Métodos Quantitativos I
Aula 5: Mínimos Quadrados Ordinários

Profs. Arthur Bragança e Daniel Grimaldi

MPAM-ENAP

16 de julho de 2025

Econometria (introd.)

- Na origem da maioria das análises de dados estão perguntas do tipo:
  - Como intervenção de política pública x influencia uma variável y?
  - Que variáveis (x_1, x_2, ...) determinam y?
  - Etc.

- A econometria é o campo da economia que trata da estimação dessas **relações causais** nos dados.

Econometria (introd.)

- O ponto de partida da econometria é a definição de um modelo populacional que descreve as relações entres as variáveis de interesse ("o que queremos explicar?").

- Depois desenvolvem-se hipóteses para a estimação desses parâmetros utilizando uma amostra da população de interesse (**hipóteses de identificação**).
  - Hipóteses de identificação são essencialmente hipóteses sobre o processo gerador de dados.
  - Não é possível avalia-las diretamente, i.e., não é possível **testa-las**.
  - Mas é **possível** (e **recomendável**) avaliar sua plausibilidade.

- Por último utiliza-se algum método de estimação (ex.: método dos momentos, máxima verossimilhança etc.) para estimar o modelo.

Econometria (introd.)

- Nas próximas **três** aulas discutiremos o modelo de regressão linear.
- Esse é o modelo econométrico mais utilizado para realizarmos estudos empíricos.
- Discutiremos a estimação, a inferência e a interpretação desse modelo.
- Esse modelo tem inúmeras limitações. Entretanto,
  - Entender suas propriedades é fundamental para compreender modelos mais complexos.
  - Ele é o melhor modelo para examinar relações de **causa-e-efeito** em contextos específicos.
  - Ele pode ser útil mesmo quando ele não é o melhor modelo para examinar relações de **causa-e-efeito**.

Modelo de Regressão Bivariado (1)

- Começaremos especificando um modelo de regressão linear simples.

- Esse modelo populacional tem apenas duas variáveis:
  - y: variável dependente ("o que queremos explicar")
  - x: variável explicativa ou regressor ("o que explica y")

- Exemplos:
  - Efeito de mudanças no preço (x) na demanda por um produto (y);
  - Efeitos de escolaridade (x) sobre os salários dos indivíduos (y);
  - Efeitos do auxílio emergencial (x) sobre a pobreza das famílias (y).

Modelo de Regressão Bivariado (2)

- No modelo de regressão linear a variável dependente (y) é uma função **aditiva** e **linear** da variável explicativa (x) e de um termo de erro não observável (u).

- Matematicamente, temos:

$$y_i = \alpha + \beta x_i + u_i$$

onde i indexa a i-ésima observação da população

Modelo de Regressão Bivariado (3)

- O parâmetro de interesse do modelo de regressão linear é β.
- Esse parâmetro mede o efeito de uma mudança unitária em x_i sobre y_i.
- Note que

$$\frac{\Delta y_i}{\Delta x_i} = \beta, \forall x_i$$

  Isso significa que o efeito de uma mudança unitária x_i sobre y_i é constante. **Hipótese muito restritiva!**
- Reinterpretação:
  - β é "média" dos efeitos de mudanças unitárias para diferentes níveis de x_i.

Hipóteses (1)

- Computar β requer hipóteses sobre relação entre regressor e termo de erro.
- Por quê?
  - Suponha que y_i é mais alto quando x_i é mais alto.
  - Isso implica que $\beta > 0$?
  - Sem hipóteses sobre a relação entre termo de erro e x_i é impossível determinar.
- Hipótese fundamental é que $E[u_i|x_i] = 0$.

Hipóteses (2)

- A hipótese fundamental é que $E[u_i|x_i] = 0$.
- **Implicação 1**: $E[u_i] = 0$

  Prova: $E[u_i] = E\left[E[u_i|x_i]\right] = 0$
- **Implicação 2**: $E[u_i x_i] = 0$

  Prova: $E[u_i x_i] = E\left[E[u_i X_i|x_i]\right] = E\left[E[u_i|x_i]x_i\right] = 0$
- Tipicamente supõe-se que $u_i \sim N(0, \sigma^2)$. Mas essa hipótese não é essencial.

O QUE $E[u_i|x_i] = 0$ SIGNIFICA?

- Considere um modelo conectando o logaritmo natural dos salários ($\log y_i$) com anos de estudo (e_i):

$$\log y_i = \beta_0 + \beta_1 e_i + u_i$$

- Considere que a habilidade dos indivíduos (a_i) é um determinante não observável dos salários:

$$u_i = a_i + \nu_i$$

- $E[u_i|e_i] = 0$ implica habilidade média dos indivíduos é idêntica independente da escolaridade. Por ex.,

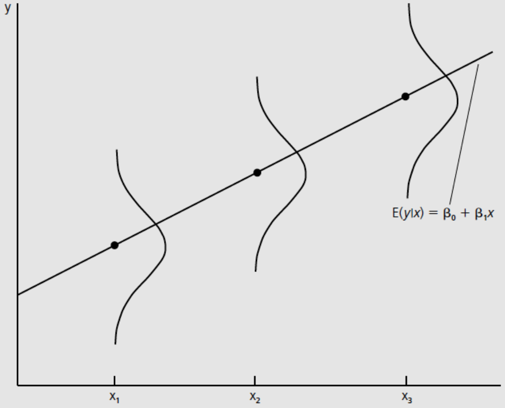
$$E[a_i|e_i = 9] = E[a_i|e_i = 12] = E[a_i|x_i = 16] = 0$$

O que $E[u_i|x_i] = 0$ significa?

Note que $E[y_i|x_i] = \alpha + \beta x_i + E[u_i|x_i] = \alpha + \beta x_i$.

Figure 2.1

E(y|x) as a linear function of x.



Parâmetro Populacional (1)

- Vimos que:

$$E[x_i|u_i] = 0 \Longrightarrow E[u_i] = 0 \text{ e } E[x_i u_i] = 0$$

- O termo de erro pode ser escrito como:

$$u_i = y_i - \alpha - \beta x_i$$

- Isso implica que:

$$E[y_i - \alpha - \beta x_i] = 0$$

$$E[x_i(y_i - \alpha - \beta x_i)] = 0$$

- Podemos utilizar o sistema de equações acima para computar os parâmetros (α, β).

Parâmetro Populacional (2)

- A primeira equação implica:

$$\alpha = E[y_i] - \beta E[x_i]$$

- A segunda equação implica:

$$E[x_i y_i] - \alpha E[x_i] - \beta E[x_i^2] = 0$$

- Substituindo α, temos:

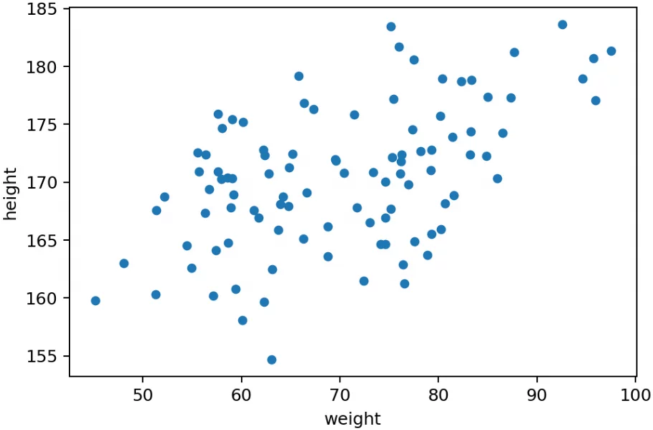
$$E[x_i y_i] - E[x_i]E[y_i] - \beta\left(E[x_i^2] - E[x_i]^2\right) = 0$$

- Portanto:

$$\beta = \frac{E[x_i y_i] - E[x_i]E[y_i]}{E[x_i^2] - E[x_i]^2} = \frac{cov(x_i, y_i)}{var(x_i)}$$

Estimação (1)

- Na prática iremos estimar os parâmetros a partir de uma amostra $\{(y_i, x_i) : i = 1, ...n\}$.



Estimação (2)

- Na prática iremos estimar os parâmetros a partir de uma amostra $\{(y_i, x_i) : i = 1, ...n\}$.
- Note que cada observação i é uma retirada da população que segue a relação entre y_i e x_i do modelo populacional.
- Portanto, o modelo amostral é:

$$y_i = \alpha + \beta x_i + u_i, \forall i = \{1, ..., n\}$$

- Como podemos estimar os parâmetros $(\hat{\alpha}, \hat{\beta})$?
- É possível estimar esses parâmetros utilizando os análogos amostrais das condições $E[u_i] = 0$ e $E[x_i u_i] = 0$.

Estimação (3)

- Podemos utilizar os análogos amostrais das condições $E[u_i] = 0$ e $E[x_i u_i] =$.

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - \alpha - \beta x_i) = 0$$

$$\frac{1}{n}\sum_{i=1}^{n}[x_i(y_i - \alpha - \beta x_i)] = 0$$

- Resolvendo o sistema:

$$\hat{\alpha} = \overline{y} - \hat{\beta}\overline{x}$$

$$\hat{\beta} = \frac{\sum(y_i - \overline{y})(x_i - \overline{x})}{\sum(x_i - \overline{x})^2} = \frac{\widehat{cov(x_i, y_i)}}{\widehat{var(x_i)}}$$

Estimação (4)

- O estimador obtido é chamado de estimador de **método dos momentos**.
- Esse nome decorre dele ser construído a partir de condições de momento ("médias" que devem ser iguais a zero).
- Essas condições de momento decorrem diretamente das hipóteses do modelo de regressão linear, i.e., de hipóteses sobre o processo gerador de dados.
- Essas hipóteses que garantem que $\hat{\beta}$ identifica o **efeito causal** de mudanças de x sobre y.

Reta de Regressão (1)

- A reta de regressão é a reta que conecta x com a expectativa condicional de y:

$$\hat{y}_i = \hat{\alpha} + \hat{\beta} x_i$$

- Os parâmetros $(\hat{\alpha}, \hat{\beta})$ podem ser interpretados como o intercepto e a inclinação dessa reta de regressão.

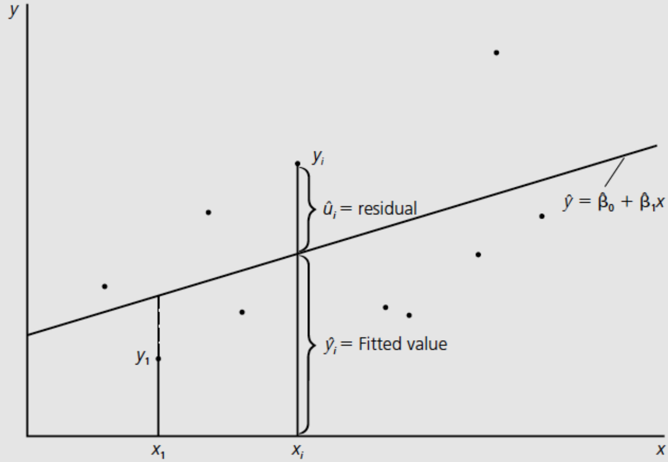
- Os resíduos podem ser definidos como a distância de cada observação para a reta de regressão:

$$\hat{u}_i = y_i - \hat{y}_i = y_i - \hat{\alpha} - \hat{\beta} x_i$$

Reta de Regressão (2)

Figure 2.4

Fitted values and residuals.



Mínimos Quadrados Ordinários (1)

- A reta de regressão nos dá uma outra intuição para os parâmetros $(\hat{\alpha}, \hat{\beta})$.
- Esses parâmetros constroem a reta mais bem ajustada ao conjunto de dados.
- Formalmente,

$$\min_{\alpha,\beta} \sum_{i=1}^{n} u_i^2 = \min_{\alpha,\beta} \sum_{i=1}^{n} (y_i - \alpha - \beta x_i)^2$$

- Os parâmetros minimizam isso significa que esses parâmetros são aqueles que minimizam uma métrica de distância entre valores observados e preditos.

Mínimos Quadrados Ordinários (2)

- As condições de primeira ordem desse problema são:

$$\sum_{i=1}^{n} (y_i - \alpha - \beta x_i) = 0$$

$$\sum_{i=1}^{n} [x_i (y_i - \alpha - \beta x_i)] = 0$$

- O sistema de equações acima é exatamente o mesmo obtido utilizando método dos momentos.

- Solução é idêntica e resulta no mesmo intercepto e inclinação.

Modelo de Regressão em R

- Estimaremos um modelo de regressão no R em três etapas:
    1. Simular um conjunto de dados.
    2. Computar estimadores de mínimos quadrados ordinários manualmente.
    3. Computar estimadores de mínimos quadrados ordinários utilizando funções do R.

- Utilizaremos o RStudio.

Pacotes

Começamos carregando um pacote para estimar regressões lineares.

```
# Instala pacote #

  install.packages("estimatr")

# Carrega pacote #

library(estimatr)
```

Coeficientes de MQO

Considere uma amostra de 10.000 observações de um conjunto de dados cujo modelo populacional é dado por $y_i = 5.5 * x_i + u_i$ em que $x_i \sim N(0,1)$ e $u_i \sim N(0,100)$.

```
## Constroi vetores y, x e u ##
set.seed(33)
x = rnorm(10000)
u = 10*rnorm(10000)
y = 5.5*x + u
## Estima parâmetros ##
beta = sum( (y- mean(y) )*(x - mean(x) ) ) / sum( (x-mean(x) )^2 )
alpha = mean(y) - beta*mean(x)
```

Comando lm_robust()

Podemos utilizar o comando lm_robust() para estimar regressões.

```
## Comando lm_robust() ##

lm_robust(y ~ x)

## Objeto com resultados...

modelo = lm_robust(y ~ x)

## O que temos no objeto?

modelo$coefficients # coeficientes #

modelo$vcov # matriz de variância #
```

Fit

- A variância de y pode ser decomposta em dois termos:

$$\begin{aligned}
SQT = \sum_{i=1}^{n}(y_i - \overline{y})^2 &= \sum_{i=1}^{n}\left[(y_i - \hat{y}_i) + (\hat{y}_i - \overline{y})\right]^2 \\
&= \sum_{i=1}^{n} u_i^2 + 2\sum_{i=1}^{n} u_i(\hat{y}_i - \overline{y}) + \sum_{i=1}^{n}(\hat{y}_i - \overline{y})^2 \\
&= \sum_{i=1}^{n} u_i^2 + \sum_{i=1}^{n}(\hat{y}_i - \overline{y})^2 \\
&= SQR + SQE
\end{aligned}$$

- Importância relativa de SQR e SQE é medida de "ajuste" da regressão.

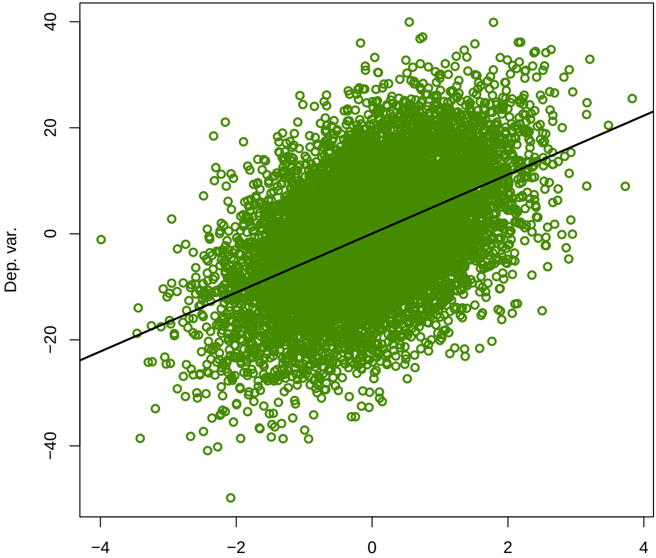
Fit

- Essa medida de ajuste é chamada de $R2$.
- Ela é definida como:

$$R2 = \frac{SQE}{SQT} = 1 - \frac{SQR}{SQT}$$

- Intuitivamente ela nos dá uma medida de quanto da variância de y é explicado (no sentido estatístico) por x.

Fit



Como eu fiz isso?

```
# inicio o arquivo onde vou salvar o grafico
pdf("fit.pdf")

# faço gráfico dos dados utilizados na estimacao
plot(x,y, type = "p", xlab = "Indep. var.", ylab = "Dep. var.", lwd = 2, col = "chartreuse4")

# adiciono a linha com o y predito
abline(a = alpha, b = beta, lwd = 2, col = "black")

# fecho o arquivo
dev.off()
```

Fit

- O *fit* não tem relação com a identificação dos parâmetros do modelo.

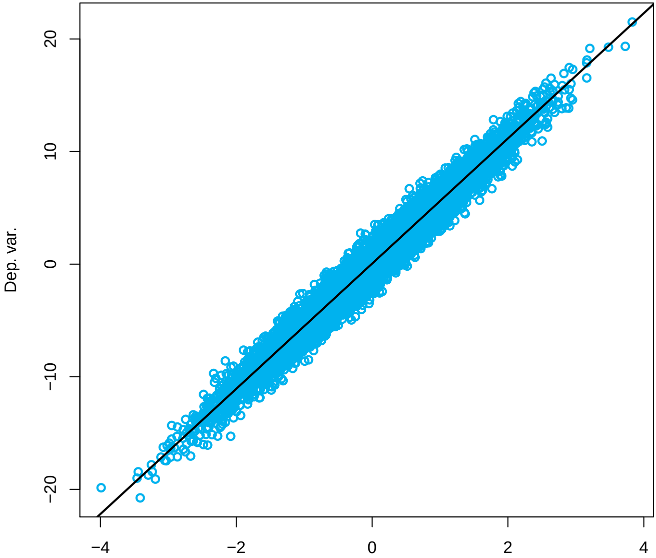
- A identificação depende apenas da expectativa do erro condicional ao regressor ser zero ($E[u_i|x_i] = 0$).

- O $R2$ se relaciona apenas com a proporção da variação em y que é explicada por x e que não é explicada por x ("explicada" por u).

β IDÊNTICO, $R2$ DIFERENTE

```
# Modelo Original ##

y = 5.5*x + u

summary(lm_robust(y ~ x))

Coefficients:
            Estimate Std. Error t value Pr(>|t|) CI Lower CI Upper   DF
(Intercept)   0.0474     0.1003  0.4725   0.6366  -0.1492    0.244 9998
x             5.5604     0.1018 54.6002   0.0000   5.3608    5.760 9998

Multiple R-squared:  0.2297 , Adjusted R-squared:  0.2296

## Menos dispersão ##

y = 5.5*x + 0.1*u

summary(lm_robust(y ~ x))

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.00474    0.01003   0.472    0.637
x            5.50604    0.01018 540.664   <2e-16 ***
---
Multiple R-squared:  0.9669,  Adjusted R-squared:  0.9669
```

Menos Dispersão



Como eu fiz isso?

```r
# inicio o arquivo onde vou salvar o grafico
pdf("fit.pdf")

# faço gráfico dos dados utilizados na estimacao
plot(x,y, type = "p", xlab = "Indep. var.", ylab = "Dep. var.", lwd = 2, col = "deepskyblue2")

# adiciono a linha com o y predito
abline(a = alpha, b = beta, lwd = 2, col = "black")

# fecho o arquivo
dev.off()
```

Interpretação de Coeficientes

- Considere o modelo:

$$y_i = \alpha + \beta x_i + u_i$$

- Coeficiente β nos dá o efeito de um aumento unitário em x_i.
- Mas muitas vezes o interesse é em outras métricas.
- **Exemplos**:
  1. Um aumento unitário em x aumenta / diminui y em quantos %?
  2. Um aumento de 1% em x aumenta / diminui y em quantos %?
  3. Um aumento de um desvio-padrão de x aumenta / diminui y em quantos desvios-padrão?
- Duas alternativas: reescalonar coeficientes ou modificar modelo.

Modelo log-nível

- Considere o modelo:

$$\log y_i = \alpha + \beta x_i + u_i$$

- Esse modelo implica:

$$y_i = \exp(\alpha + \beta x_i + u_i)$$

- O efeito de um aumento unitário de x é:

$$\frac{\Delta y_i}{\Delta x_i} = \exp(\alpha + \beta(x_i + 1) + u_i) - \exp(\alpha + \beta x_i + u_i) = (\exp(\beta) - 1)y_i$$

Modelo log-nível

- Esse efeito como proporção de y ("percentual") é:

$$\frac{\Delta y_i}{\Delta x_i} \times \frac{1}{y_i} = \exp(\beta) - 1 \approx \beta$$

- Portanto, o coeficiente desse modelo aproxima o efeito percentual de um aumento unitário de x.

- Aproximação funciona mal quando coeficiente é grande (**regra de bolso**: acima de 0.3).

Modelo log-log

- Considere o modelo:

$$\log y_i = \alpha + \beta \log x_i + u_i$$

- O coeficiente desse modelo identifica o efeito de um aumento de $\log x$ em $\log y$.
- Sabemos que uma pequena no logaritmo de uma variável aproxima uma mudança percentual nessa variável.
- Portanto, o modelo identifica efeitos percentuais em y de mudanças percentuais em x.
  - Em economia esse parâmetro é tipicamente chamado de elasticidade.

Exemplo: educação e salários

Iremos ilustrar questões relativas a forma funcional com dados reais.

```
  root = ""

  setwd(root)

# Instala e carrega pacote #

  install.packages("tidyverse")

  library(tidyverse)

  library(haven)

# Abre banco de dados #

  twins = read_dta("pubtwins.dta")

# Resume variáveis do banco de dados #

  summary(twins)
```

Modelo em nível

Começando com uma regressão em nível.

```
mod = lm_robust(hrwage ~ educ, data = twins)

summary(mod)

Call:
lm_robust(formula = hrwage ~ educ, data = twins)

Standard error type:  HC2

Coefficients:
            Estimate Std. Error t value  Pr(>|t|) CI Lower CI Upper  DF
(Intercept)  -12.705     3.5909  -3.538 4.306e-04  -19.756   -5.654 678
educ           1.935     0.2757   7.017 5.495e-12    1.393    2.476 678

Multiple R-squared:  0.09498 ,  Adjusted R-squared:  0.09364
F-statistic: 49.24 on 1 and 678 DF,  p-value: 5.495e-12
```

Interpretação

- Um ano de estudo aumenta salário por hora em \$ 1.93.

- Qual o efeito percentual?

```
mod$coefficients["educ"]/mean(twins$hrwage)

0.1340074
```

- Um ano de estudo aumenta salário por hora em cerca de 13%.

Modelo log-nível

Rodamos agora uma regressão log-nível.

```
 mod = lm_robust(lwage ~ educ, data = twins)

 summary(mod)

Call:
lm_robust(formula = lwage ~ educ, data = twins)

Standard error type:  HC2

Coefficients:
            Estimate Std. Error t value  Pr(>|t|) CI Lower CI Upper  DF
(Intercept)   1.0077    0.15791   6.381 3.257e-10  0.69761   1.3177 678
educ          0.1022    0.01151   8.880 5.934e-18  0.07958   0.1248 678

Multiple R-squared:  0.1164 , Adjusted R-squared:  0.1151
F-statistic: 78.86 on 1 and 678 DF,  p-value: < 2.2e-16
```

Interpretação

- Um ano de estudo aumenta salário por hora em torno de 10%.
- Esse efeito é uma aproximação (efeito exato é $\exp(0.1022) - 1 = 0.1076$).
- Modelo em log diminui influência dos *outliers* no resultado (mais sobre isso nas próximas aulas)

Modelo de Regressão Multivariado

- Considere agora a relação populacional entre uma variável dependente (y) e um conjunto de regressores $(x_1, \cdots, x_k)$ é:

$$y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + ... + \beta_k x_{ki} + u_i = X_i' \beta + u_i, \forall i$$

onde i indexa a i-ésima observação da população

X_i é um vetor $k \times 1$ de regressores e β é um vetor $k \times 1$ de parâmetros.

Normalmente a primeira linha de X é uma constante.

Hipóteses (revisão)

- A hipótese fundamental é que $E[u_i|X_i] = 0$.
- **Implicação 1**: $E[u_i] = 0$

  Prova: $E[u_i] = E\left[E[u_i|X_i]\right] = 0$
- **Implicação 2**: $E[u_i X_i] = 0$

  Prova: $E[u_i X_i] = E\left[E[u_i X_i|X_i]\right] = E\left[E[u_i|X_i]X_i\right] = 0$
- Tipicamente supõe-se que $u_i \sim N(0, \sigma^2)$. Mas essa hipótese não é essencial.

Interpretação (1)

- β_k é efeito de uma mudança em x_k sobre y mantendo os outros regressores fixos
- Isso significa que modelo permite estudar efeito de mudar apenas um regressor mesmo quando dados incluem variação de todos eles ao mesmo tempo.
- Tenta "emular" experimento controlado em que só um regressor é manipulado.

Interpretação (2)

- Temos interesse no efeito de múltiplos regressores em y.
- **Modelo de preços hedônicos**:
  - p_i é o preço de um bem heterogêneo (ex.: carro)
  - X_i é vetor de características observáveis desse bem (ex.: marca, potência, cor, tamanho etc.)
  - Interesse em estimar o efeito de diferentes características sobre preço ("valoração").
- Modelo de regressão multivariado conectando preço a diferentes características:

$$p_i = X_i'\beta + u_i$$

Interpretação (3)

- Temos interesse no efeito de um regressor x_{i1} sobre y.
- Mas inserimos regressores adicionais para garantir que regressor não é correlacionado com termo de erro ("controles").
- Considere o modelo:

$$y_i = \beta_0 + \beta_1 x_{i1} + \varepsilon_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + u_i$$

- x_{i1} correlacionado com termo de erro da primeira equação ($E[x_{i1}\varepsilon_i] \neq 0$).
- x_{i1} não correlacionado com termo de erro da primeira equação ($E[x_{i1}u_i] = 0$).

Parâmetro Populacional (1)

- Podemos utilizar $E[X_i u_i] = 0$ para derivar o parâmetro populacional.

- Note que $u_i = y_i - X_i'\beta$. Portanto,

$$E[X_i u_i] = E[X_i(y_i - X_i'\beta)] = E[X_i y_i] - E[X_i X_i']\beta = 0$$

- Isso implica que o parâmetro populacional é:

$$\beta = E[X_i X_i']^{-1} E[X_i y_i]$$

- Lembre que X_i é $k \times 1$ o que implica que $E[X_i X_i']^{-1}$ tem dimensão $k \times k$ e $E[X_i y_i]$ dimensão $k \times 1$.

- Logo, β tem dimensão $k \times 1$ como gostaríamos.

Parâmetro Populacional (2)

- Note que $E[u_i|X_i] = 0$ implica:

$$E[y_i|X_i] = X_i\beta$$

- Isso implica que (y_i, X_i) tem distribuição normal bivariada se $u_i \sim N(0, \sigma^2)$.

- Em estatística aprendemos que sob normalidade conjunta β é dado por $E[X_iX_i']^{-1}E[X_iy_i]$.

Estimação (1)

- Suponha que queremos estimar os parâmetros do modelo de regressão linear a partir de uma amostra $\{(y_i, x_{i1}, \ldots, x_{ik}) : i = 1, ...n\}$.

- Note que cada observação i é uma retirada da população que segue a relação entre y_i e X_i do modelo populacional.

- Logo, o modelo amostral pode ser escrito como:

$$y = X\beta + u,$$

onde $X = \begin{bmatrix} X_1' \\ \vdots \\ X_n' \end{bmatrix} = \begin{bmatrix} x_{11} & \cdots & x_{1k} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nk} \end{bmatrix}; Y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}; u = \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix}$

Estimação (2)

- Estimador de mínimos quadrados ordinários (MQO) é o conjunto de βs que minimiza a soma do quadrado dos resíduos:

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^{n} u_i^2 = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^{n} (y_i - X_i'\beta)^2$$

- Condição de primeira ordem (use regra da cadeia):

$$\sum_{i=1}^{n} 2X_i(y_i - X_i'\hat{\beta}) = 0$$

$$\sum_{i=1}^{n} (X_i y_i) - \sum_{i=1}^{n} (X_i X_i')\hat{\beta} = 0$$

$$\hat{\beta} = \left(\sum_{i=1}^{n} X_i X_i' \right)^{-1} \left(\sum_{i=1}^{n} X_i y_i \right)$$

Estimação (3)

- Em notação matricial:

$$\sum_{i=1}^{n} u_i^2 = u'u = (Y - X\beta)'(Y - X\beta)$$

- Minimizando $u'u$:

$$-2X'Y + 2X'X\hat{\beta} = 0 \Longrightarrow \hat{\beta} = (X'X)^{-1}(X'Y)$$

- Note que o estimador só existe se a matriz X tem o posto cheio (i.e., é inversível).

Estimação (3)

- Estimador equivalente é obtido se tentamos tornar condição $E[X_i u_i] = 0$ na amostra o mais próxima possível de zero (método dos momentos).

$$\frac{1}{n}\sum_{i=1}^{n} X_i(y_i - X_i'\hat{\beta}) = 0$$

$$\hat{\beta} = \left(\sum_{i=1}^{n} X_i X_i'\right)^{-1}\left(\sum_{i=1}^{n} X_i y_i\right)$$

- Estimador equivalente é obtido se maximizamos a chance de $u_i = y_i - X_i'\beta$ ser normal com média 0 e variância σ^2 (máxima verossimilhança)

Exemplo: Dados de Educação e Salários

Começando com uma regressão bivariada.

```
# Regressoes #

  modelo1 = lm_robust(lwage ~ educ, data = twins) # resultados armazenados em uma lista contendo diferentes
     objetos como: #

  # coeficientes #

  modelo1$coefficients

  # matriz de variância #

  modelo1$vcov

  # ... #
```

Podemos adicionar controles.

```
  modelo2 = lm_robust(lwage ~ educ + age + age2, data = twins) # é possível adicionar controles #

  summary(modelo2)

  modelo3 = lm_robust(lwage ~ educ + age + age2 + female + white, data = twins) # é possível adicionar controles
     #

  summary(modelo3)
```

Comando modelsummary()

É possível exportar os resultados da regressão para tabelas de diferentes formatos.

```
# instala e carrega pacote #

install.packages("modelsummary")

library(modelsummary)

  # podemos exportar resultados para tabela em formatos como "texto", "html" ou "latex" #

 modelos = list("MQO 1" = modelo1,
                "MQO 2" = modelo2,
                "MQO 3" = modelo3)

modelsummary(modelos)

  # pode salvar como Markdown, Latex ou mesmo Word (instalar pacote ‘‘flextable’’)
```

Interpretação (1)

- Que controles queremos incluir?
- **Bons controles:**
  - "Explicam" variável dependente (**regra de bolso:** sua inclusão aumenta o $R2$?);
  - Potencialmente correlacionados com regressor de interesse (mas não determinados por ele).
- **Controles ruins:**
- Não explicam variável dependente ou são muito correlacionados entre si;
- São "endógenos" (potencialmente determinados pelo regressor de interesse).

Interpretação (2)

- **Exemplo**: Considere o uma regressão conectando salários com educação:

$$\log y_i = \alpha + \beta e_i + W_i' \Gamma + u_i$$

- Que controles incluir em X_i?
  - Controles como idade, sexo, cor, estado civil, habilidade, onde reside etc. são bons controles
  - Características do emprego são controles ruins (ex,: setor de atividade).

O que os controles fazem? (1)

- Controlar por uma variável é equivalente a limpar a variação das variáveis de interesse que vem dos controles.
- É possível ver isso com a ideia de **regressão residual**.
- Considere modelo de regressão:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + u_i$$

- É possível expressar β_1 e β_2 como:

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^{n} \widehat{r}_{i1} y_i}{\sum_{i=1}^{n} \widehat{r}_{i1}^2}$$

$$\widehat{\beta}_2 = \frac{\sum_{i=1}^{n} \widehat{r}_{i2} y_i}{\sum_{i=1}^{n} \widehat{r}_{i2}^2}$$

O que os controles fazem? (2)

- r_{i1} e r_{i2} são resíduos de regressões de x_{1i} em x_{2i} e x_{2i} e x_{1i}, respectivamente.
- Intuição é que "controlar" limpa a variação de x_1 ou x_2 que vem do outro fator.
- Isso permite estimar efeitos de uma variável mantendo a outra fixa.

Exemplo

- Relação entre y e os regressores x_1 e x_2 seja descrita pelo modelo populacional:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$$

- Relação entre x_1 e x_2 é descrita por:

$$x_2 = \alpha x_1 + \nu$$

- Suponha que os parâmetros populacionais sejam $\beta_0 = 0.1$, $\beta_1 = 0.6$, $\beta_2 = 0.4$ e $\alpha = 0.5$ e as distribuições $u \sim N(0, 4)$, $\nu \sim N(0, 16)$ e $x_1 \sim N(0, 1)$.

Dados

Simulo 10.000 observações dos vetores u, ν, x_1, x_2 e y.

```
# Vetores #
set.seed(1985)
beta0 = 0.1
beta1 = 0.6
beta2 = 0.4
alpha = 0.5
u = rnorm(10000, mean = 0, sd = 2)
v = rnorm(10000, mean = 0, sd = 4)
x1 = rnorm(10000, mean = 0, sd = 1)
x2 = alpha*x1+v
y = beta0 + beta1*x1 + beta2*x2 + u
dados = data.frame(cbind(y,x1,x2))
```

Regressão Multivariada

Rodo regressão longa.

```
# Regressão #

full.model = lm_robust(y ~ x1 + x2, data = dados)

summary(full.model)

Standard error type:  HC1

Coefficients:
            Estimate Std. Error t value   Pr(>|t|) CI Lower CI Upper   DF
(Intercept)  0.06576   0.019964   3.294  9.915e-04  0.02663   0.1049 9997
x1           0.61181   0.020065  30.491 2.504e-195  0.57248   0.6511 9997
x2           0.39489   0.005066  77.950  0.000e+00  0.38496   0.4048 9997

Multiple R-squared:  0.4366 , Adjusted R-squared:  0.4364
F-statistic:  3858 on 2 and 9997 DF,  p-value: < 2.2e-16
```

Regressão Residual

Rodo regressões entre os regressores.

```
res.reg1 = lm_robust(x1 ~ x2, data = dados)

res.reg2 = lm_robust(x2 ~ x1, data = dados)

dados$r1 = dados$x1 - res.reg1$fitted.values
dados$r2 = dados$x2 - res.reg2$fitted.values
```

Regressão Residual

Rodo regressões residuais.

```
summary(lm_robust(y ~ r1, data = dados))

Call:
lm_robust(formula = y ~ r1, data = dados)

Standard error type:  HC2

Coefficients:
            Estimate Std. Error t value   Pr(>|t|)   CI Lower CI Upper   DF
(Intercept)  0.04987    0.02590   1.926  5.415e-02 -0.0008902   0.1006 9998
r1           0.61181    0.02636  23.210 4.098e-116  0.5601380   0.6635 9998

Multiple R-squared:  0.05184 ,  Adjusted R-squared:  0.05175
F-statistic: 538.7 on 1 and 9998 DF,  p-value: < 2.2e-16
```

Regressão Residual

Rodo regressões residuais.

```
summary(lm_robust(y ~ r2, data = dados))

Call:
lm_robust(formula = y ~ r2, data = dados)

Standard error type:  HC2

Coefficients:
            Estimate Std. Error t value Pr(>|t|) CI Lower CI Upper   DF
(Intercept)  0.04987   0.021497    2.32  0.02037 0.007732  0.09201 9998
r2           0.39489   0.005439   72.60  0.00000 0.384224  0.40555 9998

Multiple R-squared:  0.3466 , Adjusted R-squared:  0.3465
F-statistic:  5271 on 1 and 9998 DF,  p-value: < 2.2e-16
```

Propriedades (1)

- **Mínimos Quadrados Ordinários (MQO)**: estimadores com propriedades desejáveis.

- O que são propriedades desejáveis?
  1. **Não viesado**: expectativa do estimador do parâmetro é igual ao parâmetro verdadeiro
  2. **Eficiente**: variância do estimador é a menor possível

- Para ver essas propriedades iremos listar um conjunto de hipóteses.

- Depois derivaremos a média e a variância do estimador.

Propriedades

- **Hipóteses**:
  1. **Modelo populacional:** o modelo populacional é
  $$y = X\beta;$$
  2. **Amostragem aleatória:** $\{(y_i, x_{i1}, \ldots, x_{ik}) : i = 1, ...n\}$ é uma amostra independente e identicamente distribuída;
  3. **Independência condicional**: $E[u|X] = 0$;
  4. **Variação em X:** $X = [X_1', \cdots X_n']'$ é matriz $n \times k$ com posto cheio;
  5. **Homocedasticidade:** a variância o termo de erro é constante:
  $$V[u|X] = E[uu'|X] = \sigma^2 I$$
- As hipóteses acima muitas vezes são chamadas de hipóteses de Gauss-Markov.
  - Muitas vezes é adicionada a hipótese que $u|X \sim N(0, \sigma^2 I)$.

Média de $\widehat{\beta}$

- A média do estimador de MQO condicional a X é

$$
\begin{aligned}
E[\widehat{\beta}|X] &= E\left[(X'X)^{-1}X'y|X\right] \\
&= E\left[(X'X)^{-1}X'(X\beta + u)|X\right] \\
&= \beta + (X'X)^{-1}X'E[u|X] \\
&= \beta
\end{aligned}
$$

- Lembre que $E[\widehat{\beta}] = E\left[E[\widehat{\beta}|X]\right]$.
- Isso significa $E[\widehat{\beta}] = E\left[E[\widehat{\beta}|X]\right] = E[\beta] = \beta$.
- Logo, $\widehat{\beta}$ é um estimador não viesado de β.

Média de $\widehat{\beta}$ (bivariado)

- A média do estimador de MQO condicional a x_i é

$$
\begin{aligned}
E[\widehat{\beta}|x_i] &= E\left[\frac{\sum(x_i-\overline{x})(y_i-\overline{y})}{\sum(x_i-\overline{x})^2}|x_i\right] \\
&= E\left[\frac{\sum(x_i-\overline{x})y_i}{\sum(x_i-\overline{x})^2}|x_i\right] - E\left[\frac{\sum(x_i-\overline{x})\overline{y}}{\sum(x_i-\overline{x})^2}|x_i\right] \\
&= E\left[\frac{\sum(x_i-\overline{x})(\alpha+\beta x_i+u_i)}{\sum(x_i-\overline{x})^2}|x_i\right] \\
&= \beta + E\left[\frac{\sum(x_i-\overline{x})u_i}{\sum(x_i-\overline{x})^2}|x_i\right] \\
&= \beta + \left[\frac{\sum(x_i-\overline{x})}{\sum(x_i-\overline{x})^2}\right] E\left[u_i|x_i\right] = \beta
\end{aligned}
$$

- Isso significa $E[\widehat{\beta}] = E\left[E[\widehat{\beta}|x_i]\right] = E[\beta] = \beta$.

Variância de $\widehat{\beta}$

- A variância de $\widehat{\beta}$ condicional a X é:

$$V[\widehat{\beta}|X] = E\left[(\widehat{\beta}-\beta)(\widehat{\beta}-\beta)'|X\right]$$

- Note que:

$$\widehat{\beta} = (X'X)^{-1}X'y = (X'X)^{-1}(X'X)\beta + (X'X)^{-1}X'u$$
$$\Longrightarrow \widehat{\beta}-\beta = (X'X)^{-1}X'u$$

- Portanto:

$$\begin{aligned} V[\widehat{\beta}|X] &= E\left[((X'X)^{-1}X'u)((X'X)^{-1}X'u)'|X\right] \\ &= E\left[(X'X)^{-1}X'uu'X(X'X)^{-1}|X\right] \\ &= (X'X)^{-1}X'E[uu'|X]X(X'X)^{-1} \\ &= \sigma^2(X'X)^{-1} \end{aligned}$$

Variância de $\widehat{\beta}$

- A lei da variância total diz:

$$V[\widehat{\beta}] = E\left[V[\widehat{\beta}|X]\right] + V\left[E[\widehat{\beta}|X]\right]$$

- Temos:

$$V\left[E[\widehat{\beta}|X]\right] = V[\beta] = 0$$

- Logo,

$$V[\widehat{\beta}] = E\left[V[\widehat{\beta}|X]\right] = \sigma^2(X'X)^{-1}$$

- Variância do estimador de MQO depende da variância do termo de erro ("variância residual de y") e da variância dos regressores.

Variância de $\widehat{\beta}$ (bivariado)

- A variância de $\widehat{\beta}$ condicional a x_i é:

$$V[\widehat{\beta}|x_i] = E\left[(\widehat{\beta} - \beta)^2|x_i\right]$$

- Note que:

$$\widehat{\beta} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2} = \beta + \frac{\sum_{i=1}^{n}(x_i - \overline{x})u_i}{\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

$$\Longrightarrow \widehat{\beta} - \beta = \frac{\sum_{i=1}^{n}(x_i - \overline{x})u_i}{\sum_{i=1}^{n}(x_i - \overline{x})^2} = \sum_{i=1}^{n} w_i u_i$$

- Portanto:

$$V[\widehat{\beta}|x_i] = E\left[\left(\sum_{i=1}^{n} w_i u_i\right)^2 |x_i\right] = E\left[\left(\sum_{i=1}^{n} w_i^2 u_i^2\right)|x_i\right]$$

$$= \sigma^2 \left(\sum_{i=1}^{n} w_i\right)^2 = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

Teorema de Gauss-Markov

- Se as hipóteses 1-5 são válidas, o estimador de MQO é o melhor estimador linear não viesado do vetor β.

- O que isso significa? Considere a classe de estimadores lineares:

$$\widetilde{\beta} = \sum_{i=1}^{n} W_i y_i,$$

  em que W_i é uma função da amostra.

- Teorema de Gauss-Markov diz que $W_i = (X_i X_i')^{-}1 X_i$ é a função que:
  1. ... faz com que $\widetilde{\beta}$ tenha a menor variância possível ("melhor")
  2. ... entre os estimadores que $E[\widetilde{\beta}] = \beta$ ("não viesado")

Distribuição Normal

- **Hipótese adicional**:

$$u|X \sim N(0, \sigma^2 I)$$

- Essa hipótese adicional caracteriza o que chamamos de **modelo de regressão linear clássico**.

- Ela implica que o estimador $\widehat{\beta}$ tem distribuição normal.

- **Por quê?**

$$\widehat{\beta} = \beta + (X'X)^{-1}X'u$$
$$u|X \sim N(0, \sigma^2 I) \Longrightarrow \widehat{\beta} \sim N(\beta, \sigma^2 (X'X)^{-1})$$