Regressão Linear – Abordagem Inferencial

Adaptado de Cadima, J. (2021). O Modelos Linear. ISA-UL.

Regressão Linear - Inferência

- Até aqui a regressão linear foi usada apenas como técnica descritiva. Se as n observações forem a totalidade da população de interesse, pouco mais há a dizer.

- Mas, com frequência, as n observações são apenas uma amostra aleatória de uma população maior.

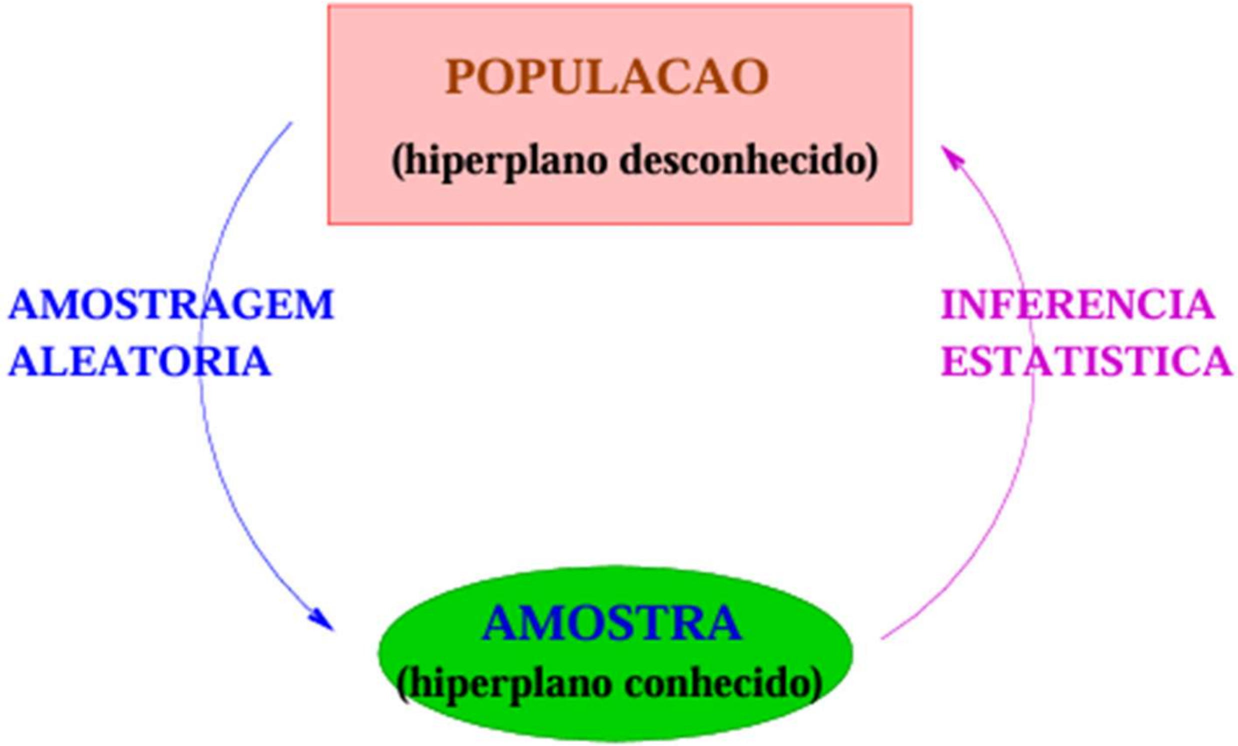
- Um hiperplano ajustado a partir duma dada amostra, $y = b_0 + b_1 x_1 + b_2 x_2 + ... + b_p x_p$, é apenas uma estimativa de um hiperplano populacional

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_p x_p \ .$$

  Outras amostras dariam hiperplanos ajustados diferentes.

- Coloca-se o problema da inferência estatística.

O problema da Inferência Estatística na Reg. Linear



MODELO - Regressão Linear

A fim de se poder fazer inferência sobre o hiperplano populacional, vamos admitir pressupostos adicionais.

Y – variável resposta aleatória.

$x_1, ..., x_p$ – variáveis preditoras não aleatórias (fixadas pelo experimentador ou trabalha-se condicionalmente aos valores de $x_1, ..., x_p$)

O modelo será ajustado com base em:

$\{(x_{1(i)}, x_{2(i)}, \dots, x_{p(i)}, Y_i)\}_{i=1}^{n}$ – n conjuntos de observações independentes das variáveis $x_1, x_2, ..., x_p$ e Y, sobre n unidades experimentais.

MODELO RL – Linearidade

Vamos ainda admitir que a relação de fundo entre Y e x_1, x_2, ..., x_p, é linear (afim), com uma variabilidade aleatória em torno dessa relação, representada por um erro aleatório ε. Para todo o $i = 1, ..., n$:

$$
\begin{array}{ccccccccccc}
Y_i & = & \beta_0 & + & \beta_1 & x_{1(i)} & + \ldots + & \beta_p & x_{p(i)} & + & \varepsilon_i \\
\downarrow & & \downarrow & & \downarrow & \downarrow & & \downarrow & \downarrow & & \downarrow \\
\text{v.a.} & & \text{cte.} & & \text{cte.} & \text{cte.} & & \text{cte.} & \text{cte.} & & \text{v.a.}
\end{array}
$$

MODELO Regressão Linear – Os erros aleatórios

Vamos ainda admitir que os erros aleatórios ε_i:

- Têm valor esperado (valor médio) nulo:

$$E[\varepsilon_i] \ = \ 0 \ , \qquad \forall \ i = 1, ..., n$$

(não é hipótese restritiva).

- Têm distribuição Normal (é restritiva, mas bastante geral).
- Homogeneidade de variâncias: têm sempre a mesma variância

$$V[\varepsilon_i] \ = \ \sigma^2 \ , \qquad \forall \ i = 1, ..., n$$

(é restritiva, mas conveniente).

- São variáveis aleatórias independentes
(é restritiva, mas conveniente).

O Modelo Linear

O modelo para inferência na regressão linear é assim:

O Modelo Linear

1. $Y_i \;=\; \beta_0 + \beta_1 x_{1(i)} + \beta_2 x_{2(i)} + \cdots + \beta_p x_{p(i)} + \varepsilon_i, \quad \forall\, i = 1, ..., n.$
2. $\varepsilon_i \frown \mathcal{N}(0\,,\,\sigma^2), \quad \forall\, i = 1, ..., n.$
3. $\{\varepsilon_i\}_{i=1}^{n}$ v.a. independentes.

NOTA: Os erros aleatórios são variáveis aleatórias independentes e identicamente distribuídas (i.i.d.).

Dado o modelo, o valor esperado (médio) de Y_i, condicional aos valores $x_1, x_2, ..., x_p$ dos preditores, é:

$$\mu_i \;=\; E[Y_i \,|\, x_1, x_2, ..., x_p] \;=\; \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_p x_p\ .$$

NOTA: β_j ($j \neq 0$) é a variação média em Y, associada a um aumento de uma unidade em x_j, mantendo os restantes preditores constantes.

O estudo do modelo

Um primeiro objectivo da inferência: os $p+1$ parâmetros do modelo, β_j $(j = 0, 1, ..., p)$.

Os parâmetros ajustados $\vec{\mathbf{b}} = (b_0, b_1, b_2, ..., b_p)$, são estimativas desses parâmetros.

Para ser possível construir intervalos de confiança e/ou efectuar testes de hipóteses sobre os valores dos parâmetros populacionais β_j, há-que:

- Definir estimadores $\hat{\beta}_j$ dos parâmetros populacionais;
- conhecer as respectivas distribuições de probabilidades (ao abrigo do Modelo);

A validade da inferência depende da validade dos pressupostos do modelo.

A notação matricial/vectorial

$$\begin{array}{ccccc}
Y_1 & = & \beta_0+\beta_1 x_{1(1)}+\beta_2 x_{2(1)}+\cdots+\beta_p x_{p(1)} & + & \varepsilon_1 \\
Y_2 & = & \beta_0+\beta_1 x_{1(2)}+\beta_2 x_{2(2)}+\cdots+\beta_p x_{p(2)} & + & \varepsilon_2 \\
Y_3 & = & \beta_0+\beta_1 x_{1(3)}+\beta_2 x_{2(3)}+\cdots+\beta_p x_{p(3)} & + & \varepsilon_3 \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
\underbrace{Y_n}_{=\vec{\mathbf{Y}}} & = & \underbrace{\beta_0+\beta_1 x_{1(n)}+\beta_2 x_{2(n)}+\cdots+\beta_p x_{p(n)}}_{=\mathbf{X}\vec{\boldsymbol{\beta}}} & + & \underbrace{\varepsilon_n}_{=\vec{\boldsymbol{\varepsilon}}}
\end{array}$$

As n equações correspondem a uma única equação vectorial:

$$\vec{\mathbf{Y}} = \mathbf{X}\vec{\boldsymbol{\beta}} + \vec{\boldsymbol{\varepsilon}} \, ,$$

onde:

$$\vec{\mathbf{Y}}=\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_n \end{bmatrix}, \quad \mathbf{X}=\begin{bmatrix} 1 & x_{1_{(1)}} & x_{2_{(1)}} & \cdots & x_{p_{(1)}} \\ 1 & x_{1_{(2)}} & x_{2_{(2)}} & \cdots & x_{p_{(2)}} \\ 1 & x_{1_{(3)}} & x_{2_{(3)}} & \cdots & x_{p_{(3)}} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1_{(n)}} & x_{2_{(n)}} & \cdots & x_{p_{(n)}} \end{bmatrix}, \quad \vec{\boldsymbol{\beta}}=\begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix}, \quad \vec{\boldsymbol{\varepsilon}}=\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

- $\vec{\mathbf{Y}}$ e $\vec{\boldsymbol{\varepsilon}}$ são vectores aleatórios,
- $\mathbf{X}$ é uma matriz não aleatória e $\vec{\boldsymbol{\beta}}$ um vector não-aleatório.

Modelo Regressão Linear - versão vectorial

O Modelo Linear em notação vectorial

1. $\vec{\mathbf{Y}} = \mathbf{X}\vec{\boldsymbol{\beta}} + \vec{\boldsymbol{\varepsilon}}$.

2. $\vec{\boldsymbol{\varepsilon}} \frown \mathcal{N}_n(\vec{\mathbf{0}}, \sigma^2 \mathbf{I}_n)$, com $\vec{\mathbf{0}} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \cdots \\ 0 \end{bmatrix}$; $\sigma^2 \mathbf{I}_n = \begin{bmatrix} \sigma^2 & 0 & 0 & \cdots & 0 \\ 0 & \sigma^2 & 0 & \cdots & 0 \\ 0 & 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \sigma^2 \end{bmatrix}$

- Cada erro aleatório individual ε_i tem distribuição Normal.
- Cada erro aleatório individual tem média zero: $E[\varepsilon_i] = 0$.
- Cada erro aleatório individual tem variância igual: $V[\varepsilon_i] = \sigma^2$.
- Erros aleatórios diferentes são independentes, porque $Cov[\varepsilon_i, \varepsilon_j] = 0$ se $i \neq j$ e, numa Multinormal, isso implica a independência.

A distribuição de $\vec{\mathbf{Y}}$

Teorema (Primeiras Consequências do Modelo)

Dado o Modelo de Regressão Linear, tem-se:

$$\vec{\mathbf{Y}} \frown \mathcal{N}_n(\mathbf{X}\vec{\boldsymbol{\beta}}\,,\,\sigma^2\,\mathbf{I}_n).$$

De facto, $\vec{\mathbf{Y}}$ é soma de vector não aleatório ($\mathbf{X}\vec{\boldsymbol{\beta}}$) e vector aleatório ($\vec{\boldsymbol{\varepsilon}}$):

$$\vec{\mathbf{Y}} = \underbrace{\mathbf{X}\vec{\boldsymbol{\beta}}}_{=''\vec{\mathbf{a}}''} + \underbrace{\vec{\boldsymbol{\varepsilon}}}_{=''\vec{\mathbf{Z}}''} .$$

- $\vec{\boldsymbol{\varepsilon}} \frown \mathcal{N}(\vec{\mathbf{0}}, \sigma^2\mathbf{I}_n)$.
- Somar vector constante ($\mathbf{X}\vec{\boldsymbol{\beta}}$) a um vector aleatório Multinormal ($\vec{\boldsymbol{\varepsilon}}$) não destrói a Multinormalidade.
- $E[\vec{\mathbf{Y}}] = E[\mathbf{X}\vec{\boldsymbol{\beta}} + \vec{\boldsymbol{\varepsilon}}] = \mathbf{X}\vec{\boldsymbol{\beta}} + E[\vec{\boldsymbol{\varepsilon}}] = \mathbf{X}\vec{\boldsymbol{\beta}}$.
- $V[\vec{\mathbf{Y}}] = V[\mathbf{X}\vec{\boldsymbol{\beta}} + \vec{\boldsymbol{\varepsilon}}] = V[\vec{\boldsymbol{\varepsilon}}] = \sigma^2\mathbf{I}_n$.

A distribuição de $\vec{\mathbf{Y}}$ (interpretação)

$$\vec{\mathbf{Y}} \frown \mathcal{N}_n(\mathbf{X}\vec{\boldsymbol{\beta}}\,,\,\sigma^2\,\mathbf{I}_n).$$

Tendo em conta as propriedades da Multinormal:

- Cada observação individual Y_i tem distribuição Normal.
- Cada observação individual Y_i tem média $\mu_i = E[Y_i] = \vec{\mathbf{x}}_{[i,]}^t\vec{\boldsymbol{\beta}} = \beta_0 + \beta_1 x_{1(i)} + \beta_2 x_{2(i)} + \ldots + \beta_p x_{p(i)}$.
- Cada observação individual tem variância igual: $V[Y_i] = \sigma^2$.
- Observações diferentes de Y são independentes, porque $Cov[Y_i, Y_j] = 0$ se $i \neq j$ e, numa Multinormal, isso implica a independência.

O vector de estimadores $\hat{\vec{\boldsymbol{\beta}}}$

O vector de estimadores $\hat{\vec{\boldsymbol{\beta}}} = (\hat{\beta}_0, \hat{\beta}_1, ..., \hat{\beta}_p)^t$ é definido a partir da equação do vector $\vec{\mathbf{b}}$ de estimativas mas substituindo o vector $\vec{\mathbf{y}}$ de valores observados de Y pelo vector aleatório $\vec{\mathbf{Y}}$.

Estimadores de Mínimos Quadrados dos parâmetros

$$\hat{\vec{\boldsymbol{\beta}}} = (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\vec{\mathbf{Y}} .$$

O vector $\hat{\vec{\boldsymbol{\beta}}}$ é de dimensão $p+1$. O seu primeiro elemento é o estimador de β_0, o seu segundo elemento é o estimador de β_1, etc...

Em geral, o estimador de β_j está na posição $j+1$ do vector $\hat{\vec{\boldsymbol{\beta}}}$.

Os resultados gerais já referidos permitem facilmente determinar a distribuição de probabilidades do estimador $\hat{\vec{\boldsymbol{\beta}}}$.

A distribuição do vector de estimadores $\hat{\vec{\beta}}$

Teorema (Distribuição do estimador $\hat{\vec{\beta}}$)

Dado o Modelo de Regressão Linear Múltipla, tem-se:

$$\hat{\vec{\beta}} \frown \mathcal{N}_{p+1}(\vec{\beta}\,,\sigma^2\,(\mathbf{X}^t\mathbf{X})^{-1})\;.$$

$\hat{\vec{\beta}}$ é produto de matriz não aleatória, $(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t$, e vector aleatório, $\vec{\mathbf{Y}}$:

$$\hat{\vec{\beta}} = \underbrace{(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t}_{\text{"}\mathbf{B}\text{"}}\underbrace{\vec{\mathbf{Y}}}_{\text{"}\vec{\mathbf{Z}}\text{"}}\;.$$

- $\vec{\mathbf{Y}} \frown \mathcal{N}_n(\mathbf{X}\vec{\beta},\sigma^2\mathbf{I}_n)$.
- Multiplicar matriz constante, $(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t$, por um vector aleatório Multinormal ($\vec{\mathbf{Y}}$) não destrói a Multinormalidade.
- $E[\hat{\vec{\beta}}] = E[(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\vec{\mathbf{Y}}] = (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^tE[\vec{\mathbf{Y}}] = (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{X}\vec{\beta} = \vec{\beta}$.
- $V[\hat{\vec{\beta}}] = V[(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\vec{\mathbf{Y}}] = (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^tV[\vec{\mathbf{Y}}][(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t]^t =$ $(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\cdot\sigma^2\mathbf{I}_n\cdot\mathbf{X}[(\mathbf{X}^t\mathbf{X})^{-1}]^t = \sigma^2\cdot(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{X}[(\mathbf{X}^t\mathbf{X})^t]^{-1} = \sigma^2(\mathbf{X}^t\mathbf{X})^{-1}$.

A distribuição de $\vec{\hat{\boldsymbol{\beta}}}$ (interpretação)

$$\vec{\hat{\boldsymbol{\beta}}} \frown \mathcal{N}_{p+1}(\vec{\boldsymbol{\beta}}\,,\, \sigma^2\,(\mathbf{X}^t\mathbf{X})^{-1})\ .$$

Tendo em conta as propriedades da Multinormal

- Cada estimador individual $\hat{\beta}_j$ tem distribuição Normal.
- Cada estimador individual tem média $E[\hat{\beta}_j] = \beta_j$, logo é centrado
- Cada estimador individual tem variância $V[\hat{\beta}_j] = \sigma^2\,(\mathbf{X}^t\mathbf{X})^{-1}_{(j+1,j+1)}$. (Note-se o desfasamento nos índices).
- Estimadores individuais diferentes não são (em geral) independentes, porque $(\mathbf{X}^t\mathbf{X})^{-1}$ não é, em geral, uma matriz diagonal: $Cov[\hat{\beta}_i, \hat{\beta}_j] = \sigma^2\,(\mathbf{X}^t\mathbf{X})^{-1}_{(i+1,j+1)}$.
- Logo, o estimador $\hat{\beta}_j$ de um parâmetro individual β_j tem distribuição $\hat{\beta}_j \frown \mathcal{N}\,(\beta_j\ ,\ \sigma^2_{\hat{\beta}_j})$, com $\sigma^2_{\hat{\beta}_j} = \sigma^2\,(\mathbf{X}^t\mathbf{X})^{-1}_{(j+1,j+1)}$.

Se $p = 1$, RLS

Estimação dos parâmetros do Modelo RLS

A recta do modelo RLS tem dois parâmetros: β_0 e β_1.

Definem-se estimadores desses parâmetros a partir das expressões amostrais obtidas para b_0 e b_1 pelo Método dos Mínimos Quadrados.

Recordar: $b_1 = \frac{cov_{xy}}{s_x^2} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{(n-1)\, s_x^2} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})\, y_i}{(n-1)\, s_x^2} = \sum_{i=1}^{n} \frac{x_i - \overline{x}}{(n-1)\, s_x^2}\, y_i$

Estimador de β_1

$$\hat{\beta}_1 = \sum_{i=1}^{n} \frac{x_i - \overline{x}}{(n-1)\, s_x^2}\, Y_i = \sum_{i=1}^{n} c_i Y_i\,, \qquad \text{com } c_i = \frac{x_i - \overline{x}}{(n-1)\, s_x^2}$$

Nota: O estimador $\hat{\beta}_1$ é combinação linear de Normais independentes, logo tem distribuição Normal.

Se $p = 1$, RLS

Estimação dos parâmetros do Modelo RLS (cont.)

Recordar: $b_0 = \overline{y} - b_1 \overline{x}$.

Estimador de β_0

$$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{x} = \frac{1}{n}\sum_{i=1}^{n} Y_i - \overline{x}\sum_{i=1}^{n} c_i Y_i = \sum_{i=1}^{n}\left(\frac{1}{n} - \overline{x} c_i\right) Y_i = \sum_{i=1}^{n} d_i Y_i \,,$$

com

$$d_i = \frac{1}{n} - \overline{x} c_i = \frac{1}{n} - \frac{(x_i - \overline{x})\overline{x}}{(n-1)\, s_x^2}.$$

Quer $\hat{\beta}_1$, quer $\hat{\beta}_0$, são combinações lineares das observações $\{Y_i\}_{i=1}^{n}$, logo são combinações lineares de variáveis aleatórias Normais independentes. Logo, ambos os estimadores têm distribuição Normal.

Se $p = 1$, RLS

Distribuição dos estimadores RLS

Distribuição dos estimadores dos parâmetros

Dado o Modelo de Regressão Linear Simples,

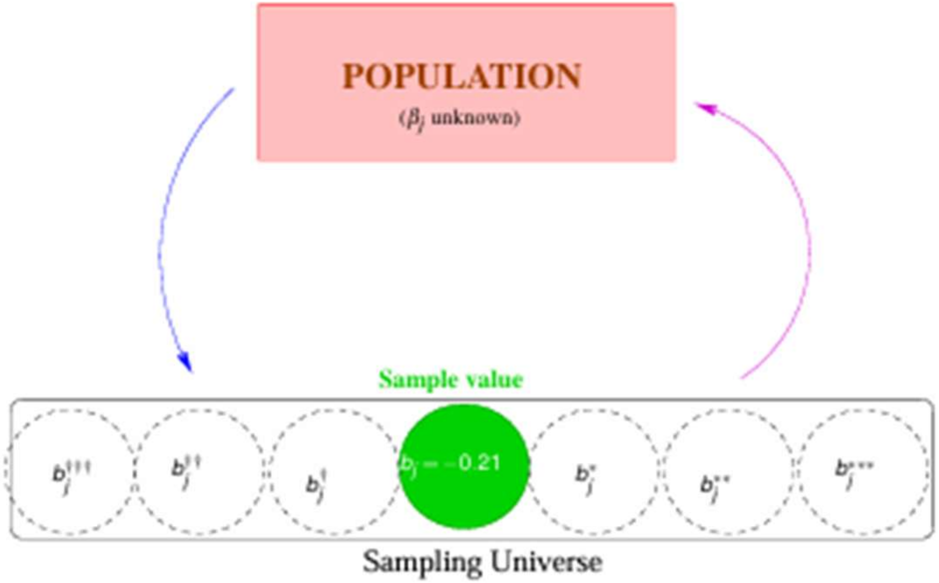
1. $\hat{\beta}_1 \frown \mathcal{N}\left(\beta_1\, ,\, \frac{\sigma^2}{(n-1)\, s_x^2}\right)$,
2. $\hat{\beta}_0 \frown \mathcal{N}\left(\beta_0\, ,\, \sigma^2 \left[\frac{1}{n} + \frac{\overline{x}^2}{(n-1)\, s_x^2}\right]\right)$

NOTAS:

1. Ambos os estimadores são centrados: $E[\hat{\beta}_1] = \beta_1$ e $E[\hat{\beta}_0] = \beta_0$.
2. Quanto maior $(n-1)\, s_x^2$, menor a variância dos estimadores.
3. A variância de $\hat{\beta}_0$ também diminui com o aumento de n, e com a maior proximidade de $\overline{x}$ de zero.

A distribuição na amostragem de $\hat{\beta}_j$ (interpretação)

$$\hat{\beta}_j \frown \mathcal{N}(\beta_j \,,\, \sigma^2_{\hat{\beta}_j}) \qquad \text{com} \qquad \sigma^2_{\hat{\beta}_j} = \sigma^2 \, (\mathbf{X}^t\mathbf{X})^{-1}_{(j+1,j+1)} \,.$$

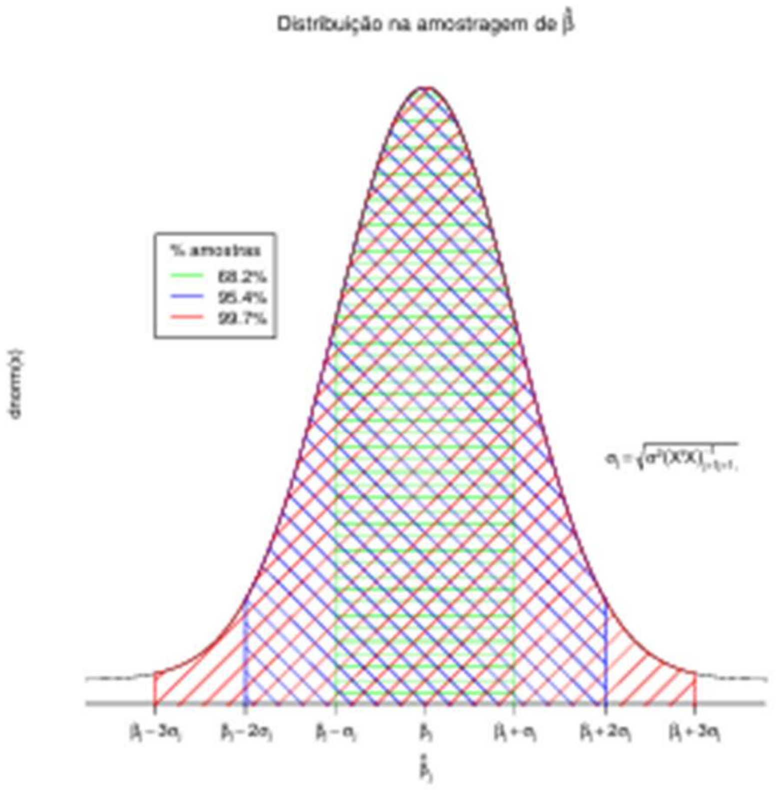


O conjunto de todas as possíveis amostras de dimensão n designa-se o Universo de Amostragem

A distribuição de probabilidades de $\hat{\beta}_j$ pode ser vista como a distribuição dos valores de b_j ao longo do Universo de Amostragem.

A distribuição na amostragem de $\hat{\beta}_j$ (interpretação)

$$\hat{\beta}_j \frown \mathcal{N}(\beta_j\ ,\ \sigma^2_{\hat{\beta}_j}) \qquad \text{com} \qquad \sigma^2_{\hat{\beta}_j} = \sigma^2\,(\mathbf{X}^t\mathbf{X})^{-1}_{(j+1,j+1)}\ .$$



A distribuição dum estimador individual

Como se viu, tem-se, $\forall j = 0, 1, ..., p$:

$$
\begin{aligned}
\hat{\beta}_j &\frown \mathcal{N}\left(\beta_j \,,\, \sigma^2 (\mathbf{X}^t\mathbf{X})^{-1}_{(j+1,j+1)}\right) \\
\Leftrightarrow \qquad \frac{\hat{\beta}_j - \beta_j}{\sigma_{\hat{\beta}_j}} &\frown \mathcal{N}(0,1) \,,
\end{aligned}
$$

com $\sigma_{\hat{\beta}_j} = \sqrt{\sigma^2 (\mathbf{X}^t\mathbf{X})^{-1}_{(j+1,j+1)}}$.

Este resultado distribucional permitiria construir intervalos de confiança ou fazer testes a hipóteses sobre os parâmetros $\vec{\boldsymbol{\beta}}$, não fosse o desconhecimento da variância σ^2 dos erros aleatórios.

Se $p = 1$, RLS

Distribuição dos estimadores RLS

Distribuição dos estimadores (cont.)

Dado o Modelo de Regressão Linear Simples,

1. $\frac{\hat{\beta}_1 - \beta_1}{\sigma_{\hat{\beta}_1}} \frown \mathcal{N}(0,1)$, com $\sigma_{\hat{\beta}_1} = \sqrt{\frac{\sigma^2}{(n-1)\,s_x^2}} = \frac{\sigma}{\sqrt{(n-1)\,s_x^2}}$
2. $\frac{\hat{\beta}_0 - \beta_0}{\sigma_{\hat{\beta}_0}} \frown \mathcal{N}(0,1)$, com $\sigma_{\hat{\beta}_0} = \sqrt{\sigma^2 \left[\frac{1}{n} + \frac{\overline{x}^2}{(n-1)\,s_x^2}\right]} = \sigma\sqrt{\frac{1}{n} + \frac{\overline{x}^2}{(n-1)\,s_x^2}}$

NOTAS:

- O desvio padrão dum estimador designa-se erro padrão (em inglês, *standard error*).
- Não confundir os erros padrão dos estimadores, $\sigma_{\hat{\beta}_1}$ e $\sigma_{\hat{\beta}_0}$, com o desvio padrão σ dos erros aleatórios.

O problema de σ^2 desconhecido

Para poder utilizar um estimador $\hat{\beta}_j$ na inferência, é preciso conhecer a sua distribuição de probabilidades, sem a presença de quantidades não-amostrais desconhecidas, além de β_j.

Para ultrapassar este problema é preciso:

- obter um estimador para σ^2; e
- ver o que acontece à distribuição de $\hat{\beta}_j$ quando σ^2 é substituído pelo seu estimador.

Como $\sigma^2 = V(\varepsilon_i)$, $\forall i$, e como os erros aleatórios ε_i são desconhecidos, é natural procurar um estimador de σ^2 através dos resíduos.

Estimando σ^2

Erros aleatórios (variáveis aleatórias – não observáveis)

$$\varepsilon_i = Y_i - (\beta_0 + \beta_1 x_{1(i)} + \beta_2 x_{2(i)} + \ldots + \beta_p x_{p(i)})$$

Resíduos (variáveis aleatórias – observáveis)

$$E_i = Y_i - \underbrace{(\hat{\beta}_0 + \hat{\beta}_1 x_{1(i)} + \hat{\beta}_2 x_{2(i)} + \ldots + \hat{\beta}_p x_{p(i)})}_{=\hat{Y}_i}$$

Resíduos (observados)

$$e_i = y_i - (b_0 + b_1 x_{1(i)} + b_2 x_{2(i)} + \ldots + b_p x_{p(i)})$$

Quadrado Médio Residual (QMRE)

Define-se o Quadrado Médio Residual como

$$QMRE = \frac{SQRE}{n-(p+1)} = \frac{\sum_{i=1}^{n} E_i^2}{n-(p+1)}$$

Dado o Modelo Linear, $\hat{\sigma}^2 = QMRE$ é um estimador centrado da variância comum dos erros aleatórios, $\sigma^2 = V[\varepsilon_i]$:

$$E[QMRE] = \sigma^2 .$$

O Quadrado Médio Residual tem como unidades de medida o quadrado das unidades de Y.

Quantidades fulcrais para a inferência sobre β_j

A estimação dos erros padrão com o QMRE transforma as distribuições normais em distribuições *t-Student*

Teorema (Distribuições para a inferência sobre β_j)

Dado o Modelo de Regressão Linear Múltipla, tem-se

$$\frac{\hat{\beta}_j - \beta_j}{\hat{\sigma}_{\hat{\beta}_j}} \frown t_{n-(p+1)} \,, \qquad \forall j = 0, 1, \ldots, p$$

com $\hat{\sigma}_{\hat{\beta}_j} = \sqrt{QMRE \cdot (\mathbf{X}^t\mathbf{X})^{-1}_{(j+1,j+1)}}$.

Este Teorema dá-nos os resultados que servem de base à construção de intervalos de confiança e testes de hipóteses para os parâmetros β_j do modelo populacional.

Se $p = 1$, RLS

Quantidades centrais para a inferência sobre β_0 e β_1

A estimação dos erros padrão com o QMRE transforma as distribuições normais em distribuições *t-Student*

Distribuições *t-Student* para a inferência sobre β_0 e β_1

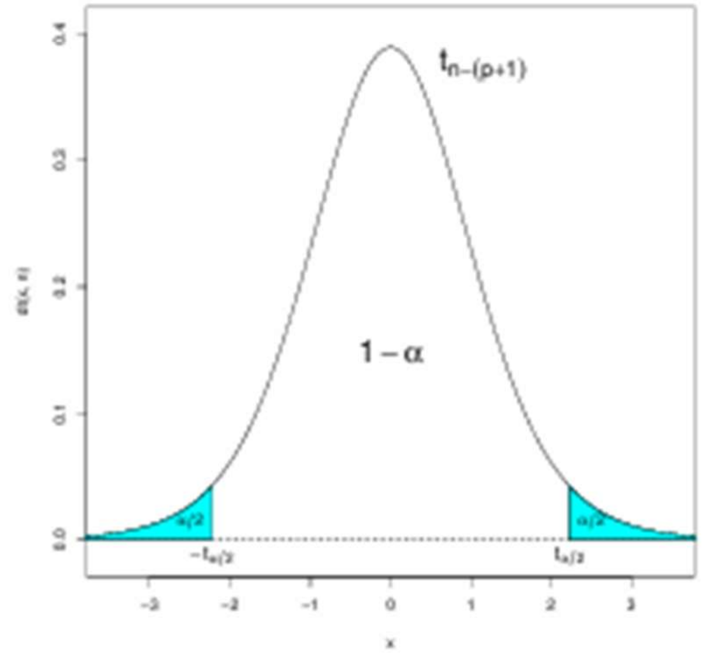
Dado o Modelo de Regressão Linear Simples, prova-se que:

1. $\frac{\hat{\beta}_1 - \beta_1}{\hat{\sigma}_{\hat{\beta}_1}} \frown t_{n-2}$, com $\hat{\sigma}_{\hat{\beta}_1} = \sqrt{\frac{QMRE}{(n-1)\, s_x^2}}$
2. $\frac{\hat{\beta}_0 - \beta_0}{\hat{\sigma}_{\hat{\beta}_0}} \frown t_{n-2}$, com $\hat{\sigma}_{\hat{\beta}_0} = \sqrt{QMRE \left[\frac{1}{n} + \frac{\overline{x}^2}{(n-1)\, s_x^2}\right]}$

Este Teorema é crucial, pois dá-nos os resultados que servirão de base à construção de intervalos de confiança e testes de hipóteses para os parâmetros da recta populacional, β_0 e β_1.

Dedução de intervalo de confiança para β_j

Sabemos que $\frac{\hat{\beta}_j - \beta_j}{\hat{\sigma}_{\hat{\beta}_j}} \frown t_{n-(p+1)}$. Logo,



$$P\left[-t_{\frac{\alpha}{2}} < \frac{\hat{\beta}_j - \beta_j}{\hat{\sigma}_{\hat{\beta}_j}} < t_{\frac{\alpha}{2}}\right] = 1 - \alpha$$

Dedução IC para β_j (cont.)

Trabalhar a dupla desigualdade até isolar β_j:

$$P\left[-t_{\frac{\alpha}{2}} \;<\; \frac{\hat{\beta}_j-\beta_j}{\hat{\sigma}_{\hat{\beta}_j}} \;<\; t_{\frac{\alpha}{2}}\right] \;=\; 1-\alpha$$

$$\begin{array}{rrcl}
 & -t_{\frac{\alpha}{2}} \cdot \hat{\sigma}_{\hat{\beta}_j} < & \hat{\beta}_j - \beta_j & < t_{\frac{\alpha}{2}} \cdot \hat{\sigma}_{\hat{\beta}_j} \\
\Leftrightarrow & t_{\frac{\alpha}{2}} \cdot \hat{\sigma}_{\hat{\beta}_j} > & \beta_j - \hat{\beta}_j & > -t_{\frac{\alpha}{2}} \cdot \hat{\sigma}_{\hat{\beta}_j} \\
\Leftrightarrow & \hat{\beta}_j - t_{\frac{\alpha}{2}} \cdot \hat{\sigma}_{\hat{\beta}_j} < & \beta_j & < \hat{\beta}_j + t_{\frac{\alpha}{2}} \cdot \hat{\sigma}_{\hat{\beta}_j} \,.
\end{array}$$

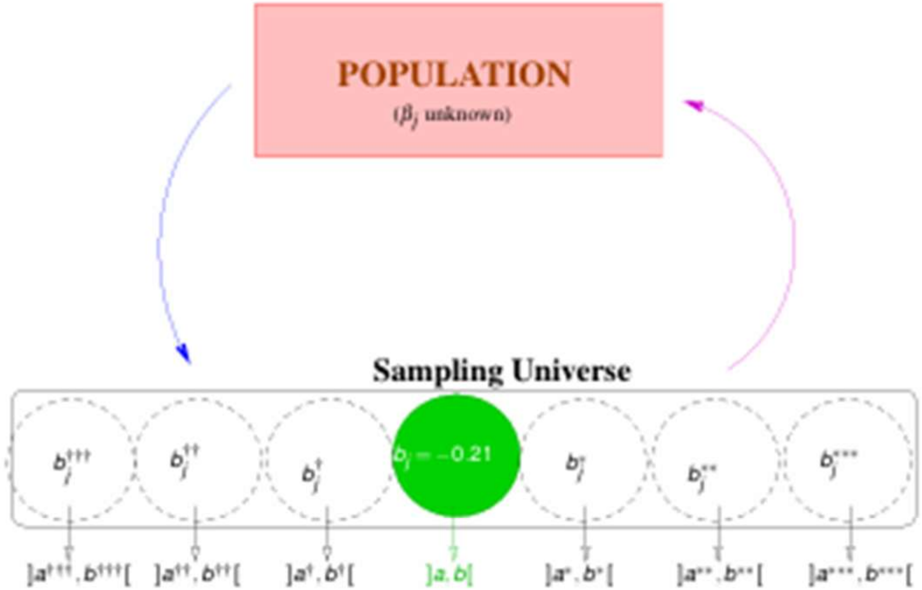
O intervalo aleatório

$$\left]\; \hat{\beta}_j - t_{\frac{\alpha}{2}} \cdot \hat{\sigma}_{\hat{\beta}_j} \;,\; \hat{\beta}_j + t_{\frac{\alpha}{2}} \cdot \hat{\sigma}_{\hat{\beta}_j} \;\right[$$

contém β_j com probabilidade $1-\alpha$.

Intervalo aleatório para β_j (interpretação)

$$\left] \ \hat{\beta}_j - t_{\frac{\alpha}{2}} \cdot \hat{\sigma}_{\hat{\beta}_j} \ , \ \hat{\beta}_j + t_{\frac{\alpha}{2}} \cdot \hat{\sigma}_{\hat{\beta}_j} \ \right[$$



Cada amostra no Universo de Amostragem gera um intervalo concreto, chamado Intervalo de Confiança

Uma proporção $1-\alpha$ desses intervalos contêm o verdadeiro valor de β_j. Os restantes α não contêm β_j.

Intervalo de confiança para β_j

Intervalo de Confiança a $(1-\alpha) \times 100\%$ para β_j

Dado o Modelo de Regressão Linear Múltipla e uma amostra, eis o intervalo a $(1-\alpha) \times 100\%$ de confiança para o parâmetro β_j:

$$\left] \quad b_j - t_{\frac{\alpha}{2}[n-(p+1)]} \cdot \hat{\sigma}_{\hat{\beta}_j} \quad , \quad b_j + t_{\frac{\alpha}{2}[n-(p+1)]} \cdot \hat{\sigma}_{\hat{\beta}_j} \quad \right[,$$

sendo:

- b_j o elemento $j+1$ do vector das estimativas $\vec{\mathbf{b}}$
- $t_{\frac{\alpha}{2}[n-(p+1)]}$ o quantil de ordem $1-\frac{\alpha}{2}$ da distribuição $t_{n-(p+1)}$;
- $\hat{\sigma}_{\hat{\beta}_j} = \sqrt{QMRE \cdot (\mathbf{X}^t\mathbf{X})^{-1}_{(j+1,j+1)}}$ (com o valor de QMRE na nossa amostra).

NOTA: A amplitude do IC aumenta com *QMRE* e o valor diagonal da matriz $(\mathbf{X}^t\mathbf{X})^{-1}$ correspondente ao parâmetro β_j.

Se $p = 1$, RLS

Intervalo de confiança para β_1

Intervalo de Confiança a $(1-\alpha)\times 100\%$ para β_1

Dado o Modelo RLS, um intervalo a $(1-\alpha)\times 100\%$ de confiança para o declive β_1 da recta de regressão populacional é dado por:

$$\left] \, b_1 - t_{\frac{\alpha}{2}[n-2]}\, \hat{\sigma}_{\hat{\beta}_1} \quad , \quad b_1 + t_{\frac{\alpha}{2}[n-2]}\, \hat{\sigma}_{\hat{\beta}_1} \, \right[\, ,$$

tendo $t_{\frac{\alpha}{2}[n-2]}$, b_1 e $\hat{\sigma}_{\hat{\beta}_1}$ sido definidos em acetatos anteriores.

NOTAS:

- O intervalo é centrado em b_1.
- A amplitude do intervalo é $2 \times t_{\frac{\alpha}{2}[n-2]}\, \hat{\sigma}_{\hat{\beta}_1}$.
- A amplitude aumenta com $QMRE$ e diminui com n e s_x^2: $\hat{\sigma}_{\hat{\beta}_1} = \sqrt{\frac{QMRE}{(n-1)\, s_x^2}}$
- A amplitude do IC aumenta para maiores graus de confiança $1-\alpha$.

Se $p = 1$, RLS

Intervalo de confiança para β_0

Intervalo de Confiança a $(1-\alpha) \times 100\%$ para β_0

Dado o Modelo de Regressão Linear Simples, um intervalo a $(1-\alpha) \times 100\%$ de confiança para a ordenada na origem, β_0, da recta populacional é:

$$\left] \, b_0 - t_{\frac{\alpha}{2}[n-2]} \cdot \hat{\sigma}_{\hat{\beta}_0} \quad , \quad b_0 + t_{\frac{\alpha}{2}[n-2]} \cdot \hat{\sigma}_{\hat{\beta}_0} \, \right[,$$

onde $t_{\frac{\alpha}{2}[n-2]}$, b_0 e $\hat{\sigma}_{\hat{\beta}_0}$ foram definidos em acetatos anteriores.

NOTAS:

- O intervalo é centrado em b_0.
- A amplitude do intervalo é $2 \times t_{\frac{\alpha}{2}[n-2]} \, \hat{\sigma}_{\hat{\beta}_0}$.
- A amplitude aumenta com $QMRE$ e com $\overline{x}^2$ e diminui com n e s_x^2:

$$\hat{\sigma}_{\hat{\beta}_0} = \sqrt{QMRE \cdot \left[\frac{1}{n} + \frac{\overline{x}^2}{(n-1)\, s_x^2}\right]}$$

- A amplitude do IC aumenta para maiores graus de confiança $1-\alpha$.

Ainda o exemplo dos lírios

RLM

```
proc reg data=iris;
    model PetalWidth = SepalLength SepalWidth PetalLength/clb;
run;
```

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	95% Confidence Limits	
Intercept	1	-0.24031	0.17837	-1.35	0.1800	-0.59283	0.11221
SepalLength	1	-0.20727	0.04751	-4.36	<.0001	-0.30115	-0.11338
SepalWidth	1	0.22283	0.04894	4.55	<.0001	0.12611	0.31955
PetalLength	1	0.52408	0.02449	21.40	<.0001	0.47568	0.57249

b_j

$$\hat{\sigma}_{\hat{\beta}_j} = \sqrt{QMRE \cdot (\mathbf{X}'\mathbf{X})^{-1}_{(j+1,j+1)}}$$

As estimativas dos desvios padrão associados à estimação de cada um dos parâmetros

Exemplo b_1: na nossa amostra estima-se que, em média, a largura da pétala diminui 0.20727 cm por cada aumento de 1 cm no comprimentos da sépala (mantendo-se as outras medições constantes).

Como $t_{0.025(146)=1.976346}$, o IC a 95% para β_1 é:

$$](-\mathbf{0.20727}) - (\mathbf{1.976346})(\mathbf{0.04751})\ ,(-\mathbf{0.20727}) + (\mathbf{1.976346})(\mathbf{0.04751})[$$

$$\Leftrightarrow\ ]-\mathbf{0.30115}, -\mathbf{0.11338}[$$

Temos 95% de confiança em como o verdadeiro valor de β_1 (na população) está compreendido entre -0.30115 e -0.11338.

Ainda o exemplo dos lírios

RLS

```
proc reg data=iris;
    model PetalWidth = SepalLength/clb;
run;
```

Parameter Estimates							
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	95% Confidence Limits	
Intercept	1	-3.20022	0.25689	-12.46	<.0001	-3.70785	-2.69258
SepalLength	1	0.75292	0.04353	17.30	<.0001	0.66690	0.83894

Um IC a 95% de confiança para o declive da recta populacional é] 0.66690, 0.83894[.

$$\hat{\sigma}_{\hat{\beta}_0} = \sqrt{QMRE\left[\frac{1}{n} + \frac{\overline{x}^2}{(n-1)\,s_x^2}\right]}$$

$$\hat{\sigma}_{\hat{\beta}_1} = \sqrt{\frac{QMRE}{(n-1)\,s_x^2}}$$

$$\hat{\sigma}_{\hat{\beta}_j} = \sqrt{QMRE \cdot (\mathbf{X}^t\mathbf{X})^{-1}_{(j+1,j+1)}}$$

Nota: O coeficiente associado ao preditor *Sepal.Length* na regressão linear simples agora ajustada é positivo, b_1 =0.75292. No modelo de regressão linear múltipla obteve-se um resultado dfiferente, pois contém, além do preditor comprimento da sépala, outros dois preditores (largura da sépala e comprimento da pétala), que contribuem para a formação dos valores ajustados. Na presença desses dois preditores, a contribuição do comprimento da sépala teve um sinal negativo. Esta aparente contradição sublinha uma ideia importante:
a introdução (ou exclusão) de preditores numa regressão linear têm efeitos sobre todos os parâmetros, não sendo possível prever qual será a equação ajustada sem refazer as contas do ajustamento.