# 講座

## 近代統計学覗見

〔2〕

阪大理学部　小川潤次郎

こゝで以下の所論に必要な補題を2つのべておく。事柄は至極簡単なことである。

**補題1**　$x_1, x_2, \cdots x_k$ は互に独立で共通分散の正規変数とする。このとき $x_1, \cdots x_k$ の二つの一次形式 $l_1x_1+\cdots+l_kx_k$, $m_1x_1+\cdots+m_kx_k$ が互に独立である為の完全條件は二つのベクトル $t=(l_1, \cdots l_k)$ と $r\pi=(m_1, \cdots m_k)$ が直交することである。

実際に $L=\sum l_ix_i$, $M=\sum m_ix_i$ は正規変数であつて、$E(x_i)=pl_i$, $i=1, \cdots k$ とおくと

$$\mathrm{Cov}(L, M)=\sum_{i,j} l_im_j\mathrm{Cov}(x_i, x_j)=\sum_{i=1}^{k} l_im_i\cdot\sigma^2$$

となるからである。

**補題2**　$T_1, \cdots T_k$ は夫々分散が $r_1\sigma^2, \cdots r_k\sigma^2$ なる正規変量で且つその平均値が相等く 0 ならば $T_1, \cdots T_k$ の2次形式

$$Q=\frac{T_1^2}{r_1}+\cdots+\frac{T_k^2}{r_k}-\frac{(T_1+\cdots+T_k)^2}{r_1+\cdots r_k} \tag{22}$$

は互に独立なこの変数の自乗の和に直すことが出来る。

これは R. A. Fisher(4) の補題4のトリビアルな拡張である。

証明：　$T_i/\sqrt{r_i}=\xi_i$, $i=1, \cdots k$ とおくと

$$D^2(\xi_i)=\frac{1}{r_i}D^2(T_i)=\sigma^2, \ i=1, \cdots k$$

となつては共通分散 $\sigma^2$ で $E(\xi_e)=\dfrac{\mu}{\sqrt{r_i}}$ である。

一般化された Helmert の直交変換を (H) とする。

$$\left.\begin{aligned}
\zeta_1&=\frac{(\sqrt{r_1}\xi_1+\sqrt{r_2}\xi_k+\cdots+\sqrt{r_k}\xi_k)}{\sqrt{r_1+\cdots+r_k}}\\
\zeta_2&=\sqrt{\frac{r_2+\cdots+r_k}{r_1+r_2+\cdots+r_k}}\,\xi_1\\
&\quad-\frac{\sqrt{r_1}}{\sqrt{(r_1+\cdots+r_k)(r_2+\cdots+r_k)}}\\
&\quad\times(\sqrt{r_2}\xi_2+\cdots+\sqrt{r_k}\xi_k\\
\zeta_3&=\sqrt{\frac{r_3+\cdots+r_k}{r_2+r_3+\cdots+r_k}}\,\xi_2\\
&\quad-\frac{\sqrt{r_2}}{\sqrt{(r_2+\cdots+r_k)(r_3+\cdots+r_k)}}\\
&\quad\times(\sqrt{r_3}\xi_2+\cdots+\sqrt{r_k}\xi_k)\\
&\cdots\cdots\cdots\cdots\\
\zeta_k&=\sqrt{\frac{r_k}{r_k+r_k}}\,\xi_{k-1}\\
&\quad-\frac{\sqrt{r_{k-1}r_k}}{\sqrt{(r_{k-1}+r_k)r_k}}\xi_k
\end{aligned}\right\} \tag{H}$$

これは直交変換だから補題1によつて $\zeta_1, \cdots \zeta_k$ は互に独立であつて

$$D^2(\zeta_i)=\sigma^2, \ E(\zeta_i)=0, \ i=1, \cdots k \tag{23}$$

そして

$$Q=\zeta_2^2+\cdots+_k^2 \tag{24}$$

これで補題2は証明されたのであるが、こゝで (H) の $\zeta_i$ を $T_i$ の一次形式として陽に求めておこう。

$$\left.\begin{aligned}
\zeta_1&=\frac{1}{\sqrt{r_1+\cdots+r_k}}(T_1+\cdots+T_k)\\
\zeta_2&=\sqrt{\frac{r_2+\cdots+r_k}{r_1(r_1+\cdots+r_k)}}\,T_1\\
&\quad-\sqrt{\frac{r_1}{(r_1+\cdots+r_k)(r_2+\cdots+r_k)}}\\
&\quad\times(T_2+\cdots+T^k)\\
\zeta_3&=\sqrt{\frac{r_2+\cdots+r_k}{r_2(r_2+\cdots+r_k)}}\,T_2\\
&\quad-\sqrt{\frac{r_2}{(r_2+\cdots+r_k)(r_3+\cdots+r_k)}}\\
&\quad\times(T_3+\cdots+T_k)\\
&\cdots\cdots\cdots\cdots\\
\zeta_k&=\sqrt{\frac{r_k}{r_k(r_{k-1}+r_k)}}\,T_{k-1}\\
&\quad-\sqrt{\frac{r_{k-1}}{(r_{k-1}+r_k)r_k}}\,T_k
\end{aligned}\right\} \tag{25}$$

今 (25) を一般的に

$$\zeta_i=l_{i1}T_1+l_{i2}T_2+\cdots+l_{ik}T_k, \quad i=1, \cdots k \tag{26}$$

とおくと $i=2, \cdots k$ については

$$\sum_{j=1}^{k} r_jl_{ij}=0 \tag{27}$$

即ち $\zeta_{i1}(l_{i1}, \cdots l_{ik})$, $i=2, \cdots k$ なる $(k-1)$ コのベクトルは $u=(r_1, \cdots r_k)$ に直交する線型部分空間を張るのである。

吾々の場合に戻つて処理和 $T_1, \cdots T_8$, $T_9$ に対応する (26)の $\zeta$ を $z$ で表わすと

$$\left.\begin{aligned}
z_1&=\frac{1}{\sqrt{48}}(T_1+T_2+\cdots+T_8+T_9)\frac{G}{\sqrt{48}}\\
z_2&=\frac{1}{\sqrt{48\cdot 11}}(11T_1-T_2-\cdots-T_8-T_9)\\
z_3&=\frac{1}{\sqrt{44\cdot 10}}(10T_2-\cdots-T_8-T_9)\\
&\cdots\cdots\cdots\cdots\\
z_9&=\frac{1}{\sqrt{20\cdot 10}}(\qquad 2T_8-T_9)
\end{aligned}\right\} \tag{28}$$

例えば $B_j$ については

$$B_j=y_{1jk_1}+y_{2jk_2}+\cdots+y_{8jk_8}$$
$$+y_{9jk_9}'+\cdots+y_{9jk_9}'''$$

の形であるから$(z_2,\cdots z_9)$の各ベクトルと$B_j$は直交し、従つて補題1によつて

$(z_2,\cdots z_9)$ と $(B_1, B_2, B_3, B_4)$

は互に独立である。$B_j$に対する$\zeta$を$u_j$とすると

$$\left.\begin{aligned} u_1&=\frac{1}{\sqrt{48}}(B_1+B_2+B_3+B_4)=\frac{G}{\sqrt{48}}\\ u_2&=\frac{1}{\sqrt{48\cdot 3}}(3B_1-B_2-B_3-B_4)\\ u_3&=\frac{1}{\sqrt{36\cdot 2}}(\quad 2B_2-B_3-B_4)\\ u_4&=\frac{1}{\sqrt{24\cdot 2}}(\qquad B_3-B_4)\end{aligned}\right\}\quad(29)$$

そして

$$\frac{B_1{}^2+\cdots+B_4{}^2}{12}-\frac{G^2}{48}=u_2{}^2+u_3{}^2+{}_4{}^2\qquad(30)$$

$(z_2,\cdots z_9,\ u_2,\ u_3,\ u_4)$ の Ortho-cemplement の base を取つて$v_1,\cdots v_{36}$ とすれば (21) は

$$\sum_{ijk}y_{ijk}{}^2=\frac{G^2}{48}+\sum_2^9 z_i{}^2+\sum_2^4 u_j{}^2+\sum_1^{36}v_i{}^2\qquad(31)$$

即ち

$$\sum_{ijk}(y_{ijk}-m-t_i-b_j)^2=\sum_1^{36}v_i{}^2\qquad(32)$$

こゝで$v_l$は$T_1,\cdots T_9,\ B_1,\cdots B_4$の張る部分空間に直交するから今

$$v_l=\sum l_{ijk}^{(l)}y_{ijk}$$

とおいて、との共変量を0とおくと

$$\sum_{i,j}l_{ijk}^{(l)}=\sum_{j,k}l_{ijk}^{(l)}=0\qquad(33)$$

つまり残差平方和の独立成分となるべき$y_{ijk}$の一次形式についてはつねに (33) が成立たねばならないことが判る。

(31)式によつて分散分析表の自由度の分割が明かにされたのである。即ち

第4表　分散分析表

| | 自由度 (d.f.) | 平方和 (s.s.) | 平均平方和 (m.s.) |
|---|---|---|---|
| 処　理 | 8 | 157448 | 19681 |
| ブロック | 3 | 289427 | 96476 |
| 誤　差 | 36 | 544690 | 15130 |
| 合　計 | 47 | 991565 | |

こゝの計算で注意すべきことは先ず G を求め、次に $m=G/48$ を次に $G^2/48=Gm$ として一これを補正項 (Correction factor) という一予め計算しておく。次に平方和

$$\sum_{ijk}y_{ijk}^2,\quad \frac{T_1{}^2+\cdots+T_8{}^2}{4}+\frac{T_9{}^2}{16},\quad \frac{B_1{}^2+\cdots+B_4{}^2}{12}$$

を計算してそれらの補正として

$$\sum_{ijk}y_{ijk}^2-Gm,\quad \frac{T_1{}^2+\cdots+T_8{}^2}{4}+\frac{T_9{}^2}{16}-Gm,$$

$$\frac{B_1{}^2+\cdots+B_4{}^2}{12}-Gm$$

を求めこの三者から

$$\text{残差平方和}=\sum_{ijk}y_{ijk}^2-G^2/48$$
$$-\left(\frac{T_1{}^2+\cdots+T_8{}^2}{4}+\frac{T_9{}^2}{16}-G^2/_{88}\right)$$
$$-\left(\frac{B_1{}^2+\cdots+B_4{}^2}{12}-G^2/48\right)=544690$$

を求めるのがよい。

今若し処理効果なし、即ち統計的仮説なる仮説

$$H_T:\ \tau_1=\cdots=\tau_8=\tau_9=0$$

を立てるとこの仮説$H_T$の下では

$$\frac{(\text{処理平方和})/(\text{その自由度}8)}{(\text{残差平方和})/(\text{その自由度}36)}$$

は自由度 (8.36) なるF分布に従う統計量でこれを吾々は $F^8{}_{36}$ で表わすことにする。第4表から

$$F^8{}_{36}=\frac{19681}{15130}=1.30$$

$F^8{}_{36}$分布の右側のパーセント点を調べると

$$F^8{}_{36}(5)=2.21,\quad F^8{}_{36}(1)=3.04$$

だから仮説$H_T$は有意水準5％で有意ではない。即ち殺虫剤は全体として見るとき効果がないということになる。これは釈のデーター文を用いた分析であつて、春にミミズの数が多かつたプロットは殺虫剤の如何に拘らず秋にも多くなつているという事実が考慮されていない為でもあり、又ミミズの数という変量は元来離散的であるのにそれを正規変量と見なすことにも問題がある。例えば仮りに1瓦の土壌の中にいるミミズの数に対して母平均人のポアツフン分布を仮定出来るならば400瓦の土壌の中のミミズの数は平均400入のポアツソン分布と考える一今簡単の為に Centagion は無しておく。そうするとこの変量Xに対しては$\lambda\gg 0$のとき

$$P\left(\frac{X-400\lambda}{\sqrt{400\lambda}}>x\right)\approx\frac{1}{\sqrt{2\Pi}}\int_{-\infty}^{x}e^{-\frac{t^2}{2}}dt$$

だから平均値が充分大きいと考えて正規分布を仮定することはよいが共通分散に問題があるこのときは一般に分散が平均値に関係するので共通分散変量を得るには適当な変数変換を考えることが必要になる。

今はこのまゝでデータの分析をつづけて見よう。以下処理平方和の更に細い分割を考察して見よう。

(1CN), (1CS),…等は処理 1CN, 1CS,…に対するプロ

ットのミミズの数の和(2CN), (2CS)…等も同様とする。第2表から

$S_0=0$ 処理に対応するミミズの数の和=5858

$S_1=(1CN)+(1CS)+(1CM)+(1CK)=4317$

$S_2=(2CN)+(2CS)+(2CM)+(2CK)=4505$

とおくと，勿論

$$G=S_0+S_1+S_2$$

である。殺虫劑1容量に対する平方和は

$$\frac{(1CN)^2+(1CS)^2+(1CM)^2+(1CK)^2}{4}-\frac{S_1^2}{16}$$

$$=\frac{(1066)^2+(928)^2+(1431)^2+(892)^2}{4}-\frac{(4317)^2}{16}=45461$$

又2容量に対する平方和は

$$\frac{(2CN)^2+(2CS)^2+(2CM)^2+(2CK)^2}{4}-\frac{S_2^2}{16}$$

$$=\frac{(1265)^2+(877)^2+(1241)^2+(1122)^2}{4}-\frac{(4505)^2}{16}=23641$$

さて

$$w_1=S_1/4$$

$$w_2=\frac{1}{\sqrt{48}}[3(1CN)-(1CS)-(1CM)-(1CK)]$$

$$w_3=\frac{1}{\sqrt{24}}[2(1CS)-(1CM)-(1CK)]$$

$$w_4=\frac{1}{\sqrt{8}}[(1CM)-(1CK)]$$

$$\chi_1=S_2/4$$

$$\chi_2=\frac{1}{\sqrt{49}}[3(2CN)-(2CS)-(2CM)-(2CK)]$$

$$\chi_3=\frac{1}{\sqrt{24}}[2(2CS)-(2CM)-(2CK)]$$

$$\chi_4=\frac{1}{\sqrt{12}}[(2CM)-(2CK)]$$

とおくとこれら8コの変数は互に独立であつて

$$1容量平方和=w_2^2+w_3^2+w_4^2$$

$$2容量平方和=\chi_2^2+\chi_3^2+\chi_4^2 \qquad (33)$$

で夫々自由度3である。(33)の分解は勿論一意的ではなく、これら3変数に直交変換を施こしたものでもよい。処理平方和の自由度は全体で8でその内(33)によつて6が考えられるから残りは2である。それは勿論（$w_2$ $w_3$, $w_4$, $\chi_2$, $\chi_3$, $\chi_4$）の張る部分空間に直交すべきであるから

$$S_0,\ S_1,\ S_2$$

の一次結合としての Comparison をとればよい。

$$\eta_1=S_2-S_0=-1353$$
$$\eta_2=S_2-2S_1+S_0=1729 \qquad (34)$$

を取つて見る。その意味は、$\eta_1$ は殺虫剤の全体としての効果で $\eta_2=(S_2-S_1)-(S_1-S_0)$ は、若し仮りに $\eta_2=0$ なら

$$S_2-S_1=S_1-S_0$$

容量に対して効果が線型ということになる、それ故 $\eta_1$ を平均効果 $\eta_2$ を平均効果の強さに対する曲率と名ずけておこう。

以上の如き Comparison に対する分散分析表は下の第5表のようになる。

第5表　処理和の分割

| | 自由度 | 平方和 | 平均平方和 | F |
|---|---|---|---|---|
| 処理 | 8 | 157449 | 19681 | $F_{36}^{8}=1.30$ |
| 平均効果 | 1 | 57207 | 57207 | $F_{36}^{1}=3.78$ |
| 平均効果の強さに対する曲率 | 1 | 31140 | 31140 | $F_{36}^{1}=2.05$ |
| 1容量間 | 3 | 45461 | 15154 | $F_{36}^{3}=1.00$ |
| 2容量間 | 3 | 23641 | 7880 | $F_{36}^{3}=1.92$ |

F欄から見るように、この結果はいずれも有意ではない。

次に $w_2^2+\cdots+w_4^2+\chi_4^2$ 即ち1容量2容量間の平方和を異つた仕方で分割して見よう。

先ず前述のことから (1CN)−(1CS), (2CN)−(2CS) の形の Comparison は $S_0$, $S_1$, $S_2$ と独立だから $\eta_1$, $\eta_2$ と独立である。さて

$$\frac{[(1CN)-(1CS)]+2^2\cdot\frac{[(2CN)-(2CS)]}{2}}{1+2^2}$$

$$=\frac{1}{5}\{[(1CN)+1(2CN)]-[(1CS)+2(2CS)]\}$$

はCNとCSの効果の差の推定値であつて(1CN)+2(2CN) (1CS)+2(2CS), (1CM)+2(2CM), (1CK)+2(2CK) の張る部分空間に属し、これは $\eta_1$, $\eta_2$ に直反する。

$$\frac{[(1CN)+2(2CN)]^2+\cdots+[(1CK)+2(2CK)]^2}{5\times4}-\frac{(S_1+2S_2)^2}{20\times4}$$

$$=\frac{(3596)^2+(2682)^2+(3913)^2+(3136)^2}{20}-\frac{(13327)^2}{80}=43408$$

の自由度は3である。上の計算に必要な補助表は次の第6表である。

第6表　(1容量)+2(2容量)

| CN | CS | CM | CK | 計 |
|---|---|---|---|---|
| 3596 | 2682 | 3913 | 3136 | 13327 |

この平方和の独立成立は次の如くである。

$$\psi_2=\frac{1}{\sqrt{240}}\Big\{3[(1CN)+2(2CN)]$$

$$-[(1CS)+2(2CS)]-[(1CM)+2(2CM)]$$

$$-[(1CK)+2(2CK)]\Big\}$$

$$\psi_2=\frac{1}{\sqrt{120}}\Big\{\quad 2[(1CS)+2(2CS)]$$

$$-[(1CM)+2(2CM)]-[(1CS)+2(2CK)]\Big\}$$

$$\psi_3=\frac{1}{\sqrt{40}}\Big\{\quad [(1CM)+2(2CM)]$$

$$-[(1CK)+2(2CK)]\Big\} \qquad (33)$$

次に

$$[(2CN)-(1CM)]-[(1CN)-(0)]$$

$$=[(2CN)-2(1CN)+(0)]$$

はCNの効果のその薬剤の強さに対する変化の曲率であつて (0) と $-2(1CN)+(2CN)$ の一次形式であつて $-2(1CN)+(2CN)$, $-2(1CS)+(2CS)$, $-2(1CM)+(2CM)$, $-2(1CK)+(2CK)$ が $\eta_1$, $\eta_2$, $\psi_2$, $\psi_3$, $\psi_4$ に直交し従つてこれらと独立なることは明かであろう。

第7表　2(1容量)−(2容量)

| CN | CS | CM | CK | 計 |
|---|---|---|---|---|
| 867 | 979 | 1621 | 662 | 4129 |

第7表から

$$\frac{[2(1CN)-(2CN)]^2+\cdots+[2(1CK)-(2CK)]^2}{20}$$

$$-\frac{(2S_1-S_2)^2}{80}$$

$$\frac{(867)^2+(979)^2+(1621)^2+(662)^2}{20}$$

$$-\frac{(4129)^2}{80}=25693$$

これの独立成分は

$$\psi_2=\frac{1}{\sqrt{240}}\Big\{3[2(1CN)-(2CS)]$$

$$-[2(1CS)-(2CS)]-[2(1CM)-2CM)]$$

$$-[2(1CK)-(2CK)]\Big\}$$

$$\psi_3=\frac{2}{\sqrt{120}}\Big\{\quad 2[2(1CS)-(2CS)]$$

$$-[2(1CM)-(2CM)]-[2(1CK)-(2CK)]\Big\}$$

$$\psi_4=\frac{1}{\sqrt{40}}\Big\{\quad [2(1CM)-(2CM)]$$

$$-[2(1CK)-(2CK)]\Big\} \qquad (36)$$

このような処理和の分割に対して次の第8表を得る。このF欄から見るようにいずれも有意ではない。これは春のミミズの数を全く無視したからであつて殺虫剤の効果がないと云うわけではないのである。

第9表　処理和の分割

| | 自由度 | 平方和 | 平均平方和 | F |
|---|---|---|---|---|
| 処　　理 | 8 | 157449 | 19681 | |
| 平均効果 | 1 | 57207 | 57207 | |
| 平均効果の強さに対する変化率 | 1 | 31140 | 31140 | |
| 各剤の効果間の変動 | 3 | 43408 | 14469 | $F_3^{36}=1.03$ |
| 各剤の効果の強さに対する変化率間の変動 | 3 | 25693 | 8564 | $F_3^{36}=1.76$ |

本稿で解説した実例は農事試験のデーターであるが、こゝに述べた解析方法は農事試験に限らず広い応用性があることは云う迄もない。

それからこゝでは所謂 Normal theony を直接ミミズの数というような enumerative なデータに適用したのであるが一般にポアツソン分布の母平均が大となれば漸近的に正規分布になるが、そのとき分数が平均に比例するのでこの難点をさけるには $\sqrt{\chi}$ なる平方根変換等を用いる方がより合理的であろう。まあしかし本稿では変量分析法の初歩的な解説という意味でこのまゝにしておいたのである。(1953. 12. 1)

# (協会だより)

## 昭和機械第3工場完成

伸線機専門メーカーKK昭和機械工作所（大阪市西淀川区御弊島中2）では、パキスタン向伸線工場のプラント輸出、其他受注多数のためかねて新工場増設の計画中であつたが、此度同社本社工場隣接地を買収第3工場を新設した。同工場は建坪200坪鉄骨建築で5屯及3屯天井グレーン及自動旋盤、シカル盤等工作機も多数新設し、同社の増産体制の新威力となるであろう。