Pesquisa Quantitativa

Lupércio França Bessegato
Mestrado em Administração/UFJF

Modelos Probabilísticos

Roteiro

1. Introdução
2. Amostragem
3. **Modelos probabilísticos**
4. Distribuições amostrais e estimação
5. Testes de hipóteses
6. Comparações de médias e variâncias
7. Análise de dados categóricos
8. Análise de regressão
9. Referências

Modelos Probabilísticos – Roteiro

2. Modelos probabilísticos:
   a) Aleatoriedade
   b) Modelos probabilísticos
   c) Outros modelos probabilísticos
   d) Distribuição da média amostral
   e) Referências

Aleatoriedade

- **Você acredita em destino?**

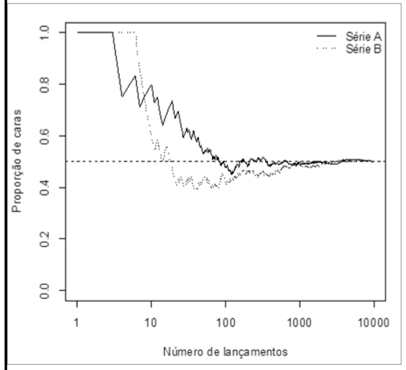
Aleatoriedade

Uma conceituação possível:

- Comportamento aleatório é imprevisível
  - √ A curto prazo
- Apresenta um padrão regular e previsível
  - √ A longo prazo

Exemplo

- 10.000 lançamentos de moeda honesta:



- Estabilização gradual da proporção de caras próximo a 0,5.
- Proporção de um número pequeno (ou moderado) de lançamentos pode produzir resultados distantes de 0,5

Fenômeno Aleatório

- Resultados individuais são incertos
- A distribuição desses resultados é regular ao longo do tempo

Aleatoriedade

- Descreve um tipo de ordem que emerge somente a longo prazo
  - √ Não é sinônimo de desorganizado

Probabilidade

- Definição frequentista:
  - √ Probabilidade de qualquer resultado de um fenômeno aleatório é a proporção de vezes em que o resultado ocorrerá em uma série muito longa de repetições
  - √ A ideia de probabilidade é empírica
    - – Baseia-se na observação e não na teoria
  - √ A probabilidade de sair cara é igual a 0,5 apenas porque a moeda tem dois lados?
    - – Giro da moeda

Estatística

- Como a incerteza atua?
  - √ Modelos probabilísticos
  - √ Se lançarmos uma moeda honesta 10 vezes, qual a probabilidade de se obter 5 caras?
- O que os dados nos indicam?
  - √ Inferência estatística
  - √ Lançamos uma moeda 10 vezes e obtivemos 5 caras. O que isso indica sobre a moeda?

Probabilidade

- Ramo da matemática preocupado com a análise de fenômenos aleatórios
  - √ Base matemática da Estatística.

Pesquisa Quantitativa - 2016

13

Probabilidade – Discussão

- Em geral, estratégias usadas para um experimento
  - √ Atribuir modelo mecanicista (função matemática) a tudo que é conhecido ou teorizado
  - √ Nos outros casos, atribuir aleatoriedade, mesmo que o processo em estudo não seja aleatório, em qualquer sentido da palavra.
  - √ Usar probabilidade para quantificar a incerteza nas conclusões, avaliando sua sensibilidade aos pressupostos do modelo

Pesquisa Quantitativa - 2016

14

- Probabilidade pode ser útil mesmo se a aleatoriedade verdadeira for uma quantidade indefinida
- Há várias interpretações possíveis de aleatoriedade
  - √ Não há concordância sobre a interpretação das probabilidades
  - √ Há concordância nas regras matemáticas que a probabilidade deve seguir

Pesquisa Quantitativa - 2016

15

Probabilidade – Interpretações

- Interpretação frequentista:
  - √ A probabilidade é a frequência de ocorrência de um evento, em repetições idênticas de um experimento
  - √ Usada na lógica do Teste de Hipótese
  - √ Até recentemente, a principal interpretação
  - √ É um Escola Estatística

Pesquisa Quantitativa - 2016

16

Probabilidade – Interpretações

- Interpretação Bayesiana:
  - √ A probabilidade é um grau subjetivo de crença
    - – Para um mesmo evento, duas pessoas poderiam atribuir diferentes probabilidades
  - √ Essa interpretação impede algumas das dificuldades filosóficas das interpretações de frequências.
  - √ Popularizada pelo avanço dos métodos e tecnologias computacionais

17

Pesquisa Quantitativa - 2016

Modelos Probabilísticos

- Descrição matemática de um fenômeno aleatório
- Descrição de um fenômeno aleatório:
  - √ Lista de resultados possíveis
  - √ Probabilidade para cada resultado
- Para modelo probabilístico é necessário:
  - √ Estabelecer um espaço amostral
  - √ Uma forma de atribuir probabilidades a eventos

18

Pesquisa Quantitativa - 2016

Modelos Probabilísticos

Espaço Amostral

- Pode ser muito simples:
  - √ Lançamento de uma moeda

ou muito complexo:

  - √ Todas as possíveis escolhas de 55.000 domicílios para formar uma amostra de 106 milhões de domicílios.

20

Pesquisa Quantitativa - 2016

Exemplo

- Resultados de pesquisa sobre satisfação no emprego
  - √ Amostra considerada representativa da população

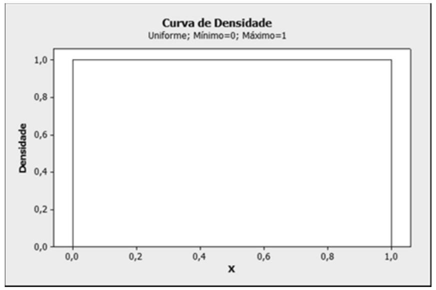
Classificação	Completamente satisfeito	Um pouco satisfeito	Um pouco insatisfeito	Completamente insatisfeito
Probabilidade	0,39	0,47	0,12	0,02

Atribuindo Probabilidades a Intervalos

- Escolha aleatória de um número entre 0 e 1
  - √ Resultados possíveis:
    - – Qualquer número neste intervalo
  - √ Não é possível atribuir probabilidade a cada valor individual e depois somá-los!
  - √ Podemos atribuir probabilidades a intervalos

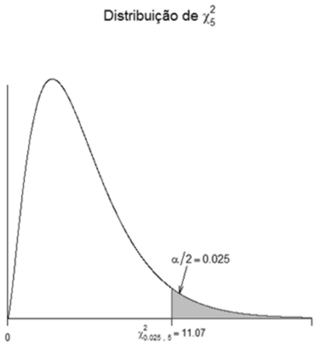
- X: valor escolhido
  - √ P{ $X \leq 0,5$}
  - √ P{$X \geq 0,5$}
  - √ P{$0,3 \leq X \leq 0,5$}
- Representação das probabilidades em retângulos
  - √ Qual a altura de cada retângulo?

Curva de Densidade - Uniforme



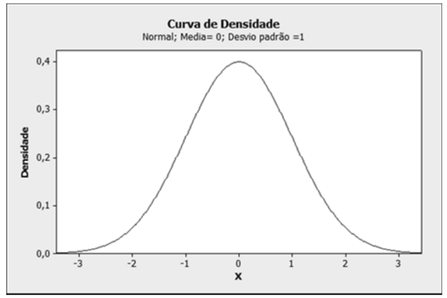
- Atribuição de probabilidade:
  - √ Área sob uma curva de densidade
- Relação entre histograma de frequências e curvas de densidade

Curvas de Densidade



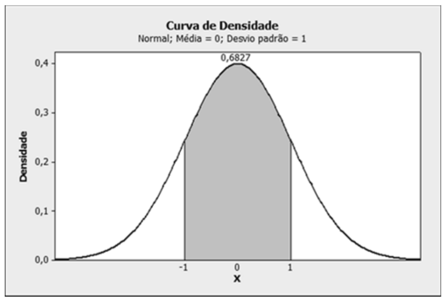
- Curvas de densidade:
  - √ Sempre acima do eixo horizontal
  - √ Área total sob a curva igual a 1

Modelo Probabilístico Normal



- Curva de densidade:
  - √ Simétrica em torno da média (μ)
  - √ Pontos de inflexão (mudança na curvatura) (μ ± σ)

Modelo Probabilístico Normal



- Probabilidade em intervalos
  - √ $\mu \pm \sigma$: 68%
  - √ $\mu \pm 2\sigma$: 95%
  - √ $\mu \pm 3\sigma$: 99,7%

Modelo Probabilístico Normal

- São importantes em Estatística:
  - √ Bons modelos de algumas distribuições de dados reais:
    - Algumas medidas de processo de produção
    - Algumas características biológicas de população
  - √ Boas aproximações dos resultados que se obtêm em muitos tipos de resultados aleatórios:
    - Lançar muitas vezes uma moeda

Modelo Probabilístico Normal

√ Diversos procedimentos de inferência estatística se baseiam na distribuição normal:

√ Funcionam bem quando aplicadas a outras distribuições que são aproximadamente simétricas

– Salários não assimétricos em uma empresa

Exemplo

- Em um dado ano, os escores de Matemática para alunos concluindo o ensino médio seguiram aproximadamente uma distribuição normal, com média 300 (em 500 possíveis) e desvio padrão 35 pontos. Obtenha uma valor aproximado de probabilidade:

√ Estudante tem um escore acima de 300

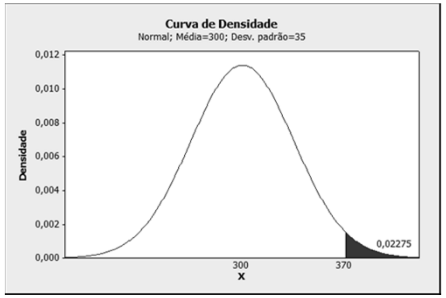
√ Estudante tem um escore acima de 370.

- Escore acima de 300:

√ 0,50 (exato)

- Escore acima de 370:

√ 0,025 (aproximado)



Variável Aleatória

- Nem todos os espaços amostrais são compostos por números
- Variável aleatória:

√ Seu valor é um resultado numérico de um fenômeno aleatório

Tipos de Variáveis Aleatórias

- Discretas:
  - √ Apresentam lacunas entre os valores possíveis
    (Em geral, números inteiros)
- Contínuas:
  - √ Não há lacunas entre os valores que a variável pode assumir

Distribuição de Probabilidade

- Distribuição idealizada das proporções dos resultados obtidos após um grande número de observações
- Informa:
  - √ Quais os valores que a variável aleatória pode assumir
  - √ Como atribuir probabilidades a esses valores

Variável Aleatória Discreta

Valores de X	x_1	x_2	**...**	x_k	**...**
Probabilidade	p_1	p_2	...	p_k	...

- As probabilidades p_i satisfazem:
  - √ Cada probabilidade p_i é um número entre 0 e 1

    $$0 \le p_i \le 1$$

  - √ A soma das probabilidades p_i é 1:

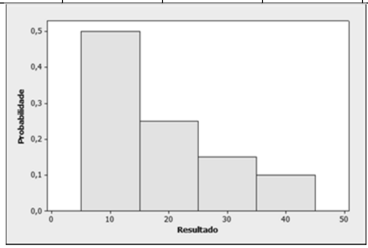
    $$p_1 + p_2 + ... + p_k + ... = 1$$

Exemplo

- Compradores de um determinado modelo de computador podem escolher modelos de HD de 10 GB, 20 GB, 30 GB ou 40 GB. Escolhe-se aleatoriamente alguns dos compradores dos últimos 60 dias. O tamanho do HD escolhido por um consumidor selecionado aleatoriamente é uma variável aleatória X.

- Distribuição de X:

Tamanho do HD	10	20	30	40
Probabilidade	0,50	0,25	0,15	0,10



√ Probabilidade de um consumidor escolher um HD de no mínimo 30 GB:

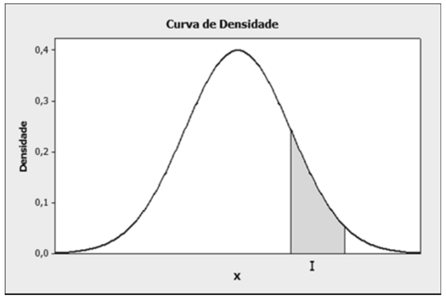
P{X = 30} + P{X = 40} = 0,15 + 0,10 = 0,25

37

Pesquisa Quantitativa - 2016

Variável Aleatória Contínua

- É descrita por uma curva de densidade
- Probabilidade de um intervalo de valores
  - √ Área abaixo da curva de densidade no intervalo



38

Pesquisa Quantitativa - 2016

Exemplo

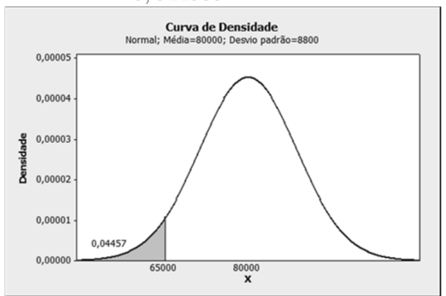
- Os pneus que teoricamente possuem uma vida útil X de 65.000 km têm, na verdade uma distribuição de probabilidade Normal com média μ = 80.000 e σ = 8.800 km.
  - √ Qual a probabilidade de que um pneu, selecionado aleatoriamente, tenha uma vida útil menor do que 65.000 km?

39

Pesquisa Quantitativa - 2016

√ Probabilidade de vida útil maior que 65.000 km

$$\begin{aligned} \mathrm{P}\{X < 65000\} &= \mathrm{P}\{\frac{X - 80.000}{8.800} < \frac{65.000 - 80.000}{8.800}\} \\ &= \mathrm{P}\{Z < -1,70\} \\ &= 0,044565 \end{aligned}$$



40

Pesquisa Quantitativa - 2016

Média de uma Variável Aleatória

- Exemplo – Jogo:
  - √ Você escolhe um número de três dígitos. Se ele for sorteado você ganha \$500. O bilhete vencedor é escolhido aleatoriamente dentre 1.000 números de três dígitos.
    - Obs.: O bilhete custa \$1,00

- X: variável aleatória associada ao prêmio pago a um bilhete
  - √ Distribuição de probabilidade de X:

Prêmio X	**\$0**	**\$500**
Probabilidade	0,999	0,001

  - √ Qual é pagamento médio feito pela loteria?
    - A longo prazo, uma pessoa seria premiada, em média, em um a cada 1.000 bilhetes comprados
    - Média dos pagamentos a longo prazo:

$$\$500\frac{1}{1000} + \$0\frac{999}{1000} = \$0,50$$

- Média da variável aleatória X (μ):
  - √ μ = \$0,50
- Importante:
  - √ A média não é um valor possível de X
  - √ Também denominada valor esperado de X
  - √ Cuidado:
    - Não quer dizer que esperamos que uma observação de X se situe próximo de sua média
- Qual o significado desta média para quem banca o jogo?

Média de Variável Aleatória Discreta

- Média dos valores possíveis ponderados pela respectivas probabilidades
- Definição:
  - √ Seja X uma variável aleatória discreta

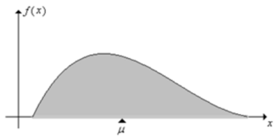
Valores de X	x_1	x_2	...	x_k	...
Probabilidade	p_1	p_2	...	p_k	...

$$\begin{aligned}\mu_X &= x_1p_1 + x_2p_2 + \cdots + x_kp_k + \ldots \\ &= \sum_{i=1}^{\infty} x_ip_i\end{aligned}$$

Média de Variável Aleatória Contínua

√ Ponto no qual a área sob a curva de densidade ficaria equilibrada

(caso fosse construída de um material sólido)



√ Em curvas simétricas (Ex. normal)

– Média encontra-se no ponto de simetria

√ Em curvas assimétricas

– Determinação requer emprego de Cálculo Integral

45

Pesquisa Quantitativa - 2016

Médias de Variáveis Aleatórias – Caso Geral

- Sejam as variáveis aleatórias X e Y e as constantes a e b:

√ W = a + b X

$$\mu_W = a + b\mu_X$$

√ W = X + Y

$$\mu_W = \mu_X + \mu_Y$$

√ W = X – Y

$$\mu_W = \mu_X - \mu_Y$$

46

Pesquisa Quantitativa - 2016

Análise de Portfolio

- Taxa de retorno de um investimento:

$$\text{taxa de retorno} = \frac{\text{variação no preço}}{\text{preço inicial}}$$

√ Variabilidade dos retornos (volatilidade)

– Medida de risco de um investimento

√ Em geral, investidores procuram retornos altos, mas também desejam segurança

- Portfolio:

√ Coleção de investimentos feitos por um indivíduo ou por uma instituição

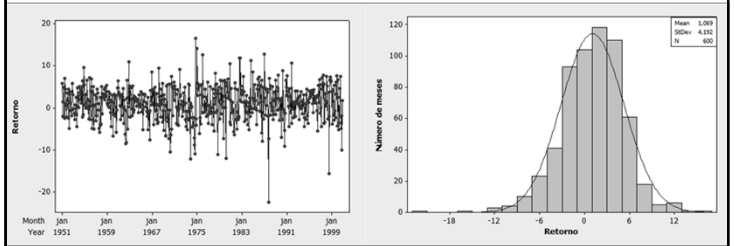
47

Pesquisa Quantitativa - 2016

Exemplo

- Retornos mensais de todas as ações ordinárias nos EUA

√ 600 meses (jan/1951 a dez/2000)



48

Pesquisa Quantitativa - 2016

- Retorno de um investimento ao longo do tempo
  - √ É uma variável aleatória
  - √ Interesse:
    - Retorno médio
    - Volatilidade

Exemplo

- Retorno anual de ativos de portfolio:
  - √ X: obrigações do Tesouro (20%)
  - √ Y: ações (80%)
- Retornos médios anuais:
  - √ Baseados nos retornos entre 1950 e 2000
  - √ $\mu_X = 5{,}2\%$ e $\mu_Y = 13{,}3\%$
- Retorno do portfolio:

$$R = 0,2X + 0,8Y$$

  - √ Retorno médio do portfolio

$$\mu_R = 0,2\mu_X + 0,8\mu_Y$$
$$= (0,2)(5,2) + (0,8)(13,3) = 11,68\%$$

Atividade

Empresa de Vendas

X: Quantidade vendida aos clientes militares

Unidades vendidas	1.000	3.000	5.000	10.000
Probabilidade	0,1	0,3	0,4	0,2

√ μ_X: número médio de unidades vendidas aos militares

$$\mu_X = (1.000)(0,1) + (3.000)(0,3) + (5.000)(0,4) + (10.000)(0,2)$$
$$= 100 + 900 + 2.000 + 2.000 = 5.000 \text{ unidades}$$

Empresa de Vendas

Y: Quantidade vendida aos clientes civis

Unidades vendidas	300	500	750
Probabilidade	0,4	0,5	0,1

√ μ_Y: número médio de unidades vendidas aos civis

$$\begin{aligned}\mu_X &= (300)(0,4) + (500)(0,5) + (750)(0,1) \\ &= 120 + 250 + 75 = 445 \text{ unidades}\end{aligned}$$

Pesquisa Quantitativa - 2016 53

Empresa de Vendas

W: Lucro total da empresa

$$W = 2.000\,X + 3.500\,Y$$

√ μ_W: lucro médio total da empresa

$$\begin{aligned}\mu_W &= 2.000\mu_X + 3.500\mu_Y \\ &= (2.000)(5.000) + 3.500(445) \\ &= \$10.000 + \$1.557.500 \\ &= \$11.557.500\end{aligned}$$

Pesquisa Quantitativa - 2016 54

Variância de Variável Aleatória

- Média ponderada dos desvios quadráticos $(X-\mu_X)^2$ da variável X em relação à sua respectiva média μ_X.
  - √ Cada resultado ponderado pela sua probabilidade
  - √ Notação σ_X^2.

Pesquisa Quantitativa - 2016 55

Variância – Variável Aleatória Discreta

- X: variável aleatória discreta com distribuição

Valores de X	x_1	x_2	...	x_k	...
Probabilidade	p_1	p_2	...	p_k	...

$$\begin{aligned}\sigma_X^2 &= (x_1-\mu)^2 p_1 + (x_2-\mu)^2 p_2 + \ldots \\ &\quad + (x_k-\mu)^2 p_k + \ldots \\ &= \sum_{i=1}^{\infty}(x_i-\mu)^2 p_i\end{aligned}$$

√ Desvio-padrão de X (σ_X): raiz quadrada da variância

Pesquisa Quantitativa - 2016 56

Exemplo – Jogo

- Bilhete com três dígitos no valor de $1
  - √ Ganha $500 se ele for sorteado
  - √ Valor esperado: $0,50

x_i	p_i	$x_i p_i$	$(x_i - \mu_X)^2 p_i$
0	0,999	0	$(0 – 0,50)^2$ (0,999) = 0,2497
500	0,001	0,50	$(500 – 0,50)^2$ (0,001) = 249,50025
	μ_X = 0,50		σ_X^2 = 249,75

  - √ Desvio-padrão: $\sigma_X = \sqrt{249{,}75}$ = $15,80
  - √ Em geral, jogos de azar apresentam grandes desvios padrão (risco)

57

Pesquisa Quantitativa - 2016

Atividade

Empresa de Vendas

X: Quantidade vendida aos clientes militares

Unidades vendidas	1.000	3.000	5.000	10.000
Probabilidade	0,1	0,3	0,4	0,2

√ Média e variância de X:

x_i	p_i	$x_i p_i$	$(x_i - \mu_X)^2 p_i$
1.000	0,1	100	$(1.000 - 5.000)^2(0{,}1) =$ 1.600.000
3.000	0,3	900	$(3.000 - 5.000)^2(0{,}3) =$ 1.200.000
5.000	0,4	2.000	$(5.000 - 5.000)^2(0{,}4) = 0$
10.000	0,2	2.000	$(10.000 - 5.000)^2(0{,}2) =$ 5.000.000
	μ_X = 5.000		σ_X^2 = 7.800.000

√ Desvio padrão de X: $\sigma_X = \sqrt{7.800.000} = 2.792,8$

59

Pesquisa Quantitativa - 2016

Empresa de Vendas

Y: Quantidade vendida aos clientes militares

Unidades vendidas	300	500	750
Probabilidade	0,4	0,5	0,1

√ Média e variância de X:

x_i	p_i	$x_i p_i$	$(x_i - \mu_X)^2 p_i$
300	0,4	120	$(300 - 445)^2(0{,}4) =$ 8.410,00
500	0,5	750	$(500 - 445)^2(0{,}5) =$ 1.512,50
750	0,1	75	$(750 - 445)^2(0{,}1) =$ 9.302,50
	μ_X = 445		σ_X^2 = 19.225,00

√ Desvio padrão de X: $\sigma_X = \sqrt{19.225} = 138,7$

60

Pesquisa Quantitativa - 2016

Empresa de Vendas

W: Lucro total da empresa

$$W = 2.000\,X + 3.500\,Y$$

√ σ_W^2: variância do lucro total da empresa

$$\begin{aligned}\sigma_W^2 &= (2.000)^2\sigma_X^2 + (3.500)\sigma_Y^2\\ &= (2.000)^2(7.800.000) + (3.500)^2(19.225)\\ &= 31.200.000.000.000 + 235.506.200.000\\ &= 31.435.500.000.000\end{aligned}$$

independência

√ σ_W: desvio padrão do lucro total da empresa

$$\sigma_W = \sqrt{31.435.500.000.000} = 5.606.737,6$$

Variáveis Aleatórias Independentes

- Conhecer qualquer evento envolvendo somente X nada nos informa sobre a ocorrência de qualquer evento envolvendo apenas Y
  - √ Pergunta importante:
    - Resultados parecem que não se relacionam uns com os outros

Correlação

- Correlação entre as variáveis X e Y (ρ):
  - √ Mede o grau de associação linear entre as duas variáveis aleatórias
  - √ $-1 \leq \rho \leq 1$
  - √ Correlação entre variáveis independentes é zero
  - √ Relação perfeitamente linear ($|\rho| = 1$)

    Ex.: Y = 100 – X

    X: gasto da renda familiar (%)

    Y: valor economizado (%)

Regras para Variâncias

- Sejam as variáveis aleatórias X e Y e as constantes a e b
  - √ W = a + bX

    $$\sigma_W^2 = b^2\sigma_X^2$$

  - √ W = X + Y
    - X e Y são variáveis aleatórias independentes

      $$\sigma_W^2 = \sigma_X^2 + \sigma_Y^2$$

    - X e Y têm correlação ρ:

      $$\sigma_W^2 = \sigma_X^2 + \sigma_Y^2 + 2\rho\,\sigma_X\sigma_Y$$

Regras para Variâncias

√ W = X – Y

- X e Y são variáveis aleatórias independentes

$$\sigma_W^2 = \sigma_X^2 + \sigma_Y^2$$

- A variância de X – Y é mais variável do que cada uma das variáveis
- X e Y têm correlação ρ:

$$\sigma_W^2 = \sigma_X^2 + \sigma_Y^2 + 2\rho\sigma_X\sigma_Y$$

65

Pesquisa Quantitativa - 2016

Exemplo – Jogo

- W: ganho líquido

  W = X – 1
- Valor médio que você ganha:

  √ $\mu_W = \mu_X - 1 = -\$0{,}50$

  √ Você perde em média \$0,50 por bilhete
- Qual a variância dos prêmios líquidos?

  √ É a mesma dos prêmios brutos (X)!

$$\sigma_X = \sqrt{249{,}75} = \$15{,}80$$

66

Pesquisa Quantitativa - 2016

- Suponha que você compre um bilhete de \$1 em dois dias diferentes

  √ Prêmios X e Y são independentes

  (sorteios independentes)

  √ Prêmio bruto total médio:

$$\mu_{X+Y} = \mu_X + \mu_Y = \$0,50 + \$0,50 = \$1,00$$

  √ Variabilidade do prêmio bruto total:

$$\sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2 = 249,75 + 249,75 = 499,5$$

$$\sigma_{X+Y} = \sqrt{\sigma_{X+Y}^2} = \sqrt{499,5} = \$22,35$$

  - As variância de variáveis aleatórias independentes se somam, os desvios padrão não

67

Pesquisa Quantitativa - 2016

Exemplo – Portfolio

- Composição de portfólio:

  √ Obrigações do Tesouro: 80%

  √ Ações: 20%
- Retornos anuais dos ativos:

  √ X: obrigações do Tesouro (20%)

  √ Y: ações (80%)
- Retorno anual do portfolio (R):

$$R = 0,2X + 0,8Y$$

68

Pesquisa Quantitativa - 2016

- Dados dos retornos:
  - √ Estimado pelos valores anuais entre 1950 e 2000

Retorno anual	Variável aleatória	Retorno médio	Risco	Correlação entre X e Y
Obrigações Tesouro	X	$\mu_X = 5{,}2\%$	$\sigma_X = 2{,}9\%$	$\rho = -0{,}1$
Ações	Y	$\mu_Y = 13{,}3\%$	$\sigma_Y = 17{,}0\%$	

  - √ Ações possuem retornos superiores às obrigações
  - √ Investimento mais volátil deve oferecer retornos maiores
    - – Compensar os investidores pelos riscos maiores

- Retorno médio anual do portfolio:

$$\mu_R = 0,2\mu_X + 0,8\mu_Y$$
$$= (0,2)(5,2) + (0,8)(13,3) = 11,68\%$$

  - √ Portfólio com retorno médio menor que se fosse composto apenas de ações
- Risco do portfolio:

$$\sigma_R^2 = \sigma_{0,2X}^2 + \sigma_{0,8Y}^2 + 2\rho\,\sigma_{0,2X}\,\sigma_{0,8Y}$$
$$= (0,2)^2\sigma_X^2 + (0,8)^2\sigma_Y^2 + 2\rho(0,2\times\sigma_X)(0,8\times\sigma_Y)$$
$$= (0,2)^2(2,9)^2 + (0,8)^2(17,0)^2 + (2)(-0,1)(0,2\times 2,9)(0,8\times 17,0)$$
$$= 183,719$$
$$\sigma_R = \sqrt{183,719} = 13,55\%$$

  - √ Portfólio é menos volátil que se fosse composto apenas de ações

Outros Modelos Probabilísticos

Modelo Binomial

- Uma loja vende 10 computadores com 1 ano de garantia.
  - √ Quantos deles não necessitarão de manutenção durante este período?
- É necessário um modelo probabilístico que faça uma contagem

Contexto Binomial

- Hipóteses
  - √ Número fixo de observações (*n*)
  - √ As *n* observações são todas independentes
    - Conhecer o resultado de uma delas nada informa sobre os demais resultados
  - √ Cada observação enquadra-se em apenas uma de duas categorias:
    - Sucesso (ocorre evento de interesse) ou fracasso
  - √ A probabilidade de sucesso (*p*) é a mesma para cada observação

Pesquisa Quantitativa - 2016

73

Exemplos

- √ Uma transação é auditada
  - Ela está em conformidade com os procedimentos ou não
- √ Você visita um cliente
  - Consegue realizar uma venda ou não
- √ O preço de um produto é reduzido
  - As vendas aumentam ou não
- √ Há um problema intermitente na rede da empresa
  - Em qualquer dia, o problema surge ou não.

Pesquisa Quantitativa - 2016

74

Distribuição Binomial

X: quantidade de sucessos em *n* observações

- √ Parâmetros:
  - *n*: número de observações
  - *p*: probabilidade de sucesso em uma observação qualquer
- √ Valores possíveis de X
  - $R_X = \{0, 1, 2, \ldots, n\}$

- Nem todos os contextos de contagem têm distribuição binomial

Pesquisa Quantitativa - 2016

75

Exemplo – Inspeção de Produto

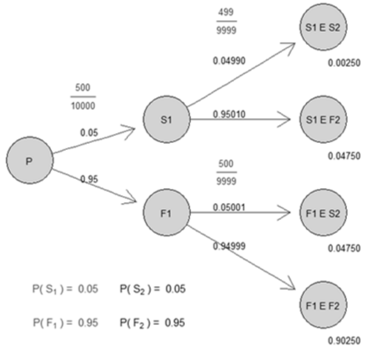
Inspeção de produto

- AAS de 10 componentes de um lote de 10.000 componentes
  - √ 500 desses componentes estão fora de especificação
- Variável aleatória de interesse:
  - √ X: quantidade de componentes não conformes na amostra.

Pesquisa Quantitativa - 2016

76

- Árvore de Probabilidade:



√ Remoção de um componente do lote altera pouco a constituição das 9.999 peças restantes

√ Pode-se considerar que $X \sim binomial(n = 10, p = 0{,}05)$

Exemplo

- Cada consumidor tem uma probabilidade de 0,25 de preferir o produto de sua empresa e não o produto da concorrência
  - √ Se escolhemos ao acaso 5 desses consumidores, qual a probabilidade de que exatamente 2 deles prefiram o seu produto?
  - √ $X \sim binomial(n = 5, p = 0{,}25)$

- Solução:
  - √ Eventos:
    - S = {consumidor prefere seu produto}
    - F = {consumidor prefere concorrência}
  - √ Configuração: | S | S | F | F | F |

$$\begin{aligned} \mathrm{P}\,(\mathrm{SSFFF}) &= \mathrm{P(S)\,P(S)\,P(F)\,P(F)\,P(F)} \\ &= (0,25)(0,25)(0,75)(0,75)(0,75) \\ &= (0,25)^2(0,75)^3 \end{aligned}$$

  - Toda disposição 2 S's e 3 F's: mesma probabilidade
  - √ Configurações possíveis:

S S F F F	S F S F F	S F F S F	S F F F S	F S S F F
F S F S F	F S F F S	F F S S F	F F S F S	F F F S S

  - √ $P\{X = 2\} = (10)(0{,}25)^2(0{,}75)^3 = 0{,}2637$

Probabilidade Binomial

- Se $X \sim binomial\ (n, p)$ então:

$$\mathrm{P}\,\{X = k\} = \binom{n}{k} p^k (1-p)^{n-k},\ k = 0, 1, \ldots, n.$$

√ com $\binom{n}{k} = \dfrac{n!}{k!(n-k)!}$

Conta a quantidade de modos diferentes que k sucessos podem ser arranjados em meio a n observações

Exemplo – Inspeção de Produto

X: quantidade de componentes não conformes na amostra

- √ $X \sim binomial(n = 10, p = 0{,}05)$

- Lote é rejeitado se pelo menos um dos itens da amostra por não conforme
  - √ Qual a probabilidade de rejeição do lote de 10.000 componentes?

- Probabilidade de pelo menos um componente estar defeituoso

$$
\begin{aligned}
\mathrm{P}\{X \geq 1\} &= 1 - \mathrm{P}\{X = 0\} \\
&= 1 - \binom{10}{0}(0,05)^0(0,95)^{10} \\
&= 1 - (1)(0,05)^0(0,95)^{10} \\
&= 1 - 0,5987 = 0,4013
\end{aligned}
$$

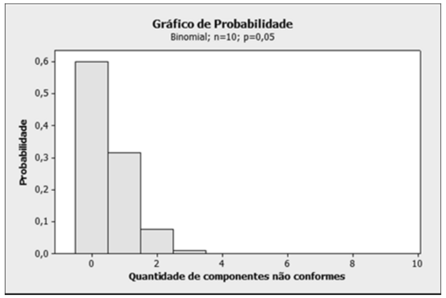
- Imagine que a venda esteja condicionada a um máximo de 5% de defeitos no lote
  - √ Essa seria uma boa regra para aceitar o lote?

- E se a regra se tornasse mais branda
  - √ No máximo um componente não conforme

$$
\begin{aligned}
\mathrm{P}\{X > 1\} &= 1 - \mathrm{P}\{X = 0\} - \mathrm{P}\{X = 1\} \\
&= 1 - \binom{10}{0}(0,05)^0(0,95)^{10} - \binom{10}{1}(0,05)^1(0,95)^9 \\
&= 1 - (1)(0,05)^0(0,95)^{10} - (10)(0,05)^1(0,95)^9 \\
&= 1 - 0,5987 - 0,3151 = 0,0861
\end{aligned}
$$

  - √ Agora, essa seria uma boa regra para aceitar o lote?
  - √ Qual a interpretação do resultado?

- Histograma de probabilidades



Distribuição Binomial

Parâmetros

	$n^{(*)}$ =	10
	p =	0,05

k	P(X=k)	P(X<=k)
0	0,5987	0,5987
1	0,3151	0,9139
2	0,0746	0,9885
3	0,0105	0,9990
4	0,0010	0,9999
5	0,0001	1,0000
6	0,0000	1,0000
7	0,0000	1,0000
8	0,0000	1,0000
9	0,0000	1,0000
10	0,0000	1,0000

- E se o lote a ser entregue estiver com o dobro de itens não conformes (p = 0,10)
  - √ Qual seria uma boa regra de decisão para rejeitar o lote?

Distribuição Binomial

Parâmetros

$n^{(*)}$ =	10
p =	0,1

k	P(X=k)	P(X<=k)
0	0,3487	0,3487
1	0,3874	0,7361
2	0,1937	0,9298
3	0,0574	0,9872
4	0,0112	0,9984
5	0,0015	0,9999
6	0,0001	1,0000
7	0,0000	1,0000
8	0,0000	1,0000
9	0,0000	1,0000
10	0,0000	1,0000

- Regra 1: pelo menos 1 defeituoso:
  - √ P{X>0} = 0,6513
- Regra 2: no máximo 1 defeituoso:
  - √ P{X>1} = 0,2639

Pesquisa Quantitativa - 2016

85

- Análise dos planejamentos amostrais:

Regra 1: pelo menos 1 defeituoso:

			Lote	
			Bom	Ruim
			p = 0,05	p = 0,10
Regra	Aceita	X = 0	0,5987	0,3487
	Rejeita	X > 0	0,4013	0,6513

Regra 2: no máximo 1 defeituoso:

			Lote	
			Bom	Ruim
			p = 0,05	p = 0,10
Regra	Aceita	X ≤ 1	0,9139	0,7361
	Rejeita	X > 1	0,0861	0,2639

- √ Em cada regra, qual o risco do vendedor e do comprador

		Risco	
		Vendedor	Comprador
Regra	1	0,4013	0,3487
	2	0,0861	0,7361

Pesquisa Quantitativa - 2016

86

- O que fazer para equilibrar os riscos?
- E se fossem exigidos lotes com uma menor quantidade de não conformes?
  - √ Exemplo: p = 0,01

Pesquisa Quantitativa - 2016

87

Binomial – Média e Desvio-Padrão

X: quantidade de componentes não conformes na amostra

- √ $X \sim binomial(n=10,\ p=0{,}05)$
- √ Média de X: $\mu_X = np$
- √ Desvio padrão de X: $\sigma_X = \sqrt{n\,p\,(1-p)}$

- Válidos apenas para as distribuições binomiais

Pesquisa Quantitativa - 2016

88

Exemplo – Inspeção de Produto

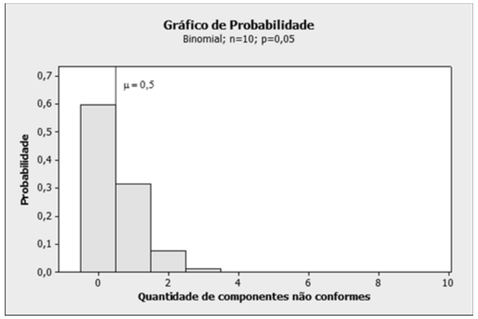
X: quantidade de componentes não conformes na amostra

√ $X \sim binomial(n{=}10,\ p{=}0{,}05)$

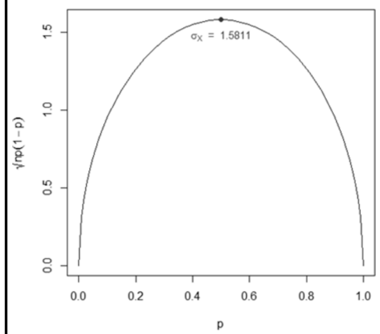
√ Média: $\mu_X = np = (10)(0,05) = 0,5$

√ Desvio-padrão: $$\sigma_X = \sqrt{n\,p\,(1-p)} = \sqrt{(10)(0,05)(0,95)} = 0,6892$$

- Histograma de probabilidades de X



- No exemplo, σ_X tem um valor máximo?
  - √ As probabilidades são limitadas.



Porque o máximo é atingido quando p = 0,5?

Distribuição de Poisson

- Há contagens que não são limitadas
- Exemplos:
  - √ Número de defeitos de fabricação da chapa de metal de uma geladeira
  - √ Quantidade de clientes que utilizam determinado caixa no final de semana
  - √ Número anual de acidentes de trabalho

Contexto de Poisson

- Número de sucessos em unidade de medida é independente do número de sucesso em quaisquer outras unidades de medida
  - √ Unidades não superpostas
- A probabilidade de ocorrência de um sucesso em uma unidade de medida é igual para todas as unidades de mesmo tamanho
  - √ Probabilidade é proporcional ao tamanho das unidades

- Probabilidade de ocorrer 2 ou mais sucessos em uma unidade aproxima-se de zero à medida que essa unidade vai se tornando menor
  - √ Em um instante ou ocorre um sucesso ou não ocorre.

Distribuição de Poisson

X: contagem de sucessos em unidade de medida

- √ $X \sim Poisson(\mu)$
- √ Parâmetros:
  - μ: número médio de sucessos por unidade de medida
- √ Valores possíveis de X
  - $R_X = \{0, 1, 2, \ldots\}$
- √ Função de probabilidade de X

$$\mathrm{P}\{X = k\} = \frac{\mathrm{e}^{\mu}\,\mu^k}{k!},\ k = 0, 1, \ldots$$

- √ Desvio padrão de X: $\sigma_X = \sqrt{\mu}$

Exemplo

- Certo tipo de material para carpete, tem a quantidade de defeitos por área variando segundo uma média de 1,6 defeito/m^2
  - √ X: número de defeitos por unidade de área
  - √ Unidade de medida: m^2 do material do carpete

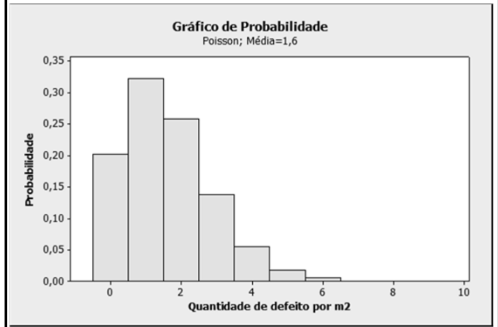
- Qual a probabilidade de que não haja mais do que 2 defeitos em 1 m²?

$$
\begin{aligned}
\mathrm{P}\{X \le 2\} &= \mathrm{P}\{X = 0\} + \mathrm{P}\{X = 1\} + \mathrm{P}\{X = 2\} \\
&= \frac{\mathrm{e}^{-1,6}(1,6)^0}{0!} + \frac{\mathrm{e}^{-1,6}(1,6)^1}{1!} + \frac{\mathrm{e}^{-1,6}(1,6)^2}{2!} \\
&= 0,2019 + 0,3230 + 0,2584 = 0,7833
\end{aligned}
$$

97

Pesquisa Quantitativa - 2016

- Distribuição de Poisson com $\mu = 1,6$



Distribuição Poisson

Parâmetro

μ = 1,6

k	P(X=k)	P(X<=k)
0	0,2019	0,2019
1	0,3230	0,5249
2	0,2584	0,7834
3	0,1378	0,9212
4	0,0551	0,9763
5	0,0176	0,9940
6	0,0047	0,9987
7	0,0011	0,9997
8	0,0002	1,0000
9	0,0000	1,0000
10	0,0000	1,0000

98

Pesquisa Quantitativa - 2016

Distribuição da Média Amostral

Distribuição Amostral

- Estatística de uma amostra aleatória:
  - √ Assume diferentes valores de amostra para amostra da mesma população
  - √ Estatística amostrais são variáveis aleatórias
    - – Variam de acordo a um padrão

101

Pesquisa Quantitativa - 2016

Média Amostral

- $\bar{X}$ varia de amostra para amostra
  - √ Apesar de sua variação, essa medida é uma medida razoável da média populacional μ.
  - √ Para amostras cada vez maiores, é garantido que a estatística $\bar{X}$ fica cada vez mais próxima do parâmetro μ.

102

Pesquisa Quantitativa - 2016

Lei dos Grandes Números

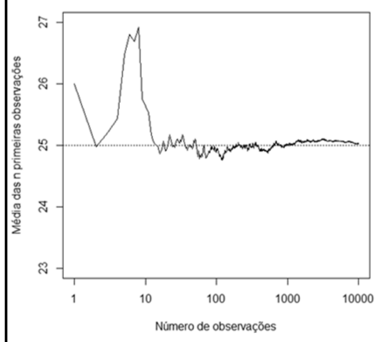
- Sejam observações aleatórias e independentes de uma mesma população, com uma média $\mu < \infty$.
  - √ À medida que o número de observações aumenta, a média $\bar{X}$ dos valores observados aproxima-se cada vez mais da média μ da população.
  - √ Os resultados médios de muitas observações são estáveis e previsíveis

103

Pesquisa Quantitativa - 2016

Exemplo

- Média acumulada de 10.000 observações
  - √ Simulação: Normal($\mu = 25, \sigma = 3$)



- Média das observações torna-se previsível a longo prazo.
- Cassinos podem contar com regularidade a longo prazo
  - √ Ganhos médios da banca em dezenas de milhares de jogadas estarão bastante próximos da média da distribuição dessas jogadas.

104

Pesquisa Quantitativa - 2016

Lei dos Grandes Números

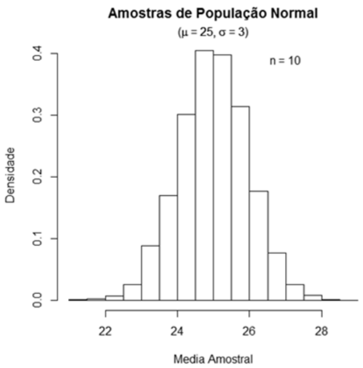
- Exemplo:
  - √ Demanda média
    - – Possível predizer, apesar de os clientes tomarem decisões independentes
- Quão grande é um número grande para assegurar que a média $\bar{X}$ dos resultados vá, de fato, ficar próxima da média μ da distribuição?
  - √ Quanto maior a variabilidade dos resultados, serão necessárias mais tentativas.

105

Pesquisa Quantitativa - 2016

Verificação Empírica

- E se o tamanho da amostra não for grande?
- Simulação:
  - √ 10.000 amostras de tamanho 10.



A que distância da média verdadeira (desconhecida) ficaram as médias de amostras de tamanho ?

Média de $\bar{X}$

$\bar{X}$: média de amostra aleatória de tamanho n

√ População grande com média μ e desvio-padrão σ.

Qual o padrão de comportamento de $\bar{X}$?

- Média da distribuição amostral de $\bar{X}$:

$$\mu_{\bar{X}} = \mu$$

√ Distribuição amostral de $\bar{X}$ é centrada em μ

(estimador não enviesado do parâmetro μ)

- Desvio-padrão da distribuição amostral de $\bar{X}$:
  - √ Quão perto o estimador ficará do parâmetro μ na maioria das amostras?

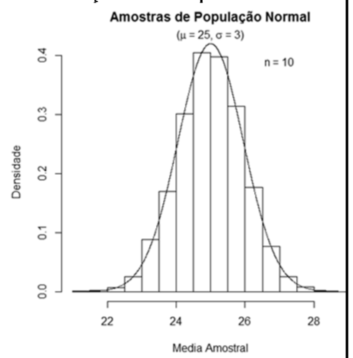
$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

  - √ Médias variam menos que as observações individuais
  - √ Menor quanto maior for a amostra
    - – Resultados de grandes amostras são menos variáveis do que os resultados de pequenas amostras

      (mais precisos!)

Distribuição Amostral de $\bar{X}$

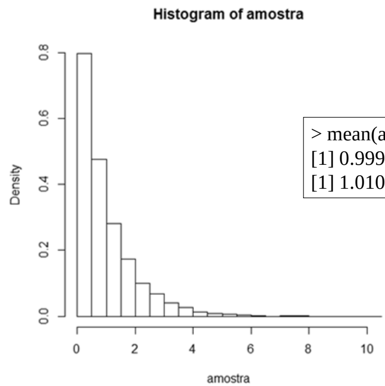
- População com distribuição N(μ,σ):
  - √ Média amostral $\bar{X}$ de n observações independentes tem distribuição N$\left(\mu, \frac{\sigma}{n}\right)$



Exemplo – Simulação

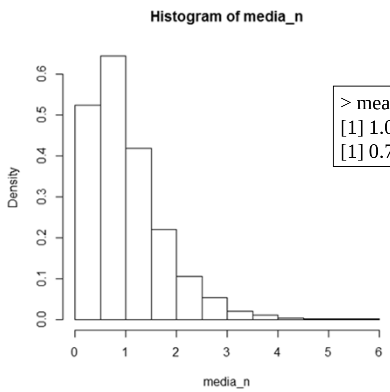
- População exponencial com média 1:
  - √ $\lambda = 1$
  - √ Geração de 10.000 valores dessa população
  - √ Amostra de tamanho 1 (n = 1)

- Amostra n =1



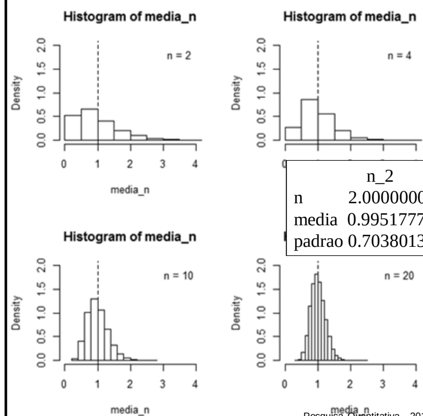
> mean(amostra); sd(amostra)
[1] 0.9990838
[1] 1.010478

- Amostra n = 2



> mean(media_n); sd(media_n)
[1] 1.012711
[1] 0.7129089

- Amostras de tamanhos 2, 4, 10 e 20

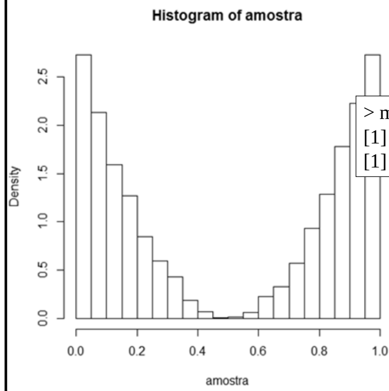


	n_2	n_4	n_10	n_20
n	2.0000000	4.0000000	10.0000000	20.0000000
media	0.9951777	1.0004503	0.9963200	0.9997876
padrao	0.7038013	0.5018562	0.3190797	0.2223573

Exemplo – Simulação

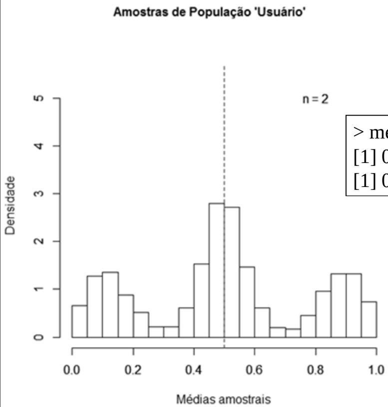
- População com densidade em U:
  - √ $f(x) = 12\,(x - 0{,}5)^2$
  - √ Geração de 10.000 valores dessa população
  - √ Amostra de tamanho 1 (n = 1)

- Amostra n =1



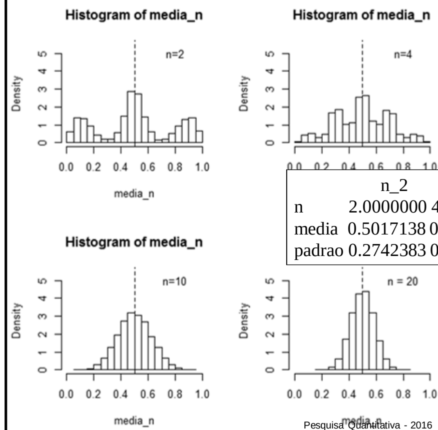

```
> mean(amostra); sd(amostra)
[1] 0.5061657
[1] 0.3877259
```


- Amostra n = 2




```
> mean(media_n); sd(media_n)
[1] 0,501576
[1] 0.2735612
```


- Amostras de tamanhos 2, 4, 10 e 20



	n_2	n_4	n_10	n_20
n	2.0000000	4.0000000	10.0000000	20.00000000
media	0.5017138	0.4986490	0.4994313	0.49993781
padrao	0.2742383	0.1957755	0.1221512	0.08599018

Teorema Central do Limite

- Amostra aleatória de tamanho n de uma população qualquer, com média μ e desvio-padrão σ:
  - √ Se *n* é suficientemente grande, a distribuição da média amostral $\bar{X}$ é aproximadamente normal.

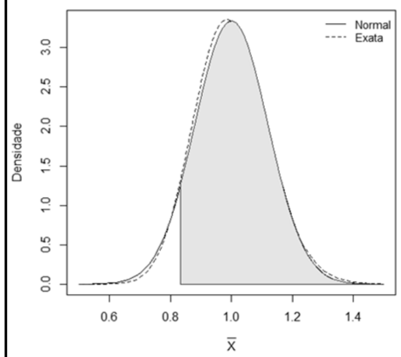
Exemplo

- X: tempo para execução de manutenção preventiva de unidade de ar condicionado
  - √ Distribuído exponencialmente
  - √ Média: $\mu = 1$ hora
  - √ Desvio padrão: $\sigma = 1$ hora
- Na empresa, há 70 dessas unidades.
- Qual a probabilidade de que o tempo médio exceda 50 minutos?

- Aproximação pelo TCL (n = 70)
  - √ Média amostral ($\bar{X}$) tem distribuição aproximadamente normal com média $\mu = 1$ e desvio padrão: $\frac{\sigma}{\sqrt{70}} = \frac{1}{\sqrt{70}} = 0,12$ hora
  - √ Cálculo da probabilidade de exceder 50 min:
    - 50 min = $\frac{50}{60}$ = 0,83 hora

$$\mathrm{P}\left\{\bar{X} > 0,83\right\} = \mathrm{P}\left\{Z > \frac{0,83-1}{0,12} = -1,42\right\}$$
$$= 1 - 0,077804 = 0,922196$$

- Distribuição exata e aproximada



- Curva Normal é uma boa aproximação
- Probabilidade
  - √ Aproximada: 0,9222
  - √ Exata: 0,9251
  - √ Diferença: 0,0029

Comentários

- √ A aproximação normal para a média amostral depende do tamanho da amostra
- √ Com população contínua, unimodal e simétrica, na maioria dos casos, o TCL trabalha bem para pequenas amostras (n = 4, 5).
- √ Em muitos casos de interesse prático, a aproximação normal será satisfatória para $n \geq 30$
- √ Se $n < 30$, o TCL funcionará se a distribuição da população não for muito diferente da normal

122

Pesquisa Quantitativa - 2016

Referências

Bibliografia Recomendada

- AGRESTI, A.; FINLAY, B. *Métodos estatísticos para as ciências sociais.* Penso, 2012.
- MOORE, D. S.; MCCABE, G. P.; DUCKWORTH, W. M.; SLOVE, S. L. *A prática da estatística empresarial: como usar dados para tomar decisões*. LTC, 2006.

124

Pesquisa Quantitativa - 2016