

EXEMPLO 5 do Cap6-parte 1

Exemplo 5

Na tabela seguinte apresentam-se os tempos de falha (em horas) de uma determinada máquina:

1476	300	98	221	157
182	499	552	1563	36
246	442	20	796	31
47	438	400	279	247
210	284	553	767	1297
214	428	597	2025	185
467	401	210	289	1024

Será que tais observações foram extraídas de uma população com distribuição Exponencial, isto é, será de admitir que os tempos de falha seguem uma distribuição Exponencial? Teste a hipótese referida considerando um nível de significância de 10%.

Pretende-se verificar se as observações da amostra dada foram extraídas de uma população Exponencial. Na seguinte tabela temos algumas orientações.

• TESTES DE AJUSTAMENTO

Condições	Teste de Ajustamento	R
Distribuição Discreta ou Distribuição Contínua (com recurso a classes)	Qui-Quadrado	chisq.test()
Distribuição Contínua (completamente especificada)	Kolmogorov-Smirnov	ks.test()
Normal e $n \geq 50$	Lilliefors	lillie.test() (library(nortest))
Normal e $n < 50$	Shapiro Wilks	shapiro.test()

*Não conhecemos o parâmetro θ , logo não se pode usar o teste de Kolmogorov-Smirnov (K-S), pois a distribuição não se pode considerar completamente especificada. Por outro lado, Exponencial não é distribuição Normal, logo não se podem usar nem o teste Lilliefors nem o teste Shapiro-Wilks. Tem que ser usado o **Teste Qui-quadrado**.*

Hipótese a ser testada

Seja X a variável aleatória que representa os tempos de falha em horas

$$H_0 : X \sim \text{Exp}(\theta) \quad \text{vs} \quad H_1 : X \not\sim \text{Exp}(\theta)$$

Dados

- Total de dados: $n = 35$
- É necessário estimar θ , como $E[X] = \theta$, então uma estimativa para θ é a média da amostra

Os dados podem ser colocados no R, da seguinte forma:

```
# Exemplo 5

# H0: X segue uma distribuição Exponencial
# contra
# H1: X Não segue uma distribuição Exponencial

amostra <- c(1476,300,98,221,157,
             182,499,552,1563,36,
             246,442,20,796,31,
             47,438,400,279,247,
             210,284,553,767,1297,
             214,428,597,2025,185,
             467,401,210,289,1024)
```

Cálculo do estimador do parâmetro θ :

```
> mean(amostra)
[1] 485.1714
```

Assim a estimativa do parâmetro $\theta = 485.1714$

- Distribuição Exponencial: $X \sim \text{Exp}(485.1714)$
- Número de parâmetros estimados: $r = 1$

Para se usar o teste do Qui-quadrado com dados contínuos é necessário construir classes.

- Como a variável é contínua é necessário definir classes \rightarrow Regra de Sturges

Recorrendo ao R, vem que:

```
#####
##### construir classes #####

# n = dimensão da amostra
(n <- length(amostra))

# número de classes ou de linhas da tabela de frequências
(k<-trunc(1+log(n)/log(2)))

(h <- (max(amostra)-min(amostra))/k) # amplitude das classes

# mínimo e máximo das classes
(valor.min <- min(amostra))
(valor.max <- valor.min + h*k)

#extremos das classes
(cortes <- seq(valor.min, valor.max, by=h))
round(cortes,2)

# intervalos abertos à esquerda e fechados à direita
# como o mínimo dos dados é igual ao primeiro valor da primeira classe
# a primeira classe tem de ser fechada nos dois lados
(classes <- cut(amostra, breaks=cortes, right=TRUE, include.lowest=TRUE))
```

Assim

- Definiram-se 6 classes, cada uma com amplitude 334.2 (considerou-se a primeira classe a iniciar no mínimo observado, intervalos abertos à esquerda e fechados à direita).

Em seguida, precisamos saber o número de elementos de cada classe, ou seja, os valores observados em cada classe (O_i). Estes valores e a tabela podem ser calculados da seguinte forma:

```
#####
## Determinar os  $O_i$  de cada classe
# frequências absolutas = frequências Observadas
( $O_i <- \text{table}(\text{classes})$ )

# é necessário estimar o tetha -> média amostral
(estimativa <- mean(amostra))

#####
## construir a tabela

r <- 1 # estimou-se um parâmetro
(gl <- k-1-r) # graus de liberdade da distribuição Qui-Quadrado
```

Para calcular os E_i é necessário saber os valores de p_i que são a probabilidade de cada uma das classes, tendo em conta uma exponencial com média 485.1714. No entanto a primeira e última classes têm de ser ajustadas.

Note-se que, por exemplo, que o cálculo do p_i da 1ª classe é $F(354.2)$ no R:

```
> pexp(354.2,1/estimativa)
[1] 0.5181157
```

Domínio: Classes $]x_i, x_{i+1}]$	Frequências Observadas $O_i = n_i$	Probabilidade $p_i = P(x_i < X \leq x_{i+1})$
$] - \infty, 354.2]^{(*)}$	18	$P(X \leq 354.2) = F(354.2) = 0.5181$
$]354.2, 688.4]$	10	$P(354.2 < X \leq 688.4) = F(688.4) - F(354.2) = 0.2399$
$]688.4, 1022.6]$	2	$P(688.4 < X \leq 1022.6) = F(1022.6) - F(688.4) = 0.1205$
$]1022.6, 1356.8]$	2	$P(1022.6 < X \leq 1356.8) = F(1356.8) - F(1022.6) = 0.0605$
$]1356.8, 1691]$	2	$P(1356.8 < X \leq 1691) = F(1691) - F(1356.8) = 0.0304$
$]1691, +\infty[^{(**)}$	1	$P(X > 1691) = 1 - P(X \leq 1691) = 1 - F(1691) = 0.0306$
	$n = 35$	1

(*) com base na amostra seria $[20, 354.2]$, mas a variável pode assumir qualquer valor em \mathbb{R}

(**) com base na amostra seria $]1691, 2025.2]$, mas a variável pode assumir qualquer valor em \mathbb{R}

Em R, um código para calcular estas probabilidades de forma a poder ser utilizada em qualquer situação de uma exponencial é:

```
# probabilidades necessárias para as frequências esperadas
(pi =pexp(cortes[2:(k+1)],1/estimativa))
for(i in 2:k){
  pi[i] <- pexp(cortes[i+1],1/estimativa)-pexp(cortes[i],1/estimativa)
}
pi[k] <- 1-pexp(cortes[k],1/estimativa) # corrigir a última probabilidade
round(pi,4)
sum(pi)
```

Fazer o teste do Qui-quadrado

```
#teste de ajustamento do Qui-Quadrado
chisq.test(x=0i,p=pi)

#dar um nome para poder ver os diversos campos
exemplo.5 <- chisq.test(x=0i,p=pi)
```

Antes de se olhar aos valores do teste, é preciso verificar as condições de aplicabilidade:

```
# verificar as regras recomendadas

# dimensão da amostra maior que 30
if(n>30){
  print("respeita a regra")
}else{
  print("a amostra é pequena")
}

# todas as frequências Esperadas >= 1
if(length(which(exemplo.4$expected<1))>0){
  print("juntar linhas da tabela de frequências")
}else{
  print("respeita a regra")
}

# não há mais de 20% das frequências Esperadas < 5
if(length(which(exemplo.4$expected<5))>(k*0.2)){
  print("juntar linhas da tabela de frequências")
}else{
  print("respeita a regra")
}
```

Como a última regra indica que se deve juntar linhas, então analisa-se as frequências esperadas, para sabermos quais linhas juntar.

```
> exemplo.5$expected
      [20,354]      (354,688]      (688,1.02e+03] (1.02e+03,1.36e+03]
      18.132891      8.396474      4.216696      2.117618
(1.36e+03,1.69e+03] (1.69e+03,2.02e+03]
      1.063464      1.072857
```

Domínio: Classes $]x_i, x_{i+1}]$	Frequências Observadas $O_i = n_i$	Probabilidade $p_i = P(x_i < X \leq x_{i+1})$	Frequências Esperadas $E_i = n \times p_i$
$] - \infty, 354.2]$	18	0.5181	$35 \times 0.5181 = 18.1341$
$]354.2, 688.4]$	10	0.2399	$35 \times 0.2399 = 8.3965$
$]688.4, 1022.6]$	2	0.1205	$35 \times 0.1205 = 4.2164^{***}$
$]1022.6, 1356.8]$	2	0.0605	$35 \times 0.0605 = 2.1173^{***}$
$]1356.8, 1691]$	2	0.0304	$35 \times 0.0304 = 1.0632^{***}$
$]1691, +\infty[$	1	0.0306	$35 \times 0.0306 = 1.0725^{***}$
	$n = 35$	1	$n = 35$

(***) Falham as condições de aplicabilidade do teste, mais de 20% ($6 \text{ classes} \times 20\% = 1.2$) das frequências esperadas são inferiores a 5, iremos agregar classes adjacentes.

Se agrupássemos as três últimas ainda não daria para cumprir a regra, assim vão-se juntar as últimas 4.

```
#####
#####
```

```
# juntar linhas, as quatro últimas linhas
#(são as que têm as frequências que não respeitam uma das regras)

# perde 3 linhas
k2 <- k-3

(gl2 <- k2-1-r)

#limites das classes (atenção que o cortes[k+1], fica com k para que
# a última classe vá até ao máximo da amostra)
(cortes2<- c(cortes[1:k2],cortes[k+1]))

#definir as classes -> cut()
(classes2 <- cut(amostra, breaks=cortes2, right=TRUE,
                include.lowest=TRUE, dig.lab = 5))

# frequências absolutas = frequências Observadas
(Oi2<-table(classes2))

# probabilidades necessárias para as frequências esperadas
(pi2 =pexp(cortes2[2:(k2+1)],1/estimativa))
for(i in 2:k2){
  pi2[i] <- pexp(cortes2[i+1],1/estimativa)-pexp(cortes2[i],1/estimativa)
}
pi2[k2] <- 1-pexp(cortes2[k2],1/estimativa)
round(pi2,4)
sum(pi2)
```


Volta-se a fazer o teste

```
#teste de ajustamento do Qui-Quadrado
chisq.test(x=0i2,p=pi2)

exemplo.51 <- chisq.test(x=0i2,p=pi2)

#####
#####

# verificar as regras recomendadas
# verificar só a que falhou anteriormente

# não há mais de 20% das frequências Esperadas < 5
if(length(which(exemplo.51$expected<5))>(k*0.2)){
  print("juntar linhas da tabela de frequências")
}else{
  print("respeita a regra")
}
```

Se construíssemos a tabela ficaria:

Domínio: Classes $]x_i, x_{i+1}]$	Frequências Observadas $O_i = n_i$	Probabilidade $p_i = P(x_i < X \leq x_{i+1})$	Frequências Esperadas $E_i = n \times p_i$
$] - \infty, 354.2]$	18	0.5181	18.1341
$]354.2, 688.4]$	10	0.2399	8.3965
$]688.4, +\infty[$	7	$P(X \geq 688.4) = 0.2420$	8.4695
	$n = 35$	1	$n = 35$

O resultado do teste foi:

```
> chisq.test(x=0i2,p=pi2)
```

Chi-squared test for given probabilities

data: 0i2

X-squared = 0.56253, df = ~~2~~, p-value = 0.~~7~~48

Observação: como foi necessário estimar parâmetros, só o valor da estatística de teste está certo.

DECISÃO:

- Pelo p-value:

```
# os graus de liberdade e o valor-p estão errados pois não tem em consideração
# que os parâmetros tiveram de ser estimados (r=1)
```

```
# valor-p correto
valorp2 <- 1-pchisq(exemplo.51$statistic, gl2)
valorp2
```

Como p - value é 0.4532, não é $\leq \alpha$, pois $\alpha = 0.1$, então não se rejeita H_0 .

- **Pela Região crítica:**

```
> #cálculo do quantil de probabilidade da a Região crítica  
> qchisq(0.90, gl2)  
[1] 2.705543
```

$$RC = [x_{1-\alpha; (k-1-r)}^2, +\infty[= [x_{0.90; (1)}^2, +\infty[= [2.71, +\infty[$$

Como $Q_{obs} = 0.5625 \notin RC$ então não se rejeita a hipótese H_0

Conclusão: Com base na amostra e ao nível de significância de 10%, conclui-se que há evidência estatística que os tempos de falha seguem uma distribuição Exponencial.