



# Métodos Quantitativos para Inferência Causal

**28 de Março, 2025**

**International Growth Centre**

# Plano



- 1.** A diferença entre correlação e causalidade
- 2.** A aplicação de inferência causal a problemas importantes de políticas públicas
- 3.** Revisão de terminologia e estatística
- 4.** Exemplos de Pesquisa

# Correlação vs Causalidade



## **O que é uma correlação?**

Duas variáveis que se movimentam juntas

## **Quando é que é importante identificar uma correlação?**

Se não conseguimos detetar uma correlação, então é improvável que haja uma relação causal

Pode ajudar a prever o movimento das duas variáveis

## **Quando é que uma correlação não é suficiente?**

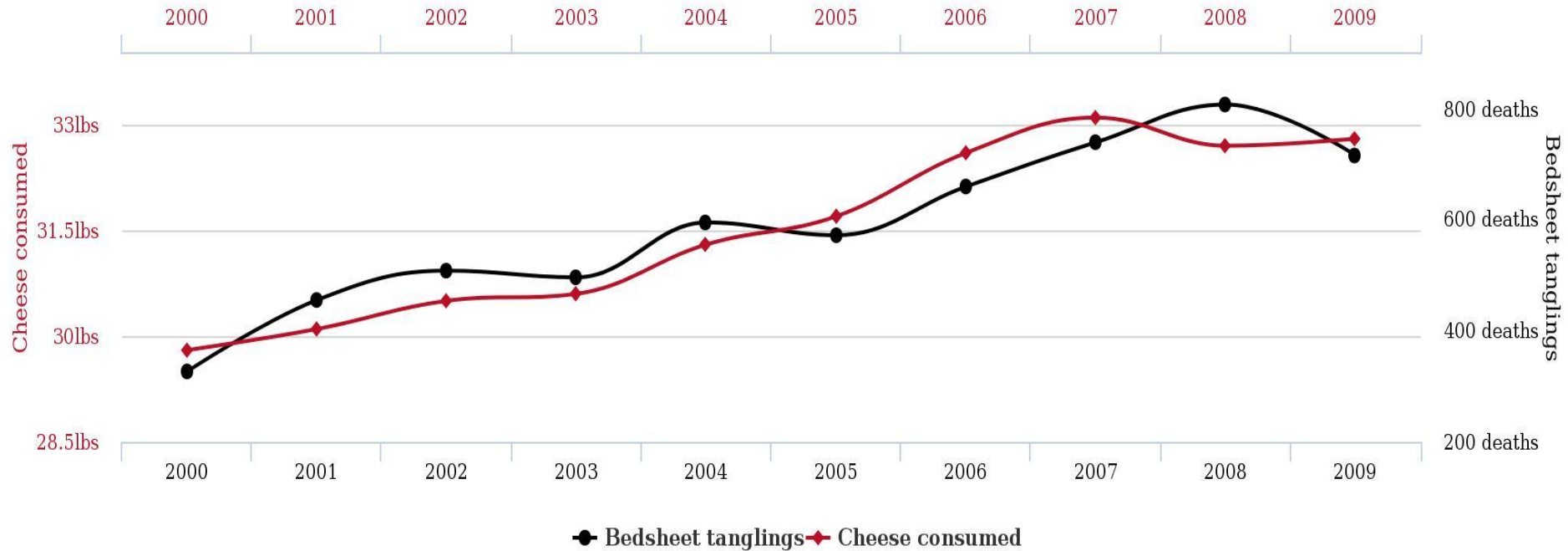
Quando é uma correlação enganosa ou falsa

Pode levar à adoção de uma política que não terá efeito na variável de interesse

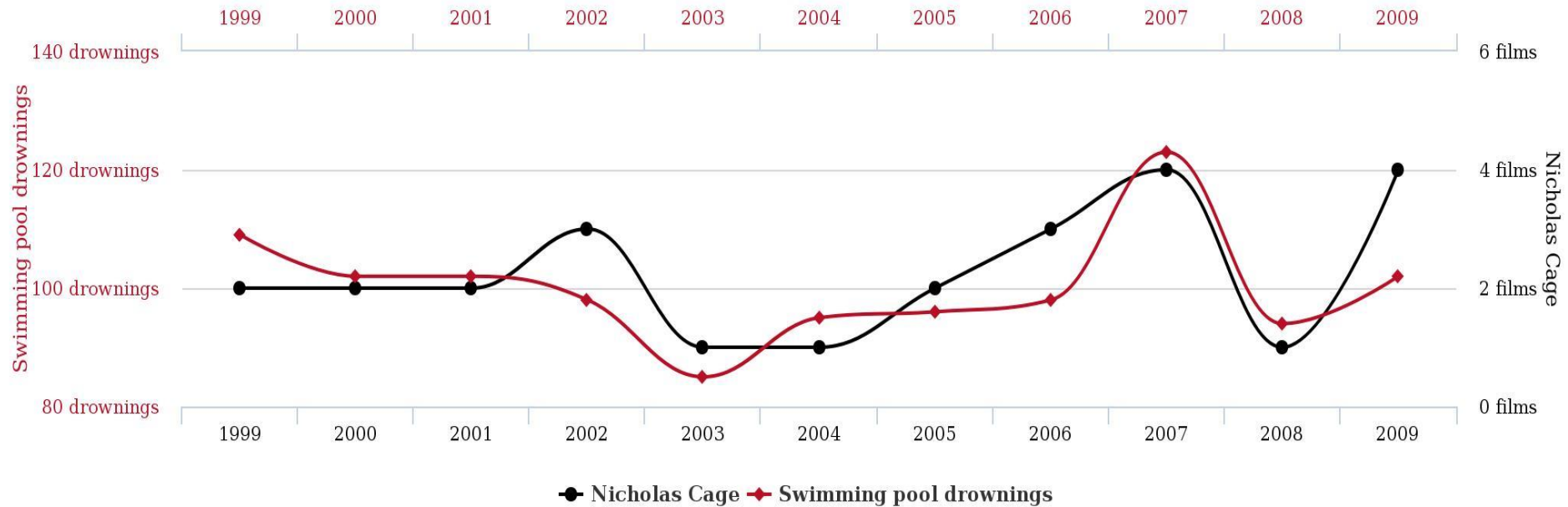
# Per capita cheese consumption

correlates with

## Number of people who died by becoming tangled in their bedsheets



# Number of people who drowned by falling into a pool correlates with Films Nicolas Cage appeared in



# A Importância da Inferência Causal

A maior parte das perguntas de políticas públicas que fazemos são causais:

- Se facilitarmos o acesso ao microcrédito numa comunidade, será que as pessoas vão criar novos negócios?
- Se aumentarmos os salários dos profissionais de saúde, será que estes vão exercer mais esforço e ser mais eficientes?
- Se reduzirmos as tarifas, será que mais empresas vão importar e exportar?

Em todos os casos, queremos saber se **X** causa **Y**

Se investirmos na política **X**, será que vamos obter o resultado **Y**?

# A Importância da Inferência Causal

## Exemplo: Avaliação do Impacto de uma Transferência Monetária

**Programa:** Uma ONG implementou um programa que compensava famílias com uma transferência monetária, se fizessem visitas regulares a clínicas



# Avaliação da Intervenção

## Três anos mais tarde

Podemos fazer uma avaliação simplista comparando os níveis de pobreza antes e depois da intervenção

Redução da pobreza + maiores investimentos em saúde = **SUCCESSO!**

A ONG tenta convencer o governo e os doadores a expandir o programa

O que deve o governo fazer?

Há evidência suficiente?

As intervenções são **dispendiosas** e há **poucos recursos**

# Explicações Possíveis

## Cenário 1

A transferência monetária aumentou os níveis de rendimento e a condição imposta pelo programa levou os participantes a recorrer às unidades de saúde.



**Neste caso, o governo deveria expandir o programa**

# Explicações Possíveis

## Cenário 2

As localidades recebem a transferência monetária e registam um rápido crescimento económico devido a grandes investimentos infraestruturais realizados na região. Estes investimentos infraestruturais levaram a criação de melhores clínicas e à melhoria dos Serviços de saúde. Há uma correlação entre os resultados de desejados.



**Será que precisamos do programa para atingir os mesmos resultados?**

# Explicações Possíveis

## Cenário 3

Ao mesmo tempo que a transferência monetária teve lugar, o governo investiu num programa de formação vocacional para ajudar na criação de pequenas empresas. O programa de transferência de dinheiro permitiu que os participantes dedicassem menos tempo à procura de emprego. Na ausência do programa, o nível de rendimento das pessoas seria ainda mais alto.



**O programa deve ser descontinuado**

# Importância de uma Relação Causal

## Porquê Inferência Causal

Recursos são escassos

Generalização de resultados

Replicabilidade

## Quando é que métodos rigorosos de avaliação causal são desejáveis e exequíveis?

Não temos evidência objectiva e rigorosa do impacto de uma política

Projecto que pode ser implementado a grande escala

O risco de investir de forma incorrecta é alto

# Avaliação da Intervenção

Os recursos são escassos e precisamos de os alocar de forma eficiente. Em qualquer avaliação, há dois tipos de preocupação:

## Validade Interna de uma análise

Podemos inferir que os dados indicam que uma política **causou** o resultado esperado? O nosso **X** causou o **Y**?

## Validade Externa de uma análise

Podemos extrapolar da avaliação que esta intervenção vai ter um impacto semelhante quando implementada noutro contexto? Será que o X causará o Y noutro contexto?

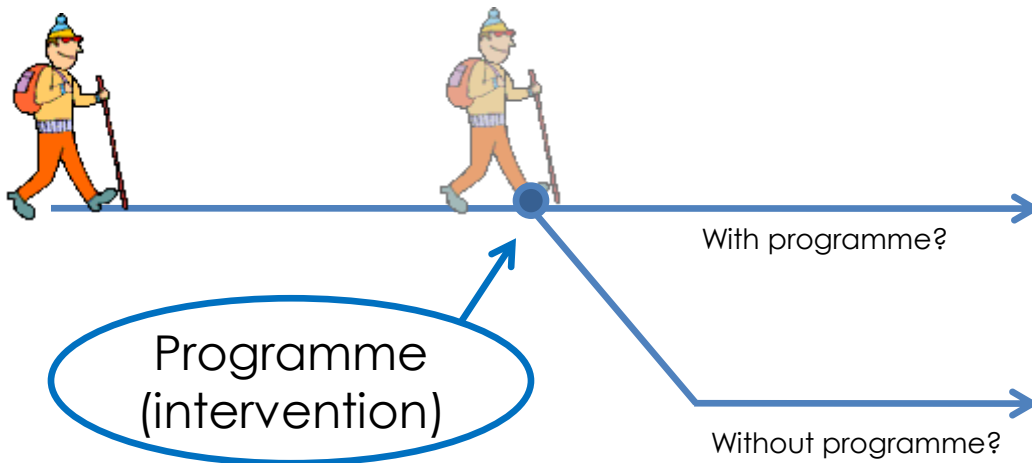
# Validade Interna: Analise Contrafactual

O que teria acontecido na ausência da intervenção?

**Cenário 1:** - melhoria de saúde

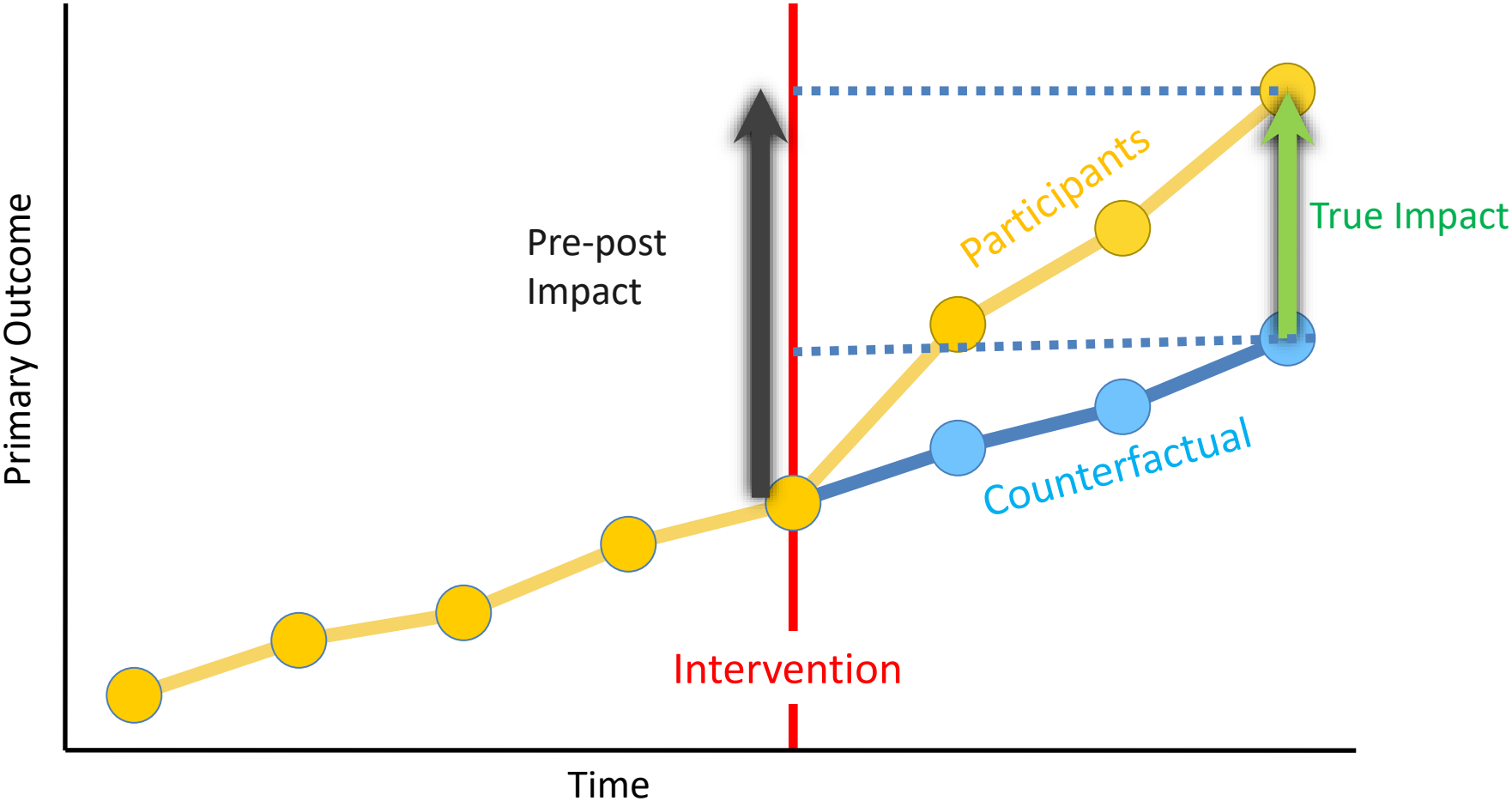
**Cenário 2:** + melhoria de saúde

**Cenário 3:** ++ melhoria de saúde



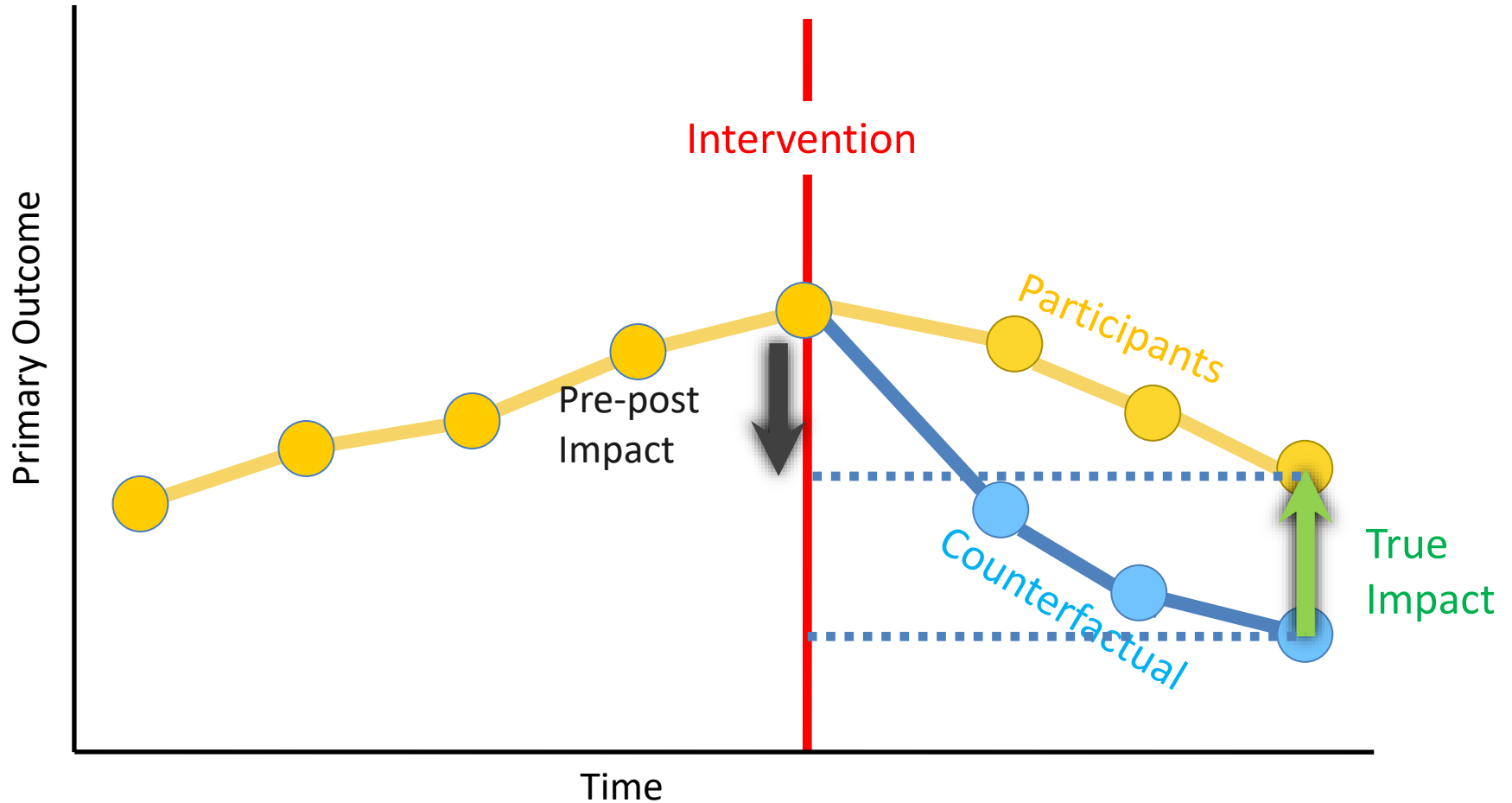
A comparação entre o resultado actual e o contrafactual permite-nos isolar o efeito do programa.

# Sobrestimar o Efeito de uma Intervenção



Source: JPAL

# Subestimar o Efeito de uma Intervenção



Source: JPAL

# Sumário até ao momento...

## Conceitos importantes:

Correlação vs Causalidade;

Validade interna e externa de uma análise;

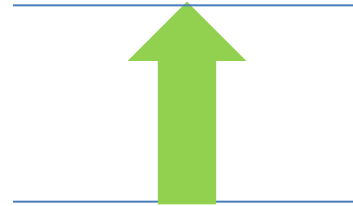
Análise **contrafactual** vs análise que compara o **antes** e o **depois**.

# Exemplo: A certificação do comércio justo ajuda os pequenos agricultores?

**ANTES**



**DEPOIS**



**CREDITO**

# Como criar um contrafactual?

## Requisito 1

**Ex Ante:** Os grupos de indivíduos têm que ser comparáveis

## Requisito 2

**Ex Post:** A única diferença entre os grupos, é a intervenção

# Como criar um contrafactual?

**Nao certificado**



**Certificado**



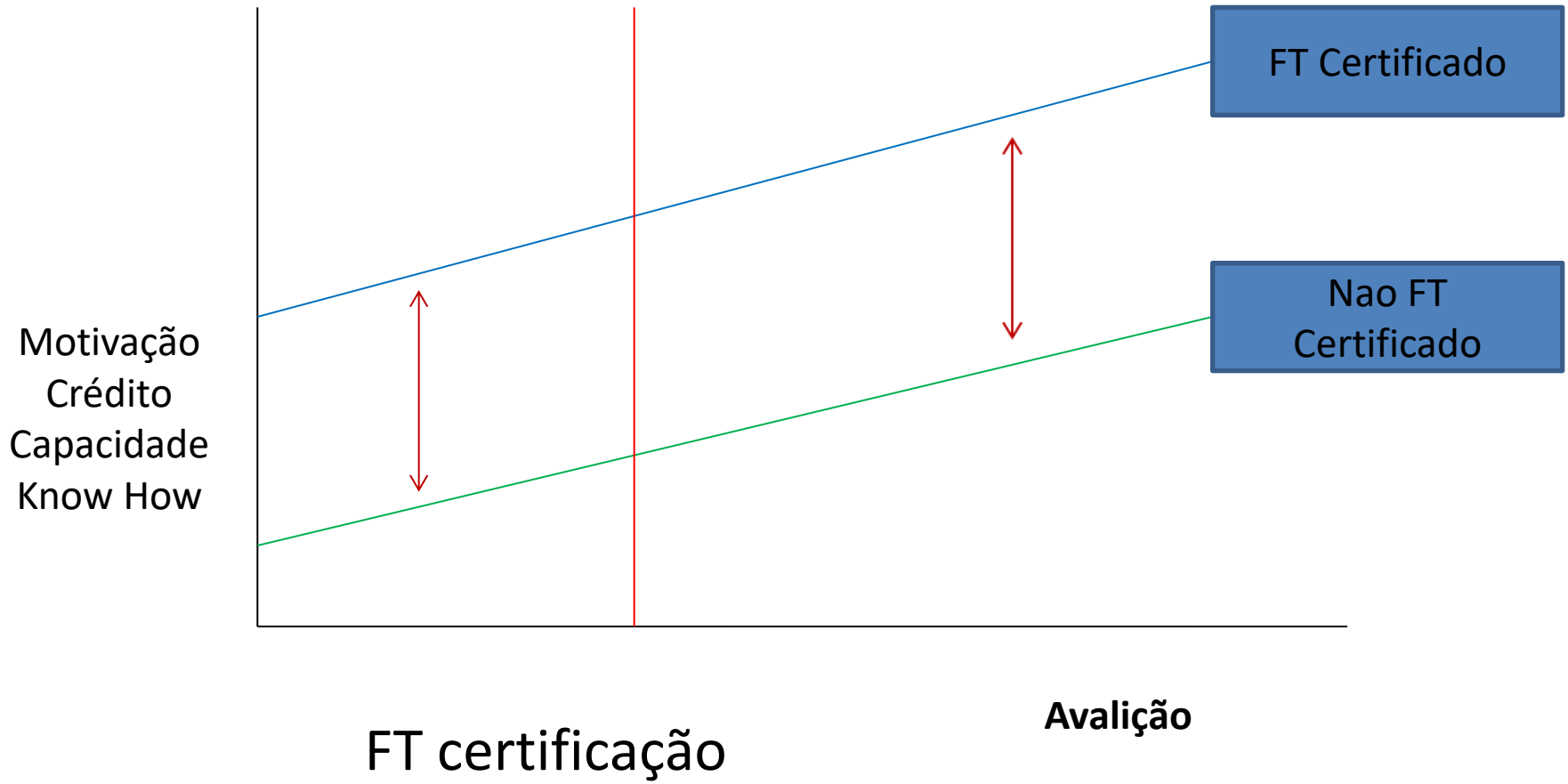
**Diferenças de motivação,  
capacidades, acesso a  
crédito**



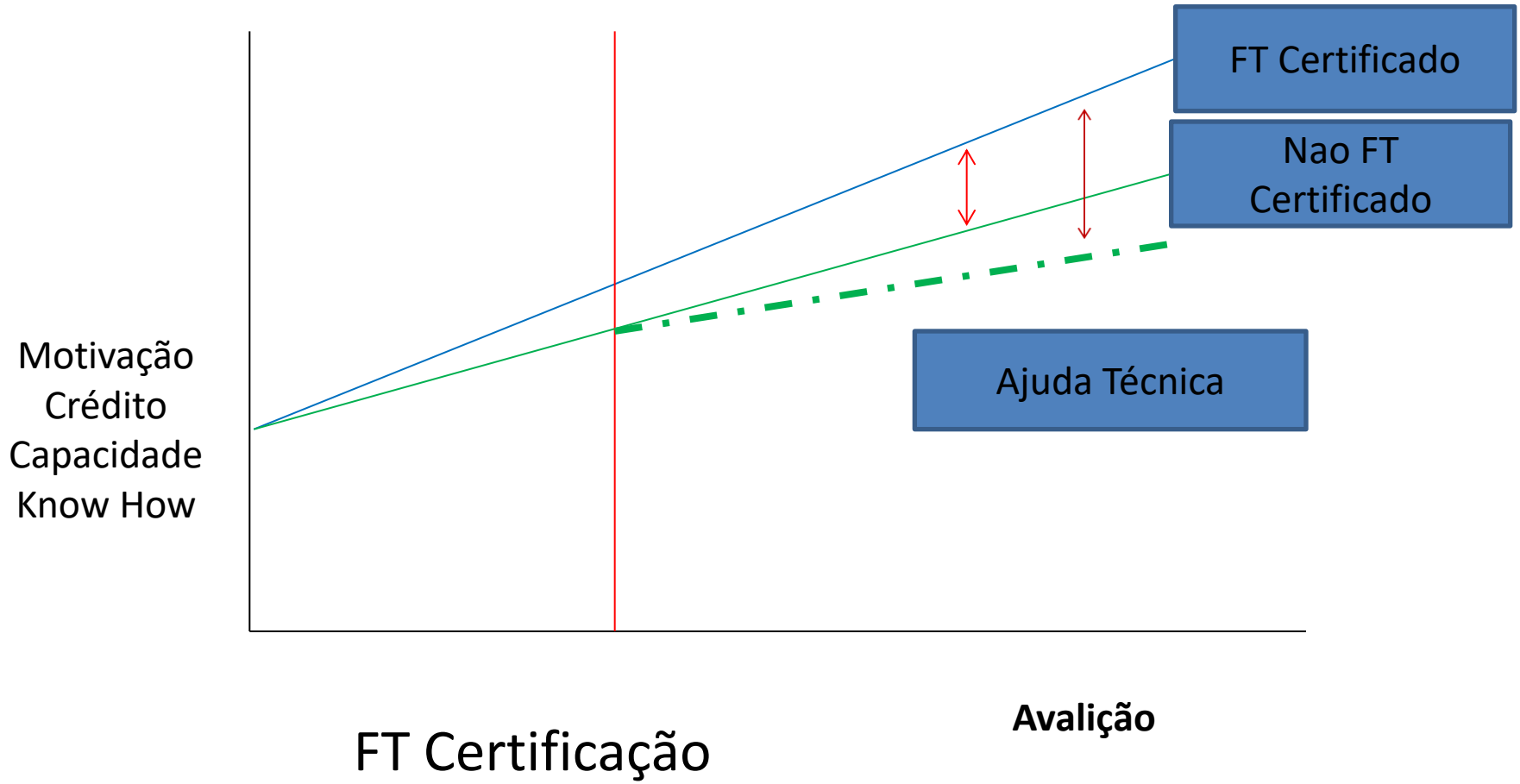
**Diferenças nos resultados**

**Em que direção é que vai o vies?**

# Requisito 1: Diferenças em condições iniciais



# Requisito 2: Contaminação do Grupo de Controlo



# Análise Contrafactual

A variável **X** (independente) afecta a variável **Y** (dependente)?

$$Y = \alpha + \beta X + \varepsilon$$


Assumindo que a variável **X** é ortogonal ao erro  $\varepsilon$   
Isto requer que a variação no **X** seja **EXÓGENA**

# Desafios Empíricos

**Causalidade Inversa:** O **Y** influencia o **X** ou o **X** influencia o **Y**?

Os agricultores de comércio justo têm um rendimento superior desde o início e por isso é que podem pagar a certificação em vez de terem um rendimento mais alto por causa da certificação.

Rendimento dos agricultores  
Certificação de comércio justo



Certificação

# Desafios Empíricos

**Variáveis Omitidas:** A variável **Z**, que foi omitida da regressão/análise, influencia a variação da variável **X**  
Exemplo: os agricultores de comércio justo são mais empreendedores e, por isso, decidem ser certificados.

**Variável omitida:** Empreendedorismo.

**Erro de Medição:** Medimos a variável **X** com erro.

Exemplo: não é claro quem está certificado (só os agricultores mais organizados e que têm os documentos da certificação em ordem).

# Importância da Variação Exógena do X

O nosso coeficiente  $\beta$  já não capta o impacto de X em Y mas também capta.....

**Causalidade Inversa:** o efeito de X em Y é o mesmo que o efeito de de Y em X.

**Vies de Variável Omitida:** o efeito de X em Y e o efeito de Z em X

**Erro de Medição:** o efeito do erro de medição de X

# Sumário

Requisitos para uma análise contrafactual

Semelhança e não-contaminação

**Desafios na identificação de relacionamentos causais**

Causalidade Inversa

Variáveis Omitidas

Erro de medição

# Como identificar uma relação causal

**Objectivo:** identificar um grupo de comparação para a análise contrafactual exige uma variação exógena na exposição ao programa/política.

## **Métodos Experimentais**

Randomized Control Trials (RCTs)

## **Métodos Quasi-experimentais**

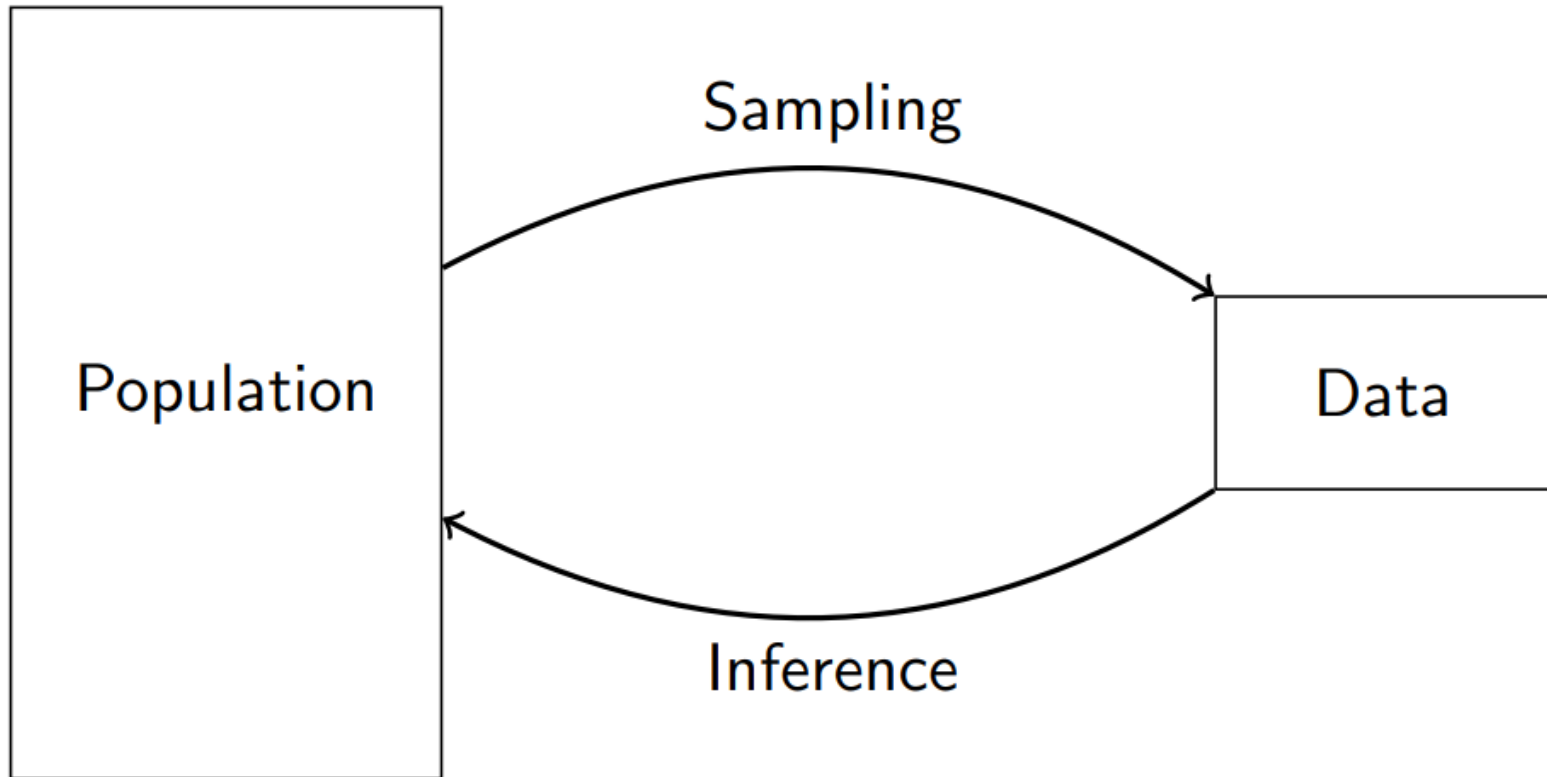
Discontinuidade de uma Regressão (RDD)

Diferenças em diferenças (DID)

# Estratégia de Identificação

- **Estratégia de Identificação:** Um desenho empírico (research design) que procura minimizar problemas de endogeneidade na análise
- Uma estratégia de identificação geralmente requer uma forma criativa do investigador/a construir um contrafactual plausível.
  - Exemplos (RCT, DID, RDD, Variável Instrumental)

# Amostragem e Inferência



# Amostragem e Inferência Causal

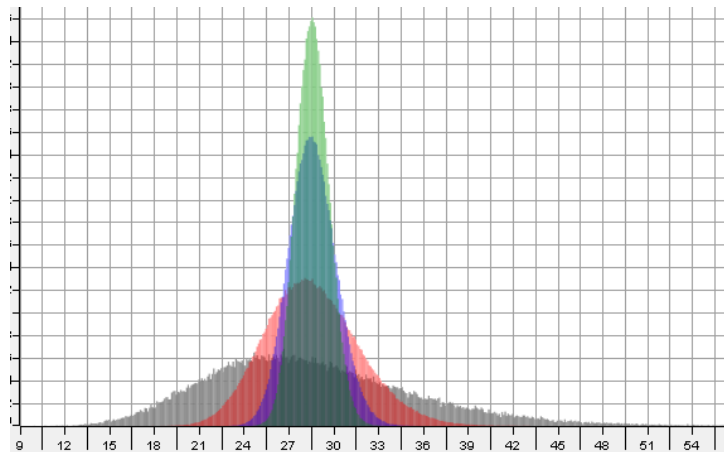
- Amostragem **Aleatória**: Evitar o vies na amostragem para obter uma amostra representativa.
- **Teorema do Limite Central**: afirma que, para uma amostra suficientemente grande, a distribuição da média amostral de uma variável aleatória se aproxima de uma **distribuição normal**, independentemente da distribuição original da população.
- Se a população tiver qualquer distribuição (uniforme, exponencial, etc.), a distribuição das médias amostrais tende a ser **normal** à medida que o tamanho da amostra aumenta.

# Amostragem e Inferência Causal

Quanto maior o tamanho da amostra ( $n$ ), melhor a aproximação a uma distribuição normal.

A média da distribuição amostral será aproximadamente igual à média populacional ( $\mu$ ). (Lei dos Grandes Números)

**O Teorema do Limite Central** é fundamental para a estatística inferencial, pois permite fazer estimativas e testes de hipóteses mesmo quando não se conhece a distribuição original da população.



# Na prática...

- **Lei dos Grande Números:** Podemos confiar em amostras grandes para obter estimativas fiáveis dos parâmetros de interesse. A média amostral aproxima-se da média populacional.
- **Teorema do Limite Central:** permite fazer afirmações probabilísticas sobre a média amostral e construir intervalos de confiança. A aproximação à distribuição normal é essencial para podermos utilizar **z-tests** e **t-tests**.

# Regressão: OLS

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

**Parâmetros:**  $\beta_0$   $\beta_1$ .

Valores constantes a serem estimados. Representam a relação de interesse a ser estimada.

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + \hat{u}_i$$

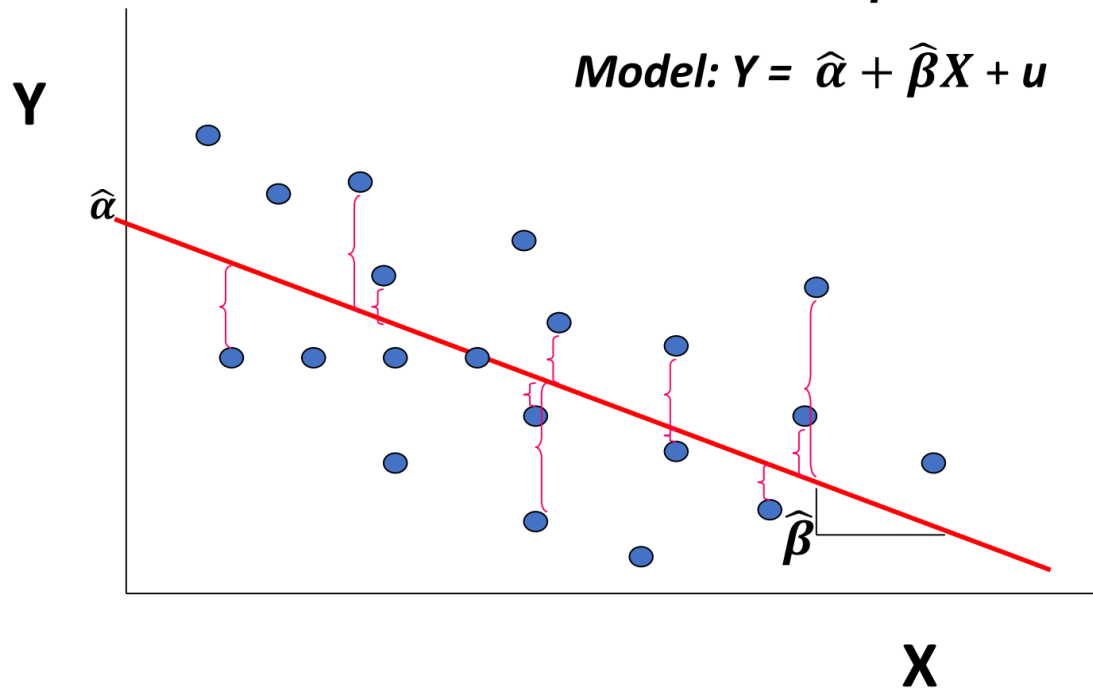
$$\min \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \min \sum_{i=1}^n (y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i))^2$$

# Regressão

$Y = m X + b$ , we re-write as:

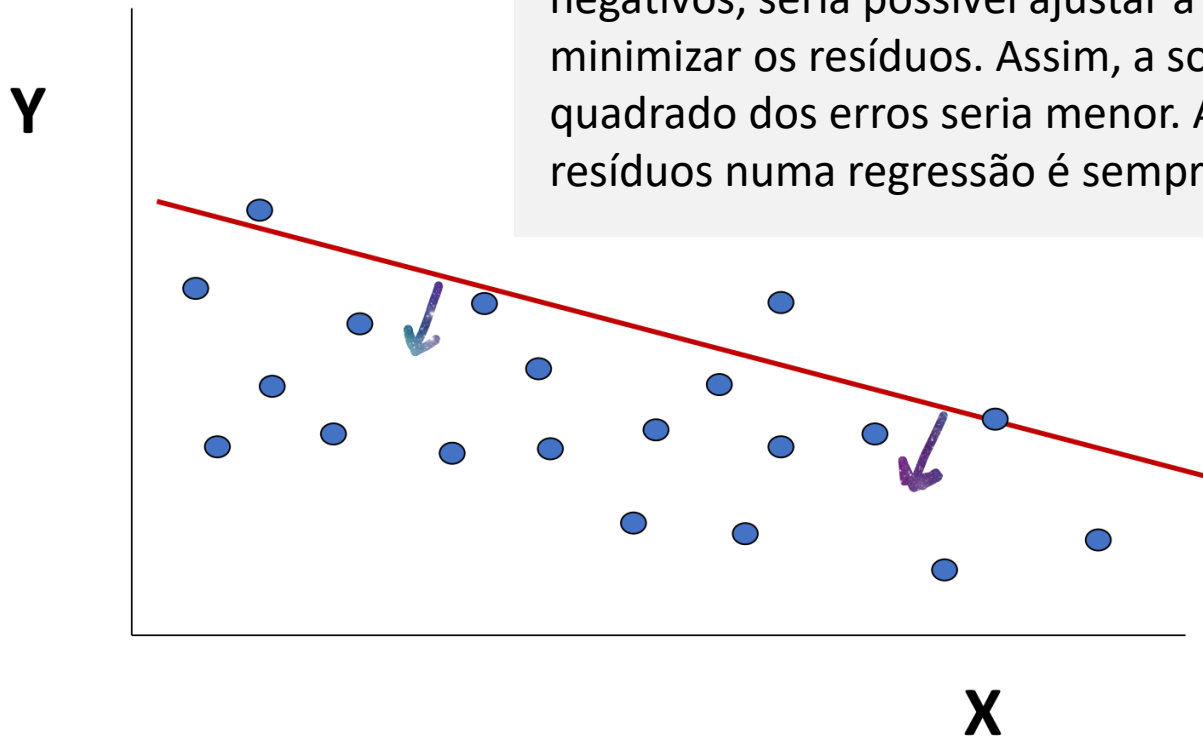
**Line:**  $\hat{Y} = \hat{\alpha} + \hat{\beta}X$

**Model:**  $Y = \hat{\alpha} + \hat{\beta}X + u$

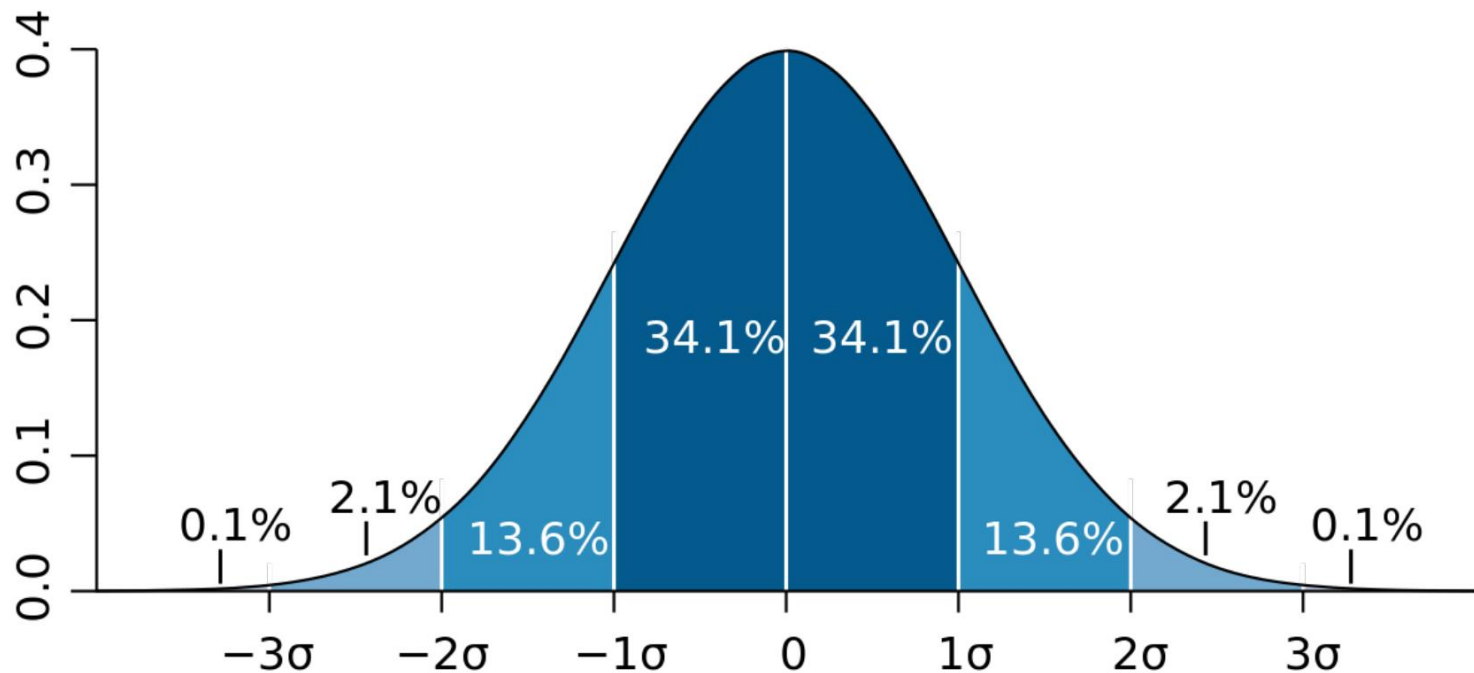


# Regressão

Se houvesse mais resíduos positivos ou negativos, seria possível ajustar a curva até minimizar os resíduos. Assim, a soma do quadrado dos erros seria menor. A média dos resíduos numa regressão é sempre zero.

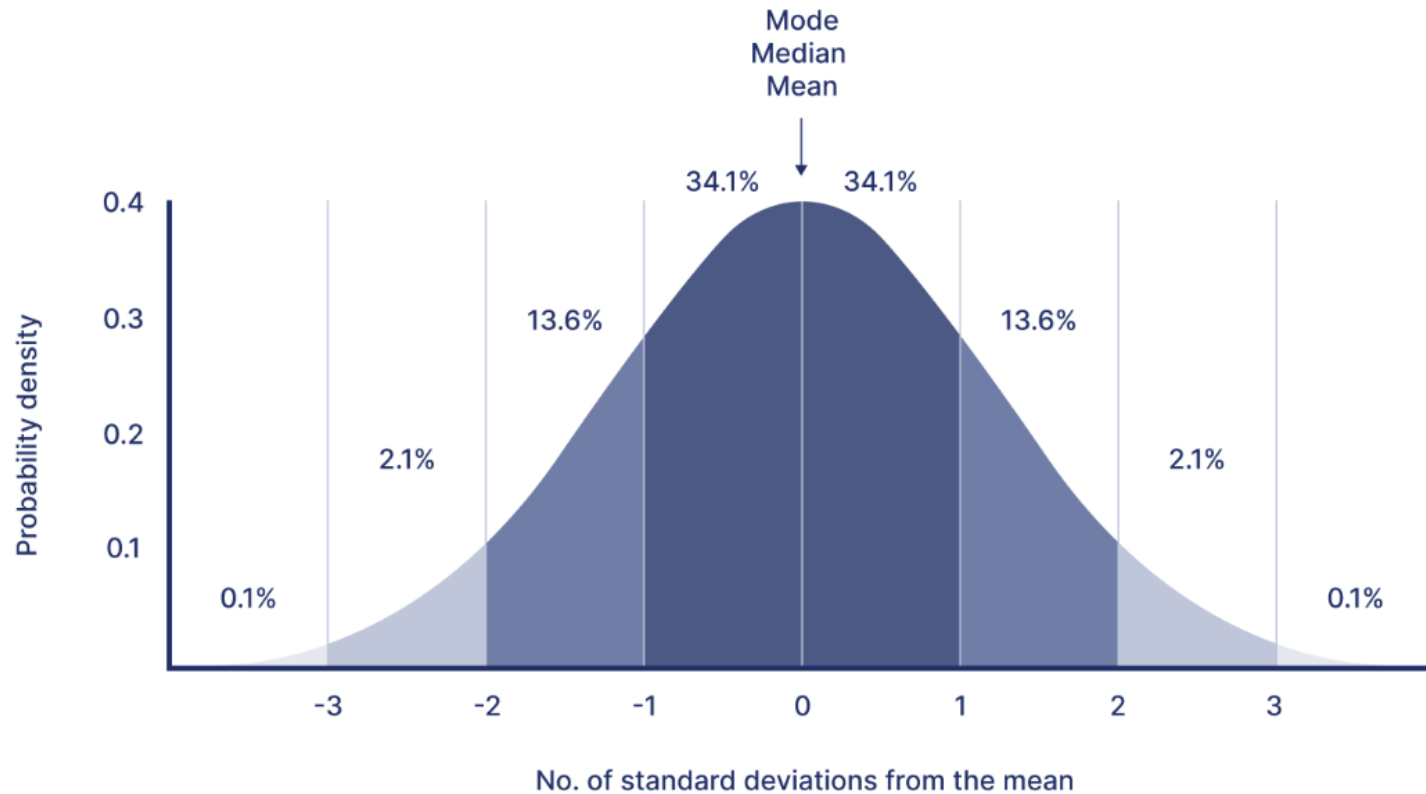


# Testes Hipóteses: Distribuição



Todas as distribuições normais têm uma relação fixa entre a média e o desvio padrão  $\sigma$ . É complicado determinar valores (cut-offs) para significância estatística, por isso normalizamos a distribuição...

# Teste de Hipóteses: Normalização

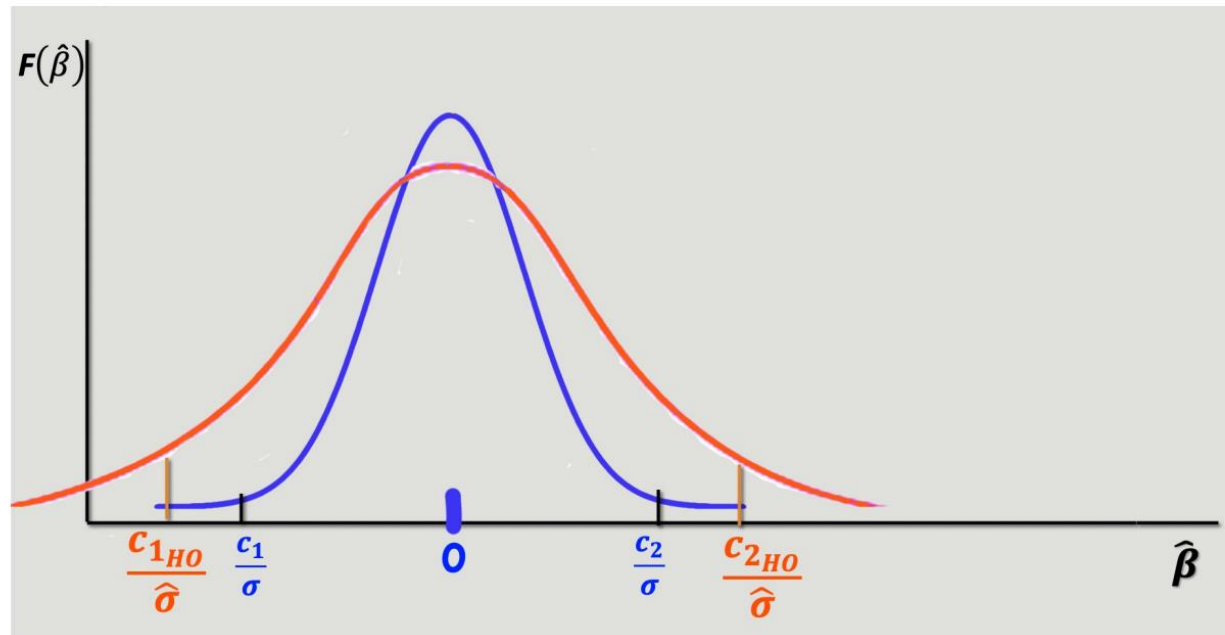


Z-Score: 
$$z = \frac{x - \mu}{\sigma}$$

# Standard Normal e t-distribution

- Para determinarmos a probabilidade de observações numa distribuição estarem acima ou abaixo de um determinado valor;
- Para descobrir a probabilidade da média amostral ser diferente da média da população;
- Para comparar z-scores de diferentes distribuições, com médias e desvio padrão diferentes.

No, entanto, o desvio padrão  $\sigma$  tem que ser estimado também, o que introduz mais incerteza.



# T-distribution

$$t = \frac{\hat{\beta} - \beta_{HO}}{\hat{\sigma}} = \frac{\hat{\beta}}{\hat{\sigma}}$$

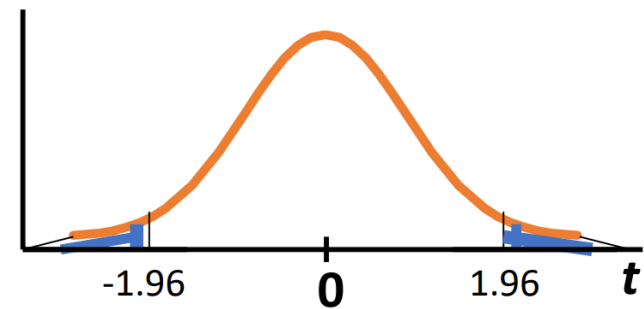
Uma vez que o desvio padrão é também estimado, há mais incerteza e uma maior massa probabilística nas caudas da distribuição em comparação com a distribuição normal padrão. O desvio padrão estimado tem a sua própria distribuição.

Esta transformação  $t$  segue uma  $t$ -distribution, que para amostras pequenas tem caudas com maior massa probabilística do que uma distribuição normal padrão, mas converge para esta à medida que a amostra aumenta.

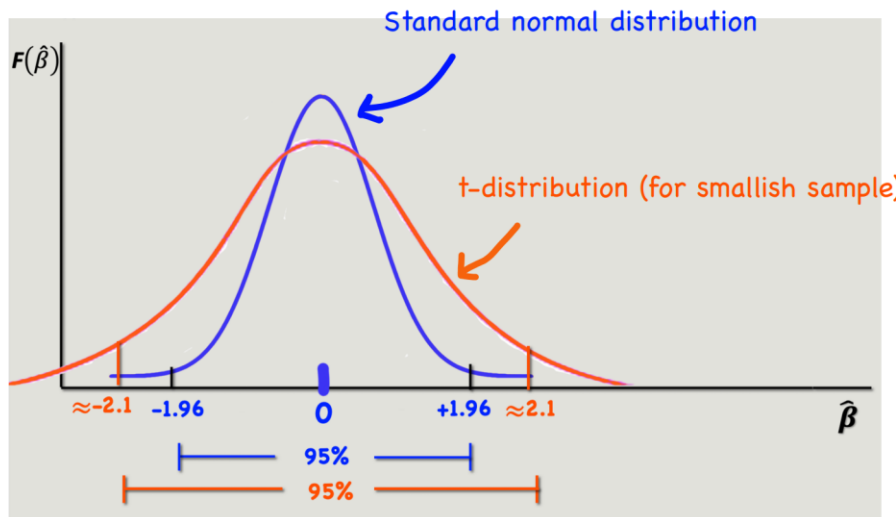
$$t = \frac{\hat{\beta} - \beta_{HO}}{\hat{\sigma}_{\hat{\beta}}}, \text{ since } \beta_{HO} = 0$$

$$\text{we get: } t = \frac{\hat{\beta}}{\hat{\sigma}_{\hat{\beta}}}$$

t - distribution:

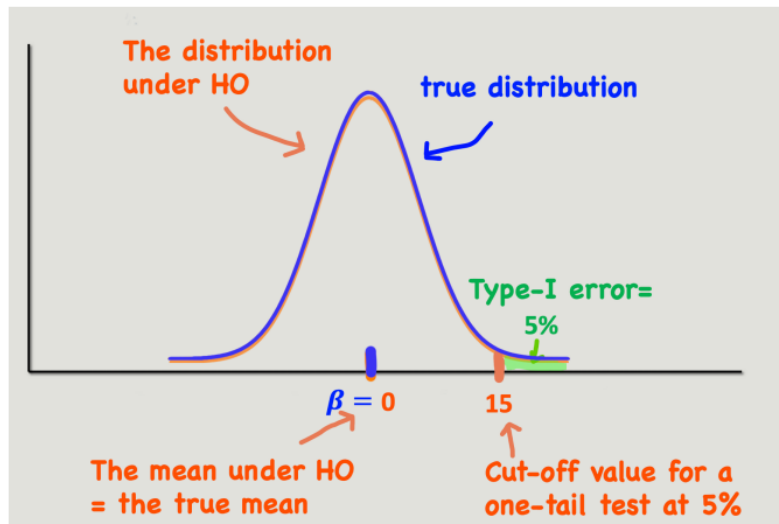


Area under curve in ■ = 5%



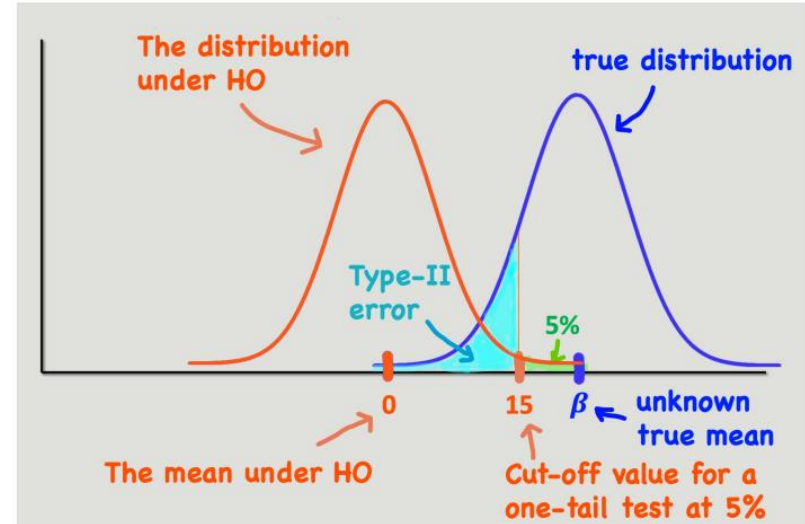
# Teste de Hipóteses

## Type I Error (Falso Positivo)



Quando rejeitamos a hipótese nula quando esta é verdadeira, detectamos um efeito que não existe na realidade.

## Type II Error (Falso Negativo)



Quando não rejeitamos uma hipótese nula que é falsa, não detectamos um efeito que existe na realidade.

# Níveis de significância

$\alpha$ : a probabilidade de rejeitar a hipótese nula quando esta existe na realidade (**Type I Error**).

Quando o nível de significância é 5% (0,05) ou 1% (0,01), estamos a dizer que estamos dispostos a aceitar uma probabilidade de 5% ou 1% de, incorretamente, rejeitar a hipótese nula.

A t-statistic  $t$  diz-nos quantos desvios padrão da distribuição amostral (ou standard errors)  $\hat{\beta}$  estão acima da média. Segundo a hipótese nula, 0. Se estiver suficientemente afastada, então rejeitamos a hipótese nula.

$$t = \frac{\hat{\beta} - \beta_{HO}}{\hat{\sigma}} = \frac{\hat{\beta}}{\hat{\sigma}}$$

# Regressão: Interpretação de Coeficientes

REGRESSION	REGRESSION	INTERPRETATION
Level-Level	$Y_i = \alpha_i + \beta_i X_i + \epsilon_i$	A unit increase in $X$ results in a $\beta$ unit increase in $Y$
Level-Log	$Y_i = \alpha_i + \beta_i \ln(X_i) + \epsilon_i$	A 1% increase in $X$ leads to a $\frac{\beta}{100}$ unit increase in $Y$
Log-Level	$\ln(Y_i) = \alpha_i + \beta_i X_i + \epsilon_i$	A unit increase in $X$ results in a $100 \times \beta\%$ increase in $Y$
Log-Log	$\ln(Y_i) = \alpha_i + \beta_i \ln(X_i) + \epsilon_i$	A 1% increase in $X$ leads to a $\beta\%$ increase in $Y$

Table 1: Regression Types and Interpretations

# Exemplos de Pesquisa

**Projecto 1:** avaliar o impacto de uma política que promove o crescimento de pequenas empresas. Ao fim de um ano, as empresas participantes têm lucros mais altos e recrutaram mais trabalhadores do que as empresas que não participaram.

**Podemos concluir que o esquema causou o crescimento das empresas participantes? Porquê?**

**Que outros factores poderão explicar esta diferença entre empresas participantes e não participantes?**

# Exemplos de Pesquisa

**Projecto 2:** Um país amplia a rede elétrica em zonas rurais. 5 anos mais tarde, as localidades com eletricidade têm um rendimento superior às localidades sem eletricidade.

**A expansão da rede elétrica terá causado estas diferenças entre localidades?**

**Que outras explicações são consistentes com esta evidência?**

# Exemplos de Pesquisa

**Projecto 3:** Um país introduziu um novo imposto com instrumentos de recolha digitais. Um ano mais tarde, a receita fiscal aumentou em 20%.

**Foi o imposto que causou este aumento de receita?**

**Que outros fatores poderão ter contribuído para este resultado?**

# Formação contínua

- IGC BREAD PhD:
  - [Urban economics](#)
  - [Environmental economics](#)
  - [Firms](#)

[Vox Dev Lit](#)