

Lista 6: Fundamentos Estatísticos para Ciência dos Dados

Ricardo Pagoto Marinho

10 de abril de 2018

- 1) A tabela a seguir é dada no problema:

k	0	1	2	3	4	5	6	7	8	9
Obs	60	62	67	68	64	56	62	44	58	67
Esp	??	??	??	??	??	??	??	??	??	??

Com ela, temos que preencher a última linha, ou seja, a da quantidade esperada de cada dígito na expansão decimal do número π . Como foi dito que os dígitos são escolhidos de forma aleatória, cada um terá a probabilidade de 10% de aparecer, ou seja, é esperado que cada dígito apareça 60.8 vezes na expansão. Logo, a última linha da tabela é completada como segue:

k	0	1	2	3	4	5	6	7	8	9
Obs	60	62	67	68	64	56	62	44	58	67
Esp	60.8	60.8	60.8	60.8	60.8	60.8	60.8	60.8	60.8	60.8

Fazendo os cálculos para o teste qui-quadrado em R, foi obtido o seguinte resultado:

```
obs<-c(60,62,67,68,64,56,62,44,58,67)
```

```
esp<-rep(60.8,10)
```

```
qqquad<-sum((obs-esp)^2/esp)
```

```
qqquad
```

```
7.493421
```

Para rejeitar ou aceitar essa distribuição, devemos olhar a tabela do teste com Grau de Liberdade 9, já que possuímos 10 classes (os 10 dígitos). Com base na seguinte tabela (encontrada em <http://www.cultura.ufpa.br/dicas/biome/biotaqui.htm>):

G.L.	χ^2
1	3.841
2	5.991
3	7.815
4	9.488
5	11.070
6	12.592
7	14.067
8	15.507
9	16.919

é possível observar que para um Grau de Liberdade 9, devemos rejeitar a hipótese caso o valor do teste seja maior do que 16.919. Como o

resultado deu 7.493421, podemos admitir que a hipótese na qual os 608 primeiro dígitos do número π sigam uma distribuição uniforme.

- 2)

```
x<-rnorm(50,5,3)
ks.test(x,"pnorm",5,3)
One-sample Kolmogorov-Smirnov test
data: x
D = 0.12562, p-value = 0.3777
alternative hypothesis: two-sided
sqrt(50)*0.12562
0.8882675
```

Como $\sqrt{n}D_n < 1.36$, não podemos ter rejeitar os dados com certeza. Agora utilizando $\mu = 5$ e $\sigma = 2$, é esperado que os dados tenham que ser rejeitados, já que é outra distribuição. Os resultados obtidos são os seguintes:

```
ks.test(x,"pnorm",5,2)
One-sample Kolmogorov-Smirnov test
data: x
D = 0.19948, p-value = 0.03211
alternative hypothesis: two-sided
sqrt(50)*0.19948
1.410537
```

Agora, $\sqrt{n}D_n > 1.36$, como esperado, os dados deverão ser rejeitados.

- 3)

```
# funcao para calcular o teste de Kolmogorov:
#      Nome: kol
#      Entradas: vec-> vetor com os valores a serem testados
#                  mu-> media utilizada para o teste
#                  sigma-> desvio padrao utilizado para o
#                          teste
#      femp-> vetor de valores empiricos a partir de 'vec'
#                  femp=#{x_i<=x}/n
#      xc-> vetor para calculo da cdf da distribuicao criada
#                  dentro da funcao
#      x->vetor de valores de uma distribuicao normal criado
#                  a partir das entradas mu e sigma
kol=function(vec,mu=mean(vec),sigma=sqrt(var(vec))){
```

```

femp<-0
xc<-0
j<-1
x<-rnorm( length( vec ),mu,sigma )
x<-sort( x )
for( i in x ){
    femp[ j ]<-sum( vec<=x[ j ] ) / length( vec )
    xc[ j ]<-sum( x<=x[ j ] ) / length( vec )
    j<-j+1
}
f<-abs( femp-xc )
return( sqrt( length( vec ) ) * max( f ) )
}

```