

CC0291 - Estatística Não Paramétrica

Testes Para Ajuste de Distribuições - 15/09/2020

Profs. Juvêncio Nobre, Maurício Mota e Gualberto Agamez

1 Introdução

Este trabalho é inteiramente calcado das notas de aula do professor Juvêncio Nobre. Vamos apresentar os principais testes usados para ajustes de uma distribuição aos dados. Estes testes são: Kolmogorov, Lilliefors, Shapiro-Wilk, Anderson-Darling, Cramer-Von Mises e Jarque-Bera para uma amostra; e Sminorv e Cramer-Von Mises para duas amostras. Usaremos o software R. Inicialmente apresentaremos a Função de Distribuição Empírica que é a ferramenta principal desses testes.

2 Teste de Lilliefors Para Normalidade

Agora teremos interesse em testar se a amostra provém de uma família de distribuições, por exemplo $N(\mu, \sigma^2)$, $Gama(\beta, \lambda)$ mas sem conhecer seus parâmetros. Devemos estimar os parâmetros usando a própria amostra e, em geral, utiliza-se a mesma estatística de teste.

Aparece uma pergunta importante: Será que podemos usar os mesmos quantis do caso anterior quando conhecíamos os parâmetros?

A resposta é **não**, pois existe uma variabilidade inerente ao(s) estimador (es) usado(s), isto é, uma variabilidade maior do que na situação anterior. Na Estatística Paramétrica temos o caso do teste da média de uma população normal quando a variância é conhecida ou não.

Vale lembrar que existem tabelas específicas para algumas classes de distribuições.

A primeira modificação do teste de Kolmogorov (e sem dúvida a mais famosa) foi realizada com o objetivo de se testar especificamente a normalidade. Isto é, a hipótese nula assume que a amostra é retirada de uma população normal sem especificar os parâmetros μ e σ^2 . Este teste foi apresentado por Lilliefors (1967). Uma característica interessante deste teste, com relação ao teste de Kolmogorov é a necessidade do uso de métodos computacionalmente intensivos para estimar os quantis da distribuição exata da estatística do teste. Vamos formalizar o teste.

Seja $\sim F(x)$ e X_1, X_2, \dots, X_n uma amostra aleatória de X .

Queremos testar:

$$H_0 : X \sim N(\mu, \sigma^2),$$

versus

$$H_1 : X \text{ não possui distribuição } \sim N(\mu, \sigma^2).$$

Dada a amostra obtemos os estimadores de máxima verossimilhança restrita de μ e σ^2 , sob H_0 , isto é,

$$\hat{\mu} = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i,$$

e

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2.$$

Depois disso obtemos os valores amostrais normalizados através de:

$$Z_i = \frac{X_i - \bar{X}}{S}, \quad i = 1, 2, \dots, n.$$

Note que:

Observação 1 Perceba que

$$\hat{F}_n(x_i) = \hat{F}_n(z_i), \quad i = 1, 2, \dots, n.$$

Observação 2 O teste é realizado usando as observações padronizadas, ao invés das observações originais.

Será que Z_1, Z_2, \dots, Z_n são independentes e identicamente distribuídas?

Vamos analisar inicialmente a independência:

$$\begin{aligned} \text{Cov}(Z_1, Z_2) &= \text{Cov}\left(\frac{X_1 - \bar{X}}{S}, \frac{X_2 - \bar{X}}{S}\right) \\ &= \text{Cov}\left(\frac{X_1}{S}, \frac{X_2}{S}\right) + \text{Var}\left(\frac{\bar{X}}{S}\right) - \text{Cov}\left(\frac{X_1}{S}, \frac{\bar{X}}{S}\right) - \text{Cov}\left(\frac{X_2}{S}, \frac{\bar{X}}{S}\right) \\ &\neq 0. \end{aligned}$$

Qual a distribuição de

$$Z_i = \frac{X_i - \bar{X}}{S}$$

no caso de Normalidade?

Sejam $z_{(1)}, z_{(2)}, \dots, z_{(n)}$ as estatísticas de ordem da nossa amostra em estudo.

A função de distribuição empírica é definida como:

Segundo o Portal Action a função de distribuição empírica é descontínua e a função de distribuição hipotética é contínua (normal).

Vamos considerar duas outras estatísticas para calcularmos a estatística de Lilliefors.

Essas estatísticas medem as distâncias (verticais) entre os gráficos das duas funções, teórica e empírica, nos pontos $z_{(i)}$ e $z_{(i-1)}$, com $F_n(z_{(0)}) = 0$.

Para $i = 1, \dots, n$

$$z_i = \frac{y_i - \bar{X}}{S}.$$

$$L^+ = \sup_{z_{(i)}} |\Phi(z_{(i)}) - F_n(z_{(i)})|$$

$$L^- = \sup_{z_{(i)}} |\Phi(z_{(i)}) - F_n(z_{(i-1)})|$$

Com isso, podemos utilizar como estatística de teste

$$L_n = \max(L^+, L^-)$$

Se L_n é maior que o valor crítico L_0 , rejeitamos a hipótese de normalidade dos dados com $(1 - \alpha)100\%$ de confiança. Caso contrário, não rejeitamos a hipótese de normalidade. Para um teste de nível de significância α , rejeitamos H_0 se $T_L \geq t$, em que

$$P_{H_0}(L_n \geq L_0) = \alpha.$$

Este valor tabelado pode ser encontrado na tabela 6 na página 289 do livro do professor Humberto de Campos. Esta tabela os limites superiores para o teste de Lilliefors para os valores do nível de significância apresentados na primeira linha da tabela.

n	0,20	0,15	0,10	0,05	0,01
4	0,300	0,319	0,352	0,381	0,417
5	0,285	0,299	0,315	0,337	0,405
6	0,265	0,277	0,294	0,318	0,364
7	0,247	0,258	0,278	0,300	0,348
8	0,233	0,244	0,261	0,285	0,331
9	0,223	0,233	0,249	0,271	0,311
10	0,215	0,224	0,239	0,258	0,294
11	0,206	0,217	0,230	0,249	0,284
12	0,199	0,212	0,223	0,242	0,275
13	0,190	0,202	0,214	0,234	0,268
14	0,183	0,194	0,207	0,227	0,261
15	0,177	0,187	0,201	0,220	0,257
16	0,173	0,182	0,195	0,213	0,250
17	0,169	0,177	0,189	0,206	0,245
18	0,166	0,173	0,184	0,200	0,239
19	0,163	0,169	0,179	0,195	0,235
20	0,160	0,166	0,174	0,190	0,231
25	0,142	0,147	0,158	0,173	0,200
30	0,131	0,138	0,144	0,161	0,187
> 30	$\frac{0,738}{\sqrt{n}}$	$\frac{0,768}{\sqrt{n}}$	$\frac{0,805}{\sqrt{n}}$	$\frac{0,886}{\sqrt{n}}$	$\frac{1,031}{\sqrt{n}}$

Para fazer o teste manualmente é interessante preparar a seguinte tabela:

amostra ord.	amostra ord. padronizada	\hat{F}_n	Φ	$L^+ = \Phi(z_i) - F_n(z_i) $	$L^- = \Phi(z_i) - F_n(z_{i-1}) $
y_1	z_1	$\frac{1}{n}$	$\Phi(z_1)$	$ \Phi(z_1) - F_n(z_1) $	$ \Phi(z_1) - 0 $
y_2	z_2	$\frac{2}{n}$	$\Phi(z_2)$	$ \Phi(z_2) - F_n(z_2) $	$ \Phi(z_2) - F_n(z_1) $
\vdots	\vdots	\vdots	\vdots	\vdots	
y_i	z_i	$\frac{i}{n}$	$\Phi(z_i)$	$ \Phi(z_i) - F_n(z_i) $	$ \Phi(z_i) - F_n(z_{i-1}) $
\vdots	\vdots	\vdots	\vdots	\vdots	
y_{n-1}	z_{n-1}	$\frac{n-1}{n}$	$\Phi(z_{n-1})$	$ \Phi(z_{n-1}) - F_n(z_{n-1}) $	$ \Phi(z_{n-1}) - F_n(z_{n-2}) $
y_n	z_n	1	$\Phi(z_n)$	$ \Phi(z_n) - F_n(z_n) $	$ \Phi(z_n) - F_n(z_{n-1}) $

Agora é só calcular

$$L_{n_{cal}} = \max(L^+, L^-) = \max(\max(L^+), \max(L^-)).$$

O nível descritivo do teste é dado por:

$$nd = P(Ln \geq L_{n_{cal}}).$$

O teste é equivalente a testar se a amostra é oriunda de uma população normal de média \bar{X} e variância S^2 , através do teste de Kolmogorov. percebe-se que a diferença é que no teste de Lilliefors padronizamos as observações, ao contrário do teste de Kolmogorov.

Vamos apresentar um exemplo extraído do livro do professor Humberto de Campos.

Exemplo 4.1

Num ensaio com bananeiras, realizado pelo Dr. Jairo Ribeiro de Mattos, da ESALQ, em Piracicaba, SP, foram obtidos os seguintes pesos médios (em kg) de cada cacho:

13,9	18,9	21,1	22,2	23,4
17,7	19,4	21,3	22,7	23,8
17,9	19,6	21,7	22,8	24,4
18,3	20,2	21,9	23,2	24,4
18,5	20,8	22,0	23,3	24,9.

Teste, ao nível de significância $\alpha = 0,05$ se os dados são oriundos de uma distribuição normal.

Solução

O valor tabelado para $n = 25$ e $\alpha = 0,05$ é $L_0 = 0,173$.

A nossa regra de decisão será:

$$Ln_{cal} > 0,173,$$

rejeitar H_0 , caso contrário não rejeitar.

Seja X o peso médio das bananas de um cacho. A amostra de tamanho 25

forneceu a média, a variância e o desvio padrão dados a seguir:

Os dados já estão ordenados. Vamos padronizar.

A média, a variância e o desvio padrão são dados por:

```
> n=length(X);n
[1] 25
> S2=var(X);S2
[1] 6.8856
> S=sd(X);S
[1] 2.624043
>
```

Vamos montar a tabela para concluir o teste.

```
> Y=sort(X);Y #####A amostra já veio ordenada!!!!
[1] 13.9 17.7 17.9 18.3 18.5 18.9 19.4 19.8 20.2 20.6 21.1 21.3 21.7 21.9 22.0
[16] 22.2 22.7 22.8 23.2 23.3 23.4 23.8 24.4 24.4 24.9
>
> Z=(Y-Xb)/S
> F_Z=pnorm(Z)
> i=1:n
> F_e=i/n
>
>
> #####Vamos calcular L^+
>
> Lmais=abs(F_Z-F_e)
>
> #####Vamos calcular L^-
>
> F1_e=c(0,F_e[1:n-1])
>
> Lmenos=abs(F_Z-F1_e)
>
> tab=cbind(Y,Z,F_Z,F_e,F1_e,Lmais,Lmenos)
>
tab=cbind(X,Z,F_Z,F_e,F1_e,Lmais,Lmenos);round(tab,3)
      X      Z    F_Z  F_e F1_e  Lmais  Lmenos
[1,] 13.9 -2.756 0.003 0.04 0.00 0.037  0.003
[2,] 17.7 -1.308 0.095 0.08 0.04 0.015  0.055
[3,] 17.9 -1.232 0.109 0.12 0.08 0.011  0.029
[4,] 18.3 -1.079 0.140 0.16 0.12 0.020  0.020
[5,] 18.5 -1.003 0.158 0.20 0.16 0.042  0.002
```

```
[6,] 18.9 -0.851 0.197 0.24 0.20 0.043 0.003
[7,] 19.4 -0.660 0.255 0.28 0.24 0.025 0.015
[8,] 19.8 -0.508 0.306 0.32 0.28 0.014 0.026
[9,] 20.2 -0.355 0.361 0.36 0.32 0.001 0.041
[10,] 20.6 -0.203 0.420 0.40 0.36 0.020 0.060
[11,] 21.1 -0.012 0.495 0.44 0.40 0.055 0.095
[12,] 21.3 0.064 0.526 0.48 0.44 0.046 0.086
[13,] 21.7 0.216 0.586 0.52 0.48 0.066 0.106
[14,] 21.9 0.293 0.615 0.56 0.52 0.055 0.095
[15,] 22.0 0.331 0.630 0.60 0.56 0.030 0.070
[16,] 22.2 0.407 0.658 0.64 0.60 0.018 0.058
[17,] 22.7 0.598 0.725 0.68 0.64 0.045 0.085
[18,] 22.8 0.636 0.738 0.72 0.68 0.018 0.058
[19,] 23.2 0.788 0.785 0.76 0.72 0.025 0.065
[20,] 23.3 0.826 0.796 0.80 0.76 0.004 0.036
[21,] 23.4 0.864 0.806 0.84 0.80 0.034 0.006
[22,] 23.8 1.017 0.845 0.88 0.84 0.035 0.005
[23,] 24.4 1.245 0.894 0.92 0.88 0.026 0.014
[24,] 24.4 1.245 0.894 0.96 0.92 0.066 0.026
[25,] 24.9 1.436 0.924 1.00 0.96 0.076 0.036
>
```

```
> L1=max(Lmais);L1
[1] 0.07550796
> L2=max(Lmenos);L2
[1] 0.1056854
>
> L_cal=max(L1,L2);L_cal
[1] 0.1056854
>
>
>
```

Temos um empate no valor 24,4. Vamos formar uma tabela com os 24 valores distintos.

```
> #####Vamos ajeitar o empate 24.4 para ele aparecer só uma vez. Na realidade
```

```
> ##vamos sumir com a linha 23.
>
>
> tab1=rbind(tab[1:22,],tab[24:25,]);tab1
      Y      Z      Fi  Fn      Lmais      Lmenos
[1,] 13.9 -2.75605273 0.002925178 0.04 0.037074822 0.002925178
[2,] 17.7 -1.30790555 0.095452673 0.08 0.015452673 0.055452673
[3,] 17.9 -1.23168728 0.109032961 0.12 0.010967039 0.029032961
[4,] 18.3 -1.07925074 0.140237984 0.16 0.019762016 0.020237984
[5,] 18.5 -1.00303246 0.157922599 0.20 0.042077401 0.002077401
[6,] 18.9 -0.85059592 0.197496928 0.24 0.042503072 0.002503072
[7,] 19.4 -0.66005024 0.254610795 0.28 0.025389205 0.014610795
[8,] 19.8 -0.50761369 0.305862142 0.32 0.014137858 0.025862142
[9,] 20.2 -0.35517715 0.361228441 0.36 0.001228441 0.041228441
[10,] 20.6 -0.20274060 0.419668892 0.40 0.019668892 0.059668892
[11,] 21.1 -0.01219492 0.495135050 0.44 0.055135050 0.095135050
[12,] 21.3 0.06402335 0.525524182 0.48 0.045524182 0.085524182
[13,] 21.7 0.21645989 0.585685358 0.52 0.065685358 0.105685358
[14,] 21.9 0.29267817 0.615115920 0.56 0.055115920 0.095115920
[15,] 22.0 0.33078730 0.629597424 0.60 0.029597424 0.069597424
[16,] 22.2 0.40700557 0.657998053 0.64 0.017998053 0.057998053
[17,] 22.7 0.59755126 0.724930301 0.68 0.044930301 0.084930301
[18,] 22.8 0.63566039 0.737501101 0.72 0.017501101 0.057501101
[19,] 23.2 0.78809694 0.784679997 0.76 0.024679997 0.064679997
[20,] 23.3 0.82620607 0.795656398 0.80 0.004343602 0.035656398
[21,] 23.4 0.86431521 0.806292621 0.84 0.033707379 0.006292621
[22,] 23.8 1.01675175 0.845364231 0.88 0.034635769 0.005364231
[23,] 24.4 1.24540657 0.893508830 0.96 0.066491170 0.026491170
[24,] 24.9 1.43595225 0.924492036 1.00 0.075507964 0.035507964
>
>
```

Vamos transformar a tabela 1 usando o Latex .

O valor máximo de L^+ observado foi:

$$L_1 = \max(L^+) = 0,07550.$$

O valor máximo de L^- observado foi:

$$L_2 = \max(L^-) = 0,10569.$$

Finalmente o valor da estatística do teste é

$$L_{n_{cal}} = \max(L^+, L^-) = \max(\max(L^+), \max(L^-)) = 0,106.$$

	X	Z	F_Z	F_e	F1_e	Lmais	Lmenos
1	13.90	-2.76	0.00	0.04	0.00	0.04	0.00
2	17.70	-1.31	0.10	0.08	0.04	0.02	0.06
3	17.90	-1.23	0.11	0.12	0.08	0.01	0.03
4	18.30	-1.08	0.14	0.16	0.12	0.02	0.02
5	18.50	-1.00	0.16	0.20	0.16	0.04	0.00
6	18.90	-0.85	0.20	0.24	0.20	0.04	0.00
7	19.40	-0.66	0.25	0.28	0.24	0.03	0.01
8	19.80	-0.51	0.31	0.32	0.28	0.01	0.03
9	20.20	-0.36	0.36	0.36	0.32	0.00	0.04
10	20.60	-0.20	0.42	0.40	0.36	0.02	0.06
11	21.10	-0.01	0.50	0.44	0.40	0.06	0.10
12	21.30	0.06	0.53	0.48	0.44	0.05	0.09
13	21.70	0.22	0.59	0.52	0.48	0.07	0.11
14	21.90	0.29	0.62	0.56	0.52	0.06	0.10
15	22.00	0.33	0.63	0.60	0.56	0.03	0.07
16	22.20	0.41	0.66	0.64	0.60	0.02	0.06
17	22.70	0.60	0.72	0.68	0.64	0.04	0.08
18	22.80	0.64	0.74	0.72	0.68	0.02	0.06
19	23.20	0.79	0.78	0.76	0.72	0.02	0.06
20	23.30	0.83	0.80	0.80	0.76	0.00	0.04
21	23.40	0.86	0.81	0.84	0.80	0.03	0.01
22	23.80	1.02	0.85	0.88	0.84	0.03	0.01
23	24.40	1.25	0.89	0.92	0.88	0.03	0.01
24	24.40	1.25	0.89	0.96	0.92	0.07	0.03
25	24.90	1.44	0.92	1.00	0.96	0.08	0.04

```

> L1=max(Lmais);L1
[1] 0.07550796
> L2=max(Lmenos);L2
[1] 0.1056854
> L_cal=max(L1,L2);L_cal
[1] 0.1056854
> L1=max(tab1[,5])
> L2=max(tab1[,6])
> max(L1,L2)
[1] 0.1056854
>

```

Como $0,106 < 0,173$ não podemos rejeitar a normalidade dos dados.

Usando a tabela dada temos :

Como $P(L_n \leq 0,106) > P(L_n \leq 0,142) = 0,20$ temos que $nd > 0,20$.

Na realidade ele vale 0,6689 usando a função `lillie.test` do pacote `nortest` do software *R* faz diretamente o teste. A saída do *R* foi:

```
>
> require(nortest)
>
>
> lillie.test(X)

Lilliefors (Kolmogorov-Smirnov) normality test

data:  X
D = 0.10569, p-value = 0.6689

>
```

Como $\alpha = 0,05 < 0,6689$ não temos razões para descrever da normalidade dos dados.

3 Caso Exponencial

O teste de Lilliefors também pode ser usado para a distribuição exponencial. vamos contar um pouco dessa história.

Lilliefors em 1967 comparou o poder do teste de qui-quadrado para muitas distribuições não normais e obteve que seu teste foi mais poderoso em todas as situações consideradas.

Um procedimento para se construir bandas de confiança para a distribuição normal foi obtido em Srinivasan e Wharton (1973, Biometrika). Uma segunda modificação do teste Kolmogorov foi feita por Lilliefors (1967, JASA). O teste consiste em verificar se uma determinada amostra é oriunda de uma população como função de distribuição acumulada dada

$$F_0(x) = \left[1 - \exp\left(-\frac{x}{\lambda}\right)\right] I_{(0,\infty)}(x), \lambda > 0,$$

com λ não especificado, sendo estimado através da amostra obtida.

O estimador de máxima verossimilhança de λ é dado por:

$$\hat{\lambda} = \frac{\sum_{i=1}^n X_i}{n} = \bar{X}.$$

Assim o nosso interesse é verificar se a amostra vem de uma distribuição exponencial de média λ .

Embora Lilliefors tenha obtido a distribuição da estatística do teste sob H_0 usando técnicas de simulação Durbin (1975, Biometrika) e Margolin Maurer (1976, Biometrika) obtiveram a distribuição exata da estatística do teste.

A distribuição exponencial é utilizada para descrever o tempo de espera entre a ocorrência de 2 eventos consecutivos quando eles ocorrem aleatoriamente ao longo do tempo. A taxa de falhas é constante e é dada por:

$$r(t) = \frac{f(t)}{1 - F(t)} = \frac{1}{\lambda},$$

e isto caracteriza a propriedade da falta de memória da exponencial.

Assim o teste para a distribuição exponencial pode ser utilizado para verificar a aleatoriedade de uma sequência.

Uma segunda aplicação do teste é verificar de A taxa de falhas de um determinado produto é constante no tempo de vida útil do mesmo. Esta propriedade originou a distribuição exponencial.

Um outro fato interessante usado no teste é o seguinte: Se $X \sim \text{Exp}(\lambda)$ então

$$Y = \frac{X}{\lambda} \sim \text{Exp}(1).$$

Seja $G(y)$ a função de distribuição de Y . Para $y > 0$ temos :

$$G(y) = P(Y \leq y) = P\left(\frac{X}{\lambda} \leq y\right) = F(\lambda y) = 1 - \exp(-y),$$

que é a acumulada da exponencial com média 1.

Agora vamos descrever o teste de Lilliefors para a distribuição exponencial.

A suposição do teste é que temos:

$$X_1, X_2, \dots, X_n \stackrel{iid}{\sim} F(x)$$

O nosso objetivo é:

$$H_0 : F(x) = F_0(x) = \left[1 - \exp\left(-\frac{x}{\lambda}\right)\right] I_{(0,\infty)}(x)$$

versus

$$H_1 : F(x) \neq F_0(x), \text{ para algum } x > 0.$$

Lilliefors (1969) propôs padronizar as observações da seguinte forma:

$$Z_i = \frac{X_i}{\bar{X}}, \quad i = 1, 2, \dots, n.$$

Agora vamos pegar as estatísticas de ordem da nossa amostra de Z .

$$z_{(1)} < z_{(2)} < \dots < z_{(n)}.$$

Considere F_0 a acumulada da exponencial de média 1. A estatística do teste

$$LE^+ = \sup_{z_{(i)}} |F_0(z_{(i)}) - F_n(z_{(i)})|$$

$$LE^- = \sup_{z_{(i)}} |F_0(z_{(i)}) - F_n(z_{(i-1)})|$$

Com isso, podemos utilizar como estatística de teste

$$LE_n = \max(LE^+, LE^-).$$

Rejeitamos H_0 , ao nível α se $lE_n \geq L_0$ de forma que

$$P_{H_0}(L_n \geq L_0) = \alpha.$$

O livro do Conover traz a tabela **A15** para nos ajudar a tomar uma decisão.

TABLE A15 Quantiles of the Lilliefors Test Statistic for the Exponential Distribution*

	$p = 0.05$	0.10	0.20	0.30	0.50	0.70	0.80	0.90	0.95	0.99	0.999
$n = 2$	0.3127	0.3200	0.3337	0.3617	0.4337	0.5034	0.5507	0.5934	0.6133	0.6284	0.6317
3	0.2299	0.2544	0.2899	0.3166	0.3645	0.4122	0.4508	0.5111	0.5508	0.6003	0.6296
4	0.2072	0.2281	0.2545	0.2766	0.3163	0.3685	0.4007	0.4442	0.4944	0.5374	0.6215
5	0.1884	0.2052	0.2290	0.2483	0.2877	0.3317	0.3603	0.4045	0.4420	0.5127	0.5814
6	0.1726	0.1882	0.2102	0.2290	0.2645	0.3045	0.3320	0.3732	0.4085	0.4748	0.5497
7	0.1604	0.1750	0.1961	0.2136	0.2458	0.2838	0.3098	0.3481	0.3811	0.4459	0.5181
8	0.1506	0.1646	0.1845	0.2006	0.2309	0.2671	0.2914	0.3274	0.3590	0.4208	0.4913
9	0.1426	0.1561	0.1746	0.1897	0.2186	0.2529	0.2758	0.3101	0.3404	0.3995	0.4679
10	0.1359	0.1486	0.1661	0.1805	0.2082	0.2407	0.2626	0.2955	0.3244	0.3813	0.4473
12	0.1249	0.1364	0.1524	0.1657	0.1912	0.2209	0.2411	0.2714	0.2981	0.3511	0.4132
14	0.1162	0.1268	0.1418	0.1542	0.1778	0.2054	0.2242	0.2525	0.2774	0.3272	0.3858
16	0.1091	0.1191	0.1332	0.1448	0.1669	0.1929	0.2105	0.2371	0.2606	0.3076	0.3632
18	0.1032	0.1127	0.1260	0.1369	0.1578	0.1824	0.1990	0.2242	0.2445	0.2911	0.3441
20	0.0982	0.1073	0.1199	0.1303	0.1501	0.1735	0.1893	0.2132	0.2345	0.2771	0.3277
22	0.0939	0.1025	0.1146	0.1245	0.1434	0.1657	0.1809	0.2038	0.2241	0.2649	0.3135
24	0.0901	0.0984	0.1099	0.1195	0.1376	0.1590	0.1735	0.1954	0.2150	0.2542	0.3010
26	0.0868	0.0947	0.1058	0.1150	0.1324	0.1530	0.1670	0.1881	0.2069	0.2447	0.2899
28	0.0838	0.0914	0.1021	0.1110	0.1278	0.1477	0.1611	0.1815	0.1997	0.2362	0.2799
30	0.0811	0.0885	0.0988	0.1074	0.1236	0.1428	0.1559	0.1756	0.1932	0.2286	0.2709
35	0.0754	0.0822	0.0918	0.0997	0.1148	0.1326	0.1447	0.1630	0.1793	0.2123	0.2517
40	0.0707	0.0771	0.0861	0.0935	0.1077	0.1243	0.1356	0.1528	0.1681	0.1990	0.2361
45	0.0668	0.0729	0.0814	0.0884	0.1017	0.1174	0.1281	0.1443	0.1588	0.1880	0.2231
50	0.0636	0.0693	0.0774	0.0840	0.0966	0.1116	0.1217	0.1371	0.1509	0.1787	0.2121
60	0.0582	0.0635	0.0708	0.0769	0.0895	0.1021	0.1114	0.1255	0.1381	0.1635	0.1943
70	0.0541	0.0589	0.0658	0.0714	0.0821	0.0946	0.1033	0.1164	0.1281	0.1517	*
80	0.0507	0.0553	0.0616	0.0669	0.0769	0.0887	0.0968	0.1090	0.1200	0.1421	*
90	0.0479	0.0522	0.0582	0.0632	0.0726	0.0838	0.0914	0.1029	0.1132	0.1341	*
$n = 100$	0.0455	0.0496	0.0553	0.0600	0.0690	0.0796	0.0868	0.0977	0.1075	0.1274	*
Approximation	0.4550	0.4959	0.5530	0.6000	0.6998	0.7957	0.8678	0.9773	1.0753	1.2743	*
or $n > 100$	\sqrt{n}	\sqrt{n}	\sqrt{n}	\sqrt{n}	\sqrt{n}	\sqrt{n}	\sqrt{n}	\sqrt{n}	\sqrt{n}	\sqrt{n}	\sqrt{n}

SOURCE: Adapted from Durbin (1975), with permission from the Biometrika Trustees.
The entries in this table are selected quantiles w_α of the Lilliefors test statistic T_n as given by Equation 6.2.6. Reject at the level of significance α if T_n is greater than the $1 - \alpha$ quantile given in the table. The approximation for $n > 100$ is merely the exact value for $n = 100$. More accurate approximations for $n > 100$ may be obtained from Table 54 of Pearson and Hartley (1972).
These quantiles are not presently available.

Figura 1:

Exemplo 4.2 Considere que a secretaria do DEMA-UFC recebeu na última sexta-feira 10 telefonemas entre 10:00 e 11:00 horas, especificamente nos instantes:

10 : 06, 10 : 08, 10 : 16, 10 : 22, 10 : 23, 10 : 34, 10 : 44, 10 : 47, 10 : 51, 10 : 57.

Teste, ao nível de significância de 5% se as chamadas chegam de maneira aleatória.

Solução:

Sabemos que os instantes em que chamadas chegam são aleatórios, se e somente se, a distribuição do tempo entre duas chegadas consecutivas for exponencialmente distribuída. Seja X a variável aleatória representando o tempo entre chegadas sucessivas em minutos. Assim,

$$x_1 = 6, x_2 = 2, x_3 = 8, x_4 = 6, x_5 = 1, x_6 = 11, x_7 = 10, x_8 = 3, x_9 = 4, x_{10} = 6.$$

As estatísticas de ordem da amostra são:

$$y_1 = 1, y_2 = 2, y_3 = 3, y_4 = 4, y_5 = 6, y_6 = 6, y_7 = 6, y_8 = 8, y_9 = 10, y_{10} = 11.$$

A média da amostra vale $\bar{X} = 5,7$. Os dados padronizados são dados por:

Veja a saída do R:

```
> #####Exponencial
>
> X=c(6,2,8,6,1,11,10,3,4,6)
> Xb=mean(X);Xb
[1] 5.7
> n=length(X);n
[1] 10
> Y=sort(X);Y
[1] 1 2 3 4 6 6 6 8 10 11
> mean(Y)
[1] 5.7
> ##k=8 valores distintos
> y=c(1,2,3,4,6,8,10,11)
> z=y/Xb;z
[1] 0.1754386 0.3508772 0.5263158 0.7017544 1.0526316 1.4035088 1.7543860
[8] 1.9298246
>
> F_e=c(1,2,3,4,5,8,9,10)/10##### distribuição empírica
> F1_e=c(0,1,2,3,4,5,8,9)/10
> F_0=pexp(z,1)
>
> L_mais=abs(F_0-F_e)
> L_menos=abs(F_0-F1_e)
>
> tab=cbind(y,z,F_e,F_0,F1_e,L_mais,L_menos)
> round(tab,4)
y      z F_e    F_0 F1_e L_mais L_menos
[1,]  1 0.1754 0.1 0.1609  0.0 0.0609  0.1609
[2,]  2 0.3509 0.2 0.2959  0.1 0.0959  0.1959
[3,]  3 0.5263 0.3 0.4092  0.2 0.1092  0.2092
[4,]  4 0.7018 0.4 0.5043  0.3 0.1043  0.2043
[5,]  6 1.0526 0.5 0.6510  0.4 0.1510  0.2510
[6,]  8 1.4035 0.8 0.7543  0.5 0.0457  0.2543
[7,] 10 1.7544 0.9 0.8270  0.8 0.0730  0.0270
[8,] 11 1.9298 1.0 0.8548  0.9 0.1452  0.0452
>
> L1=max(L_mais);L1
[1] 0.1509819
> L2=max(L_menos);L2
[1] 0.2542668
>
> L=max(L1,L2);L;round(L,4)
[1] 0.2542668
```

```
[1] 0.2543
> require(nortest)
>
> alfa=0.05#####(Percentil 95)
> LE_0=0.3244#####(tabela A15 do Conover)
> L <LE_0#####Não podemos rejeitar que a distribuição seja exponencial.
[1] TRUE
>
>
> #####Vamos calcular o nível descritivo por interpolação
> #####0,2407 < 0,2543 < 0,2686
> ### nd=P(LE_n >=0,2543)
>
> ##p_1=P(LE_n>=0,2686)=0,20 e p_2=P(LE_n>=0,2407)=0,30
>
> p1=0.20;p1
[1] 0.2
>
> p2=0.30
> ap=p2-p1;ap
[1] 0.1
>
> a1=0.2686
> a2=0.2407
>
> a=abs(a1-a2);a
[1] 0.0279
>
> a_0=0.2543
>
> aux=abs(a_0 -a1);aux
[1] 0.0143
>
>
> nd=p1+ (aux/a)*ap;nd
[1] 0.2512545
>
>
```

O nível descritivo aproximado é $nd = 0,2513$.
Vamos retomar o exemplo 14.6 do Bussab&Morettin.

>

```
> #####Amostra
>
> X=c(104,173,393,444,637,651,761,764,818,848,857,865,971,987,995,1001,1052,1069,
+ 1172,1217,1261,1298,1303,1316,1411,1460,1464,1475,1668,2214)/100;X
[1]  1.04  1.73  3.93  4.44  6.37  6.51  7.61  7.64  8.18  8.48  8.57  8.65
[13]  9.71  9.87  9.95 10.01 10.52 10.69 11.72 12.17 12.61 12.98 13.03 13.16
[25] 14.11 14.60 14.64 14.75 16.68 22.14
>
>
>
> n=length(X);n #####tamanho da amostra
[1] 30
>
>
>
> ###Análise Exploratória
>
> xb=mean(X);xb
[1] 10.21633
> s2=var(X);s2
[1] 19.9545
> s=sd(X);s
[1] 4.467046
>
> require(DescTools)
>
> LillieTest(X)

Lilliefors (Kolmogorov-Smirnov) normality test

data:  X
D = 0.088407, p-value = 0.7953

>
>
> Z= scale(X, center = T)
> Z
[,1]
[1,] -2.05422853
[2,] -1.89976404
[3,] -1.40726855
[4,] -1.29309914
[5,] -0.86104628
```

```
[6,] -0.82970566
[7,] -0.58345791
[8,] -0.57674207
[9,] -0.45585681
[10,] -0.38869833
[11,] -0.36855079
[12,] -0.35064186
[13,] -0.11334858
[14,] -0.07753073
[15,] -0.05962180
[16,] -0.04619011
[17,] 0.06797930
[18,] 0.10603577
[19,] 0.33661320
[20,] 0.43735092
[21,] 0.53585002
[22,] 0.61867880
[23,] 0.62987188
[24,] 0.65897389
[25,] 0.87164240
[26,] 0.98133457
[27,] 0.99028904
[28,] 1.01491381
[29,] 1.44696667
[30,] 2.66925093
attr(,"scaled:center")
[1] 10.21633
attr(,"scaled:scale")
[1] 4.467046
> i=1:n;i
[1] 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
[26] 26 27 28 29 30
> Fe=i/n
>
> j=0:(n-1);j
[1] 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
[26] 25 26 27 28 29
> Fem=j/n
> F=pnorm(Z)
>
> DM=abs(F -Fe)
> Dm=abs(F -Fem)
>
```



```
>
> tab1=cbind(i,X,Z, F,Fe,DM)
>
> round(tab1,3)
i      X      Fe
[1,]  1  1.04 -2.054 0.020 0.033 0.013
[2,]  2  1.73 -1.900 0.029 0.067 0.038
[3,]  3  3.93 -1.407 0.080 0.100 0.020
[4,]  4  4.44 -1.293 0.098 0.133 0.035
[5,]  5  6.37 -0.861 0.195 0.167 0.028
[6,]  6  6.51 -0.830 0.203 0.200 0.003
[7,]  7  7.61 -0.583 0.280 0.233 0.046
[8,]  8  7.64 -0.577 0.282 0.267 0.015
[9,]  9  8.18 -0.456 0.324 0.300 0.024
[10,] 10  8.48 -0.389 0.349 0.333 0.015
[11,] 11  8.57 -0.369 0.356 0.367 0.010
[12,] 12  8.65 -0.351 0.363 0.400 0.037
[13,] 13  9.71 -0.113 0.455 0.433 0.022
[14,] 14  9.87 -0.078 0.469 0.467 0.002
[15,] 15  9.95 -0.060 0.476 0.500 0.024
[16,] 16 10.01 -0.046 0.482 0.533 0.052
[17,] 17 10.52  0.068 0.527 0.567 0.040
[18,] 18 10.69  0.106 0.542 0.600 0.058
[19,] 19 11.72  0.337 0.632 0.633 0.002
[20,] 20 12.17  0.437 0.669 0.667 0.002
[21,] 21 12.61  0.536 0.704 0.700 0.004
[22,] 22 12.98  0.619 0.732 0.733 0.001
[23,] 23 13.03  0.630 0.736 0.767 0.031
[24,] 24 13.16  0.659 0.745 0.800 0.055
[25,] 25 14.11  0.872 0.808 0.833 0.025
[26,] 26 14.60  0.981 0.837 0.867 0.030
[27,] 27 14.64  0.990 0.839 0.900 0.061
[28,] 28 14.75  1.015 0.845 0.933 0.088
[29,] 29 16.68  1.447 0.926 0.967 0.041
[30,] 30 22.14  2.669 0.996 1.000 0.004
>
> DMmax=max(DM);DMmax  #### bate com a saída do R
[1] 0.08840679
>
> Dmmax=max(Dm);Dmmax
[1] 0.07979254
>
> L_cal=max(DMmax,Dmmax);L_cal
```

```
[1] 0.08840679
>
> ##### Compare com o valor tabela!!!!
>
>
>
>
```

Com $n = 30$ e $\alpha = 0,05$ a tabela 6 do Humberto nos dá:

$$L_{tab} = 0,161$$

Como

$$L_{cal} = 0,088 < 0,161,$$

não rejeitamos H_0 , isto é, os dados seguem normal.

O nível descritivo do teste é dado por:

Como

$$0,0,88 < 0,0,131$$

temos que

$$nd < 0,80$$

Na saída do teste temos:

$$nd = 0,79 < 0,80.$$