複變值直線迴歸分析理論及其應用於水稻品種地域適應性之分析

魏　應　澤

(一) 緒　　論

稻穀產量受各地域之環境及氣候因子變化之影響相當大，各水稻品種有其最適應的地域及氣候。爲了尋求各地域氣候因子與稻穀產量之關係，我們通常採用複迴歸分析法 (Multiple Regression Analysis)，但是除了各地域氣候因子能影響稻穀產量之外，尚有水稻品種本身之農藝性狀諸如株高，有效分蘖數等等皆可影響穀產量之高低，這些農藝性狀仍受各地域氣候因子之影響，所以我們倘若僅考慮穀產量依隨氣候因子之複迴歸或僅考慮穀產量依隨其他農藝性狀之複迴歸，則將得不到完整而精確之推論結果。因此，我們將考慮穀產量和其他農藝性狀依隨氣候因子之複變值迴歸分析方法 (Multivariate Regression Analysis)。

本研究首先討論複變值迴歸模式 (Multivariate Regression Model) 中各種族群未知常數 (Parameters of a Population) 之推算問題及各種統計假說 (Statistical Hypothesis) 之測驗問題。然後將其應用於水稻品種地域適應性之研究。

(二) 複變值直線迴歸模式及複變值常態分佈之誘導 (Multivariate Linear Regression Model and the Derivation of Multivariate Normal Distribution)

玆設從每一抽樣單位 (Sampling Unit) 上可獲得 p 種具有相關關係之逢機變值 (Random Variates)，以向量 **y** 表示之，並且在同一抽樣單位上可獲得 q 種共軛變值 (Concomitant Variates)，以向量 **x** 表示之，則 **y** 可用下列之直線迴歸模式表示之：

$$\mathbf{y}=\mathbf{B}\mathbf{x}+\mathbf{d} \tag{2·1}$$

式中 B 爲 $p\times q$ 之係數行列 (Coefficient Matrix)，**d** 爲 $p\times 1$ 之剩餘機差向量 (Residual Error Vector)，其在 **x** 固定下的附件期望值 (Conditional Expected Value) 及附件分散量行列 (Conditional Dispersion Matrix) 如次：

$$\left.\begin{array}{l} E(\mathbf{d}|\mathbf{x})=0 \\ E(\mathbf{d}\mathbf{d}'|\mathbf{x})=\Sigma \end{array}\right\} \tag{2·2}$$

式中 Σ 爲 $p\times p$ 之分散量行列，該行列是一個正值行列 (Positive Definite Matrix)：

$$\Sigma=\begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots\cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots\cdots & \sigma_{2p} \\ \vdots & \vdots & & \\ \sigma_{p1} & \sigma_{p2} & \cdots\cdots & \sigma_{pp} \end{pmatrix}$$

式中 $\sigma_{ij}=\sigma_{ji}$ 係 y_i 與 y_j 之變積 (Covariance) 而 σ_{ii} 爲 y_i 之變方 (Variance)。

現在假定 **y** 中 p 種具有相關關係之逢機變值是從具有不同期望值及變方之 p 種常態族群 (Normal Population) 逢機抽出的觀測值。我們欲誘導此 p 種逢機變值之聯合機率密度函數 (Joint Probability

Density Function)。令 $z_1, \cdots\cdots, z_p$ 爲 p 種互相獨立之逢機變值，每一變值之分布都是期望值等於零而變方等於 1 之常態分布，並令 A 爲行列式值 (Determinant) 不等於零之 $p\times p$ 行列 (Nonsingular Matrix)，則 $\mathbf{y}$ 與 $\mathbf{z}$ 間可設立下列之函數關係：

$$\mathbf{y}=\mathbf{A}\mathbf{z}+\mathbf{B}\mathbf{x}$$

根據 $\mathbf{z}$ 之定義，我們可得 $\mathbf{z}'=(z_1, \cdots\cdots, z_p)$ 之聯合機率密度函數如下：

$$g(\mathbf{z}')=(2\pi)^{-\frac{p}{2}}e^{-\frac{1}{2}\mathbf{z}'\mathbf{z}}$$

玆因 A 爲行列式值不等於零之行列，其逆行列 (Inverse Matrix) 存在，故我們可得下列之變數轉換 (Transformation)：

$$\mathbf{z}=\mathbf{A}^{-1}(\mathbf{y}-\mathbf{B}\mathbf{x})$$

若以 a^{ij} 表示 A^{-1} 之元素，則顯而易見 $\partial z_i/\partial y_j=a^{ij}$，故變數轉換之 Jacobian 爲 A^{-1} 之行列式值的絕對值。因而 $\mathbf{y}'=(y_1, \cdots\cdots, y_p)$ 之聯合機率密度函數可得如下：

$$f(\mathbf{y}')=g(\mathbf{z}')\left|\frac{\partial \mathbf{z}'}{\partial \mathbf{y}'}\right|=(2\pi)^{-\frac{p}{2}}e^{-\frac{1}{2}(\mathbf{y}-\mathbf{B}\mathbf{x})'(\mathbf{A}^{-1})'(\mathbf{A}^{-1})(\mathbf{y}-\mathbf{B}\mathbf{x})}\,|\mathbf{A}^{-1}|$$

令 $\mathbf{C}=\mathbf{A}\mathbf{A}'$，則因 $(\mathbf{A}')^{-1}=(\mathbf{A}^{-1})'$，我們有下列之關係：

$$\mathbf{C}^{-1}=(\mathbf{A}\mathbf{A}')^{-1}=(\mathbf{A}')^{-1}\mathbf{A}^{-1}=(\mathbf{A}^{-1})'\mathbf{A}^{-1}$$

且 $$|\mathbf{C}^{-1}|=|(\mathbf{A}^{-1})'||\mathbf{A}^{-1}|=|\mathbf{A}^{-1}|^2$$

∴ $$\big||\mathbf{A}^{-1}|\big|=\sqrt{|\mathbf{C}^{-1}|}$$

故 $$f(\mathbf{y}')=\frac{\sqrt{|\mathbf{C}^{-1}|}}{(2\pi)^{p/2}}e^{-\frac{1}{2}(\mathbf{y}-\mathbf{B}\mathbf{x})'\mathbf{C}^{-1}(\mathbf{y}-\mathbf{B}\mathbf{x})}$$

式中 $\mathbf{C}^{-1}$ 是對稱正值行列 (Symmetric and Positive Definite Matrix)，蓋因

$$(\mathbf{C}^{-1})'=[(\mathbf{A}\mathbf{A}')^{-1}]'=[(\mathbf{A}\mathbf{A}')']^{-1}=(\mathbf{A}\mathbf{A}')^{-1}=\mathbf{C}^{-1}$$

且在所有不等於零之向量 $\mathbf{w}$ ($\neq 0$) 與 $\mathbf{C}^{-1}$ 構成的二次式 (Quadratic Form) 恆大於零故也，即

$$\mathbf{w}'\mathbf{C}^{-1}\mathbf{w}=\mathbf{w}'(\mathbf{A}^{-1})'(\mathbf{A}^{-1})\mathbf{w}=(\mathbf{A}^{-1}\mathbf{w})'(\mathbf{A}^{-1}\mathbf{w})>0$$

在 $\mathbf{x}$ 固定下，$\mathbf{y}$ 之期望值如次：

$$E(\mathbf{y}|\mathbf{x})=E(\mathbf{A}\mathbf{z}+\mathbf{B}\mathbf{x}|\mathbf{x})=\mathbf{B}\mathbf{x}$$

因 $E(\mathbf{z})=\mathbf{0}$ 故也。又 $\mathbf{y}$ 之分散量行列如下：

$$\Sigma=E[(\mathbf{y}-\mathbf{B}\mathbf{x})(\mathbf{y}-\mathbf{B}\mathbf{x})'|\mathbf{x}]$$

$$=\int_{-\infty}^{\infty}\cdots\cdots\int_{-\infty}^{\infty}(\mathbf{y}-\mathbf{B}\mathbf{x})(\mathbf{y}-\mathbf{B}\mathbf{x})'f(\mathbf{y}')d\mathbf{y}'$$

$$=\frac{\sqrt{|\mathbf{C}^{-1}|}}{(2\pi)^{p/2}}\int_{-\infty}^{+\infty}\cdots\cdots\int_{-\infty}^{\infty}(\mathbf{y}-\mathbf{B}\mathbf{x})(\mathbf{y}-\mathbf{B}\mathbf{x})'e^{-\frac{1}{2}(\mathbf{y}-\mathbf{B}\mathbf{x})'\mathbf{C}^{-1}(\mathbf{y}-\mathbf{B}\mathbf{x})}d\mathbf{y}'$$

令 $\mathbf{y}=\mathbf{A}\mathbf{z}+\mathbf{B}\mathbf{x}$，則變數轉換之 Jacobian 爲

$$\left|\frac{\partial \mathbf{y}'}{\partial \mathbf{z}'}\right|=|\mathbf{A}|=\sqrt{|\mathbf{C}|}$$

然而 $$\Sigma=\frac{\sqrt{|\mathbf{C}^{-1}|}}{(2\pi)^{p/2}}\int_{-\infty}^{\infty}\cdots\cdots\int_{-\infty}^{\infty}\mathbf{A}\mathbf{z}\mathbf{z}'\mathbf{A}'e^{-\frac{1}{2}\mathbf{z}'\mathbf{z}}\sqrt{|\mathbf{C}|}\,d\mathbf{z}'$$

$$=A\left\{(2\pi)^{-p/2}\int_{-\infty}^{\infty}\cdots\cdots\int_{-\infty}^{\infty} zz'e^{-\frac{1}{2}z'z}\,dz'\right\}A'$$

因 $E(z_i)=0$，$E(z_i^2)=1$ 及 $E(z_iz_j)=0$，故

$$(2\pi)^{-\frac{p}{2}}\int_{-\infty}^{\infty}\cdots\cdots\int_{-\infty}^{\infty} zz'e^{-\frac{1}{2}z'z}\,dz'=E(zz')=I$$

故 $$\Sigma=AIA'=AA'=C$$

由此可見，y 之分散量行列爲對稱正值行列。玆將上列所得結果歸納於下列之定理：

【定理 2·1】 在複變值直線廻歸模式 (2·1) 式中 y' 之複變值常態機率密度函數 (Multivariate Normal Probability Density Function) 爲

$$f(y')=\frac{1}{(2\pi)^{p/2}\sqrt{|\Sigma|}}e^{-\frac{1}{2}(y-Bx)'\Sigma^{-1}(y-Bx)} \tag{2·3}$$

其期望值等於 Bx 而分散量 (Dispersion) 等於對稱正值行列 Σ (Symmetric and Positive Definite Matrix)。

今後爲方便計以 $N(Bx,\Sigma)$ 表示具有期望值等於 Bx 而分散量等於 Σ 之複變值常態分布。

(三) 複變值直線廻歸未知常數之推算 (Estimation of Parameters in Multivariate Linear Regression)

玆欲應用最大機率推算法 (Maximum Likelihood Estimation) 來推算複變值直線廻歸之未知常數。設從 $N(Bx,\Sigma)$ 中逢機抽出 n 組觀測值 (y_h', x_h'), $h=1,\cdots\cdots,n$，y_h 爲 $p\times1$ 逢機變值向量，x_h 爲 $q\times1$ 無機差變值向量，B 爲 $p\times q$ 未知係數行列，Σ 爲 $p\times p$ 未知分散量之對稱正值行列。我們假定 $n\geq p+q$ 及 $X=(x_1,\cdots\cdots,x_n)$ 之階數(Rank) 等於 q。$(y_1',\cdots\cdots,y_n')$ 之聯合機率密度函數或簡稱爲機率函數 (Likelihood Function) 如次：

$$L=(2\pi)^{-\frac{1}{2}np}|\Sigma^{-1}|^{\frac{n}{2}}\exp\left\{-\frac{1}{2}\sum_{h=1}^{n}(y_h-Bx_h)'\Sigma^{-1}(y_h-Bx_h)\right\}$$

將上式取自然對數則得

$$\begin{aligned}\log L&=-\frac{1}{2}np\log(2\pi)+\frac{n}{2}\log|\Sigma^{-1}|-\frac{1}{2}\sum_{h=1}^{n}(y_h-Bx_h)'\Sigma^{-1}(y_h-Bx_h)\\&=-\frac{1}{2}np\log(2\pi)+\frac{n}{2}\log|\Sigma^{-1}|\\&\quad-\frac{1}{2}\,\mathrm{tr}\left[\Sigma^{-1}\left\{\sum_{h=1}^{n}y_hy_h'-2B\sum_{h=1}^{n}x_hy_h'+B\sum_{h=1}^{n}x_hx_h'B'\right\}\right]\end{aligned}$$

式中 tr 係 "trace" 之縮寫。現在就 log L 對 B 偏微分並令其結果等於零，則得

$$\frac{\partial\log L}{\partial B}=\Sigma^{-1}\left[-2\sum_{h=1}^{n}y_hx_h'+2B\sum_{h=1}^{n}x_hx_h'\right]=0$$

從上式可得下列之常規方程式 (Normal Equations)：

$$B\sum_{h=1}^{n}x_hx_h'=\sum_{h=1}^{n}y_hx_h'$$

上式可簡寫爲

$$BXX'=YX'$$

式中 $X=(x_1,\cdots\cdots,x_n)$，$Y=(y_1,\cdots\cdots,y_n)$。

因爲我們假設 X 之階數爲 q，故 XX′ 爲行列式值不等於零的行列，其逆行列存在，所以 B 之最大機率推算式 (Maximum Likelihood Estimator) 如下：

$$\hat{B}=YX'(XX')^{-1} \tag{3.1}$$

倘若我們將上式中之各種行列改變成下列之形式時

$$\underset{q\times p}{B}=\begin{pmatrix}\beta_1'\\ \vdots\\ \beta_q'\end{pmatrix},\quad \underset{n\times q}{X}=\begin{pmatrix}x_1'\\ \vdots\\ x_n'\end{pmatrix},\quad \underset{n\times p}{Y}=\begin{pmatrix}y_1'\\ \vdots\\ y_n'\end{pmatrix}$$

則 (3.1) 式可寫成如下：

$$\hat{B}=(X'X)^{-1}X'Y \tag{3.2}$$

可見上式中 $\hat{B}$ 行列所含各向量 β_i' 與複迴歸模式中迴歸係數之推算式同形。

其次，我們欲求 Σ 之最大機率推算式。將 log L 寫成如下：

$$\log L=-\frac{1}{2}np\log(2\pi)+\frac{n}{2}\log|\sigma^{ij}|-\frac{1}{2}\sum_{i=1}^{p}\sum_{j=1}^{p}\sigma^{ij}d_{ij}$$

式中 σ^{ij} 爲 Σ^{-1} 之第 (i, j) 元素，

$$d_{ij}=\sum_{h=1}^{n}(y_{ih}-\sum_{k=1}^{q}\beta_{ik}x_{kh})(y_{jh}-\sum_{l=1}^{q}\beta_{jl}x_{lh})$$

現在就 log L 對 σ^{tt} 及 σ^{tu} 分別求偏微分如下：

$$\frac{\partial\log L}{\partial\sigma^{tt}}=\frac{1}{2}\frac{n}{|\Sigma^{-1}|}\frac{\partial|\Sigma^{-1}|}{\partial\sigma^{tt}}-\frac{1}{2}d_{tt}$$

$$=\frac{1}{2}\frac{n}{|\Sigma^{-1}|}\operatorname{cof}(\sigma^{tt})-\frac{1}{2}d_{tt}$$

式中 $\operatorname{cof}(\sigma^{tt})$ 是 $|\Sigma^{-1}|$ 中 σ^{tt} 之子行列式 (Cofactor)。

$$\frac{\partial\log L}{\partial\sigma^{tu}}=\frac{n}{|\Sigma^{-1}|}\operatorname{cof}(\sigma^{tu})-d_{tu}$$

令上列偏微分結果等於零，則得

$$\frac{\operatorname{cof}(\sigma^{tu})}{|\Sigma^{-1}|}=\frac{1}{n}d_{tu}$$

但上式等號左邊係 Σ 之第 (u, t) 元素，故 Σ 之最大機率推算式如下：

$$\hat{\Sigma}^*=\frac{1}{n}\hat{D}=\frac{1}{n}\sum_{h=1}^{n}(y_h-\hat{B}x_h)(y_h-\hat{B}x_h)'$$

$$=\frac{1}{n}\Big[\sum_{h=1}^{n}y_hy_h'-\hat{B}\sum_{h=1}^{n}x_hy_h'\Big]$$

$$=\frac{1}{n}\Big[\sum_{h=1}^{n}y_hy_h'-\hat{B}\sum_{h=1}^{n}x_hx_h'\hat{B}'\Big] \tag{3.3}$$

上式可簡寫爲

$$\hat{\Sigma}^*=\frac{1}{n}[YY'-\hat{B}(XX')\hat{B}'] \tag{3.4}$$

式中 $Y=(y_1,\cdots\cdots,y_n)$，$X=(x_1,\cdots\cdots,x_n)$。

我們可察知複迴歸模式中 σ^2 之最大機率推算式與 (3.4) 式中之元素具有相同之形式。

（四） 最大機率推算式 $\hat{B}$ 及 $\hat{\Sigma}^*$ 之分布 (Distribution of the Maximum Likelihood Estimates $\hat{B}$ and $\hat{\Sigma}^*$)

因爲 $\hat{B}$ 中之每一元素 β_{ij} $(i=1,\cdots\cdots,p;\ j=1,\cdots\cdots,q)$ 都是常態變值 y 之直線函數，故 $\hat{B}$ 之分布

亦爲常態分布。玆求 $\hat{B}$ 之期望值及分散量於次：根據（2•1）式可將（3•1）式中之 Y 行列寫成爲

$$Y=BX+D \tag{4•1}$$

玆爲方更計，令

$$\left.\begin{aligned}
Y&=(y_1,\cdots\cdots,y_n)=\begin{pmatrix}w_1'\\ \vdots\\ w_p'\end{pmatrix}=W'\\
X&=(x_1,\cdots\cdots,x_n)=\begin{pmatrix}z_1'\\ \vdots\\ z_q'\end{pmatrix}=Z'\\
B&=(\beta_1,\cdots\cdots,\beta_q)=\begin{pmatrix}\theta_1'\\ \vdots\\ \theta_p'\end{pmatrix}=\theta'\\
D&=(d_1,\cdots\cdots,d_n)=\begin{pmatrix}\eta_1'\\ \vdots\\ \eta_p'\end{pmatrix}=\eta'
\end{aligned}\right\} \tag{4•2}$$

則（4•1）式可寫成

$$W=Z\theta+\eta \tag{4•3}$$

然而 θ 之最大機率推算式可從（3•1）式得如下：

$$\hat{\theta}=(Z'Z)^{-1}Z'W=\hat{B}' \tag{4•4}$$

現在求 $\hat{\theta}$ 之期望值於次：

$$\begin{aligned}E(\hat{\theta}|Z)&=E[(Z'Z)^{-1}Z'W|Z]=E[(Z'Z)^{-1}Z'(Z\theta+\eta)|Z]\\&=E[\theta+(Z'Z)^{-1}Z'\eta|Z]=\theta\end{aligned} \tag{4•5}$$

可見 $\hat{B}$ 爲 B 之無偏推算式 (Unbiased Estimator)。

繼之，我們欲求 $\hat{\theta}$ 中 $\hat{\theta}_i$ 與 $\hat{\theta}_j$ 之分散量。從（4•3）式得

$$w_i=Z\theta_i+\eta_i$$

$$w_j=Z\theta_j+\eta_j$$

而從（4•4）式得 θ_i 及 θ_j 之最大機率推算式如次：

$$\hat{\theta}_i=(Z'Z)^{-1}Z'w_i$$

$$\hat{\theta}_j=(Z'Z)^{-1}Z'w_j$$

由（4•5）式得知 $E(\hat{\theta}_i)=\theta_i$，$E(\hat{\theta}_j)=\theta_j$，因此 $\hat{\theta}_i$ 與 $\hat{\theta}_j$ 之分散量如次：

$$\begin{aligned}E[(\hat{\theta}_i-\theta_i)(\hat{\theta}_j-\theta_j)'|Z]&=E[(Z'Z)^{-1}Z'\eta_i\eta_j'Z(Z'Z)^{-1}|Z]\\&=(Z'Z)^{-1}Z'(\sigma_{ij}I)Z(Z'Z)^{-1}\\&=\sigma_{ij}(Z'Z)^{-1}\end{aligned} \tag{4•6}$$

因此，$\hat{B}$ 中第 i 及第 j 橫向量之分散量爲 $\sigma_{ij}(XX')^{-1}$。今若將 $\hat{B}$ 之橫向量排成一行卽 $(\theta_1',\cdots\cdots,\theta_p')$，則其分散量行列如次：

$$\begin{pmatrix}\sigma_{11}S^{-1} & \sigma_{12}S^{-1} & \cdots\cdots & \sigma_{1p}S^{-1}\\ \sigma_{21}S^{-1} & \sigma_{22}S^{-1} & \cdots\cdots & \sigma_{2p}S^{-1}\\ \vdots & \vdots & & \vdots\\ \sigma_{p1}S^{-1} & \sigma_{p2}S^{-1} & \cdots\cdots & \sigma_{pp}S^{-1}\end{pmatrix}=\Sigma\otimes(XX')^{-1} \tag{4•7}$$

式中 $S=XX'$，$\Sigma\otimes(XX')^{-1}$ 表示 Σ 與 $(XX')^{-1}$ 之 Kronecker 乘積或直接乘積 (Direct Product)。

其次，我們欲求者爲 $n\hat{\Sigma}^*=YY'-\hat{B}XX'\hat{B}'$ 之分布。因爲 $S=XX'$ 爲行列式值不等於零之行列 (Nonsingular Matrix)，我們可找出一個能滿足下列關係之 $q\times q$ 行列 G

$$GSG'=I \qquad G'^{-1}S^{-1}=I$$

令 $H_2=GX$，則 $X=G^{-1}H_2$，又

$$H_2H_2'=GXX'G'=GSG'=I$$

因此，H_2 中所含 q 個 $1\times n$ 橫向量互成直交 (Orthogonal)。我們可另外找出一個 $(n-q)\times n$ 之行列 H_1 與 H_2 合併而成爲一個直交行列 H，即

$$H=\begin{bmatrix}H_1\\H_2\end{bmatrix}$$

令 $U=(u_1,\cdots\cdots,u_n)=YH'$，則

$$\begin{aligned}YH'&=(y_1,\cdots\cdots,y_n)H'\\&=\begin{pmatrix}w_1'\\\vdots\\w_p'\end{pmatrix}H'=\begin{pmatrix}w_1'H'\\\vdots\\w_p'H'\end{pmatrix}=\begin{pmatrix}\lambda_1'\\\vdots\\\lambda_p'\end{pmatrix}=(u_1,\cdots\cdots,u_n)\end{aligned}$$

玆求 U 之期望值及分散量於次：

$$E(U)=E(YH')=BXH'=BG^{-1}H_2(H_1'|H_2')=BG^{-1}(O|I)=(O|BG^{-1})$$

可見 U 行列中居首 $(n-q)$ 個列向量 (Column Vectors) $u_i, i=1,\cdots\cdots,n-q$ 之期望值都等於 0。因

$$\begin{aligned}(w_1'H',\cdots\cdots,w_p'H')&=(w_1',\cdots\cdots,w_p')\begin{pmatrix}H' & & O\\ & H' & \\ O & & H'\end{pmatrix}\\&=(w_1',\cdots\cdots,w_p')I\otimes H'\\&=(\lambda_1',\cdots\cdots,\lambda_p')\end{aligned}$$

故

$$\begin{pmatrix}\lambda_1\\\vdots\\\lambda_p\end{pmatrix}=I\otimes H\begin{pmatrix}w_1\\\vdots\\w_p\end{pmatrix}$$

因爲 $(w_1',\cdots\cdots,w_p')$ 之分散量爲 $\sum\otimes I$，故 $(\lambda_1',\cdots\cdots,\lambda_p')$ 之分散量如下：

$$(I\otimes H)(\textstyle\sum\otimes I)(I\otimes H)=(I\textstyle\sum I)\otimes(HIH')=\textstyle\sum\otimes I \qquad (4\cdot9)$$

故 $(u_1',\cdots\cdots,u_n')$ 之分散量等於 $I\otimes\sum$。由此可見 u_i 之分布爲互相間成獨立的常態分布，而在 U 行列中居首之 $(n-q)$ 個 u_i 之分布爲 $N(0,\sum)$。

繼之，欲要證明的是

$$n\hat{\textstyle\sum}^*=YY'-\hat{B}S\hat{B}'=\sum_{i=1}^{n-q}u_iu_i'$$

因 $U=YH'$，故 $Y=UH$。然而

$$YY'=UHH'U'=UU'=\sum_{i=1}^{n}u_iu_i'$$

又

$$\begin{aligned}\hat{B}S\hat{B}'&=(YX'S^{-1})S(S^{-1}XY')=UHH_2'(G^{-1})'S^{-1}G^{-1}H_2H'U'\\&=U\begin{bmatrix}H_1\\H_2\end{bmatrix}H_2'H_2(H_1'H_2')U'=U\begin{bmatrix}O\\I\end{bmatrix}(O\ \ I)U'\\&=\sum_{i=n-q+1}^{n}u_iu_i'\end{aligned}$$

故

$$n\hat{\textstyle\sum}^*=\sum_{i=1}^{n}u_iu_i'-\sum_{i=n-q+1}^{n}u_iu_i'=\sum_{i=1}^{n-q}u_iu_i' \qquad (4\cdot10)$$

由上式可見 $n\hat{\sum}^*$ 與 $\hat{B}$ 互相獨立。現在因 $u_i(i=1,\cdots\cdots,n-q)$ 之分布爲互相獨立之 $N(0,\sum)$ 的緣故，$n\hat{\sum}^*$ 之分布爲自由度等於 $(n-q)$ 的 Wishart 氏分布，其機率密度函數可參閱 Andersen, T. W. (1958) 如次：

$$f(n\hat{\textstyle\sum}^*)=\frac{|n\hat{\sum}^*|^{\frac{1}{2}(n-p-q-1)}\exp\left[-\frac{n}{2}\operatorname{tr}(\hat{\sum}^*\sum^{-1})\right]}{2^{\frac{1}{2}p(n-q)}\pi^{\frac{1}{4}p(p-1)}|\sum|^{\frac{1}{2}(n-q)}\prod_{i=1}^{p}\Gamma\left[\frac{1}{2}(n-q+1-i)\right]} \qquad (4\cdot11)$$

今後將具有 $(n-q)$ 自由度的 Wishart 分布以 $W(\sum,n-q)$ 表示之。因爲 u_i 之分布是互相獨立的 $N(0,\sum)$，故從 (4·10) 式求期望值時可得

$$E(n\hat{\Sigma}^*)=(n-q)\Sigma$$

因此，我們可得 Σ 之無偏推算式 (Unbiased Estimator) 如次：

$$\hat{\Sigma}=\frac{n}{n-q}\hat{\Sigma}^*=\frac{1}{n-q}[YY'-\hat{B}XX'\hat{B}'] \tag{4·12}$$

玆將上面所得結果歸納於下列之定理：

【定理 4·1】 今設從 $N(Bx_i,\Sigma)$ 逢機抽出 n 個 $p\times1$ 複變值向量 $y_i(i=1,\cdots\cdots,n)$，則

（1）B 之最大機率推算式 (Maximun Likelihood Estimator) $\hat{B}$ 的分布爲常態，其期望值爲 B 而其第 i 及第 j 橫向量之分散量行列爲 $\sigma_{ij}(XX')^{-1}$，式中 σ_{ij} 爲 Σ 之第 (i, j) 元素而 $X=(x_1,\cdots\cdots,x_n)$，x_i 爲 $q\times1$ 向量。

（2）$n\Sigma$ 之最大機率推算式 $n\hat{\Sigma}^*=YY'-\hat{B}XX'\hat{B}'$ 的分布爲 Wishart 分布 $W(\Sigma,n-q)$。

（3）$n\hat{\Sigma}^*$ 與 $\hat{B}$ 互相獨立。

(五) 複變值廻歸係數之統計擬說測驗 (Testing Statistical Hypotheses About the Multivariate Regression Coefficients)

（1）測驗單一廻歸係數向量之顯著性

在廻歸係數行列 B 中，我們欲測驗其某一向量 β_i 是否與統計擬說之向量 β_{i0} 相等，卽測驗統計擬說 $H_0:\beta_i=\beta_{i0}$ 之顯著性。

我們知道 β_i 之無偏推算式爲 $\hat{B}$ 行列中第 i 列向量 $\hat{\beta}_i$。在第四節已證明將 $\hat{B}$ 之橫向量排成一行 $(\hat{\theta}_1',\cdots\cdots,\hat{\theta}_p')$ 時，其分散量行列等於 $\Sigma\otimes(XX')^{-1}$，今若將 $\hat{B}$ 之列向量排成一行 $(\hat{\beta}_1',\cdots\cdots,\hat{\beta}_q')$，則其分散量行列應等於 $(XX')^{-1}\otimes\Sigma$。因此，若令 c_{ij} 爲 $(XX')^{-1}$ 之第 (i, j) 元素，則 $\hat{\beta}_i$ 之分散量行列等於 $c_{ii}\Sigma$，故 $\hat{\beta}_i$ 之分布爲 $N(\beta_i,c_{ii}\Sigma)$。從第四節定理 4·1 得知 $\hat{\beta}_i$ 與 $(n-q)\hat{\Sigma}=YY'-\hat{B}XX'\hat{B}'$ 互相獨立且 $(n-q)\hat{\Sigma}$ 之分布爲 $W(\Sigma,n-q)$。因此，我們可利用 Hotelling 氏 T^2 來測驗 H_0。將 Hotelling 氏一般化 T^2 統計值之分布定理摘錄於次：

【定理 5·1】 設 $m\hat{\Sigma}$ 之分布爲 $W(\Sigma,m)$ 而 $p\times1$ 向量 **d** 之分布爲 $N\left(\partial,\frac{1}{c}\Sigma\right)$，並設 $m\hat{\Sigma}$ 與 **d** 互相獨立，則 Hotelling 氏一般化 T^2 統計值 (Generalized T^2-statistic) 如下：

$$T^2=cd'\hat{\Sigma}^{-1}d=\frac{d'\hat{\Sigma}^{-1}d}{d'\Sigma^{-1}d}(cd'\Sigma^{-1}d) \tag{5·1}$$

（i）$d'\Sigma^{-1}d/d'\hat{\Sigma}^{-1}d$ 之分布爲自由度等於 $m-p+1$ 之 χ^2 分布。

（ii）$d'\Sigma^{-1}d/d'\hat{\Sigma}^{-1}d$ 與 **d** 互相獨立。

（iii）$cd'\Sigma^{-1}d$ 之分布爲自由度等於 p 及非中心性常數 (Non-centrality Parameter) 等於 $c\partial'\Sigma^{-1}\partial$ 之 χ^2 分布。

（iv）$\frac{m-p+1}{p}\frac{T^2}{m}$ 之分布爲自由度等於 p 及 $m-p+1$ 之非中心 F 分布 (Noncentral F Distribution)，其非中心性常數爲 $c\partial'\Sigma^{-1}\partial$，當 $\partial=0$ 時，便成中心 F 分布。

上列定理之證明可參閱 Wijsman, R. A (1957)。其他有關文獻可參閱 Hotelling, H. (1931)，Bose, R. C. and Roy, S. N. (1938)，Hsu P. L. (1938)，及 Rao, C. R. (1946, 1949) 等。

現在利用上列定理來測驗 $H_0:\beta_i=\beta_{i0}$。T^2 統計值爲

$$T^2=c_{ii}^{-1}(\hat{\beta}_i-\beta_{i0})'\hat{\Sigma}^{-1}(\hat{\beta}_i-\beta_{i0}) \tag{5·2}$$

然而

$$\frac{n-q-p+1}{p}\frac{T^2}{n-q} \tag{5·3}$$

之分布爲自由度等於 p 及 $n-q$ 之非中心 F 分布，其非中心性常數爲 $c_{11}^{-1}(\beta_1-\beta_{10})'\Sigma^{-1}(\beta_1-\beta_{10})$。當 $H_0:\beta_1=\beta_{10}$ 屬眞時，其分布便是中心 F 分布。因此，我們可根據 Snedecor 氏之理論 F 值以測驗 H_0 之顯著性，即若 (5•3) 式所得數值大於理論 F 值，則我們可根據顯著機率水準 (Significance Level) 丢棄 $H_0:\beta_1=\beta_{10}$。

（2）測驗一組廻歸係數向量之顯著性

我們將 $p\times q$ 之 B 行列劃分爲二個行列如下：

$$B=(B_1|B_2)$$

式中 B_1 爲 $p\times q_1$ 行列而 B_2 爲 $p\times q_2$ 行列，$q_1+q_2=q$。現在我們擬採用機率比值軌範 (Likelihood Ratio Criterion) 以測驗下列之統計擬說

$$H_0:B_1=B_1^*$$

B_1^* 爲已知行列。令 Ω 表示未知常數空間 (Parameter Space) 而 ω 表示在 H_0 下之未知常數副空間 (Subspace)，按照第三節，機率函數爲

$$L=(2\pi)^{-\frac{1}{2}np}|\Sigma^{-1}|^{\frac{n}{2}}\exp\left[-\frac{1}{2}\sum_{h=1}^{n}(\mathbf{y}_h-B\mathbf{x}_h)'\Sigma^{-1}(\mathbf{y}_h-B\mathbf{x}_h)\right]$$

而我們已求得在 Ω 下 B 及 Σ 之最大機率推算式分別如 (3•1) 及 (3•4) 兩式。因此，若以 $L(\hat{\Omega})$ 表示在 Ω 下之最大的機率函數值，則

$$\begin{aligned}L(\hat{\Omega})&=(2\pi)^{-\frac{1}{2}np}|\hat{\Sigma}^*|^{-\frac{n}{2}}\exp\left[-\frac{1}{2}\sum_{h=1}^{n}(\mathbf{y}_h-\hat{B}\mathbf{x}_h)'\hat{\Sigma}^{*-1}(\mathbf{y}_h-\hat{B}\mathbf{x}_h)\right]\\&=(2\pi)^{-\frac{1}{2}np}|\hat{\Sigma}^*|^{-\frac{n}{2}}\exp\left[-\frac{1}{2}\operatorname{tr}\{\hat{\Sigma}^{*-1}\sum_{h=1}^{n}(\mathbf{y}_h-\hat{B}\mathbf{x}_h)(\mathbf{y}_h-\hat{B}\mathbf{x}_h)'\}\right]\\&=(2\pi)^{-\frac{1}{2}np}|\hat{\Sigma}^*|^{-\frac{n}{2}}\exp\left[-\frac{n}{2}\operatorname{tr}\{\hat{\Sigma}^{*-1}\hat{\Sigma}^*\}\right]\\&=(2\pi)^{-\frac{1}{2}np}|\hat{\Sigma}^*|^{-\frac{n}{2}}\exp\left[-\frac{n}{2}\operatorname{tr}(I)\right]\\&=(2\pi)^{-\frac{1}{2}np}|\hat{\Sigma}^*|^{-\frac{n}{2}}e^{-\frac{1}{2}np}\end{aligned}$$

爲了要區別在 Ω 及 ω 下之最大機率推算式起見，擬在推算式附加 Ω 及 ω 之脚註以資區別。所以，將上式寫成

$$L(\hat{\Omega})=(2\pi)^{-\frac{1}{2}np}|\hat{\Sigma}_\Omega^*|^{-\frac{n}{2}}e^{-\frac{1}{2}np}\tag{5•4}$$

其次，欲在 ω 下求 B_2 及 Σ 之最大機率推算式。玆按照 B 之劃分方式將 $\mathbf{x}_h$ 分成如下：

$$\mathbf{x}_h=\begin{pmatrix}\mathbf{x}_h^{(1)}\\\mathbf{x}_h^{(2)}\end{pmatrix}$$

式中 $\mathbf{x}_h^{(1)}$ 爲 $q_1\times1$ 向量而 $\mathbf{x}_h^{(2)}$ 爲 $q_2\times1$ 向量。令

$$\mathbf{z}_h=\mathbf{y}_h-B_1^*\mathbf{x}_h^{(1)}$$

則在 H_0 下 $\mathbf{z}_h$ 可視爲從 $N(B_2\mathbf{x}_h^{(2)},\Sigma)$ 逢機抽出之向量，因此 B_2 之最大機率推算式可做第三節求得如次：

$$\begin{aligned}\hat{B}_{2\omega}&=\sum_{h=1}^{n}\mathbf{z}_h\mathbf{x}_h^{(2)\prime}S_{22}^{-1}=\sum_{h=1}^{n}(\mathbf{y}_h-B_1^*\mathbf{x}_h^{(1)})\mathbf{x}_h^{(2)\prime}S_{22}^{-1}\\&=(C_2-B_1^*S_{12})S_{22}^{-1}\end{aligned}\tag{5•5}$$

式中 C_2，S_{12} 及 S_{22} 係按照 B 之割分方式將 $YX'=C$ 及 $XX'=S$ 割分所得之行列，即

$$YX'=C=(C_1|C_2)$$

$$XX'=S=\left(\frac{S_{11}\,|\,S_{12}}{S_{21}\,|\,S_{22}}\right)$$

在 H_0 下，$\sum$ 之最大機率推算式爲

$$n\hat{\sum}_\omega{}^*=\sum_{h=1}^{n}(z_h-\hat{B}_{2\omega}x_h{}^{(2)})(z_h-\hat{B}_{2\omega}x_h{}^{(2)})'$$

$$=\sum_{h=1}^{n}z_h z_h{}'-\hat{B}_{2\omega}S_{22}\hat{B}'_{2\omega}$$

$$=\sum_{h=1}^{n}(y_h-B_1{}^*x_h{}^{(1)})(y_h-B_1{}^*x_h{}^{(1)})'-\hat{B}_{2\omega}S_{22}\hat{B}_{2\omega}{}' \tag{5·6}$$

因此，在 ω 下之最大的機率函數值如下：

$$L(\hat{\omega})=(2\pi)^{-\frac{1}{2}np}\left|\hat{\sum}_\omega{}^*\right|^{-\frac{n}{2}}e^{-\frac{1}{2}np} \tag{5·7}$$

玆以 (5·7) 式 $L(\hat{\Omega})$ 除 (5·4) 式 $L(\hat{\omega})$，則得測驗 H_0 之機率比值，即

$$\lambda=\frac{L(\hat{\omega})}{L(\hat{\Omega})}=\frac{|\hat{\sum}_\Omega{}^*|^{n/2}}{|\hat{\sum}_\omega{}^*|^{n/2}} \tag{5·8}$$

我們可適當地選擇某一常數 λ_0 以作爲丟棄 H_0 與否之根據；即當 $\lambda<\lambda_0$ 時我們丟棄 H_0。此 λ_0 需根據所需要之顯著機率水準 (Significance Level) α 來選定，即

$$\int_0^{\lambda_0}g(\lambda\,|\,H_0)\,d\lambda=\alpha$$

於此，我們需要 λ 之機率密度函數 $g(\lambda|H_0)$。

(六) 測驗統計假說 $H_0:B_1=B_1{}^*$ 之機率比值的分布 (The Distribution of the Likelihood Ratio for Testing $H_0:B_1=B_1{}^*$)

爲方便計，我們擬求 $\lambda^{n/2}$ 之分布。首先欲要求出 $n\hat{\sum}_\omega{}^*$ 與 $n\hat{\sum}_\Omega{}^*$ 之關係。令

$$X=\left(\frac{X_1}{X_2}\right),\qquad \underset{q_1\times n}{X_1}=(x_1{}^{(1)},\cdots\cdots,x_n{}^{(1)}),\qquad \underset{q_2\times n}{X_2}=(x_1{}^{(2)},\cdots\cdots,x_n{}^{(2)})$$

$$\hat{B}_\Omega=(\underset{p\times q_1}{\hat{B}_{1\Omega}}\;|\;\underset{p\times q_2}{\hat{B}_{2\Omega}})$$

則我們有下列之恒等式。

$$Y-BX=(X-\hat{B}_\Omega X)+(\hat{B}_{2\Omega}-B_2)X_2+(\hat{B}_{1\Omega}-B_1{}^*)X_1$$

$$=(Y-\hat{B}_\Omega X)+(\hat{B}_{2\omega}-B_2)X_2-(\hat{B}_{2\omega}-\hat{B}_{2\Omega})X_2+(\hat{B}_{1\Omega}-B_1{}^*)X_1$$

從 (3·1) 式可得

$$\hat{B}_\Omega(XX')=YX'$$

或 $$(\hat{B}_{1\Omega}\,|\,\hat{B}_{2\Omega})\left(\frac{S_{11}\,|\,S_{12}}{S_{21}\,|\,S_{22}}\right)=(C_1\,|\,C_2)$$

因此 $$\hat{B}_{1\Omega}S_{12}+\hat{B}_{2\Omega}S_{22}=C_2$$

∴ $$\hat{B}_{2\Omega}=C_2S_{22}{}^{-1}-\hat{B}_{1\Omega}S_{12}S_{22}{}^{-1}$$

將 (5·5) 式之 $\hat{B}_{2\omega}=(C_2-B_1{}^*S_{12})S_{22}{}^{-1}$ 減去上式，則得

$$\hat{B}_{2\omega}-\hat{B}_{2\Omega}=(C_2-B_1{}^*S_{12})S_{22}{}^{-1}-(C_2-\hat{B}_{1\Omega}S_{12})S_{22}{}^{-1}=(\hat{B}_{1\Omega}-B_1{}^*)S_{12}S_{22}{}^{-1} \tag{6·1}$$

故 $$Y-BX=(Y-\hat{B}_{\Omega}X)+(\hat{B}_{2\omega}-B_2)X_2+(\hat{B}_{1\Omega}-B_1^*)(X_1-S_{12}S_{22}^{-1}X_2)$$

將上式等號兩邊各減去 $(\hat{B}_{2\omega}-B_2)X_2$，則得

$$Y-B_1^*X_1-\hat{B}_{2\omega}X_2=(Y-\hat{B}_{\Omega}X)+(\hat{B}_{1\Omega}-B_1^*)(X_1-S_{12}S_{22}^{-1}X_2)$$

因此，

$$\begin{aligned}n\hat{\Sigma}_{\omega}^*&=(Y-B_1^*X_1-\hat{B}_{2\omega}X_2)(Y-B_1^*X_1-\hat{B}_{2\omega}X_2)'\\&=(Y-\hat{B}_{\Omega}X)(Y-\hat{B}_{\Omega}X)'\\&\quad+(\hat{B}_{1\Omega}-B_1^*)(X_1-S_{12}S_{22}^{-1}X_2)(X_1-S_{12}S_{22}^{-1}X_2)'(\hat{B}_{1\Omega}-B_1^*)'\\&=n\hat{\Sigma}_{\Omega}^*+(\hat{B}_{1\Omega}-B_1^*)S_{11\cdot2}(\hat{B}_{1\Omega}-B_1^*)'\end{aligned}\tag{6.2}$$

式中 $$S_{11\cdot2}=S_{11}-S_{12}S_{22}^{-1}S_{21}$$

現在我們可將 $\lambda^{2/n}$ 寫成如次：

$$U=\lambda^{2/n}=\frac{|\hat{\Sigma}_{\Omega}^*|}{|\hat{\Sigma}_{\omega}^*|}=\frac{|n\hat{\Sigma}_{\Omega}^*|}{|n\hat{\Sigma}_{\Omega}^*+(\hat{B}_{1\Omega}-B_1^*)S_{11\cdot2}(\hat{B}_{1\Omega}-B_1^*)'|}\tag{6.3}$$

在定理 4·1 中，我們已知 $n\hat{\Sigma}_{\Omega}^*$ 之分布爲 $W(\Sigma, n-q)$ 而與 $\hat{B}_{\Omega}$ 互相獨立；$\hat{B}_{\Omega}$ 之分布爲常態，其期望值爲 B 而其第 i 及第 j 橫向量之分散量爲 $\sigma_{ij}S^{-1}$。玆擬證明下列之關係成立：

$$(\hat{B}_{1\Omega}-B_1^*)S_{11\cdot2}(\hat{B}_{1\Omega}-B_1^*)'=\sum_{i=1}^{q_1}z_i z_i'\tag{6.4}$$

式中 z_i 互相獨立，其分布爲 $N(0,\Sigma)$。

玆令 $A=S^{-1}$，則

$$\left(\begin{array}{c|c}A_{11}&A_{12}\\A_{21}&A_{22}\end{array}\right)\left(\begin{array}{c|c}S_{11}&S_{12}\\S_{21}&S_{22}\end{array}\right)=\left(\begin{array}{c|c}I&O\\O&I\end{array}\right)$$

由上式之關係可得 $A_{11}^{-1}=S_{11}-S_{12}S_{22}^{-1}S_{21}=S_{11\cdot2}$。因此，$\hat{B}_{1\Omega}$ 之第 i 及第 j 橫向量之分散量可由 (4·6) 式而得知爲 $\sigma_{ij}S_{11\cdot2}^{-1}$。因爲 $S_{11\cdot2}$ 是一種行列式值不等於零之行列，我們可找出能滿足下列關係之行列 H：

$$HS_{11\cdot2}H'=I$$

令 $$\hat{B}_{1\Omega}-B_1^*=ZH=(z_1,\cdots\cdots,z_{q_1})H$$

則 $$(\hat{B}_{1\Omega}-B_1^*)S_{11\cdot2}(\hat{B}_{1\Omega}-B_1^*)'=ZHS_{11\cdot2}H'Z'=ZIZ'=ZZ'=\sum_{i=1}^{q_1}z_i z_i'$$

因爲在 $H_0:B_1=B_1^*$ 屬眞之下，$E(\hat{B}_{1\Omega})=B_1^*$，故

$$E(Z)=E[(\hat{B}_{1\Omega}-B_1^*)H^{-1}]=0$$

令 $\hat{\theta}_{1i}$ 爲 $\hat{B}_{1\Omega}$ 之第 i 橫向量，且令 w_i 爲 Z 之第 i 橫向量，則

$$\begin{aligned}E(w_i'w_j)&=E(H^{-1})'(\hat{\theta}_{1i}-\theta_{1i}^*)'(\hat{\theta}_{1j}-\theta_{1j}^*)H^{-1}=\sigma_{ij}(H^{-1})'S_{11\cdot2}^{-1}H^{-1}\\&=\sigma_{ij}(HS_{11\cdot2}H')^{-1}=\sigma_{ij}I\end{aligned}$$

故 z_i 之分布爲 $N(0,\Sigma)$。

玆令 $V=n\hat{\Sigma}_{\Omega}^*$，$T=V+\sum_{i=1}^{q_1}z_i z_i'$，則

$$U=\frac{|V|}{|V+\sum_{i=1}z_i z_i'|}=\frac{|V|}{|T|}$$

我們先求 U 之第 h 級動差 (The h^{th} Moment)，然後由其探出 U 之分布。因爲 V 與 $\sum_i z_i z_i'$ 互相獨立，而 V 之分布爲 $W(\Sigma, n-q)$，z_i 之分布爲互相獨立之 $N(0,\Sigma)$，故 V 與 z_i's 之聯合機準密度函數

等於 V 之機率密度函數與 z_i's 之聯合機率密度函數相乘積，可從 (2•3) 及 (4•11) 兩式得之。所以

$$E(U^h)=E(|V|^h|T|^{-h})$$

$$=\int\cdots\cdots\int K(\Sigma, n-q)\,|V|^h\,|T|^{-h}\,|V|^{\frac{1}{2}(n-q-p-1)}\exp\left(-\frac{1}{2}\,\mathrm{tr}\,\Sigma^{-1}V\right)$$

$$\cdot(2\pi)^{\frac{1}{2}q_1 p}\,|\Sigma|^{-\frac{1}{2}q_1}\exp\left(-\frac{1}{2}\sum_{i=1}^{q_1} z_i'\Sigma^{-1}z_i\right)dVdZ$$

式中 $$K(\Sigma, n-q)=2^{\frac{1}{2}(n-q)p}\,\pi^{\frac{p(p-1)}{4}}\,|\Sigma|^{\frac{1}{2}(n-q)}\prod_{i=1}^{p}\Gamma\left[\tfrac{1}{2}(n-q+1-i)\right]$$

我們可將上式寫成如次：

$$E(U^h)=\frac{K(\Sigma, n-q)}{K(\Sigma, n-q+2h)}\int\cdots\cdots\int|T|^{-h}K(\Sigma, n-q+2h)|V|^{\frac{1}{2}(n-q+2h-p-1)}$$

$$\cdot\exp\left(-\frac{1}{2}\,\mathrm{tr}\,\Sigma^{-1}V\right)(2\pi)^{-\frac{1}{2}q_1 p}\,|\Sigma|^{-\frac{1}{2}q_1}\exp\left(-\frac{1}{2}\sum_{i=1}^{q_1}z_i'\Sigma^{-1}z_i\right)dV\,dZ$$

上列積分符號內含有 V 之部份爲 $W(\Sigma, n-q+2h)$ 之機率密度函數。我們知道 $\sum_{i=1}^{q_1}z_i z_i'$ 之分布爲 $W(\Sigma, q_1)$，其與 V 之分布互相獨立，因此，T 之分布可視爲 $W(\Sigma, n-q+2h+q_1)$。然而，上列之積分部份可視爲 $|T|^{-h}$ 之期望值，因此

$$E(U^h)=\frac{K(\Sigma, n-q)}{K(\Sigma, n-q+2h)}\int\cdots\cdots\int|T|^{-h}f(T|\Sigma, n-q+2h+q_1)\,dT$$

$$=\frac{K(\Sigma, n-q)\,K(\Sigma, n-q+2h+q_1)}{K(\Sigma, n-q+2h)\,K(\Sigma, n-q+q_1)}$$

$$=\prod_{i=1}^{p}\left\{\frac{\Gamma\left[\frac{1}{2}(n-q+1-i)+h\right]\Gamma\left[\frac{1}{2}(n-q+q_1+1-i)\right]}{\Gamma\left[\frac{1}{2}(n-q+1-i)\right]\Gamma\left[\frac{1}{2}(n-q+q_1+1-i)+h\right]}\right\}$$

$$=\prod_{i=1}^{p}\left\{\frac{\Gamma\left[\frac{1}{2}(n-q_1-q_2+1-i)+h\right]\Gamma\left[\frac{1}{2}(n-q_2+1-i)\right]}{\Gamma\left[\frac{1}{2}(n-q_1-q_2+1-i)\right]\Gamma\left[\frac{1}{2}(n-q_2+1-i)+h\right]}\right\}\tag{6•5}$$

詳察上式可知其爲 Beta 分布的動差相乘積。Beta 分布之機率密度函數爲

$$\beta\left(x:\tfrac{1}{2}a, \tfrac{1}{2}b\right)=\frac{\Gamma\left(\frac{a+b}{2}\right)}{\Gamma\left(\frac{a}{2}\right)\Gamma\left(\frac{b}{2}\right)}x^{\frac{a}{2}-1}(1-x)^{\frac{b}{2}-1}\qquad \text{if } 0<x<1$$

$$=0\qquad \text{if } x<0,\ x>1$$

x 之第 h 級動差爲

$$E(X^h)=\int_0^1\frac{\Gamma\left(\frac{a+b}{2}\right)}{\Gamma\left(\frac{a}{2}\right)\Gamma\left(\frac{b}{2}\right)}x^{\frac{a}{2}+h-1}(1-x)^{\frac{b}{2}-1}dx$$

$$=\frac{\Gamma\left(\frac{a+b}{2}\right)\Gamma\left(\frac{a}{2}+h\right)}{\Gamma\left(\frac{a+b}{2}+h\right)\Gamma\left(\frac{a}{2}\right)}$$

因此，從 (6•5) 式得知

$$E(U^h)=\prod_{i=1}^{p}E(x_i{}^h)$$

式中 x_i 之機率密度函數爲 $\beta\left[x_i:\frac{1}{2}(n-q+1-i),\frac{1}{2}q_1\right]$ 而 $x_1,\cdots\cdots,x_p$ 互相獨立，故

$$E(U^h)=\prod_{i=1}^{p}E(x_i{}^h)=E\{\prod_{i=1}^{p}x_i{}^h\}=E\{(\prod_{i=1}^{p}x_i)^h\}$$

玆將上面所得結果歸納於下列之定理：

【定理 6•1】 (6•3) 式 U 之分布爲 p 種互相獨立的 Beta 變值之聯合分布，即 $U=\prod_{i=1}^{p}x_i$，式中 x_i 之機率密度函數爲 $\beta\left[x_i:\frac{1}{2}(n-q+1-i),\ \frac{1}{2}q_1\right]$。當 $H_0:B_1=B_1{}^*$ 屬眞時，U 之第 h 級動差如 (6•5) 式所示。

因爲 U 之分布與 p，q_1，及 $n-q$ 有關，故今後在 U 上要附加脚註如 $U_{p,q_1,n-q}$。$n-q$ 爲 $n\hat{\Sigma}_\Omega{}^*$ 之分布 $W(\Sigma,n-q)$ 的自由度，q_1 爲 H_0 中 B_1 所含之列向量個數，p 爲分散量行列 Σ 之大小。

以上所論之 U 分布誘導方法係先求出其動差後，從動差判斷其分布。Anderson, T. W. (1958), Wilks, S. S. (1962), Kendall, M. G. and Stuart, A. (1966) 等都採用此法。Rao, C. R. (1965) 根據最小自乘法理論直接求出 U 之分布。下列之定理可顯示 $U_{p,q_1,n-q}$ 之分布特性。

【定理 6•2】 當 $H_0:B_1=B_1{}^*$ 屬眞時，$U_{p,q_1,n-q}$ 之分布與 $U_{q_1,p,n-q+q_1-p}$ 之分布相同。

證明：假定 $q_1<p$，則 (6•5) 式可寫成如下：

$$E\left(U^h_{p,q_1,n-q}\right)=\prod_{i=1}^{q_1}\frac{\Gamma\left[\frac{1}{2}(n-q_2+1-i)\right]}{\Gamma\left[\frac{1}{2}(n-q_2+1-i)+h\right]}\prod_{i=q_1+1}^{p}\frac{\Gamma\left[\frac{1}{2}(n-q_2+1-i)\right]}{\Gamma\left[\frac{1}{2}(n-q_2+1-i)+h\right]}$$

$$\cdot\prod_{i=1}^{p-q_1}\frac{\Gamma\left[\frac{1}{2}(n-q+1-i)+h\right]}{\Gamma\left[\frac{1}{2}(n-q+1-i)\right]}\prod_{i=p-q_1+1}^{p}\frac{\Gamma\left[\frac{1}{2}(n-q+1-i)+h\right]}{\Gamma\left[\frac{1}{2}(n-q+1-i)\right]}$$

$$=\prod_{i=1}^{q_1}\frac{\Gamma\left[\frac{1}{2}(n-q_2+1-i)\right]}{\Gamma\left[\frac{1}{2}(n-q_2+1-i)+h\right]}\left\{\prod_{i=1}^{p-q_1}\frac{\Gamma\left[\frac{1}{2}(n-q+1-i)\right]}{\Gamma\left[\frac{1}{2}(n-q+1-i)+h\right]}\right.$$

$$\left.\cdot\prod_{i=1}^{p-q_1}\frac{\Gamma\left[\frac{1}{2}(n-q+1-i)+h\right]}{\Gamma\left[\frac{1}{2}(n-q+1-i)\right]}\right\}\prod_{i=1}^{q_1}\frac{\Gamma\left[\frac{1}{2}(n-q_2-p+1-i)+h\right]}{\Gamma\left[\frac{1}{2}(n-q_2-p+1-i)\right]}$$

$$=\prod_{i=1}^{q_1}\left\{\frac{\Gamma\left[\frac{1}{2}(n-q_2+1-i)\right]\Gamma\left[\frac{1}{2}(n-q_2-p+1-i)+h\right]}{\Gamma\left[\frac{1}{2}(n-q_2+1-i)+h\right]\Gamma\left[\frac{1}{2}(n-q_2-p+1-i)\right]}\right\}$$

$$=E\left(U^h_{q_1,p,n-p-q_2}\right)$$

同理可證當 $q_1>p$ 時上式亦可成立。 Q. E. D.

Wilks, S. S. (1932), Nair, U. S. (1939), Anderson, T. W. (1958) 等利用變數轉換 (Transformation) 而發現在某些特殊場合下之分布爲 F 分布。玆將其整理如下表：

表 (6·1)：在特殊場合下之分布

p，q_1 值	F 值	F 之自由度
$p=1$，q_1 任何數	$\dfrac{1-U_{1,q_1,n-q}}{U_{1,q_1,n-q}}\cdot\dfrac{n-q}{q_1}$	q_1, $n-q$
$p=2$，q_1 任何數	$\dfrac{1-\sqrt{U_{2,q_1,n-q}}}{\sqrt{U_{2,q_1,n-q}}}\cdot\dfrac{n-q-1}{q_1}$	$2q_1$, $2(n-q-1)$
$q_1=1$，p 任何數	$\dfrac{1-U_{p,1,n-q}}{U_{p,1,n-q}}\cdot\dfrac{n-q-p+1}{p}$	p, $n-q-p+1$
$q_1=2$，p 任何數	$\dfrac{1-\sqrt{U_{p,2,n-q}}}{\sqrt{U_{p,2,n-q}}}\cdot\dfrac{n-q-p+1}{p}$	$2p$, $2(n-q-p+1)$

可見在上表諸特殊場合下，統計擬說 $H_0: B_1=B_1^*$ 之顯著性測驗可利用 Snedecor 氏 F 分布進行之。但在一般情形之下，$U_{p,q_1,n-q}$ 之累加分布函數 (Cumulative Distribution Function) 爲極複雜之函數，尙無積分表可供顯著性測驗之用，因此，需要考慮 $U_{p,q_1,n-q}$ 之近似分布。

(七) $U_{p,q_1,n-q}$ 之近似分布

Box, G. E. P. (1949) 曾經誘導一般化之近似分布理論，凡是動差 (Moments) 爲甘馬函數 (Gamma Function) 之逢機變數 (Random Variable) 皆可適用。其所討論之逢機變數爲 $W(0\leq W\leq 1)$，具有下列之第 h 級動差：

$$E(W^h) = K\left(\frac{\prod_{j=1}^{b} y_j^{\,y_j}}{\prod_{k=1}^{a} x_k^{\,x_k}}\right)\frac{\prod_{k=1}^{a}\Gamma\left[x_k(1+h)+d_k\right]}{\prod_{j=1}^{b}\Gamma\left[y_j(1+h)+e_j\right]} \tag{7·1}$$

式中 K 爲調整 $E(W^0)=1$ 之常數，又

$$\sum_{k=1}^{a} x_k = \sum_{j=1}^{b} y_j$$

當 $x_k=\frac{n}{2}=y_j$，$d_k=\frac{1}{2}(-q+1-k)$，$e_j=\frac{1}{2}(-q_2+1-j)$，$a=b=p$ 時，(7·1) 式便是之 $\lambda-U^{n/2}_{p,q_1,n-q}$ 之第 h 級動差。令

$$M = -2\log W \tag{7·2}$$

然而求 $\rho M(0\leq\rho<1)$ 之特徵函數 (Characteristic Function)

$$\phi(t)=E\left(e^{it\rho M}\right)=E\left(W^{-2it\rho}\right)$$

式中 ρ 爲待後要決定之任意常數。經過對數變換 $\log\phi(t)$ 後利用甘馬函數之展開公式而得到 $\phi(t)$ 之近似公式，然而根據下式獲得 M 之分布函數：

$$\begin{aligned}
G_M(M_0)&=P(M\leq M_0)=P(\rho M\leq\rho M_0)\\
&=P(\chi^2_f\leq\rho M_0)+\omega_1\left[P(\chi^2_{f+2}\leq\rho M_0)-P(\chi^2_f\leq\rho M_0)\right]\\
&\quad+\omega_2\left[P(\chi^2_{f+4}\leq\rho M_0)-P(\chi^2_f\leq\rho M_0)\right]\\
&\quad+\frac{\omega_1^2}{2}\left[P(\chi^2_{f+4}\leq\rho M_0)-2P(\chi^2_{f+2}\leq\rho M_0)\right.\\
&\quad\left.+P(\chi^2_f\leq\rho M_0)\right]+\text{etc.}
\end{aligned} \tag{7·3}$$

式中
$$f=-2\left[\sum_k d_k-\sum_j e_j-\frac{1}{2}(a-b)\right]$$
$$\omega_r=\frac{(-1)^{r+1}}{r(r+1)}\left\{\sum_k\frac{B_{r+1}(b_k+d_k)}{(\rho x_k)^r}-\sum_j\frac{B_{r+1}(c_j+e_j)}{(\rho y_j)^r}\right\}$$
$$b_k=(1-\rho)x_k,\qquad c_j=(1-\rho)y_j$$

$B_r(u)$ 爲 r 次巴奴里多項式 (Bernoulli Polynomial of Degree r)，例如
$$B_0(u)=1$$
$$B_1(u)=u-\frac{1}{2}$$
$$B_2(u)=u^2-u+\frac{1}{6}$$
$$B_3(u)=u^3-\frac{3}{2}u^2+\frac{1}{2}u$$

假如我們選擇使 (7·3) 式中之 $\omega_1=0$ 的 ρ 而僅採用第一項 $P(\chi^2_f\leq\rho M_0)$ 做爲 $G_M(M_0)$ 之近似值時，其誤差爲 θ^{-2} 級若 $x_k\geq p_k\theta$，$y_j\geq q_j\theta$（p_k，q_j 爲大於零之常數）。詳細情形可參閱 Box, G. E. P. (1949) 及 Anderson, T. W. (1958)。

玆利用上列之結果誘導第六節所討論的機率比値 λ 之 $-2\rho\log\lambda$ 的近似分布。令 $W=\lambda$，則其第 h 級動差爲
$$E(\lambda^h)=K\frac{\prod_{k=1}^{p}\Gamma\left[\frac{1}{2}(n-q+1-k+nh)\right]}{\prod_{j=1}^{p}\Gamma\left[\frac{1}{2}(n-q_2+1-j+nh)\right]}$$

令
$$a=b=p,\qquad x_k=\frac{n}{2},\qquad y_j=\frac{n}{2},\qquad d_k=\frac{1}{2}(-q+1-k)$$
$$e_j=\frac{1}{2}(-q_2+1-j),\qquad b_k=\frac{1}{2}(1-\rho)n,\qquad c_j=\frac{1}{2}(1-\rho)n$$

則
$$\begin{aligned}\omega_1&=\frac{1}{\rho n}\sum_{k=1}^{p}\Big[\left\{\frac{1}{2}[(1-\rho)n-q+1-k]\right\}^2-\frac{1}{2}[(1-\rho)n-q+1-k]\\&\qquad-\left\{\frac{1}{2}[(1-\rho)n-q_2+1-k]\right\}^2+\frac{1}{2}[(1-\rho)n-q_2+1-k]\Big]\\&=\frac{1}{\rho n}\sum_{k=1}^{p}\left\{\frac{-2[(1-\rho)n-q_2+1-k]q_1+q_1^2}{4}+\frac{q_1}{2}\right\}\\&=\frac{pq_1}{4\rho n}[-2(1-\rho)n+2q_2+p+q_1+1]\end{aligned}$$

爲了使 $\omega_1=0$ 起見，我們需要 ρ 等於下式：
$$\rho=\frac{1}{n}\left[n-q_2-\frac{1}{2}(p+q_1+1)\right]=\frac{1}{n}\left[n-q-\frac{1}{2}(p-q_1+1)\right]$$

令
$$m=\rho n=n-q-\frac{1}{2}(p-q_1+1)$$

則由 (7·3) 式可得：
$$\begin{aligned}P\left[-2\frac{m}{n}\log\lambda\leq z\right]&=P[-m\log U_{p,q_1,n-q}\leq z]\\&=P(\chi^2_{pq_1}\leq z)+\frac{\nu_2}{m^2}[P(\chi^2_{pq_1+4}\leq z)-P(\chi^2_{pq_1}\leq z)]\\&\quad+\frac{1}{m^4}[\nu_4\{P(\chi^2_{pq_1+8}\leq z)-P(\chi^2_{pq_1}\leq z)\}\\&\quad-\nu_2^2\{P(\chi^2_{pq_1+4}\leq z)-P(\chi^2_{pq_1}\leq z)\}]+\text{etc.}\end{aligned}\tag{7·4}$$

式中 $$\nu_2=\frac{1}{48}pq_1(p^2+q_1^2-5)$$

$$\nu_4=\frac{1}{2}\nu_2^2+\frac{pq_1}{1920}[3p^4+3q_1^4+10p^2q_1^2-50(p^2+q_1^2)+159]$$

我們若採用 (7•4) 式等號右邊之第一項時，其誤差爲 n^{-2} 級，採用至第二項時爲 n^{-4} 級，而採用至第三項時爲 n^{-6} 級。因此，當 n 適當大時，$-m\log U_{p,q_1,n-q}$ 之近似分布可視爲自由度等於 pq_1 之 χ^2 分布。Bartlett, M.S. (1947) 已提倡此說。

【定理 7•1】 當樣品大小 (Sample Size) n 適當大時，測驗統計擬說 $H_0: B_1=B_1^*$ 的統計值

$$-\left[(n-q)-\frac{1}{2}(p-q_1+1)\right]\log U_{p,q_1,n-q} \tag{7•5}$$

之近似分布爲自由度等於 pq_1 之 χ^2 分布，其誤差爲 n^{-2} 級。式中 $U_{p,q_1,n-q}=|\hat{\Sigma}_\Omega^*|/|\hat{\Sigma}_\omega^*|$。

當樣品大小 n 不大時，Box, G.E.P. (1949) 誘導近似 F 分布而得到較正確之結果。玆錄其結果如下：

【定理 7•2】 設測驗統計擬說 $H_0: B_1=B_1^*$ 之統計值爲

$$F_B=-\frac{1}{pq_1}\left[n-q_2-\frac{1}{2}(p+q_1+1)-(n-q_2)\frac{f_1}{f_2}\right]\log U_{p,q_1,n-q} \tag{7•6}$$

式中 $f_1=pq_1$， $f_2=\dfrac{12(n-q_1)^2(pq_1+2)}{p^2+q_1^2-5}$

則 F_B 之近似分布爲自由度等於 f_1 及 f_2 之 F 分布，其誤差爲 n^{-3} 級。

上列定理中之 $q_1>2$。因當 $q_1=1$ 及 $q_1=2$ 時，我們已在第六節表 (6•1) 中獲得測驗統計值之確實分布。

(八) 數個分散量行列之均等性測驗 (Test of Homoscedasticity of Several Dispersion Matrices)

令 $p\times1$ 向量 $\mathbf{y}_i^{(h)}(i=1,\cdots\cdots,n_h;\ h=1,\cdots\cdots,m)$ 爲第 h 族群 (Population) $N(B_h\mathbf{x}_i^{(h)},\Sigma_h)$ 之觀測值，B_h 爲 $p\times q$ 係數行列，$\mathbf{x}_i^{(h)}$ 爲 $q\times1$ 向量，Σ_h 爲 $p\times p$ 分散量行列。我們欲測驗的統計擬說是

$$H_1:\Sigma_1=\Sigma_2=\cdots\cdots=\Sigma_m$$

我們仍採用機率比值軌範 (Likelihood Ratio Criterion) 以測驗 H_1。在未知常數空間 (Parameter Space) Ω 中 Σ_h 爲正值行列 (Positive Definite Matrix) 而 B_h 爲任何係數行列，在副空間 (Subspace) ω 中 $\Sigma_1=\cdots\cdots=\Sigma_m$ 而 B_h 爲任何係數行列。機率函數如下：

$$L=\prod_{h=1}^{m}\frac{|\Sigma_h^{-1}|^{\frac{1}{2}n_h}}{(2\pi)^{\frac{1}{2}pn_h}}\exp\left[-\frac{1}{2}\sum_{i=1}^{n_h}(\mathbf{y}_i^{(h)}-B_h\mathbf{x}_i^{(h)})'\Sigma_h^{-1}(\mathbf{y}_i^{(h)}-B_h\mathbf{x}_i^{(h)})\right]$$

從第三節得知 Σ_h 及 B_h 之最大機率推算式爲

$$\hat{B}_h=Y_hX_h'(X_hX_h')^{-1} \tag{8•1}$$

$$\hat{\Sigma}_h^*=\frac{1}{n_h}\sum_{i=1}^{n_h}(\mathbf{y}_i^{(h)}-\hat{B}_h\mathbf{x}_i^{(h)})(\mathbf{y}_i^{(h)}-\hat{B}_h\mathbf{x}_i^{(h)})'$$

$$=\frac{1}{n_h}[Y_hY_h'-\hat{B}_h(X_hX_h')\hat{B}_h'] \tag{8•2}$$

式中 $Y_h=(\mathbf{y}_1^{(h)},\cdots\cdots,\mathbf{y}_{n_h}^{(h)})$， $X_h=(\mathbf{x}_1^{(h)},\cdots\cdots,\mathbf{x}_{n_h}^{(h)})$

因此，機率函數之最大值爲

$$L(\hat{\Omega})=(2\pi)^{-\frac{1}{2}np}\left\{\prod_{h=1}^{m}|\hat{\Sigma}_h^*|^{-\frac{1}{2}n_h}\right\}e^{-\frac{1}{2}np} \tag{8•3}$$

式中　　$n=\sum_{h=1}^{m} n_h$

在 ω 中 B_h 之最大機率推算式仍與 (8·1) 式相同。當 $\Sigma_1=\Sigma_2=\cdots\cdots=\Sigma_m=\Sigma$ 時，機率函數爲

$$(2\pi)^{-\frac{1}{2}np}|\Sigma|^{-\frac{n}{2}}\exp\left[-\frac{1}{2}\sum_{h=1}^{m}\sum_{i=1}^{n_h}(y_i^{(h)}-\hat{B}_h x_i^{(h)})'\Sigma^{-1}(y_i^{(h)}-\hat{B}_h x_i^{(h)})\right]$$

從第三節之結果得 Σ 之最大機率推算式如下：

$$\hat{\Sigma}^*=\frac{1}{n}\sum_{h=1}^{m}\sum_{i=1}^{n_h}(y_i^{(h)}-\hat{B}_h x_i^{(h)})(y_i^{(h)}-\hat{B}_h x_i^{(h)})'$$

$$=\frac{1}{n}\sum_{h=1}^{m} n_h \hat{\Sigma}_h^* \tag{8·4}$$

然而，機率函數之最大值爲

$$L(\hat{\omega})=(2\pi)^{-\frac{1}{2}np}|\hat{\Sigma}^*|^{-\frac{n}{2}}e^{-\frac{1}{2}np} \tag{8·5}$$

因此，測驗 H_1 之機率比値軌範爲

$$\lambda_1=\left\{\prod_{h=1}^{m}|\hat{\Sigma}_h^*|^{\frac{n_h}{2}}\right\}\Big/|\hat{\Sigma}^*|^{\frac{n}{2}}$$

$$=\frac{\prod_{h=1}^{m}|n_h\hat{\Sigma}_h^*|^{\frac{1}{2}n_h}}{|n\hat{\Sigma}^*|^{\frac{1}{2}n}}\cdot\frac{n^{\frac{1}{2}np}}{\prod_{h=1}^{m}n_h^{\frac{1}{2}n_h p}} \tag{8·6}$$

令　　$S_h=n_h\hat{\Sigma}_h^*$，　　$S=\sum_{h=1}^{m}S_h$

則 S_h 之自由度爲 $f_h=n_h-q$ 而 S 之自由度爲 $f_0=\sum_{h=1}^{m}f_h=n-mq$。然而 (8·6) 式之 λ_1 可用下式取而代之：

$$W_1=\frac{\prod_{h=1}^{m}|S_h|^{\frac{1}{2}f_h}}{|S|^{\frac{1}{2}f_0}}\cdot\frac{f_0^{\frac{1}{2}pf_0}}{\prod_{h=1}^{m}f_h^{\frac{1}{2}pf_h}} \tag{8·7}$$

其第 r 級動差 (The r^{th} moment) 爲 (參閱 Anderson, T. W. 1958)

$$E(W_1^r)=K\left(\frac{\prod_{j=1}^{p}\left(\frac{1}{2}f_0\right)^{\frac{1}{2}f_0}}{\prod_{h=1}^{m}\prod_{i=1}^{p}\left(\frac{1}{2}f_h\right)^{\frac{1}{2}f_h}}\right)^r\frac{\prod_{h=1}^{m}\prod_{i=1}^{p}\Gamma\left[\frac{1}{2}f_h(1+r)+\frac{1}{2}(1-i)\right]}{\prod_{j=1}^{p}\Gamma\left[\frac{1}{2}f_0(1+r)+\frac{1}{2}(1-j)\right]}$$

查上式與 (7·1) 式同型，兹令

$b=p$，　$y_j=\frac{1}{2}f_0$，　$e_j=\frac{1}{2}(1-j)$，　$j=1,\cdots\cdots,p$

$a=mp$，　$x_k=\frac{1}{2}f_h$，　$d_k=\frac{1}{2}(1-i)$，　$i=1,\cdots\cdots,p$；$h=1,\cdots\cdots,m$

則從 (7·3) 式可得

$$f=-2\left[\sum_k d_k-\sum_j e_j-\frac{1}{2}(a-b)\right]$$

$$=-\left[m\sum_{i=1}^{p}(1-i)-\sum_{j=1}^{p}(1-j)-(mp-p)\right]$$

$$=-\left[-\frac{1}{2}mp(p-1)+\frac{1}{2}p(p-1)-(m-1)p\right]$$

$$=\frac{1}{2}(m-1)p(p+1) \tag{8•8}$$

再令 $b_k=\frac{1}{2}(1-\rho)f_h$，$c_j=\frac{1}{2}(1-\rho)f_0$，則可得使 ω_1 等於零之 ρ 如下：

$$\rho=1-\left(\sum_{h=1}^{m}\frac{1}{f_h}-\frac{1}{f_0}\right)\frac{2p^2+3p-1}{6(m-1)(p+1)} \tag{8•9}$$

又

$$\omega_2=\frac{p(p+1)}{48\rho^2}\left[(p-1)(p+2)\left(\sum_{h=1}^{m}\frac{1}{f_h}-\frac{1}{f_0}\right)-6(m-1)(1-\rho)^2\right] \tag{8•10}$$

因此，$-2\rho\log W_1$ 之累加分布函數 (Cumulavetive Distribution Function) 如下：

$$P\{-2\rho\log W_1\leqq z\}=P(\chi^2_f\leqq z)+\omega_2\left[P(\chi^2_{f+4}\leqq z)-P(\chi^2_f\leqq z)\right]+0(f_0^{-3}) \tag{8•11}$$

我們若僅採用上式等號右邊第一項時，其誤差爲 f_0^{-2} 級，簡寫爲 $0(f_0^{-2})$。茲將上面所得結果歸納於下列之定理：

【定理 8•1】 統計擬說 $H_1:\sum_1=\cdots\cdots=\sum_m$ 之測驗統計值

$$-2\left[1-\left(\sum_{h=1}^{m}\frac{1}{f_h}-\frac{1}{f_0}\right)\frac{2p^2+3p-1}{6(m-1)(p+1)}\right]\log\frac{\prod_{h=1}^{m}|S_h|^{\frac{1}{2}f_h}}{|S|^{\frac{1}{2}f_0}}\cdot\frac{f_0^{\frac{1}{2}pf_0}}{\prod_{h=1}^{m}f_h^{\frac{1}{2}pf_h}}$$

之近似分布爲自由度等於 $f=\frac{1}{2}(m-1)p(p+1)$ 之 χ^2 分布，其誤差爲 f_0^{-2} 級，若照 (8•11) 式計算累加分布函數值時，則其誤差可減小至 f_0^{-3} 級。式中 $S_h=n_h\sum_h^*$，$S=\sum_{h=1}^{m}S_h$，$f_h=n_h-q$ 爲 S_h 之自由度，$f_0=\sum_{h=1}^{m}f_h=n-mq$。

Box, G. E. P. (1949) 誘導近似 F 分布而得到較正確之結果。茲錄其結果於下：

【定理 8•2】 統計擬說 $H_1:\sum_1=\cdots\cdots=\sum_m$ 之測驗統計值

$$-\frac{2}{df_1}\left(\rho-\frac{df_1}{df_2}\right)\log W_1$$

之近似分布爲自由度等於 df_1 及 df_2 之 F 分布，其誤差爲 f_0^{-3} 級。式中 ρ 如 (8•9) 式，W_1 如 (8•7) 式，

$$df_1=f=\frac{1}{2}(m-1)p(p+1)$$

$$df_2=(df_1+2)/\left[\theta-(1-\rho)^2\right]$$

$$\theta=\frac{(p-1)(p+2)}{6(m-1)}\left(\sum_{h=1}^{m}\frac{1}{f_h^2}-\frac{1}{f_0^2}\right)$$

(九) 數個複變值直線廻歸之常態分布的一致性測驗 (Testing the Hypothesis that Several Normal Distributions of the Multivariate Linear Regression are Identical)

設從第 h 族群 $N(B_h x_i^{(h)},\sum_h)$ 逢機抽出 $p\times1$ 向量 $y_i^{(h)}$ $(i=1,\cdots\cdots,n_h;\ h=1,\cdots\cdots,m)$，$B_h$ 爲 $p\times q$ 係數行列，$x_i^{(h)}$ 爲 $q\times1$ 向量，$\sum_h$ 爲 $p\times p$ 分散量行列。我們欲測驗的統計擬說是

$$H_2 : B_1 = B_2 = \cdots\cdots = B_m,\ \Sigma_1 = \Sigma_2 = \cdots\cdots = \Sigma_m$$

我們仍應用機率比值軌範 (Likelihood Ratio Criterion) 來測驗 H_2 之顯著性。在未知常數空間 (Parameter Space) Ω 中之機率函數 (Likelihood Function) 如下：

$$L(\Omega) = \prod_{h=1}^{m} \frac{|\Sigma_h|^{-\frac{1}{2}n_h}}{(2\pi)^{\frac{1}{2}pn_h}} \exp\left[-\frac{1}{2}\sum_{i=1}^{n_h}(\mathbf{y}_i^{(h)} - B_h\mathbf{x}_i^{(h)})'\Sigma_h^{-1}(\mathbf{y}_i^{(h)} - B_h\mathbf{x}_i^{(h)})\right]$$

倣照第三節之步驟可得 B_h 及 Σ_h 之最大機率推算式如次：

$$\hat{B}_h = Y_hX_h'(X_hX_h')^{-1} \tag{9•1}$$

$$\hat{\Sigma}_h^* = \frac{1}{n_h}\sum_{i=1}^{n_h}(\mathbf{y}_i^{(h)} - \hat{B}_h\mathbf{x}_i^{(h)})(\mathbf{y}_i^{(h)} - \hat{B}_h\mathbf{x}_i^{(h)})'$$

$$= \frac{1}{n_h}[Y_hY_h' - \hat{B}_hX_hX_h'\hat{B}_h'] \tag{9•2}$$

式中 $Y_h = (\mathbf{y}_1^{(h)}, \cdots\cdots, \mathbf{y}_{n_h}^{(h)})$， $X_h = (\mathbf{x}_1^{(h)}, \cdots\cdots, \mathbf{x}_{n_h}^{(h)})$

因此，機率函數之最大值爲

$$L(\hat{\Omega}) = (2\pi)^{-\frac{1}{2}np} \prod_{h=1}^{m} |\hat{\Sigma}_h^*|^{-\frac{1}{2}n_h} e^{-\frac{1}{2}np} \tag{9•3}$$

式中 $n = \sum_{h=1}^{m} n_h$

在 ω 中，即當 $B_1 = \cdots\cdots = B_m = B$ 及 $\Sigma_1 = \cdots\cdots = \Sigma_m = \Sigma$ 時之機率函數爲

$$L(\omega) = \frac{|\Sigma|^{-\frac{n}{2}}}{(2\pi)^{\frac{1}{2}np}} \exp\left[-\frac{1}{2}\sum_{h=1}^{m}\sum_{i=1}^{n_h}(\mathbf{y}_i^{(h)} - B\mathbf{x}_i^{(h)})'\Sigma^{-1}(\mathbf{y}_i^{(h)} - B\mathbf{x}_i^{(h)})\right]$$

應用第三節之結果可得 B 及 Σ 之最大機率推算式如下：

$$\hat{B} = \sum_{h=1}^{m} Y_hX_h'\left(\sum^{m} X_hX_h'\right)^{-1} \tag{9•4}$$

$$\hat{\Sigma}^* = \frac{1}{n}\sum_{h=1}^{m}\sum_{i=1}^{n_h}(\mathbf{y}_i^{(h)} - \hat{B}\mathbf{x}_i^{(h)})(\mathbf{y}_i^{(h)} - \hat{B}\mathbf{x}_i^{(h)})'$$

$$= \frac{1}{n}\left[\sum_{h=1}^{n} n_h\hat{\Sigma}_h^* + \sum_{h=1}^{m}(\hat{B}_h - \hat{B})X_hX_h'(\hat{B}_h - \hat{B})'\right] \tag{9•5}$$

因此，機率函數之最大值爲

$$L(\hat{\omega}) = (2\pi)^{-\frac{1}{2}np}|\hat{\Sigma}^*|^{-\frac{n}{2}}e^{-\frac{1}{2}np} \tag{9•6}$$

於是，測驗 H_2 之機率比值軌範爲

$$\lambda_2 = \left[\prod_{h=1}^{m}|\hat{\Sigma}_h^*|^{\frac{1}{2}n_h}\right] \Big/ |\hat{\Sigma}^*|^{\frac{n}{2}}$$

$$= \frac{\prod^{m}|n_h\hat{\Sigma}_h^*|^{\frac{1}{2}n_h}}{|n\hat{\Sigma}^*|^{\frac{1}{2}n}} \cdot \frac{n^{\frac{1}{2}np}}{\prod_{h=1}^{m} n_h^{\frac{1}{2}pn_h}} \tag{9•7}$$

令 $S_h=n_h\hat{\Sigma}_h^*$，$S=\sum_{h=1}^{m}S_h$，$R=\sum_{h=1}^{m}(\hat{B}_h-\hat{B})X_hX_h'(\hat{B}_h-\hat{B})'$，則 λ_2 可用下式取而代之：

$$W_2=\frac{\prod_{h=1}^{m}|S_h|^{\frac{1}{2}f_h}}{|S+R|^{\frac{1}{2}f_0}}\cdot\frac{f_0^{\frac{1}{2}pf_0}}{\prod_{h=1}^{m}f_h^{\frac{1}{2}pf_h}} \tag{9.8}$$

式中 f_h 為 S_h 之自由度，$f_0=\sum_{h=1}^{m}f_h=n-mq$ 為 S 之自由度。

在 H_2 下，S 之分布為 $W(\Sigma,f_0)$。根據定理 4.1 得知 S 與 $\hat{B}_h$ 及 $\hat{B}$ 互相獨立。R 可寫成如次：

$$\begin{aligned}R&=\sum_{h=1}^{m}[(\hat{B}_h-B)-(\hat{B}-B)]X_hX_h'[(\hat{B}_h-B)-(\hat{B}-B)]'\\&=\sum_{h=1}^{m}(\hat{B}_h-B)X_hX_h'(\hat{B}_h-B)'-(\hat{B}-B)\Big(\sum_{h=1}^{m}X_hX_h'\Big)(\hat{B}-B)'\end{aligned}$$

於 H_2 屬眞之下，根據定理 4.1 上式等號右邊第一項之分布為 $\sum_{i=1}^{mq}z_iz_i'$ 而第二項之分布為 $\sum_{i=mq+1}^{(m+1)q}z_iz_i'$，式中 $z_i's$ 互相獨立，其分布為 $N(0,\Sigma)$。因此，R 之分布為 $W(\Sigma,(m-1)q)$。倣照第六節求 $E(U^h)$ 之方法，我們可得 W_2 之第 r 級動差如下：

$$E(W_2^r)=K\left[\frac{\prod_{j=1}^{p}\left(\frac{1}{2}f_0\right)^{\frac{1}{2}f_0}}{\prod_{h=1}^{m}\prod_{i=1}^{p}\left(\frac{1}{2}f_h\right)^{\frac{1}{2}f_h}}\right]^r\frac{\prod_{h=1}^{m}\prod_{i=1}^{p}\Gamma\left[\frac{1}{2}f_h(1+r)+\frac{1}{2}(1-i)\right]}{\prod_{j=1}^{p}\Gamma\left[\frac{1}{2}f_0(1+r)+\frac{1}{2}\{(m-1)q+1-j\}\right]} \tag{9.9}$$

茲令　$b=p$，　$y_j=\frac{1}{2}f_0$，　$e_j=\frac{1}{2}\{(m-1)q+1-j\}$，　$j=1,\cdots\cdots,p$

$a=mp$，　$x_k=\frac{1}{2}f_h$，　$d_k=\frac{1}{2}(1-i)$，　$i=1,\cdots\cdots,p;\ h=1,\cdots\cdots,m$

則可應用 (7.3) 式求出 $-2\rho\log W_2$ 之近似累加分布函數。

$$\begin{aligned}f&=-2\Big[\sum_k d_k-\sum_j e_j-\frac{1}{2}(a-b)\Big]\\&=-\Big[m\sum_{i=1}^{p}(-i)-\sum_{j=1}^{p}\{(m-1)q+1-j\}-(mp-p)\Big]\\&=\frac{1}{2}(m-1)p(p+2q+1)\end{aligned} \tag{9.10}$$

在 (7.3) 式中之 ω_1 可簡化如下：

$$\omega_1=\frac{1}{2\rho}\left(-(1-\rho)f+\sum_k\frac{d_k^2-d_k+\frac{1}{6}}{x_k}-\sum_j\frac{e_j^2-e_j+\frac{1}{6}}{y_j}\right) \tag{9.11}$$

於是，本測驗之 ω_1 如下：

$$\begin{aligned}\omega_1&=\frac{1}{2\rho}\Big[-(1-\rho)f+\Big(\sum_{h=1}^{m}\frac{1}{f_h}-\frac{1}{f_0}\Big)\sum_{i=1}^{p}\Big\{\frac{1}{2}(1-i)^2-(1-i)+\frac{1}{3}\Big\}\\&\qquad-\frac{1}{f_0}\sum_{j=1}^{p}\Big\{\frac{1}{2}(m-1)^2q^2+(m-1)q(1-i)+(m-1)q\Big\}\Big]\\&=\frac{1}{2\rho}\Big[-(1-\rho)f+\Big(\sum_{h=1}^{m}\frac{1}{f_h}-\frac{1}{f_0}\Big)\frac{p(2p^2+3p-1)}{12}\\&\qquad+\frac{1}{2f_0}(m-1)pq(p-(m-1)q+1)\Big]\end{aligned}$$

令 $\omega_1=0$，則得

$$\rho=1-\left(\sum_{h=1}^{m}\frac{1}{f_h}-\frac{1}{f_0}\right)\frac{2p^2+3p-1}{6(m-1)(p+2q+1)}-\frac{q[p-(m-1)q+1]}{f_0(p+2q+1)} \tag{9.12}$$

當 $q=1$ 時，上式便成

$$\rho=1-\left(\sum_{h=1}^{m}\frac{1}{f_h}-\frac{1}{f_0}\right)\frac{2p^2+3p-1}{6(m-1)(p+3)}-\frac{p-m+2}{f_0(p+3)} \tag{9.13}$$

此式與 Anderson, T. W. (1958) 所導出結果相吻合。現在我們可得 $-2\rho\log W_2$ 之近似累加分布函數如下：

$$P(-2\rho\log W_2\leq z)=P(\chi^2_f\leq z)+O(f_0^{-2}) \tag{9.14}$$

當 f_0 相當大時，$-2\rho\log W_2$ 之近似分布爲自由度等於 (9•10) 式之 f 的 χ^2 分布。玆將上列所得結果歸納於下列之定理：

【定理 9•1】 設有 m 個族群之分布爲 $N(B_h x_i^{(h)}, \Sigma_h)$。我們從各族群分別獨立地逢機抽出樣品 $y_i^{(h)}(i=1,\cdots\cdots,n_h;\ h=1,\cdots\cdots,m)$ 以測驗下列之統計擬說：

$$H_1: B_1=\cdots\cdots=B_m,\ \Sigma_1=\cdots\cdots=\Sigma_m$$

按照機率比值軌範，我們得到下列之測驗統計值

$$-2\rho\log\frac{\prod_{h=1}^{m}|n_h\Sigma_h^*|^{\frac{1}{2}f_h}}{|n\Sigma^*|^{\frac{1}{2}f_0}}\cdot\frac{f_0^{\frac{1}{2}pf_0}}{\prod_{h=1}^{m}f_h^{\frac{1}{2}pf_h}} \tag{9.15}$$

其近似分布爲自由度等於 $f=\frac{1}{2}(m-1)p(p+2q+1)$ 之 χ^2 分布，而其誤差爲 f_0^{-2} 級。式中 ρ 如 (9•12) 式。

（十） 水稻品種地域適應性分析 (Analysis of Rice Varietal Adaptation)

搜集農業試驗所，臺中、嘉義、及高雄農業改良場過去十五年間所舉行之水稻豐歉試驗結果記錄資料及歷年氣象因子觀測記錄資料做爲本研究之資料。各試驗場每年第一及第二期作都有蓬萊稻及在萊稻各兩品種參加試驗。除了蓬萊稻臺中 65 號品種在各試驗場均有參加試驗外，各試驗場採用當地所推廣之品種參加試驗。每一期作之各品種農藝性狀僅有其觀測平均值可供分析之用。本研究採用之農藝性狀有株高 (y_1)、穗數 (y_2)、及穀產量 (y_3) 三種。因爲每一品種在每一期作內僅有此三種性狀之三個觀測平均值的緣故，氣象因素亦採用自水稻挿秧至收穫之平均值與其相配。本研究所採用之氣象因素有溫度 (x_1)、日照 (x_2)、濕度 (x_3)、溫差 (x_4)、及雨量 (x_5) 等五種。玆擬僅就臺中 65 號一品種加以分析之。

（1）臺中 65 號第一期作在四個不同地域之農藝性狀的差異顯著性測驗

在臺北、臺中、嘉義、及屏東四個地域第一期作都有栽培臺中 65 號，連續栽培十五年。玆擬先探討臺中 65 號之農藝性狀是否因地域而有顯著之差異。以 $y_i^{(h)}(h=1,\cdots\cdots,m;\ i=1,\cdots\cdots,n)$ 代表第 h 地域第 i 年之 $p\times 1$ 農藝性狀觀測值向量，其第一元素爲株高，第二元素爲穗數，第三元素爲穀產量。以 μ_h 代表 $y_i^{(h)}$ 之期望值，以 Σ_h 代表 $p\times p$ 分散量行列 $(p=3)$，在常態分布之前提下，其複變值機率密度函數可令 (2•3) 式中之 Bx 等於 μ_h 而得之如下：

$$f(y_i^{(h)})=(2\pi)^{-\frac{p}{2}}|\Sigma_h|^{-\frac{1}{2}}\exp\left[-\frac{1}{2}(y_i^{(h)}-\mu_h)'\Sigma_h^{-1}(y_i^{(h)}-\mu_h)\right]$$

首先我們欲測驗的統計擬說是各地域分散量之均等性，卽

$$H_1: \Sigma_1=\cdots\cdots=\Sigma_m$$

在未知常數空間 Ω 中 Σ_h 爲正値行列而 μ_h 爲任何向量。在 Ω 中之機率函數爲

$$L(\Omega)=\prod_{h=1}^{m}\frac{|\Sigma_h|^{-\frac{n}{2}}}{(2\pi)^{np/2}}\exp\left[-\frac{1}{2}\sum_{h=1}^{n}(\mathbf{y}_i^{(h)}-\boldsymbol{\mu}_h)'\Sigma_h^{-1}(\mathbf{y}_i^{(h)}-\boldsymbol{\mu}_h)\right] \tag{10·1}$$

倣照第三節可得 $\boldsymbol{\mu}_h$ 及 Σ_h 之最大機率推算式如下：

$$\hat{\boldsymbol{\mu}}_h=\bar{\mathbf{y}}_h=\frac{1}{n}\sum_{i=1}^{n}\mathbf{y}_i^{(h)} \tag{10·2}$$

$$\hat{\Sigma}_h^*=\frac{1}{n}\sum_{i=1}^{n}(\mathbf{y}_i^{(h)}-\bar{\mathbf{y}}_h)(\mathbf{y}_i^{(h)}-\bar{\mathbf{y}}_h)' \tag{10·3}$$

即 $\hat{\boldsymbol{\mu}}_h$ 爲各地域內三種農藝性狀十五年觀測值之平均而 $n\hat{\Sigma}_h^*$ 爲地域內三種農藝性狀之平方和及乘積和 (Sum of Squares and Products) 之行列。因此，機率函數 $L(\Omega)$ 之最大值如下：

$$L(\hat{\Omega})=(2\pi)^{-\frac{1}{2}mnp}\left[\prod_{h=1}^{m}|\hat{\Sigma}_h^*|^{-\frac{n}{2}}\right]e^{-\frac{1}{2}mnp} \tag{10·4}$$

玆將四個地域按照 (10·2) 及 (10·3) 兩式求得之最大機率推算值列如次：

	$\hat{\boldsymbol{\mu}}_h$	$n\hat{\Sigma}_h^*$		
臺 北：	103.7600	192.076000	27.522000	−0.375720
	20.1867		44.917333	−2.487840
	4.2118			1.472722
臺 中：	104.9933	500.489333	83.660000	5.405267
	16.8000		40.840000	5.500000
	5.3927			2.035093
嘉 義：	102.5933	665.429333	−11.475333	42.374352
	16.3467		108.457333	−1.763934
	3.0180			5.312011
屏 東：	102.8733	313.709333	35.394000	15.182987
	16.3600		19.336000	2.085680
	5.5605			4.352856

在未知常數副空間 ω 中，$\Sigma_1=\cdots\cdots=\Sigma_m=\Sigma$ 而 $\boldsymbol{\mu}_h$ 爲任何向量。其機率函數爲

$$L(\omega)=(2\pi)^{-\frac{1}{2}mnp}|\Sigma|^{-\frac{1}{2}mn}\exp\left[-\frac{1}{2}\sum_{h=1}^{m}\sum_{i=1}^{n}(\mathbf{y}_i^{(h)}-\boldsymbol{\mu}_h)'\Sigma^{-1}(\mathbf{y}_i^{(h)}-\boldsymbol{\mu}_h)\right] \tag{10·5}$$

式中 $\boldsymbol{\mu}_h$ 之最大機率推算式仍爲 $\hat{\boldsymbol{\mu}}_h=\bar{\mathbf{y}}_h$，而 Σ 之推算式如下：

$$\hat{\Sigma}^*=\frac{1}{mn}\sum_{h=1}^{m}\sum_{i=1}^{n}(\mathbf{y}_i^{(h)}-\bar{\mathbf{y}}_h)(\mathbf{y}_i^{(h)}-\bar{\mathbf{y}}_h)'=\frac{1}{mn}\sum_{h=1}^{m}(n\hat{\Sigma}_h^*) \tag{10·6}$$

即 $mn\hat{\Sigma}^*$ 爲四個地域內平方和及乘積和行列之總計行列，計算所得結果如下：

$$mn\hat{\Sigma}^*=\begin{bmatrix}1671.703999 & 135.100667 & 62.586886\\ & 213.550666 & 3.333906\\ & & 13.172682\end{bmatrix}$$

因此，機率函數 $L(\omega)$ 之最大值如下：

$$L(\hat{\omega})=(2\pi)^{-\frac{1}{2}mnp}|\hat{\Sigma}^*|^{-\frac{1}{2}mn}e^{-\frac{1}{2}mnp} \tag{10·7}$$

故得機率比值軌範如下：

$$\lambda_1=\frac{\prod_{h=1}^{m}|\hat{\Sigma}_h^*|^{-\frac{1}{2}n}}{|\hat{\Sigma}^*|^{\frac{1}{2}mn}}=\frac{\prod_{h=1}^{m}|n\hat{\Sigma}_h^*|^{-\frac{1}{2}n}}{|mn\hat{\Sigma}^*|^{\frac{1}{2}mn}}\cdot\frac{(mn)^{\frac{1}{2}mnp}}{\prod_{h=1}^{m}n^{\frac{1}{2}np}}$$

令 $S_h=n\hat{\Sigma}_h^*$，$S=\sum_{h=1}^{m}S_h$，則 S_h 之自由度 $f_h=n-1$ 而 S 之自由度 $f_o=\sum_{h=1}^{m}f_h=m(n-1)$。然而上式可用下式取而代之：

$$W_1=\frac{\prod_{h=1}^{m}|S_h|^{\frac{1}{2}(n-1)}}{|S|^{\frac{1}{2}m(n-1)}}\cdot\frac{[m(n-1)]^{\frac{1}{2}m(n-1)p}}{\prod_{h=1}^{m}(n-1)^{\frac{1}{2}(n-1)p}} \tag{10.8}$$

我們可利用定理 8•1 以測驗 H_1 之顯著性。由 (8•9) 式得

$$\rho=1-\left(\sum_{h=1}^{m}\frac{1}{f_h}-\frac{1}{f_o}\right)\frac{2p^2+3p-1}{6(m-1)(p+1)}=1-\frac{(m+1)(2p^2+3p-1)}{6m(n-1)(p+1)}$$

由 (8•12) 式得測驗統計值如下：

$$M_1=-2\rho\log W_1-=2\left[1-\frac{(m+1)(2p^2+3p-1)}{6m(n-1)(p+1)}\right]\log W_1 \tag{10.9}$$

其近似分布爲自由度等於 $f=\frac{1}{2}(m-1)p(p+1)$ 之 χ^2 分布。由實際資料求得之 M_1 值爲 37.7037。根據自由度 $f=18$ 查得 $\chi^2_{.01}=34.805$。因此，我們丟棄 H_1，即四個地域之分散量不具有均等性。

其次，我們欲測驗的統計擬說是四個地域之複變值常態分布互相一致，即

$$H_2:\ \mu_1=\cdots\cdots=\mu_m,\ \Sigma_1=\cdots\cdots=\Sigma_m$$

從 H_1 之測驗結果已知四個地域之分散量不相等，足可說明四地域之複變值常態分布並非一致，無需再做 H_2 之測驗，但爲了要說明 H_2 之測驗方法起見，仍要測驗之，我們可預知 H_2 之測驗結果必定否認四個地域之複變值常態分布之一致性。在 Ω 中之機率函數 $L(\Omega)$ 與 (10•1) 式相同，因此，μ_h 及 Σ_h 之最大機率推算式分別如 (10•2) 及 (10•3) 兩式。在 ω 即 $\mu_1=\cdots\cdots=\mu_m=\mu$ 及 $\Sigma_1=\cdots\cdots=\Sigma_m=\Sigma$ 之機率函數如次：

$$L(\omega)=(2\pi)^{-\frac{1}{2}mnp}|\Sigma|^{-\frac{1}{2}mn}\exp\left[-\frac{1}{2}\sum_{h=1}^{m}\sum_{i=1}^{n}(\mathbf{y}_i^{(h)}-\mu)'\Sigma^{-1}(\mathbf{y}_i^{(h)}-\mu)\right] \tag{10.10}$$

式中 μ 及 Σ 之最大機率推算式分別如下：

$$\hat{\mu}=\frac{1}{mn}\sum_{h=1}^{m}\sum_{i=1}^{n}\mathbf{y}_i^{(h)}=\frac{1}{m}\sum_{h=1}^{m}\bar{\mathbf{y}}_h=\bar{\mathbf{y}} \tag{10.11}$$

$$\hat{\Sigma}^*=\frac{1}{mn}\sum_{h=1}^{m}\sum_{i=1}^{n}(\mathbf{y}_i^{(h)}-\bar{\mathbf{y}})(\mathbf{y}_i^{(h)}-\bar{\mathbf{y}})' \tag{10.12}$$

上兩式之計算方法係將四個地域合併成一資料後計算各農藝性狀之總平均值，總平方和及總乘積和。茲將計算結果列如下：

$$\hat{\mu}=\begin{pmatrix}103.5550\\17.4233\\4.5458\end{pmatrix}\qquad mn\hat{\Sigma}^*=\begin{pmatrix}1724.208499 & 156.552967 & 78.511360\\ & 368.267366 & -24.474314\\ & & 43.922096\end{pmatrix}$$

將上列兩推算式代入 (10•10) 式則得機率函數之最大值如次：

$$L(\hat{\omega})=(2\pi)^{-\frac{1}{2}mnp}|\hat{\Sigma}^*|^{-\frac{1}{2}mn}e^{-\frac{1}{2}mnp} \tag{10.13}$$

由 (10•4) 及 (10•13) 兩式得機率比值如下：

$$\lambda_2=\frac{\prod_{h=1}^{m}|\hat{\Sigma}_h^*|^{\frac{1}{2}n}}{|\hat{\Sigma}^*|^{\frac{1}{2}mn}}=\frac{\prod_{h=1}^{m}|n\hat{\Sigma}_h^*|^{\frac{1}{2}n}}{|mn\hat{\Sigma}^*|^{\frac{1}{2}mn}}\cdot\frac{(mn)^{\frac{1}{2}mnp}}{\prod_{h=1}^{m}n^{\frac{1}{2}np}}$$

令 $n\Sigma_h^*=S_h$，$\sum_{h=1}^{m}S_h=S$，$R=n\sum_{h=1}^{m}(\bar{y}_h-\bar{y})(\bar{y}_h-\bar{y})'$ 及 $T=mn\Sigma^*$，則我們不難證明 $T=S+R$。就變方分析原理而言，則 S 爲組內平方和乘積和之行列，R 爲組間平方和乘積和之行列，而 T 爲總平方和總乘積和之行列。令 $f_h=n-1$ 爲 S_h 之自由度而 $f_o=\sum_{h=1}^{m}f_h=m(n-1)$ 爲 S 之自由度。現在 λ_2 可用下式取而代之：

$$W_2=\frac{\prod_{h=1}^{m}|S_h|^{\frac{1}{2}(n-1)}}{|S+R|^{\frac{1}{2}m(n-1)}}\cdot\frac{[m(n-1)]^{\frac{1}{2}m(n-1)p}}{\prod_{h=1}^{m}(n-1)^{\frac{1}{2}(n-1)p}}$$

$$=\frac{\prod_{h=1}^{m}|S_h|^{\frac{1}{2}(n-1)}}{|S+R|^{\frac{1}{2}m(n-1)}}\cdot m^{\frac{1}{2}m(n-1)p} \tag{10·14}$$

因上式與定理 9·1 之 W 同型，即爲定理 9·1 之 q=1 時的特殊場合，所以我們可利用測驗統計值 $-2\rho\log W_2$ 於測驗 H_2，式中 ρ 如 (9·12) 式，即

$$\rho=1-\left(\sum_{h=1}^{m}\frac{1}{f_h}-\frac{1}{f_o}\right)\frac{2p^2+3p-1}{6(m-1)(p+3)}-\frac{p-m+2}{f_o(p+3)}$$

$$=1-\frac{(m+1)(2p^2+3p-1)}{6m(n-1)(p+3)}-\frac{p-m+2}{m(n-1)(p+3)} \tag{10·15}$$

因此，H_2 之測驗統計值如次：

$$M_2=-2\rho\log W_2 \tag{10·16}$$

其近似分布爲自由度等於 $f=\frac{1}{2}(m-1)p(p+3)$ 之 χ^2 分布。由實際資料求得之 M_2 值爲 136.5171，根據自由度 f=27 查得 $\chi^2_{.01}=46.963$。因此，我們丟棄 H_2，即四個地域之複變值常態分布並不相一致，查其原因可能是四個地域之氣候相異以致影響水稻農藝性狀呈顯著之差異情形。我們待後擬更進一步討論水稻農藝性狀依隨氣象因子而變之複變值廻歸。

（2）臺中 65 號品種第二期作在四個不同地域之農藝性狀的差異顯著性測驗

（i）H_1：$\Sigma_1=\cdots\cdots=\Sigma_h$

依照第一期作之測驗方法，我們求出下列之最大機率推算值：

$$\begin{array}{lcc}
 & \hat{\mu}_h & n\Sigma_h^*\ (n=15) \\
臺\ 北: & \begin{pmatrix}98.8667\\12.9733\\3.0031\end{pmatrix} & \begin{pmatrix}814.433333 & 159.316667 & 47.787033\\ & 65.009333 & 16.924327\\ & & 6.378807\end{pmatrix} \\
臺\ 中: & \begin{pmatrix}105.2000\\14.3333\\4.3071\end{pmatrix} & \begin{pmatrix}518.980000 & 81.800000 & 59.396400\\ & 36.853333 & 12.708467\\ & & 12.091767\end{pmatrix} \\
嘉\ 義: & \begin{pmatrix}113.6667\\12.2467\\3.9909\end{pmatrix} & \begin{pmatrix}542.153333 & 40.963333 & 64.158023\\ & 53.037333 & 16.440981\\ & & 14.644643\end{pmatrix} \\
屏\ 東: & \begin{pmatrix}112.8733\\11.5400\\3.9275\end{pmatrix} & \begin{pmatrix}477.589333 & 25.196000 & 11.003210\\ & 49.636000 & 1.949380\\ & & 4.545256\end{pmatrix}
\end{array}$$

$$mn\Sigma^*=\begin{pmatrix}2353.155999 & 307.276000 & 182.344666\\ & 204.535999 & 48.023155\\ & & 37.660473\end{pmatrix}$$

應用 (10•9) 式計算所得之 M_1 值爲 22.4291。根據自由度 $f=18$ 查得 $\chi^2_{.01}=34.805$。因此，我們接受 H_1，即四地域之分散量具有均等性，此結果與第一期作測驗結果相反。

(ii) $H_2: \mu_1=\cdots\cdots=\mu_4,\ \sum_1=\cdots\cdots=\sum_4$

我們從 H_1 之測驗知悉第二期作灣中 65 號在四地域之分散量具有均等性，現在想更進一步知道其農藝性狀平均值是否在四地域亦相等。原來我們可在四地域之分散量相等之前提下測驗 $H_0: \mu_1=\cdots\cdots=\mu_4$，但四地域之分散量相等之前提是否確實成立却非經測驗 H_1 無從而知，爲了確定此前提成立與否而舉行 H_1 之測驗，然後再進行 H_0 之測驗，則顯著機率水準不能維持我們事先決定之 $\alpha\%$ 如 5% 或 1%，隨自由度之大小而增加。所以，直接測驗 H_2 較爲適當。在 Ω 中之最大機率推算值與 (i) 者相同，而在 ω 中之最大機率推算值如次：

$$\hat{\mu} = \begin{bmatrix} 107.6517 \\ 12.7733 \\ 3.8072 \end{bmatrix} \qquad mn\hat{\sum}{}^* = \begin{bmatrix} 4552.649899 & 79.432700 & 295.928587 \\ & 268.617299 & 53.630028 \\ & & 51.831463 \end{bmatrix}$$

應用 (10•16) 式計算所得之 M_2 值爲 88.6271，根據自由度 $f=27$ 查得 $\chi^2_{.01}=46.963$。可見四個地域之複變值常態分布並不相一致。我們從 H_1 之測驗得知 $\sum_1=\cdots\cdots=\sum_4$，所以四個地域之複變值常態分布具有相同的分散量行列但不具相同的期望值向量。如同第一期作，此期望值向量之差異，可能是由於不同氣候之影響所致。

(3) 第一期作臺中 65 號在四地域農藝性狀隨氣象因子而變之複變值直線迴歸分析

現在我們欲探知株高 (y_1)、穗數 (y_2)、及穀產量 (y_3) 等三種農藝性狀依隨溫度 (x_1)、日照 (x_2)、濕度 (x_3)、溫差 (x_4)、及雨量 (x_5) 等五種氣象因子而變之複變值直線廻歸。利用 (3•1) 及 (4•12) 兩式我們分別可得到四地域內之 B_h 及 $\sum_h$ 的無偏推算值。在此兩式中未經矯正之平方和乘積和可用矯正之平方和乘積和取而代之。玆將計算結果列如下：

臺　北：

$$\begin{array}{cccccccc} & & x_1 & x_2 & x_3 & x_4 & x_5 & \\ \hat{B}_1 = & \Bigg[& 6.7615 & -24.7692 & -0.2735 & 25.2408 & 6.0997 & \Bigg] \; y_1 \\ & & 13.6079 & -0.1875 & -0.7166 & 14.4585 & 3.9099 & \quad y_2 \\ & & -0.9793 & 1.4179 & -0.0728 & 0.6742 & 0.0209 & \quad y_3 \end{array}$$

$$S_1 = \begin{bmatrix} 145.076000 & 14.959027 & 2.359860 \\ & 21.874603 & -2.695691 \\ & & 0.970962 \end{bmatrix}$$

臺　中：

$$\hat{B}_2 = \begin{bmatrix} 14.8624 & -9.3933 & 2.3655 & 35.5272 & 1.2772 \\ 11.6458 & -4.4308 & -0.3641 & 8.1615 & 0.1758 \\ 1.3257 & -0.4161 & -0.0674 & -0.5580 & -0.8666 \end{bmatrix}$$

$$S_2 = \begin{bmatrix} 308.318631 & 15.248768 & -2.150375 \\ & 12.003349 & 1.557210 \\ & & 0.770226 \end{bmatrix}$$

嘉　義：

$$\hat{B}_3 = \begin{bmatrix} -0.7425 & -21.8582 & 0.4334 & -7.5999 & -6.8977 \\ 3.3702 & 2.9908 & 0.8268 & -6.2201 & -0.0814 \\ 0.4883 & -1.9577 & -0.1952 & -1.6608 & -1.1940 \end{bmatrix}$$

$$S_3 = \begin{bmatrix} 517.861745 & 7.986976 & 27.542429 \\ & 85.497935 & 1.633434 \\ & & 2.883861 \end{bmatrix}$$

屏　東：

$$\hat{B}_4 = \begin{bmatrix} -2.2146 & -15.2390 & -0.7680 & 20.4083 & 3.6436 \\ 4.2988 & 0.8710 & -0.6375 & -2.0449 & -1.3972 \\ -2.6055 & -2.7507 & 0.0981 & 4.7997 & 0.9253 \end{bmatrix}$$

$$S_4 = \begin{bmatrix} 274.644727 & 33.237807 & 8.974802 \\ & 13.128425 & 2.028377 \\ & & 2.428314 \end{bmatrix}$$

上列之符號中 $S_h=(n_h-q-1)\hat{\sum}_h\ (n_h=15, q=5, h=1, \cdots\cdots, 4)$，$\hat{\sum}_h$ 爲 $\sum_h$ 之無偏推算值而 $\sum_h$ 之最大機率推算值爲 $n_h^{-1}S_h$。

觀察上列之計算結果可知各地域農藝性狀隨氣象因子而變之廻歸係數並不相一致，此說是否確實容後測驗擬說而求證之。現在我們想測驗四地域之剩餘機差 (Residual Error) 分散量之均等性，即

$$H_1: \sum_1=\sum_2=\sum_3=\sum_4$$

玆利用定理 8•1 得到下列之測驗統計值：

$$M_1=-2\rho\log\left(\frac{\prod_{h=1}^{m}|S_h|^{\frac{1}{2}(n-q-1)}}{|S|^{\frac{1}{2}m(n-q-1)}}\cdot m^{\frac{1}{2}m(n-q-1)p}\right) \tag{10•17}$$

式中 $\rho=1-\dfrac{(m+1)(2p^2+3p-1)}{6m(n-q-1)(p+1)}$，　$n_h=n$

$$S=\sum_{h=1}^{m}S_h$$

M_1 之近似分布爲自由度等於 $f=\dfrac{1}{2}(m-1)p(p+1)$ 之 χ^2 分布。由實際資料求得之 M_1 值爲 31.9784。根據自由度 $f=18$ 查得 $\chi^2_{.01}=34.805$。因此，在顯著機率水準爲 1% 之下我們接受 H_1，即四個地域之剩餘機差分散量具有均等性。我們現在更進一步測驗四個地域之複變值直線廻歸之分布是否相一致，即欲測驗下列之擬說：

$$H_2: B_1=\cdots\cdots=B_4,\ \sum_1=\cdots\cdots=\sum_4$$

我們可應用 (9•4) 及 (9•5) 兩式求出在 ω 下之 B 及 $\sum$ 的最大機率推算值。在此兩式中雖然用無矯正之平方和乘積和，但我們亦可用矯正之平方和乘積和取而代之。例如行列 $\sum_{h=1}^{m}X_hX_h'$ 可用四個地域之氣象記錄資料合併後求得之矯正平方和乘積和代之。玆將計算結果列如下：

$$\hat{B}=\begin{bmatrix} -0.8124 & 0.8549 & 0.5350 & 1.3472 & -0.4483 \\ 3.0746 & -0.9257 & 0.3932 & -5.2965 & -0.9766 \\ -1.5365 & 0.2155 & 0.0641 & -0.6507 & -1.0477 \end{bmatrix}$$

$$S=\begin{bmatrix} 1703.439136 & 173.223966 & 59.488264 \\ & 288.498163 & 12.885429 \\ & & 11.539705 \end{bmatrix}$$

上列之 $S=(mn-q-1)\hat{\sum}$，$\hat{\sum}$ 爲 $\sum$ 之無偏推算值而 $\sum$ 之最大機率推算值爲 $(mn)^{-1}S$。

利用定理 9•1 求得 H_2 之測驗統計值如次：

$$M_2=-2\rho\log\left(\frac{\prod_{h=1}^{m}|S_h|^{\frac{1}{2}(n-q-1)}}{|S|^{\frac{1}{2}m(n-q-1)}}\,m^{\frac{1}{2}m(n-q-1)p}\right) \tag{10•18}$$

式中 $\rho=1-\dfrac{(m+1)(2p^2+3p-1)}{6m(n-q-1)(p+2q+1)}-\dfrac{q[p-(m-1)q+1]}{m(n-q-1)(p+2q+1)}$

M_2 之近似分布爲自由度等於 $f=\dfrac{1}{2}(m-1)p(p+2q+1)$ 之 χ^2 分布。由實際資料求得之 M_2 值爲 96.5627，因爲 χ^2 之自由度 $f=63$ 爲相當大之數值，故經過變數轉換後所得之 $\sqrt{2\chi^2}-\sqrt{2f-1}$ 的分布可視爲 $N(0,1)$。於是我們計算 $\sqrt{2M_1}-\sqrt{2f-1}$ 得 2.7167，而在顯著機率水準爲 1% 下之常態變值爲 2.58，因此，我們丟棄 H_2，即四個地域之複變值直線廻歸之分布並不相一致。根據 H_1 及 H_2 之測驗結果我們獲悉四個地域之複變值直線廻歸之分布爲不具有相等之廻歸係數行列而具有相等剩餘機差分散量行列之複變值常態分布。可見第一期作臺中 65 號品種之農藝性狀隨氣象因子而變化之情形因地域不同而異。

(4) 第二期作臺中 65 號品種在四地域農藝性狀隨氣象因子而變之複變値直線廻歸分析

(i) $H_1: \Sigma_1 = \cdots\cdots = \Sigma_4$ 之測驗

按照第一期之測驗方法，我們求出四地域之 B_h 及 Σ_h 的無偏推算值如下：

臺 北：

$$\hat{B}_1 = \begin{array}{c} \begin{array}{ccccc} x_1 & x_2 & x_3 & x_4 & x_5 \end{array} \\ \begin{bmatrix} -35.8458 & -13.3337 & -5.4685 & -65.5533 & 14.1489 \\ -17.0308 & -5.6589 & -0.3450 & -0.3117 & 1.3512 \\ -5.3414 & -1.3563 & -0.1360 & -5.2984 & -0.5014 \end{bmatrix} \end{array} \begin{array}{c} y_1 \\ y_2 \\ y_3 \end{array}$$

$$S_1 = \begin{bmatrix} 312.935532 & 92.574912 & 29.667328 \\ & 44.991323 & 11.922745 \\ & & 4.729463 \end{bmatrix}$$

臺 中：

$$\hat{B}_2 = \begin{bmatrix} -29.0716 & 30.5913 & 11.3015 & -22.3891 & -12.9357 \\ -7.2249 & 8.2999 & 2.8744 & -4.9400 & -2.5532 \\ -0.9773 & 1.9363 & 3.0422 & -2.6474 & -1.1755 \end{bmatrix}$$

$$S_2 = \begin{bmatrix} 303.959179 & 34.956274 & 35.395057 \\ & 25.619424 & 7.254976 \\ & & 6.502143 \end{bmatrix}$$

嘉 義：

$$\hat{B}_3 = \begin{bmatrix} -2.9046 & -21.4032 & -0.5497 & 32.4084 & -3.9546 \\ 8.1409 & 2.1768 & -0.5393 & 7.1427 & 0.0128 \\ -0.0351 & 0.0416 & -1.2506 & 7.6818 & 0.0745 \end{bmatrix}$$

$$S_3 = \begin{bmatrix} 357.750087 & 29.603458 & 39.024354 \\ & 28.009171 & 7.080369 \\ & & 6.866171 \end{bmatrix}$$

屏 東：

$$\hat{B}_4 = \begin{bmatrix} -46.1217 & -23.2068 & -1.5716 & -5.7577 & -7.1125 \\ -2.3410 & 3.5738 & -0.4546 & -5.6159 & -1.3012 \\ -3.0732 & -4.6325 & 0.2113 & 7.0827 & -1.1225 \end{bmatrix}$$

$$S_4 = \begin{bmatrix} 242.619781 & -14.677770 & 5.204928 \\ & 38.379706 & -0.071698 \\ & & 2.505369 \end{bmatrix}$$

上列之 $S_h = (n-q-1)\hat{\Sigma}_h$，$\hat{\Sigma}_h$ 爲 Σ_h 之無偏推算值而 Σ_h 之最大機率推算值爲 $n^{-1}S_h$，利用 (10•17) 式計算所得之 M_1 值爲 12.5838，根據自由度 $f=18$ 査得之 $\chi_{.01}^2 = 34.805$。因此，我們在顯著機率水準爲 1% 下接受 H_1，即四地域之剩餘機差分散量具有均等性。

(ii) $H_2: B_1 = \cdots\cdots = B_4,\ \Sigma_1 = \cdots\cdots = \Sigma_4$

倣照第一期作之測驗方法，利用 (9•4) 及 (9•5) 兩式求出在 ω 下之 B 及 Σ 的最大機率推算值如下：

$$\hat{B} = \begin{bmatrix} -16.3061 & -7.6873 & -2.2982 & -8.2283 & -2.1852 \\ 0.0778 & 2.7796 & -0.0159 & 5.4356 & -1.0208 \\ -1.9044 & 0.1457 & -0.0307 & 2.2069 & -0.4785 \end{bmatrix}$$

$$S = \begin{bmatrix} 2083.270392 & 271.740561 & 175.686575 \\ & 202.540390 & 43.924064 \\ & & 37.264807 \end{bmatrix}$$

上列之 $S = (mn-q-1)\hat{\Sigma}$，$\hat{\Sigma}$ 爲 Σ 之無偏推算值而 Σ 之最大機率推算值爲 $(mn)^{-1}S$，在 Ω 下之 B_h 及 Σ_h 的最大機率推算值與 (i) 同，將其代入 (10•18) 式得 M_2 值爲 76.5227。因 χ^2 之自由度 $f=63$ 相當大，故我們計算近似常態變值 $\sqrt{2M_2} - \sqrt{2f-1}$ 得 1.1909，我們知道在顯著機率水準爲 1% 下之常態變值爲 2.58，因此，我們接受 H_2，即四個地域之複變值直線廻歸之分布相一致，由此可見第二期作臺中 65 號品種之農藝性狀隨氣象因子而變化之情形不因地域之不同而異，此與第一期作之情形完全相反。

綜合以上之分析結果可知臺中 65 號品種於第一期作栽培時，其農藝性狀受氣象因子之影響因地域之不同而異，可是在第二期作栽培時則不然。農藝性狀中最重要的是穀產量，在第一期作時以臺中及屏東兩地之

產量爲最高而在第二期作時除臺中及屏東兩地外，嘉義地域之產量尚稱良好，在臺北地域兩期作之穀產量都不如臺中及屏東兩地。總之，臺中65號水稻品種最適應的地域爲臺中及屏東兩地。

（十一）摘　　要

本研究係應用常態分布下之複變值直線廻歸 (Multivariate Linear Regression) 於探討水稻品種之地域適應性。首先誘導複變值直線廻歸 $y=Bx+d$ 之複變值常態機率密度函數 (Multivariate Normal Probability Density Function)，然後求出該函數所含之未知常數 (Parameters) 的最大機率推算式 (Maximum Likelihood Estimators)。繼之，討論最大機率推算值之分布並利用機率比值軌範 (Likelidood Ratio Criterion) 測驗各種統計擬說 (Statistical Hypothesis) 之顯著性。測驗數個複變值常態直線廻歸分布之一致性的統計擬說，即 $H_2: B_1=\cdots\cdots=B_m$, $\sum_1=\cdots\cdots=\sum_m$，尚未發現於任何之文獻上。因此，著者曾應用機率比值軌範導出該統計擬說之測驗統計值 (Test Statistic) 並求其近似分布以供測驗之用。最後搜集臺灣省農業試驗所、臺中、嘉義、及高雄農業改良場過去十五年間所舉行之水稻品種結果記錄資料及歷年氣象因子觀測記錄資料做爲本研究之資料，用於探討氣象因子及地域環境與稻作之關係，以便選出各地域最適應之品種。

誌謝：本研究承蒙國家科學委員會之補助，又在水稻品種地域適應性之分析時承臺灣大學生物統計研究室吳果高講師之電腦計算程序設計，農藝研究所沈明來同學及助理員曾潔芳小姐協助計算工作，臺灣省農業試驗所，臺中區，高雄區農業改良場及嘉義農業試驗分所供給資料，於此一併致謝。

參考文獻 (Reference)

1. Anderson, T. W. (1958) An Introduction to Multivariate Statistical Analysis. John Wiley & Sons, Inc., New York.
2. Bartlett, M. S. (1947) Multivariate analysis. Journal of Royal Statistical Society, Series B, 9:176-197.
3. Bose, R. C. and Roy, S. N. (1938) The distribution of Studentised D^2-statistic. Sankhya 4:19-38.
4. Box, G. E. P. (1949) A general distribution theory for a class of likelihood criteria. Biometrika 36:317-346.
5. Hotelling, H. (1931) The generalization of Student's ratio. Annals of Mathematical Statistics 2:360-378.
6. Hsu, P. L. (1938) Notes on Hotelling's generalized T^2. Annals of Mathematical Statistics 9:231-243.
7. Kendall, M. G. and Stuart, A. (1966) The Advanced Theory of Statistics, Vol. 3. Hafner Publishing Co., New York.
8. Nair, U. S. (1939) The application of the moment function in the study of distribution laws in statistics. Biometrika 30:274-294.
9. Rao, C. R. (1946) Tests with discrimenant functions in multivariate analysis. Sankhya 7:407-414.
10. Rao, C. R. (1948) Tests of significance in multivariate analysis. Biometrika 35:58-79.
11. Rao, C. R. (1949) On some problems arising out of discrimination with multiple characters. Sankhya 9:343-364.
12. Rao, C. R. (1965) Linear Statistical Inference and Its Applications. John Wiley & Sons, Inc., New York.
13. Wijsman, R. A. (1957) Random orthogonal transformations and their use in some classical distribution problems in multivariate analysis. Annals of Mathematical Statistics 28:415-423.
14. Wilks, S. S. (1932) Certain generalizations in the analysis of variance. Biometrika 24:471-494.
15. Wilks, S. S. (1962) Mathematical Statistics. John Wiley & Sons, Inc., New York.

Some Theory of Multivariate Linear Regression Analysis and Its Application to Analysing Rice Varietal Adaptation

by

Ing-Tzer Wey

Summary

The rice varietal adaptation in Taiwan is discussed in this study by using the multivariate linear regression under the assumption of normality. First of all, the multivariate normal probability density function of the linear regression $\mathbf{y}=\mathbf{Bx}+\mathbf{d}$ is derived and the estimatiors of parameters in the function are obtained by the method of maximum likelihood. The sampling distribution of the maximum likelihood estimates is then studied and several kinds of statistical hypothesis are tested by using likelihood ratio criterion. Test of hypothesis that several multivariate normal distributions of linear regression are identical, namely

$$H_2: B_1=\cdots\cdots=B_m, \ \Sigma_1=\cdots\cdots=\Sigma_m$$

is not found in literature. Therefore, the author has obtained from likelihood ratio criterion the test statistic for H_2 and its asymptotic distribution provided for testing the hypothesis in practice. Finally, the multivariate linear regression analysis is applied to the field experiment data obtained from different locations and different crop seasons for identification of rice varieties adaptable to various climates and locations.