Variance Stabilization Transformations

David Allen
University of Kentucky

January 22, 2013
1 Variance stabilization transformations

If the assumptions for a linear model are not satisfied, transformation of the data may help. Here we describe the variance stabilization transformation that is applied to the response variable.
Suppose we have a random variable $Y$ with mean $\mu$ and variance $g(\mu)$. Our objective is to find a monotone function $h(Y)$ such that $\text{Var}(h(Y))$ is nearly constant. We approximate $h(Y)$ by

$$h(Y) \approx h(\mu) + h'(\mu)(Y - \mu)$$

where $h'$ is the derivative of $h$. 
Variance Approximation

The variance of the approximation is

\[(h'(\mu))^2 g(\mu)\].

Setting this equal to a constant \(c\), rearranging the expression, and replacing \(\mu\) with a more conventional variable \(t\), gives the differential equation

\[h'(t) = \frac{c}{\sqrt{g(t)}}.\]

We now consider some specific \(g(t)\).
The square root transformation

Suppose \( g(t) = t \) as it is with the Poisson distribution. The differential equation is

\[
h'(t) = ct^{-\frac{1}{2}}
\]

with solution

\[
h(t) = 2ct^{\frac{1}{2}}.
\]

For convenience, set \( c = 0.5 \) to yield the square root transformation.
The logarithmic transformation

Suppose \( g(t) = t^2 \) as it does when there are multiplicative errors. The differential equation is

\[
h'(t) = \frac{c}{t}
\]

and its solution is

\[h(t) = c \log(t)\].

For convenience, set \( c = 1 \) to yield the logarithmic transformation.
A generalized power transformation

Suppose we have a random variable $Y$ with mean $\mu$ and variance $\mu^k$. We have looked at this situation when $k = 1$ and $k = 2$; the same methodology can be applied for non-integer values of $k$. The differential equation is

$$h'(t) = \frac{c}{t^{k/2}}.$$ 

When $k = 2$ we have $h(t) = \log(t)$ as previously shown. The solution of the equation for $k \neq 2$ is

$$\frac{c}{-k/2 + 1} t^{-k/2 + 1} + \alpha$$

where $\alpha$ is a constant of integration.
To simplify the expression, define $\lambda = -k/2 + 1$. Set $c = 1$ and $a = -1/\lambda$. The transformation is

$$h(t) = \begin{cases} \frac{(t^\lambda - 1)}{\lambda} & \text{if } \lambda \neq 0 \\ \log(t) & \text{if } \lambda = 0 \end{cases}$$

This transformation is due to Box and Cox [1].

As an exercise, show that $\lim_{\lambda \to 0} (t^\lambda - 1)/\lambda = \log(t)$. Indeed, the constant $a$ was chosen to provide this continuity.
The arc sine square root transformation

If $\hat{p}$ is a sample binomial proportion, then

$$g(t) = \frac{t(1-t)}{n}$$

where $n$ is the sample size. The differential equation is

$$h'(t) = \frac{c\sqrt{n}}{\sqrt{t(1-t)}}.$$ 

This equation is most easily solved using the trigonometric substitution $\sqrt{t} = \sin(\theta)$ and $\sqrt{1-t} = \cos(\theta)$. 
Relationship between \( t \) and \( \theta \)

The relationship between \( t \) and \( \theta \) is emphasized in the triangle:

Diagram goes here
Solving the Equations

The equation in terms of $\theta$ and the steps for solving the equation are

$$h'(\sin^2(\theta)) = \frac{c\sqrt{n}}{\sin(\theta)\cos(\theta)}$$

$$h'(\sin^2(\theta))2\sin(\theta)\cos(\theta) = 2c\sqrt{n}$$

$$h(\sin^2(\theta)) = 2c\sqrt{n}\theta + a$$

where $a$ is a constant of integration. For convenience we let $c = 0.5$ and $a = 0.0$. Back substitute $t$ for $\theta$ to obtain

$$h(t) = \sqrt{n}\arcsin\left(\sqrt{t}\right).$$
Exercise 1.1. Suppose

\[ y = \frac{1}{\theta + \epsilon} \]

where \( \theta \) is a parameter and \( \epsilon \) is a random variable with mean zero and variance one. Approximate \( y \) with a linear function of \( \epsilon \). (Hint: first two terms of Taylor series about zero) Give the mean and variance of the approximation. Give \( \lambda \) of the appropriate Box-Cox transformation on \( y \).