

Aprendendo Estatística con LearningStats

Elena Fernández, Mercedes Conde e Maribel Borrajo

- O recente incremento polo estudo e análise de datos converteu o coñecemento da **Estatística** nunha **competencia imprescindible** para os profesionais de diversos ámbitos.
- Para facilitar o proceso de ensino-aprendizaxe desenvolveuse en  o paquete LearningStats.



- Preséntase como unha **ferramenta didáctica**.

- Preséntase como unha ferramenta didáctica.
- Prioriza a intuitividade e sinxeleza de uso.

- Preséntase como unha **ferramenta didáctica**.
- Prioriza a **intuitividade** e **sinxeleza de uso**.
- Proporciona mensaxes de erro ou aviso facilmente comprensibles para o usuario.

- Preséntase como unha **ferramenta didáctica**.
- Prioriza a **intuitividade** e **sinxeleza de uso**.
- Proporciona mensaxes de erro ou aviso facilmente comprensibles para o usuario.
- É **versatil**: cubre as áreas principais da Estatística,

Estatística
Descritiva

Modelos de
Probabilidade

Inferencia
Estatística

Modelos de
Regresión

- Preséntase como unha **ferramenta didáctica**.
- Prioriza a **intuitividade** e **sinxeleza de uso**.
- Proporciona mensaxes de erro ou aviso facilmente comprensibles para o usuario.
- É **versatil**: cubre as áreas principais da Estatística,

Estatística
Descritiva

Modelos de
Probabilidade

Inferencia
Estatística

Modelos de
Regresión

- Contén conxuntos de datos que permiten ilustrar o uso na práctica dos distintos procedementos e ferramentas.

Obxectivo

Avaliar se os datos dunha mostra se axustan a unha distribución de probabilidade teórica.

Obxectivo

Avaliar se os datos dunha mostra se axustan a unha distribución de probabilidade teórica.

- Ferramentas gráficas.

Obxectivo

Avaliar se os datos dunha mostra se axustan a unha distribución de probabilidade teórica.

- Ferramentas gráficas.
- Contrastes de hipóteses

$$H_0 : F = F_0$$

$$H_a : F \neq F_0$$

$$H_0 : F \in \{F_\theta : \theta \in \Theta\}$$

$$H_a : F \notin \{F_\theta : \theta \in \Theta\}$$

- Test de Kolmogorov-Smirnov: emprégase só con distribucións continuas.
 - Test de Lilliefors: contrasta normalidade.

$$D_n = \sup_{x \in \mathbb{R}} |F_n(x) - F_\theta(x)|$$

$$D_n = \sup_{x \in \mathbb{R}} |F_n(x) - F_{\hat{\theta}}(x)|$$

- Test de Kolmogorov-Smirnov: emprégase só con distribucións continuas.
 - Test de Lilliefors: contrasta normalidade.

$$D_n = \sup_{x \in \mathbb{R}} |F_n(x) - F_\theta(x)| \qquad D_n = \sup_{x \in \mathbb{R}} |F_n(x) - F_{\hat{\theta}}(x)|$$

- Test chi-cadrado: contrasta a bondade de axute tanto en distribucións continuas como categóricas. Baséase na comparación de frecuencias observadas e esperadas baixo H_0 .

$$T = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}$$

- `plotDistribution`. Elabora representacións gráficas que permiten comparar as funcións de distribución e de densidade teóricas e estimadas. O usuario obtén unha primeira intuición acerca do contraste.

- `plotDistribution`. Elabora representacións gráficas que permiten comparar as funcións de distribución e de densidade teóricas e estimadas. O usuario obtén unha primeira intuición acerca do contraste.
- `gof.ks.test`. Realiza o test de Kolmogorov-Smirnov e se contempla o caso onde os parámetros da distribución teórica precisen ser estimados, seguindo a idea do test de Lilliefors.
- `gof.chisq.test`. Realiza o test chi-cadrado de bondade de axuste contemplando tamén a estimación de parámetros.

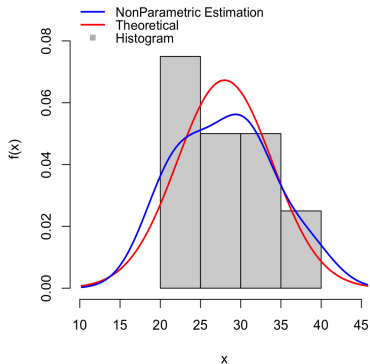
```
1  datos <- c(38,30,24,27,20,31,22,32)
2  plotDistribution(x = datos, dist = "norm", type = "b",
   col.est = "blue", col.theo = "red")
```

- x: mostra de datos.
- dist: nome da distribución teórica.
- type: caracter que identifica o tipo de representación a realizar.
- col.est e col.theo: cores asociadas ás funcións estimadas e teóricas.
- theta: vector de parámetros.

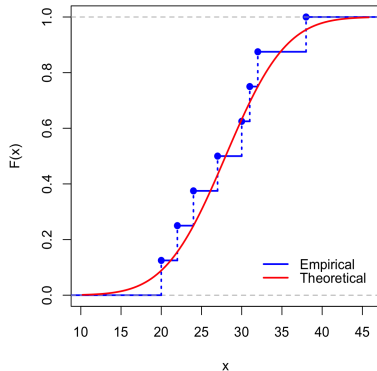
```
1  datos <- c(38,30,24,27,20,31,22,32)
2  plotDistribution(x = datos, dist = "norm", type = "b",
3    col.est = "blue", col.theo = "red")
4
5  Distribution: 'norm'
6  Estimated parameters: 28, 5.9281
7
8  Warning message:
9  In plotDistribution(x = datos, dist = "norm", type = "b",
   col.est = "blue", :
   As 'theta' was no provided, parameters have been estimated
   from the sample.
```

Exemplo 1

Density Function



Distribution Function



```
1  gof.ks.test(x = datos, dist = "norm", plot = TRUE,  
    alpha = 0.05, lwd = 1, B = 2000)
```

- x: mostra de datos.
- dist: nome da distribución teórica continua.
- plot: caracter que determina se se realiza ou non a representación gráfica.
- alpha: nivel de significación do contraste.
- lwd: valor numérico que determina o grosor dos trazos na representación gráfica.
- B: número de repeticións da simulación Monte Carlo.
- theta: vector de parámetros.

```
1  gof.ks.test(x = datos, dist = "norm", plot = TRUE,  
2    alpha = 0.05, lwd = 1, B = 2000)
```

```
3    Kolmogorov-Smirnov test (unknown parameters)
```

```
4  H0: data follows 'norm' distribution
```

```
5  Ha: data does not follow 'norm' distribution
```

```
6  
7  Estimated parameters:
```

```
8  [1] 28.0000  5.9281
```

```
9  
10 data:  datos
```

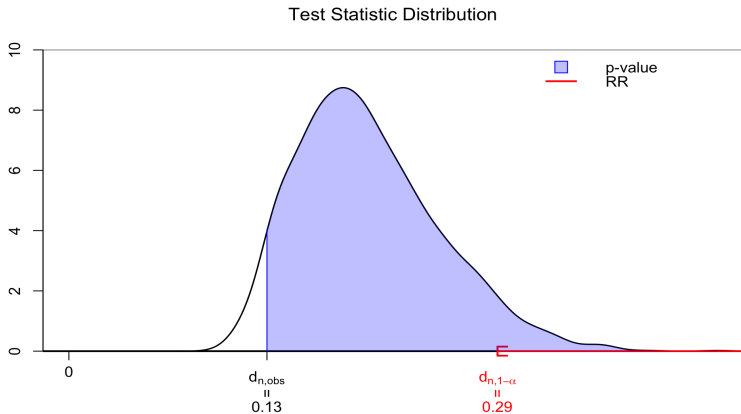
```
11 Dn = sup|Fn(x)-F0(x)|
```

```
12 alpha =  0.05
```

```
13 RR = (0.28582, +Inf)
```

```
14 d_n,obs = 0.13208, n = 8, p-value = 0.959
```

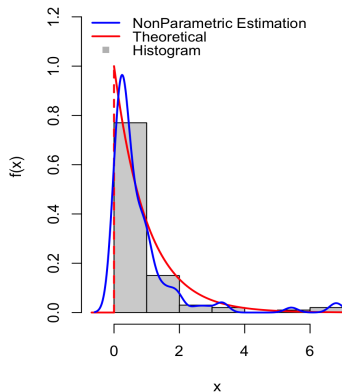
Exemplo 1



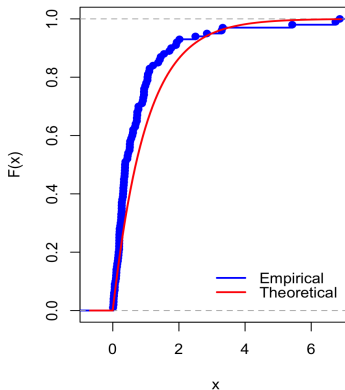
Exemplo 2

```
1 datos2 <- c(0.208, 0.039, 0.330, ..., 0.466, 0.135, 0.273)  
2 plotDistribution(x = datos2, "exp", theta = 1, type = "b")
```

Density Function



Distribution Function

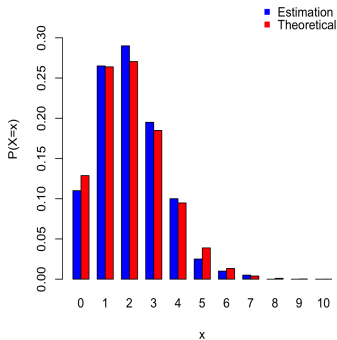


```
1  gof.ks.test(x = datos2, "exp", theta = 1)
2
3      Kolmogorov-Smirnov test (known parameters)
4
5  H0: data come from a Exp(1) distribution
6  Ha: data do not come from a Exp(1) distribution
7
8  data:  datos2
9  Dn = sup |Fn(x) - F0(x)|
10 alpha = 0.05
11 RR = (0.13358, +Inf)
    d_n, obs = 0.19187, n = 100, p-value < 2.2e-16
```

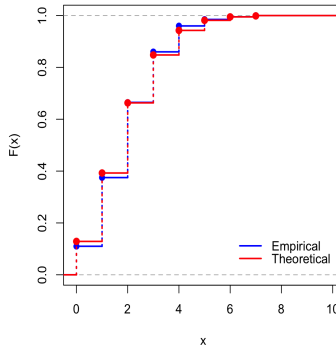
Exemplo 3

```
1 datos3 <- c(22, 53, 58, 39, 20, 5, 2, 1)
2 plotDistribution(x = c(rep(0:7,datos3)), dist = "pois"  
  , type = "b")
```

Probability Mass Function



Distribution Function



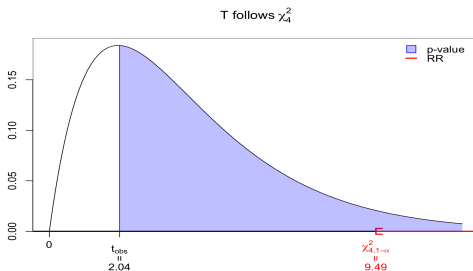
```
1  gof.chisq.test(x = datos3, dist = "pois", xnames = 0:7,  
    alpha = 0.05, plot = TRUE, lwd = 1)
```

- x: mostra de datos.
- dist: nome da distribución teórica.
- xnames: categorías nas que se divide a mostra.
- alpha: nivel de significación do contraste.
- plot: caracter que determina se se realiza ou non a representación gráfica.
- lwd: valor numérico que determina o grosor dos trazos na representación gráfica.
- theta: vector de parámetros.

```
1  gof.chisq.test(x = datos3, dist = "pois", xnames = 0:7,  
2      alpha = 0.05, plot = TRUE, lwd = 1)  
3  
4  Chi-squared test for goodness-of-fit  
5  
6  H0: data follows 'pois' distribution  
7  Ha: data does not follow 'pois' distribution  
8  
9  Estimated parameters:  
10 [1] 2.05  
11  
12 Chi.squared = sum {(Oi - Ei)2 / Ei}  
13 Chi.squared follows X2{k-1-q}  
14 alpha = 0.05  
15 t_obs = 2.03967  
16 RR = (9.48773, +Inf)  
17 p-value = 0.72846
```

Exemplo 3

```
16 Observed frequencies, Oi:
17 {0}      {1}      {2}      {3}      {4} {5,6,7,...}
18 22       53       58       39       20      8
19
20 Expected frequencies, Ei:
21 {0}      {1}      {2}      {3}      {4} {5,6,7,...}
22 25.75     52.78     54.10     36.97     18.95     11.46
23 Warning message:
24 In gof.chisq.test(x = datos3, dist = "pois", xnames = 0:7, alpha = 0.05, :
25 Grouping has been carried out as some Ei was less than 5
```



```
1  gof.ks.test(x=seq(1,2,by=0.01), dist="pois")
2
3  Error in gof.ks.test(x = seq(1, 2, by = 0.01), dist = "pois"):  
4  'dist' must be the name of a continuous probability distribution  
   . The allowed ones are: 'norm', 'exp', 'unif', 'gamma' and '  
   beta'.
```

```
1  gof.ks.test(x=seq(1,2,by=0.01), dist="pois")
2
3  Error in gof.ks.test(x = seq(1, 2, by = 0.01), dist = "pois"):  
4  'dist' must be the name of a continuous probability distribution  
   . The allowed ones are: 'norm', 'exp', 'unif', 'gamma' and '  
   beta'.
```

```
1  gof.chisq.test(datos_exp,"exp",alpha=3)
2
3  Error in gof.chisq.test(datos_exp, "exp", alpha = 3):'alpha'  
   must be a single number in (0,1)
```

```
1  gof.ks.test(x=seq(1,2,by=0.01), dist="pois")
2
3  Error in gof.ks.test(x = seq(1, 2, by = 0.01), dist = "pois"):  
4  'dist' must be the name of a continuous probability distribution  
   . The allowed ones are: 'norm', 'exp', 'unif', 'gamma' and '  
   beta'.
```

```
1  gof.chisq.test(datos_exp, "exp", alpha=3)
2
3  Error in gof.chisq.test(datos_exp, "exp", alpha = 3): 'alpha'  
   must be a single number in (0,1)
```

```
1  plotDistribution(pois_data, "pois", type="h")
2
3  Error in plotDistribution(pois_data, "pois", type = "h"): 'type'  
   must be 'dis'(distribution), 'den'(density) or 'b'(both)
```

- Ferramenta didáctica para o ensino e aprendizaxe da Estatística.
- Facilita a comprensión mediante exemplos e representacións gráficas.
- Promove unha aprendizaxe activa, intuitiva e consciente.

Aprendendo Estatística con LearningStats

Elena Fernández, Mercedes Conde e Maribel Borrajo