Estatística Econômica II

Lupércio França Bessegato
Departamento de Estatística/UFJF

Roteiro Geral

1. Introdução
2. **Modelos probabilísticos**
3. Amostragem
4. Distribuições amostrais e estimação
5. Testes de significância
6. Comparações de médias
7. Tópicos em inferência
8. Referências

Estatística Econômica II - 2018 2

Modelos Probabilísticos Univariados

Roteiro do Módulo

3. Modelos probabilísticos:
   a) Aleatoriedade
   b) Modelos probabilísticos
   c) Variáveis aleatórias
   d) Médias e variâncias de variáveis aleatóris
   e) Modelo Binomial e de Poisson
   f) Distribuição da média amostral
   g) Modelo Normal
   h) Referências

Estatística Econômica II - 2018 4

Aleatoriedade

- **Você acredita em destino?**

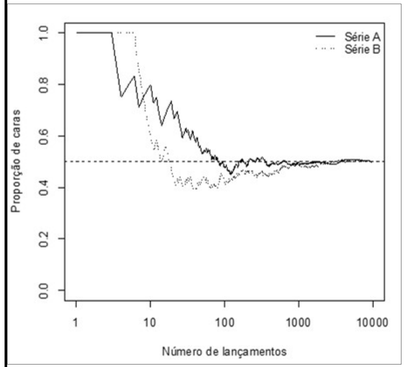
Aleatoriedade

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.

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.

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

- 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

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

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

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

Variáveis Aleatórias

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	$\mathbf{x_1}$	$\mathbf{x_2}$	**...**	$\mathbf{x_k}$	**...**
Probabilidade	p_1	p_2	...	p_k	...

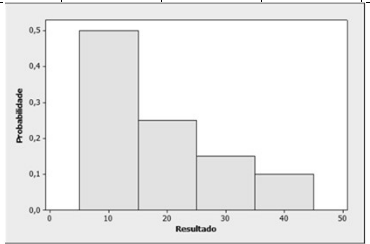
- As probabilidades p_i satisfazem:
  - √ Cada probabilidade p_i é um número entre 0 e 1
    - $0 \leq p_i \leq 1$
  - √ A soma das probabilidades p_i é 1:
    - $p_1 + p_2 + ... + p_k + ... = 1$

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:

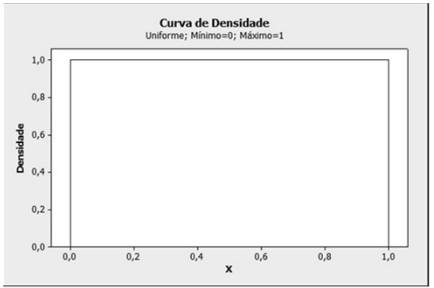
$P\{X = 30\} + P\{X = 40\} = 0,15 + 0,10 = 0,25$

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

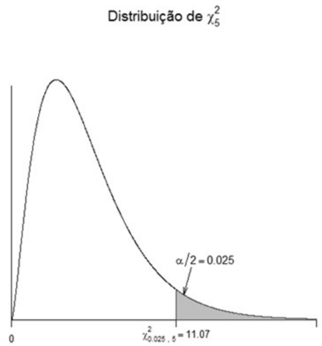


- Atribuição de probabilidade:
  - √ Área sob uma curva de densidade
- Relação entre histograma de frequências e curvas de densidade

35

Estatística Econômica II - 2018

Curvas de Densidade



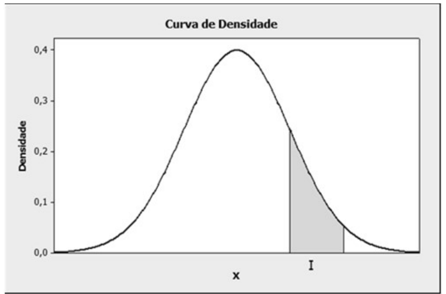
- Curvas de densidade:
  - √ Sempre acima do eixo horizontal
  - √ Área total sob a curva igual a 1

36

Estatística Econômica II - 2018

Variável Aleatória Contínua

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



37

Estatística Econômica II - 2018

Médias e Variâncias de Variáveis Aleatórias

Considerações

- Probabilidade
  - √ A longo prazo, descreve regularidade do comportamento de fenômenos aleatórios.
- Distribuição de probabilidade:
  - √ Função idealizada de frequências relativas

- Frequências relativas:
  - √ Médias e desvios-padrão
- Distribuições de variáveis aleatórias:
  - √ Médias e desvios-padrão
  - √ Ex.:
    - Ganho médio de um jogo de azar
    - Desvio-padrão da quantidade de chamadas telefônicas por hora

Média de Variável Aleatória

- Conjunto de observações
  - √ Média observada (média amostral)
  - √ Notação: $\bar{x}$
- Variável aleatória
  - √ Média dos valores possíveis de X
  - √ Leva-se em conta que nem todos os valores precisam ser igualmente prováveis
  - √ Notação: μ

Exemplo

- Jogo de Loteria:
  - √ 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:

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

    - A longo prazo, você ganha $500, em média, uma vez a cada 1.000 apostas e 0 nas 999 restantes

- 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	...

$$\mu_X = x_1p_1 + x_2p_2 + \cdots + x_kp_k + \dots$$
$$= \sum_{i=1}^{\infty} x_ip_i$$

Exemplo

- Estudo de falhas nas pinturas de geladeiras
  - √ Tipos de falhas: ondulações e afundamentos
- Quantidade de falhas em uma geladeira é uma variável aleatória
- Estimativas de médias efetuadas por inspetor de qualidade:
  - √ Ondulações: 0,7 por geladeira
  - √ Afundamentos: 1,4 por geladeira
- Qual número médio de falhas por geladeira

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

Exercício 4 – Lista nº 2
(MOORE et al – seção 4.4)

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

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

Empresa de Vendas

W: Lucro total da empresa

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

√ μ_W: lucro médio total da empresa

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

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.

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	...

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

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

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_ip_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	$(500 - 0{,}50)^2\,(0{,}001) = 249{,}50025$
	$\mu_X = 0{,}50$		$\sigma_X^2 = 249{,}75$

√ Desvio-padrão: $\sigma_X = \sqrt{249{,}75}$ = $15,80

√ Em geral, jogos de azar apresentam grandes desvios padrão (risco)

Continuação exercício 4 - Lista 2 (MOORE et al – seção 4.4)

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	$(10.000 - 5.000)^2(0{,}2) = 5.000.000$
$\mu_X = 5.000$			$\sigma_X^2 = 7.800.000$

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

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	$(750 - 445)^2(0{,}1) = 9.302{,}50$
$\mu_X = 445$			$\sigma_X^2 = 19.225{,}00$

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

Empresa de Vendas

W: Lucro total da empresa

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

√ σ_W^2: variância do lucro total da empresa

independência

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

√ 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:

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

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)!

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

63

Estatística Econômica II - 2018

- 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:

$$\sigma^2_{X+Y} = \sigma^2_X + \sigma^2_Y = 249,75 + 249,75 = 499,5$$

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

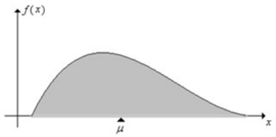
  – As variância de variáveis aleatórias independentes se somam, os desvios padrão não

64

Estatística Econômica II - 2018

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

65

Estatística Econômica II - 2018

Modelo Binomial

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

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.

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
  - $R_X = \{0, 1, 2, \ldots, n\}$

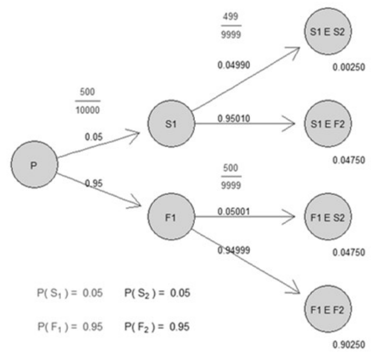
- Nem todos os contextos de contagem têm distribuição binomial

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.

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

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

- 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 $\binom{n}{k} = \frac{n!}{k!(n-k)!}$

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

- 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

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

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

Exemplo – Inspeção de Produto

X: quantidade de componentes não conformes na amostra

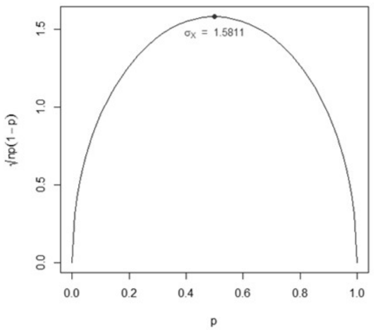
√ $X \sim binomial(n=10,\ p=0{,}05)$

√ 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?

Modelo de Poisson

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: $\sigma_X = \sqrt{\mu}$

92

Estatística Econômica II - 2018

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

93

Estatística Econômica II - 2018

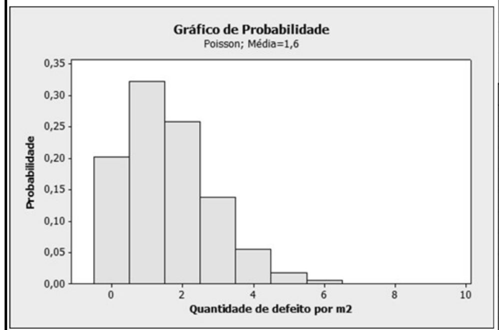
- Qual a probabilidade de que não haja mais do que 2 defeitos em 1 m^2?

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

94

Estatística Econômica II - 2018

- Distribuição de Poisson com μ = 1,6



Distribuição Poisson

Parâmetro

μ = 1,6

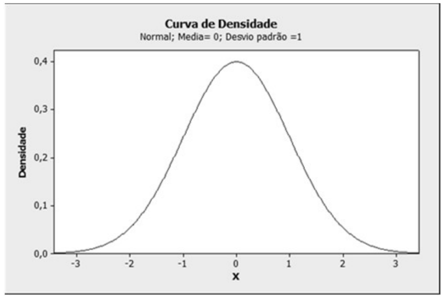
k	P(X=k)	P(X<=k)
0	0,2019	0,2019
1	0,3230	0,5249
2	0,2584	0,7834
3	0,1378	0,9212
4	0,0551	0,9763
5	0,0176	0,9940
6	0,0047	0,9987
7	0,0011	0,9997
8	0,0002	1,0000
9	0,0000	1,0000
10	0,0000	1,0000

95

Estatística Econômica II - 2018

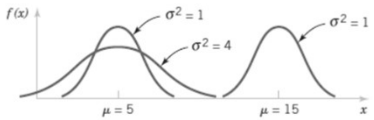
Modelo Normal

Modelo Probabilístico Normal



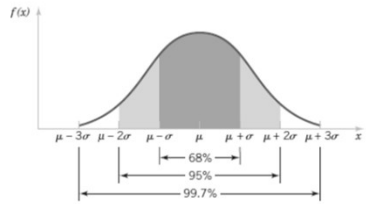
- Curva de densidade:
  - √ Simétrica em torno da média (μ)
  - √ Pontos de inflexão (mudança na curvatura) (μ ± σ)

- Efeitos dos parâmetros μ e σ na curva:



- √ μ é a média
  - parâmetro de locação)
- √ σ é o desvio padrão
  - parâmetro de escala
- √ Para qualquer valor de μ e σ a forma da curva é a mesma
  - formato de sino

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

Normal Padronizada

- Variável aleatória normal padronizada (Z)

$$Z \sim N(\mu = 0, \sigma = 1)$$

  - √ Média é zero e desvio padrão é 1
  - √ Áreas sob a curva normal padronizada são tabeladas
    - Possível calcular probabilidades associadas à variável aleatória normal padrão

- Tabela da distribuição normal padrão:

TABLE • III Cumulative Standard Normal Distribution

z	−0.09	−0.08	−0.07	−0.06	−0.05	−0.04	−0.03	−0.03	−0.01	−0.00
−3.9	0.000033	0.000034	0.000036	0.000037	0.000039	0.000041	0.000042	0.000044	0.000046	0.000048
−3.8	0.000050	0.000052	0.000054	0.000057	0.000059	0.000062	0.000064	0.000067	0.000069	0.000072
−3.7	0.000075	0.000078	0.000082	0.000085	0.000088	0.000092	0.000096	0.000100	0.000104	0.000108
−3.6	0.000112	0.000117	0.000121	0.000126	0.000131	0.000136	0.000142	0.000147	0.000153	0.000159
−3.5	0.000165	0.000172	0.000179	0.000185	0.000193	0.000200	0.000208	0.000216	0.000224	0.000233
−3.4	0.000242	0.000251	0.000260	0.000270	0.000280	0.000291	0.000302	0.000313	0.000325	0.000337
−3.3	0.000350	0.000362	0.000376	0.000390	0.000404	0.000419	0.000434	0.000450	0.000467	0.000483
−3.2	0.000501	0.000519	0.000538	0.000557	0.000577	0.000598	0.000619	0.000641	0.000664	0.000687
−3.1	0.000711	0.000736	0.000762	0.000789	0.000816	0.000845	0.000874	0.000904	0.000935	0.000968

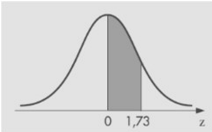
$\Phi(-3{,}01) = 0{,}001306$

TABLE • III Cumulative Standard Normal Distribution (*Continued*)

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.500000	0.503989	0.507978	0.511967	0.515953	0.519939	0.532922	0.527903	0.531881	0.535856
0.1	0.539828	0.543795	0.547758	0.551717	0.555760	0.559618	0.563559	0.567495	0.571424	0.575345
0.2	0.579260	0.583166	0.587064	0.590954	0.594835	0.598706	0.602568	0.606420	0.610261	0.614092
0.3	0.617911	0.621719	0.625516	0.629300	0.633072	0.636831	0.640576	0.644309	0.648027	0.651732
0.4	0.655422	0.659097	0.662757	0.666402	0.670031	0.673645	0.677242	0.680822	0.684386	0.687933
0.5	0.691462	0.694974	0.698468	0.701944	0.705401	0.708840	0.712260	0.715661	0.719043	0.722405
0.6	0.725747	0.729069	0.732371	0.735653	0.738914	0.742154	0.745373	0.748571	0.751748	0.754903
0.7	0.758036	0.761148	0.764238	0.767305	0.770350	0.773373	0.776373	0.779350	0.782305	0.785236
0.8	0.788145	0.791030	0.793892	0.796731	0.799546	0.802338	0.805106	0.807850	0.810570	0.813267
0.9	0.815940	0.818589	0.821214	0.823815	0.826391	0.828944	0.831472	0.833977	0.836457	0.838913
1.0	0.841345	0.843752	0.846136	0.848495	0.850830	0.853141	0.855428	0.857690	0.859929	0.862143

$\Phi(0{,}97) = 0{,}833977$

- Cálculos de probabilidade com a tabela:
  - √ P{0 ≤ Z ≤ 1,73}



$$\begin{aligned} P\{0 \le Z \le 1,73\} &= \Phi(1,73) - \Phi(0) \\ &= 0,958185 - 0,500000 \\ &= 0,458185 \end{aligned}$$

  - √ P{Z ≥ 1,73}

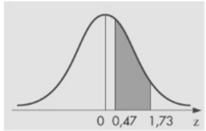
$$\begin{aligned} P\{Z \ge 1,73\} &= 1 - \Phi(1,73) \\ &= 1 - 0,958185 \\ &= 0,041815 \end{aligned}$$

√ P{−1,73 ≤ Z ≤ 1,73}

- $P\{Z \le -1,73\} = P\{Z \ge 1,73\}$

$$\begin{aligned} P\{-1,73 \le Z \le 1,73\} &= \Phi(1,73) - \Phi(-1,73) \\ &= 0,958185 - 0,041815 \\ &= 0,916370 \end{aligned}$$

√ P{0,47 ≤ Z ≤ 1,73}



$$\begin{aligned} P\{0,47 \le Z \le 1,73\} &= \Phi(1,73) - \Phi(0,47) \\ &= 0,958185 - 0,680822 \\ &= 0,277363 \end{aligned}$$

- Probabilidades da Normal Padrão

1. $P\{Z \ge 1,26\}$

$$\begin{aligned} P\{Z \ge 1,26\} &= 1 - \Phi(1,26) \\ &= 1 - 0,896165 \\ &= 0,103835 \end{aligned}$$

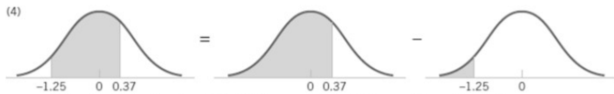
2. P{Z < −0,86}

$$P\{Z < -0,86\} = \Phi(-0,86) = 0,194894$$

3. P{Z > −1,37}

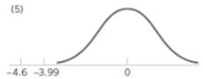
$$\begin{aligned} P\{Z > -1,37\} &= P\{Z < 1,37\} \\ &= \Phi(1,37) = 0,914657 \end{aligned}$$

4. $P\{-1,25 \le Z \le 0,37\}$



$$\begin{aligned} P\{-1,25 \le Z \le 0,37\} &= \Phi(0,37) - \Phi(-1,25) \\ &= 0,644309 - 0,1055650 \\ &= 0,537525 \end{aligned}$$

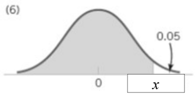
5. P{Z ≤ −4,6}



$$P\{Z \le -4,6\} < P\{Z \le -3,99\} = \Phi(-3,99) = 0,000033$$

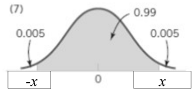
- Determinação de percentis (ou quantis)

6. Encontre $\boldsymbol{x}$ tal que $P\{Z > x\} = 0{,}05$



$$\Phi(1,65) = 0,950529$$
$$x \approx 1,65$$

7. Encontre $\boldsymbol{x}$ tal que $P\{-x \leq Z \leq x\}$ 0,99



$$\Phi(2,58) = 0,995060$$
$$x \approx 2,58$$

Padronização

- Se X for uma variável aleatória normal, com média μ e desvio padrão σ, então a variável aleatória

$$Z = \frac{(X - \mu)}{\sigma}$$

é uma variável aleatória normal com média 0 e desvio padrão 1.

√ Transformação válida para qualquer variável aleatória normal X

√ Áreas sob a normal padronizada são tabeladas

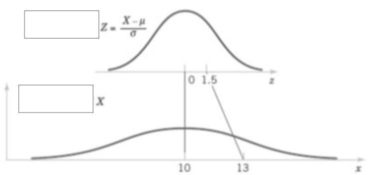
Exemplo

- X é uma variável aleatória normal com média 10 e desvio padrão 2

$$X \sim N(\mu = 10, \sigma = 2)$$

√ $P\{X > 13\}$

$$P\{X > 13\} = P\{\frac{X - 10}{2} > \frac{13 - 10}{2}\}$$
$$= P\{Z > 1,5\} = 1 - \Phi(1,5)$$
$$= 1 - 0,933193 = 0,066807$$

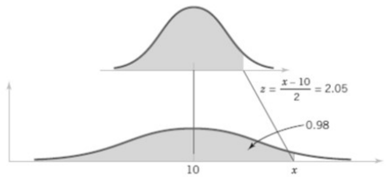


$$X \sim N(\mu = 10, \sigma = 2)$$

√ Encontre o valor $\boldsymbol{x}$ tal que $P\{X < x\}$ 0,98

$$P\{X < x\} = P\{\frac{X - 10}{2} < \frac{x - 10}{2}\} \Rightarrow P\{Z < 2,05\} = 0,979818$$
$$= P\left\{Z < \frac{x - 10}{2}\right\}$$
$$= 0,980$$

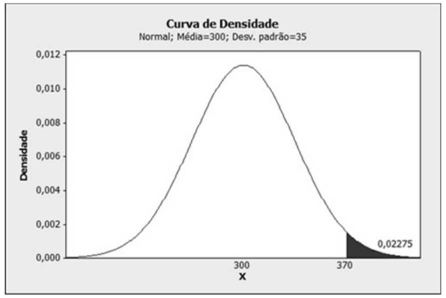
$$x = \mu + z\sigma$$
$$= 10 + (2,05)(2)$$
$$= 14,1$$



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:
  - √ ≈ 0,023

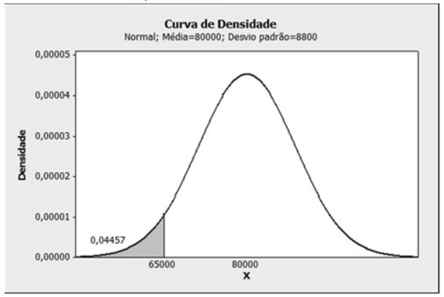


Exemplo

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

√ Probabilidade de vida útil menor que 65.000 km

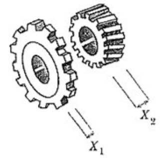
$$\begin{aligned} \mathrm{P}\{X < 65000\} &= \mathrm{P}\{\frac{X - 80.000}{8.800} < \frac{65.000 - 80.000}{8.800}\} \\ &= \mathrm{P}\{Z < -1,70\} \\ &= 0,044565 \end{aligned}$$



Combinações Lineares de Variáveis Aleatórias Normais

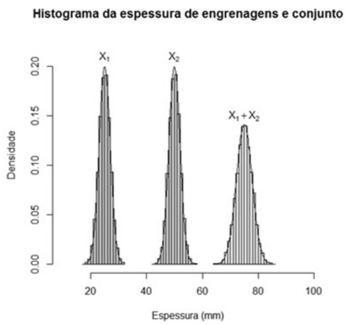
Soma de Variáveis Aleatórias

- Exemplo:
  - √ Espessura de engrenagens
  - √ $X_1 \sim N(\mu_1 = 25\ mm, \sigma_1 = 2\ mm)$
  - √ $X_2 \sim N(\mu_2 = 50\ mm, \sigma_2 = 2\ mm)$
  - √ X_1 e X_2 são independentes
  - √ Como se parece a distribuição da espessura total, $X_1 + X_2$?



119
Estatística Econômica II - 2018

- Simulação de rodada de produção de cada um dos 10.000



- Soma apresenta
  - √ Simetria
  - √ Maior dispersão
- Aparenta ser normal

120
Estatística Econômica II - 2018

- Média da soma $X_1 + X_2$:

$$\mu_{Soma} = \mu_1 + \mu_2$$
$$= 25 + 70 = 75 \text{ mm.}$$

- Desvio padrão da soma:

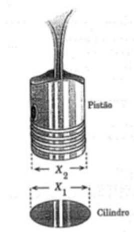
$$\sigma_{Soma} = \sqrt{\sigma_1^2 + \sigma_2^2} = \sqrt{2^2 + 2^2} = \sqrt{8}$$
$$= 2,828$$

  - √ $\sigma_{soma} > \sigma_1, \sigma_{soma} > \sigma_2$
  - √ Conjuntos são mais variáveis que as peças individuais

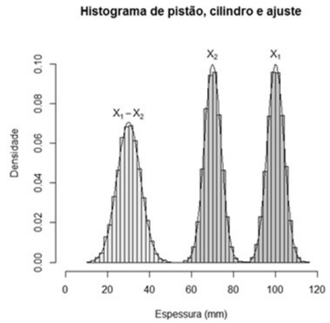
121
Estatística Econômica II - 2018

Diferença de Variáveis Aleatórias

- Exemplo:
  - √ Pistões de motor e cilindro
  - √ $X_1 \sim N(\mu_1 = 100\ mm, \sigma_1 = 4\ mm)$
  - √ $X_2 \sim N(\mu_2 = 70\ mm, \sigma_2 = 4\ mm)$
  - √ X_1 e X_2 são independentes
  - √ Como se parece a distribuição da folga entre o cilindro e o pistão, $D = X_1 - X_2$?



- Simulação de rodada de produção de cada um dos 10.000



- Diferença apresenta
  - √ Simetria
  - √ Maior dispersão
- Aparenta ser normal

- Média da diferença $D = X_1 - X_2$:

$$\mu_D = \mu_1 - \mu_2$$
$$= 100 - 70 = 30 \text{ mm}.$$

- Desvio padrão da soma:

$$\sigma_D = \sqrt{\sigma_1^2 + \sigma_2^2} = \sqrt{4^2 + 4^2} = \sqrt{32}$$
$$= 5,657$$

  - √ $\sigma_D > \sigma_1, \sigma_D > \sigma_2$
  - √ Folgas são mais variáveis que as peças individuais

Combinações Lineares

- Somas e diferenças de variáveis aleatórias independentes e normalmente distribuídas também tem distribuição normal.

Exemplo – 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

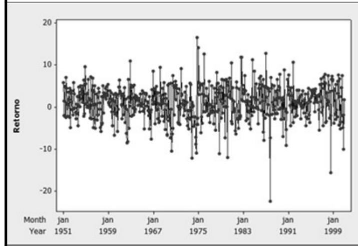
- Portfolio:

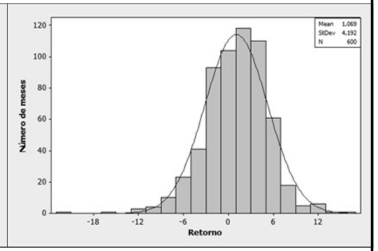
  √ Coleção de investimentos feitos por um indivíduo ou por uma instituição

Exemplo

- Retornos mensais de todas as ações ordinárias nos EUA

  √ 600 meses (jan/1951 a dez/2000)





- Retorno de um investimento ao longo do tempo

  √ É uma variável aleatória

  √ Interesse:
    – Retorno médio
    – Volatilidade

Exemplo – Portfolio

- Composição de portfólio:

  √ Obrigações do Tesouro: 80%

  √ Ações: 20%

- Retornos anuais dos ativos:

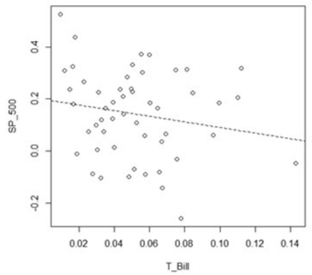
  √ X: obrigações do Tesouro (20%)

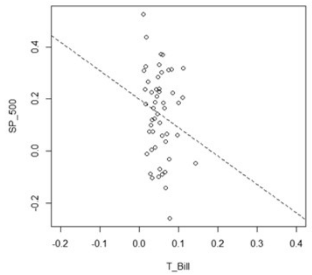
  √ Y: ações (80%)

- Retorno anual do portfolio (R):

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

- Retornos ações vs. obrigações do Tesouro





- 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

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.