

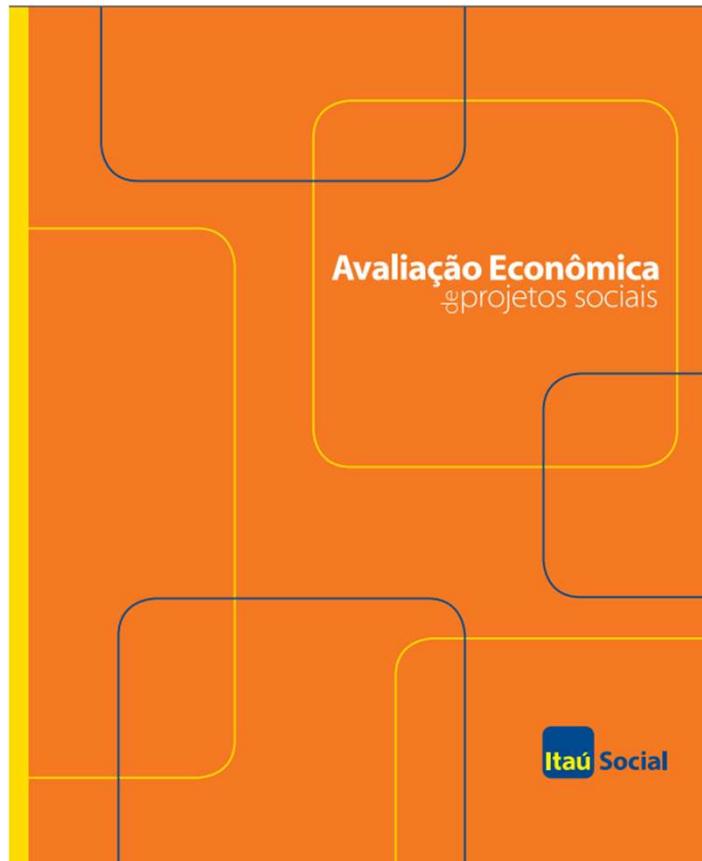
**CURSO BÁSICO DE ESTATÍSTICA E
ECONOMETRIA APLICADAS À
AVALIAÇÃO DE POLÍTICAS PÚBLICAS**

Prof. Francis Petterini (PPGEco/UFSC)

Curso para o TCE/SC

2023

Bibliografia básica



ÍNDICE

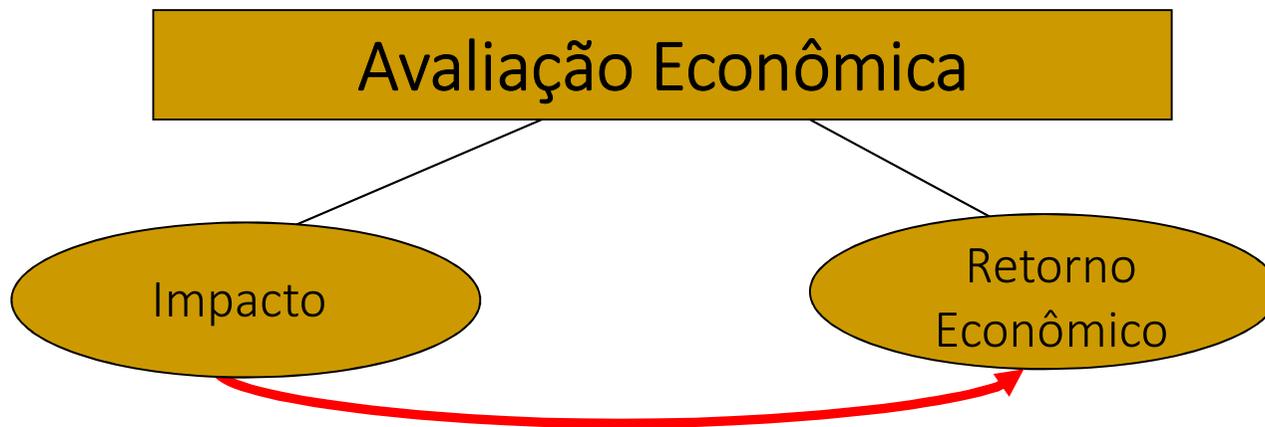
Apresentação	8
Prefácio	9
1. Avaliação de Impacto de Programas Sociais. Por que, para que e quando fazer?	13
2. Modelo de Resultados Potenciais	39
3. Método de Aleatorização	55
4. Diferenças em Diferenças	85
5. Pareamento	111
6. Variáveis Instrumentais	145
7. Regressão Descontínua	167
8. O Cálculo do Retorno Econômico	193
Referências Bibliográficas	226
Apêndice	237

SUMÁRIO DO CURSO

1. Conceitos básicos em avaliação de impacto
2. Revisão de estatística e uso avançado de planilhas eletrônicas
3. Modelo de Resultados Potenciais
4. Econometria: revisão do modelo de mínimos quadrados
5. Desenho de uma avaliação para uma política
6. Modelo de Diferenças em Diferenças
7. Técnicas de Pareamento
8. Estudo de caso e aplicações de diferentes técnicas econométricas
9. Discussão sobre os trabalhos de conclusão de curso
10. Revisão de matemática financeira
11. O Cálculo do Retorno Econômico
12. Conclusão do curso: apresentação dos trabalhos

INTRODUÇÃO

- A avaliação econômica é um componente importante de qualquer política pública: “Apesar de bilhões de dólares serem gastos todos os anos com projetos sociais, muito pouco se sabe do seu impacto sobre beneficiários e sua viabilidade econômica” (Banco Mundial)
- A avaliação econômica pode ser dividida em duas partes: avaliação de impacto e avaliação de retorno econômico



AVALIAÇÃO ECONÔMICA

- A avaliação de resultado e impacto usa ferramentas estatísticas para estimar o **efeito causal** do projeto sobre o beneficiários, ou seja, se ações do projeto atingiram ou não seus objetivos iniciais e qual a magnitude desse efeito – para essa análise é preciso indicadores objetivos e mensuráveis
- O cálculo do retorno econômico estima se os benefícios gerados superam os custos do projeto durante a elaboração, execução e monitoramento

Existem outros tipos de avaliação (diferentes e/ou complementares a “avaliação econômica”). Por exemplo, “avaliações qualitativas por entrevistas”, “pesquisas de disponibilidade a pagar” etc.

PROJETOS SOCIAIS E POPULAÇÃO-ALVO

- Um projeto (programa ou política) social sempre parte do pressuposto de que se trará benefícios para uma **população-alvo** (ou público-alvo)
- Assim, para realizarmos uma avaliação econômica, primeiro é preciso definir claramente o grupo ao qual o projeto se destina, isto é, o grupo de pessoas que o projeto pretende beneficiar (o **grupo de tratamento**)
- Em geral, para a identificação clara da população-alvo, analisam-se os chamados **critérios de elegibilidade** ao projeto
 - ✓ *Ex.: o Programa Bolsa Família atende todas as famílias que tenham renda familiar per capita abaixo do nível de pobreza*

FOCALIZAÇÃO

- Tendo em mente o público-alvo, é possível avaliar a focalização do programa
- O programa está incluindo de fato as pessoas que deveria incluir? Ou será que dentre os beneficiários estão indivíduos que a principio não deveriam estar participando?
- Quando um programa beneficia pessoas que não deveriam ser contempladas e/ou quando deixa de fora pessoas que pertencem ao público-alvo, dizemos que o programa incorreu nos chamados erros de focalização
 - ✓ *excluir do programa pessoas pertencentes ao público-alvo (ineficiência no alcance ou na cobertura do programa)*
 - ✓ *incluir no programa pessoas não pertencentes ao público-alvo (vazamento do programa)*

AVALIAÇÃO DE IMPACTO: RELAÇÕES CAUSAIS

- A avaliação de impacto objetiva quantificar **relações causais** entre as ações do projeto e os indicadores de interesse
- Relações causais são as relações de causa e efeito (importante: **correlação não implica necessariamente em causalidade**)
- Ter tido impacto causal significa dizer que na ausência do projeto as alterações observadas nos indicadores de interesse não teriam acontecido
- A avaliação de impacto nos permitirá concluir se foi de fato o projeto o responsável pelas alterações observadas nos indicadores de interesse
- A avaliação também nos permite estudar se programa está causando algum impacto não pensado inicialmente

EXEMPLO DE UMA PROVÁVEL RELAÇÃO CAUSAL

Suponha um programa que tem por objetivo melhorar a qualidade do ensino

- ✓ Ação: capacitação de professores
- ✓ Indicador escolhido: desempenho escolar



Se a avaliação de impacto mostrar que os resultados obtidos são significativamente positivos, justifica-se a implementação do programa em outras escolas parecidas

CONTRAFACTUAL: O PROBLEMA FUNDAMENTAL DA AVALIAÇÃO

- O problema central da área de avaliação de programas é construir o contrafactual do grupo tratado pelo programa
- Na forma mais simples de apresentar o problema, podemos pensar que **qualquer indivíduo está sempre em uma de duas situações mutuamente excludentes: ter sido ou não ter sido tratado**
- O melhor grupo de comparação para os indivíduos tratados seria formado pelos mesmos indivíduos na situação em que eles não fossem tratados. Mas dado que essas situações são mutuamente excludentes, é impossível observar este grupo
- O desafio do avaliador é encontrar um grupo de indivíduos que represente adequadamente a situação de não tratamento, ou seja, um grupo que funcione como um bom contrafactual do grupo tratado (grupo de tratamento), chamado de **grupo de controle**

RETORNO ECONÔMICO

- Tendo feito a avaliação de impacto, outras questões surgirão, tais como:
 - ✓ Foram os custos justificados para os resultados observados?
 - ✓ Qual a dimensão destes resultados em relação a outras alternativas?
- Estas perguntas são importantes porque muitas vezes, embora o projeto tenha impacto positivo, ele não é grande o suficiente diante dos custos incorridos em sua implementação
- Pode ser que outros programas alcançassem resultados semelhantes mas com um custo menor
- Então o cálculo do retorno econômico deve comparar os benefícios observados com os custos incorridos na implementação do programa

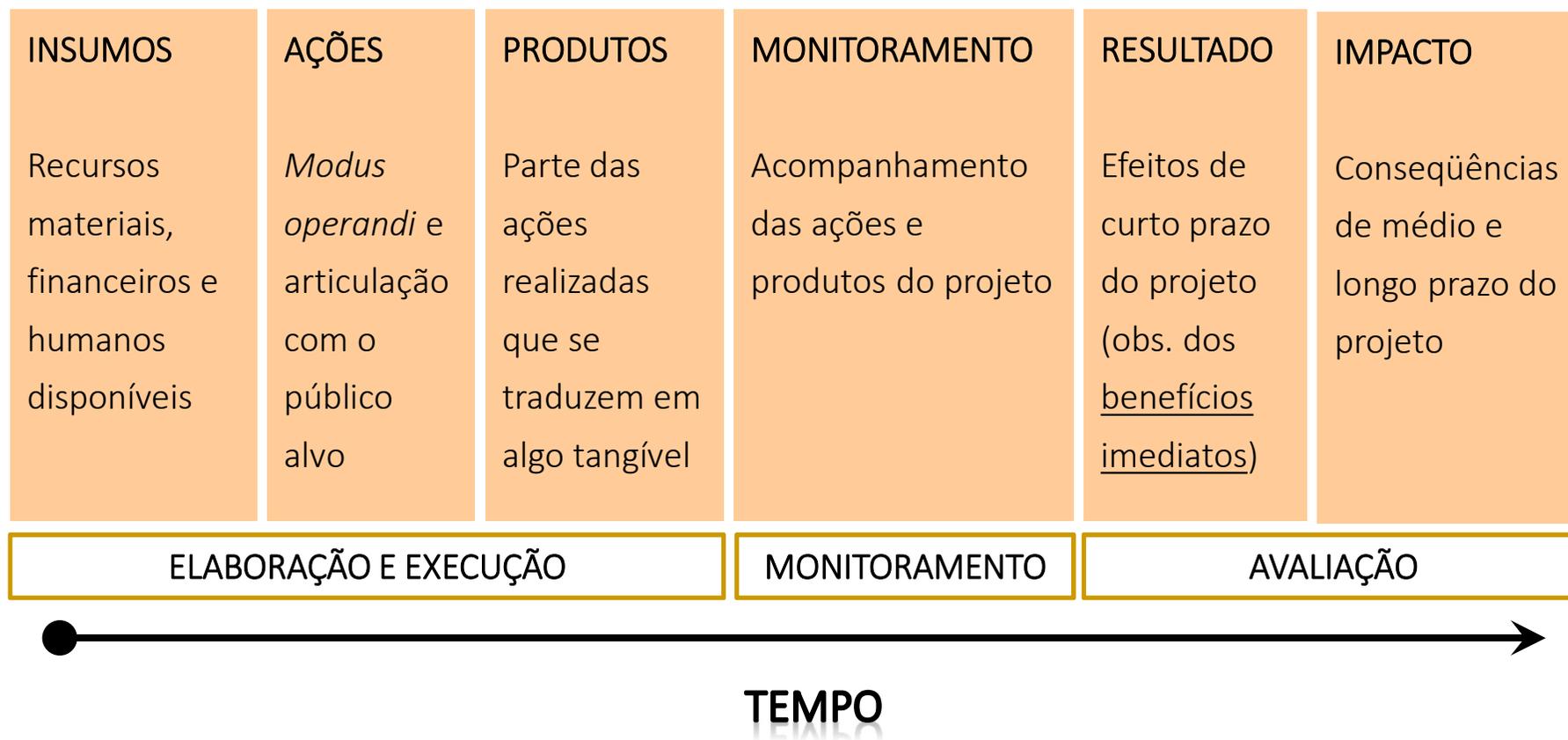
MARCO LÓGICO E AVALIAÇÕES EX-ANTE E EX-POST

- A decisão pela implementação de um programa é geralmente precedida pela apresentação de argumentos que justificam sua existência. Esta argumentação é chamada de “Marco Lógico”
- Quando a argumentação é feita com algum tipo de simulação quantitativa, diz-se que foi feita uma avaliação *ex-ante*
- A metodologia que será discutida no decorrer deste curso é a *ex-post*
- As avaliações *ex-ante* e *ex-post* se baseiam cada qual em um conjunto distinto de informações
 - ✓ Na avaliação *ex-ante*, o papel da teoria é primordial, uma vez que as estimativas de impacto são obtidas a partir de simulações sobre o comportamento hipotético dos futuros beneficiários do programa
 - ✓ Na avaliação *ex-post*, quando uma grande quantidade de informações foi coletada, a teoria será testada pela evidência empírica

MARCO LÓGICO E AVALIAÇÃO

- O Marco Lógico é uma ferramenta que deve facilitar o processo de justificativa, elaboração, execução e avaliação de projetos em gerais (sociais em particular)
- O objetivo do Marco Lógico é dar estrutura ao processo de planejamento e informação essencial relativa ao projeto
- Pode ser utilizado em todas as etapas de preparação, execução e avaliação do projeto
- Na perspectiva da avaliação, especificamente, é no Marco Lógico onde se deve encontrar os elementos de análise das relações causais e a forma de construção do grupo de controle

DESCRIÇÃO DO MARCO LÓGICO



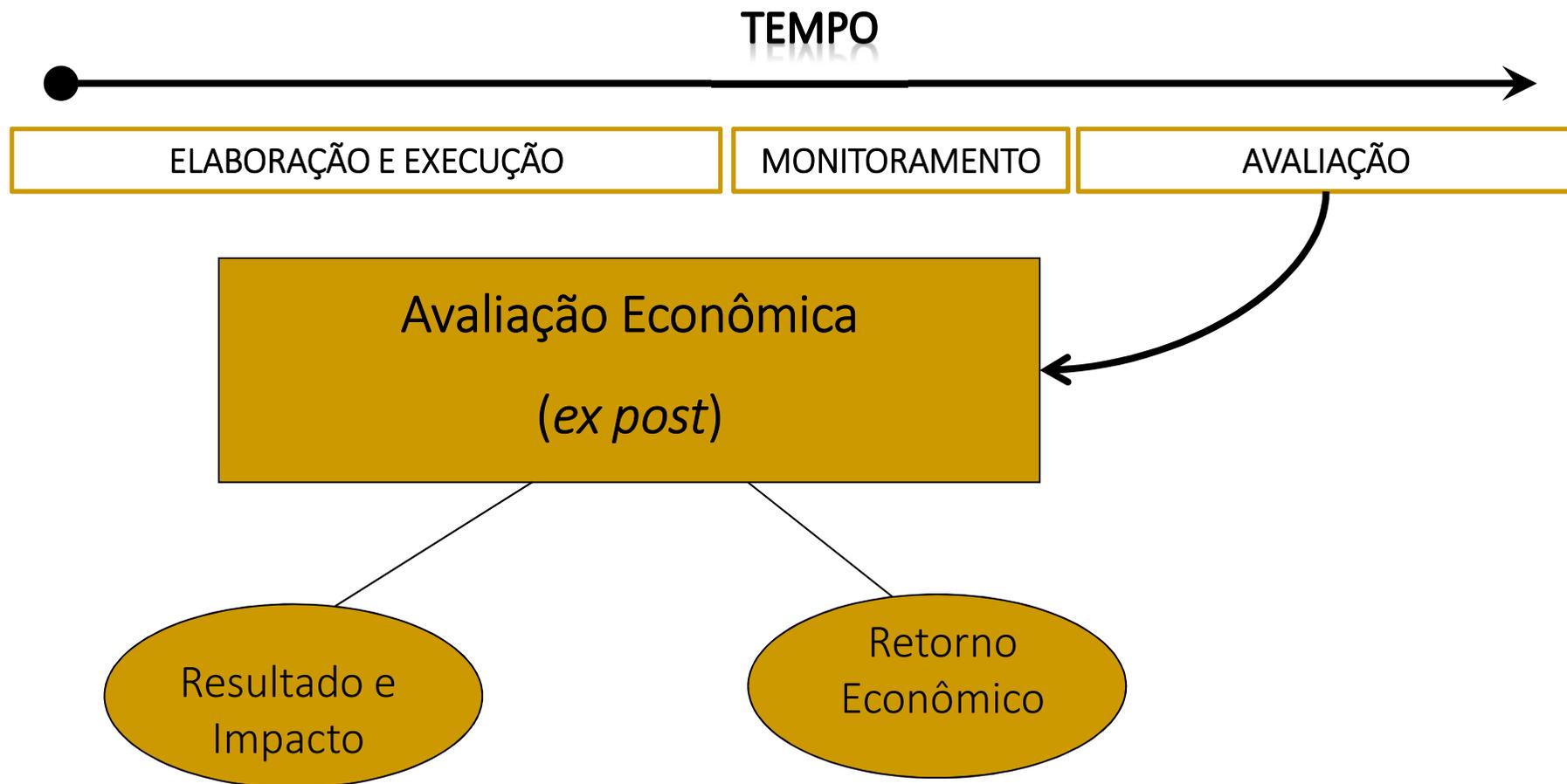
MARCO LÓGICO – EXEMPLO: PROGRAMA SAÚDE DA FAMÍLIA

INSUMOS	AÇÕES	PRODUTOS	MONITORAMENTO	RESULTADO	IMPACTO
Equipes de atenção básica, postos de saúde, alimentos, remédios etc.	Profissionais de saúde observando as famílias, distribuição de remédios	Visitas para prevenir o início de doenças evitáveis	Acompanhamento do número de visitas das equipes e pesagem das crianças, entrevistas com as mães etc.	Redução da taxa de mortalidade infantil e internação de idosos	Melhoria do capital humano e redução de custos para o SUS
ELABORAÇÃO E EXECUÇÃO			MONITORAMENTO	AVALIAÇÃO	



TEMPO

MARCO LÓGICO E AVALIAÇÃO



INDICADORES

- Como dito, para a avaliação econômica é preciso indicadores objetivos e mensuráveis
 - ✓ Indicadores de **oferta** estão associados aos recursos necessários para a realização do projeto
 - ✓ Indicadores de **monitoramento** estão associados aos recursos utilizados para a realização do projeto
 - ✓ Indicadores de **resultado** estão associados aos objetivos de curto prazo do projeto (coletados para os grupos de tratamento e controle)
 - ✓ Indicadores de **impacto** estão associados aos objetivos de médio e longo prazo do projeto (coletados para os grupos de tratamento e controle)

INDICADORES – EXEMPLO: PROGRAMA SAÚDE DA FAMÍLIA



POR QUE AVALIAR ECONOMICAMENTE UM PROGRAMA SOCIAL?

- Uma das motivações da avaliação é, precisamente, a validação de um “Marco Lógico” e das relações causais nele descritas. Por este motivo é importante a avaliação de resultado e impacto de programas em escala piloto
- Mesmo que tenhamos convicção da existência de impacto, tão ou mais importante do que saber se há algum impacto é conhecer a magnitude do impacto
- Validar e mensurar os impactos é indispensável para o aperfeiçoamento do programa e para a melhoria na sua adequação às necessidades de seus usuários

A RELAÇÃO ENTRE IMPACTO, AMBIENTE SOCIOECONÔMICO E CARACTERÍSTICAS DA POPULAÇÃO BENEFICIADA

- Projetos sociais são desenhados para uma população-alvo em um ambiente pré-estabelecido
- É de se esperar que uma mesma ação dirigida a outra população, outro local e outro momento do tempo não beneficie da mesma forma os que dela participam
- Uma das utilidades da avaliação econômica é, precisamente, determinar em que medida a eficiência, a eficácia e a efetividade de um programa depende das características da população beneficiada, do momento no tempo e da natureza do ambiente socioeconômico em que ocorre

EFICIÊNCIA, EFICÁCIA E EFETIVIDADE

- Eficiência é uma medida relativa, usada para comparar duas ações de um mesmo projeto ou dois projetos diferentes. Refere-se à relação entre os resultados obtidos (indicadores de resultado e/ou impacto) e os recursos empregados (indicadores de oferta)
- Eficácia é uma medida relacionada a uma meta. Diz-se que um projeto social foi eficaz se ele alcançou um objetivo mensurável por um indicador de impacto e/ou resultado
- Efetividade é uma medida de duração de um impacto. Isto é, se o projeto social for efetivo, se ele deixar de existir os resultados alcançados não serão desfeitos

EFICIÊNCIA, EFICÁCIA E EFETIVIDADE: EXEMPLO

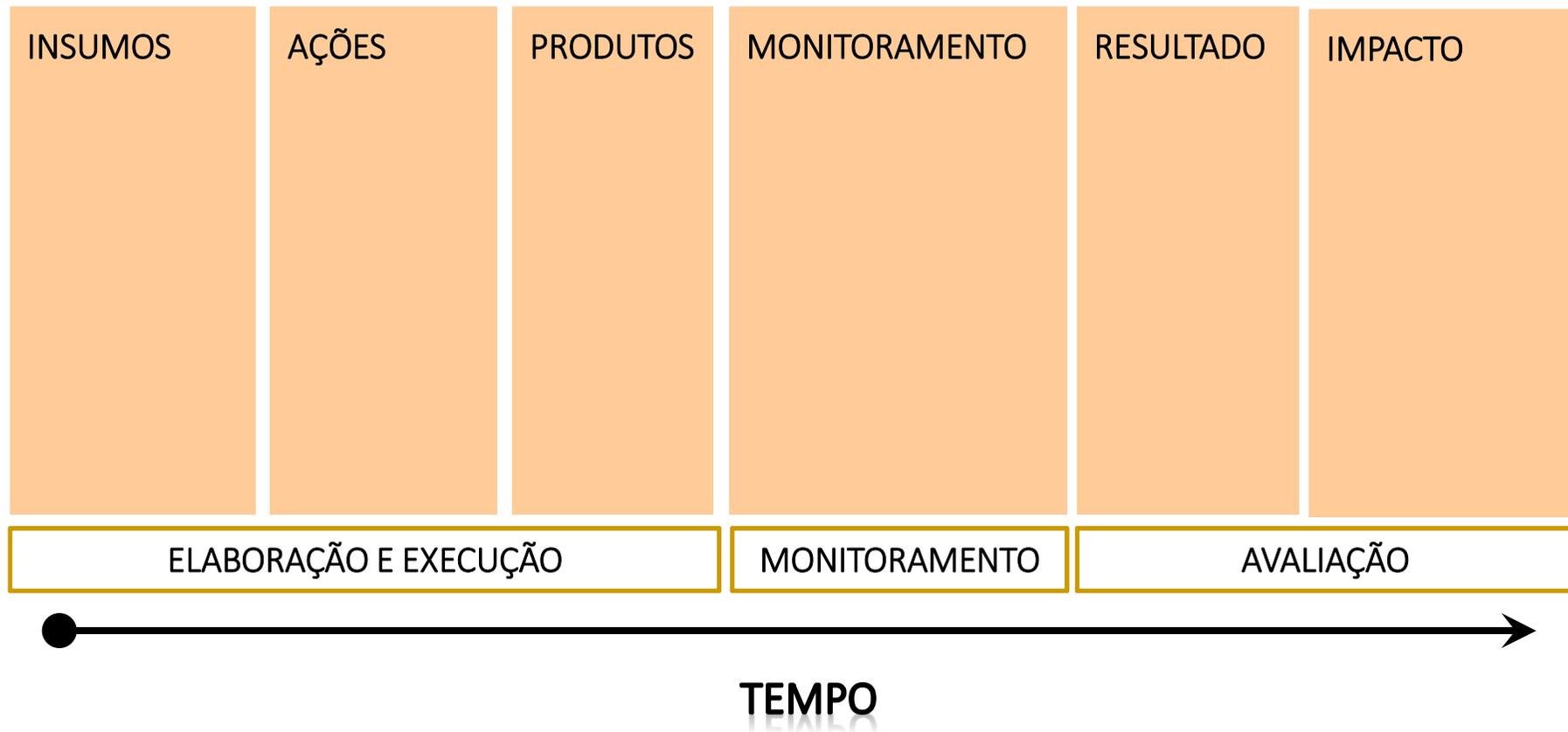
- Por exemplo, o Objetivo 1 da Declaração do Milênio é “erradicar a pobreza extrema e a fome”
- Suponha dois programas, A e B, executados em locais diferentes, mas com mesma população-alvo e mesma quantidade de recursos. Se A consegue retirar mais pessoas da condição de pobreza que B, em um mesmo período de tempo, diz-se que A está sendo mais eficiente que B
- Se ambos os programas erradicarem a pobreza até 20??, ambos foram eficazes
- Se no local onde A foi executado nunca mais ninguém passar fome, diz-se que A foi efetivo

Exercícios complementares

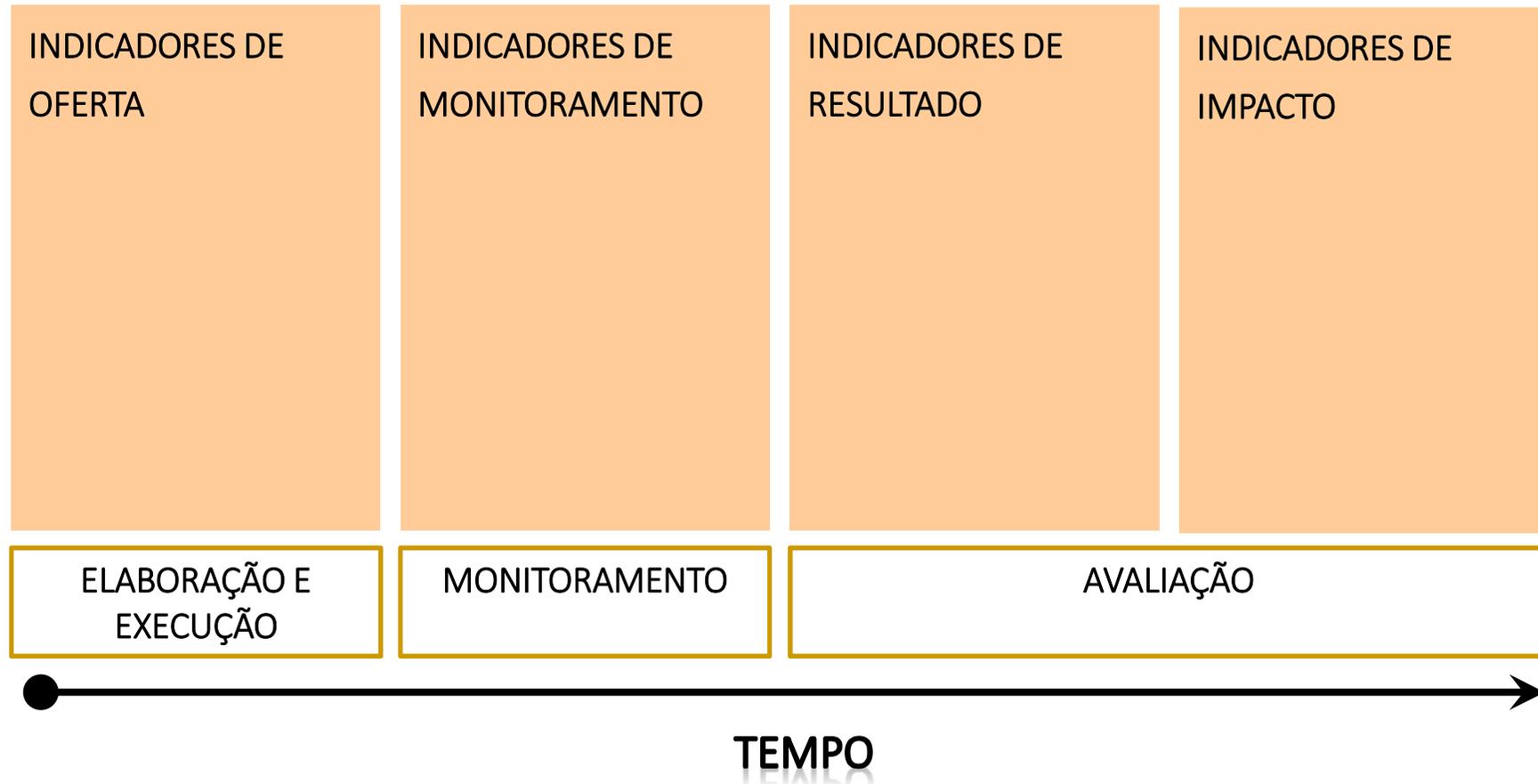
EXERCÍCIO 1: PARA PENSAR

- Quais são os objetivos dos seguintes projetos? Qual é a população-alvo? Como podem ser dadas as relações causais? Como se montariam os grupos de tratamento e controle?
 - ✓ Programa Saúde da Família
 - ✓ Projeto TAMAR
 - ✓ Programa Bolsa Família

2) Preencha o Marco Lógico do Projeto TAMAR e do Programa Bolsa Família



3) Preencha os indicadores do Projeto TAMAR e do Programa Bolsa Família



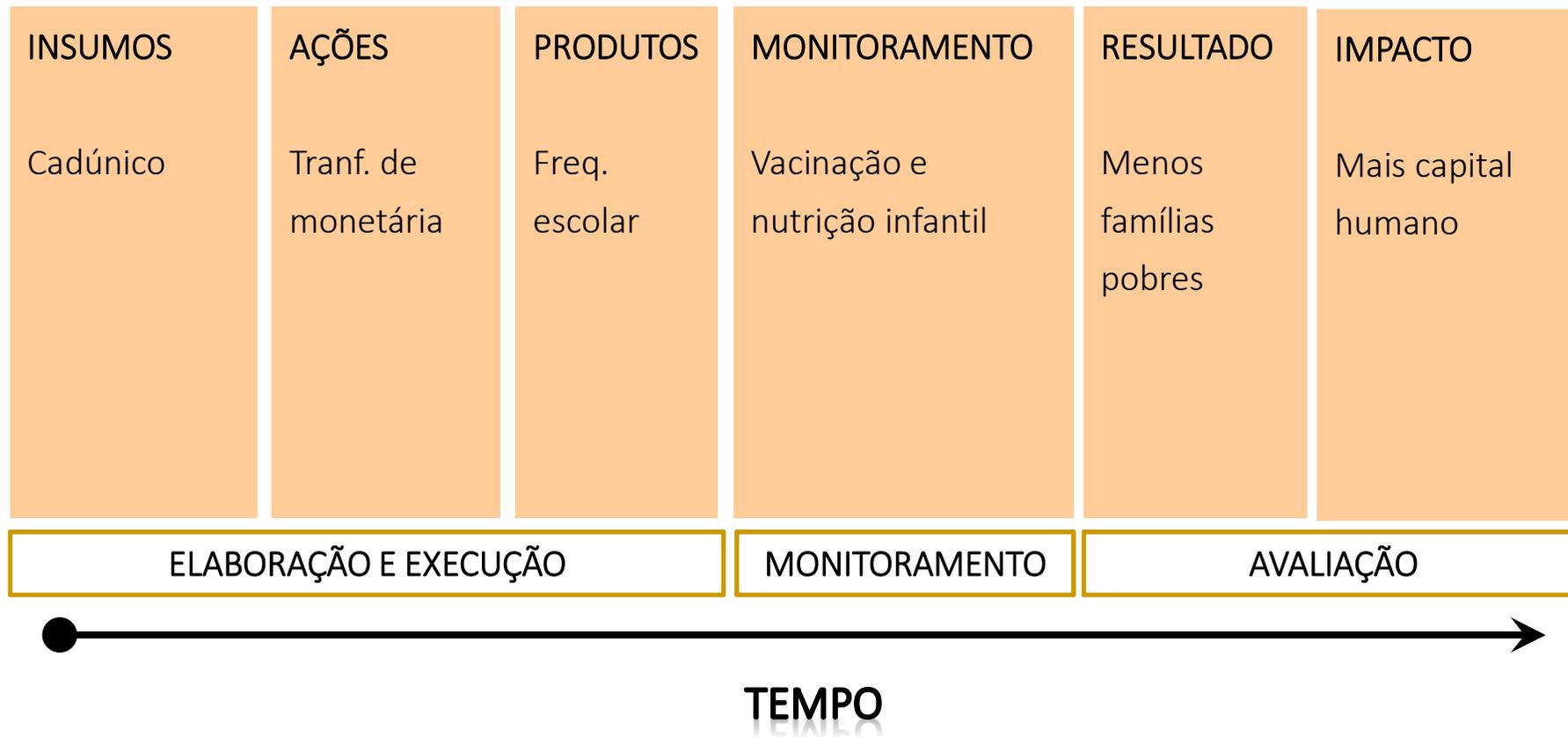
Exercícios complementares : indicações das respostas

MARCO LÓGICO – PROJETO TAMAR



TEMPO

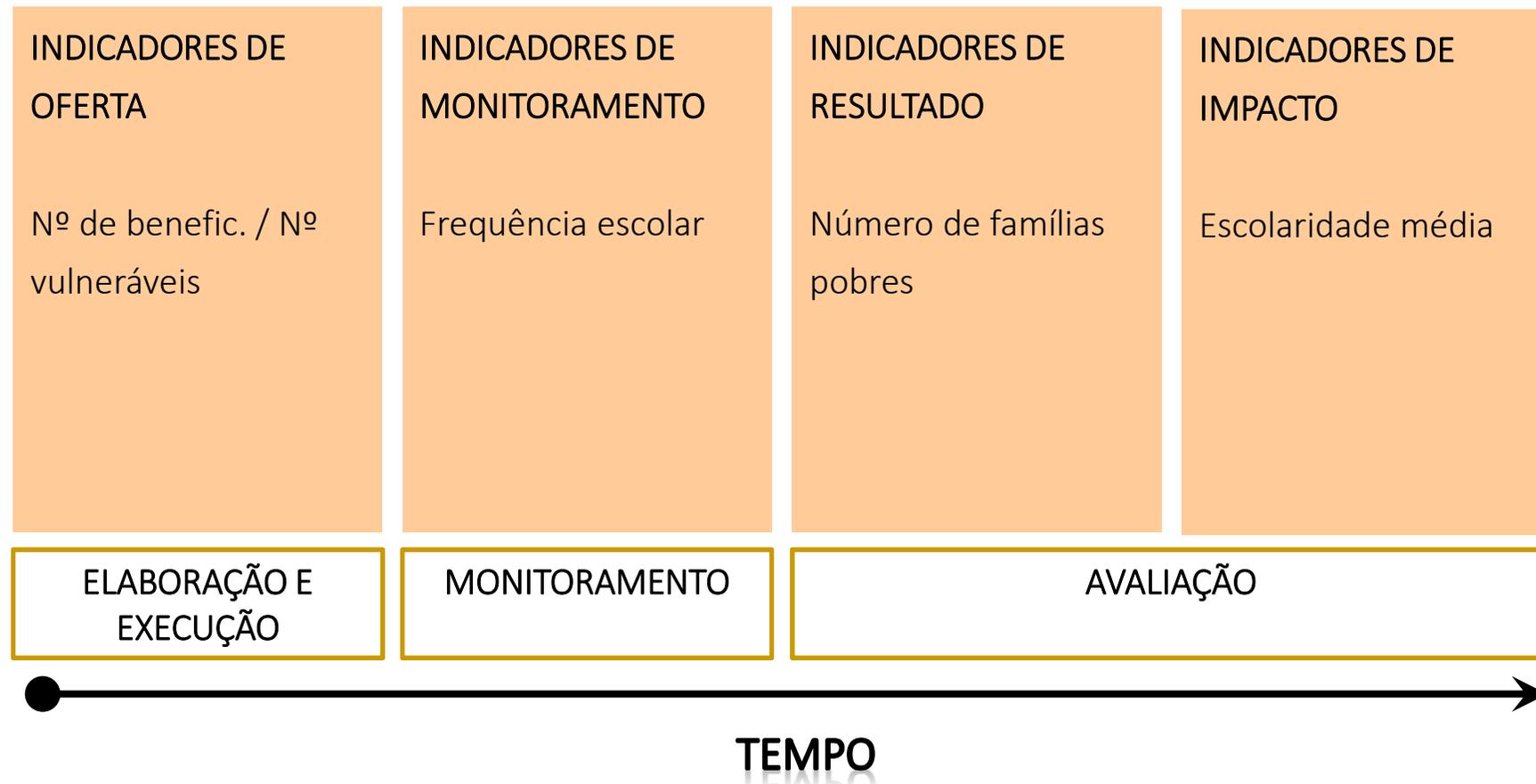
MARCO LÓGICO – BOLSA FAMÍLIA



INDICADORES – PROJETO TAMAR



INDICADORES – BOLSA FAMÍLIA



INDICADORES – EXEMPLO: PROGRAMA SAÚDE DA FAMÍLIA

- Relembrando, como dito na aula passada, para a avaliação econômica é preciso indicadores objetivos e mensuráveis

INDICADORES DE OFERTA

Número de assistentes sociais, médicos e dentistas;
Número de remédios para se distribuir

INDICADORES DE MONITORAMENTO

Razão entre o número de visitas das equipes por família e o número ideal de visitas

INDICADORES DE RESULTADO

Taxa de mortalidade infantil = número de óbitos de menores de um ano em relação a 1.000 nascidos vivos

INDICADORES DE IMPACTO

Atendimentos no SUS decorrentes a doenças relacionadas a má nutrição na infância

SUMÁRIO DO CURSO

- ~~1. Conceitos básicos em avaliação de impacto~~
2. Revisão de estatística e uso avançado de planilhas eletrônicas
3. Modelo de Resultados Potenciais
4. Econometria: revisão do modelo de mínimos quadrados
5. Desenho de uma avaliação para uma política
6. Modelo de Diferenças em Diferenças
7. Técnicas de Pareamento
8. Estudo de caso e aplicações de diferentes técnicas econométricas
9. Discussão sobre os trabalhos de conclusão de curso
10. Revisão de matemática financeira
11. O Cálculo do Retorno Econômico
12. Conclusão do curso: apresentação dos trabalhos

INDICADORES QUANTITATIVOS E QUALITATIVOS

- A avaliação de resultado e impacto ensinada neste curso usa ferramentas estatísticas para estimar o efeito causal do projeto sobre o beneficiários
- Nesse sentido, note que podemos classificar os indicadores estudados (de oferta, monitoramento, resultado ou impacto) em duas categorias:
 - ✓ Quantitativos (renda familiar, peso das crianças, frequência escolar, taxa de mortalidade infantil, número de tartarugas observadas etc.)
 - ✓ Qualitativos ou categóricos (gênero masculino/feminino, região norte/sul/leste/oeste, nível de satisfação alta/média/baixa etc.)

INDICADORES MENSURÁVEIS

- Os indicadores quantitativos são diretamente mensuráveis (renda, altura, peso etc.), enquanto que os qualitativos podem ser convertidos em escalas mensuráveis (números)
- Por exemplo:
 - ✓ “Sexo = 1” se a pessoa for do gênero masculino, e “Sexo = 0” caso contrário
 - ✓ “Região = 1” se for norte, “Região = 2” se for sul, “Região = 3” se for leste e “Região = 4” se for oeste
 - ✓ “Satisfação = +1” se o nível de satisfação for alto, “Satisfação = 0” se o nível de satisfação for médio e “Satisfação = -1” se o nível de satisfação for baixo

O “BRILHO NOS OLHOS”

- Alguns projetos sociais objetivam causar impactos psicológicos na população-alvo. Por exemplo, melhorar a autoestima ou reduzir o nível de tensão em ambientes extremos (como é o caso do projeto “Doutores da Alegria”)
- Nestes casos, os avaliadores precisam medir o que popularmente ficou conhecido por “brilho nos olhos” das pessoas
- Para tanto, é possível usar escalas psicométricas para medir algo que parece tão intangível como o “brilho nos olhos”. Dentre estas escalas, é possível citar:
 - ✓ A escala Likert que explora o nível de concordância com alguma afirmação. Por exemplo, em relação a afirmação “você está feliz” é possível dizer que: (1) concordo; ou, (0) discordo
 - ✓ A escala Rosenberg, que é mais complexa que a Likert mas é muito popular em estudos psicométricos

INDICADORES E UNIDADE DE OBSERVAÇÃO

- Tendo em mente que os indicadores são (ou podem ser transformados em) números, note agora que eles sempre estão associados a uma unidade de observação:
 - ✓ Renda de um trabalhador que passou pelo treinamento
 - ✓ Peso de uma criança acompanhada pelo PSF
 - ✓ Número de tartarugas nascidas na região Nordeste (projeto TAMAR)
 - ✓ Nível de satisfação de uma pessoa com a Prefeitura
- De agora em diante passaremos a denotar um indicador (renda, peso, nível de satisfação etc.) associado a uma unidade de observação (trabalhador, região etc.) por y_i , onde y representa o indicador e i a unidade de observação
- Na avaliação, sendo y_i um indicador de impacto ou resultado, estaremos interessados em estudar estatisticamente seu comportamento nos grupos de tratamento e controle
- Tentaremos identificar se o programa causou mudanças de comportamento em relação a y_i

VARIÁVEL ALEATÓRIA

- Um indicador qualquer associado a uma unidade de observação qualquer, y_i , é uma variável aleatória
- Ilustrativamente, imagine uma criança de um ano acompanhada pelo Programa Saúde da Família (PSF). Se y_i representa o peso dela, seu valor pode ser, em princípio, qualquer valor não negativo
- Mais ainda, a experiência pode nos dizer que y_i deve estar próximo de 10 ou 12 Kg, mas não temos como saber seu valor exato sem colocar a criança na balança
- Se o peso pode ser perto de 10 ou 12 Kg é variável, e se não sabemos ao certo é aleatório

POPULAÇÃO E AMOSTRA

- O grupo completo de observações de pessoas, objetos ou coisas é denominado por população, e um subconjunto de observações da população é denominado por amostra
- Seja a variável aleatória y_i . Ela pode ser observada em toda uma população de tamanho N , ou em uma amostra de tamanho n , com $n < N$
- Por exemplo, se estamos estudando a altura do brasileiro, temos que $N \approx 200$ milhões de pessoas. Mas por motivos de custo de pesquisa, disponibilidade de tempo etc. deveremos estudar uma amostra de n brasileiros
- **Neste curso trabalharemos com amostras !**

MEDIDAS DE TENDÊNCIA E DE DISPERSÃO

- Ao estudarmos uma variável aleatória y_i , geralmente estaremos interessados em medidas de tendência e de dispersão de seu comportamento

Tendência		Dispersão	
Média	$\bar{y} = \frac{\sum_{i=1}^n y_i}{n}$	Variância	$s^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n - 1}$
Moda	O valor de y_i mais comum	Desvio Padrão	$s = \sqrt{s^2}$
Mediana	O y_i cuja metade da amostra registra valor menor que ele	Amplitude	Distância entre o maior e o menor valor de y_i

Exercícios sobre o cálculo de
medidas de tendência e dispersão:
Usaremos, em Excel, o arquivo “Ex1”

DESCRIÇÃO DA BASE DE DADOS

O programa “Mulher do Futuro” – fictício – objetiva aumentar a renda de famílias chefiadas por mulheres. Para isso, oferece cursos profissionalizantes em diversas áreas e ajuda na inserção das participantes no mercado de trabalho, por meio de convênios com empresas. Os cursos duram em média seis meses. A avaliação de impacto do programa foi pensada desde sua concepção. Assim, realizou-se uma pesquisa com as participantes e não participantes antes da implementação. Dois anos após o início do programa, o grupo de avaliação voltou a campo e novamente coletou informações sobre as mulheres. O banco de dados contém as informações em dois momentos no tempo – escolaridade, idade, se o domicílio possui luz elétrica, se a mulher vive com cônjuge e a renda familiar per capita.

DDM.xml - Microsoft Excel

Arquivo Página Inicial Inserir Layout da Página Fórmulas Dados Revisão Exibição Acrobat

Área de Tran... Fonte Alinhamento Número

Formatação Condicional Formatar como Tabela Estilos de Célula Inserir Excluir Formatar Células

AutoSoma Preencher Limpar Classificar e Filtrar Localizar e Selecionar Edição

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	ID	escolaridad	idade	luz	renda_pc	tratado	conju	tempo	tra_tem					
2	567	2	14	0	85	0	1	0	0					
3	585	2	14	1	128	0	1	0	0					
4	571	1	14	1	255									
5	628	2	14	1	285									
6	626	1	14	1	316									
7	630	1	14	1	340									
8	564	2	14	1	340									
9	649	3	14	1	340									
10	652	1	14	1	383									
11	620	3	14	1	402									
12	622	3	14	1	418									
13	623	2	14	1	470									
14	644	1	14	1	510	0	1	0	0					
15	629	3	14	1	510	0	1	0	0					
16	635	3	14	1	510	0	1	0	0					
17	640													
18	643													
19	621	1	14	1	755	0	1	0	0					

Classificar

Adicionar Nível Excluir Nível Copiar Nível Opções... Meus dados contêm cabeçalhos

Coluna	Classificar em	Ordem
Classificar por	tempo	Valores
		Do Menor para o Maior

OK Cancelar

Selezione apenas as observações referentes ao período inicial

Sheet1

Pronto Média: 221.1048194 Contagem: 54009 Soma: 11939660.25 140%

DDM.xlsx - Microsoft Excel

Arquivo Página Inicial Inserir Layout da Página Fórmulas Dados Revisão Exibição Acrobat

Calibri 11

Quebrar Texto Automaticamente

Fonte Alinhamento Número Estilo Células

MODO.ÚNICO X ✓ f_x =SOMA(B2:B3002)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	ID	escolaridade	idade	luz	renda_pc	tratado	conjuje	tempo	tra_tem					
2993	815	3	58	1	510	1	0	0	0					
2994	818	2	58	1	550	1	1	0	0					
2995	802	2	58	1	570	1	1	0	0					
2996	779	2	58	1	590	1	1	0	0					
2997	807	2	58	1	633	1	1	0	0					
2998	808	2	58	1	765	1	0	0	0					
2999	775	7	58	1	925	1	1	0	0					
3000	783	2	58	1	1000	1	1	0	0					
3001	778	3	58	1	1005	1	1	0	0					
3002														
3003	Somatório	=SOMA(B2:B3002)												
3004	Média	SOMA(núm1; [núm2]; ...)												
3005														
3006														
3007														
3008														
3009														
3010														

Aponte

140%

Compute o numerador da fórmula da média:

$$\sum_{i=1}^n y_i = \sum_{i=1}^{3000} \text{escolaridade}_i$$

DDM.xlsx - Microsoft Excel

Arquivo Página Inicial Inserir Layout da Página Fórmulas Dados Revisão Exibição Acrobat

Área de Tran... Fonte Alinhamento Número

Formatação Condicional Estilos de Célula

Inserir Excluir Formatar Células

AutoSoma Preencher Limpar Classificar e Filtrar Localizar e Selecionar Edição

MODO.ÚNICO X ✓ fx =B3003/3000

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	ID	escolaridade	idade	luz	renda_pc	tratado	conjuge	tempo	tra_tem					
2993	815	3	58	1	510	1	0	0	0					
2994	818	2	58	1	550	1	1	0	0					
2995	802	2	58	1	570	1	1	0	0					
2996	779	2	58	1	590	1	1	0	0					
2997	807	2	58	1	633	1	1	0	0					
2998	808	2	58	1	765	1	0	0	0					
2999	775	7	58	1	925	1	1	0	0					
3000	783	2	58	1	1000	1	1	0	0					
3001	778	3	58	1	1005	1	1	0	0					
3002														
3003	Somatório	8794												
3004	Média	=B3003/3000												
3005														
3006														
3007														
3008														
3009														
3010														

Compute a fórmula da média:

$$\bar{y} = \frac{\sum_{i=1}^n y_i}{n} = \frac{\sum_{i=1}^{3000} escolaridade_i}{3000}$$

Digite

140%

DDM.xml - Microsoft Excel

Arquivo | Página Inicial | Inserir | Layout da Página | Fórmulas | Dados | Revisão | Exibição | Acrobat

Calibri 11

Fonte | Alinhamento | Número | Estilos de Célula

PROJ.LIN X ✓ fx =MÉDIA(B2:B6001)

	A	B	C	D	E	F	G	H	I	J	K
1	ID	escolaridade	idade	luz	renda_pc	tratado	conjugue	tempo	tra_tem		
5989	1572	8	50	1	250.399994	1	1	1	1		
5990	1574	8	50	1	256	1	1	1	1		
5991	1535	8	50	1	2400	1	1	1	1		
5992	1454	8	51	1	3843.33325	0	1	1	0		
5993	1490	8	51	1	615	1	1	1	1		
5994	1420	8	52	1	1358	1	0	1	1		
5995	1282	8	53	1	600	0	1	1	0		
5996	1292	8	53	1	2860	0	1	1	0		
5997	1327	8	53	1	600.666687	1	1	1	1		
5998	1002	8	57	1	1260	0	1	1	0		
5999	924	8	58	1	860	0	0	1	0		
6000	895	8	58	1	1485	0	1	1	0		
6001	940	8	58	1	1053	1	0	1	1		
6002											
6003	Somatório	17689									
6004	Média	2.94816667									
6005	Média	=MÉDIA(B2:B6001)									
6006											

Sheet1

Edita

Use o assistente de funções para implementar a fórmula MÉDIA do Excel e verifique que os valores são exatamente iguais. Depois repita o processo para as outras variáveis

DDM.xlsx - Microsoft Excel

Arquivo Página Inicial Inserir Layout da Página Fórmulas Dados Revisão Exibição Acrobat

Calibri 11 A A

Quebrar Texto Automaticamente

Mesclar e Centralizar

Área de Tran... Fonte Alinhamento

L3003

Use o assistente de fórmulas do Excel para computar as principais medidas de tendência e dispersão

	A	B	C	D	E				
1	ID	escolaridade	idade	luz	renda_pc				
2993	815	3	58	1	510				
2994	818	2	58	1	550	1	1	0	0
2995	802	2	58	1	570	1	1	0	0
2996	779	2	58	1	590	1	1	0	0
2997	807	2	58	1	633	1	1	0	0
2998	808	2	58	1	765	1	0	0	0
2999	775	7	58	1	925	1	1	0	0
3000	783	2	58	1	1000	1	1	0	0
3001	778	3	58	1	1005	1	1	0	0
3002									
3003	Média	2.93	39.17	0.97	335.12	0.53	0.86	0.00	0.00
3004	Moda	2.00	25.00	1.00	510.00	1.00	1.00	0.00	0.00
3005	Mediana	2.00	40.00	1.00	255.00	1.00	1.00	0.00	0.00
3006	Desvio	1.73	11.51	0.17	372.16	0.50	0.34	0.00	0.00
3007	Amplitude	7.00	44.00	1.00	4000.00	1.00	1.00	0.00	0.00
3008									
3009									
3010									

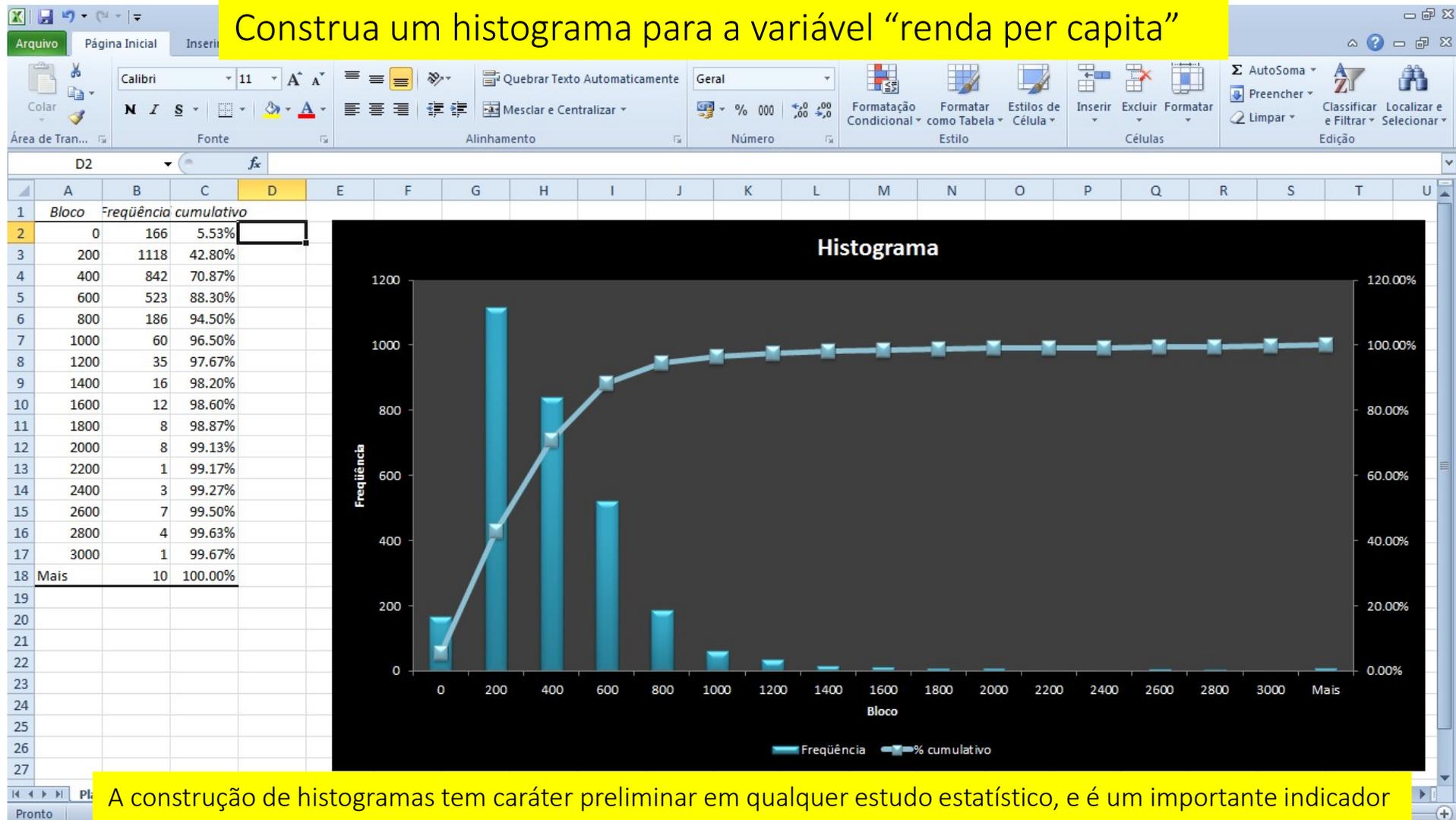
Pronto

140%

Uma ferramenta útil para uma inspeção visual das medidas de tendência e dispersão é a construção de histogramas. Para tanto, abra o assistente de “Análise de dados”, na aba “Dados”.

3003	Média	2.93	39.17	0.97	335.12	0.53	0.86	0.00	0.00
3004	Moda	2.00	25.00	1.00	510.00	1.00	1.00	0.00	0.00
3005	Mediana	2.00	40.00	1.00	255.00	1.00	1.00	0.00	0.00
3006	Desvio	1.73	11.51	0.17	372.16	0.50	0.34	0.00	0.00
3007	Amplitude	7.00	44.00	1.00	4000.00	1.00	1.00	0.00	0.00

Construa um histograma para a variável “renda per capita”



A construção de histogramas tem caráter preliminar em qualquer estudo estatístico, e é um importante indicador da função de distribuição de probabilidade de dados

Inferência estatística

INTERVALO DE CONFIANÇA

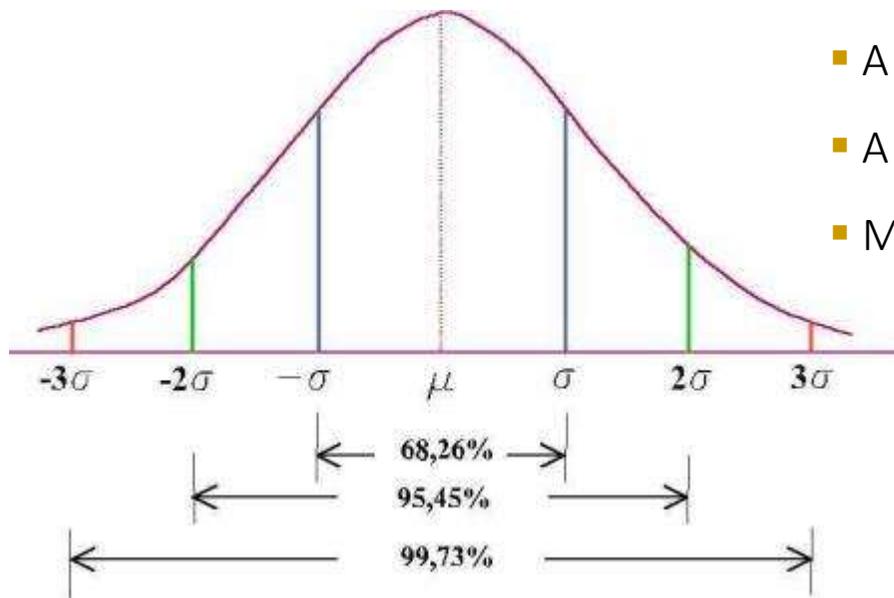
- Em avaliação econômica, com frequência desejamos saber se a média populacional de um indicador de impacto difere entre os grupos de tratamento e controle. Como trabalhamos com amostras, a ferramenta adequada para esta análise é o “intervalo de confiança”
- Para introduzir a ideia de “intervalo de confiança”, um bom exemplo começa por lembrar que em épocas de eleição seguido lemos notícias do tipo: o candidato Fulano tem 50% das intenções de voto, com margem de erro de 2% para cima e para baixo. Ou seja, espera-se que o candidato tenha de 48% a 52% dos votos com algum nível de confiança
- Para entender como podemos construir intervalos de confiança, precisamos entender dois resultados importantes em estatística: o primeiro é a chamada “Lei dos grandes números” (LGN); e o segundo é o “Teorema do Limite Central” (TLC)

LEI DOS GRANDES NÚMEROS

- A LGN é intuitiva. Ela nos diz que, sendo μ a média populacional de uma variável aleatória, $\bar{y} = \frac{\sum_{i=1}^n y_i}{n} \rightarrow \mu$ se $n \rightarrow N$ (quanto maior a amostra, mais a média amostral estará próxima da média populacional)
- Mais do que escrever $n \rightarrow N$, é comum se apresentar nos livros o resultado $n \rightarrow \infty$, chamado de resultado assintótico. Mas a ideia é a mesma: quanto maior a amostra, mais a média amostral estará próxima da média populacional!

TEOREMA DO LIMITE CENTRAL E A DISTRIBUIÇÃO NORMAL

- A distribuição normal é uma das mais importantes distribuições de probabilidade
- O Teorema do Limite Central (TLC) postula que, numa população de tamanho N , se você retirar A amostras de tamanho n , você verificará que $\frac{\bar{y}_a - \mu}{S_a} \sim N(0, 1)$, onde $a = 1, 2, \dots, A$ indica a amostragem. $S_a = \sigma / \sqrt{n}$ é o desvio padrão de \bar{x}_a



- A representação formal é $N(\mu, \sigma^2)$
- A distribuição é simétrica
- Média = Mediana = Moda

INTERVALO DE CONFIANÇA: CONECTANDO A LGN E O TLC

- Voltando para o exemplo das eleições. Usando técnicas de amostragem, que não serão abordadas aqui, os institutos de pesquisa primeiro definem um tamanho n para atender a LGN
- Em seguida, perguntam ao eleitor i se ele votará no candidato. Marcando $y_i = 1$ para o caso afirmativo e $y_i = 0$ para o caso contrário
- Então, pela LNG sabe-se que $\bar{y}_a \rightarrow \mu$, onde a é a amostragem, \bar{y}_a é o percentual de intenções de voto nesta amostragem e μ é a verdadeira (e desconhecida) votação (e.g., 50%)
- Para calcular a margem de erro, usa-se o TLC observando-se que

$$z_a = \frac{\bar{y}_a - \mu}{\sigma / \sqrt{n}} \text{ e que } z_a \sim N(0, 1)$$

- Pelas características da distribuição normal ocorre que:

$$Pr\left(\bar{y}_a - z^* \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{y}_a + z^* \frac{\sigma}{\sqrt{n}}\right) = \alpha\%$$

INTERVALO DE CONFIANÇA E DISTRIBUIÇÃO NORMAL PADRÃO

Pelas características da Normal Padrão sabemos que com $\alpha\%$ de confiança um valor z_α encontra-se no intervalo $[-z^*; +z^*]$ (valores tabelados)

$$z_\alpha \sim N(0, 1) \Rightarrow Pr(-z^* \leq z_\alpha \leq z^*) = \alpha\%$$

$$Pr\left(-z^* \leq \frac{\bar{y}_a - \mu}{\sigma/\sqrt{n}} \leq z^*\right) = \alpha\%$$

$$Pr\left(-z^* \frac{\sigma}{\sqrt{n}} \leq \bar{y}_a - \mu \leq z^* \frac{\sigma}{\sqrt{n}}\right) = \alpha\%$$

$$Pr\left(-\bar{y}_a - z^* \frac{\sigma}{\sqrt{n}} \leq -\mu \leq -\bar{y}_a + z^* \frac{\sigma}{\sqrt{n}}\right) = \alpha\%$$

$$Pr\left(\bar{y}_a - z^* \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{y}_a + z^* \frac{\sigma}{\sqrt{n}}\right) = \alpha\%$$

Com um pouco de manipulação algébrica podemos determinar com $\alpha\%$ de confiança que μ se encontra no intervalo $\left[\bar{y}_a - z^* \frac{\sigma}{\sqrt{n}}; \bar{y}_a + z^* \frac{\sigma}{\sqrt{n}}\right]$

INTERVALO DE CONFIANÇA E DISTRIBUIÇÃO t

- Existe um problema de ordem prática para se determinar com $\alpha\%$ de confiança que μ se encontra no intervalo $\left[\bar{y}_a - z^* \frac{\sigma}{\sqrt{n}}; \bar{y}_a + z^* \frac{\sigma}{\sqrt{n}} \right]$. Esse intervalo está escrito em função de σ , o desvio padrão populacional, que é desconhecido do pesquisador
- Felizmente, sabemos que o TLC também vale para a estimativa amostral do desvio padrão, S . E neste caso podemos determinar com $\alpha\%$ de confiança que μ se encontra no intervalo $\left[\bar{y}_a - t^* \frac{S}{\sqrt{n}}; \bar{y}_a + t^* \frac{S}{\sqrt{n}} \right]$. Onde t^* são os valores tabelados da distribuição t

INTERVALO DE CONFIANÇA E TESTE DE HIPÓTESES

- Sabendo que com $\alpha\%$ de confiança que μ se encontra no intervalo $\left[\bar{y}_a - t^* \frac{s}{\sqrt{n}}; \bar{y}_a + t^* \frac{s}{\sqrt{n}}\right]$, então podemos testar a hipótese de que μ seja um determinado valor

- Por exemplo, se queremos testar a hipótese de que $\mu = 0$, basta observar se 0 pertence ao intervalo

$$IC_{\alpha\%} = \left[\bar{y}_a - t^* \frac{s}{\sqrt{n}}; \bar{y}_a + t^* \frac{s}{\sqrt{n}}\right]$$

- ✓ Se $0 \in IC_{\alpha\%}$ não podemos rejeitar a hipótese de que $\mu = 0$

INTERVALO DE CONFIANÇA E TESTE DE DIFERENÇA DE MÉDIA

- Quando desejamos saber se a média de uma variável aleatória é diferente entre dois grupos (homens e mulheres, atendidos e não atendidos pelo projeto social etc.), não basta comparar os valores das duas médias para saber se há diferença. Por quê? Porque essas médias são médias amostrais e, portanto, são variáveis aleatórias
- É preciso saber se essas médias são estatisticamente diferentes a um dado nível de confiança
- Um procedimento simples de teste consiste em construir dois intervalos, um para cada grupo, e verificar se o limite superior de um está descolado do limite inferior do outro

Exercícios sobre inferência

estatística:

Novamente usaremos o arquivo

“ex1”

DDM.xlsx - Microsoft Excel

Arquivo Página Inicial Inserir Layout da Página Fórmulas Dados Revisão Exibição Acrobat

Calibri 11 A A Quebrar Texto Automaticamente Geral

Colar Área de Tran... Fonte Alinhamento Número

Formatação Condicional Formatar como Tabela Estilos de Célula

Inserir Excluir Formatar Células

AutoSoma Preencher Limpar Classificar e Filtrar Localizar e Selecionar Edição

MODO.ÚNICO X ✓ fx =INT.CONFIANÇA.T(5%;B3006;3000)

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	ID	escolaridade	idade	luz	renda_pc	tratado	conjuge	tempo	tra_tem				
2997	807	2	58	1	633	1	1	0	0				
2998	808	2	58	1	765	1	0	0	0				
2999	775	7	58	1	925	1	1	0	0				
3000	783	2	58	1	1000	1	1	0	0				
3001	778	3	58	1	1005	1	1	0	0				
3002													
3003	Média	2.93	39.17	0.97	335.12	0.53	0.86	0.00	0.00				
3004	Moda	2.00	25.00	1.00	510.00	1.00	1.00	0.00	0.00				
3005	Mediana	2.00	40.00	1.00	255.00	1.00	1.00	0.00	0.00				
3006	Desvio	1.73	11.51	0.17	372.16	0.50	0.34	0.00	0.00				
3007	Amplitude	7.00	44.00	1.00	4000.00	1.00	1.00	0.00	0.00				
3008	Int	=INT.CONFIANÇA.T(5%;B3006;3000)											
3009	Inf	=INT.CONFIANÇA.T(alfa; desv_padrao; tamanho)											
3010	Sup												
3011													
3012													
3013													
3014													

Edita

A função INT.CONFIANÇA.T computa o valor $t^* \frac{s}{\sqrt{n}}$, onde você informa o nível de significância $1 - \alpha\%$, no exemplo 5%, e o programa usa o valor tabelado de t^* para este nível de confiança/significância

DDM.xlsx - Microsoft Excel

Arquivo Página Inicial Inserir Layout da Página Fórmulas Dados Revisão Exibição Acrobat

Colar Calibri 11 A A Quebrar Texto Automaticamente Geral Formatar Estilos de Inserir Excluir Formatar AutoSoma Preencher Limpar Classificar e Filtrar Localizar e Selecionar Edição

É possível aceitar a hipótese de que a média da renda per capita da população é de R\$ 340?

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	ID	escolaridade	idade	luz	renda_pc	tratado	conjuce	tempo	tra_tem				
2997	807	2	58	1	633	1	1	0	0				
2998	808	2	58	1	765	1	0	0	0				
2999	775	7	58	1	925	1	1	0	0				
3000	783	2	58	1	1000	1	1	0	0				
3001	778	3	58	1	1005	1	1	0	0				
3002													
3003	Média	2.93	39.17	0.97	335.12	0.53	0.86	0.00	0.00				
3004	Moda	2.00	25.00	1.00	510.00	1.00	1.00	0.00	0.00				
3005	Mediana	2.00	40.00	1.00	255.00	1.00	1.00	0.00	0.00				
3006	Desvio	1.73	11.51	0.17	372.16	0.50	0.34	0.00	0.00				
3007	Amplitude	7.00	44.00	1.00	4000.00	1.00	1.00	0.00	0.00				
3008	Int	0.06	0.41	0.01	13.32	0.02	0.01	#NÚM!	#NÚM!				
3009	Inf	2.87	38.75	0.96	321.80	0.51	0.85	#NÚM!	#NÚM!				
3010	Sup	2.99	39.58	0.98	348.45	0.55	0.88	#NÚM!	#NÚM!				
3011													
3012													
3013													
3014													

Observe que os intervalos das últimas duas variáveis não puderam ser computados, já que estamos trabalhando apenas com o primeiro período de tempo

Em seguida você cria os limites superiores e inferiores do intervalo...

$$\left[\bar{y}_a - t^* \frac{S}{\sqrt{n}}; \bar{y}_a + t^* \frac{S}{\sqrt{n}} \right]$$

Agora vamos fazer alguns testes de diferenças de média usando o assistente de Tabela Dinâmica para a construção de estatísticas de diferentes grupos

The screenshot shows Microsoft Excel with a PivotTable and a dialog box. The PivotTable displays statistical data for two groups: 't=1' and 't=0'. The dialog box 'Criar Tabela Dinâmica' is open, showing the data source as '=0!\$A\$1:\$I\$3001' and the location as 'Local:'. The PivotTable data is as follows:

		t=1	t=0						
3001		3	58	1	1005	1			
3003	Média	2.93	39.17	0.97	335.12	0.53	0.01	#NÚM!	#NÚM!
3004	Moda	2.00	25.00	1.00	510.00	1.00	1.00	#NÚM!	#NÚM!
3005	Mediana	2.00	40.00	1.00	255.00	1.00	1.00	#NÚM!	#NÚM!
3006	Desvio	1.73	11.51	0.17	372.16	0.50	0.01	#NÚM!	#NÚM!
3007	Amplitude	7.00	44.00	1.00	4000.00	1.00	1.00	#NÚM!	#NÚM!
3008	Int	0.06	0.41	0.01	13.32	0.02	0.01	#NÚM!	#NÚM!
3009	Inf	2.87	38.75	0.96	321.80	0.51	0.85	#NÚM!	#NÚM!
3010	Sup	2.99	39.58	0.98	348.45	0.55	0.88	#NÚM!	#NÚM!

DDM.xlsx - Microsoft Excel

Ferramentas de Tabela Dinâmica

Arquivo | Página Inicial | Inserir | Layout da Página | Fórmulas | Dados | Revisão | Exibição | Acrobat

Opções | Design

Nome da Tabela Dinâmica: Tabela dinâmica1 | Campo Ativo: conjuge

Opções | Configurações do Campo

Tabela Dinâmica | Campo Ativo | Agrupar

Agrupar Seleção | Desagrupar | Agrupar Campo

Classificar | Inserir Segmentação de Dados

Atualizar | Alterar Fonte de Dados

Limpar | Selecionar | Mover Tabela Dinâmica

Cálculos | Gráfico Dinâmico | Ferramentas OLAP | Teste de Hipóteses

Lista de Campos | Botões +/- | Cabeçalhos de Campos

Mostrar

	A	B	C	D	E
2					
3	Rótulos de Linha	Contagem de idade	Média de idade	DesvPad de idade	
4	0	410	37.25853659	11.77152039	
5	1	2590	39.46872587	11.44012182	
6	Total Geral	3000	39.16666667	11.50905622	
7					
8					
9					
10					
11					
12					
13					
14					
15					
16					
17					
18					
19					
20					

Lista de campos da tabela dinâmica

Escolha os campos para adicionar ao relatório:

- ID
- escolaridade
- idade
- luz
- renda_pc
- tratado
- conjuge
- tempo
- tra_tem

Arraste os campos entre as áreas abaixo:

Filtro de Relatório

Rótulos de Linha

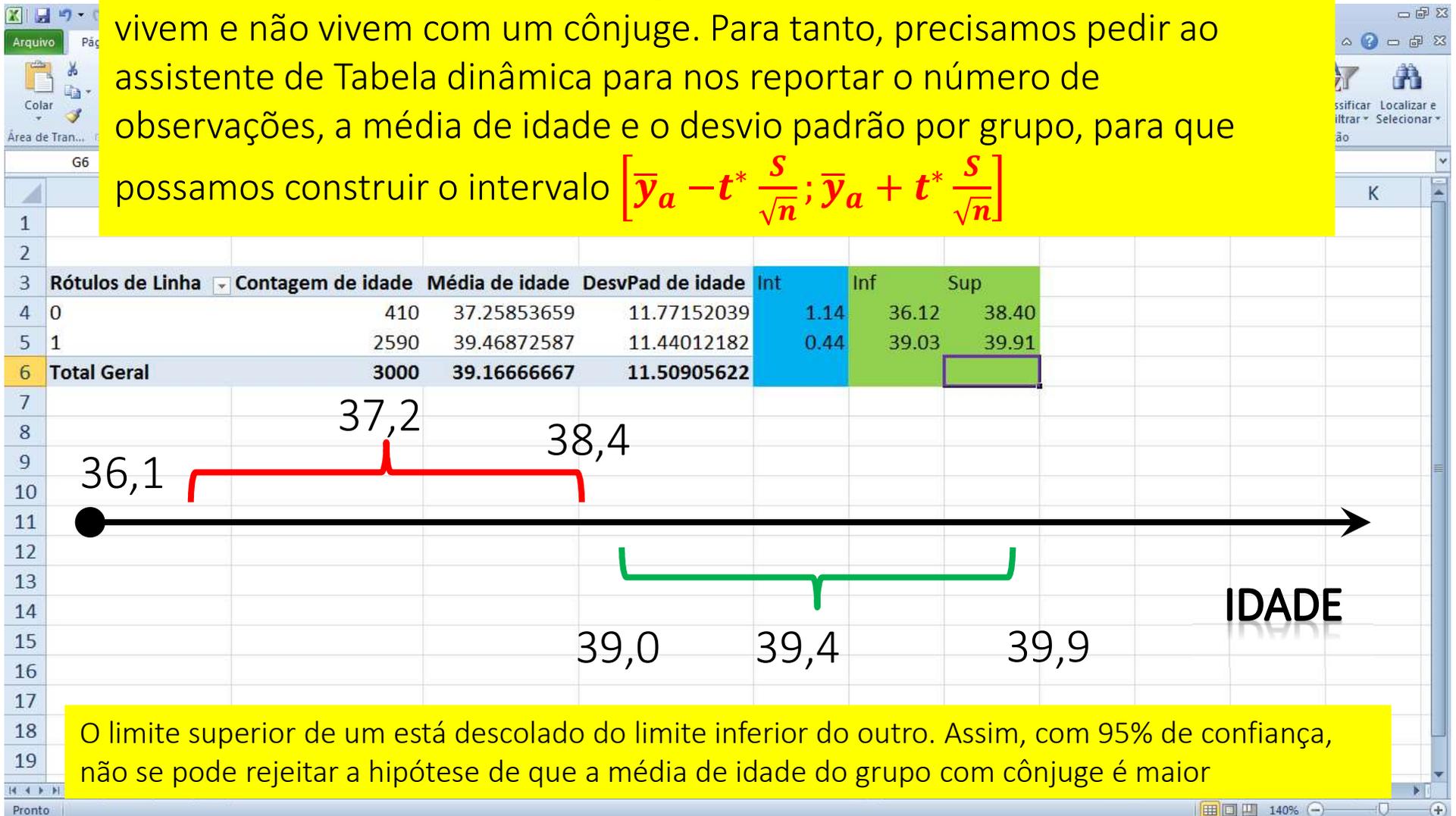
Rótulos de Coluna

Adiar Atualização do Layout

Atualizar

Posicione na linha a variável que indica o estado civil da mulher: "1" se vive com um cônjuge e "0" no caso contrário (detalhes p. 172)

Testaremos a hipótese de diferença de idade média entre mulheres que vivem e não vivem com um cônjuge. Para tanto, precisamos pedir ao assistente de Tabela dinâmica para nos reportar o número de observações, a média de idade e o desvio padrão por grupo, para que possamos construir o intervalo $\left[\bar{y}_a - t^* \frac{S}{\sqrt{n}}; \bar{y}_a + t^* \frac{S}{\sqrt{n}} \right]$



O limite superior de um está descolado do limite inferior do outro. Assim, com 95% de confiança, não se pode rejeitar a hipótese de que a média de idade do grupo com cônjuge é maior

Retorne agora para base de dados completa do programa “Mulher do Futuro”. O exercício é construir intervalos de 95% de confiança para médias da renda per capita “antes versus depois” e “tratamento versus controle” do programa.

The image shows an Excel spreadsheet with a data table and a dialog box for creating a dynamic table. The data table has the following columns: ID, escolaridade, idade, luz, renda_pc, tratado, conjuge, tempo, and tra_tem. The dialog box is titled "Criar Tabela Dinâmica" and has the following options:

- Escolha os dados que deseja analisar:
 - Selecionar uma tabela ou intervalo: Tabela/Intervalo: Sheet1!\$A\$1:\$I\$6001
 - Usar uma fonte de dados externa: Escolher Conexão...
- Nome da conexão:
- Escolha onde deseja que o relatório de tabela dinâmica seja colocado:
 - Nova Planilha
 - Planilha Existente: Local: []

Buttons: OK, Cancelar

ddm.xlsx - Microsoft Excel

Ferramentas de Tabela Dinâmica

Arquivo Página Inicial Inserir Layout da Página Fórmulas Dados Revisão Exibição Acrobat

Opções Design

Nome da Tabela Dinâmica: Tabela dinâmica1 Campo Ativo:

Opções Tabela Dinâmica Configurações do Campo Campo Ativo

Agrupar Seleção Desagrupar Agrupar Campo Agrupar

Classificar Inserir Segmentação de Dados Classificar e Filtrar

Atualizar Alterar Fonte de Dados Dados

Limpar Selecionar Mover Tabela Dinâmica Ações

Cálculos Gráfico Dinâmico Ferramentas OLAP Teste de Hipóteses Ferramentas

Lista de Campos Botões +/- Cabeçalhos de Campos Mostrar

Rótulos de Linha	Média de renda_pc	Contagem de renda_pc	DesvPad de renda_pc	Média de renda_pc	Contagem de r
0	310.3276596	1410	347.598396	457.7524594	
1	357.1106918	1590	391.4607943	634.2272192	
Total Geral	335.1226667	3000	372.1634222	551.2840821	

Lista de campos da tabela dinâmica

Escolha os campos para adicionar ao relatório:

- ID
- escolaridade
- idade
- luz
- renda_pc
- tratado
- conjuge
- tempo
- tra_tem

Arraste os campos entre as áreas abaixo:

Filtro de Relatório

Rótulos de Coluna

- tempo
- Σ Valores

Rótulos de Linha

- tratado
- Σ Valores
- Média de renda_pc
- Contagem de renda_pc
- DesvPad de renda_pc

Adiar Atualização do Layout

Atualizar

Pronto

100%

Faça a relação “antes versus depois” na coluna e a relação “tratamento versus controle” na linha, computando a média, o número de observações e o desvio padrão para a construção do intervalo $\left[\bar{y}_a - t^* \frac{s}{\sqrt{n}}; \bar{y}_a + t^* \frac{s}{\sqrt{n}} \right]$

ddm.xlsx - Microsoft Excel

Arquivo Página Inicial Inserir Layout da Página Fórmulas Dados Revisão Exibição Acrobat

Calibri 11

Fonte Alinhamento Número

Formatação Condicional Formatar como Tabela Estilos de Célula

Inserir Excluir Formatar Células

AutoSoma Preencher Limpar Classificar e Filtrar Localizar e Selecionar Edição

H15

	B	C	D	E	F	G	H	I	J	K
1										
2										
3	Rótulos de Colu									
4	0			1			Total Média de re	Total Contagem d	Total DesvPad de	renda_pc
5	Média de renda_ Contagem de ren DesvPad de renda Média de renda_ Contagem de ren DesvPad de renda_pc									
6	310.3276596	1410	347.598396	457.7524594	1410	342.5028914	384.0400595	2820	352.788387	
7	357.1106918	1590	391.4607943	634.2272192	1590	378.549182	495.6689555	3180	409.1800451	
8	335.1226667	3000	372.1634222	551.2840821	3000	372.5597917	443.2033744	6000	387.7028237	
9										
10										
11		Antes		Depois						
12		Int		Int						
13	Controle	18.16		17.89						
14	Tratado	19.26		18.62						
15										
16										
17		Antes		Depois						
18		Inf	Sup	Inf	Sup					
19	Controle	292.17	328.49	439.86	475.65					
20	Tratado	337.85	376.37	615.61	652.85					
21										
22										
23										

Observe que, com 95% de confiança, as médias dos quatro grupos de observação são diferentes

Plan2 Sheet1

Pronto 120%

Exercício complementar

Considere uma pesquisa eleitoral onde perguntam ao eleitor i se ele votará em determinado candidato, marcando $y_i = 1$ para o caso afirmativo e $y_i = 0$ para o caso contrário. Logo, trata-se de um experimento Bernoulli com $\mu = p$ e $\sigma = \sqrt{p(1-p)}$. Considere que se perguntassem a todos os eleitores o candidato teria 60% das intenções de voto, ou seja, $p = 0,6$. Simule 1.000 amostragens de 100 eleitores e compute $z_a = \frac{\bar{y}_a - \mu}{\sigma/\sqrt{n}} = \frac{\bar{y}_a - 0,6}{\sqrt{0,6(1-0,6)}/100}$. Faça um histograma desta variável para, outra vez, verificar a validade do TLC. Compute a média e o desvio padrão de z_a . Eles se aproximam de 0 e de 1, respectivamente? Qual a razão disto?

Gere as amostragens através de uma distribuição Bernoulli

Geração de número aleatório

Número de variáveis:

Número de números aleatórios:

Distribuição:

Parâmetros

Valor p =

Semente aleatória:

Opções de saída

Intervalo de saída:

Nova planilha:

Nova pasta de trabalho

Pasta1 - Microsoft Excel

Arquivo Página Inicial Inserir Layout da Página Fórmulas Dados Revisão Exibição Acrobat

Obter Dados Externos Conexões Existentes Atualizar tudo Editar Links Conexões Classificar Filtro Reaplicar Avançado Ferramentas de Dados Estrutura de Tópicos Análise

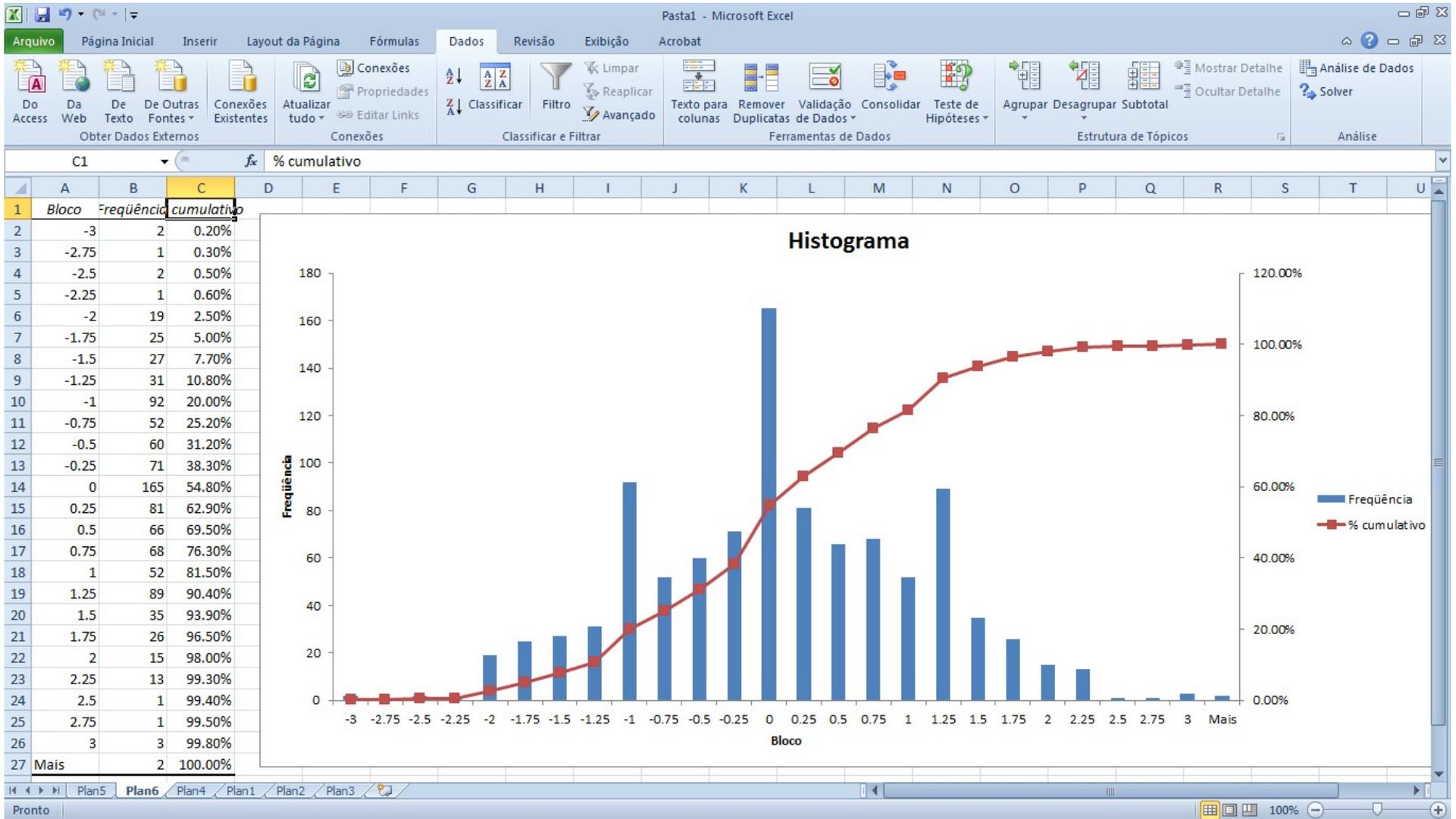
INT.CONFIANÇA.T $\text{=(MÉDIA(A1:CV1)-0.6)/RAIZ(0.6*(1-0.6)/100)}$

	CN	CO	CP	CQ	CR	CS	CT	CU	CV	CW	CX	CY	CZ	DA	D
1	0	1	1	0	1	1	1	1	1	$\text{=(MÉDIA(A1:CV1)-0.6)/RAIZ(0.6*(1-0.6)/100)}$					
2	1	1	1	1	0	1	0	0	1	MÉDIA(núm1; [núm2]; ...)					
3	1	1	0	0	0	1	1	1	1						
4	1	0	0	0	1	1	1	1	1						
5	1	1	1	1	1	1	1	0	1						
6	0	1	1	0	0	1	0	0	1						
7	0	1	1	1	0	0	1	0	1						
8	1	1	1	1	1	1	1	0	0						
9	1	1	0	0	0	0	1	0	1						
10	1	1	1	1	0	1	0	0	0						
11	0	1	0	0	0	1	0	0	0						
12	1	1													
13	0	0													
14	0	0													
15	1	0													
16	0	0													
17	0	1													
18	1	0	1	0	1	1	0	1	1						
19	0	0	1	0	1	1	1	1	1						

Plan4 Plan1 Plan2 Plan3

Edita 140%

Compute $Z_a = \frac{\bar{y}_a - \mu}{\sigma / \sqrt{n}} = \frac{\bar{y}_a - 0,6}{\sqrt{0,6(1-0,6)/100}}$



SUMÁRIO DO CURSO

- ~~1. Conceitos básicos em avaliação de impacto~~
- ~~2. Revisão de estatística e uso avançado de planilhas eletrônicas~~
3. Modelo de Resultados Potenciais
4. Econometria: revisão do modelo de mínimos quadrados
5. Desenho de uma avaliação para uma política
6. Modelo de Diferenças em Diferenças
7. Técnicas de Pareamento
8. Estudo de caso e aplicações de diferentes técnicas econométricas
9. Discussão sobre os trabalhos de conclusão de curso
10. Revisão de matemática financeira
11. O Cálculo do Retorno Econômico
12. Conclusão do curso: apresentação dos trabalhos

CONTRAFACTUAL

- O problema central da área de avaliação de programas é construir o **contrafactual** do grupo tratado pelo programa
- Na forma mais simples de apresentar o problema, podemos pensar que qualquer indivíduo está sempre em uma de duas situações mutuamente excludentes: ter sido ou não ter sido tratado
- Idealmente, o melhor grupo de comparação para os indivíduos tratados seria formado pelos mesmos indivíduos na situação em que eles não fossem tratados. Contudo, claramente não é possível observar os mesmos indivíduos na condição de tratados e de não tratados ao mesmo tempo

CONTRAFACTUAL, GRUPO DE CONTROLE E RESULTADO POTENCIAL

- O desafio do avaliador é encontrar um grupo de indivíduos que represente adequadamente a situação de não tratamento, ou seja, um grupo que funcione como um bom contrafactual do grupo tratado, chamado de **grupo de controle**
- Um elemento crucial para um bom entendimento desse arcabouço é ter sempre em mente a **potencial existência** de uma situação contrafactual
- Por isso usamos o nome de “**Modelo de Resultados Potenciais**” para o que veremos daqui a pouco

MÉTODOS INGÊNUOS DE COMPARAÇÃO

- Antes de apresentarmos formalmente o Modelo de Resultados Potenciais, vale a pena analisar dois métodos “ingênuos”, porém amplamente utilizados por não especialistas para construir o grupo de controle
 - ✓ O primeiro é baseado na comparação do que ocorre com o grupo de tratamento antes e depois da intervenção
 - ✓ O segundo consiste em comparar o grupo que escolhe passar pelo programa com um grupo que decide não participar da intervenção
- Como veremos, esses métodos dificilmente são capazes de fornecer o impacto do programa

COMPARANDO OS TRATADOS ANTES E DEPOIS

- Considere uma situação em que um programa de treinamento profissional foi oferecido para trabalhadores de baixa escolaridade numa certa localidade
- Esse programa consistia de um curso ministrado por pessoal especializado e tinha como objetivo aumentar o rendimento do trabalho dos participantes
- Um pouco antes do começo do programa, todos os participantes responderam um questionário que levantou informações sobre suas características socioeconômicas, incluindo a situação de emprego e rendimento laboral naquele momento

COMPARANDO OS TRATADOS ANTES E DEPOIS (CONTINUAÇÃO)

- Suponha que o rendimento médio do trabalho para todos os indivíduos que entraram no programa de treinamento foi de R\$ 1.000
- Um ano depois, os mesmos tipos de informações foram novamente coletadas com todos os participantes do programa. O rendimento médio encontrado nesse segundo momento foi igual a R\$ 1.100
- Alguns diriam que o efeito do programa foi de R\$ 100, o que equivale a um aumento de 10% no rendimento médio inicial. Mas será que podemos afirmar que esse foi o verdadeiro efeito do programa?
- Para responder essa pergunta, temos que nos questionar se o grupo de tratamento antes do programa representa o contrafactual adequado para a situação de não tratamento depois do programa

COMPARANDO OS TRATADOS ANTES E DEPOIS (CONTINUAÇÃO)

- Existe uma série de fatores que podem ter afetado o rendimento médio do grupo tratado além do programa
- Por exemplo, é possível que a situação do mercado de trabalho da localidade onde residem os participantes do programa tenha melhorado ao longo do tempo entre as coletas das informações
- A menos que possamos controlar a influência do mercado de trabalho e de todos os outros fatores que afetam o rendimento médio do trabalho, a simples comparação entre o que ocorre com o grupo de tratados antes e depois do programa não identifica corretamente o efeito causal de uma intervenção

COMPARANDO OS TRATADOS ANTES E DEPOIS (CONTINUAÇÃO)

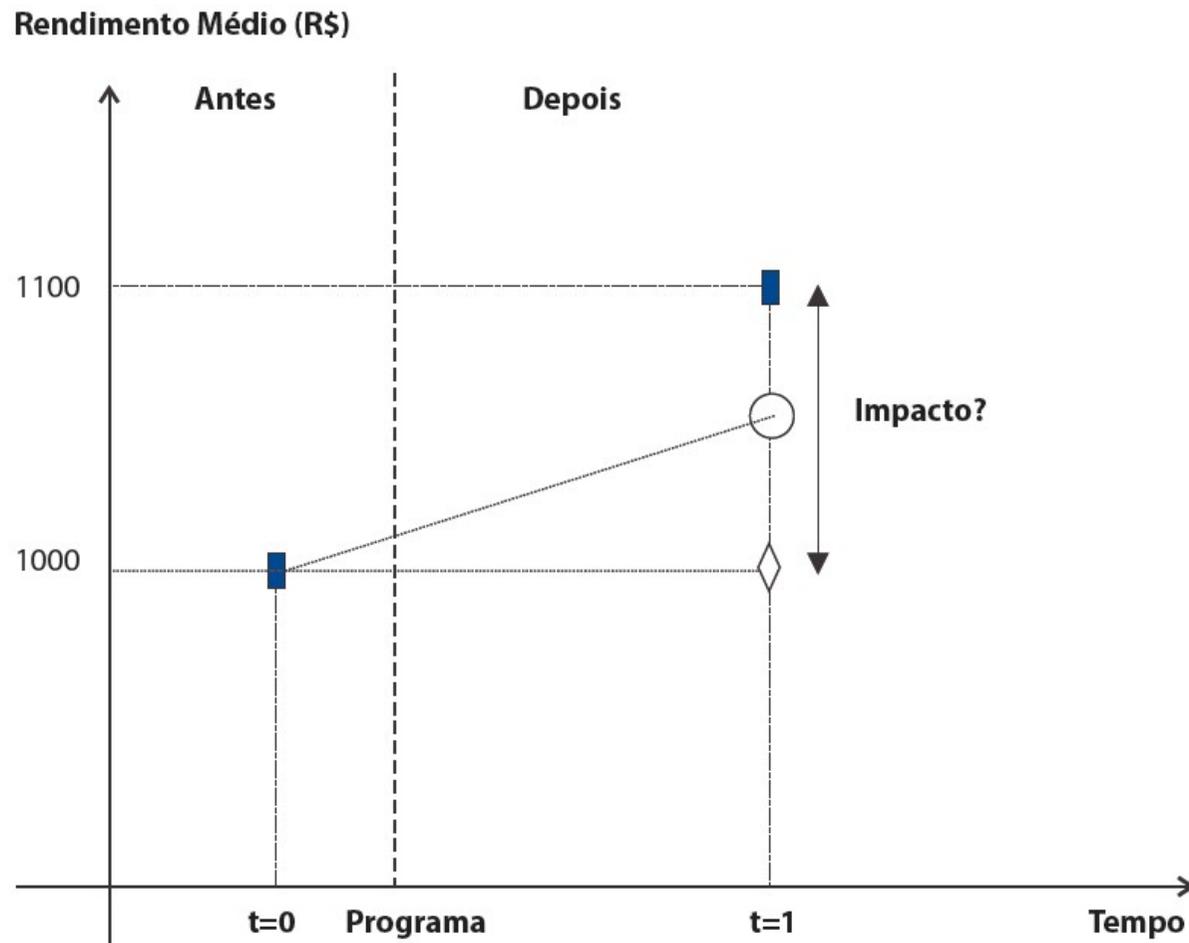


Figura da p. 37.

O verdadeiro impacto é ilustrado pela distância entre o retângulo e o círculo

COMPARANDO OS TRATADOS COM OS NÃO TRATADOS

- Outra prática “ingênua” é a que compara o grupo de tratados diretamente com um grupo de não tratados. Apenas por não ter passado pela intervenção não significa que o grupo de não tratados representa bem o que ocorreria com o grupo de tratamento caso este não tivesse sido tratado
- As principais razões para isso estão relacionadas com as diferenças nas características observáveis e não observáveis (pelo avaliador). É possível que, mesmo que os grupos sejam parecidos em atributos observáveis (gênero, idade, escolaridade, renda familiar etc.), eles difiram em atributos não observáveis (habilidades, motivação, esforço etc.)
- Na medida em que muitas das características podem afetar a variável de resultado sobre a qual se quer medir o impacto, o uso da comparação direta entre os grupos não necessariamente mede o efeito causal do programa. Afinal, as diferenças nas características não observáveis poderiam estar misturadas ao efeito do programa

O ARCABOUÇO DE RESULTADOS POTENCIAIS

- Para apresentar o arcabouço de modo formal, utilizaremos o algarismo “1” para denotar a situação em que o indivíduo é tratado e o algarismo “0” para a situação contrafactual na qual esse mesmo indivíduo não é tratado
- Estamos interessados em avaliar o impacto de um programa (e.g., um programa de qualificação de mão de obra) sobre uma variável de resultado (e.g., o rendimento do trabalho)
- Denotando por y essa variável, podemos pensar que seu valor para indivíduo i será $y_i(1)$ caso ele esteja na situação de tratado, mas poderia assumir, pelo menos potencialmente, o valor $y_i(0)$ caso ele não seja tratado
- A vantagem da formalização é explicitar que cada uma das situações está associada a um resultado potencialmente distinto para um mesmo indivíduo. Assim, se fosse possível observar o indivíduo nas duas situações, a diferença $\beta_i = y_i(1) - y_i(0)$ seria o impacto do programa para i

O ARCABOUÇO DE RESULTADOS POTENCIAIS (CONTINUAÇÃO)

- Seja T_i uma variável discreta que assume valor igual a 1 se o indivíduo participa do programa e valor 0 caso ele não participe. Trata-se, portanto, do que chamamos de uma variável binária (ou *dummy*, em inglês), cujos valores unitário ou nulo podem depender da decisão do indivíduo de participar ou não da intervenção, respectivamente
- Dado o apresentado até aqui, podemos expressar a equação de resultados potenciais como:

$$\begin{aligned} y_i &= T_i y_i(1) + (1 - T_i) y_i(0) \\ &= y_i(0) + T_i (y_i(1) - y_i(0)) \\ &= y_i(0) + \beta_i T_i \\ &= \alpha_0 + \alpha_1 x_{1i} + \dots + \alpha_K x_{Ki} + \varepsilon_i + \beta_i T_i \end{aligned}$$

Eq. (5) da p. 41

Cada “x” é uma covariada (um atributo observável que pode explicar parte do indicador de resultado), e “epsilon” representa os fatores não observáveis

ATT

- Em geral, estamos interessados em computar o que se costuma denominar de efeito médio do programa sobre os tratados (EMPT ou ATT, do inglês *average treatment effect on the treated*)

$$ATT = E[y_i(\mathbf{1})|T_i = \mathbf{1}] - E[y_i(\mathbf{0})|T_i = \mathbf{1}] = E[\beta_i]$$

- ✓ Onde $E[\dots]$ é o operador “esperança matemática”
- O problema enfrentado pelo avaliador na estimação do ATT é que não se observa $y_i(\mathbf{0})|T_i = \mathbf{1}$, logo não se observa $E[y_i(\mathbf{0})|T_i = \mathbf{1}]$

VIÉS DE AUTO SELEÇÃO

- Por outro lado, é possível observar $y_i(0)|T_i = 0$
- Notando isso, ingenuamente o avaliador pode tentar substituir $E[y_i(0)|T_i = 1]$ por $E[y_i(0)|T_i = 0]$ na estimação do ATT, viesando a estimativa do resultado

$$ATT = E[y_i(1)|T_i = 1] - E[y_i(0)|T_i = 1]$$

$$ATT = E[y_i(1)|T_i = 1] - E[y_i(0)|T_i = 0] + E[y_i(0)|T_i = 0] - E[y_i(0)|T_i = 1]$$

$$ATT = E[y_i(1)|T_i = 1] - E[y_i(0)|T_i = 0] + \underbrace{E[y_i(0)|T_i = 0] - E[y_i(0)|T_i = 1]}_{\text{viés}}$$

- Este viés também é chamado de “viés de auto seleção”

VIÉS DE AUTO SELEÇÃO: EXEMPLO

- Considere uma situação em que um programa de treinamento profissional
- y_i é o salário observado após o treinamento
- Existe uma variável não observada, x_i , que representa “motivação”. Se $x_i = 1$ o indivíduo é motivado, e se $x_i = 0$ o indivíduo não é motivado
- Considere que apenas indivíduos motivados busquem o treinamento.

Então ocorre que

$$\underbrace{E[y_i(0)|T_i = 0] - E[y_i(0)|T_i = 1]}_{\text{viés}} = \underbrace{E[y_i(0)|x_i = 0] - E[y_i(0)|x_i = 1]}_{\text{viés}}$$

- Portanto, se o indivíduos motivados tiverem salários maiores (na média), existirá um viés de auto seleção

Exercícios do Cap. 2

1. Um jornal de grande circulação apresentou numa matéria um gráfico com a série histórica da incidência de diarreia em crianças de certo município. O gráfico mostrava que a incidência de diarreia permanecia inalterada e a matéria dizia que o programa de saneamento do governo não teve nenhum impacto sobre a doença. Essa afirmativa do jornal está correta? Por quê?
2. O prefeito de uma capital implantou um programa de alfabetização de adultos no terceiro ano de seu mandato. No ano seguinte, em campanha eleitoral, ele comparou a taxa de analfabetismo da sua cidade com a do interior do estado para mostrar que o programa era um sucesso. Seria o interior do estado um bom grupo de controle para avaliar o programa implantado na capital? Discuta.
3. Qual é a mais importante função do grupo de controle em uma avaliação de impacto?
4. Tipicamente, a escolha do curso universitário é uma decisão que envolve uma análise sobre as motivações, interesses, aptidões, os custos pessoais e financeiros de fazer o curso, além de uma previsão sobre a renda que a profissão escolhida pode oferecer no futuro. Tomando o curso universitário como um programa (de treinamento), responda os seguintes itens:
 - Seria o rendimento do trabalho de pessoas com ensino superior influenciado por características normalmente não observadas pelo avaliador? Cite algumas dessas características.
 - A escolha por entrar em certo curso universitário equivale à decisão de participar de certo programa. Seria essa decisão correlacionada com o rendimento futuro do trabalho dos formados?
 - Qual o principal tipo de problema que pode surgir se compararmos diretamente os rendimentos do trabalho dos estudantes formados em diferentes cursos?

Método de Aleatorização

MÉTODO DE ALEATORIZAÇÃO

- O principal problema da avaliação é encontrar um grupo de controle que represente adequadamente o contrafactual do grupo tratado
- A situação contrafactual pode ser bastante bem aproximada pelo uso do método de aleatorização. Este método, muitas vezes chamado de método experimental, é baseado na seleção aleatória dos que farão parte dos grupos de tratamento e controle
- Quando bem implementado, o mecanismo da aleatorização fornece um balanceamento satisfatório tanto das características observadas quanto das não observadas das unidades que compõem os dois grupos
- Por construção, o método permite criar uma situação na qual não há correlação entre ser ou não tratado e os atributos das unidades de observação
- Assim, o problema do viés de auto seleção é mitigado

ALEATORIZAÇÃO TENDE A ELIMINAR O VIÉS DE AUTO SELEÇÃO

- Pela própria natureza da aleatorização, as esperanças condicionais serão idênticas as esperanças não condicionais
- Portanto ocorre:

$$\underbrace{E[y_i(\mathbf{0})|T_i = \mathbf{0}] - E[y_i(\mathbf{0})|T_i = \mathbf{1}]}_{\text{viés}} = E[y_i(\mathbf{0})] - E[y_i(\mathbf{0})] = \mathbf{0}$$

CUSTOS DA ALEATORIZAÇÃO

- Por conta de fatores éticos, financeiros e técnicos o método de aleatorização está longe de ser o predominantemente utilizado
- Sem dúvida, não se cogita utilizar aleatorização para casos de intervenções de ampla escala (e.g., programas de vacinação), ou para públicos-alvo muito específicos (e.g., pessoas com algum tipo de deficiência). No entanto, há pelo menos duas situações importantes em que esse problema ético não se coloca:
 - ✓ Quando existe escassez de vagas para atender toda a demanda do programa. Nessa situação, pode-se inclusive justificar o uso da aleatorização como o critério mais justo de alocação das vagas existentes entre os candidatos ao programa
 - ✓ Quando os recursos do programa são limitados e, apesar de se planejar atender toda a população-alvo ao final da implementação, só é possível incluir os elegíveis ao longo do tempo. Em casos como esse, ninguém será privado de participar do programa; o que se aleatoriza é somente a ordem com que os elegíveis entram no programa

Exercícios sobre aleatorização

Para o exercício, construa um conjunto de 1.000 candidatos a serem qualificados por um programa de treinamento, indexados por $i = 1, 2, \dots, 999, 1000$ como na figura

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	i													
985	984													
986	985													
987	986													
988	987													
989	988													
990	989													
991	990													
992	991													
993	992													
994	993													
995	994													
996	995													
997	996													
998	997													
999	998													
1000	999													
1001	1000													
1002														

Excel interface showing a spreadsheet with the following data:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	i	motivação												
2	1	0												
3	2	1												
4	3	=B2												
5	4													
6	5													
7	6													
8	7													
9	8													
10	9													
11	10													
12	11													
13	12													
14	13													
15	14													
16	15													
17	16													
18	17													
19	18													

Crie uma variável “motivação” igual a 1 se se *i* for ímpar, e 0 no caso contrário. A interpretação é que se **motivação = 1** o indivíduo é motivado, caso contrário não é.

The image shows a screenshot of the Microsoft Excel interface. The spreadsheet has columns labeled A through J and rows numbered 1 through 13. The formula bar at the top displays the formula `=SE(ALEATÓRIO())<1/10;1;0`. The spreadsheet data is as follows:

	A	B	C	D	E	F	G	H	I	J
1	i	motivação	tratamento							
2	1	0	<code>=SE(ALEATÓRIO())<1/10;1;0</code>							
3	2	1	<small>SE(teste_lógico; [valor_se_verdadeiro]; [valor_se_falso])</small>							
4	3	0								
5	4	1								
6	5	0								
7	6	1								
8	7	0								
9	8	1								
10	9	0								
11	10	1								
12	11	0								
13	12	1								

A yellow text box is overlaid on the spreadsheet, containing the following text:

A função “aleatório” do Excel gera um número aleatório oriundo de uma distribuição uniforme no intervalo unitário. Assim, vamos criar o seguinte sorteio: se o indivíduo receber um número menor que 1/10 ele será tratado, caso contrário não será. Tratamento = 1 indicada que ele foi sorteado.

Pasta1 - Microsoft Excel

Arquivo Página Inicial Inserir Layout da Página Fórmulas Dados Revisão Exibição Acrobat

Área de Tran... Fonte Alinhamento Número

Formatação Condicional Formatar como Tabela Estilos de Célula Inserir Excluir Formatar Células

AutoSoma Preencher Limpar Classificar e Filtrar Localizar e Selecionar Edição

MODO.ÚNICO X ✓ f_x =1000+500*B2+1000*C2

	A	B	C	D	E	F	G	H	I	J
1	i	motivação	tratamento	salário						
2	1	0	0	=1000+500*B2+1000*C2						
3	2	1	0							
4	3	0	0							
5	4	1	0							
6	5	0								
7	6	1								
8	7	0								
9	8	1								
10	9	0								
11	10	1								
12	11	0								
13	12	1	0							

Plan1 Plan2 Plan3

Aponte 200%

Crie uma salário fictício para todos.
 Independente de ser motivado ou tratado o indivíduo receberá R\$ 1.000; se ele for motivado o salário aumenta em R\$ 500; e se ele foi tratado o salário aumenta em mais R\$ 1.000

Pasta1 - Microsoft Excel

Arquivo Página Inicial Inserir Layout da Página Fórmulas Dados Revisão Exibição Acrobat

Calibri 11 A A Quebrar Texto Automaticamente Geral

Colar Fonte Alinhamento Número

Formatação Condicional Formatar como Tabela Estilos de Célula Inserir Excluir Formatar Células

AutoSoma Preencher Limpar Classificar e Filtrar Localizar e Selecionar Edição

MODO.ÚNICO X ✓ f =CONT.SE(C2:C1001;B1005)

	A	B	C	D	E	F	G	H	I	J
1	i	motivação	tratamento	salário						
995	994	1	1	2500						
996	995	0	0	1000						
997	996	1	0	1500						
998	997	0	0	1000						
999	998	1	0	1500						
1000	999	0	0	1000						
1001	1000	1	0	1500						
1002										
1003										
1004	Tratado	1	112							
1005	Controle	0	=CONT.SE(C2:C1001;B1005)							
1006			CONT.SE(intervalo; critérios)							

Plan1 Plan2 Plan3

Edita 200%

Use a função CONT.SE do Excel para contar quantos indivíduos e passam e não passam pelo tratamento.

Você já deve ter observado que este número muda a toda hora. É claro! É um número aleatório.

Pasta1 - Microsoft Excel

Arquivo Página Inicial Inserir Layout da Página Fórmulas Dados Revisão Exibição Acrobat

Área de Tran... Fonte Alinhamento Número

Formatação Condicional Formatar como Tabela Estilos de Célula Inserir Excluir Formatar Células

AutoSoma Preencher Limpar Classificar e Filtrar Localizar e Selecionar Edição

MODO.ÚNICO $\text{=SOMASE}(C2:C1001;B1005;D2:D1001)/C1005$

	A	B	C	D	E	F	G	H	I	J
1	i	motivação	tratamento	salário						
996	995	0	0	1000						
997	996	1	0	1500						
998	997	0	0	1000						
999	998	1	0	1500						
1000	999	0	0	1000						
1001	1000	1	0	1500						
1002										
1003										
1004	Tratado	1	116	2241.38						
1005	Controle	0	884	$\text{=SOMASE}(C2:C1001;B1005;D2:D1001)/C1005$						
1006										
1007										

Use a função SOMASE do Excel para calcular o salário médio dos grupos de tratamento e controle.

Plan1 Plan2 Plan3

Edita 200%

Pasta1 - Microsoft Excel

Arquivo Página Inicial Inserir Layout da Página Fórmulas Dados Revisão Exibição Acrobat

Calibri 11 A A

Colar

Área de Tran...

Fonte

Alinhamento

Número

D1007 =D1004-D1005

	A	B	C	D	E
1	i	motivação	tratamento	salário	
998	997	0	0	1000	
999	998	1	0	1500	
1000	999	0	0	1000	
1001	1000	1	0	1500	
1002					
1003					
1004	Tratado	1	97	2221.65	
1005	Controle	0	903	1253.05	
1006					
1007				968.604	
1008					
1009					

Plan1 Plan2 Plan3

Pronto

200%

Calcule a diferença das médias e veja como ela se aproxima dos R\$ 1.000 – que é o verdadeiro ATT estipulado no exercício.

Perceba que você, na condição do pesquisador, não conhece o verdadeiro ATT. Mas na condição de “feitor” do exercício conhece. Essa é uma versão simples de um experimento de Monte Carlo

OBSERVAÇÕES SOBRE O EXERCÍCIO

- Porque não observamos exatamente a estimação do ATT de R\$ 1.000? Porque mesmo aleatorizando o processo de seleção, ainda existe a probabilidade de termos mais (ou menos) pessoas motivadas na seleção. Então o ATT estimado recebe um pequeno efeito do fator “motivação”
- Pela Lei dos Grandes Números, conforme aumentarmos os tamanhos dos grupos de tratamento e controle o ATT estimado se aproximará dos R\$ 1.000
- Pelo Teorema do Limite Central, se retirarmos seguidos e diferentes grupos de tratamento desta população, a média do ATT estimado se aproximará dos R\$ 1.000
- Quanto maior o efeito do tratamento em relação ao efeito da motivação, mais o ATT estimado se aproximará dos R\$ 1.000

SUMÁRIO DO CURSO

- ~~1. Conceitos básicos em avaliação de impacto~~
- ~~2. Revisão de estatística e uso avançado de planilhas eletrônicas~~
- ~~3. Modelo de Resultados Potenciais~~
4. Econometria: revisão do modelo de mínimos quadrados
5. Desenho de uma avaliação para uma política
6. Modelo de Diferenças em Diferenças
7. Técnicas de Pareamento
8. Estudo de caso e aplicações de diferentes técnicas econométricas
9. Discussão sobre os trabalhos de conclusão de curso
10. Revisão de matemática financeira
11. O Cálculo do Retorno Econômico
12. Conclusão do curso: apresentação dos trabalhos

OS RESULTADOS POTENCIAIS E A ANÁLISE DE REGRESSÃO

- No modelo de resultados potenciais temos y_i e T_i como um indicador de resultado e uma *dummy* que indica “tratamento”, respectivamente. Lembre-se da aula passada que o modelo é baseado na seguinte equação:

$$y_i = T_i y_i(1) + (1 - T_i) y_i(0) = y_i(0) + T_i (y_i(1) - y_i(0)) = y_i(0) + \beta_i T_i$$

- Definimos que o valor de y para indivíduo i será $y_i(1)$ caso ele esteja na situação de tratado, mas poderia assumir, potencialmente, o valor $y_i(0)$ caso ele não seja tratado. Assim explicitamos que cada uma das situações está associada a um resultado potencialmente distinto para um mesmo indivíduo e, se fosse possível observar o indivíduo nas duas situações, a diferença β_i seria o impacto do programa para i

EXEMPLO DE ESTRATÉGIA DE IDENTIFICAÇÃO

- Como a variável estudada assume o valor $y_i(\mathbf{0})$ caso o indivíduo não seja tratado, ela será função apenas de características específicas dele. Por exemplo, suponha um contexto de treinamento profissional, onde $y_i(\mathbf{0})$ seria o salário do indivíduo. O salário deve variar em função atributos observáveis (gênero, idade, escolaridade, renda familiar etc.) e atributos não observáveis (habilidades, motivação, esforço etc.)
- Para simplificar, suponha que exista um salário mínimo dado por α_0 (R\$). E que o salário em média aumente em α_1 (R\$) para cada ano de escolaridade x_{1i} (que assume os valores $x_{1i} = 0, 1, 2, \dots$)
- Suponha também que a escolaridade é o único atributo observado pelo pesquisador, e que escrevemos um índice ε_i representando todo o mais que não observamos
- Nestas condições, o pesquisador pode definir o seguinte modelo: $y_i(\mathbf{0}) = \alpha_0 + \alpha_1 x_{1i} + \varepsilon_i$

EXEMPLO DE ESTRATÉGIA DE IDENTIFICAÇÃO (CONTINUAÇÃO)

- Sob a hipótese de que $y_i(0) = \alpha_0 + \alpha_1 x_{1i} + \varepsilon_i$, o salário dos indivíduos (tratados ou não tratados pelo programa) deve ser representado pela seguinte equação:

$$y_i = y_i(0) + \beta_i T_i = \alpha_0 + \alpha_1 x_{1i} + \varepsilon_i + \beta_i T_i$$

- Para simplificar, faz-se a hipótese adicional de que o impacto do programa é igual para todos os indivíduos, ou seja, $\beta_i = \beta$ para qualquer indivíduo da população. Assim, a equação acima se transformará na equação (5) da p. 41 do livro. Onde β é o **ATT** do programa

$$y_i = \alpha_0 + \alpha_1 x_{1i} + \beta T_i + \varepsilon_i$$

- Na sequência da aula nos dedicaremos a entender a técnica dos “mínimos quadrados”, que usaremos para calcular de α_0 , α_1 e β e verificar se nossas hipóteses de fato mimetizam a realidade, ou seja, se construímos uma boa “estratégia de identificação”

CONSIDERAÇÕES SOBRE AS LIMITAÇÕES DO MÉTODO

- O uso do método da regressão linear que veremos a seguir possui duas grandes limitações na avaliação de impacto:
 - ✓ A primeira, discutida em seguida, é que ele mede efeitos de correlação entre a variável de tratamento (T_i) e a variável de impacto (y_i). Então a conclusão sobre a causalidade dependerá da razoabilidade das hipóteses feitas na estratégia de identificação (que nem sempre são testáveis)
 - ✓ E a segunda, que é vista em cursos avançados de econometria, é que a análise de regressão pressupõe a ausência de correlação entre ε_i e todos os outros regressores (x_i e T_i). Se essa hipótese for violada, as estimativas dos parâmetros α_0 , α_1 e β serão viesadas
- Quanto a esta segunda limitação, as técnicas de “diferenças em diferenças” e “pareamento”, que serão vistas nas próximas aulas, tem o propósito de resolver a questão

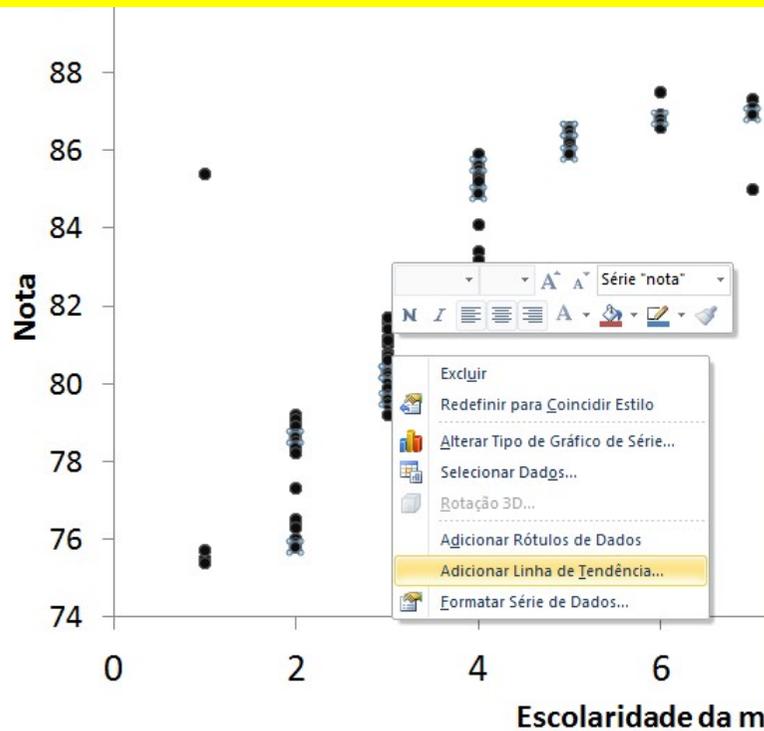
PROGRAMA “BRASIL MELHOR”

O programa “Brasil Melhor” – fictício – visa melhorar o desempenho escolar dos alunos do ensino fundamental. Esse programa foi desenvolvido com alunos da Escola Estadual João dos Santos. O desenho do programa não abrangia a sua avaliação de impacto que foi implementada após um ano de seu funcionamento. Dessa forma, só foi possível obter informações em um momento no tempo – depois do programa implementado. O banco de dados “ex2” apresenta informações sobre a nota no exame de proficiência realizado em toda a escola fundamental, a participação no programa, o sexo, a cor e os anos de estudos da mãe do aluno.

GRÁFICO DE DISPERSÃO E IDEIA DE CORRELAÇÃO

- Observe no arquivo “PSM1” que a nota média dos alunos é de 83,6
- Observe também que neste arquivo podemos observar a escolaridade da mãe
- Podemos melhorar a nossa ideia de “nota média” ao observar que existe uma correlação entre a nota e a “escolaridade da mãe”? Se “nota média” e “escolaridade da mãe” forem covariadas podemos!
- Façamos um gráfico de dispersão entre a nota do aluno e a escolaridade da mãe

Vamos aprender agora a usar as ferramentas de “análise de regressão”, que nos ajudam a melhorar a descrição da média condicionando-a a algumas covariadas!



Clique com o botão direito do mouse sobre o gráfico de dispersão e observe a ferramenta “adicionar linha de tendência”

PSM1.xml - Microsoft Excel

Ferramentas de Gráfico

Arquivo Página Inicial Inserir Layout da Página Fórmulas Dados Revisão Exibição Acrobat Design Layout Formatar

Área de Tran... Fonte

Colar

Quebrar Texto Automaticamente

Mesclar e Centralizar

Formatação Formatar Estilos de

Inserir Excluir Formatar

Células

AutoSoma

Preencher

Limpar

Classificar e Filtrar

Localizar e Selecionar

Edição

100

Brilhantes

Formatar Linha de Tendência

Opções de Linha de Tendência

Cor da Linha

Estilo da Linha

Opções de Linha de Tendência

Tipo de Tendência/Regressão

Exponencial

Linear

Logarítmica

Polinomial Ordem: 2

Potência

Média Móvel Período: 2

Nome da Linha de Tendência

Automático: Linear (nota)

Personalizado:

Previsão

Avançar: 0.0 períodos

Recuar: 0.0 períodos

Definir Interseção = 0.0

Exibir Equação no gráfico

Exibir valor de R-quadrado no gráfico

Fechar

30

20

10

0

0 2 4 6 8 10 12

Escolaridade da mãe

Gráf1 Sheet1

Pronto

89%

Primeiramente use uma tendência linear ajustando:

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

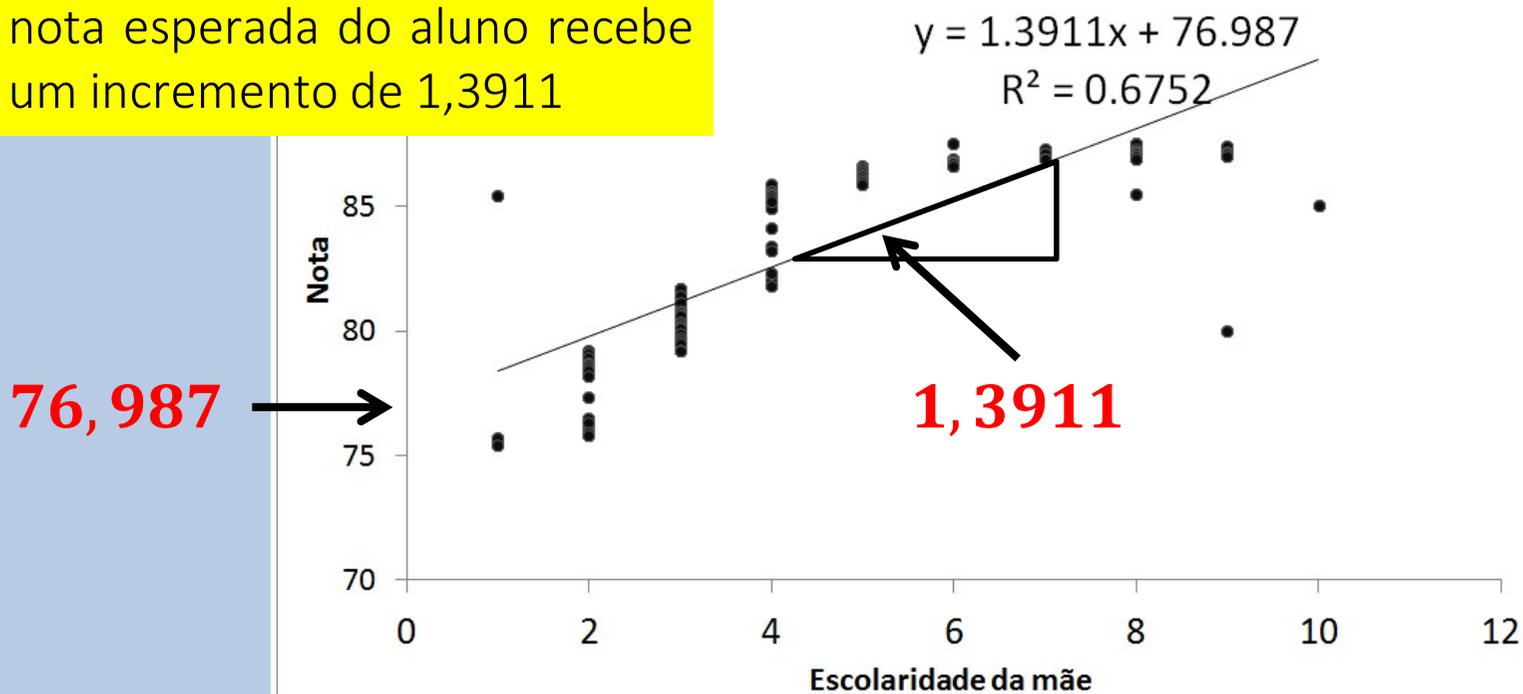
Onde y_i é a nota e x_i é a escolaridade da mãe do aluno i

Interpretação: se a mãe tem escolaridade 0, a nota esperada do aluno é de 76,987; para cada ano de escolaridade da mãe, a nota esperada do aluno recebe um incremento de 1,3911

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

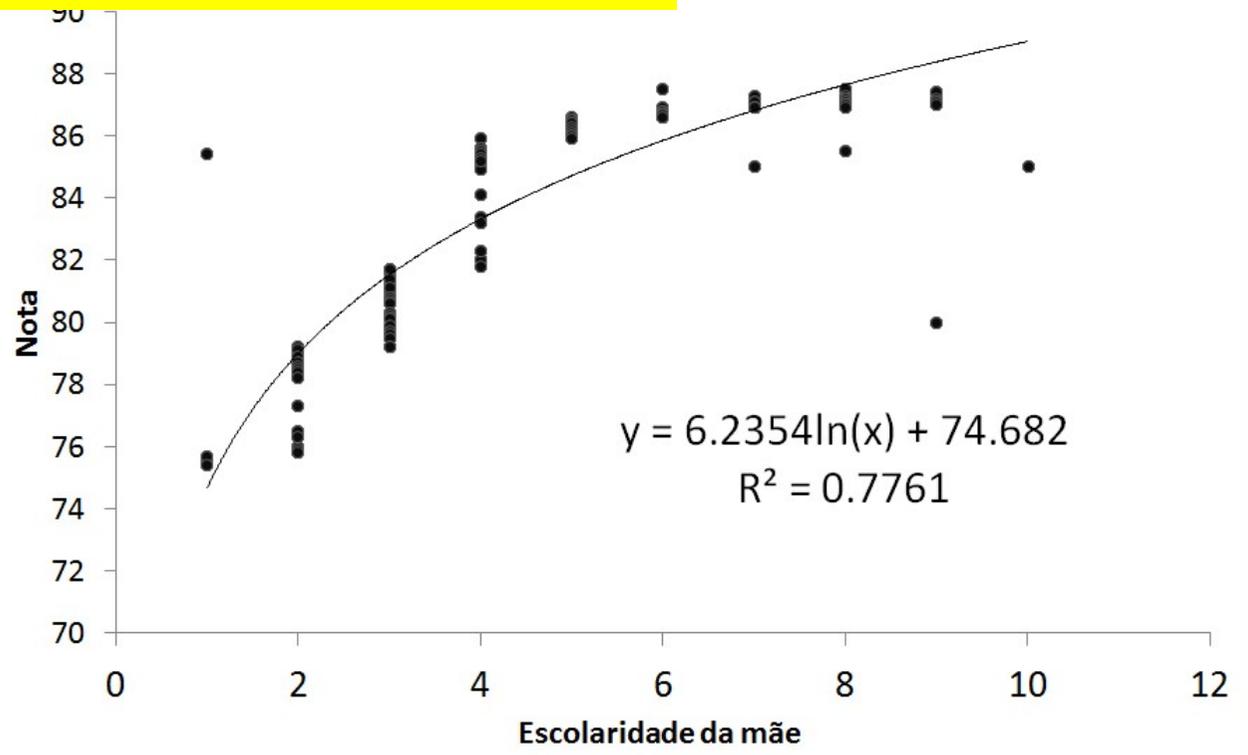
Pressupõe-se que $E(\varepsilon_i | x_i) = 0$

$$E(y_i | x_i) = \hat{\beta}_0 + \hat{\beta}_1 x_i$$
$$E(y_i | x_i) = 76,987 + 1,3911 x_i$$



Pelo modelo, a média condicionada de uma criança com mãe com 4 anos de escolaridade é de $76,98 + 1,39 \times 4 \approx 83,6$

Note que podemos fazer o ajuste de outras formas funcionais entre y_i e x_i , mas que elas sempre serão lineares em β_0 e β_1



CAUSALIDADE

- Casualidade (também referida como causação) é a relação entre um evento (a causa) e um segundo evento (o efeito), em que o segundo acontecimento é entendida como uma consequência do primeiro
- No exercício que acabamos de fazer, estamos pressupondo que exista uma relação causal entre a escolaridade da mãe e a nota da criança. O acontecimento pressuposto, por exemplo, é que em média mães com maior escolaridade ajudam os filhos com as tarefas escolares e assim eles conseguem melhorar seu desempenho

CORRELAÇÃO E CAUSALIDADE

- Uma importante observação a ser feita é que a análise de regressão é, em última instância, uma análise de correlação!
- Se uma variável causa outra em todo ou em parte (e.g., se a escolaridade da mãe leva a uma maior nota do aluno, ou se a participação no programa leva a um aumento de renda) então se observará correlação entre as variáveis
- Porém, observar correlação então significa observar causalidade. Por exemplo, o fato da população de cegonhas estar diminuindo ano após ano não é o causador da redução da taxa de fecundidade nas sociedades modernas (esta potencial correlação só existe nas histórias infantis, mas é um exemplo clássico de relação espúria)

A IDEIA DE MODELAR UM COMPORTAMENTO REPRESENTATIVO

- Com a ferramenta de ajuste de linha de tendência, o Excel está usando o método dos “mínimos quadrados”. Que é a pedra-angular da Econometria e de toda análise de regressão!
- Para entendermos a técnica, precisamos pensar um pouco mais sobre o significado de média
- Todos os conceitos de média refletem a ideia de que existe um ponto em torno do qual todas as observações oscilam

Uma observação qualquer da variável aleatória

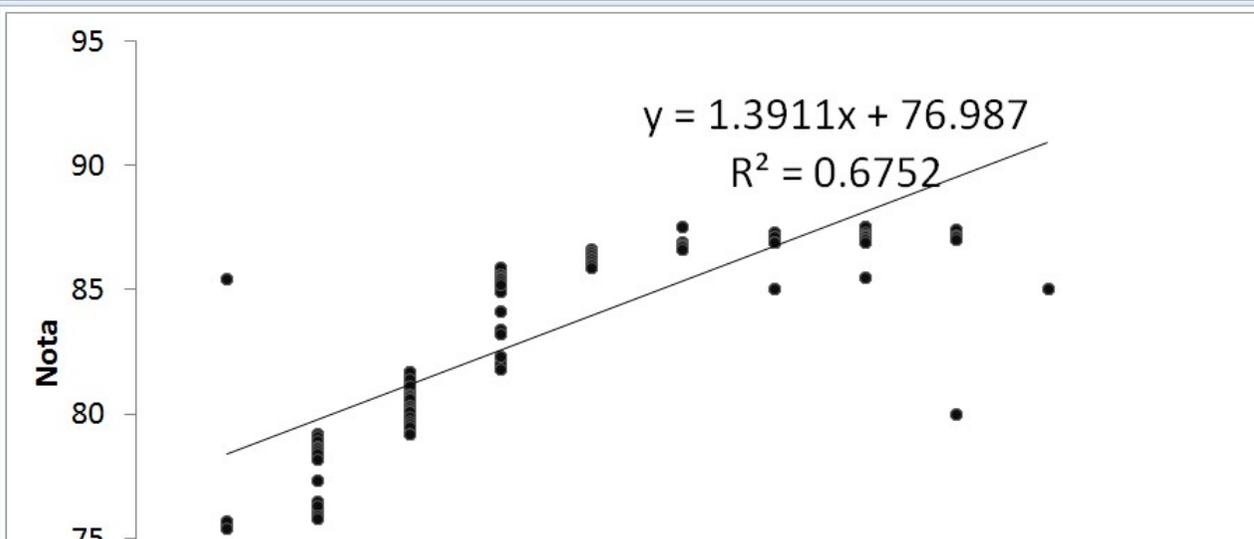
$$y_i = \bar{y} + \varepsilon_i$$

Um desvio da observação

A LÓGICA DOS MÍNIMOS QUADRADOS

- Suponha que você detém de n observações de y_i . Assim, você pode tentar minimizar um “erro médio” resolvendo o problema $\text{Min.}_{\bar{y}} \frac{\sum_{i=1}^n \varepsilon_i}{n} = \frac{\sum_{i=1}^n y_i}{n} - \bar{y}$. Mas rapidamente perceberá que isto não é possível sem alguma imposição *ad-hoc* (e.g., que a média dos erros terá de ser zero)
- Alternativamente, você pode tentar minimizar um “erro quadrado médio” resolvendo o problema $\text{Min.}_{\bar{y}} \frac{\sum_{i=1}^n \varepsilon_i^2}{n} = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n}$. E três resultados interessantes surgirão sem nenhuma imposição:
 - ✓ A média aritmética $\bar{y} = \frac{\sum_{i=1}^n y_i}{n}$ é o estimador que minimiza o “erro quadrado médio”
 - ✓ E assim, o erro médio será zero sem precisar impor nenhuma condição: $\frac{\sum_{i=1}^n \varepsilon_i}{n} = 0$
 - ✓ E assim, média aritmética é quem também minimizará a variância do erro:
$$\text{Min.}_{\bar{y}} \frac{\sum_{i=1}^n \varepsilon_i^2}{n} = \text{Min.}_{\bar{y}} \frac{\sum_{i=1}^n (\varepsilon_i - \frac{\sum_{i=1}^n \varepsilon_i}{n})^2}{n}$$
- *Por isso os mínimos quadrados são um bom critério de ajuste de médias*

Como o Excel ajustou esta reta? Minimizando a soma dos quadrados do erro em relação aos dois parâmetros que definem a reta



$$\begin{aligned} & \text{Min.}_{\beta_0, \beta_1} \sum_{i=1}^n \varepsilon_i^2 \\ & = \text{Min.}_{\beta_0, \beta_1} \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2 \end{aligned}$$

sss.xlsx - Microsoft Excel

Arquivo | Página Inicial | Inserir | Layout da Página | Fórmulas | **Dados** | Revisão | Exibição | Acrobat

Obter Dados Externos | Conexões | Atualizar tudo | Propriedades | Editar Links | Classificar e Filtrar | Ferramentas de Dados | Estrutura de Tópicos | Análise

	A	B	C	D	E	F	G	H	I	J	K	L
	ID	nota	tratado	mulher	cor	estudo_mae						
1												
2	1	86.900002	1	0	0	7						
3	2	86.900002	1	0	1	6						
4	3	86.800003	1	0	0	6						
5	4	86.699997	1	0	1	6						
6	5	86.699997				6						
7	6	86.699997				6						
8	7	86.599998				6						
9	8	86.599998				6						
10	9	86.5				5						
11	10	86.5				5						
12	11	86.400002				5						
13	12	86.400002	1	0	1	5						
14	13	86.400002	1	1	1	5						
15	14	86.400002	1	1	0	5						
16	15	86.300003	1	0	1	5						
17	16	86.099998	1	0	1	5						
18	17	86	1	0	0	5						
19	18	85.900002	1	0	0	4						

Análise de dados

Ferramentas de análise

- Estadística descritiva
- Ajuste exponencial
- Teste-F: duas amostras para variâncias
- Análise de Fourier
- Histograma
- Média móvel
- Geração de número aleatório
- Ordem e percentil
- Regressão**
- Amostragem

OK | Cancelar | Ajuda

Pronto | Gráf1 | Sheet1 | 140%

A ferramenta de “Regressão” é bem mais ampla que o simples ajuste de tendência.

Vamos estimar a linha da tendência usando esta ferramenta

sss.xlsx - Microsoft Excel

Arquivo | Página Inicial | Inserir | Layout da Página | Fórmulas

Do Access | Da Web | De Texto | De Outras Fontes | Conexões Existentes | Atualizar tudo | Conexões

Obter Dados Externos | Propriedades | Editar Links | Classificar e Filtrar | Ferramentas de Dados | Estrutura de Tópicos | Análise

Mostrar Detalhes | Ocultar Detalhes | Análise de Dados | Solver

B1

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

	A	B	C	D	E	F	G	H	I	J	K	L
1	ID	nota	tratado	mulher	cor	estudo_mae						
729	728	79.5	0	0	1	3						
730	729	79.199997	0	1	1	3						
731	730	79.199997	0	1	1	2						
732	731	79.199997	0	0	1	3						
733	732	79.099998	0	0	1	2						
734	733	78.900002	0	0	1	2						
735	734	78.699997	0	0	1	2						
736	735	78.699997	0	1	0	2						
737	736	78.599998	0	0	0	2						
738	737	78.5	0	0	1	2						
739	738	78.400002	0	0	1	2						
740	739	78.199997	0	1	0	2						
741	740	76.5	0	1	0	2						
742	741	76.300003	0	0	1	2						
743	742	75.900002	0	0	0	2						
744	743	75.800003	0	1	0	2						
745	744	75.699997	0	0	0	1						
746	745	75.400002	0	1	0	1						

Regressão

Entrada

Intervalo Y de entrada: \$B\$1:\$B\$746

Intervalo X de entrada: \$F\$1:\$F\$746

Rótulos Constante é zero

Nível de confiança: 95 %

Opções de saída

Intervalo de saída: []

Nova planilha: reg1

Nova pasta de trabalho

Resíduos

Resíduos Plotar resíduos

Resíduos padronizados Plotar ajuste de linha

Probabilidade normal

Plotagem de probabilidade normal

OK | Cancelar | Ajuda

A planilha de resultados da regressão apresenta uma série de estatísticas.

Neste curso focaremos nossa análise nos: coeficientes; erro padrão; estatística t ; e, R-Quadrado

Estadística de regressão	
R múltiplo	0.821718911
R-Quadrado	0.675221968
R-quadrado ajustado	0.674784851
Erro padrão	2.164328892
Observações	745

ANOVA					
	<i>gl</i>	<i>SQ</i>	<i>MQ</i>	<i>F</i>	<i>F de significação</i>
Regressão	1	7235.944818	7235.944818	1544.716311	1.2757E-183
Resíduo	743	3480.449428	4.684319553		
Total	744	10716.39425			

	<i>Coefficientes</i>	<i>Erro padrão</i>	<i>Stat t</i>	<i>valor-P</i>	<i>95% inferiores</i>	<i>95% superiores</i>	<i>Inferior 95.0%</i>	<i>Superior 95.0%</i>
Interseção	76.9871	0.1859	414.0404	0.0000	76.6221	77.3522	76.6221	77.3522
estudo_mae	1.3911	0.0354	39.3029	0.0000	1.3216	1.4606	1.3216	1.4606

Os coeficientes são as estimativas $\hat{\beta}_0$ e $\hat{\beta}_1$ da linha de tendência

O ERRO PADRÃO ASSOCIADO AOS COEFICIENTES

- Relembre que uma das versões do Teorema do Limite Central (TLC) postula que, se você retirar A amostras de tamanho n , você verificará que $\frac{\bar{x}_a - \mu}{s_a} \sim t$, onde $a = 1, 2, \dots, A$ indica a amostragem e $s_a = s / \sqrt{n}$ é o desvio padrão de \bar{x}_a
- Analogamente, $\hat{\beta}$ é uma estimativa amostral do verdadeiro parâmetro populacional β , e o erro padrão reportado na planilha de resultados da regressão é o equivalente do desvio padrão. Ou seja:
 - ✓ $\hat{\beta}$ é o equivalente de \bar{x}_a
 - ✓ β é o equivalente de μ
 - ✓ E o erro padrão é o equivalente de s_a

$$\frac{\hat{\beta} - \beta}{\text{erro padrão}} \sim t$$

sss.xlsx - Microsoft Excel

Arquivo Página Inicial Inserir Layout da Página Fórmulas Dados Revisão Exibição Acrobat

Obter Dados Externos

Do Access Da Web De Texto De Outras Fontes

Conexões Existentes

Atualizar tudo

Propriedades Editar Links

Conexões

Classificar e Filtrar

Limpar Reaplicar Avançado

Texto pa coluna

Mostrar Detalhe Análise de Dados

MODO.ÚNICO \times \checkmark f_x =B17/C17

	A	B	C	D
1	RESUMO DOS RESULTADOS			
2				
3	<i>Estadística de regressão</i>			
4	R múltiplo	0.821718911		=B17/C17
5	R-Quadrado	0.675221968		
6	R-quadrado ajustado	0.674784851		
7	Erro padrão	2.164328892		
8	Observações	745		
9				
10	ANOVA			
11		<i>gl</i>	<i>SQ</i>	<i>MQ</i>
12	Regressão	1	7235.944818	7235.944818
13	Resíduo	743	3480.449428	4.684319553
14	Total	744	10716.39425	
15				
16		<i>Coefficientes</i>	<i>Erro padrão</i>	<i>Stat t</i>
17	Interseção	76.9871	0.1859	414.0404
18	estudo_mae	1.3911	0.0354	39.3029
19				
20				
21				

Para uma variável qualquer, divida o valor do coeficiente pelo erro padrão. E, assim, verifique que é exatamente o valor da estatística t . Porque isso acontece? Porque é um teste de hipóteses, cuja hipótese nula é $H_0: \beta = 0$

$$\frac{\hat{\beta} - \beta}{\text{erro padrão}} \sim t$$

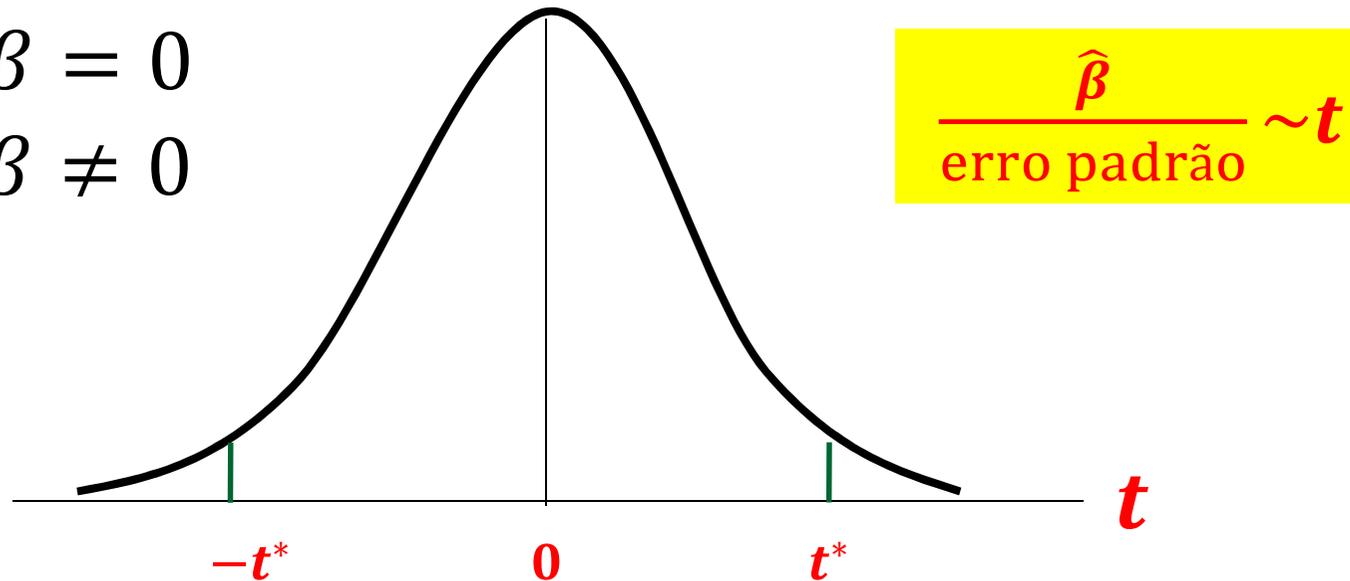
reg1 Sheet1

Edita

130%

O ESTATÍSTICA t ASSOCIADA AOS COEFICIENTES

$$\begin{cases} H_0: \beta = 0 \\ H_1: \beta \neq 0 \end{cases}$$



- A estatística t deve ser maior (menor) que o valor tabelado t^*
- A regra de bolso é $|t^*| > 2$

O R-QUADRADO

- O R-Quadrado é a medida mais popular de ajustamento da reta aos dados. Com um pouco de manipulação algébrica podemos deduzi-lo

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + \hat{\varepsilon}_i = \hat{y}_i + \hat{\varepsilon}_i$$

← Valor observado
← Valor esperado
← Resíduo

$$y_i - \bar{y} = \hat{y}_i - \bar{y} + \hat{\varepsilon}_i$$

$$(y_i - \bar{y})^2 = (\hat{y}_i - \bar{y})^2 + 2\hat{\varepsilon}_i(\hat{y}_i - \bar{y}) + (\hat{\varepsilon}_i)^2$$

$$\Sigma(y_i - \bar{y})^2 = \Sigma(\hat{y}_i - \bar{y})^2 + \Sigma(\hat{\varepsilon}_i)^2$$

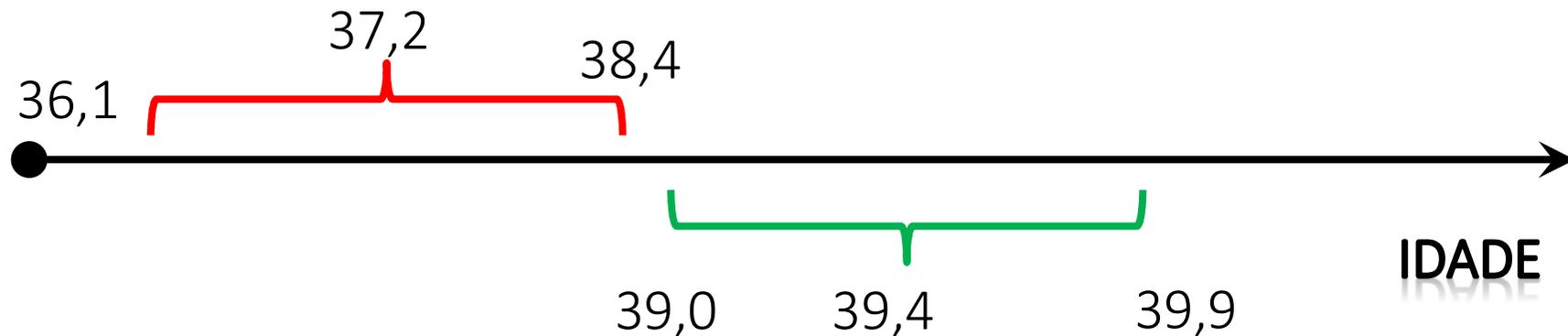
$$R^2 = \frac{\Sigma(\hat{y}_i - \bar{y})^2}{\Sigma(y_i - \bar{y})^2} = 1 - \frac{\Sigma(\hat{\varepsilon}_i - 0)^2/n}{\Sigma(y_i - \bar{y})^2/n}$$

$$R^2 = 1 - \frac{\text{Variância do resíduo}}{\text{Variância da nota}}$$

Quanto mais próximo de 1 o R-Quadrado estiver, mais bem ajustada aos dados estará a reta

REGRESSÃO E TESTE DE DIFERENÇA DE MÉDIAS

- Relembre do teste de diferença da média da idade entre participantes com e sem cônjuge no “Mulheres do Futuro” – arquivo “ex1”
- Podemos executar este teste usando uma regressão, ao invés de construir intervalos de confiança, fazendo $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ com y_i sendo a idade e x_i uma variável binária indicando “cônjuge”
- Os intervalos que calculamos são replicados abaixo (usando o período inicial)



$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ com y_i sendo a idade e x_i uma variável binária indicando “cônjuge”

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	ID	escolaridade	idade	luz	renda_pc	tratado	conjuge	tempo	tra_tem				
2984	787	2	58	1	400	1	1	0	0				
2985	795	3	58	1	403	1	1	0	0				
2986	805	2	58	1	475	1							
2987	801	2	58	1	500	1							
2988	796	4	58	1	500	1							
2989	816	2	58	1	503	1							
2990	792	1	58	1	507	1							
2991	797	1	58	1	510	1							
2992	817	2	58	1	510	1							
2993	815	3	58	1	510	1							
2994	818	2	58	1	550	1							
2995	802	2	58	1	570	1							
2996	779	2	58	1	590	1							
2997	807	2	58	1	633	1							
2998	808	2	58	1	765	1							
2999	775	7	58	1	925	1	1	0	0				
3000	783	2	58	1	1000	1	1	0	0				
3001	778	3	58	1	1005	1	1	0	0				

Regressão

Entrada

Intervalo Y de entrada: \$C\$1:\$C\$3001

Intervalo X de entrada: \$G\$1:\$G\$3001

Rótulos Constante é zero

Nível de confiança: 95 %

Opções de saída

Intervalo de saída:

Nova planilha: reg3

Nova pasta de trabalho

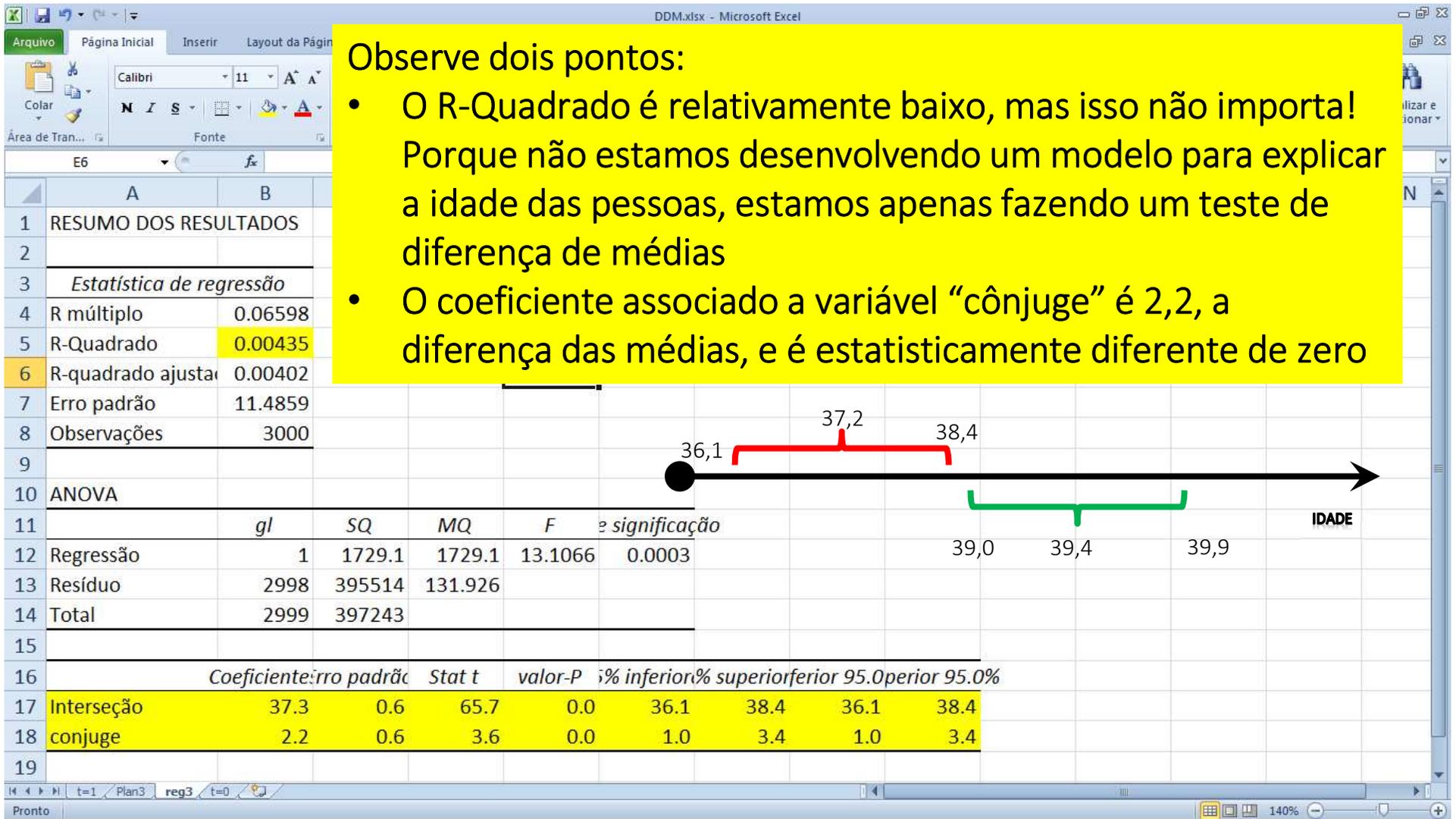
Resíduos

Resíduos Plotar resíduos

Resíduos padronizados Plotar ajuste de linha

Probabilidade normal

Plotagem de probabilidade normal



Observe dois pontos:

- O R-Quadrado é relativamente baixo, mas isso não importa! Porque não estamos desenvolvendo um modelo para explicar a idade das pessoas, estamos apenas fazendo um teste de diferença de médias
- O coeficiente associado a variável “cônjuge” é 2,2, a diferença das médias, e é estatisticamente diferente de zero

REGRESSÃO MÚLTIPLA

- A técnica da análise de regressão não se restringe a uma única variável explicativa. Com pequenos ajustes algébricos ela pode ser adaptada para muitas variáveis explicativas

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \cdots + \beta_K x_{Ki} + \varepsilon_i$$

- Na regressão inicial, relacionando nota e escolaridade da mãe, vamos acrescentar as demais covariadas disponíveis no banco de dados

Agora x_i será um vetor contendo as variáveis “tratado”, “mulher”, “cor” e “estudo_mae”

The screenshot shows the Microsoft Excel interface with a spreadsheet containing data for a regression analysis. The spreadsheet has the following data:

ID	nota	tratado	mulher	cor	estudo_mae
1	86.900002	1	0	0	7
2	86.900002	1	0	1	6
3	86.800003	1	0	0	6
4	86.699997	1	0	1	6
5	86.699997	1	0	1	6
6	86.699997	1	1	1	6
7	86.599998	1	1	1	6
8	86.599998	1	1	1	6
9	86.5	1	1	0	5
10	86.5	1	0	1	5
11	86.400002	1	0	0	5
12	86.400002	1	0	1	5
13	86.400002	1	1	1	5
14	86.400002	1	1	0	5
15	86.300003	1	0	1	5
16	86.099998	1	0	1	5
17	86	1	0	0	5
18	85.900002	1	0	0	4

The 'Regressão' dialog box is open, showing the following settings:

- Entrada: Intervalo Y de entrada: $\$B\$1:\$B\746 ; Intervalo X de entrada: $\$C\$1:\$F\746
- Rótulos
- Constante é zero
- Nível de confiança: 95 %
- Opções de saída: Intervalo de saída; Nova planilha: reg2; Nova pasta de trabalho
- Resíduos: Resíduos; Resíduos padronizados; Plotar resíduos; Plotar ajuste de linha
- Probabilidade normal: Plotagem de probabilidade normal

Note que os coeficientes estimados para “mulher” e “cor” não são estatisticamente diferentes de zero!

Por outro lado, o fato de ter participado do programa faz com que a nota média se eleve em aproximadamente 1 ponto (e é estatisticamente significativa).

O programa causou impacto na nota? Dada a estratégia de identificação disponível, temos indícios que o ATT do programa é de 1,076

Estatística de regressão	
R múltiplo	0.83264
R-Quadrado	0.69328
R-quadrado ajustado	0.69162
Erro padrão	2.10755
Observações	745

ANOVA					
	gl	SQ	MQ	F	e significação
Regressão	4	7429.47	1857.37	418.158	3E-188
Resíduo	740	3286.92	4.44179		
Total	744	10716.4			

	Coefficiente	erro padrão	Stat t	valor-P	% inferior	% superior	inferior 95.0%	superior 95.0%
Interseção	76.322	0.246	310.836	0.000	75.840	76.804	75.840	76.804
tratado	1.076	0.171	6.306	0.000	0.741	1.411	0.741	1.411
mulher	-0.176	0.156	-1.132	0.258	-0.482	0.129	-0.482	0.129
cor	0.202	0.155	1.299	0.194	-0.103	0.507	-0.103	0.507
estudo_mae	1.452	0.036	40.288	0.000	1.381	1.523	1.381	1.523

REGRESSÃO MÚLTIPLA E MUDANÇA DE FORMA FUNCIONAL

- Estime o modelo de regressão múltipla feito anteriormente mudando a forma funcional para?

$$Nota_i = \beta_0 + \beta_1 \ln(esc_mae_i) + \beta_2 tratado_i + \varepsilon_i$$

- Este modelo se ajusta melhor aos dados?
- A variável tratado é uma “variável *dummy*”, qual seu significado nesta regressão?

PSM1.xml - M

Arquivo Página Inicial Inserir Layout da Página Fórmulas Dados Revisão Exibição Acrobat

Calibri 11 A A

Quebrar Texto Automaticamente

Fonte Alinhamento

L11 fx

	A	B	C	D	E	F	G
1	RESUMO DOS RESULTADOS						
2							
3	<i>Estatística de regressão</i>						
4	R múltiplo	0.88743					
5	R-Quadrado	0.78752					
6	R-quadrado ajustado	0.78695					
7	Erro padrão	1.75177					
8	Observações	745					
9							
10	ANOVA						
11		<i>gl</i>	<i>SQ</i>	<i>MQ</i>	<i>F</i>	<i>e significação</i>	
12	Regressão	2	8439.42	4219.71	1375.08	3E-250	
13	Resíduo	742	2276.98	3.0687			
14	Total	744	10716.4				
15							
16		<i>Coefficiente</i>	<i>erro padrão</i>	<i>Stat t</i>	<i>valor-P</i>	<i>% inferior</i>	<i>% superior</i>
17	Interseção	74.14	0.20	367.03	0.00	73.74	74.54
18	tratado	0.89	0.14	6.32	0.00	0.61	1.16
19	ln(estudo_mae)	6.41	0.12	52.14	0.00	6.17	6.65
20							
21							

reg4 Plan3 Sheet1

Pronto 130%

Pelo critério do R-Quadrado, este modelo se ajusta melhor aos dados.

E quanto a “variável *dummy*”, ela é um deslocador do intercepto (intersecção).

O valor positivo e estatisticamente significativo está apontando uma correlação positiva entre a participação no programa e o aumento das notas.

O ATT é estimado agora em 0,89.

SUMÁRIO DO CURSO

- ~~1. Conceitos básicos em avaliação de impacto~~
- ~~2. Revisão de estatística e uso avançado de planilhas eletrônicas~~
- ~~3. Modelo de Resultados Potenciais~~
- ~~4. Econometria: revisão do modelo de mínimos quadrados~~
5. Desenho de uma avaliação para uma política
6. Modelo de Diferenças em Diferenças
7. Técnicas de Pareamento
8. Estudo de caso e aplicações de diferentes técnicas econométricas
9. Discussão sobre os trabalhos de conclusão de curso
10. Revisão de matemática financeira
11. O Cálculo do Retorno Econômico
12. Conclusão do curso: apresentação dos trabalhos

PLANEJANDO A AVALIAÇÃO

Em linhas gerais, o planejamento da avaliação segue os seguintes passos:

1. Descrever o projeto social recuperando seu Marco Lógico e definindo precisamente seus objetivos e população-alvo
2. Pensar na estratégia de identificação dando foco na identificação de quais **indicadores** serão analisados e como será construído o **grupo de controle**
3. Identificar como serão coletados os dados (cadastro do projeto, dados primários e/ou dados secundários)
4. Descrever a técnica econométrica a ser utilizada (teste de diferença de médias, regressão, diferenças em diferenças e/ou pareamento)
5. Coletar os indicadores de custos para calcular o retorno econômico
6. Refletir sobre as limitações da avaliação realizada

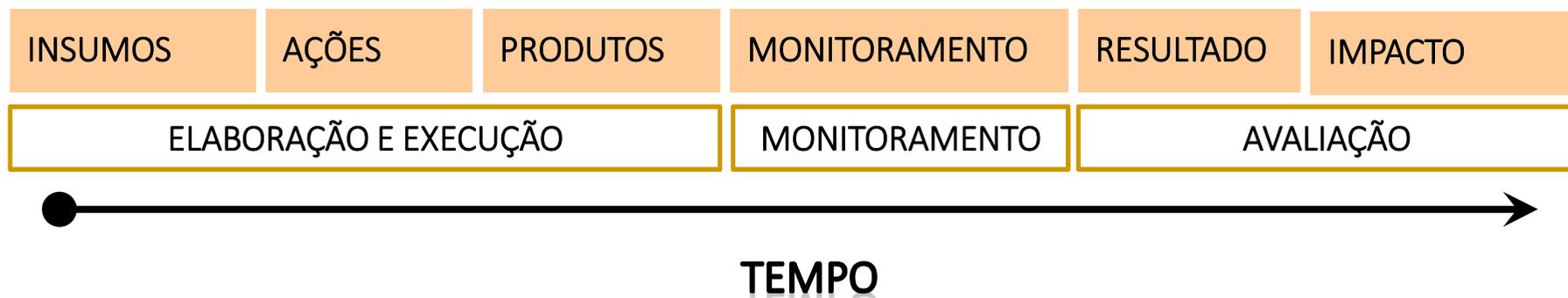
1. A descrição do projeto

A QUEM SE DESTINA O PROGRAMA?

- Parte da avaliação se dá na capacidade de contar a história do projeto. Se ele possui uma história consistente é mais provável que tenha resultados consistentes
- Para que possamos fazer uma boa avaliação econômica, é muito importante que definamos claramente o grupo a que ele se destina, isto é, o grupo de pessoas que o programa pretende beneficiar
- Depois de definidas quais são as características que delimitam o público-alvo, é importante conhecer melhor este grupo, isto é, conhecer as condições de vida e necessidades desse grupo
 - ✓ Por exemplo, se um programa tiver como público-alvo os jovens em idade escolar, é interessante saber onde moram estes jovens, saber se frequentam a escola, se trabalham, se moram com os pais e assim por diante

O MARCO LÓGICO

- Se o Marco Lógico foi documentado essa tarefa será fácil. Mas na maioria das vezes o avaliador terá de recuperar essa descrição do projeto
- Nesse sentido, as seguintes questões podem ser úteis: (1) Porque o projeto foi iniciado?; (2) Para qual população-alvo ele foi pensado?; (3) Como os beneficiários estão sendo atendidos (*modus operandi* do projeto)?; (4) Como foram definidas as ações?; (5) Quais os produtos gerados por estas ações?; (6) Existe um monitoramento?; (7) A população-alvo parece estar sendo adequadamente atingida de forma a observarmos algum resultado?



ESPECIFICIDADES SOBRE CRITÉRIOS DE ELEGIBILIDADE AO PROGRAMA SOCIAL AVALIADO E SUA FOCALIZAÇÃO

- Em geral, para a identificação do público-alvo utilizam-se os chamados critérios de elegibilidade
 - ✓ Por exemplo, o Programa Bolsa Família atende todas as famílias que tenham renda familiar per capita abaixo do nível de pobreza.
Definido o público-alvo, é possível posteriormente avaliar a focalização do programa
- Quando um programa beneficia pessoas que não deveriam ser contempladas e/ou quando deixa de fora pessoas que pertencem ao público-alvo, dizemos que o programa incorreu nos chamados erros de focalização (este é um importante componente da avaliação)

AS RELAÇÕES CAUSAIS SÃO CONSISTENTES COM OS OBJETIVOS?

- Para que possamos realizar a avaliação econômica, tão importante quanto a definição do público-alvo é a definição dos objetivos de impacto do projeto
- O objetivo deve dizer para que servirá o projeto, ou seja, o que se pretende alcançar ao final do programa
- Faça uma análise intuitiva (e racional) das pressupostas relações causais:
 - ✓ Entre produtos e resultados
 - ✓ Entre produtos e impactos
 - ✓ Entre resultados e impactos

2. A estratégia de identificação

A ESTRATÉGIA DE IDENTIFICAÇÃO

- Lembre-se que o modelo de resultados potenciais pode ser descrito pelas seguintes equações:

$$y_i = T_i y_i(1) + (1 - T_i) y_i(0)$$

$$y_i(0) + T_i (y_i(1) - y_i(0))$$

$$y_i(0) + \beta_i T_i$$

$$\alpha_0 + \alpha_1 x_{1i} + \dots + \alpha_K x_{Ki} + \varepsilon_i + \beta_i T_i$$

- Então deve-se cuidadosamente escolher quais indicadores y_i , x_{1i} , ..., x_{Ki} serão escolhidos de forma a:
 - ✓ Serem coerentes com o Marco Lógico do programa
 - ✓ Minimizarem a magnitude das características importantes para explicar y_i , mas que não são observadas (ε_i)

INDICADORES

- Os objetivos do projeto devem ser “mensuráveis”. Para que isto seja feito precisamos definir bem seus indicadores de resultado e impacto (y_i), bem como suas covariadas (x_{1i}, \dots, x_{Ki})
- É preciso que fique claro que y_i está intimamente relacionado ao objetivo proposto e, portanto, é determinado pelas especificidades de cada projeto
- Vejamos os exemplos a seguir:
 - ✓ Exemplo 1: Para o programa que tem por objetivo melhorar a qualidade do ensino entre os jovens de ensino médio, a taxa de frequência escolar desses jovens pode ser um indicador importante. Um outro indicador possível é a média das notas que esses adolescentes obtiveram na prova de matemática (podem existir mais de um y_i)
 - ✓ Exemplo 2: Para o programa de combate à desnutrição infantil, a média de peso das crianças de 0 a 3 anos pode ser o indicador escolhido

QUALIDADES DOS INDICADORES

- Dada a importância dos indicadores para o conhecimento da situação antes e depois do projeto, é importante que estes apresentem algumas características
 - ✓ Entendimento simples
 - ✓ Facilidade de construção
 - ✓ Confiabilidade
 - ✓ Consistência com os dados disponíveis
 - ✓ Relação direta com as ações efetuadas
- Para alguns tipos de projetos, pode ser interessante incorporar indicadores qualitativos em sua análise
- Respostas qualitativas são subjetivas, expressando referências e valores. Devido à sua própria natureza, pode existir maior dificuldade para classificar e tabular as respostas qualitativas

CONSTRUÇÃO DO GRUPO DE CONTROLE

- Em toda a avaliação o desafio central do avaliador é encontrar um bom grupo de controle, ou seja, um grupo que funcione como um bom contrafactual do grupo tratado
- Muito especificamente, **o avaliador precisa de uma coluna no Excel onde exista o preenchimento de uma variável T_i com zeros e uns para as n observações disponíveis**
- Com uma frequência muito maior que se desejaria, será simplesmente inviável avaliar uma série de projetos sociais por não se conseguir encontrar um grupo de controle

QUALIDADE DO GRUPO DE CONTROLE

- A qualidade do grupo de controle é quem determina a qualidade da avaliação
- Se o grupo de controle for de boa qualidade, a única diferença de resultados deve partir do fato de que alguns receberam o tratamento
- Resgatando o modelo de resultados potenciais, o componente $y_i(0)$ deve ser estatisticamente semelhante em média entre os grupos de tratamento e controle. Isto significa que as médias das covariadas x_{1i}, \dots, x_{Ki} devem ser semelhantes entre os grupos de tratamento e controle

3. A base de dados

FONTES DE INFORMAÇÃO

- Para que a avaliação econômica possa ser realizada é necessário que existam informações ou dados para isso. Esse trabalho será facilitado se os gestores do projeto tiverem um cadastro dos participantes do projeto
- O cadastramento dos participantes deve trazer todas as informações e características consideradas relevantes para a avaliação
- Pode existir uma tendência a não se dar a devida importância a essa fase. Ou, então, os gestores do programa podem não construir um cadastro adequado, por julgarem que demanda muito tempo e recursos
- Sem a identificação adequada, muito do trabalho posterior de avaliação pode ser impossibilitado, por não se ter empregado algum tempo e esforço no cadastramento.

QUE INFORMAÇÕES DEVEM CONSTAR NO CADASTRO?

- A resposta a essa pergunta depende em grande parte dos objetivos do programa. Um cadastro pode ser feito com o emprego de uma planilha eletrônica, que tem os recursos necessários ao armazenamento das informações relevantes
- Cadastros contém informações do tipo:
 - ✓ Nome
 - ✓ Idade
 - ✓ Gênero
 - ✓ Cor ou raça
 - ✓ Série em que estuda
 - ✓ Escolaridade da mãe
 - ✓ Renda familiar
 - ✓ Número de componentes da família
 - ✓ Indicação se trabalha

FONTES EXTERNAS DE DADOS

- Algumas vezes pode ser necessário recorrer ao que denominamos “fontes externas de dados”. Esses dados são encontrados em pesquisas e avaliações feitas usualmente por órgãos governamentais
- Por meio dessas fontes, é possível conhecer de forma detalhada algumas características relevantes, como por exemplo, educação, rendimentos, hábitos de consumo e outras características socioeconômicas, de alguma população que se tenha interesse

EXEMPLOS DE FONTES EXTERNAS DE DADOS

- A **PNAD** é uma extensiva pesquisa feita pelo IBGE em mais de 100 mil domicílios brasileiros, abrangendo dados de cerca de 350 mil indivíduos, com o objetivo de gerar informações sobre a situação socioeconômica do país
- A PNAD é uma pesquisa amostral representativa do conjunto da população do Brasil
- A pesquisa começou a ser feita na década de 70 e vai a campo anualmente, com exceção dos anos de censo demográfico (1980, 1991, 2000 e 2010), quando deixa de ser executada
- O conjunto de temas pesquisados é bastante amplo e inclui tópicos como educação, emprego, rendimentos, sindicalização, condições de habitação, posse de bens duráveis e condições de saneamento
- Com periodicidade variável, são feitos suplementos, com perguntas relativas a outros temas, como migração, fecundidade, nupcialidade, saúde e nutrição, de acordo com as necessidades de informação para o país

EXEMPLOS DE FONTES EXTERNAS DE DADOS

- O **Censo Escolar** é realizado conjuntamente pelo MEC e pelas secretarias estaduais de educação para avaliar as condições de todas as escolas de educação básica do país
- Sua finalidade é compor um quadro sobre as condições de oferta do ensino básico no Brasil
- É aplicado anualmente desde 1991
- O questionário é composto de nove blocos, bastante extensos
- Há questões sobre a existência de bibliotecas, salas de estudo, microcomputadores, acesso à internet, emprego de livros didáticos e instalações físicas
- Também se procura saber a escolaridade do diretor e do corpo docente da instituição de ensino, bem como o número de profissionais envolvidos no processo educativo
- Há perguntas sobre o número de matriculados, horários de estudo, idade dos alunos por série, etc.

E SE FOR PRECISO APLICAR QUESTIONÁRIO?

- O questionário é um instrumento de pesquisa constituído por uma série de questões sobre um determinado tema
- Em geral as respostas são transformadas em estatísticas
- O desafio é fazer com que as pessoas falem (a verdade)



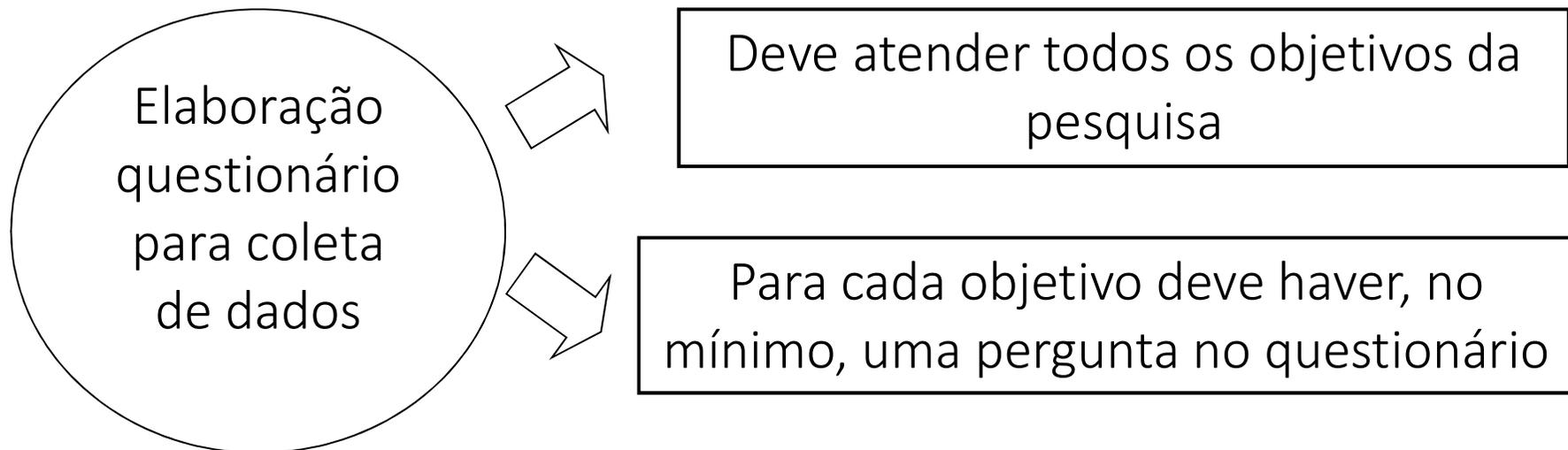
FRASEOLOGIA DAS PERGUNTAS

- Linguagem das perguntas pode influenciar fortemente a validade e a confiabilidade do questionário
- Usa palavras compreensíveis, claras, de acordo com o nível cultural dos entrevistados
- Não faça perguntas que requeiram respostas generalizadas. A pergunta deve ser o mais específica possível

FORMAS DE APLICAÇÃO

- Por correspondência, e-mail, formulário eletrônico ou telefone a vantagem é a facilidade de aplicação. A desvantagem é que talvez poucas respostas sejam retornadas ou os números de telefone tenham mudado
- Por trabalho de campo – face a face. É mais trabalhosa, mas mais confiável

ELABORAÇÃO DE QUESTIONÁRIOS



Portanto, é necessário determinar inicialmente, e com bastante clareza, os objetivos da pesquisa e, posteriormente elaborar o questionário para a realização das entrevistas

4. A técnica econométrica e
ser utilizada

A TÉCNICA UTILIZADA PARA SE ESTIMAR O IMPACTO

- Neste curso vimos ou ainda veremos quatro técnicas
 1. Aleatorização: a se usar testes de diferenças de média quando os grupos de tratamento e controle foram definidos por **sorteio**
 2. Regressão: quando não houve aleatorização e dispomos de **poucas informações** dos grupos de tratamento e controle **depois** de início do projeto
 3. Diferenças em diferenças: quando não houve aleatorização e dispomos de informações dos grupos de tratamento e controle **antes e depois** de início do projeto
 4. Pareamento: quando não houve aleatorização e dispomos de **muitas informações** dos grupos de tratamento e controle antes de início do projeto ou apenas depois do projeto

5. Informações para o cálculo do retorno econômico

CUSTOS

- Um último elemento essencial para a avaliação econômica, especificamente para o cálculo do retorno econômico do projeto, é uma medida dos custos incorridos na realização do projeto
- É importante ter em mente que todos os custos precisam ser incluídos
- Custos econômicos = custos contábeis + o custo de oportunidade daqueles que participam do programa
 - ✓ Custo de oportunidade é o rendimento que se deixa de obter quando se realiza uma determinada escolha
 - ✓ Dito de outra forma, o custo de oportunidade de um item é aquilo que você abre mão para poder obter tal item

EXEMPLO DE CUSTO DE OPORTUNIDADE

- Suponha um programa de melhoria da qualidade do ensino. Suponha que seu foco esteja no combate à evasão escolar
- Neste caso, há um custo para esse jovem ou para sua família, quando ele passa a freqüentar a escola (mesmo que isso seja desejado pelos gestores do programa)
- Ao ir à escola, o jovem pode ter que deixar de trabalhar ou então passar a trabalhar apenas em meio-período
- Isto é, há uma queda na renda que ele gerava para sua família
- Mesmo quando o jovem não trabalha ou quando seu trabalho não é remunerado, ir à escola pode significar ter que encontrar alguém para fazer os trabalhos domésticos ou cuidar dos irmãos mais novos
- O salário perdido ou o custo da babá são os custos de oportunidade dessas famílias

OUTRO EXEMPLO DE CUSTO DE OPORTUNIDADE

- Há diversos projetos que são desenvolvidos dentro de uma instituição que pertence aos próprios gestores do programa
- Portanto, não há custos associados ao aluguel do imóvel
- No entanto, podemos pensar que essa casa ou escola ou quadra esportiva onde o projeto é desenvolvido poderia ser alugada, gerando uma renda para os gestores.
- Renda de aluguel perdida = custo de oportunidade importante associado ao projeto, que deve ser considerado

TAXA DE JUROS DE MERCADO = CUSTO DE OPORTUNIDADE DO PROJETO SOCIAL

- Quando investimos recursos em um projeto social, abrimos mão de aplicar esses recursos em outro tipo de projeto social ou mesmo numa aplicação financeira
- É por essa razão que mais à frente, iremos fazer referência à taxa de juros de mercado como sendo o custo de oportunidade do projeto social

6. As limitações

REFLEXÃO SOBRE AS LIMITAÇÕES DA AVALIAÇÃO

- Um programa tem geralmente uma variedade de impactos, alguns intencionais e muitos outros colaterais. Então, uma clara limitação das avaliações é que elas sempre são uma análise de **equilíbrio parcial**
- Mesmo que fosse possível estimar a magnitude de todos os impactos, ainda seria preciso estimar o valor que cada beneficiário atribui a cada um deles. O que é intangível

SUMÁRIO DO CURSO

- ~~1. Conceitos básicos em avaliação de impacto~~
- ~~2. Revisão de estatística e uso avançado de planilhas eletrônicas~~
- ~~3. Modelo de Resultados Potenciais~~
- ~~4. Econometria: revisão do modelo de mínimos quadrados~~
- ~~5. Desenho de uma avaliação para uma política~~
6. Modelo de Diferenças em Diferenças
7. Técnicas de Pareamento
8. Estudo de caso e aplicações de diferentes técnicas econométricas
9. Discussão sobre os trabalhos de conclusão de curso
10. Revisão de matemática financeira
11. O Cálculo do Retorno Econômico
12. Conclusão do curso: apresentação dos trabalhos

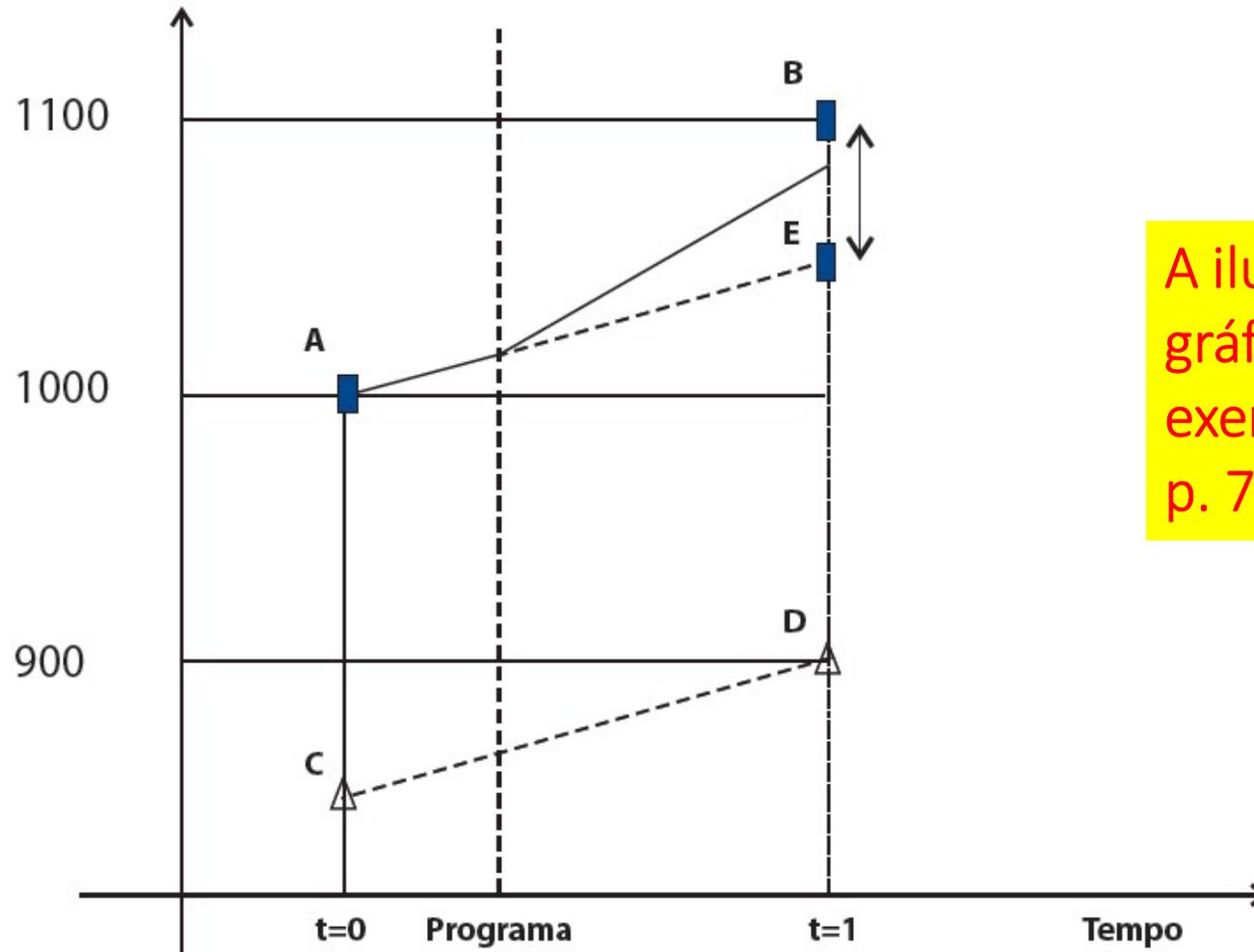
MÉTODOS DE AVALIAÇÃO

- Os métodos de avaliação de impacto são geralmente divididos em dois grupos: **experimental** e **não-experimental**
 - ✓ O primeiro é baseado na seleção aleatória dos participantes e não participantes do programa e é considerado o método de referência na área de avaliação
 - ✓ O segundo grupo é composto por uma variedade de métodos distintos que, embora façam uso de hipóteses específicas para identificar o efeito causal do programa, procuram mimetizar o método experimental
- A técnica das diferenças em diferenças (DD) é um método não-experimental que requer dados amostrais para os tratados e não tratados em pelo menos dois períodos no tempo: antes e depois da intervenção

HIPÓTESE

- A principal hipótese do método de DD é que a trajetória temporal da variável de resultado para o grupo de controle representa o que ocorreria com o grupo tratado caso não houvesse a intervenção
- Naturalmente, essa hipótese não pode ser testada diretamente nos dados, mas uma indicação de sua validade aparece quando as trajetórias dos dois grupos são parecidas pré-programa
- A ideia básica é que, se as trajetórias se assemelham durante o período antes do programa, então parece razoável supor que a evolução do grupo de controle após o programa represente com fidedignidade o que ocorreria com o grupo de tratados

Rendimento Médio (R\$)



A ilustração gráfica do exemplo da p. 77

UNIDADES DE OBSERVAÇÃO

- O procedimento de DD pode ser utilizado com dados de indivíduos/famílias, mas também com dados em níveis mais agregados, por exemplo, setores de atividade, ocupações, municípios ou estados
- Sua aplicabilidade com dados agregados é uma importante vantagem do método de DD, pois permite que o impacto de uma série de intervenções possam ser avaliadas apenas com base em informações rotineiramente coletadas por governos e institutos de pesquisa e estatística

OUTRA VANTAGEM PARA O CASO DE DADOS EM PAINEL

- Outra importante vantagem do método de DD é que ele permite controlar para características não observáveis dos indivíduos que sejam invariantes no tempo (por exemplo, habilidades inatas)
- Embora o método de DD só consiga levar em conta esse tipo de atributo que é fixo no tempo, esse controle é capaz de resolver o problema de auto seleção para um conjunto amplo de aplicações
- Essa capacidade confere ao DD uma vantagem relativa a outros métodos não-experimentais que não conseguem controlar para vieses de auto seleção decorrentes de atributos não observáveis, fixos ou não no tempo

O MODELO DE DIFERENÇAS EM DIFERENÇAS

- A forma mais simples de expressar o estimador do método de DD é calculando uma dupla diferença de médias da variável de resultado. Se denotarmos por $T = \{1,0\}$ a participação ou não no programa e por $t = \{1,0\}$ os períodos posterior e anterior à intervenção, respectivamente, o estimador de DD será dado por:

$$\beta_{DD} = \{E[Y|T = 1, t = 1] - E[Y|T = 1, t = 0]\} \\ - \{E[Y|T = 0, t = 1] - E[Y|T = 0, t = 0]\}$$

- Embutida nesse estimador está a hipótese de que a variação temporal na variável de resultado para o grupo de controle representa a variação contrafactual do grupo tratado. Em decorrência dessa hipótese, a diferença entre a variação efetivamente observada para o grupo de tratamento e a variação contrafactual fornecida pelo grupo de controle vai capturar o efeito causal da intervenção

UMA IMPORTANTE OBSERVAÇÃO

- A média da variável de resultado para o grupo de controle no período anterior ao programa não precisa coincidir com a média correspondente para o grupo de tratamento, ou seja, os grupos podem ou não partir de um mesmo ponto
- Na realidade, na maior parte das aplicações do procedimento de DD, os dois grupos têm essas médias distintas, um fato que tipicamente reflete as diferentes influências dos atributos observáveis e não observáveis dos indivíduos sobre a variável de resultado
- O que o método de fato requer é que a variação temporal do que ocorre com o grupo de controle antes e depois do programa reflita corretamente a variação temporal do grupo de tratados na situação contrafactual de não tratamento

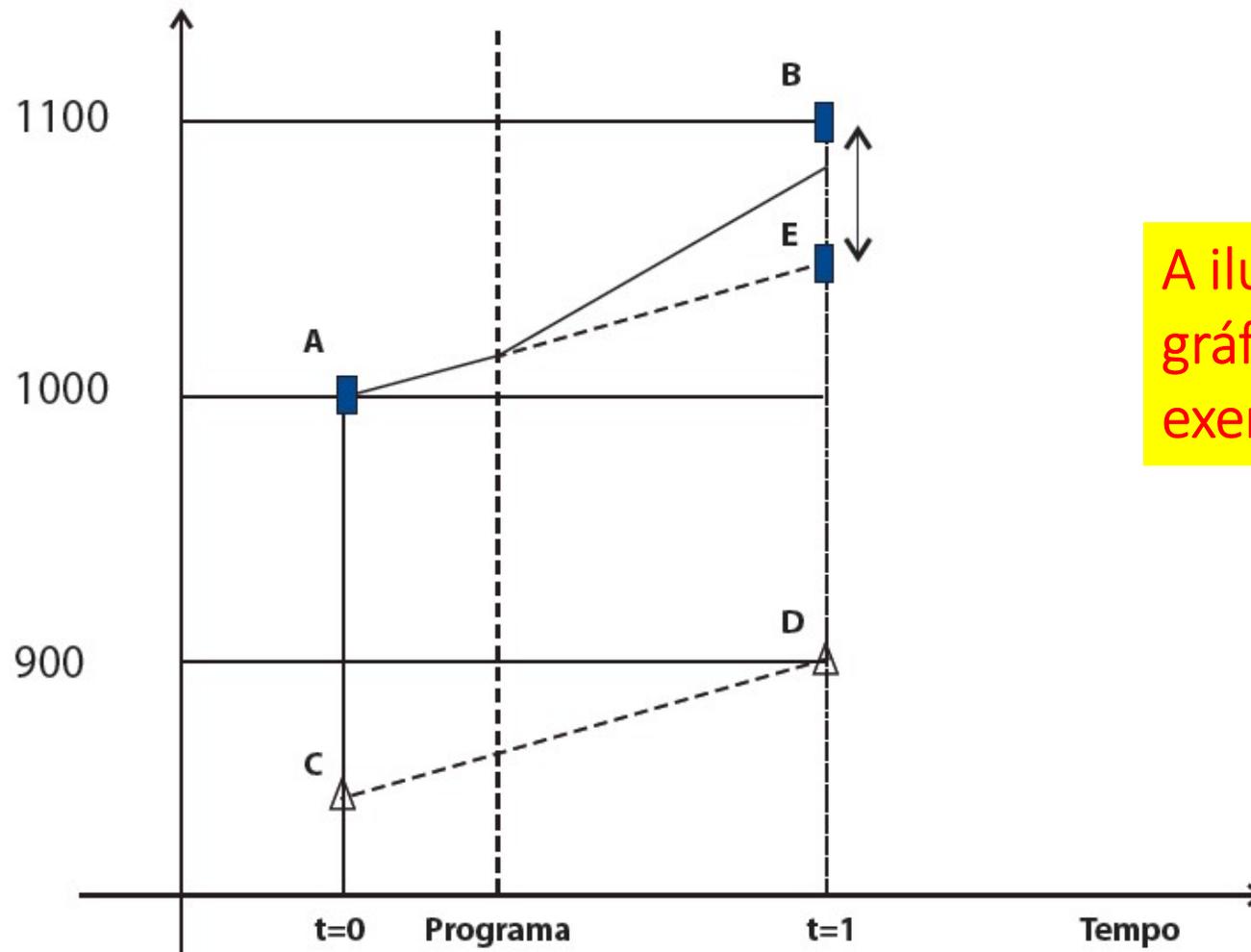
EXEMPLO DO LIVRO

- Considere uma situação em que um programa de treinamento profissional foi oferecido para trabalhadores de baixa escolaridade numa certa localidade
- Esse programa consistia de um curso ministrado por pessoal especializado e tinha como objetivo aumentar o rendimento do trabalho dos participantes
- O interesse recaí sobre o impacto do programa sobre o salário médio dos tratados, que era de R\$ 1.000 um pouco antes do início do curso e de R\$ 1.100 alguns meses após o programa
- A tabela a seguir contém os valores dos salários médios dos grupos de tratamento e controle – respectivamente, (T_1, T_0) – para os períodos anterior e posterior ao programa – (t_0, t_1) , respectivamente

	Tratados	Controle	Diferença
Antes	$\bar{Y}_{T_1,t_0} = 1000$	$\bar{Y}_{T_0,t_0} = 850$	$\Delta\bar{Y}_{T_1-T_0,t_0} = 150$
Depois	$\bar{Y}_{T_1,t_1} = 1100$	$\bar{Y}_{T_0,t_1} = 900$	$\Delta\bar{Y}_{T_1-T_0,t_1} = 200$
Diferença	$\Delta\bar{Y}_{T_1,t_1-t_0} = 100$	$\Delta\bar{Y}_{T_0,t_1-t_0} = 50$	$\begin{aligned} \Delta\bar{Y}_{T_1,t_1-t_0} - \Delta\bar{Y}_{T_0,t_1-t_0} \\ &= \Delta\bar{Y}_{T_1-T_0,t_1} \\ &\quad - \Delta\bar{Y}_{T_1-T_0,t_0} \\ &= 50 \end{aligned}$

- A notação \bar{Y} representa a média amostral do salário e Δ significa o operador de diferença

Rendimento Médio (R\$)



A ilustração gráfica do exemplo

UTILIZANDO A REGRESSÃO LINEAR

- ✓ Renda = $\alpha + \beta_1$ 'programa' + β_2 'depois' + β_3 'programa*depois' + erro
- ✓ β_3 = a diferença das diferenças, ou seja, é o coeficiente que mede o impacto do programa.
- ✓ β_1 captura se os grupos são diferentes, independentemente do programa.
- ✓ β_2 captura se o indicador muda no tempo, independentemente do programa.

PORQUE β_3 É O ESTIMADOR DE DIFERENÇAS EM DIFERENÇAS?

Médias por grupo	Antes	Depois	Variação
Tratamento	TA	TD	TD - TA
Controle	CA	CD	CD - CA
Variação das variações	$(TD - TA) - (CD - CA)$		

$$\text{Renda}^{\text{est}} = \alpha + \beta_1 \times \text{'programa'} + \beta_2 \times \text{'depois'} + \beta_3 \times \text{'programa*depois'}$$

✓ $\text{TA} = \alpha + \beta_1 \times \text{'1'} + \beta_2 \times \text{'0'} + \beta_3 \times 0$

✓ $\text{TA} = \alpha + \beta_1$

✓ $\text{TD} = \alpha + \beta_1 \times \text{'1'} + \beta_2 \times \text{'1'} + \beta_3 \times 1$

✓ $\text{TD} = \alpha + \beta_1 + \beta_2 + \beta_3$

✓ $\text{TD} - \text{TA} = \beta_2 + \beta_3$

$$\text{Renda}^{\text{est}} = \alpha + \beta_1 \times \text{'programa'} + \beta_2 \times \text{'depois'} + \beta_3 \times \text{'programa*depois'}$$

✓ $CA = \alpha + \beta_1 \times '0' + \beta_2 \times '0' + \beta_3 \times 0$

✓ $CA = \alpha$

✓ $CD = \alpha + \beta_1 \times '0' + \beta_2 \times '1' + \beta_3 \times 0$

✓ $CD = \alpha + \beta_2$

✓ $CD - CA = \beta_2$

$$\text{Renda}^{\text{est}} = \alpha + \beta_1 \times \text{'programa'} + \beta_2 \times \text{'depois'} + \beta_3 \times \text{'programa*depois'}$$

✓ $(\text{TD} - \text{TA}) = \beta_2 + \beta_3$

✓ $(\text{CD} - \text{CA}) = \beta_2$

✓ $(\text{TD} - \text{TA}) - (\text{CD} - \text{CA}) = \text{dif em dif} = (\beta_2 + \beta_3 - \beta_2) = \beta_3$

EXERCÍCIO USANDO A BASE “EX1”

O programa “Mulher do Futuro” – fictício – objetiva aumentar a renda de famílias chefiadas por mulheres. Para isso, oferece cursos profissionalizantes em diversas áreas e ajuda na inserção das participantes no mercado de trabalho, por meio de convênios com empresas. Os cursos duram em média seis meses. A avaliação de impacto do programa foi pensada desde sua concepção. Assim, realizou-se uma pesquisa com as participantes e não participantes antes da implementação. Dois anos após o início do programa, o grupo de avaliação voltou a campo e novamente coletou informações sobre as mulheres. O banco de dados contém as informações em dois momentos no tempo – escolaridade, idade, se o domicílio possui luz elétrica, se a mulher vive com cônjuge e a renda familiar per capita.

ddm.xlsx - Microsoft Excel

Arquivo Página Inicial Inserir Layout da Página Fórmulas Dados Revisão Exibição Acrobat

Área de Tran... Fonte Alinhamento Número

Formatação Condicional Formatar como Tabela Estilos de Célula Inserir Excluir Formatar Células

AutoSoma Preencher Limpar Classificar e Filtrar Localizar e Selecionar Edição

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	ID	escolaridad	idade	luz	renda_pc	tratado	conjuge	tempo	tra_tem					
2	567	2	14	0	85	0	1	0	0					
3	585	2	14	1	128	0	1	0	0					
4	571	1	14	1	255	0	0	0	0					
5	628	2	14	1	285	0	1	0	0					
6	626	1	14	1	316	0	1	0	0					
7	630	1	14	1	340	0	1	0	0					
8	564	2	14	1	340	0	0	0	0					
9	649	3	14	1	340	0	1	0	0					
10	652	1	14	1	383	0	1	0	0					
11	620	3	14	1	402	0	1	0	0					
12	622	3	14	1	418	0	1	0	0					
13	623	2	14	1	470	0	1	0	0					
14	644	1	14	1	510	0	1	0	0					
15	629	3	14	1	510	0	1	0	0					
16	635	3	14	1	510	0	1	0	0					
17	640	3	14	1	608	0	1	0	0					
18	643	7	14	1	750	0	1	0	0					
19	621	1	14	1	755	0	1	0	0					

Sheet1

Pronto

140%

Abra a base, e observe que a última coluna é o produto entre as *dummies* de tempo e tratamento

DIGRESSÃO: A INTERPRETAÇÃO DE UM PARÂMETRO LINEAR DE UMA VARIÁVEL BINÁRIA QUANDO A VARIÁVEL DEPENDENTE É UM LOGARITMO NATURAL

- Se o pesquisador estiver trabalhando com uma forma funcional $\ln y = \alpha + \beta D$, onde D é uma *dummy*, note que ao se derivar implicitamente y em relação a D ocorre que:

$$\beta \cong \frac{1}{y} \frac{y' - y}{D' - D} = \frac{y' - y}{y}$$

- Ou seja, a leitura do parâmetro β é direta e feita da seguinte forma: se β estimado é de, por exemplo, 0,05, então se conclui que no caso $D = 1$ ocorre um aumento de 5% em y
- Este é o conceito de “semi-elasticidade”

Vamos trabalhar com o indicador de resultado “logaritmo natural da renda per capita” para que possamos fazer uma leitura direta dos resultados de nossas regressões

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	ID	escolaridade	idade	luz	renda_pc	tratado	conjugue	tempo	tra_tem	ln_renda_pc				
2	567	2	14	0	85	0	1	0	0	=ln(E2)				
3	585	2	14	1	128	0	1	0	0					
4	571	1	14	1	255	0	0	0	0					
5	628	2	14	1	285	0	1	0	0					
6	626	1	14	1	316	0	1	0	0					
7	630	1	14	1	340	0	1	0	0					
8	564	2	14	1	340	0	0	0	0					
9	649	3	14	1	340	0	1	0	0					
10	652	1	14	1	383	0	1	0	0					
11	620	3	14	1	402	0	1	0	0					
12	622	3	14	1	418	0	1	0	0					
13	623	2	14	1	470	0	1	0	0					
14	644	1	14	1	510	0	1	0	0					
15	629	3	14	1	510	0	1	0	0					
16	635	3	14	1	510	0	1	0	0					
17	640	3	14	1	608	0	1	0	0					
18	643	7	14	1	750	0	1	0	0					
19	621	1	14	1	755	0	1	0	0					

ddm.xlsx - Microsoft Excel

Arquivo Página Inicial Inserir Layout da Página Fórmulas Dados Revisão Exibição Acrobat

Calibri

Área de Tran... Fonte

K5

Lembre que não existe logaritmo de zero, então não poderemos usar todas as informações da base de dados

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	ID	escolaridade	idade	luz	renda_pc	tratado	conjuge	tempo	tra_tem	ln_renda_pc				
50	691	2	16	1	462	0	1	0	0	6.14				
51	695	2	16	1	503	0	1	0	0	6.22				
52	2989	8	16	1	603	0	1	0	0	6.40				
53	2990	5	16	1	0	1	0	0	0	#NÚM!				
54	2991	7	16	1	148	1	0	0	0	5.00				
55	723	2	16	1	255	1	1	0	0	5.54				
56	715	3	16	1	318	1	1	0	0	5.76				
57	720	1	16	1	340	1	1	0	0	5.83				
58	731	5	16	1	349	1	1	0	0	5.86				
59	727	2	16	1	368	1	1	0	0	5.91				
60	712	4	16	1	403	1	1	0	0	6.00				
61	735	3	16	1	455	1	1	0	0	6.12				
62	2985	3	17	0	78	0	1	0	0	4.36				
63	2984	7	17	1	107	0	0	0	0	4.67				
64	2983	4	17	1	153	0	1	0	0	5.03				
65	2986	7	17	1	360	0	0	0	0	5.89				
66	2987	2	17	1	338	1	1	0	0	5.82				
67	2988	7	17	1	900	1	1	0	0	6.80				

Sheet1

Pronto

140%

ddm.xlsx - Microsoft Excel

Arquivo Página Inicial Inserir Layout da Página Fórmulas Dados Revisão Exibição Acrobat

Calibri

Colar

Área de Tran...

Fonte

AutoSoma

Insérer

Compar

Classificar e Filtrar

Localizar e Selecionar

Edição

Delete as observações cuja renda per capita é zero

A221 1798

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	ID	escolaridade	idade	luz	renda_pc	tratado	conjuge	tempo	tra_tem	ln_renda_pc				
212	2085	3	43	1	0	0	1	1	0	#NÚM!				
213	1895	3	45	1	0	0	1	1	0	#NÚM!				
214	1905	3	45	1	0	0	1	1	0	#NÚM!				
215		46		1	0	0	1	1	0	#NÚM!				
216		47		1	0	0	1	1	0	#NÚM!				
217		1	37	1	0	1	1	1	1	#NÚM!				
218		1	43	1	0	1	1	1	1	#NÚM!				
219		1	44	1	0	1	1	1	1	#NÚM!				
220		1	45	1	0	1	1	1	1	#NÚM!				
221		1	47	0	0	1	1	1	1	#NÚM!				
222		2	54	1	4	1	1	0	0	1.39				
223		3	45	1	11	0	1	0	0	2.40				
224		2	42	1	11	1	1	0	0	2.40				
225		2	52	1	11	1	1	0	0	2.40				
226		1	45	1	14	1	1	0	0	2.64				
227	2921	4	23	1	15	0	1	0	0	2.71				
228	2031	3	42	1	16	1	1	0	0	2.77				
229	2808	1	28	1	17	0	1	0	0	2.83				

Recortar

Copiar

Opções de Colagem:

Colar Especial...

Inserir

Excluir

Limpar conteúdo

Formatar células...

Altura da Linha...

Ocultar

Re-exibir

Sheet1

Pronto

Contagem: 2200

140%

ddm.xlsx - Microsoft Excel

Arquivo | Página Inicial | Inserir | Layout da Página | Fórmulas | Dados | Revisão | Exibição | Acrobat

Colar | Arial | 10 | Quebrar Texto Automaticamente | Geral | Formatação Condicional | Formatar como Tabela | Estilos de Célula | Inserir | Excluir | Formatar | AutoSoma | Preencher | Limpar | Classificar e Filtrar | Localizar e Selecionar | Edição

Área de Tran... | Fonte | Alinhamento | Número

A2 | fx | 1780

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	ID	scolaridade	idade	luz	renda_pc	tratado	conjugue	tempo	tra_tem	ln_renda_pc				
2	1780	3	45	1	11	0	1	0	0	2.40				
3	2921	4	23	1	15	0	1	0	0	2.71				
4	2808	1	28	1	17	0	1	0	0	2.83				
5	2074									2.89				
6	2812									2.94				
7	2138									2.94				
8	2955									3.00				
9	1222									3.00				
10	1214									3.04				
11	2577									3.09				
12	1694									3.09				
13	2155									3.14				
14	1289									3.14				
15	2846	4	26	1	25	0	1	0	0	3.22				
16	1920	2	43	1	25	0	1	0	0	3.22				
17	770	3	58	1	25	0	1	0	0	3.22				
18	2671	3	31	1	26	0	1	0	0	3.26				
19	2408	2	36	1	26	0	1	0	0	3.26				

Classificar

Adicionar Nível | Excluir Nível | Copiar Nível | Opções... | Meus dados contêm cabeçalhos

Coluna	Classificar em	Ordem
Classificar por: tratado	Valores	Do Menor para o Maior
E depois por: tempo	Valores	Do Menor para o Maior
E depois por: ln_renda_pc	Valores	Do Menor para o Maior

OK | Cancelar

Ordene as observações por condição de tratamento e data da observação

Pronto | Média: 200.2671697 | Contagem: 57810 | Soma: 11575442.41 | 140%

ddm.xlsx - Microsoft Excel

Arquivo Página Inicial Inserir Layout da Página Fórmulas Dados Revisão Exibição Acrobat

Calibri

Colar

Área de Tran...

AutoSoma Preencher Limpar Classificar e Filtrar Localizar e Selecionar Edição

G2706

1 ID escolaridade idade luz renda_pc tratado conjugue tempo tra_tem ln_renda_pc

1	ID	escolaridade	idade	luz	renda_pc	tratado	conjugue	tempo	tra_tem	ln_renda_pc
2697	1069	4	56	1	2610	0	1	1	0	7.87
2698	2941	5	24	1	2610	0	1	1	0	7.87
2699	2222	5	41	1	2610	0	1	1	0	7.87
2700	1292	8	53	1	2860	0	1	1	0	7.96
2701	131	6	42	1	3540	0	1	1	0	8.17
2702	1897	7	45	1	3613.33	0	1	1	0	8.19
2703	1454	8	51	1	3843.33	0	1	1	0	8.25
2704	374	2	25	1	4110	0	1	1	0	8.32
2705	2338	3	39	1	4110	0	1	1	0	8.32
2706										
2707	1109	2	54	1	4	1	1	0	0	1.39
2708	2040	2	42	1	11	1	1	0	0	2.40
2709	1268	2	52	1	11	1	1	0	0	2.40
2710	1818	1	45	1	14	1	1	0	0	2.64
2711	2031	3	42	1	16	1	1	0	0	2.77
2712	2116	2	41	1	17	1	1	0	0	2.83
2713	2057	2	42	1	17	1	1	0	0	2.83
2714	2232	1	39	1	19	1	1	0	0	2.94

Sheet1

Pronto

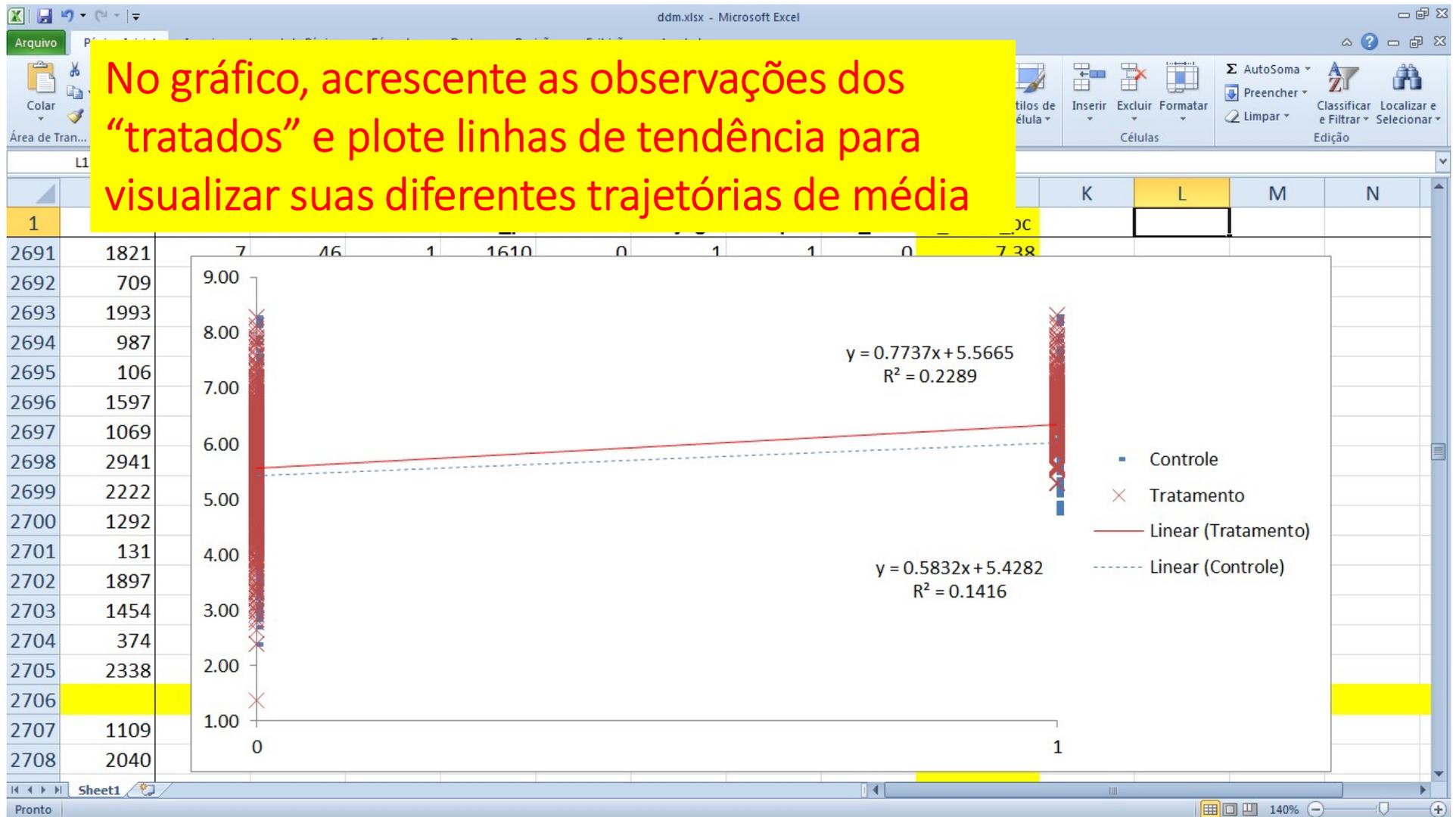
140%

Separe a base de dados com uma linha extra, dividindo as observações entre tratados e controles

Faça um gráfico de dispersão entre “tempo” e o logaritmo natural da renda per capita das observações não tratadas

	A	B	C	D	E	F			I	J	K	L	M	N
1	ID	scolaridad	idade	luz	renda_pc	tratado			ra_tem	ln_renda_pc				
2696	1597	4	49	1	2310	0		0	7.75					
2697	1069	4	56	1	2610	0		0	7.87					
2698	2941	5	24	1	2610	0	1	0	7.87					
2699	2222	5	41	1	2610	0	1	0	7.87					
2700	1292	8	53	1	2860	0	1	0	7.96					
2701	131	6	42	1	3540	0	1	0	8.17					
2702	1897	7	45	1	3613.33	0	1	0	8.19					
2703	1454	8	51	1	3843.33	0	1	0	8.25					
2704	374	2	25	1	4110	0	1	0	8.32					
2705	2338	3	39	1	4110	0	1	0	8.32					
2706														
2707	1109	2	54	1	4	1	1	0	1.39					
2708	2040	2	42	1	11	1	1	0	2.40					
2709	1268	2	52	1	11	1	1	0	2.40					
2710	1818	1	45	1	14	1	1	0	2.64					
2711	2031	3	42	1	16	1	1	0	2.77					
2712	2116	2	41	1	17	1	1	0	2.83					
2713	2057	2	42	1	17	1	1	0	2.83					

No gráfico, acrescente as observações dos “tratados” e plote linhas de tendência para visualizar suas diferentes trajetórias de média



Desfaça a separação usada para a análise gráfica

The screenshot shows a Microsoft Excel spreadsheet with a data table. The table has columns labeled A through N and rows numbered 1 through 2708. Column J, labeled 'ln_renda_pc', is highlighted in yellow. A dialog box titled 'Excluir' (Exclude) is open, showing options to exclude the entire row, column, or shift cells. The 'Linha inteira' (Entire row) option is selected. The status bar at the bottom indicates 'Pronto' and '140%' zoom.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	ID	escolaridad	idade	luz	renda_pc	tratado	conjuje	tempo	tra_tem	ln_renda_pc				
2691	1821	7	46	1	1610	0	1	1	0	7.38				
2692	709	3	20	1	1865	0	1	1	0	7.53				
2693	1993	6	44	1	1943.33	0	1	1	0	7.57				
2694	987	2	57	1	2010	0	1	1	0	7.61				
2695	106	2	26	1	2110	0	1	1	0	7.65				
2696	1597	4	49	1			1	1	0	7.75				
2697	1069	4	56	1			1	1	0	7.87				
2698	2941	5	24	1			1	1	0	7.87				
2699	2222	5	41	1			1	1	0	7.87				
2700	1292	8	53	1			1	1	0	7.96				
2701	131	6	42	1			1	1	0	8.17				
2702	1897	7	45	1	3613.33	0	1	1	0	8.19				
2703	1454	8	51	1	3843.33	0	1	1	0	8.25				
2704	374	2	25	1	4110	0	1	1	0	8.32				
2705	2338	3	39	1	4110	0	1	1	0	8.32				
2706														
2707	1109	2	54	1	4	1	1	0	0	1.39				
2708	2040	2	42	1	11	1	1	0	0	2.40				

Desfaça a fórmula do logaritmo usando a ferramenta do "colar especial"

The screenshot shows the Microsoft Excel interface with a spreadsheet containing data. The columns are labeled A through N, and rows are numbered 1 through 2708. The data includes fields like ID, escolaridade, idade, luz, renda_pc, tratado, conjugue, tempo, and tra_tem. A yellow banner at the top contains the text 'Desfaça a fórmula do logaritmo usando a ferramenta do "colar especial"'. The 'Colar Especial' (Paste Special) menu is open, showing options like 'Recortar', 'Copiar', 'Opções de Colagem', and 'Colar Especial...'. The 'Colar Especial...' option is highlighted, and a sub-menu is visible with options like 'Inserir células copiadas', 'Excluir', 'Limpar conteúdo', 'Formatar células...', 'Largura da Coluna...', 'Ocultar', and 'Re-exibir'. The status bar at the bottom shows 'Média: 5.851238811', 'Contagem: 5781', 'Soma: 33820.16033', and '140%'.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	ID	escolaridade	idade	luz	renda_pc	tratado	conjugue	tempo	tra_tem	In_re				
2691	1821	7	46	1	1610	0	1	1	0					
2692	709	3	20	1	1865	0	1	1	0					
2693	1993	6	44	1	1943.33	0	1	1	0					
2694	987	2	57	1	2010	0	1	1	0					
2695	106	2	26	1	2110	0	1	1	0					
2696	1597	4	49	1	2310	0	1	1	0					
2697	1069	4	56	1	2610	0	1	1	0					
2698	2941	5	24	1	2610	0	1	1	0					
2699	2222	5	41	1	2610	0	1	1	0					
2700	1292	8	53	1	2860	0	1	1	0					
2701	131	6	42	1	3540	0	1	1	0					
2702	1897	7	45	1	3613.33	0	1	1	0					
2703	1454	8	51	1	3843.33	0	1	1	0					
2704	374	2	25	1	4110	0	1	1	0					
2705	2338	3	39	1	4110	0	1	1	0					
2706	1109	2	54	1	4	1	1	0	0					
2707	2040	2	42	1	11	1	1	0	0					
2708	1268	2	52	1	11	1	1	0	0					

Exclua a coluna da "renda_pc", para que não atrapalhe ao se usar a ferramenta de "regressão" do Excel

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	ID	escolaridade	idade	luz	renda_pc	tratado	coniuqe	tempo	tra_tem	ln_renda_pc				
2691	1821	7	46	1	1610			1	1	0	7.38			
2692	709	3	20	1	1865			1	1	0	7.53			
2693	1993	6	44	1	1943.33			1	1	0	7.57			
2694	987	2	57	1	2010			1	1	0	7.61			
2695	106	2	26	1	2110			1	1	0	7.65			
2696	1597	4	49	1	2310			1	1	0	7.75			
2697	1069	4	56	1	2610			1	1	0	7.87			
2698	2941	5	24	1	2610			1	1	0	7.87			
2699	2222	5	41	1	2610			1	1	0	7.87			
2700	1292	8	53	1	2860			1	1	0	7.96			
2701	131	6	42	1	3540	0	1	1	1	0	8.17			
2702	1897	7	45	1	3613.33	0	1	1	1	0	8.19			
2703	1454	8	51	1	3843.33	0	1	1	1	0	8.25			
2704	374	2	25	1	4110	0	1	1	1	0	8.32			
2705	2338	3	39	1	4110	0	1	1	1	0	8.32			
2706	1109	2	54	1	4	1	1	0	0	0	1.39			
2707	2040	2	42	1	11	1	1	0	0	0	2.40			
2708	1268	2	52	1	11	1	1	0	0	0	2.40			

Para uma primeira regressão, sem usar os controles (escolaridade, idade etc.), agregue as variáveis de tratamento e tempo

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	ID	escolaridade	idade	luz	conjugue	tratado	tempo	tra_tem	ln_renda_pc					
2	1780	3	45	1	1	0	0	0	2.40					
3	2921	4	23	1	1	0	0	0	2.71					
4	2808	1	28	1	1	0	0	0	2.83					
5	2074	3	41	1	1	0	0	0	2.89					
6	2812	6	27	1	1	0	0	0	2.94					
7	2138	2	40	1	1	0	0	0	2.94					
8	2955	5	21	1	1	0	0	0	3.00					
9	1222	2	52	1	1	0	0	0	3.00					
10	1214	2	52	1	1	0	0	0	3.04					
11	2577	2	33	1	1	0	0	0	3.09					
12	1694	1	46	1	1	0	0	0	3.09					
13	2155	5	40	1	1	0	0	0	3.14					
14	1289	3	51	1	0	0	0	0	3.14					
15	2846	4	26	1	1	0	0	0	3.22					
16	1920	2	43	1	1	0	0	0	3.22					
17	770	3	58	1	1	0	0	0	3.22					
18	2671	3	31	1	1	0	0	0	3.26					
19	2408	2	36	1	1	0	0	0	3.26					

Vamos usar a ferramenta de regressão do Excel para estimar um modelo simples de DD: $\ln(\text{Renda}^{\text{est}}) = \alpha + \beta_1 \times \text{'programa'} + \beta_2 \times \text{'depois'} + \beta_3 \times \text{'programa*depois'}$

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	ID	escolaridade	idade	luz	conjuje	tratado	tempo	tra_tem	ln_renda_pc					
2	1780	3	45	1	1	0	0	0	2.40					
3	2921	4	23	1	1	0	0	0	2.71					
4	2808	1	28	1	1	0	0	0	2.83					
5	2074	3	41	1	1	0	0	0	2.89					
6	2812								2.94					
7	2138								2.94					
8	2955								3.00					
9	1222								3.00					
10	1214								3.04					
11	2577								3.09					
12	1694								3.09					
13	2155	5	40	1	1	0	0	0	3.14					
14	1289	3	51	1	0	0	0	0	3.14					
15	2846	4	26	1	1	0	0	0	3.22					
16	1920	2	43	1	1	0	0	0	3.22					
17	770	3	58	1	1	0	0	0	3.22					
18	2671	3	31	1	1	0	0	0	3.26					
19	2408	2	36	1	1	0	0	0	3.26					

Análise de dados

Erreamentas de análise

- Estadística descritiva
- Ajuste exponencial
- Teste-F: duas amostras para variâncias
- Análise de Fourier
- Histograma
- Média móvel
- Geração de número aleatório
- Ordem e percentil
- Regressão
- Amostragem

OK Cancelar Ajuda

ddm.xlsx - Microsoft Excel

Arquivo | Página Inicial | Inserir | Layout da Página | Fórmulas | **Dados** | Revisão | Exibição | Acrobat

Do Access | Da Web | De Texto | De Outras Fontes | Conexões Existentes | Atualizar tudo | Conexões | Propriedades | Editar Links | Classificar | Filtro | Limpar | Reaplicar | Avançado | Texto para colunas | Remover Duplicatas | Validação de Dados | Consolidar | Teste de Hipóteses | Agrupar | Desagrupar | Subtotal | Mostrar Detalhe | Ocultar Detalhe | Análise de Dados | Solver | Análise

F1 | tratado

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	ID	escolaridade	idade	luz	conjuge	tratado	tempo	tra_tem	ln_renda_pc					
5764	1399					1	1	1	7.67					
5765	2253					1	1	1	7.67					
5766	1535					1	1	1	7.78					
5767	135					1	1	1	7.82					
5768	441					1	1	1	7.82					
5769	113					1	1	1	7.87					
5770	202					1	1	1	7.88					
5771	156					1	1	1	7.89					
5772	615					1	1	1	7.90					
5773	2298					1	1	1	7.94					
5774	1961					1	1	1	7.97					
5775	611					1	1	1	7.97					
5776	1809					1	1	1	8.06					
5777	246					1	1	1	8.13					
5778	1259	4	54	1	1	1	1	1	8.18					
5779	1108	3	56	1	1	1	1	1	8.19					
5780	603	2	28	1	1	1	1	1	8.33					
5781	1819	2	47	1	0	1	1	1	8.33					

Regressão

Entrada

Intervalo Y de entrada: \$I\$1:\$I\$5781

Intervalo X de entrada: \$F\$1:\$H\$5781

Rótulos Constante é zero

Nível de confiança 95 %

Opções de saída

Intervalo de saída:

Nova planilha: reg1

Nova pasta de trabalho

Resíduos

Resíduos Plotar resíduos

Resíduos padronizados Plotar ajuste de linha

Probabilidade normal

Plotagem de probabilidade normal

OK | Cancelar | Ajuda

Selecione adequadamente as matrizes e salve o resultado desta regressão como, por exemplo, "reg1"

Sheet1 | Pronto | 140%

ddm.xlsx - Microsoft Excel

Arquivo Página Inicial Inserir Layout da Página Fórmulas Dados Revisão Exibição Acrobat

Calibri 11 A A

Quebrar Texto Automaticamente

Fonte Alinhamento Número

Formatação Condicional Formatar como Tabela Estilos de Célula

Inserir Excluir Formatar Células

AutoSoma Preencher Limpar Classificar e Filtrar Localizar e Selecionar Edição

D5

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	RESUMO DOS RESULTADOS														
2															
3	<i>Estatística de regressão</i>														
4	R múltiplo	0.45662													
5	R-Quadrado	0.2085													
6	R-quadrado ajustado	0.20809													
7	Erro padrão	0.71372													
8	Observações	5780													
9															
10	ANOVA														
11		<i>gl</i>	<i>SQ</i>	<i>MQ</i>	<i>F</i>	<i>e significação</i>									
12	Regressão	3	775.077	258.35888	507.187	1E-292									
13	Resíduo	5776	2942.27	0.5093954											
14	Total	5779	3717.34												
15															
16		<i>Coefficiente</i>	<i>Erro padrã</i>	<i>Stat t</i>	<i>valor-P</i>	<i>% inferior</i>	<i>% superior</i>	<i>ferior 95.0</i>	<i>perior 95.0%</i>						
17	Interseção	5.4282	0.0195	278.7167	0.0000	5.3900	5.4663	5.3900	5.4663						
18	tratado	0.1383	0.0269	5.1506	0.0000	0.0857	0.1909	0.0857	0.1909						
19	tempo	0.5832	0.0275	21.2439	0.0000	0.5294	0.6370	0.5294	0.6370						
20	tra_tem	0.1905	0.0376	5.0622	0.0000	0.1167	0.2643	0.1167	0.2643						
21															
22															
23															

reg1 Sheet1

Pronto

120%

Observe que nesta primeira regressão que o ATT estimado é de 19,05%

A idade do trabalhador é uma *proxy* para sua experiência. No contexto, um resultado bastante conhecido em Economia é que a experiência apresenta retornos decrescentes de escala no mercado de trabalho

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	ID	escolaridade	idade	idade^2	luz	conjuje	tratado	tempo	tra_tem	ln_renda_pc				
2	1780	3	45	=C2^2	1	1	0	0	0	2.40				
3	2921	4	23		1	1	0	0	0	2.71				
4	2808	1	28		1	1	0	0	0	2.83				
5	2074	3	41		1	1	0	0	0	2.89				
6	2812	6	27		1	1	0	0	0	2.94				
7	2138	2	40		1	1	0	0	0	2.94				
8	2955	5	21		1	1	0	0	0	3.00				
9	1222	2	52		1	1	0	0	0	3.00				
10	1214	2	52		1	1	0	0	0	3.04				
11	2577	2	33		1	1	0	0	0	3.09				
12	1694	1	46		1	1	0	0	0	3.09				
13	2155	5	40		1	1	0	0	0	3.14				
14	1289	3	51		1	0	0	0	0	3.14				
15	2846	4	26		1	1	0	0	0	3.22				
16	1920	2	43		1	1	0	0	0	3.22				
17	770	3	58		1	1	0	0	0	3.22				
18	2671	3	31		1	1	0	0	0	3.26				
19	2408	2	36		1	1	0	0	0	3.26				

Para considerar esta observação "clássica", acrescentaremos o quadrado da idade em nossa análise

ddm.xlsx - Microsoft Excel

Arquivo | Página Inicial | Inserir | Layout da Página | Fórmulas | **Dados** | Revisão | Exibição | Acrobat

Obter Dados Externos | Conexões Existentes | Atualizar tudo | Editar Links | Classificar e Filtrar | Ferramentas de Dados | Estrutura de Tópicos | Análise

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	ID	escolaridade	idade	idade^2	luz	conjuge	tratado	tempo	tra_tem	ln_renda_pc				
5764	1399	5	52	2704	1	1	1	1	1	7.67				
5765	2253						1	1	1	7.67				
5766	1535						1	1	1	7.78				
5767	135						1	1	1	7.82				
5768	441						1	1	1	7.82				
5769	113						1	1	1	7.87				
5770	202						1	1	1	7.88				
5771	156						1	1	1	7.89				
5772	615						1	1	1	7.90				
5773	2298						1	1	1	7.94				
5774	1961						1	1	1	7.97				
5775	611						1	1	1	7.97				
5776	1809						1	1	1	8.06				
5777	246						1	1	1	8.13				
5778	1259	4	54	2916	1	1	1	1	1	8.18				
5779	1108	3	56	3136	1	1	1	1	1	8.19				
5780	603	2	28	784	1	1	1	1	1	8.33				
5781	1819	2	47	2209	1	0	1	1	1	8.33				

Regressão

Entrada

Intervalo Y de entrada:

Intervalo X de entrada:

Rótulos Constante é zero

Nível de confiança: 95 %

Opções de saída:

Intervalo de saída:

Nova planilha:

Nova pasta de trabalho

Resíduos:

Resíduos Plotar resíduos

Resíduos padronizados Plotar ajuste de linha

Probabilidade normal:

Plotagem de probabilidade normal

OK Cancelar Ajuda

Vamos agora
estimar o modelo
DD por regressão
acrescentando todos
os controles que
dispomos

Calculando Regressão ...

ddm.xlsx - Microsoft Excel

Arquivo Página Inicial Inserir Layout da Página Fórmulas Dados Revisão Exibição Acrobat

Calibri 11 A A

Quebrar Texto Automaticamente

Fonte Alinhamento Número

Formatação Condicional Formatar como Tabela Estilos de Célula

Inserir Excluir Formatar Células

AutoSoma Preencher Limpar Classificar e Filtrar Localizar e Selecionar Edição

Q22

RESUMO DOS RESULTADOS									
<i>Estatística de regressão</i>									
R múltiplo	0.518106								
R-Quadrado	0.268434								
R-quadrado ajustado	0.26742								
Erro padrão	0.686464								
Observações	5780								
ANOVA									
	gl	SQ	MQ	F	e significação				
Regressão	8	997.8603	124.7325	264.6941	0				
Resíduo	5771	2719.484	0.471233						
Total	5779	3717.344							
	Coefficiente	erro padrão	Stat t	valor-P	% inferior	% superior	ferior 95.0 perior	95.0%	
Interseção	6.224	0.116	53.428	0.000	5.996	6.452	5.996	6.452	
escolaridade	0.064	0.005	12.251	0.000	0.054	0.074	0.054	0.074	
idade	-0.064	0.005	-11.869	0.000	-0.074	-0.053	-0.074	-0.053	
idade^2	0.001	0.000	11.140	0.000	0.001	0.001	0.001	0.001	
luz	0.459	0.055	8.318	0.000	0.351	0.568	0.351	0.568	
conjuge	-0.240	0.026	-9.265	0.000	-0.291	-0.189	-0.291	-0.189	
tratado	0.162	0.026	6.261	0.000	0.111	0.213	0.111	0.213	
tempo	0.586	0.026	22.165	0.000	0.535	0.638	0.535	0.638	
tra_tem	0.187	0.036	5.156	0.000	0.116	0.258	0.116	0.258	

Com esta nova especificação, observe que nesta primeira regressão que o ATT estimado é de 18,7%

reg1 reg2 Sheet1

Pronto 100%

SUMÁRIO DO CURSO

- ~~1. Conceitos básicos em avaliação de impacto~~
- ~~2. Revisão de estatística e uso avançado de planilhas eletrônicas~~
- ~~3. Modelo de Resultados Potenciais~~
- ~~4. Econometria: revisão do modelo de mínimos quadrados~~
- ~~5. Desenho de uma avaliação para uma política~~
- ~~6. Modelo de Diferenças em Diferenças~~
7. Técnicas de Pareamento
8. Estudo de caso e aplicações de diferentes técnicas econométricas
9. Discussão sobre os trabalhos de conclusão de curso
10. Revisão de matemática financeira
11. O Cálculo do Retorno Econômico
12. Conclusão do curso: apresentação dos trabalhos

MÉTODO DE AVALIAÇÃO POR PAREAMENTO

- A técnica do “pareamento” é um método não-experimental que busca construir um grupo de controle semelhante ao grupo de tratamento em termos de determinadas características observáveis pelo pesquisador
- De acordo com as hipóteses do método, cada membro do grupo de tratamento teria um par no grupo de controle que representa o resultado que ele teria obtido caso não fosse tratado
- De outra forma, as hipóteses do pareamento postulam que, ao comparar dois indivíduos, um no grupo de controle e outro no grupo de tratamento, com as mesmas características observáveis, o único fator que diferencia os resultados destes indivíduos é a participação ou não no programa

1ª HIPÓTESE: SELEÇÃO NOS OBSERVÁVEIS

- A hipótese principal deste método, que tem como objetivo estimar o efeito médio do tratamento sobre os tratados (ATT), é que o vetor de variáveis observáveis X contém todas as informações que descrevem o resultado potencial na ausência do tratamento, $Y(0)$
- Ao levar em consideração estas características observáveis, o avaliador controla por todas as variáveis que estão relacionadas ao resultado potencial na ausência de tratamento e que também afetam a decisão do indivíduo em participar ou não
- Assim, ao controlar pelo vetor X , a variável $Y(0)$ torna-se independente de T
- Esta hipótese é conhecida como seleção nos observáveis, ou ignorabilidade, ou ainda não-confundimento

2ª HIPÓTESE: SUPORTE COMUM

- Precisamos que a região do vetor X que engloba as características dos indivíduos tratados também represente as características dos indivíduos que estão no grupo de não-tratados
- Então é preciso que, ao observar apenas o vetor X , o pesquisador não saiba ao certo se o indivíduo participou ou não do tratamento
- Formalmente, isto significa que
$$0 < Pr[T_i = 1 | X_i] < 1$$

PROBLEMAS COMUNS COM O MÉTODO DE PAREAMENTO

Este método deve lidar com dois tipos de problemas:

1. O problema da dimensão:
 - Com K variáveis **binárias**, o número de potenciais pares é 2^k e cresce exponencialmente com K
 - Com K variáveis **contínuas**, a dimensionalidade é intratável
2. Falta de suporte comum: isto ocorre quando não é possível encontrar uma unidade no grupo de controle que tenha os mesmos valores de X para uma unidade do grupo de tratados

RESOLVENDO O PROBLEMA DA DIMENSIONALIDADE: ESCORE DE PROPENSÃO

- O “escore de propensão” transforma o problema multidimensional em um problema unidimensional
- O “escore de propensão” é a probabilidade condicional de receber o tratamento dadas as variáveis observadas X antes do tratamento:

$$p(X) = \Pr\{T = 1 | X\} = E\{T | X\}$$

ESTRATÉGIA DE ESTIMAÇÃO DO ATT USANDO PAREAMENTO

✓ Passos a seguir:

1. Estimação do “escore de propensão”
2. Identificação do suporte comum
3. Identificação da estimativa de efeito individual
 $\beta_i = y_i(1) - y_i(0)$ dado o “escore de propensão”
4. Estimação do efeito médio de tratamento (ATT)

EXERCÍCIO: PROGRAMA “BRASIL MELHOR” (EX2)

O programa “Brasil Melhor” – fictício – visa melhorar o desempenho escolar dos alunos do ensino fundamental...

psm1.xlsx - Microsoft Excel

Arquivo Página Inicial Inserir Layout da Página Fórmulas Dados Revisão Exibição Acrobat

Calibri 11 A A

Quebrar Texto Automaticamente

Fonte Alinhamento Número

Formatação Condicional Formatar como Tabela Estilos de Célula

Inserir Excluir Formatar Células

AutoSoma Preencher Limpar Classificar e Filtrar Localizar e Selecionar Edição

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	ID	nota	tratado	mulher	cor	estudo_mae								
2	1	86.9	1	0	0	7								
3	2	86.9	1	0	1	6								
4	3	86.8	1	0	0	6								
5	4	86.7	1	0	1	6								
6	5	86.7	1	0	1	6								
7	6	86.7	1	1	1	6								
8	7	86.6	1	1	1	6								
9	8	86.6	1	1	1	6								
10	9	86.5	1	1	0	5								
11	10	86.5	1	0	1	5								
12	11	86.4	1	0	0	5								
13	12	86.4	1	0	1	5								
14	13	86.4	1	1	1	5								
15	14	86.4	1	1	0	5								
16	15	86.3	1	0	1	5								
17	16	86.1	1	0	1	5								
18	17	86	1	0	0	5								
19	18	85.9	1	0	0	4								

Sheet1

Pronto

140%

Neste exercício, a variável de resultado estudada é a nota.

O vetor X é formado pelas variáveis “mulher”, “cor” e “estudo_mae”.

O primeiro passo para a implementação da técnica é a estimação do “escore de propensão”: $p(X) = Pr\{T = 1 | X\} = E\{T | X\}$

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	ID	nota	tratado	mulher	cor	estudo_mae								
729	728	79.5	0	0	1	3								
730	729	79.2	0	1	1	3								
731	730	79.2	0	1	1	2								
732	731	79.2	0	0	1	3								
733	732	79.1	0	0	1	2								
734	733	78.9	0	0	1	2								
735	734	78.7	0	0	1	2								
736	735	78.7	0	1	0	2								
737	736	78.6	0	0	0	2								
738	737	78.5	0	0	1	2								
739	738	78.4	0	0	1	2								
740	739	78.2	0	1	0	2								
741	740	76.5	0	1	0	2								
742	741	76.3	0	0	1	2								
743	742	75.9	0	0	0	2								
744	743	75.8	0	1	0	2								
745	744	75.7	0	0	0	1								
746	745	75.4	0	1	0	1								

Regressão

Entrada

Intervalo Y de entrada:

Intervalo X de entrada:

Rótulos Constante é zero

Nível de confiança: 95 %

Opções de saída

Intervalo de saída:

Nova planilha:

Nova pasta de trabalho

Resíduos

Resíduos Plotar resíduos

Resíduos padronizados Plotar ajuste de linha

Probabilidade normal

Plotagem de probabilidade normal

Em um primeiro momento faremos isso usando um modelo de regressão linear

psm1.xlsx - Microsoft Excel

Arquivo Página Inicial Inserir Layout da Página

Calibri 11

Área de Tran...

Fonte

I9

RESUMO DOS RESULTADOS

Estatística de regressão

4	R múltiplo	0.26821
5	R-Quadrado	0.07194
6	R-quadrado ajustado	0.06818
7	Erro padrão	0.45381
8	Observações	745

ANOVA

	gl	SQ	MQ	F	e significação	
12	Regressão	3	11.8284	3.9428	19.1454	5.8E-12
13	Resíduo	741	152.601	0.20594		
14	Total	744	164.43			

	Coefficiente	erro padrão	Stat t	valor-P	% inferior	% superior	inferior 95.0%	superior 95.0%	
17	Interseção	0.5884	0.0483	12.1939	0.0000	0.4936	0.6831	0.4936	0.6831
18	mulher	-0.0130	0.0335	-0.3892	0.6973	-0.0789	0.0528	-0.0789	0.0528
19	cor	0.0222	0.0335	0.6641	0.5068	-0.0435	0.0879	-0.0435	0.0879
20	estudo_mae	-0.0558	0.0075	-7.4500	0.0000	-0.0704	-0.0411	-0.0704	-0.0411

reg1 Sheet1

Pronto

120%

Observe que as estatísticas t associadas aos coeficientes das variáveis “mulher” e “cor” não suportam a hipótese de que os coeficientes são estatisticamente diferentes de zero aos níveis de significância padrão (5% e 10%).

Porém, cabe lembrar que essa estatística de teste assume que o erro possui suporte irrestrito. E isso não acontece no presente caso, pois o erro está limitado no intervalo unitário.

O modelo gera estimativas consistentes para $p(X) = Pr\{T = 1 | X\} = E\{T | X\}$, mas as estatísticas de teste reportadas devem ser desconsideradas.

psm1.xlsx - Microsoft Excel

Arquivo Página Inicial Inserir Layout da Página Fórmulas Dados Revisão Exibição Acrobat

Colar Fonte Alinhamento Número Estilo Células Edição

FISHER $=\$C\$1+\$D\$1*D4+\$E\$1*E4+\$F\$1*F4$

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1			0.5884	-0.0130	0.0222	-0.0558								
2														
3	ID	nota	tratado	mulher	cor	estudo_mae	p(x)							
4	1	86.9	1	0	0	7	$=\$C\$1+\$D\$1*D4+\$E\$1*E4+\$F\$1*F4$							
5	2	86.9	1	0	1	6								
6	3	86.8	1	0	0	6								
7	4	86.7	1	0	1	6								
8	5	86.7	1	0	1	6								
9	6	86.7	1	1	1	6								
10	7	86.6	1	1	1	6								
11	8	86.6	1	1	1	6								
12	9	86.5	1	1	0	5								
13	10	86.5	1	0	1	5								
14	11	86.4	1	0	0	5								
15	12	86.4	1	0	1	5								
16	13	86.4	1	1	1	5								
17	14	86.4	1	1	0	5								
18	15	86.3	1	0	1	5								
19	16	86.1	1	0	1	5								

O próximo passo é computar o escore de propensão ao tratamento para cada unidade de observação: $p(X_i)$

reg1 Sheet1

Aponte 140%

The image shows a Microsoft Excel spreadsheet with a data table. The table has columns labeled ID, nota, tratado, mulher, cor, estudo_mae, and p(X). The first row of data (row 4) is highlighted in yellow. A dialog box titled 'Mover ou copiar' is open, showing options to move the selected data to a specific location in the workbook. The dialog box has the following fields and options:

- Para pasta: psm1.xlsx
- Antes da planilha: reg1, Sheet1, (mover para o final)
- Criar uma cópia
- Buttons: OK, Cancelar

The spreadsheet data is as follows:

	A	B	C	D	E	F	G	H	I
1			0.5884	-0.0130	0.0222	-0.0558			
2									
3	ID	nota	tratado	mulher	cor	estudo_mae	p(X)		
4	1	86.9	1	0	0	7	0.19806		
5	2	86.9	1	0	1	6	0.27603		
6	3	86.8	1	0	0	6	0.25382		
7	4	86.7	1	0	1	6	0.27603		
8	5	86.7	1	0	1	6	0.27603		
9	6	86.7	1	1	1	6	0.26299		
10	7	86.6	1	1	1	6	0.26299		
11	8	86.6	1	1	1	6	0.26299		
12	9	86.5	1	1	0	5	0.29653		
13	10	86.5	1	0	1	5	0.33179		
14	11	86.4	1	0	0	5	0.30957		
15	12	86.4	1	0	1	5	0.33179		
16	13	86.4	1	1	1	5	0.31874		
17	14	86.4	1	1	0	5	0.29653		
18	15	86.3	1	0	1	5	0.33179		
19	16	86.1	1	0	1	5	0.33179		

Para facilitar a análise, copie a planilha usando a ferramenta “colar especial”.

Use somente os valores numéricos, isto é, apague as fórmulas.

psm1.xlsx - Microsoft Excel

Arquivo Página Inicial Inserir Layout da Página Fórmulas Dados Revisão Exibição Acrobat

Calibri 11 A A

Quebrar Texto Automaticamente Geral

Colar

Fonte

Alinhamento

Número

Formatação Condicional

Formatar como Tabela

Estilos de Célula

Inserir Excluir Formatar

Células

AutoSoma

Preencher

Limpar

Classificar e Filtrar

Localizar e Selecionar

Edição

L7

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	ID	nota	tratado	p(X)		ID	nota	tratado	p(X)						
2	50	87.5	0	0.25382		1	86.9	1	0.19806						
3	51	87.5	0	0.14230		2	86.9	1	0.27603						
4	52	87.5	0	0.14230		3	86.8	1	0.25382						
5	53	87.5	0	0.12926		4	86.7	1	0.27603						
6	54	87.4	0	0.07350		5	86.7	1	0.27603						
7	55	87.3	0	0.22028		6	86.7	1	0.26299						
8	56	87.3	0	0.12926		7	86.6	1	0.26299						
9	57	87.2	0	0.14230		8	86.6	1	0.26299						
10	58	87.2	0	0.14230		9	86.5	1	0.29653						
11	59	87.2	0	0.15147		10	86.5	1	0.33179						
12	60	87.2	0	0.14230		11	86.4	1	0.30957						
13	61	87.2	0	0.08655		12	86.4	1	0.33179						
14	62	87.2	0	0.14230		13	86.4	1	0.31874						
15	63	87.2	0	0.15147		14	86.4	1	0.29653						
16	64	87.1	0	0.20723		15	86.3	1	0.33179						
17	65	87.1	0	0.18501		16	86.1	1	0.33179						
18	66	87.1	0	0.16452		17	86	1	0.30957						
19	67	87.1	0	0.09572		18	85.9	1	0.36533						

T=0 T=1

Mínimo 0.0177 0.1850

Máximo 0.5326 0.5326

O próximo passo é separar as observações dos grupos de tratamento e controle e verificar que existe um suporte comum para $p(X_i)$.

Neste caso o suporte comum é o intervalo [0,1850; 0,5326]

reg1 reg2 Sheet1 ATT Sheet1 (2)

Pronto

psm1.xlsx - Microsoft Excel

Arquivo | Página Inicial | Inserir | Layout da Página | Fórmulas | Dados | Revisão | Exibição | Acrobat

Área de Tran... | Fonte | Arial | 10

Classificar

Adicionar Nível | Excluir Nível | Copiar Nível | Opções... | Meus dados contêm cabeçalhos

Coluna	Classificar em	Ordem
Classificar por	p(X)	Do Menor para o Maior
E depois por	nota	Do Menor para o Maior

	A	B	C	D
1	ID	nota	tratado	p(X)
2	50	87.5	0	0.25382
3	51	87.5	0	0.14230
4	52	87.5	0	0.14230
5	53	87.5	0	0.12926
6	54	87.4	0	0.07350
7	55	87.3	0	0.22028
8	56	87.3	0	0.12926
9	57	87.2	0	0.14230
10	58	87.2	0	0.14230
11	59	87.2	0	0.15147
12	60	87.2	0	0.14230
13	61	87.2	0	0.08655
14	62	87.2	0	0.14230
15	63	87.2	0	0.15147
16	64	87.1	0	0.20723
17	65	87.1	0	0.18501
18	66	87.1	0	0.16452
19	67	87.1	0	0.09572

5	86.7	1	0.27603
6	86.7	1	0.26299
7	86.6	1	0.26299
8	86.6	1	0.26299
9	86.5	1	0.29653
10	86.5	1	0.33179
11	86.4	1	0.30957
12	86.4	1	0.33179
13	86.4	1	0.31874
14	86.4	1	0.29653
15	86.3	1	0.33179
16	86.1	1	0.33179
17	86	1	0.30957
18	85.9	1	0.36533

OK | Cancelar

Pronto | Média: 120.4133005 | Contagem: 2004 | Soma: 240826.6009 | 140%

Classifique as observações do grupo de controle por $p(X_i)$ e "nota".

psm1.xlsx - Microsoft Excel

Arquivo Página Inicial Inserir Layout da Página Fórmulas Dados Revisão Exibição Acrobat

Colar Área de Tran... Fonte Alinhamento Número

Formato... Formatar Estilos... Formatação Condicional como Tabela Estilos de Célula

Classificar Localizar e Filtrar Selecionar

F2 1

	A	B	C	D	E	F	G	H	I	J	N	C
1	ID	nota	tratado	p(X)		ID	nota	tratado	p(X)			
2	109	85	0	0.01774		1	86.9	1	0.19806			
3	258	85	0	0.01774		2	86.9	1	0.27603			
4	407	85	0	0.01774		3	86.					
5	556	85	0	0.01774		4	86.					
6	705	85	0	0.01774		5	86.					
7	54	87.4	0	0.07350		6	86.					
8	203	87.4	0	0.07350		7	86.					
9	352	87.4	0	0.07350		8	86.					
10	501	87.4	0	0.07350		9	86.					
11	650	87.4	0	0.07350		10	86.					
12	128	80	0	0.08655		11	86.					
13	277	80	0	0.08655		12	86.					
14	426	80	0	0.08655		13	86.4	1	0.31874			
15	575	80	0	0.08655		14	86.4	1	0.29653			
16	724	80	0	0.08655		15	86.3	1	0.33179			
17	72	87	0	0.08655		16	86.1	1	0.33179			
18	221	87	0	0.08655		17	86	1	0.30957			
19	370	87	0	0.08655		18	85.9	1	0.36533			

Classifique as observações do grupo de tratamento por $p(X_i)$ e "nota".

Classificar

Adicionar Nível Excluir Nível Copiar Nível Opções... Meus dados contêm cabeçalhos

Coluna	Classificar em	Ordem
Classificar por p(X)	Valores	Do Menor para o Maior
E depois por nota	Valores	Do Menor para o Maior

OK Cancelar

Pronto Média: 101.8657131 Contagem: 984 Soma: 99828.39889 140%

psm1.xlsx - Microsoft Excel

Arquivo Página Inicial Inserir Layout da Página Fórmulas Dados Revisão Exibição Acrobat

Calibri 11 A A Quebrar Texto Automaticamente Geral

Colar Área de Tran... Fonte Alinhamento Número Formatação Condicional Formata como Tabe Estilo

M24

	A	B	C	D	E	F	G	H	I	J
1	ID	nota	tratado	p(X)		ID	nota	tratado	p(X)	
14	426	80	0	0.08655		301	86.8	1	0.25382	
15	575	80	0	0.08655		450	86.8	1	0.25382	
16	724	80	0	0.08655		599	86.8	1	0.25382	
17	72	87	0	0.08655		7	86.6	1	0.26299	
18	221	87	0	0.08655		8	86.6	1	0.26299	
19	370	87	0	0.08655		156	86.6	1	0.26299	
20	519	87	0	0.08655		157	86.6	1	0.26299	
21	668	87	0	0.08655		305	86.6	1	0.26299	
22	61	87.2	0	0.08655		306	86.6	1	0.26299	
23	210	87.2	0	0.08655		454	86.6	1	0.26299	
24	359	87.2	0	0.08655		455	86.6	1	0.26299	
25	508	87.2	0	0.08655		603	86.6	1	0.26299	
26	657	87.2	0	0.08655		604	86.6	1	0.26299	
27	67	87.1	0	0.09572		6	86.7	1	0.26299	
28	68	87.1	0	0.09572		155	86.7	1	0.26299	
29	216	87.1	0	0.09572		304	86.7	1	0.26299	
30	217	87.1	0	0.09572		453	86.7	1	0.26299	
31	365	87.1	0	0.09572		602	86.7	1	0.26299	

reg1 Sheet1 Sheet1 (2)

Pronto 140%

Numa inspeção visual, note que tanto no grupo de controle quanto no de tratamento existem observações com mesmo escore de propensão mas com notas diferentes.

Portanto, precisamos definir um critério de comparação de notas das observações com mesmo $p(X_i)$

PAREAMENTO POR VIZINHO MAIS PRÓXIMO

Pela observação anterior, precisamos agora de um critério para parear as unidades de tratamento e controle. Usaremos aqui o critério do “vizinho mais próximo” (detalhes nas pág. 90 a 92 do livro).

- **Vantagens:**
 - ❑ Para cada unidade tratada sempre é encontrado um par, não tratado, evitando a exclusão de observações tratadas
 - ❑ É facilmente implementado usando as fórmulas do Excel
- **Principal Problema:** em algumas bases de dados pode gerar pareamento de indivíduos com escore de propensão muito diferentes, pois o vizinho mais próximo pode não ser tão próximo. O que não será o caso deste exercício em particular, já que para toda combinação **X** no grupo de tratamento encontra-se uma igual combinação no grupo de controle.

psm1.xlsx - Microsoft Excel

Arquivo Página Inicial Inserir Layout da Página Fórmulas Dados Revisão Exibição Acrobat

Colar 11 A A Quebrar Texto Automaticamente Geral Formatação Condicional Formatar como Tabela Estilos de Célula Inserir Excluir Formatar

Área de Tran... Fonte Alinhamento Número Estilo Células

FISHER X ✓ f =B2

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	ID	nota	tratado	p(X)	Y(0)_i	ID	nota	tratado	p(X)						
2	109	85	0	0.01774	=B2	26	85	1	0.18501						
3	258	85	0	0.01774		175	85	1	0.18501						
4	407	85	0	0.01774		324	85	1	0.18501						
5	556	85	0	0.01774		473	85	1	0.18501						
6	705	85	0	0.01774		622	85	1	0.18501						
7	54	87.4	0	0.07350		1	86.9	1	0.19806						
8	203	87.4	0	0.07350		150	86.9	1	0.19806						
9	352	87.4	0	0.07350		299	86.9	1	0.19806						
10	501	87.4	0	0.07350		448	86.9	1	0.19806						
11	650	87.4	0	0.07350		597	86.9	1	0.19806						
12	128	80	0	0.08655		3	86.8	1	0.25382						
13	277	80	0	0.08655		152	86.8	1	0.25382						
14	426	80	0	0.08655		301	86.8	1	0.25382						
15	575	80	0	0.08655		450	86.8	1	0.25382						
16	724	80	0	0.08655		599	86.8	1	0.25382						
17	72	87	0	0.08655		7	86.6	1	0.26299						
18	221	87	0	0.08655		8	86.6	1	0.26299						
19	370	87	0	0.08655		156	86.6	1	0.26299						

Para iniciar o procedimento de pareamento, é útil criar uma coluna para $y_i(0)$ junto ao grupo de controle.

psm1.xlsx - Microsoft Excel

Arquivo | Página Inicial | Inserir | Layout da Página | Fórmulas | Dados | Revisão | Exibição | Acrobat

Calibri 11

Fonte | Alinhamento | Número

Formatação Condicional | Formatar como Tabela | Estilos de Célula | Inserir | Excluir | Formatar | Células

AutoSoma | Preencher | Limpar | Classificar e Filtrar | Localizar e Selecionar | Edição

FISHER | X ✓ f_x =SOMASE(\$D\$2:\$D\$501;I2;\$E\$2:\$E\$501)/CONT.SE(\$D\$2:\$D\$501;I2)

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	ID	nota	tratado	p(X)	Y(0)_i	ID	nota	tratado	p(X)	contrafactual			
2	109	85	0	0.01774	85	26	85	1	0.18501	=SOMASE(\$D\$2:\$D\$501;I2;\$E\$2:\$E\$501)/			
3	258	85	0	0.01774	85	175	85	1	0.18501	CONT.SE(\$D\$2:\$D\$501;I2)			
4	407	85	0	0.01774	85	324	85	1	0.18501	SOMASE(intervalo; critérios; [intervalo_soma])			
5	556	85	0	0.01774	85	473	85	1	0.18501				
6	705	85	0	0.01774	85	622	85	1	0.18501				
7	54	87.4	0	0.07350	87.4	1	86.9	1	0.19806				
8	203	87.4	0	0.07350	87.4	150							
9	352	87.4	0	0.07350	87.4	299							
10	501	87.4	0	0.07350	87.4	448							
11	650	87.4	0	0.07350	87.4	597							
12	128	80	0	0.08655	80	3							
13	277	80	0	0.08655	80	152							
14	426	80	0	0.08655	80	301							
15	575	80	0	0.08655	80	450							
16	724	80	0	0.08655	80	599							
17	72	87	0	0.08655	87	7							
18	221	87	0	0.08655	87	8	86.6	1	0.26299				

reg1 | Sheet1 | Sheet1 (2)

Edita | 150%

O passo seguinte é computar o contrafactual estimado da unidade tratada pela média da nota dos controles que possuem mesmo escore de propensão (a fórmula $\hat{Y}_i(0) = \frac{1}{M} \sum_{j \in H_M(i)} Y_j$ da pág. 91 do livro). Para tanto, podemos combinar as fórmulas SOMASE e CONT.SE do Excel.

psm1.xlsx - Microsoft Excel

Arquivo Página Inicial Inserir Layout da Página Fórmulas Dados Revisão Exibição Acrobat

Calibri 11

Fonte Alinhamento Número

Formatação Condicional Formatar como Tabela Estilos de Célula

Inserir Excluir Formatar Células

AutoSoma Preencher Limpar Classificar e Filtrar Localizar e Selecionar Edição

FISHER X ✓ f_x =G246-J246

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	ID	nota	tratado	p(X)	Y(0)_i	ID	nota	tratado	p(X)	contrafactual	beta_i		
241	693	86.4	0	0.29653	86.4	619	85.4	1	0.51956	75.40	10.00		
242	101	86	0	0.30957	86	48	75.5	1	0.53260	75.70	-0.20		
243	250	86	0	0.30957	86	197	75.5	1	0.53260	75.70	-0.20		
244	399	86	0	0.30957	86	346	75.5	1	0.53260	75.70	-0.20		
245	548	86	0	0.30957	86	495	75.5	1	0.53260	75.70	-0.20		
246	697	86	0	0.30957	86	644	75.5	1	0.53260	75.70	=G246-J246		
247	100	86.1	0	0.30957	86.1								
248	249	86.1	0	0.30957	86.1								
249	398	86.1	0	0.30957	86.1								
250	547	86.1	0	0.30957	86.1								
251	696	86.1	0	0.30957	86.1								
252	98	86.3	0	0.30957	86.3								
253	247	86.3	0	0.30957	86.3								
254	396	86.3	0	0.30957	86.3								
255	545	86.3	0	0.30957	86.3								
256	694	86.3	0	0.30957	86.3								
257	94	86.5	0	0.30957	86.5								

reg1 Sheet1 Sheet1 (2)

Edita

150%

O próximo passo é a identificação da estimativa de efeito individual $\beta_i = y_i(1) - y_i(0)$ dado o “escore de propensão”

psm1.xlsx - Microsoft Excel

Arquivo Página Inicial Inserir Layout da Página Fórmulas Dados Revisão Exibição Acrobat

Calibri 11

Colar

Área de Tran...

Fonte

Alinhamento

Número

Estilo

Células

Classificar e Filtrar Selecionar Edição

K253

fx

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	ID	nota	tratado	p(X)	Y(0)_i	ID	nota	tratado	p(X)	contrafactual	beta_i		
239	544	86.4	0	0.29653	86.4	321	85.4	1	0.51956	75.40	10.00		
240	692	86.4	0	0.29653	86.4	470	85.4	1	0.51956	75.40	10.00		
241	693	86.4	0	0.29653	86.4	619	85.4	1	0.51956	75.40	10.00		
242	101	86	0	0.30957	86	48	75.5	1	0.53260	75.70	-0.20		
243	250							1	0.53260	75.70	-0.20		
244	399							1	0.53260	75.70	-0.20		
245	548							1	0.53260	75.70	-0.20		
246	697							1	0.53260	75.70	-0.20		
247	100												
248	249									ATT	0.18811		
249	398									Desvio	1.79		
250	547									Intervalo	0.23		
251	696									Limite Inf.	-0.04		
252	98									Limite Sup.	0.41		
253	247												
254	396												
255	545	86.3	0	0.30957	86.3								

Pronto

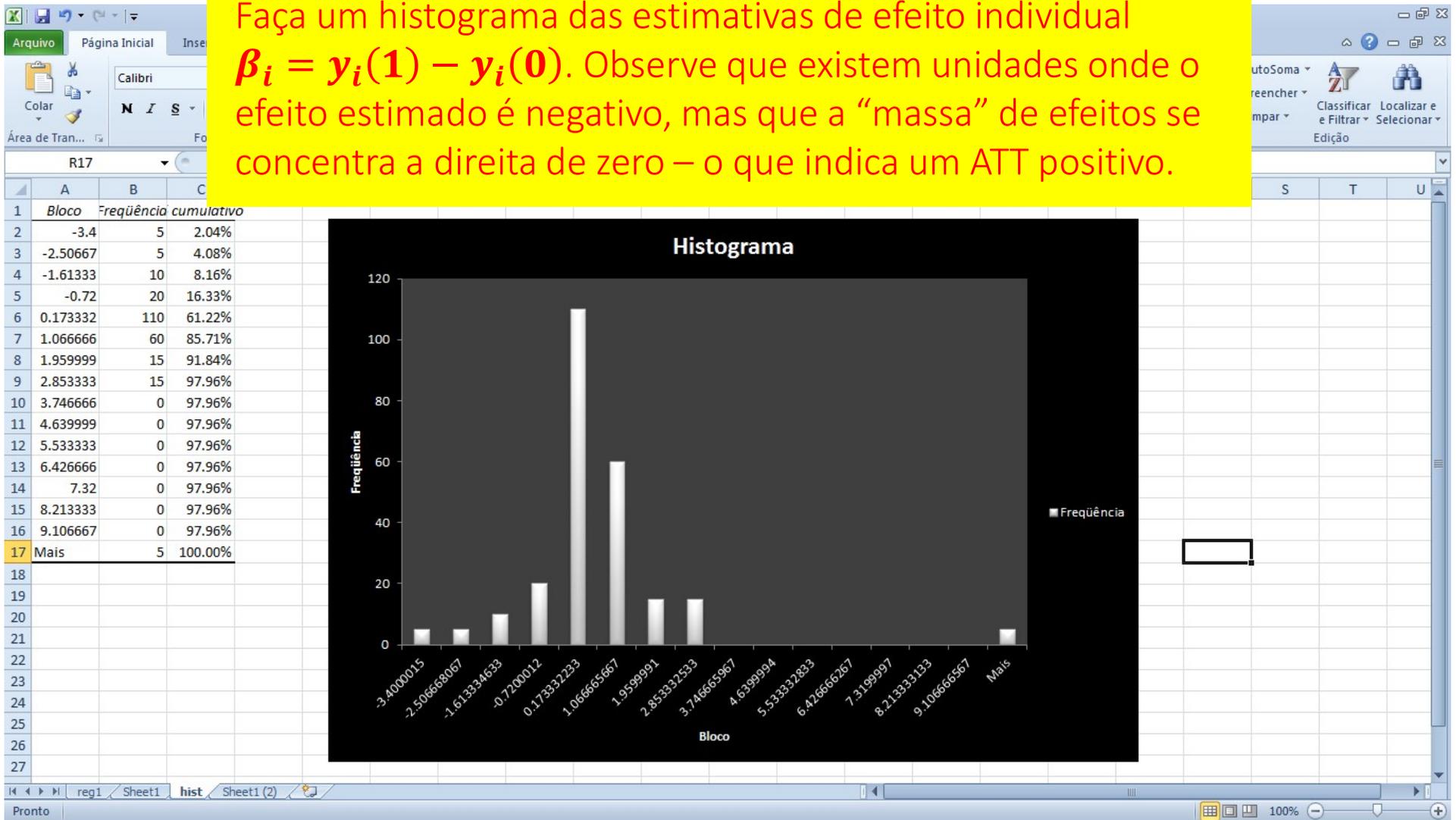
reg1 Sheet1 hist Sheet1 (2)

150%

Calcule o ATT a partir da média das estimativas de β_i

Compute um intervalo de 95% de confiança para o ATT estimado aqui em 0,188107826 pontos (o valor gerado pelo procedimento no STATA estimando $p(X_i)$ com um Logit é de 0.188107789). Observa-se que não se pode aceitar a hipótese de que o programa gerou efeito médio positivo sobre os tratados.

Faça um histograma das estimativas de efeito individual $\beta_i = y_i(1) - y_i(0)$. Observe que existem unidades onde o efeito estimado é negativo, mas que a “massa” de efeitos se concentra a direita de zero – o que indica um ATT positivo.



Vamos comparar o ATT estimado pelo método do pareamento com o ATT estimado pelo método da regressão linear.

psm1.xlsx - Microsoft Excel

Arquivo | Página Inicial | Inserir | Layout da Página | Fórmulas | Dados | Revisão | Exibição | Acrobat

Do Access | Da Web | De Texto | De Outras Fontes

Obter Dados Externos

Análise de Dados | Solver

Análise

C3

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1			0.5884	-0.0130	0.0222	-0.0558								
2														
3	ID	nota	tratado	mulher	cor	estudo_mae	p(X)							
733	730	79.2	0	1	1	2	0.48601							
734	731	79.2	0	0	1	3	0.44330							
735	732	79.1	0	0	1	2	0.49906							
736	733	78.9	0	0	1	2	0.49906							
737	734	78.7	0	0	1	2	0.49906							
738	735	78.7	0	1	0	2	0.46380							
739	736	78.6	0	0	0	2	0.47684							
740	737	78.5	0	0	1	2	0.49906							
741	738	78.4	0	0	1	2	0.49906							
742	739	78.2	0	1	0	2	0.46380							
743	740	76.5	0	1	0	2	0.46380							
744	741	76.3	0	0	1	2	0.49906							
745	742	75.9	0	0	0	2	0.47684							
746	743	75.8	0	1	0	2	0.46380							
747	744	75.7	0	0	0	1	0.53260							
748	745	75.4	0	1	0	1	0.51956							

Regressão

Entrada

Intervalo Y de entrada: \$B\$3:\$B\$748

Intervalo X de entrada: \$C\$3:\$F\$748

Rótulos Constante é zero

Nível de confiança: 95 %

Opções de saída

Intervalo de saída:

Nova planilha: reg2

Nova pasta de trabalho

Resíduos

Resíduos Plotar resíduos

Resíduos padronizados Plotar ajuste de linha

Probabilidade normal

Plotagem de probabilidade normal

OK | Cancelar | Ajuda

reg1 | Sheet1 | ATT | Sheet1 (2)

Aponte

140%

psm1.xlsx - Microsoft Excel

Arquivo Página Inicial Inserir Layout da Página Fórmulas Dados Revisão Exibição Acrobat

Calibri 11 A A

Quebrar Texto Automaticamente Geral

Colar

Fonte Alinhamento

Área de Tran... Fonte Alinhamento

G10

	A	B	C	D	E	F	G
1	RESUMO DOS RESULTADOS						
2							
3	<i>Estatística de regressão</i>						
4	R múltiplo	0.83264					
5	R-Quadrado	0.69328					
6	R-quadrado ajustado	0.69162					
7	Erro padrão	2.10755					
8	Observações	745					
9							
10	ANOVA						
11		<i>gl</i>	<i>SQ</i>	<i>MQ</i>	<i>F</i>	<i>e significação</i>	
12	Regressão	4	7429.47	1857.37	418.158	3E-188	
13	Resíduo	740	3286.92	4.44179			
14	Total	744	10716.4				
15							
16		<i>Coefficiente</i>	<i>erro padrão</i>	<i>Stat t</i>	<i>valor-P</i>	<i>% inferior</i>	<i>% superior</i>
17	Interseção	76.322	0.246	310.836	0.000	75.840	76.804
18	tratado	1.076	0.171	6.306	0.000	0.741	1.411
19	mulher	-0.176	0.156	-1.132	0.258	-0.482	0.129
20	cor	0.202	0.155	1.299	0.194	-0.103	0.507
21	estudo_mae	1.452	0.036	40.288	0.000	1.381	1.523
22							

reg1 reg2 Sheet1 ATT Sheet1 (2)

Pronto

120%

ATT estimado pelo método do pareamento = 0,18 (e não é estatisticamente diferente de zero)

ATT estimado pelo método da regressão linear = 1,07 (e é estatisticamente diferente de zero)

COMO EXPLICAR ISSO? (pág. 92 e 93)

PAREAMENTO E REGRESSÃO LINEAR

- Primeiro, porque o método de regressão linear usou todas as informações da base de dados, e não apenas aquelas que se encontravam no suporte comum. Isso fez com que se ponderasse uma parte do grupo de controle que não mimetiza satisfatoriamente o contrafactual do tratamento, gerando um viés.
- Segundo, porque o estimador de mínimos quadrados (regressão linear) parte do pressuposto de que não existe correlação entre os regressores e o termo de erro. Mas a escolaridade da mãe (observada) pode estar correlacionada com a habilidade (não observada) do aluno, gerando outra fonte de viés.
- Terceiro, se tivéssemos usado uma outra regra de pareamento, diferente do “vizinho mais próximo”, teríamos encontrado outro ATT

O PROGRAMA TEM EFEITO POSITIVO?

- As técnicas utilizadas (método do pareamento e regressão linear) apontam um ATT não negativo
- Podemos afirmar que há indícios de efeito positivo do programa, muito embora seja pequeno
- Na melhor das hipóteses, um ATT de 1,07 ponto represente apenas cerca de 25% de um desvio padrão da nota média (a nota média e o desvio padrão foram estimados na Aula 2 em 83,6 e 3,8 pontos, respectivamente). O que é um impacto pequeno.
- Conclusão: há indícios de que o programa causa um impacto muito pequeno, que deve ser confrontado com seus custos para concluirmos sobre sua viabilidade

SUMÁRIO DO CURSO

- ~~1. Conceitos básicos em avaliação de impacto~~
- ~~2. Revisão de estatística e uso avançado de planilhas eletrônicas~~
- ~~3. Modelo de Resultados Potenciais~~
- ~~4. Econometria: revisão do modelo de mínimos quadrados~~
- ~~5. Desenho de uma avaliação para uma política~~
- ~~6. Modelo de Diferenças em Diferenças~~
- ~~7. Técnicas de Pareamento~~
8. Estudo de caso e aplicações de diferentes técnicas econométricas
9. Discussão sobre os trabalhos de conclusão de curso
10. Revisão de matemática financeira
11. O Cálculo do Retorno Econômico
12. Conclusão do curso: apresentação dos trabalhos

APRESENTAÇÃO

- Este estudo de caso é um material de apoio baseado no projeto “Os pioneiros: recuperação de Matas Ciliares na Microbacia do Córrego Taquara Branca”, no Município de Sumaré – SP
- O estudo de caso tem apenas uma finalidade didática, sendo que os dados utilizados foram modificados
- O exercício tem como objetivo calcular o impacto de uma intervenção ambiental (ainda não se fará a parte de retorno econômico)
- Pela natureza dos dados, a principal metodologia a ser utilizada é a regressão de diferenças em diferenças (esta mostrou impacto positivo e estatisticamente significativo)

CONTEXTUALIZAÇÃO

I – Projeto:

- a. Quais são os objetivos?
- b. Quais são as ações realizadas?
- c. Qual é o público-alvo?

- O governo de Minas Gerais, preocupado com o nível de desmatamento, decidiu lançar, em 2000, um edital para contratação de uma organização não governamental (ONG) com o intuito de realizar um trabalho relacionado ao meio ambiente
- A ONG selecionada, Instituição Filantrópica Sustentar, possuía experiência na área ambiental e apresentou uma proposta consistente, mostrando que seria possível estimar o impacto do programa
- Dentre as atividades propostas pela ONG estavam a recuperação de matas ciliares e a educação ambiental. Esta última estava relacionada à capacitação de jovens, envolvendo conscientização, conservação etc.
- O programa Verde Novo, implantado pela ONG Sustentar, teve início em 2000

CONTEXTUALIZAÇÃO (CONTINUAÇÃO)

I – Projeto:

- a. Quais são os objetivos?
- b. Quais são as ações realizadas?
- c. Qual é o público-alvo?

- O programa Verde Novo possuía os seguintes objetivos:
 - i) conservação de matas ciliares
 - ii) redução dos processos de erosão e assoreamento dos corpos hídricos
 - iii) redução da temperatura local
- A ideia seria contribuir para a recuperação e manutenção ambiental e também oferecer trabalho para a população carente da região beneficiada com o programa
- Dada a necessidade de escolha de uma região prioritária para implementação, as áreas atendidas não puderam ser sorteadas para participação no programa
- A região selecionada para receber o projeto deveria possuir baixos indicadores sociais, áreas degradadas e com baixo número de árvores por metro quadrado

CONTEXTUALIZAÇÃO (CONTINUAÇÃO)

- I – Projeto:
- Quais são os objetivos?
 - Quais são as ações realizadas?
 - Qual é o público-alvo?

Tabela 1 – Dados socioeconômicos e ambientais das mesorregiões mineiras em 2000

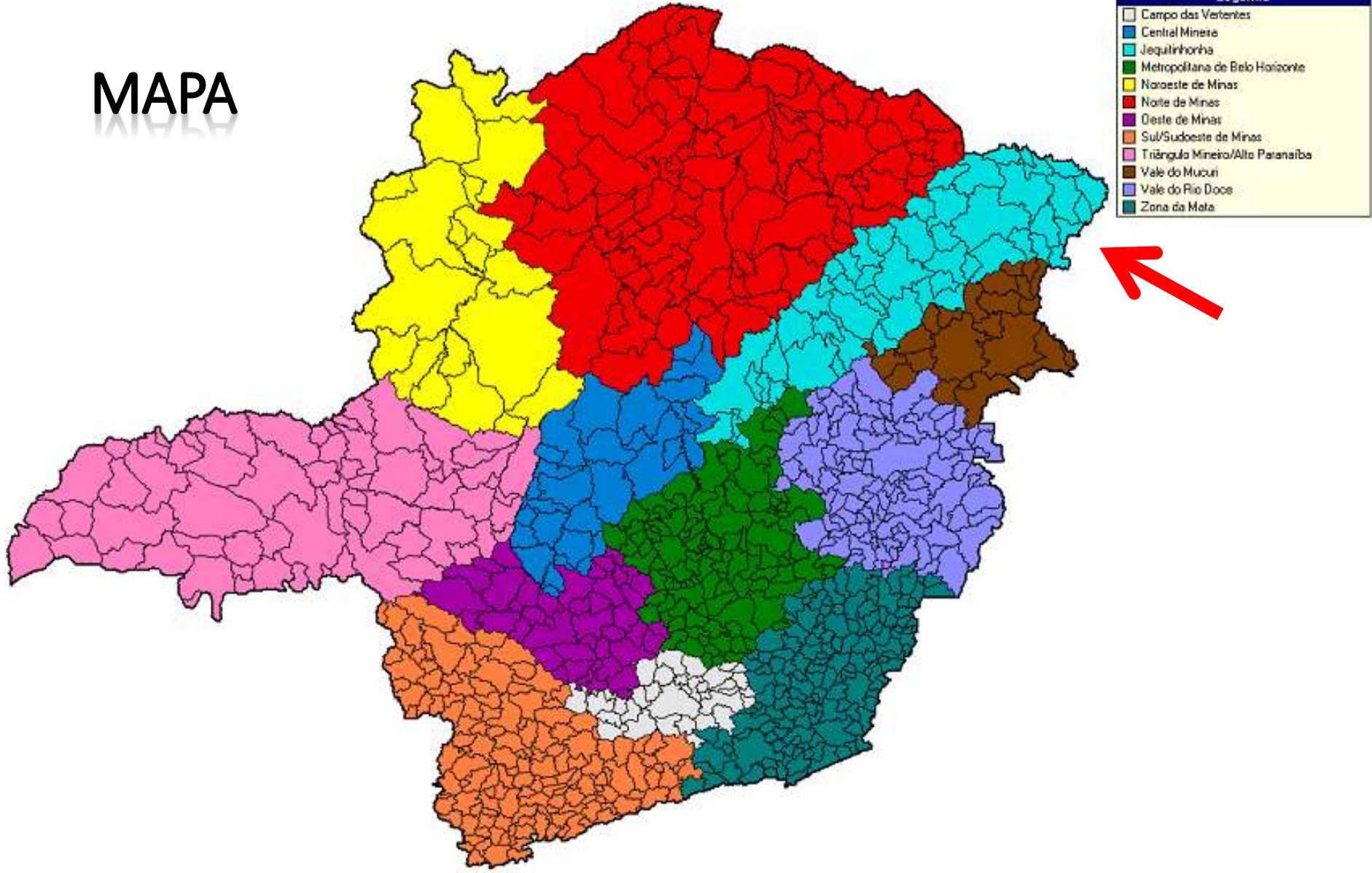
Fonte: colunas 2-5, Atlas do Desenvolvimento Humano no Brasil (2003); coluna 6, fictício.

Nota: a: proporção de pessoas com menos de 4 anos de estudo.

Mesorregião	Renda per capita média	Escolaridade ^a	% pobres	Índice de Gini	Índice de Theil	Nº árvores M ²
Jequitinhonha	113,97	60,30	60,63	0,62	0,83	70
Norte de Minas	133,03	49,30	55,39	0,63	0,81	80
Vale do Mucuri	152,42	50,80	52,67	0,64	0,87	112
Vale do Rio Doce	210,54	48,80	37,53	0,61	0,79	270
Central Mineira	226,79	46,50	27,69	0,57	0,77	290
Noroeste de Minas	229,60	46,80	34,00	0,64	1,19	120
Campo das Vertentes	237,91	42,70	28,02	0,58	0,68	280
Zona da Mata	255,12	44,40	27,97	0,59	0,74	305
Oeste de Minas	269,34	42,50	17,64	0,55	0,76	282
Minas Gerais	276,44	39,40	27,88	0,61	0,82	209
Sul/Sudoeste Minas	291,10	45,50	18,47	0,57	0,72	300
Triângulo Mineiro	330,85	35,50	51,37	0,58	0,76	100
Metrop. B.H.	355,32	29,90	21,23	0,62	0,81	295

Diante do cenário, os técnicos chegaram à conclusão de que o Vale do Jequitinhonha, situado no Nordeste mineiro, deveria ser a região selecionada para participar do projeto, pois apresentava, em média, os piores indicadores socioeconômicos e ambientais a época.

MAPA



CONTEXTUALIZAÇÃO (CONTINUAÇÃO)

I – Projeto:

- a. Quais são os objetivos?
- b. Quais são as ações realizadas?
- c. Qual é o público-alvo?

- Para cumprir os objetivos propostos foram contratados um técnico ambiental e cinco jovens da região
- Dentre as atividades desenvolvidas por esta equipe estavam: conscientizar a população em torno das áreas selecionadas acerca da importância da preservação ambiental; conservar as matas ciliares; conciliar a agricultura local com o desenvolvimento do solo; orientar a população local para a destinação adequada dos resíduos sólidos e, por fim, acompanhar, durante todo o programa, o desenvolvimento das áreas selecionadas
- Segundo a ONG, a duração do projeto seria de 10 anos, tempo necessário para verificar a evolução das áreas tratadas

SELEÇÃO DO INDICADOR DE IMPACTO E DO GRUPO CONTROLE

II – A avaliação:

a. Grupo de tratamento e controle

i. Quem são?

ii. Como foram escolhidos?

iii. Quantos são?

b. Indicadores de impacto

i. Quais são?

ii. Como se relacionam com os objetivos?

- Sabemos que a mesorregião do Vale do Jequitinhonha foi eleita para receber o projeto de recuperação da mata ciliar devido aos seus baixos indicadores socioeconômicos e ambientais
- Para receber o tratamento, foram selecionadas, nesta região, 50 áreas degradadas de 1.000 m² cada
- Restava escolher o grupo de controle

SELEÇÃO DO INDICADOR DE IMPACTO E DO GRUPO CONTROLE (CONTINUAÇÃO)

- Foi necessário selecionar uma área semelhante ao grupo de tratamento, mas que não fosse receber a intervenção realizada pela ONG
- Ao analisar a tabela 1 percebemos que a região com maior semelhança ao Vale do Jequitinhonha na época, dentre as analisadas, era a mesorregião Norte de Minas e, portanto, esta foi selecionada como o grupo de controle
- Assim, foram escolhidas 50 áreas de 1.000m² na mesorregião do Norte de Minas, com características similares às daquelas do Vale do Jequitinhonha

SELEÇÃO DO INDICADOR DE IMPACTO E DO GRUPO CONTROLE (CONTINUAÇÃO)

- Os indicadores de impacto devem estar intimamente relacionados com o objetivo e também serem de fácil mensuração, baixo custo, confiabilidade e, obviamente, estarem relacionados com as ações efetuadas pelo programa que se quer avaliar
- Este projeto possuía vários objetivos e, para verificar se o programa atingiu todos estes, seriam necessárias várias avaliações (uma para cada objetivo)
- A ONG decidiu avaliar inicialmente o objetivo i) conservação das matas ciliares. Mas para verificar posteriormente se as matas foram realmente recuperadas, era necessário encontrar um indicador que estivesse relacionado com este objetivo. O número de árvores seria um indicador factível para este caso. Assim, o indicador de impacto escolhido seria o número de árvores por 1.000 m²

COLETA DE DADOS, ESTATÍSTICA DESCRITIVA E INTERVALO DE CONFIANÇA

III – Dados:

a. Instrumentos de coleta

i. Como os dados foram coletados?

b. Informações coletadas

i. Quais variáveis explicativas foram coletadas?

ii. Por que estas variáveis foram escolhidas?

III. Estatística descritiva e intervalo de confiança

i. Qual é o valor e interpretação da média e desvio-padrão?

ii. Qual a interpretação do intervalo de confiança?

- As áreas de tratamento e controle foram demarcadas utilizando GPS e cada área recebeu estacas indicativas totalizando 100 áreas, entre tratamento e controle
- Demarcadas as áreas, restaria contar o número de árvores dentro de cada perímetro dos grupos de tratamento e controle antes e após a execução do projeto

COLETA DE DADOS, ESTATÍSTICA DESCRITIVA E INTERVALO DE CONFIANÇA

- Como o programa não foi aleatorizado, se fez necessário coletar informações adicionais além do próprio indicador
- Quais variáveis explicativas deveriam ser utilizadas? Para saber quais variáveis coletar, podemos proceder do seguinte modo:
 - i) encontrar variáveis que possibilitaram ao grupo de tratamento receber o benefício, ou seja, aquelas variáveis que fizeram com que este grupo fosse selecionado como tratamento; esgotadas estas variáveis, devemos
 - ii) coletar variáveis que sejam diferentes entre tratamento e controle e que afetam o indicador de interesse
- Como discutido anteriormente, a mesorregião do Vale do Jequitinhonha foi selecionada com base em suas características socioeconômicas e ambientais

COLETA DE DADOS, ESTATÍSTICA DESCRITIVA E INTERVALO DE CONFIANÇA

Tabela 2 – Descrição, definição e sinais esperados dos parâmetros das variáveis utilizadas

* Sinal esperado conforme literatura e técnicos.

Variável	Definição	Sinal esperado do parâmetro*
Programa	1 se a área participa do programa e 0 se não participa	Positivo
Cercamento	1 se a área é cercada e 0 caso contrário	Positivo
Agricultura	1 se a área possui algum tipo de agricultura e 0, caso contrário	Negativo
Fogo	1 se a área pegou fogo há menos de um ano 0, caso contrário	Negativo
Capim Braquiária	1 se a área possui capim braquiária e 0, caso contrário	Negativo
Animais	Número de animais de pasto por 1.000 m ²	Negativo
Distância até o centro urbano	Distância medida em km do centro urbano mais próximo	Positivo
Renda per capita média	Renda per capita média da região em que a área está situada.	Positivo

- Usando a base de dados “verde_novo” vamos construir agora intervalos de confiança (com 95% de confiança, pela distribuição normal para replicar os resultados do texto) para cada uma das variáveis coletadas nos anos de 2000 (antes) e 2010 (depois do projeto)

verde_novo.xlsx [Modo de Compatibilidade] - Microsoft Excel

Arquivo Página Inicial Inserir Layout da Página Fórmulas Dados Revisão Exibição Acrobat

Área de Tran... Fonte Alinhamento Número Estilo Células Edição

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1		areas	indicador	programa	cercamento	agricultura	fogo	braquiaria	animais	dist_centro_urbanc	renda			
96		95	1	0	0	1	1	1	98	3	11.58			
97		96	0	0	0	1	1	1	99	2	35.58			
98		97	26	0	0	1	0	1	25	12	46.42			
99		98	9	0	0	1	1	1	99	2	10.58			
100		99	1	0	0	1	1	1	80	5	28.42			
101		100	11	0	0	1	1	1	99	1	33.58			
102														
103		Média	21.04	1.00	0.30	0.54	0.70	0.78	16.82	11.18	31.41			
104		DP	14.49	0.00	0.46	0.50	0.46	0.42	11.13	6.77	20.08			
105		n	50.00	51.00	52.00	53.00	54.00	55.00	56.00	57.00	58.00			
106		IC	4.02	#NÚM!	0.13	0.14	0.12	0.11	2.91	1.76	5.17			
107		Inf	17.02	#NÚM!	0.17	0.40	0.58	0.67	13.91	9.42	26.24			
108		Sup	25.05	#NÚM!	0.43	0.68	0.82	0.89	19.73	12.94	36.57			
109		Média	16.25	0.00	0.36	0.58	0.78	0.88	56.24	4.04	30.89			
110		DP	17.36	0.00	0.48	0.50	0.42	0.33	34.53	3.42	26.14			
111		n	50.00	51.00	52.00	53.00	54.00	55.00	56.00	57.00	58.00			
112		IC	4.81	#NÚM!	0.13	0.13	0.11	0.09	9.04	0.89	6.73			
113		Inf	11.44	#NÚM!	0.23	0.45	0.67	0.79	47.20	3.15	24.16			
114		Sup	21.06	#NÚM!	0.49	0.71	0.89	0.97	65.28	4.93	37.61			
115														
116														
117														

2000 2010

Pronto

Estatísticas descritivas para o grupo de tratamento no ano 2000

verde_novo.xlsx [Modo de Compatibilidade] - Microsoft Excel

Arquivo | Página Inicial | Inserir | Layout da Página | Fórmulas | Dados | Revisão | Exibição | Acrobat

Tabela Dinâmica | Tabela | Imagem | Clip-Art | Formas | SmartArt | Instantâneo | Colunas | Linhas | Pizza | Barras | Área | Dispersão | Outros Gráficos | Linha | Coluna | Ganhos/Perdas | Segmentação de Dados | Hiperlink | Caixa de Texto e Rodapé | Cabeçalho e Rodapé | WordArt | Linha de Assinatura | Objeto | Equação | Símbolo

M97

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1		areas	indicador	programa	cercamento	agricultura	fogo	braquiaria	animais	dist_centro_urbanc	renda			
96		95	1	0	0	1	1	1	98	3	159			
97		96	0	0	0	1	1	1	99	2	135			
98		97	37	0	1	0	0	1	25	12	217			
99		98	2	0	1	1	1	1	99	2	160			
100		99	10	0	1	0	0	1	80	5	199			
101		100	0	0	0	1	1	1	99	1	137			
102														
103		Média	77.24	1.00	0.60	0.26	0.30	0.60	14.70	11.18	265.42			
104		DP	25.71	0.00	0.49	0.44	0.46	0.49	11.25	6.77	37.32			
105		n	50.00	51.00	52.00	53.00	54.00	55.00	56.00	57.00	58.00			
106		IC	7.13	#NÚM!	0.13	0.12	0.12	0.13	2.95	1.76	9.60			
107		Inf	70.11	#NÚM!	0.47	0.14	0.18	0.47	11.75	9.42	255.82			
108		Sup	84.37	#NÚM!	0.73	0.38	0.42	0.73	17.65	12.94	275.02			
109		Média	11.24	0.00	0.66	0.46	0.66	0.78	55.84	4.04	170.58			
110		DP	18.12	0.00	0.48	0.50	0.48	0.42	35.03	3.42	40.71			
111		n	50.00	51.00	52.00	53.00	54.00	55.00	56.00	57.00	58.00			
112		IC	5.02	#NÚM!	0.13	0.14	0.13	0.11	9.18	0.89	10.48			
113		Inf	6.22	#NÚM!	0.53	0.32	0.53	0.67	46.66	3.15	160.10			
114		Sup	16.26	#NÚM!	0.79	0.60	0.79	0.89	65.02	4.93	181.06			
115														
116														
117														

2000 2010

Selecione o local de destino e te

140%

Estatísticas descritivas para o grupo de tratamento no ano 2010

MAIS SOBRE AS ESTATÍSTICAS DESCRITIVAS E INTERVALOS DE CONFIANÇA

- Ao analisar os resultados calculados, é possível perceber que, antes do programa, o número de árvores por 1.000 m² é maior no grupo de tratamento, no entanto, esta diferença não é estatisticamente significativa, já que a média do grupo de tratamento, 21,04, está dentro do intervalo de confiança do grupo de controle [11,44;21,06]. Ou seja, os dois grupos são estatisticamente iguais em quantidade de árvores antes do programa
- Diante disto, será possível planejar ações de modo que o número de árvores no grupo tratado seja estatisticamente maior do que o grupo de controle em 2010? Por exemplo, percebemos que 70% das áreas do grupo de tratamento sofreram queimadas em 2000, será que ao monitorar estas áreas fazendo prevenção, controle e fiscalização para que não ocorram queimadas, o número de árvores, comparativamente ao grupo de controle, seria maior e estatisticamente diferente em 2010?
- Percebemos que em 2010 o grupo de tratamento possui, em média, maior quantidade de árvores que o grupo de controle. No entanto, essas diferenças não podem ser atribuídas inteiramente ao programa. Podemos ter, por exemplo, uma área tratada mais distante de um centro urbano que a área de controle, possibilitando maior conservação devido à menor ocupação do homem
- É necessário utilizar uma metodologia que possa “limpar” o efeito dessas influências e obter o efeito líquido do programa

METODOLOGIA E RESULTADOS

IV – Avaliação de Impacto:

- a. Qual a técnica utilizada? Qual a justificativa para utilização desta técnica?
- b. Quais foram os resultados encontrados? O programa teve impacto?

- Sabemos que para selecionar a metodologia a ser utilizada é necessário verificar como os indivíduos foram selecionados para participar do programa
- Neste caso temos pelo menos três possibilidades: se o programa foi aleatorizado podemos i) usar o teste de médias. Caso o programa não tiver sido sorteado, teremos mais duas possibilidades: ii) utilizar variáveis observáveis para o pesquisador, e estimar uma regressão (ou usar pareamento); e se acreditarmos que variáveis não observáveis influenciaram o indicador de interesse, e além disso, tais variáveis forem diferentes entre o grupo de tratamento e controle, poderemos iii) empregar a metodologia de diferenças em diferenças tendo informações antes e depois do programa
- Como o programa não foi aleatorizado e a própria ONG coletou os dados nos dois períodos de tempo, antes e após, podemos testar a robustez dos resultados e utilizar ambas as metodologias de avaliação: regressão em variáveis observáveis e regressão utilizando diferenças em diferenças

Primeiro faremos uma análise de regressão após o tratamento

verde_novo.xlsx [Modo de Compatibilidade] - Microsoft Excel

Arquivo Página Inicial Inserir Layout da Página Fórmulas Dados Revisão Exibição Acrobat

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1		areas	indicador	programa	cercamento	agricultura	fogo	braquiaria	animais	dist_centro_urbanc	renda			
92		91	0	0	0	1	1	1	92	3	149			
93		92	0	0	1	1	1	1	90	3	150			
94		93	0	0	1	1	1	1	90	1	130			
95		94	0	0	0	1	1	1	90	3	120			
96		95	1	0	0	1	1	1						
97		96	0	0	0	1	1	1						
98		97	37	0	1	0	0	1						
99		98	2	0	1	1	1	1						
100		99	10	0	1	0	0	1						
101		100	0	0	0	1	1	1						
102														
103		Média	77.24	1.00	0.60	0.26	0.30	0.60						
104	T	DP	25.71	0.00	0.49	0.44	0.46	0.49	11.25	6.77	37.32			
105	T	n	50.00	51.00	52.00	53.00	54.00	55.00	56.00	57.00	58.00			
106	T	IC	7.13	#NÚM!	0.13	0.12	0.12	0.13	2.95	1.76	9.60			
107	T	Inf	70.11	#NÚM!	0.47	0.14	0.18	0.47	11.75	9.42	255.82			
108	T	Sup	84.37	#NÚM!	0.73	0.38	0.42	0.73	17.65	12.94	275.02			
109	C	Média	11.24	0.00	0.66	0.46	0.66	0.78	55.84	4.04	170.58			
110	C	DP	18.12	0.00	0.48	0.50	0.48	0.42	35.03	3.42	40.71			
111	C	n	50.00	51.00	52.00	53.00	54.00	55.00	56.00	57.00	58.00			
112	C	IC	5.02	#NÚM!	0.13	0.14	0.13	0.11	9.18	0.89	10.48			
113	C	Inf	6.22	#NÚM!	0.53	0.32	0.53	0.67	46.66	3.15	160.10			

Análise de dados

- Ferramentas de análise
- Ajuste exponencial
- Teste-F: duas amostras para variâncias
- Análise de Fourier
- Histograma
- Média móvel
- Geração de número aleatório
- Ordem e percentil
- Regressão
- Amostragem
- Teste-T: duas amostras em par para médias

OK Cancelar Ajuda

verde_novo.xlsx [Modo de Compatibilidade] - Microsoft Excel

Arquivo Página Inicial Inserir Layout da Página Fórmulas Dados Revisão Exibição Acrobat

Colar Arial 10 Quebrar Texto Automaticamente Geral

Área de Tran... Fonte Alinhamento

Para saber quais variáveis foram significativas, basta verificar se o número zero está contido no intervalo de confiança

Podemos dizer que três variáveis foram estatisticamente significativas: programa, distância e renda per capita

	gl	SQ	MQ	F	le significação
12 Regressão	8	151274.7	18909.34	282.2055	9.24E-61
13 Resíduo	91	6097.507	67.00558		
14 Total	99	157372.2			

	Coefficiente	Erro padrã	Stat t	valor-P	5% inferior	% superior	inferior 95.0%	superior 95.0%
17 Interseção	-33.077	9.066	-3.648	0.000	-51.085	-15.068	-51.085	-15.068
18 programa	28.251	3.212	8.795	0.000	21.870	34.632	21.870	34.632
19 cercamento	1.202	2.401	0.501	0.618	-3.567	5.970	-3.567	5.970
20 agricultura	-1.503	2.316	-0.649	0.518	-6.104	3.098	-6.104	3.098
21 fogo	-1.972	3.171	-0.622	0.536	-8.270	4.327	-8.270	4.327
22 braquiaria	-2.992	2.593	-1.154	0.252	-8.143	2.159	-8.143	2.159
23 animais	-0.005	0.047	-0.107	0.915	-0.098	0.088	-0.098	0.088
24 dist_centro_urba	1.902	0.266	7.142	0.000	1.373	2.430	1.373	2.430
25 renda	0.237	0.042	5.667	0.000	0.154	0.320	0.154	0.320

2000 reg1 2010

Pronto 130%

Para estimar o resultado pela técnica das diferenças em diferenças primeiro precisamos consolidas as informações de “antes” e “depois” em uma única planilha. A seguir, devemos criar as variáveis que indicam “tempo” e “tempo x programa”.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	areas	indicador	programa	percament	agricultura	fogo	braquiaria	animais	dist_centro	renda	t	programa_t		
2	1	61	1	1	1	0	1	2	20	84.58	0	0		
3	2	10	1	0	1	1	1	25	10	16.42	0	0		
4	3	10	1	0	1	1	1	27	6	15.42	0	0		
5	4	2	1	0	1	1	1	15	11	4.58	0	0		
6	5	15	1	0	1	1	1	15	18	24.58	0	0		
7	6	14	1	0	1	0	1	2	20	34.58	0	0		
8	7	13	1	0	1	0	1	5	20	54.58	0	0		
9	8	28	1	0	0	0	0	3	18	36.58	0	0		
10	9	18	1	1	0	1	1	9	19	25.58	0	0		
11	10	14	1	1	0	1	1	14	20	22.58	0	0		
12	11	6	1	0	1	1	1	10	12	0.58	0	0		
13	12	25	1	1	1	0	0	3	20	33.58	0	0		
14	13	18	1	0	0	1	1	13	18	26.58	0	0		
15	14	13	1	0	1	1	1	27	5	15.42	0	0		
16	15	22	1	1	0	0	0	1	20	33.58	0	0		
17	16	47	1	0	0	0	0	2	21	76.58	0	0		
18	17	4	1	0	0	1	1	20	12	20	0	0		
19	18	31	1	0	1	1	1	30	2	35.42	0	0		
20	19	14	1	0	1	1	1	27	3	43.42	0	0		
21	20	3	1	1	0	1	1	35	9	5.42	0	0		
22	21	19	1	0	0	1	1	27	5	16.42	0	0		
23	22	26	1	0	1	1	1	25	5	26.42	0	0		

O passo seguinte é repetir a regressão usando as informações consolidadas e acrescentar as variáveis “tempo” e “tempo x programa”.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	areas	indicador	programa	mercament	agricultura	fogo	braquiaria	animais	dist_centro	renda	t	programa_t		
181	80	3	0	1	1	1	1							
182	81	1	0	1	1	1	1							
183	82	0	0	0	1	1	1							
184	83	0	0	0	0	1	1							
185	84	0	0	1	0	1	1							
186	85	0	0	1	0	1	1							
187	86	0	0	0	0	1	1							
188	87	0	0	0	0	1	1							
189	88	0	0	0	1	1	1							
190	89	0	0	0	0	1	1							
191	90	0	0	0	0	1	1							
192	91	0	0	0	1	1	1							
193	92	0	0	1	1	1	1							
194	93	0	0	1	1	1	1							
195	94	0	0	0	1	1	1							
196	95	1	0	0	1	1	1	98	3	139				
197	96	0	0	0	1	1	1	99	2	135	1			
198	97	37	0	1	0	0	1	25	12	217	1			
199	98	2	0	1	1	1	1	99	2	160	1			
200	99	10	0	1	0	0	1	80	5	199	1			
201	100	0	0	0	1	1	1	99	1	137	1			
202														

verde_novo.xlsx [Modo de Compatibilidade] - Microsoft Excel

Arquivo Página Inicial Inserir Layout da Página Fórmulas

Colar Área de Tran... Fonte

N25 fx

A B C D

2

3 *Estatística de regressão*

4 R múltiplo 0.94246

5 R-Quadrado 0.88824

6 R-quadrado ajustac 0.88233

7 Erro padrão 11.2953

8 Observações 200

9

10 ANOVA

	gl	SQ	MQ	F	e significação
12 Regressão	10	191643	19164.3	150.21	2.7E-84
13 Resíduo	189	24113.2	127.583		
14 Total	199	215756			

15

	Coefficiente	Erro padrã	Stat t	valor-P	% inferior	% superior	inferior 95.0%	superior 95.0%
17 Interseção	13.604	4.164	3.267	0.001	5.391	21.818	5.391	21.818
18 programa	-3.731	2.993	-1.247	0.214	-9.635	2.173	-9.635	2.173
19 cercamento	-0.876	2.061	-0.425	0.671	-4.942	3.189	-4.942	3.189
20 agricultura	0.168	1.934	0.087	0.931	-3.646	3.982	-3.646	3.982
21 fogo	-2.618	2.726	-0.961	0.338	-7.996	2.759	-7.996	2.759
22 braquiaria	-2.394	2.615	-0.915	0.361	-7.552	2.765	-7.552	2.765
23 animais	-0.106	0.041	-2.550	0.012	-0.187	-0.024	-0.187	-0.024
24 dist_centro_urbano	0.515	0.214	2.414	0.017	0.094	0.937	0.094	0.937
25 renda	0.352	0.034	10.454	0.000	0.285	0.418	0.285	0.418
26 t	-54.476	5.107	-10.666	0.000	-64.551	-44.401	-64.551	-44.401
27 programa_t	26.947	4.396	6.130	0.000	18.276	35.618	18.276	35.618

28

reg1 2000 2010 reg2 tudo

Pronto

120%

Ao analisar os resultados percebemos que as variáveis estatisticamente significantes, em que o zero não está no intervalo de confiança, são: animais, distância, renda, depois e a interação entre esta e a variável programa

Assim, ao retirar o efeito de fatores não observados, temos que para as áreas que participaram do programa, o número de árvores por 1.000m² é, em média, 26,95 unidades maior do que na área de controle.

CONSIDERAÇÕES FINAIS

- Os técnicos perceberam outros resultados que não foram medidos utilizando a regressão, mas foram observados no meio ambiente. Estes resultados são caracterizados como externalidades
- De modo geral, podemos dizer que existe uma externalidade quando a ação de um indivíduo afeta o bem-estar de outro(s) indivíduo(s). Como exemplo, podemos ter uma externalidade negativa quando a fumaça de cigarro afeta o bem-estar de um não fumante. Teremos uma externalidade positiva quando, por exemplo, um parque ambiental é reformado e proporciona melhora de bem-estar à população
- No caso do programa Verde Novo a externalidade, obviamente, é positiva, dado que foi possível verificar progressos ambientais além da quantidade de árvores
- Quais progressos foram estes? Foram percebidas as seguintes externalidades: redução da temperatura local, aumento do fluxo de água e redução da erosão
- Assim, a ação da ONG afetou, positivamente, o bem-estar da população através do melhor desenvolvimento ambiental

ÁREA DE TRATAMENTO NA REGIÃO DO VALE DO JEQUITINHONHA



EM 2000



EM 2010

SUMÁRIO DO CURSO

- ~~1. Conceitos básicos em avaliação de impacto~~
- ~~2. Revisão de estatística e uso avançado de planilhas eletrônicas~~
- ~~3. Modelo de Resultados Potenciais~~
- ~~4. Econometria: revisão do modelo de mínimos quadrados~~
- ~~5. Desenho de uma avaliação para uma política~~
- ~~6. Modelo de Diferenças em Diferenças~~
- ~~7. Técnicas de Pareamento~~
- ~~8. Estudo de caso e aplicações de diferentes técnicas econométricas~~
- 9. Discussão sobre os trabalhos de conclusão de curso**
- 10. Revisão de matemática financeira**
11. O Cálculo do Retorno Econômico
12. Conclusão do curso: apresentação dos trabalhos

TAXA DE JUROS

- A taxa de juros é a taxa que determina o valor dos juros, isto é, da remuneração que um indivíduo recebe por realizar algum investimento durante um certo período de tempo
- Outra forma de definir Juros: “prêmio” recebido por adiar consumo presente
- A taxa de juros i pode se referir a um período de um ano, mês ou em relação a um período qualquer
- Por exemplo, se investimos R\$ 100 por um ano e obtemos R\$ 110 após esse tempo, a taxa de juros é de 10% ao ano. Como fazemos esse cálculo?

VALOR PRESENTE E VALOR FUTURO

- ✓ “Valor Presente” = valor investido hoje ou o capital inicial [VP]
- ✓ “Valor Futuro” = o valor a ser recebido após o período do investimento [VF]

$$i\% = \left(\frac{(VF - VP)}{VP} \right) * 100 = \left(\frac{VF}{VP} - 1 \right) * 100$$

- ✓ Voltando ao exemplo do investimento de R\$ 100 por um ano e obtemos R\$ 110 após esse tempo, a taxa de juros é de 10% ao ano

$$i\% = \left(\frac{110}{100} - 1 \right) * 100 = 10\%$$

GENERALIZANDO

- ✓ Se realizamos um investimento hoje por n períodos, sendo a taxa de juros definida em $i\%$ por período, teremos ao final um total de:

$$VF = VP \times (1+i)^n$$

Diagram illustrating the components of the future value formula:

- Valor futuro (Future Value) points to VF
- Número de períodos (Number of periods) points to n
- Valor Presente (Present Value) points to VP
- Taxa de juros (Interest rate) points to i

Ex. 1. Calcular o valor futuro de um capital de \$ 5.000 aplicado à taxa de 4% ao mês, durante 5 meses

The image shows a screenshot of a Microsoft Excel spreadsheet. The spreadsheet has a grid with columns labeled A through H and rows numbered 1 through 10. The formula bar at the top shows the formula $=5000*(1+4\%)^5$. The spreadsheet content is as follows:

	A	B	C	D	E	F	G	H
1	1. Calcular o valor futuro de um capital de \$ 5.000 aplicado à taxa de 4% ao mês, dur							
2	$=5000*(1+4\%)^5$							
3	2. Qual é o juro obtido através da aplicação de capital de \$2.500 a 7% ao ano durante							
4								
5	3. Calcular o valor futuro ou montante de uma aplicação financeira de \$15.000, admi							
6	uma taxa de 30% a.a. para um período de 18 meses.							
7								
8	4. Calcular o valor presente ou capital de uma aplicação que gerou um montante de \$							
9	efetuada pelo prazo de 6 meses a uma taxa de 1,85% ao mês.							
10								

Ex. 2. Qual é o juro obtido através da aplicação de capital de \$2.500 a 7% ao ano durante 3 anos?

FISHER =2500*(1+7%)^3-2500

	A	B	C	D	E	F	G
1	1. Calcular o valor futuro de um capital de \$ 5.000 aplicado à taxa de 4% ao mês, dur						
2	R\$ 6,083.26						
3	2. Qual é o juro obtido através da aplicação de capital de \$2.500 a 7% ao ano durante						
4	=2500*(1+7%)^3-2500						
5	3. Calcular o valor futuro ou montante de uma aplicação financeira de \$15.000, admi						
6	uma taxa de 30% a.a. para um período de 18 meses.						
7							
8	4. Calcular o valor presente ou capital de uma aplicação que gerou um montante de \$						
9	efetuada pelo prazo de 6 meses a uma taxa de 1,85% ao mês.						
10							

VP, VF, taxa de juros VPL TIR LEMBRETES 260%

Ex. 3. Calcular o valor futuro ou montante de uma aplicação financeira de \$15.000, admitindo-se uma taxa de 30% a.a. para um período de 18 meses

	A	B	C	D	E	F	G
1	1. Calcular o valor futuro de um capital de \$ 5.000 aplicado à taxa de 4% ao mês, durante						
2	R\$ 6,083.26						
3	2. Qual é o juro obtido através da aplicação de capital de \$2.500 a 7% ao ano durante						
4	R\$ 562.61						
5	3. Calcular o valor futuro ou montante de uma aplicação financeira de \$15.000, admitindo						
6	uma taxa de 30% a.a. para um período de 18 meses.						
7	<u>=15000*(1+30%)^(18/12)</u>						
8	4. Calcular o valor presente ou capital de uma aplicação que gerou um montante de \$						
9	efetuada pelo prazo de 6 meses a uma taxa de 1,85% ao mês.						
10							

Ex. 4. Calcular o valor presente ou capital de uma aplicação que gerou um montante de R\$ 98.562,25, efetuada pelo prazo de 6 meses a uma taxa de 1,85% ao mês.

	A	B	C	D	E	F	G
6	uma taxa de 30% a.a. para um período de 18 meses.						
7	R\$ 22,233.42						
8	4. Calcular o valor presente ou capital de uma aplicação que gerou um montante de \$						
9	efetuada pelo prazo de 6 meses a uma taxa de 1,85% ao mês.						
10	$=98562.25/(1+1.85\%)^6$						
11	5. Um imóvel foi adquirido por R\$ 3.000,00 em uma determinada data, sendo vendido						
12	quatro anos depois. Qual o retorno obtido por este proprietário?						
13							
14	6. Para um poupador que deseja ganhar 2,5% ao mês, o que é mais interessante:						
15	a) receber \$18.500 de hoje a 4 meses; ou b) \$25.500 de hoje a doze meses?						

Ex. 5. Um imóvel foi adquirido por R\$ 3.000,00 em uma determinada data, sendo vendido por R\$ 30.000 quatro anos depois. Qual o retorno (taxa de juros) obtido por este proprietário?

	A	B	C	D	E	F	G
6	uma taxa de 30% a.a. para um período de 18 meses.						
7	R\$ 22,233.42						
8	4. Calcular o valor presente ou capital de uma aplicação que gerou um montante de \$						
9	efetuada pelo prazo de 6 meses a uma taxa de 1,85% ao mês.						
10	R\$ 88,296.69						
11	5. Um imóvel foi adquirido por R\$ 3.000,00 em uma determinada data, sendo vendido						
12	quatro anos depois. Qual o retorno obtido por este proprietário?						
13	$= (30000/3000)^{(1/4)} - 1$						
14	6. Para um poupador que deseja ganhar 2,5% ao mês, o que é mais interessante:						
15	a) receber \$18.500 de hoje a 4 meses; ou b) \$25.500 de hoje a doze meses?						

Ex. 6. Para um poupador que deseja ganhar 2,5% ao mês, o que é mais interessante: a) receber \$18.500 de hoje a 4 meses; ou b) \$25.500 de hoje a doze meses?

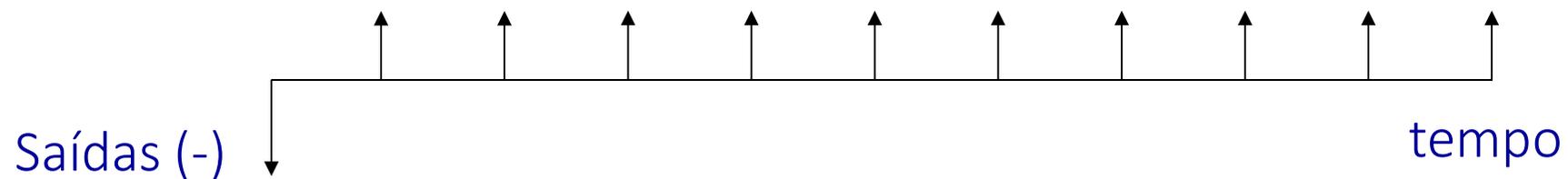
	A	B	C	D	E
60					
61	Quanto ele terá que investir em cada situação?				
62	r	2.5%			
63		opção 1		opção 2	
64	VF	R\$ 18,500.00		R\$ 25,500.00	
65	n		4	12	
66	VP	R\$ 16,760.09		=D64/((1+B62)^D65)	
67					
68	opção 1				
69					

FLUXO DE CAIXA

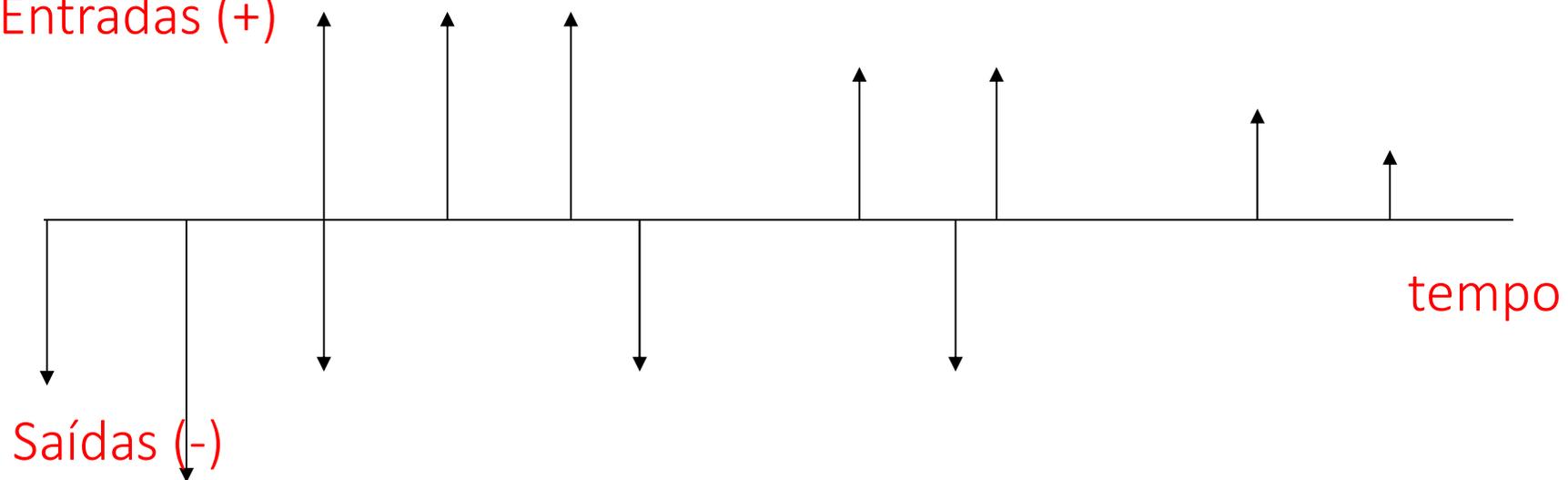
- É a representação dos pagamentos e recebimentos ao longo do tempo
- Os fluxos de caixa podem diferir com relação a vários aspectos:
 - Quanto ao tempo:
 - ✓ Temporária
 - ✓ Infinita
 - Quanto à constância ou periodicidade:
 - ✓ Periódicas
 - ✓ Não periódicas
 - Quanto ao valor dos pagamentos:
 - ✓ Fixos ou uniformes
 - ✓ Variáveis

EXEMPLOS DE FLUXOS DE CAIXA

Entradas (+)



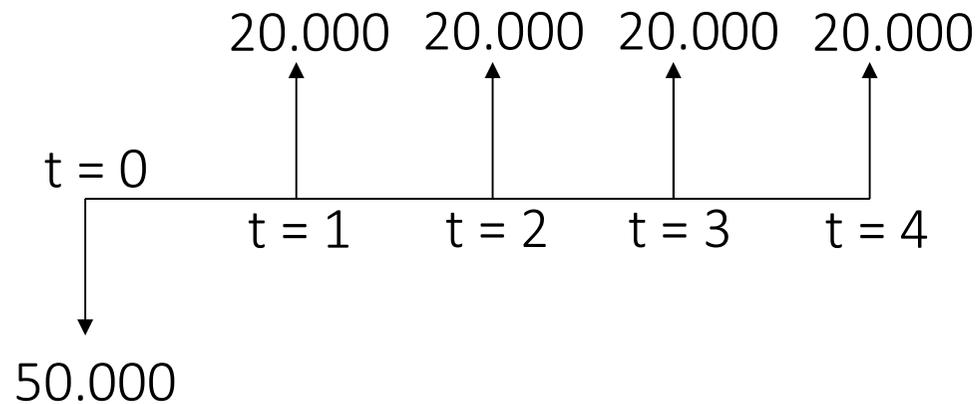
Entradas (+)



VALOR PRESENTE LÍQUIDO (VPL)

- Os valores dos recebimentos e pagamentos de um projeto estão distribuídos ao longo do tempo
- Esses valores, no entanto, **não** podem ser prontamente comparados visto que **estão avaliados em momentos diferentes no tempo**. “Receber R\$ 100 hoje é, provavelmente, diferente de receber R\$ 100 daqui a 3 meses”
- Como então comparar valores? A ideia é “trazer” esses valores todos para uma mesma data, por exemplo, para a data relativa ao início do projeto. Em outros termos, calcular o “valor presente” do investimento
- “Líquido” porque trazemos para valor presente recebimentos e pagamentos
- Da mesma maneira que acrescentamos juros quando queremos saber o VF de um investimento realizado hoje, quando calculamos o valor presente desse retorno futuro do investimento temos que descontar os juros
- Quando temos um fluxo de pagamentos e recebimentos, precisamos trazer para a mesma data cada uma das parcelas: se trouxermos todos para o 1º período, temos o VPL

Considere a série de pagamentos e recebimentos abaixo:



$$VPL = -50.000 + \frac{20.000}{(1+i)^1} + \frac{20.000}{(1+i)^2} + \frac{20.000}{(1+i)^3} + \frac{20.000}{(1+i)^4}$$

Outro exemplo

- ✓ Os custos de um projeto estão distribuídos ao longo dos dois primeiros anos do projeto. Por outro lado, há recebimentos por 3 anos consecutivos, conforme tabela abaixo:

Ano	2006	2007	2008	2009	2010
Valor (\$)	-50.000	-10.000	30.000	30.000	30.000

- ✓ Qual é o valor presente líquido do projeto ($i=15\%$ ao ano)?

$$VPL = -50000 - \frac{10000}{1,15} + \frac{30000}{(1,15)^2} + \frac{30000}{(1,15)^3} + \frac{30000}{(1,15)^4} = 866,74$$

Ex. 1. Um projeto de investimento inicial de \$120 mil gera entradas de caixa anuais de R\$ 25 mil. Nos próximos 10 anos, em cada ano será necessário um gasto de R\$ 5 mil para manutenção. Determinar o fluxo de caixa e o valor presente líquido da operação (i = 5% a.a.)

	A	B	C	D	E	F	G
4							
5	tempo	saidas	entradas	liquido	VPL		
6	0	-120000	0	-120000	=VPL(0.05;D7:D16)+D6		
7	1	-5000	25000	20000	VPL(taxa; valor1; [valor2]; [valor3]; ...)		
8	2	-5000	25000	20000			
9	3	-5000	25000	20000			
10	4	-5000	25000	20000			
11	5	-5000	25000	20000			
12	6	-5000	25000	20000			
13	7	-5000	25000	20000			
14	8	-5000	25000	20000			
15	9	-5000	25000	20000			
16	10	-5000	25000	20000			
17							

Ex. 2. Um investimento inicial de \$ 200 mil com entradas anuais de \$ 300 mil nos próximos 10 anos, no final do 10º ano terá o ativo vendido por \$ 50 mil. As saídas de caixa devem ser de R\$ 20 mil, exceto no 6º ano quando uma reforma exigirá uma saída de caixa complementar de R\$ 500 mil. Determinar o fluxo de caixa e o VPL da operação (considerar $r=5\%$)

	A	B	C	D	E	F	G	H
24	taxa de juros	5%						
25	tempo	custo	entradas	beneficio				
26	0	-200000	0	-200000				
27	1	-20000	300000	280000		=VPL(B24;D27:D36)+D26		
28	2	-20000	300000	280000				
29	3	-20000	300000	280000				
30	4	-20000	300000	280000				
31	5	-20000	300000	280000				
32	6	-520000	300000	-220000				
33	7	-20000	300000	280000				
34	8	-20000	300000	280000				
35	9	-20000	300000	280000				
36	10	-20000	350000	330000				

TAXA INTERNA DE RETORNO (TIR)

- ✓ É a taxa de juros (ou desconto) que faz com que o VPL de um projeto seja zero
- ✓ VPL igual a zero significa que os custos avaliados no tempo zero são iguais aos benefícios, também avaliados no tempo zero
- ✓ Idéia: você quer descobrir qual a taxa de juros que está embutida no seu investimento, isto é, qual a taxa de juros que a partir do seu investimento inicial gerou as entradas futuras do seu fluxo de caixa

Retomando o exemplo anterior ($i = ? \%$) :

Ano	2005	2006	2007	2008	2009
Valor (\$)	-50.000	-10.000	30.000	30.000	30.000

$$VPL = -50.000 - \frac{10.000}{(1 + TIR)} + \frac{30.000}{(1 + TIR)^2} + \frac{30.000}{(1 + TIR)^3} + \frac{30.000}{(1 + TIR)^4} = 0$$

- ✓ Como o fluxo está em anos, a TIR calculada será uma taxa anual.
- ✓ Para fazer o cálculo, vamos utilizar o Excel – o cálculo “algébrico” só pode ser feito por aproximação.

Ex. 1. Um projeto de investimento inicial de \$ 70 mil gera entradas de caixa de R\$ 25 mil nos próximos 5 anos. Em cada ano será necessário um gasto de \$ 5 mil para manutenção, considerando um custo de oportunidade de 8% a.a. Determinar o fluxo de caixa , o VPL e a TIR da operação.

	A	B	C	D	E	F
4						
5		inv. inicial	beneficios	custos	líquido	
6	0	-70000			-70000	
7	1		25000	-5000	20000	
8	2		25000	-5000	20000	
9	3		25000	-5000	20000	
10	4		25000	-5000	20000	
11	5		25000	-5000	20000	
12						
13				VPL	R\$ 9,854.20	
14						
15						
16				TIR	=TIR(E6:E11)	
17						

aula9_-_exercicios_mat_financ_resolvido.xls [Modo de Compatibilidade] - Microsoft Excel

Arquivo Página Inicial Inserir Layout da Página Fórmulas Dados Revisão Exibição Acrobat

Colar Arial 10 Quebrar Texto Automaticamente Porcentagem

Área de Tran... Fonte Alinhamento Número Estilo Células Edição

FISHER X ✓ f_x =TIR(D28:D31)

	A	B	C	D	E	F
26						
27		custos	beneficios	liquido		TIR
28	0	47000		-47000		=TIR(D28:D31)
29	1		12000	12000		TIR(valores; [estimativa])
30	2		15000	15000		
31	3		23000	23000		
32						
33						
34						
35						
36						

VP, VF, taxa de juros VPL TIR LEMBRETES

Edita 270%

Ex. 2. 2) Uma aplicação financeira envolve uma saída de caixa de R\$ 47.000 no momento inicial, e os seguintes benefícios esperados de caixa ao final dos três meses posteriores: R\$ 12.000; R\$ 15.000 e R\$ 23.000. Determinar a taxa interna de retorno deste investimento.

SUMÁRIO DO CURSO

- ~~1. Conceitos básicos em avaliação de impacto~~
- ~~2. Revisão de estatística e uso avançado de planilhas eletrônicas~~
- ~~3. Modelo de Resultados Potenciais~~
- ~~4. Econometria: revisão do modelo de mínimos quadrados~~
- ~~5. Desenho de uma avaliação para uma política~~
- ~~6. Modelo de Diferenças em Diferenças~~
- ~~7. Técnicas de Pareamento~~
- ~~8. Estudo de caso e aplicações de diferentes técnicas econométricas~~
- ~~9. Discussão sobre os trabalhos de conclusão de curso~~
- ~~10. Revisão de matemática financeira~~
11. O Cálculo do Retorno Econômico
12. Conclusão do curso: apresentação dos trabalhos

BENEFÍCIOS DO PROJETO SOCIAL EM VALORES MONETÁRIOS

- O benefício do projeto social é o valor monetário do seu impacto
- Portanto, a avaliação de impacto antecede a avaliação de retorno econômico e somente se realiza quando existe evidencia que o projeto gerou resultados
- A transformação do impacto estimado em benefício é uma das maiores dificuldades da avaliação de retorno econômico. Ela exige a associação de um valor monetário a um impacto, muitas vezes, não monetário
- A única exceção são os projetos para geração de renda, em que o indicador de impacto é a renda, ou seja, já é estimado em valor monetário, não exigindo nenhuma transformação
- Para os demais projetos sociais, que não tenham como indicador de impacto a renda dos participantes, a transformação do impacto em benefício precisa ser realizada

BENEFÍCIOS DO PROJETO EM VALORES MONETÁRIOS: DESAFIO

- A pergunta que se faz é como efetuar essa transformação em projetos sociais com objetivos tão diversos, como, por exemplo, a melhora na qualidade da educação, na saúde, a redução da criminalidade, entre outros
- A resposta não é simples. Para efetuarmos essas transformações, temos que fazer hipóteses que permitam atribuir um valor monetário ao impacto. Essas hipóteses variam de projeto para projeto e exigem um profundo conhecimento sobre a temática à qual se aplica
- O avaliador precisa conhecer a literatura a respeito do indicador de impacto. Ele precisa saber quanto a mudança causada no indicador, pelo programa, economiza para a sociedade e quanto aumenta o bem-estar, devido à contribuição advinda das mudanças geradas nos beneficiários e nas circunstâncias que os envolvem

EXEMPLO: INDICADOR DE IMPACTO LIGADO A ESCOLARIDADE

- Suponha um programa para o aumento da escolaridade, no qual o indicador de impacto é a taxa de jovens com ensino superior completo
- A literatura mostra que escolaridade mais alta propicia salários maiores, menor probabilidade de acessar programas sociais, menor probabilidade de encarceramento, entre outros efeitos
- Na transformação do impacto em benefício o avaliador tem que considerar:
 - os ganhos salariais devido ao aumento da escolaridade;
 - a probabilidade daqueles jovens tratados acessarem programas sociais e a média dos custos em prover esses programas que seriam acessados caso o jovem não tivesse aumentado sua escolaridade – isso equivale a uma economia para a sociedade;
 - a probabilidade de os jovens tratados serem presos e o custo dessas prisões que teriam acontecido se a escolaridade não tivesse sido aumentada – é o valor que a sociedade deixou de gastar, também uma economia

EXEMPLO: INDICADOR DE IMPACTO LIGADO A SAÚDE

- No caso de projetos na área de saúde, por exemplo, para prevenir a incidência de diabetes, a transformação do impacto em benefício é também baseada em uma série de hipóteses
- Digamos que o impacto do projeto seja mensurado em termos da diminuição da taxa de incidência de diabetes na população
- Uma menor taxa de incidência de diabetes leva a:
 - uma diminuição dos gastos com o sistema de saúde para tratamento das doenças derivadas dessa;
 - uma melhora na vida das pessoas que adquiriram hábitos saudáveis por causa do projeto e que ficariam doentes caso ele não existisse;
 - economia para a sociedade com os dias de trabalho que seriam perdidos se as pessoas ficassem doentes;
 - diminuição dos gastos com previdência social que teria que ser paga aos doentes e inválidos

QUANTO AS HIPÓTESES NECESSÁRIAS

- Para determinar o valor ganho ou economizado pela sociedade devido ao projeto o avaliador terá que conhecer, profundamente, a literatura
- Só assim será capaz de estimar as várias facetas afetadas pelo projeto e adotar boas hipóteses para a transformação
- O avaliador terá que considerar na estimativa do benefício do projeto é que o impacto, em geral, perdura mais do que um período de tempo
- Algumas vezes o benefício perdura por toda a vida produtiva do indivíduo, como no caso dos projetos de qualificação profissional; ou por séculos, como o caso de projetos de reflorestamento
- Outras vezes, os benefícios são grandes logo após a implementação e reduzem-se ao longo do tempo, por exemplo, campanhas de prevenção a doenças epidemiológicas como a dengue
- Existem também projetos que só geram efeitos muitos anos após sua implementação, como aqueles para o aumento da escolaridade
- O que todos esses casos têm em comum é que os benefícios ocorrem em momentos distintos no tempo

O PAPEL DA MATEMÁTICA FINANCEIRA

- Se os benefícios são recebidos em momentos diferentes, é preciso considerar o tempo no cálculo do benefício total. Para isso, montamos o fluxo de caixa do benefício e utilizamos a taxa de desconto intertemporal para deduzir do montante o fato de que o valor só será recebido no futuro
- A taxa de desconto intertemporal indica quanto o indivíduo valoriza o consumo presente em relação ao futuro, ou seja, é o percentual pelo qual ele está disposto a abrir mão do consumo no presente para só fazê-lo no futuro
- Para efetuar o desconto intertemporal aplicamos, então, os conceitos de “valor presente” e “valor futuro” oriundos da matemática financeira
- Muitas vezes, os avaliadores adotam taxas de desconto intertemporal menores que a taxa de juros corrente. Isso acontece, porque, como a maioria dos projetos tem impactos de longo prazo, os avaliadores utilizam uma taxa de juros de longo prazo. A taxa de juros de longo prazo, em geral, é menor que a taxa de juros corrente

CUSTO DO PROJETO

- Uma vez calculado o benefício total, a outra dimensão necessária para a avaliação de retorno econômico é o custo do projeto
- O custo considerado nesse tipo de análise é denominado “custo econômico” e transcende o custo contábil por considerar o custo de oportunidade

Custo econômico = custo contábil + custo de oportunidade

- O custo contábil é todo dispêndio que, via de regra, aparece nos demonstrativos contábeis. É todo desembolso monetário necessário para o funcionamento do projeto, por exemplo, pagamento de conta de luz, de pessoal, aquisição de material etc.
- O custo de oportunidade difere do custo contábil por ser um custo implícito, ou seja, em geral não aparece nos demonstrativos contábeis. O conceito do custo de oportunidade foi cunhado pela economia para designar o rendimento que se deixa de obter ao realizar determinada escolha. Em outras palavras, é o ganho que poderia advir da aplicação alternativa de um determinado recurso

CUSTO DE OPORTUNIDADE

- Muitos projetos funcionam com um exíguo orçamento próprio, mas utilizam muitos recursos extras da sociedade. Por exemplo, um programa de fornecimento de sopa para pessoas carentes que utiliza mão de obra voluntária, que recebe os alimentos de um doador, que utiliza a cozinha cedida de uma instituição beneficente, entre outros recursos
- O custo contábil provavelmente será muito baixo e se somente ele fosse computado no cálculo do retorno, esse estaria distorcido. O custo de oportunidade resolve o problema da distorção do retorno, uma vez que computa quanto valeria: a aplicação da mão de obra em outra atividade produtiva; os alimentos doados se fossem aplicados de outra maneira, por exemplo, vendidos; o aluguel da cozinha para um restaurante
- Nesse caso específico, o custo econômico do projeto tem como maior parcela o custo de oportunidade. Ao utilizá-lo para o cálculo do retorno econômico, esse não mais estaria distorcido

MEDIDAS DO RETORNO ECONÔMICO

- A viabilidade econômica se refere ao fato de o projeto gerar um “lucro social”, ou seja, seu benefício ser maior que seu custo para a sociedade
- A comparação de dois projetos indica que aquele com maior retorno é o mais atrativo como investimento. Mas é preciso ter cuidado com essa comparação
- Existem várias formas de calcularmos o retorno e cada uma delas constitui uma estatística de retorno econômico
- A diferença entre as estatísticas é apenas a ótica da sua construção, pois todas medem a relação entre o custo e o benefício do projeto.
- Aqui apresentamos quatro estatísticas de retorno econômico:
 - ✓ Valor Presente Líquido (VPL)
 - ✓ Taxa Interna de Retorno (TIR)
 - ✓ Razão Custo-Benefício
 - ✓ Razão Custo-Efetividade

VALOR PRESENTE LÍQUIDO – VPL

- Considerando que os benefícios e os custos acontecem em momentos diferentes no tempo, para realizarmos essa subtração, precisamos que ambos estejam a valor presente de um mesmo período. Considerando t_0 o primeiro ano do investimento, trazemos os valores do benefício e do custo de cada ano a valor presente desse período, por meio da fórmula do VPL:

$$VPL_{t_0} = \left(VB_{t_0} + \sum \frac{VFB_{t_n}}{(1+i)^n} \right) - \left(VC_{t_0} + \sum \frac{VFC_{t_n}}{(1+i)^n} \right)$$

- Se $VPL > 0$ – retorno positivo, o projeto é viável economicamente, pois seu benefício supera o custo
- Se $VPL < 0$ – retorno negativo, o projeto não é viável economicamente, pois seu custo supera o benefício gerado
- Se $VPL = 0$ – retorno neutro, valor do benefício gerado é igual ao custo incorrido

VALOR PRESENTE LÍQUIDO – VPL

- Considerando que os benefícios e os custos acontecem em momentos diferentes no tempo, para realizarmos essa subtração, precisamos que ambos estejam a valor presente de um mesmo período. Considerando t_0 o primeiro ano do investimento, trazemos os valores do benefício e do custo de cada ano a valor presente desse período, por meio da fórmula do VPL:

$$VPL_{t_0} = \left(VB_{t_0} + \sum \frac{VFB_{t_n}}{(1+i)^n} \right) - \left(VC_{t_0} + \sum \frac{VFC_{t_n}}{(1+i)^n} \right)$$

VPL dos benefícios **VPL dos custos**

- Se $VPL > 0$ – retorno positivo, o projeto é viável economicamente, pois seu benefício supera o custo
- Se $VPL < 0$ – retorno negativo, o projeto não é viável economicamente, pois seu custo supera o benefício gerado
- Se $VPL = 0$ – retorno neutro, valor do benefício gerado é igual ao custo incorrido

TAXA INTERNA DE RETORNO - TIR

- A taxa interna de retorno é a taxa de oportunidade de investir o recurso no projeto. Ela representa o retorno implícito do investimento. Matematicamente, a TIR é a taxa de desconto intertemporal implícita que iguala o benefício do projeto ao seu custo, de forma que o valor presente líquido seja zero
- Para interpretarmos o resultado da TIR, é necessária a comparação com a TIR de outro projeto ou com uma taxa mínima de atratividade - TMA. A taxa mínima de atratividade é a taxa de juros mínima que um investidor pretende obter com o seu investimento. Assim a TMA pode ser a taxa de juros básica do mercado, a taxa de remuneração da poupança ou outra taxa que se considerar a mínima para remunerar um investimento.
- São três os resultados possíveis:
 - $TIR > TMA$ – o projeto é viável economicamente, pois o retorno gerado é maior que a taxa mínima de atratividade
 - $TIR < TMA$, o projeto não é viável economicamente, porque seu retorno é menor que a mínima de atratividade
 - $TIR = TMA$, o projeto está em uma zona de indiferença de retorno, mas é economicamente viável

RAZÃO CUSTO-BENEFÍCIO

- A Razão Custo-Benefício é uma das estatísticas de retorno econômico mais utilizadas, devido à sua fácil compreensão
- Para o seu cálculo basta dividir o benefício total pelo custo econômico total . Tanto o benefício quanto o custo têm que estar a valor presente

$$\text{Razão Custo-Benefício} = \text{VPTB}_{t_n} / \text{VPTC}_{t_n}$$

- Onde t_n indica o tempo n , no qual o benefício e o custo foram calculados
- A interpretação do resultado nos fornece o montante de benefício que será gerado por cada real investido. Por exemplo, se a Razão Custo-Benefício foi de 1,5, significa que a cada real investido o benefício gerado será de um real e cinquenta centavos. Portanto, o projeto é viável economicamente
- Por outro lado, se a Razão Custo-Benefício for de 0,60, significa que cada real investido no projeto valerá no futuro apenas sessenta centavos. Ou seja, o dinheiro investido perdeu valor com o projeto, pois esse não foi capaz de gerar benefícios suficientes para compensar os custos

RAZÃO CUSTO-EFETIVIDADE

- A Razão Custo-Efetividade é uma alternativa para a comparação dos custos com os benefícios dos projetos, quando não se deseja, ou quando não é possível, valorar monetariamente o impacto
- Alguns projetos têm indicadores de impacto de difícil valoração. Por exemplo, o Projeto TAMAR. O impacto será medido em termos de número de tartarugas salvas. Para transformar o impacto em benefício teríamos que saber quanto vale para a sociedade cada tartaruga salva. Isso exigiria hipóteses nada triviais
- A Razão Custo-Efetividade é calculada pela divisão do impacto estimado (na avaliação de impacto) pelo custo econômico total

$$\text{Razão Custo-Efetividade} = \text{impacto estimado} / \text{VPTC}_{tn}$$

- Esta estatística fornece quanta unidade de impacto se obtém por real gasto. Por exemplo, no caso do TAMAR, a Razão Custo-Efetividade fornece quantas tartarugas foram salvas para cada real gasto
- Se a Razão Custo-Efetividade é de 0,000003, significa que a cada real investido salva 0,000003 tartarugas (a fim de tornar o resultado mais intuitivo, podemos multiplicar o resultado por 1.000.000)

ANÁLISE DE SENSIBILIDADE

- A análise de sensibilidade é a última etapa da avaliação econômica. Seu objetivo é testar a sensibilidade (variação) dos resultados da avaliação em relação aos parâmetros utilizados
- Como a avaliação econômica é baseada em métodos estatísticos, que estão sujeitos a erros, e em parâmetros escolhidos às vezes de forma discricionária, faz-se necessário testar se os resultados encontrados sofreriam alteração caso a amostra e os parâmetros da avaliação fossem diferentes
- Assim, a análise de sensibilidade é uma técnica de verificação de robustez dos resultados da avaliação. O termo robustez em avaliação é empregado no sentido figurado de “força” da avaliação, ou seja, sua confiabilidade
- Então, na prática, a análise de sensibilidade implica em refazer a avaliação econômica alterando os parâmetros utilizados (e.g. taxa de juros) e verificar se as estatísticas de retorno econômico continuam com o mesmo resultado
- Se o resultado é mantido em vários cenários, dizemos que ele é robusto

EXEMPLO FICTÍCIO

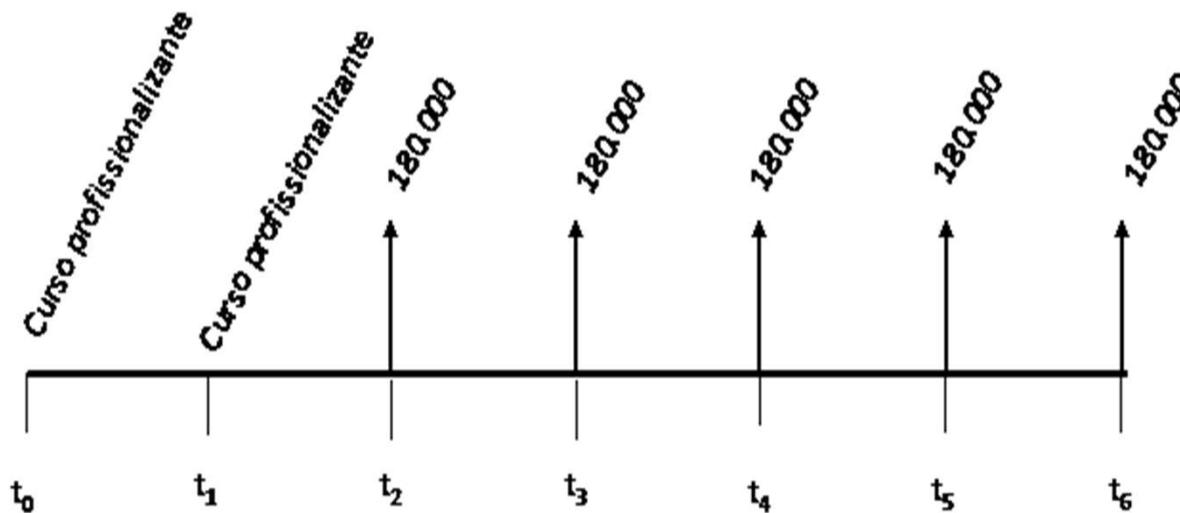
- Para fixar o conteúdo e servir de roteiro para futuras avaliações de retorno econômico, apresentamos a seguir a avaliação econômica de um projeto fictício, realçando todas as etapas e cálculos necessários à estimação do retorno do projeto
 - ✓ Nome do projeto: “Trabalha Brasil”
 - ✓ Objetivo: inserir jovens no mercado de trabalho.
 - ✓ Público alvo: jovens de 17 anos nos dois últimos anos do ensino médio.
 - ✓ Ações: curso profissionalizante com duração de dois anos, oferecido no contra turno escolar.
 - ✓ Número de beneficiários: 150 indivíduos.
 - ✓ Indicador de impacto: empregabilidade dos jovens.
 - ✓ Resultado da estimação de impacto um ano após o fim do projeto: em média, 10% dos tratados conseguiram emprego por causa do projeto

EXEMPLO FICTÍCIO: TRANSFORMAÇÃO DO IMPACTO EM BENEFÍCIO

- Dos 150 participantes 10% conseguiram emprego por causa do projeto
- Então, o projeto conseguiu inserir 15 jovens no mercado de trabalho, que não o acessariam se não tivessem participado do curso profissionalizante
- Para transformar o impacto em benefício considere as seguintes hipóteses:
 - i. O benefício de colocar o jovem no mercado de trabalho é o salário que ele recebe. Esse equivale ao que ele produz para a sociedade;
 - ii. Na média os beneficiários empregados recebem R\$ 1.000,00. Esse valor foi calculado com base nas informações de salário informadas pelos jovens.
- Para calcular o benéfico do projeto por um ano multiplicamos o número de jovens que estão empregados por causa do projeto (15) pelo salário médio que recebem por mês (R\$ 1.000,00) por 12 meses – perfazendo um total de R\$ 180.000,00 por ano

EXEMPLO FICTÍCIO: FLUXO DE CAIXA DOS BENEFÍCIOS

- Para elaborar o fluxo de caixa considere que:
 - i. Se não fosse o projeto, os jovens beneficiários entrariam no mercado de trabalho, de qualquer forma, cinco anos após o fim do projeto. Então, o benefício perdura durante cinco anos;
 - ii. A taxa de desconto intertemporal é de 5%.



EXEMPLO FICTÍCIO: FLUXO DE CAIXA (CONTINUAÇÃO)

Para somar os valores em momentos diferentes no tempo, aplicamos a taxa de desconto intertemporal e calculamos o Valor Presente Total do Benefício em t_0 :

$$VPTB_{t_0} = VB_{t_0} + \sum \frac{VFB_{t_n}}{(1+i)^n}$$

$$VPTB_{t_0} = VB_{t_0} + \frac{VFB_{t_1}}{(1+i)^1} + \frac{VFB_{t_2}}{(1+i)^2} + \frac{VFB_{t_3}}{(1+i)^3} + \frac{VFB_{t_4}}{(1+i)^4} + \frac{VFB_{t_5}}{(1+i)^5} + \frac{VFB_{t_6}}{(1+i)^6}$$

$$VPTB_{t_0} = 0 + 0 + \frac{180.000}{(1+0,05)^2} + \frac{180.000}{(1+0,05)^3} + \frac{180.000}{(1+0,05)^4} + \frac{180.000}{(1+0,05)^5} + \frac{180.000}{(1+0,05)^6} = 742.196,00$$

As externalidades advindas do salário dos jovens inseridos no mercado de trabalho por causa do projeto não foram computadas. Dessa forma, sabemos de antemão que esse benefício está subestimado.

CÁLCULO DO CUSTO ECONÔMICO

Para o funcionamento dos cursos, a gestão do projeto despense recursos com salário dos professores, material de laboratório, apostilas e lanches fornecidos para os alunos

Custo Contábil	Valor (R\$)
Salário dos professores e coordenador	60.000,00
Material de laboratório	100.000,00
Apostilas	5.000,00
Lanches	72.000,00
Total	137.000,00

CÁLCULO DE OPORTUNIDADE

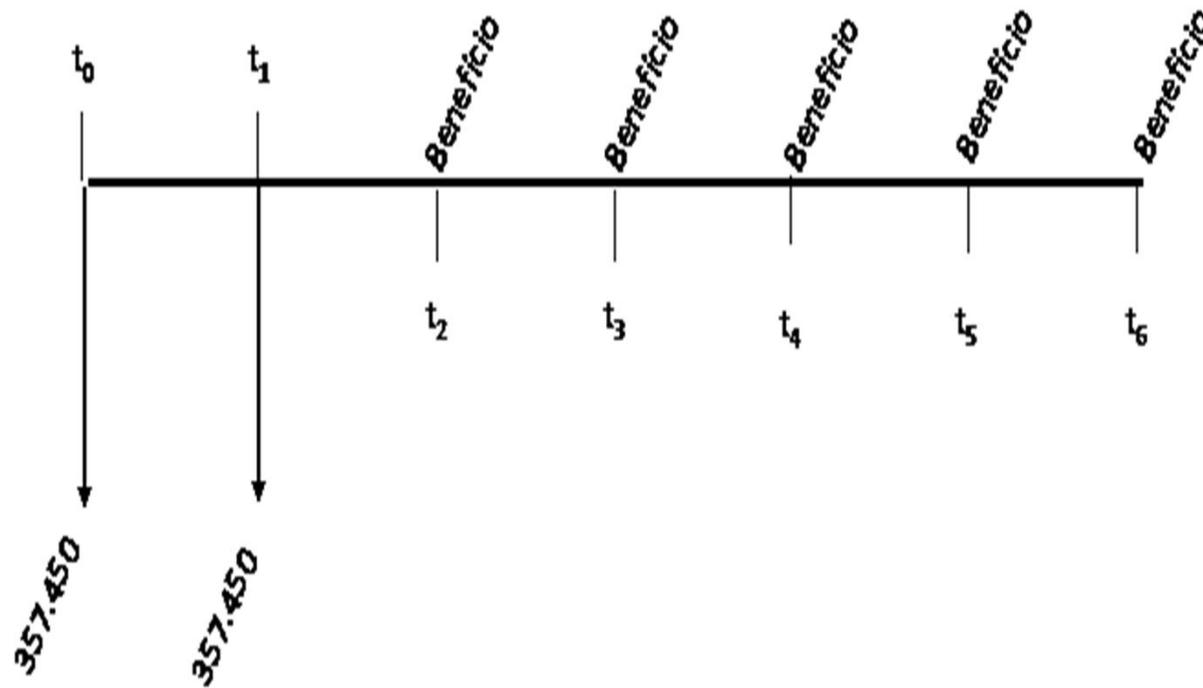
Também os alunos, ao decidirem alocar seu tempo nas aulas em vez de investi-lo em outras atividades, incorrem no custo de oportunidade

Custo de oportunidade	Valor (R\$)
Aluguel do laboratório	10.000,00
Conta de luz	300,00
Conta de água	150,00
Tempo dos alunos	210.000,00
Total	220.450,00

Custo econômico = custo contábil + custo de oportunidade

Custo econômico = 137.000,00 + 220.450,00 = 357.450,00

EXEMPLO FICTÍCIO: FLUXO DE CAIXA DOS BENEFÍCIOS E CUSTOS



CÁLCULO DO VALOR PRESENTE LÍQUIDO

Considere a taxa de desconto intertemporal de 5%. O Valor Presente Líquido é:

$$VPL_{t_0} = \left(VB_{t_0} + \sum \frac{VFB_{t_n}}{(1+i)^n} \right) - \left(VC_{t_0} + \sum \frac{VFC_{t_n}}{(1+i)^n} \right)$$

$$VPL_{t_0} = \left(\frac{180.000}{(1+0,05)^2} + \frac{180.000}{(1+0,05)^3} + \frac{180.000}{(1+0,05)^4} + \frac{180.000}{(1+0,05)^5} + \frac{180.000}{(1+0,05)^6} \right) - \left(357.450 + \frac{357.450}{(1+0,05)^1} \right)$$

$$VPL_{t_0} = 44.317,43$$

O resultado significa que os benefícios do projeto excedem os custos em 44.317,43 reais. Então, podemos concluir que o projeto apresenta um retorno econômico positivo para a sociedade, sendo viável economicamente.

CÁLCULO DA TAXA INTERNA DE RETORNO

No exemplo do projeto “Trabalha Brasil” a TIR é de 7%. A fórmula para seu cálculo é:

$$\left(VB_{t_0} - VC_{t_0} \right) + \left(\frac{VFB_{t_1}}{(1 + TIR)^1} + \dots + \frac{VFB_{t_n}}{(1 + TIR)^n} \right) - \left(\frac{VFC_{t_1}}{(1 + TIR)^1} + \dots + \frac{VFC_{t_n}}{(1 + TIR)^n} \right) = 0$$

A interpretação da TIR depende da taxa mínima de atratividade que considerarmos. Suponha que a taxa mínima de atratividade seja de 5%. Então, a TIR de 7% indica que o projeto é viável economicamente.

	A	B	C	D	E	F	G
11							
12							
13		tempo	custo	beneficio	beneficio-custo		
14		0	-357450		-357450	= TIR (E14:E20)	
15		1	-357450		-357450	7%	
16		2		180000	180000		
17		3		180000	180000		
18		4		180000	180000		
19		5		180000	180000		
20		6		180000	180000		
21							
22							

CÁLCULO DA RAZÃO CUSTO-BENEFÍCIO

Dos cálculos anteriores temos que:

$$\begin{aligned} VPTB_{t_0} &: 742.196,00 \\ VPTC_{t_0} &: 697.878,57 \end{aligned}$$

$$\begin{aligned} \text{Razão Custo - Benefício} &= \frac{VPTB_{t_0}}{VPTC_{t_0}} \\ \text{Razão Custo - Benefício} &= \frac{742196}{697878,57} = 1,06 \end{aligned}$$

O resultado indica que para cada real gasto no projeto, o retorno para a sociedade será de 1,06 reais. Lembre-se que esse valor foi calculado considerando uma taxa de desconto intertemporal de 5%.

No exemplo: Retorno = $(1,06 - 1) * 100 = 6,0\%$

Ou seja, sob uma taxa de desconto intertemporal de 5%, o retorno do projeto é de 6%.

CÁLCULO DA RAZÃO CUSTO-EFETIVIDADE

- Esta estatística fornece quanta unidade de impacto se obtém por real gasto
- Concluimos que o projeto conseguiu inserir 15 jovens no mercado de trabalho, que não o acessariam se não tivessem participado do curso profissionalizante
- Concluimos também que o valor presente dos custos totais é dado por:

$$\left(357.450 + \frac{357.450}{(1+0,05)^1} \right)$$

Razão Custo-Efetividade= impacto estimado / VPTC_{tn}

Razão Custo-Efetividade= 15 / 697.878,57 = 0.0000215

SUMÁRIO DO CURSO

- ~~1. Conceitos básicos em avaliação de impacto~~
- ~~2. Revisão de estatística e uso avançado de planilhas eletrônicas~~
- ~~3. Modelo de Resultados Potenciais~~
- ~~4. Econometria: revisão do modelo de mínimos quadrados~~
- ~~5. Desenho de uma avaliação para uma política~~
- ~~6. Modelo de Diferenças em Diferenças~~
- ~~7. Técnicas de Pareamento~~
- ~~8. Estudo de caso e aplicações de diferentes técnicas econométricas~~
- ~~9. Discussão sobre os trabalhos de conclusão de curso~~
- ~~10. Revisão de matemática financeira~~
- ~~11. O Cálculo do Retorno Econômico~~
- 12. Conclusão do curso: apresentação dos trabalhos**