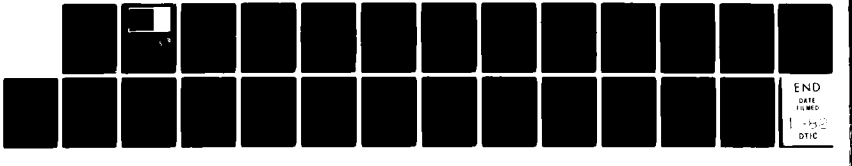
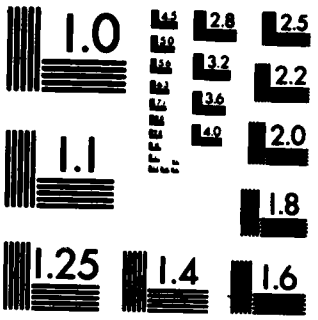
AD-A120 997 TESTING THE NORMALITY OF RESIDUALS(U) WISCONSIN 1/1
UNIV-MADISON MATHEMATICS RESEARCH CENTER
N R DRAPER ET AL. SEP 82 MRC-TSR-2426 DAAG29-80-C-0041
UNCLASSIFIED F/G 12/1 NL





MRC Technical Summary Report # 2426

TESTING THE NORMALITY OF RESIDUALS

N. R. Draper and J. A. John

Mathematics Research Center
University of Wisconsin—Madison
610 Walnut Street
Madison, Wisconsin 53706

September 1982

(Received August 4, 1982)

Approved for public release
Distribution unlimited

Sponsored by

U. S. Army Research Office
P. O. Box 12211
Research Triangle Park
North Carolina 27709

UNIVERSITY OF WISCONSIN - MADISON
MATHEMATICS RESEARCH CENTER

TESTING THE NORMALITY OF RESIDUALS

N. R. Draper and J. A. John*

Technical Summary Report #2426

September 1982

ABSTRACT

The use of residuals to test the assumption of normality of the errors in a linear model is considered. Standard tests for normality typically require an assumption of independence; however the residuals are correlated. An investigation of the Shapiro-Wilk test shows that it is affected by these correlations, but the problem can be overcome by a simple adjustment to the test procedure.

AMS (MOS) Subject Classifications: 62J05, 62J10, 62J99

Key Words: Linear model, Residuals, Shapiro-Wilk test, Tests of normality

Work Unit Number 4 - Statistics and Probability

* Mathematics Department, University of Southampton, SO9 - 5NH, U.K.

Sponsored by the United States Army under Contract Nos. DAAG29-80-C-0041 and DAAG29-80-C-0113.

SIGNIFICANCE AND EXPLANATION

In a regression situation, the residuals provide (correlated and often unequal variance) estimates of the errors in the postulated model. We anticipate that, if the errors are normally distributed as is usually assumed, the residuals will show a "similar" behaviour. Thus the departure from, or consonance with, normality of the residuals is of interest. Standard tests for normality typically require an assumption of independence; even if the residuals are standardized to be of equal variance, they will be correlated. We discuss various test possibilities and then focus on the well-known Shapiro-Wilk test. Although it is affected by the correlations, this difficulty can be overcome by a simple adjustment to the test procedure.

The responsibility for the wording and views expressed in this descriptive summary lies with MRC, and not with the authors of this report.

TESTING THE NORMALITY OF RESIDUALS

N. R. Draper and J. A. John*

1. INTRODUCTION

Consider the linear model

$$\underset{\sim}{y} = \underset{\sim}{X}\underset{\sim}{\beta} + \underset{\sim}{\varepsilon} \tag{1.1}$$

where $\underset{\sim}{y}$ is an $n \times 1$ vector of observations, $\underset{\sim}{X}$ is an $n \times p$ matrix of specified predictor variables, $\underset{\sim}{\beta}$ is a $p \times 1$ vector of parameters and $\underset{\sim}{\varepsilon}$ is an $n \times 1$ vector of unknown errors assumed to be $N(\underset{\sim}{0}, \sigma^2\underset{\sim}{I})$. The vector of residuals obtained from a least squares (LS) fit of (1.1) is given by

$$\underset{\sim}{e} = (\underset{\sim}{I} - \underset{\sim}{R})\underset{\sim}{y} \tag{1.2}$$

where $\underset{\sim}{R} = \underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}$. Examination of these residuals to check the basic model assumptions is essential. With the advent of large scale computing, a huge literature has grown up on ways to examine and test the residuals. For basic references see, for example Seber (1977), Barnett and Lewis (1978), Belsley, Kuh and Welsch (1980), Draper and Smith (1981) and Hawkins (1980).

In this paper, the use of residuals to test the assumption of the normality of errors is examined. There are several tests for normality in use, but typically they require the assumption of independence. These tests are considered in the next section. However, they may not be appropriate when applied to LS residuals since these residuals are not independent. They are based on $(n - p)$ rather than n degrees of freedom and are correlated, their variance

* Mathematics Department, University of Southampton, S09 - 5NH, U.K.

Sponsored by the United States Army under Contract Nos. DAAG29-80-C-0041 and DAAG29-80-C-0113.

covariance matrix being $V(\underset{\sim}{e}) = (\underset{\sim}{I} - \underset{\sim}{R})\sigma^2$. This lack of independence raises a number of interesting questions. To what extent are the standard normality tests affected by the presence of correlations amongst the residuals? Can these tests be amended or are new tests necessary? Alternatively, can a subset of the $(n - p)$ residuals or, more generally, $(n - p)$ linearly and statistically independent linear functions of the residuals be found to which the standard tests can be applied?

A number of ways of transforming residuals to independence have been proposed and some of the methods are described in section 3. For a number of reasons we feel that such transformed residuals are not, in general, entirely suitable for testing normality. Instead, an investigation into the use of one of the standard tests, namely the Shapiro-Wilk test, on the LS residuals (1.2) is reported in the remaining sections. Our conclusion is that this test, properly interpreted, gives a suitable basis for testing the normality of residuals.

2. TESTS FOR NORMALITY

A number of tests have been proposed to check for normality. A popular one that has stood up well in various comparative investigations is due to Shapiro and Wilk (1965, 1968); see formula (4.1). Extensive comparisons of the Shapiro-Wilk statistic with competitors were made by Shapiro, Wilk and Chen (1968). They concluded that it

was a "generally superior omnibus measure of non-normality". Similar conclusions transfer to a modification of the Shapiro-Wilk test statistic for sample sizes exceeding 50, described by Shapiro and Francia (1972). Other investigations favouring the Shapiro-Wilk statistic are given in Dyer (1974) and Huang and Bolch (1974).

For samples of size 50 or more, D'Agostino (1971) suggests a statistic D which is "up to a constant the ratio of Downton's linear unbiased estimator of the population standard deviation to the sample standard deviation". The null distribution of D can be approximated by Cornish-Fisher expansions.

D'Agostino and Rosman (1974) in an investigation of Geary's computationally simple test based on the ratio of the mean deviation to the standard deviation conclude that it is a possibly useful test but that "there appears to be no specific situation where Geary's test clearly and for practical purposes dominates all other tests...".

Spiegelhalter (1977, 1980) investigates an omnibus test for normality "based on the posterior probability of the normal shape" under various assumptions about the sampled population. Power comparisons with other tests for $n = 20$ and 50 show "good overall performance when $n = 20$" with a drop in power "against moderately asymmetric alternatives for $n = 50$".

Lin and Mudholkar (1980) offer a test statistic Z which makes use of the fact that it is only for the normal distribution that the mean and variance are distributed independently. This test, against symmetric alternatives, is summarized by Nelson (1981) who also provides a table of critical values.

Martinez and Inglewicz (1981) suggest a test statistic which is the ratio of two estimators of variance and make a power comparison in which the "proposed test outperforms all considered competitors for long-tailed symmetric alternatives and performs well for all other cases considered". However, it is not universally superior to all other tests.

3. UNCORRELATED TRANSFORMED RESIDUALS

The residuals $\underset{\sim}{e} = (\underset{\sim}{I} - \underset{\sim}{R})\underset{\sim}{y}$ are linear combinations of the observations and are based on $n - p$ degrees of freedom. Since $\underset{\sim}{I} - \underset{\sim}{R}$ is of rank $n - p$ it is possible to find an $(n - p) \times n$ matrix $\underset{\sim}{A}$ of rank $n - p$ such that

$$\underset{\sim}{e}_a = \underset{\sim}{A}\underset{\sim}{e} \tag{3.1}$$

is $N_{n-p}(\underset{\sim}{0}, \sigma^2\underset{\sim}{I})$. The vector $\underset{\sim}{e}_a$ thus consists of $(n - p)$ uncorrelated homoscedastic residuals. The matrix $\underset{\sim}{A}$ necessarily satisfies $\underset{\sim}{A}\underset{\sim}{X} = \underset{\sim}{0}$ and $\underset{\sim}{A}\underset{\sim}{A}' = \underset{\sim}{I}$.

There are, however, many possible choices for $\underset{\sim}{A}$. The overall aim is to make a "good" choice which will facilitate some stated objective. The following choices have been suggested:

(a) Theil (1965) proposed the use of Best Linear Unbiased Scaled (BLUS) residuals. Let the vector $\underset{\sim}{J}\underset{\sim}{\varepsilon}$ contain $(n - p)$ selected components of $\underset{\sim}{\varepsilon}$; $\underset{\sim}{J}$ is a submatrix of $\underset{\sim}{I}_n$. Then BLUS residuals are obtained by choosing $\underset{\sim}{A}$ so that the expected sum of squares of the discrepancies in the vector $\underset{\sim}{e}_a - \underset{\sim}{J}\underset{\sim}{\varepsilon}$ is minimised, that is, so that

$$E\{(\underset{\sim}{A}\underset{\sim}{e} - \underset{\sim}{J}\underset{\sim}{\varepsilon})' (\underset{\sim}{A}\underset{\sim}{e} - \underset{\sim}{J}\underset{\sim}{\varepsilon})\}$$

is minimised. For a succinct, excellent account, see Grossman and Styan (1972); alternatively see Dent and Styan (1978). For a program which computes BLUS residuals, see Farebrother (1976). For related work, see Koerts (1967) and Abrahamse and Koerts (1971)

(b) Tiao and Guttman (1967) suggested augmenting $\underset{\sim}{e}_a$ in (3.1) by p independent random variables $z_1, z_2, \ldots, z_p$ which are uncorrelated with $\underset{\sim}{\varepsilon}$ and such that $E(\underset{\sim}{z}) = \underset{\sim}{0}$, $V(\underset{\sim}{z}) = \sigma^2\underset{\sim}{I}$. If $(\underset{\sim}{K}, \underset{\sim}{J})$ and $\underset{\sim}{G}$ are $n \times n$ permutation and orthogonal matrices respectively, then a general form of such residuals can be written (see Dent and Styan, 1978) as

$$\underset{\sim}{e}_g = \underset{\sim}{G}(\underset{\sim}{J}\underset{\sim}{e}_a + \underset{\sim}{K}\underset{\sim}{z})$$

Following Hildreth (1971), the choice of $\underset{\sim}{G}$ to minimise the trace of the variance covariance matrix of $\underset{\sim}{e}_g - \underset{\sim}{\varepsilon}$ leads to the so-called Best Augmented Unbiased Scaled (BAUS) residuals. The $\underset{\sim}{e}_a$ residual vector is said to be "unaugmented", the $\underset{\sim}{e}_g$ is "augmented". For additional details see Dent and Styan (1978).

For properties of uncorrelated transformed residuals, see also Godolphin and Tullio (1978).

Other alternative approaches are as follows:

(c) Hedayat and Robson (1970) and Brown, Durbin and Evans (1975) have suggested the use of "stepwise" (or recursive) residuals; see also Farebrother (1978). The uth $(u \geq p)$ such residual is calculated

as the deviation of the uth observation from its predicted value based on a LS fit to only the first u observations, normalized to have variance σ^2. The n - p stepwise residuals are not only mutually independent and homoscedastic but are also independent of all the calculated regression functions. They are not, of course, a transformed set of overall LS residuals and their vector is not restricted to the same space as the vector $\underline{e}$ but they are clearly linear combinations of the observations. Moreover, each stepwise residual has a clear identification with one point of the design. However, the set of stepwise residuals obtained is entirely dependent on the order selected for the entry of the observation. Hedayat and Robson argue that the ordered fitted values might be used to specify the ordering. Brown et. al. suggest the use of such residuals in time series where there is a natural ordering.

(d) Gentleman and Wilk (1975) and John and Draper (1978) have suggested the use of "adjusted" residuals as an aid to the detection of outliers. The procedure is as follows. Obtain the residuals from a LS analysis of all data points and choose one of them, normalized to have variance σ^2, as the first adjusted residual. Then the point corresponding to this residual is deleted and the procedure repeated on the reduced data set to give a second adjusted residual; again normalized to have variance σ^2. A second point is then deleted from the data set, and so on. This procedure is repeated until n - p

adjusted residuals have been obtained; the remaining p residuals will then be zero. John and Draper have shown that a missing value procedure can be used as an alternative to deleting observations. The resulting adjusted residuals are homoscedastic and uncorrelated. Observations can be selected for deletion in any order; the order used in outlier detection is in terms of a decreasing modulus size of residuals obtained when the observation for the largest residual at the previous stage is deleted.

Note the reciprocity between stepwise and adjusted residuals; they are based on forward and backward selection procedures respectively.

The standard tests of normality, discussed in the previous section, could be applied to any of the uncorrelated transformed residuals given above. For a number of reasons it was decided that such procedures would not, in general, be particularly appropriate.

The arbitrariness of transformed residuals is one of the main drawbacks. For testing normality, there appears to be no reason in general why one particular $\underset{\sim}{A}$ matrix in (3.1) should be preferred to any other matrix. If the order of observations is predetermined, as in time series, then the use of stepwise or adjusted residuals may be appropriate. However, in other cases the use of such residuals is dubious especially as the resulting residuals may not be normally distributed even when the error distribution is normal. This is certainly the case, for example, when adjusted residuals are used with an ordering based on a decreasing modulus size of residuals.

Further, the fact that the residuals are uncorrelated does not imply that they are independent in general. Huang and Bolch (1974) point this out for BLUS residuals and add that "... the lack of independence among BLUS residuals when the disturbances [errors] are not normal may be as least as great as the lack of independence among [ordinary] LS residuals". In other related work, Ramsey (1969) and Ramsey and Gilbert (1972) investigate tests for detection of regression specification errors such as omitted variables, incorrect functional forms, simultaneous equation problems and heteroscedasticity. Ramsey and Gilbert note that "... there does not seem to be any universal simple solution to the problem of choosing between [BLUS] and [ordinary] LS residuals".

Another disadvantage of the transformed residuals is the computational burden involved in calculating them. Tests for normality based on such residuals would have to enjoy considerable benefits over tests using the LS residuals for this disadvantage to be overcome.

Hence, in this paper, we have investigated the possibility of applying a standard test, namely the Shapiro-Wilk test, to the LS residuals. As is shown in the next section, such a test appears to be appropriate if the test percentage points are modified by making a simple adjustment. The virtues of this procedure are that it is easy to carry out, needs no additional tables, and its accuracy seems adequate for the situations we have investigated.

4. SIMULATION RESULTS FROM SELECTED EXPERIMENTAL SITUATIONS

A simulation study was carried out on a number of selected experimental situations to assess the appropriateness of using the Shapiro-Wilk test to examine the normality of LS residuals. Table 1 shows results obtained using two-way r x c tables. $N = rc$ standard normal variates were generated and residuals $e_i (i = 1, \ldots, N)$ evaluated assuming the usual additive model. The Shapiro-Wilk statistic

$$W = \left\{ \sum_i a_{N-i+1} (e_{N-i+1} - e_i) \right\}^2 / \sum e_i^2 \qquad (4.1)$$

was then calculated, where the a's are constants given by Shapiro and Wilk (1965). For each r x c table, this procedure was repeated 3,000 times. Note that, in general, standardized residuals should be used in (4.1) to give uniform variances, but this is not necessary here since all residuals are estimated with equal precision. As a check on the procedure and calculations, the Shapiro-Wilk statistic was also calculated for the errors ε_i themselves. The 5% and 10% percentage points of the W statistic for N observations, given in Shapiro and Wilk (1965), were then used to obtain the proportion of sampled W values falling in the corresponding tail area. For the errors ε_i these proportions should be exactly 5% and 10%; any discrepancies reflecting sampling error. It can be seen from Table 1 that these results are satisfactory. For the residuals e_i on the other hand the proportions are clearly too small. The tables of percentage points can, however, be

examined to determine what value of N should have been used in order to produce the correct tail areas. In table 1, these values of N which lead to the correct 5% and 10% points are denoted by N_5 and N_{10} respectively. This means that for the 3 x 5 two-way table, for example, if the Shapiro-Wilk test had been applied at the 5% level with $N_5 = 18$, instead of $N = 15$, an appropriate test would have been made. Similarly, $N_{10} = 17.5$ would have been appropriate for a 10% level. Note that $N_5 > N$ and $N_{10} > N$ for all cases in Table 1.

Table 2 shows a parallel set of figures for the residuals e_i from main effect models for various types of factorial designs as indicated by the first column. Again $N_5 > N$ and $N_{10} > N$ in most but (perhaps surprisingly) not all cases. Table 3 examines the case where two factor interactions are estimated as well, thus further reducing the residual degrees of freedom. The results are in line with the previous tables except for the cases $2^2 \times 4$, $2^2 \times 5$ and $2^2 \times 6$ which were found to be completely anomalous; the 3,000 simulations produced many more small W values than did neighbouring cases. Close examination of these anomalous cases revealed no assignable cause for this puzzling and atypical behaviour. Additional computions were made for the $2^2 \times 7$ and $2^2 \times 8$ cases and these confirmed the anomaly, as shown in Table 3. These cases are excluded from the discussion in the next section.

Table 4 provides parallel results and a broadly similar pattern for a series of rotatable response surface designs. These consist of a 2^k factorial ($k = 3, 4$) or a 2^{k-1} fractional

Table 1. Generations for selected r x c two-way tables showing the results of applying the Shapiro-Wilk normality test to errors, and to residuals from an additive main effects model without interactions.

DESIGN	N	5% POINTS			10% POINTS		
		% less than 5% ERRORS	% less than 5% RESIDUALS	N_5	% less than 10% ERRORS	% less than 10% RESIDUALS	N_{10}
3 x 5	15	.052	.023	18	.107	.059	17.5
4 x 4	16	.056	.028	19	.108	.064	18
3 x 6	18	.055	.021	22	.100	.059	21
4 x 5	20	.055	.033	22	.104	.076	22
3 x 7	21	.049	.020	26	.096	.054	25.5
3 x 8	24	.058	.028	29	.111	.068	27
4 x 6	24		.032	27		.078	26
5 x 5	25	.050	.028	29	.093	.066	28
3 x 9	27	.059	.019	33	.109	.063	32
4 x 7	28	.061	.035	32	.110	.072	32
3 x 10	30	.051	.023	36	.104	.073	34
5 x 6	30		.035	33		.089	32
4 x 8	32	.059	.036	35	.114	.083	34
5 x 7	35	.048	.038	38	.095	.086	37
4 x 9	36	.043	.034	38	.094	.087	38
6 x 6	36		.037	39		.088	38
4 x 10	40	.045	.025	45	.093	.078	44
5 x 8	40		.046	41		.094	41
6 x 7	42	.047	.048	43	.094	.099	43
5 x 9	45	.043	.042	47	.096	.096	46
6 x 8	48	.048	.043	>50	.102	.087	>50
7 x 7	49		.037	>50		.082	>50

Table 2. Generations for selected factorial designs showing the results of applying the Shapiro-Wilk normality test to the residuals from an additive main effects model without interactions.

DESIGN	N	5% POINTS		10% POINTS	
		Proportion less than 5%	N_5	Proportion less than 10%	N_{10}
2^4	16	.040	17	.087	17
$2^2 \times 4$	16	.027	18	.077	17.5
2×3^2	18	.039	19	.087	19
$2^2 \times 5$	20	.037	22	.081	21
$2^3 \times 3$	24	.046	25	.092	25
$2 \times 3 \times 4$	24	.040	26	.095	25
$2^2 \times 6$	24	.033	27	.084	26
3^3	27	.050	27	.108	26.5
$2 \times 3 \times 5$	30	.037	33	.084	32
$2^3 \times 4$	32	.039	35	.088	34
2×4^2	32	.042	34	.084	34
$2^2 \times 3^2$	36	.036	38	.083	38
$3^2 \times 4$	36	.038	39	.083	38
$2 \times 3 \times 6$	36	.039	38	.086	38
$2^3 \times 5$	40	.049	41	.095	41
$2 \times 4 \times 5$	40	.046	41	.095	40
$3^2 \times 5$	45	.044	47	.098	45
$2^2 \times 3 \times 4$	48	.058	47	.106	46
3×4^2	48	.052	48	.109	46
$2 \times 4 \times 6$	48	.051	48	.096	48
$2^3 \times 6$	48	.042	> 50	.092	49

Table 3. Generations for selected factorial designs showing the results of applying the Shapiro-Wilk normality test to the residuals from an additive main effects model with first order interactions.

DESIGN	N	Residual d.f. ν	5% POINTS		10% POINTS	
			Proportion less than 5%	N_5	Proportion less than 10%	N_{10}
2^4	16	5	.033	19	.061	20
$2^2 \times 4$	16	3	.289	6	.399	6
2×3^2	18	4	.024	22	.054	22
$2^2 \times 5$	20	4	.294	9	.412	8
$2^3 \times 3$	24	9	.025	29	.058	30
$2 \times 3 \times 4$	24	6	.026	29	.064	28
$2^2 \times 6$	24	5	.289	12	.439	12
3^3	27	8	.029	34	.057	35
$2^2 \times 7$	28	6	.322	13	.459	13
$2 \times 3 \times 5$	30	8	.038	33	.084	33
$2^3 \times 4$	32	13	.027	38	.061	38
2×4^4	32	9	.051	32	.089	35
$2^2 \times 8$	32	7	.331	15	.449	15
$2^2 \times 3^2$	36	16	.022	44	.056	43
$3^2 \times 4$	36	12	.027	44	.055	44
$2 \times 3 \times 6$	36	10	.045	37	.088	37
$2^3 \times 5$	40	17	.037	43	.079	44
$2 \times 4 \times 5$	40	12	.054	39	.097	41
$3^2 \times 5$	45	16	.032	>50	.069	>50
$2 \times 3 \times 4$	48	23	.036	>50	.079	>50
3×4^2	48	18	.028	>50	.069	>50
$2 \times 4 \times 6$	48	15	.062	46	.100	48
$2^3 \times 6$	48	21	.044	50	.089	50

Table 4. Generations for selected second order rotatable composite designs showing the results of applying the Shapiro-Wilk normality test to residuals from a full second order model fit.

No of variables	No of Centre Pts.	No of Observ.	Residual d.f.	At Shapiro-Wilk			
				5% point		10% point	
k	n_o	N	ν	Observed %	N_5	Observed %	N_{10}
3	2	16	6	.041	17½	.070	18
	4	18	8	.042	19	.078	21
	6	20	10	.047	20½	.081	22
	8	22	12	.047	22½	.082	24
	10	24	14	.043	25	.081	26
4	3	27	12	.026	34	.061	33
	4	28	13	.025	35	.060	35
	6	30	15	.032	34	.075	34
	8	32	17	.029	38	.064	38
	10	34	19	.032	38	.071	38
5	2	28	7	.026	36	.052	40
	4	30	9	.032	37	.058	40
	6	32	11	.032	40	.055	42
	8	34	13	.028	42	.054	42
6	1	45	17	.042	49	.083	49
	2	46	18	.038	> 50	.082	50
	4	48	20	.036	> 50	.082	> 50

factorial ($k = 5, 6$) with points coded as $(\pm 1, \pm 1, \ldots, \pm 1)$ plus $2k$ axial points at a distance $2^{(k-p)/4}$ from the centre, where $p = 0$ for $k = 3,4$ and $p = 1$ for $k = 5,6$, plus n_o centre points (as tabulated). Again the results in tables 2 - 4 are based on 3,000 simulations.

5. APPROXIMATIONS TO N_5 AND N_{10} AND RECOMMENDATIONS

The calculations in Section 4 indicate that the Shapiro-Wilk test for normality is incorrect when applied in the usual fashion to correlated residuals but that the effects can be adjusted for by using the Shapiro-Wilk percentage points with values N_5 and N_{10} rather than with N. These alternative values are somewhat higher than N so that, with no adjustment to N, the presence of correlations among the residuals means that the assumption for normality will be rejected less often for independent samples of N observations. That is, use of the Shapiro-Wilk test without adjustment will usually result in a more conservative test.

An important but difficult question is whether there is an attributable pattern to results of this type, in general. For an approximation it seems sensible to seek an adjustment to N which decreases to zero as the residual degrees of freedom ν tends to N, so that no adjustment would be made in the limiting (but impractical) case where the model contained no parameters and so the residuals were independent. For that reason formulas of the type

$$\hat{N} = N + \theta(1 - \nu/N) \qquad (5.1)$$

were examined (amongst others). It turned out that the (rather appealing) use of $\theta = 5$ for $\hat{N}_5$ and $\theta = 10$ for $\hat{N}_{10}$ provided an adequate approximation considering the sampling error that occurs naturally in the generations. Table 5 shows the probability levels that would have been obtained, for the generations of Table 1, had the $\hat{N}_5$ and $\hat{N}_{10}$ values of (5.1) been employed. The agreement is excellent for $\hat{N}_{10}$, less so (but conservative) for $\hat{N}_5$. Similarly conclusions are obtained from the generations in Tables 2 - 4.

As an overall practical recommendation for this work, we thus give the following rule of thumb:

<u>To check the normality of a set of N standardized residuals from a regression model, apply the Shapiro-Wilk test but use the percentage point for $\hat{N}_\theta = N + \theta(1 - \nu/N)$, where ν is the residual degrees of freedom and $\alpha = \theta/100$ is the desired significance level.</u>

Table 5. Actual tail proportions found (for the generations of Table 1) lying below the Shapiro-Wilk α percentage points when values $\hat{N}_\theta = N + \theta(1 - \nu/N)$ are used instead of N in the Shapiro-Wilk tables for the $\alpha = \theta/100$ cases, $\theta = 5$ and 10. (Ideally the values in the last two columns should be 0.050 and 0.100 respectively).

Design	Tail areas achieved by using $\hat{N}_5$	$\hat{N}_{10}$
3 x 5	.043	.092
4 x 4	.045	.100
3 x 6	.034	.082
4 x 5	.049	.104
3 x 7	.030	.078
3 x 8	.035	.083
4 x 6	.042	.094
5 x 5	.041	.091
3 x 9	.027	.072
4 x 7	.040	.086
3 x 10	.032	.084
5 x 6	.046	.101
4 x 8	.045	.098
5 x 7	.045	.100
4 x 9	.048	.099
6 x 6	.045	.097
4 x 10	.035	.096
5 x 8	.052	.107
6 x 7	.051	.101
5 x 9	.046	.103
6 x 8	.043	.095
7 x 7	.041	.094

REFERENCES

Abrahamse, A.P.J. and Koerts, J. (1971). New estimators of disturbances in regression analysis. J. Amer. Statist. Assoc., 66, 71-74.

Barnett, V. and Lewis, T. (1978). Outliers in Statistical Data. New York, Wiley.

Belsley, D.A., Kuh, E., and Welsch, R.E. (1980). Regression Diagnostics: Identifying Influential Data and Sources of Collinearity. New York: Wiley.

Brown, R.L., Durbin, J., and Evans, J.M. (1975). Techniques for testing the constancy of regression relationships over time. J. Roy. Statist. Soc., B-37, 149-163, discussion 163-192.

D'Agostino, R.B. (1971). An omnibus test of normality for moderate and large size samples. Biometrika, 58, 341-348.

D'Agostino, R.B. and Rosman, B. (1974). The power of Geary's test of normality. Biometrika, 61, 181-184.

Dent, W.T. and Styan, G.P.H. (1978). Uncorrelated residuals from linear models. J. of Econometrics, 7, 211-225

Draper, N.R. and Smith, H.(1981). Applied Regression Analysis, 2nd Ed. New York: Wiley.

Dyer, A.R. (1974). Comparisons of tests for normality with a cautionary note. Biometrika, 61, 185-189.

Farebrother, R.W. (1976). BLUS residuals. Algorithm AS 104, Applied Statistics, 25, 317-322.

Farebrother, R.W. (1978). A historical note on recursive residuals. J. Roy. Statist. Soc., B-40, 373-375.

Gentleman, J.F. and Wilk, M.B. (1975). Detecting outliers. II. Supplementing the direct analysis of residuals. Biometrics, 31, 387-410.

Godolphin, E.J. and de Tullio, M. (1978). Invarience properties of uncorrelated residual transformations. J. Roy. Statist. Soc., B-40, 313-321.

Grossman, S.I. and Styan, G.P.H. (1972). Optimality properties of Theil's BLUS residuals. J. Amer. Statist. Assoc., 67, 672-673.

Hawkins, D.M. (1980). Identification of Outliers, London: Chapman and Hall.

Hedayat, A. and Robson, D.S. (1970). Independent stepwise residuals for testing homoscedasticity. J. Amer. Statist. Assoc., 65, 1573-1581.

Hildreth, C. (1971). A note on approximate regression disturbances. Discussion Paper No. 2, Center for Economic Research, Department of Economics, University of Minnesota, Minneapolis, MN.

Huang, C.J. and Bolch, B.W. (1974). On the testing of regression disturbances for normality. J. Amer. Statist. Assoc., 69, 330-335.

John, J.A. and Draper, N.R. (1978). On testing for two outliers or one outlier in two-way tables. Technometrics, 20 69-78.

Koerts, J. (1967). Some further notes on disturbance estimates in regression analysis. J. Amer. Statist. Assoc., 62 169-183.

Lin, C-C and Mudholkar, G.S. (1980). A simple test for normality against asymmetric alternatives. Biometrika, 67, 455-461.

Martinez, J. and Iglewicz, B. (1981). A test for departure from normality based on a biweight estimator of scale. Biometrika, 68, 331-333.

Nelson, L.S. (1981). A simple test for normality. J. of Quality Technology, 13 , 76-77.

Ramsey, J.B. (1969). Tests for specification errors in classical linear least-squares regression analysis. J. Roy. Statist. Soc., B-31, 350-370.

Ramsey, J.B. and Gilbert, R. (1972). A Monte Carlo study of some small properties of tests for specification error. J. Amer. Statist. Assoc., 67, 180-186,

Seber, G.A.F. (1977). Linear Regression Analysis. New York: Wiley.

Shapiro, S.S. and Francia, R.S. (1972). An approximate analysis of variance test for normality. J. Amer. Statist. Assoc., 67, 215-216.

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

Shapiro, S.S. and Wilk, M.B. (1968). Approximations for the null distribution of the W statistic. Technometrics, 10, 861-866.

Shapiro, S.S., Wilk, M.B., and Chen, H.J. (1968). A comparative study of various tests for normality. J. Amer. Statist. Assoc., 63, 1343 - 1372.

Spiegelhalter, D.J. (1977). A test for normality against symmetric alternatives. Biometrika, 64, 415-418.

Spiegelhalter, D.J. (1980). An omnibus test for normality for small samples. Biometrika, 67, 493-496.

Theil, H. (1965). The analysis of disturbances in regression analysis. J. Amer. Statist. Assoc., 60, 1067-1079.

Tiao, G.C. and Guttman, I. (1967). Analysis of outliers with adjusted residuals. Technometrics, 9, 541-559.

NRD/JAJ/jvs

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER # 2426	2. GOVT ACCESSION NO. AD-A120997	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Testing the Normality of Residuals		5. TYPE OF REPORT & PERIOD COVERED Summary Report - no specific reporting period
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) N. R. Draper and J. A. John		8. CONTRACT OR GRANT NUMBER(s) DAAG29-80-C-0041 and DAAG29-80-C-0113
9. PERFORMING ORGANIZATION NAME AND ADDRESS Mathematics Research Center, University of Wisconsin 610 Walnut Street Madison, Wisconsin 53706		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS Work Unit Number 4 - Statistics and Probability
11. CONTROLLING OFFICE NAME AND ADDRESS U. S. Army Research Office P.O. Box 12211 Research Triangle Park, North Carolina 27709		12. REPORT DATE September 1982
		13. NUMBER OF PAGES 21
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Linear model, Residuals, Shapiro-Wilk test, Tests of normality

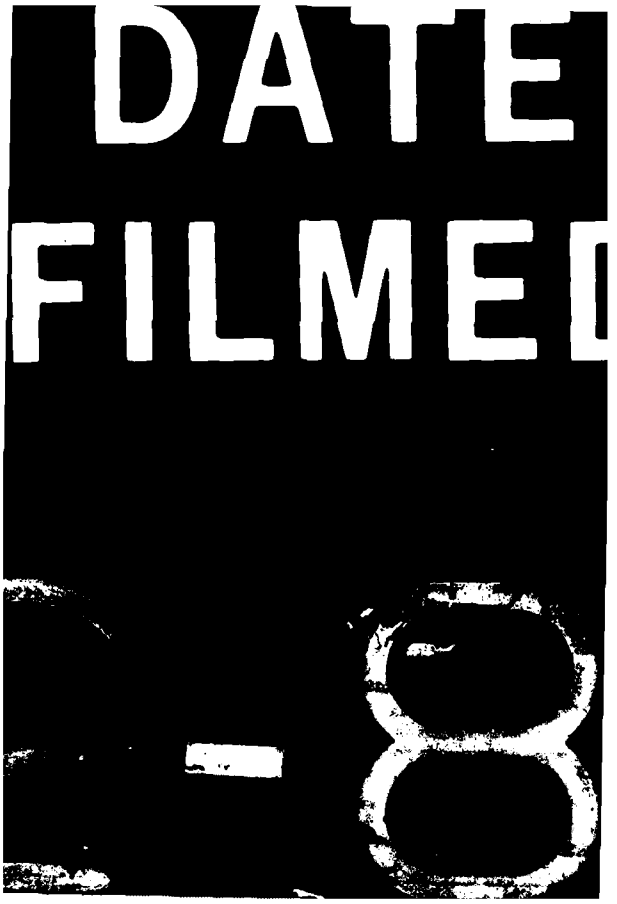
20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The use of residuals to test the assumption of normality of the errors in a linear model is considered. Standard tests for normality typically require an assumption of independence; however the residuals are correlated. An investigation of the Shapiro-Wilk test shows that it is affected by these correlations, but the problem can be overcome by a simple adjustment to the test procedure.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)


DATE
FILME