



### **Giovanna Machado Flores**

Autora aluna da Estatística - IMECC / Unicamp

### **Laetícia R. De Souza**

Orientadora pesquisadora do NEPO / Unicamp

### **Guilherme P. Telles**

Co-orientador pesquisador do IC / Unicamp

## A Olimpíada Brasileira de Informática: Uma análise comparativa entre as escolas que tiveram alunas premiadas e aquelas nas quais apenas alunos foram premiados entre 2007 e 2013.

### **Introdução**

As causas pelas quais mulheres são sub-representadas nas áreas de Ciência, Tecnologia, Engenharias e Matemática (CTEM) no Brasil são sociais [1]. Uma das principais possibilidades é que a escolha da carreira seja afetada pelos estereótipos de gênero, apresentados a meninas e meninos já na infância. As meninas são influenciadas pelas escolhas de brinquedos e brincadeiras, falta de incentivo ao uso de tecnologias e falta de representatividade nas mídias [2]. A falta de modelos femininos em posições influentes na área da tecnologia também pode gerar falta de interesse de outras mulheres para a escolha dessa carreira, e acredita-se que, antes mesmo disso, ter meninas que se interessam por ciências, matemática e tecnologia dentro do convívio social de outras meninas pode influenciá-las positivamente. Isso porque os interesses que uma criança desenvolve estarão dentro de um grupo de possibilidades que lhe for apresentado. Portanto,

um papel importante da família e da escola para combater a desigualdade de gênero nas profissões é mostrar para meninas, desde cedo, que elas possuem liberdade para se interessarem por qualquer área do conhecimento.

Neste contexto, as Olimpíadas Científicas são um ótimo método para gerar interesse de crianças e adolescentes nas áreas de CTEM, já que os (as) estudantes as vêem como um desafio para resolver problemas complexos e a usar seus conhecimentos de maneira mais prática do que na sala de aula.

A OBI (Olimpíada Brasileira de Informática) é realizada anualmente por alunos nos anos finais do ensino fundamental (a partir do 4º ano) até o primeiro ano do Ensino Superior, com a finalidade de despertar nesses alunos o interesse pela ciência da computação ao testar seus conhecimentos em lógica, raciocínio matemático e programação [3]. A OBI é organizada pela Sociedade Brasileira de Computação (SBC) em parceria com o Instituto de Computação da Universidade

Estadual de Campinas (IC - UNICAMP), e acredita-se que é uma olimpíada científica de muita importância, por fomentar a educação em informática nas escolas do Brasil, apresentando essa área para crianças desde cedo.

Moreira, 2019 [5] investigou se o reconhecimento de um aluno impacta seu próprio desempenho acadêmico e o de seus colegas, e chegou a conclusão que sim a partir de dados da OBMEP. Com este projeto, pretendemos comparar as caracte-

terísticas das escolas que tiveram alunas premiadas àquelas em que apenas alunos foram premiados na OBI. Também tentaremos verificar se o fato de ter alunas premiadas numa escola pode despertar interesse (inscrição) de mais outras alunas na OBI.

---

## Objetivos

O projeto pretende responder às seguintes questões:

a) Quais as características diferenciam escolas:

- Participantes da OBI vs. não participantes;
- Participantes sem premiação vs. participantes com alunos(as) premiados(as);
- Nas quais apenas alunos foram premiados vs. nas quais houve pelo menos uma aluna premiada.

b) Essas características têm mudado ao longo do tempo? Como?

c) Entre as escolas com alunas premiadas, houve aumento significativo no número de meninas inscritas em anos subsequentes?

---

## Metodologia

Esse projeto faz parte de uma pesquisa mais ampla na qual se pretende estimar o impacto que um prêmio na OBI possa ter sobre alguns indicadores educacionais. Por isso, é importante conhecer bem as bases de dados a serem utilizadas na análise, de forma a selecionar as variáveis, harmonizar os bancos de dados ano a ano e, finalmente, construir e analisar o banco de dados final. Utilizou-se os dados do Censo Escolar coletados entre 2007 e 2013 e os dados das alunas e alunos inscritos e premiados na OBI para os mesmos anos.

O Censo Escolar é uma pesquisa realizada anualmente pelo INEP em parceria com as secretarias estaduais/municipais de edu-

cação. O objetivo é coletar informações sobre todas as escolas públicas e privadas do país, desde a Educação Infantil até o Ensino Médio, passando por Ensino Profissionalizante e EJA. Entre as informações coletadas estão: número de alunos (as) matriculado (as), turmas, docentes, infraestrutura das escolas, entre outras [4].

Inicialmente, foi preciso manipular os dados do Censo Escolar de forma a facilitar a interpretação dos dados. Foi necessária uma homogeneização dos bancos de dados de 2007 a 2013, já que o censo escolar sofreu alterações nesse período. Foram filtradas apenas as escolas em funcionamento pleno nestes anos, e retiradas as escolas que possuem apenas ensino infantil (que atende crian-

ças na faixa etária de 0 a 5 anos e 11 meses), já que a OBI não contempla essa faixa etária. Além disso, variáveis com poucas respostas ou pouco relevantes para o projeto foram retiradas. Ao analisar as variáveis restantes, havia inconsistências em algumas contagens máximas, que demonstraram erro de preenchimento (por exemplo: 9999 ou 8888 para representar falta de informação) e que foram tratadas.

Para alguns tópicos, houve o interesse de criar novas variáveis a partir das existentes. Por exemplo: Proporção de meninas na educação básica, que é a razão entre a Quantidade de meninas na educação básica e a Quantidade total de alunos na educação básica.

Um desafio para fazer a junção dos dados do Censo Escolar com os da OBI foi que os dados da OBI não possuíam o código único oficial do INEP para escolas. Assim, a junção teve que ser feita com o nome da escola, auxiliada por informações adicionais como CEP e endereço. Essas informações derivam do preenchimento de funcionários(as) e, por isso, podem não coincidir para muitas escolas. Foi criado um script que, para os mesmos cidade e estado, comparou os nomes das escolas existentes nos bancos de dados e computou a porcentagem de combinação que os nomes e os CEPs possuíam. Caso essa comparação fosse perfeita, assumiu-se que se tratava da mesma escola.

Para os casos em que não houve uma porcentagem de combinação suficiente, criou-se uma planilha com as informações dos dois censos para que a confitimação fosse manual. Para cada escola presente no banco de dados da OBI sem combinação, eram mostradas as escolas no banco de dados do Censo Escolar mais prováveis de serem a mesma. O trabalho manual foi checar e investigar, algumas vezes com pesquisa na internet,

de qual escola se tratava. Para que fosse possível seguir o cronograma do projeto, essa combinação pode ser feita apenas para os anos de 2007 e 2013, diminuindo a quantidade de escolas analisadas. A escolha foi do primeiro e último ano do período, para ser possível avaliar algum tipo de progressão das variáveis. Essa medida fez a taxa de escolas não encontradas no Censo Escolar se tornar extremamente baixa.

Após a junção dos dados, além da análise descritiva que inclui tabulações simples e cruzadas, foram empregados testes estatísticos para verificar a existência de diferenças entre os grupos de escolas: participantes versus não participantes; premiadas versus não premiadas; e com alunas premiadas versus apenas alunos premiados.

Para as variáveis categóricas, deve ser utilizado o teste estatístico de independência qui-quadrado, que testa a interação (dependência) entre duas variáveis qualitativas. Neste teste, é calculado o valor esperado de cada par de categorias e uma medida global, chamada  $\chi^2$  de Pearson. A hipótese nula do teste é a de que não há dependência entre as variáveis (não-associação), que é aceita quando o valor da estatística de teste é próximo de zero. Para validar o teste de independência qui-quadrado, é necessário que os dados sejam selecionados aleatoriamente e que todas as frequências esperadas sejam maiores ou iguais a 1. Além disso, não deve haver mais de 20% das frequências esperadas inferiores a 5. [6]

Já para variáveis quantitativas, o teste estatístico utilizado foi o Teste t de Student, que tem finalidade de comparar

as médias de duas amostras independentes. O valor t do teste é calculado a partir da diferença entre as duas médias das amostras e seu desvio padrão. É necessário que essa diferença tenha distribuição aproximadamente normal. O nível de significância escolhido (5%) determina a probabilidade máxima permitida para rejeitar a hipótese nula, que, nesse caso, é a de que os grupos têm

Exemplo de possibilidades de escolha de nomes de escola para "match manual", a partir do nome das escolas:



médias iguais. Se o p-valor obtido for menor do que 5%, rejeita-se H0 e conclui-se que há diferença significativa entre as duas amostras. [7]

Exemplo de transformação de variáveis booleanas para variável categórica (a partir de junção):

ESCOLA	ÁGUA ENCANADA	POÇO ARTESIANO	SEM ÁGUA
A	sim	não	não
B	sim	não	não
C	não	não	sim

➔

ESCOLA	ABASTECIMENTO DE ÁGUA
A	encanada
B	encanada
C	sem abastecimento

## Resultados

O principal resultado do trabalho efetuado foi um banco de dados harmonizado que traz conjuntamente variáveis derivadas do Censo Escolar e informações sobre participação, premiação e quantidade de alunos e alunas premiados (as) para as edições de 2007 e 2013 da OBI.

Na Tabela abaixo, é possível perceber que a porcentagem de escolas com meninas premiadas caiu em 2013 em relação a 2007, mesmo aumentando em número absoluto de escolas. Já a porcentagem de meninas premiadas teve um pequeno aumento, mas ainda está longe de representar metade dos premiados na competição.

	Escolas não participantes	Escolas participantes	Escolas premiadas	Escolas com meninas premiadas	Total de premiados(as)	Meninas premiadas
2007	162355	151 (0,09%)	65 (43%)	35 (53,8%)	227	58 (25,6%)
2013	151373	292 (0,19%)	92 (31,5%)	39 (42,4%)	337	91 (27%)

Ao final da manipulação e homogeneização dos dados do censo escolar, restaram 215 variáveis para a análise. Por ser um número muito elevado, houve uma seleção a mais de variáveis para realizar a análise descritiva (levando em consideração a relevância e a possibilidade de correlação entre as variáveis e a participação/premiação de alunos), restando assim 128 variáveis.

A partir dos resultados sumarizados da análise descritiva e dos testes estatísticos dessas variáveis, é possível extrair informações que levam a algumas hipóteses sobre as diferenças entre as escolas participantes, premiadas e com meninas premiadas na OBI.

A primeira hipótese levantada é a de que as variáveis relacionadas a estrutura da escola tem correlação com a participação na OBI, mas não necessariamente tem grande influência na premiação. Alguns exemplos são as variáveis "Tratamento de esgoto sanitário", "Oferece alimentação para os alunos", "Biblioteca e/ou Sala de leitura existente e utilizada na escola", "Laboratório de informática existentes e utilizado na escola" e "Acesso à Internet", dentre muitas outras omitidas. A partir da variável "Conveniada com o poder público", é possível observar a discrepância entre a proporção de escolas públicas e particulares para participação, premiação e premiação de meninas, que podem ser indicadores de

diferença no incentivo à participação (tanto geral quanto feminina) e qualidade da educação.

Sobre a participação feminina, chama a atenção a diferença na quantidade e proporção de matrículas no ensino profissionalizante técnico. Uma hipótese criada é a de que escolas com ensino técnico em informática (ou similares) possuem grande interesse de participação e ótimo desempenho, porém possuem poucas meninas estudantes. Essas e muitas outras hipóteses criadas a partir da análise estão sendo testadas com o banco de dados criado.

## Referências Bibliográficas

- [1] Blickenstaff, J. C. (2005). Women and science careers: leaky pipeline or gender filter? *Gender and Education*. Vol. 17, No. 4, October 2005, pp. 369–386.
- [2] DA COSTA, B. S. L. M. C. *Gender, Science and Technology: Traveled Paths and New Challenges*. 2016. DOI: 10.1590/18094449201600480005.
- [3] "Sobre a OBI". Autor desconhecido (site oficial). Disponível em: <<https://olimpiada.ic.unicamp.br/info>>. Acesso em: 10/03/2023.
- [4] Ministério da Educação. *Censo Escolar - Resultados*. Disponível em: <<https://www.gov.br/inep/pt-br/areas-de-atuacao/pesquisas-estatisticas-e-indicadores/censo-escolar/resultados>>. Acesso em: 10/03/2023.

- [5] Moreira, D. (2019). Success Spills Over: How Awards Affect Winners' and Peers' Performance in Brazil. Artigo. Disponível em:

<<https://dianamoreira.com/#workingpapers>>. Acesso em 03/05/2023.

- [6] Hazra A, Gogtay N. Biostatistics Series Module 4: Comparing Groups - Categorical Variables. *Indian J Dermatol*. 2016 Jul-Aug; DOI: 10.4103/0019-5154.185700.

- [7] Torman VBL, Coster R, Riboldi J. Normality of variables: diagnosis methods and comparison of some nonparametric tests by simulation. *Rev HCPA*. 2012;32:227-34.

**"Eu quero que toda menina saiba que suas vozes podem mudar o mundo."**

**- Malala Yousafzai**