Testing Transformations to Achieve Approximate Normality

by

Raymond J. Carroll*
University of North Carolina

Abstract

We propose a competitor to likelihood and significance methods for power transformations to achieve approximate normality in a linear model. The new method is shown in theory and a Monte-Carlo experiment to produce more robust inferences than the likelihood method and considerably more powerful (although possibly slightly less robust) inferences than the significance method.

Key Words and Phrases: Transformations, Robustness, Power Family, Monte-Carlo, Likelihood.

*This work was supported by the Air Force Office of Scientific Research under Contract AFOSR-75-2796.

1. Introduction

Box and Cox (1964) considered methods in the linear model for achieving approximate normality, homoscedasticity and additivity by transformations. They hypothesized that for some λ,

$$\underset{\sim}{Y}^{(\lambda)} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon} ,$$

where X is an n×p full rank matrix, $\underset{\sim}{\beta}$ is a ρ×1 vector of constants, the errors $\underset{\sim}{\varepsilon}$ are normal with covariance matrix $\sigma^2 I$, and individual elements $Y_i^{(\lambda)}$ are defined by

$$\begin{aligned} Y_i^{(\lambda)} &= (Y_i^{\lambda} - 1)/\lambda \qquad && \lambda \neq 0 \\ &= \log Y_i && \lambda = 0 \quad . \end{aligned}$$

They proposed maximum likelihood for estimating $\underset{\sim}{\beta}$, σ and λ and likelihood ratio tests for hypohteses of the form $H_0\colon \lambda = \lambda_0$.

Andrews (1971) noted that these methods could be criticized on two counts. First, the likelihood ratio tests are only approximately valid. Second, he showed by example that the maximum likelihood estimate of λ could be heavily influenced by one deviant observation, confirming the well-known fact that maximum likelihood at the normal model tends not be robust. To meet these criticisms, Andrews developed a method (the significance method) based on F-tests with exactly specified Type I error at the normal model and, in his example, relative insensitivity to a deviant observation.

Atkinson (1973) criticized the significance method by showing in a Monte-Carlo experiment that, at the normal distribution, the likelihood ratio test is much more powerful than Andrew's F-test. Atkinson also proposed a method for testing $H_0\colon \lambda = \lambda_0$ which his Monte-Carlo study indicates is essentially equivalent at normality to the likelihood ratio test. Atkinson does not address Andrews' point about robustness.

We thus have a situation that the methods which are powerful for testing in the normal model are not robust, while the method which appears to be robust is not powerful at the normal model. Since power transformations are intended to achieve only approximate normality, the methods mentioned above are deficient in not considering the possibility that the errors $\underline{\varepsilon}$ might be only close to normality, with perhaps heavier tail behavior. The purpose of this paper is to present a computationally feasible method which strikes a middle path between the existing methods, being powerful at models including (but not limited to) the normal model and also being relatively robust for Andrews' example.

In Section 2 we present our method. In Section 3 we discuss a Monte-Carlo experiment in which likelihood based methods do poorly in terms of Type I error, Andrews' significance method does poorly in terms of power, and our methods appear to be dominant.

2. A New Method

To motivate the method, we first consider the problem of estimating location with a symmetric error distribution. Huber (1964) considered the model

$$Y_i = \theta + \varepsilon_i \quad ,$$

where the ε_i are independent and identically distributed with distribution function F belonging to a class $\mathcal{F} = \{G: G = (1-\varepsilon)\Phi + \varepsilon H, \text{ H symmetric}\}$, $\varepsilon > 0$ and Φ the standard normal distribution function. Without considerations of scale, the least favorable member of $\mathcal{F}$ was shown to have density function proportional to $\exp\{-\rho(x)\}$, where

$$\begin{aligned} \rho(x) &= \tfrac{1}{2} x^2 & |x| \le k \\ &= k(|x| - k/2) & |x| > k \ . \end{aligned}$$

The class of Huber M-estimates of θ are maximum likelihood estimates of θ under the model with the least favorable density.

Our method follows this idea and attempts to stay as close to maximum likelihood as is computationally feasible. The corresponding "least favorable" likelihood is proportional to

$$L(\underset{\sim}{\beta},\sigma,\lambda) = \sigma^{-n} \sum_{i=1}^{n} \exp\{-\rho\left(\frac{Y_i^{(\lambda)} - \underset{\sim}{x}_i \underset{\sim}{\beta}}{\sigma}\right) + (\lambda-1)\log Y_i\} \quad . \tag{2.1}$$

For any given λ, we will find $\hat{\underset{\sim}{\beta}}(\lambda)$ and $\hat{\sigma}(\lambda)$ by the robust regression method proposed by Huber (1977, p. 37), so that our method is computationally feasible. Thus, we will maximize (2.1) subject to constraints defining $\hat{\underset{\sim}{\beta}}(\lambda)$ and $\hat{\sigma}(\lambda)$, namely for $\psi(x) = d\rho(x)/dx$,

$$n^{-1} \sum_{i=1}^{n} \psi\left[\frac{Y_i^{(\lambda)} - \underset{\sim}{x}_i \underset{\sim}{\beta}}{\sigma}\right] \underset{\sim}{x}_i = 0 \tag{2.2}$$

$$(n-p)^{-1} \sum_{i=1}^{n} \psi^2\left(\frac{Y_i^{(\lambda)} - \underset{\sim}{x}_i \beta}{\sigma}\right) = E_\Phi \psi^2(Z) \ , \tag{2.3}$$

where Φ is the standard normal distribution and the expectation is taken under Φ.

We denote the estimates of λ and $\underset{\sim}{\beta}$ obtained in this fashion by λ_R and $\hat{\underset{\sim}{\beta}}(\lambda_R)$. In testing a hypothesis of the form H_0: $\lambda = \lambda_0$, the likelihood ratio statistic under the model (2.1) is

$$\Lambda_R = -2 \log \left(\frac{L(\hat{\underset{\sim}{\beta}}(\lambda_0),\ \hat{\sigma}(\lambda_0),\ \lambda_0)}{L(\hat{\underset{\sim}{\beta}}(\lambda_R),\ \hat{\sigma}(\lambda_R),\ \lambda_R)} \right) \ ,$$

under appropriate conditions which is asymptotically chi-square with one degree of freedom.

In terms of inference concerning the $\underset{\sim}{\beta}$ vector, computationally it is most feasible to decide upon a transformation and then do a robust regression analysis as in Huber (1977), perhaps using F-tests as in Shrader (1976) or pseudo-observations as in Bickel (1976).

A theoretical assessment of the influence of outliers on this method and on normal theory maximum likelihood is difficult. A useful technique is the influence curve (Hampel (1974)), which has been used in location problems to describe the effect of a single observation on an estimator. If one can write the estimator as a function $T(F_n)$ of the empirical distribution function, then the influence curve evaluated at a point y_0 and an underlying distribution F is

$$IC(y_0,F) = \lim_{\varepsilon\to 0} (T((1-\varepsilon)F + \varepsilon\, \delta(y_0)) - T(F))/\varepsilon \;,$$

where $\delta(y_0)$ is a distribution with mass one at the point y_0.

To get an idea of the effect of an observation on the new method, we consider the simplified case that $\sigma = 1$ and $\underset{\sim}{\beta} = 0$ and consider the model (2.1). If F_n is the empirical distribution function, differentiation yields that $\lambda_R = \lambda(F_n)$, where in general $\lambda(F)$ can be written as the solution to

$$\int \psi(y^{(\lambda(F))})G(y,\ \lambda(F))d\ F_n(y) = \int \log y\ d\ F_n(y) \;,$$

where

$$G(y,\ \lambda(F)) = dy^{(\lambda)}/d\lambda\Big|_{\lambda=\lambda(F)} = (y^{\lambda} \log y - y^{(\lambda)})/\lambda \;.$$

Applying the definition of the influence curve, one can show that the influence curve is proportional to

$$\psi(y_0{}^{(\lambda(F))})\ G(y_0,\ \lambda(F)) \quad .$$

We thus obtain that if $\lambda(F) = 1$ (no transformation necessary), for large values of y,

(1) The influence curve of normal model maximum likelihood ($\psi(x) = x$) is proportional to $y^2 \log y$.

(2) The influence curve of the new method is proportional to $y \log y$.

Thus, we expect our new method to be somewhat sensitive to outliers but not nearly as much as maximum likelihood at the normal model.

To verify this hypothesis, we followed Andrews' lead and changed the response .23 for poison II, treatment A (in the Box and Cox biological example) to .13. The results are given in Table 1.

Table 1

The effect of one outlier on estimates of λ

	Original Data	Changed Data
Likelihood Method		
75% confidence interval	$-1.01 < \lambda < -0.55$	$-0.30 < \lambda < 0.05$
Maximum likelihood estimate	-0.75	-0.15
Andrews Method		
75% confidence interval	$-0.90 < \lambda < 0.05$	$-1.20 < \lambda < 0.00$
Minimum F estimate	-0.50	-0.50
New Method		
75% confidence interval	$-1.02 < \lambda < -0.57$	$-0.64 < \lambda < -0.20$
Estimate	-0.78	-0.43

Table 1 backs up our influence curve calculations in showing that the new method's estimate λ_R is more robust than the normal theory maximum likelihood estimate. Andrews' minimum F estimate appears slightly more robust than λ_R but the issue is still in doubt because

(1) This may be a reflection (see Section 3) of the lack of power of the F-test used in the significance method.

(2) Whether one uses $\lambda = -1.00$ or $\lambda = -.50$ appears immaterial, as the latter has significance levels of .49 for interaction and less than 0.0001 for main effects.

Certainly λ_R is not completely insensitive to outliers, but the results in Table 1 suggest that it may be of real use when compared to the very conservative minimum F estimate.

3. A Monte-Carlo Study

For purposes of nomenclature, we will call the method proposed in this article the Robust-type method. In order to assess the small sample performance of the four procedures discussed here, we performed a Monte-Carlo experiment, similar to the one described by Atkinson (1973), for computing the power of tests of H_0: $\lambda = -1$ at nominal significance level $\alpha = .05$ and various true values of λ. The specifics were as follows:

(1) The parameter values were the estimates from an additive model on the observations y^{-1}, using the Box and Cox biological data.

(2) In order to assess the methods over a wide class of distributions, four error models were considered.

(a) $N(0,1)$ - standard normal, $\mu = 0$, $\sigma^2 = 1$.
(b) $.90N(0,1) + .10N(0,\sigma^2 = 9)$ - standard normal with probability .90, $N(0,\sigma^2 = 9)$ with probability .10.
(c) $.80N(0,1) + .20N(0,\sigma^2 = 9)$.
(d) The t distribution with 8 degrees of freedom.

(3) For each of the true values λ, the observations for each simulation were $(\text{expected cell mean} + C * \text{error})^{1/\lambda}$. Here the errors are the four distributions mentioned above and $C = (.237)^{1/2}$ was chosen so that most observations were positive. If an observation was less than 0.10, it was set to 0.10.

(4) The sample sizes were N=1000 for Atkinson's and the significance method, N=600 for maximum likelihood, and $150 \leq N \leq 250$ for the robust-type, the latter due to time considerations.

In Table 2 we list the results of this study. The conclusions can be summarized as follows:

(1) The Atkinson and likelihood approaches are quite similar. It is significant to note that they both have Type I errors much higher than the nominal $\alpha = .05$ under the two contaminated normal distributions.

(2) The Robust-type method proposed here appears preferable to the Atkinson and Likelihood methods, because it has comparable power, with the pleasant feature that its Type I error adheres much closer to the nominal $\alpha = .05$.

(3) The Robust-type method also appears preferable to the significance method because of the lack of power of the latter over all the distributions, which can be best seen by looking at the slopes of a plot of λ against power.

Table 2

Power of the four tests for H_0: $\lambda_0 = -1$ under three distributions. The number of iterations is N=1000 for Andrews and Atkinson, N=600 for Maximum Likelihood, and $150 \leq N \leq 250$ for the Robust-Type method. Nominal type I error (when $\lambda = -1.00$) is $\alpha = .05$.

N(0,1)

λ	Andrews	Atkinson	Maximum Likelihood	Robust-Type
-2.0	.539	.762	.779	.772
-1.5	.268	.443	.450	.461
-1.0	.044	.058	.045	.048
- .75	.305	.453	.440	.402
- .60	.807	.931	.952	.942

.90 N(0,1) + .10 N(0,9)

λ	Andrews	Atkinson	Maximum Likelihood	Robust-Type
-2.0	.363	.615	.599	.636
-1.5	.195	.404	.368	.379
-1.0	.048	.178	.167	.100
- .75	.209	.393	.388	.353
- .60	.589	.832	.844	.864

.80 N(0,1) + .20 N(0,9)

λ	Andrews	Atkinson	Maximum Likelihood	Robust-Type
-2.0	.268	.580	.548	.549
-1.5	.152	.406	.375	.348
-1.0	.048	.198	.200	.116
- .75	.163	.371	.338	.279
- .60	.466	.780	.728	.744

t,8 degrees of freedom

λ	Andrews	Atkinson	Maximum Likelihood	Robust Type
-2.00	-	-	.688	.686
-1.5	-	-	.376	.341
-1.0	-	-	.072	.047
- .75	-	-	.388	.410
- .60	-	-	.952	.888

The simulation results show that the Robust-type method proposed here is preferable to the likelihood method for inference about the parameter λ. The same preference held when making inferences about $\tilde{\beta}$ when the null hypothesis H_0: $\lambda = -1$ is true but the errors are slightly heavier-tailed than the normal model. In Table 3 below we list for the intercept parameter the values of $n^{\frac{1}{2}}$ (# of simulations)$^{\frac{1}{2}}$ (standard error for intercept) when the true value of $\lambda = -1.00$, with $n = 48$.

Table 3

Normal standard errors for intercept when $\lambda = -1.00$.

	Maximum Likelihood	Robust-type Method
Standard Normal	3.04	3.12
.90 N(0,1) + .10 N(0,9)	4.48	4.01
.80 N(0,1) + .20 N(0,9)	5.17	4.62

Thus, the Robust-type method leads to more robust estimates of the $\tilde{\beta}$ vector than that obtained by maximum likelihood.

4. Discussion

We have proposed a method which is essentially a constrained likelihood method for a model with heavier tails than the normal model. This Robust-type method appears preferable to normal theory likelihood approaches in terms of inference about λ and $\tilde{\beta}$, at least when one considers the possibility of error models other than the normal. The Robust-type method seems preferable to the significance method because of the lack of power of the latter and the ability of the former to stay near its nominal Type I error (for inferences about λ) over a wide range of distributions. The new method appears slightly less robust against outliers than the significance method, but this defect does not appear to be very serious.

References

Andrews, D.F. (1971). A note on the selection of data transformations. *Biometrika 58*, 249-254.

Atkinson, A.C. (1973). Testing transformations to normality. *J.R. Statist. Soc. B. 35*, 473-479.

Bickel, P.J. (1976). Another look at robustness. *Scand. J. Statist.* 3, 145-168.

Box, G.E.P. and Cox, D.R. (1964). An analysis of transformations. *J.R. Statist. Soc. B 26*, 211-252.

Hampel, F.R. (1974). The influence curve and its role in robust estimation. *J. Am. Statist. Assoc. 69*, 383-393.

Hinkley, D.V. (1975). On power transformations to symmetry. *Biometrika 62*, 101-111.

Huber, P.J. (1964). Robust estimation of a location parameter. *Ann. Math. Statist.* 35, 73-101.

Huber, P.J. (1977). *Robust Statistical Procedures.* SIAM, Philadelphia.

Rao, C.R. (1973). *Linear Statistical Inference and its Applications, Second Edition.* John Wiley and Sons, New York.

Schrader, R.M. (1976). Robust inference based on M-estimates in the linear model. Ph.D. thesis, Pennsylvania State University.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Testing Transformations to Achieve Approximate Normality		5. TYPE OF REPORT & PERIOD COVERED TECHNICAL
		6. PERFORMING ORG. REPORT NUMBER Mimeo Series No. 1190
7. AUTHOR(s) Raymond J. Carroll		8. CONTRACT OR GRANT NUMBER(s) AFOSR-75-2796
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics, University of North Carolina, Chapel Hill, North Carolina 27514		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research, Bolling AFB, Washington, D.C.		12. REPORT DATE August, 1978
		13. NUMBER OF PAGES 11
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)

Transformations, Robustness, Power Family, Monte-Carlo, Likelihood

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

We propose a competitor to likelihood and significance methods for power transformations to achieve approximate normality in a linear model. The new method is shown in theory and a Monte-Carlo experiment to produce more robust inferences than the likelihood method and considerably more powerful (although possibly slightly less robust) inferences than the significance method.