

Doce Estatística: análise da distribuição de cores do confeito $M\&M^{\text{®}}$ utilizando Python

Matheus C. de Souza,¹ Marcos W. S. Oliveira,² Rosemeire A. Rosa Oliveira
Grupo de Pesquisa em Matemática Científica e Computacional, Instituto Federal de São Paulo IFSP-SJC
Eder A. Milani
IME/UFG, Goiânia, GO
Amanda Buosi Gazon
DEs/UFSCar, São Carlos, SP

Este trabalho descreve um experimento de levantamento e análise de dados adaptável para sala de aula, ações de extensão e eventos, como feiras científicas. O levantamento consiste na contagem de doces disponíveis no mercado, organizando-os conforme suas cores tradicionais, enquanto a análise resulta da construção de gráficos e aplicação de teste de hipótese para proporções, avaliando a distribuição de cores. Com isso, vislumbra-se auxiliar letramento [4] e divulgação científica [5] de conceitos de Matemática e Estatística.

Destaca-se que letramento científico envolve capacitar o público em geral a compreender as características do pensamento científico e seus métodos, conduzindo ao desenvolvimento da cidadania de todos [1]. Enquanto divulgação científica compreende o conjunto de atividades de transmissão de informações sobre inovações científicas e tecnológicas direcionadas ao público leigo [2]. Com isso, ações no contexto deste trabalho provocam a população a refletir criticamente sobre como se realizam levantamentos estatísticos e sobre conceitos de amostragem e margem de erro.

Os doces $M\&M^{\text{®}}$ são produzidos em 6 cores: amarelo, azul, laranja, marrom, verde e vermelho. Nesse contexto, o trabalho [3] se utiliza desses doces para explorar técnicas de amostragem e estatística, com uma amostra de 1196 doces de diferentes produtos da marca. Ainda em [3], indica-se como proporção, reportada pela fábrica: $p_{\text{amarelo}} = 21,4\%$, $p_{\text{azul}} = 14,3\%$, $p_{\text{laranja}} = 21,4\%$, $p_{\text{marrom}} = 14,3\%$, $p_{\text{verde}} = 14,3\%$ e $p_{\text{vermelho}} = 14,3\%$.

O conjunto de dados construído nesta pesquisa advém da contagem de 1819 doces coloridos de 34 diferentes embalagens de mesmo lote do produto $M\&M^{\text{®}}$ de apresentação 45 gramas e sabor chocolate ao leite. Essa amostra garante 90% de chances de representar a realidade com erro de $\pm 2,1\%$, uma vez que é preciso uma amostra de 1805 doces para essa configuração, segundo [7].

Desse levantamento resulta a distribuição apresentada na Tabela 1 que contém as proporções de doces obtidas na amostra. Nota-se que a proporção das cores amarelo, azul, laranja e vermelho aproxima-se de 16%, enquanto que a das cores marrom e verde é de aproximadamente 22% e 13%, respectivamente.

Tabela 1: Proporções dos doces nas respectivas cores obtidas no experimento.

	Amarelo	Azul	Laranja	Marrom	Verde	Vermelho
p	16,49%	16,87%	15,71%	22,34%	13,20%	15,39%

Para avaliação da distribuição de cores obtida na Tabela 1 aplicou-se um teste de hipóteses qui-quadrado (χ^2) de proporções para medir se as proporções da amostra se encaixam às proporções

¹costa.matheus@aluno.ifsp.edu.br

²oliveiramw@ifsp.edu.br

dadas para a população [8]. Utilizou-se sua implementação em Python na biblioteca `scipy.stats` [6]. Com isso, sendo H_0 : a amostra segue as proporções especificadas e H_1 : a amostra não segue as proporções especificadas e considerando 90% de confiança e 5 graus de liberdade, obtem-se o valor crítico $V_c = 9,236$.

Inicialmente, aplicou-se o teste citado considerando a proporção especificada igual à proporção padrão de fábrica, segundo [3]. Obteve-se o valor da estatística $\chi^2 = 141,21$ e o p-valor $< 0,01$, ao nível de significância de 10%, rejeita-se H_0 . Devido a esse resultado, investigou-se o caso da proporção ser uniforme, ou seja, uma proporção igual a 16,67% para cada uma das 6 cores. Neste caso, obteve-se $\chi^2 = 50,87$ e um p-valor $< 0,01$ e, novamente, rejeita-se H_0 , ao nível de significância de 10%.

Naturalmente, diversas características intrínsecas ao processo de produção dos doces influenciam na distribuição das cores. Outrossim, a distribuição apresentada na Tabela 1 advém de uma amostra representativa, nas configurações descritas. Portanto, apesar das abordagens iniciais não identificarem ajuste às distribuições esperadas, uma investigação ao longo do tempo, com análise de diferentes lotes de produção, pode indicar se a distribuição encontrada representa a proporção populacional.

Agradecimentos

M.C.S. agradece apoio do Programa Institucional de Iniciação Científica (PIBIFSP) 2021.

Referências

- [1] C. Attico. “Alfabetização Científica: uma possibilidade para a inclusão social”. Em: **Revista Brasileira de Educação** 22 (2003), pp. 89–100.
- [2] W. D. C. Bueno. “Jornalismo científico: revisitando o conceito”. Em: **Jornalismo científico e desenvolvimento sustentável** (2009), pp. 157–178.
- [3] L. S. Canaes et al. “Using Candy Samples To Learn about Sampling Techniques and Statistical Data Evaluation”. Em: **Journal of Chemical Education** 8 (2008), pp. 1083–1088.
- [4] D. Motta-Roth. “Letramento científico: sentidos e valores”. Em: **Revistas notas de pesquisa** 1 (2011), pp. 12–25.
- [5] A. Sarita. “Divulgação científica: informação científica para cidadania”. Em: **Ciência da Informação** 25 (1996), pp. 396–404.
- [6] SCIPY. **Numpy and Scipy Documentation**. Online. Acessado em 01/03/2021, <https://docs.scipy.org/doc/>.
- [7] S. K. Thompson. “Sample Size for Estimating Multinomial Proportions”. Em: **The American Statistician** 41 (1987), pp. 42–46.
- [8] M. F. Triola. **Elementary Statistics**. 10a. ed. Pearson, 2007. ISBN: 0321331834.