

Notas de aulas de Estatística Econômica

Marcos Minoru Hasegawa

2021-05-26

Sumário

| | |
|--|-----------|
| Licença | 5 |
| Sobre o material | 7 |
| Sobre o Autor | 9 |
| 1 Estatística Descritiva | 11 |
| 1.1 Medidas de posição | 11 |
| 1.2 Medidas de dispersão | 20 |
| 1.3 Medidas de relação linear entre duas variáveis | 31 |
| 2 Medidas de desigualdade | 35 |
| 2.1 Princípio de Pigou-Dalton | 35 |
| 2.2 Transferência Regressiva | 35 |
| 2.3 Curva de Lorenz | 35 |
| 2.4 Índice Gini | 38 |
| 2.5 Discrepância Máxima | 46 |
| 2.6 Redundância e Índice de Theil | 48 |
| 2.7 Variância dos Logaritmos | 55 |
| 3 Números-Índices | 57 |
| 3.1 Introdução | 57 |
| 3.2 Preços Relativos | 57 |
| 3.3 Índice Simples de Preços Agregados | 59 |
| 3.4 Média Aritmética dos Preços Relativos | 61 |
| 3.5 Índice de Preços de Laspeyres | 62 |
| 3.6 Índice de Preços de Paasche | 65 |
| 3.7 Índice de Preços de Fisher | 66 |
| 3.8 Índice de Preços de Marshall-Edgeworth | 67 |
| 3.9 Deflacionamento | 68 |
| 4 Variável Aleatória Discreta | 71 |
| 4.1 Esperança matemática | 71 |
| 4.2 Variável Aleatória Discreta | 74 |

| | | |
|----------|---|------------|
| 4.3 | Distribuição Uniforme | 74 |
| 4.4 | Distribuição de Bernoulli | 75 |
| 4.5 | Distribuição Binomial | 77 |
| 4.6 | Distribuição de Poisson | 83 |
| 5 | Variável Aleatória Contínua | 87 |
| 5.1 | Distribuições contínuas | 87 |
| 5.2 | Distribuição Normal | 98 |
| 5.3 | Teorema de Tchebichev | 117 |
| 5.4 | Momentos de uma distribuição | 119 |
| 6 | Distribuição de variáveis aleatórias conjunta | 121 |
| 6.1 | Distribuição de probabilidade conjunta de variáveis aleatórias discretas | 121 |
| 6.2 | Distribuição de probabilidade conjunta de variáveis aleatórias contínuas | 136 |
| 7 | Estimadores | 155 |
| 7.1 | Introdução | 155 |
| 8 | Teste de hipótese | 157 |
| 9 | Regressão Linear | 159 |

Licença

Como está descrito no repositório, os poucos códigos originais desenvolvidos ao longo do texto estão sob a licença **GNU GPLv3** .

O texto e as artes gráficas elaboradas de forma original estão sob licença **Creative Commons BY-NC-SA 4.0**.

Sobre o material

A situação especial causada pela pandemia da COVID-19 forçou a muitos professores criarem materiais para facilitar aulas remotas das suas disciplinas. A disciplina SE305 Estatística Econômica e Introdução à Econometria da UFPR não poderia ser diferente. Então, o objetivo deste material é de suprir a falta das bibliografias básicas na sua versão digital com a disponibilização de forma digital e gratuita o que seria o material das notas das aulas da disciplina de Estatística Econômica. Não é o ideal, mas a ideia é melhorar o material com tempo.

Sobre o Autor

Professor do Departamento de Economia da Universidade Federal do Paraná. Engenheiro Agrônomo pela UNESP/Jaboticabal, Mestrado em Economia Agrária pela ESALQ/USP e Doutorado em Economia Aplicada pela ESALQ/USP, é um dos professores responsáveis pelas disciplinas de SE305 Estatística Econômica e Introdução à Econometria e SE308 Econometria ambas do curso de Economia da Universidade Federal do Paraná (UFPR).

Capítulo 1

Estatística Descritiva

O conteúdo deste tópico foi elaborado com base no capítulo 2 de Sartoris (2013) e nos capítulos 4 e 5 de Hoffmann (2006).

1.1 Medidas de posição

Trata-se de medidas de tendência central ou resumo. Como os nomes dizem, tratam-se de medidas que tratam de resumir a massa de valores e um único número.

1.1.1 Variável Aleatória

- variável aleatória (v.a.) é uma variável que está associada a uma *distribuição de probabilidade*.
- Ou seja, cada valor da v.a. está associada a uma probabilidade.
- O resultado do lançamento de uma dado, que poder ser qualquer número de 1 a 6, está associada a uma probabilidade de 1/6.

1.1.2 Média Aritmética Simples

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad (1.1)$$

onde $i = 1, \dots, n$

Exemplo 1

Qual é a média aritmética de um grupo de cinco pessoas cujas idades são em ordem crescente, 21, 23, 25, 28 e 31. Para responder, basta aplicar (1.1).

$$\bar{X} = \frac{21 + 23 + 25 + 28 + 31}{5} = 25,6$$

Exemplo 1 no R

```
X <- c(21, 23, 25, 28, 31)
X
```

```
## [1] 21 23 25 28 31
```

```
mediaX <- mean(X)
mediaX
```

```
## [1] 25,6
```

Exemplo 2

Qual é a média aritmética de três provas realizadas por um aluno, cujas notas foram 4,6 e 8. Para responder, basta aplicar (1.1).

$$\bar{X} = \frac{4 + 6 + 8}{3} = 6$$

Exemplo 2 no R

```
X2 <- c(4, 6, 8)
X2
```

```
## [1] 4 6 8
```

```
mediaX2 <- mean(X2)
mediaX2
```

```
## [1] 6
```

1.1.3 Média Aritmética Ponderada

Na média aritmética ponderada, cada valor pode ter importância diferentes dos outros valores considerados no computo. A frequência dos valores é muito comumente usada para dar maior ou menor importância relativa entre os valores considerados no computo da média aritmética ponderada. Veja como fica a fórmula para o cálculo da média aritmética ponderada em (1.2)

$$\bar{X} = \frac{1}{\sum_{i=1}^n w_i} \sum_{i=1}^n w_i X_i \quad (1.2)$$

onde w_i é a ponderação ou peso associado a i ésimo valor de X .

Podemos escrever na forma de frequência relativa dos valores da variável X :

$$f_i = \frac{w_i}{\sum_{i=1}^n w_i} \quad (1.3)$$

Exemplo 3

Qual é a média aritmética de um grupo de vinte alunos, oito com 22 anos, sete de 23 anos, três de 25 anos, um de 28 anos e um de 30 anos. Para responder, basta aplicar (1.2).

$$\overline{X} = \frac{22 \times 8 + 23 \times 7 + 25 \times 3 + 28 \times 1 + 30 \times 1}{20} = 23,5$$

Exemplo 3 no R

```
X3 <- c(22, 23, 25, 28, 30)
X3
```

```
## [1] 22 23 25 28 30
```

```
w3 <- c(8, 7, 3, 1, 1)
w3
```

```
## [1] 8 7 3 1 1
```

```
wX3 <- w3 * X3
mediaX3 <- sum(wX3)/sum(w3)
mediaX3
```

```
## [1] 23,5
```

Exemplo 4

Qual é a média ponderada de três provas realizadas por um aluno, cujas notas foram 4, 6 e 8. A primeira prova tem peso igual a 1, a segunda tem peso igual a 2 e a terceira tem peso igual a 3. Para responder, basta aplicar (1.2).

$$\overline{X} = \frac{4 \times 1 + 6 \times 2 + 8 \times 3}{1 + 2 + 3} \cong 6,7$$

Exemplo 4 no R

```
X4 <- c(4, 6, 8)
X4
```

```
## [1] 4 6 8
```

```
w4 <- c(1, 2, 3)
w4
```

```
## [1] 1 2 3
```

```
wX4 <- w4 * X4
mediaX4 <- sum(wX4)/sum(w4)
round(mediaX4, digits = 1)
```

```
## [1] 6,7
```

1.1.4 Média Geométrica Simples

Na média geométrica simples, a forma de obter uma medida resumo ou de tendência central é multiplicar todos os n valores e tirar a raiz enésima do resultado do produtório. Assim é possível ter duas fórmulas para a média geométrica a (1.4) e (1.5).

$$G = \left(\prod_{i=1}^n X_i \right)^{\frac{1}{n}} \quad (1.4)$$

ou

$$G = \sqrt[n]{X_1 \times X_2 \times \dots \times X_n} \quad (1.5)$$

O que acontece se um dos valores de X for igual a zero? E se um dos valores for negativo?

Exemplo 5

Sejam três valores 4, 6 e 8. Calcule a média geométrica simples.

$$\sqrt[3]{4 \times 6 \times 8} \cong 5,7690$$

Exemplo 5 no R

```
X5 <- c(4, 6, 8)
X5

## [1] 4 6 8

n <- length(X5)
mediaX5 <- prod(X5)^(1/n)
round(mediaX5, digits = 1)

## [1] 5,8
```

1.1.5 Média Geométrica Ponderada

Na média geométrica ponderada que podem ser calculadas através de duas fórmulas (1.6) e (1.7), cada valor pode ter uma importância diferente em relação aos outros valores no compute da média geométrica. Muito comumente, esta maior ou menor importância pode estar associada a frequência dos valores considerados no cálculo.

$$G = \left(\prod_{j=1}^k X_j^{w_j} \right)^{\frac{1}{n}} \quad (1.6)$$

ou

$$G = \sqrt[n]{X_1^{w_1} \times X_2^{w_2} \times \dots \times X_k^{w_k}} \quad (1.7)$$

onde a $\sum_{j=1}^k w_j = n$

Exemplo 6

tomando os valores do exemplo 5 e ponderando por 1,2 e 3, temos:

$$\sqrt[6]{4^1 \times 6^2 \times 8^3} \cong 6,5$$

O exemplo 6 no R

```
x6 <- c(4, 6, 8)
class(x6)

## [1] "numeric"
x6

## [1] 4 6 8
w6 <- c(1, 2, 3)
w6

## [1] 1 2 3
G2 <- round((prod(x6^w6))^(1/sum(w6)), 1)
G2

## [1] 6,5
```

1.1.6 Média Harmônica

É o inverso da média dos inversos dos valores da variável que pode ser calculada através das fórmulas (1.8) e (1.9).

$$H = \frac{n}{\sum_{i=1}^n \frac{1}{X_i}} \quad (1.8)$$

$$H = \frac{n}{\frac{1}{X_1} + \frac{1}{X_2} + \dots + \frac{1}{X_n}} \quad (1.9)$$

O que acontece se um dos valores de X for igual a zero? Para entender essa situação, use o conceito de limite fazendo o valor tender a zero.

Exemplo 7

Tomando o exemplo das notas, temos:

$$H = \frac{3}{\frac{1}{4} + \frac{1}{6} + \frac{1}{8}} \cong 5,5.$$

1.1.7 Média Harmônica Ponderada

Na média harmônica ponderada, assim como na média aritmética ponderada e na média geométrica ponderada, cada valor pode ter uma importância em relação aos outros valores considerados no seu cálculo. Comumente, a frequência do valor pode associar uma maior ou menor importância no cálculo da média harmônica ponderada que pode ser calculada através das fórmulas (1.10) e (1.11)

$$H = \frac{n}{\sum_{j=1}^k w_j \frac{1}{X_j}} \quad (1.10)$$

ou

$$H = \frac{n}{w_1 \frac{1}{X_1} + w_2 \frac{1}{X_2} + \dots + w_k \frac{1}{X_k}} \quad (1.11)$$

onde $\sum_{j=1}^k w_j = n$

Exemplo 8

Tomando o exemplo das notas

$$H = \frac{6}{\frac{1}{4} \times 1 + \frac{1}{6} \times 2 + \frac{1}{8} \times 3} \cong 6,3.$$

Observação

Tanto para as médias simples como para as ponderadas, a média aritmética é maior do que a média geométrica e essa, por sua vez, é maior que a harmônica. Isso só não vale quando todos os valores são iguais. Veja de forma esquemática em (1.12)

$$\overline{X} \geq G \geq H \quad (1.12)$$

Exemplo 9

O aluno tira as seguintes notas bimestrais: 3,4,5,7 e 8,5. Determine qual seria sua média final se esta fosse calculada dos três modos, aritmética, geométrica e harmônica, em cada um dos seguintes casos: i) as notas têm o mesmo peso e; ii) as notas têm pesos diferentes.

i) As notas dos bimestres têm os mesmos pesos.

$$\overline{X} = \frac{3 + 4,5 + 7 + 8,5}{4} = 23/4 = 5,75$$

$$G = \sqrt[4]{3 \times 4,5 \times 7 \times 8,5} = \sqrt[4]{803,25} \cong 5,32$$

$$H = \frac{4}{\frac{1}{3} + \frac{1}{4,5} + \frac{1}{7} + \frac{1}{8,5}} \cong 4,90$$

- ii) Suponha que agora os pesos para as notas bimestrais sejam, 30%, 25%, 25% e 20%.

$$\overline{X} = 0,3 \times 3 + 0,25 \times 4,5 + 0,25 \times 7 + 0,20 \times 8,5 = 5,475$$

$$G = 3^{0,3} \times 4,5^{0,25} \times 7^{0,25} \times 8,5^{0,2} \cong 5,05$$

$$H = \frac{1}{0,3\frac{1}{3} + 0,25\frac{1}{4,5} + 0,25\frac{1}{7} + 0,2\frac{1}{8,5}} \cong 4,66$$

1.1.8 Mediana

é o valor que divide um conjunto de dados ordenados ao meio, ou seja, dois grupos de valores de igual tamanho. Com base na definição de mediana, o valor da mediana pode ser obtida através da sua posição que proporciona duas situações: i) o número de valores é ímpar e ii) o número de valores é par.

- i) Quando o número de valores é ímpar, a posição do valor correspondente a mediana é obtida através de (1.13):

$$P\text{Mediana}_{\text{ímpar}} = \frac{n+1}{2} \quad (1.13)$$

onde n é o número de valores considerado no cálculo.

- ii) Quando o número de valores é par, a posição da mediana é obtida através da média entre os dois valores centrais do conjunto de valores ordenados de menor a maior. O primeiro valor central é definido pela posição obtida através de (1.14)

$$P1\text{Mediana}_{\text{par}} = \frac{n}{2} \quad (1.14)$$

onde n é o número de valores considerado para o cálculo.

O segundo valor central é definido pela posição obtida através de (1.15)

$$P2\text{Mediana}_{\text{par}} = \frac{n}{2} + 1 \quad (1.15)$$

onde n é o número de valores considerado para o cálculo.

Assim, a mediana quando o número de valores é par é obtida através da média aritmética simples dos valores correspondentes às posições obtidas por (1.14) e por (1.15) através de (1.16)

$$Mediana_{par} = \frac{ValorCentral_1 + ValorCentral_2}{2} \quad (1.16)$$

Exemplo numérico de Mediana quando o número de valores é ímpar

Seja um conjunto de valores 2,-3,1,-2,0,-1,3. Obtenha a mediana.

Primeiramente ordena-se do menor para o maior.

-3,-2,-1,0,1,2,3

Como se trata de número ímpar de valores o valor central que divide o conjunto de valores em dois subconjuntos de igual tamanho é o valor da mediana. Neste caso é o zero.

Mediana no R

```
w <- c(-3, -2, -1, 0, 1, 2, 3)
mediana1 <- median(w)
print(mediana1)
```

```
## [1] 0
```

Exemplo numérico de Mediana quando o número de valores é par

No exemplo anterior o conjunto de dados era composto por um número ímpar de valores. Neste exemplo o número de valores ordenado de menor a maior é par. Nesse caso, apesar de existir vários critérios, o mais usual é tirar a média aritmética simples entre os dois valores centrais do conjunto de valores ordenados de menor a maior. Uma vez que não existe um valor que separe dois subconjuntos de igual tamanho, a média aritmética simples destes dois valores é o valor da mediana quando o número total de valores não é ímpar.

Sejam os valores -2,1,3,2,-3,1. Obtenha a mediana.

Primeiramente ordena-se os seis valores.

-3,-2,-1,1,2,3

Note que trata-se de conjunto com um número par de valores.

Dessa forma, toma-se os dois valores centrais que são -1 e 1 e calcula-se a média aritmética simples. Ou seja, a mediana para este conjunto com seis valores é igual a zero.

O exemplo do número par de valores no R

```
v <- c(-3, -2, -1, 1, 2, 3)
mediana2 <- median(v)
print(mediana2)
```

```
## [1] 0
```

1.1.9 Quartis ou Quartiles

são os valores que dividem o conjunto de dados ordenados em quatro subconjuntos de igual tamanho. Ou seja são valores do conjunto que definem o primeiro quarto dos dados (25%), a metade dos dados (50%) que coincide com a mediana, os três quartos dos dados (75%).

Dessa forma para obter os valores que dividem o conjunto de dados ordenados de menor a maior e quatro subconjuntos de igual tamanho, é necessário definir qual é a posição desses valores. Uma vez definido as suas posições pode-se obter os valores corretamente.

A posição do valor que separa o primeiro do segundo quartil é definido por (1.17).

$$PQ_1 = \frac{(n+1)}{4} \quad (1.17)$$

onde n é o número de valores. A posição do valor que separa o segundo do terceiro quartil é definido por (1.18).

$$PQ_3 = \frac{3(n+1)}{4} \quad (1.18)$$

onde n é o número de valores.

Note que o termo genérico é percentil. Por exemplo, o quintis são os valores que dividem o conjunto de dados ordenados de menor a maior em cinco subconjuntos de igual tamanho.

Quartis no R

No R tem uma função específica para a obtenção dos quartis.

```
p <- c(0:100)
length(p)

## [1] 101

quantile(p)

##    0%   25%   50%   75%  100%
##     0    25    50    75   100

faixainterquant <- quantile(p, 0.75) - quantile(p,
  0.25)
faixainterquant

## 75%
## 50
```

Quartis no R

No R tem uma função específica para a obtenção dos quartis.

```

p2 <- c(1:100)
length(p2)

## [1] 100
quantile(p2)

##      0%      25%      50%      75%     100%
##    1,00  25,75  50,50  75,25 100,00
faixainterquant2 <- quantile(p2, 0.75) - quantile(p2,
  0.25)
faixainterquant2

## 75%
## 49,5

```

1.1.10 Moda

Moda é o elemento de maior frequência, ou seja, que aparece o maior número de vezes. Pode haver mais de uma moda em um conjunto de valores:

- Unimodal
- Bimodal
- Multimodal
- Amodal

Moda no R Não existe uma função da moda para pronto uso no R. É necessário criar uma função segue abaixo.

```

# criando a função moda no R

getmode <- function(v) {
  uniqv <- unique(v)
  uniqv[which.max(tabulate(match(v, uniqv)))]
}
z <- c(2, 1, 2, 3, 1, 2, 3, 4, 1, 5, 5, 3, 2, 3)
moda1 <- getmode(z)
print(modal)

## [1] 2

```

1.2 Medidas de dispersão

Este tópico está baseado nos materiais de Hoffmann (2006), Morettin and Bussab (2013) e Sartoris (2013).

As medidas de dispersão medem como os dados estão **agrupados**, mais ou menos próximos entre si, ou seja, mais ou menos dispersos.

1.2.1 Amplitude

A amplitude de um conjunto de valores é a diferença entre o maior elemento e o menor elemento desse conjunto.

1.2.2 Variância

A variância é a somatória dos quadrados dos desvios em relação a média, dividido pelo número de observações. Note que a ideia inicial de dispersão foi a distância de cada valor do conjunto dados da variável em relação à média da variável. Mas como trata-se da distância relativa de cada valor em relação a média dos valores da variável, a sua somatória sempre resulta zero. Pois os desvios em relação a médias são compostos de valores positivos e negativos por estarem acima ou abaixo da média e assim a somatória das mesmas resulta zero, sempre. Portanto, a soma dos desvios não tem utilidade como medida de dispersão. Mas se a soma for dos quadrados dos desvios, isso é resolvido. Por isso, a variância é o valor médio dos quadrados dos desvios em relação à média. Ou seja,

$$var(X) = \sigma^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n} \text{ (população)}$$

$$var(X) = \sigma^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1} \text{ (amostra)}$$

Note que a variância da amostra é um estimador não viesado da variância populacional. A diferença entre variância populacional e variância amostral será apresentada mais adiante. A interpretação intuitiva da diferença entre ambas as variâncias é de que quando se trabalha com amostra, está tendo acesso a parte das informações e isso precisa ser penalizado. Note que esta penalização se dá para amostras pequenas pois se trata de subtrair uma unidade do número de observações. O que acontece com a diferença entre variância populacional e a variância amostral quando i) o número de observações torna-se muito grande, tipo bem maior que 30 e; ii) o número de observações tende ao infinito.

Variância no R

Aproveitando os dados de alturas de 30 pessoas:

```
tail(X)
```

```
## [1] 21 23 25 28 31
```

```
varX <- round(var(X), 4)
varX
```

```
## [1] 15,8
```

Note que a variância no R é a variância amostral, cujo denominador é $(n - 1)$.

Em termos práticos, a variância tem uma desvantagem: a unidade do seu resultado é o quadrado da unidade original da variável. Portanto, se a variável em questão é preço de uma mercadoria em Reais, a sua variância será Reais ao quadrado. Tal fato dificulta a sua interpretação. Por isso é apresentado o desvio padrão como medida de dispersão na sequência.

1.2.3 Desvio Padrão

É a raiz quadrada da variância. No desvio padrão, denotado como $d.p.(X)$ ou σ , não tem o efeito do quadrado.

$$d.p.(X) \cong \sigma = \sqrt{var(X)}$$

Portanto, a sua interpretação é clara e direta por ter a mesma unidade da sua variável original. Desta forma, o desvio padrão facilita a sua análise juntamente com as medidas de posição como a média aritmética simples, por exemplo.

Desvio Padrão no R

```
tail(X)
```

```
## [1] 21 23 25 28 31
```

```
dpX <- round(sd(X), 4)
dpX
```

```
## [1] 3,9749
```

Note que, da mesma forma que a variância no R, o desvio padrão calculado no R tem como base a variância cujo numerador é $(n - 1)$.

Fórmula alternativa da Variância

Desenvolvendo a fórmula da definição da variância tem-se:

$$\begin{aligned}
 \text{var}(X) &= \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \\
 &= \frac{1}{n} \sum_{i=1}^n (X_i^2 - 2X_i\bar{X} + \bar{X}^2) \\
 &= \frac{1}{n} \sum_{i=1}^n X_i^2 - \frac{1}{n} \sum_{i=1}^n 2X_i\bar{X} + \frac{1}{n} \sum_{i=1}^n \bar{X}^2 \\
 &= \frac{1}{n} \sum_{i=1}^n X_i^2 - 2\bar{X} \frac{1}{n} \sum_{i=1}^n X_i + \frac{1}{n} n\bar{X}^2 \\
 &= \frac{1}{n} \sum_{i=1}^n X_i^2 - 2\bar{X}\bar{X} + \bar{X}^2 \\
 &= \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2.
 \end{aligned}$$

Em outras palavras

$$\text{var}(X) = \text{média dos quadrados} - \text{quadrado da média}.$$

Exemplo de variância e desvio padrão no R

Tomando o exemplo numérico da tabela 2.7 (Sartoris, 2013, p.40) sobre notas do aluno A tem-se:

| Aluno A | notas | notas^2 |
|------------------|------------|------------------|
| Economia | 3 | 9 |
| Contabilidade | 2 | 4 |
| Administração | 4 | 16 |
| Matemática | 1 | 1 |
| Somatória | 10 | 30 |
| Média | 2,5 | 7,5 |

$$\text{var}(X) = 7,5 - (2,5)^2 = 1,25$$

$$\text{dp}(X) = \sqrt{1,25} = 1,12$$

```
X3 <- c(3, 2, 4, 1)
mediaX3e2 <- sum(X3^2)/length(X3)
```

```
mediaX3 <- sum(X3)/length(X3)
varX3 <- mediaX3e2 - mediaX3^2
varX3
```

```
## [1] 1,25
```

```
dpX3 <- round(sqrt(varX3), 4)
dpX3
```

```
## [1] 1,118
```

1.2.4 Desvio Absoluto Médio

Por definição, o desvio absoluto médio de um conjunto de dados

$$X_1, X_2, \dots, X_n$$

é a média aritmética dos valores absolutos dos desvios em relação a média de X_i ,

$$\delta = \frac{1}{n} \sum_{i=1}^n |X_i - \bar{X}|.$$

Se os valores de X estiverem em ordem crescente

$$X_1 \leq X_2 \leq \dots, X_n,$$

sendo válida pelos menos uma desigualdade, e se h é um inteiro positivo tal qual que

$$X_i < \bar{X} \text{ para } 1 \leq i \leq h$$

e

$$X_i \geq \bar{X} \text{ para } h < i \leq n,$$

temos

$$\delta = \frac{1}{n} \left[- \sum_{i=1}^h (X_i - \bar{X}) + \sum_{i=h+1}^n (X_i - \bar{X}) \right]$$

Como a soma dos n desvios em relação á média \bar{X} é zero

$$\sum_{i=1}^n (X_i - \bar{X}) = 0$$

segue-se que

$$\sum_{i=h+1}^n (X_i - \bar{X}) = - \sum_{i=1}^h (X_i - \bar{X}).$$

Substituindo (1.2.4) em (1.2.4) se tem

$$\delta = \frac{2}{n} \sum_{i=h+1}^n (X_i - \bar{X}) = \frac{2}{n} \sum_{i=1}^h (\bar{X} - X_i)$$

Note que se pode economizar cálculos se optar por uma das fórmulas do desvio absoluto médio dado por (1.2.4). O desvio absoluto médio será utilizado mais adiante para definir a fórmula da discrepância máxima na seção sobre o índice de Gini.

1.2.5 Diferença Absoluta Média

Por definição, a diferença absoluta média de um conjunto de dados

$$X_1, X_2, \dots, X_n$$

é dado por

$$\Delta = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n |X_i - X_j|.$$

Se

$$X_1 \leq X_2 \leq \dots, X_n$$

é possível escrever os valores de $X_i - X_j$, com $i = 1, \dots, n$ e $j = 1, \dots, n$, com segue

$$\begin{array}{ccccc} X_n - X_n & X_n - X_{n-1} & \cdots & X_n - X_2 & X_n - X_1 \\ X_{n-1} - X_{n-1} & X_{n-1} - X_{n-1} & \cdots & X_{n-1} - X_2 & X_{n-1} - X_1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ X_n - X_2 & X_{n-1} - X_2 & \cdots & X_2 - X_2 & X_1 - X_1 \\ X_n - X_1 & X_{n-1} - X_1 & \cdots & X_2 - X_1 & X_1 - X_1 \end{array}$$

Coletando os termos positivos e os negativos, se tem

$$\begin{aligned} \Delta &= \frac{1}{n^2} [X_1 + 3X_2 + \dots + (2n-3)X_{n-1} + (2n-1)X_n] - \\ &\quad - \frac{1}{n^2} [X_n - 3X_{n-1} + \dots + (2n-3)X_2 + (2n-1)X_1] \end{aligned}$$

Note que o segundo membro de (1.2.5) (lado direito) é a soma algébrica de duas parcelas de sinais contrários, seu valor não se altera se for adicionado a cada uma das expressões entre colchetes a expressão

$$[X_n + 3X_{n-1} + \dots + (2n-3)X_2 + (2n-1)X_1].$$

Dessa forma se obtém

$$\begin{aligned} \Delta &= \frac{1}{n^2} 2n [X_1 + X_2 + \dots + X_{n-1} + X_n] - \\ &\quad - \frac{2}{n^2} [X_n + 3X_{n-1} + \dots + (2n-3)X_2 + (2n-1)X_1]. \end{aligned}$$

Como

$$\sum_{i=1}^n X_i = n\bar{X}$$

se tem

$$\Delta = 2\bar{X} - \frac{2}{n^2} [(2n-1)X_1 + (2n-3)X_2 + \dots + 3X_{n-1} + X_n].$$

Somando e subtraindo

$$\sum_{i=1}^n X_i = n\bar{X}$$

à expressão entre colchetes em (1.2.5) se tem

$$\Delta = 2\bar{X} \left(1 + \frac{1}{n}\right) - \frac{4}{n^2} [nX_1 + (n-1)X_2 + \dots + 2X_{n-1} + X_n]$$

ou

$$\Delta = 2\bar{X} \left(1 + \frac{1}{n}\right) - \frac{4}{n^2} \sum_{i=1}^n (n-i+1)X_i.$$

Com mais algumas transformações algébricas de (1.2.5) se tem

$$\Delta = \frac{4}{n^2} \sum_{i=1}^n iX_i - 2\bar{X} \left(1 + \frac{1}{n}\right)$$

Note que a fórmula da diferença média absoluta será utilizada mais adiante para a obtenção da fórmula do índice de Gini.

1.2.6 Histograma

O histograma é uma ferramenta da estatística descritiva para mostrar visualmente, de forma bastante simples, como os valores da variável estão distribuídos. Mas também permite ter uma ideia visual da dispersão do conjunto de valores. Portanto, não se trata de uma medida de dispersão. Mas deveria ser a primeira coisa a se obter das variáveis de interesse em um trabalho de pesquisa.

Considere a altura de 30 pessoas medidas em centímetros.

Tabela 2.1 - Altura de 30 pessoas em cm.

| | | | | |
|-----|-----|-----|-----|-----|
| 159 | 168 | 172 | 175 | 181 |
| 161 | 168 | 173 | 176 | 183 |
| 162 | 169 | 173 | 177 | 185 |
| 164 | 170 | 174 | 178 | 190 |

| | | | | |
|-----|-----|-----|-----|-----|
| 166 | 171 | 174 | 179 | 194 |
| 167 | 171 | 174 | 180 | 201 |

Usando o R para construir o histograma do exemplo numérico

Os dados são inputados na variável X.

```
X <- c(159, 161, 162, 164, 166, 167, 168, 168, 169,
      170, 171, 171, 172, 173, 173, 174, 174, 174, 175,
      176, 177, 178, 179, 180, 181, 183, 185, 190, 194,
      201)
head(X)
```

```
## [1] 159 161 162 164 166 167
```

```
obsX <- length(X)
obsX
```

```
## [1] 30
```

```
faixaX <- range(X)
faixaX
```

```
## [1] 159 201
```

Usando a função hist do R para elaborar o histograma de altura

```
grafico1 <- hist(
  X,
  main="Histograma da Altura",
  xlab="cm",
  ylab="frequência",
  border="blue",
  col="green",
  xlim=c(150,210),
  las=1,
  breaks=5,
  right=FALSE
)

grafico1
```

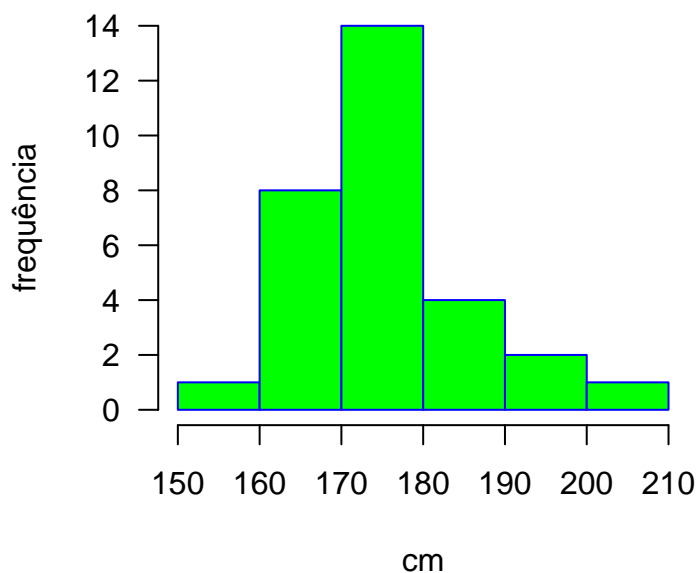
onde

- **main**="Histograma da Altura de 30 pessoas" título do histograma
- **xlab**="cm" rótulo do eixo horizontal
- **ylab**="frequência" rótulo do eixo vertical
- **border**="blue" cor do contorno das barras

- **col="green"** cor das barras
- **xlim=c(150,210)** limite inferior e superior
- **las=1** rotacao do rotulo dos numeros do eixo vertical
- **breaks=5** número de classes
- **right=FALSE** define intervalo do tipo $[a,b)$, se FALSE, e $(a,b]$, se TRUE.

Obtendo o histograma

Histograma da Altura de 30 pessoas



O pacote **ggplot2** gera gráficos e histogramas melhor elaborados.

Obtendo o histograma usando uma forma alternativa

Agrupando essas pessoas em **classes** de 10 cm temos:

| classes | frequência |
|-------------|------------|
| [150 ; 160[| 1 |
| [160 ; 170[| 8 |
| [170 ; 180[| 14 |
| [180 ; 190[| 4 |
| [190 ; 200[| 2 |
| [200 ; 210[| 1 |

Fazendo isso no R:

```

nobs <- c(1:30)
dataX <- as.data.frame(cbind(nobs, X))
# transformando em data frame
tail(dataX)

##      nobs    X
## 25    25 181
## 26    26 183
## 27    27 185
## 28    28 190
## 29    29 194
## 30    30 201

# mostrando as seis últimas observações
quebras <- seq(150, 210, by = 10)
# definindo os intervalos
quebras

## [1] 150 160 170 180 190 200 210

dataX.cut <- cut(dataX$X, quebras, right = FALSE)
# construindo as classes fechado a esq e aberto a
# direita
dataX.freq <- table(dataX.cut)
# obtendo a frequência para cada classe.
dataXfreq <- cbind(dataX.freq)
# colocando os dados em colunas
dataXfreq

##           dataX.freq
## [150,160)           1
## [160,170)           8
## [170,180)          14
## [180,190)           4
## [190,200)           2
## [200,210)           1

```

1.2.7 Diagrama de caixa (Boxplot)

O texto sobre o diagrama de caixa foi baseado em Morettin and Bussab (2013).

Boxplot ou caixa de bigode também é uma ferramenta da estatística descritiva que permite visualizar a dispersão dos valores da variável em análise. O que define o diagrama de caixa são os quartis. A parte inferior e superior da caixa, são respectivamente o primeiro quartil (Q_1) e o terceiro quartil (Q_3). A linha que corta da caixa é a mediana ou o segundo quartil (Q_2). Os bigodes que são as linhas que se estendem a partir da caixa, são calculado com base na amplitude

interquartil (AIQ). A amplitude interquartil é a diferença entre os valores do terceiro e do primeiro quartis. Ou seja,

$$AIQ = Q_3 - Q_1$$

O bigode inferior denominado LI é calculado subtraindo $1,5 \times AIQ$ do valor do primeiro quartil Q_1 . Ou seja,

$$LI = Q_1 - 1,5 \times AIQ$$

O bigode superior, denominado LS , é calculado somando $1,5 \times AIQ$ ao valor da terceiro quartil Q_3 . Ou seja,

$$LS = Q_3 + 1,5 \times AIQ$$

Os valores que forem menor que o LI ou maior que o LS são denominados valores discrepantes ou *outliers*. Os valores discrepantes, quando existentes, são colocados separadamente no diagrama de caixa mantendo a distancia relativa do limite inferior ou do limite superior.

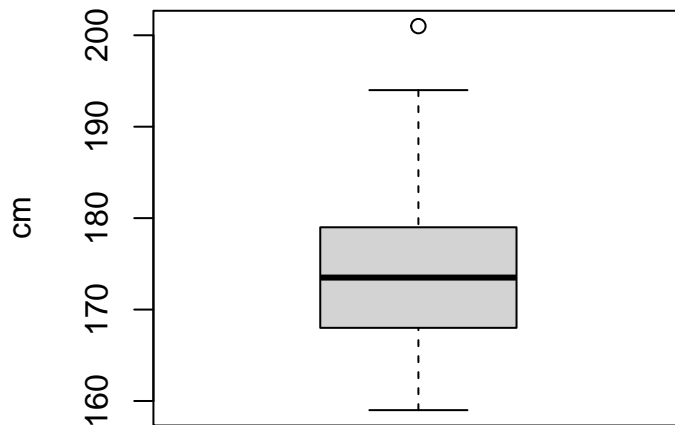
Toma-se o mesmo exemplo da altura de 30 pessoas para apresentar o boxplot.

O código seria:

```
boxplot(X, data = dataX, main = "Diagrama de Caixa",  
        ylab = "cm", xlab = "altura de 30 pessoas")
```

e o resultado segue abaixo.

Diagrama de Caixa



altura de 30 pessoas

1.3 Medidas de relação linear entre duas variáveis

Este assunto tem como base o material de Sartoris (2013).

Parece um pouco estranho incluir esse tópico logo depois das medidas de dispersão. Mas a variância é um caso especial da covariância que é a primeira medida de relação linear entre duas variáveis.

O coeficiente de correlação utiliza a covariância e o desvio padrão para resolver o problema de interpretação do resultado da covariância.

1.3.1 Covariância

pode ser estendida como uma *variância conjunta* entre duas variáveis. Ou seja,

$$cov(X, Y) = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})$$

Fórmula alternativa da Variância

Também existe a fórmula alternativa da covariância.

$$\text{cov}(X, Y) = \frac{1}{n} \sum_{i=1}^n X_i Y_i - \overline{XY}.$$

Fórmula alternativa da Covariância

Em outras palavras

$$\text{cov}(X, Y) = \text{média dos produtos de X e Y} - \text{produto das médias de X e Y}.$$

Covariância no R

Tomando o exemplo de consumo e renda da tabela 2.11 (Sartoris, 2013, p.42) tem-se

| Ano | Consumo(X) | Renda(Y) | (XY) |
|------------------|--------------|--------------|------------------|
| 1 | 600 | 1.000 | 600.000 |
| 2 | 700 | 1.100 | 770.000 |
| 3 | 800 | 1.300 | 1.040.000 |
| 4 | 900 | 1.400 | 1.260.000 |
| Somatória | 3.000 | 4.800 | 3.670.000 |
| Média | 750 | 1.200 | 917.500 |

Covariância no R

```
C1 <- c(600, 700, 800, 900)
R1 <- c(1000, 1100, 1300, 1400)
mediaC1 <- sum(C1)/length(C1)
mediaR1 <- sum(R1)/length(R1)
mediaC1R1 <- sum(C1 * R1)/length(C1)
covC1R1 <- mediaC1R1 - mediaC1 * mediaR1
covC1R1
```

```
## [1] 17500
```

```
cov(C1, R1)
```

```
## [1] 23333,33
```

Note que a função covariância no R é calculada dividindo por $(n - 1)$ e não por n .

1.3.2 Coeficiente de Correlação

É obtido dividindo a covariância pelos desvios padrões das variáveis, retirando-se o efeito dos valores de cada variável. Como as unidades das variáveis se cancelam

matematicamente, o coeficiente de correlação é um número puro que varia entre -1 e +1. Essa característica o torna mais fácil e claro a sua interpretação. Ou seja,

$$\text{corr}(X, Y) \cong \rho_{xy} = \frac{\text{cov}(X, Y)}{dp(X) \times dp(Y)}$$

onde

$$-1 \leq \rho \leq +1$$

Portanto, quando o coeficiente de correlação é igual a zero ou muito próximo a zero, significa que as duas variáveis analisadas não tem relação do tipo linear entre elas. Quando o coeficiente de correlação é igual a -1 ou próximo de -1, tal fato indica que a existência de uma relação do tipo linear entre as duas variáveis analisadas, sendo que as variações ocorrem no sentido oposto. Ou seja, quando uma das variáveis aumenta de valor, a outra diminui. Quando o coeficiente de correlação é igual a +1 ou muito próximo de um positivo, tal fato indica que as duas variáveis tem uma relação do tipo linear, sendo que as variações em ambas as variáveis ocorrem no mesmo sentido. Ou seja, quando uma das variáveis aumenta de valor, a outra aumenta também. O que significa o coeficiente de correlação ser: i) exatamente igual a zero; ii) ser exatamente igual a -1 e; exatamente igual a +1?

Correlação no R

```
medC1 <- sum(C1)/length(C1)
medR1 <- sum(R1)/length(R1)
varC1 <- (sum((C1 - medC1)^2))/length(C1)
varC1

## [1] 12500

varR1 <- (sum((R1 - medR1)^2))/length(R1)
varR1

## [1] 25000

dpC1 <- abs(sqrt(varC1))
dpR1 <- abs(sqrt(varR1))
corrC1R1 <- round(covC1R1/(dpC1 * dpR1), 4)
corrC1R1

## [1] 0,9899

Ou simplesmente

round(cor(C1, R1), 4)

## [1] 0,9899
```


Capítulo 2

Medidas de desigualdade

O assunto sobre medidas de desigualdade está baseada na sua totalidade no capítulo 17 de Hoffmann (2006)

2.1 Princípio de Pigou-Dalton

A condição de Pigou-Dalton define que as medidas de desigualdades devem ter seus valores aumentados quando há transferência regressivas de renda. Para entender a condição de Pigou-Dalton, considere uma população com apenas duas pessoas cujas rendas são X_1 e X_2 . Então, $\mu = \frac{X_1 + X_2}{2}$. No caso de perfeita igualdade, $X_1 = X_2 = \mu$. No caso de uma distribuição com $X_1 \neq X_2$, uma transferência de renda do mais pobre para o mais rico, mantendo a renda média constante, aumenta o grau de desigualdade.

2.2 Transferência Regressiva

Essa transferência de renda do mais pobre para o mais rico, mantida a renda média constante, é denominada como **transferência regressiva** de renda. Portanto, uma **transferência progressiva** é a transferência de renda do mais rico para o mais pobre.

2.3 Curva de Lorenz

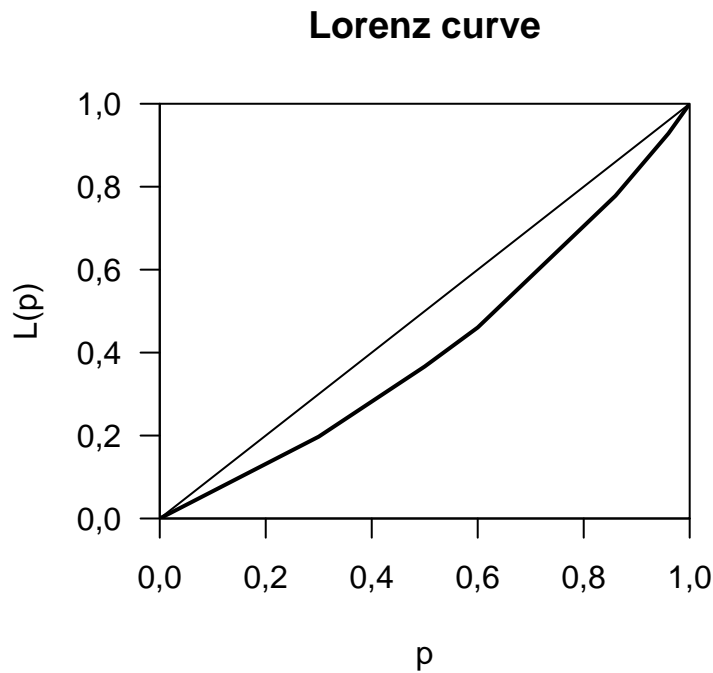
Tabela 2.1: Distribuição de pessoas ocupadas conforme renda obtida na atividade exercida no Brasil, de acordo com a PNAD 2003

| estrato | % no estrato da população (%) | % no estrato da renda (%) | % acumulada da população (100p) | % acumulada da renda (100Φ) |
|---------|-------------------------------|---------------------------|---------------------------------|-----------------------------|
| I | 30 | 7 | 30 | 7 |
| II | 20 | 9 | 50 | 16 |
| III | 20 | 13 | 70 | 29 |
| IV | 10 | 10 | 80 | 39 |
| V | 10 | 16 | 90 | 55 |
| VI | 5 | 13 | 95 | 68 |
| VII | 4 | 19 | 99 | 87 |
| VIII | 1 | 13 | 100 | 100 |

Considere os dados da tabela 2.1. Na coluna de porcentagem acumulada podemos observar que 70% da população possui 29% da renda. Os percentuais acumulados da população p e da renda Φ formam um plano cartesiano (p, Φ) originando a Cuirva de Lorenz.

```
library(ineq)
# usando os valores do exemplo em porcentagem
# mesmo
p <- c(30, 20, 20, 10, 10, 5, 4, 1)
r <- c(7, 9, 13, 10, 16, 13, 19, 13)

# calcula o mínimo da curva de Lorenz
Lcmin <- Lc(r, n = p)
# Desenha a curva de Lorenz em um gráfico
plot(Lcmin)
```



Considerando a curva de Lorenz, figura 2.1, que é basicamente a obtida pelo R, figura ??, mas com algumas indicações, é possível obter algumas definições.

```
knitr::include_graphics("lorenz3.png")
```

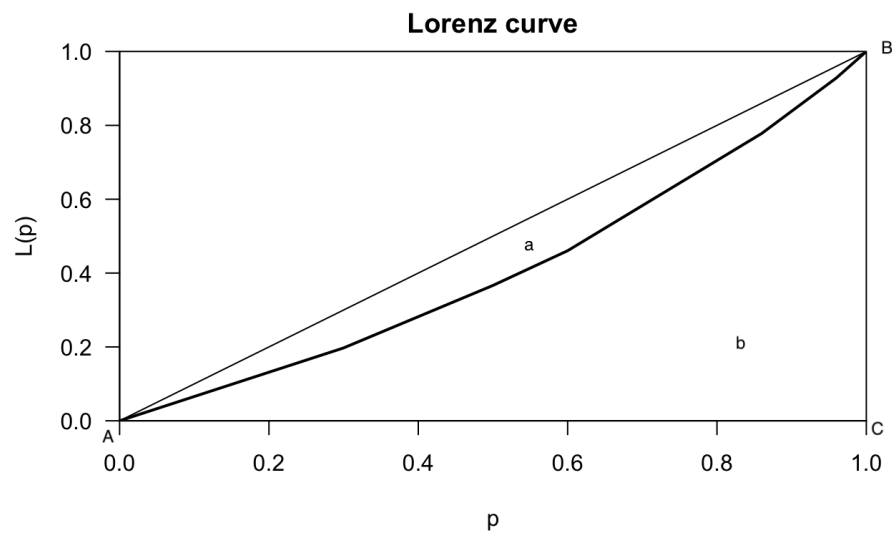


Figura 2.1: A curva de Lorenz com algumas indicações

A área que corresponde a letra a é denominada área de desigualdade. o seguimento de retas \overline{AB} é chamado de *linha de perfeita igualdade* onde $p = \Phi$ e a área de de desigualdade é zero.

Analisando o casos de máxima desigualdade:

- excluindo-se o fato de renda negativa, considere que apenas um de n indivíduos receba toda a renda e os demais $n - 1$ indivíduos recebam zero de renda.
- Neste caso a porcentagem de renda é zero até o ponto $\frac{n-1}{n}$ no eixo horizontal, tornando-se $\Phi = 1$ ao se incluir o último indivíduo.
- Neste caso, a Curva de Lorenz é dada pela poligonal \widehat{ABC} e a área de desigualdade máxima é o triângulo ABC .

2.4 Índice Gini

Considere os dados da tabela 2.1. Seja p o valor da proporção acumulada da população até certo estrato e seja Φ o valor da correspondente proporção acumulada da renda. Os pares de valores (p, Φ) , para os diversos estratos, definem pontos em um sistema de eixos cartesianos como aparece na figura 2.1. Estes pontos estão sobre a curva de Lorenz, que mostra como a proporção acumulada da renda (Φ) varia em função da proporção acumulada da população (p), com as pessoas ordenadas de acordo com valores crescentes da renda. A área correspondente a a que está entre a reta AB e a curva de Lorenz na figura 2.1, é denominada **área de desigualdade**.

Para entender como ocorre a variação desta área de desigualdade, a área a , primeiro considere uma situação de distribuição de renda com perfeita igualdade, ou seja, uma população em que todos recebem a mesma renda. Nesta situação, a uma população p da população corresponde uma igual proporção Φ da renda total, ou seja, $\Phi = p$. Portanto, a curva de Lorenz dessa distribuição coincide com a reta AB da figura 2.1, denominado, por isso, de **linha de perfeita igualdade**. Neste caso a área de desigualdade é igual a zero.

Considere agora uma outra situação, uma distribuição de renda com o máximo de desigualdade. Considerando que **não** existe a possibilidade de renda negativa, esse seria o caso de uma população com n pessoas, em que uma delas recebe toda a renda e as $n - 1$ restante receba zero de renda. Nesta situação, a proporção acumulada da renda é igual a zero até o ponto do eixo horizontal (abscissa) $\frac{(n-1)}{n}$, tornando-se $\Phi = 1$ quando se se inclui a pessoa que recebe toda a renda. Neste caso, a curva de Lorenz passa a ser a poligonal ABC da figura 2.1. Que é numericamente igual a 0,5 (Por quê?).

Por definição, o **índice de Gini (G)** é uma relação entre a área de desigualdade, indicada por a que passar a ser denominada de α , e a área do triângulo ABC que é numericamente igual a 0,5, ou seja,

$$G = \frac{\alpha}{0,5} = 2\alpha$$

A fórmula (2.4) é uma das fórmulas de Gini que tem utilidade do ponto de vista teórico. Uma vez que

$$0 \leq \alpha \leq 0,5$$

tem-se que

$$0 \leq G \leq 1$$

Ou seja de que o índice de Gini varia entre zero, ausência de desigualdade, e um, máxima desigualdade. Adicionalmente, o índice de Gini é um número adimensional.

Uma fórmula alternativa e mais prática do ponto de vista do cálculo do índice de Gini pode ser obtida considerando-se uma distribuição discreta.

Seja uma variável aleatória discreta X_i para $i = 1, \dots, n$, cujos valores estão em ordem crescente

$$X_1 \leq X_2 \leq \dots \leq X_{n-1} \leq X_n$$

admitindo-se que os n valores são igualmente prováveis.

a proporção acumulada do número de elementos, até o i -ésimo elemento, é

$$p_i = \frac{i}{n}, \text{ para } i = 1, \dots, n$$

A correspondente proporção acumulada de X , até o i -ésimo elemento é

$$\Phi_i = \frac{\sum_{j=1}^i X_j}{\sum_{j=1}^n X_j} = \frac{1}{n\mu} \sum_{j=1}^i X_j$$

onde

$$\mu = \frac{1}{n} \sum_{j=1}^n X_j$$

Se X representa a renda individual e se $X_j < X_{j+1}$, Φ_i representa a fração da renda total apropriada pelas pessoas com renda inferior ou igual a X_i . As expressões (2.4) e (2.4) definem as coordenadas (p_i, Φ_i) com $i = 1, \dots, n$ de n pontos da curva de Lorenz. A rigor não existe, nesse caso, uma curva, mas uma poligonal cujos vértices são a origem dos eixos e os pontos de coordenadas (p_i, Φ_i) .

Na sequência é apresentada de forma resumida como se calcula o índice de Gini a partir dos valores de X_i para $i = 1, \dots, n$ da variável. Na figura 2.1 a soma das áreas a e b totaliza a área do polígono ABC que numericamente é igual a 0,5. Portanto, $a = 0,5 - b$. Ou seja, colocando na notação mais elegante,

$$\alpha = 0,5 - \beta.$$

Substituindo (2.4) em (2.4) obtém-se

$$G = \frac{0,5 - \beta}{0,5} = 1 - 2\beta.$$

Note que a área abaixo da “curva” de Lorenz pode ser representada, de forma aproximada, como a soma das áreas de n trapézios um do lado do outro. Desta forma, a área b da figura 2.1, compreendida entre a poligonal de Lorenz e o eixo das abscissas, é obtida somando-se a área dos n trapézios. Ou seja, a área do i -ésimo trapézio é

$$S_i = \frac{\Phi_{i-1} + \Phi_i}{2} \times \frac{1}{n}$$

onde Φ_{i-1} é a base menor do i -ésimo trapézio; Φ_i é a base maior do i -ésimo trapézio e; $1/n$ é a altura do trapézio que corresponde a pessoa da população composta por n pessoas.

Note que o valor de $\Phi_0 = 0$, ou seja, o valor de Φ_{i-1} para $i = 1$. Com base na fórmula (2.4) é possível obter a área corresponde a b na figura 2.1 ou β nas notações matemáticas no texto

$$\beta = \sum_{i=1}^n S_i = \frac{1}{2n} \sum_{i=1}^n (\Phi_{i-1} + \Phi_i)$$

Substituindo (2.4) em (2.4), obtém-se

$$G = 1 - \frac{1}{n} \sum_{i=1}^n (\Phi_{i-1} + \Phi_i)$$

Considerando a fórmula (2.4) e que $\Phi_0 = 0$, se obtém o índice de Gini em termos da variável X_i . Ou seja,

$$G = 1 - \frac{1}{n^2\mu} [(2n-1)X_1 + (2n-3)X_2 + \dots + 3X_{n-1} + 1X_n]$$

Na parte sobre Estatística Descritiva foi apresentada a medida de dispersão chamada Diferença Absoluta Média que é dada por

$$\Delta = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n |X_i - X_j|$$

Trabalhando algebricamente a fórmula (2.4)

$$\Delta = 2\mu - \frac{2}{n^2}[(2n-1)X_1 + (2n-3)X_2 + \cdots + 3X_{n-1} + 1X_n]$$

Se dividir (2.4) por 2μ obtém a fórmula do índice de Gini em termos da medida de dispersão Diferença Absoluta Média

$$G = \frac{\Delta}{2\mu}$$

Tratando-se da distribuição da renda em uma população, a relação (2.4) mostra que o índice de Gini, como medida de do grau de desigualdade, apresenta a vantagem de medir diretamente as diferenças de rendal, levando em consideração diferenças entre as rendas de **todos** os pares de pessoas.

Como Δ é uma medida de dispersão da distribuição, a relação (2.4) mostra que o índice de Gini é uma medida de dispersão relativa. Assim, o conceito de desigualdade de uma distribuição se confunde com o conceito de dispersão relativa.

Com um desenvolvimento algébrico de Δ é possível transformar a fórmula (2.4) em

$$G = \frac{2}{n^2\mu} \sum_{i=1}^n iX_i - \frac{1}{n} - 1$$

A relação (2.4) mostra que, no cálculo do índice de Gini, cada valor de X_i da variável aparece poderado por i . Ou seja, X_i aparece poderada pelo respectivo número de ordem na sequência dos valores ordenados.

Exemplo numérico

Para aplicar a fórmula do índice de Gini, utiliza-se os dados apresentados na tabela abaixo, obtidos de Hoffmann (2006).

Tabela 2.2: Valores de X_i , p_i , Φ_i e $\Phi_{i-1} + \Phi_i$ para a população hipotética de 8 elementos

| i | p_i | X_i | $\sum_{j=1}^n X_j$ | Φ_i | $\Phi_{i-1} + \Phi_i$ |
|-----|-------|-------|--------------------|----------|-----------------------|
| 1 | 0,125 | 1 | 1 | 0,02 | 0,02 |
| 2 | 0,250 | 1 | 2 | 0,04 | 0,06 |
| 3 | 0,375 | 1 | 3 | 0,06 | 0,10 |
| 4 | 0,500 | 2 | 5 | 0,10 | 0,16 |
| 5 | 0,625 | 4 | 9 | 0,18 | 0,28 |
| 6 | 0,750 | 8 | 17 | 0,34 | 0,52 |
| 7 | 0,875 | 13 | 30 | 0,60 | 0,94 |

| i | p_i | X_i | $\sum_{j=1}^n X_j$ | Φ_i | $\Phi_{i-1} + \Phi_i$ |
|-----|-------|-------|--------------------|----------|-----------------------|
| 8 | 1,000 | 20 | 50 | 1,00 | 1,60 |

Com essas informações é possível calcular o índice de Gini, através de (2.4)

$$G = 1 - \frac{1}{n} \sum_{i=1}^n (\Phi_{i-1} + \Phi_i);$$

através de (2.4)

$$G = \frac{\Delta}{2\mu},$$

onde

$$\Delta = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n |X_i - X_j|$$

e

$$\mu = \sum_{i=1}^n X_i;$$

e através de (2.4)

$$G = \frac{2}{n^2\mu} \sum_{i=1}^n iX_i - \frac{1}{n} - 1$$

Usando (2.4), é necessário totalizar a coluna $\Phi_{i-1} + \Phi_i$ na tabela 2.2. Ou seja,

```
options(OutDec = ",")
somaphis <- c(0.02, 0.06, 0.1, 0.16, 0.28, 0.52, 0.94,
             1.6)
somasomaphis <- sum(somaphis)
somasomaphis
```

```
## [1] 3,68
```

$$\sum_{i+1}^8 (\Phi_{i-1} + \Phi_i) = 3,68$$

Portanto

```
giniphis <- 1 - 1/length(somaphis) * somasomaphis
giniphis
```

```
## [1] 0,54
```

$$G = 1 - \frac{1}{8} \times 3,68 = 0,54$$

Para aplicar a fórmula (2.4) que é a fórmula do índice de Gini em termos de diferença absoluta média, Δ , é necessário calcular a diferença absoluta média com base nos dados de X_i da tabela 2.2.

```
xi <- c(1, 1, 1, 2, 4, 8, 13, 20)

XC <- matrix(xi, nrow = length(xi), ncol = length(xi),
             byrow = FALSE)
XC
```

```
##      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8]
## [1,]    1    1    1    1    1    1    1    1
## [2,]    1    1    1    1    1    1    1    1
## [3,]    1    1    1    1    1    1    1    1
## [4,]    2    2    2    2    2    2    2    2
## [5,]    4    4    4    4    4    4    4    4
## [6,]    8    8    8    8    8    8    8    8
## [7,]   13   13   13   13   13   13   13   13
## [8,]   20   20   20   20   20   20   20   20
```

```
XL <- matrix(xi, nrow = length(xi), ncol = length(xi),
             byrow = TRUE)
XL
```

```
##      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8]
## [1,]    1    1    1    2    4    8   13   20
## [2,]    1    1    1    2    4    8   13   20
## [3,]    1    1    1    2    4    8   13   20
## [4,]    1    1    1    2    4    8   13   20
## [5,]    1    1    1    2    4    8   13   20
## [6,]    1    1    1    2    4    8   13   20
## [7,]    1    1    1    2    4    8   13   20
## [8,]    1    1    1    2    4    8   13   20
```

```
DIFX <- XC - XL
DIFX
```

```
##      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8]
## [1,]    0    0    0   -1   -3   -7  -12  -19
## [2,]    0    0    0   -1   -3   -7  -12  -19
## [3,]    0    0    0   -1   -3   -7  -12  -19
## [4,]    1    1    1    0   -2   -6  -11  -18
## [5,]    3    3    3    2    0   -4   -9  -16
## [6,]    7    7    7    6    4    0   -5  -12
## [7,]   12   12   12   11    9    5    0   -7
```

```
## [8,] 19 19 19 18 16 12 7 0
ABSDIFX <- abs(DIFX)
ABSDIFX

##      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8]
## [1,] 0 0 0 1 3 7 12 19
## [2,] 0 0 0 1 3 7 12 19
## [3,] 0 0 0 1 3 7 12 19
## [4,] 1 1 1 0 2 6 11 18
## [5,] 3 3 3 2 0 4 9 16
## [6,] 7 7 7 6 4 0 5 12
## [7,] 12 12 12 11 9 5 0 7
## [8,] 19 19 19 18 16 12 7 0

iota <- matrix(1, nrow = length(xi), ncol = 1, byrow = TRUE)
iota

##      [,1]
## [1,] 1
## [2,] 1
## [3,] 1
## [4,] 1
## [5,] 1
## [6,] 1
## [7,] 1
## [8,] 1

somalunhaABSDIFX <- t(iota) %*% ABSDIFX
somalunhaABSDIFX

##      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8]
## [1,] 42 42 42 40 40 48 68 110

somalinhasomacolunaABSDIFX <- somacolunaABSDIFX %*%
iota
somalinhasomacolunaABSDIFX

##      [,1]
## [1,] 432

obs <- length(xi)

delta <- obs^(-2) * somalinhasomacolunaABSDIFX
delta

##      [,1]
## [1,] 6,75
```

Ou seja,

$$\Delta = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n |X_i - X_j| = \frac{1}{(8)^2} \times 432 = 6,75$$

Com a diferença absoluta média de X_i devidamente calculada, aplica-se a fórmula (2.4)

```
ximedio <- sum(xi)/length(xi)
ximedio
```

```
## [1] 6,25
```

```
ginidelta <- delta/(2 * ximedio)
ginidelta
```

```
##      [,1]
## [1,] 0,54
```

$$G = \frac{\Delta}{2\mu} = \frac{6,75}{2 \times 6,25} = 0,54$$

Para aplicar a fórmula (2.4) na obtenção do índice de Gini é necessário ponderar cada valor de X_i pela sua respectiva ordem i e soma todos os respectivos produtos

$$\sum_{i=1}^n iX_i.$$

usando os dados da tabela 2.2

```
is <- matrix(1:length(xi), nrow = length(xi), ncol = 1,
             byrow = TRUE)
is
```

```
##      [,1]
## [1,]    1
## [2,]    2
## [3,]    3
## [4,]    4
## [5,]    5
## [6,]    6
## [7,]    7
## [8,]    8
```

```
ixi <- t(is) * xi
ixi
```

```
##      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8]
## [1,]    1    2    3    8   20   48   91  160
```

```
somaixi <- sum(ixi)
somaixi

## [1] 333
ginifinal <- 2/(obs^2 * ximedio) * somaixi - obs^(-1) -
1
ginifinal

## [1] 0,54

obtem
```

$$G = \frac{2}{n^2\mu} \sum_{i=1}^n iX_i - \frac{1}{n} - 1 = \frac{2}{8^2 \times 6,25} \times 333 - \frac{1}{8} - 1 = 0,54$$

2.5 Discrepância Máxima

Discrepância Máxima é a maior distância entre a linha AB e a curva de Lorenz da figura 2.1. Portanto, a discrepância máxima é a diferença máxima entre a relação da porcentagem acumulada da população e a sua respectiva porcentagem acumulada da renda numa situação de exata igualdade dada pela reta AB e a relação da porcentagem acumulada da população e a sua respectiva porcentagem acumulada da renda numa situação de desigualdade entre as pessoas dessa população que na figura corresponde a poligonal da curva de Lorenz. De acordo com Hoffmann (2006)

$$D = p_h - \Phi_h$$

Portanto, o cálculo de D através de (2.5) depende de h , um número inteiro positivo. Encontrando-se $i = h$ encontra-se a discrepância máxima D .

Seja uma sequência de valores ordenados em ordem crescente de uma variável discreta X_i

$$X_1 \leq X_2 \leq \dots \leq X_n$$

sendo válida pelos menos uma desigualdade.

Para o cálculo da discrepância máxima, é importante entender que a mesma ocorre quando a inclinação do segmento da poligonal da curva de Lorenz passa de um valor menor que um para um valor maior que um. De acordo com Hoffmann (2006), a inclinação do segmento da poligonal é dada por

$$d_i = \frac{X_i}{\mu}$$

Essa mudança é identificada quando o valor de X_i ordenada em ordem crescente passa de um valor menor que a média μ para um valor maior que a média μ , ou seja,

$$X_i < \mu \text{ para } 1 \leq i \leq h$$

e

$$X_i \geq \mu \text{ para } h < i \leq n$$

Nestas condições percorre-se a sequência de valores em ordem crescente, o valor de $p_i - \Phi_i$ aumenta até a inclusão do h -ésimo elemento, que corresponde a (2.5).

Considerando (2.5), (2.4) e (2.4) chega-se em

$$D = \frac{h}{n} - \frac{1}{n\mu} \sum_{i=1}^h X_i$$

que depois de algumas manobras algébricas torna-se

$$D = \frac{1}{n\mu} \sum_{i=1}^h (\mu - X_i)$$

Na parte de estatística descritiva foi apresentado a medida de dispersão desvio absoluto médio

$$\delta = \frac{1}{n} \sum_{i=1}^n |X_i - \mu|$$

e considerando que

$$\delta = \frac{1}{n} \left[- \sum_{i=1}^h (X_i - \mu) + \sum_{i=h+1}^n (X_i - \mu) \right]$$

e que

$$\sum_{i=h+1}^n (X_i - \mu) = - \sum_{i=1}^h (X_i - \mu).$$

Portanto

$$\delta = \frac{2}{n} \sum_{i=h+1}^n (X_i - \mu) = \frac{2}{n} \sum_{i=1}^h (\mu - X_i).$$

Comparando (2.5) e (2.5) obtém-se

$$D = \frac{\delta}{2\mu}.$$

Se o Desvio Absoluto Médio, δ , é uma medida de dispersão da distribuição, a fórmula (2.5) mostra que discrepância máxima, da mesma forma que o índice de Gini, é uma medida de dispersão relativa. Retomando o exemplo numérico da tabela 2.2, é possível obter o valor da sua discrepância máxima através de (2.5).

$$D = p_h - \Phi_h = 0,625 - 0,180 = 0,445.$$

O mesmo resultado poderia ser obtido através da fórmula (2.5). Para isso é necessário calcular o desvio absoluto médio, δ , para os valores de X_i do exemplo numérico. Ou seja,

```
somadesvioabsolutomedioxi <- sum(abs(xi - ximedio))
somadesvioabsolutomedioxi
```

```
## [1] 44,5
```

```
desvioabsolutomedioxi <- (length(xi))^(-1) * sum(abs(xi -
  ximedio))
desvioabsolutomedioxi
```

```
## [1] 5,5625
```

$$\delta = \frac{1}{n} \sum_{i=1}^n |X_i - \mu| = \frac{1}{8} \times 44,5 = 5,5625.$$

Portanto,

```
discrepanciamaximaxi <- desvioabsolutomedioxi/(2 *
  ximedio)
discrepanciamaximaxi
```

```
## [1] 0,445
```

$$D = \frac{\delta}{2\mu} = \frac{5,5625}{2 \times 6,25} = 0,445.$$

2.6 Redundância e Índice de Theil

2.6.1 Teoria da Informação

Para entender melhor as medidas de desigualdades de Theil, é necessário introduzir alguns conceitos da teoria da informação.

Seja x , a probabilidade de ocorrer o evento E .

- Para $x = 1$, a mensagem **evento E ocorreu** não tem nenhum conteúdo informativo.

- Para $x \rightarrow 0$, ou seja, para valores muito pequenos de x , a mensagem **evento E ocorreu** tem alto valor informativo.

A segunda situação seria, por exemplo, o caso de uma notícia que nos causas surpresa ou de um *furo* de imprensa. Quando x tende a zero, o conteúdo informativo da mensagem **evento E ocorreu** tende a infinito.

Matematicamente, o conteúdo informativo da mensagem que afirma que determinado evento ocorreu é dado por

$$h(x) = \log \frac{1}{x} = \log x^{-1} = -\log x$$

De acordo com Hoffmann (2006), a escolha da função logarítmica é devido a propriedade de atividade do conteúdo informativo no caso de eventos independentes. Portanto, se E_1 e E_2 são dois eventos independentes com probabilidades x_1 e x_2 , respectivamente, a probabilidade de que ambos ocorram é $x_1 x_2$. O conteúdo informativo da mensagem de que ambos os eventos ocorreram é

$$h(x_1 x_2) = \log \frac{1}{x_1 x_2} = \log \frac{1}{x_1} + \log \frac{1}{x_2} = h(x_1) + h(x_2)$$

Em teoria da informação, normalmente se utiliza logaritmos na base 2 ou logaritmos naturais. Desta forma:

- Logaritmos na base 2: o conteúdo informativo é medido em **bits**.
- Logaritmos naturais: o conteúdo informativo é medido em **nits**.
- 1 bit = 0,693 nit.
- 1 nit = 1,443 bit.

Generalizando o conceito de informação, é apresentando, na sequência, como se mede o conteúdo informativo de uma **mensagem sujeita a erro**, ou **mensagem incerta**.

Para isso, admita-se que a a probabilidade de chover em um determinado dia, em certo local, estabelecida com base em séries históricas, seja $x_1 = 0,5$. Nesse caso o conteúdo da informação **chove** é de

$$h(x_1) = \log \frac{1}{0,5} = \log 2^1 = 1 \text{ bit}$$

Suponha agora que uma previsão de tempo estabeleceu que iria chover. Suponha, também, que, com base nos resultados anteriores de tais previsões, probabilidade de que realmente shova passa a ser $y_1 = 0,68$. De acordo com as novas suposições, o conteúdo da informação **chove** é

$$h(y_1) = \log \frac{1}{0,68} + 0,5564 \text{ bit}$$

O conteúdo informativo da previsão é

$$h(x_1) - h(y_1) = \log \frac{1}{x_1} - \log \frac{1}{y_1} = 1 - 0,5564 = 0,4436 \text{ bit}$$

ou

$$h(x_1) - h(y_1) = \log \frac{1}{x_1} - \log \frac{1}{y_1} = \log \frac{1}{x_1} + \log \left(\frac{1}{y_1} \right)^{-1} = \log \frac{y_1}{x_1} = \log \left(\frac{0,50}{0,68} \right) = 0,4436 \text{ bit}.$$

Ou seja, o conteúdo informativo **chove**, com base na probabilidade x_i nos dados históricos e na probabilidade y_i do histórico de previsões, é de 0,4436 bit.

Generalizando, o conteúdo informativo de uma mensagem sujeita a erro ou mensagem incerta, como é o caso da previsão, é dado por

$$\log \frac{y}{x}$$

onde

- x é a probabilidade *ex-ante* ou a probabilidade de que o evento ocorra antes de recebida a mensagem;
- y é a probabilidade *ex-post* ou a probabilidade de que o evento ocorra uma vez recebida a mensagem.

Na sequência é apresentado o conceito de *entropia*.

Entropia de uma distribuição $H(x)$

Seja o universo de n possíveis eventos E_i , para $i = 1, \dots, n$, mutuamente exclusivos aos quais associa-se as probabilidades x_i . Sabe-se que

$$\sum_i^n x_i = 1.$$

A informação esperada de uma mensagem certa, ou seja, a esperança matemática do conteúdo informativo da mensagem **ocorreu** E_i , também denominada entropia da distribuição, é

$$H(x) = E[h(x_i)] = \sum_{i=1}^n x_i h(x_i) = \sum_{i=1}^n x_i \log \frac{1}{x_i} = - \sum_{i=1}^n x_i \log x_i$$

Para o caso particular de $x_i = 0$, adota-se a definição

$$x \log x = 0, \text{ se } x = 0$$

uma vez que

$$\lim_{x \rightarrow 0} (x \log x) = 0$$

Para $0 < x_i \leq 1$ se tem

$$\frac{1}{x_i} \geq 1$$

e

$$\log \frac{1}{x_i} \geq 0.$$

Conclui-se que

$$H(x) = \sum_{i=1}^n x_i \log \frac{1}{x_i} = - \sum_{i=1}^n x_i \log x_i \geq 0$$

Valor mínimo de $H(x)$

O valor mínimo de $H(x)$ ocorre quando uma das probabilidades é 1 e as demais, conseqüentemente, são nulas. Nesse caso $H(x) = 0$. Ou seja, na somatória há um único $x_i = 1$ e o restante $x_i = 0$. Portanto,

- quando $x = 0$

$$x \log x = 0$$

de acordo com (2.6.1);

- quando $x = 1$ se tem $\log 1 = 0$ e também

$$x \log x = 0.$$

Assim o valor mínimo do valor esperado do conteúdo informativo $H(x)$ é

$$H(x) = - \sum_{i=1}^n x_i \log x_i = 0$$

Valor Máximo de $H(x)$

Para encontrar o valor máximo de $H(x)$ sujeito a condição de que $\sum x_i = 1$, utiliza-se o método do multiplicador de Lagrange, escrevendo a seguinte função

$$\max H(x) = - \sum_{i=1}^n x_i \log x_i$$

sujeito a

$$\sum_{i=1}^n x_i = 1$$

então

$$\mathcal{L} = - \sum_{i=1}^n x_i \log x_i - \lambda \left(\sum_{i=1}^n x_i - 1 \right)$$

Igualando a zero as derivadas parciais de (2.6.1) em relação a x_i e admitindo-se que se usa os logaritmos naturais, se tem:

$$\log x_i = -(1 + \lambda), \quad \text{para } i = 1, \dots, n$$

sendo que

$$x_i = e^{-(1+\lambda)} = \frac{1}{e^{(1+\lambda)}}.$$

Note que λ é constante neste caso.

O valor máximo de $H(x)$ acontece quando todos os valores de x_i , ou seja, todas as probabilidades são iguais entre si e, portanto, igual a $\frac{1}{n}$. Nesse caso,

$$H(x) = \sum_{i=1}^n x_i \log \frac{1}{x_i} = \sum_{i=1}^n \frac{1}{n} \log n = n \frac{1}{n} \log n = \log n$$

Resumindo, o valor esperado da informação ou a entropia da distribuição $H(x)$ varia entre 0 e $\log n$. Ou seja,

$$0 \leq H(x) \leq \log n.$$

A entropia da distribuição é máxima, ou seja, há um máximo de incerteza a respeito do que pode ocorrer, quando todos os possíveis eventos são igualmente prováveis, ou seja, quando há um máximo de *desordem* no sistema.

Informação de uma mensagem incerta

Finalmente é apresentado o conceito de informação de uma mensagem incerta. Dado o universo de n possíveis eventos E_i , mutuamente exclusivos, com probabilidades x_i , para $i = 1, \dots, n$, considera-se uma mensagem incerta que poderia ser uma previsão ou uma mensagem duvidosa, que transforma as probabilidades *a priori* x_i em probabilidade *a posteriori* y_i , onde y_i é a probabilidade de ocorrência do evento E_i depois de recebido a mensagem. Lembrando (2.6.1), verifica-se que a esperança matemática do conteúdo informativo da mensagem é

$$I(y : x) = \sum_{i=1}^n y_i \log \frac{y_i}{x_i}$$

A definição (2.6.1), do conteúdo informativo de uma mensagem certa, é somente um caso especial de (2.6.1), em que uma probabilidade *a posteriori* é igual a um e todas as outras são iguais a zero, ou seja, $y_j = 1$ e $y_i = 0$ para todo $i \neq j$.

2.6.2 Índice T de Theil

Seja uma população com n pessoas em que cada uma recebe uma fração não negativa da renda total,

$$y_i \geq 0, \text{ com } i = 1, \dots, n.$$

Se a renda média é μ e X_i é a renda i -ésima pessoa,

$$y_i = \frac{X_i}{n\mu}.$$

Obviamente,

$$\sum_{i=1}^n y_i = 1.$$

Os valores de y_i tem as mesmas propriedades que as probabilidades x_i associadas a um universo de eventos E_i da teoria da informação. Assim sendo, pode-se, com base em (2.6.1), definir a **entropia** da distribuição de renda considerada como sendo

$$H(y) = \sum_{i=1}^n y_i \log \frac{1}{y_i}.$$

De acordo com (2.6.1), se tem

$$0 \leq H(y) \leq \log n.$$

Assim é possível definir as duas situações extremas:

- o caso de perfeita igualdade na distribuição da renda,

$$y_i = \frac{1}{n} \text{ para } i = 1, \dots, n,$$

se tem $H(y) = \log n$;

- o caso de perfeita desigualdade na distribuição de renda,

$$y_i = 1, \text{ para } i = 1, \dots, n,$$

se tem $H(y) = 0$.

A **entropia** é, portanto, uma medida do grau de igualdade da distribuição. Mas como o objeto de análise é desigualdade, é muito mais interessante uma medida de desigualdade. Para isto basta subtrair a entropia do seu valor próprio máximo, $\log n$. Essa medida, denominada **Índice T de Theil** da distribuição é dada por

$$T = \log n - H(y) = \sum_{i=1}^n y_i \log ny_i$$

Para o cálculo do índice T de Theil, pode-se usar os logaritmos naturais ou os logaritmos na base 2, obtendo-se o valor de T em *nits* ou *bits*, respectivamente. Na prática, utiliza-se mais o logaritmo natural.

Note que

$$0 \leq T \leq \log n$$

sendo que:

- $T = 0$ corresponde ao caso de uma distribuição da renda com perfeita igualdade e;
- $T = \log n$ corresponde ao caso de uma distribuição da renda com perfeita desigualdade.

De (2.6.2) se tem

$$T = \sum_{i=1}^n y_i \log \frac{y_i}{\frac{1}{n}}.$$

Comparando essa equação com (2.6.1), verifica-se que o **índice T de Theil** corresponde à esperança do valor informativo de uma mensagem incerta, em que as probabilidades *a posteriori* são as frações da renda total y_i apropriadas pelas pessoas, e as probabilidades *a priori* são iguais a $1/n$, ou seja, iguais à fração da população correspondente a cada pessoa.

2.6.3 Índice de L de Theil

A outra medida de desigualdade proposta por Theil, o **índice L de Theil**, corresponde à esperança do valor informativo de uma mensagem incerta, em que as probabilidades *a posteriori** são as frações da população $1/n$ e as probabilidades *a priori* são as frações da renda y_i . Ou seja,

$$L = \sum_{i=1}^n \frac{1}{n} \log \frac{\frac{1}{n}}{y_i} = \frac{1}{n} \sum_{i=1}^n \log \frac{1}{ny_i}.$$

Verifica-se que o **índice L de Theil** é igual a zero no caso de perfeita igualdade

$$y_i = \frac{1}{n}$$

para todo i .

Basta uma das renda aproximar-se de zero para que o valor de L tenda a infinito, fazendo que o índice L seja inútil quando se trata de comparar distribuições de renda que incluem valores nulos.

Uma vantagem importante das medidas de desigualdades de Theil na análise da distribuição de renda ou da riqueza é que, quando os dados podem ser agrupados com base em um critério qualquer, por exemplo regiões, os valores de T e L podem ser decompostos em uma medida de desigualdade **entre** grupos, por exemplo inter-regional, e uma média ponderada das medidas de desigualdades **dentro** de grupos, por exemplo dentro das regiões.

2.7 Variância dos Logaritmos

A variância dos logaritmos das rendas é frequentemente utilizada como medida da desigualdade da distribuição da renda em uma população. Para uma população com n pessoas, em que a renda da i -ésima pessoa é indicada por X_i para $i = 1, \dots, n$, a variância dos logaritmos das rendas é dada por

$$V(Z) = \frac{1}{n} \sum_{i=1}^n (Z_i - \bar{Z})^2$$

onde

$$Z_i = \log X_i$$

e

$$\bar{Z} = \frac{1}{n} \sum_{i=1}^n Z_i.$$

Nota-se que $V(Z)$ só é definida para $X_i \geq 0$ para $i = 1, \dots, n$.

Indicando-se por X^* a média geométrica dos X_i , se tem

$$V(Z) = \frac{1}{n} \sum_{i=1}^n \left(\log \frac{X_i}{X^*} \right)^2$$

A variância dos logaritmos, da mesma forma que as médias T e L de Theil, é uma medida de desigualdade que, quando os dados podem ser agrupados segundo um critério qualquer, pode ser decomposta em um componente que corresponde à desigualdade entre os grupos e uma média ponderada das variâncias dos logaritmos dentro dos grupos.

Capítulo 3

Números-Índices

O conteúdo deste tópico foi elaborado com base no capítulo 16 de Hoffmann (2006) e no capítulo 11 de Sartoris (2013).

3.1 Introdução

Os números-índices, ou simplesmente índices, são proporções estatísticas, geralmente expressas em porcentagem, idealizadas para comparar as situações de um conjunto de variáveis em épocas ou localidades diversas.

Para quem vai fazer uso do números-índices na análise de um problema, é importante saber como são obtidos. Mesmo que não seja necessário calcular um novo índice, é interessante conhecer os métodos de cálculo, pois isso permite interpretar melhor os índices publicados e avaliar suas limitações.

3.2 Preços Relativos

O número-índice preços relativos é a relação entre o preço de um produto em determinado período (ano ou mês, geralmente) e o preço no período escolhido como base. Esse índice se destina a acompanhar a evolução do preço de determinado produto. O preço relativo pode ser denominado também de índice relativo de preço ou número-índice simples de preço.

Se P_0 é o preço da mercadoria no período-base e P_t é o preço em um período t , o preço relativo da mercadoria no ano t é dado por

$$I(P_t|P_0) = \frac{P_t}{P_0}. \quad (3.1)$$

Usualmente, o valor do índice é dado em porcentagem, calculando-se

$$I^*(P_t|P_0) = \frac{P_t}{P_0} \cdot 100. \quad (3.2)$$

Exemplo numérico de índice relativo

Seja os dados da tabela 3.1 abaixo

Tabela 3.1: dados hipotético de preços e quantidades para três produtos no período de 2001 a 2005

| ano | P_{1t} | Q_{1t} | P_{2t} | Q_{2t} | P_{3t} | Q_{3t} |
|------|----------|----------|----------|----------|----------|----------|
| 2001 | 12 | 3 | 5 | 7 | 20 | 3 |
| 2002 | 15 | 4 | 10 | 9 | 25 | 4 |
| 2003 | 18 | 5 | 20 | 8 | 35 | 5 |
| 2004 | 24 | 5 | 30 | 7 | 45 | 6 |
| 2005 | 30 | 6 | 60 | 6 | 50 | 5 |

Fonte: Hoffmann (2006).

Tomando com base o ano 2002, o preço relativo do produto 1 em 2004, de acordo com (3.2) é

$$I^*(P_t|P_0) = \frac{24}{15} \cdot 100 = 160. \quad (3.3)$$

Com base na tabela 3.1, os demais preços relativos podem ser calculados, tomando como base o ano de 2002, que são apresentados na tabela 3.2.

Tabela 3.2: Preços relativos dos três produtos no período de 2001 a 2005, tomando como base o ano de 2002

| ano | Produto 1 | Produto 2 | Produto 3 |
|------|-----------|-----------|-----------|
| 2001 | 80 | 50 | 80 |
| 2002 | 100 | 100 | 100 |
| 2003 | 120 | 200 | 140 |
| 2004 | 160 | 300 | 180 |
| 2005 | 200 | 600 | 200 |

Exemplo numérico de preço relativo no R

Carrega-se os preços dos três produtos e calcula os respectivos preços relativos tendo como base o ano de 2002.

```

prd1 <- c(12, 15, 18, 24, 30)
prd2 <- c(5, 10, 20, 30, 60)
prd3 <- c(20, 25, 35, 45, 50)

iprprd1 <- prd1/prd1[2] * 100
iprprd2 <- prd2/prd2[2] * 100
iprprd3 <- prd3/prd3[2] * 100

ipr161 <- cbind(iprprd1, iprprd2, iprprd3)
rownames(ipr161) <- c(2001, 2002, 2003, 2004, 2005)

ipr161

```

```

##      iprprd1 iprprd2 iprprd3
## 2001      80      50      80
## 2002     100     100     100
## 2003     120     200     140
## 2004     160     300     180
## 2005     200     600     200

```

Note que na linha da ano 2002, os três preços relativos em porcentagem são iguais a 100.

O preço relativo mostra como está evoluindo o preço de cada um dos produtos. Mas quando analisa-se um conjunto de mercadorias, interessa-se em obter um único índice que nos mostre como está evoluindo o nível geral dos preços dessas mercadorias.

3.3 Índice Simples de Preços Agregados

O índice Simples de Preços Agregados é a relação entre o somatório dos preços das mercadorias no período t e o somatório dos preços das mercadorias no período escolhido como base. O índice simples de preços agregados é também chamado de índice agregativo de preços. Defini-se que o índice i , variando de 1 a n , indique as n diferentes mercadorias do conjunto considerado. Note que $I_A(\mathbf{p}_t|\mathbf{p}_0)$ é o valor do índice agregativo para o vetor ou conjunto de preços $\mathbf{p}_t = P_{it}, i = 1, \dots, n$ quando comparado ao vetor de preços do período-base, $\mathbf{p}_0 = P_{i0}, i = 1, \dots, n$. Ou seja,

$$I_A(\mathbf{p}_t|\mathbf{p}_0) = \frac{\sum_{i=1}^n P_{it}}{\sum_{i=1}^n P_{i0}}, \quad (3.4)$$

ou em porcentagem

$$I_A^*(\mathbf{p}_t|\mathbf{p}_0) = \frac{\sum_{i=1}^n P_{it}}{\sum_{i=1}^n P_{i0}} \cdot 100. \quad (3.5)$$

Exemplo numérico do Índice simples de preços agregados

om base nos dados da tabela 3.1, calcula-se o índice simples de preços agregados para os três produtos em 2004, tomando 2002 como ano-base. De acordo com (3.5),

$$I_A^*(\mathbf{p}_4|\mathbf{p}_2) = \frac{24 + 30 + 45}{15 + 10 + 25} \cdot 100 = 198. \quad (3.6)$$

Analogamente, obtém-se os demais índices que são apresentados na tabela 3.3.

Tabela 3.3: Índice simples de preços agregados dos três produtos no período 2001-2005 tomando como base o ano de 2002

| ano | $I_A^*(P_t P_0)$ |
|------|------------------|
| 2001 | 74 |
| 2002 | 100 |
| 2003 | 146 |
| 2004 | 198 |
| 2005 | 280 |

A seguir é calculado o Índice simples de preços agregados do exemplo numérico no R.

Exemplo numérico do índice simples de preços agregados no R

```
p1 <- c(12, 15, 18, 24, 30)
p2 <- c(5, 10, 20, 30, 60)
p3 <- c(20, 25, 35, 45, 50)

p161 <- cbind(p1, p2, p3)
rownames(p161) <- c(2001, 2002, 2003, 2004, 2005)

IAp1p2 <- sum(p161[1, ])/sum(p161[2, ]) * 100
IAp2p2 <- sum(p161[2, ])/sum(p161[2, ]) * 100
IAp3p2 <- sum(p161[3, ])/sum(p161[2, ]) * 100
IAp4p2 <- sum(p161[4, ])/sum(p161[2, ]) * 100
IAp5p2 <- sum(p161[5, ])/sum(p161[2, ]) * 100

ispa161 <- rbind(IAp1p2, IAp2p2, IAp3p2, IAp4p2, IAp5p2)
colnames(ispa161) <- c("ISPA")

ispa161

##          ISPA
## IAp1p2    74
## IAp2p2   100
## IAp3p2   146
```

IAp4p2 198

IAp5p2 280

3.4 Média Aritmética dos Preços Relativos

A Média Aritmética dos Preços Relativos é um índice geral de preços que se obtém calculando a média aritmética dos preços relativos, isto é

$$I_M(\mathbf{p}_t|\mathbf{p}_0) = \frac{1}{n} \sum_{i=1}^n \frac{P_{it}}{P_{i0}} \quad (3.7)$$

ou, em porcentagem,

$$I_M^*(\mathbf{p}_t|\mathbf{p}_0) = \frac{1}{n} \sum_{i=1}^n \frac{P_{it}}{P_{i0}} \cdot 100. \quad (3.8)$$

Na tabela 3.3 abaixo são apresentados os valores desse índice para os dados da tabela 3.1

Tabela 3.4: Índice média aritmética dos preços relativos dos três produtos no período 2001-2005 tomando como base o ano de 2002

| ano | $I_M^*(P_t P_0)$ |
|------|------------------|
| 2001 | 70,0 |
| 2002 | 100,0 |
| 2003 | 153,3 |
| 2004 | 213,3 |
| 2005 | 333,3 |

Note que os valores da tabela 3.4 é a média aritmética dos preços relativos da tabela 3.1 para cada ano.

Para o ano de 2003, por exemplo, foi calculada da seguinte forma:

$$I_M^*(P_{2003}|P_{2002}) = \frac{1}{3} \times (120 + 200 + 140) = 153,3$$

Exemplo numérico da média aritmética dos preços relativos no R

Como temos calculado o preço relativo para cada um dos três produtos com ano base 2002, é só calcular a média aritmética para cada ano.

ipr161

```
##      iprprd1 iprprd2 iprprd3
## 2001      80      50      80
## 2002     100     100     100
## 2003     120     200     140
## 2004     160     300     180
## 2005     200     600     200

maprp1p2 <- sum(ipr161[1, ])/length(ipr161[1, ])
maprp2p2 <- sum(ipr161[2, ])/length(ipr161[2, ])
maprp3p2 <- sum(ipr161[3, ])/length(ipr161[3, ])
maprp4p2 <- sum(ipr161[4, ])/length(ipr161[4, ])
maprp5p2 <- sum(ipr161[5, ])/length(ipr161[5, ])

mapr161 <- round(rbind(maprp1p2, maprp2p2, maprp3p2,
  maprp4p2, maprp5p2), 1)
colnames(mapr161) <- c("MAPR")

mapr161

##      MAPR
## maprp1p2  70,0
## maprp2p2 100,0
## maprp3p2 153,3
## maprp4p2 213,3
## maprp5p2 333,3
```

Considerando que a expressão (3.4) pode ser escrita como em (3.9), pode-se entender por que o índice cresceu mais na média aritmética dos preços relativos na tabela 3.4 que o índice simples de preços agregados na tabela 3.3.

$$I_A(\mathbf{p}_t|\mathbf{p}_0) = \frac{\sum_{i=1}^n \left(\frac{P_{it}}{P_{i0}} \right) P_{i0}}{\sum_{i=1}^n P_{i0}} \quad (3.9)$$

Tanto o índice simples de preços agregados como a média aritmética dos preços relativos são índices gerais de preços. No seus cálculos não se leva em consideração a importância econômica de cada mercadoria que é dada pelo valor monetário da quantidade vendida.

Portanto, os índices simples de preços agregados bem como a média aritmética do preços relativos só devem ser usados quando não se tem disponível as informações de quantidades. Caso contrário, deve ser utilizados os índices ponderados de preço.

3.5 Índice de Preços de Laspeyres

O índice de preços de Laspeyres para um conjunto de mercadorias, em um período t , é a média ponderada dos preços relativos dessas mercadorias, utilizando, como

fatores de ponderação, os valores monetários das quantidades de cada mercadoria vendidas no período base.

$$I_L(\mathbf{p}_t|\mathbf{p}_0) = \frac{\sum_{i=1}^n \left(\frac{P_{it}}{P_{i0}} \right) P_{i0} Q_{i0}}{\sum_{i=1}^n P_{i0}} \quad (3.10)$$

onde

- Q_{i0} é a quantidade da i -ésima mercadoria vendida no período-base e;
- $P_{i0}Q_{i0}$ é o valor monetário da i -ésima mercadoria ao preço do período-base.

Simplificando

$$I_L(\mathbf{p}_t|\mathbf{p}_0) = \frac{\sum_{i=1}^n P_{it} Q_{i0}}{\sum_{i=1}^n P_{i0} Q_{i0}} \quad (3.11)$$

ou, em porcentagem,

$$I_L^*(\mathbf{p}_t|\mathbf{p}_0) = \frac{\sum_{i=1}^n P_{it} Q_{i0}}{\sum_{i=1}^n P_{i0} Q_{i0}} \cdot 100. \quad (3.12)$$

Exemplo numérico de Índice de preços de Laspeyres

Com base na tabela 3.1 e na fórmula (3.12), o índice de preços de Laspeyres para os três produtos, em 2004, tomando 2002 como ano-base é

$$I_L^*(\mathbf{p}_4|\mathbf{p}_2) = \frac{24 \cdot 4 + 30 \cdot 9 + 45 \cdot 4}{15 \cdot 4 + 10 \cdot 9 + 25 \cdot 4} \cdot 100 = \frac{546}{250} \cdot 100 = 218,4 \quad (3.13)$$

Os valores do índice de preços de Laspeyres para 2001, 2003 e 2005 podem ser obtidos da mesma maneira e são apresentados na tabela 3.5.

Tabela 3.5: Índice de Laspeyres dos três produtos no período 2001-2005 tomando como base o ano de 2002

| ano | $I_L^*(P_t P_0)$ |
|------|------------------|
| 2001 | 69,2 |
| 2002 | 100,0 |
| 2003 | 156,8 |
| 2004 | 218,4 |
| 2005 | 344,0 |

Exemplo numérico de Índice de preços de Laspeyres no R

```

q1 <- c(3, 4, 5, 5, 6)
q2 <- c(7, 9, 8, 7, 6)
q3 <- c(3, 4, 5, 6, 5)
q161 <- cbind(q1, q2, q3)
rownames(q161) <- c(2001, 2002, 2003, 2004, 2005)

q161

```

```

##      q1 q2 q3
## 2001  3  7  3
## 2002  4  9  4
## 2003  5  8  5
## 2004  5  7  6
## 2005  6  6  5

```

```

iplp1p2 <- sum(p161[1, ] * q161[2, ])/sum(p161[2, ] *
  q161[2, ]) * 100
iplp2p2 <- sum(p161[2, ] * q161[2, ])/sum(p161[2, ] *
  q161[2, ]) * 100
iplp3p2 <- sum(p161[3, ] * q161[2, ])/sum(p161[2, ] *
  q161[2, ]) * 100
iplp4p2 <- sum(p161[4, ] * q161[2, ])/sum(p161[2, ] *
  q161[2, ]) * 100
iplp5p2 <- sum(p161[5, ] * q161[2, ])/sum(p161[2, ] *
  q161[2, ]) * 100

ipl161 <- rbind(iplp1p2, iplp2p2, iplp3p2, iplp4p2,
  iplp5p2)
colnames(ipl161) <- c("IPL")

ipl161

```

```

##      IPL
## iplp1p2 69,2
## iplp2p2 100,0
## iplp3p2 156,8
## iplp4p2 218,4
## iplp5p2 344,0

```

Uma interpretação econômica do Índice de preços de Laspeyres

O índice de preços de Laspeyres é uma relação entre o custo de aquisição da **cesta de mercadorias** q_0 no período t e o custo de aquisição dessa mesma cesta de mercadorias no período-base.

3.6 Índice de Preços de Paasche

O índice de preços de Paasche para o período t pode ser interpretado como uma média ponderada dos preços relativos, utilizando com fatores de ponderação os valores monetários das quantidades vendidas no período t , considerando os preços do período-base.

$$I_P(\mathbf{p}_t|\mathbf{p}_0) = \frac{\sum_{i=1}^n \left(\frac{P_{it}}{P_{i0}} \right) P_{i0} Q_{it}}{\sum_{i=1}^n P_{i0} Q_{it}} \quad (3.14)$$

onde

- Q_{it} é a quantidade da i -ésima mercadoria vendida no período t e;
- $P_{i0}Q_{it}$ é o valor monetário da i -ésima mercadoria no período t considerando o preço do período-base.

Simplificando

$$I_P(\mathbf{p}_t|\mathbf{p}_0) = \frac{\sum_{i=1}^n P_{it} Q_{it}}{\sum_{i=1}^n P_{i0} Q_{it}} \quad (3.15)$$

ou, em porcentagem,

$$I_P^*(\mathbf{p}_t|\mathbf{p}_0) = \frac{\sum_{i=1}^n P_{it} Q_{it}}{\sum_{i=1}^n P_{i0} Q_{it}} \cdot 100. \quad (3.16)$$

De acordo com os dados da tabela 3.1 e a fórmula (3.16), o índice de preços de Paasche para os três produtos em 2004, tomando 2002 como ano-base é

$$I_P^*(\mathbf{p}_4|\mathbf{p}_2) = \frac{24 \cdot 5 + 30 \cdot 7 + 45 \cdot 6}{15 \cdot 5 + 10 \cdot 7 + 25 \cdot 6} \cdot 100 = \frac{600}{295} \cdot 100 = 203,4 \quad (3.17)$$

Os valores dos índices de preços de Paasche para os demais anos podem ser calculados na mesma forma e são apresentados na tabela 3.6.

Tabela 3.6: Índice de Paasche dos três produtos no período 2001-2005 tomando como base o ano de 2002

| ano | $I_P^*(P_t P_0)$ |
|------|------------------|
| 2001 | 68,9 |
| 2002 | 100,0 |
| 2003 | 151,8 |
| 2004 | 203,4 |
| 2005 | 287,3 |

Exemplo numérico de índice de preços de Paasche no R

```

ipp1p2 <- sum(p161[1, ] * q161[1, ])/sum(p161[2, ] *
  q161[1, ]) * 100
ipp2p2 <- sum(p161[2, ] * q161[2, ])/sum(p161[2, ] *
  q161[2, ]) * 100
ipp3p2 <- sum(p161[3, ] * q161[3, ])/sum(p161[2, ] *
  q161[3, ]) * 100
ipp4p2 <- sum(p161[4, ] * q161[4, ])/sum(p161[2, ] *
  q161[4, ]) * 100
ipp5p2 <- sum(p161[5, ] * q161[5, ])/sum(p161[2, ] *
  q161[5, ]) * 100

ipp161 <- round(rbind(ipp1p2, ipp2p2, ipp3p2, ipp4p2,
  ipp5p2), 1)
colnames(ipp161) <- c("IPP")

ipp161

##          IPP
## ipp1p2  68,9
## ipp2p2 100,0
## ipp3p2 151,8
## ipp4p2 203,4
## ipp5p2 287,3

```

Interpretação econômica do índice de preços de Paasche

O índice de preços de Paasche é uma relação entre o custo de aquisição da **cesta de mercadorias** q_t no período t com o custo de aquisição dessa mesma cesta de mercadorias no período-base.

Comparando os índices de preços de Laspeyres e Paasche

O método de Paasche exige mais informações do que o método de Laspeyres no cálculo do índice ponderado de preços.

3.7 Índice de Preços de Fisher

Vimos que o método de Laspeyres e o método de Paasche entregam, em geral, resultados diferentes quando utilizados para avaliar a variação no nível dos preços de um conjunto de produtos.

Para tentar superar essa divergência de resultados, foram criados índices que conduzem a valores intermediários entre o índice de Laspeyres e o índice de Paasche.

O índice de preços de Fisher, é por definição, a média geométrica entre o índice de preços de Laspeyres e o índice de preços de Paasche.

Ou seja,

$$I_F(\mathbf{p}_t|\mathbf{p}_0) = \sqrt{I_L(\mathbf{p}_t|\mathbf{p}_0)I_P(\mathbf{p}_t|\mathbf{p}_0)} = \left(\frac{\sum_{i=1}^n P_{it}Q_{i0}}{\sum_{i=1}^n P_{i0}Q_{i0}} \frac{\sum_{i=1}^n P_{it}Q_{it}}{\sum_{i=1}^n P_{i0}Q_{it}} \right)^{1/2} \quad (3.18)$$

ou, em porcentagem,

$$I_F^*(\mathbf{p}_t|\mathbf{p}_0) = I_F(\mathbf{p}_t|\mathbf{p}_0) \cdot 100. \quad (3.19)$$

Exemplo numérico do Índice de preços de Fisher

Para calcular o índice de preços de Fischer de 2004, tomando como 2002 com ano-base, com base nos dados da tabela 3.1, basta tomar os valores dos índices de preços de Laspeyres e de Paasche nas tabelas 3.5 e 3.6 respectivamente.

Ou seja,

$$I_F^*(\mathbf{p}_t|\mathbf{p}_0) = \sqrt{218,4 \cdot 203,4} = 210,8 \quad (3.20)$$

3.8 Índice de Preços de Marshall-Edgeworth

O índice de preços de Marshall-Edgeworth pode ser interpretado como uma relação entre o custo de aquisição no período t e o custo de aquisição no período-base de uma cesta de mercadorias que contém, para cada produto, a média aritmética das quantidades vendidas no período -base e no período t .

$$I_E(\mathbf{p}_t|\mathbf{p}_0) = \frac{\sum_{i=1}^n P_{it}(Q_{i0} + Q_{it})}{\sum_{i=1}^n P_{i0}(Q_{i0} + Q_{it})} \quad (3.21)$$

ou, em porcentagem.

$$I_E^*(\mathbf{p}_t|\mathbf{p}_0) = I_E(\mathbf{p}_t|\mathbf{p}_0) \cdot 100. \quad (3.22)$$

Exemplo numérico do índice de preços de Marshall-Edgeworth

Para calcular o índice de preços de Marshall-Edgeworth de 2004, tomando 2002 como ano base, calcula-se utilizando a fórmula (3.22) usando os dados da tabela 3.1

$$I_E^*(\mathbf{p}_t|\mathbf{p}_0) = \frac{24(4+5) + 30(9+7) + 45(4+6)}{15(4+5) + 10(9+7) + 25(4+6)} \cdot 100 = 210,3 \quad (3.23)$$

3.9 Deflacionamento

Considere dois valores monetários pagos ou recebidos em diferentes datas:

- $x = \text{R\$}120,00$ recebidos em setembro e 1997 e;
- $y = \text{R\$}240,00$ recebidos em setembro de 2003.

Aparentemente, não há a necessidade de modificar a unidade medida de nenhum dos dois valores, pois ambos estão em reais.

Mas devido à inflação ou desvalorização da moeda, o real de setembro de 2003 é uma unidade de medida de valor de troca bastante diferente do real de setembro de 1997.

Por isso, antes de fazer qualquer comparação ou operação aritmética envolvendo os valores de x e y é necessário fazer uniformizar a unidade de medida.

Isso se faz através de um índice de preços que possa ser utilizado como uma medida de desvalorização da moeda. Essa é uma das principais aplicações dos números-índices em uma economia inflacionária.

O índice de preços utilizados como medida da inflação ou desvalorização da moeda é denominada **deflator**.

Os valores de x e y medidos em reais da data em que o pagamento é efetuado são denominados **valores nominais** ou **valores em moeda corrente**. Caso se trate de preços de um produto, são denominados **preços correntes**.

Então, antes de fazer qualquer comparação ou operação aritmética envolvendo os valores em moeda corrente, é necessário calcular os **valores reais** ou **valores deflacionados** utilizando um **deflator**.

Através da um regra de três simples,

$$\frac{V_r}{V_t} = \frac{I_k}{I_t} \quad (3.24)$$

temos

$$V_r = \frac{I_k}{I_t} V_t. \quad (3.25)$$

Se, por exemplo, a inflação entre o período k e o período t tivesse sido de 100%,

$$I_t = 2I_k \quad (3.26)$$

então

$$V_r = \frac{V_t}{2}. \quad (3.27)$$

Ou seja, o valor real, medido em moeda do ano k , seria igual à metade do valor em moeda corrente.

Geralmente, quando se calculam valores reais, a finalidade é apenas uniformizar a unidade de medida.

Então, por facilidade de cálculo, são obtidos os valores reais medidos em moeda do período-base do deflator, isto é, em lugar de I_k , considera-se $I_0 = 100$. Nesse caso, a expressão (3.25) fica

$$V_r = \frac{V_t}{I_t} 100. \quad (3.28)$$

Como exemplos de índices comumente usados em trabalhos de pesquisa em economia se tem:

- o Índice Geral de Preços - Disponibilidade Interna (IGP-DI) é usado como deflator para preços de produtos e serviços.
- o índice de custo de vida é o mais adequado para deflacionar salários.

Exemplo para o cálculo de valores reais

Considere $x = \text{R\$ } 120,00$ o valor corrente do salário mínimo vigente em setembro de 1997 e $y = \text{R\$ } 240,00$ o valor corrente do salário mínimo vigente em setembro de 2003.

Um deflator apropriado, nesse caso, é o **Índice Nacional de Preços ao Consumidor Restrito** (INPC) calculado pelo IBGE para medir o custo de vida das famílias cujos chefes são assalariados em sua ocupação principal e cujo rendimento monetário disponível situe-se entre 1 e 5 salários mínimos.

Com base em dezembro de 1993, o INPC é igual a R\$ 1.415,18 em setembro de 1997 e igual a R\$ 2.288,16 em setembro de 2003

Então, de acordo com (3.25), o valor deflacionado de y em reais de setembro de 1997 é

$$V_{1997} = \frac{I_{1997}}{I_{2003}} \times V_{2003}, \quad (3.29)$$

ou seja,

$$V_{1997} = \frac{1.415,18}{2.288,16} \times 240 = 148,44 \quad (3.30)$$

Alternativamente pode-se calcular o valor real de x em moeda de setembro de 2003,

$$V_{2003} = \frac{I_{2003}}{I_{1997}} \times V_{1997}, \quad (3.31)$$

ou seja,

$$V_{2003} = \frac{2.228,16}{1.415,18} \times 120 = 194,02. \quad (3.32)$$

Capítulo 4

Variável Aleatória Discreta

A associação de cada um dos valores de uma variável a uma probabilidade correspondente de ocorrer corresponde a uma **distribuição de probabilidade**.

Essa distribuição de probabilidade pode ser de uma variável aleatória discreta ou de uma variável aleatória contínua. No capítulo 03 do Sartoris (2013) apresenta-se a distribuições de probabilidade de variáveis aleatórias discretas.

4.1 Esperança matemática

Quando se calcula uma média do que **pode** acontecer com a variável, baseada em sua distribuição de probabilidade, está obtendo-se um valor médio esperado. Ou seja, a **esperança matemática**, ou, simplesmente, **esperança**.

A esperança de uma variável aleatória discreta X , denotada $E(X)$, é definida como

$$E(X) = X_1P(X_1) + X_2P(X_2) + \dots + X_nP(X_n) = \sum_{i=1}^n X_iP(X_i)$$

A probabilidade aqui tem o mesmo papel da frequência relativa.

Note que,

- quando se fala em frequência relativa, usualmente refere-se a uma quantidade obtida.
- quando se fala em probabilidade, nos remete a ideia de proporções em que a variável pode assumir determinado valor.

4.1.1 Função de Probabilidade

$P(X)$ é a função que associa o valor de X à sua probabilidade, que é chamada de **função de probabilidade**. É representada como $f(X)$. Desta forma pode-se escrever:

$$E(X) = \sum_{i=1}^n X_i f(X_i) \text{ para } i = 1, \dots, n.$$

4.1.2 Função distribuição acumulada

A função que fornece a probabilidade acumulada dado o valor de X é a **função distribuição acumulada** ou, simplesmente, função distribuição, representada por $F(X)$.

$$F(X_k) = \sum_{i=1}^k f(X_i) \text{ para } i = 1, \dots, n$$

Exemplo numérico sobre Função de Probabilidade (Sartoris, 2013, pg.55)

Este exemplo numérico bem básico é sobre probabilidade que está na página 55 de Sartoris (2013). Num sorteio de números inteiros de 1 a 5, a probabilidade de um número ser sorteado é proporcional a esse número. Qual é a probabilidade de cada número ser sorteado?

Resposta: Se considerarmos que a $P(1)$ é igual a $1A$, onde A é uma constante desconhecida, temos:

$$P(1) = A$$

$$P(2) = 2A$$

$$P(3) = 3A$$

$$P(4) = 4A$$

$$P(5) = 5A.$$

Sendo os eventos mutuamente exclusivos, a soma de todas as probabilidades tem de ser igual a 1. Portanto,

$$P(1) + P(2) + P(3) + P(4) + P(5) = 1$$

$$A + 2A + 3A + 4A + 5A = 1$$

$$15A = 1$$

$$A = \frac{1}{15}.$$

4.1.3 Esperanças, Variâncias e Covariâncias

Note que

$$E[aX + b] = aE(X) + b \quad (4.1)$$

$$E[X + Y] = E(X) + E(Y) \quad (4.2)$$

$$Var(X) = E[X - E(X)]^2 = E(X^2) - [E(X)]^2 \quad (4.3)$$

$$Cov(X, Y) = E[X - E(X)][Y - E(Y)] = E(XY) - E(X)E(Y) \quad (4.4)$$

Exemplo Numérico sobre Esperança e Variância (Sartoris, 2013, p.56)

Este exemplo numérico é o 3.1.2 do Sartoris (2013) que está na página 56. Uma ação comprada por R\$10,00 pode assumir, após 30 dias, os seguintes valores: R\$5,00 com 20% de probabilidade; R\$10,00 com 30% de probabilidade; R\$16,00 com 25% de probabilidade ou R\$20,00 com 25% de probabilidade. Determine o valor esperado da ação e sua variância.

Resposta: O valor esperado da ação é

$$E(X) = 5 \times 0,2 + 10 \times 0,3 + 16 \times 0,25 + 20 \times 0,25 \quad (4.5)$$

$$E(X) = 1 + 3 + 4 + 5 = 13 \quad (4.6)$$

Como o preço da ação foi de R\$ 10,00, o lucro médio esperado dessa ação é de R\$3,00.

Para calcular a variância pela forma alternativa, é necessário calcular somente $E(X^2)$ uma vez que $E(X)$ já foi calculado.

$$E(X^2) = (5)^2 \times 0,2 + (10)^2 \times 0,3 + (16)^2 \times 0,25 + (20)^2 \times 0,25 \quad (4.7)$$

$$E(X^2) = 25 \times 0,2 + 100 \times 0,3 + 256 \times 0,25 + 400 \times 0,25 \quad (4.8)$$

$$E(X^2) = 5 + 30 + 64 + 100 \quad (4.9)$$

$$E(X^2) = 199. \quad (4.10)$$

Assim,

$$Var(X) = E(X^2) - [E(X)]^2 \quad (4.11)$$

$$Var(X) = 199 - 13^2 \quad (4.12)$$

$$Var(X) = 30. \quad (4.13)$$

Note que, ao medir a dispersão dos possíveis valores da ação, a variância é uma medida do risco da ação.

4.2 Variável Aleatória Discreta

Quando os valores da variável são específicas, ou seja, não pode assumir qualquer valor dentro do conjunto dos números reais.

4.3 Distribuição Uniforme

Na **distribuição uniforme** todos os elementos têm a mesma probabilidade de ocorrer.

Exemplo: obter um número de 1 a 6, lançando um dado não viciado. A probabilidade de obter qualquer um dos seis números é de $1/6$.

Exemplo numérico sobre Distribuição Uniforme (Sartoris, 2013, p.57)

Este exemplo numérico sobre distribuição uniforme está na página 57 de Sartoris (2013).

Joga-se um dado uma única vez. Qual o valor esperado do número obtido? E sua variância?

Resposta: O valor esperado é

$$E(X) = 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + 3 \times \frac{1}{6} + 4 \times \frac{1}{6} + 5 \times \frac{1}{6} + 6 \times \frac{1}{6} \quad (4.14)$$

$$E(X) = 3,5. \quad (4.15)$$

Para calcular a variância pela forma alternativa é necessário $E(X^2)$.

$$E(X^2) = (1)^2 \frac{1}{6} + (2)^2 \frac{1}{6} + (3)^2 \frac{1}{6} + (4)^2 \frac{1}{6} + (5)^2 \frac{1}{6} + (6)^2 \frac{1}{6} \quad (4.16)$$

$$E(X^2) = 1 \frac{1}{6} + 4 \frac{1}{6} + 9 \frac{1}{6} + 16 \frac{1}{6} + 25 \frac{1}{6} + 36 \frac{1}{6} \quad (4.17)$$

$$E(X^2) = \frac{91}{6}. \quad (4.18)$$

Desta forma a variância é

$$Var(X) = E(X^2) - [E(X)]^2 \quad (4.19)$$

$$Var(X) = \frac{91}{6} - \left(\frac{21}{6}\right)^2 \quad (4.20)$$

$$Var(X) = \frac{105}{36} \quad (4.21)$$

$$Var(X) \cong 2,92 \quad (4.22)$$

4.4 Distribuição de Bernoulli

A **distribuição de Bernoulli** caracteriza-se pela existência de apenas dois eventos, mutuamente exclusivos, denominados **sucesso** e **fracasso**, num experimento que é realizado uma única vez.

Se a probabilidade de **sucesso** é p , a probabilidade de **fracasso** é $(1 - p)$, uma vez que só existem esses dois eventos e eles são mutuamente exclusivos.

Exemplos de distribuição de Bernoulli

- **lançamento de uma moeda uma única vez.** Se o resultado cara é o sucesso, associado a uma probabilidade $p = 1/2$, então o resultado coroa é o fracasso associado a uma probabilidade $(1 - p) = 1/2$.
- **lançamento de um dado uma única vez.** Se o número escolhido é três, então o resultado 3 é o sucesso, associado a uma probabilidade $p = 1/6$ e qualquer um dos outros cinco números é o fracasso, associado a uma probabilidade $(1 - p) = 5/6$.

Exemplo numérico sobre distribuição de Bernoulli (Sartoris, 2013, p.58)

Este exemplo numérico sobre distribuição de Bernoulli está na página 58 de Sartoris (2013).

No caso da cara ou coroa, atribuindo o valor 1 para o sucesso e 0 para o fracasso, determine a média e a variância do resultado após uma jogada.

Resposta: A média é

$$E(X) = 1 \times \frac{1}{2} + 0 \times \frac{1}{2} \quad (4.23)$$

$$E(X) = \frac{1}{2} \quad (4.24)$$

$$E(X) = 0,5 \quad (4.25)$$

E a variância

$$E(X^2) = 1^2 \times \frac{1}{2} + 0^2 \times \frac{1}{2} \quad (4.26)$$

$$E(X^2) = \frac{1}{2} \quad (4.27)$$

$$E(X^2) = 0,5 \quad (4.28)$$

$$Var = E(X^2) - [E(X)]^2 \quad (4.29)$$

$$Var = 0,5 - (0,5)^2 \quad (4.30)$$

$$Var = 0,25. \quad (4.31)$$

Outro exemplo numérico sobre distribuição de Bernoulli (Sartoris, 2013, p.59)

No caso do dado, em que se aposta em um único número, atribuindo o valor 1 para o sucesso e 0 para o fracasso, determine a média e a variância do resultado após uma jogada.

Resposta: A média é

$$E(X) = 1 \times \frac{1}{6} + 0 \times \frac{5}{6} \quad (4.32)$$

$$E(X) = \frac{1}{6} \quad (4.33)$$

E a variância

$$E(X^2) = 1^2 \times \frac{1}{6} + 0^2 \times \frac{5}{6} \quad (4.34)$$

$$E(X^2) = \frac{1}{6} \quad (4.35)$$

$$Var = E(X^2) - [E(X)]^2 \quad (4.36)$$

$$Var = \frac{1}{6} - \left(\frac{1}{6}\right)^2 \quad (4.37)$$

$$Var = \frac{5}{36} \quad (4.38)$$

Média e Variância da Distribuição de Bernoulli

Pelos exemplos numéricos apresentado, verifica-se que numa distribuição de Bernoulli:

$$E(X) = p \quad (4.39)$$

$$Var(X) = p(1 - p) \quad (4.40)$$

4.5 Distribuição Binomial

A **distribuição Binomial** é a generalização da distribuição de Bernoulli. Ou seja, há um sucesso com probabilidade p e um fracasso com probabilidade $(1 - p)$. Mas o número de experimentos ou jogadas pode ser qualquer um.

Retomemos o exemplo da cara ou coroa, mas agora com três jogadas ou experimentos.

Associa-se cara a uma letra C e uma coroa a uma letra K para facilitar a visualização.

| jogada 1 | jogada 2 | jogada 3 |
|----------|----------|----------|
| C | C | C |
| C | C | K |
| C | K | C |
| C | K | K |
| K | C | C |
| K | C | K |
| K | K | C |
| K | K | K |

Na primeira jogada só há dois resultados possíveis: cara ou coroa. Portanto

$$P(C) = p = 1/2$$

$$P(K) = (1 - p) = 1/2$$

Na segunda jogada há quatro combinações para três resultados possíveis que são (cara, cara); (cara, coroa) e (coroa, coroa). Portanto

$$P(C, C) = pp = 1/4$$

$$P(C, K) + P(K, C) = p(1 - p) + (1 - p)p = 2p(1 - p) = 2/4$$

$$P(K, K) = (1 - p)(1 - p) = 1/4$$

Na terceira jogada há oito combinações para quatro resultados possíveis que são (cara, cara, cara); (cara, cara, coroa); (cara, coroa, coroa) e (coroa, coroa, coroa). Portanto

$$\begin{aligned}
P(C, C, C) &= 1/8 \\
P(C, C, K) + P(C, K, C) + P(K, C, C) &= 3/8 \\
P(K, K, C) + P(K, C, K) + P(C, K, K) &= 3/8 \\
P(K, K, K) &= 1/8
\end{aligned}$$

Ou em detalhes:

$$\begin{aligned}
P(3 \text{ caras}) &= ppp = 1/8 \\
P(2 \text{ caras, 1 coroa}) &= pp(1-p) + p(1-p)p + (1-p)pp = 3pp(1-p) = 3/8 \\
P(2 \text{ coroas, 1 cara}) &= (1-p)(1-p)p + (1-p)p(1-p) + p(1-p)(1-p) \\
&= 3(1-p)(1-p)p = 3/8 \\
P(3 \text{ coroas}) &= (1-p)(1-p)(1-p) = 1/8
\end{aligned}$$

Ou seja,

$$\begin{aligned}
P(3 \text{ caras}) &= 1 \times p \times p \times p \\
P(2 \text{ caras, 1 coroa}) &= 3 \times p \times p \times (1-p) \\
P(2 \text{ coroas, 1 cara}) &= 3 \times (1-p) \times (1-p) \times p \\
P(3 \text{ coroas}) &= 1 \times (1-p) \times (1-p) \times (1-p)
\end{aligned}$$

usando o binômio de newton,

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

$$\begin{aligned}
P(3 \text{ caras}) &= \binom{3}{3} \times p \times p \times p \\
P(2 \text{ caras, 1 coroa}) &= \binom{3}{2} \times p \times p \times (1-p) \\
P(2 \text{ coroas, 1 cara}) &= \binom{3}{1} \times (1-p) \times (1-p) \times p \\
P(3 \text{ coroas}) &= \binom{3}{0} \times (1-p) \times (1-p) \times (1-p)
\end{aligned}$$

Generalizando, nos leva a função de probabilidade,

$$P(X = k) = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$$

onde

- k é o número de sucessos;
- n é o número de jogadas ou experimentos;
- p é a probabilidade de sucesso;
- $(1-p)$ é a probabilidade de fracasso.

O primeiro exemplo numérico de distribuição binomial

O primeiro exemplo numérico de distribuição binomial é o exemplo numérico 3.2.3.1 da página 62 do Sartoris (2013). Suponha um jogo de dados em que aposta em um único número. Determine as seguintes probabilidades:

- em três jogadas, ganhar duas.

$$\begin{aligned} P(X = 2) &= \binom{3}{2} \times \left(\frac{1}{6}\right)^2 \times \left(\frac{5}{6}\right)^{(3-2)} \\ P(X = 2) &= \frac{3!}{2!(3-2)!} \times \frac{1}{36} \times \frac{5}{6} \\ P(X = 2) &= 3 \times \frac{1}{36} \times \frac{5}{6} \\ P(X = 2) &= \frac{15}{216} = 0,0694 \end{aligned}$$

- em quatro jogadas, ganhar duas.

$$\begin{aligned} P(X = 2) &= \binom{4}{2} \times \left(\frac{1}{6}\right)^2 \times \left(\frac{5}{6}\right)^{(4-2)} \\ P(X = 2) &= \frac{4!}{2!(4-2)!} \times \frac{1}{36} \times \frac{25}{36} \\ P(X = 2) &= 6 \times \frac{1}{36} \times \frac{25}{36} \\ P(X = 2) &= \frac{150}{1.296} = 0,1157 \end{aligned}$$

- Em cinco jogadas, ganhar três.

$$\begin{aligned}
P(X=3) &= \binom{5}{3} \times \left(\frac{1}{6}\right)^3 \times \left(\frac{5}{6}\right)^{(5-3)} \\
P(X=3) &= \frac{5!}{3!(5-3)!} \times \frac{1}{216} \times \frac{25}{36} \\
P(X=3) &= 10 \times \frac{1}{216} \times \frac{25}{36} \\
P(X=3) &= \frac{250}{7.776} = 0,0321
\end{aligned}$$

Um segundo exemplo numérico de distribuição binomial

O segundo exemplo numérico sobre distribuição binomial é da página 63 de Sartoris (2013). Calcule a média e a variância no jogo de cara ou coroa, atribuindo valor um para cara e zero para coroa.

Considerando uma ,duas e três jogadas.

- Para uma jogada, que é o caso particular da distribuição de Bernoulli.

$$E(X) = p = \frac{1}{2}$$

$$var(X) = p(1-p) = \frac{1}{2} \left(1 - \frac{1}{2}\right) = \frac{1}{4}$$

- Para duas jogadas: CC; CK; KC; KK. Ou seja,

$$E(X) = 2 \times \frac{1}{4} + 1 \times \frac{2}{4} + 0 \times \frac{1}{4} = \frac{4}{4} = 1$$

$$E(X^2) = 2^2 \times \frac{1}{4} + 1^2 \times \frac{2}{4} + 0^2 \times \frac{1}{4} = \frac{6}{4} = 1,5$$

$$var(X) = 1,5 - [1]^2 = 0,5$$

- Para três jogadas: CCC; CCK; CKC; KCC; CKK; KCK; KKC; KKK. Ou seja,

$$E(X) = 3 \times \frac{1}{8} + 2 \times \frac{3}{8} + 1 \times \frac{3}{8} + 0 \times \frac{1}{8} = \frac{12}{8} = 1,5$$

$$E(X^2) = 3^2 \times \frac{1}{8} + 2^2 \times \frac{3}{8} + 1^2 \times \frac{3}{8} + 0^2 \times \frac{1}{8} = \frac{24}{8} = 3$$

$$var(X) = 3 - [1,5]^2 = 0,75$$

4.5.1 Média e Variância da distribuição binomial

Com base nos resultados do exemplo numérico anterior tem-se:

A média da distribuição binomial

$$E(X) = np$$

A variância da distribuição binomial

$$\text{var}(X) = np(1 - p)$$

4.5.2 Distribuição Binomial usando o R

Considerando que:

x é um vetor de números;

p é um vetor de probabilidades;

n é o número de observações;

tamanho é o número de experimentos ou tentativas e;

prob é a probabilidade de sucesso de cada experimento ou tentativa.

Seguem-se quatro funções no R para a distribuição binomial.

dbinom(x, tamanho, prob):

esta função entrega a distribuição densidade de probabilidade a cada ponto.

pbinom(x, tamanho, prob):

esta função entrega a probabilidade cumulativa de um evento. Ou seja, é um único valor representando probabilidade.

qbinom(p, tamanho, prob):

esta função toma o valor da probabilidade e entrega um número cujo valor cumulativo coincide com o valor da probabilidade.

rbinom(n, tamanho, prob):

esta função gera o número requerido de valores aleatórios de uma dada probabilidade.

Exemplo numérico de distribuição binomial resolvido usando o R

O primeiro exemplo numérico 3.2.3.1 da página 62 de Sartoris (2013) resolvido anteriormente é agora resolvido usando-se o R. Suponha um jogo de dados em que aposta em um único número. Determine as seguintes probabilidades:

- em três jogadas, ganhar duas.

```
library(MASS)

p1 <- dbinom(2, 3, 1/6)
p1
```

```
## [1] 0,06944444

p1frac <- fractions(p1)
p1frac
```

```
## [1] 5/72
```

- em quatro jogadas, ganhar duas.

```
library(MASS)

p2 <- dbinom(2, 4, 1/6)
p2
```

```
## [1] 0,1157407

p2frac <- fractions(p2)
p2frac
```

```
## [1] 25/216
```

- Em cinco jogadas, ganhar três.

```
library(MASS)

p3 <- dbinom(3, 5, 1/6)
p3
```

```
## [1] 0,03215021

p3frac <- fractions(p3)
p3frac
```

```
## [1] 125/3888
```

O segundo exemplo numérico de distribuição binomial usando R

O segundo exemplo numérico resolvido é o exemplo 3.2.3.2 da página 63 de Sartoris (2013) é agora resolvido usando-se o R. Calcule a média e a variância no jogo de cara ou coroa, atribuindo valor um para cara e zero para coroa. Considerando uma ,duas e três jogadas.

- Para uma jogada, que é o caso particular da distribuição de Bernoulli.

```
media1 <- 1 * 0.5
media1
```

```
## [1] 0,5
```

```
var1 <- 1 * 0.5 * (1 - 0.5)
var1
```

```
## [1] 0,25
```

- Para duas jogadas: CC; CK; KC; KK.

```
media2 <- 2 * 0.5
media2
```

```
## [1] 1
```

```
var2 <- 2 * 0.5 * (1 - 0.5)
var2
```

```
## [1] 0,5
```

- Para três jogadas: CCC; CCK; CKC; KCC; CKK; KCK; KKC; KKK.

```
media3 <- 3 * 0.5
media3
```

```
## [1] 1,5
```

```
var3 <- 3 * 0.5 * (1 - 0.5)
var3
```

```
## [1] 0,75
```

4.6 Distribuição de Poisson

- Morte por patada de cavalo, mesmo para a época em que o principal meio de transporte era a cavalo, era um acontecimento pouco usual.
- Celular do professor tocar durante a aula de Estatística Econômica poderia ser um exemplo de evento muito difícil de ocorrer.
- Para estas situações a probabilidade associada ao sucesso p é muito pequena, mesmo considerando muitos meses ou muitas aulas. Ou seja, n ser muito grande.
- Por isso, pode se pensar que p tende a zero enquanto que n tende ao ∞ .

4.6.1 Média e a Variância da Distribuição de Poisson

Partindo da distribuição binomial e como $p \rightarrow 0$ e $n \rightarrow \infty$ então

$$E(X) = np = \lambda.$$

Ou seja, np é igual a um número finito λ que significa o número médio de vezes que o evento ocorre.

Para a variância tem-se que

$$\text{var}(X) = \underbrace{np}_{\lambda} \underbrace{(1-p)}_1 = \lambda.$$

Ou seja, a média e a variância da distribuição de Poisson são iguais a λ .

4.6.2 A função de probabilidade da distribuição de Poisson

A função de probabilidade para a distribuição de Poisson pode ser obtida a partir da função de probabilidade da distribuição binomial fazendo $p \rightarrow 0$ e $n \rightarrow \infty$.

$$P(X = k) = \frac{e^{-\lambda} \lambda^k}{k!}$$

A derivação matemática está no apêndice do capítulo 3 do Sartoris (2013).

Primeiro exemplo numérico de distribuição de Poisson

Exemplo 3.2.6.1 da página 67 de Sartoris (2013). Suponha que, em média, o telefone toque quatro vezes ao dia em uma casa. Qual a probabilidade de que, em certo dia, ele toque, no máximo, duas vezes?

Resposta: Trata-se de um problema sobre distribuição de Poisson cujo parâmetro é $\lambda = 4$. Note que pede-se a probabilidade de tocar **no máximo** duas vezes: $P(0) + P(1) + P(2)$.

$$\begin{aligned} P(X = 0) &= \frac{e^{-4} 4^0}{0!} = 1e^{-4} \\ P(X = 1) &= \frac{e^{-4} 4^1}{1!} = 4e^{-4} \\ P(X = 2) &= \frac{e^{-4} 4^2}{2!} = 8e^{-4} \end{aligned}$$

Portanto:

$$P(X \leq 2) = 13e^{-4} \cong 0,2381 = 23,81\%$$

Segundo exemplo numérico de distribuição de Poisson

Exemplo 3.2.6.2 da página 67 de Sartoris (2013). Um candidato tem apenas 2% das intenções de voto. Qual a probabilidade de que, em 100 eleitores escolhidos ao acaso, encontrem-se cinco que desejem votar nesse candidato?

Resposta: Usando a binomial pura e simplesmente

$$P(X = 5) = \frac{100!}{5!(100-5)!} \times 0,02^5 \times 0,98^{95} \cong 0,0353 = 3,53\%$$

Note que o exercício pode ser resolvido usando a distribuição de Poisson como aproximação, tendo como parâmetro $\lambda = np = 100 \times 0,02 = 2$.

$$P(X = 5) = \frac{e^{-2}2^5}{5!} \cong 0,0361 = 3,61\%$$

4.6.3 Distribuição de Poisson usando o R

Considerando que:

x é um vetor de números;

q é um vetor de quantiles;

p é um vetor de probabilidades;

n é o número de observações;

lambda é o vetor de médias não negativas;

Seguem-se quatro funções no R para a distribuição de Poisson.

dpois(x, lambda) :

esta função entrega a distribuição densidade de probabilidade a cada ponto.

ppois(q, lambda):

esta função entrega a probabilidade cumulativa de um evento. Ou seja, é um único valor representando probabilidade.

qpois(p, lambda):

esta função toma o valor da probabilidade e entrega um número cujo valor cumulativo coincide com o valor da probabilidade.

rpois(n, lambda):

esta função gera o número de requerido de valores aleatórios de uma dada probabilidade a partir de uma dada amostra.

O primeiro exemplo numérico de distribuição de Poisson resolvido usando R

Exemplo 3.2.6.1 da página 67 de Sartoris (2013). Suponha que, em média, o telefone toque quatro vezes ao dia em uma casa. Qual a probabilidade de que, em certo dia, ele toque, no máximo, duas vezes?

```
p4 <- round(ppois(2, 4), 4)
p4
```

```
## [1] 0,2381
```

O segundo exemplo numérico de distribuição de Poisson resolvido usando R

Exemplo 3.2.6.2 da página 67 de Sartoris (2013). Um candidato tem apenas 2% das intenções de voto. Qual a probabilidade de que, em 100 eleitores escolhidos ao acaso, encontrem-se cinco que desejem votar nesse candidato?

Usando distribuição binomial:

```
p5 <- round(dbinom(5, 100, 0.02), 4)
p5
```

```
## [1] 0,0353
```

Usando a distribuição de Poisson como aproximação e sabendo que $\lambda = np = 100 \times 0,02 = 2$

```
p6 <- round(dpois(5, 2), 4)
p6
```

```
## [1] 0,0361
```

Capítulo 5

Variável Aleatória Contínua

5.1 Distribuições contínuas

Para entender a diferença entre variável aleatória discreta e a variável aleatória contínua é necessário definir cada uma delas:

- Variável Aleatória Discreta: assume valores específicos. Exemplo: número de clientes que entram na loja, número de mortes, número de nascimentos, etc.
- Variável Aleatória Contínua: assume qualquer valor no domínio dos números reais. Por isso entre os valores 2 e 3, por exemplo, pode existir infinitos valores para uma variável aleatória contínua. Exemplo: peso, altura, preço de commodities, temperatura, quantidade de chuva em milímetros, etc.

Sartoris (2013) exemplifica as duas situações de variáveis discretas e contínuas usando o exemplo de relógio digital e de relógio analógico considerando somente a hora como valor da variável. No caso do relógio digital a mudança de valor ocorre de forma discreta, diferentemente da situação do relógio analógico que a mudança de horas é de forma gradual.

Para o caso discreto é possível calcular a probabilidade para os valores específicos. Por isso

$$P(2 < x < 3) \neq P(2 \leq x \leq 3) \neq P(2 < x \leq 3) \neq P(2 \leq x < 3)$$

Já para o caso contínuo não é possível calcular a probabilidade para um valor específico da variável. Inclusive, no limite, a probabilidade de um valor específico para uma variável aleatória contínua é igual zero. Portanto,

$$P(2 < x < 3) = P(2 \leq x \leq 3) = P(2 < x \leq 3) = P(2 \leq x < 3)$$

A situação do exemplo do relógio analógico do Sartoris (2013), como a mudança de hora ocorre de forma gradual e a probabilidade é igual para qualquer valor da hora, trata-se de uma distribuição de probabilidade uniforme contínua. Gráfico gerado usando a função `ggplot` do pacote `ggplot2`.

```
library(ggplot2)

hora <- c(1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12)
prob <- c(1/12, 1/12, 1/12, 1/12, 1/12, 1/12, 1/12, 1/12, 1/12, 1/12, 1/12, 1/12)
horaanalogica <- as.data.frame(cbind(hora, prob))
horaanalogica

##      hora      prob
## 1      1 0,08333333
## 2      2 0,08333333
## 3      3 0,08333333
## 4      4 0,08333333
## 5      5 0,08333333
## 6      6 0,08333333
## 7      7 0,08333333
## 8      8 0,08333333
## 9      9 0,08333333
## 10     10 0,08333333
## 11     11 0,08333333
## 12     12 0,08333333

ggplot(horaanalogica, aes(hora, prob)) + geom_line() +
  xlim(0, 13) + ylim(0, 0.2) + scale_x_continuous(breaks = seq(1,
    12, by = 1))
```

Nesta situação apresentada na figura 5.1, matematicamente faz sentido calcular a probabilidade para um intervalo de valores da variável e não para valores específicos com é no caso da variável aleatória discreta. Dessa forma com a função de probabilidade da distribuição, denominada de **função densidade de probabilidade** (f.d.p.), pode-se obter a probabilidade obtendo a área abaixo da curva desta função. Matematicamente, a probabilidade pode ser obtida através da integral definida desta função para o intervalo de valores em questão.

Note que para $f(x)$ ser função densidade de probabilidade, precisa atender as seguintes propriedades definidas pela teoria da probabilidade:

- a soma das probabilidades para todo os valores definidos da variável x dever ser igual a 1.
- a probabilidade obtida através de $f(x)$ não pode ser negativa.

No exemplo da relógio analógico a probabilidade da hora estar entre 2 e 3 é a área da curva ou reta da distribuição uniforme contínua cujo valor constante

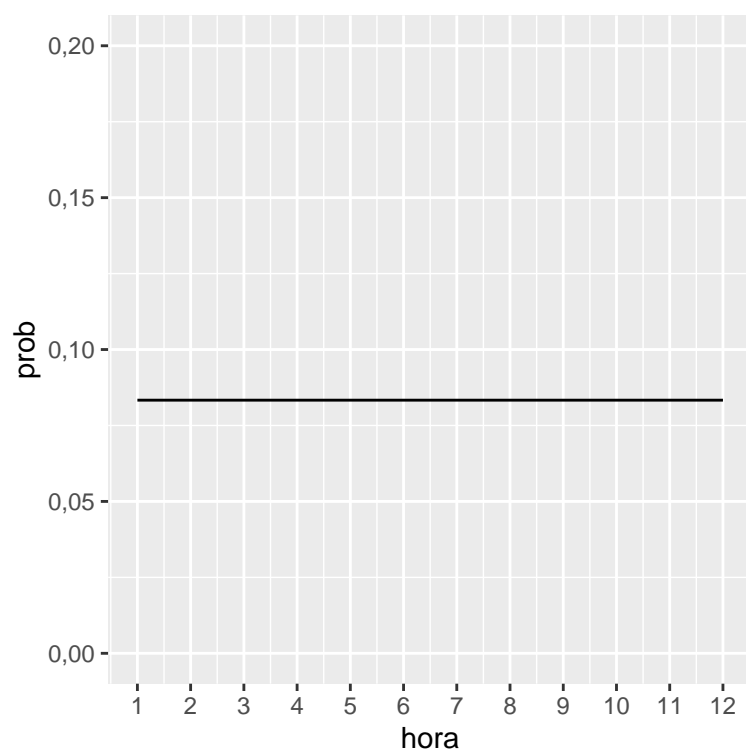


Figura 5.1: Um exemplo de distribuição uniforme Contínua.

da probabilidade é $1/12$. O intervalo entre 2 e 3 proporciona um intervalo equivalente a unidade, uma vez que intervalo pode ser calculado entre a diferença entre o valor final e o valor inicial. Dessa forma:

$$P(2 < x < 3) = \frac{1}{12} \times (3 - 2) = \frac{1}{12},$$

ou seja,

$$P(2 < x < 3) = P(2 \leq x \leq 3) = P(2 < x \leq 3) = P(2 \leq x < 3) = \frac{1}{12}$$

Exemplo numérico sobre função densidade de probabilidade

Exemplo 4.1.3 da página 79 do Sartoris (2013). Dada a função densidade de probabilidade da variável aleatória contínua mostrada a seguir

$$f(x) = \begin{cases} Ax & 0 \leq x \leq 3 \\ 0, & X < 0 \text{ ou } x > 3 \end{cases}$$

a) Determine o valor de A.

Resposta: Dado que $f(x) = Ax$

$$f(3) = 3A$$

para $x = 3$ e

$$f(0) = 0$$

para $x = 0$

Graficamente seria um triângulo de base 3 e altura de $3A$. Como a função definida de modo que

$$f(x) \geq 0$$

ou seja resulte valores não-negativos, basta igualar a área a 1, pois a soma das probabilidades para todos os valores definidos para a variável x deve ser igual a 1. Portanto a probabilidade é a área correspondente ao triângulo de base 3 e altura $3A$. Ou seja,

$$\frac{3A \times 3}{2} = 1$$

e por isso

$$A = \frac{2}{9}$$

Ou seja,

$$f(x) = \frac{2}{9}x$$

b) Determine a probabilidade de que x esteja entre 2 e 3.

Para $x = 2$ se tem

$$f(2) = \frac{2}{9} \times 2 = \frac{4}{9}$$

e para $x = 3$

$$f(3) = \frac{2}{9} \times 3 = \frac{2}{3}$$

A área correspondente a probabilidade de x estar entre 2 e 3 pode ser obtida subtraindo a área do triângulo maior que seria com x entre 0 e 3 e o triângulo menor com x entre 2 e 3. Ou seja, a área do triângulo maior é calculado como

$$\text{triângulo maior} = 3 \times \frac{2}{3} \times \frac{1}{2} = 1$$

e a área do triângulo menor é calculado como

$$\text{triângulo menor} = 2 \times \frac{4}{9} \times \frac{1}{2} = \frac{4}{9}.$$

Assim a diferença das áreas do triângulo maior e do triângulo menor é igual a $5/9$. Ou seja,

$$P(2 < x < 3) = 1 - \frac{4}{9} = \frac{5}{9}$$

Segundo exemplo numérico sobre função densidade de probabilidade

Exemplo 4.1.4 da página 80 do Sartoris (2013). Dada da f.d.p. de uma variável aleatória contínua,

$$f(x) = \begin{cases} Ax^2 & \text{para } 0 \leq x \leq 1 \\ 0, & \text{para } x < 0 \text{ ou } x > 1 \end{cases}$$

- a) Determine o valor da constante A. Note que a função já não é mais do tipo linear e por isso não é necessário utilizar a integral da função em questão para calcular a probabilidade. Considerando novamente que a soma das probabilidades para todos os valores definidos para a variável x deve ser igual a um, se tem

$$\int_{-\infty}^{+\infty} f(x)dx = 1$$

Para este exemplo, a função $f(x)$ só está definido para os valores entre 0 e 1, sendo que para valores de x menor que 0 e valores de x maior que 1 a função $f(x)$ assume valor igual a zero. Portanto,

$$\int_0^1 Ax^2 dx = 1$$

Como A é constante, pode-se colocar para fora da integral

$$A \int_0^1 x^2 dx = 1$$

Calculando a integral definida

$$A \left[\frac{x^3}{3} \right]_0^1 = 1$$

Avaliando para os valores 1 e 0 e depois realizando a diferença

$$A \left[\frac{1}{3} - \frac{0}{3} \right]_0^1 = 1$$

Ou seja,

$$A = 3.$$

Ou seja a função é

$$f(x) = 3x^2.$$

b) Dtermine a probabilidade de que x esteja entre 0,5 e 1.

Primeiramente verifica-se que tanto 0,5 como 1 são valores que estão dentro do intervalo de x definido para a função em questão. Por exemplo, se um dos limites fosse igual a 2, consideraria $x = 1$ pois $f(x) = 0$ para $x > 1$. Por isso basta calcular a integral da função para os valores de x entre 0,5 e 1. Ou seja,

$$P(0,5 \leq x \leq 1) = \int_{0,5}^1 3x^2 dx$$

$$P(0,5 \leq x \leq 1) = \left[\frac{3x^3}{3} \right]_{0,5}^1$$

$$P(0,5 \leq x \leq 1) = 1^3 - 0,5^3 = 1 - 0,125 = 0,875.$$

Resumindo, para que uma função qualquer seja uma função densidade de probabilidade precisa atender duas condições:

$$\int_{-\infty}^{+\infty} f(x) dx = 1$$

e

$$f(x) \geq 0 \text{ para todos os valores de } x$$

Exemplo numérico sobre a distribuição exponencial

O exemplo 4.1.5 da página 82 de Sartoris (2013) é uma simples aplicação para a distribuição exponencial. Dada a f.d.p. da variável aleatória contínua x

$$f(x) = \begin{cases} Ae^{-\alpha x}, & \text{para } x \geq 0 \\ 0, & \text{para } x < 0 \end{cases}$$

a) Determine o valor de A. Então como foi informado no anúncio do exemplo numérico, essa distribuição especial é conhecida com distribuição exponencial. Como

$$\int_{-\infty}^{+\infty} f(x)dx = 1$$

e a função assume valores iguais a zero para valores negativos de x , se tem

$$\int_0^{+\infty} Ae^{-\alpha x} = 1.$$

Como A é constante, pode-se colocar par fora da integral

$$A \int_0^{+\infty} e^{-\alpha x} dx = 1$$

e sabendo que a integral da função é um pouco diferente, considerando que há um α multiplicando x , temos

$$A \left[\frac{e^{-\alpha x}}{-\alpha} \right]_0^{+\infty} = 1$$

$$A \left[0 - \left(-\frac{1}{\alpha} \right) \right] = 1$$

Portanto,

$$A = \alpha$$

5.1.1 Função de distribuição de variáveis contínuas

A semelhança do caso das variáveis discretas, a **função de distribuição acumulada** ou **função de distribuição** $F(x)$ é a soma da probabilidades de todos os valores possíveis que a variável x pode assumir até o valor de x propriamente dito. Isso é feito através da integral da seguinte forma

$$F(x) = \int_{-\infty}^x f(t)dt.$$

Portanto, matematicamente, $f(x)$ é a derivada da função $F(x)$

$$f(x) = \frac{dF(x)}{dx}.$$

Exemplo numérico da função de distribuição

O exemplo 4.2.1 da página 83 do Sartoris (2013), trata-se da obtenção da função distribuição exponencial através da sua respectiva função densidade de probabilidade. Dada a função f.d.p. da distribuição exponencial, determine a função de distribuição correspondente.

$$f(x) = \begin{cases} e^{-x}, & \text{para } x \geq 0 \\ 0, & \text{para } x < 0 \end{cases}$$

Dado que a função só é definida para $x \geq 0$, o limite de integração inferior é zero.

$$F(x) = \int_0^{+\infty} f(t) dt$$

substituindo a f.d.p.

$$F(x) = \int_0^{+\infty} e^{-t} dt$$

lembrando a integral de e^{-x} e que todo número elevado a zero é igual a um

$$F(x) = [-e^{-t}]_0^x.$$

Assim,

$$F(x) = -e^{-x} + e^0$$

e

$$F(x) = 1 - e^{-x}$$

Portanto a função distribuição é dada por

$$F(x) = \begin{cases} 1 - e^{-x}, & \text{para } x \geq 0 \\ 0, & \text{para } x \leq 0 \end{cases}$$

Segundo Exemplo numérico sobre função de distribuição

Exemplo 4.2.2 da página 84 de Sartoris (2013). Seja a função distribuição

$$F(x) = \begin{cases} 0,5(x^3 + 1), & \text{para } -1 \leq x \leq 1 \\ 0, & \text{para } x < -1 \\ 1, & \text{para } x > 1 \end{cases}$$

determine a função densidade de probabilidade correspondente.

Resposta:

A função densidade de probabilidade é dado por

$$f(x) = \frac{dF(x)}{dx}$$

$$f(x) = \frac{d(0,5x^2 + 1)}{dx}$$

$$f(x) = 3 \times 0,5x^2 + 0$$

$$f(x) = 1,5x^2$$

Portanto, a função densidade de probabilidade é:

$$f(x) = \begin{cases} 1,5x^2, & \text{para } -1 \leq x \leq 1 \\ 0, & \text{para } x < -1 \text{ ou } x > 1 \end{cases}$$

A função distribuição $F(x)$, assim como a função densidade, deve atender dois requisitos:

- não pode ser negativa e deve ser menor ou igual a 1.

$$0 \leq F(x) \leq 1$$

,

- a soma das probabilidades de todos os valores da variável é igual a 1.

$$\lim_{x \rightarrow \infty} F(x) = 1$$

.

Estes requisitos são atendidos pelas funções distribuições $F(x)$ dos exemplos numéricos anteriores.

5.1.2 Esperança e variância de variáveis aleatórias contínuas

A esperança matemática para uma variável aleatória contínua é a soma contínua de todos os valores da variável com suas respectivas probabilidades. Uma soma contínua é a integral e, por sua vez, a probabilidade é encontrada pela função densidade de probabilidade. Dessa forma, matematicamente a esperança é

$$E(x) = \int_{-\infty}^{+\infty} xf(x) \, dx.$$

A variância por sua vez,

$$Var(x) = E[x - E(x)]^2$$

ou simplesmente

$$Var(x) = E[x - \mu]^2$$

onde μ é igual a média de x . Em termos de integral fica

$$Var(x) = \int_{-\infty}^{+\infty} (x - \mu)^2 dx$$

Alternativamente

$$Var(x) = E(x^2) - [E(x)]^2$$

que em termos de integral fica

$$Var(x) = \int_{-\infty}^{+\infty} x^2 f(x) dx - \left[\int_{-\infty}^{+\infty} xf(x) dx \right]^2$$

Exemplo numérico sobre esperança e variância

Exemplo 4.3.1. da página 85 do Sartoris (2013). Seja a f.d.p.

$$f(x) = \begin{cases} 3x^2 & \text{para } 0 \leq x \leq 1 \\ 0, & \text{para } x < 0 \text{ ou } x > 1 \end{cases}$$

a) Calcule o valor médio de x .

Resposta:

Como a média pode ser calculada através de $E(x)$

$$E(x) = \int_{-\infty}^{+\infty} xf(x) \, dx.$$

se tem

$$E(x) = \int_0^1 x 3x^2 dx$$

$$E(x) = 3 \int_0^1 x^3 dx$$

Calculando-se a integral

$$E(x) = 3 \left[\frac{x^4}{4} \right]_0^1$$

$$E(x) = 3 \times \frac{1^4}{4}$$

$$E(x) = \frac{3}{4} = 0,75$$

b) Calcule a variância de x .

Resposta:

Tomando a fórmula da variância alternativa

$$Var(x) = E(x^2) - [E(x)]^2$$

que em termos de integral é

$$Var(x) = \int_{-\infty}^{+\infty} x^2 f(x) dx - \left[\int_{-\infty}^{+\infty} x f(x) dx \right]^2$$

verifica-se que só falta só calcular $E(x^2)$ para poder calcular a variância de x .
Ou seja,

$$E(x^2) = \int_0^1 X^2 3x^2 dx$$

$$E(x^2) = 3 \int_0^1 x^4 dx$$

calculando a integral

$$E(x^2) = 3 \left[\frac{x^5}{5} \right]_0^1$$

$$E(x^2) = 3 \times \frac{1}{5}$$

$$E(x^2) = \frac{3}{5} = 0,6.$$

Assim a variância é

$$Var(x) = E(x^2) - [E(x)]^2$$

$$Var(x) = 0,6 - (0,75)^2 = 0,6 - 0,5625$$

$$Var(x) = 0,0375$$

c) Calcule o desvio padrão de x .

$$dp(x) = \sqrt{0,0375}$$

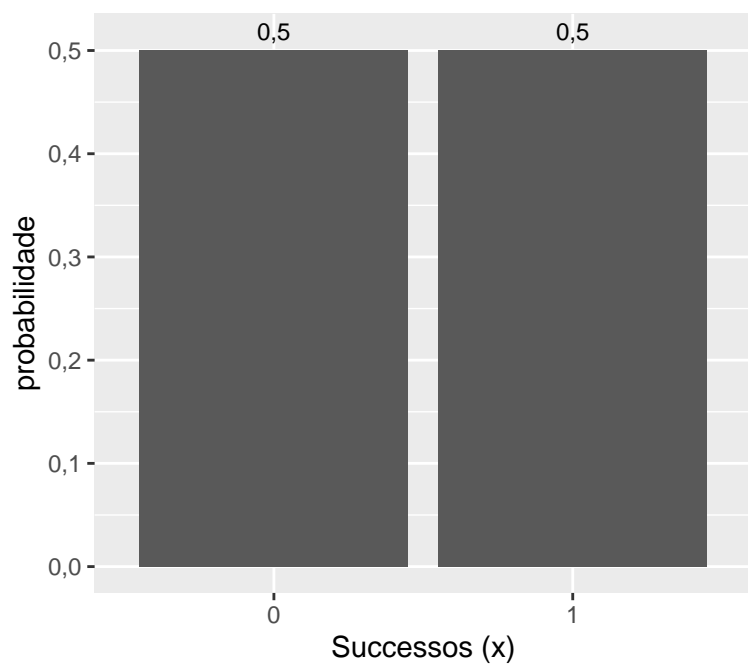
$$dp(x) \cong 0,194$$

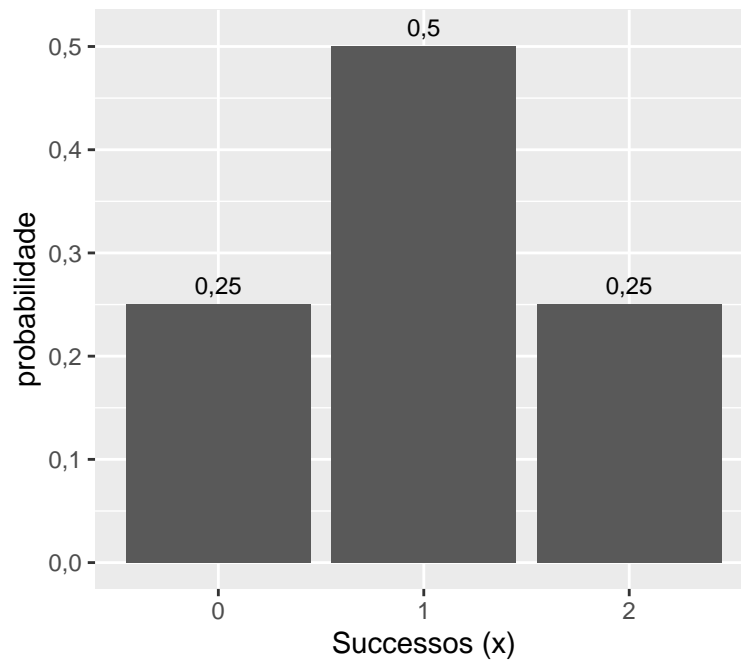
5.2 Distribuição Normal

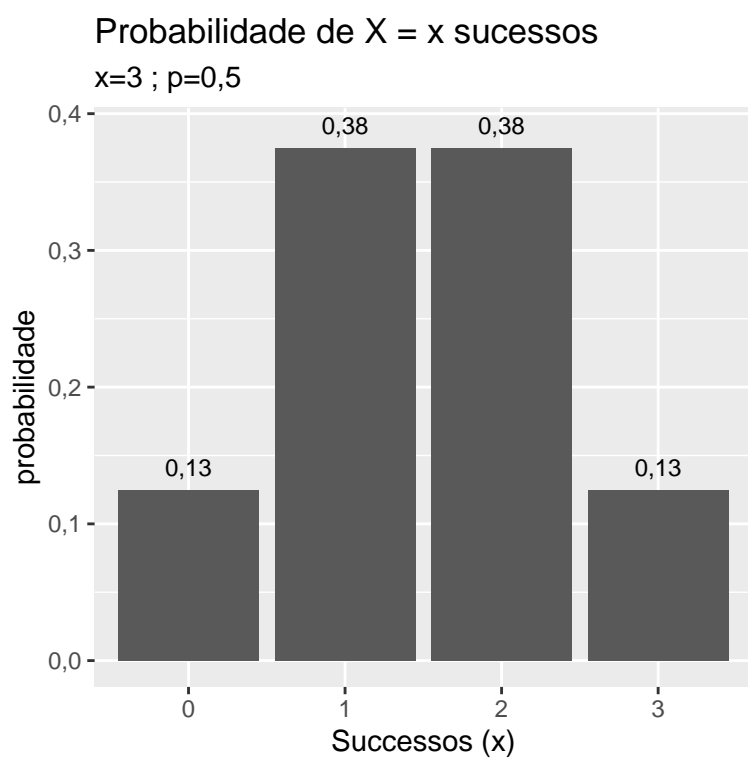
Para entender a distribuição normal, tome a distribuição binomial com uma probabilidade p igual 0,5 e n tendendo ao infinito. Gráficamente exemplifica-se com $n = 1, 2, 3, 5, 10$.

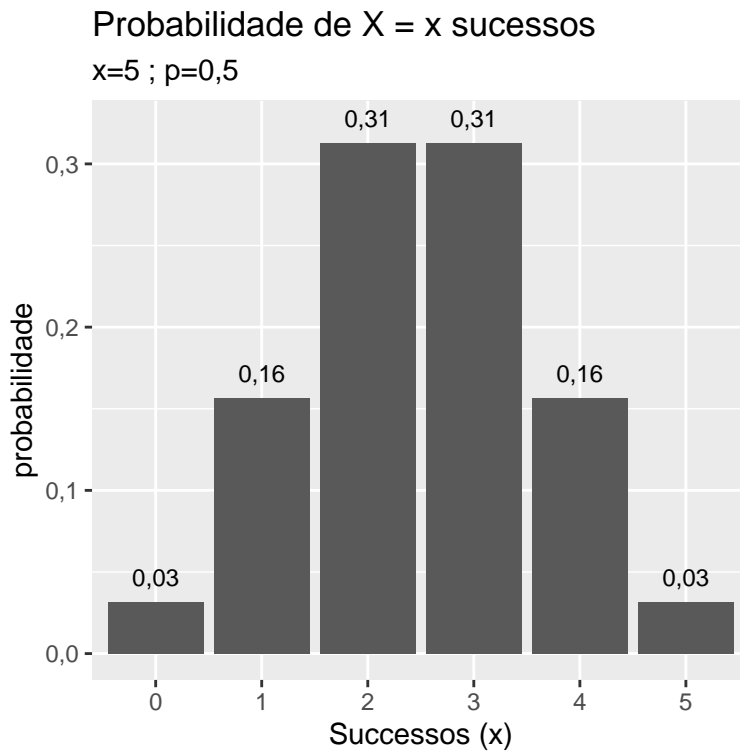
Probabilidade de $X = x$ sucessos

$x=1$; $p=0,5$



Probabilidade de $X = x$ sucessos $x=2$; $p=0,5$ 





Note na figura 5.2 que a medida que n aumenta a distribuição binomial assume um formato que é a de uma típica distribuição normal. Se no limite n tende ao infinito, a distribuição binomial perde o aspecto discreto na forma de “escada” e passa a ter um formato e aspecto de distribuição normal, figura 5.3, que é a referência para uma distribuição de variável aleatória contínua.

Essa distribuição de probabilidade é conhecida com **normal** ou **gaussiana** e a sua função densidade de probabilidade é dada por:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

onde μ é a média e σ é o desvio padrão da variável x . Se a variável x tem distribuição normal ou é normalmente distribuída, é usual escrever com notação matemática da seguinte forma

$$x \sim N(\mu, \sigma)$$

e deve ser lida da seguinte forma: x segue uma distribuição normal com média μ e desvio padrão σ . Em vez de desvio padrão essa notação matemática pode ter a variância σ^2 no lugar do desvio padrão σ .

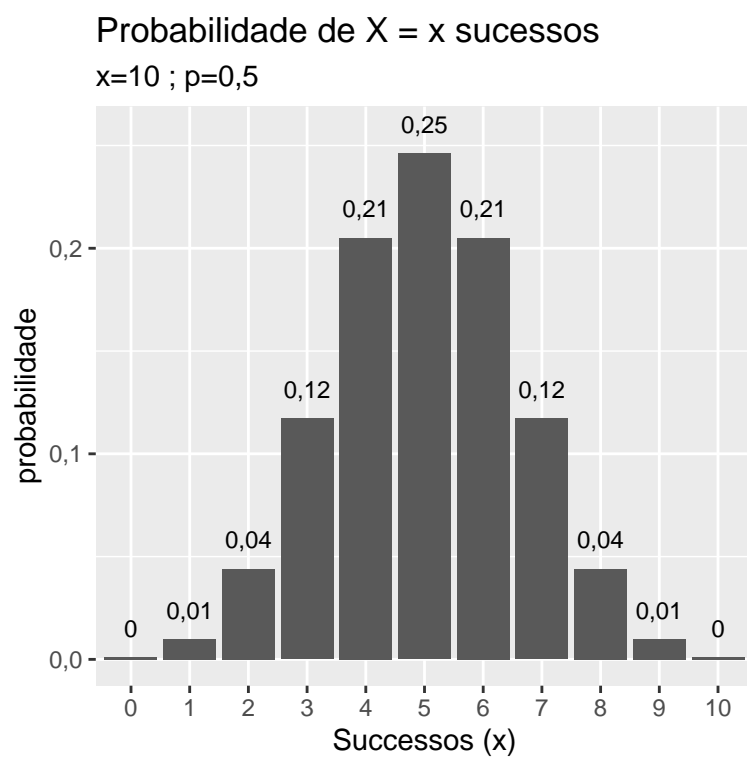


Figura 5.2: Da distribuição Binomial para a Distribuição Normal

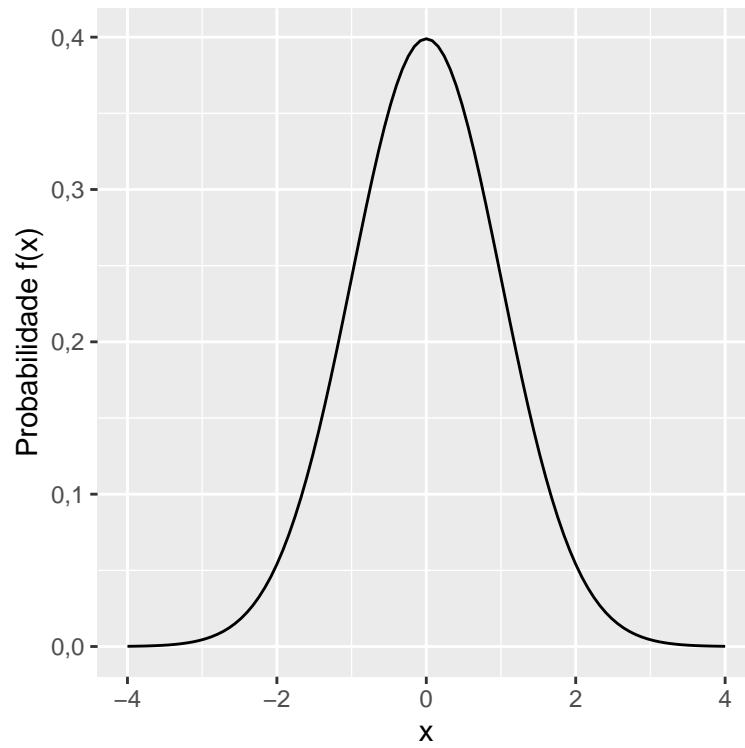


Figura 5.3: A distribuição Normal

Na função densidade de probabilidade da distribuição normal (5.2) os parâmetros média μ e desvio padrão σ definem uma família de distribuições normais para a variável x .

O parâmetro média μ determina a posição da curva em relação à origem, enquanto o desvio padrão σ determina se a curva será mais “**gorda**”, ou seja, mais dispersa com maior desvio padrão, ou mais “**magra**”, ou seja mais concentrada com um desvio padrão menor.

O cálculo das probabilidades sob uma distribuição normal é bastante trabalhosa, uma vez que não existe uma função cuja derivada é e^{-x^2} e, por isso, o cálculo é feito por métodos numéricos. Por isso, tabela-se os valores das integrais da distribuição normal padronizada, ou seja, a distribuição normal com média μ igual a zero e desvio padrão σ igual a zero.

$$\mu = 0$$

e

$$\sigma = 1$$

As variáveis normalmente distribuídas com média μ diferente de zero e desvio padrão σ diferente de um

$$\mu \neq 0$$

e

$$\sigma \neq 1$$

podem ter os valores das probabilidades obtidas das tabelas de valores para uma distribuição normal padronizada, padronizando a variável x , por exemplo, da seguinte forma

$$z = \frac{x - \mu}{\sigma}$$

onde a variável z tem distribuição normal com média μ igual a zero e desvio padrão σ igual a 1. Ou seja possui um distribuição normal padronizada para a qual pode-se usar os valores calculados e disponibilizados em tabelas que se encontram no apêndice da maioria dos livros de estatística e econometria. Na tabela com os valores da distribuição normal padronizada, usualmente denota-se a variável normal padronizada como z . Outro detalhe da tabela da distribuição normal padronizada é a disponibilidade dos valores das probabilidades para

a metade da distribuição, ou seja, para os valores positivos de z pelo fato da distribuição normal ser simétrica. Portanto,

$$P(0 < z < 1,23) = P(-1,23 < z < 0).$$

Dessa forma, o maior valor de probabilidade encontrada na tabela para uma distribuição normal padronizada usualmente é aproximadamente igual a 0,5. Claro, existem tabelas que podem apresentar os valores para toda a distribuição e nestes casos o maior valor será aproximadamente igual a 1, pois os valores são em termos de função distribuição de probabilidade $F(z)$. Usando o R, a probabilidade de z estar entre zero e 1,23 é calculada através da função `pnorm` que entrega o valor de $F(z)$.

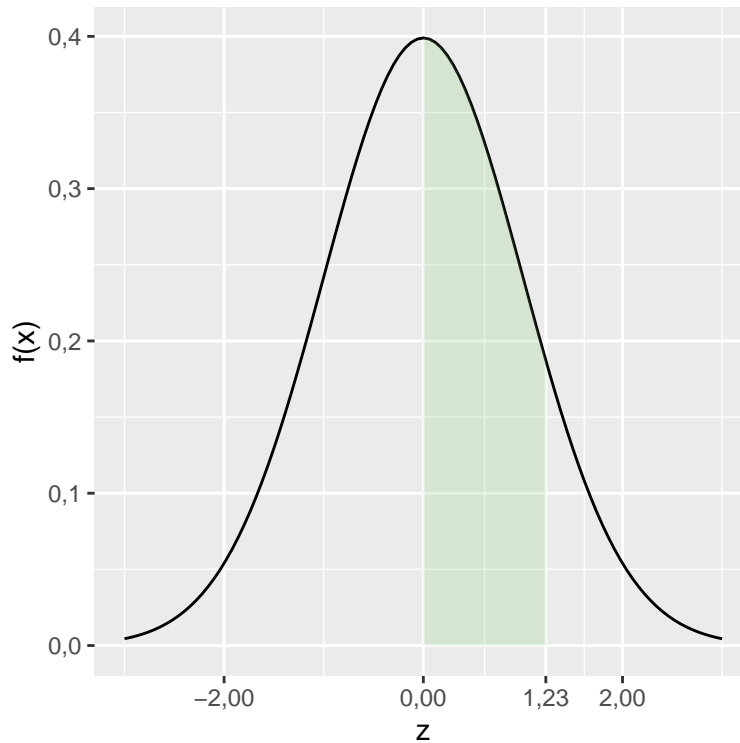
```
p0z1.23 <- round(pnorm(1.23, mean = 0, sd = 1) - pnorm(0,
                    mean = 0, sd = 1), 4)
p0z1.23
```

```
## [1] 0,3907
```

Ou seja

$$P(0 < z < 1,23) = F(1,23) - F(0) \cong 0,3907.$$

Graficamente, usando o pacote `ggplot2`, fica mais fácil de visualizar a probabilidade de z estar entre zero e 1,23 que de é de 0,3907 na figura ?? que corresponde à área colorida.



Note que o valor 0,3907 é o valor encontrado diretamente na tabela que apresenta valores de probabilidade somente para metade da distribuição, neste caso para $0 < z < +\infty$.

Quando os valores de z são maiores que zero, se tem uma outra situação que comumente aparece nos exercícios. Seja a probabilidade de z estar entre 0,27 e 1,43. Ou seja,

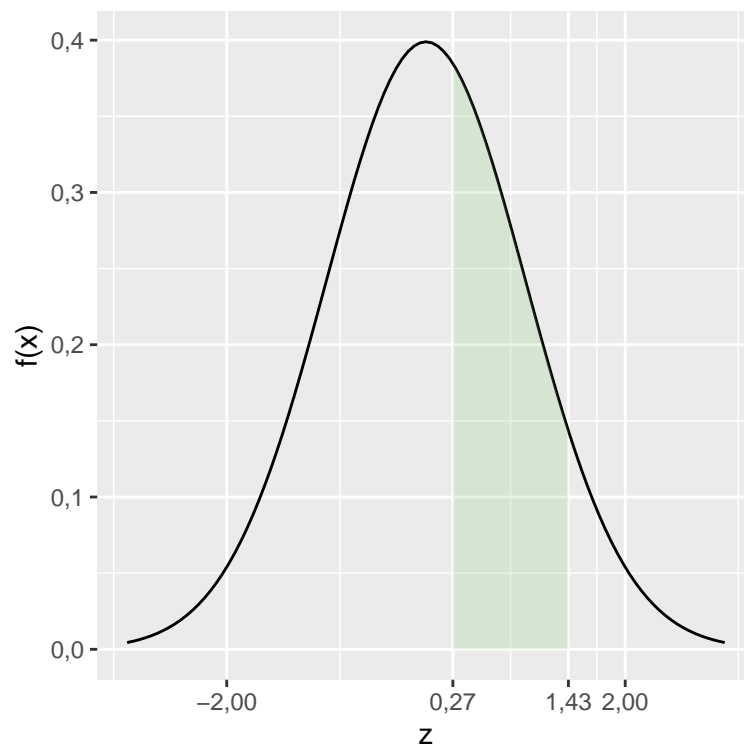
$$P(0,27 < z < 1,43)$$

Note que a probabilidade corresponde à área que está abaixo da curva da função densidade de probabilidade para z entre 0,27 e 1,43. Ou seja, a integral de $f(z)$ entre 0,27 e 1,43. Usando-se a função distribuição de probabilidade $F(z)$ corresponde a diferença entre $F(1,43)$ e $F(0,27)$

$$P(0,27 < z < 1,43) = F(1,43) - F(0,27)$$

Na figura 5.4 fica claro o porquê da diferença entre $F(z)$. Note que aqui a ideia é válida tanto para a metade da distribuição como para a distribuição completa incluindo a parte em que $z < 0$.

Para a situação em que se tem os valores das probabilidades tabulados somente para metade da distribuição se tem

Figura 5.4: Probabilidade z entre 0,27 e 1,43

$$P(0,27 < z < 1,43) = P(0 < z < 1,43) - P(0 < z < 0,27),$$

ou seja,

$$P(0,27 < z < 1,43) = [F(1,43) - F(0)] - [F(0,27) - F(0)].$$

Usando o R, a probabilidade de z estar entre 0,27 e 1,43 pode ser calculada da seguinte forma

```
# Calculo de F(1,43) - F(0)

p2 <- round(pnorm(1.43, mean = 0, sd = 1, lower.tail = T) -
  pnorm(0, mean = 0, sd = 1, lower.tail = T), 4)

p2

## [1] 0,4236

# Calculo de F(0,27) - F(0)

p1 <- round(pnorm(0.27, mean = 0, sd = 1, lower.tail = T) -
  pnorm(0, mean = 0, sd = 1, lower.tail = T), 4)

p1

## [1] 0,1064

# Calculo de P(0,27 < z < 1,43)

p027z143 <- p2 - p1

p027z143

## [1] 0,3172
```

Portanto,

$$P(0,27 < z < 1,43) = 0,4236 - 0,1064 = 0,3172.$$

Alternativamente, se tem os valores das probabilidades para a distribuição completa na tabela ou se está usando o R para os cálculos

$$P(0,27 < z < 1,43) = F(1,43) - F(0,27)$$

Usando o R, o cálculo da probabilidade de z estar entre 0,27 e 1,43 é simplesmente

```

p027z143r <- round(pnorm(1.43, mean = 0, sd = 1, lower.tail = TRUE) -
  pnorm(0.27, mean = 0, sd = 1, lower.tail = TRUE),
  4)

# Calculo de F(1,43)

p2r <- round(pnorm(1.43, mean = 0, sd = 1, lower.tail = T),
  4)

p2r

## [1] 0,9236
# Calculo de F(0,27)

p1r <- round(pnorm(0.27, mean = 0, sd = 1, lower.tail = T),
  4)

p1r

## [1] 0,6064
# Calculo de P(0,27 < z < 1,43)

p027z143r <- p2r - p1r

p027z143r

## [1] 0,3172

```

Portanto,

$$P(0,27 < z < 1,43) = 0,9236 - 0,6064 = 0,3172.$$

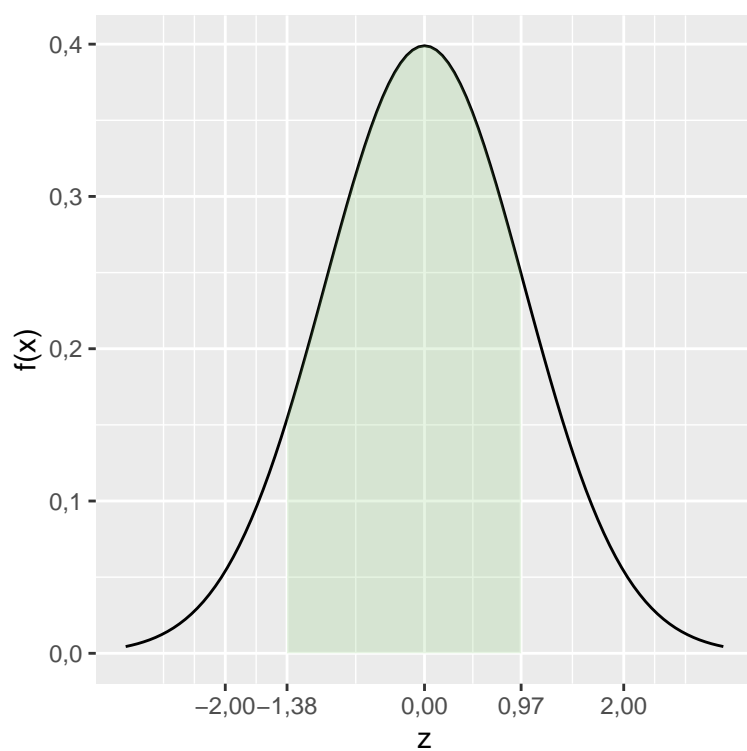
Uma outra situação bastante comum nos exercícios é quando o limite inferior do intervalo é negativo e valor do limite superior é positivo. Considere o cálculo da probabilidade de z estar entre -1,38 e 0,97. Na figura 5.5 fica mais fácil visualizar qual a área a ser calculada que corresponde a probabilidade solicitada.

Considerando a situação em que a tabela tem apenas metade dos valores de probabilidade para a distribuição normal padronizada, o cálculo é

$$P(-1,38 < z < 0,97) = P(-1,38 < z < 0) + P(0 < z < 0,97).$$

Como a distribuição normal é simétrica, pode-se escrever

$$P(-1,38 < z < 0,97) = P(0 < z < 1,38) + P(0 < z < 0,97)$$

Figura 5.5: Probabilidade z entre $-1,38$ e $0,97$

ou

$$P(-1,38 < z < 0,97) = [F(1,38) - F(0)] + [F(0,97) - F(0)]$$

Usando o R, a probabilidade de z estar entre 0,27 e 1,43 pode ser calculada da seguinte forma

```
# Calculo de F(1,38) - F(0)

pf138f0 <- round(pnorm(1.38, mean = 0, sd = 1, lower.tail = T) -
  pnorm(0, mean = 0, sd = 1, lower.tail = T), 4)

pf138f0

## [1] 0,4162

# Calculo de F(0,97) - F(0)

pf097f0 <- round(pnorm(0.97, mean = 0, sd = 1, lower.tail = T) -
  pnorm(0, mean = 0, sd = 1, lower.tail = T), 4)

pf097f0

## [1] 0,334

# Calculo de P(1,38 < z < 0,97)

p138z097 <- pf138f0 + pf097f0

p138z097

## [1] 0,7502

Portanto,
```

$$P(0,27 < z < 1,43) = 0,4162 + 0,334 = 0,7502.$$

Alternativamente, se tem os valores das probabilidades para a distribuição completa na tabela ou se está usando o R para os cálculos, calcula-se

$$P(-0,38 < z < 0,97) = F(0,97) - F(-1,38).$$

Por que o valor de $F(0,97)$ é maior que $F(-1,38)$ e nesta situação se calcula a diferença e não a soma?

Usando o R, o cálculo da probabilidade de z estar entre -1,38 e 0,97 é simplesmente

```
p138z097r <- round(pnorm(-1.38, mean = 0, sd = 1, lower.tail = TRUE) -
  pnorm(0.97, mean = 0, sd = 1, lower.tail = TRUE),
```



```

4)

# Calculo de F(-1,38)

pf138r <- round(pnorm(-1.38, mean = 0, sd = 1, lower.tail = T),
4)

pf138r

## [1] 0,0838

# Calculo de F(0,97)

pf097r <- round(pnorm(0.97, mean = 0, sd = 1, lower.tail = T),
4)

pf097r

## [1] 0,834

# Calculo de P(-1,38 < z < 0,97)

p138z097r <- pf097r - pf138r

p138z097r

## [1] 0,7502

```

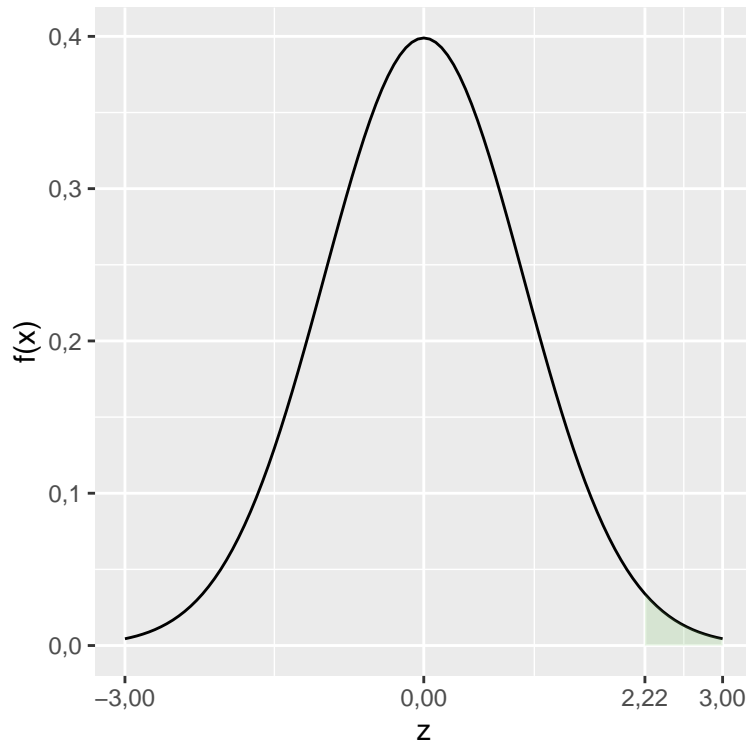
Portanto,

$$P(0,27 < z < 1,43) = 0,834 - 0,0838 = 0,7502.$$

Uma outra situação bastante interessante e comum nos exercícios sobre a aplicação de distribuição normal padronizada é quando se pede o cálculo da probabilidade de z ser maior que um determinado valor. Nestas situações, a palavra chave é **pelo menos**. Considere a situação na qual se pede o cálculo da probabilidade de que z seja maior que 2,22. Ou seja,

$$P(z > 2,22)$$

Na figura ?? fica claro onde fica a área correspondente a probabilidade solicitada.



Dessa forma, o cálculo considerando a disponibilidade da metade dos valores da distribuição normal padronizada é

$$P(z > 2,22) = 0,5 - P(0 < z < 2,22) = 0,5 - [F(2,22) - F(0)]$$

Usando o R, o cálculo é

```
f222 <- round(pnorm(2.22, mean = 0, sd = 1, lower.tail = T),
4)
f222
```

```
## [1] 0,9868
```

```
f0 <- round(pnorm(0, mean = 0, sd = 1, lower.tail = T),
4)
f0
```

```
## [1] 0,5
```

```
pf222f0 <- f222 - f0
```

```
pf222f0
```

```
## [1] 0,4868
```

```
pz222 <- 0.5 - pf222f0
```

```
pz222
```

```
## [1] 0,0132
```

Portanto,

$$P(z > 2, 22) = 0,5 - P(0 < z < 2, 22) = 0,5 - 0,4868 = 0,0132.$$

Alternativamente se tem os valores da probabilidade para a distribuição completa ou se está usando o R para os cálculos

$$P(z > 2, 22) = 1 - F(2, 22)$$

Usando o R para o cálculo,

```
pz222r <- 1 - round(pnorm(2.22, mean = 0, sd = 1, lower.tail = T),
4)
```

```
pz222r
```

```
## [1] 0,0132
```

Como nem todas as variáveis com distribuição normal tem $\mu = 0$ e $\sigma = 1$, deve-se padronizá-las para usar os valores da tabela com a distribuição normal padronizada. Seja a variável x com distribuição normal com $\mu \neq 0$ e $\sigma \neq 1$. A transformação de x para que se torne padronizada é

$$z = \frac{x - \mu}{\sigma}$$

onde z , com $\mu = 0$ e $\sigma = 1$, é a variável x transformada.

Prova:

Como σ e μ são constantes

$$E(z) = \sum_{i=1}^n \left(\frac{x - \mu}{\sigma} \right) = \frac{1}{\sigma} \sum_{i=1}^n (x - \mu) = \frac{1}{\sigma} \times 0 = 0$$

Pois a soma dos desvios em relação à média μ é sempre zero.

Aplicando a definição de variância em z sabendo-se que $E(z) = 0$ e $E[x - \mu]^2 = \sigma^2$ ou $dp(x) = \sigma$

$$Var(z) = E[z - E(z)]^2 = E \left[\frac{x - \mu}{\sigma} \right]^2 = \frac{1}{\sigma^2} E[x - \mu]^2 = \frac{1}{\sigma^2} \times \sigma^2 = 1.$$

Assim, através de (5.2), pode-se padronizar qualquer variável que segue uma distribuição normal e que tenha $\mu \neq 0$ e $\sigma \neq 1$.

Exemplo numérico sobre a padronização

Exemplo 4.4.1 da página 93 do Sartoris (2013). O faturamento mensal de uma loja segue uma distribuição normal com média R\$20.000,00 e desvio padrão R\$4.000,00. Calcule a probabilidade de que, em determinado mês, o faturamento esteja entre R\$19.000,00 e R\$25.000,00.

Resposta:

A variável em questão, faturamento mensal, é normal mas não é padronizada. Por isso, é necessário padronizar os valores antes de utilizar a tabela, cujos valores são para uma distribuição normal padronizada com $\mu = 0$ e $\sigma = 1$. Para 19.000

$$z_1 = \frac{x_1 - \mu}{\sigma} = \frac{19.000 - 20.000}{4.000} = -0,25$$

e para 25.000

$$z_2 = \frac{x_2 - \mu}{\sigma} = \frac{25.000 - 20.000}{4.000} = 1,25.$$

Portanto

$$P(19.000 < x < 25.000) = P(-0,25 < z < 1,25)$$

Como a distribuição normal é simétrica

$$P(19.000 < x < 25.000) = P(-0,25 < z < 0) + P(0 < z < 1,25)$$

$$P(19.000 < x < 25.000) = P(0 < z < 0,25) + P(0 < z < 1,25)$$

Se a tabela apresenta valores para a distribuição completa

$$P(19.000 < x < 25.000) = [F(0,25) - F(0)] + [F(1,25) - F(0)]$$

ou se a tabela só apresenta valores para a metade da distribuição

$$P(19.000 < x < 25.000) = F(0,25) + F(1,25)].$$

Mas para qualquer um dos dois caminhos

$$P(19.000 < x < 25.000) = F(0,25) + F(1,25)$$

Usando o R para os cálculos intermediários, se tem

```
# Calculo de F(1,25) - F(0)

pf125f0 <- round(pnorm(1.25, mean = 0, sd = 1, lower.tail = T) -
  pnorm(0, mean = 0, sd = 1, lower.tail = T), 4)

pf125f0
```

```
## [1] 0,3944
# Calculo de F(0,25) - F(0)

pf025f0 <- round(pnorm(0.25, mean = 0, sd = 1, lower.tail = T) -
  pnorm(0, mean = 0, sd = 1, lower.tail = T), 4)

pf025f0

## [1] 0,0987
# Calculo de P(-0,25 < z < 1,25)

p025z125 <- pf025f0 + pf125f0

p025z125

## [1] 0,4931
Ou seja,
```

$$P(19.000 < x < 25.000) = F(0, 25) + F(1, 25) = 0,0987 + 0,3944$$

$$P(19.000 < x < 25.000) = 0,4931.$$

5.3 Teorema de Tchebichev

Se a função densidade de uma variável aleatória é conhecida, pode-se conhecer sua média e variância. Por outro lado, se são conhecidas somente a média e a variância de uma variável aleatória, não é possível conhecer a sua respectiva função densidade de probabilidade. Mas é possível definir um limite para uma distribuição de probabilidade qualquer, seja discreta ou contínua, que é dado pelo teorema de Tchebichev.

Teorema de Tchebichev

seja uma variável aleatória x com média μ e desvio padrão σ . A probabilidade de que a variável x esteja acima ou abaixo da média que não supere, em módulo, k desvio padrão, é maior ou pelo menos igual a $1 - 1/k^2$. Ou seja,

$$P(|x - \mu| < k\sigma) \geq 1 - \frac{1}{k^2}.$$

Consequentemente, a probabilidade de que a variável x esteja acima ou abaixo da média supere ou seja igual, em módulo, a k desvio padrão é menor ou pelo menos igual a $1/k^2$. Ou seja,

$$P(|x - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

Essas duas relações dão um limite universal ao desvio $|x - \mu|$ em termos de σ .

Assumindo, por exemplo, que $k = 2$, se tem

$$P(|x - \mu| < 2\sigma) \geq 1 - \frac{1}{2^2}$$

$$P(|x - \mu| < 2\sigma) \geq 0,75 = 75\%$$

ou

$$P(|x - \mu| \geq 2\sigma) \leq \frac{1}{2^2}$$

$$P(|x - \mu| \geq 2\sigma) \leq 0,25 = 25\%$$

Exemplo numérico sobre teorema de Tchebichev

Exemplo 4.6.1 na página 96 do Sartoris (2013). Seja uma variável aleatória contínua x com média 50 e desvio padrão 10. Calcule a probabilidade mínima de que x esteja entre 35 e 65.

Resposta:

Note que a diferença em módulo tanto do valor do limite inferior 35

$$|x - \mu| = |35 - 50| = 15$$

como a diferença em módulo do valor do limite superior 65

$$|x - \mu| = |65 - 50| = 15$$

são iguais a 15. Dividindo a diferença em módulo pelo desvio padrão,

$$\frac{|x - \mu|}{\sigma} = \frac{15}{10} = 1,5$$

obtem-se o valor de k do teorema de Tchebichev que neste caso é 1,5. Portanto, usando a relação do Teorema de Tchebichev

$$P(35 < x < 65) = P(|x - \mu| < 1,5\sigma) \geq 1 - \frac{1}{1,5^2}$$

$$P(35 < x < 65) = P(|x - \mu| < 1,5\sigma) \geq 0,6656 = 55,56\%$$

Ou seja, de forma aproximada, a probabilidade de x estar entre 35 e 65 é pelo menos igual a 55,56%.

5.4 Momentos de uma distribuição

O **momento de uma distribuição** de uma variável aleatória x de ordem k , em relação à média M_k é

$$M_k = E(x - \mu)^k$$

Como já foi visto anteriormente que a somatória dos desvios é sempre igual a zero, o primeiro momento em relação à média também é sempre igual a zero. Ou seja,

$$M_1 = E(x - \mu)^1 = E(x) - \mu = \mu - \mu = 0$$

Já segundo momento em relação à média é a variância

$$M_2 = E(x - \mu)^2 = \sigma^2$$

O terceiro momento em relação à média

$$M_3 = E(x - \mu)^3$$

entrega o grau de simetria da distribuição. Por exemplo, uma distribuição simétrica como a distribuição normal tem

$$M_3 = 0.$$

Com M_3 é possível definir um coeficiente de assimetria

$$\alpha_3 = \frac{M_3}{\sigma^3}$$

que é tão maior em módulo quanto mais assimétrica for a distribuição analisada.

O quarto momento em relação à média

$$M_4 = E(x - \mu)^4$$

entrega o **grau de achatamento** de uma distribuição e por isso está relacionado com a **curtose**.

- Se uma distribuição é muito achatada, ela é dita **platicúrtica**.
- Se uma distribuição é mais pontiaguda, ela é dita **leptocúrtica**
- Se a distribuição se assemelha a distribuição normal, ela é dita **mesocúrtica**.

Com M_4 define-se o coeficiente de curtose

$$\alpha_4 = \frac{M_4}{\sigma^4}$$

que define

- Se $\alpha_4 = 3$, a distribuição é normal, portanto mesocúrtica.
- Se $\alpha_4 > 3$, a distribuição é leptocúrtica.
- Se $\alpha_4 < 3$, a distribuição é platicúrtica.

Os momentos podem ser definidos em relação à origem. O momento de uma distribuição em relação à origem de ordem k em torno da origem é dado por

$$M'_k = E(x)^k$$

Assim

- para $k = 1$, o momento em torno da origem de ordem 1 é a própria média.
- para $k = 2$, o momento em torno da origem de ordem 2 é a média dos quadrados.

5.4.1 Posição da média, da mediana e da moda numa distribuição assimétrica

Dado que foi apresentado o conceito de assimetria, é possível realizar uma rápida discussão sobre o impacto da assimetria sobre a posição da média, da mediana e da moda numa distribuição.

- Se a distribuição é simétrica, ou seja, com a cauda esquerda fosse a imagem da cauda direita em um espelho e vice-versa,

$$\bar{X} = \text{Mediana} = \text{Moda}.$$

- Se a distribuição é assimétrica à direita, ou seja, cauda maior no lado direito,

$$\text{Moda} < \text{Mediana} < \bar{X}.$$

- Se a distribuição é assimétrica à esquerda, ou seja, cauda maior no lado esquerdo,

$$\bar{X} < \text{Mediana} < \text{Moda}.$$

Capítulo 6

Distribuição de variáveis aleatórias conjunta

Probabilidade conjunta é a probabilidade que se refere a duas ou mais variáveis aleatórias simultaneamente.

A distribuição de probabilidade de um vetor (X, Y) com duas variáveis, por exemplo, seria o caso bidimensional. Como o material sobre este assunto é baseado em Sartoris (2013), a apresentação da distribuição de variáveis aleatórias conjuntas será o caso bidimensional.

As variáveis da distribuição de probabilidade conjunta podem ser discretas ou contínuas.

6.1 Distribuição de probabilidade conjunta de variáveis aleatórias discretas

Com base em Sartoris (2013), é apresentado o assunto da seção através da apresentação de um exemplo numérico prático.

Seja um time de volei que vai disputar um campeonato muito equilibrado, em que a probabilidade de ganhar ou perder uma partida é de 0,5. O técnico pede ao estatístico da equipe que faça uma análise das probabilidades das três primeiras partidas consideradas vitais para o restante do campeonato. Em particular, a vitória na primeira partida é considerada decisiva pela comissão técnica.

O estatístico define duas variáveis, X e Y , sendo que

- X é o número de vitórias obtidas nos três primeiros jogos e;
- Y é igual a 1, caso ocorra vitória no primeiro jogo e zero, caso ocorra o contrário.

Por enquanto considera-se que X e Y são variáveis independentes.

Como são três jogos com dois resultados possíveis, vitória ou derrota com 0,5 de probabilidade cada um, existem oito possibilidades entre os três primeiros jogos em análise. Na tabela 6.1 abaixo é apresentado os valores de X e Y de acordo com os resultados possíveis.

Tabela 6.1: Resultados prováveis em três partidas, considerando vitória (V) ou derrota(D) como resultado possível de cada partida com 0,5 de probabilidade

| resultados possíveis | X | Y |
|----------------------|---|---|
| VVV | 3 | 1 |
| VVD | 2 | 1 |
| VDV | 2 | 1 |
| VDD | 1 | 1 |
| DVV | 2 | 0 |
| DDV | 1 | 0 |
| DVD | 1 | 0 |
| DDD | 0 | 0 |

Fonte: Sartoris (2013).

Onde o resultado possível nos três primeiros jogos definido como VDV significa que o time teve vitória na primeira e na terceira partidas e derrota na segunda partida. Esse mesmo resultado VDV define que $X = 2$ e $Y = 1$ lembrando que X é o número de vitórias entre as três partidas e Y é igual a 1 se o time conseguiu vitória na primeira partida e zero se não conseguiu ganhar a primeira partida.

Na sequência o estatístico constrói uma tabela que apresenta as probabilidades conjuntas de x e Y cujo preenchimento é feito com base na tabela 6.2

Tabela 6.2: Probabilidades conjuntas de X e Y

| | X=0 | X=1 | X=2 | X=3 |
|------------|-----|-----|-----|-----|
| Y=0 | 1/8 | 2/8 | 1/8 | 0 |
| Y=1 | 0 | 1/8 | 2/8 | 1/8 |

Com base na tabela 6.2 é possível obter a probabilidade para um resultado com duas vitórias sendo que uma das duas vitórias foi na primeira partida:

$$P(X = 2 \text{ e } Y = 1) = 2/8$$

Ou seja em oito resultados possíveis há duas combinações possíveis com resultado.

Veja que se o time ganha as três partidas $X = 3$, não é possível que $Y = 0$ e por isso

$$P(X = 3 \text{ e } Y = 0) = 0$$

O inverso também é válido. Ou seja, se o time não ganhou nenhuma das três partidas, não é possível que $Y = 1$ e por isso

$$P(X = 0 \text{ e } Y = 1) = 0$$

Ainda com base na tabela 6.2 é possível obter probabilidades só para valores de X e probabilidades só para valores de Y . Ou seja, para obter a probabilidade de $X = 1$ é necessário somar todas as probabilidades conjuntas que tem $X = 1$

$$P(X = 1) = P(X = 1 \text{ e } Y = 0) + P(X = 1 \text{ e } Y = 1) = 2/8 + 1/8 = 3/8.$$

Na tabela 6.2, $P(X = 1)$ é obtida somando os valores da coluna para $X = 1$. Também é possível obter o valor da probabilidade de $Y = 1$ que é simplesmente a soma de todas as probabilidades conjuntas que tem $Y = 1$

$$P(Y = 1) = P(X = 0 \text{ e } Y = 1) + P(X = 1 \text{ e } Y = 1) + P(X = 2 \text{ e } Y = 1) + P(X = 3 \text{ e } Y = 1)$$

ou seja,

$$P(Y = 1) = 0 + 1/8 + 2/8 + 1/8 = 1/2$$

Na tabela 6.2, $P(Y = 1)$ é obtida somando os valores da linha para $Y = 1$. Assim, usando este raciocínio, é possível agregar mais uma linha e mais uma coluna com as probabilidades marginais de X e de Y . As probabilidades marginais de X são as somas de cada uma das colunas da tabela 6.2. As probabilidades marginais de Y são as somas de cada uma das linhas da tabela 6.2. Desta forma se obtém a tabela 6.3.

Tabela 6.3: Probabilidades marginais de X e Y

| | X=0 | X=1 | X=2 | X=3 | P(Y) |
|-------------|------------|------------|------------|------------|------------|
| Y=0 | 1/8 | 2/8 | 1/8 | 0 | 1/2 |
| Y=1 | 0 | 1/8 | 2/8 | 1/8 | 1/2 |
| P(X) | 1/8 | 3/8 | 3/8 | 1/8 | 1 |

Com base na tabela 6.3 é possível calcular a probabilidade condicional, embora não possa ser obtida diretamente da fonte.

Suponha a seguinte pergunta com base neste exemplo numérico prático: qual é a probabilidade de o time ganhar apenas uma partida entre as três dado que essa partida ocorre na primeira partida? Em notação matemática seria

$$P(X = 1|Y = 1) = ??$$

Lembrando das aulas de teoria da probabilidade, a probabilidade condicional é dada da seguinte forma:

$$P(X = x|Y = y) = \frac{P(X = x \cap Y = y)}{P(Y = y)}$$

onde $P(X = x \cap Y = y)$ é a probabilidade conjunta entre $X = x$ e $Y = y$ e $P(Y = y)$ é a probabilidade marginal de $Y = y$.

Assim, podemos calcular a probabilidade condicional $P(X = 1|Y = 1)$

$$P(X = 1|Y = 1) = \frac{P(X = 1 \cap Y = 1)}{P(Y = 1)} = \frac{1/8}{1/2} = 1/4$$

Dado que $Y = 1$, só existe quatro possibilidades para essa situação, dos quais apenas uma situação tem $X = 1$.

Note que o contrário também é possível obter. Ou seja considere a probabilidade condicional da seguinte forma:

$$P(Y = y|X = x) = \frac{P(Y = y \cap X = x)}{P(X = x)}.$$

Com base no exemplo do time de vôlei, qual é probabilidade condicional de que o time não vença na primeira partida dado que o time ganhe duas entre as três partidas?

$$P(Y = 0|X = 2) = \frac{P(Y = 0 \cap X = 2)}{P(X = 2)}.$$

Consultando a tabela 6.3 se tem

$$P(Y = 0|X = 2) = \frac{P(Y = 0 \cap X = 2)}{P(X = 2)} = \frac{1/8}{3/8} = 1/3$$

Agora sabendo calcular a probabilidade condicional é possível verificar se as variáveis são independentes ou não. Para isso toma-se as situações de probabilidade conjunta igual a zero na tabela 6.3. Note que se o time ganha as três partidas, não há possibilidade da variável Y assumir o valor 0. Veja o outro valor zero. Quando o time não ganha nenhuma das três primeiras partidas não é possível a variável Y assumir valor igual a 1. A variável Y sendo igual a 1 necessariamente o time tem que ter ganho a primeira partida. Por isso as variáveis X e Y do exemplo do time de vôlei não são independentes.

Então uma forma de verificar se as variáveis são independentes ou não é verificar se a probabilidade condicional é igual a probabilidade marginal. Ou seja,

$$P(X = x|Y = y) = P(X = x)$$

ou

$$P(Y = y|X = x) = P(Y = y)$$

Caso as probabilidades condicionais não sejam iguais a sua respectiva probabilidade marginal, então as duas variáveis X e Y não são independentes. Com base na tabela 6.3, toma-se, por exemplo, a probabilidade conjunta para $X = 1$ e $Y = 1$

$$P(X = 1|Y = 1) = 1/4$$

e a probabilidade marginal $X = 1$

$$p(X = 1) = 3/8$$

são diferentes. Portanto X e Y não são independentes. Basta somente um par de valores de X e Y apresentar essa diferença para poder concluir que as variáveis são dependentes uma da outra. Para verificar se as variáveis são independentes, é necessário verificar se todos os pares possíveis tem probabilidade condicional igual a probabilidade incondicional. Note que a probabilidade marginal é denominada também de probabilidade incondicional.

Ou seja, basta somente uma situação com

$$P(X = x|Y = y) \neq P(X = x)$$

para que X e Y seja dependentes.

6.1.1 Covariância e correlação no contexto da distribuição de probabilidade conjunta

O cálculo do valor esperado bem como da variância de variáveis distribuídas conjuntamente não é problema pois é possível obter as probabilidades marginais facilmente a partir das probabilidades conjuntas. O que há de novo, mas nem tanto, é sobre o cálculo da covariância e da correlação, que é apresentado na forma de exemplo numérico a seguir.

Exemplo numérico sobre covariância e correlação para a distribuição conjunta

Exemplo 5.1.1 da página 112 do Sartoris (2013). Calcule com base nos dados do exemplo numérico sobre o time de vôlei:

- o valor esperado de X e Y ;
- as variâncias de X e Y ;
- a covariância entre X e Y ;

- a correlação entre X e Y .

Para calcular $E(X)$ utiliza-se as probabilidades marginais de X que estão na tabela 6.3

$$E(X) = \sum_{i=1}^n P(X_i)X_i$$

$$E(X) = P(X_1) \times X_1 + P(X_2) \times X_2 + P(X_3) \times X_3 + P(X_4) \times X_4$$

Portanto,

$$E(X) = 1/8 \times 0 + 3/8 \times 1 + 3/8 \times 2 + 1/8 \times 3 = 12/8 = 1,5$$

A esperança matemática de x do exemplo numérico é 1,5.

Para Y

$$E(Y) = \frac{1}{2} \times 0 + \frac{1}{2} \times 1 = 0,5$$

Para calcular a variância de X usando a fórmula alternativa

$$Var(X) = E(X^2) - [E(X)]^2$$

Assim só está faltando calcular a esperança do quadrado de X

$$E(X^2) = \sum_{i=1}^n P(X_i) \times X_i^2 = P(X_1) \times X_1^2 + \dots + P(X_n) \times X_n^2$$

Portanto,

$$E(X^2) = \frac{1}{8} \times 0^2 + \frac{3}{8} \times 1^2 + \frac{3}{8} \times 2^2 + \frac{1}{8} \times 3^2 = \frac{24}{8} = 3$$

Com $E(X)$ e $E(X^2)$

$$Var(X) = E(X^2) - [E(X)]^2 = 3 - (1,5)^2 = 3 - 2,25 = 0,75.$$

Para a variância de Y é necessário calcular $E(Y^2)$

$$E(Y^2) = \frac{1}{2} \times 0^2 + \frac{1}{2} \times 1^2 = 0,5$$

Assim,

$$\text{Var}(Y) = E(Y^2) - [E(Y)]^2 = 0,5 - (0,5)^2 = 0,5 - 0,25 = 0,25$$

Para calcular a covariancia entre X e Y , com base na fórmula alternativa,

$$\text{Cov}(XY) = E(XY) - E(X)E(Y)$$

é necessário calcular o produto entre X e Y que segue na tabela 6.4.

Tabela 6.4: Produto entre X e Y

| X | Y | XY |
|---|---|----|
| 3 | 1 | 3 |
| 2 | 1 | 2 |
| 2 | 1 | 2 |
| 1 | 1 | 1 |
| 2 | 0 | 0 |
| 1 | 0 | 0 |
| 1 | 0 | 0 |
| 0 | 0 | 0 |

Com base nos resultados de XY apresentados na tabela 6.4, é possível obter as seguintes probabilidades

$$P(XY = 0) = \frac{4}{8}$$

$$P(XY = 1) = \frac{1}{8}$$

$$P(XY = 2) = \frac{2}{8}$$

$$P(XY = 3) = \frac{1}{8}.$$

Assim

$$E(XY) = \sum_{i=1}^n P(X_i Y_i) \times XY$$

$$E(XY) = \frac{4}{8} \times 0 + \frac{1}{8} \times 1 + \frac{2}{8} \times 2 + \frac{1}{8} \times 3 = \frac{8}{8} = 1$$

como já se tem calculado $E(X)$ e $E(Y)$

$$Cov(XY) = E(XY) - E(X)E(Y) = 1 - (1,5)(0,5) = 1 - 0,75 = 0,25$$

Para o cálculo da correlação se tem todas as partes da sua fórmula calculadas

$$corr(XY) = \rho_{XY} = \frac{Cov(X, Y)}{\sqrt{Var(X) \times Var(Y)}}$$

Portanto

$$\rho_{XY} = \frac{0,25}{\sqrt{0,75 \times 0,25}} \cong 0,5774$$

6.1.2 Esperança condicionada

A esperança condicionada é similar a esperança marginal ou incondicional sendo que as probabilidades associadas a variável em questão é a probabilidade condicionada que precisa ser previamente calculada, dado que é necessário ter a probabilidade conjunta e a probabilidade marginal ou incondicional para o seu cálculo. Ou seja se é probabilidade de X dado Y

$$P(X = x|Y = y) = \frac{P(X = x \cap Y = y)}{P(Y = y)}$$

para poder calcular a esperança condicionada de X dado Y

$$E(X_i|Y = c) = \sum_{i=1}^n P(X_i|Y = c) \times X_i.$$

Exemplo numérico sobre Esperança condicional

Exemplo 5.1.2 da página 113 do Sartoris (2013). Seja as variáveis aleatórias X e Y definidas no exemplo sobre o time de volei, determine $E(X|Y = 0)$.

Resposta:

Para o cálculo da esperança condicionada são necessárias as probabilidades condicionais para todos os valores de X . Pois

$$E(X|Y = 0) = \sum_{i=1}^n P(X_i|Y = 0) \times X_i$$

As probabilidades condicionais são

$$P((X = 0|Y = 0) = \frac{P(X = 0 \text{ e } Y = 0)}{P(Y = 0)} = \frac{\frac{1}{8}}{\frac{1}{2}} = \frac{1}{4}$$

$$P((X = 1|Y = 0) = \frac{P(X = 1 \text{ e } Y = 0)}{P(Y = 0)} = \frac{\frac{2}{8}}{\frac{1}{2}} = \frac{1}{2}$$

$$P((X = 2|Y = 0) = \frac{P(X = 2 \text{ e } Y = 0)}{P(Y = 0)} = \frac{\frac{1}{8}}{\frac{1}{2}} = \frac{1}{4}$$

$$P((X = 3|Y = 0) = \frac{P(X = 3 \text{ e } Y = 0)}{P(Y = 0)} = \frac{0}{\frac{1}{2}} = 0$$

Com essas probabilidades condicionais calcula-se a esperança de X condicionada a $Y = 0$

$$E(X|Y = 0) = \frac{1}{4} \times 0 + \frac{1}{2} \times 1 + \frac{1}{4} \times 2 + 0 \times 3 = 1$$

Portanto a $E(X|Y = 0) = 1$.

6.1.3 Lei das Expectativas Iteradas

A lei das expectativas iteradas diz que o valor esperados das esperanças condicionais é igual a esperança incondicional. Ou seja,

$$E[E(X|Y)] = E(X)$$

Exemplo sobre a Lei das Expectativas Iteradas

O Exemplo 5.1.3 da página 114 do Sartoris (2013) tem o objetivo de aplicar a lei das expectativas iteradas através de (6.1.3). Seja as variáveis X e Y do exemplo sobre o time de volei, determine $E[E(X|Y)]$.

Resposta:

Note que o cálculo da esperança das esperanças condicionais é necessário os todos as esperanças condicionais de X e as respectivas probabilidades de Y . Como já foi calculado $E(X|Y = 0)$ no exemplo numérico anterior, fica faltando $E(X|Y = 1)$, uma vez que os valores de Y são zero e um. Ou seja,

Para o cálculo da esperança condicionada de $\#X$ dado que $Y = 1$ são necessárias as probabilidades condicionais para todos os valores de X . Pois

$$E(X|Y = 1) = \sum_{i=1}^n P(X_i|Y = 1) \times X_i$$

sendo que

$$P(X|Y = 1) = \frac{P(X \text{ e } Y = 0)}{P(Y = 0)}.$$

As probabilidades condicionais são

$$P((X = 0|Y = 1) = \frac{P(X = 0 \text{ e } Y = 1)}{P(Y = 1)} = \frac{0}{\frac{1}{2}} = 0$$

$$P((X = 1|Y = 1) = \frac{P(X = 1 \text{ e } Y = 1)}{P(Y = 1)} = \frac{\frac{1}{8}}{\frac{1}{2}} = \frac{1}{4}$$

$$P((X = 2|Y = 1) = \frac{P(X = 2 \text{ e } Y = 1)}{P(Y = 1)} = \frac{\frac{2}{8}}{\frac{1}{2}} = \frac{1}{2}$$

$$P((X = 3|Y = 1) = \frac{P(X = 3 \text{ e } Y = 1)}{P(Y = 1)} = \frac{\frac{1}{8}}{\frac{1}{2}} = \frac{1}{4}$$

Com essas probabilidades condicionais calcula-se a esperança de X condicionada a $Y = 1$

$$E(X|Y = 1) = 0 + \frac{1}{4} \times 1 + \frac{1}{2} \times 2 + \frac{1}{4} \times 3 = 2$$

Portanto a $E(X|Y = 1) = 2$.

O valor de $E[E(X|Y)]$ é calculado da seguinte forma

$$E[E(X|Y)] = \sum_{j=1}^k P(Y_j) \times E(X|Y_j)$$

onde $j = 1, \dots, k$. Assim sendo, para o exemplo numérico fica

$$E[E(X|Y)] = P(Y = 0) \times E(X|Y = 0) + P(Y = 1) \times E(X|Y = 1)$$

colocando os valores calculados

$$E[E(X|Y)] = \frac{1}{2} \times 1 + \frac{1}{2} \times 2 = 1,5$$

Note que $E(X)$ calculando anteriormente é, de fato, 1,5. Portanto confirma-se através do exemplo numérico que

$$E[E(X|Y)] = E(X).$$

6.1.4 Variância condicionada

A variância condicional como o próprio nome diz é a variância de uma variável condicionada ao valor de uma outra variável. Por isso o seu cálculo necessita a probabilidade condicional e a esperança condicionada. Neste caso, além da esperança condicional da variável em si, é necessário calcular a esperança condicional da variável elevada ao quadrado. Ou seja,

$$P(X = x|Y = y) = \frac{P(X = x \cap Y = y)}{P(Y = y)}$$

para calcular a esperança condicionada de X dado Y

$$E(X_i|Y = c) = \sum_{i=1}^n P(X_i|Y = c) \times X_i.$$

e a esperança condicionada de X^2 dado Y

$$E(X_i^2|Y = c) = \sum_{i=1}^n P(X_i|Y = c) \times X_i^2.$$

para então calcular

$$Var(X|Y = c) = E(X^2|Y = c) - [E(X|Y = c)]^2.$$

Exemplo sobre a variância condicionada

Exemplo 5.1.4 da página 114 do Sartoris (2013). Com base nas variáveis do exemplo do time de volei, calcule $Var(Y|X = 1)$.

Resposta:

É necessário calcular $P(Y = 0|X = 1)$ e $P(Y = 1|X = 1)$

$$P(Y = 0|X = 1) = \frac{P(Y = 0 \cap X = 1)}{P(X = 1)} = \frac{\frac{2}{8}}{\frac{3}{8}} = \frac{2}{3}$$

e

$$P(Y = 1|X = 1) = \frac{P(Y = 1 \cap X = 1)}{P(X = 1)} = \frac{\frac{1}{8}}{\frac{3}{8}} = \frac{1}{3}.$$

Assim

$$E(Y|X = 1) = P(Y = 0|X = 1) \times 0 + P(Y = 1|X = 1) \times 1$$

$$E(Y|X = 1) = \frac{2}{3} \times 0 + \frac{1}{3} \times 1 = \frac{1}{3}.$$

Para Y^2

$$E(Y^2|X = 1) = P(Y = 0|X = 1) \times 0^2 + P(Y = 1|X = 1) \times 1^2$$

$$E(Y^2|X = 1) = \frac{2}{3} \times 0^2 + \frac{1}{3} \times 1^2 = \frac{1}{3}.$$

Portanto,

$$Var(Y|X = 1) = E(Y^2|X = 1) - [E(Y|X = 1)]^2$$

$$Var(Y|X = 1) = \frac{1}{3} - \left[\frac{1}{3}\right]^2$$

$$Var(Y|X = 1) = \frac{1}{3} - \frac{1}{9} = \frac{2}{9}$$

Outro exemplo numérico sobre variância condicionada

Exemplo 5.1.5 na página 114 do Sartoris (2013). Sejam as variáveis X e Y do exemplo numérico do time de volei. Calcule $Var(X|Y = 0)$ e $Var(X|Y = 1)$.

Resposta:

É necessário calcular

$$P(X = x|Y = y) = \frac{P(X = x \cap Y = y)}{P(Y = y)}$$

para poder calcular a esperança condicionada de X dado Y

$$E(X_i|Y = c) = \sum_{i=1}^n P(X_i|Y = c) \times X_i.$$

e a esperança condicionada de X^2 dado Y

$$E(X_i^2|Y = c) = \sum_{i=1}^n P(X_i|Y = c) \times X_i^2.$$

para então calcular

$$Var(X|Y = c) = E(X^2|Y = c) - [E(X|Y = c)]^2.$$

Já estão calculados

$$P((X = 0|Y = 0) = \frac{P(X = 0 \text{ e } Y = 0)}{P(Y = 0)} = \frac{\frac{1}{8}}{\frac{1}{2}} = \frac{1}{4}$$

$$P((X = 1|Y = 0) = \frac{P(X = 1 \text{ e } Y = 0)}{P(Y = 0)} = \frac{\frac{2}{8}}{\frac{1}{2}} = \frac{1}{2}$$

$$P((X = 2|Y = 0) = \frac{P(X = 2 \text{ e } Y = 0)}{P(Y = 0)} = \frac{\frac{1}{8}}{\frac{1}{2}} = \frac{1}{4}$$

$$P((X = 3|Y = 0) = \frac{P(X = 3 \text{ e } Y = 0)}{P(Y = 0)} = \frac{0}{\frac{1}{2}} = 0$$

e

$$P((X = 0|Y = 1) = \frac{P(X = 0 \text{ e } Y = 1)}{P(Y = 1)} = \frac{0}{\frac{1}{2}} = 0$$

$$P((X = 1|Y = 1) = \frac{P(X = 1 \text{ e } Y = 1)}{P(Y = 1)} = \frac{\frac{1}{8}}{\frac{1}{2}} = \frac{1}{4}$$

$$P((X = 2|Y = 1) = \frac{P(X = 2 \text{ e } Y = 1)}{P(Y = 1)} = \frac{\frac{2}{8}}{\frac{1}{2}} = \frac{1}{2}$$

$$P(X = 3|Y = 1) = \frac{P(X = 3 \text{ e } Y = 1)}{P(Y = 1)} \frac{\frac{1}{8}}{\frac{1}{2}} = \frac{1}{4}.$$

Também já estão calculados

$$E(X|Y = 0) = \frac{1}{4} \times 0 + \frac{1}{2} \times 1 + \frac{1}{4} \times 2 + 0 \times 3 = 1$$

e

$$E(X|Y = 1) = 0 + \frac{1}{4} \times 1 + \frac{1}{2} \times 2 + \frac{1}{3} \times 3 = 2.$$

Faltam ser calculadas as $E(X^2|Y = 0)$ e $E(X^2|Y = 1)$.

$$\begin{aligned} E(X^2|Y = 0) &= P(X = 0|Y = 0) \times 0^2 + P(X = 1|Y = 0) \times 1^2 + \\ &\quad P(X = 2|Y = 0) \times 2^2 + P(X = 3|Y = 0) \times 3^2 \end{aligned} \quad (6.1)$$

$$E(X^2|Y = 0) = \frac{1}{4} \times 0^2 + \frac{1}{2} \times 1^2 + \frac{1}{4} \times 2^2 + 0 \times 3^2 = 1,5.$$

e

$$\begin{aligned} E(X^2|Y = 1) &= P(X = 0|Y = 1) \times 0^2 + P(X = 1|Y = 1) \times 1^2 + \\ &\quad P(X = 2|Y = 1) \times 2^2 + P(X = 3|Y = 1) \times 3^2 \end{aligned} \quad (6.2)$$

$$E(X^2|Y = 1) = 0 \times 0^2 + \frac{1}{4} \times 1^2 + \frac{1}{2} \times 2^2 + \frac{1}{4} \times 3^2 = 4,5.$$

Portanto, as variâncias condicionadas são

$$Var(X|Y = 0) = E(X^2|Y = 0) - [E(X|Y = 0)]^2 = 1,5 - 1^2 = 0,5.$$

$$Var(X|Y = 1) = E(X^2|Y = 1) - [E(X|Y = 1)]^2 = 4,5 - 2^2 = 0,5.$$

Note que a variância condicional é sempre ou quase sempre menor do que a variância incondicional, ou seja,

$$Var(X|Y) \leq Var(X)$$

e

$$Var(Y|X) \leq Var(Y).$$

Por outro lado, Esperança condicional pode ser menor ou maior ou igual a esperança incondicional. Ou seja,

$$E(X|Y) \geq E(X)$$

e

$$E(Y|X) \geq E(Y)$$

Para entender essas relações, toma-se o exemplo do time de volei.

Se é garantido que o primeiro jogo vitória, ou seja, dado que $Y = 1$, o número de vitórias esperado aumenta. Ao contrário, se o primeiro jogo é, com certeza, derrota, o número de vitórias esperado diminui.

No caso da variância, condicional ou não, é uma média de dispersão. O fato do resultado do primeiro jogo ser dado, seja o resultado qual for, diminui o número de resultados possíveis. Por isso, a variância condicional ser quase sempre menor.

Se as variáveis X e Y forem independentes, isso significa

$$P(X|Y) = P(X)$$

e

$$E(X|Y) = E(X)$$

e

$$Var(X|Y) = Var(X).$$

Obviamente, também

$$P(Y|X) = P(Y)$$

e

$$E(Y|X) = E(Y)$$

e

$$Var(Y|X) = Var(Y).$$

6.1.5 Decomposição da Variância

A decomposição da variância entrega uma relação interessante

$$Var(X) = Var[E(X|Y)] + E[Var(X|Y)]$$

Exemplo sobre a decomposição da variância

Exemplo 5.1.6 da página 115 do Sartoris (2013). Tomando o exemplo do time de volei, calcule $E[Var(X|Y)]$ e $Var[E(X|Y)]$

Resposta:

$$E[Var(X|Y)] = Var(X|Y = 0) \times P(Y = 0) + Var(X|Y = 1) \times P(Y = 1)$$

$$E[Var(X|Y)] = 0,5 \times 0,5 + 0,5 \times 0,5$$

$$E[Var(X|Y)] = 0,5.$$

No caso da $Var[E(X|Y)]$ é necessário calcular $E\{[E(X|Y)]^2\}$, pois

$$Var[E(X|Y)] = E\{[E(X|Y)]^2\} - \{E[E(X|Y)]\}^2$$

Então,

$$E\{[E(X|Y)]^2\} = [E(X|Y = 0)]^2 \times P(Y = 0) + [E(X|Y = 1)]^2 \times P(Y = 1) \quad (6.3)$$

$$E\{[E(X|Y)]^2\} = 1^2 \times 0,5 + 2^2 \times 0,5$$

$$E\{[E(X|Y)]^2\} = 2,5$$

Portanto,

$$Var[E(X|Y)] = E\{[E(X|Y)]^2\} - \{E[E(X|Y)]\}^2$$

$$Var[E(X|Y)] = 2,5 - (1,5)^2$$

$$Var[E(X|Y)] = 2,5 - 2,25$$

$$Var[E(X|Y)] = 0,25$$

6.2 Distribuição de probabilidade conjunta de variáveis aleatórias contínuas

Se as variáveis aleatórias forem contínuas, o procedimento é similar a da situação com uma variável. Neste caso haverá duas integrais no caso de duas variáveis.

6.2.1 Função densidade de probabilidade conjunta

Define-se uma função densidade de probabilidade (f.d.p.) conjunta $f(x, y)$ de tal forma que a probabilidade de x estar entre os valores a e b e de y estar entre c e d é dada por:

$$P(a < x < b \text{ e } c < y < d) = \int_c^d \int_a^b f(x, y) \, dx dy$$

Ou seja, a função densidade probabilidade conjunta, assim como a distribuição de probabilidade conjunta discreta, entrega a probabilidade para os intervalos definidos entre as duas ou mais variáveis.

Diferentemente da situação das variáveis aleatórias discretas, no caso de uma ou mais variáveis aleatórias contínuas a probabilidade só pode ser calculada para um intervalo pois,

$$P(X = x_0 \text{ e } Y = y_0) = 0,$$

mesmo que $X = x_0$ e $Y = y_0$ sejam eventos possíveis.

A função densidade de probabilidade conjunta deve seguir as mesmas propriedades da f.d.p. para uma variável, ou seja,

- não pode ser negativa e;

$$f(x, y) \geq 0$$

- a soma de todas as probabilidades tem de ser igual a 1.

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) \, dx dy = 1.$$

Exemplo numérico sobre a f.d.p. contínua e conjunta

Exemplo 5.2.1 da página 118 do Sartoris (2013). seja a função

$$f(x, y) = \begin{cases} Axy, & \text{para } 0 < x < 1 \text{ e } 0 < y < 1 \\ 0, & \text{para os demais valores} \end{cases} \quad (6.4)$$

Calcule o valor de A para que $f(x, y)$ seja uma função densidade de probabilidade.

Resposta:

Para que $f(x, y)$ seja uma f.d.p., deve atender a

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) \, dx dy = 1.$$

No caso específico do exemplo numérico, sabendo que tanto x como y variam entre zero e um se tem

$$\int_0^1 \int_0^1 f(x, y) \, dx dy = 1$$

$$\int_0^1 \int_0^1 Axy \, dx dy = 1$$

$$\int_0^1 Ay \int_0^1 x \, dx dy = 1$$

a regra de integração para uma função como a do exemplo numérico é de somar 1 a potência da variável para a qual está sendo integrada e dividir pelo resultado da soma da potência.

$$\int_0^1 Ay \left[\frac{x^2}{2} \right]_0^1 dy = 1$$

assim se calcula a diferença para $x = 1$ e $x = 0$

$$\int_0^1 Ay \left[\frac{1^2}{2} - \frac{0^2}{2} \right] dy = 1$$

e obtém-se a diferença

$$\int_0^1 Ay \frac{1}{2} dy = 1.$$

Como A e $1/2$ são constantes, coloca-se para fora da integral

$$\frac{A}{2} \int_0^1 y \, dy = 1$$

a integral para a variável y fica

$$\frac{A}{2} \left[\frac{y^2}{2} \right]_0^1 dy = 1$$

Calcula-se para $y = 1$ e $y = 0$

$$\frac{A}{2} \left[\frac{1^2}{2} - \frac{0^2}{2} \right] = 1$$

e obtém-se a diferença

$$\frac{A}{2} \frac{1}{2} = 1$$

$$\frac{A}{4} = 1$$

Portanto

$$A = 4$$

Assim a função $f(x, y)$ do exemplo numérico fica

$$f(x, y) = \begin{cases} 4xy, & \text{para } 0 < x < 1 \text{ e } 0 < y < 1 \\ 0, & \text{para os demais valores} \end{cases} \quad (6.5)$$

Outro exemplo numérico sobre a função densidade de probabilidade conjunta

Exemplo 5.2.2 da página 119 do Sartoris (2013). Dada a f.d.p. do exemplo numérico anterior

$$f(x, y) = \begin{cases} 4xy, & \text{para } 0 < x < 1 \text{ e } 0 < y < 1 \\ 0, & \text{para os demais valores} \end{cases} \quad (6.6)$$

Calcule a probabilidade de x estar entre 0,2 e 0,4 e de y estar entre 0,6 e 0,8.

Resposta:

A probabilidade conjunta é dada diretamente pela integral da função densidade de probabilidade

$$P(0,2 < x < 0,4 \text{ e } 0,6 < y < 0,8) = \int_{0,6}^{0,8} \int_{0,2}^{0,4} f(x, y) \, dx dy$$

$$P(0,2 < x < 0,4 \text{ e } 0,6 < y < 0,8) = \int_{0,6}^{0,8} \int_{0,2}^{0,4} 4xy \, dx dy$$

$$P(0,2 < x < 0,4 \text{ e } 0,6 < y < 0,8) = \int_{0,6}^{0,8} 4y \int_{0,2}^{0,4} x \, dx dy$$

integrando para x

$$P(0,2 < x < 0,4 \text{ e } 0,6 < y < 0,8) = \int_{0,6}^{0,8} 4y \left[\frac{x^2}{2} \right]_{0,2}^{0,4} dy$$

calculando para $x = 0,2$ e $x = 0,4$

$$P(0,2 < x < 0,4 \text{ e } 0,6 < y < 0,8) = \int_{0,6}^{0,8} 4y \left[\frac{(0,4)^2}{2} - \frac{(0,2)^2}{2} \right] dy$$

calculando a diferença para intervalo

$$P(0,2 < x < 0,4 \text{ e } 0,6 < y < 0,8) = \int_{0,6}^{0,8} 4y [0,08 - 0,02] dy$$

$$P(0,2 < x < 0,4 \text{ e } 0,6 < y < 0,8) = \int_{0,6}^{0,8} 0,24y dy$$

Como 0,24 é constante, coloca-se para fora da integral

$$P(0,2 < x < 0,4 \text{ e } 0,6 < y < 0,8) = 0,24 \int_{0,6}^{0,8} y dy$$

$$P(0,2 < x < 0,4 \text{ e } 0,6 < y < 0,8) = 0,24 \left[\frac{y^2}{2} \right]_{0,6}^{0,8}$$

$$P(0,2 < x < 0,4 \text{ e } 0,6 < y < 0,8) = 0,24 \left[\frac{(0,8)^2}{2} - \frac{(0,6)^2}{2} \right]$$

$$P(0,2 < x < 0,4 \text{ e } 0,6 < y < 0,8) = 0,24 \left[\frac{0,64}{2} - \frac{0,36}{2} \right]$$

$$P(0,2 < x < 0,4 \text{ e } 0,6 < y < 0,8) = 0,24 \times 0,14$$

$$P(0,2 < x < 0,4 \text{ e } 0,6 < y < 0,8) = 0,0336$$

6.2.2 Função densidade de probabilidade marginal

Para variáveis aleatórias discretas, a distribuição marginal de x é encontrada somando-se as probabilidades para todos os valores de y , e vice-versa.

No caso de variáveis aleatórias contínuas, a função densidade de probabilidade marginal de x , denominada de $g(x)$, é encontrada de forma análoga, ou seja, integrando-se $f(x, y)$ em y .

De maneira geral, a f.d.p. marginal de x pode ser encontrada da seguinte forma,

$$g(x) = \int_{-\infty}^{+\infty} f(x, y) dy$$

Exemplo numérico sobre função densidade de probabilidade marginal

Exemplo 5.2.3 da página 120 do Sartoris (2013). Seja a função densidade de probabilidade conjunta do exemplo numérico anterior, calcule as f.d.p. marginais de x e de y .

Resposta:

Como se encontra a f.d.p. marginal de x integrando $f(x, y)$ em y

$$g(x) = \int_{-\infty}^{+\infty} f(x, y) dy$$

no caso do exemplo numérico é

$$g(x) = \int_0^1 4xy \, dy.$$

Ou seja,

$$g(x) = 4x \int_0^1 y \, dy$$

$$g(x) = 4x \left[\frac{y^2}{2} \right]_0^1$$

$$g(x) = 4x \frac{1}{2}$$

$$g(x) = 2x$$

De forma análoga, a f.d.p. marginal de y , denotada como $h(y)$ é a integral de $f(x, y)$ em x

$$h(y) = \int_{-\infty}^{+\infty} f(x, y) dx$$

Assim, para o exemplo numérico fica

$$h(y) = \int_0^1 4xy \, dx.$$

$$h(y) = 4y \int_0^1 x \, dx$$

$$h(y) = 4y \left[\frac{x^2}{2} \right]_0^1$$

$$h(y) = 4y \frac{1}{2}$$

$$h(y) = 2y$$

Outro exemplo numérico sobre a f.d.p. marginal

Exemplo 5.2.4 da página 121 do Sartoris (2013). Seja a f.d.p. conjunta apresentada anteriormente

$$f(x, y) = \begin{cases} 4xy, & \text{para } 0 < x < 1 \text{ e } 0 < y < 1 \\ 0, & \text{para os demais valores} \end{cases} \quad (6.7)$$

calcule a probabilidade de x estar entre 0,3 e 0,7.

Resposta:

Note que é necessário integrar a f.d.p. conjunta em y nos seus limites definidos para a função, para obter a f.d.p. marginal de x o que permite calcular a probabilidade somente de x . No exemplo numérico anterior já foi calculado o $g(x)$ que é

$$g(x) = \int_0^1 4xy \, dy = 2x$$

Portanto,

$$P(0,3 < x < 0,7) = \int_{0,3}^{0,7} 2x \, dx$$

$$P(0,3 < x < 0,7) = [x^2]_{0,3}^{0,7} = 0,7^2 - 0,3^2$$

$$P(0,3 < x < 0,7) = 0,48 - 0,09$$

$$P(0,3 < x < 0,7) = 0,4.$$

6.2.3 Probabilidade condicional

A probabilidade condicional conjunta definido no contexto discreto é

$$P(X = x|Y = y) = \frac{P(X = x \cap Y = y)}{P(Y = y)}.$$

A $P(X = x \cap Y = y)$ corresponde a probabilidade conjunta, que no contexto contínuo corresponde a probabilidade conjunta dada pela função $f(x, y)$.

A $P(Y = y)$ corresponde a probabilidade marginal, que no contexto contínuo corresponde a probabilidade marginal dada pela função $h(y)$.

Portanto a probabilidade condicional no contexto contínuo, denotado como $f_{x|y}$ pode ser definido da seguinte forma

$$f_{x|y} = \frac{f(x, y)}{h(y)}$$

Ou seja, a f.d.p. condicional de x dado y .

De forma análoga, a f.d.p. condicional de y dado x , denotado como $f_{y|x}$ pode ser definido como

$$f_{y|x} = \frac{f(x, y)}{g(x)}$$

onde $g(x)$ é f.d.p. marginal de x .

Exemplo numérico sobre probabilidade marginal

Exemplo 5.2.5 da página 121 do Sartoris (2013).

Seja a f.d.p. dos exemplos numéricos anteriores

$$f(x, y) = \begin{cases} 4xy, & \text{para } 0 < x < 1 \text{ e } 0 < y < 1 \\ 0, & \text{para os demais valores} \end{cases} \quad (6.8)$$

calcule a f.d.p. condicional de x e y .

Resposta:

No caso da f.d.p. condicional de x dado y é necessário calcular

$$f_{x|y} = \frac{f(x, y)}{h(y)}$$

e já estão disponíveis $f(x, y)$

$$f(x, y) = 4xy$$

para $0 < x < 1$ e $0 < y < 1$, dado pelo exemplo corrente, e a f.d.p. marginal de y ,

$$h(y) = 2y$$

que foi calculada no exemplo numérico anterior para $0 < y < 1$.

Portanto,

$$f_{x|y} = \frac{f(x, y)}{h(y)} = \frac{4xy}{2y}$$

$$f_{x|y} = 2x$$

De forma análoga, a f.d.p. condicional de y dado x

$$f_{y|x} = \frac{f(x, y)}{g(x)}$$

Note que já estão disponíveis as partes necessárias para o cálculo de $f_{y|x}$

$$f_{y|x} = \frac{f(x, y)}{g(x)} = \frac{4xy}{2x}$$

$$f_{y|x} = 2y$$

Verificação da independência de duas variáveis

Note que quando ocorre

$$f_{x|y} = g(x)$$

e

$$f_{y|x} = h(y)$$

ou seja, as f.d.p. condicionais são iguais a f.d.p. incondicionais ou marginais, pode-se concluir que as variáveis são independentes.

Dessa forma, é válido afirmar que

$$f(x, y) = g(x) \times h(y).$$

Note que no último exemplo numérico, verifica-se que

$$f(x, y) = g(x) \times h(y) = 2x \times 2y = 4xy$$

A igualdade (6.2.3) é válida sempre que as variáveis forem independentes. Uma demonstração dessa igualdade é apresentada no apêndice do capítulo 5 de Sartoris (2013).

Do ponto de vista matemático, é possível verificar se as variáveis em uma função densidade de probabilidade conjunta são independentes, é verificar se é possível ser fatorada em funções para cada variável separadamente. Ou seja, o produto de uma função só para x e de uma função só para y , como se obteve no último exemplo numérico.

6.2.4 Esperança matemática de x

Para o cálculo da esperança matemática de x , a semelhança da esperança matemática para uma única variável, é possível realizar diretamente através de um f.d.p. conjunta $f(x, y)$, ou através de uma f.d.p. marginal de x

$$g(x) = \int_{-\infty}^{+\infty} f(x, y) dy,$$

Partindo-se da f.d.p. conjunta, obtém-se $E(x)$ calculando

$$E(x) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x f(x, y) dx dy.$$

Já partindo de uma f.d.p. marginal de x , obtém-se calculando

$$E(x) = \int_{-\infty}^{+\infty} x g(x) dx.$$

Exemplo numérico sobre o valor esperado de uma variável a partir de uma f.d.p. conjunta

Exemplo 5.2.6 da página 122 do Sartoris (2013). Seja a f.d.p. conjunta dos exemplos numéricos anteriores,

$$f(x, y) = \begin{cases} 4xy, & \text{para } 0 < x < 1 \text{ e } 0 < y < 1 \\ 0, & \text{para os demais valores} \end{cases} \quad (6.9)$$

calcule $E(x)$.

Resposta:

Partindo da f.d.p. conjunta, é necessário calcular

$$E(x) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x f(x, y) dx dy$$

$$E(x) = \int_0^1 \int_0^1 x \, 4xy \, dx dy$$

$$E(x) = 4 \int_0^1 \int_0^1 x^2 y \, dx dy$$

$$E(x) = 4 \int_0^1 y \int_0^1 x^2 \, dx dy$$

$$E(x) = 4 \int_0^1 y \left[\frac{x^3}{3} \right]_0^1 dy$$

$$E(x) = \frac{4}{3} \int_0^1 y \, dy$$

$$E(x) = \frac{4}{3} \left[\frac{y^2}{2} \right]_0^1 dy$$

$$E(x) = \frac{4}{3} \times \frac{1}{2}$$

$$E(x) = \frac{2}{3}$$

Partindo da f.d.p. marginal $g(x)$

$$g(x) = 2x$$

é necessário calcular

$$E(x) = \int_{-\infty}^{+\infty} x \, g(x) \, dx.$$

$$E(x) = \int_0^1 x \, 2x \, dx.$$

$$E(x) = 2 \int_0^1 x^2 \, dx.$$

$$E(x) = 2 \left[\frac{x^3}{3} \right]_0^1$$

$$E(x) = 2 \times \frac{1}{3}$$

$$E(x) = \frac{2}{3}$$

que é o mesmo valor do cálculo partindo da f.d.p. conjunta.

Note que o cálculo para $E(y)$ é similar ao cálculo de $E(x)$, tanto a partir da f.d.p. conjunta como a partir da f.d.p. marginal de y . Ou seja,

$$E(y) = \frac{2}{3}.$$

A verificação é deixada como exercício.

6.2.5 Variância de x

Assim como a esperança matemática de x , a variância de x também pode ser obtida através da f.d.p. conjunta $f(x, y)$ de x e y , ou através da f.d.p. marginal de x , denotada como $g(x)$. A semelhança da variância para uma única variável, pode-se usar a fórmula da definição

$$Var(x) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} [x - E(x)]^2 f(x, y) \, dx dy$$

ou a versão alternativa

$$Var(x) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x^2 f(x, y) \, dx dy - \left[\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x f(x, y) \, dx dy \right]^2$$

que, apesar das integrais, é a velha e conhecida média dos quadrados menos o quadrado da média.

Partindo da f.d.p. marginal $g(x)$, pode-se usar a fórmula da definição

$$Var(x) = \int_{-\infty}^{+\infty} [X - E(x)]^2 g(x) \, dx$$

ou a versão alternativa

$$Var(x) = \int_{-\infty}^{+\infty} x^2 g(x) \, dx - \left[\int_{-\infty}^{+\infty} x g(x) \, dx \right]^2$$

Exemplo sobre a variância de x

Exemplo 5.2.7 da página 127 do Sartoris (2013). Seja a f.d.p. conjunta dos exemplos numéricos anteriores

$$f(x, y) = \begin{cases} 4xy, & \text{para } 0 < x < 1 \text{ e } 0 < y < 1 \\ 0, & \text{para os demais valores} \end{cases} \quad (6.10)$$

calcule $Var(x)$.

Resposta:

Como já foi calculado a esperança de x utilizando a f.d.p. marginal de x

$$E(x) = \int_{-\infty}^{+\infty} x g(x) dx = \frac{2}{3}$$

e usando a fórmula alternativa da variância

$$Var(x) = \int_{-\infty}^{+\infty} x^2 g(x) dx - \left[\int_{-\infty}^{+\infty} x g(x) dx \right]^2$$

$$Var(x) = \int_0^1 x^2 g(x) dx - \left[\frac{2}{3} \right]^2$$

e falta calcular a esperança dos quadrados, a semelhaça do cálculo da variância de uma única variável.

$$Var(x) = \int_0^1 x^2 2x dx - \left[\frac{4}{9} \right]$$

$$Var(x) = 2 \int_0^1 x^3 dx - \left[\frac{4}{9} \right]$$

$$Var(x) = 2 \left[\frac{x^4}{4} \right]_0^1 - \left[\frac{4}{9} \right]$$

$$Var(x) = \frac{1}{2} - \frac{4}{9}$$

$$Var(x) = \frac{2-1}{18}$$

$$Var(x) = \frac{1}{18}$$

A variância de y é deixado como exercício.

6.2.6 Covariância entre x e y

Lembrando que a covariância apresentada na parte de estatísticas descritivas é dado por

$$\text{cov}(x, y) = E[x - E(x)][y - E(y)] = E(xy) - E(x)E(y).$$

Em termos de f.d.p. conjunta fica da seguinte forma considerando a definição de covariância

$$\text{Cov}(x, y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} [x - E(x)][y - E(y)] f(x, y) \, dx \, dy$$

e a forma alternativa

$$\text{Cov}(x, y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xy f(x, y) \, dx \, dy - \int_{-\infty}^{+\infty} x g(x) \, dx \int_{-\infty}^{+\infty} y h(y) \, dy$$

Exemplo numérico sobre a covariância entre x e y

Exemplo 5.2.8 da página 124 do Sartoris (2013). Seja a f.d.p. conjunta dos exemplos numéricos anteriores

$$f(x, y) = \begin{cases} 4xy, & \text{para } 0 < x < 1 \text{ e } 0 < y < 1 \\ 0, & \text{para os demais valores} \end{cases} \quad (6.11)$$

calcule a covariância entre x e y .

Resposta:

Como já foram calculadas as médias de x e de y , utiliza-se a forma alternativa da covariância. Assim só falta calcular $E(xy)$.

$$\text{Cov}(x, y) = \int_0^1 \int_0^1 xy \, 4xy \, dx \, dy - \frac{2}{3} \times \frac{2}{3}$$

$$\text{Cov}(x, y) = 4 \int_0^1 y^2 \int_0^1 x^2 \, dx \, dy - \frac{4}{9}$$

$$\text{Cov}(x, y) = 4 \int_0^1 y^2 \left[\frac{x^3}{3} \right]_0^1 dy - \frac{4}{9}$$

$$\text{Cov}(x, y) = 4 \int_0^1 y^2 \left[\frac{1}{3} \right] dy - \frac{4}{9}$$

$$\text{Cov}(x, y) = \frac{4}{3} \int_0^1 y^2 \, dy - \frac{2}{3} \times \frac{2}{3}$$

$$\text{Cov}(x, y) = \frac{4}{3} \left[\frac{y^3}{3} \right]_0^1 - \frac{4}{9}$$

$$\text{Cov}(x, y) = \frac{4}{9} - \frac{4}{9}$$

$$\text{Cov}(x, y) = 0$$

Note que já era esperado que a covariância fosse igual a zero, uma vez que verificou-se que as variáveis x e y do exemplo numérico desenvolvido são independentes.

Um outro exemplo numérico sobre f.d.p. conjunta

Exemplo 5.2.9 da página 125 do Sartoris (2013). Seja a função

$$f(x, y) = \begin{cases} B(x^2 + y^2), & \text{para } 0 < x < 1 \text{ e } 0 < y < 1 \\ 0, & \text{para os demais valores} \end{cases} \quad (6.12)$$

- Calcule o valor da constante B , de modo que função seja f.d.p. conjunta;
- Calcule as f.d.p. marginais de x e y ;
- Calcule as f.d.p. condicionais de x e y ;
- Verifique se x e y são variáveis independentes;
- Calcule $P(x < 0,5 | y = 0,5)$;

Respostas:

- Para ser uma função f.d.p. precisa atender a seguinte condição

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) \, dx dy = 1$$

como x e y variam entre zero e um é possível definir os limites das integrais

$$\int_0^1 \int_0^1 f(x, y) \, dx dy = 1$$

incluindo a função dada

$$\int_0^1 \int_0^1 B(x^2 + y^2) \, dx dy = 1$$

Como B é constante

$$B \int_0^1 \int_0^1 (x^2 + y^2) \, dx dy = 1$$

lembre-se que $x^0 = 1$

$$B \int_0^1 \int_0^1 (x^2 + y^2) \, dx dy = 1$$

$$B \int_0^1 \left[\frac{x^3}{3} + y^2 x \right]_0^1 dy = 1$$

$$B \int_0^1 \left[\frac{1}{3} + y^2 \right] dy = 1$$

$$B \left[\frac{1}{3} y + \frac{y^3}{3} \right]_0^1 = 1$$

$$B \left[\frac{1}{3} + \frac{1}{3} \right] = 1$$

$$B \times \frac{2}{3} = 1$$

$$B = \frac{3}{2}$$

Portanto, a f.d.p. conjunta deste exemplo numérico fica

$$f(x, y) = \int_0^1 \int_0^1 \frac{3}{2} (x^2 + y^2) \, dx dy$$

- b) Para obter a f.d.p. marginal de x , $g(x)$, é necessário integrar $f(x, y)$ em y .
Ou seja,

$$g(x) = \int_0^1 \frac{3}{2} (x^2 + y^2) \, dy$$

lembrando que $y^0 = 1$

$$g(x) = \frac{3}{2} \left[x^2 y + \frac{y^3}{3} \right]_0^1 dy$$

$$g(x) = \frac{3}{2} \left[x^2 + \frac{1}{3} \right]$$

Procede-se de forma similar para obter a f.d.p. marginal de y , $h(y)$, ou seja, integrando $f(x, y)$ em x

$$h(y) = \int_0^1 \frac{3}{2}(x^2 + y^2) \, dx$$

lembrando que $x^0 = 1$

$$h(y) = \frac{3}{2} \left[\frac{x^3}{3} + y^2 x \right]_0^1$$

$$h(y) = \frac{3}{2} \left[\frac{1}{3} + y^2 \right]$$

- c) Como se tem $f(x, y)$ e já foram calculadas as f.d.p. marginais de x e de y , fica fácil obter as f.d.p. condicionais de x e de y .

Para $f_{x|y}$

$$f_{x|y} = \frac{f(x, y)}{h(y)} = \frac{\frac{3}{2}(x^2 + y^2)}{\frac{3}{2} \left(\frac{1}{3} + y^2 \right)}$$

$$f_{x|y} = \frac{(x^2 + y^2)}{\left(\frac{1}{3} + y^2 \right)}$$

e para $f_{y|x}$

$$f_{y|x} = \frac{f(x, y)}{g(x)} = \frac{\frac{3}{2}(x^2 + y^2)}{\frac{3}{2} \left(x^2 + \frac{1}{3} \right)}$$

$$f_{y|x} = \frac{(x^2 + y^2)}{\left(x^2 + \frac{1}{3} \right)}$$

- d) Como

$$f_{x|y} \neq g(x)$$

e

$$f_{y|x} \neq h(y)$$

conclui-se que as variáveis x e y neste exemplo numérico não são independentes.

Mas, com resolução de (b) e de (c), já seria possível verificar que a função $(x^2 + y^2)$ não é fatorável de modo a obter uma função só de x ou outra só de y .

- e) Como se trata de uma probabilidade condicional, é necessário fazer uso da f.d.p. condicional de x dado $y = 0,5$. A $f_{x|y}$ é

$$f_{x|y} = \frac{(x^2 + y^2)}{\left(\frac{1}{3} + y^2\right)}$$

fazendo $y = 1/2$ a $f_{x|y}$ fica

$$f_{x|y} = \frac{x^2 + \left(\frac{1}{2}\right)^2}{\frac{1}{3} + \left(\frac{1}{2}\right)^2} = \frac{x^2 + \frac{1}{4}}{\frac{1}{3} + \frac{1}{4}} = \frac{x^2 + \frac{1}{4}}{\frac{7}{12}}$$

$$f_{x|y} = \frac{12}{7} \left(x^2 + \frac{1}{4}\right)$$

Agora falta definir que $x < 0,5$. uma vez que já é dado que $y = 0,5$. Assim,

$$P(x < 0,5 \mid y = 0,5) = \frac{12}{7} \int_0^{0,5} \left(x^2 + \frac{1}{4}\right) dx = \frac{12}{7} \left[\frac{x^3}{3} + \frac{1}{4}x\right]_0^{0,5} = \frac{12}{7} \left[\frac{1}{24} + \frac{1}{8}\right]$$

$$P(x < 0,5 \mid y = 0,5) = \left[\frac{1}{14} + \frac{3}{14}\right]$$

$$P(x < 0,5 \mid y = 0,5) = \frac{2}{7} \cong 0,2857$$

Exemplo sobre esperança condicional

Exemplo 5.2.10 da página 126 do Sartoris (2013). Seja a f.d.p. conjunta do exemplo numérico

$$f(x, y) = \begin{cases} \frac{3}{2} (x^2 + y^2), & \text{para } 0 < x < 1 \text{ e } 0 < y < 1 \\ 0, & \text{para os demais valores} \end{cases} \quad (6.13)$$

Calcule $E(x|y = 0,5)$.

Resposta:

Do exemplo numérico anterior sabe-se que

$$f_{x|y=0,5} = \frac{12}{7} \left(x^2 + \frac{1}{4} \right)$$

A esperança condicional de x é dada por:

$$E(x|y = y_0) = \int_{-\infty}^{+\infty} x f_{x|y} dx$$

Dessa forma

$$E(x|y = 0,5) = \frac{12}{7} \int_0^1 x \left(x^2 + \frac{1}{4} \right) dx$$

$$E(x|y = 0,5) = \frac{12}{7} \int_0^1 \left(x^3 + \frac{1}{4}x \right) dx$$

$$E(x|y = 0,5) = \frac{12}{7} \left(\frac{x^4}{4} + \frac{x^2}{8} \right)_0^1$$

$$E(x|y = 0,5) = \frac{12}{7} \left(\frac{1}{4} + \frac{1}{8} \right)$$

$$E(x|y = 0,5) = \frac{9}{14}$$

Capítulo 7

Estimadores

O conteúdo sobre estimadores é baseado no capítulo 6 do Sartoris (2013).

7.1 Introdução

Inferência Estatística ou **Estatística inferencial** é a parte da Estatística que faz inferência usadas para a descrição de uma população. Exemplos:

- colherada pra provar um molho que está sendo preparado;
- coleta de sangue para realizar vários exames do sangue;
- pesquisa de intenção de votos em uma cidade.

População e Amostra

O todo é denominado **população** e a parte é denominada **amostra**. Por isso, a **inferência estatística** trata de, a partir da amostra, obter informações sobre a população.

Quando o tamanho da amostra é muito pequeno em relação ao tamanho da população, se diz que a população é **infinita**.

Um exemplo seria a amostra de três mil eleitores para a pesquisa de intenção de votos na capital do Estado de São Paulo que tem atualmente um pouco mais de oito milhões e quatrocentos mil eleitores.

Continuará. Texto em construção

Capítulo 8

Teste de hipóste

Capítulo 9

Regressão Linear

Referências Bibliográficas

- Hoffmann, R. (2006). *Estatística para Economistas*. Cengage Learning, São Paulo, 4 edition.
- Morettin, P. A. and Bussab, W. d. O. (2013). *Estatística Básica*. Saraiva, São Paulo, 8 edition.
- Sartoris, A. (2013). *Estatística e Introdução à Econometria*. Saraiva, São Paulo, 2 edition.