Chapter 2

DISTRIBUTIONS CONCERNING THE TEST PROCEDURES FOR MULTICOLLINEARITY OF MEANS.

2.1: Introduction and summary:-

In this chapter, we shall derive the distributions connected with the statistics, obtained in chapter I [(1.5)(iii)], connected with the tests for multicollinearity of means of the first kind. We shall define these statistics as:

$$(2.1.1) \qquad \text{(i) } \lambda_p = \text{maximum root of } (W_{1.2}S_{1.2}^{-1} - I_p),$$

$$\text{(ii) generalised } U = \operatorname{tr} W_{1.2}S_{1.2}^{-1} - p, \text{ and}$$

$$\text{(iii) (generalised } R)^{-1} = \operatorname{tr} W_{1.2}^{-1} S_{1.2} + r - p$$

where $W_{1.2} = W_{11} - W_{12}W_{22}^{-1}W'_{12}$ & $S_{1.2} = S_{11} - S_{12}S_{22}^{-1}S'_{12}$,

$$S = \begin{array}{cc} \begin{pmatrix} S_{11} & S_{12} \\ S'_{12} & S_{22} \end{pmatrix} & \begin{matrix} p \\ q \end{matrix} \\ \begin{matrix} p & \quad q \end{matrix} & \end{array}$$ is the S.P.M. due to error based on

m degrees of freedom (d.f.), $m \geqslant (p+q)$ and $W = S + Q = \begin{array}{cc} \begin{pmatrix} W_{11} & W_{12} \\ W'_{12} & W_{22} \end{pmatrix} & \begin{matrix} p \\ q \end{matrix} \\ \begin{matrix} p & \quad q \end{matrix} & \end{array}$

is the S.P.M. due to null hypothesis, based on r d.f. . The statistic due to likelihood criterion has been considered by e/ Bartlett (10,11), Hsu(33), Lawly (47) and Rao(68). Hence its consideration is omitted here. But for the null hypothesis the mutual independence of certain statistics in the sense given by the author (40) and K.S.Rao (72) are here established, and a numerical example is worked out to illustrate the technique of the computation of the statistics and the test procedure (27).

2.2: Distributions of the statistics defined in (2.1.1):

In order to include the different cases like (1.5.14) and (1.5.15), we assume without any loss of generality, in view of (1.6.1) and (1.6.5), that

$$(2.2.1)\quad \begin{array}{cc} \begin{pmatrix} Y \\ X \end{pmatrix} & \begin{array}{c} p \\ q \end{array} \\ r & \end{array} \quad \text{is} \quad MN\left[\begin{array}{c} p \\ q \end{array} \begin{array}{c} \begin{pmatrix} \mu \\ \nu \end{pmatrix} \\ r \end{array}, \Sigma = \begin{array}{cc} \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma'_{12} & \Sigma_{22} \end{pmatrix} & \begin{array}{c} p \\ q \end{array} \\ \begin{array}{cc} p & \quad q \end{array} & \end{array}\right]$$

and $S = \begin{array}{cc} \begin{pmatrix} S_{11} & S_{12} \\ S'_{12} & S_{22} \end{pmatrix} & \begin{array}{c} p \\ q \end{array} \\ \begin{array}{cc} p & \quad q \end{array} & \end{array}$ is Wishart $(m, p+q; \Sigma; S)$ and that they are independently distributed. We also note that $W = S + \begin{pmatrix} Y \\ X \end{pmatrix}(Y' \quad X')$.

We shall first prove a number of lemmas:

<u>Lemma 1</u>:- Given the distributions of (2.2.1), the joint distribution of $W = S + \begin{pmatrix} Y \\ X \end{pmatrix}(Y' \quad X')$, $Z = \tilde{T}_3'^{-1} X : q\text{x}r$, and $F : p\text{x}r = \tilde{T}_1'^{-1}(Y - T_2'\tilde{T}_3'^{-1}X)\,\tilde{D}^{-1}$ where $W = \tilde{T}'\tilde{T}$, $\tilde{D}'\tilde{D} = I - Z'Z$ and $\tilde{T} = \begin{array}{cc} \begin{pmatrix} \tilde{T}_1 & 0 \\ T_2 & \tilde{T}_3 \end{pmatrix} & \begin{array}{c} p \\ q \end{array} \\ \begin{array}{cc} p & \quad q \end{array} & \end{array}$ is given by

$$(2.2.2)\quad c\,|W|^{(m+r-p-q-1)/2}\,|I_r - Z'Z|^{(m-q-1)/2}\,|I_r - F'F|^{(m-p-q-1)/2}\,.$$

$$\exp\left\{-\tfrac{1}{2}\mathrm{tr}\Sigma^{-1}W + \mathrm{tr}\Sigma^{-1}(\mu' \quad \nu')'(\tilde{D}'F' \quad Z')\tilde{T}\right\},$$

where $c^{-1}=2^{(p+q)(r+m)/2}\,|\Sigma|^{(m+r)/2}\,\pi^{(p+q)(p+q+2r-1)/4}$.

$$\prod_{i=1}^{p+q}\Gamma\left(\frac{m-i+1}{2}\right) \text{ if } p+q\leq m.$$

Proof: The joint distribution of Y,X and S is

(2.2.3) $$c|S|^{(m-p-q-1)/2}\exp\left\{-\tfrac{1}{2}\mathrm{tr}\,\Sigma^{-1}S-\tfrac{1}{2}\mathrm{tr}\,\Sigma^{-1}\begin{pmatrix}Y-\mu\\X-\nu\end{pmatrix}\begin{pmatrix}Y-\mu\\X-\nu\end{pmatrix}'\right\}$$

where c is defined in (2.2.2) .

(2.2.4) If $W=S+\begin{pmatrix}Y\\X\end{pmatrix}(Y' \quad X')=\tilde{T}'\,\tilde{T}$ [See (A.1.4a)],

$$\tilde{T}=\begin{array}{cc}\begin{pmatrix}\tilde{T}_1 & 0\\ T_2 & \tilde{T}_3\end{pmatrix} & \begin{matrix}p\\q\end{matrix}\\ \begin{matrix}p & \quad q\end{matrix} & \end{array} \text{ and } \begin{pmatrix}Z_1\\Z\end{pmatrix}=\tilde{T}'^{-1}\begin{pmatrix}Y\\X\end{pmatrix}, \text{ then } Z=\tilde{T}_3'^{-1}X,$$

$$\tilde{T}_3'\tilde{T}_3=S_{22}+XX', \text{ taking } S=\begin{array}{cc}\begin{pmatrix}S_{11} & S_{12}\\ S_{12}' & S_{22}\end{pmatrix} & \begin{matrix}p\\q\end{matrix}\\ \begin{matrix}p & \quad q\end{matrix} & \end{array} \text{ and}$$

$$\tilde{T}^{-1}=\begin{array}{cc}\begin{pmatrix}\tilde{T}_1^{-1} & 0\\ -\tilde{T}_3^{-1}T_2\tilde{T}_1^{-1} & \tilde{T}_3^{-1}\end{pmatrix} & \begin{matrix}p\\q\end{matrix}\\ \begin{matrix}p & \qquad\quad q\end{matrix} & \end{array} \text{ by using (A.1.19b) .}$$

Hence, $I_r-Z'Z=I_r-X'(S_{22}+XX')^{-1}X=(I_r+X'S_{22}^{-1}X)^{-1}$ [See(A.1.18b)].

$$=\tilde{D}'\,\tilde{D} \quad \text{(say) [See(A.1.4a)] .}$$

Similarly $I_r-Z'Z-Z_1'Z_1=\left\{I_r+(Y' \quad X')S^{-1}\begin{pmatrix}Y\\X\end{pmatrix}\right\}^{-1}$.

(2.2.5) Also $|S|=\left|W-\begin{pmatrix}Y\\X\end{pmatrix}(Y' \quad X')\right|$

$$=|W|\,.\left|I_p-\tilde{T}'^{-1}\begin{pmatrix}Y\\X\end{pmatrix}(Y' \quad X')\tilde{T}^{-1}\right|.$$

i.e. $|S| = |W| \cdot \left| I_p - \begin{pmatrix} Z_1 \\ Z \end{pmatrix} (Z_1' \quad Z') \right|$

$= |W| \cdot |I_r - Z_1'Z_1 - Z'Z|$ using(A.1.14).

Let us transform S, Y and X to W, Z_1 and F as $W = S + \begin{pmatrix} Y \\ X \end{pmatrix} (Y' \quad X')$, $Z = \tilde{T}_3'^{-1}X$ and $F = \tilde{T}_1'^{-1}(Y - T_2'\tilde{T}_3'^{-1}X)\tilde{D}^{-1}$.

By using (A.2.12) and (A.2.1), the jacobian of the transformation is

(2.2.6) $J(Y,X,S;F,Z,W) = |\tilde{T}_3'|^r \, |\tilde{T}_1'|^r \, |\tilde{D}|^p = |W|^{r/2} |I - Z'Z|^{p/2}$,

and $|S| = |W| \cdot |I_r - Z'Z| \cdot |I_r - F'F|$ by (2.2.5) & (2.2.4).

By using (2.2.6) in (2.2.3), the joint distribution of F, Z, and **W** can be proved to be (2.2.2).

<u>Lemma 2:-</u> Under null hypothesis $(\mu = \Sigma_{12}\Sigma_{22}^{-1}\nu)$,

$F = \tilde{T}_1'^{-1}(Y - T_2'\tilde{T}_3'^{-1}X)\tilde{D}^{-1}$ and $W = S + \begin{pmatrix} Y \\ X \end{pmatrix} (Y' \quad X')$ are independently distributed and their distributions are given respectively as

(2.2.7) $c_1 |I_r - F'F|^{(m-p-q-1)/2}$ where

$$c_1^{-1} = \pi^{-pr/2} \prod_{i=1}^{r} \Gamma\left(\frac{m-q+r-i+1}{2}\right) \Big/ \Gamma\left(\frac{m-q-p+r-i+1}{2}\right)$$ and

(2.2.8) $c_2 |W|^{(m+r-p-q-1)/2} \underset{X:q\times t}{\int \cdot \int} |W_{22} - XX'|^{(m-q-1)/2} e^{\operatorname{tr} \Sigma_{22}^{-1} \nu} \, dX'$.

$\cdot \exp\{-\tfrac{1}{2} \operatorname{tr} \Sigma^{-1} W\}$ where $c_2^{-1} = c^{-1} c_1$.

<u>Proof:-</u> If $\Sigma = \tilde{B}'\tilde{B}$ [See (A.1.4a)] and $\tilde{B} = \begin{array}{cc} \begin{pmatrix} \tilde{B}_1 & 0 \\ B_2 & \tilde{B}_3 \end{pmatrix} & \begin{array}{c} p \\ q \end{array} \\ \begin{array}{cc} p & \quad q \end{array} & \end{array}$,

then by using (A.1.19b), $\tilde{B}^{-1}=\begin{pmatrix}\tilde{B}_1^{-1} & 0\\ -\tilde{B}_3^{-1}B_2\tilde{B}_1^{-1} & \tilde{B}_3^{-1}\end{pmatrix}$ and using (A.1.4b), $B_2'\tilde{B}_3'^{-1}=\Sigma_{12}\Sigma_{22}^{-1}$, $\tilde{B}_3'\tilde{B}_3=\Sigma_{22}$ and so $\Sigma^{-1}\begin{pmatrix}\mu\\ \nu\end{pmatrix}=$

$$\tilde{B}^{-1}\tilde{B}'^{-1}\begin{pmatrix}\mu\\ \nu\end{pmatrix}=\tilde{B}^{-1}\begin{pmatrix}\tilde{B}_1'^{-1}(\mu-\Sigma_{12}\Sigma_{22}^{-1}\nu)\\ \tilde{B}_3^{-1}\nu\end{pmatrix}.$$

Hence under null hypothesis ($\mu=\Sigma_{12}\Sigma_{22}^{-1}\nu$), we have

$$(2.2.9)\qquad \operatorname{tr}\Sigma^{-1}\begin{pmatrix}\mu\\ \nu\end{pmatrix}(\tilde{D}'F'\quad Z')\tilde{T}=\operatorname{tr}\Sigma_{22}^{-1}\nu Z'\tilde{T}_3 .$$

With the help of (2.2.9), (2.2.2) is equal to

$$(2.2.10)\qquad f_2(W,Z).f_1(F) \text{ where}$$

$f_1(F)$ is the same as defined in (2.2.7) and

$$(2.2.11)\quad f_2(W,Z)=c_2\,|I_r-ZZ'|^{(m-q-1)/2}|W|^{(m+r-p-q-1)/2}$$

$$\exp(-\tfrac{1}{2}\operatorname{tr}\Sigma^{-1}W+\operatorname{tr}\Sigma_{22}^{-1}\nu Z'\tilde{T}_3).$$

Hence (2.2.10) means, F and W are independently distributed. If we integrate over Z, then we can rewrite (2.2.11) as (2.2.8).

<u>Lemma 2</u>:- With the same notations as above,

(i) λ_p=maxi.root of $\{I_r-(I_r-F'F)^{-1}\}$ or $\{I_p-(I_p-FF')^{-1}\}$,

(ii) generalised $U=\operatorname{tr}FF'(I_p-FF')^{-1}=\operatorname{tr}F'F(I_r-F'F)^{-1}$

and (iii) (generalised $R)^{-1}=r-\operatorname{tr}FF'=r-\operatorname{tr}F'F$.

Proof: We have by the use of (2.2.4)

$$F'F = \tilde{D}'^{-1} Z_1' Z_1 \tilde{D}^{-1} .$$

Hence by (A.1.16b), roots of $F'F$ are the roots of $(\tilde{D}'\tilde{D})^{-1} Z_1' Z_1$ except for some zero roots. i.e.

(2.1.12) roots of $F'F$ are the roots of

$$\left\{ I_r - (I_r + X'S_{22}^{-1} X) \left[I_r + (Y' \quad X') S^{-1} \begin{pmatrix} Y \\ X \end{pmatrix} \right]^{-1} \right\}$$ except

for some zero roots, using (2.3.4).

But by (A.1.22a) roots of $\left\{ (I_r + X'S_{22}^{-1}X)^{-1} \left[I_r + (Y' \quad X') S^{-1} \begin{pmatrix} Y \\ X \end{pmatrix} \right] - I_r \right\}$ are the roots of $\left\{ (W_{11} - W_{12} W_{22}^{-1} W_{12}') (S_{11} - S_{12} S_{22}^{-1} S_{12}')^{-1} - I_p \right\}$ except for some zero roots. Therefore

(2.2.13) roots of $\left\{ (I_r - F'F)^{-1} - I_r \right\}$ are the roots of

$$\left\{ (W_{11} - W_{12} W_{22}^{-1} W_{12}') (S_{11} - S_{12} S_{22}^{-1} S_{12}')^{-1} - I_p \right\}$$ except for some zero roots .

By the results (2.2.13) and (2.2.14), we at once arrive at the proof of the results (i) and (ii) of lemma 3.

Also from (A.1.22b), roots of $\left\{ (I_r + X'S_{22}^{-1} X) \left[I_r + (Y' \quad X') S^{-1} (Y' \quad X')' \right]^{-1} - I_r \right\}$ are the roots of $\left\{ (W_{11} - W_{12} W_{22}^{-1} W_{12}')^{-1} (S_{11} - S_{12} S_{22}^{-1} S_{12}') - I_p \right\}$ except for some zero roots. Hence applying this to (2.1.12), we can at once prove the result (iii) of the above lemma .

Lemma 4: With the notations as above, under the null hypothesis ($\mu = \Sigma_{12} \Sigma_{22}^{-1} \nu$), the distribution of G=F'F is

$$(2.2.15) \qquad c_3 \, |G|^{(p-r-1)/2} \, |I_r - G|^{(m-q-p-1)/2} \quad \text{if } p \geq r$$

where $c_3 = c_1 \prod_{i=1}^{r} \left\{ \Gamma\left(\frac{p-i+1}{2}\right) \right\}^{-1} \pi^{\frac{1}{2}pr - \frac{1}{4}r(r-1)}$, while the distribution of H=FF':pxp is

$$(2.2.16) \qquad c_4 \, |H|^{(r-p-1)/2} \, |I_p - H|^{(m-q-p-1)/2} \quad \text{if } r \geq p$$

where $c_4 = c_1 \, \pi^{\frac{1}{2}pr - \frac{1}{4}p(p-1)} \prod_{i=1}^{p} \left\{ \Gamma\left(\frac{r-i+1}{2}\right) \right\}^{-1}$.

Proof: Under null hypothesis ($\mu = \Sigma_{12} \Sigma_{22}^{-1} \nu$), the distribution of F:pxr is given in (2.2.7). Then applying the result (A.3.1) in each separate case, we establish lemma 4. Now we can obtain the distributions of the statistics given in (2.1.1) under null hypothesis:-

(i) Let θ_p be the maximum root of G or H. Then $\theta_p = \lambda_p/(1+\lambda_p)$ and the distributions of θ_p is given by various persons and their tables are given for various values of m-q, p & r. The distribution of G or H is given in lemma 4 and we can apply the results of Roy (73,74,79), Pillai (61,63) and Nanda (52). For the tables of θ_p, we can refer to Bantegui (7,8), Pillai (8,64), Foster (20,21,22), Sen Pentro (82) and

Ventura (84).

<u>(ii) & (iii)</u>: By using the result (A.3.6) in (2.2.15) and (2.2.16), we have the approximate distributions of g_1, the generalised U, and $g_2 = r - (\text{generalised } R)^{-1}$ given for two different cases. Also we can use the tables of Pillai (64,65) for generalised U.

Case (a):- If $p \geq r$,

(2.2.16) $$\phi_1(g_1) = \frac{(g_1/r)^{(pr-2)/2}\,(1+g_1/r)^{-\{r(m-q-1)+2\}/2}}{rB\{\frac{1}{2}pr, \frac{1}{2}r(m-q-p-1)+1\}}$$

for $0 \leq g_1 \leq \infty$ and.

(2.2.17) $$\phi_2(g_2) = \frac{(g_2/r)^{(pr-2)/2}\,(1-g_2/r)^{\{r(m-p-q+r)-2\}/2}}{rB\{\frac{1}{2}pr, \frac{1}{2}r(m-p-q+r)\}} \quad \text{for } 0 \leq g_2 \leq r \,.$$

The approximation given in (2.2.16) is valid if $m \geq 57+p+q$ when $r=2$ and when r increases by one unit, $(m-p-q)$ must increase by 18. The approximation given in (2.2.17) is valid if $m \geq 64+q$ when $r=2$ and as r increases by one unit, $(m-q)$ must increase by 21.

Case (b):- If $p \leq r$,

(2.2.18) $$\phi_3(g_1) = \frac{(g_1/p)^{(pr-2)/2}\,(1+g_1/p)^{-\{p(m+r-p-q-1)+2\}/2}}{pB\{\frac{1}{2}pr, \frac{1}{2}p(m-p-q-1)+1\}} \quad \text{for } 0 \leq g_1 \leq \infty$$

and

$$(2.2.19) \qquad \emptyset_4(g_2) = \frac{(g_2/p)^{(pr-2)/2}(1-g_2/p)^{\{p(m-q)-2\}/2}}{p\,B\{\frac{1}{2}pr,\frac{1}{2}p(m-q)\}} \quad \text{for } 0 \leq g_2 \leq p.$$

The approximation given in (2.2.18) is valid if $m \geq 59+q$ when $p=2$ and when p increases by one unit, $(m-q)$ must increase by 19. The approximation given in (2.2.19) is valid if $m \geq 66+r+q$ when $p=2$ and if p increases by one unit, $(m-r-q)$ must increase by 22.

(2.2.20) In short, we obtain the following test procedures:-

(i) $\theta_p = \lambda_p/(1+\lambda_p)$ is distributed as $\theta_p\{\frac{1}{2}(p-r-1),\frac{1}{2}(m-q-p-1)\}$ defined by Pillai (8,64) or Foster (20,21,22) if $p \geq r$, while it is distributed as $\theta_p\{\frac{1}{2}(r-p-1),\frac{1}{2}(m-q-p-1)\}$ if $p \leq r$.

(ii) $g_1\{r(m-p-q-1)+2\}/pr^2$ is distributed as $F_{pr,r(m-q-p-1)+2}$ if $p \geq r$ while $g_1\{p(m-p-q-1)+2\}/p^2r$ is distributed as $F_{pr,p(m-q-p-1)+2}$ if $p \leq r$.

(iii) (r. generalised R-1)(m-p-q+r)/p is distributed as $F_{pr,r(m-p-q+r)}$ if $p \geq r$, while $\frac{(m-q)(r.\text{ generalised } R-1)}{r\{1-(r-p)\text{ generalised } R\}}$ is distributed as $F_{pr,p(m-q)}$ if $p \leq r$.

2.3: Power-Distribution of g_1 or g_2 or λ_p for the particular case when r=1:-

This problem has already been considered by Narain(54) but we present here a different approach.

Let $\boldsymbol{\Sigma}=\tilde{B}'\tilde{B}$ where $\tilde{B}=\begin{array}{cc} \begin{pmatrix} \tilde{B}_1 & 0 \\ B_2 & \tilde{B}_3 \end{pmatrix} & \begin{matrix} p \\ q \end{matrix} \\ \begin{matrix} p & \quad q \end{matrix} & \end{array}$ and let $\tilde{B}'^{-1}\begin{pmatrix} \underline{\boldsymbol{\mu}} \\ \underline{\boldsymbol{\nu}} \end{pmatrix}=\begin{pmatrix} \underline{\boldsymbol{\delta}}_1 \\ \underline{\boldsymbol{\delta}}_2 \end{pmatrix}$.

Hence we have $\underline{\boldsymbol{\delta}}_1'\,\underline{\boldsymbol{\delta}}_1=(\underline{\boldsymbol{\mu}}' \quad \underline{\boldsymbol{\nu}}')\boldsymbol{\Sigma}^{-1}\begin{pmatrix} \underline{\boldsymbol{\mu}} \\ \underline{\boldsymbol{\nu}} \end{pmatrix}-\underline{\boldsymbol{\nu}}'\,\boldsymbol{\Sigma}_{22}^{-1}\,\underline{\boldsymbol{\nu}}$=increased distance in the sense of Rao (68) and $\underline{\boldsymbol{\delta}}_2'\,\underline{\boldsymbol{\delta}}_2=\underline{\boldsymbol{\nu}}'\boldsymbol{\Sigma}_{22}^{-1}\,\underline{\boldsymbol{\nu}}$.

Let us transform W to $\tilde{G}$ by the relation:

$W=\tilde{T}'\tilde{T}$ and $\tilde{G}=\tilde{T}\tilde{B}^{-1}$. Then by using (A.2.5) and (A.2.3), the jacobian of the transformation is

$$J(W;\tilde{G})=J(W;\tilde{T})J(\tilde{T};\tilde{G})=2^{p+q}\prod_{i=1}^{p+q} t_{ii}^{i}\, b_{ii}^{p-i+1}$$

$$=2^{p+q}\prod_{i=1}^{p+q} g_{ii}^{i}\, |\boldsymbol{\Sigma}|^{(p+1)/2} .$$

Hence we can rewrite the equation (2.2.2) as

$$(2.3.1) \qquad c\,2^{p+q}\prod_{i=1}^{p+q} g_{ii}^{m-p-q+i}\,(1-\underline{z}'\underline{z})^{(m-q-1)/2}\,(1-\underline{f}'\underline{f})^{\frac{1}{2}(m-p-q-1)} .$$

$$\exp\left\{-\tfrac{1}{2}\mathrm{tr}\tilde{G}'\tilde{G}+(d\underline{f}' \quad \underline{z}')\,\tilde{G}\begin{pmatrix} \underline{\boldsymbol{\delta}}_1 \\ \underline{\boldsymbol{\delta}}_2 \end{pmatrix}\right\}$$

where $\underline{f}: p\times 1, \underline{z}: q\times 1, \underline{\boldsymbol{\delta}}_1: p\times 1, \underline{\boldsymbol{\delta}}_2: q\times 1$ and $\tilde{G}=\begin{array}{cc} \begin{pmatrix} \tilde{G}_1 & 0 \\ G_2 & \tilde{G}_3 \end{pmatrix} & \begin{matrix} p \\ q \end{matrix} \\ \begin{matrix} p & \quad q \end{matrix} & \end{array}$, $d=(1-\underline{z}'\underline{z})^{\frac{1}{2}}$

and $c^{-1} = \pi^{(p+q)(p+q+1)/4} \prod_{i=1}^{p+q} \Gamma\left(\frac{m-i+1}{2}\right) . 2^{(p+q)(m+1)/2}$.

Then performing the following integrations one by one, we shall get the required result.

(i) Integrate over $\underline{f}$ such that $u \leq \underline{f}'\underline{f} \leq u+du$,

(ii) integrate over the values of $\tilde{G}_1$ and G_2 ,

(iii) integrate over $\underline{z}$ such that $v \leq \underline{z}'\underline{z} \leq v+dv$

and (iv) integrate over $\tilde{G}_3$.

Note that $1-u=(1+\underline{x}'S_{22}^{-1}\underline{x})/\left\{1+(\underline{y}' \; \underline{x}')S^{-1}\begin{pmatrix}\underline{y}\\ \underline{x}\end{pmatrix}\right\}$ and $v=1-(1+\underline{x}'S_{22}^{-1}\underline{x})^{-1}$. The integrations utilised are (A.3.2) and (A.3.4) or (A.3.5).

The joint distribution of u and v is

$$(2.3.2) \quad \exp\left[-\left\{\underline{\delta}_2'\underline{\delta}_2+(1-v)\underline{\delta}_1'\underline{\delta}_1\right\}/2\right]\sum_{i,j=0}^{\infty}\frac{(\underline{\delta}_2'\underline{\delta}_2/2)^i}{i!} .$$

$$\frac{(\underline{\delta}_1'\underline{\delta}_1/2)^j}{j!} \; \frac{v^{\frac{1}{2}q+i-1}(1-v)^{\frac{1}{2}(m-q+1)+j-1}}{B\left\{\frac{1}{2}q+i,\frac{1}{2}(m-q+1)\right\}} ,$$

$$u^{\frac{1}{2}p+j-1}(1-u)^{\frac{1}{2}(m-q-p+1)-1} \Big/ B\left\{\frac{1}{2}p+j,\frac{1}{2}(m-q-p+1)\right\}.$$

Integrating over v, we get the power distribution of u and if $\underline{\delta}_1'\underline{\delta}_1$ =increased distance=0, we get the same distribution ~~of~~ as (2.2.17) after putting r=1.

2.4: Mutual Independence of Certain Statistics under null hypothesis:-

Lemma 5:- If $S=\begin{array}{cc} \left(\begin{array}{cc} S_{11} & S_{12} \\ S'_{12} & S_{22} \end{array}\right) & \begin{array}{c} p \\ q \end{array} \\ \begin{array}{cc} p & q \end{array} & \end{array}$, $\begin{array}{cc} \left(\begin{array}{c} Y_i \\ X_i \end{array}\right) & \begin{array}{c} p \\ q \end{array} \\ r_i & \end{array}$ are independently distributed as Wishart $(m,p+q;\Sigma;S)$ and

$$MN\left[\begin{array}{c} p \\ q \end{array}\begin{array}{c}\left(\begin{array}{c} \mu_i \\ \nu_i \end{array}\right) \\ r_i \end{array} ,\Sigma=\begin{array}{cc} \left(\begin{array}{cc} \Sigma_{11} & \Sigma_{12} \\ \Sigma'_{12} & \Sigma_{22} \end{array}\right) & \begin{array}{c} p \\ q \end{array} \\ \begin{array}{cc} p & q \end{array} & \end{array}\right], \quad i=1,2,3,\ldots,k, \text{ then}$$

$$S_k=S+\sum_{i=1}^{k}\left(\begin{array}{c} Y_i \\ X_i \end{array}\right)(Y'_i \quad X'_i) \quad \text{and} \quad F_j=\tilde{T}'^{-1}_{1.j}(Y_j-T'_{2.j}\tilde{T}'^{-1}_{3.j}X_j)\tilde{D}^{-1}_j$$

$$\text{where } S_j=S+\sum_{i=1}^{j}\left(\begin{array}{c} Y_i \\ X_i \end{array}\right)(Y'_i \quad X'_i)=\begin{array}{cc} \left(\begin{array}{cc} S_{11.j} & S_{12.j} \\ S'_{12.j} & S_{22.j} \end{array}\right) & \begin{array}{c} p \\ q \end{array} \\ \begin{array}{cc} p & q \end{array} & \end{array}=\tilde{T}'_j\,\tilde{T}_j\,,$$

$$\tilde{T}_j=\begin{array}{cc} \left(\begin{array}{cc} \tilde{T}_{1.j} & 0 \\ T_{2.j} & \tilde{T}_{3.j} \end{array}\right) & \begin{array}{c} p \\ q \end{array} \\ \begin{array}{cc} p & q \end{array} & \end{array} \quad \text{and} \quad \tilde{D}'_j\,\tilde{D}_j=I_{r_j}-X'_j\,S^{-1}_{22.j}\,X_j \quad \text{for}$$

$j=1,2,\ldots,k$ are all independently distributed under the null hypothesis $(\mu_i=\Sigma_{12}\Sigma^{-1}_{22}\nu_i,\ i=1,2,\ldots,k)$ and their distributions are respectively given by

(2.4.1) $$c_o|S_k|^{(m+n_k-p-q-1)/2}\exp(-\tfrac{1}{2}\mathrm{tr}\,\Sigma^{-1}S_k)\ .$$

$$\underset{\substack{X_i:\,q\times t_i\\ i=1,2,\dots,k}}{\int\cdots\int}\prod_{t=1}^{k}\left|S_{22.t}-\sum_{s=1}^{t}X_sX_s'\right|^{(m+n_{t-1}-q-1)/2}e^{\sum_{t=1}^{k}\mathrm{tr}\,\Sigma_{22}^{-1}\mathcal{U}_tX_t'}\prod_{t=1}^{k}dX_t$$

and

(2.4.2) $$c_j|I_p-F_jF_j'|^{(m+n_{j-1}-p-q-1)/2}\quad\text{for } j=1,2,\dots,k$$

where $c_j^{-1}=\pi^{-pr_j/2}\prod_{t=1}^{p}\Gamma\left(\frac{m+n_j-q-t+1}{2}\right)\Big/\Gamma\left(\frac{m+n_{j-1}-q-t+1}{2}\right)$

for $j=1,2,\dots k$; $n_t=\sum_{s=1}^{t}r_s$, $n_o=1$ and $c_o^{-1}=c_{k+1}^{-1}c_1c_2\cdots c_k$,

$c_{k+1}^{-1}=\pi^{(p+q)(p+q+2n_k-1)/4}|2\Sigma|^{(m+n_k)/2}\prod_{i=1}^{p+q}\Gamma\left(\frac{m-t+1}{2}\right)$.

Proof follows on the lines similar as in lemma 2.

<u>Corollary 3</u>: If $S:p\times p$, $Y_i:p\times r_i$ are independently distributed as $W(m,p;\Sigma;S)$ and $MN(0,\Sigma)$ $i=1,2,\dots,k$ then $S_k=S+\sum_{i=1}^{k}Y_iY_i'$ and $F_j=\tilde{T}_j'^{-1}Y_j$ where $S_j=S+\sum_{i=1}^{j}Y_iY_i'=\tilde{T}_j'\tilde{T}_j$, are all independently distributed and their distributions are respectively given by

(2.4.3) Wishart $(m+n_k,\ p;\Sigma;\ S_k)$ and

(2.4.4) $c_{1j}\,|I-F_jF_j'|^{(m+n_{j-1}-p-1)/2}$ where

where $\sum_{s=1}^{t} r_s = n_t$, $n_0=1$ and

$$c_{1j}=\pi^{-pr_j/2}\prod_{t=1}^{p}\Gamma\left(\frac{m+n_j-t+1}{2}\right)\Big/\Gamma\left(\frac{m+n_{j-1}-t+1}{2}\right) \text{for } j=1,\dots,k.$$

Lemma 6:- If the distributions of S and $\begin{pmatrix} Y_i \\ X_i \end{pmatrix}$ $i=1,2,\dots,k$ are the same as stated in lemma 5, then the distributions of nonzero characteristic roots of

$$P_j=\left[I_{r_j}-(I_{r_j}+X_j' S_{22.j}^{-1} X_j)\left\{I_{r_j}+(Y_j' \quad X_j')S_{j-1}^{-1}\begin{pmatrix} Y_j \\ X_j \end{pmatrix}\right\}^{-1}\right]$$

or $Q_j=\left\{I_p-(S_{11.j}-S_{12.j}S_{22.j}^{-1}S_{12.j}')^{-1}(S_{11.j-1}-S_{12.j-1}S_{22.j-1}^{-1}S_{12.j-1}')\right\}$

for $j=1,2,\dots,k$ and S_k are all independently distributed and the distributions of nonzero characteristic roots of P_j (or Q_j) are the same as those of the characteristic roots of F_jF_j' if $r_j \geq p$ (or $F_j'F_j$ if $r_j \leq p$), and the distribution of S_k is the same as given in (2.4.2). The distributions of F_jF_j' if $r_j \geq p$ are given by

$$(2.4.5) \quad c_j' \, |F_jF_j'|^{(r_j-p-1)/2} |I_p-F_jF_j'|^{(m+n_{j-1}-q-p-1)/2}$$

where $c_j'=c_j\pi^{-\frac{1}{4}p(p-1)+\frac{1}{2}pr_j}\prod_{t=1}^{p}\left\{\Gamma\left(\frac{r_j-t+1}{2}\right)\right\}^{-1}$

and those of $F_j'F_j$ if $p \geq r_j$ are given by

(2.4.6) $c''_j \left|F'_jF_j\right|^{(p-r_j-1)/2}\left|I_{r_j}-F'_jF_j\right|^{(m+n_{j-1}-q-p-1)/2}$

where $c''_j=c_j\pi^{-\frac{1}{4}r_j(r_j-1)+\frac{1}{2}pr_j}\prod_{t=1}^{r_j}\left\{\Gamma\left(\frac{p-t+1}{2}\right)\right\}^{-1}$.

This follows immediately from lemma 5 by using (A.3.1) and a similar result to the maximum root of F'_jF_j or $F_jF'_j$ can be established.

<u>Corollary 4</u>: If in lemma 6, all $r_j=1$ for $j=1,2,\dots,k$ then $P_j=1-R_j$ where R_j has the same definition as given in (68,40) or $R_j=(1+x'_jS^{-1}_{22.j-1}x_j)/\left\{1+(y'_j\ \ x'_j)S^{-1}_{j-1}\begin{pmatrix}y_j\\ x_j\end{pmatrix}\right\}$, $j=1,2,\dots,k$ are all independently distributed under the null hypothesis and their distributions are given by

(2.4.7) constant $P_j^{\frac{1}{2}p-1}(1-P_j)^{(m+j-q-p-2)/2}$.

<u>Corollary 5</u>: If the distributions of $S:p\times p$ and $Y_j:p\times r_j$ $j=1,2,\dots,k$ are the same as stated in corollary 3, then the distributions of the characteristic roots of $Q_j=(Y_jY'_j)S^{-1}_j$ if $r_j\geq p$ or $P_j=Y'_jS^{-1}_jY_j$ if $p\geq r_j$ and $S_k=S+\sum_{i=1}^{k}Y_iY'_i$ are independently distributed and the distributions of Q_j,P_j and S_k are respectively given by

(2.4.8) $c'_{1j} |Q_j|^{(r_j-p-1)/2} |I-Q_j|^{(m+n_{j-1}-p-1)/2}$ if $r_j \geqslant p$,

(2.4.9) $c''_{1j} |P_j|^{(p-r_j-1)/2} |I-P_j|^{(m+n_{j-1}-p-1)/2}$ if $r_j \leq p$

and S_k is Wishart $(m+n_k, p; \Sigma; S_k)$ where

$$c''_{1j}=c_{1j}\,\pi^{-\frac{1}{4} r_j(r_j-1)+\frac{1}{2}pr_j} \prod_{t=1}^{r_j}\left\{ \Gamma\left(\frac{p-t+1}{2}\right)\right\}^{-1} \quad \text{and}$$

$$c'_{1j}=c_{1j}\,\pi^{-\frac{1}{4}p(p-1)+\frac{1}{2}pr_j} \prod_{t=1}^{p}\left\{\Gamma\left(\frac{r_j-t+1}{2}\right)\right\}^{-1} .$$

<u>Corollary 6:</u> If in corollary 5, all $r_j=1$ for $j=1,2,\ldots,k$ then the distribution of $P_j=\underline{y}_j' S_j^{-1} \underline{y}_j=T_j^2/(1+T_j^2)$ where $T_j^2=\underline{y}_j' S_{j-1}^{-1} \underline{y}_j$ [Hotelling's T^2 (28)] for $j=1,2,\ldots,k$ are all independently distributed and their distributions are given by

(2.4.10) constant $P_j^{\frac{1}{2}p-1} (1-P_j)^{(m+j-p-2)/2}$ or

constant $(T_j^2)^{\frac{1}{2}p-1} (1+T_j^2)^{-(m+j)/2}$.

On a different line, K.S.Rao (72) has established only the corollary 6. The results given in corollary 3, corollary 4, corollary 5 and corollary 6 were established by the author earlier and published (40). More general results are established here.

2.5: A numerical example to illustrate the technique of computation (37):

The following is an example of the Egyptian skulls taken from Barnard (9), Bartlett (11), Rao (68), Kendall (35,36) and Anderson (5). Barnard has four series of skulls, 91 from late Predynastic, 162 from the Sixth-to-Twelfth, 70 from the Twelfth-and-Thirteenth, and 75 from Ptolemaic dynasties. On each four measurements were taken: x_1=maximum breadth, x_2=basial-veolar leng-th, x_3=nasal height, and x_4=basibregmatic height. The relevant data are summarised in Tables 1 and 2 which give the means for the four series and the analysis of dispersion.

Table 1 : Means for the four series

Character	I N_1=91	II N_2=162	III N_3=70	IV N_4=75
x_1	133.582418	134.265432	134.371429	135.306667
x_2	98.307692	96.462968	95.857143	95.040000
x_3	50.835165	51.148148	50.100000	52.093333
x_4	133.000000	134.882716	133.642857	131.466667

Table 2 : Analysis of Dispersion (S.P.M.)

	Q: between, 3 d.f.	S: Within(error), 394 d.f.	W: Total, 397 d.f.
x_1^2	123.180628	9661.997470	9785.178098
$x^2{}_2$	486.345863	9073.115027	9559.460890

x_3^2	150.411505	3938.320351	4088.731856
x_4^2	640.733891	8741.508829	9382.242720
x_1x_2	-231.375635	445.573301	214.197666
x_1x_3	87.305348	1130.623900	1217.929248
x_1x_4	-128 .763994	2148.584210	2019.820216
x_2x_3	-107.505618	1239.221990	1131.716372
x_2x_4	125.313318	2255.812722	2381.126040
x_3x_4	-137.580764	1271.054662	1133.473898

Q.I : Do the characters x_3 and x_4 show significant variation in the four series independently of the variation due to the characters x_1 and x_2 ?

A: Computational procedure:-

We shall adopt the abbreviated Doolittle Technique (17) to obtain $W_{22}-W'_{12}W^{-1}_{11}W_{12}$ and $S_{22}-S'_{12}S^{-1}_{11}S_{12}$ matrices from W and S.

On matrix W

9785.178098	214.197666	1217.929248	2019.820216
1	.021890012	.124466743	.206416296
	9554.772101	1105.055886	2336.912151
	1	.115654867	.244580627

Hence $$W_{22}-W'_{12}\,W^{-1}_{11}\,W_{12} = \begin{pmatrix} 3809.335078 & 611.798190 \\ 611.798190 & 8393.755473 \end{pmatrix}$$

Similarly on matrix S

9661.997470	445.573301	1130.623900	2148.584210
1	.046116065	.117017615	.222374744
	9052.567120	1187.082065	2156.728473
	1	.131132092	.238244958

Hence $S_{22}-S'_{12}\,S_{11}^{-1}\,S_{12} = \begin{pmatrix} 3650.352884 & 736.816146 \\ 736.816146 & 7749.888281 \end{pmatrix}$

Hence $(S_{22}-S'_{12}S_{11}^{-1}S_{12})^{-1}$ and $(S_{22}-S'_{12}S_{11}^{-1}S_{12})^{-1}(W_{22}-W'_{12}W_{11}^{-1}W_{12})$

are

$$10^{-4}\begin{pmatrix} 2.79306163 & -0.26554872 \\ -0.26554872 & 1.31558807 \end{pmatrix} \text{ and } \begin{pmatrix} 1.0477245416 & -0.052016097 \\ -0.0206689654 & 1.088026234 \end{pmatrix}.$$

Hence the determinantal equation for $\left\{ (S_{22}-S'_{12}S_{11}^{-1}\,S_{12})^{-1}\,(W_{22}-W'_{12}\,W_{11}^{-1}\,W_{12})-I \right\}$ is

$$\lambda^2 -0.1357507753\,\lambda + 0.0031258927 = 0\,.$$

i.e. $\lambda_1 = 0.10636143$ and $\lambda_2 = 0.029389345$.

Therefore g_1=generalised U=0.1357507753, λ_{max}=0.10636143 and

$$\frac{1}{\text{generalised R}} = \frac{1}{1.10636143} + \frac{1}{1.029389345} + 3-2$$

$$= 2.875313487$$

i.e. generalised R = 0.347788165 .

From the results given in ([illegible].2.20), taking m=394,

$p=q=2$, and $r=3$, F-test for generalised U is $F_{6,780}=8.8238$, F-test for generalised R is $F_{6,784}=8.6878$ and Θ_{max}-test for λ_{max} is $\Theta_{max}(0,194.5)=0.96136$.

All are highly significant, so that x_3 and x_4 may be considered as discriminating the series independently of x_1 and x_2.

(II): One of the topics discussed by Barnard (9) was the use of these measurements in discriminating between the four periods and possibility of the variates having a linear regression on time. The intervals between the four series were taken in the proportion 2:1:2 and may conveniently take the values of t as -5, -1, 1, 5.

Q.II(a): Can the variation of the four characters be accounted for by the linear regressions of individual characters on time?

(b): If (a) is not true, can the variation of the characters x_1 and x_2 removing the variations due to x_3 and x_4 be accounted for by the linear regressions on time?

A: The calculation of individual regressions involves the quantities: $\bar{t}=-0.432160804$, $\sum_{t}(t_t-\bar{t})^2=4307.668342$.

$\sum_{t} x_i(t_t-\bar{t})$: 718.7475357, -1407.2712179, 410.1003118, -733.4420851 .

Table 3 : Analysis of Dispersion(S.P.M.)

Dippersion due to

	Regression 1 d.f.	Q:Deviation from regression 2 d.f.	S:Within (Error) 394 d.f.	W=S+Q 396 d.f.
x_1^2	119.9252076	3.2554204	9661.997470	9665.2528904
x_2^2	459.7411229	26.6047401	9073.115027	9099.7199471
x_3^2	39.0425289	111.3689761	3938.320351	4049.6893271
x_4^2	124.8789947	515.8548963	8741.508829	9257.3637253
x_1x_2	-234.8074736	3.4318386	445.573301	449.0051396
x_1x_3	68.4264816	18.8788664	1130.623900	1149.5027664
x_1x_4	-122.3770377	-6.3869563	2148.584210	2142.1972537
x_2x_3	-133.9755802	26.4699622	1239.221990	1265.6919522
x_2x_4	239.6080325*	-114.2947145	2255.812722	2141.5180075
x_3x_4	-69.8254378	-67.7553262	1271.054662	1203.2993358

(i) Hence, we have S^{-1}

$$10^{-4}\begin{pmatrix} 1.11805532 & 0.04157560 & -0.25381309 & -0.24863127 \\ 0.04157560 & 1.21076884 & -0.30299437 & -0.27861027 \\ -0.25381309 & -0.30299437 & 2.79306163 & -0.26554872 \\ -0.24863127 & -0.27861027 & -0.26554872 & 1.31558807 \end{pmatrix}$$

---

* This item was wrongly computed in (11) and so its error was carried to in a number of books (68),(36) and (35).

Now to obtain the maximum root of QS^{-1} which is a 4x4 non-symmetric matrix. Here Q is based on 2 d.f. and so we shall write $Q=BD_qB'$ where $B':2x4 = \begin{pmatrix} 1 & b_{21} & b_{31} & b_{41} \\ 0 & 1 & b_{32} & b_{42} \end{pmatrix}$

and $D_q:2x2 = \begin{pmatrix} q_1 & 0 \\ 0 & q_2 \end{pmatrix}$. Then nonzero roots of QS^{-1} are the nonzero roots of $D_qB'S^{-1}B$. The matrix B and D_q are obtained by Doolittle method as follows:-

On matrix Q

3.2554204	3.4318386	18.8788664	-6.3869563
1	1.05419214	5.79920996	-1.96194516
	22.98692282	6.56800963	-107.56163537
	1	.28572809	-4.67925334

If we still proceed further, we must have zero throughout, for rank Q=2, but we do not find it so. This happens because the matrix Q is obtained by rounding the last digits. Hence omitting the latter part we write

$$B'=\begin{pmatrix} 1 & 1.05419214 & 5.79920996 & -1.96194516 \\ 0 & 1 & .28572809 & -4.67925334 \end{pmatrix}$$

and $D_q=\begin{pmatrix} 3.25542040 & 0.00000000 \\ 0.00000000 & 22.98692282 \end{pmatrix}$.

Hence $B'S^{-1}B = \begin{pmatrix} 0.010307049847 & 0.002654196333 \\ 0.002654196333 & 0.003338844070 \end{pmatrix}$.

Hence the nonzero roots of QS^{-1} are obtained from the determinantal equation of $D_q B' S^{-1} B$, i.e. they are the roots of

$$\lambda^2 - 0.11030353129\,\lambda + 0.002048070255 = 0 .$$

i.e. $\lambda_1 = 0.0866739$ and $\lambda_2 = 0.0236296$.

Note that $\mathrm{tr}QS^{-1}$ by this method is 0.11030353129 while that by direct multiplying Q by S^{-1}, we get 0.11030751. The difference is at the sixth place which is due to rounding off.

Hence $\lambda_{max} = 0.0866739$, $g_1 = 0.11030353$ and

$$\frac{1}{\text{generalised } R} = \frac{1}{1.0866739} + \frac{1}{1.02362960} + 1+1+2-4 = 1.897155142$$

Therefore generalised R = 0.527105 .

From the results given in (3.2.20), taking m=394, p=4, q=0, r=2, F-test for g_1 gives $F_{8,780} = 5.3775$ and that for g_2 gives $F_{8,784} = 5.31258$ and θ_{max}-test for λ_{max} is

$\theta(0.5, 148.5) = 0.07976$.

All are highly significant. Hence the regressions cannot be considered linear.

(ii) By the Doo-little technique given above, we have as in case (I), $S_{11} - S_{12} S_{22}^{-1} S'_{12}$ and $W_{11} - W_{12} W_{22}^{-1} W'_{12}$ respectively as

$$\begin{pmatrix} 8955.536550 & -307.516836 \\ -307.516836 & 8269.774508 \end{pmatrix} \text{and} \begin{pmatrix} 8974.655369 & -267.450256 \\ -267.450256 & 8353.933861 \end{pmatrix}$$

and the inverse of $S_{11}-S_{12}\,S_{22}^{-1}\,S'_{12}$ is

$$10^{-4}\begin{pmatrix} 1.1805532 & .0415756 \\ .0415756 & 1.2107688 \end{pmatrix}.$$

The maximum root of $S_{1.2}^{-1}\,W_{1.2}$ $(S_{1.2}=S_{11}-S_{12}S_{22}^{-1}S'_{12}$, $W_{1.2}=W_{11}-W_{12}W_{22}^{-1}\,W'_{12})$ is obtained from the maximum root of

$$\lambda^2 -2.0126605181\lambda+1.0126605677=0$$

i.e. $\lambda_1=1.012656646$ and $\lambda_2=1.0000038721$.

Therefore $\lambda_{max}=0.012656646$ and $g_1=0.0126605181$ and

$$\frac{1}{\text{generalised R}} = \frac{1}{1.012656646} + \frac{1}{1.0000038721} +2-2= 1.98749771 .$$

From the results given in (3.2.20), taking $m=394$, $p=q=r=2$, F-test for g_1 gives $F_{4,780}=1.2344$, that for g_2 gives $F_{4,784}=1.2329$ and θ_{max}-test for λ_{max} gives $\theta(-0.5,\ 148.5)=0.0124985$.

Hence we find that the results are not significant and so the variations of x_1 and x_2, removing those due to x_3 and x_4, can be accounted for by the linear regressions on time.

*****