HOLOS
ISSN: 1518-1634
holos@ifrn.edu.br
Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Norte
Brasil

SOUZA, G. F.
CONSIDERAÇÕES SOBRE A ALEATORIEDADE DOS CONCURSOS DA MEGA SENA
HOLOS, vol. 8, 2017, pp. 65-75
Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Norte
Natal, Brasil

CONSIDERAÇÕES SOBRE A ALEATORIEDADE DOS CONCURSOS DA MEGA SENA

G. F. SOUZA
Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande de Norte
gustavo.fontoura@ifrn.edu.br

Submetido 02/10/2017 - Aceito 26/12/2017

DOI: 10.15628/holos.2017.5647

RESUMO

As Loterias tem atraído uma grande qualidade de pessoas com esperanças de fortuna. Os jogos são relativamente simples e, provavelmente em função dos prêmios milionários tem atraído o interesse de muitas pessoas. Os jogos dessa natureza, como a Mega Sena, no Brasil, são bons exemplos de distribuição de probabilidade num espaço equiprovável e permitem o entendimento de diversos aspectos de teorias estatísticas. Neste trabalho, é proposto uma pequena análise do jogo Mega Sena, bem como análise dos resultados dos concursos já realizados. Um teste de hipótese é utilizado para verificar a característica equiprovável do jogo.

PALAVRAS-CHAVE: Mega Sena, distribuição uniforme, teste de aderência.

CONSIDERATIONS ABOUT RANDOMIZATION OF MEGA SENA COMPETITIONS

ABSTRACT

The Lottery has attracted a lot of people with Fortune hopes. The games are relatively simple and, probably because of the millionaires awards has attracted the interest of many people. Games such as the Mega Sena in Brazil are good examples of probability distribution in an equiprobable space and allow the understanding of various aspects of statistical theories. In this work, a little analysis of the Mega Sena game is proposed, as well as analysis of the results of competitions already made. A hypothesis test is used to verify the equiprobable character of the game.

KEYWORDS: Mega Sena, uniform distribution, Goodness of fit test.

HOLOS
ISSN 1807 - 1600

1 INTRODUÇÃO

As loterias têm atraído uma grande qualidade de pessoas com esperanças de fortuna. Os jogos de loteria são relativamente simples e, provavelmente em função dos prêmios milionários têm atraído o interesse de muitas pessoas. Os jogos baseados em "bingo", como a Mega Sena que consiste em acertar seis números dentro de um conjunto de números possíveis, de um a sessenta.

Jogos dessa natureza são exemplos de distribuição de probabilidade uniforme, ou seja, cada número do conjunto apresenta a mesma probabilidade de ser sorteado. A distribuição uniforme é extremante importante para a teoria das probabilidades (Meyer, 1983).

A Mega Sena é um jogo de apostas lotéricas do Brasil. Talvez a mais famosa modalidade dentre as atuais loterias da Caixa Econômica Federal. Os sorteios, também chamados de concursos, acontecem duas vezes por semana e já é realizado desde 1996. Já foram realizados 1990 concursos (até 22/11/2017) (Caixa Econômica Federal, 2017). Os Prêmios pagos para o acertador da Mega Sena, geralmente, são altos e pode-se ganhar também acertando 5 números (quina) e 4 números (quadra).

Neste trabalho será investigada a aleatoriedade dos sorteios da Mega Sena utilizando os resultados já obtidos nos 1990 concursos realizados até o dia 22/11/2017. Além disso, o artigo propõe entender as regras e o processo da Mega Sena. Será utilizando para tanto simulações computacionais realizadas com o software R (R Development Core Team,2017), que é um software livre desenvolvido com vistas a realizações de aplicações em estatísticas.

O artigo está organizado em tópicos de modo que cada um deles aborda uma temática diferente sobre o jogo da Mega Sena. A seção 2 apresenta um detalhamento das condições na qual o jogo se apresenta, como as probabilidade de acertar. Já na seção 3 é apresentado o formalismo matemático a respeito do modelo de probabilidade do jogo da Mega Sena. Na Seção 4 é apresentado os dados dos sorteios anteriores e realizados testes a fim de verificar a aderência dos dados ao modelo estatístico proposto, bem como uma simulação afim de verificar o que acontece, assintoticamente, com a ampliação do número de sorteios. Por fim, a seção 6 apresenta as considerações finais.

2 ANÁLISE DO JOGO DA MEGA SENA

Características do Jogo

Em cada concurso da Maga Sena, os jogadores devem escolher 6 números entre um grupo de 60 possíveis (1,2,...,59,60). A probabilidade de uma pessoa que realizou uma aposta simples (de 6 números) é um cálculo simples e o resultado está disponível no site da Caixa Econômica Federal (http://loterias.caixa.gov.br/wps/portal/loterias/landing/megasena). A Equação (1) apresenta a formula que a calcula e o seu resultado.

$$P(X=x)=\frac{1}{C_{60,6}}=\frac{1}{50.063.860}\approx 2.10^{-8}=0{,}000002\% \tag{1}$$

Na qual $C_{60,6}$ indica a quantidade de combinações que podem ser feitas com 60 números agrupados 6 a 6.

Percebe-se que, como qualquer jogo de azar (ou sorte?), é extremamente improvável que uma pessoa ganhe com uma aposta, pois, a probabilidade de não ganhar é 99,999998 %. No jogo na Mega Sena é possível, entretanto, ganhar acertando apenas 5 ou 4 números (quina e quadra), o que torna o jogo mais interessante para o apostador, contudo ainda pouco provável. A Tabela 1 apresenta as probabilidade de ganhar acertando 6, 5 ou 4 números sorteados, com apenas uma aposta de 6 números.

Tabela 1: Probabilidades de acertar 6, 5 ou 4 números com uma aposta de 6 números.

Número de acertos	Probabilidade acertar	Probabilidade de não acertar
6	0,000002 %	99,999998 %.
5	0,00065 %	99,99935 %
4	0,043 %	99.95712 %

Fonte: Autoria Própria

Em Rodrigues (2004) é citada duas comparações interessantes para que os leitores percebam, de fato a dimensão dessas probabilidades, comparando com eventos conhecidos. Uma delas afirma que é mais provável, ao jogar uma moeda 25 vezes obter 25 caras do que acertar os 6 números da Mega Sena com uma aposta de 6 números. A comprovação dessa comparação é simples. Jogando-se uma moeda "honesta", pode-se obter dois resultados equiprováveis (cara ou coroa), ou seja, a probabilidade de sair uma cara (ou coroa) é de 0,5. Se repetirmos o experimento n vezes, como os lançamentos são independentes, teremos probabilidade igual a $0{,}5^n$. Assim, para obtermos o resultado apontado por Rodrigues (2004) basta igualar as duas probabilidade e identificar o valor de n, conforme a Equação (2).

$$\frac{1}{2^n} = \frac{1}{50.063.860} \qquad (2)$$

O resultado da equação (2) é, aproximadamente, n = 25,577. Assim a quantidade (1/50.063.860) é menor que a probabilidade de obtermos 25 caras em 25 lançamentos de uma moeda. Em uma simulação realizada por computador, utilizando-se o software R, na qual lança-se uma moeda 25 vezes e observamos o número de caras, só obteve-se 25 caras após 2.936.819 repetições. O que mostra quão improvável é o acerto da Mega Sena.

Uma abordagem interessante é feito em FREITAS (2013), na qual é calculada a probabilidade de acerto dos seis números, caso fosse possível marcar qualquer quantidade de números. A Figura 1 mostra a relação entre a quantidade de números apostados e a probabilidade de acertar os 6 números sorteados. Nessa Figura, pode-se observar que para se ter probabilidade de 20% de acertar os seis números é necessários escolher 47 números, ou seja, 78% dos números disponíveis.

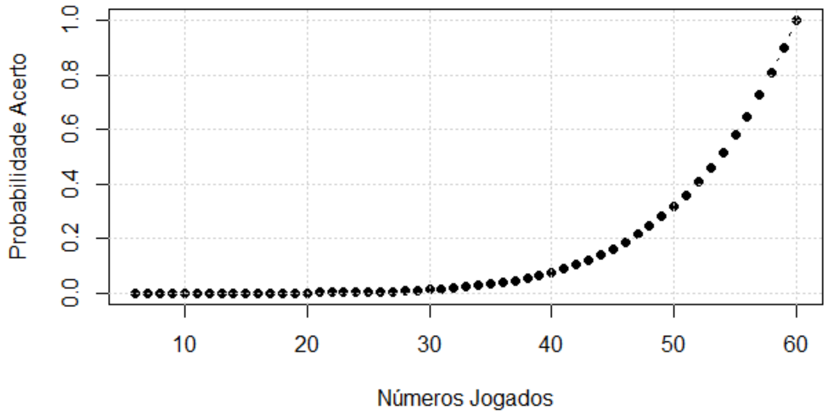

Figura 1: Probabilidade de acerto dos seis números em função da quantidade de números escolhidos.

Fonte: Elaboração Própria.

Em Francisco (2012), o autor amplia os cálculos para muitos concursos e concluí o que os números da Tabela 1 já sugerem: que é improvável ganhar na Mega Sena. Neste artigo, concluí-se que se durante 50 anos, apostássemos 100 jogos simples (de 6 números) por dia de aposta (duas vezes por semana), nossa chance de acertar ao menos uma vez seria de pouco mais de 1%, isto é teríamos uma probabilidade de quase 99% de não ganhar nada nesses 50 anos.

Múltiplas apostas

É possível também apostar com 7, 8 ou mais números num mesmo cartão, contudo os valores pagos por essas apostas são maiores e as probabilidades de acertar um pouco maior. Essa é uma outra curiosidade a respeito da Mega Sena, se um apostador quiser apostar 7 números fazendo apostas simples (de 6 números) pagará mais ou menos que se fizer uma única aposta de 7 números?, isto é, é mais barato jogar um único cartão com 7 números ou todas as combinações dos 7 números com apostas simples?

O calculo, nesse caso, também é simples, pois se foi escolhido 7 números e podemos jogar apenas 6, teremos a combinação de 7, 6 a 6, cujo resultado é 7. Assim, para se apostar 7 números em apostas simples precisa-se apostar 7 cartões (apostas), cujo valor é R$ 24,50 (7 x R$ 3,50), logo não há diferença entre o preço de 7 apostas de 6 números ou de uma aposta com 7 números. A Tabela 2 mostra valores das apostas e combinações necessárias em cartões simples.

Tabela 2: Valor cobrado de acordo com número de apostas

Quantidade de números (n)	Quantidade de combinações (p) (apostas simples)	Valor cobrado por um jogo com n números (R$)	Valor cobrado por p apostas (R$)	Probabilidade de não acertar
6	1	3,50	3,50	99,999998 %
7	7	24,50	24,50	99,999997 %
8	28	98,00	98,00	99,99994 %
9	84	294,00	294,00	99,99983 %
10	210	735,00	735,00	99,99958 %
11	462	1.617,00	1.617,00	99,99908 %
12	924	3.234,00	3.234,00	99,99815 %
13	1716	6.006,00	6.006,00	99,99657 %
14	3003	10.510,50	10.510,50	99,99400 %
15	5005	17.517,50	17.517,50	99,99000 %

Fonte: (Caixa Econômica Federal2017)

Apesar de facilitar o processo para pessoas que queiram apostar mais de 6 números, ainda é improvável que se venha a ganhar. Uma aposta de 15 números, por exemplo, que custa R$ 17.517,50, tem uma probabilidade de 99,99 % de não acertar os 6 números.

3 MODELO ESTATÍSTICO

O concurso da Mega Sena consiste em um modelo de distribuição denominado uniforme discreto, com parâmetros (1,2...,59,60). A função de probabilidade desse modelo é apresentado na Equação (3). O gráfico desta função é apresentado na Figura 2. Observa-se que para o modelo discreto, somente os valores 1,2...,59,60 apresentam valores de probabilidade.

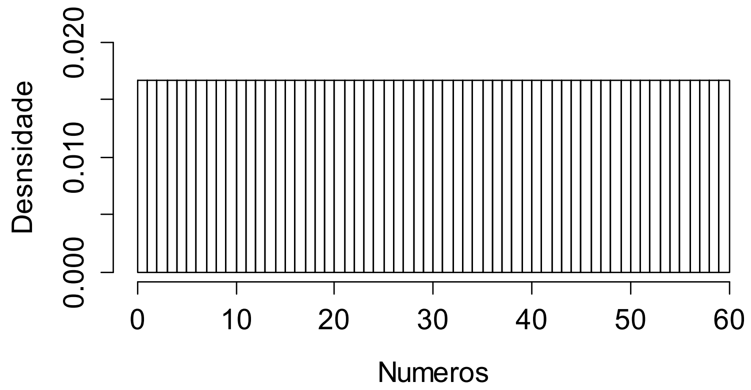

Figura 2: Gráfico da função de probabilidade da distribuição uniforme discreta

Fonte: Elaboração Própria.

A principal característica desde modelo de distribuição de probabilidade é que cada resultado possível tem a mesma probabilidade de acontecer, neste caso 1/60.

$$P(X = x) = \frac{1}{60} I_{[1,2,..59,60]}(x) \tag{3}$$

Na qual $I_{[1,2,...,59,60]}$ (x) é a variável indicadora, cujo valor é igual a 1 se x for igual a um dos valores possíveis e zero, caso contrário. A Esperança e a Variância da distribuição uniforme discreta, podem ser calculadas, conforme as definições de Esperança e Variância, obtendo os valores indicados nas equações (4) e (5) (Bussab e Moretin, 2010).

$$E(X) = \sum_{i=1}^{60} x_i.P(X = x_i) = 30{,}5 \tag{4}$$

$$V(X) = E(X^2) - E^2(X) = 1230{,}167 - 30{,}5^2 = 299{,}92 \tag{5}$$

O Valor Esperado ou valor médio de uma variável aleatória representa o valor médio “esperado” de uma experiência se ela for repetida muitas vezes (Magalhães, 2006). Já a variância indica como os dados estão distribuídos em torno do valor esperado.

4 ANÁLISE DOS DADOS

Análise de ocorrência simples

No site da Caixa Econômica Federal, encontram-se disponíveis para download os resultados dos concursos anteriores. Foi utilizando neste artigo os resultados dos concursos de 1 a 1990. A Figura 3 apresenta o histograma da distribuição de probabilidade amostral desses dados. A média é 30,485 e a variância é 300,4199. Os valores que mais foram sorteados são 5 e

53 com 230 e 226 vezes, respectivamente, e os menos sorteados são 26 e 55 com 163 e 172 concursos. Ao contrário do que se pode pensar a diferença entre a quantidade de vezes que o número 5 foi sorteado e a quantidade de vezes que o 26 foi sorteado, não é significativa do ponto de vista estatístico, pois se a distribuição é uniforme os dados experimentais (concursos, neste caso) dificilmente apresentarão rigorosamente a mesma quantidade de sorteios para todos os valores.

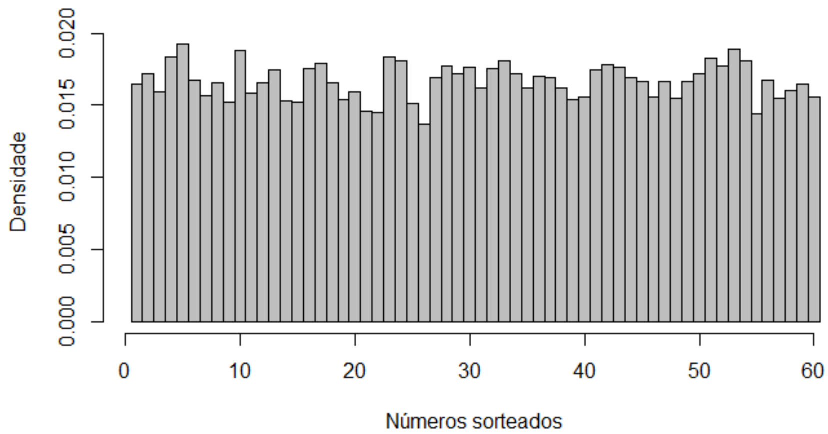

Figura 3: Histograma dos dados já sorteados.

Fonte: Elaboração Própria.

Uma simulação para 50 mil concursos da Mega Sena foi realizado e os dados obtidos são apresentados no histograma da Figura 4. Observa-se que o número 30 foi o mais sorteado com 5165 sorteios, contudo a diferença percentual entre ele e o 13, número menos vezes sorteado (com 4853 sorteios), que foi o menos sorteado, foi de 6,04 %. Assim, quanto maior o número de concursos menos significativo é a diferença entre o mais sorteado e o menos sorteado. A Figura 5 (apresentada na seção seguinte) mostra o gráfico da diferença percentual entre o valor mais sorteado e o menos sorteado para dados simulados de 10 mil a 10 milhões de concursos.

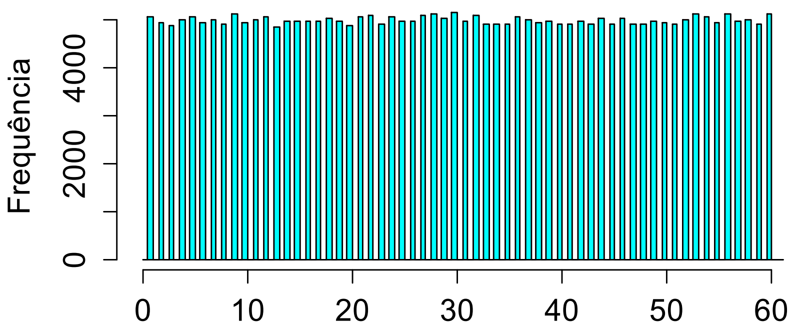

Figura 4: Histograma dos dados de simulação de 50 mil concursos. Fonte: Elaboração Própria.

Observa-se que não se tem uma distribuição exatamente uniforme discreta, pois alguns valores apresentam frequência maior que outros. Com a finalidade de testar a aderência dos dados já sorteados nos concurso da Mega Sena ao modelo uniforme, aplicou-se dois testes de hipóteses não-parametricos, muitos utilizados para este fim (Siegel e Castellan,2006). O Teste de aderência qui-quadrado e o teste de kolmogorov-smirnov.

O Teste qui-quadrado apresenta uma estatística teste dada pela equação (6), já o teste de kolmogorov-smirnov é apresentado na equação (7) (Siegel e Castellan, 2006). Em ambos os teste as hipóteses a serem testadas são as mesmas:

$$\begin{cases} H_0 : \text{Os dados apresentam distribuição uniforme discreta} \\ H_A : \text{Os dados não apresentam distribuição uniforma discreta} \end{cases}$$

$$\chi^2 = \sum_{i=1}^{60} \frac{(O_i - E_i)^2}{E_i} \quad (6)$$

Na qual, O_i é o valor observado para o i-ésimo elemento e E_i é o valor esperado para o i-ésimo elemento.

$$Dn = \sup_x |F(x) - F_n(x)| \quad (7)$$

Onde, F(x) é a função de distribuição acumulada assumida para os dados e $F_n(x)$ é a função de distribuição acumulada empírica dos dados. Sup é o supremo.

Tabela 3 - Testes estatísticos não-parametricos realizados.

Teste	Qui-Quadrado	Kolmogorov-Smirnov
Estatística teste	62,121	0,0167
Graus de liberdade	59	-
Quantil crítico	77,931	0,1756
Alfa	5%	5%
p-valor	0,3656	≈ 1
Conclusão do teste	Não Rejeitar Ho	Não Rejeitar Ho

Fonte: Elaboração Própria.

Logo, de acordo com os testes apresentados na Tabela 3, não há evidências estatísticas de que os dados não sigam uma distribuição de probabilidade uniforme. Esse resultado é muito importante, pois corrobora com a premissa de que os concursos não apresentam vícios. É importante entender os resultados dos testes de aderência, eles indicam que não pode-se rejeitar a hipótese de que os dados tenham uma distribuição uniforme discreta (1,60).

HOLOS
ISSN 1807 - 1600

Análise de co-ocorrência

Outra análise que pode ser feita diz respeito a ocorrências conjunta, ou seja, os números que são sorteados juntos mais vezes. Os números 5 e 33 foram sorteados no mesmo concurso 45 vezes, já os pares (39 e 32) e (57 e 36) foram sorteados juntos apenas 6 vezes e não existe nenhuma combinação de dois números que nunca foram sorteados. A Figura 4 apresenta de forma gráfica, utilizando uma escala de tons de cinza a matriz de co-ocorrência 2 a 2 dos valores sorteados até agora. Na Figura 4, os valores mais escuros representam menor ocorrência e os valores mais claros, maior ocorrência, conforme legenda da Figura que vai de 0 a 45. A diagonal da matriz é sempre zero (preta) uma vez que o mesmo número não pode ser sorteado duas vezes no mesmo concurso.

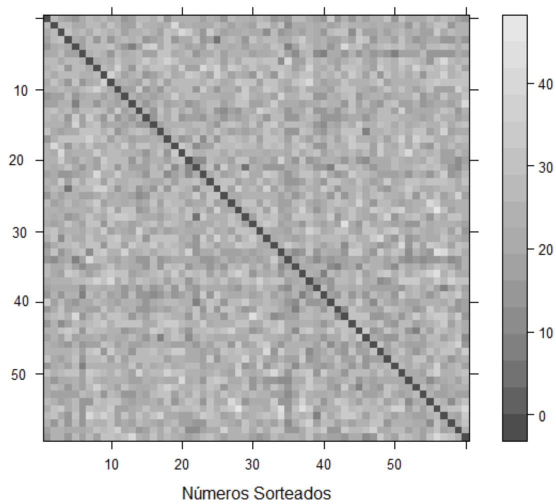

Figura 4: Matriz de co-ocorrência 2 a 2 dos valores sorteados na Mega Sena.
Fonte: Elaboração própria.

Em função de um grupo muito pequeno de sorteios já realizados não é adequado fazer qualquer teste ou tirar qualquer conclusão sobre as combinações de dados. Os números (ou pares de números) mais sorteados têm 32 ocorrências, os trios, quadras e quinas terão frequências de ocorrência ainda menores. É necessária uma quantidade muito maior de sorteios para que se tenha alguma confiança nos analises feitas. É bom salientar que, teoricamente, somente após o concurso número 50.063.860 é que teremos uma combinação de todos os resultados possíveis (se não houver repetições). Como hoje se dispõe apenas de, aproximadamente, 0,0040 % (1990) desse total (50.096.860) de concursos realizados, podemos concluir que não há resultados (dados) suficientes.

Apesar disso, pode-se realizar uma simulação dos sorteios posteriores utilizando o modelo de probabilidade uniforme discreta. Nesta simulação realizada com software estatístico R, foi obtido o resultado para 10 milhões de concurso, isto é 20% do total de combinações (50.096.860). O gráfico da Figura 5 apresenta a diferença percentual entre o número de vezes que o par de números mais frequente aconteceu e o par menos frequente. O gráfico mostra que quando se dispões de poucos concursos a diferença é maior, próximo de 15%, contudo quando o número de sorteios aumenta a diferença se aproxima de zero, que é esperado pelo modelo de probabilidade uniforme discreto. Esse resultado corrobora com a premissa de que a diferenças observadas nos dados tenderão para zero para um maior número de concursos.

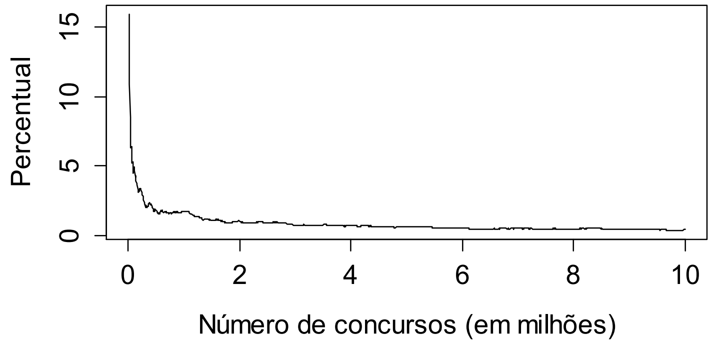

Figura 5: Diferença percentual entre o número mais sorteado e o menos sorteado.

Fonte: Elaboração própria.

5 CONSIDERAÇÕES FINAIS

É possível perceber a quantidade de informações e análises que podem ser feitas com resultados dos concursos da Mega Sena. Neste trabalho, pode-se observar a natureza aleatória dos valores sorteados, verificando-se os resultados dos testes de hipóteses. Ambos os teste realizados apontam para que os resultados dos concurso anteriores foram obtidos de uma distribuição de probabilidade uniforme discreta. As ocorrências conjuntas dão uma ideia de como o processo acontece em cada concurso. A análise descritiva dos resultados da Mega Sena melhora o entendimento do processo.

A análise das combinações que possibilitem o acerto em cada concurso em outros jogos de loteria pode ser feitas e talvez até a análise de ocorrência conjunta, já que existem outros jogos de loteria com probabilidades acerto maiores que a Mega Sena. Esse tema pode ser trabalhado futuramente como continuidade deste trabalho.

6 REFERÊNCIAS

Bussab, W. D. O., & Morettin, P. A. (2010). *Estatística básica*. Saraiva.

Caixa Econômica Federal (2017), *Informações sobre a Mega Sena*. Recuperado em 24/11/2017, de http://loterias.caixa.gov.br/wps/portal/loterias/landing/megasena.

Francisco, J. A. (2012). *Como não ganhar na Mega-Sena*. Recuperado em 12/09/2015 de http://www.sigmasociety.com/artigos/jose_antonio_francisco_megasena.pdf.

Freitas, M. A. (2013). *Aspectos históricos e teóricos das loterias*. Dissertação de Mestrado. Instituto de Matemática e Estatística - IME (RG). Universidade Federal de Goiás. Goiana, GO, Brasil.

Magalhães, M. N. (2006). *Probabilidade e variáveis aleatórias*. Edusp.

Meyer, P. L. (2011). Probabilidade: aplicações à estatística; tradução do Prof. *Ruy de CB Lourenço Filho. Rio de Janeiro, Livros Técnicos e Científicos*.

R Development Core Team (2017). *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, www.R-project.org.

Rodrigues, F. W. (2004) *A mídia e a Mega Sena acumulada*. Coleção Explorando o Ensino da Matemática. Vol. 1. Cap. 1. Brasília: MEC, Secretaria de Educação Básica.

Siegel, S., & Castellan Jr, N. J. (2006). *Estatística não-paramétrica para ciências do comportamento.* trad. *Sara Ianda Correa Carmona*, *2*. 2. Ed.