

# Relatório Probabilidade e Estatística

## **Software simulador**

Julia Rodrigues Gubolin, Gustavo Pereira Pazzini e Paulo Tavares  
Borges

22 de fevereiro de 2021

Universidade Estadual Paulista – Unesp/IBILCE. São José do Rio  
Preto - SP

## Sumário

<b>1.0 Introdução.....</b>	<b>2</b>
<b>2.0 Objetivos.....</b>	<b>2</b>
<b>3.0 Metodologia .....</b>	<b>2</b>
<b>4.0 Resultados .....</b>	<b>3</b>
<b>4.1 Experimento 1.....</b>	<b>3</b>
<b>4.2 Números aleatórios .....</b>	<b>7</b>
<b>4.3 Experimento 2.....</b>	<b>9</b>
<b>5.0 Considerações Finais .....</b>	<b>10</b>
<b>6.0 Referências.....</b>	<b>10</b>

## 1.0 Introdução

Neste relatório, a equipe irá expor todo o processo de construção e análise dos dados obtidos em cada simulação. Serão apresentados gráficos, tabelas e equações as quais permitiram a coleta de valores e a calibragem do software.

Dentro da aplicação, mais especificamente dentro da área reservada para o experimento 1, existe um exemplo pronto com uma análise completa de valores que possuem certa tendência a uma distribuição exponencial. Todo o processo de desenvolvimento, análise estatística, cálculo do teste Qui-quadrado, testagem e conclusão está descrito com detalhes neste relatório. Na área 2 do experimento 1, o usuário pode fazer estes testes manualmente. Os detalhes sobre esta parte também serão expostos aqui.

Foi reservado, também, um tópico apenas para tratar da verificação da qualidade do software quanto a geração de números aleatórios. Além disso, com os números gerados aleatoriamente é possível construir uma distribuição exponencial e uma distribuição normal.

Por fim, também foi dedicado um espaço para comentar sobre a construção do experimento 2, baseado no Teorema do Limite Central e suas aplicações.

## 2.0 Objetivo

No experimento 1, foi feita uma simulação a qual verifica se determinada variável  $X$  segue uma distribuição escolhida pelo usuário. Para isso, são calculados os valores esperados, baseado no modelo escolhido, e o teste Qui-quadrado, além de serem confeccionados 2 gráficos. Um gráfico representa os dados observados, inseridos pelo usuário, e o outro representa os dados esperados.

No experimento 2, o usuário pode conferir uma prova completa para o Teorema do Limite Central, baseada em um exemplo previamente preparado.

Baseado nisso, o objetivo primordial é construir uma aplicação didática, simples e completa para que o usuário possa confiar nos resultados.

## 3.0 Metodologia

Para a criação da aplicação, foi utilizado o ambiente de desenvolvimento da Microsoft Visual Studio 2019, com a linguagem C#.

Com relação aos gráficos e tabelas da seção “Experimento 1”, foi utilizado o editor de planilhas da Microsoft, o Excel.

Com o objetivo de verificar os valores e gráficos mostrados no software, bem como verificar as equações corretas para uso em cada modelo, utilizamos livros e exemplos de exercícios presentes nos materiais.

## 4.0 Resultados

### 4.1 Experimento 1

O experimento 1 foi dividido em duas seções: uma seção para a simulação previamente preparada e uma seção para que o usuário faça sua própria simulação. Os resultados e testes serão descritos abaixo.

Com relação a simulação preparada, a variável X de estudo é “Tempo para absorção do fármaco Clonazepam (4 mg) em horas”. Os dados reais colhidos indicam que esta variável tende a uma distribuição exponencial. Tendo isso em mente, os valores esperados foram calculados utilizando  $\alpha = 0.07709$  e a equação  $\int_{x_1}^{x_2} \alpha e^{-\alpha \cdot x} dx \cdot \sum_{i=1}^n Oi$ . O resultado de  $\sum_{i=1}^n Oi$  é 90,8.

Para utilizar o teste de Aderência Qui-Quadrado, é necessário criar 2 hipóteses  $H_0$  e  $H_1$ :

- 1)  $H_0$ : a variável X segue o modelo proposto.
- 2)  $H_1$ : a variável X não segue o modelo proposto.

Categorias	Freq. Observada (Oi)	Freq. Esperada (Ei)	Qui-quadrado	P(X=x)
[0; 6)	41,1	33,62	1,66	0,37
[6; 12)	25	21,17	0,69	0,23
[12; 18)	10	13,33	0,83	0,15
[18; 24)	7,5	8,39	0,09	0,09
[24; 30)	4,5	5,28	0,11	0,06
[30; 36)	2,1	3,33	0,45	0,04
[36; 42)	0,6	2,09	1,26	0,02
<b>TOTAL</b>	<b>90,8</b>	<b>87,21</b>	<b>4,89</b>	<b>0,96</b>

Tabela 1: Valores calculados para determinar se X segue o modelo exponencial.

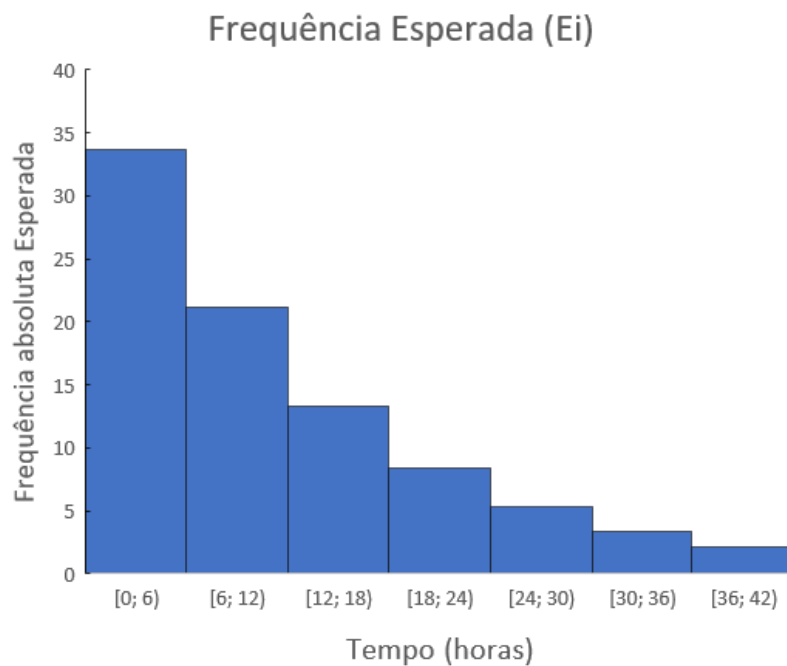


Figura 1: Gráfico da frequência esperada da variável X.

O gráfico da figura 1 é formado pelo histograma que representa os valores esperados para a variável X. Já o gráfico da figura 2 é composto por um histograma que representa os dados reais e por uma curva que apenas demonstra a tendência destes valores. Ela ajuda também a corroborar a teoria de que estes valores seguem uma tendência exponencial, mas não pode ser usada como parâmetro para confirmar esta hipótese de fato.

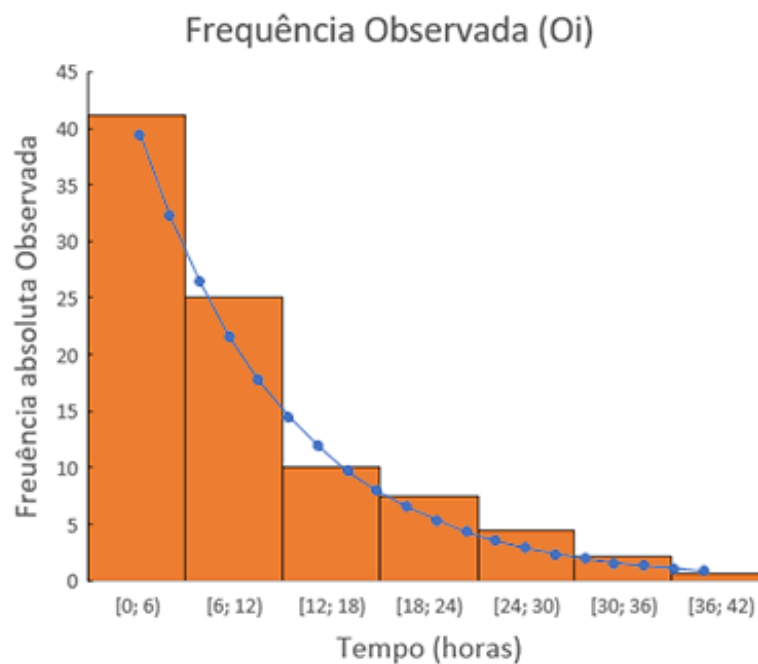


Figura 2: Gráfico dos valores observados da variável X.

Na tabela de valores Qui-quadrado, para  $\alpha = 0.05$  (5%) e graus de liberdade (gl) = 6, Qui-quadrado tem o valor de 12,592. Como teste de Aderência Qui-Quadrado calculado < teste de Aderência Qui-Quadrado tabelado, podemos concluir que a variável X segue de fato uma distribuição exponencial de parâmetro  $\alpha = 0.07709$  e que é possível, portanto, aceitar a hipótese  $H_0$ .

Na área reservada para que o utente possa fazer sua própria simulação, são disponibilizados 3 modelos possíveis: o modelo normal, o modelo exponencial e o modelo uniforme. Para corroborar o funcionamento de cada um, foram utilizados exercícios prontos e previamente solucionados.

Para o modelo exponencial, nos baseamos na simulação preparada. O software recebe como entrada os dados observados e calcula os dados esperados e o teste de Aderência Qui-Quadrado. Na figura 3, está o resultado mostrado pelo software. Ele ratifica que o modelo exponencial é um modelo adequado para representar a variável X.

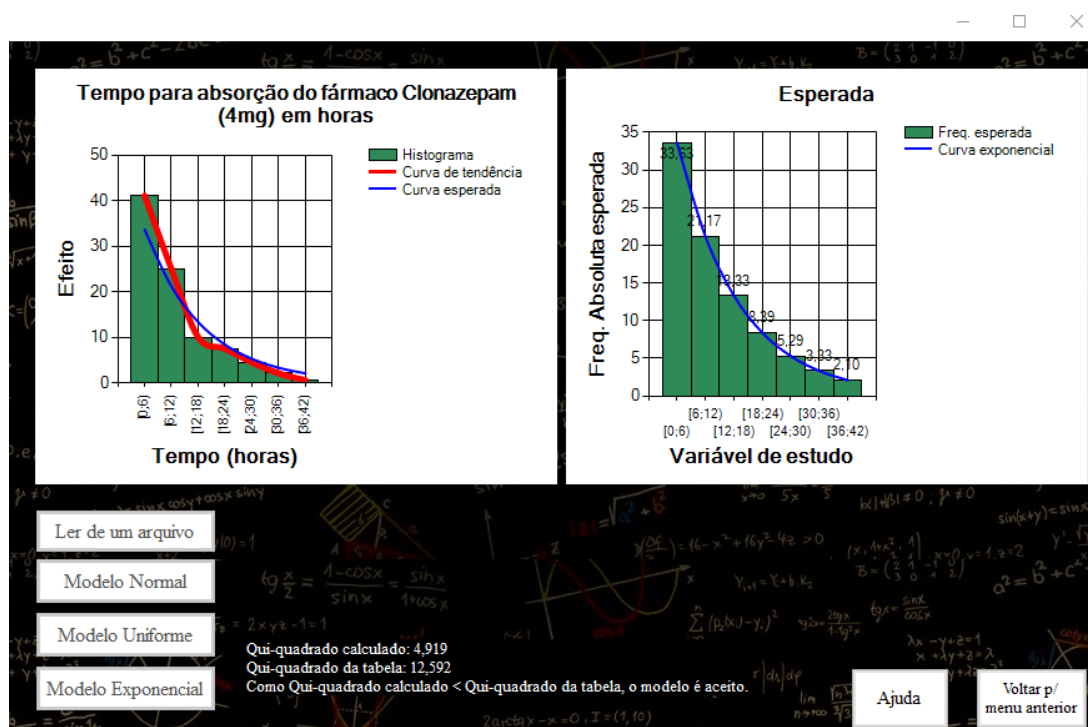


Figura 3: Resultado da variável X para um modelo exponencial com o software.

Com o intuito de corroborar o funcionamento para o modelo normal, foi utilizado o exercício: Um pesquisador está interessado em fazer um estudo sobre estaturas de recém nascidos de mães fumantes. Para isso, usou dados de hospitais da região Sul. Os recém nascidos originados no Sul possuíam estaturas distribuídas segundo a tabela a baixo. É possível afirmar que os dados das estaturas seguem uma distribuição aproximadamente normal com média de 45 cm e desvio padrão de 6 cm?

Estatura	Frequência observada
[30;35)	5
[35;40)	18
[40;45)	33
[45;50)	26
[50;55)	16
[55;60)	2

Tabela 2: Tabela referência para o exercício acima.

A variável de estudo X do problema é: “Estatura de recém nascidos”.

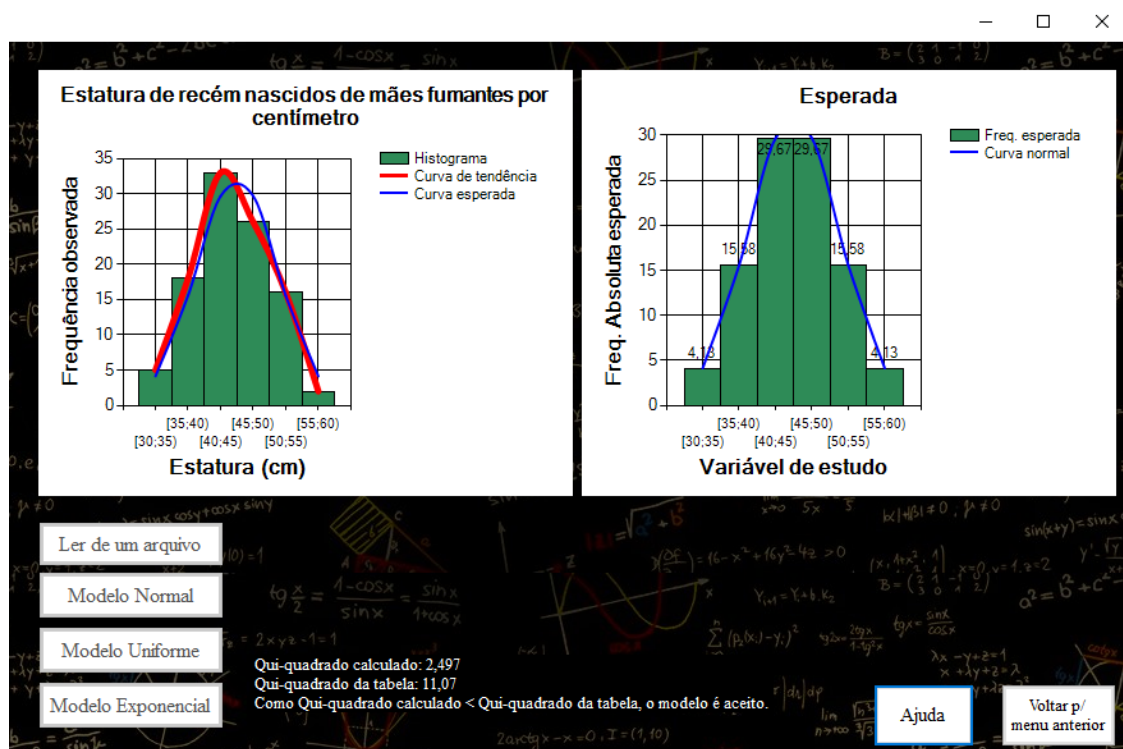


Figura 4: Resultado da variável X para um modelo normal com o software.

Por fim, a verificação para o modelo uniforme se baseou apenas em colocar valores em um intervalo curto, no caso entre 10 e 12, e verificar se o software aceitava o modelo uniforme para este caso. Os valores inseridos foram: 10.5, 10.7, 10.2, 11.09, 11.7, 11.5, 11.4.

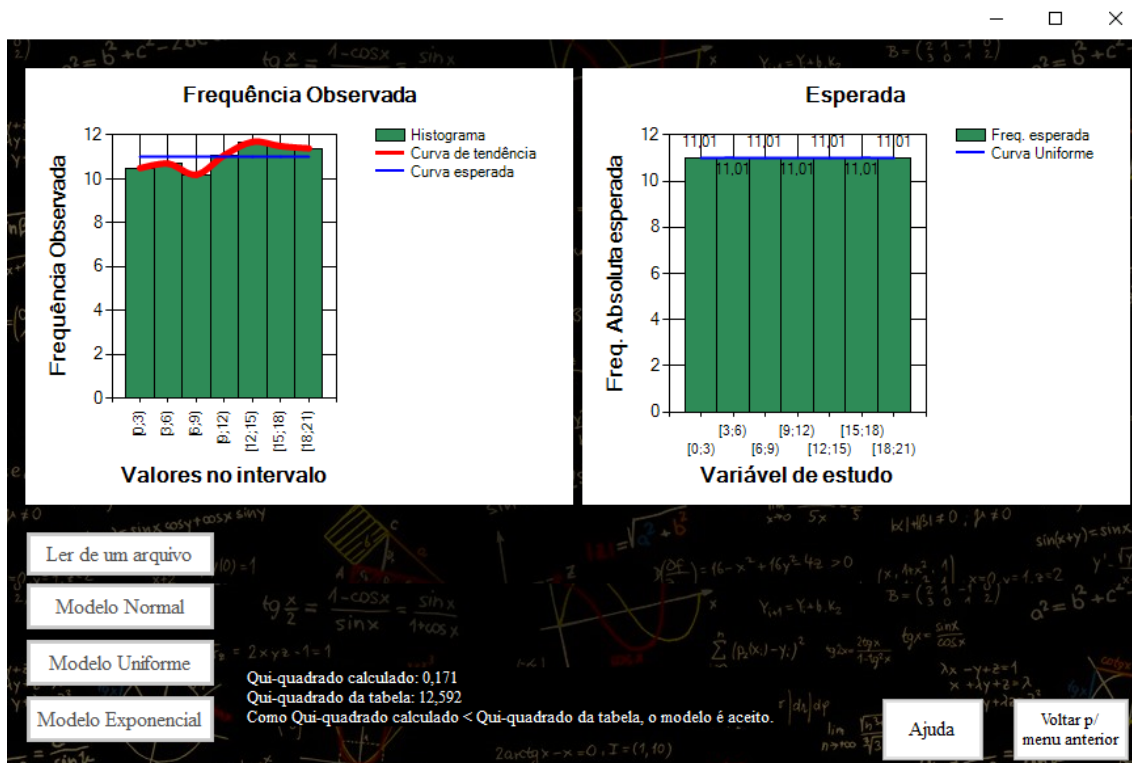


Figura 5: Resultado de uma sequência de valores para um modelo uniforme com o software.

## 4.2 Números aleatórios

Nesta área do projeto, é mostrado ao usuário que o software é adequado para a geração de números randômicos, além de possibilitar a plotagem de curvas normais e exponenciais baseadas nestes números. Para demonstrar, foram utilizados os dados: 150 para quantidade de valores e  $\alpha = 2$ .

Na figura 6, o programa gerou 150 números aleatórios entre 0 e 1 e os colocou em intervalos de 0.1. O valor 12, por exemplo, indica que foram gerados 12 números que estão no intervalo [0.3; 0.4]. Comparando o que foi gerado com a média e baseando-se no modelo uniforme, podemos ver que o teste de Aderência Qui-Quadrado aceitou o modelo uniforme para representar estes dados. Logo, o software possui uma boa geração de números aleatórios.

Nas figuras 7 e 8, a aplicação gerou números aleatórios entre 0 e 1, os quais representam os valores de  $x$ , e os colocou nas equações de cada modelo respectivamente. As equações são:  $-\frac{1}{\alpha} \cdot \text{Log}(x)$ , para o modelo exponencial e  $\frac{1}{\sqrt{2\pi}\sigma^2} \exp\left[-\frac{1}{2} * (x - \mu/\sigma)^2\right]$  para o modelo normal.



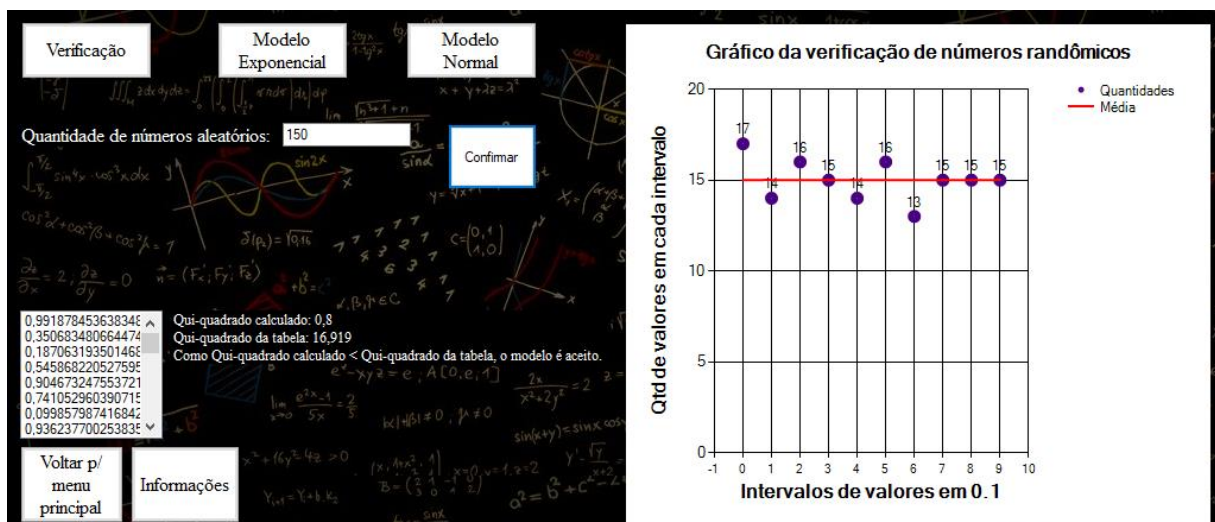


Figura 6: Gerando 150 valores aleatórios para X.

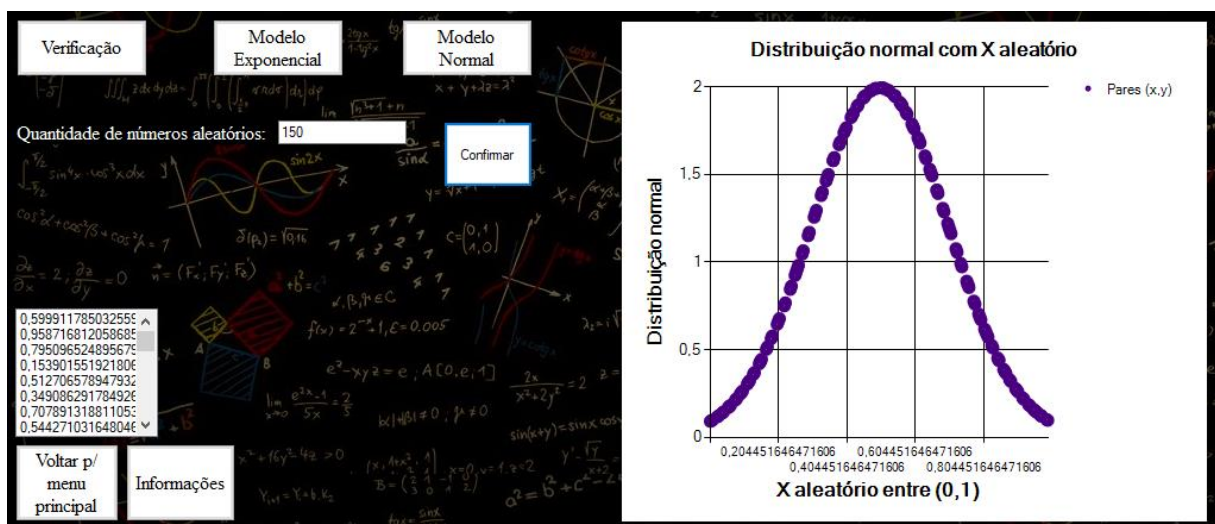


Figura 7: Curva normal com valores aleatórios para x.

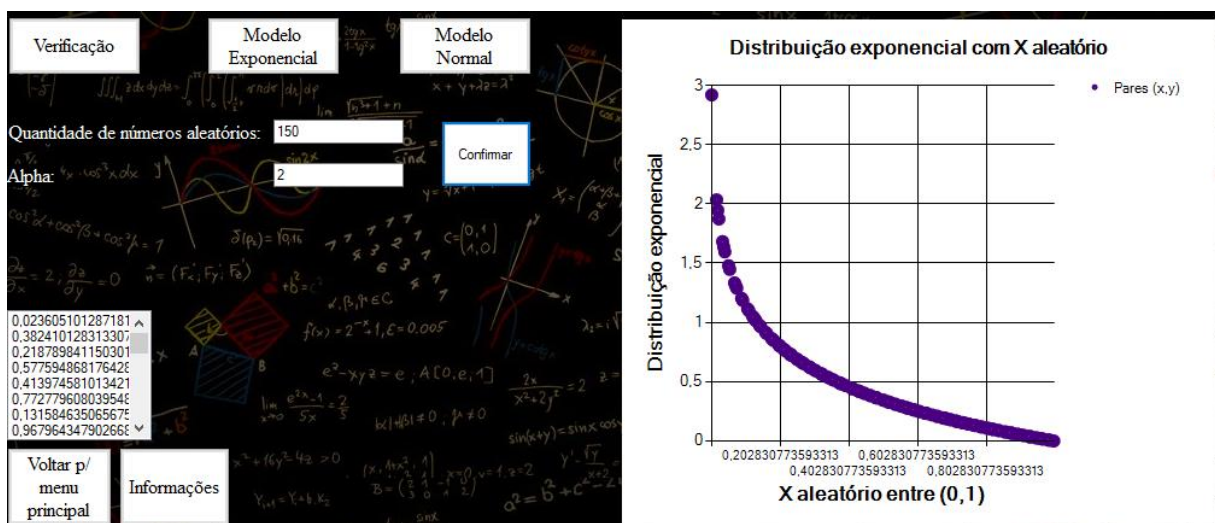


Figura 8: Curva exponencial com valores aleatórios para x.

## 4.3 Experimento 2

No experimento 2, o usuário não necessita entrar com valores, já que o software possui uma simulação preparada pra ser visualizada.

Ele pode navegar pelas partes desta solução através das setas  $\leftarrow \rightarrow$ . A divisão foi feita baseada no enunciado que inspirou o experimento. As divisões são: Simule, Média Amostral, Histograma, Examinando Gráficos e Verificação.

**5. Verificação**

**a) Há simetria no histograma e forma de sino?**  
Sim, a forma de sino é evidente e a distribuição do gráfico é simétrica, já que o este se concentra no centro, não tendendo nem a direita nem a esquerda.

**b) A média, a mediana e a moda são próximas?**  
Média: 100, Mediana: 100, Moda: 101.  
Como podemos ver pelos valores acima, sim, os valores são próximos.

**c) A amplitude não ultrapassa 6 vezes o desvio padrão?**  
Amplitude:  $103 - 95 = 8$ , Desvio padrão: 1.85864075457  
 $1.85864075457 * 6 = 11.1518445274$   
Como  $8 < 11.1518445274$ , a resposta é não.

**d) A proporção de outliers é bem pequena e não ultrapassa 0,5%?**  
Sim, pois temos apenas um outlier, como visto no boxplot.

Figura 9: Perguntas respondidas ao final do experimento 2.

Na área “1. Simule”, o usuário pode ver em quais números se basearam as informações e cálculos deste experimento. Na área “2. Média Amostral”, foram calculadas as médias amostrais para  $n = 1$ ,  $n = 2$ ,  $n = 5$  e  $n = 10$ , baseadas em um pequeno exemplo para  $n = 1$ . Em “3. Histograma”, o usuário pode ver o que acontece com o gráfico conforme  $n$  cresce. Para isso, basta mudar o valor apontado pela seta azul e verificar o que ocorre no gráfico. Os valores foram retirados da seção anterior “2. Média Amostral”. Na área “4. Examinando gráficos” existem 2 gráficos: um boxplot e um histograma. O histograma mostra quantas médias existem em cada intervalo. Exemplo: no intervalo de  $(98,99]$ , temos apenas uma média. O boxplot se baseia nos valores da área 1. Ele mostra um outlier em dourado, o valor 95, a amplitude,  $103 - 95 = 8$ , dentre outros dados. Por fim, na área “5. Verificação”, provamos que o gráfico da área 4 de fato é uma curva normal respondendo às perguntas da figura 9.

Para o cálculo da média amostral, foram coletadas linhas de acordo com o valor de  $n$ . Exemplo: para  $n = 1$ , a apenas a primeira linha foi escolhida. A partir disso, colocamos a linha em ordem crescente e contamos quantas vezes cada valor aparece para plotar o histograma.

## 5.0 Considerações Finais

A experiência com a construção do projeto auxiliou no entendimento do assunto abordado, já que para confeccionar um software simples e didático, é necessário entender quais são as possíveis dificuldades do usuário, bem como possíveis caminhos que levariam a erros na simulação. Além disso, é gratificante saber que o projeto criado auxiliaria estudantes em seus projetos pessoais e no entendimento da disciplina como um todo.

## 6.0 Referências

Wilton O. Bussab e Pedro A. Morettin. **Estatística Básica**. Editora Saraiva. Oitava Edição, Sétima Tiragem, 2015.

Silvia. Distribuição da Média Amostral. **UFPR**, 2012. Disponível em: <http://leg.ufpr.br/~shimakur/CE055/node58.html>. Acesso em: 18 de fev. de 2021.

Gouveia, Rosimar. Desvio Padrão. **Toda Matéria**, 2018. Disponível em: <https://www.todamateria.com.br/desviopadrao/> Acesso em: 20 de fev. 2021.