A MANUAL OF TRANSFORMATIONS (AND GRAPHS) USEFUL IN STATISTICS

J. P. Mandeli and D. B. Hall

BU-524-M First Draft, July, 1974

ABSTRACT

A manual is given for the two transformations log Y and e^{aY}. The manual in its final form would provide easy reference to many transformations that can be useful in the statistical analysis of data.

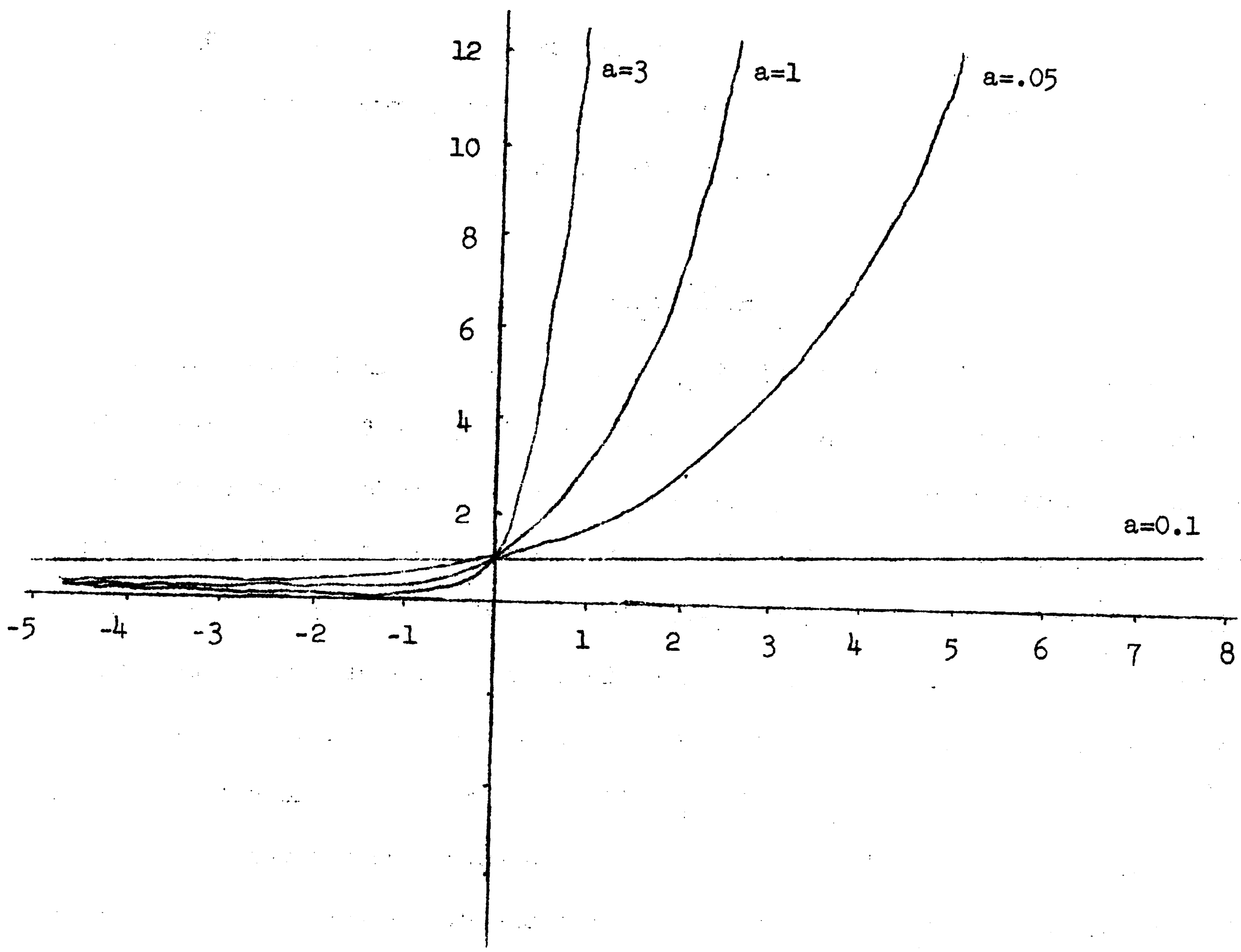
INTRODUCTION

On many occasions the statistical analysis of data is improved by making a transformation on the data, e.g., $Y \rightarrow \log Y$. The object of this manual is to provide easy reference to information that helps one decide what transformation(s) might be suitable for experimental data. One use of the manual is that when data can be plotted the resulting plot should be compared with the graphs in Part 1. From this comparison (including consideration of properties of the curves in Part 1 with the known properties of your data, e.g. limits, maxima, minima, slope, differential equations and some specific values) select a function $y = f(x)$ which appears to fit your data. Then use the transformation(s) suggested for the function chosen and use Part 2 for the properties of that transformation(s).

Part 1

Fitting Data

Graph of $y = e^{ax}$



Equation: $y = e^{ax}$ $(a > 0)$

$x = 0$, $y = 1$ for all a

$y = 0$, x does not exist

Limits: $y \to 0$ as $x \to -\infty$

$y \to \infty$ as $x \to \infty$

Differential Equations: $\frac{dy}{dx} = ay$ and in general $\frac{d^n y}{dx^n} = a^n y$ (n positive integer)

Maxima, Minima, Points of Inflexion: There exists no maximum or point of inflexion. The minimum is $x = 0$. The function is monotone increasing.

Transformation to make regression linear: For the model $y = e^{ax}\epsilon$, where ϵ is a multiplicative error term, the transformation $z = \log Y$ has the model $z = ax + \epsilon'$ with $\epsilon' = \log \epsilon$. (See page 8 for properties of the transformation $z = \log Y$.)

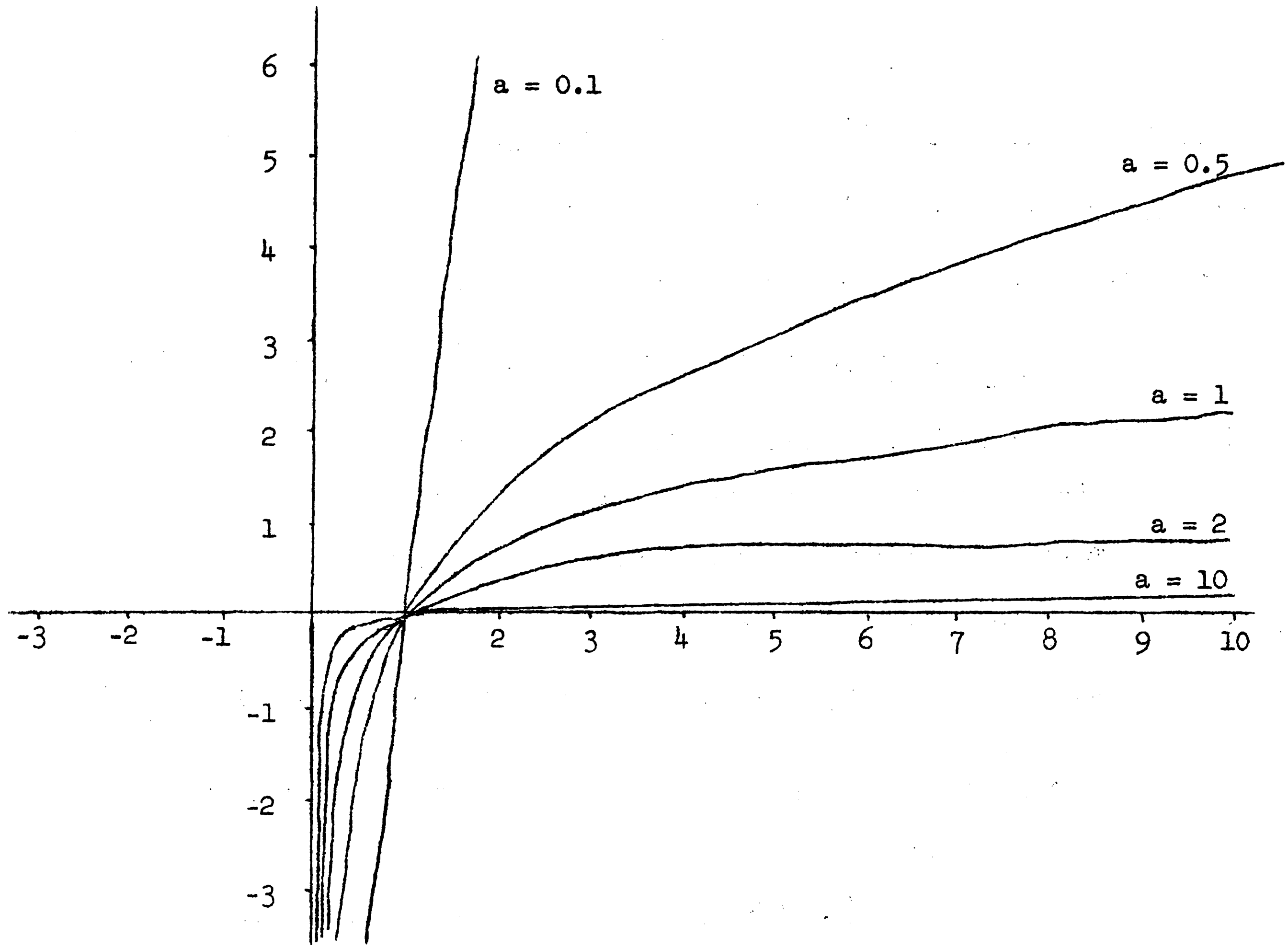
Transformation to get homogeneous variance: When $Y = e^{aX}$ with $X \sim (\mu_x, \sigma_x^2)$, $Y \sim [e^{a\mu_x}, a^2\sigma_x^2(e^{a\mu_x})^2]$ with the variance dependent on the mean. But $\log Y \sim (a\mu_x, a^2\sigma_x^2)$ (See page 8 for properties of the transformation $\log Y$.)

Values of y for certain x and a

$$y = e^{ax}$$

x \ a	0.1	0.5	1	2	4	10
-10	0.368	0.007	0.00005	$\to 0$	$\to 0$	$\to 0$
-8	0.449	0.018	0.0003	$\to 0$	$\to 0$	$\to 0$
-6	0.549	0.0498	0.0025	$\to 0$	$\to 0$	$\to 0$
-4	0.670	0.135	0.018	0.0003	$\to 0$	$\to 0$
-2	0.819	0.368	0.135	0.018	0.0003	$\to 0$
0	1.00	1.00	1.00	1.00	1.00	1.00
1	1.105	1.649	2.718	7.389	54.60	22026
2	1.221	2.718	7.389	54.60	2981.0	$\to \infty$
4	1.492	7.389	54.60	2981.0	$\to \infty$	$\to \infty$
6	1.822	20.086	403.43	$\to \infty$	$\to \infty$	$\to \infty$
8	2.2255	54.60	2981.0	$\to \infty$	$\to \infty$	$\to \infty$
10	2.718	148.41	22026	$\to \infty$	$\to \infty$	$\to \infty$

Graph of $y = \frac{\log x}{a}$



Equation: $y = \frac{\log x}{a}$ $(a > 0)$

$x = 0$, y does not exist

$y = 0$, $x = 1$ for all a

Limits: $y \to -\infty$ as $x \to 0$

$y \to \infty$ as $x \to \infty$

Differential Equations: $\frac{dy}{dx} = \frac{1}{ax}$ and $\frac{d^n y}{dx^n} + \frac{1}{x}\frac{d^{n-1}y}{dx^{n-1}} = 0$ for all integers $n > 1$.

Maxima, Minima, Points of Inflexion: There exist none for $x > 0$.

Continuity: $y = \frac{\log x}{a}$ is continuous for all positive x.

Maximum Gradient: As x approaches 0 the slope of $\frac{\log x}{a} = y$ increases without limit.

Transformation to make regression linear: For the model $y = \frac{\log x}{a} + \epsilon$ the transformation $z = e^{aY}$ has the model $z = x\epsilon'$ with $\epsilon' = e^{a\epsilon}$, where ϵ and ϵ' are error terms. (See page 9 for properties of the transformation $z = e^{aY}$.

Transformation to get homogeneous variance: When $Y = \frac{\log X}{a}$ with $X \sim (\mu_X, \sigma_X^2)$

$$Y \sim \left(\frac{\log \mu_x}{a}, \frac{\sigma_x^2}{a^2\mu_x^2}\right)$$ with the variance dependent on the mean.

But $e^{aY} \sim (\mu_x, \sigma_x^2)$. (See page 9 for properties of the transformation e^{aY}.)

Values of y for certain x and a

$$y = \frac{\log x}{a}$$

x \ a	0.1	0.5	1	2	4	10
0	--	--	--	--	--	--
0.1	-23.03	-4.61	-2.30	-1.15	-0.58	-0.23
0.5	-6.93	-1.386	-0.693	-0.347	-0.173	-0.069
1	0	0	0	0	0	0
2	6.93	1.386	0.693	0.347	0.173	0.069
4	13.86	2.77	1.386	0.693	0.347	0.139
6	17.92	3.584	1.792	0.8959	0.448	0.179
8	20.79	4.159	2.079	1.0397	0.520	0.208
10	23.026	4.6052	2.303	1.151	0.576	0.230
20	29.957	5.991	2.996	1.498	0.749	0.2996
30	34.012	6.802	3.401	1.701	0.850	0.340
40	36.889	7.378	3.689	1.84	0.922	0.369

Part 2

Transformations

Transformation: Log Y

Moments of log Y, given the moments of Y:

For

$$Y \sim (\mu_Y, \sigma_Y^2)$$

$$\log Y \mathrel{\dot\sim} \left(\log \mu_Y, \frac{\sigma_Y^2}{\mu_Y^2}\right)$$

using 1-term Taylor expansion.

Or

$$E(\log Y) \mathrel{\dot\sim} \log \mu_Y - \frac{\sigma_Y^2}{2\mu_Y^2}$$

using 2-term Taylor expansion.

Distribution of log Y, given distribution of Y:

Let $F(y)$ be the distribution function of Y.

Then $G(y) = F(e^y)$ is the distribution function of log Y.

Transformation: e^{aY} $\quad a > 0$

Moments of e^{aY}, given the moments of Y:

For

$$Y \sim (\mu_Y, \sigma_Y^2)$$

$$e^{aY} \mathrel{\dot\sim} (e^{a\mu_Y}, a^2 e^{2a\mu_Y} \sigma_Y^2)$$

using 1-term Taylor expansion.

Or

$$E(e^{aY}) \doteq e^{a\mu_Y} \left(1 + \frac{a^2}{2} \sigma_Y^2\right)$$

using 2-term Taylor expansion.

Distribution of e^{aY}, given distribution of Y:

Let F(y) be the distribution function of Y.

Then $G(y) = F(\frac{\log Y}{a})$ is the distribution function of e^{aY}.

(i) When Y has the exponential distribution with parameter λ,

$$F(y) = \begin{cases} 1 - e^{-\lambda y}, & y \geq 0 \\ 0, & y < 0. \end{cases}$$

e^{aY} has the distribution

$$G(y) = F(\frac{\log y}{a}) = \begin{cases} 1 - y^{-\lambda/a}, & y \geq 1 \\ 0, & y < 1 \end{cases}$$

with

$$E(e^{aY}) = \begin{cases} \frac{\lambda}{\lambda - a}, & \lambda > a \\ \infty, & \lambda \leq a \end{cases}$$

$$\mathrm{Var}\,(e^{aY}) = \begin{cases} \frac{\lambda a^2}{(\lambda-2a)(\lambda-a)^2}, & \lambda > 2a \\ \infty, & \lambda \leq 2a \end{cases}$$

(ii) When Y has the normal distribution with mean μ_Y and variance σ_Y^2 e^{aY} has the lognormal distribution

$$G(y) = \begin{cases} \dfrac{1}{a\sigma_Y y\sqrt{2\pi}} \, e^{-\dfrac{(\log y - a\mu_Y)^2}{2a^2\sigma_Y^2}} & , \quad y > 0 \\ 0 & , \quad y \le 0 \end{cases}$$

with

$$E(e^{aY}) = e^{(a\mu_Y + \frac{1}{2}a^2\sigma_Y^2)}$$

$$\mathrm{Var}\,(e^{aY}) = e^{(2a\mu_Y + 2a^2\sigma_Y^2)} - e^{(2a\mu_Y + a^2\sigma_Y^2)}$$

(iii) When Y has the uniform distribution on the interval (y_1, y_2) e^{aY} has the distribution

$$G(y) = \begin{cases} \dfrac{\log y - ay_1}{a(y_2 - y_1)} & , \quad e^{ay_1} \le y \le e^{ay_2} \\ 0 & , \quad \text{elsewhere} \end{cases}$$

with

$$E(e^{aY}) = \frac{e^{ay_2} - e^{ay_1}}{a(y_2 - y_1)}$$

$$\mathrm{Var}\,(e^{aY}) = \frac{e^{2ay_2} - e^{2ay_1}}{2a(y_2 - y_1)} - \frac{1}{a^2(y_2 - y_1)^2}(e^{ay_2} - e^{ay_1})^2$$