CONTRIBUIÇÕES PARA O ENSINO DA DISTRIBUIÇÃO NORMAL OU CURVA DE GAUSS EM CURSOS DE GRADUAÇÃO

Hélio Radke Bittencourt
Departamento de Estatística, PUCRS
Lori Viali
Departamento de Estatística, PUCRS e UFRGS

RESUMO

A distribuição normal ou curva de Gauss-Moivre-Laplace é o principal modelo probabilístico contínuo, pois serve de base para a principal área da Estatística: a Inferência. A distribuição normal faz parte do currículo de praticamente todas as disciplinas que envolvem Probabilidade, Estatística ou Estocástica. Desta forma a sua compreensão, além da simples memorização, é fundamental para que o processo de ensino-aprendizagem tenha sucesso. Neste estudo o ensino da distribuição normal e do teorema do limite central é feito com o recurso de três exemplos práticos. O primeiro explora a geração de números aleatórios e o teorema do limite central; o segundo mostra a relação da média com o desvio padrão e o terceiro utiliza resultados de um concurso vestibular para ilustrar cálculos que envolvem a distribuição.

Palavras-chave: Educação Estatística, distribuição Normal, curva de Gauss.

ABSTRACT

The normal distribution or Gauss-Moivre-Laplace model is the main continuous model in the Probability theory, because it is the base to the Statistical Inference area. The normal distribution is included in practically all disciplines involving probability theory, statistics or randomness. In this way, it real understanding, beyond simple memorization, is basic to a successful teach-learning process. In this study the teaching of normal model and the central limit theorem are realized using three practical examples as a resource. The first explores how to generate random numbers in Excel and the central limit theorem; the second shows the relation between mean and standard deviations in Normal distribution and the third uses data from university entrance examinations to illustrate the calculation of probabilities.

Keywords: Statistical Education, Normal distribution, Gaussian model.

1 Introdução

A distribuição normal ou curva de Gauss é um modelo que descreve o comportamento de vários fenômenos aleatórios. Os conteúdos referentes a este modelo geralmente estão inseridos em disciplinas de Probabilidade e/ou Estatística sob um tópico denominado “variáveis aleatórias contínuas”.

Antes de prosseguir, façamos a distinção entre Probabilidade e Estatística. Existe uma grande confusão entre estes dois termos, apresentados de maneira equivocada numa grande quantidade de livros texto de Estatística. Contrariando muitos pensamentos, Probabilidade não é uma área da Estatística, mas sim uma área da Matemática que trata da modelagem de fenômenos aleatórios. Os fenômenos ou a forma de tratá-los podem ser divididos em dois tipos: os determinísticos e os não-determinísticos ou probabilísticos. Os fenômenos determinísticos são modelados pela matemática através de equações, equações diferenciais ou sistemas de equações que fazem parte de grande quantidade de disciplinas, como o cálculo ou a álgebra, entre outras. Os fenômenos não-determísticos são assuntos da teoria da probabilidade que, quase invariavelmente, está inserida em disciplinas de Estatística. Salvo eventuais exceções, os únicos cursos de graduação brasileiros que têm disciplinas dedicadas apenas à Probabilidade são os de Bacharelado em Estatística. Bayer et al. (2005) ressaltam que conteúdos de Probabilidade são geralmente lecionados em disciplinas de Estatística no Ensino Superior. Já no ensino fundamental e médio, os conteúdos de Probabilidade e de Estatística estão incluídos em disciplinas de matemática.

Mas, se Probabilidade é uma área da Matemática, por que ela está tão relacionada à Estatística? A resposta para esta pergunta é simples: a principal área da Estatística – denominada Estatística Inferencial – tem a sua base alicerçada sobre a Teoria da Probabilidade. A Estatística Inferencial é a área que permite ao pesquisador transcender os resultados encontrados numa amostra para toda a população da qual ela foi extraída. Mesmo numa grande amostra o processo de inferência está sujeita à erros e, em sendo ela probabilística, é possível dimensionar esse erro.

A Inferência Estatística e a Teoria da Amostragem são tópicos dependentes da Teoria das Probabilidades e destas o principal papel é desempenhado pela distribuição Normal. Esse modelo serve de base na construção de intervalos de confiança e na realização dos testes de hipóteses. De fato, na prática, a distribuição *t* de *Student* é até mais utilizada que o modelo normal, mas é mais fácil apresentar e entender o modelo de *Student* com a compreensão anterior do modelo Normal.

2 Histórico

A distribuição normal foi introduzida inicialmente pelo matemático francês Abraham de Moivre (1667 – 1754) em um artigo que foi reimpresso na segunda edição do seu livro “A doutrina do acaso” de 1738. Ele percebeu que à medida que número de eventos do lançamento de moedas aumentava a distribuição binomial se aproximava de uma curva suave.

Seus resultados foram estendidos pelo francês Pierre Simon de Laplace (1748 – 1827) em seu livro “Teoria Analítica da Probabilidade” de 1812, num resultado que hoje é conhecido como Teorema de Moivre-Laplace. Laplace utilizou a distribuição na análise de erros de experimentos. Ele também mostrou que mesmo uma distribuição não sendo normal a média de repetidas amostras dessa distribuição é aproximadamente normal e que quanto maior for o tamanho da amostra melhor será essa aproximação.

O italiano Galileo Galilei (1564 - 1642) já havia notado que esses erros eram simétricos e que os valores pequenos apresentavam uma freqüência de ocorrência maior do que os valores grandes.

A primeira pessoa a aplicar a distribuição normal na área social foi o belga Adolph Quetelet (1796-1874) que coletou dados sobre medidas do peito de soldados escoceses e da altura de soldados franceses e verificou que elas podiam ser modeladas pela distribuição.

De forma independente os matemáticos alemães Robert Adrian (1775-1843) em 1808 e Carl Friedrich Gauss (1777 - 1855) em 1809 desenvolveram a equação da distribuição e mostraram que modelava bastante bem os erros de observações astronômicas. Gauss utilizou a equação em 1809 para justificar o MMQ (Método dos Mínimos Quadrados), introduzido por Adrien Marie Legendre (1752 - 1833), em 1809, mas já utilizado por ele desde 1794 (Stigler,1999; Daw e Pearson, 1972)

A importância dessa distribuição reside principalmente no fato de que muitos fenômenos naturais apresentam uma distribuição normal ou aproximadamente normal. Além disso, as médias de amostras retiradas de distribuição qualquer tendem a apresentar comportamento normal à medida que o número de observações (tamanho da amostra) aumenta.

O nome de distribuição normal ou curva normal foi dado de forma independente pelo filósofo americano Charles S. Peirce (1839 – 1914), pelo antropólogo e geneticista britânico Francis Galton (1822 – 1911) e pelo economista alemão Wilhelm Lexis (1837 – 1914) por volta de 1875. A terminologia de curva em forma de sino foi cunhada em 1872 pelo francês Esprit Pascal Jouffret que denominou a normal bivariada de superfície campanular (*bell surface*).

A distribuição normal é mais conhecida como curva de Gauss dando razão a lei da eponímia de Stigler de afirma “nenhuma descoberta científica é batizada com o nome do seu descobridor original”.

3 O ensino do modelo normal

Vários pesquisadores já se preocuparam com o ensino da distribuição Normal. Batanero et al (1999, 2001, 2004) conduziu estudos voltados ao ensino e aprendizagem do modelo enfocando também a parte inferencial. Doane (2004) mostrou que o ensino de distribuições probabilísticas pode ser mais eficaz com o uso de simulações, procedimento que também é defendido por Viali (2004, 2005).

Em cursos de graduação o modelo aparece inserido no tópico dedicado as variáveis aleatórias contínuas quando o assunto é Probabilidade e na amostragem quando se tratar de Estatística. Em geral, os modelos probabilísticos contínuos são vistos após os conceitos iniciais de probabilidade e de variáveis aleatórias discretas. Sugere-se que a apresentação do modelo normal seja iniciada com o seu resgate históricos seguido de uma exploração gráfica. Essa exploração deve ressaltar o efeito dos parâmetros na forma do modelo. A Figura 2, mostra um exemplo que os autores utilizam em disciplinas da área de Exatas.

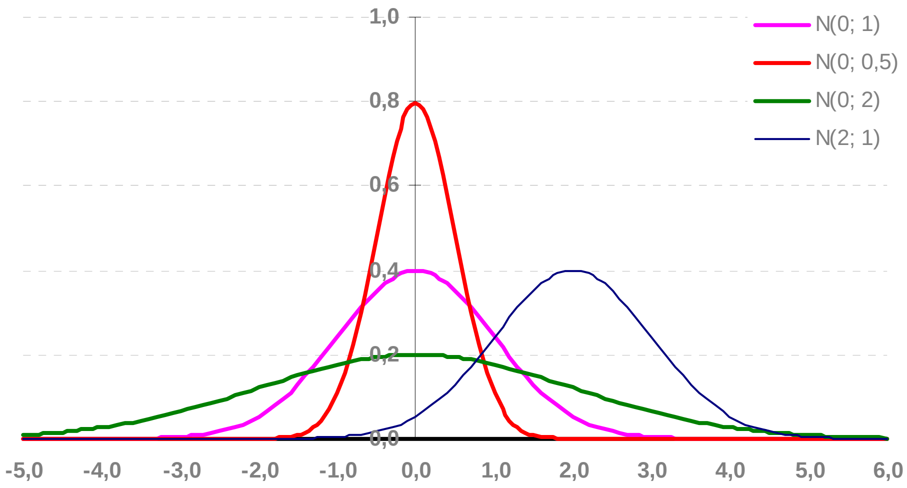


Figura 2 – Modelos normais desenhados com o auxílio da planilha

O cálculo das probabilidades, áreas sob a curva, é talvez o exemplo clássico do ensino tradicional que poderia ser classificado como “espaguete”. Quase que invariavelmente os livros didáticos apresentam regras sem nenhuma especificação ou contexto de como essas áreas (probabilidades) são obtidas e, freqüentemente, utilizam tabelas da parte positiva da normal padrão reforçando o ensino memorialistico e mecânico.

O que propomos é a utilização da função de distribuição acumulada que pode ser associada com as freqüências acumuladas da estatística descritiva. É possível e conveniente fazer um paralelo entre os modelos probabilísticos, isto é a teoria, com os histogramas de variáveis contínuas da estatística descritiva. A associação é sempre benéfica, pois está evocando os conhecimentos anteriores do aluno numa abordagem construtivista. Óbvio que essa abordagem supõe que os conteúdos de descritiva tenham sido vistos antes dos de probabilidade o que nem sempre é o caso. Muitos cursos de Estatística Básica envolvendo conteúdos de probabilidade começam por ela, desperdiçando, assim, uma ótima oportunidade de tirar proveito da sinergia entre a prática e a teoria e optando por um ensino mecânico ao invés de construtivo.

O uso da Função de Distribuição Acumulada, denotada por F(X), pode ser feito com todos os modelos probabilísticos, sejam eles discretos ou contínuos. No entanto ele se mostra de maior utilidade com os modelos contínuos, especialmente com a curva normal, que não é integrável e exige a utilização de uma tabela ou de um software para o cálculo das áreas (probabilidades)

Assim se for estabelecido que f(x) é a densidade do modelo a F(x) é a função de distribuição ou de repartição, isto é, $F(x) = P(X \leq x)$. Como existem infinitas curvas normais não é possível tabelar a todas, assim, escolhe-se uma para ser tabelada. A escolhida é denominada normal padrão e qualquer outro modelo pode ser reduzido a ela através da mudança $Z = (X - \mu)/\sigma$. Novamente aqui pode ser evocado a Estatística Descritiva, lembrando o que ocorre quando um conjunto de dados sofre a transformação acima. Fixado essa simbologia pode-se então utilizar a notação grega (convencional) para a densidade $\varphi(z)$ e para a função de distribuição $\Phi(z)$. Assim a área à esquerda (integral) da curva normal padrão será dada por $\Phi(z) = P(Z \leq z)$. Com essa providência é possível ter uma regra de fácil entendimento que pode ser generalizada para as demais situações.

Evidentemente a carga simbólica e de cálculo pode ser maior ou menor dependendo do curso que se está trabalhando sendo maior para a área de exatas e mais informal com as demais áreas. Adotando uma tabela envolvendo os percentis de -4 a mais 4 para a curva normal padrão as demais situações podem ser colocadas da seguinte forma: cálculo da área à direita da curva, isto é, $P(Z > z) = 1 - P(Z \leq z) = 1 - \Phi(z) = \Phi(-z)$, devido a simetria da curva.

O cálculo da área entre dois valores segue o mesmo raciocínio. Assim $P(z_1 < Z < z_2) = F(z_2) - F(z_1) = \Phi(z_2) - \Phi(z_1)$, evidenciando assim o teorema fundamental do cálculo. Essa associação não é feita nos livros didáticos, perdendo-se assim uma boa oportunidade de integração de conteúdos. O aluno, mesmo o da área de exatas, acaba vendo esse tipo de conteúdo apenas como um conjunto de regras com pouco ou nenhum significado deixando de perceber assim a beleza subjacente e a riqueza de idéias e conceitos envolvidos. Com esta forma de abordagem um assunto normalmente aborrecido passa a ter vida e o aluno verifica que o esforço empregado no estudo do Cálculo é agora compensado em parte.

Como exercício, o professor pode solicitar aos alunos que obtenham as seguintes probabilidades a partir de uma tabela ou com o recurso de um software. É importante ressaltar que, independentemente dos valores de μ e σ, estas probabilidades se mantêm.

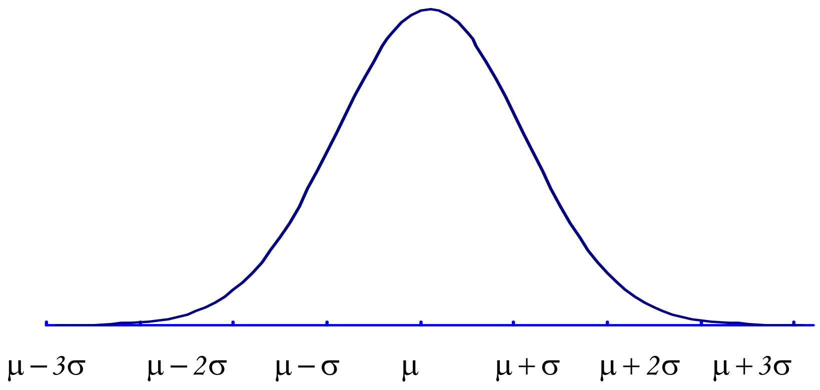
$$P(\mu - \sigma \leq X \leq \mu + \sigma) = P(-1 \leq Z \leq 1) \cong 68{,}26\%$$

$$P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) = P(-2 \leq Z \leq 2) \cong 95{,}45\%$$

$$P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) = P(-3 \leq Z \leq 3) \cong 99{,}73\%$$

Quadro 1 – Características da distribuição Normal

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\} \text{, para} \quad x \in \Re \text{;} \quad \mu \in \Re \text{;} \quad \sigma > 0$$



Valor esperado ou média = μ = E(X) = $\int xf(x)dx$

Desvio-padrão = DP(X) = $\sigma = \sqrt{\int x^2 f(x)dx - \mu^2} = \sigma$

P(X<x) = $\int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\} dx = \phi(x)$

Quadro 2 – Características da distribuição Normal-padronizada (Z)

$$f(z) = \frac{1}{\sqrt{2\pi}} exp\left\{-\frac{x^2}{2}\right\} \text{ , para } \quad z \in \Re; \quad \mu = 0; \quad \sigma = 1$$

Valor esperado ou média = $\mu = 0$

Desvio-padrão = DP(X) = $\sigma = 1$

$$P(Z<z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} exp\left\{-\frac{z^2}{2}\right\} dz = \phi(z) \text{ (tabelada)}$$

4 O Teorema do Limite Central

O teorema do limite central pode ser enunciado de diferentes formas, com maior ou menor grau de rigor matemático. Para aulas de Probabilidade e Estatística direcionadas à maior parte dos cursos de graduação, o seguinte enunciado é suficiente:

Seja x_1 , ..., x_n , n variáveis aleatórias independentes e identicamente distribuídas com média μ e desvio-padrão σ. A distribuição da soma das n variáveis tende a apresentar um comportamento Normal, como segue:

$$X_1 + X_2 + ... + X_n = \sum X_i \sim Normal(n\mu; \sqrt{n\sigma 2})$$

Este resultado pode ser utilizado para compreender o comportamento probabilístico da média amostral $\overline{X}$ utilizando-se propriedades da Esperança e da Variância de variáveis aleatórias:

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} \sim Normal\left(\mu; \frac{\sigma}{\sqrt{n}}\right)$$

5 Atividades Práticas

Neste item são apresentadas três atividades práticas que podem ser apresentadas em disciplinas de Estatística ministradas em cursos de graduação.

Atividade 1 – O Teorema do Limite central na prática

É devido ao teorema do limite central (TLC) que a distribuição Normal tem grande destaque na Estatística. O TLC garante que o comportamento probabilístico de vários estimadores possa ser descrito com boa aproximação pela distribuição normal. Além disso, muitas variáveis encontradas na natureza tendem a apresentar comportamento normal como outra conseqüência visível deste teorema. Quando observamos variáveis na natureza ou em qualquer outro experimento observacional, um dado valor observado pode ser considerado como a resultante de um grande número de variáveis e suas interações, o que está em sintonia com o TLC.

Para exemplificar o teorema é utilizado o gerador de números pseudo-aleatórios da planilha Excel e algumas de suas funções. A primeira atividade consiste em gerar k = 10, 30, 120 e 250 colunas com 1000 observações cada uma utilizando a função ALEATÓRIO(), conforme mostrado na Figura 2. Para cada valor de k cria-se uma coluna adicional com a soma e/ou a média dos k valores. O recomendável e utilizar à média para se poder comparar os diferentes valores de *k.*

A função ALEATÓRIO() gera valores no intervalo [0;1] segundo uma distribuição uniforme. O valor esperado e a variância da Uniforme, neste intervalo, são iguais a 1/2 e 1/12, respectivamente. A média de $k \to \infty$ variáveis uniformes converge para uma normal de média *1/2* e variância igual a 1/12k. O aluno, dessa forma, pode comprovar que, de fato, os valores aleatórios gerados se comportam de acordo com o modelo.

A Tabela 1 resume os resultados da média e do desvio-padrão esperados e observados para diferentes valores de k. Os resultados comprovam o teorema que deixa assim de ser uma questão de fé para se tornar uma experiência vivenciada. Pode-se observar que a soma e/ou média de uniformes já é aproximadamente normal para valores a partir de dez como ilustram os histogramas da Figura 3.

Microsoft Excel - Pasta2

I989 =ALEATÓRIO()

	A	B	C	D	E	F	G	H	I	J	K	L
1	Var_1	Var_2	Var_3	Var_4	Var_5	Var_6	Var_7	Var_8	Var_9	Var_10	**Média**	
2	0,216	0,327	0,203	0,536	0,065	0,046	0,021	0,974	0,720	0,844	**0,395**	
3	0,472	0,367	0,354	0,095	0,477	0,897	0,200	0,679	0,260	0,423	**0,422**	
985	...	...	...	...	...	...	...	...	...	...	**...**	
986	...	...	...	...	...	...	...	...	...	...	**...**	
987	0,052	0,083	0,179	0,137	0,151	0,328	0,656	0,789	0,776	0,185	**0,334**	
988	0,488	0,651	0,528	0,509	0,723	0,839	0,042	0,815	0,604	0,198	**0,540**	
989	0,475	0,831	0,943	0,835	0,186	0,133	0,754	0,157	0,718	0,182	**0,521**	
990	0,042	0,983	0,873	0,352	0,956	0,937	0,949	0,217	0,506	0,736	**0,655**	
991	0,725	0,625	0,170	0,879	0,574	0,780	0,371	0,669	0,538	0,504	**0,584**	
992	0,077	0,029	0,967	0,270	0,568	0,864	0,639	0,318	0,410	0,194	**0,434**	
993	0,377	0,048	0,897	0,163	0,186	0,344	0,289	0,471	0,482	0,564	**0,382**	
994	0,832	0,178	0,172	0,601	0,881	0,487	0,020	0,485	0,911	0,818	**0,538**	
995	0,937	0,924	0,376	0,824	0,464	0,944	0,927	0,385	0,648	0,037	**0,646**	
996	0,459	0,902	0,387	0,672	0,345	0,301	0,221	0,541	0,804	0,456	**0,509**	
997	0,953	0,301	0,291	0,589	0,005	0,931	0,225	0,886	0,227	0,571	**0,498**	
998	0,558	0,928	0,676	0,233	0,810	0,071	0,693	0,795	0,319	0,450	**0,553**	
999	0,537	0,954	0,239	0,930	0,192	0,390	0,944	0,591	0,575	0,748	**0,610**	
1000	0,509	0,204	0,094	0,475	0,237	0,087	0,456	0,155	0,775	0,856	**0,385**	
1001	0,610	0,643	0,125	0,594	0,861	0,536	0,785	0,576	0,716	0,760	**0,621**	
1002												

Plan1 Plan2 Plan3

Pronto NUM

Figura 2 – Geração de variáveis uniformemente distribuídas na planilha

Tabela 1 – Valores da média e da variância esperados e observados para diferentes valores de k

k	Esperados		Observados	
	Esperança	Variância	Média	Variância
10	0,5	0,00833	0,497	0,00739
30	0,5	0,00278	0,500	0,00273
120	0,5	0,00069	0,500	0,00069
250	0,5	0,00033	0,500	0,00039

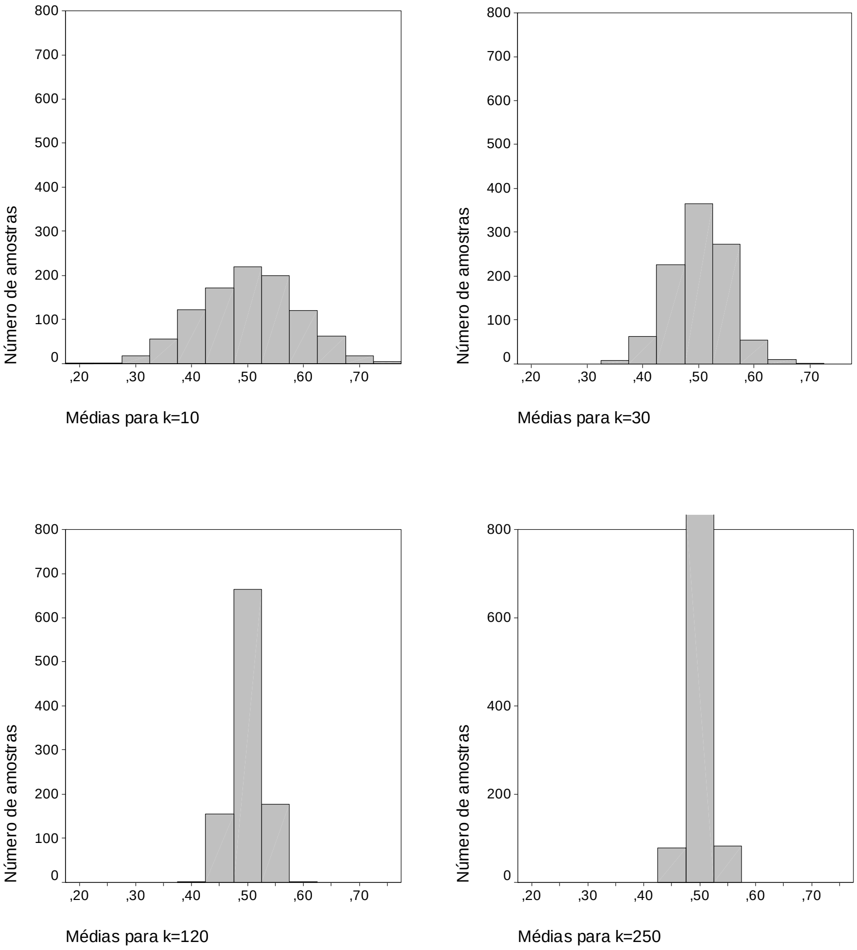


Figura 3 – Histogramas da média de k: = 10, 30, 120 e 250 variáveis uniformes

Atividade 2 – A relação média ± *desvios-padrão nos dados simulados*

A Atividade 1 consistiu de uma simulação na qual foi verificada que a convergência para a distribuição Normal ocorre rapidamente. Nesta segunda atividade o aluno deverá verificar se os valores gerados por simulação obedecem às probabilidades do modelo Normal. A função da planilha utilizada para contagem do número de observações é a CONT.SE.

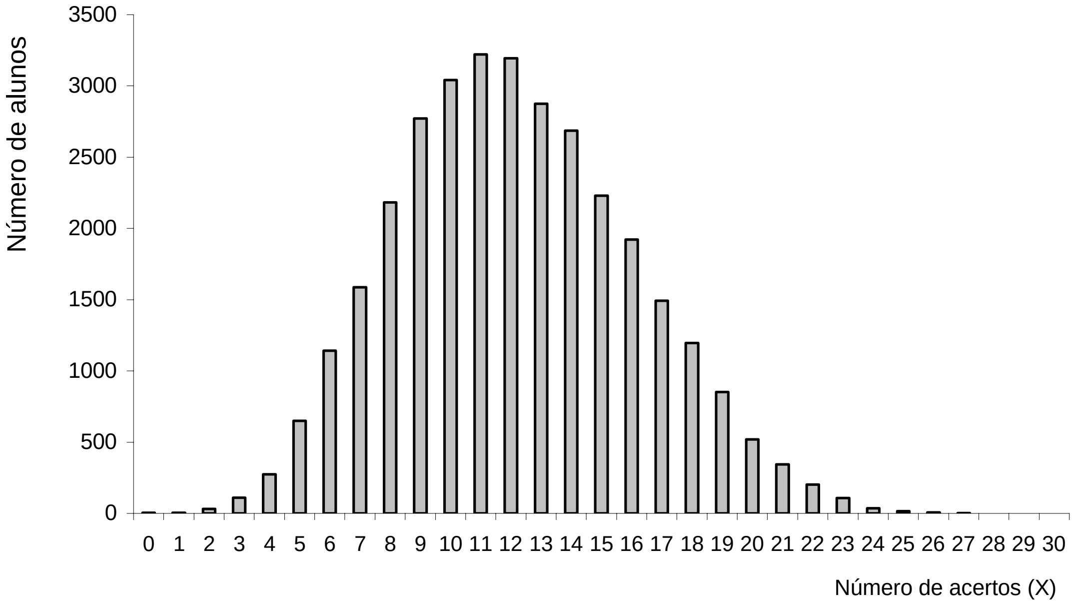
Neste caso, pode-se contar quantas observações estão acima da média mais *k* desvios-padrão e abaixo da média menos *k* desvios. Assim, como foram simulados 1000 valores, eliminam-se aqueles que estão fora do intervalo da média mais ou menos *k* desvios-padrão. A Tabela 2 mostra que os resultados simulados correspondem ao comportamento esperado para a curva Normal.

Tabela 2 – Percentual de valores simulados no intervalo média mais ou menos um, dois e três desvios padrão

k	Média	Desvio-padrão	Intervalo	Esperado (%)	Observado (%)
10	0,497	0,0860	média ± 1 desvio	68,27	671/1000 = 67,10
			média ± 2 desvios	95,44	937/1000 = 93,70
			média ± 3 desvios	99,74	997/1000 = 99,70
30	0,500	0,0523	média ± 1 desvio	68,27	679/1000 = 67,90
			média ± 2 desvios	95,44	951/1000 = 95,10
			média ± 3 desvios	99,74	995/1000 = 99,50
120	0,500	0,0263	média ± 1 desvio	68,27	681/1000 = 68,10
			média ± 2 desvios	95,44	959/1000 = 95,90
			média ± 3 desvios	99,74	998/1000 = 99,8
250	0,500	0,0198	média ± 1 desvio	68,27	683/1000 = 68,30
			média ± 2 desvios	95,44	950/1000 = 95,00
			média ± 3 desvios	99,74	997/1000 = 99,70

Atividade 3 – Calculando probabilidades utilizando dados de concurso vestibular

Nos últimos oito anos o número de acertos de algumas matérias do concurso vestibular da Universidade Federal do Rio Grande do Sul (UFRGS) vem apresentando um comportamento que pode ser aproximado pela distribuição Normal. Apesar do número de acertos ser claramente uma variável discreta, pode-se mostrar que o modelo Normal representa o fenômeno com fidelidade. A título de exemplo optou-se pela prova de Literatura do ano de 2006. A Figura 4 apresenta os resultados. Os alunos podem ter livre acesso a esses dados no *site* da universidade (ver fonte da Figura 4).



Fonte: http://www.ufrgs.br/cvresultados/histogramas/

Figura 4 – Gráfico de colunas simples do número de acertos na prova de literatura da UFRGS-2006 (N = 32710 candidatos)

O professor pode apresentar aos alunos a tabela de freqüências que originou o gráfico e solicitar aos alunos que encontrem a média e o desvio-padrão dos acertos utilizando os conhecimentos prévios de estatística descritiva. A Tabela 3 apresenta os resultados onde pode ser visto que a média é 12,16 acertos e o desvio-padrão é 3,94 acertos.

Tabela 3 – Distribuição de freqüências absolutas e relativas para o número de acertos na prova de Literatura da UFRGS, 2006

x_i	Número de Alunos (fi)	% do número de alunos (fri)	Acumulado do nº de alunos (Fi)	% acumulado do de alunos (Fri)
0	5	0,015	5	0,015
1	5	0,015	10	0,031
2	32	0,098	42	0,128
3	111	0,339	153	0,468
4	275	0,841	428	1,308
5	650	1,987	1078	3,296
6	1142	3,491	2220	6,787
7	1588	4,855	3808	11,642
8	2182	6,671	5990	18,312
9	2772	8,474	8762	26,787
10	3042	9,300	11804	36,087
11	3222	9,850	15026	45,937
12	3194	9,765	18220	55,702
13	2876	8,792	21096	64,494
14	2687	8,215	23783	72,709
15	2231	6,821	26014	79,529
16	1923	5,879	27937	85,408
17	1492	4,561	29429	89,969
18	1197	3,659	30626	93,629
19	851	2,602	31477	96,231
20	520	1,590	31997	97,820
21	345	1,055	32342	98,875
22	202	0,618	32544	99,493
23	107	0,327	32651	99,820
24	35	0,107	32686	99,927
25	16	0,049	32702	99,976
26	6	0,018	32708	99,994
27	2	0,006	32710	100,000
28	0	0,000	32710	100,000
29	0	0,000	32710	100,000
30	0	0,000	32710	100,000
Total	32710	100,000	-	-

Com base na média e desvio-padrão da prova o aluno pode calcular pela distribuição Normal a probabilidade de um aluno escolhido ao acaso ter nota superior a digamos 20 acertos. A partir da Tabela 3 pode calcular a proporção de candidatos com número de acertos superior a 20. O Quadro 3 apresenta os cálculos e a comparação dos resultados, onde é possível constatar que os resultados foram bastante semelhantes.

Quadro 3 – Cálculo de probabilidades supondo normalidade e comparação com os resultados da prova de literatura da UFRGS, 2006

Proporção real de alunos com 21 acertos ou mais: 713 / 32.710 ≅ 0,0218 = 2,18%

Proporção esperada de candidatos X ~ N(12,16; 3,94)

$$P(X > 20) = P\left(Z > \frac{20 - 12{,}16}{3{,}94}\right) = P(Z > 1{,}99) = 1 - \phi(1{,}99) \cong 0{,}0233 = 2{,}33\%$$

6 Considerações Finais

A distribuição Normal é lecionada nos cursos de graduação de todas as áreas que trabalham com investigação científica. Apesar de ser freqüente a grande quantidade de conteúdo para poucas horas-aula semanais, julgamos ser muito difícil o ensino da distribuição normal se for ignorado os conceitos básicos de Probabilidade. Sem um conhecimento prévio de variáveis aleatórias discretas e contínuas, esperança e variância, fica muito difícil conduzir aulas que realmente colaborem para o verdadeiro aprendizado do aluno.

Neste artigo estimula-se que professores de Probabilidade e Estatística usufruam os recursos da planilha com atividades que ilustrem conceitos probabilísticos e estatísticos, especialmente os recursos de simulação. A planilha Excel permite a geração de números pseudo-aleatórios que seguem uma distribuição Uniforme no intervalo [0;1]. A partir da Uniforme pode-se simular quaisquer outras distribuições, ilustrando de uma forma concreta conteúdos de outra forma ficariam apenas na memória de curto prazo e que poderiam ser esquecidos tão logo não fossem mais solicitados.

Com o recurso da planilha o aluno é o agente da sua aprendizagem e desta forma toma parte ativa na construção do seu conhecimento tornando dessa forma indelével.

Referências

BATANERO, Carmen, TAUBER, L. SÁNCHEZ, M. V. Significado y compreensión de la distribución normal en un curso introductorio de análisis de datos. Quadrante. v. 10, n. 1, p. 59-92, 2001.

BATANERO, Carmen, TAUBER, L., MEYER, R. From Data Analysis to Inference: A Research Project on Students' Understanding of The Normal Distribution. ISI 52 Session, Helsinki, Finland. 1999.

BATANERO, Carmen, TAUBER, L., SÁNCHEZ, M. V. Student's reasoning about the normal distribution. In D. Ben-Zvi y J.B. Garfield (Eds) The Challenge of Developing Statistical Literacy, Reasoning, and Thinking. Dordrecht: Kluwer. p. 257-76, 2004.

BAYER, A.; ECHEVESTE, S.; BITTENCOURT, H. R. ; ROCHA, J. . Preparação do formando em Matemática-licenciatura plena para lecionar Estatística no Ensino Fundamental e Médio. In: V Encontro Nacional de Pesquisa em Educação em Ciências, 2005, Bauru, SP. Anais... , 2005. v.

DAW, R. H., PEARSON, E. S. Studies in the history of probability and statistics: Abraham de Moivre's 1733 derivation of the normal curve: a bibliographical note, Biometrika 59 (1972), 677-680.

DOANE D. P. Using Simulation to Teach Distributions. Journal of Statistics Education. v. 12, n. 1, 2004. http://www.amstat.org/publications/jse/v12n1/doane.html

STIGLER, Stephen. M. Statistics on the Table. Cambridge (MA): Harvard University Press. 1999 (p. 404-30)

VIALI, L. Utilizando recursos computacionais (planilhas) no ensino do cálculo de probabilidades. In: Helena Noronha Cury. (Org.). Disciplinas Matemáticas em Cursos Superiores: reflexões, relatos, propostas.. 1 ed. Porto Alegre: EDIPUCRS, 2004, v. 13, p. 351-395.

VIALI, L. O ensino de probabilidade com recurso da planilha. In: V CIBEM (Congresso Ibero-Americano de Educação Matemática), 2005, Porto. ACTAS, 2005.