Econometria Aplicada

Regressão linear: simples e múltipla

João Ricardo Costa Filho

Econometria Aplicada

Gestão de expectativas

O que vocês esperam deste curso?

O que é Econometria

"The most important questions of life are, for the most part, really only problems in probability."

Laplace (1812)

O que é Econometria

"In God we trust. All others must bring data."
William Edwards Deming

Tipos de dados

- Cross-section.

Tipos de dados

- Cross-section.
- Dados em painel.

Tipos de dados

- Cross-section.
- Dados em painel.
- Série de tempo

A nossa jornada

- Aula 1 - Regressão linear: simples e múltipla

A nossa jornada

- Aula 1 - Regressão linear: simples e múltipla
- Aula 2 - Regressão linear múltipla e formas funcionais

A nossa jornada

- Aula 1 - Regressão linear: simples e múltipla
- Aula 2 - Regressão linear múltipla e formas funcionais
- Aula 3 - Modelos de probabilidade (Probit e Logit)

A nossa jornada

- Aula 1 - Regressão linear: simples e múltipla
- Aula 2 - Regressão linear múltipla e formas funcionais
- Aula 3 - Modelos de probabilidade (Probit e Logit)
- Aula 4 - Variáveis Instrumentais

A nossa jornada

- Aula 1 - Regressão linear: simples e múltipla
- Aula 2 - Regressão linear múltipla e formas funcionais
- Aula 3 - Modelos de probabilidade (Probit e Logit)
- Aula 4 - Variáveis Instrumentais
- Aula 5 - Modelos com dados em painel

A nossa jornada

- Aula 1 - Regressão linear: simples e múltipla
- Aula 2 - Regressão linear múltipla e formas funcionais
- Aula 3 - Modelos de probabilidade (Probit e Logit)
- Aula 4 - Variáveis Instrumentais
- Aula 5 - Modelos com dados em painel
- Aula 6 - Introdução à séries temporais

A nossa jornada

- Aula 1 - Regressão linear: simples e múltipla
- Aula 2 - Regressão linear múltipla e formas funcionais
- Aula 3 - Modelos de probabilidade (Probit e Logit)
- Aula 4 - Variáveis Instrumentais
- Aula 5 - Modelos com dados em painel
- Aula 6 - Introdução à séries temporais
- Aula 7 - Modelos ARIMA

A nossa jornada

- Aula 1 - Regressão linear: simples e múltipla
- Aula 2 - Regressão linear múltipla e formas funcionais
- Aula 3 - Modelos de probabilidade (Probit e Logit)
- Aula 4 - Variáveis Instrumentais
- Aula 5 - Modelos com dados em painel
- Aula 6 - Introdução à séries temporais
- Aula 7 - Modelos ARIMA
- Aula 8 - Discussão sobre os trabalhos

Livro-texto

Wooldridge, J. M. (2006). Introdução à econometria: uma abordagem moderna. Pioneira Thomson Learning.

Livro-texto

Wooldridge, J. M. (2006). Introdução à econometria: uma abordagem moderna. Pioneira Thomson Learning.

- Aula 1 - Regressão linear: simples e múltipla [capítulo 2, 3 e 4]
- Aula 2 - Regressão linear múltipla e formas funcionais [capítulo 2, 3 e 4]

Livro-texto

Wooldridge, J. M. (2006). Introdução à econometria: uma abordagem moderna. Pioneira Thomson Learning.

- Aula 1 - Regressão linear: simples e múltipla [capítulo 2, 3 e 4]
- Aula 2 - Regressão linear múltipla e formas funcionais [capítulo 2, 3 e 4]
- Aula 3 - Modelos de probabilidade (Probit e Logit) [capítulo 17]

Livro-texto

Wooldridge, J. M. (2006). Introdução à econometria: uma abordagem moderna. Pioneira Thomson Learning.

- Aula 1 - Regressão linear: simples e múltipla [capítulo 2, 3 e 4]
- Aula 2 - Regressão linear múltipla e formas funcionais [capítulo 2, 3 e 4]
- Aula 3 - Modelos de probabilidade (Probit e Logit) [capítulo 17]
- Aula 4 - Variáveis Instrumentais [capítulo 15]

Livro-texto

Wooldridge, J. M. (2006). Introdução à econometria: uma abordagem moderna. Pioneira Thomson Learning.

- Aula 1 - Regressão linear: simples e múltipla [capítulo 2, 3 e 4]
- Aula 2 - Regressão linear múltipla e formas funcionais [capítulo 2, 3 e 4]
- Aula 3 - Modelos de probabilidade (Probit e Logit) [capítulo 17]
- Aula 4 - Variáveis Instrumentais [capítulo 15]
- Aula 5 - Modelos com dados em painel [capítulo 13]

Livro-texto

Wooldridge, J. M. (2006). Introdução à econometria: uma abordagem moderna. Pioneira Thomson Learning.

- Aula 1 - Regressão linear: simples e múltipla [capítulo 2, 3 e 4]
- Aula 2 - Regressão linear múltipla e formas funcionais [capítulo 2, 3 e 4]
- Aula 3 - Modelos de probabilidade (Probit e Logit) [capítulo 17]
- Aula 4 - Variáveis Instrumentais [capítulo 15]
- Aula 5 - Modelos com dados em painel [capítulo 13]
- Aula 6 - Introdução à séries temporais [capítulo 18]

Livro-texto

Wooldridge, J. M. (2006). Introdução à econometria: uma abordagem moderna. Pioneira Thomson Learning.

- Aula 1 - Regressão linear: simples e múltipla [capítulo 2, 3 e 4]
- Aula 2 - Regressão linear múltipla e formas funcionais [capítulo 2, 3 e 4]
- Aula 3 - Modelos de probabilidade (Probit e Logit) [capítulo 17]
- Aula 4 - Variáveis Instrumentais [capítulo 15]
- Aula 5 - Modelos com dados em painel [capítulo 13]
- Aula 6 - Introdução à séries temporais [capítulo 18]
- Aula 7 - Modelos ARIMA [capítulo 18]

Metodologia ativa

- **Vocês são os protagonistas do próprio aprendizado**.

Metodologia ativa

- **Vocês são os protagonistas do próprio aprendizado**.
- Há evidências de que os alunos aprendem mais com métodos ativos,

Metodologia ativa

- **Vocês são os protagonistas do próprio aprendizado**.
- Há evidências de que os alunos aprendem mais com métodos ativos, embora muitas vezes prefiram aulas meramente expositivas.

Metodologia ativa

- **Vocês são os protagonistas do próprio aprendizado**.
- Há evidências de que os alunos aprendem mais com métodos ativos, embora muitas vezes prefiram aulas meramente expositivas. Surge o nosso primeiro conflito.

Metodologia ativa

- **Vocês são os protagonistas do próprio aprendizado**.
- Há evidências de que os alunos aprendem mais com métodos ativos, embora muitas vezes prefiram aulas meramente expositivas. Surge o nosso primeiro conflito.
- A importância de (saber) resolver problemas.

Metodologia ativa

- **Vocês são os protagonistas do próprio aprendizado**.
- Há evidências de que os alunos aprendem mais com métodos ativos, embora muitas vezes prefiram aulas meramente expositivas. Surge o nosso primeiro conflito.
- A importância de (saber) resolver problemas.
- **Marquem uma conversa comigo!** (quero saber sobre você, seu interesse no programa e o seu **plano de estudos** para a disciplina).

Avaliação

Avaliação

- Atividades em sala.

Avaliação

- Atividades em sala.
- Trabalho – Aplicação.

Avaliação

- Atividades em sala.
- Trabalho – Aplicação.
- Fórum de discussão.

Trabalho

- O **principal produto da disciplina** é desenvolver a capacidade de desenhar e implementar um projeto aplicado.

Trabalho

- O **principal produto da disciplina** é desenvolver a capacidade de desenhar e implementar um projeto aplicado. Por isso, o trabalho consiste na **apresentação**

Trabalho

- O **principal produto da disciplina** é desenvolver a capacidade de desenhar e implementar um projeto aplicado. Por isso, o trabalho consiste na **apresentação** de (i) uma pergunta,

Trabalho

- O **principal produto da disciplina** é desenvolver a capacidade de desenhar e implementar um projeto aplicado. Por isso, o trabalho consiste na **apresentação** de (i) uma pergunta, (ii) descrição e análise dos dados que serão utilizados para responder essa pergunta,

Trabalho

- O **principal produto da disciplina** é desenvolver a capacidade de desenhar e implementar um projeto aplicado. Por isso, o trabalho consiste na **apresentação** de (i) uma pergunta, (ii) descrição e análise dos dados que serão utilizados para responder essa pergunta, (iii) estimativas econométricas

Trabalho

- O **principal produto da disciplina** é desenvolver a capacidade de desenhar e implementar um projeto aplicado. Por isso, o trabalho consiste na **apresentação** de (i) uma pergunta, (ii) descrição e análise dos dados que serão utilizados para responder essa pergunta, (iii) estimativas econométricas e (iv) conclusões.

Trabalho

- O **principal produto da disciplina** é desenvolver a capacidade de desenhar e implementar um projeto aplicado. Por isso, o trabalho consiste na **apresentação** de (i) uma pergunta, (ii) descrição e análise dos dados que serão utilizados para responder essa pergunta, (iii) estimativas econométricas e (iv) conclusões.
- A última aula da disciplina será destinada à apresentação do trabalho e entrega (apresentação – um arquivo no formato 'pdf' – e códigos + dados, enviados pelo Teams).
- Sugestão: acesse a "Base dos dados" para encontrar dados para a economia brasileira que ajudem a responder a pergunta de pesquisado do trabalho.

Ferramental

Ferramental

- Linguagem: R
- Como?
  - RStudio
  - Google Colab: https://colab.research.google.com/#create=true&language=r

A regressão linear

Motivação (tudo começa com uma pergunta)

Qual é a exposição de ações individuais ao portfólio de mercado?

Visualização dos dados (super importante!)

Vamos coletar os dados e "olhar" para eles.

Por que visualizar os dados é tão importante assim?

O quarteto de Anscombe

O quarteto de Anscombe (Anscombe 1973)

Imagine quatro conjuntos de dados.

O quarteto de Anscombe (Anscombe 1973)

Imagine quatro conjuntos de dados. Em cada um deles, temos duas variáveis (X e Y). Em todos, temos...

O quarteto de Anscombe (Anscombe 1973)

Imagine quatro conjuntos de dados. Em cada um deles, temos duas variáveis (X e Y). Em todos, temos...

- ...a mesma média e o mesmo desvio-padrão de X.

O quarteto de Anscombe (Anscombe 1973)

Imagine quatro conjuntos de dados. Em cada um deles, temos duas variáveis (X e Y). Em todos, temos...

- ... a mesma média e o mesmo desvio-padrão de X.
- ... a mesma média e o mesmo desvio-padrão de Y.

O quarteto de Anscombe (Anscombe 1973)

Imagine quatro conjuntos de dados. Em cada um deles, temos duas variáveis (X e Y). Em todos, temos...

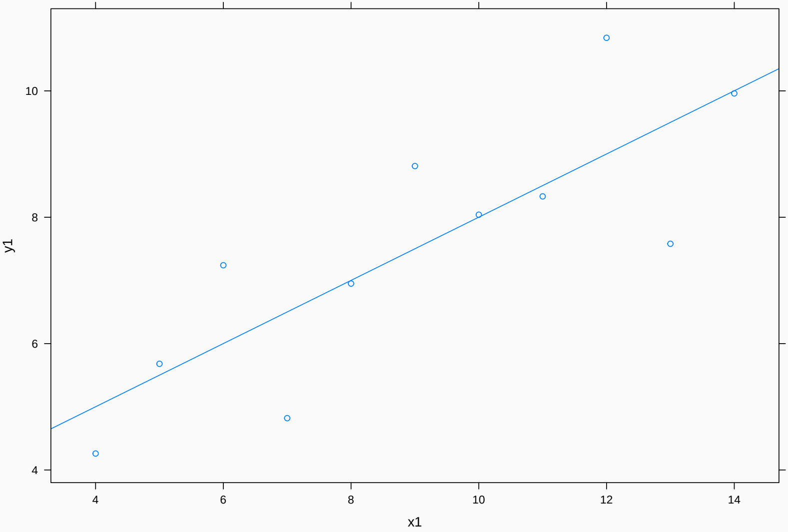
- ...a mesma média e o mesmo desvio-padrão de X.
- ...a mesma média e o mesmo desvio-padrão de Y.
- ...a mesma correlação entre X e Y.

O quarteto de Anscombe (Anscombe 1973)

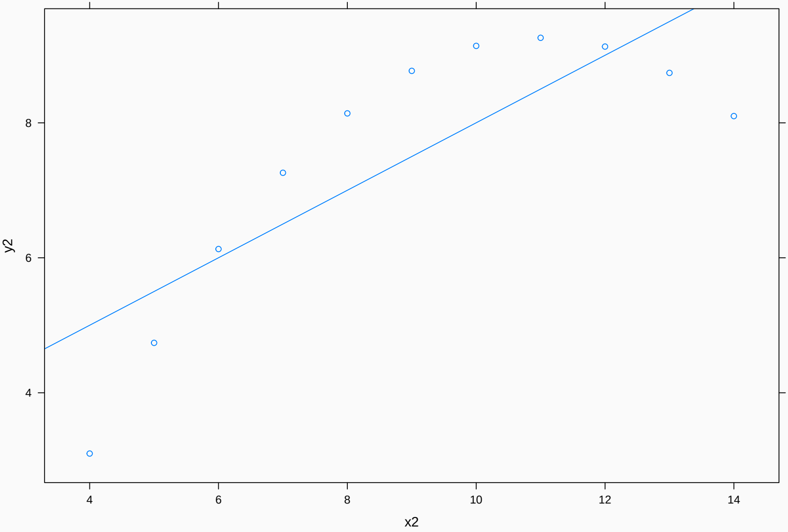
Imagine quatro conjuntos de dados. Em cada um deles, temos duas variáveis (X e Y). Em todos, temos...

- ...a mesma média e o mesmo desvio-padrão de X.
- ...a mesma média e o mesmo desvio-padrão de Y.
- ...a mesma correlação entre X e Y.
- ...os mesmos coeficientes estimados para uma regressão linear de Y em X.

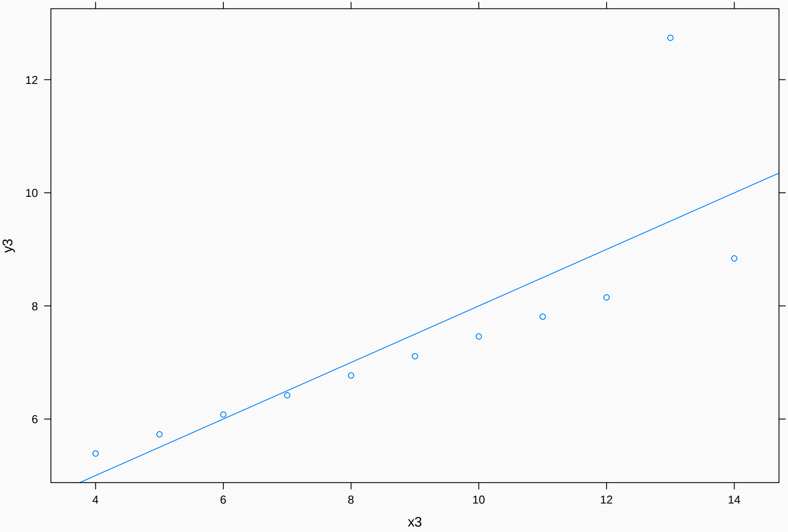
O quarteto de Anscombe - olhem para os dados!



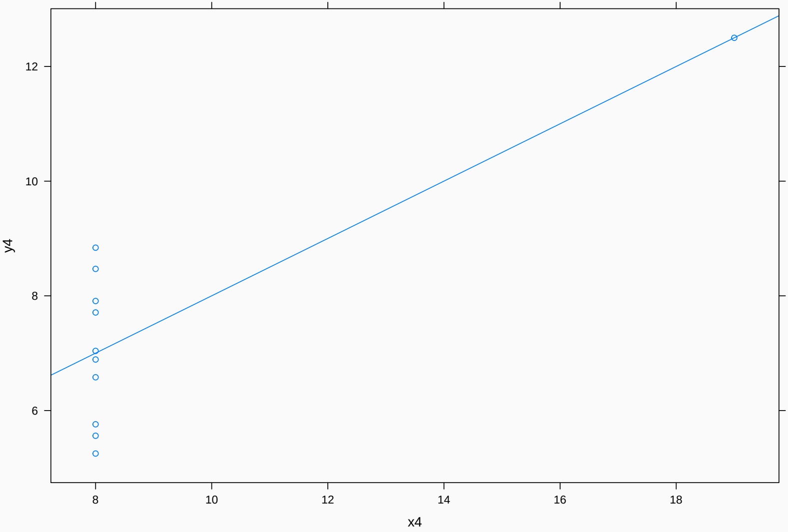
O quarteto de Anscombe - olhem para os dados!



O quarteto de Anscombe - olhem para os dados!



O quarteto de Anscombe - olhem para os dados!



Voltemos à questão dos retornos das ações frente ao retorno do mercado.

Como responder a questão que motivou a nossa análise?

A regressão linear

Assuma que possamos relacionar os retornos diários de uma ação "a" com os retornos diários do Ibovespa da seguinte forma:

$$r_i^{\mathrm{a}} = \beta_0 + \beta_1 r_i^{\mathrm{ibov}} + \varepsilon_i$$

A regressão linear

Assuma que possamos relacionar os retornos diários de uma ação “a” com os retornos diários do Ibovespa da seguinte forma:

$$r_i^{\text{a}} = \beta_0 + \beta_1 r_i^{\text{ibov}} + \varepsilon_i$$

O que os parâmetros significam?

Assuma que possamos relacionar os retornos diários de uma ação “a” com os retornos diários do Ibovespa da seguinte forma:

$$r_i^{\text{a}} = \beta_0 + \beta_1 r_i^{\text{ibov}} + \varepsilon_i$$

O que os parâmetros significam? Como estimá-los?

Assuma que busquemos um estimador que minimize o erro quadrado. Por quê erro quadrado?

Regressão linear com MQO

Assuma que busquemos um estimador que minimize o erro quadrado. Por quê erro quadrado?

- Erro: $\varepsilon_i = r_i^{\text{a}} - \beta_0 - \beta_1 r_i^{\text{ibov}}$.

Regressão linear com MQO

Assuma que busquemos um estimador que minimize o erro quadrado. Por quê erro quadrado?

- Erro: $\varepsilon_i = r_i^{\mathrm{a}} - \beta_0 - \beta_1 r_i^{\mathrm{ibov}}$.
- $\min_{\hat{\beta}_0, \hat{\beta}_1} \sum_{i=1}^{n} \left(\epsilon_i\right)^2 = \sum_{i=1}^{n} \left(r_i^{\mathrm{a}} - \hat{\beta}_0 - \hat{\beta}_1 r_i^{\mathrm{ibov}}\right)^2$

Regressão linear com MQO

Assuma que busquemos um estimador que minimize o erro quadrado. Por quê erro quadrado?

- Erro: $\varepsilon_i = r_i^{\text{a}} - \beta_0 - \beta_1 r_i^{\text{ibov}}$.
- $\min_{\hat{\beta}_0, \hat{\beta}_1} \sum_{i=1}^{n} \left(\epsilon_i\right)^2 = \sum_{i=1}^{n} \left(r_i^{\text{a}} - \hat{\beta}_0 - \hat{\beta}_1 r_i^{\text{ibov}}\right)^2$
    - $\hat{\beta}_1 = \frac{\sum_{i=1}^{n} \left(r_i^{\text{ibov}} - \overline{r^{\text{ibov}}}\right)\left(r_i^{\text{a}} - \overline{r^{\text{a}}}\right)}{\sum_{i=1}^{n} \left(r_i^{\text{ibov}} - \overline{r^{\text{ibov}}}\right)^2}$

Regressão linear com MQO

Assuma que busquemos um estimador que minimize o erro quadrado. Por quê erro quadrado?

- Erro: $\varepsilon_i = r_i^{\text{a}} - \beta_0 - \beta_1 r_i^{\text{ibov}}$.
- $\min_{\hat{\beta}_0, \hat{\beta}_1} \sum_{i=1}^{n} \left(\epsilon_i\right)^2 = \sum_{i=1}^{n} \left(r_i^{\text{a}} - \hat{\beta}_0 - \hat{\beta}_1 r_i^{\text{ibov}}\right)^2$
    - $\hat{\beta}_1 = \frac{\sum_{i=1}^{n} \left(r_i^{\text{ibov}} - \overline{r^{\text{ibov}}}\right)\left(r_i^{\text{a}} - \overline{r^{\text{a}}}\right)}{\sum_{i=1}^{n} \left(r_i^{\text{ibov}} - \overline{r^{\text{ibov}}}\right)^2}$
    - $\hat{\beta}_0 = \overline{r^{\text{a}}} - \hat{\beta}_1 \overline{r^{\text{ibov}}}$

Regressão linear com MQO

Genericamente

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} \left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{\sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2} = \frac{cov(X, Y)}{var(X)} = corr(X, Y)\frac{s_X}{s_y}$$

Regressão linear com MQO

Genericamente

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2} = \frac{cov(X, Y)}{var(X)} = corr(X, Y)\frac{s_X}{s_y}$$

e

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1\bar{X}$$

Vamos estimar as regressões!

Inferência

Definição importante

Vocês aceitam errar quantas vezes para cada 100 tentativas?

Inferência

Como verificar se a associação entre as variáveis é estatísticamente significativa?

Inferência

Como verificar se a associação entre as variáveis é estatísticamente significativa?Realizados testes de hipótese sobre os parâmetros!

Inferência

Como verificar se a associação entre as variáveis é estatísticamente significativa?Realizados testes de hipótese sobre os parâmetros!

- Para $\hat{\beta}_0$:

Inferência

Como verificar se a associação entre as variáveis é estatísticamente significativa?Realizados testes de hipótese sobre os parâmetros!

- Para $\hat{\beta}_0$:

$$\mathcal{H}_0 : \beta_0 = 0$$
$$\mathcal{H}_a : \beta_0 \neq 0$$

Inferência

Como verificar se a associação entre as variáveis é estatísticamente significativa?Realizados testes de hipótese sobre os parâmetros!

- Para $\hat{\beta}_0$:

$$\mathcal{H}_0 : \beta_0 = 0$$
$$\mathcal{H}_a : \beta_0 \neq 0$$

- Para $\hat{\beta}_1$:

$$\mathcal{H}_0 : \beta_1 = 0$$
$$\mathcal{H}_a : \beta_1 \neq 0$$

(Não precisam ser apenas com $\neq$ e nem com zero!)

Vamos simular o comportamento de β_0 e β_1 em diferentes amostras?

Vamos simular para entender o que significa um teste de hipótese

Imagine que tenhamos 500 amostras de tamanho 200 com duas variáveis, X e Y.

Vamos simular para entender o que significa um teste de hipótese

Imagine que tenhamos 500 amostras de tamanho 200 com duas variáveis, X e Y. E que saibamos que $Y_i = 2 + 3X_i + \epsilon_i$.

Vamos simular para entender o que significa um teste de hipótese

Imagine que tenhamos 500 amostras de tamanho 200 com duas variáveis, X e Y. E que saibamos que $Y_i = 2 + 3X_i + \epsilon_i$. Ou seja, que $\beta_0 = 2$ e $\beta_1 = 3$.

Vamos simular para entender o que significa um teste de hipótese

Imagine que tenhamos 500 amostras de tamanho 200 com duas variáveis, X e Y. E que saibamos que $Y_i = 2 + 3X_i + \epsilon_i$. Ou seja, que $\beta_0 = 2$ e $\beta_1 = 3$. Quais seriam os resultados dos estimadores ($\hat{\beta}_0$ e $\hat{\beta}_1$) em cada uma delas?

Vamos simular para entender o que significa um teste de hipótese

Imagine que tenhamos 500 amostras de tamanho 200 com duas variáveis, X e Y. E que saibamos que $Y_i = 2 + 3X_i + \epsilon_i$. Ou seja, que $\beta_0 = 2$ e $\beta_1 = 3$. Quais seriam os resultados dos estimadores ($\hat{\beta}_0$ e $\hat{\beta}_1$) em cada uma delas? Podemos identificar algum padrão?

Vamos simular para entender o que significa um teste de hipótese

```r
# Para replicarmos as variáveis pseudo aleatórias
set.seed(1301)

# Definindo os parâmetros

amostras <- 500 # número de amostras

n <- 200        # tamanho de cada amostra

b0 <- 2

b1 <- 3
```

Vamos simular para entender o que significa um teste de hipótese

```
# Criando as amostras

X <- replicate( amostras, rnorm( n, mean = 10, sd = 2 ) )

e <- replicate( amostras, rnorm( n, mean = 0, sd = 1 ) )

Y <- b0 + b1 * X + e
```

Vamos simular para entender o que significa um teste de hipótese

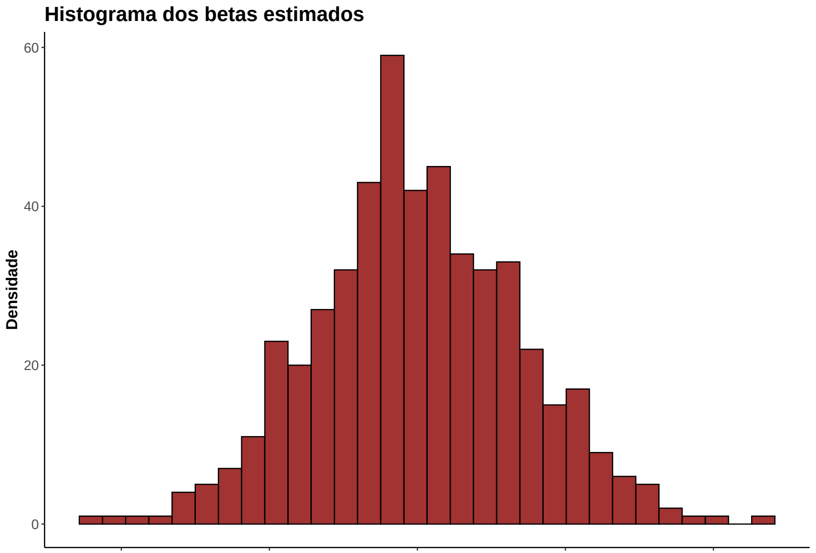
```r
# Fazendo as regressões

regressoes <- lapply( 1:amostras,
                      function(i) lm( Y[ , i ] ~ X[ , i ] )

betas <- sapply(regressoes,
                function(modelo) coef(modelo)[2])

beta1 = mean( betas )
```

Vamos simular para entender o que significa um teste de hipótese



Distribuição amostral

Ou seja, tanto $\hat{\beta}_0$ quanto $\hat{\beta}_1$ são **estatísticas** (i.e. funções dos valores amostrais) e cada estatística possui uma **distribuição**. Em função disso, podemos (i) definir um nível de significância e (ii) fazer um teste de hipótese sobre o parâmetro de interesse.

Teste t

Teste t

- Para $\hat{\beta}_1$:

Teste t

- Para $\hat{\beta}_1$:

$$\mathcal{H}_0 : \beta_1 = \mu$$
$$\mathcal{H}_a : \beta_1 \neq \mu$$

Teste t

- Para $\hat{\beta}_1$:

$$\mathcal{H}_0 : \beta_1 = \mu$$
$$\mathcal{H}_a : \beta_1 \neq \mu$$

A estatística do teste é dada por:

$$t_{\hat{\beta}_1} = \frac{\hat{\beta}_1 - \mu}{se(\hat{\beta}_1)}$$

porque $t_{\hat{\beta}_1} \sim T_{n-k-1}$.

Como fica o teste para as regressões que estimamos?

Quanto os modelos explicam a variação dos retornos diários das empresas?

Goodness of fit

Goodness of fit

Quanto eu consigo explicar sobre a variação dos retorno diários das ações com base nas variações do portfólio de mercado?

Goodness of fit

Quanto eu consigo explicar sobre a variação dos retorno diários das ações com base nas variações do portfólio de mercado?

- Do total da soma (dos quadrados) dos residuos, $\sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2 = (n-1)s_Y^2 \ldots$

Goodness of fit

Quanto eu consigo explicar sobre a variação dos retorno diários das ações com base nas variações do portfólio de mercado?

- Do total da soma (dos quadrados) dos residuos, $\sum_{i=1}^{n} \left(Y_i - \bar{Y}\right)^2 = (n-1)s_Y^2 \ldots$
- $\ldots$ uma parte é explicada pelo modelo, $\sum_{i=1}^{n} \left(\hat{Y}_i - \bar{Y}\right)^2 = s_{\hat{Y}}^2 \ldots$

Goodness of fit

Quanto eu consigo explicar sobre a variação dos retorno diários das ações com base nas variações do portfólio de mercado?

- Do total da soma (dos quadrados) dos residuos, $\sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2 = (n-1)s_Y^2 \ldots$
- ...uma parte é explicada pelo modelo, $\sum_{i=1}^{n}\left(\hat{Y}_i - \bar{Y}\right)^2 = s_{\hat{Y}}^2 \ldots$
- ...e outra parte é explicada pelo erro, $\sum_{i=1}^{n}\left(\varepsilon_i - 0\right)^2 = s_{\varepsilon}^2 \ldots$

Goodness of fit

Quanto eu consigo explicar sobre a variação dos retorno diários das ações com base nas variações do portfólio de mercado?

- Do total da soma (dos quadrados) dos residuos, $\sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2 = (n-1)s_Y^2 \ldots$
- ... uma parte é explicada pelo modelo, $\sum_{i=1}^{n}\left(\hat{Y}_i - \bar{Y}\right)^2 = s_{\hat{Y}}^2 \ldots$
- ... e outra parte é explicada pelo erro, $\sum_{i=1}^{n}\left(\varepsilon_i - 0\right)^2 = s_{\varepsilon}^2 \ldots$
- Assim, podemos definir uma estatística que avalia quâo aderente é o modelo aos dados: $R^2 = \frac{s_{\hat{Y}}^2}{s_Y^2} = 1 - \frac{s_{\varepsilon}^2}{s_Y^2}$

Por quê MQO?

BLUE

- Sob algumas hipóteses (Gauss-Markov), o estimador de mínimos quadros é **BLUE** (*best linear unbiased estimator*). Mesmo sem assumirmos a normalidade dos erros!

BLUE

- Sob algumas hipóteses (Gauss-Markov), o estimador de mínimos quadros é **BLUE** (*best linear unbiased estimator*). Mesmo sem assumirmos a normalidade dos erros!
- Sob a hipótese de normalidade dos erros, o estimador de MQO é o mais eficiente entre os estimadores lineares e não-lineares (Cramér–Rao)!

BLUE

- Sob algumas hipóteses (Gauss-Markov), o estimador de mínimos quadros é **BLUE** (*best linear unbiased estimator*). Mesmo sem assumirmos a normalidade dos erros!
- Sob a hipótese de normalidade dos erros, o estimador de MQO é o mais eficiente entre os estimadores lineares e não-lineares (Cramér–Rao)!
- E quais são essas hipóteses?

Hipóteses

Hipóteses

- **Linearidade**: $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ (linear nos parâmetros, as variáveis podem ser não-lineares).

Hipóteses

- **Linearidade**: $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ (linear nos parâmetros, as variáveis podem ser não-lineares).
- **Exogeneidade**: $E[\varepsilon_i | X_i] = E[\varepsilon_i] = 0$.

Hipóteses

- **Linearidade**: $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ (linear nos parâmetros, as variáveis podem ser não-lineares).
- **Exogeneidade**: $E[\varepsilon_i | X_i] = E[\varepsilon_i] = 0$.
- **Multicolinearidade não-perfeita**: se tivermos mais de uma variável X (e.g. $X_1, X_2, \cdots, X_k$), elas não podem ser perfeitamente correlacionadas.

Hipóteses

- **Linearidade**: $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ (linear nos parâmetros, as variáveis podem ser não-lineares).
- **Exogeneidade**: $E[\varepsilon_i | X_i] = E[\varepsilon_i] = 0$.
- **Multicolinearidade não-perfeita**: se tivermos mais de uma variável X (e.g. $X_1, X_2, \cdots, X_k$), elas não podem ser perfeitamente correlacionadas.
- **Homocedasticidade**: $Var[\varepsilon_i | X_i] = \sigma^2$ e $Cov[\varepsilon_i, \varepsilon_j | X_i] = 0$.

Exogeneidade

Exogeneidade

Esse é um ponto crucial para nós.

Exogeneidade

Esse é um ponto crucial para nós.

- O termo erro (ε_i) inclui, por definição, tudo o que não está no modelo. Ele **não** pode influenciar as variáveis explicativas (X). Se isso acontecer, é porque temos:
  - Variáveis omitidas.

Exogeneidade

Esse é um ponto crucial para nós.

- O termo erro (ε_i) inclui, por definição, tudo o que não está no modelo. Ele **não** pode influenciar as variáveis explicativas (X). Se isso acontecer, é porque temos:
    - Variáveis omitidas.
    - Erro de mensuração das variáveis explicativas.

Exogeneidade

Esse é um ponto crucial para nós.

- O termo erro (ε_i) inclui, por definição, tudo o que não está no modelo. Ele **não** pode influenciar as variáveis explicativas (X). Se isso acontecer, é porque temos:
  - Variáveis omitidas.
  - Erro de mensuração das variáveis explicativas.
  - Simultaneidade

BLUE

BLUE

Sob exogeneidade e Multicolinearidade não-perfeita, o estimador de MQO é **consistente** e **não-viesado**.

BLUE

Sob exogeneidade e Multicolinearidade não-perfeita, o estimador de MQO é **consistente** e **não-viesado**.

- Consistência: $plim_{n \to \infty} |\hat{\beta}_1 - \beta| = 0$.

Sob exogeneidade e Multicolinearidade não-perfeita, o estimador de MQO é **consistente** e **não-viesado**.

- Consistência: $plim_{n \to \infty} |\hat{\beta}_1 - \beta| = 0$.
- Não-viesado: $E[\hat{\beta}_1] = \beta$

Regressão múltipla

Será que podemos melhorar a maneira como respondemos a questão proposta?

Regressão múltipla

- Generalização da regressão simples na qual incluímos mais de uma variável explicativa.

Regressão múltipla

- Generalização da regressão simples na qual incluímos mais de uma variável explicativa.
- Podemos ter (i) a variável explicada,

Regressão múltipla

- Generalização da regressão simples na qual incluímos mais de uma variável explicativa.
- Podemos ter (i) a variável explicada, (ii) a(s) variável(is) de interesse

Regressão múltipla

- Generalização da regressão simples na qual incluímos mais de uma variável explicativa.
- Podemos ter (i) a variável explicada, (ii) a(s) variável(is) de interesse e (iii) variáveis de controle.

Regressão múltipla

- Generalização da regressão simples na qual incluímos mais de uma variável explicativa.
- Podemos ter (i) a variável explicada, (ii) a(s) variável(is) de interesse e (iii) variáveis de controle.
- A diferença é que agora temos diversas dimensões, mas continuamos com uma reta que se ajusta ao minimizar a soma do erro quadrado.

Regressão múltipla

Regressão múltipla

$$Y_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \cdot + \beta_k X_{k,i} + \varepsilon_i.$$

Regressão múltipla

$$Y_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \cdot + \beta_k X_{k,i} + \varepsilon_i.$$

Clique aqui para a matemática do estimador

Quais outras perguntas podemos responder com uma regressão múlitpla?

R^2 e R^2 ajustado

R^2 e R^2 ajustado

- Como podemos comparar modelos? A estatística R^2 **não** é uma boa maneira.

R^2 e R^2 ajustado

- Como podemos comparar modelos? A estatística R^2 **não** é uma boa maneira.
- Podemos utilizar o R^2 ajustado, no entanto:

$$R^2_{\text{adj}} = 1 - \left(1 - R^2\right) \frac{n-1}{n-k-1} \tag{1}$$

onde n é o número de observações da amostra e k representa o número de variáveis independentes do modelo.

Teste-F

Podemos testar a significância conjunta dos estimadores:

$$\mathcal{H}_0 : \beta_0 = \beta_1 = \beta_2 = \cdots = \beta_k = 0$$
$$\mathcal{H}_a : \beta_j \neq 0, \text{ para menos um valor de} j$$

$$F = \frac{\sum_i \epsilon_i^2 - \sum_i e_i^2}{\sum_i e_i^2} \frac{n - k_2}{k_2 - k_1} \sim F_{k_2 - k_1, n - k_2} \tag{2}$$

onde k_2 é o número de parâmetros do modelos irrestrito e k_1 o número de parâmetros do modelo restrito.

Apêndice

Regressão múltipla – estimador de MQO

- $\min_{\hat{\beta}_0,\hat{\beta}_1,\hat{\beta}_2} \sum_{i=1}^{n} \left(\epsilon_i\right)^2 = \sum_{i=1}^{n} \left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1,i} - \hat{\beta}_2 X_{2,i}\right)^2$

Regressão múltipla – estimador de MQO

- $\min_{\hat{\beta}_0,\hat{\beta}_1,\hat{\beta}_2} \sum_{i=1}^{n} \left(\epsilon_i\right)^2 = \sum_{i=1}^{n} \left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1,i} - \hat{\beta}_2 X_{2,i}\right)^2$
  - $\hat{\beta}_1 = \frac{\rho_{X_1,Y} - \rho_{X_1,X_2} \times \rho_{X_2,Y}}{1-\rho^2_{X_1,X_2}}$
  - $\hat{\beta}_2 = \frac{\rho_{X_2,X_1} - \rho_{X_1,X_2} \times \rho_{X_1,Y}}{1-\rho^2_{X_1,Y}}$
  - $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \overline{X_1} - \hat{\beta}_2 \overline{X_2}$

Retornar

Regressão múltipla – estimador de MQO

E com mais variáveis?

Regressão múltipla – estimador de MQO

E com mais variáveis? Vamos utilizar álgebra matricial! Podemos escrever o modelo da seguinte forma:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ \vdots \\ Y_n \end{bmatrix}_{n\times 1} = \begin{bmatrix} 1 & X_{11} & X_{21} & \dots & X_{k1} \\ 1 & X_{12} & X_{22} & \dots & X_{k2} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & X_{1n} & X_{2n} & \dots & X_{kn} \end{bmatrix}_{n\times k} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \vdots \\ \beta_n \end{bmatrix}_{k\times 1} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \vdots \\ \epsilon_n \end{bmatrix}_{n\times 1}$$

Ou, simplesmente

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

◂ Retornar

Regressão múltipla – estimador de MQO

O resíduo (não o erro) pode ser definido como

$$\boldsymbol{e} = \boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta} \tag{3}$$

A soma dos quadrados dos resíduos pode ser escrita como:

$$\begin{bmatrix} e_1 & e_2 & \dots & e_n \end{bmatrix}_{1\times n} \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}_{n\times 1} = [e_1 \times e_1 + e_2 \times e_2 + \ldots + e_n \times e_n]_{1\times 1}$$

Retornar

Regressão múltipla – estimador de MQO

$$
\begin{aligned}
\boldsymbol{e}'\boldsymbol{e} &= (\boldsymbol{Y} - \boldsymbol{X}\hat{\beta})'(\boldsymbol{Y} - \boldsymbol{X}\hat{\beta}) \\
&= \boldsymbol{Y}'\boldsymbol{Y} - \hat{\beta}'\boldsymbol{X}'\boldsymbol{Y} - \boldsymbol{Y}'\boldsymbol{X}\hat{\beta} + \hat{\beta}'\boldsymbol{X}'\boldsymbol{X}\hat{\beta} \\
&= \boldsymbol{Y}'\boldsymbol{Y} - 2\hat{\beta}'\boldsymbol{X}'\boldsymbol{Y} + \hat{\beta}'\boldsymbol{X}'\boldsymbol{X}\hat{\beta}
\end{aligned} \tag{4}
$$

Assim, temos

$$
\frac{\partial \boldsymbol{e}'\boldsymbol{e}}{\partial \hat{\beta}} = -2\boldsymbol{X}'\boldsymbol{Y} + 2\boldsymbol{X}'\boldsymbol{X}\hat{\beta} = 0 \tag{5}
$$

é igual a

$$
\hat{\boldsymbol{\beta}} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\left(\boldsymbol{X}'\boldsymbol{Y}\right).
$$

◂ Retornar

Anscombe, Francis J. 1973. "Graphs in Statistical Analysis." *The American Statistician* 27 (1): 17–21.