



CÁLCULO DE PODER ESTATÍSTICO

Cálculo de Poder Estatístico

Organizadores

André Portela Souza
Lycia Lima

Equipe técnica

Alei Fernandes Santos
Caio de Souza Castro
Hugo Villarinho Pereira de Carvalho
Luigi Garzon
Michel Szklo
Pedro Molina Ogeda

Citação sugerida:

LIMA, L.; SOUZA, A. P. (Orgs.). Cálculo de poder estatístico. São Paulo: Fundação Getulio Vargas, 2025.

APRESENTAÇÃO

Este guia integra a série de publicações *Avaliação na Prática*, desenvolvida pelo FGV CLEAR com o objetivo de ampliar o acesso a conhecimentos sobre monitoramento e avaliação com foco em políticas públicas.

Fundado em 2015, o FGV CLEAR tem se dedicado ao fortalecimento da cultura de gestão orientada por evidências no Brasil e em países lusófonos. Com sede na Escola de Economia de São Paulo da Fundação Getúlio Vargas (FGV EESP), o FGV CLEAR atua como centro regional da Iniciativa CLEAR (Centers for Learning on Evaluation and Results).

A Iniciativa CLEAR, criada em 2010, é um programa de desenvolvimento de capacidades em monitoramento e avaliação que congrega instituições acadêmicas e parceiros doadores de modo a contribuir para o uso de evidências na tomada de decisões em países em desenvolvimento. Ao todo, são seis centros regionais CLEAR, coordenados pela *Global Evaluation Initiative (GEI)*, um programa liderado pelo Banco Mundial e pelo Programa das Nações Unidas para o Desenvolvimento (PNUD).

O FGV CLEAR atua com governos, organizações multilaterais, sociedade civil e academia, oferecendo capacitação, assistência técnica, geração e disseminação de conhecimento científico na área.

Saiba mais sobre o FGV CLEAR e acesse outras publicações em: **www.fgvclear.org**

CÁLCULO DE PODER ESTATÍSTICO

SUMÁRIO

1	INTRODUÇÃO	2
2	CÁLCULO DE PODER ESTATÍSTICO	4
2.1	DERIVANDO O MDE	6
2.2	PRATICANDO: FÓRMULA ANALÍTICA	9
2.3	PRATICANDO: SIMULAÇÕES	13
3	IMPERFECT COMPLIANCE	15
3.1	PRATICANDO: FÓRMULA ANALÍTICA	16
3.2	PRATICANDO: SIMULAÇÕES	18
4	RCT AGRUPADO	21
4.1	PRATICANDO: FÓRMULA ANALÍTICA	23
4.2	PRATICANDO: SIMULAÇÕES	26
5	ESTRATIFICAÇÃO	27
5.1	PRATICANDO: SIMULAÇÕES	30
6	MÚLTIPLAS HIPÓTESES	32
6.1	PRATICANDO: FÓRMULA ANALÍTICA	33
7	CONCLUSÃO	35

1 INTRODUÇÃO

Em avaliações de impacto, busca-se determinar se um programa ou intervenção foi efetivo. Para isso, comparam-se os resultados entre dois grupos (tratamento e controle) por meio de testes estatísticos. Esses testes ajudam a discernir se as diferenças observadas são fruto do acaso, de erros de amostragem ou se representam evidências concretas de impacto. Em regra, parte-se de uma hipótese nula, que pressupõe não haver diferença entre as médias dos grupos ou que a diferença média é zero.

A estatística t e o valor p são instrumentos analíticos que permitem avaliar se as diferenças observadas entre as médias dos grupos são estatisticamente significativas ou meramente produto do "acaso" ou de erros de amostragem. Esses testes fundamentam a decisão sobre a rejeição ou não da hipótese nula, que postula a ausência de diferenças significativas entre os grupos.

Dois tipos de erros podem comprometer os resultados de testes estatísticos. O primeiro, denominado **Erro Tipo I** (ou falso positivo), ocorre quando se rejeita indevidamente a hipótese nula. Em outras palavras, interpretamos que existe um impacto ou mudança, mas esse resultado é apenas um acaso do teste, sem efeito real. Esse erro leva à interpretação incorreta de flutuações aleatórias como se fossem efeitos reais da intervenção. O segundo, classificado como **Erro Tipo II** (ou falso negativo), ocorre quando não se rejeita a hipótese nula apesar da existência de um efeito real. Nessa situação, o teste falha em identificar um impacto que deveria ser identificado.

Logo, para minimizar esses erros, é necessário estabelecer um tamanho amostral adequado, capaz de limitar a ocorrência tanto de falsos positivos quanto de falsos negativos a níveis mínimos aceitáveis na medição da diferença média "esperada" ou "presumida" entre duas populações. Esses parâmetros tornam o teste mais confiável e preciso na detecção de impactos reais.

Como garantir que o tamanho de amostra é adequado para detectar um efeito, caso ele realmente exista? A solução está no conceito de Efeito Mínimo Detectável (**MDE**), que representa a menor diferença entre as médias dos grupos que o estudo é capaz de medir, considerando o tamanho da amostra, o poder do teste e a taxa de erro Tipo I

definidos. Quando a diferença observada é inferior ao MDE, o teste pode não alcançar significância estatística, mesmo que o efeito real exista. O cálculo do tamanho amostral necessário deve considerar os seguintes fatores: o **nível de confiança** (tipicamente fixado em 95%), que reflete o grau desejado de certeza; e o **poder do teste**, que representa a probabilidade de detectar um efeito real e costuma ser definido em 80%, dependendo da dispersão (ou variância) da variável de interesse do estudo.

Compreender o conceito de MDE é de extrema importância por diversas razões que afetam diretamente a condução e a eficácia de experimentos. Primeiramente, o MDE desempenha um papel crucial no planejamento do experimento, pois possibilita calcular o tamanho amostral necessário e definir a duração ideal para obter resultados confiáveis. Sem esse entendimento, há o risco de planejar experimentos subdimensionados ou excessivamente longos, comprometendo a eficiência e a validade das análises.

Além disso, compreender o MDE é fundamental para uma alocação eficiente de recursos. Investir tempo, dinheiro e esforços em experimentos mal calibrados pode levar a desperdícios significativos ou à incapacidade de detectar efeitos reais.

O MDE também oferece um contexto indispensável para a interpretação dos resultados obtidos nos experimentos. Ele estabelece uma referência clara sobre o tamanho mínimo do efeito que deve ser detectado, evitando interpretações equivocadas ou exageradas dos dados.

Por fim, ter uma noção clara do MDE é essencial para gerenciar riscos de forma adequada, permitindo avaliar o impacto potencial das intervenções testadas e os possíveis riscos associados a elas. Dessa forma, o entendimento do MDE não é apenas uma ferramenta técnica, mas uma estratégia essencial para garantir o sucesso de experimentos bem planejados e bem executados.

O **Poder do teste** é um conceito intimamente relacionado à taxa de erro Tipo II, pois é definido como 1 menos a taxa de erro Tipo II. Por exemplo, um teste com taxa de erro Tipo II de 20% possui poder de 80%. Em termos práticos, o poder representa a probabilidade de o teste identificar corretamente um efeito quando ele de fato existe.

Este guia apresenta o ferramental necessário para calcular o poder de experimentos.

Na Seção 2, primeiro é derivado o MDE para um caso padrão de *Randomized Controlled Trials* (RCT), prosseguindo com a demonstração de como o cálculo do MDE se modifica em diferentes desenhos experimentais nas seções subsequentes. A Seção 3 aborda o cálculo de MDE para situações de *imperfect compliance*, enquanto a Seção 4 examina o caso de *Clustered RCT*. A Seção 5 trata de experimentos estratificados, e a Seção 6 aborda contextos onde são testadas múltiplas hipóteses. Nestas seções, são apresentados tanto os fundamentos teóricos por trás da fórmula do MDE quanto exemplos práticos de implementação no software R. Os cálculos do MDE são executados tanto por meio de fórmulas analíticas quanto através de simulações.

2 CÁLCULO DE PODER ESTATÍSTICO

Para derivar a fórmula do Efeito Mínimo Detectável (MDE), é necessário primeiro compreender o ambiente do experimento. Para isso, inicia-se com um RCT ¹ simples como referência, conforme o seguinte exemplo prático: Avaliar os impactos da matrícula em uma escola de ensino técnico em comparação com escolar regulares. Nesse cenário, a aleatorização ocorre na distribuição dos alunos entre esses dois tipos de instituições de ensino.

Considere os resultados potenciais, $Y_i(d)$, para os valores de tratamento $d \in \{0, 1\}$, os quais serão entendidos nesse exemplo como as notas potenciais do aluno i em uma prova específica². Neste exemplo, a variável de tratamento, D_i , indica a matrícula em uma escola técnica **específica**: $D_i = 1$ quando o aluno i está matriculado na escola técnica em questão, e $D_i = 0$ quando matriculado em qualquer outra escola. O parâmetro de

¹Um RCT (*Randomized Controlled Trial*) é um método experimental utilizado em econometria para estimar relações causais. Nesse tipo de experimento, indivíduos ou unidades de análise, como pessoas, escolas ou empresas, são aleatoriamente alocados a um grupo de tratamento, que recebe a intervenção, ou a um grupo de controle, que não recebe. A aleatorização é fundamental para garantir que quaisquer diferenças observadas entre os grupos após o tratamento sejam atribuíveis exclusivamente ao próprio tratamento, eliminando o viés de seleção e assegurando validade interna. Por exemplo, para avaliar o impacto de um programa de bolsas de estudo no desempenho acadêmico, uma população de alunos elegíveis seria dividida aleatoriamente em dois grupos: um que recebe a bolsa e outro que não recebe. Após um período, mede-se o desempenho acadêmico de ambos os grupos, e a diferença média nos resultados reflete o impacto causal do programa. Veja Duflo et al. (2007).

²Uma prova como o Saeb, por exemplo.

interesse é o Efeito Médio de Tratamento (ATE)³, $\beta = \mathbb{E}[Y_i(1) - Y_i(0)]$, que pode ser interpretado como o efeito causal de estar matriculado nessa escola de ensino técnico específica.

Para fins de simplificação, considera-se que, se o resultado da loteria for $D_i = 1$, o estudante necessariamente se matricula na escola de ensino técnico e, se $D_i = 0$, ele não consegue se matricular (*perfect compliance*)⁴. Assume-se também a observação completa dos resultados de todos os alunos que participaram do experimento (não existe atrito)⁵.

O ATE pode ser estimado através do coeficiente β_1 na seguinte regressão de Mínimos Quadrados Ordinários (MQO):⁶

$$Y_i = \beta_0 + \beta_1 D_i + u_i \quad (2.1)$$

O estimador de MQO para β_1 é dado pela diferença de médias de Y_i entre o grupo de controle e o grupo de tratamento:

$$\hat{\beta}_1 = \frac{1}{N_1} \sum_{i:D_i=1} Y_i - \frac{1}{N_0} \sum_{i:D_i=0} Y_i \quad (2.2)$$

em que N_1 é o número de alunos tratados e N_0 é o número de alunos não tratados.⁷ Ao assumir homoscedasticidade dos erros, isto é, $\text{Var}(u_i) = \sigma^2$, para todo $i \in \{1, \dots, N\}$,

³O ATE (*Average Treatment Effect*), ou Efeito Médio do Tratamento, é uma medida que quantifica o impacto médio de uma intervenção em uma população, comparando o resultado esperado de quem recebe o tratamento com o de quem não recebe. Em experimentos como RCTs, o ATE pode ser estimado diretamente pela diferença nas médias dos resultados entre os grupos de tratamento e controle, pois a randomização garante que esses grupos sejam comparáveis.

⁴*Perfect compliance* ocorre quando todos os participantes de um estudo seguem exatamente as atribuições de tratamento ou controle conforme planejado, ou seja, quem foi designado para o tratamento o recebe, e quem foi designado ao controle não o recebe. Nesse cenário ideal, os efeitos observados refletem diretamente o impacto causal do tratamento. Na prática, a falta de *compliance* é comum, e métodos como o ITT (*intention to treat*) ou estimadores instrumentais são usados para lidar com desvios e evitar vieses.

⁵Atrito é o termo que se refere à perda de unidades observadas de uma dada amostra. Neste exemplo prático, haveria atrito caso algum aluno incluído na amostra inicial não realizasse a prova utilizada como medida de resultado para avaliação do efeito do tratamento.

⁶O método Mínimos Quadrados Ordinários (MQO) é uma técnica de otimização matemática que procura encontrar o melhor ajuste para um conjunto de dados mediante a minimização da soma dos quadrados das diferenças entre o valor estimado e os dados observados. Consultar [Angrist and Pischke \(2008\)](#) para mais detalhes.

⁷É comum ter o mesmo número de tratados e controles, isto é, $N_1 = N_0 = N/2$. Desta forma, pode-se escrever o estimador de MQO como $\hat{\beta}_1 = \frac{2}{N} \sum_{i:D_i=1} Y_i - \frac{2}{N} \sum_{i:D_i=0} Y_i$.

pode-se derivar a variância do estimador de MQO:

$$\text{Var}(\hat{\beta}_1) = \frac{1}{p(1-p)} \frac{\sigma^2}{N} \quad (2.3)$$

em que p é a proporção de indivíduos no grupo de tratamento e N é o número de indivíduos na amostra. Na próxima subseção, será utilizada a fórmula da variância do estimador de MQO para construir o teste e, a partir dele, derivar o Efeito Mínimo Detectável (MDE).

2.1 DERIVANDO O MDE

Para testar se o efeito do tratamento é zero, isto é, $\beta = 0$ são definidas as hipóteses nula e alternativa:

$$H_0 : \beta = 0$$

$$H_1 : \beta \neq 0$$

Conforme o exemplo empírico descrito anteriormente, a hipótese nula, H_0 , significa aqui que o efeito de ingressar em uma escola de ensino técnico nas notas dos alunos é 0, isto é, o aluno teria a mesma nota tanto em uma escola de ensino técnico, quanto em uma escola convencional. Já a hipótese alternativa, H_1 , afirma que há efeito, seja ele positivo ou negativo. Durante o desenho de um experimento, é essencial antecipar e considerar os possíveis erros que podem ocorrer durante o teste dessas hipóteses. A tabela 2.1 sintetiza os cenários possíveis.

	Não-Rejeita H_0	Rejeita H_0
H_0 é verdadeira	Decisão correta ($1 - \alpha$)	Erro do tipo I (α)
H_1 é verdadeira	Erro do tipo II ($1 - \kappa$)	Decisão correta (κ)

Table 2.1: Erros estatísticos

O Erro de Tipo I, α , também chamado de nível de significância, é a probabilidade de rejeitar a hipótese nula quando esta é verdadeira. Já o Erro de Tipo II, $1 - \kappa$, é a

probabilidade de não rejeitar a hipótese nula quando a hipótese alternativa é verdadeira. O poder do teste, κ , representa a probabilidade de rejeitar a hipótese nula quando a hipótese alternativa é verdadeira, correspondendo assim à decisão estatisticamente correta. Estes conceitos são importantes, pois eles estão diretamente relacionados com o cálculo do MDE. Para testar a hipótese de que $\beta = 0$, primeiro precisam ser definidos os valores de α e κ . Tipicamente, adota-se o valor $\alpha = 0.05$. Em seguida, se constrói o teste estatístico:

$$t = \frac{\hat{\beta}_1 - \beta}{\sqrt{\text{Var}(\hat{\beta}_1)}}$$

A estatística t compara o efeito médio calculado com a amostra com o efeito verdadeiro. Sob a hipótese de que o efeito não existe, H_0 , esse teste se torna:

$$t = \frac{\hat{\beta}_1}{\sqrt{\text{Var}(\hat{\beta}_1)}} \xrightarrow{d} \mathcal{N}(0, 1) \quad (2.4)$$

Considere um $\bar{\beta} > 0$, onde o efeito verdadeiro é dado por $\beta = \bar{\beta}$, indicando que a hipótese alternativa é verdadeira. Neste cenário, a distribuição do teste estatístico sob a hipótese alternativa segue uma distribuição normal com média deslocada para a direita em relação à distribuição nula:

$$t = \frac{\hat{\beta}_1}{\sqrt{\text{Var}(\hat{\beta}_1)}} = \underbrace{\frac{\hat{\beta}_1 - \bar{\beta}}{\sqrt{\text{Var}(\hat{\beta}_1)}}}_{\xrightarrow{p} \mathcal{N}(0,1)} + \frac{\bar{\beta}}{\sqrt{\text{Var}(\hat{\beta}_1)}} \quad (2.5)$$

A Figura 2.1 ilustra as distribuições dos testes sob H_0 (preto) e H_1 (vermelho). O valor $t_{\frac{\alpha}{2}}$ é o valor crítico do teste, isto é, quando o valor absoluto do teste excede esse limiar ($|t| > t_{\frac{\alpha}{2}}$), a hipótese nula é rejeitada. Portanto, área preenchida em azul representa a região de rejeição do teste. Se a estatística t estiver nessa região, a probabilidade do valor verdadeiro de β ser 0 é baixa. Caso a hipótese alternativa seja verdadeira, H_0 é corretamente rejeitada quando $|t| > t_{\frac{\alpha}{2}}$. Isso corresponde ao poder do teste, denotado por, κ e representado pela área vermelha. Esta relação quantifica a capacidade do teste

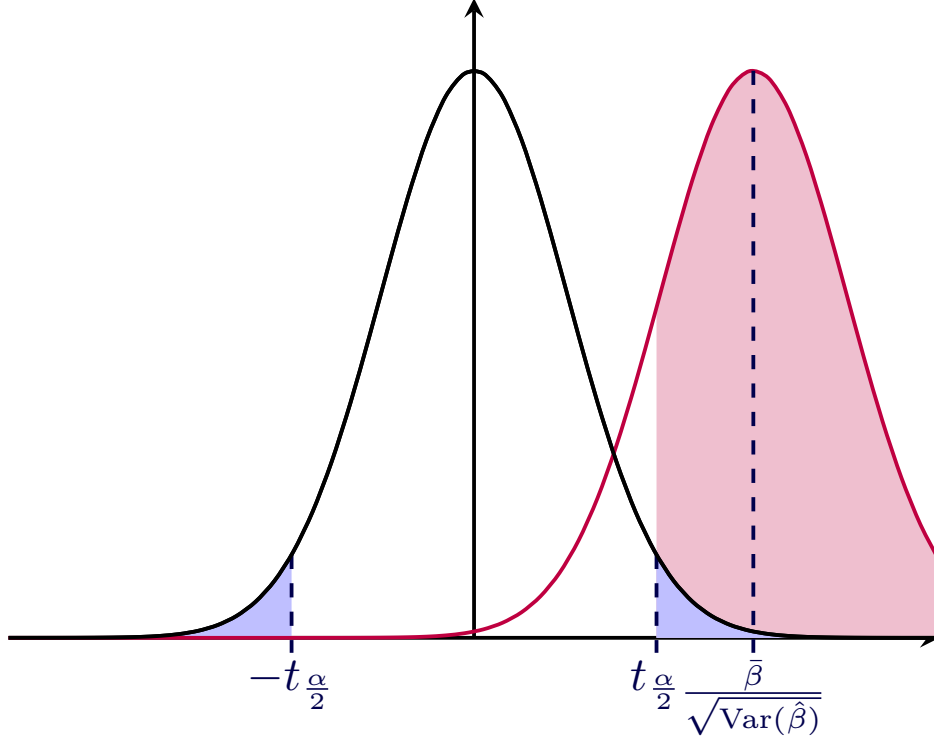


Figure 2.1: Distribuições dos testes

de detectar um efeito quando ele de fato existe. Assim, é possível estabelecer a seguinte relação:

$$\frac{\bar{\beta}}{\sqrt{\text{Var}(\hat{\beta}_1)}} - t_{\frac{\alpha}{2}} = t_{1-\kappa} \quad (2.6)$$

Da equação 2.6, combinando com a fórmula da variância do estimador de MQO obtida na equação 2.3, é possível isolar $\bar{\beta}$ para obter o Efeito Mínimo Detectável (MDE):

$$MDE := \bar{\beta} = (t_{\frac{\alpha}{2}} + t_{1-\kappa}) \sqrt{\frac{1}{p(1-p)} \frac{\sigma^2}{N}} \quad (2.7)$$

A equação 2.7 estabelece a relação dos parâmetros $\alpha, \kappa, \sigma^2, N$, e p para um valor de $\bar{\beta}$, que pode ser interpretado como o menor efeito de tratamento que detectável dado esses parâmetros. Na prática, é comum adotar $t_{\frac{\alpha}{2}} + t_{1-\kappa} = 2.8$, que corresponde aos valores convencionais de $\alpha = 0.05$ e $\kappa = 0.8$.

Vale destacar que a fórmula analítica derivada para o MDE depende das hipóteses assumidas. Aqui, presume-se uma situação de *perfect compliance*, não existência de atrito

e homoscedasticidade dos erros. É possível, por exemplo, alterar a presunção em torno da homoscedasticidade dos erros assumindo que a variância dos erros depende do status de tratamento, isto é, $\sigma_d^2 = \text{Var}(Y_i(d))$. Neste exemplo prático, isso ocorre quando se acredita que a variância das notas dos alunos tratados na prova é diferente da variância das notas dos alunos que não receberam o tratamento. Isso modifica a variância do estimador de MQO, resultando na seguinte expressão para o MDE:

$$MDE = (t_{\frac{\alpha}{2}} + t_{1-\kappa}) \sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_0^2}{N_0}} \quad (2.8)$$

Na equação 5.5, N_1 e N_0 são o número de pessoas tratadas e não tratadas, respectivamente, enquanto σ_1^2 e σ_0^2 equivalem às variâncias de Y_i para os grupos de tratamento e controle, respectivamente. Nas próximas subseções, serão modificadas outras hipóteses e, com isso, será derivado o MDE para diferentes contextos.

Em casos de resultados binários, isto é, $Y \in \{0, 1\}$, é possível derivar um limite superior para sua variância:

$$\text{Var}(Y) = \mathbb{E}[Y^2] - \mathbb{E}[Y]^2 = p(1 - p)$$

em que p é a probabilidade de $Y = 1$. Ao maximizar essa expressão, encontra-se que $\max \text{Var}(Y) = \frac{1}{4}$. Portanto, é possível derivar um limite superior para o MDE para os valores de α, κ, N_1, N_0 como:

$$MDE = (t_{\frac{\alpha}{2}} + t_{1-\kappa}) \sqrt{\frac{\frac{1}{4}}{N_1} + \frac{\frac{1}{4}}{N_0}}$$

Nas próximas seções, serão modificadas diferentes hipóteses a respeito do desenho do experimento, o que fará com que as fórmulas analíticas do MDE mudem.

2.2 PRATICANDO: FÓRMULA ANALÍTICA

Existem duas formas de calcular o MDE. A primeira é pela fórmula analítica, utilizando as equações derivadas na seção anterior. Esta é recomendada para quando não há acesso a dados prévios ao experimento e, portanto, não é possível estimar o desvio padrão da

variável de interesse, Y_i . Nesse caso, o MDE é calculado em termos de desvios padrões.⁸ Quando os dados necessários estão disponíveis, é possível optar por calcular o MDE através de simulações, conforme será explicado mais a frente nesta seção.

Considere novamente o exemplo empírico, em que o tratamento D_i , se refere a receber uma vaga em uma escola de ensino técnico específica e a variável de interesse, Y_i , representa a nota dos alunos em uma prova específica.

O software R será utilizado para a realização dos cálculos. Inicialmente, o ambiente de trabalho será limpo e, posteriormente, o pacote `tidyverse` será carregado. Esse pacote facilita a manipulação dos dados e auxilia na criação de gráficos.

```
# Limpar o ambiente
rm(list=ls())

# Carregando pacotes necessários
library(tidyverse)
```

O próximo passo envolve a definição das variáveis fixas no cálculo do poder. Conforme a equação (2.7), estabelece-se $t_{\frac{\alpha}{2}} = 1.96$, correspondente a um nível de significância de 5%; $t_{1-\kappa} = 0.84$, para um poder estatístico de 80%; e $p = 0.5$, indicando que 50% dos alunos na amostra serão aleatorizados para receberem uma vaga na escola de ensino técnico.

```
# Parâmetros para cálculo analítico do MDE

t.kappa = 0.84 # poder de 80%
t.alpha = 1.96 # nível de significância de 5%
P = 0.5 # Porcentagem de tratados de 50%
```

Agora, com base na equação (2.7), é criada a função para o cálculo do MDE para um determinado tamanho da amostra, N , utilizando os valores pré-definidos no último bloco

⁸Também é possível realizar este cálculo com simulações sem esses dados, onde as distribuições das variáveis são presumidas. No entanto, esse método está além do escopo desse guia.

de código.

```
# Fórmula do MDE como função de N

MDE = function(N){(t.kappa+t.alpha)*((1/(P*(1-P)))^0.5)*((1/N)
  ^0.5)}
```

Com essa função, é possível calcular o MDE para diferentes tamanhos de amostra. Com $N = 1000$, por exemplo, o MDE assume um valor de 0.177 desvios padrões.

```
# Valor do MDE quando temos N = 1000

MDE(1000)
```

Os resultados do MDE podem ser representados graficamente, conforme representado na Figura 2.2. O código apresentado a seguir gera o gráfico ilustrado, plotando os valores do MDE no eixo vertical (Y) em função do tamanho amostral, que varia entre 500 e 3500 observações, no eixo horizontal (X).

```
# Gráfico do MDE variando N

ggplot() + xlim(500, 3500) +
  geom_function(fun = MDE, colour = "blue", linewidth = 0.8) +
  labs(y = "MDE", x = "N") +
  theme_classic()
```

Conforme mencionado anteriormente, o cálculo do MDE é normalmente realizado em termos de desvio padrão quando se utiliza a fórmula analítica. Porém, como demonstrado na seção anterior, quando o Y_i é binário, é possível determinar o valor máximo de sua variância. Assim, no bloco de código abaixo, inicialmente defini-se o valor máximo da variância como $\sigma^2 = 0.25$, e implementa-se uma função para calcular o MDE máximo

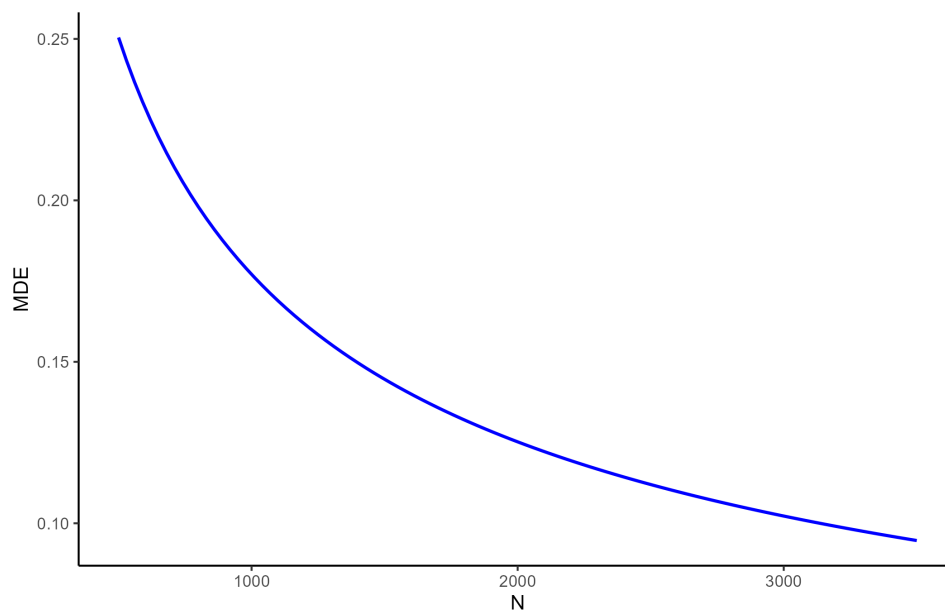


Figure 2.2: Gráfico com valores do MDE variando o tamanho da amostra de 500 a 3500.

no caso binário, utilizando esse parâmetro pré-estabelecido. As últimas quatro linhas do bloco de código geram o gráfico da Figura 2.3.

```
sigma.max = 1/4 # Variância máxima com Y binário

# Fórmula do MDE com Y binário como função de N

MDE.max = function(N){(t.kappa+t.alpha)*((1/(P*(1-P)))^0.5)*((
  sigma.max/N)^0.5)}

# Gráfico do MDE com Y binário variando N

ggplot() + xlim(500, 3500) +
  geom_function(fun = MDE.max, colour = "blue", linewidth = 0.8) +
  labs(y = "MDE", x = "N") +
  theme_classic()
```

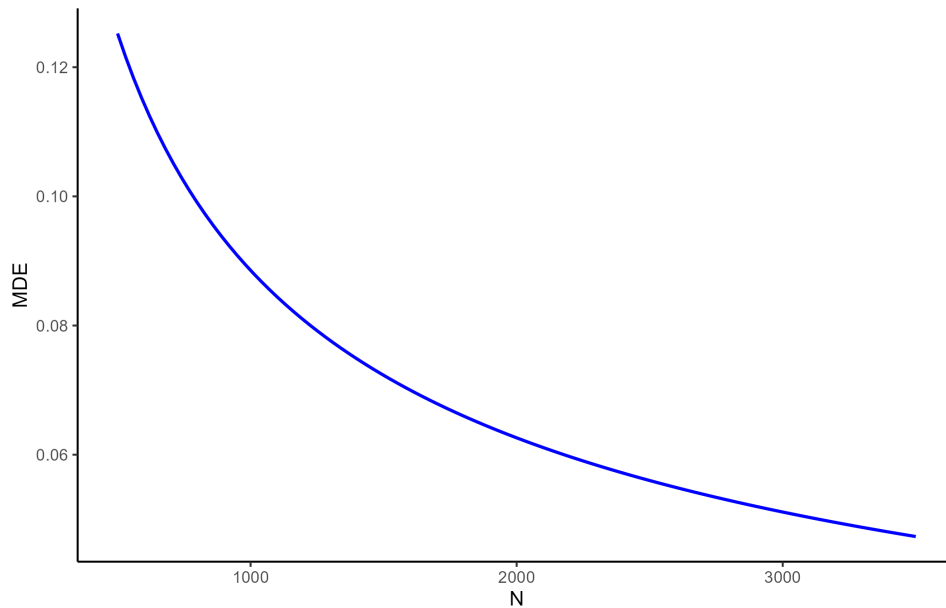


Figure 2.3: Gráfico com valores máximos do MDE no caso em que Y_i é binário, variando o tamanho da amostra de 500 a 3500.

2.3 PRATICANDO: SIMULAÇÕES

A fórmula analítica do MDE, permite seu cálculo em termos de desvios padrão, conforme demonstrado na seção anterior. No entanto, quando há disponibilidade de dados pré-experimentais, é possível estimar o desvio padrão por meio de simulações.

bloco de código a seguir realiza a limpeza do ambiente de trabalho e carrega os pacotes necessários para análise. O pacote **randomizr** é empregado para facilitar o processo de aleatorização dos alunos. Para que os resultados sejam reproduzíveis, define-se uma *seed*,⁹ isto é, um algoritmo de aleatorização, através da função **set.seed**. Aqui, será adotado o valor 42. Por fim, carrega-se a base de dados **base_rct.csv**, que localizada na pasta **Base**.

⁹As *seeds* de aleatorização funcionam como "pontos de partida" para a geração de números aleatórios em um computador. Podem ser pensadas como uma senha ou um código que determina a sequência de números gerada, pois computadores não geram números realmente aleatórios, mas sim pseudoaleatórios e que seguem uma lógica definida. Ao fixar uma *seed*, garante-se que a mesma sequência de números aleatórios será gerada em diferentes execuções do código, permitindo a reprodução exata dos resultados. Este mecanismo é particularmente útil em pesquisas científicas, onde a replicabilidade dos experimentos é essencial. Por exemplo, se um pesquisador realiza um sorteio em um código com uma *seed* específica, outra pessoa utilizando o mesmo código com a mesma *seed* obterá exatamente os mesmos resultados.

```

# Limpar o ambiente
rm(list=ls())

# Carregando pacotes necessários

library(tidyverse)
library(randomizr)

# Fazendo com que as simulações se tornem reproduzíveis
set.seed(42)

# Base de dados
base_rct <- read.csv("Base/base_rct.csv")

```

Em seguida, realiza-se a simulação de 100 alocações aleatórias diferentes do tratamento e, a partir disso, calcula-se a média do desvio padrão do estimador do MQO. Cada iteração do *loop* irá (i) aleatorizar o tratamento na amostra, (ii) realizar uma regressão de MQO das notas dos alunos, Y_i , no tratamento, D_i , e, por fim, (iii) armazenar o desvio padrão do estimador de MQO em um vetor. Para obter o MDE, calcula-se a média dos desvios padrões simulados e multiplica-se pela soma dos valores críticos, $t_{\frac{\alpha}{2}} + t_{1-\kappa} = 2.8$.

```

# Criando vetor com 100 entradas para armazenar os erros padrões
se <- c(rep(NA, 100))

# Simulando 100 vezes o erro padrão
for (i in 1:100) {

  D <- simple_ra(N = nrow(base_rct), prob = 0.5) # aleatorizando
  os alunos

  df <- cbind(base_rct, D) # juntando base

  reg <- lm(y ~ D, data = df) # regressão

```

```

se[i] <- summary(reg)$coefficients[2,2] # armazenando erro padrão
o

}

# Calculando MDE
MDE = 2.8*mean(se)
print(MDE)

```

3 IMPERFECT COMPLIANCE

Nesta seção, será considerado um cenário com *imperfect compliance*¹⁰, caracterizado pela adesão incompleta ao tratamento. Neste cenário, $Z_i = 1$ indica a oferta de vaga em escola de ensino técnico, enquanto $Z_i = 0$ representa a não oferta. Por outro lado, $D_i = 1$ corresponde à matriculado do aluno na escola de ensino técnico, e $D_i = 0$ à não matrícula. Aqui, a situação de *imperfect compliance* significa que alguns alunos que receberam uma vaga aleatoriamente não se matricularam ($Z_i = 1$, mas $D_i = 0$), enquanto outros alunos, que não receberam uma vaga de forma aleatória, conseguiram se matricular ($Z_i = 0$, mas $D_i = 1$).

Neste caso, ao utilizar a oferta de vaga em uma escola de ensino técnico, Z_i , como variável de tratamento, identifica-se o parâmetro *Intention-to-Treat* (ITT). Este estimador captura o efeito de oferecer o tratamento, independente da adesão. Para calcular o MDE do ITT, utiliza-se a mesma fórmula dada em 2.7. No entanto, em casos de *imperfect compliance*, o parâmetro de interesse frequentemente passa a ser o *Local Average Treatment Effect* (LATE), definido por $\beta^{\text{LATE}} = \mathbb{E}[Y_i(1) - Y_i(0) \mid C]$, onde C é o grupo

¹⁰Imperfect compliance ocorre em experimentos, como RCTs, quando os indivíduos não seguem perfeitamente as alocações designadas, resultando em situações em que alguns no grupo de tratamento não recebem o tratamento, e outros no grupo de controle acabam recebendo-o. Isso compromete a adesão completa ao design experimental e pode introduzir viés nas estimativas causais. Para lidar com esse problema, utilizam-se métodos como o *Intention-to-Treat* (ITT), que mede o efeito com base nas alocações originais, independentemente da adesão, refletindo o impacto de oferecer o tratamento; e o *Local Average Treatment Effect* (LATE), que isola o efeito causal sobre os *compliers* (indivíduos que seguem as alocações designadas), geralmente estimado com variáveis instrumentais. Essas abordagens permitem obter inferências robustas mesmo diante de conformidade imperfeita, comum em experimentos aplicados.

de *compliers*, isto é, quando $Z_i = D_i$. Ele pode ser identificado por

$$\beta^{\text{LATE}} = \frac{\mathbb{E}[Y_i | Z_i = 1] - \mathbb{E}[Y_i | Z_i = 0]}{\mathbb{E}[D_i | Z_i = 1] - \mathbb{E}[D_i | Z_i = 0]}$$

O parâmetro β^{LATE} pode ser estimado através da divisão da diferença de médias de Y_i entre o grupo de tratamento e controle, $\bar{Y}_1 - \bar{Y}_0$, pela diferença de médias de D_i entre o grupo de tratamento e controle, $\bar{D}_1 - \bar{D}_0$, denominada de *take up* diferencial. No exemplo prático utilizado neste guia, o denominador do estimador do LATE, $\bar{D}_1 - \bar{D}_0$, pode ser interpretado como a diferença de matrícula na escola de ensino técnico entre os alunos do grupo de tratamento e controle. Para calcular o MDE do estimador do LATE, basta dividir a fórmula apresentada em 2.7 por essa diferença de adesão ao tratamento:

$$MDE = (t_{\frac{\alpha}{2}} + t_{1-\kappa}) \sqrt{\frac{1}{p(1-p)} \frac{\sigma^2}{N} \frac{1}{(\bar{D}_1 - \bar{D}_0)}} \quad (3.1)$$

Em cenários de *imperfect compliance*, o MDE pode ser substancialmente maior comparado a situações de *perfect compliance*, tais como na subseção anterior. Para ilustrar, considere um caso em que apenas 50% dos alunos do grupo de tratamento efetivamente se matriculam na escola técnica, enquanto nenhum aluno do grupo de controle o faz ($\bar{D}_1 - \bar{D}_0 = 50\%$). Nesse experimento, seria necessário quadruplicar o tamanho amostral para alcançar o mesmo MDE obtido em um experimento com adesão completa (100% de *compliance*). Essa relação decorre do fato de que o MDE varia linearmente com a proporção de *compliance*, enquanto sua dependência com o tamanho amostral N segue uma escala de raiz quadrada.

3.1 PRATICANDO: FÓRMULA ANALÍTICA

Agora, será calculado o MDE através da fórmula analítica para o caso de *imperfect compliance*. Tal como na seção anterior, o primeiro passo consiste em limpar o ambiente e carregar o pacote `tidyverse`. Posteriormente, definem-se os parâmetros fixos necessários para o cálculo do MDE.

```
# Limpar o ambiente
rm(list=ls())

# Carregando pacotes necessários
library(tidyverse)

# Parâmetros para cálculo analítico do MDE
t.kappa = 0.84 # poder de 80%
t.alpha = 1.96 # nível de significância de 5%
P = 0.5 # Porcentagem de tratados de 50%
```

Observa-se que o tamanho do parâmetro de *take up* diferencial não foi previamente definido, uma vez que a função do cálculo do MDE será implementada considerando tanto o tamanho da amostra quanto o *take up* diferencial. No bloco de código abaixo, a função `MDE.IV` se refere à equação (3.1), em que o argumento `N` se refere ao tamanho da amostra, N , e o argumento `take.up` corresponde ao *take up* diferencial, $\bar{D}_1 - \bar{D}_0$.

```
# Fórmula analítica do MDE em função de N e da % de take up

MDE.IV = function(N,take.up){(t.kappa+t.alpha)*((1/(P*(1-P)))^0.5)
  *((1/N)^0.5*(1/take.up))}
```

O bloco de código a seguir gera a Figura 3.1, que apresenta a relação entre o MDE (em termos de desvio padrão) no eixo Y, e o tamanho da amostra (variando de 500 até 3500) no eixo X. Três curvas distintas são plotadas, correspondentes a cada valor de *take up* diferencial: 50%, 60% e 70%. Observa-se que, quanto menor o *take up* diferencial, maior o valor do MDE.

```
# Gráfico do MDE variando N, para take up de 50%, 60% e 70%
```

```

ggplot() + xlim(500, 3500) +
  geom_function(fun = MDE.IV, args = list(take.up = 0.50), aes(
    colour = "Take Up = 50%"), linewidth = 0.8) +
  geom_function(fun = MDE.IV, args = list(take.up = 0.60), aes(
    colour = "Take Up = 60%"), linewidth = 0.8) +
  geom_function(fun = MDE.IV, args = list(take.up = 0.70), aes(
    colour = "Take Up = 70%"), linewidth = 0.8) +
  labs(y = "MDE", x = "N") +
  theme_classic() +
  guides(colour=guide_legend(title="Take Up"))

```

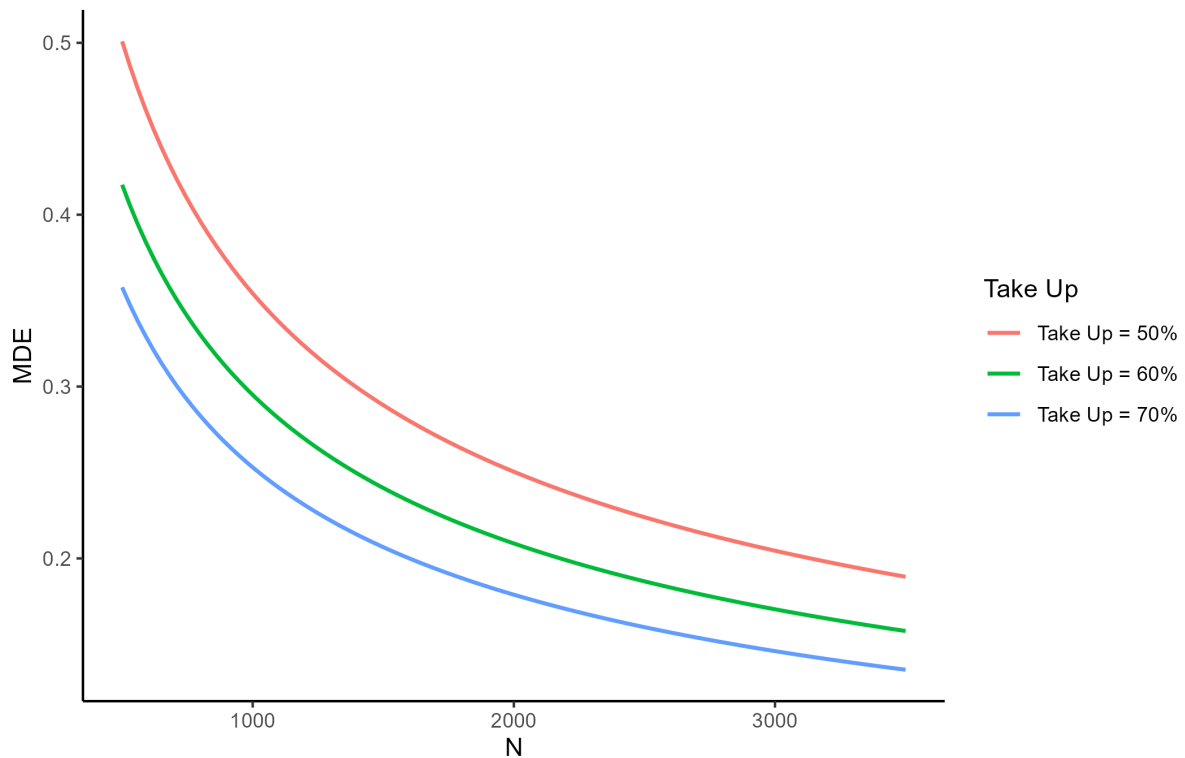


Figure 3.1: Gráfico do MDE em termos de desvio padrão variando o tamanho da amostra de 500 a 3500 para diferentes tamanhos de *take up*.

3.2 PRATICANDO: SIMULAÇÕES

A seguir, o cálculo do MDE para o cenário de *imperfect compliance* será realizado mediante simulações. Primeiramente, procede-se com a limpeza do ambiente e carregamento

dos pacotes necessários. Desta vez, também será utilizado o pacote `ivreg`, que permite realizar regressões de 2SLS.¹¹ Novamente, estabelece-se a *seed* seed 42. Por último, carrega-se novamente a base de dados `base_rct.csv`.

```
# Limpar o ambiente
rm(list=ls())

# Carregando pacotes necessários

library(tidyverse)
library(randomizr)
library(ivreg)

# Fazendo com que as simulações se tornem reproduzíveis
set.seed(42)

# Base de dados
base_rct <- read.csv("Base/base_rct.csv")
```

O processo para calcular o erro padrão do estimador de 2SLS é semelhante ao caso do MQO na seção anterior, com a adaptação necessária para incorporar a aleatorização do instrumento Z_i . Para determinar o status de tratamento, D_i , cria-se uma variável auxiliar, `aux`, que assume uma distribuição uniforme $[0,1]$. Aqui, a variável de tratamento depende dos valores do instrumento e da variável auxiliar. Essa variável determina a probabilidade com que o aluno se matricula de fato em uma escola de ensino técnico. No exemplo de código abaixo, caso o aluno não receba o incentivo da vaga na escola de ensino técnico ($Z_i = 0$), a probabilidade dele se matricular na escola equivale a 20%, enquanto que, caso o aluno receba o incentivo ($Z_i = 1$), a probabilidade da matrícula chega a 75%. Portanto, o *take up* para cada iteração da simulação será aproximadamente de $75\% - 20\% = 55\%$.

¹¹Uma regressão de MQO em dois estágios (*Two-Stage Least Squares* – 2SLS) é realizada quando há um problema de endogeneidade em um modelo de regressão. O 2SLS funciona como um “filtro” das partes problemáticas do tratamento (estar matriculado em uma escola de ensino técnico), usando uma variável auxiliar (ter recebido a oferta de um sorteio aleatório) para isolar o efeito puro da variável que realmente interessa para o experimento.

O restante da simulação é análogo ao cenário de *perfect compliance*, pois segue as seguintes etapas: (i) execução da regressão de 2SLS para cada iteração; (ii) armazenamento do erro padrão do estimador de 2SLS em um vetor e (iii) cálculo do MDE utilizando a média dos erros padrão estimados.

```
# Criando vetor com 100 entradas para armazenar os erros padrão
se_late <- c(rep(NA, 100))

# Simulando 100 vezes o erro padrão
for (i in 1:100) {

  Z <- simple_ra(N = nrow(base_rct), prob = 0.5) # Aleatorizando o
  instrumento

  df <- cbind(base_rct, Z) # juntando bases

  df <- df %>% # criando variável auxiliar para criar variável de
  tratamento

  mutate(aux = runif(nrow(base_rct))) %>%
  mutate(D = case_when(Z==0 & aux<0.2 ~ 1,
                        Z==0 & aux>0.2 ~ 0,
                        Z==1 & aux<0.75 ~ 1,
                        Z==1 & aux>0.75 ~ 0)) %>%

  select(-aux)

  reg_late <- ivreg(y ~ D|Z, data = df) # regressão 2sls
  se_late[i] <- summary(reg_late)$coefficients[2,2] # armazenando
  erro padrão

}

# Calculando MDE
MDE_late = 2.8*mean(se_late)
print(MDE_late)
```

4 RCT AGRUPADO

Os desenhos experimentais discutidos até o momento pressupõem a aleatorização no nível individual, onde cada unidade de análise (como um aluno específico) é aleatorizada para receber ou não a intervenção. Por exemplo, no caso empírico analisado, os alunos são aleatorizados para determinar quem receberá uma vaga em uma escola de ensino técnico, permitindo uma análise causal direta no nível individual. Contudo, existem contextos alternativos em que, apesar da disponibilidade de dados individuais, a aleatorização ocorre em nível agregado. Um exemplo ilustrativo seria a randomização de escolas inteiras para implementação de um sistema de ensino técnico, mantendo-se a coleta de dados no nível dos alunos pertencentes a essas instituições. Nesses casos, embora a unidade de aleatorização seja a escola, a análise ainda pode ser realizada utilizando informações individuais dos estudantes.

Esse delineamento experimental exige ajustes metodológicos na especificação econométrica, pois a aleatorização em *clusters* (como escolas) cria dependências entre as observações dentro de cada unidade. Tais dependências são qualificadas pela correlação intraclasse (*Intraclass Correlation* - ICC), que mensura o grau de similaridade entre as observações dentro de um mesmo grupo (*cluster*). Se os alunos da mesma escola tendem a ter notas mais parecidas entre si (por exemplo, devido ao mesmo professor, estrutura da escola ou ambiente socioeconômico), a ICC assume um valor mais elevado. Se há maior variedade entre as notas dos alunos dentro da mesma escola, a ICC é baixa.

Para capturar essa estrutura e garantir a robustez das inferências estatísticas, recomenda-se especificar uma regressão que considere as correlações intraclasse da seguinte maneira:

$$Y_{ij} = \beta_0 + \beta_1 D_j + \nu_j + \eta_{ij} \quad (4.1)$$

onde i se refere ao aluno e j se refere à escola. Por simplicidade, assume-se que ν_j equivale a *i.i.d* e $\text{Var}(\nu_j) = \tau^2$, e que η_{ij} é *i.i.d* e $\text{Var}(\eta_{ij}) = \omega^2$. Ainda, presume-se que há J escolas com n alunos em cada uma. Novamente, o objetivo é estimar β_1 . Nesse caso, a variância do estimador de MQO se torna

$$\text{Var}(\hat{\beta}_1) = \frac{1}{p(1-p)} \frac{n\tau^2 + \omega^2}{nJ} \quad (4.2)$$

A partir da equação 4.2, é possível derivar a fórmula do MDE para o caso de aleatorização a nível de grupo:

$$MDE = (t_{\frac{\alpha}{2}} + t_{1-\kappa}) \sqrt{\frac{1}{p(1-p)} \frac{n\tau^2 + \omega^2}{nJ}} \quad (4.3)$$

Também é possível derivar uma fórmula do MDE que se compare com o caso de um desenho de experimento em que a aleatorização tenha sido feita a nível de indivíduo. A notação introduzida nesta subsecção permite escrever a variância do estimador de MQO para o caso de aleatorização a nível de indivíduo como

$$\text{Var}(\hat{\beta}_{1, \text{indiv\u00edduo}}) = \frac{1}{p(1-p)} \frac{\tau^2 + \omega^2}{nJ} \quad (4.4)$$

A divisão da equação 4.2 pela equação 4.4, resulta no efeito do desenho (*Design Effect* - DE), isto é, o efeito na variância do estimador quando a aleatorização é realizada no nível grupal comparado com a aleatorização no nível individual:

$$DE = 1 + (n-1)\rho \quad (4.5)$$

onde $\rho = \frac{\tau^2}{\tau^2 + \omega^2}$ corresponde à ICC. Nota-se que a aleatorização em nível grupal amplifica a variância do estimador de MQO de tanto a ρ quanto a n . Essa relação permite derivar uma expressão para o MDE em função: (i) do MDE sob aleatorização individual (Equação 2.7) e (ii) do efeito do desenho:

$$MDE = (t_{\frac{\alpha}{2}} + t_{1-\kappa}) \sqrt{\frac{1}{p(1-p)} \frac{\sigma^2}{nJ} (1 + (n-1)\rho)} \quad (4.6)$$

A desconsideração da ICC em desenhos com aleatorização a nível de grupo resulta na subestimação da variância do estimador e, portanto, do MDE. Como há maior semelhança entre as observações dentro do mesmo *cluster*, é como se houvesse menos "dados úteis" do que o número total de observações sugere.

Os resultados do MDE apresentados dependem da hipótese de que J e n estejam fixos. Porém, muitas vezes, por restrições da natureza do experimento, é possível que tanto J , quanto n variem por grupo de tratamento. Por exemplo, limitações orçamentárias podem exigir a aleatorização de menos escolas para receber o tratamento. Isso faria com que $J_1 < J_0$, alterando a variância do estimador de MQO e, por consequência, o MDE. Neste caso, a fórmula do MDE seria

$$MDE = (t_{\frac{\alpha}{2}} + t_{1-\kappa}) \sqrt{\sigma^2(1 + (n-1)\rho) \left(\frac{1}{nJ_0} + \frac{1}{nJ_1} \right)} \quad (4.7)$$

O mesmo pode ocorrer com o número médio de indivíduos por *cluster*, n . Se esse número for diferente a depender do grupo de tratamento, o MDE equivale a

$$MDE = (t_{\frac{\alpha}{2}} + t_{1-\kappa}) \sqrt{\sigma^2 \left(\frac{1-\rho}{n_1J} + \frac{1+(2n_0-1)\rho}{n_0J} \right)} \quad (4.8)$$

4.1 PRATICANDO: FÓRMULA ANALÍTICA

Esta seção apresenta o cálculo do MDE através da formula analítica para o caso de *Clustered RCT*. Como nas seções anteriores, inicia-se com a limpeza do ambiente, seguida pelo carregamento dos pacotes necessários e a definição dos parâmetros fixos para o cálculo do MDE.

```
# Limpar o ambiente
rm(list=ls())

# Carregando pacotes necessários
library(tidyverse)

# Parâmetros para cálculo analítico do MDE
t.kappa = 0.84 # poder de 80%
t.alpha = 1.96 # nível de significância de 5%
P = 0.5 # Porcentagem de tratados de 50%
```

Em seguida, cria-se a função `MDE.cluster`, baseada na equação (4.6), permitindo a variação dos seguintes três argumentos na função: `N`, referente ao número médio de alunos em uma escola, N ; `J`, que equivale à quantidade de escolas na amostra, J ; e `rho`, referente ao valor da ICC, ρ .

```
# Fórmula do MDE variando N, J e ICC

MDE.cluster = function(N,J,rho){(t.kappa+t.alpha)*((1/(P*(1-P)))
  ^0.5)*((1/(N*J))^0.5*(1+(N - 1)*rho)^0.5)}
```

A seguir serão gerados dois gráficos: o primeiro exibe a variação do número médio de alunos nas escolas (N), enquanto o segundo mostra a variação do número de escolas na amostra (J). Ambos os gráficos contam com três curvas para diferentes valores de ICC.

O bloco de código abaixo gera o gráfico da Figura 4.1a, em que o MDE é calculado variando o número médio de alunos, N , de 10 a 70. Neste gráfico, o número de escolas está fixado em 50, e os três valores de ICC utilizados são 0.05, 0.1 e 0.15.

```
# Gráfico do MDE variando N, para ICC de 0.05, 0.1 e 0.15, com J =
  50

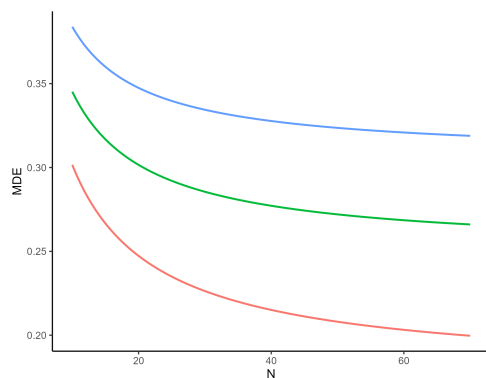
ggplot() + xlim(10, 70) +
  geom_function(fun = MDE.cluster, args = list(J = 50, rho = 0.05)
    , aes(colour = "ICC = 0.05"), linewidth = 0.8) +
  geom_function(fun = MDE.cluster, args = list(J = 50, rho = 0.1),
    aes(colour = "ICC = 0.1"), linewidth = 0.8) +
  geom_function(fun = MDE.cluster, args = list(J = 50, rho = 0.15)
    , aes(colour = "ICC = 0.15"), linewidth = 0.8) +
  labs(y = "MDE", x = "N") +
  theme_classic() +
  guides(colour=guide_legend(title="ICC"))
```

O segundo gráfico, gerado pelo bloco de código abaixo, está representado na Figura

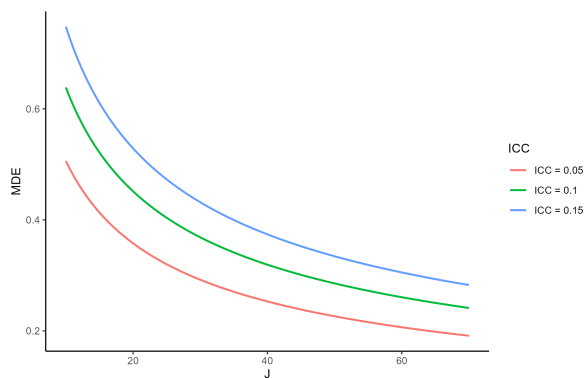
4.1b. Nela, o MDE é calculado com base na variação do número de escolas da amostra, J , de 10 a 70, mantendo o quantitativo médio de alunos por escola fixo em 30. Os valores da ICC são os mesmos três utilizados para o primeiro gráfico. Observa-se que, em ambos os casos, quanto maior a ICC, maior o MDE.

```
# Gráfico do MDE variando J, para ICC de 0.05, 0.1 e 0.15, com N =
30

ggplot() + xlim(10, 70) +
  geom_function(fun = MDE.cluster, args = list(N = 30, rho = 0.05),
    , aes(colour = "ICC = 0.05"), linewidth = 0.8) +
  geom_function(fun = MDE.cluster, args = list(N = 30, rho = 0.1),
    , aes(colour = "ICC = 0.1"), linewidth = 0.8) +
  geom_function(fun = MDE.cluster, args = list(N = 30, rho = 0.15),
    , aes(colour = "ICC = 0.15"), linewidth = 0.8) +
  labs(y = "MDE", x = "J") +
  theme_classic() +
  guides(colour=guide_legend(title="ICC"))
```



(a) Gráfico de MDE variando N para diferentes valores de ICC.



(b) Gráfico de MDE variando J para diferentes valores de ICC.

Figure 4.1: Gráficos de MDE variando o número médio de alunos e o número de escolas da amostra para diferentes valores de ICC.

4.2 PRATICANDO: SIMULAÇÕES

Nesta seção, o cálculo do MDE para *Clustered RCT* é realizado por meio de simulações. Segue-se as etapas estabelecidas nas simulações anteriores, começando com a limpeza do ambiente de trabalho, seguida pelo carregamento dos pacotes necessários, a escolha de uma *seed* aleatória e o carregamento da base de dados. Desta vez, também será utilizado o pacote `lfe`, garantindo que as estimativas dos erros padrões considerem o nível do *cluster*. Por fim, a base `base_rct_cluster.csv` é importada.

```
# Limpar o ambiente
rm(list=ls())

# Carregando pacotes necessários

library(tidyverse)
library(randomizr)
library(lfe)

# Fazendo com que as simulações se tornem reproduzíveis
set.seed(42)

# Base de dados
base_rct_cluster <- read.csv("Base/base_rct_cluster.csv")
```

O bloco de código abaixo estima o erro padrão do estimador de efeito fixo por meio de simulações. Aqui, o tratamento é aleatorizado no nível das escolas, o que representa a grande diferença em relação às seções anteriores. Com isso, realiza-se uma regressão clusterizando o erro padrão a nível das escolas, armazenando os erros padrões para que estes sejam utilizados em seguida durante o cálculo do MDE.

```
# Criando vetor com 100 entradas para armazenar os erros padrões
se_cluster <- c(rep(NA, 100))
```

```

# Simulando 100 vezes o erro padrão
for (i in 1:100) {

  # aleatorizando escolas
  D <- simple_ra(N = length(unique(base_rct_cluster$id_escola)),
    prob = 0.5)

  # criando base para dar match na base principal
  match <- data.frame(id_escola = unique(base_rct_cluster$id_
    escola),

                      D = D)

  df <- left_join(base_rct_cluster, match, by = "id_escola") #
    juntando bases

  reg_cluster <- felm(y ~ D|0|0|id_escola, data = df) # regressão
  se_cluster[i] <- summary(reg_cluster)$coefficients[2,2] #
    armazenando se

}

# Calculando MDE
MDE_cluster = 2.8*mean(se_cluster)
print(MDE_cluster)

```

5 ESTRATIFICAÇÃO

A aleatorização em blocos constitui um método frequentemente empregado para garantir maior controle no balanceamento de covariadas pré-tratamento entre grupos de tratamento e controle. Essa abordagem, também conhecida como estratificação, é especialmente útil para assegurar que certos fatores sejam equilibrados antes da aplicação do tratamento, como idade, gênero ou características socioeconômicas. Além de promover

um balanceamento mais preciso, a aleatorização em blocos tende a reduzir a variância do estimador de interesse, o que resulta em ganhos significativos de eficiência na análise. Essa característica torna os desenhos estratificados particularmente atrativos em contextos onde a precisão das estimativas é crucial, maximizando o poder estatístico do experimento sem aumentar o tamanho da amostra.

No exemplo empírico utilizado ao longo desse guia, considere o objetivo de garantir balanceamento para variáveis demográficas pré-tratamento, como gênero, raça e faixa de renda. Para tanto, criam-se grupos homogêneos de alunos segundo essas características e, dentro de cada grupo g , realiza-se a aleatorização das vagas em escolas de ensino técnico. Essa abordagem permite estimar o efeito do tratamento para cada um desses grupos como

$$\hat{\beta}_g = \bar{Y}_{1,g} - \bar{Y}_{0,g} \quad (5.1)$$

onde $\bar{Y}_{d,g}$ equivale à média dos alunos com status de tratamento $d \in \{0, 1\}$ dentro do bloco g . A variância do estimador para cada grupo g é dada por

$$\text{Var}(\hat{\beta}_g) = \frac{\sigma_{1,g}^2}{N_{1,g}} + \frac{\sigma_{0,g}^2}{N_{0,g}} \quad (5.2)$$

onde $\sigma_{d,g}^2 = \text{Var}(Y_{i,g}(d))$ e $N_{d,g}$ se refere ao número de alunos com status de tratamento $d \in \{0, 1\}$ no grupo g . É possível estimar o efeito do tratamento para todos os grupos calculando a média de todos os estimadores por grupo e ponderando pela proporção de alunos em cada grupo, $\frac{N_g}{N}$,

$$\hat{\beta} = \sum_{g=1}^G \hat{\beta}_g \frac{N_g}{N} \quad (5.3)$$

Com isso, a variância do estimador do efeito médio de tratamento é dada por

$$\text{Var}(\hat{\beta}) = \sum_{g=1}^G \text{Var}(\hat{\beta}_g) \left(\frac{N_g}{N} \right)^2 \quad (5.4)$$

A variância do estimador para cada grupo, dada pela equação 5.2, permite calcular o MDE para cada bloco g ,

$$MDE_g = (t_{\frac{\alpha}{2}} + t_{1-\kappa}) \sqrt{\frac{\sigma_{1,g}^2}{N_{1,g}} + \frac{\sigma_{0,g}^2}{N_{0,g}}} \quad (5.5)$$

O MDE do efeito de tratamento para todos os grupos pode ser obtido por meio da média dos MDEs de cada grupo, ponderando pela proporção de alunos em cada grupo,

$$MDE = \sum_{g=1}^G MDE_g \frac{N_g}{N} \quad (5.6)$$

A fórmula dada pela equação 5.6 equivale à derivação do MDE a partir da variância do estimador do efeito de tratamento para todos os grupos, como especificado na equação 5.4.

Um caso extremo da aleatorização em blocos é a **aleatorização por pares**¹², em que os alunos são pareados de acordo com características em comum e, dentro de cada par, um aluno é aleatorizado para receber o tratamento, enquanto o outro é alocado para o grupo de controle. O procedimento para calcular o MDE nesse caso é igual ao da aleatorização em blocos. No entanto, não é possível estimar a variância do estimador para cada um dos pares, dada pela equação 5.2, uma vez que há apenas um aluno tratado e um não tratado em cada par. Isso faz com que o teste t seja conservador, ou seja, a probabilidade de rejeitar a hipótese nula é menor do que o α especificado. Bai et al. (2022) derivam inferência exata para esse caso, propondo ajustes no teste. Em situações de aleatorização por pares, o foco será calcular o MDE por meio de simulações, conforme abordado na próxima seção.

O MDE pode ser derivado para casos de **efeitos heterogêneos e múltiplos tratamentos** seguindo a mesma abordagem utilizada na aleatorização estratificada. No caso de efeitos heterogêneos, a amostra pode ser dividida entre os grupos para os quais se deseja medir a heterogeneidade do tratamento, e o MDE do efeito total é obtido por meio de uma média ponderada dos MDEs de cada grupo. Por exemplo, ao analisar a heterogeneidade do efeito de matrícula em uma escola de ensino técnico entre diferentes níveis de renda dos alunos, basta calcular os MDEs separadamente para cada estrato de

¹² *Matched-pairs randomization.*

renda.

No caso de múltiplos tratamentos, o MDE é calculado separadamente para cada grupo de tratamento. No exemplo empírico, se houver dois tipos distintos de escolas de ensino técnico, é possível calcular o MDE para cada uma delas individualmente.

5.1 PRATICANDO: SIMULAÇÕES

Nesta seção, o MDE para experimentos estratificados será calculado exclusivamente por meio de simulações. Na aleatorização estratificada, as variâncias dos erros diferem entre os blocos. Dessa forma, a fórmula analítica é útil para calcular o MDE em termos de desvio padrão em um bloco específico, mas não para o resultado agregado.¹³

O procedimento inicia com a limpeza do ambiente de trabalho e o carregamento dos pacotes necessários. Utiliza-se o pacote `lfe`, para estimar regressões com efeitos fixos. Define-se uma *seed* para reprodução dos resultados e, finalmente, carrega-se a base de dados `base_rct_block.csv`.

```
# Limpar o ambiente
rm(list=ls())

# Carregando pacotes necessários

library(tidyverse)
library(randomizr)
library(lfe)

# Fazendo com que as simulações se tornem reproduzíveis
set.seed(42)

# Base de dados
base_rct_block <- read.csv("Base/base_rct_block.csv")
```

¹³Para calcular o MDE pela fórmula analítica em um grupo específico, basta replicar o procedimento da Seção 2.2 com os valores correspondentes ao grupo em questão.

Na simulação, emprega-se a função `block_ra` do pacote `randomizr` para realizar a aleatorização em blocos. A base de dados contém diversas escolas distribuídas por diferentes municípios. A estratificação é realizada ao nível municipal, com 50% das escolas em cada município sendo aleatorizadas para receber o programa de ensino técnico. Considerando a estratificação por município, a análise requer uma regressão com efeitos fixos municipais. Novamente, os erros padrão são clusterizados ao nível escolar, já que este também é um *Clustered RCT*. Ao final do processo, armazenam-se os erros padrão estimados em cada uma das 100 iterações, calculando-se o MDE com base na média dos erros padrão obtidos.

```
# Criando vetor com 100 entradas para armazenar os erros padr es
se_block <- c(rep(NA, 100))

# Simulando 100 vezes o erro padrão
for (i in 1:100) {

  # aleatorizando por blocos (municípios)
  df <- base_rct_block %>%
    mutate(D = block_ra(blocks=id_mun, prob=0.5))

  reg_block <- felm(y~D|id_mun|0|id_escola,data=df) #regressão com
    efeito fixo de mun
  se_block[i] <- reg_block$se[1] # armazenando erro padrão

}

# Calculando MDE
MDE_block = 2.8*mean(se_block)
print(MDE_block)
```

6 MÚLTIPLAS HIPÓTESES

Em diversas pesquisas, o interesse recai sobre o efeito do tratamento em múltiplas variáveis de resultado. Nesses casos, recomenda-se a correção para testes de múltiplas hipóteses. Ao testar hipóteses para diferentes variáveis, a probabilidade de rejeição de pelo menos uma hipótese nula verdadeira torna-se superior ao nível de significância, α , estabelecido para testes individuais. ao conduzir 10 testes independentes com nível de significância de 5%, a probabilidade de rejeitar pelo menos uma hipótese nula verdadeira aproxima-se de 40%¹⁴.

Um método comum para correção de múltiplas hipóteses consiste na divisão das variáveis de interesse em famílias distintas. A tabela 6.1 ilustra essa abordagem no contexto do exemplo empírico utilizado nesse guia. A **Família I** se refere às variáveis de mercado de trabalho, enquanto a **Família II** engloba as variáveis de desempenho escolar. Essa estratificação por famílias permite testar se o efeito do tratamento difere significativamente de zero para cada grupo de variáveis (veja Anderson, 2008).

Table 6.1: FAMÍLIAS DE VARIÁVEIS DE INTERESSE

	Tipo	Variáveis
Família I.	Mercado de Trabalho	1. Está empregado ou não 2. Formal ou informal
Família II.	Desempenho Escolar	3. Nota na prova 4. Frequência 5. Está matriculado ou não

Para incorporar o teste de múltiplas hipóteses no cálculo do poder estatístico, pode-se ajustar o nível de significância utilizado na estimação do MDE. Uma abordagem prática consiste em modificar o nível de significância de acordo com o tamanho de cada família de variáveis, aplicando a correção de Šidák:

$$\alpha_h = 1 - (1 - 0.05)^{\frac{1}{h}} \quad (6.1)$$

¹⁴A probabilidade de não rejeitar nenhuma hipótese nula verdadeira corresponde a $(1 - 0.05)^{10}$. Portanto, a probabilidade de rejeitar pelo menos uma hipótese nula verdadeira é dada por $1 - (1 - 0.05)^{10} \approx 40\%$.

onde h representa o número de variáveis na família de hipóteses sendo testadas. Observe que quando $h = 1$, obtém-se o valor de $\alpha = 0.05$. À medida que h aumenta, utilizado diminui, garantindo que a probabilidade de rejeitar pelo menos uma hipótese nula verdadeira permaneça em 5%¹⁵. Existem outros métodos que podem ser utilizados, como a correção de Bonferroni ou a correção de Holm (veja [Anderson, 2008](#)). Contudo, nos exemplos empíricos apresentados na próxima seção, será utilizada exclusivamente a correção de Šidák.

6.1 PRATICANDO: FÓRMULA ANALÍTICA

Nesta seção, o MDE para múltiplas hipóteses é calculado exclusivamente mediante a fórmula analítica. Considera-se apenas a correção do nível de significância, sem abordar em detalhes os procedimentos para testes de múltiplas hipóteses¹⁶.

O processo inicia com a preparação do ambiente de trabalho, envolvendo sua limpeza, o carregamento dos pacotes necessários e a definição dos parâmetros constantes no cálculo do MDE.

```
# Limpar o ambiente
rm(list=ls())

# Carregando pacotes necessários
library(tidyverse)

# Parâmetros para cálculo analítico do MDE
t.kappa = 0.84 # poder de 80%
t.alpha = 1.96 # nível de significância de 5%
P = 0.5 # Porcentagem de tratados de 50%
```

¹⁵Isso acontece quando as hipóteses são independentes. Caso elas tenham correlação positiva, o teste é conservador e, se a correlação for negativa, o teste é inválido. Uma maneira comum de evitar correlações negativas entre os teste é mudar o sinal das variáveis para que sua direção positiva indique uma "melhoria" na variável (veja [Anderson, 2008](#)).

¹⁶Diversas abordagens existem para testar múltiplas hipóteses conjuntamente. Na prática, além da correção do nível de significância, é necessário considerar métodos adequados para agregação das variáveis nos testes conjuntos. Maiores detalhes podem ser encontrados em [Anderson \(2008\)](#).

Em seguida, cria-se uma função para calcular o nível de significância ajustado e outra para calcular o valor crítico do nível de significância ajustado, dependendo do número de variáveis dentro de uma família, h .

```
# Correção de Sidák

alpha.h = function(h){1 - (1 - 0.05)^(1/h)}

# Valor crítico variando com h

t.alpha.h = function(h){qnorm(1-alpha.h(h)/2)}
```

Com a função do valor crítico ajustado, é possível criar uma função para o cálculo do MDE variando tanto o tamanho da amostra, N , quanto o número de hipóteses a serem testadas conjuntamente, h .

```
# Fórmula analítica variando N e h

MDE.h = function(N,h){(t.kappa+t.alpha.h(h))*((1/(P*(1-P)))^0.5)*
  ((1/N)^0.5)}
```

O bloco de código abaixo gera a Figura 6.1, que ilustra a relação entre o MDE (expresso em desvios padrão) e o tamanho amostral, variando de 500 a 3500 observações. O gráfico apresenta três curvas distintas, correspondentes a diferentes números de hipóteses testadas simultaneamente. Como evidenciado pelos resultados, o MDE aumenta proporcionalmente ao número de hipóteses conjuntas analisadas.

```
# Gráfico do MDE variando N, para h = {1,2,3}.

ggplot() + xlim(500, 3500) +
  geom_function(fun = MDE.h, args = list(h = 1), aes(colour = "h =
```

```

1"), linewidth = 0.8) +
geom_function(fun = MDE.h, args = list(h = 2), aes(colour = "h =
2"), linewidth = 0.8) +
geom_function(fun = MDE.h, args = list(h = 3), aes(colour = "h =
3"), linewidth = 0.8) +
labs(y = "MDE", x = "N") +
theme_classic() +
guides(colour=guide_legend(title=""))

```

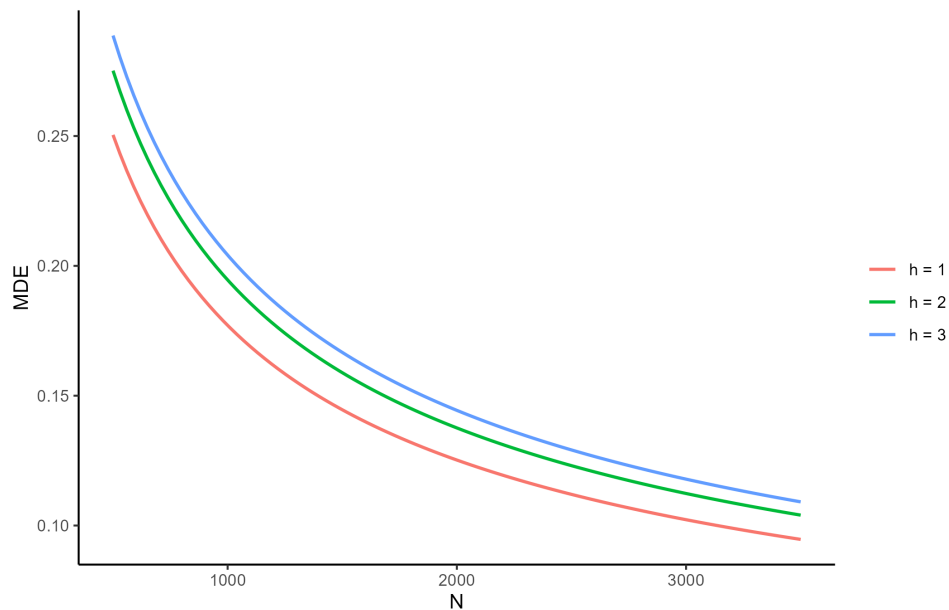


Figure 6.1: Gráfico do MDE em termos de desvio padrão variando o tamanho da amostra de 500 a 3500 para diferentes valores de hipóteses a serem testadas conjuntamente.

7 CONCLUSÃO

O cálculo do poder estatístico é uma etapa fundamental no planejamento de experimentos, pois garante que os recursos sejam utilizados de forma eficiente e que os resultados obtidos sejam confiáveis. Um experimento com baixo poder estatístico pode levar a conclusões equivocadas, seja pela incapacidade de detectar um efeito real (**Erro Tipo II**) ou pela superestimação de um efeito observado em amostras pequenas.

Pesquisadores normalmente buscam desenhar experimentos com poder estatístico de

80%, ou seja, quando há um efeito real, o experimento tem 80% de chance de detectá-lo. Esse valor depende de parâmetros que podem ser controlados na fase de planejamento, como o tamanho da amostra e a proporção de indivíduos tratados, e de parâmetros fora do controle, como a variância da variável de interesse. Por isso, a análise prévia ao experimento é essencial para garantir a credibilidade dos resultados obtidos posteriormente.

Este guia apresenta detalhadamente a análise de poder estatístico para diversos delineamentos experimentais. Observa-se que, à medida que a complexidade dos desenhos experimentais aumenta, maior se torna a presença de ruído, resultando consequentemente na redução do poder estatístico.

Existem duas formas de realizar a análise de poder: a fórmula analítica ou a utilização de simulações. Essencialmente, o cálculo do MDE por meio da fórmula analítica requer o conhecimento prévio da expressão matemática para a variância do estimador empregado. Este método é útil na ausência de dados antes da realização do experimento, permitindo a estimação do MDE em termos de desvio padrão.

Quando há acesso aos dados, o MDE pode ser calculado mediante simulações. Através da simulação da alocação entre grupos de tratamento e controle, é possível estimar a variância do estimador a ser utilizado. Esse método mostra-se vantajoso quando as fórmulas analíticas da variância do estimador são muito complexas.

A determinação prévia do poder estatístico permite o ajuste do tamanho amostral, a consideração da variabilidade dos dados e a garantia de detectabilidade dos efeitos de interesse com confiança estatística adequada. Essa abordagem é especialmente relevante em estudos aplicados, onde limitações de recursos financeiros, temporais e de disponibilidade de participantes impõem restrições ao desenho experimental.

Por fim, o planejamento adequado do poder estatístico aprimora a reprodutibilidade dos estudos, reduz o risco de obter resultados inconclusivos e contribui para a credibilidade das análises realizadas. Dessa forma, o cálculo do poder não é apenas uma formalidade estatística, mas uma prática essencial para conduzir experimentos robustos e com impacto significativo.

REFERENCES

- ANDERSON, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the abecedarian, perry preschool, and early training projects. *Journal of the American Statistical Association*, 103(484):1481–1495.
- ANGRIST, J. D. AND PISCHKE, J.-S. (2008). Mostly Harmless Econometrics. *Princeton University Press*.
- BAI, Y., ROMANO, J. P., AND SHAIKH, A. M. (2022). Inference in experiments with matched pairs. *Journal of the American Statistical Association*, 117(540):1726–1737.
- DUFLO, E., GLENNERSTER, R., AND KREMER, M. (2007). Using randomization in development economics research: A toolkit. *Handbook of development economics*, 4:3895–3962.