

CC0291 - Estatística Não Paramétrica

Testes Para Ajuste de Distribuições - 25/04/2023

Profs. Juvêncio Nobre, Maurício Mota e Gualberto Agamez

1 Introdução

Este trabalho é inteiramente calcado das notas de aula do professor Juvêncio Nobre e também no livro do Conover Practical NonParametric Statistics. Vamos apresentar o teste de Shapiro-Wilk. Usaremos o software R .

2 Teste de Shapiro -Wilk

Vamos apresentar o teste de Shapiro -Wilk para a normalidade.

A nossa suposição básica é que temos uma amostra aleatória

$$X_1, X_2, \dots, X_n \sim F_X(x)$$

em que $F_X(x)$ é a acumulada de uma normal com média μ e variância σ^2

As nossas hipóteses são:

$$H_0 : X \xrightarrow{D} N(,),$$

a hipótese alternativa é a não normalidade.

O teste foi proposto por Samuel Shapiro e Martin Wilk(1965, Biometrika) e é um teste competidor ao teste de Lilliefors.

Estudos empíricos indicam que o teste de Shapiro -Wilk é mais poderoso que o teste de Lilliefors (Shapiro et al,1968,JASA).

Originalmente,o teste (distribuição exata da estatística e seus respectivos quantis) foi proposto para um tamanho amostral $n \leq 50$, todavia existem algumas propostas de extensão para situações em que n é grande, veja por exemplo D'Agostino (1971, Biometrika), Shapiro e Francia (1972,JASA) e Royston (1982,Applied Statistics).

Vamos discutir as ideias por trás da construção do teste.

Considere as estatísticas de ordem de uma amostra aleatória de uma distribuição normal padrão:

$$Z_{(1)}, Z_{(2)}, \dots, Z_{(n)}.$$

e vamos representar:

$$\alpha(i) = \mathbb{E}(Z_{(i)}), \quad i = 1, 2, \dots, n$$

$$\gamma_{ij} = \text{Cov}(Z_{(i)}, Z_{(j)}), \quad 1 \leq i < j \leq n.$$

Seja X_1, X_2, \dots, X_n uma amostra aleatória de $X \sim N(\mu, \sigma^2)$
então:

$$Z_{(i)} = \frac{X_{(i)} - \mu}{\sigma},$$

de forma que:

$$\begin{aligned} X_{(i)} &= \mu + \sigma Z_{(i)} \\ &= \mu + \sigma X_i + \sigma(Z_{(i)} - X_i) \\ &= \mu + \sigma X_i + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma^2). \end{aligned}$$

Os parâmetros μ e σ podem ser estimados através do Método de Mínimos Quadrados Generalizados.

A estatística proposta Shapiro e Wilk(1965, Biometrika) é:

$$W = \frac{\left[\sum_{i=1}^n m_i X_{(i)} \right]^2}{\sum_{i=1}^n (X_i - \bar{X})^2} = R^2(x, m) = R^2(x, \hat{x}),$$

em que

$$(\hat{x}_{(1)}, \hat{x}_{(2)}, \dots, \hat{x}_{(n)}) = \hat{x}_{()},$$

é o vetor v com os valores estimados das estatísticas de ordem que dependem de μ e σ^2 e $m' = (m_{1,2}, \dots, m_n)$ é tal que:

$$m' = \frac{\alpha' \Gamma^{-1}}{\sqrt{\alpha' \Gamma^{-1} \Gamma^{-1} \alpha}}.$$

Note que a norma do vetor m é unitária. Se a amostra for retirada de uma população com distribuição normal, então espera-se que:

$$W \approx 1,$$

enquanto em que situações em que a população geradora não possui distribuição normal, W tende a assumir um valor **pequeno**. Portanto devemos rejeitar H_0 se:

$$W \leq w_0,$$

quando w_0 for pequeno.

Para a nossa felicidade não é necessário calcular X_i e γ_{ij} diretamente uma vez que a estatística do teste pode ser reescrita como:

$$W = \frac{\left[\sum_{i=1}^k a_i (X_{(n-i+1)} - X_{(i)}) \right]^2}{\sum_{i=1}^n (X_i - \bar{X})^2},$$

em que os coeficientes a_1, a_2, \dots, a_k são apresentados na tabela **A16** com $k \approx \frac{n}{2}$.

Logo, rejeitamos H_0 ao nível de significância α se,

$$W \leq W_{tab},$$

com

$$\mathbb{P}_{H_0}[W \leq W_{tab}] = \alpha.$$

O nível descritivo pode ser obtido por simulação ou então de forma aproximada usando interpolação.

A estatística do Teste W é basicamente o quadrado do coeficiente de correlação entre as estatísticas de ordem da amostra e os escores que dem ser próximos se a população for normal. Se W é perto de 1 a amostra comporta-se como normal. Se W é muito pequeno a amostra parece ser não Normal. Quantis de W são dados na tabela **A17**.

Assim devemos rejeitar H_0 se W é menor ou igual ao quantil α dado pela tabela **A17**. Para um valor do nível descritivo mais preciso, as instruções da tabela **A18** permitem aproximar W para uma distribuição normal.

Vamos apresentar as várias tabelas que estão no livro do Conover:

TABLE A16 Coefficients for the Shapiro-Wilk Test*

n	2	3	4	5	6	7	8	9	10
1	0.7071	0.7071	0.6872	0.6646	0.6431	0.6233	0.6052	0.0588	0.5739
2	—	0.0000	0.1667	0.2413	0.2806	0.3031	0.3164	0.3244	0.3291
3	—	—	—	0.0000	0.0875	0.1401	0.1743	0.1976	0.2141
4	—	—	—	—	—	0.0000	0.0561	0.0947	0.1224
5	—	—	—	—	—	—	—	0.0000	0.0399

n	11	12	13	14	15	16	17	18	19	20
1	0.5601	0.5475	0.5359	0.5251	0.5150	0.5056	0.4968	0.4886	0.4808	0.4734
2	0.3315	0.3325	0.3325	0.3318	0.3306	0.3290	0.3273	0.3253	0.3232	0.3211
3	0.2260	0.2347	0.2412	0.2460	0.2495	0.2521	0.2540	0.2553	0.2561	0.2565
4	0.1429	0.1586	0.1707	0.1802	0.1878	0.1939	0.1988	0.2027	0.2059	0.2085
5	0.0695	0.0922	0.1099	0.1240	0.1353	0.1447	0.1524	0.1587	0.1641	0.1686
6	0.0000	0.0303	0.0539	0.0727	0.0880	0.1005	0.1109	0.1197	0.1271	0.1334
7	—	—	0.0000	0.0240	0.0433	0.0593	0.0725	0.0837	0.0932	0.1013
8	—	—	—	—	0.0000	0.0196	0.0359	0.0496	0.0612	0.0711
9	—	—	—	—	—	—	0.0000	0.0163	0.0303	0.0422
10	—	—	—	—	—	—	—	—	0.0000	0.0140

n	21	22	23	24	25	26	27	28	29	30
1	0.4643	0.4590	0.4542	0.4493	0.4450	0.4407	0.4366	0.4328	0.4291	0.4254
2	0.3185	0.3156	0.3126	0.3098	0.3069	0.3043	0.3018	0.2992	0.2968	0.2944
3	0.2578	0.2571	0.2563	0.2554	0.2543	0.2533	0.2522	0.2510	0.2499	0.2487
4	0.2119	0.2131	0.2139	0.2145	0.2148	0.2151	0.2152	0.2151	0.2150	0.2148
5	0.1736	0.1764	0.1787	0.1807	0.1822	0.1836	0.1848	0.1857	0.1864	0.1870
6	0.1399	0.1443	0.1480	0.1512	0.1539	0.1563	0.1584	0.1601	0.1616	0.1630
7	0.1092	0.1150	0.1201	0.1245	0.1283	0.1316	0.1346	0.1372	0.1395	0.1415
8	0.0804	0.0878	0.0941	0.0997	0.1046	0.1089	0.1128	0.1162	0.1192	0.1219
9	0.0530	0.0618	0.0696	0.0764	0.0823	0.0876	0.0923	0.0965	0.1002	0.1036
10	0.0263	0.0368	0.0459	0.0539	0.0610	0.0672	0.0728	0.0778	0.0822	0.0862
11	0.0000	0.0122	0.0228	0.0321	0.0403	0.0476	0.0540	0.0598	0.0650	0.0697
12	—	—	0.0000	0.0107	0.0200	0.0284	0.0358	0.0424	0.0483	0.0537
13	—	—	—	—	0.0000	0.0094	0.0178	0.0253	0.0320	0.0381
14	—	—	—	—	—	—	0.0000	0.0084	0.0159	0.0227
15	—	—	—	—	—	—	—	—	0.0000	0.0076

Figura 1:

No pacote **nortest** do R função **shapiro.test** executa o teste de normalidade.

Vamos apresentar o exemplo que consta do material:

Estatística não paramétrica básica no software R: uma abordagem por resolução de problemas.

Oa autores são:

Magda Carvalho Pires Matheus Barros Castro Zaba Valtuille Lieber Thais Pacheco Menezes
Raquel Yuri da Silveira Aoki

Exemplo 2.1 *Exemplo 1.5 Testar se os dados guardados em Y seguem a distribuição Normal.*

```
>
> Y=c(1.90642, 2.10288, 1.52229, 2.61826, 1.42738, 2.22488, 1.69742, 3.15435,
+ 1.98492, 1.99568)
>
>
> shapiro.test(Y)

Shapiro-Wilk normality test
```

TABLE A16 (Continued)

$\frac{n}{i}$	31	32	33	34	35	36	37	38	39	40
1	0.4220	0.4188	0.4156	0.4127	0.4096	0.4068	0.4040	0.4015	0.3989	0.3964
2	0.2921	0.2898	0.2876	0.2854	0.2834	0.2813	0.2794	0.2774	0.2755	0.2737
3	0.2475	0.2462	0.2451	0.2439	0.2427	0.2415	0.2403	0.2391	0.2380	0.2368
4	0.2145	0.2141	0.2137	0.2132	0.2127	0.2121	0.2116	0.2110	0.2104	0.2098
5	0.1874	0.1878	0.1880	0.1882	0.1883	0.1883	0.1883	0.1881	0.1880	0.1878
6	0.1641	0.1651	0.1660	0.1667	0.1673	0.1678	0.1683	0.1686	0.1689	0.1691
7	0.1433	0.1449	0.1463	0.1475	0.1487	0.1496	0.1505	0.1513	0.1520	0.1526
8	0.1243	0.1265	0.1284	0.1301	0.1317	0.1331	0.1344	0.1356	0.1366	0.1376
9	0.1066	0.1093	0.1118	0.1140	0.1160	0.1179	0.1196	0.1211	0.1225	0.1237
10	0.0899	0.0931	0.0961	0.0988	0.1013	0.1036	0.1056	0.1075	0.1092	0.1108
11	0.0739	0.0777	0.0812	0.0844	0.0873	0.0900	0.0924	0.0947	0.0967	0.0986
12	0.0585	0.0629	0.0669	0.0706	0.0739	0.0770	0.0798	0.0824	0.0848	0.0870
13	0.0435	0.0485	0.0530	0.0572	0.0610	0.0645	0.0677	0.0706	0.0733	0.0759
14	0.0289	0.0344	0.0395	0.0441	0.0484	0.0523	0.0559	0.0592	0.0622	0.0651
15	0.0144	0.0206	0.0262	0.0314	0.0361	0.0404	0.0444	0.0481	0.0515	0.0546
16	0.0000	0.0068	0.0131	0.0187	0.0239	0.0287	0.0331	0.0372	0.0409	0.0444
17	—	—	0.0000	0.0062	0.0119	0.0172	0.0220	0.0264	0.0305	0.0343
18	—	—	—	—	0.0000	0.0057	0.0110	0.0158	0.0203	0.0244
19	—	—	—	—	—	—	0.0000	0.0053	0.0101	0.0146
20	—	—	—	—	—	—	—	—	0.0000	0.0049

$\frac{n}{i}$	41	42	43	44	45	46	47	48	49	50
1	0.3940	0.3917	0.3894	0.3872	0.3850	0.3830	0.3808	0.3789	0.3770	0.3751
2	0.2719	0.2701	0.2684	0.2667	0.2651	0.2635	0.2620	0.2604	0.2589	0.2574
3	0.2357	0.2345	0.2334	0.2323	0.2313	0.2302	0.2291	0.2281	0.2271	0.2260
4	0.2091	0.2085	0.2078	0.2072	0.2065	0.2058	0.2052	0.2045	0.2038	0.2032
5	0.1876	0.1874	0.1871	0.1868	0.1865	0.1862	0.1859	0.1855	0.1851	0.1847
6	0.1693	0.1694	0.1695	0.1695	0.1695	0.1695	0.1695	0.1693	0.1692	0.1691
7	0.1531	0.1535	0.1539	0.1542	0.1545	0.1548	0.1550	0.1551	0.1553	0.1554
8	0.1384	0.1392	0.1398	0.1405	0.1410	0.1415	0.1420	0.1423	0.1427	0.1430
9	0.1249	0.1259	0.1269	0.1278	0.1286	0.1293	0.1300	0.1306	0.1312	0.1317
10	0.1123	0.1136	0.1149	0.1160	0.1170	0.1180	0.1189	0.1197	0.1205	0.1212
11	0.1004	0.1020	0.1035	0.1049	0.1062	0.1073	0.1085	0.1095	0.1105	0.1113
12	0.0891	0.0909	0.0927	0.0943	0.0959	0.0972	0.0986	0.0998	0.1010	0.1020
13	0.0782	0.0804	0.0824	0.0842	0.0860	0.0876	0.0892	0.0906	0.0919	0.0932
14	0.0677	0.0701	0.0724	0.0745	0.0765	0.0783	0.0801	0.0817	0.0832	0.0846
15	0.0575	0.0602	0.0628	0.0651	0.0673	0.0694	0.0713	0.0731	0.0748	0.0764

Figura 2:

```
data: Y
W = 0.92669, p-value = 0.4162
```

Vamos fazer manualmente no R .

```
>
> #####Vamos ordenar em forma crescente:
> X=sort(Y);X
[1] 1.42738 1.52229 1.69742 1.90642 1.98492 1.99568 2.10288 2.22488 2.61826
[10] 3.15435
> #####Vamos ordenar em forma decrescente:
>
> X_d=sort(X,decreasing = TRUE);X_d
```

TABLE A16 (Continued)

n	41	42	43	44	45	46	47	48	49	50
16	0.0476	0.0506	0.0534	0.0560	0.0584	0.0607	0.0628	0.0648	0.0667	0.0685
17	0.0379	0.0411	0.0442	0.0471	0.0497	0.0522	0.0546	0.0568	0.0588	0.0608
18	0.0283	0.0318	0.0352	0.0383	0.0412	0.0439	0.0465	0.0489	0.0511	0.0532
19	0.0188	0.0227	0.0263	0.0296	0.0328	0.0357	0.0385	0.0411	0.0436	0.0459
20	0.0094	0.0136	0.0175	0.0211	0.0245	0.0277	0.0307	0.0335	0.0361	0.0386
21	0.0000	0.0045	0.0087	0.0126	0.0163	0.0197	0.0229	0.0259	0.0288	0.0314
22	—	—	0.0000	0.0042	0.0081	0.0118	0.0153	0.0185	0.0215	0.0244
23	—	—	—	—	0.0000	0.0039	0.0076	0.0111	0.0143	0.0174
24	—	—	—	—	—	—	0.0000	0.0037	0.0071	0.0104
25	—	—	—	—	—	—	—	—	0.0000	0.0035

SOURCE: Reprinted from Vol. 2 of Pearson and Hartley (1976), with permission from the Biometrika Trustees.
*The entries in this table are the coefficients a_i for use in the Shapiro-Wilk test statistic for normality given by Equation 6.2.9.

TABLE A17 Quantiles of the Shapiro-Wilk Test Statistic*

n	0.01	0.02	0.05	0.10	0.50	0.90	0.95	0.98	0.99
3	0.753	0.756	0.767	0.789	0.959	0.998	0.999	1.000	1.000
4	0.687	0.707	0.748	0.792	0.935	0.987	0.992	0.996	0.997
5	0.686	0.715	0.762	0.806	0.927	0.979	0.986	0.991	0.993
6	0.713	0.743	0.788	0.826	0.927	0.974	0.981	0.986	0.989
7	0.730	0.760	0.803	0.838	0.928	0.972	0.979	0.985	0.988
8	0.749	0.778	0.818	0.851	0.932	0.972	0.978	0.984	0.987
9	0.764	0.791	0.829	0.859	0.935	0.972	0.978	0.984	0.986
10	0.781	0.806	0.842	0.869	0.938	0.972	0.978	0.983	0.986
11	0.792	0.817	0.850	0.876	0.940	0.973	0.979	0.984	0.986
12	0.805	0.828	0.859	0.883	0.943	0.973	0.979	0.984	0.986
13	0.814	0.837	0.866	0.889	0.945	0.974	0.979	0.984	0.986
14	0.825	0.846	0.874	0.895	0.947	0.975	0.980	0.984	0.986
15	0.835	0.855	0.881	0.901	0.950	0.975	0.980	0.984	0.987
16	0.844	0.863	0.887	0.906	0.952	0.976	0.981	0.985	0.987
17	0.851	0.869	0.892	0.910	0.954	0.977	0.981	0.985	0.987
18	0.858	0.874	0.897	0.914	0.956	0.978	0.982	0.986	0.988
19	0.863	0.879	0.901	0.917	0.957	0.978	0.982	0.986	0.988
20	0.868	0.884	0.905	0.920	0.959	0.979	0.983	0.986	0.988
21	0.873	0.888	0.908	0.923	0.960	0.980	0.983	0.987	0.989
22	0.878	0.892	0.911	0.926	0.961	0.980	0.984	0.987	0.989
23	0.881	0.895	0.914	0.928	0.962	0.981	0.984	0.987	0.989
24	0.884	0.898	0.916	0.930	0.963	0.981	0.984	0.987	0.989

Figura 3:

```
[1] 3.15435 2.61826 2.22488 2.10288 1.99568 1.98492 1.90642 1.69742 1.52229
[10] 1.42738
> n=length(Y);n
[1] 10
> k=n/2;k
[1] 5
>
> Xb=mean(X);Xb #####média amostral.
[1] 2.063448
>
>
> #####Calcular o denominador da Estatística:
>
>
> den=sum((X-Xb)^2);den
[1] 2.392327
>
> #####Pegar os coeficientes na tabela A16 do Conover com n=10.
```

TABLE A17 (Continued)

n	0.01	0.02	0.05	0.10	0.50	0.90	0.95	0.98	0.99
25	0.888	0.901	0.918	0.931	0.964	0.981	0.985	0.988	0.989
26	0.891	0.904	0.920	0.933	0.965	0.982	0.985	0.988	0.989
27	0.894	0.906	0.923	0.935	0.965	0.982	0.985	0.988	0.990
28	0.896	0.908	0.924	0.936	0.966	0.982	0.985	0.988	0.990
29	0.898	0.910	0.926	0.937	0.966	0.982	0.985	0.988	0.990
30	0.900	0.912	0.927	0.939	0.967	0.983	0.985	0.988	0.990
31	0.902	0.914	0.929	0.940	0.967	0.983	0.986	0.988	0.990
32	0.904	0.915	0.930	0.941	0.968	0.983	0.986	0.988	0.990
33	0.906	0.917	0.931	0.942	0.968	0.983	0.986	0.989	0.990
34	0.908	0.919	0.933	0.943	0.969	0.983	0.986	0.989	0.990
35	0.910	0.920	0.934	0.944	0.969	0.984	0.986	0.989	0.990
36	0.912	0.922	0.935	0.945	0.970	0.984	0.986	0.989	0.990
37	0.914	0.924	0.936	0.946	0.970	0.984	0.987	0.989	0.990
38	0.916	0.925	0.938	0.947	0.971	0.984	0.987	0.989	0.991
39	0.917	0.927	0.939	0.948	0.971	0.984	0.987	0.989	0.991
40	0.919	0.928	0.940	0.949	0.972	0.985	0.987	0.989	0.991
41	0.920	0.929	0.941	0.950	0.972	0.985	0.987	0.989	0.991
42	0.922	0.930	0.942	0.951	0.972	0.985	0.987	0.989	0.991
43	0.923	0.932	0.943	0.951	0.973	0.985	0.987	0.990	0.991
44	0.924	0.933	0.944	0.952	0.973	0.985	0.987	0.990	0.991
45	0.926	0.934	0.945	0.953	0.973	0.985	0.988	0.990	0.991
46	0.927	0.935	0.945	0.953	0.974	0.985	0.988	0.990	0.991
47	0.928	0.936	0.946	0.954	0.974	0.985	0.988	0.990	0.991
48	0.929	0.937	0.947	0.954	0.974	0.985	0.988	0.990	0.991
49	0.929	0.937	0.947	0.955	0.974	0.985	0.988	0.990	0.991
50	0.930	0.938	0.947	0.955	0.974	0.985	0.988	0.990	0.991

SOURCE: Reprinted from Pearson and Hartley (1976), with permission from the Biometrika Trustees.

* The entries in this table are quantiles w_p of the Shapiro-Wilk test statistic given by Equation 6.2.9. Reject H_0 at the level p if $T_1 < w_p$.

Figura 4:

TABLE A18 A Method for Converting the Shapiro-Wilk Statistic to Approximate Normality

n	3	4	5	6	n	3	4	5	6
$v(d_n)$	(0.7500)	(0.6297)	(0.5521)	(0.4963)	$v(d_n)$	(0.7500)	(0.6297)	(0.5521)	(0.4963)
-7.0	-3.29	—	—	—	2.2	0.52	0.74	0.75	0.64
-5.4	-2.81	—	—	—	2.6	0.67	1.00	1.09	1.06
-5.0	-2.68	—	—	—	3.0	0.81	1.23	1.40	1.45
-4.6	-2.54	—	—	—	3.4	0.95	1.44	1.67	1.83
-4.2	-2.40	—	—	—	3.8	1.07	1.65	1.91	2.17
-3.8	-2.25	-3.50	—	—	4.2	1.19	1.85	2.15	2.50
-3.4	-2.10	-3.27	—	—	4.6	1.31	2.03	2.47	2.77
-3.0	-1.94	-3.05	-4.01	—	5.0	1.42	2.19	2.85	3.09
-2.6	-1.77	-2.84	-3.70	—	5.4	1.52	2.34	3.24	3.54
-2.2	-1.59	-2.64	-3.38	—	5.8	1.62	2.48	3.64	—
-1.8	-1.40	-2.44	-3.11	—	6.2	1.72	2.62	—	—
-1.4	-1.21	-2.22	-2.87	—	6.6	1.81	2.75	—	—
-1.0	-1.01	-1.96	-2.56	-3.72	7.0	1.90	2.87	—	—
-0.6	-0.80	-1.66	-2.20	-2.88	7.4	1.98	2.97	—	—
-0.2	-0.60	-1.31	-1.81	-2.27	7.8	2.07	3.08	—	—
0.2	-0.39	-0.94	-1.41	-1.85	8.2	2.15	3.22	—	—
0.6	-0.19	-0.57	-0.97	-1.38	8.6	2.23	3.36	—	—
1.0	0.00	-0.19	-0.51	-0.84	9.0	2.31	—	—	—
1.4	0.18	0.15	-0.06	-0.33	9.4	2.38	—	—	—
1.8	0.35	0.45	0.37	0.18	9.8	2.45	—	—	—

For $3 \leq n \leq 6$, first compute $v = \ln[(T - d_n)/(1 - T)]$ where d_n is given at the top of the table and T is the Shapiro-Wilk statistic. Then enter the table with v and n to find G , which is approximately normal.

Figura 5:

```
> a=c(5739,3291,2141,1224,399)/10000
>
> ###Calcular a diferença:
>
```

TABLE A18 (Continued)

n	b_n	c_n	d_n	n	b_n	c_n	d_n
7	-2.356	1.245	0.4533	29	-6.074	1.934	0.1907
8	-2.696	1.333	0.4186	30	-6.150	1.949	0.1872
9	-2.968	1.400	0.3900				
10	-3.262	1.471	0.3600	31	-6.248	1.965	0.1840
				32	-6.324	1.976	0.1811
11	-3.485	1.515	0.3451	33	-6.402	1.988	0.1781
12	-3.731	1.571	0.3270	34	-6.480	2.000	0.1755
13	-3.936	1.613	0.3111	35	-6.559	2.012	0.1727
14	-4.155	1.655	0.2969				
15	-4.373	1.695	0.2842	36	-6.640	2.024	0.1702
				37	-6.721	2.037	0.1677
16	-4.567	1.724	0.2727	38	-6.803	2.049	0.1656
17	-4.713	1.739	0.2622	39	-6.887	2.062	0.1633
18	-4.885	1.770	0.2528	40	-6.961	2.075	0.1612
19	-5.018	1.786	0.2440				
20	-5.153	1.802	0.2359	41	-7.035	2.088	0.1591
				42	-7.111	2.101	0.1572
21	-5.291	1.818	0.2264	43	-7.188	2.114	0.1552
22	-5.413	1.835	0.2207	44	-7.266	2.128	0.1534
23	-5.508	1.848	0.2157	45	-7.345	2.141	0.1516
24	-5.605	1.862	0.2106				
25	-5.704	1.876	0.2063	46	-7.414	2.155	0.1499
				47	-7.484	2.169	0.1482
26	-5.803	1.890	0.2020	48	-7.555	2.183	0.1466
27	-5.905	1.905	0.1980	49	-7.615	2.198	0.1451
28	-5.988	1.919	0.1943	50	-7.677	2.212	0.1436

SOURCE: Reprinted from Vol. 2 of Pearson and Hartley (1976), with permission from the Biometrika Trustees.
For $7 \leq n \leq 50$, enter the table above with n to find the coefficients b_n , c_n , and d_n . Then compute

$$G = b_n + c_n \ln \{(T - d_n)/(1 - T)\}$$
which is approximately standard normal.

Figura 6:

```
> D=X_d[1:5] - X[1:5]
> aD=a*D
> tab2=cbind(X_d[1:5],X[1:5],a,D,aD);tab2
              a      D      aD
[1,] 3.15435 1.42738 0.5739 1.72697 0.991108083
[2,] 2.61826 1.52229 0.3291 1.09597 0.360683727
[3,] 2.22488 1.69742 0.2141 0.52746 0.112929186
[4,] 2.10288 1.90642 0.1224 0.19646 0.024046704
[5,] 1.99568 1.98492 0.0399 0.01076 0.000429324
> S=sum(aD);S
[1] 1.489197
> num=S^2;num
[1] 2.217708
> W_cal=num/den;W_cal
[1] 0.9270088
>
>
> ####Comente a diferença entre os valores obtidos da estatística.
>
```


Vamos apresentar um exemplo do Conover na página 246.

Exemplo 2.2 *Uma amostra aleatória de tamanho 50 de números de dois algarismos de uma agenda de telefones:*

```
>
> Y=c(23,23,24,27,29,31,32,33,33,35,36,37,40,42,43,43,44,45,48,48,54,54,
56,57,57,58,58,58,58,59,61,61,62,63,64,65,66,68,68,70,73,73,74,75,77,81,
87,89,93,97)
>

> n=length(Y);n
[1] 50
>
> X=sort(Y);X
[1] 23 23 24 27 29 31 32 33 33 35 36 37 40 42 43 43 44 45 48 48
   54 54 56 57 57 58 58 58 58 59 61 61 62 63 64 65 66 68 68 70
   73 73 74 75 77 81 87 89 93 97
>
> X_d=sort(X,decreasing = TRUE)
> Xb=mean(X);Xb
[1] 55.04
>
> den=sum((X-Xb)^2);den
[1] 17697.92
>
> k=n/2;k
[1] 25
> a=c(3751,2574,2260,2032,1847,1691,1554,1430,1317,1212,1113,1020,932,
846,764,685,608, 532,459,386,314,244,174,104,35)/10000
>
> tab=cbind(a,X_d[1:25],X[1:25]);tab
      a
[1,] 0.3751 97 23
[2,] 0.2574 93 23
[3,] 0.2260 89 24
[4,] 0.2032 87 27
[5,] 0.1847 81 29
[6,] 0.1691 77 31
[7,] 0.1554 75 32
[8,] 0.1430 74 33
[9,] 0.1317 73 33
[10,] 0.1212 73 35
[11,] 0.1113 70 36
```

```
[12,] 0.1020 68 37
[13,] 0.0932 68 40
[14,] 0.0846 66 42
[15,] 0.0764 65 43
[16,] 0.0685 64 43
[17,] 0.0608 63 44
[18,] 0.0532 62 45
[19,] 0.0459 61 48
[20,] 0.0386 61 48
[21,] 0.0314 59 54
[22,] 0.0244 58 54
[23,] 0.0174 58 56
[24,] 0.0104 58 57
[25,] 0.0035 58 57
>
> D=X_d[1:25]-X[1:25]
>
> aD=a*D
>
> tab1=cbind(a,D,aD);tab1
      a  D      aD
[1,] 0.3751 74 27.7574
[2,] 0.2574 70 18.0180
[3,] 0.2260 65 14.6900
[4,] 0.2032 60 12.1920
[5,] 0.1847 52  9.6044
[6,] 0.1691 46  7.7786
[7,] 0.1554 43  6.6822
[8,] 0.1430 41  5.8630
[9,] 0.1317 40  5.2680
[10,] 0.1212 38  4.6056
[11,] 0.1113 34  3.7842
[12,] 0.1020 31  3.1620
[13,] 0.0932 28  2.6096
[14,] 0.0846 24  2.0304
[15,] 0.0764 22  1.6808
[16,] 0.0685 21  1.4385
[17,] 0.0608 19  1.1552
[18,] 0.0532 17  0.9044
[19,] 0.0459 13  0.5967
[20,] 0.0386 13  0.5018
[21,] 0.0314  5  0.1570
[22,] 0.0244  4  0.0976
```

```
[23,] 0.0174  2  0.0348
[24,] 0.0104  1  0.0104
[25,] 0.0035  1  0.0035
> S=sum(aD);S
[1] 130.6261
> num=S^2;num
[1] 17063.18
>
> W_cal=num/den;W_cal
[1] 0.9641347
>
> round(W_cal,4)
[1] 0.9641
>
```

Usando

$$0,955 < 0,964 < 0,974,$$

temos que:

$$0,10 < nd < 0,50.$$

Vamos fazer a interpolação na tabela A.17

```
>
>
> alfa1=0.10;alfa2=0.50
>
> w_1=0.955;w_2=0.974
>
> w=0.9642
> dalfa=alfa2-alfa1;dalfa
[1] 0.4
>
> dw=w_2-w_1;dw
[1] 0.019
>
>
>
> nd=0.10+ dalfa*(w-w_1)/dw;nd
[1] 0.2936842
>
>
> round(nd,2)
```

```
[1] 0.29
```

Resolvendo direto temos:

```
require(nortest)
> shapiro.test(Y)

Shapiro-Wilk normality test

data:  Y
W = 0.97444, p-value = 0.3473
```

Vamos apresentar um exemplo do livro do Peter Sprent e Nigel Smeeton *Applied Nonparametric Statistical Methods*, na página 92.

Exemplo 2.3 *Uma amostra aleatória da idade ao morrer de homens escoceses são dadas em ordem crescente:*

11	13	14	22	29	30	41	41	52	55	56	59	65	65	66
74	74	75	77	81	82	82	82	82	83	85	85	87	87	88

É razoável supor que as idades ao morrer são normalmente distribuídas?

```
> Y=c(11,13,14,22,29,30,41,41,52,55,56,59,65,65,66,74,74,75,77,81,82,82,
82,82,83,85, 85,87,87,88)
>
>
> shapiro.test(Y)

Shapiro-Wilk normality test

data:  Y
W = 0.86294, p-value = 0.001171
```

Há um forte indicativo que os dados não são normais.
Faça agora utilizando o passo a passo utilizado em sala.

```
> X=sort(Y);X
[1] 11 13 14 22 29 30 41 41 52 55 56 59 65 65 66 74 74 75 77 81 82
[26] 82 82 82 83 85 85 87 87 88
>
> X_d=sort(X,,decreasing = TRUE);X_d
>
[1] 88 87 87 85 85 83 82 82 82 82 81 77 75 74 74 66 65 65 59 56 55 52
>
[26] 41 41 30 29 22 14 13 11
>
>
> D=X_d-X;D
[1] 77 74 73 63 56 53 41 41 30 27 25 18 10 9 8 -8 -9 -10 -18
[20] -25 -27 -30 -41 -41 -53 -56 -63 -73 -74 -77
> n=length(Y);n
[1] 30
> k=n/2;k
[1] 15
>
> X_M=X_d[1:15]
> X_m=X[1:15]
>
> a=c(4254,2944,2487,2148,1870,1630,1415,1219,1036,862,697,537,381,227,76)/10000
> D_Mm=X_M-X_m
>
>
> AD=a*D_Mm
> S=sum(AD);S
[1] 124.9299
>
> num=S^2;num
[1] 15607.48
> Xb=mean(X);Xb
[1] 61.43333
> den=sum((X-Xb)^2);den
[1] 18187.37
>
>
> W_cal=num/den;W_cal
[1] 0.8581495
>
```

Note que o valor da estatística do teste 0,8581495 não bate com a saída da função `shapiro.test` 0,86294. Na interpolação para o nível descritivo

$$nd < 0,01,$$

para $w_{tab} = 0,900$.

3 Exercícios

Para cada exercício proposto faça duas soluções: A primeira usando uma função pronta do R . A segunda fazendo o passo a passo incluindo o uso das tabelas do Conover:

1. O retorno do investimento por 12 meses em 20 ações selecionadas aleatoriamente é o seguinte:

9,1	5,0	7,3	7,4	5,5
8,6	7,0	4,3	4,7	8,0
4,0	8,5	6,4	6,1	5,8
9,5	5,2	6,7	8,3	9,2

Teste a hipótese nula de Normalidade usando o teste de Shapiro-Wilk.

2. Quinze calouros do curso de Estatística tiveram as seguintes pontuações no ENEM.

481	620	642	515	740
562	395	615	596	618
525	584	540	580	598

Teste a hipótese nula de Normalidade usando o teste de Shapiro-Wilk.

3. Use o teste de Lilliefors para tratar a normalidade da questão 1.
4. Use o teste de Lilliefors para tratar a normalidade da questão 2.