

O trabalho realizado foi baseado em um artigo publicado em 2006 chamado Indicador de Efeito Escola: uma metodologia para a identificação dos sucessos escolares a partir dos dados da Prova Brasil de autoria dos professores Amaury Patrick Gremaud, Fabiana de Felicio e Roberta Loboda Biondi.

A partir de dados da Prova Brasil (hoje incorporado pelo Saeb – Sistema de Avaliação da Educação Básica) apresentava uma metodologia de avaliação de eficácia escolar baseada no quanto a escola agregou à proficiência de seus alunos, e não simplesmente na nota final obtida por ela no exame. Para isso, ele tem por premissa a ideia que a proficiência está associado a dois grandes fatores: os fatores escolares (qualidade do professor, da direção da escola, ambiente escolar, etc) e os fatores extra-escolares (questões financeiras, estruturais da família e até da própria região) e são facilmente obtidas através de dados do censo e dos questionários socioeconômicos aplicados pelo próprio Saeb.

Utilizando a nota média das escolas e os dados socioeconômicos conseguimos estimar a parcela da nota que cabe aos fatores socioeconômicos e conseqüentemente, de forma residual, o valor agregado pelas escolas ao rendimento de seus alunos (IEE - Indicador de Efeito Escola). Para a criação do IEE, foram realizadas as seguintes etapas do trabalho, que serão melhor detalhadas posteriormente:

1. Definição do público
2. Pré-seleção de variáveis (fontes de dados)
3. Definição da variável resposta
4. Criação de Books das variáveis transformadas
5. Seleção de Variáveis por Book
6. Criação de Macro variáveis/Pilares
7. Criação do modelo
8. Criação do IEE (Indicador Efeito Escola) e Ranking

1. Definição do Público

Constrói uma base única de escolas dos 5EF, 9EF e 3EM participantes do Saeb, onde cada linha representa uma escola, série e sua nota média no Saeb. Ou seja, uma escola poderia aparecer 3x na base caso obtivesse turmas das 3 séries.

Foi decidido fazer um modelo e conseqüentemente, um ranking por escola, pois as faixas de pontuação são graduais (aumentam conforme a evolução no ensino) e porque a relação entre as variáveis e a nota varia conforme a série.

Metodologia

A seguir estão os passos para criação de base de público.

- 1) Base resultado das escolas no Saeb 2019

- 2) Filtra escolas estaduais de SP (estaduais) por série que tenham algum resultado divulgado (Língua Portuguesa ou Matemática).
- 3) Empilha base das escolas (pode ter duplicidade)
- 4) Cruza com Censo Escolar para obter nome da escola e flag de ensino integrado
- 5) Cruza com Censo para obter nome dos municípios
- 6) Cruza com base de PEI e ETI
- 7) Marca tipo de ensino e exclui as ETEC.

Resultado

Foi gerado um arquivo de público com 7.249 escolas das 3 séries.

- 5EF: 1431
- 9EF: 3247
- 3EM: 2571

Vale ressaltar que o Saeb possui critérios próprios para aplicação do exame e divulgação de resultado por escolas:

Critérios de Aplicação

São excluídas as seguintes populações da Aplicação SAEB:

Escolas com menos de 10 estudantes matriculados nas etapas avaliadas;

- *Turmas multisseriadas;*
- *Turmas de correção de fluxo;*
- *Turmas de Educação de Jovens e Adultos;*
- *Turmas de Ensino Médio Normal/Magistério;*
- *Classes, escolas ou serviços especializados de Educação Especial não integrantes do ensino regular;*
- *Escolas indígenas em que o ensino não é ministrado em Língua Portuguesa como primeira língua.*

Critérios para divulgação dos resultados

- Apenas dados de estudantes que responderam aos testes e cujos dados estavam consistentes com censo escolar foram considerados para apuração do resultado por escola. ID_SITUACAO_CENSO = 0 e IN_PREENCHIMENTO_XX = 1, para cada área de conhecimento.
- Foram considerados válidos apenas os cadernos de teste daqueles estudantes que responderam, no mínimo, três itens no caderno de prova. IN_PROFICIENCIA_LP = 1 OU IN_PROFICIENCIA_MT = 1
- Para ter resultado divulgado, as escolas participantes do Saeb devem ter no mínimo 80% de estudantes participantes (razão entre o total de alunos participantes no Saeb 2019 e o total de alunos declarados no Censo Escolar 2019). São considerados participantes os alunos presentes no momento da aplicação o teste (IN_PRESENCA_LP = 1 ou IN_PRESENCA_MT = 1) e que estavam declarados no Censo Escolar 2019 (ID_SITUACAO_CENSO = 1).

2. Pré-seleção de variáveis (fontes de dados)

Nessa etapa foi feita uma seleção de variáveis que refletiam os fatores extra-escolares dos alunos e da região em que estavam inseridos. As fontes utilizadas foram:

- Questionários Socioeconômico aplicados aos alunos - Saeb
- Questionários Socioeconômico aplicados aos alunos - Saesp
- Questionários Socioeconômico aplicados aos pais - Saesp
- IMP (Informações dos Municípios Paulistas) do Seade
- IPVS (Indicador Paulista de Vulnerabilidade Social)

Como o objetivo era manter apenas as questões que refletiam os fatores extraescolares, foram excluídas quaisquer variáveis que pudessem refletir direta ou indiretamente algum fator escolar, como por exemplo, a frequência com que o aluno faz dever de casa, pois apesar de ter uma parcela de influência socioeconômica (incentivo da família, tempo disponível para estudo, etc.) há também influência da gestão pedagógica e liderança dos docentes que deveriam passar e estimular os alunos a fazerem o dever de casa.

Resultados

Questionário Socioeconômico aplicados aos alunos - Saeb

Id_questao	questao	enunciado	selecao
TX_RESP_Q001	Questão 1	Qual língua você fala com mais frequência em sua casa?	Mantida
TX_RESP_Q002	Questão 2	Qual é a sua cor ou raça?	Mantida
TX_RESP_Q003a	Questão 3a	Normalmente, quem mora na sua casa? - Mãe (mães ou madrasta).	Mantida
TX_RESP_Q003b	Questão 3b	Normalmente, quem mora na sua casa? - Pai (pais ou padrasto).	Mantida
TX_RESP_Q003c	Questão 3c	Normalmente, quem mora na sua casa? - Irmão(s) ou irmã(s).	Mantida
TX_RESP_Q003d	Questão 3d	Normalmente, quem mora na sua casa? - Avô ou avó.	Mantida
TX_RESP_Q003e	Questão 3e	Normalmente, quem mora na sua casa? - Outros (tios, primos etc.).	Mantida
TX_RESP_Q004	Questão 4	Qual é a maior escolaridade da sua mãe (ou mulher responsável por você)?	Mantida
TX_RESP_Q005	Questão 5	Qual é a maior escolaridade de seu pai (ou homem responsável por você)?	Mantida
TX_RESP_Q006a	Questão 6a	Com que frequência seus pais ou responsáveis costumam: - Conversar com você sobre o que acontece na escola.	Mantida
TX_RESP_Q006b	Questão 6b	Com que frequência seus pais ou responsáveis costumam: - Incentivar você a estudar.	Mantida
TX_RESP_Q006c	Questão 6c	Com que frequência seus pais ou responsáveis costumam: - Incentivar você a fazer a tarefa de casa.	Excluída
TX_RESP_Q006d	Questão 6d	Com que frequência seus pais ou responsáveis costumam: - Incentivar você a comparecer às aulas.	Mantida
TX_RESP_Q006e	Questão 6e	Com que frequência seus pais ou responsáveis costumam: - Ir às reuniões de pais na escola.	Mantida
TX_RESP_Q007	Questão 7	Com que frequência sua família paga alguém para auxiliar nos trabalhos domésticos (faxina ou limpeza)?	Mantida
TX_RESP_Q008a	Questão 8a	Na região que você mora tem: - Rua pavimentada (asfalto ou calçamento).	Mantida
TX_RESP_Q008b	Questão 8b	Na região que você mora tem: - Água tratada da rua.	Mantida
TX_RESP_Q008c	Questão 8c	Na região que você mora tem: - Iluminação na rua.	Mantida
TX_RESP_Q009a	Questão 9a	Dos itens relacionados abaixo, quantos existem na sua casa? - Geladeira.	Mantida
TX_RESP_Q009b	Questão 9b	Dos itens relacionados abaixo, quantos existem na sua casa? - Tablet.	Mantida

TX_RESP_Q009c	Questão 9c	Dos itens relacionados abaixo, quantos existem na sua casa? - Computador (ou notebook).	Mantida
TX_RESP_Q009d	Questão 9d	Dos itens relacionados abaixo, quantos existem na sua casa? - Quartos para dormir.	Mantida
TX_RESP_Q009e	Questão 9e	Dos itens relacionados abaixo, quantos existem na sua casa? - Televisão.	Mantida
TX_RESP_Q009f	Questão 9f	Dos itens relacionados abaixo, quantos existem na sua casa? - Banheiro.	Mantida
TX_RESP_Q009g	Questão 9g	Dos itens relacionados abaixo, quantos existem na sua casa? - Carro.	Mantida
TX_RESP_Q010a	Questão 10a	Na sua casa tem: - Tv a cabo (Ex.:Net©, Sky© etc.).	Mantida
TX_RESP_Q010b	Questão 10b	Na sua casa tem: - Rede Wi-Fi.	Mantida
TX_RESP_Q010c	Questão 10c	Na sua casa tem: - Um quarto só seu.	Mantida
TX_RESP_Q010d	Questão 10d	Na sua casa tem: - Mesa para estudar (ou escrivaninha).	Mantida
TX_RESP_Q010e	Questão 10e	Na sua casa tem: - Garagem.	Mantida
TX_RESP_Q010f	Questão 10f	Na sua casa tem: - Forno de microondas.	Mantida
TX_RESP_Q010g	Questão 10g	Na sua casa tem: - Aspirador de pó.	Mantida
TX_RESP_Q010h	Questão 10h	Na sua casa tem: - Máquina de lavar roupa.	Mantida
TX_RESP_Q010i	Questão 10i	Na sua casa tem: - Freezer (independente ou segunda porta da geladeira).	Mantida
TX_RESP_Q011	Questão 11	Quanto tempo você demora para chegar à sua escola?	Mantida
TX_RESP_Q012	Questão 12	Considerando a maior distância percorrida, normalmente de que forma você chega à sua escola?	Mantida
TX_RESP_Q013	Questão 13	Com que idade você entrou na escola?	Excluída
TX_RESP_Q014	Questão 14	A partir do primeiro ano do ensino fundamental, em que tipo de escola você estudou?	Excluída
TX_RESP_Q015	Questão 15	Você já foi reprovado?	Excluída
TX_RESP_Q016	Questão 16	Alguma vez você abandonou a escola deixando de frequentá-la até o final do ano escolar?	Excluída
TX_RESP_Q017a	Questão 17a	Fora da escola em dias de aula, quanto tempo você usa para: - Lazer (TV, internet, jogar bola, música etc.).	Mantida
TX_RESP_Q017b	Questão 17b	Fora da escola em dias de aula, quanto tempo você usa para: - Fazer cursos.	Mantida
TX_RESP_Q017c	Questão 17c	Fora da escola em dias de aula, quanto tempo você usa para: - Fazer trabalhos domésticos (lavar louça, limpar quintal, cuidar dos irmãos).	Mantida
TX_RESP_Q017d	Questão 17d	Fora da escola em dias de aula, quanto tempo você usa para: - Estudar (lição de casa, trabalhos escolares, etc.).	Excluída

TX_RESP_Q017e	Questão 17e	Fora da escola em dias de aula, quanto tempo você usa para: - Trabalhar fora de casa (recebendo ou não um salário).	Mantida
TX_RESP_Q018a	Questão 18a	Com que frequência você costuma: - Ler notícias (jornais, revistas, internet etc.).	Mantida
TX_RESP_Q018b	Questão 18b	Com que frequência você costuma: - Ler livros que não sejam das matérias escolares.	Mantida
TX_RESP_Q018c	Questão 18c	Com que frequência você costuma: - Ler histórias em quadrinhos (mangás, gibis etc.).	Mantida

Questionário Socioeconômico aplicados aos alunos - Saresp

questao	enunciado	1_selecao
Q01	Qual seu grau de satisfação com sua escola ?	Excluída
Q02	Qualidade do Ensino	Excluída
Q03	Localização da Escola	Excluída
Q04	Segurança da Escola	Excluída
Q05	Salas de Aula	Excluída
Q06	Espaço para Esportes	Excluída
Q07	Refeitório	Excluída
Q08	Banheiros	Excluída
Q09	Quando você começou a estudar?	Excluída
Q10	Você já foi reprovado(a)?	Excluída
Q11	Seus pais ou seus responsáveis incentivam você a ler?	Excluída
Q12	Seus pais ou seus responsáveis incentivam você a estudar?	Excluída
Q13	Seus pais ou seus responsáveis ajudam você na lição de casa ou conferem se você fez a lição ?	Excluída
Q14	Seus pais ou seus responsáveis incentivam você a ir à escola e não faltar às aulas?	Excluída
Q15	Seus pais ou seus responsáveis perguntam a você sobre o que acontece na escola ?	Excluída
Q16	Seus pais ou seus responsáveis têm o costume de ler (jornal, revistas, livros etc.)?	Mantida
Q17	Você usa a Internet para : fazer pesquisas ou trabalhos?	Mantida
Q18	Você usa a Internet para : reforçar os estudos?	Mantida
Q19	Você usa a Internet para : entrar nas redes sociais ou conversar com seus amigos?	Mantida
Q20	Você usa a Internet para : jogar?	Mantida
Q21	Você usa a Internet para : assistir vídeos ou séries?	Mantida

Q22	Você gosta de estudar Matemática ?	Excluída
Q23	Você faz a lição de casa de Matemática?	Excluída
Q24	Você gosta de Língua Portuguesa?	Excluída
Q25	Você faz a lição de casa de Língua Portuguesa?	Excluída
Q26	Quantos livros você leu este ano?	Excluída
Q27	Qual sua forma preferida de estudar? Prestando atenção às aulas.	Excluída
Q28	Qual sua forma preferida de estudar? Lendo minhas anotações em casa .	Excluída
Q29	Qual sua forma preferida de estudar? Estudando com colegas/amigos.	Excluída
Q30	Qual sua forma preferida de estudar? Estudando pelos livros da escola .	Excluída
Q31	Qual sua forma preferida de estudar? Fazendo pesquisas na internet, utilizando programas no computador ou aplicativos no celular.	Excluída
Q32	Quanto tempo você passa por dia fazendo lição de casa ?	Excluída
Q33	Como você se prepara para as provas?	Excluída
Q34	Você já sofreu algum tipo de discriminação na escola ?	Excluída
Q35	Neste ano, fui chamado de nomes feios por outros alunos	Excluída
Q36	Neste ano, fui ameaçado por outros alunos	Excluída
Q37	Neste ano, Outros alunos me deixaram de fora das coisas de propósito.	Excluída
Q38	Neste ano, Outros alunos zombaram de mim.	Excluída
Q39	Neste ano, Outros alunos tiraram de mim ou destruíram coisas que me pertenciam.	Excluída
Q40	Neste ano, Eu apanhei ou fui empurrado por outros alunos.	Excluída
Q41	Neste ano, Outros alunos espalharam boatos desagradáveis a meu respeito.	Excluída
Q42	O que você acha que vai acontecer no futuro ?	Excluída
Q43	O que você acha das seguintes relações em sua escola ? Entre professores e alunos.	Excluída
Q44	O que você acha das seguintes relações em sua escola ? Entre os alunos.	Excluída
Q45	O que você acha das seguintes relações em sua escola ? Entre professores e direção .	Excluída

Q46	O que você acha das seguintes relações em sua escola ? Entre os professores.	Excluída
Q47	O que você acha das seguintes relações em sua escola ? Entre alunos e direção .	Excluída
Q48	O que você acha das seguintes relações em sua escola ? Entre alunos e funcionários.	Excluída
Q49	O que você acha das seguintes relações em sua escola ? Entre professores e funcionários.	Excluída

Questionário Socioeconômico aplicados aos pais - Saesp

questao	enunciado	selecao
Q01	Qual seu grau de satisfação com a escola de seu filho?	Excluída
Q02	Qualidade do Ensino	Excluída
Q03	Localização da Escola	Excluída
Q04	Segurança da Escola	Excluída
Q05	Salas de Aula	Excluída
Q06	Espaço para Esportes	Excluída
Q07	Refeitório	Excluída
Q08	Banheiros	Excluída
Q09	A escola passa lição de casa para o seu filho?	Excluída
Q10	Seu filho faz lição de casa?	Excluída
Q11	Você ou seu filho utilizam o app Minha Escola SP no celular?	Excluída
Q12	Converso com meu filho sobre a escola.	Mantida
Q13	Acompanho as lições de casa do meu filho.	Excluída
Q14	Converso com os professores do meu filho.	Excluída
Q15	Participo das reuniões de pais.	Excluída
Q16	Participo das reuniões do Conselho de Escola.	Excluída
Q17	Participo de passeios, festas, campeonatos esportivos ou apresentações culturais promovidas pela escola.	Excluída
Q18	Participo de outras formas.	Mantida
Q19	Até que série/nível de ensino o pai (ou responsável) estudou?	Mantida
Q20	Até que série/nível de ensino a mãe (ou responsável) estudou?	Mantida
Q21	Qual é a situação de trabalho do pai (ou responsável)?	Mantida

Q22	Qual é a situação de trabalho da mãe (ou responsável)?	Mantida
Q23	Qual é a renda familiar de seu domicílio, ou seja, a soma dos salários e rendimentos (valor bruto) de todas as pessoas que moram em sua casa?	Mantida
Q24	Na sua casa você tem: Jornal de notícias.	Mantida
Q25	Na sua casa você tem: Revista de informação geral (Veja, Época, Isto É etc.).	Mantida
Q26	Na sua casa você tem: Dicionário.	Mantida
Q27	Na sua casa você tem: Livros (romances, poesias, contos, etc.).	Mantida
Q28	Na sua casa você tem: Gibis e histórias em quadrinhos	Mantida
Q29	Na sua casa você tem: Revistas educativas ou de divulgação científica (Ciência Hoje, Galileu, etc.).	Mantida
Q30	Na sua casa você tem: Acesso à internet.	Mantida
Q31	Na sua casa você tem: Serviço de TV por assinatura (cabo, satélite, streaming etc.).	Mantida
Q32	Indique quais e quantos bens e serviços domésticos listados abaixo você possui em casa. Banheiro	Mantida
Q33	Indique quais e quantos bens e serviços domésticos listados abaixo você possui em casa. Quarto para dormir	Mantida
Q34	Indique quais e quantos bens e serviços domésticos listados abaixo você possui em casa. Televisão	Mantida
Q35	Indique quais e quantos bens e serviços domésticos listados abaixo você possui em casa. Geladeira	Mantida
Q36	Indique quais e quantos bens e serviços domésticos listados abaixo você possui em casa. Freezer	Mantida
Q37	Indique quais e quantos bens e serviços domésticos listados abaixo você possui em casa. Microondas	Mantida
Q38	Indique quais e quantos bens e serviços domésticos listados abaixo você possui em casa. Telefone fixo	Mantida
Q39	Indique quais e quantos bens e serviços domésticos listados abaixo você possui em casa. Telefone celular	Mantida
Q40	Indique quais e quantos bens e serviços domésticos listados abaixo você possui em casa. Computador	Mantida
Q41	Indique quais e quantos bens e serviços domésticos listados abaixo você possui em casa. Tablet	Mantida
Q42	Indique quais e quantos bens e serviços domésticos listados abaixo você possui em casa. Máquina de lavar roupa	Mantida

Q43	Indique quais e quantos bens e serviços domésticos listados abaixo você possui em casa. Máquina de secar roupa	Mantida
Q44	Indique quais e quantos bens e serviços domésticos listados abaixo você possui em casa. Máquina de lavar louça	Mantida
Q45	Indique quais e quantos bens e serviços domésticos listados abaixo você possui em casa. Aparelho de DVD	Mantida
Q46	Indique quais e quantos bens e serviços domésticos listados abaixo você possui em casa. Aspirador de pó	Mantida
Q47	Indique quais e quantos bens e serviços domésticos listados abaixo você possui em casa. Automóvel	Mantida
Q48	Indique quais e quantos bens e serviços domésticos listados abaixo você possui em casa. Motocicleta	Mantida
Q49	Indique quais e quantos bens e serviços domésticos listados abaixo você possui em casa. Empregada diarista	Mantida
Q50	Indique quais e quantos bens e serviços domésticos listados abaixo você possui em casa. Empregada mensalista	Mantida

Seade

No caso de IMP Seade, já foram selecionadas diretamente as variáveis e interesse:

Índice de Desenvolvimento Humano Municipal - IDHM Longevidade 2010
Índice de Desenvolvimento Humano Municipal - IDHM Renda 2010
PIB per Capita (Em reais correntes) 2018
Rendimento Médio do Total de Empregos Formais (Em reais correntes) 2019
Empregos Formais das Pessoas com Ensino Fundamental Incompleto (%) 2019
Empregos Formais das Pessoas com Ensino Médio Completo (%) 2019
Empregos Formais das Pessoas com Ensino Superior Completo (%) 2019
Empregos formais de pessoas até 24 anos (%) 2019
Empregos Formais de Mulheres (%) 2019
Nascidos Vivos de Mães com Menos de 18 Anos (Em %) 2019
Taxa de Mortalidade da População de 15 a 34 Anos (Por cem mil habitantes nessa faixa etária) 2019
Taxa de Mortalidade por Agressões (Por cem mil habitantes) 2017
Taxa de Mortalidade por Causas Externas (Por cem mil habitantes) 2017
Índice de Envelhecimento (Em %) 2019
Taxa de Mortalidade Infantil (Por mil nascidos vivos) 2019
Abastecimento de Água - Nível de Atendimento - Censo Demográfico (Em %) 2010
Coleta de Lixo - Nível de Atendimento - Censo Demográfico (Em %) 2010
Esgoto Sanitário - Nível de Atendimento - Censo Demográfico (Em %) 2010

Variável	Descrição	seleção
v10	Grupo do IPVS	Mantida
v11	Domicílios particulares e coletivos	Mantida
v12	Domicílios particulares permanentes	Mantida
v13	Total de domicílios particulares improvisados	Mantida
v14	Moradores em domicílios particulares permanentes	Mantida
v15	Média de moradores em domicílios particulares permanentes	Mantida
v16	Proporção de crianças de 0 a 5 anos na população	Mantida
v17	Total de pessoas responsáveis	Mantida
v18	Rendimento médio domiciliar dos domicílios particulares permanentes	Mantida
v19	Renda per capita nos domicílios particulares permanentes do setor censitário	Mantida
v20	Proporção de Domicílios sem Renda per Capita	Mantida
v21	Proporção de Domicílios com Renda per Capita até 1/8 SM - Proxy de até 70	Mantida
v22	Proporção de Domicílios com Renda per Capita de 1/8 a 1/2 SM - Proxy de 70 até 261	Mantida
v23	Proporção de Domicílios com Renda per Capita de 1/2 a 2 SM - Proxy de 261 a 914	Mantida
v24	Proporção de Domicílios com Renda per Capita de mais de 2 SM - Proxy de mais de 914	Mantida
v25	Proporção de domicílios particulares com rendimento nominal mensal de até 1/2 S.M.	Mantida
v26	Proporção de domicílios particulares com rendimento nominal mensal de até 1/4 S.M.	Mantida
v27	Idade média das pessoas responsáveis	Mantida
v28	Rendimento médio do responsável pelo domicílio	Mantida
v29	Proporção de pessoas responsáveis alfabetizadas	Mantida
v30	Proporção de pessoas responsáveis de 10 a 29 anos	Mantida
v31	Participação da renda do responsável pelo domicílio na renda domiciliar (em porcentagem)	Mantida
v32	Idade média das mulheres responsáveis pelo domicílio	Mantida
v33	Proporção de mulheres responsáveis pelo domicílio alfabetizadas	Mantida
v34	Proporção de mulheres responsáveis pelo domicílio com menos de 30 anos	Mantida
v35	Rendimento médio das mulheres responsáveis pelo domicílio	Mantida
v36	Fator 1	Excluída
v37	Fator 2	Excluída
v38	Fator 1 rural	Excluída
v39	Fator 2 rural	Excluída

3. Definição da variável resposta

O objetivo dessa etapa era escolher a nota da proficiência do Saeb que seria utilizada para construir o modelo e consequentemente, o IEE.

Foram feitos 6 testes rápidos de modelo utilizando como variáveis explicativas as questões do questionário Alunos e como resposta as seguintes alternativas com os respectivos resultados de performance (R2) na tabela abaixo:

1. Língua Portuguesa (LP)
2. Matemática (MT)
3. Média de LP e MT como proxy de desempenho do aluno resumida
4. Língua Portuguesa (LP) padronizada
5. Matemática (MT) padronizada
6. Média de LP padronizada e MT padronizada

	5EF	9EF	3EM
LP	41%	42%	58%
MT	35%	40%	57%
Média	40%	43%	59%
LP std	41%	42%	58%
MT std	35%	40%	57%
Media std	40%	43%	59%

A decisão pelo uso da variável Média justificou-se tanto pela apresentação de bons resultados em performance quanto pela facilitação do uso pela equipe de negócio já que seria necessário apenas um ranking por ano de ensino ao invés de 2 (um para LP e outro para MT).

4. Criação de Books das variáveis transformadas

Transformação das variáveis dos questionários dos alunos a nível escola por meio de sumarização pelo cálculo de proporção de alunos que responderam determinada alternativa de uma questão. As variáveis do Seade e do IPBS

I. Saeb

Criação de 3 Books por ID_ESCOLA e para cada série:

- Saeb5ef_prop
- Saeb9ef_prop
- Saeb3em_prop

II. Saesp Pais

Criação de 3 Books por ID_ESCOLA e para cada série:

- Saesp5ef_prop
- Saesp9ef_prop
- Saesp3em_prop

III. Saesp Alunos

Como a maioria das variáveis foi excluída na fase de pré-seleção (Seleção qualitativa) e as variáveis restantes não se mostraram relevantes na análise exploratória, não foram criados books para esse questionário.

Metodologia

Primeiramente, é feita uma transformação das variáveis das questões em Dummies/Binárias. Exemplo: A Q1 vira Q1_A, Q1_B, Q1_C, Q1_D, Q1_E, que possui preenchimento = 1 ou =0 caso o aluno tenha assinalado aquela alternativa ou não, respectivamente.

Input:

ID_ESCOLA	Aluno	Questão 1
3500001	1	A
3500001	2	A
3500001	3	B
3500001	4	C

Output:

ID_ESCOLA	Aluno	Questão 1	Q1 Alt. A	Q1 Alt. B	Q1 Alt. C
3500001	1	A	1	0	0
3500001	2	A	1	0	0
3500001	3	B	0	1	0
3500001	4	C	0	0	1

Após isso, resume-se os dados em proporções por escola. Por exemplo: A Q1_A agora indicará a quantidade de alunos daquela escola que assinalaram a alternativa A na Q1.

Input:

ID_ESCOLA	Aluno	Questão 1	Q1 Alt. A	Q1 Alt. B	Q1 Alt. C
3500001	1	A	1	0	0
3500001	2	A	1	0	0
3500001	3	B	0	1	0
3500001	4	C	0	0	1

Output:

ID_ESCOLA	Q1_A	Q1_B	Q1_C
3500001	2/4 = 0,5	¼ = 0,25	¼ = 0,25

IV. IPVS

No caso da base do IPVS, o nível de observação era o setor censitário (subdivisão dentro do distrito). Como o foco de interesse é o distrito (subdivisão dentro do município), transformamos as variáveis de duas maneiras:

- Numéricas: resumimos a média por distrito. Exemplo: idade média das pessoas responsáveis
- Categórica: transformamos em dummy/binária e resumimos os dados à proporção daquele grupo por distrito (assim como foi feito no Saeb). Ex: Grupo de Vulnerabilidade Social

V. Seade

Nesse caso, as informações na fonte de origem estão por município, então não há o que se falar em transformação; o cruzamento entre bases de dados é feita de forma direta.

5. Seleção de Variáveis por Book

Exclusão de variáveis menos relevantes a explicar a variabilidade de notas entre as escolas.

Metodologia

- 1) Remove colunas com variância zero ou próxima de zero
- 2) Cruza com base de público para adicionar a média (variável resposta)
- 3) Roda uma regressão univariada e mensura o r-quadrado
- 4) Seleciona as melhores variáveis:
 - Saeb e Saresp: Seleciona as top 20 variáveis que obedeçam ao corte mínimo de 10% de r2
 - IPVS: Seleciona as top 10 variáveis que obedeçam ao corte mínimo de 5% de r2
 - Seade: Seleciona as variáveis que obedeçam ao corte mínimo de 10% de r2
- 5) Gera base final cruzando df_publico com variáveis selecionadas

6. Criação Dos Pilares

Criação de macro variáveis categóricas através da interação entre variáveis relacionadas (por tema).

Metodologia

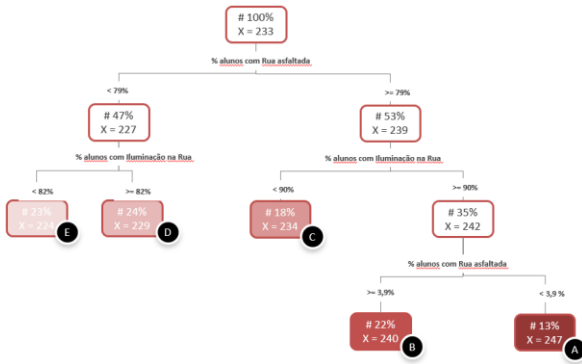
A partir de uma base consolidada por série com as questões do Saeb e Saresp:

1. Seleciona as questões de acordo com cada tema (apresentado abaixo)
2. Treina uma Árvore de Decisão
3. De acordo com a média esperada em cada folha (grupo final em cada árvore), classifica-se as observações naquele grupo em uma categoria, onde A indica o grupo de escolas com maior nota esperada, B o segundo maior grupo e assim por diante
4. Isso foi feito para cada ano de ensino, a seguir estão apenas os do 5ª ano do Ensino Fundamental a título de exemplo.

Pilar de Região

Condições básicas da região em que o aluno mora (questões tx_resp_q008)

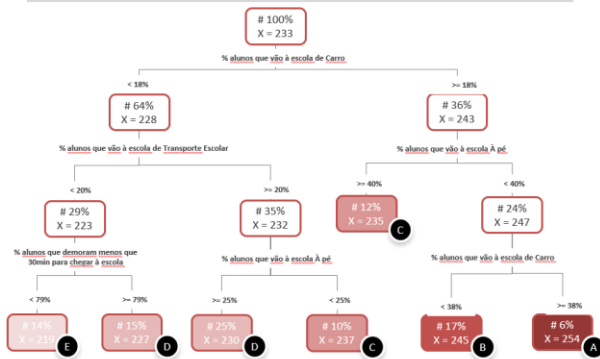
Pilar Região – 5EF



Pilar de Localização

Acesso à escola: tempo de deslocamento e meio de transporte utilizado (questões tx_resp_q011 e tx_resp_q012)

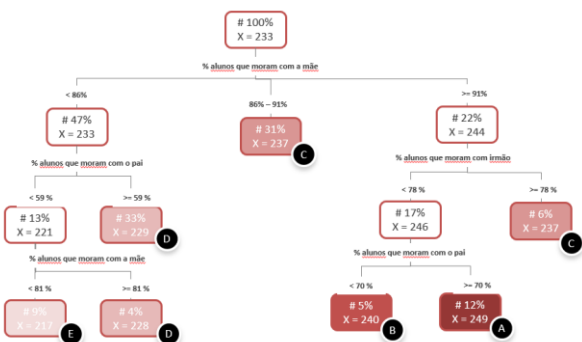
Pilar Localização – 5EF



Pilar de Família

Estrutura da família: com quem o aluno mora (questões tx_resp_q003)

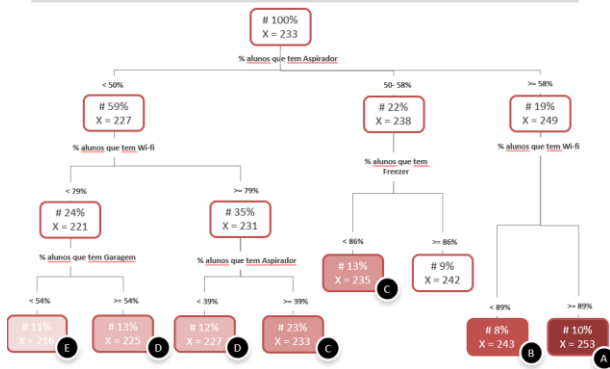
Pilar família – 5EF



Pilar de Bens

Posse de bens materiais e imateriais (questões tx_resp_q007, tx_resp_q009 e tx_resp_q010)

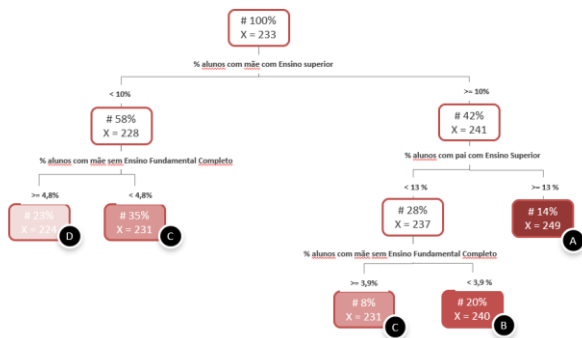
Pilar Bens – 5EF



Pilar de Letramento

Escolaridade do pai e da mãe (questões tx_resp_q004 e tx_resp_q005)

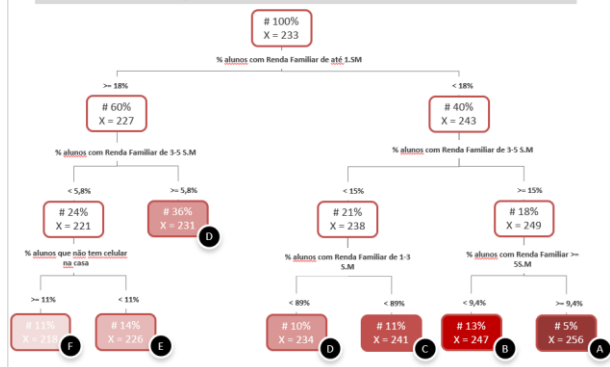
Pilar Letramento – 5EF



Pilar do Saresp

Variáveis de Renda, Situação de trabalho dos pais e alguns bens

Pilar Saresp – 5EF



7. Criação do Modelo

Ao todo, foram treinados 3 modelos para cada ano de ensino: v0, v1 e v2 conforme melhorias foram sendo incrementadas. Foi utilizado cross-validation com 5 Folds e os algoritmos testados foram: Linear Regression, Random Forest e Gradient Boosting. A métrica de avaliação utilizada para mensurar a performance do modelo foi R-quadrado e nas 3 séries o algoritmo com melhor performance foi a Regressão Linear.

5º EF

Equação final

media ~ nivel_socio_economico + pilar_ipvs+ pilar_familia + pilar_letramento + pilar_localizacao + pilar_bens_superfluos + pilar_bens_eletrodomesticos + l(tx_resp_q017e_A^2) + sqrt(tx_resp_q002_A)+ q23_D_spp * q23_E_spp

R2

1. Regressão: 0.6029943
2. Random Forest: 0.5987986
- 3.Boosting: 0.5393571

9º EF

Equação final

media ~ nivel_socio_economico + pilar_ipvs+ pilar_familia + pilar_letramento + pilar_localizacao + pilar_bens_superfluos + pilar_bens_eletrodomesticos + l(tx_resp_q017e_A^2) + sqrt(tx_resp_q002_A)+ q23_D_spp * q23_E_spp

R2

1. Regressão: 0.5861684
2. Random Forest: 0.5608338
- 3.Boosting: 0.5394691

3º EM

Equação final

media ~ nivel_socio_economico + pilar_familia + tx_resp_q004_E+tx_resp_q005_E + tx_resp_q004_F + tx_resp_q005_F + pilar_regiao + pilar_localizacao + pilar_bens_superfluos + pilar_bens_comodos + l(tx_resp_q017e_A^2) + tx_resp_q002_A + q23_D_spp

R2

1. Regressão: 0.6017677
2. Random Forest: 0.5835313
- 3.Boosting: 0.5662748

8. Criação do IEE (Indicador Efeito Escola e Ranking)

O IEE é criado a partir do resíduo do modelo, ou seja, a diferença entra a nota real tirada pela escolada no Saeb e a nota estimada pelo modelo considerando apenas fatores socioeconômicos dos alunos. Ou seja, é uma estimativa da parcela da nota que cabe aos fatores escolares, uma medida de eficácia escolar.

Exemplo: Escola com Nota de 260 pontos no Saeb cuja nota estimada pelo modelo é 250 pontos tem o IEE de 10 pontos. Podemos interpretar da seguinte forma: aquela escola tirou 10 pontos a mais do que o esperado para escolas com perfil socioeconômico de alunos semelhante.

Anexos

Fontes de Dados

Saeb

<https://www.gov.br/inep/pt-br/aceso-a-informacao/dados-abertos/microdados/saeb>

Saresp

<https://dados.educacao.sp.gov.br/dataset/microdados-de-alunos-do-sistema-de-avalia%C3%A7%C3%A3o-de-rendimento-escolar-do-estado-de-s%C3%A3o-paulo>

Censo Escolar

<https://www.gov.br/inep/pt-br/aceso-a-informacao/dados-abertos/microdados/censo-escolar>

ESCOLAS PEI/ETI

<https://dados.educacao.sp.gov.br/dataset/programa-de-ensino-integral-pei>

Seade

<http://www.imp.seade.gov.br/frontend/#/tabelas>

IPVS

<http://ipvs.seade.gov.br/view/index.php>

R-quadrado das variáveis

coluna	label	r2_5ef	r2_9ef	r2_3em	r2_pilar_3em
tx_resp_q001_A	Qual língua você fala com mais frequência em sua casa? Português.	6%	4%	4%	
tx_resp_q001_B	Qual língua você fala com mais frequência em sua casa? Espanhol.	2%	0%	0%	
tx_resp_q001_C	Qual língua você fala com mais frequência em sua casa? Outra língua.	1%	1%	0%	
tx_resp_q002_A	Qual é a sua cor ou raça? Branca.	19%	24%	35%	
tx_resp_q002_B	Qual é a sua cor ou raça? Preta.	13%	8%	9%	
tx_resp_q002_C	Qual é a sua cor ou raça? Parda.	1%	4%	16%	
tx_resp_q002_D	Qual é a sua cor ou raça? Amarela.	1%	0%	0%	
tx_resp_q002_E	Qual é a sua cor ou raça? Indígena.	0%	0%	1%	
tx_resp_q002_F	Qual é a sua cor ou raça? Não quero declarar.	1%	5%	9%	
tx_resp_q003a_A	Normalmente, quem mora na sua casa? - Mãe (mães ou madrasta). Não.	5%	7%	9%	33%
tx_resp_q003a_B	Normalmente, quem mora na sua casa? - Mãe (mães ou madrasta). Sim.	22%	21%	21%	33%
tx_resp_q003b_A	Normalmente, quem mora na sua casa? - Pai (pais ou padrasto). Não.	2%	2%	3%	33%
tx_resp_q003b_B	Normalmente, quem mora na sua casa? - Pai (pais ou padrasto). Sim.	18%	16%	17%	33%
tx_resp_q003c_A	Normalmente, quem mora na sua casa? - Irmão(s) ou irmã(s). Não.	10%	11%	9%	33%
tx_resp_q003c_B	Normalmente, quem mora na sua casa? - Irmão(s) ou irmã(s). Sim.	0%	0%	0%	33%
tx_resp_q003d_A	Normalmente, quem mora na sua casa? - Avô ou avó. Não.	6%	4%	5%	33%
tx_resp_q003d_B	Normalmente, quem mora na sua casa? - Avô ou avó. Sim.	0%	0%	0%	33%
tx_resp_q003e_A	Normalmente, quem mora na sua casa? - Outros (tios, primos etc.). Não.	13%	10%	13%	33%
tx_resp_q003e_B	Normalmente, quem mora na sua casa? - Outros (tios, primos etc.). Sim.	3%	4%	8%	33%
tx_resp_q004_A	Qual é a maior escolaridade da sua mãe (ou mulher responsável por você)? Não completou o 5º ano do Ensino Funda	10%	7%	8%	46%
tx_resp_q004_B	Qual é a maior escolaridade da sua mãe (ou mulher responsável por você)? Ensino Fundamental, até o 5º ano.	5%	5%	8%	46%
tx_resp_q004_C	Qual é a maior escolaridade da sua mãe (ou mulher responsável por você)? Ensino Fundamental completo.	3%	1%	2%	46%
tx_resp_q004_D	Qual é a maior escolaridade da sua mãe (ou mulher responsável por você)? Ensino Médio completo.	3%	19%	14%	46%
tx_resp_q004_E	Qual é a maior escolaridade da sua mãe (ou mulher responsável por você)? Ensino Superior completo (faculdade ou g	22%	23%	40%	46%
tx_resp_q004_F	Qual é a maior escolaridade da sua mãe (ou mulher responsável por você)? Não sei.	0%	18%	31%	46%
tx_resp_q005_A	Qual é a maior escolaridade de seu pai (ou homem responsável por você)? Não completou o 5º ano do Ensino Fundar	6%	5%	7%	46%
tx_resp_q005_B	Qual é a maior escolaridade de seu pai (ou homem responsável por você)? Ensino Fundamental, até o 5º ano.	4%	2%	3%	46%
tx_resp_q005_C	Qual é a maior escolaridade de seu pai (ou homem responsável por você)? Ensino Fundamental completo.	3%	0%	0%	46%
tx_resp_q005_D	Qual é a maior escolaridade de seu pai (ou homem responsável por você)? Ensino Médio completo.	3%	23%	24%	46%
tx_resp_q005_E	Qual é a maior escolaridade de seu pai (ou homem responsável por você)? Ensino Superior completo (faculdade ou gr	19%	21%	38%	46%
tx_resp_q005_F	Qual é a maior escolaridade de seu pai (ou homem responsável por você)? Não sei.	0%	20%	40%	46%
tx_resp_q007_A	Com que frequência sua família paga alguém para auxiliar nos trabalhos domésticos (faxina ou limpeza)? Nunca ou q	5%	3%	0%	
tx_resp_q007_B	Com que frequência sua família paga alguém para auxiliar nos trabalhos domésticos (faxina ou limpeza)? De vez em q	3%	4%	11%	
tx_resp_q007_C	Com que frequência sua família paga alguém para auxiliar nos trabalhos domésticos (faxina ou limpeza)? Sempre ou c	8%	6%	2%	
tx_resp_q008a_A	Na região que você mora tem: - Rua pavimentada (asfalto ou calçamento). Não.	5%	3%	2%	26%
tx_resp_q008a_B	Na região que você mora tem: - Rua pavimentada (asfalto ou calçamento). Sim.	13%	6%	4%	26%
tx_resp_q008b_A	Na região que você mora tem: - Água tratada da rua. Não.	2%	4%	4%	26%
tx_resp_q008b_B	Na região que você mora tem: - Água tratada da rua. Sim.	8%	11%	8%	26%
tx_resp_q008c_A	Na região que você mora tem: - Iluminação na rua. Não.	1%	2%	2%	26%
tx_resp_q008c_B	Na região que você mora tem: - Iluminação na rua. Sim.	10%	7%	6%	26%
tx_resp_q009a_A	Dos itens relacionados abaixo, quantos existem na sua casa? - Geladeira. Nenhum.	0%	0%	0%	48%
tx_resp_q009a_B	Dos itens relacionados abaixo, quantos existem na sua casa? - Geladeira. 1.	5%	0%	0%	48%
tx_resp_q009a_C	Dos itens relacionados abaixo, quantos existem na sua casa? - Geladeira. 2.	3%	2%	2%	48%
tx_resp_q009a_D	Dos itens relacionados abaixo, quantos existem na sua casa? - Geladeira. 3 ou mais.	1%	1%	1%	48%
tx_resp_q009b_A	Dos itens relacionados abaixo, quantos existem na sua casa? - Tablet. Nenhum.	2%	1%	0%	48%
tx_resp_q009b_B	Dos itens relacionados abaixo, quantos existem na sua casa? - Tablet. 1.	10%	7%	3%	48%
tx_resp_q009b_C	Dos itens relacionados abaixo, quantos existem na sua casa? - Tablet. 2.	5%	1%	0%	48%
tx_resp_q009b_D	Dos itens relacionados abaixo, quantos existem na sua casa? - Tablet. 3 ou mais.	1%	0%	1%	48%
tx_resp_q009c_A	Dos itens relacionados abaixo, quantos existem na sua casa? - Computador (ou notebook). Nenhum.	18%	19%	29%	48%
tx_resp_q009c_B	Dos itens relacionados abaixo, quantos existem na sua casa? - Computador (ou notebook). 1.	18%	13%	9%	48%
tx_resp_q009c_C	Dos itens relacionados abaixo, quantos existem na sua casa? - Computador (ou notebook). 2.	24%	19%	31%	48%
tx_resp_q009c_D	Dos itens relacionados abaixo, quantos existem na sua casa? - Computador (ou notebook). 3 ou mais.	10%	10%	20%	48%
tx_resp_q009d_A	Dos itens relacionados abaixo, quantos existem na sua casa? - Quartos para dormir. Nenhum.	9%	0%	0%	48%
tx_resp_q009d_B	Dos itens relacionados abaixo, quantos existem na sua casa? - Quartos para dormir. 1.	14%	8%	12%	48%

coluna	label	r2_5ef	r2_9ef	r2_3em
tx_resp_q009d_C	Dos itens relacionados abaixo, quantos existem na sua casa? - Quartos para dormir. 2.	2%	0%	0%
tx_resp_q009d_D	Dos itens relacionados abaixo, quantos existem na sua casa? - Quartos para dormir. 3 ou mais.	15%	7%	9%
tx_resp_q009e_A	Dos itens relacionados abaixo, quantos existem na sua casa? - Televisão. Nenhum.	1%	0%	0%
tx_resp_q009e_B	Dos itens relacionados abaixo, quantos existem na sua casa? - Televisão. 1.	8%	3%	3%
tx_resp_q009e_C	Dos itens relacionados abaixo, quantos existem na sua casa? - Televisão. 2.	4%	0%	0%
tx_resp_q009e_D	Dos itens relacionados abaixo, quantos existem na sua casa? - Televisão. 3 ou mais.	18%	10%	7%
tx_resp_q009f_A	Dos itens relacionados abaixo, quantos existem na sua casa? - Banheiro. Nenhum.	4%	2%	3%
tx_resp_q009f_B	Dos itens relacionados abaixo, quantos existem na sua casa? - Banheiro. 1.	9%	11%	21%
tx_resp_q009f_C	Dos itens relacionados abaixo, quantos existem na sua casa? - Banheiro. 2.	19%	11%	16%
tx_resp_q009f_D	Dos itens relacionados abaixo, quantos existem na sua casa? - Banheiro. 3 ou mais.	21%	15%	23%
tx_resp_q009g_A	Dos itens relacionados abaixo, quantos existem na sua casa? - Carro. Nenhum.	25%	20%	20%
tx_resp_q009g_B	Dos itens relacionados abaixo, quantos existem na sua casa? - Carro. 1.	21%	13%	12%
tx_resp_q009g_C	Dos itens relacionados abaixo, quantos existem na sua casa? - Carro. 2.	22%	15%	16%
tx_resp_q009g_D	Dos itens relacionados abaixo, quantos existem na sua casa? - Carro. 3 ou mais.	6%	2%	0%
tx_resp_q010a_A	Na sua casa tem: - Tv a cabo (Ex.:Net@, Sky@ etc.). Não.	9%	2%	0%
tx_resp_q010a_B	Na sua casa tem: - Tv a cabo (Ex.:Net@, Sky@ etc.). Sim.	22%	4%	0%
tx_resp_q010b_A	Na sua casa tem: - Rede Wi-Fi. Não.	9%	9%	11%
tx_resp_q010b_B	Na sua casa tem: - Rede Wi-Fi. Sim.	20%	13%	14%
tx_resp_q010c_A	Na sua casa tem: - Um quarto só seu. Não.	2%	8%	10%
tx_resp_q010c_B	Na sua casa tem: - Um quarto só seu. Sim.	13%	12%	13%
tx_resp_q010d_A	Na sua casa tem: - Mesa para estudar (ou escrivaninha). Não.	6%	23%	36%
tx_resp_q010d_B	Na sua casa tem: - Mesa para estudar (ou escrivaninha). Sim.	16%	32%	40%
tx_resp_q010e_A	Na sua casa tem: - Garagem. Não.	19%	12%	15%
tx_resp_q010e_B	Na sua casa tem: - Garagem. Sim.	27%	16%	17%
tx_resp_q010f_A	Na sua casa tem: - Forno de microondas. Não.	11%	10%	9%
tx_resp_q010f_B	Na sua casa tem: - Forno de microondas. Sim.	25%	17%	13%
tx_resp_q010g_A	Na sua casa tem: - Aspirador. Não.	21%	20%	21%
tx_resp_q010g_B	Na sua casa tem: - Aspirador. Sim.	38%	27%	26%
tx_resp_q010h_A	Na sua casa tem: - Máquina de lavar roupa. Não.	11%	9%	7%
tx_resp_q010h_B	Na sua casa tem: - Máquina de lavar roupa. Sim.	26%	20%	12%
tx_resp_q010i_A	Na sua casa tem: - Freezer (independente ou segunda porta da geladeira). Não.	7%	23%	20%
tx_resp_q010i_B	Na sua casa tem: - Freezer (independente ou segunda porta da geladeira). Sim.	17%	33%	26%
tx_resp_q011_A	Quanto tempo você demora para chegar à sua escola? Menos de 30 minutos.	3%	2%	16%
tx_resp_q011_B	Quanto tempo você demora para chegar à sua escola? Entre 30 minutos e uma hora.	8%	5%	19%
tx_resp_q011_C	Quanto tempo você demora para chegar à sua escola? Mais de uma hora.	1%	0%	6%
tx_resp_q012_A	Considerando a maior distância percorrida, normalmente de que forma você chega à sua escola? À pé.	25%	15%	27%
tx_resp_q012_B	Considerando a maior distância percorrida, normalmente de que forma você chega à sua escola? De ônibus urbano.	0%	3%	15%
tx_resp_q012_C	Considerando a maior distância percorrida, normalmente de que forma você chega à sua escola? De transporte escolar.	9%	1%	1%
tx_resp_q012_D	Considerando a maior distância percorrida, normalmente de que forma você chega à sua escola? De barco.	0%	0%	0%
tx_resp_q012_E	Considerando a maior distância percorrida, normalmente de que forma você chega à sua escola? De bicicleta.	0%	0%	0%
tx_resp_q012_F	Considerando a maior distância percorrida, normalmente de que forma você chega à sua escola? De carro.	31%	27%	29%
tx_resp_q012_G	Considerando a maior distância percorrida, normalmente de que forma você chega à sua escola? Outros meios de transporte.	3%	4%	0%
tx_resp_q017c_A	Fora da escola em dias de aula, quanto tempo você usa para: - Fazer trabalhos domésticos (lavar louça, limpar quintal).	1%	4%	1%
tx_resp_q017c_B	Fora da escola em dias de aula, quanto tempo você usa para: - Fazer trabalhos domésticos (lavar louça, limpar quintal).	14%	6%	14%
tx_resp_q017c_C	Fora da escola em dias de aula, quanto tempo você usa para: - Fazer trabalhos domésticos (lavar louça, limpar quintal).	2%	7%	3%
tx_resp_q017c_D	Fora da escola em dias de aula, quanto tempo você usa para: - Fazer trabalhos domésticos (lavar louça, limpar quintal).	3%	3%	13%
tx_resp_q017e_A	Fora da escola em dias de aula, quanto tempo você usa para: - Trabalhar fora de casa (recebendo ou não um salário).	19%	12%	25%
tx_resp_q017e_B	Fora da escola em dias de aula, quanto tempo você usa para: - Trabalhar fora de casa (recebendo ou não um salário).	4%	1%	1%
tx_resp_q017e_C	Fora da escola em dias de aula, quanto tempo você usa para: - Trabalhar fora de casa (recebendo ou não um salário).	1%	0%	0%
tx_resp_q017e_D	Fora da escola em dias de aula, quanto tempo você usa para: - Trabalhar fora de casa (recebendo ou não um salário).	1%	3%	16%