

Aula de Revisão

Econometria II · UFBA

Prof. Pablo Castro

Parte I — Revisão de Estatística

- Variáveis aleatórias e distribuições
- Operador esperança: definição e propriedades
- Operador variância: definição e propriedades
- Covariância e correlação
- Esperança condicional $\mathbb{E}(Y | X)$
- Lei dos Grandes Números e TLC
- Distribuição Normal

Parte II — MQO e a transição

- O modelo de regressão linear
- Hipóteses de Gauss-Markov
- A hipótese que quebra nas séries de tempo
- Motivação para Econometria II

PARTE I

Revisão de Estatística

Referência: Morettin & Bussab — Estatística Básica

Uma **variável aleatória** (v.a.) X é uma função que associa a cada resultado de um experimento aleatório um número real.

Discreta

- Assume valores enumeráveis:
 x_1, x_2, \dots
- Descrita pela função de probabilidade $P(X = x_i) = p_i$
- Exemplo: número de desempregados no mês

Contínua

- Assume valores em intervalos de \mathbb{R}
- Descrita pela função de densidade $f(x)$, com
$$P(a \leq X \leq b) = \int_a^b f(x) dx$$
- Exemplo: salário, taxa de câmbio, inflação

A **função de distribuição acumulada** é $F(x) = P(X \leq x)$.

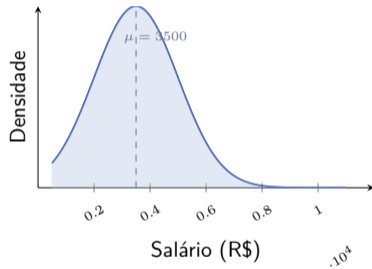
Exemplo — Variáveis Aleatórias

Discreta: lançamento de um dado

Face	$X(\omega)$	$P(X = x)$
■	1	1/6
■ ■	2	1/6
■ ■ ■	3	1/6
...
Soma		1

Cada face é igualmente provável. O resultado é aleatório, mas a distribuição é conhecida.

Contínua: distribuição de salários



$P(2000 \leq X \leq 5000) =$ área sob a curva nesse intervalo.

Operador Esperança

Definição

- Discreta: $\mathbb{E}(X) = \sum_i x_i p_i$
- Contínua: $\mathbb{E}(X) = \int_{-\infty}^{+\infty} x f(x) dx$

Propriedades (válidas para quaisquer v.a. X, Y e constantes a, b)

- $\mathbb{E}(b) = b$
- $\mathbb{E}(aX) = a \mathbb{E}(X)$
- $\mathbb{E}(aX + b) = a \mathbb{E}(X) + b$
- $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$ (linearidade, sem hipótese de independência)
- Se X e Y são independentes: $\mathbb{E}(XY) = \mathbb{E}(X) \mathbb{E}(Y)$

Observação. $\mathbb{E}(g(X)) \neq g(\mathbb{E}(X))$ em geral — a esperança não comuta com funções não lineares.

Exemplo — Operador Esperança

Dado honesto: $X \in \{1, 2, 3, 4, 5, 6\}$,
cada com $p = 1/6$

$$\mathbb{E}(X) = \sum_{x=1}^6 x \cdot \frac{1}{6} = \frac{21}{6} = 3,5$$

O valor 3,5 nunca ocorre — é o valor médio esperado após muitos lançamentos.

Aplicando $\mathbb{E}(aX + b) = a\mathbb{E}(X) + b$:
Ganho $W = 10X + 5$ (R\$ por ponto, mais taxa fixa):

$$\mathbb{E}(W) = 10 \times 3,5 + 5 = 40$$

Salários de 5 trabalhadores

Trabalhador	Salário (x_i)	p_i
Ana	R\$ 2.000	1/5
Bruno	R\$ 3.000	1/5
Carla	R\$ 3.500	1/5
Davi	R\$ 4.000	1/5
Eva	R\$ 7.500	1/5
Média	R\$ 4.000	

A Eva distorce a média. Isso ilustra como a esperança é sensível a valores extremos.

Operador Variância

Definição

$$\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}(X))^2] = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2$$

Propriedades

- $\text{Var}(b) = 0$
- $\text{Var}(aX) = a^2 \text{Var}(X)$
- $\text{Var}(aX + b) = a^2 \text{Var}(X)$
- $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2 \text{Cov}(X, Y)$
- Se X e Y são independentes: $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$

O desvio padrão é $\sigma(X) = \sqrt{\text{Var}(X)}$.

Por que isso aparece em Econometria II? A variância de um processo temporal $\{y_t\}$ precisa ser *constante* no tempo para que a série seja estacionária.

Exemplo — Operador Variância

Dois grupos de trabalhadores com a **mesma média salarial** (R\$ 3.000), mas **dispersão diferente**.

Grupo A — salários homogêneos

Trabalhador	Salário
1	R\$ 2.800
2	R\$ 3.000
3	R\$ 3.200
Média	R\$ 3.000
Var	26.667
DP	R\$ 163

Grupo B — salários heterogêneos

Trabalhador	Salário
1	R\$ 1.000
2	R\$ 3.000
3	R\$ 5.000
Média	R\$ 3.000
Var	2.666.667
DP	R\$ 1.633

Média idêntica, variância 100× maior no Grupo B. A variância captura o *risco* em torno da média que a média sozinha não revela.

Propriedade $\text{Var}(aX + b) = a^2\text{Var}(X)$: aumentar todos os salários em R\$ 500 (soma constante b) não altera a variância. Dobrar todos os salários ($a = 2$) quadruplica a variância.

Covariância e Correlação

Covariância

$$\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))] = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$

Propriedades da covariância

- $\text{Cov}(X, X) = \text{Var}(X)$
- $\text{Cov}(aX, bY) = ab \text{Cov}(X, Y)$
- $\text{Cov}(X + c, Y) = \text{Cov}(X, Y)$
- $X \perp Y \Rightarrow \text{Cov}(X, Y) = 0$ (a volta não vale em geral)

Correlação (covariância normalizada)

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma(X)\sigma(Y)}, \quad -1 \leq \rho \leq 1$$

Conexão com o curso. Em séries de tempo, a *autocovariância* de ordem k é $\gamma_k = \text{Cov}(y_t, y_{t-k})$ — mesma fórmula, mas aplicada à mesma série em dois instantes distintos.

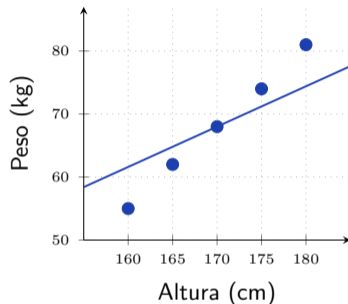
Exemplo — Covariância

Cinco alunos: $X =$ altura (cm), $Y =$ peso (kg).

Aluno	x_i	y_i	$x_i - \bar{x}$	$y_i - \bar{y}$
1	160	55	-10	-12
2	165	62	-5	-5
3	170	68	0	1
4	175	74	5	7
5	180	81	10	14

$\bar{x} = 170$ $\bar{y} = 68$

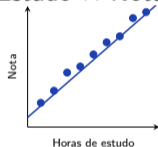
$$\widehat{\text{Cov}} = \frac{120 + 25 + 0 + 35 + 140}{5} = 64$$



Exemplo — Correlação: três casos

Positiva ($\rho \approx +1$)

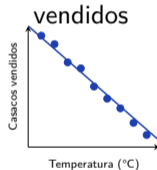
Estudo \times Nota



Mais estudo \rightarrow nota maior

Negativa ($\rho \approx -1$)

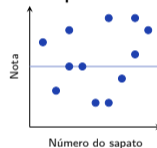
Temperatura \times Casacos



Mais calor \rightarrow menos casacos

Nula ($\rho \approx 0$)

N.º do sapato \times Nota



Nenhuma relação linear

Atenção: $\rho = 0$ não implica independência — apenas ausência de relação *linear*.

Esperança Condicional $\mathbb{E}(Y | X)$

Definição informal. A esperança de Y dado que observamos $X = x$ é

$$\mathbb{E}(Y | X = x) = \int y f_{Y|X}(y | x) dy$$

Propriedades

- **Lei das expectativas iteradas (L.E.I.):** $\mathbb{E}(Y) = \mathbb{E}[\mathbb{E}(Y | X)]$
- **Linearidade:** $\mathbb{E}(aY + b | X) = a\mathbb{E}(Y | X) + b$
- Se X e Y são independentes: $\mathbb{E}(Y | X) = \mathbb{E}(Y)$
- **Fatoração:** $\mathbb{E}(g(X)Y | X) = g(X)\mathbb{E}(Y | X)$

Relevância para o curso. A hipótese central do MQO é $\mathbb{E}(u | X) = 0$. Em séries de tempo, trabalharemos com $\mathbb{E}(y_t | y_{t-1}, y_{t-2}, \dots)$, a esperança condicional à história passada da série.

Exemplo — Esperança Condicional

Salário condicionado à escolaridade

Trabalhador	Educ. (X)	Salário (Y)
Ana	Médio	R\$ 2.000
Bruno	Médio	R\$ 2.500
Carla	Médio	R\$ 1.500
Davi	Superior	R\$ 5.000
Eva	Superior	R\$ 4.000
Fábio	Superior	R\$ 6.000

$$\mathbb{E}(Y \mid X = \text{Médio}) = \mathbf{2.000}$$

$$\mathbb{E}(Y \mid X = \text{Superior}) = \mathbf{5.000}$$

Verificando a lei das expectativas iteradas

Média geral calculada diretamente:

$$\mathbb{E}(Y) = \frac{2000 + 2500 + \dots + 6000}{6} = \mathbf{3.500}$$

Pela L.E.I., com

$$P(\text{Médio}) = P(\text{Superior}) = \frac{1}{2}:$$

$$\frac{1}{2} \cdot 2.000 + \frac{1}{2} \cdot 5.000 = \mathbf{3.500} \checkmark$$

Podemos calcular a média geral como uma média ponderada das médias de cada grupo.

Lei dos Grandes Números e Teorema Central do Limite

Lei dos Grandes Números (LGN). Sejam X_1, \dots, X_n i.i.d. com $\mathbb{E}(X_i) = \mu$. Então a média amostral converge em probabilidade para a média populacional:

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{p} \mu \quad \text{quando } n \rightarrow \infty$$

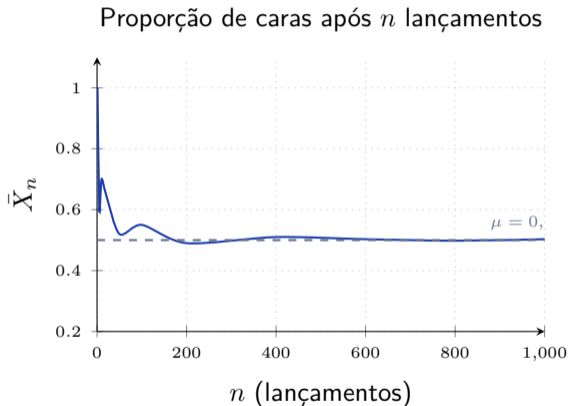
Teorema Central do Limite (TCL). Nas mesmas condições, com $\text{Var}(X_i) = \sigma^2 < \infty$:

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow{d} N(0, 1)$$

Por que aparecem em Econometria II?

- As propriedades assintóticas dos estimadores (MQO, MLE) dependem de versões da LGN e do TCL para *processos dependentes* — que é exatamente o caso das séries de tempo.
- Para séries estacionárias, resultados análogos valem, mas a variância assintótica precisa levar em conta a autocorrelação.

Exemplo — Lei dos Grandes Números



$X_i = 1$ (cara) ou $X_i = 0$ (coroa),
com $\mathbb{E}(X_i) = 0,5$.

n	\bar{X}_n
1	1,000
5	0,600
10	0,700
50	0,520
200	0,490
1000	0,503

Os desvios iniciais se diluem. Com n
grande, $\bar{X}_n \rightarrow 0,5$.

A Distribuição Normal

Definição. $X \sim N(\mu, \sigma^2)$ tem densidade

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right)$$

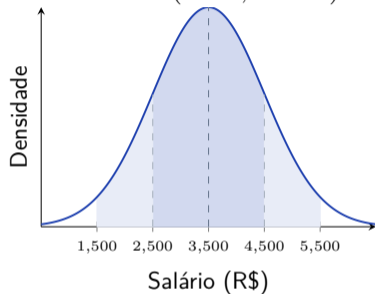
Propriedades úteis para o curso

- $\mathbb{E}(X) = \mu, \quad \text{Var}(X) = \sigma^2$
- Se $X \sim N(\mu, \sigma^2)$, então $aX + b \sim N(a\mu + b, a^2\sigma^2)$
- Combinações lineares de normais são normais
- Padronização: $Z = (X - \mu)/\sigma \sim N(0, 1)$
- Se $X \sim N(0, 1)$: $X^2 \sim \chi^2(1)$
- Razão normal/raiz-qui-quadrado gera a distribuição t de Student

Em econometria, assume-se com frequência $u | X \sim N(0, \sigma^2 I)$, o que torna os estimadores MQO normalmente distribuídos em amostras finitas.

Exemplo — Distribuição Normal

Salários $\sim N(3500, 1000^2)$ — R\$



Regra empírica (68–95–99,7)

Intervalo	Probabilidade
$[\mu - \sigma, \mu + \sigma]$	68%
$[\mu - 2\sigma, \mu + 2\sigma]$	95%
$[\mu - 3\sigma, \mu + 3\sigma]$	99,7%

Com $\mu = 3500$, $\sigma = 1000$:

- 68% ganham entre R\$ 2.500 e R\$ 4.500
- 95% ganham entre R\$ 1.500 e R\$ 5.500

Padronização:

$$Z = (X - 3500)/1000 \sim N(0, 1).$$

PARTE II

MQO e a Transição para Séries de Tempo

O Modelo de Regressão Linear

O modelo populacional é

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + u_i, \quad i = 1, \dots, n$$

ou, em notação matricial, $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$.

O estimador de Mínimos Quadrados Ordinários (MQO) minimiza $\sum_i \hat{u}_i^2$:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

Configuração típica em Econometria I (cross-section). O índice i representa unidades distintas observadas *num mesmo instante*: trabalhadores, firmas, municípios. As observações são assumidas independentes entre si.

Econometria II. O índice passa a ser t , representando o *tempo*. A mesma unidade é observada em $t = 1, 2, \dots, T$. Essa mudança viola hipóteses do modelo clássico.

Hipóteses de Gauss-Markov

- H1 Linearidade:** o modelo é linear nos parâmetros β .
- H2 Amostra aleatória:** as observações (y_i, \mathbf{x}_i) são i.i.d. (implica independência entre i e j).
- H3 Posto completo:** \mathbf{X} tem posto coluna $k + 1$ (sem multicolinearidade perfeita).
- H4 Exogeneidade estrita:** $\mathbb{E}(u_i | \mathbf{X}) = 0$ para todo i .
- H5 Homocedasticidade:** $\text{Var}(u_i | \mathbf{X}) = \sigma^2$ constante.
- H6 Não-autocorrelação:** $\text{Cov}(u_i, u_j | \mathbf{X}) = 0$ para $i \neq j$.

Sob H1–H6, o estimador $\hat{\beta}$ é **BLUE** (*Best Linear Unbiased Estimator*) — Teorema de Gauss-Markov.

A Hipótese que Quebra nas Séries de Tempo

Em dados **cross-section**, H_6 é plausível: o erro de salário de um trabalhador não tem razão óbvia para estar correlacionado com o de outro.

Em **séries de tempo**, a observação de hoje y_t tipicamente carrega informação da de ontem y_{t-1} . Por isso:

$$\text{Cov}(u_t, u_{t-s}) \neq 0 \quad \text{para algum } s \neq 0$$

Exemplo: IPCA mensal (Brasil)

Mês	IPCA (%)
Janeiro	0,42
Fevereiro	0,83
Março	1,32
Abril	0,61
Maio	0,46
Junho	0,21

Se a inflação de março foi 1,32%, ela reflete pressões de custos, expectativas e inércia que *não se dissipam* até abril.

Ou seja: y_{abr} depende de y_{mar} . Os erros de um modelo que ignora isso serão correlacionados ao longo do tempo:

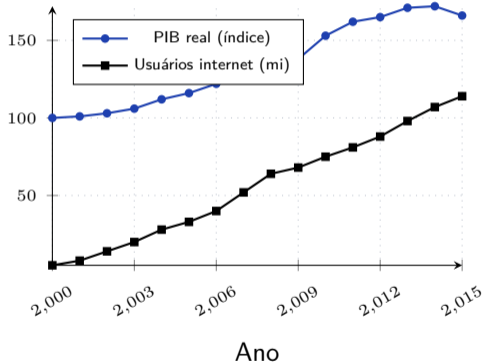
$$\text{Cov}(u_{abr}, u_{mar}) \neq 0$$

Consequências de ignorar a autocorrelação dos erros

- $\hat{\beta}$ permanece não viesado (H4 pode ainda valer).
- A fórmula usual de $\text{Var}(\hat{\beta})$ fica errada — erros padrão são inconsistentes.
- Testes t e F clássicos perdem validade.
- Risco de **regressão espúria**: duas séries sem relação real podem apresentar R^2 alto e t -stats elevados por ambas terem tendência ou memória de longo prazo.

Exemplo — Regressão Espúria

PIB real do Brasil × Usuários de internet (2000–2015)



Rodando a regressão

$$\text{PIB}_t = \alpha + \beta \text{Internet}_t + u_t$$

Estatística	Valor
$\hat{\beta}$	1,37***
R^2	0,97
t-stat	21,4

Parece que internet “explica” 97% da variação do PIB. Mas ambas as séries simplesmente *crecem no tempo* — a correlação é um artefato da tendência comum, não de uma relação causal.

O que muda quando o índice é o tempo?

- As observações não são mais independentes: y_t depende de y_{t-1}, y_{t-2}, \dots
- A ordem das observações importa — não podemos reordená-las livremente.
- A variável pode ter *tendência* (média não constante) ou *variância crescente*, violando a estacionariedade.

Pergunta central. Dada a sequência y_1, y_2, \dots, y_T , como modelar a dependência temporal e fazer inferência?

Dúvidas?

Próxima aula: Processos Estacionários e Não Estacionários

Referência: Notas de Aula