

- [Tutorial](#)
- [Exercícios](#)
- [Apostila](#)

## 6. Tutoriais de Testes de Significância

### Chacal

Vamos entrar os dados de comprimento da mandíbula de machos e fêmeas de chacal. Esses dados reais farão parte do nosso tutorial. Veja cada um dos objetos criados, sua estrutura e sua classe.

```
macho=c(120,107,110,116, 114, 111, 113,117,114,112)
femea=c(110,111,107, 108,110,105,107,106,111,111)

media.m=mean(macho)
media.m
media.f=mean(femea)
media.f
chacal=c(macho,femea)
sexo=factor(rep(c("macho","femea"),each=10))
```

### Dois Gráfico para ver os mesmos dados

Produza o boxplot dos dados com a função `boxplot` e compare com a outra função (mais de uma linha). Em qual gráfico vc. se sente mais seguro para afirmar que há diferenças entre os grupos? Você acredita que uma diferença média de cerca de 5mm na mandíbula possa ser importante?

```
boxplot(chacal~sexo)

plot(1:20,chacal, pch=rep(c(15,16),each=10),col=rep(1:2,each=10))
  for(i in 1:10)
  {
    lines(c(i,i),c(chacal[i],media.m),col=1)
  }
  for(j in 11:20)
  {
    lines(c(j,j),c(chacal[j],media.f),col=2)
  }

lines(c(1,10),c(media.m,media.m),col=1)
lines(c(11,20),c(media.f,media.f),col=2)
```

# Iniciando a Simulação

## Calculando a diferença entre médias

```
mean(chacal)
sd(chacal)
mean(femea) - mean(macho)
mean(macho) - mean(femea)
dif=round(abs(mean(femea) - mean(macho)), 1)
dif
```

### Perguntas:

- Essa diferença pode ser gerada pelo acaso?
- Qual a possibilidade dessa diferença ser gerada pelo acaso?

### Nos comandos do gráfico, abaixo:

- O histograma é a amostra, e a curva é uma simulação da população que a média e o desvio padrão da amostra produzem.

```
hist(chacal, freq=FALSE, xlim=c(95,125))
curve(exp=dnorm(x, mean=mean(chacal), sd=sd(chacal)), from=95, to=125,
col="red", add=T)
```

## Simulando Dados

Utilizando as estimativas de parâmetros da amostra podemos simular uma amostra aleatória de uma população:

```
rnorm(10, mean=mean(chacal), sd=sd(chacal))
round(rnorm(10, mean=mean(chacal), sd=sd(chacal)))
abs(round(rnorm(10, mean=mean(chacal), sd=sd(chacal))))
abs(round(mean(rnorm(10, mean=mean(chacal), sd=sd(chacal))) -
mean(rnorm(10, mean=mean(chacal), sd=sd(chacal)))))
```

### PERGUNTA: Qual cenário estamos simulando???

## Criando Ciclos

A função `for()` cria ciclos de eventos concatenados, seu código é simples, basta eleger uma variável no caso abaixo `i` e dizer quais os valores que essa variável vai assumir em cada ciclo, aqui os valores de 1 a 10. Entre `{ }` concatenamos os eventos que queremos desencadear. A cada ciclo o valor de `i` muda, portanto ele é usado como índice para as operações.


Veja o efeito do código abaixo, a função `cat()` mostra na tela (R Console) os objetos. Os símbolos “\t” e “\n”, são os códigos do R para tabulação e nova linha respectivamente.

```
for(i in 1:10)
{
cat("\n\t", i)
}
```

Agora vamos fazer isso para a nossa simulação. Aqui o pulo do gato é preparar um objeto antes de iniciar o ciclo para poder guardar os resultados. A variável `i` serve tanto para contar os ciclos, quanto para posicionar o resultado no vetor `RESULTA`

```
resulta=rep(NA,10)
for (i in 1:10)
{
  resulta[i]=
    abs(mean( round( rnorm(10,mean=mean(chacal),sd=sd(chacal))) -
mean( round( rnorm(10,mean=mean(chacal),sd=sd(chacal))) )
})
```

Vamos agora usar a **animada** função chamada [simula.r](#), de autoria do Prof. Adalardo.



Atenção usuários do RStudio: a função `simula()` não é tão animada assim na janela do RStudio. Para que ela funcione, antes de usá-la abra uma janela gráfica com o comando `x11()`

Primeiro precisamos carregá-la no espaço de trabalho, com o comando `source` Depois é só simular!

```
source("simula.r")
dif.chacal=simula(macho,femea)
```

Agora mais vezes...

```
dif.chacal=simula(macho,femea,nsim=2000)
table(dif.chacal$diferencas)
n=sum(abs(dif.chacal$diferencas)>=dif)
n
```

## Calculando a Probabilidade do Acaso

```
n/length(dif.chacal$diferencas)
```

## Bicaudal e Unicaudal

Veja agora a mesma simulação para diferenças que não estão em módulo.

```
dif.chacal.uni=simula(macho,femea, teste="uni")
dif.chacal.uni=simula(macho,femea, teste= "uni", nsim=2000)
table(dif.chacal.uni$diferencas)
```

### Podemos Responder às perguntas.

*As Mandibulas de machos e fêmeas são diferentes?*

```
n.maior=sum(round(dif.chacal.uni$diferencas,1)>=round(dif,1))
n.menor=sum(round(dif.chacal.uni$diferencas,1)<=round((dif*-1),1))
n.maior
n.menor
n.sim=2000

## Qual a probabilidade de errar ao fazer a afirmação que as mandibulas são
diferentes?

p.bi=(n.maior+n.menor)/2000
p.bi
```

*A Mandibula de machos são maiores que as de fêmeas ??*

```
## Qual a probabilidade de errar ao fazer a afirmação que as mandibulas dos
machos são maiores?
p.uni=n.maior/n.sim
p.uni
```

## Simulando o teste T

Vamos calcular a estatística t e depois simular a distribuição t.

```
v1=var(macho)
v1
v2=var(femea)
v2
n1=length(macho)
n1
n2=length(femea)
n2

## desvio padrão das diferenças
```

```
s12=sqrt((v1/n1)+(v2/n2))
s12
dif
tvalor=dif/s12
tvalor
```

Agora simulando a distribuição com a função `simula()`

```
res.t= simula(macho,femea,nsim=2000, teste="t")
##
## agora vamos calcular as probabilidades
maior.menor.t=sum(res.t$diferencas>=tvalor | res.t$diferencas<=-tvalor)
maior.menor.t
prob.t=maior.menor.t/2000
prob.t
## compare com o resultado do teste t da função t.test()
t.test(macho,femea)
```

## ANOVA

### Interpretando uma tabela

Aqui calculamos os valores de somatórias quadráticas que são a base da análise de anova. O ponto importante é que essa variação é aditiva e portanto, pode ser decomposta. A variação total é decomposta em variação relacionada ao tratamento(entre grupos) e uma variação interna dos grupos (chamada de erro). A estatística F é a razão entre essas variações após dividir cada uma delas pelo seus respectivos graus de liberdade. Nosso objetivo é construir a tabela de anova abaixo:



Segue o código para cálculo das variações das somatórias quadráticas e das médias quadráticas.

```
are=c(6,10,8,6,14,17,9,11,7,11)
are
arg=c(17,15,3,11,14,12,12,8,10,13)
arg
hum=c(13,16,9,12,15,16,17,13,18,14)
hum
solos=data.frame(are,arg,hum)
solos
str(solos)
boxplot(solos)
media.solos<-apply(solos,2,mean)
vetor.obs=1:30
vetor.dados=c(are,arg,hum)
media.geral=mean(c(are,arg,hum))
media.geral
dif.geral=solos-media.geral
dif.geral
sum(dif.geral)
```

```
round(sum(dif.geral),10)
ss.solos=dif.geral^2
ss.solos
ss.total=sum(ss.solos)
ss.total
vetor.cor<-rep(1:3, each=10)
vetor.medias<-rep(media.solos, each=10)
```

## Calculo da SS total



O código para a produção do gráfico acima.

```
plot(vetor.obs,vetor.dados,ylim=c(0,20),pch=(rep(c(15,16,17),each=10)),col=v
etor.cor,ylab="Variável Resposta", xlab="Observações")
  for(i in 1:30)
  {
    lines(c(i,i),c(vetor.dados[i],mean(vetor.dados)),col=vetor.cor[i])
  }
  abline(h=media.geral)
```

## Somatória Quadráticas Intra grupo



### Código Gráfico

```
plot(vetor.obs,vetor.dados,ylim=c(0,20),pch=(rep(c(15,16,17),each=10)),col=v
etor.cor,main="Variação Intra Grupos",ylab="Variável Resposta",
xlab="Observações")
  for(i in 1:30)
  {
    lines(c(i,i),c(vetor.medias[i],vetor.dados[i]),col=vetor.cor[i])
  }
  lines(c(1,10),c(media.solos[1],media.solos[1]),col=1)
  lines(c(11,20),c(media.solos[2],media.solos[2]),col=2)
  lines(c(21,30),c(media.solos[3],media.solos[3]),col=3)
```

Cálculo dos valores

```
solos
media.solos
ss.are=sum((are-media.solos["are"])^2)
ss.are
ss.arg=sum((arg-media.solos["arg"])^2)
ss.arg
ss.hum=sum((hum-media.solos["hum"])^2)
```

```
ss.hum
ss.intra=ss.are+ss.arg+ss.hum
ss.intra
```



## Somatória Quadráticas Entre grupos



```
plot(vetor.obs,vetor.medias,ylim=c(5,16),pch=(rep(c(15,16,17),each=10)),col=
vetor.cor,main="Variação Entre Grupos",ylab="Variável Resposta",
xlab="Observações")
  for(i in 1:30)
  {
    lines(c(i,i),c(vetor.medias[i],mean(vetor.medias)),col=vetor.cor[i])
  }
  abline(h=media.geral)
points(vetor.obs,vetor.dados,ylim=c(0,20),pch=(rep(c(0,1,2),each=10)),col=ve
tor.cor,cex=0.5)

#### Cálculo dos valores

media.solos=apply(solos,2,mean)
media.solos
media.geral
ss.entre=10*sum((media.solos-media.geral)^2)
ss.entre

ss.intra+ss.entre
ss.total
```

## Tabela Final de Anova



## Cálculo do F

```
ms.entre=ss.entre/2
ms.intra=ss.intra/27
ms.entre
ms.intra
F.solos=ms.entre/ms.intra
F.solos
```

## Distribuição de F

Os gráficos de outras aulas apresentam a distribuição de densidade probabilística, onde a área do gráfico é relacionada à probabilidade de cada valor em intervalos muito pequenos, portanto o valor de p é a área da curva, conceitualmente mais correto que o anterior. Note que os gráficos são muito parecidos, e podem causar confusão!



```
curve(expr=df(x, 2,27),main="Distribuição F de Fisher (df=2,27)",
xlab="Valor F",ylab="Densidade Probabilística (df)",xlim=c(0,10))
abline(v=F.solos,col="red")
abline(h=0, lty=2)
xf=seq(F.solos,10,0.01)
ydf=df(xf,2,27)
polygon(c(F.solos,xf),c(0,ydf),col="red")
text(locator(1),paste("pf(x)
=",round(pf(F.solos,2,27,lower.tail=F),4)),cex=0.8, col="red")
savePlot("densPF.jpeg", "jpeg")
```

## Cálculo do P

```
p.solos=pf(F.solos,2,27, lower.tail=FALSE)
p.solos
```

## Anova no R

```
str(solos)
var.resp=c(solos$are,solos$arg,solos$hum)
var.resp
solos.f=factor(rep(c("are", "arg", "hum"),each=10))
solos.f
res.anova=aov(var.resp~solos.f)
res.anova
summary(res.anova)
```

From:

<http://labtrop.ib.usp.br/> - **Laboratório de Ecologia de Florestas Tropicais**

Permanent link:

[http://labtrop.ib.usp.br/doku.php?id=cursos:ecor:02\\_tutoriais:tutorial6:start](http://labtrop.ib.usp.br/doku.php?id=cursos:ecor:02_tutoriais:tutorial6:start)



Last update: **2020/07/27 18:49**