An Analysis of Transformations

G. E. P. Box, D. R. Cox

Journal of the Royal Statistical Society. Series B (Methodological), Volume 26, Issue 2 (1964), 211-252.

Stable URL:
http://links.jstor.org/sici?sici=0035-9246%281964%2926%3A2%3C211%3AAAOT%3E2.0.CO%3B2-6

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at http://www.jstor.org/about/terms.html. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

Journal of the Royal Statistical Society. Series B (Methodological) is published by Royal Statistical Society. Please contact the publisher for further permissions regarding the use of this work. Publisher contact information may be obtained at http://www.jstor.org/journals/rss.html.

Journal of the Royal Statistical Society. Series B (Methodological)
©1964 Royal Statistical Society

JSTOR and the JSTOR logo are trademarks of JSTOR, and are Registered in the U.S. Patent and Trademark Office. For more information on JSTOR contact jstor-info@umich.edu.

©2002 JSTOR

http://www.jstor.org/
Fri Sep 6 10:21:25 2002

An Analysis of Transformations

By G. E. P. Box and D. R. Cox

University of Wisconsin *Birkbeck College, University of London*

[Read at a RESEARCH METHODS MEETING of the SOCIETY, April 8th, 1964, Professor D. V. LINDLEY in the Chair]

SUMMARY

In the analysis of data it is often assumed that observations $y_1, y_2, \ldots, y_n$ are independently normally distributed with constant variance and with expectations specified by a model linear in a set of parameters $\boldsymbol{\theta}$. In this paper we make the less restrictive assumption that such a normal, homoscedastic, linear model is appropriate after some suitable transformation has been applied to the y's. Inferences about the transformation and about the parameters of the linear model are made by computing the likelihood function and the relevant posterior distribution. The contributions of normality, homoscedasticity and additivity to the transformation are separated. The relation of the present methods to earlier procedures for finding transformations is discussed. The methods are illustrated with examples.

1. INTRODUCTION

THE usual techniques for the analysis of linear models as exemplified by the analysis of variance and by multiple regression analysis are usually justified by assuming

(i) simplicity of structure for $E(y)$;
(ii) constancy of error variance;
(iii) normality of distributions;
(iv) independence of observations.

In analysis of variance applications a very important example of (i) is the assumption of additivity, i.e. absence of interaction. For example, in a two-way table it may be possible to represent $E(y)$ by additive constants associated with rows and columns.

If the assumptions (i)–(iii) are not satisfied in terms of the original observations, y, a non-linear transformation of y may improve matters. With this in mind, numerous special transformations for use in the analysis of variance have been examined in the literature; see, in particular, Bartlett (1947). The main emphasis in these studies has tended to be on obtaining a constant error variance, especially when the variance of y is a known function of the mean, as with binomial and Poisson variates.

In multiple regression problems, and in particular in the analysis of response surfaces, assumption (i) might be that $E(y)$ is adequately represented by a rather simple empirical function of the independent variables $x_1, x_2, \ldots, x_l$ and we would want to transform so that this assumption, together with assumptions (ii) and (iii), is approximately satisfied. In some cases transformation of independent as well as of dependent variables might be desirable to produce the simplest possible regression model in the transformed variables. In all cases we are concerned not merely to find a transformation which will justify assumptions but rather to find, where possible, a metric in terms of which the findings may be succinctly expressed.

Each of the considerations (i)–(iii) can, and has been, used separately to select a suitable candidate from a parametric family of transformations. For example, to achieve additivity in the analysis of variance, selection might be based on

(a) minimization of the F value for the degree of freedom for non-additivity (Tukey, 1949); or

(b) minimization of the F ratio for interaction versus error; or

(c) maximization of the F ratio for treatments versus error (Tukey, 1950).

Tukey and Moore (1954) used method (a) in a numerical example, plotting contours of F against (λ_1, λ_2) for transformations in the family $(y+\lambda_2)^{\lambda_1}$. They found that in their particular example the minimizing values were very imprecisely determined.

In both (a) and (b) the general object is to look for a scale on which effects are