Probabilidade II
Aula 13

Junho de 2009

Mônica Barros, D.Sc.

Conteúdo

- O Teorema Central do Limite
- O Teorema de DeMoivre e Laplace
- Correção de Continuidade

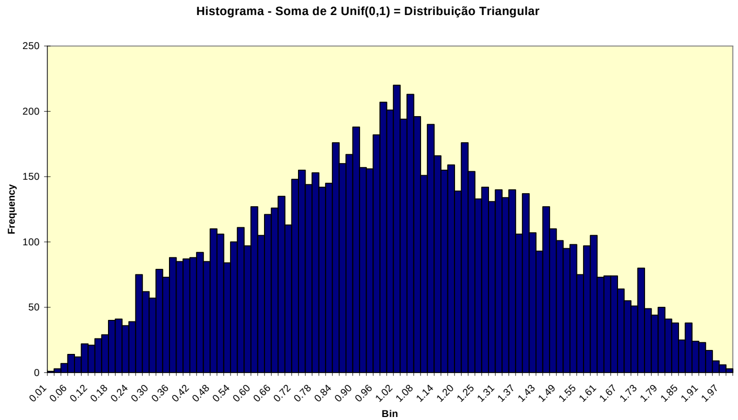
Teorema Central do Limite

- O Teorema Central do Limite é um resultado importantíssimo porque nos diz que:
  - Somas de variáveis independentes são aproximadamente Normais
  - Médias de variáveis independentes são aproximadamente Normais
- E o mais surpreendente é – NÃO INTERESSA qual a densidade das variáveis que estão sendo somadas!

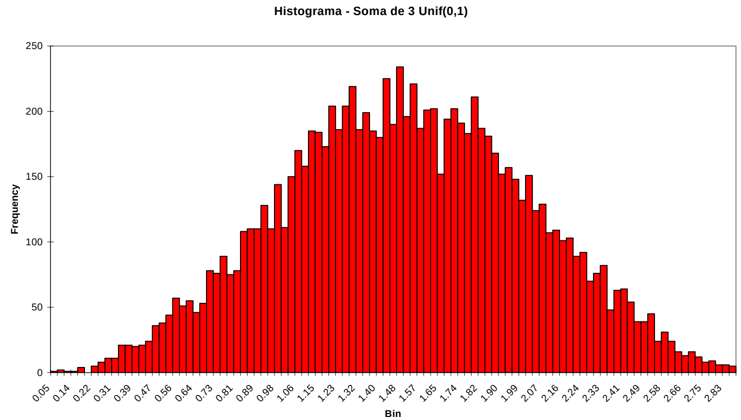
Teorema Central do Limite

- Para ver o TCL em ação, e se convencer de que ele funciona, veja os próximos gráficos.
- Simulamos diversas amostras de tamanho 10000 da Uniforme(0,1).
- Em seguida, criamos novas variáveis que são a soma destas Uniformes.
- Veja o resultado a seguir.

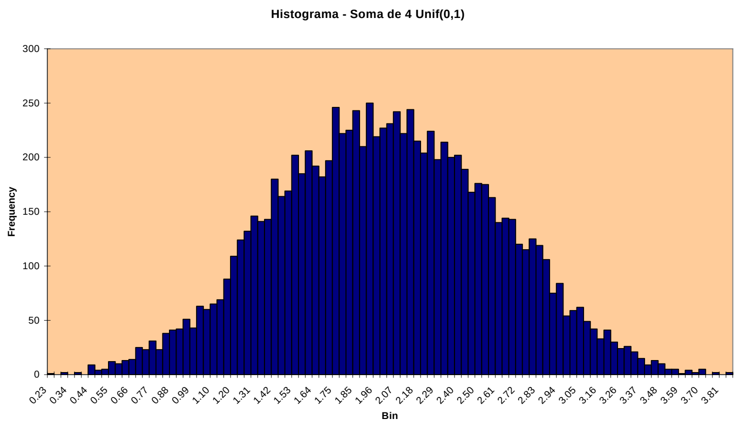
Soma de DUAS Unif(0,1)



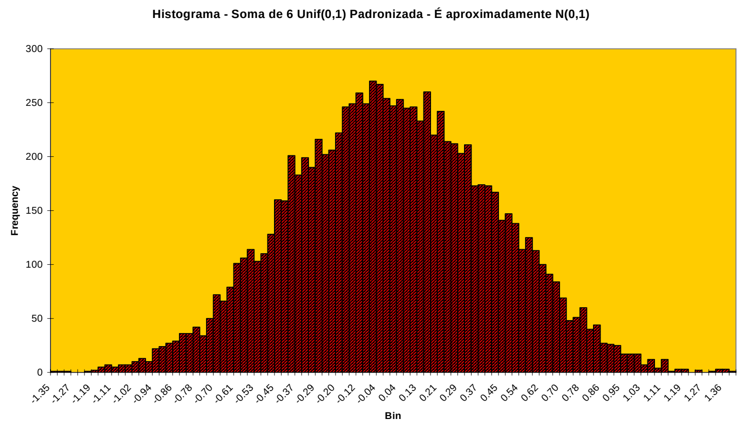
Soma de TRÊS Unif(0,1)



Soma de QUATRO Unif(0,1)



Soma de SEIS Unif(0,1)



Teorema Central do Limite

- O exemplo mostra que...
  - Quanto mais variáveis eu uso na soma, mais "Normal" parece o histograma....
  - E note que eu comecei com uma "coisa" absolutamente "não-Normal", a Uniforme!
  - De uma maneira MUITO informal, é isso mesmo o que o Teorema Central do Limite diz!

Teorema Central do Limite

- O teorema central do limite é um dos resultados fundamentais da teoria de probabilidades.
- **Por que ele é tão importante?**
- Em palavras, o teorema diz que **toda soma de variáveis aleatórias independentes é aproximadamente Normal**, desde que o número de termos na soma seja suficientemente grande.

Teorema Central do Limite

- Teorema
- Sejam $X_1, X_2, ..., X_n$ variáveis aleatórias independentes tais que $E(X_i) = \mu_i$ e $VAR(X_i) = \sigma_i^2$, ambas finitas. Seja $Y = X_1 + X_2 + ... + X_n$. Então, sob condições bastante gerais podemos afirmar que:

$$Z = \frac{Y - \sum_{i=1}^{n} \mu_i}{\sqrt{\sum_{i=1}^{n} \sigma_i^2}}$$

- Tem aproximadamente uma distribuição N(0,1).
- Esta aproximação torna-se cada vez melhor à medida que n cresce.
- A importância do teorema decorre do fato da distribuição dos X_i 's ser qualquer!!

Teorema Central do Limite

- O teorema central do limite pode ser escrito de maneira equivalente usando-se a média de variáveis aleatórias (ao invés da soma).
- Esta nova versão do teorema é especialmente importante quando os X_i 's são identicamente distribuídos, além de independentes, isto é, todos têm a mesma distribuição de probabilidade.

ENCE Escola Nacional de Ciências Estatísticas IBGE

Teorema Central do Limite

- Definição
  - Dizemos que $X_1, X_2, ..., X_n$ são identicamente distribuídos quando todos têm a mesma distribuição de probabilidade.
  - Neste caso, a média e a variância são as mesmas para todos os X_i, isto é, $E(X_i) = \mu$ e $VAR(X_i) = \sigma^2$ para i = 1, 2, ..., n.
  - Se, além de identicamente distribuídos, os X_i 's são independentes, dizemos que eles são i.i.d. (independentes e identicamente distribuídos).

ENCE Escola Nacional de Ciências Estatísticas IBGE

Teorema Central do Limite

- Teorema Central do Limite (versão iid)
- Sejam $X_1, X_2, ..., X_n$ variáveis aleatórias independentes e identicamente distribuídas (iid) tais que $E(X_i) = \mu$ e $VAR(X_i) = \sigma^2$, ambas finitas.
- Seja $Y = X_1 + X_2 + ... + X_n$. Então, sob condições bastante gerais:

$$Z = \frac{Y - n\mu}{\sqrt{n\sigma^2}}$$

- Tem aproximadamente uma distribuição N(0,1).
- Esta aproximação torna-se cada vez melhor à medida que n cresce.

ENCE Escola Nacional de Ciências Estatísticas IBGE

Teorema Central do Limite

- O teorema pode ser reescrito em termos da média amostral, $\overline{X}$, como a seguir.
- Teorema Central do Limite (versão iid em termos da média amostral)
- Sejam $X_1, X_2, ..., X_n$ variáveis aleatórias independentes e identicamente distribuídas (iid) tais que $E(X_i) = \mu$ e $VAR(X_i) = \sigma^2$, ambas finitas. Seja $\overline{X}$ a média amostral. Então:

$$Z = \frac{\overline{X} - \mu}{\sqrt{\sigma^2 / n}} = \sqrt{n}.\left(\frac{\overline{X} - \mu}{\sigma}\right) \text{ é aproximadamente N(0,1)}$$

- Como nos outros casos, a aproximação melhora com o aumento do tamanho da amostra.

ENCE Escola Nacional de Ciências Estatísticas IBGE

Teorema Central do Limite

- Pode-se provar que estas duas últimas versões do Teorema Central do Limite são totalmente equivalentes.
- Uso do teorema central do limite na prática:
  - o número de termos na soma é relativamente grande (n maior que 20 ou 30)
  - a distribuição dos X_i 's é "bem comportada", isto é, não é muito assimétrica. Se a distribuição de cada X_i é muito diferente de uma Normal geralmente é necessário aumentar n para garantir uma aproximação razoável.

Teorema Central do Limite

- Note que estamos definindo Y como a soma dos X_i's, e o teorema central do limite é aplicado a uma transformação linear de Y, e não ao próprio Y.
- Para que serve esta transformação?
- O seu objetivo é padronizar Y, transformando-o numa variável aleatória com média zero e variância 1. Esta nova variável tem a "cara" de uma N(0,1).

Teorema Central do Limite

- Como transformar uma variável aleatória com média μ e variância σ^2 numa variável com média 0 e variância 1?
- Suponha que W é uma variável aleatória qualquer com média E(W) = μ e variância VAR(W) = σ^2. Então:

$$Z = \frac{W - E(W)}{\sqrt{VAR(W)}}$$

- é uma variável com média zero e variância 1.

Teorema Central do Limite

- Este tipo de transformação foi usado para gerar uma variável Z com média zero e variância 1 a partir da soma (ou média) de X_1, X_2, ..., X_n.
- Esta nova variável Z tem, quando n é grande, aproximadamente a densidade N(0,1).

Teorema Central do Limite

- Exemplo
- Um teste padrão é aplicado a todos os formandos de Engenharia.
- Baseado nos testes aplicados num passado recente sabe-se que a nota no teste é uma variável aleatória com média 60 e desvio padrão 8.
- A nota média dos 100 alunos da Universidade XYZ que fizeram o teste neste ano foi 58.
- Existe evidência para sugerir que a nota dos alunos da Universidade XYZ foi pior que a dos outros? Para verificar se isso é verdade, calcule a probabilidade de que a média amostral seja menor ou igual a 58 para uma amostra de tamanho 100.

Teorema Central do Limite

- **Solução**
- **A nota na população de alunos de Engenharia é uma variável com média 60 e desvio padrão 8.**
- **Seja $\overline{X}$ a nota média dos alunos da Universidade XYZ baseada numa amostra de n = 100 alunos.**

Teorema Central do Limite

- **Desejamos calcular (aproximadamente, pois a distribuição das notas só tem média e variância conhecidas) Pr{ $\overline{X}$ ≤ 58}.**
- **Pelo Teorema Central do Limite, esta probabilidade pode ser aproximada usando-se uma distribuição Gaussiana, e então:**

$$\Pr\left\{\overline{X} \le 58\right\} \approx \Pr\left\{\frac{\overline{X} - 60}{\sqrt{64/100}} \le \frac{58-60}{\sqrt{64/100}}\right\} = \Pr\left\{Z \le -2.5\right\} = \Phi\left(-2.5\right) = 0.0062$$

Teorema Central do Limite

- **Exemplo**
- **Uma caixa de madeira contém 250 peças pequenas.**
- **O peso de cada peça é uma variável aleatória com média 0.20 Kg e desvio padrão 0.08 kg.**
- **20 caixas são colocadas num guindaste, que tem a capacidade de agüentar 1010 kg sem problemas. Acima deste peso recomenda-se o uso de outro equipamento.**
- **Qual a probabilidade da substituição por outro guindaste ser necessária?**

Teorema Central do Limite

- **Solução**
- **Esta é uma aplicação típica do teorema central do limite.**
- **A distribuição dos pesos das peças não foi especificada, sabemos apenas a sua média e variância.**
- **O teorema central do limite pode ser usado por que estamos interessados no peso total, isto é, na soma dos pesos individuais, que são supostos independentes e identicamente distribuídos.**

Teorema Central do Limite

- Seja X_i o peso em kg da i-ésima peça, onde i = 1, 2, ..., 5000. Então $Y = \Sigma X_i$ é o peso total das 20 caixas de peças (excluindo-se o peso das caixas propriamente ditas).

- Note que E(Y) = n.(0.2) = 5000(0.2) = 1000 kg e VAR(Y) = n.$(0.08)^2$ = 5000.$(0.08)^2$ e então o desvio padrão de Y é $0.08\sqrt{5000} \approx \sqrt{32}$ = 5.65 kg.

Teorema Central do Limite

- Seja:

$$Z = \frac{Y - 5000(0.20)}{(0.08)\sqrt{5000}} = \frac{Y - 1000}{(0.08).\sqrt{5000}} = 0.1768(Y - 1000)$$

- Pelo Teorema Central do Limite, Z é aproximadamente N(0,1). A probabilidade de interesse é:

$$\Pr(Y > 1010) = \Pr(Z > 0.1768(1010 - 1000)) =$$
$$= \Pr(Z > 1.768) = 1 - \Phi(1.768) = 1 - 0.9615 = 0.0385 = 3.85\%$$

Teorema Central do Limite (exemplo de simulação)

- O Teorema Central do Limite é usado na prática para gerar v.a. aproximadamente Normais a partir de v.a. Unif(0,1), conforme o algoritmo descrito a seguir.

- Sejam $X_1, X_2, ..., X_n$ iid Unif(0,1). Então cada X_i tem média 0.5 e variância 1/12.
- Seja Y a soma de 12 v.a. Unif(0,1). Então Y tem uma distribuição (não Uniforme) com média 12(0.5) = 6 e variância (12)*(1/12)) = 1.

Teorema Central do Limite (exemplo de simulação)

- Crie uma nova variável Z da seguinte maneira:

$$Z = (Y - 6) = \left(\sum_{i=1}^{12} X_i - 6\right)$$

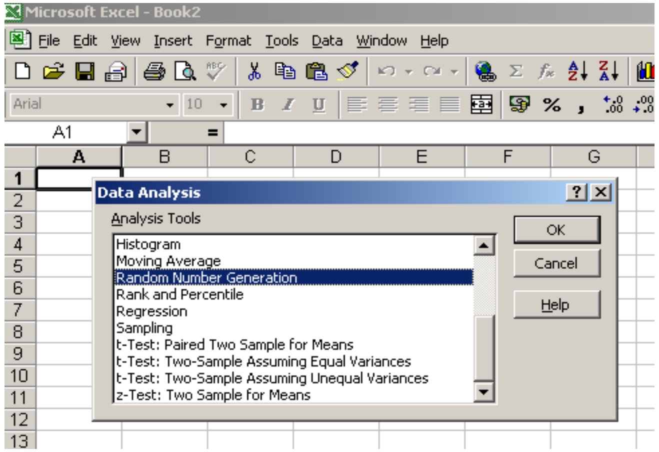
- Pelo TCL, a variável Z tem aproximadamente uma distribuição N(0,1).

- Será que isso funciona na prática? Veremos o resultado no Excel.

Teorema Central do Limite (exemplo de simulação)

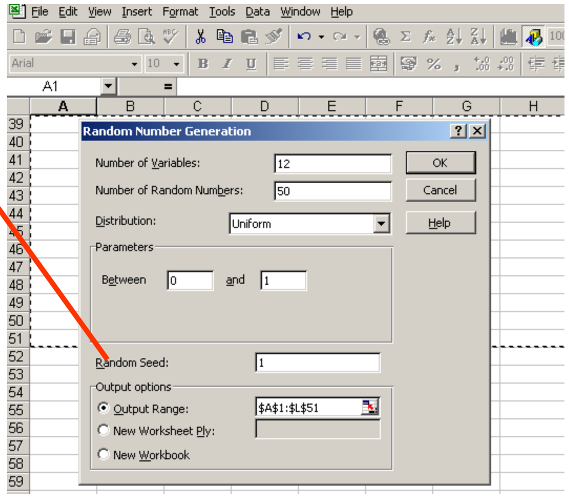
- A seguir geramos 12 colunas contendo, cada uma, 50 números aleatórios produzidos pelo procedimento de geração de números aleatórios existente no suplemento "Análise de dados".
- A 13a. Coluna da planilha conterá a soma das 12 anteriores (isto é, a variável Y), e a 14a. Coluna será a variável Z (Y padronizado).
- Se tudo der certo, o histograma de Z será uma curva em formato de sino, simétrica em torno de zero.

Teorema Central do Limite (exemplo de simulação)



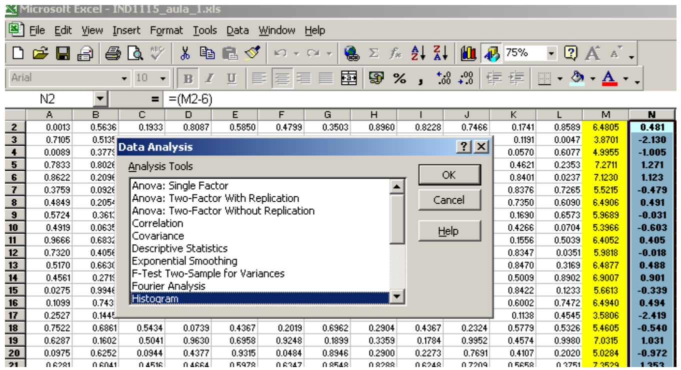
Teorema Central do Limite (exemplo de simulação)

Coloquei a semente do gerador de números aleatórios igual a um número determinado (1 neste caso) para poder repetir a geração dos MESMOS números em outra ocasião



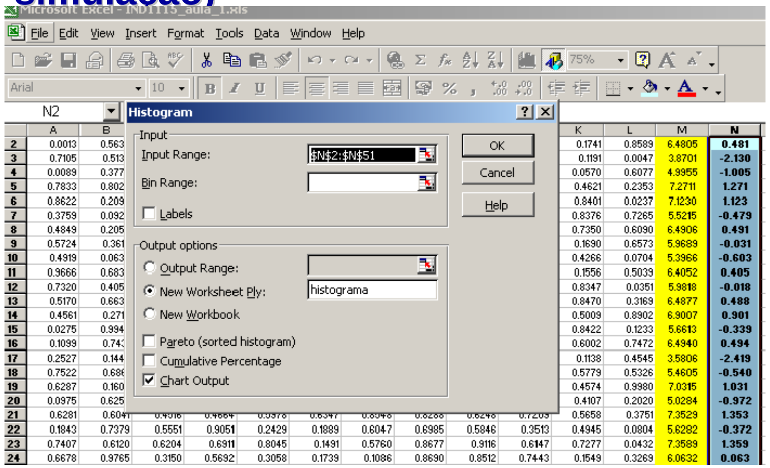
Teorema Central do Limite (exemplo de simulação)

- Procedimento para obter o histograma (também requer a instalação prévia da ferramenta de análise de dados)



ENCE Escola Nacional de Ciências Estatísticas IBGE

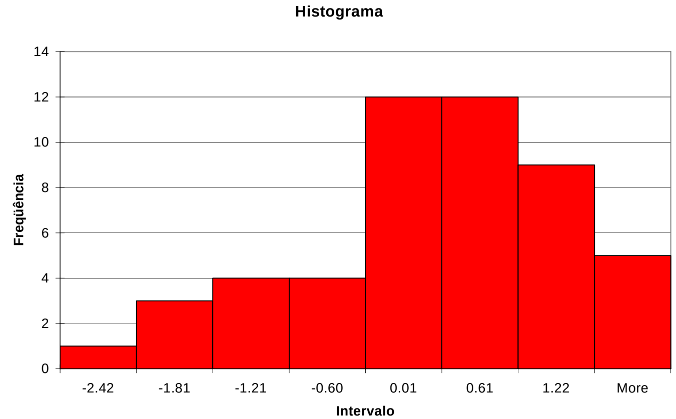
Teorema Central do Limite (exemplo de simulação)



ENCE Escola Nacional de Ciências Estatísticas IBGE

Teorema Central do Limite (exemplo de simulação)

- Histograma de Z



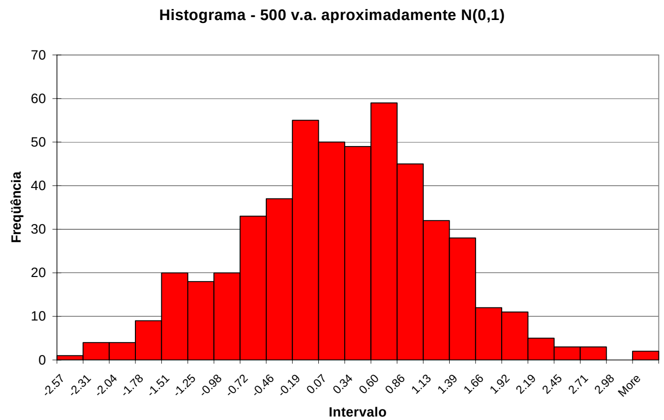
ENCE Escola Nacional de Ciências Estatísticas IBGE

Teorema Central do Limite (exemplo de simulação)

- O histograma não ficou perfeito, mas já tem uma "cara" de distribuição Normal (embora ainda um pouco assimétrico).
- Tente repetir este exemplo gerando 500 variáveis aproximadamente Normais. Como fica o histograma?
- No próximo gráfico está o resultado de uma simulação que eu fiz usando 500 v.a.

ENCE Escola Nacional de Ciências Estatísticas IBGE

Teorema Central do Limite (exemplo de simulação)



Teorema Central do Limite

- O resultado do próximo exemplo pode parecer surpreendente à primeira vista (mas não é, se você entendeu o Teorema Central do Limite!).
- Imagine que você peça a 100 pessoas para jogarem dados 5 vezes, e que elas calculem a média nas 5 jogadas. Qual será a "cara" da média?
- Como você já sabe, a média é aproximadamente Normal (pelo TCL).

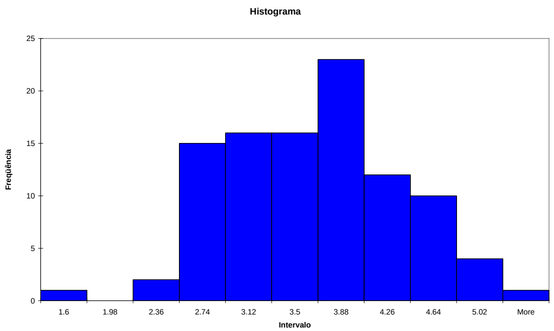
Teorema Central do Limite

- Abaixo estão parte dos resultados de uma simulação...

PESSOA	Jogada de 5 dados por 100 pessoas					média
1	4	1	4	2	5	3.2
2	3	6	3	6	1	3.8
3	2	2	5	3	2	2.8
4	5	2	6	6	6	5.0
5	6	1	1	2	4	2.8
6	6	1	5	2	3	3.4
7	2	5	3	2	6	3.6
8	3	1	1	3	5	2.6
9	4	3	4	6	4	4.2
10	4	1	2	5	3	3.0
11	6	6	1	6	3	4.4
12	6	3	6	2	1	3.6
13	3	4	4	4	3	3.6
14	6	6	3	1	6	4.4
15	6	6	4	2	4	4.4
16	1	4	6	6	1	3.6
17	2	6	4	5	1	3.6
18	2	2	3	2	6	3.0
19	5	3	1	4	1	2.8
20	4	5	2	3	4	3.6

Teorema Central do Limite

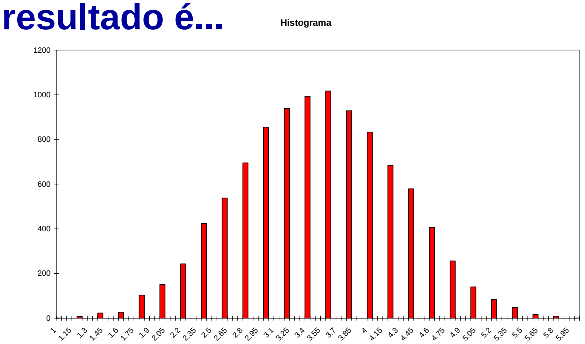
- Se você fizer o histograma da coluna marcada "Média" você encontrará...



Pode-se até argumentar que o gráfico não parece "muito" Normal, mas as observações originais (jogadas do dado) eram Uniformes nos números 1, 2,3, 4, 5, 6, o que é bem diferente disso!

Teorema Central do Limite

- Se você agora simula 10 mil pessoas jogando dados e calculando a média, o resultado é...



Teorema Central do Limite (para casa)

- Um computador, ao adicionar números, arredonda cada número para o inteiro mais próximo. Suponha que os erros de arredondamento são independentes e identicamente distribuídos com densidade Uniforme no intervalo (-0.5, +0.5).
- Suponha que 144 números são somados. Qual a probabilidade de que o módulo da soma dos erros ultrapasse 2?
- *Dica:* Use o teorema central do limite. Qual é a média e a variância de cada um dos 144 erros de arredondamento?

Teorema Central do Limite (para casa)

- O saldo devedor médio dos usuários de um certo cartão de crédito é uma variável aleatória com média R$ 112 e desvio padrão R$ 56.
- Toma-se uma amostra de 50 portadores do cartão. Qual a probabilidade do saldo devedor médio na amostra estar entre R$ 100 e R$ 130 ?

Teorema Central do Limite (para casa)

- Você está encarregado da compra de champagne para uma festa de reveillon. Sabe-se que, na média, uma pessoa consome 360 ml de champagne, com um desvio padrão de 90ml. Existem 50 convidados na sua festa.
- a) Você comprou 22 garrafas de champagne, cada uma com 750 ml. Qual a probabilidade da champagne não ser suficiente, isto é, do consumo total exceder as 22 garrafas?
- b) Qual a quantidade (em ml) de champagne que você deve comprar para ter certeza que o consumo total só vai exceder esta quantidade 5% do tempo? E 10% do tempo?

Teorema de DeMoivre e Laplace

- O Teorema de DeMoivre e Laplace - aproximação da Binomial pela Normal
- Este teorema é **apenas um caso particular do teorema central do limite**, pois uma variável aleatória com distribuição Binomial pode ser encarada como a soma de n variáveis Bernoulli(p) independentes.
- Na verdade, o teorema de DeMoivre e Laplace foi formulado muito antes do teorema central do limite.

Teorema de DeMoivre e Laplace

- Sabemos que a distribuição Binomial com parâmetros n e p pode ser aproximada pela distribuição Poisson quando n é grande e p é pequeno.
- A aproximação da Binomial pela Normal (Teorema de DeMoivre e Laplace) "funciona" quando n é grande e p não é necessariamente pequeno (por exemplo, p próximo de 1/2).

Teorema de DeMoivre e Laplace

- O fato da distribuição Binomial ser discreta e da Normal ser contínua cria a necessidade de corrigir os valores das probabilidades encontrados via DeMoivre e Laplace.
- Esta correção será chamada de correção de continuidade.

Teorema de DeMoivre e Laplace

- A correção de continuidade permite uma melhora significativa na aproximação dada pelo teorema de DeMoivre e Laplace.
- Teorema
- Seja Y ~ Bin(n, p) onde n é "grande" e p não está próximo de zero. Então:

$$Z = \frac{Y - E(Y)}{\sqrt{VAR(Y)}} = \frac{Y - np}{\sqrt{np(1-p)}}$$

- tem aproximadamente uma distribuição N(0,1).

Teorema de DeMoivre e Laplace

- Em particular, podemos calcular probabilidades de valores específicos da variável Y usando a tabela da distribuição Normal padrão.
- Este resultado segue diretamente do Teorema Central do Limite, pois Y pode ser pensado como a soma de n variáveis aleatórias iid, cada uma com distribuição Bernoulli(p).

Teorema de DeMoivre e Laplace

- ***Nota***
- **A aproximação Normal fornecida pelo Teorema de DeMoivre e Laplace funciona bem se n > 10 e p está próximo de 1/2.**
- **Para outros valores de p é necessário aumentar o valor de n.**
- **Em geral o teorema de DeMoivre e Laplace produz resultados aceitáveis se n.p > 5 quando p < 1/2 e se n.(1-p) > 5 quando p > 1/2.**

Teorema de DeMoivre e Laplace

- **Exemplo**
- **Suponha que desejamos calcular Pr(Y ≤ y), onde Y ~ Bin(n, p). Como poderemos fazê-lo usando o teorema de DeMoivre e Laplace?**

$$\Pr(Y \le y) = \Pr\left(\frac{Y-np}{\sqrt{npq}} \le \frac{y-np}{\sqrt{npq}}\right) = \Pr\left(Z \le \frac{y-np}{\sqrt{npq}}\right) \approx \Phi\left(\frac{y-np}{\sqrt{npq}}\right)$$

- **O resultado desta aproximação pode ser sensivelmente melhorado com a correção de continuidade indicada a seguir.**

Teorema de DeMoivre e Laplace

- **Correção de Continuidade**

Quantidade desejada na distribuição Binomial	Quantidade Calculada através da correção de continuidade	Expressão aproximada usando a densidade Normal
Pr(Y = y)	Pr(y - 0.5 ≤ Y ≤ y + 0.5)	$\Phi\left(\frac{y+0.5-np}{\sqrt{npq}}\right) - \Phi\left(\frac{y-0.5-np}{\sqrt{npq}}\right)$
Pr(Y ≤ y)	Pr(Y ≤ y +0.5)	$\Phi\left(\frac{y+0.5-np}{\sqrt{npq}}\right)$
Pr(Y < y) = Pr(Y ≤ y-1)	Pr(Y ≤ y - 1 + 0.5)	$\Phi\left(\frac{y-1+0.5-np}{\sqrt{npq}}\right) = \Phi\left(\frac{y-0.5-np}{\sqrt{npq}}\right)$
Pr(Y ≥ y)	Pr(Y ≥ y - 0.5)	$1-\Phi\left(\frac{y-0.5-np}{\sqrt{npq}}\right)$
Pr(Y > y) = Pr(Y ≥ y +1)	Pr(Y ≥ y + 1 - 0.5)	$1-\Phi\left(\frac{y+1-0.5-np}{\sqrt{npq}}\right)$
Pr(a ≤ Y ≤ b)	Pr(a - 0.5 ≤ Y ≤ b + 0.5)	$\Phi\left(\frac{b+0.5-np}{\sqrt{npq}}\right) - \Phi\left(\frac{a-0.5-np}{\sqrt{npq}}\right)$

Teorema de DeMoivre e Laplace

- **Exemplo**
- **Uma fábrica produz amortecedores para carros, dos quais 20% são defeituosos.**
- **Uma amostra de 100 amortecedores é selecionada diariamente. Seja X o número de amortecedores defeituosos na amostra.**
- **Calcule a probabilidade de X ser menor ou igual a 15 usando a aproximação Normal com correção de continuidade.**

ENCE Escola Nacional de Ciências Estatísticas IBGE

Teorema de DeMoivre e Laplace

- **Solução**
- **Neste caso X ~ Bin(100, 0.2) e então E(X) = np = 20, VAR(X) = npq = 20(0.8) = 16. Usando a correção de continuidade:**

$$\Pr(X \le 15) \approx \Phi\left(\frac{15+0.5-np}{\sqrt{npq}}\right) = \Phi\left(\frac{15.5-20}{\sqrt{16}}\right) = \Phi\left(\frac{-4.5}{4}\right) =$$
$$= \Phi(-1.125) = 0.1292$$

- **A aproximação sem a correção de continuidade seria:** $\Pr(X \le 15) \approx \Phi\left(\frac{15-20}{\sqrt{16}}\right) = \Phi\left(\frac{-5}{4}\right) = \Phi(-1.25) = 0.1056$

ENCE Escola Nacional de Ciências Estatísticas IBGE

Teorema de DeMoivre e Laplace

- **O valor exato desta probabilidade seria calculado através da fórmula:**

$$\Pr(X \le 15) = \sum_{x=0}^{15} \binom{100}{x} .(0.2)^x .(0.8)^{100-x} = 0.1285$$

- **O resultado com a correção de continuidade é muito mais próximo do exato do que o cálculo sem a correção.**
- **A correção de continuidade serve para melhorar a "performance" da aproximação da Binomial pela Normal.**

ENCE Escola Nacional de Ciências Estatísticas IBGE

Teorema de DeMoivre e Laplace

- **Além disso, se desejamos aproximar através de DeMoivre e Laplace a probabilidade da variável Binomial ser igual a um número qualquer, isto só pode ser feito usando a correção de continuidade.**
- **Por que? A distribuição Normal é contínua, e portanto a probabilidade de qualquer valor é zero, e assim precisamos encontrar a probabilidade de um intervalo em torno do ponto desejado.**

ENCE Escola Nacional de Ciências Estatísticas IBGE

Teorema de DeMoivre e Laplace

- **Exemplo**
- **Seja X ~ Bin(20, 1/2). Calcule a probabilidade dos valores 5, 6, ...10 diretamente (usando a própria distribuição Binomial) e usando a aproximação Normal com correção de continuidade. Compare os resultados.**
- **Solução**
- **O cálculo exato é dado pela fórmula:**

$$\Pr(X = x) = \binom{20}{x} . \left(\frac{1}{2}\right)^x . \left(\frac{1}{2}\right)^{20-x} = \binom{20}{x} . \left(\frac{1}{2}\right)^{20}$$

Teorema de DeMoivre e Laplace

- O cálculo aproximado com correção de continuidade é:

$$\Pr(X = x) \approx \Phi\left(\frac{x+0.5-np}{\sqrt{npq}}\right) - \Phi\left(\frac{x-0.5-np}{\sqrt{npq}}\right) =$$

$$= \Phi\left(\frac{x+0.5-10}{\sqrt{5}}\right) - \Phi\left(\frac{x-0.5-10}{\sqrt{5}}\right)$$

- A próxima tabela compara os resultados exato e aproximado:

x	Pr(X =x) exata	Pr(X =x) aproximada
5	0.01479	0.01513
6	0.03696	0.03668
7	0.07393	0.07301
8	0.12013	0.11939
9	0.16018	0.16036
10	0.17620	0.17694

Teorema de DeMoivre e Laplace

- **Exemplo**
- No exemplo anterior calcule Pr(5 ≤ X ≤ 9) de 3 maneiras:
  - Exatamente,
  - Aproximadamente com correção de continuidade,
  - Aproximadamente sem correção de continuidade.
- **Solução**

  Cálculo exato

$$\Pr(5 \le X \le 9) = \Pr(X = 5) + \Pr(X = 6) + \ldots + \Pr(X = 9) = 0.40599$$

Teorema de DeMoivre e Laplace

- Aproximação com correção de continuidade

$$\Pr(5 \le X \le 9) \approx \Pr(4.5 \le X \le 9.5) = \Phi\left(\frac{9.5-np}{\sqrt{npq}}\right) - \Phi\left(\frac{4.5-np}{\sqrt{npq}}\right) =$$

$$= \Phi\left(\frac{9.5-10}{\sqrt{5}}\right) - \Phi\left(\frac{4.5-10}{\sqrt{5}}\right) = 0.41153 - 0.00695 = 0.40458$$

- Aproximação sem correção de continuidade

$$\Pr(5 \le X \le 9) \approx \Phi\left(\frac{9-np}{\sqrt{npq}}\right) - \Phi\left(\frac{5-np}{\sqrt{npq}}\right) =$$

$$= \Phi\left(\frac{9-10}{\sqrt{5}}\right) - \Phi\left(\frac{5-10}{\sqrt{5}}\right) = 0.32736 - 0.01267 = 0.31469$$

- Que é um resultado bem pior que a aproximação com correção de continuidade.

Teorema de DeMoivre e Laplace

- O Teorema de DeMoivre e Laplace pode ser combinado com a Lei Fraca dos Grandes Números para fornecer uma estimativa mais refinada do tamanho de amostra necessário.

- Suponha que $X_1, X_2, ..., X_n$ são iid Bernoulli(p) e $Y = X_1 + X_2 + ... + X_n$. Então Y é Bin(n, p).

ENCE Escola Nacional de Ciências Estatísticas IBGE

Teorema de DeMoivre e Laplace

- Pelo Teorema de DeMoivre e Laplace:

$$\frac{Y-np}{\sqrt{np(1-p)}}=\frac{n.\hat{p}-np}{\sqrt{np(1-p)}}=\frac{\sqrt{n}\left(\hat{p}-p\right)}{\sqrt{p(1-p)}} \text{ é aprox. N(0,1)}$$

- Assim:

$$\Pr\left(\left|\frac{\sqrt{n}(\hat{p}-p)}{\sqrt{p(1-p)}}\right|\geq c\right)=\Pr\left(|\hat{p}-p|\geq c\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right)\approx 2\{1-\Phi(c)\}$$

$$\text{Seja } k=\frac{c\sqrt{p(1-p)}}{\sqrt{n}}\Rightarrow c=\frac{k\sqrt{n}}{\sqrt{p(1-p)}}$$

ENCE Escola Nacional de Ciências Estatísticas IBGE

Teorema de DeMoivre e Laplace

- Exemplo
- Uma amostra de tamanho n é coletada para determinar a porcentagem de eleitores a favor de um certo candidato.
- Seja X_i = 1 se o i-ésimo eleitor vota no candidato, e X_i = 0 se ele não vota no candidato.
- Então $Y = X_1 + X_2 + ... + X_n$ é o número de eleitores na amostra que votam no candidato.

ENCE Escola Nacional de Ciências Estatísticas IBGE

Teorema de DeMoivre e Laplace

- Seja p a probabilidade de um eleitor votar no candidato.
- Note que p^= Y/n é a proporção de eleitores na amostra que votam no candidato.
- Então E(p^) = p e VAR(p^) = p.(1-p)/n
- Suponha que o valor real de p é suficientemente próximo de ½ para que p(1-p) possa ser aproximado por ¼.

ENCE Escola Nacional de Ciências Estatísticas IBGE

Teorema de DeMoivre e Laplace

- p^ pode ser usado para estimar p.
- 1) Suponha que n = 900. Ache:

  $\Pr\left(|\hat{p}-p|\geq 0.025\right)$
- Dos slides anteriores:

$$\Pr\left(\left|\frac{\sqrt{n}(\hat{p}-p)}{\sqrt{p(1-p)}}\right|\geq c\right)=\Pr\left(|\hat{p}-p|\geq k\right)\approx 2\{1-\Phi(c)\}$$

$$c\frac{\sqrt{p(1-p)}}{\sqrt{n}}=c\frac{1}{2\sqrt{900}}=\frac{c}{60}=0.025\Rightarrow c=1.50\Rightarrow \Pr\left(|\hat{p}-p|\geq 0.025\right)\approx 2\{1-\Phi(1.5)\}=2(0.067)=0.134$$

Teorema de DeMoivre e Laplace

- 2) Suponha que n = 900. Ache c tal que:

$$\Pr\left(\left|\hat{p}-p\right|\geq 0.025\right)=0.01$$

- Aqui:

$$2\{1-\Phi(c)\}=0.01 \Rightarrow \Phi(c)=0.995 \Rightarrow c=2.58$$

- Então:

$$c=2.58=\frac{k\sqrt{n}}{\sqrt{p(1-p)}}=\frac{k\sqrt{900}}{1/2}=2k(30)\Rightarrow 60k=2.58\Rightarrow k=0.043$$

Teorema de DeMoivre e Laplace

- 3) Ache n tal que:

$$\Pr\left(\left|\hat{p}-p\right|\geq 0.025\right)=0.01$$

- Da mesma forma que no item 2), c = 2.58

$$c=2.58=\frac{k\sqrt{n}}{\sqrt{p(1-p)}}=\frac{k\sqrt{n}}{1/2}=2k(\sqrt{n})$$

- Também, k = 0.025 e substituindo acima:

$$c=2.58=2(0.025)(\sqrt{n})\Rightarrow \sqrt{n}=\frac{2.58}{0.05}=51.6\Rightarrow n=2663 \text{ aproximadamente}$$

Teorema de DeMoivre e Laplace

- Na verdade o item 3 deste exemplo é típico e responde a pergunta: qual o tamanho de amostra necessário para que a diferença entre p e seu estimador ("chute") baseado na amostra esteja dentro de certo limite k com uma probabilidade especificada.

Teorema de DeMoivre e Laplace (para casa)

- A probabilidade de uma pessoa com mais de 65 anos pegar uma gripe no outono é 75%. Toma-se uma amostra de 60 pessoas na "3a. idade" e seja X o número destas pessoas com gripe na amostra. Calcule as seguintes probabilidades:
  - Pr{X ≥ 40} aproximadamente pelo teorema de DeMoivre e Laplace *com* correção de continuidade.
  - Pr{X ≥ 40} exatamente (só se você tiver acesso a um computador, pois do contrário será bastante trabalhoso!)
  - Pr{45 ≤ X ≤ 50} aproximadamente pelo teorema de DeMoivre e Laplace *com* correção de continuidade.

Teorema de DeMoivre e Laplace (para casa)

- Seja Y ~ Bin(12 , 1/2).
- Calcule $\Pr(Y \geq 9)$ exatamente.
- Calcule $\Pr(Y \geq 9)$ aproximadamente pelo teorema de DeMoivre e Laplace com correção de continuidade.
- Calcule $\Pr(Y = 7)$ exatamente.
- Calcule $\Pr(Y= 7)$ aproximadamente pelo teorema de DeMoivre e Laplace com correção de continuidade.