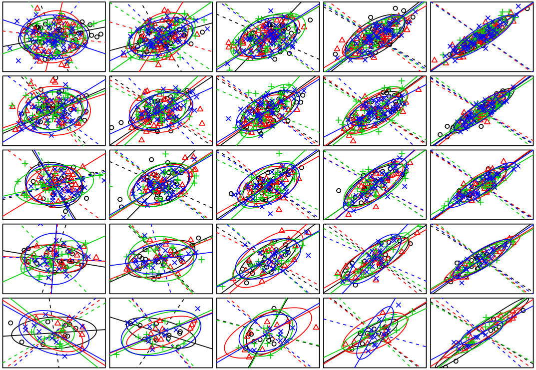
Estatística Multivariada

Slides de apoio às aulas



Faculdade de Ciências e Tecnologia
Universidade Nova de Lisboa
2018/19

Apresentação da Unidade Curricular

Apresentação da Unidade Curricular

- **Objectivo** Esta unidade curricular aborda os principais métodos estatísticos de análise de dados multivariados. Em particular, serão abordadas as técnicas de inferência para valores médios multivariados e matrizes de covariância e modelos lineares em populações Gaussianas. Serão também focados métodos de redução da dimensionalidade, de discriminação e classificação de dados (*clustering*).

- **Programa**:

  1. Breves revisões (álgebra linear, variáveis e vetores aleatórios e inferência estatística)
  2. Introdução à estatística multivariada
  3. Distribuição normal multivariada e distribuição de Whishart
  4. Inferência sobre médias multivariadas
     1. Inferência sobre um vetor de médias
     2. Comparação de dois vetores de médias
     3. Comparação de k ($k \geq 2$) vetores de médias
  5. Inferência sobre matrizes de covariâncias
  6. Análise da estrutura de covariância
     1. Análise em componentes principais
     2. Análise de correlação canónica
  7. Análise classificatória e de *clustering*
     1. Análise discriminante
     2. Análise de *clusters*

Método de ensino e material de apoio

- **Método de ensino:** As aulas de EM assentam numa exposição teórico-prática, com recurso à resolução prática de exercícios (*papel-e-lápis*) e análise de dados em ambiente R. Softwares a instalar:
  - R (download a partir de https://cran.r-project.org/bin/windows/base/)
  - RStudio (download a partir de https://www.rstudio.com/products/rstudio/download/)
  - MikteX (download a partir de https://miktex.org/download)
  - Versão Java 64 bits
- **Material de apoio:**
  - Slides de apoio às aulas
  - Texto de apoio (Professor Filipe Marques)
  - Caderno de exercícios (Professor Filipe Marques)

Bibliografia

- Base:
  - Johnson, R. and Wichern, D. W. (2007), Applied Multivariate Statistical Analysis, 6th Edition, Prentice Hall, New Jersey
- Complementar:
  - Flury, B. (1997), A First Course in Multivariate Statistics, Springer. New York
  - Morrison, D. F. (2004), Multivariate Statistical Methods, 4th Edition, Duxbury Press
  - Rencher, A. C. (1998), Multivariate Statistical Inference and Applications, John Wiley & Sons
  - Rencher, A. C. and Christensen, W. F. (2012). Methods of Multivariate Analysis, Third Edition, John Wiley & Sons
- Avançada:
  - Anderson, T. W. (2003), An Introduction to Multivariate Statistical Analysis, 3rd ed., J. Wiley & Sons, New York
  - Muirhead, R. J. (1982), Aspects of multivariate statistical theory, Wiley, New York
  - Kshirsagar, A. M. (1972). Multivariate Analysis, Dekker, New York
- Aplicações em R:
  - Everitt *et al* (2011). An Introduction to Applied Multivariate Analysis with R. Springer
  - Zelterman, D. (2015). Applied Multivariate Statistics with R. Springer

Avaliação

1. **Frequência:**
   - Em todas as aulas teórico-práticas serão assinaladas as presenças dos alunos. Os alunos que queiram justificar as suas faltas devem entregar o respetivo comprovativo de justificação no prazo de 5 dias úteis, a contar da data em que ocorreram essas mesmas faltas.
   - Só serão admitidos a avaliação na disciplina alunos que tenham um total de presenças superior ou igual a 2/3 das aulas leccionadas durante o semestre
   - Esta regra é válida para todos os alunos, com exceção de alunos com o estatuto de trabalhador/estudante, ou qualquer outro reconhecido pelas regras de avaliação da faculdade.
2. **Avaliação contínua:**
   - A avaliação contínua será feita através de três elementos de avaliação:
     - 1º avaliação - Teste a realizar no período de aulas com uma ponderação de 40%. O teste terá a duração de 2h. O teste é classificado numa escala de 0 a 20 valores.
     - 2º avaliação - Teste a realizar no período de aulas com uma ponderação de 40%. O teste terá a duração de 2h. O teste é classificado numa escala de 0 a 20 valores.
     - 3º avaliação - Trabalho individual a realizar com o apoio do software R (é valorizado o uso de RMarkdown). O trabalho terá uma ponderação de 20% e será entregue na última aula do semestre. O trabalho é classificado numa escala de 0 a 20 valores.
   - O aluno obtém aprovação na disciplina em época normal (avaliação contínua) se as notas dos testes forem superior ou iguais a 7.0 valores e se a média ponderada dos três elementos de avaliação for superior ou igual a 9.5 valores.
   - Caso um aluno não compareça a uma das avaliações, esse elemento de avaliação terá nota 0.0 para a classificação final.

Avaliação

1. **Recurso e melhoria de nota:**
   - A avaliação da época de recurso (recurso ou melhoria) é feita por exame, numa única data dentro da época de recurso prevista no calendário letivo.
   - O exame é classificado numa escala de 0 a 20 valores. O aluno obtém aprovação à cadeira se conseguir nota superior ou igual a 9.5 valores no exame.
   - Os alunos que pretenderem realizar o exame de recurso, com vista à melhoria de nota, devem, antecipadamente, requerer essa melhoria junto dos serviços académicos.

2. **Outras notas importantes:**
   - A inscrição em qualquer prova de avaliação escrita é obrigatória devendo ser feita online através da página no CLIP da disciplina. As inscrições para as diferentes provas são independentes.
   - As inscrições para cada teste ou exame deverão ser feitas dentro do prazo assinalado no CLIP.
   - É obrigatório que os alunos se façam acompanhar do seu bilhete de identidade e de um caderno de exame para a realização dos testes ou exames.
   - As tabelas estatísticas ou outro tipo de material de apoio serão fornecidos pelos professores durante as provas.

Aula 1

Revisões - Algebra linear

Vetores

- Um vetor $\mathbf{x}_{n\times 1}$, ou simplesmente $\mathbf{x}$, de n elementos $x_1, ..., x_n$ (conjunto de números - ou variáveis - organizados em linhas) representa-se por

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

- Geometricamente, um vetor com n elementos identifica um ponto num espaço de dimensão n. Os elementos do vetor são as coordenadas do ponto.
- $\mathbf{x}' = (x_1, ..., x_n)$ representa o vetor *transposto* de $\mathbf{x}$

Exemplo

$$\mathbf{x} = \begin{bmatrix} 4 \\ 5 \\ -3 \end{bmatrix} \qquad \mathbf{x}' = (4, 5, -3)$$

Código R:

```
> x<-matrix(c(4,5,-3),3,1)
> t(x)

     [,1] [,2] [,3]
[1,]    4    5   -3
```

Operações básicas com vetores

- **Soma**: $\mathbf{z} = \mathbf{x} + \mathbf{y} \Leftrightarrow z_i = x_i + y_i \; (i = 1, ..., n)$

Exemplo

$$\begin{bmatrix} 4 \\ 5 \\ -3 \end{bmatrix} + \begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 5 \\ 8 \\ -2 \end{bmatrix}$$

Código R:

```
> matrix(c(4,5,-3),3,1)+matrix(c(1,3,1),3,1)

     [,1]
[1,]    5
[2,]    8
[3,]   -2
```

- **Produto por um escalar**: $c\,\mathbf{x} = c \times x_i = x_i \times c \; (i = 1, ..., n)$

Exemplo

$$2 \begin{bmatrix} 4 \\ 5 \\ -3 \end{bmatrix} = \begin{bmatrix} 8 \\ 10 \\ -6 \end{bmatrix}$$

Operações básicas com vetores

Código R:

```
> x<-matrix(c(4,5,-3),3,1)
> 2*x

     [,1]
[1,]    8
[2,]   10
[3,]   -6
```

- **Produto interno**: $\mathbf{x}'_{1\times n}\mathbf{y}_{n\times 1} = \sum_{i=1}^{n} x_i y_i$

Exemplo

$$\mathbf{x} = \begin{bmatrix} 4 \\ 5 \\ -3 \end{bmatrix} \qquad \mathbf{y} = \begin{bmatrix} 2 \\ 1 \\ -2 \end{bmatrix} \qquad \mathbf{x}'\mathbf{y} = 19$$

Código R:

```
> x<-matrix(c(4,5,-3),3,1)
> y<-matrix(c(2,1,-2),3,1)
> crossprod(x,y)

     [,1]
[1,]   19
```

Comprimento de um vetor

- Os vetores têm propriedades geométricas de comprimento e direção
- O comprimento (ou *norma*) de um vetor é dado por: $\|\mathbf{x}\| = \sqrt{\mathbf{x}'\mathbf{x}} = \sqrt{\sum_{i=1}^{n} x_i^2}$

Exemplo

$$\mathbf{x}' = (4, 5, -3) \qquad \|\mathbf{x}\| = 7.071$$

- Se $\mathbf{x}'\mathbf{x} = 1$ (norma igual a 1) então o vetor diz-se *normalizado*
- Todo o vetor pode ser normalizado dividindo-o pelo seu comprimento

$$\mathbf{e} = \frac{\mathbf{x}}{\sqrt{\mathbf{x}'\mathbf{x}}}$$

Exemplo

$e_1 = \frac{4}{\sqrt{4^2+5^2+(-3)^2}} = 0.566$ $\quad e_2 = \frac{5}{\sqrt{4^2+5^2+(-3)^2}} = 0.707$ $\quad e_3 = \frac{-3}{\sqrt{4^2+5^2+(-3)^2}} = -0.424$

Código R:

```
> x<-c(4,5,-3)
> round(x/sqrt(sum(x^2)),3)

[1]  0.566  0.707 -0.424
```

Direção de um vetor

- O ângulo θ formado por dois vetores $\mathbf{x}$ e $\mathbf{y}$ ambos de dimensão n é definido por

$$\cos(\theta) = \frac{\mathbf{x}'\mathbf{y}}{\sqrt{\mathbf{x}'\mathbf{x}}\sqrt{\mathbf{y}'\mathbf{y}}}$$

Exemplo

$\mathbf{x}' = (-1, 5, 2, -2)$ $\quad$ $\mathbf{y}' = (4, -3, 0, 1)$ $\quad$ $\cos(\theta) = -0.706 \Leftrightarrow \theta = 135^\circ$

Código R:

```
> x1<-c(-1,5,2,-2)
> y1<-c(4,-3,0,1)
> theta<-round(sum(x1*y1)/(sqrt(sum(x1^2))*sqrt(sum(y1^2))),3)
> theta

[1] -0.706

> rad2deg <- function(rad) {(rad * 180) / (pi)}
> rad2deg(acos(theta))

[1] 134.9104
```

Combinação linear de vetores

- $\mathbf{x}$ e $\mathbf{y}$ dizem-se *ortogonais* (geometricamente perpendiculares) se $\mathbf{x}'\mathbf{y} = x_1y_1 + ... + x_ny_n = 0$

Exemplo

$$\mathbf{x}' = (2, 5, -6) \qquad \mathbf{y}' = (1, 2, 2) \qquad \mathbf{x}'\mathbf{y} = 0$$

- O vetor $c_1\mathbf{x}_1 + ... + c_n\mathbf{x}_n$ é dito uma *combinação linear* dos vetores $\mathbf{x}_1, ..., \mathbf{x}_n$
- $\mathbf{x}_1,..., \mathbf{x}_n$, são *linearmente dependentes* se existirem constantes $(c_1, ..., c_n)$ <u>não todas nulas</u>, tal que $c_1\mathbf{x}_1 + ... + c_n\mathbf{x}_n = 0$, ou seja, um vetor pode ser escrito como combinação linear dos restantes (serão linearmente independentes no caso contrário)

Exemplo

$$\mathbf{x}_1' = (1, 2, 1) \qquad \mathbf{x}_2' = (1, 0, -1) \qquad \mathbf{x}_3' = (1, -2, 1)$$

$c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + c_3\mathbf{x}_3 = 0 \Leftrightarrow \mathbf{c}' = (0, 0, 0) \Rightarrow$ vetores linearmente independentes.

<u>Código R</u>:

```
> A<-matrix(c(1,2,1,1,0,-1,1,-2,1),3,3)
> solve(A,c(0,0,0))

[1] 0 0 0
```

Matrizes

- Uma matriz $\mathbf{A}_{n\times p}$, ou simplesmente $\mathbf{A}$ (conjunto de números - ou variáveis - organizados numa tabela em linhas e colunas) representa-se por

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1p} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \dots & a_{np} \end{bmatrix} \equiv [a_{ij}]\,(i = 1, ..., n;\ j = 1, ..., p)$$

- Uma matriz com $n = p$ diz-se *quadrada*; se $n \neq p$ a matriz diz-se *retangular*

- A matriz transposta de $\mathbf{A}_{n\times p}$, representa-se por $\mathbf{A}'_{p\times n}$ e obtem-se trocando as linhas pelas colunas da matriz $\mathbf{A}$

$$\mathbf{A}' = [a_{ji}]\,(j = 1, ..., p;\ i = 1, ..., n)$$

Exemplo

$$\mathbf{A} = \begin{bmatrix} 3 & -2 & 1 \\ -2 & 5 & 1 \\ 4 & 8 & -2 \end{bmatrix} \qquad \mathbf{A}' = \begin{bmatrix} 3 & -2 & 4 \\ -2 & 5 & 8 \\ 1 & 1 & -2 \end{bmatrix}$$

Código R:

```
> A<-matrix(c(3,-2,4,-2,5,8,1,1,-2),3,3)
> A

     [,1] [,2] [,3]
[1,]    3   -2    1
[2,]   -2    5    1
[3,]    4    8   -2

> t(A)

     [,1] [,2] [,3]
[1,]    3   -2    4
[2,]   -2    5    8
[3,]    1    1   -2

> B<-matrix(c(1,4,2,5,2,2,-5,0,1),3,3)
> B

     [,1] [,2] [,3]
[1,]    1    5   -5
[2,]    4    2    0
[3,]    2    2    1

> t(B)

     [,1] [,2] [,3]
[1,]    1    4    2
[2,]    5    2    2
[3,]   -5    0    1
```

- **Soma**:

$$\mathbf{C}_{n\times p} = \mathbf{A}_{n\times p} + \mathbf{B}_{n\times p} \Leftrightarrow c_{ij} = a_{ij} + b_{ij} \; (i = 1, ..., n;\; j = 1, ..., p)$$

Exemplo

$$\begin{bmatrix} 3 & -2 & 1 \\ -2 & 5 & 1 \\ 4 & 8 & -2 \end{bmatrix} + \begin{bmatrix} 1 & 5 & -5 \\ 4 & 2 & 0 \\ 2 & 2 & 1 \end{bmatrix} = \begin{bmatrix} 4 & 3 & -4 \\ 2 & 7 & 1 \\ 6 & 10 & -1 \end{bmatrix}$$

Código R:

```
> A+B

     [,1] [,2] [,3]
[1,]    4    3   -4
[2,]    2    7    1
[3,]    6   10   -1
```

- **Produto por um escalar**:

$$c\,\mathbf{A}_{n\times p} = \mathbf{A}_{n\times p}\,c = \mathbf{B}_{n\times p} = \{b_{ij}\} \Leftrightarrow b_{ij} = c \times a_{ij} = a_{ij} \times c\,(i = 1, \ldots, n;\; j = 1, \ldots, p)$$

Exemplo

$$2\begin{bmatrix} 3 & -2 & 1 \\ -2 & 5 & 1 \\ 4 & 8 & -2 \end{bmatrix} = \begin{bmatrix} 6 & -4 & 2 \\ -4 & 10 & 2 \\ 8 & 16 & -4 \end{bmatrix}$$

Código R:

```
> A

     [,1] [,2] [,3]
[1,]    3   -2    1
[2,]   -2    5    1
[3,]    4    8   -2

> 2*A

     [,1] [,2] [,3]
[1,]    6   -4    2
[2,]   -4   10    2
[3,]    8   16   -4
```

- **Produto matricial**:

$$\mathbf{AB}_{n\times m} = \mathbf{A}_{n\times p}\mathbf{B}_{p\times m} \Leftrightarrow c_{ij} = (a_i^{\text{linha}})' \times (b_j^{\text{coluna}}) = \sum_{k=1}^{p} a_{ik}\, b_{kj}\; (i = 1, ..., n;\; j = 1, ..., m)$$

Exemplo

$$\mathbf{A} = \begin{bmatrix} 2 & 1 & 3 \\ 4 & 6 & 5 \\ 7 & 2 & 3 \\ 1 & 3 & 2 \end{bmatrix} \qquad \mathbf{B} = \begin{bmatrix} 1 & 4 \\ 2 & 6 \\ 3 & 8 \end{bmatrix} \qquad \mathbf{AB} = \begin{bmatrix} 13 & 38 \\ 31 & 92 \\ 20 & 64 \\ 13 & 38 \end{bmatrix}$$

Código R:

```
> A<-matrix(c(2,4,7,1,1,6,2,3,3,5,3,2),4,3)
> B<-matrix(c(1,2,3,4,6,8),3,2)
> A%*%B

     [,1] [,2]
[1,]   13   38
[2,]   31   92
[3,]   20   64
[4,]   13   38
```

Operações com matrizes

- O produto matricial não é comutativo
- Se $\mathbf{x}$ é um vetor $p \times 1$, $\mathbf{Ax}$ é uma combinação linear dos elementos coluna da matriz $\mathbf{A}$
- Se $\mathbf{x}$ é um vetor $n \times 1$, $\mathbf{x}'\mathbf{A}$ é uma combinação linear dos elementos linha da matriz $\mathbf{A}$

Código R:

```
> A<-matrix(c(2,4,7,1,1,6,2,3,3),3,3)
> A

     [,1] [,2] [,3]
[1,]    2    1    2
[2,]    4    1    3
[3,]    7    6    3

> t(x)

     [,1] [,2] [,3]
[1,]    4    5   -3

> A%*%x

     [,1]
[1,]    7
[2,]   12
[3,]   49

> t(x)%*%A

     [,1] [,2] [,3]
[1,]    7   -9   14
```

- **Matriz triangular**:

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ 0 & a_{22} & a_{23} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & a_{nn} \end{bmatrix} \qquad \mathbf{A} = \begin{bmatrix} a_{11} & 0 & 0 & \dots & 0 \\ 0 & a_{22} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \dots & a_{nn} \end{bmatrix}$$

- **Matriz diagonal**:

$$\mathbf{A} = \begin{bmatrix} a_{11} & 0 & 0 & \dots & 0 \\ 0 & a_{22} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & a_{nn} \end{bmatrix}$$

- **Matriz identidade** (caso particular):

$$\mathbf{I} = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}$$

- **Determinante** de uma matriz $\mathbf{A}_{p\times p}$ é o escalar

$$|\mathbf{A}| = a_{11} \text{ se } p = 1$$
$$|\mathbf{A}| = \sum_{j=1}^{p} a_{1j}|\mathbf{A}_{1j}|(-1)^{1+j} \text{ se } p > 1$$

sendo $\mathbf{A}_{1j}$ a matriz $\mathbf{A}_{(p-1)\times(p-1)}$, resultante de eliminar a primeira linha e j-ésima coluna de $\mathbf{A}$

- Matriz 2×2

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22}(-1)^2 + a_{12}a_{21}(-1)^3 = a_{11}a_{22} - a_{12}a_{21}$$

- Matriz 3×3

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix}(-1)^2 + a_{12}\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix}(-1)^3 + a_{13}\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}(-1)^4$$

Exemplo

$$\mathbf{A} = \begin{bmatrix} 3 & 4 \\ 2 & 6 \end{bmatrix} \qquad |\mathbf{A}| = 10$$

$$\mathbf{B} = \begin{bmatrix} 3 & 1 & 6 \\ 7 & 4 & 5 \\ 2 & -7 & 1 \end{bmatrix} \qquad |\mathbf{B}| = -222$$

Código R:

```
> A<-matrix(c(3,2,4,6),2,2)
> det(A)

[1] 10

> B<-matrix(c(3,7,2,1,4,-7,6,5,1),3,3)
> det(B)

[1] -222
```

Matrizes quadradas

- **Matriz inversa**: $\mathbf{A}^{-1}$ é matriz inversa de $\mathbf{A}$ se $\mathbf{A}^{-1}\mathbf{A} = \mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$. Em particular, a inversa de uma matriz 2×2 é dada por

$$\mathbf{A}^{-1} = \frac{1}{|A|} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}$$

Exemplo

$$\mathbf{A} = \begin{bmatrix} 3 & 4 \\ 2 & 6 \end{bmatrix} \qquad \mathbf{A}^{-1} = \begin{bmatrix} 0.6 & -0.4 \\ -0.2 & 0.3 \end{bmatrix}$$

Código R:

```
> solve(A)

     [,1] [,2]
[1,]  0.6 -0.4
[2,] -0.2  0.3
```

- **Matriz ortogonal**: $\mathbf{A}_{p \times p}$ é matriz ortogonal se $\mathbf{A}\mathbf{A}' = \mathbf{A}'\mathbf{A} = \mathbf{I}$, ou seja, sse $\mathbf{A}' = \mathbf{A}^{-1}$

- **Matriz simétrica**: $\mathbf{A}' = \mathbf{A} \Leftrightarrow a_{ij} = a_{ji}\ (i = 1, ..., n;\ j = 1, ..., p)$

- **Traço** de uma matriz (quadrada)

$$tr(\mathbf{A}) = \sum_{i=1}^{n} a_{ii}$$

Exemplo

$$\mathbf{A} = \begin{bmatrix} 3 & -2 \\ 2 & 4 \end{bmatrix} \qquad tr(\mathbf{A}) = 7$$

Código R:

```
> library(psych)
> A<-matrix(c(3,2,-2,4),2,2)
> tr(A)

[1] 7
```

Valores e vetores próprios

- Seja $\mathbf{A}$ uma matriz quadrada $p \times p$ e $\mathbf{I}$ a matriz identidade com a mesma dimensão. Os escalares $\ell_1, ..., \ell_p$ que satisfazem a equação (*característica*)

$$|\mathbf{A} - \ell\,\mathbf{I}| = 0$$

designam-se por **valores próprios** da matriz $\mathbf{A}$ (raízes características).

- O vetor não-nulo $\mathbf{x}_{p\times 1}$ ($\mathbf{x}_{p\times 1} \neq \mathbf{0}_{p\times 1}$) tal que

$$\mathbf{A}\mathbf{x} = \ell\mathbf{x}$$

é chamado o **vetor próprio** da matriz $\mathbf{A}$ associado ao valor próprio ℓ

- Em regra, toma-se o vetor normalizado, $\mathbf{e}_{p\times 1}$, de comprimento unitário, como o vetor próprio correspondente ao valor próprio ℓ, i.e.,

$$e_i = \frac{x_i}{\sqrt{\mathbf{x}'\mathbf{x}}} \; (i = 1, ..., p)$$

Note-se que se $\mathbf{e}$ é vetor próprio então $-\mathbf{e}$ também o é (as soluções definem direções).

Exemplo

Considere-se a matriz

$$\boldsymbol{S} = \begin{bmatrix} 2.459 & 1.019 \\ 1.019 & 2.472 \end{bmatrix}$$

Os valores próprios de $\boldsymbol{S}$ são os valores ℓ tal que $|\boldsymbol{S} - \ell \mathbf{I}| = 0$:

$$|\boldsymbol{S} - \ell \mathbf{I}| = 0 \Leftrightarrow \begin{vmatrix} 2.459 - \ell & 1.019 \\ 1.019 & 2.472 - \ell \end{vmatrix} = 0 \Leftrightarrow$$

$$(2.459 - \ell)(2.472 - \ell) - 1.019^2 = 0 \Rightarrow \ell_1 = 3.485 \wedge \ell_2 = 1.446$$

Os vetores próprios da matriz $\boldsymbol{S}$ são determinados tal que $\boldsymbol{S}\mathbf{x} = \ell \mathbf{x}$, sujeito à restrição $\|\mathbf{x}\| = 1$. Para $\ell_1 = 3.484$, tem-se

$$\begin{bmatrix} 2.459 & 1.019 \\ 1.019 & 2.472 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 3.484 \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

com

$$x_1^2 + x_2^2 = 1$$

Resolvendo em ordem a x_1 e x_2, obtém-se o vetor próprio, associado ao valor próprio ℓ_1, de elementos $x_1 = 0.705$ e $x_2 = 0.709$.
Procedendo da mesma forma para ℓ_2, obtém-se o segundo vetor próprio de elementos, $x_1 = -0.709$ e $x_2 = 0.705$.

Código R:

```
> S<-matrix(c(2.459,1.019,1.019,2.472),2,2)
> S

      [,1]  [,2]
[1,] 2.459 1.019
[2,] 1.019 2.472

> round(eigen(S)$values,3)

[1] 3.485 1.446

> round(eigen(S)$vectors,3)

      [,1]   [,2]
[1,] 0.705 -0.709
[2,] 0.709  0.705
```

- Se $\mathbf{A}_{p\times p}$ for uma matriz *simétrica* então os seus valores/vetores próprios possuem as seguinte propriedades:
  - Os valores/vetores próprios são sempre reais
  - Vectores próprios associados a valores próprios diferentes são sempre *ortogonais*
  - $tr(\mathbf{A}) = \sum_{i=1}^{n} \ell_i$ e $|\mathbf{A}| = \prod_{i=1}^{n} \ell_i$

Decomposição espetral

- Uma matriz $\mathbf{A}_{p\times p}$ simétrica pode ser representada usando os seus valores e vetores próprios $(\ell_i, \mathbf{e}_i)$ através da expressão

$$\mathbf{A} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}' = \sum_{j=1}^{p} \ell_j \mathbf{e}_j \mathbf{e}_j'$$

onde ℓ_j $(j = 1, ..., p)$ são os valores próprios de $\mathbf{A}$, $\boldsymbol{\Lambda} = diag(\ell_1, ..., \ell_p)$ e $\mathbf{e}_j$ $(j = 1, ..., p)$ os vetores próprios normalizados associados aos valores próprios ℓ_j. $\mathbf{P}$ é uma matriz ortogonal cujas colunas são os vetores próprios normalizados.

Exemplo

Considere-se a matriz simétrica

$$\mathbf{A} = \begin{bmatrix} 2.2 & 0.4 \\ 0.4 & 2.8 \end{bmatrix}$$

Os valores próprios de $\mathbf{A}$ são os valores $\ell_1 = 3$ e $\ell_2 = 2$
Os vetores próprios normalizados da matriz $\mathbf{A}$ são $\mathbf{e}_1' = (0.447, 0.894)$ e $\mathbf{e}_2' = (-0.894, 0.447)$, logo

$$\mathbf{A} = \begin{bmatrix} 2.2 & 0.4 \\ 0.4 & 2.8 \end{bmatrix} = 3 \times \begin{bmatrix} 0.447 \\ 0.894 \end{bmatrix} \begin{bmatrix} 0.447 & 0.894 \end{bmatrix} + 2 \times \begin{bmatrix} -0.894 \\ 0.447 \end{bmatrix} \begin{bmatrix} -.8940 & 0.447 \end{bmatrix}$$

Decomposição espetral

Código R:

```
> A<-matrix(c(2.2,0.4,0.4,2.8),2,2)
> A

     [,1] [,2]
[1,]  2.2  0.4
[2,]  0.4  2.8

> round(eigen(A)$values,3)

[1] 3 2

> round(eigen(A)$vectors,3)

      [,1]   [,2]
[1,] 0.447 -0.894
[2,] 0.894  0.447

> round(eigen(A)$values[1]*eigen(A)$vectors[,1]%*%
   t(eigen(A)$vectors[,1])+
   eigen(A)$values[2]*eigen(A)$vectors[,2]%*%
   t(eigen(A)$vectors[,2]),1)

     [,1] [,2]
[1,]  2.2  0.4
[2,]  0.4  2.8
```

Matriz definida positiva

- Sejam $\mathbf{A}_{p\times p}$ uma matriz simétrica e $\mathbf{x}$ de dimensão p. Então $\mathbf{x}'\mathbf{A}\mathbf{x}$ designa-se por *forma quadrática* de $\mathbf{A}$

- A matriz simétrica $\mathbf{A}$ é designada por *definida não negativa* se para todo o vetor $\mathbf{x}\,(\mathbf{x} \neq \mathbf{0})$

$$\mathbf{x}'\mathbf{A}\mathbf{x} \geq 0$$

Uma matriz simétrica é definida não negativa *sse* os **valores próprios são todos não negativos**.

- A matriz simétrica $\mathbf{A}$ é designada por *definida positiva* se para todo o vetor $\mathbf{x}\,(\mathbf{x} \neq \mathbf{0})$

$$\mathbf{x}'\mathbf{A}\mathbf{x} > 0$$

Uma matriz simétrica é definida positiva *sse* os **valores próprios são todos positivos**.

- A decomposição espetral da matriz $\mathbf{A}_{p\times p}$ permite expressar a matriz inversa em termos dos valores e vetores próprios

$$\mathbf{A}^{-1} = \mathbf{P}\boldsymbol{\Lambda}^{-1}\mathbf{P}' = \sum_{j=1}^{p} \frac{1}{\ell_j}\mathbf{e}_j\mathbf{e}_j'$$

onde $\boldsymbol{\Lambda}^{-1} = diag(1/\ell_1, ..., 1/\ell_p)$

Exemplo

$$\mathbf{A}^{-1} = \frac{1}{3} \times \begin{bmatrix} 0.447 \\ 0.894 \end{bmatrix} \begin{bmatrix} 0.447 & 0.894 \end{bmatrix} + \frac{1}{2} \times \begin{bmatrix} -0.894 \\ 0.447 \end{bmatrix} \begin{bmatrix} -.894 & 0.447 \end{bmatrix}$$

Código R:

```
> round((1/eigen(A)$values[1])*eigen(A)$vectors[,1]%*%
   t(eigen(A)$vectors[,1])+
   (1/eigen(A)$values[2])*eigen(A)$vectors[,2]%*%
   t(eigen(A)$vectors[,2]),3)

       [,1]   [,2]
[1,]  0.467 -0.067
[2,] -0.067  0.367
```

Matriz raiz quadrada (por decomposição espetral)

- Permite também definir a *matriz raiz-quadrada*

$$\mathbf{A}^{1/2} = \mathbf{P}\mathbf{\Lambda}^{1/2}\mathbf{P}' = \sum_{j=1}^{p} \sqrt{\ell_j}\mathbf{e}_j\mathbf{e}_j'$$

onde $\mathbf{\Lambda}^{1/2} = diag(\sqrt{\ell_1}, ..., \sqrt{\ell_p})$

Exemplo

$$\mathbf{A}^{1/2} = \sqrt{3} \times \begin{bmatrix} 0.447 \\ 0.894 \end{bmatrix} \begin{bmatrix} 0.447 & 0.894 \end{bmatrix} + \sqrt{2} \times \begin{bmatrix} -0.894 \\ 0.447 \end{bmatrix} \begin{bmatrix} -.8940 & 0.447 \end{bmatrix}$$

Código R:

```
> round(sqrt(eigen(A)$values[1])*eigen(A)$vectors[,1]%*%
   t(eigen(A)$vectors[,1])+
   sqrt(eigen(A)$values[2])*eigen(A)$vectors[,2]%*%
   t(eigen(A)$vectors[,2]),3)

      [,1]  [,2]
[1,] 1.478 0.127
[2,] 0.127 1.668
```

Revisões - Variáveis e vetores aleatórios

Notação

1. Constantes/valores fixos: Letras minúsculas a itálico

   Exemplos: a, c, ℓ, x, y, z

2. Variáveis: Letras minúsculas

   Exemplos: x, y, z

3. Parâmetros (população): Letras gregas minúsculas

   Exemplos: λ, μ, ρ, σ

4. Vetores (de números, variáveis ou parâmetros): Letras minúsculas a negrito

   Exemplos:

   $\boldsymbol{a}$ ($\boldsymbol{a}'=(a_1,...,a_n)$), $\boldsymbol{c}$ ($\boldsymbol{c}'=(c_1,...,c_n)$), $\boldsymbol{\ell}$ ($\boldsymbol{\ell}' = (\ell_1, ..., \ell_n)$)
   $\boldsymbol{x}$ ($\boldsymbol{x}'=(x_1,...,x_n)$), $\boldsymbol{y}$ ($\boldsymbol{y}'=(y_1,...,y_n)$), $\boldsymbol{z}$ ($\boldsymbol{z}'=(z_1,...,z_n)$)
   $\mathbf{x}$ ($\mathbf{x}' = (\mathrm{x}_1, ..., \mathrm{x}_n)$), $\mathbf{y}$ ($\mathbf{y}' = (\mathrm{y}_1, ..., \mathrm{y}_n)$), $\mathbf{z}$ ($\mathbf{z}' = (\mathrm{z}_1, ..., \mathrm{z}_n)$)
   $\boldsymbol{\lambda}$ ($\boldsymbol{\lambda}' = (\lambda_1, ..., \lambda_n)$), $\boldsymbol{\mu}$ ($\boldsymbol{\mu}' = (\mu_1, ..., \mu_n)$)
   Caso particular: $\mathbf{0}$ ($\mathbf{0}' = (0, ..., 0)$)

Variáveis aleatórias

- Em muitas situações, a análise dos resultados de uma dada experiência aleatória passa pela descrição numérica dos resultados dessa experiência aleatória.
- Porem, o espaço de resultados Ω não é um conjunto numérico $\Rightarrow$ Descrição dos resultados através de valores assumidos por uma **variável** no decurso de uma experiência aleatória.
- Em resumo, pretende-se passar de Ω para $\mathbb{R}$ (ou mais geralmente para $\mathbb{R}^n$)
- Designa-se por **váriável aleatória** a função x de Ω para $\mathbb{R}$

$$x : \Omega \rightarrow \mathbb{R}$$

- A lei de probabilidades ou distribuição de uma v.a. x pode ser descrita através de diversas funções: *função de distribuição*, *função massa de probabilidade* e *função densidade de probabilidade*:
  - *Função de distribuição*: $F(x) = P(x \leq x)$ $(x \in \mathbb{R};\ 0 \leq F(x) \leq 1)$ (sendo $F(x)$ diferenciável)
  - No caso de v.a. discretas, *função massa de probabilidade*: $f(x) = P(x = x)$ verificando as condições $f(x) \geq 0, \forall x$ e $\sum_{x \in \mathbb{K}} f(x) = 1$.
  - No caso de v.a. contínuas, *função densidade de probabilidade*: $f(x) = \frac{d}{dx}F(x) \Leftrightarrow F(x) = \int_{-\infty}^{x} f(t)dt$, verificando as condições $f(x) \geq 0, \forall x$ e $\int_{-\infty}^{+\infty} f(x)dx = 1$.

Exemplo

Seja $p\,(0 < p < 1)$ a probabilidade de sucesso de uma certa experiência aleatória e x a v.a. que representa o *número de provas necessárias até à ocorrência do primeiro sucesso*.
A variável em causa é discreta sendo a sua distribuição de probabilidade (fmp) definida por

$$f(x) = \left\{ \begin{array}{cccc} 1 & 2 & 3 & \ldots \\ p & (1-p)p & (1-p)^2 p & \ldots \end{array} \right.$$

$$f(x) = \left\{ \begin{array}{c} x \\ p_x = p(1-p)^x \end{array} \right. \quad x = 0, 1, 2, \ldots$$

É fácil verificar que

- $p(1-p)^x \geq 0, \forall x \in \mathbb{N}_0$
- $\sum_{i=1}^{+\infty} p(1-p)^i = \frac{p}{1-(1-p)} = 1$

A função de distribuição é dada por

$$F(x) = P(x \leq x) = \sum_{i=1}^{x} p(1-p)^i$$

Exemplo

Suponhamos que sobre um dado segmento de recta (a, b) se escolhe um ponto ao acaso, i.e., tal que a probabilidade de escolha seja independente da posição. A densidade de probabilidade deve ser então considerada constante, i.e.,

$$f(x) = \begin{cases} c & a < x < b \\ 0 & x \leq a \text{ ou } x \geq b \end{cases}$$

Para que $f(x)$ seja fdp deve ser não negativa e verificar

$$\int_a^b f(x)dx = 1 \Rightarrow c = \frac{1}{b-a}$$

A função de distribuição é dada por

$$F(x) = \begin{cases} a & x < a \\ \frac{x-a}{b-a} & a \leq x < b \\ 1 & x \geq b \end{cases}$$

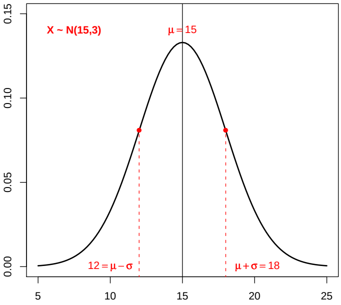
Valor médio e variância

- Dada uma v.a. x chama-se *valor médio*, valor esperado ou média e representa-se por $E(x)$, μ_x ou μ a quantidade definida por,
  - no caso discreto,
    $$\mu = E(x) = \sum_{i=1}^{n} x_i p_i$$
    sendo x uma v.a. com distribuição (x_i, p_i) $(i = 1, \ldots, n)$
  - no caso contínuo,
    $$\mu = E(x) = \int_{-\infty}^{+\infty} x f(x) dx$$
    sendo x uma v.a. com densidade $f(x)$
- Dada uma v.a. x chama-se *variância* e representa-se por $Var(x)$, σ_x^2 ou σ^2 a quantidade definida por
  $$\sigma^2 = Var(x) = E[(x-\mu)^2] = E(x^2) - \mu^2$$

- Uma variável aleatória contínua, x diz-se ter distribuição gaussiana ou normal com valor médio μ e desvio-padrão σ, i.e., $x \frown N(\mu, \sigma)$, se a sua *fdp* corresponde a

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2},\ -\infty < x < +\infty,\ -\infty < \mu < +\infty,\ 0 < \sigma < +\infty$$

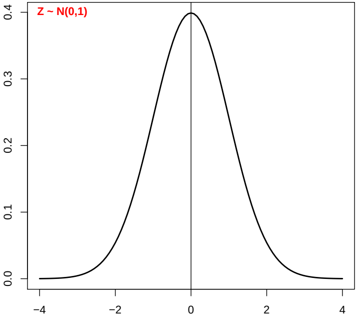
- Função densidade de probabilidade:



- Curva unimodal, sendo a moda $x = \mu$ (maximizante da fdp)
- Simétrica em relação ao eixo $x = \mu$
- Pontos de inflexão em $x = \mu \pm \sigma$
- Eixo das abcissas como assimptota

- Uma v.a. z diz-se ser uma gaussiana padrão, normal padrão ou normal reduzida, isto é, uma gaussiana de parâmetros $\mu = 0$ (centrada em zero) e $\sigma = 1$ (com escala unitária), $z \frown N(0,1)$, se a sua fdp for

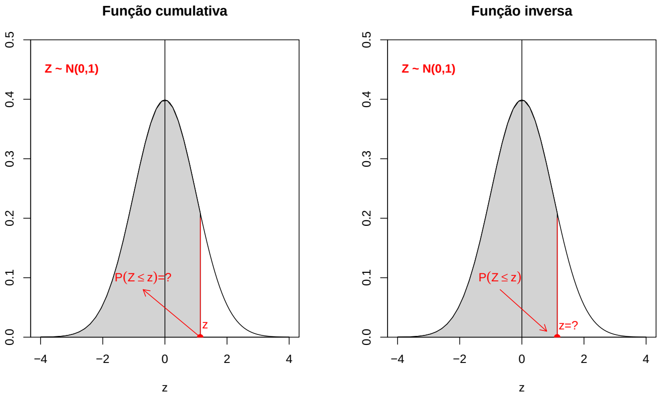
$$z \frown N(0,1) \Rightarrow f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}, \; -\infty < z < +\infty$$



- Padronização da gaussiana: $x \frown N(\mu, \sigma) \Rightarrow z = \frac{x-\mu}{\sigma} \frown N(0,1)$

Tabela da distribuição da normal padrão

- Tabela da função de distribuição da variável normal padrão:
  - Função de distribuição (Função cumulativa)$\longrightarrow$ $\Phi(z) = P(z \leq z)$
  - Inversa da função de distribuição (Função inversa)$\longrightarrow$ $\Phi^{-1}(p) = z_p$ $(0 < p < 1)$ (quantis da guassiana). Genericamente z_p representa o quantil de probabilidade p (ou percentil $p \times 100$) da distribuição normal padrão, tal que $P(z < z_p) = p$



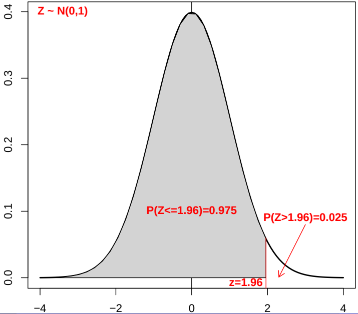
- Note-se que porque a curva é simétrica em relação a $z = 0$, $\Phi(-z) = 1 - \Phi(z)$

Exemplo

$\Phi(1.96) = P(z \leq 1.96) = 0.975$
$z_{0.975} = \Phi^{-1}(0.975) = 1.96$, denota o quantil de probabilidade 0.975 da normal padrão

Código R:

```
> round(pnorm(1.96),3)
[1] 0.975
> round(qnorm(0.975),3)
[1] 1.96
```



- **Alguns teoremas relevantes**:

  1. Seja $x \frown N(\mu, \sigma)$ e $y = a \pm bx$, então $y \sim N(a \pm b\mu, |b|\sigma)$
  2. Sejam $x_1 \frown N(\mu_1, \sigma_1)$ e $x_2 \frown N(\mu_2, \sigma_2)$ independentes, então
  $$(x_1 \pm x_2) \frown N(\mu_1 \pm \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2})$$
  3. Generalizando $x_i \frown N(\mu_i, \sigma_i)\,(i = 1...n)$ independentes, então
  $$\sum_i X_i \frown N\Big(\sum_i \mu_i, \sqrt{\sum_i \sigma_i^2}\Big)$$
  4. Sejam $x_i \frown N(\mu, \sigma)$ n v.a, iid, então $S_n = \sum_i x_i \frown N(n\mu, \sigma\sqrt{n})$

- **Teorema Limite Central**: Sejam $\{x_1, x_2, ..., x_n\}$ variáveis aleatórias independentes e identicamente distribuídas com valor médio μ e variância finita σ^2 e $S_n = \sum_{i=1}^n$ então para $n \to \infty$ tem-se
$$S_n \overset{a}{\sim} N\left(n\mu, \sigma\sqrt{n}\right)$$

- Para n inteiro positivo a v.a. x $(0 \leq x < +\infty)$ com *fdp*

$$f(x) = \begin{cases} \frac{1}{2^{n/2}\Gamma(n/2)}\, x^{n/2-1}\, e^{-y/2} & x > 0 \\ 0 & \text{c.c.} \end{cases}$$

com $\Gamma(t) = \int_0^{\infty} x^{t-1} e^{-x} dx$, $t > 0$, diz-se ter distribuição qui-quadrado com n graus de liberdade, i.e., $x \frown \chi^2_n$

- Função densidade de probabilidade:

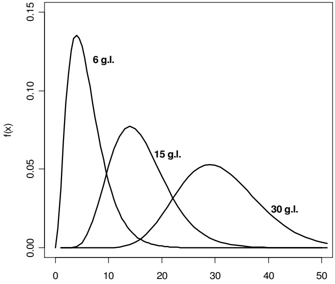


Tabela da distribuição Qui-quadrado

- Tabela da função de distribuição da variável $x \frown \chi^2_n$
  - Função de distribuição (Função cumulativa)$\longrightarrow P(x \leq \chi^2_n)$
  - Inversa da função de distribuição (Função inversa) $\longrightarrow \chi^2_{n,p}$ (quantil de probabilidade p da distribuição χ^2_n)

Exemplo

Considere-se a v.a. $x \frown \chi^2_{10}$
$P(x \leq 5) = 0.109$
$\chi^2_{10,0.90} = 15.987$, denota o quantil de probabilidade 0.90 da distribuição χ^2_{10}

Código R:

```
> round(pchisq(5,10),3)
[1] 0.109
> round(qchisq(0.90,10),3)
[1] 15.987
```

- **Alguns teoremas relevantes**:

1. Sejam $x_i \frown \chi_{(n_i)}$ $(i = 1 \ldots p)$ independentes, então
$$\sum_{i=1}^{p} x_i \frown \chi^2_{(\sum n_i)}$$

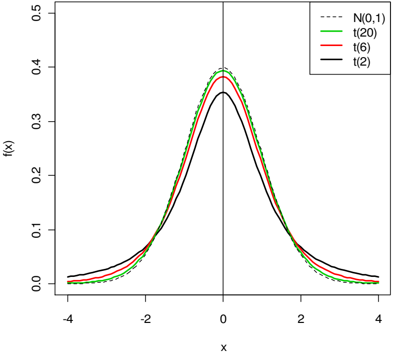
2. Seja $z \frown N(0,1)$ então $z^2 \frown \chi^2_1$

3. Generalizando, $z_i \frown N(0,1)$ $(i = 1, \ldots, n)$ então
$$\sum_{i=1}^{n} z_i^2 \frown \chi^2_n$$

- Sejam $z \frown N(0,1)$ e $y \frown \chi^2_n$, então

$$x = \frac{z}{\sqrt{y/n}} \frown t_n$$

- Uma variável aleatória x $(-\infty < x < +\infty)$ cuja distribuição é bem modelada por uma distribuição t-Student, representa-se por $x \frown t_n$ onde n é o número de graus de liberdade (g.l.).
- Função densidade de probabilidade:



- Tabela da função de distribuição da variável $X \frown t_n$
  - Função de distribuição (Função cumulativa) $\longrightarrow P(x \leq t_n)$
  - Inversa da função de distribuição (Função inversa) $\longrightarrow t_{n,p}$ (quantil de probabilidade p da distribuição t_n)

Exemplo

Considere-se a v.a. $x \frown t_5$
$P(x \leq 1) = 0.818$
$t_{5,0.90} = 1.476$, denota o quantil de probabilidade 0.90 da distribuição t_5

Código R:

```
> round(pt(1,5),3)
[1] 0.818
> round(qt(0.90,5),3)
[1] 1.476
```

- Uma variável aleatória x $(0 \leq x < +\infty)$ tem distribuição F-Snedecor com n (numerador) e d (denominador) graus de liberdade, e representa-se por $x \frown F(n, d)$, se a variável aleatória for tal que

$$x = \frac{y/n}{w/d}$$

onde $y \frown \chi^2_n$ e $w \frown \chi^2_d$ são v.a. independentes.
- Função densidade de probabilidade:

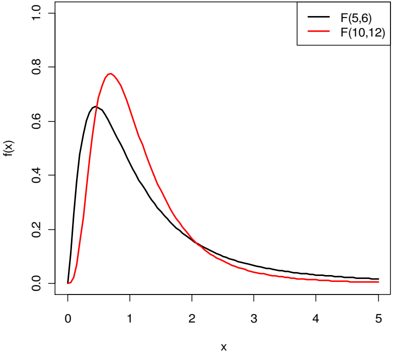


Tabela da distribuição F-Snedecor

- Tabela da função de distribuição da variável $x \frown F_{n,d}$
  - Função de distribuição (Função cumulativa)$\longrightarrow P(x \leq F_{(n,d)})$
  - Inversa da função de distribuição (Função inversa) $\longrightarrow F_{(n,d),p}$ (quantil de probabilidade p da distribuição $F_{(n,d)}$)

Exemplo

Considere-se a v.a. $x \frown F_{(5,8)}$
$P(x \leq 1) = 0.525$
$f_{(5,8),0.90} = 2.726$, denota o quantil de probabilidade 0.90 da distribuição $F_{(5,8)}$

Código R:

```
> round(pf(1,5,8),3)
[1] 0.525
> round(qf(0.90,5,8),3)
[1] 2.726
```

- $(x_1, x_2, ..., x_n)$ diz-se uma *amostra aleatória* de dimensão n se as variáveis são mutuamente independentes e têm a mesma função de distribuição $F(x)$ (cada x_i $(i = 1, ..., n)$ tem a mesma fdp/fmp $f(x)$)

- $x_1, x_2, ..., x_n$ dizem-se variáveis independentes e identicamente distribuídas, abreviadamente, variáveis *iid*

- Note-se que uma a.a., tal como definida, não é mais que o *vetor aleatório* $\mathbf{x} = (x_1, x_2, ..., x_n)$

- Atendendo à independência entre as variáveis numa amostra aleatória, a *distribuição de probabilidade conjunta* do vetor $(x_1, ..., x_n)$ pode ser definida como

$$f_{x_1,...,x_n}(x_1, ..., x_n) = \prod_{i=1}^{n} f(x_i)$$

onde $f(x_i)$ $(i = 1, ..., n)$ representa a função densidade de probabilidade (fdp) da população quando esta é contínua e a função massa de probabilidade (fmp) quando é discreta.

Função verosimilhança de uma amostra

- Na maioria das situações a fdp/fmp da população é membro de uma *família paramérica*, isto é, a forma funcional da função depende de um parâmetro ou de um vetor de parâmetros $\theta = (\theta_1, ..., \theta_q)$, tal que

$$f_{x_1,\ldots,x_n}(x_1, ..., x_n|\theta) = \prod_{i=1}^{n} f(x_i|\theta)$$

- A Estatística analisa os resultados de uma amostra aleatória procurando explicar os valores observados em função dos parâmetros que definem a população

- Assim, para um dado ponto amostral fixo, observado, define-se a *função verosimilhança de uma amostra*, $L(\theta)$, como

$$L(\theta|x_1, ..., x_n) = \prod_{i=1}^{n} f(x_i|\theta)$$

Exemplo

Seja $(x_1, x_2, ..., x_n)$ uma a.a. da população *Exponencial*(β) $(\beta > 0)$ com fdp

$$f(x) = \frac{1}{\beta} \exp\{-x/\beta\} \, (x \geq 0)$$

A fdp conjunta é dada por

$$f(\mathbf{x}) = f(x_1, ..., x_n) = \prod_{i=1}^{n} \frac{1}{\beta} \exp(-x/\beta) = \beta^{-n} \exp\left\{ -\sum_{i=1}^{n} x_i/\beta \right\}$$

que fixando β, depende apenas de $(x_1, x_2, ..., x_n)$.

A função verosimilhança é portanto dada por

$$L(\beta) = \beta^{-n} \exp\left\{ -\sum_{i=1}^{n} x_i/\beta \right\}$$

que fixando $(x_1, x_2, ..., x_n)$ depende apenas de β.

Revisões - Inferência estatística

- Em regra o objetivo último da análise estatística é inferir sobre a população de estudo com base numa amostra aleatória

- O desconhecimento sobre a população pode ser total, i.e., nada se sabe sobre a sua distribuição de probabilidade

- O desconhecimento sobre a população pode ser apenas parcial, i.e., é conhecida a forma funcional da função de distribuição da população mas não se conhecem os valores do(s) parâmetro(s) que caracterizam a distribuição

- Supondo conhecida a forma geral da distribuição de probabilidade da população, a sua completa especificação implica a *estimação* do(s) parâmetro(s), $\boldsymbol{\theta} \in \Theta$ (sendo Θ o espaço dos parâmetros), que lhe estão associados

- A estimação de parâmetros diz-se *pontual*, quando se obtém um só valor para representar $\boldsymbol{\theta}$ e *intervalar* quando se pretende encontrar um intervalo de valores para representar $\boldsymbol{\theta}$
- A *estimação pontual* é um procedimento estatístico que utiliza a informação contida numa amostra para obter um número, ou ponto, que representa o valor do(s) parâmetro(s) que se pretende estimar
- O objetivo principal é portanto o de encontrar uma *estimativa* (função de uma amostra observada) para $\boldsymbol{\theta}$
- A estimativa de $\boldsymbol{\theta}$ não é mais do que a realização de uma *estatística* (função de uma amostra aleatória) a que chamamos *estimador* (contradomínio definido pelo espaço de parâmetros)

Exemplo

$\boldsymbol{x} = (x_1, ..., x_n)$ é uma realização da a.a. $\mathbf{x} = (x_1, ..., x_n)$ onde x_i $(i = 1, .., n)$ são v.a. iid.

$\bar{x}$ (função da amostra observada) é uma estimativa para μ, i.e., é uma realização da estatística $\bar{x} = \sum_i x_i / n$ (função de uma amostra aleatória)

s^2 (função da amostra observada) é uma realização da estatística $s^2 = \sum_i (x_i - \bar{x})^2 / (n-1)$ (função de uma amostra aleatória)

$\bar{x}$ e s^2 são estimadores de μ e σ^2 respetivamente.

- Variáveis aleatórias definidas em função de um estimador e do parâmetro, cuja distribuição não depende de nenhum parâmetro desconhecido designam-se por variáveis *fulcrais* (ou *pivot*)

Exemplo

Seja $(x_1, ..., x_n)$ v.a. iid da população $N(\mu, \sigma)$. Então

$$\frac{\bar{x} - \mu}{s/\sqrt{n}} \frown t_{(n-1)}$$

$$\frac{(n-1)s^2}{\sigma^2} \frown \chi^2_{n-1}$$

- As variáveis fulcrais são usadas para construir *estimadores intervalares*
- Um estimador intervalar de um parâmetro θ desconhecido é im intervalo aleatório que, com uma certa probabilidade de cobertura $1 - \alpha$, contem/captura o parâmetro θ

Exemplo

$$\left(\bar{x} - t_{(n-1);1-\alpha/2}\frac{s}{\sqrt{n}}; \bar{x} + t_{(n-1);1-\alpha/2}\frac{s}{\sqrt{n}}\right)$$

- Genericamente, um **testes de hipóteses** é um conjunto de procedimentos estatísticos que visam determinar se certas afirmações (hipóteses), feitas sobre uma população (ou mais do que uma) são ou não suportadas pelos dados de uma amostra concreta.

- Elementos de um teste de hipóteses:

1. Hipóteses: Hipótese nula (H_0) e Hipótese alternativa (H_1)

2. Estatística do teste

3. Definição da região de rejeição (ou região crítica) e decisão sobre H_0

Hipóteses estatísticas

- Hipótese estatística é toda a proposição feita acerca da(s) população(ções) em estudo.
- Num teste de hipóteses têm-se 2 hipóteses: uma que acreditamos ser verdadeira — **Hipótese Nula** — e a sua contrária — **Hipótese Alternativa**.
- O objectivo final dum teste estatístico é decidir acerca da plausibilidade das hipóteses formuladas.
- Na formulação das hipóteses, deve ter-se em conta que:

    1. O teste é constituído por duas hipóteses complementares, a hipótese nula, H_0 e a hipótese alternativa, H_1;
    2. H_0 é admitida como verdadeira ao longo do procedimento (sob H_0);
    3. Existindo evidências contra H_0 diz-se que se rejeita a favor de H_1;
    4. As hipóteses nula e alternativa são sempre complementares e no seu conjunto estabelecem o universo de possibilidades.

Estatística do teste

- Depois de formuladas as hipóteses, coloca-se a questão "*Como decidir acerca da plausibilidade da hipótese colocada?*"
- O processo de decisão vai basear-se numa estatística — *estatística do teste* (genericamente $T \equiv T(\mathbf{x})$)— para a qual se conhece a respectiva distribuição amostral, função dos dados e que sob H_0 não depende de parâmetros desconhecidos.
- Esta estatística, que é calculada admitindo que H_0 é verdadeira e tendo por os valores observados, vai permitir testar a plausibilidade das hipótese.

Exemplo

Suponha-se um teste ao valor médio — Hipóteses: $H_0 : \mu = \mu_0$ vs. $H_1 : \mu \neq \mu_0$
Neste caso, o parâmetro em análise é o valor médio. Recorde que, numa população onde $x \frown N(\mu, \sigma)$ (σ desconhecido), a estatística $\bar{x}$ (estimador de μ) tem distribuição t_{n-1}:

$$\frac{\bar{x} - \mu}{S/\sqrt{n}} \sim t_{n-1}$$

Logo, atendendo ao valor médio fixado em H_0, i.e. para $\mu = \mu_0$, e para uma amostra concreta é possível calcular a estatística $\frac{\bar{x}-\mu}{S/\sqrt{n}}$.

Região crítica (ou região de rejeição)

- A decisão sobre a plausibilidade das hipóteses formuladas vai basear-se no conhecimento da distribuição de probabilidade da estatística de teste, sob H_0.

- Os valores mais extremos da estatística são aqueles cuja probabilidade de ocorrência é menor e, portanto, correspondem aos valores para os quais se considera improvável a validade da hipótese nula. Constituem por isso a **região crítica** ou região de rejeição

- Considere-se $H_0 : \theta \leq \theta_0$ vs. $H_1 :: \theta > \theta_0$. A hipótese H_0 é rejeitada sse $T > c_\alpha$, sendo este um teste de nível α, com $P(T > c_\alpha) = \alpha$ sob H_0

- Considere-se $H_0 : \theta \geq \theta_0$ vs. $H_1 :: \theta < \theta_0$. A hipótese H_0 é rejeitada sse $T < c_\alpha$, sendo este um teste de nível α, com $P(T < c_\alpha) = \alpha$ sob H_0

- Considere-se $H_0 : \theta = \theta_0$ vs. $H_1 :: \theta \neq \theta_0$. A hipótese H_0 é rejeitada sse $|T| > c_{\alpha/2}$, sendo este um teste de nível α, com $P(|T| > c_{\alpha/2}) = \alpha/2$ sob H_0

Exemplo

Nma amostra de 16 indivíduos com glaucoma de ângulo aberto registaram-se as idades dos pacientes:

62 62 68 48 51 60 51 57 57 41 62 50 53 34 62 61

a) Podemos concluir que a idade média da população a partir da qual a amostra é 60 anos? ($\alpha = 0.01$).
b) Calcule o IC da idade média desta população a 95%.

Código R:

```
> idade<-c(62, 62, 68, 48, 51, 60, 51, 57, 57, 41, 62, 50, 53, 34, 62, 61)
> t.test(x = idade,mu = 60)

        One Sample t-test

data:  idade
t = -2.2822, df = 15, p-value = 0.03749
alternative hypothesis: true mean is not equal to 60
95 percent confidence interval:
 50.20944 59.66556
sample estimates:
mean of x
  54.9375
```