



Treinamento em Monitoria e Avaliação para o Governo de Moçambique

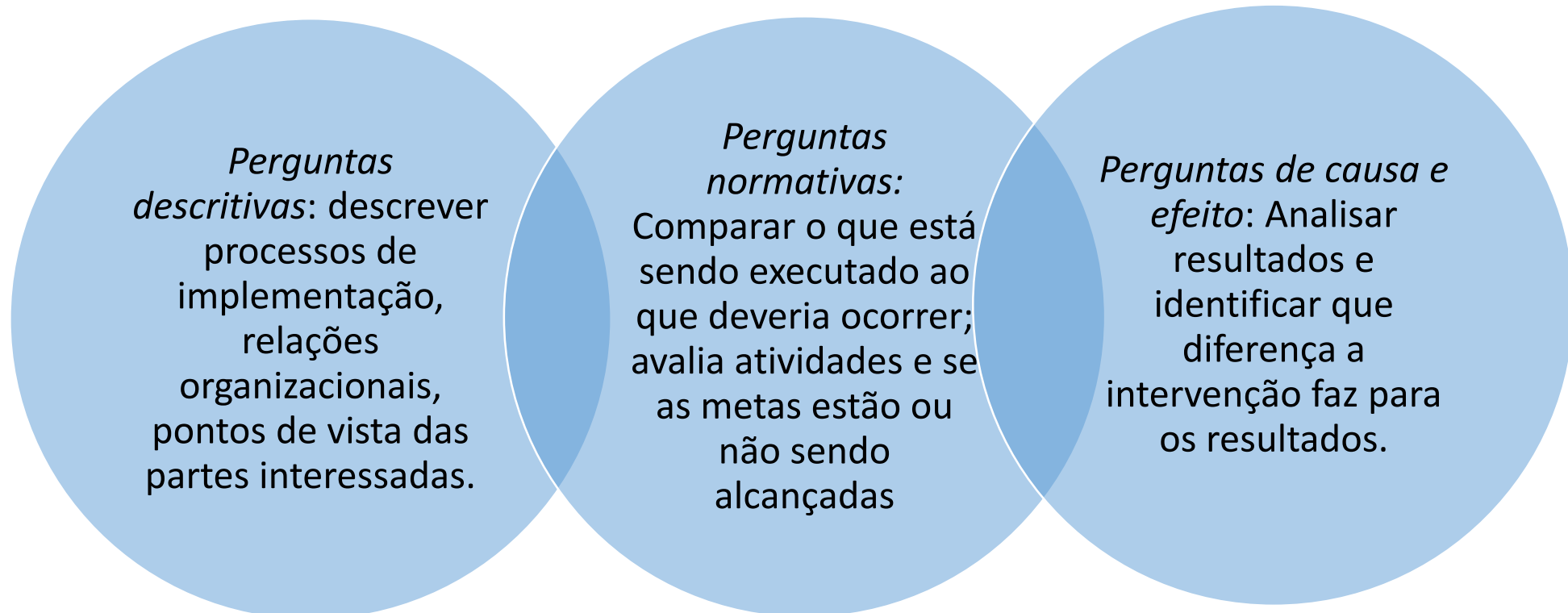
27 de fevereiro a 03 de março de 2017

Maputo

Objetivo da avaliação de impacto

Um avaliação de impacto procura responder a perguntas de causa e efeito.

Em geral, as avaliações podem abordar três tipos de questões (Imas e Rist 2009)



Objetivo da avaliação de impacto

Pergunta básica da avaliação de impacto:

Qual é o impacto de um programa/política sobre um resultado de interesse?



= Efeito causal

Objetivo da avaliação de impacto

Pergunta básica da avaliação de impacto:

Qual é o impacto de um programa/política sobre um resultado de interesse?



= Efeito causal

Por exemplo:

- O impacto de um programa de transferência de renda na frequência escolar das crianças*
- O impacto de um programa de transferência de renda na segurança alimentar dos beneficiários*

Monitoria e Avaliação

Insumos → Atividades → Produtos → Resultados → Impactos

Inputs
Recursos financeiros / humanos

Activities
Ações tomadas para converter insumos em produtos específicos

Outputs
Produtos resultantes da conversão em produtos tangíveis

Outcomes
Uso dos produtos pelo público-alvo

Impact
Mudanças na vida dos beneficiários / não beneficiários

Monitoria & Avaliação de Processos

Avaliação de Impacto








Exercício: Quais das perguntas seguintes são perguntas de monitoria e quais são perguntas de avaliação

Exemplo: O subsídio de criança - SC (0-2 anos) a ser criado como uma das vertentes do pacote de subsídios para crianças da ENSSB-2016-2024 tem como objetivo diminuir melhorar a alimentação das crianças, reduzir indicadores de desnutrição e morbimidade, facilitar acesso aos serviços de saúde.

Exercício: Quais das perguntas seguintes são perguntas de monitoria e quais são perguntas de avaliação

- O SC causou uma melhora no estado nutricional das crianças?
- Quantas crianças foram cobertas pelo SC?
- Quanto dinheiro foi transferido às delegações para o pagamento do subsídio?
- O SC mudou os hábitos alimentares das crianças que participaram e dos pais delas? (consumo de alimentos mais saudáveis, com maior frequência etc)
- O SC levou um aumento ou redução de gastos com a saúde das crianças
- Houve efeito sobre crianças maiores de 2 anos no agregado familiar?

Exercício: Quais das perguntas seguintes são perguntas de monitoria e quais são perguntas de avaliação

- O SC causou uma melhora no estado nutricional das crianças?  Avaliação de impacto
- Quantas crianças foram cobertas pelo SC?  Monitoria / Avaliação de processo
- Quanto dinheiro foi transferido às delegações para o pagamento do subsídio?  Monitoria / Avaliação de processo
- Os pagamentos foram realizados com a periodicidade planejada?  Monitoria / Avaliação de processo
- O SC mudou os hábitos alimentares das crianças que participaram e dos pais delas? (consumo de alimentos mais saudáveis, com maior frequência etc)  Avaliação de impacto
- O SC levou um aumento ou redução de gastos com a saúde das crianças  Avaliação de impacto
- Houve efeito sobre crianças maiores de 2 anos no agregado familiar?  Avaliação de impacto

Monitoria e Avaliação

Conclusão:

Monitoria e avaliações de processo/resultados podem....

- Avaliar a quantidade e a qualidade das atividades implementadas para promover a “accountability” (prestação de contas) interna e externa dos recursos utilizados e dos resultados obtidos;
- Informar de maneira contínua em que medida o projeto está atingindo seus objetivos;
- Identificar potenciais problemas no estágio inicial da política e propor soluções/alternativas possíveis;
- Monitorar os níveis de acesso de cada setor da população-alvo à política/projeto (por exemplo, as nível de participação, etc.)
- Monitorar a eficiência dos diferentes componentes de um projeto em implementação e sugerir melhorias (por ex.: custos, preço dos insumos).

Monitoria e Avaliação

Conclusão:

A avaliação de impacto pode...

- Estimar o impacto (= efeito causal) de um programa sobre um resultado de interesse
- Responder à perguntas de causa e efeito sobre um programa
- Buscar as mudanças que vão além dos “objetivos específicos” do programa e que são atribuíveis diretamente a um programa – efeitos não intencionais
- Investigar o impacto do programa sobre não beneficiários que pode ser afetados pelo programa: “spillover effects”

2. Avaliação de impacto – como?

- **As avaliações de impacto tratam as perguntas de uma pesquisa de avaliação como um problema de inferência causal**
- A questão chave de uma avaliação de impacto é: qual é a relação de causa-efeito entre o programa P e o resultado de interesse Y?
- Assim, o impacto do programa será a diferença entre os resultados com o programa (P=1) e sem o programa (P=0)

$$\text{Impacto} = (Y | P=1) - (Y | P=0)$$

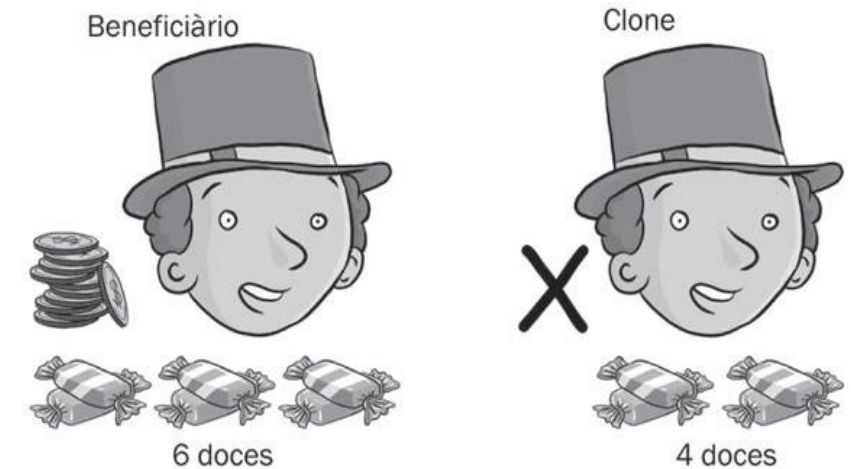
Obs: Para medir o efeito causal, o avaliador deve ter como objetivo eliminar/neutralizar todas as possibilidades e fatores que podem influenciar nos indicadores de resultado que não digam respeito à política/programa sendo avaliada/o.

O contrafactual

Consequentemente, se queremos medir o impacto de um programa (= efeito causal) sempre precisamos responder à pergunta seguinte: “Qual teria sido o resultado para os participantes do programa se eles não tivessem participado do programa”



Para avaliar o impacto de um programa, o ideal seria comparar a situação do beneficiário com um “clone” dele mesmo, que não tivesse participado do programa.



$$\text{Impacto} = 6 - 4 = 2 \text{ doces}$$

$$\text{Impacto} = (Y | P=1) - (Y | P=0)$$

O contrafactual (2)

Na prática, a avaliação de impacto exige que o avaliador encontre um grupo de comparação para estimar o que teria acontecido aos participantes do programa na ausência do programa.



O contrafactual (3)

Para criar um grupo de comparação válido, este precisa ser semelhante ao grupo de beneficiários (grupo de tratamento) em três maneiras:

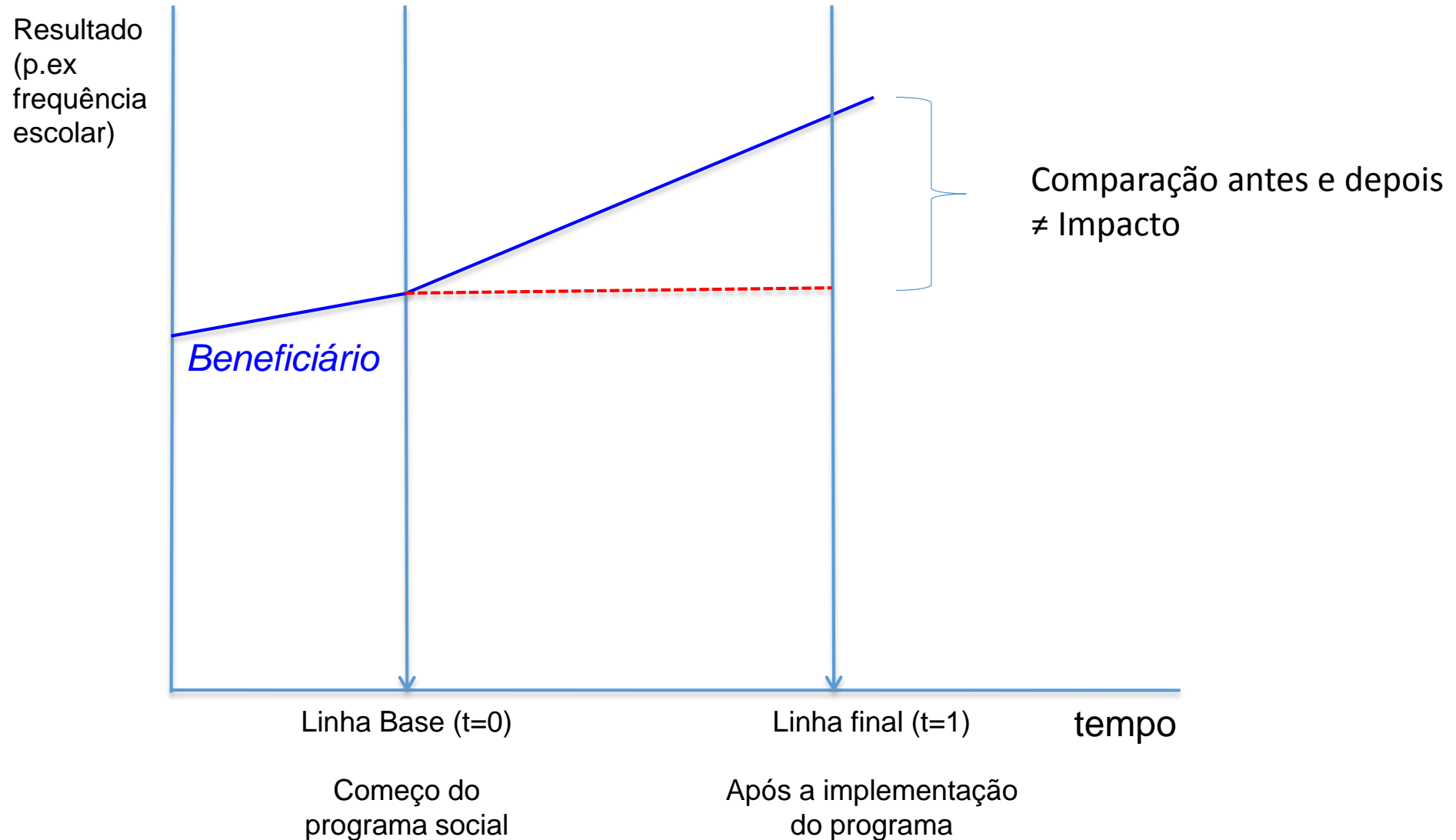
- Os grupos devem ser (em média) idênticos na ausência do programa;
- Os grupos devem reagir da mesma maneira ao benefício oferecido pelo programa;
- Os grupos não podem estar sujeitos a diferentes intervenções durante o período da avaliação



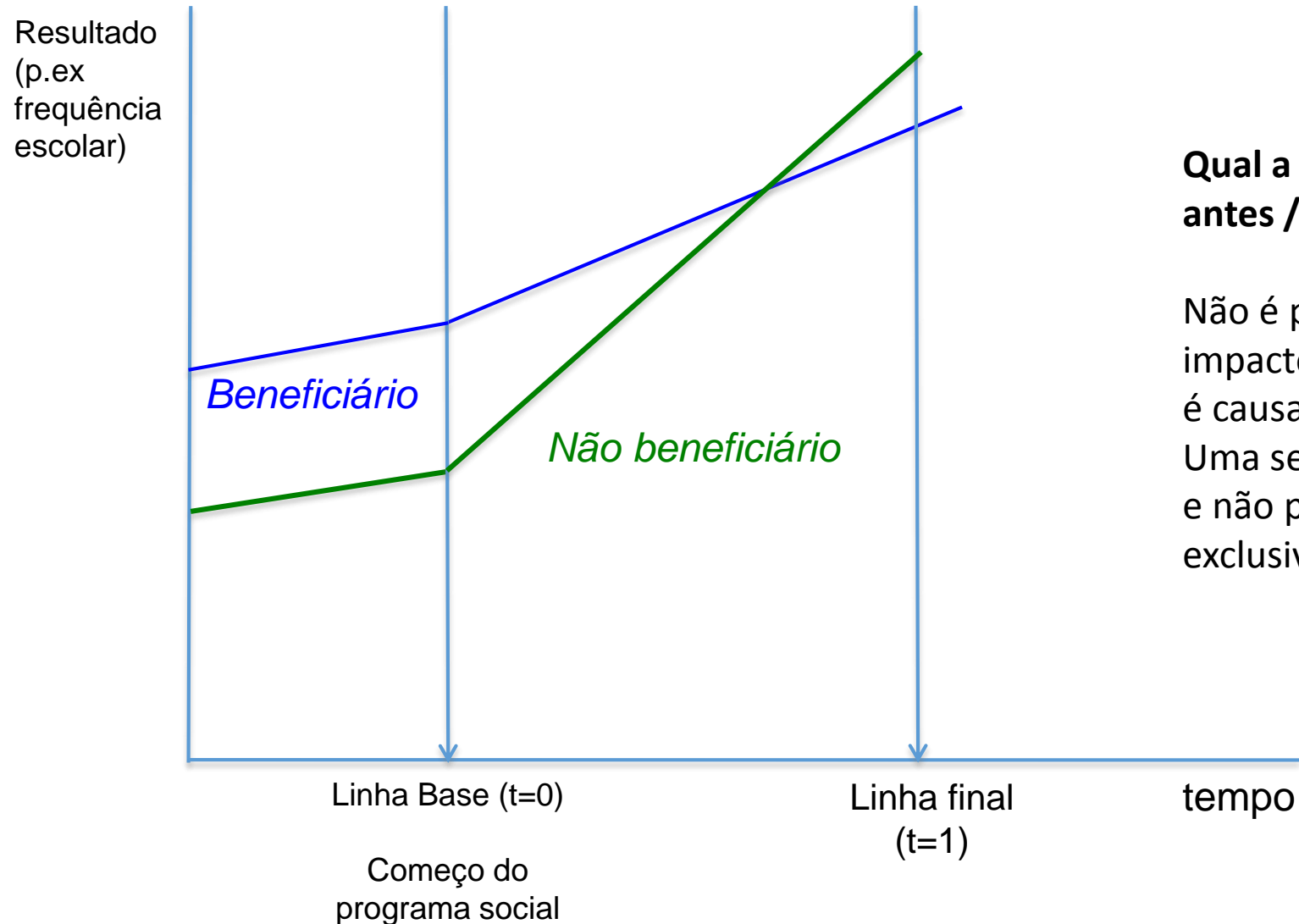
Avaliação de impacto – como?

É possível avaliar o impacto de um programa social observando as mudanças nos indicadores de resultado apenas do grupo de beneficiários, ao longo do tempo?

Avaliação de impacto – como?



Avaliação de impacto – como?



Qual a limitação da comparação antes / depois ?

Não é possível determinar se o impacto no indicador de resultado é causado por outros fatores (ex. Uma seca/um boom na economia), e não pelo programa exclusivamente.

Avaliação de impacto – como?

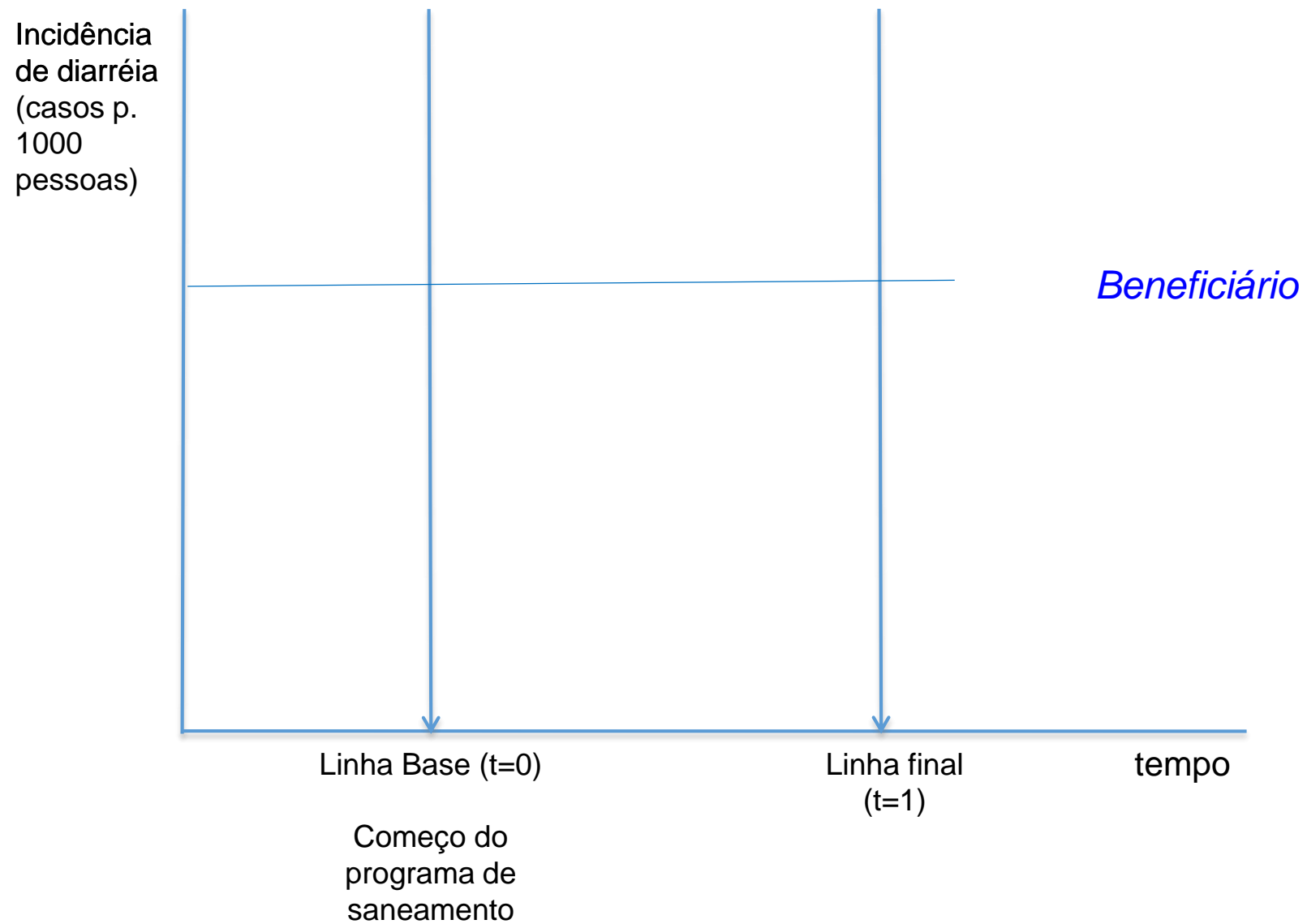
- Exercício: Um jornal de grande circulação apresentou em uma matéria um gráfico com a série histórica da incidência de diarreia em crianças da província de Ouarzazate. O gráfico mostrava que a incidência da doença permanecia inalterada após a implementação do piloto do SC em 2018.

Exercício:

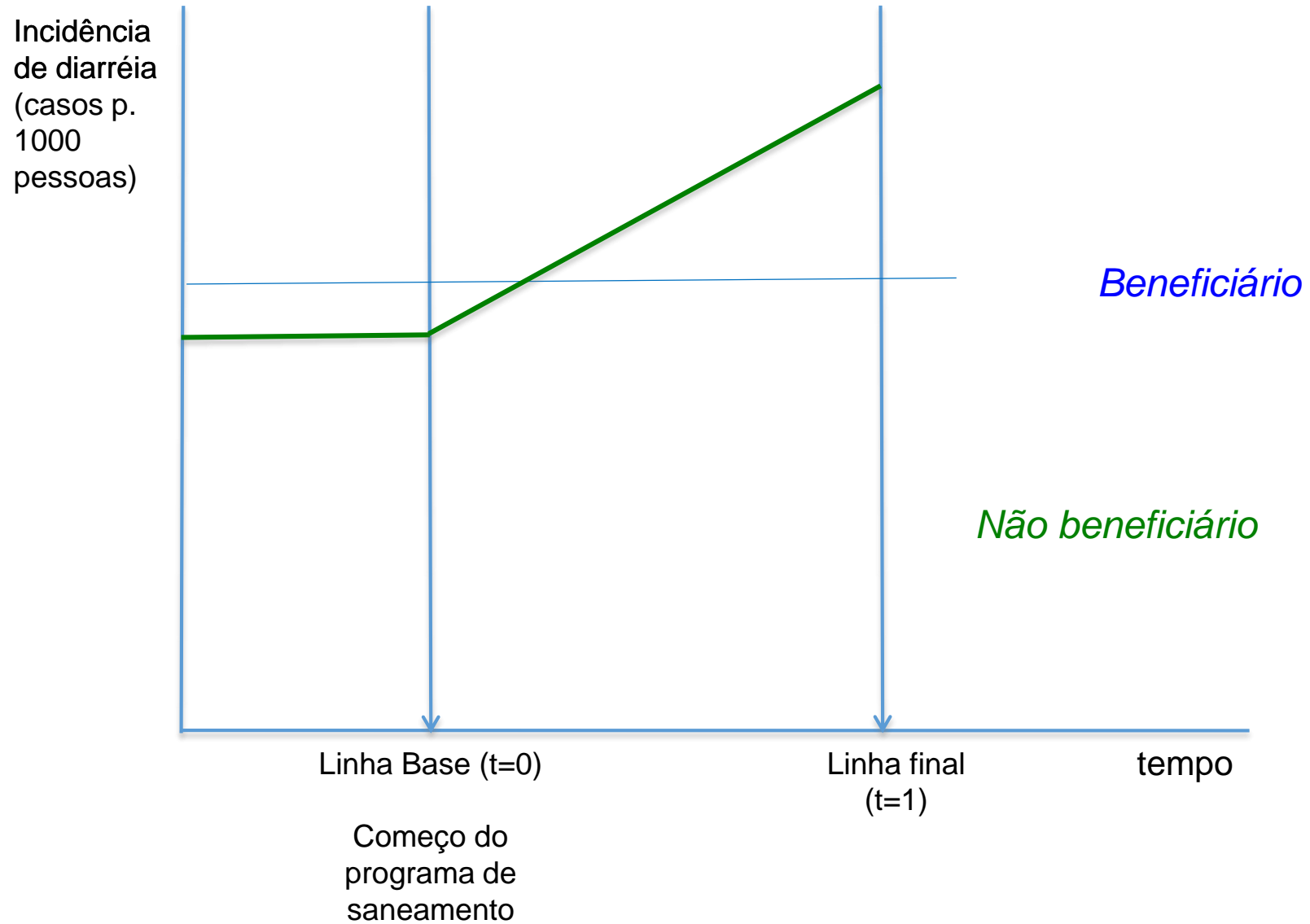


Essa afirmativa do jornal está correta? Por quê?

Exercício



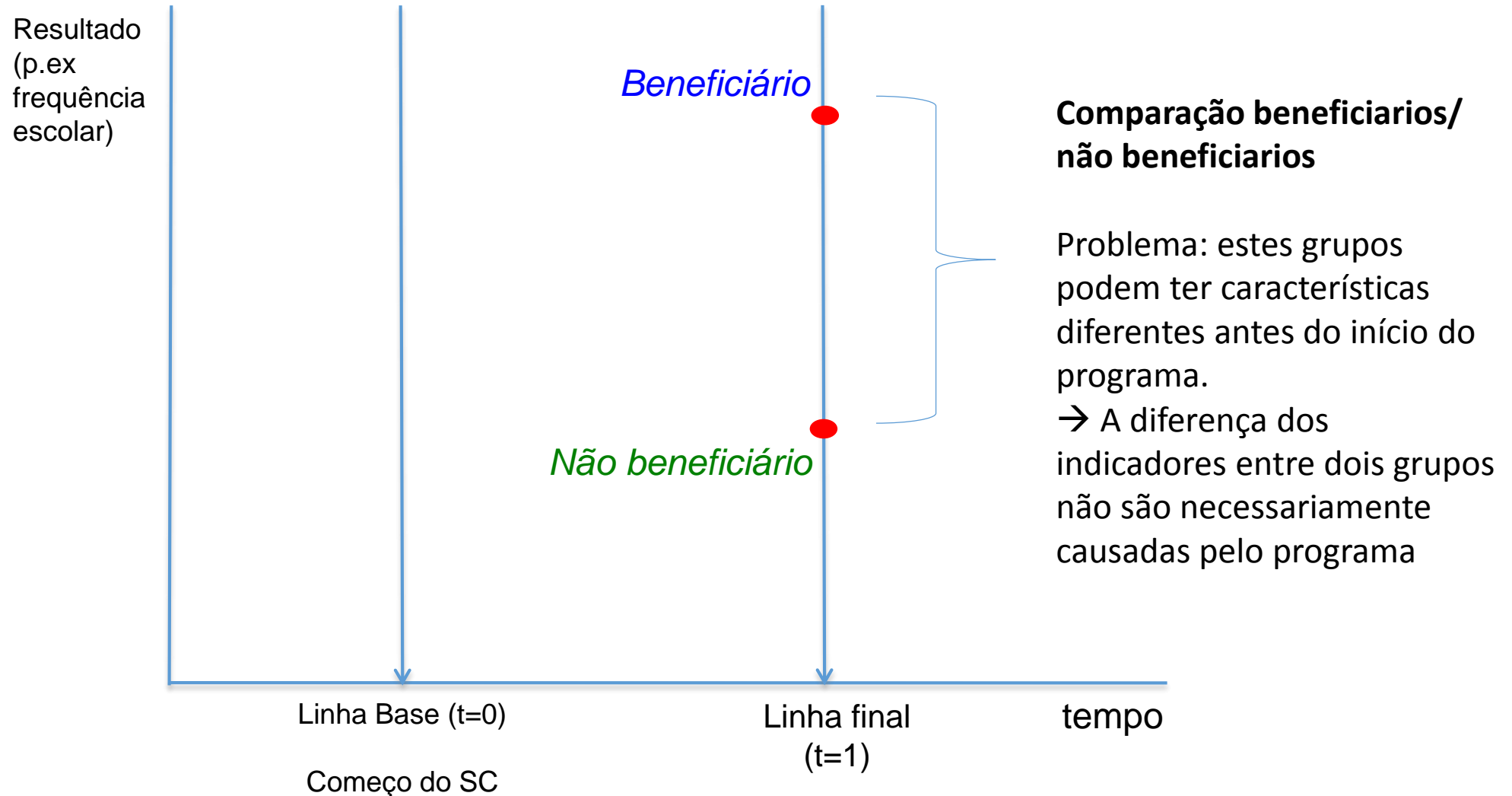
Exercício



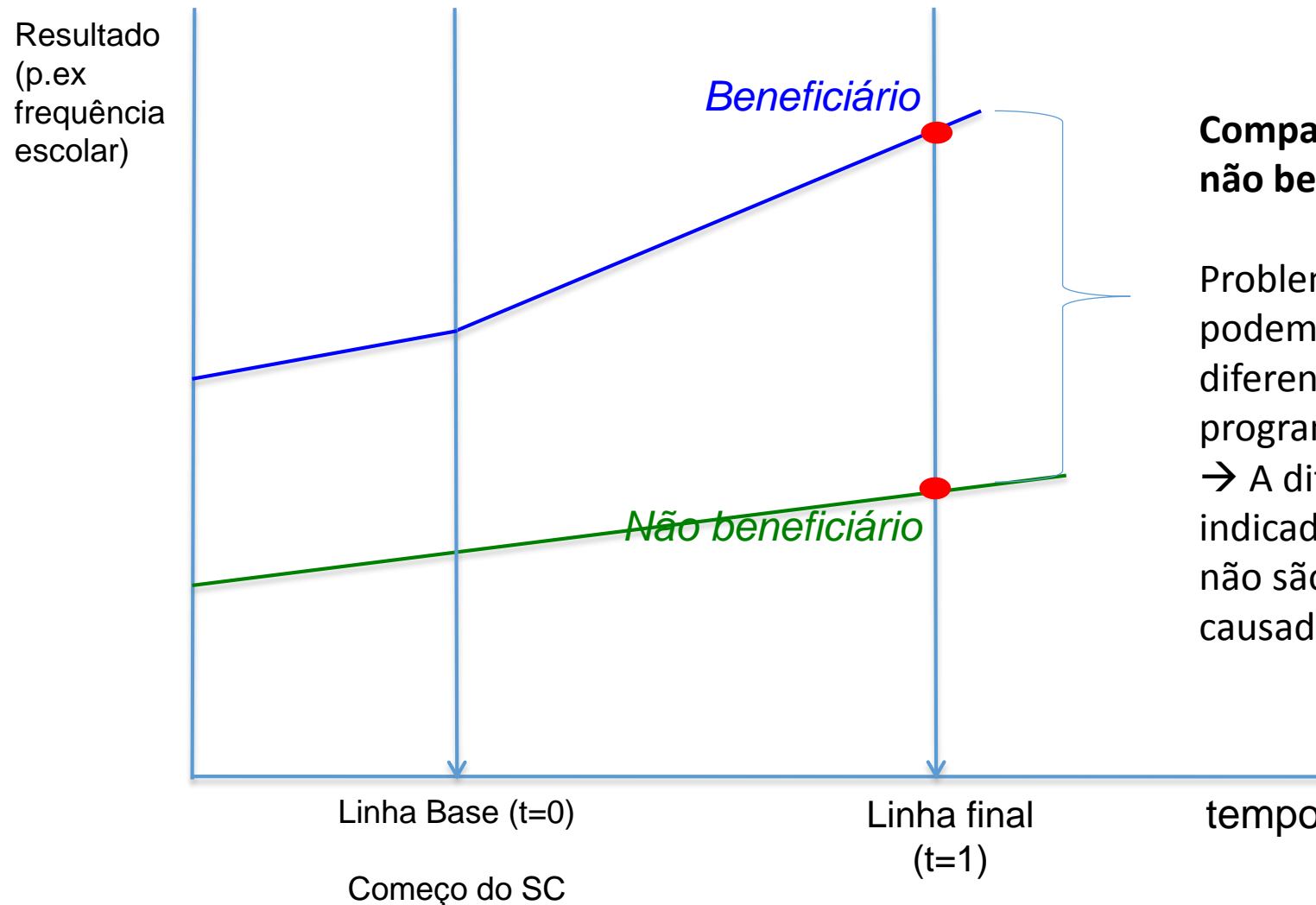
Avaliação de impacto – como?

Por que não
comparar
simplesmente
beneficiários e
não
beneficiários?

Avaliação de impacto – como?



Avaliação de impacto – como?

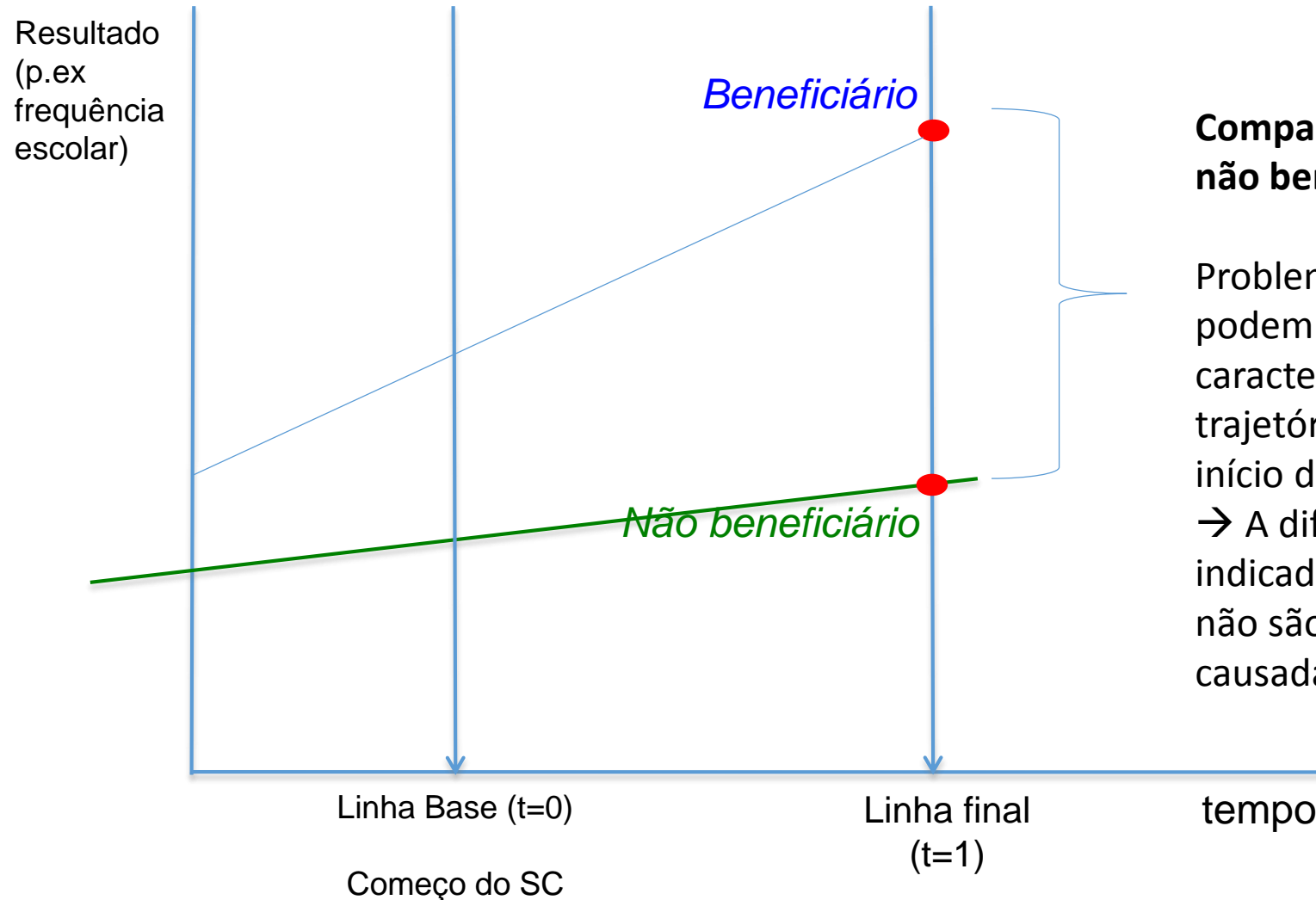


Comparação beneficiarios/ não beneficiarios

Problema: estes grupos podem ter características diferentes antes do início do programa.

→ A diferença dos indicadores entre dois grupos não são necessariamente causadas pelo programa

Avaliação de impacto – como?

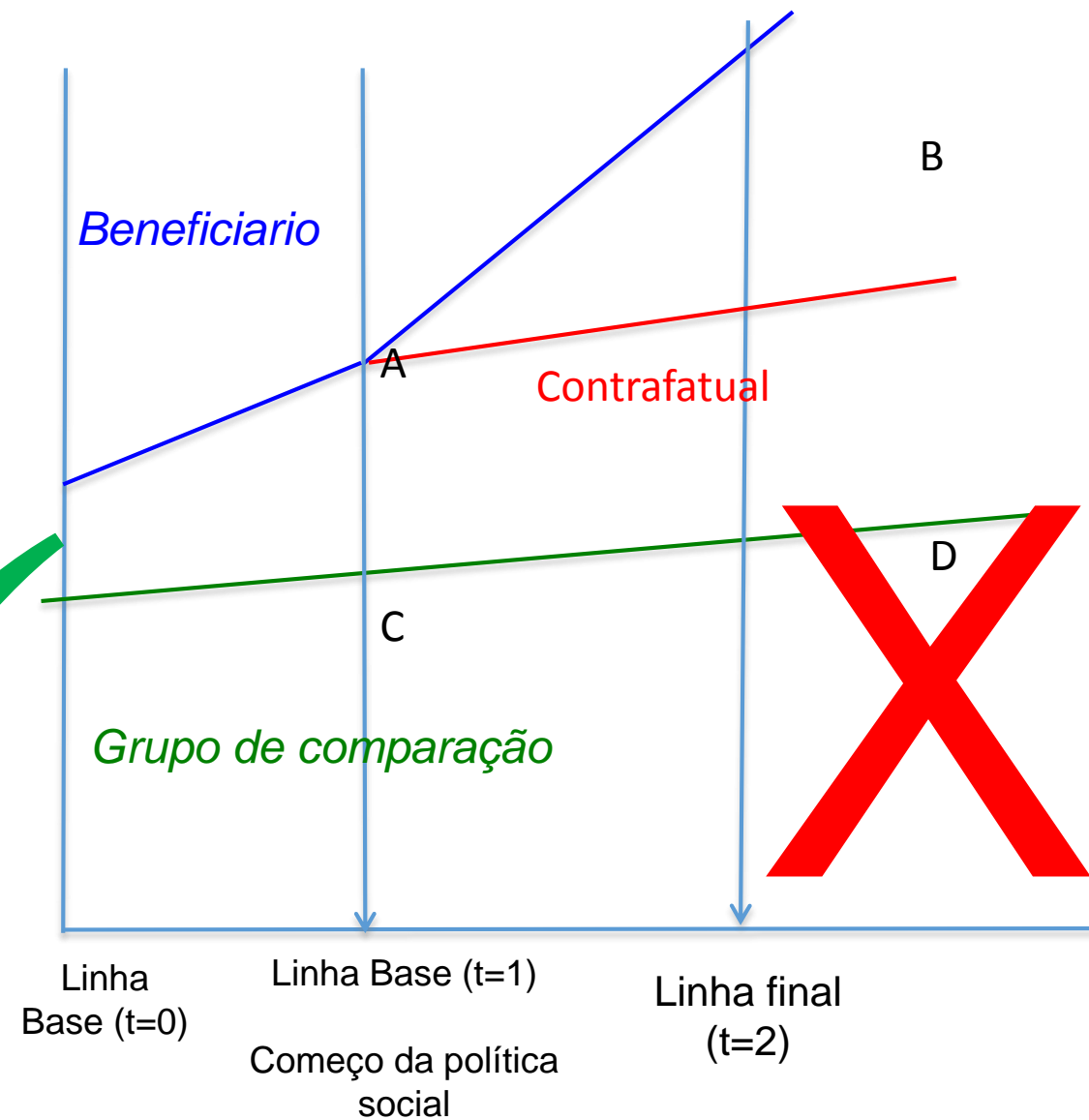
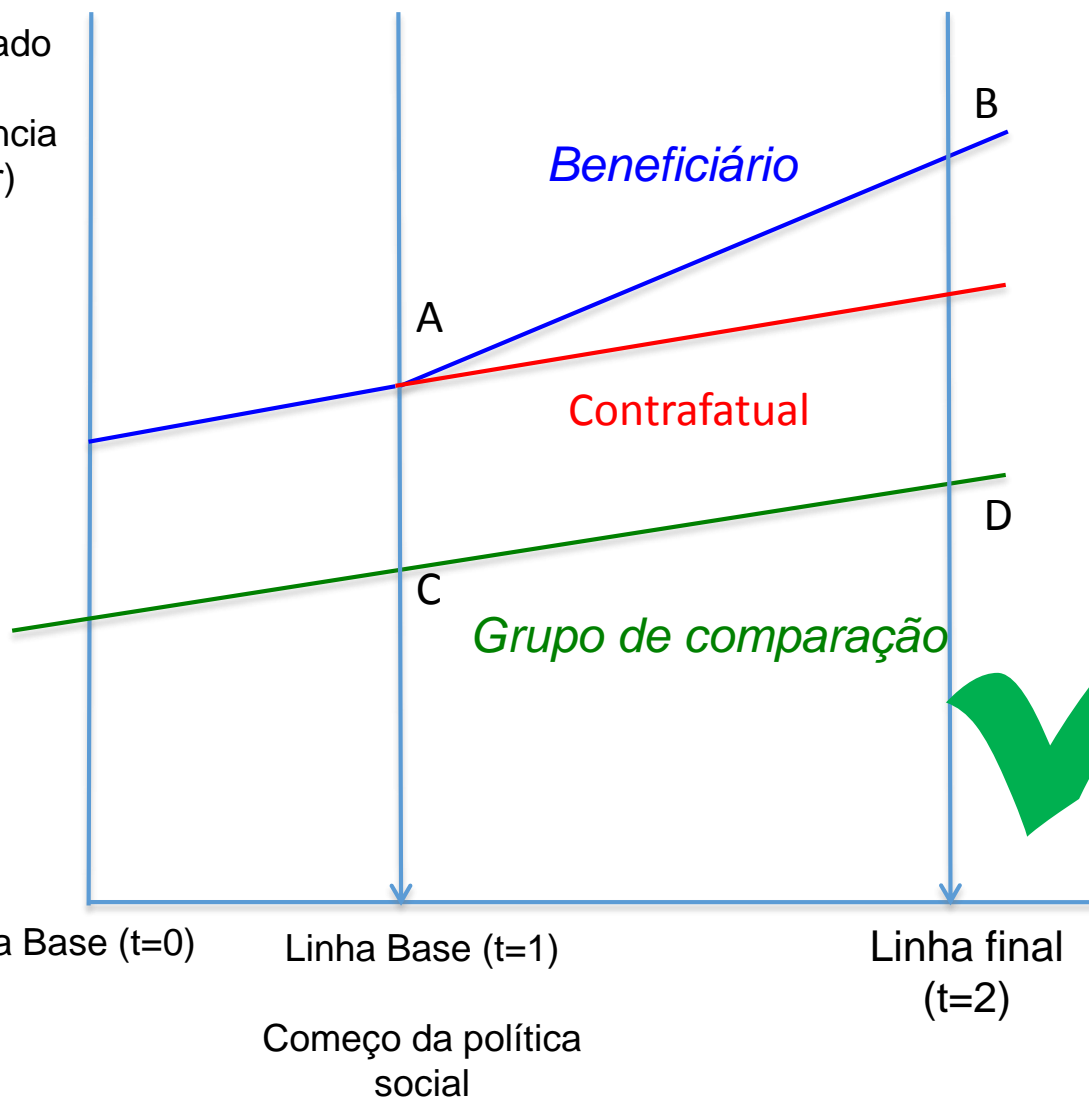


Comparação beneficiarios/ não beneficiarios

Problema: estes grupos podem ter não só características diferentes, mas trajetórias diferentes antes do início do programa.
→ A diferença dos indicadores entre dois grupos não são necessariamente causadas pelo programa

“Tendências similares” – Antes da Intervenção

Resultado
(p.ex
frequência
escolar)



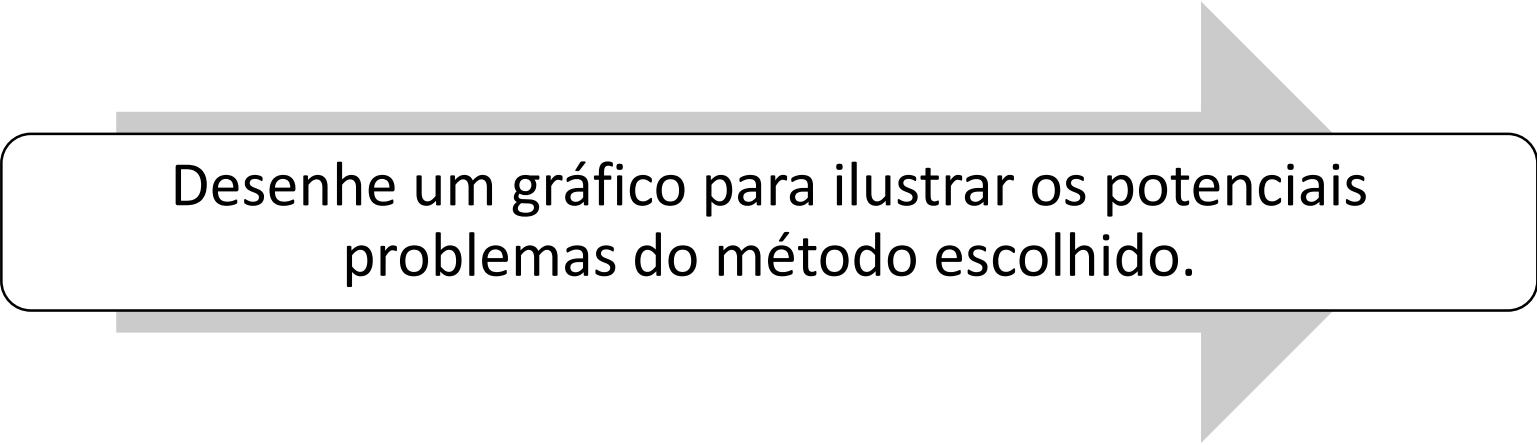
Como testar a hipótese da tendência comum

- Dois valores de linha de base (observe ambos grupos por 2 períodos anteriores à introdução do tratamento)
 - Vários períodos: checar se as tendências antes do tratameto são coincidentes entre os dois grupos.
 - Use um grupo de comparação diferente (o resultado deveria ser o mesmo)
 - Use uma variável de resultado que NÃO deve ser afetada pela intervenção.
 - Teste de Placebo: use um grupo de tratamento „falso“ (não afetado pelo programa) ou faça de conta que o programa foi implementado antes de sua implementação real.
- Se a estimativa de DD para estes placebos não for 0, então há um problema!

Avaliação de impacto – como?

Exercício: O prefeito da capital de Ourolândia implantou um programa de trabalhos públicos PASP OURO no terceiro ano de seu mandato. No ano seguinte, em campanha eleitoral, ele comparou a taxa de desemprego da sua cidade com cidades menores do interior da província para mostrar que o programa era um sucesso. Seria o interior da província bom grupo de comparação para avaliar o programa implantado na capital?

Exercício:



Desenhe um gráfico para ilustrar os potenciais problemas do método escolhido.

Como encontrar um grupo de comparação válido? (1)

- O método mais direto para construir um grupo de comparação é através de métodos de seleção aleatório (desenho experimental)
- Para isso, o acesso dos beneficiários ao programa deve ser feito de **forma aleatória**. Todos os indivíduos que possuem as características necessárias para acessar o programa são alocados de forma aleatória em dois grupos: beneficiários (grupo de tratamento) e não beneficiários (grupo de comparação)
- A alocação aleatória dos indivíduos em dois grupos faz com estes tenham as mesmas características (observáveis e não observáveis), e portanto iriam **reagir (em média) ao acesso ao benefício (ou ao não acesso) da mesma maneira**.

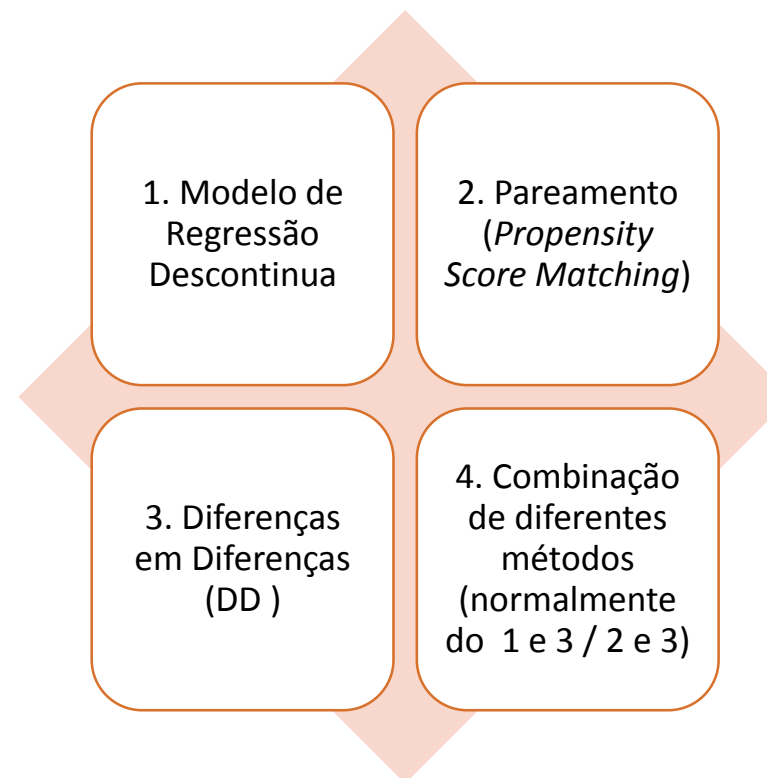
Como encontrar um grupo de comparação válido? (2)

E quando o formato do programa não prevê uma alocação aleatória dos beneficiários?

A segunda opção para selecionar um grupo de comparação é aplicar uma metodologia quase-experimental, que usa métodos econométricos que visam assegurar resultados semelhantes a um método experimental (isto é, que os grupos possuem características semelhantes).

Desafio: hipóteses não verificáveis

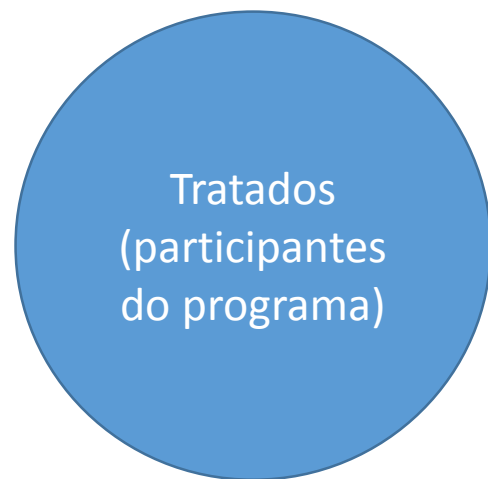
Métodos de Avaliação de Impacto



Estimar o impacto – a ideia básica (1)

Impacto

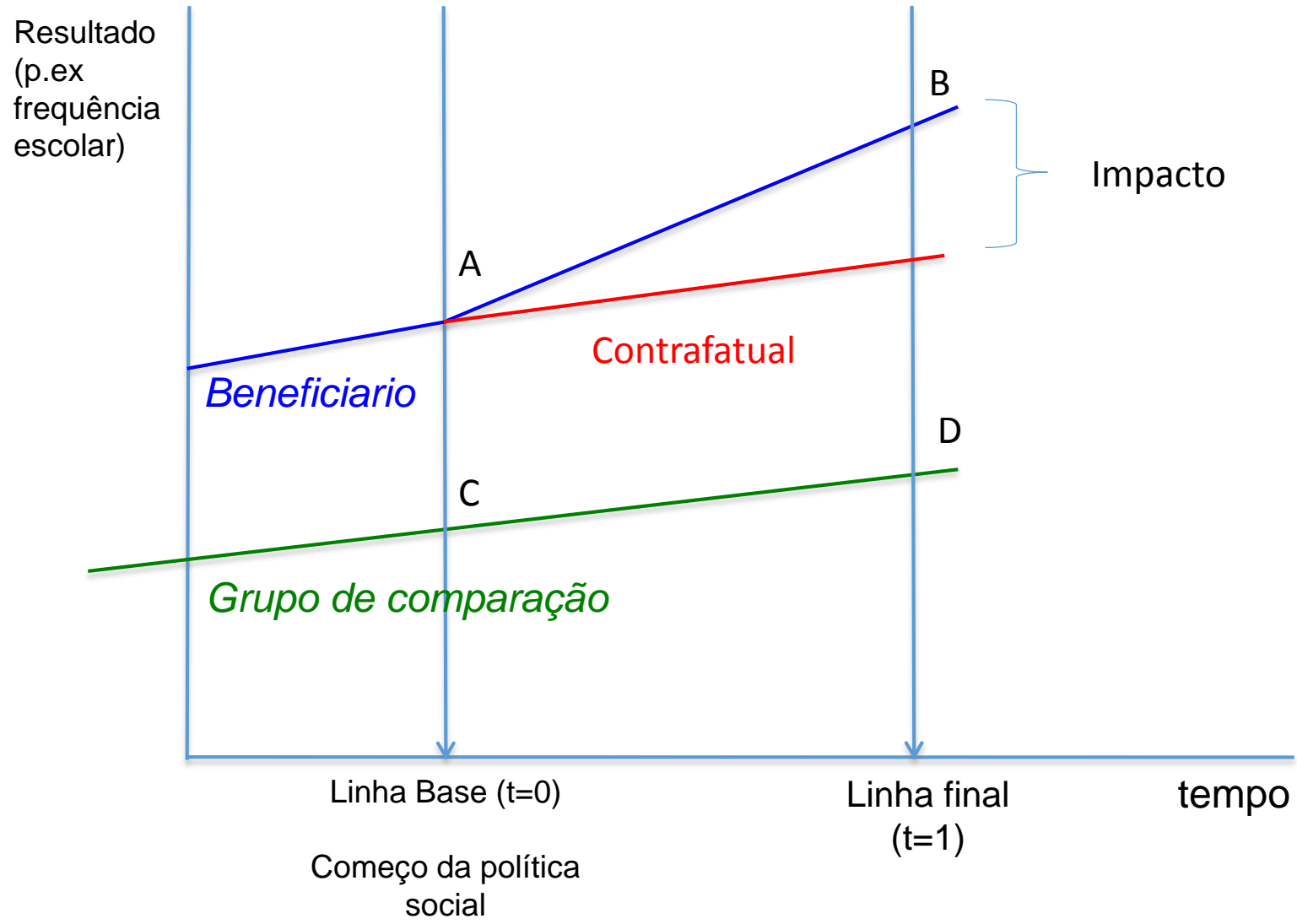
Depois de ter construído um grupo de comparação adequado, o impacto de um programa pode ser estimado comparando:



De que forma comparar?

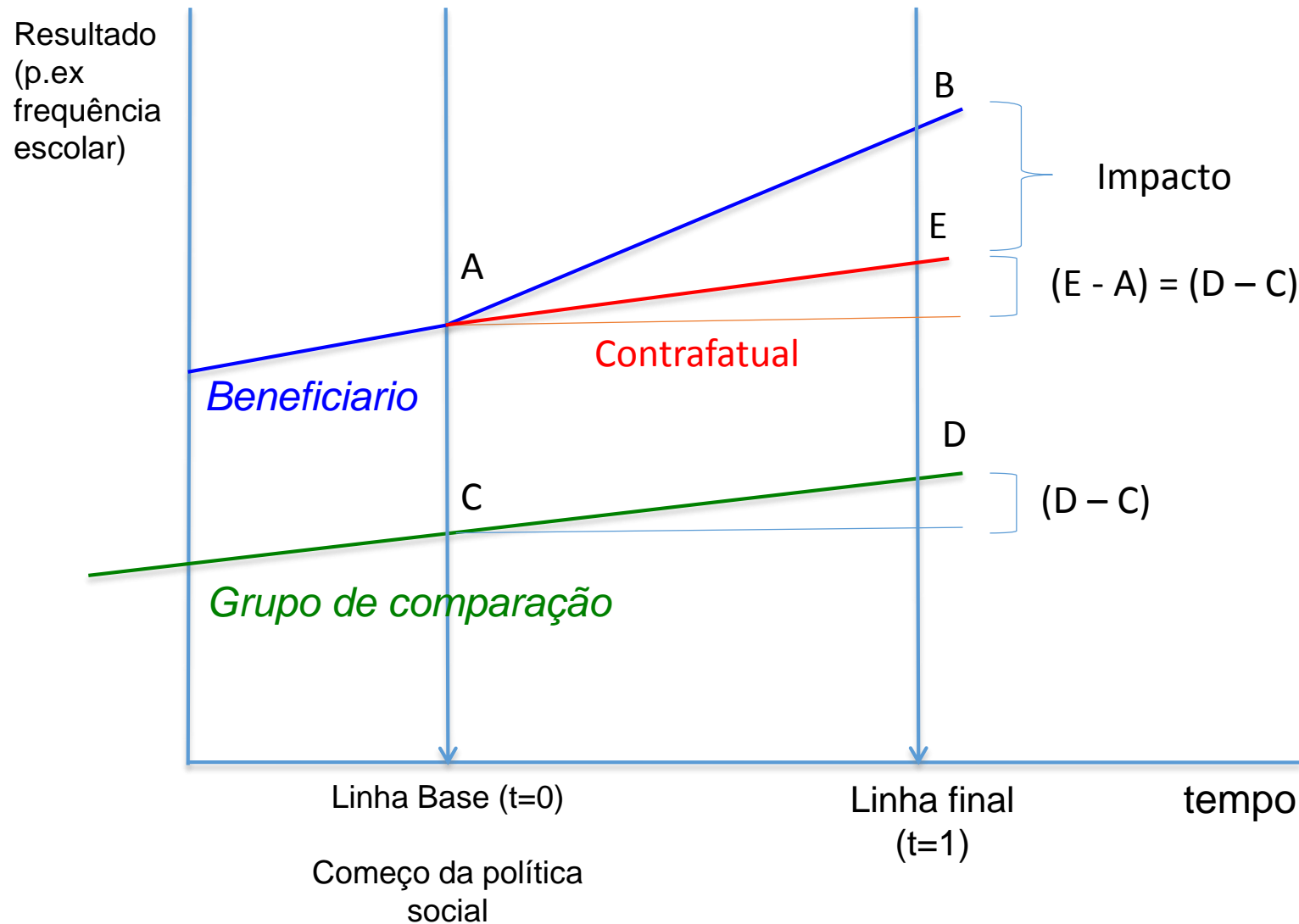
- ❖ Geralmente: Compara-se a situação antes/depois dos tratados com a situação antes/depois do grupo de comparação (DD).
- ❖ Alguns estudos podem comparar a situação 'depois' dos tratados com a situação 'depois' no grupo de comparação, mas precisam fazer hipóteses não testáveis mais fortes do que quando há uma linha de base para estabelecer a relação causal.

Estimar o impacto – ideia básica (2)



$$\text{Impacto} = \underbrace{(B - A)}_{\text{Mudança na situação dos tratados}} - \underbrace{(D - C)}_{\text{Mudança na situação do grupo de comparação}}$$

Estimar o impacto – ideia básica (2)



$$\text{Impacto} = \underbrace{(B - A)}_{\text{Mudança na situação dos tratados}} - \underbrace{(D - C)}_{\text{Mudança na situação do grupo de comparação}}$$

Estimar o impacto – notação formal (1)

Na prática, se compara a média de um indicador de interesse entre o grupo de tratamento com a **média no grupo de comparação**:

- ATT (Efeito Médio de Tratamento)

$$\sum_{i=1}^n \frac{(Y_{1i} | P = 1) - (Y_{1i} | P = 0)}{n}$$

- Contudo, $(Y_{1i} | P = 0)$ é a situação inexistente (beneficiários caso não recebesse o programa). Encontraremos este valor através do contrafactual.

Estimar o impacto – notação formal (2)

Assim, o ATT será estimado como

$$ATT = \sum_{i=1}^n \frac{(Y_{1i} | P=1) - (Y_{0i} | P=0)}{n}, \text{ onde}$$

$(Y_{0i} | P = 0)$ é o resultado do contrafactual

Estimar o impacto – notação formal (3)

- E para uma comparação antes e depois:

$$ATT = \sum_{i=1}^n \frac{[(Y_{1i} | P=1, t=1) - (Y_{1i} | P=1, t=0)] - [(Y_{0i} | P=0, t=1) - (Y_{0i} | P=0, t=0)]}{n}$$

2. Conceitos básicos de estatística para ler uma tabela com resultados de uma avaliação de impacto:

- Média aritmética
- Variância
- Desvio Padrão
- Significância estatística: p-valor

A. Média Aritmética

O impacto de um programa social pode ser estimado comparando as **médias aritméticas** (médias amostrais) entre grupo de tratamento e grupo de comparação.

Media aritmética da População = Somatório de todos os valores na população (Y_i) dividido pelo tamanho da população (N)

$$\mu = \frac{\sum_{i=1}^N Y_i}{N}$$

Media aritmética da População = Somatório de todos os valores na amostra (Y_i) dividido pelo tamanho da amostra (n)

$$\bar{Y} = \frac{\sum_{i=1}^n Y_i}{n}$$

A média aritmética visa a descrever o centro de uma distribuição

Exemplo:

O gráfico mostra as notas em uma amostra de alunos de um curso em Monitoria e Avaliação.

Por favor calcule o tamanho da amostra e a média amostral.

ID aluno (i)	Nota
1	4
2	8
3	8
4	3
5	7
6	7
7	5
8	5
9	5
10	9
11	2
12	9
13	2
14	7
15	4
16	5
17	6
18	5
19	4
20	4

B.) Variância populacional e variância amostral

A variância populacional (σ^2) descreve quão longe os membros de uma amostra estão da média da população

$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \mu)^2}{N}$$

x_i = valor do indicador x para o indivíduo i na população

μ = média aritmética do indicador x na população

N = tamanho da população

A variância amostral (S^2) estima quão longe os membros de uma amostra estão da média da “população”, mas utilizando-se os parâmetros de uma amostra. Há duas opções para se calcular este estimador.

$$S^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n} \quad \text{ou} \quad S^{2'} = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$$

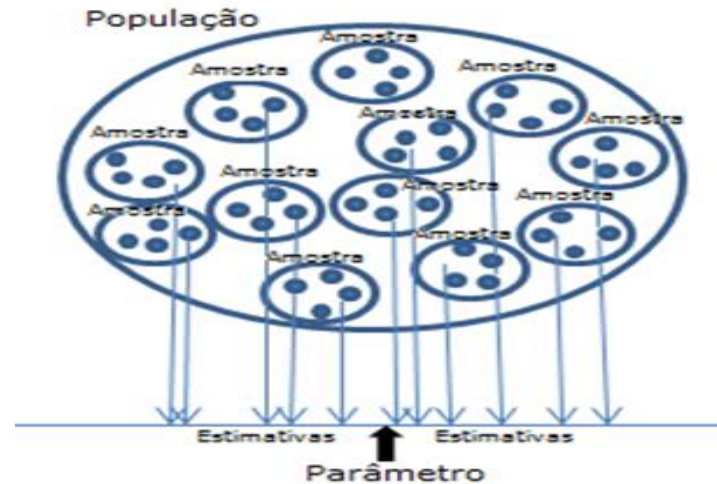
x_i = valor do indicador x para o indivíduo i na amostra

\bar{x} = média aritmética do indicador x na amostra

n = tamanho da amostra

Noção de distribuição amostral...

Distribuição de amostragem de um **estimador** (ou **estatística**) é a distribuição dos valores que o **estimador** assume para todas as possíveis amostras, da mesma dimensão, que se podem selecionar da população.



Estimador centrado Um estimador diz-se *centrado* quando a média das estimativas obtidas para todas as amostras possíveis (da mesma dimensão) que se podem extrair da população, segundo o esquema de amostragem considerado, coincide com o parâmetro a estimar. Quando se tem um estimador *centrado*, também se diz que é *não enviesado*.

Observe agora as escolas 2 e 3.

Calculando-se a média das duas amostras, alcançaremos o mesmo valor de média: 5,45

Pode-se afirmar que os alunos da escola 2 tem o mesmo desempenho que os alunos da escola 3 em matemática?

É claro que na Escola 2 os alunos tem notas mais distantes de 5,45, enquanto na escola 3 quase toda a amostra tem notas mais próximas da média.

Exercício: Calcule a variância das duas amostras

Escola 2		Escola 3	
ID aluno (i)	Nota	ID aluno (i)	Nota
1	8	21	6
2	7	22	5
3	8	23	6
4	10	24	6
5	8	25	5
6	7	26	5
7	1	27	5
8	1	28	5
9	1	29	5
10	7	30	6
11	10	31	5
12	1	32	6
13	3	33	5
14	0	34	5
15	1	35	6
16	4	36	5
17	4	37	6
18	9	38	5
19	10	39	6
20	9	40	6

ID aluno (i)	Nota	ID aluno (i)	Nota
1	8	21	6
2	7	22	5
3	8	23	6
4	10	24	6
5	8	25	5
6	7	26	5
7	1	27	5
8	1	28	5
9	1	29	5
10	7	30	6
11	10	31	5
12	1	32	6
13	3	33	5
14	0	34	5
15	1	35	6
16	4	36	5
17	4	37	6
18	9	38	5
19	10	39	6
20	9	40	6

$$\text{Var}(\text{escola 2}) = \frac{(8-5,45)^2 + \dots + (9-5,45)^2}{20-1} = 13,3$$

$$\text{Var}(\text{escola 3}) = \frac{(6-5,45)^2 + \dots + (6-5,45)^2}{20-1} = 0,26$$

C. Desvio padrão

- O **desvio padrão** é a raiz quadrada da variância
 - Deve-se entender a **variância** e o **desvio padrão** como medidas complementares à média, que nos ajudam a descrever uma amostra.

$$DP = \sqrt{\text{Variância}} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$$

- O desvio padrão tem a mesma 'escala' que os valores originais (p.ex notas)

ID aluno (i)	Nota	ID aluno (i)	Nota
1	8	21	6
2	7	22	5
3	8	23	6
4	10	24	6
5	8	25	5
6	7	26	5
7	1	27	5
8	1	28	5
9	1	29	5
10	7	30	6
11	10	31	5
12	1	32	6
13	3	33	5
14	0	34	5
15	1	35	6
16	4	36	5
17	4	37	6
18	9	38	5
19	10	39	6
20	9	40	6

Exercício: Calcular o desvio padrão das duas amostras.

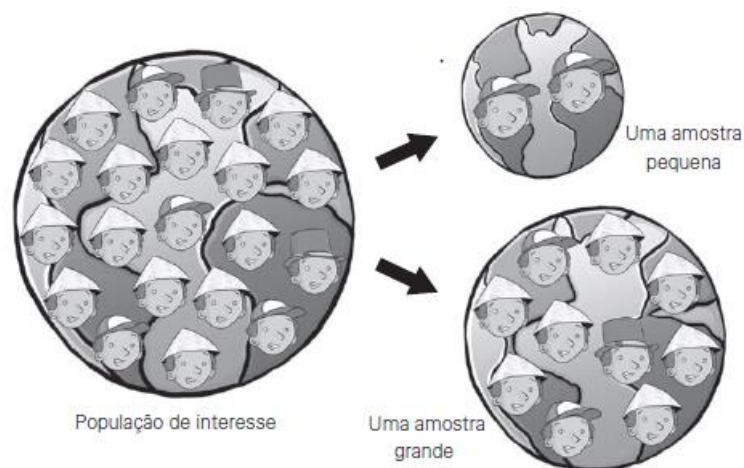
$$DP (escola 2) = \sqrt{Var (escola 2)} = \sqrt{Var (13,3)} = 3,647$$

$$DP (escola 3) = \sqrt{Var (escola 3)} = \sqrt{Var (0,26)} = 0,51$$

3. Inquéritos domiciliares e amostragem

- Para avaliar o impacto de um programa quase sempre vamos trabalhar com uma amostra da população de interesse (não a população inteira)

Figura 11.1 Uma Amostra Grande Representará Melhor a População



Fonte: Autores.

P.ex.: população pobre

Avaliação de Impacto na Prática

P.ex.: população pobre

Amostra da avaliação (beneficiários e não beneficiários)

3. Inquéritos domiciliares e amostragem

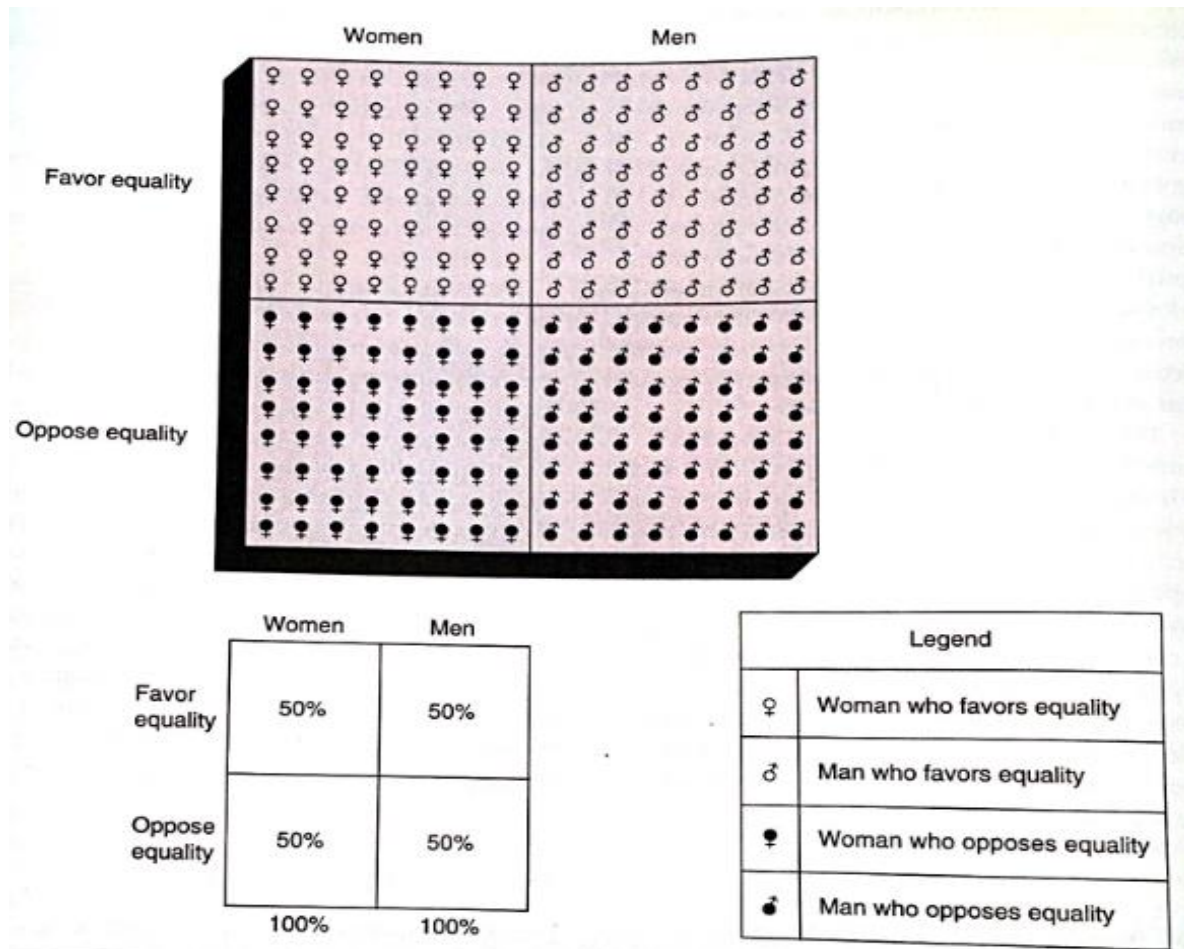
- É importante que a amostra da avaliação seja representativa da população

→ Amostragem aleatória

- Amostra grande: Mais representativa da população
- Amostra pequena: Menores custos de coleta de dado

3. Inquéritos domiciliares e amostragem

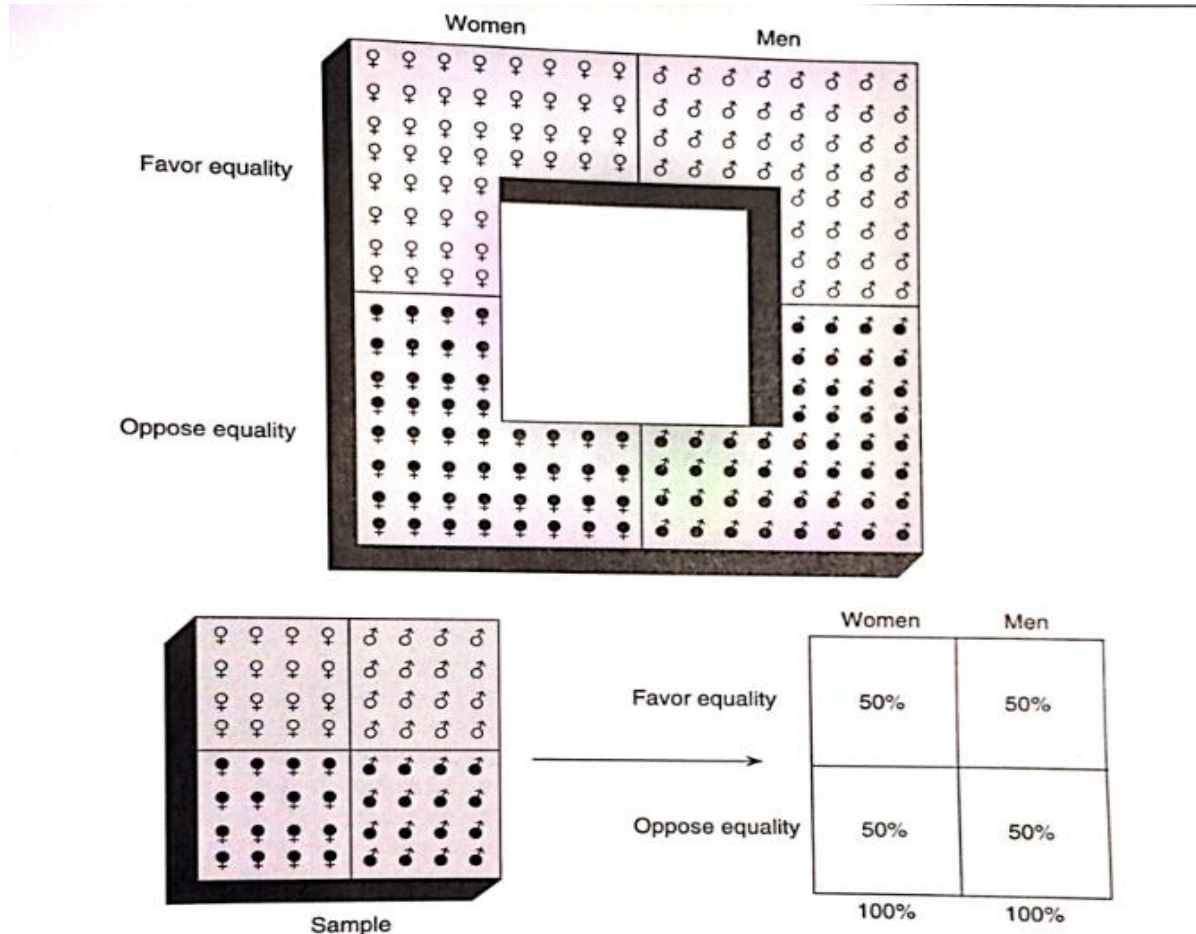
1. População hipotética de homens e mulheres que são a favor ou contra igualdade de gênero



- O diagrama apresenta a população hipotética de 256 pessoas. Metade são homens, metade são mulheres.
- Opinião sobre igualdade de gênero:
 - Os que são a favor tem círculos abertos
 - Os que são contra tem círculos fechados
- Questão: qual a relação entre gênero e a opinião sobre igualdade entre homens e mulheres?
- Hipótese: mulheres tendem a ser mais favoráveis a igualdade de gênero do que os homens.

3. Inquéritos domiciliares e amostragem

2. Amostra representativa



Amostra: ¼ da população total

16 indivíduos de cada tipo: metade homens e metade mulheres;
Metade é a favor da igualdade de gênero, metade é contra.

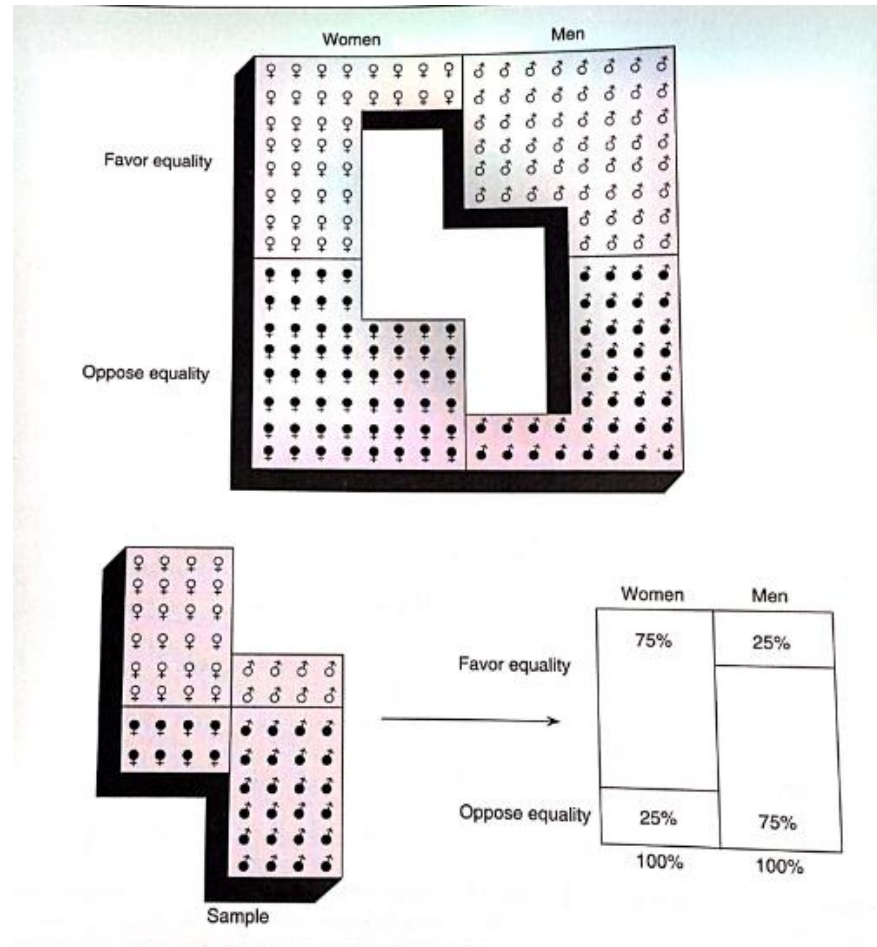
Com os resultados desta investigação, assume-se que não há relação entre a opinião sobre igualdade de gênero e o sexo do indivíduo.

Contudo, inquéritos amostrais são raramente representações perfeitas da população estudada (Babbie, 2004:462).

Veja o exemplo a seguir:

3. Inquéritos domiciliares e amostragem

3. Amostra não representativa



Neste caso, observa-se que a amostra selecionou muitas mulheres a favor e muitos homens contra a igualdade de gênero.

Ao se observar esta amostra (da mesma população) é possível é o avaliador chegue a outra conclusão: de que a variável gênero **influencia** a opinião sobre igualdade de gênero.

3. Inquéritos domiciliares e amostragem

- Na prática, nunca sabemos qual é o valor real de um indicador na população. Sempre precisamos analisar com métodos estatísticos
 - se uma relação observada reflete a situação da população
 - ou se essa relação pode ser equivocada, devido a problemas na amostragem

→ Significância estatística

4. Significância estatística

- Ideia intuitiva (\neq definição acadêmica !)
- A significância estatística descreve a probabilidade que uma característica observada numa amostra é também observável na população
- Geralmente são usadas duas ferramentas para ‘medir’ a significância estatística
 - O ‘p-value’ (valor p)
 - O erro padrão da estimativa ou estimador

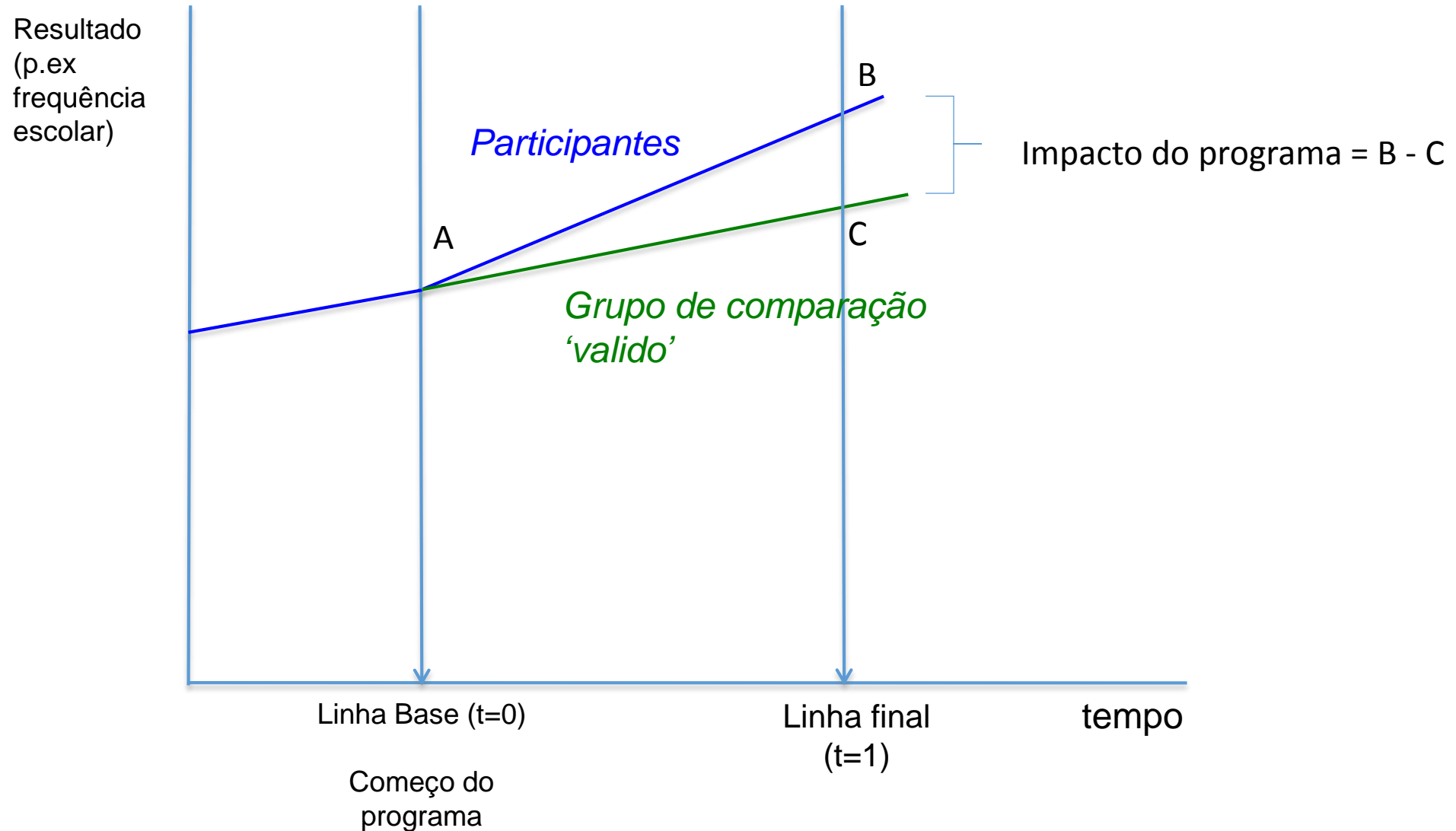
4. Significância estatística: o p-value

- O 'p-value' (valor p) mede a significância estatística: Ele mostra a probabilidade que uma relação observada na amostra não seja observada na população.
- P.ex. p-value = 0.05
 - Interpretação: Com uma probabilidade de 5%, a característica observada na amostra, não é observada na população.
 - E ao mesmo tempo: Com uma probabilidade de 95% a característica observada na amostra é observada na população.

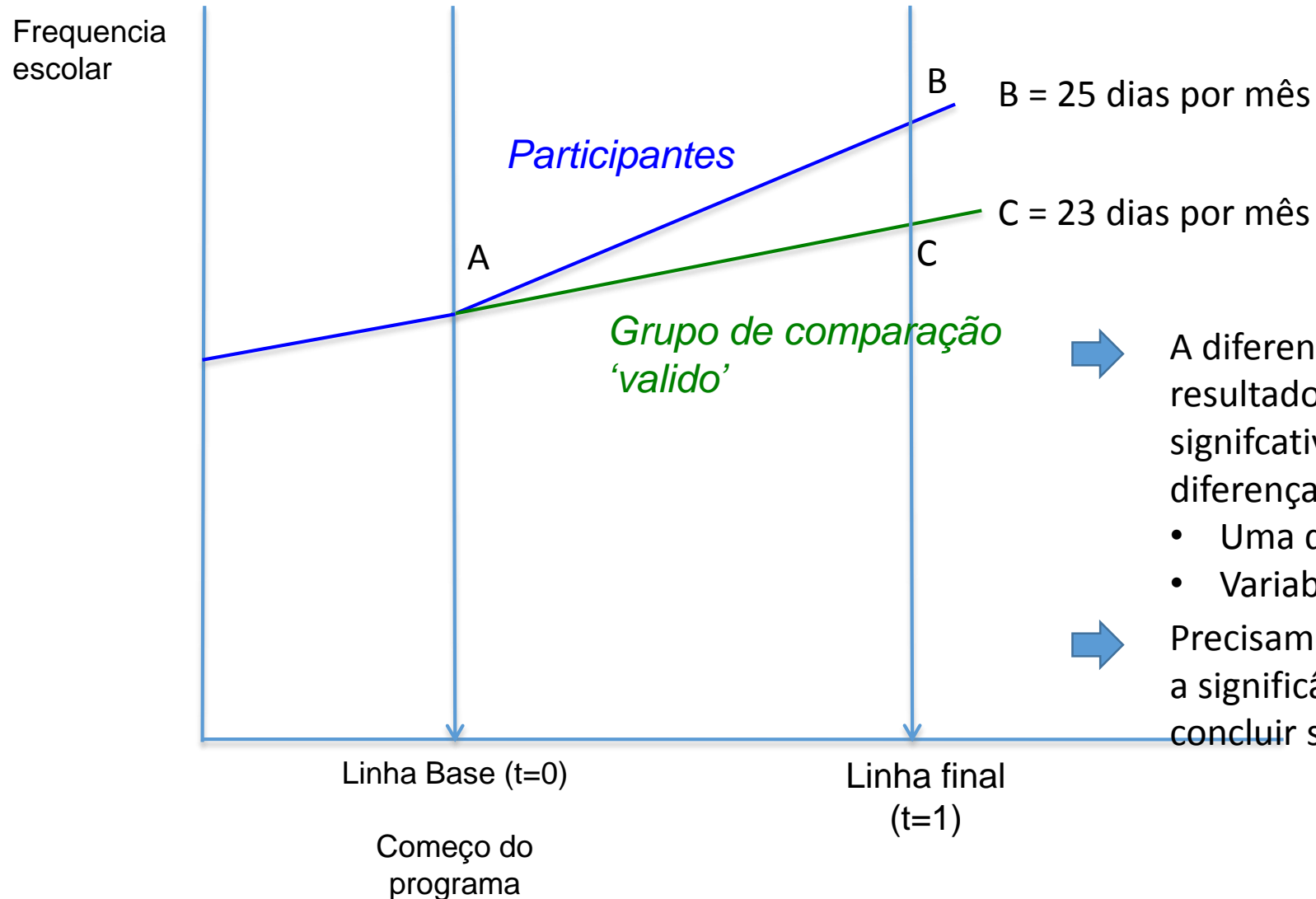
4. Significância estatística

- Nas avaliações de impacto um p-value ≤ 0.05 é, em geral, considerado aceitável.
- Com esse valor temos uma significância estatística ao nível de 5%
→ sabemos com 95% de certeza que o resultado observado também está presente na população.

Exemplo – Significância Estatística



Exemplo – Significância Estatística



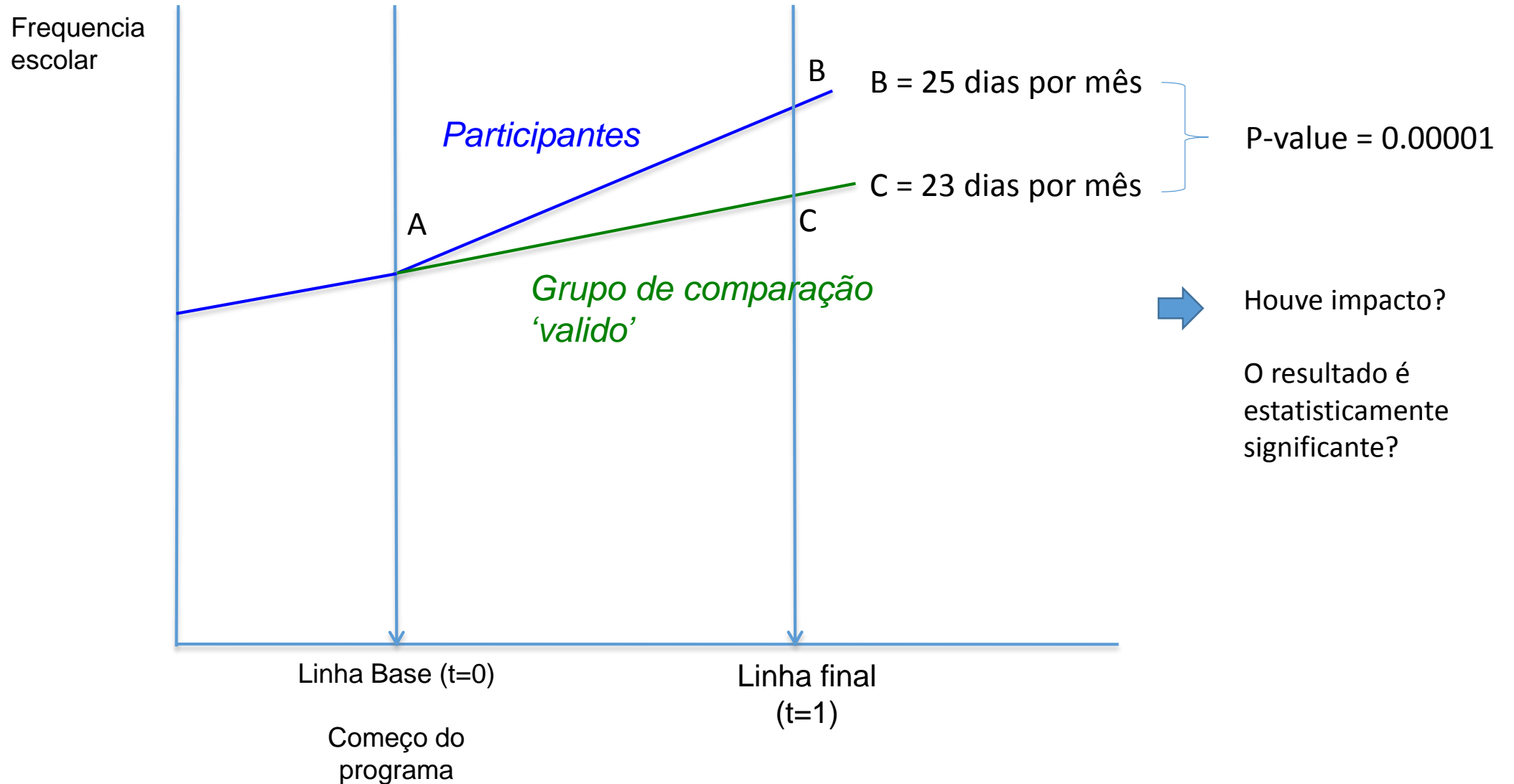
A diferença como tal não mostra se o resultado é estadisticamente significativo. Não sabemos se a diferença è devida a

- Uma diferença real na população
- Variabilidade amostral

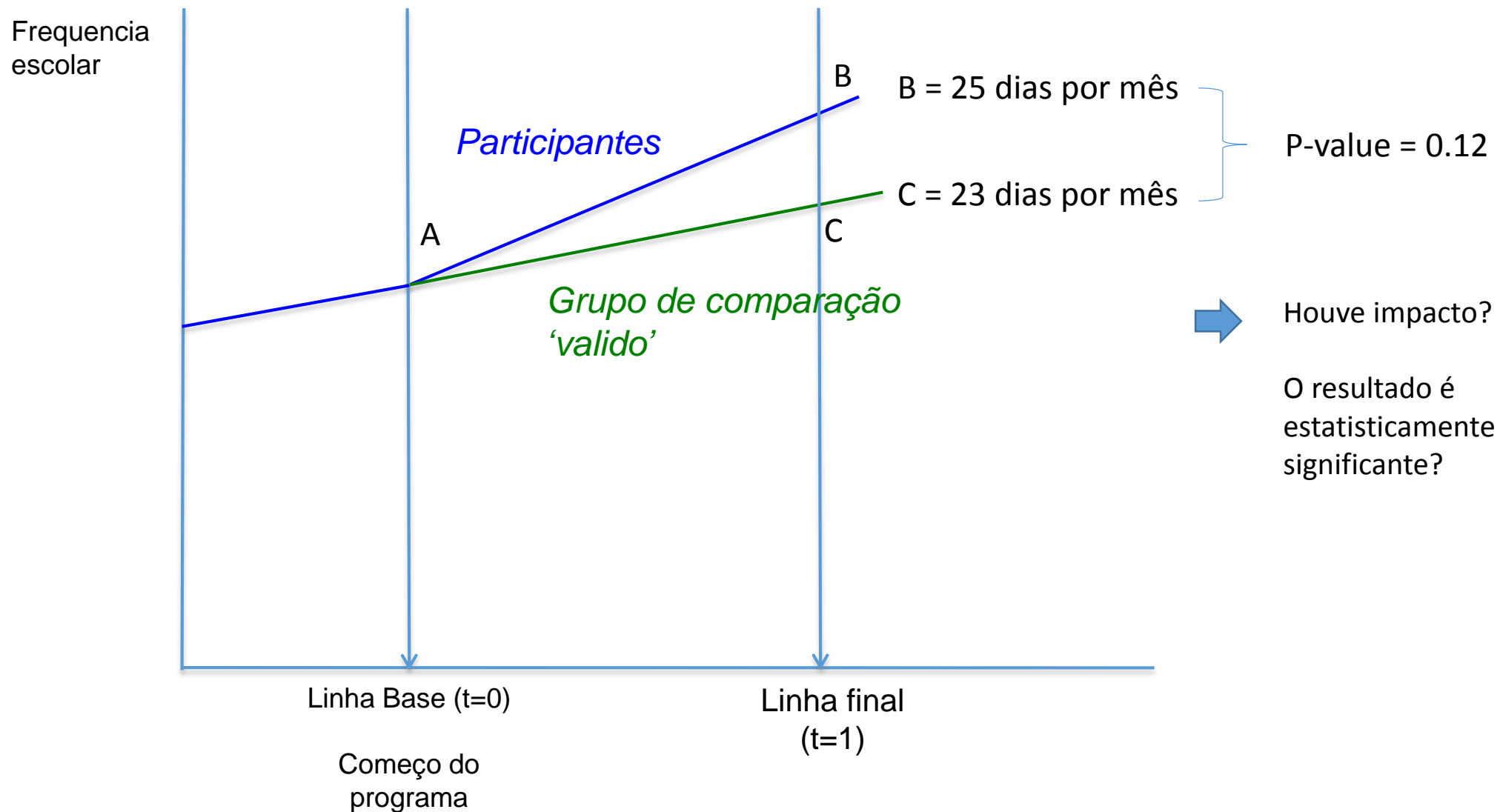


Precisamos conhecer o p-value e a significância estatística para concluir se há impacto

Exemplo – Significância Estatística



Exemplo – Significância Estatística



4. Significância estatística

Nos relatórios de avaliação de impacto, o nível de significância estatística é frequentemente destacado com asteriscos (*)

- Notações diferentes dependendo do autor (é necessário verificar as legendas das tabelas respectivas)
- Geralmente
 - *** = estatisticamente significativo ao nível de 1%
 - ** = estatisticamente significativo ao nível de 5%
 - * = estatisticamente significativo ao nível de 10%

4. Significância estatística

Table 5. Effect on School Participation

	Mean in Control Group	Impact of LCT to Fathers
	(1)	(2)
<u>Panel A: Household sample</u>		
Attending school by end of year 2, among those 6-15 at baseline	0.737 [0.44]	0.074 (0.016)***
Dropped out by end of year 2, among those enrolled in grades 1-4 at baseline	0.1 [0.3]	-0.076 (0.012)***
Attending school by end of year 2 if had dropped out at any time before baseline	0.147 [0.355]	0.121 (0.041)***
Never Enrolled in school by end of year 2, among those 6-15 in year 0	0.035 [0.185]	-0.011 (0.008)

- Aumento de 7.4% na frequência escolar
- Estatisticamente significativo ao nível de 1%

- Diminuição de 1.1% da taxa de crianças que nunca estavam inscritos numa escola
- **Estatisticamente não significativo**

4. Significância estatística – erro padrão do estimador

A segunda ferramenta usada para descrever a significância estatística é o ‘erro padrão’ da estimativa.

O erro padrão nos ajuda a construir um ‘intervalo de confiança’ de possíveis valores que a estimativa poderia alcançar com 95% de confiança.

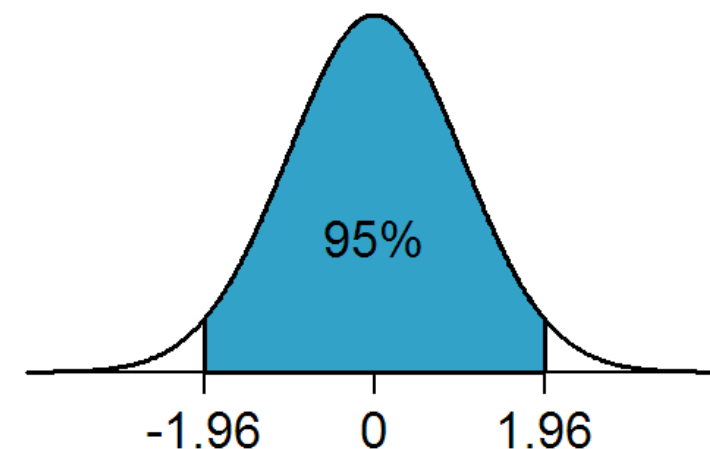
4. Significância estatística – erro padrão do estimador

Para um intervalo de confiança ao nível de 95%, sabemos que:

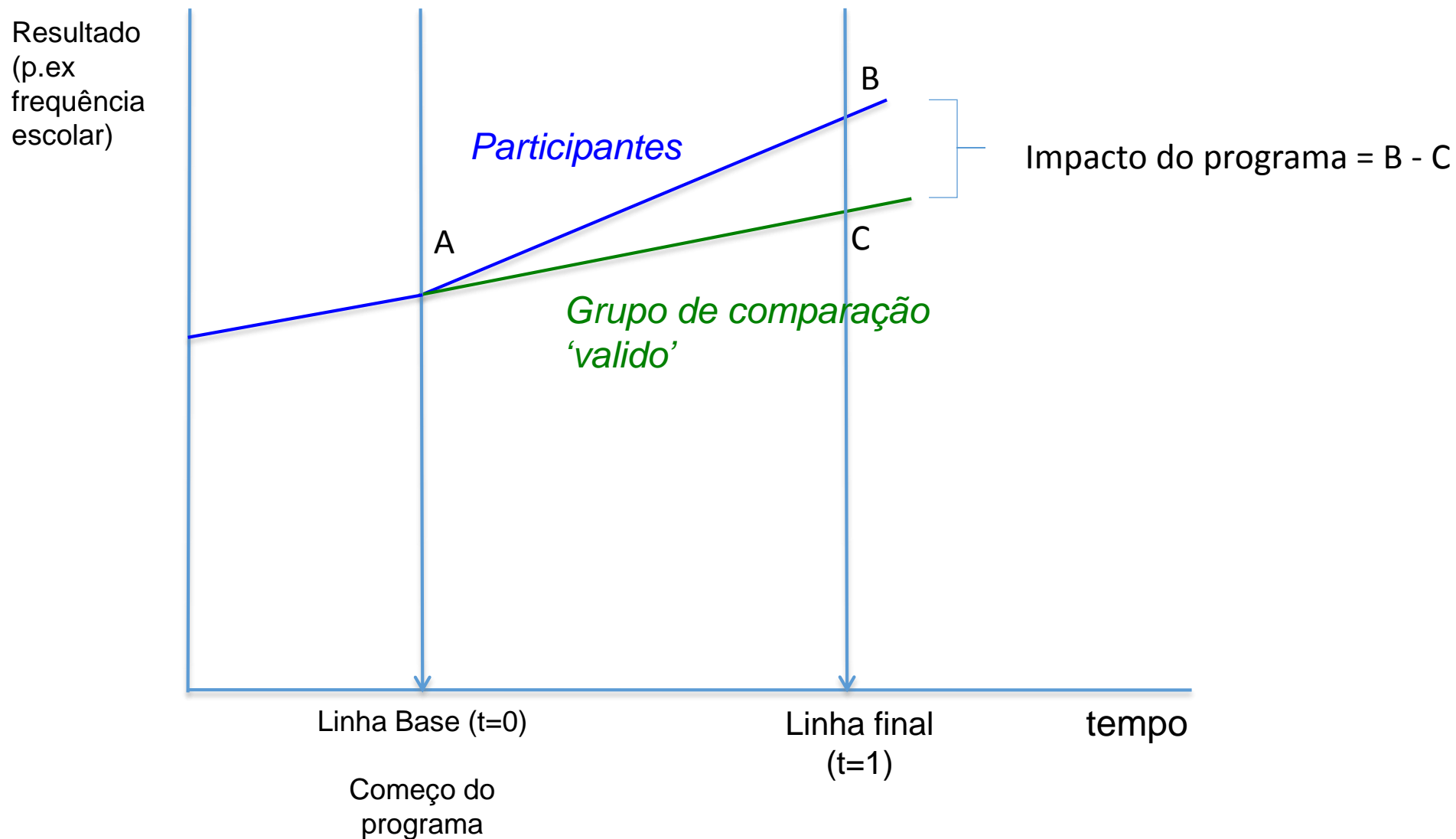
- Limite superior = Valor estimado + (1.96 * erro padrão)
- Limite inferior = Valor estimado – (1.96 * erro padrão)

Regra geral:

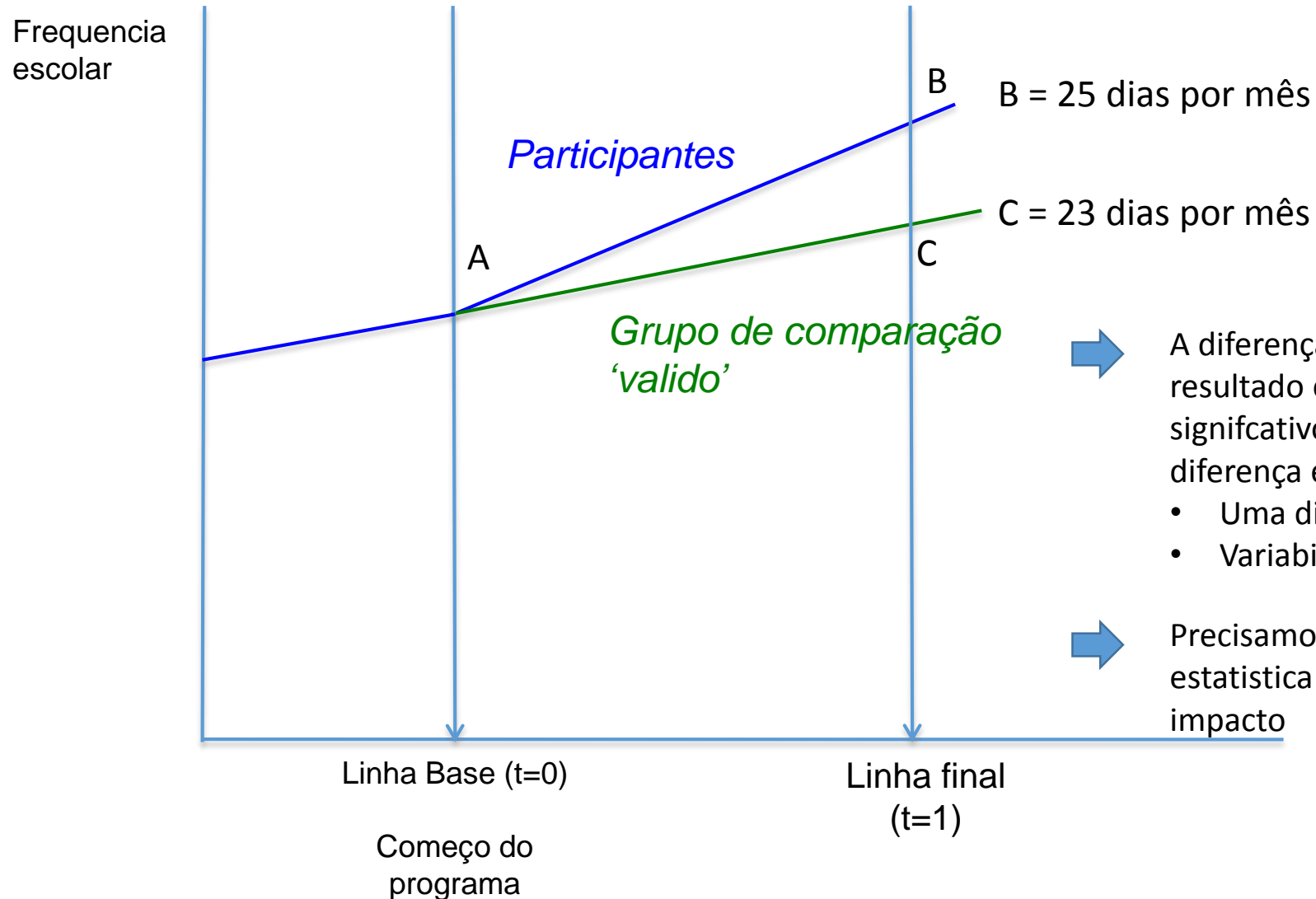
- Se nenhum valor dentro do intervalo de confiança é igual a 0, a estimação está estatisticamente significativa.



Exemplo – Significância Estatística



Exemplo – Significância Estatística



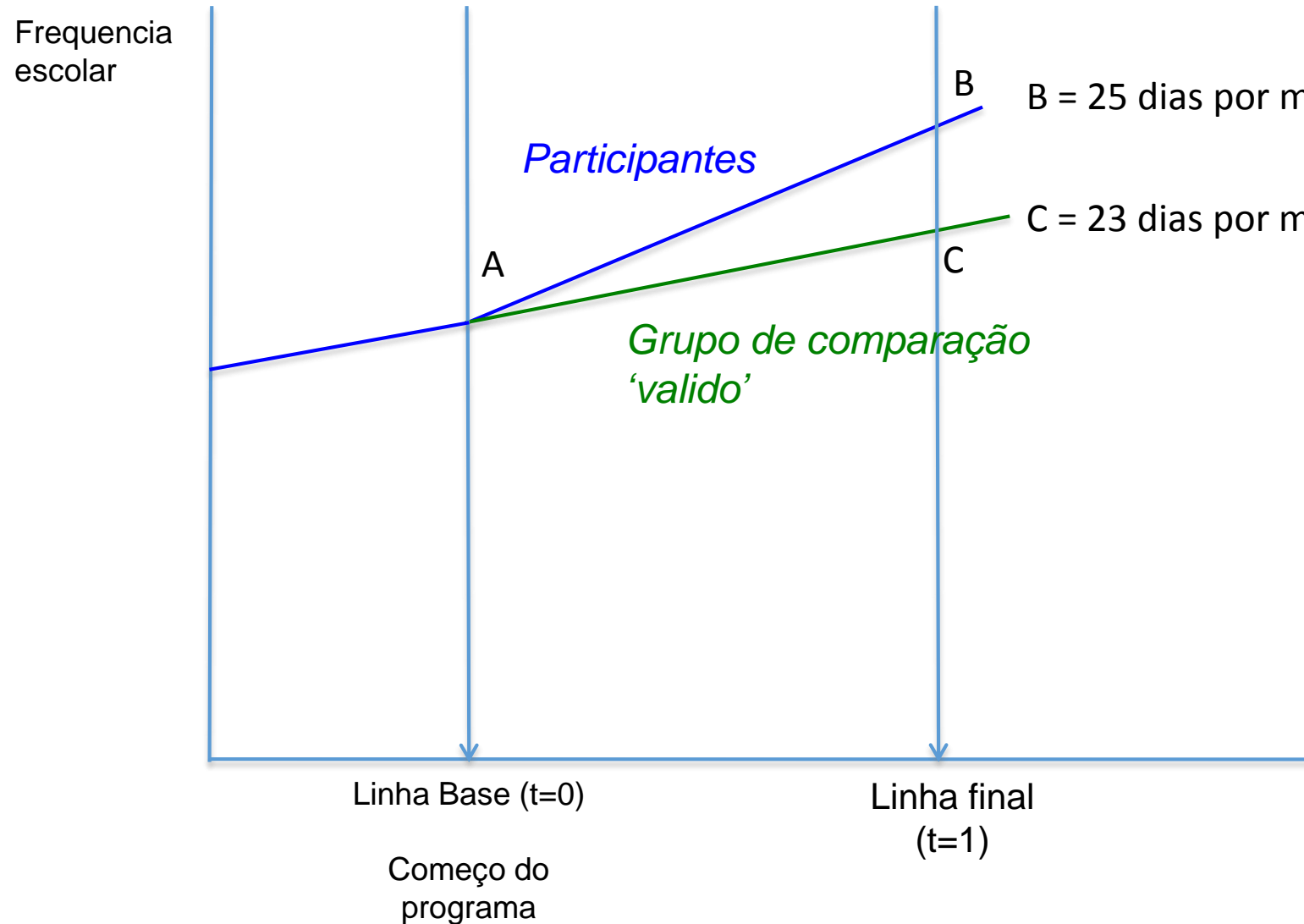
A diferença como tal não mostra se o resultado é estatisticamente significativo. Não sabemos se a diferença é devida a

- Uma diferença real na população
- Variabilidade amostral



Precisamos conhecer a significância estatística para concluir se há impacto

Exemplo – Significância Estatística



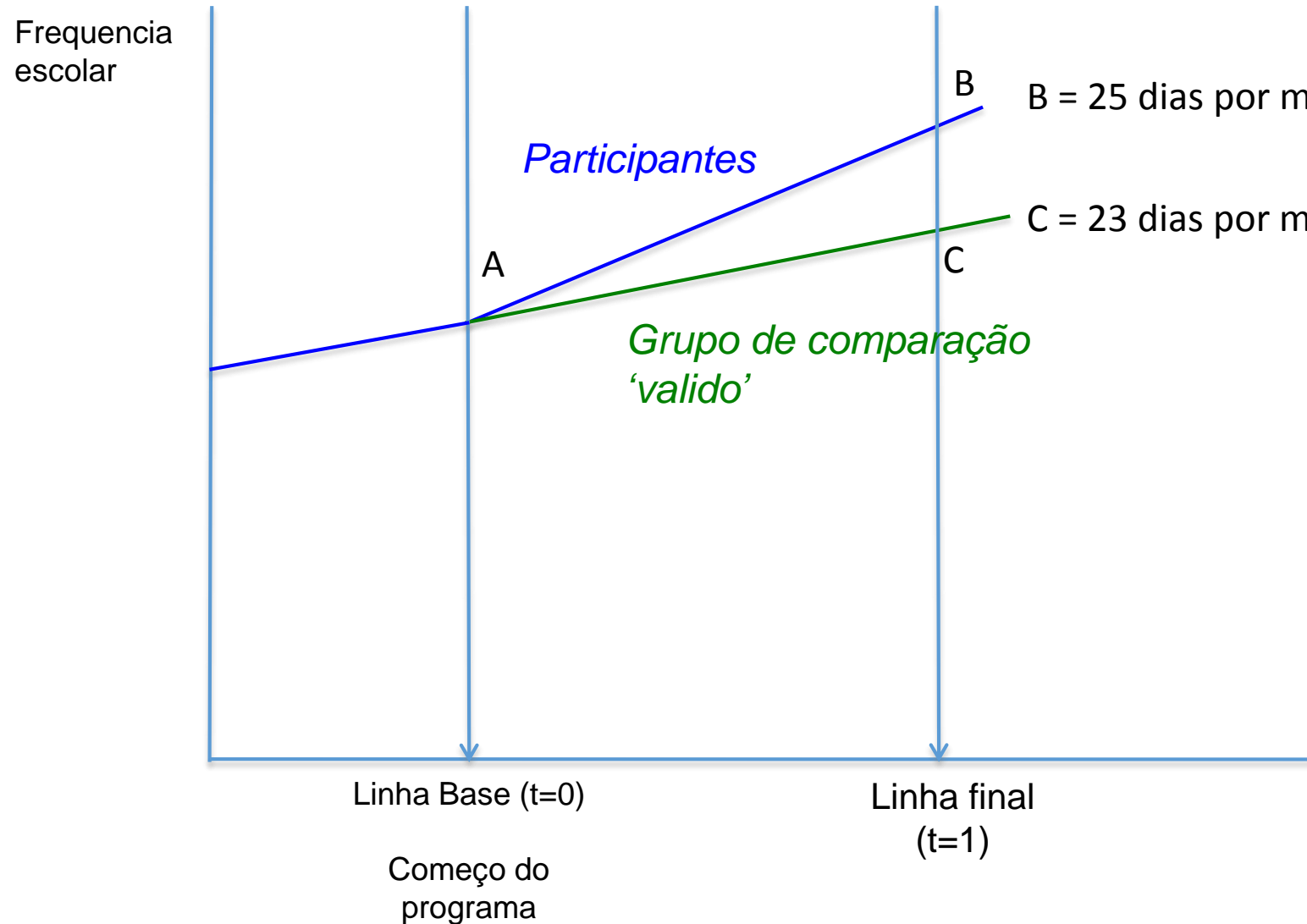
Impacto estimado: 2 dias
Erro padrão: 5 dias



Há impacto?

O resultado é estatisticamente significativo?

Exemplo – Significância Estatística



Impacto estimado: 2 dias
Erro padrão: 0.5 dias



Há impacto?

O resultado é estatisticamente significativo?

4. Significância estatística

TABELA 5
Impacto do PBF sobre as horas trabalhadas no setor informal

	Amostra total Todos os domicílios que recebem PBF		Filtro I Domicílios com RDPC < R\$ 700		Filtro II Domicílios com RDPC > R\$ 50 e RDPC < R\$ 700	
	Coefficiente	Janela ótima	Coefficiente	Janela ótima	Coefficiente	Janela ótima
Bolsa Família (erro-padrão)	-2,006 (1,224)	1,67	-1,778 (1,028)	2,08	-2,152 (1,215)	2,00
Bolsa Família (<i>bandwidth</i> = 110) (erro-padrão)	-2,392 (1,497)	1,84	-1,866 (1,178)	2,29	-2,463 (1,525)	2,20
Bolsa Família (<i>bandwidth</i> = 125) (erro-padrão)	-1,965 (1,134)	2,09	-2,633 (1,568)	2,60	-2,072 (1,327)	2,50
Bolsa Família (<i>bandwidth</i> = 150) (erro-padrão)	-1,431 (0,975)	2,50	-1,752 (1,425)	3,12	-1,638 (1,313)	3,00

Fonte: PNAD de 2006.

- Diminuição de 2 horas nas horas trabalhadas no setor informal
 - Estatisticamente significativo?
 $\text{Erro padrão} * 1.96 = 1,224 * 1.96 = 2.39$
 - Intervalo de confiança:
 $2.006 \pm 2.39 = [-0.384; 4.396]$
- O valor de '0' está dentro desse intervalo
- Estatisticamente não significativo

Avaliação experimental e não experimental

Os 3 métodos mais comuns para encontrar um grupo de comparação válido:

Experimental: Estudos Controlados Randomizados / Métodos de seleção aleatória

Não experimental:

Diferença em Diferenças; Técnicas de Pareamento, Método de Regressão Descontínua e combinação entre estes métodos.

Particularmente diferença-em-diferenças e pareamento

Projeto Transferência e desenhos de avaliação

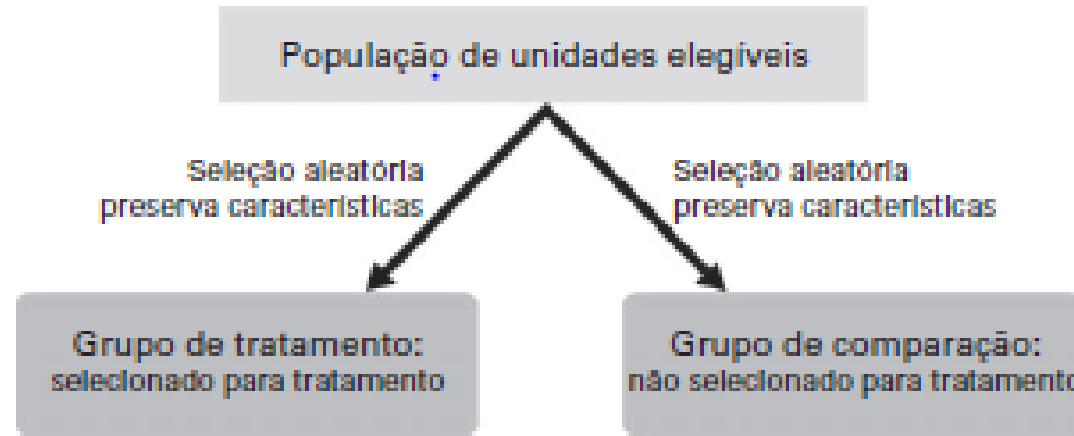
País	Tipo de Desenho	Nível de aleatorização (ou do pareamento)	Tamanho da amostra do inquérito de seguimento	Amostra de inelegíveis?
Etiópia	Não experimental: Escore de Propensão (PSM e IPW)	Agregado Familiar dentro da localidade	3.351	Sim
Gana	Não experimental: Escore de Propensão (IPW)	Agregado Familiar e Região	1.504	Não
Quênia	Experimental + PSM e IPC	Localidade	1.811	Não
Lesoto	Experimental	Distrito Eleitoral	2.150	Sim
Malauí	Experimental	Cluster de localidades	3.369	Apenas na linha de base
Malauí	Experimental	Cluster de localidades	766	Não
África do Sul	Não experimental: Escore de Propensão (GSM)	Agregado Familiar	2.964	Não
Zâmbia (modelo CG)	Experimental	Comitês de Assistência Social Comunitária	2.519	Não
Zâmbia (modelo MCTG)	Experimental	Comitês de Assistência Social Comunitária	3.077	Não
Zimbábue	Não experimental: pareamento	Distritos e localidades (wards)	2.630	Apenas na linha de base

Métodos de Aleatorização

→ Métodos de Aleatorização em ação!

Métodos de Aleatorização

- Ideia básica: selecionar aleatoriamente quem será membro do grupo de controle e quem será membro do grupo de comparação



Fonte: Autores.



A aleatorização resolve o problema do viés de seleção:
Em média, os grupos de tratamento e de comparação deveriam ser idênticos em todas as características (observáveis e não observáveis)

Questões éticas em uma avaliação aleatorizada

Selecionar aleatoriamente as unidades que receberão o tratamento levanta questões éticas. O quão ético é negar o acesso a um potencial (benéfico) programa de forma aleatória?

Aleatorizar “eticamente”:

- a) Para programas que possuem muita procura (cotas): a aleatorização pode aumentar a transparência nas regras de alocação no programa:
- B) Para programas que, em seu desenho, preveem um escalonamento da sua ampliação – Programas do Projeto Transferência
- C) Programas nacionais – pode-se aleatorizar a comunicação feita sobre o programa, ao invés do acesso ao programa. Hipótese: a comunicação é efetiva e gera diferenças nas taxas de inscrição entre tratados e controles.

A aleatorização em níveis maiores de agregação (cidades, distritos, escolas) tende a ser mais aceitáveis do que a aleatorização de indivíduos. Aqui há outro objetivo da avaliação de impacto: medir os “efeitos de transbordamento” (spillover effects)

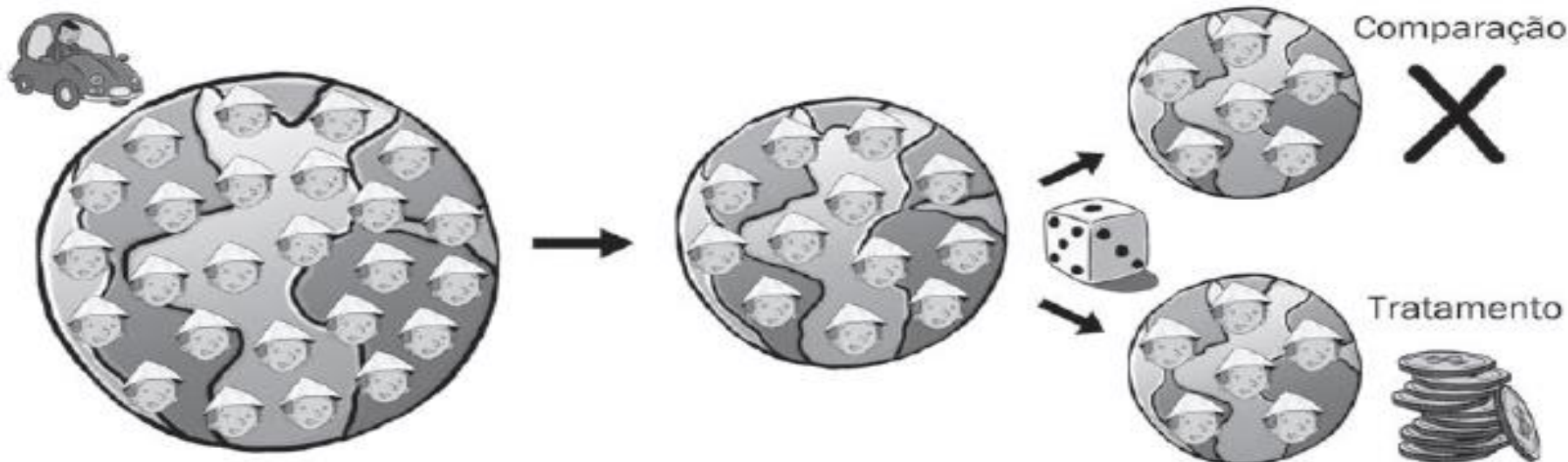
Dois conceitos importantes

- Validade Interna: validade do grupo de comparação – ausência de viés na estimativa que encontro para a minha amostra.
- Validade Externa: representatividade da população de interesse. Em que medida meus resultados de impacto são generalizáveis para toda a população de interesse, caso eu queria fazer a ampliação do programa.

Passo 1:
Unidades Elegíveis

Passo 2:
Amostra da Avaliação

Passo 3:
Aleatorização ao Tratamento



Validade Externa

Validade Interna



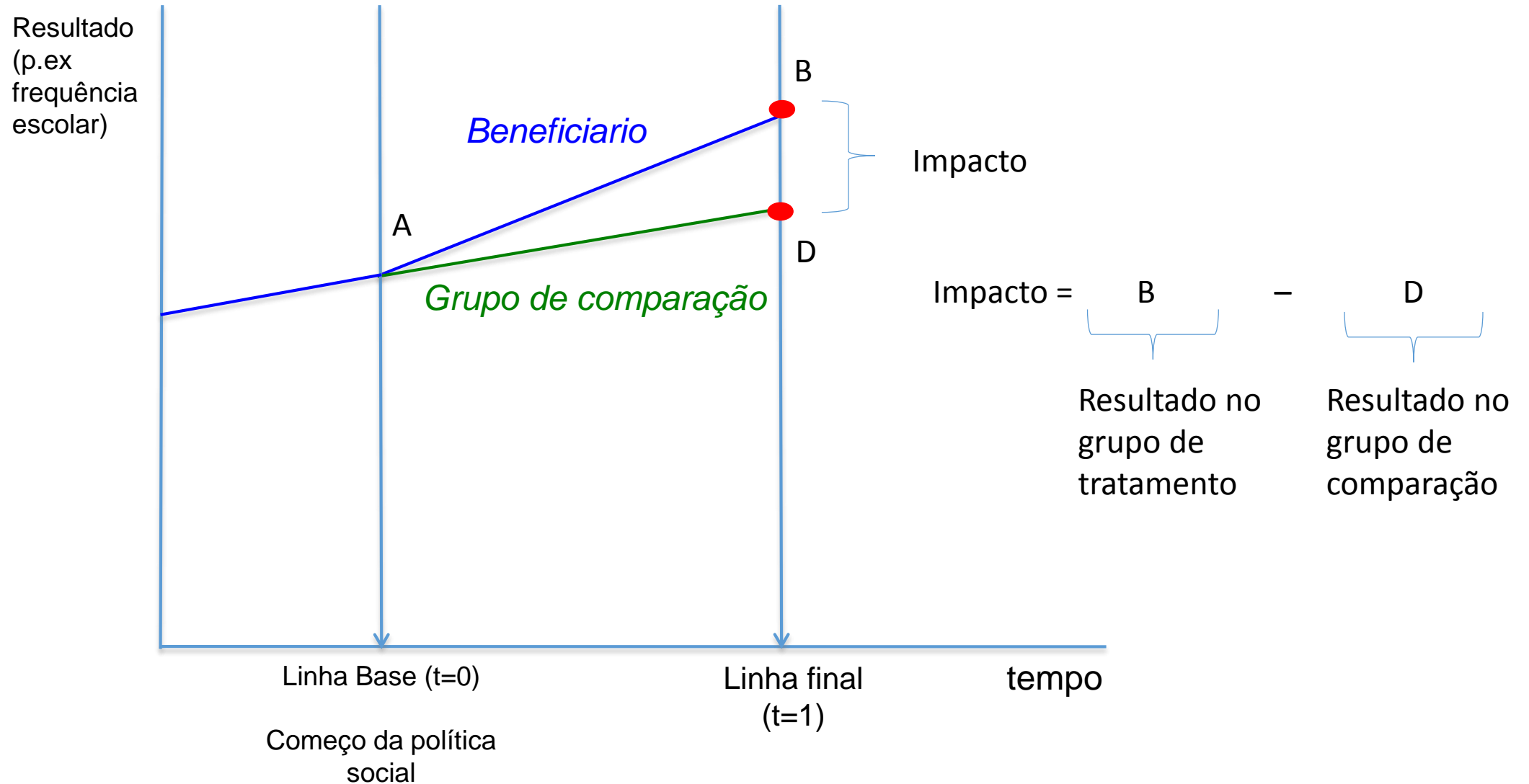
Inelegível



Elegível

Fonte: Autores.

Métodos de Aleatorização



Efeito da Intenção-de-tratar (ITT) e Efeito do Tratamento sobre os tratados (ATT)

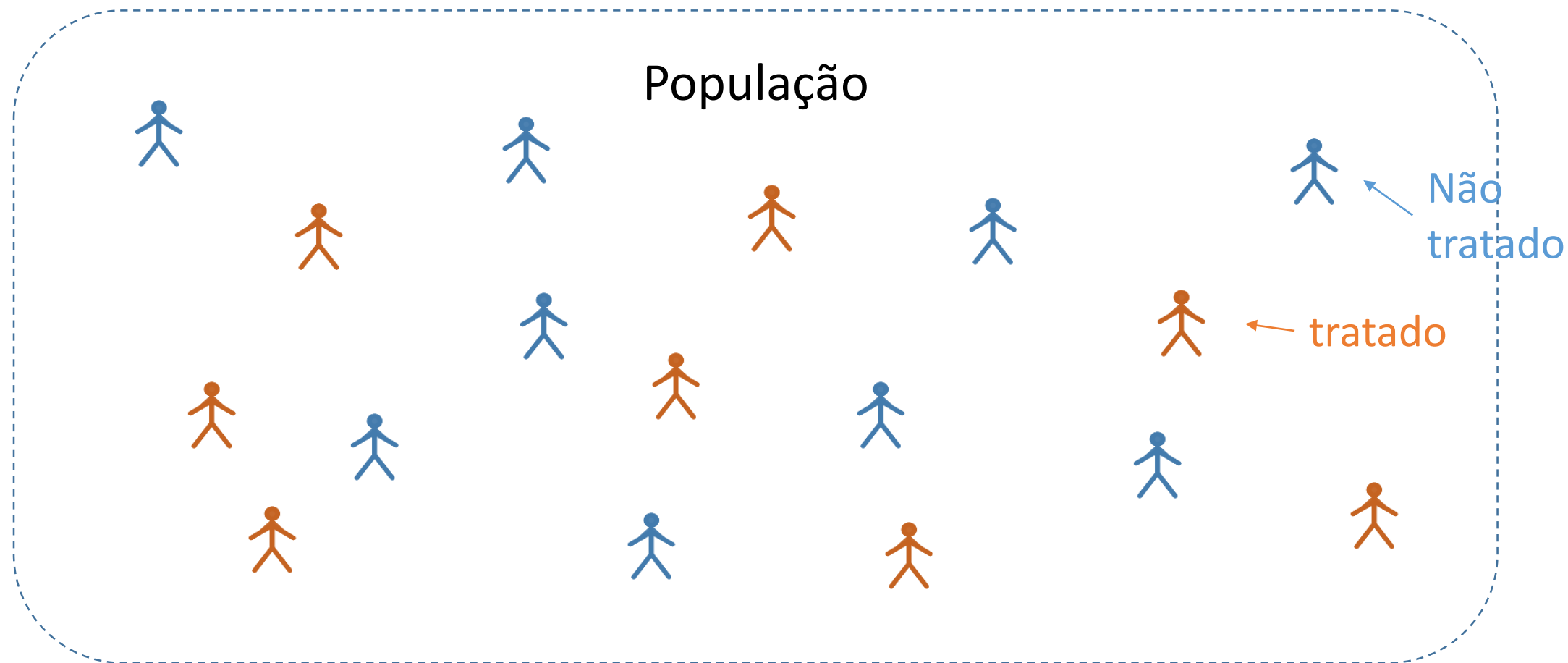
É importante saber que existem dois tipos diferentes de estimativas de impacto.

O impacto pode ser estimado sobre todos aqueles para quem o programa foi oferecido independentemente de terem recebido o programa/subsídio ou não – Efeito da Intenção de tratar... ou apenas para aqueles que realmente receberam – Efeito do Tratamento sobre os tratados.

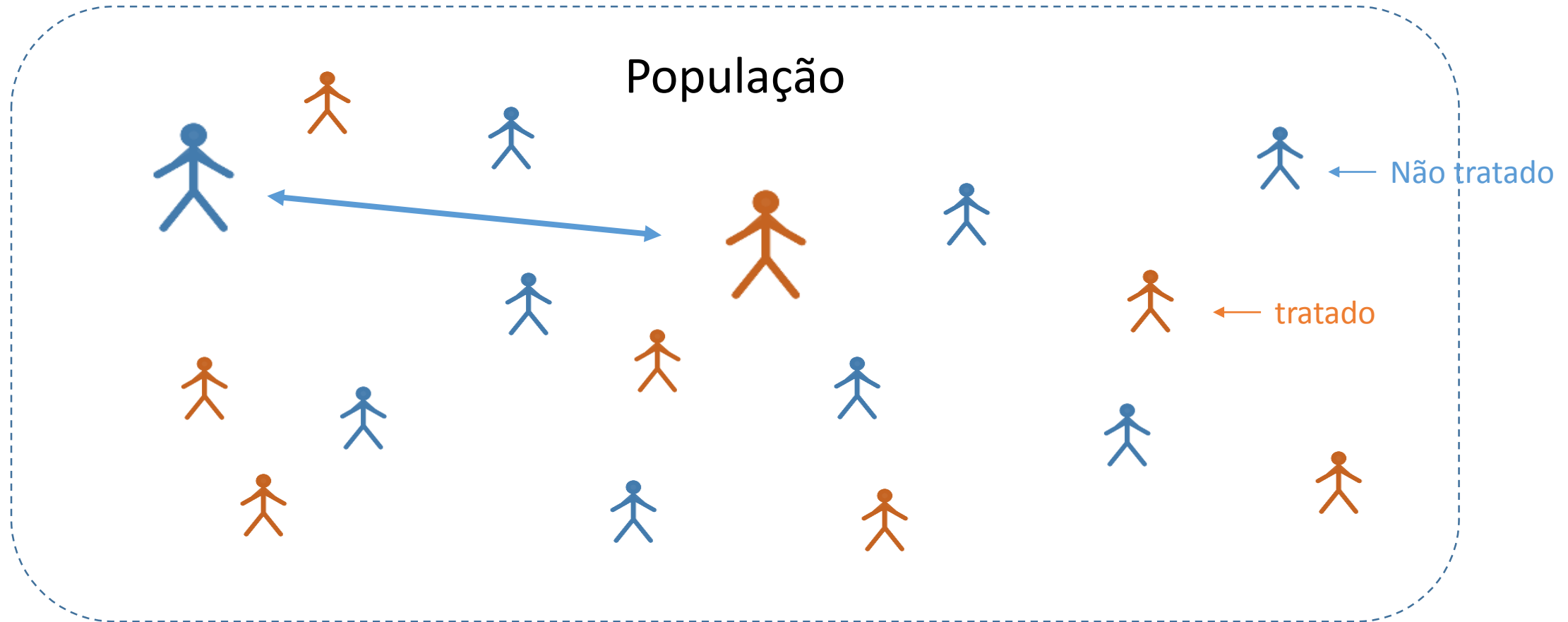
Estimadores baseados em pareamento

- Pareamento– método popular para estimar impactos causais
- Útil quando a aleatorização é impossível
- Ideia principal é encontrar dois (ou pares de) indivíduos idênticos na base de dados, com a exceção de que um é tratado e o outro não é tratado.

Pareamento



Pareamento: Dois indivíduos idênticos



- Qualquer diferença nas variáveis de resultados entre indivíduos tratados e não tratados que foram pareados é atribuível ao tratamento.

Hipótese importante

- Seleção em observáveis

= seleção para os tratamento é completamente determinada por variáveis observáveis.

=dadas estas variáveis observáveis, a alocação para o tratamento pode ser considerada aleatória.

→ Entre duas pessoas com as mesmas características observáveis, mas em que uma foi tratada e a outra não. A não tratada fornece o contrafactual para a que foi tratada.

Grupo de controle artificial

- Métodos de pareamento fornecem um método cuidado através do qual as observações que compõem o grupo de comparação são selecionadas.
- Um grupo de controle artificial é gerado.
- Após isso o método de DD pode ser implementado

- **Duas hipóteses importantes precisam ser feitas.**

Hipótese 1: Seleção em observáveis

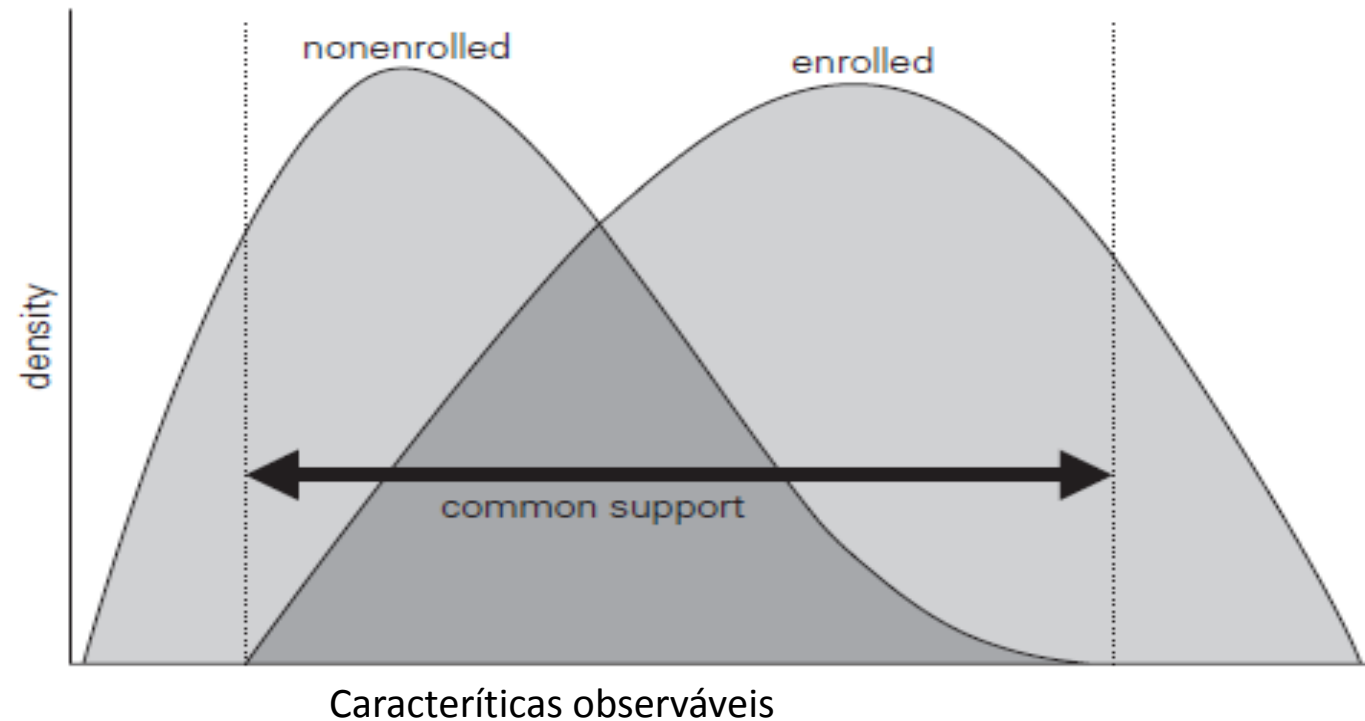
- Todas as variáveis que são relevantes para determinar o tratamento são observáveis.
 - Assim, o resultado potencial do tratamento torna-se observável (ambos, o grupo de tratamento e comparação tem o mesmo resultado na ausência do tratamento)
 - Similar à hipótese da tendência comum no caso da DD
 - Ambos os grupos evoluíam de maneira semelhante na ausência do tratamento.
- Dados os observáveis, a alocação do tratamento não é correlacionada com a variável de interesse.

Hipótese 2: Suporte comum

- Grande parte dos indivíduos tratados têm uma contraparte entre os indivíduos do grupo de controle, i.e., há interseção entre os dois grupos.
- Esta hipótese é testável.

Pareamento (seleção das variáveis observáveis)

- Quanto maior o suporte comum, melhor



Pareamento - estimando o impacto

