UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ - UTFPR
MESTRADO PROFISSIONAL EM MATEMÁTICA EM REDE NACIONAL
PROFMAT

GUILHERME MIGUEL ROSA

O PROBLEMA DO COLECIONADOR DE CUPONS: QUANTO CUSTA COMPLETAR O ÁLBUM DE FIGURINHAS DA COPA DO MUNDO?

CURITIBA

2019

GUILHERME MIGUEL ROSA

O PROBLEMA DO COLECIONADOR DE CUPONS: QUANTO CUSTA COMPLETAR O ÁLBUM DE FIGURINHAS DA COPA DO MUNDO?

Dissertação apresentada ao Mestrado Profissional em Matemática em Rede Nacional da Universidade Tecnológica Federal do Paraná em Curitiba - PROFMAT-UTCT como requisito parcial para obtenção do grau de Mestre.
Orientador: Roy Wilhelm Probst

CURITIBA

2019

Dados Internacionais de Catalogação na Publicação

Rosa, Guilherme Miguel
O problema do colecionador de cupons [recurso eletrônico] : quanto custa completar o álbum de figurinhas da Copa do Mundo? / Guilherme Miguel Rosa.-- 2019.
1 arquivo texto (30 f.) : PDF; 304 KB.

Principais resultados disponíveis em https://youtu.be/EvO4nJ9LyjY
Modo de acesso: World Wide Web
Título extraído da tela de título (visualizado em 23 ago. 2019)
Texto em português com resumo em inglês
Dissertação (Mestrado) - Universidade Tecnológica Federal do Paraná. Programa de Mestrado Profissional em Matemática em Rede Nacional, Curitiba, 2019
Bibliografia: f. 29-30

1. Matemática - Dissertações. 2. Copas do mundo (Futebol). 3. Copa do mundo (21 : 2018 : Rússia). 4. Figurinhas - Custos. 5. Álbuns de figurinhas - Colecionadores e coleções - Custos. 6. Figurinhas - Colecionadores e coleções - Custos. 7. Probabilidade (Matemática). 8. Variáveis aleatórias. I. Probst, Roy Wilhelm. II. Universidade Tecnológica Federal do Paraná. Programa de Mestrado Profissional em Matemática em Rede Nacional. III. Título.

CDD: ed. 23 -- 510

Biblioteca Central da UTFPR, Câmpus Curitiba
Bibliotecário: Adriano Lopes CRB-9/1429

Ministério da Educação
Universidade Tecnológica Federal do Paraná
Diretoria de Pesquisa e Pós-Graduação

TERMO DE APROVAÇÃO DE DISSERTAÇÃO Nº 69

A Dissertação de Mestrado intitulada **O Problema do Colecionador de Cupons: quanto custa completar o álbum de figurinhas da Copa do Mundo?**, defendida em sessão pública pelo candidato **Guilherme Miguel Rosa**, no dia **16 de agosto de 2019**, foi julgada para a obtenção do título de Mestre em **Matemática**, área de concentração **Matemática Aplicada**, e aprovada em sua forma final, pelo Programa de Pós-Graduação em **Matemática em Rede Nacional**.

BANCA EXAMINADORA:

Prof. Dr. Roy Wilhelm Probst - Presidente - UTFPR
Prof. Dr. Rafael Santos de Oliveira Alves - UFABC
Prof. Dr. Ronie Peterson Dario - UTFPR

A via original deste documento encontra-se arquivada na Secretaria do Programa, contendo a assinatura da Coordenação após a entrega da versão corrigida do trabalho.

Curitiba, 16 de agosto de 2019.

Carimbo e Assinatura do(a) Coordenador(a) do Programa

AGRADECIMENTOS

Aos meus pais Lidio Rosa e Maria Alves Rosa por todo o apoio fornecido ao longo dessa jornada.

A todos os meus alunos virtuais espalhados pelo Brasil inteiro, que direta ou indiretamente continuam possibilitando meu trabalho com vídeo aulas na internet.

Em especial ao meu orientador Roy Wilhelm Probst por todas as sugestões e contribuições fundamentais para a conclusão dessa dissertação.

RESUMO

ROSA, Guilherme Miguel. **O Problema do Colecionador de Cupons:** quanto custa completar o álbum de figurinhas da Copa do Mundo? 30 f. Dissertação - Programa de Mestrado Profissional em Matemática em Rede Nacional - PROFMAT, Universidade Tecnológica Federal do Paraná. Curitiba, 2019.

Este trabalho apresenta um estudo sobre o Problema do Colecionador de Cupons aplicado ao desafio de colecionar as figurinhas do álbum oficial da Copa do Mundo de futebol de 2018. O custo médio para completar um álbum é determinado considerando diferentes cenários, como a colaboração entre colecionadores. As soluções obtidas são validadas através de resultados numéricos que simulam o processo de coleção das figurinhas.

Palavras-chave: Probabilidade. Valor esperado. Problema do Colecionador de Cupons.

ABSTRACT

ROSA, Guilherme Miguel. **The Coupon Collector's Problem:** how much does it cost to complete the World Cup stickers album? 30 pg. Dissertation - Programa de Mestrado Profissional em Matemática em Rede Nacional - PROFMAT, Universidade Tecnológica Federal do Paraná. Curitiba, 2019.

This work presents a study on the Coupon Collector's Problem applied to the challenge of collecting the soccer stickers of the official World Cup album of 2018. The average cost to complete an album is determined by considering different scenarios such as collector collaboration. The obtained solutions are validated through numerical results that simulate the collection process of the stickers.

Keywords: Probability. Expected value. Coupon Collector's Problem.

SUMÁRIO

INTRODUÇÃO

A cada quatro anos acontece a Copa de Mundo de futebol e com ela retorna uma prática muito popular entre os apaixonados pelo esporte. Trata-se da coleção de figurinhas do álbum oficial do mundial, cujo desafio consiste em conseguir todas as figurinhas do álbum comprando pacotes com figurinhas aleatórias. Como são adquiridas aleatoriamente, é natural que conforme a coleção avança apareçam figurinhas repetidas, isto é, figurinhas já obtidas pelo colecionador anteriormente. Com isso é preciso seguir comprando até conseguir todas as figurinhas necessárias. É nesse momento que surge uma questão incluindo a matemática na brincadeira e tornando-a ainda mais interessante: quantas figurinhas é necessário comprar em média para conseguir pelo menos uma de cada e completar uma coleção?

Esse é um problema que já foi estudado na área da teoria das probabilidades e ficou conhecido na literatura matemática como O Problema do Colecionador de Cupons. Sua descrição é simples: considere uma pessoa que coleciona objetos (cupons) e suponha que existe um número finito de objetos (cupons) diferentes para serem colecionados. Adquirindo-se aleatoriamente um por vez, quantas aquisições são necessárias em média para completar uma coleção?

O problema tem uma longa história, que de acordo com (FERRANTE; SALTALAMACCHIA, 2014) iniciou em 1708 com De Moivre na obra *De Mensura Sortis* e teve contribuições de Laplace e Euler. Desde o surgimento até não muito distante, diversas variações e abordagens distintas foram exploradas. Considerando que todos os elementos são igualmente prováveis, Newman e Shepp desenvolveram o modelo que generaliza o problema para o caso envolvendo as trocas entre colecionadores, determinando a quantidade esperada de aquisições quando deseja-se m cópias de cada elemento da coleção (NEWMAN; SHEPP, 1960). Em (KOBZA; JACOBSON; VAUGHAN, 2007) os autores discutem o problema com a suposição de que em cada aquisição obtém-se uma quantidade aleatória de elementos igualmente prováveis. Quando os objetos da coleção não são todos igualmente prováveis, ou seja, quando alguns são mais difíceis de conseguir, tem-se os trabalhos de (SCHELLING, 1954), (FLAJOLET; GARDY; THIMONIER, 1992) e (PAPANICOLAOU; KOKOLAKIS; BONEH, 1998). Nesse penúltimo inclusive, considera-se também a coleção de um subconjunto de objetos. Supondo agora que os elementos chegam aleatoriamente em grupos de tamanho constante não havendo repetições em um mesmo grupo, tem-se os trabalhos de (STADJE, 1990) e (FERRANTE; SALTALAMACCHIA, 2014). Outra adaptação do problema está presente em (MAY, 2008) e (SHANK; YANG, 2013): ela considera a situação em que deseja-se colecionar uma quantidade específica de cada objeto, denominada cota, com as probabilidades de ocorrência não necessariamente iguais e as quantidades de cada objeto não necessariamente constantes.

Em português, foi publicado um artigo na Revista do Professor de Matemática (RPM) falando sobre a quantidade esperada de figurinhas compradas para completar um certo álbum sem trocas e com probabilidades iguais (CARVALHO, 2010). Outro trabalho mais recente em português analisa a relação entre as figurinhas faltantes de cada álbum com simulações em quatro modelos (BASSANEZI, 2018). No primeiro, deseja-se completar três álbuns respeitando uma hierarquia de álbuns na colagem das figurinhas, que possuem a mesma probabilidade de ocorrência. O segundo modelo considera a troca de figurinhas repetidas entre dois colecionadores. No terceiro considera-se dois tipos de figurinhas com probabilidades diferentes de ocorrência, onde deseja-se completar um álbum. E no último modelo o objetivo é encontrar a menor quantidade de figurinhas necessárias para completar três álbuns, trocando duas repetidas por uma não colecionada. Ao final, simula o custo para completar o álbum da última Copa do Mundo.

O clássico Problema do Colecionador de Cupons despertou interesse com a motivação de muitas aplicações possíveis em diversas áreas da ciência. Algumas aplicações representativas estão na biologia (POON; DAVIS; CHAO, 2005), linguística (RAFFERTY; GRIFFITHS; KLEIN, 2009), ciência da computação (O'CONNOR, 2008) e criptografia (SASAKI et al., 2013). Muitas aplicações aparecem em publicações recentes, reforçando a relevância do problema.

Nesta dissertação será estudado o Problema do Colecionador de Cupons em sua versão mais simples, que é o caso de elementos equiprováveis chegando um por vez. O objetivo geral é determinar o custo médio para completar o álbum de figurinhas oficial da última Copa do Mundo, realizada em 2018 na Rússia. Os objetivos específicos são:

- Modelar e resolver o Problema do Colecionador de Cupons.
- Generalizar o problema para o caso envolvendo as trocas de figurinhas repetidas.
- Implementar a solução geral para obter o número esperado de figurinhas por colecionador.
- Realizar simulações observando a quantidade de figurinhas necessárias por colecionador, para validar os resultados teóricos.

O texto está organizado em dois capítulos. No Capítulo 1 será modelado e resolvido o Problema do Colecionador de Cupons, tanto na sua versão clássica como na generalização envolvendo as trocas de figurinhas. Em ambos os casos é necessário o auxílio de recursos computacionais para a obtenção dos resultados numéricos, é o que será realizado no Capítulo 2 com a implementação das fórmulas. Ainda no segundo capítulo, são implementados algoritmos que simulam o processo de coleção das figurinhas para comparar com os resultados teóricos.

1 O PROBLEMA DO COLECIONADOR DE CUPONS

O objetivo desse capítulo é encontrar uma maneira de calcular quantas figurinhas é necessário comprar em média para completar o álbum oficial da Copa do Mundo de 2018. Inicialmente, a primeira seção apresenta alguns conceitos necessários e, na segunda seção, a solução para o Problema do Colecionador de Cupons será construída. Posteriormente, o problema de interesse na versão que não envolve a troca de figurinhas repetidas é resolvido simplesmente como um caso particular da solução geral. Porém, completar uma coleção sem trocar os elementos repetidos certamente não é a estratégia mais interessante, ainda mais uma coleção grande como as figurinhas do álbum da Copa do Mundo. Assim, o capítulo apresenta também um estudo acerca da relação cooperativa entre dois ou mais colecionadores, o que torna o Problema do Colecionador de Cupons mais próximo da realidade e mais interessante no aspecto matemático também. Nesse contexto, é preciso determinar quantos elementos espera-se adquirir para completar m coleções de n elementos. A resposta para esse problema já foi dada por Newman e Shepp (NEWMAN; SHEPP, 1960), as últimas seções desse capítulo baseiam-se nesses resultados, trazendo uma releitura de todos eles.

1.1 CARACTERÍSTICAS DO MODELO

No Problema do Colecionador de Cupons existe um experimento aleatório com características específicas. Cada aquisição aleatória de um novo elemento para a coleção apresenta dois resultados possíveis, denominados de sucesso ou fracasso. O sucesso é quando o colecionador consegue um elemento inédito na coleção, e o fracasso é ele obter um dos já colecionados. Quando ocorre fracasso, a coleção não avança, ou seja, o colecionador precisa repetir o experimento até alcançar sucesso. Resumindo, tem-se um experimento aleatório com as seguintes características:

- Resulta sucesso ou fracasso.
- Será repetido até ocorrer o primeiro sucesso.
- Todas as tentativas são independentes.
- A probabilidade de sucesso é constante em cada tentativa.

Essas características descrevem uma distribuição geométrica de probabilidades. Como o experimento será repetido até que haja o primeiro sucesso, é interessante saber o número de tentativas necessárias para isso. O número exato de tentativas não existe, mas uma média sim, ela é conhecida como valor médio ou valor esperado. Uma definição de valor esperado é dada a seguir (CARVALHO, 2010).

Definição 1.1. *Suponha que um determinado resultado aleatório pode assumir os valores x_1, x_2, x_3, $\cdots$, com probabilidades p_1, p_2, p_3, $\cdots$. Define-se seu valor médio (ou esperado) como:*

$$E = p_1 \cdot x_1 + p_2 \cdot x_2 + p_3 \cdot x_3 + \cdots .$$

Essa definição considera a possibilidade de ter-se uma infinidade de resultados possíveis. A soma é uma média ponderada dos valores de x sendo as respectivas probabilidades os pesos. Porém, a soma das probabilidades resulta um, assim o denominador da média não aparece.

Comparando a definição com o experimento caracterizado anteriormente, a probabilidade p_i é a probabilidade do primeiro sucesso ocorrer na tentativa de número i, $i = 1, 2, 3, \cdots$. Basta determinar quais são as probabilidades $p_1, p_2, p_3, \cdots$, e por meio da definição obter a quantidade de tentativas necessárias, em média, até haver o primeiro sucesso.

Sendo p a probabilidade do sucesso, com $0 < p < 1$, a probabilidade do fracasso será $(1-p)$. Com isso, a probabilidade do primeiro sucesso ocorrer na n-ésima tentativa é $(1-p)^{n-1}p$, pois certamente houve fracasso em $(n-1)$ tentativas e sucesso em uma, a última. Portanto, pela definição vista anteriormente, o número médio ou esperado de tentativas até o primeiro sucesso é dado por:

$$E = p + 2 \cdot (1-p)^1 p + 3 \cdot (1-p)^2 p + \cdots .$$

Para efetuar a soma e determinar o valor esperado E, uma possibilidade é reagrupar os termos escrevendo todos com coeficiente igual a 1, dispondo em linhas e colunas da seguinte maneira:

$$\begin{aligned} E = p + (1-p)^1 p + (1-p)^2 p + (1-p)^3 p + \cdots \\ +(1-p)^1 p + (1-p)^2 p + (1-p)^3 p + \cdots \\ +(1-p)^2 p + (1-p)^3 p + \cdots \\ +(1-p)^3 p + \cdots . \end{aligned}$$

Pode-se mostrar que a série é absolutamente convergente utilizando o teste da razão (STEWART, 2013), portanto é possível reagrupar os termos desta forma. Na primeira coluna, coloca-se os termos p, como só existe um ele aparece somente na primeira linha. Na segunda coluna estão os termos $(1-p)^1 p$, que aparecem nas duas primeiras linhas já que o coeficiente desse termo é 2. Na sequência, o termo $(1-p)^2 p$ está nas três primeiras linhas já que seu coeficiente é 3, e assim segue sucessivamente. Olhando para as linhas, na primeira todos os termos de E aparecem uma vez. Na próxima, não há o termo p, em seguida, não há também o termo $(1-p)^1 p$, e assim por diante. Em cada linha, basta remover o primeiro termo para obter a seguinte.

Note que em todas as linhas surge uma soma de infinitos termos em progressão geométrica de razão $(1-p)$. As somas resultam um número real pois $0 < (1-p) < 1$. Realizando as somas das primeiras linhas, os resultados são:

$$\frac{p}{1-(1-p)},\ \frac{(1-p)p}{1-(1-p)},\ \frac{(1-p)^2p}{1-(1-p)},\ \cdots .$$

Esses números também estão em progressão geométrica, ou seja, somando todas as linhas o resultado é uma nova soma de infinitos termos em progressão geométrica, agora com o primeiro termo igual a 1 e a razão novamente igual a $(1-p)$. Portanto:

$$\begin{aligned} E &= \frac{p}{1-(1-p)} + \frac{(1-p)p}{1-(1-p)} + \frac{(1-p)^2p}{1-(1-p)} + \cdots \\ &= 1 + (1-p) + (1-p)^2 + \cdots \\ &= \frac{1}{1-(1-p)} \\ &= \frac{1}{p}. \end{aligned}$$

Conclui-se que em uma distribuição geométrica, o valor esperado correspondente ao número médio de tentativas até conseguir o primeiro sucesso é o inverso da probabilidade do sucesso.

Exemplo 1.2. Uma urna contém dez bolas numeradas de 1 a 10 e uma bola será retirada da urna ao acaso, com reposição. Quantas retiradas são esperadas até que se consiga uma bola de número múltiplo de 4?

A probabilidade do sucesso em cada retirada é $2/10$, pois tanto a bola 4 como a 8 servem, logo, são necessárias em média $10/2 = 5$ retiradas.

$\square$

1.2 COMPLETANDO UM ÁLBUM

Completar uma coleção finita de n objetos sem trocar os adquiridos repetidamente consiste no clássico Problema do Colecionador de Cupons. Sua solução, que corresponde ao número de unidades que precisam ser adquiridas em média para concluir o desafio, é

$$n \cdot \sum_{i=1}^{n} \frac{1}{i}.$$

Para chegar na solução, basta observar a probabilidade de se obter um elemento ainda não colecionado em cada etapa da coleção, tomar o inverso das probabilidades e realizar o somatório. Seja k o número de elementos já obtidos entre os n totais da coleção, a probabilidade de se obter um elemento ainda não colecionado na próxima aquisição é

$$\frac{n-k}{n}.$$

Adquirido um novo elemento, ele pode ser o $(k+1)$-ésimo inédito, o que representa o sucesso do experimento, mas também pode ser um mesmo elemento dentre os k já colecionados. Nesse caso, o colecionador repetirá o experimento, ou seja, terá que adquirir mais um elemento e seguir assim até alcançar o sucesso. Significa que o número médio de tentativas até obter o $(k+1)$-ésimo elemento da coleção é

$$\frac{n}{n-k}.$$

Por exemplo, para $k=0$, é necessário somente uma tentativa para ganhar o primeiro elemento da coleção, o que faz sentido visto que qualquer um será inédito. Vale observar que o último valor assumido por k é $n-1$, já que o próximo elemento inédito adquirido completa a coleção. Fazendo $k = 0, 1, 2, ..., n-1$, tem-se a sequência

$$\frac{n}{n}, \frac{n}{n-1}, \frac{n}{n-2}, ..., \frac{n}{n-(n-1)}.$$

A soma da sequência acima corresponde ao número médio ou esperado de unidades adquiridas para se conseguir todos os n que completam a coleção. Seja E tal soma, ela é dada por:

$$E = \frac{n}{n} + \frac{n}{n-1} + \frac{n}{n-2} + ... + \frac{n}{n-(n-1)}.$$

Assim:

$$E = \sum_{i=0}^{n-1} \frac{n}{n-i}. \tag{1.1}$$

Uma maneira mais compacta de escrever a soma é percebendo que os denominadores são naturais de 1 até n, logo:

$$E = \sum_{i=1}^{n} \frac{n}{i}. \tag{1.2}$$

Porém, a forma usada em (1.1) é mais interessante para observar o número médio de elementos a serem adquiridos até colecionar um certo número deles, ou ainda, para saber a quantidade a ser adquirida para conseguir apenas os últimos da coleção por exemplo.

1.3 COMPLETANDO MAIS DE UM ÁLBUM

Deseja-se completar m coleções iguais de n elementos. Cada coleção pode ser interpretada como um álbum cujos elementos são as figurinhas que o compõem. Seja p_i a probabilidade de não completar m álbuns após adquirir i figurinhas e $E_m(n)$ a quantidade esperada de figurinhas compradas para completar os m álbuns de n figurinhas. Logo:

Lema 1.3.

$$E_m(n) = \sum_{i=0}^{\infty} p_i.$$

Demonstração. A demonstração para o lema baseia-se em (FELLER, 1950). Pela Definição (1.1) tem-se:

$$E = \sum_{j=1}^{\infty} j \cdot p_j.$$

Fazendo $p_{k+1} + p_{k+2} + \cdots = q_k$, $k \geq 0$, o valor esperado E será igual a

$$\sum_{k=0}^{\infty} q_k.$$

Isso é facilmente verificado, pois $q_0 = p_1+p_2+p_3+\cdots$, $q_1 = p_2+p_3+\cdots$, $q_2 = p_3+\cdots$, e assim por diante. Logo

$$q_0 + q_1 + q_2 + \cdots = p_1 + 2p_2 + 3p_3 + \cdots .$$

Considerando uma distribuição geométrica de probabilidades com probabilidade p de sucesso, já foi mostrado que o valor esperado é $1/p$ usando a probabilidade do sucesso para tal. Para provar o lema basta mostrar que

$$\sum_{k=0}^{\infty} q_k = \frac{1}{p},$$

onde q_k é a probabilidade de fracasso em cada tentativa. De fato, basta observar na igualdade $q_k = p_{k+1} + p_{k+2} + \cdots$ que q_k é a probabilidade de fracasso na k-ésima tentativa. Para haver fracasso na k-ésima tentativa, pode ocorrer sucesso na tentativa $k+1$, ou $k+2$, ou $k+3$ e assim por diante.

Dessa maneira, $q_k = (1-p)^k$, já que todas as k primeiras tentativas devem fracassar, e com isso

$$E = \sum_{k=0}^{\infty} q_k = \sum_{k=0}^{\infty} (1-p)^k = \frac{1}{p}.$$

□

Exemplo 1.4. Quantas vezes espera-se lançar uma moeda até obter uma coroa?

A probabilidade de obter coroa em cada lançamento é $1/2$. Portanto, espera-se que sejam necessários 2 lançamentos. Pelo Lema (1.3):

$$E = \sum_{n=0}^{\infty} \left(\frac{1}{2}\right)^n = 2.$$

□

A quantidade de maneiras diferentes de se adquirir i figurinhas, uma por vez, é n^i, pois existem n possibilidades em cada retirada. De todas as possibilidades, N_i representa o número de maneiras nas quais não se completam os m álbuns desejados. Assim:

$$p_i = \frac{N_i}{n^i}.$$

Exemplo 1.5. Seja $n = 2$ (2 figurinhas, representadas por a e b) e $m = 2$ (2 álbuns).

Para $i = 2$, tem-se $N_2 = 4$, pois as possibilidades aa, ab, ba, bb não completam os dois álbuns. Esses são exatamente os termos que surgem no desenvolvimento de $(a + b)^2$. Para N_4 por exemplo, basta analisar $(a + b)^4$:

$$(a + b)^4 = a^4 + 4a^3b + 6a^2b^2 + 4ab^3 + b^4.$$

O termo $6a^2b^2$ indica 6 maneiras de comprar duas figurinhas a e duas figurinhas b, o que completa os dois álbuns, eis as possibilidades: $aabb, bbaa, abab, baba, abba, baab$. Conclui-se que $N_4 = 10$, pois só contam as possibilidades que não completam os dois álbuns.

De modo geral, N_i é a soma dos coeficientes de $(a + b)^i$ desconsiderando os termos onde ambos os expoentes são maiores ou iguais a 2. Assim:

i	$(a+b)^i$	N_i
0	1	1
1	$a + b$	2
2	$a^2 + 2ab + b^2$	4
3	$a^3 + 3a^2b + 3ab^2 + b^3$	8
4	$a^4 + 4a^3b + 6a^2b^2 + 4ab^3 + b^4$	10
5	$a^5 + 5a^4b + 10a^3b^2 + 10a^2b^3 + 5ab^4 + b^5$	12
6	$a^6 + 6a^5b + 15a^4b^2 + 20a^3b^3 + 15a^2b^4 + 6ab^5 + b^6$	14
7	$a^7 + 7a^6b + \cdots + 7ab^6 + b^7$	16
8	$a^8 + 8a^7b + \cdots + 8ab^7 + b^8$	18
$\vdots$	$\vdots$	$\vdots$
k	$a^k + ka^{k-1}b + \cdots + kab^{k-1} + b^k$	$2k + 2$

Tabela 1 – Desenvolvimento $(a + b)^i$

Logo, pode-se determinar a quantidade esperada de figurinhas compradas para completar 2 álbuns com 2 figurinhas. Pelo Lema (1.3):

$$\begin{aligned} E_2(2) &= \frac{1}{2^0} + \frac{2}{2^1} + \frac{4}{2^2} + \frac{8}{2^3} + \frac{10}{2^4} + \frac{12}{2^5} + \frac{14}{2^6} + \cdots \\ &= 1 + 1 + 1 + 1 + \sum_{k=4}^{\infty} \frac{2k+2}{2^k} \\ &= 4 + 1{,}5 \\ &= 5{,}5. \end{aligned}$$

□

O somatório foi calculado utilizando o software WolframAlpha (WOLFRAM, 2009). Mais detalhes sobre o cálculo de somatórios com essa ferramenta serão dados no Capítulo 2.

1.4 COMPLETANDO m ÁLBUNS DE n FIGURINHAS

Generalizando para m álbuns de n figurinhas, representadas por $(x_1, x_2, \cdots, x_n)$, N_i é a soma dos coeficientes do desenvolvimento de $(x_1 + x_2 + \cdots + x_n)^i$ após removidos os termos com todos os expoentes $\geq m$. Alternativamente, N_i é $(x_1 + x_2 + \cdots + x_n)^i$ expandido e avaliado em $x_1 = \cdots = x_n = 1$ depois que os termos com todos os expoentes $\geq m$ são removidos.

Definição 1.6. *Para m fixo, se $P(x_1, x_2, \cdots, x_n)$ é um polinômio ou série de potências, então $\{P(x_1, x_2, \cdots, x_n)\}$ é o polinômio ou série de potências resultante quando os termos com todos os expoentes $\geq m$ são removidos.*

Em termos dessa definição:

$$p_i = \frac{\{(x_1 + x_2 + \cdots + x_n)^i\}}{n^i},$$

avaliado em $x_1 = \cdots = x_n = 1$.

Agora faz-se necessário a utilização da série de Maclaurin da função exponencial:

$$e^t = \sum_{k=0}^{\infty} \frac{t^k}{k!} = 1 + t + \frac{t^2}{2!} + \frac{t^3}{3!} + \cdots .$$

Definição 1.7.

$$S_m(t) = \sum_{k<m} \frac{t^k}{k!} = 1 + t + \frac{t^2}{2!} + \cdots + \frac{t^{m-1}}{(m-1)!}.$$

De acordo com a definição acima, $S_m(t)$ é obtida removendo os termos da série e^t que apresentam expoente $\geq m$, ou seja, é um truncamento da série da função exponencial.

Definição 1.8. $F = e^{x_1 + \cdots + x_n} - (e^{x_1} - S_m(x_1))(e^{x_2} - S_m(x_2)) \cdots (e^{x_n} - S_m(x_n)).$

Para explorar a expressão F, o exemplo a seguir traz o caso $m = n = 2$.

Exemplo 1.9. Considere $n = 2$ (por exemplo, a e b) e $m = 2$ (ou seja $S_2(t) = 1 + t$).

Pela Definição (1.8):

$$F = e^{a+b} - [e^a - (1 + a)][e^b - (1 + b)].$$

Mas note que:

$$
\begin{aligned}
\{e^{a+b}\} = \{e^a e^b\} &= \left\{\left(1+a+\frac{a^2}{2!}+\cdots\right)\left(1+b+\frac{b^2}{2!}+\cdots\right)\right\} \\
&= (1+a)\left(1+b+\frac{b^2}{2!}+\cdots\right)+(1+b)\left(\frac{a^2}{2!}+\frac{a^3}{3!}+\cdots\right) \\
&= S_2(a)e^b + S_2(b)(e^a - S_2(a)) \\
&= e^a e^b - e^a e^b + S_2(a)e^b + S_2(b)(e^a - S_2(a)) \\
&= e^{a+b} - e^b(e^a - S_2(a)) + S_2(b)(e^a - S_2(a)) \\
&= e^{a+b} - (e^a - S_2(a))(e^b - S_2(b)) \\
&= F.
\end{aligned}
$$

Logo, $F = \{e^{a+b}\}$.

$\square$

Exemplo 1.10. Suponha $n = 2$ (a e b) e m qualquer.

Pela Definição (1.8):

$$
F = e^{a+b} - \left[e^a - \left(1+a+\frac{a^2}{2!}+\cdots+\frac{a^{m-1}}{(m-1)!}\right)\right]\left[e^b - \left(1+b+\frac{b^2}{2!}+\cdots+\frac{b^{m-1}}{(m-1)!}\right)\right].
$$

Mas note que:

$$
\begin{aligned}
\{e^{a+b}\} = \{e^a e^b\} &= \left\{\left(1+a+\frac{a^2}{2!}+\cdots\right)\left(1+b+\frac{b^2}{2!}+\cdots\right)\right\} \\
&= \left(1+a+\cdots+\frac{a^{m-1}}{(m-1)!}\right)(1+b+\cdots) \\
&\quad + \left(1+b+\cdots+\frac{b^{m-1}}{(m-1)!}\right)\left(\frac{a^m}{m!}+\frac{a^{m+1}}{(m+1)!}+\cdots\right) \\
&= S_m(a)e^b + S_m(b)(e^a - S_m(a)) \\
&= e^a e^b - e^a e^b + S_m(a)e^b + S_m(b)(e^a - S_m(a)) \\
&= e^{a+b} - e^b(e^a - S_m(a)) + S_m(b)(e^a - S_m(a)) \\
&= e^{a+b} - (e^a - S_m(a))(e^b - S_m(b)) \\
&= F.
\end{aligned}
$$

Logo, $F = \{e^{a+b}\}$.

$\square$

O próximo resultado generaliza os exemplos anteriores para n e m qualquer:

Lema 1.11. $F = \{e^{x_1+\cdots+x_n}\} = e^{x_1+\cdots+x_n} - (e^{x_1} - S_m(x_1))\cdots(e^{x_n} - S_m(x_n))$.

Demonstração. A demonstração se faz por indução sobre n. Para $n = 1$:

$$\begin{aligned} e^{x_1} - (e^{x_1} - S_m(x_1)) &= S_m(x_1) \\ &= 1 + x_1 + \cdots + \frac{x_1^{m-1}}{(m-1)!} \\ &= \{e^{x_1}\}. \end{aligned}$$

Agora, pela hipótese de indução, suponha válido para $n = k$, ou seja:

$$\{e^{x_1+\cdots+x_k}\} = e^{x_1+\cdots+x_k} - (e^{x_1} - S_m(x_1))\cdots(e^{x_k} - S_m(x_k)).$$

Para $n = k+1$:

$$\{e^{x_1+\cdots+x_k+x_{k+1}}\} = e^{x_1+\cdots+x_{k+1}} - (e^{x_1} - S_m(x_1))\cdots(e^{x_{k+1}} - S_m(x_{k+1})).$$

Pela hipótese de indução, $e^{x_1+\cdots+x_k} - (e^{x_1} - S_m(x_1))\cdots(e^{x_k} - S_m(x_k))$ não tem termos com todos os expoentes $\geq m$. Basta analisar como podem ser gerados todos os termos de $\{e^{x_1+\cdots+x_k+x_{k+1}}\}$ por meio de $\{e^{x_1+\cdots+x_k}\}$. Inicialmente, pode-se multiplicar $e^{x_{k+1}}$ por todos os termos de $\{e^{x_1+\cdots+x_k}\}$, já que nenhum deles apresenta todos os expoentes $\geq m$. Além disso, a diferença $e^{x_1+\cdots+x_k} - \{e^{x_1+\cdots+x_k}\}$ (que gera apenas os termos de $e^{x_1+\cdots+x_k}$ com todos os expoentes $\geq m$) pode ser multiplicada por $S_m(x_{k+1})$ pois esse traz os termos de $e^{x_{k+1}}$ com expoentes $< m$. Logo:

$$\begin{aligned} &\{e^{x_1+\cdots+x_k+x_{k+1}}\} \\ &= e^{x_{k+1}}\{e^{x_1+\cdots+x_k}\} + S_m(x_{k+1})[e^{x_1+\cdots+x_k} - \{e^{x_1+\cdots+x_k}\}] \\ &= e^{x_{k+1}}\{e^{x_1+\cdots+x_k}\} + S_m(x_{k+1})e^{x_1+\cdots+x_k} - S_m(x_{k+1})\{e^{x_1+\cdots+x_k}\} \\ &= \{e^{x_1+\cdots+x_k}\}[e^{x_{k+1}} - S_m(x_{k+1})] + S_m(x_{k+1})e^{x_1+\cdots+x_k} \\ &= [e^{x_1+\cdots+x_k} - (e^{x_1} - S_m(x_1))\cdots(e^{x_k} - S_m(x_k))][e^{x_{k+1}} - S_m(x_{k+1})] + S_m(x_{k+1})e^{x_1+\cdots+x_k} \\ &= e^{x_1+\cdots+x_k+x_{k+1}} - e^{x_{k+1}}(e^{x_1} - S_m(x_1))\cdots(e^{x_k} - S_m(x_k)) - S_m(x_{k+1})e^{x_1+\cdots+x_k} \\ &\quad + S_m(x_{k+1})(e^{x_1} - S_m(x_1))\cdots(e^{x_k} - S_m(x_k)) + S_m(x_{k+1})e^{x_1+\cdots+x_k} \\ &= e^{x_1+\cdots+x_k+x_{k+1}} - e^{x_{k+1}}(e^{x_1} - S_m(x_1))\cdots(e^{x_k} - S_m(x_k)) \\ &\quad + S_m(x_{k+1})(e^{x_1} - S_m(x_1))\cdots(e^{x_k} - S_m(x_k)) \\ &= e^{x_1+\cdots+x_k+x_{k+1}} - (e^{x_1} - S_m(x_1))\cdots(e^{x_k} - S_m(x_k))[e^{x_{k+1}} - S_m(x_{k+1})]. \end{aligned}$$

□

Voltando ao problema de achar a quantidade esperada de figurinhas para completar m álbuns de n figurinhas, pelo Lema (1.3):

$$E_m(n) = \sum_{i=0}^{\infty} p_i = \sum_{i=0}^{\infty} \frac{\{(x_1 + x_2 + \cdots + x_n)^i\}}{n^i},$$

avaliado em $x_1 = x_2 = \cdots = x_n = 1$.

Por outro lado, pelo Lema (1.11) e da série de Maclaurin da função exponencial:

$$F = \sum_{i=0}^{\infty} \frac{\{(x_1 + x_2 + \cdots + x_n)^i\}}{i!}.$$

Para usar F no cálculo de $E_m(n)$ é preciso trocar $1/i!$ por $1/n^i$, e isso é obtido com a identidade a seguir.

Lema 1.12.

$$n \int_0^{\infty} \frac{t^i}{i!} e^{-nt} dt = \frac{1}{n^i},\ i = 0, 1, 2,\ \ldots\ .$$

Demonstração. Para demonstrar o lema, utiliza-se a indução. Para $i = 0$:

$$\begin{aligned} n \int_0^{\infty} \frac{t^0}{0!} e^{-nt} dt &= n \lim_{T \to \infty} \int_0^T e^{-nt} dt \\ &= n \lim \left. \frac{e^{-nt}}{-n} \right|_0^T \\ &= \lim_{T \to \infty} [-e^{-nT} + e^0] \\ &= 1 \\ &= \frac{1}{n^0}. \end{aligned}$$

Supondo válido para $i = k$:

$$n \int_0^{\infty} \frac{t^k}{k!} e^{-nt} dt = \frac{1}{n^k},$$

será mostrado que continua válido para $i = k + 1$:

$$n \int_0^{\infty} \frac{t^{k+1}}{(k+1)!} e^{-nt} dt = \frac{1}{n^{k+1}}.$$

Integrando por partes, $\displaystyle\int u\ dv = uv - \int v\ du$, onde:

$$u = \frac{t^{k+1}}{(k+1)!},\ du = \frac{t^k}{k!} dt,\ v = \frac{e^{-nt}}{-n}\ \text{ e }\ dv = e^{-nt} dt,$$

tem-se:

$$\begin{aligned} n \int_0^{\infty} \frac{t^{k+1}}{(k+1)!} e^{-nt} dt &= n \lim_{T \to \infty} \int_0^T \frac{t^{k+1}}{(k+1)!} e^{-nt} dt \\ &= n \left[\lim_{T \to \infty} \frac{t^{k+1}}{(k+1)!} \cdot \left. \frac{e^{-nt}}{-n} \right|_0^T - \lim_{T \to \infty} \int_0^T \frac{t^u}{u!} \cdot \frac{e^{-nt}}{-n} dt \right] \\ &= \frac{n}{n} \int_0^{\infty} \frac{t^k}{k!} e^{-nt} dt \\ &= \frac{1}{n} \cdot \frac{1}{n^k} \\ &= \frac{1}{n^{k+1}}. \end{aligned}$$

□

Finalmente, agora é possível demonstrar o principal resultado desse capítulo.

Teorema 1.13. $E_m(n) = n\int_0^\infty 1 - [1 - e^{-t}S_m(t)]^n dt.$

Demonstração. Pelos Lemas (1.12) e (1.11):

$$\begin{aligned}\sum_{i=0}^{\infty}\frac{\{(x_1+\cdots+x_n)^i\}}{n^i} &= \sum_{i=0}^{\infty}\left[\{(x_1+\cdots+x_n)^i\}n\int_0^\infty \frac{t^i}{i!}e^{-nt}dt\right]\\ &= n\int_0^\infty \sum_{i=0}^{\infty}\left[\frac{\{(x_1+\cdots+x_n)^i\}t^i}{i!}\right]e^{-nt}dt\\ &= n\int_0^\infty [e^{t(x_1+\cdots+x_n)} - [e^{tx_1} - S_m(tx_1)]\cdots[e^{tx_n} - S_m(tx_n)]]e^{-nt}dt.\end{aligned}$$

Avaliando em $x_1 = x_2 = \cdots = x_n = 1$, tem-se:

$$\begin{aligned}E_m(n) &= n\int_0^\infty [e^{nt} - (e^t - S_m(t))\cdots(e^t - S_m(t))]e^{-nt}dt\\ &= n\int_0^\infty [1 - (e^t - S_m(t))^n(e^{-t})^n]dt\\ &= n\int_0^\infty 1 - [1 - e^{-t}S_m(t)]^n dt.\end{aligned}$$

□

Essa é a solução para o problema do Colecionador de Cupons com trocas: a quantidade esperada de figurinhas compradas para completar m álbuns de n figurinhas é dada por

$$E_m(n) = n\int_0^\infty \left[1 - \left(1 - \frac{1 + t + \dfrac{t^2}{2!} + \cdots + \dfrac{t^{m-1}}{(m-1)!}}{e^t}\right)^n\right] dt.$$

A integral presente no teorema necessita de métodos computacionais para sua resolução, assunto que será abordado no Capítulo 2.

2 RESULTADOS NUMÉRICOS

Nesse capítulo, a partir das soluções encontradas no Capítulo 1 e com o auxílio de recursos computacionais, serão obtidos diversos resultados numéricos acerca da coleção de figurinhas do álbum oficial da Copa do Mundo de 2018.

2.1 VALOR ESPERADO SEM TROCAS

Conforme abordado no primeiro capítulo, a quantidade esperada de unidades a serem adquiridas aleatoriamente até completar uma coleção de n elementos é dada por:

$$E = \sum_{i=0}^{n-1} \frac{n}{n-i}. \tag{2.1}$$

Onde a mesma expressão pode ser escrita como:

$$E = \sum_{i=1}^{n} \frac{n}{i}. \tag{2.2}$$

No álbum oficial da Copa do Mundo de 2018, do grupo Panini, existem 682 figurinhas a serem colecionadas. Dessa forma, a quantidade média ou esperada de figurinhas compradas para completar o álbum sem trocas é dada por:

$$E = \sum_{i=1}^{682} \frac{682}{i} \approx 4844.$$

Para resolver esse somatório, assim como os demais a seguir, foi utilizado o software WolframAlpha (WOLFRAM, 2009). Para calcular somatórios nesse ambiente o comando tem o formato $sum\ f(x),\ x = a\ to\ b$, onde $f(x)$ é a expressão em x a qual deseja-se somar para $x = a$ até $x = b$. Nesse caso específico, o comando utilizado foi

$$sum\ 682/i,\ i = 1\ to\ 682.$$

Do total de figurinhas, é interessante observar que são necessárias menos de 10% para chegar na metade do álbum, ou seja, colar 341 figurinhas. Por (2.1):

$$\sum_{i=0}^{340} \frac{682}{682-i} \approx 472.$$

Significa que mais de 90% das figurinhas são necessárias para completar a segunda parte do álbum, algo que apesar de contraintuitivo faz sentido, pois quanto mais próximo do final da coleção menor a probabilidade de conseguir uma figurinha inédita, com isso é preciso comprar cada vez mais. Para se ter uma ideia, quase duas mil figurinhas são necessárias, em média, para obter apenas as 10 últimas da coleção:

$$\sum_{i=672}^{681} \frac{682}{682-i} \approx 1998.$$

Para facilitar a tarefa do colecionador, a Panini (responsável pela produção dos álbuns e figurinhas) oferece a possibilidade de encomendar as últimas 40 figurinhas faltantes. Utilizando o serviço, é necessário coletar 642 figurinhas das 682, com isso a quantidade média de figurinhas a serem adquiridas de maneira aleatória reduz consideravelmente. Por (2.1):

$$\sum_{i=0}^{641} \frac{682}{682-i} \approx 1926.$$

As figurinhas são vendidas em pacotes com cinco unidades cada, por R$ 2,00 o pacote no Brasil, onde o comprador não deve encontrar duplicatas em um pacote. Porém, no presente trabalho, por questão de simplicidade, adota-se a hipótese de que as figurinhas possam ser compradas aleatoriamente de forma unitária, por R$ 0,40 cada uma. Conclui-se que o custo para completar o álbum sem trocas é, em média, aproximadamente $4844 \cdot 0{,}40 = 1937{,}60$ reais, sem encomendar da Panini as últimas 40 figurinhas faltantes. Aproveitando-se do benefício, o custo é, em média, $1926 \cdot 0{,}40 = 770{,}40$ reais, mais as despesas de manuseio e postagem das 40 figurinhas. Por esse serviço, a Panini cobra $25{,}00$ reais, logo, o preço total é de $770{,}40 + 25{,}00 = 795{,}40$ reais.

2.2 VALOR ESPERADO COM TROCAS

A expressão $E_m(n)$ obtida no primeiro capítulo fornece a quantidade esperada de figurinhas para completar m álbuns de n figurinhas. Dividindo esse valor pelo número de álbuns tem-se a quantidade esperada de figurinhas por álbum, em média. Assim, supondo que um grupo de amigos deseja completar m álbuns trocando as figurinhas repetidas entre si, a quantidade média de figurinhas que cada amigo precisa para completar seu álbum é:

$$\frac{E_m(n)}{m} = \frac{n}{m} \int_0^{\infty} 1 - \left[1 - e^{-t} S_m(t)\right]^n \, dt.$$

Buscando resolver $E_m(n)/m$ em geral com a maior precisão possível, será criada uma função no software livre Octave (EATON et al., 2018). Para criar funções no Octave, o comando é da forma:

```
function resultado = nome da função(argumentos)
desenvolvimento
endfunction
```

Os argumentos são uma lista de valores separados por vírgula utilizados no desenvolvimento da função, e o resultado é o valor que a função criada retorna. Por exemplo, uma função que calcula a média aritmética simples entre dois números pode ser criada da seguinte maneira:

```
function m = media(a,b)
valor=(a+b)/2;
endfunction
```

Os argumentos de entrada são os números os quais se deseja calcular a média, representados por a e b. O desenvolvimento é a regra que calcula a média entre a e b, ou seja, a metade da soma dos valores. Essa regra é atribuída a m pois a função deve retornar a média. Para calcular a média entre 3 e 7 por exemplo, basta executar o comando $x = media(3, 7)$ após criada a função que ela retornará $x = 5$.

A expressão que fornece $E_m(n)/m$ depende da quantidade de álbuns/colecionadores e do número de figurinhas. Primeiro é criada uma função relacionando $E_m(n)$ com

$$n \int_0^\infty 1 - [1 - e^{-t} S_m(t)]^n \, dt,$$

que será representado a princípio por z. Para escrever o integrando, define-se outra função, agora $y = f(t)$. Em seguida, a representação de $S_m(t)$ necessita de uma variável k que percorra os naturais de 0 a $m - 1$ para gerar o somatório

$$\sum_{k=0}^{m-1} \frac{t^k}{k!}.$$

O somatório é chamado de S e iniciado em zero. Define-se $S = S + t^k/k!$. Na primeira etapa, onde $k = 0$, $S = 0 + 1$. Depois, $k = 1$ e $S = 1 + t$, em seguida, $k = 2$ onde $S = (1 + t) + t^2/2!$, e assim segue sucessivamente até $k = m - 1$ gerando $S = S_m(t)$. Assim, é possível escrever $y = f(t)$ definida anteriormente: $y = 1 - (1 - e^{-t}S)^n$. Finalmente, aplica-se a integração numérica em f e o resultado é multiplicado por n/m gerando $z = E(m, n)$ e concluindo a aplicação. A integração numérica, que é uma aproximação do valor da integral, é realizada no Octave por meio do comando $quad$, que baseia-se em quadratura gaussiana. Detalhes sobre integração numérica via quadratura gaussiana podem ser vistos em (BURDEN; FAIRES, 2008). O código da implementação realizada no software é o seguinte:

```
function z=E(m,n)
function y=f(t)
S=0;
for k=0:m-1
    S=S+t^k/factorial(k);
end
```

```
y=1-(1-exp(-t)*S)^n;
endfunction
z=n*quad("f",0,inf)/m;
endfunction
```

A quantidade esperada de figurinhas por colecionador para completar o álbum da Copa do Mundo é determinada pela função criada com a entrada $E(m, 682)$, sendo m o número de amigos que trocarão as figurinhas repetidas entre si. A Tabela 2 apresenta a quantidade esperada de figurinhas e o custo por pessoa em algumas situações:

Tabela 2 – Quantidade esperada de figurinhas por colecionador

Álbuns/colecionadores	Figurinhas	Custo por pessoa
1 (sozinho)	4844	R$ 1937,60
2	3219	R$ 1287,60
5	2050	R$ 820,00
10	1563	R$ 625,20
20	1262	R$ 504,80
50	1025	R$ 410,00

A Figura 1 mostra o gráfico da quantidade esperada de figurinhas por colecionador para 1 até 50 colecionadores:

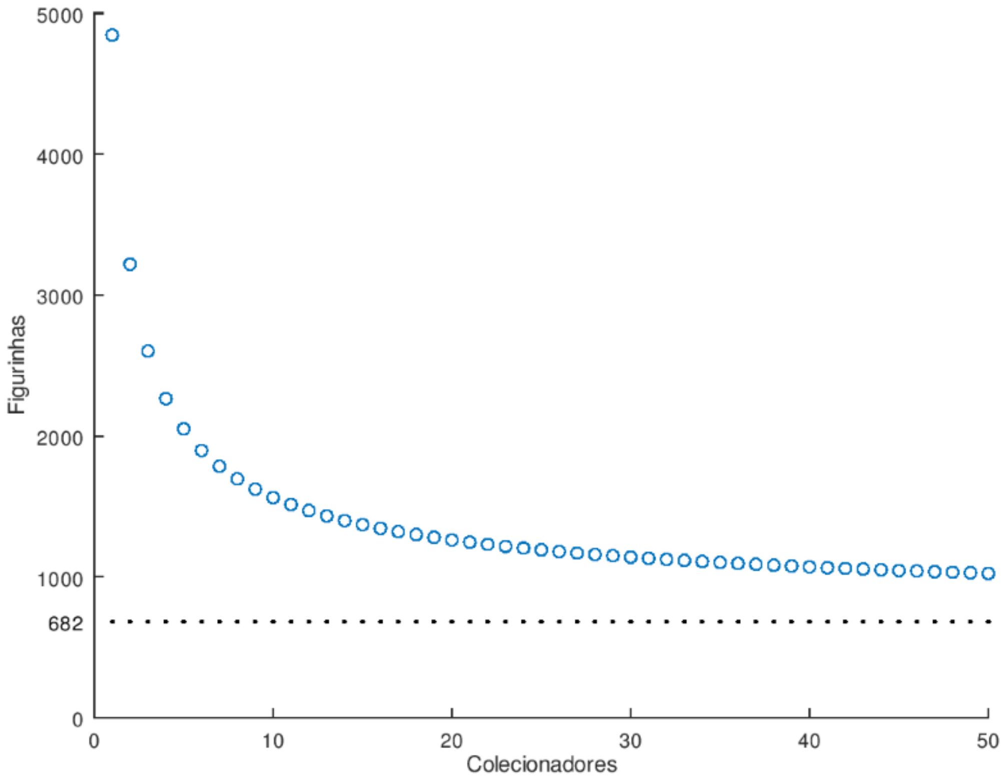

Figura 1 – Quantidade esperada por colecionador em comparação com 682

Por fim, a Tabela 3 compara diferentes maneiras para se completar um álbum:

Tabela 3 – Custo para completar um álbum em diferentes situações

Situação	Figurinhas	Custo
Sozinho	4844	R$ 1937,60
Com o serviço da Panini	1926	R$ 795,40
Trocando com 1 amigo	3219	R$ 1287,60
Trocando com 9 amigos	1563	R$ 625,20

2.3 SIMULAÇÕES

Visando comparar os resultados teóricos com uma situação mais próxima da realidade, será implementado no Octave um algoritmo que simula o processo de coleção das figurinhas. O algoritmo deve gerar aleatoriamente números inteiros maiores ou iguais a 1 e menores ou iguais a n até que sejam completados os m conjuntos desejados, computando a quantidade necessária de números gerados para tal.

Para isso, são gerados e armazenados em uma matriz linha $20 \cdot m \cdot n$ números inteiros aleatórios correspondentes às figurinhas, por meio do comando $1 + (floor(rand(1, 20 * m * n) * n))$. O comando $rand(a, b)$ gera uma matriz com a linhas e b colunas, cujos elementos são números aleatórios entre 0 e 1. Multiplicando $rand(a, b)$ por um inteiro n, os números gerados aleatoriamente passam a estar entre 0 e n, e fazendo $floor((rand(1, b) * n))$ é gerada uma matriz linha com b números inteiros entre 0 e n, uma vez que $floor(x)$ realiza o arredondamento de x para o maior inteiro menor que x. Nesse caso, o menor número gerado é 0 e o maior $(n - 1)$, porém, pode-se somar uma unidade ficando entre 1 e n, incluindo estes.

Na melhor das hipóteses, o número mínimo de figurinhas necessárias para completar m álbuns de n figurinhas é $m \cdot n$, portanto a análise inicia-se para os primeiros k-ésimos elementos da matriz, com $k = m \cdot n$, sendo k o contador de figurinhas utilizadas. Para cada figurinha, ou seja, para cada f inteiro de 1 a 682, a quantidade de elementos que vale f desde o primeiro até o k-ésimo deve ser maior ou igual a m. Caso seja menor que m, é necessário acrescentar mais uma figurinha para as já coletadas, ou seja, $k = k + 1$, onde o ciclo se repete até que seja alcançado o objetivo. Ao final do ciclo, o valor de k indica a quantidade de figurinhas que foram necessárias no total e, dividindo esse valor por m, tem-se a quantidade média de figurinhas por álbum/colecionador.

Para realizar a simulação diversas vezes e obter uma média dos resultados, a implementação no Octave foi realizada com o código a seguir:

```
function z=simula(m,n,s)
r=zeros(1:s);
for i=1:s
```

```
    t=1+(floor(rand(1,20*m*n)*n));
    k=n*m;
    for f=1:n
        while sum(t(1:k)==f)<m
        k=k+1;
        end
    end
r(i)=k/m;
end
z=mean(r);
endfunction
```

Os resultados obtidos com 100 simulações estão na Tabela 4.

Tabela 4 – Quantidade esperada x Média das simulações

Álbuns/colecionadores	Quantidade esperada	Média das simulações	Variação
1 (sozinho)	4844	4940	1,94%
2	3219	3217	0,06%
5	2050	2075	1,20%
10	1563	1556	0,45%
20	1262	1265	0,24%
50	1025	1023	0,19%

Sobre o custo computacional, o maior caso testado, isto é, $simula(50, 682, 100)$ demorou aproximadamente 5 minutos em um computador com processador de 1.8 GHz e $6,00$ GB de memória RAM.

Na prática, existe um cenário ainda melhor: o colecionador pode trocar as figurinhas repetidas e ainda aproveitar-se da possibilidade de comprar as últimas 40 faltantes com o serviço da Panini. A próxima simulação considera essa situação, computando o número necessário de figurinhas até que o total de faltantes em todos os álbuns corresponda a quantidade que pode ser comprada diretamente com a fabricante.

Inicialmente, funciona como na simulação anterior: são gerados $20 \cdot m \cdot n$ números inteiros aleatórios maiores ou iguais a 1 e menores ou iguais a n, e a análise inicia-se com o contador de figurinhas compradas k valendo $(n - 40) \cdot m$, pois na melhor das hipóteses esse é o número mínimo de figurinhas necessárias. A contagem de quantas figurinhas faltam de cada uma para completar todos os álbuns é realizada com a diferença de duas matrizes linha com n colunas, cada coluna indicando a quantidade de determinada figurinha. Na primeira matriz tem-se a quantidade necessária de cada figurinha e na segunda estão as quantidades já obtidas, de modo que a diferença dessas matrizes retorna a quantidade de faltantes da primeira até a n-ésima figurinha. Portanto, todos os elementos da primeira matriz valem m, isso é obtido fazendo-se

$m * ones(1, n)$. Na segunda matriz, por meio do comando $histc(t(1 : k), 1 : n)$ os elementos são tais que o n-ésimo indica quantas vezes o número n apareceu entre os k primeiros sorteados. É necessário que o valor mínimo da diferença entre as matrizes seja zero para qualquer elemento, pois mesmo que sejam obtidas mais de m unidades de certa figurinha a quantidade de faltantes é zero. Assim, é preciso utilizar o comando $max(m * ones(1, n) - histc(t(1 : k), 1 : n), 0)$, que compara cada elemento com o zero retornando aquele que é maior ou igual. Finalmente, com $sum(max(m * ones(1, n) - histc(t(1 : k), 1 : n), 0))$ faz-se a soma dos elementos da matriz diferença que é a soma das figurinhas faltantes. Enquanto a soma for maior que $40 \cdot m$ tem-se $k = k + 1$, o que equivale a adicionar mais um número para a análise ou comprar uma nova figurinha. A implementação determina a média de figurinhas compradas por colecionador para obter 642 figurinhas diferentes, ou seja, antes de utilizar o serviço da Panini.

```
function z=panini(m,n,s)
r=zeros(1,s);
for i=1:s
    t=1+(floor(rand(1,20*m*n)*n));
    k=(n-40)*m;
    while sum(max(m*ones(1,n)-histc(t(1:k),1:n),0))>40*m
          k=k+1;
    end
    r(i)=k/m;
end
z=mean(r);
endfunction
```

A Tabela 5 apresenta a média dos resultados de 100 simulações, onde o custo por pessoa é o custo total, que inclui o gasto com as figurinhas compradas para obter as 642 mais o valor cobrado pela Panini pelo envio das 40 faltantes:

Tabela 5 – Simulações do custo para completar um álbum com trocas e utilizando o serviço da Panini

Álbuns/colecionadores	Figurinhas	Custo por pessoa
1 (sozinho)	1925	R$ 795,00
2	1337	R$ 559,80
5	951	R$ 405,40
10	811	R$ 349,40
20	732	R$ 317,80
50	678	R$ 296,20

Nessa simulação, o custo computacional com o caso $panini(50, 682, 100)$ demorou aproximadamente 9 minutos.

3 CONCLUSÃO

Este trabalho tem como objetivo determinar o custo médio para se completar o álbum de figurinhas oficial da Copa do Mundo de 2018, a partir da quantidade esperada de figurinhas necessárias.

Para isso foi construída a solução do Problema do Colecionador de Cupons, primeiro na versão clássica e depois na versão envolvendo as trocas, provando-se todos os resultados de Newman e Shepp. Adotando as hipóteses de que todas as figurinhas possuem a mesma probabilidade de aparecer e que elas podem ser compradas aleatoriamente uma a uma, juntamente com a implementação das fórmulas, foi possível determinar computacionalmente o número de figurinhas necessárias e consequentemente o custo médio para um colecionador completar sua coleção em diferentes cenários, seja sozinho ou trocando as figurinhas repetidas. As simulações realizadas comprovaram que a teoria funciona à medida que o experimento é repetido muitas vezes. Quanto mais simulações são realizadas, mais a média tende ao valor esperado teórico. Para fim de divulgação, os principais resultados dessa dissertação estão em um vídeo disponível em: https://youtu.be/EvO4nJ9LyjY.

Visando trabalhos futuros, é interessante olhar para o problema sob outra perspectiva, que é a probabilidade de completar uma ou mais coleções comprando um certo número de figurinhas. Mais precisamente, seria responder a pergunta: qual é a probabilidade de completar m álbuns de n figurinhas ao comprar f figurinhas? Nesse contexto, uma simulação computacional foi realizada por (TREBBIN, 2014), apresentando as probabilidades de completar $1, 2, 3, \ldots, 19$ coleções de 42 elementos, para 1 até 1000 objetos adquiridos aleatoriamente.

Sobre o PROFMAT, ele foi muito importante em minha formação acadêmica trazendo novos conhecimentos e solidificando muitos outros já existentes. A graduação, sobretudo na licenciatura, deixa muitas lacunas na formação matemática do professor, e as disciplinas desse mestrado proporcionaram um aprofundamento dos assuntos, com todo o rigor necessário, me tornando mais preparado e seguro para o exercício da docência.

REFERÊNCIAS

BASSANEZI, R. C. Problema do colecionador de figurinhas. 2018. DOI: 10.13140/RG.2.2.25234.96960.

BURDEN, R. L.; FAIRES, J. D. **Análise Numérica**. 8. ed. [S.l.]: Cengage Learning, 2008.

CARVALHO, P. C. P. Quantas figurinhas comprar para completar o álbum da copa? **Revista do Professor de Matemática**, SBM, v. 73, 2010.

EATON, J. W. et al. **GNU Octave version 4.4.0 manual: a high-level interactive language for numerical computations**. [S.l.], 2018. Disponível em: <https://www.gnu.org/software/octave/doc/v4.4.1/>. Acesso em: 15 abr. 2019.

FELLER, W. **An introduction to probability theory and its applications**. [S.l.]: New York, 1950. v. 1.

FERRANTE, M.; SALTALAMACCHIA, M. The coupon collector's problem. **MATerials MATemàtics**, 2014.

FLAJOLET, P.; GARDY, D.; THIMONIER, L. Birthday paradox, coupon collectors, caching algorithms and self-organizing search. **Discrete Applied Mathematics**, n. 39, p. 207–229, 1992.

KOBZA, J. E.; JACOBSON, S. H.; VAUGHAN, D. E. A survey of the coupon collector's problem with random sample sizes. **Methodology And Computing In Applied Probability**, v. 9, p. 573–584, 2007.

MAY, R. Coupon collecting with quotas. **The Electronic Journal of Combinatorics**, v. 15, n. 31, 2008.

NEWMAN, D. J.; SHEPP, L. The double dixie cup problem. **The American Mathematical Monthly**, Mathematical Association of America, v. 67, n. 1, p. 58–61, 1960.

O'CONNOR, L. Entropy bounds for traffic confirmation. **Cryptology ePrint Archive**, 2008.

PAPANICOLAOU, V. G.; KOKOLAKIS, G. E.; BONEH, S. Asymptotics for the random coupon collector problem. **Journal of Computational and Applied Mathematic**, n. 93, p. 95–105, 1998.

POON, A.; DAVIS, B. H.; CHAO, L. The coupon collector and the suppressor mutation: Estimating the number of compensatory mutations by maximum likelihood. **Genetics**, v. 170, p. 1323–1332, 2005.

RAFFERTY, A. N.; GRIFFITHS, T. L.; KLEIN, D. Convergence bounds for language evolution by iterated learning. **Proceed. 31st Annual Conference of the Cognitive Science Society**, 2009.

SASAKI, Y. et al. Coupon collector's problem for fault analysis against aes — high tolerance for noisy fault injections. **Financial Cryptography and Data Security**, 2013.

SCHELLING, H. V. Coupon collecting for unequal probabilities. **The American Mathematical Monthly**, v. 61, n. 5, p. 306–311, 1954.

SHANK, N.; YANG, H. Coupon collector problem for non-uniform coupons and random quotas. **The electronic journal of combinatorics**, v. 20, n. 33, 2013.

STADJE, W. The collector's problem with group drawings. **Advances in Applied Probability**, v. 22, p. 866–882, 1990.

STEWART, J. **Cálculo**. 7. ed. [S.l.]: Cengage Learning, 2013. v. 2.

TREBBIN, G. **Probability of Collecting Multiple Full Sets**. [S.l.], 2014. Disponível em: <https://www.grant-trebbin.com/2014/03/probability-of-collecting-multiple-full.html>. Acesso em: 04 jun. 2019.

WOLFRAM, S. **Wolfram|Alpha**. [S.l.], 2009. Disponível em: <https://www.wolframalpha.com/>. Acesso em: 15 abr. 2019.