Estatística Não Paramétrica

Teste de aderência Qui-quadrado e Kolmogorov-Smirnof (KS)

Danielle blabla blabla e Luiz Fernando Coelho Passos

Teste de Aderência

Teste Qui-Quadrado

. Explicação .

Exemplo 1 - v.a. discretas

Um estudo sobre a distribuição dos acidentes de trabalho numa indústria nos cincos dias da semana revelou que, em 150 acidentes: segunda: 32, terça: 40, quarta: 20, quinta: 25 e sexta: 33. O objetivo é testar a hipótese que os acidentes ocorrem com igual frequência nos cinco dias da semana (considere $\alpha = 5\%$).

Hipóteses: $H_0 : p_1 = p_2 = p_3 = p_4 = p_5 = 1/5$ versus $H_1 : p_j \neq 1/5$, para pelo menos um j

```
Oi <- c(seg=32, ter=40, qua=20, qui=25, sex=33)
Ei <- sum(Oi)*1/length(Oi)                      # esperados sob H_0
X2 <- sum((Oi-Ei)^2/Ei)                         # estatística do teste
nu <- length(Oi)-1                              # graus de liberdade
pchisq(X2, df=nu, lower.tail=FALSE)             # p-valor do teste
```

```
## [1] 0.09405103
```

Usando a função `chisq.test()`, temos:

```
chisq.test(Oi)
```

```
##
##  Chi-squared test for given probabilities
##
## data:  Oi
## X-squared = 7.9333, df = 4, p-value = 0.09405
```

Conclusão

O p-valor obtido foi de 0.09405. Assim, não rejeitamos H_0 ao nível de significância de 5%, ou seja, podemos considerar que os acidentes ocorrem com igual frequência nos 5 dias da semana.

Exemplo 2 - v.a. contínuas

Considere os dados abaixo, que supostamente são uma amostra de tamanho n = 30 de uma distribuição normal, de média $\mu = 10$ e variância $\sigma^2 = 25$. Os dados já estão ordenados:

```
df <- 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,
```

```r
        11.72,  12.17,  12.61,  12.98,  13.03,  13.16,
        14.11,  14.60,  14.64,  14.75,  16.68,  22.14)
```

Categorizaremos os dados do exemplo da seguinte forma:

```r
Oi <- c()
Oi[1]<-sum(df < 6.63)
Oi[2]<-sum(df > 6.63 & df < 10)
Oi[3]<-sum(df > 10 & df < 13.37)
Oi[4]<-sum(df > 13.37)

Ei <- sum(Oi)*1/length(Oi)                        # esperados sob H_0
X2 <- sum((Oi-Ei)^2/Ei)                           # estatística do teste

# Tabela
cbind(
  rbind(Oi, Ei),
  c(sum(Oi), Ei*4)
) %>% `colnames<-`(c("(-Inf;6.63]", "(6.63;10]", "(10;13.37]", "(13.37;+Inf)", "Total"))
```

```
##    (-Inf;6.63] (6.63;10] (10;13.37] (13.37;+Inf) Total
## Oi         6.0       9.0        9.0          6.0    30
## Ei         7.5       7.5        7.5          7.5    30
```

```r
nu <- length(Oi)-1                                # graus de liberdade
pchisq(X2, df=nu, lower.tail=FALSE)               # p-valor do teste
```

```
## [1] 0.7530043
```

Usando a função `chisq.test()`, temos:

```r
chisq.test(Oi)
```

```
##
##  Chi-squared test for given probabilities
##
## data:  Oi
## X-squared = 1.2, df = 3, p-value = 0.753
```

Conclusão

O p-valor obtido foi de 0.753. Assim, não rejeitamos H_0 ao nível de significância de 5%, ou seja, podemos considerar que temos uma amostra de uma normal com média 10 e variância 25.

Teste Kolmogorov-Smirnof

. Explicação .

Exemplo

Retomemos o `Exemplo 2`, onde queríamos testar se 30 valores observados provinham de uma distribuição normal, com média 10 e variância 25.

A hipótese a ser testada pode ser escrita na forma

$H_0 : F(x) = F_0(x), \forall x$

$H_1 : F(x) = F_0(x), \text{ para algum } x,$

onde $F_0(x)$ é a função de densidade acumulada da v.a. $X \sim N(10, 25)$.

O teste de Kolmogorov-Smirnov (KS) é calculado, no `R`, através da função `ks.test`. Assim, para testar se os valores são aderentes à distribuição normal:

```r
ks.test(df, y="pnorm", mean=10, sd=5)
```

```
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  df
## D = 0.11633, p-value = 0.769
## alternative hypothesis: two-sided
```

Conclusão

O p-valor obtido foi de 0.769. Assim, não rejeitamos H_0 ao nível de significância de 5%, ou seja, podemos considerar que temos uma amostra de uma normal com média 10 e variância 25.

Referência

- MORETTIN, Pedro A.; BUSSAB, Wilton O. **Estatística Básica**. São Paulo: Saraiva, 2010. cap. 14, 399-419.