

Atividade Prática Supervisionada 3

Microeconomia IV

João Gabriel Cerqueira Gomes

Pedro Xavier Marino

Victor Manoel Alves

Pedro Ongaratto

João Alonso Casella

Inspier - Instituto de Ensino e Pesquisa

2022

PARTE 1

A) Em linha com o modelo de formação de habilidades de Cunha & Heckman visto em aula, a teoria de Attanasio (2015) propõe um modelo de produção de capital humano, com foco especialmente na acumulação de capital humano nos primeiros anos de vida. Estruture formalmente (com equações e suas palavras) o modelo microeconômico de Attanasio (2015), deixando claro quais as premissas, o problema de maximização dos pais, e qual o principal resultado teórico do modelo. Finalize sua exposição elencando a hipótese econômica do artigo. (Dica: Em sua resolução, sugerimos que tenha em mente qual a pergunta de investigação deste artigo. Não é necessário trazer o desenvolvimento integral do modelo. A ideia aqui também não é fazer um resumo do modelo, e sim estruturá-lo para ter um início, meio e fim. O desfecho será a hipótese econômica.)

O artigo de Attanasio (2015) busca identificar e analisar o impacto de intervenções e restrições de investimento por parte dos pais/responsáveis para o capital humano das pessoas, quando feitas nos primeiros anos de vida. O autor, na intenção de evitar variadas inconsistências, levou em consideração uma série de premissas para a estruturação do modelo microeconômico.

Attanasio procura com seu modelo microeconômico entender como o capital humano como um objetivo de estudo, caracterizado por cognição e habilidades socioemocionais e saúde, evolui ao longo do desenvolvimento da criança associando fatores que são responsáveis por isso. Dentre os fatores observáveis, o autor denota que a função de produção do capital humano depende do nível inicial de capital humano, background socioeconômico das crianças, expõe as diferentes influências que o pai e a mãe têm no desenvolvimento do capital humano da criança, dos investimentos, capital, despendido em recursos humanos.

Vale ressaltar que o Capital, inclui tanto em questões materiais como brinquedos e livros e tempo gasto pelos pais com as crianças e é bem mais complexo do que isso ainda, pois também são influenciadas pelas *crenças sobre práticas parentais ideais*, ademais, *a decisão dos pais de comprar objetos para seus filhos depende do quão eficaz eles veem essas atividades no desenvolvimento da criança e se acreditam se compensa passar tempo com elas em relação ao*

tempo que seria utilizado com trabalho e lazer. Por último, um vetor de choques aleatórios que não são observados diretamente. Dessa maneira, dispõe-se uma breve legenda das variáveis utilizadas e fatores de influência escolhidos pelo autor para o modelo:

Variáveis utilizadas:

$H_{i,t}$ = capital humano da criança

$Z_{i,t}$ = background dos responsáveis

$X_{i,t}$ = investimento em capital humano

$e_{i,t}^H$ = choques

i = famílias

t = idade da criança

Fatores de influência:

c = habilidades cognitivas

s = habilidades socioemocionais

h = saúde

M = materiais

T = tempo

m = background da mãe

f = background do pai

r = outros

Dados os fatores de influência, dispõe-se, também, como estes se relacionam com as variáveis utilizadas:

$$H_{i,t+1} = g_t(H_{i,t}, X_{i,t}, Z_{i,t}, e_{i,t}^H)$$

$$H_{i,t} = \{\theta_{i,t}^c, \theta_{i,t}^s, \theta_{i,t}^h\}$$

$$Z_{i,t} = \{\theta_{i,t}^m, \theta_{i,t}^f, \theta_{i,t}^r\}$$

$$X_{i,t} = \{\theta_{i,t}^M, \theta_{i,t}^T\}$$

Assim, estruturou-se o seguinte modelo de maximização da utilidade dos pais, e função de investimento:

Função Maximização Utilidade dos Pais

$$\max_{\{C_{i,t}, X_{i,t}\}} U(C_{i,t}, H_{i,t+1})$$

Sujeito a:

$$C_{i,t} + P_t^X X_{i,t} = Y_{i,t} \quad e \quad H_{i,t+1} = g_t(H_{i,t}, X_{i,t}, Z_{i,t}, e_{i,t})$$

Função de Investimento

$$X_{i,t} = (H_t^i, P_t^i, Z_t^i, e_{i,t}^X, Y_t^i, \pi)$$

O autor usa-se de um modelo simples onde não existe poupança e os pais somente tem um único filho para descrever a função de maximização a qual os pais querem maximizar sua utilidade que está sujeita ao seu consumo e o capital humano das crianças, restrito ao seu orçamento, o quanto eles podem investir de capital em seus filhos, e restrito também à função de capital humano descrita anteriormente, ou seja, o capital humano irá aumentar estritamente de acordo com essa função.

Uma vez construído o modelo microeconômico, evidencia-se o seu principal resultado teórico: confirma-se a noção de que o investimento feito pelos pais é fator relevante na definição do capital humano de uma criança nos primeiros anos de vida e que esse investimento é moldado pela maneira à qual os pais interpretam a natureza da função de produção. Foi também obtido que o investimento em materiais afeta o desenvolvimento cognitivo da criança; que a disposição de tempo na relação entre pais e filhos afeta o desenvolvimento de fatores sócio-emocionais; e que o histórico dos pais tem efeito relevante no nível de investimento.

Resultado importante do modelo foi que os investimentos são endógenos, pois, caso ocorra uma intervenção por meio das políticas públicas, as percepções dos pais em relação à importância do capital humano muda. Assim, eles se tornam mais propensos a gastar mais dinheiro e tempo com

as crianças; e o investimento dos pais em tempo e material é concluído como consideravelmente maior nas aldeias de tratamento do que nas de controle.

Por fim, propõe-se a seguinte hipótese econômica para o modelo teórico microeconômico disposto no artigo, implícita pelo discurso do autor: Com base no modelo proposto por Attanasio et al. (2015a), aumentos no investimento feito pelos pais, tanto em questões monetárias quanto em tempo (de qualidade) despendido com a criança em seus primeiros anos de vida, resultarão em aumentos no capital humano da mesma.

B) Explique, com as suas palavras, qual a estratégia de identificação empregada na análise empírica do artigo de Afridi (2010). Explique também o racional que baseou a exposição dos resultados nas Tabelas 2 a 9. Em seguida, interprete ao menos 1 coeficiente estimado (aquele que seu Grupo julgar mais importante para a análise) de cada Tabela. Finalize identificando qual é a principal tabela e o principal resultado deste artigo. Justifique.

A estratégia de identificação se divide em três partes.

Primeiramente, é necessário dizer que existe mais de uma variável resposta no estudo. Todas elas são nutrientes ingeridos pelas crianças analisadas e se dividem em Calorias, Carboidratos, Proteínas, Cálcio, Ferro e Zinco. A principal variável explicativa é a refeição escolar promovida pela escola, e seu efeito na ingestão de cada um dos nutrientes.

Sendo assim, a primeira parte da estratégia empírica utilizada foi uma análise dos dados em cross-section. Nessa parte, foram feitas quatro regressões, todas em 2LSL, cada uma delas se diferenciando das outras por alguma mudança estrutural, como a introdução de uma dummy ou utilização de um instrumento. Na primeira regressão, por exemplo, uma dummy para dia escolar foi utilizada como instrumento para presença escolar.

Na segunda parte, os dados em painel foram realmente considerados em um modelo de efeitos fixos para comunidades, ou tribos como chama o estudo. Aqui são feitas mais três regressões, uma em OLS e as outras duas em 2LSL, variando novamente entre a utilização de dummies e

instrumentos específicos ou não. O objetivo dessa parte da estratégia é identificar efeitos não observáveis entre diferentes comunidades do estudo e compará-los com os outros achados.

Por fim, a terceira parte, e a que a autora considera de suma importância, regride os dados em um modelo de efeito fixo individual, analisando as crianças questionadas pelo estudo. Nesse modelo, é possível capturar o efeito do programa na ingestão de nutrientes, em média, de cada uma das crianças afetadas. A regressão dessa parte foi apenas uma, e foi um mínimo quadrados ordinários.

Dentro das duas últimas regressões da segunda parte e da regressão da terceira parte, o estudo ainda especifica resultados para horas do dia. A autora, portanto, faz essas três regressões apenas para as horas cujas quais a criança se encontra na escola, e faz essas três regressões apenas para as horas depois da escola, quando a criança normalmente está em casa.

Para testar o efeito de outras variáveis na transferência de calorias do programa para a criança, a autora interagiu o consumo de calorias diárias com sexo, idade e tamanho da família. Os resultados dessas interações poderiam ser positivos ou negativos, dando a indicar como essas variáveis afetam a alocação de recursos dentro de uma família. O estudo faz isso tanto para o consumo total de calorias quanto para o consumo de calorias apenas em horário escolar.

Considerando por todo tipo de endogeneidade com instrumentos, o estudo garante a hipótese de endogeneidade estrita. Todas as outras suposições são consideradas pelos seguintes fatores:

Madhya Pradesh, a localização indiana escolhida para o estudo, compôs um experimento natural. O fato de que algumas escolas nessa localidade em específico adotaram um programa nacional de alimentação diferente do usado anteriormente, e outras não, criaram um perfeito grupo controle e grupo teste. Primeiro, é possível comparar a mesma escola antes e depois da adoção do programa, e ainda é possível comparar o resultado de escolas que adotaram o programa com escolas que nunca adotaram.

O questionário utilizado para a montagem da base de dados foi posto em prática em dias escolares e não escolares de maneira aleatória, eliminando qualquer viés na diferença dos dias e garantindo a randomização da amostra.

As crianças do estudo foram questionadas sobre sua ingestão de calorias em dois dias em sequência, sendo um dia escolar e o outro não. E foi possível acompanhar a mesma criança entre um dia e outro, usando os dados de painel. Isso permite utilizar o modelo de efeitos fixos individuais, o que potencializa os resultados estatísticos a respeito da eficiência do programa alimentar, sendo possível analisar o impacto da política em uma mesma criança, e não na média de crianças. Isso elimina heterogeneidades individuais que possam comprometer os resultados.

O questionário ainda foi feito em um período de dois meses, não dando espaço a nenhum tipo de contaminação por mudanças sazonais que possam ser causadas por diferentes estações ou mudanças políticas. Esse método minimiza o viés nos resultados.

O estudo ainda analisa, por meio dos questionários, a qualidade dos nutrientes sendo ingeridos pelas crianças questionadas, ao invés de apenas questionar a quantidade de calorias ingeridas. Isso dá espaço para comparar a qualidade de comida ingerida em dias escolares e dias não escolares, quando a criança come com a família.

Por acaso, a maneira como o programa alimentar é conduzido na Índia, com refeições sendo preparadas e entregues na mão da criança que necessita, permite uma análise de impacto de política pública onde o público alvo realmente recebe diretamente o que foi proposto.

Tabela 2

A tabela 2 consiste de um sumário de estatísticas das variáveis da pesquisa. Além disso, mostra a diferença desses dados para um dia sem escola (*Non-school day*) e um dia com escola (*School day*).

Variable	Type of day		
	Non-school day	School day	Difference
	(N = 182)	(N = 723)	
	(1)	(2)	

Total annual household income (Rs.)	21,416.94 (1011.09)	20,776.35 (680.142)	640.59 (1447.65)
-------------------------------------	---------------------	---------------------	------------------

Essa estatística representa a média da renda anual das famílias das crianças que participaram da pesquisa.

Tabela 3

A tabela 3 mostra as médias do consumo total de determinados nutrientes no dia de referência e o montante desses respectivos nutrientes que foram advindas da merenda escolar.

Nutrient (unit)	Total daily consumption	Consumption from school meal	Percentage of total daily consumption from school meal
(N=448)	(1)	(2)	[(2)/(1)]*100
Calories (kcal)	1379.80 (440.99)	263.06 (128.22)	19.07

Neste caso, a tabela mostra que, em dias escolares, em média, 19,07% das calorias consumidas por uma criança participante do programa vem da refeição escolar.

Tabela 4

A tabela 4 traz à tona dados referentes às atividades diárias feitas pelas crianças em um dia escolar e não-escolar. O objetivo de interpretar esses dados é para identificar a diferença entre a quantidade necessária de calorias em um dia escolar e não-escolar.

Activity	Mean hours per day		Difference
	Non-school day (N= 183)	School day (N= 747)	
Livestock care	0.47 (0.074)	0.36 (0.032)	0.11 (0.074)

Aqui podemos observar que em dias não escolares as crianças, em média, cuidam dos animais 11% a mais do que nos dias que elas atendem a escola. Isso pode significar que uma criança precisaria de mais calorias em um dia não-escolar do que escolar.

Tabela 5

A tabela 5 mostra o consumo diário de nutrientes de crianças que não fazem parte do programa de merendas em dias escolares e não-escolares.

Nutrient (unit)	Non-school day (N=93)	School day (N=143)	Difference
Calories (kcal)	1335.53 (44.347)	1254.45 (32.727)	81.08 (54.087)

Esse dado pode ser visto como uma evidência para provar o que foi descrito na tabela 4, sendo que ele mostra que o consumo de calorias é, em média, maior em dias onde a criança não atende a escola. O que, como visto na tabela 4, pode ser explicado pela maior demanda de atividades físicas presentes no dia-dia da criança em dias não-escolares.

Tabela 6

Essa tabela demonstra o impacto da refeição escolar no consumo diário total de nutrientes de um indivíduo.

Specification	Coefficient on quantity of nutrient intake from school meal					N
	Calories	Carbohydrates	Proteins	Calcium	Iron	
<i>Individual fixed effects</i>						
(7) OLS	0.76* (0.405)	0.66 (0.399)	0.62 (0.443)	0.15 (0.671)	0.43 (0.342)	546

Esse coeficiente no valor de 0,76 representa que para cada 1 quilograma de caloria que o Estado provê para a criança se é observado, em média, um aumento de 0,76 quilogramas de caloria na dieta total desse mesmo indivíduo.

Entretanto, vale ressaltar que o nível de significancia desse coeficiente eh de 10%.

Tabela 7

A sétima tabela mostra o impacto da refeição escolar no consumo total de nutrientes durante o período escolar.

Specification	Calories		Carbohydrates		Proteins		Calcium		Iron		N
	Quantity of nutrient from school meal	School attendance	Quantity of nutrient from school meal	School attendance	Quantity of nutrient from school meal	School attendance	Quantity of nutrient from school meal	School attendance	Quantity of nutrient from school meal	School attendance	
(7) OLS-FE	0.89*** (0.229)	-233.12*** (55.927)	0.85*** (0.214)	-42.61*** (10.119)	0.83*** (0.216)	-6.74*** (1.722)	0.75*** (0.206)	-15.73*** (4.499)	0.68*** (0.209)	-1.32** (0.597)	500

Aqui devemos interpretar apenas o sinal desses coeficientes. Assim, isso indica que os indivíduos que atendem uma escola que faz parte do programa de merendas têm um maior consumo de calorias durante o período escolar, já que 0,89 apresenta um sinal positivo.

Tabela 8

A oitava tabela mostra o impacto da refeição escolar no consumo total de nutrientes durante o período não-escolar.

Specification	Calories		Carbohydrates		Proteins		Calcium		Iron		N
	Quantity of nutrient from school meal	School attendance	Quantity of nutrient from school meal	School attendance	Quantity of nutrient from school meal	School attendance	Quantity of nutrient from school meal	School attendance	Quantity of nutrient from school meal	School attendance	
(7) OLS-FE	-0.06 (0.356)	106.64 (73.777)	-0.13 (0.351)	23.43 (12.082)	-0.17 (0.386)	5.08* (2.561)	-0.39 (0.604)	20.14 (15.729)	-0.33 (0.318)	2.18** (1.056)	500

Assim como a tabela 7, o fator que importa nesses dados são os sinais. Dessa forma, em concordância com os resultados citados acima, o sinal negativo no coeficiente de calorias representa que os alunos têm um consumo menor de calorias durante o período não-escolar.

Tabela 9

A tabela 9 apresenta os determinantes da alocação de recursos dentro de uma família (*Household*).

Interaction of quantity of calorie intake from school meal with...	Dependent variables			
	Total individual daily calorie intake		Total individual calorie intake during school hours	
	(1)	(2)	(3)	(4)
Household size	0.20***	0.28**	0.03	0.02

O único efeito no consumo diário de calorias da tabela 9 que o autor julgou significativo é o do tamanho da *Household*, que se mostra positivo. Portanto, quanto maior o tamanho da *Household* maior seria a alocação de recursos representada pelo valor monetário de refeições escolares que chegam na criança.

A tabela principal do artigo é a 6, já que mede diretamente, em média, o quão impactante é o programa de refeições escolares na nutrição de uma criança.

C) Explique por que o Experimento investigado por Dynarski et al. (2013), de alocação dos alunos em turmas de diferentes tamanhos, é considerado um marco importante na identificação do efeito causal do tamanho da turma no desempenho educacional dos alunos. Explique quais seriam as limitações se, ao invés de realizar um experimento aleatório, fossem coletados dados observacionais como do Censo Escolar (que possui características dos estudantes e escolas) e da Prova Brasil (que contém as notas dos estudantes) para identificar o efeito do tamanho da turma nas notas dos estudantes.

O modelo proposto em Dynarski et al. (2013) aborda métodos econométricos para verificar os efeitos de tamanho da turma no desempenho educacional dos alunos, em especial seu efeito sobre a entrada no ensino superior e escolha de curso. Os autores utilizam da base de dados STAR, a qual acompanha os alunos participantes do projeto e por meio de dados da NCS observam o desempenho destes alunos ao longo da sua trajetória acadêmica.

Ademais, os autores também propõem a utilização de variáveis instrumentais como forma de traçar outros efeitos. Por exemplo, eles propõem o uso da variável instrumental de anos que poderiam ter sido passados em classes pequenas para regredir a variável de anos passados em classes pequenas. Tal estratégia tem como objetivo contornar possíveis efeitos advindos da alta exposição do indivíduo ao programa, e também na possível presença de seleção em variáveis que não são observadas mas afetarão a exposição ao programa, afetando assim os resultados potenciais do programa.

Diferente dos dados do projeto STAR, a utilização de dados não formulados por meio de um experimento aleatório, como o Censo Escolar e a Prova Brasil, se dá logo de início no processo de identificação dos indivíduos tratados e de controle. Caso estes não sejam aleatorizados e os grupos não sejam semelhantes na média, torna-se extremamente difícil a observação de efeitos médios sobre características dos indivíduos, invalidando a observação de um contrafactual. Como ambos testes são aplicados sobre o grupo geral de alunos de ensino público (um grupo heterogêneo), a observação de efeitos de um possível programa de tamanho de turmas nestes alunos não seria identificável.

PARTE 2

A) Estabeleça a Pergunta de Pesquisa a ser investigada por sua Equipe com base nas informações disponíveis. Formalize uma Teoria Microeconômica que fundamente os argumentos teóricos da sua pergunta. Deixe claro quais as referências da literatura que serviram de base para a construção de seu argumento teórico. Identifique a Hipótese econômica resultante de sua Teoria Micro.

Entendido que a neurociência da pedagogia tem certo consenso na preferência por turmas menores do que turmas maiores, podemos desenvolver uma Pergunta de Pesquisa que seria dada por “Seriam salas menores melhores para o aprendizado de cada aluno?”. Para resolver essa pergunta, vamos nos basear em um modelo pré-existente.

No modelo de Edward Lazear, o tamanho da sala influencia a probabilidade de uma sala de aula prestar atenção à aula. Essa atenção leva a uma melhora no desempenho dos alunos, o que inevitavelmente aumenta o lucro da escola como um todo, visto que ela se prova melhor em prover o serviço que oferece. O foco de Lazear é justamente no lucro da escola, como dado por:

$$\pi = Z * V * (p^n) - W * m$$

Onde ‘p’ é a probabilidade de um aluno em específico se comportar, e ‘n’ é o número de alunos por sala. A potência resultante dessas duas variáveis gera uma variável que representa o quanto cada sala como um todo está prestando atenção na aula, no geral. É essa potência, por consequência do funcionamento do mercado, que vai ter efeito positivo no lucro.

A fim de analisar o impacto do tamanho de uma sala de aula no desempenho de um aluno em específico, é possível utilizar essa variável criada por Lazear. Isto porque, quanto mais a sala de

aula desse aluno foca na aula, menos esse aluno é prejudicado pelos seus colegas, e mais ele mesmo presta atenção na aula, o que acaba por aumentar seu desempenho.

Além disso, o aprendizado de um aluno em específico também é dado por variáveis específicas desse aluno, intrínseca à sua genética e seus estímulos quando criança e criação parental. No modelo, esses fatores serão capturados por uma variável Alpha. A renda do aluno também aumenta sua capacidade de aprendizado, aumentando sua habilidade de procurar conhecimento, seu tempo livre para estudar, e diminuir seu estresse cotidiano, além de que alunos de melhor renda estão melhor alimentados. A renda será capturada no modelo pela variável Psi. Uma função E dada por Alpha e Psi captura os efeitos específicos a cada aluno.

É possível que o aprendizado μ_i de cada aluno i dentro de uma escola qualquer com n alunos por turma, e uma função E qualquer, seja dado por:

$$\mu_i = (p^n) * E(\alpha_i, \psi_i)$$

Podemos ir além, e desenvolver a probabilidade dos alunos se comportarem (p), como a razão de professores por aluno. Assim, a probabilidade ‘ p ’ deixa de ser dada exógenamente e passa a depender das variáveis presentes dentro da sala de aula em si. Quanto mais professores por aluno, maior a concentração de alunos. Sendo a quantidade de professores por sala dado por ‘ m ’, então temos:

$$\mu_i = \left(\frac{m}{n}\right)^n * E(\alpha_i, \psi_i)$$

Se derivarmos ‘ μ_i ’ em ‘ n ’, podemos encontrar qual a relação do número de alunos no aprendizado de cada aluno.

$$\mu'(n) = \frac{E * m}{n} * \left[\ln\left(\frac{m}{n}\right) - 1 \right]$$

Sendo E uma variável sempre positiva, m e n também sempre sendo variáveis positivas (não é possível ter professores ou alunos negativos por sala), temos que o primeiro termo $\frac{Em}{n^n}$ é sempre positivo. Se o segundo termo da equação, portanto, for negativo, então a derivada é negativa; e se for positivo, a derivada é positiva.

Para que ela seja positiva $\ln(m/n) > 1$, e isso se mantém sempre que $m > n$. E para ser negativa, $\ln(m/n) < 1$, e isso se mantém sempre que $m < n$.

Sob a adequada suposição de que o número de alunos sempre excede o número de professores dentro de uma sala de aula, podemos assumir que essa derivada é sempre negativa.

$$\frac{\delta\mu}{\delta n} < 0$$

Baseado nisso, podemos assumir uma hipótese econômica que deriva diretamente dos achados do modelo microeconômico:

“Quanto maior o número de alunos, em média, por sala, menor a capacidade de aprendizado de cada aluno individual”.

B) Elabore algumas estatísticas descritivas para caracterizar os três tipos de turmas.

Observe que, com exceção das variáveis de nota, as médias das demais variáveis são muito próximas independentemente do tipo de turma. Por quê?

Observando os gráficos 2 ao 8 nota-se que os três grupos apresentam comportamentos similares para a maioria das variáveis observadas, em especial quando se olha para o comportamento da média das variáveis numéricas. Tal comportamento é esperado, já que para o bom funcionamento da aleatorização do programa, os grupos precisam apresentar características similares na média, permitindo a verificação de uma mudança média pela implementação de um determinado tratamento. Por exemplo, observando o gráfico 2 é facilmente observável que o número de faltas por alunos nos três grupos é bastante próximo, mostrando descritivamente uma possível

semelhança dos grupos. Outro exemplo pode ser observado pelo gráfico 7, sendo notável que a proporção de meninos e meninas das diferentes classes é similar.

C) Considere os seguintes modelos:

$$\begin{cases} \text{Reading score}_i = \alpha_0 + \alpha_1 \text{small}_i + \alpha_2 \text{aide}_i + \sum_{k=3}^K \alpha_k x_{ki} + \varepsilon_{1i} \\ \text{Math score}_i = \beta_0 + \beta_1 \text{small}_i + \beta_2 \text{aide}_i + \sum_{k=3}^K \beta_k x_{ki} + \varepsilon_{2i} \end{cases}$$

Estime os modelos incluindo também as variáveis x_{3i} , x_{4i} , ..., x_{Ki} – fatores de sua escolha que expliquem as notas dos alunos. Explique suas escolhas e interprete os resultados estimados. Neste caso, a não inclusão destes fatores é fonte de viés? Justifique.

Realizando a regressão de *readscore* em função do tamanho da sala, presença de um professor assistente, experiência do professor, presença em aulas, gênero, antecedência asiática, lanche grátis e a variável “*schurban*”. Escolheu-se as variáveis para analisar se esses respectivos fatores afetam o desempenho do aluno com foco no tamanho da sala para verificar se isso influencia nas suas notas. Observando a figura 1, nos anexos, nota-se que o coeficiente de *small* é de 5,734, ou seja, positivo para essa dummy. Nesse sentido, a sala *small* que é definida por ter entre 13 a 17 alunos, tem relação positiva com o desempenho do aluno. Ademais, a não inclusão das demais variáveis *x*, como experiência do professor, presença em aulas, gênero, etc faz com que esses fatores fiquem no termo de erro, nesse sentido se o termo de erro é significativamente relacionado com o desempenho então teremos endogeneidade.

D) Adicione, aos modelos resultantes do item anterior, variáveis indicativas de cada escola:

$$escola_j = \begin{cases} 1, & \text{se o estudante é aluno da escola } j \\ 0, & \text{caso contrário} \end{cases}$$

Explique o que representa a inclusão das variáveis escola nestas regressões. Teste a significância conjunta dos efeitos fixos das escolas. A inclusão destas variáveis altera as estimativas dos coeficientes de *small* e *aide*? Interprete seus resultados.

Ao adicionar efeitos da escola sobre o desempenho do aluno verifica-se na tabela 11 que há um aumento do efeito de *small* e *aide* aumenta. Tal inclusão serve como forma de obter efeitos particulares e intrínsecos de cada escola (como infraestrutura, método de ensino e outras coisas). Este aumento segue um padrão esperado, dado que certas características da escola poderiam estar sendo ignoradas numa regressão que exclua a importância da escola na determinação do desempenho.

Observando a tabela 12 e a tabela 13, nota-se que para o teste F parcial para a regressão referente a teste de leitura e matemática têm-se evidências, a um nível de significância de 5%, de que as variáveis referentes aos efeitos das escolas possuem relevância para explicar o desempenho do aluno em cada um dos testes.

E) Uma forma de checar se randomização dos estudantes foi bem-executada é estimar um modelo de probabilidade linear que regreda *small* sobre as outras características explicativas. **Se a aleatorização foi bem-feita, podemos esperar coeficientes não significativos nesta regressão. Por quê? Estime este modelo e discuta seus resultados.**

A tabela 14 demonstra os resultados da regressão de *small* pelas demais variáveis, nesta pode-se notar que o maior p-valor é o da variável experiência do professor com p-valor menor do que 10%, sendo portanto relevante, para explicar *small*. As variáveis *absent*, *techwhite1* e *tchmaster1* tem relevância caso medida com significância de 5% e *tchexper* tem relevância com significância de 1%, somente *boy1*, *freelunch1*, *whiteasian1* e *schurban1* são não irrelevantes para explicar *small*.

Nesse sentido, observa-se que muitos coeficientes são significantes. Se a aleatorização foi bem feita então o impacto dos regressores em *small* deve ser irrelevante pois nesse caso o viés de seleção seria nulo. O viés de seleção ocorre quando as razões pelas quais a variável *small* é determinada estão correlacionadas a outros fatores de forma que essa seleção não é estritamente

exógena. Garantir que o impacto estimado esteja livre de viés de seleção é um dos principais objetivos e requer que *small* seja exógena. Por conseguinte, devido aos dados da tabela 14, nota-se que vários regressores na regressão de *small* tem relevância estatística, assim, temos que existe viés de seleção, dessa forma, a aleatorização, possivelmente, tem problemas. Nessa perspectiva, somente se todos os coeficientes fossem irrelevantes que não existiria viés de seleção consequentemente gerando uma aleatorização bem feita, como isso não ocorre coloca-se em xeque a eficácia da aleatorização da variável *small*.

<i>Dependent variable:</i>	
<i>as.numeric(small)</i>	
<i>tchexper</i>	−0.002* (0.001)
<i>absent</i>	−0.002*** (0.001)
<i>boy1</i>	−0.002 (0.012)
<i>white_asian1</i>	0.003 (0.018)
<i>tchwhite1</i>	0.067*** (0.018)
<i>tchmasters1</i>	−0.051*** (0.013)
<i>freelunch1</i>	−0.005 (0.014)
<i>schurban1</i>	0.005 (0.016)
Constant	1.302*** (0.025)
Observations	5,745
R ²	0.007
Adjusted R ²	0.006
Residual Std. Error	0.458 (df = 5736)
F Statistic	5.402*** (df = 8; 5736)
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01	

ANEXO

Tabela 10 - Regressões para desempenho em teste de leitura e matemática

	<i>Dependent variable:</i>	
	readscore	mathscore
	(1)	(2)
small1	5.734*** (0.992)	8.170*** (1.500)
aide1	1.036 (0.955)	0.459 (1.444)
tchexper	0.526*** (0.071)	0.654*** (0.108)
absent	-0.288*** (0.044)	-0.471*** (0.066)
boy1	-6.147*** (0.795)	-7.891*** (1.202)
white_asian1	6.585*** (1.172)	10.193*** (1.771)
tchwhite1	-1.724 (1.216)	-6.416*** (1.838)
tchmasters1	-1.499* (0.869)	-2.826** (1.314)
freelunch1	-14.144*** (0.908)	-17.444*** (1.374)
schurban1	1.953* (1.074)	-3.503** (1.623)
Constant	439.638*** (1.694)	494.828*** (2.562)
Observations	5,745	5,745
R ²	0.105	0.095
Adjusted R ²	0.103	0.093
Residual Std. Error (df = 5734)	30.087	45.490
F Statistic (df = 10; 5734)	67.288***	60.125***
Note:	*p<0.1; **p<0.05; ***p<0.01	

Fonte: Elaboração Própria

Tabela 11 - Regressões com efeito fixo de escola para desempenho em teste de leitura e matemática

	<i>Dependent variable:</i>	
	readscore	mathscore
	(1)	(2)
small1	6.392*** (0.917)	9.095*** (1.373)
aide1	1.143 (0.885)	0.947 (1.326)
tchexper	0.292*** (0.073)	0.418*** (0.109)
absent	-0.263*** (0.041)	-0.466*** (0.061)
boy1	-5.439*** (0.727)	-6.695*** (1.089)
white_asian1	8.590*** (1.540)	18.380*** (2.306)
tchwhite1	-0.047 (1.437)	-0.949 (2.153)
tchmasters1	-0.737 (0.956)	-2.194 (1.432)
freelunch1	-13.927*** (0.888)	-18.725*** (1.330)
schurban1	-29.420*** (4.804)	-35.002*** (7.197)
Constant	443.131*** (3.562)	492.740*** (5.335)
Observations	5,745	5,745
R ²	0.271	0.276
Adjusted R ²	0.260	0.265
Residual Std. Error (df = 5657)	27.338	40.951
F Statistic (df = 87; 5657)	24.176***	24.834***
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01		

Fonte: Elaboração Própria

2022

Tabela 12 - Teste de significância para a regressão com efeito fixo de escola para desempenho em teste de leitura

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	5, 734	5,190,612.000				
2	5, 657	4,227,798.000	77	962,814.000	16.731	0

Fonte: Elaboração Própria

2022

Tabela 13 - Teste de significância para a regressão com efeito fixo de escola para desempenho em teste de matemática

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	5, 734	11,865,598.000				
2	5, 657	9,486,636.000	77	2,378,963.000	18.423	0

Fonte: Elaboração Própria

2022

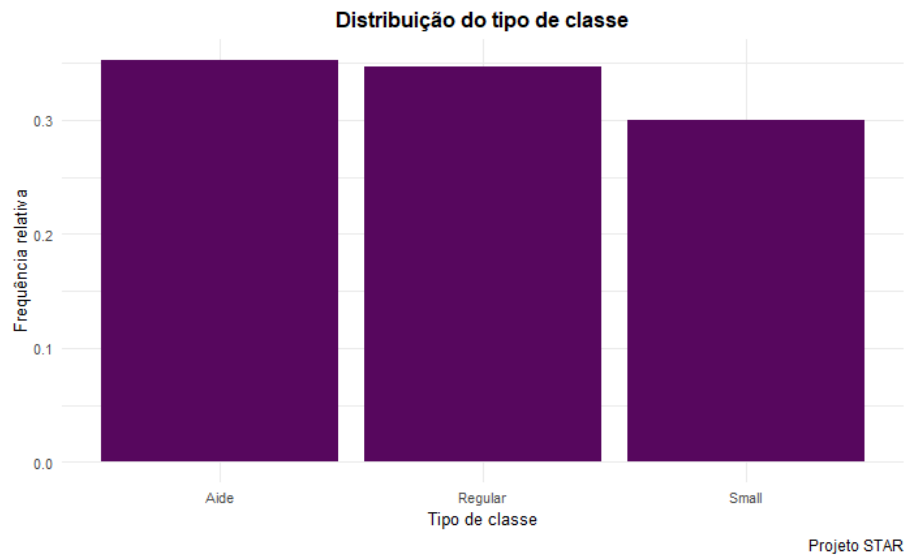
Tabela 14 - Regressão MPL para a variável small

	<i>Dependent variable:</i>
	as.numeric(small)
tchexper	-0.002* (0.001)
absent	-0.002*** (0.001)
boy1	-0.002 (0.012)
white_asian1	0.003 (0.018)
tchwhite1	0.067*** (0.018)
tchmasters1	-0.051*** (0.013)
freelunch1	-0.005 (0.014)
schurban1	0.005 (0.016)
Constant	1.302*** (0.025)
Observations	5,745
R ²	0.007
Adjusted R ²	0.006
Residual Std. Error	0.458 (df = 5736)
F Statistic	5.402*** (df = 8; 5736)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

Fonte: Elaboração Própria

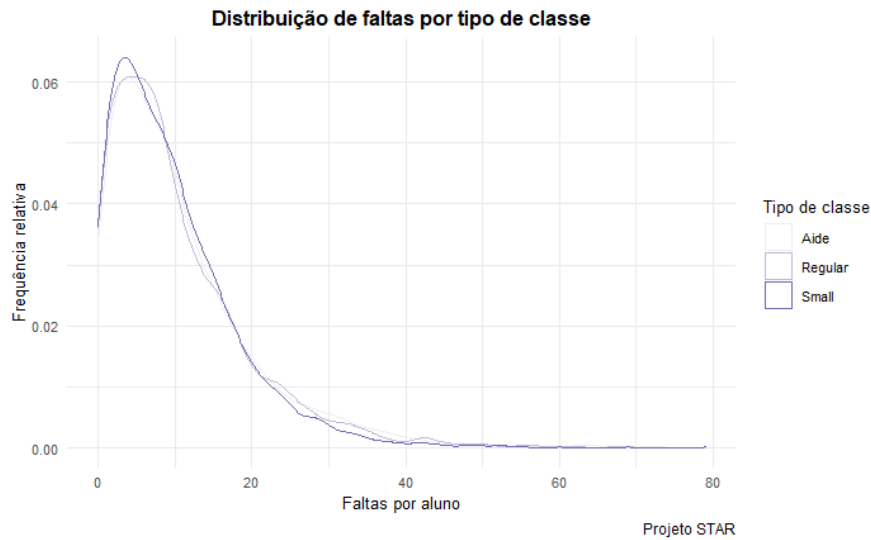
2022

Gráfico 1 - Distribuição do tipo de classe



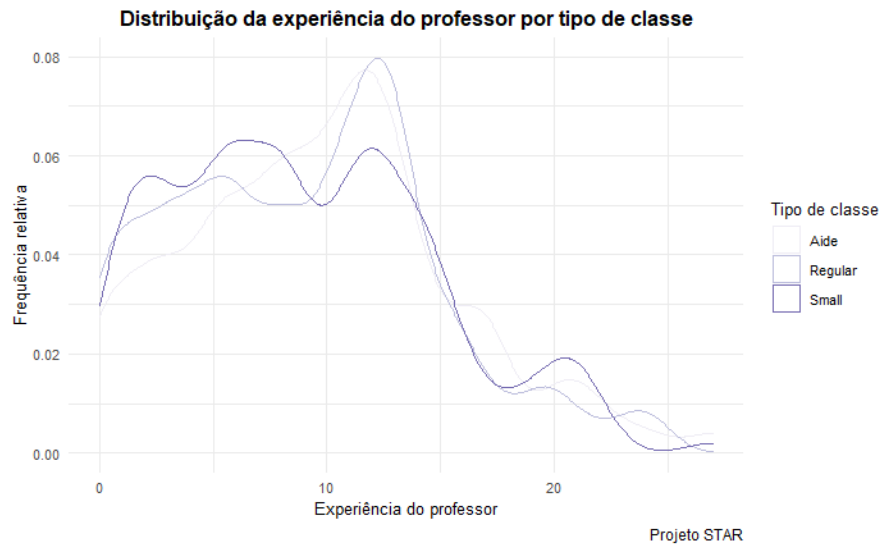
Fonte: Elaboração Própria
2022

Gráfico 2 - Distribuição de faltas por tipo de classe



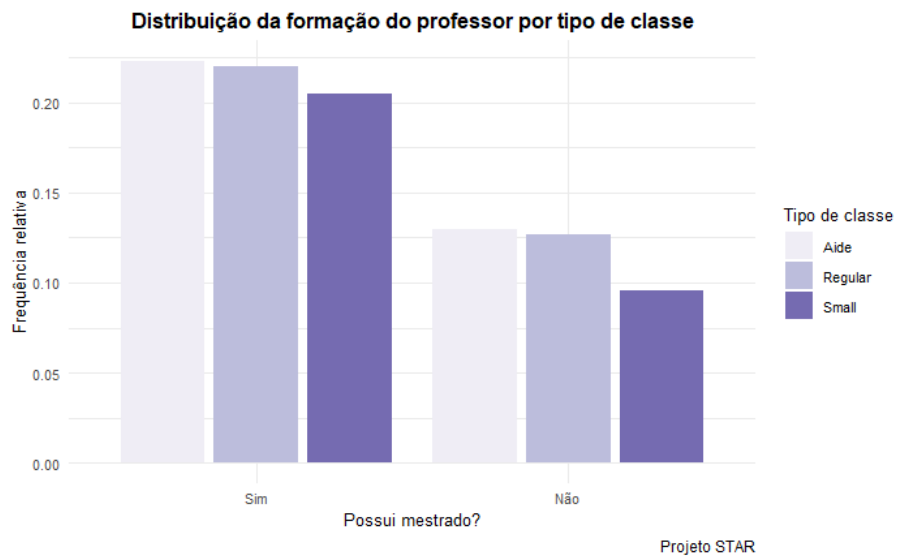
Fonte: Elaboração Própria
2022

Gráfico 3 - Distribuição da experiência do professor por tipo de classe



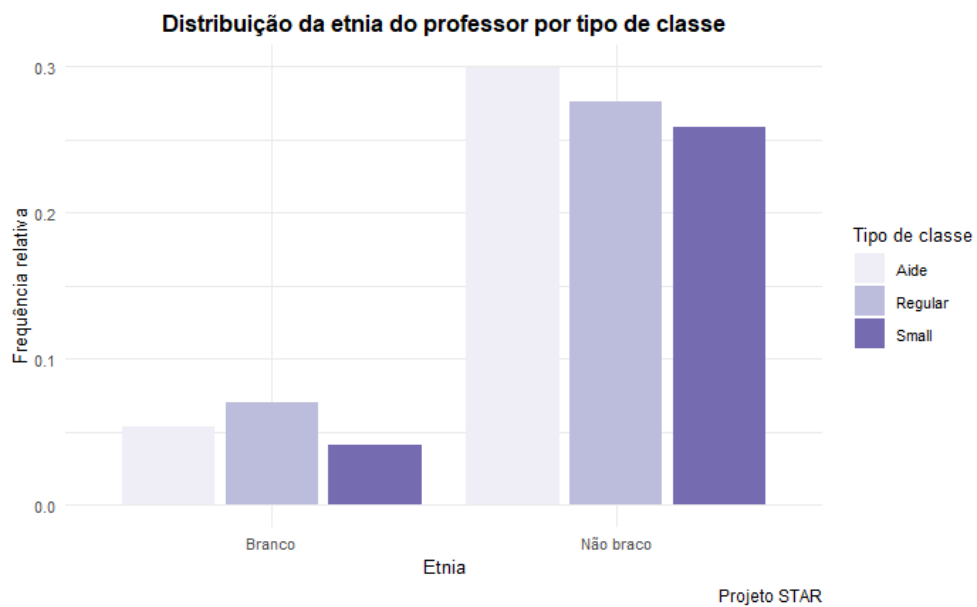
Fonte: Elaboração Própria
2022

Gráfico 4 - Distribuição da formação do professor por tipo de classe



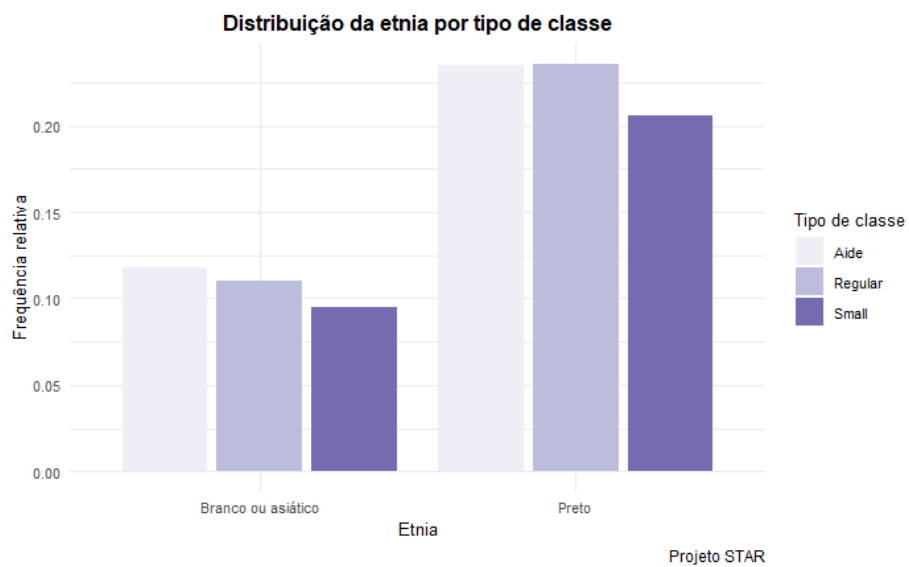
Fonte: Elaboração Própria
2022

Gráfico 5 - Distribuição da etnia do professor por tipo de classe



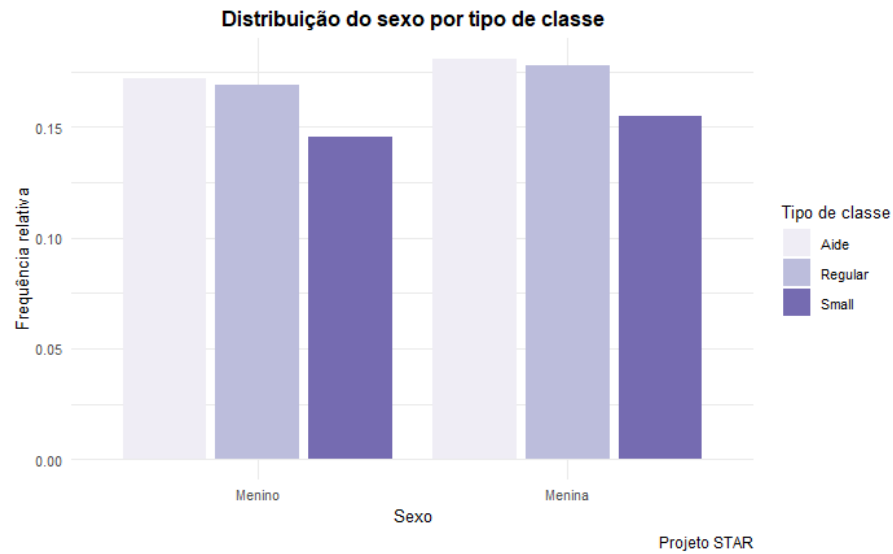
Fonte: Elaboração Própria
2022

Gráfico 6 - Distribuição da etnia por tipo de classe



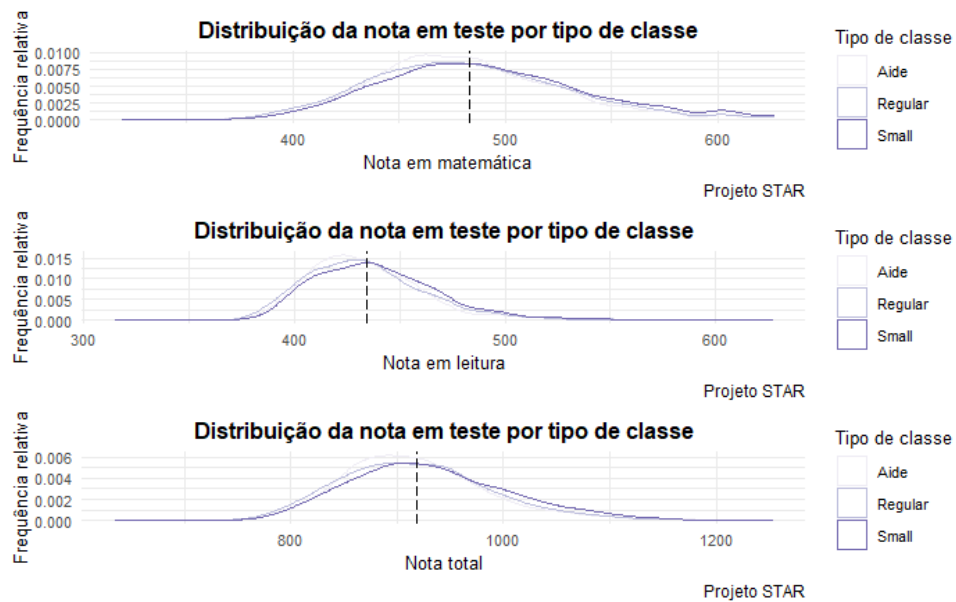
Fonte: Elaboração Própria
2022

Gráfico 7 - Distribuição do sexo por tipo de classe



Fonte: Elaboração Própria
2022

Gráfico 8 - Distribuição do desempenho por tipo de classe



Fonte: Elaboração Própria
2022