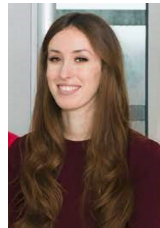


LearningStats: un paquete en R para a docencia

Sabela Varela-Rey, María Isabel Borrajo, Mercedes
Conde-Amboage e Alejandra López-Pérez





**CIT
MAGA**

CENTRO DE INVESTIGACIÓN
E TECNOLOXÍA MATEMÁTICA
DE GALICIA

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

FACULTAD DE MATEMÁTICAS

- Para complementar a súa formación, os estudantes precisan aprender ferramentas estatísticas, e moitos empregan .
- A necesidade de codificación e as saídas difíciles de entender supoñen unha traba para algúns usuarios.
- Co obxectivo de evitar isto e guiar aos usuarios en cada paso, creouse *LearningStats*, un paquete cunha finalidade didáctica.
- Ademais das novas versións de funcións xa existentes, a librería *LearningStats* inclúe novos comandos non dispoñibles ata agora en .

```
x <- rnorm(50, mean = 4, sd = 2)
t.test(x,mu=3.5,conf.level=0.99)

##
## One Sample t-test
##
## data: x
## t = 0.58024, df = 49, p-value = 0.5644
## alternative hypothesis: true mean is not equal to 3.5
## 99 percent confidence interval:
## 2.915576 4.407426
## sample estimates:
## mean of x
## 3.661501
```

```
dat1=rnorm(20,mean=2,sd=1) ; dat2=rnorm(30,mean=2,sd=1)
x1=mean(dat1) ; x2=mean(dat2)
sc1=sd(dat1) ; sc2=sd(dat2)
n1=length(dat1) ; n2=length(dat2)
diffmean.CI(x1,x2,sc1=sc1,n1=n1,n2=n2,paired=FALSE,var.equal=FALSE,
conf.level=0.9)
```

```
## Error in diffmean.CI(x1, x2, sc1 = sc1, n1 = n1, n2 = n2, paired
= FALSE, : For sample 2, either the sample standard deviation or the
cuasi-standard deviation needs to be provided
```

```
x <- rnorm(20, mean = 1, sd = 2)
y <- rnorm(21, mean = 1.5, sd = 1.3)
sample.cov(x,y)
```

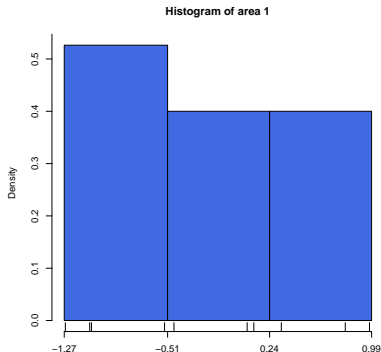
```
## Error in sample.cov(x, y): The length of 'x' and 'y' must be the
same
```

O paquete cobre cinco eidos principais dentro da Estatística:

- Conxuntos de datos.
- Estatística Descritiva: *Histogram*, *summary.freq.table*.
- Modelos de Probabilidade: *plotFS*, *AproxBinomNorm*.
- Inferencia Estatística: *Mean.CI*, *diffproportion.test*.
- Regresión: *plotAnova*.

```
x=rnorm(10)  
Histogram(x, freq=FALSE,col.fill="royalblue", xlab="")
```

```
##  
##          Frequency table  
##  
##          ni  fi  Ni  Fi  
## [-1.27,-0.51)  4 0.4  4 0.4  
## [-0.51,0.24)  3 0.3  7 0.7  
## [0.24,0.99)   3 0.3 10 1.0
```



```
x=rnorm(20)
summary.freq.table(x,cont=TRUE)

##      Min. 1st Qu. Median Mean      Mode 3rd Qu. Max. Variance
## 1 -2.25   -0.86   0.01 -0.17 [-0.4,0.52)   0.52 1.44      4.62
##      Coefficient of Variation
## 1                      12.57

x=sample(1:5,size=20,replace=TRUE)
summary.freq.table(x,cont=FALSE)

##      Min. 1st Qu. Median Mean Mode 3rd Qu. Max. Variance Coefficient of Variation
## 1      1      1      3 2.6      1      3      5      1.74                      0.51

x=sample(c("yellow","red","blue","green"),size=20,replace=TRUE)
summary.freq.table(x,cont=FALSE)

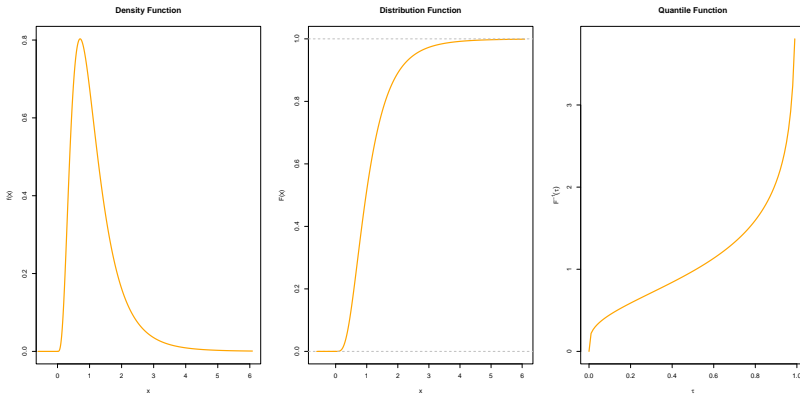
##      Mode
## 1 yellow

x=sample(c("high","small","medium"),size=20,replace=TRUE)
summary.freq.table(x,cont=FALSE,ord=c("small","medium","high"))

##      Min. 1st Qu. Median Mode 3rd Qu. Max.
## 1 small   small medium small   high high
```

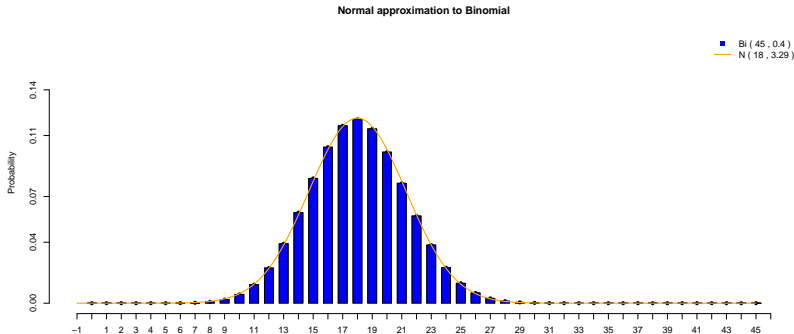


```
df1=10;df2=15  
plotFS(df1, df2, type = "b", col = "orange")
```



```
n=45; p=0.4
```

```
AproxBinomNorm(n, p, legend = TRUE, xlab = "", lwd = 2,  
ylab = "Probability", col.fill="blue", col.line="orange",  
main = "Normal approximation to Binomial")
```



```
dat=rnorm(20,mean=2,sd=1)
Mean.CI(dat, sigma=1, conf.level=0.95)

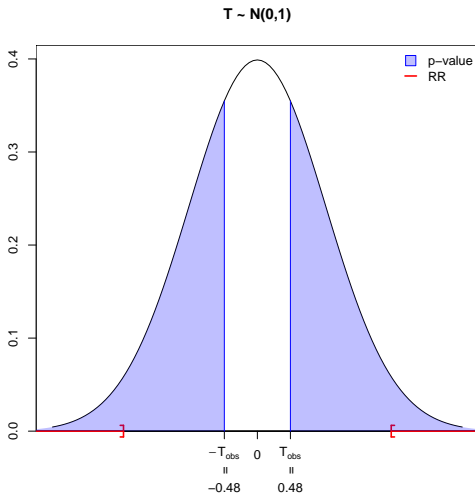
##
## Sample mean: 1.60665 units
##
## 95% confidence interval for the mean of a Normal population with known variance
##
## (bar.X - z_(1-alpha/2) * sigma/sqrt(n) , bar.X + z_(1-alpha/2) * sigma/sqrt(n))
##
## (1.16839 , 2.04491) units

dat=rnorm(20,mean=2,sd=1)
Mean.CI(mean(dat),sc=sd(dat),n=20,conf.level=0.95)

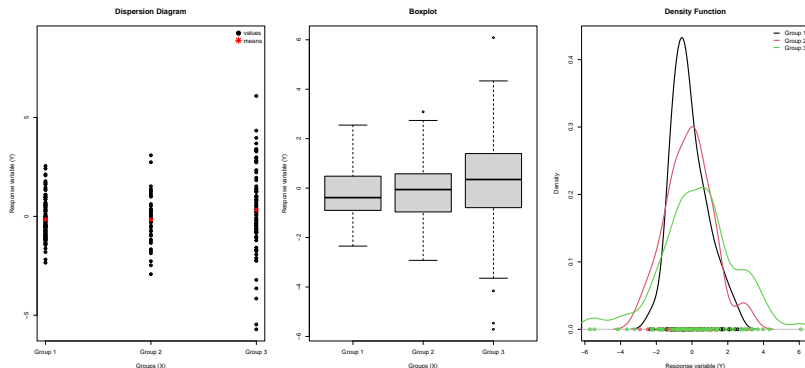
##
## Sample mean: 2.14547 units
##
## 95% confidence interval for the mean of a Normal population with unknown variance
##
## (bar.X - t_(n-1,1-alpha/2)*Sc/sqrt(n) , bar.X + t_(n-1,1-alpha/2)*Sc/sqrt(n))
##
## (1.79434 , 2.4966) units
```

```
x1 <- rbinom(1, 120, 0.6)
x2 <- rbinom(1, 100, 0.6)
diffproportion.test(x1 = x1, x2 = x2, n1 = 120, n2 = 100,
alternative="two.sided", alpha=0.05, plot=TRUE, lwd=1)

##
## Test for a proportion difference
##
## H0:  $p_1 - p_2 = 0$ 
## Ha:  $p_1 - p_2 \neq 0$ 
##  $T = (\hat{p}_1 - \hat{p}_2) / \sqrt{\hat{p}w * (1 - \hat{p}w) * (1/n_1 + 1/n_2)}$ ,
## with  $\hat{p}w = (n_1 * \hat{p}_1 + n_2 * \hat{p}_2) / (n_1 + n_2)$ 
##  $T \sim N(0,1)$ 
##  $\alpha = 0.05$ 
##  $T_{obs} = 0.48368$ 
##  $RR = (-\infty, -1.95996] \cup [1.95996, +\infty)$ 
## p-value = 0.62861
```

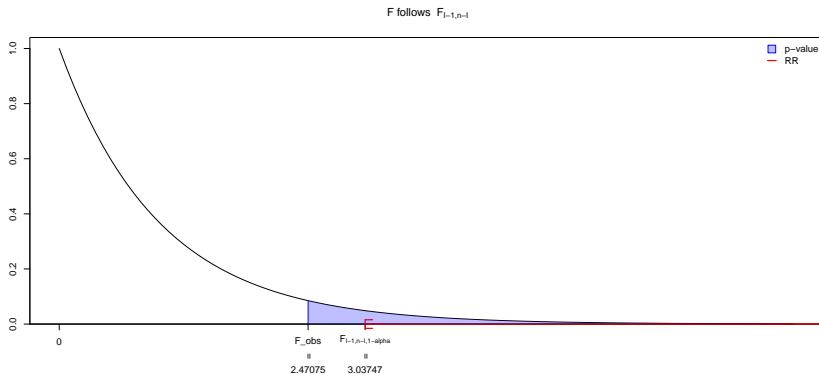


```
n1=100; n2=40; n3=80
set.seed(1234)
y=c(rnorm(n1,mean=0,sd=1),rnorm(n2,mean=0,sd=1.5),rnorm(n3,mean=0,sd=2))
x=c(rep("Group 1",n1),rep("Group 2",n2),rep("Group 3",n3))
plotAnova(x,y,xlab="Groups (X)",ylab="Response variable (Y)",
col.points="black",col.mean="red",pch.points=19,pch.mean=8,lwd=2)
```



```
alpha=0.05
Anova.test(x,y,alpha,plot=TRUE)

##
## Analysis of Variance
##
## H0: mu1 = mu2 = ... = muI
## Ha: mui != muj for some i,j
## F = Between Group Variation / Within-Group Variation ,
## where:
## Between Group Variation = (sum(i in 1:I){sum(j in 1:ni){(bar.yi-bar.y)^2}})/(I-1)
## Within-Group Variation = (sum(i in 1:I){sum(j in 1:ni){(yij-bar.yi)^2}})/(n-I)
## F follows F_{I-1,n-I}
## alpha = 0.05
## F_obs = 2.47075
## RR = [3.03747, +inf)
## p-value = 3.03747
## ANOVA table:
##           Between.Group Within.Group
## Variation      5.88413      2.38152
## DF             2.00000     217.00000
```



LearningStats: un paquete en R para a docencia

Sabela Varela-Rey, María Isabel Borrajo, Mercedes
Conde-Amboage e Alejandra López-Pérez