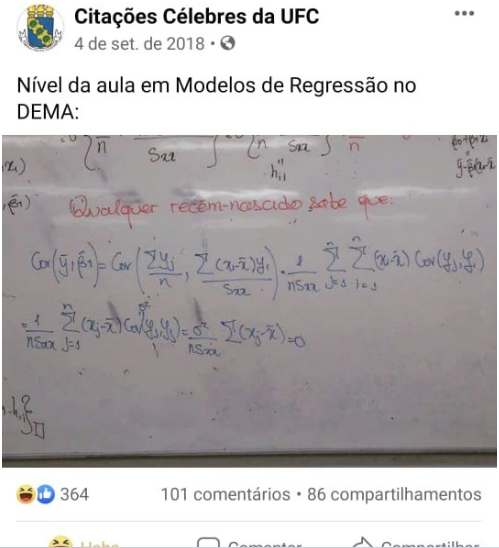
$$\begin{aligned}
\widehat{\mathrm{Var}}\left[\widehat{\beta}_0\right] &= \widehat{\sigma}^2\left(\frac{1}{n}+\frac{\bar{x}_n^2}{S_{xx}}\right) \text{ e } \widehat{\mathrm{EP}}(\widehat{\beta}_0) = \widehat{\sigma}\sqrt{\left(\frac{1}{n}+\frac{\bar{x}_n^2}{S_{xx}}\right)} \\
\widehat{\mathrm{Var}}\left[\widehat{\beta}_1\right]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

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

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

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

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

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 $\bar{y}_n$:

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

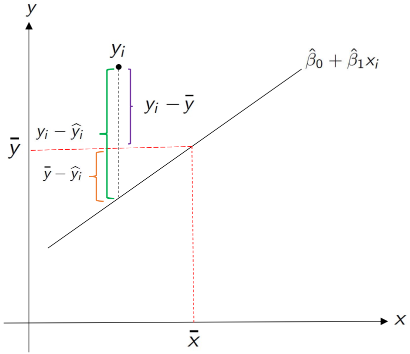
que pode ser reescrito como

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

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

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

implicando que

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

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

<u>**Ideia**</u>: 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

$$\frac{\mathrm{SQReg}}{\sigma^2} = \frac{\widehat{\beta}_1^2 S_{xx}}{\sigma^2} \sim \chi_1^2.$$

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, $\mathrm{SQReg} := \sum_{i=1}^{n}(\widehat{y}_i - \bar{y}_n)^2 = \dfrac{\widehat{\beta}_1^2 S_{xx}}{\sigma^2} \sim \chi^2_{(1, \beta_1^2 S_{xx}/\sigma^2)}$, 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

- 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,

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

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

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

- 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

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

- 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 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

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

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

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

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

- 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 $\mathcal{H}_0 : \beta_1 = 0$

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

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

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

- 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 $\mathcal{H}_0 : \beta_1 = 0$

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

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?

Valores preditos

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

$$\widehat{y_i} = \widehat{\beta}_0 + \widehat{\beta}_1 x_i = (1, x_i)\begin{pmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \end{pmatrix} \sim \mathcal{N}\left[\beta_0 + \beta_1 x_i; \sigma^2 \left\{\frac{1}{n} + \frac{(x_i - \bar{x})^2}{S_{xx}}\right\}\right].$$

- Perceba que a variância de $\widehat{y_i}$ aumenta a medida que x_i se **distancia** do centro das observações $\bar{x}$.

- Tem-se que

$$\mathrm{Var}[\widehat{y_i}] = \sigma^2 \left\{\frac{1}{n} + \frac{(x_i - \bar{x})^2}{S_{xx}}\right\} = \sigma^2 h_{ii},$$

em que h_{ii} é denotada de **alavancagem** (*leverage*) da i-ésima observação. Estudaremos isso em detalhes posteriormente, em especial quando formos estudar a respeito de métodos de diagnóstico do modelo.😎

Valores preditos

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

$$\widehat{y}_i = \widehat{\beta}_0 + \widehat{\beta}_1 x_i = (1, x_i) \begin{pmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \end{pmatrix} \sim \mathcal{N}\left[\beta_0 + \beta_1 x_i ; \sigma^2 \left\{ \frac{1}{n} + \frac{(x_i - \bar{x})^2}{S_{xx}} \right\}\right].$$

- Perceba que a variância de $\widehat{y}_i$ aumenta a medida que x_i se **distancia** do centro das observações $\bar{x}$.

- Tem-se que

$$\mathrm{Var}[\widehat{y}_i] = \sigma^2 \left\{ \frac{1}{n} + \frac{(x_i - \bar{x})^2}{S_{xx}} \right\} = \sigma^2 h_{ii},$$

em que h_{ii} é denotada de **alavancagem** (*leverage*) da i-ésima observação. Estudaremos isso em detalhes posteriormente, em especial quando formos estudar a respeito de métodos de diagnóstico do modelo.😎

Valores preditos

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

$$\widehat{y}_i = \widehat{\beta}_0 + \widehat{\beta}_1 x_i = (1, x_i)\begin{pmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \end{pmatrix} \sim \mathcal{N}\left[\beta_0 + \beta_1 x_i; \sigma^2\left\{\frac{1}{n} + \frac{(x_i - \bar{x})^2}{S_{xx}}\right\}\right].$$

- Perceba que a variância de $\widehat{y}_i$ aumenta a medida que x_i se **distancia** do centro das observações $\bar{x}$.

- Tem-se que

$$\mathrm{Var}[\widehat{y}_i] = \sigma^2\left\{\frac{1}{n} + \frac{(x_i - \bar{x})^2}{S_{xx}}\right\} = \sigma^2 h_{ii},$$

em que h_{ii} é denotada de **alavancagem** (*leverage*) da i-ésima observação. Estudaremos isso em detalhes posteriormente, em especial quando formos estudar a respeito de métodos de diagnóstico do modelo.

Exercício - Entregar próxima aula

Exercício: Sob o MRLS e suas pressuposições, mostre que

$$(\mathrm{Cor}[y_i, \widehat{y}_i])^2 = h_{ii}.$$

Com base no resultado acima, interprete o valor h_{ii}.

Valores preditos

- Dado que $\widehat{\boldsymbol{\beta}}$ e SQRes são independentes, então $\widehat{y}_i = g(\widehat{\boldsymbol{\beta}})$ e SQRes também são independentes.

- Além disso, sob a suposição de normalidade, tem-se que

$$\frac{\dfrac{\widehat{y}_i - (\beta_0 + \beta_1 x_i)}{\sqrt{\sigma^2 h_{ii}}}}{\sqrt{\dfrac{\mathrm{SQRes}}{\sigma^2(n-2)}}} = \frac{\widehat{y}_i - (\beta_0 + \beta_1 x_i)}{\sqrt{h_{ii}\mathrm{QMRes}}} \sim t_{(n-2)}, \tag{13}$$

de forma que $\widehat{\mathrm{Var}}(\widehat{y}_i) = h_{ii}\mathrm{QMRes}$.

- De agora em diante, denotar-se-á $\mathbb{E}[y_i|X_i = x_i] = \beta_0 + \beta_1 x_i = \mu_{x_i}$ por questão de simplicidade.

Valores preditos

- Dado que $\widehat{\boldsymbol{\beta}}$ e SQRes são independentes, então $\widehat{y}_i = g(\widehat{\boldsymbol{\beta}})$ e SQRes também são independentes.
- Além disso, sob a suposição de normalidade, tem-se que

$$\frac{\dfrac{\widehat{y}_i - (\beta_0 + \beta_1 x_i)}{\sqrt{\sigma^2 h_{ii}}}}{\sqrt{\dfrac{\mathrm{SQRes}}{\sigma^2(n-2)}}} = \frac{\widehat{y}_i - (\beta_0 + \beta_1 x_i)}{\sqrt{h_{ii}\mathrm{QMRes}}} \sim t_{(n-2)}, \tag{13}$$

de forma que $\widehat{\mathrm{Var}}(\widehat{y}_i) = h_{ii}\mathrm{QMRes}$.

- De agora em diante, denotar-se-á $\mathbb{E}[y_i|X_i = x_i] = \beta_0 + \beta_1 x_i = \mu_{x_i}$ por questão de simplicidade.

Valores preditos

- Dado que $\widehat{\boldsymbol{\beta}}$ e SQRes são independentes, então $\widehat{y}_i = g(\widehat{\boldsymbol{\beta}})$ e SQRes também são independentes.
- Além disso, sob a suposição de normalidade, tem-se que

$$\frac{\dfrac{\widehat{y}_i - (\beta_0 + \beta_1 x_i)}{\sqrt{\sigma^2 h_{ii}}}}{\sqrt{\dfrac{\mathrm{SQRes}}{\sigma^2(n-2)}}} = \frac{\widehat{y}_i - (\beta_0 + \beta_1 x_i)}{\sqrt{h_{ii}\mathrm{QMRes}}} \sim t_{(n-2)}, \tag{13}$$

de forma que $\widehat{\mathrm{Var}}(\widehat{y}_i) = h_{ii}\mathrm{QMRes}$.

- De agora em diante, denotar-se-á $\mathbb{E}[y_i|X_i = x_i] = \beta_0 + \beta_1 x_i = \mu_{x_i}$ por questão de simplicidade.

Exercício - Entregar próxima aula

Exercício: Com base na quantidade pivotal (13) apresente um teste de hipóteses, especificando a região crítica e como obter o valor-p, nos seguintes casos:

i) $\mathcal{H}_0 : \mu_{x_i} = \mu_0$ vs $\mathcal{H}_1 : \mu_{x_i} \neq \mu_0$, μ_0 especificado.

ii) $\mathcal{H}_0 : \mu_{x_i} = \mu_0$ vs $\mathcal{H}_1 : \mu_{x_i} > \mu_0$, μ_0 especificado.

iii) $\mathcal{H}_0 : \mu_{x_i} = \mu_0$ vs $\mathcal{H}_1 : \mu_{x_i} < \mu_0$, μ_0 especificado.

IC valor médio

- Considerando (13) podemos utilizar o método da quantidade pivotal para determinar um IC com nível de confiança $(1-\alpha)$ para μ_{x_i}.
- Além disso, como a distribuição de referência é simétrica em torno da origem,então o IC simétrico corresponde ao **melhor** IC com nível de confiança $(1-\alpha)$ para μ_{x_i} e o mesmo é dado por

$$\mathrm{IC}_{1-\alpha}(\mu_{x_i}) = \left[\widehat{y}_i \pm t_{(1-\alpha/2,n-2)}\sqrt{h_{ii}\mathrm{QMRes}}\right],$$

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

IC valor médio

- Considerando (13) podemos utilizar o método da quantidade pivotal para determinar um IC com nível de confiança $(1-\alpha)$ para μ_{x_i}.
- Além disso, como a distribuição de referência é simétrica em torno da origem,então o IC simétrico corresponde ao **melhor** IC com nível de confiança $(1-\alpha)$ para μ_{x_i} e o mesmo é dado por

$$\mathrm{IC}_{1-\alpha}(\mu_{x_i}) = \left[\widehat{y}_i \pm t_{(1-\alpha/2,n-2)}\sqrt{h_{ii}\mathrm{QMRes}}\right],$$

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

Prevendo novas observações

- Frequentemente, temos interesse em predizer o valor de uma nova observação y_h, relativamente ao valor x_h da variável explicativa, i.e., queremos prever o valor da variável resposta em uma nova observação com $x = x_h$.
- O preditor de

$$y_h = \beta_0 + \beta_1 x_h + e_h$$

é

$$\widehat{y}_h = \widehat{\beta}_0 + \widehat{\beta}_1 x_h$$

associado ao **erro de previsão**

$$\widehat{y}_h - y_h = (\widehat{\beta}_0 - \beta_0) + (\widehat{\beta}_1 - \beta_1) x_h - e_h.$$

- Dizemos que $\widehat{y}_h$ é um **preditor** não viesado de y_h pois o valor esperado do erro de previsão é igual a zero, i.e.,

$$\mathbb{E}\left[\widehat{y}_h - y_h\right] = 0,$$

implicando que $\mathbb{E}[\widehat{y}_h] = \mathbb{E}[y_h]$. Note, entretanto que $\mathbb{E}[\widehat{y}_h] = \beta_0 + \beta_1 x_h \neq y_h$ que é uma variável aleatória.

Prevendo novas observações

- Frequentemente, temos interesse em predizer o valor de uma nova observação y_h, relativamente ao valor x_h da variável explicativa, i.e., queremos prever o valor da variável resposta em uma nova observação com $x = x_h$.
- O preditor de
$$y_h = \beta_0 + \beta_1 x_h + e_h$$
é
$$\widehat{y}_h = \widehat{\beta}_0 + \widehat{\beta}_1 x_h$$
associado ao **erro de previsão**
$$\widehat{y}_h - y_h = (\widehat{\beta}_0 - \beta_0) + (\widehat{\beta}_1 - \beta_1)x_h - e_h.$$
- Dizemos que $\widehat{y}_h$ é um **preditor** não viesado de y_h pois o valor esperado do erro de previsão é igual a zero, i.e.,
$$\mathbb{E}\left[\widehat{y}_h - y_h\right] = 0,$$
implicando que $\mathbb{E}[\widehat{y}_h] = \mathbb{E}[y_h]$. Note, entretanto que $\mathbb{E}[\widehat{y}_h] = \beta_0 + \beta_1 x_h \neq y_h$ que é uma variável aleatória.

Prevendo novas observações

- Frequentemente, temos interesse em predizer o valor de uma nova observação y_h, relativamente ao valor x_h da variável explicativa, i.e., queremos prever o valor da variável resposta em uma nova observação com $x = x_h$.
- O preditor de
$$y_h = \beta_0 + \beta_1 x_h + e_h$$
é
$$\widehat{y}_h = \widehat{\beta}_0 + \widehat{\beta}_1 x_h$$
associado ao **erro de previsão**
$$\widehat{y}_h - y_h = (\widehat{\beta}_0 - \beta_0) + (\widehat{\beta}_1 - \beta_1)x_h - e_h.$$
- Dizemos que $\widehat{y}_h$ é um **preditor** não viesado de y_h pois o valor esperado do erro de previsão é igual a zero, i.e.,
$$\mathbb{E}\left[\widehat{y}_h - y_h\right] = 0,$$
implicando que $\mathbb{E}[\widehat{y}_h] = \mathbb{E}[y_h]$. Note, entretanto que $\mathbb{E}[\widehat{y}_h] = \beta_0 + \beta_1 x_h \neq y_h$ que é uma variável aleatória.

Prevendo novas observações

- Considere que a fonte de variação associada a **nova observação** e_h seja independente de $e_1, e_2, \ldots, e_n$ e que $\mathrm{Var}(e_h) = \sigma^2$.
- Isso implica que $\widehat{y}_h = \widehat{\beta}_0 + \widehat{\beta}_1 x_h = g(y_1, \ldots, y_n) = g(\boldsymbol{e})$ é independente de e_h, desde que $\widehat{y}_h$ é uma função linear de $y_1, \ldots, y_n$ ou equivalentemente de $e_1, \ldots, e_n$.
- Desta forma, temos que a variância do erro de previsão é dada por

$$\begin{aligned}
\mathrm{Var}\left(\widehat{y}_h - y_h\right) &= \mathrm{Var}(\widehat{y}_h) + \mathrm{Var}(y_h) = \mathrm{Var}(\widehat{y}_h) + \mathrm{Var}\left(\beta_0 + \beta_1 x_h + e_h\right) \\
&= \sigma^2 \left\{ \frac{1}{n} + \frac{(x_h - \bar{x}_n)^2}{S_{xx}} \right\} + \mathrm{Var}(e_h) \\
&= \sigma^2 \left\{ 1 + \frac{1}{n} + \frac{(x_h - \bar{x}_n)^2}{S_{xx}} \right\}.
\end{aligned}$$

- Denotando o erro de previsão por $\widehat{e}_h := \widehat{y}_h - y_h$, então sob as suposições usuais do MRLS (sem normalidade) temos

$$\widehat{e}_h \sim \left(0, \sigma^2 \left\{ 1 + \frac{1}{n} + \frac{(x_h - \bar{x}_n)^2}{S_{xx}} \right\}\right).$$

Prevendo novas observações

- Considere que a fonte de variação associada a **nova observação** e_h seja independente de $e_1, e_2, \ldots, e_n$ e que $\mathrm{Var}(e_h) = \sigma^2$.
- Isso implica que $\widehat{y}_h = \widehat{\beta}_0 + \widehat{\beta}_1 x_h = g(y_1, \ldots, y_n) = g(\boldsymbol{e})$ é independente de e_h, desde que $\widehat{y}_h$ é uma função linear de $y_1, \ldots, y_n$ ou equivalentemente de $e_1, \ldots, e_n$.
- Desta forma, temos que a variância do erro de previsão é dada por

$$\begin{aligned}\mathrm{Var}\left(\widehat{y}_h - y_h\right) &= \mathrm{Var}(\widehat{y}_h) + \mathrm{Var}(y_h) = \mathrm{Var}(\widehat{y}_h) + \mathrm{Var}\left(\beta_0 + \beta_1 x_h + e_h\right) \\ &= \sigma^2\left\{\frac{1}{n} + \frac{(x_h - \bar{x}_n)^2}{S_{xx}}\right\} + \mathrm{Var}(e_h) \\ &= \sigma^2\left\{1 + \frac{1}{n} + \frac{(x_h - \bar{x}_n)^2}{S_{xx}}\right\}.\end{aligned}$$

- Denotando o erro de previsão por $\widehat{e}_h := \widehat{y}_h - y_h$, então sob as suposições usuais do MRLS (sem normalidade) temos

$$\widehat{e}_h \sim \left(0, \sigma^2\left\{1 + \frac{1}{n} + \frac{(x_h - \bar{x}_n)^2}{S_{xx}}\right\}\right).$$

Prevendo novas observações

- Considere que a fonte de variação associada a **nova observação** e_h seja independente de $e_1, e_2, \ldots, e_n$ e que $\mathrm{Var}(e_h) = \sigma^2$.
- Isso implica que $\widehat{y}_h = \widehat{\beta}_0 + \widehat{\beta}_1 x_h = g(y_1, \ldots, y_n) = g(\boldsymbol{e})$ é independente de e_h, desde que $\widehat{y}_h$ é uma função linear de $y_1, \ldots, y_n$ ou equivalentemente de $e_1, \ldots, e_n$.
- Desta forma, temos que a variância do erro de previsão é dada por

$$\begin{aligned}
\mathrm{Var}\left(\widehat{y}_h - y_h\right) &= \mathrm{Var}(\widehat{y}_h) + \mathrm{Var}(y_h) = \mathrm{Var}(\widehat{y}_h) + \mathrm{Var}\left(\beta_0 + \beta_1 x_h + e_h\right) \\
&= \sigma^2 \left\{ \frac{1}{n} + \frac{(x_h - \bar{x}_n)^2}{S_{xx}} \right\} + \mathrm{Var}(e_h) \\
&= \sigma^2 \left\{ 1 + \frac{1}{n} + \frac{(x_h - \bar{x}_n)^2}{S_{xx}} \right\}.
\end{aligned}$$

- Denotando o erro de previsão por $\widehat{e}_h := \widehat{y}_h - y_h$, então sob as suposições usuais do MRLS (sem normalidade) temos

$$\widehat{e}_h \sim \left(0, \sigma^2 \left\{ 1 + \frac{1}{n} + \frac{(x_h - \bar{x}_n)^2}{S_{xx}} \right\}\right).$$

Prevendo novas observações

- Considere que a fonte de variação associada a **nova observação** e_h seja independente de $e_1, e_2, \ldots, e_n$ e que $\mathrm{Var}(e_h) = \sigma^2$.
- Isso implica que $\widehat{y}_h = \widehat{\beta}_0 + \widehat{\beta}_1 x_h = g(y_1, \ldots, y_n) = g(\boldsymbol{e})$ é independente de e_h, desde que $\widehat{y}_h$ é uma função linear de $y_1, \ldots, y_n$ ou equivalentemente de $e_1, \ldots, e_n$.
- Desta forma, temos que a variância do erro de previsão é dada por

$$\begin{aligned}
\mathrm{Var}\left(\widehat{y}_h - y_h\right) &= \mathrm{Var}(\widehat{y}_h) + \mathrm{Var}(y_h) = \mathrm{Var}(\widehat{y}_h) + \mathrm{Var}\left(\beta_0 + \beta_1 x_h + e_h\right) \\
&= \sigma^2 \left\{ \frac{1}{n} + \frac{(x_h - \bar{x}_n)^2}{S_{xx}} \right\} + \mathrm{Var}(e_h) \\
&= \sigma^2 \left\{ 1 + \frac{1}{n} + \frac{(x_h - \bar{x}_n)^2}{S_{xx}} \right\}.
\end{aligned}$$

- Denotando o erro de previsão por $\widehat{e}_h := \widehat{y}_h - y_h$, então sob as suposições usuais do MRLS (sem normalidade) temos

$$\widehat{e}_h \sim \left(0, \sigma^2 \left\{ 1 + \frac{1}{n} + \frac{(x_h - \bar{x}_n)^2}{S_{xx}} \right\}\right).$$

Prevendo novas observações

- Adicionalmente, sob a suposição clássica de normalidade temos que
$$\widehat{e}_h \sim \mathcal{N}\left(0, \sigma^2\left\{1 + \frac{1}{n} + \frac{(x_h - \bar{x}_n)^2}{S_{xx}}\right\}\right).$$
- Usando o fato de que $\widehat{\beta} \perp \mathrm{SQRes}$ e considerando que $e_h \perp (e_1, \ldots, e_n)^\top$, temos que $\widehat{e}_h = \widehat{y}_h - y_h = g(\widehat{\beta}, e_h) \perp \mathrm{SQRes}$, de tal forma que obtemos a seguinte quantidade pivotal
$$\frac{\dfrac{\widehat{y}_h - y_h}{\sqrt{\sigma^2 \mathrm{Var}(\widehat{e}_h)}}}{\sqrt{\dfrac{\mathrm{SQRes}}{\sigma^2(n-2)}}} = \frac{\widehat{y}_h - y_h}{\sqrt{\mathrm{QMRes}\left\{1 + \frac{1}{n} + \frac{(x_h - \bar{x}_n)^2}{S_{xx}}\right\}}} \sim t_{(n-2)}. \quad (14)$$

Prevendo novas observações

- Adicionalmente, sob a suposição clássica de normalidade temos que

$$\widehat{e}_h \sim \mathcal{N}\left(0, \sigma^2\left\{1 + \frac{1}{n} + \frac{(x_h - \bar{x}_n)^2}{S_{xx}}\right\}\right).$$

- Usando o fato de que $\widehat{\boldsymbol{\beta}} \perp \mathrm{SQRes}$ e considerando que $e_h \perp (e_1, \ldots, e_n)^\top$, temos que $\widehat{e}_h = \widehat{y}_h - y_h = g(\widehat{\boldsymbol{\beta}}, e_h) \perp \mathrm{SQRes}$, de tal forma que obtemos a seguinte quantidade pivotal

$$\frac{\dfrac{\widehat{y}_h - y_h}{\sqrt{\sigma^2 \mathrm{Var}(\widehat{e}_h)}}}{\sqrt{\dfrac{\mathrm{SQRes}}{\sigma^2(n-2)}}} = \frac{\widehat{y}_h - y_h}{\sqrt{\mathrm{QMRes}\left\{1 + \dfrac{1}{n} + \dfrac{(x_h - \bar{x}_n)^2}{S_{xx}}\right\}}} \sim t_{(n-2)}. \tag{14}$$

Intervalo de predição

- Usando o método da quantidade pivotal, encontramos o seguinte intervalo de previsão/predição ótimo para uma nova observação $y_h = \beta_0 + \beta_1 x_h + e_h$ ao nível $1-\alpha$:

$$\mathrm{IP}_{1-\alpha}(y_h) = \left[\widehat{y}_h \pm t_{(1-\alpha/2,n-2)}\sqrt{\mathrm{QMRes}\left\{1+\frac{1}{n}+\frac{(x_h-\bar{x}_n)^2}{S_{xx}}\right\}}\right].$$

- O conceito de intervalo de predição é análogo ao de intervalo de confiança, com a diferença de que, enquanto o IC se refere a um parâmetro (constante, p.e., β_0), o intervalo de previsão se refere a uma variável aleatória, y_h no caso.
- Por esta razão, o comprimento do intervalo de predição para y_h é **sempre maior** que o do respectivo IC para $\mathbb{E}[y_h|X=x_h]$, para um mesmo nível de confiança $1-\alpha$, mais especificamente o erro-padrão do erro de predição é uma unidade de QMRes **maior**.

Intervalo de predição

- Usando o método da quantidade pivotal, encontramos o seguinte intervalo de previsão/predição ótimo para uma nova observação $y_h = \beta_0 + \beta_1 x_h + e_h$ ao nível $1-\alpha$:

$$\mathrm{IP}_{1-\alpha}(y_h) = \left[\widehat{y}_h \pm t_{(1-\alpha/2, n-2)} \sqrt{\mathrm{QMRes}\left\{1 + \frac{1}{n} + \frac{(x_h - \bar{x}_n)^2}{S_{xx}}\right\}}\right].$$

- O conceito de intervalo de predição é análogo ao de intervalo de confiança, com a diferença de que, enquanto o IC se refere a um parâmetro (constante, p.e., β_0), o intervalo de previsão se refere a uma variável aleatória, y_h no caso.

- Por esta razão, o comprimento do intervalo de predição para y_h é **sempre maior** que o do respectivo IC para $\mathbb{E}[y_h|X = x_h]$, para um mesmo nível de confiança $1-\alpha$, mais especificamente o erro-padrão do erro de predição é uma unidade de QMRes **maior**.

Intervalo de predição

- Usando o método da quantidade pivotal, encontramos o seguinte intervalo de previsão/predição ótimo para uma nova observação $y_h = \beta_0 + \beta_1 x_h + e_h$ ao nível $1-\alpha$:

$$\mathrm{IP}_{1-\alpha}(y_h) = \left[\widehat{y}_h \pm t_{(1-\alpha/2, n-2)} \sqrt{\mathrm{QMRes}\left\{1 + \frac{1}{n} + \frac{(x_h - \bar{x}_n)^2}{S_{xx}}\right\}}\right].$$

- O conceito de intervalo de predição é análogo ao de intervalo de confiança, com a diferença de que, enquanto o IC se refere a um parâmetro (constante, p.e., β_0), o intervalo de previsão se refere a uma variável aleatória, y_h no caso.

- Por esta razão, o comprimento do intervalo de predição para y_h é **sempre maior** que o do respectivo IC para $\mathbb{E}[y_h|X = x_h]$, para um mesmo nível de confiança $1-\alpha$, mais especificamente o erro-padrão do erro de predição é uma unidade de QMRes **maior**.

Comentários

- Analisando o intervalo de confiança para $\mathbb{E}[y_h|x = x_h] = \beta_0 + \beta_1 x_h = \mu_{x_h}$

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

de forma que podemos afirmar que a precisão da estimativa de μ_{x_h} é tanto maior quanto:

- Menor for o QMRes, i.e., quanto menor for a dispersão dos valores observados da variável resposta em torno da reta de regressão estimada.
- Maior for o tamanho da amostra (n).
- Maior for S_{xx}, i.e., maior for a dispersão dos valores $x_1, \ldots, x_n$ em torno de $\bar{x}_n$.
- Logo, podemos concluir que:
- O número de observações deve ser o maior possível e se possível, devemos escolher os valores de $x_1, \ldots, x_n$ de forma mais **dispersa** possível, obtendo um alto valor de S_{xx}.

Comentários

- Analisando o intervalo de confiança para $\mathbb{E}[y_h|x = x_h] = \beta_0 + \beta_1 x_h = \mu_{x_h}$

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

de forma que podemos afirmar que a precisão da estimativa de μ_{x_h} é tanto maior quanto:

- Menor for o QMRes, i.e., quanto menor for a dispersão dos valores observados da variável resposta em torno da reta de regressão estimada.
- Maior for o tamanho da amostra (n).
- Maior for S_{xx}, i.e., maior for a dispersão dos valores $x_1, \ldots, x_n$ em torno de $\bar{x}_n$.
- Logo, podemos concluir que:
- O número de observações deve ser o maior possível e se possível, devemos escolher os valores de $x_1, \ldots, x_n$ de forma mais **dispersa** possível, obtendo um alto valor de S_{xx}.

Comentários

- Analisando o intervalo de confiança para $\mathbb{E}[y_h|x = x_h] = \beta_0 + \beta_1 x_h = \mu_{x_h}$

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

de forma que podemos afirmar que a precisão da estimativa de μ_{x_h} é tanto maior quanto:

- Menor for o QMRes, i.e., quanto menor for a dispersão dos valores observados da variável resposta em torno da reta de regressão estimada.
- Maior for o tamanho da amostra (n).
- Maior for S_{xx}, i.e., maior for a dispersão dos valores $x_1, \ldots, x_n$ em torno de $\bar{x}_n$.
- Logo, podemos concluir que:
- O número de observações deve ser o maior possível e se possível, devemos escolher os valores de $x_1, \ldots, x_n$ de forma mais **dispersa** possível, obtendo um alto valor de S_{xx}.

Comentários

- Analisando o intervalo de confiança para $\mathbb{E}[y_h|x = x_h] = \beta_0 + \beta_1 x_h = \mu_{x_h}$

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

de forma que podemos afirmar que a precisão da estimativa de μ_{x_h} é tanto maior quanto:

- Menor for o QMRes, i.e., quanto menor for a dispersão dos valores observados da variável resposta em torno da reta de regressão estimada.
- Maior for o tamanho da amostra (n).
- Maior for S_{xx}, i.e., maior for a dispersão dos valores $x_1, \ldots, x_n$ em torno de $\bar{x}_n$.
- Logo, podemos concluir que:
- O número de observações deve ser o maior possível e se possível, devemos escolher os valores de $x_1, \ldots, x_n$ de forma mais **dispersa** possível, obtendo um alto valor de S_{xx}.

Comentários

- Analisando o intervalo de confiança para $\mathbb{E}[y_h|x = x_h] = \beta_0 + \beta_1 x_h = \mu_{x_h}$

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

de forma que podemos afirmar que a precisão da estimativa de μ_{x_h} é tanto maior quanto:

- Menor for o QMRes, i.e., quanto menor for a dispersão dos valores observados da variável resposta em torno da reta de regressão estimada.
- Maior for o tamanho da amostra (n).
- Maior for S_{xx}, i.e., maior for a dispersão dos valores $x_1, \ldots, x_n$ em torno de $\bar{x}_n$.
- Logo, podemos concluir que:
- O número de observações deve ser o maior possível e se possível, devemos escolher os valores de $x_1, \ldots, x_n$ de forma mais **dispersa** possível, obtendo um alto valor de S_{xx}.

Comentários

- Analisando o intervalo de confiança para $\mathbb{E}[y_h|x = x_h] = \beta_0 + \beta_1 x_h = \mu_{x_h}$

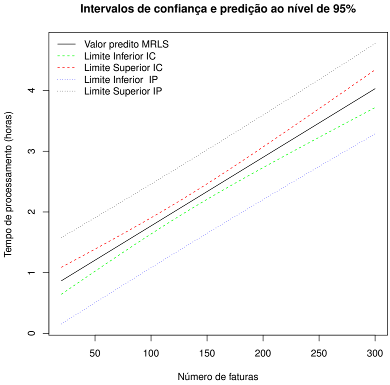
$$\mathrm{IC}_{1-\alpha}(\mu_{x_h}) = \left[\widehat{y}_h \pm t_{(1-\alpha/2, n-2)} \sqrt{\mathrm{QMRes}\left\{\frac{1}{n} + \frac{(x_h - \bar{x}_n)^2}{S_{xx}}\right\}}\right],$$

de forma que podemos afirmar que a precisão da estimativa de μ_{x_h} é tanto maior quanto:

- Menor for o QMRes, i.e., quanto menor for a dispersão dos valores observados da variável resposta em torno da reta de regressão estimada.
- Maior for o tamanho da amostra (n).
- Maior for S_{xx}, i.e., maior for a dispersão dos valores $x_1, \ldots, x_n$ em torno de $\bar{x}_n$.
- Logo, podemos concluir que:
- O número de observações deve ser o maior possível e se possível, devemos escolher os valores de $x_1, \ldots, x_n$ de forma mais **dispersa** possível, obtendo um alto valor de S_{xx}.

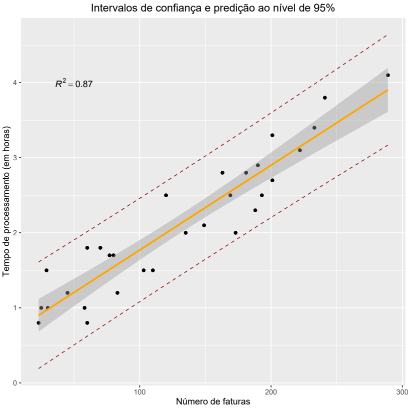
Exemplo IP x IC

Figura: Ilustração gráfica do IP e IC para um exemplo de dados reais.



Exemplo IP x IC

Figura: Ilustração gráfica do IP e IC para um exemplo de dados reais.



Comentários

- Perceba que o intervalo de predição (confiança) tende a ter um comprimento maior a medida que x_h se distancia de $\bar{x}_n$. Isso é esperado? Por quê?
- Ao fazer uma extrapolação é necessário considerar ainda um outro problema, provavelmente mais sério que o aumento da amplitude do IP/IC à medida que x_h se afasta de $\bar{x}_n$, que está associado ao **perigo da extrapolação** já discutido anteriormente.
- Frequentemente o modelo (linear) ajustado só oferece uma boa aproximação para a amplitude coberta pela amostra, mas pode ser absolutamente inapropriado para a região no qual se pretende ou se tem interesse em fazer a extrapolação.

Comentários

- Perceba que o intervalo de predição (confiança) tende a ter um comprimento maior a medida que x_h se distancia de $\bar{x}_n$. Isso é esperado? Por quê?
- Ao fazer uma extrapolação é necessário considerar ainda um outro problema, provavelmente mais sério que o aumento da amplitude do IP/IC à medida que x_h se afasta de $\bar{x}_n$, que está associado ao **perigo da extrapolação** já discutido anteriormente.
- Frequentemente o modelo (linear) ajustado só oferece uma boa aproximação para a amplitude coberta pela amostra, mas pode ser absolutamente inapropriado para a região no qual se pretende ou se tem interesse em fazer a extrapolação.

Comentários

- Perceba que o intervalo de predição (confiança) tende a ter um comprimento maior a medida que x_h se distancia de $\bar{x}_n$. Isso é esperado? Por quê?
- Ao fazer uma extrapolação é necessário considerar ainda um outro problema, provavelmente mais sério que o aumento da amplitude do IP/IC à medida que x_h se afasta de $\bar{x}_n$, que está associado ao **perigo da extrapolação** já discutido anteriormente.
- Frequentemente o modelo (linear) ajustado só oferece uma boa aproximação para a amplitude coberta pela amostra, mas pode ser absolutamente inapropriado para a região no qual se pretende ou se tem interesse em fazer a extrapolação.

Modelos de regressão passando pela origem

- Existem uma série de situações práticas, exemplo das escovas/número de faturas (comentar no quadro), no qual o interesse é ajustar um modelo de regressão passando pela origem (intercepto nulo) do tipo

$$y_i = \beta_1 x_i + e_i, \; i = 1, \ldots, n, \tag{15}$$

considerando válida todas as suposições usuais do MRLS.

- Lógico que para ajustar tal classe de modelos, deve-se levar em consideração a estrutura das variáveis, bem como se as observações estão próximas da origem. Caso isso não ocorra, pode-se ajustar o MRLS usual e posteriormente através do teste de hipóteses para $\mathcal{H}_0 : \beta_0 = 0$ reduzir para o modelo de intercepto nulo.

- Sob a forma funcional (15) e as suposições consideradas, tem-se que o EMQ de β_1 é obtido através da minimização de

$$S(\beta_1) := \sum_{i=1}^{n} (y_i - \beta_1 x_i)^2.$$

Modelos de regressão passando pela origem

- Existem uma série de situações práticas, exemplo das escovas/número de faturas (comentar no quadro), no qual o interesse é ajustar um modelo de regressão passando pela origem (intercepto nulo) do tipo

$$y_i = \beta_1 x_i + e_i, \; i = 1, \ldots, n, \tag{15}$$

considerando válida todas as suposições usuais do MRLS.

- Lógico que para ajustar tal classe de modelos, deve-se levar em consideração a estrutura das variáveis, bem como se as observações estão próximas da origem. Caso isso não ocorra, pode-se ajustar o MRLS usual e posteriormente através do teste de hipóteses para $\mathcal{H}_0 : \beta_0 = 0$ reduzir para o modelo de intercepto nulo.

- Sob a forma funcional (15) e as suposições consideradas, tem-se que o EMQ de β_1 é obtido através da minimização de

$$S(\beta_1) := \sum_{i=1}^{n} (y_i - \beta_1 x_i)^2.$$

Modelos de regressão passando pela origem

- Existem uma série de situações práticas, exemplo das escovas/número de faturas (comentar no quadro), no qual o interesse é ajustar um modelo de regressão passando pela origem (intercepto nulo) do tipo

$$y_i = \beta_1 x_i + e_i, \ i = 1, \ldots, n, \tag{15}$$

considerando válida todas as suposições usuais do MRLS.

- Lógico que para ajustar tal classe de modelos, deve-se levar em consideração a estrutura das variáveis, bem como se as observações estão próximas da origem. Caso isso não ocorra, pode-se ajustar o MRLS usual e posteriormente através do teste de hipóteses para $\mathcal{H}_0 : \beta_0 = 0$ reduzir para o modelo de intercepto nulo.

- Sob a forma funcional (15) e as suposições consideradas, tem-se que o EMQ de β_1 é obtido através da minimização de

$$S(\beta_1) := \sum_{i=1}^{n} (y_i - \beta_1 x_i)^2.$$

Modelos de regressão passando pela origem

Logo,

$$\frac{\partial S(\beta_1)}{\partial \beta_1} = \sum_{i=1}^{n}(-2x_i)(y_i - \beta_1 x_i)$$

resultando na equação normal

$$\sum_{i=1}^{n} x_i(y_i - \widehat{\beta}_1 x_i) = 0 \ \Leftrightarrow \ \widehat{\beta}_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i$$

$$\therefore \widehat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}, \qquad (16)$$

se $\sum_{i=1}^{n} x_i^2 > 0$, i.e., se tiver ao menos um valor da variável explicativa diferente de zero.

Modelos de regressão passando pela origem

Perceba que ao contrário do MRLS usual, a equação normal fornece apenas a identidade

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

Por outro lado, dado que

$$\frac{\partial^2 S(\beta_1)}{\partial \beta_1^2} = 2\sum_{i=1}^{n} x_i^2 > 0,$$

então $\widehat{\beta}_1$ dado em (16) é o EMQ de β_1.

Exercícios - Entregar na próxima aula

Exercício: Mostre que $\widehat{\beta}_1$ dado em (16) é uma combinação linear do vetor de variáveis respostas $\boldsymbol{y} = (y_1, \ldots, y_n)^\top$ e que

$$\mathbb{E}[\widehat{\beta}_1] = \beta_1 \text{ e } \mathrm{Var}[\widehat{\beta}_1] = \frac{\sigma^2}{\sum_{i=1}^{n} x_i^2}.$$

Exercício: Sob a hipótese de normalidade, mostre que

$$\widehat{\beta}_1 \sim \mathcal{N}\left(\beta_1, \sigma^2 / \sum_{i=1}^{n} x_i^2\right).$$

Exercício: Prove o teorema de Gauss-Markov para o MRLS de intercepto nulo, i.e., mostre que que $\widehat{\beta}_1$ dado em (16) é o BLUE de β_1.

Exercícios - Entregar na próxima aula

Exercício: Mostre que

$$\frac{\sum_{i=1}^{n}(y_i - \widehat{\beta}_1 x_i)^2}{n-1}.$$

é um estimador não viesado de σ^2.

Exercício: Sob a suposição de normalidade, mostre que β_1 e σ^2 são ortogonais. Além disso, mostre que os EMVs são dados, respectivamente, por

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2} \text{ e } \widehat{\sigma}^2 = \frac{\sum_{i=1}^{n}(y_i - \widehat{\beta}_1 x_i)^2}{n}.$$

MINQUE

O MINQUE de σ^2 é dado por

$$
\begin{aligned}
\widehat{\sigma}^2 &= \frac{\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2}{n-1} = \frac{\sum_{i=1}^{n}(y_i - \widehat{\beta}_1 x_i)^2}{n-1} := \frac{\mathrm{SQRes}}{n-1} = \mathrm{QMRes} \\
&= \frac{\sum_{i=1}^{n}(y_i^2 - 2\widehat{\beta}_1 y_i x_i + \widehat{\beta}_1^2 x_i^2)}{n-1} \\
&= \frac{\sum_{i=1}^{n} y_i^2 - 2\widehat{\beta}_1 \sum_{i=1}^{n} y_i x_i + \widehat{\beta}_1^2 \sum_{i=1}^{n} x_i^2}{n-1} \\
&= \frac{\sum_{i=1}^{n} y_i^2 - \widehat{\beta}_1 \sum_{i=1}^{n} y_i x_i}{n-1},
\end{aligned}
$$

dado que $\widehat{\beta}_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i$.

Exercício- Entregar próxima aula

Exercício: Sob a suposição de normalidade, mostre que

$$\frac{\text{SQRes}}{\sigma^2} \sim \chi^2_{(n-1)}.$$

e que $\widehat{\beta}_1 \perp \text{SQRes}$.

Quantidade pivotal

Logo, temos que uma quantidade pivotal para β_1 é dada por

$$t_1^* := \frac{\dfrac{\widehat{\beta}_1 - \beta_1}{\sqrt{\dfrac{\sigma^2}{\sum_{i=1}^n x_i^2}}}}{\sqrt{\dfrac{\mathrm{SQRes}}{\sigma^2(n-1)}}} = \frac{\widehat{\beta}_1 - \beta_1}{\sqrt{\dfrac{\mathrm{QMRes}}{\sum_{i=1}^n x_i^2}}} \sim t_{(n-1)}. \tag{17}$$

Exercício- Entregar próxima aula

Exercício: Utilizando o método da quantidade pivotal baseado em (17), mostre que os melhores intervalos de confiança de nível $1-\alpha$ para β_1, $\mu_{x_0} = \beta_1 x_0$ e intervalo de predição para y_{x_0}, são dados respectivamente por:

$$\begin{aligned}
\mathrm{IC}_{1-\alpha}(\beta_1) &= \left[\widehat{\beta}_1 \pm t_{(1-\alpha/2,n-1)}\sqrt{\frac{\mathrm{QMRes}}{s^*_{xx}}}\right], \\
\mathrm{IC}_{1-\alpha}(\mu_{x_0}) &= \left[\widehat{\mu}_{x_0} \pm t_{(1-\alpha/2,n-1)}\sqrt{\frac{x_0^2\mathrm{QMRes}}{s^*_{xx}}}\right] \ \mathrm{e} \\
\mathrm{IP}_{1-\alpha}(y_{x_0}) &= \left[\widehat{y}_{x_0} \pm t_{(1-\alpha/2,n-1)}\sqrt{\mathrm{QMRes}\left\{1+\frac{x_0^2}{s^*_{xx}}\right\}}\right],
\end{aligned}$$

em que $s^*_{xx} = \sum_{i=1}^{n} x_i^2$.

Exercício - Entregar próxima aula

Exercício: Considere o MRLS de intercpeto nulo 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

Podemos utilizar a quantidade pivotal (17) como estatística de teste para testar as hipóteses

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

Teste de hipóteses

- Para testar $\mathcal{H}_0 : \beta_1 = b_1$ vs $\mathcal{H}_1 : \beta_1 \neq b_1$, com 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 $t_1^* \sim t_{(n-1)}$.
- Rejeita-se $\mathcal{H}_0$ ao nível α se $|t_1^*| > t_{(1-\alpha/2, n-1)}$.
- O valor-p associado é dado por

$$\text{Valor-p} := 2\min\{\mathbb{P}[t_{(n-1)} > t_1^*]; \mathbb{P}[t_{(n-1)} < t_1^*]\}.$$

Teste de hipóteses

- Para testar $\mathcal{H}_0 : \beta_1 = b_1$ vs $\mathcal{H}_1 : \beta_1 \neq b_1$, com 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 $t_1^* \sim t_{(n-1)}$.
- Rejeita-se $\mathcal{H}_0$ ao nível α se $|t_1^*| > t_{(1-\alpha/2, n-1)}$.
- O valor-p associado é dado por

$$\text{Valor-p} := 2\min\{\mathbb{P}[t_{(n-1)} > t_1^*]; \mathbb{P}[t_{(n-1)} < t_1^*]\}.$$

Teste de hipóteses

- Para testar $\mathcal{H}_0 : \beta_1 = b_1$ vs $\mathcal{H}_1 : \beta_1 \neq b_1$, com 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 $t_1^* \sim t_{(n-1)}$.
- Rejeita-se $\mathcal{H}_0$ ao nível α se $|t_1^*| > t_{(1-\alpha/2, n-1)}$.
- O valor-p associado é dado por

$$\text{Valor-p} := 2\min\{\mathbb{P}[t_{(n-1)} > t_1^*]; \mathbb{P}[t_{(n-1)} < t_1^*]\}.$$

Teste de hipóteses

- Para testar $\mathcal{H}_0 : \beta_1 = b_1$ vs $\mathcal{H}_1 : \beta_1 \neq b_1$, com 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 $t_1^* \sim t_{(n-1)}$.
- Rejeita-se $\mathcal{H}_0$ ao nível α se $|t_1^*| > t_{(1-\alpha/2, n-1)}$.
- O valor-p associado é dado por

$$\text{Valor-p} := 2\min\{\mathbb{P}[t_{(n-1)} > t_1^*]; \mathbb{P}[t_{(n-1)} < t_1^*]\}.$$

Exercício - Entregar próxima aula

Exercício: Baseado na estatística de teste apresentada anteriormente, especifique a região crítica e como obter o valor-p nos seguintes casos:

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

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

Exercício: Considerando o MRLS de intercepto nulo e a suposição de normalidade, obtenha uma estatística de teste para testar as hipóteses

$$\mathcal{H}_0 : \mu_{x_0} = \mu_0 \text{ versus } \mathcal{H}_1 : \mu_{x_0} \neq (>,<)\mu_0,\ \mu_0 \text{ especificado},$$

ao nível de significância α, especificando região crítica e a forma de se obter o respectivo valor-p associado.

Decomposição da Soma de Quadrados Total

- No MRLS, utiliza-se a seguinte identidade

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

de forma que obtemos a seguinte decomposição

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

dado que $\sum_{i=1}^{n}(y_i - \widehat{y}_i)(\widehat{y}_i - \bar{y}_n) = 0$.

- A partir desta decomposição, foi definido o coeficiente de determinação

$$R^2 = \frac{\sum_{i=1}^{n}(\widehat{y}_i - \bar{y}_n)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_n)^2} = \frac{\text{SQReg}}{\text{SQT}},$$

que fornece a proporção da variabilidade da variável resposta, em torno de $\bar{y}_n$, explicada pelo modelo de regressão (variável explicativa).

Decomposição da Soma de Quadrados Total

- No MRLS, utiliza-se a seguinte identidade

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

de forma que obtemos a seguinte decomposição

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

dado que $\sum_{i=1}^{n}(y_i - \widehat{y}_i)(\widehat{y}_i - \bar{y}_n) = 0$.

- A partir desta decomposição, foi definido o coeficiente de determinação

$$R^2 = \frac{\sum_{i=1}^{n}(\widehat{y}_i - \bar{y}_n)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_n)^2} = \frac{\text{SQReg}}{\text{SQT}},$$

que fornece a proporção da variabilidade da variável resposta, em torno de $\bar{y}_n$, explicada pelo modelo de regressão (variável explicativa).

Decomposição da Soma de Quadrados Total

- No MRLS de intercepto nulo, a identidade $\sum_{i=1}^{n}(y_i - \widehat{y}_i) = 0$ **pode não se verificar**. portanto a decomposição padrão utilizada na ANOVA no MRLS nem sempre é válida.

- Uma proposta é considerar

$$y_i = \widehat{y}_i + (y_i - \widehat{y}_i), i = 1, \ldots, n$$

de forma que

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

- Dado que

$$\begin{aligned}\sum_{i=1}^{n} \widehat{y}_i (y_i - \widehat{y}_i) &= \sum_{i=1}^{n} \widehat{\beta}_1 x_i (y_i - \widehat{y}_i) \\ &= \widehat{\beta}_1 \sum_{i=1}^{n} x_i (y_i - \widehat{y}_i) = 0,\end{aligned}$$

pelo sistema de equações normais, resultante do processo de estimação associado ao MRLS de intercepto nulo.

Decomposição da Soma de Quadrados Total

- No MRLS de intercepto nulo, a identidade $\sum_{i=1}^{n}(y_i - \widehat{y}_i) = 0$ **pode não se verificar**. portanto a decomposição padrão utilizada na ANOVA no MRLS nem sempre é válida.
- Uma proposta é considerar

$$y_i = \widehat{y}_i + (y_i - \widehat{y}_i), i = 1, \ldots, n$$

de forma que

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

- Dado que

$$\begin{aligned}\sum_{i=1}^{n} \widehat{y}_i(y_i - \widehat{y}_i) &= \sum_{i=1}^{n} \widehat{\beta}_1 x_i(y_i - \widehat{y}_i) \\ &= \widehat{\beta}_1 \sum_{i=1}^{n} x_i(y_i - \widehat{y}_i) = 0,\end{aligned}$$

pelo sistema de equações normais, resultante do processo de estimação associado ao MRLS de intercepto nulo.

Decomposição da Soma de Quadrados Total

- No MRLS de intercepto nulo, a identidade $\sum_{i=1}^{n}(y_i - \widehat{y}_i) = 0$ **pode não se verificar**. portanto a decomposição padrão utilizada na ANOVA no MRLS nem sempre é válida.
- Uma proposta é considerar

$$y_i = \widehat{y}_i + (y_i - \widehat{y}_i), i = 1, \ldots, n$$

de forma que

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

- Dado que

$$\begin{aligned} \sum_{i=1}^{n} \widehat{y}_i(y_i - \widehat{y}_i) &= \sum_{i=1}^{n} \widehat{\beta}_1 x_i (y_i - \widehat{y}_i) \\ &= \widehat{\beta}_1 \sum_{i=1}^{n} x_i (y_i - \widehat{y}_i) = 0, \end{aligned}$$

pelo sistema de equações normais, resultante do processo de estimação associado ao MRLS de intercepto nulo.

Decomposição da Soma de Quadrados Total

- Então,

$$\begin{aligned} \sum_{i=1}^{n} y_i^2 &= \sum_{i=1}^{n} \widehat{y}_i^2 + \sum_{i=1}^{n} (y_i - \widehat{y}_i)^2 \\ \mathrm{SQT}_0 &= \mathrm{SQReg}_0 + \mathrm{SQRes}, \end{aligned} \tag{18}$$

com $\mathrm{SQT}_0, \mathrm{SQReg}_0$ e SQRes possuindo interpretação **similar** aos seus correspondentes no MRLS.

- De forma análoga, pode-se definir o coeficiente de determinação para o MRLS de intercepto nulo

$$R_0^2 = \frac{\sum_{i=1}^{n} \widehat{y}_i^2}{\sum_{i=1}^{n} y_i^2} = \frac{\mathrm{SQReg}_0}{\mathrm{SQT}_0},$$

que representa a proporção da variabilidade da variável resposta, **em torno da origem**, que é explicada pelo modelo de regressão.

Decomposição da Soma de Quadrados Total

- Então,

$$\begin{aligned} \sum_{i=1}^{n} y_i^2 &= \sum_{i=1}^{n} \widehat{y}_i^2 + \sum_{i=1}^{n} (y_i - \widehat{y}_i)^2 \\ \mathrm{SQT}_0 &= \mathrm{SQReg}_0 + \mathrm{SQRes}, \end{aligned} \tag{18}$$

com $\mathrm{SQT}_0, \mathrm{SQReg}_0$ e SQRes possuindo interpretação **similar** aos seus correspondentes no MRLS.

- De forma análoga, pode-se definir o coeficiente de determinação para o MRLS de intercepto nulo

$$R_0^2 = \frac{\sum_{i=1}^{n} \widehat{y}_i^2}{\sum_{i=1}^{n} y_i^2} = \frac{\mathrm{SQReg}_0}{\mathrm{SQT}_0},$$

que representa a proporção da variabilidade da variável resposta, **em torno da origem**, que é explicada pelo modelo de regressão.

Observações

- Em geral, $R_0^2 > R^2$, i.e., o coeficiente de determinação usual *subestima* a proporção da variabilidade explicada em modelos de intercepto nulo.
- Existem outras propostas de coeficiente de determinação em modelos de intercepto nulo. Para detalhes, veja por exemplo Montgomery et al. (2012).
- Existem vários papers versando sobre classes de modelos de intercepto nulo, por exemplo:
- Aoki, R., Achcar, J., Bolfarine, H. and Singer, J.M. (2003). Bayesian analysis of null intercept errors-in-variables regression for pretest/post-test data. Journal of Applied Statistics, **30**, 3-12. (doi: 10.1080/0266476022000018466).
- Singer, J.M. , **Nobre, J.S.** and Henry, C.S. (2004). Regression models for pretest/posttest data in blocks. Statistical Modelling, **4**, 324-338. (doi: 10.1191/1471082X04st083oa).
- Vários outros...

Observações

- Em geral, $R_0^2 > R^2$, i.e., o coeficiente de determinação usual *subestima* a proporção da variabilidade explicada em modelos de intercepto nulo.
- Existem outras propostas de coeficiente de determinação em modelos de intercepto nulo. Para detalhes, veja por exemplo Montgomery et al. (2012).
- Existem vários papers versando sobre classes de modelos de intercepto nulo, por exemplo:
- Aoki, R., Achcar, J., Bolfarine, H. and Singer, J.M. (2003). Bayesian analysis of null intercept errors-in-variables regression for pretest/post-test data. Journal of Applied Statistics, **30**, 3-12. (doi: 10.1080/0266476022000018466).
- Singer, J.M. , **Nobre, J.S.** and Henry, C.S. (2004). Regression models for pretest/posttest data in blocks. Statistical Modelling, **4**, 324-338. (doi: 10.1191/1471082X04st083oa).
- Vários outros...

Observações

- Em geral, $R_0^2 > R^2$, i.e., o coeficiente de determinação usual *subestima* a proporção da variabilidade explicada em modelos de intercepto nulo.
- Existem outras propostas de coeficiente de determinação em modelos de intercepto nulo. Para detalhes, veja por exemplo Montgomery et al. (2012).
- Existem vários papers versando sobre classes de modelos de intercepto nulo, por exemplo:
- Aoki, R., Achcar, J., Bolfarine, H. and Singer, J.M. (2003). Bayesian analysis of null intercept errors-in-variables regression for pretest/post-test data. Journal of Applied Statistics, **30**, 3-12. (doi: 10.1080/0266476022000018466).
- Singer, J.M. , **Nobre, J.S.** and Henry, C.S. (2004). Regression models for pretest/posttest data in blocks. Statistical Modelling, **4**, 324-338. (doi: 10.1191/1471082X04st083oa).
- Vários outros...

Observações

- Em geral, $R_0^2 > R^2$, i.e., o coeficiente de determinação usual *subestima* a proporção da variabilidade explicada em modelos de intercepto nulo.
- Existem outras propostas de coeficiente de determinação em modelos de intercepto nulo. Para detalhes, veja por exemplo Montgomery et al. (2012).
- Existem vários papers versando sobre classes de modelos de intercepto nulo, por exemplo:
- Aoki, R., Achcar, J., Bolfarine, H. and Singer, J.M. (2003). Bayesian analysis of null intercept errors-in-variables regression for pretest/post-test data. Journal of Applied Statistics, **30**, 3-12. (doi: 10.1080/0266476022000018466).
- Singer, J.M. , **Nobre, J.S.** and Henry, C.S. (2004). Regression models for pretest/posttest data in blocks. Statistical Modelling, **4**, 324-338. (doi: 10.1191/1471082X04st083oa).
- Vários outros...

Observações

- Em geral, $R_0^2 > R^2$, i.e., o coeficiente de determinação usual *subestima* a proporção da variabilidade explicada em modelos de intercepto nulo.
- Existem outras propostas de coeficiente de determinação em modelos de intercepto nulo. Para detalhes, veja por exemplo Montgomery et al. (2012).
- Existem vários papers versando sobre classes de modelos de intercepto nulo, por exemplo:
- Aoki, R., Achcar, J., Bolfarine, H. and Singer, J.M. (2003). Bayesian analysis of null intercept errors-in-variables regression for pretest/post-test data. Journal of Applied Statistics, **30**, 3-12. (doi: 10.1080/0266476022000018466).
- Singer, J.M. , **Nobre, J.S.** and Henry, C.S. (2004). Regression models for pretest/posttest data in blocks. Statistical Modelling, **4**, 324-338. (doi: 10.1191/1471082X04st083oa).
- Vários outros...

Observações

- Em geral, $R_0^2 > R^2$, i.e., o coeficiente de determinação usual *subestima* a proporção da variabilidade explicada em modelos de intercepto nulo.
- Existem outras propostas de coeficiente de determinação em modelos de intercepto nulo. Para detalhes, veja por exemplo Montgomery et al. (2012).
- Existem vários papers versando sobre classes de modelos de intercepto nulo, por exemplo:
- Aoki, R., Achcar, J., Bolfarine, H. and Singer, J.M. (2003). Bayesian analysis of null intercept errors-in-variables regression for pretest/post-test data. Journal of Applied Statistics, **30**, 3-12. (doi: 10.1080/0266476022000018466).
- Singer, J.M. , **Nobre, J.S.** and Henry, C.S. (2004). Regression models for pretest/posttest data in blocks. Statistical Modelling, **4**, 324-338. (doi: 10.1191/1471082X04st083oa).
- Vários outros...

Exercício - Entregar próxima aula

Exercício: Considere o MRLS de intercepto nulo adicionada a suposição de normalidade. Enuncie e prove o Teorema de Cochran correspondente, i.e., sob $\mathcal{H}_0 : \beta_1 = 0$, temos que

i) SQRes e SQReg_0 são independentes.

ii) $\dfrac{\mathrm{SQT}_0}{\sigma^2} \sim \chi^2_{(n)}$.

iii) $\dfrac{\mathrm{SQRes}}{\sigma^2} \sim \chi^2_{(n-1)}$.

iv) $\dfrac{\mathrm{SQReg}_0}{\sigma^2} \sim \chi^2_1$.

Exercício: Baseado na decomposição (18) e sob a suposição de normalidade, obtenha a tabela de ANOVA, inclusive estatística F, para o modelo de regressão linear simples com intercepto nulo.

Transformações estabilizadoras da variância

- Considere o MRLS com as suas pressuposições básicas e sua forma funcional

$$y_i = \beta_0 + \beta_1 x_i + e_i,\ e_i \sim \mathrm{RB}(0, \sigma^2),\ i = 1, \ldots, n,$$

implicando que $y_i \overset{\text{nc}}{\sim} (\beta_0 + \beta_1 x_i, \sigma^2),\ i = 1, \ldots, n$, em que $\overset{\text{nc}}{\sim}$ denota que as variáveis são não-correlacionadas.

- Por simplicidade, a partir de agora, denotar-se-á $\mu_i := \beta_0 + \beta_1 x_i$.
- Considere que a forma funcional permanece válida, mas que exista heteroscedasticidade, especificamente

$$y_i \overset{\text{nc}}{\sim} (\mu_i, \mu_i),\ \mu_i > 0,\ i = 1, \ldots, n,$$

o que ocorre, por exemplo, se $y_i \sim \mathrm{P}(\mu_i)$.

- Em uma situação como essa, não se pode utilizar o método de mínimos quadrados usual para estimar os parâmetros de regressão, dado que neste caso não há garantia alguma de bons resultados (o método não irá gerar o BLUE, por exemplo).

Transformações estabilizadoras da variância

- Considere o MRLS com as suas pressuposições básicas e sua forma funcional

$$y_i = \beta_0 + \beta_1 x_i + e_i,\ e_i \sim \mathrm{RB}(0, \sigma^2),\ i = 1, \ldots, n,$$

implicando que $y_i \overset{\mathrm{nc}}{\sim} (\beta_0 + \beta_1 x_i, \sigma^2),\ i = 1, \ldots, n$, em que $\overset{\mathrm{nc}}{\sim}$ denota que as variáveis são não-correlacionadas.

- Por simplicidade, a partir de agora, denotar-se-á $\mu_i := \beta_0 + \beta_1 x_i$.

- Considere que a forma funcional permanece válida, mas que exista heteroscedasticidade, especificamente

$$y_i \overset{\mathrm{nc}}{\sim} (\mu_i, \mu_i),\ \mu_i > 0,\ i = 1, \ldots, n,$$

o que ocorre, por exemplo, se $y_i \sim \mathrm{P}(\mu_i)$.

- Em uma situação como essa, não se pode utilizar o método de mínimos quadrados usual para estimar os parâmetros de regressão, dado que neste caso não há garantia alguma de bons resultados (o método não irá gerar o BLUE, por exemplo).

Transformações estabilizadoras da variância

- Considere o MRLS com as suas pressuposições básicas e sua forma funcional

$$y_i = \beta_0 + \beta_1 x_i + e_i,\ e_i \sim \mathrm{RB}(0, \sigma^2),\ i = 1, \ldots, n,$$

implicando que $y_i \overset{\mathrm{nc}}{\sim} (\beta_0 + \beta_1 x_i, \sigma^2),\ i = 1, \ldots, n$, em que $\overset{\mathrm{nc}}{\sim}$ denota que as variáveis são não-correlacionadas.

- Por simplicidade, a partir de agora, denotar-se-á $\mu_i := \beta_0 + \beta_1 x_i$.
- Considere que a forma funcional permanece válida, mas que exista heteroscedasticidade, especificamente

$$y_i \overset{\mathrm{nc}}{\sim} (\mu_i, \mu_i),\ \mu_i > 0,\ i = 1, \ldots, n,$$

o que ocorre, por exemplo, se $y_i \sim \mathrm{P}(\mu_i)$.

- Em uma situação como essa, não se pode utilizar o método de mínimos quadrados usual para estimar os parâmetros de regressão, dado que neste caso não há garantia alguma de bons resultados (o método não irá gerar o BLUE, por exemplo).

Transformações estabilizadoras da variância

- Considere o MRLS com as suas pressuposições básicas e sua forma funcional

$$y_i = \beta_0 + \beta_1 x_i + e_i,\ e_i \sim \mathrm{RB}(0, \sigma^2),\ i = 1, \ldots, n,$$

implicando que $y_i \overset{\text{nc}}{\sim} (\beta_0 + \beta_1 x_i, \sigma^2),\ i = 1, \ldots, n$, em que $\overset{\text{nc}}{\sim}$ denota que as variáveis são não-correlacionadas.

- Por simplicidade, a partir de agora, denotar-se-á $\mu_i := \beta_0 + \beta_1 x_i$.
- Considere que a forma funcional permanece válida, mas que exista heteroscedasticidade, especificamente

$$y_i \overset{\text{nc}}{\sim} (\mu_i, \mu_i),\ \mu_i > 0,\ i = 1, \ldots, n,$$

o que ocorre, por exemplo, se $y_i \sim \mathrm{P}(\mu_i)$.

- Em uma situação como essa, não se pode utilizar o método de mínimos quadrados usual para estimar os parâmetros de regressão, dado que neste caso não há garantia alguma de bons resultados (o método não irá gerar o BLUE, por exemplo).

Resultado importante

Método Delta

Considere X uma variável aleatória com $\mathbb{E}[X] = \mu$ e $\mathrm{Var}[X] = \sigma^2$, então para $g : \mathcal{X} \mapsto \mathbb{R}$ diferenciável com $g^{(1)}(\mu) \neq 0$, tem-se que

$$\mathbb{E}[g(X)] \approx g(\mu) \text{ e } \mathrm{Var}[g(X)] \approx (g^{(1)}(\mu))^2\sigma^2.$$

- Este resultado segue diretamente do uso da expansão em série de Taylor estocástica de primeira ordem, especificamente

$$g(X) = g(\mu) + g^{(1)}(\mu)(X - \mu) + o_p(1),$$

no qual o resultado é obtido por operações elementares.

- O resultado é aproximado no sentido de ser assintoticamente válido.

Resultado importante

Método Delta

Considere X uma variável aleatória com $\mathbb{E}[X] = \mu$ e $\mathrm{Var}[X] = \sigma^2$, então para $g : \mathcal{X} \mapsto \mathbb{R}$ diferenciável com $g^{(1)}(\mu) \neq 0$, tem-se que

$$\mathbb{E}[g(X)] \approx g(\mu) \text{ e } \mathrm{Var}[g(X)] \approx (g^{(1)}(\mu))^2\sigma^2.$$

- Este resultado segue diretamente do uso da expansão em série de Taylor estocástica de primeira ordem, especificamente

$$g(X) = g(\mu) + g^{(1)}(\mu)(X - \mu) + o_p(1),$$

  no qual o resultado é obtido por operações elementares.

- O resultado é aproximado no sentido de ser assintoticamente válido.

Resultado importante

Método Delta

Considere X uma variável aleatória com $\mathbb{E}[X] = \mu$ e $\mathrm{Var}[X] = \sigma^2$, então para $g : \mathcal{X} \mapsto \mathbb{R}$ diferenciável com $g^{(1)}(\mu) \neq 0$, tem-se que

$$\mathbb{E}[g(X)] \approx g(\mu) \text{ e } \mathrm{Var}[g(X)] \approx (g^{(1)}(\mu))^2\sigma^2.$$

- Este resultado segue diretamente do uso da expansão em série de Taylor estocástica de primeira ordem, especificamente

$$g(X) = g(\mu) + g^{(1)}(\mu)(X - \mu) + o_p(1),$$

  no qual o resultado é obtido por operações elementares.

- O resultado é aproximado no sentido de ser assintoticamente válido.

Transformações estabilizadoras da variância

- Voltando ao MRLS, considere que $y_i \overset{\text{ind}}{\sim} \mathrm{P}(\mu_i)$.
- Para utilizarmos o método de Mínimos Quadrados Ordinários é necessário que a fonte de variação/variável resposta seja **homoscedástica**.
- Para este fim, uma alternativa é considerar uma transformação $g(y_i)$, tal que $\mathrm{Var}[g(y_i)] = \sigma^2$ (constante).
- No nosso caso, $y_i \overset{\text{ind}}{\sim} \mathrm{P}(\mu_i)$, então bastar considerar g tal que

$$\begin{aligned}
&\mathrm{Var}[g(y_i)] = (g^{(1)}(\mu_i))^2\mu_i = c > 0 \\
&\Rightarrow g^{(1)}(\mu_i) = k\mu_i^{-1/2}, \text{i.e., } \frac{d}{d\mu}g(\mu)\bigg|_{\mu=\mu_i} = k\mu_i^{-1/2} \\
&\Leftrightarrow g(\mu) = \int \frac{d}{d\mu}g(\mu)d\mu = \int k\mu^{-1/2}d\mu = \frac{k}{2}\sqrt{\mu}
\end{aligned}$$

de forma que

$$\therefore g(\mu_i) \propto \sqrt{\mu_i}.$$

Transformações estabilizadoras da variância

- Voltando ao MRLS, considere que $y_i \overset{\text{ind}}{\sim} \mathrm{P}(\mu_i)$.
- Para utilizarmos o método de Mínimos Quadrados Ordinários é necessário que a fonte de variação/variável resposta seja **homoscedástica**.
- Para este fim, uma alternativa é considerar uma transformação $g(y_i)$, tal que $\mathrm{Var}[g(y_i)] = \sigma^2$ (constante).
- No nosso caso, $y_i \overset{\text{ind}}{\sim} \mathrm{P}(\mu_i)$, então bastar considerar g tal que

$$\begin{aligned} &\mathrm{Var}[g(y_i)] = (g^{(1)}(\mu_i))^2 \mu_i = c > 0 \\ &\Rightarrow g^{(1)}(\mu_i) = k\mu_i^{-1/2}, \text{i.e.}, \left.\frac{d}{d\mu} g(\mu)\right|_{\mu=\mu_i} = k\mu_i^{-1/2} \\ &\Leftrightarrow g(\mu) = \int \frac{d}{d\mu} g(\mu) d\mu = \int k\mu^{-1/2} d\mu = \frac{k}{2}\sqrt{\mu} \end{aligned}$$

de forma que

$$\therefore g(\mu_i) \propto \sqrt{\mu_i}.$$

Transformações estabilizadoras da variância

- Voltando ao MRLS, considere que $y_i \overset{\text{ind}}{\sim} \mathrm{P}(\mu_i)$.
- Para utilizarmos o método de Mínimos Quadrados Ordinários é necessário que a fonte de variação/variável resposta seja **homoscedástica**.
- Para este fim, uma alternativa é considerar uma transformação $g(y_i)$, tal que $\mathrm{Var}[g(y_i)] = \sigma^2$ (constante).
- No nosso caso, $y_i \overset{\text{ind}}{\sim} \mathrm{P}(\mu_i)$, então bastar considerar g tal que

$$\begin{aligned}
&\mathrm{Var}[g(y_i)] = (g^{(1)}(\mu_i))^2\mu_i = c > 0\\
&\Rightarrow g^{(1)}(\mu_i) = k\mu_i^{-1/2}, \text{i.e.}, \left.\frac{d}{d\mu}g(\mu)\right|_{\mu=\mu_i} = k\mu_i^{-1/2}\\
&\Leftrightarrow g(\mu) = \int \frac{d}{d\mu}g(\mu)d\mu = \int k\mu^{-1/2}d\mu = \frac{k}{2}\sqrt{\mu}
\end{aligned}$$

de forma que

$$\therefore g(\mu_i) \propto \sqrt{\mu_i}.$$

Transformações estabilizadoras da variância

- Voltando ao MRLS, considere que $y_i \stackrel{\text{ind}}{\sim} \mathrm{P}(\mu_i)$.
- Para utilizarmos o método de Mínimos Quadrados Ordinários é necessário que a fonte de variação/variável resposta seja **homoscedástica**.
- Para este fim, uma alternativa é considerar uma transformação $g(y_i)$, tal que $\mathrm{Var}[g(y_i)] = \sigma^2$ (constante).
- No nosso caso, $y_i \stackrel{\text{ind}}{\sim} \mathrm{P}(\mu_i)$, então bastar considerar g tal que

$$
\begin{aligned}
&\mathrm{Var}[g(y_i)] = (g^{(1)}(\mu_i))^2\mu_i = c > 0 \\
&\Rightarrow g^{(1)}(\mu_i) = k\mu_i^{-1/2}, \text{i.e., } \frac{d}{d\mu}g(\mu)\bigg|_{\mu=\mu_i} = k\mu_i^{-1/2} \\
&\Leftrightarrow g(\mu) = \int \frac{d}{d\mu}g(\mu)d\mu = \int k\mu^{-1/2}d\mu = \frac{k}{2}\sqrt{\mu}
\end{aligned}
$$

de forma que

$$\therefore g(\mu_i) \propto \sqrt{\mu_i}.$$

Transformações estabilizadoras da variância

- Logo, se $y_i \overset{\text{ind}}{\sim} \mathrm{P}(\mu_i), i = 1, \ldots, n$, então

$$\sqrt{y_i} \overset{\text{ind}}{\sim} (\sqrt{\mu_i}, \sigma^2), i = 1, \ldots, n.$$

- Se $\mu_i \to \infty$, então pelo TCL

$$\sqrt{y_i} \sim \mathcal{AN}(\sqrt{\mu_i}, \sigma^2), i = 1, \ldots, n.$$

- Apesar de ser possível encontrar uma transformação que estabiliza a variância, o processo em si pode ser extremamente complicado, em especial, ao que tange a interpretabilidade do modelo a ser considerado, fugindo assim da ideia da parcimônia.

Transformações estabilizadoras da variância

- Logo, se $y_i \overset{\text{ind}}{\sim} \mathrm{P}(\mu_i), i = 1, \ldots, n$, então

$$\sqrt{y_i} \overset{\text{ind}}{\sim} (\sqrt{\mu_i}, \sigma^2), i = 1, \ldots, n.$$

- Se $\mu_i \to \infty$, então pelo TCL

$$\sqrt{y_i} \sim A\mathcal{N}(\sqrt{\mu_i}, \sigma^2), i = 1, \ldots, n.$$

- Apesar de ser possível encontrar uma transformação que estabiliza a variância, o processo em si pode ser extremamente complicado, em especial, ao que tange a interpretabilidade do modelo a ser considerado, fugindo assim da ideia da parcimônia.

Transformações estabilizadoras da variância

- Logo, se $y_i \overset{\text{ind}}{\sim} \mathrm{P}(\mu_i), i = 1, \ldots, n$, então

$$\sqrt{y_i} \overset{\text{ind}}{\sim} (\sqrt{\mu_i}, \sigma^2), i = 1, \ldots, n.$$

- Se $\mu_i \to \infty$, então pelo TCL

$$\sqrt{y_i} \sim A\mathcal{N}(\sqrt{\mu_i}, \sigma^2), i = 1, \ldots, n.$$

- Apesar de ser possível encontrar uma transformação que estabiliza a variância, o processo em si pode ser extremamente complicado, em especial, ao que tange a interpretabilidade do modelo a ser considerado, fugindo assim da ideia da parcimônia.

Exercício - Entregar próxima aula

Exercício: Para cada um dos casos abaixo, determine a transformação estabilizadora da variância

i) $y_i \stackrel{\text{ind}}{\sim} \mathrm{B}(\mu_i), \mu \in (0,1), i = 1, \ldots, n.$

ii) $y_i \stackrel{\text{ind}}{\sim} (\mu_i, \mu_i^2), \mu_i \in \mathbb{R}, i = 1, \ldots, n.$

iii) $y_i \stackrel{\text{ind}}{\sim} (\mu_i, \mu_i^3), \mu_i > 0, i = 1, \ldots, n.$

iv) $y_i \stackrel{\text{ind}}{\sim} (\mu_i, \mu_i^4), \mu_i \in \mathbb{R}, i = 1, \ldots, n.$

v) $y_i \stackrel{\text{ind}}{\sim} (\mu_i, (1 + \mu_i^2)^2), \mu_i \in \mathbb{R}, i = 1, \ldots, n.$

Transformações estabilizadoras da variância

- A técnica utilizada anteriormente, se preocupa única e exclusivamente na estabilização da variância.
- Durante muitos anos, os modelos baseados na **suposição de normalidade** foram utilizados para descrever a maioria dos fenômenos de interesse. Mesmo quando o fenômeno sob estudo não apresentava características de normalidade, tentava-se algum tipo de transformação no sentido de alcançar a normalidade procurada. 🙂
- Provavelmente a transformação mais conhecida foi proposta por Box and Cox (1964, JRSS B), a qual transforma a variável resposta em

$$z := \begin{cases} \dfrac{y^{\lambda}-1}{\lambda} & , \text{ se } \lambda \neq 0 \\ \ln y & , \text{ se } \lambda = 0, \end{cases}$$

sendo λ uma constante desconhecida.

Transformações estabilizadoras da variância

- A técnica utilizada anteriormente, se preocupa única e exclusivamente na estabilização da variância.
- Durante muitos anos, os modelos baseados na **suposição de normalidade** foram utilizados para descrever a maioria dos fenômenos de interesse. Mesmo quando o fenômeno sob estudo não apresentava características de normalidade, tentava-se algum tipo de transformação no sentido de alcançar a normalidade procurada. 😕
- Provavelmente a transformação mais conhecida foi proposta por Box and Cox (1964, JRSS B), a qual transforma a variável resposta em

$$z := \begin{cases} \dfrac{y^\lambda - 1}{\lambda} & , \text{ se } \lambda \neq 0 \\ \ln y & , \text{ se } \lambda = 0, \end{cases}$$

sendo λ uma constante desconhecida.

Transformações estabilizadoras da variância

- A técnica utilizada anteriormente, se preocupa única e exclusivamente na estabilização da variância.
- Durante muitos anos, os modelos baseados na **suposição de normalidade** foram utilizados para descrever a maioria dos fenômenos de interesse. Mesmo quando o fenômeno sob estudo não apresentava características de normalidade, tentava-se algum tipo de transformação no sentido de alcançar a normalidade procurada. 😕
- Provavelmente a transformação mais conhecida foi proposta por Box and Cox (1964, JRSS B), a qual transforma a variável resposta em

$$z := \begin{cases} \dfrac{y^{\lambda}-1}{\lambda} & , \text{ se } \lambda \neq 0 \\ \ln y & , \text{ se } \lambda = 0, \end{cases}$$

sendo λ uma constante desconhecida.

Transformações estabilizadoras da variância

- Box and Cox (1964) sugerem estimar os parâmetros de regressão, λ e σ^2 usando o método de máxima verosimilhança, e neste caso, deve-se conhecer a verossimilhança original, bem como fazer uso da suposição de independência. Quando a estimativa de λ estiver próxima de um é um indicativo que a variável original possui distribuição normal.
- Outras classes de transformações se popularizaram, como a classe Aranda-Ordaz, Misturas-Skew, etc...
- Acreditava-se que a transformação de Box-Cox gerava/produzia normalidade aproximada e também estabilizava a variância, no entanto, isso raramente ocorre para um único valor de λ.
- **Felizmente**, hoje em dia, não é mais necessário se preocupar com transformações, pois podemos modelar diretamente a variável resposta, mesmo nos casos heteroscedásticos, por exemplo fazendo uso de modelos lineares generalizados, modelos simétricos, modelos simétricos heteroscedásticos, modelos lineares generalizados simétricos, modelos assimétricos heteroscedásticos, modelo de regressão Beta, Gama Unitária, etc...

Transformações estabilizadoras da variância

- Box and Cox (1964) sugerem estimar os parâmetros de regressão, λ e σ^2 usando o método de máxima verosimilhança, e neste caso, deve-se conhecer a verossimilhança original, bem como fazer uso da suposição de independência. Quando a estimativa de λ estiver próxima de um é um indicativo que a variável original possui distribuição normal.
- Outras classes de transformações se popularizaram, como a classe Aranda-Ordaz, Misturas-Skew, etc...
- Acreditava-se que a transformação de Box-Cox gerava/produzia normalidade aproximada e também estabilizava a variância, no entanto, isso raramente ocorre para um único valor de λ.
- **Felizmente**, hoje em dia, não é mais necessário se preocupar com transformações, pois podemos modelar diretamente a variável resposta, mesmo nos casos heteroscedásticos, por exemplo fazendo uso de modelos lineares generalizados, modelos simétricos, modelos simétricos heteroscedásticos, modelos lineares generalizados simétricos, modelos assimétricos heteroscedásticos, modelo de regressão Beta, Gama Unitária, etc...

Transformações estabilizadoras da variância

- Box and Cox (1964) sugerem estimar os parâmetros de regressão, λ e σ^2 usando o método de máxima verosimilhança, e neste caso, deve-se conhecer a verossimilhança original, bem como fazer uso da suposição de independência. Quando a estimativa de λ estiver próxima de um é um indicativo que a variável original possui distribuição normal.
- Outras classes de transformações se popularizaram, como a classe Aranda-Ordaz, Misturas-Skew, etc...
- Acreditava-se que a transformação de Box-Cox gerava/produzia normalidade aproximada e também estabilizava a variância, no entanto, isso raramente ocorre para um único valor de λ.
- **Felizmente**, hoje em dia, não é mais necessário se preocupar com transformações, pois podemos modelar diretamente a variável resposta, mesmo nos casos heteroscedásticos, por exemplo fazendo uso de modelos lineares generalizados, modelos simétricos, modelos simétricos heteroscedásticos, modelos lineares generalizados simétricos, modelos assimétricos heteroscedásticos, modelo de regressão Beta, Gama Unitária, etc...

Transformações estabilizadoras da variância

- Box and Cox (1964) sugerem estimar os parâmetros de regressão, λ e σ^2 usando o método de máxima verosimilhança, e neste caso, deve-se conhecer a verossimilhança original, bem como fazer uso da suposição de independência. Quando a estimativa de λ estiver próxima de um é um indicativo que a variável original possui distribuição normal.
- Outras classes de transformações se popularizaram, como a classe Aranda-Ordaz, Misturas-Skew, etc...
- Acreditava-se que a transformação de Box-Cox gerava/produzia normalidade aproximada e também estabilizava a variância, no entanto, isso raramente ocorre para um único valor de λ.
- **Felizmente**, hoje em dia, não é mais necessário se preocupar com transformações, pois podemos modelar diretamente a variável resposta, mesmo nos casos heteroscedásticos, por exemplo fazendo uso de modelos lineares generalizados, modelos simétricos, modelos simétricos heteroscedásticos, modelos lineares generalizados simétricos, modelos assimétricos heteroscedásticos, modelo de regressão Beta, Gama Unitária, etc...

Modelos linearizáveis

- Por vezes, podemos nos deparar com modelos **não lineares nos parâmetros**, mas que se tornam lineares através de simples transformações e podem ser ajustados de forma similar ao MRLS.

- Por exemplo, considere o modelo multiplicativo

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

que é não-linear, todavia é linearizável, pois se $y_i > 0$

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

ou equivalentemente,

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

- O modelo acima pode ser ajustado como um MRLS com variável resposta y^*, variável explicativa x_i e fonte de variação e_i^*, desde que as pressuposições sejam atendidas.

Modelos linearizáveis

- Por vezes, podemos nos deparar com modelos **não lineares nos parâmetros**, mas que se tornam lineares através de simples transformações e podem ser ajustados de forma similar ao MRLS.

- Por exemplo, considere o modelo multiplicativo

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

que é não-linear, todavia é linearizável, pois se $y_i > 0$

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

ou equivalentemente,

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

- O modelo acima pode ser ajustado como um MRLS com variável resposta y^*, variável explicativa x_i e fonte de variação e_i^*, desde que as pressuposições sejam atendidas.

Modelos linearizáveis

- Por vezes, podemos nos deparar com modelos **não lineares nos parâmetros**, mas que se tornam lineares através de simples transformações e podem ser ajustados de forma similar ao MRLS.

- Por exemplo, considere o modelo multiplicativo

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

que é não-linear, todavia é linearizável, pois se $y_i > 0$

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

ou equivalentemente,

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

- O modelo acima pode ser ajustado como um MRLS com variável resposta y^*, variável explicativa x_i e fonte de variação e_i^*, desde que as pressuposições sejam atendidas.

Modelos linearizáveis

- Considerando válida as pressuposições, podemos determinar os EMQ de β_0^* e β_1 de forma usual.
- Se a amostra não for grande, então precisa-se supor normalidade por parte de $e_i^* = \ln e_i$, ou de forma equivalente, supor que e_i são iid com distribuição log-normal, $i = 1, \ldots, n$, para conseguir fazer inferência de segunda ordem (IC, testes de hipóteses, etc...).
- Para o parâmetro β_1, nenhuma operação é necessária, dado que ele permanece inalterado no modelo.
- Já com relação a β_0 isso não ocorre, dado que o parâmetro em si agora é $\ln \beta_0$. Dentro do contexto de estimação por máxima verossimilhança, basta usar a propriedade da **equivariância/invariância** para, dado o EMV de $\ln \beta_0$, obtermos o estimador de β_0. Considerando outros métodos de estimação, isso em geral não se verifica.
- Para obtenção de IC, testes de hipóteses relacionados a β_0, se faz necessário o uso do método delta, em especial para obter o erro-padrão de $\widehat{\beta}_0$, pois obtemos diretamente o erro-padrão de $\widehat{\beta}_0^* = \widehat{\ln \beta_0}$.

Modelos linearizáveis

- Considerando válida as pressuposições, podemos determinar os EMQ de β_0^* e β_1 de forma usual.
- Se a amostra não for grande, então precisa-se supor normalidade por parte de $e_i^* = \ln e_i$, ou de forma equivalente, supor que e_i são iid com distribuição log-normal, $i = 1, \ldots, n$, para conseguir fazer inferência de segunda ordem (IC, testes de hipóteses, etc...).
- Para o parâmetro β_1, nenhuma operação é necessária, dado que ele permanece inalterado no modelo.
- Já com relação a β_0 isso não ocorre, dado que o parâmetro em si agora é $\ln \beta_0$. Dentro do contexto de estimação por máxima verossimilhança, basta usar a propriedade da **equivariância/invariância** para, dado o EMV de $\ln \beta_0$, obtermos o estimador de β_0. Considerando outros métodos de estimação, isso em geral não se verifica.
- Para obtenção de IC, testes de hipóteses relacionados a β_0, se faz necessário o uso do método delta, em especial para obter o erro-padrão de $\widehat{\beta}_0$, pois obtemos diretamente o erro-padrão de $\widehat{\beta}_0^* = \widehat{\ln \beta}_0$.

Modelos linearizáveis

- Considerando válida as pressuposições, podemos determinar os EMQ de β_0^* e β_1 de forma usual.
- Se a amostra não for grande, então precisa-se supor normalidade por parte de $e_i^* = \ln e_i$, ou de forma equivalente, supor que e_i são iid com distribuição log-normal, $i = 1, \ldots, n$, para conseguir fazer inferência de segunda ordem (IC, testes de hipóteses, etc...).
- Para o parâmetro β_1, nenhuma operação é necessária, dado que ele permanece inalterado no modelo.
- Já com relação a β_0 isso não ocorre, dado que o parâmetro em si agora é $\ln \beta_0$. Dentro do contexto de estimação por máxima verossimilhança, basta usar a propriedade da **equivariância/invariância** para, dado o EMV de $\ln \beta_0$, obtermos o estimador de β_0. Considerando outros métodos de estimação, isso em geral não se verifica.
- Para obtenção de IC, testes de hipóteses relacionados a β_0, se faz necessário o uso do método delta, em especial para obter o erro-padrão de $\widehat{\beta}_0$, pois obtemos diretamente o erro-padrão de $\widehat{\beta}_0^* = \widehat{\ln \beta}_0$.

Modelos linearizáveis

- Considerando válida as pressuposições, podemos determinar os EMQ de β_0^* e β_1 de forma usual.
- Se a amostra não for grande, então precisa-se supor normalidade por parte de $e_i^* = \ln e_i$, ou de forma equivalente, supor que e_i são iid com distribuição log-normal, $i = 1, \ldots, n$, para conseguir fazer inferência de segunda ordem (IC, testes de hipóteses, etc...).
- Para o parâmetro β_1, nenhuma operação é necessária, dado que ele permanece inalterado no modelo.
- Já com relação a β_0 isso não ocorre, dado que o parâmetro em si agora é $\ln \beta_0$. Dentro do contexto de estimação por máxima verossimilhança, basta usar a propriedade da **equivariância/invariância** para, dado o EMV de $\ln \beta_0$, obtermos o estimador de β_0. Considerando outros métodos de estimação, isso em geral não se verifica.
- Para obtenção de IC, testes de hipóteses relacionados a β_0, se faz necessário o uso do método delta, em especial para obter o erro-padrão de $\widehat{\beta}_0$, pois obtemos diretamente o erro-padrão de $\widehat{\beta}_0^* = \widehat{\ln \beta_0}$.

Modelos linearizáveis

- Considerando válida as pressuposições, podemos determinar os EMQ de β_0^* e β_1 de forma usual.
- Se a amostra não for grande, então precisa-se supor normalidade por parte de $e_i^* = \ln e_i$, ou de forma equivalente, supor que e_i são iid com distribuição log-normal, $i = 1, \ldots, n$, para conseguir fazer inferência de segunda ordem (IC, testes de hipóteses, etc...).
- Para o parâmetro β_1, nenhuma operação é necessária, dado que ele permanece inalterado no modelo.
- Já com relação a β_0 isso não ocorre, dado que o parâmetro em si agora é $\ln \beta_0$. Dentro do contexto de estimação por máxima verossimilhança, basta usar a propriedade da **equivariância/invariância** para, dado o EMV de $\ln \beta_0$, obtermos o estimador de β_0. Considerando outros métodos de estimação, isso em geral não se verifica.
- Para obtenção de IC, testes de hipóteses relacionados a β_0, se faz necessário o uso do método delta, em especial para obter o erro-padrão de $\widehat{\beta}_0$, pois obtemos diretamente o erro-padrão de $\widehat{\beta}_0^* = \widehat{\ln \beta}_0$.

Modelos linearizáveis

- Dado que $\beta_0 = \exp\{\beta_0^*\}$, então o EMV é tal que $\widehat{\beta}_0 = \exp\{\widehat{\beta}_0^*\}$, implicando que

$$\mathrm{EP}(\widehat{\beta}_0) = \exp\{\widehat{\beta}_0^*\}\mathrm{EP}(\widehat{\beta}_0^*).$$

- Obtido o $\mathrm{EP}(\widehat{\beta}_0)$ procedemos de forma similar ao que já foi estudado para obter IC e testes de hipóteses associados ao parâmetro β_0 de interesse.

Modelos linearizáveis

- Dado que $\beta_0 = \exp\{\beta_0^*\}$, então o EMV é tal que $\widehat{\beta}_0 = \exp\{\widehat{\beta}_0^*\}$, implicando que

$$\mathrm{EP}(\widehat{\beta}_0) = \exp\{\widehat{\beta}_0^*\}\mathrm{EP}(\widehat{\beta}_0^*).$$

- Obtido o $\mathrm{EP}(\widehat{\beta}_0)$ procedemos de forma similar ao que já foi estudado para obter IC e testes de hipóteses associados ao parâmetro β_0 de interesse.

Exercício - Entregar próxima aula

Exercício: Faça o gráfico da função de regressão associado ao modelo multiplicativo

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

para situações em que $\beta_1 > 0$ e $\beta_1 < 0$.

Exercício: Nos casos abaixo, encontre a forma linearizável e faça o gráfico da função de regressão. Adicionalmente, descreva os procedimentos e suposições necessários para obter os IC's para os parâmetros de regressão.

i) $y_i = \beta_0 x_i^{\beta_1} e_i, i = 1, \ldots, n.$

ii) $\dfrac{1}{y_i} = \beta_0 + \dfrac{1}{x_i}\beta_1 + e_i, i = 1, \ldots, n.$

iii) $y_i = \beta_0 + \beta_1 \ln x_i + e_i, i = 1, \ldots, n.$

iv) $\mu(\boldsymbol{x}; \boldsymbol{\beta}) = \dfrac{x_i}{\beta_0 x_i - \beta_1}, i = 1, \ldots, n.$