Universidade do Minho
Escola de Ciências

Pedro da Silva Ximenes

Conceitos da Estatística Explorados com o Software R no Ensino Secundário em Timor-Leste

agosto de 2014

Pedro da Silva Ximenes **Conceitos da Estatística Explorados com o Software R no Ensino Secundário em Timor-Leste**

UMinho | 2014

Universidade do Minho
Escola de Ciências

Pedro da Silva Ximenes

Conceitos da Estatística Explorados com o Software R no Ensino Secundário em Timor-Leste

Dissertação de Mestrado
Mestrado em Ciências – Formação Contínua de Professores
Área de Especialização em Matemática

Trabalho realizado sob a orientação da
Professora Doutora Ana Paula Amorim

agosto de 2014

Pedro da Silva Ximenes

Endereço eletrónico: pedrodasilvaximenes@ymail.com

Título da dissertação: **Conceitos da Estatística Explorados com o Software R no Ensino Secundário em Timor-Leste**

Orientadora: **Professora Doutora Ana Paula Amorim**

Ano de conclusão: 2014

Mestrado em Ciências - Formação Contínua de Professores - Área Especialização em Matemática

Universidade do Minho, agosto de 2014

Assinatura:

AGRADECIMENTOS

Manifesto o meu sincero agradecimento e a minha gratidão a todos os que contribuíram para a realização deste trabalho.

- À Professora Doutora Ana Paula Amorim, que aceitou orientar este trabalho pela sua disponibilidade sem limites para apoiar e corrigir este trabalho.

- Aos professores da Escola de Ciências da Universidade do Minho que me apoiaram durante o meu curso de Mestrado.

- Ao gabinete da Bolsa de estudo de Ministério da Educação República Democrática de Timor-Leste pelo apoio à concretização deste Mestrado.

- À minha mulher e aos meus filhos que esperam o meu sucesso.

- Aos meus pais e família que apoiaram a realização deste mestrado.

- A todos aqueles que contribuíram para a realização deste trabalho.

RESUMO

A disciplina de Matemática é uma das mais importantes não só pela aplicação no dia a dia, mas também pelo seu contributo para o desenvolvimento do raciocínio. O raciocínio necessário para a resolução dos problemas matemáticos, pode ser utilizado em muitas áreas do conhecimento. A Matemática destaca-se como a disciplina mais importante do mundo moderno com aplicações diversificadas na área financeira, na indústria, na investigação e na informática. Esta universalidade exige um grande investimento no seu ensino. A Probabilidade, normalmente associa-se a algumas palavras como sorte, risco, azar, incerteza e dúvida. A Teoria das Probabilidades tenta quantificar a noção de provável. A Estatística recorrendo a números, tabelas e gráficos procura resumir, organizar e representar os dados das mais diversas áreas de atividade. O desenvolvimento das tecnologias disponibiliza à Estatística computadores com grande capacidade de processamento. Esta capacidade computacional permite tratar uma grande quantidade de dados estatísticos complexos que era uma tarefa muito morosa e custosa. A presente tese propõe a utilização do software R para o estudo das Probabilidades e Estatística nas aulas do Ensino Secundário em Timor-Leste. O software R é um programa computacional livre desenvolvido em 1993 por Robert Genglemet e Ross Ihaka. Nesta tese apresentam-se de um modo detalhado os temas constituintes das unidades curriculares de Probabilidades e Estatística da disciplina de Matemática do 12º ano: noções e conceitos das Probabilidades, da Estatística descritiva e indutiva, modelos paramétricos e estimação pontual de parâmetros desconhecidos dos modelos Bernoulli, Binomial, Poisson e Normal e ainda a estimação dos coeficientes do modelo de regressão linear simples e sua aplicação.

Palavras-chave: Matemática, Probabilidades e Estatística, Software R.

ABSTRACT

The Mathematics is one of the most important course not only for application in everyday life, but also for its contribution to the development of reasoning. The reasoning required to solve mathematical problems can be used in many areas of knowledge. Mathematics stands out as the most important discipline of the modern world with several applications in finance, industry, research and informatics. This universality requires a great investment in their teaching. The Probability is associated with some words like luck, risk, chance, uncertainty and doubt. The Probability Theory attempts to quantify the notion of probable. The Statistics using numbers, tables and graphs looking summarize, organize and represent data from various fields of activity. The development of technological allows to Statistics use computers with high processing capacity. This computational power allows analyzing a large amount of data using more complex statistical methods that was a very tiring and expensive task. This thesis proposes the use of the R software for the study of Probability and Statistics in classes of Secondary Education in Timor-Leste. The R software is a free software developed in 1993 by Ross Ihaka and Robert Genglemet. In this thesis are presented, in a detailed way, the subjects of the courses of Probability and Statistics in Mathematics of Year 12: notions and concepts of Probability, descriptive and inductive Statistics, parametric models and point estimation of unknown parameters of Bernoulli, Binomial, Poisson and Normal models and also the estimation of the coefficients of the linear regression model and its application.

Key words: Mathematics, Probability and Statistics, Software R.

Sumário

Capítulo 1

Introdução

1.1 Enquadramento da Tese

A República Democrática de Timor-Leste encontra-se numa fase de profundo investimento no sistema educativo, tendo como principal objetivo garantir o acesso de todas as crianças e jovens a uma formação bem estruturada e sólida, ao nível das melhores práticas internacionais, contribuíndo deste modo para o desenvolvimento do país. Para atingir este propósito, o governo de Timor-Leste está a realizar um forte investimento na formação contínua e inicial de professores nos diferentes níveis de ensino, garantindo deste modo uma preparação científica e pedagógica do corpo docente fundamental para a reforma profunda do processo ensino-aprendizagem.

Nos diferentes níveis de ensino foram elaborados novos programas com novas metodologias e nomeadamente na disciplina de matemática do ensino secundário é proposta uma abordagem exploratória dos conteúdos sempre que possível com o uso de calculadoras gráficas e/ou computadores.

O Ministério da Educação de Timor-Leste, no seu documento orientador a Lei de Bases da Educação (2008), salienta como objetivos a atingir na Matemática ao nível do ensino secundário:

- Assegurar e aprofundar as competências e os conteúdos fundamentais de uma formação e de uma cultura humanística, artística, científica e técnica, como suporte cognitivo e metodológico necessário ao prosseguimento de estudos superiores ou à inserção na vida activa;
- Assegurar o desenvolvimento do raciocínio, da reflexão e da curiosidade científica;
- Desenvolver as competências necessárias à compreensão das manifestações culturais e estéticas e possibilitar o aperfeiçoamento da expressão artística;

- Fomentar a aquisição e aplicação de um saber cada vez mais aprofundado, assente na leitura, no estudo, na reflexão crítica, na observação e na experimentação;

- Fomentar, a partir da realidade, e no apreço pelos valores permanentes da sociedade, em geral, e da cultura timorense, em particular, pessoas activamente empenhadas na concretização das opções estratégicas de desenvolvimento de Timor-Leste e sensibilizadas, criticamente, para a realidade da comunidade internacional;

- Assegurar a orientação e formação vocacional, através da preparação técnica e tecnológica adequada ao ingresso no mundo do trabalho;

- Facultar contactos e experiências com o mundo do trabalho, fortalecendo os mecanismos de aproximação entre a escola, a vida activa e a comunidade e dinamizando a função inovadora e interventora da escola;

- Assegurar a existência de hábitos de trabalho, individual e em grupo, e fomentar o desenvolvimento de atitudes de reflexão metódica, de abertura de espírito, de sensibilidade e de disponibilidade e adaptação à mudança.

O Ensino Secundário de Timor-Leste está organizado em duas áreas: Ciências e Tecnologias e Ciências Sociais e Humanidades, cada uma dela com um conjunto de disciplinas específicas. A Estatística integra a unidade temática Tratamentos de Dados existente no plano curricular do 7º ano e 8º ano de escolaridade. No Ensino Secundário as Probabilidades e Estatística correspondem à unidade temática 9 do programa de Matemática, disciplina obrigatória, para os alunos da área das Ciências e Tecnologias e cujos objetivos de aprendizagem apresentados no Plano Curricular do Ensino Secundário Geral (2011) são:

1. Compreender a relação entre o avanço científico e o progresso da Humanidade;

2. Aprofundar uma cultura científica e humanística que constitua suporte para o prosseguimento de estudos como para a inserção na vida activa;

3. Contribuir para o desenvolvimento da existência de uma consciência crítica e interventiva em áreas como o ambiente, a saúde e a economia entre outras formando para uma cidadania ativa e participativa;

4. Desenvolver a capacidade de usar a Matemática como instrumento de interpretação e intervenção no real;

5. Desenvolver as capacidades de formular e resolver problemas, de comunicar, assim como a memória, o espírito crítico e a criatividade;

6. Desenvolver a compreensão da Matemática como elemento da cultura humana, incluindo aspetos da sua história;
7. Analisar situações da vida real identificando modelos matemáticos que permitam a sua interpretação e resolução;
8. Interpretar fenómenos e resolver problemas recorrendo a funções e seus gráficos por via intuitiva e analítica;
9. Desenvolver a capacidade de formular hipóteses e prever resultados, assim como validar conjeturas e fazer raciocínios demonstrativos usando métodos adequados;
10. Desenvolver atitudes positivas face à Matemática e a capacidade de apreciar esta ciência.

No anexo A encontra-se a Unidade Temática 9 - Estatística Descritiva e Indutiva, do programa de Matemática do 12º ano de escolaridade em vigor em Timor-Leste.

1.2 Objetivos da Tese

A presente trabalho tem os seguintes objetivos principais:

- Conhecer os conceitos e resultados relativos à Estatística, Análise Combinatória e Probabilidades usados no Ensino Secundário;
- Conhecer a perspetiva histórica dos conceitos matemáticos envolvidos;
- Aprender as Probabilidades e a Estatística com instrumentos de interpretação e intervenção na realidade;
- Utilizar o software R apropriado para tratamento dos temas estudados.

1.3 Estrutura da Tese

Esta tese está organizada em sete capítulos. No primeiro capítulo apresenta-se a realidade da educação em Timor-Leste e o esforço do governo deste país para atingir uma boa qualidade de educação de nível internacional. São também referidos para além dos objetivos, as principais dificuldades na aprendizagem das Probabilidades e Estatística. No segundo capítulo faz-se uma breve introdução ao programa computacional R que é um programa livre com um número crescente de utilizadores e usado na área da Estatística. No terceiro capítulo apresentam-se as noções e os conceitos das Probabilidades. O capítulo 4 é dedicado à Estatística Descritiva do programa do 12º ano com métodos de recolha, apresentação e interpretação dos dados através de tabelas e gráficos. No capítulo 5 os modelos paramétricos incluídos no programa são apresentados e estudados com algum detalhe. O capítulo 6 é reservado às técnicas de estimação de parâmetros desconhecidos de modelos (Bernoulli, Binomial, Poisson e Normal). O modelo de regressão linear simples também é aqui estudado com detalhe. As conclusões do trabalho e o trabalho futuro são apresentadas no capítulo 7.

1.4 Dificuldades na Aprendizagem das Probabilidades e Estatística

Os alunos no Ensino Secundário têm normalmente uma visão parcial da utilidade da Estatística associando-a apenas à organização de dados númericos de uma amostra ou aos cálculos de média aritmética, desvio padrão, percentagem ou à elaboração de gráficos. A Estatística pode ser utilizada em todas as áreas do conhecimento como ferramenta e é considerada a tecnologia da ciência, auxiliando a pesquisa desde o planeamento até à interpretação dos dados.

A visão restrita da Estatística pelos alunos pode ser explicada pelo facto de que na disciplina de Matemática geralmente se ensinar apenas a Estatística descritiva. A Estatística inferencial, é geralmente precedida de muita teoria das probabilidades e não é aprofundada.

A recomendação para o ensino da estatística de acordo com Garfield e Ahlgren (1988) é introduzir os tópicos com atividades e simulações concretas, tentando explicar aos alunos que a Matemática não é apenas símbolos, regras e convenções mas tem a sua utilidade na vida real. As ilustrações e os métodos de exploração de dados permitem apresentar a Estatística descritiva sem os conceitos de probabilidades.

De um modo geral, os alunos sentem-se mais motivados para trabalhar dados que lhes estão associados, nomeadamente o peso, altura, distância da casa à escola e número de irmãos. Estes dados são um bom instrumento de trabalho para introduzir os conceitos estatísticos. De acordo com Stuart (1995) e Garfield e Chance (2000), deve-se iniciar a aprendizagem facilitando a linguagem e priveligiando a visualização gráfica, para que o aluno possa, então, comparar a sua intuição, a sua habilidade visual e o conceito estatístico. Fernandes e Barros (2005) questionam os conhecimentos dos professores dado que para o ensino das Probabilidades e Estatística deixa à sua compreensão de forma a levar os alunos a raciocinar corretamente.

A propósito do curriculum da disciplina de Matemática no documento *Plano Curricular do Ensino Secundário Geral*, em Timor-Leste, é referido que "o professor deve propor aos alunos a realização de diferentes tipos de tarefas, dando-lhes indicações claras em relação ao que se espera da atividade a desenvolver e apoiando-os na sua realização. Ao aluno podem proporcionar-se experiências matemáticas diversificadas, nomeadamente, resolver problemas, realizar atividade de natureza exploratória, desenvolver pequenos projetos, participar em jogos e ainda resolver exercícios práticos."

Ghinis et. al. (2009) nas suas investigações sobre as dificuldades na apreendizagem estatística conseguiu descobrir dois tipos de dificuldades. O primeiro tipo de dificuldade associada aos alunos prende-se com a compreensão dos conceitos básicos da Estatística, a suposição e conclusão do problema estatístico, a definição do método para obter a solução, a aplicação do método apropriado da Estatística, a validação do método da solução, a utilização dos cenceitos da Estatística na vida real, a interpretação dos resultados de um teste estatístico e a operação matemática para obter a solução. O segundo tipo de dificuldades está associada aos professores e é a apresentação dos conceitos básicos da Estatística, a organização dos dados na sala de aula, a análise dos dados, a intrepretação dos resultados de um teste estatístico e o ensino com o projetor ou slides. Garfield e Ahlgren (1988) e Ghinis at. al. (2009) defendem o ensino das Probabilidades e Estatística com recurso ao software sugerindo aos professores o uso do software na sala de aula e o trabalho com dados concretos envolvendo os alunos na recolha, organização e interpretação desses dados.

Fernandes (1999) analizou o desenvolvimento e a introdução das Probabilidades e Estatística no ensino aprendizagem em vários países. O autor afirma que na Áustria até 1970 faziam parte do currículo escolar o cálculo combinatório que em seguida era aplicado ao cálculo de probabilidades. Na Alemanha em 1980 verificaram-se alterações importantes resultantes da introdução das probabilidades e da estatística inferencial com uma pequena parte da estatística descritiva nos dois últimas anos do ensino secundário, e recentemente foi incluída a análise exploratória de dados. Em Inglaterra, as deficiências do ensino da Estatística, foram apontadas pelos profissionais de estatística desde a década de 70, como alvo o ensino da estatística , enquanto parte da sua educação geral dos alunos dos 11 aos 16 anos. Nos Estados Unidos a *estocástica* é o termo usado para designar conjuntamente o estudo das Probabilidades e Estatística, e tem sido utilizado particularmente na Europa Continental. Fernandes (1999), refere que ainda não constitui um hábito no ensino da Matemática, e presentemente muito pouca estatística é ensinada aos alunos antes de entrarem no Ensino Superior. A Hungria introduziu a *estocástica* ao nível de escolaridade básica em todos os anos (do 7º ao 8º ano de escolaridade) que fazia parte do respetivo programa de matemática com de a designação de *estocástica*. Em Portugal, a questão do ensino das Probabilidades e Estatística tem sido influenciada pelas opções dos outros países da Europa. No âmbito da Estatística e Probabilidades o programa da disciplina de Matemática da área científico-naturais de 1979/80 incluía no 11º ano de escolaridade tópicos de Cálculo Combinatório e de introdução à Estatística e às Probabilidades. Atualmente a Estatística tem vindo a ganhar protagonismo a nível dos programas da Matemática desde os primeiros anos do ensino básico até ao secundário.

Ponte & Fonseca (2001) referem que em Inglaterra, um dos países pioneiros deste campo, a Estatística começou a ser incluída na Matemática do Ensino Secundário no final dos anos 50, estreitamente ligada ao estudo das probabilidades e com uma orientação marcadamente teórica (com especial relevo para o estudo de testes de hipóteses). Este autor faz também uma distinção do lugar ou tendência da Estatística no currículo, de três países da Europa.

1. Ênfase no processo de Análise de Dados, na perspetiva em que esta ciência é utilizada na sociedade, tendo em conta que o uso de dados faz parte da vida de todos os dias (tendência predominante em países como a Inglaterra);

2. No currículo da Matemática, por vezes designada por *Estocástica*, enfatizando aspetos conceptuais e/ou computacionais (abordagem seguida, por exemplo em França);

3. Como "state"istics, ou seja, como uma ferramenta auxiliar para o estudo de diversos assuntos e disciplinas escolares (tendência visível, por exemplo, na Suécia).

Com o desenvolvimento das tecnologias, os computadores são muito usados como recurso ou ferramenta por vários cientistas nas mais variadas áreas de estudo, isto é para calcular, desenhar, para simular, etc. Em termos de Matemática, Ponte (1991) afirma:

"As relações entre a Matemática e o computador são complexas e interativas, desenvolvendo-se nos dois sentidos. Por um lado a Matemática é responsável por contributos decisivos para o seu surgimento e contínuo aperfeiçoamento, de tal forma espetacular que as suas capacidades em certos tarefas, ultrapassam as do próprio pensamento humano. Por outro lado a Matemática como ciência dinâmica e em constante evolução, vê o seu desenvolvimento já hoje influenciado pela sua existência, tanto no que respeita aos problemas como aos métodos de investigação."

Atualmente existem muitos programas de computadores para apoiar os alunos na apreendizagem. Na Probabilidade e Estatística, o R é um recurso importante no tratamento de dados. O R desenvolvido por Ross Ihaka e Robert Gentleman fornece uma linguagem básica com um bom desempenho no processo de análise, representação e interpretação os dados. Athayde (2013), propõe um manual de utilização do software R, no ensino da Estatística para o ensino superior, que vai sustentar toda a parte gráfica e computacional imprescindível à aplicação da Estatística. As vantagens de usar o software R de acordo com Muenchen (2011) são:

- O R oferece um vasto conjunto de métodos de análise estatística;
- O R oferece frequentemente implementações de novos métodos;
- O R tem uma vasta lista de livrarias recomendadas;
- O R rapidamente se transformou numa linguagem universal para tratamento de dados;
- Os gráficos em R são extremamente flexíveis e apresentam uma boa resolução;
- O R é muito flexível no tipo de dados que podem ser analisados;

- O R permite desenvolver programas especificos próprios;
- O R permite alterações às funções disponíveis;
- As funções específicas são tratadas de igual modo que as funções existentes no programa;
- O R corre nas plaformas, Windows, Macintosh, Linux, e UNIX;
- O R é livre e pode ser facilmente descarregado da internet.

Capítulo 2

O Software R

2.1 Introdução

O R é uma linguagem de programação e um ambiente de computação estatística e construção de gráficos. Esta linguagem é uma variante da linguagem comercial S desenvolvida no Bell Laboratories por John Chambers que ganhou o prestigiado prémio de software da organização ACM[1]. A linguagem R foi criada originalmente por Ross Ihaka e por Robert Gentleman no departamento de Estatística da Universidade de Auckland, Nova Zelândia e tem sido desenvolvida por um esforço colaborativo de pessoas em vários locais do mundo. A designação R está associada às iniciais do nome dos 2 criadores, como refere Muenchen (2011).

Este software contêm uma linguagem de programação que permite a computação de uma grande variedade de métodos de estatísticos e técnicas gráficas. Um dos pontos fortes do R é a facilidade com que produz gráficos bem delineados e de alta qualidade para impressão com possibilidade de inclusão de fórmulas e símbolos matemáticos quando necessário. Além disso, o software R também apresenta uma série de recursos gráficos que permitem a descrição detalhada de todos os aspetos que se podem querer personalizar num gráfico, como a cor, tipo e tamanho da letra, títulos e sub-títulos, pontos, linhas, legendas e planos de fundo.

2.2 Instalação do R

O processo de instalação do R depende de sistema operativo onde se pretende efetuar essa operação. O R está disponibilizado como software livre e aberto para todos os sistemas operativos (Linux, Unix, Windows, MacOs, etc). Para descarregar o R, deve-se aceder ao site www.r.project.org,

[1] Association for Computing Machinery

depois click em CRAN[2], escolher o servidor mais próximo e fazer o *download*. Após o termino da instalação, aparecerá a janela de finalização do instalador onde deve optar por ”Concluir”. A partir desse momento, o R já pode ser usado. A intalação do R está descrita no anexo B e para mais detalhe consultar Torgo (2009).

2.3 Leitura de ficheiros

A maneira mais fácil de inserir dados em objetos no R é a leitura de arquivos. Ele pode ler arquivos de estruturas simples com as extenções .txt. Também é possível importar outros tipos de arquivos mais complexos como $.xls$ mas nesse caso aconselha-se a salvá-lo como $.txt$.

Quando se salva uma área de trabalho, guarda-se o nome e o conteúdo dos objetos. Todos os comandos executados e todos os resultados não armazenados em objetos são perdidos.

Esta característica do R recomenda que se trabalhe no R em associação com um editor de texto da sua preferência. Alguns editores de texto muito úteis são: o script do R, o Bloco de notas do Windows, o Tinn-R, o WinEdt e o Emacs. Esses editores são usados tanto para elaborar os arquivos de dados que serão lidos pelo R, como para armazenar rotinas (conjuntos de linhas de comando) com vista à utilização futura.

Para ler uma tabela de dados no R usa-se a função read.table(). Esta função lê o arquivo e armazena-o na forma de data frame num objeto. O primeiro argumento dessa função refere-se ao nome do arquivo a ser lido. Esse argumento deve vir entre aspas. O endereço desse arquivo também deve ser passado ao R. Para isso, tem-se duas opções: (1) Na barra de menu, botão Arquivo, mudar diretório para o lugar onde se encontra o arquivo; (2) Escrever todo o endereço do arquivo dentro do primeiro argumento da função read.table(). O segundo argumento dessa função refere-se ao cabeçalho (nome) das colunas de dados contidas no arquivo. Se as colunas tiverem cabeçalho (header), então deve-se digitar $h = TRUE$, caso contrário, $h = FALSE$.

Exemplos de comando de leitura de arquivo quando se muda o diretório de leitura para o lugar onde o arquivo está armazenado

$> read.table(''nome.txt'', h = TRUE)$

e quando o endereço completo é passado na função

$> read.table(''C : \backslash \backslash \text{ Meus Documentos } \backslash \backslash nome.txt'', h = TRUE)$

[2]Compreensive R Archive Network

2.4 Help

O método mais fácil de se aprender a usar R é consultar os seus tópicos de ajuda. Os tipos de ajuda no R são basicamente:

- help('função()'): Esta ajuda deve ser solicitada quando se sabe da existência de uma função (sabe-se seu nome exato), mas existe dúvidas em como usá-la. Se o pacote que contém essa função estiver instalado será aberta a respectiva documentação;
- help.search(' '): quando se deseja investigar a existência de uma função, esta ajuda recebe uma palavra-chave (em Inglês) e retorna todas aqueles funções que contém aquela palavra na sua documentação. A busca é feita nos pacotes existentes no computador em questão, ou seja, se uma busca não retornar nenhum resultado adequado, não significa que a função não existe, mas sabemos pelo menos que não está instalada naquele computador.

2.5 Objetos

Mais que um software que realiza análises estatísticas o R é um ambiente de trabalho e uma linguagem de programação orientada a objetos. Nesta linguagem números, vetores, matrizes, arrays, data frames e listas podem ficar armazenados como objetos. Para criar um objeto é só atribuir um valor a um nome, ou seja, quando se coloca um valor dentro de um objeto, este passa a existir automaticamente. Uma atribuição pode ser feita usando o sinal de $=$ ou $< -$.

1. **Número**

   É possível atribuir apenas um número a um objeto.

   Por exemplo, o seguinte comando atribui o número 6 ao objeto a

   ```
   > a<-6
   ```

   e o número 3 ao objeto x.

   ```
   > x<-3
   ```

   Para verificar quanto vale o objeto, digite apenas o seu nome e faça enter.

   ```
   > a
   [1] 6
   > x
   [1] 3
   ```

Uma vez criados, os objetos podem ser usados em contas, equações, funções e sistemas.

```
> a+x # soma
[1] 9
> x-a # subtracao
[1] -3
> a+6
[1] 12
> x*a # produto de escalares
[1] 18
> a/x # divisao
[1] 2
> x^a # potenciacao
[1] 729
> sqrt(x) # raiz quadrada
[1] 1.732051
```

O resultado de uma operação matemática pode, por sua vez, ser guardado dentro de um terceiro objeto.

```
> d<-3*a+72/x
> d
[1] 42
```

2. **Vetor**

Uma das vantagens do R é a possibilidade de transformar um vetor de dados num vetor de resultados obtido pelo uso de uma função. Os elementos dos vetores podem ser números, palavras ou valores lógicos (F (falso) ou V (verdadeiro)). Para se atribuir um conjunto de valores a um objeto pode-se usar o comando c(), onde os valores aparecem separados por vírgulas, dentro de parênteses.

```
> v<-c(5, 8, 22, 32.12, 11.14, 5)
> x<-sqrt(v)-2*v
> x
[1]  -7.763932 -13.171573 -39.309584 -58.572549 -18.942336  -7.763932
> x1<-round(x,2) # arredonda o x para 2 digitos
> x1
```

```
[1]  -7.76 -13.17 -39.31 -58.57 -18.94  -7.76
> x1[3] # elemento na posicao 3 do vetor x1
[1] -39.31
> x1[5] # elemento na posicao 5 do vetor x1
[1] -18.94
```

3. **Matriz**

   Uma matriz pode ser criada usando a função $matrix()$. Essa função tem como argumentos o conjunto de dados, o número de linhas e o número de colunas da matriz.

```
> b<-matrix(c(3,5,7,6,8,-2,4,11,6),3,3)
> b
     [,1] [,2] [,3]
[1,]    3    6    4
[2,]    5    8   11
[3,]    7   -2    6

> diag(b) # diagonal principal da matriz b
[1] 3 8 6

> b1<-matrix(rep(5,9),3,3)
> b1
     [,1] [,2] [,3]
[1,]    5    5    5
[2,]    5    5    5
[3,]    5    5    5

> b2<-rbind(b[1,], b1[3,])

# criar nova matriz com a linha 1 da matriz b e linha 3 da matriz b1

> b2
     [,1] [,2] [,3]
[1,]    3    6    4
[2,]    5    5    5

> b3<-cbind(b[,1], b1[,3])
```

```
# criar nova matriz com a coluna 1 da matriz b e coluna 3 da matriz b1

> b3
     [,1] [,2]
[1,]    3    5
[2,]    5    5
[3,]    7    5

> 2*b2 # 2 vezes da matriz b2
     [,1] [,2] [,3]
[1,]    6   12    8
[2,]   10   10   10

> b+b1 # adicao da matriz b e b1
     [,1] [,2] [,3]
[1,]    8   11    9
[2,]   10   13   16
[3,]   12    3   11

> t(b) # transposta da matriz b
     [,1] [,2] [,3]
[1,]    3    5    7
[2,]    6    8   -2
[3,]    4   11    6

> b*b1 # multiplicacao da matriz b e b1
     [,1] [,2] [,3]
[1,]   15   30   20
[2,]   25   40   55
[3,]   35  -10   30

> b2*b3 # multiplicacao da matriz b2 e b3
     [,1] [,2]
[1,]    9   25
[2,]   25   25
[3,]   49   25

> det(b) # determinante da matriz b
[1] 228
```

```
> det(b+b1) # determinante de adicao da matriz b e b1
[1] 528

> b1-b  # subtracao da matriz b e b1
     [,1] [,2] [,3]
[1,]    2   -1    1
[2,]    0   -3   -6
[3,]   -2    7   -1
```

4. **Array**
   Os arrays representam uma generalização de uma matriz ou seja extensões das matrizes para mais do que duas dimensões. Quando tem três dimensões, um array pode ser entendido como um conjunto de matrizes de mesma dimensão. O comando para o array é o $array()$.

```
> d<-array(50:100, dim=c(2,5,5))
# criar uma matriz com numeros de 50 a 100,
# em 5 matrizes de 2 linhas e 5 colunas
> d
, , 1
     [,1] [,2] [,3] [,4] [,5]
[1,]   50   52   54   56   58
[2,]   51   53   55   57   59
, , 2
     [,1] [,2] [,3] [,4] [,5]
[1,]   60   62   64   66   68
[2,]   61   63   65   67   69
, , 3
     [,1] [,2] [,3] [,4] [,5]
[1,]   70   72   74   76   78
[2,]   71   73   75   77   79
, , 4
     [,1] [,2] [,3] [,4] [,5]
[1,]   80   82   84   86   88
[2,]   81   83   85   87   89
, , 5
     [,1] [,2] [,3] [,4] [,5]
[1,]   90   92   94   96   98
```

```
[2,]   91   93   95   97   99
```

5. **Lista**
   Uma lista é uma coleção ordenada de objetos de tamanhos e naturezas diferentes. A função para a lista é *list*(). Os objetos da lista são designadas por números entre dois parênteses [[ ]].
   Vamos exemplificar uma lista constituída por um número na primeira posição, uma matriz na segunda, uma palavra na terceira e uma vetor na quarta.

```
> p<-list(3,matrix(c(6,3,7,4),2,2),"numero",c(1,2,3,4))
> p
[[1]]
[1] 3
[[2]]
     [,1] [,2]
[1,]    6    7
[2,]    3    4
[[3]]
[1] "numero"
[[4]]
[1] 1 2 3 4
> dados<-list(n=23582, nome="Pedro da Silva Ximenes",
+ data=25121977, peso=70, altura=169)
# criar listas de dados
> dados
$n
[1] 23582
$nome
[1] "Pedro da Silva Ximenes"
$data
[1] 25121977
$peso # ou [[4]]
[1] 70
$altura # ou [[5]]
[1] 169
```

6. **Data frame** Uma data frame é uma espécie de tabela, de estrutura bidemensional de dados. A sua função é $data.frame()$. Num data frame podemos ter números e *strings* e podem ser dados nomes às colunas. Um exemplo deste tipo de objeto do R.

```
    > mes<-data.frame("ano2016"=c("janeiro", "fevereiro","marco","abril",
    +"maio","junho", "julho","agosto","setembro","outubro" ,"novembro",
    +"dezembro"), "dias"=c(31,29,31,30,31,30,31,31,30,31,30,31))
> mes
     ano2016  dias
1    janeiro   31
2  fevereiro   29
3      marco   31
4      abril   30
5       maio   31
6      junho   30
7      julho   31
8     agosto   31
9   setembro   30
10   outubro   31
11  novembro   30
12  dezembro   31
```

Capítulo 3

Probabilidade

3.1 Introdução

A probabilidade é um ramo de matemática que estuda fenómenos observáveis, influenciados pelo acaso ou seja, fenómenos aleatórios. Como introdução faremos uma breve referência histórica dos probabilistas mais notáveis e serão apresentados os seus principais contributos para o desenvovimento desta área do saber. A incerteza, associada aos fenómenos aleatórios, foi desde sempre a razão principal do estudo das probabilidades.

Giordamo Cardano (1501-1576) era Físico, Astrólogo e Matemático, de nacionalidade italiana escreveu uma vasta coleção de livros em diferentes áreas do saber. Das suas obras destaca-se o *Liber de Ludo Aleae* traduzido como (Livro de Jogos de Azar) que é considerado o primeiro livro completo dedicado às probabilidades.

O interesse pelos jogos de cartas e dados foram na altura o motivo para manter presente a discussão e reflexão sobre conceitos, definições e problemas concretos em torno das probabilidades. No século XVII, a troca de correspondência científica entre Pierre DeFermat (1601-1665) e Blaise Pascal (1623-1662) acerca de um problema colocado a Pascal por, Antonie Gombaud (1610-1685), mais conhecido por o Chavalier De Méré foram um forte incentivo no cálculo de probabilidades. O problema colocado pelo Chavalier De Méré consistia no seguinte: dois jogadores A e B estão a jogar os dados. Cada um aposta num determinado número e ganha o primeiro que obtiver pela terceira vez o número em que apostou. A aposta foi de 64 moedas (32 moedas de cada jogador) e o jogo foi interrompido quando o jogador A tinha dois sucessos contra um sucesso do adversário. A questão colocada foi a seguinte: como dividir o valor apostado? O jogador A (De Méré) entendia que tinha direito a 48 moedas ficando 16 para o adversário. O jogador B não tinha a mesma opinião e defendia que tinha direito a 1/3 das moedas, 21 moedas, ficando De Méré com as restantes 43. Pascal na correspondência trocada com Fermat, fez o seguinte raciocínio: "Ora eu (fez-se passar por De Méré) estou tão seguro de ter 32 moedas porque mesmo perdendo (entende-se por perder sair o número do adversário na próxima jogada) as ganho; quanto às outras 32, talvez eu as tenha, talvez vós as tinhais: o azar é igual. Partilhemos pois essas 32 moedas pela metade e assim receberei 16 para além das 32 que já me estão asseguradas". Foi assim que Pascal expôs o seu raciocínio a Fermat, atribuindo 48 moedas a De Méré.

Vários contributos na área das probabilidades foram dados por ilustres homens da ciência nomeadamente Cristian Huygens (1629-1695), Jacob Bernoulli (1654-1705), Abraham De Moivre (1667-1754), Thomas Bayes (1702-1761), Pierre-Simon Marquis de Laplace (1749-1827), Johann Carl Gauss (1777-1855) e Andrey Kolmogorov (1903-1987) com a conhecida axiomática de probabilidades, ver em Azevedo (2004) e Katz (2010).

Destaque para a obra *Ars Conjectandi* (Arte de Conjeturar) de Jacob Bernoulli, onde o autor apresentou o conceito frequencista de probabilidade, a conhecida Lei dos Grandes Números.

Em Portugal, Daniel Augusto da Silva (1814-1878) nasceu em Lisboa e licenciou-se na Universidade de Coimbra, em 1839. Autor de três memórias notáveis, que apresentou à Academia de Ciências de Lisboa entre 1850 e 1852. Na terceira memória intitulada de *Propriedades Gerais e Resolução das Congruências Binómias*, entre vários resultados apresentados destaca-se a fórmula do cardinal da reunião de n conjuntos quaisquer. Devido ao isolamento da ciência portuguesa em relação à ciência de outros países, muitos dos resultados a que este matemático chegou e publicou estão atribuídos a outros matemáticos que as obtiveram mais tarde, referido como curiosidade em Ministério da Educação (2013).

No enquadramento da tese este capítulo corresponde ao subtema de probabilidades, da unidade temática - Organização e tratamento de dados - do 12º ano do programa atual da disciplina de Matemática na República Democrática de Timor-Leste. Os conteúdos lecionados são: experiência aleatória, conjunto de resultados, acontecimentos, classificação de acontecimentos, operações com a contecimentos, aproximações conceptuais de probabilidade, aproximações frequencista, dfenição clássica de Laplace, definição axiomática (caso finito), propriedades da probabilidade, probabilidade condicionada e independência. A bibliografia base usada na elaboração deste capítulo foram os livros de Murteira et al.(2010) e Pestana & Velosa (2010).

Experiência Aleatória

Uma experiência aleatória é qualquer processo que gera um resultado que pode ser diferente de cada vez que o processo é executado em iguais condições e em que é conhecido o conjunto dos resultados possíveis.

Assim, uma experiência aleatória verifica as seguintes características:

1. possibilidade de repetição de experiência em condições iguais;

2. o conjunto Ω de todos os resultados possíveis é conhecido;

3. em cada realização da experiência não se sabe qual o resultado que irá ocorrer (fenómeno aleatório).

Exemplo 3.1 Exemplos de experiências aleatórias:

1. Lançamento de um dado e observação do número da face voltada para cima.
2. Retirar uma carta de um baralho e registar a cor.

Espaço de Resultados ou Espaço Amostra

O conjunto de todos os resultados possíveis de uma experiência aleatória é designado por espaço de resultados ou espaço amostra, e representa-se habitualmente por Ω ou S ou E.

Exemplo 3.2 A experiência aleatória que consiste em dois lançamentos ao ar de uma moeda de 50 centavos e registar em cada lançamento a face voltada para cima. Consideremos na moeda a face anverso designada por (A) e a face verso representada por (V). Um diagrama de árvore é útil no registo de todos os resultados possíveis desta experiência:

Figura 3.1: Resultados da experiência aleatória

O conjunto de todos os resultados possíveis é:

$$\Omega = \{(A, A), (A, V), (V, A), (V, V)\}.$$

Nota: O Exemplo 3.2 pode ser visto de uma outra forma. Podemos pensar em observar o número de vezes que ocorreu por exemplo a face verso, após os dois lançamentos da moeda de 50 centavos. Neste caso o espaço amostral será $\Omega = \{0, 1, 2\}$. Os valores do espaço amostra correspodem à não saída da face verso, saída de uma face verso e saída de duas faces verso.

Exemplo 3.3 Para as seguintes experiências aleatórias determinar o espaço de resultados:

1. Lançamento de um dado e observação do número da face voltada para cima. O espaço de resultados é $\Omega = \{1, 2, 3, 4, 5, 6\}$;
2. Lançamento uma moeda de 50 centavos (com faces "A" e "V") e registo da face voltada para cima. O espaço de resultados é $\Omega = \{A, V\}$;
3. Dois lançamentos de uma moeda de 50 centavos (com faces "A" e "V")) e observação das duas faces voltadas para cima. O espaço de resultados é $\Omega = \{AA, AV, VA, VV\}$.

Acontecimentos

Dada uma experiência aleatória em que o espaço amostra é Ω, chama-se **acontecimento** a todo o subconjunto de Ω.

Exemplo 3.4 Considere a experiência aleatória que consiste em lançar um dado equilibrado com as faces numeradas de 1 a 6 e registar a face voltada para cima.

O espaço amostra associado a esta experiência aleatória é: $\Omega = \{1, 2, 3, 4, 5, 6\}$.

Considere os seguintes acontecimentos:

A: "O número da face voltada para cima é par";

B: "O número da face voltada para cima é múltiplo de 6";

C: "O número da face voltada para cima é múltiplo de 10";

D: "O número da face voltada para cima é divisor de 420";

Para cada um dos acontecimentos está definido um subconjunto do espaço amostra

$$A = \{2, 4, 6\}, \quad B = \{6\}, \quad C = \{\} = \emptyset, \quad D = \{1, 2, 3, 4, 5, 6\}$$

Observe que:

Quando a um acontecimento corresponde o conjunto vazio, diz-se que é um **acontecimento impossível**.

Quando a um acontecimento corresponde o conjunto Ω, diz-se que é um **acontecimento certo**.

Quando a um acontecimento corresponde o conjunto que tem apenas um e um só elemento do espaço amostra, diz-se que é um **acontecimento elementar**.

Quando a um acontecimento corresponde o conjunto com mais do que um elemento do espaço amostra, diz-se que é um **acontecimento composto**.

Espaço de Acontecimentos é o conjunto formado por todos os subconjuntos do espaço amostra e designa-se por $P(\Omega)$.

Exemplo 3.5 Numa caixa estão três bolas numeradas de 1 a 3.

O espaço amostral é $\Omega = \{1, 2, 3\}$.

O espaço de acontecimentos é:

$$P(\Omega) = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\}$$

3.2 Operação com Acontecimentos

Como os acontecimentos estão associados a conjuntos, a maneira de operar com acontecimentos decorre do modo de se operar com conjuntos. Vamos relembrar as principais propriedades da teoria de conjuntos recorrendo sempre que possível a diagramas de Venn.

União de Acontecimentos

Sejam A e B dois acontecimentos definidos no espaço Ω, o acontecimento união (reunião) de A com B representa-se como $A \cup B$. Podemos escrever: $A \cup B = \{\omega : \omega \in A \vee \omega \in B\}$.

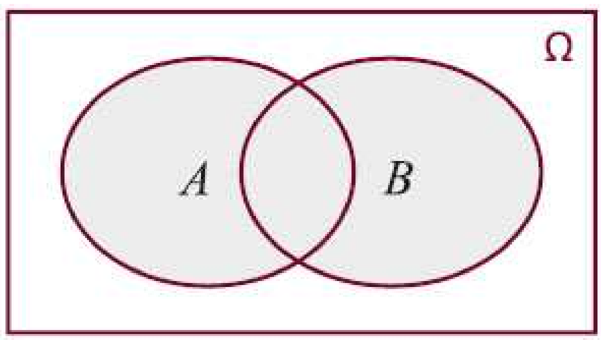


Figura 3.2: União dos acontecimentos A e B

Exemplo 3.6 Numa caixa com bolas numeradas de 1 a 6, consideramos os seguintes acontecimentos:

A: ”ser um número primo”;

B: ”ser um número divisor de 4”;

$\Omega = \{1, 2, 3, 4, 5, 6\}$ $A = \{2, 3, 5\}$ $B = \{1, 2, 4\}$

O acontecimento união é o acontecimento, $A \cup B$: ”ser um número primo ou divisor de 4”

$A \cup B = \{1, 2, 3, 4, 5\}$.

Interseção de Acontecimentos

A interseção de dois acontecimento A e B é o acontecimento que se realiza-se se e só se A e B acontecem simultaneamente. Representa-se por $A \cap B$ e $A \cap B = \{\omega : \omega \in A \wedge \omega \in B\}$.

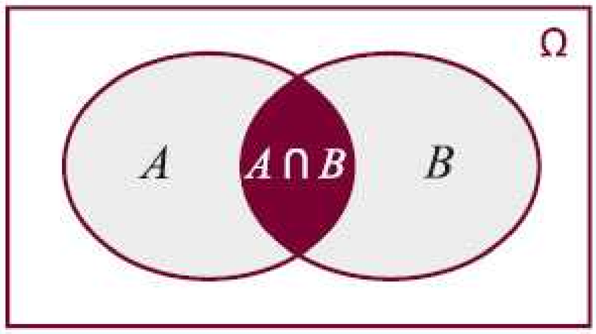


Figura 3.3: Interseção dos acontecimentos A e B

Exemplo 3.7 Continuação do Exemplo 1.6, consideramos agora o acontecimento interseção que é o acontecimento:

$A \cap B$: "ser um número primo e divisor de 4"

$A \cap B = \{2\}$.

Acontecimentos Incompatíveis ou Disjuntos

Acontecimentos incompatíveis ou disjuntos são acontecimentos que não têm resultados comuns. A e B são incompatíveis se e ó se $A \cap B = \varnothing$.

Exemplo 3.8 No lançamento de um dado consideramos os seguintes acontecimentos:

A: "ser um número múltiplo de 2";

B: "ser um número ímpar";

$\Omega = \{1, 2, 3, 4, 5, 6\}$ $A = \{2, 4, 6\}$ $B = \{1, 3, 5\}$

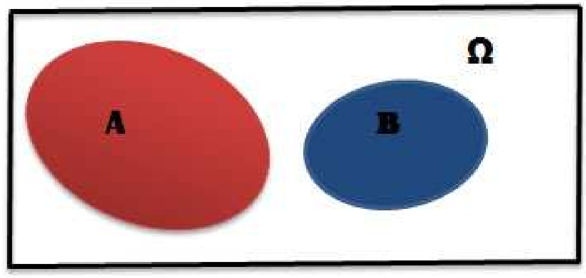


Figura 3.4: Acontecimentos incompatíveis

Os acontecimentos A e B são incompatíveis, dado que não existem múltiplos de 2 que sejam primos, donde $A \cap B = \emptyset$.

Acontecimento Contrário ou Complementar

O acontecimento contrário ou complementar a A é o acontecimento constituído por todos os resultados do espaço amostral que não pertencem a A e representa-se por $\overline{A}$. Consequentemente são verificadas as seguintes propriedades entre os conjuntos:

$A \cup \overline{A} = \Omega$

e

$A \cap \overline{A} = \varnothing$.

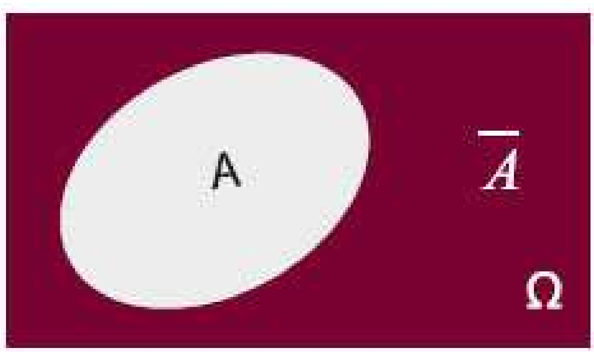


Figura 3.5: Acontecimento contrário ou complementar

Exemplo 3.9 No lançamento de um dado consideramos o acontecimento A e o seu contrário $\overline{A}$:

A: ”ser um número inferior a 4”;

$\overline{A}$: ”ser um número superior ou igual a 4”;

$\Omega = \{1, 2, 3, 4, 5, 6\}$ $A = \{1, 2, 3\}$ $\overline{A} = \{4, 5, 6\}$

Diferença de Acontecimentos

O acontecimento diferença entre A e B é o acontecimento que se realiza sempre que se realiza A e não se realiza o acontecimento B. Será assim, o acontecimento constituído por todos os elementos de A que simultaneamente não pertencem de B, ou seja:

$$A - B = A \setminus B = A \cap \overline{B}.$$

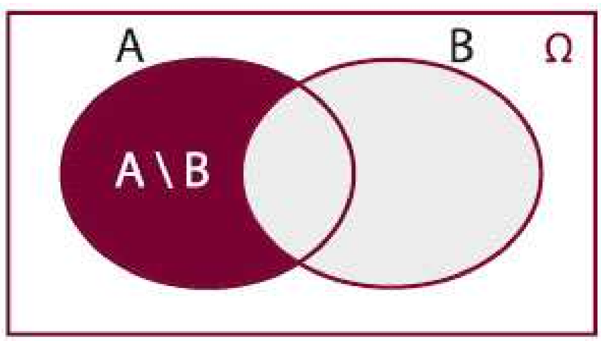


Figura 3.6: Diferença de acontecimentos

3.3 Relação entre Conjuntos e Acontecimentos

Na primeira tabela 3.1 faz-se a correspondência entre a notação de conjuntos e a notação de acontecimentos, na tabela 3.2 relembram-se as propriedades dos conjuntos para as operações de união e interseção. Considere-se uma experiência aleatória com A, B e C acontecimentos quaisquer e Ω o espaço de resultados.

Tabela 3.1: Correspondência entre Conjuntos e Acontecimentos

Notação de Conjuntos	Notação de Acontecimentos
Ω - Universo: conjunto de todos os pontos ou elementos.	Ω - Espaço de resultados: conjunto de todos resultados.
$\varnothing$ -conjunto vazio: conjunto que não contém elementos.	$\varnothing$ - Acontecimento impossível: inexistência de resultados.
$\overline{A}$ - Conjunto complementar: conjunto de pontos que são de $\overline{A}$ e não são de A.	$\overline{A}$ - Não ocorrência do acontecimento A.
$A \cup B$ - União: conjunto de pontos que são de A, que são de B e em que são de ambos.	$A \cup B$ - Pelo menos um: ocorrência de pelo menos um acontecimento.
$A \cap B$ - Interseção: conjunto de pontos que são de A e de B.	$A \cap B$ - Simultâneo: ocorrência em simultâneo de dois acontecimentos.
$A - B$ - Diferença: conjunto de pontos que são de A e não são de B.	$A - B$ - Ocorre A e não B.

Tabela 3.2: Propriedades dos Conjuntos

Propriedades	União	Interseção
Comutativa	$A \cup B = B \cup A$	$A \cap B = B \cap A$
Associativa	$A \cup (B \cup C) = (A \cup B) \cup C$	$A \cap (B \cap C) = (A \cap B) \cap C$
Distribuitiva	$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$	$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
Indempotência	$A \cup A = A$	$A \cap A = A$
Lei do Complemento	$A \cup \overline{A} = \Omega$	$A \cap \overline{A} = \varnothing$
Elemento Neutro	$A \cup \varnothing = A$	$A \cap \Omega = A$
Elemento Absorvente	$A \cup \Omega = \Omega$	$A \cap \varnothing = 0$

Leis De Morgan

A acrescentar ao conjunto de operações entre conjuntos/acontecimentos temos as chamadas leis De Morgan:

1. Negar que se realiza pelo menos um dos acontecimentos é afirmar que não se realiza nem

um nem outro.

$$\overline{A} \cup \overline{B} = \overline{A \cap B}$$

2. Negar que se realizam simultaneamente dois acontecimentos é dizer que não se realiza pelo menos um deles.

$$\overline{A \cap B} = \overline{A} \cup \overline{B}$$

3.4 Conceito de Probabilidade

Teoria frequencista da probabilidade

Para introduzir o conceito de frequência relativa de um acontecimento, vamos considerar a experiência do lançamento de um dado octaédrico com as faces numeradas de 1 a 8 e o registo do número da face voltada para baixo.

O espaço amostral associado a esta experiência é:

$$\Omega = \{1, 2, 3, 4, 5, 6, 7, 8\}$$

e os acontecimentos elementares são:

$$\{1\}, \quad \{2\}, \quad \{3\}, \quad \{4\}, \quad \{5\}, \quad \{6\}, \quad \{7\}, \quad \{8\}$$

Após a repetição da experiência várias vezes registou-se que o acontecimento A: "saída da face 8"ocorreu 40 vezes. O número 40 por si só é pouco informativo, contudo enquadrado num total de realizações da experiência ganha outra dimensão. Se a experiência foi repetida 50 vezes, o número 40 significa que o acontecimento A ocorreu muitas vezes. Se porém a experiência se realizou 300 vezes então a interpretação é de que o acontecimento A ocorreu poucas vezes.

O número de vezes que ocorreu o acontecimento A representa a **frequência absoluta** do acontecimento. Com o conhecimento do número total de experiências realizadas podemos definir a **frequência relativa** do acontecimento A como $\frac{40}{300} = 0.133$, em percentagem, $13,3\%$.

Se uma experiência é realizada n vezes e o acontecimento A ocorre m vezes ($m \leq n$), define-se frequência relativa do acontecimento A como sendo o quociente $\frac{m}{n}$.

Usualmente representa-se por: $f_r(A) = \frac{m}{n}$

Exemplo 3.10 Consideremos a experiência aleatória que consiste no lançamento de dois dados equilibrados e determinemos o valor absoluto da diferença dos pontos das faces voltadas para cima. No quadro seguinte encontra-se a representação do espaço de resultados desta experiência, para o acontecimento A: "valor absoluto da diferença dos pontos das faces":

A	1	2	3	4	5	6
1	0	1	2	3	4	5
2	1	0	1	2	3	4
3	2	1	0	1	2	3
4	3	2	1	0	1	2
5	4	3	2	1	0	1
6	5	4	3	2	1	0

No quadro está representada a distribuição de frequências relativas.

Acontecimento	0	1	2	3	4	5
Freq. relativa	$\frac{6}{36}$	$\frac{10}{36}$	$\frac{8}{36}$	$\frac{6}{36}$	$\frac{4}{36}$	$\frac{2}{36}$

Propriedades da frequência relativa de um acontecimento

- Se A é um acontecimento impossível, então $f_r(A) = 0$.
- Se A é um acontecimento certo, então $f_r(A) = 1$.
- Se A é um acontecimento qualquer, então $0 \leq f_r(A) \leq 1$.

- Se A é um acontecimento composto, $A = \{A_1, A_2, A_3, \ldots\}$, com $A_1, A_2, \ldots$ disjuntos dois a dois,

  então
$$f_r(A) = f_r(A_1) + f_r(A_2) + f_r(A_3) + \ldots$$

- A soma das frequências relativas de todos os acontecimentos elementares é 1.

- Se A e $\overline{A}$ são acontecimentos contrários, então $f_r(A) + f_r(\overline{A}) = 1$.

Lei dos grandes números

A teoria frequencista assume que a frequência relativa de um acontecimento tende a estabilizar num determinado valor, à medida que aumenta o número de repetições da experiência aleatória e é esse o valor que se assume para a probabilidade de um determinado acontecimento ocorrer.

Definição frequencista de probabilidade

A probabilidade (empírica ou frequencista) do acontecimento A representa-se por $P(A)$ e corresponde ao valor para o qual a frequência relativa tende a estabilizar quando o número de experiências tende para infinito.

Definição clássica de probabilidade ou de Laplace

Consideremos a experiência aleatória de lançamento de um dado com as faces numeradas de 1 a 6 sendo registado o número da face voltada para cima. Seja A o acontecimento: "saída de um número inferior a 4".

O espaço amostral é $\Omega = \{1, 2, 3, 4, 5, 6\}$. Qualquer um dos acontecimentos elementares do espaço amostral tem igual probabilidade de ocorrer, isto significa que todas as faces do dado são equiprováveis.

Sendo A o acontecimento: ”saída de um número inferior a 4”, $A = \{1,2,3\}$. Existem três resultados favoráveis em seis resultados possíveis. Reparemos que o número de casos favoráveis é igual ao número de acontecimentos elementares que fazem parte do acontecimento A. Pretendemos determinar a probabilidade de ocorrer o acontecimento A:

$$\begin{aligned} P(A) = &\ P(\{1\}) + P(\{2\}) + P(\{3\}) \\ = &\ \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2} \end{aligned}$$

Reparemos que:

$$P(A) = \frac{\text{número de casos favoráveis a A}}{\text{número de casos possíveis}}$$

Lei de Laplace

Consideremos uma experiência aleatória onde o espaço amostral Ω é constituído por n elementos, sendo equiprováveis os n acontecimentos elementares.

Se um acontecimento A é constituído por m acontecimentos elementares, sendo $m \leq n$, a probabilidade de A é dada pelo quociente entre o número de casos favoràveis e o número de casos possíveis, isto é

$$P(A) = \frac{m}{n}$$

Nota 1: A lei de Laplace é conhecida como a primeira definição de probabilidade e por isso também ser conhecida como definição clássica. A aplicação desta regra exige que os acontecimentos elementares sejam equiprováveis.

Nota 2: A probabilidade de $P(A)$ escrita de modo equivalente:

$$P(A) = \frac{\text{número de casos favoráveis a A}}{\text{número de casos possíveis}} = \frac{\#A}{n}$$

onde $\#A$ = ”número de casos favoráveis a A” e n= ”número de resultados possíveis”. Dado um conjunto A, diferente do vazio e finito, ao número de elementos de A chamamos cardinal de A e representa-se por $\#A$.

Exemplo 3.11 Considere uma experiência que consiste no lançamento de um dado equilibrado, cujas faces estão numeradas de um a seis. Sejam A, B e C os acontecimentos:

A: "sair um número superior a 5"

B: "sair um número ímpar"

C: "sair um número ímpar e primo"

Qual a probabilidade de cada acontecimento?

Resolução

O espaço de resultados é $\Omega = \{1, 2, 3, 4, 5, 6\}$ com n=$\#\Omega = 6$.

Acontecimento A: $A = \{6\}$, então $\#A = 1$, logo

$$P(A) = \frac{\#A}{n} = \frac{1}{6}$$

Acontecimento B: $B = \{1, 3, 5\}$, então $\#B = 3$, logo

$$P(B) = \frac{\#B}{n} = \frac{3}{6} = \frac{1}{2}$$

Acontecimento C: $C = \{3, 5\}$, então $\#C = 2$, logo

$$P(C) = \frac{\#C}{n} = \frac{2}{6} = \frac{1}{3}$$

Exemplo 3.12 Considere o lançamento de dois dados equilibrados, ambos com as faces numeradas de um a seis. Calcule as probabilidades dos seguintes acontecimentos:

A: "a soma dos pontos das faces dos dois dados é 10"

B: "a soma dos pontos das faces dos dois dados é maior ou igual a 10"

C: "a soma dos pontos das faces dos dois dados é um número primo"

D: "a soma dos pontos das faces dos dois dados é um quadrado perfeito"

Resolução

O espaço de resultados da soma dos números das faces ocorridas no lançamento de dois dados equilibrados: $\#\Omega = 6 \times 6 = 36$

A: "a soma dos pontos das faces dos dois dados é 10"

$$A = \{(4,6),(5,5),(6,4)\} \Rightarrow \#A = 3$$

então a probabilidade de

$$P(A) = \frac{3}{36} = \frac{1}{12}$$

Todos os possíveis resultados que podem ocorrer nesta experiência estão representados na tabela seguinte:

Tabela 3.3: Espaço de resultados associado à soma dos pontos das faces

+	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

De igual modo, podemos construir tabelas idênticas para restantes alíneas do exercício.

B: "a soma dos pontos das faces dos dados é maior ou igual a 10"

$$B = \{(4,6),(5,5),(5,6),(6,4),(6,5),(6,6)\} \Rightarrow \#B = 6$$

então

$$P(B) = \frac{6}{36} = \frac{1}{6}$$

C: "a soma dos pontos das faces dos dados é um número primo"

$$\begin{aligned} C = \{&(1,1),(1,2),(1,4),(1,6),(2,1),(2,3),(2,5) \\ &(3,2),(3,4),(4,1),(4,3),(5,2),(5,6),(6,1),(6,5)\} \Rightarrow \#C = 15 \end{aligned}$$

então

$$P(C) = \frac{15}{36} = \frac{5}{12}$$

D: "a soma dos pontos das faces dos dados é um quadrado perfeito"

$$D = \{(1,3),(2,2),(3,1),(3,6),(4,5),(5,4),(6,3)\} \Rightarrow \#D = 7$$

logo

$$P(B) = \frac{7}{36}$$

3.5 Princípio fundamental de contagem

O Princípio fundamental de contagem aplica-se quando queremos realizar k escolhas sucessivas em que na primeira há n_1 alternativas, e na segunda há n_2 alternativas e assim sucessivamente.

Podemos afirmar que o número total de alternativas é dado por

$$n_1 \times n_2 \times \ldots \times n_k.$$

Exemplo 3.13 Um saco contém quatro bolas numeradas de 1 a 4. A Joana retira sucessivamente três bolas, sem reposição, e escreve o número de três algarismos, em que o algarismo das centenas é o número da primeira bola retirada, o algarismo das dezenas é o número da segunda bola retirada e por último o algarismo das unidades é o número da terceira bola retirada. Pretendemos saber qual é a probabilidade de o algarismo 4 não aparecer no número escrito pela Joana?

Seja A o acontecimento, tal que A: "número de três algarismos escrito pela Joana, não contém o algarismo 4".

O exercício pode ser resolvido recorrendo a uma representação em diagrama em árvore, onde a contagem dos casos possíveis e favoráveis fica facilitada, mas iremos optar por simplesmente contar o número de casos possíveis e favoráveis fazendo o seguinte raciocínio:

O número pretendido é da forma CDU, em que C corresponde ao algarismo das centenas, D corresponde ao algarismo das dezenas e U ao algarismo das unidades.

Casos possíveis: para ocupar o lugar C no número temos 4 possibilidades; para D há 3 possibilidades e para U há apenas 2 possibilidades. Obtemos então o número de casos possíveis igual a 24 e que foi obtido por $4 \times 3 \times 2$.

Casos favoráveis: para ocupar o lugar C no número temos 3 possibilidades; para D há 2 possibilidades e para U há apenas 1 possibilidades. Obtemos então o número de casos favoráveis igual a 6 e que foi obtido por $3 \times 2 \times 1$.

$$P(A) = \frac{3 \times 2 \times 1}{4 \times 3 \times 2} = \frac{1}{4}.$$

Exemplo 3.14 Um saco tem 5 bolas, 2 vermelhas(V), uma azul (A) e duas brancas (B). Considere a experiência aleatória que consiste em retirar, sucessivamente sem reposição, 2 bolas e verificar as cores.

Qual a probabilidade do acontecimento, do acontecimento em que ambas as bolas são vermelhas?

C: ”Ambas as bolas são vermelhas”

Consideremos que: V:”saída de bola vermelha”; A:”saída de bola azul”e B:”saída de bola branca”.

O espaço amostral é:

$$\Omega = \{VV; VA; VB; AV; AB; BV; BA; BB\}$$

Estruturamos o exercício recorrendo a uma representação em diagrama em árvore e em cada ramo colocamos a probabilidade do acontecimento:

Reparemos que os acontecimentos elementares apresentam as seguintes probabilidades:

$P(V) = 2/5$; $P(A) = 1/5$ e $P(B) = 2/5$.

A probabilidade do acontecimento C é:

$$P(C) = \frac{2}{5} \times \frac{1}{4} = \frac{1}{10}.$$

Curiosidade: Conta-se que D'Alembert cometeu um erro de raciocínio com o seguinte problema:

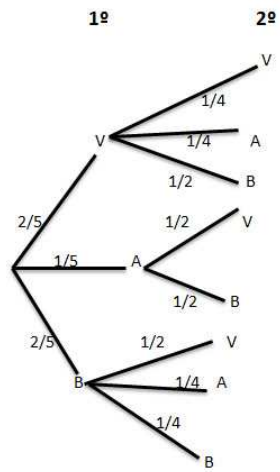


Figura 3.7: Diagrama em árvore

"Se lançarmos uma moeda ao ar duas vezes seguidas, qual é a probabilidade de obtermos pelo menos uma vez a face verso?"

D'Alembert respondeu que a probabilidade era 2 em 3 e justificou dizendo que havia três possibilidades: dois reversos, dois versos ou um reverso e um verso e só uma das possibilidades era desfavorável. Qual o erro do raciocínio D'Alembert?

D'Alembert contou mal os casos possíveis e os favoráveis da experiência.

Exemplo 3.15 Consideremos a experiência que consta do lançamento de uma moeda três vezes. Qual a probabilidade de não obter a mesma face duas vezes consecutivas?

Seja o acontecimento de interesse representado por C: "não obter a mesma face duas vezes consecutivas". Vamos ilustrar o espaço amostral através de um diagrama em árvore com a probabilidade do acontecimento no ramo.

O acontecimento C é a reunião de dois acontecimentos elementares $C1 = \{F1F2F1\}$ e $C2 = \{F2F1F2\}$, sendo

$$P(C1) = \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2}$$

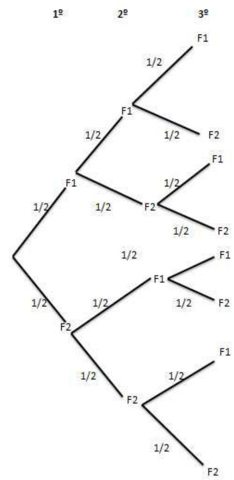


Figura 3.8: Diagrama em árvore

$$P(C2) = \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2}.$$

Então, $P(C) = P(C1) + P(C2) = \frac{1}{8} + \frac{1}{8} = \frac{1}{4}$.

3.6 Propriedades da probabilidade de Laplace

Do ponto de vista da corrente Laplaciana o conceito de probabilidade é um quociente em cujo denominador é o número de casos possíveis e o numerador é o número de casos favoráveis, podemos observar como exercício que esta definição de probabilidade verifica o seguinte conjunto de regras fundamentais:

- A probabilidade de um acontecimento $\overline{A}$ contrário (ou complementar) de A é $P(\overline{A}) = P(\Omega - A) = 1 - P(A)$, pois se houver k casos favoráveis a A em n possíveis há $n - k$ favoráveis a $\overline{A}$ em n possíveis. No caso de $A \equiv \Omega$ obtém-se $P(\varnothing) = 0$;

- Se $A_1, \cdots, A_n$ são acontecimentos disjuntos dois a dois, então é válida a regra da adição

$$P\left(\bigcup_{k=1}^{n} A_k\right) = \sum_{k=1}^{n} P(A_k),$$

porque, devido à disjunção, o número de casos favoráveis à união é a soma dos números de casos favoráveis a cada um dos acontecimentos A_k.

- Se $A \cap B$, então $P(A) \leq P(B)$ Consequentemente, como $\varnothing \cap A \subseteq \Omega$ segue-se que $0 \leq P(A) \leq 1$ Basta notar que o número de casos favoráveis a B não pode ser inferior ao número de casos favoráveis a A.

- A probabilidade de que se verifique A sem se verificar B é

$$P(A - B) = P(A) - P(A \cap B).$$

  Apenas há que excluir os casos favoráveis a B que eram favoráveis a A.

- $P(A \bigcup B) = P(A) + P(B) - P(A \bigcap B)$. Estamos apenas a eliminar a duplicação de casos favaráveis que se deve aos que são favoráveis simultaneamente a A e a B

- Se em n casos possíveis houver k_1 casos favoráveis ao acontecimento A, e k_2 casos favoráveis ao acontecimento B, e K favoráveis simultaneamente a A e a B, a probabilidade de $A \cap B$ é $\frac{K}{n}$. Se A não tiver influência sobre a realização de B, e vice-versa, então é natural admitir que a proporção de casos favoráveis a B que estão em A, $\frac{K}{k_1}$, é igual à proporção de casos favoráveis a B no universo, $\frac{k_2}{n}$.

  Deduz-se então $K = \frac{k_1 k_2}{n}$, e consequentemente a regra da multiplicação

$$P(A \cap B) = \frac{k_1}{n} \times \frac{k_2}{n} = P(A) \times P(B),$$

  se A e B não forem mutuamente informativos (dizemos que são acontecimentos independentes).

3.7 Definição axiomática de probabilidade

O russo Andrey Kolmogorov, em 1933, propôs a primeira definição formal da axiomática de probabilidade (caso finito) baseada em três axiomas.

Chama-se probabilidade a toda a aplicação P de domínio Ω e conjunto de chegada $\mathbb{R}_0^+$ tal que, a todo o acontecimento A é associado um número real maior ou igual que zero que se designa por probabilidade do acontecimento A.

$$\begin{aligned} P: \quad & \Omega \longrightarrow \mathbb{R}_0^+ \\ & A \longrightarrow P(A) \end{aligned}$$

1. A probabilidade do acontecimento certo é 1.

$$P(\Omega) = 1$$

2. A probabilidade de qualquer acontecimento A é não negativa.

$$P(A) \geq 0$$

3. Se A e B são acontecimentos incompatíveis, a probabilidade do acontecimento $A \cup B$ é a soma das probabilidades de A e de B.

$$A \cap B = \emptyset \Rightarrow P(A \cup B) \leq P(A) + P(B)$$

   pois o número de casos favoráveis a A ou a B não excede a soma do número de casos favoráveis a A com o número de casos favoráveis a B. A igualdade é verificada quando A e B forem disjuntos.

Teorema 1

Se A é um acontecimento impossível, então $P(A) = 0$.

Demonstração

Hipótese: $A = \{\}$

Tese: $P(A) = 0$

O espaço amostral Ω pode ser escrito como: $\Omega = \Omega \cup \{\}$.

Pelo axioma 1, tem-se $P(\Omega) = 1$, logo $P(\Omega \cup \{\}) = 1$. (i).

Mas como Ω e $\{\}$ são acontecimentos incompatíveis, então

pelo axioma 3 tem-se que $P(\Omega \cup \{\}) = P(\Omega) + P(\{\})$. (ii)

De (i) e (ii), resulta que $1 = P(\Omega) + P(\{\})$, ou seja, $1 = 1 + P(\{\})$, donde se conclui que $P(\{\}) = 0$. (c.q.d.)

Teorema 2

Se $\overline{A}$ é o acontecimento contrário do acontecimento A, então $P(\overline{A}) = 1 - P(A)$.

Demonstração

Hipótese: A e $\overline{A}$ são acontecimentos contrários.

Tese: $P(\overline{A}) = 1 - P(A)$

Sendo A e $\overline{A}$ acontecimentos contrários, tem-se que $\Omega = A \cup \overline{A}$, pelo axioma 1.

$P(\Omega) = P(A \cup \overline{A})$ e A e $\overline{A}$ são acontecimentos incompatíveis.

Pelos axiomas, tem-se:

$$1 = P(A) + P(\overline{A}) \Leftrightarrow P(\overline{A}) = 1 - P(A) \qquad \text{(c.q.d.)}$$

Teorema 3

Se A e B são acontecimentos tais que $B \subset A$, então $P(B) \leq P(A)$.

Demonstração

Hipótese: A e B são acontecimentos tais que $B \subset A$.

Tese: $P(B) \leq P(A)$

Se $B \subset A$, então existe C tal que $B \cap C = \{\}$ e $B \cup C = A$.

Por aplicação do axioma 3, tem-se:

$$P(B \cup C) = P(B) + P(C) = P(A)$$

Pelo axioma 2, sabe-se que $P(C) \geq 0$, então pode-se concluir que $P(B) \leq P(A)$. (c.q.d.)

Teorema 4

Para qualquer acontecimento A, tem-se $0 \leq P(A) \leq 1$.

Demonstração

Hipótese: A é um acontecimento.

Tese: $0 \leq P(A) \leq 1$.

Pelo axioma 2, sabe-se que $P(A) \geq 0$. (i)

Como $A \subset \Omega$, por aplicaçaõ do teorema 3 conclui-se que

$$P(A) \leq P(\Omega).$$

Pelo axioma 1, tem-se $P(\Omega) = 1$. Então, $P(A) \leq 1$. (ii)

De (i) e (ii), conclui-se que $0 \leq P(A) \leq 1$. (c.q.d.)

Teorema 5

Se A e B são dois acontecimentos compatíveis, então

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Demonstração

Hipótese: A e B são dois acontecimentos compatíveis.

Tese: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Os acontecimentos $A - B$ e $A \cap B$ são incompatíveis e

$$A = (A - B) \cup (A \cap B)$$

Pelo axioma 3, tem-se:

$$P(A) = P(A - B) + P(A \cap B).$$

Daqui resulta que $P(A - B) = P(A) - P(A \cap B)$. (i)

Os acontecimentos $A - B$ e B são incompatíveis e $A \cup B = (A - B) \cup B$.

Pelo axioma 3, tem-se:

$$P(A \cup B) = P(A - B) + P(B).$$

Daqui resulta que $P(A - B) = P(A \cup B) - P(B)$. (ii)

Comparando (i) e (ii), tem-se $P(A \cup B) - P(B) = P(A) - P(A \cap B)$.

Ou seja, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. (c.q.d.)

3.8 Probabilidade Condicionada e Independência

Suponha-se que A e B são acontecimentos associados a uma experiência aleatória e tais que $P(B) \neq 0$, chama-se probabilidade condicionada de A, dado B, e representa-se por $P(A|B)$, ao valor $\dfrac{P(A \cap B)}{P(B)}$.

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \quad P(B) \neq 0$$

A probabilidade $P(A|B)$ é designada por probabilidade condicionada de A, dado B, uma vez que a probabilidade de A está condicionada por uma informação adicional, o facto de se saber que o acontecimento B ocorreu. Da igualdade resulta que $P(A \cap B) = P(A|B)P(B)$.

Vejamos agora como se relacionam os conceitos de probabilidade e independência:

Num espaço amostral Ω, consideremos dois acontecimentos A e B, tais que $P(A) \neq 0$ e $P(B) \neq 0$.

Diz-se que o acontecimento A é independente do acontecimento B se $P(A|B) = P(A)$ ou $P(B|A) = P(B)$. Reparemos que se $P(A|B) = P(A)$, significa que o facto de termos conhecimento de que B ocorreu não influência a probabilidade de A. Por outro lado, usando a definições de probabilidade condicionada podemos escrever que:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

Se os acontecimentos A e B são independentes então, $P(A|B) = P(A)$

donde resulta que:

$$P(A) = P(A|B) = \frac{P(A \cap B)}{P(B)} \Leftrightarrow P(A \cap B) = P(A) \times P(B).$$

Concluímos que se A e B são independentes então

$$P(A \cap B) = P(A) \times P(B).$$

Exemplo 3.16 Mostre que se A e B são acontecimentos independentes, então A e $\overline{B}$ também o são.

Resolução:

Se A e B são acontecimentos independentes $\Rightarrow P(A \cap B) = P(A) \times P(B)$. Por outro lado observamos que

$A = A \cap \Omega = A \cap (B \cup \overline{B}) = (A \cap B) \cup (A \cap \overline{B})$

Aplicando a probabilidade, obtemos:

$P(A) = P\left((A \cap B) \cup (A \cap \overline{B})\right)$

como os acontecimentos $A \cap B$ e $A \cap \overline{B}$ são disjuntos resulta,

$P(A) = P(A \cap B) + P(A \cap \overline{B})$

$P(A) = P(A) \times P(B) + P(A \cap \overline{B})$; A e B são acontecimentos independentes

$P(A \cap \overline{B}) = P(A) - P(A) \times P(B)$

$P(A \cap \overline{B}) = P(A)[1 - P(B)]$

$P(A \cap \overline{B}) = P(A)P(\overline{B})$

Concluimos então que se os acontecimentos A e B são independentes então também os acontecimentos A e $\overline{B}$ são independentes.

Exemplo 3.17 Uma urna contém 10 bolas vermelhas (V) e 5 bolas brancas (B). Calcule a probabilidade de

1. sair uma bola vermelha seguida de uma bola branca, em extrações sem reposição.
2. sair uma bola vermelha seguida de uma bola branca, em extrações com reposição.

Resolução

Consideremos os acontecimentos V: "sair uma bola vermelha" e B: "sair uma bola branca".

1. A probabilidade de sair uma bola vermelha é

$$P(V) = \frac{10}{15}$$

   Na segunda extração pretende-se calcular a probabilidade de sair uma bola branca sabendo que na primeira extração saiu uma bola vermelha, e o espaço amostral alterou-se visto que a extração foi realizada sem reposição, sendo o número total de bolas na urna igual a 14.

$$P(B|V) = \frac{5}{14}$$

   Pela propriedade

$$P(V \cap B) = P(V) \times P(B|V)$$

   Logo $P(V \cap B) = P(V) \times P(B|V) = \frac{10}{15} \times \frac{5}{14} = 0.238$

2. A probabilidade de sair uma bola vermelha é $P(V) = \frac{10}{15}$. Na segunda extração pretende-se calcular a probabilidade de sair uma bola branca sabendo que na primeira extração saiu uma bola vermelha, sendo a primeira extração realizada com reposição, equivale a dizer que o espaço amostral não se alterou mantendo-se um total de 15 bolas na urna
$$P(V \cap B) = P(V) \times P(B) = \frac{10}{15} \times \frac{5}{15} = \frac{2}{9}$$

Exemplo 3.18 No quadro seguinte está representada a situação ao nível de emprego e sexo dos habitantes (adultos) de um subdistrito do concelho de Matatuto de Timor-Leste.

1. Seleciona-se, ao caso, um dos habitantes:
   (a) Qual a probabilidade de ser mulher?
   (b) Qual a probabilidade de estar desempregado?
   (c) Qual a probabilidade de ser mulher e desemprego?
2. Seleciona-se, ao caso, um dos habitantes e verifica-se que é mulher. Qual a probabilidade de estar desempregada?
3. Seleciona-se, ao caso, um dos habitantes e verifica-se que é desempregado. Qual a probabilidade de ser mulher?

Tabela 3.4: Situação de emprego por sexo dos habitantes (adultos)

	$N^{\underline{O}}$ empregados	$N^{\underline{O}}$ desempregados	Total
Homens	900	100	1000
Mulheres	820	910	1730
Total	1720	1010	2730

Resolução

Consideremos os acontecimentos, M: ”ser mulher”, H: ”ser homem” e D: ”estar desempregado”:

1. (a) $P(M) = \frac{1730}{2730} = 0.63370$
   (b) $P(D) = \frac{191}{2730} = 0.06996$
   (c) $P(M \cap D) = \frac{91}{2730} = 0.03333$
2. $P(D \mid M) = \frac{P(D \cap M)}{P(M)} = \frac{910}{1730} = 0.52601$

3. $P(M \mid D) = \dfrac{P(M \mid D)}{P(D)} = \dfrac{910}{1010} = 0.90099.$

Capítulo 4

Estatística Descritiva

4.1 Introdução

A Estatística é atualmente muito importante dado que é uma metodologia usada em muitas áreas científicas e necessária na tomada de decisões. A Estatística é um ramo de Matemática aplicada que trata da recolha, organização, análise e apresentação de uma forma útil da informação contida nos dados. A estatística é a ciência que trata da obtenção de informação recorrendo a técnicas de amostragem e planeamento de experiências, de modo a assegurar uma recolha de dados com uma correta qualidade de informação. No tratamento inicial dos dados são usadas técnicas como: a ordenação, o cálculo das característica amostrais, o agrupamento em classes (se o número de dados o justificar), representações gráficas informativas de uma população. Esta parte da estatística é designada como estatística descritiva e análise exploratória de dados. Por sua vez a inferência estatística permite também inferir a partir de uma amostra as características de uma população (estimação de parâmetros populacionais a partir das características amostrais, decisão sobre hipóteses, comparação de populações, relacionamento de uma variável resposta com variáveis controladas). Para finalizar, não podemos deixar de referir que também compete à Estatística num contexto de incerteza e variabilidade a tomada de decisões estratégicas. A elaboração deste capítulo foi orientada com base em Athayde (2013), Martins et al. (1997), Murteira et al. (2010) e Pestana & Velosa (2010). A seguir faz-se uma breve introdução aos conceitos de recenseamento, sondagem, população, amostra e tipos de amosragem como introdução à Estatística.

Recenseamento e Sondagem

Na linguagem da Estatística os termos recenseamento e sondagem são termos muitos comuns, mas aplicam-se em contextos diferentes. A palavra recenseamento está associada à contagem oficial e periódica dos indivíduos de um país. O recenseamento permite conhecer diferentes características da população, nomeadamente situação civil, habitacional, rendimentos, classes etárias, mortalidade, natalidade, estudar atitudes e muitos outros aspetos da vida e hábitos dessa população, tudo que é informação relevante para que os governantes desse país possam tomar decisões em relação às áreas da saúde, educação, habitação, etc. A sondagem analisa apenas uma parte de uma população em estudo com o objetivo de generalizar as conclusões estatísticas a todos os elementos da população.

O recenseamento das populações, os inquéritos sobre a produção anual de produtos essenciais como o trigo e a recolha de dados para fins militares constituíram as primeiras aplicações das técnicas estatísticas nas civilizações mais relevantes como a chinesa, a egípcia, a assíria e a grega.

População e Amostra

Ao grupo de todos os elementos que se pretende estudar e que possuem uma ou mais características em comum chama-se **população**. O termo população não significa população humana, mas sim uma coleção de entidades, que podem ser pessoas, animais, resultados experimentais, todos com uma ou mais características em comum, que se pretende analisar. Altura dos alunos, cor dos olhos, números de irmãos de cada aluno são exemplos de características da população que pode haver interesse em estudar.

A **amostra** é um subconjunto da população que se analisa com o objectivo de tirar conclusões para a população de onde foi recolhida. A validade dessas conclusões depende da qualidade da amostra e portanto do processo usado para construir a amostra. A amostragem é a área da estatística se ocupa das metodologias necessárias para a qualidade (representatividade) de uma amostra. As amostragens mais usuais são:

Amostragem aleatória simples: cada indivíduo da população tem a mesma probabilidade de ser escolhido;

Amostragem aleatória sistemática: trata-se de escolher os elementos da amostra por uma regra previamente definida;

Amostragem estratificada: consiste em considerar a população dividida em pequenos grupos ou estratos, pelo que a escolha da amostra requer um número de elementos de cada estrato proporcional à dimensão do grupo.

4.2 Estatística descritiva

Estatística descritiva trata e calcula um conjunto de medidas que tem por objetivo descrever e resumir a informação subjacente aos dados. Ela tem por finalidade descrever certas propriedades relativas de uma amostra ou um população. Mas quando realizamos uma sondagem para além de organizar os dados e descrever as características da amostra, inferimos a propriedades para toda população. O processo ou tipos de representação os dados na estatística descritiva através de métodos numéricos (envolvendo apresentação de medidas de posição ou dispersão) método gráfico (envolvendo gráfico ou tabular). A importância de tabelas fornecem uma ideia mais precisa e possibilitam uma inspecção mais rigoroso aos dados e os gráficos são mais indicados em situações que objectivam de uma visão mais rápida e fácil respeita das variações as quais se referem os dados (constituem uma das formas mais eficientes de representação de dados). Uma forte análise das estatísticas descritivas dos dados fornece os alicerces para uma correta estatística indutiva (identificar características da população a partir das características amostrais). A elaboração desta seção baseou-se no programa do 12º ano atualmente em vigor em Timor-Leste.

Atributos estatísticos

Designam-se por atributos todas as características da população ou amostra que são objeto de estudo.

Um **atributo qualitativo** é uma qualidade, uma característica da população não mensurável e que vai ser objeto de estudo.

Um **atributo quantitativo** é uma característica da população que é mensurável e que vai ser objeto de estudo.

Quando um atributo é mensurável é designado por variável estatística. A variável estatística deve estar definida de um modo preciso e claro. A variável estatística classifica-se como discreta se assume apenas um número finito de valores em qualquer intervalo limitado. Caso contrário é uma variável contínua e assume qualquer valor no seu intervalo de variação (estas definições são retomadas no capítulo 6).

Organização de dados

Os valores observados que formam a amostra chamam-se dados estatísticos. Uma boa organização dos dados permite uma maior facilidade na leitura e interpretação da informação. A contrução de tabelas, gráficos e diagrama permite evdenciar as características dos dados.

Tabelas de frequências

Depois de recolhida a amostra é importante organizar os dados em tabelas de modo a facilitar a visão global da informação amostral, leitura e interpretação.

Distribuição Estatística

Exemplo 4.1 Na turma A do 12º ano, da Escola Secundária 28 de Novembro, em Timor-Leste, realizou-se um estudo sobre as alturas (em cm) dos alunos dessa turma. Os dados obtidos estão registados na seguinte tabela:

Tabela 4.1: Alturas dos alunos da turma A do 12º Ano

145	140	167	165	150	145	167	150	140	165	168	140
150	160	168	167	145	165	160	167	145	150	168	160

A variável estatística em estudo é X: ”altura de um aluno da turma A do 12º ano”, sendo a população em estudo a turma A do 12º ano. Neste exemplo os valores que a variável estatística assume são: 140, 145, 150, 160, 165, 167 e 168.

Temos uma **distribuição estatística** sempre que o valor da variável estatística é conhecido para cada elemento da população ou da amostra em estudo. Designando por X a variável estatística, os diferentes valores que a variável assume são representados por $x_1, x_2, x_3, \cdots$.

No caso da variável estatística X ser quantitativa e assumir apenas um número limitado de valores $x_1, x_2, \cdots, x_n$ que devem ser ordenados por ordem crescente

$$x_1 < x_2 < \cdots < x_n$$

No exemplo 4.1 temos:

$$x_1 = 140,\ x_2 = 145,\ x_3 = 150,\ x_4 = 160,\ x_5 = 165,\ x_6 = 167,\ x_7 = 168.$$

Frequência Absoluta

A frequência absoluta de um dado estatístico, representa-se por f_i, e é igual ao número de vezes que esse valor se repete na amostra. Quando organizamos os dados numa tabela, na coluna da esquerda colocamos os diferentes valores x_i que a variável em estudo pode tomar. Vamos exemplificar com caso de estudo.

Frequência Relativa

Frequência relativa de um dado estatístico é o quociente entre a frequência absoluta e o número total de observações e representa-se por fr_i onde

$$fr_i = \frac{f_i}{n}.$$

Nota: Ao multiplicar a frequência relativa por 100 esta aparece expressa em termos de percentagem.

Exemplo 4.2 Os alunos da turma B do 12º ano, da Escola Secundária 28 de Novembro, foram inquiridos relativamente às suas preferências a nível da Gastronomia de Timor-Leste. Os resultados obtidos foram os seguintes:

Tabela 4.2: Preferência Gastronómica

Tipo de gastronomia	Freq. absoluta
Tukir de Cabrito	3
Kadaca	8
Manu Salar	10
Singa de Camarão	9

Frequência Acumulada

Existem dois tipos de frequência acumulada, a frequência absoluta acumulada (designada por F_i , e cujo valor se obtém adicionando as frequências absolutas até ao valor considerado da variável estatística) e frequência relativa acumulada (designada por Fr_i, e cujo valor se obtém adicionando as frequências relativas até ao valor considerado da variável estatística).

Tabela 4.3: Frequências simples e acumuladas

x_i	f_i	fr_i	F_i	Fr_i
Tukir de Cabrito	3	0.10	3	0.10
Kadaca	8	0.27	11	0.37
Manu Salar	10	0.33	21	0.70
Singa de Camarão	9	0.30	30	1.00

Função Cumulativa

No caso da frequência absoluta acumulada (função cumulativa das frequências absolutas), esta função faz corresponder a cada valor de x_i o número total de dados observados com valor menor ou igual a x_i. No caso da frequência relativa acumulada (função das frequências relativas), esta função faz corresponder a cada valor de x_i a frequência relativa do total de dados observados com valor menor ou igual a x_i.

Exemplo 4.3 Uma pesquisa de Saúde Pública investigou o número de filhos em 48 casais para analizar a evolução da natalidade no distrito Viqueque. Na tabela estão registados os valores obtidos para a variável X : ”número de filhos por casal”:

1	2	3	2	1	2	1	2	1	2	1	3	2	1	3	1
1	4	3	2	3	1	3	2	3	1	2	3	2	0	4	2
2	1	2	4	3	2	1	3	4	2	1	3	2	4	2	1

Vamos determinar as frequências absolutas e relativas (simples e acumuladas) para os dados da tabela:

x_i	f_i	fr_i	F_i	Fr_i
0	1	0.021	1	0.021
1	14	0.292	15	0.313
2	17	0.354	32	0.667
3	11	0.229	43	0.896
4	5	0.104	48	1.000

Para os dados do exemplo 4.3, a função cumulativa das frequências relativas define-se da seguinte forma:

$$F(x) = \begin{cases} 0.000 & \text{se} \quad x < 0 \\ 0.021 & \text{se} \quad 0 \leq x < 1 \\ 0.313 & \text{se} \quad 1 \leq x < 2 \\ 0.667 & \text{se} \quad 2 \leq x < 3 \\ 0.896 & \text{se} \quad 3 \leq x < 4 \\ 1.000 & \text{se} \quad x \geq 4 \end{cases}$$

Agrupamentos de dados em classes

Quando a variável estatística (contínua ou discreta) pode tomar uma grande diversidade de valores, então procede-se ao agrupamento dos dados em classes (intervalos). As classes têm que verificar as seguintes condições:

1. Serem disjuntas;
2. Incluirem todos os valores possíveis da variável que está em estudo.

À diferença entre o extremo superior e o inferior chama-se amplitude da classe. Ao ponto médio de cada classe damos o nome de marca da classe e representa-se por x_m. Num intervalo do tipo $[a, b[$ a marca da classe obtém-se do seguinte modo: $x_m = \dfrac{a+b}{2}$.

Existem algumas regras que é necessário ter em conta na formação das classes:

1. Todas devem ter a mesma amplitude;
2. Não se devem sobrepor para que cada dado pertença exatamente a uma e só uma classe;
3. O limite superior de uma classe deve coincidir com o limite inferior da seguinte;
4. O valor mínimo da amostra deve pertencer à primeira classe e o máximo à última;
5. O número de classes é obtido utilizando a seguinte regra: para uma amostra de dimensão n o número de classes k é o menor número inteiro tal que: $2^k \geq n$.

Exemplo 4.4 Numa prova de aptidão para acesso a uma empresa pública, os 40 candidatos tiveram as seguintes classificações na prova (de escala 0 a 100):

75	70	75	60	65	60	45	55	75	70
60	65	60	55	75	65	65	80	75	83
80	75	65	65	75	80	65	65	75	65
80	65	70	75	75	65	83	83	65	75

Resolução

Para este conjunto de dados temos:

1. A dimensão da amostra é 40;
2. O valor máximo é 83;
3. O valor mínimo é 45;
4. O número de classes k, onde $2^k \geq n \Leftrightarrow 2^k \geq 40 \Leftrightarrow 2^6 \cong 40 \Leftrightarrow k \cong 6$;
5. Amplitude das classes: $\dfrac{83-45}{6} = \dfrac{38}{6} = 6.33 \cong 6$.

No exemplo, os dados são agrupados em 6 classes e vamos calcular os pontos médios das classes, as frequências simples (absolutas e relativas) e as frequências acumuladas (absolutas e relativas) para elaborar uma tabela com todas as frequências.

4.3 Representações gráficas

A representação gráfica de conjuntos alargados ou não de dados é um meio eficaz, prático, simples, preciso e apelativo de transmitir informação.

Classes	x_m	f_i	fr_i	F_i	Fr_i
[44;50[	47	1	0.025	1	0.025
[50;56[	53	2	0.05	3	0.075
[56;62[	59	4	0.100	7	0.175
[62;68[	65	12	0.300	19	0.475
[68;74[	71	3	0.075	22	0.550
[74;80[	77	11	0.275	33	0.825
[80;86[	83	7	0.157	40	1

Tabela 4.4: Distribuição dos dados em classes

De entre os métodos gráficos usados para representar um conjunto de dados, os principais são o diagrama de linhas, o diagrama de barras, o diagrama circular e o histograma. O diagrama de linhas é usado para dados de natureza qualitativa ou quantitativa discreta, com um número pequeno de valores distintos. A altura de cada linha deverá ser proporcional à frequência que lhe corresponde. No R, a organização dos dados em tabelas de frequências absolutas/frequências relativas é feita pelo comando `table(x)` e `table(x)/length(x)` e os diagramas de linhas correspondentes são construídos usandos os comandos `plot(table(x))` e `plot(table(x)/length(x))`. No caso de uma variável qualitativa também se pode usar o diagrama de barras (idêntico ao diagrama de linhas) e o diagrama circular. O diagrama circular é constituída por um círculo, em que se apresentam vários setores circulares, tantos quanto as classes (categorias) consideradas na tabela de frequências da amostra em estudo. Os ângulos dos setores são proporcionais às frequências relativas das classes (categorias). No R, o diagrama de barras é obtido com a instrução `barplot(table(x))` e o diagrama circular com `pie(table(x))`.O histograma é uma representação gráfica (um gráfico de barras verticais ou barras horizontais) da distribuição de frequências de um conjunto de dados quantitativos contínuos ou discretos com muitos valores. O histograma pode ser um gráfico por frequências absolutas ou frequências relativas. No caso de densidade, a frequência relativa do intervalo i, (fr_i), é representada pela área de um retângulo que é colocado acima do ponto médio da classe i. Consequentemente, a área total do histograma (igual a soma das áreas de todos os retângulos) será igual a 1. Assim, ao construir o histograma, cada retângulo deverá ter área proporcional à frequência relativa (ou à frequência absoluta, o que é indiferente) correspondente. No caso em que os intervalos têm amplitudes iguais, as alturas dos retângulos serão iguais às frequências relativas (ou iguais às frequências absolutas) dos intervalos correspondentes. No R os histogramas são obtidos pelo comando `hist(x)`. Por defeito, as classes têm a mesma amplitude e a altura dos retângulos é a frequência. O histograma terá uma área igual a 1, se for acrescentada a opção `freq=F`. Vamos ilustrar a construção dos gráficos diagrama de barras e diagrama de circular, com os dados do exemplo 4.2:

Diagrama de Barras

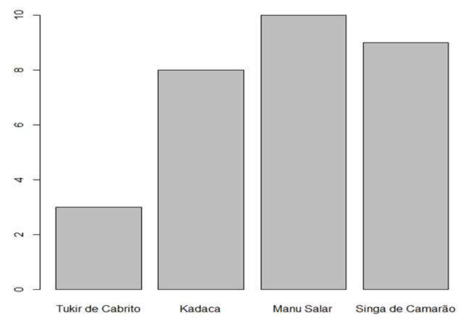


Figura 4.1: Diagrama de barras para X: "Preferências gastronómicas"

Resolução em R

```
>GTl=c(3,8,10,9)
> gtl=c(rep(1,3),rep(2,8),rep(3,10),rep(4,9))
> gtl.d=table(gtl)
> names(gtl.d)=c("Tukir de Cabrito","Kadaca","Manu Salar","Singa de Camar\~{a}o")
> par(mfrow=c(1,2))
>barplot(gtl.d)
```

Diagrama de linhas

Para os dados do exemplo 4.3, realizou-se uma representação gráfica em diagrama de linhas.

Resolução em R

```
> x<-scan()
1: 1 2 3 2 1 2 1 2 1 2 1 3 2 1 3 1 2 1 4 3 2 3 1 3
2: 2 1 2 4 3 2 1 3 4 2 1 3 2 4 2 1 3 1 2 3 2 0 4 2
49:
Read 48 items
> table(x)
x
```

```
0  1  2  3  4
 1 14 17 11  5
> freq<-table(x)
>plot(freq, type="h", xlab="Numeros de filhos")
```

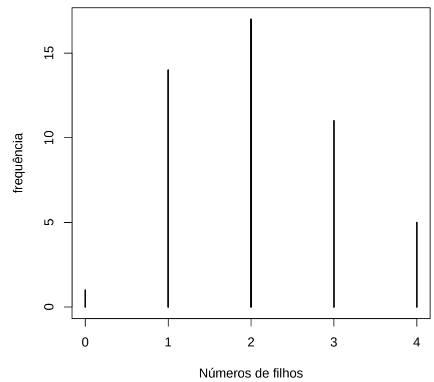


Figura 4.2: Diagrama de linhas para X: ”Número de filhos”.

Diagrama Circular

Para construir o diagrama circular manualmente, precisamos de calcular a informação da tabela.

Tabela 4.5: Tabela das frequências de Gastronomia

x_i	f_i	fr_i	$fr_i \times 360^o$
Tukir de Cabrito	3	0.10	36^o
Kadaca	8	0.27	96^o
Manu Salar	10	0.33	120^o
Singa de Camarão	9	0.30	108^o
Total	30	1.00	360^o

A função `pie()` permite obter a representação do diagrama circular.

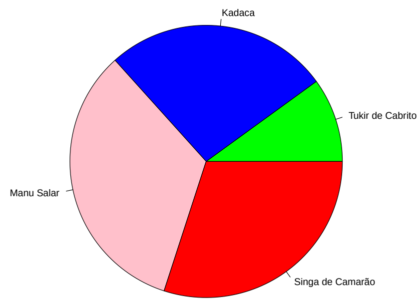


Figura 4.3: Diagrama circular

Resolução em R

```
>GTl=c(3,8,10,9)
> gtl=c(rep(1,3),rep(2,8),rep(3,10),rep(4,9))
> gtl.d=table(gtl)
> names(gtl.d)=c("Tukir de Cabrito","Kadaca","manu Salar","Singa de Camarao")
> par(mfrow=c(1,2))
>pie(gtl.d, radius=1.2, col=c("green","blue","pink","red"))
```

Histograma

Para o exemplo 4.4, obtemos o seguinte histograma:

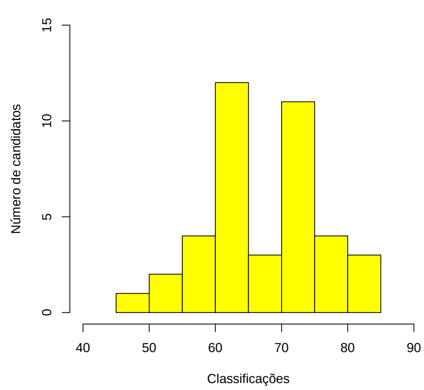


Figura 4.4: Histograma para X: ”Classificação no teste de aptidão”

Resolução em R

```
>RE=c(75, 70, 75, 60, 65, 60, 45, 55, 75, 70, 60, 65, 60, 55, 75, 65, 65, 80, 75, 85,
>hist(RE,right = T, xlab="Classificacoes", ylab="Numero de candidatos", xlim=c(40,90)
```

Observação: Um tipo de gráfico diferente é o pictograma que é uma ilustração que usa símbolos sugestivos da variável (ou variáveis) em estudo.

Polígono de Frequências

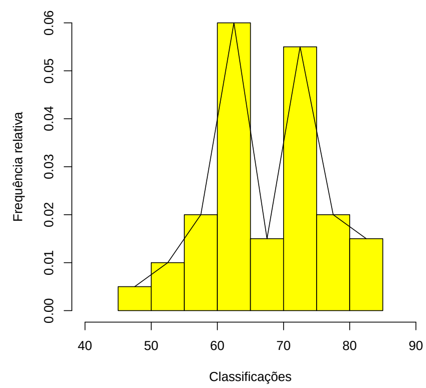


Figura 4.5: Histograma e polígono de frequências relativas para X: ”Classificação no teste de aptidão”

Resolução em R

```
> points(h$mids, h$density, "l")
```

O polígono de frequências é um gráfico de linhas em que no eixo vertical pode-se utilizar as frequências absolutas ou frequências relativas e no eixo horizontal o ponto médio de cada classe. A linha é construída unindo-se os pontos de coordenadas: as abscissas correspondentes aos pontos médios de cada classe e as ordenadas às frequências absolutas ou relativas dessas mesmas classes.

Diagrama de caule-e-folhas

No gráfico de caule-e-folhas os dados estão colocados de um modo ordenado em linhas horizontais, apresentando uma visualização idêntica à de um histograma. Consideremos um conjunto de dados, em que os registos são formados por dois algarismos, o algarismo das dezenas será designado como caule e o das unidades como folha. As folhas são colocadas em linhas horizontais à frente do respetivo caule. O diagrama deve apresentar tantas folhas quantos os dados da amostra, permitindo vizualizar a forma da distribuição dos dados (Nota: esta é uma apresentação simplista, para maior detalhe de construção deste tipo de diagrama consultar Pestana & Velosa (2010)).

```
3 | 0
4 | 2 9
5 | 0 3 5 6 6 7 8 8 9 9
6 | 0 0 1 2 2 3 5 5 5 5 5 5 6 6 7 7 7 8
7 | 0 1 3 4 4 4 5 5 5 6 6 6 6 6 7 7 8 8
8 | 0 0 1 3 4 4 4 4 5 6 7 8 9 9
9 | 0 0 2 3 5 8 9
```

Como a amostra está ordenada no diagrama este tipo gráfico é bastante útil no cáculo de medidas estatísticas que envolvam ordens (posições) dos dados. O comando em R para executar o gráfico é `stem()`.

4.4 Medidas de localização

Medida de localização ou de tendência central é uma grandeza numérica cujo valor referência a posição de um conjunto de dados numa escala de valores possíveis. As medidas de localização mais usadas são a média, a moda e a mediana. Pretendemos estudar o efeito de alteração dos dados a nível destas medidas descritivas.

Média

A média é uma medida de localização e representa o ponto de equilíbrio de um conjunto de dados. A notação é $\overline{x}$ e define-se para um conjunto de dados, x_1, x_2,$\cdots$,x_n, como:

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Nota: Quando nos referimos à média estamos a falar da média aritmética, onde todos os valores são igualmente ponderados. Existem médias que aplicam ponderações diferentes aos valores amostrais.

Dada uma amostra de dimensão n onde cada variável x_i toma k valores diferentes, sendo f_i e fr_i a frequência absoluta e relativa respetivamente do valor x_i, tem-se:

$$\overline{x} = \frac{f_1 \times x_1 + f_2 \times x_2 + \cdots + f_k \times x_k}{n} = \frac{\sum_{i=1}^{k} f_i \times x_i}{n}.$$

$$\overline{x} = fr_1 \times x_1 + fr_2 \times x_2 + \cdots + fr_k \times x_k = \sum_{i=1}^{k} fr_i \times x_i.$$

Propriedades da Média

Propriedade 1 Seja x_1, x_2,$\cdots$,x_n uma amostra cuja média é $\overline{x}$. Adicionando uma constante k a todos os dados observados, a nova amostra passa a ser $x_1 + k, x_2 + k, \cdots, x_n + k = x'_1, x'_2, \cdots, x'_n$ sendo a nova média igual a:

$$\overline{x'} = \overline{x} + k, \quad k \in \mathbb{R}.$$

Demonstração

A média dos novos dados é $\overline{x'}$, então:

$$\begin{aligned}
\overline{x'} &= \frac{\sum_{i=1}^{n}(x_i + k)}{n} \\
&= \frac{\sum_{i=1}^{n} x_i + k \times n}{n} \\
&= \frac{\sum_{i=1}^{n} x_i}{n} + \frac{k \times n}{n} \\
&= \overline{x} + k
\end{aligned}$$

Exemplo 4.5 Num restaurante, os valores pagos por um cliente (em dólares) pelas refeições de uma semana (7 dias) foram os seguintes: 10 10,50 11 12 12,50 13 13.

A média do preço diário por refeição foi de 11,71 dólares.

Se ao preço inicial da refeição, lhe acrescentarmos o consumo diário de uma bebida e uma sobremesa cujo preço é de 9 dólares, então o cliente pagaria na semana os seguintes montantes: 19 19,50 20 21 21,50 22 22 Cuja média é 20,71 dólares.

Repare que se verifica a propriedade referida, a média final é a primeira média adicionada do valor que aumentou o custo da refeição, neste caso de 9 dólares.

Propriedade 2 Seja x_1, x_2,$\cdots$,x_n uma amostra cuja média é $\overline{x}$. Multiplicando uma constante k a todos os dados observados, a nova amostra passa a ser $x_1 \times k, x_2 \times k, \cdots, x_n \times k = x_1^{'}, x_2^{'}, \cdots, x_n^{'}$ sendo a nova média igual a:

$$\overline{x'} = \overline{x} \times k, \quad k \in \mathbb{R} \setminus \{0\}.$$

Demonstração

Sendo a média dos novos dados $\overline{x'}$, então:

$$\begin{aligned}
\overline{x'} &= \frac{\sum_{i=1}^{n}(x_i \times k)}{n} \\
&= \frac{k \times \sum_{i=1}^{n} x_i}{n} \\
&= k \times \frac{\sum_{i=1}^{n} x_i}{n} \\
&= \overline{x} \times k
\end{aligned}$$

Exemplo 4.6 A Escola Secundária 28 de Novembro no fim do ano letivo tem por norma distribuir prémios de criatividade pelos estudantes. No ano letivo de 2013 os resultados foram registados na tabela seguinte:

A média dos valores dos prémios é de 64,58 dólares.

Tabela 4.6: Tabela de Prémios

Prémio (em dólares)	Números dos Alunos
150	1
125	1
100	1
75	2
50	3
25	4

Tabela 4.7: Tabela dos novos Prémios

Prémio (em dólares)	Números dos Alunos
300	1
250	1
200	1
150	2
100	3
50	4

No ano letivo seguinte o valor dos prémios duplicou, mantendo-se o mesmo número de alunos por categoria de prémio.
A média é 129,20 dólares.
Repare que se verifica a propriedade referida, a média de 2014 é igual à média de 2013 multiplicada por 2.

Propriedade 3 Seja x_1, x_2,$\cdots$,x_n uma amostra cuja média é $\overline{x}$. Considere que $d_i = x_i - \overline{x}$ representa o i-ésimo desvio, então para $i = 1, \cdots, n$ temos que $\sum_{i=1}^{n} d_i = 0$.

Demonstração

$$\begin{aligned}
\sum_{i=1}^{n} d_i &= \sum_{i=1}^{n} (x_i - \overline{x}) \\
&= \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \overline{x} \\
&= n \times \frac{\sum_{i=1}^{n} x_i}{n} - \sum_{i=1}^{n} \overline{x} \\
&= n \times \overline{x} - n \times \overline{x} = 0
\end{aligned}$$

Moda

A moda de um conjunto de dados é o valor que aparece mais vezes, ou seja, é aquele que apresenta a maior frequência observada. Há situações nas quais ela não é única, pois pode acontecer de se ter, em uma série estatística, duas ou mais observações que tenham-se destacado de forma idêntica, isto é, que tenham ocorrido com a mesma frequência máxima. Então, conforme o caso, teremos distribuições bimodais (duas modas) ou multimodais (multimodal). Também é possível acontecer que todos os elementos tenham apresentado exactamente o mesmo número de ocorrências. Isso significa que não há moda, pois nenhum dado se destacou; o conjunto é, então, chamado amodal. No caso de os dados estarem agrupados em classes, à classe com maior frequência absoluta dá-se o nome de classe modal e vamos considerar, nesse caso, a moda como o ponto médio da classe. A notação usual para a moda é M_o. No R, esta medida estatística não se encontra implementada.

Mediana

A mediana de um conjunto com n observações, é o valor que ocupa a posição central da distribuição ordenada (por ordem crescente ou decrescente). Trata-se portanto de uma medida de localização e representa-se por M_e ou $\tilde{X}$. Na escolha do valor central há que ter em conta o seguinte:

- Se n é ímpar, a mediana é o elemento que ocupa a posição $\dfrac{n+1}{2}$.

  Ou seja a mediana é: $m_e = x_{\frac{n+1}{2}}$.

- Se n é par, existem dois valores no meio, sendo a mediana a semissoma dos elementos que ocupam as posições $\dfrac{n}{2}$ e $\dfrac{n}{2}+1$.

  Ou seja a mediana é: $m_e = \dfrac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2}$.

Exemplo 4.7 Numa aula de matemática, das turmas I e II do 10º do programa Ciência e Tecnologia da Escola Secundária 28 de Novembro, o professor fez o registo da altura dos alunos. Os resultados (em cm) obtidos foram os seguintes:

Turma I
139, 143, 142, 155, 145, 138, 158, 159, 163, 162, 142, 140, 140, 166, 139, 148, 151, 149, 168, 156, 142, 154, 137, 148, 167, 145, 154, 149, 152, 143, 151, 150, 153, 146, 147, 150, 139, 144, 153, 142, 164, 146, 149, 153, 159, 147, 164, 157, 145, 163, 155.

Turma II
164, 161, 142, 156, 143, 138, 144, 139, 160, 156, 150, 146, 161, 144, 140, 158, 157, 149, 165, 157, 150, 145, 155, 158, 147, 137, 154, 159, 152, 143, 164, 151, 153, 166, 147, 151, 140, 154, 157, 162, 164, 146, 149, 153, 152, 167, 154, 146.

Determine a altura mediana dos alunos de cada uma das turmas.

Resolução

Vamos utilizar um diagrama de caule-de-folha com caules comuns para ordenar em simultâneo os dados das duas turmas. Do lado direito são colocados os dados da turma I e do lado esquerdo os dados da turma II:

9 8 7	13	7 8 9 9 9
4 4 3 3 2 0 0	14	0 0 2 2 2 2 3 3 4
9 9 7 7 6 6 6 5	14	5 5 5 6 6 7 7 8 8 9 9 9
4 4 4 3 3 2 2 1 1 0 0	15	0 0 1 1 2 3 3 3 4 4
9 8 8 7 7 7 6 6 5	15	5 5 6 7 8 9 9
4 4 4 2 1 1 0	16	2 3 3 4 4
7 6 5	16	6 7 8

Como a turma I tem um número ímpar de alunos, $n = 51$, então a mediana é dada por:

$$m_e = x_{\frac{51+1}{2}} = x_{\frac{52}{2}} = x_{26} = 149$$

Resolução em R

```
> median(T1)
[1] 149
```

Na turma II há um número par alunos, $n = 48$, então a mediana é dada por:

$$m_e = \frac{x_{24} + x_{25}}{2} = \frac{152 + 153}{2} = 152.5$$

Resolução com R

```
> median(T2)
[1] 152.5
```

As alturas medianas são 149 cm e 152.5 cm respetivamente para a turma I e II.

Quando os dados estão agrupados em classes, identifica-se a classe mediana e localiza-se graficamente a mediana. A interpretação geométrica da mediana para dados agrupados em classes é o valor do eixo das abcissas tal que a reta vertical que passa nesse ponto divide a área do histograma em duas regiões com áreas iguais. Uma aproximação para o valor da mediana, pode ser obtida fazendo o gráfico da função cumulativa e determinando o ponto de interseção da reta $y = 0.5$ com o gráfico. A abcissa do ponto corresponde ao valor aproximado da mediana.

Nota: A amostra ordenada por ordem crescente é usual ser representada como

$$x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}$$

onde $x_{(1)}$ e $x_{(n)}$ representam o mínimo e o máximo da amostra respetivamente. O uso desta notação sugere que as ordens das observações usadas no cálculo da mediana ou outras medidas estatísticas que envolvam ordens estejam entre parênteses curvos.

Quartis

Dada uma amostra é importante conhecer como os dados da amostra se distribuiem no seu intervalo de valores. Os quartis são valores obtidos a partir da amostra ordenada em ordem crescente, que dividem a distribuição dos dados em quatro partes iguais. O primeiro quartil, Q_1, é o número que indica que 25 por cento das observações são menores ou iguais e 75 por cento das observações são maiores ou iguais. O terceiro quartil, Q_3, indica que 75 são das observações são inferiores ou iguais ao seu valor e 25 por cento são maiores ou iguais. A mediana, Q_2 indica que 50 por cento das observações são menores ou iguais e 50 por cento da amostra apresenta valores maiores ou iguais ao valor da mediana. As definições apresentadas para os quartis são as dadas no manual do 12º adoptado em Timor-Leste. Como no caso da mediana existem fórmulas para determinar os quartis, também aqui vamos apresentar as respetivas expressões de cálculo para o Q_1 e Q_3.

Tabela 4.8: Tabela dos quartis para n par

Localização	Quartil
$k = \frac{n+2}{2}$	$Q_1 = x_k$
$k = \frac{n}{2}$	$Q_2 = \tilde{x}$
$k = \frac{3n+2}{4}$	$Q_3 = x_k$

No caso do número de dados da amostra ser ímpar temos:

A definição usada no R para o quartil-p amostral é o valor que separa os $p \times (100)\%$ valores menores da amostra dos $(1-p) \times (100)\%$ valores maiores da amostra. No programa R, o quantil-p da amostra `x` é dado por `quantile(x,p)`.

Tabela 4.9: Tabela dos quartis para n ímpar

Localização	Quartil
$k=\frac{n+1}{4}$	$Q_1=x_k$
$k=\frac{n+1}{2}$	$Q_2=\tilde{x}$
$k=3\times\frac{n+1}{4}$	$Q_3=x_k$

Diagrama de extremos e quartis

O diagrama de extremos e quartis (ou caixa-com-bigodes) é um tipo de representação gráfica em que se realçam algumas características da amostra, nomeadamente a amplitude amostral, a dispersão dos dados e as possíveis assimetrias da distribuição de dados. Para a sua construção precisamos de calcular a partir da amostra os quartis :q_1, q_2 e q_3 e os extremos amostrais ($x_{(1)}$ e $x_{(n)}$). O diagrama de extremos e quartis é um gráfico que apresenta uma caixa central limitada pelos quartis q_1 e q_3 sendo dividida por uma barra vertical com o valor de q_2. A largura da caixa não dá qualquer informação e a partir dos meios dos lados da caixa partem duas linhas até aos extremos da amostra. Para construir o diagrama de extremos e quartis procedemos do seguinte modo:

1. Determinar na amostra os extremos ($x_{(1)}$ e $x_{(n)}$) amostrais, e os quartis (q_1, q_2 e q_3);
2. Traçar um eixo com escala para assinalar os valores determinados anteriormente;
3. Traçar dois segmentos de reta correspondentes aos extremos da amostra;
4. Construir uma caixa em que os dois lados correspondem a q_1 e q_3;
5. Dividir a caixa anterior em duas partes usando um segmento de reta que corresponde ao valor da mediana.

   Finalizar a construção do diagrama unindo os valores extremos à caixa.

Nota: Este diagrama também é conhecido por caixa-com-bigodes e no R é realizado com o comando (`boxplot()`).

Valores muito pequenos ou muito grandes quando comparados com a maioria dos valores centrais da amostra, denominam-se de outliers.

Os outliers dividem-se em outliers moderados ou severos. Vejamos qual o critério de classificação:

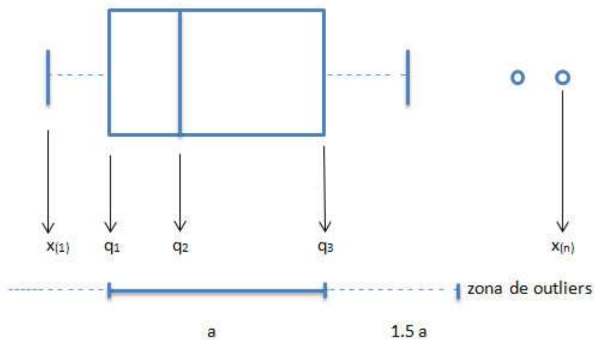


Figura 4.6: Diagrama de extremos e quartis

Uma observação é um outlier se está fora do intervalo

$$(q_1 - 1.5(q_3 - q_1),\, q_3 + 1.5(q_3 - q_1))$$

Um outlier é severo se está fora do intervalo $(q_1 - 3(q_3 - q_1),\, q_3 + 3(q_3 - q_1))$. Caso contrário diz-se moderado.

No R, o parâmetro `range=1.5` permite verificar a existência de outliers na amostra e está por defeito no `boxplot()`. Se for introduzida a opção `range=3` a existirem outliers são severos.

Exemplo 4.8
Os dados seguintes representam as classificações obtidas por 75 estudantes, num teste de Estatística (de escala 0-100).
75 98 42 75 84 87 50 65 59 63 86 78 37 99 66 90 79 80 89 68 57 95 55 79 88 76 60 77 49 92 83 71 78 53 81 77 58 93 85 70 62 80 74 69 90 62 84 74 73 61 74 65 56 67 68 56 65 76 65 60 76 89 76 84 58 67 59 67 75 76 65 68 66 65 84.

a) Determine as seguintes medidas de localização: média, moda e quartis;

b) Represente os dados num diagrama de extremos e quartis.

Resolução

a)

A média de amostra é $\overline{x} = \dfrac{\sum_{i=1}^{n} x_i}{n} = \dfrac{5403}{75} = 72.04$

A representação em caule-e-folhas é útil pois a amostra está ordenada permitindo identificar de um modo simples as observações amostrais a usar no cálculo da mediana, primeiro e segundo quartis.

```
3 | 7
4 | 2 9
5 | 0 3 5 6 6 7 8 8 9 9
6 | 0 0 1 2 2 3 5 5 5 5 5 5 6 6 7 7 7 8 8 8 9
7 | 0 1 3 4 4 4 5 5 5 6 6 6 6 6 7 7 8 8 9 9
8 | 0 0 1 3 4 4 4 4 5 6 7 8 9 9
9 | 0 0 2 3 5 8 9
```

Resolução em R

```
> stem(dados)
```

O diagrama caule-e-folhas fornece a informação necessária ao cálculo de algumas medidas estatísticas de localização.

A moda é 65, pois é o valor com maior frequência na amostra.

Como temos uma amostra de dimensão ímpar, n=75, a mediana é o elemento da amostra que se encontra na posição 38 ou utilizando a fórmula de mediana obtemos: $m_e = x_{\frac{n+1}{2}} = x_{\frac{75+1}{2}} = x_{38} = 74$.

Os quartis são as medianas de cada uma das partes em que ficou dividido o conjunto dos dados pela mediana.

O 1º quartil é Q_1:
$k = \frac{n+1}{4}$, então o valor de 1º quartil é $q_1 = x_{19} = 63$.

O 2º quartil é Q_2 é igual à mediana ($q_2 = 74$).

O 3º quartil é Q_3, é obtido pela fórmula:

$$k = 3 \times \frac{n+1}{4} = 3 \times \frac{75+1}{4} = \frac{228}{4} = 57$$

então o valor de $q_3 = x_{57} = 81$.

Estas estatísticas são calculadas no R com o comando `quantile()`.

Resolução em R

Tabela 4.10: Quadro resumo das principais estatísticas

Média	72.04
Moda	65
1º Quartil	63
2º Quartil ou mediana	74
3º Quartil	81

```
> > quantile(x,probs=seq(0,1,0.25))
  0%  25%  50%   75%  100%
37.0  64.0 74.0  80.5  99.0
```

Os diferentes valores obtidos para alguns dos quartis amostrais devem-se ao facto do manual escolar e do software R usarem definições diferentes para a mesma medida estatística.

b) Construção do diagrama de extremos e quartis:

Com a informação da tabela anterior, podemos construir o diagrama de extremos e quartis:

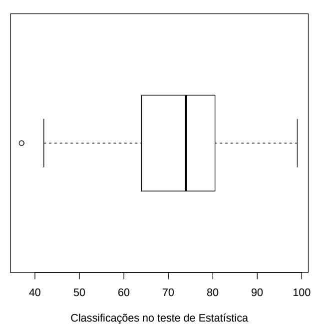


Figura 4.7: Diagrama de extremos e quartis

Da observação do diagrama concluímos que existe um outlier à esquerda. Como $14.5 < 37 < 39.25$, o valor 37 é um outlier moderado.

Resolução em R

```
>boxplot(dados)
```

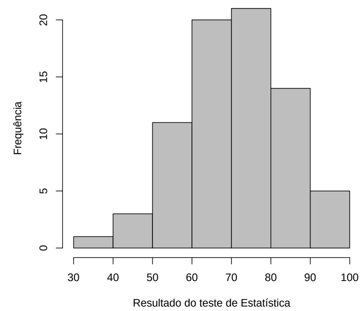


Figura 4.8: Histograma para X:"Classificação no teste de Estatística"

4.5 Medidas de Dispersão

As medidas de dispersão medem a variabilidade de um conjunto de dados a um parâmetro, a média.

Amplitude

Num conjunto de dados, chama-se amplitude, à diferença entre a maior e a menor das observações. Esta medida representa o intervalo de variação No R, esta medida de dispersão é obtida fazendo `range(x)` e `diff(range(x))`.

Amplitude Interquartil

A amplitude interquartil é dada por $q_3 - q_1$. O seu valor é um indicador do intervalo de variação na parte central da amostra.

Nota: No R, usa-se o comando `IQR(x)`.

Variância e Desvio padrão

A variância amostral é a média (corrigida) dos quadrados dos desvios em relação à média amostral. A variância representa-se por s^2.

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

Como a unidade da variância não é a mesma dos dados, torna-se útil determinar o desvio padrão que vem expresso na mesma unidade que os dados e que a média amostral.

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

Nota 1: No R, para a variância temos `var(x)` e para o desvio padrão `sd(x)`.

Nota 2: Quando $s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$ (variância não corrigida) então $s = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2}$

Nota 3: Para dados agrupados em classes temos a seguinte expressão:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{k} f_i(x_i - \overline{x})^2 \Rightarrow s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{k} f_i(x_i - \overline{x})^2}$$

com f_i a frequência absoluta e x_i o ponto médio da classe i, $i = 1, \cdots, k$.

Nota 4: No programa do Secundário em Timor-Leste, a variância amostral assume qualquer uma das notações: s^2 e σ^2. Neste trabalho reservamos σ^2 para a variância populacional.

Coeficiente de variação

O coeficiente de variação representa-se por $CV = \frac{s}{\overline{x}}$. O coeficiente de dispersão é uma medida de dispersão relativa, obtida por divisão de uma medida de dispersão por uma de localização, sendo deste modo invariante para mudanças de escala (mudança das unidades não altera o valor do coeficiente).

Nota: No R, esta medida de dispersão é dada por `sd(x)/mean(x)`.

Um quadro resumo é apresentado com o possível efeito nas medidas de dispersão quando se realiza uma mudança de variável.

Medida de dispersão	M. variável	Efeito
Amplitude	$Y = X + k$	$Amp(Y) = Amp(X)$
Amplitude	$Y = X \times k$	$Amp(Y) = Amp(X) \times k$
Variância	$Y = X + k$	$Var(X) = Var(Y)$
Variância	$Y = X \times k$	$Var(Y) = Var(X) \times k^2$
Desvio padrão	$Y = X + k$	$S_Y = S_X$
Desvio padrão	$Y = X \times k$	$S_Y = S_X \times k$
Coef. variação	$Y = X + k$	depende do sinal de k
Coef. variação	$Y = X \times k$	$CV(Y) = CV(X)$

Observação: as demonstrações destas propriedades das medidas de dispersão foram omitidas dado que são idênticas às realizadas para a medida de localização média.

4.6 Medidas de Forma

As medidas de forma servem para classificar a distribuição dos dados em relação ao achatamento e à assimetria.

Comecemos por definir momento central de ordem r:

$$m_r \quad = \frac{1}{n}\sum (x_i - \overline{x})^r$$

O coeficiente de assimetria é dado por:

$$b_1 = \frac{m_3}{m_2^{3/2}}$$

Diz-se que a distribuição dos dados apresenta uma assimetria negativa se $b_1 < 0$, positiva se $b_1 > 0$ e diz-se simétrica se $b_1 = 0$.

Quadro auxiliar de classificação de distribuições dos dados quanto à assimetria, relacionando as três medidas de localização:

Distribuição simétrica	média=mediana = moda
Distribuição assimétrica	positiva: média>mediana>moda negativa: média<mediana<moda

O coeficiente de achatamento é dado por:

$$b_2 = \frac{m_4}{m_2^2}$$

quando comparado com o modelo normal ($b_2 = 3$), a distribuição é dita platicúrtica se $b_2 < 3$, mesocúrtica se $b_2 = 3$ e leptocúrtica se $b_2 > 3$.

Principais características dos coeficientes de achatamento e assimetria:

O coeficiente de achatamento é sempre positivo;

O coeficiente de assimetria pode ser nulo, positivo ou negativo.

Os dois coeficientes são invariantes para mudanças de localização e escala dos dados.

Nota: No software R as medidas de forma encontram-se na library(moments). No exemplo 4.9 optamos por definir no R as funções correspondentes aos coeficientes de achatamento e assimetria.

Exemplo 4.9 Continuação do exemplo 4.8

Vamos determinar os coeficientes de assimetria e achatamento b_1 e b_2,

```
> Valor
> summary(Valor)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  37.00   64.00   74.00   72.04   80.50   99.00
```

```
> var(Valor)
[1] 169.4714
> n=length(Valor)
> n
[1] 75
> b1<-mean((Valor-mean(Valor))^3)/(var(Valor)*(n-1)/n)^(3/2)
> b1
[1] -0.1581455
> b2<-mean((Valor-mean(Valor))^{4)}/(var(Valor)*(n-1)/n)^(4/2)
> b2
[1] 2.749683
>IQR(Valor)
> gama
       75%
-0.2121212
```

Exemplo 4.10

O número de golos marcados nas 30 jornadas do clube da cidade de Gleno no distrito de Ermera, foram registados na tabela seguinte:

Tabela 4.11: Tabela de frequências

Número de golos	0	1	2	3	4	5
Número de jogos	3	10	7	4	5	1

Comecemos por determinar a média:

$$\begin{aligned}\bar{x} &= \frac{0\times 3+1\times 10+2\times 7+3\times 4\times 5+5\times 1}{30}\\ &= \frac{61}{30}\\ &= 2.03\end{aligned}$$

A tabela seguinte facilita a organização de dados para calcular a variância:

Tabela 4.12: Tabela de um Estudo

x_i	f_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	$f_i \times (x_i - \bar{x})^2$
0	3	-2.03	4.126	12.36
1	10	-1.03	1.06	10.60
2	7	-0.03	0.0009	0.0063
3	4	0.97	0.94	3.76
4	5	1.97	3.88	19.40
5	1	2.97	8.82	8.82
Total	30			$\sum_{i=1}^{n}(x - \bar{x})^2 = 54.95$

Com base nos cálculos da tabela temos:

$$\begin{aligned} s^2 &= \frac{\sum_{i=1}^{n} f_i(x_i - \overline{x})^2}{n} \\ &= \frac{54.95}{30} \\ &= 1.8 \end{aligned}$$

Retomamos o nosso exemplo 4.4

Tabela 4.13: Frequência Relativa

Classes	x_i	f_i	$x_i f_i$	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	$f_i \times (x_i - \bar{x})^2$
[44; 50[	47	1	47	-23	529	529
[50; 56[	53	2	106	-17	289	578
[56; 62[	59	4	236	-11	121	484
[62; 68[	65	12	780	-5	25	300
[68; 74[	71	3	213	1	1	3
[74; 80[	77	11	847	7	49	539
[80; 86[	83	7	581	13	169	1183
Total		40	$\sum_{i=1}^{n} x_i \times f_i = 2810$			$\sum_{i=1}^{n} f_i(x_i - \bar{x})^2 = 3616$

A média é

$$\bar{x} = \frac{\sum_{i=1}^{n} f_i \times x_i}{\sum_{i=1}^{n} f_i} = \frac{2810}{40} \cong 70$$

O desvio padrão é

$$s = \sqrt{\frac{\sum_{i=1}^{n} f_i (x_i - \overline{x})^2}{n}} = \sqrt{\frac{3616}{40}} \cong 9$$

4.7 Dados Bidimensionais

Este tópico do programa será aprofundado no capítulo reservado à estimação pontual contudo, faremos uma passagem breve ao conteúdo da regressão linear simples como estudado no 12º ano.

Muitos vezes, a análise estatística tem como objetivo estudar, em simultâneo, duas características do mesmo indíviduo dando origem a amostras bidimensionais. Estas amostras são constituídas por pares de dados. Cada coordenada do par é uma variável estatística, que vai ser observada e registada. As variáveis podem ser ambas quantitativas, qualitativos ou uma de cada tipo. Na análise de uma amostra bidimensional para além do estudo individual de cada uma das variáveis, interessa também verificar se existe algum tipo de associação entre elas e no caso afirmativo que tipo de relação. Considere o par de variáveis (X, Y), tal que X: "peso de um aluno da escola, em quilogramas" e Y: "altura de um aluno da escola, em centímetros". Para cada aluno vamos registar um par de valores, estes valores estão relacionados entre si porque são valores do mesmo indívíduo, mas os pares de valores são independentes entre si. Num estudo de dados bidimensionais a primeira abordagem é construír uma representação gráfica designada como diagrama de dispersão para se verificar se existe uma relação entre as variáveis.

Apresentamos como exemplo os seguintes diagramas de dispersão:

Caso 1: Existe correlação linear positiva entre duas variáveis, pois verifica-se que se uma variável cresce a outra também cresce.

Caso 2: Existe correlação Linear negativa entre duas variáveis, isto significa, que as variáveis evoluem em sentido contrário (se uma cresce a outra decresce).

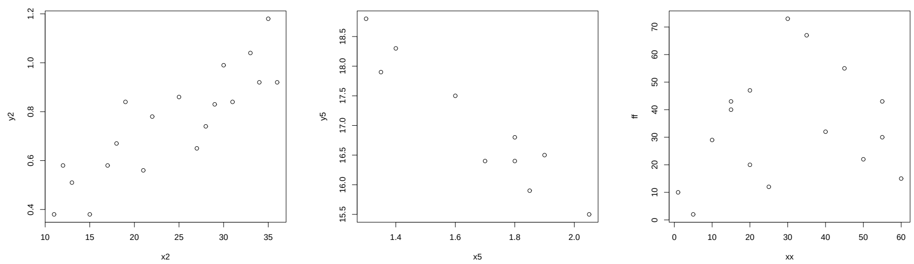


Figura 4.9: Diagrama de dispersão, da esquerda para a direita: caso 1; caso 2 e caso 3

Caso 3: Existe correlação nula se não há qualquer influência de uma variável na outra e neste caso a nuvem apresenta uma dispersão sem uma tendência definida.

Ao ponto de coordenadas ($\overline{x}$, $\overline{y}$) chama-se ponto médio da nuvem de pontos (ou centro de gravidade) e a reta que passa nesse ponto é a que melhor se ajusta à nuvem de pontos e chama-se reta de regressão linear. A reta de regressão linear faz sentido ser ajustada apenas nos casos 1 e 2.

Exemplo 4.11 Num dado estudo bivariado foram observadas os seguintes registos:

Tabela 4.14: Tabela de um Estudo

x_i	1	2	3	4	5	6
y_i	3	6	7	10	10	12

Esboce o diagrama de dispersão e ajuste a reta de regressão linear a este conjunto de dados recorrendo ao software R.

Resolução com R

A representação do diagrama de dispersão é feita com o comando `plot()`, como se pode observar na resolução:

```
> x<-c(1,2,3,4,5,6)
> y<-c(3,6,7,10,10,12)
> reta<-lm(y~x)
> reta
Call:
```

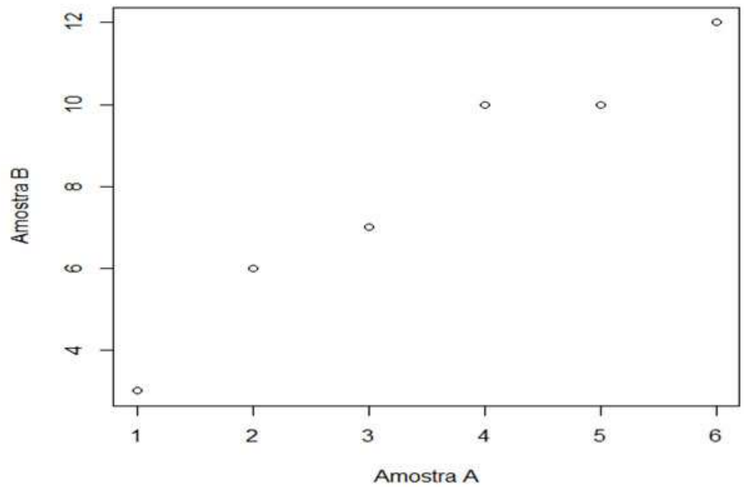


Figura 4.10: Diagrama de dispersão

```
lm(formula = y ~ x)
Coefficients:
(Intercept)            x
      2.000        1.714
> plot(x,y,ylim=c(0,15), xlab="Amostra A", ylab="Amostra B")
> mean(x)
[1] 3.5
> mean(y)
[1] 8
```

A reta obtida pelo comando `lm()` é a reta que melhor se ajuste à nuvem de pontos:

Resolução em R

```
> abline(reta)
> lines (x,1.7143*x+2.0000)
> text(4,7,"y = 1.71x + 2") # ou text(locator(1), "y = 1.71x + 2")
```

Nota: A regressão linear no programa do 12º não é lecionada com a profundidade suficiente para que os alunos a possam determinar pelo que, o uso do software R é uma ferramenta indispensável para a sua determinação e representação.

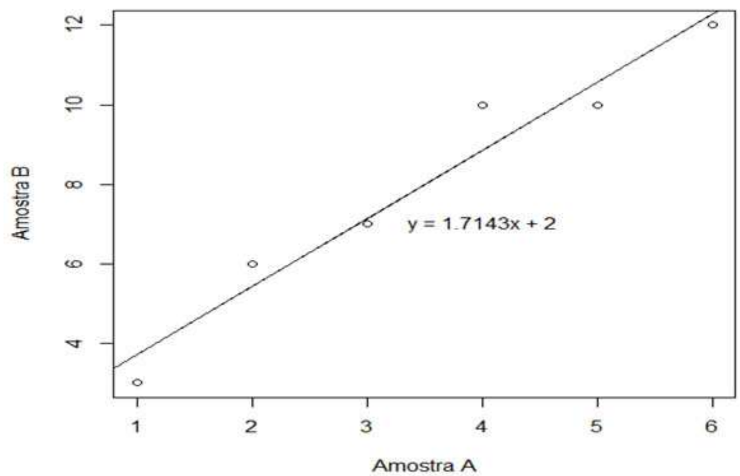


Figura 4.11: Regressão Linear

Para quantificar a associação do tipo linear entre duas variáveis vamos definir o coeficiente de correlação amostral de Pearson.

O grau de associação linear entre duas variáveis é transformado matematicamente por uma estatística a que chamamos coeficiente correlação linear e é designada por r. Seja $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$ os valores observados correspondentes ao par de variáveis (X, Y), define-se o coeficiente de correlação do seguinte modo:

$$
\begin{aligned}
r &= \frac{\displaystyle\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\displaystyle\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} \\
&= \frac{\displaystyle\sum_{i=1}^{n}(x_i - \bar{x})y_i}{\sqrt{\displaystyle\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} \\
&= \frac{\displaystyle\sum_{i=1}^{n}x_i(y_i - \bar{y})}{\sqrt{\displaystyle\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}
\end{aligned}
$$

$$
\begin{aligned}
r &= \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2\sum_{i=1}^{n}(y_i-\bar{y})^2}} \\
&= \frac{\sum_{i=1}^{n}x_iy_i - n\bar{x}\bar{y}}{\sqrt{(\sum_{i=1}^{n}x_i^2-n\bar{x}^2)(\sum_{i=1}^{n}y_i^2-n\bar{y}^2)}}
\end{aligned}
$$

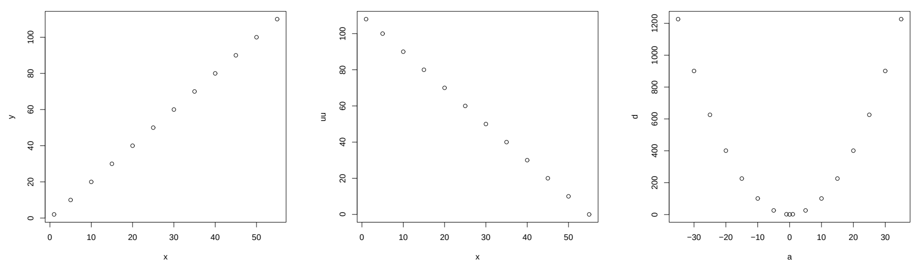


Figura 4.12: Exemplos de correlações, da esquerda para a direita: $r = 1$, $r = -1$ e $r = 0$.

Exemplo 4.12 Pretende-se averiguar a existência ou não de correlação entre a variável X ="Total de alunos", e Y= "Total de professores" do ensino básico de Timor-Leste nos 13 distritos no ano letivo de 2012, cuja amostra se encontra na seguinte tabela. Verifique se existe uma relação linear entre duas variáveis. Para saber se há ou não correlação estre duas variáveis, vamos construir uma tabela com os cálculos intermédios da expressão da correlação amostral de Pearson:

Tabela 4.15: Professor e os Alunos EB de Timor-Leste

Distritos	Alunos	Professores
Ainaro	18939	480
Aileu	13311	447
Baucau	34387	1063
Bobonaro	26001	802
Covalima	19188	650
Dili	54045	1254
Ermera	34711	727
Liquica	16941	490
Lautem	20519	657
Manufahi	13737	522
Manatuto	13238	403
Oecusse	16704	437
Viqueque	22675	800
Total	303396	8732

Tabela 4.16: Cálculos Intermédios da Expressão da Correlação

Distritos	x	y	$x \times y$	x^2	y^2
Ainaro	18939	480	9090720	358685721	230400
Aileu	13311	447	5950017	177182721	199809
Baucau	34387	1063	36553381	1182465769	1129969
Bobonaro	26001	802	20852802	676052001	643204
Covalima	19188	650	12472200	368179344	422500
Dili	54045	1254	67772430	2920862025	1572516
Ermera	34711	727	25234897	1204853521	528529
Liquica	16941	490	8301090	286997481	240100
Lautem	20519	657	13480983	421029361	431649
Manufahi	13737	522	7170714	188705169	272484
Manatuto	13238	403	5334914	175244644	162409
Oecusse	16704	437	7299648	279023616	190969
Viqueque	22675	800	18140000	514155625	640000
Total	303396	8732	237653796	8753436998	6664538

Com base na tabela obtemos:

$$\begin{aligned} r &= \frac{n(x_i \times y_i) - (\sum x_i) \times (\sum y_i)}{\sqrt{n \sum x_i^2 - (x_i^2)} \times \sqrt{n \sum y_i^2 - (y_i)^2}} \\ &= \frac{431513476}{4 \times 68664 \times 10^8} \\ &= 0.92073 \end{aligned}$$

O valor de $r \approx 0.921$ mostra uma correlação elevada e positiva entre as duas variáveis. O gráfico de regressão monstra seguinte:

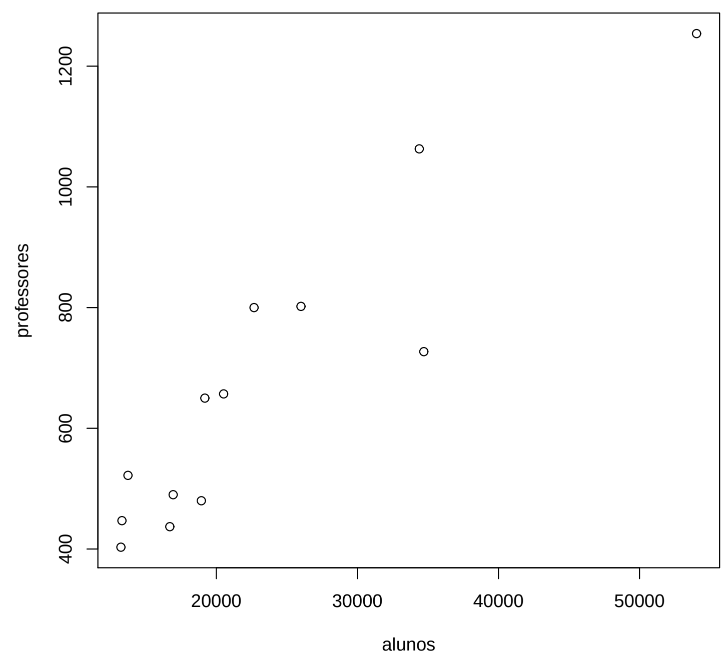


Figura 4.13: Diagrama de dispersão alunos vs professores

Resolução em R

```
> a<-c(18939,13311,34387,26001,19188,54045,34711,16941,20519,13737,13238,16704,22675)
> p<-c(480,447,1063,802,650,1254,727,490,657,522,403,437,800)
> plot(p~a,xlab="alunos",ylab="professores")
> cor(a,p)
```

```
[1] 0.9207309
```

Nota 1: O tópico das distribuições de probabilidade pela sua importância e destaque no programa do 12º será remetido para o capítulo seguinte.

Nota 2: O tópico da correlação e regressão linear simples encontra-se desenvolvido com mais detalhe no capítulo 6 numa secção 6.3 de dados bidimensionais.

Capítulo 5

Modelos Paramétricos

5.1 Variáveis Aleatórias

Nesta secção vamos fazer uma breve introdução a alguns conceitos associados às variáveis aleatórias.

Dada uma experiência aleatória, existem situações em que estamos interessados em associar valores numéricos aos resultados da experiência.

Exemplo 5.1 Considere a experiência aleatória de dois lançamentos de uma moeda equilibrada. Considere os acontecimentos: C= "saída da face cara" e E= "saída da face euro". O espaço de resultados associado a esta experiência aleatória é $\Omega = \{CC, CE, EC, EE\}$.

Para estabelecer uma correspondência entre o resultado ocorrido e um valor numérico é necessário definir variável aleatória. Assim, define-se: uma variável aleatória é uma função X, definida num espaço amostral e com valores em $\mathbb{R}$, que associa a cada elemento ω de Ω um valor real, que representamos por $X(\omega)$. Temos assim,

$$\begin{array}{rccc} X: & \Omega & \longrightarrow & \mathbb{R} \\ & \omega & \longmapsto & X(\omega) \end{array}$$

Exemplo 5.1 (cont.): Considere X a v.a. que representa o "número de caras". Para este exemplo obtemos a seguinte tabela:

ω	CC	CE	EC	EE
$X(\omega)$	2	1	1	0

Como a moeda é equilibrada, cada um dos 4 resultados possíveis tem probabilidade 1/4, e portanto a v.a. X pode assumir os valores 0, 1 ou 2, respetivamente com probabilidades 1/4, 1/2 e 1/4.

As variáveis aleatórias podem ser discretas (assume um número finito ou infinito numerável de valores) ou contínuas (assume um número finito ou infinito numerável de valores).

Seja A o conjunto de valores que uma variável aleatória discreta X assume. A função massa probabilidade (f.m.p.) da variável aleatória X é definida por:

$$f(x) = \begin{cases} P(X = x) & \text{se} \quad x \in A \\ 0 & \text{outros valores} \end{cases}$$

De modo equivalente podemos escrever a função massa de probabilidade de uma v.a. X que toma os valores $\{x_1, x_2, ...\}$, designando por p_i a probabilidade de cada elemento x_i,

$$p_i = P(X = x_i), \quad i = 1, 2, \ldots$$

ou

$$X : \begin{cases} x_1 & x_2 & x_3 & \ldots \\ p_1 & p_2 & p_3 & \ldots \end{cases}$$

Exemplo 5.1 (cont.)

$$X : \begin{cases} 0 & 1 & 2 \\ 1/4 & 1/2 & 1/4 \end{cases}$$

Define-se função distribuição de X, à função real de variável real, F, com domínio $\mathbb{R}$ tal que:

$$F(x) = P(X \leq x)$$

Propriedades de $F(x)$:

$0 \leq F(x) \leq 1, \quad \forall \in \mathbb{R}$

$\lim\limits_{x \to -\infty} F(x) = 0$ e $\lim\limits_{x \to +\infty} F(x) = 1$

$F(x) = \sum_{x_i \leq x} f(x_i)$

$f(x_i) = P(X = x_i) = F(x_i) - F(x_{i-1})$

Para uma variável contínua temos definições semelhantes às de uma variável discreta mas adaptadas à natureza da variável. Uma variável aleatória diz-se contínua se e só se existir uma função real, $f(x)$ não negativa tal que:

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt$$

f(.) designa-se por função densidade de probabilidade

F(.) designa-se por função de distribuição

$f(x) \geq 0, \quad \forall x \in \mathbb{R}$

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

$f(x) = \dfrac{dF(x)}{dx}$

Para as variáveis contínuas verificam-se:

$P(X \leq a) = F(a)$

$P(a < X \leq b) = \int_a^b f(t)dt = F(b) - F(a)$

$P(X > a) = 1 - P(X \leq a) = 1 - F(a)$

$P(X = a) = 0 \quad \forall a$

$P(a \leq X \leq b) = P(a < X < b) = P(a \leq X < b) = P(a < X \leq b)$

As distribuições têm características teóricas correspondentes às características amostrais estudadas na Estatística Descritiva.

O valor médio ou valor esperado de uma variável aleatória X é definido como:

$$E[X] = \begin{cases} \sum_i x_i f(x_i) & \text{v.a. discreta} \\ \\ \int_{-\infty}^{+\infty} x f(x)dx & \text{v.a. contínua} \end{cases}$$

$$\mu = \sum_i x_i f(x_i) \quad \text{ou} \quad \mu = \int x f(x)dx$$

O valor médio de X que corresponde à média amostral designa-se por μ e é uma média pesada (de acordo com f.m.p. ou f.d.) dos valores de X.

As expressões anteriores podem ser generalizadas para uma função de uma variável aleatória, $Y = h(X)$, obtendo-se:

$$E[Y] = \begin{cases} \sum_i h(x_i) f(x_i) & \text{v.a. discreta} \\ \\ \int_{-\infty}^{+\infty} h(x) f(x)dx & \text{v.a. contínua} \end{cases}$$

Propriedades do valor esperado:

Sejam X e Y duas v.a. independentes e a e b duas constantes reais.

$E[a] = a$

$E[aX + b] = aE[X] + b$

$E[X + Y] = E[X] + E[Y]$

$E[XY] = E[X]E[Y]$

Vejamos outras medidas com interesse das distribuições:

Moda: valor x para o qual a função f(x) é máxima.

Mediana (M): é o menor valor de x tal que $F(x) \geq 0.5$.

No caso contínuo, corresponde ao valor que de x que acumula à sua esquerda uma área de 0.5 da densidade.

A variância de uma variável aleatória X é definida por:

$$Var[X] = E[(X - \mu)^2] = \begin{cases} \sum_i (x_i - \mu)^2 f(x_i) & \text{v.a. discreta} \\ \\ \int_{-\infty}^{+\infty} (x - \mu)^2 f(x) dx & \text{v.a. contínua} \end{cases}$$

$Var[X] = E[X^2] - E[X]^2$

$\sigma^2 = Var[X]$

Propriedades da variância: Sejam X e Y duas v.a. independentes e a e b duas constantes reais.

$Var[X] \geq 0$

$Var[a] = 0$

$Var[aX + b] = a^2 Var[X]$

$Var[X + Y] = Var[X] + Var[Y]$

Vamos em seguida apresentar com algum detalhe as distribuições discretas de Bernoulli, Binomial e Poisson e a distribuição contínua normal.

5.2 Distribuição de Bernoulli

A distribuição de Bernoulli é uma distribuição de variável aleatória discreta que está associada a um processo de Bernoulli. Um processo de Bernoulli é qualquer experiência estatística com as seguintes propriedades:

- Consiste em n tentativas repetidas;
- Cada tentativa tem dois resultados possíveis: sucesso ou insucesso;
- A probabilidade de sucesso p é a mesma em qualquer tentativa;
- As tentativas repetidas são independentes (e, portanto, a probabilidade de sucesso não é afetada pelo possível conhecimento do resultado obtido em tentativas anteriores).

Considere-se X uma variável aleatória (v.a.) que admite dois valores possíveis (sucesso ou insucesso). Normalmente, considera-se $X = 0$ quando o resultado é um insucesso e $X = 1$ para o resultado sucesso. Ao sucesso está associado a probabilidade p e q é a probabilidade de ocorrer um insucesso, com $q = 1 - p$. A v.a. discreta X segue uma distribuição de Bernoulli, se a sua função massa de probabilidade é:

$$f(x) = P(X = x) = \begin{cases} p^x(1-p)^{r-x}, & x \in \{0, 1\} \\ 0, & \text{outros valores} \end{cases}$$

A notação abreviada é $X \sim Bernoulli(p)$.

Propriedades da distribuição de Bernoulli

Parâmetro: $p \in]0, 1[$

Conjunto de valores de X: $\{0, 1\}$

Valor médio:
$E(X) = \mu_X = p$.
Demonstração

$$
\begin{aligned}
E(X) &= \sum_{x=0}^{1} x f(x) \\
&= \sum_{x=0}^{1} x p^x (1-p)^{n-x} \\
&= 0 \times (1-p) + 1 \times p \\
&= p
\end{aligned}
$$

Variância:
$Var(X) = \sigma_X^2 = p(1-p)$.
Demostração

$$
\begin{aligned}
Var(X) &= E(X^2) - [E(X))]^2 \\
&= [1^2 \times p] - p^2 \\
&= p - p^2 \\
&= p(1-p)
\end{aligned}
$$

Exemplo 5.2
Considere-se o lançamentos de um dado em que o acontecimento de sucesso é **"saída de face superior a 4"**. Defina a variável aleatória associada à experiência.
Resolução

Seja X a v.a. discreta que assume o valor 1 se ocorre o sucesso **"saída de face superior a 4"** e 0 se ocorre o insucesso.

$$p = P(\text{"sucesso"}) = P(\text{"saída de face superior a 4"}) = \frac{2}{6} = \frac{1}{3}$$

Seja,

$$1 - p = P(\text{"insucesso"}) = 1 - \frac{1}{3} = \frac{2}{3}$$

Define-se a v.a. X como $X \sim \text{Bernoulli}(\frac{1}{3})$.

5.3 Distribuição Binomial

A distribuição binomial é pode ser vista como uma generalização da distribuição Bernoulli para o caso de n tentativas independentes e portanto está também ela associada a um processo de Bernoulli. A distribuição binomial ou modelo binomial modela experiências com as seguintes características:

- Em cada tentativa considera-se somente a ocorrência ou não de um certo evento que será denominado sucesso e cuja não ocorrência é denominada por insucesso;
- As tentativas são independentes;
- A probabilidade de sucesso p é a mesma para cada tentativa. A probabilidadede insucesso será denotada por $1 - p = q$.

Como já foi referido, a distribuição binomial pode ser vista como uma generalização da distribuição de Bernoulli para o caso de uma sequência de n tentativas de Bernoulli. Assim, se X_i representar o sucesso/insucesso obtidos na tentativa i (X_i toma o valor 1 ou 0, respetivamente), e cada $X_i \sim Bernoulli(p)$, $i = 1, 2, \cdots, n$, sendo o número de sucessos em n tentativas de Bernoulli independentes uma variável aleatória X tal que

$$X = X_1 + X_2 + \cdots + X_n \sim Bin(n, p).$$

A função massa de probabilidade da distribuição binomial é definida:

$$f(x) = P(X = x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x}, & x \in \{0, 1, 2, \cdots, n\} \\ 0, & x \notin \{0, 1, 2, \cdots, n\} \end{cases}$$

De forma abreviada escrevemos que $X \sim Bin(n, p)$.

Os coeficientes binomiais $\binom{n}{x}$ que aparecem na expressão da f.m.p. do modelo binomial definem-se como:

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}, \quad x = 0, 1, \ldots, n$$

e verificam as seguintes propriedades:

i $\binom{n}{0} = \binom{n}{n} = 1;\ \ 0! = 1$

ii $\binom{n}{1} = \binom{n}{n-1} = n;$

iii $\binom{n}{x} = \binom{n}{x-1} = \binom{n+1}{x+1};$

iv $(a+b)^n = \sum_{x=0}^{n} \binom{n}{x} a^x b^{n-x}$, para todos os a,b $\in \mathbb{R}$

Propriedades da distribuição binomial

Parâmetros: $n \in N$ e $p \in]0,1[$

Conjunto de valores de X: $\{0, 1, 2, \cdots, n\}$

Observação: Nas demonstrações do valor médio e da variância da v.a. $X \sim Bin(n,p)$, vamos considerar $X = \sum X_i$ com $X_i \sim Bernoulli(p)$, $i = 1, \ldots, n$ v.a.'s independentes e identicamente distribuidas com $E(X_i) = p$ e $Var(X_i) = pq$, sendo $q = 1-p$.

Valor médio: $E(X) = \mu_X = np$
Demonstração

Sendo X uma soma de variáveis independentes o seu valor médio é igual à soma dos valores médios das variáveis X_i, $i = 1, \ldots, n$.

$$\begin{aligned} \mu_X &= E(X_1 + X_2 + \cdots + X_n) \\ &= E(X_1) + E(X_2) + \cdots + E(X_n) \\ &= p + \cdots + p \\ &= np \end{aligned}$$

Variância: $Var(X) = \sigma_X^2 = np(1-p)$

Sendo X uma soma de variáveis independentes a sua variância é igual à soma das variâncias das variáveis X_i, $i = 1, \ldots, n$.
Demonstração

$$\begin{aligned} \sigma_X^2 &= Var(X_1 + X_2 + \cdots + X_n) \\ &= Var(X_1) + Var(X_2) + \cdots + Var(X_n) \\ &= p(1-p) + \cdots + p(1-p) \\ &= np(1-p) \end{aligned}$$

A determinação do valor médio e da variância através do modelo binomial vai ser apresentada recorrendo à função geradora de momentos que passamos a descrever:

Função Geradora de Momento (f.g.m)

A função geradora de momento de uma variável X é definida por

$$M_X(t) = E(e^{tX})$$

desde que valor médio seja finito para t real em algum intervalo $-t_0 < t < t_0$ com $t_0 > 0$.

Recordamos que e^x pode ser escrito como uma expansão em série de potências

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$$

temos então que

$$e^{tx} = 1 + tx + \frac{(tx)^2}{2!} + \frac{(tx)^3}{3!} + \cdots$$

Aplicando o valor médio em ambos os lados, obtemos do lado esquerdo a f.g.m. $M_X(t)$.

$$M_X(t) = E(e^{tX}) = E(1 + tX + \frac{(tX)^2}{2!} + \frac{(tX)^3}{3!} + \cdots)$$

$$M_X(t) = 1 + tE(X) + \frac{t^2 E(X^2)}{2!} + \frac{t^3 E(X^3)}{3!} + \cdots$$

Admitimos ser possível permutar soma infinita e valor médio.

Como $M_X(t)$ é uma função na variável t, é possível derivar $M_X(t)$ em ordem a t. (Suponhamos agora que o lado direito pode ser escrito uma soma infinita das respetivas derivadas)

$$M_X^{'}(t) = \frac{d}{dt} M_X(t) = 0 + E(X) + \frac{2tE(X^2)}{2!} + \cdots$$

Para $t = 0$ obtemos:

$$M_X^{'}(0) = E(X)$$

Calculando a segunda derivada de $M_X(t)$ temos que,

$$M_X^{''}(t) = \frac{d\, M_X^{'}(t)}{dt} = E(X^2) + tE(X^3) + \cdots$$

Para $t = 0$ obtemos:

$$M_X''(0) = E(X^2)$$

Podemos então calcular a variância de X

$$Var(X) = E(X^2) - (E(X)^2) = M_X''(0) - (M_X'(0))^2.$$

Vamos aplicar este método ao modelo binomial,

$$\begin{aligned} M_X(t) &= E(e^{tX}) = \sum_{k=0}^{n} e^{tk} \binom{n}{k} p^k (1-p)^{n-k} \\ &= \sum_{k=0}^{n} \binom{n}{k} (e^t p)^k (1-p)^{n-k} \\ &= (pe^t + 1 - p)^n \end{aligned}$$

$$M_X'(t) = \frac{d}{dt}(pe^t + 1 - p)^n = n(pe^t + 1 - p)^{n-1} pe^t$$

$$E(X) = M_X'(0) \text{ então } E(X) = n(p + 1 - p)^{n-1} p = np$$

Para determinarmos a variância derivamos mais uma vez a função $M_X(t)$

$$M_X'' = \frac{d^2}{dt^2}(pe^t + 1 - p)^n = n(n-1)(pe^t + 1 - p)^{n-2}(pe^t)^2 + npe^t(pe^t + 1 - p)^{n-1}$$

$$M_X''(0) = n(n-1)(p + 1 - p)^{n-2} \times p^2 + np(p + 1 - p)^{n-1} = n(n-1)p^2 + np$$

e portanto, obtemos que

$$E(X^2) = M_X''(0) = n(n-1)p^2 + np.$$

Deste modo, $Var(X)$ pode ser calculado por:

$$\begin{aligned} Var(X) &= E(X^2) - (E(X))^2 \\ &= \left[n(n-1)p^2 + np\right] - (np)^2 \\ &= n^2p^2 - np^2 + np - n^2p^2 \\ &= np(1-p) \end{aligned}$$

Moda:
A distribuição é unimodal se $(n+1)p$ não é um número inteiro. Neste caso, a moda é dada por $[(n+1)p]$, ou seja, a parte inteira de $(n+1)p$. A distribuição é bimodal se $(n+1)p$ é um número inteiro. Neste caso as modas são $(n+1)p$ e $(n+1)p-1$.

Representam-se de seguida os gráficos correspondentes às f.m.p.'s de distribuições binomiais com parâmetros $n=19$ e $p=1/5$ e $n=20$ e $p=1/5$ para ilustrar as situações unimodal e bimodal do modelo binomial.

Coeficiente de assimetria: $\beta_1 = \dfrac{1-2p}{\sqrt{np(1-p)}}$

Coeficiente de achatamento: $\beta_2 = 3 + \dfrac{1-6p(1-p)}{\sqrt{np(1-p)}}$

Nota 1: Define-se $[x]$ como a parte inteira de x.

Nota 2: $X \sim Bernoulli(p) \Leftrightarrow X \sim Bin(1,p)$

Nota 3: $X_i \sim Bin(n_i,p)$,$i=1,2,\cdots,m$ e se $X_1, X_2, \cdots, X_m$ são variáveis independentes então,

$$X = \sum_{i=1}^{m} X_i = X_1 + X_2 + \cdots + X_n \sim Bin(n_1+n_2+\ldots+n_m, p).$$

Exemplo 5.3
Sendo X uma v.a. discreta tal que $X \sim Bin(25,p)$ e $E(X)=7$.

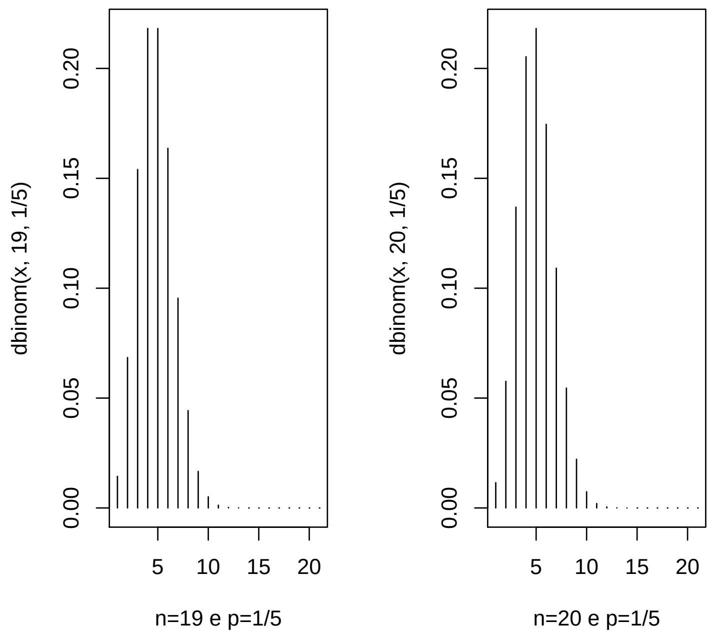


Figura 5.1: $X \sim Bin(19, 1/5)$ e $X \sim Bin(20, 1/5)$

1. Determine o valor de p;

2. Calcule:

   (a) $P(X = 10)$;

   (b) $P(X \geq 10)$;

   (c) $P(4 \leq X \leq 10)$.

Resolução

Seja $X \sim Bin(25, p)$, como $E(X) = 7$ então podemos escrever,

1. $E(X) = 7 \Leftrightarrow np = 7 \Leftrightarrow 25p = 7 \Leftrightarrow p = \frac{7}{25} \Leftrightarrow p = 0.28$

   (a) $X \sim Bin(25, 0.28) \Leftrightarrow P(X = r) = \binom{25}{r} 0.28^r (1 - 0.28)^{25-r} \quad r = 0, \ldots, 25$

$$
\begin{aligned}
P(X=10) &= \frac{25!}{10!(25-10)!} \times 0.28^{10} \times (1-0.28)^{25-10} \\
&= \frac{25!}{10!15!} \times 0.28^{10} \times 0.72^{15} \\
&= \frac{25!}{10!15!} \times 0.28^{10\times} 0.72^{15} \\
&= 0.070
\end{aligned}
$$

Resolução em R

```
> n=25
> p=0.28
> dbinom(10,n,p)
[1] 0.070
```

(b) $P(X \geq 10)$

$$
\begin{aligned}
P(X \geq 10) &= 1 - P(X < 10) \\
&= 1 - P(X \leq 9) \\
&= 1 - [P(X=0) + \cdots + P(X=9)] \\
&= 1 - (0.000 + 0.003 + 0.012 + \cdots + 0.153 + 0.113) \\
&= 1 - 0.865 \\
&= 0.134
\end{aligned}
$$

Resolução em R

```
> 1-pbinom(9,n,p)
[1] 0.134
```

ou de um modo equivalente,

```
> pbinom(9,25,0.28, lower.tail=F)
[1] 0.134
```

(c) $P(4 \leq X \leq 10)$

$$\begin{aligned} P(4 \leq X \leq 10) &= P(X=4) + \cdots + P(X=10) \\ &= 0.078 + 0.128 + 0.166 + 0.175 + 0.153 + 0.113 + 0.070 \\ &= 0.884 \end{aligned}$$

Resolução em R

```
> pbinom(10,n,p)-pbinom(3,n,p)
[1] 0.884
```

Exemplo 5.4
Suponha que numa linha de produção a probabilidade de obter uma unidade defeituosa (sucesso) é $p = 0.2$. Toma-se uma amostra de 20 unidades para serem inspecionadas. Qual é a probabilidade de obter:

1. Uma unidade defeituosa;
2. No máximo três unidades defeituosas.

Resolução
Seja X o número de unidades defeituosas, então
$X \sim Bin(n = 20, p = 0.2)$

1. Uma unidade defeituosa:

$$\begin{aligned} P(X=1) &= \binom{20}{1} \times 0.2^1 \times (1-0,2)^{20-1} \\ &= \frac{20!}{1!(20-1)!} \times 0.2 \times 0.8^{19} \\ &= 0.06 \end{aligned}$$

Resolução em R

```
>dbinom(1,20,0.2)
[1] 0.06
```

Com o comando plot podemos representar o gráfico da função de distribuição através do seguinte comando:

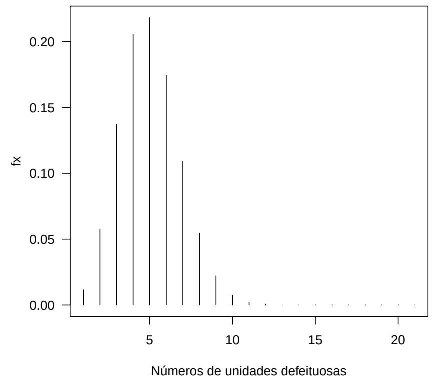


Figura 5.2: Função massa de probabilidade

```
> plot(dbinom(seq(0,20, by=1),size=20, prob=0.2), type="h",
+ xlab="Numero de unidades defeituosas",
+ ylab="Probabilidade", main="Funcao massa de probabilidade")
```

A função de probabilidade acumulada pode ser representada com seguinte instrução:

```
plot(pbinom(seq(0,20, by=1),size=20, prob=0.2), type="h",
+ xlab="Numero de unidades defeituosas",
+ ylab="Fx")
```

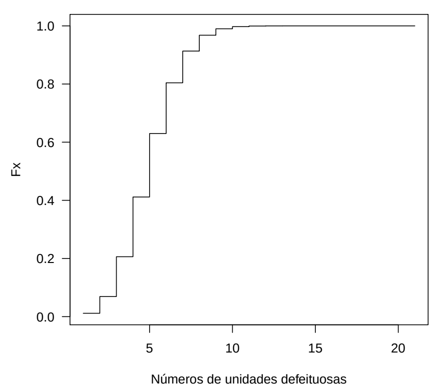


Figura 5.3: Função de distribuição

2. No máximo três unidades defeituosas:

$$\begin{aligned} P(X \leq 2) &= P(X=0)+P(X=1)+P(X=2) \\ &= (0,016+0,058+0,137) \\ &= 0,211 \end{aligned}$$

Resolução em R

```
> pbinom(2,20,0.2)
[1] 0.21
```

5.4 Distribuição de Poisson

A distribuição de Poisson é um modelo probabilístico adequado para descrever os fenómenos em que os acontecimentos se repetem no tempo ou no espaço. Um modelo de Poisson verifica as seguintes condições:

- O número de ocorrências em intervalos de tempo não sobrepostos são variáveis aleatórias independentes;
- A probabilidade de um certo número de ocorrências se verificar é a mesma para intervalos da mesma dimensão, isto é, aquela probabilidade depende apenas da amplitude do intervalo e não dá posição em que se situa esse intervalo;
- A probabilidade de se registarem duas ou mais, ocorrências num intervalo suficientemente pequeno é desprezável, quando comparada com a probabilidade de se verificar apenas uma ocorrência.

A variável aleatória discreta X, que designa o número de ocorrências num determinado intervalo de tempo, quando os eventos são independentes uns dos outros, segue a distribuição de Poisson e escreve-se:
$X \sim Poisson(\lambda)$, sendo λ o parâmetro, a função massa de probabilidade é representada por:

$$f(x) = P(X=x) = \begin{cases} \frac{e^{-\lambda}\lambda^x}{x!}, & x \in \{0,1,2,\cdots\} \\ \\ 0, & x \notin \{0,1,2,\cdots\} \end{cases}$$

Outra forma de calcular $f(x)$ é usando a seguinte fórmula de recorrência:

$$\begin{cases} f(0) &= e^{-\lambda} \\ f(x) &= f(x-1)\dfrac{\lambda}{x}, \quad x \in \{1,2,3,\cdots\} \end{cases}$$

Esta maneira de definir $f(x)$ é útil para justificar o seguinte: Considera-se um processo de Poisson definido num intervalo de tempo t. Este intervalo é dividido em subintervalos muitos pequenos de modo que:

(a) a probabilidade de ocorrer um evento num dessas subintervalos é proporcional ao seu comprimento;

(b) a probabilidade de ocorrer mais do que um evento num desses subintervalos é desprezável;

(c) a probabilidade de ocorrer um evento num desses subintervalos é independente de ter ou não ocorrido um evento noutro qualquer subintervalo.

Suponha-se que λ representa o número médio de eventos que ocorrem no intervalo de tempo t. Podemos considerar um processo de Bernoulli; $X_i \sim Bernoulli(p = \lambda/n)$, para $i = 1, \ldots, n$ em que $X_i = 1$ ou $X_i = 0$ consoante ocorre ou não um evento no i-ésimo subintervalo, para calcular

$$P(X_1 + X_2 + \cdots + X_n = x)$$

em que $X_1 + X_2 + \cdots + X_n \sim Bin(n, p = \lambda/n)$ representa o número de eventos que ocorrem no intervalo de tempo t. Neste caso; $X_1 + X_2 + \cdots + X_n$ o que implica que

$$P(X_1 + X_2 + \cdots + X_n = x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x \in \{0, 1, 2, \cdots, n\}.$$

Vamos mostrar que, quando $n \to +\infty$, esta distribuição tende para a distribuição de Poisson:

$$\begin{aligned}
P(X = x) = \binom{n}{x} p^x (1-p)^{n-x} &= \frac{n!}{x!(n-x)!} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x} \\
&= \frac{n!}{x!(n-x)!} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{-x} \left(1 - \frac{\lambda}{n}\right)^{n} \\
&= \frac{n(n-1)\cdots(n-x+1)}{x!} \left(\frac{\lambda}{n}\right)^x \left(\frac{n}{n-\lambda}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n} \\
&= \frac{n(n-1)\cdots(n-x+1)}{x!} \frac{\lambda^x}{n^x} \left(\frac{1}{1-\frac{\lambda}{n}}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n} \\
&= \frac{n(n-1)\cdots(n-x+1)}{n^x} \frac{\lambda^x}{x!} \left(\frac{1}{1-\frac{\lambda}{n}}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n}
\end{aligned}$$

A expressão $\frac{n(n-1)\cdots(n-x+1)}{n^x}$ é um quociente de dois polinómios de grau x, pelo que

$$\lim_{n\to+\infty} \frac{n(n-1)\cdots(n-x+1)}{n^x} = 1$$

$$\lim_{n\to+\infty} \left(\frac{1}{1-\frac{\lambda}{n}}\right)^x = 1$$

$$\lim_{n\to+\infty} \left(1-\frac{\lambda}{n}\right)^n = e^{-\lambda}$$

Assim, obtemos o resultado pretendido:

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

No caso de $Y \sim Bin(n,p)$ e de n ser suficientemente grande, podemos aproximar $P(Y=y)$ usando a distribuição de Poisson com $\lambda = np$.

Propriedades da distribuição de Poisson

Parâmetro:
O parâmetro representa o número médio de eventos que ocorrem num intervalo de tempo ou numa região espacial, e designado por:$\lambda \in]0,+\infty[$.

Conjunto de valores de X: $\{0,1,2,\cdots\}$

Valor médio: $E(X)$=$\mu_X = \lambda$.
Demostração

$$\begin{aligned}
E(X) &= \sum_{x=0}^{+\infty} x\frac{e^{-\lambda}\lambda^x}{x!} \\
&= \lambda e^{-\lambda}\sum_{x=0}^{+\infty}\frac{\lambda^{x-1}}{(x-1)!} \\
&= \lambda e^{-\lambda}\sum_{x=1}^{+\infty}\frac{\lambda^s}{s!} \\
&= \lambda e^{-\lambda}e^{\lambda} \\
&= \lambda
\end{aligned}$$

Variância: $Var(X) = \sigma_X^2 = \lambda$

Demostração

$$
\begin{aligned}
E(X^2) &= \sum_{x=0}^{+\infty} x^2 \frac{e^{-\lambda}\lambda^x}{x!} \\
&= \lambda e^{-\lambda} \sum_{x=1}^{+\infty} x \frac{\lambda^{x-1}}{(x-1)!}, \text{considerando } x-1=s, \text{ obtemos} \\
&= \lambda e^{-\lambda} \sum_{s=0}^{+\infty} (s+1) \frac{\lambda^s}{s!}
\end{aligned}
$$

Mas

$$
\begin{aligned}
\sum_{x=0}^{+\infty} (s+1) \frac{\lambda^s}{s!} &= \sum_{s=0}^{+\infty} s \frac{\lambda^s}{s!} + \sum_{s=0}^{+\infty} \frac{\lambda^s}{s!} \\
&= \sum_{s=1}^{+\infty} \frac{\lambda \lambda^{s-1}}{(s-1)!} + e^{\lambda} \\
&= \lambda \sum_{s=0}^{+\infty} \frac{\lambda^s}{s!} + e^{\lambda} \\
&= e^{\lambda}(\lambda+1)
\end{aligned}
$$

Assim

$$
\begin{aligned}
E(X^2) &= \lambda e^{-\lambda} e^{\lambda}(\lambda+1) \\
&= \lambda^2 + \lambda
\end{aligned}
$$

e

$$
\begin{aligned}
Var(X) &= E(X^2) - [E(X)]^2 \\
&= (\lambda^2 + \lambda) - \lambda^2 \\
&= \lambda + \lambda^2 - \lambda^2 \\
&= \lambda
\end{aligned}
$$

Moda:
Se λ não for um número inteiro então a distribuição tem a unimodal. Neste caso a moda é dada por $[\lambda]$, ou seja a parte inteira de λ. Se λ é um número inteiro então a moda da distribuição é bimodal. Neste caso as modas são $\lambda - 1$ e λ.

Representam-se de seguida os gráficos correspondentes às f.m.p.'s de distribuições de Poisson com parâmetros $\lambda = 3.5$ e $\lambda = 2$ para ilustrar as situações unimodal e bimodal do modelo de Poisson.

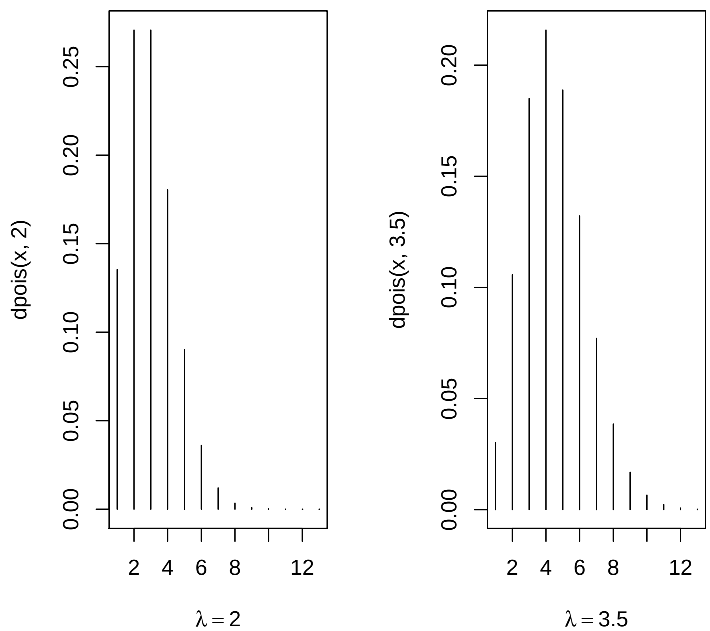


Figura 5.4: $X \sim Poisson(2)$ e $X \sim Poisson(3.5)$

Coeficiente de assimetria: $\beta_1 = \dfrac{1}{\sqrt{\lambda}}$

Coeficiente de achatamento: $\beta_2 = 3 + \dfrac{1}{\sqrt{\lambda}}$

Exemplo 5.5 A radiação de um determinado material emite partículas γ a uma taxa de duas por segundo. Determine a probabilidade de:

(a) não serem emitidas partículas num período de 0.5 segundo

(b) serem emitidas duas partículas num segundo.

(c) serem emitidas pelo menos 3 partículas em dois segundos.

Resolução

$\lambda = 2$

X : "número de partículas γ emitidas por segundo"

$X \sim Poisson(2)$

$P(X = k) = \dfrac{e^{-2}2^2}{k!}$, $k = 0, 1, 2, \cdots$

(a) serem emitidas duas partículas num segundo:

$t = 1$,

$P(X_1 = 2) = \frac{e^{-2(1)}(2(1))^2}{2!} = 0.271.$

Resolução em R

```
> dpois(2,2)
[1] 0.271
```

(b) não serem emitidas partículas num período de 0.5 segundo

$Y \sim Poisson(1)$

y : "número de partículas γ emitidas em 0.5 segundos"

$P(y = 0) = \frac{e^{-1}1^0}{0!} = e^{-1} = 0.369$

Resolução em R

```
> dpois(0,1)
[1] 0.3678794
```

(c) serem emitidas pelo menos 3 partículas em dois segundos:

T : "número de partículas γ emitidas em 2 segundos".

$T \sim Poisson(4)$

$$\begin{aligned} P(T \geq 3) &= 1 - P(T < 3) \\ &= 1 - P(T \leq 2) \\ &= 1 - 0.2381033 \\ &= 0.762 \end{aligned}$$

Resolução em R

```
> 1-ppois(2,4)
[1] 0.762
```

5.5 Distribuição Normal

A distribuição normal ou distribuição gaussiana é uma distribuição contínua com forma de sino e desempenha a nível da Estatística um papel primordial pelas suas propriedades e aplicações. Vejamos algumas aplicações que a tornam especial:

- É um modelo adequado para representar muitos dos fenómenos do mundo real (características humanas como a altura e o peso, características mensuráveis, etc);
- É muito usada na inferência estatística. Mesmo quando a distribuição da população não é normal, a distribuição das médias amostrais é aproximadamente normal (teorema do Limite Central);
- Muitas técnicas desenvolvidas na área da Estatística são exatas no caso de distribuições normais;
- Algumas variáveis aleatórias (como por exemplo, a binomial e a de Poisson) podem ser aproximadas por uma variável aleatória normal.

Seja X uma variável aleatória contínua que tem distribuição normal com valor médio μ e variância σ^2, então escrevemos:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2},$$

onde $x \in R$. Vejamos algumas características da distribuição normal.

Características da curva normal

- Forma em sino ou simétrica
  Tem um máximo para $x = \mu$

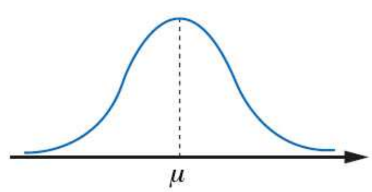


Figura 5.5: Curva Normal

- Ouanto maior for o desvio padrão, σ, mais achatada é a curva.
  $\sigma_2 > \sigma_1$

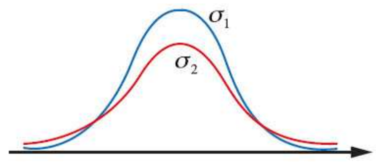


Figura 5.6: Curva Normal

- A área compreendida entre a curva e o eixo $0x$ é igual a 1

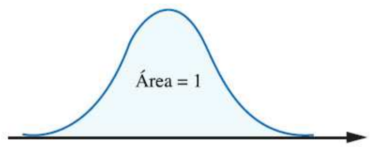


Figura 5.7: Curva Normal

- A probabilidade de que a variável tome valores no intervalo $[x_i, x_j[$ é igual à área definida pelo eixo Ox, pelo gráfico da função densidade e pelas retas $x = x_i$ e $x = x_j$.

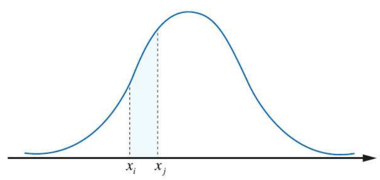


Figura 5.8: Curva Normal

- A concavidade da curva muda de sentido para

$$x_1 = \mu - \sigma \text{ e } x_2 = \mu + \sigma.$$

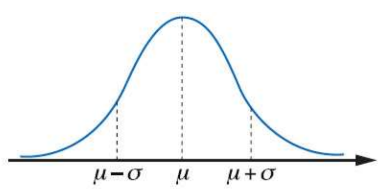


Figura 5.9: Curva Normal

- O eixo das abcissas é uma assintota da curva. A área abaixo da curva distribui-se em intervalos da seguinte forma:

$$P(\mu - \sigma \leq X \leq \mu + \sigma) \approx 0,6827\%$$

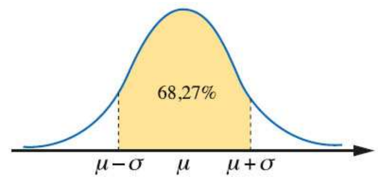


Figura 5.10: Curva Normal

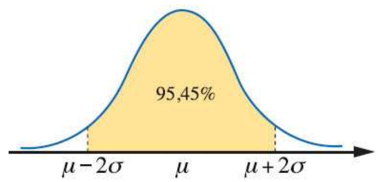


Figura 5.11: Curva Normal

$$P(\mu - \sigma \leq X \leq \mu + \sigma) \approx 95,45\%$$

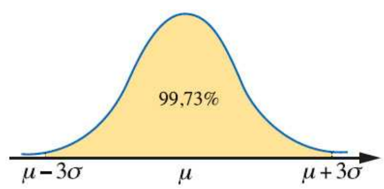


Figura 5.12: Curva Normal

$$P(\mu - \sigma \leq X \leq \mu + \sigma) \approx 0,9973\%$$

Propriedades da distribuição Normal

Parâmetro:
A distribuição normal é uma distribuição que tem dois parâmetros, o valor médio μ (parâmetro de localização) e a variância σ^2 (parâmetro de escala).

Conjunto de valores de X: $]-\infty, +\infty[$

Valor Médio:
$E(X) = \mu$ (parâmetro de localização).
Demostração

$$E(X) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} x e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} dx,$$

Fazendo a mudança de variável: $z = \frac{x-\mu}{\sigma} \Leftrightarrow x = z\sigma + \mu$ e
$dx = \sigma dz$, obtemos

$$\begin{aligned}
E(X) &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} x e^{-z^2/2} dx \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} (\sigma z + \mu) e^{-z^2/2} dz \\
&= \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} z e^{-z^2/2} dz + \mu \frac{\mu}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-z^2/2} dz \\
&= \frac{\sigma}{\sqrt{2\pi}} \times 0 + \mu \times 1 \\
&= \mu
\end{aligned}$$

Variância:
Demonstração

$$E(X^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} x^2 e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} dx,$$

Fazendo a mudança de variável: $z = \frac{x-\mu}{\sigma} \Leftrightarrow x = z\sigma + \mu$ e $dx = \sigma dz$, obtemos

$$\begin{aligned}
E(X^2) &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} (\sigma z + \mu)^2 e^{-z^2/2} dz \\
&= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} z^2 e^{-z^2/2} dz + \frac{2\mu\sigma}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} z e^{-z^2/2} dz + \frac{\mu^2}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-z^2/2} dz \\
&= \mu
\end{aligned}$$

como vimos, o segundo integral é nulo e o terceiro integral é um,

$$\frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} z e^{-z^2/2} dz \;\text{ e }\; \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-z^2/2} dz = 1.$$

Assim,

$$E(X^2) = \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} z^2 e^{-z^2/2} dz + \mu^2$$

A resolução deste integral é feita pelo método de integração por partes:
$dv = ze^{-z^2/2}dz$, $v = -ze^{-z^2/2}$ e $u = z$, $du = dz$
obtemos

$$E(X^2) \quad = \sigma^2 \left\{ \left[\frac{-ze^{-z^2/2}}{\sqrt{2\pi}} \right]_{-\infty}^{+\infty} + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-z^2/2} dz \right\} + \mu^2.$$

Portanto:
$Var(X)=E(X^2) - [E(X)]^2=(\sigma^2 + \mu^2) - \mu^2=\sigma^2$

Moda: μ

Coeficiente de assimetria: $\beta_1 = 0$

Coeficiente de achatamento:$\beta_2 = 3$

Exemplo 5.6
Considere a v.a. $Z \sim N(\mu = 0, \sigma^2 = 1)$. Determine a probabilidades $P(Z > 1.21)$.
Resolução
$Z \sim N(\mu = 0, \sigma^2 = 1)$, σ=1

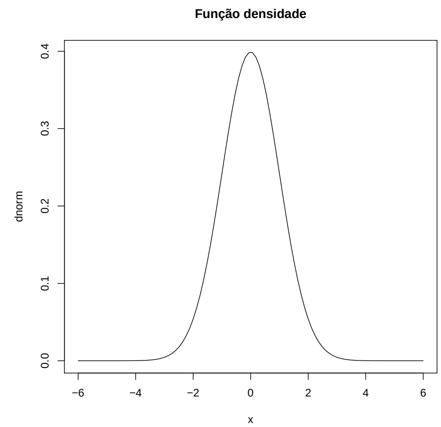
$$\begin{aligned} P(Z > 1.21) &= 1 - P(Z > \tfrac{X-\mu}{\sigma}) \\ &= 1 - P(Z > \tfrac{1.21-0}{1}) \\ &= 1 - \Phi(1.21) \\ &= 1 - 0.88686 \\ &= 0.1131 \end{aligned}$$

Resolução em R

```
> 1-pnorm(1.21,0,1)
[1] 0.1131
```

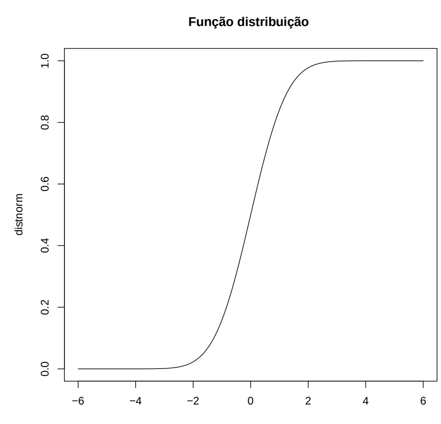
Determine o gráfico da função de densidade de probabilidade através do seguinte comando em R:

```
> x=seq(-3,3, length=100)
> y=dnorm(x, mean=0, sd=1)
> plot(x,y, type="l", lwd=2, col="black", ylab="Probabilidade",
 + main="Funcao densidade ")
```



O gráfico da função de distribuição executa-se com a seguinte sequência de instruções:

```
> x=seq(-3,3, length=100)
> y=dnorm(x, mean=0, sd=1)
> y=pnorm(x, mean=0, sd=1)
> plot(x,y, type="l", lwd=2, col="black", ylab="Probabilidade",
 + main="Funcao distribuicao")
```



5.5.1 Variáveis normais

Dado o destaque das variáveis aleatórias normais, iremos apresentar alguns resultados respeitantes à soma e média de variáveis aleatórias.

- Se $X \frown N(\mu, \sigma)$ e $Y = aX + b$ com a e b constantes, então
$$Y \frown N(a\mu + b, \sqrt{a^2\sigma^2})$$

- Se $X_i \frown N(\mu_i, \sigma_i), i = 1, 2, \ldots, n$ então
$$X = X_1 + X_2 + \cdots + X_n \frown N(\mu, \sigma)$$
com $\mu = \mu_1 + \mu_2 + \cdots + \mu_n$ e $\sigma^2 = \sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2$;

- Se $X_i \frown N(\mu, \sigma), i = 1, 2, \ldots, n$ são variáveis aleatórias independentes e identicamente distribuídas então
$$X = X_1 + X_2 + \cdots + X_n \frown N(\mu_1, \sigma_1)$$
com $\mu_1 = n\mu$ e $\sigma_1^2 = n\sigma^2$;

- Se $X_i \frown N(\mu, \sigma), i = 1, 2, \ldots, n$, são são variáveis aleatórias independentes e identicamente distribuídas então
$$\overline{X} \frown N(\mu, \sigma/\sqrt{n})$$

Lei dos Grandes Números (LGN): Esta lei estabelece que dada uma amostra de variáveis aleatórias $X_1, \cdots, X_n$ independentes e identicamente distribuídas com X, tal que $\mu = E[X]$, então a média amostral $\overline{X}$ converge em probabilidade para μ, i.e., a probabilidade de $\overline{X}$ estar próximo de μ (tão próximo quanto se queira) tende para 1.

$$\lim_{n\to\infty} P(|\overline{X} - \mu| < \epsilon) = 1$$

Uma consequência da LGN é que a frequência relativa de um acontecimento A converge para a sua probabilidade, $P(A)$.

O teorema seguinte garante que a soma de várias variáveis aleatórias independentes e identicamente distribuídas tem distribuição que se aproxima da Normal Reduzida para n grande ($n \geq 30$).

Teorema do limite central (TLC): Sejam $X_1, X_2, \ldots, X_n$ variáveis aleatórias independentes e identicamente distribuídas com valor médio μ e variância σ^2 e consideremos a variável aleatória $\sum_{i=1}^{n} X_i$. Então,

$$\frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}} \frown N(0,1) \Leftrightarrow \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \frown N(0,1)$$

Pelo TLC, podemos aproximar probabilidade referentes às v.a. $\sum_{i=1}^{n} X_i$ e $\overline{X}$, calculadas a partir do modelo normal, qualquer que seja a distribuição subjacente às v.a. X_i (variância finita), para $n \to \infty$.

Exemplo 5.7 O peso de um homem é uma variável aleatória com distribuição $N(75, 5)$. Qual a probabilidade do peso de 4 homens (com pesos independentes) não exceder 320kg?

Resolução

Considerando $Y = \sum_{i=1}^{4} X_i$ com $X_i \frown N(75, 5)$, $\quad i = 1, ..., 4$, $Y \frown N(\mu, \sigma)$ com $\mu = 4 \times 75 = 300$ e $\sigma^2 = 4 \times 25 = 100 \Rightarrow \sigma = 10$;

$$P(Y \leq 320) = P(Z \leq \frac{320 - 300}{10}) = \Phi(2) = 0.9772$$

Resolução em R:

$P(Y \leq 320) =$`pnorm(320,300,10)=0.9772499`

5.5.2 Aproximações de distribuições discretas

- Se $X \frown bi(n, p)$ com n grande e p pequeno então $X \frown Poisson(\lambda = np)$;

- Aproximação da distribuição binomial pela distribuição Normal:

  Se $X_i \frown Ber(p)$, $i = 1, ..., n$, então $S_n \frown Bin(n, p)$ e o TLC garante

  $$S_n \frown N(np, \sqrt{np(1-p)})$$

  Esta aproximação é considerada boa para $n > 30$, $np \geq 5$ e $n(1-p) \geq 5$.

- Aproximação da distribuição de Poisson pela Normal:

  Uma v.a. que segue $P(\lambda)$, com λ inteiro, pode ser considerada como a soma de λ variáveis aleatórias de Poisson de parâmetro unitário. Assim, para λ grande (≥ 5) pode-se aproximar $P(\lambda)$ pela distribuição $N(\lambda, \sqrt{\lambda})$.

As aproximações entre diferentes distribuições serão visualizadas através de gráficos. Nas três figuras seguintes pretendemos mostrar como a distribuição binomial é bem aproximada pela distribuição de Poisson para valores grandes de n.

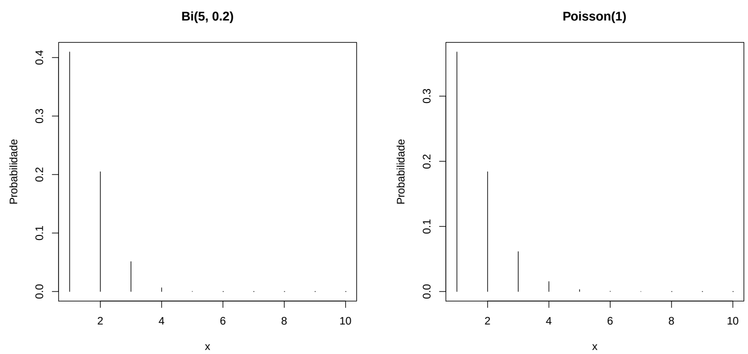


Figura 5.13: Binomial vs Poisson com $n = 6$, $p = 0.5$ e $\lambda = 1$

Apresentam-se a seguir as probabilidades dadas pelas duas distribuições que confirmam a vizualização gráfica observada na Figura 5.13.

Resolução em R

```
> dbinom(0:10,5,0.2)
 [1] 0.32768 0.40960 0.20480 0.05120 0.00640
 [6] 0.00032 0.00000 0.00000 0.00000 0.00000
[11] 0.00000 0.00000

> dpois(0:10,1)
 [1] 0.36788 0.36788 0.18394 0.06131 0.01533
 [6] 0.00307 0.00051 0.00007 0.00000 0.00000
[11] 0.00000
```

Apresentam-se a seguir as probabilidades dadas pelas duas distribuições que confirmam a vizualização gráfica observada na Figura 5.14.

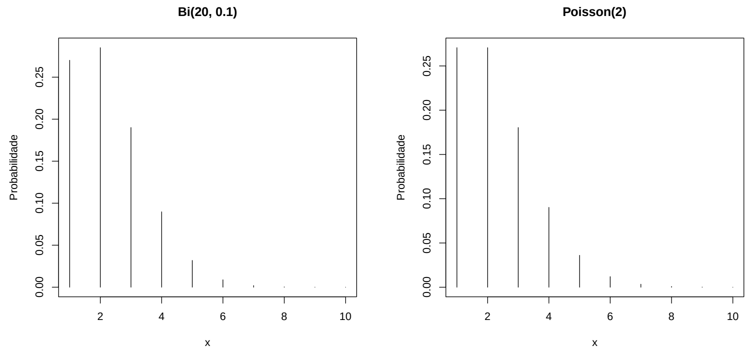


Figura 5.14: Binomial vs Poisson com $n = 20$, $p = 0.1$ e $\lambda = 2$

Resolução em R

```
> dbinom(0:10,20,0.1)
 [1] 0.12158 0.27017 0.28518 0.19012 0.08978
 [6] 0.03192 0.00887 0.00197 0.00036 0.00005
[11] 0.00000

> dpois(0:10,2)
 [1] 0.13534 0.27067 0.27067 0.18045 0.09022
 [6] 0.03609 0.01203 0.00344 0.00086 0.00019
[11] 0.00000
```

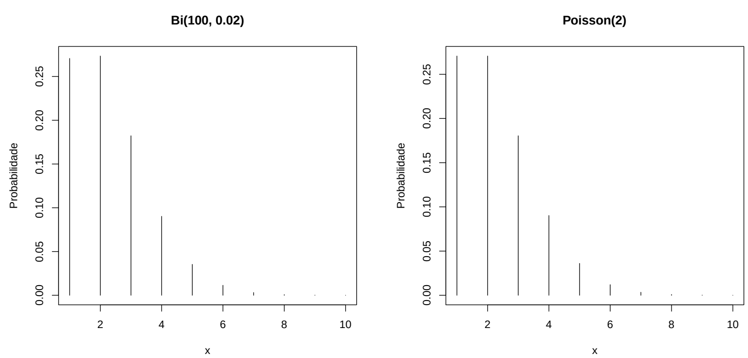


Figura 5.15: Binomial vs Poisson com $n = 100$, $p = 0.02$ e $\lambda = 2$

Apresentam-se a seguir as probabilidades dadas pelas duas distribuições que confirmam a vizualização gráfica observada na Figura 5.15 e comparamos os valores obtidos com os da distribuição Binomial com n=1000 e p=0.002 (situação com n grande e p pequeno).

Resolução em R

```
>dbinom(0:10,100,0.02)
 [1] 0.13262 0.27065 0.27341 0.18228 0.09021
 [6] 0.03535 0.01142 0.00313 0.00074 0.00015
[11] 0.00002

> dpois(0:10,2)
 [1] 0.13533 0.27067 0.27067 0.18045 0.09022
 [6] 0.03609 0.01203 0.00344 0.00086 0.00019
[11] 0.00004

> dbinom(0:10,1000,0.002)
 [1] 0.13506 0.27067 0.27094 0.18063 0.09022
 [6] 0.03602 0.01197 0.00341 0.00085 0.00019
[11] 0.00004
```

Terminamos a visualização gráfica com as aproximações das distribuições Binomial e Poisson pela distribuição Normal.

Para mostrar que as distribuições discretas Binomial e Poisson são bem aproximadas pela distribuição contínua Normal, foram calculados os valores das respetivas distribuições para vários valores das variáveis aleatórias:

Resolução em R

```
> x1<-c(10,15,20,25,30,35,40,45,50,55)
> pbinom(x1,100,0.3)
 [1] 0.00002 0.00040 0.01646 0.16313 0.54912
 [6] 0.88392 0.98750 0.99946 0.99999 1.00000

> pnorm(x1,30,sqrt(21))
 [1] 0.00001 0.00053 0.01455 0.13762 0.50000
 [6] 0.86238 0.98545 0.99947 0.99999 1.00000


> x2<-c(20,25,30,35,40,50,55,60,65,70)
> ppois(x2,50)
 [1] 0.00000 0.00001 0.00159 0.01621 0.08607
 [6] 0.53752 0.78447 0.92784 0.98274 0.99703
```

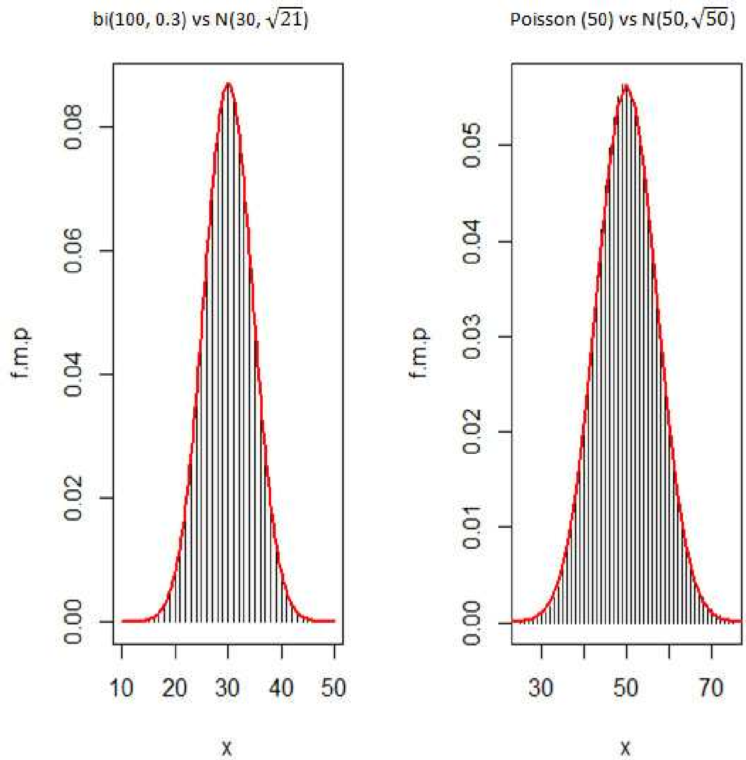


Figura 5.16: Binomial vs Normal e Poisson vs Normal

```
> pnorm(x2,50,sqrt(50))
 [1] 0.00000 0.00020 0.00233 0.01695 0.07865
 [6] 0.50000 0.76025 0.92135 0.98305 0.99766
```

Capítulo 6

Estimação Pontual

6.1 Introdução

Neste capítulo, vamos considerar que se observa uma amostra aleatória de uma variável X, e pretendemos identificar os parâmetros da distribuição (Binomial, Poisson, Normal, ...) que gerou esses dados. Numa prespetiva paramétrica é suposto conhecermos a forma do modelo do qual os dados são provenientes, o que na prática se traduz por adaptar à variável um conjunto de possíveis modelos e a partir da informação prévia disponível e da análise dos dados da amostra identificar o elemento da família que melhor se adapta aos dados em estudo.

Amostra aleatória simples

Uma amostra diz-se aleatória simples quando: (i) cada elemento da população tem a mesma probabilidade de ser escolhido; (ii) as observações realizam-se com reposição, de modo que a população é idêntica em todas as extrações.

Numa amostra aleatória simples, cada observação tem a distribuição de probabilidade (função massa de probabilidade - f.m.p.; ou função densidade de probabilidade - f.d.p.) da população. Seja $f(x;\theta)$ a distribuição de probabilidade da variável observada X, e represente-se uma amostra de dimensão n como $(x_1, x_2, \ldots, x_n)$, onde x_i representa o valor de x no i-ésimo elemento. Na amostra observada todos os $x_i,\ i = 1, \ldots, n$ são independentes e identicamente distribuidos pelo que é válida a seguinte igualdade:

$$f(x_1, x_2, \ldots, x_n; \theta) = f(x_1; \theta) f(x_2; \theta) \ldots f(x_n; \theta).$$

Sempre que se mencionar amostra aleatória (a.a.) está subentendido que a amostra é aleatória simples.

Exemplo 6.1 Seja X uma variável aleatória que segue uma distribuição de Poisson com $\lambda = 2$. Pretende-se determinar a probabilidade de obter a amostra $(3, 1, 0, 2, 1)$.

Resolução

Sendo X uma v.a. discreta, a probabilidade de se obter a amostra é igual ao produto da função massa de probabilidade em cada valor observado da amostra, isto é

$$\begin{aligned} f(3,1,0,2,1) &= f(3)f(1)f(0)f(2)f(1) = \\ &= P(X_1 = 3, X_2 = 1, X_3 = 0, X_4 = 2, X_5 = 1) = \\ &= P(X = 3)P(X = 1)P(X = 0)P(X = 2)P(X = 1) \end{aligned}$$

Como $f(x) = P(X = x) = \dfrac{e^{-2}2^x}{x!}$, $x = 0, 1, \ldots$, para a amostra observada obtemos,

$$P(X = 3)P(X = 1)P(X = 0)P(X = 2)P(X = 1) =$$

$$= \frac{e^{-2}2^3}{3!}\frac{e^{-2}2^1}{1!}\frac{e^{-2}2^0}{0!}\frac{e^{-2}2^2}{2!}\frac{e^{-2}2^1}{1!} =$$

$$= e^{-10}2^7\frac{1}{3!}\frac{1}{1!}\frac{1}{0!}\frac{1}{2!}\frac{1}{1!} = 0.00048$$

A probabilidade da amostra $(3, 1, 0, 2, 1)$ ocorrer é de 4.8×10^{-4}.

Resolução em R

$> dpois(0, 2) * dpois(1, 2)^2 * dpois(2, 2) * dpois(3, 2)$

[1]0.00048

6.2 Estimadores Pontuais e Métodos

Dada uma amostra proveniente de um modelo paramétrico em que um ou mais parâmetros não estão especificados, torna-se necessário encontrar estimadores para esses parâmetros. Os estimadores são expressões que usam apenas a informação amostral para atribuir valores, estimativas, para os parâmetros desconhecidos do modelo. Vamos apresentar dois métodos para obter estimadores: o método dos momentos e o método de máxima verosimilhança, para mais detalhe consultar, por exemplo, Pestana e Velosa (2010) e Murteira et al. (2010).

Método dos Momentos

Consideremos uma a.a. $(x_1, x_2, \ldots, x_n)$ de uma população com função densidade de probabilidade (ou f.m.p.) $f(x;\theta)$, em que $\theta = (\theta_1, \theta_2, \ldots, \theta_r)$ depende de r parâmetros. Pretende-se estimar $\theta_1, \theta_2, \ldots, \theta_r$.
O método dos momentos consiste em igualar os r primeiros momentos simples da amostra, aos r primeiros momentos simples da população, i.e,

$$\begin{cases} E(X) = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} x_i \\ \vdots \\ E(X^r) = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} x_i^r \end{cases}$$

Exemplo 6.2 Seja $(X_1, X_2, \ldots, X_n)$ uma a.a. de uma população $N(\mu, \sigma)$ com dois parâmetros desconhecidos, μ e σ. Pretende-se determinar estimadores para μ e σ.

Resolução

Temos que $Var(X) = E(X^2) - (E(X))^2$ pelo que, sendo $Var(X) = \sigma^2$ e $E(X) = \mu$, $E(X^2) = \sigma^2 + \mu^2$. Então, o sistema é

$$\begin{cases} E(X) = \mu = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} x_i = \overline{x} \\ \\ E(X^2) = \sigma^2 + \mu^2 = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} x_i^2 \end{cases}$$

Para os parâmetros μ e σ^2, os estimadores obtidos pelo método dos momentos, são representados como $\tilde{\mu}$ e $\tilde{\sigma}^2$.

$$\begin{cases} \tilde{\mu} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} x_i = \overline{x} \\ \\ \tilde{\sigma}^2 = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} x_i^2 - (\overline{x})^2 \end{cases}$$

ou seja o estimador para μ é a média amostral $\overline{x}$ e para σ^2 é a variância amostral não corrigida.

Método da Máxima Verosimilhança

O método da máxima verosimilhança consiste em encontrar o estimador $\hat{\theta}$ que maximiza o valor da função de verosimilhança para uma determinada amostra. Este método pode ser aplicado para estimar mais do que um parâmetro em simultâneo.

Seja $(x_1, x_2, \cdots, x_n)$, uma amostra aleatória obtida de uma população com função densidade de probabilidade, $f(x;\theta)$, em que $\theta = (\theta_1, \theta_2, \ldots, \theta_r)$. A função densidade de probabilidade conjunta (ou f.m.p. conjunta) das variáveis que constituem a amostra é dada por:

$$f(x_1, x_2, \cdots, x_n; \theta) = f(x_1;\theta) f(x_2;\theta) \cdots f(x_n;\theta) = \prod_{i=1}^{n} f(x_i;\theta).$$

dado que os x_i $i = 1, \cdots, n$ são independentes e identicamente distribuídos. Para a amostra $\mathbf{x} = (x_1, x_2, \cdots, x_n)$, designa-se por função de verosimilhança a função de θ e da amostra tal que:

$$L(\theta; x_1, x_2, \cdots, x_n) = L(\theta; \mathbf{x}) = \prod_{i=1}^{n} f(x_i;\theta).$$

Os passos a seguir permitem obter o estimador de maxíma verosimilhança:

1. Determinar a função de verosimilhança $L(\theta;\mathbf{x})$;
2. Se necessário aplicar a transformação logarítmica à função de verosimilhança $ln\, L(\theta;\mathbf{x}) = l(\theta)$. Esta transformação, em geral, torna o problema da maximização mais simples;
3. Determinar os pontos onde a 1ª derivada da função $L(\theta;\mathbf{x})$ ou ln $L(\theta;\mathbf{x})$ em ordem a θ_j se anula (condição de primeira ordem) com $j = 1, \cdots, r$:
$$\frac{\partial L(\theta;\mathbf{x})}{\partial \theta_j} = 0 \quad \text{ou} \quad \frac{\partial ln\, L(\theta;\mathbf{x})}{\partial \theta_j} = 0$$
4. Verificar se a 2ª derivada em ordem a θ_j é negativa (condição de segunda ordem) com $j = 1, \cdots, r$:

$$\frac{\partial^2 L(\theta;\mathbf{x})}{\partial\theta_j^2} < 0 \quad \text{ou} \quad \frac{\partial^2 ln\, L(\theta;\mathbf{x})}{\partial\theta_j^2} < 0$$

Exempo 6.3 Uma sondagem realizada no distrito de Dili, a 600 pessoas mostrou que 150 são a favor do imposto extra, cujo valor será usado no programa de desenvolvimento tecnológico "Para além de Dili". Deduzir o estimador de máxima verosimilhança para a probabilidade p de uma pessoa escolhida ao acaso no distrito de Dili ser a favor do imposto extra.

Resolução

Seja X: "Ser a favor do imposto extra para o desenvolvimento". Os valores que X pode tomar são: 1 (resposta sim), 0 (resposta não), como a variável é discreta e dicotómica, $X \frown Bernoulli(p)$ com parâmetro $p = P(X = 1)$:

$$P(X = x) = p^x\,(1-p)^{1-x}, \quad x = 0, 1$$

Cálculo da função verosimilhança:

$$L(p; x_1, x_2, \ldots, x_n) = L(p;\mathbf{x}) = \prod_{i=1}^{n} P(X = x_i)$$

$$= \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i}$$

Cálculo da função log-verosimilhança:

$$\ln\, L(p;\mathbf{x}) = l(p) = \ln\left(p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i}\right) =$$

$$= \ln(p)\sum_{i=1}^{n} x_i + \ln(1-p)(n - \sum_{i=1}^{n} x_i)$$

o estimador de máxima verosimilhança para o parâmetro p, representa-se por $\hat{p}$ e obtém-se resolvendo a condições de primeira ordem:

$$\frac{d\,l(p)}{d\,p} = 0 \Leftrightarrow \frac{d\left[\ln\,p\sum_{i=1}^{n}x_i + \ln\,(1-p)\left(n-\sum_{i=1}^{n}x_i\right)\right]}{d\,p} = 0$$

$$\Leftrightarrow \frac{\sum_{i=1}^{n}x_i}{p} - \frac{n-\sum_{i=1}^{n}x_i}{1-p} = 0$$

$$\Leftrightarrow \frac{\sum_{i=1}^{n}x_i}{p} - \frac{n-\sum_{i=1}^{n}x_i}{1-p} = 0$$

$$\Leftrightarrow (1-p)\sum_{i=1}^{n}x_i - p(n-\sum_{i=1}^{n}x_i) = 0 \Leftrightarrow p = \frac{1}{n}\sum_{i=1}^{n}x_i$$

Condições de segunda ordem:

$$\frac{d^2\,l(p)}{d\,p^2} = \frac{d^2\left[\ln\,p\sum_{i=1}^{n}x_i + \ln\,(1-p)\left(n-\sum_{i=1}^{n}x_i\right)\right]}{d\,p^2}$$

$$= \frac{-\sum_{i=1}^{n}x_i}{p^2} - \frac{n-\sum_{i=1}^{n}x_i}{(1-p)^2}$$

$$= \frac{-\sum_{i=1}^{n}x_i}{p^2} - \frac{n-\sum_{i=1}^{n}x_i}{(1-p)^2}$$

$$= \frac{-\sum_{i=1}^{n}x_i}{p^2} - \frac{n-\sum_{i=1}^{n}x_i}{(1-p)^2}$$

$$= \frac{-\left[(1-p)^2\sum_{i=1}^{n}x_i + np^2 + p^2\sum_{i=1}^{n}x_i\right]}{p^2(1-p)^2} < 0, \quad 0 \leq p \leq 1$$

Concluimos então, que o estimador de máxima verosilhança de p é a média amostral de uma variável de Bernoulli:

$$\hat{p} = \overline{X}$$

Neste exemplo temos

$$\hat{p} = \frac{160}{600} = 0.25$$

Podemos escrever então, que $X \frown Bernoulli(0.25) = bin(1, 0.25)$.

Exemplo 6.4

Seja $\mathbf{x} = (x_1, x_2, \cdots, x_n)$ uma amostra aleatória de uma distribuição normal, $X \sim N(\mu; \sigma)$, com μ e σ parâmetros cujos valores são desconhecidos . Determinar os estimadores para os parâmetros μ e σ^2 pelo método da máxima verosimilhança.

Resolução

Função densidade de probabilidade (f.d.p.):

$$f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad \sigma > 0$$

Função verosimilhança:

$$\begin{aligned} L(\mu, \sigma^2; \mathbf{x}) &= \prod_{i=1}^{n} f(x_i; \mu, \sigma^2) \\ &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2} \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\,\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2} \\ &= \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2} \end{aligned}$$

Logaritmo da função verosimilhança:

$$\ln\, L(\mu, \sigma^2; \mathbf{x}) = l(\mu, \sigma^2) = -\frac{n}{2}(\ln\, 2 + \ln\, \pi + \ln\, \sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2.$$

Condições de primeira ordem:

$$\begin{cases} \dfrac{\partial l(\mu,\sigma^2)}{\partial \mu} = 0 \\ \dfrac{\partial l(\mu,\sigma^2;x)}{\partial \sigma^2} = 0 \end{cases} \Leftrightarrow \begin{cases} -\dfrac{1}{2\sigma^2}(-2\displaystyle\sum_{i=1}^{n} x_i + 2n\mu) = 0 \\ -\dfrac{n}{2}\dfrac{1}{\sigma^2} + \displaystyle\sum_{i=1}^{n}(x_i-\mu)^2\dfrac{1}{2\sigma^4} = 0 \end{cases} \Leftrightarrow$$

$$\begin{cases} \displaystyle\sum_{i=1}^{n} x_i - n\mu = 0 \\ -n\sigma^2 + \displaystyle\sum_{i=1}^{n}(x_i-\mu)^2 = 0 \end{cases} \Leftrightarrow \begin{cases} \mu = \dfrac{\sum_{i=1}^{n} x_i}{n} \\ \sigma^2 = \dfrac{\sum_{i=1}^{n}(x_i-\mu)^2}{n} \end{cases} \Leftrightarrow$$

$$\begin{cases} \mu = \overline{x} \\ \sigma^2 = \dfrac{n-1}{n}\dfrac{\sum_{i=1}^{n}(x_i-\mu)^2}{n-1} \end{cases} \Leftrightarrow \begin{cases} \mu = \overline{x} \\ \sigma^2 = \dfrac{n-1}{n}\, s^2 \end{cases}$$

Condições de segunda ordem:

$$\begin{cases} \dfrac{\partial^2 l(\mu,\sigma^2;x)}{\partial \mu^2} = -\dfrac{1}{2\sigma^2}2n = -\dfrac{n}{\sigma^2} < 0 \\ \dfrac{\partial^2 l(\mu,\sigma^2;x)}{\partial \sigma^4} = \dfrac{n}{2}\dfrac{1}{\sigma^4} - \displaystyle\sum_{i=1}^{n}(x_i-\mu)^2\dfrac{2\sigma^2}{2\sigma^8} = \dfrac{n}{2\sigma^4}\left[1 - 2\dfrac{n-1}{n}\dfrac{\sum_{i=1}^{n}(x_i-\mu)^2}{n-1}\dfrac{1}{\sigma^2}\right] < 0 \end{cases}$$

Portanto, os estimadores de máxima verosimilhança obtidos foram:

$$\begin{cases} \hat{\mu} = \overline{X} = \dfrac{\sum_{i=1}^{n} X_i}{n} \\ \hat{\sigma}^2 = \displaystyle\sum_{i=1}^{n}\dfrac{(X_i-\overline{X})^2}{n} = \dfrac{n-1}{n}\displaystyle\sum_{i=1}^{n}\dfrac{(X_i-\overline{X})^2}{n-1} = \dfrac{n-1}{n}S^2 \end{cases}$$

Exemplo 6.5 Seja $\mathbf{x} = (x_1, x_2, \cdots, x_n)$ uma amostra aleatória de uma distribuição de Poisson, tal que $X \sim Poisson(\lambda)$. Determinar o estimador de máxima verosimilhança (e.m.v.) para o parâmetro λ.

Resolução Seja X é uma v.a. com distribuição de Poisson de parâmetro λ. Tomemos uma amostra aleatória $\mathbf{x} = (x_1, x_2, \cdots, x_n)$ então a função de probabilidade de X é:

$$P(X = x) = f(x) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad x = 0, 1, 2, \cdots$$

A função verosimilhança é dada por:

$$L(\lambda; \mathbf{x}) = \prod_{i=1}^{n}\frac{\lambda^{x_i} e^{-\lambda}}{x_i!} = \frac{\lambda^{\sum_{i=1}^{n} x_i} e^{-n\lambda}}{\prod_{i=1}^{n} x_i!}$$

Para encontrar o estimador de máxima verosimilhança para λ, devemos encontrar o valor de λ para o qual a função de verosimilhança $L(\lambda; \mathbf{x})$ é máxima. Apliquemos a função logarítmica com o objetivo de isolar o parâmetro λ:

$$l(\lambda) = ln\, L(\lambda; \mathbf{x}) = ln \left[\frac{\lambda^{\sum x_i}\, e^{-n\lambda}}{\prod_{i=1}^{n} x_i!} \right] = ln\left[\lambda^{\sum x_i}\, e^{-n\lambda}\right] - ln\left[\prod_{i=1}^{n} x_i!\right]$$

$$= ln\left(\lambda^{\sum x_i}\right) + ln\left(e^{-n\lambda}\right) - \sum_{i=1}^{n} ln\, x_i!$$

$$= \sum_{i=1}^{n} x_i ln\, \lambda - n\lambda - \sum_{i=1}^{n} ln\, x_i!$$

Derivando a última expressão em ordem a λ e igualando o resultado a zero, obtemos:

$$\frac{d\, l(\lambda)}{d\, \lambda} = \frac{\sum_{i=1}^{n} x_i}{\lambda} - n = 0 \quad \Leftrightarrow \frac{\sum_{i=1}^{n} x_i}{\lambda} = n$$

$$\Leftrightarrow \frac{\sum_{i=1}^{n} x_i}{n} = \lambda$$

Assim, concluimos que o estimador do parâmetro λ é:

$$\hat{\lambda} = \frac{\sum_{i=1}^{n} X_i}{n} = \overline{X}$$

O estimador de máxima verosimilhança do valor médio μ é $\overline{X}$. Vejamos se $\overline{x}$ é um ponto de máximo:

$$\frac{d^2 l(\lambda)}{d\, \lambda^2} = -\frac{\sum_{i=1}^{n} x_i}{\lambda^2} < 0$$

como o segunda derivada é negativa então a função admite um máximo.

No quadro seguinte estão os estimadores de máxima verosimilhança para os parâmetros dos modelos estudados:

Modelo	Parâmetros	Estimadores
$Bin(1, p)$	p	$\hat{p} = \overline{X}$
$P(\lambda)$	λ	$\hat{\lambda} = \overline{X}$
$N(\mu, \sigma)$	$\mu, \quad \sigma$	$\hat{\mu} = \overline{X}, \quad \hat{\sigma} = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2}$

Vejamos agora as propriedades mais relevantes dos estimadores e que nos permitem escolher entre vários possíveis estimadores para um parâmetro o estimador que apresenta as melhores características.

Propriedade da Invariância

Se $\hat{\theta}$ é o estimador de máxima verosimilhança de θ e g é uma função bijectiva de θ, então o estimador de máxima verosimilhança de $g(\theta)$ é $g(\hat{\theta})$.

Estimador centrado

Um estimador $\hat{\theta}$ do parâmetro θ é centrado ou não enviesado se e só se $E(\hat{\theta}) = \theta$.

Viés ou Bias

Um estimador $\hat{\theta}$ de θ que não é centrado diz-se que é enviesado. O viés é dado por

$$E(\hat{\theta}) - \theta$$

Exemplo 6.6 Sejam $X_1, \cdots, X_n$ variáveis aleatórias independentes e identicamente distribuídas com $E(X_i) = \mu$ para $i = 1, 2, \cdots, n$. Os estimadores do tipo

$$\hat{\mu} = a_1 X_1 + \cdots + a_n X_n \ \text{ com } \ \sum_{i=1}^{n} a_i = 1$$

são centrados para para o valor médio, μ, qualquer que seja a distribuição.

Exemplo 6.7 Sejam $X_1, \cdots, X_n$ variáveis aleatórias independentes e identicamente distribuídas com $E(X_i) = \mu$ para $i = 1, 2, \cdots, n$. Mostre que $\overline{X}$ é um estimador centrado para o valor médio, μ.

Resolução Sejam $X_1, \cdots, X_n$ variáveis aleatórias independentes e identicamente distribuídas, com $E(X_i) = \mu \;\; i = 1, \cdots, n$.

$$\begin{aligned}
E[\overline{X}] &= E\left[\frac{X_1 + X_2 + \cdots + X_n}{n}\right] \\
&= \frac{1}{n}\left[E(X_1) + E(X_2) + \cdots + E(X_n)\right] \\
&= \frac{1}{n}\left[\mu + \mu + \cdots + \mu\right] \\
&= \frac{n\mu}{n} \\
&= \mu
\end{aligned}$$

Concluimos então, que o $E[\overline{X}] = \mu$, logo $\overline{X}$ é um estimador centrado para o parâmetro μ.

Exemplo 6.8 Mostre que

$$Var[\overline{X}] = \frac{\sigma^2}{n}$$

Quaisquer que sejam $X_1, \cdots, X_n$ variáveis aleatórias independentes e identicamente distribuídas, com $Var[X_i] = \sigma^2$ para $i = 1, 2, \cdots, n$.

Resolução

$$\begin{aligned}
Var[\overline{X}] &= Var\left[\frac{X_1 + X_2 + \cdots + X_n}{n}\right] \\
&= \frac{Var[X_1 + X_2 + \cdots + X_n]}{n^2} \\
&= \frac{Var(X_1) + Var(X_2) + \cdots + Var(X_n)}{n^2} \\
&= \frac{n\sigma^2}{n^2} \\
&= \frac{\sigma^2}{n}
\end{aligned}$$

Concluimos então, que $Var[\overline{X}] = \dfrac{\sigma^2}{n}$, isto é, a variância da média amostral é igual à variância populacional a dividir pela dimensão da amostra.

Estimador assintoticamente centrado

Uma sucessão de estimadores $\{\hat{\theta}_n\}$, para o parâmetro θ diz-se assintoticamente centrada

$$\lim_{n\to\infty} E(\hat{\theta}_n) = \theta$$

Erro quadrático médio

Num estimador não centrado $\hat{\theta}$ de θ uma medida da proximidade de $\hat{\theta}$ em relação a θ é dada pelo erro quadrático médio (EQM) definido por:

$$EQM\left(\hat{\theta}\right) = E\left[\left(\hat{\theta} - \theta\right)^2\right] = Var(\hat{\theta}) + \left[\text{viés}(\hat{\theta})\right]^2 = Var(\hat{\theta}) + \left[E(\hat{\theta}) - \theta\right]^2$$

Note que se $\hat{\theta}$ for centrado o seu EQM coincide com a variância do estimador.

Exemplo 6.9 Seja X_1 a única observação de uma distribuição de Bernoulli de parâmetro p. Considere os seguintes estimadores de p: $T_1 = X_1$ e $T_2 = \dfrac{X_1}{2}$.

a) Indique qual destes estimadores é centrado.

b) Determine as variâncias e os erros quadráticos médios para os estimadores propostos.

Resolução

a) T_1 é centrado pois o $E[T_1] = p$.
   T_2 não é centrado, ou seja, é enviesado com viés igual a $E[T_2] - p = -\dfrac{p}{2}$.

b) $Var[T_1] = p\,(1-p)\,;\ Var[T_2] = p\,(1-p)\,/4$. Logo, $Var[T_1] > Var[T_2]$.
   $EQM\,[T_1] = p(1-p)$; $EQM\,[T_2] = p/4$.

Consistência

Um estimador é consistente quando à medida que a dimensão da amostra, n, aumenta, os estimativas aproximam-se do verdadeiro valor do parâmetro.

Condições suficientes de consistência: se $\hat{\theta}$ é centrado ou assintoticamente centrado e $\lim_{n\to\infty} Var(\hat{\theta}) = 0$ então $\hat{\theta}_n$ é um estimador consistente de θ.

Eficiência

Entre dois estimadores centrados do mesmo parâmetro o **mais eficiente** é aquele que apresentar **menor variância**.

E entre dois estimadores assintoticamente centrados do mesmo parâmetro o **mais eficiente** é aquele que apresentar **menor** EQM.

Entre os estimadores centrados, o estimador mais eficiente é o que apresenta menor variância quando comparado com qualquer outro estimador centrado para o mesmo parâmetro (eficiência absoluta).

Características dos estimadores

Os estimadores obtidos pelo método dos momentos são de um modo geral consistentes, têm distribuição assintótica normal e não são os estimadores assintoticamente mais eficientes.

Os estimadores obtidos pelo método de máxima verosimilhança são, assintoticamente centrados, consistentes, invariantes, têm distribuição assintotica normal e são assintoticamente mais eficientes.

Exemplo 6.10 Seja X uma v.a. discreta que toma os valores $1, 2$, e 3 com probabilidade $\frac{1}{3}, \frac{1}{2}, \frac{1}{6}$, respetivamente. Considere todas as amostras de dimensão 2.

a) Determine a distribuição amostral da média;

b) Mostre que $\overline{X}$ é um estimador centrado para a média populacional μ.

c) Mostre que $Var(\overline{X}) = \frac{\sigma^2}{2}$

Resolução:

a) Cálculo da distribuição amostral da média:

A probabilidade de obter cada uma das amostras é dada por:

Tabela 6.1: Distribuição amostral da média

Amostra	Nº de Amostras	Probabilidade	Média Amostral ($\overline{x}$)
(1,1)	1	1/9	1
(1,2)	2	2/6	1.5
(1,3)	2	2/18	2
(2,2)	1	1/4	2
(2,3)	2	2/12	2.5
(3,3)	1	1/36	3

$$P((1,1)) = P(X_1 = 1)P(X_2 = 1) = \frac{1}{3} \times \frac{1}{3} = \frac{1}{9} = \frac{4}{36}$$

$$P((1,2)) = 2(P(X_1 = 1)P(X_2 = 2)) = 2 \times (\frac{1}{3} \times \frac{1}{2}) = 2(\frac{1}{6}) = \frac{12}{36}$$

$$P((1,3)) = 2(P(X_1 = 1)P(X_2 = 3)) = 2(\frac{1}{3} \times \frac{1}{6}) = 2(\frac{1}{18}) = \frac{4}{36}$$

$$P((2,2)) = P(X_1 = 2)P(X_2 = 2) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4} = \frac{9}{36}$$

$$P((2,3)) = 2(P(X_1 = 2)P(X_2 = 3)) = 2(\frac{1}{2} \times \frac{1}{6}) = 2(\frac{1}{12}) = \frac{6}{36}$$

$$P((3,3)) = P(X_1 = 3)P(X_2 = 3) = \frac{1}{6} \times \frac{1}{6} = \frac{1}{36}$$

A probabilidade de obter cada valor da média amostral é dada por:

$P(\overline{X} = 1) = P((1,1)) = 4/36$
$P(\overline{X} = 1.5) = P((1,2)) = 12/36$
$P(\overline{X} = 2) = P((1,3)) + P((2,2)) = 4/36 + 9/36 = 13/36$
$P(\overline{X} = 2.5) = P((2,3)) = 6/36$
$P(\overline{X} = 3) = P((3,3)) = 1/36$

$$\overline{X} : \begin{cases} 1 & 1.5 & 2 & 2.5 & 3 \\ \frac{4}{36} & \frac{12}{36} & \frac{13}{36} & \frac{6}{36} & \frac{1}{36} \end{cases}$$

b) Vamos agora mostrar que $E(\overline{X}) = \mu$.

O valor de $E(\overline{X})$ é dado por:

$$E(\overline{X}) = 1 \times \frac{4}{36} + 1.5 \times \frac{12}{36} + 2 \times \frac{13}{36} + 2.5 \times \frac{6}{36} + 3 \times \frac{1}{36} = 1.083$$

$$X : \begin{cases} 1 & 2 & 3 \\ \frac{1}{3} & \frac{1}{2} & \frac{1}{6} \end{cases}$$

A média populacional é igual a:

$$\mu = 1 \times \frac{1}{3} + 2 \times \frac{1}{2} + 3 \times \frac{1}{6} = 1.083$$

Concluimos então que $E(\overline{X}) = \mu$.

c) Pretendemos mostrar que $Var(\overline{X}) = \frac{\sigma^2}{2}$.

A variância populacional é dada por:

$$\sigma^2 = (1 \times \frac{1}{3} + 4 \times \frac{1}{2} + 9 \times \frac{1}{6}) - (1.083)^2 = 0.484$$

Para determinar o valor da variância da média amostral, vamos começar por calcular $E(\overline{X}^2)$, que é dado por:

$$E(\overline{X}^2) = 1 \times \frac{4}{36} + 1.5^2 \times \frac{12}{36} + 4 \times \frac{13}{36} + 2.5^2 \times \frac{6}{36} + 9 \times \frac{1}{36} = 3.57$$

para obtermos o valor da variância da média amostral, calculamos:

$$\begin{aligned} Var(\overline{X}) &= E(\overline{X}^2) - (E(\overline{X}))^2 \\ &= 3.57 - 1.083^2 \\ &= 0.242 \end{aligned}$$

Concluimos então que,

$$Var(\overline{X}) = \frac{\sigma^2}{2} = \frac{0.484}{2} = 0.242$$

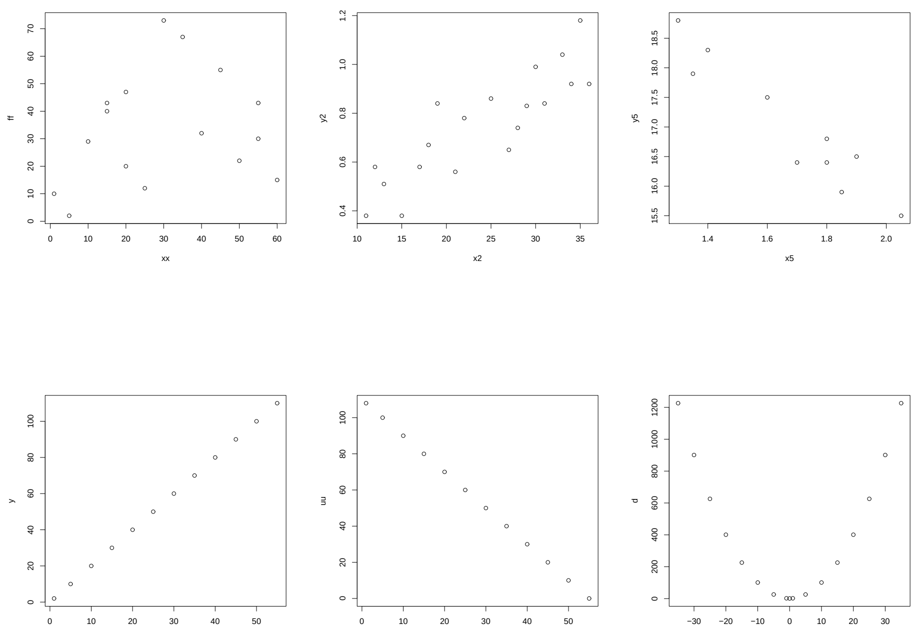
como pretendiamos mostrar.

6.3 Dados Bidimensionais

6.3.1 Correlação

Consideremos uma amostra bivariada (x_i, y_i), $\quad i = 1, \cdots, n$, resultante da observação de duas variáveis x e y. Com o objetivo de identificar a tendência e relação entre as duas variáveis de interesse, comecemos por realizar uma representação gráfica dos pontos (x_i, y_i) num sistema de eixos perpendiculares a que chamamos diagrama de dispersão.

Vejamos alguns exemplos de possíveis relações e tendências entre os pares de valores correspondentes às observações das variáveis x e y.



Se os valores de ambas as variáveis apresentam o mesmo sentido, isto é, ambas aumentam ou diminuem, temos uma associação positiva. No caso de apresentarem sentidos inversos, isto é, uma variável aumenta e a outra diminui, temos uma associação negativa.

Vejamos brevemente a classificação das variáveis que estão associadas às características em estudo:

Variável qualitativa - quando apresenta várias categorias.

- **Escala nominal**: se a ordem das categorias não tem significado (ex. estado civil, grupo sanguíneo).
- **Escala ordinal**: se a ordem das categorias tem significado (ex.acidez do azeite, nível económico de uma população).

Variável quantitativa - resulta da medição de uma certa quantidade.

- **variável contínua**: assume qualquer valor num intervalo real.
- **variável discreta**: assume valores num subconjunto finito ou uma infinidade numerável de valores.

Para as variáveis quantitativas podemos definir

- **Escala intervalar**: os valores numéricos possuem ordem e é possível falar em diferença entre pontos da escala. Não há um zero absoluto, isto é, o zero não significa a ausência da característica (ex. escala de temperatura em graus Fahrenheit).
- **Escala percentual ou de razão**: tem as mesmas características da escala anterior. Há um zero absoluto, isto é, o zero significa a ausência da característica (ex. peso, volume).

6.3.2 Coeficiente de correlação amostral de Pearson

Para quantificarmos o grau da relação de linearidada entre as duas variáveis vamos usar uma estatística conhecida como, coeficiente de correlação amostral de Pearson:

$$r = \frac{1}{n-1}\sum_{i=1}^{n}\frac{x_i - \overline{x}}{s_x}\frac{y_i - \overline{y}}{s_y}$$

Observações

Este coeficiente é aplicado a dados quantitativos (escala intervalar ou percentual);

$-1 \leq r \leq 1$;

$r = \pm 1$ se e só se existir uma relação linear perfeita entre as duas variáveis x e y, definida por $y_i = a + bx_i$;

$r = 0$ significa ausência de relação linear entre as duas variáveis, podendo existir uma relação não linear entre as duas variáveis;

Para transformações lineares dos dados r é invariante;

O valor de r permanece igual permutando as duas variáveis;

Variáveis independentes têm correlação nula;

Correlação forte pode não significar uma relação de causa e efeito entre as variáveis;

Correlação nula significa ausência de relação linear, não invalida outro tipo de relação funcional entre as variáveis.

No software R: $cor(x, y)$ ou $cor(x, y, method = "pearson")$.

6.3.3 Coeficiente de correlação ordinal de Spearman

Quando pelo menos uma das duas variáveis se encontra em escala ordinal, para medir a associação entre elas usa-se o coeficiente de correlação ordinal de Spearman

Cada par (x_i, y_i) é substituído pelas respetivas ordens $(ord(x_i), ord(y_i))$ e $d_i = ord(x_i) - ord(y_i)$

$$r_S = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$$

Observações

Este coeficiente de correlação é aplicado a dados que estejam em escala ordinal, intervalar ou percentual.

A aplicação deste coeficiente de correlação a dados em que apenas uma das variáveis está em escala ordinal, obriga à conversão da outra variável numa escala ordinal;

$-1 \leq r \leq 1$;

$r = 1$ corresponde a uma mesma ordenação;

$r = -1$ corresponde a uma ordenação contrária;

No software R: $cor(x, y, method = "spearman")$.

6.3.4 Coeficiente de correlação τ de Kendall amostral

Outra medida de associação entre duas variáveis em escala ordinal, intervalar ou percentual é o coeficiente de correlação τ de Kendall.

$$\tau = \frac{1}{\binom{n}{2}} \sum_{i<j} a_{ij} \qquad a_{ij} = \begin{cases} 1, & \text{se} \quad (x_i, y_i) \text{ e } (x_j, y_j) \text{ concordantes} \\ -1, & \text{se} \quad (x_i, y_i) \text{ e } (x_j, y_j) \text{ discordantes} \\ 0, & \text{caso contrário} \end{cases}$$

Os pares de dados (x_i, y_i) e (x_j, y_j) são:

concordantes se $x_i < x_j$ e $y_i < y_j$ ou $x_i > x_j$ e $y_i > y_j$

discordantes se $x_i < x_j$ e $y_i > y_j$ ou $x_i > x_j$ e $y_i < y_j$

Observações

Este coeficiente de correlação é aplicado a dados que estejam em escala ordinal, intervalar ou percentual.

$-1 \leq \tau \leq 1$;

$\tau = 1$ concordância perfeita, $\quad \tau = -1$ discordância perfeita.

No software R: $cor(x, y, method = "kendall")$.

6.4 Regressão linear simples

Vamos analisar agora uma situação de estudo em que estão presentes duas variáveis quantitativas, que designamos por x e y e em que pode verificar-se uma relação funcional entre elas. A variável x vai representar a variável independente (existem estudos em que esta variável é controlada pelo experimentador) e a variável y representa a variável dependente.

Através de um diagrama de dispersão é possível verificar a existência de uma possível relação funcional entre x e y. Se essa relação for do tipo linear, os pontos encontram-se dispersos aleatoriamente em torno de uma reta, e o ajuste de um modelo de regressão linear de y em x será adequado aos dados.

Na regressão linear simples a relação entre as variáveis x e y é da forma $y = a + b\,x$. Se o valor de y estiver afetado de um erro aleatório, e, então escrevemos $y = a + b\,x + e$.

Para um conjunto de dados estatísticos (x_i, y_i) $i = 1, \cdots, n$ temos então

$$y_i = a + b\,x_i + e_i, \quad i = 1, \cdots, n$$

O resíduo, também designado como erro ou desvio, e_i associado à $i-$ésima observação (x_i, y_i), é igual à diferença entre o valor de y_i e o valor $\widehat{y}_i = a + bx_i$ ajustado pelo modelo:

$$y_i = a + bx_i + e_i = \widehat{y}_i + e_i$$

De um modo intuitivo podemos escrever que

dados = ajustamento + resíduos

As constantes a e b são designadas como coeficientes de regressão. Para ajustarmos uma reta de regressão a um conjunto de dados é necessário conhecer estimadores para a e b que serão obtidos por aplicação do método dos mínimos quadrados a apresentar a seguir.

Nota: o termo linear no modelo de regressão está associado aos coeficientes, donde todo o polinómio do tipo $\widehat{y} = a_0 + a_1 x + a_2 x^2 + \cdots + a_p x^p$ corresponde a uma regressão linear com os coeficientes (parâmetros) $a_0, a_1, \cdots, a_p$ e a variável x.

6.4.1 Método dos mínimos quadrados

O método dos mínimos quadrados permite obter os melhores estimadores para os coeficientes de regressão a e b.

Os estimadores para a e b são obtidos pela minimização da função de duas variáveis $F(a,b)$, sendo esta função dada por:

$$F(a,b) = \sum_{i=1}^{n} (y_i - (a + bx_i))^2 = \sum_{i=1}^{n} (y_i - \widehat{y}_i)^2 = \sum_{i=1}^{n} e_i^2$$

Se existir solução do sistema esta corresponde a um mínimo.

$$\begin{cases} \dfrac{\partial\, F(a,b)}{\partial\, a} = 0 \\ \dfrac{\partial\, F(a,b)}{\partial\, b} = 0 \end{cases}$$

A minimização da função $F(a,b)$ (corresponde à minimização da soma dos quadrados dos resíduos) tem como solução

$$\begin{cases} a = \overline{y} - b\,\overline{x} \\ b = \dfrac{\displaystyle\sum_{i=1}^{n} x_i y_i - n\,\overline{x}\,\overline{y}}{(n-1)s_x^2} \end{cases}$$

Nota: A prova de que a solução do sistema é um mínimo, exige mostrar que a matriz Hessiana (matriz quadrada das derivadas de segunda ordem de F) é semidefinida positiva.

6.4.2 Qualidade do ajustamento

A qualidade do ajustamento da reta de regressão é medida pela decomposição da variância total dos dados y_i (s_y^2) ou da soma dos quadrados total $\text{SST} = \displaystyle\sum_{i=1}^{n} (y_i - \overline{y})^2$:

$$\sum_{i=1}^{n} (y_i - \overline{y})^2 = \sum_{i=1}^{n} (\widehat{y}_i - \overline{y})^2 + \sum_{i=1}^{n} (y_i - \widehat{y}_i)^2$$

$$\text{SST} = \text{SSA} + \text{SSE}$$

se igualdade anterior for dividida por $(n-1)$ obtemos a seguinte igualdade

$$s_y^2 = s_{\widehat{y}}^2 + s_e^2$$

que em linguagem corrente se traduz como: variância dos dados y_i = variância explicada pela regressão de y em x + variância residual.

Observações:

- SSE $= 0$, o ajustamento é perfeito, ou seja, a relação linear entre as variáveis é perfeita;
- SSE = SST , o ajustamento linear não é adequado;
- 0 <SSE<SST, o ajustamento é classificado como mediano;

$$1 - \frac{SSE}{SST} = \frac{SSA}{SST} = r^2$$

- r^2 é o coeficiente de determinação ($0 \leq r^2 \leq 1$);
- $b = r\dfrac{s_y}{s_x}$, onde r é o coeficiente de correlação de Pearson;
- $$\frac{SSA}{SST} = \frac{s_{\widehat{y}}^2}{s_y^2}$$

  representa a fração da variância total que é devida ao ajustamento do modelo de regressão linear;
- quanto maior for SSA melhor é o ajustamento;
- quanto mais próximo de 1 estiver o quociente melhor é o ajustamento;
- $\sum e_i = 0$; $\sum y_i = \sum \widehat{y}_i$;
- a reta dos mínimos quadrados é muito sensível à presença de outliers (se possível devem ser excluídos do ajustamento);
- o par $(\overline{x}, \overline{y})$ pertence à reta de regressão;
- a identificação correta da variável independente (controlada) x e da variável dependente (resposta) y é muito importante;
- as retas de regressão de y em x e de x em y não coincidem;
- a previsão do modelo de regressão linear para x^* é dado por $\widehat{y} = a + bx^*$.

6.4.3 Análise de Resíduos

A inferência estatística baseada no modelo de regressão linear assenta no pressuposto que os erros de ajustamento são normais, com valor médio nulo, não estão correlacionados e têm variância constante. Assim, numa análise gráfica dos resíduos devemos observar que estes:

- não devem apresentar padrões ou tendências: uma representação gráfica dos pontos (x_i, e_i) deve ter um aspeto aleatório;
- devem estar numa banda horizontal, dado que a variância deve ser constante;
- devem formar uma nuvem de pontos simétrica em relação ao eixo dos xx uma vez que a média deve ser próxima de 0.

Vamos mostrar com um exemplo a aplicação do modelo de regressão linear resolvido apenas no R.

Exemplo 6.11 Considere o seguinte conjunto de dados bivariados

x	10	10	11	11	12	15	17	19	20	20	23	25	27	30
y	21.1	19.9	22.5	23.7	25.0	30.3	36.1	38.6	41.5	42.7	45.0	50.0	53.9	62.1

a) Esboce o diagrama de dispersão para os pontos (x, y).

b) Determine os valores de diferentes coeficiente de correlação. Justifique qual o mais adequado.

c) Estime a reta de regressão linear.

d) Avalie a qualidade do ajustamento da regressão linear.

e) Estime o valor de y para $x = 18$.

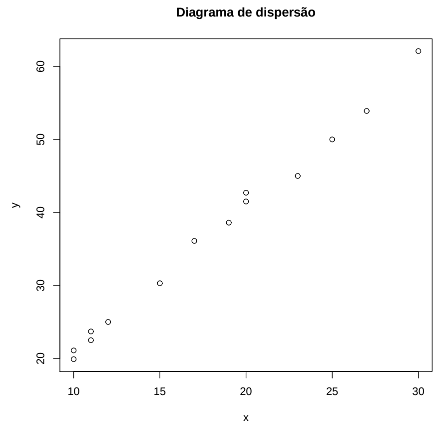
Resolução:

A resolução deste exemplo será realizada apenas com o software R.

a) O diagrama de dispersão para os pontos (x, y) mostra que os dados apresentam uma disposição em quase linha reta, pelo que o ajuste de uma reta de regressão é adequada.

   No R:

```
> x<-c(10, 10, 11 ,11, 12 ,15, 17, 19, 20, 20, 23, 25 ,27, 30)
> y<-c(21.1, 19.9, 22.5, 23.7, 25.0, 30.3, 36.1, 38.6, 41.5,
+ 42.7, 45.0 ,50.0,53.9, 62.1)
> plot(y~x,xlab="x",ylab="y",main="Diagrama de dispersao")
```



b) Cálculo dos diferentes coeficientes de correlação no R:

```
> cor(x,y,method="pearson")
[1] 0.9969425
> cor(x,y,method="spearman")
[1] 0.9966978
> cor(x,y,method="kendall")
[1] 0.9833783
```

O coefiente de correlação de Pearson é o mais adequado para o tipo de dados do exemplo, dado que ambas as variáveis são quantitativas e essa informação é usada no cálculo do coeficiente. Este coeficiente quantifica a relação linear entre as duas variáveis e o valor obtido foi de 0.997 o que indica uma correlação forte entre as variáveis. Em relação aos outros dois coeficientes observamos que o coeficiente de Spearman é aproximadamente igual ao de Pearson, enquanto o coeficiente de Kendall apresenta o valor menor de 0.983.

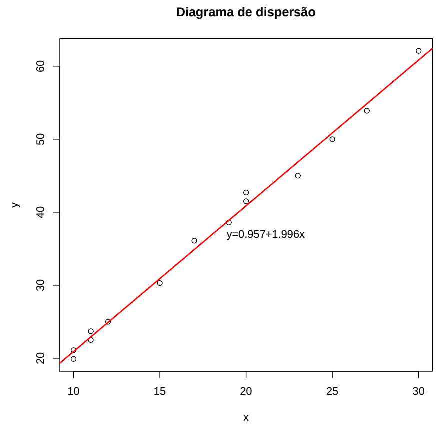
c) Estime a reta de regressão linear.

A reta estimada pelo método dos mínimos quadrados é $\hat{y} = 0.957 + 1.996x$.

No R:

```
> aj<-lm(y~x)
> aj
Call:
lm(formula = y ~ x)

Coefficients:
```




```
(Intercept)           x
     0.9574      1.9960
```

d) A avaliação da qualidade do ajustamento da reta de regressão linear, $y = 0.957 + 1.996x$ vai ser feita usando o diagrama de dispersão dos resíduos e o coeficiente de determinação. Da análise do diagrama de dispersão de resíduos resultam os seguintes comentários:

-os pontos (x_i, e_i) apresentam um aspeto aleatório;

-estão dispostos numa banda horizontal;

-formam uma nuvem de pontos simétrica em relação ao eixo dos xx uma vez que a média deve ser próxima de 0.

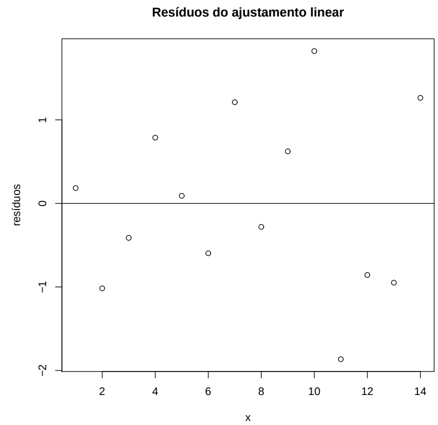
-apresentam uma pequena dispersão com uma amplitude amostral dos resíduos (≈ 3.69).

Concluimos então que o diagrama dos resíduos apresenta a configuração desejada para um bom ajustamento.

Cálculo do coeficiente de determinação:

No R:

```
> m<-mean(y)
> sst<-sum((y-m)^2)
>  ssa<-sum((fitted(aj)-m)^2)
> r2<-ssa/sst
> r2
[1] 0.9938944
```



Observamos que o coeficiente de determinação r^2 é próximo de 1 (≈ 0.9939), o que é indicativo de um bom ajustamento.

e) O valor estimado de y para $x = 18$ é dado por:

No R:

```
>predict(aj,list(x=18))
       1
36.88514
```

Concluimos que o valor estimado de y para $x = 18$ é $\hat{y} = 36.9$.

Capítulo 7

Conclusões e Trabalho Futuro

7.1 Conclusões

Nesta secção são apresentadas as conclussões e sugestões para o trabalho na sala de aula de Probabilidade e a Estatística. O desenvolvimento do pensamento estatístico do aluno será facilitado se for introduzido o ensino da estatística a partir de 7º ano e 8º ano do ensino básico. Esta antecipação segue a tendência mundial e melhora depois o ensino das Probabilidades e Estatística no 12º ano do ensino secundário.

As Probabilidades e a Estatística no ensino secundário podem ainda valorizar o estudo de matemática discreta, muitos vezes relegada para segundo plano, através da simulação de experiências aleatórias que envolvem variáveis aleatórias discretas e a comparação dos resultados amostrais com os teóricos, obtidos a partir dos processos de contagem previstos no Cálculo das Probabilidades.

A introdução ao pensamento probabilístico já realizado no ensino secundário é importante para o aluno adquirir uma correta intuição probabilística e ao ingressar no curso superior não tenha uma intuição viciada dos fenômenos aleatórios discretao dada a pouca familiaridade com as variações amostrais e com o estudo dos fenômenos aleatórios em geral.

Na vida quotidiana o volume de dados a que se tem acesso aumentou significativamente. Uma educação que favoreça o exercício consciente da cidadania, exige necessariamente que os alunos tenham uma formação estatística que lhes permita desenvolver o seu raciocínio crítico na análise de dados, interpretação de gráficos, de médias e de outras informações estatísticas e utilizarem a Estatística como um instrumento de tomada de decisão.

O software R revelou-se uma ferramenta de aprendizagem poderosa para visualizar e interpretar os dados com os alunos a poderem chegar às conclusões de um modo autónomo.

No presente trabalho sugerem-se algumas atividades didáticas na sala de aula:

1. O conceito de probabilidade é introduzido na sala de aula com grupos de 2 ou 3 alunos a realizar expeiências com moedas e dados;

2. Os alunos da turma registam a informação relevante sobre eles próprios, como por exemplo, a cor dos olhos de cada aluno e dos respetivos pais. O objetivo é saber se a relação de depedência linear entre a cor dos olhos dos pais e dos filhos é ou não linear;

3. Os alunos listam as suas idades. O objetivo é estudar algumas propriedades como por exemplo valor mínimo e máximo, média, mediana e moda;

4. Os alunos apresentam os dados das idades numa tabela para calcular a frequência relativa e cumulativa;

5. Os alunos registam as notas da uma disciplina de duas turmas do mesmo ano e do mesmo professor para analisar se as turmas têm o mesmo aproveitamento;

6. Os alunos analisam a dificuldade do exame nacional da disciplina Matemática recolhendo informação sobre as notas de exame nacional do ano letivo anterior e as respetivas notas do exame nacional de Matemática;

7. Nas distribuições de probabilidades, os alunos analisam os conjuntos de dados através de tabelas ou de histogramas, para associar os modelos matemáticos mais adequados;

8. A Estatística Inferêncial (estimações, intervalos de confiança e testes de hipóteses) são introduzidos a partir de exemplos contextualizados nas áreas de interesse dos alunos, sempre que possível com dados reais, fornecidos pelos próprios alunos;

9. Os alunos devem utilizar o software R para a apresentação, visualização e intrepretação dos dados de modo a familizarem-se com as tecnologias.

Estas sugestões para o ensino das Probabilidades e da Estatística exigem que as escolas secundários disponham de equipamentos (Computadores) com o software R instalado e os professores preparados para ensinar com recurso a este software.

Esta tese pretende ser um recurso didático para a formação em Estatística dos professores.

7.2 Trabalho Futuro

No futuro um dos objetivos será a realização de formações para os professores do nível secundário de modo a conhecerem e usarem o R na aprendizagem das Probabilidades e da Estatística.

Dinamizar a criação de laboratórios computacionais na rede do ensino em Timor-Leste.

O sucesso de qualquer formação depedente essencialmente do formador (professor), do seu empenho, dedicação e conhecimentos.

Um estudo sobre o nível de utilização e aceitação do software R é também importante para propôr novas estratégias de desenvolvimento do ensino das Probabilidades e da Estatística.

Bibliografia

1. Athayde, M. E., (2013). *Estatística. R*, Escola de Ciências, Universidade do Minho, Braga, Portugal.

2. Azevedo, Cecília, (2004). *O que é a probabilidade? Interpretações da probabilidade*, Departamento de Matemática, Universidade do Minho, Braga, Portugal.

3. Fernandes, J. P., (1999). *Intuições e Aprendizagem de Probabilidades, Uma Proposta de ensino de Probabilidade no* 9º *Ano de Escolaridade*, Tese de Doutoramento em Educação, Universidade do Minho, Braga, Portugal.

4. Fernandes, J.P. & Barros, P.M., (2005). *Dificuldades em estocástica de uma futura professora do* 1º *e* 2º *ciclos do Ensino Básico.* Revista Portuguesa da Educação, Vol. 18 número 1, Instituto de Educação e Psicologia, Universidade do Minho, Braga, Portugal.

5. Garfield, J. & Chance, B., (2000),*Assesment in Statistics Education: Issues and Challanges*, p.102.

6. Garfield, J. & Ahlgren, A., (1988). *Difficulties in Learning Basic Concepts in Probability and Statistics: Implication for Research*, Journal for Research in Mathematics Education,Vol. 19,No. 1, pp.44-43.

7. Ghinis, D., Korres, K. & Bersimis, S., (2009). *Difficulties Greek Senior High School Students Identify in Learning and the Teaching of Statistics: The case of Experimental and Private High Schools*, University of Piraeus, Greece.

8. Katz, V. J., (2010). *História da Matemática*, Fundação Calouste Gulbenkian, Lisboa, Portugal.

9. Ministério da Educação, (2008). *Artigo* 14º *Lei Bases da Educação*, http://www.me.gov.tl/lei-de-base-da-educao.

10. Ministério da Educação de Timor-Leste, (2011). *Plano Curricular do Ensino Secundário Geral.*

11. Ministério da Educação de Timor-Leste, (2014). *Matemática*, Manual do Aluno, 12º ano de escolaridade, Universidade de Aveiro, Portugal.

12. Ministério da Educação e Ciência, (2013). *Matemática A, Questões de Exames Nacional e de Teste Intermédios do 12º Ano 1997-2013*, Volume I, Probabilidade e Combinatória, Editorial do, Lisboa, Portugal.

13. Martins, M. E. G., Monteiro, C., Viana, J. P. & Turkman, M. A. A., (1997). *Estatística: Matemática*, 10º ano de escolaridade, Ministério da Educação, Departamento do Ensino Secundário, Lisboa, Portugal.

14. Muenchen, R. A., (2011). R for SAS and SPSS Users, Statistics and computing, Second Edition, Springer, London.

15. Murteira, B., Ribeiro, C. S., Silva, J. A. & Pimenta, C., (2010). *Introdução à Estatística*, Escolar Editora, Lisboa, Portugal.

16. Pestana, D. D. & Velosa, S. F., (2010). *Introdução à Probabilidade e à Estatística*, Volume 1, 4ª Edição, Fundação Calouste Gulbenkian, Lisboa, Portugal.

17. Ponte, J. P., (1991). *O Computador no Ensino de Matematica, Um Processo de Investigação e Formação de Professores*, Faculdade da Ciência de Universidade de Lisboa, Lisboa, Portugal.

18. Ponte, J. P. & Fonseca, H., (2001). *Orientações Curriculares para o Ensino da Estatística análise comparativo de três paises*, Faculdade da Ciência de Universidade de Lisboa, Lisboa, Portugal.

19. Stordahl, K., (2007). *The History Behind the Probability Theory and the Queuing Theory.*

20. Stuart, T., (1995). *Changing the Teaching of Statistics*, Source: The Statistician, Vol. 44, No. 1, pp. 45-54.

21. Torgo, L., (2009). *A Linguagem R*, Programação para a análise de dados, Escolar Editora, Lisboa, Portugal.

Anexo A

Do programa de Matemática do 12º de Timor-Leste foi transcrita a unidade temática 9: Organização e Tratamento de Dados, que se refere aos conteúdos abordados neste trabalho.

As probabilidades fornecem conceitos e métodos para estudar casos de incerteza e para interpretar previsões baseadas em incertezas.

Este estudo, que pode ser em grande parte experimental, fornece uma base conceptual que capacita para interpretar, de forma crítica toda a comunicação que utiliza a linguagem das Probabilidades, bem como a linguagem Estatística. As técnicas de contagem que aqui aparecem como auxiliar do cálculo de probabilidades constituem uma aprendizagem significativa por si só, especialmente por se desenvolverem as capacidades do raciocínio e as conexões matemáticas e menos a aplicação das fórmulas.

Considera-se ainda que o tema das Probabilidades constitui uma boa oportunidade para a introdução de uma axiomática, uma das formas de organizar uma teoria matemática, permitindo que os estudantes tenham uma melhor compreensão do que é a atividade demonstrativa em Matemática. Por outro lado qualquer destes assuntos é bom para prosseguir objetivos de trabalho em aspetos da História da Matemática.

Abordagem da Estatística e das Probabilidades completará as aprendizagem básicas, com algumas novas noções e ferramentas que não podiam ser compreendidas no ensino pré-secundário. A Estatística é uma área favorável ao desenvolvimento de certas capacidades expressas nos currículos, tais como interpretar e intervir no real; formular e resolver problemas; manifestar rigor e espírito crítico. Outro aspeto importante no ensino da Estatística é a compreensão da importância da ciência e da investigação como um meio de resolver problemas do homem e obter benefícios para a sociedade.

Subtema 1: Probabilidades

Conteúdos	Metas de aprendizagem	Atividades práticas
Experiencia aleatória. Conjunto de resultados. Acontecimentos Classificação de acontecimentos Operações com acontecimentos. Aproximações concetuais para probabilidade: Aproximação frequencista Definição clássica ou de Laplace Definição axiomática (caso finito) Propriedades da probabilidade. Probabilidade condicionada e independência	O aluno: • Define experiência aleatória; • Identifica o conjunto de resultados ou espaço amostral de uma experiência aleatória; • Identifica e classifica diferentes tipos de acontecimentos; • Realiza operações com acontecimentos; • Estabelece a aproximação frequencista de probabilidade; • Utiliza a lei de Laplace na resolução de problemas; • Determina a probabilidade condicionada de dois ou mais acontecimentos; • Identifica acontecimentos independentes.	Determinar o espaço amostral de uma experiência aleatória. Comparar e classificar diferentes tipos de acontecimentos. Resolver exercícios que envolvam operações com acontecimentos. Resolver problemas usando as diferentes abordagens concetuais para a probabilidade. Determinar a probabilidade condicionada de dois ou mais elementos. Identificar acontecimentos independentes.

Subtema 2: Estatística descritiva e indutiva

Conteúdos	Metas de aprendizagem	Atividades práticas
Estatística - Generalidades Organização e interpretação de caracteres estatísticos (qualitativos e quantitativos) Medidas de tendência central e medidas de dispersão Distribuições bidimensionais (abordagem gráfica e intuitiva) Reta de regressão Variáveis discretas e contínuas Função massa de probabilidade Distribuição de probabilidades Modelo Binomial Modelo de Poisson Modelo Normal	O aluno: • Organiza os dados em tabelas e gráficos; • Representa graficamente os dados em estudo; • Determina as medidas de tendência central e de dispersão de uma amostra; • Retira conclusões das medidas encontradas; • Constrói a reta de regressão; • Identifica variáveis aleatórias discretas e contínuas; • Constrói a função massa de probabilidade de uma distribuição dada; • Utiliza os modelos Binomial, Poisson e Normal; • Resolve problemas usando os modelos referidos.	Organizar em tabelas os dados obtidos. Representar graficamente os dados obtidos. Resolver exercícios para determinar as medidas de tendência central e as de dispersão, analisando convenientemente os resultados obtidos. Elaborar diagramas de dispersão e a respetiva regra de regressão e realizar predições usando a reta de regressão Resolver exercícios usando variáveis aleatórias discretas e contínuas. Usar os modelos estabelecidos na resolução de problema.

Anexo B

Processo de Instalação do R: Siga as instruções de instalação apresentados a seguir escolhendo a opção assinalada pela seta encarnada.

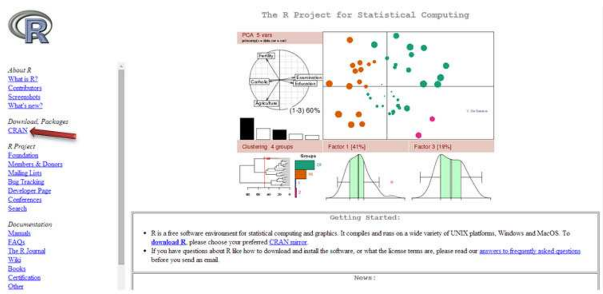


Figura 7.1: Passo 1

Figura 7.2: Passo 2

Figura 7.3: Passo 3

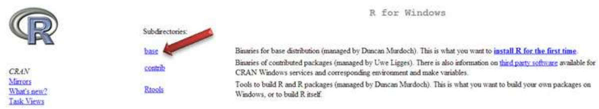


Figura 7.4: Passo 4

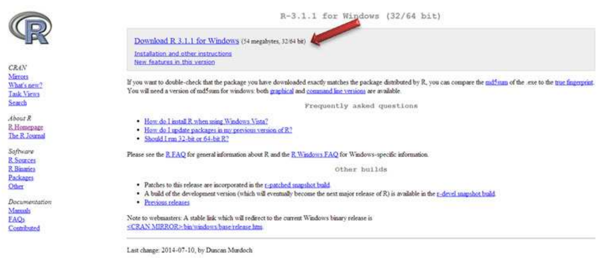


Figura 7.5: Passo 5

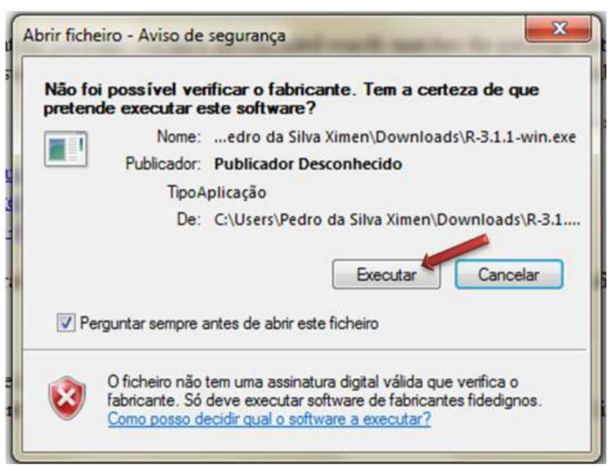


Figura 7.6: Passo 6

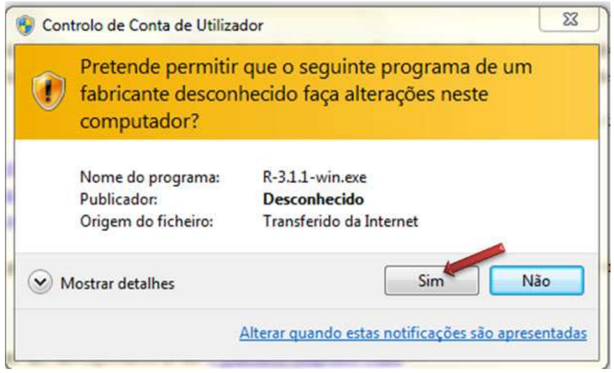


Figura 7.7: Passo 7

Figura 7.8: Passo 8

Figura 7.9: Passo 9

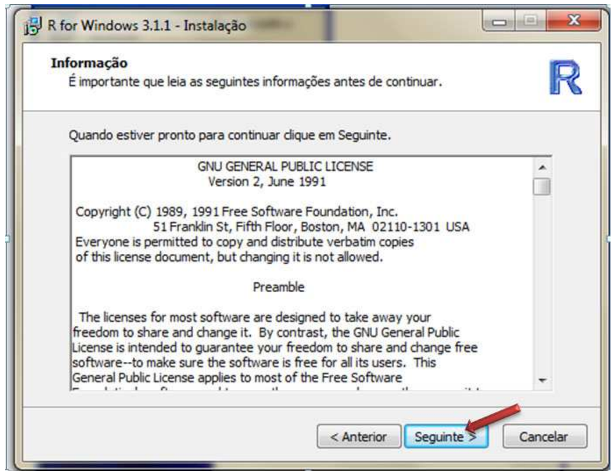


Figura 7.10: Passo 10

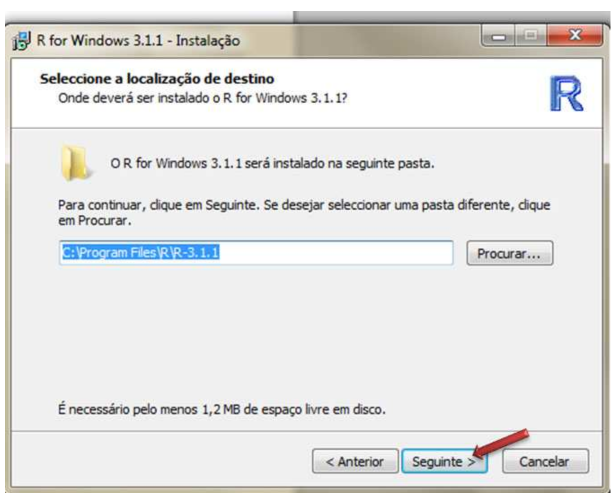


Figura 7.11: Passo 11

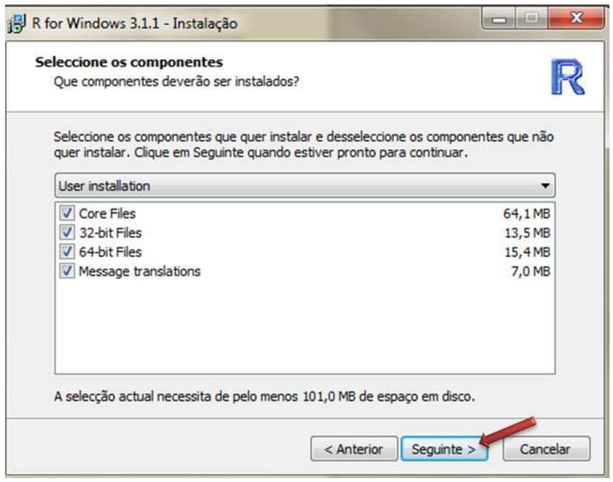


Figura 7.12: Passo 12

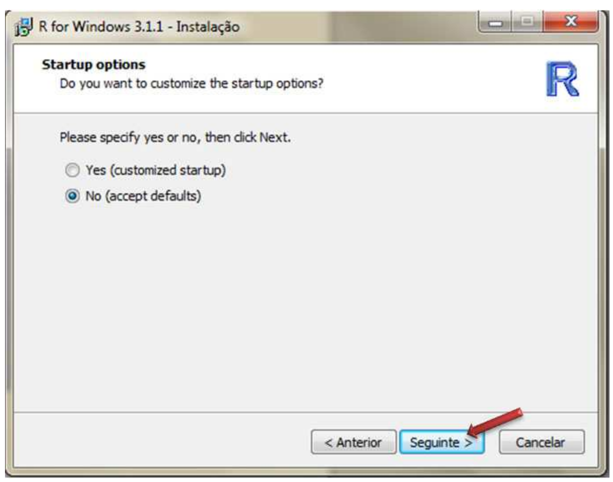


Figura 7.13: Passo 13

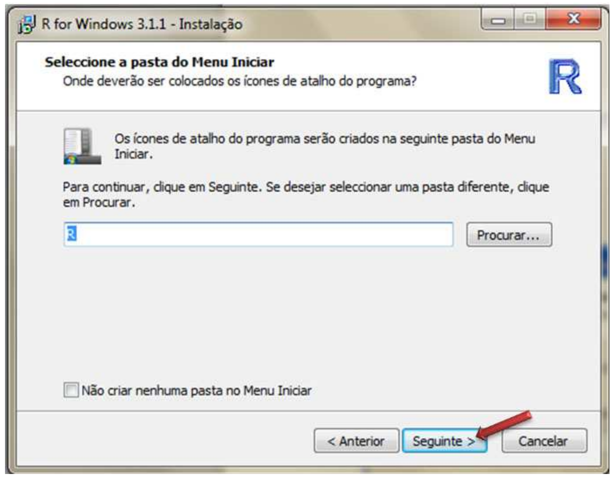


Figura 7.14: Passo 14

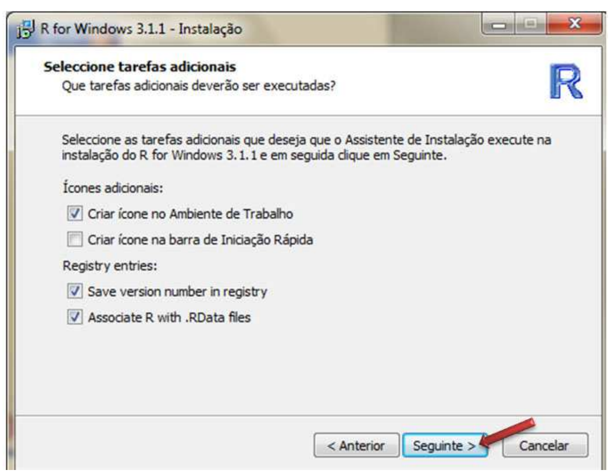


Figura 7.15: Passo 15

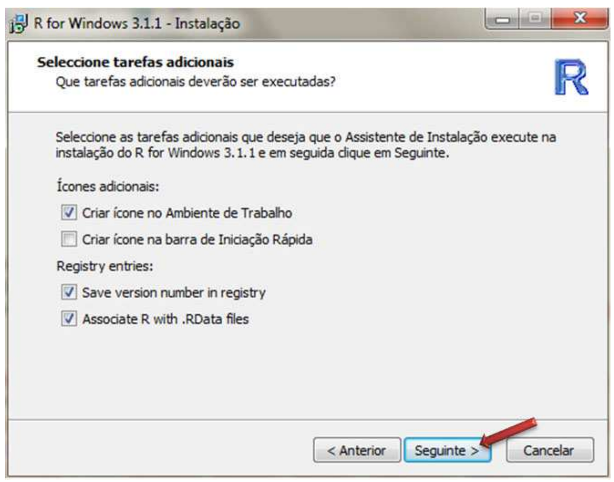


Figura 7.16: Passo 16

Anexo C

Apresentamos uma lista de funções básicas, comandos para gráficos e distribuições de probabilidade, pertencentes aos pacotes básicos do R.

FUNÇÕES BÁSICAS

- combinação das teclas `CTRL+L`: limpa todos os comandos da consola
- `rm(x,y)`: apaga os objetos x e y
- `NA`: dado ausente (not available)
- `help(comandoX)`: retorna ajuda sobre o comando X
- `library()`: lista todas as packages instalados
- `library(help=nome)`: dá informação sobre a package "nome"
- `help(package=datasets)`: obter ajuda sobre (p.ex) a package datasets
- `library(name)`: carrega a package "nome"
- ls(`package:nome`): lista as funções e operações da package "nome"
- `attach(nome)`: adiciona a package "nome"ao path do R
- `sum(x)`: soma todos os elementos de um objeto x
- `mean(x,opcoes)`: média amostral
- `quantile(x,p)`: quantil-p amostral
- `summary(x)`: valores de $x_{(1)}, q_{1/4}, q_{1/2}, \overline{x}, q_{3/4}, x_{(n)}$
- `fivenum(x)`: valores de extremos, quartos e mediana (por ordem crescente)
- `boxplot.stat(x)`: valores dos outliers e das estatísticas associadas ao diagrama

- `length(x)`: retorna o comprimento de um objeto x
- `rep(x,n)`: repete o número x, n vezes
- `seq(a,b,by=c)`: gera uma sequência de números contidos entre a e b, distantes c unidades um do outro.
- `table(x)` retorna uma tabela com as frequências absolutas de ocorrência da cada elemento de x
- `sort(x)`: ordena os elementos de x
- `sort(X,decreasing=T)`: ordena os elementos de X a decrescer
- `rank(x)`: ordem de cada elemento de x
- `range(x)`: extremos da amostra $x_{(1)}$ e $x_{(n)}$
- `diff(range(x))`: amplitude amostral $r_n = x_{(n)} - x_{(1)}$
- `IQR(x)`: amplitude interquartis $q_{3/4} - q_{1/4}$
- `min(x)`: mínimo de x
- `max(x)`: máximo de x
- `sum(x)`: soma das componentes de x
- `prod(x)`: produto das componentes de x
- `cumprod(x)`: produtos acumulados das componentes de x
- `cumsum(x)`: somas acumuladas das componentes de x
- `median(x)`: mediana de x
- `var(x)`: variância de x
- `sd(x)`: desvio padrão de x
- `sd(x)/mean(x)`: coeficiente de dispersão $s/\overline{x}$
- `mean((x-mean(x))^r)`: momento central de ordem $r : m_r$
- `m3/(m2^(3/2))`: coeficiente de assimetria b_1
- `m4/(m2^(4/2))`: coeficiente de assimetria b_2
- `prod(a,b)`: função para multiplicação de ”a” e ”b”

- `sqrt(x)`: função raiz quadrada de x
- `factorial(a)`: fatorial de a
- rep("a",b): repetir a letra "a" b vezes
- x<-c(1,3,5,3): atribui a x a sequência indicada
- `c(1,3,5,3)`: devolve a sequência introduzida
- X<-c("S","N","S"): atribui a X a sequência indicada
- `y<-x`: atribui a y o objeto x
- `z<-c(a1=7, a2=5)`: atribui a z uma sequência com nomes
- c(X, "N"): acrescenta "N" à sequência X
- `c(x,4,3)`: junta novos elementos a x
- `x[c(2,4)]`: seleciona o 2º e 4º elementos de x
- `x[-c(2,4)]`: exclui o 2º e 4º elementos de x
- `x[x>=2]`: seleciona elementos de x que são maiores ou iguais a 2
- `x>=2`: testa se cada elemento de x é maior ou igual a 2
- `x==3|x==1`: testa se cada elemento de x é 3 ou 1
- `X>=`Q: testa se cada elemento de X é maior ou igual a "Q"
- `which(x==max(x))`: posição do maior elemento de x
- `which(x>=3)`: posição dos elementos de x maiores ou iguais a 3
- `length(x)`: comprimento do objeto x
- `1:7 ou  seq(1,7)`: forma sequência de inteiros de 1 a 7
- `2*1:4 ou seq(2,8,2)`: duplica os valores da sequência de 1 a 4
- `seq(3,1,-0.5)`: sequência de 3 a 1 com passo de -0.5
- `seq(from=3,to=1,len=5)`: sequência de 3 a 1 com 5 elementos equidistantes
- `rep(1,7)`: sequência de 7 elementos iguais a 1
- `rep(x,c(1,1,2,4))`: repete elementos de x com determinada frequência

GRÁFICOS

- `par(mfrow=c(a,b))`: apresenta as figuras distribuídas em a linhas e b colunas
- `plot(x,y, opcoes)`: representa graficamente os pontos de coordenadas (x_i, y_i) onde x_i e y_i são as componentes dos vetores x e y
- `curve(f,a,b)`: representa o gráfico de f no intervalo $[a, b]$
- `plot(table(x), opcoes)`: diagrama de linhas para amostra x
- `pie(table(x), opcoes)`: gráfico circular para amostra x
- `stem(x, opcoes)`: diagrama de caule-e-folhas para amostra x
- `hist(x, opcoes)`: histograma para amostra x
- `barplot(x)`: gráfico de barras para amostra x
- `boxplot(table(x))`: diagrama de caixa-com-bigodes para amostra x
- `help(par)`: informação sobre todas as opções disponíveis para os gráficos
- `pch`: determina a símbolo a usar na representação gráfica de pontos; pode tomar valores de 1 a 25
- `col`: código da cor ou nome da cor a usar
- `las`: orienta os "labels"dos eixos
- `xlab=`*nome*/`ylab=`*nome*: coloca o texto nome no eixo dos xx/yy
- `xlim=c(a,b)` / `ylim=c(a,b)`: fixa $[a, b]$ como o intervalo do eixo dos xx/yy
- `main=`*nome*: coloca o título nome no gráfico
- `x<-c()`: variável indepedente
- `y<-c()`: variável depedente
- `plot(y~x,`$\cdots$`)`: diagrama de dispersão
- `lm(y~x,`$\cdots$`)`: regressão linear simples, $y = a + bx$
- `lm(y~x-1,`$\cdots$`)`: regressão (passando pela origem), $y = bx$
- `abline(lm(`$\cdots$`))`: acrescenta reta ajustada no plot dos dados
- `resid()`: resíduos do ajustamento linear e_i

- `cor(x,y,method=c(pearson,spearman,kendall))`: coeficiente de correlação de Pearson, Spearman ou Kendall

- sample `(x, r, replace=FALSE,prob=NULL)`: simula uma amostra de dimensão r

DISTRIBUIÇÕES DE PROBABILIDADE

Distribuição Binomial

- `dbinom(x, n, p)`: $P(X = x)$ f.m.p.
- `pbinom(x,n,p)`: $P(X \leq x)$ f.d.
- `pbinom(x,n,p, lower.tail=F)`: $P(X > x)$
- `qbinom(a,n,p)`: quantil de ordem a
- `rbinom(r,n,p)`: simula uma amostra de dimensão r

Distribuição de Poisson

- `dpois(x,lambda)`: $P(X = x)$ f.m.p.
- `ppois(x,lambda)`: $P(X \leq x)$ f.d.
- `ppois(x,lambda, lower.tail=FALSE)`: $P(X > x)$
- `ppois(p,lambda)`: quantil de ordem p
- `rpois(r,lambda)`: simula uma amostra de dimensão r

Distribuição Normal

- `dnorm(x, m, s)`: função densidade
- `pnorm(x, m, s)`: $P(X \leq x)$ função distribuição
- `pnorm(x,lambda, lower.tail=FALSE)`: $P(X > x)$
- `qnorm(p, m, s)`: quantil de ordem p
- `rnorm(r, m, s)`: simula uma amostra de dimensão r