CHAPTER - II

DESCRIPTIVE STATISTICS FOR MULTIVARIATE MODELS

2.1 Introduction :

As discussed in the previous chapter, the descriptive statistics for univariate models have been investigated in many works. It is difficult to generalise to these concepts for multivariate cases. In this chapter we discuss such measures. In this chapter we discuss the following points

1) Some preliminary definitions of matrix, random vector and some their preliminary results,
2) Measure for location,
3) Measure for scatter,
4) Measure for skewness, and
5) Measure for kurtosis.

Generally multivariate statistics represents the expansion of more familiar univariate and bivariate statistics. Univariate and bivariate statistics are special cases or just simplifications of more general multivariate models. With the help of multivariate statistics we can analyze and study simultaneously more than one variable. To begin with we discuss some preliminary results which are used in defining the statistics for multivariate models :

2.2 Preliminaries :

In this section we give some elementary definitions related to matrices, random vector, and some results on variance co-variance matrix, which will be used in the subsequent sections.

2.2.1 Matrix :

'A' is said to a matrix of order p x n if pn elements arrange in p rows and n columns.

When n = 1 ' A ' is called as column vector. When p = 1 'A' is called as row vector. $A_{p \times p}$ = symmetric matrix.

2.2.2 Some special matrices :

1) If $a_{ij} = \begin{cases} 1 & \text{for } i = j. \\ 0 & \text{for } i \neq j. \end{cases}$

Then ' A ' is called as an identity matrix.

2) If $a_{ii} = \begin{cases} a_{ij} & \text{for } i = j. \\ 0 & \text{for } i \neq j. \end{cases}$

Then ' a ' is a diagonal matrix it is denoted as

$A = \text{diag}[a_{11}, a_{22}, \ldots, a_{pp}]$

3) If $A^2 = A$, then A is called as an idem potent matrix.

4) If $A' = A$, then A is called as symmetric matrix.

5) If $AA' = I$, then A is called as an orthogonal matrix.

6) If $a_{ij} = 1$, ∀ i,j = 1, 2, ...p then A is called as unit matrix.

7) $\text{Trace}(A) = \sum_{i=1}^{p} a_{ii}$

8) If $A_{p\times p}$ $P_{p\times p}$ are orthogonal matrix then ,

$$tr(PAP') = tr(P'AP) = tr(A)$$

9) If A is idempotent matrix then,

$$tr(A) = \rho(A)$$
$$= R(A)$$

10) If $A_{p\times p}$ is any non-singular matrix then,

$$tr(AA^{-1}) = p$$

Definition (2.2.1) : Random vector

Let Ω together with a σ-field be a measurable space. A real valued measurable function defined on the space Ω is a random vector.

Definition (2.2.2) : Distribution function of r.v.

Let $\underline{X}$ denote a p-variate random vector. Then its distribution function is defined as

$$F_{\underline{X}}(\underline{x}) = \Pr [X_1 \le x_1,\ X_2 \le x_2,\ \ldots,\ X_p \le x_p] \qquad \forall\ \underline{x} \in \mathbb{R}^k$$

Let $f(x_1, \ldots, x_p)$ be the joint density function of $\underline{X}$. Then the marginal distribution of one or more x_i's say $X_1, X_2, \ldots, X_k$ ($k < p$) is given by,

$$f_{(X_1,\ldots,X_k)}(x_1, x_2, \ldots, x_k)$$

$$= \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f(x_1, \ldots, x_p)\ dx_{k+1}, \ldots, dx_p \qquad (2.2.1)$$

The integration is over the appropriate range of $X_{k+1}, \ldots, X_p$.

If the joint distribution is a discrete one,we shall be dealing with the joint probability function and integration will be replaced by summation over the relevant variables,while finding the marginal distribution.

2.2.3 : Moments of $\underline{X}$:

a) We shall use E to denote the expectation operator. So that $E(\underline{X})$ will be the column vector with elements $E(X_i)$ $(i = 1,2,..p)$.

Thus,

$$E(\underline{X}) = [\ E(X_1),\ E(X_2),\ \ldots,\ E(X_p)\ ]' \qquad (2.2.2)$$

where,

$$E(X_i) = \int_{-\infty}^{\infty} x_i\ f(x_i)\ dx_i$$

b) Let $Z = A\ \underline{X}$ where A be any matrix of constants, of order mxp. Then,

$$E(Z) = E(A\ \underline{X})$$

$$= A\ E(\underline{X}).$$

c) The symbols V, COV will be used to denote variance and co-variance. Then the symmetric matrix

$$\begin{bmatrix} V(X_1) & COV(X_1,X_2) & \ldots & COV(X_1,X_p) \\ COV(X_2,X_1) & V(X_2) & \ldots & COV(X_2,X_p) \\ \vdots & \vdots & & \vdots \\ COV(X_p,X_1) & COV(X_p,X_2) & \ldots & V(X_p) \end{bmatrix} \qquad (2.2.3)$$

is known as the variance co variance matrix of the vector $\underline{X}$ and shall be denoted by $V(\underline{X})$. From the definitions of variance and co variances it is obvious that ,

$$V(\underline{X}) = E\ [\ \underline{X} - E(\underline{X})\ ]\ \ [\ \underline{X} - E(\underline{X})]'$$

This implies that,

$$\Sigma = V_{p\times p}$$

$$= E(\underline{X}\underline{X}') - E(\underline{X})E(\underline{X}')$$

Now we discuss some elementary results which are used in finding multivariate statistics.

__Lemma__ (2.2.1) : Σ is the positive definite matrix.

__Proof__ : Let $Z = \underline{\alpha}'\underline{X}$ where $\underline{\alpha}$ is any p-dimensional vector of constants. Note that Z is a random variable. Then,

$$V(\underline{Z}) = V(\underline{\alpha}'\underline{X})$$

$$= E\left[[\underline{\alpha}'\underline{X} - \underline{\alpha}'E(\underline{X})]\,[\underline{\alpha}'\underline{X} - \alpha'E(\underline{X})]'\right]$$

$$= E\left[(\alpha'[\underline{X} - E(\underline{X})])(\underline{\alpha}'[\underline{X} - E(\underline{X})])'\right]$$

$$= \alpha' E\left[[\underline{X} - E(\underline{X})][\underline{X} - E(\underline{X})])'\right]\alpha.$$

$$= \alpha'\Sigma\ \alpha.$$

$$\geq 0 \qquad\qquad \forall\ \underline{\alpha}.$$

This implies that, Σ is positive definite matrix.

Lemma (2.2.2) : Σ is symmetric matrix.

Proof : We know that,

$$\Sigma = V(\underline{X})$$

$$= E\left[\underline{X} - E(\underline{X})\right]\left[\underline{X} - E(\underline{X})\right]'$$

$$= E\left[\underline{X}\underline{X}' - E(\underline{X})\underline{X}' - XE(\underline{X}') + E(\underline{X})E(\underline{X}')\right]$$

Hence,

$$\Sigma = E(\underline{X}\underline{X}') - E(\underline{X})E(\underline{X}')$$

$$= V_{pxp}$$

$$= \begin{bmatrix} V(X_1) & COV(X_1, X_2) & \cdots & COV(X_1, X_p) \\ COV(X_2, X_1) & V(X_2) & \cdots & COV(X_2, X_p) \\ \vdots & \vdots & & \vdots \\ COV(X_p, X_1) & COV(X_p, X_2) & \cdots & V(X_p) \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix} \qquad (2.2.4)$$

This implies that, Σ is symmetric matrix. There are [p (p + 1)]/2 distinct parameters in Σ.

2.3 Measures of location :

In this section we discuss some measures of location introduced by Oja (1983). Before going to the definition of measures of location for multivariate models we introduce the following notations.

Notations and preliminaries :

Let, $X_1 = (x_{11},....x_{1k})'$,...., $X_{k+1} = (x_{k+1,1}.....x_{k+1,k})'$ be the points in $\mathbb{R}^k$. These points determine a k-dimensional simplex. The volume of this simplex is denoted by $\Delta(x_1,...x_{k+1})$ and given by

$$\Delta(x_1, \ ..., \ x_{k+1}) = \text{abs}\left[\frac{1}{k!}\begin{vmatrix} 1 & 1 & & 1 \\ x_{1,1} & x_{2,1} & \cdots & x_{k+1,1} \\ x_{1,k} & x_{2,k} & & x_{k+1,1} \end{vmatrix}\right] \qquad (2.3.1)$$

Note : i) If k= 1 then $\Delta(x_1,x_2)$ indicates the distance between the two points x_1,x_2 as follows:

$$\Delta(x_1, \ x_2) = \text{abs}\left[\frac{1}{1!}\begin{vmatrix} 1 & 1 \\ x_{11} & x_{21} \end{vmatrix}\right] \qquad (2.3.2)$$

$$= \text{abs}(x_{21} - x_{11})$$

ii) If $k = 2$, then $\Delta(x_1, x_2, x_3)$ gives the area of triangle in $\mathbb{R}^2$ which is given by,

$$\Delta(x_1, x_2, x_3) = \text{abs}\left[\frac{1}{2!} \begin{vmatrix} 1 & 1 & 1 \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{vmatrix} \right]$$

<u>Lemma</u> (2.3.1) : If $L:\mathbb{R}^k \longrightarrow \mathbb{R}^k$ is an affine transformation of the form $L(x) = Ax + \mu$ where A is a kxk matrix and μ is a k-vector then ,

$$\Delta(L(x_1), L(x_2), \ldots, L(x_{k+1})) = \text{abs}(\ |A|\ )\ \Delta(x_1, x_2, \ldots, x_{k+1}) \tag{2.3.3}$$

<u>Proof</u> : Let $(x_1, \ldots . x_{k+1})$ be the points in k-dimensional simplex. Then,

$$\Delta(x_1, x_2, \ldots, x_{k+1}) = \text{abs}\left[\int_{x \in S} dx_1\ dx_2\ \ldots\ dx_{k+1} \right] \tag{2.3.4}$$

where S is the simplex formed by the points $x_1, x_2, \ldots, x_{k+1}$ and $\Delta(x_1, x_2, \ldots, x_{k+1})$ indicates the k-dimensional volume of the simplex. Let us transform X to Y using transformation

$$Y = L(x)$$

$$= Ax + \mu \tag{2.3.5}$$

where A is kxk matrix and μ is a k-vector.

Differentiating equation (2.3.5) with respect to x, we have,

$$dy = A\,dx \tag{2.3.6}$$

Then $\Delta(y_1, y_2, \ldots, y_{k+1})$ is given by,

$$\Delta(L(x_1), L(x_2), \ldots, L(x_{k+1})) = \text{abs}\left[\int_{x \in S^*} dy_1\ dy_2, \ldots dy_{k+1}\right]$$

where S^* is the simplex formed by the points $(L(x_1), L(x_2), \ldots, L(x_{k+1}))$

Hence,

$$\Delta(L(x_1), L(x_2), \ldots L(x_{k+1})) = \text{abs}\left[|A| \int_{x \in S} dx_1\ dx_2, \ldots dx_{k+1}\right]$$

Thus,

$$\Delta(L(x_1), L(x_2), \ldots L(x_{k+1})) = \text{abs}(|A|)\ \Delta(x_1, x_2, \ldots x_{k+1})$$

<u>Definition</u> (2.3.1) : Let P be the class of probability distributions in $\mathbb{R}^k$, then $P \in \mathbb{P}$ is said to be stochastically smaller than $Q \in \mathbb{P}$ if,

$$\int f\, dp \leq \int f\ dQ. \tag{2.3.7}$$

for all real bounded co-ordinate wise increasing f. This is denoted by

$$P <_L Q \tag{2.3.8}$$

Lemma (2.3.2) : P $<_L$ Q is equivalent to the existence of two $\mathbb{R}^k$ valued random vectors on the same probability space with respect to distributions P and Q such that X ≤ Y (a.s.).

Proof : The property P $<_L$ Q in terms of the distribution functions of corresponding random variables X and Y can be explained by taking F as the indicator function given by ,

$$f(u) = [\ \underline{u} > \underline{t}\ ]$$

Now (2.3.8) implies,

$$\int_{\underline{t}}^{\infty} dp \leq \int_{\underline{t}}^{\infty} dQ$$

That is

$$S_X(t) \leq S_Y(t) \tag{2.3.9}$$

where $S_X(t)$ and $S_Y(t)$ indicate the survival function of an random variables X and Y having Probability distribution P and Q respectively. Hence if X is an random variable having distribution P and distribution function F & Y is an random variable having distribution Q and distribution function G, then, P $<_L$ Q implies F(t) ≤ G(t). Now by defining,

$Y = G^{-1}(F(x)) \geq G^{-1}(G(x)) = X$ we have,

$$Y \geq X \tag{2.3.10}$$

We can write,

$x = (x_1, \ldots x_k)' \leq y = (y_1 . y_2, \ldots\ldots y_k)'$

<u>Lemma</u> (2.3.3) : If $\phi : \mathbb{R}^k \longrightarrow \mathbb{R}^k$ is a function satisfying $\phi(x) \geq x$ $\forall$ $x \in \mathbb{R}^k$, then the distribution of X is stochastically smaller than the distribution of Y.

<u>Proof</u> : Consider,

$A = \{ x : \phi(x) \leq t \} \subseteq \{ x : x \leq t \} = B$

Hence

$P[\phi(x) \leq t] \leq P[x \leq t]$.

That is the distribution of x is stochastically smaller than the distribution of $\phi(x)$. This implies that X is stochastically smaller than $\phi(x)$. ■.

<u>Note</u> : It is natural to associate with the distributions in $\mathbb{P}$ in a consistent way with the stochastic ordering , an $\mathbb{R}^k$ valued measures of location. This can be done in the following way.

<u>Definition</u> (2.3.2) : <u>Measure of Location</u>

Suppose $\psi : \mathbb{R}^k \longrightarrow \mathbb{R}^k$ is a function such that

i) $P, Q \in \mathbb{P}$ $\quad P <_L Q$

This implies that, $\psi(P) \leq \psi(Q)$.

ii) For any affine transformation L : $\mathbb{R}^k \longrightarrow \mathbb{R}^k$ and $P \in \mathbb{P}$ such that $PL^{-1} \in \mathbb{P}$.

$\psi(PL^{-1}) = L\psi(P)$.

Then ψ is said to be a measure of location. If P is the distribution corresponding to X then PL^{-1} is the distribution corresponding to LX.

<u>Lemma</u> (2.3.4) : By using $\psi(X)$ interchangably with $\Psi(P)$ whenever X has the distribution P, the requirement (ii) of definition (2.2.2) can be stated as

$$\psi(L(X)) = L(\Psi(X)) \tag{2.3.11}$$

<u>Proof</u> : Let $y = L(X) = aX + b$ be an affine transformation. Let $Q = PL^{-1}$ be the distribution of Y and $Q \in \mathbb{P}$, the class of all distributions. Now for simplicity let us define

$\psi(P) = EX$, where $X \sim P$

Then,

$$\begin{aligned} \text{Mean } (Q) &= \text{Mean } (PL^{-1}) \\ &= E(Y) \\ &= E(aX + b) \end{aligned}$$

Thus,

$$\text{Mean }(Q) = AE(X) + b$$
$$= LE(X).$$

Hence the proof. ■.

<u>Remark</u> : If k = 1 then the property of function ψ defined in the definition (2.2.2) satisfies the property of a measure of location in the non-parametric models defined in Bickel and Lehamman (1975, 1976). We will discuss about location parameters for the non-parametric models in the next chapter.

<u>Definition</u> (2.3.3) : <u>Symmetric Distribution</u>

If an random variables X and (-X) have the same distribution, that is, f(x) = f(-x), then the random variable is said to have a symmetric distribution (about zero).

<u>Lemma</u> (2.3.5) : If the distribution is symmetric about $\mu \in \mathbb{R}^k$, That is, the random variables $(X-\mu)$ and $(\mu-X)$ have the same distributions, then necessarily, $\psi(x) = \mu$.

<u>Proof</u> : First we prove this lemma for the distribution of X which is symmetric about zero. Then using this fact we write it for μ as general case. Since X and (-X) have the same

distribution we have,

$$\psi(X) = \psi(-X)$$
$$= 0$$
$$= \psi(-IX)$$

where I is the identity matrix.

Hence, by using condition (ii) of definition (2.3.2) we get,

$$\psi(X) = -I\psi(X)$$
$$= -\psi(X).$$

This implies that,

$$\psi(X) = 0$$

Thus in general if X is symmetric about μ then $\psi(X) = \mu$. Below we describe the class of measures of location introduced by Oja (1983). Suppose the functions

$$\mu_\alpha : \mathbb{P} \longrightarrow \mathbb{R}^k \qquad \text{for } 0 < \alpha < \infty$$

are such that

$$E\left\{[\Delta(x_1, x_2, \dots, x_k, \mu_\alpha(P))]^\alpha\right\} = \inf_{\mu \in \mathbb{R}^k} E\left\{[\Delta(x_1, x_2, \dots, x_k, \mu(P))]^\alpha\right\} \qquad (2.3.12)$$

where Δ is defined in (2.3.1) and $x_1, x_2, \dots x_k$ is a random sample of size k from p. Then the functions μ_α are the measures of location.

Lemma (2.3.6) : The lemma (2.3.1) implies that for any affine transformation $L : \mathbb{R}^k \longrightarrow \mathbb{R}^k$ of the form $L(x) = Ax + \mu$ where A is a kxk matrix and μ is a k-vector.

$$\psi(PL^{-1}) = L\psi(p)$$

where $P \in \mathbb{P}$ such that $P\bar{L}^{-1} \in \mathbb{P}$

Proof : To prove this lemma we have to prove

$$\psi(Y) = a\ \psi(X)+b \tag{2.3.13}$$

Now consider

$$E\left\{[\Delta(y_1, y_2, \ldots, y_k, \mu_\alpha(Q)]^\alpha\right\}$$

$$= \inf_{\mu^* \in \mathbb{R}^k} E\left\{[\ \Delta(y_1, y_2, \ldots, y_k, \mu^*\ )]^\alpha\right\}$$

where, $\mu^* = L(\mu)$

$$= a\ \mu + b$$

Hence,

$$E\left\{\ [\Delta(y_1, y_2, \ldots, y_k, \mu_\alpha(Q)]^\alpha\right\}$$

$$= \text{abs}|A|^\alpha \inf_{\mu \in \mathbb{R}^k} E\left\{\ [\Delta(x_1, x_2, \ldots x_k, \mu(P))]^\alpha\right\}$$

Thus we have,

$$E\left\{ [\Delta(y_1, y_2, \dots, y_k, \mu_\alpha(Q)]^\alpha \right\}$$

$$= \text{abs}|A|^\alpha \inf_{\mu \in \mathbb{R}^k} E\left\{ [\Delta(x_1, x_2, \dots x_k, \mu(P))]^\alpha \right\}$$

$$= \text{abs}|A|^\alpha \; E\left\{ [\Delta(x_1, x_2, \dots x_k, \mu(p))]^\alpha \right\}$$

That is,

$\mu(Q) = a\mu(P) + b$ ∎.

<u>Note</u> : Thus functions μ_α $0 < \alpha < \infty$ are measures of location. If the value of $\mu_\alpha(P)$ $\alpha > 1$ exists ,it is unique.

i) If $k = 1$ and $\alpha = 1$ then,

$$E [\Delta(x_1, \mu_1(P))]^1 = E [| x_1 - \mu_1(P) |]$$

$$= \inf_{\mu \in \mathbb{R}^1} E| x_1 - \mu |$$

The value of above derivation is infimum if μ is equal to the median of distribution P.(by the property of median.)

Hence,

μ = median.

Therefore when we put k=1 and $\alpha = 1$ in (2.3.3) then we get the median of univariate distribution.

ii) Similarlly if k = 1 and $\alpha = 2$ then,

$$E\,[\Delta(x_1,\mu_2(P))]^2 = \inf_{\mu\in\mathbb{R}^1} E\,|(x_1 - \mu(P))|^2$$

The infimum is obtained when $\mu(P)$ is equal to the mean of distribution P

<u>Remark</u> : The sample version of $\mu_\alpha(P)$ is obtained by $\mu_\alpha(\hat{P})$,whenever $\hat{P}$ the estimate of P, is the empirical distribution based on the sample $X_1, X_2, \ldots, X_n$. Thus, if $X_1, X_2, \ldots X_n$ is an observed sample from P, natural estimates $\hat{\mu}_\alpha(P)$,of $\mu_\alpha(P)$ (for $0 < \alpha < \infty$) are given by,

$$\sum [\Delta(x_{i_1}, x_{i_2}, \ldots x_{i_k}, \hat{\mu}_\alpha]^\alpha_\mu$$

$$= \inf_{\mu\in\mathbb{R}^k} \sum [\Delta(x_{i_1}, x_{i_2}, \ldots x_{i_k}, \mu]^\alpha \qquad (2.3.14)$$

where the summation is over $1 \le i_1 \le \ldots \le i_k \le n$.

2.4 <u>Scatter</u> :

<u>Definition</u> (2.4.1) : A symmetric univariate distribution P (symmetric about μ) is said to more dispersed about μ than another symmetric distribution Q about υ if ,

$P(\mu-a,\mu+a) \leq Q(\nu-a,\nu+a). \quad \forall\ a > 0.$ (2.4.1)

For example let P corresponds to N(0,1) distribution and Q corresponds to N(0,2). Then it is clear that for 0 <a <1

$$P\left[(\mu-a),\ (\mu+a) \right] \leq Q\left[(\mu-a),\ (\mu+a) \right]$$

and for a > 1

$$Q\left[(\mu-a),\ (\mu+a) \right] \leq P\left[(\mu-a),\ (\mu+a) \right]$$

Hence P and Q are not comparable. In case of k-variate this statement can be stated as

$P(c+\mu) \quad \geq \quad Q(c+\nu)$ (2.4.2)

for all convex symmetrical sets $c \subset \mathbb{R}^k$.

<u>Definition</u> (2.4.2) : Suppose $\phi : \mathbb{R}^k \longrightarrow \mathbb{R}^k$ is a function such that

$$\Delta(\ \phi(x_1),\ \ldots,\ \phi(x_{k+1}))\ \geq\ \Delta(\ x_1,\ \ldots,\ x_{k+1})$$

$$\forall\ x_1,\ \ldots,\ x_{k+1} \in \mathbb{R}^k$$

and let P be the distribution of X and Q be the distribution of $\phi(x)$. Then we say that Q is more scattered than P. This is denoted by ,

$P \ <_s\ Q.$ (2.4.3)

Remark : i) If we take $\phi(x) = y = Ax + b$ where A is an orthogonal matrix then we have,

$$\Delta(\phi(x_1), \ldots \phi(x_{k+1})) = \Delta(x_1, \ldots x_{k+1}) \qquad (2.4.4)$$

since A is an orthogonal, matrix. Thus by orthogonal transformation spread does not change.

Lemma (2.4.1) : If P and Q are multinormal distributions and $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_k$ and $\gamma_1 \geq \gamma_2 \geq \ldots \geq \gamma_k$ are ordered eigen values of the respective covariance matrices then ,

$\lambda_i \geq \gamma_i \; \forall \; i = 1,2,\ldots k.$

This implies that $P <_s Q$. (2.4.5)

Proof : Let $X \sim N_k(\mu, \Lambda)$ where $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_k)$.

Let $Y = \Lambda^{-1}\nu X$ be the transformation of X. Then,

$$Y \sim N_k(\Lambda^{-1}\nu\mu, \nu)$$

and

$$\Delta(y_1, \ldots, y_{k+1}) = |\Lambda^{-1}\nu| \; \Delta(x_1, \ldots, x_{k+1})$$

$$= \frac{|\nu|}{|\Lambda|} \Delta(x_1, \ldots, x_{k+1})$$

That is,

$$\Delta(y_1, \ldots, y_{k+1}) = \frac{\Pi\nu_i}{\Pi\lambda_i}\Delta(x_1,\ldots x_{k+1}) > \Delta(x_1,\ldots x_{k+1}).$$

This implies that , $P <_s Q$

<u>Definition</u> (2.4.3) : <u>Measure of Scatter</u>

The function $\psi : \mathbb{P} \longrightarrow \mathbb{R}^k$ is a measure of scatter if

i) $P <_s Q \Rightarrow \psi(P) \leq \psi(Q) \quad \forall \quad P, Q \in \mathbb{P}$.

ii) For any affine transformation L: $\mathbb{R}^k \longrightarrow \mathbb{R}^k$ of the form

$L(x) = Ax + \mu$ and $p \in \mathbb{P}$ such that $PL^{-1} \in \mathbb{P}$

$$\psi(PL^{-1}) = \text{abs}(|A|)\ \psi(P).$$

<u>Note</u> : i) Condition (ii) in above definition can be stated as

$$\psi(Ax + \mu) = \text{abs}\ (|A|)\ \psi(x).$$

ii) If A is a singular matrix then, $|A| = 0$.
Hence $\psi(Ax+\mu) = 0$.

iii) For $k = 1$ the usual measure of scatter like standard deviation satisfies the above properties.

The functions $\sigma_\alpha : \mathbb{P} \longrightarrow \mathbb{R}^k$ $\alpha > 0$ defined as

$$\sigma_\alpha(P) = \sqrt[\alpha]{E\ [\ \Delta\ (x_1, x_2, \ldots x_k, \mu_\alpha(P)\ ]^\alpha} \tag{2.4.6}$$

where $x_1, x_2, \ldots, x_k$ is a random sample form P are also a measure of scatter.

Lemma (2.4.2) : The function defined above satisfies the condition (ii) of Definition (2.4.3).

Proof : Let $Y = Ax + \mu$ be an affine transformation where A is a nonsingular matrix and μ is a mean vector and let Q be the distribution of Y. Then by definition (2.4.3) we have,

$$\sigma_\alpha(Q) = \sqrt[\alpha]{E[\Delta(y_1, y_2, \ldots y_k, \upsilon_\alpha(Q)]^\alpha}$$

$$= \sqrt[\alpha]{E[\Delta(Ax_2 + \mu, \ldots Ax_k + \mu, A\mu_\alpha(p) + \mu)}$$

$$= \sqrt[\alpha]{abs|A|^\alpha E[\Delta(x_1 \ldots x_k, \mu_\alpha(p)]^\alpha}$$

$$= abs|A| \sqrt[\alpha]{E[\Delta(x_1, x_2, \ldots x_k, \mu_\alpha(P)]^\alpha}$$

$$= abs|A| \sigma_\alpha(p) \qquad (2.4.7)$$

Hence the condition (ii) satisfied. ■.

The function,

$$\sigma^*_\alpha(P) = \sqrt[\alpha]{E[\Delta(x_1, \ldots x_{k+1})]^\alpha} \qquad (2.4.8)$$

is also a measure of scatter, where $X_1, X_2, \ldots, X_{k+1}$ is a random sample from P.

Following are some simple measures of scatter;

1) For k = 1, $\alpha = 1$ we get,

$$\sigma(p) = E [\Delta (x_1, \mu_1(p))]$$

$$= E \mid X_1 - \mu \mid = \inf E \mid X_1 - \mu \mid$$

The value of $\sigma(p)$ is minimum if μ is median of P and in this case $\sigma(p)$ becomes mean deviation about median.

2) For k = 1 and $\alpha = 2$ we get, it is minimum if μ is the mean of P.

<u>Note</u> : 1) The sample version of this measure is given as follows:

If $X_1, X_2, \ldots, X_n$ is a random sample of size n form P then the natural estimates s_α of σ_α are given by,

$$\hat{\sigma}_\alpha = \sqrt[\alpha]{\binom{n}{k}^{-1} \Sigma \, [\Delta(x_{i_1}, \ldots x_{i_k}, \hat{\mu}_\alpha)]^\alpha} \qquad (2.4.9)$$

where the sum is over $1 \le i_1 \le \ldots \le i_k \le n$.

Similarlly the estimate for $\sigma^*_\alpha(P)$ is given by,

$$\hat{\sigma}^{*}_{\alpha} = \sqrt[\alpha]{\binom{n}{k+1}^{-1} \Sigma \left[\Delta(x_{i_1}, \ldots x_{i_{k+1}}) \right]^{\alpha}} \qquad (2.4.10)$$

where the sum is over $1 \le i_1 \le \ldots \le i_k \le n$.

2) A smallest convex set containing set S is called as convex hull. The convex hull generated by S is the minimal convex set, which contains the set S.Convex hull can be interpreted as a scatter set. The area or volume of this convex hull R ,which is generalized range is a measure of scatter.

3) If S is a subset of lesser dimensional space ($S \subset \mathbb{R}^{k-1}$) then the value of convex hull of S is zero.

4) With the help of measure of scatter we can compare the two distributions with each other,for example if we comparing normal with Caushy using variances, then we can not compare them because for Caushy distribution variance does not exist,but if we take the measure of scatter we can compare both the distributions.

For example let X_1, X_2, ..., X_n be a k-valued sample of size n, then for k = 1 the closed interval $[x_{(1)}, x_{(n)}]$ and its length $(x_{(n)} - x_{(1)})$ is a measure of scatter.

2.5 : Measure for skewness :

In this section we discuss a measures of multivariate skewness introduced by K. Mardia (1970).

Definition (2.5.1) : Cumulants

The cumulants $k_1, k_2, \ldots . k_r$ are defined by identity in t.

$$\exp\left\{ k_1 t + \frac{k_2 t^2}{2!} + \ldots + \frac{k_r t^r}{r!} + \ldots \right\}$$

$$= 1 + \mu_1' t + \frac{\mu_2' t^2}{2!} + \ldots + \frac{\mu_r' t^r}{r!} + \ldots .$$

It is some time convenient to write the same equation with (it) for t. Thus,

$$\exp\left\{ k_1 (i) t + \frac{k_2 (it)^2}{2!} + \ldots + \frac{k_r (it)^r}{r!} + \ldots \right\}$$

$$= 1 + \mu_1' it + \frac{\mu_2' (it)^2}{2!} + \ldots + \frac{\mu_r' (it)^r}{r!} + \ldots$$

$$= \int_{-\infty}^{\infty} \exp\{i\, tx\} dF$$

$$= \phi(t)$$

where μ_r' is the coefficient of $\frac{(it)^r}{r!}$ in $\phi(t)$ characteristic function k_r is the coefficient of $\frac{(it)^r}{r!}$ in log[$\phi(t)$] if the expansion is in the power series. [Ref. Kendal and Stuart, 1948].

<u>Lemma</u> (2.5.1) : $\text{corr}(\bar{X}, S^2) \approx \left\{ \beta_1/(2+\gamma_2) \right\}^{1/2}$ (2.5.1)

<u>Proof</u> : Let X_1, X_2, ..., X_n be a random sample of size n from a non-normal population with mean μ and finite variance . Let $\bar{X}$ and S^2 be the sample mean and variances. Now by definition of correlation coefficent we have,

$$\text{corr}(\bar{X}, S^2) = \frac{\text{cov}(\bar{X}, S^2)}{\sqrt{V(\bar{x})}\ \sqrt{V(S^2)}}$$

Now for large n,

$$V(\bar{X}) = \frac{\mu_2}{n},$$

$$V(S^2) \approx \frac{\mu_4 - \mu_2^2}{n} \quad \text{and}$$

$$\text{cov}(\bar{X}, S^2) \approx \frac{\mu_3}{n}$$

where μ_2, μ_3 and μ_4 are central moments. [Ref. Rhohatgi, 1976, Page 304]. Therefore,

$$\mathrm{corr}(\bar{X}, S^2) = \frac{\mu_3}{\sqrt{\mu_2(\mu_4 - \mu_2^{\ 2})}}$$

$$= \left\{\mu_3^{\ 2}/\mu_2^{\ 3}(\mu_4/\mu_2^{\ 2}-1)\right\}^{1/2}$$

$$= \left\{\beta_1/(\beta_2 - 3 + 3 - 1)\right\}^{1/2}$$

Hence,

$$\mathrm{corr}(\bar{X}, S^2) = \left\{\beta_1/(2+\gamma_2)\right\}^{1/2}$$

Thus correlation between $\bar{X}$ and S^2 itself is a measure of skewness. Mardia uses this concept to define measure of skewness for multivariate distribution. When developing the measure he has considered the canonical correlations between $\bar{X}$ and S^2.

Note : If the non-normal population is symmetric then odd ordered central moments are zero. Therefore, corr($\bar{X}, S^2$) = 0. Hence $\bar{X}$ and S^2 are exactly uncorrelated and independent. So that normal theory will be hold for n large enough.

Further assume that μ_4 is negligible. Then the equation (2.5.1) reduces to,

$$\mathrm{corr}(\bar{X}, S^2) = \left\{\beta_1/2\right\}^{1/2} \qquad (2.5.2)$$

which can be regarded as measure of univariate skewness.

The above discussion suggests the following extension of β_1 to the multivariate case as follows. Before introducing the measure of multivariate skewness we consider a definition of cannocical correlations.

Definition (2.5.2) : Cannonical Variables

Let $\underline{X}$ and $\underline{Y}$ be two column vectors of p and q random variables respectively with $p \leq q$. Let variance co-variance matrix of all these (p + q) elements be

$$V\left[\frac{\underline{X}}{\underline{Y}}\right] = \left[\begin{array}{c|c} \Sigma_{11} & \Sigma_{12} \\ \hline \Sigma_{21} & \Sigma_{22} \end{array}\right]\begin{matrix} p \\ q \end{matrix} \qquad (2.5.3)$$

Let us consider the transformations, $\underline{U}=L\underline{X}$ and $\underline{V}=M\underline{Y}$ from $\underline{X}$, $\underline{Y}$ to $\underline{U}$, $\underline{V}$ such that

$$V\left[\frac{\underline{U}}{\underline{V}}\right] = \left[\begin{array}{c|c} I & p \\ \hline p & I_p \end{array}\right]\begin{matrix} p \\ q \end{matrix} \qquad (2.5.4)$$

where,

$$p = [\text{diag}\,(\rho_1, \rho_2, \ldots, \rho_r, 0, 0, \ldots, 0) \mid 0_{q-p}]$$

r being the rank of Σ_{12}. Σ_{11} and Σ_{22} are non-singular. The $\underline{U}$ are linear combinations of $x_1, x_2, \ldots, x_p$; the elements of $\underline{X}$ and $\underline{V}$

are combinations of y_1, y_2, ..., y_q; the elements of $\underline{Y}$. $\underline{U}$ are called cannonical variables of X-space and $\underline{V}$ are called canonical variables of Y-space. The columns of L' and M' are called canonical vectors. If ρ_1^2, ρ_2^2, ...,ρ_p^2 are so arranged that $\rho_1^2 \geq \rho_2^2 \geq \ldots \geq \rho_p^2$ than U_1 is called first canonical variable of $\underline{X}$-space and V_1 is called as first canonical variable of Y-space. U_2 will be second and so on. The entire relationship between p-variables of $\underline{X}$ and q variables of $\underline{Y}$ is expressed only in terms of r parameters ρ_1^2, ρ_2^2, ..., ρ_r^2. Hence the name canonical variables.

Form (2.5.4) we can observe that the variances of each u_i and v_i is 1 and the co variances are correlations. Form (2.5.4) we find that,

$$\begin{aligned} Corr(u_i, v_j) &= \rho_i && \text{only when } i = j \\ &= 0 && \text{otherwise.} \end{aligned}$$

Also we observe that,

$$Corr(u_i, u_j) = 0 \qquad \text{for } i \neq j$$

and

$$Corr(v_i, v_j) = 0, \qquad \text{for } i \neq j$$

In other words, the first canonical variable of x-space is correlated only with first canonical variable of the Y-space, and so on.

Definition (2.5.3) : Measure for skewness

Let us denote the measure of multivariate skewness by $\beta_{1,p}$. This measure is obtained by considering the canonical correlations between $\bar{X}$ and S^2 under the following assumptions.

i) The second ordered moments of $\bar{X}$ and S^2 are taken to order n^{-1}.

ii) The cumulants of order higher than 3 of X are negligible.

Let,

$\underline{\mu}$ = the mean vector of population,

Σ = the covariance matrix of population and

$X_i' = (x_{1_i}, x_{2_i}, \ldots, x_{p_i})$ for i = 1, 2, ..., n be a random sample of size n from p-variate population with random vector

$X' = (x_1, x_2, \ldots, x_p)$.

$\bar{X} = (\bar{X}_1, \bar{X}_2, \ldots, \bar{X}_p)$ be sample mean vector.

$S = \{S_{ij}\}$ be the sample co variance matrix.

For simplification let us assume that $\mu = 0$ and Σ be a non-singular matrix. Now write the elements of $S = \{S_{ij}\}$ as vector $\underline{U}$ as follows;

$\underline{U} = (S_{11}, \ldots S_{pp}, S_{12}, \ldots . S_{2p}, \ldots . S_{p-1} S_p)'$ which has p + q elements where q = p(p-1)/2. Since S has p(p+1)/2 elements.

The first p components of U are diagonal elements of S while the remaining q components are the elements above the diagonal.

Let

$$\Lambda = \begin{bmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{bmatrix} \tag{2.5.5}$$

be the co variance matrix of $(\overline{X},U)$ where Λ_{11} is the co variance matrix of $\overline{X}$ and so on. Then the canonical correlations $\lambda_1, \lambda_2, \ldots, \lambda_p$ of $\overline{X}$ and U are found by finding the roots of determinantal equation,

$$|\Lambda_{11}^{-1} \Lambda_{12} \Lambda_{22}^{-1} \Lambda_{21} - \lambda^2 I| = 0, \tag{2.5.6}$$

where I is the identity matrix of p x p. The roots of this equation which are $\lambda_1, \lambda_2, \ldots, \lambda_p$ are canonical correlations of $\overline{X}$ and U [Ref. Kendal and Stuart, 1968 p.305].

Now any function of $\lambda_1^2, \lambda_2^2, \ldots, \lambda_p^2$ can be taken as a measure of skewness $\beta_{1,p}$, provided that this function satisfies an invariance property. Let us choose a specific function given by,

$$\beta_{1,p} = 2\sum_{i=1}^{p} \lambda_i^2. \tag{2.5.7}$$

Here multiplier 2 is taken because for p = 1 (2.5.7) reduces to (2.5.1).

Case I : Consider the case $\Sigma = I$.

We have, from (2.5.6) and (2.5.7)

$$|\Lambda_{11}^{-1}\Lambda_{12}\Lambda_{22}^{-1}\Lambda_{21} - \lambda^2 I| = 0, \text{ and } \beta_{1,p} = 2\sum_{i=1}^{p}\lambda_i^2.$$

Now since trace is equal to the sum of characteristic root we can write

$$\beta_{1,p} = 2\,\mathrm{tra}(\Lambda_{11}^{-1}\Lambda_{12}\Lambda_{22}^{-1}\Lambda_{21}) \qquad (2.5.8)$$

Since $\Lambda_{11} = I/n$ the above equation reduces to

$$\beta_{1,p} = 2\,n\,\mathrm{tra}(\Lambda_{11}^{-1}\Lambda_{21}\Lambda_{12}) \qquad (2.5.9)$$

Now from result (2.5.1) we have,

$$V(S_{11}) = \frac{\mu_4 - \mu_2^2}{n}$$

and by normal approximation it is equal to 2/n, and

$$V(S_{12}) = \left[\frac{1}{n}\sum X_{1i}X_{2i}\right] \qquad X_i \sim N(0, I) \qquad (2.5.10)$$

$$= \frac{1}{n^2} E\left[\sum X_{1i}X_{2i}\right]^2$$

$$= \frac{1}{n^2} E\left[\sum X_{1i}^2 X_{2i}^2 + 2\sum_{i\#j} X_{1i}X_{2i}X_{1j}X_{2j}\right]$$

Hence,

$$V(S_{12}) = \frac{1}{n^2}\left[\, n - 2\ .0 \,\right]$$

$$= \frac{1}{n} \qquad (2.5.11)$$

Hence approximation up to order 1/n we have,

Λ_{22} = diag(2, 2, ..., 2, 1, ...,1)/n with p 2's and q 1's.

Substituting Λ_{22} in $\beta_{1,p}$ we get,

$$\beta_{1,p} = n^2 \sum_{i,j,k=1}^{p} \text{cov} \left\{\bar{X}_i , S_{jk}\right\}^2 \qquad (2.5.12)$$

Under the assumptions (i), (ii) and μ = 0 we have,

$$\text{cov}\ (\bar{X}_i,\ S_{jk}) = E(X_i\ X_j\ X_k)/n \qquad \forall\ i,j,k.$$

So that,

$$\beta_{1,p} = \sum_{i,j}^{p} \left\{E(X_i\ X_j\ X_k)\right\}^2 \qquad (2.5.13)$$

To verify invariance property of this measure consider the following lemma. Befor we start lemma we define some functions.

Let $\phi(x)$ be a functional defined on random vector $\underline{X}$ with mean zero and co variance matrix Σ. Let Σ be positive definite matrix. Therefore there exist a nonsingular matrix U such that U

$\Sigma U' = I$. Define a random vector $\underline{Y}$ such that

$$X = UY$$

equivalently,

$$Y = U^{-1} X \qquad (2.5.14)$$

corresponding to ϕ as,

$$\phi^*(\underline{Y}) = \phi(U\underline{Y}) \qquad (2.5.15)$$

keeping in mind $U \Sigma U' = I$ the co variance matrix of $\underline{Y}$ is Σ.

The equation $U \Sigma U' = I$ does not define U uniquely for given Σ. So ϕ^* defined above may not extend ϕ uniquely. For this situation we have to discuss the following lemma which gives sufficient condition under which ϕ^* will be unique.

<u>Lemma</u> (2.5.2) : The function $\phi^*(\underline{Y})$ is uniquely defined for given Σ if $\phi(X)$ is invariant under orthogonal transformations. Furthermore $\phi^*(Y)$ will be then invariant under non-singular transformations.

<u>Proof</u> : Let V be another matrix satisfying $V \Sigma V' = I$.

For the first part of lemma it is sufficient to show that

$$\phi(UY) = \phi(VY) \qquad (2.5.16)$$

We have, $U \Sigma U' = I$ which implies that

$$\Sigma = U^{-1}(U^{-1})'$$

$= BB'$ (2.5.17)

where, $B = U^{-1}$.

Now,

$V \Sigma V' = I$

$= V^{-1}(V^{-1})'$

$= DD'$ (2.5.18)

where, $D = V^{-1}$.

Thus from equation (2.5.17) and (2.5.18) we have,

$BB' = DD'$

This happens only when either B and D are identical or D is multiple of any orthogonal matrix C. That is, B = DC. Putting these values of B and D we get ,

$V U^{-1} U = V V^{-1} C U$

This implies that,

$V = C U$ (2.5.19)

Under the assumption of lemma (2.5.2) we have,

$\phi(U Y) = \phi(X)$

$= \phi(C X)$.

Therefore from (2.5.14) we obtain,

$\phi(UY) = \phi(C UY)$

Also using equation (2.5.19) we get,

$$\phi(UY) = \phi(VY). \qquad (2.5.20)$$

Now we shall prove the invariance property. Let

$$Y_1 = A\,Y \qquad (2.5.21)$$

be a non-singular transformation and let Σ_1 denote the covariance matrix of Y_1. Using equation (2.5.14) and (2.5.21) we have,

$$(U\,\bar{A}^1)\,\Sigma_1\,(U\,\bar{A}^1)' = I \qquad (2.5.22)$$

Therefore from euqation (2.5.15) and (2.5.22) we get,

$$\phi^*(Y_1) = \phi(U\,\bar{A}^1\,Y_1)$$

Using equation (2.5.21) and (2.5.15), above equation reduces to,

$$\phi^*(Y_1) = \phi(UY)$$

$$= \phi^*(Y).$$

Hence the result. ■.

<u>Invariance property of $\beta_{1,p}$</u> :

Form (2.5.13) we have,

$$\beta_{1,p} = \sum_{i\,j\,k}^{p} \left\{E(X_i X_j X_k)\right\}^2$$

Consider the orthogonal transformation, $Y = CX$.

Let $C = \{ C_{ij} \}$ then, $X_{ij} = \sum_{r=1}^{p} C_{ri} Y_r$.

Substituting X_i in $\beta_{1,p}$ we get ,

$$\beta_{1,p} = \sum_{ijk}^{p} \left\{ E\ (C_{ri} Y_i,\quad C_{rj} Y_j,\quad C_{rk} Y_k) \right\}^2 \tag{2.5.23}$$

After simplification we get

$$\beta_{1,p} = \sum_{r,s,t} \sum_{r's't'} \left(\sum_{i=1}^{p} C_{ri} C_{r'i} \right) \left(\sum_{j=1}^{p} C_{sj} C_{s'j} \right) \left(\sum_{k=1}^{p} C_{tk} C_{t'k} \right)$$

$$.\quad E(Y_r Y_s Y_t) E(Y_{r'} Y_{s'} Y_{t'})$$

Since C is an orthogonal matrix the above equation becomes,

$$\beta_{1,p} = \sum_{rst}^{p} \left\{ E(Y_r Y_s Y_t) \right\}^2 . \tag{2.5.24}$$

Hence β_{1p} is invariant under an orthogonal transformation.

Case II : Consider the case for general Σ

We know that for positive defintite matrix Σ there exist an orthogonal matrix such that,

$C' \Sigma C = D$

where $D = \mathrm{diag}(d_1, d_2, \ldots, d_p)$ and $d_i > 0$; $i = 1,2,\ldots p$.

Consequently we may take, $U = D^{-1/2} C'$ in $U \Sigma U' = I$. Then from equation (2.5.14) we find,

$$X_i = \sum_{=1}^{p} C_{ri} \; Y_r \; / \; d_i^{1/2}$$

Let $\Sigma^{-1} = \left\{ \sigma^{rr'} \right\}$, then from $C' \; \Sigma \; C = D$ we get the result,

$$\sigma^{rr'} = \sum_{i=1}^{p} C_{ri} C_{r'i} / \; d_i \qquad (2.5.25)$$

Substituting for X_i in (2.5.13) and (2.5.25) we get

$$\beta_{1,p} = \sum_{ijk} \left\{ E \left[(C_{ri} \; Y_r / d_i^{1/2}) \; (C_{rj} \; Y_r / d_j^{1/2}) \; (C_{rk} \; Y_r / d_k^{1/2}) \right] \right\}^2$$

$$= \sum_{rst} \qquad \sum_{r's't'} \left\{ \sum_{i=1}^{p} C_{ri} C_{r'i} / d_i \right\}, \left\{ \sum_{j=1}^{p} C_{rj} C_{r'j} / d_j \right\}, \left\{ \sum_{k=1}^{p} C_{rk} C_{r'k} / d_k \right\},$$

$$\cdot \, E(Y_r Y_s Y_t) E(Y_{r'} Y_{s'} Y_{t'})$$

$$= \sum_{rst} \; \sum_{r's't'} \sigma^{rr'} \sigma^{ss'} \sigma^{tt'} \; E(Y_r Y_s Y_t) E(Y_{r'} Y_{s'} Y_{t'}) \qquad (2.5.26)$$

Therefore from random vector $\underline{X}$ with mean vector $\underline{\mu} = (\mu_1, \mu_2, \ldots, \mu_p)'$ and co variance matrix Σ we have ,

$$\beta_{1,p} = \sum_{rst} \sum_{r's't'} \sigma^{rr'} \sigma^{ss'} \sigma^{tt'} \; \mu_{111}(rst) \; \mu_{111}(r's't') \qquad (2.5.27)$$

where $\mu_{111}(rst) = E\left\{(X_r - \mu_r)(X_s - \mu_s)(X_t - \mu_t)\right\}$

Note : Form the lemma (2.5.2) we can show that this measure posses an invariance property under orthogonal transformation $X = A\ Y + b$.

2.6 Measure for Kurtosis :

To find the measure for multivariate kurtosis Mardia (1970) used the following result

Result : In the univariate case for one sample Pitman's test Box and Anderson (1955) have shown that the square of t-statistic has approximately F distribution with δ and $\delta(n-1)$ degrees of freedom, where,

$$\delta = 1 + \frac{\beta_2 - 3}{n} + o\ (1/n). \qquad (2.6.1)$$

The coefficient of 1/n in δ provides a measure of univariate kurtosis,β_2. An extension of it is given by Arnold (1964) which gives a sensible measure for multivariate kurtosis. Let $\beta_{2,p}$ denote the measure for multivariate kurtosis. Now let $X_1, X_2, \ldots, X_n$ be the random sample from p-variate population with random vector $\underline{X}$ and μ be the population mean vector, Σ be the variance-covariance matrix of population.

Definition (2.6.1) : Hotelling T^2 statistics

If X has p-variate non-singular normal distributions with mean 0 and variance covariance matrix Σ, that is, $\underline{X} \sim N_p(0,\Sigma)$ and if positive definite symmetric matrix D has Wishart distribution with f degrees of freedom, that is, $D \sim W_p(f,\Sigma)$ and X and D are independent then T^2 statistics which is known as Hotelling's T^2statistics is defined as,

$$T^2 = \underline{X}' (D/f)^{-1}\underline{X}$$

$$= f . (\underline{X}' D^{-1}\underline{X}). \qquad (2.6.2)$$

Theorem (2.6.1) : If $\underline{X} \sim N_p(0, \Sigma)$ distribution and $D \sim W_p(f, \Sigma)$ then $\frac{f - p + 1}{p} \frac{T^2}{f}$ has an F-distribution with p and f-p+1 degrees of freedom.

Proof : We have, $T^2 = f (\underline{X}' D^{-1} \underline{X})$

where $\underline{X} \sim N_p(0, \Sigma)$ and $D \sim W_p(f, \Sigma)$ $\underline{X}$and D are independent.

Consider the transformation, $\underline{Z} = C^{-1} \underline{X}$where C is lower triangular matrix such that $\Sigma = C C'$.

Then $Z \sim N_p(0,I_p)$, that is, components of Z_i are distributed as independent normal.

Therefore

$$\sum_{i=1}^{p} Z_i^2 = Z\,Z' \sim \chi_p^2 \qquad (2.6.3)$$

But $Z'Z = (\underline{X}'C^{-1})'C^{-1}\underline{X} = \underline{X}'(CC')^{-1}\underline{X} = \underline{X}'\Sigma^{-1}\underline{X}$.

Hence $\underline{X}'\Sigma^{-1}\underline{X} \sim \chi_p^2$. (2.6.4)

Now consider T^2/f

$$T^2/f = \underline{X}'D^{-1}\underline{X} \left[\frac{\underline{X}'\Sigma^{-1}\underline{X}}{\underline{X}'\Sigma^{-1}\underline{X}} \right] \text{ provided } \underline{X}'\Sigma^{-1}\underline{X} \neq 0 \qquad (2.6.5)$$

We know that, $[\underline{X}'\ \Sigma^{-1}\ \underline{X}]/[\underline{X}'\ D^{-1}\ \underline{X}]$ is distributed as χ^2_{f-p+1} and it is independent of $\underline{X}$, hence it is independent of $\underline{X}'\Sigma^{-1}\underline{X}$. Therefore,

$$\frac{T^2}{f} = \frac{\chi_p^2}{\chi^2_{f-p+1}}$$

Since χ_p^2 and χ^2_{f-p+1} are independently distributed,

$$\frac{T^2}{f} \sim \beta_2(p/2,\ (f-p+1)/2) \qquad (2.6.6)$$

The probability density function (p.d.f.) can be obtained by writing joint density function of χ_p^2 and χ^2_{f-p+1} and making transformation

$$\frac{T^2}{f} = \frac{\chi^2_p}{\chi^2_{f-p+1}}$$

Hence the p.d.f. of $\frac{T^2}{f}$ is

$$h\left(\frac{T^2}{f}\right) = 1/\beta_2(p/2,(f-p+1)/2)\,.\ \frac{(T^2/f)^{p/2-1}}{\left[1+T^2/f\right]^{(f+1)/2}} \qquad (2.6.7)$$

Also,

$$\frac{T^2}{f} = \frac{\left(\chi^2_p\ /\ p\right)p}{\left(\chi^2_{f-p+1/(f-p+1)/(f-p+1)}\right)(f-p+1)}$$

Thus,

$$\frac{f-p+1}{p}\,\frac{T^2}{f} \sim F\,(p,\ f-p+1) \qquad (2.6.8)$$

If f+1 = n then,

$$\frac{n-p}{p(n-1)}\,\frac{T^2}{f} \sim F(p,n-p)\,. \qquad (2.6.9)$$

Hence the proof. ■.

Mardia (1970) remarks that if we use the permutation moments of T^2 and the method of Box and Anderson then the distribution of

$(n-p)T^2/[p(n-1)]$ is approximately F distribution with δp and $\delta(n-p)$ degrees of freedom where δ is given by

$$\delta = 1 + \frac{E(b^*_{2,p}) - (p+2)}{n\left[1 - \left\{E(b^*_{2,p})/(n+2)\right\}\right]} \tag{2.6.10}$$

with

$$b^*_{2,p} = \left\{(n+2)/n^2 p\right\} \sum_{i=1}^{n} \left\{(X_i - \mu)' S^{-1} (X_i - \mu)\right\}^2 \tag{2.6.11}$$

Further S is the sample co-variance matrix about μ. We can write δ as

$$\delta = 1 + (1/n) \left[\left\{\beta_{2,p} - p(p+2)\right\}/p\right] + o(1/n) \tag{2.6.12}$$

where,

$$\beta_{2,p} = E\left\{ (X-\mu)'\Sigma^{-1}(X-\mu) \right\}^2 \tag{2.6.13}$$

If we compare (2.6.12) with (2.6.10) we get $\beta_{2,p}$ as measure of multivariate kurtosis.

Properties of $\beta_{2,p}$:

1) For $\mu = o$ and $\Sigma = I$ equation (2.6.13) reduces to,

$$\beta_{2,p} = E\left\{ (X'X)^2 \right\} \tag{2.6.14}$$

This measure is invariant under orthogonal transformation.

Hence by lemma (2.5.1),

$$\beta_{2,p} = E\left\{(X - \mu)'\Sigma^{-1}(X - \mu)\right\}^2$$

is invariant under nonsingular transformation.

Further,

$$E\left[a_0 + a_1\sum_{i=1}^{p} X_i + a_2\sum_{i=1}^{p} X_i^2\right]^2 \geq 0.$$

$$E\left[(a_0)^2 + \left(a_1\sum_{i=1}^{p} X_i\right)^2 + \left(a_2\sum_{i=1}^{p} X_i^2\right)^2 + 2\ a_0 a_1\sum_{i=1}^{p} X_i + 2\ a_0 a_2\sum_{i=1}^{p} X_i^2\right.$$

$$\left. + 2\ a_1 a_2\sum_{i=1}^{p} X_i \sum_{=1}^{p} X_i^2\right] \geq 0$$

$$E(a_0)^2 + a_1^2 E\left(\sum_{=1}^{p} X_i\right)^2 + a_2^2 E\left(\sum_{=1}^{p} X_i^2\right)^2 + 2\ a_0 a_1 E\left[\sum_{=1}^{p} X_i\right] + 2\ a_0 a_2 E\left[\sum_{=1}^{p} X_i^2\right]$$

$$+ 2\ a_1 a_2 E\left[\sum_{=1}^{p} X_i \sum_{=1}^{p} X_i^2\right] \geq 0$$

In particular by taking,

$a_0 = p$, $a_1 = -A/p$, $a_2 = -1$ where $A = E\left[\sum_{=1}^{p} X_i \sum_{=1}^{p} X_i^2\right]$ we get

$$p^2 + A^2/p + \beta_{2,p} - 2\ p^2 - 2\ (A/p)A \geq 0$$

That is $\beta_{2,p} \geq p^2 + A^2/p$. This implies $\beta_{2,p} \geq p^2$.

Remark : For p = 1 we get $\beta_{2,p} \geq (1 + \beta_1^2)$ which is well known inequality in univariate case. But if we try to find such relationship between $\beta_{1,p}$ and $\beta_{2,p}$ we found that these two measures are not related in such way.

Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from k-variate population. Then the measure for kurtosis is given by,

$$b_{2,p} = \frac{1}{n}\sum_{i=1}^{n}\left\{(X_i - \bar{X})' S^{-1} (X_i - \bar{X})\right\}^2$$

For example let X has bivariate normal distribution with mean μ and variance co variance matrix Σ. Then $\beta_{2,p}$ for this distribution is given in the following;

By definition of $\beta_{2,p}$ we have,

$$\beta_{2,p} = E\left\{ (X - \mu)' \Sigma^{-1} (X - \mu) \right\}^2$$

Since, X has $N_2(\mu, \Sigma)$ we get,

$$E\left\{ (X - \mu)' \Sigma^{-1} (X - \mu) \right\}^2 = E\left[\chi^2 \right]^2$$

$$= \int_0^\infty t^2 f(t)\, dt$$

$$= 2.$$

where $t = \chi^2$.

* * * * *