# Estatística Básica - Bussab e Morettin (2017)

## Comandos R para análises estatísticas

## Capítulo 1: Introdução

O R é um software livre de larga utilização em Estatística e seu número de usuários aumenta a cada dia. O objetivo deste texto é, de forma simples, descrever de que forma alternativa podemos realizar análises presentes no livro *Estatística Básica* (Bussab e Morettin, 2017) utilizando o software livre R. Como o R é de código aberto ele possui um amplo número de implementações que nos auxiliam em tal tarefa. Por conta disso, as soluções nem sempre são únicas, no sentido de que muitas vezes possuímos a mesma função implementada (não necessariamente da mesma forma), em diversos pacotes diferentes. Sempre que possível, descreveremos diferentes formas de realizar as análises, para que o leitor tenha a opção de escolher aquela que melhor lhe agrada. É importante ainda estar familiarizado com a linguagem R. Você pode encontrar detalhes sobre ela [nos manuais do R](https://cran.r-project.org/manuals.html). Uma série de outras referências e comandos básicos também pode ser encontrada [aqui](https://rpubs.com/samejimakim/IntrodR). Se o leitor desejar se aprofundar na utilização do R e dos exemplos do livro com este programa, há uma página detalhada deste conteúdo com comandos e saídas que pode ser acessada clicando [aqui](https://rpubs.com/EstatBasica/Introd).

## Pacotes (bibliotecas, *packages*, ou *libraries*) utilizados neste texto:

**library**(akima)

**library**(boot)

**library**(bootstrap)

**library**(colorRamps)

**library**(MASS)

**library**(ggExtra)

**library**(ggplot2)

**library**(gmodels)

**library**(knitr)

**library**(mvtnorm)

**library**(pander)

**library**(reshape2)

**library**(scales)

**library**(scatterplot3d)

**library**(stats)

**library**(stats4)

**library**(tables)

**library**(xtable)

## Dados e funções programadas ao longo dos capítulos:

Se o leitor deseja apenas acessar os scripts das funções programadas num arquivo texto(extensão .R), clique [aqui](http://www.ime.usp.br/~pam/EstBas9.R). Os dados e funções programadas em uma imagem do R (arquivo .RData) podem ser acessados [aqui](http://www.ime.usp.br/~pam/dados.RData) ou simplesmente colando a seguinte linha no console do R:

**load**(**url**(description = "https://www.ime.usp.br/~pam/dados.RData"))

## Capítulo 2: Resumo de Dados

### Tabela 2.1

tab2\_1<-**read.table**("tabela2\_1.csv", dec=",", sep=";",h=T)

**names**(tab2\_1)

**summary**(tab2\_1$salario)

### Exemplo 2.2

### Tabela 2.2

ni<-**table**(tab2\_1$grau\_instrucao) # Calcula a tabela de frequências absolutas e armazena o resultado em 'mytab'

fi<-**prop.table**(ni) # Tabela de frequências relativas (f\_i)

p\_fi<-100\***prop.table**(ni) # Porcentagem (100 f\_i)

# Adiciona linhas de total

ni<-**c**(ni,**sum**(ni))

fi<-**c**(fi,**sum**(fi))

p\_fi<-**c**(p\_fi,**sum**(p\_fi))

**names**(ni)[4]<-"Total"

tab2\_2<-**cbind**(ni,fi=**round**(fi,digits=2),p\_fi=**round**(p\_fi,digits=2))

tab2\_2

### Tabela 2.3:

#quebras de linha apenas ilustrativas para facilitar a leitura

tab2\_3<-**as.data.frame**(

**t**(**rbind**(

ni=**c**(650,1020,330,2000),

p\_fi=**c**(32.5,51,16.5,1)

))

,row.names =**c**("Fundamental","Médio","Superior","Total")

)

tab2\_3

### Exemplo 2.3

### Tabela 2.4

ni<-**table**(**cut**(tab2\_1$salario, breaks = **seq**(4,24,by=4),right=FALSE)) *# Frequencias por categorias*

tab2\_4 <- **rbind**(ni, p\_fi = 100\***prop.table**(ni)) *# Frequencias relativas em %*

*#quebras de linha apenas ilustrativas para facilitar a leitura*

tab2\_4 <- **as.data.frame**(

**t**(**cbind**(

tab2\_4,

**c**(**sum**(tab2\_4[1,]),**sum**(tab2\_4[2,])

))),row.names =**c**(**colnames**(tab2\_4),"Total")) *#Construcao da tabela*

tab2\_4<-**transform**(tab2\_4,p\_fi=**round**(p\_fi,digits=2))

tab2\_4

## 2.3 Gráficos

### Exemplo 2.4

### Figura 2.2

#quebras de linha apenas ilustrativas para facilitar a leitura

**barplot**(

**table**(tab2\_1$grau\_instrucao),

ylab="Frequência",

cex.names=0.7,

names.arg = **c**("Fundamental","Médio", "Superior"),

col="darkgrey",

border=NA,

main="Figura 2.2: Gráfico em barras para a variável Y: grau de instrução.",

axes=TRUE,

ylim=**c**(0,20)

)

### Figura 2.3

labs<- **paste(1:3,"(",tab2\_1[1:3,1],";",paste(tab2\_1[1:3,3],1),"%)",sep="")**

**pie**(**table**(tab2\_1$grau\_instrucao),labels=labs)

*title("Figura 2.3: Gráfico em setores para a variável Y: grau de instrução")*

**legend**(-1.1,-0.8,legend=**c**("1=Fundamental, 2=Médio, 3=Superior"),border=NA,box.col=NA)

### Exemplo 2.5

### Figura 2.4

#quebras de linha apenas ilustrativas para facilitar a leitura

**barplot**(

**table**(tab2\_1$n\_filhos),

ylab="Frequência",

cex.names=0.7,

col="darkgrey",

main="Figura 2.4: Gráfico em barras para a variável Z: Número de filhos.",

border=NA)

### Figura 2.5

**par**(mfrow=**c**(1,3),pin=**c**(2,2))

tbs<-**as.data.frame**(**cbind**(x=0:5,y=**c**(1,1,1,1,NA,1)),row.names = **as.integer**(**c**(4,5,7,3,NA,1)))

**plot**(tbs,ylim=**c**(0,7),pch=19,ylab=NA,bty="n",yaxt="n", col="darkblue",xlab="(a)")

**text**(x=tbs$x,y=tbs$y,**rownames**(tbs),pos=3)

**stripchart**(tab2\_1$n\_filhos,method = "stack", offset = 1, pch = 19, col="darkblue",ylim=**c**(0,7),ylab=NA,bty="n",yaxt="n",xlab="(b)",cex=1)

**plot**(**table**(tab2\_1$n\_filhos),type="p", col="darkblue",pch = 19,bty="n",ylab=NA,xlab="(c)")

### Tabela 2.5

ni<-**table**(tab2\_1$n\_filhos) *# Frequencias absolutas*

tab2\_5 <- **rbind**(ni, p\_fi = 100\***prop.table**(ni)) *# Frequencias relativas em %*

*#quebras de linha apenas ilustrativas para facilitar a leitura*

tab2\_5 <- **as.data.frame**(

**t**(**cbind**(

tab2\_5,

**c**(**sum**(tab2\_5[1,]),**sum**(tab2\_5[2,])

))),row.names =**c**(**colnames**(tab2\_5),"Total")) *#Construcao da tabela*

tab2\_5<-**transform**(tab2\_5,p\_fi=**round**(p\_fi,digits=2))

tab2\_5

### Exemplo 2.6

### Figura 2.6

#quebras de linha apenas ilustrativas para facilitar a leitura

**barplot**(

**table**(**cut**(tab2\_1$salario, breaks = **seq**(4,24,by=4),right=FALSE)),

ylab="Frequência",

xaxt="n",

cex.names=0.7,

col="darkgrey",

border=NA,

#main="Figura 2.6: Gráfico em barras para a variável S: salários" )

**axis**(1,at=**c**(.75,1.9,3.1,4.3,5.5),labels=**seq**(6,22,4),tick=F)

### Tabela 2.6

ni<-**table**(**cut**(tab2\_1$salario, breaks = **seq**(4,24,by=4),right=FALSE)) *# Frequencias por categorias*

*#quebras de linha apenas ilustrativas para facilitar a leitura*

tab2\_6 <- **rbind**(si=**seq**(6,22,by=4),ni, p\_fi = 100\***prop.table**(ni)) *# Frequencias relativas em %*

tab2\_6 <- **as.data.frame**(

**t**(**cbind**(tab2\_6,

**c**(NA,**sum**(tab2\_6[2,]),**sum**(tab2\_6[3,])

))),row.names =**c**(**colnames**(tab2\_6),"Total")) *#Construcao da tabela*

tab2\_6<-**transform**(tab2\_6,p\_fi=**round**(p\_fi,digits=2))

tab2\_6

### Exemplo 2.7

### Figura 2.7

fig27<-**hist**(tab2\_1$salario, breaks = **seq**(4,24,by=4),right=FALSE,probability = T,plot=F)

aux<-**with**(fig27, 100 \* density\* **diff**(breaks)[1])

labs <- **paste**(**round**(aux), "%", sep="")

*#quebras de linha apenas ilustrativas para facilitar a leitura*

**plot**(fig27,

freq = FALSE, labels = labs,

ylab="Densidade de Frequência",

xlab="Salário",

col="darkgrey",

border="white",

*labels=T,*

*main="Figura 2.7: Histograma da variável S: salários",*

xlim=**c**(0,24), xaxp=**c**(0,24,6),

ylim=**c**(0,.1))

### Figura 2.8

fig28<-**hist**(tab2\_1$n\_filhos, right=F, breaks=**seq**(-.5,5.5,1),plot=F)

aux<-**with**(fig28, 100 \* density\* **diff**(breaks)[1])

labs <- **paste**(**round**(aux), "%", sep="")

*#quebras de linha apenas ilustrativas para facilitar a leitura*

**plot**(fig28,

ylab="Densidade de Frequência",

xlab="Número de Filhos",

col="darkgrey",

border="white",

bty="n",yaxt="n",ylim=**c**(0,8),

*main="Figura 2.8: Histograma da variável Z: número de filhos"*

labels=labs)

## 2.4 Ramo-e-Folhas

### Exemplo 2.8

### Figura 2.9

print("Figura 2.9: Ramo-e-folhas para a Variável S: salários.")

**stem**(tab2\_1$salario,scale=2)

### Exemplo 2.9

### Figura 2.10

#quebras de linha apenas ilustrativas para facilitar a leitura

dureza<-**c**(53 ,70.2,84.3,69.5,77.8,87.5,53.4,82.5,67.3,54.1,

70.5,71.4,95.4,51.1,74.4,55.7,63.5,85.8,53.5,64.3,

82.7,78.5,55.7,69.1,72.3,59.5,55.3,73 ,52.4,50.7

)

**stem**(**as.integer**(dureza),scale=.5)

### Figura 2.11

print("Figura 2.11: Ramo-e-folhas para dados de dureza, com ramos divididos.")

**stem**(**as.integer**(dureza),scale=1)

## 2.5 Exemplos Computacionais

## Exemplo 2.10

### Figura 2.12

cd\_notas<-**read.table**("cd-notas.csv",h=T,skip=4,sep=";",dec=",")

**hist**(cd\_notas$nota,col="darkgrey",

*main="Figura 2.12: Histograma para o CD-Notas. R.",*

xlab="Notas",ylab="Frequência",border="white")

**Figura 2.12:** Histograma para o CD-Notas. R.

#### Figura 2.13

**stripchart**(cd\_notas$nota,method = "stack", offset = 2, at=0,

*main="Figura 2.13: Gráfico de dispersão unidimensional para CD-Notas. R.",*

pch = 19, col="darkblue",ylab=NA,cex=0.5)

Figura 2.14

**stem**(cd\_notas$nota)

Exemplo 2.11

Figura 2.15

cd\_poluicao<-**read.table**("cd-poluicao.csv",h=T,skip=8,sep=";",dec=",")

*#quebras de linha apenas ilustrativas para facilitar a leitura*

**plot.ts**(cd\_poluicao$temp,

*main="Figura 2.15: Dados de Temperatura de São Paulo. R.",*

xlab="Dia", ylab="Grau",col="darkgrey")

Figura 2.16

*#quebras de linha apenas ilustrativas para facilitar a leitura*

**hist**(cd\_poluicao$temp,col="darkgrey",xlab="Temperatura",border="white",

*main="Figura 2.16: Histograma dos dados de Temperatura de São Paulo. R.",*

ylab="")

Figura 2.17

**stripchart**(cd\_poluicao$temp, method = "stack", offset = 2, at=0,

*main="Figura 2.17: Gráfico de dispersão unidimensional \n para os dados de temperatura de São Paulo. R.",*

pch = 19, col="darkblue",ylab=NA,cex=0.5)

Figura 2.18

*print("Figura 2.18: Ramo-e-Folha para os dados de temperatura de São Paulo. R.")*

**stem**(cd\_poluicao$temp, scale=.5)

Capítulo 3: Medidas-Resumo

3.1 Medidas de Posição

Exemplo 3.1

*#mean(x, na.rm = FALSE, ...)*

*# na.rm: Se este argumento for TRUE, então os valores NA são desconsiderados para o calculo da média.*

**mean**(tab2\_1$n\_filhos,na.rm=T)

Exemplo 3.2

fig27<-**hist**(tab2\_1$salario, breaks = **seq**(4,24,by=4),right=FALSE, plot=F)

**median**(tab2\_1$salario)

**moda2**(tab2\_1$grau\_instrucao)

tab2\_1$grau\_instr\_num<-1\***c**(tab2\_1$grau\_instrucao=="ensino fundamental")+2\***c**(tab2\_1$grau\_instrucao=="ensino médio")+3\***c**(tab2\_1$grau\_instrucao=="superior")

**median**(tab2\_1$grau\_instr\_num)

**levels**(**as.factor**(tab2\_1$grau\_instrucao))[**median**(tab2\_1$grau\_instr\_num)]

**mean**(tab2\_1$salario[

tab2\_1$salario>**quantile**(tab2\_1$salario,probs=0.01)

&

tab2\_1$salario<**quantile**(tab2\_1$salario,probs=0.99)

])

*#Note que este valor é menor que o valor da média da variável sem a "truncagem":*

**mean**(tab2\_1$salario)

3.2 Medidas de Dispersão

Exemplo 3.3

n<-**sum**(!**is.na**(tab2\_1$n\_filhos))

*#Variância:*

**var**(tab2\_1$n\_filhos,na.rm=T)\*(n-1)/n

*#Desvio padrão:*

**sd**(tab2\_1$n\_filhos,na.rm=T)\*(n-1)/n

*#Desvio médio:*

**mean**(**abs**(tab2\_1$n\_filhos[!**is.na**(tab2\_1$n\_filhos)]-**mean**(tab2\_1$n\_filhos,na.rm=T)))

*# Exemplo:*

x<-**c**(0,1,NA,NA,0,-2,pi,NA)

**is.na**(x)

!**is.na**(x)

Exemplo 3.4

**mean**(tab2\_1$salario,na.rm=T)

**dm**(tab2\_1$salario)

*#n<-sum(!is.na(tab2\_1$n\_filhos))*

**varp**(tab2\_1$salario,na.rm=T)

3.3 Quantis Empíricos

Exemplo 3.5

x<-**c**(15,5,3,8,10,2,7,11,12)

**median**(x)

**mean**(x)

**quantile**(x,probs=**c**(.25,.50,.75),type=5) *# "type 6" indica o mesmo método de cálculo que o SPSS e Minitab*

*#Agora, adicionando a observação 67 ao vetor x, temos:*

x<-**c**(**sort**(x),67) *# A função `sort` ordena o vetor*

x

**median**(x)

**mean**(x)

**quantile**(x,probs=**c**(.25,.50,.75),type=5)

*# 20% das observações a esquerda*

**quantile**(x,probs=0.2)

Exemplo 3.6

fig27<-**hist**(tab2\_1$salario, breaks = **seq**(4,24,by=4),right=FALSE,probability = T,plot=F)

aux<-**with**(fig27, 100 \* density\* **diff**(breaks)[1])

labs <- **paste**(**round**(aux), "%", sep="")

*#quebras de linha apenas ilustrativas para facilitar a leitura*

**plot**(fig27,

freq = FALSE, labels = labs,

xlab="Salário",

ylab="",

col="darkgrey",

border="white",

yaxt="n",

xlim=**c**(0,24), xaxp=**c**(0,24,6),

ylim=**c**(0,.1), main="")

**median**(tab2\_1$salario)

**quantile**(tab2\_1$salario,probs=**c**(.20,.5,.95,.75))

q<-**quantile**(tab2\_1$salario,probs=**c**(.25,.5,.75)) *# armazenando as informações de quantis em um objeto `q`*

q

dq=q[3]-q[1]

**print**(**paste**("Distância inter-quartis (d) = ", dq))

Figuras 3.2 e 3.3

*# Utilizando os dados do exemplo 3.5 :*

x<-**c**(15,5,3,8,10,2,7,11,12)

**summary**(x)

Exemplo 3.7

cd\_municipios<-**read.table**("cd-municipios.csv",h=T,skip=4,sep=";",dec=",")

*# Medidas resumo para a base de dados dos municípios (n=30)*

**summary**(cd\_municipios$populacao,digits = 4)

*# Medidas resumo para a base de dados dos municípios, desconsiderando São Paulo e Rio de Janeiro (n=28)*

**summary**(cd\_municipios$populacao[-**c**(1,2)],digits = 4)

3.4 *Box plots*

Exemplo 3.8

munic<-cd\_municipios[**order**(cd\_municipios$populacao,decreasing = TRUE),]

munic

munic<-cd\_municipios[1:15,]

munic

*# Medidas resumo para a base de dados dos 15 municípios mais populosos (n=15)*

**summary**(munic$populacao)

Figuras 3.6 e 3.7

**boxplot**(munic$populacao,

pch="\*", *# tipo de marcador dos outliers*

col="lightblue", *# cor do preenchimento do box plot*

border="darkgrey", *# cor da linha do box plot*

boxwex=0.3 *# Tamanho da caixa*

)

*# Os comandos 'text'a seguir imprimem no box plot os nomes das cidades dos 4 pontos destacados*

**text**(x=cd\_municipios$populacao[1],label=cd\_municipios$municipio[1],pos=4,cex=0.7) *# Máximo*

**text**(x=cd\_municipios$populacao[2],label=cd\_municipios$municipio[2],pos=4,cex=0.7) *# 2a. maior observacao*

**text**(x=cd\_municipios$populacao[3],label=cd\_municipios$municipio[3],pos=4,cex=0.7) *# 3a. maior observacao*

**text**(x=cd\_municipios$populacao[15],label=cd\_municipios$municipio[15],pos=4,cex=0.7) *# 15a. observacao*

*#Para imprimir o box plot desconsiderando aqueles pontos que foram considerados outliers de acordo com seu critério, faça:*

**boxplot**(munic$populacao, col="lightblue", border="darkgrey", boxwex=0.3, outline=FALSE)

**title**("Box plot para os quinze maiores \n municípios do Brasil, desconsiderando outliers")

3.5 Gráficos de Simetria

Exemplo 3.9

x<-**c**(0.5,2.3,4,6.4,8,15.3,13.5,12,9.8)

*# Mediana de x*

**median**(x)

*# Calculando os ui e vi*

u<-**sort**((**median**(x)-x)[(**median**(x)-x)>0],decreasing = T)

v<-**sort**((x-**median**(x))[(x-**median**(x))>0],decreasing = T)

u

v

Figura 3.11

u<-**median**(cd\_municipios$populacao)-cd\_municipios$populacao

v<-cd\_municipios$populacao-**median**(cd\_municipios$populacao)

**plot**(**sort**(u),**sort**(v), pch=19, xlab="ui", ylab="vi",col="darkblue",xlim=**c**(0,**max**(u)),ylim=**c**(0,**max**(v)))

*#title("Figura 3.11: Gráfico de simetria para o CD-Municípios.")*

**abline**(0,1)

Transformações

Exemplo 3.10 e Figura 3.12

xp<-**list**() *# declarando xp como uma lista vazia*

xp[[1]]<-**log**(cd\_municipios$populacao) *# Transformação log*

xp[[2]]<-(cd\_municipios$populacao)^(1/4) *# Transformação p=1/4*

xp[[3]]<-(cd\_municipios$populacao)^(1/2) *# Transformação p=1/2*

xp[[4]]<-(cd\_municipios$populacao)^(1/3) *# Transformação p=1/3*

**par**(mfrow=**c**(2,2))

**hist**(xp[[1]], main="log", ylab="", xlab="", col="darkgrey", border="white",cex.axis=0.8,cex.main=0.8)

**hist**(xp[[2]], main="p=1/4", ylab="", xlab="", col="darkgrey", border="white",cex.axis=0.8,cex.main=0.8)

**hist**(xp[[3]], main="p=1/2", ylab="", xlab="", col="darkgrey", border="white",cex.axis=0.8,cex.main=0.8)

**hist**(xp[[4]], main="p=1/3", ylab="", xlab="", col="darkgrey", border="white",cex.axis=0.8,cex.main=0.8)

Figura 3.13

**boxplot**(xp,pch="-", col="lightblue", border="darkgrey", boxwex=0.3, names=**c**("p=0","p=1/4","p=1/2","p=1/3"))

3.7 Exemplos Computacionais

Exemplo 2.10 (continuação)

*#Podemos construir uma nova função, digamos summary2, que conterá mais informação do que a summary atual:*

summary2<-function(x,limtrim=**c**(.01,.99), pop=FALSE, na.rm=TRUE, ...){

*# x : argumento numerico para o qual calcularemos as estatísticas descritivas*

*# limtrim: Limites para a média truncada*

*# pop: se esta variável é TRUE, entao utiliza-se 'n' como denominador da estimativa da variância. caso contrário, 'n-1'.*

*# na.rm: ação para os outliers*

pop=TRUE

na.rm=TRUE

newsumm<-**matrix**(NA,nrow=**dim**(**as.matrix**(x))[2],ncol=11)

x<-**as.matrix**(x)

newsumm[,1]<-**apply**(x,2,function(...){**return**(**dim**(x)[1])})

newsumm[,2:7]<-**t**(**apply**(**as.data.frame**(x),2,summary))

newsumm[,8]<-**apply**(x,2,function(...){**mean**(x[x>**quantile**(x,probs=0.01) &

x<**quantile**(x,probs=0.99)])})

if (pop==TRUE & na.rm==TRUE){ *# calcula var\_n desconsiderando os missings*

n<-**sum**(!**is.na**(x))

newsumm[,9]<-**var**(x,na.rm=na.rm)\*(n-1)/n

}

else

{

if(pop==TRUE & na.rm==FALSE ){

n<-**length**(x)

newsumm[,9]<-**var**(x)\*(n-1)/n

}

else {

if(pop==FALSE & na.rm==TRUE){

n<-**sum**(!**is.na**(x))

newsumm[,9]<-**var**(x,na.rm=TRUE)

}

else

{

n<-**length**(x)

newsumm[,9]<-**var**(x)

}

}

}

newsumm[,1]<-n

newsumm[,10]<-**sqrt**(newsumm[,9])

newsumm[,11]<-**sqrt**(newsumm[,10]/newsumm[,1])

**colnames**(newsumm)<-**c**("N", "Min.","1st Qu.","Median","Mean","3rd Qu.","Max.","Tr Mean","Var","StDev","SE Mean")

**return**(**t**(newsumm))

}

cd\_notas<-**read.table**("cd-notas.csv",header=TRUE,skip=4,sep=";",dec=",") *# Leitura dos dados*

**summary2**(cd\_notas$nota)

Figura 3.14

**boxplot**(cd\_notas$nota,pch="-", col="lightblue", border="darkgrey")

Figura 3.15

u<-**median**(cd\_notas$nota)-cd\_notas$nota

v<-cd\_notas$nota-**median**(cd\_notas$nota)

**plot**(**sort**(u),**sort**(v), pch=19, xlab="ui", ylab="vi",col="darkblue",xlim=**c**(0,**max**(u)),ylim=**c**(0,**max**(v)))

**title**("Figura 3.15: Gráfico de simetria para o CD-Notas.")

**abline**(0,1)

Exemplo 2.11 (continuação)

*# Quadro 3.4 Medidas descritivas para temperaturas. R.*

**summary**(cd\_poluicao$temp)

Figura 3.16

**boxplot**(cd\_poluicao$temp,pch="-", col="lightblue", border="darkgrey")

Figura 3.17

u<-**median**(cd\_poluicao$temp)-cd\_poluicao$temp

v<-cd\_poluicao$temp-**median**(cd\_poluicao$temp)

**plot**(**sort**(u),**sort**(v), pch=19, xlab="ui", ylab="vi",col="darkblue",xlim=**c**(0,**max**(u)),ylim=**c**(0,**max**(v)))

**abline**(0,1)

Figura 3.18

x<-**c**(15,5,3,8,10,2,7,11,12)

**plot**(**ecdf**(x),xlim=**c**(0,16),ylim=**c**(0,1),main="",ylab="",pch=20,cex=0.7) *# 'ecdf' é a função que calcula a função distribuição empírica*

**lines**(y = **c**(0.5,0.5),x=**c**(-1,8),col="grey")

**lines**(y = **c**(0,0.5),x=**c**(8,8),col="grey")

**text**(-.5,0.5,"p=9/18",pos=4,col="darkblue")

**text**(8,0,"8",pos=3,col="darkblue")

*#title("Figura 3.18: Funções de distribuição empírica(F\_e) \n e f.d.e. alisada (Fs\_e) para o Exemplo 3.5")*

**par**(new=T) *# Utilize este comando para sobrepor o gráfico a seguir no gráfico anterior*

*# Construiremos a seguir uma função 'scdf', que será a ecdf acima suavisada*

scdf<-function(x){

f<-t<-**table**(x)

st<-**sum**(t[1:**length**(x)])

for(i in 1:**length**(x))

f[i]<-**sum**(t[1:i])/st - (t[i]/st)/2

**return**(f)

}

**par**(new=T)

**plot**(**scdf**(x),pch=20,cex=0.7,lty=2, type="b",xlab="",ylab="",col="grey",xaxt="n",yaxt="n",xlim=**c**(0,16),ylim=**c**(0,1))

**legend**(12,.4,**c**("F\_e","Fs\_e"),lty=**c**(1,2),col=**c**("black","grey"),lwd=2,cex=0.8)

Capítulo 4: Análise Bidimensional

4.2 Variáveis Qualitativas

Exemplo 4.1

Tabela 4.2

**attach**(tab2\_1)

tab4\_2<-**table**(reg\_procedencia,grau\_instrucao)

tab4\_2

Total\_linha<-**margin.table**(tab4\_2,2) *# O argumento 2 define a marginal da linha*

Total\_coluna<-**margin.table**(tab4\_2,1) *# O argumento 1 define a marginal da coluna*

tab4\_2\_<-**rbind**(**cbind**(tab4\_2,Total\_coluna),**c**(Total\_linha, **sum**(Total\_coluna)))

**dimnames**(tab4\_2\_)[[1]][4]<-"Total\_linha"

tab4\_2\_

Total\_linha<-**apply**(tab4\_2, MARGIN=2, FUN=sum)

Total\_coluna<-**apply**(tab4\_2, MARGIN=1, FUN=sum)

tab4\_2\_<-**rbind**(**cbind**(tab4\_2,Total\_coluna),**c**(Total\_linha, **sum**(Total\_coluna)))

**dimnames**(tab4\_2\_)[[1]][4]<-"Total\_linha"

Tabela 4.3

tab4\_3<-**prop.table**(tab4\_2)

tab4\_3

Total\_linha<-**margin.table**(tab4\_3,2)

Total\_coluna<-**margin.table**(tab4\_3,1)

tab4\_3\_<-**rbind**(**cbind**(tab4\_3,Total\_coluna),**c**(Total\_linha, **sum**(Total\_coluna)))

**dimnames**(tab4\_3\_)[[1]][4]<-"Total\_linha"

tab4\_3\_

Tabela 4.4

tab4\_4<-**prop.table**(tab4\_2,2)

tab4\_4

Total\_linha<-**margin.table**(tab4\_4,2)

Total\_coluna<-**margin.table**(tab4\_2,1)/**sum**(**margin.table**(tab4\_2,1))

tab4\_4\_<-**rbind**(**cbind**(tab4\_4,Total\_coluna),**c**(Total\_linha, **sum**(Total\_coluna)))

**dimnames**(tab4\_4\_)[[1]][4]<-"Total\_linha"

tab4\_4\_

Tabelas 4.2 a 4.4 utilizando o pacote gmodels

*#Tabela 4.2*

**CrossTable**(reg\_procedencia,grau\_instrucao,

prop.r=FALSE, *# Se TRUE, entao retorna as proporções nas linhas*

prop.c=FALSE, *# Se TRUE, entao retorna as proporções nas colunas*

prop.t=FALSE, *# Se TRUE, entao retorna as proporções em relação ao total*

prop.chisq=FALSE *# Se TRUE, entao retorna a contribuição de cada casela para a estatística de Qui-quadrado*

)

*#Tabela 4.3*

**CrossTable**(reg\_procedencia,grau\_instrucao,prop.r=FALSE, prop.c=FALSE,

prop.t=TRUE, prop.chisq=FALSE)

*#Tabela 4.4*

**CrossTable**(reg\_procedencia,grau\_instrucao,prop.r=FALSE, prop.c=TRUE,

prop.t=FALSE, prop.chisq=FALSE)

Figura 4.1

**ggplot**(**melt**(tab4\_2\_[1:3,],value.name = "contagem",

varnames = **c**("reg\_procedencia","grau\_instrucao") ))+ ## `melt` empilha os dados no formato necessário para o ggplot

**aes**(x=grau\_instrucao,y=contagem,fill=reg\_procedencia) + ## Variáveis a serem plotadas.

**geom\_bar**(stat="identity", position = "fill") + ## Define o gráfico de barras percentual empilhado

**scale\_fill\_brewer**(name="Região de\n Procedência")+ ## Opções do preenchimento do gráfico (label e paleta de cores)

**scale\_y\_continuous**(labels = **percent\_format**()) + ## Formato do eixo Y em porcentagem

**theme\_bw**()+ ## Define a cor do fundo do gráfico: neste caso, branco

*#theme(legend.position="bottom") + ## Define a posição da legenda abaixo do gráfico*

*#ggtitle("Figura 4.1: Distribuição da região de procedência por grau de instrução")+*

**xlab**("Grau de Instrução") + **ylab**("") ## Define os `labels` dos eixos

4.3 Associação entre Variáveis Qualitativas

Exemplo 4.2

Tabela 4.5

tab4\_5<-**as.table**(**matrix**(**c**(85,55,35,25),ncol=2))

**dimnames**(tab4\_5)[[1]] = **c**("Economia", "Administração")

**dimnames**(tab4\_5)[[2]] = **c**("Masculino", "Feminino")

tab4\_5

Total\_linha<-**margin.table**(tab4\_5,2)

Total\_coluna<-**margin.table**(tab4\_5,1)

tab4\_5\_<-**rbind**(**cbind**(tab4\_5,Total\_coluna),**c**(Total\_linha, **sum**(Total\_coluna)))

**dimnames**(tab4\_5\_)[[1]][3]<-"Total\_linha"

Tabela 4.6

tab4\_6<-**prop.table**(tab4\_5,2)

tab4\_6

Total\_linha<-**margin.table**(tab4\_6,2)

Total\_coluna<-**margin.table**(tab4\_6,1)/**sum**(**margin.table**(tab4\_6,1))

tab4\_6\_<-**rbind**(**cbind**(tab4\_6,Total\_coluna),**c**(Total\_linha, **sum**(Total\_coluna)))

**dimnames**(tab4\_6\_)[[1]][3]<-"Total\_linha"

Tabela 4.7

*# Construindo os dados do exemplo:*

dados.tab4\_7<-**data.frame**(**rbind**(

**matrix**(**rep**(**c**("1.Masculino","1.Física"),times=100),ncol=2,byrow=T),

**matrix**(**rep**(**c**("2.Feminino","1.Física"),times=20),ncol=2,byrow=T),

**matrix**(**rep**(**c**("1.Masculino","2.Ciências Sociais"),times=40),ncol=2,byrow=T),

**matrix**(**rep**(**c**("2.Feminino","2.Ciências Sociais"),times=40),ncol=2,byrow=T)))

**colnames**(dados.tab4\_7)<-**c**("sexo","curso")

**CrossTable**(dados.tab4\_7$curso,dados.tab4\_7$sexo,prop.r=FALSE, prop.c=TRUE,

prop.t=FALSE, prop.chisq=FALSE)

4.4 Medidas de Associação entre Variáveis Qualitativas

Exemplo 4.3

*# Construindo os dados do exemplo:*

dados.tab4\_8<-**data.frame**(**rbind**(

**matrix**(**rep**(**c**("1.Consumidor","1.São Paulo"),times=214),ncol=2,byrow=T),

**matrix**(**rep**(**c**("1.Consumidor","2.Paraná"),times=51),ncol=2,byrow=T),

**matrix**(**rep**(**c**("1.Consumidor","3.Rio G. do Sul"),times=111),ncol=2,byrow=T),

**matrix**(**rep**(**c**("2.Produtor","1.São Paulo"),times=237),ncol=2,byrow=T),

**matrix**(**rep**(**c**("2.Produtor","2.Paraná"),times=102),ncol=2,byrow=T),

**matrix**(**rep**(**c**("2.Produtor","3.Rio G. do Sul"),times=304),ncol=2,byrow=T),

**matrix**(**rep**(**c**("3.Escola","1.São Paulo"),times=78),ncol=2,byrow=T),

**matrix**(**rep**(**c**("3.Escola","2.Paraná"),times=126),ncol=2,byrow=T),

**matrix**(**rep**(**c**("3.Escola","3.Rio G. do Sul"),times=139),ncol=2,byrow=T),

**matrix**(**rep**(**c**("4.Outras","1.São Paulo"),times=119),ncol=2,byrow=T),

**matrix**(**rep**(**c**("4.Outras","2.Paraná"),times=22),ncol=2,byrow=T),

**matrix**(**rep**(**c**("4.Outras","3.Rio G. do Sul"),times=48),ncol=2,byrow=T)))

**colnames**(dados.tab4\_8)<-**c**("tipo\_de\_cooperativa","estado")

**attach**(dados.tab4\_8)

Tabela 4.8: Cooperativas autorizadas a funcionar por tipo e estado, junho de 1974.

**CrossTable**(estado,tipo\_de\_cooperativa,

prop.r=TRUE, prop.c=FALSE, prop.t=FALSE, prop.chisq=FALSE,

digits=2)

Tabela 4.9: Valores esperados na Tabela 4.8 assumindo a independência entre as duas variáveis.

**CrossTable**(estado,tipo\_de\_cooperativa,

prop.r=FALSE, prop.c=FALSE, prop.t=FALSE, prop.chisq=FALSE, expected=TRUE,

digits=0)

Tabela 4.10: Desvios observados e esperados.

**CrossTable**(estado,tipo\_de\_cooperativa,

prop.r=FALSE, prop.c=FALSE, prop.t=FALSE, resid=TRUE, prop.chisq=TRUE,

format="SPSS", digits=2)

tab4\_8<-**table**(estado,tipo\_de\_cooperativa)

testequi<-**chisq.test**(tab4\_8)

testequi

Expressões (4.5) e (4.6):

### (4.5) ###

C = **sqrt**(testequi$statistic/(testequi$statistic+1551))

**round**(C,digits=2) *# Arredondando para duas casas decimais*

### (4.6) ###

T = **sqrt**(testequi$statistic/(1551\*(4-1)\*(3-1)))

**round**(T,digits=2) *# Arredondando para duas casas decimais*

4.5 Associação entre Variáveis Quantitativas

Exemplo 4.4

dados.ex4\_4<-**data.frame**(agente=**c**("A","B","C","D","E","F","G","H","I","J"),

anos\_servico=**c**(2,3,4,5,4,6,7,8,8,10),

n\_clientes=**c**(48,50,56,52,43,60,62,58,64,72))

Tabela 4.12

**print**(**kable**(dados.ex4\_4, caption="\*\*Tabela 4.12\*\* Número de anos de serviço(X) por número de clientes(Y) de agentes de uma companhia de seguros."))

Figura 4.2

**attach**(dados.ex4\_4)

**plot**(n\_clientes,anos\_servico,pch=20, col="darkblue",

xlab="Anos de Serviço", ylab="Número de Clientes",cex=2)

Exemplo 4.5

Figura 4.3

cd\_brasil<-**read.table**("cd-brasil.csv",h=T,skip=7,sep=";")

**names**(cd\_brasil)

**attach**(cd\_brasil)

**plot**(pop\_urbana,pop\_rural,pch=20, col="darkblue",

xlab="População urbana",ylab="População rural",cex=2)

Exemplo 4.6

dados.ex4\_6<-**data.frame**(

familia=**c**("A","B","C","D","E","F","G","H","I","J"),

renda\_bruta=**c**(12,16,18,20,28,30,40,48,50,54),

gasto\_saude=**c**(7.2,7.4,7,6.5,6.6,6.7,6,5.6,6,5.5))

**attach**(dados.ex4\_6)

**plot**(renda\_bruta,gasto\_saude,pch=20, col="darkblue",

xlab="Renda Bruta",ylab="% de gasto com saúde",cex=2)

Tabela 4.14

dados.tab4\_14<-**data.frame**(

individuo=**c**("A","B","C","D","E","F","G","H"),

resultado=**c**(45,52,61,70,74,76,80,90),

tempo=**c**(343,368,355,334,337,381,345,375))

**attach**(dados.tab4\_14)

**plot**(resultado,tempo,pch=20, col="darkblue",

xlab="Resultado teste",ylab="Tempo",cex=2)

Exemplo 4.7

**attach**(dados.ex4\_4)

tab4\_15<-dados.ex4\_4

tab4\_15$dsvx<-anos\_servico-**mean**(anos\_servico) *# Desvio da variável X em relação a sua média*

tab4\_15$dsvy<-n\_clientes-**mean**(n\_clientes) *# Desvio da variável Y em relação a sua média*

tab4\_15$zx<-tab4\_15$dsvx/**sqrt**(**varp**(anos\_servico)) *# Desvios de X padronizados pelo seu desvio-padrão*

tab4\_15$zy<-tab4\_15$dsvy/**sqrt**(**varp**(n\_clientes)) *# Desvios de Y padronizados pelo seu desvio-padrão*

tab4\_15$zxzy<-tab4\_15$zx\*tab4\_15$zy *# Coeficiente de correlação entre X e Y*

**attach**(tab4\_15)

**par**(mfrow=**c**(1,2))

**plot**(dsvx,dsvy,pch=20, col="darkblue",

xlab=**expression**(x-**bar**(x)),ylab=**expression**(y-**bar**(y)))

**abline**(v=0, lty=2,col="grey")

**abline**(h=0, lty=2,col="grey")

**plot**(zx,zy,pch=20, col="darkblue",

xlab=**expression**(z[x]),ylab=**expression**(z[y]))

**abline**(v=0, lty=2,col="grey")

**abline**(h=0, lty=2,col="grey")

*#cor(X,Y) # Calcula a correlação linear entre X e Y*

*#cov(X,Y) # Calcula a covariância entre X e Y*

4.6 Associação entre Variáveis Qualitativas e Quantitativas

Exemplo 4.8

**attach**(tab2\_1)

**tapply**(salario,grau\_instrucao,summary)-> summary.tab4\_16 *# Calculando medidas de posição por grupo*

**tapply**(salario,grau\_instrucao,varp)-> varp.tab4\_16 *# Calculando variancia por grupo*

**tapply**(salario,grau\_instrucao,length)-> n.tab4\_16 *# Calculando n por grupo*

summary.tab4\_16[[4]]<-**summary**(salario) *# Calculando medidas de posição de toda a amostra*

varp.tab4\_16[[4]]<-**varp**(salario) *# Calculando variancia de toda a amostra*

n.tab4\_16[[4]]<-**length**(salario) *# Calculando n*

tab4\_16 <- **matrix**(**unlist**(summary.tab4\_16), ncol = 6, byrow = TRUE)

tab4\_16<-**cbind**(n.tab4\_16,tab4\_16,**sqrt**(varp.tab4\_16),varp.tab4\_16)

**dimnames**(tab4\_16)<-**list**(

**c**("Fundamental","Medio","Superior", "Todos"),

**c**("n", "Min.","Q1","Mediana","Média","Q3","Max.","dp(S)","Var(S)"))

Tabela 4.16

**print**(**kable**(tab4\_16,digits = 2,caption="\*\*Tabela 4.16:\*\* Medidas-resumo para a variável salário, segundo grau de instrução, na Companhia MB."))

Figura 4.8

**ggplot**(tab2\_1, **aes**(grau\_instrucao, salario)) + **geom\_boxplot**(fill = "darkblue", colour = "grey")

**attach**(tab2\_1)

**tapply**(salario,reg\_procedencia,summary)-> summary.tab4\_17 *# Calculando medidas de posição por grupo*

**tapply**(salario,reg\_procedencia,varp)-> varp.tab4\_17 *# Calculando variancia por grupo*

**tapply**(salario,reg\_procedencia,length)-> n.tab4\_17 *# Calculando n por grupo*

summary.tab4\_17[[4]]<-**summary**(salario) *# Calculando medidas de posição de toda a amostra*

varp.tab4\_17[[4]]<-**varp**(salario) *# Calculando variancia de toda a amostra*

n.tab4\_17[[4]]<-**length**(salario) *# Calculando n*

tab4\_17 <- **matrix**(**unlist**(summary.tab4\_17), ncol = 6, byrow = TRUE)

tab4\_17<-**cbind**(n.tab4\_17,tab4\_17,**sqrt**(varp.tab4\_17),varp.tab4\_17)

**dimnames**(tab4\_17)<-**list**(

**c**("Capital","Interior","Outra", "Todos"),

**c**("n", "Min.","Q1","Mediana","Média","Q3","Max.","dp(S)","Var(S)"))

Tabela 4.17

**print**(**kable**(tab4\_17,digits = 2,caption="\*\*Tabela 4.17:\*\* Medidas-resumo para a variável salário segundo a região de procedência, na Companhia MB."))

Figura 4.9

**ggplot**(tab2\_1, **aes**(reg\_procedencia, salario)) + **geom\_boxplot**(fill = "darkblue", colour = "grey") *# + geom\_jitter()*

Exemplo 4.9

*#R^2 para Salário por grau de instrução*

varS1=**t**(tab4\_16[1:3,1])%\*%tab4\_16[1:3,9]/**sum**(tab4\_16[1:3,1])

R2.1=1-varS1/tab4\_16[4,9]

**print**(**paste**("media(var(S)) :",**round**(varS1,2)))

**print**(**paste**("R^2 :",**round**(R2.1,2)))

*#R^2 para Salário por região de procedência*

varS2=**t**(tab4\_17[1:3,1])%\*%tab4\_17[1:3,9]/**sum**(tab4\_17[1:3,1])

R2.2=1-varS2/tab4\_17[4,9]

**print**(**paste**("media(var(S)) :",**round**(varS2,2)))

**print**(**paste**("R^2 :",**round**(R2.2,2)))

4.7 Gráficos *q x q*

Exemplo 4.10

tab4\_18<-**data.frame**(aluno=1:20,prova1=**c**(8.5,3.5,7.2,5.5,9.5,7,4.8,6.6,2.5,7,7.4,5.6,6.3,3,8.1,3.8,6.8,10,4.5,5.9),prova2=**c**(8,2.8,6.5,6.2,9,7.5,5.2,7.2,4,6.8,6.5,5,6.5,3,9,4,5.5,10,5.5,5))

**plot**(tab4\_18$prova1,tab4\_18$prova2,xlab="Quantis da 1a. prova",ylab="Quantis da 2a. prova",pch=16,col="darkblue")

**abline**(a=0,b=1)

Tabela 4.18

**print**(**kable**(tab4\_18,caption="\*\*Tabela 4.18:\*\* Notas de 20 alunos em duas provas de Estatística."))

Exemplo 4.11

cd\_temperaturas<-**read.table**("cd-temperaturas.csv",h=T,skip=4,sep=";",dec=",") *# Leitura dos dados*

**attach**(cd\_temperaturas)

q\_cananeia<-**quantile**(cananeia,probs = **seq**(0,1,1/120)) *# seq(0,1,1/120) constroi uma sequencia de 0 a 1 com saltos de 1/120 de distância.*

q\_ubatuba<-**quantile**(ubatuba,probs = **seq**(0,1,1/120))

**plot**(q\_cananeia,q\_ubatuba,xlab="Quantis Cananéia",ylab="Quantis Ubatuba",pch=16,col="darkblue")

**abline**(a=0,b=1)

4.8 Exemplos computacionais

Exemplo 4.12

Figura 4.12

tab2\_1$idade<-tab2\_1$idade\_anos\*12+tab2\_1$idade\_meses

**attach**(tab2\_1)

**par**(mfrow=**c**(1,3),pin=**c**(2,2))

**plot**(idade[grau\_instrucao=="ensino fundamental"],salario[grau\_instrucao=="ensino fundamental"],main="Fundamental",xlab="Idade",ylab="Salário",pch=16,col="darkblue")

**plot**(idade[grau\_instrucao=="ensino médio"],salario[grau\_instrucao=="ensino médio"],main="Médio",xlab="Idade",ylab="Salário",pch=16,col="darkblue")

**plot**(idade[grau\_instrucao=="superior"],salario[grau\_instrucao=="superior"],main="Superior",xlab="Idade",ylab="Salário",pch=16,col="darkblue")

Exemplo 4.13

Leitura dos dados:

cd\_mercado <- **read.table**("cd-mercado.csv",h=T,skip=4,sep=";",dec=",") *# Leitura dos dados*

**attach**(cd\_mercado)

**plot**(telebras[1:39],indice[1:39],xlab="Telebrás",ylab="Ibovespa",pch=16,col="darkblue")

**abline**(**lm**(indice[1:39]~telebras[1:39]))

**print**(**paste**("Corr(X,Y) = ",**round**(**cor**(indice[1:39],telebras[1:39]),2)))

Exemplo 4.14

Figura 4.14

cd\_veiculos <- **read.table**("cd-veiculos.csv",h=T,skip=4,sep=";",dec=",") *# Leitura dos dados*

**attach**(cd\_veiculos)

**ggplot**(cd\_veiculos, **aes**(comprimento,preco)) + **geom\_point**(**aes**(shape =N\_I,colour=N\_I), size = 4)

Capítulo 6: Variáveis Aleatórias Discretas

6.1 Introdução

Figura 6.1

Figura 6.4

A<-**matrix**(0,7,7) *# Matriz de setas: colunas são as origens e linhas, os destinos*

A[2,1]<-"2/5"

A[3,1]<-"3/5"

A[4,2]<-"2/4"

A[5,2]<-"2/4"

A[6,3]<-"3/4"

A[7,3]<-"1/4"

col <- A

col[] <- "blue"

pos=**rbind**(**c**(0.05,0.5),

**c**(0.3,0.33),

**c**(0.3,0.67),

**c**(0.7,0.2),

**c**(0.7,0.4),

**c**(0.7,0.6),

**c**(0.7,0.8)) *# Matriz de coordenadas dos nós*

knots=**c**(" ","V","B","V","B","V","B") *# labels dos nós.*

pp <- **plotmat**(A, pos = pos, name = knots,

box.cex = **c**(2,2,2,1,1,1,1), box.type = "none", box.prop = 1.0, *# Parâmetros do nó: tamanho da letra, tipo de borda e proporção*

cex.txt = 1, dtext=0, *#parâmetros do texto das setas: tamanho e posição*

arr.lcol = col, arr.col = col, arr.pos = 0.7, arr.type = "simple", arr.len = 0.3,

curve = 0, segment.from = 0.8, segment.to = 0.1,lwd = 2, *# Parametros complementares de cor, posição da ponta da seta, tipo de seta, espessura da linha, curvatura, comprimento da linha*

main = "Diagrama de árvore" *# titulo*

)

Figura 6.5

A<-**matrix**(0,7,7) *# Matriz de setas: colunas são as origens e linhas, os destinos*

A[2,1]<-"1/2"

A[3,1]<-"1/2"

A[4,2]<-"1/2"

A[5,2]<-"1/2"

A[6,3]<-"1/2"

A[7,3]<-"1/2"

col <- A

col[] <- "blue"

pos=**rbind**(**c**(0.05,0.5),

**c**(0.3,0.33),

**c**(0.3,0.67),

**c**(0.7,0.2),

**c**(0.7,0.4),

**c**(0.7,0.6),

**c**(0.7,0.8)) *# Matriz de coordenadas dos nós*

knots=**c**(" ","R","C","R","C","R","C") *# labels dos nós.*

pp <- **plotmat**(A, pos = pos, name = knots,

box.cex = **c**(2,2,2,1,1,1,1), box.type = "none", box.prop = 1.0, *# Parâmetros do nó: tamanho da letra, tipo de borda e proporção*

cex.txt = 1, dtext=0, *#parâmetros do texto das setas: tamanho e posição*

arr.lcol = col, arr.col = col, arr.pos = 0.7, arr.type = "simple", arr.len = 0.3,

curve = 0, segment.from = 0.8, segment.to = 0.1,lwd = 2, *# Parametros complementares de cor, posição da ponta da seta, tipo de seta, espessura da linha, curvatura, comprimento da linha*

main = "Diagrama de árvore" *# titulo*

)

Figura 6.7

lucro<-**data.frame**(x=**c**(15,10,5,-5), px=**c**(0.56 , 0.23, 0.02, 0.19))

**plot**(lucro$x,lucro$px, pch=16, col="darkblue", xlab="Lucro", ylab="p(x)", bty="n", ylim=**c**(0,0.7))

*# Alternativamente, pode-se modificar a posição do eixo Y utilizando o comando:*

*# axis(2,at=seq(0,0.6,0.1),labels=seq(0,0.6,0.1),pos=0)*

*# Se desejar isso, é necessário adicionar a opção `yaxt="n"` no comando que gera o gráfico.*

Figura 6.8

o.lucro<-lucro[**order**(lucro$x),] *# Ordenando o dataset*

o.lucro<-**cbind**(o.lucro,fda\_x=**cumsum**(o.lucro$px)) *# Calculando a fda de X*

**plot**(**stepfun**(o.lucro$x,**c**(0,o.lucro$fda\_x)),

main="f.d.a para a v.a. X = Lucro por montagem",

xlab="Lucro", ylab="p(x)", bty="n",col="darkblue",pch=20)

**lines**(x=**c**(5,5),y=**c**(0,o.lucro$fda\_x[1]),lty=2,col="gray")

**lines**(x=**c**(10,10),y=**c**(0,o.lucro$fda\_x[2]),lty=2,col="gray")

**lines**(x=**c**(15,15),y=**c**(0,o.lucro$fda\_x[3]),lty=2,col="gray")

Tabela 6.12

x = **c**(0,1,2,3)

px<-**dbinom**(x=x,size=3,p=1/2)

table612<-**data.frame**(x,px)

**print**(**kable**(table612,digits=3, align = **c**("r","c"),

caption="\*\*Tabela 6.12:\*\* Probabilidades binomiais para n=3 e p=1/2"))

Figura 6.12

**plot**(table612$x,table612$px, pch=16, col="darkblue", xlab="Lucro", ylab="p(x)", bty="n")

Exemplo 6.15

*# x: numero de peças defeituosas desejadas*

*# m: Número total de peças defeituosas*

*# n: Número total de peças não defeituosas*

*# k: Número de bolas retiradas*

**dhyper**(x=0 , m = 10, n = 90, k = 5)

Exemplo 6.17

x<-0:9

p.5=0.5

p.005=0.005

n=10

pbinomial.5<-**dbinom**(x,n,p.5)

ppoisson.5<-**dpois**(x,lambda=n\*p.5)

pbinomial.005<-**dbinom**(x,n,p.005)

ppoisson.005<-**dpois**(x,lambda=n\*p.005)

table613a<-**data.frame**(x,pbinomial.5,ppoisson.5,pbinomial.005,ppoisson.005)

table613a<-**rbind**(table613a,**c**(10,1-**sum**(pbinomial.5),1-**sum**(ppoisson.5),1-**sum**(pbinomial.005),1-**sum**(ppoisson.005)))

**print**(**kable**(table613a))

**attach**(table613a)

**par**(mfrow=**c**(1,2))

**plot.ts**(table613a[,2:3], plot.type = "single", ylab="p(x)", xlab="x",col=**c**("blue","red"), main="p=0.5")

**plot.ts**(table613a[1:3,4:5], plot.type = "single", ylab="p(x)", xlab="x",col=**c**("blue","red"), main="p=0.005")

Exemplo 6.19

x<-0:12

px<-**dbinom**(x,size=14,p=0.3)

fdax<-**cumsum**(px)

quadro61<-**data.frame**(x,px,fdax)

**print**(**kable**(quadro61,caption="\*\*Quadro 6.1\*\*: Probabilidades Binomiais geradas pelo R"))

Quadro 6.2

x<-0:17

px<-**dpois**(x,lambda=5.2)

fdax<-**cumsum**(px)

quadro62<-**data.frame**(x,px,fdax)

**print**(**kable**(quadro62,caption="\*\*Quadro 6.1\*\*: Probabilidades de Poisson geradas pelo R"))

Capítulo 7: Variáveis Aleatórias Contínuas

7.1 Introdução

x<-**seq**(-3,3,0.1) *# Cria um intervalo de -3 a 3*

fdnorm<-**dnorm**(x = x, mean = 0, sd=1) *# Calcula a fdp da distr. normal para o intervalo x*

fdanorm<-**pnorm**(q = x, mean = 0, sd=1) *# Calcula a fda da distr. normal para o intervalo x*

## Imprimindo os gráficos da fdp e fda:

**par**(mfrow=**c**(1,2))

**plot**(x=x,y=fdnorm,type="l", col="blue",lwd=2, main="f.d.p. da Distrib. Normal padrão",xlab="z")

**plot**(x=x,y=fdanorm,type="l", col="blue",lwd=2, main="f.d.a. da Distrib. Normal padrão",xlab="z")

**lines**(x=**c**(0,0),y=**c**(0,fdanorm[x==0]),lty=2, col="gray")

**par**(mfrow=**c**(1,2))

**curve**(**dnorm**(x = x, mean = 0, sd=1), xlim=**c**(-3,3), col="blue",lwd=2, main="f.d.p. da Distrib. Normal padrão",xlab="z")

**curve**(**pnorm**(q = x, mean = 0, sd=1), xlim=**c**(-3,3), col="blue",lwd=2, main="f.d.a. da Distrib. Normal padrão",xlab="z")

**lines**(x=**c**(0,0), y=**c**(0,fdanorm[x==0]), lty=2, col="gray")

Figura 7.8

z\_a=-1.96

z\_b=1.96

*#Cálculo da f.d.a até z\_a = -1.96*

**pnorm**(q = z\_a, mean = 0, sd = 1)

*#Cálculo da f.d.a até z\_b = 1.96*

**pnorm**(q = z\_b, mean = 0, sd = 1)

*#Cálculo da f.d.a entre z\_a = -1.96 e z\_b = 1.96*

**pnorm**(q = z\_b, mean = 0, sd = 1) - **pnorm**(q = z\_a, mean = 0, sd = 1)

**par**(mfrow=**c**(1,3))

regiao=**seq**(-3,-1.96,0.01)

cord.x <- **c**(**min**(regiao),regiao,**max**(regiao))

cord.y <- **c**(0,**dnorm**(regiao),0)

**curve**(**dnorm**(x,0,1),xlim=**c**(-3,3),main='Normal Padrão(até -1.96)',xlab="z")

**polygon**(cord.x,cord.y,col='lightgray')

regiao=**seq**(-3,1.96,0.01)

cord.x <- **c**(**min**(regiao),regiao,**max**(regiao))

cord.y <- **c**(0,**dnorm**(regiao),0)

**curve**(**dnorm**(x,0,1),xlim=**c**(-3,3),main='Normal Padrão(até +1.96)',xlab="z")

**polygon**(cord.x,cord.y,col='lightgray')

regiao=**seq**(-1.96,1.96,0.01)

cord.x <- **c**(**min**(regiao),regiao,**max**(regiao))

cord.y <- **c**(0,**dnorm**(regiao),0)

**curve**(**dnorm**(x,0,1),xlim=**c**(-3,3),main='Normal Padrão(de -1.96 a 1.96)',xlab="z")

**polygon**(cord.x,cord.y,col='lightgray')

Figura 7.12

z<-**seq**(-3,3,0.1) *# Cria um intervalo de -3 a 3*

fdnorm<-**dnorm**(x = z, mean = 0, sd=1) *# Calcula a fdp da distr. normal para o intervalo x*

**plot**(x=z,y=fdnorm,type="l", col="blue",lwd=2, main="f.d.p. da Distrib. Normal padrão")

**lines**(x=**c**(1,1),y=**c**(0,fdnorm[x==1]),lty=2, col="gray") *# Linha cinza tracejada em z=-1*

**lines**(x=**c**(-1,-1),y=**c**(0,fdnorm[x==-1]),lty=2, col="gray") *# Linha cinza tracejada em z=1*

Figura 7.13

regiao=**seq**(-1,3,0.01) *# Regiao a ser sombreada na figura*

cord.x <- **c**(**min**(regiao),regiao,**max**(regiao))

cord.y <- **c**(0,**dnorm**(regiao,mean=2),0)

**curve**(**dnorm**(x,2,1),xlim=**c**(-1,5),bty="n", main='',xlab="",ylab="",xaxt="n") *#A opção `xaxt="n"` suprime o eixo x*

**axis**(1,at=3,labels = "y") *# Comando para imprimir o eixo x com labels especificos*

**polygon**(cord.x,cord.y,col='lightgray') *# imprime a regiao sombreada*

Figura 7.14

**curve**(**pnorm**(q = x, mean = 0, sd=1), xlim=**c**(-3,3), col="blue",lwd=2,ylab="",xlab="z")

**lines**(x=**c**(0,0), y=**c**(0,fdanorm[x==0]), lty=2, col="gray")

Figura 7.15

regiao=**seq**(1,2,0.01)

cord.x <- **c**(**min**(regiao),regiao,**max**(regiao))

cord.y <- **c**(0,**dnorm**(regiao),0)

**curve**(**dnorm**(x,0,1),xlim=**c**(-3,3),main='',xaxt="n",ylab="")

**axis**(1,at=**c**(1,2),labels = **c**("a","b"))

**polygon**(cord.x,cord.y,col='lightgray')

Figura 7.16

regiao=**seq**(0,2,0.01)

cord.x <- **c**(**min**(regiao),regiao,**max**(regiao))

cord.y <- **c**(0,**dnorm**(regiao),0)

**curve**(**dnorm**(x,0,1),xlim=**c**(-3,3),main='',xaxt="n",xlab="z") *#A opção `xaxt="n"` suprime o eixo x*

**axis**(1,at=**c**(0,2),labels = **c**("0","z\_c")) *# Comando para imprimir o eixo x com labels especificos*

**polygon**(cord.x,cord.y,col='lightgray') *# Comando para sombrear a região no gráfico*

Exemplo de cálculo da f.d.a. Normal

**pnorm**(1.73)-**pnorm**(0)

1. P(−1.73≤Z≤0)P(−1.73≤Z≤0)

**pnorm**(0)-**pnorm**(-1.73)

1. P(Z≥1.73)P(Z≥1.73)

1-**pnorm**(1.73)

1. P(Z<−1.73)P(Z<−1.73)

**pnorm**(-1.73)

1. P(0.47≤Z≤1.73)P(0.47≤Z≤1.73)

**pnorm**(1.73)-**pnorm**(0.47)

Figura 7.17

**par**(mfrow=**c**(1,3))

regiao=**seq**(0,1.73,0.01)

cord.x <- **c**(**min**(regiao),regiao,**max**(regiao))

cord.y <- **c**(0,**dnorm**(regiao),0)

**curve**(**dnorm**(x,0,1),xlim=**c**(-3,3),main='',xlab="z")

**polygon**(cord.x,cord.y,col='lightgray')

regiao1=**seq**(-3,-1.73,0.01)

regiao2=**seq**(1.73,3,0.01)

cord.x1 <- **c**(**min**(regiao1),regiao1,**max**(regiao1))

cord.x2 <- **c**(**min**(regiao2),regiao2,**max**(regiao2))

cord.y1 <- **c**(0,**dnorm**(regiao1),0)

cord.y2 <- **c**(0,**dnorm**(regiao2),0)

**curve**(**dnorm**(x,0,1),xlim=**c**(-3,3),main='',xlab="z")

**polygon**(cord.x1,cord.y1,col='lightgray')

**polygon**(cord.x2,cord.y2,col='lightgray')

regiao=**seq**(0.47,1.73,0.01)

cord.x <- **c**(**min**(regiao),regiao,**max**(regiao))

cord.y <- **c**(0,**dnorm**(regiao),0)

**curve**(**dnorm**(x,0,1),xlim=**c**(-3,3),main='',xlab="z")

**polygon**(cord.x,cord.y,col='lightgray')

**pnorm**(5,mean=3,sd=4)-**pnorm**(2,mean=3,sd=4)

Figura 7.18

**par**(mfrow=**c**(1,2))

regiao=**seq**(2,5,0.01)

cord.x <- **c**(**min**(regiao),regiao,**max**(regiao))

cord.y <- **c**(0,**dnorm**(regiao,3,4),0)

**curve**(**dnorm**(x,3,4),xlim=**c**(-9,15),main='X - Normal(3,16)')

**polygon**(cord.x,cord.y,col='lightgray')

**lines**(x=**c**(3,3),y=**c**(0,**dnorm**(3,3,4)))

regiao=**seq**(-0.25,0.5,0.01)

cord.x <- **c**(**min**(regiao),regiao,**max**(regiao))

cord.y <- **c**(0,**dnorm**(regiao),0)

**curve**(**dnorm**(x,0,1),xlim=**c**(-3,3),main='Z - Normal Padrão',xlab="z")

**polygon**(cord.x,cord.y,col='lightgray')

**lines**(x=**c**(0,0),y=**c**(0,fdnorm[x==0]))

regiao=**seq**(2,5,0.01)

cord.x <- **c**(**min**(regiao),regiao,**max**(regiao))

cord.y <- **c**(0,**dnorm**(regiao,3,4),0)

**curve**(**dnorm**(x,3,4),xlim=**c**(-9,15),ylim=**c**(0,0.4),xlab="",ylab="",xaxs="i",yaxs="i",col="lightblue3",lwd=2)

**polygon**(cord.x,cord.y,col='lightblue3')

**lines**(x=**c**(3,3),y=**c**(0,**dnorm**(3,3,4)),col="darkgray",lty=2)

**par**(new=TRUE)

regiao=**seq**(-0.25,0.5,0.01)

cord.x <- **c**(**min**(regiao),regiao,**max**(regiao))

cord.y <- **c**(0,**dnorm**(regiao),0)

**curve**(**dnorm**(x,0,1),xlim=**c**(-9,15),ylim=**c**(0,0.4),xlab="",ylab="",xaxs="i",yaxs="i",col="orange2",lwd=2)

**polygon**(cord.x,cord.y,col='orange2')

**lines**(x=**c**(0,0),y=**c**(0,fdnorm[x==0]),col="gray",lty=2)

**legend**(5,0.35,**c**("N(3,16)","N(0,1)"),lwd=**c**(2,2), col=**c**("lightblue3","orange2"),pch=**c**(15,15),bty="n")

Exemplo 7.9

1. P(X≤10000)P(X≤10000)

**pnorm**(10000,mean=10000,sd=1500)

1. P(X≥10000)P(X≥10000)

1-**pnorm**(10000,mean=10000,sd=1500)

1. P(12000<X<15000)P(12000<X<15000)

**pnorm**(15000,mean=10000,sd=1500)-**pnorm**(12000,mean=10000,sd=1500)

1. P(X>20000)P(X>20000)

1-**pnorm**(20000,mean=10000,sd=1500)

Exemplo 7.10

**par**(mfrow=**c**(1,2))

**curve**(**dexp**(x,1),xlim=**c**(0,5),main="f.d.p para X~exponencial(1)",xlab="(a)")

**curve**(**pexp**(x,1),xlim=**c**(0,5),main="f.d.a para X~exponencial(1)",xlab="(b)")

Para o cálculo da probabilidade P(T>500)P(T>500), o comando é:

1-**pexp**(500,rate=1/500)

7.5 Aproximação Normal à Binomial

1-**pbinom**(6,size=10,prob = 1/2)

1-**pnorm**(6.5,mean=5, sd=**sqrt**(2.5))

Figura 7.20

ymax=0.3

**barplot**(height = **dbinom**(0:10,size=10,prob = 1/2),col = "white",ylim=**c**(0,ymax))

**par**(new=T)

**barplot**(height = **c**(**rep**(0,7),**dbinom**(7:10,size=10,prob = 1/2)),ylim=**c**(0,ymax),

border=**c**(**rep**(NA,7),**rep**("black",4)), col = **c**(**rep**(NA,7),**rep**("gray",4)))

**par**(new=T)

**curve**(**dnorm**(x,mean=5, sd=**sqrt**(2.5)),xlim=**c**(-0.8,10.8),ylim=**c**(0.0,ymax),xaxs="i",yaxs="i",ylab="")

P(3<Y≤6)=P(3<Y≤6)=

**pbinom**(6,size=10,prob=1/2)-**pbinom**(3,size=10,prob=1/2)

P(3.5≤X≤6.5)=P(3.5≤X≤6.5)=

**pnorm**(6.5,mean=5,sd=**sqrt**(2.5))-**pnorm**(3.5,mean=5,sd=**sqrt**(2.5))

Figura 7.21

ymax=0.3

**barplot**(height = **dbinom**(0:10,size=10,prob = 1/2),col = "white",ylim=**c**(0,ymax))

**par**(new=T)

**barplot**(height = **c**(**rep**(0,4),**dbinom**(4:6,size=10,prob = 1/2),**rep**(0,4)),ylim=**c**(0,ymax),

border=**c**(**rep**(NA,4),**rep**("black",3),**rep**(NA,4)), col = **c**(**rep**(NA,4),**rep**("gray",3),**rep**(NA,4)))

**par**(new=T)

**curve**(**dnorm**(x,mean=5, sd=**sqrt**(2.5)),xlim=**c**(-0.8,10.8),ylim=**c**(0.0,ymax),xaxs="i",yaxs="i",ylab="")

Figura 7.25

**curve**(**dgamma**(x,shape = 3, scale = 1),xlim=**c**(0,10))

Figura 7.26

**par**(mfrow=**c**(1,3))

**curve**(**dchisq**(x,df=1),xlim=**c**(0,10),xlab="(a) df=1")

**curve**(**dchisq**(x,df=2),xlim=**c**(0,10),xlab="(b) df=2")

**curve**(**dchisq**(x,df=3),xlim=**c**(0,10),xlab="(c) df=3")

Exemplo 7.13

Y∼χ2(10)Y∼χ2(10), então:  
P(Y>y0)=0.99⇒y0=2.558P(Y>y0)=0.99⇒y0=2.558.

**qchisq**(1-.99,df=10)

regiao=**seq**(2.6,20,0.01)

cord.x <- **c**(**min**(regiao),regiao,**max**(regiao))

cord.y <- **c**(0,**dchisq**(regiao,10),0)

**curve**(**dchisq**(x,10),xlim=**c**(0,20),xlab="",ylim=**c**(0,0.15),ylab="",

col="lightblue3",lwd=2,xaxt="n")

**axis**(1,at=2.6,labels = "y\_0=2.558")

**polygon**(cord.x,cord.y,col='lightblue3')

Figura 7.27

**qchisq**(1-0.05,df=10)

regiao=**seq**(18.3,20,0.01)

cord.x <- **c**(**min**(regiao),regiao,**max**(regiao))

cord.y <- **c**(0,**dchisq**(regiao,10),0)

**curve**(**dchisq**(x,10),xlim=**c**(0,20),ylim=**c**(0,0.15),xlab="",ylab="",

col="lightblue3",lwd=2,xaxt="n")

**axis**(1,at=18.3,labels = "y\_0=18.307")

**polygon**(cord.x,cord.y,col='lightblue3')

**print**(**paste**("Aproximação Normal: ",**round**(**qnorm**(1-0.099),3)))

**print**(**paste**("Valor /qui-quadrado: ",**round**(**qchisq**(1-0.1,df=30),3)))

Figura 7.28

**curve**(**dnorm**(x),ylim=**c**(0,0.5),xlim=**c**(-3,3),xlab="",col="darkred",ylab="")

**par**(new=TRUE)

**curve**(**dt**(x,df=1),ylim=**c**(0,0.5),xlim=**c**(-3,3),xlab="",col="blue",lty=2,ylab="")

**legend**(-3,0.5,lty=**c**(1,2), col=**c**("darkred","blue"),legend=**c**("normal padrão", "t-student"),bty="n",lwd=**c**(2,2),cex=0.75)

Exemplo 7.15

P(−1.943<t(6)<1.943)=P(−1.943<t(6)<1.943)=

**pt**(1.943,df = 6)-**pt**(-1.943,df = 6)

P(t(6)>2.447)=P(t(6)>2.447)=

1-**pt**(2.447,df = 6)

Exemplo 7.16

F∼F(5,7)F∼F(5,7), então P(F>3.97)=P(F>3.97)=

P\_F=**pf**(3.97,df1 = 5, df2 = 7)

f=**qf**(0.05,df1 = 5, df2 = 7)

Exemplo 7.17

**print**(**paste**("Q(0.5)=",**round**(**qnorm**(0.5),3)))

**print**(**paste**("Q(0.25)=",**round**(**qnorm**(0.25),3)))

**print**(**paste**("Q(0.3)=",**round**(**qnorm**(0.3),3)))

**print**(**paste**("Q(0.75)=",**round**(**qnorm**(0.75),3)))

Exemplo 7.18

**print**(**paste**("Q(0.5)=",**round**(**qexp**(0.5,rate = 1/2),3)))

Exemplo 7.19

**pnorm**(8.65,mean=10,sd=25)

**qnorm**(0.8269,mean=10,sd=25)

**pexp**(0.85,rate=2)

**qexp**(0.345,rate=2)

Exemplo 7.21

**curve**(**pnorm**(q = x, mean = 0, sd=1), xlim=**c**(-4,4), col="blue",lwd=2,ylab="",xlab="z")

**lines**(x=**c**(0,0), y=**c**(0,fdanorm[x==0]), lty=2, col="gray")

Capítulo 8: Variáveis Aleatórias Multidimensionais

8.1 Introdução

tab<-**tabular**( estado\_civil ~ grau\_instrucao\*idade\_anos\*(mean) )

**pander**(tab)

tab<-**tabular**( estado\_civil\*grau\_instrucao ~ idade\_anos\*(mean+sd) )

**pander**(tab)

**tabular**( (estado\_civil + 1)\*(grau\_instrucao+1) ~ idade\_anos\*(mean+sd) )

tab<-**tabular**( (estado\_civil + 1)\*(grau\_instrucao+1) ~ idade\_anos\*(mean+sd))

**pander**(tab)

**tabular**( (**factor**(estado\_civil) + 1)\*(grau\_instrucao+1) ~ idade\_anos\*(mean+sd) )

Gráficos

Gráfico de dispersão com marginais:

**scatter.marginal**(tab2\_1$idade\_anos, tab2\_1$salario, ylab="salario", xlab="idade")

**scatter.marginal**(tab2\_1$idade\_anos, tab2\_1$salario, ylab="salario", xlab="idade",type="boxplot")

**scatter.marginal**(tab2\_1$idade\_anos, tab2\_1$salario, ylab="salario", xlab="idade",type="dispersion")

dispersao = **ggplot**(tab2\_1, **aes**(idade\_anos,salario)) + **geom\_point**() *#Dispersao*

**ggMarginal**(dispersao, type="histogram") *# Marginais com histograma*

dispersao = ggplot2::**ggplot**(tab2\_1, ggplot2::**aes**(idade\_anos, salario)) + ggplot2::**geom\_point**() *#Dispersao*

**ggMarginal**(dispersao,type = "boxplot") *# Marginais com boxplot*

dispersao = ggplot2::**ggplot**(tab2\_1, ggplot2::**aes**(idade\_anos, salario)) + ggplot2::**geom\_point**() *#Dispersao*

scatter <- **ggplot**(tab2\_1, **aes**(idade\_anos,salario)) + **geom\_point**() + *#dispersão*

**scale\_x\_continuous**(limits=**c**(**min**(idade\_anos),**max**(idade\_anos))) + *# marginal X*

**scale\_y\_continuous**(limits=**c**(**min**(salario),**max**(salario))) + *# marginal Y*

**geom\_rug**(size=0.1) + **theme\_set**(**theme\_minimal**(base\_size = 18))

scatter

**ggMarginal**(dispersao,type = "density") *# Marginais com função densidade empírica*

Gráficos de perfis - Gráficos de dispersão Cruzados

**pairs**( ~ n\_filhos + salario + idade\_anos, pch=16, col="blue")

**pairs**( ~ n\_filhos + salario + idade\_anos, pch=16, col="blue",upper.panel = NULL)

Gráficos de curva de nível e mapas de calor

**contour**(x=1:**nrow**(volcano), y=1:**ncol**(volcano), volcano, col="darkred",xlab = "Sul - Norte", ylab = "Leste - Oeste")

**title**("Mapa topográfico: Maunga Whau", font = 4) *# Titulo do Gráfico*

**Figura 8.A**: Gráfico de curvas de nível do vulcão Maunga Whau em função da latitude e longitude.

**library**(colorRamps) *# Paleta de cores que contem a sequencia de cores `blue2red`.*

**filled.contour**(x=1:**nrow**(volcano), y=1:**ncol**(volcano), z=volcano, color = blue2red,

plot.title = **title**(main = "Mapa topográfico: Maunga Whau \n ",

xlab = "Sul - Norte", ylab = "Leste - Oeste"),

plot.axes = { **axis**(1, **seq**(100, 800, by = 100))

**axis**(2, **seq**(100, 600, by = 100)) },

key.axes = **axis**(4, **seq**(90, 190, by = 10)))*# outra sugestão de cor: `color = terrain.colors` ou não declarar esta opção para visualizar a cor padrão.*

8.4 Covariância amostral

**cov**(salario,idade\_anos)

**cov**(idade\_anos,salario)

8.8 Distribuição Normal Bidimensional

Sigma=**cbind**(**c**(1,0.6),**c**(0.6,1)) *# matriz de variancias e covariancias*

mu=**c**(0,0) *# Vetor de médias*

Z.sim<-**rmvnorm**(n = 1000, mean=mu, sigma=Sigma) *# simula 100 observações de uma normal bivariada*

fdamvnorm<-function(upper,lower=**c**(-Inf,-Inf),mu=**rep**(0,times=**length**(upper)),Sigma=**diag**(**length**(upper))){

**return**(**pmvnorm**(lower=**c**(-Inf,-Inf),upper=upper,mean=mu,sigma=Sigma)[[1]]) *# Calcula a integral dupla da fdp da normal multivariada entre no retangulo `lower X upper`.*

}

**fdamvnorm**(upper=**c**(0,0)) *# FDA considerando o terceiro quadrante*

**fdamvnorm**(upper=**c**(Inf,0)) *# FDA considerando o terceiro e quarto quadrantes*

**fdamvnorm**(**c**(0,Inf)) *# FDA considerando o segundo e terceiro quadrantes*

**fdamvnorm**(lower=**c**(-1.96,-1.96),upper=**c**(1.96,1.96)) *#FDA considerando o quadrado (-1.96,-1.96)X(1.96,1.96)*

Sigma=**cbind**(**c**(1,0.6),**c**(0.6,1)) *# matriz de variancias e covariancias*

mu=**c**(0,0) *# Vetor de médias*

**dmvnorm**(x=**c**(0,0), mean=mu, sigma=Sigma)

Figura 8.15

xy<-**seq**(-4,4,0.1)

mu=**c**(0,0)

Sigma=**diag**(2)

Z1<-**matrix**(NA,**length**(xy),**length**(xy))

for(i in 1:**length**(xy)){

for(j in 1:**length**(xy)){

Z1[i,j]<-**dmvnorm**(x=**c**(xy[i],xy[j]), mean=mu, sigma=Sigma)

}

}

Figura 8.15(a)

**persp**(Z1, phi = 30, theta = -45,col=**heat.colors**(**ncol**(Z1)\***nrow**(Z1),.8), ticktype = "detailed",xlab="X")

*# `phi` e `theta` são os parâmetros de deslocamento da perspectiva da figura.*

*# heat.colors(.) modifica a paleta de cores do gráfico*

*# box=FALSE elimina a caixa que envolve o gráfico*

*# scale =FALSE define que as variáveis não devem ser padronizadas*

*# ticktype: tipo de eixos. Pode assumir valores "simple" ou "detailed"*

Figura 8.15(b)

xy<-**seq**(-4,4,0.1)

mu=**c**(0,0)

Sigma=**cbind**(**c**(1,0.6),**c**(0.6,1))

Z2<-**matrix**(NA,**length**(xy),**length**(xy))

for(i in 1:**length**(xy)){

for(j in 1:**length**(xy)){

Z2[i,j]<-**dmvnorm**(x=**c**(xy[i],xy[j]), mean=mu, sigma=Sigma)

}

}

**persp**(Z2, phi = 30, theta = -45,col=**heat.colors**(**ncol**(Z2)\***nrow**(Z2),.8), ticktype = "detailed", xlab="X")

Figura 8.17

**par**(mfrow=**c**(1,2))

**contour**(x=1:**nrow**(Z2), y=1:**ncol**(Z2), Z2, col="darkred",xlab = "X", ylab = "Y")

**contour**(x=1:**nrow**(Z1), y=1:**ncol**(Z1), Z1, col="darkred",xlab = "X", ylab = "Y")

Capítulo 9: Noções de Simulação

9.1 Introdução

Exemplo 9.4

u<-**runif**(10,0,1) *# gera 10 observações aleatórias de uma v.a. uniforme[0,1]*

u

u <- **runif**(1000,0,1) *# gera 1000 observações aleatórias de uma v.a. uniforme[0,1]*

x <- **qnorm**(u,mean=0, sd = 1) *# Calcula os quantis para o vetor u simulado da uniforme.*

Figura 9.A

**par**(mfrow=**c**(1,2))

**hist**(u, freq=FALSE, main="Histograma da amostra da \n distribuição Uniforme simulada", col="lightblue4",border="gray")

**hist**(x, freq=FALSE, main="Histograma da variável X simulada a partir \n do resultado do Teorema 9.1", col="lightblue4",border="gray")

Exemplo 9.6

Figura 9.B

u <- **runif**(1000,0,1) *# gera 1000 observações aleatórias de uma v.a. uniforme[0,1]*

x <- **sqrt**(u) *# Calcula a f.d.a. inversa para o vetor u gerado.*

**par**(mfrow=**c**(1,2))

**hist**(u, freq=FALSE, main="Histograma da amostra da \n distribuição Uniforme simulada", col="lightblue4",border="gray")

**hist**(x, freq=FALSE, main="Histograma da variável x simulada a partir \n do resultado do Teorema 9.1", col="lightblue4",border="gray")

Figura 9.4

**par**(mfrow=**c**(1,2))

x<-**seq**(0,1.75,0.01) *# Produzindo o vetor da abscissa*

f\_x<-**c**(2\*x[x<=1],0\*x[x>1]) *# Calculando f(x)*

**plot**(x,f\_x, col='darkred', ylab="f(x)",xaxt="n",type="l") *#Grafico de f(x)*

**lines**(x=**c**(0,1),y=**c**(2,2), col="gray",lty=2)

F\_x<-**c**((x[x<=1])^2,**rep**(1,**length**(x))[x>1]) *# Calculando F(x)*

**plot**(x,F\_x, col='darkblue', ylab="F(x)",xaxt="n",type="l") *#Grafico de F(x)*

**lines**(**c**(0,1),**c**(1,1), col="gray",lty=2) *# Adicionando linas explicativas tracejadas*

**lines**(**c**(1,1),**c**(0,1), col="gray",lty=2)

**lines**(**c**(0,.71),**c**(0.51,.51), col="gray",lty=2)

**lines**(**c**(.71,.71),**c**(0,.51), col="gray",lty=2)

**text**(0,.51,"u=0.5", pos=4) *# inserindo texto no grafico*

**text**(0.71,0,"x=0.71")

Exemplo 9.8

**rbernoulli**(p=0.52)

u<-**rbernoulli**(n=10,p=0.52)

u

aa<-**rbernoulli**(n=1000, p=0.52)

**sum**(aa)/1000

Exemplo 9.9

xbinomial<-**numeric**()

for( i in 1:20){

xbinomial[i]<-**sum**(**rbernoulli**(n=10,p=0.52))

}

xbinomial

**rbinom**(n=20, size = 10, p=0.52)

Exemplo 9.10

u<-**runif**(n=5,0,1)

**qexp**(u,rate=1/2)

**rexp**(n=5,rate=1/2)

Exemplo 9.11

**qnorm**(0.230,mean=10, sd = .4) *# Calcula os quantis para a normal com média 10 e sd = 0.4*

Exemplo 9.13

x<-**rbinom**(20,10,0.5)

x

y<-**rpois**(20,1.7)

y

Exemplo 9.14

x<-**rbinom**(20,10,0.5)

y<-**rpois**(20,1.7)

z<-**runif**(100,0,1)

b<-**rbinom**(15,1,0.7)

Figura 9.8

**par**(mfrow=**c**(2,2))

**hist**(x, col="lightblue4",border="white")

**hist**(y, col="lightblue4",border="white")

**hist**(z, col="lightblue4",border="white")

**hist**(b, col="lightblue4",border="white")

Exemplo 9.16

z<-**rnorm**(500,0,1)

y<-**rnorm**(200,10,3)

t<-**rt**(500,35)

Exp<-**rexp**(500,2)

w<-**rchisq**(300,5)

f<-**rf**(500,10,12)

Figura 9.16

``{r, eval=FALSE} par(mfrow=c(3,2)) hist(z, col=“lightblue4”,border=“white”) hist(y, col=“lightblue4”,border=“white”) hist(t, col=“lightblue4”,border=“white”) hist(Exp, col=“lightblue4”,border=“white”) hist(w, col=“lightblue4”,border=“white”) hist(f, col=“lightblue4”,border=“white”) ```

Exemplo 9.18

u1<-**runif**(200,0,1)

u2<-**runif**(200,0,1)

Figura 9.1

**plot**(u1,u2, pch=20, col="darkblue")

Exemplo 9.19

u1<-**runif**(10000,0,1)

u2<-**runif**(10000,0,1)

P<-**data.frame**(u1,u2)

x<-**qnorm**(u1)

y<-**qnorm**(u2)

Figura 9.C

**scatter.marginal**(x,y,breaks=30) *# Função criada no Capitulo 8 para construir este tipo de gráfico*

Figura 9.11

**library**(mvtnorm)

**library**(scatterplot3d)

**scatterplot3d**(x,y, **dmvnorm**(x=**cbind**(x,y)), highlight.3d=TRUE, pch=16, zlab="Z")

Figura 9.11

**contour**(**interp**(x,y, **dmvnorm**(x=**cbind**(x,y))))

Capítulo 10: Introdução à Inferência Estatística

10.5 Amostra Aleatória Simples

##### Amostragem sem reposição: Utilize a opção 'replace = FALSE'

AASs = **sample**(x = 1:**dim**(tab2\_1)[1], size=15, replace = FALSE )

*# Elementos selecionados na amostra*

AASs

##### Amostragem com reposição: Utilize a opção 'replace = TRUE'

AASc = **sample**(x = 1:**dim**(tab2\_1)[1], size=15, replace = TRUE )

*# Elementos selecionados na amostra*

AASc

**print**(**kable**(tab2\_1[AASs,], caption = '\*\*Tabela 10.A\*\*: Amostra Aleatória Simples sem Reposição da Tabela 2.1'))

**print**(**kable**(tab2\_1[AASc,], caption = '\*\*Tabela 10.B\*\*: Amostra Aleatória Simples com Reposição da Tabela 2.1'))

Exemplo 10.8

x<-**rnorm**(5,mean=167, sd= 5)

x

10.7 Distribuições Amostrais

x\_normal<-**rnorm**(10000,mean=167, sd= 5)

**hist**(x\_normal, col="lightblue4", border="white",freq = FALSE)

*#Inicializando as variaveis como vetores numericos*

## Media e variancia para amostras de tam 15

xbar15<-**numeric**()

var\_amostral15<-**numeric**()

## Media e variancia para amostras de tam 30

xbar30<-**numeric**()

var\_amostral30<-**numeric**()

## Media e variancia para amostras de tam 100

xbar100<-**numeric**()

var\_amostral100<-**numeric**()

for ( i in 1:200){

*# Extraindo amostras de tamanho 15 e calculando a média e variancia*

smp<-**sample**(x\_normal,size = 15)

xbar15[i]<-**mean**(x\_normal[smp])

var\_amostral15[i]<-**var**(x\_normal[smp])

*# Extraindo amostras de tamanho 30 e calculando a média e variancia*

smp<-**sample**(x\_normal,size = 30)

xbar30[i]<-**mean**(x\_normal[smp])

var\_amostral30[i]<-**var**(x\_normal[smp])

*# Extraindo amostras de tamanho 100 e calculando a média e variancia*

smp<-**sample**(x\_normal,size = 100)

xbar100[i]<-**mean**(x\_normal[smp])

var\_amostral100[i]<-**var**(x\_normal[smp])

}

Figura 10.A

**par**(mfrow=**c**(2,3))

**hist**(xbar15, col="lightblue4", border="white",freq = FALSE, breaks = 15, main="Média com n=15")

**hist**(xbar30, col="lightblue4", border="white",freq = FALSE, breaks = 15, main="Média com n=30")

**hist**(xbar100, col="lightblue4", border="white",freq = FALSE, breaks = 15, main="Média com n=100")

**hist**(var\_amostral15, col="lightblue4", border="white",freq = FALSE, breaks = 15, main="Variância com n=15")

**hist**(var\_amostral30, col="lightblue4", border="white",freq = FALSE, breaks = 15, main="Variância com n=30")

**hist**(var\_amostral100, col="lightblue4", border="white",freq = FALSE, breaks = 15, main="Variância com n=100")

Figura 10.B

**plot** (**density**(x\_normal), ylim=**c**(0,1), xlab="", main = "Funções densidades empíricas")

**lines** (**density**(xbar15),col="red")

**lines** (**density**(xbar30),col="blue")

**lines** (**density**(xbar100),col="orange")

**legend**("topright", **c**("X~N(167,25)", "Média com n=15", "Média com n=30", "Média com n=100"),

col=**c**("black", "red", "blue", "orange"), lty=**c**(1,1,1,1))

Exemplo 10.8

## Mediana para amostras de tam 5

md5<-**numeric**()

for ( i in 1:200){

*# Extraindo amostras de tamanho 5 e calculando a mediana*

smp<-**sample**(x\_normal,size = 5)

md5[i]<-**median**(x\_normal[smp])

}

**print**(**summary**(md5))

**cat**("E(md)=",**mean**(md5), ", Var(md)=",**var**(md5),", dp(md)=",**sd**(md5),", min(md)=",**min**(md5),", max(md)=",**max**(md5))

Figura 10.3

**hist**(md5, col="lightblue3", border="white",freq = FALSE, main="Mediana com n=5")

*# breaks=seq(min(md5)-1,max(md5)+1,2),*

Figura 10.5

n=**cbind**(**c**(2, 5,25), *# numero de amostras para os gráficos da primeira linha*

**c**(2,10,30), *# numero de amostras para os gráficos da segunda linha*

**c**(3, 5,10), *# numero de amostras para os gráficos da terceira linha*

**c**(5,15,25)) *# numero de amostras para os gráficos da quarta linha*

X<-**list**(

x1=**matrix**(NA,nrow=200,ncol=3), *# Matriz que armazenara as médias da primeira linha*

x2=**matrix**(NA,nrow=200,ncol=3), *# Matriz que armazenara as médias da segunda linha*

x3=**matrix**(NA,nrow=200,ncol=3), *# Matriz que armazenara as médias da terceira linha*

x4=**matrix**(NA,nrow=200,ncol=3)) *# Matriz que armazenara as médias da quarta linha*

for ( i in 1:3){ *# iterando ao longo do número de amostras de cada linha*

for ( j in 1: 200){ *# Construindo 200 observacoes para cada gráfico*

X$x1[j,i] = **mean**( **runif**( n=n[i,1])) *# Calculando a média pra cada uma das 200 simulacoes de uma distribuicao uniforme*

X$x2[j,i] = **mean**( **rexp**( n=n[i,2], rate=2)) *# Calculando a média pra cada uma das 200 simulacoes de uma distribuicao exponencial*

X$x3[j,i] = **mean**(**rchisq**( n=n[i,3], df=1)) *# Calculando a média pra cada uma das 200 simulacoes de uma distribuicao qui-quadrado*

X$x4[j,i] = **mean**( **rbeta**( n=n[i,4], shape1 = .2, shape2 = .2)) *# Calculando a média pra cada uma das 200 simulacoes de uma distribuicao beta*

}

}

**par**(mfrow=**c**(4,4))

**plot**(**dunif**(x = **seq**(0,1,0.05)),type="l", ylab="f(x)", main="fdp-Uniforme(0,1)", xlab="X")

**hist**(X$x1[,1], col = "lightblue3", border = "white", main = **paste**("n = ", n[1,1]), freq = FALSE, xlab="X\_bar")

**hist**(X$x1[,2], col = "lightblue3", border = "white", main = **paste**("n = ", n[2,1]), freq = FALSE, xlab="X\_bar")

**hist**(X$x1[,3], col = "lightblue3", border = "white", main = **paste**("n = ", n[3,1]), freq = FALSE, xlab="X\_bar")

**plot**(**dexp**(x = **seq**(0,1,0.05),rate=2),type="l", ylab="f(x)", main="fdp-Exponencial(1/2)", xlab="X")

**hist**(X$x2[,1], col = "lightblue3", border = "white", main = **paste**("n = ", n[1,2]), freq = FALSE, xlab="X\_bar")

**hist**(X$x2[,2], col = "lightblue3", border = "white", main = **paste**("n = ", n[2,2]), freq = FALSE, xlab="X\_bar")

**hist**(X$x2[,3], col = "lightblue3", border = "white", main = **paste**("n = ", n[3,2]), freq = FALSE, xlab="X\_bar")

**plot**(**dchisq**(x = **seq**(0,1,0.05),df=1),type="l", ylab="f(x)", main="fdp-QuiQuad(1)", xlab="X")

**hist**(X$x3[,1], col = "lightblue3", border = "white", main = **paste**("n = ", n[1,3]), freq = FALSE, xlab="X\_bar")

**hist**(X$x3[,2], col = "lightblue3", border = "white", main = **paste**("n = ", n[2,3]), freq = FALSE, xlab="X\_bar")

**hist**(X$x3[,3], col = "lightblue3", border = "white", main = **paste**("n = ", n[3,3]), freq = FALSE, xlab="X\_bar")

**plot**(**dbeta**(x = **seq**(0,1,0.05), shape1 = .2, shape2 = .2),type="l", ylab="f(x)", main="fdp-Beta(0.2, 0.2)", xlab="X")

**hist**(X$x4[,1], col = "lightblue3", border = "white", main = **paste**("n = ", n[1,4]), freq = FALSE, xlab="X\_bar")

**hist**(X$x4[,2], col = "lightblue3", border = "white", main = **paste**("n = ", n[2,4]), freq = FALSE, xlab="X\_bar")

**hist**(X$x4[,3], col = "lightblue3", border = "white", main = **paste**("n = ", n[3,4]), freq = FALSE, xlab="X\_bar")

Exemplo 10.11:

x\_barr<-**sapply**(1:200, function(i){**return**(**mean**(**rnorm**(n=100,mean=500,sd=10)))})

**hist**(x\_barr, col = "lightblue3", border = "white", freq=FALSE)

**sum**( x\_barr >= 498 & x\_barr <= 502 )/200

Exemplo 10.12

x\_barr <- **numeric**()

x\_var <- **numeric**()

for (i in 1:200){

smp <- **rbinom**(n = 10, size = 1, prob = 0.3)

x\_barr[i] <- **mean**(smp)

x\_var[i] <- **var**(smp)

}

p\_hat = **sum**( x\_barr ) / 200

p\_hat

**hist**(x\_barr, col="lightblue3", border="white", breaks=**seq**(0,1,0.02))

x\_barr <- **numeric**()

x\_var <- **numeric**()

for (i in 1:200){

smp <- **rbinom**(n = 100, size = 1, prob = 0.3)

x\_barr[i] <- **mean**(smp)

x\_var[i] <- **var**(smp)

}

p\_hat = **sum**( x\_barr ) / 200

p\_hat

**hist**(x\_barr, col="lightblue3", border="white", breaks=**seq**(0,1,0.02))

Exemplo 10.13

x<-**c**(1,3,5,5,7)

smp<-**matrix**(NA,nrow=200,ncol=2)

for ( i in 1:200){

smp[i,] <- **sample**(x, size=2, replace=TRUE)

}

**hist**(smp, col="lightblue3", border="white")

x<-**c**(1,3,5,5,7)

S2<-**numeric**()

md<-**numeric**()

var\_hat<-**numeric**()

for ( i in 1:200){

smp<-**sample**(x, size=3, replace=TRUE)

S2[i] <- **var**(smp)

var\_hat[i] <- **var**(smp)\*2/3

md[i] <- **median**(smp)

}

Figuras 10.6 a 10.8

**hist**(S2, col="lightblue3", border="white", main="Figura 10.6")

**hist**(md, col="lightblue3", border="white", main="Figura 10.7")

**hist**(var\_hat, col="lightblue3", border="white", main="Figura 10.8")

Exemplo 10.16

x<-**c**(1,2,3,4,5,6,7,8,9,10,11,12,13,14,15)

**sample**(x, 7)

**sample**(x, 7, replace=T)

Capítulo 11: Estimação

Exemplo 11.9

tab11.1<- **data.frame**(

x=**c**(1.2,1.5,1.7,2,2.6),

y=**c**(3.9,4.7,5.6,5.8,7.0))

**mean**(tab11.1$x)

**mean**(tab11.1$y)

**plot**(tab11.1$x,tab11.1$y,pch=16,col= "darkblue")

tab11.1$x\_3=3\*tab11.1$x

tab11.1$y\_3x=tab11.1$y - tab11.1$x\_3

tab11.1$y\_3x\_2=tab11.1$y\_3x^2

**print**(**kable**(tab11.1,col.names = **c**("$X$","$Y$", "$3X$","$Y-3X$","$(Y-3X)^2$"), caption = "\*\*Tabela 11.1\*\*: Análise do modelo $Ŷ=3X$"))

**print**(**paste**("Total da soma de quadrados dos resíduos(sum(tab11.1$y\_3x\_2)):",**sum**(tab11.1$y\_3x\_2)))

theta\_hat=**sum**(tab11.1$x\*tab11.1$y)/**sum**(tab11.1$x^2)

Exemplo 11.11

x\_sim <- **rexp**(1000, rate = 2) *# 1000 observações amostradas de um modelo exponencial de média 1/2*

**mean**(x\_sim)

**hist**(x\_sim, col="lightblue3", border="white", freq = FALSE)

MV <- function(rate) {

R = **dnorm**(x\_sim, rate)

**return**(-**sum**(**log**(R))) *# devolvendo '-log(verossimilhanca)'*

}

emv\_rate\_x <- **mle**(MV, *# MV é a função que se deseja otimizar*

start = **list**(rate=1)) *# start é uma lista com 'chutes' iniciais para o processo de otimização.*

emv\_rate\_x

Exemplo 11.12

Figura 11.4

samples=20 *# Numero de amostras*

N=25 *# Numero de observações por amostra*

x\_sim<-**matrix**(NA,N,samples) *# x\_sim armazenara as amostras simuladas*

for( i in 1:samples){

x\_sim[,i]<-**rnorm**(n = N, mean= 5, sd=3) *# Simulando 20 N(5,3^2)*

}

cont=0

for( i in 1:samples){

llim=**mean**(x\_sim[,i])-1.96\***sd**(x\_sim[,i])/**sqrt**(N) *# limite inferior do IC da i-ésima amostra*

ulim=**mean**(x\_sim[,i])+1.96\***sd**(x\_sim[,i])/**sqrt**(N) *# limite inferior do IC da i-ésima amostra*

if (llim>5 || ulim<5){

line\_col="red"

cont=cont+1

}

else line\_col="black"

if(i!=1)

**par**(new=T)

**plot**(x=i,y=**mean**(x\_sim[,i]), xlim=**c**(0,21), ylim=**c**(0,10), pch=20,cex=1, col=line\_col, ylab="Médias", xlab="Amostras") *# plotando as médias de cada amostra*

**lines**(x = **c**(i,i), y =**c**(llim,ulim), lwd=1, col=line\_col) *# Adicionando as linhas dos ICs de cada amostra*

**abline**(h=5,lty = 'dashed', col="blue") *# Valor real da média*

}

Exemplo 11.13

**IC**(n=25, media = 485, desvpad=**sqrt**(100), gama=0.95)

Exemplo 11.15

**IC**(n=400, media = 0.6, desvpad=**sqrt**(1/4), gama=0.95)

Exemplo 11.16

p=80/400

dp=**sqrt**(p\*(1-p))

**IC**(n=400, media = p, desvpad=dp, gama=0.9)

*# IC Conservador*

p=1/2

dp=**sqrt**(p\*(1-p))

**IC**(n=400, media = 80/400, desvpad=dp, gama=0.9)

11.9 Exemplos Computacionais

Bootstrap

amostra\_original<-**rexp**(50, rate=3) *# Vamos simular uma exponencial de média 1/3*

n=**length**(amostra\_original)

md=**numeric**() *# vetor que armazenará as medianas calculadas ao longo das amostras bootstrap*

b=30 *# Número de amostras bootstrap a serem resorteadas baseadas na amostra original.*

for( i in 1:b){ *# Vamos calcular 30 valores da mediana para construir sua dist. empirica*

boot<-**sample**(amostra\_original, size = n, replace = TRUE) *# Tomando amostras de tamanho n, com reposicao*

md[i]<-**median**(boot)

}

**hist**(md, col="lightblue3", border= "white")

ep\_md=**numeric**() *# vetor que armazenará os erros padrões da mediana para cada rodada do bootstrap*

md=**numeric**() *# vetor que armazenará as medianas calculadas ao longo das amostras bootstrap*

b=**c**(30,50,100,500,1000) *# Número de amostras bootstrap a serem resorteadas baseadas na amostra original.*

for (j in 1:5){ *# neste for, percorreremos os valores de b*

for( i in 1:b[j]){ *# Vamos calcular 30 valores da mediana para construir sua dist. empirica*

boot<-**sample**(amostra\_original, size = n, replace = TRUE) *# Tomando amostras de tamanho n, com reposicao*

md[i]<-**median**(boot)

}

ep\_md[j]=**sd**(md)/b[j]

}

**plot**(1:5, ep\_md, cex=2, type="b", ylab="Erro Padrão", xlab="b: tamanho do bootstrap")

Exemplo 11.19

amostra\_original<-**c**(2,5,3,4,6)

n=**length**(amostra\_original)

b=5

boot<-**matrix**(NA,n,b) *# armazena as amostras geradas*

md=**numeric**()

xbarr=**numeric**()

for( i in 1:b){

boot[,i]<-**sample**(amostra\_original, size = n, replace = TRUE)

md[i]<-**median**(boot[,i])

xbarr[i]<-**mean**(boot[,i])

}

Tabela 11.2

tab11.2<-**data.frame**(x1\_boot=boot[,1],

x2\_boot=boot[,2],

x3\_boot=boot[,3],

x4\_boot=boot[,4],

x5\_boot=boot[,5],

md=md,

media=xbarr)

**print**(**kable**(tab11.2,col.names = **c**("Bootstrap:x1\*","Bootstrap:x2\*","Bootstrap:x3\*","Bootstrap:x4\*","Bootstrap:x5\*", "md(x\*)","média(x\*)"),caption = "\*\*Tabela 11.2\*\*: Procedimento \*bootstrap\*."))

### Calculando agora o EP para a mediana a partir das amostras geradas:

ep=0

for( i in 1:b){

ep=ep+(md[i]-**mean**(md))^2/(b-1)

}

ep=**sqrt**(ep)

ep

Exemplo 11.20

### Calculando agora o EP para a mediana a partir das amostras geradas:

ep\_media=0

for( i in 1:b){

ep\_media=ep\_media+(xbarr[i]-**mean**(xbarr))^2/(b-1)

}

ep\_media=**sqrt**(ep\_media)

ep\_media

**IC**(n = 1,media = 5.2, desvpad = ep, gama = 0.95)

amostra\_original<-**c**(2,5,3,4,6)

n=**length**(amostra\_original)

b=50

boot<-**matrix**(NA,n,b) *# armazena as amostras geradas*

md=**numeric**()

xbarr=**numeric**()

for( i in 1:b){

boot[,i]<-**sample**(amostra\_original, size = n, replace = TRUE)

md[i]<-**median**(boot[,i])

xbarr[i]<-**mean**(boot[,i])

}

### Calculando o EP para a mediana a partir das 50 amostras geradas:

ep=0

for( i in 1:b){

ep=ep+(md[i]-**mean**(md))^2/(b-1)

}

ep=**sqrt**(ep)

ep

### Calculando o EP para a media a partir das50 amostras geradas:

ep\_media=0

for( i in 1:b){

ep\_media=ep\_media+(xbarr[i]-**mean**(xbarr))^2/(b-1)

}

ep\_media=**sqrt**(ep\_media)

ep\_media

Exemplos 11.19 e 11.20 utilizando o pacote bootstrap

amostra\_original<-**c**(2,5,3,4,6)

### Exemplo 11.19: Mediana

boot\_11\_19<-**bootstrap**(x=amostra\_original, nboot=5, theta=function(x){**median**(x)})

boot\_11\_19$thetastar

### Exemplo 11.19: Mediana

boot\_11\_20<-**bootstrap**(x=amostra\_original, nboot=5, theta=function(x){**mean**(x)})

boot\_11\_20$thetastar

*#Se desejarmos, como nos exemplos, a estimativa do erro padrão, faríamos:*

err\_pad<- function(x){**sqrt**(**var**(x)/**length**(x))}

**bootstrap**(x=amostra\_original, nboot=15, theta=function(x){**median**(x)}, func=err\_pad)

**bootstrap**(x=amostra\_original, nboot=15, theta=function(x){**mean**(x)}, func=err\_pad)

Capítulo 12: Testes de Hipóteses

Exemplo 12.1

**qnorm**(p = 0.05, mean=155, sd =4)

Figura 12.2

x<-**seq**(125,167,0.1) *# eixo X do gráfico*

regiao1=**seq**(148.42,167,0.01) *# Regiao a ser sombreada na figura*

cord.x <- **c**(**min**(regiao1),regiao1,**max**(regiao1)) *# coordenadas X da regiao*

cord.y <- **c**(0,**dnorm**(regiao1,mean=145,sd=5.76),0) *# coordenadas Y da regiao*

**curve**(**dnorm**(x,mean=145,sd=5.76),xlim=**c**(125,167),ylim=**c**(0,0.12),xlab="",ylab="",xaxs="i",yaxs="i",col="black",lwd=2, xaxt='n') *# Curva da distribuicao normal*

**polygon**(cord.x,cord.y,col='orange2') *# plotando a regiao*

**lines**(x=**c**(145,145),y=**c**(0,**dnorm**(145,145,5.76)),col="gray",lty=2) *# linha vertical da media*

**par**(new=TRUE) *# inserindo a segunda curva*

regiao2=**seq**(125,148.42,0.01) *# Regiao a ser sombreada na figura*

cord.x <- **c**(**min**(regiao2),regiao2,**max**(regiao2))

cord.y <- **c**(0,**dnorm**(regiao2,mean=155,sd=4),0)

**curve**(**dnorm**(x,mean=155,sd=4),xlim=**c**(125,167),ylim=**c**(0,0.12),xlab="",ylab="",xaxs="i",yaxs="i",col="black",lwd=2, xaxt='n')

**polygon**(cord.x,cord.y,col='lightblue3')

**lines**(x=**c**(155,155),y=**c**(0,**dnorm**(155,155,4)),col="darkgray",lty=2)

**axis**(side=1,at = **c**(145,148.42,155), labels = **c**("145","148.42","155")) *# 'labels' do eixo X*

**legend**("topleft", legend=**c**("RC=5%", "Poder=7.93%"), pch=**c**(15,15), col=**c**("lightblue3","orange2"), cex=2, bty = "n") *# legenda*

Figura 12.3

x<-**seq**(125,167,0.1)

regiao1=**seq**(125,148.42,0.01) *# Regiao a ser sombreada na figura*

cord.x <- **c**(**min**(regiao1),regiao1,**max**(regiao1))

cord.y <- **c**(0,**dnorm**(regiao1,mean=155,sd=4),0)

**curve**(**dnorm**(x,mean=155,sd=4),xlim=**c**(125,167),ylim=**c**(0,0.15),

xlab="",ylab="",xaxs="i",yaxs="i",col="black",lwd=2, xaxt='n')

**polygon**(cord.x,cord.y,col='lightblue3')

**lines**(x=**c**(155,155),y=**c**(0,**dnorm**(155,155,4)),col="darkgray",lty=2)

**axis**(side=1,at = **c**(148.42,155), labels = **c**("148.42","mu=155"))

**par**(new=TRUE)

*# Curva tracejada*

**curve**(**dnorm**(x,mean=148.42,sd=3),xlim=**c**(125,167),ylim=**c**(0,0.15),

xlab="",ylab="",xaxs="i",yaxs="i",col="gray",lty=2, lwd=2, xaxt='n')

**par**(new=TRUE)

*# Curva tracejada*

**curve**(**dnorm**(x,mean=145,sd=3),xlim=**c**(125,167),ylim=**c**(0,0.15),

xlab="",ylab="",xaxs="i",yaxs="i",col="gray",lty=2, lwd=2, xaxt='n')

Figura 12.4

x<-**seq**(130,190,0.1)

regiao2=**seq**(130,147.16,0.01)

regiao1=**seq**(162.84,190,0.01)

cord.x1 <- **c**(**min**(regiao1),regiao1,**max**(regiao1))

cord.x2 <- **c**(**min**(regiao2),regiao2,**max**(regiao2))

cord.y1 <- **c**(0,**dnorm**(regiao1,mean=155,sd=4),0)

cord.y2 <- **c**(0,**dnorm**(regiao2,mean=155,sd=4),0)

**curve**(**dnorm**(x,mean=155,sd=4),xlim=**c**(125,190),ylim=**c**(0,0.15),

xlab="",ylab="",xaxs="i",yaxs="i",lwd=2, xaxt='n')

**polygon**(cord.x1,cord.y1,col='lightblue3')

**polygon**(cord.x2,cord.y2,col='lightblue3')

**lines**(x=**c**(155,155),y=**c**(0,**dnorm**(155,155,4)),col="darkgray",lty=2)

**axis**(side=1,at = **c**(147.16,155,162.84),

labels = **c**("X\_c1","155","X\_c2"))

**par**(new=TRUE)

*# Curva tracejada*

**curve**(**dnorm**(x,mean=145,sd=4.5),xlim=**c**(130,190),ylim=**c**(0,0.15),

xlab="",ylab="",xaxs="i",yaxs="i",col="gray",lty=2, lwd=2, xaxt='n')

**par**(new=TRUE)

*# Curva tracejada*

**curve**(**dnorm**(x,mean=140,sd=3),xlim=**c**(130,190),ylim=**c**(0,0.15),

xlab="",ylab="",xaxs="i",yaxs="i",col="gray",lty=2, lwd=2, xaxt='n')

**par**(new=TRUE)

*# Curva tracejada*

**curve**(**dnorm**(x,mean=170,sd=3),xlim=**c**(130,190),ylim=**c**(0,0.15),

xlab="",ylab="",xaxs="i",yaxs="i",col="gray",lty=2, lwd=2, xaxt='n')

**par**(new=TRUE)

*# Curva tracejada*

**curve**(**dnorm**(x,mean=175,sd=4.5),xlim=**c**(130,190),ylim=**c**(0,0.15),

xlab="",ylab="",xaxs="i",yaxs="i",col="gray",lty=2, lwd=2, xaxt='n')

Exemplo 12.2

Figura 12.5

**RC**(tipo = "bilateral",480,520,500,5,0.01)

Exemplo 12.3

Figura 12.7

pc<-**RC**(tipo = "inferior", x\_min = 0.45,x\_max = 0.75, media = 0.60, desvpad = **sqrt**(0.24/200), alpha = 0.05)

Poder de um Teste

limites<-**RC**(tipo="bilateral", alpha=0.01, media=500, desvpad=5, grafico = FALSE)

Tabela 12.1

mu\_abaixo=**c**(500,498,495,492,490,487,485,480,475)

mu\_acima=**c**(500,502,505,508,510,513,515,520,525)

tab12.1<-**data.frame**(mu\_abaixo,mu\_acima,

pi\_mu=(1-**sapply**(mu\_acima,function(mu){**return**(**pnorm**(limites[2],mean=mu,sd=5)-**pnorm**(limites[1],mean=mu,sd=5))}))\*100,

beta\_mu=**sapply**(mu\_acima,function(mu){**return**(**pnorm**(limites[2],mean=mu,sd=5)-**pnorm**(limites[1],mean=mu,sd=5))})\*100)

**print**(**kable**(tab12.1,col.names = **c**("À esquerda de 500", "À direita de 500", "pi(mu) (em %)","beta(mu) (em %)"),digits = 1))

Figura 12.8

**par**(mfrow=**c**(6,1))

**RC\_**(tipo="bilateral", alpha=0.01, media=500, desvpad=5, x\_min=460,x\_max=540,media\_=500)

**RC\_**(tipo="superior" , alpha=0.005, media=502, desvpad=5, x\_min=460,x\_max=540,media\_=500)

**RC\_**(tipo="superior" , alpha=0.005, media=510, desvpad=5, x\_min=460,x\_max=540,media\_=500)

**RC\_**(tipo="superior" , alpha=0.005, media=515, desvpad=5, x\_min=460,x\_max=540,media\_=500)

**RC\_**(tipo="inferior" , alpha=0.005, media=495, desvpad=5, x\_min=460,x\_max=540,media\_=500)

**RC\_**(tipo="inferior" , alpha=0.005, media=487, desvpad=5, x\_min=460,x\_max=540,media\_=500)

Figura 12.9

x=**c**(475,480,485,487,490,492,495,498,500,502,505,508,510,513,515,520,525)

**plot**(x,(1-**sapply**(x,function(mu){**return**(**pnorm**(limites[2],mean=mu,sd=**sqrt**(400/16))-**pnorm**(limites[1],mean=mu,sd=**sqrt**(400/16)))}))\*100,type="b", ylab="", xlim=**c**(475,525),ylim=**c**(0,100))

**par**(new=T)

limites\_100<-**RC**(tipo="bilateral", alpha=0.01, media=500, desvpad=**sqrt**(400/100), grafico = FALSE)

**plot**(x,(1-**sapply**(x,function(mu){**return**(**pnorm**(limites\_100[2],mean=mu,sd=**sqrt**(400/100))-**pnorm**(limites\_100[1],mean=mu,sd=**sqrt**(400/100)))}))\*100,type="l", ylab="", lty=2, xlim=**c**(475,525), ylim=**c**(0,100))

Exemplo 12.4

limites\_100<-**RC**(tipo="bilateral", alpha=0.01, media=500, desvpad=**sqrt**(400/100), grafico = FALSE)

Exemplo 12.5

val\_p=**pnorm**(0.52, 0.6,**sqrt**(0.24/200))

Figura 12.11

**RC**(tipo = "inferior", x\_min = 0.5,x\_max=0.7,media = 0.6,desvpad = **sqrt**(0.24/200),alpha = val\_p)

Exemplo 12.7:

(1-**pnorm**(314,300,30/**sqrt**(10)))\*2

Exemplo 12.8

*#Calculo da estatistica observada:*

q\_obs=15\*169/100

q\_obs

*#Calculo da RC:*

alpha=0.05 *# nivel de significancia desejado*

n=16 *# numero de pacotes*

qc\_inf=**qchisq**(p = alpha/2,df=n-1) *#valor critico da regiao inferior*

qc\_sup=**qchisq**(p = 1- alpha/2,df=n-1) *#valor critico da regiao superior*

*#Valores críticos:*

**c**(qc\_inf,qc\_sup)

Figura 12.12

*# Plotando o gráfico:*

x\_min=0

x\_max=35

x<-**seq**(x\_min,x\_max,0.1)

regiao1=**seq**(0,**qchisq**(p=alpha/2, df=n-1),0.001)

regiao2=**seq**(**qchisq**(p=1-alpha/2, df=n-1),x\_max,0.001)

cord.x1 <- **c**(**min**(regiao1),regiao1,**max**(regiao1))

cord.x2 <- **c**(**min**(regiao2),regiao2,**max**(regiao2))

cord.y1 <- **c**(0,**dchisq**(regiao1,df=n-1),0)

cord.y2 <- **c**(0,**dchisq**(regiao2,df=n-1),0)

**curve**(**dchisq**(x,df=n-1),xlim=**c**(x\_min,x\_max),xlab="",ylab="",xaxs="i",yaxs="i",lwd=2, xaxt='n')

**polygon**(cord.x1,cord.y1,col='orange2')

**polygon**(cord.x2,cord.y2,col='orange2')

**axis**(side=1,at = **c**(**qchisq**(p=alpha/2,df=n-1), **qchisq**(p=1-alpha/2,df=n-1)),

labels = **round**(**c**(**qchisq**(p=alpha/2,df=n-1), **qchisq**(p=1-alpha/2,df=n-1)),3))

Exemplo 12.9

x<-**c**(253,187,96,450,320,105)

s\_o<-**var**(x)

*#Variancia observada:*

s\_o

alpha=0.1 *# nivel de significancia desejado*

n=6

qc\_inf=**qchisq**(p = alpha/2,df=n-1) *#valor critico da regiao inferior*

qc\_sup=**qchisq**(p = 1- alpha/2,df=n-1) *#valor critico da regiao superior*

*#Valores críticos:*

**c**(qc\_inf,qc\_sup)

**pchisq**(qc\_sup,df=n-1)-**pchisq**(qc\_inf,df=n-1)

Exemplo 12.10

t\_c=**qt**(p=1-0.05,df=24)

t\_c

*# Valor observado da estatística t:*

T=**sqrt**(25)\*(31.5-30)/3

T

1-**pt**(T,df=24)

*# IC para mu*

alpha=0.05

n=25

lim\_inf=31.5+**qt**(alpha/2,df=n-1)\*(3/**sqrt**(n))

lim\_sup=31.5+**qt**(1-alpha/2,df=n-1)\*(3/**sqrt**(n))

**print**(**paste**("IC(mu;",1-alpha,")=[",**round**(lim\_inf,2),";",**round**(lim\_sup,2),"]"))

Capítulo 13: Inferência para duas populações

Exemplo 13.2

maqA<-**c**(145,127,136,142,141,137)

maqB<-**c**(143,128,132,138,142,132)

s2\_a<-**var**(maqA)

s2\_b<-**var**(maqB)

teste\_13\_2<-**var.test**(x = maqA, y = maqB, conf.level = 0.9, alternative = "two.sided")

teste\_13\_2

Exemplo 13.3

**teste\_var**(s1=85,s2=8,n1 = 6,n2=6,alpha = 0.1)

Exemplo 13.4

**teste\_t**(x\_barra1=68,x\_barra2=76,s1=50,s2=52,n1=12,n2=15,alpha=0.05, H\_1="A<B",var\_iguais=TRUE)

**IC\_diff\_medias**(x\_barra1=68,x\_barra2=76,s1=50,s2=52,n1=12,n2=15,gama=0.95)

Exemplo 13.5

**teste\_t**(x\_barra1=70.5,x\_barra2=84.3,s1=81.6,s2=210.8,

n1=15,n2=20,alpha=0.05, H\_1="A!=B", var\_iguais=FALSE)

Exemplo 13.7

**pwilcox**(q=87-10\*11/2,m=10,n=10)

Exemplo 13.8

controle<-**c**(1.3, 1.5, 2.1)

tratamento<-**c**(1.5,2.5)

**pwilcox**(q=7.5,m=3,n=2)

**wilcox.test**(controle,tratamento, alternative = "less")

Exemplo 13.9

Figura 13.4

T<-**c**(0.6,0.63,0.83,0.85,0.91,0.95,1.01,1.03,1.03,1.16,

1.19,1.2,1.26,1.28,1.3,1.37,1.45,1.54,1.68,2.2)

C<-**c**(0.52,0.77,0.79,0.79,0.81,0.81,0.89,0.98,1.01,1.18,

1.19,1.2,1.34,1.36,1.38,1.43,1.64,1.71,2.16,2.25)

**hist**(T,breaks=**seq**(0,2.8,0.4),col="lightgray", main="", freq = FALSE,yaxt='n',ylab="")

Figura 13.5

**hist**(C,breaks=**seq**(0,2.8,0.4),col="lightgray", main="", freq = FALSE,yaxt='n',ylab="")

Tabela 13.7

media<-(**c**(C,T))

tipo<-**factor**(**c**(**rep**("C",20),**rep**("T",20)))

posto<-**rank**(**c**(C,T))

tab13\_7<-**data.frame**(media,tipo,posto)

tab13\_7<-tab13\_7[**order**(tab13\_7$posto),]

**print**(**kable**(**t**(tab13\_7[ 1:10,]),col.names = **rep**("\\",10), caption="\*\*Tabela 13.7\*\*: Postos para o Exemplo 13.9"))

**print**(**kable**(**t**(tab13\_7[11:20,]),col.names = **rep**("\\",10)))

**print**(**kable**(**t**(tab13\_7[21:30,]),col.names = **rep**("\\",10)))

**print**(**kable**(**t**(tab13\_7[31:40,]),col.names = **rep**("\\",10)))

**attach**(tab13\_7)

**wilcox.test**(media~tipo, alternative="less")

Exemplo 13.10

A<-**c**(80,72,65,78,85)

B<-**c**(75,70,60,72,78)

operador<-1:5

tab13\_8<-**data.frame**(operador,A,B)

**t.test**(A, B, alternative="greater", paired=TRUE, conf.level = 0.90)

Exemplo 13.12

**p.test**(p1=168/400, p2=180/600,n1=400,n2=600,alpha=0.05, H1="p1!=p2")

**IC\_proporcoes**(p1=168/400, p2=180/600,n1=400,n2=600,gama=0.95)

tab13\_12A<-**matrix**(**c**(168,232,180,420),ncol=2,nrow=2,byrow = TRUE)

**prop.test**(x=tab13\_12A, alternative = "two.sided")

Exemplo Computacional

tab13\_12<-**data.frame**(sujeito=1:26,**rbind**(

**c**(2.18 , 0.43),**c**(2.05 , 0.08),**c**(1.05 , 0.18),**c**(1.95 , 0.78),**c**(0.28 , 0.03),

**c**(2.63 , 0.23),**c**(1.5 , 0.2 ),**c**(0.45 , 0 ),**c**(0.7 , 0.05),**c**(1.3 , 0.3 ),

**c**(1.25 , 0.33),**c**(0.18 , 0 ),**c**(3.3 , 0.9 ),**c**(1.4 , 0.24),**c**(0.9 , 0.15),

**c**(0.58 , 0.1 ),**c**(2.5 , 0.33),**c**(2.25 , 0.33),**c**(1.53 , 0.53),**c**(1.43 , 0.43),

**c**(3.48 , 0.65),**c**(1.8 , 0.2 ),**c**(1.5 , 0.25),**c**(2.55 , 0.15),**c**(1.3 , 0.05),

**c**(2.65 , 0.25)))

**names**(tab13\_12)[2:3]<-**c**("antes", "depois")

tab13\_12$d<-tab13\_12$antes-tab13\_12$depois

tab13\_12$postos<-**rank**(**abs**(tab13\_12$d))

**print**(**kable**(tab13\_12,caption="\*\*Tabela 13.12\*\*: Índices de placa bacteriana.",col.names = **c**("Sujeito","Antes(x\_i)","Depois(y\_i)","d\_i=x\_i-y\_i","postos de |d\_i|")))

Figura 13.6

xp<-**list**(tab13\_12$antes,tab13\_12$depois)

**boxplot**(xp,pch="-", col="lightblue", border="black", boxwex=0.3, names=**c**("x","y"))

Quadro 13.1

diff\_trats<-**t.test**(tab13\_12$antes,tab13\_12$depois, alternative="two.sided", paired=TRUE, conf.level = 0.95)

diff\_trats

Figura 13.7

**stripchart**(tab13\_12$d,method = "overplot", offset = 2, at=1,

pch = 1, col="darkred",ylab=NA,

xlim=**c**(0,3))

**par**(new=TRUE)

**plot**(x=**c**(0,**mean**(tab13\_12$d)),

y=**c**(0.8,0.8),xlim=**c**(0,3),pch=**c**(20,20),ylim=**c**(0,2),ylab="",xlab="", col="darkblue")

**par**(new=TRUE)

**plot**(x=**c**(diff\_trats$conf.int[1],diff\_trats$conf.int[2]), col="blue",

y=**c**(0.8,0.8),xlim=**c**(0,3),pch=**c**("|","|"),ylim=**c**(0,2), type="b",ylab="",xlab="")

**text**(x=**c**(0,**mean**(tab13\_12$d)),y=**c**(0.7,0.7), labels=**c**("H\_0","Média(d)"), col="lightblue4")

Figura 13.8

**boxplot**(tab13\_12$d,pch="|", col="lightblue", border="black", boxwex=0.3,horizontal = TRUE )

**par**(new=TRUE)

**plot**(x=**c**(0,**mean**(tab13\_12$d)),

y=**c**(0.8,0.8),xlim=**c**(0,3),pch=**c**(20,20),ylim=**c**(0,2),ylab="",xlab="", col="darkblue",xaxt='n', yaxt='n')

**par**(new=TRUE)

**plot**(x=**c**(diff\_trats$conf.int[1],diff\_trats$conf.int[2]), col="blue",

y=**c**(0.8,0.8),xlim=**c**(0,3),pch=**c**("|","|"),ylim=**c**(0,2), type="b",ylab="",xlab="",

xaxt='n', yaxt='n')

**text**(x=**c**(0,**mean**(tab13\_12$d)),y=**c**(0.7,0.7), labels=**c**("H\_0","Média(d)"), col="lightblue4")

Capítulo 14: Análise de Aderência e Associação

Exemplo 14.1

f\_obs = **c**(43,49,56,45,66,41)

f\_esp = **c**(50,50,50,50,50,50)

q\_obs = **sum**((f\_obs-f\_esp)^2)/50

q\_c=**qchisq**(0.95,5)

if(q\_obs > q\_c){

**print**("Rejeito H0")

}else **print**("Aceito H0")

**chisq.test**(f\_obs)

Exemplo 14.2

c\_hum<-**c**(15,20,30,20,15)

c\_bio<-**c**( 8,23,18,34,17)

tab14\_2<-**rbind**(c\_hum,c\_bio)

test\_tab14\_2=**chisq.test**(tab14\_2)

test\_tab14\_2

tab14\_8 = **rbind**(test\_tab14\_2$expected,

total=**apply**(test\_tab14\_2$expected,2,sum))

tab14\_8=**rbind**(test\_tab14\_2$expected, total=**apply**(test\_tab14\_2$expected,2,sum))

**print**(**kable**(tab14\_8, caption = "\*\*Tabela 14.8\*\*: Frequências absolutas sob $H\_0$", col.names = **c**("A","B","C","D","E")))

Exemplo 14.3

M<-**data.frame**(uso\_hospital=**c**("usaram\_hospital", "nao\_usaram\_hospital"),homens=**c**(100,900),mulheres=**c**(150,850), row.names = 1)

M

Exemplo 14.4

Tabela 14.5

acidentes<-**c**(32,40,20,25,33)

test\_tab14\_5<-**chisq.test**(acidentes)

q\_i<-(acidentes-test\_tab14\_5$expected)^2/test\_tab14\_5$expected

tab14\_5<- **rbind**(acidentes, e\_i=test\_tab14\_5$expected, q\_i)

tab14\_5<-**cbind**(tab14\_5,**c**(**sum**(tab14\_5[1,]),**sum**(tab14\_5[2,]),**sum**(tab14\_5[3,])))

test\_tab14\_5

Exemplo 14.5

x<-0:10

n\_k<- **c**(57,203,383,525,532,408,273,139,45,27,16)

np\_k<-**c**(54.399,210.523,407.361,525.496,508.418,393.515,253.817,140.325,67.882,29.189,17.075)

tab6\_13<-**cbind**(x,n\_k,np\_k)

**teste\_quiquad**(n\_k,np\_k,0.05,10)

Exemplo 14.6

dados.ex14\_6<-**c**( 1.04, 1.73, 3.93, 4.44, 6.37, 6.51,

7.61, 7.64, 8.18, 8.48, 8.57, 8.65,

9.71, 9.87, 9.95, 10.01, 10.52, 10.69,

11.72, 12.17, 12.61, 12.98, 13.03, 13.16,

14.11, 14.60, 14.64, 14.75, 16.68, 22.14)

classe<-**numeric**()

classe[1]<-**sum**(dados.ex14\_6<6.63)

classe[2]<-**sum**(dados.ex14\_6 > 6.63 & dados.ex14\_6 < 10)

classe[3]<-**sum**(dados.ex14\_6 > 10 & dados.ex14\_6 < 13.37)

classe[4]<-**sum**(dados.ex14\_6 > 13.37)

test\_tab14\_6<- **chisq.test**(classe)

Tabela 14.6

tab14\_6<-**cbind**(**rbind**(classe,e\_i=test\_tab14\_6$expected),**c**(**sum**(classe),**sum**(test\_tab14\_6$expected)))

**teste\_quiquad**(classe,test\_tab14\_6$expected,0.05,3)

**chisq.test**(classe)

Figura 14.1

**par**(mfrow=**c**(1,2))

**hist**(dados.ex14\_6,col="lightblue3",border = "white",xlab="Dados Normal")

**boxplot**(dados.ex14\_6, xlab="Dados Normal",col="lightblue", pch=16, border="darkgray")

Exemplo 14.7

tab14\_10<-**rbind**(P\_1T=**c**(29,60,9,2),P\_2C=**c**(37,44,13,6))

**chisq.test**(tab14\_10)

Exemplo 14.8

*# Dados das tabelas 4.8 e 4.9*

dados.tab4\_8<-**data.frame**(**rbind**(

**matrix**(**rep**(**c**("1.Consumidor","1.São Paulo"),times=214),ncol=2,byrow=T),

**matrix**(**rep**(**c**("1.Consumidor","2.Paraná"),times=51),ncol=2,byrow=T),

**matrix**(**rep**(**c**("1.Consumidor","3.Rio G. do Sul"),times=111),ncol=2,byrow=T),

**matrix**(**rep**(**c**("2.Produtor","1.São Paulo"),times=237),ncol=2,byrow=T),

**matrix**(**rep**(**c**("2.Produtor","2.Paraná"),times=102),ncol=2,byrow=T),

**matrix**(**rep**(**c**("2.Produtor","3.Rio G. do Sul"),times=304),ncol=2,byrow=T),

**matrix**(**rep**(**c**("3.Escola","1.São Paulo"),times=78),ncol=2,byrow=T),

**matrix**(**rep**(**c**("3.Escola","2.Paraná"),times=126),ncol=2,byrow=T),

**matrix**(**rep**(**c**("3.Escola","3.Rio G. do Sul"),times=139),ncol=2,byrow=T),

**matrix**(**rep**(**c**("4.Outras","1.São Paulo"),times=119),ncol=2,byrow=T),

**matrix**(**rep**(**c**("4.Outras","2.Paraná"),times=22),ncol=2,byrow=T),

**matrix**(**rep**(**c**("4.Outras","3.Rio G. do Sul"),times=48),ncol=2,byrow=T)))

**colnames**(dados.tab4\_8)<-**c**("tipo\_de\_cooperativa","estado")

**attach**(dados.tab4\_8)

Tabela 14.11

**CrossTable**(estado,tipo\_de\_cooperativa,

prop.r=FALSE, prop.c=FALSE, prop.t=FALSE, prop.chisq=FALSE, expected=TRUE,

digits=0)

Exemplo 14.9

anos\_exp<-**c**(2,4,5,6,8)

n\_clientes<-**c**(48,56,64,60,72)

**cor.test**(anos\_exp,n\_clientes,alternative="two.sided", method="pearson",conf.level=0.95)

14.6 Outro teste de Aderência

Exemplo 14.6 (Continuação)

Figura 14.4

**ks.test**(dados.ex14\_6,y="pnorm",mean=10,sd=5)

*# Gráfico qxq da distribuição emprírica e da distribuição Normal(10,25)*

**qqnorm**(dados.ex14\_6,cex=2,pch=20,col="darkblue", xlab="Quantis da Normal Padrão", ylab="Quantis dos dados", main="")

**qqline**(dados.ex14\_6)

Capítulo 15: Inferência para várias populações

Exemplo 15.1

Tabela 15.1

**attach**(tab15\_1)

**summary2**(tempo\_Y[sexo\_W=="H"])

**summary2**(tempo\_Y[sexo\_W=="M"])

summary\_sexo<-**sapply**(**levels**(sexo\_W),function(sex){**summary2**(tempo\_Y[sexo\_W==sex])})

summary\_sexo

summary\_idade<-**sapply**(**levels**(idade\_X),function(age){**summary2**(tempo\_Y[idade\_X==age])})

summary\_idade

Tabela 15.4

mean\_idade<-**rep**(0,20)

for(age in **levels**(idade\_X)){

mean\_idade[mean\_idade==0]<-((idade\_X==age)\***mean**(tempo\_Y[idade\_X==age]))[mean\_idade==0]

}

mean\_sexo<-**rep**(0,**length**(sexo\_W))

for(sex in **levels**(sexo\_W)){

mean\_sexo[mean\_sexo==0]<-((sexo\_W==sex)\***mean**(tempo\_Y[sexo\_W==sex]))[mean\_sexo==0]

}

e1<-tempo\_Y-**mean**(tempo\_Y)

e2<-tempo\_Y-mean\_sexo

e3<-tempo\_Y-mean\_idade

tab15\_4<-**cbind**(tab15\_1[,1:4],e1,e2,e3)

15.2.3 Intervalos de Confiança

**IC\_t**( **mean**(tempo\_Y[sexo\_W=="H"]),**mean**(tempo\_Y[sexo\_W=="M"]),

**var**(tempo\_Y[sexo\_W=="H"]),**var**(tempo\_Y[sexo\_W=="M"]),10,10,gama=0.95)

**IC\_diff\_medias**(**mean**(tempo\_Y[sexo\_W=="M"]),**mean**(tempo\_Y[sexo\_W=="H"]),

**var**(tempo\_Y[sexo\_W=="M"]),**var**(tempo\_Y[sexo\_W=="H"]),10,10,gama=0.95)

15.2.4 Tabela de Análise de Variância

fit15\_4\_sexo<-**anova**(**lm**(tempo\_Y~sexo\_W))

**par**(mfrow=**c**(2,2))

**plot**(**aov**(**lm**(tempo\_Y~sexo\_W)))

Exemplo 15.4

**qf**(0.95,1,18)

if(fit15\_4\_sexo$`F value`[1]>**qf**(0.95,1,18)){

**print**("Rejeito H\_0")

} else **print**( "Aceito H0")

Modelo para mais de duas populações

Figura 15.2

boxplot(tempo\_Y~idade\_X, col="lightblue3")

fit15\_4\_idade<-**anova**(**lm**(tempo\_Y~idade\_X))

**par**(mfrow=**c**(2,2))

**plot**(**aov**(**lm**(tempo\_Y~idade\_X)))

Figura 15.3

**plot**(**as.vector**(idade\_X),e3,col="darkred",pch=20,xaxt="n")

**axis**(1,at=**c**(19.6,24.6,29.6,34.6,39.6),labels=**c**("20 anos","25 anos","30 anos","35 anos","40 anos"))

**abline**(v=**c**(19.6,24.6,29.6,34.6,39.6), col="darkgrey", lty=2)

15.4 Comparações entre médias

**qf**(0.95,4,15)

if(fit15\_4\_idade$`F value`[1]>**qf**(0.95,4,15)){

**print**("Rejeito H\_0")

} else **print**( "Aceito H0")

Exemplo 15.5

medias\_idade<-**sapply**(**levels**(idade\_X),function(age){**mean**(tempo\_Y[idade\_X==age])})

medias\_idade

**diff**(medias\_idade)

Exemplo 15.6

v<-**levels**(idade\_X)

SQ<-fit15\_4\_idade$`Sum Sq`[1] *#fit15\_4\_idade ;e o modelo ajustado e $`Sum Sq`[1] é a SQ para a idade.*

ics\_diff\_idades<-**matrix**(NA,10,2)

cont=1

lab\_ics<-**character**(0)

vec\_diff\_idades<-**numeric**()

for( i in 1:4)

for(j in (i+1):5){

ics\_diff\_idades[cont,]<-**IC\_t\_diff**(**mean**(tempo\_Y[idade\_X==v[i]]),**mean**(tempo\_Y[idade\_X==v[j]]),SQ,20,4,4,gama = 0.995)

vec\_diff\_idades[cont]<-**mean**(tempo\_Y[idade\_X==v[i]])-**mean**(tempo\_Y[idade\_X==v[j]])

lab\_ics[cont]<-**paste**(v[i],v[j],sep="\_")

cont=cont+1

}

**rownames**(ics\_diff\_idades)<-lab\_ics

ics\_diff\_idades

### Grafico com os ICs

**plot**(1:10,vec\_diff\_idades, ylim=**c**(**min**(ics\_diff\_idades)-1,**max**(ics\_diff\_idades)+1), type="p",xaxt="n",xlab="pares",ylab="ICs")

**axis**(1,at=1:10,labels=lab\_ics) *# labels no eixo x*

for (i in 1:10) {

if(ics\_diff\_idades[i,1]>0 || ics\_diff\_idades[i,2]<0) {color="red"} else color="darkblue"

**lines**(x =**c**(i,i), y=**t**(ics\_diff\_idades)[,i], col=color) *# Plotando os ICs*

}

Exemplo 15.7

**bartlett.test**(x=tempo\_Y, g=idade\_X)

Quadro 15.1

fit15\_4\_idade

ics\_idades<-**matrix**(NA,5,2)

cont=1

for(age in **levels**(idade\_X)){

ics\_idades[cont,]<-**IC**(n = 4, **mean**(tempo\_Y[idade\_X==age]),**sd**(tempo\_Y[idade\_X==age]),gama=0.95)

cont=cont+1

}

**print**(**kable**(**cbind**(medias\_idade,ics\_idades),caption="Média e Intervalos de confiança por grupo de idade",col.names = **c**("Média","Limite Inferior", "Limite Superior")))

### gráfico dos ICs

**plot**(**c**(20,25,30,35,40),medias\_idade, ylim=**c**(**min**(ics\_idades)-1,**max**(ics\_idades)+1), type="p",xlab="Idade",ylab="ICs")

for (i in 1:5) **lines**(x =**c**(v[i],v[i]), y=**t**(ics\_idades)[,i]) *# Plotando os ICs*

Figuras 15.4 e 15.5

**boxplot**(e3~**as.vector**(idade\_X),col="lightblue3",pch=20)

**boxplot**(e3,col = "lightblue3")

Capítulo 16: Regressão Linear Simples

Figura 16.1

**attach**(tab15\_1)

n\_idade\_X<-**as.numeric**(**as.character**(idade\_X))

**plot**(n\_idade\_X,tempo\_Y, pch=20, xlab="Idade(x)", ylab="Estímulo(y)", col="darkblue")

**abline**(**lm**(tempo\_Y~n\_idade\_X), lwd=2, col="red")

Exemplo 16.1

fit16\_1<-**lm**(tempo\_Y~n\_idade\_X)

fit16\_1

*# fit16\_1$coefficients nos dá os coeficientes ajustados, sendo fit16\_1$coefficients[1] o intercepto e fit16\_1$coefficients [2] a inclinação.*

y\_20=fit16\_1$coefficients[1]+fit16\_1$coefficients[2]\*20

y\_20

**predict**(object = fit16\_1)

y\_33=fit16\_1$coefficients[1]+fit16\_1$coefficients[2]\*33

y\_33

*# Na função `predict` é possivel também calcular os valores preditos para valores específicos das variáveis explicativas:*

newdata=**data.frame**(n\_idade\_X=20) *# `data.frame` contém os dados das variáveis explicativas utilizados para calcular as previsões.*

**predict**(fit16\_1, newdata)

Tabela 16.1

tab16\_1<-**cbind**(tab15\_1[,1:3],resid\_fit16\_1=**resid**(fit16\_1))

**print**(**kable**(tab16\_1,caption="\*\*Tabela 16.1\*\*: Resíduos para o modelo (16.18).", col.names = **c**("Tempo de Reação", "Sexo", "Idade", "Resíduos")))

anova\_fit16\_1<-**anova**(fit16\_1)

SQRes=**sum**(tab16\_1$resid\_fit16\_1^2)

S2e=SQRes/(20-2)

SQRes

S2e

**sqrt**(S2e)

2\***sqrt**(S2e)

Exemplo 16.2

s2<-**var**(tempo\_Y)

s2

Exemplo 16.3

**summary**(fit16\_1)$r.squared

Tabela 16.3

**anova**(fit16\_1)

Exemplo 16.4

**confint**(fit16\_1)

Exemplo 16.5

**summary**(fit16\_1)

Exemplo 16.6

newdata=**data.frame**(n\_idade\_X=28) *# `data.frame` contém os dados das variáveis explicativas utilizados para calcular as previsões.*

**predict**(fit16\_1, newdata, interval="predict")

Exemplo 16.7

**attach**(tab16\_1)

nui=1/**length**(n\_idade\_X)+(n\_idade\_X-**mean**(n\_idade\_X))^2/**sum**((n\_idade\_X-**mean**(n\_idade\_X))^2)

ri=resid\_fit16\_1/(**sqrt**(S2e)\***sqrt**(1-nui))

tab16\_4<-**cbind**(tab16\_1[3:4],zi=resid\_fit16\_1/**sqrt**(S2e),ri)

Figura 16.4

**par**(mfrow=**c**(1,2))

**plot**(n\_idade\_X,resid\_fit16\_1, cex=2, col="darkblue", pch=20, xlab="Idade", ylab="Resíduos",main="(a)")

**abline**(h=0)

**plot**(n\_idade\_X,tab16\_4$zi, cex=2, col="darkblue", pch=20, xlab="Idade", ylab="Resíduos padronizados",main="(b)")

**abline**(h=0)

Figura 16.8

**hist**(resid\_fit16\_1, xlab="Resíduos", ylab="", col="lightblue3", border="white", main="")

Figura 16.9

**qqnorm**(resid\_fit16\_1, cex=1.5, col="darkblue", xlab="Quantis da normal padrão", ylab="Quantis dos resíduos", pch=20)

**qqline**(resid\_fit16\_1, col="darkred")

Análise de Resíduos no R:

**par**(mfrow=**c**(2,2))

**plot**(fit16\_1)

Exemplo 16.8

Tabela 16.5

fit16\_5<-**lm**(tab16\_5$yi~tab16\_5$xi)

fit16\_5

Figura 16.10

**attach**(tab16\_5)

**plot**(xi,yi, pch=20, xlab="Temperatura", ylab="Vapor", col="darkblue",main="(a)")

**abline**(**lm**(yi~xi), lwd=2, col="red")

**par**(mfrow=**c**(1,2))

**plot**(xi,**resid**(fit16\_5), cex=2, col="darkblue", pch=20, xlab="Temperatura", ylab="Resíduos",main="(b)")

**abline**(h=0)

**qqnorm**(**resid**(fit16\_5), cex=1.5, col="darkblue", xlab="Quantis da normal padrão", ylab="Quantis dos resíduos", pch=20,main="(c)")

**qqline**(**resid**(fit16\_5), col="darkred")

Exemplo 16.9

*#Ajuste do modelo para os dados do Exemplo 16.9 sem o intercepto é ajustado adicionando `-1` à equação:*

fit16\_9<-**lm**(tab16\_9A$y~tab16\_9A$x-1)

fit16\_9

**summary**(fit16\_9)

S2e=**sum**(**resid**(fit16\_9)^2)/(9-1)

S2e

**sqrt**(S2e)

**confint**(fit16\_9)

**plot**(tab16\_9A$x,tab16\_9A$y, pch=20, xlab="x", ylab="y", col="darkblue")

**abline**(**lm**(tab16\_9A$y~tab16\_9A$x), lwd=2, col="red")

16.6.2 Modelo Não Lineares

Tabela 16.6

ano=**seq**(1961,1979,2)

inflacao\_Y=**c**(9,24,72,128,192,277,373,613,1236,2639)

tab16\_6<-**data.frame**(

ano=ano,

t=ano-**mean**(ano),

inflacao\_Y=inflacao\_Y,

log\_Y=**log**(inflacao\_Y))

Exemplo 16.10

Figura 16.12

**plot**(t,inflacao\_Y, pch=20, col="darkblue", ylim=**c**(0,**max**(inflacao\_Y)+5),ylab="Inflação",xlab="Ano(t)")

**par**(new=T)

**plot**(t,**exp**(**predict**(fit16\_10)),pch="+", col="darkred", ylim=**c**(0,**max**(inflacao\_Y)+5),ylab="",xlab="")

Exemplo 16.10

fit16\_10<-**lm**(log\_Y~t)

fit16\_10

**exp**(fit16\_10$coefficients[1])

Figura 16.13

**plot**(tab16\_6$t,tab16\_6$log\_Y, pch=20, xlab="x", ylab="y", col="darkblue")

**abline**(**lm**(tab16\_6$log\_Y~tab16\_6$t), lwd=2, col="red")

resid\_exp=inflacao\_Y-**exp**(**predict**(fit16\_10))

tab16\_7<-**data.frame**(t=tab16\_6$t,resid\_reta=**resid**(fit16\_10),resid\_exp)

**print**(**kable**(tab16\_7,caption="\*\*Tabela 16.7\*\*: Resíduos para os modelos linear e exponencial.", col.names=**c**("t", "Resíduos Reta", "Resíduos Exponencial")))

Figura 16.14

**plot**(tab16\_7$t,tab16\_7$resid\_reta, xlab="ano", ylab="Resíduos reta", pch=20)

**abline**(h=0,lty=2, col="gray")

Figura 16.15

**plot**(tab16\_7$t,tab16\_7$resid\_exp, xlab="ano", ylab="Resíduos exponencial", pch=20)

**abline**(h=0,lty=2, col="gray")

Figura 16.16

**par**(mfrow=**c**(1,2))

**hist**(tab16\_7$resid\_reta,col="lightblue3",border="white", main="(a)",xlab="",ylab="Resíduos reta")

**hist**(tab16\_7$resid\_exp,col="lightblue3",border="white", main="(b)",xlab="",ylab="Resíduos exponencial")

Figura 16.16

**par**(mfrow=**c**(1,2))

**qqnorm**(tab16\_7$resid\_reta,col="darkblue", main="(a)",xlab="Quantis da normal padrão",ylab="Resíduos reta", pch=20,cex=1.5)

**qqline**(tab16\_7$resid\_reta, col="darkred")

**qqnorm**(tab16\_7$resid\_exp,col="darkblue", main="(b)",xlab="Quantis da normal padrão",ylab="Resíduos exponencial", pch=20,cex=1.5)

**qqline**(tab16\_7$resid\_exp, col="darkred")

Exemplo 16.11

Figura 16.19

*# Definin*

grupo\_E=**c**(2,1,4,3,5,8,6)

grupo\_C=**c**(7,12,10,11,9,14)

grupo\_D=**c**(16,13,15,18,17,20,19)

fit16\_11<-**reg\_resist**(**data.frame**(x=n\_idade\_X,y=tab15\_1$tempo\_Y))

fit16\_11

Exemplo 16.12

Figura 16.20

cd\_mercado <- **read.table**("cd-mercado.csv",h=T,skip=4,sep=";",dec=",") *# Leitura dos dados*

**attach**(cd\_mercado)

fit16\_12<-**lm**(telebras[1:39]~indice[1:39])

**plot**(indice[1:39],telebras[1:39],ylab="Tel",xlab="IBV",pch=16,col="darkblue")

**abline**(fit16\_12)

Quadro 16.1

**summary**(fit16\_12)

**anova**(fit16\_12)

Figura 16.21

resid\_fit16\_12<-**resid**(fit16\_12)

**par**(mfrow=**c**(2,2))

**qqnorm**(resid\_fit16\_12, pch=20, col="darkblue")

**qqline**(resid\_fit16\_12, col="darkred")

**plot.ts**(resid\_fit16\_12, pch=20, col="darkblue")

**abline**(h=**c**(-0.7932,0.7932), col="red", lty=2)

**hist**(resid\_fit16\_12, col="lightblue3", border = "white")

**plot**(**fitted**(fit16\_12),resid\_fit16\_12, pch=20, col="darkblue")

**abline**(h=0)

**par**(mfrow=**c**(2,2))

**plot**(fit16\_12)

Exemplo 16.21:

Figura 16.22

**attach**(tab16\_8)

fit\_16\_13<-**lm**(vt~t)

fit\_16\_13

fit\_resist16\_13<-**reg\_resist**(**data.frame**(x=t,y=vt))

fit\_resist16\_13

Referências

* Bussab W.O. e Morettin P.A., *Estatística Básica*, Saraiva, Sao Paulo, 9ed, 2017.