Colloquium Biometricum 42
2012, 43—50

ADAPTATION OF SHAPIRO-WILK TEST TO THE CASE OF KNOWN MEAN

Zofia Hanusz[1], Joanna Tarasińska[1], Wojciech Zieliński[2]

[1]Department of Applied Mathematics and Computer Science
University of Life Sciences in Lublin
Akademicka 13, 20-950 Lublin
[2]Department of Econometrics and Statistics
Warsaw University of Life Sciences
Nowoursynowska 166, 02-787 Warszawa
e-mail: zofia.hanusz@up.lublin.pl; joanna.tarasinska@up.lublin.pl
wojtek.zielinski@statystyka.info

Summary

Shapiro-Wilk *W* test is widely used for checking normality of data. The paper considers its modification to the case of normality with known mean. The table with critical values of modified test for different sample sizes and several significance levels is given. An application for residuals in two-way ANOVA model is presented.

Keywords and phrases: Shapiro-Wilk *W* test, normality, two-way experimental layout

Classification AMS 2010: 62G10

1. Introduction

Shapiro and Wilk (1965) introduced the W test for normality based on statistic

$$W = \frac{\left(\sum_{i=1}^{n} a_i X_{(i)}\right)^2}{\sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2},$$

where $X_{(1)} \le X_{(2)} \le \ldots \le X_{(n)}$ are the ordered values of the sample $(X_1, X_2, \ldots, X_n)$ and a_i are tabulated coefficients. The *W* test is considered as very powerful for the hypothesis that a random variable *X* is normally distributed with unknown parameters μ and σ^2.

However, frequently we are interested in testing null hypothesis that distribution of *X* is normal with known expectation μ_0. Adaptation of the Shapiro-Wilk *W* test to the case of known mean is described in Section 2. In Section 3 we give two examples illustrating applications of the Shapiro-Wilk *W* test and its modification. Some concluding remarks are presented in Section 4.

2. Description of W_0 statistic

Let us consider the null hypothesis of the form:

H_0: *X* is normally distributed with a known expectation μ_0.

To test the H_0 hypothesis we propose modification of the Shapiro-Wilk *W* statistic in the following form

$$W_0 = \frac{\left(\sum_{i=1}^{n} a_i X_{(i)}\right)^2}{\sum_{i=1}^{n} \left(X_i - \mu_0\right)^2}.$$

The hypothesis H_0 is rejected at a significance level α if W_0 is less then the critical value $W_0(\alpha; n)$. The critical values of W_0 can be evaluated in simulation study. For each sample size of $n = 3, 4, \ldots, 50$; $N = 1{,}000{,}000$ pseudorandom samples from $N(0, 1)$ were generated and for each sample the value W_0 was calculated, so the sample $w_1, \ldots, w_N$ of values of the W_0 statistic were obtained. The critical value $W_0(\alpha; n)$ was taken as the α-th quantile of $w_1, \ldots, w_N$. All calculations were done independently in Mathematica and in R program. In program R we used the procedure "shapiro.test" in which Royston's procedure is applicated (Royston 1992; Hanusz, Tarasinska 2011). The results are given in Table 1.

Table 1. Critical values of W_0 statistic for sample sizes *n* and significance level α

n	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$	*n*	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$
3	0.0184	0.0881	0.1714	27	0.7379	0.8232	0.8601
4	0.0721	0.2037	0.3127	28	0.7463	0.8287	0.8645
5	0.1419	0.3086	0.4190	29	0.7539	0.8340	0.8688
6	0.2090	0.3867	0.4952	30	0.7611	0.8394	0.8730
7	0.2742	0.4525	0.5543	31	0.7677	0.8437	0.8765
8	0.3299	0.5051	0.5998	32	0.7746	0.8482	0.8800
9	0.3785	0.5493	0.6374	33	0.7804	0.8524	0.8834
10	0.4233	0.5852	0.6682	34	0.7871	0.8565	0.8863
11	0.4606	0.6165	0.6935	35	0.7917	0.8602	0.8894
12	0.4940	0.6431	0.7154	36	0.7969	0.8634	0.8921
13	0.5246	0.6661	0.7346	37	0.8008	0.8670	0.8947
14	0.5494	0.6862	0.7504	38	0.8063	0.8701	0.8972
15	0.5739	0.7038	0.7651	39	0.8109	0.8731	0.8996
16	0.5954	0.7196	0.7778	40	0.8145	0.8760	0.9018
17	0.6126	0.7337	0.7890	41	0.8194	0.8787	0.9040
18	0.6319	0.7476	0.7998	42	0.8227	0.8816	0.9061
19	0.6478	0.7590	0.8088	43	0.8271	0.8839	0.9081
20	0.6626	0.7696	0.8176	44	0.8301	0.8862	0.9100
21	0.6761	0.7792	0.8250	45	0.8343	0.8887	0.9120
22	0.6876	0.7875	0.8319	46	0.8374	0.8911	0.9138
23	0.7008	0.7965	0.8390	47	0.8403	0.8931	0.9154
24	0.7104	0.8034	0.8446	48	0.8433	0.8951	0.9169
25	0.7205	0.8103	0.8501	49	0.8470	0.8974	0.9187
26	0.7296	0.8170	0.8553	50	0.8491	0.8989	0.9200

The statistic W_0 has similar properties to the *W* statistic, namely, W_0 is scale invariant and the maximum of W_0 is one. However, the minimum of W_0 is zero, whereas the minimum of *W* is $\varepsilon = \frac{na_1}{n-1}$ (Shapiro and Wilk, 1965). It is sufficient to consider the maximization of $\sum_{i=1}^{n}(x_i-\mu_0)^2$ subject to $\sum_{i=1}^{n}a_i x_i=1$ and note that $\sum_{i=1}^{n}(x_i-\mu_0)^2$ may be arbitrarily large.

3. Application of W_0 test

The advantage of using of the W_0 test we illustrate with numerical examples. Let us consider a two-way experimental layout which involves two treatment factors *A* and *B*. Let us assume that the factor *A* has *a* levels $A_1, A_2, \ldots, A_a$, and the factor *B* has *b* levels $B_1, B_2, \ldots, B_b$. For each possible value of *i* $(i=1,\ldots,a)$ and *j* $(j=1,\ldots,b)$, let x_{ijk} be a *k*th observation of *X* $(k=1,\ldots,n)$ affected by levels A_i and B_j.

Let us assume that observations x_{ijk} fulfill the following model

$$x_{ijk}=\mu_{ij}+e_{ijk} \tag{3.1}$$

with $\mu_{ij}=\alpha_i+\beta_j+(\alpha\beta)_{ij}$, where α_i denotes an effect of *i*th level of *A*, β_j denotes an effect of *j*th level of *B*, $(\alpha\beta)_{ij}$ denotes an interaction between *i*th level of *A* and *j*th level of *B* ($i=1,\ldots,a$, $j=1,\ldots,b$, $k=1,\ldots,n$). We assume that e_{ijk} 's are independent $N(0,\sigma^2)$ variables. If the model (3.1) is adequate to the experimental data then for each combination (i,j) the residuals $\hat{e}_{ijk}=x_{ijk}-\overline{x}_{ij}$, where $\overline{x}_{ij}=\frac{1}{n}\sum_{k=1}^{n}x_{ijk}$, should be distributed as $N(0,\sigma^2)$.

When the interaction in model (3.1) is neglected, then the following model is considered:

$$x_{ijk}=\alpha_i+\beta_j+e_{ijk}\,. \tag{3.2}$$

In model (3.2), residuals are equal to $\hat{\hat{e}}_{ijk} = x_{ijk} - \bar{x}_{i\cdot} - \bar{x}_{\cdot j} + \bar{x}_{ij}$, where $\bar{x}_{i\cdot} = \frac{1}{b}\sum_{j=1}^{b}\bar{x}_{ij}$ and $\bar{x}_{\cdot j} = \frac{1}{a}\sum_{i=1}^{a}\bar{x}_{ij}$.

Example 1. Let us consider an experiment with two levels of a factor *A* $(a = 2)$ and three levels of a factor *B* $(b = 3)$. One set of data with $n = 10$ replications was generated according to model (3.1) with $\mu_{11} = 1$, $\mu_{12} = 2$, $\mu_{13} = 3$, $\mu_{21} = 2$, $\mu_{22} = 4$, $\mu_{23} = 4$ and $\sigma^2 = 1$. The results of analysis of variance are given in Table 2.

Table 2. The results of analysis of variance for model (3.1)

Source	d.f.	Sum of Squares	Mean Square	*F* Test	*p*-value
A	1	14.2789	14.2789	15.3288	0.00026
B	2	31.5567	15.7783	16.9385	0.00002
*A*x*B*	2	4.5072	2.2536	2.4193	0.09857
Error	54	50.3015	0.9315		

Thus, for our data the interaction between factors turned out to be insignificant. In spite of the fact that with given μ_{ij} 's interaction was involved in the model.

If we consider model (3.2) then we get the results given in Table 3.

Table 3. The results of analysis of variance for model (3.2)

Source	d.f.	Sum of Squares	Mean Square	*F* Test	*p*-value
A	1	14.2789	14.2789	14.5892	0.00034
B	2	31.5567	15.7783	16.1213	0.000003
Error	56	54.8087	0.9787		

Now, we will focus on checking whether for each combination (i, j) the residuals in both models (3.1) and (3.2) are normally distributed with null mean i.e. the hypothesis that residuals are $N(0, \sigma^2)$ should be verified. Results, rounded to the third decimal place, are given in Table 4. The values of *W* are the same in (3.1) and (3.2) models. In model (3.1) they are also the same as W_0 values.

For $n = 10$ and $\alpha = 0.05$, critical value of the W_0 test is equal to 0.585 (see Table 1) and of the Shapiro-Wilk *W* test is equal to 0.8449 (Hanusz, Tarasinska, 2011). In the case of model (3.2), the hypothesis of normality with null mean was rejected for observations from the cell A_1B_1 (bold number), while in the case of model (3.1) was never rejected. Let us notice that the Shapiro-Wilk W test never rejected the normality of residuals in both models.

Table 4. Results of checking normality for residuals in ANOVA.

Cell	A_1B_1	A_1B_2	A_1B_3	A_2B_1	A_2B_2	A_2B_3
Residuals in model (3.1)	-0.382	1.321	1.573	-0.932	-0.484	-1.160
	-0.044	0.982	0.756	-0.090	1.337	0.621
	0.798	-1.525	0.725	-0.123	0.375	0.731
	-0.314	-0.616	-0.550	-0.694	0.717	0.141
	0.404	0.357	0.177	1.132	-1.211	0.736
	0.218	-2.714	-0.561	0.947	-0.609	-0.406
	-0.259	1.554	0.396	0.658	-0.505	1.027
	-1.196	-0.458	-0.255	0.391	-0.018	-0.501
	0.112	-1.010	-2.865	-0.748	0.350	-0.408
	-0.339	2.109	0.604	-0.541	0.049	-0.783
W_0	0.891	0.968	0.871	0.924	0.982	0.925
W	0.891	0.968	0.871	0.924	0.982	0.925
Residuals in model (3.2)	-0.036	0.996	1.553	-1.277	-0.159	-1.139
	0.302	0.658	0.736	-0.436	1.662	0.642
	1.143	-1.850	0.704	-0.468	0.700	0.752
	0.032	-0.941	-0.571	-1.039	1.042	0.162
	0.750	0.032	0.156	0.787	-0.886	0.757
	0.564	-3.039	-0.582	0.601	-0.285	-0.385
	0.087	1.229	0.376	0.312	-0.180	1.048
	0.150	-0.783	-0.276	0.046	0.307	-0.480
	0.458	-1.334	-2.886	-1.093	0.675	-0.387
	0.007	1.784	0.583	-0.887	0.374	-0.762
W_0	**0.469**	0.922	0.871	0.746	0.807	0.924
W	0.891	0.968	0.871	0.924	0.982	0.925

Example 2. In this example we consider similar model as in Example 1, just taking $\mu_{22} = 5$ instead of $\mu_{22} = 4$. A set of data with $n = 10$ replications was generated. The results of analysis of variance for model (3.1) are presented in Table 5.

In this example, the interaction between factors is significant. The results of testing normality of the residuals for each combination (i, j) for models (3.1) and (3.2) are presented in Table 6.

Table 5. The results of analysis of variance for model (3.1)

Source	d.f.	Sum of Squares	Mean Square	*F* Test	*p*-value
A	1	31.4958	31.4958	64.1261	$9.58\ 10^{-11}$
B	2	83.8075	41.9038	85.3171	$1.93\ 10^{-11}$
*A*x*B*	2	20.3664	10.1832	20.7332	$2.08\ 10^{-7}$
Error	54	26.5223	0.4912		

Table 6. Results of checking normality for residuals in ANOVA.

Cell	A_1B_1	A_1B_2	A_1B_3	A_2B_1	A_2B_2	A_2B_3
Residuals in model (3.1)	-1.264	0.102	-0.478	-0.019	0.230	0.672
	-0.468	-0.115	-0.410	0.019	-0.366	-1.193
	0.410	-0.670	0.037	0.143	-0.364	-0.056
	-0.191	0.490	0.472	-0.576	0.516	-0.434
	-0.568	0.493	-0.718	2.006	-1.257	0.377
	0.040	-0.931	0.555	-0.214	0.471	0.232
	0.923	0.579	0.086	-0.996	-0.009	0.249
	0.007	1.291	0.255	1.103	0.421	0.619
	0.351	-1.647	0.123	-0.024	0.425	-0.082
	0.761	0.436	0.074	-1.440	-0.067	-0.384
W_0	0.973	0.942	0.931	0.934	0.853	0.930
W	0.973	0.942	0.931	0.934	0.853	0.930
Residuals in model (3.2)	-0.976	-0.710	0.046	-0.308	1.042	0.148
	-0.179	-0.928	0.114	-0.270	0.447	-1.717
	0.699	-1.512	0.561	-0.146	0.448	-0.580
	0.098	-0.322	0.996	-0.865	1.329	-0.958
	-0.280	-0.319	-0.194	1.717	-0.444	-0.147
	0.328	-1.743	1.079	-0.503	1.283	-0.292
	1.211	-0.234	0.610	-1.285	0.803	-0.275
	0.295	0.479	0.779	0.814	1.234	0.094
	0.639	-2.460	0.652	-0.312	1.238	-0.608
	1.050	-0.376	0.598	-1.729	0.746	-0.908
W_0	0.802	**0.477**	**0.334**	0.852	**0.251**	**0.474**
W	0.973	0.942	0.931	0.934	0.853	0.930

For the model (3.1) both tests W_0 and W did not reject the null hypothesis about normality. However, when we consider not adequate model (3.2), the W_0 test rejected normality of residuals in four cells (bold numbers), while the Shapiro-Wilk W test never rejected the normality of residuals.

4. Conclusions

In the paper some modification of the Shapiro-Wilk W test for testing normality is proposed. This test should be applied when the mean of random variable is known. In general, suggested test should be recommended for testing normality of residuals, when observation are affected by some factors. In the paper we show by the examples that the W_0 test can reject normality when data does not fulfill the theoretical model, contrary to the Shapiro-Wilk W test which does not reject normality in such a situation.

References

Hanusz Z., Tarasinska J. (2011). Tables for Shapiro-Wilk W statistic according to Royston approximation. *Colloquium Biometricum* 41, 211–219.

Royston P. (1992). Approximating the Shapiro-Wilk W-test for non-normality. *Statistics and Computing* 2, 117–119.

R Development Core Team (2008). R: *A language and environment for statistical computing.* R Foundation for Statistical Computing. Vienna, Austria. ISBN 3–900051–07–0, URL http://www.R-project.org.

Shapiro S.S., Wilk M.B. (1965). An analysis of variance test for normality (complete samples). *Biometrika* 52, 591–611.