Applications of a new power normal family

Isogai, Takafumi

(Citation)
Journal of Applied Statistics,32(4):421-436

(Issue Date)
2005-06

(Resource Type)
journal article

(Version)
Accepted Manuscript

(URL)
https://hdl.handle.net/20.500.14094/90000135

Applications of a new power normal family

Takafumi Isogai
Faculty of Maritime Sciences,
Kobe University,
Fukae, Higashinada, Kobe 658-0022, Japan
E-mail:isogai@maritime.kobe-u.ac.jp

Abstract

The main purpose of this paper is to give an algorithm to attain joint normality of non-normal multivariate observations through a new power normal family introduced by the author (1999). The algorithm tries to transform each marginal variable simultaneously to joint normality, but due to a large number of parameters it repeats a maximization process with respect to the conditional normal density of one transformed variable given the other transformed variables. A non-normal data set is used to examine performance of the algorithm, and the degree of achievement of joint normality is evaluated by measures of multivariate skewness and kurtosis. Besides the above topic, making use of properties of our power normal family, we discuss not only a normal approximation formula of noncentral F distributions in the frame of regression analysis but also some decomposition formula of a power parameter appeared in a Wilson-Hilferty power transformation setting.

Keywords

Power normal family, non-normality, joint normality, measures of multivariate skewness and kurtosis

1. Introduction

A new power normal family has been constructed by Isogai (1999) from the generalized F distribution (see Prentice (1975)) through a Wilson-Hilferty type power transformation of an F random variable. Let $F_{\phi_2}^{\phi_1}$ be the random variable of a central F distribution with degrees of freedom (ϕ_1, ϕ_2) . When $\phi_1 > 2$ and $\phi_2 > 2$, our Wilson-Hilferty transformation of the $F_{\phi_2}^{\phi_1}$ variable is given by

$$\frac{\left(F_{\phi_2}^{\phi_1}\right)^{\tilde{h}} - \left(\tilde{F}_{\phi_2}^{\phi_1}(0.5)\right)^{\tilde{h}}}{\tilde{h}\left[2\left(1/\phi_1 + 1/\phi_2\right)\right]^{1/2}}, \qquad (1)$$

where $\tilde{h}$ is a power parameter defined by

$$\tilde{h} = -\frac{1}{3}\left(\frac{\phi_1 - \phi_2}{\phi_1 + \phi_2}\right)$$

and $\tilde{F}_{\phi_2}^{\phi_1}(0.5)$ is an approximate formula for the median of $F_{\phi_2}^{\phi_1}$ defined by

$$\tilde{F}_{\phi_2}^{\phi_1}(0.5) = \exp\left[-\frac{2}{3}\left(\frac{1}{\phi_1} - \frac{1}{\phi_2}\right)\right]. \tag{2}$$

On the other hand, a random variable X of the generalized F distribution is defined by

$$X = \eta\left(F_{2\phi_2}^{2\phi_1}\right)^{\gamma} \tag{3}$$

for $\eta > 0, \gamma > 0, \phi_1 > 1$ and $\phi_2 > 1$. Substituting (3) into (1) and reparametrizing parameters in X, our new power normal family is given by

$$\frac{1}{\delta}\left\{\left(\frac{X}{\eta}\right)^{\delta/\sigma} - \exp\left(-\delta^2\right)\right\} \tag{4}$$

for $\eta > 0, \sigma > 0$ and

$$-\frac{1}{3} < \delta < \frac{1}{3}, \tag{5}$$

where we put

$$\begin{cases} \delta = \delta_1\delta_2, \\ \sigma = \gamma\delta_2, \\ \delta_1 = \frac{1}{3}\left(\frac{1}{\phi_1} - \frac{1}{\phi_2}\right) / \left(\frac{1}{\phi_1} + \frac{1}{\phi_2}\right), \\ \delta_2 = \left(\frac{1}{\phi_1} + \frac{1}{\phi_2}\right)^{1/2}. \end{cases} \tag{6}$$

The quantity (4) is approximately distributed as the standard normal variate. As δ tends to zero, $\log X$ is nearly distributed as the normal distribution with mean μ (= $\log\eta$) and variance σ^2. When δ is near $1/3$, the distribution of X is approximated by the Weibull distribution, of which distribution function is given by

$$\Pr\left(X \le x\right) = 1 - \exp\left\{-\left(\frac{x}{\eta}\right)^{\frac{1}{\sigma}}\right\}$$

for $x \ge 0$. On the other hand, when δ is near $-1/3$, the distribution of X is approximated by the extreme value distribution of Type 2 (see Johnson & Kotz (1970)):

$$\Pr\left(X \le x\right) = \exp\left\{-\left(\frac{x}{\eta}\right)^{-\frac{1}{\sigma}}\right\}$$

for $x \ge 0$.

Here, in (4), from (5) we can replace $\exp\left(-\delta^2\right)$ by one, and further we set

$$\frac{\delta}{\sigma} = \rho, \frac{\eta^{\rho} - 1}{\rho} = \xi, \sigma\eta^{\rho} = \tau. \tag{7}$$

Then, we have the Box & Cox (1964) power normal family:

$$\frac{X^{\rho}-1-\rho\xi}{\rho\tau} \tag{8}$$

for $\tau > 0$. It is well known that $(X^{\rho}-1)/\rho$ is nearly distributed as the normal distribution with mean ξ and variance τ^2. Parametrization (7) indicates some functional relationship among ξ , τ and ρ in (8) such that

$$\tau = \sigma(1+\rho\xi).$$

This relationship suggests that there exist high correlations among estimates of ξ , τ , and ρ in the Box & Cox power normal family.

Under the scale transformation of $X \mapsto \theta X$ ($\theta > 0$), parameters δ and σ of the model (4) are invariant and μ $(=\log\eta)$ is converted to $\mu + \log\eta$. Under the same transformation of X, ρ of the model (8) is invariant, and ξ and τ are converted to $\theta^{\rho}\xi + (\theta^{\rho}-1)/\rho$ and $\theta^{\rho}\tau$ respectively. Difference in responses of parameters of the above two models leads to the following serious result. Under the scale transformation of X, the correlation structure of estimates (especially, maximum likelihood estimates) of δ, σ and μ in the model (4) is invariant, but that of estimates of ρ, τ and ξ in the model (8) is not. The last fact means that changes of units of measurement in X give us different correlation structures of estimates of ρ, τ and ξ in the model (8).

Also note that under the power transformation of $X \mapsto X^{\theta}$, the parameter δ of the model (4) is invariant and σ and μ $(=\log\eta)$ are converted to $\theta\sigma$ and $\theta\mu$ respectively. Under the same power transformation of X, parameters ρ, τ and ξ of the model (8) are converted to $\rho/\theta, \theta\tau$ and $\theta\xi$ respectively. Although correlation structures of m.l.e.'s with respect to both models are invariant under the power transformation of X, the value of an estimate of ρ in the model (8) is not meaningful unless the dimension of a scale of measurement is known.

In the section 2 we shall discuss some further properties of our new power normal family (4). Using our new power normal family, we first consider normal approximations of noncentral F distributions. Our basic idea is to use F approximations of noncentral F distributions, because we have very good F approximation formulas (see Patnaik (1949) and Tiku (1966)). The Patnaik's F approximation formula enables us to deal with regression structures in the noncentral parameter of noncentral F distributions like those in the scale parameter η of our new power normal family.

Besides, we try to express parameters δ, σ and η in terms of the first three cumulants κ_i ($i = 1,2,3$) of X in (4). Such descriptions of δ and σ give us a fundamental formula of the power parameter h (= δ/σ) :

$$h = 1 - \frac{1}{3}\frac{\kappa_3\kappa_1}{(\kappa_2)^2}. \tag{9}$$

Thus, our expressions of δ and σ in terms of cumulants give us a decomposition of the power parameter h.

In the section 3, for a given random sample of multivariate observations we try to achieve joint normality of it by applying our power normal family (4) to each marginal component. Let $X' = \left(X_1, X_2, ..., X_p\right)$ be a p-dimensional random vector defined on the positive domain, and let $\left(\phi_1(X_1), \phi_2(X_2), ..., \phi_p(X_p)\right)$ be the transformed random vector, where $\phi_j(X_j)$ $\left(j = 1, 2, ..., p\right)$ is given by

$$\phi_j(X_j) = \frac{1}{\delta_j}\left[\exp\left\{\delta_j\left(\frac{\log X_j - \mu_j}{\sigma_j}\right)\right\} - \exp\left(-\delta_j^2\right)\right], \qquad (10)$$

where δ_j, μ_j and σ_j $\left(j = 1, ..., p\right)$ are parameters chosen to make the joint distribution of $\left(\phi_1(X_1), \phi_2(X_2), ..., \phi_p(X_p)\right)$ nearly equal to a multivariate normal distribution.

As an increase of the dimension of a random vector causes a rapid increase of the number of parameters to be estimated, we have to avoid a direct application of the likelihood function of the joint distribution. Thus, our estimation procedure is based on the well known fact that for a given subset of the random vector distributed as a multivariate normal distribution, the conditional distribution of the rest subset of the random vector is multivariate normal. Especially, our iteration algorithm for estimation adopts a special case of the above property where one of the two subsets consists of one component and the other has $p-1$ components, and repeats maximization process with respect to the conditional likelihood function of $\phi_j(X_j)$ given $\phi_i(X_i)$ $\left(i = 1, 2, ..., p; i \neq j\right)$ from $j = 1$ to p.

We give an example to show performance of our estimation algorithm. The degree of achievement of joint normality is evaluated by measures of multivariate skewness and kurtosis (see Mardia (1970) and Isogai (1983a, 1983b, 1989)).

Finally we give some implications of achievement of joint normality. One of our conclusions is that to attain joint normality brings us with a high correlation structure among transformed variables, which sometimes causes near multi-collinearity. As a consequence, the minimum eigenvalue, which is about zero, of the correlation matrix introduces a linear relationship

$$\alpha_1\phi_1\left(X_1\right) + \alpha_2\phi_2\left(X_2\right) + \cdots + \alpha_p\phi_p\left(X_p\right) \approx 0,$$

where $\alpha = \left(\alpha_1, ..., \alpha_p\right)'$ is the eigenvector corresponding to the minimum eigenvalue. This relationship defines a hypersurface in a p-dimensional space, on which the random vector X is distributed. So, we may call it a principal surface.

2. Some properties of a new power normal family

2.1. Normal approximation to noncentral F distributions

Let $F_{\phi_2}^{\phi_1}(\lambda)$ be the random variable of the noncentral F distribution with (ϕ_1, ϕ_2) degrees of freedom and noncentrality parameter λ. From (1) we know that good F approximations to $F_{\phi_2}^{\phi_1}(\lambda)$ give us good normal approximations to that. Tiku (1966) has obtained a quite accurate F approximation such as $b + cF_{\phi_2}^{\phi_1'}$, where $F_{\phi_2}^{\phi_1'}$ is the central F variable with $(\phi_1^{'}, \phi_2)$ degrees of freedom and b, c and $\phi_1^{'}$ are chosen so as to make the first three moments of $F_{\phi_2}^{\phi_1}(\lambda)$ and those of $b + cF_{\phi_2}^{\phi_1^{'}}$ agree. Tiku (1966) has also shown in his comparative study that the accuracy of the Patnaik's (1949) F approximation is fairly good among various approximations of $F_{\phi_2}^{\phi_1}(\lambda)$. The Patnaik's (1949) result is that the distribution of $F_{\phi_2}^{\phi_1}(\lambda)$ can be approximated by the distribution of $(1+\lambda/\phi_1)F_{\phi_2}^{\phi^*}$, where $\phi^* = (\phi_1+\lambda)^2(\phi_1+2\lambda)^{-1}$.

Our concern here is to make a statistical inference on the noncentrality parameter λ rather than to obtain an accurate normal approximation to $F_{\phi_2}^{\phi_1}(\lambda)$. Due to its simple form the Patnik's approximation formula is suitable for our purpose. From (2) we obtain an approximate median formula of $F_{\phi_2}^{\phi_1}(\lambda)$ such as

$$\tilde{F}_{\phi_2}^{\phi_1}(\lambda; 0.5) = (1+\lambda/\phi_1)\exp\left[-\frac{2}{3}\left(\frac{1}{\phi^*} - \frac{1}{\phi_2}\right)\right]$$

and also from (1) we have a normal approximation formula of $F_{\phi_2}^{\phi_1}(\lambda)$ given by

$$\frac{\left(\frac{F_{\phi_2}^{\phi_1}(\lambda)}{\eta}\right)^{\frac{\delta}{\sigma}} - \exp(-\delta^2)}{\delta}, \tag{11}$$

where we put

$$\eta = 1 + \lambda/\phi_1,$$

$$\sigma = \left\{2\left(\frac{1}{\phi^*} + \frac{1}{\phi_2}\right)\right\}^{1/2},$$

$$\delta = \left(\frac{\sqrt{2}}{3}\right)\left(\frac{1}{\phi^*} - \frac{1}{\phi_2}\right) \Big/ \left(\frac{1}{\phi^*} + \frac{1}{\phi_2}\right)^{1/2}.$$

Let us consider a condition that we can introduce a regression structure into η through λ with the other parameters δ and σ regarded as nuisance terms. At least it is necessary that the distributional shape of $F_{\phi_2}^{\phi_1}(\lambda)$ in (11) is stable when observations and covariates are varying. This requirement is satisfied as long as the sign of δ is unchanged. The meaningful condition is $\phi^* \geq \phi_2$, which is attained by large enough λ. In this case δ has a negative sign, and σ moves around a constant. Thus, even when data come from a noncentral F population and we have to make an inference about covariates in sufficiently large λ of $F_{\phi_2}^{\phi_1}(\lambda)$, we can apply our new power normal family (4) approximately to such a population.

Typical examples in the above setting are yield or percentage data appearing in productive activities of manufacturing. Let us denote an observation of yield data by B ($0 < B < 1$), and in order to examine factors having effects on variations of B, we shall suppose that B is approximately distributed as a noncentral beta distribution with shape parameters ϕ_1 and ϕ_2 and noncentrality parameter λ. Using two mutually independent central and noncentral χ^2 variables, the noncentral beta variable $B_{\phi_2}^{\phi_1}(\lambda)$ is defined by $B_{\phi_2}^{\phi_1}(\lambda) = \chi_{2\phi_1}^2(\lambda) / \left(\chi_{2\phi_1}^2(\lambda) + \chi_{2\phi_2}^2 \right)$. Large λ indicates the large amount of yield, and here we suppose that λ is involved in various covariates such as control variables and environmental factors. Thus, we can use the model (4) with respect to $B/(1-B)$ instead of X to examine the influence of covariates through η, because $B/(1-B)$ is approximately distributed as $\left(\phi_1/\phi_2\right) F_{2\phi_2}^{2\phi_1}(\lambda)$.

In the preceding discussion, by approximating a noncentral F distribution by an appropriate central F, we eventually use a central beta distribution to describe the distribution of yield data. Thus, if from the beginning we suppose that B is approximately distributed as a central beta distribution, that is, we put $\lambda = 0$ in $B_{\phi_2}^{\phi_1}(\lambda)$, large ϕ_1 indicates the large amount of yield. In this case, to use the model (4) with respect to $B/(1-B)$ instead of X is to introduce a regression structure into ϕ_1.

2.2. Relationship between cumulants and parameters

Isogai (2001) has shown that under the assumptions:(i) $0 < \delta_2 < 1$ (i.e. $0 < 1/\phi_1 + 1/\phi_2 < 1$), (ii) $0 < \sigma < 1$ and (iii) $\sigma = O(\delta_2)$, expectation $E[X]$ and variance $V(X)$ of the model (4) are approximately given by

$$E[X] = \eta \exp\left(-\frac{3}{2}\sigma\delta + \frac{\sigma^2}{2} \right) (= \kappa_1, say), \tag{12}$$

$$V(X) = \eta^2 \exp\left(-3\sigma\delta + \sigma^2 \right) \left\{ \exp \sigma^2 - 1 \right\} (= \kappa_2, say). \tag{13}$$

Similarly, the third cumulant κ_3 of the model (4) is approximately given by

$$\kappa_3 = 3\eta^3\sigma^3 \left(\sigma - \delta \right). \tag{14}$$

For the derivation of κ_3, including κ_1 and κ_2, see Appendix. Here we remark that the condition (ii) $0 < \sigma < 1$ ensures unimodality of the distribution of (3) on the positive domain.

In the following we shall consider meanings of parameters of the model (4). First, from (12) we know a coarse relationship

$$\eta \approx \kappa_1.$$

Then, from (12) and (13), we have

$$\sigma \approx \frac{\kappa_2^{1/2}}{\kappa_1}. \tag{15}$$

The parameter σ has a meaning of the coefficient of variation. Furthermore, using the fact that $\kappa_2 \approx \eta^2\sigma^2$, from (14) we obtain

$$\delta \approx \sigma - \frac{1}{3}\frac{\kappa_3}{\kappa_2^{3/2}} \approx \frac{\kappa_2^{1/2}}{\kappa_1} - \frac{1}{3}\frac{\kappa_3}{\kappa_2^{3/2}} \tag{16}$$

and also

$$\frac{\delta}{\sigma} \approx 1 - \frac{1}{3}\frac{\kappa_3 \kappa_1}{\kappa_2^2}. \tag{17}$$

The expression of a power parameter h (= δ/σ) in (17) has been found first by Sankaran (1959) to obtain a normal approximation of noncentral χ^2 distributions, and rediscovered again by Mudholkar & Trivedi (1980) to obtain a normal approximation of the distribution of the Wilks' Λ statistic in multivariate analysis of variance. Therefore, expressions (15) and (16) of our parameters σ and δ are regarded as a decomposition of the power parameter h.

3. Attainment of joint normality

Let $\left(X_1, X_2, ..., X_p\right)$ be a p-dimensional random vector defined on the positive domain, and let $\left(\phi_1(X_1), \phi_2(X_2), ..., \phi_p(X_p)\right)$ be the transformed random vector, where $\phi_j(X_j)$ $\left(j = 1, 2, ..., p\right)$ is given by (10).

Now we suppose that $\left(\phi_1(X_1), \phi_2(X_2), ..., \phi_p(X_p)\right)$ has a p-variate normal density. Then, the conditional distribution of $\phi_j(X_j)$ is a univariate normal under the condition that the rest variables $\phi_i(X_i)$ $\left(i = 1, 2, ..., p; i \neq j\right)$ are given. Our estimation method consists of making the conditional distribution of $\phi_j(X_j)$ close to a univariate normal distribution.

A sketch of our iteration algorithm is as follows:

(Initialization) Initial values of parameters $\mu_j^{(0)}, \sigma_j^{(0)}$ and $\delta_j^{(0)}$ $\left(j = 1, ..., p\right)$ are chosen so as to make their marginal distributions of $\phi_j^{(0)}(X_j)$ near to univariate normal distributions.

(Update) Let $\phi_j^{(k)}(X_j)$ $\left(j = 1, ..., p\right)$ be solutions of the k-th iteration. New values of parameters $\mu_j^{(k+1)}, \sigma_j^{(k+1)}$ and $\delta_j^{(k+1)}$ of $\phi_j^{(k+1)}(X_j)$ are obtained so as to maximize the likelihood function with respect to the conditional univariate normal density of $\phi_j^{(k+1)}(X_j)$ under the condition that $\phi_i^{(k+1)}(X_i)$ $\left(i = 1, 2, ..., j-1\right)$ and $\phi_i^{(k)}(X_i)$ $\left(i = j+1, j+2, ..., p\right)$ are given.

(Judgment of convergence) Stop the further iteration if $\left|\delta_j^{(k+1)} - \delta_j^{(k)}\right|$ $\left(j = 1, 2, ..., p\right)$ are less than some specified value ε, otherwise continue the update. In our case, an candidate of the value of ε is 0.0001, which is small enough.

Here, let x_i = $\left(x_{i1}, x_{i2}, ..., x_{ip}\right)'(i = 1, ..., n)$ be a p-variate random sample of size n. We shall describe the update algorithm with respect to $\phi_j^{(k+1)}(x_{ij})$ $\left(i = 1, ..., n\right)$ in detail. The conditional univariate density of x_{ij} $\left(i = 1, ..., n\right)$ is given by

$$L = \frac{1}{\left(\sqrt{2\pi\tau^2}\right)^n} \exp\left\{-\frac{1}{2\tau^2}\left(\sum_{i=1}^{n}\phi_j^{(k+1)}(x_{ij}) - \sum_{m=1}^{j-1}\beta_m\phi_m^{(k+1)}(x_{im}) - \sum_{m=j+1}^{p}\beta_m\phi_m^{(k)}(x_{im})\right)^2\right\}$$

$$\times\prod_{i=1}^{n} d\phi_j^{(k+1)}(x_{ij}),$$

where

$$d\phi_j^{(k+1)}(x_{ij}) = \frac{1}{\sigma_j^{(k+1)}}\exp\left\{\delta_j^{(k+1)}\left(\frac{\log x_{ij} - \mu_j^{(k+1)}}{\sigma_j^{(k+1)}}\right)\right\}\frac{dx_{ij}}{x_{ij}}.$$

We first maximize $\log L$ with respect to τ^2 and $\left(\beta_1, \beta_2, ..., \beta_{j-1}, \beta_{j+1}, ..., \beta_p\right)(=\beta',$ say). Straight calculation leads to solutions:

$$\begin{cases} \hat{\beta} = \left(\Phi'\Phi\right)^{-1}\Phi'\tilde{\phi}_j^{(k+1)}\left(\tilde{x}_j\right), \\ \hat{\tau}^2 = (1/n)\tilde{\phi}_j^{(k+1)}\left(\tilde{x}_j\right)'\left\{I - \Phi\left(\Phi'\Phi\right)^{-1}\Phi'\right\}\tilde{\phi}_j^{(k+1)}\left(\tilde{x}_j\right), \end{cases} \tag{18}$$

where $\tilde{x}_j$ and $\tilde{\phi}_j^{(k+1)}\left(\tilde{x}_j\right)$ are $n\times 1$ vectors given by

$$\tilde{x}_j = \left(x_{1j}, x_{2j}, \cdots, x_{nj}\right)',$$

$$\tilde{\phi}_j^{(k+1)}\left(\tilde{x}_j\right) = \left(\phi_j^{(k+1)}(x_{1j}), \phi_j^{(k+1)}(x_{2j}), \cdots, \phi_j^{(k+1)}(x_{nj})\right)',$$

and also I is the identity matrix of order n, and Φ is an $n\times\left(p-1\right)$ matrix defined by

$$\Phi = \left(\tilde{\phi}_1^{(k+1)}(\tilde{x}_1) \quad \cdot\cdot \quad \tilde{\phi}_{j-1}^{(k+1)}(\tilde{x}_{j-1}) \quad \tilde{\phi}_{j+1}^{(k)}(\tilde{x}_{j+1}) \quad \cdot\cdot \quad \tilde{\phi}_p^{(k)}(\tilde{x}_p)\right).$$

Substituting the solutions (18) into $\log L,$ we obtain our target function

$$Q = -\frac{n}{2}\log 2\pi - \frac{n}{2} - \frac{n}{2}\log\hat{\tau}^2 + \delta_j^{(k+1)}\sum_{i=1}^{n}\left(\frac{\log x_{ij} - \mu_j^{(k+1)}}{\sigma_j^{(k+1)}}\right) - n\log\sigma_j^{(k+1)} - \sum_{i=1}^{n}\log x_{ij},$$

which is to be maximized with respect to $\mu_j^{(k+1)}, \sigma_j^{(k+1)}$ and $\delta_j^{(k+1)}$. In order to make the sample variance of $\phi_j^{(k+1)}(x_{ij})$ $\left(i = 1, ..., n\right)$ equal to one, we apply the Lagrange multiplier method and minimize

$$-Q + \lambda\left(\frac{\tilde{\phi}_j^{(k+1)}\left(\tilde{x}_j\right)'\tilde{\phi}_j^{(k+1)}\left(\tilde{x}_j\right)}{n-1} - 1\right)^2$$

with respect to $\mu_j^{(k+1)}, \sigma_j^{(k+1)},$ $\delta_j^{(k+1)}$ and $\lambda.$

3.1. Example

For a brief explanation of the above methods we shall examine a data set listed in the journal of Japan Automobile Federation (1993). The data was collected to investigate the coverage of cars and the other 6 related indices over 29 countries. A part of it is given in Table 1(called JAF data briefly below), where we have taken four indices from the original data: the number of persons per car (Per/Car, in brief), the number of persons per doctor (Per/Doc), GNP ($) per person (GNP) and consumer price index (CPI).

*Table*1

Apparently distributions of three variables Per/Car, Per/Doc, GNP and CPI are anti J-shaped and far from normality. Descriptive statistics of Per/Car, Per/Doc, GNP and CPI are given in Table 2. We also apply our power normal model (4) to each variable and give values of estimates $\hat{\delta}$, $\hat{\sigma}$ and $\hat{\mu}$ in Table 2.

*Table*2

Using estimates $\hat{\delta}$, $\hat{\sigma}$ and $\hat{\mu}$ in Table 2 as initial values, we have applied our algorithm to improve joint normality of JAF data. To evaluate joint normality of data sets, we have taken up four measures of multivariate skewness and kurtosis. Two ones are measures $b_{1,p}$ and $tr(S_2)$ for multivariate skewness, and the other two are measures $b_{2,p}$ and $tr(K_2)$ for multivariate kurtosis. Measures $b_{1,p}$ and $b_{2,p}$ have been defined by Mardia (1970). Alternatively, Isogai (1983a) has introduced measures $tr(S_2)$ and $tr(K_2)$ and examined their properties (also see Isogai (1983b, 1989)).

For a given random sample $x_i = \left(x_{i1}, x_{i2}, ..., x_{ip}\right)'(i=1,...,n)$ of size n, measures $b_{1,p}, b_{2,p}, tr(S_2)$ and $tr(K_2)$ are expressed in terms of 2nd, 3rd and 4th order multivariate k statistics, i.e. k_{ab}, k_{abc} and k_{abcd} $\left(a,b,c,d = 1,2,...,p\right)$ as follows:

$$b_{1,p} = k^{aa'} k^{bb'} k^{cc'} k_{abc} k_{a'b'c'},$$

$$b_{2,p} = k^{aa'} k^{bb'} k_{aa'bb'},$$

$$tr\left(S_2\right) = k^{aa'} k^{bb'} k^{cc'} k_{aa'b} k_{b'cc'},$$

$$tr\left(K_2\right) = k^{aa'} k^{bb'} k^{cc'} k^{dd'} k_{abcd} k_{a'b'c'd'},$$

where indices a, a', b, b', c, c', d and d' run from 1 through p, the symbol of summation $\sum$ is omitted for simplicity of expression, and a $p \times p$ matrix $\left(k^{ab}\right)$ is the inverse of $\left(k_{ab}\right)$ (for multivariate k statistics, see Kaplan (1952)).

Under the null hypothesis that data comes from a p-variate normal population, asymptotic distributions of measures $b_{1,p}, b_{2,p}, tr(S_2)$ and $tr(K_2)$ are summarized below:

(1) $\left\{\left(n-1\right)\left(n-2\right)/(6n)\right\} b_{1,p}$ $(= TestB_1$, say), $\left\{\left(n-1\right)\left(n-2\right)/(2\left(p+2\right)n)\right\} tr(S_2)$ $(= TestS_2$, say) and $\left\{\left(n-1\right)\left(n-2\right)\left(n-3\right)/(24n\left(n+1\right))\right\} tr(K_2)$ $(= TestK_2$, say) are asymptotically distributed as $\chi^2_{p(p+1)(p+2)/6}$, χ^2_p and $\chi^2_{p(p+1)(p+2)(p+3)/24}$ respectively, where χ^2_q means a χ^2 variable with q degrees of freedom.

(2) $b_{2,p} / \left\{ 8n(n+1)p(p+2) / \left((n-1)(n-2)(n-3) \right) \right\}^{1/2}$ ($= TestB_2$, say) is asymptotically distributed as a unit normal $\mathrm{N}(0,1)$.

Four measures $b_{1,p}, b_{2,p}, tr(S_2)$ and $tr(K_2)$ have different sensitivities to departures from joint normality. Roughly speaking, $b_{1,p}$ has a good sensitivity to a skewed distribution with a narrow tail. On the other hand, $tr(S_2)$ is well sensitive to a skewed distribution with a spread tail. As for measures of kurtosis, $b_{2,p}$ has a good power to detect a distribution with a spread tail in all the directions. On the contrary, $tr(K_2)$ detects well the departure from joint normality when marginal normality is assured.

In Table 3 we give values of measures $b_{1,p}, b_{2,p}, tr(S_2)$ and $tr(K_2)$ and their test statistics for 4 cases:(i) Original 4-dimensional case; (ii) Log transformation case, where each variable is transformed by the natural logarithm; (iii) Marginal power transformation case, where each variable is transformed to normality by our new power normal model (4); (iv) Joint power transformation case, where we apply our algorithm to attain joint normality of JAF data.

*Table*3

From Table 3 we know that the original JAF data has extreme non-normality as we have expected. Log transformation,which is a usual technique to transform an anti-J shaped distribution to a symmetric one, of JAF data improves non-normality a little bit. Marginal and joint power transformations have good performance to attain joint normality of JAF data. A main difference between both transformations appears in the value of $tr(K_2)$. The algorithm of the joint power transformation works well, especially to decrease non-normality in the meaning of $tr(K_2)$.

3.2. Notion of a principal surface

In Table 4 we give correlations and principal components of original and jointly power transformed cases of JAF data. Large differences appear in the values of correlations and proportions of eigenvalues between both cases. Decreasing non-normality of the original data seems to bring us a high correlation structure of the transformed data. Thus we can find a useful regression structure among variables of the transformed data, even though we could not find a meaningful regression structure at a first stage of the analysis

*Table*4

of the original data . For example, we take up the variable Per/Car as a dependent variable and the other variables Per/Doc, GNP and CPI as independent variables, and perform regression analysis using the usual least squares method. Then, using the original data, the result is 18% in R^2. On the other hand, using the transformed data to carry out the same analysis, we have 87% proportion in R^2.

We shall pick up some X_j of the random vector $X = \left(X_1, \ldots, X_p \right)'$ as a dependent variable and denote it by Y and the other variables by $Z_j \left(j = 1, \ldots, q \right)$ as independent variables. Then the regression structure of the transformed case is expressed as

$$\phi(Y) = \beta_1\phi_1(Z_1) + \beta_2\phi_2(Z_2) + \cdots + \beta_q\phi_q(Z_q),$$

where $\phi(Y)$ and $\phi_j(Z_j)(j = 1,...,q)$ are given by (10) and β_j 's are regression parameters. Unlike Box & Cox power normal family, our power normal family in the above model preserves the correlation structure of all estimates invariant under scale transformations of the all variables $Y, Z_1,..., Z_q$.

Regression analysis of the transformed variables gives us a linear relationship within them. We have another linear relationship among the transformed variables. The existence of a very small minimum eigenvalue in the correlation matrix of the transformed data indicates multi-collinearity in the transformed variables, which leads to the following approximate relationship:

$$\alpha_1\phi_1(X_1) + \alpha_2\phi_2(X_2) + \cdots + \alpha_p\phi_p(X_p) \approx 0,$$

where $\alpha = (\alpha_1,...,\alpha_p)'$ is the eigenvector corresponding to the minimum eigenvalue. Reexpressing the above relationship, we have

$$\frac{\alpha_1}{\delta_1}\left(\frac{X_1}{\eta_1}\right)^{\delta_1/\sigma_1} + \frac{\alpha_2}{\delta_2}\left(\frac{X_2}{\eta_2}\right)^{\delta_2/\sigma_2} + \cdots + \frac{\alpha_p}{\delta_p}\left(\frac{X_p}{\eta_p}\right)^{\delta_p/\sigma_p}$$

$$= \frac{\alpha_1}{\delta_1}\exp(-\delta_1^2) + \frac{\alpha_2}{\delta_2}\exp(-\delta_2^2) + \cdots + \frac{\alpha_p}{\delta_p}\exp(-\delta_p^2),$$

This gives us a typical hypersurface in the p-dimensional Euclidean space, on which our random vector $X = (X_1,..., X_p)'$ is distributed. Thus we may call it a principal surface.

From the result of the power transformed data in Table 4, we have an example of a principal surface such as

$$0.7592\phi_1(Per/Car) - 0.1307\phi_2(Per/Doc) + 0.6374\phi_3(GNP) - 0.0149\phi_4(CPI) \approx 0.$$

Disregarding the term $\phi_4(CPI)$ because of its small coefficient, we give graphics of the scatter plot of JAF data and two aspects of the surface of $0.7592\phi_1(Per/Car) - 0.1307\phi_2(Per/Doc) + 0.6374\phi_3(GNP) = 0$ in Figures 1(a) , 1(b) and 1(C). From Figure 1(a) we see that points of JAF data are scattered along three axes, and Figures 1(b) and 1(c) indicate that data points stick to the surface.

*Figure*1(*a*)

*Figure*1(*b*)

*Figure*1(*c*)

4. References

Box, G.E.P. & Cox, D.R. (1964) An analysis of transformations, *Journal of Royal Statistical Society*, B, 26, pp.211-252.

Isogai, T. (1983a) On measures of multivariate skewness and kurtosis, *Mathematica Japonica*, 28, pp.251-261.

Isogai, T. (1983b) Monte Carlo study on some measures for evaluating multinormality,

Reports of Statistical Application Research, JUSE, 30, pp.1-10.
Isogai, T. (1989) On using influence functions for testing multivariate normality, *Annals of The Institute of Statistical Mathematics*, 41, pp.169-186.
Isogai, T. (1999) Power transformation of the F distribution and a power normal family, *Journal of Applied Statistics*, 26, pp.355-371.
Isogai, T. (2001) Applications of power transformation formulas of the F distribution and a power normal family, *Journal of the Japanese Society for Quality Control*, 31, pp.89-104, (in Japanese).
Johnson, N. L. & Kotz, S. (1970) *Continuous univarite distributions*-1, Wiley-Interscience.
Kaplan, E. L. (1952) Tensor notation and the sampling cumulants of k statistics, *Biometrika*, 39, pp.319-323.
Mardia, K. V. (1970) Measures of multivariate skewness and kurtosis with applications, *Biometrika*, 57, pp.519-530.
Mudholkar, G. S. & Trivedi, M. C. (1980) A normal approximation for the distribution of the likelihood ratio statistic in multivariate analysis of variance, *Biometrika*, 67, pp.485-488.
Patnaik, P. B. (1949) The non-central χ^2 and F-distributions and their applications, *Biometrika*, 36, pp.202-232.
Prentice, R. L. (1975) Discrimination among some parametric models, *Biometrika*, 62, pp.607-614.
Sankaran, M. (1959) On the non-central χ^2 distribution, *Biometrika*, 46, pp.235-237.
Tiku, M. L. (1966) A note on approximating to the noncentral F-distribution, *Biometrika*, 53, pp.606-610.
Wishart, J. (1947) The cumulants of the z and of the logarithmic χ^2 and t distributions, *Biometrika*, 34, pp.170-178.

5. Appendix

First we remark that the cumulant generating function of $\log F_{2\phi_2}^{2\phi_1}$ for the first four cumulants is given by

$$\log E\left[\left(F_{2\phi_2}^{2\phi_1}\right)^t\right] = t\tilde{\kappa}_1 + \frac{t^2}{2}\tilde{\kappa}_2 + \frac{t^3}{3!}\tilde{\kappa}_3 + \frac{t^4}{4!}\tilde{\kappa}_4 + t^5 O_1(\Delta^4) \tag{19}$$

with

$$\tilde{\kappa}_1 = -\frac{1}{2}\left(\frac{1}{\phi_1} - \frac{1}{\phi_2}\right) - \frac{1}{12}\left(\frac{1}{\phi_1^2} - \frac{1}{\phi_2^2}\right) + O_2\left(\Delta^4\right),$$

$$\tilde{\kappa}_2 = \left(\frac{1}{\phi_1} + \frac{1}{\phi_2}\right) + \frac{1}{2}\left(\frac{1}{\phi_1^2} + \frac{1}{\phi_2^2}\right) + \frac{1}{6}\left(\frac{1}{\phi_1^3} + \frac{1}{\phi_2^3}\right) + O_3\left(\Delta^5\right),$$

$$\tilde{\kappa}_3 = -\left(\frac{1}{\phi_1^2} - \frac{1}{\phi_2^2}\right) - \left(\frac{1}{\phi_1^3} - \frac{1}{\phi_2^3}\right) + O_4\left(\Delta^4\right),$$

$$\tilde{\kappa}_4 = 2\left(\frac{1}{\phi_1^3} + \frac{1}{\phi_2^3}\right) + 3\left(\frac{1}{\phi_1^4} + \frac{1}{\phi_2^4}\right) + O_5\left(\Delta^5\right),$$

where we put $\Delta = \max\left(1/\phi_1, 1/\phi_2\right)$. For the derivation of the above cumulants, see Wishart (1947).

To express the cumulants of $\log F_{2\phi_2}^{2\phi_1}$ in terms of δ_2 , δ_1 and δ in (6), we note that the domain of δ_1 is

$$-\frac{1}{3} < \delta_1 < \frac{1}{3} \tag{20}$$

and ϕ_1 and ϕ_2 are written as

$$\begin{cases} \phi_1^{-1} = (1/2)\delta_2^2\left(1+3\delta_1\right), \\ \phi_2^{-1} = (1/2)\delta_2^2\left(1-3\delta_1\right). \end{cases} \tag{21}$$

Here, if we assume that

$$0 < \delta_2 < 1, \tag{22}$$

from (21) we have a little bit narrower restriction : $0 < 1/\phi_1 + 1/\phi_2 < 1$ than the original restriction : $\phi_1 > 1$ and $\phi_2 > 1$. In what follows, to obtain asymptotic expansions of cumulants in terms of δ_2, we assume that the condition (22) holds.

The property (20) means that $\left|\delta_1\right| = O(1)$. Thus, $\delta = \delta_1\delta_2 = O(\delta_2)$ and $\Delta = \max\left(1/\phi_1, 1/\phi_2\right) =$ $O(\delta_2^2)$. Substituting relations (21) into the cumulants of $\log F_{2\phi_2}^{2\phi_1}$, we can evaluate the first four cumulants $\tilde{\kappa}_1, \tilde{\kappa}_2, \tilde{\kappa}_3$ and $\tilde{\kappa}_4$ of $\log F_{2\phi_2}^{2\phi_1}$ with respect to the order of δ_2. The results are

$$\tilde{\kappa}_1 = -\frac{3\delta\delta_2}{2} - \frac{\delta\delta_2^3}{4} + O_2(\delta_2^8),$$

$$\tilde{\kappa}_2 = \delta_2^2 + \frac{9\delta^2\delta_2^2}{4} + \frac{\delta_2^4}{4} + O_3(\delta_2^6),$$

$$\tilde{\kappa}_3 = -3\delta\delta_2^3 - \frac{27\delta^3\delta_2^3}{4} - \frac{9\delta\delta_2^5}{4} + O_4(\delta_2^8),$$

$$\tilde{\kappa}_4 = \frac{27\delta^2\delta_2^4}{2} + \frac{\delta_2^6}{2} + O_5(\delta_2^8).$$

The first four cumulants K_1, K_2, K_3 and K_4 of $(1/\delta_2)\log F_{2\phi_2}^{2\phi_1}$ with respect to the order of δ_2 are given by relations $K_i = \tilde{\kappa}_i/(\delta_2)^i \quad (i = 1, 2, 3, 4)$. That is,

$$K_1 = -\frac{3\delta}{2} - \frac{\delta\delta_2^2}{4} + O_2(\delta_2^7),$$

$$K_2 = 1 + \frac{9\delta^2}{4} + \frac{\delta_2^2}{4} + O_3(\delta_2^4),$$

$$K_3 = -3\delta - \frac{27\delta^3}{4} - \frac{9\delta\delta_2^2}{4} + O_4(\delta_2^5),$$

$$K_4 = \frac{27\delta^2}{2} + \frac{\delta_2^2}{2} + O_5(\delta_2^4).$$

To obtain the first three approximate cumulants κ_1, κ_2 and κ_3 of $X = \eta\left(F_{2\phi_2}^{2\phi_1}\right)^{\gamma}$, we reexpress X into the form $X = \eta\left(F_{2\phi_2}^{2\phi_1}\right)^{\sigma/\delta_2}$. Using the formula (19), we have

$$E[X^r] = \eta^r \exp\left\{ r\sigma K_1 + \frac{(r\sigma)^2}{2} K_2 + \frac{(r\sigma)^3}{3!} K_3 + \frac{(r\sigma)^4}{4!} K_4 + \sigma^5 O_1(\delta_2^3) \right\}$$

for $r = 1, 2, 3$. Here, we add further the following assumption

$$\sigma = O(\delta_2)$$

to the original assumption $0 < \sigma < 1$. Then, neglecting terms higher order than δ_2^4, we have

$$E[X] = \eta \exp\left(-\frac{3\delta\sigma}{2} + \frac{\sigma^2}{2} - \frac{\sigma\delta\delta_2^2}{4} + \frac{9\delta^2\sigma^2}{8} + \frac{\delta_2^2\sigma^2}{8} - \frac{\delta\sigma^3}{2} \right),$$

$$E[X^2] = \eta^2 \exp\left(-3\delta\sigma + 2\sigma^2 - \frac{\sigma\delta\delta_2^2}{2} + \frac{9\delta^2\sigma^2}{2} + \frac{\delta_2^2\sigma^2}{2} - \frac{4\delta\sigma^3}{3} \right),$$

$$E[X^3] = \eta^3 \exp\left(-\frac{9\delta\sigma}{2} + \frac{9\sigma^2}{2} - \frac{3\sigma\delta\delta_2^2}{4} + \frac{81\delta^2\sigma^2}{8} + \frac{9\delta_2^2\sigma^2}{8} - \frac{27\delta\sigma^3}{2} \right).$$

From these expressions we can get the first three approximate cumulants κ_1, κ_2 and κ_3 of $X = \eta\left(F_{2\phi_2}^{2\phi_1}\right)^{\gamma}$ straightforwardly as follows:

$$\kappa_1 = \eta\left(1 - \frac{3\delta\sigma}{2} + \frac{\sigma^2}{2} + O_6(\delta_2^4) \right),$$

$$\kappa_2 = \eta^2\left(\sigma^2 + O_7(\delta_2^4) \right),$$

$$\kappa_3 = \eta^3\left(3(\sigma^4 - \delta\sigma^3) + O_8(\delta_2^6) \right).$$

Table 1. Survey on the coverage of cars over 29 countries in 1993.

Countries	Per/Car	Per/Doc	GNP ($)	CPI (%)
U.S.A.	1.8	404	22550	2.9
Canada	2.1	467	21500	1.3
Old West Germany	2.1	378	21475	4.4
Italy	2.1	234	19511	2.0
New Zealand	2.1	332	11875	1.0
Switzerland	2.3	329	35100	3.4
Australia	2.3	438	16310	0.9
France	2.4	399	21188	2.0
United Kingdom	2.5	611	17738	3.2
Japan	3.6	609	27326	1.7
Singapore	9.7	753	15030	2.3
Malaysia	9.9	2656	2965	4.7
Taiwan	11.2	961	10215	4.5
South Africa	12.0	1340	2970	15.3
Saudi Arabia	12.2	852	6600	2.2
Mexico	13.2	600	3200	11.9
Brazil	16.3	684	2200	1055.0
South Korea	20.8	1078	6635	4.5
Hong Kong	27.1	933	15750	9.4
Turkey	35.5	1189	1670	56.3
Egypt	46.6	616	730	19.8
Thailand	68.7	4361	1660	4.5
The Philippines	144.7	1016	835	8.2
Indonesia	149.6	7238	645	8.0
Nigeria	158.4	5997	330	13.0
Pakistan	159.2	1987	400	12.7
Kenya	185.6	7122	340	14.8
India	349.2	2075	310	12.8
China	989.2	724	360	6.4

The data was listed in the June issue of the monthly magazine "JAFMATE" , Journal of Japan Automobile Federation, 1993, p.38 (in Japanese).

Table 2(1). Marginal analyses of JAF data

Methods	Per/Car	Per/Doc	GNP	CPI
mean	84.2207	1599.414	9910.97	44.4517
standard deviation	192.1107	1994.019	10035.16	194.6416
skewness ($\sqrt{b_{1,1}}$)	4.0761	2.095	0.7772	5.3603
kurtosis ($b_{2,1}$)	18.6340	11.3395	0.1790	28.8110
New Power Normal Model				
$\hat{\delta}$	-0.2401	-0.3134	0.2433	-0.3040
$\hat{\sigma}$	1.6093	0.7217	1.4124	0.9886
$\hat{\mu}$	2.1967	6.4760	8.8662	1.2762

Table 2(2). Joint analysis of JAF data

Methods	Per/Car	Per/Doc	GNP	CPI
New Power Normal Model				
$\hat{\delta}$	-0.3152	-0.3628	0.2322	-0.3576
$\hat{\sigma}$	1.5305	0.6906	1.4536	0.9439
$\hat{\mu}$	2.0164	6.4218	8.8459	1.1955

Table 3. Tests for joint normality of JAF data in four transformation cases.

Test statistics		Original	Log	Marginal	Joint
Skewness	$b_{1,p}$	55.7154	16.2727	6.6429	5.6444
	$TestB_1$	<u>242.0738</u>	<u>70.7020</u>	28.8623	24.5240
	$tr(S_2)$	43.8022	8.7313	1.9651	1.5121
	$TestS_2$	<u>95.1564</u>	<u>18.9681</u>	4.2690	3.2848
Kurtosis	$b_{2,p}$	43.1835	11.4741	2.0969	0.5804
	$TestB_2$	<u>14.8134</u>	<u>3.9360</u>	0.7193	0.1991
	$tr(K_2)$	1.39e+03	200.096	52.8282	41.5091
	$TestK_2$	<u>1.31e+03</u>	<u>188.367</u>	49.7314	39.0758

Underlines indicate that values of test statistics are 5% significant.

$\chi^2_{20}(0.05)=31.4$, $\chi^2_{4}(0.05)=9.45$, $\chi^2_{35}(0.05)=49.80$;

$\chi^2_{20}(0.01)=37.6$, $\chi^2_{4}(0.01)=13.28$, $\chi^2_{35}(0.01)=57.34$

Table 4. Correlations and principal components of JAF data

(a) Original case

Correlations

Variables	Per/Car	Per/Doc	GNP	CPI
Per/Car	1	0.1744	-0.3998	-0.0642
Per/Doc	0.1744	1	-0.5071	-0.0795
GNP	-0.3998	-0.5071	1	-0.1722
CPI	-0.0642	-0.0795	-0.1722	1

Principal Components

Eigenvalues	1.7390	1.0604	0.8303	0.3704
Percent	43.5	26.5	20.8	9.2
Eigenvectors				
Per/Car	-0.4878	-0.1825	-0.7816	0.3432
Per/Doc	-0.5649	-0.1561	0.6187	0.5231
GNP	0.6635	-0.1934	-0.0524	0.7208
CPI	-0.0514	0.9513	-0.0591	0.2982

(b) Joint power transformation case

Correlations

Variables	$\phi_1(Per/Car)$	$\phi_2(Per/Doc)$	$\phi_3(GNP)$	$\phi_4(CPI)$
$\phi_1(Per/Car)$	1	0.8051	-0.9199	0.6920
$\phi_2(Per/Doc)$	0.8051	1	-0.7565	0.5535
$\phi_3(GNP)$	-0.9199	-0.7565	1	-0.6892
$\phi_4(CPI)$	0.6920	0.5535	-0.6892	1

Principal Components

Eigenvalues	3.2212	0.4627	0.2406	0.0756
Percent	80.5	11.6	6.0	1.9
Eigenvectors				
$\phi_1(Per/Car)$	0.5340	0.1445	0.3429	0.7592
$\phi_2(Per/Doc)$	0.4850	0.5249	-0.6872	-0.1307
$\phi_3(GNP)$	-0.5261	-0.0840	-0.5567	0.6374
$\phi_4(CPI)$	0.4504	-0.8346	-0.3168	-0.0149

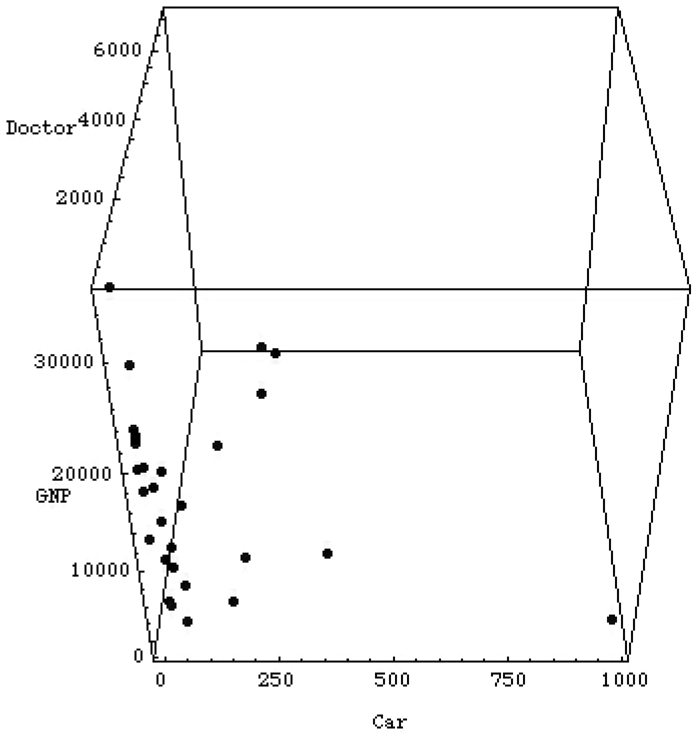

Figure 1(a). Scatter plot of JAF data.

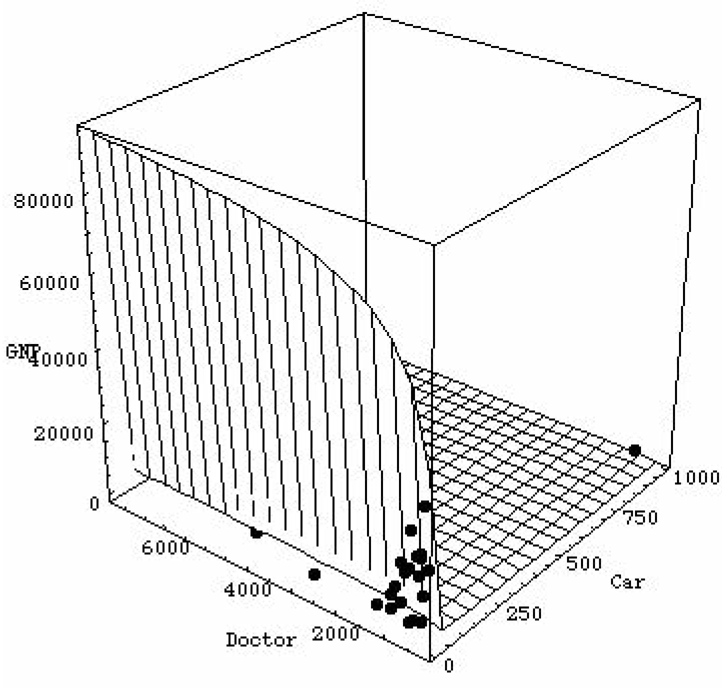

Figure 1(b). Principal surface of JAF data.

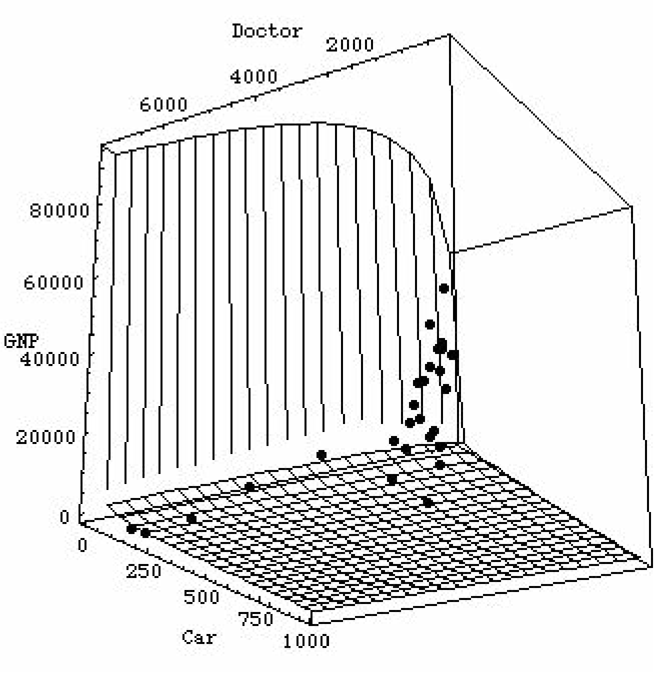

Figure 1(c). Principal surface of JAF data.