

Análise exploratória de dados com o software R

Simulação de números aleatórios de uma distribuição específica e cálculo de probabilidades com o software R (www.r-project.org)

1. Em R, obtenha amostras aleatórias de dimensão 100 das populações seguintes e para as distribuições contínuas, represente no mesmo gráfico as funções densidade de probabilidade e o histograma das amostras obtidas:
 - a) Binomial com $n=20$ e $p=0.6$
 - b) Poisson de parâmetro $\lambda=4$
 - c) Normal com parâmetros $\mu=0$ e $\sigma=1$
 - d) Normal com parâmetros $\mu=7$ e $\sigma=0.5$
 - e) Normal com parâmetros $\mu=3$ e $\sigma=6$
 - f) Exponencial de parâmetro 5
2. Calcule a probabilidade para cada um dos seguintes eventos:
 - a) Uma variável normal padronizada é maior do que 3
 - b) Uma variável normal com média 35 e desvio padrão 6 é maior do que 42
 - c) $X \leq 0.9$ quando X segue uma distribuição normal padronizada

R-project homepage: <http://www.r-project.org>

Análise exploratória de dados com o software R

Análise inicial dos dados: Estatística Descritiva e Análise Exploratória dos Dados

3. Em *Applied Life Data Analysis* (Wiley, 1982), Wayne Nelson apresenta o tempo em minutos de quebra de um fluído isolante entre eléctrodos a 34kV.

0.19	2.78	4.85	8.27	33.91
0.78	3.16	6.50	12.06	36.71
0.96	4.15	7.35	31.75	72.89
1.31	4.67	8.01	32.52	

- a) Construir a tabela de frequências e o histograma.
- b) Determinar a média, variância, desvio-padrão e mediana da amostra.
- c) Desenhe à mão o diagrama caixa-dos-bigodes para esta amostra.

4. Considere a seguinte tabela de frequências

x_i	115	116	117	118	119	120	121	122	123	124
f_i	4	6	9	13	15	19	20	18	15	10

- a) Determinar a média, variância e desvio-padrão da amostra.
- b) Determinar a mediana e a moda.
- c) Calcule o coeficiente de variação.

5. Um artigo em *Technometrics* (Vol. 19 1977, p. 425) apresenta os seguintes dados para a taxa de octanas presente em vários tipos de gasolina (fazer download dos dados da página web da disciplina).

88.5	94.7	88.2	88.5	93.3	87.4	91.1	90.5
87.7	91.1	90.8	90.1	91.8	88.4	92.6	93.7
83.4	91.0	88.3	89.2	92.3	88.9	89.8	92.7
86.7	94.2	98.8	88.3	90.4	91.2	90.6	92.2

Análise exploratória de dados com o software R

87.5	87.8	94.2	85.3	90.1	89.3	91.1	92.2
91.5	89.9	92.7	87.9	93.0	94.4	90.4	91.2
88.6	88.3	93.2	88.6	88.7	92.7	89.3	91.0
100.3	87.6	91.0	90.9	89.9	91.8	89.7	92.2
95.6	84.3	90.3	89.0	89.8	91.6	90.3	90.0
93.3	86.7	93.4	96.1	89.6	90.4	91.6	90.7

- a) Construir um gráfico de caule-e-folhas.
 - b) Construir a tabela de frequências e o histograma.
 - c) Calcular a média, variância e desvio-padrão da amostra.
6. O aumento de peso, em gramas, em dois conjuntos de animais submetidos a diferentes dietas foram:

Grupo A	56	67	42	48	55	61	52	39	47	58	50	40	59	62	44	57
Grupo B	78	34	37	72	58	68	27	55	65	40	75	33	66			

Utilize os diagramas de caixa (à mão e no R) no estudo da diferença entre os dois tipos de dieta.

Análise exploratória de dados com o software R

R – Exemplos simples de utilização (versão 2.2.0) (R disponível em www.r-project.org)

Operações sobre matrizes e vectores

x=c(1,2,3,4,5)	(definição de um vector linha)
x%*%x	(produto escalar de x)
x*x	(produto elemento a elemento)

A=matrix(c(1,3,2,4),2,2)	
A	
B= matrix(c(5,7,6,8),2,2)	
B	
A+B	
A%*%B	(produto matricial)
A*B	(produto elemento a elemento)
t(A)	(matriz transposta de A)

c=A[,1]	
c	
d=matrix(A[,1],2,1)	
d	
dim(A)	
length(c)	(a instrução “dim(c)” dá erro)
dim(d)	

Definição de vectores e representação gráfica de funções

x=seq(1,5,0.2)	
x	
y=seq(0,2*pi,pi/8)	
y	
z=seq(10,0,-2)	
z	
plot(y,sin(y),”n”)	(com o argumento “n” o gráfico surge vazio, mas esta instrução
	é
lines(y,sin(y))	necessária para de seguida indicar o comando “lines”)

Operações aritméticas simples

3+7
x=4*5
x
sqrt(x)
x=c(-1/3,sqrt(3),(1+2+3)*4/5,(-1)^3)
x
x[4]=log10(1005)
x
x[5]=abs(x[1])
x

Utilização de “strings” de texto

Análise exploratória de dados com o software R

```
s="Ola"  
s  
s=paste(s," , boa tarde",sep="")
```

Exemplo com função polinomial

```
polyroot(c(-1,0,1))
```

 (raízes do polinómio $-1+0*x+1*x^2$)

Definição de um procedimento

Na janela principal do R ("R Console"), seleccionar "File-New script".
Carregar no botão para maximizar a janela para que apareçam as barras de ferramentas.
Escrever as seguintes linhas de comando:

```
a=2  
b=3  
y=a+b  
print(paste(a,"+",b,"=",y))
```

Guardar o ficheiro ("File-Save as" ou clicar no botão respectivo) e dar um nome ao ficheiro seguido da extensão ".R".

Na janela principal do R, seleccionar "File-Change dir" e seleccionar o caminho onde foi guardado o ficheiro, e de seguida "File-Source R code" e seleccionar o ficheiro guardado. O procedimento é executado.

Definição de uma função

Na janela principal do R ("R Console"), seleccionar "File-New script".
Carregar no botão para maximizar a janela para que apareçam as barras de ferramentas.
Escrever as seguintes linhas de comando:

```
soma=function(a,b){  
  resultado=a+b  
  print(paste(a,"+",b,"=",resultado))  
}
```

Guardar o ficheiro ("File-Save as" ou clicar no botão respectivo) e dar um nome ao ficheiro seguido da extensão ".R", neste caso o mais indicado seria "soma.R".

Na janela principal do R, seleccionar "File-Change dir" e seleccionar o caminho onde foi guardado o ficheiro, e de seguida "File-Source R code" e seleccionar o ficheiro guardado. A função passa a ser reconhecida pelo R.

Na janela principal do R, executar "soma(2,3)". Surge de seguida
[1] "2 + 3 = 5"

Comandos de ajuda

Através da janela principal do R ("R Console"):

help.search("*palavras-chave separadas por espaço*") (pesquisa por palavras-chave)
help("*nomedafunção*") ou ?*nomedafunção* (ajuda específica de uma função)

Ou através do menu "Help" da janela principal do R