

多次元の場合の Range について

塩 谷 実

(1959 年 2 月 受付)

On the Range in the Multivariate Case

Minoru SIOTANI

ABSTRACT: The range in the multivariate case is defined as the maximum value of the generalized distances between any two sample points in p -dimensional space under consideration. In the univariate case, our statistic becomes the square of the ordinary range. As an example of using the range in the multivariate case, the multivariate confidence bounds relating to mean differences in k multivariate populations, which due to S. N. Roy, is explained. The main works in this paper is to give the method of obtaining the approximate upper percentage points of our range when the population distribution is normal and is to check the accuracy of the approximation numerically. The result is satisfactory for the practical application.

Institute of Statistical Mathematics

§ 1. Range の定義

x_1, x_2, \dots, x_n を p -次元の (縦) ベクトル変量 x についての n 個の独立な観測値とする。 x の従う分布は平均 m , 分散・共分散行列 A をもつものとする。任意の 2 つ x_α, x_β ($\alpha \neq \beta$) の間の一般化された距離は (常数を除いて)

$$R^2_{\alpha\beta} = (x_\alpha - x_\beta)' A^{-1} (x_\alpha - x_\beta) \dots\dots\dots(1.1)$$

で定義される。 A^{-1} は A の逆行列である。 (1.1) は変数を $y = (x_\alpha - x_\beta)$ と考えると、 y -空間に於ける原点を中心とする楕円体を表わして、 x_α と x_β の距りが大きい程、原点より遠い y -楕円体面上の点になる。即ち $R^2_{\alpha\beta}$ は大きな値になる。 x の方の空間に於ける、 m からの距りの度合を表わす、 m を中心とする楕円体 $c^2 = (x - m)' A^{-1} (x - m)$ の方で、たとえ同一楕円体面上にある二点 x_α, x_β であつても、 (x -空間で m からの距りという意味では同等である。) 遠く離れているもの程、 y -空間では、原点より遠くはなれた楕円体面上にある。また x_α と x_β が、 x -空間で異つた楕円体面上にあつても、その距りが近ければ、 y の方では、原点に近い楕円体面になることになる。つまり $R^2_{\alpha\beta}$ は小さい値になる。このように $R^2_{\alpha\beta}$ は二点 x_α, x_β の距りの度合を表わしているのである。そこで多次元の場合の Range を

$$R^2_{\text{MAX}} = \max_{\alpha < \beta} \{ (x_\alpha - x_\beta)' A^{-1} (x_\alpha - x_\beta) \} \dots\dots\dots(1.2)$$

で定義しよう。しかし母集団における共分散行列 A が未知の時には (1.1), (1.2) を用いること

ができないで、その推定行列をもつてくる必要がある。今差 $x_\alpha - x_\beta$ と統計的に独立な A の不偏推定行列 L が利用できるものとする。しかる時、 A が未知の場合の Range を

$$R^2_{MAX} = \max_{\alpha < \beta} \{R^2_{\alpha\beta}\} = \max_{\alpha < \beta} \{(x_\alpha - x_\beta)' L^{-1} (x_\alpha - x_\beta)\} \dots\dots\dots (1.3)$$

で定義する。 $p=1$ とすれば (1.2), (1.3) はそれぞれ

$$R^2_{MAX} = \max_{\alpha < \beta} \{(x_\alpha - x_\beta)^2 / \sigma^2\} = \{(\text{最大値} - \text{最小値}) / \sigma\}^2 \dots\dots\dots (1.4)$$

$$R^2_{MAX} = \max_{\alpha < \beta} \{(x_\alpha - x_\beta)^2 / s^2\} = \{(\text{最大値} - \text{最小値}) / s\}^2 \dots\dots\dots (1.5)$$

となり、これ等の正の平方根は一次元の場合の普通の Range である。

これらは次の k 個の平均の変動をみる場合に拡張される。 $\bar{x}_\alpha (\alpha=1, 2, \dots, k)$ を平均 m_α , 共分散行列 A をもつ k 個の母集団からの大きさ n_α の標本における平均値とする。しかる時は、 $z_\alpha = \bar{x}_\alpha - m_\alpha$ は平均 0, 共分散行列 $\frac{1}{n_\alpha} A$ を持ち、 $z_\alpha - z_\beta = (\bar{x}_\alpha - \bar{x}_\beta) - (m_\alpha - m_\beta)$ の共分散行列は $\frac{n_\alpha + n_\beta}{n_\alpha n_\beta} A$ となる。故にこの場合 z_1, z_2, \dots, z_k の間の Range は

$$R^2_{MAX} = \max_{\alpha < \beta} \left\{ \frac{n_\alpha n_\beta}{n_\alpha + n_\beta} (z_\alpha - z_\beta)' A^{-1} (z_\alpha - z_\beta) \right\} \dots\dots\dots (1.6)$$

と定義される。 L が利用される時には

$$R^2_{MAX} = \max_{\alpha < \beta} \left\{ \frac{n_\alpha n_\beta}{n_\alpha + n_\beta} (z_\alpha - z_\beta)' L^{-1} (z_\alpha - z_\beta) \right\} \dots\dots\dots (1.7)$$

となり、これは $m_1 = m_2 = \dots = m_k$ のテスト及び m_α の任意の比較の同時信頼限界を求めるときに使うことができる。特に $n_1 = n_2 = \dots = n_k \equiv n$ のときは (1.6), (1.7) より $\frac{n}{2}$ を除いて差支えない。

§ 2. Range の用いられる一つの例

§ 1 で定義された Range の用いられる例として S. N. Roy [4] による多次元信頼限界を求める一つの問題を考える。これは Tukey [6] 及び Scheffé [5] が一次元の場合に取扱つた問題の多次元への拡張である。今 k 個の p 変数母集団 $\Pi_i(m_i, A)^*$ ($i=1, 2, \dots, k$) から大きさ $n_i (i=1, 2, \dots, k)$ のランダム・サンプルをとる。 $\bar{x}_i (i=1, 2, \dots, k)$ を標本平均、 L を共通な共分散行列 A の不変推定行列とする。

$$\left(\sum_{i=1}^k n_i - k \right) L = \sum_{i=1}^k \left\{ \sum_{\alpha=1}^{n_i} x_{i\alpha} x_{i\alpha}' - n_i \bar{x}_i \bar{x}_i' \right\}. \dots\dots\dots (2.1)$$

これ等の知識をもとに

$$a' (m_i - m_j) \dots\dots\dots (2.2)$$

すべての零でない a' ,
すべての i, j

に対する同時信頼限界の組を求めることを考える。このため先づ

$$\begin{aligned} & (\bar{x}_i - \bar{x}_j - m_i + m_j)' L^{-1} (\bar{x}_i - \bar{x}_j - m_i + m_j) \\ & = \text{Sup}_a a' (\bar{x}_i - \bar{x}_j - m_i + m_j) (\bar{x}_i - \bar{x}_j - m_i + m_j)' a / a' L a \dots\dots\dots (2.3) \end{aligned}$$

であることに注意する。

[(2.3) の証明] $\eta = a' (\bar{x}_i - \bar{x}_j - m_i + m_j) (\bar{x}_i - \bar{x}_j - m_i + m_j)' a / a' L a$ とおき両辺を $a_i (i=1, 2, \dots, p)$ で微分して 0 とおくことにより η の極大値を与える a は

* 平均ベクトル m_i , 共分散行列 A をもつ母集団を意味する。

$$((\bar{x}_i - \bar{x}_j - m_i + m_j)(\bar{x}_i - \bar{x}_j - m_i + m_j)' - \eta L)a = 0 \quad \dots\dots\dots(2.4)$$

を充さねばならない。かつ零でない a を考えているから、(2.4) が解を持つためには

$$|(\bar{x}_i - \bar{x}_j - m_i + m_j)(\bar{x}_i - \bar{x}_j - m_i + m_j)' - \eta L| = 0 \quad \dots\dots\dots(2.5)$$

が成立する。これを充す η は求める極大値である。(2.5) の根は p 個あるが、明らかに $(\bar{x}_i - \bar{x}_j - m_i + m_j)(\bar{x}_i - \bar{x}_j - m_i + m_j)'$ の階数は 1 であるから、 $(p - 1)$ 個の根は 0 で、0 でない根は唯 1 個である。この 0 でない根を η_{\max} とすれば、(2.5) より、よく知られた行列論の定理を使つて

$$\begin{aligned} \text{tr}(\bar{x}_i - \bar{x}_j - m_i + m_j)(\bar{x}_i - \bar{x}_j - m_i + m_j)'L^{-1} &= (\eta \text{ の } p \text{ 個の根の和}) \\ &= \eta_{\max} \end{aligned}$$

である。これ即ち (2.3) である。(終)。

$$\mathfrak{R}_{ij}^2 = (\bar{x}_i - \bar{x}_j - m_i + m_j)' L^{-1}(\bar{x}_i - \bar{x}_j - m_i + m_j) \quad \dots\dots\dots(2.6)$$

とおけば、与えられた (i, j) にたいして又あるスカラー r^2 にたいして、 $\mathfrak{R}_{ij}^2 \leq r^2$ を考えると、これは (2.3) によつて

$$\begin{aligned} a'(\bar{x}_i - \bar{x}_j) - [r^2 a'La]^{\frac{1}{2}} &\leq a'(m_i - m_j) \leq a'(\bar{x}_i - \bar{x}_j) + [r^2 a'La]^{\frac{1}{2}} \quad \dots\dots\dots(2.7) \\ &\text{すべての零でない } a' \end{aligned}$$

と同等である。 $a'(m_i - m_j)$ の信頼限界を求める a' の方の要請は (2.7) で解決される。すべての i, j という要求にたいしては k 個の標本 (母集団) のすべての対を考えて

$$\text{すべての } \mathfrak{R}_{ij}^2 \leq r^2 \quad \dots\dots\dots(2.8)$$

なる事象をとればよい。(2.8) は勿論

$$\mathfrak{R}_{\text{MAX}}^2 \equiv \max_{i < j} \mathfrak{R}_{ij}^2 \leq r^2 \quad \dots\dots\dots(2.9)$$

と同等である。斯くして求める信頼限界にたいしては、(2.9) を基礎にし (2.7) と組合せて考えればよい。即ち今信頼係数を $1 - \alpha$ とするならば、 $\nu = \sum_{i=1}^k n_i - k$ として

$$P_r\{\mathfrak{R}_{\text{MAX}}^2 > r^2(\alpha; p, k, \nu)\} = \alpha \quad \dots\dots\dots(2.10)$$

を充す $r^2(\alpha; p, k, \nu)$ を求める。しかる時は、信頼係数 $1 - \alpha$ の、すべての零でない a' 、すべての i, j にたいする $a'(m_i - m_j)$ の同時信頼限界は

$$\begin{aligned} a'(\bar{x}_i - \bar{x}_j) - [r^2(\alpha; p, k, \nu)a'La]^{\frac{1}{2}} &\leq a'(m_i - m_j) \\ &\leq a'(\bar{x}_i - \bar{x}_j) + [r^2(\alpha; p, k, \nu)a'La]^{\frac{1}{2}} \quad \dots\dots(2.11) \end{aligned}$$

である。

(2.11) が実用になるためには (2.10) の解 $r^2(\alpha; p, k, \nu)$ が得られなければならない。しかし一般的にこれを求めることは $k=2$ の場合を除いてはきわめて困難である。若し $\Pi_i(m_i, A) = N(m_i, A)^{*}$ ($i=1, 2, \dots, k$) でかつ $n_1 = n_2 = \dots = n_k \equiv n$ の場合には、 $\bar{x}_i - m_i \equiv z_i$ は $N(0, \frac{1}{n}A)$ に従う独立な確率変数となり、次節で議論されるように、実用になる $r^2(\alpha; p, k, \nu)$ のよい近似を得ることができる。広く使用される Analysis of dispersion では、上の場合が多く現われるので、§ 3 の方法で数表が作成されれば大いに役に立つであろう。最後に本節の問題で A が既知ならば当然 $R_{\text{MAX}}^2 = \max_{i < j} \{(\bar{x}_i - \bar{x}_j - m_i + m_j)' A^{-1}(\bar{x}_i - \bar{x}_j - m_i + m_j)\}$ が用いられることを注意しておく。

§ 3. Range- R_{MAX}^2 と $\mathfrak{R}_{\text{MAX}}^2$ の上方臨界点

R_{MAX}^2 及び $\mathfrak{R}_{\text{MAX}}^2$ の標本分布を正確に求めることは、極めて困難であるので、ここでは筆者が

* 平均ベクトル m_i 、共分散行列 A をもつ正規母集団を意味する。

The Annals of the Institute of Mathematical Statistics, Vol. 10, No. 3 に発表する予定である***) 一般の場合の応用として, R^2_{MAX} と \mathfrak{R}^2_{MAX} の上方臨界点を近似的に求める方法を示す.

3.1 近似解の求め方. x_1, x_2, \dots, x_n を互に独立に p -変量正規分布 $N(m, A)$ に従う(縦)ベクトル変量とする. § 1 で定義した任意の二つのベクトル変量の間の汎距離

$$R^2_{\alpha\beta} = (x_\alpha - x_\beta)' A^{-1} (x_\alpha - x_\beta), \quad (\alpha \neq \beta), \quad (A \text{ が既知の時})$$

$$\mathfrak{R}^2_{\alpha\beta} = (x_\alpha - x_\beta)' L^{-1} (x_\alpha - x_\beta), \quad (\alpha \neq \beta), \quad (A \text{ が未知の時})$$

について考える. 先づ A の既知の場合の臨界点を求める方法を述べよう. $R^2_{\alpha\beta}$ の個数は $N = \frac{1}{2} n \times (n-1)$ であるが, これを一定の順序例えば, $R^2_{12}, R^2_{13}, \dots, R^2_{1n}, R^2_{23}, \dots, R^2_{2n}, \dots, R^2_{(n-1),n}$ と並べ, 改めて $R_1^2, R_2^2, \dots, R_N^2$ と命名する. しかる時は, 一定の値 r^2 に対して, $R^2_{MAX} > r^2$ なる事象は N 個の事象 $R_1^2 > r^2, R_2^2 > r^2, \dots, R_N^2 > r^2$ の少くとも 1 つが起ることと同等で, よく知られた確率論の公式 [1] により

$$Pr\{R^2_{MAX} > r^2\} = \sum_{i=1}^N Pr\{R_i^2 > r^2\} - \sum_{i < j} Pr\{R_i^2 > r^2, R_j^2 > r^2\} + \sum_{i < j < k} Pr\{R_i^2 > r^2, R_j^2 > r^2, R_k^2 > r^2\} - \dots \dots \dots (3.1)$$

を得る.

$$\begin{aligned} \sum_{i=1}^N Pr\{R_i^2 > r^2\} &= NP_r\{R_1^2 > r^2\} \\ &= NP_r\{(x_1 - x_2)' A^{-1} (x_1 - x_2) > r^2\} \dots \dots \dots (3.2) \end{aligned}$$

は明らかである. 第二項については, $Pr\{R_i^2 > r^2, R_j^2 > r^2\}$ が, i, j のすべての値に対して対称ではないが次の2つのグループに分かれている.

第1のグループ (G_1): 2つの R^2 が共通の x を含んでいない.

第2のグループ (G_2): 2つの R^2 が共通の x を含んでいる.

例えば, 2つの R^2 が, $(x_1 - x_2)' A^{-1} (x_1 - x_2)$ と $(x_3 - x_4)' A^{-1} (x_3 - x_4)$ である場合は G_1 に属し, 2つの R^2 が, $(x_1 - x_2)' A^{-1} (x_1 - x_2)$ と $(x_1 - x_3)' A^{-1} (x_1 - x_3)$ である時は G_2 に属す.

今 G_1, G_2 に含まれる要素の個数をそれぞれ M_1, M_2 で表わす. しかる時

$$M_1 = 0, \quad n = 3 \dots \dots \dots (3.3)$$

$$M_1 = \frac{1}{2} \sum_{i=1}^{n-3} i(i+1)(i+2), \quad n \geq 4 \dots \dots \dots (3.4)$$

である. これは数学的帰納法により, 容易に証明される.

[(3.4) の証明] $n = m$ の時 (3.4) が正しいと仮定し, $n = m + 1$ の時にも成立することを示す. $(x_\alpha - x_\beta)$ を (α, β) で表わすと, $n = m + 1$ の時に, $n = m$ の時に比し新たに加わる対は, $(1, m+1), (2, m+1), \dots, (m, m+1)$ の m 個である. この中の任意の $(i, m+1)$ と共通の数字を持たない対の個数は, $n = m$ の時の対の総数 $\frac{1}{2} m(m-1)$ から, 共通の数字を持つ対の個数 $(m-1)$ を引いたもの即ち $\frac{1}{2} m(m-1) - (m-1) = \frac{1}{2} (m-1)(m-2)$ である. 各 i について同数ずつあるから, $n = m$ の時の M_1 に新たに加わるものは $\frac{1}{2} m(m-1)(m-2)$ 個である. 故に $n = m + 1$ の時の M_1 は

$$\frac{1}{2} m(m-1)(m-2) + \frac{1}{2} \sum_{i=1}^{m-3} i(i+1)(i+2) = \frac{1}{2} \sum_{i=1}^{m-2} i(i+1)(i+2)$$

***) 以下に於いて此の論文を [3] として参照する.

即ち (3.4) が成立する。 $n = 4$ の時、 $M_1 = 3$ であることは明らかであるから、すべての n に対して (3.4) は証明されたことになる。

(3.4) の和の計算を遂行すれば、実際に M_1 を求める式として

$$M_1 = \binom{\frac{1}{2}(n-1)(n-2)}{2} \dots\dots\dots(3.5)$$

を得る。従つて M_2 は

$$\begin{aligned} M_2 &= \frac{1}{2}N(N-1) - M_1 = \binom{\frac{1}{2}n(n-1)}{2} - \binom{\frac{1}{2}(n-1)(n-2)}{2} \\ &= \frac{1}{2}n(n-1)(n-2) \dots\dots\dots(3.6) \end{aligned}$$

となる。

G_1 に属する同時確率、 G_2 に属する同時確率は、それぞれのグループ内では、事象の対にたいして symmetric であるから

$$\begin{aligned} \sum_{i < j}^N P_r\{R_i^2 > r^2, R_j^2 > r^2\} &= M_1 P_r\{(x_1 - x_2)'A^{-1}(x_1 - x_2) > r^2, (x_3 - x_4)'A^{-1}(x_3 - x_4) > r^2\} \\ &+ M_2 P_r\{(x_1 - x_2)'A^{-1}(x_1 - x_2) > r^2, (x_1 - x_3)'A^{-1}(x_1 - x_3) > r^2\} \end{aligned} \quad (3.7)$$

と書くことができる。

以上は A が既知の場合であるが、未知の場合にも全く同様に次の式を得る。

$$\begin{aligned} P_r\{\mathfrak{R}_{MAX}^2 > s^2\} &= \sum_{i=1}^N P_r\{\mathfrak{R}_i^2 > s^2\} - \sum_{i < j}^N P_r\{\mathfrak{R}_i^2 > s^2, \mathfrak{R}_j^2 > s^2\} + \dots \\ &= NP_r\{(x_1 - x_2)'L^{-1}(x_1 - x_2) > s^2\} \\ &- M_1 P_r\{(x_1 - x_2)'L^{-1}(x_1 - x_2) > s^2, (x_3 - x_4)'L^{-1}(x_3 - x_4) > s^2\} \\ &- M_2 P_r\{(x_1 - x_2)'L^{-1}(x_1 - x_2) > s^2, (x_1 - x_3)'L^{-1}(x_1 - x_3) > s^2\} \\ &+ \dots\dots\dots(3.8) \end{aligned}$$

下の表は各 n について N, M_1, M_2 を与えるものである。

第 1 表 N, M_1, M_2 の値

| n | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|-------|---|----|----|----|-----|-----|-----|-----|-----|
| N | 3 | 6 | 10 | 15 | 21 | 28 | 36 | 45 | 55 |
| M_1 | 0 | 3 | 15 | 45 | 105 | 210 | 378 | 630 | 990 |
| M_2 | 3 | 12 | 30 | 60 | 105 | 168 | 252 | 360 | 495 |

3.2 A が既知の場合 — R_{MAX}^2 の上方臨界点 —

普通に用いられる範囲の有意水準 α にたいして、 R_{MAX}^2 の上方 $100\alpha\%$ 点 $r^2(\alpha; p, N)$ の第 1 次近似は、(3.1) 式の第 1 項を用いて

$$\alpha = NP_r\{(x_1 - x_2)'A^{-1}(x_1 - x_2) > r_1^2(\alpha; p, N)\} \dots\dots\dots(3.9)$$

を充す $r_1^2(\alpha; p, N)$ で与えられる。 [3], § 3 の (3.6) 式により、この第 1 次近似は

$$r_1^2(\alpha; p, N) = 2\chi^2(\alpha/N; p) \dots\dots\dots(3.10)$$

で計算される。ここに $N = \frac{1}{2}n(n-1)$ であり、 $\chi^2(\alpha/N; p)$ は自由度 p のカイ—自乗分布の上方 $100\alpha/N\%$ 点である。 $r_1^2(\alpha; p, N)$ の精度を見るために次の Bonferroni の不等式を利用

する。

$$NP_r\{(x_1-x_2)'A^{-1}(x_1-x_2) > r_1^2(\alpha; p, N)\} - \beta(\alpha; p, N) < P_r\{R^2_{MAX} > r_1^2(\alpha; p, N)\} \\ < NP_r\{(x_1-x_2)'A^{-1}(x_1-x_2) > r_1^2(\alpha; p, N)\}$$

即ち

$$\alpha - \beta(\alpha; p, N) < P_r\{R^2_{MAX} > r_1^2(\alpha; p, N)\} < \alpha \dots\dots\dots (3.11)$$

ただし $\beta(\alpha; p, N) = M_1 P_r\{(x_1-x_2)'A^{-1}(x_1-x_2) > r_1^2(\alpha; p, N), (x_3-x_4)'A^{-1}(x_3-x_4) > r_1^2(\alpha; p, N)\} + M_2 P_r\{(x_1-x_2)'A^{-1}(x_1-x_2) > r_1^2(\alpha; p, N), (x_1-x_3)'A^{-1}(x_1-x_3) > r_1^2(\alpha; p, N)\} \dots\dots\dots (3.12)$

従つて $\beta(\alpha; p, N)$ を評価する必要がある。G₁ に属する方は簡単で

$$P_r\{(x_1-x_2)'A^{-1}(x_1-x_2) > r_1^2(\alpha; p, N), (x_3-x_4)'A^{-1}(x_3-x_4) > r_1^2(\alpha; p, N)\} \\ = \left[P_r\{(x_1-x_2)'A^{-1}(x_1-x_2) > r_1^2(\alpha; p, N)\} \right]^2 \\ = (\alpha/N)^2 = \left[\alpha / \frac{1}{2} n(n-1) \right]^2 \dots\dots\dots (3.13)$$

G₂ に属する方は [3] の (4.8) の $\gamma=2, \delta=1$ としたもので

$$P_r\{(x_1-x_2)'A^{-1}(x_1-x_2) > r_1^2(\alpha; p, N), (x_1-x_3)'A^{-1}(x_1-x_3) > r_1^2(\alpha; p, N)\} \\ = \left(\frac{3}{4}\right)^{\frac{1}{2}p} \left[\Gamma\left(\frac{1}{2}p\right) \right]^{-1} \sum_{j=0}^{\infty} \frac{\Gamma\left(\frac{1}{2}p+j\right)}{j!} \left(\frac{1}{2}\right)^{2j} \left(\int_{\frac{1}{2}r_1^2}^{\infty} g^*_{p+2j}(x^2) dx^2 \right)^2 \dots\dots (3.14)$$

ただし $\eta = \frac{1}{3} r_1^2(\alpha; p, N)$ でかつ $g^*_{p+2j}(x^2) = \left[\Gamma\left(\frac{1}{2}p+j\right) \right]^{-1} 2^{-(\frac{1}{2}p+j)} (\chi^2)^{\frac{p}{2}+j-1} e^{-\frac{\chi^2}{2}}$

即ち自由度 $(p+2j)$ のカイ—自乗分布の密度函数である。(3.13), (3.14) により $\beta(\alpha; p, N)$ を $p=1, 2, 4$ の時に数値的に調べると下表のようになる。

第 2 表 $\beta(\alpha; p, N)$

| $p \backslash n$ | 3 | 4 | 5 | 7 | 10 |
|------------------|----------|----------|----------|----------|----------|
| $\alpha = 0.05$ | | | | | |
| 1 | 0.00603 | 0.00947 | 0.01183 | 0.01507 | 0.01822 |
| 2 | 0.00531 | 0.00818 | 0.01008 | 0.01262 | 0.01502 |
| 4 | 0.00473 | 0.00715 | 0.00851 | 0.01065 | 0.01245 |
| $\alpha = 0.01$ | | | | | |
| 1 | 0.000657 | 0.001019 | 0.001275 | 0.001623 | 0.001966 |
| 2 | 0.000551 | 0.000843 | 0.001040 | 0.001305 | 0.001562 |
| 4 | 0.000465 | 0.000706 | 0.000853 | 0.001048 | 0.001237 |

この表より、 $\beta(\alpha; p, N)$ は p が大きくなると小さくなり、 n が増せば大きくなる事がわかる。更にその値の大きさも α に対して無視することのできぬものである。従つて R^2_{MAX} の上方 100 α % 点を求めるには (3.1) 式の第 2 項まで考慮した第 2 近似による必要がある。しかし実際の計算はわづらわしいので、便法としてさきに求めた第 1 近似 $r_1^2(\alpha; p, N)$ に対する $\beta(\alpha; p, N)$ を用い

$$NP_r\{R^2_{MAX} > r_2^2(\alpha; p, N)\} = \alpha + \beta(\alpha; p, N) \dots\dots\dots (3.15)$$

の解 $r_2^2(\alpha; p, N)$ を R^2_{MAX} の上方 $100\alpha\%$ 点の第2近似とする. (3.1) 式の第2項までの範囲で, 正確に α となるような第2近似は, $r^2(\alpha; p, N)$ の under estimate となるので, 上の簡便法は可成りよい近似を与えるものと思われる. $p=1$ の時には, Range の $\%$ 点の表が既に利用できるので, これとわれわれの方法で求めたものと比較してみよう.

第3表 $p=1$ の時の $\sqrt{r_2^2(\alpha; 1, N)}$ と Bio. Table のものとの比較

| n | | 3 | 4 | 5 | 7 | 10 |
|----------------|----------|------|------|------|------|------|
| (3.15) によるもの | α = 0.05 | 3.33 | 3.65 | 3.87 | 4.18 | 4.49 |
| Bio. Table のもの | | 3.31 | 3.63 | 3.86 | 4.17 | 4.47 |
| (3.15) によるもの | α = 0.01 | 4.12 | 4.41 | 4.60 | 4.88 | 5.16 |
| Bio. Table のもの | | 4.12 | 4.40 | 4.60 | 4.88 | 5.16 |

Bio. Table = Biometrika Tables for Statisticians, Vol. I の Table 29.

非常によくあつている. $p > 1$ の時は $\beta(\alpha; p, N)$ が $p=1$ の時に比べ小さくなるので, もつと精度がよくなるのが期待される.

3.3 A が未知の場合 — \mathfrak{R}^2_{MAX} の上方臨界点 —

前節と同様に, \mathfrak{R}^2_{MAX} の上方 $100\alpha\%$ 点 $s^2(\alpha; p, N, \nu)$ の第1近似は (3.8) より

$$\alpha = NP_r\{(x_1 - x_2)'L^{-1}(x_1 - x_2) > s_1^2(\alpha; p, N, \nu)\} \dots\dots\dots(3.16)$$

を充す $s_1^2(\alpha; p, N, \nu)$ であたえられる. $x_1 - x_2$ は平均 0, 共分散行列 $2A$ を持つ正規分布に従うベクトル変量であり, A の不偏推定行列 L が自由度 ν の Wishart 分布に従うものとすれば, よく知られているように $\frac{1}{2}(x_1 - x_2)'L^{-1}(x_1 - x_2)$ は Hotelling の T^2 -分布に従うから, [3] の (3.7) により

$$s_1^2(\alpha; p, N, \nu) = 2\nu \left\{ \frac{1}{C\left(\alpha/N; \frac{1}{2}(\nu+1-p), \frac{1}{2}p\right)} - 1 \right\} \dots\dots\dots(3.17)$$

を得る. ここに $C(\alpha^*; a, b)$ は係数 a, b を持つベータ分布の下方 $100\alpha^*\%$ 点である. Bonferroni の不等式により

$$\alpha - \beta^*(\alpha; p, N, \nu) < P_r\{\mathfrak{R}^2_{MAX} > s_1^2(\alpha; p, N, \nu)\} < \alpha \dots\dots\dots(3.18)$$

を得るが 3.2 節と同様 $\beta^*(\alpha; p, N, \nu)$ を評価する必要がある. ただし

$$\begin{aligned} \beta^*(\alpha; p, N, \nu) = & M_1 P_r\{(x_1 - x_2)'L^{-1}(x_1 - x_2) > s_1^2(\alpha; p, N, \nu), (x_3 - x_4)'L^{-1}(x_3 - x_4) \\ & > s_1^2(\alpha; p, N, \nu)\} + M_2 P_r\{(x_1 - x_2)'L^{-1}(x_1 - x_2) > s_1^2(\alpha; p, N, \nu), \\ & (x_1 - x_3)'L^{-1}(x_1 - x_3) > s_1^2(\alpha; p, N, \nu)\} \dots\dots\dots(3.19) \end{aligned}$$

$\beta^*(\alpha; p, N, \nu)$ は $\beta(\alpha; p, N, \nu)$ の時と異なり正確に求めることが困難で, 自由度 ν についての漸近的な評価を行うしかない. この漸近的評価は [3] の定理 5.1 によつてできる. その結果の式を再録する.

$$\begin{aligned} \chi^2 &\equiv \chi^2(\alpha/N; p), \quad 2\eta \equiv [\gamma/(\gamma^2 - \delta^2)]r_1^2(\alpha; p, N) = 2[\gamma/(\gamma^2 - \delta^2)]\chi^2(\alpha/N; p), \\ g_m(\eta) &\equiv [\Gamma(m)]^{-1}\eta^{m-1}e^{-\eta}, \quad G_m(\eta) = \int_{\eta}^{\infty} g_m(t)dt, \end{aligned}$$

の記号を用いる. (y_1', y_2') を $(p+p)$ 次元のベクトル正規変量とし, $E(y_1) = E(y_2) = 0, E(y_1 y_1') = E(y_2 y_2') = \gamma A, E(y_1 y_2') = \delta A$ とするとき, $P_r\{y_1' L^{-1} y_1 > B_1^2(\alpha; p, N, \nu), y_2' L^{-1} y_2 > B_1^2(\alpha; p, N, \nu)\}$ は ν^{-2} の order の項まで次のように得られる. ただし $\gamma > |\delta|, L$ は A の自由度 ν をもつ

不偏推定行列でかつ y_i とは独立であり, $B_1^2(\alpha; p, N, \nu)$ は

$$\alpha = NP_r\{y_1'L^{-1}y_1 > B_1^2(\alpha; p, N, \nu)\} \dots\dots\dots (3.20)$$

を充すものである.

$$\begin{aligned} &P_r\{y_1'L^{-1}y_1 > B_1^2(\alpha; p, N, \nu), y_2'L^{-1}y_2 > B_1^2(\alpha; p, N, \nu)\} \\ &= O_0 + O_1 + O_2 + O(\nu^{-3}) \dots\dots\dots (3.21) \end{aligned}$$

$$O_0 = \left[1 - (\delta/r)^2\right]^{\frac{1}{2}p} \left[\Gamma\left(\frac{1}{2}p\right)\right]^{-1} \sum_{j=0}^{\infty} \frac{\Gamma\left(\frac{1}{2}p+j\right)}{j!} (\delta/r)^{2j} G_{\frac{p}{2}+j}^2(\eta),$$

$$O_1 = O_{11} + O_{12}$$

$$O_{11} = -\frac{1}{\nu} \frac{\chi^2(\chi^2+p)}{2} \left[1 - (\delta/r)^2\right]^{\frac{1}{2}p-1} \left[\Gamma\left(\frac{1}{2}p\right)\right]^{-1} \sum_{j=0}^{\infty} \frac{\Gamma\left(\frac{1}{2}p+j\right)}{j!} (\delta/r)^{2j} g_{\frac{p}{2}+j}(\eta) G_{\frac{p}{2}+j}^2(\eta),$$

$$\begin{aligned} O_{12} = &\frac{1}{\nu} \left[1 - (\delta/r)^2\right]^{\frac{p}{2}} \left[\Gamma\left(\frac{1}{2}p\right)\right]^{-1} \sum_{j=0}^{\infty} \frac{\Gamma\left(\frac{1}{2}p+1+j\right)}{j!} \left(\frac{\delta}{r}\right)^{2j} \left\{ (2\eta+p-2j) g_{\frac{p}{2}+j+1}(\eta) \times \right. \\ &\left. \times G_{\frac{p}{2}+j}^2(\eta) + (2j+1) g_{\frac{p}{2}+1+j}^2(\eta) \right\}, \end{aligned}$$

$$O_2 = O_{21} + O_{22} + O_{23} + O_{24} + O_{25}$$

$$\begin{aligned} O_{21} = &-\frac{1}{24\nu^2} \left[1 - (\delta/r)^2\right]^{\frac{1}{2}p-1} \left[\Gamma\left(\frac{1}{2}p\right)\right]^{-1} \chi^2 \left\{ 4\chi^4 + (13p-2)\chi^2 + 7p^2 - 4 \right\} \times \\ &\times \sum_{j=0}^{\infty} \frac{\Gamma\left(\frac{1}{2}p+j\right)}{j!} \left(\frac{\delta}{r}\right)^{2j} g_{\frac{p}{2}+j}(\eta) G_{\frac{p}{2}+j}^2(\eta), \end{aligned}$$

$$\begin{aligned} O_{22} = &\frac{1}{16\nu^2} \left[1 - (\delta/r)^2\right]^{\frac{p}{2}-2} \left[\Gamma\left(\frac{1}{2}p\right)\right]^{-1} \chi^4 (\chi^2+p)^2 \sum_{j=0}^{\infty} \frac{\Gamma\left(\frac{1}{2}p+j\right)}{j!} \left(\frac{\delta}{r}\right)^{2j} \times \\ &\times \left[\left\{ g_{\frac{p}{2}+j}(\eta) - g_{\frac{p}{2}+j-1}(\eta) \right\} G_{\frac{p}{2}+j}^2(\eta) + g_{\frac{p}{2}+j}^2(\eta) \right], \end{aligned}$$

$$\begin{aligned} O_{23} = &-\frac{1}{2\nu^2} \left[1 - (\delta/r)^2\right]^{\frac{1}{2}p-1} \left[\Gamma\left(\frac{1}{2}p\right)\right]^{-1} \chi^2 (\chi^2+p) \sum_{j=0}^{\infty} \frac{\Gamma\left(\frac{1}{2}p+1+j\right)}{j!} \left(\frac{\delta}{r}\right)^{2j} \times \\ &\times \left[\left(\eta + \frac{1}{2}p - j\right) \left\{ \left(g_{\frac{p}{2}+1+j}(\eta) - g_{\frac{p}{2}+j}(\eta) \right) G_{\frac{p}{2}+j}^2(\eta) + g_{\frac{p}{2}+1+j}(\eta) g_{\frac{p}{2}+j}(\eta) \right\} \right. \\ &\left. + (2j+1) \left\{ g_{\frac{p}{2}+1+j}(\eta) - g_{\frac{p}{2}+j}(\eta) \right\} g_{\frac{p}{2}+1+j}(\eta) - g_{\frac{p}{2}+1+j}(\eta) G_{\frac{p}{2}+j}^2(\eta) \right]. \end{aligned}$$

$$\begin{aligned} O_{24} = &-\frac{1}{\nu^2} \left[1 - (\delta/r)^2\right]^{\frac{p}{2}} \left[\Gamma\left(\frac{1}{2}p\right)\right]^{-1} \sum_{j=0}^{\infty} \frac{1}{j!} \left(\frac{\delta}{r}\right)^{2j} \times \\ &\times \left[\frac{8}{3} \Gamma\left(\frac{1}{2}p+3+j\right) \left\{ \left(g_{\frac{p}{2}+3+j}(\eta) - 2g_{\frac{p}{2}+2+j}(\eta) + g_{\frac{p}{2}+1+j}(\eta) \right) G_{\frac{p}{2}+j}^2(\eta) \right. \right. \\ &\left. \left. + 3 \left(g_{\frac{p}{2}+2+j}(\eta) - g_{\frac{p}{2}+1+j}(\eta) \right) g_{\frac{p}{2}+1+j}(\eta) \right\} \right. \\ &\left. + 4(p+3) \Gamma\left(\frac{1}{2}p+2+j\right) \left\{ \left(g_{\frac{p}{2}+2+j}(\eta) - g_{\frac{p}{2}+1+j}(\eta) \right) \times \right. \right. \\ &\left. \left. \times \left(G_{\frac{p}{2}+j}^2(\eta) - g_{\frac{p}{2}+1+j}(\eta) \right) + g_{\frac{p}{2}+1+j}^2(\eta) \right\} \right] \end{aligned}$$

$$\begin{aligned}
 & +2(p^2+3p+4)\Gamma\left(\frac{1}{2}p+1+j\right)\left\{G_{\frac{p}{2}+j}(\eta)-g_{\frac{p}{2}+1+j}(\eta)\right\}g_{\frac{p}{2}+1+j}(\eta)\Big], \\
 O_{25} = & \frac{1}{\nu^2}\left[1-(\delta/r)^2\right]^{\frac{p}{2}}\left[\Gamma\left(\frac{1}{2}p\right)\right]^{-1}\sum_{j=0}^{\infty}\frac{1}{j!}\left(\frac{\delta}{r}\right)^{2j}\times \\
 & \times\left[\Gamma\left(\frac{1}{2}p+4+j\right)\left\{\left(g_{\frac{p}{2}+4+j}(\eta)-3g_{\frac{p}{2}+3+j}(\eta)+3g_{\frac{p}{2}+2+j}(\eta)-g_{\frac{p}{2}+1+j}(\eta)\right)G_{\frac{p}{2}+j}(\eta)\right. \right. \\
 & \qquad \qquad \qquad \left. \left. +3\left(g_{\frac{p}{2}+2+j}(\eta)-g_{\frac{p}{2}+1+j}(\eta)\right)^2\right. \right. \\
 & \left. \left. +4\left(g_{\frac{p}{2}+3+j}(\eta)-2g_{\frac{p}{2}+2+j}(\eta)+g_{\frac{p}{2}+1+j}(\eta)\right)g_{\frac{p}{2}+1+j}(\eta)\right\}\right. \\
 & \left. +2(p+5)\Gamma\left(\frac{1}{2}p+3+j\right)\left\{\left(g_{\frac{p}{2}+3+j}(\eta)-2g_{\frac{p}{2}+2+j}(\eta)+g_{\frac{p}{2}+1+j}(\eta)\right)G_{\frac{p}{2}+j}(\eta)\right. \right. \\
 & \qquad \qquad \qquad \left. \left. -\left(g_{\frac{p}{2}+2+j}(\eta)-g_{\frac{p}{2}+1+j}(\eta)\right)^2\right. \right. \\
 & \left. \left. -\left(g_{\frac{p}{2}+3+j}(\eta)-5g_{\frac{p}{2}+2+j}(\eta)+4g_{\frac{p}{2}+1+j}(\eta)\right)g_{\frac{p}{2}+1+j}(\eta)\right\}\right. \\
 & \left. +\left(p^2+12p+23\right)\Gamma\left(\frac{1}{2}p+2+j\right)\left\{\left(g_{\frac{p}{2}+2+j}(\eta)-g_{\frac{p}{2}+1+j}(\eta)\right)\times\right. \right. \\
 & \qquad \qquad \qquad \left. \left. \times\left(G_{\frac{p}{2}+j}(\eta)-2g_{\frac{p}{2}+1+j}(\eta)\right)+g_{\frac{p}{2}+1+j}^2(\eta)\right\}\right. \\
 & \left. +\frac{1}{2}\left(p^2+4p+7\right)\Gamma\left(\frac{1}{2}p+2+j\right)\left\{g_{\frac{p}{2}+2+j}(\eta)-g_{\frac{p}{2}+1+j}(\eta)\right\}^2\right. \\
 & \left. +\left(2p^2+5p+5\right)\Gamma\left(\frac{1}{2}p+1+j\right)\left\{2\left(G_{\frac{p}{2}+j}(\eta)-g_{\frac{p}{2}+1+j}(\eta)\right)g_{\frac{p}{2}+1+j}(\eta)\right. \right. \\
 & \qquad \qquad \qquad \left. \left. -g_{\frac{p}{2}+1+j}^2(\eta)\right\}\right].
 \end{aligned}$$

上の公式の O_{22} の中の $g_{\frac{p}{2}+j-1}(\eta)$ は例外で

$$g_{\frac{p}{2}+j-1}(\eta) = \begin{cases} 0, & p=2, j=0 \text{ のとき} \\ -\frac{1}{2\sqrt{\pi}}\eta^{-\frac{5}{2}}e^{-\eta}, & p=1, j=0 \text{ のとき} \end{cases}$$

を表わす。ただしこれ以外の p と j にたいしては普通の表示である。

$\beta^*(\alpha; p, N, \nu)$ 即ち (3.19) 式の第1項の確率を求めるのに上の漸近公式を適用するには、 $y_1 = x_1 - x_2$, $y_2 = x_3 - x_4$ が互に独立であることに注意して、 $\gamma=2$, $\delta=0$ とおく。即ち O_0, O_1, O_2 は $j=0$ のときのものだけが残る。更に

$$G_{\frac{p}{2}}(\eta) = \int_{\eta}^{\infty} g_{\frac{p}{2}}(t)dt = \int_{2\eta}^{\infty} g_p^*(\chi^2)d\chi^2 = \int_{\chi^2(\alpha/N; p)}^{\infty} g_p^*(\chi^2)d\chi^2 = \alpha/N$$

となり

$$O_0 = (\alpha/N)^2,$$

$$O_1 = \frac{1}{\nu} \frac{p}{2} \left\{ (2\eta+p)(\alpha/N)g_{\frac{p}{2}+1}(\eta) + g_{\frac{p}{2}+1}^2(\eta) \right\} - \frac{1}{\nu} \frac{\chi^2(\chi^2+p)}{2} (\alpha/N) g_{\frac{p}{2}}(\eta)$$

等と簡単になる。

(3.19) の第2項の方は

$$\begin{aligned}
 E\{(x_1-x_2)(x_1-x_2)'\} & = E\{(x_1-x_3)(x_1-x_3)'\} = 2A, \\
 E\{(x_1-x_2)(x_1-x_3)'\} & = A
 \end{aligned}$$

であるから, (3.21) において $\gamma=2, \delta=1$ において計算される. $p=2, \alpha=0.05$ に対して実際に計算した結果を示すと下の表のようになる. 計算は FACOM-128 (リレー自動計算機) によつて行つた.

$$\beta^*(0.05; 2, N, \nu) = M_1(O_0^{(1)} + O_1^{(1)} + O_2^{(1)}) + M_2(O_0^{(2)} + O_1^{(2)} + O_2^{(2)})$$

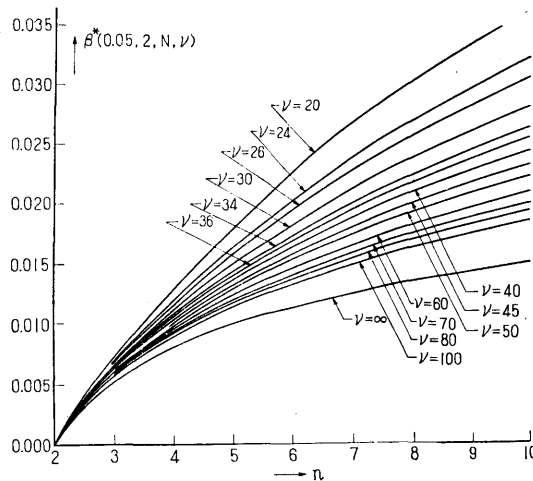
として示す.

第4表 $p=2, \alpha=0.05$ の時の $O_0^{(1)}, O_1^{(1)}, O_2^{(1)}$ の表

| n | $O_0^{(1)}$ | $\nu O_1^{(1)}$ | $\nu^2 O_2^{(1)}$ | $O_0^{(2)}$ | $\nu O_1^{(2)}$ | $\nu^2 O_2^{(2)}$ |
|-----|-------------|-----------------|-------------------|-------------|-----------------|-------------------|
| 3 | * * | * * | * * | 0.0017712 | 0.012769 | 0.019316 |
| 4 | 0.00006944 | 0.001591 | 0.02109 | 0.00066411 | 0.0063927 | 0.015512 |
| 5 | 0.00002500 | 0.0007018 | 0.01189 | 0.00032360 | 0.0039736 | 0.011401 |
| 7 | 0.000005669 | 0.0002068 | 0.004617 | 0.00011448 | 0.0018393 | 0.0074030 |
| 10 | 0.000001235 | 0.00005713 | 0.0004436 | 0.000039570 | 0.00081030 | 0.0044169 |

($n=3$ の時の $O_0^{(1)}, O_1^{(1)}, O_2^{(1)}$ は, $M_1=0$ であるから求める必要がない).

この結果をみると, β^* の第1項の計算の方は, $\nu=20$ の附近で漸近公式 (3.21) による近似があまりよくない. 併し第2項の値にくらべ, order が1つ低いので, β^* 全体としての大勢には左程ひびかないと思われる. $\beta^*(0.05; 2, N, \nu)$ を図に示すと下のグラフのようになる. これよりわかるように, β^* の値は $\alpha = 0.05$ にくらべ著しく有意な値を示しており, R_{MAX}^2 の上方臨界点を評価するには, 第1近似即ち (3.16) に基づくことができない. 第2近似を考慮する必要があるが, それは非常に厄介な計算を必要としかつ R_1^2, R_2^2, R_3^2 と3つの同時確率によるチェックが極



めて困難である. 従つてここでは, 3.2 節の (3.15) にならつて, α に対する $\beta^*(\alpha; p, N, \nu)_x^2$ を用い

$$NP_r\{(x_1 - x_2)' L^{-1}(x_1 - x_2) > s_2^2(\alpha; p, N, \nu)\} = \alpha + \beta^*(\alpha; p, N, \nu) \dots\dots\dots (3.22)$$

によつて $s_2^2(\alpha; p, N, \nu)$ を求めることを考える. $p=1$ の場合既に得られている臨界点と (3.22) によるものとを比較して, (3.22) による方法のチェックとしよう.

これで見ると, R_{MAX}^2 の臨界点の時程よくないが, (3.22) による簡便第2近似は, 実際問題に

第 5 表 $p=1, \alpha=0.05$ の時の $\sqrt{s^2(0.05; 1, N, \nu)}$ と Bio. Table のものとの比較

| ν | | ν | | | | | | | | |
|----------------|-------|-------|------|------|------|------|------|------|------|------|
| | | 10 | 14 | 18 | 20 | 24 | 30 | 40 | 60 | 120 |
| (3.22)によるもの | $n=3$ | 3.91 | 3.73 | 3.63 | 3.60 | 3.55 | 3.50 | 3.46 | 3.41 | 3.36 |
| Bio. Table のもの | | 3.88 | 3.70 | 3.61 | 3.58 | 3.53 | 3.49 | 3.44 | 3.40 | 3.36 |
| (3.22)によるもの | $n=5$ | 4.70 | 4.44 | 4.31 | 4.26 | 4.20 | 4.13 | 4.06 | 4.00 | 3.93 |
| Bio. Table のもの | | 4.65 | 4.41 | 4.28 | 4.23 | 4.17 | 4.10 | 4.04 | 3.98 | 3.92 |

$$n = 3 : O_0^{(2)} = 0.002010, \quad \nu O_1^{(2)} = 0.01155, \quad \nu^2 O_2^{(2)} = 0.00473$$

$$n = 5 : \begin{cases} O_0^{(1)} = 0.000025, & \nu O_1^{(1)} = 0.000535, & \nu^2 O_2^{(1)} = 0.00463 \\ O_0^{(2)} = 0.000382, & \nu O_1^{(2)} = 0.00379, & \nu^2 O_2^{(2)} = 0.00505 \end{cases}$$

たいして充分使用することのできるものである。 $\alpha = 0.01$ の方はもつと近似がよくなり、更に $p > 1$ の時には $p=1$ の時に比べ β^* が小さくなるので近似は実用的に充分なものである。

(統計数理研究所)

最後にこの論文の種々の数値計算を担当していただいた吉田薫嬢に感謝の意を表します。

参 考 文 献

- [1] W. FELLER, *An Introduction to Probability Theory and its Applications, Vol. I*, John Wiley and Sons, New York (1950), Chap. 4, § 1.
- [2] E. S. PEARSON AND H. O. HARTLEY, (ed.), *Biometrika Tables for Statisticians, Vol. I*, Cambridge University Press (1956), Table 29.
- [3] M. SIOTANI, "The extreme value of the generalized distances of the individual points in the multivariate normal sample", *Ann. Inst. Math. Stat.*, Vol. 10. (to be published).
- [4] S. N. ROY. *Some Aspects of Multivariate Analysis*, John Wiley and Sons, New York, (1957), § 14.4.
- [5] H. SCHEFFÉ, "A method for judging all contrasts in the analysis of variance", *Biometrika*, Vol. 40 (1953), pp. 87-104.
- [6] J. W. TUKEY, "Quick and dirty methods in statistics. Part II. Simple analyses for standard designs", *Proc. Fifth Annual Convention, Amer. Soc. for Quality Control*, (1951), pp. 189-197. (筆者未見)