To the Graduate Council:

I am submitting herewith a thesis written by Nancy H. Willis entitled "On Transformations of Statistical Data." I recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Science, with a major in Mathematics.

Albert T.-M. Wang

Albert Tso-Ming Wang, Major Professor

We have read this thesis
and recommend its acceptance:

Ray Balsom

Julius Smith

Accepted for the Council:

L Evans Roth

Vice Chancellor
Graduate Studies and Research

ON TRANSFORMATIONS OF STATISTICAL DATA

A Thesis

Presented for the

Master of Science

Degree

The University of Tennessee, Knoxville

Nancy H. Willis

June 1977

1324173

DEDICATION

This thesis is dedicated to my husband, for his encouragement and support.

ACKNOWLEDGEMENT

The author extends her appreciation to Dr. Albert T. M. Wang for his patience and help throughout this study.

She also expresses her gratitude to Dr. Julius Smith and Dr. Raj Pal Soni for their suggestions.

ABSTRACT

The purpose of this study was to investigate transformations of statistical data, namely

1. transformations of data (dependent variables) for analysis of variance;
2. transformations of data (dependent and/or independent variables) for regression analysis.

For the analysis of variance, transformations are devised to provide a change of scale of measurement so that standard analysis techniques can be used. For regression analysis, we seek to transform data so that a simple regression model is appropriate. Methods for determining transformations are discussed.

TABLE OF CONTENTS

CHAPTER		PAGE
I.	INTRODUCTION	1
	Transformation of Data for Analysis of Variance	1
	Transformations of the Independent Variables	3
II.	TRANSFORMATIONS OF THE DEPENDENT VARIABLE TO STABILIZE VARIANCE	6
	Square Root Transformation	7
	Logarithmic Transformation	11
	The Arcsine or Angular Transformation	16
	Summary	17
III.	STRENGTHS OF TRANSFORMATIONS AND THE SIMPLE FAMILY OF TRANSFORMATIONS APPLIED TO DEPENDENT VARIABLES	19
	Strength of a Transformation	20
	The Simple Family and Their Strengths	24
IV.	A QUICK METHOD FOR DETERMINING A TRANSFORMATION OF THE SIMPLE FAMILY FOR THE INDEPENDENT VARIABLES	29
V.	AN ITERATIVE METHOD FOR DETERMINING TRANSFORMATIONS OF THE INDEPENDENT VARIABLES	38
BIBLIOGRAPHY		47
VITA		49

LIST OF TABLES

TABLE		PAGE
II-1.	Number of Poppy Plants in Oats (Plants per $3\frac{3}{4}$ Square Feet)	8
II-2.	Square Roots of the Poppy Numbers in Table II-1	9
II-3.	Analysis of Variance of the Transformed Data in Table II-2	10
II-4.	Macrolepidoptera Catches by Three Traps	13
II-5.	Analysis of Variance and Test for Non-additivity for Data in Table II-4	14
II-6.	Transformed Data of Table II-4 (Using Transformation $\text{Log}_{10}(x)$)	15
II-7.	Analysis of Variance of Table II-6	15
II-8.	Summary of Transformations	18
IV-1.	Amount of Sodium in Biological Fluid	33
IV-2.	First and Second Differences of Data in Table IV-1	33
IV-3.	Track Records	36
IV-4.	Selected Data Points from Table IV-3 and Their Differences	36
V-1.	Generated Data and Application of Iterative Method	43
V-2.	Quantity of Polymer Formed in One Hour Under Steady State Conditions at Equally Spaced Charges of Catalyst	45
V-3.	Analysis of Variance of Original Data in Table V-2	45
V-4.	Analysis of Variance of Transformed Data in Table V-2.	46

CHAPTER I

INTRODUCTION

In statistics, transformations are used to manipulate data so that it can be analyzed by using standard methods of analysis. Transformations of data for two types of analysis will be considered:

(1) transformations of data (dependent variables) for analysis of variance;

(2) transformations of independent and/or dependent variables in regression analysis.

Transformation of Data for Analysis of Variance

Ideally, a set of data which is to be analyzed should have the following characteristics:

(1) a homogeneous variance;

(2) a normal distribution of errors (or one that is approximately normal);

(3) the arithmetic average as an efficient estimator of the true mean;

(4) linear and additive real effects.

Normality is desired since its assumption is the basis for many standard statistical techniques. Homogeneity of variance is needed for a similar reason--heterogeneity may invalidate certain statistical techniques. The arithmetic average is the basis for computations involved in analysis of variance. Finally, additivity is desired since, in certain

designs, non-additivity can inflate the error term and mask treatment effects, or cause other discrepancies.

In practice, departures from these assumptions can be recognized by certain tests. Normality can be tested for by plotting the data, or by using the chi square goodness of fit test, along with tests for skewness and kurtosis. Homogeneity of variance may be detected by such tests as the Hartley or Cochran test. Tukey's test for non-additivity can be used to detect removable non-additivity.

Box (1953) has shown that the distribution of the F ratio in analysis of variance is affected relatively little by moderate departures from normality and by inequalities in variance which have been pooled in the experimental error. However, serious departures from the above assumptions may have to be dealt with either by modifying the method of analysis or by transforming the data to fit the assumptions more closely. Transformations used in this way actually provide a change of scale of measurement for the analysis of variance. It is fortunate that a transformation which bends the data closer to one assumption often improves the degree to which the other assumptions hold, as well; for example, a transformation which stabilizes the variance frequently improves the closeness of the data to normality, and vice versa. Transformations devised by Bartlett (1947) to stabilize variance (but which may also improve the violation of other properties) are presented in Chapter II. Chapter III presents a large and important family of transformations of the dependent variables.

Transformations of the Independent Variables

In regression, transformations are applied to the independent variables so that a regression model of simple form may be used instead of a more difficult model. For example, if a satisfactory linear model (that is, linear in the parameters) can be found, the estimation of the parameters proceeds more easily. In the most general form, a linear model in the independent variables $(X_1, X_2, \ldots, X_k)$ can be written in the form

$$\vec{Y} = Z\vec{\beta} + \vec{\varepsilon}.$$

$\vec{Y} = (Y_1, Y_2, \ldots, Y_n)'$ is the response vector, and $\vec{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)'$ and $\vec{\varepsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)'$ are the parameter and error vectors, respectively.

$$Z = \begin{bmatrix} Z_{10} & Z_{11} & \cdots & Z_{1p} \\ \vdots & \vdots & & \vdots \\ Z_{n0} & Z_{n1} & & Z_{np} \end{bmatrix}$$

is a matrix of functions of the n settings of the independent variables where $Z_{i0} = 1$ for all i; $Z_{ij} = Z_j (X_{i1}, X_{i2}, \ldots, X_{ik})$, is a general function of the k levels of the independent variables corresponding to the i^{th} observation.

If the Z_j are integral powers and cross-products of the levels of the independent variables, the general regression form becomes the

polynomial regression case. If the Z_j's involve other parameters then a nonlinear model is actually being proposed; but by transforming the data through the use of the transformations Z_j, the model reduces to the form of the general linear case. In this paper we have $Z_j = Z_j(X_j)$, $j = 1, 2, \ldots, k$, where each Z_j transforms only one independent variable. Our main concern is the nonlinear case, where Z_j involves other parameters; that is, $Z_j = Z_j(X_i, \vec{\alpha}_i)$, where $\vec{\alpha}_i$ is a vector of parameters. Chapter IV presents a method devised by Dolby (1963) of determining p for the transformation of the form

$$y = a + b(c + x)^p.$$

The model is nonlinear; c and p must be estimated by nonlinear methods. When these two parameters are estimated and the transformation $z = (c + x)^p$ is applied to the data, the model is of the form

$$y = a + bz.$$

Chapter V deals with an iterative method for determining transformations of the form

$$z_i = x_i^{\alpha_i}, \quad i = 1, 2, \ldots k$$

to produce models of the form

$$E(y) = \beta_0 + \sum_{i=1}^{k} \beta_i z_i$$

or

$$E(y) = \beta_0 + \sum_{i=1}^{k} \sum_{\substack{j=1 \\ j \geq i}}^{k} \beta_{ij} z_i z_j .$$

The iterative method produces an estimate a_i of α_i, which is used to transform the data. The motivation behind the derivation of these methods is finding a regression equation of simple form, possibly a polynomial of low order or other elementary function in the transformed variables.

CHAPTER II

TRANSFORMATIONS OF THE DEPENDENT VARIABLE TO STABILIZE VARIANCE

In non-normal distributions, the variance often depends on the mean. If the variance changes with the mean level and is some known function of the mean, say $\sigma_x^2 = f(\mu)$, then the variance may be stabilized by an appropriate change in scale. To determine an appropriate transformation, $g(x)$, we expand $g(x)$ about the mean in a one term Taylor's expansion:

$$g(x) = g(\mu) + g'(\mu)(x - \mu) + \frac{g''(\xi)}{2!}(x - \mu)^2,$$

where ξ is between x and μ. If the error term is relatively small, it can be disregarded, leaving the approximation

$$g(x) \cong g(\mu) + g'(\mu)(x - \mu).$$

Since $E(x - \mu) = 0$, then $E[g(x)] \cong g(\mu)$. Thus,

$$\sigma^2_{g(x)} \cong E[g(x) - g(\mu)]^2$$

$$\cong \{[g'(\mu)]^2(x - \mu)^2$$

$$\cong [g'(\mu)]^2 \sigma_x^2$$

$$\cong [g'(\mu)]^2 f(\mu).$$

Also it can be proven (Rao, 1973, p. 426) that if $\{T_n\}$, $n = 1, 2, \ldots$, is a sequence of statistics and if $\sqrt{n}\,(T_n - \theta) \xrightarrow{L} X \sim N[0,\sigma,\theta)]$, then $\sqrt{n}\,[g(T_n) - g(\theta)] \xrightarrow{L} X \sim N(0,[g'(\theta)\sigma(\theta)]^2)$, assuming g is a function for which $g'(x)$ exists and $g'(\theta) \neq 0$. In order for the variance to be constant and therefore independent of the mean level,

$$g'(\mu) = \frac{c}{\sqrt{f(\mu)}}\,.$$

This leads to the variance stabilizing transformation

$$g(x) = \int \frac{c\,dx}{\sqrt{f(x)}}\,.$$

This result can be applied to various known distributions.

Square Root Transformation

It is known that for the Poisson distribution, the variance equals the mean. Using the fact that $\sigma_x^2 = f(\mu) = \mu$ in the above formula, the transformation

$$g(x) = \int \frac{c}{\sqrt{x}}\,dx = \sqrt{x}$$

for appropriate choice of c, will lead to a stable variance. The transformed data will have mean $\sqrt{\mu}$ and variance $\sigma^2_{g(x)} = \frac{1}{4}$.

Frequently, statistical data consisting of integers or counts of rare events will have a Poisson distribution. For small counts, Bartlett (1947) recommends the use of $\sqrt{x + \frac{1}{2}}$; other recommendations include $\sqrt{x + 1}$ or $\sqrt{x} + \sqrt{x + 1}$ for this case. The square root transformation can also be used for data in which the variance seems to be proportional to the mean, that is, $\sigma_x^2 = k\mu$.

Table II-1 shows a set of data which responds to the square root transformation.

Table II-1. Number of Poppy Plants in Oats (Plants per 3 $\frac{3}{4}$ Square Feet)

	Treatment				
Block	A	B	C	D	E
1	438	538	77	17	18
2	442	422	61	31	26
3	319	377	157	87	77
4	380	315	52	16	20
Mean	394.75	413.00	86.75	37.75	35.25
Range	123	223	105	71	51
Variance	10,059	26,606	6,901	3,375	2,359

Source: George W. Snedecor and William G. Cochran, Statistical Methods, 6th ed. Ames, Iowa: The Iowa State University Press, 1967, p. 326.

The range and variance of the treatments vary widely, suggesting heterogeneity of variance. This would seriously affect the comparison of pairs or subgroups of treatments. After the square root transformation is applied to the data, the ranges and variances of the treatments are more stable (see Table II-2) so that the analysis of variance and tests for differences between means are more valid.

Table II-2. Square Roots of the Poppy Numbers in Table II-1

	Treatment				
Block	A	B	C	D	E
1	20.9	23.2	8.8	4.1	4.2
2	21.0	20.5	7.8	5.6	5.1
3	17.9	19.4	12.5	9.3	8.8
4	19.5	17.7	7.2	4.0	4.5
Mean	19.8	20.2	9.1	5.8	5.6
Range	3.1	5.5	5.3	5.3	4.6
Variance	6.3	16.0	17.0	18.4	13.6

Source: George W. Snedecor and William G. Cochran, Statistical Methods, 6th ed. Ames, Iowa: The Iowa State University Press, 1967, p. 326.

An analysis of variance of the transformed data (see Table II-3) indicates a difference between treatments at the 5% level.

Table II-3. Analysis of Variance of the Transformed Data in Table II-2

Source	D.F.	S.S.	M.S.	F
Treatments	4	865.44	216.36	53.29*
Blocks	3	22.65		
Error	12	48.69	4.06	
Total	19	936.78		

*Significant at .05 level.

Source: George W. Snedecor and William G. Cochran, Statistical Methods, 6th ed. Ames, Iowa: The Iowa State University Press, 1967, p. 326.

Applying the L.S.D. test to determine which means are significantly different, we find that the 5% L.S.D. value is $\left(\sqrt{\frac{2(4.06)}{4}}\right)(2.179) = 3.10$.

Treatments	E	D	C	B	A
(Ascending order)	5.6	5.8	9.1	19.8	20.2

According to this test, treatments D and E are superior to treatment C, which is superior to treatments A and B. If the means are transformed to the original scale, we have:

Treatment	A	B	C	D	E
Mean	392.0	408.0	82.8	33.64	31.36

Logarithmic Transformation

If the standard deviation varies directly as the mean so that $\sigma_x^2 = k^2\mu^2$, or if the variance depends on the mean and the mean level itself fluctuates so that $\sigma_x^2 = \mu + \sigma_\mu^2$, the logarithm transformation may be appropriate. For example, in biological populations, increases in numbers are often proportional to the present size of the population, so that $\sigma_\mu^2 = \lambda^2\mu^2$, giving a variance law

$$\sigma_x^2 = \mu + \lambda^2\mu^2,$$

which often fits a population following the negative binomial distribution. The logarithm transformation can be derived for both $\sigma_x^2 = k^2\mu^2$ and $\sigma_x^2 = \mu + \lambda^2\mu^2$, for λ or μ large. If λ is relatively large compared to μ, $\sigma_x^2 = \mu + \lambda^2\mu^2 \cong \lambda^2\mu^2$. If μ is large, we have

$$\mu + \lambda^2\mu^2 = \mu^2\left(\frac{1}{\mu} + \lambda^2\right).$$

Since μ is large, $\frac{1}{\mu}$ is small, so in this case

$$\sigma_x^2 = \mu + \lambda^2\mu^2 \cong \lambda^2\mu^2.$$

Integrating, we have

$$g(x) = \int \frac{c}{\sqrt{k^2x^2}}\,dx = \log x,$$

for appropriate c. For transformations of integers or of data where zeros are likely to occur, the use of log (x + 1) is suggested to avoid difficulty.

For the variance law $\sigma_x^2 = \mu + \lambda^2\mu^2$, if λ can be estimated well enough, a more exact transformation can be determined:

$$g(x) = \int \frac{c}{\sqrt{x + \lambda^2 x^2}}\,dx = \frac{1}{\lambda} \sinh^{-1} [\lambda\sqrt{x}]$$

or

$$\frac{1}{\lambda} \log [\sqrt{x + \lambda^2 x} + \lambda\sqrt{x}],$$

for appropriate choice of c. This transformation can also be applied to the more general variance law

$$\sigma_x^2 = m^2(\mu + \lambda^2\mu^2).$$

For small numbers, the use of $\frac{1}{\lambda} \sinh^{-1} [\lambda\sqrt{x + \frac{1}{2}}]$ is recommended.

The more exact transformation just discussed has the disadvantage of requiring an approximation of λ. The use of log [1 + x] apparently works sufficiently well, and has an approximate linear relationship with $\sinh^{-1} [\lambda\sqrt{x + \frac{1}{2}}]$ for values of λ which are likely to occur in practice.

In dealing with biological populations, in addition to the fact that change in population size often depends on the mean, it is also true that the measurement we are using may have some limiting factor. For example, if the measurement is the fraction of area covered by some

species of plant, the measurement can never exceed one. In this situation, the transformation $\log \{\frac{x}{1 - x}\}$ can be used to take this limiting factor into account.

Another set of data which may respond to the logarithmic transformation is a set of sample variances or standard deviations. Since the mean of a set of sample variances is proportional to σ^2 and the variance is proportional to σ^4 the variance is a function of the square of the mean, which is the variance law used to derive the logarithmic transformation.

Table II-4 illustrates a set of data for which the logarithmic transformation is useful.

Table II-4. Macrolepidoptera Catches by Three Traps

	Trap		
Period	1	2	3
1	19.1	50.1	123.0
2	23.4	166.1	407.4
3	29.5	223.9	398.1
4	23.4	58.9	229.1
5	16.6	64.6	251.2
Range	12.9	173.8	284.4

Source: George W. Snedecor and William G. Cochran, Statistical Methods, 6th ed. Ames, Iowa: The Iowa State University Press, 1967, p. 333.

The range of the treatments (traps) varies widely. Also, if Tukey's test for non-additivity is applied, the sum of squares for non-addivity is found to be 24,327, which is significant at the 5% level (see Table II-5). Thus, we reject the hypothesis for additivity.

Table II-5. Analysis of Variance and Test for Non-additivity for Data in Table II-4

Source	D.F.	S.S.	M.S.	F
Traps	2	173,333		
Periods	4	52,066		
Error	8	30,607		
Non-additivity	1	24,327	24,327	27.1*
Remainder	7	6,280	897	

*Significant at .05 level.

Source: George W. Snedecor and William G. Cochran, Statistical Methods, 6th ed. Ames, Iowa: The Iowa State University Press, 1967, p. 334.

After the data is transformed using the base 10 logarithm, the ranges of the traps are more stable (Table II-6). In addition, Tukey's test for non-addivity applied to the transformed data indicates that we should not reject the hypothesis of addivity at the 5% level (see Table II-7). Also, we should reject the hypothesis that there is no difference in treatment means. Applying the Q test for difference in

Table II-6. Transformed Data of Table II-4 (Using Transformation $Log_{10}(x)$)

	Trap		
Period	1	2	3
1	1.2810	1.6998	2.0899
2	1.3692	2.2204	2.6100
3	1.4698	2.3501	2.6000
4	1.3692	1.7701	2.3600
5	1.2201	1.8102	2.4000
Range	0.2497	0.6503	0.5201

Table II-7. Analysis of Variance of Table II-6.

Source	D.F.	S.S.	M.S.	F
Traps	2	2.8918	1.4459	83.378
Periods	4	.4247	.1062	
Error	8	.1386	.0173	
Non-additivity	1	1.0395	.0395	2.801
Remainder	7	.0991	.0141	

means, we find D = $(\sqrt{.01735})$ (3.77) = 0.222. Arranging the means in ascending order,

Treatment	1	2	3
Mean	1.3419	1.9701	2.4120

it is obvious that treatment 3 is superior to treatments 1 and 2 and treatment 2 is superior to treatment 1.

The Arcsine or Angular Transformation

The binomial proportion x has mean p and variance p(1-p)/n, where n is the number of individuals in the sample. Using the integration technique, the following transformation is obtained:

$$g(x) = \int \frac{c}{\sqrt{x(1-x)}}\,dx$$
$$= \int \frac{c}{\sqrt{x}} \cdot \frac{1}{\sqrt{1-x}}\,dx$$
$$= \sin^{-1}\sqrt{x},$$

for appropriate c. This transformation is equivalent to replacing a proportion x by the angle whose sine is $\sqrt{x}$. It has the effect of spreading out proportions near 0 or 1, thus increasing their variance. However, it has little effect on the conclusions if most of the proportions lie between 30% and 70%, since p(1-p) changes little in this range. The transformation will produce a theoretical mean and variance of $\sin^{-1}\sqrt{p}$, and $\frac{1}{4n}$, respectively.

Anscombe (1948) proposes the transformation replacing the proportion $x = \frac{a}{n}$ by $\sin^{-1}\sqrt{(a + \frac{3}{8}) \div (n + \frac{3}{4})}$ to achieve, theoretically at least, more stability in variance. This more difficult transformation could be replaced by the simpler transformation $\sin^{-1}\sqrt{\frac{a}{n}}$ for large n.

If n is relatively small (less than 50), an empirical correction

is replacing 0 by $\frac{1}{4}$ and n by n - $\frac{1}{4}$ (Bartlett, 1947). This is equivalent to replacing a zero proportion by $\frac{1}{4}$ n and a 100% proportion by $\frac{n - \frac{1}{4}}{n}$. This correction tends to improve the equality of variance.

The arcsine transformation can also be tried on proportions subject to other sources of variation besides the binomial, if the variance is thought to be proportional to p(1-p).

Summary

The previously discussed transformations are summarized in Table II-8.

Table II-8. Summary of Transformations

Variance as Function of Mean m	Transformation	Approximate Variance on New Scale	Relevant Distribution
μ	$\sqrt{x}$ or $\sqrt{x + \frac{1}{2}}$	$\frac{1}{4}$	Poisson
$\lambda^2\mu$	for small integers	$\frac{1}{4\lambda^2}$	Empirical
$\lambda^2\mu^2$	$\log_e x$, $\log_e (x + 1)$	λ^2	Empirical
	$\log_{10} x$, $\log_{10}(x + 1)$	$.189\lambda^2$	
$\mu + \lambda^2\mu^2$ μ or λ large	$\log_e x$, $\log_{10} x$	$\simeq \lambda^2$	
$\mu + \lambda^2\mu^2$	$\sinh^{-1} \lambda\sqrt{x}$ or	$\frac{1}{4}$	Negative binomial
λ known	$\sinh^{-1}\lambda\sqrt{x + \frac{1}{2}}$		
$m^2(\mu + \lambda^2\mu^2)$	for small integers	$\mu^2/4$	Empirical
$\mu(1-\mu)/n$	$\sin^{-1} \sqrt{x}$	$\frac{1}{4n}$(radians)	Binomial
		$\frac{821}{n}$(degrees)	

Source: M. S. Bartlett, "The Use of Transformations," Biometrics 3 (1947), 52.

CHAPTER III

STRENGTHS OF TRANSFORMATIONS AND THE SIMPLE FAMILY OF TRANSFORMATIONS APPLIED TO DEPENDENT VARIABLES

As discussed in Chapter I, it is desirable, for analysis of data, to have

(1) additive effects;

(2) a constant error variance; and

(3) a symmetric and approximately normal distribution of errors.

The failure of any of these properties to hold can be recognized by the inequality of two differences, $y_B - y_A$ and $y_D - y_C$, where it is assumed that $y_A < y_B \leq y_C < y_D$. For example, the failure of the above three properties can be recognized by the following differences:

<u>non-additivity</u> - $y(u_1,v_2) - y(u_1,v_1) \neq y(u_2,v_2) - y(u_2,v_1)$, where $y(u,v)$ usually represents a typical value of the distribution of values obtained for fixed (u,v), such as the population mean or median;

<u>nonconstant variance</u> - $p(u_1) - q(u_1) \neq p(u_2) - q(u_2)$, where $p(u_1)$ is the p% point of the distribution of values when $u = u_1$ and $q(u_1)$ is the q% point.

<u>assymmetry</u> - $q(u) - y(u) \neq y(u) - p(u)$, where $p(u)$ and $q(u)$ are the upper and lower percentage points cutting off equal fail areas, and $y(u)$ is the median.

In practice, the required quantities y_A, y_B, y_C, and y_D can be estimated using observed order statistics.

Tukey (1957) measures the degree of failure of one or more properties to hold for a set of data by use of the quantity $\log \frac{y_B - y_A}{y_D - y_C}$. The ratio $\frac{y_B - y_A}{y_D - y_C}$ is invariant under a change of scale from y to ay, as it should be, since multiplying a set of data by some constant is not expected to change the degree to which the above properties hold. The use of this quantity is useful in determining the closeness of a transformed set of data to the possession of the desired properties.

Strength of a Transformation

The strength of a transformation is assessed by the extent to which it changes the violation of some property. If y is transformed to z, then for a given quartet, we have altered $\log \frac{y_B - y_A}{y_D - y_C}$ into $\log \frac{z_B - z_A}{z_D - z_C}$, assuming that $y_A < y_B \leq y_C < y_D$ and $z_A < z_B \leq z_C < z_D$. The strength of the transformation can be measured by:

$$k(y_A, y_B; y_C, y_D) = \log \frac{z_B - z_A}{z_D - z_C} - \log \frac{y_B - y_A}{y_D - y_C}$$

$$= \log \frac{z_B - z_A}{y_B - y_A} - \log \frac{z_D - z_C}{y_D - y_C} .$$

If the difference quotients are replaced by their limits, one pair of arguments can coalesce to produce the forms

$$k(y_1; y_C, y_D) = k(y_1, y_1; y_C, y_D) = [\log \frac{dz}{dy}]_{y_1} - \log \frac{z_D - z_C}{y_D - y_C},$$

$$k(y_A, y_B; y_2) = k(y_A, y_B; y_2, y_2) = \log \frac{z_B - z_A}{y_B - y_A} - \log [\frac{dz}{dy}]_{y_2}.$$

Both pairs of arguments can coalesce to produce the confluent strength

$$k(y_1, y_2) = k(y_1, y_1; y_2, y_2) = [\log \frac{dz}{dy}]_{y_1} - \log [\frac{dz}{dy}]_{y_2}.$$

If successive transformations from y to z and from z to w, and the composition of these transformations from y to w are considered, the strengths are

$$k_{yz}(y_A, y_B; y_C, y_D) = \log \frac{z_B - z_A}{y_B - y_A} - \log \frac{z_D - z_C}{y_D - y_C};$$

$$k_{zw}(z_A, z_B; z_C, z_D) = \log \frac{w_B - w_A}{z_B - z_A} - \log \frac{w_D - w_A}{z_B - z_A};$$

$$k_{yw}(y_A, y_B; y_C, y_D) = \log \frac{w_B - w_A}{y_A - y_B} - \log \frac{w_D - w_C}{y_D - y_C}.$$

It is obvious that

$$k_{yz}(y_A, y_B; y_C, y_D) + k_{zw}(z_A, z_B; z_C, z_D) = k_{yw}(y_A, y_B; y_C, y_D),$$

and that strengths of composed transformations are additive. However, if the same transformation is applied twice in succession to a certain quartet, the strengths of the two applications will not be the same,

and the strength of the composition will not be twice either strength.

If the derivative of the confluent strength is taken, we have

$$k_{yz}(y) = \lim_{h\to 0} \frac{k(y;\, y+h) - k(y,y)}{h}$$

$$= \lim_{h\to 0} \frac{1}{h} \left[\log \left(\frac{dz}{dy}\right) - \log \left(\frac{dz}{dy}\right)_{y+h} + \log \left(\frac{dz}{dy}\right) - \log \left(\frac{dz}{dy}\right)\right]$$

$$= - \left(\frac{d^2z}{dy^2}\right) / \left(\frac{dz}{dy}\right).$$

We cannot use this directly as a local measure of the strength since the distinction between the underlying points has been lost. However, applying this result to the composition of strengths, if $z = f(y)$ and $z + \delta = f(y + h)$, then

$$k_{yw}(y;\, y+h) = k_{yz}(y;\, y+h) + k_{zw}(z;\, z+\delta)$$

and

$$\lim_{h\to 0} \frac{1}{h}[k_{yw}(y;\, y+h] = \lim_{h\to 0} \frac{1}{h}\left[\log \left(\frac{dz}{dy}\right) - \log \left(\frac{dz}{dy}\right)_{y+h} + \log \left(\frac{dw}{dz}\right)_{f(y)}\right.$$

$$\left. - \log \left(\frac{dw}{dz}\right)_{f(y+\delta)}\right]$$

$$= k_{yz}(y) + \frac{dz}{dy}\, k_{zw}(z).$$

Consider the transformations $z = \sqrt{y}$ and $w = \sqrt{z}$. Applying the derivative of the confluent strength we have

$$k_{yz}(y) = \frac{1/(4y\sqrt{y})}{1/2\sqrt{y}} = \frac{1}{2y},$$

$$k_{zw}(z) = \frac{1}{2z},$$

$$\frac{dz}{dy} = \frac{1}{2\sqrt{y}}.$$

From the formula,

$$k_{yw} = \frac{1}{2y} + (\frac{1}{2\sqrt{y}})(\frac{1}{2z})$$

$$= \frac{1}{2y} + \frac{1}{4y} \quad (\text{since } z = \sqrt{y})$$

$$= \frac{3}{4y}.$$

This is exactly the answer which would have been obtained if k_{yw} had been calculated directly.

Having developed the idea of the strength of the general transformation to measure the extent it "bends" some set of data to fit a certain assumption, the concept of strength is now applied to an important family of transformations.

The Simple Family and Their Strengths

The simple family of transformations studied by Tukey are defined to include the transformations of the form where y is replaced by

$$z = (y + c)^p.$$

This family includes the limits of the above transformations, thereby including the transformations $z = e^{my}$ and $z = \log (y + c)$, and also contains all linear transformations of the above transformations. In practice, members of the simple family with $p \leq 1$ are used more frequently.

Consider the pure power transformations of the simple family

where $z = \begin{cases} y^p, & p \neq 0 \\ \log y, & p = 0 \end{cases}$. In this case,

$$k(y_i; y_2) = [\log \frac{dz}{dy}]_{y_1} - [\log \frac{dz}{dy}]_{y_2}$$

$$= \begin{cases} \log p + (p-1) \log y_1 - \log p - (p-1) \log y_2, & p \neq 0, \\ -\log y_1 + \log y_2, & p = 0 \end{cases}$$

$$= (p-1) \log (\frac{y_1}{y_2}), \text{ for all } p.$$

The confluent strength is proportional to (p-1) for any pair of confluent arguments.

Considering the composition of pure power transformations, if $w = z^q$ and $z = y^p$,

$$k_{yz}(y_1; y_2) = (p-1) \log \left(\frac{y_1}{y_2}\right)$$

$$k_{zw}(z_1; z_2) = (q-1) \log \left(\frac{z_1}{z_2}\right) = (q-1)p \log \left(\frac{y_1}{y_2}\right)$$

$$k_{yw}(y_1; y_2) = (pq-1) \log \left(\frac{y_1}{y_2}\right).$$

The additive decomposition of the numerical coefficients of the strengths is

$$(p-1) + (q-1)p = pq - 1.$$

Thus, if the square root transformation is applied twice in succession so that $p = q = \frac{1}{2}$, the numerical coefficient of the strength of the composition is

$$pq - 1 = -\frac{3}{4} = \left(-\frac{1}{2}\right) + \left(-\frac{1}{4}\right).$$

That is, the second square root transformation has only half the strength of the first.

The previously stated decomposition of numerical coefficients still holds when the second transformation is the logarithm transformation ($q = 0$), since

$$pq - 1 = -1 = (p-1) + (0-1)p.$$

However, if the logarithm is chosen as the first transformation (p = 0) and is composed with a power transformation, the decomposition fails to hold since $\log \frac{z_1}{z_2} \neq p \log \frac{y_1}{y_2}$. In this case, the strength depends strongly on the particular values y_1 and y_2. However, the transformations resulting from these compositions, such as $\sqrt{\log y}$, are not power transformations and are rarely used.

The power strength is defined as the ratio

$$\frac{k_{yz}(y_1,y_2)}{-\log (y_1/y_2)} .$$

The power strength of the pure power transformation is 1-p and the sequence of transformations

$$\ldots, y, \sqrt{y}, \log y, \frac{1}{\sqrt{y}}, \frac{1}{y}, \ldots$$

have equally spaced power strengths

$$\ldots, 0, \frac{1}{2}, 1, \frac{3}{2}, 2, \ldots .$$

Now consider the modified power transformations of the simple family,

$$z = a_p(y + c)^p, \quad p \neq 0,$$

$$\log (y + c), \quad p = 0,$$

where a_p may be chosen so that sgn a_p = sgn p. We have

$$\log \frac{dz}{dy} = (p-1) \log (y + c) + b$$

where

$$b = 0, \qquad p = 0,$$

$$b = \log [p(a_p)], \quad p \neq 0.$$

Then $k(y_1; y_2) = (p-1) \log \frac{(y_1 + c)}{(y_2 + c)}$ and the power strength will be

$$\frac{(p-1) \log ((y_1 + c)/(y_2 + c))}{- \log (y_1/y_2)} = \frac{(1-p) \log ((y_1 + c)/(y_2 + c))}{\log (y_1/y_2)} .$$

Thus, for a transformation which is not a pure power, the power strength depends on the particular values of y_1 and y_2.

The concept of power strengths can be used to express a relationship between two otherwise unrelated transformations. For example, it is known that a Poisson distribution with mean λ can be approximately normalized by either the transformation $z = \sqrt{y}$ or $z = \log (y + \lambda)$. These transformations can be shown to have approximately the same power strengths at $y = \lambda$. The power strength of $z = \sqrt{y}$ is $(1-p) = 1 - \frac{1}{2} = \frac{1}{2}$. To approximate the power strength of $z = \log (y + \lambda)$, we utilize the Taylor's expansion of $\log \frac{(y_1 + c)/(y_2 + c)}{(y_1/y_2)}$ to obtain

$$\frac{k(y_1,y_2)}{- \log (y_1,y_2)} = (1-p)[\frac{y}{y + c} - \frac{\delta^2}{3} \frac{c(2y + c)}{y(y + c)^3} + \ldots],$$

where $y = y_1 + k\delta$ and $y = y_2 - \ell\delta$. Using this expansion, the power strength of $z = \log (y + \lambda)$ at $y = \lambda$ is approximately

$$\frac{k(y_1,y_2)}{-\log (y_1,y_2)} \cong (1-0)\left[\frac{\lambda}{x + \lambda}\right] = \frac{1}{2}.$$

This approximate equality of power strength indicates the rough equivalence of the transformations.

CHAPTER IV

A QUICK METHOD FOR DETERMINING A TRANSFORMATION OF THE SIMPLE FAMILY FOR THE INDEPENDENT VARIABLES

Dolby (1963) applies the simple family of transformations to the independent variables and makes use of a differential equation (and corresponding difference equation) characteristic of this family to obtain a quick rough estimate of p to be used directly or as a starting point for obtaining a more exact solution. Written in the most general form, if

$$y = a + b(c + x)^p,$$

it is obvious that $\frac{d}{dx}\left(\frac{y'}{y''}\right) = \frac{1}{p-1}$. Let f be a solution of this differential equation, and let

$$A_f = .5 + \frac{d}{dx}\frac{f'}{f''}.$$

If g is the inverse function of f, the identity $f[g(x)] = x$ can be differentiated successively as follows:

4.1 $$f'(g)g' = 1$$

4.2 $$f'(g)g'' + f''(g)(g')^2 = 0$$

4.3 $$f''(g)g'g'' + f'(g)g''' + 2f''(g)g'g'' + f'''(g)(g')^3 = 0.$$

If we divide 4.3 by $f''(g)g''$, we obtain the statement

$$4.4 \qquad 3g' + \frac{f'(g)g'''}{f''(g)g''} + \frac{f'''(g)(g')^3}{f''(g)g''} = 0.$$

From 4.2, the identities

$$f''(g) = -\frac{f'(g)g''}{(g')^2}$$

$$(g')^2 = -\frac{f'(g)g''}{f''(g)}$$

are obtained, which on substitution into 4.4 yields

$$\frac{f'f'''}{(f'')^2} + \frac{g'g'''}{(g'')^2} = 3.$$

Since $A_f = .5 + \frac{d}{dx}\frac{f'}{f''} = .5 + 1 - \frac{f'f'''}{(f'')^2}$

$$= 1.5 + (\frac{g'g'''}{(g'')^2} - 3)$$

$$= -(.5 + \frac{(g'')^2 - g'g'''}{(g'')})$$

$$= -(.5 + \frac{d}{dx}\frac{g'}{g''}),$$

it is obvious that $A_f = -A_g$.

For $A_f = A_g = 0$, we find $p = -1$, which indicates f is of the form $a + \frac{b}{c+x}$, that is, f is the reciprocal transformation. The inverse

of the reciprocal transformation has the same form. This is also true for the linear transformation where A_f and A_g are infinite. Because of these properties and the relation $A_f = -A_g$, Dolby uses A_f as an index of the simple family. He maps the real line onto the circle by the stereographic map with $A_f = 0$ at the south pole and A_f infinite at the north pole (see Figure IV-1). The transformation f and its inverse g are symmetrically placed around the circle. Choosing a diameter of $\sqrt{3}/2$ units insures that the transformations of greatest interest are equally spaced around the circle.

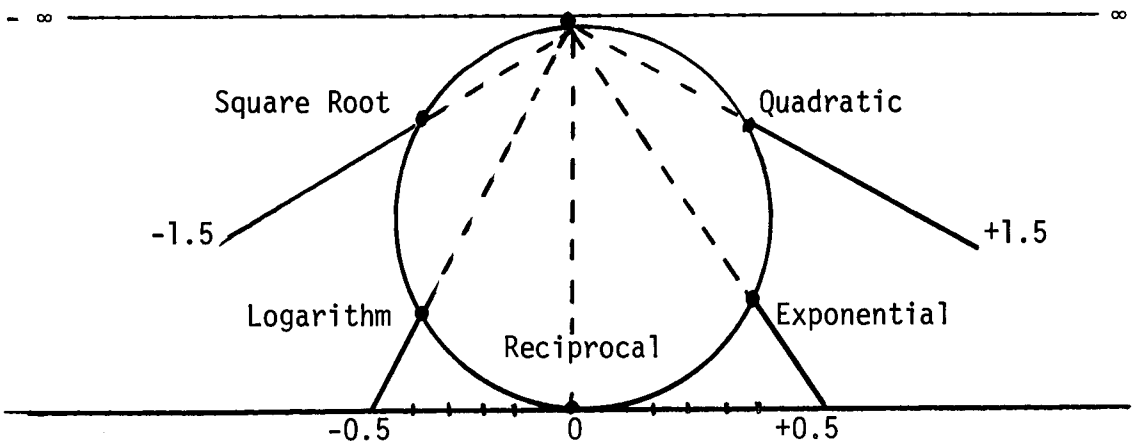


Figure IV-1. Stereographic Mapping of A

Source: J. L. Dolby, "A Quick Method for Choosing a Transformation," Technometrics 5 (August, 1963), 320.

The difference approximation of the differential equation

$$A_f = .5 + \frac{d}{dx}\frac{f'}{f''}$$

is

$$a_f = .5 + \left[\frac{.5(\Delta_3 + \Delta_2)}{\Delta_3 - \Delta_2} - \frac{.5(\Delta_2 + \Delta_1)}{\Delta_2 - \Delta_1} \right]$$

where $\Delta_i = f(x_{i+1}) - f(x_i)$. It can be shown that

$$a_{\text{Quadratic}} = 1.5 = A_{\text{Quadratic}}$$

$$a_{\text{Exponential}} = .5 = A_{\text{Exponential}}$$

$$a_{\text{Reciprocal}} = 0 = A_{\text{Reciprocal}}$$

$$a_{\text{Log}} \tilde{=} .5 = A_{\text{Log}}$$

$$a_{\text{Square Root}} \tilde{=} -1.5 = A_{\text{Square Root}}$$

} provided the smallest value of c + x is safely away from 0.

Also, $a_{\text{linear}} \to \infty$ if approached from the quadratic side and $a_{\text{linear}} \to -\infty$ if approached from the square root side.

The difference equation can be applied to the data in Table IV-1 to determine an approximation of p.

The first and second differences are calculated in Table IV-2. If the difference equation is applied to the first four points and their differences, and to the last four points and their differences, the estimates are

Table IV-1. Amount of Sodium in Biological Fluid

Sodium Present (in mg/100ml)	Total Sodium Found in Six Analyses
100	607
200	1208
300	1806
400	2394
500	2943

Source: J. L. Dolby, "A Quick Method for Choosing a Transformation," Technometrics 5 (August, 1963), 321.

Table IV-2. First and Second Differences of Data in Table IV-1

Data	First Difference	Second Difference
607		
	601	
1208		-3
	598	
1806		-10
	588	
2394		-39
	549	
2943		

Source: J. L. Dolby, "A Quick Method for Choosing a Transformation," Technometrics 5 (August, 1963), 321.

$$A_1 = .5(1 + \frac{598 + 588}{-10} - \frac{601 + 598}{-3}) = 141.0$$

$$A_2 = .5(1 + \frac{588 + 549}{-39} - \frac{598 + 588}{-10}) = 54.2.$$

Neither of these values, projected onto the circle in Figure IV-2, indicate nonlinearity.

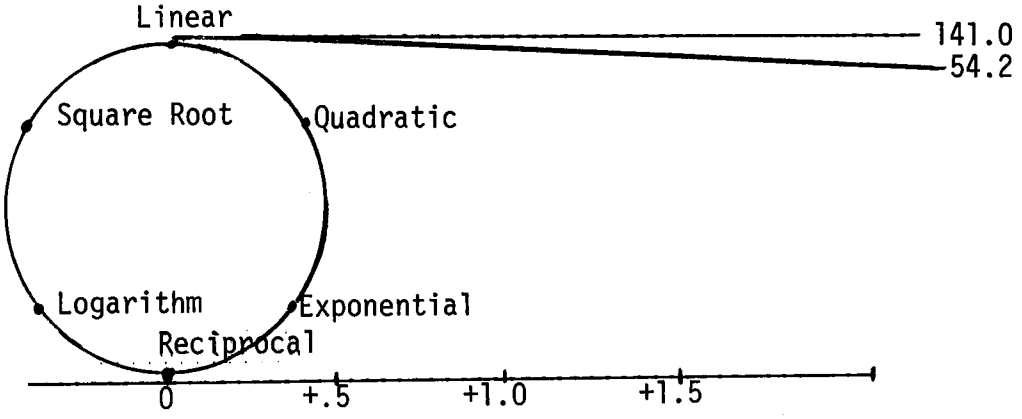


Figure IV-2. Two Values of A for the Data in Table IV-1.

Source: J. L. Dolby, "A Quick Method for Choosing a Transformation," Technometrics 5 (August 1963), 320.

Use of the relation $p = \frac{2A + 1}{2A - 1}$ yields estimates $p_1 = 1.007$ and $p_2 = 1.022$, both of which are very close to $p = 1$.

To apply a similar procedure to data which is not equally spaced, the difference formula must be redefined. Let

$[2,1] = \frac{f_2 - f_1}{x_2 - x_1}$ and $[3,2,1] = \frac{[3,2] - [2,1]}{x_3 - x_1}$. Define

$$a_f = \frac{1}{2}\left\{1 + \frac{1}{x_4 + x_3 - x_1}\left(\frac{[4,3] + [3,2]}{[4,3,2]} - \frac{[3,2] + [2,1]}{[3,2,1]}\right)\right\}.$$

If this procedure is applied to the data in Table IV-3, it must first be decided which data points are to be used in the calculation of a. Since three slopes are used in the definition, it is suggested that the two endpoints and the two interior points which most nearly divide the data into three equal intervals be chosen. From this set of points and their differences (see Table IV-4), we get

$$A = .5[1 + \frac{1}{4473.6}(\frac{-.00045031}{.00000005532} - \frac{-.00524691}{.00000274647})$$

$$= -.1963.$$

The value of A is projected onto the circle (see Figure IV-3) indicating that the appropriate transformation lies between the reciprocal transformation and the logarithmic transformation. The corresponding estimate of p is $p = \frac{2A + 1}{2A - 1} \cong -.436$. This compares favorably with the estimate $p = -.426$ obtained by another method (Dolby, 1963, p. 322).

Table IV-3. Track Records

Time (x)	Speed (y)
20.0	10.06
45.2	8.85
105.7	7.57
139.0	7.19
237.8	6.77
302.2	6.62
472.8	6.35
816.8	6.12
1710.4	5.85
3600.0	5.57

Source: J. L. Dolby, "A Quick Method for Choosing a Transformation," Technometrics 5 (August, 1963), 322.

Table IV-4. Selected Data Points from Table IV-3 and Their Differences

x	y	[·, ·]	[·, ·, ·]
20.0	10.06		
		-0.00494477	
816.8	6.12		0.00000274647
		-0.00030214	
1710.4	5.85		0.00000005532
		-0.00014817	
3600.0	5.57		

Source: J. L. Dolby, "A Quick Method for Choosing a Transformation," Technometrics 5 (August, 1963), 323.

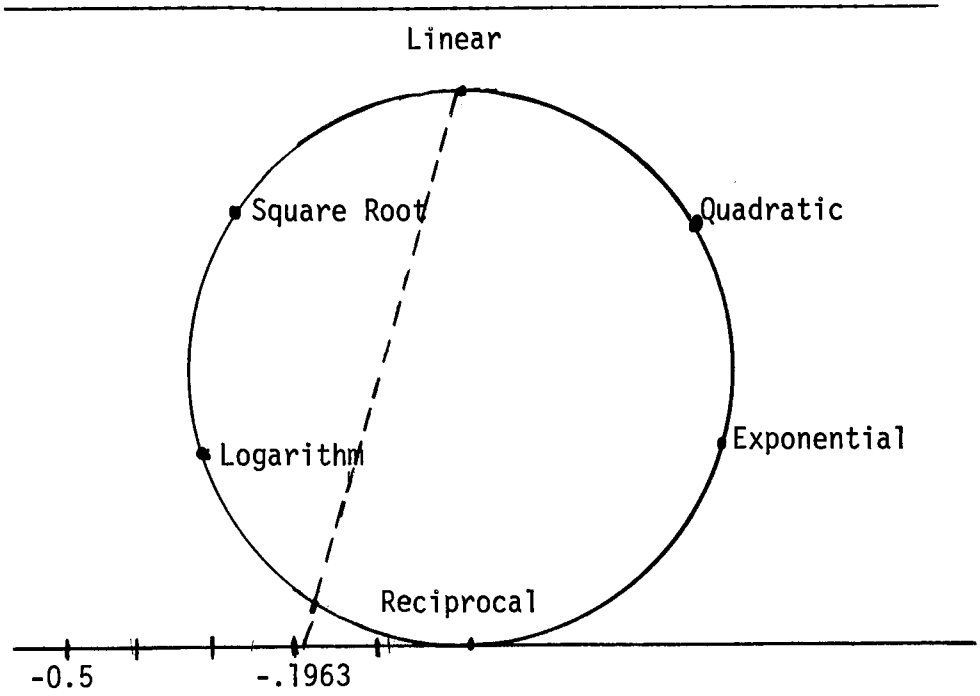


Figure IV-3. Value of A for the Data in Table IV-3.

Source: J. L. Dolby, "A Quick Method for Choosing a Transformation," Technometrics 5 (August, 1963), 324.

CHAPTER V

AN ITERATIVE METHOD FOR DETERMINING TRANSFORMATIONS OF THE INDEPENDENT VARIABLES

Box and Tidwell (1962) have presented an iterative method for determining suitable transformations of data, so that a first or second degree polynomial in the transformed data can be used. Suppose for a particular set of data, the observations $y_1, y_2, \ldots, y_n$ are available at n sets of conditions $\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_n$ where $\vec{x}_u$ is the $1 \times k$ vector containing the k levels of the independent variables for the u^{th} observation, and assume that

$$E(y_i) = \eta_i$$

and

$$E(y_i - \eta_i)(y_j - \eta_j) = \begin{cases} \sigma^2, & i = j, \\ 0 & i \neq j, \end{cases}$$

for all i and j, where σ^2 is unknown. Also suppose that the response η can be closely represented over a region of interest by some simple function

$$\eta = f(\vec{\xi}, \vec{\beta}),$$

where $\vec{\xi} = [\xi_1, \xi_2, \ldots, \xi_k]$ and the ξ_i are some transformations of the x's such that

$$\xi_i = \xi_i\ (x_i, \vec{\alpha}_i),\ i = 1, 2, \ldots, k.$$

$\vec{\alpha}_i = [\alpha_{i1}, \alpha_{i2}, \ldots, \alpha_{ip_i}]$ is a vector of unknown constants of the transformation and $\vec{\beta} = [\beta_0, \beta_1, \beta_2, \ldots, \beta_\ell]$ is a vector of unknown constants depending on the choice of the transformations ξ_i.

If $\vec{\alpha}_i^{(0)} = [\alpha_{i1}^{(0)}, \alpha_{i2}^{(0)}, \ldots, \alpha_{ip_i}^{(0)}]$ are the first guesses for the constants of the i^{th} transformation, we can use the Taylor's expansion to expand about these guessed values to get

$$5.1 \qquad \eta_u = f(\vec{\xi}_u^{(0)}, \vec{\beta}) + \sum_{i=1}^{k} \sum_{j=1}^{p_i} (\alpha_{ij} - \alpha_{ij}^{(0)}) \left\{ \frac{\partial f(\vec{\xi}, \vec{\beta})}{\partial \alpha_{ij}} \right\}_{\substack{\vec{\xi} = \vec{\xi}_u^{(0)} \\ \vec{\alpha}_i = \vec{\alpha}_i^{(0)}}}$$

where $\vec{\xi}_u^{(0)}$ is the vector with i^{th} element $\xi_i(x_{ui}\vec{\alpha}_i^{(0)})$. Now

$$\left\{ \frac{\partial f(\vec{\xi}, \vec{\beta})}{\partial \alpha_{ij}} \right\}_{\substack{\vec{\xi} = \vec{\xi}_u^{(0)} \\ \vec{\alpha}_i = \vec{\alpha}_i^{(0)}}} = \left\{ \frac{\partial f(\vec{\xi}, \vec{\beta})}{\partial \xi_i} \right\}_{\vec{\xi} = \vec{\xi}_u^{(0)}} \left\{ \frac{\partial \xi_{ui}}{\partial \alpha_{ij}} \right\}_{\vec{\alpha}_i = \vec{\alpha}_i^{(0)}}.$$

The quantities $\left\{ \frac{\partial \xi_{ui}}{\partial \alpha_{ij}} \right\}$ can be calculated from knowledge of the transformations $\xi_i(x_i, \vec{\alpha}_i)$. Since the β_i are unknown parameters, the quantities $\left\{ \frac{\partial f(\vec{\xi}, \vec{\beta})}{\partial \xi_i} \right\}_{\vec{\xi} = \vec{\xi}_u^{(0)}}$ must be approximated. This approximation can be accomplished by fitting the observations to $\eta_u = f(\vec{\xi}_u^{(0)}, \vec{\beta})$, $u = 1, \ldots, n$, using some method such as least squares. The fitted expressions

$$y_u = f(\vec{\xi}_u^{(0)}, \vec{b})$$

where $\vec{b}$ is the estimate of $\vec{\beta}$, may then be differentiated to obtain the desired approximated values. Using the approximations we have obtained, we can refit the observations to equation 5.1, letting the adjusted constants take the place of the initial guesses.

This method can be applied to the power transformation where $\xi_i = \begin{cases} x_i^{\alpha_i}, & \alpha_i \neq 0 \\ \ln x_i, & \alpha_i = 0 \end{cases}$. It is convenient to choose the initial guesses $\alpha_i^{(0)} = 1$, which is equivalent to no transformation of the data. We begin with the model

$$\eta = f(\vec{\xi}, \vec{\beta}) \text{ with } \xi_i = x_i^{\alpha_i}, \text{ and } \alpha_i^{(0)} = 1, \; i = 1, 2, \ldots, k.$$

Adapting from Equation 5.1,

$$\eta_u = f(\vec{x}_u, \vec{\beta}) + \sum_{i=1}^{k} (\alpha_i - 1) \left\{ \frac{\partial f(\vec{\xi}, \vec{\beta})}{\partial \alpha_i} \right\}_{\substack{\vec{\xi} = \vec{x}_u \\ \alpha_i = 1}} .$$

Since

$$\left\{ \frac{\partial \xi_{ui}}{\partial \alpha_i} \right\} = x_{ui} \ln x_{ui},$$

the above statement becomes

$$\eta_u = f(\vec{x}_u, \vec{\beta}) + \sum_{i=1}^{k} (\alpha_i - 1) \left\{ \frac{\partial f(\vec{\xi}, \vec{\beta})}{\partial \xi_i} \right\}_{\vec{\xi} = \vec{x}_u} x_{ui} \ln x_{ui}.$$

The quantities $\left\{\frac{\partial f(\vec{\xi}, \vec{\beta})}{\partial \xi_i}\right\}$ are estimated by fitting

$$\hat{y}_u = f(\vec{x}_u, \vec{b}),$$

then differentiating as described previously. Let

$$z_{ui} = \left\{\frac{\partial f(\vec{x}_u \vec{b})}{\partial x_{ui}}\right\} x_{ui} \ln x_{ui}.$$

To obtain estimates $a_i - 1$ of $\alpha_i - 1$, fit the equation

$$\eta_u = f(\vec{x}_u, \vec{\beta}) + \sum_{i=1}^{k} (\alpha_i - 1) z_{ui}.$$

Once an estimation a_i of α_i is obtained, new independent variables are computed using the transformation $x' = x^{a_i}$, and the process can be repeated.

If we are working with a linear function in the ξ_i's, the response function is of the form

$$\eta_u = f(\vec{\xi}_u, \vec{\beta}) = \beta_0 + \sum_{i=1}^{k} \beta_i \xi_{ui},$$

where $\xi_{ui} = x_{ui}^{\alpha_i}$. From the previously described procedure, we would first fit

$$\hat{y}_u = b_0 + \sum_{i=1}^{k} b_i x_{ui}$$

to approximate $f(\vec{\xi}_u, \vec{\beta})$, then differentiate with respect to $\xi_i = x_i$ to

obtain

$$\left\{ \frac{\partial f(\vec{\xi}_u, \vec{b})}{d\vec{x}_u} \right\} = [b_1, b_2, \ldots, b_k].$$

Then we would fit

$$\hat{y}'_u = b'_0 + \sum_{i=1}^{k} b'_i x_{ui} + \sum_{i=1}^{k} c_i x_{ui} \ln x_{ui} ,$$

where $c_i = (a_i - 1)b_i$. The approximation of α_i would be provided by $a_i = \frac{c_i}{b_i} + 1$. A second iteration of the process can then be performed, using as data the "new" independent variables $x'_{ui} = x_{ui}^{a_i}$. An example of this procedure is provided by a set of data generated, without error, by the functions

$$\eta_1 = 10 + \sqrt{x} ,$$

$$\eta_2 = 10 + \ln x,$$

$$\eta_3 = 10 + x^{-1}$$

$$\eta_4 = 10 + x^{-2},$$

for values of x = 1, 2, 3, 4, 5. This data is presented in Table V-1. We first obtain estimates of a = .48, -.07, -1.13, -1.97 for η_1, η_2, η_3, and η_4 respectively. These first approximations are close to the

Table V-1. Generated Data and Application of Iterative Method

x		η_1	η_2	η_3	η_4
1		11.00	10.00	11.00	11.00
2		11.41	10.69	10.50	10.25
3		11.73	11.09	10.33	10.11
4		12.00	11.38	10.25	10.06
5		12.24	11.60	11.20	10.04
	b =	0.3058	0.3912	-0.1850	-0.2108
	c =	-0.1593	-0.4180	0.3940	0.6260
$a = \frac{c}{b} + 1$		0.48	-0.07	-1.13	-1.97

Source: G. E. P. Box and P. W. Tidwell, "Transformations of the Independent Variables," Technometrics 4 (November, 1962), 537.

actual values α = .50, 0, -1, -2. The approximations can be refined by setting $\alpha^{(1)}$ = a for each η_i, then performing a second iteration as described.

If the representative function $f(\vec{x}_u, \vec{\beta})$ is quadratic in the ξ_i's and the power transformation is used, then the equation will be of the form

$$\eta_u = f(\vec{\xi}_u, \vec{\beta}) = \sum_{i=0}^{k} \sum_{\substack{j=i \\ j \geq i}}^{k} \beta_{ij}, \xi_{ui} \xi_{uj}$$

where $\xi_{ui} = x_{ui}^{\alpha_i}$, $\beta_{0j} = \beta_j$, $x_{u0} = 1$ and $\beta_{00} = \beta_0$. Following the iterative

procedure, we first fit

$$\hat{y}_u = \sum_{i=1}^{k} \sum_{\substack{j=1 \\ j \geq i}}^{k} b_{ij} x_{ui} x_{uj}.$$

we differentiate to approximate $\frac{\partial f(\vec{x}_u, \vec{\beta})}{\partial x_{ui}}$, obtaining

$$\frac{\partial f(x_{ui}, b_i)}{\partial x_{ui}} = b_i + 2b_{ii} x_{ui} + \sum_{j=i+1}^{k} b_{ij} x_{uj},$$

and construct the new independent variables

$$z_{ui} = \{b_i + 2b_{ii} x_{ui} + \sum_{j=i+1}^{k} b_{ij} x_{uj}\} x_{ui} \ln x_{ui}.$$

The equations

$$\hat{y}_u = \sum_{i=0}^{k} \sum_{\substack{j=i \\ j \geq i}}^{k} b'_{ij} x_{ui} x_{uj} + \sum_{i=1}^{k} (a_i - 1) z_{ui}$$

can now be refit to obtain estimates a_i of α_i.

Considering an example of data with error, if a linear model is fit to the data in Table V-2, the equation

$$\hat{y} = 25.4575 + 13.729x$$

is obtained. The analysis of variance in Table V-3 shows F = 26.8 for lack of fit, significant at the .05 level, so that there is reason to question the adequacy of the linear fit. If the iterative

Table V-2. Quantity of Polymer Formed in One Hour Under Steady State Conditions at Equally Spaced Charges of Catalyst

Grams of Catalyst Charged to Reactor x	Transformed Data x	Quantity of Polymer Formed y_1	y_2
0.5	.7071	27.86	29.04
1.5	1.2247	47.70	47.86
2.5	1.5811	65.78	62.92
3.5	1.8708	74.18	75.16
4.5	2.1213	82.83	84.47

Source: G. E. P. Box and P. W. Tidwell, "Transformations of the Independent Variables," Technometrics 4 (November, 1962), 548.

Table V-3. Analysis of Variance of Original Data in Table V-2

Source	S.S.	D.F.	M.S.	F
Regression	39506.1928	2		
Lack of Fit	106.1768	3	35.3923	26.8*
Error	6.6238	5	1.3248	
Total	39618.9934	10		

*Significant at .05 level.

Source: G. E. P. Box and P. W. Tidwell, "Transformations of the Independent Variables," Technometrics 4 (November, 1962), 549.

method is applied to the data, the first estimate of α is $a = -\frac{8.0844}{13.729} + 1 = 0.411$. Other iterations lead to an estimate of a close to .5, suggesting the square root transformation. After the data is transformed in this fashion (see Table V-2) the model

$$\hat{y} = .3654 + 39.5826x^{\frac{1}{2}}$$

is obtained. The analysis of variance in Table V-4 gives little reason to doubt the adequacy of the model.

Table V-4. Analysis of Variance of Transformed Data in Table V-2.

Source	S.S.	D.F.	M.S.	F
Regression	39605.4044	3		
Lack of Fit	6.9652	2	3.4876	2.63
Error	6.6238	5	1.3248	
Total	39618.9934	10		

Source: G. E. P. Box and P. W. Tidwell, "Transformations of the Independent Variables," Technometrics 4 (November, 1962), 549.

BIBLIOGRAPHY

BIBLIOGRAPHY

Anscombe, F. J., "The Transformation of Poisson, Binomial, and Negative-Binomial Data," Biometrika, Vol. 35 (1948), 246-254.

Bartlett, M. S., "The Use of Transformations," Biometrics, Vol. 3 (1947), 39-52.

Box, G. E. P., "Non-normality and Tests on Variance," Biometrika, Vol. 40, (1953), 318-355.

Box, G. E. P. and P. W. Tidwell, "Transformation of the Independent Variables," Technometrics, Vol. 4 (1962), 531-550.

Dolby, J. L., "A Quick Method for Choosing a Transformation," Technometrics, Vol. 5 (1963), 317-325.

Draper, N. R. and H. Smith, Applied Regression Analysis, New York: John Wiley and Sons, Inc., 1966.

Rao, C. R., Linear Statistical Inference and Its Applications, 2nd. ed., New York: John Wiley and Sons, Inc., 1973.

Snedecor, George W. and William G. Cochran, Statistical Methods, 6th ed., Ames, Iowa: The Iowa State University Press, 1967.

Tukey, J. W., "On the Comparative Anatomy of Transformations," Annals of Mathematical Statistics, Vol. 28 (1957), 602-632.

Winer, B. J., Statistical Principles in Experimental Design, New York: McGraw-Hill, Inc., 1971.

VITA

Nancy Holcomb Willis was born in Toccoa, Georgia, on February 26, 1954. She graduated from Westminster High School in Westminster, South Carolina in May, 1971 and the following summer entered the University of North Carolina at Asheville. In 1973 she married John G. Willis, and in September, 1974 she and her husband transferred to East Carolina University in Greenville, North Carolina. She was elected to membership in Pi Mu Epsilon and Phi Kappa Phi while attending there, and was awarded a Bachelor of Arts degree with a major in Mathematics in May, 1975. In the fall of 1975, she accepted a teaching assistantship at the University of Tennessee and began study toward a Master of Science with a major in Mathematics. This degree was awarded in June, 1977.