Econometria

prof. Danielle Carusi Machado

Aula 1

PPGE/UFF

1o. Semestre de 2023

Econometria

- "[...] a econometria pode ser definida como a análise quantitativa dos fenômenos econômicos ocorridos com base no desenvolvimento paralelo da teoria e das observações e com o uso de métodos de inferência adequados." (Samuelson, P. A.; Koopmans, T. C.; Stone, J. R. N. *Report of the evaluative committee for econometrica. Econométrica*.Abr. 1954, v. 22, n. 2)
- "A econometria diz respeito à determinação empírica das leis econômicas." (Theil, H. *Principles of econometrics.* Nova York: John Wiley & Sons, 1971).
- "A econometria pode ser definida como a ciência social em que as ferramentas da teoria econômica, da matemática e da inferência estatística são aplicadas à análise dos fenômenos econômicos." (Goldberger, Arthur S. *Econometric theory.* Nova York: John Wiley & Sons, 1964.)

Utilidade de Econometria

- Instrumental para analisar:
  - Fenômenos econômicos;
  - Fenômenos sociais e políticos;
  - Testar teorias;
  - Produzir conhecimento empírico.
- Basicamente, queremos entender relações entre variáveis.

Utilidade da Econometria

- A teoria econômica faz hipóteses de natureza qualitativa.

- Exemplos:
  - Se a renda das famílias aumenta, o consumo de determinados bens aumenta.
  - Se o preço de um bem aumenta, a quantidade demandada do bem irá cair.

Utilidade da Econometria

- A teoria econômica, por exemplo, postula uma relação negativa entre o preço de um bem e sua quantidade demandada.
- Não fornece, contudo, uma medida quantitativa da relação entre as duas variáveis.
- O econometrista faz esta estimativa numérica.
- Ele calcula quanto a quantidade aumentará ou diminuirá com a variação do preço.
- Econometria: dá conteúdo prático a teoria econômica.

Porque uma disciplina separada?

- Econometria é diferente de Estatística e Matemática.
- Precisamos de métodos especiais para estimar relações econômicas dada a natureza da maior parte dos dados econômicos.
- Dados Observacionais (não experimentais) X Dados experimentais (dados controlados diretamente).
- Em Economia, trabalhamos com dados observacionais usualmente.

Dados experimentais

- São produzidos em ambientes laboratoriais.
- Em um experimento tradicional, o pesquisador participa ativamente do processo de geração dos dados.
- Nas ciências sociais é mais difícil obter dados deste tipo (viabilidade, custo financeiro ou objeções morais).
- Mais comuns recentemente: experimentos aleatórios

https://economics.mit.edu/sites/default/files/publications/experimental%20approach.pdf

Dados observacionais

- Dados que são coletados a partir de uma *survey,* de registros administrativos ou produtos de uma atividade (*big data*).
- São usualmente restropectivos, sendo coletados depois do evento ter acontecido.
- Algumas técnicas recentes (*web scraping*) permitem que o dado coletado seja o mais próximo possível do momento exato que a ação ocorreu.
- O pesquisador é um agente passivo ao processo de geração do dado.
- O pesquisador observa ações/resultados mas não interfere no ambiente onde isto acontece.

Relações causais

- Com dados observacionais, as correlações não refletem necessariamente relações causais.

- Relações Causais e análise *ceteris paribus*
  - *1 ano a mais de educação em quanto aumenta o salário mensal?*
  - *Reduzir o tamanho da classe aumenta a performance do aluno na escola?*
  - *Reduzir imposto aumenta a atividade econômica?*

Relações causais

- **Ceteris Paribus:** mantendo todos os demais fatores fixos (fatores relevantes) – *crucial para estabelecer relações causais.*
- ***Correlação*** *é diferente de causalidade.*
- ***Métodos econométricos:*** *outros fatores fixos – inferir causalidade.*

Correlação vs. Causalidade

- Pode existir correlação entre duas variáveis, contudo, uma variável não necessariamente tem efeito causal sobre a outra. Ex: o sol nasce quando o galo canta.... Mas o galo não causa o sol nascer...
- Pode não existir correlação entre duas variáveis, mas uma pode ter efeito causal sobre a outra.

Relações entre as variáveis

- Relações determinísticas: podem ser estabelecidas pela teoria. → Matemática

- Relações não determinísticas: estocásticas (função de probabilidade conjunta entre as variáveis). → Estatística e Econometria

- Relações estocásticas podem ser usadas para testar teorias! → Econometria

Metodologia econométrica tradicional

- Exposição da teoria ou hipótese
- Exposição do **modelo** matemático da teoria/hipótese
- Exposição do **modelo** estatístico econométrico
- Obtenção dos dados
- Estimação dos parâmetros
- Testes de hipóteses
- Previsão/projeção
- Uso do modelo para fins de política.

Modelo teórico

- Descreve **matematicamente uma relação**, um comportamento e um processo de interesse.
- Estrutura que descreve as relações pelas quais estamos interessados.
- Fornece a intuição dos resultados e as hipóteses testáveis.
- É uma construção teórica abstrata, que representa de forma simples a realidade.

Modelo econométrico

- Pode ser estimado diretamente a partir dos dados.
- Estabelecer a forma functional: como as variáveis se relacionam.
- Como escolher as variáveis que serão *proxy* das variáveis teóricas.
- E o que fazer na inexistência de variáveis – formular hipóteses.

Relações entre variáveis

- Foco: na média, ou melhor, na **resposta média esperada.**

$$E\left(y / w, c\right)$$

- Vetor *c : conjunto de variáveis de controle que estarão explicitamente fixas quando estudamos o efeito de w em y.*
- *O vetor w é correlacionado com outros fatores que também influenciam y.*

Relações entre variáveis

- Se w é contínuo, o interesse recai em:

$$\frac{\partial E(y/w,c)}{\partial w}$$

  - Efeito parcial de w em $E(y/w,c)$

- Se w é discreto, o interesse recai em:

  - $E(y/w,c)$ valorado para diferentes valores de w

Relação entre variáveis

- Quais são os controles?
  - Muitos elementos de *c* não são observáveis.

Exemplo: $E(y/w,c)$

  - Y – salário
  - W – anos de estudo
  - C – *habilidade* e experiência
- Não é possível obter dados de todos controles desejados!

Relação entre variáveis

- Outros problemas que interferem na estimação da relação causal:

  - Podemos não ter boas medidas para y e w (erro de medida)

  - Podemos observar apenas valores de equilíbrio de y e w (simultaneidade).

Como escolher a relação entre Y e W e c?

- Construção das hipóteses do modelo que explica a relação populacional entre as variáveis – **forma funcional**.
- Olhar os dados e inferir algumas hipóteses sobre o modelo que explicaria a geração de dados.
- Identificar a relação funcional entre as variáveis Y e W, c.

O Modelo de Regressão Linear Múltipla

- Utilizado para estudar a relação entre uma variável dependente e uma ou mais variáveis independentes.
- Forma genérica do modelo de regressão linear:

$$y = f(x_1, x_2, \ldots, x_K, \beta_1, \beta_2, \ldots \beta_K) + \varepsilon$$
$$= x_1\beta_1 + x_2\beta_2 + \ldots + x_K\beta_K + \varepsilon$$

- $f(x_1, x_2, \ldots, x_K, \beta_1, \beta_2, \ldots \beta_K)$ é a equação de regressão populacional de y em $x_1, x_2, \ldots, x_K$.
- Y é o regressando
- $x_1, x_2, \ldots, x_K$ regressores ou controles
- **ε é o distúrbio aleatório**

Betas são parâmetros

Erros de medida, variáveis omitidas

Exemplo da função de consumo keynesiana

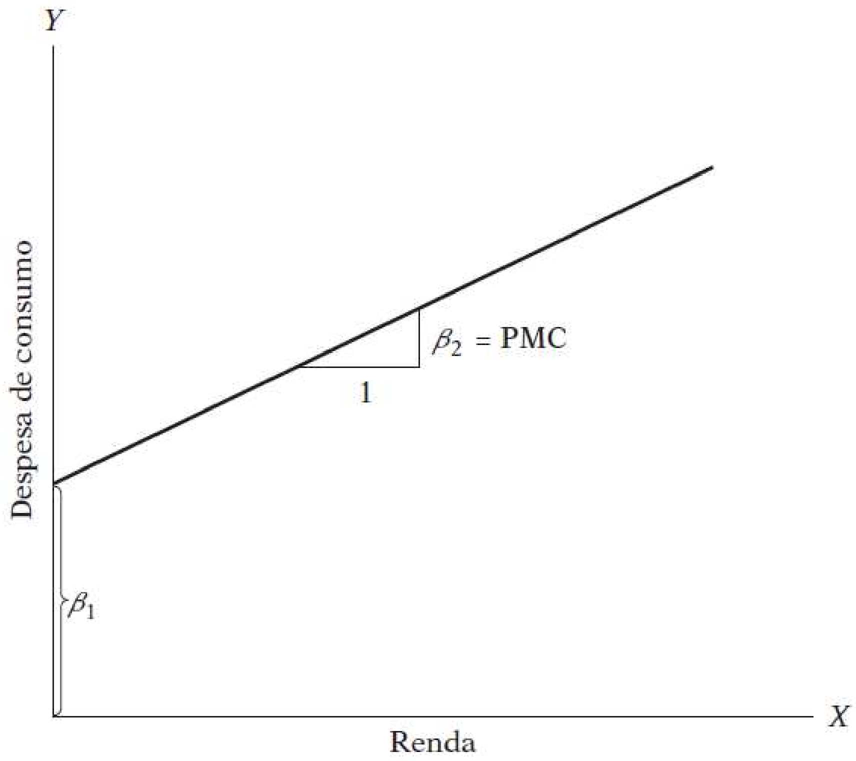
- “A lei psicológica fundamental é que os homens [as mulheres] estão dispostos, como regra e em média, a aumentar seu consumo conforme sua renda aumenta, mas não na mesma proporção que o aumento na renda.” (Keynes, J.M. Teoria Geral, 1936).

- A **propensão marginal a consumir (PMC)**, a taxa de variação do consumo por variação de uma unidade (digamos, um dólar) de renda, é maior que zero, mas menor que 1.

Exemplo da função de consumo keynesiana

- Função de consumo keynesiana
  - $C = \beta 1 + \beta 2X$ → **Modelo teórico**
  - Onde $0<\beta 2<1$ → **Hipótese sobre PMC**
  - C: despesas com consumo
  - X: renda
  - $\beta 1$ e $\beta 2$ são parâmetros do modelo, intercepto e coeficiente angular, respectivamente.

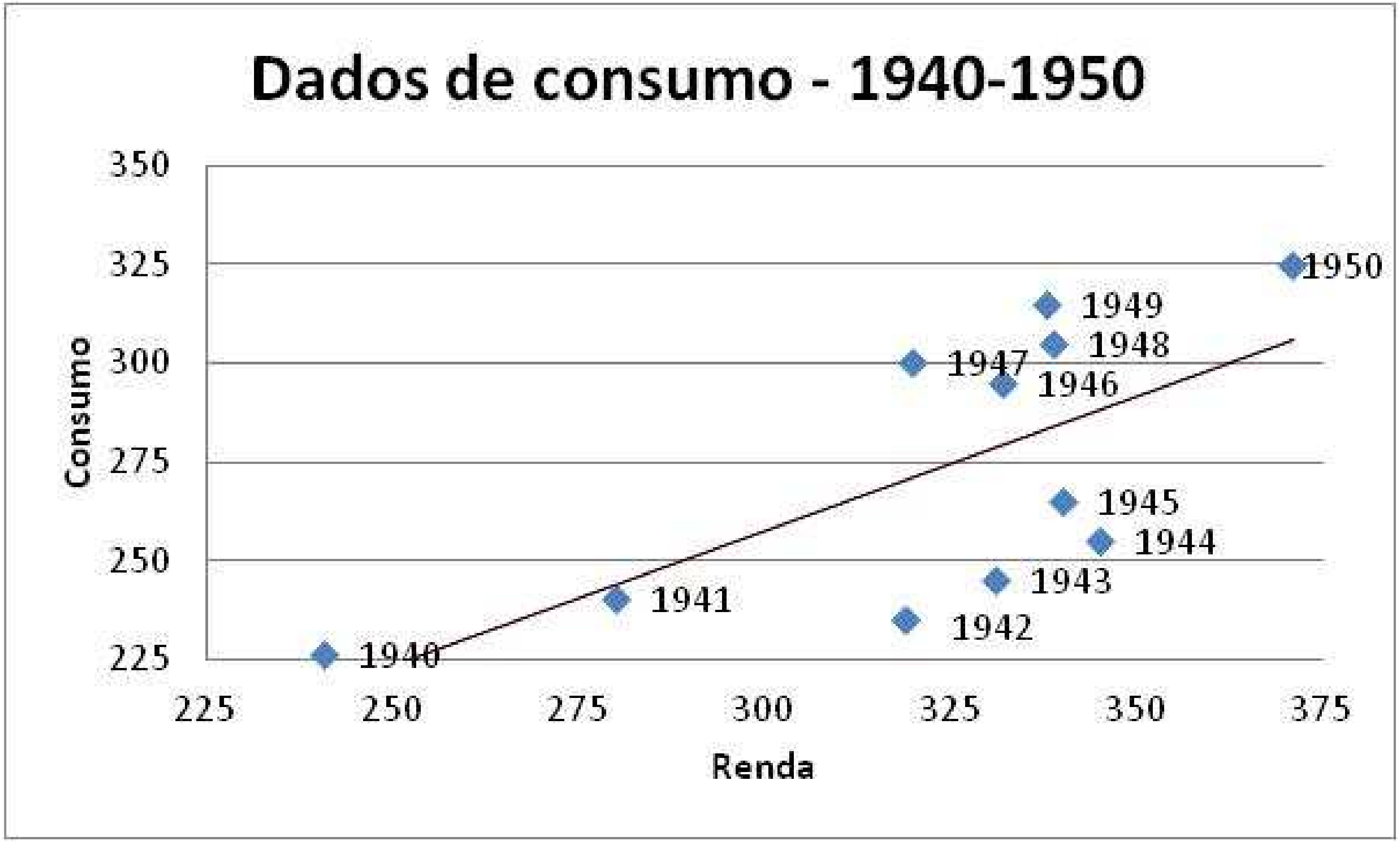
Representação gráfica



Exemplo da função de consumo keynesiana

- Função de consumo keynesiana
  - Não existe uma relação determinística entre consumo e renda.
  - $C = f(X, \varepsilon)$
    - Onde ε é o elemento estocástico
    - Como incorporar este elemento estocástico ao modelo? De forma aditiva:
    - $C = \beta 1 + \beta 2X + \varepsilon$
    - **Contrapartida empírica** do modelo teórico de Keynes.

Exemplo



Dados de consumo - 1940-1950
Consumo
Renda
350
325
300
275
250
225
225
250
275
300
325
350
375
1940
1941
1942
1943
1944
1945
1946
1947
1948
1949
1950

Exemplo

- A reta do gráfico anterior é distorcida pelo racionamento do período de guerra.
- Especificação mais apropriada: acomodar a natureza estocástica do dado e as circunstâncias especiais dos anos 1942-1945.
- Dummy que identifica este período

$$C = \beta_1 + \beta_2 X + \gamma_w d_{anoguerra} + \varepsilon$$

Estimando o modelo de consumo

Variável dependente Consumo	(1)	(2)
	mqo1	mqo2
Renda	0.685**	0.858***
	(0.249)	(0.0853)
Dummy anos de Guerra		-50.69***
		(5.932)
Constant	51.90	14.50
	(80.84)	(27.30)
Observations	11	11
R-squared	0.457	0.946

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Alguns conceitos de estatística

Variável aleatória e experimentos

- **Variável aleatória**: assume valores numéricos e seus resultados são determinados por um experimento.
- **Experimento aleatório (Processo aleatório)**: procedimento que pode ser repetido e tem um conjunto de resultados possíveis e bem definidos.
- **Exemplos:**

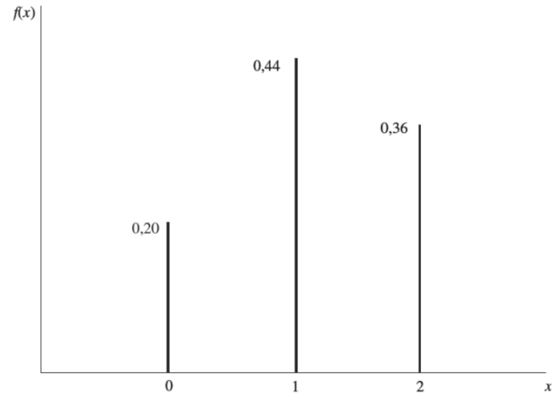
Experimento: jogamos uma moeda para cima dez vezes e contamos o número de vezes que aparece coroa.

Variável aleatória: número de vezes que aparece coroa.

Tipos de variáveis aleatórias

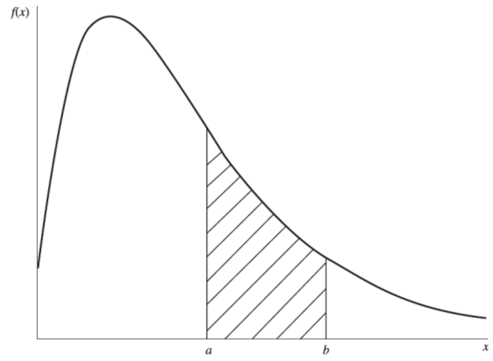
- **Variável aleatória discreta**:

possuem um número finito de valores.



- **Variável aleatória contínua**:

assume tantos valores possíveis que não podemos enumerá-los.



Função densidade de probabilidade

- Função densidade de probabilidade **marginal**: resume as informações relativas aos possíveis valores de X e suas probabilidades correspondentes.

$$f\left(X = x_j\right) = p_j$$

- **Função de distribuição de probabilidade acumulada**:

$$F(X = x_j) \equiv P(X \leq x_j)$$

Função de distribuição acumulada

- Para variáveis discretas: somar todas as probabilidades abaixo dos possíveis valores de xj (X<xj).
- Para variáveis contínuas: F(x) é a área abaixo da curva de densidade de probabilidade à esquerda do ponto xj.
- F(x) é uma probabilidade, logo, sempre estará entre 0 e 1.
- F(x) é uma função não decrescente de x.

Função de distribuição acumulada

- *Para qualquer número* c:

$$P(X > c) = 1 - P(X \leq c) = 1 - F(c)$$

- Para q□aisq□er números a e b (a<b):

$$P(a < X \leq b) = F(b) - F(a)$$

- Em Econometria, usualmente, para variáveis contínuas:

$$P(X > c) = P(X \geq c)$$

$$P(a < X < b) = P(a \leq X \leq b) = P(a \leq X < b)$$
$$= P(a < X \leq b)$$

OBS: as principais funções de distribuições acumuladas são tabuladas (normal, t-student, F, Quiquadrada).

Funções de probabilidades conjunta e condicional

- Normalmente, estamos interessados em ocorrências de eventos que envolvem mais de uma variável aleatória.
- **Função densidade de probabilidade conjunta**

$$f_{xy}\left(X = x_j, Y = y_j\right)$$

- **Função de densidade de probabilidade condicional**

$$f_{Y|X}\left(Y|X = x_j\right) = \frac{f_{xy}}{f_x}$$

Função de distribuição condicional

- Em Econometria, usualmente, estamos interessados em entender o comportamento de uma variável aleatória Y, dado o comportamento de uma ou mais variáveis X.

- Logo, a função de distribuição condicional é muito importante!!

Independência entre variáveis aleatórias

- Para duas variáveis aleatórias independentes, temos que a função de distribuição conjunta das variáveis X e Y é igual a multiplicação das respectivas funções de densidade marginais de X e Y:

$$f_{xy}(x,y) = f_x f_y$$

Com X e Y independentes, conhecer o resultado de X não muda a probabilidade de possíveis valores de Y.

Alguns aspectos das funções de distribuição

- **Esperança**: Se X é uma variável aleatória, o valor esperado de X é uma média ponderada de todos possíveis valores de X $(x_1, x_2, \ldots, x_k)$. Os pesos são dados pela probabilidade de cada valor na população $(f(x_1), \ldots, f(x_k))$.

$$\mathrm{E}(X) = x_1 f(x_1) + x_2 f(x_2) + \ldots + x_k f(x_k) \equiv \sum_{j=1}^{k} x_j f(x_j).$$

- Representa a tendência central de uma variável aleatória.

$$E(X) = \mu$$

$$\mathrm{E}(X) = \int_{-\infty}^{\infty} x f(x) dx, \longrightarrow \text{Se X é contínuo}$$

Propriedades da Esperança

$$E[g(X)] = \sum_{j=1}^{k} g(x_j) f_X(x_j) \qquad E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)\,dx.$$

$$E(c) = c. \qquad E(aX + b) = aE(X) + b.$$

$$E(a_1X_1 + a_2X_2 + \ldots + a_nX_n) = a_1E(X_1) + a_2E(X_2) + \ldots + a_nE(X_n).$$

$$E\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i E(X_i). \qquad E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i).$$

Exemplo

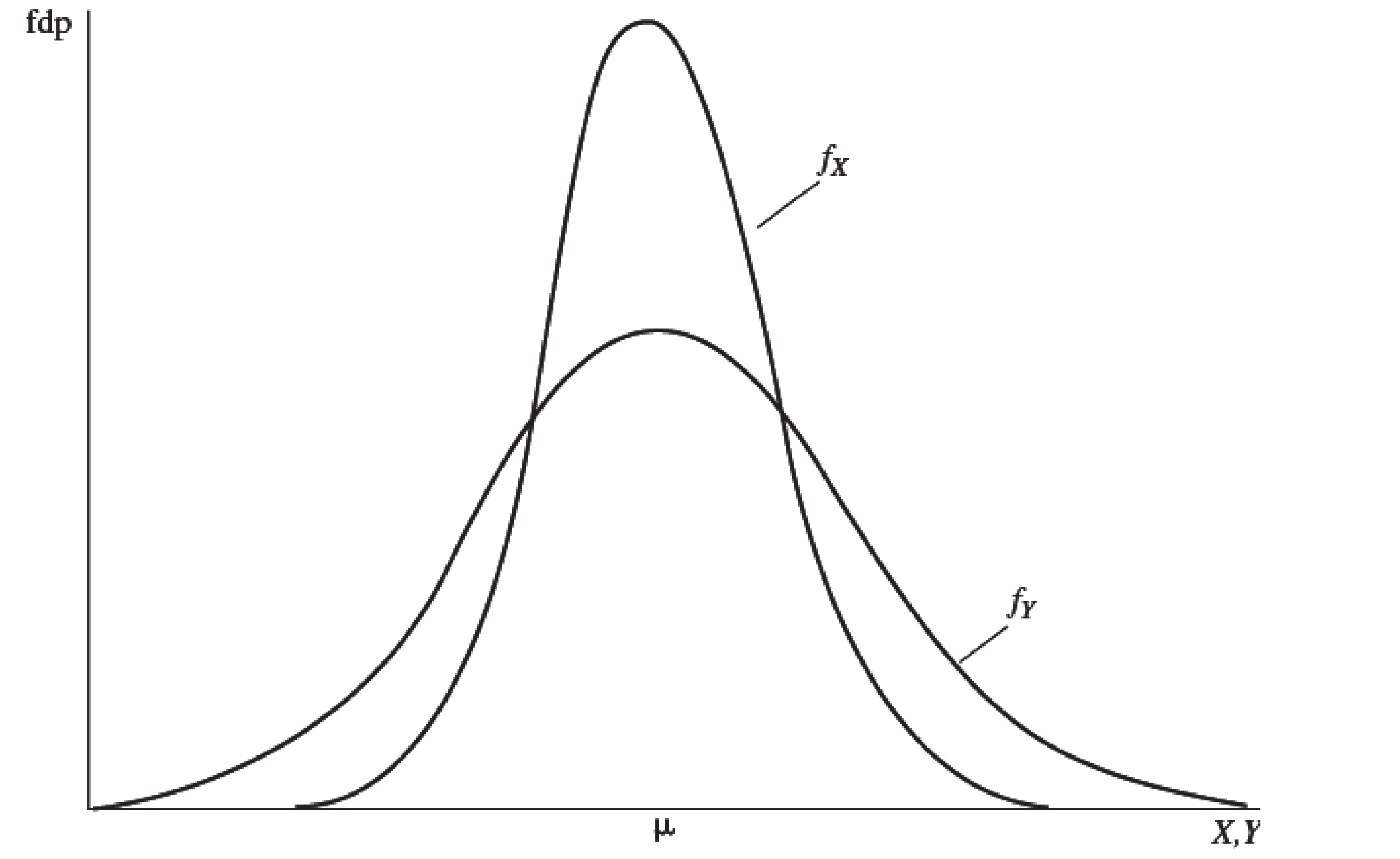
- Suponha que X1, X2 e X3 são os totais de pizzas pequenas, médias e grandes vendidas durante algum dia numa pizzaria.
- Estas variáveis aleatórias tem os seguintes valores esperados e preços:
- $E(X_1) = 25$, R$ 5,50
- $E(X_2) = 57, R\$ \, 7{,}60$
- $E(X_3) = 40, R\$ \, 9{,}15$
- Qual a receita esperada com a venda das pizzas?

Medidas de dispersão

- **Variância e desvio padrão**: Retrata a dispersão dos valores em torno da média. Representa uma média que nos diz o quão distante os valores de X estão da média.

$$var(X) = E(X-\mu)^2 = \sigma^2$$
$$dp(X) = +\sqrt{E(X-\mu)^2} = \sigma$$

Alguns conceitos importantes!



Propriedades da variância

- $var(X) = 0$ s.s.s existe uma constante c tal que P(X=c)=1.

- Para qualquer constante a e b:

$$var(aX + b) = a^2 var(X)$$

$$var(aX + bY)$$
$$= a^2 var(X) + b^2 var(Y) + 2abcov(XY)$$

Covariância

- Medida resumida da relação linear entre duas variáveis X e Y (tem a ver com o grau de dependência linear entre duas variáveis.

$$cov(X,Y) = E[(X-\mu_X)(Y-\mu_Y)] = \sigma_{XY}$$

$$cov(X,Y) = E(XY) - E(X)E(Y)$$

- Se $\sigma_{XY} > 0$, em média, quando X estiver acima da média, Y também estará, e vice-versa quando $\sigma_{XY} < 0$.

Propriedades da covariância

- Se X e Y são independentes: cov(X,Y)=0

- Para quaisquer constantes a, b, c, d:

$$cov(aX + b, cY + d) = ac.\,cov(X, Y)$$

Coeficiente de correlação

- **Coeficiente de Correlação:** mostra a relação entre duas variáveis que não depende das suas respectivas unidades de medida (ex. educação em anos de estudos e salário, em reais):

$$corr(X,Y) = \frac{cov(X,Y)}{dp(X)dp(Y)} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

$$-1 \leq corr(X,Y) \leq 1$$

Esperança condicional

- **Esperança condicional**

Covariância e correlação medem a relação linear entre duas variáveis aleatórias e as tratam de forma simétrica.

Se X assume um valor particular, podemos calcular o valor médio de Y dado este valor de X:

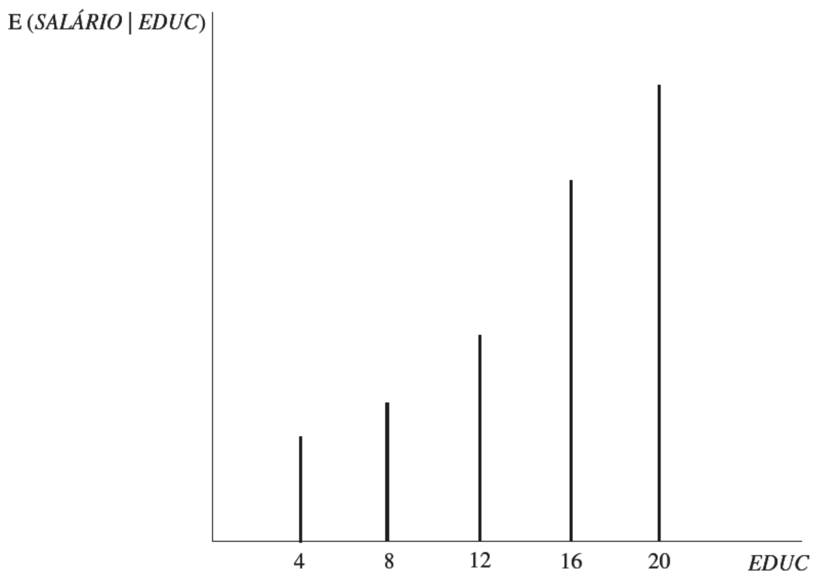
$$E(Y|X = x)$$

$$\mathrm{E}(Y|x) = \sum_{j=1}^{m} y_j f_{Y|X}(y_j|x).$$

Esperança condicional

- Gostaríamos de saber com a distribuição de salários muda com o nível educacional. Podemos sumarizar esta distribuição utilizando o conceito de esperança condicional.

O valor esperado do salário por hora considerando vários níveis de educação.



Propriedades da Esperança condicional

- Se X e Y são independentes: $E(X|Y)=E(X)$

- $E(c(X)|X)=c(X)$ para qq função $c()$

- $E(a(X)Y+b(X).X|X) = a(X)E(Y|X)+b(X)X$

- Lei das expectativas iteradas:

$$E(E(Y|X))=E(Y)$$

Operadores de somatórios

Operador de somatório

- Expressões que envolvem a soma de diversos números.

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + \cdots + x_n$$

- O subscrito i apenas representa um valor específico da variável aleatória x.
- Propriedades:

$$\sum_{i=1}^{n} c = nc. \qquad \sum_{i=1}^{n} cx_i = c\sum_{i=1}^{n} x_i.$$

$$\sum_{i=1}^{n} (ax_i + by_i) = a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} y_i. \qquad \sum_{i=1}^{n} (x_i/y_i) \neq \left(\sum_{i=1}^{n} x_i\right) \Big/ \left(\sum_{i=1}^{n} y_i\right). \qquad \sum_{i=1}^{n} x_i^2 \neq \left(\sum_{i=1}^{n} x_i\right)^2.$$

Operador de somatório

- Considerando n valores para x_i:

- Média: $\bar{x} = (1/n)\sum_{i=1}^{n} x_i.$

- Desvios em relação à média:

$$\sum_{i=1}^{n} d_i = \sum_{i=1}^{n} (x_i - \bar{x}) = \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \bar{x} = \sum_{i=1}^{n} x_i - n\bar{x} = n\bar{x} - n\bar{x} = 0$$

$$\sum_{i=1}^{n} (x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - n(\bar{x})^2.$$

Operador de somatório

- Considere uma sequência de dois números:

$\{y_1, y_2, \dots, y_m\}$ e $\{x_1, x_2, \dots, x_n\}$, e o caso em que n=m=2:

$$\sum_{i=1}^{2} x_i \sum_{j=1}^{2} y_j = \sum_{i=1}^{2} x_i \,(y_1 + y_2) = x_1\,(y_1 + y_2) + x_2(y_1 + y_2) = \sum_{i=1}^{2} \sum_{j=1}^{2} x_i \;\; y_j$$

Operador de somatório

- Correlação usa este somatório:

$$\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})=\sum_{i=1}^{n}x_i(y_i-\bar{y})$$

$$=\sum_{i=1}^{n}(x_i-\bar{x})y_i=\sum_{i=1}^{n}x_iy_i-n(\bar{x}\cdot\bar{y});$$

Econometria e esperança condicional

Exemplo

$$E(SALÁRIO|EDUC) = 1{,}05 + 0{,}45\ EDUC.$$

- O valor esperado do salário pode ser calculado para cada nível de escolaridade.
- Em econometria, especifica-se a função que relaciona a variável Y com a variável X ou mais de uma variável X.
- Especificação geral:

Variáveis explicativas

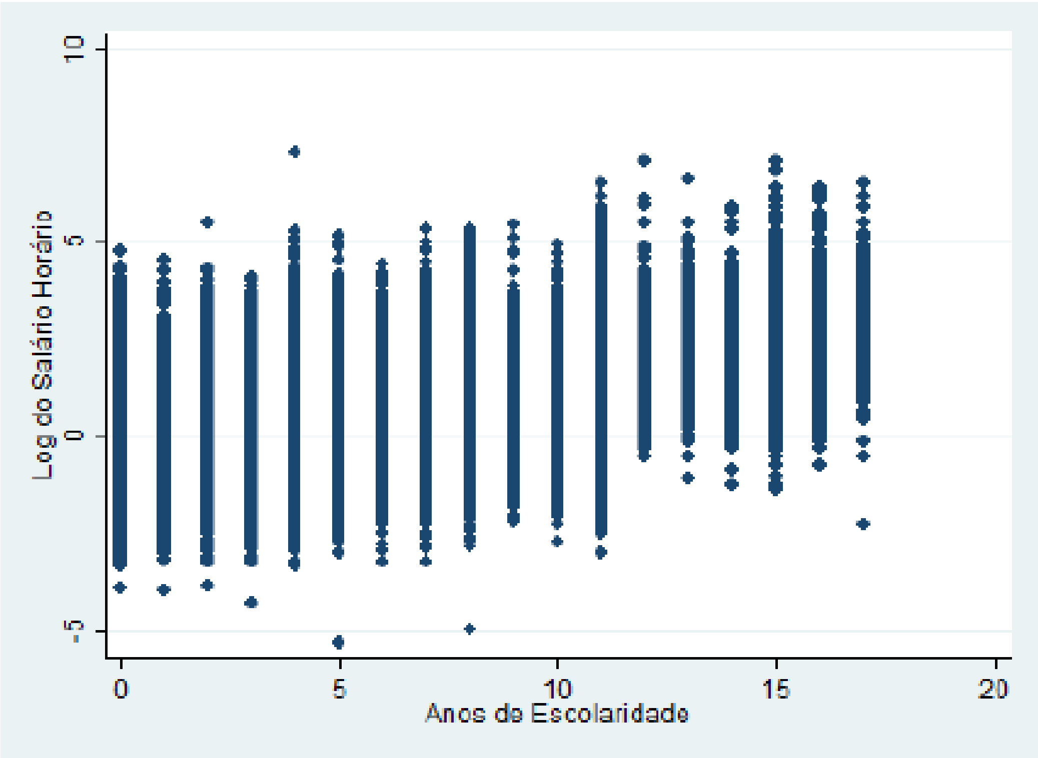
Parâmetros

Termo de erro ou distúrbios aleatórios

$$y = f(x_1, x_2, \ldots, x_K, \beta_1, \beta_2, \ldots \beta_K) + \varepsilon$$
$$= x_1\beta_1 + x_2\beta_2 + \ldots + x_K\beta_K + \varepsilon$$

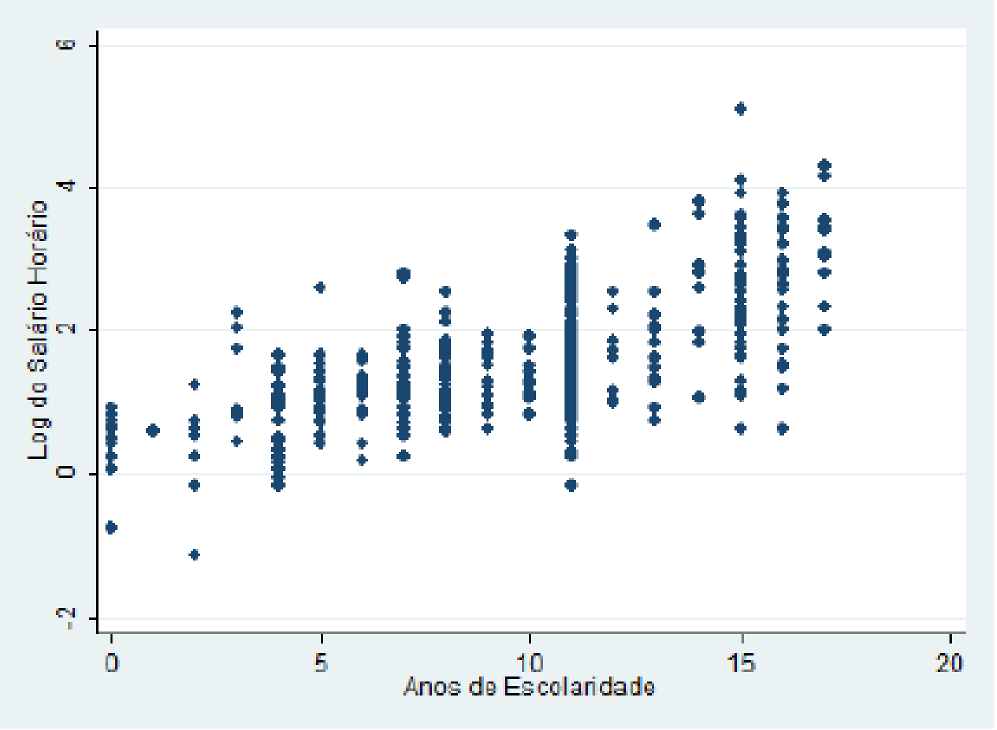
Outro exemplo...

Brasil, amostra de 167.117 trabalhadores de acordo com a Pnad/IBGE. Para um dado nível de educação, encontramos uma distribuição de salários...



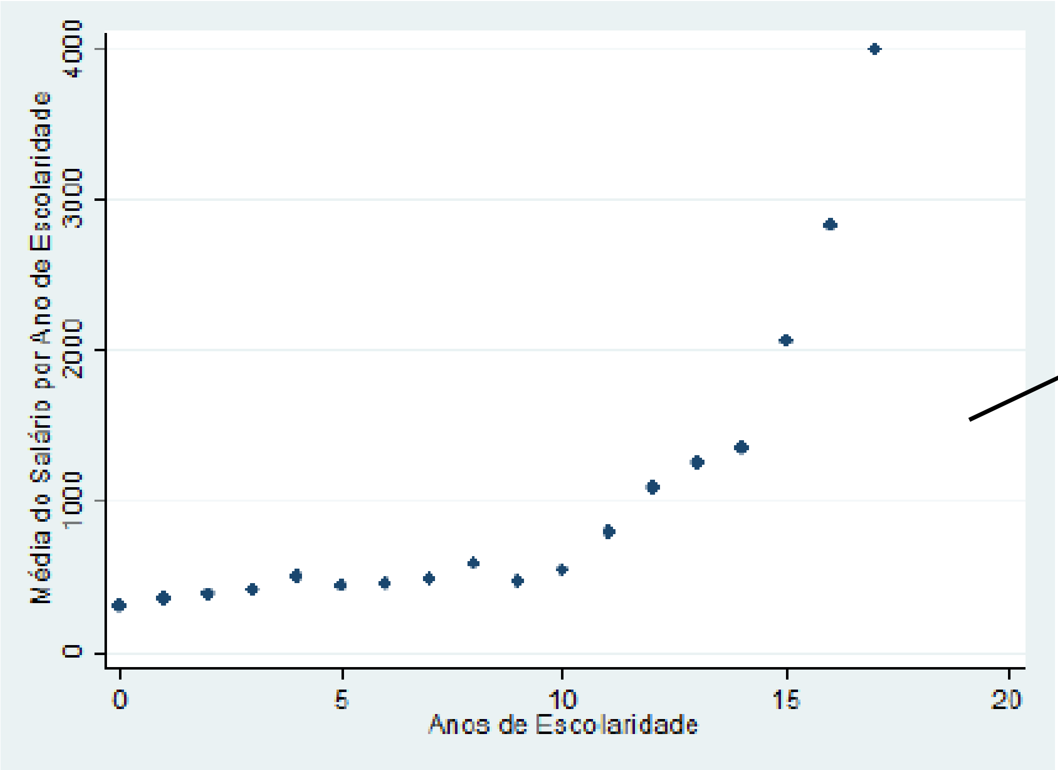
Refinar um pouco para entender os dados...

Brasil, amostra de 488 trabalhadores homens, brancos, urbanos, idade=30



Tentamos entender a média para um dado X que é escolaridade...

Brasil, amostra completa: média de salários por ano de escolaridade



Valor esperado de salário dado o valor de escolaridade