Artigo

O USO DO SOFTWARE EXCEL COMO APOIO DIDÁTICO AO ENTENDIMENTO DO TEOREMA DO LIMITE CENTRAL

Anna Paula Matsuoka[1], *Tatielih Pardim de Oliveira, Emerson Wruck*

Resumo: O Teorema do Limite Central é um resultado muito importante na inferência estatística e sua aplicabilidade é essencial no entendimento da teoria da estimação. Seu uso é verificado em áreas como Biologia, Engenharia, Química, Física, entre outras (Meyer, P, 1982). No entanto, esse resultado costuma gerar certo ceticismo nos alunos principiantes em probabilidade e estatística, necessitando de outras motivações para o seu melhor entendimento.

Este trabalho constituiu-se em utilizar o software Excel como ferramenta computacional no apoio ao entendimento do teorema através do uso de simulação de dados e construção de gráficos, buscando ilustrar o comportamento das distribuições amostrais.

Ferramentas de análise de dados do software, como geração de números aleatórios, medidas descritivas e histogramas, foram utilizadas. Através da construção dos resultados gerados como conseqüência do Teorema do Limite Central, os alunos têm assimilado melhor esse teorema, facilitando dessa forma o aprendizado.

INTRODUÇÃO:

A distribuição Normal ou Gaussiana, sem dúvida, é a distribuição de probabilidade mais importante na família exponencial, pois uma enorme quantidade de fenômenos químicos, físicos e biológicos, entre outros, se comporta de forma normal e pode ser modeladas através dessa distribuição. Para uma variável aleatória X com distribuição normal, com média m e variância s^2, sua densidade é dada por:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$

para $-\infty < x < \infty$, $-\infty < \mu < \infty$ e $\sigma^2 > 0$. Abaixo, ilustramos o histograma do modelo normal com média $\mu = 10$ e $\sigma^2 = 2$.

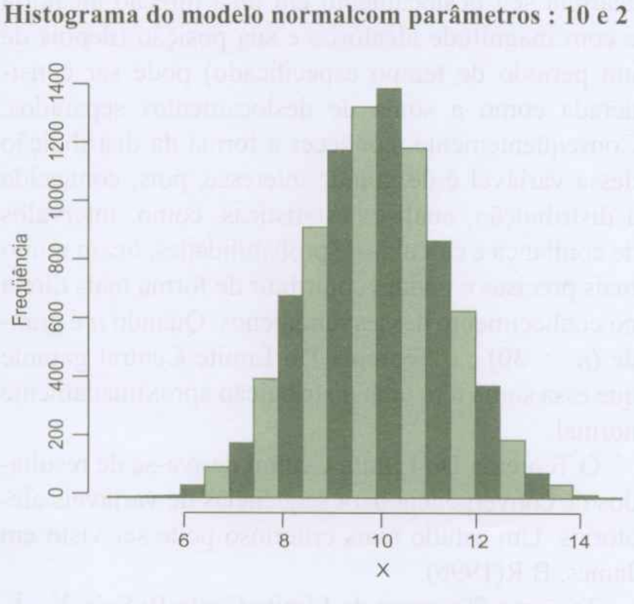


Figura 1: Histograma de uma Distribuição Normal com $\mu = 10$ e $\sigma^2 = 2$

Percebam-se o gráfico em forma de sino e a concentração em torno da média $m = 10$.

O Teorema do Limite Central, que é um resultado da lei dos grandes números (James, B.R., 1996), garante que, seja qual for a distribuição da variável de interesse, para grandes amostras, num caso particular, a distribuição das médias amostrais será aproximadamente distribuída de forma normal, e tendera a uma distribuição normal exata à medida que o tamanho da amostra cresce. Essa afirmação tem causado grande ceticismo entre os alunos no primeiro momento. Visando diminuir esse ceticismo, o trabalho teve como objetivo apresentar uma aplicação experimental com apoio computacional no software Excel (Laponi, J.C.,1997), através do uso de medidas descritivas da a construção de histogramas, buscando evidenciar de forma mais didática esse teorema.

[1] japapriest@hotmail.com

Teorema do Limite Central

Verifica-se que, em muitos problemas, a variável aleatória em estudo poderá ser representada pela soma de n variáveis aleatórias independentes. Por exemplo, o consumo da eletricidade em uma cidade, em uma época qualquer, é a soma das procuras de um grande número de consumidores individuais. A quantidade de água em um reservatório pode ser pensada como o resultado da soma de um grande número de contribuições individuais. E o erro de mensuração em um experimento físico é composto de muitos erros pequenos, não observáveis, os quais podem ser admitidos como aditivos. O bombardeamento molecular, que uma partícula suspensa em um líquido está sofrendo, acarreta seu deslocamento em uma direção aleatória e com magnitude aleatória, e sua posição (depois de um período de tempo especificado) pode ser considerada como a soma de deslocamentos separados. Conseqüentemente, conhecer a forma da distribuição dessa variável é de grande interesse, pois, conhecida a distribuição, análises estatísticas, como, intervalos de confiança e cálculos de probabilidades, ficam muito mais precisas e podem contribuir de forma mais direta no conhecimento desses fenômenos. Quando n é grande ($n > 30$), o Teorema Do Limite Central garante que essa soma terá uma distribuição aproximadamente normal.

O Teorema Do Limite Central deriva-se de resultados de convergência para seqüências de variáveis aleatórias. Um estudo mais criterioso pode ser visto em James, B.R(1996).

Teorema (Teorema do Limite Central): Seja $X_1, X_2, \ldots$ uma seqüência de variáveis aleatórias independentes e identicamente distribuídas com média μ e variância σ^2. Seja $S_n = X_1 + X_2 + \ldots + X_n$ então $Z_n = \dfrac{S_n - n\mu}{\sigma\sqrt{n}}$ converge em distribuição para Z que tem distribuição normal com média zero e variância 1.

A demonstração desse resultado através da função geradora de momentos pode ser vista em Dantas, C.A.B.(1997).

Algumas observações pertinentes:

- No caso particular em que cada $X_i = 1$ com probabilidade p e $X_i = 0$ com probabilidade $(1 - p)$, $S_n = \sum_{i=1}^{n} X_i$ é o número de sucessos (ocorrências de 1) em n ensaios de Bernoulli. Sabe-se que S_n tem distribuição binomial com média np e variância $np(1-p)$. Como conseqüência, o Teorema Do Limite Central diz que $Z = \dfrac{S_n - np}{\sqrt{np(1-p)}}$ tem, para n suficientemente grande, distribuição aproximadamente normal com $\mu = 0$ e $\sigma^2 = 1$, ou seja, $Z \sim N(0,1)$.
- A razão $\dfrac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}}$ pode ser reescrita, dividindo-se numerador e denominador por n. Dessa forma, tem-se:

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

e o Teorema Do Limite Central diz que, para n suficientemente grande, a média amostral tem distribuição aproximadamente normal com média μ e a variância σ^2/n, ou seja, $\bar{X} \sim N(\mu, \sigma^2/n)$, o que é equivalente a dizer que $\dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$.

Essa última observação é muito utilizada para resolução de problemas aplicados e causa grande ceticismo entre os alunos, pois o teorema garante que, para qualquer distribuição de X, se n é suficientemente grande, tem-se que a distribuição das médias amostrais será normalmente distribuída com a mesma média de X e variância σ^2/n, ou seja, $\bar{X} \sim N(\mu, \sigma^2/n)$. Abaixo serão propostos dois experimentos com apoio computacional para ilustração desse resultado.

METODOLOGIA

No trabalho, buscou-se evidenciar o Teorema Do Limite Central através da geração de variáveis aleatórias bem diferentes da distribuição normal e estudar o comportamento das médias dessas amostras.

O modelo escolhido para a geração foi o modelo uniforme, pelo fato de ele ter uma característica muito particular. A distribuição uniforme é um modelo contínuo, empregado quando um fenômeno pode ocorrer, com igual chance, em qualquer ponto de uma região ou intervalo definido $[a,b]$. A função de densidade de probabilidade de uma variável aleatória X que obedece a uma distribuição uniforme é dada por: $f(x) = \dfrac{1}{b-a}$ para $a \le x \le b$ e $f(x) = 0$, fora desse intervalo.

Tem-se que $E(X) = \dfrac{a+b}{2}$ e $Var(X) = \dfrac{(b-a)^2}{12}$

O gráfico da função de densidade de probabilidade de X é uma função constante em $\dfrac{1}{b-a}$. Abaixo, ilustramos o histograma do modelo uniforme para $a = 0$ e $b = 2$.

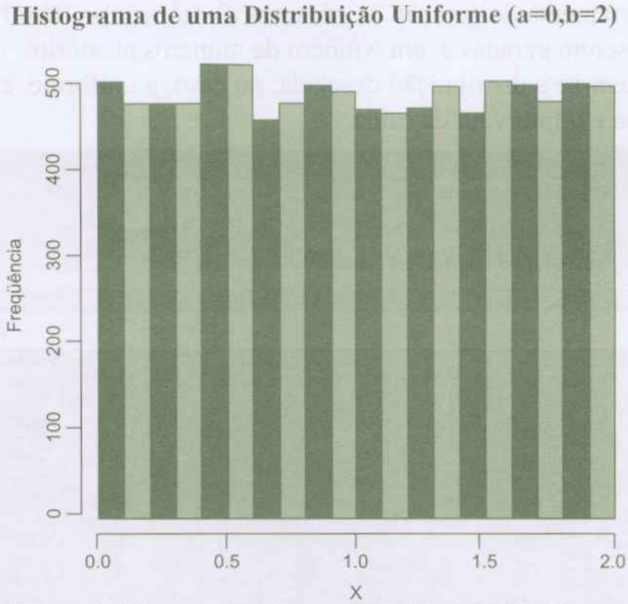


Figura 2: Histograma de uma Distribuição Uniforme

Pelo Teorema Do Limite Central, espera-se que, para $n > 30$, as médias amostrais dessa população apresentem uma distribuição normal com forma de sino, com média igual à média populacional e variância menor, inversamente proporcional a n.

Na simulação com o Software Excel, foram geradas uma seqüência de 250 amostras independentes de tamanho 15 de um modelo uniforme com parâmetros $a = 0$ e $b = 2$ e uma outra seqüência com o mesmo número de amostras, só que, agora, com tamanho 60. Assim, com o apoio de medidas descritivas e o histograma, vamos evidenciar que, quando o tamanho da amostra aumenta de 15 para 60, o histograma da distribuição das médias tenderá a forma normal com média $\mu_{\bar{x}} = \dfrac{a+b}{2} = 1$ e variância $\sigma^2_{\bar{x}} = \dfrac{\frac{(b-a)^2}{12}}{n}$.

Gerações de variáveis aleatórias no Excel

Para geração de números aleatórios no Software Exce,l usamos a seqüência através do menu Ferramentas + Análise de dados + Geração de Números Aleatórios e selecionamos o modelo uniforme, como na ilustração da Figura 3.

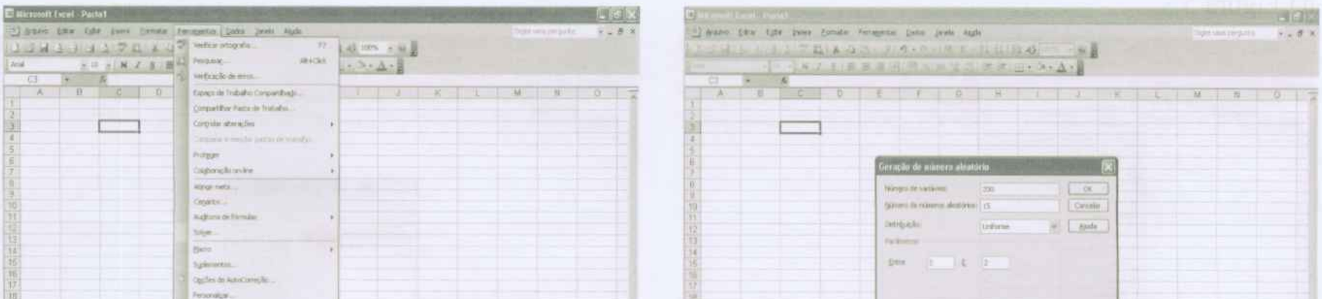


Figura 3: Geração de Números Aleatórios no Excel

Na caixa de diálogo da ferramenta de geração de números aleatórios, no item Número de variáveis, indiquemos o número de amostras a serem geradas e, em Número de números aleatórios, o tamanho da amostra (n). No campo Distribuição, selecionem-se a distribuição desejada, no caso, a uniforme, e, na seqüência, os parâmetros da mesma, a semente aleatória e o Intervalo de saída.

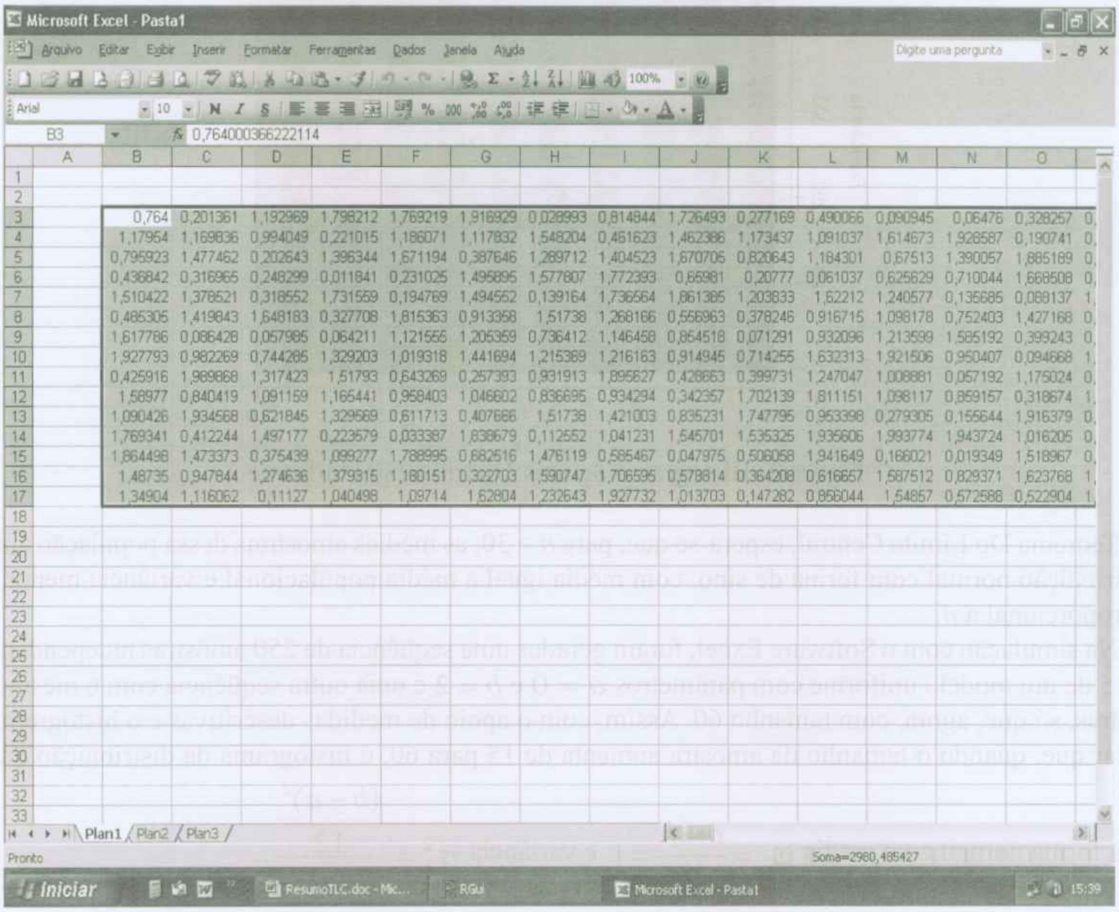
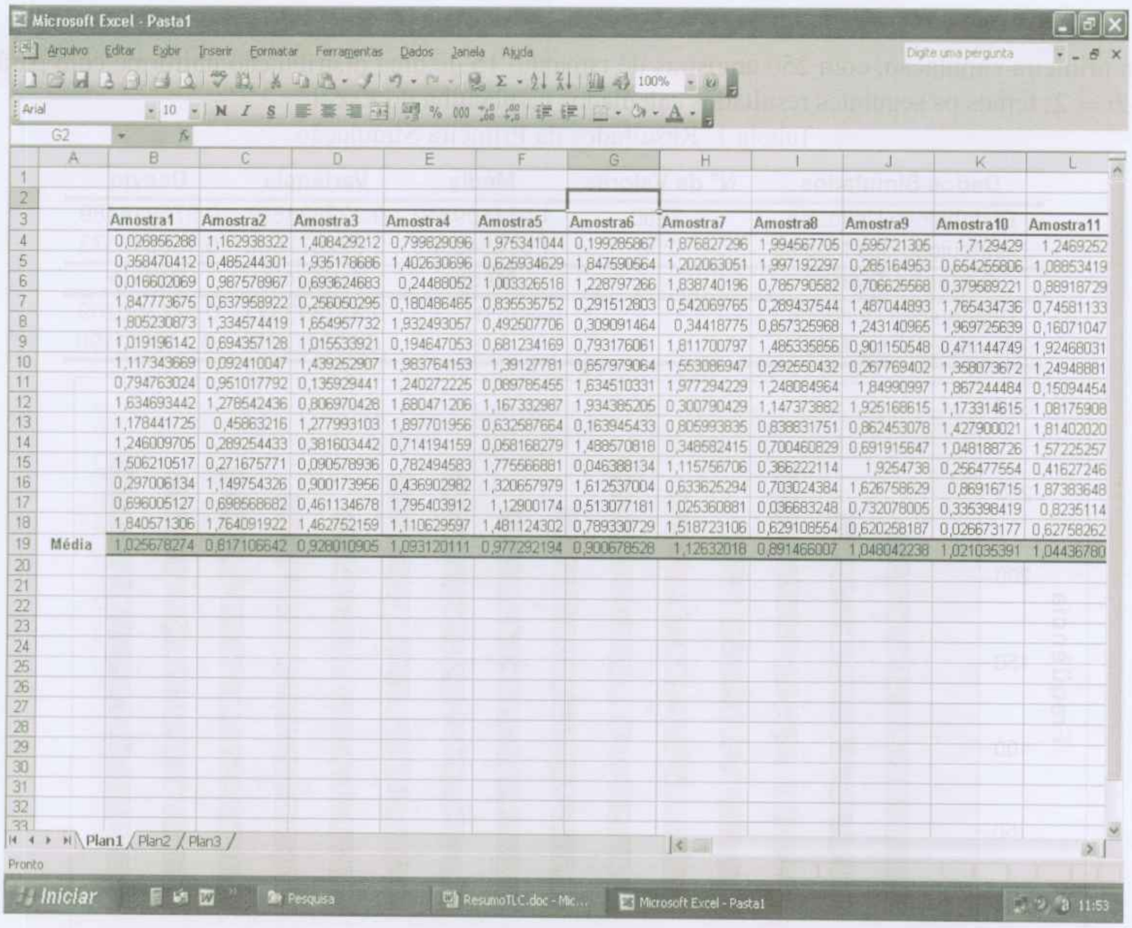


Figura 4: Resultado da Geração de Números.

O resultado dessa simulação está ilustrado na Figura 4. Nesta planilha foram geradas 250 colunas (amostras) e 15 linhas (tamanho das amostras). Pode ocorrer que a Ferramenta Análise de Dados não esteja disponível, nesse caso deve-se habilitá-lá através da seqüência: Ferramentas + Suplementos + Ferramentas de Análise, como segue na Figura 5 .

Figura 5: Habilitação da Ferramenta Análise de Dados

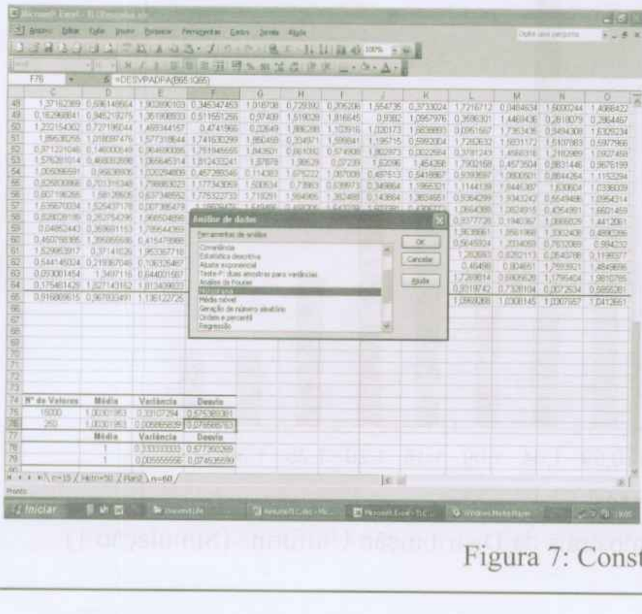
Para cada amostra gerada, calculamos a média amostral, gerando um novo conjunto de dados, o conjunto das médias amostrais da distribuição uniforme que, nesse caso, é composto de 250 observações. Abaixo ilustramos a disposição dos dados na planilha para a simulação 1.



	A	B	C	D	E	F	G	H	I	J	K	L
3		Amostra1	Amostra2	Amostra3	Amostra4	Amostra5	Amostra6	Amostra7	Amostra8	Amostra9	Amostra10	Amostra11
4		0,026856288	1,162938322	1,408429212	0,799829096	1,975341044	0,199285867	1,876827296	1,994567705	0,595721305	1,7129429	1,2469252
5		0,358470412	0,485244301	1,935178686	1,402630696	0,625934629	1,847590564	1,202063051	1,997192297	0,285164953	0,654255806	1,08853419
6		0,016602069	0,987578967	0,693624683	0,24488052	1,003326518	1,228797266	1,838740196	0,785790582	0,706625568	0,379589221	0,88918729
7		1,847773675	0,637958922	0,256050295	0,180486465	0,835535752	0,291512803	0,542069765	0,289437544	1,487044893	1,765434736	0,74581133
8		1,805230873	1,334574419	1,654957732	1,932493057	0,492507706	0,309091464	0,34418775	0,857325968	1,243140965	1,969725639	0,16071047
9		1,019196142	0,694357128	1,015533921	0,194647053	0,681234169	0,793176061	1,811700797	1,485335856	0,901150548	0,471144749	1,92468031
10		1,117343669	0,092410047	1,439252907	1,983764153	1,39127781	0,657979064	1,553086947	0,292550432	0,267769402	1,358073672	1,24948881
11		0,794763024	0,951017792	0,135929441	1,240272225	0,089785455	1,634510331	1,977294229	1,248084964	1,84990997	1,867244484	0,15094454
12		1,634693442	1,278542436	0,806970428	1,680471206	1,167332987	1,934385205	0,300790429	1,147373882	1,925168615	1,173314615	1,08175908
13		1,178441725	0,45863216	1,277993103	1,897701956	0,632587664	0,163945433	0,805993835	0,838831751	0,862453078	1,427900021	1,81402020
14		1,246009705	0,289254433	0,381603442	0,714194159	0,058168279	1,488570818	0,348582415	0,700460829	0,691915647	1,048188726	1,57225257
15		1,506210517	0,271675771	0,090578936	0,782494583	1,775566881	0,046388134	1,115756706	0,366222114	1,9254738	0,256477554	0,41627246
16		0,297006134	1,149754326	0,900173956	0,436902982	1,320657979	1,612537004	0,633625294	0,703024384	1,626758629	0,86916715	1,87383648
17		0,696005127	0,698568682	0,461134678	1,795403912	1,12900174	0,513077181	1,025360881	0,036683248	0,732078005	0,335398419	0,8235114
18		1,840571306	1,764091922	1,462752159	1,110629597	1,481124302	0,789330729	1,518723106	0,629108554	0,620258187	0,026673177	0,62758262
19	Média	1,025678274	0,817106642	0,928010905	1,093120111	0,977292194	0,900678528	1,12632018	0,891466007	1,048042238	1,021035391	1,04436780

Figura 6: Dados gerados na simulação 1

O Próximo passo é calcular a média, a variância e o desvio-padrão dos dados. No Excel, os comandos são, respectivamente: =media(intervalo de dados), =varp(intervalo de dados), =desvpadpa(intervalo de dados). Para a construção do Histograma, obedeça-se a seqüência : Ferramentas + Análise de dados + Histograma, como ilustrado na Figura7 abaixo:



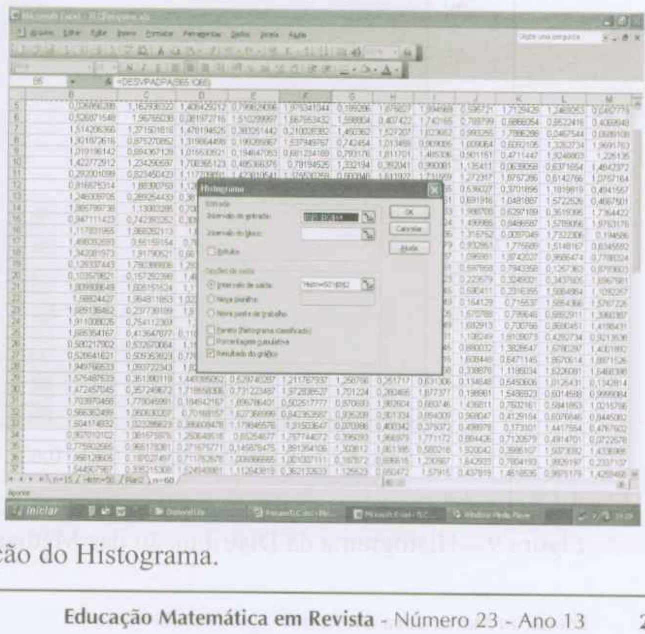


Figura 7: Construção do Histograma.

Na caixa de diálogo da ferramenta Histograma, no campo Seqüência de dados, entre-se com os dados de onde será gerado o histograma, no campo Intervalo de bloco, entre-se com o intervalo de bloco (intervalo de classes), se tiver, habite-se o campo Resultado gráfico e escolha-se o Intervalo de saída do resultado.

RESULTADOS

Para a primeira simulação, com 250 amostras de tamanho 15 de uma distribuição uniforme com parâmetros $a = 0$ e $b = 2$, temos os seguintes resultados calculados na planilha do Excel.

Tabela 1: Resultados da Primeira Simulação

Dados Simulados	Nº de Valores	Média	Variância	Desvio
Distribuição Uniforme	3750	1,014409549	0,332834597	0,576918189
Distribuição das Médias	250	1,014540369	0,022577605	0,150545673
Situação Teórica		**Média**	**Variância**	**Desvio**
Distribuição Uniforme		1	0,333333333	0,577350269
Distribuição das Médias		1	0,022222222	0,149071198

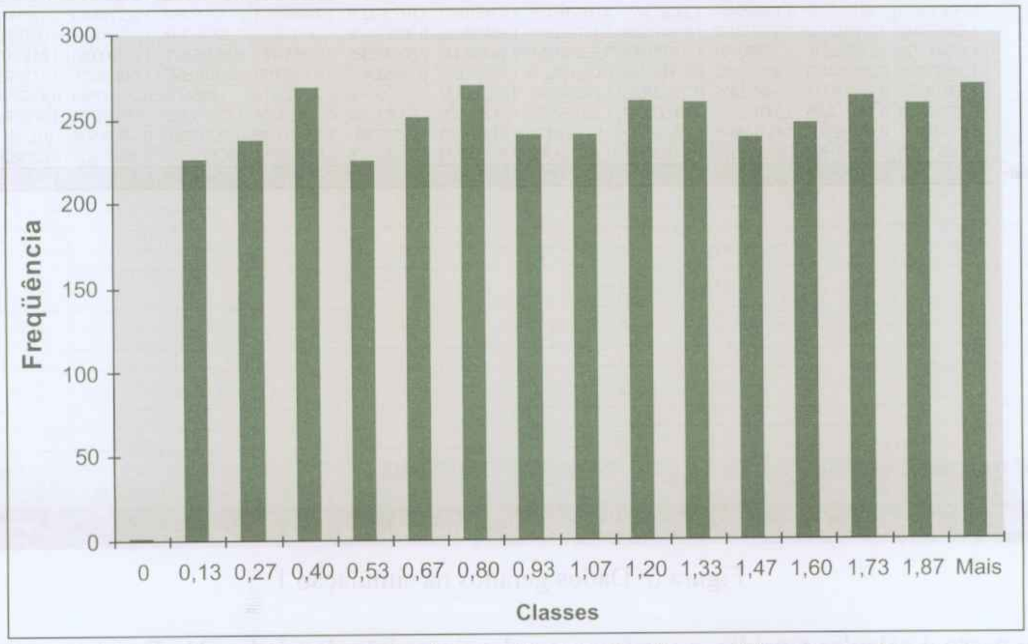


Figura 8 – Histograma da Distribuição Uniforme (Simulação 1)

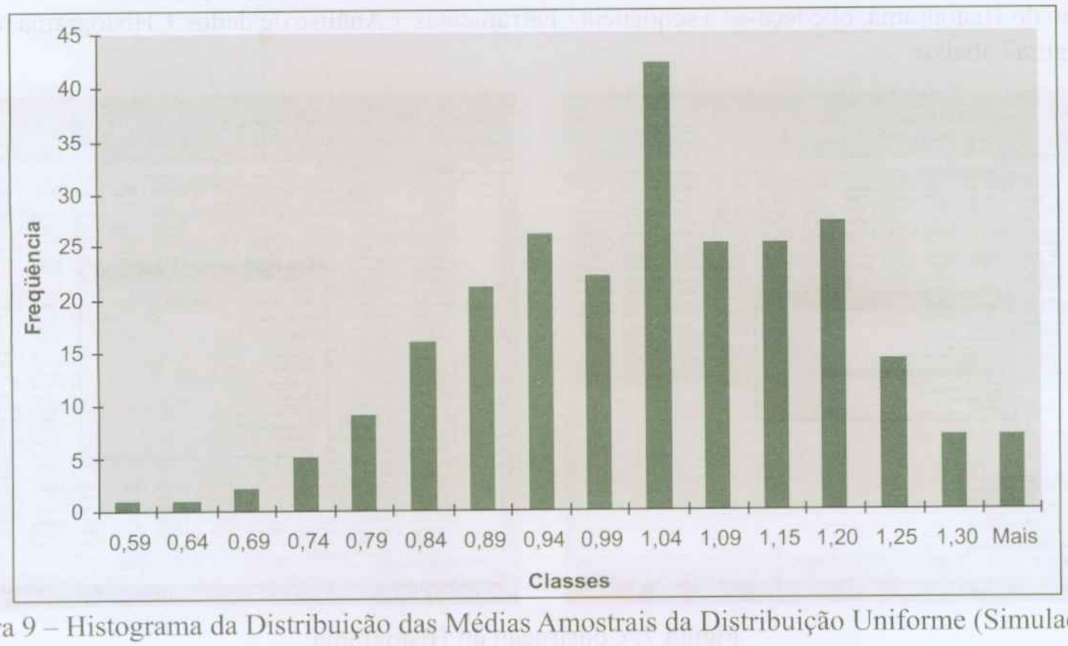


Figura 9 – Histograma da Distribuição das Médias Amostrais da Distribuição Uniforme (Simulação 1)

Perceba-se que os resultados, através da simulação, se aproximam muito dos resultados teóricos, o que evidencia o teorema. No entanto, a forma da distribuição das médias amostrais ainda não se parece muito com a forma normal. Para a segunda simulação, com 250 amostras de tamanho 60 de uma distribuição uniforme com parâmetros $a = 0$ e $b = 2$, temos os seguintes resultados calculados na planilha do Excel.

Tabela 2: Resultados da Segunda Simulação

Dados Simulados	Nº de Valores	Média	Variância	Desvio
Distribuição Uniforme	15000	1,00301953	0,33107294	0,575389381
Distribuição das Médias	250	1,00301953	0,005865839	0,076588763
Situação Teórica		**Média**	**Variância**	**Desvio**
Distribuição Uniforme		1	0,333333333	0,577350269
Distribuição das Médias		1	0,005555556	0,074535599

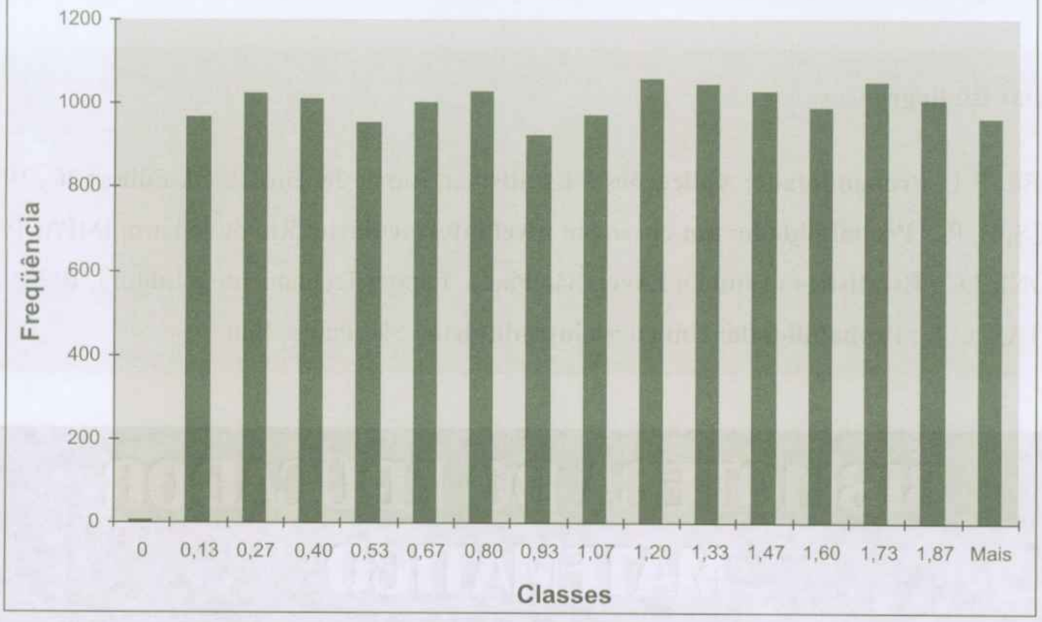


Figura 10 – Histograma da Distribuição Uniforme (Simulação 2)

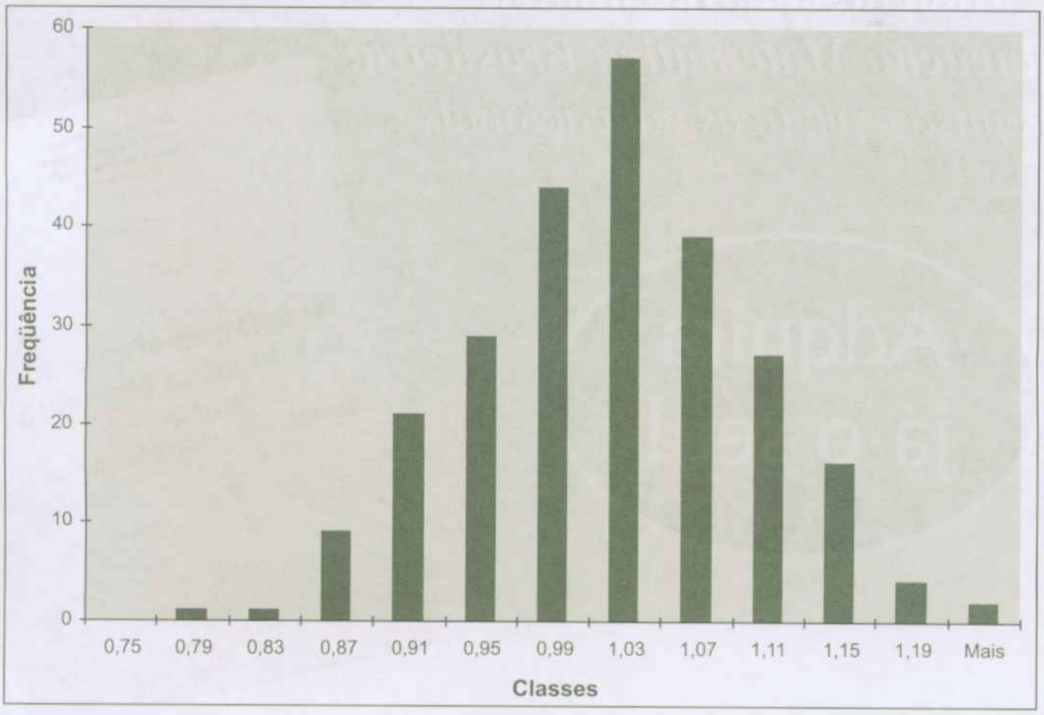


Figura 11 – Histograma da Distribuição das Médias Amostrais da Distribuição Uniforme (Simulação 2)

Nessa segunda simulação, com $n = 60$, percebe-se uma melhor aproximação aos resultados teóricos e, o mais importante, fica bem caracterizada a forma normal para a distribuição das médias, evidenciando experimentalmente o Teorema do Limite Central.

CONCLUSÕES E DISCUSSÃO

Neste trabalho, conseguiu-se uma ilustração fiel dos resultados do Teorema Do Limite Central através da simulação computacional com o software Excel. A construção do experimento, com o uso da simulação, faz com que o aluno entenda melhor as condições do problema, além de proporcionar uma maior familiarização com o Software, que é de grande aplicação no mercado. A ilustração gráfica do resultado acaba com o ceticismo existente, facilita a aprendizagem e incentiva o uso do resultado nas situações práticas.

Dessa forma, o uso desse procedimento se mostrou eficaz como ferramenta didática e, o mais importante, pode ser utilizado em qualquer laboratório de informática que disponha de tal software.

Referências Bibliográficas:

MEYER, P. L, **Probabilidade: Aplicações à Estatística**, Rio de Janeiro; 2º ed; Editora JC, 1982.

JAMES, B. R.,: **Probabilidade: um curso em nível intermediário**; Rio de Janeiro; IMPA, 1996.

LAPONI, J.C.; **Estatística usando o Excel**; São Paulo; Laponi Treinamento e Editora, 1996.

DANTAS, C.A.; **Probabilidade: Um curso introdutório**. São Paulo, Edu