



# IND 1115

## Inferência Estatística

### Aula 7

Setembro 2004

Mônica Barros

monica@mbarros.com

1



## Conteúdo

- ❑ Por que a revisão de probabilidades até agora?
- ❑ A importância da distribuição Normal
- ❑ O Teorema Central do Limite

monica@mbarros.com

2

## Por que uma revisão de Probabilidades?



- ❑ Ao longo deste curso estaremos desenvolvendo inúmeros resultados e testes baseados na distribuição Normal.
- ❑ A questão importante é: por que esta ênfase na Normal? O que leva a isso? A resposta vem dos resultados que iremos demonstrar nesta parte inicial do curso.
- ❑ O objetivo nesta aula é apresentar o Teorema Central do Limite, que justifica o uso tão intenso da distribuição Normal.

monica@mbarros.com

3



## 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.

monica@mbarros.com

4

## Teorema Central do Limite



- Teorema
- Sejam  $X_1, X_2, \dots, X_n$  variáveis aleatórias independentes tais que  $E(X_i) = \mu_i$  e  $\text{VAR}(X_i) = \sigma_i^2$ , ambas finitas. Seja  $Y = X_1 + X_2 + \dots + X_n$ . 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!!**

monica@mbarros.com

5

## 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.
- Definição
- Dizemos que  $X_1, X_2, \dots, X_n$  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  $\text{VAR}(X_i) = \sigma^2$  para  $i = 1, 2, \dots, n$ .

monica@mbarros.com

6

## Teorema Central do Limite



- Se, além de identicamente distribuídos, os  $X_i$ 's são independentes, dizemos que eles são **i.i.d.** (independentes e identicamente distribuídos).
- Teorema Central do Limite (versão iid)
- Sejam  $X_1, X_2, \dots, X_n$  variáveis aleatórias independentes e identicamente distribuídas (iid) tais que  $E(X_i) = \mu$  e  $\text{VAR}(X_i) = \sigma^2$ , ambas finitas. Seja  $Y = X_1 + X_2 + \dots + X_n$ . 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.

monica@mbarros.com

7

## Teorema Central do Limite



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

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

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

monica@mbarros.com

8

## 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  $\mu$  e variância  $\sigma^2$  numa variável com média 0 e variância 1?
- ❑ Suponha que W é uma variável aleatória qualquer com média  $E(W) = \mu$  e variância  $VAR(W) = \sigma^2$ . 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  $X_1, X_2, \dots, X_n$ .
- ❑ 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  $\bar{X}$  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)  $\Pr\{\bar{X} \leq 58\}$ .
- ❑ Pelo Teorema Central do Limite, esta probabilidade pode ser aproximada usando-se uma distribuição Gaussiana, e então:

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

## 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  $i = 1, 2, \dots, 5000$ . Então  $Y = \sum X_i$  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:

$$\begin{aligned} \Pr(Y > 1010) &= \Pr(Z > 0.1768(1010 - 1000)) = \\ &= \Pr(Z > 1.768) = 1 - \Phi(1.768) = 1 - 0.9615 = 0.0385 = 3.85\% \end{aligned}$$

## 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  $X_1, X_2, \dots, X_n$  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  $(12)^2(1/12) = 12$ .

## Teorema Central do Limite (exemplo de simulação)



- ❑ Crie uma nova variável Z da seguinte maneira:

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

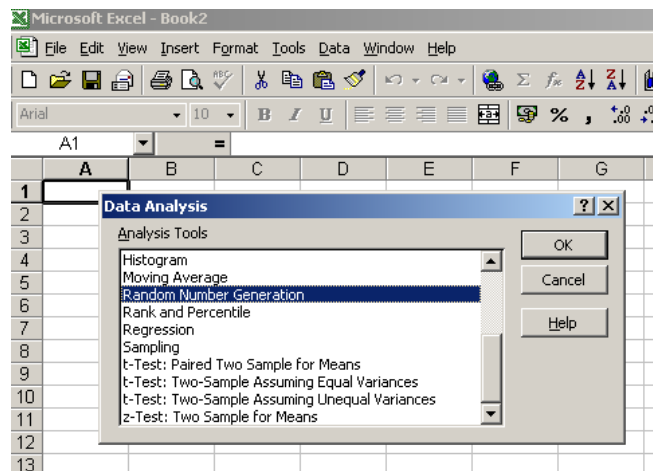
- ❑ Pelo CLT, a variável Z tem aproximadamente uma distribuição  $N(0,1)$ .
- ❑ **Será que isso funciona na prática? Veremos o resultado no Excel.**

## 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.

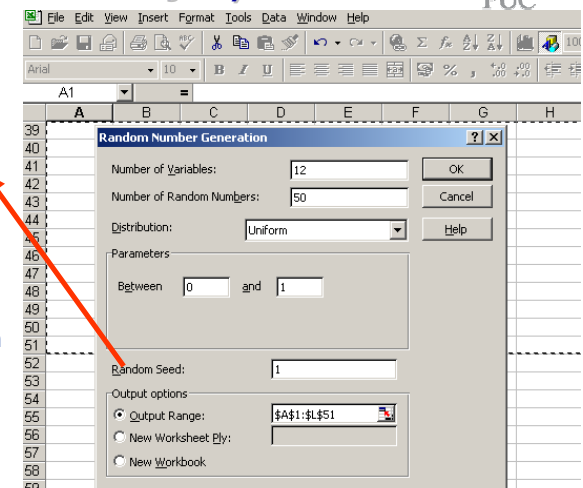
## Teorema Central do Limite (exemplo de simulação)



## Teorema Central do Limite (exemplo de simulação)



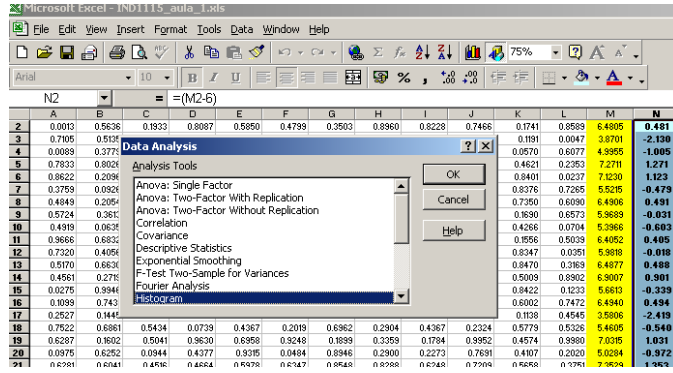
Coloquei a semente do gerador de números aleatórios igual a um número determinado (1 neste caso) para poder repetir a geração dos MESMOS números em outra ocasião



## Teorema Central do Limite (exemplo de simulação)



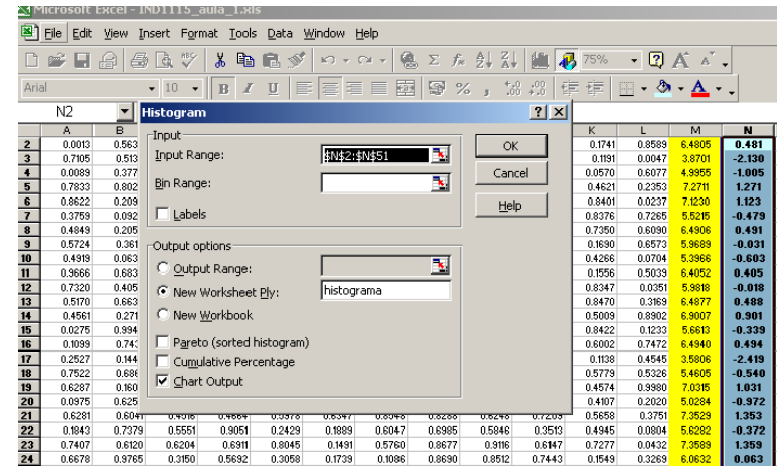
- Procedimento para obter o histograma (também requer a instalação prévia da ferramenta de análise de dados)



monica@mbarros.com

25

## Teorema Central do Limite (exemplo de simulação)



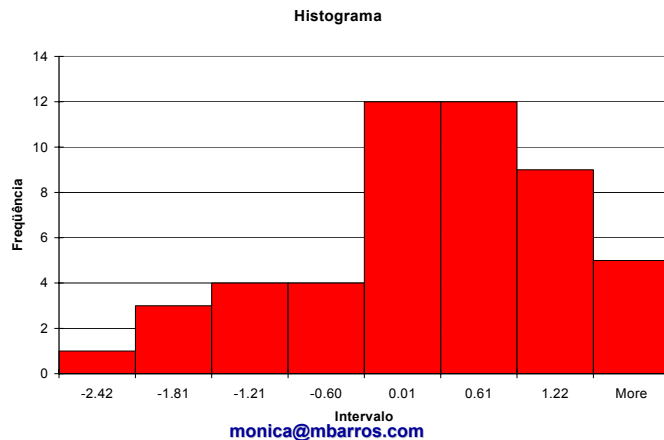
monica@mbarros.com

26

## Teorema Central do Limite (exemplo de simulação)



- Histograma de Z



monica@mbarros.com

27

## 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.

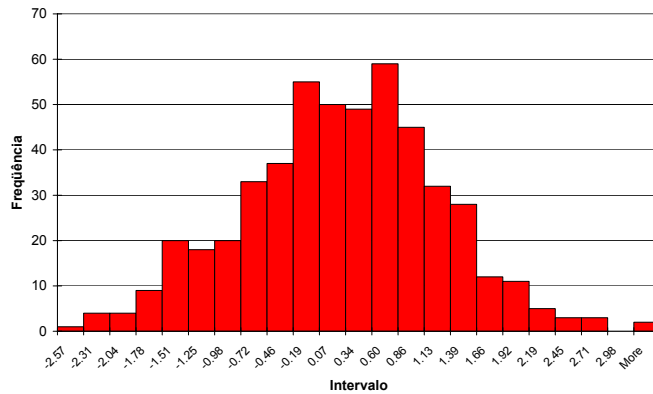
monica@mbarros.com

28

## Teorema Central do Limite (exemplo de simulação)



Histograma - 500 v.a. aproximadamente  $N(0,1)$



monica@mbarros.com

29

## 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?

monica@mbarros.com

30

## 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 ?

monica@mbarros.com

31

## 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?

monica@mbarros.com

32