

統計講習会テキスト

|    |    |
|----|----|
| 経営 | 47 |
| 測定 | 13 |

# 統計解析

昭和31年12月

林業試験場経営部

# 目 次

|                                                                                                                                                                                                                                 | 頁  |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 第 1 章 緒 論 .....                                                                                                                                                                                                                 | 1  |
| 1. 統計学と科学者、2. 数学的な考え方は統計学の基礎である                                                                                                                                                                                                 |    |
| 第 2 章 変動、統計量及び度数表 .....                                                                                                                                                                                                         | 13 |
| 1. 変動、2. 母集団、3. 標本、4. 算術平均、5. 標準偏差、6. 平均値の標準誤差、<br>7. 分散、8. 度数表、9. 度数表の作成、10. 度数表のグラフ表示、11. 度数表か<br>らの平均値と標準偏差の計算、12. 変動係数                                                                                                      |    |
| 第 3 章 理論的度数分布 .....                                                                                                                                                                                                             | 23 |
| 1. 生物的量の一つの実際の分布の特性、2. 正規分布の実際の偏差、3. 正規分<br>布、4. 正規曲線のあてはめ、5. 非正規型、6. 非正規型の尺度、7. 標本平均の分<br>布、8. 二項分布、9. ポアソン分布、                                                                                                                 |    |
| 第 4 章 有意性の検定 .....                                                                                                                                                                                                              | 39 |
| 1. 有意性の検定の論理、2. いづれかの方向における偏倚の検定、3. $t$ 分布を<br>用いる有意性の検定、4. 変量が対でない場合の検定の応用、5. 例 4-1 変量<br>が対でない場合の $t$ 検定の応用、6. 例 4-2 $n_1 \neq n_2$ の場合、二つの母集団<br>の分散が違っていると仮定したときの $t$ 検定の応用、7. 帰無仮説、8. 信頼限界                                 |    |
| 第 5 章 分散分析 .....                                                                                                                                                                                                                | 51 |
| 1. 定義、2. 基礎原理、3. 有意性の検定、4. 直交性、5. 平方和及び自由度の割当<br>6. 例 5-1 変量の二重分類、7. 実験誤差、8. 妥当な誤差の選択、9. 交互作用<br>10. 高次の交互作用、11. 平方和の計算法、12. 個々の自由度に応じて偏差平<br>方和を成分に分割する方法、13. 例 5-2 妥当な誤差の選択、14. 例 5-3<br>変量の 3 重分類、15. 分散分析の限界、16. 変換         |    |
| 第 6 章 直線回帰分析 .....                                                                                                                                                                                                              | 84 |
| 1. 一般的な観測、2. 回帰方程式の誘導、3. 回帰線の適合度、4. 回帰係数の信頼<br>限界、5. 推定量 $Y$ の標準誤差、6. 両方の変量が共に誤差を 5 ける場合の直線回<br>帰、7. 群にとりまとめていない資料に対する直線回帰の計算、8. 回帰表に配列<br>した資料に対する直線回帰、9. 回帰表に対する計算、10. 例 6-1 対の値の<br>系列での回帰分析、11. 例 6-2 回帰表にした大きな対の系列に対する回帰分析 |    |



## 第 7 章 相 関 ..... 100

1. 共変動、2. 相関の定義、3. 相関の尺度、4. 相関係数の範囲および解釈、  
5. 相関係数の抽出分布、6. 相関係数の計算

## 第 8 章 偏回帰と重回帰および偏相関と重相関 ..... 109

1. 偏回帰および重回帰の問題、2. 偏回帰法、3. 偏回帰の意味、4. 偏回帰係数の計算、5. 例 8-1 偏回帰係数の計算、6. 従属変数の色々な組に対する偏回帰係数の計算、7. 例 8-2 ガウス乗数を用いた偏回帰係数の計算、8. 偏回帰係数の有意性の検定、9. 変量の削除、10. 例 8-3  $Y, X_1, X_2$  および  $X_3$  に対する回帰分析からの  $X_2$  の削除、11. 偏相関と重相関、12. 例 8-4 偏相関係数および重相関係数の計算、13. 重相関係数の解釈と有意性、14. 特別な応用

## 第 9 章 共分散分析法 ..... 125

1. 共分散分析の原理、2. 誤差のコントロールへの共分散法の応用、3. 回帰係数の系列に対する不斉一性の検定、4. 誤差のコントロールに対する変量の選択  
5. 積和の計算、6. 例 9-1 共分散分析、7. 例 9-2 回帰の不斉一性の検定

## 第 10 章 非直線回帰 ..... 135

1. 直線および非直線回帰成分の分析、2. 回帰成分の分析の実例、3. 非直線回帰における基礎因子、4. 曲線を適合するための前提、5. 傾向が非直線的である場合の回帰方程式の選択、6. 多項式式のあてはめ、7. 直交多項式、8. 例 10-1  $Y$  の値が等荷重でない資料への多項式のあてはめ、9. 直交多項式あてはめのためのフィッシャーの加算法、10. 例 10-2 直交多項式の適合—加算法、  
11. フィッシャー及びイエーツの  $\xi$  の表を用いた直交多項式のあてはめ、  
12. 例 10-3 フィッシャーおよびイエーツの  $\xi$  の表による多項式のあてはめ、

## 第 1 章 結 論

### 1 統計学と科学者

統計学は科学者にとって特に関心をもたれている課題である。科学者は統計学は彼が日常生活において見出しそうな分野の大抵のものに実際的な応用性がある事を知っているし、とにかく彼の生涯を通じて自分の身のまわりに密接なつながりをもっている自然現象のよりよい理解と評価の助けとなるであろう。第一の場合においては、統計学は変動の性質を理解する場合の助けとなる。これは一組の条件の下に観測された現象は他の組の条件の下では決して重複しないので、実際的な応用では重要である。調査研究者はすべての変動因子を考慮に入れ、彼の実験を計画しそれによつて彼の結果を解釈しなければならないので、変動の性質を理解におくことは調査の仕事において重要である。変動を生ずる因子は生産品の質に変動をおこさせるから、産業に対して重要である。

第二の場合には、統計学はそれ自体意味を持たない個々の測定のおつまりから一般法則を如何にして導くかを教えてくれる。運動粒子によつて作られた模型を追跡すると、早い速度で色々な方向に粒子が飛び交っている事が分る。任意の点では、ある一つの粒子は与えられた速度で与えられた方向に移動しているといふことができる。一つの単一粒子としての動向は全く予測できない。もしも考えている物質が密閉した部屋内のガスであれば、温度が上昇すると粒子は力を増して部屋の壁を圧迫し、したがつて壁上の圧力は増加する。圧力を測定すると、ある境界内では圧力の変化は温度に比例するといふことがわかる。換言すれば、個々の粒子の移動はでたらめであるにもかかわらず、粒子の数が非常に多いので生じた結果は一様で一定している。あらゆる研究者が研究室で観察したことのあるこの簡単な例は統計的法則の一つの例である。これと類似した法則はその結果は上の例の様に一定ではないけれども実験から導くことができる。二つの値の処理の間の比較が行われたとする。土壌及び天候の変動のために、穀物収穫の処理の効果に関してこれ等二つの処理間の差は一定ではないが、もしも真の差が存在し、充分な数で比較できれば、この真の差は明らかとなるであろう。そのときの結果はもしも全実験が充分な回数くりかえされたとすれば、類似の結果が得られるといふことが確信されるから一般的な法則として表わすことができる。

統計学の基礎的な機能の一つは非常に多くの測定によつて表わされた事実をまとまつた形で表示することである。したがつて平均値がしばしば一つの群を表わすために最良な単一の値として用いられる。観測群の変動性は標準偏差の値によつて表わされる。肥料の量を増すと収穫が増す様な傾向はまづ第一に図的な形で表わされ、最後に回帰係数の値によつて表わされる。これは非常に多くの情報を単一のグラフ又は単一の値に圧縮する統計学の記述的な価値及びその力を強調するものである。生物学的現象を取扱う際に、変動が例外であるというよりもむしろ一つの規則である様な場合には、統計学のこの様な面は広い応用性をもっていることは明らか



である。

研究しようという目的をもつて更に何か進んだ学問をしようと考えている研究者にとつては統計学の勉強は、その必要性は明らかであろう。これに対する理由はすでに述べたことから明らかであろうが、よりはつきりした理由は統計的な原則は結果の解明を含むと同様に実験の有効性および経済的な計画を含むということである。統計学者が統計的方法は実験計画に必要であるという原則を明らかにでき、また一般的に受け入れることができたのはごく最近のことに過ぎない。それ迄は実験が行われ、資料が表示された後で調査研究者が統計学者に相談するという様なことはごく普通の出来事であつた。多くの場合、その結果はほとんど価値がなく、他の場合には実験はもつと容易な方法で行うことができたであろうし、または与えられた量の努力に対してはより多くの情報を得ることができたであろうことがわかつた。

この本の大部分は実験計画および分析に対する統計学の応用を取扱っているが、統計学を応用する場合の最も重要な分野であると思ひがたいとはいけない。統計学の原則およびそれが取扱う色々な理論はすべての応用に対して共通である。ある特殊化された方法は他のものよりも仕事の一つの分野においてより普通なものであるが、それらが何時全く違つた分野で応用できるかということは誰にも決してわからない。章を追つてゆくと、工業製品の品質管理の統計的原則の応用がわかるであろう。これは統計学の重要な部門で多くの研究者が最大の関心をもっているものである。

統計学の一つの面はそれから標本が抽出される群すなわち母集団についての情報を得るために標本をとる正しい方法を取扱う。この様な抽出問題が一般的なものである。それらは事務製造、社会研究、経済、土地評価、調査実験、医学および多くの他のものに应用される。例えばある例では色々な社会層における人間の社会調査を指導する方法とまた他の例で与えられた野外で平方フット当りカリギリスの卵の平均数を決定するために採用する抽出方法との間には類似性がある。

一般的な結論は統計的原則は非常に広い応用通性をもっているということである。従つて、生徒の主として関心をもたなければならないのは限られた分野において特別な価値をもつ技術というよりはむしろこれ等の原則の研究であるということである。特に科学者は非常に広い対象に関心があり、したがつて特にその基礎的な原則に関しての統計的な知識は明らかに必要である。

## 2. 数学的な考え方は統計学の基礎である

統計学の研究を企画する前に、研究者が必要とする数学的な訓練に対して非常に多くの論議がある。この本での統計的方法および論議は進んだ数学の知識を必要としないが、初等的な代数、幾何、三角の基礎数学のしつかりした基礎が必要である。確率の代数は特に重要である。復習の目的のため以下二三の重要な数学的な考え方を論じよう。

## a. 確率

数学者は確率の定義に関してくどくどと論議してきた。我々の目的に対してはIntroduction to the Theory of Probability and Statistics において Arley and Buch が与えた定義をうけ入れるのが便利である。この定義は実際には二三の言葉で述べるよりはむしろ記述しなければならない概念である。n事象の系列の内  $n_A$  がある特別なクラスに属すると仮定しよう。例えば農場地帯で観測された穀物品が一つの事象である。これらの品の中でいくらかは小麦であり、したがつて特別なクラスに属する。他の燕麦、裸麦等においても然りである。観測したn個の品の中  $n_A$  が小麦畑であれば

$$f(A) = n_A / n \quad (1)$$

を観測したシリーズにおけるAの相対度数という。さてnが増すと  $f(A)$  が一定になつてくるといふことが実際に観測をくりかえすとわかる。小標本に対しては変動が大であるが、大標本では  $f(A)$  の観測値はある中央の値の附近に集まるであろう。観測数を増すとこの集中する傾向は無作為法則と呼ばれるものである。観測値は一定すなわち真値であるところのある物理常数の推定量であるとして理解される。

Arley and Buch を引用すると、“相対的度数が実験値であるこの物理常数は事象Aの確率と呼ばれ、 $p(A)$ をもつて表わす”。

上の定義は確率の正確な値を見出すというわけにはゆかないが、その内部で条件についてある仮定をすることによつて確率を導くことができるものである。例えば貨幣を投げるとき、表と裏は同じ様に出るすなわちより特殊化していえば、一つのものが他のものよりも多くおこることを仮定するにはよい理由がないとすることが合理的である。したがつて非常に多くの回数くりかえすと表が出る相対度数は近似的に投げた場合の半分は表、半分は裏であるから  $1/2$  に近づくといふことが期待される。つまり表の確率は次の簡単な定義から導くことができる。すなわち

$$p(A) = \frac{n_1}{n_1 + n_2} \quad (2)$$

ここに  $(n_1 + n_2)$  は事象が起り得るすべての可能な方法で、 $n_1$  は特別な事象Aが起り得る数である。この場合には  $n_1 = 1$   $n_2 = 1$  で  $p(A) = 1/2$

統計的理論においては、必ずしも正確に確率を評価する必要はないということは一般的に真である。典型的な統計的問題の例として偏倚の程度及び存在を定めるためのサイコロの検定を考えよう。もしもサイコロを1200回投げればまたそれが正しいサイコロであれば、期待値は1が200回出る。この特別なサイコロで1が出る正確な確率は必要でないことは注意を要する。確率はサイコロが真であるとしてたてた仮説の結果として生じてくる。そのときサイコロが非常に多くの回数なげられ、1がでた理論度数と実際度数との間に一致が認められる。



もしも真のサイコロでの期待結果とその結果が一致すればこの特別な場合については少なくとも偏倚はないと結論する。一方もしも結果が期待から全くはづれていれば、サイコロは偏倚があると結論する。

#### b. 確率の基礎的な規則

これらの規則の中の最初のもは次の様にごく簡単に述べれる。 $P(A)$ を事象Aの確率、 $P(B)$ を事象Bの確率としよう。そのときもしも一つの事象が起つたとき他のものが起らないならばA又はB何れかがおこる確率は

$$P(A, B) = P(A) + P(B) \quad (3)$$

サイコロの問題を例にとると、Aは1の目の生起であり、Bは5の目の生起である。別々の確率は  $P(A) = 1/6$  及び  $P(B) = 1/6$  である。一度投げて1であれば5は起り得ない。また5であれば1は起り得ない。そのとき1又は5の何れかを得る確率は  $1/6 + 1/6 = 1/3$  である。この規則は如何なるお互に排斥な事象の場合にも拡張できる。したがって三つの事象の場合には

$$P(A, B, C) = P(A) + P(B) + P(C) \quad (4)$$

この規則を1, 2, 3, 4, 5又は6の目が出る確率に應用すると

$$P = 1/6 + 1/6 + 1/6 + 1/6 + 1/6 + 1/6 = 1$$

これは六つの可能な事象の中一つまたは他のものは必らず起るはずであるから明らかである。

第二の一般的な定理はお互に排斥でない事象の場合である。すなわち例えば二つのサイコロを投げて1及び5を得る確率をきめるような場合である。ここに二つの事象は独立である。もしも事象をA及びBとすれば共に起る確率は

$$P(A+B) = P(A) \times P(B) \quad (5)$$

サイコロの問題では

$$P(1+5) = 1/6 \times 1/6 = 1/36$$

この問題の変形は第二の事象が第一の事象の生起について条件つきである様な場合に遭遇する。例えばカードの一束から2枚ひいて2枚共エースである確率を考えよう。第一のエースをひく確率は  $4/52$  で第二のエースをひく確率は第一のエースをひいた後であるから  $3/51$  であることに注意せよ。したがって2枚のエースをひく確率は  $4/52 \times 3/51 = 12/2652 = 1/221$  である。

一般的な論述はこの型の問題に対して次の様に行うことができる。

$$p = p(\text{第一の事象がおこる確率})$$

$$+ p(\text{第一の事象がおこつて後第二の事象がおこる確率})$$

$$\text{すなわち } p = P(A) \times P_A(B) \quad (6)$$

ここに  $P_A(B)$  はAなる条件の下にBのおこる確率を表わす。

#### C. 順列及び組合せ

4つの文字 a, b, c, d をとりこれ等の中2つの文字の順列は ab, ba, ac, ca,

ad, da, bc, cb, bd, db, cd, dc, である。同じ文字を含んでいても順序が異なれば順列が異なることに注意せよ。n個のものから同時にr個のものをとる順列の全数は次式で与えられる。

$${}_n P_r = n(n-1)(n-2) \cdots (n-r+1) \quad (7)$$

もしもすべてのn個のものをとるとすれば

$${}_n P_n = n(n-1)(n-2) \cdots 3, 2, 1 \quad (8)$$

連続した積  $n(n-1)(n-2) \cdots 3, 2, 1$  はnの階乗として知られているし、これを  $n!$  なる記号で表わす。

7つの文字 aabbbcc の順列の数をきめる際のように、配列されるべきすべてが異つたものでなければ2つのa, 3つのb, 2つのcがあることに注意を要する。そのときの順列の全数は

$$\frac{7!}{2!3!2!} = 210$$

もしもnで文字の数を表わし、p, q, r, ... で各種の数を表わすと、すべてが同時にとられる順列の総数に対する一般公式は

$$P = \frac{n!}{p!q!r!} \quad (9)$$

組合せは種類に関してであるが、順序に関係なく配列できるもののグループに関するものである。したがって4つの文字 a, b, c, d のすべての組合せは ab, ac, ad, bc, bd, cd である。n個のものから同時にr個のものをとると  $\binom{n}{r}$  によつて表わされる組合せの数は

$$\binom{n}{r} = \frac{n(n-1)(n-2) \cdots (n-r+1)}{r!} = \frac{n!}{r!(n-r)!} \quad (10)$$

0, 1, 2, ... 又はnを同時にとつたn個の異つた物の組合せの全数は  $2^n$  である。

順列および組合せは確率問題を取扱ふ際に重要である。したがってしばしば数理統計学の手段に取扱われる。順列を含む簡単な問題はツボの中に1から12迄の数がかかっている紙片を仮定することである。もしも2枚の紙片を無作為に引き出して紙片上の数字の和が5となる確率はいくらか? 二枚の紙片は上の方程式(7)を用いて12枚から  $12(12-2+1) = 12 \times 11$  通り引き出すことができる。和が5になるのは数が4と1, 3と2, 2と3, 又は1と4の4通りである。そのとき必要な確率は  $4/132 = 1/33$  である。

組合せを含む重要な定理はn個の貨幣をなげた際に表の違つた数が出る確率の様な問題を考えるときに起つてくる。4枚の貨幣がなげられるとしよう。可能な表及び裏の違つた数は明らかに4枚表で0枚裏; 3枚表で1枚裏; 2枚表で2枚裏; 1枚表で3枚裏; 0枚表で4枚裏である。各場合の確率をきめるため下に示す様に4枚の貨幣が1, 2, 3, 4なる数の位置を占



める列に配列されると考えると便利である。0なる表から始めると配列の総数は  $\binom{4}{0} = 1$ 。1つの表は  $\binom{4}{1} = 4$ 通りである。同様に2つの表は  $\binom{4}{2} = 6$ 通り、3および4の表も同様である

|                | 位 置 |   |   |   | $\binom{n}{r}$ | p              |
|----------------|-----|---|---|---|----------------|----------------|
|                | 1   | 2 | 3 | 4 |                |                |
| 0, 1, 2, 3及び4の | 0   | 0 | 0 | 0 | 1              | $\frac{1}{16}$ |
| 表の可能な配列の       | H   | 0 | 0 | 0 | 4              | $\frac{1}{4}$  |
| 内の一つ           | H   | H | 0 | 0 | 6              | $\frac{3}{8}$  |
|                | H   | H | H | 0 | 4              | $\frac{1}{4}$  |
|                | H   | H | H | H | 1              | $\frac{1}{16}$ |
|                | 計   |   |   |   | 16             | 1              |

最後の欄にあたえてある確率pは容易に得られる。

同様な問題に対して、我々は確率を直接に書くことができる。分母は  $2^n$  である。ここにnは投げられる貨幣の数を表す。そして分子は

$$1. \quad \binom{n}{n-1}, \binom{n}{n-2}, \dots, \binom{n}{n-r}, \dots, \binom{n}{1}, 1 \quad (11)$$

すなわち

$$1. \quad n, \frac{n!}{2!(n-2)!}, \dots, \frac{n!}{r!(n-r)!}, \dots, n, 1$$

である。

#### d. 二項定理

この定理は通常次の形のものの一つで述べられる。

$$(q+p)^n = q^n + \binom{n}{n-1} q^{n-1} p + \binom{n}{n-2} q^{n-2} p^2 + \dots + \binom{n}{n-r} q^{n-r} p^r + \dots + p^n \quad (12)$$

$$= q^n + n q^{n-1} p + \frac{n(n-1)}{2!} q^{n-2} p^2 + \dots + \frac{n!}{r!(n-r)!} q^{n-r} p^r + \dots + p^n \quad (13)$$

これは確率問題で非常によく用いられる定理である。簡単な例としてpを一つの事象が起る確率とし、 $q=1-p$ を起らない確率としよう。そのときn回の試行において正確に事象がr回起る確率は

$$p(r) = \frac{n!}{r!(n-r)!} q^{n-r} p^r \quad (14)$$

これは明らかに二項展開の一般項である。換言すれば、二項展開の次々の項はn回の試行に於ける事象の0, 1, 2, 3, ..., r, ..., nの生起の確率を与える。

再びサイコロの問題を考えると、1が出る確率は $\frac{1}{6}$ で、1が出ない確率は $\frac{5}{6}$ である。したがって  $p=\frac{1}{6}$ ,  $q=\frac{5}{6}$  である。6個のサイコロを投げて正確に2個の1が出る確率は

$$\binom{6}{2} \left(\frac{5}{6}\right)^4 \left(\frac{1}{6}\right)^2 = \frac{6!}{2!4!} \left(\frac{5}{6}\right)^4 \left(\frac{1}{6}\right)^2 = 0.2009$$

であり、0, 1又は2個の1の目の何れかを得る確率は

$$\left(\frac{5}{6}\right)^6 + \frac{6!}{1!5!} \left(\frac{5}{6}\right)^5 \left(\frac{1}{6}\right) + \frac{6!}{2!4!} \left(\frac{5}{6}\right)^4 \left(\frac{1}{6}\right)^2 = 0.9378$$

#### e. 代数函数

統計研究者は  $y=2x^2$  又は  $y=e^x$  のような函数の意味を理解しなければならない。特別な値をとるyはxの値如何によることがわかる。この場合には、我々はyはxの函数であるといふ、言葉でいうのを数学的に短かくして  $y=f(x)$  といふ。統計学においては、我々はよく度数函数を取扱う。例えば

$$y = c e^{-u^2} \quad (15)$$

ここにyは度数で、uは函数であり、eはlogのNapierian systemの底である。量uは他の量xの函数であり、方程式(15)は次の様に書ける

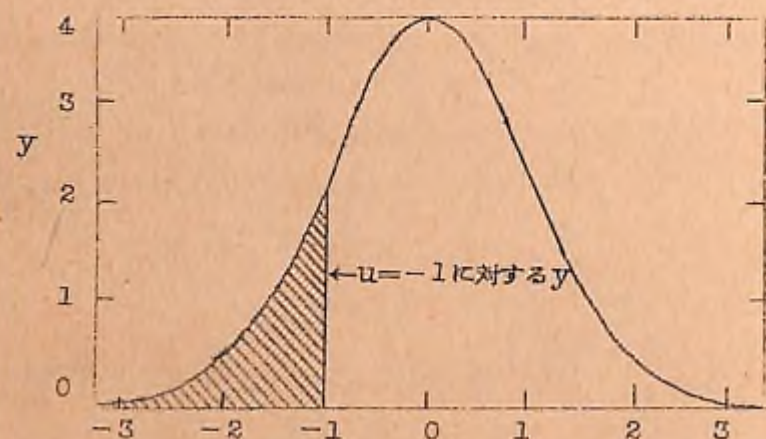
$$y = c e^{-\frac{x^2}{\sigma^2}} \quad (16)$$

特別な問題に対してはσは常数であることが期待され、この制限内ではyは厳密にxの函数である。度数函数においてxは穀物の収量、又は動物の重量等のような変量を表す。xの与えられた値に対して度数函数は理論度数yを与える。統計学における我々の問題の一つは実験資料に適合する様に度数函数を選ぶことである。よりしばしば度数函数はその基礎の下に検定が実験資料に適用される数学的モデルとして役に立つ。

勿論、函数は図で表わすことができる。そして第1-1図は方程式(15)に対して得られた図の型を示す。このような図によつて二つの重要な事実があらわされる。まづ第一に、それはuの与えられた値に対してのyの値を示す。例えば、 $u=-1$ でyの値は2.4である。第二に曲線と基線との間にある色々な部分の面積の間の関係を示す。例えば斜線をつけた部分は-1と-∞との間にある面積を示す。



第 1-1 図 函数  $y = ce^{-u^2}$  の図



この場合にはこの特別な方程式に対しては、曲線は実際には決して横軸に交らないという言葉を私は使ったがこのことは何等混乱をひき起さないであろう。すべての実際目的に対して、 $u = -4$  以上の面積は殆んど 0 と考えてよい。曲線の全面積を 1 で表わすと、 $-\infty$  から  $-1$  迄の割合は全体の一つの割合である。数学的な言葉では、それは  $-\infty$  から  $-1$  まで方程式(15)を積分したものであるという。

#### f. 連立方程式

$$\begin{aligned} 3x + 5y &= 22 & \text{なる二つの式があればこれ等は二つの未知数 } x \text{ 及び } y \text{ を} \\ 7x - 4y &= 20 & \text{もつ連立方程式の対と呼ばれる。これは未知数の一つの}$$

係数を両方の方程式の係数が同じになるように乗じたり割つたりして解くことができる。

例えば  $7(3x + 5y = 22) = 21x + 15y = 154$

$$3(7x - 4y = 20) = 21x - 12y = 60$$

したがって上から下を引くと

$$47y = 94$$

すなわち  $y = 2$

方程式の一つに  $y = 2$  を代入すると  $x$  に対して解くことが出来る新しい方程式を得る。

連立方程式の組は未知数と同数の方程式がなければ解くことができないということは記憶しておかねばならない。しかし、もしも一つだけ方程式を欠き、しかも右辺の項が 0 であれば、未知数の相対的値を得ることができる。例えば

$$8x + 5y + 2z = 0$$

$$11x + 2y + 6z = 0$$

$x$  で割ると

$$8 + 5\frac{y}{x} + 2\frac{z}{x} = 0$$

$$11 + 2\frac{y}{x} + 6\frac{z}{x} = 0$$

今  $\frac{y}{x} = y_1, \frac{z}{x} = z_1$  とおいて

$$5y_1 + 2z_1 = -8$$

$$2y_1 + 6z_1 = -11$$

これを解くと

$$15y_1 + 6z_1 = -24$$

$$2y_1 + 6z_1 = -11$$

$$13y_1 = -13$$

$$y_1 = -1$$

$5y_1 + 2z_1 = -8$  に  $y_1 = -1$  を代入して

$$-5 + 2z_1 = -8$$

$$2z_1 = -3$$

$$z_1 = -\frac{3}{2}$$

この方程式の解は  $\frac{y}{x} = -1$  および  $\frac{z}{x} = -\frac{3}{2}$ 、したがって

$\frac{y}{z} = \frac{y}{x} \times \frac{x}{z} = (-1) \times (-\frac{2}{3}) = \frac{2}{3}$  を得る。ある応用においてはこの種の解で必要充分である場合がある。

#### g. 行列式及び行列

次の連立方程式を考えよう。

$$a_1x + b_1y = c_1$$

$$a_2x + b_2y = c_2$$

これを解くと

$$x = \frac{b_2c_1 - b_1c_2}{a_1b_2 - a_2b_1}$$



$$y = \frac{a_1 c_2 - a_2 c_1}{a_1 b_2 - a_2 b_1}$$

を得る。

両方の分数の分母は同じであることに注意せよ。これを便利な形で次のように書く。

$$\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}$$

この形は二次の行列式として知られているものである。これは常に第一の斜めの積から第二の斜めの積を差引いて、直接型  $a_1 b_2 - a_2 b_1$  を作ることができる。ここに斜めは次の様に示す。



分子はまた次の様に行列式で表わすことができる。

$$x = \frac{\begin{vmatrix} c_1 & b_1 \\ c_2 & b_2 \end{vmatrix}}{\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}} \quad y = \frac{\begin{vmatrix} a_1 & c_1 \\ a_2 & c_2 \end{vmatrix}}{\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}} \quad (17)$$

最初の方程式において相応する記号と行列式の記号の位置的関係は容易に明らかで、これらに必要な未知数を与える行列式を直接書くことができる。

三次の行列式は次のような方程式から導くことができる。

$$\begin{aligned} a_1 x + b_1 y + c_1 z &= d_1 \\ a_2 x + b_2 y + c_2 z &= d_2 \\ a_3 x + b_3 y + c_3 z &= d_3 \end{aligned}$$

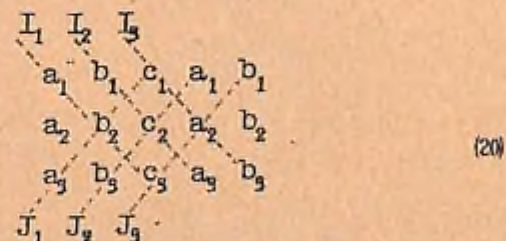
これから次の様に示すことができる。

$$x = \frac{\begin{vmatrix} d_1 & b_1 & c_1 \\ d_2 & b_2 & c_2 \\ d_3 & b_3 & c_3 \end{vmatrix}}{\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}} \quad (18)$$

これらは規則によつて書くことができる。何故なら分子  $d$  は未知の係数のところにおかれ、分母はすべて同じだからである。したがつて

$$y = \frac{\begin{vmatrix} a_1 & d_1 & c_1 \\ a_2 & d_2 & c_2 \\ a_3 & d_3 & c_3 \end{vmatrix}}{\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}} \quad (19)$$

三次の行列式もまた簡単な規則によつて解くことができる。前と同様に下の様な図表で書くことができる。I 対角線は加え、J 対角線は引く。



何次もの行列式は存在するが、四次以上の高次のものは二次及び三次の様に簡単な規則で解くことはできない。(附録参照)

次の様な形に配列された記号の組

$$\begin{pmatrix} a_1 & b_1 & c_1 & d_1 \\ a_2 & b_2 & c_2 & d_2 \\ a_3 & b_3 & c_3 & d_3 \\ a_4 & b_4 & c_4 & d_4 \end{pmatrix} \quad (20)$$

を行列 (マトリックス) と呼ぶ。これは単に行及び列に配列された記号の組に関するものであ



る。そして大括弧又は二重の線によつて行列として示される。また列の数は必ずしも行の数と一致しない。行列として記号の組に関係することができ、したがつてある組織的な配列に対して一つの記号で全行列を表わすことができるのは数学的な書き方において非常に便利である。行列は行と列が同じ数でも行列式とはならないことに注意されたい。しかし正方行列は行列式を決定する。従つて行列

$$\begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \end{pmatrix}$$

に対して

$$\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}$$

を行列の行列式と呼ぶことができる。

## 第 2 章 変動、統計量及び度数表

### 1. 変 動

統計的方法の必要性は我々が取扱わねばならない資料の変動性から生ずる。誰でもが木についているリンゴの大きさは同じでないし、穀物の各エーカーは同じブッシェル数を産しないし、すべての人が同じ高さでないし、同じ群の動物は同量の飼料を食べない等の事実はよく知つてゐる。同じような性質をもつた多くの他の例を引用することができる。

変動が存在するので、群の特性を決定しようとする問題が生じてくるのは明らかである。測定さるべき特性は変動をうけているから、それをあらわす際に、常に有効であるような数字を作ることは不可能であるといふことができる；しかしこの様な尺度がかなりの信頼度で長く用いられる時には明らかに不合理である。個々の群の特性を記述する最良の方法を定め、いかにしてこのような記述がそれ等の信頼性の客観的な検定に役立ちうるかを示すのが統計学の目的である。一般的にいうと、方法は統計量を計算するために表および図を用意し、有意性の検定を適用することである。

### 2. 母 集 団

母集団なる言葉が統計的にどのような意味をもつかをきめることは困難であるが、例を与えることは幾分容易である。もしも我々がカナダの男の市民の平均収入をきめようとするならば我々の心に抱いている母集団は厳密には収入の母集団であり、男の市民の母集団ではない。このような母集団は数理統計学の展開において、我々の思考の背景を形作る無限母集団から明らかに区別された有限母集団に関するものである。R. A. Fisher を引用すれば、“一つ以上の特性に関して度数分布において分布した無限母集団の考え方はすべての統計的仕事の基礎である”。

統計学の主目的の一つは標本として知られている無作為に抽出された個々のものの群の研究から、母集団に関する推論を導くことである。

### 3. 標 本

上に指摘したように、標本を抽出する目的はそれが抽出された母集団について情報を得ることにある。この情報はもしも得られた結果が不偏であるように標本が抽出されるとすれば、そ



のときにのみ真の値をもっている。事象の通常の過程において、このことは標本を無作為に抽出することによつて得られる。即ち標本を作り上げる際に、母集団における個々のものが含まれる機会が等しいようにして得られる。標本の大きさは費用、便利さ、必要な正確度および多くの他の因子によつて変化する。事実、標本抽出は統計の書物において多くの論議が行われているものであり、全問題に対するすぐれた本は W. Edwards Deming の "Some Theory of Sampling" で、この本に対して研究者は多くの抽出問題の詳細な論議を読みとらねばならない。

#### 4. 算術平均

算術平均は統計量のまづ第一の例である。それは標本から計算されるから統計量と呼ばれる。そして我々の目的に対してそれから標本が抽出される母集団平均値を推定するのに用いられる。標本から計算されたすべての統計量について、この重要な命題を心にとめておくことは統計学を研究する上で必要なことである。したがつて実験圃場におけるプロットからとられたとき的小麦の品種の収量は、もしも同様な条件の下で大面積上に生育していればそれがどのような品種が収穫されるかの単なる推定量である場合にはそれそのものには特に興味はない。

N個の変量をもつ標本に対して、 $X_1$  で一つの変量を表わすと、平均値  $\bar{X}$  は次によつて与えられる。

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_1 + \dots + X_N}{N}$$

これは短かく次のように書ける。

$$\bar{X} = \frac{\sum_{i=1}^N (X_i)}{N} \quad (1)$$

短かい公式を用いるということは、単に量の和が理解され、我々が和のすべてを書く必要がなく、和の符号でそれ等を開係させるという意味である。表示  $\sum_{i=1}^N (X_i)$  は N個の値を加えあわしたということを示す；しかし加え合す値の和が明らかなきときには  $\sum (X)$  なる簡単な形が用いられる。

平均値の最も興味ある特性の一つは、平均値からの個々の変量の偏差の和が0であるということである。これは個々の変量を  $X_1$  で表わし、 $x_1 = (X_1 - \bar{X})$  で個々の偏差を表わすと次の

ようにして証明できる。

$$\begin{aligned} \sum (x_1) &= (X_1 - \bar{X}) + (X_2 - \bar{X}) + \dots + (X_1 - \bar{X}) + \dots + (X_N - \bar{X}) \\ &= (X_1 + X_2 + \dots + X_1 + \dots + X_N) - N\bar{X} \\ \text{又 } N\bar{X} &= \frac{N(X_1 + X_2 + \dots + X_1 + \dots + X_N)}{N} \end{aligned} \quad (2)$$

であるから  $\sum (x_1) = 0$  であることは明らかである。

代数的に短かくして和の符号を用いると

$$\begin{aligned} \sum (x_1) &= \sum (X_1 - \bar{X}) = \sum (X) - N\bar{X} \\ N\bar{X} &= \frac{N\sum (X)}{N} \end{aligned}$$

であるから再び  $\sum (x_1) = 0$  であることは明らかである。

平均値の他の興味ある特性は平均値からの偏差の平方和即ち  $\sum (X_1 - \bar{X})^2 = \sum (x_1)^2$  はいかなる他の値からの偏差平方和より小さいということである。これは公式的に次のようにして述べることができる。

$$\sum (X_1 - \bar{X})^2 = \sum (X_1 - M)^2 - N(\bar{X} - M)^2 \quad (3)$$

従つてもしも X の代りに M を代入すると偏差平方和は  $N(\bar{X} - M)^2$  だけ増大する\*

註\*  $M = \bar{X} \pm \Delta$  としよう。

$$\begin{aligned} \text{そのとき } \sum (X_1 - M)^2 &= \sum (X_1 - \bar{X} \pm \Delta)^2 = \sum [(X_1 - \bar{X}) \pm \Delta]^2 \\ &= \sum (X_1 - \bar{X})^2 \pm 2\Delta \sum (X_1 - \bar{X}) + N\Delta^2 \end{aligned}$$

上の(2)から  $\sum (X_1 - \bar{X}) = 0$  であることがわかるそして表示  $2\Delta \sum (X_1 - \bar{X})$  において各偏差は常数である  $2\Delta$  倍されたものであるから全表示は0となる。従つて

$$\begin{aligned} \sum (X_1 - M)^2 &= \sum (X_1 - \bar{X})^2 + N\Delta^2 \\ &= \sum (X_1 - \bar{X})^2 + N(\bar{X} - M)^2 \end{aligned}$$

これは上の(3)に等しい。

#### 5. 標準偏差

標本を表わすため標本平均をとると、この平均の信頼性は標本を作り上げている個々の変量間の変動の程度によるということが明らかにされねばならない。もしも変動がなければ平均値は全標本を完全に表わす。しかし変動が大きくなると平均値の単一の値は全群を表わすことが少くなり、もつと完全に標本を表わすためにはある変動の尺度を作ることが必要である。平均値



からの平均偏差が暗示されるが、平均値からの偏差の和は0である。(4節)したがって偏差の平均はまた0となる。我々は常に0となる統計量を取扱うことはできない。そして他の一つの方法は符号を考えないで絶対偏差の平均をとることであるが、この型の統計量は代数的操作が容易でない。

選ばれるべき変動の他の尺度は  $\sum (X_1 - \bar{X})^2$ ,  $\sum (X_1 - \bar{X})^3$ ,  $\sum (X_1 - \bar{X})^4$  等の値に基くものであろう。偶数幕は常に正であるが、奇数幕は正、0又は負である。選択をする際に注意すべき最初の点は、偏差の幕が大となればなる程個々のかけはなれの偏差の影響が大となるということである。これは必要以上に計算を複雑にするので望ましいことではないという事を考えあわせて、 $\sum (X_1 - \bar{X})^2$  の値が変動の標準の尺度の基礎として一般に選ばれてきた。

記号  $s'$  によつて標準偏差を表わすと、これは次によつて与えられる。

$$s' = \sqrt{\frac{\sum (X_1 - \bar{X})^2}{N}} = \sqrt{\frac{\sum (X_1)^2}{N}} \quad (4)$$

標準偏差を計算する直接法は、平均値からの偏差を決定し、平方し、加えあわせて  $N$  で割つて平方根を求める。標本に二三の変量があり特に偏差が小数を含む場合にはもつと短かい方法が利用できる。その操作の主部分は偏差平方和を求めることであり、次のように非常に容易に示すことができる。

$$\sum X_1^2 = \sum X_1^2 - \frac{(\sum X_1)^2}{N} \quad (5)$$

この公式は特に計算機に利用でき、現在統計研究室で広く用いられているものである。

統計的方法の実際的应用において非常に重要である点を考えることがここでは必要である。標本平均はそれから標本が抽出された母平均の最良可能な推定量としてとられる。これに対する理由は平均値は偏倚平均値ではないということである。それはもしも等しい大きさをもつ標本からの非常に多くの平均値のグループがあり、それらの平均値を決定したとすれば、この値は丁度群々の配列を考えないで個々のものの全体の集まりの平均値を決定したと同じであるという意味で母平均の最良推定量である。さて抽出さるべき母集団に対して、この母集団に対する平均値および標準偏差は一定値であり、従つて母数と呼ばれる。もしも母平均を  $\mu$  で表わすと、標本平均  $\bar{X}$  は母数  $\mu$  の不偏推定量である。同様に母標準偏差を  $\sigma$  で表わすと、標本から計算されるものはこの母数の不偏推定量である。実際この推定量は上に定義した平方平均平方根ではない。これは  $\mu$  を母平均とすれば  $\sigma^2$  の不偏推定量は

$$\frac{\sum (X_1 - \mu)^2}{N} \quad (6)$$

で与えられるが  $\mu$  は未知で、標本平均  $\bar{X}$  で代用しなければならないということから分る。上の(3)から次のことが明らかである。

$$s'^2 = \frac{\sum (X_1 - \mu)^2}{N} = \frac{\sum (X - \mu)^2}{N} \quad (7)$$

換言すれば  $s'^2$  は  $(X - \mu)^2$  なる量だけ  $\frac{\sum (X - \mu)^2}{N}$  より小さい。更にこの考え方を代

数的に展開していくと、 $\sigma$  の不偏推定量は

$$s^2 = \frac{\sum (X_1 - \bar{X})^2}{N - 1} \quad (8)$$

によつて与えられることが証明される。

数学的期待値に基いたこの簡単な証明は  $H_0: \theta = 1$  によつて与えられている。

そのとき不偏推定量は

$$s = \sqrt{\frac{\sum (X_1 - \bar{X})^2}{N - 1}} \quad (9)$$

によつて与えられるが実際にはこれは少し偏倚があり、そしてこの偏倚はもしも母集団が第3章で述べるような正規分布をしているということが分つておれば、そのときのみ完全に除去できる。 $\sigma$  の不偏推定量を与える公式は

$$\hat{\sigma} = s \sqrt{\frac{2}{N} \frac{(\frac{N-2}{2})!}{(\frac{N-3}{2})!}} \quad \text{又は} \quad s \sqrt{\frac{2}{N-1} \frac{(\frac{N-2}{2})!}{(\frac{N-3}{2})!}} \quad (10)$$

である。ここに  $\sigma$  は  $\sigma$  の推定値を意味する。実際我々はしばしば一般に偏倚は重要である程大きくないので  $s$  を  $\sigma$  の不偏推定量として用いる。

ここで採用した記号は母数に対してはギリシャ文字、母数の推定量である統計量にはラテン文字を用いたことを注意しておく。これはこれから以下ずつと用いる記号である。

上の(8)式の分母  $(N-1)$  は標準偏差推定に利用できる標本での自由度 ( $D, F$ ) として知られている。最初標本に利用できる  $N$  個の自由度は1つの統計量  $\bar{X}$  そのものは標本から計算され標準偏差の推定量決定に用いられるから1だけ減ずる。

## 6 平均値の標準誤差

標本平均は標本が変れば変化し、その変動の程度は個々の変量間の変動の程度に関係するとは明らかである。個々の変量の変動及び標本平均の変動間の一般的な関係は次によつて与えられる。

$$s_{\bar{X}} = \frac{s}{\sqrt{N}} \quad (11)$$



ここに  $s_{\bar{x}}$  は標本平均の誤差であり、 $s$  は標本に対する全体としての標準偏差で、 $N$  は標本における数である。換言すれば、もしも  $s$  を母集団の標準偏差の推定量とすると、 $s/N$  は同じ母集団から抽出された大きさ  $N$  なる標本平均の母集団の標準偏差の推定量である。平均値の標準誤差は従つて標本における数の平方根に逆比例する。

## 7 分散

分散なる言葉は現在実験からの結果の統計的分析に関係して広く用いられている。従つて研究者は最初からそれに精通しなければならない。母分散は単に標準偏差の平方であり、従つて通常  $\sigma^2$  なる記号で表わす。同様に通常偏差の平方平均として知られている  $\sigma^2$  の不偏推定量は  $s^2$  で表わす。

標準偏差と同様分散は変動の尺度であるが、統計的分析におけるその特別な価値は第5章、分散分析法で詳細に論ずる。

## 8 度数表

度数表は研究されている標本に対して変量がある明らかなきまつた級に属する度数を示すものである。もしも標本が小さければ、度数表は通常必要でないが、大きいか又は適当な大きさの標本であれば、この種の表でもつて資料の縮小をやることを望ましい。度数表はグラフ表示を容易にするために価値がある。又それから平均値や標準偏差のような統計量が個々の値をもとの組から計算するよりも非常にたやすく計算される。

## 9 度数表の作成

第一の仕事は級の値をきめることである。これは表からの統計量計算の正確度、変動の範囲（これは勿論標本の最大と最小の値の差である）、標本における数即ち総度数およびこれらの級を用いて計算を取扱う場合の容易さによる。第一の場合は、級の数が多くなればなる程、表から行う計算の正確度は大となる。しかし我々が取扱いに便利な級の数には限度がある。これらの相反する二つの因子はバランスしなければならない。良好な一般的規則は級間隔を標準偏差の  $1/4$  以下にすることである。勿論我々は一般に表を作成する前に標準偏差を知ることにはできないが、変動の範囲からそれを大ざつかに推定することが可能である。Tippet は変動範囲と標準偏差の関係について詳細な表を作成した。そしてこれを Snedecor が短かい表にまとめあげた。2-1表は Snedecor の表からとつた数字を二桁の有効数字にまとめたものである。

2-1 表

20~1000 なる標本の大きさに対する  
範囲/標準偏差 の値

| 標本数 | 範囲/ $\sigma$ | 標本数  | 範囲/ $\sigma$ |
|-----|--------------|------|--------------|
| 20  | 3.7          | 200  | 5.5          |
| 30  | 4.1          | 300  | 5.8          |
| 50  | 4.5          | 400  | 5.9          |
| 75  | 4.8          | 500  | 6.1          |
| 100 | 5.0          | 700  | 6.3          |
| 150 | 5.3          | 1000 | 6.5          |

いま500の標本があり、変動の

範囲が0.25~2.63即ち2.38で

あるとしよう。そのとき期待標準偏

差は2.38/6.1で適当な間隔は  $1/4$

(2.38/6.1)=0.098である

う。間隔の中央点は級範囲をきめる

値におけるより以上の小数を必要としないから、級間隔に対して偶数よりも奇数にする方が便利であろう。例えば、上例ではおそらく0.11間隔とした方がよいであろう。

級を作る際に、最小の値より少し小さいところに最初の級の境界をするのが普通である。したがつて級および中央点は最後には次のようになる。

| 級 範 囲     | 級の値即ち級範囲の中央点 | これらの級範囲は次の小数位をも   |
|-----------|--------------|-------------------|
| 0.19~0.29 | 0.24         | つ0.296のような値はおこらない |
| 0.30~0.40 | 0.35         | と仮定している。もしも0.296の |
| 0.41~0.51 | 0.46         | ような数がでてくればそれを0.30 |
| 0.52~0.62 | 0.57         | ~0.40の級にいれる。      |
| etc       | etc          | 上の規則に従うことによつて度数   |

表から計算される統計量は充分な正確度を得る。しかし、もしも表が主として下に述べるようなグラフを作るために必要だとすれば、この方法は級のいくつかは非常に少ない度数しかないか又は全くないようなあまりにも小さい級を与える。そのときは級の間隔を標準偏差の  $1/4$  から  $1/2$  間隔にすることが望ましい。

配列は取扱いに便利なカード上に各変量の値を書くことによつて非常に容易になる。級範囲をカード上に書き、表上の数字の順番に並べると配列は迅速に行うことができ、分布を完成する際に東にざつと眼を通し、チェックすることが容易である。誤つた位置におかれたカードは最後の段階で非常に面倒になるので、この点で完全に正確に直すことが大切である。度数表は最後にそれらに応ずる級の値に度数を入れることによつて作られる。

2-2表は度数表の例である。これは500の牡牛の精液の範囲についての資料を表わす、範囲は任意の単位で与えてある。この標本に対する標準偏差は2.58である；したがつて計算に高度の正確度をもたせるためには級間隔は約0.6でなければならない。しかしこれほどの正確度は必要でなかつたから、級間隔は1.0が選ばれた。範囲および標準偏差間の関係が2-1



表で与えたものと一致していることに注意せよ。ここに範囲は近似的には0.6~16.5で15.9であり、2-1表から標準偏差は近似的に15.9/6.1=2.61でなければならない。

## 2-2表

500の牡牛の精液の任意単位面積に対する  
度数表

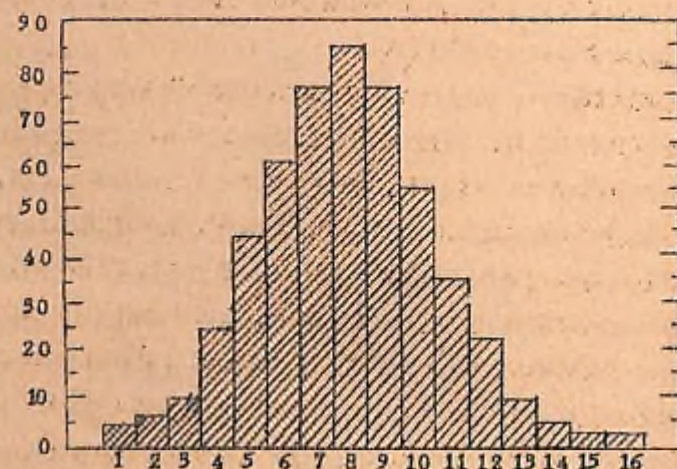
| 級 範 囲   | 度 数 | 級 範 囲     | 度 数 |
|---------|-----|-----------|-----|
| 0.6-1.5 | 5   | 8.6-9.5   | 69  |
| 1.6-2.5 | 6   | 9.6-10.5  | 54  |
| 2.6-3.5 | 10  | 10.6-11.5 | 35  |
| 3.6-4.5 | 25  | 11.6-12.5 | 22  |
| 4.6-5.5 | 44  | 12.6-13.5 | 12  |
| 5.6-6.5 | 61  | 13.6-14.5 | 4   |
| 6.6-7.5 | 68  | 14.6-15.5 | 1   |
| 7.6-8.5 | 83  | 15.6-16.5 | 1   |

## 10 度数表のグラフ表示

度数表は通常二つの型のグラフで表わされる。ごく一般的なものはヒストグラムである。それは級を横軸にとり、度数を縦軸上の適当な位置に立てた縦軸上にとつた度数のグラフ表示である。例として2-2表の資料に対するヒストグラムを2-1図に示す。

グラフのもう一つの型は度数多角形と呼ばれるものである。各度数に應ずる級の中央に垂直な直線を立て、これらの端を直線によつて順々に結びつけたものである。これは図的にはヒストグラムほど正確ではないが、同じグラフ上の二つの分布の比較するにはヒストグラムよりよいものである。

2-1図  
500の牡牛の精液の面積に  
対するヒストグラム



## 11 度数表からの平均値と標準偏差の計算

第一段階は下に示す例でわかるように、度数表に二つの欄を加えることである。平均値は次によつて与えられる。

$$\bar{X} = \frac{\sum (f_1 X_1)}{N} \quad (12)$$

また標準偏差は次によつて与えられる

$$S = \sqrt{\frac{\sum (f_1 X_1)^2 - [\sum (f_1 X_1)]^2 / N}{N-1}} \quad (13)$$

| 級の値   | 度 数   | 度数X級の値               | 度数X級の値の平方               |
|-------|-------|----------------------|-------------------------|
| $X_1$ | $f_1$ | $f_1 X_1$            | $f_1 X_1^2$             |
| 1     | 2     | 2                    | 2                       |
| 2     | 4     | 8                    | 16                      |
| 3     | 7     | 21                   | 63                      |
| 4     | 6     | 24                   | 96                      |
| 5     | 1     | 5                    | 25                      |
| 計     | 20=N  | 60= $\sum (f_1 X_1)$ | 202= $\sum (f_1 X_1^2)$ |

標準偏差の分子は $\sum (X_1^2)$ を含んでいることを思い出して頂きたい。標準偏差はこの場合は(13)式により計算してある。

多くの場合、級の値は2とか4などの文字を含む数である。その場合、計算に当つて1, 2, 3, 4, ...等の自然数列にそれ等を置き換えることによつて労力の節約ができる。そのとき $\bar{X}'$ および $S'$ で表わされる計算から平均値と標準偏差が得られる。これ等は次の等式によつて真の値に変換できる。

$$\bar{X} = (\bar{X}' - 1) C + X_1 \quad (14)$$

$$S = S' C \quad (15)$$

ここにCはもとの級間隔で $X_1$ は最初のもとの級の値である。

級の数がかなり多い時は、平均値に近い任意の原点をとつて更に計算を簡単にできる。表は次のようにして作る。ここに実際の数字は例としてとつたものである。

| $X_1$ | $u_1$ | $f_1$ | $f_1 u_1$      | $f_1 u_1^2$      |
|-------|-------|-------|----------------|------------------|
| 18    | -2    | .     | .              | .                |
| 21    | -1    | .     | .              | .                |
| 24    | 0     | .     | .              | .                |
| 27    | 1     | .     | .              | .                |
| 30    | 2     | .     | .              | .                |
|       |       | N     | $\sum f_1 u_1$ | $\sum f_1 u_1^2$ |



ここに  $X_0$  は任意の原点 = 24

$$X_1 = C u_1 + X_0$$

なることに注意すると

$$\bar{X} = C \bar{u} + X_0 \quad (16)$$

ここに  $\bar{u} = \sum f_1 u_1 / N$ 。標準偏差は次によつて与えられる。

$$C = \frac{\sum f_1 (u_1 - \bar{u})^2}{N - 1} \quad (17)$$

ここに

$$\sum f_1 (u_1 - \bar{u})^2 = \sum f_1 u_1^2 - \frac{(\sum f_1 u_1)^2}{N}$$

## 1.2 変動係数

変動係数なる言葉は標本平均の百分率で表わした標準偏差のことである。それは統計的方法によつてその信頼性をきめることが困難であるため、限られた応用をもつ一つの統計量である。

公式は

$$C (\text{変動係数}) = S \left( \frac{100}{\bar{X}} \right) \quad (18)$$

## 第3章 理論的度数分布

### 1 生物的変量の一つの実際的分布の特性

第2章で500の牡牛の精液に対する度数表の結果を図示して得られたヒストグラムを示した。さてもしも500の代りに10,000の精虫が測定されたならば、級間隔はより小さくすることができ、基線を同じ長さにとると、端の端は殆んど平滑な曲線になるであろう。その理論的な結論にこの論法を適用すると、無限に大なる母集団における牡牛の精液の範囲はこのような曲線で表わすことができるということがわかる。これは我々が連続的な変量の母集団に統計的方法を適用する際に心に抱いているものの実際的な図である。不連続変量はヒストグラムによつてのみ無限母集団で表わすことができる。

少くとも一つの生物的変量の分布の模様に注目して、このような分布が生ずる理論的条件について探究することは興味あることである。これは変動の原因となる作用の現われ方についての一つの考え方を導く。できてくる精液は明らかにその究極の大きさに影響する多くの因子によつて左右される。それが腺の復位組織からできるときは、あるものは精虫を大きくし、あるものは精虫を小さくするような大きさに影響する無数の因子がある。これらの因子が作用する模相に対しての手がかりを得る基礎として、すべてが等しい強さで作用すると仮定し、生み出される理論分布に関してある結論をうることが有用である。このやり方は等しい強さで作用する多数の因子があるという仮定からの数学的な誘導法である。それは普通正規分布と呼ばれている分布を導く。このやり方は純数学的であるから、生物学者はおそらく一般化された分布型を実際の試行によつて作つたり決定したりすることにより興味があるであろう。

### 2 正規分布の実験的偏差

Tippett Fisher および Yates が与えた乱数表の組を調べる際、もしも10のグループをとるとこのグループに対する総計は0から90迄変化するということに注意しよう。

しかし0なる総計は我々が0を10選んだときのみ得られる。又90なる総計は9を10選んだときのみ得られる。表から無作為に1つの数字をとる際に0を選ぶ確率は0.1であるから無作為に10のグループをとつて0を10選ぶ確率は  $0.1^{10} = 0.0000000001$  である。勿論9を10選ぶ確率は0を10選ぶ確率に同じである。

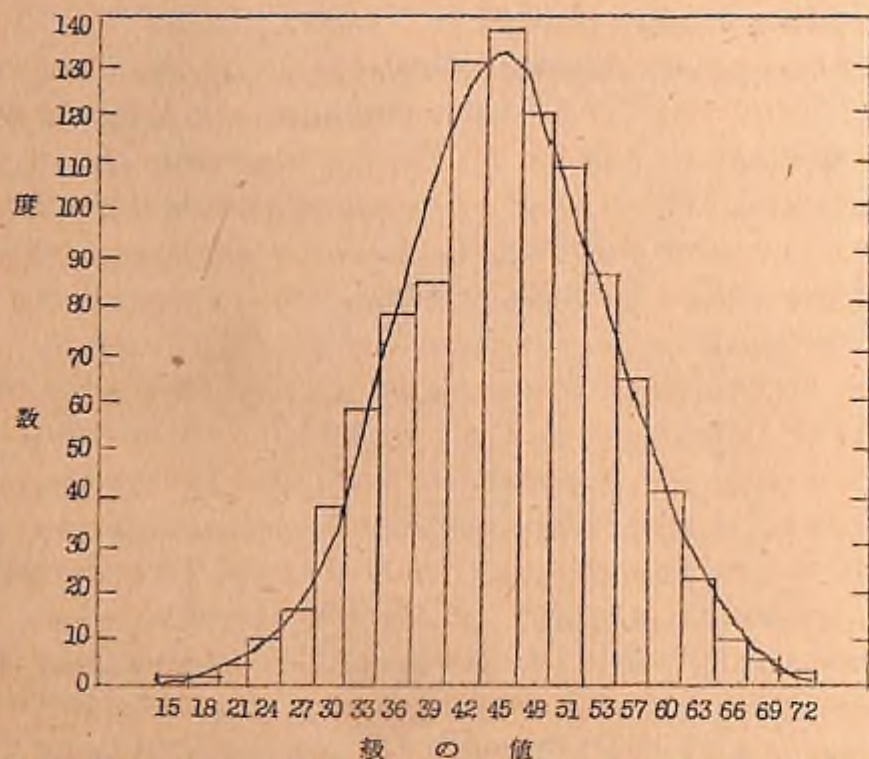
しかし45のような総計は数字の非常に多くの組合せから作り得ることは明らかであるから10の数字のグループを無作為に選ぶ際には、この総計は0や9の総計よりもずっと大きな度数でおこるであろう。したがつて範囲の中央の数は最大度数でおこり、他は端の限界に近づくにつれてより小さい度数でおこる。

ここに設けた条件は大きつた方法で生物的変量に影響するものと比較できる。各数字は原因としてまた、グループ総計はこれ等の原因が生物的変量に与える作用の究極の影響と考える



3-1 図

乱数の1000のグループの総計に対するヒストグラムおよび適合された正規曲線



ことができる。したがって仮りにこの仮定が合理的であるとすれば我々は乱数のグループの総計から生物学的変量を測定してうるものと非常によく似た型の分布を得るであろう。

3-1図は Fisher & Yates の表からとつた10の乱数の1000グループの総計の度数分布をヒストグラムで表わしたものである。

後に例3-1で述べる方法で平滑な正規曲線をあてはめた。ヒストグラムは2-1図のように二つの条件の組を大体比較できるという結論を示すに充分である。

以下正規曲線の基礎的な数学的原則のいくつかを論議し、それに含まれる統計量のいくつかを定義しよう。

### 3 正規分布

正規分布は次式によつて定義される

$$Y = \frac{N}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{x - \mu}{\sigma}\right)^2} \quad (1)$$

ここに  $\sigma$  は母集団標準偏差、 $N$  は変量総計、 $e$  は log の Napierian system の底で、 $Y$  は母平均から測定した  $x$  なる点での度数密度である。したがって曲線は従属変数としての  $Y$  と  $x$  との関係を表わす。3-2図は牡牛の精液の範囲の度数に対して2-1図のヒストグラム上に画いた正規

曲線を示す。

方程式(1)は次のように書ける

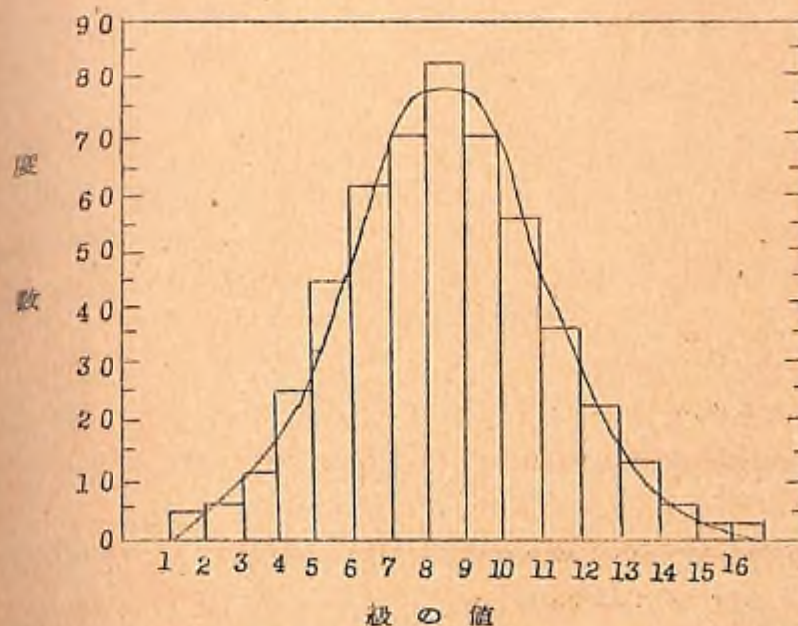
$$Y \left(\frac{\sigma}{N}\right) = \frac{e^{-\frac{1}{2} \left(\frac{x - \mu}{\sigma}\right)^2}}{\sqrt{2\pi}}$$

$$Y \left(\frac{\sigma}{N}\right) = Z \text{ とおいて}$$

$$Z = \frac{e^{-\frac{1}{2} \left(\frac{x - \mu}{\sigma}\right)^2}}{\sqrt{2\pi}} \quad (3)$$

さて  $x/\sigma < -6$  又は  $x/\sigma > 6$  なる値に対しては  $Z$  の値は非常に小さい。したがって実際には  $Z$  が  $x/\sigma$  について0から6まで0.01間隔で表わされるならば我々が求める  $Z$  に対するすべての情報をうるることができる。これは確率積分の Sheppard の表に与えてあり Pearson の Tables for Statisticians & Biometricians, Part 1 (Table A-2 に短かくした形でのせてある) にも印刷されている。

3-2 図 2-1図のヒストグラムの上に画いた正規曲線



我々が取扱っている実際の母集団では  $Z$  の値は級間隔を  $c$  とすると  $cN/\sigma$  を乗じて  $Y$  に変換できる。換言すれば、 $N$  および  $\sigma$  がわかっている特別な母集団に対しては  $Z$  の表の組で理論的な平滑な正規曲線をプロットしてゆくことができる。

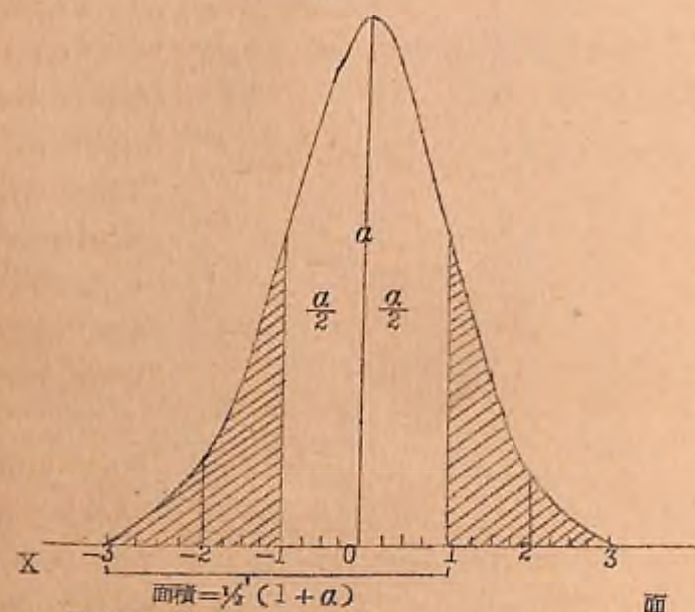
上述の方法によつてプロットした

平滑な曲線はそれから標本が抽出された無限母集団の型の推定量である；しかししばしば必要となるものは標本の実際度数に應ずる理論的な度数分布である。即ち我々は実際度数の任意に選ばれた級に対して理論的な正規度数を必要とする。この目的のためもしも  $N$  を1とすれば方程式(1)は次のようになる

$$Y = \frac{e^{-\frac{1}{2} \left(\frac{x - \mu}{\sigma}\right)^2}}{\sigma \sqrt{2\pi}} \quad (4)$$



3-3 図  $X'_\sigma = +1$  及び  $-1$  での縦軸を示す正規曲線  
影をつけない部分が  $a$  で、影をつけた部分が  $(1-a)$  である



#### 4 正規曲線のあてはめ

正規曲線のあてはめは次の例3-1で示される。

### 例3-1

実際の度数分布に正規曲線を適合させて、実際度数に應ずる正規度数を決定するに必要な計算は3-1表に与えてある。資料は軽い貧血をおこした病人からとつた400人の赤血球の透明度に対するものである。透明度は赤血球の切断面を通過する全光線の比としてとつた。この分布に対しては $\bar{X}=7.06$  及び $S^*=2.45$  で $S^*$ は $\sigma$ の代りにとられる。計算は表の各欄を考えるとよく分る。これらは参考になるため表の始めに番号がつけられている。

(1)欄 第2章で述べた級範囲である計算を簡単にするため実際の級間隔の代りに1なる単位  
の級間隔をとつた。級間隔がきまつた後、実際度数を(1)欄の様にのせるが、この段階では計算  
には用いない。

± $\gamma/\sigma$  の限界内の面積は  $\gamma/\sigma$  の定められた値に対して  $\alpha$  で表わされる。したがつてもしも曲線の全面積が1であれば  $\gamma/\sigma = \infty$  から

$x/g$  の定められた値までの面積は  $\frac{1}{2} + \frac{1}{2}a$  である。この方法の実際例えの応用は次の4節で述べる。

3-1 表 正規曲線の適合及び理論度数の計算

| (1)<br>級間隔 | (2)<br>級限界 | (3)<br>d | (4)<br>$\frac{d}{s}$ | (5)<br>Z | (6)<br>$Y = \frac{N}{Z \left( \frac{N}{s} \right)}$ | (7)<br>$\frac{1}{2}Q + w$ | (8)<br>$\frac{N}{2}Q + w$ | (9)<br>理論正規<br>度 數 | (10)<br>實際度 數 |
|------------|------------|----------|----------------------|----------|-----------------------------------------------------|---------------------------|---------------------------|--------------------|---------------|
|            |            | 9.56     | 3.90                 | 0.0002   | 0.03                                                | 1.0000                    | 400.00                    | 0.08               |               |
|            |            | 8.56     | 3.49                 | 0.0009   | 0.15                                                | 0.9998                    | 399.92                    | 0.32               |               |
|            |            | 7.56     | 3.08                 | 0.0035   | 0.57                                                | 0.9990                    | 399.60                    | 1.08               |               |
| 0.6—1.5    |            | 6.56     | 2.68                 | 0.0110   | 1.80                                                | 0.9963                    | 398.52                    | 3.16               | 4             |
| 1.6—2.5    | 1.5        | 5.56     | 2.27                 | 0.0303   | 4.95                                                | 0.9884                    | 395.36                    | 7.92               | 11            |
| 2.6—3.5    | 2.5        | 4.56     | 1.86                 | 0.0707   | 11.54                                               | 0.9686                    | 387.44                    | 16.84              | 17            |
| 3.6—4.5    | 3.5        | 3.56     | 1.45                 | 0.1394   | 22.76                                               | 0.9265                    | 370.60                    | 30.28              | 29            |
| 4.6—5.5    | 4.5        | 2.56     | 1.04                 | 0.2323   | 37.92                                               | 0.8508                    | 340.32                    | 44.76              | 43            |
| 5.6—6.5    | 5.5        | 1.56     | 0.64                 | 0.3251   | 53.08                                               | 0.7389                    | 295.56                    | 59.16              | 56            |
|            | 6.5        | 0.56     | 0.23                 | 0.3985   | 63.43                                               | 0.5910                    | 236.40                    |                    |               |
| 6.6—7.5    | 7.06       | 0.00     | 0.00                 | 0.3989   | 65.12                                               | 0.5000                    | 200.00                    | 64.96              | 58            |
|            | 7.5        | 0.44     | 0.18                 | 0.3925   | 64.08                                               | 0.5714                    | 228.56                    |                    |               |
| 7.6—8.5    | 8.5        | 1.44     | 0.59                 | 0.3352   | 54.72                                               | 0.7224                    | 288.96                    | 60.40              | 63            |
| 8.6—9.5    | 9.5        | 2.44     | 1.00                 | 0.2420   | 39.51                                               | 0.8413                    | 335.52                    | 47.56              | 61            |
| 9.6—10.5   | 10.5       | 3.44     | 1.40                 | 0.1497   | 24.44                                               | 0.9192                    | 367.68                    | 31.16              | 25            |
| 10.6—11.5  | 11.5       | 4.44     | 1.81                 | 0.0775   | 12.65                                               | 0.9649                    | 385.96                    | 18.28              | 20            |
| 11.6—12.5  | 12.5       | 5.44     | 2.22                 | 0.0339   | 5.53                                                | 0.9868                    | 394.72                    | 8.76               | 9             |
| 12.6—13.5  | 13.5       | 6.44     | 2.63                 | 0.0126   | 2.06                                                | 0.9957                    | 398.28                    | 3.56               | 4             |
| 13.6—14.5  |            | 7.44     | 3.04                 | 0.0039   | 0.64                                                | 0.9988                    | 399.52                    | 1.24               |               |
|            |            | 8.44     | 3.44                 | 0.0011   | 0.18                                                | 0.9997                    | 399.88                    | 0.36               |               |
|            |            | 9.44     | 3.85                 | 0.0002   | 0.03                                                | 0.9999                    | 399.96                    | 0.08               |               |
|            |            | 10.44    | 4.26                 | 0.0000   | 0.00                                                | 1.0000                    | 400.00                    | 0.04               |               |
|            |            |          |                      |          |                                                     |                           | 計                         | 400                | 400           |

(2) 標 識限界\* なる言葉の意味をつかむために第2章2-1図のヒストグラムを参考にしよう。この限界はヒストグラムの欄を区切る線に応ずるものである。

標本平均はそれが属する範囲におかれる。この場合には平均値は7.06であるから、級範囲6.6~7.5におかねばならない。残りの限界は平均値から両方の方面に向つて作られる、平均値が属する級は二つの限界を入れなければならないが、他の各々に対しては平均値から最も離れたもののみをのせればよい。

(3) 欄 平均値からの級限界の偏差

(4) 欄 前欄の同じ行の値を次の公式から計算した標準偏差で割つたもの



$$\sqrt{\frac{\sum (X - \bar{X})^2}{N}}$$

ここに偏差  $(X - \bar{X})$  は単位級間隔に基づくものである。

- (5) 欄 正規曲線の縦軸についてのA-2表からとつたZの値。
- (6) 欄 Zを  $N/s$  倍したもの、完全な公式は  $C N/s$  であるが  $C=1$  である。
- (7) 欄 A-1表からの  $\frac{1}{2}(1+a)$  の値。
- (8) 欄  $\frac{1}{2}(1+a)$  をN倍したもの。
- (9) 欄 端からはじめて中に向つて計算した8欄の次々の値の差。中央の級では二つの差を加える。この欄の総計はNとなり計算のチェックを与える。
- (10) 欄 実数度数

### 5. 非正規型

度数分布の正規型からの変動の型は大ざつばに次の三つの級に分けられる。これらは次の通りである。

#### a. 歪度

分布の歪度は近似的に次式によつて与えられる。

$$\text{歪度} = \frac{\text{平均値} - \text{モード}}{\sigma}$$

ここにモードは曲線の最高点に引いた垂線の基線即ちX軸での位置である。この尺度は正規分布では曲線は対称で平均値とモードは一致するから明らかに0である。モードが平均値よりも大きいときは負の歪みをもち、小さいときは正の歪みをもつ。負の歪みは曲線の端が左側にのびている場合で、正の歪みは右側にのびている場合である。

#### b. 平坦性

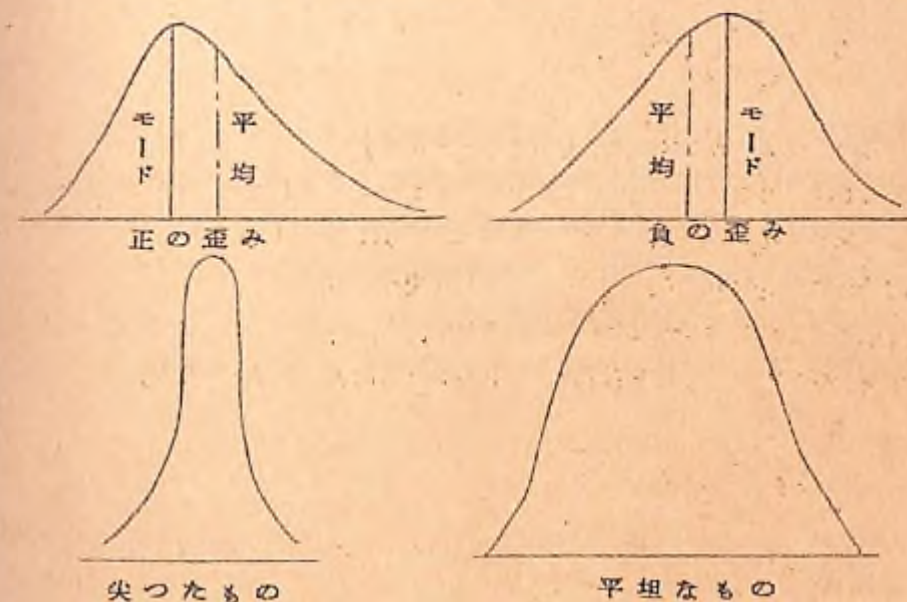
平坦な曲線では肩は正規曲線よりも十分広がっているし、端はすぼんでいる。

#### c. 尖度

先の尖つた曲線に対しては正規型より中央が高く、端は広がっている。ある分布においては平坦性および尖度とともに歪みをもつことがある。この例を3-4図 b, cに示してある。これらは非正規型の通常の型を示すものである。

\*注) 級限界は正確でないことに注意せよ；例えば0.6~1.5および1.6~2.5なる級間隔の間では正確な級限界は1.55である。各級で0.05を除くと計算は簡単になり、計算された縦軸および度数は少ししか違わない。

3-4図 非正規型の例



### 6. 非正規型の尺度

歪みの近似的な尺度は5節で述べた方法で得ることができるが、より有効な尺度は標本のモーメントの値によつて定められた統計量を計算して得られる。特別な原点mに関して分布のk次モーメントは  $[\sum (X-m)^k] / N$  で与えられる。したがつて原点として平均値をとるとそのまわりの二次モーメントは  $[\sum (X-m)^2] / N$  である。四次までのモーメントは正規性からのへだたりの便利な指標である。例えばもしも分布が対称であれば、平均値のまわりのモーメントは奇数巾に対しては0である。そしてもしもそれが広がっていれば正規分布に対するよりも大なる偶数の高次モーメントを与える傾向がある。

原点として平均値をとつた場合のモーメントの計算に対してはまづ下に示す量の系列を得ることが最も便利である。

$$a_1 = \frac{\sum X}{N}, \quad a_2 = \frac{\sum X^2}{N}, \quad a_3 = \frac{\sum X^3}{N}, \quad a_4 = \frac{\sum X^4}{N} \quad (5)$$

そのとき、モーメントは

$$\left. \begin{aligned} \text{一次モーメント} \quad \nu_1 &= a_1 \\ \text{二次モーメント} \quad \nu_2 &= a_2 - a_1^2 \\ \text{三次モーメント} \quad \nu_3 &= a_3 - 3a_1 a_2 + 2a_1^3 \\ \text{四次モーメント} \quad \nu_4 &= a_4 - 4a_1 a_3 + 6a_1^2 a_2 - 3a_1^4 \end{aligned} \right\} \quad (6)$$



モーメントに基いた分布型の二つの尺度は

$$a_3 = \frac{\nu_3}{\nu_2^{3/2}}, \quad a_4 = \frac{\nu_4}{\nu_2^2} \quad (7)$$

(これらを  $\beta_1$  および  $\beta_2$  とかくことがある)

対称分布では奇数市のモーメントは0となるので正規型又は他の対称分布では  $a_3$  は0となる。正規曲線では  $a_4 = 3$  で、この値からのへだたりは非正規性の指標と考えられる。 $a_4 < 3$  であれば分布は頂上が平たく、 $a_4 > 3$  であれば尖っている。

正規性からの度数分布のへだたりの尺度として、R. A. Fisher はモーメントに関する組直値を計算して  $k$  統計量と名付けた。これらは次によつて与えられる。

$$\left. \begin{aligned} k_1 &= \nu_1 \\ k_2 &= \left( \frac{N}{N-1} \right) \nu_2 \\ k_3 &= \left[ \frac{N^2}{(N-1)(N-2)} \right] \nu_3 \\ k_4 &= \frac{N^2}{(N-1)(N-2)} \left[ \frac{(N+1)\nu_4 - 3(N-1)\nu_2^2}{N-3} \right] \end{aligned} \right\} \quad (8)$$

二つの  $k$  統計量  $k_2$  および  $k_4$  には度数分布の級間隔に対する補正が必要である。単位の間隔では補正値は次のようになる。

$$k_2' = k_2 - \frac{1}{12}, \quad k_4' = k_4 + \frac{1}{120} \quad (9)$$

他の間隔については計算には単位間隔を用いることが常に可能であるから補正は必ずしも必要でないであろう。

$k$  統計量での  $a_3$  および  $a_4$  と比較できる曲り方の尺度は

$$g_1 = \frac{k_3}{(k_2')^{3/2}}, \quad g_2 = \frac{k_4'}{(k_2')^2} \quad (10)$$

である。

これらは母数  $\gamma_1$  および  $\gamma_2$  に対する標本からの推定量と考えられる。

$g_1$  および  $g_2$  の標準誤差は

$$s_{g_1} = \sqrt{\frac{6N(N-1)}{(N-2)(N+1)(N+3)}} \quad (11)$$

$$s_{g_2} = \sqrt{\frac{24N(N-1)^2}{(N-3)(N-2)(N+3)(N+5)}} \quad (12)$$

研究者は曲線が正規型に近づくにつれて  $g_2$  は0に近づき、 $a_4$  は3に近づくことを注意しなければならない。正規分布に対しては  $g_1$  および  $g_2$  共に0である。 $g_1$  は対称性の尺度であり、(平均値-モード)と同じ符号をもつ。3-4図は  $g_1$  の正および負の値によつて正および負の歪みを示したものである。 $g_2$  の正の値は先の尖つた曲線を示し、負の値は平坦な曲線を示す。これら二つの型は3-4図にも示した。

$k$  統計量の計算および歪みの尺度  $g_1$  および  $g_2$  の計算を例3-2に示す。

例 3-2

3-2表はこの例となる度数分布とその初歩的な計算を与える。偏差は任意原点5からとつてあることに注意せよ。任意原点をとることは本質的ではないが特に度数表の級の数がかなり大きいときは計算の仕事が非常に簡単になる。上で与えた  $a_1$   $a_2$   $a_3$  および  $a_4$  に対する公式で  $X$  を仮平均からの偏差  $d$  で置きかえる。

$k$  統計量から  $g_1$  および  $g_2$  を次のように計算する。

$$g_1 = \frac{0.697258}{(2.42663)^{3/2}} = +0.1844 \quad s_{g_1} = 0.226$$

$$g_2 = \frac{0.1094}{(2.4266)^2} = +0.0186 \quad s_{g_2} = 0.451$$



3-2 表

仮平均を用いた k 統計量の計算

| x                                            | f  | 仮想平均からの<br>偏差 (d) | fd         | fd <sup>2</sup> | fd <sup>3</sup> | fd <sup>4</sup> |
|----------------------------------------------|----|-------------------|------------|-----------------|-----------------|-----------------|
| 1                                            | 1  | -4                | -4         | 16              | -64             | 256             |
| 2                                            | 6  | -3                | -18        | 54              | -162            | 486             |
| 3                                            | 13 | -2                | -26        | 52              | -104            | 208             |
| 4                                            | 25 | -1                | -25        | 25              | -25             | 25              |
| 5                                            | 30 |                   |            |                 |                 |                 |
| 6                                            | 22 | 1                 | 22         | 22              | 22              | 22              |
| 7                                            | 9  | 2                 | 18         | 36              | 72              | 144             |
| 8                                            | 5  | 3                 | 15         | 45              | 135             | 405             |
| 9                                            | 2  | 4                 | 8          | 32              | 128             | 512             |
| $\Sigma(d) \cdots \Sigma(d^4) \quad (N=113)$ |    |                   | -10        | 282             | 2               | 2058            |
| $a_1, a_2, a_3, a_4$                         |    |                   | -0.0884956 | 2.49558         | 0.0176991       | 18.2124         |
|                                              |    |                   |            | -0.00783        | 0.6625435       | 0.0063          |
|                                              |    |                   |            |                 | -0.0013861      | 0.1173          |
|                                              |    |                   |            |                 |                 | -0.0001         |
| * $\nu_2, \nu_3, \nu_4$                      |    |                   |            | 2.48775         | 0.678865        | 18.3359         |
| $k_2, k_3, k_4$                              |    |                   |            | 2.50996         | 0.6972583       | 0.1011          |
| 補正項                                          |    |                   |            | -0.06333        |                 | 0.0083          |
| $k'_2, k'_3, k'_4$                           |    |                   |            | 2.42663         | 0.6972583       | 0.1094          |

\*  $\nu_2, \nu_3, \nu_4$  の計算には公式(1)を用いる

## 7 標本平均の分布

5節で  $a_3$  および  $a_4$  は実際の度数分布の非正規性の尺度であることを注意した。これらの統計量に関連して標本平均に重要な関係をもつ数学的統計量に関する良く知られた定理がある。この定理はもしも我々が母集団から大きさ  $N$  なる標本を抽出し、標本平均の度数分布を決定すれば、この分布に対する  $a_{3\bar{x}}$  および  $a_{4\bar{x}}$  の値が次によつて与えられるというのである。

$$a_{3\bar{x}} = \frac{a_3}{\sqrt{N}}, \quad a_{4\bar{x}} = \frac{a_4 - 3}{N} + 3 \quad \text{ここに } a_3 \text{ および } a_4 \text{ は標本が抽出された母集団の値である。したがつて母集団に存在する非正規性に関係なくその母}$$

集団から抽出された標本平均の分布は正規型に近づく。極端な歪分布をもつ母集団に対してさえ、すべての実際目的には  $n$  が 4 位の標本平均でも正規に分布する。Shewhart は 2 つ

の非正規母集団の各々から 1,000 からなる 4 つの標本を抽出する実験を行つた。これらの中一つは矩形であり、他は直角三角形である。両方の場合に 4 つの標本平均の分布は正規分布に近づいた。この関係について Shewhart は次のように述べている。「もつと厳密な解析的方法によつて得られるこのような結論から、実用上殆んどの場合我々は 4 つ又はそれ以上の標本平均の抽出限界を正規法則の理論から作り得るといふ確信をもつことができる。」と。

このことから我々は乱数表から無作為に抽出した 10 ケの数字の総計から作られた 2 節の大體正規に近い分布についても一度注意する必要がある。乱数は本質的に矩形分布であるすなわち曲線の頂上は平坦である。そして数字の範囲は 0~9 に限られるので  $a_3$  および  $a_4$  の理論値を導くことができる。これらの値は  $a_3 = 0$  および  $a_4 = 2.01$  である。したがつて 10 ケの数字の総計の分布に対しては  $a_3 = 0$ ,  $a_4 = [(2.01 - 3) / 10] + 3 = 2.90$  である。実験で得られた  $a_4$  の値は 2.88 であつた。

## 8. 二項分布

二項分布は不連続変量を取扱う上に最も有用な分布の一つである。実際的には次のように定義される：小麦の分離母集団において、もしも芒のあるものが生ずる確率を  $p$ , 芒のないものの生ずる確率を  $q$  とすれば、大きさ  $n$  の非常に多数の標本中で芒のないものと芒のあるものが生ずる理論分布は  $(q + p)^n$  なる二項展開の次々の項で与えられる。 $q = \frac{1}{4}$ ,  $p = \frac{3}{4}$  と仮定して各植物が 6 つから成立つているとすれば二項  $(\frac{1}{4} + \frac{3}{4})^6$  を展開して (第 1 章 2 節 d) 次を得る。

$$\left(\frac{1}{4}\right)^6 + 6\left(\frac{1}{4}\right)^5\left(\frac{3}{4}\right) + 15\left(\frac{1}{4}\right)^4\left(\frac{3}{4}\right)^2 + 20\left(\frac{1}{4}\right)^3\left(\frac{3}{4}\right)^3 + 15\left(\frac{1}{4}\right)^2\left(\frac{3}{4}\right)^4 + 6\left(\frac{1}{4}\right)\left(\frac{3}{4}\right)^5 + \left(\frac{3}{4}\right)^6$$

これは

$$\left(\frac{1}{4}\right)^6 + 6 \times 3 \left(\frac{1}{4}\right)^6 + 15 \times 3^2 \left(\frac{1}{4}\right)^6 + 20 \times 3^3 \left(\frac{1}{4}\right)^6 + 15 \times 3^4 \left(\frac{1}{4}\right)^6 + 15 \times 3 \left(\frac{1}{4}\right)^6 + 6 \times 3 \left(\frac{1}{4}\right)^6 + 3 \left(\frac{1}{4}\right)^6$$

とおくと計算が少し容易になる。

これらの項は対数を用いて次のように計算される：

$$\text{項1の } \log = \log \left(\frac{1}{4}\right)^6 = 6 \log \left(\frac{1}{4}\right)$$

$$\text{項2の } \log = \log 6 + \log 3 + 6 \log \left(\frac{1}{4}\right)$$

$$\text{項3の } \log = \log 15 + 2 \log 3 + 6 \log \left(\frac{1}{4}\right)$$

etc



まず  $6 \log (1/4)$  および  $\log 3$  を求めておくことと便利である。これは

$$6 \log (1/4) = 4.3876400$$

$$\log 3 = 0.4771213$$

次の段階ではこれらの項の対数および真数を次のように書く。

| 6つの標本の芒のある植物 | $\log$ の項 | antilogの項即ち確率 |
|--------------|-----------|---------------|
| 0            | 4.3876400 | 0.000244      |
| 1            | 3.6429126 | 0.004395      |
| 2            | 2.5179739 | 0.032959      |
| 3            | 1.1200339 | 0.131836      |
| 4            | 1.4722165 | 0.296631      |
| 5            | 1.5513978 | 0.355957      |
| 6            | 1.2503678 | 0.177979      |

対数によらずに計算器を用いて二項展開の項を計算する直接法はSnedecorにより与えられている。 $(1/4 + 3/4)^6$  の展開においてこれを例示しよう。

| 芒のある植物 | $(q)^{n-r}$<br>$= (0.25)^{n-r}$ | $p^r$<br>$= (0.75)^r$ | $\binom{n}{r}$ | 項        |
|--------|---------------------------------|-----------------------|----------------|----------|
| 0      | 0.000244                        |                       | 1              | 0.000244 |
| 1      | 0.000977                        | 0.75                  | 6              | 0.004396 |
| 2      | 0.003906                        | 0.5625                | 15             | 0.032957 |
| 3      | 0.015625                        | 0.421875              | 20             | 0.131836 |
| 4      | 0.0625                          | 0.316406              | 15             | 0.296631 |
| 5      | 0.25                            | 0.237305              | 6              | 0.355958 |
| 6      |                                 | 0.177979              | 1              | 0.177979 |

最後から2番目の欄の0.25から出発して、次々に0.25を乗じてゆくと $(q)^{n-r}$ の欄が計算できる。同様に上から2番目の欄の0.75から出発して次々に0.75を乗じてゆくと $p^r$ の欄が計算できる。三番目の欄は $n! / r! (n-r)!$ である。二欄、三欄、四欄の数字を乗じると五欄の値を得る。

計算器の場合に用いられるもう一つの方法は次の事実から導かれたものである。即ち $q^n$  から始めて次々に $(p/q)$  を乗すると次を得る。

$$q^n, q^{n-1}p, q^{n-2}p^2, q^{n-3}p^3, \dots$$

そしてこれらに対応する $\binom{n}{r}$ の値を乗すれば二項展開の次々の項を得る。同じ例に対して計算を示してみよう。

| 芒のある植物 | $q^n \left(\frac{p}{q}\right)^r$ | 項        |
|--------|----------------------------------|----------|
| 0      | 0.000244141 $\times 1 =$         | 0.00024  |
| 1      | 0.000732423 $\times 6 =$         | 0.00439  |
| 2      | 0.00219727 $\times 15 =$         | 0.03296  |
| 3      | 0.00659181 $\times 20 =$         | 0.13184  |
| 4      | 0.0197754 $\times 15 =$          | 0.296631 |
| 5      | 0.0593262 $\times 6 =$           | 0.35596  |
| 6      | 0.177979 $\times 1 =$            | 0.17798  |

さてこの分布の実験的な応用を示してみよう。もし $p$ および $q$ が与えられた母集団に対して正しい確率であれば、6つの植物の無作為標本をとる際に2つ以下の芒のある植物を得る確率をきめることは容易である。これは明らかに0, 1および2つの芒のある植物を得る確率の和で計算すると0.038である。6つの植物の1000の標本をとると、2つ以下の芒のある植物を含む標本の期待数は38である。

二項分布の平均値および標準偏差は度数分布の場合と同様に計算できる。即ち6つの植物の1000の標本があると仮定して理論分布は次のようになる。

| 標本における芒のある植物 | 度数    |
|--------------|-------|
| X            | f     |
| 0            | 0.2   |
| 1            | 4.4   |
| 2            | 33.0  |
| 3            | 131.8 |
| 4            | 296.6 |
| 5            | 356.0 |
| 6            | 178.0 |
| 計            | 1000  |

$$\Sigma(X) = (0 \times 0.2) + (1 \times 4.4) + \dots + (6 \times 178.0) = 450.2$$

$$\text{芒のある植物の標本あたり平均} = \frac{450.2}{1000} = 4.50$$

$$\Sigma(X^2) = (0.2 \times 0^2) + (4.4 \times 1^2) + \dots + (178.0 \times 6^2) = 2137.62$$

$$\frac{[\Sigma(X)]^2}{1000} = \frac{(450.2)^2}{1000} = 20251.8$$

$$\text{従つて } \Sigma(X^2) = 21376.2 - 20251.8 = 1124.4$$

$$\text{又 } \sigma = \sqrt{\frac{1124.4}{1000}} = 1.06$$



実際には、この例では我々は  $p, q$  および  $n$  が分っている理論分布の二つの母数を測定している。 $p$  および  $n$  は完全に二項分布を定義するので、平均値および標準偏差は数学的にはそれらに關係している。したがって数学的な計算によつて

$$\bar{X} = pn \quad (13)$$

なることが証明できる。ここに  $\bar{X}$  はそれにたいする確率が  $p$  である事象の標本当り生起の平均数を変換す。標準偏差は

$$\sigma = \sqrt{pqn} \quad (14)$$

によつて変換されることがまた証明できる。

上例にこれらの公式を適用して

$$\bar{X} = \left(\frac{3}{4}\right) 6 = 4.5$$

期待したように直接計算の結果と完全に一致している。

$$\sigma = \sqrt{\frac{3}{4} \times \frac{1}{4} \times 6} = 1.06$$

勿論、ある応用においては  $p$  および  $q$  の値が未知であることが分るであろう。この場合

には最初に示したような直接法で資料から  $p$  および  $q$  の値の推定量を得ることができる。 $n$  が小さいとき二項展開の各項の  $p$  および  $q$  の値の影響はかなり大である。 $p$  と  $q$  が等しいときには、項は対称であるが、等しくないときは分布は歪んでいる。

二項分布の形が  $p, q$  および  $n$  によつてどのように影響されるかは二項分布の  $\alpha_3$  および  $\alpha_4$  に対する公式からよくわかる。即ち

$$\alpha_3 = \frac{q-p}{\sqrt{pqn}} \quad \alpha_4 = 3 + \frac{1-6pq}{npq} \quad (15)$$

$p=q$  のときは  $\alpha_3 = 0$  であり、 $p, q$  が違つてくるとその値が増してくることに注意せよ。しかし  $p$  および  $q$  の差に無關係に  $\alpha_3$  は  $n$  が増すと全く小さくなる。 $\alpha_4$  を調べると、 $p, q$  および  $n$  が變動しても比較的にその値の变化は小さいが、 $p$  および  $q$  の差が著しく、 $n$  が小さいときは頂上が平坦な分布を示す傾向がある。

もしも  $n=9$  になる二項分布に対してヒストグラムを作ると、ヒストグラムには 10 の級および 10 の欄がある。基線を同じ長さにすると、 $n=29$  に対するヒストグラムは 30 の欄となり、欄は  $n=9$  のものの幅の  $\frac{1}{3}$  となる。  $n$  を増してゆき、同じ基線にヒストグラムを作ると、その頂上は連続曲線になる傾向があることは明らかである。実際には、 $n$  が増すと二項分布は我々が既に研究した正規分布に近づき遂に正規分布になつてしまう。

## 9. ポアソン分布

ポアソン分布は實際的に非常に重要な理論分布である。それは特殊な例によつて最もよく表わすことができる。

車輦に積まれた小麦から標本が抽出され各標本は 1 ポンド又は約 16,000 の種子からなつていと仮定しよう。小麦の中にサハギタの種子がある。そして各標本はポンド当りサハギタ

の種子がどれだけあるかをきめるため注意深く検定される。もしも貨車積小麦でのポンド当りサハギタの種子の実際の数が 3 であれば多くの標本をとるとき、この数字まわりのサハギタの種子の数の變動が期待されるであろう。二項分布についての我々の研究から 16,000 の種子当りサハギタの理論分布は

$$(q+p)^{16,000}$$

で与えられることは明らかである。ここに  $p$  はサハギタの種子が出てくる確率、 $q$  は出てこない確率である。しかし次の二つの点に關して二項分布の場合の問題とは違つていふことに注意しなければならぬ。即ち第一に標本に屬する数が非常に大であるといふこと、第二に確率  $p$  は非常に小さいといふことである。これらの条件の下で、分布の次々の連続した項は次によつて表わされることが数学的に証明できる。

$$e^{-m} (1, m, \frac{m^2}{2!}, \dots, \frac{m^r}{r!}, \dots) \quad (16)$$

ここに  $e$  は対数の Napierian system の底で、 $m$  は標本当り事象の平均生起数である。これは通常ポアソン分布と呼ばれる。更に詳細にのべると二項に対するポアソン分布と呼ばれるものである。

ポアソン分布の最も有用な特性の一つは、分散が平均値に等しいといふことである。この理由から、ポアソン分布は平均値によつて完全に規定され、したがつて上に表示した分布の各項に対しては平均値が与えられた唯一の媒介変数である。

ポアソン分布の型に  $p$  および  $n$  がどのように影響するかといふこともまた興味がある。

$$\sigma_3 = \frac{1}{np} \quad \text{及び} \quad \sigma_4 = + \frac{1}{np} \quad (17)$$

であるから、 $p$  が増すにつれて歪みが大きくなり、 $n$  が増すにつれて分布が対称に近づくことは明らかである。 $\sigma_4$  の公式は  $\sigma_4$  は決して 3 以下にならないことを示している；したがつて分布は正規分布よりも決して平坦にはならない。量  $np$  がかなり大きくなると曲線の尖りの程度は非常に小さくなる。

ポアソン分布が応用できる例は多くの場合抽出問題の性質によつてきめることができる。与えられた例で、車輦から無作為に一つの種子を選ぶ際にサハギタの種子を得る確率は全く小さいが、1 ポンドの標本ではサハギタの種子の實際数は評価できる位である。 $n$  を標本に屬する種子の数とすると平均値は  $pn$  であるから、ポアソン分布となるための条件は次のように定義できる。事象の生起の確率は全く小さくなければならないが、 $n$  は  $pn$  が評価できる位大でなければならない。例えば 200 プワットの系列においてプロット当り約 10,000 小麦の穂があると仮定しよう。プロットを調べると、一部分の穂が小麦の稈によつて枯死していることが注意される。数えてみると、プロット当り害をうけた穂の平均数は 6 であつた。 $pn=6$  であるから  $p = \frac{6}{10,000}$  であることが分り、直ちに害をうけた穂の分布はポアソン分布に従う



ということがわかる。

ポアソン分布の応用は多くの標本抽出法において重要である。更に進んだ応用は後章で与えるが、種子抽出にポアソン分布を応用することを考えるのはこの点で興味あることである。農夫がポンド当り3つのサハギタの入っている種子をもつていと仮定しよう。積荷のロットからとられた種子の1ポンド標本が6以上のサハギタの種子を含む確率はいくらか？ただし許容限界はポンド当り5である。これは換言すれば、農夫のもつ種子が1ポンド標本の検定からの結果を基にしてすてられる確率である。この問題を解くために我々は  $m=3$  である場合の理論分布を作らねばならない。これは Pearson の Tables for Statisticians & Biometricians を利用して得られた。表をみて  $m=3$  のところをみると次の値が分る。

| 標本当りサハギタの種子の数 | 生起確率  |
|---------------|-------|
| 0             | 0.050 |
| 1             | 0.149 |
| 2             | 0.224 |
| 3             | 0.224 |
| 4             | 0.168 |
| 5             | 0.101 |
| 6             | 0.050 |
| 7             | 0.022 |
| 8             | 0.008 |
| 9             | 0.003 |
| 10            | 0.001 |
|               | 0.000 |
| 計             | 1.000 |

$P$  = 標本をうけいれる確率、 $Q = 1 - P$  = すてる確率 とすると、 $Q$  は標本当り 6, 7, 8, 以上のサハギタの種子をうる確率の和である。これは 0.084 で従つて  $P = 0.916$  である。したがつてもし 5 をある級の種子に対する許容限界とし、単一の 1ポンド標本がとられたとすると、ロットがすてられる確率は 0.084 である。これを別ないい方でいうとすてる確率は 1:12 であるといえる；又は調査してこのような標本の非常に多くが得られると、12回のうち1回だけすてられるということが期待される。

## 第 4 章 有意性の検定

統計的方法の最も重要な役目の一つは有意性の検定である。ここに引用した有意性の検定という言葉は一見すると完全な一定の考え方を表わしているようであるが、更に深く考えると、その意味を明らかにし特に統計的意味で有意性検定とは如何なるものであるかを定義することが必要である。有意という言葉の日常使用している意味は広義の意味においてのみ統計的な言葉と同じである。

今、我々が貨幣をもつていとしよう。それが偽つてゐるか否か即ちそれを投げて表と裏が出る確率が等しいか否かについて疑問があると仮定しよう。まづ思いつくことは何回も貨幣を投げてそれが表か裏かを記録することである。もし我々がそれをおこなつて、表か裏かのどちらかが多く出れば、貨幣には偏倚があるという証拠になる。しかし更に分析しなければ、この試験には非常に不満足な状態が残る、というのは表か裏が多く出るとは貨幣が本当に偏倚しているということを示すのか、又はもしも貨幣が間違ひなくできていても、表と裏の出る回数が必ずしも等しくないが、その変動についてはつきりしることができないからである。この点で我々の数学的な知識を利用すると同時に我々の考え方を明らかにする論理的方法を採用するのである。

### 1. 有意性の検定の論理

貨幣を 16 回投げて、13 回表、3 回裏であつたとしよう。有意性の検定の論理を展開する第一の操作は次の質問をすることである：もしも貨幣が不偏であれば、このような結果はどの位得られるか？真の貨幣では表が出る確率  $p$  と裏が出る確率  $q$  とは相等しい、ここに  $p+q=1$  である。表および裏の分布はしたがつて二項展開  $(p+q)^{16} = (\frac{1}{2} + \frac{1}{2})^{16}$  で与えられる。最初の五項を求めるため展開すると次のようになる。

| 表の数    | 16       | 15       | 14       | 13       | 12       |
|--------|----------|----------|----------|----------|----------|
| 確率 $p$ | 0.000015 | 0.000244 | 0.001831 | 0.008545 | 0.027771 |

真の貨幣を 16 回投げて 13 回以上の表を得る確率は最初の四項の和である。即ち  $0.000015 + 0.000244 + 0.001831 + 0.008545 = 0.0111$

観測された結果は真の貨幣では各試行（16 回投げる試行）を 100 回を行つてただ一回のみみるようなものである。即ち実際に投げて最初にそれが起るといふことは貨幣の真実性についてかなりの疑を抱かせるものである。

真の貨幣の表と裏の出る理論分布についての我々の知識から今我々がやつたことで実際の貨幣での結果の有意性を判断する明らかな尺度を得ることは明らかであろう。もしも理論分布が  $p=0.4$  であれば、得られた結果はとにかく異常なものではないし偏倚は存在しないことは明らかである。一方もしも  $p=0.001$  であれば貨幣は真でないといふことの確信を抱かせるであろう。



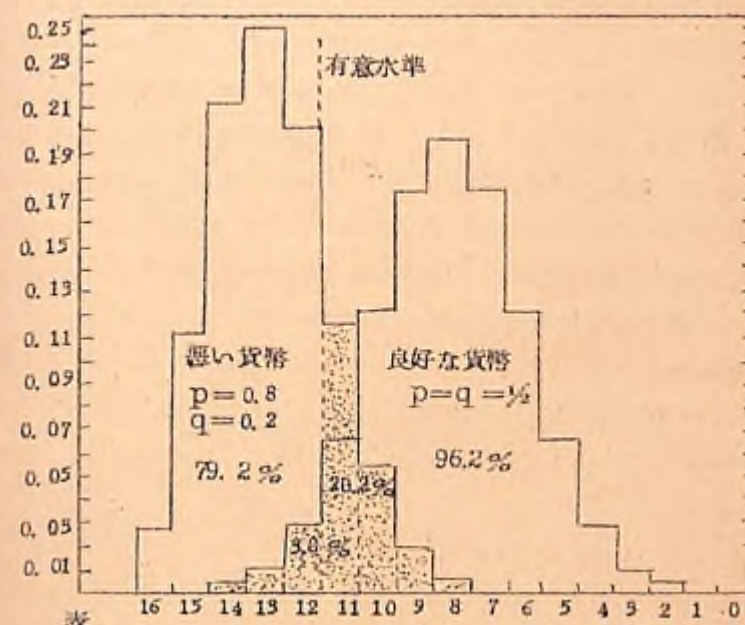
これは  $p$  の値が考慮を要することを示す。どの位の水準で我々は有意といえるか？この質問に答えるために最も現実的な方法は、不備な貨幣で非常に多くの試行をして得られる結果を考えることである。もしも  $p \leq 0.05$  であれば貨幣は偏倚しており、 $p > 0.05$  であれば不偏であるときめよう。いま真の貨幣で 1000 回試行して、50 回が  $p \leq 0.05$  の値を与えるということが期待される。したがって我々は貨幣が実際に不偏である時に試行の 50% だけ偏倚あるものとして分類するであろう。もしも試行の 5% は誤まつてもよいということであれば、 $p$  の値は 0.05 が満足なものであるが、もしそれほど誤まりたくなければもつと小さい値を用いることが必要である。

ここで、 $p$  の小さい値が誤まつた判断をすることを少なくすると考えてそれが最も望ましいものであるときめこむようなことがあつてはならない。これは一つの重要な因子をまだ考慮に置いていないからである。実際面においては完全な貨幣の、多くの系列を検定することはあまり好ましくない。検定の問題が起るのは、少くとも、いくつかの悪い貨幣がある場合であり検定の目的はそれらを除くことにあるのである。我々がおかす誤りには二つのものがあることは明らかである。J. Neyman & E. S. Pearson の適切な暗示によれば、それらは第一種の過誤および第二種の過誤と呼ばれるものである。我々が良好な貨幣を悪い貨幣として分類したとき第一種の過誤をおかし、悪い貨幣を良好な貨幣として分類したときに第二種の過誤をおかす。さて有意水準がどのようにこの二種類の過誤に影響するかをみてみよう。

色々な有意水準を用いた場合の影響を示す簡単な方法は、検定される貨幣が  $p=q=1/2$  なる真の貨幣から作られ、悪い貨幣は  $p=0.8$  および  $q=0.2$  から作られていると仮定しよう。4-1 図は今の二種類の貨幣で表の出る理論分布を示す。任意の有意水準として  $p=0.038$  をとり 16 回投げて 12 回表が出たとしよう。 $(0.5+0.5)^{16}$  および  $(0.8+0.2)^{16}$  にたいする分布を 4-1 図に示す。良好な貨幣の分配で全体の 3.8% を表わす薄い部分の面積は図の左下にある。これは良好な貨幣で悪いと呼ばれるものの割合である。濃い部分は悪い貨幣で良好な貨幣と呼ばれる割合を表わす。

これによつて有意水準が変つた場合の影響は明らかである。もし一つ以上の欄を含むよう有意水準を右に動かせば悪い貨幣のうち 8.2% だけが良好な貨幣と考えられるが、良好な貨幣の中 10.5% を悪い貨幣と考えることになる。一欄を左に動かせば、悪い貨幣の中 40.2% が良好な貨幣として分類され、良好な貨幣の中 1.1% が悪い貨幣として分類される。これは次のように非常に都合よくとりまとめることができる。

4-1 図 二種類の貨幣を 16 回投げた場合の表の分布



| 有意水準<br>16 投げた表の数 | 良好な貨幣を悪い貨幣<br>として分類する割合 | 悪い貨幣を良好な貨幣<br>として分類する割合 | 全誤差  |
|-------------------|-------------------------|-------------------------|------|
| 15: $p=0.0003$    | 0.03                    | 89.5                    | 89.5 |
| 14: $p=0.002$     | 0.2                     | 64.8                    | 65.0 |
| 13: $p=0.011$     | 1.1                     | 40.2                    | 41.3 |
| 12: $p=0.038$     | 3.8                     | 20.2                    | 24.0 |
| 11: $p=0.105$     | 10.5                    | 8.2                     | 18.7 |
| 10: $p=0.227$     | 22.7                    | 2.7                     | 25.4 |
| 9: $p=0.402$      | 40.2                    | 0.7                     | 40.9 |

この特別な例では、誤差の最小点は  $p=0.105$  のところであるが、これはいつもそうだということではない。もしも悪い貨幣の分布が  $(0.7+0.3)^{16}$  であれば結果は異つてくるであろう。そして悪い貨幣の偏倚のばらつきが非常に広範囲にわたつていれはますます複雑な問題となる。更に二種類の誤差の重大さの問題がある。ある場合には第一種の過誤は第二種の過誤よりも重大である。このとき有意水準としてなるべく小さい  $p$  をとることが必要となるであろう。第二種の過誤がより重要である場合には、 $p$  は大きくとつた方がよいであろう。

この問題について厳密に考えると有意水準を設ける場合ひかえ目になりやすい。どんな事象においても  $p$  の値をあまりに小さくとると第二種の過誤は増加する。これは最も重要な一般論であり、且有意性検定を適用する際にみおとしてはならないことである。一般的規則としては 0.05 の  $p$  が満足なものと考えられる。それは普通 5% 有意水準と呼ばれている。



## 2. いづれかの方向における偏倚の検定

貨幣を投げる前節の論議において、質問は次の通りであつた：表が期待したよりも多く出るような貨幣は偏倚があるか？、いいかえれば我々は貨幣は偏倚しているか？、というような簡単な質問をしているのではない。後者は表又は裏いづれかが出るような偏倚にたいする検定を含んでいる。この質問に答えるためには良好な貨幣で観察されるより以上悪い結果をうる度数を決定することが必要である。したがつてもし表が13回裏が3回出るとすれば一方の側の尺度で同程度の偏倚を表わす結果は3つの表、13の裏であろう。観察されたものよりも悪い結果の度数は分布の両端の理論度数を加えて得られる。この例では分布は対称であるから13以上の表にたいしての度数を単に2倍すればよい。即ち  $p = 0.022$  となる。

もつと簡単な偏倚が存在するかしないかという問題はより一般的なものであつて有意性の検定において質問される問題は普通このようなものである。もしも質問の内容がよくわかつていればいかにして正しい  $p$  をきめるかは困難ではない。しかしこの点の理解がなければ非常に混乱するおそれがある。

## 3. $t$ 分布を用いる有意性の検定

実験的方法、調査計画等において用いられる有意性の検定は、理論的には貨幣投げについて今迄述べてきたものよりもかなり複雑である。しかしこのような検定を実際に行うことは、以前から行われてきた多くの理論およびその目的のため用意された表を用いればさして困難なことではない。ここでは我々の主として検定の根底にある論理に関連してその見地から話をすすめてみよう。

4-1 表に与える二品種の小麦のプロット試験の結果があり、統計研究室でこの資料から2品種の収穫量に本質的な差があるかどうかを検定したいと仮定しよう。

4-1 表 12箇地での小麦の2品種の収量

| 箇地       | 1    | 2    | 3    | 4    | 5    | 6    | 7    | 8    | 9    | 10   | 11   | 12   |
|----------|------|------|------|------|------|------|------|------|------|------|------|------|
| 品種 $X_1$ | 24.4 | 27.9 | 28.2 | 19.8 | 23.1 | 22.9 | 25.6 | 23.7 | 26.2 | 25.7 | 37.0 | 31.5 |
| 品種 $X_2$ | 17.5 | 15.1 | 21.6 | 18.2 | 21.6 | 13.7 | 24.8 | 27.8 | 25.2 | 19.2 | 34.0 | 25.2 |

まづ第一に以前の研究から小麦品種の収量は殆んど正規に分布するということがわかつている。更に野外プロットの穀物の収量の分布にたいする文献を調べて、この型の変量の分布は明らかに正規に近づく傾向があることがわかつた。とにかく我々は、12品種の標本平均は、標本が抽出された母集団の分布がどうであろうと、近似的には正規分布をすることを知つてゐる。したがつてもし品種  $X_1$  又は  $X_2$  の、各々12の収量からなる多くの標本をとればこれらの標本の平均値は正規に分布すると仮定しても安全である。

検定のために設ける論理的な仮説は2品種間の平均差は0であるということである。したがつて我々は、仮説から平均値0のまわりに分布する差  $\bar{X}_1 - \bar{X}_2$  の非常に大きな正規分布を想像することになる。そうすると我々の正規分布の知識(第3章第3節)から平均値  $\mu$  および

標準偏差 ( $\sigma_{md}$ ) (添字は平均値の標準誤差を表わす) で観察したものと同じか、又はそれよりも大きい  $\bar{X}_1 - \bar{X}_2$  の値をうる確率を、決定することができるとを知つてゐる。例えば、もしも標準偏差が  $\sigma_{md} = 2$  であり、 $\bar{X}_1 - \bar{X}_2 = 4$  であれば、この二つの比即ち、これは標準偏差の項で表わした  $\bar{X}_1 - \bar{X}_2$  の尺度にすぎないが、この比は2となり、正規曲線の確率積分の表をみるとこの比より  $\frac{1}{2}(1 + \alpha) = 0.9773$  をうる。したがつて、観察したものと同じか、又はそれより大きい  $\bar{X}_1 - \bar{X}_2$  の正の値をうる確率は、 $1 - 0.9773 = 0.0227$  である。

しかしこの方法を実例に適用するとき、今迄概略述べた検定を行うためには、適切な事実をもたないことは明らかである。仮説から我々が想像しているものは、与えられた標準偏差  $\sigma_{md}$  をもち、平均値0のまわりに正規に分布する母集団であることを再び思いおこすである。検定を完全にするためにはこの標準偏差を知ることが必要である。換言すれば標準偏差がわかれば各々が  $\bar{X}_1 - \bar{X}_2$  なる値の非常に多くの標本を抽出できる。それから各々の標本にたいして比  $(\bar{X}_1 - \bar{X}_2) / \sigma_{md}$  を計算することができ、観察されたものと同じか又はそれより大きい  $\bar{X}_1 - \bar{X}_2$  の値をうる確率  $p$  の値を表から見出すことができる。仮説が真であるとき非常に多くの標本系列での結果をチェックすると標本の5%が  $p < 0.05$  の値を得ることが分る。これは正確に理論と一致しているから検定が正しいことが分るであろう。しかし我々は  $\sigma_{md}$  は分らないから、検定は標本からの  $\sigma_{md}$  の推定量に基いて行われなければならない。

第一の問題は  $\sigma_{md}$  を推定する方法をきめることである。これは標本が抽出された個々の値をもつ母集団にたいして  $\sigma$  を推定することであつて、これは  $\sigma_{md}$  を推定することと同じである。論理的にいうと2つの標本からえられた  $\sigma$  の推定量を一緒にするということである。 $\sigma$  の推定量を  $S$  で表わすと第1の標本から次のようになる。

$$S_1^2 = \frac{\sum (X_1 - \bar{X}_1)^2}{n-1} = \frac{\sum X_1^2}{n-1} \quad (1)$$

第2の標本にたいして

$$S_2^2 = \frac{\sum (X_2 - \bar{X}_2)^2}{n-1} = \frac{\sum X_2^2}{n-1} \quad (2)$$

さて差の分散の推定量は

$$S_d^2 = \frac{\sum [(X_1 - X_2) - (\bar{X}_1 - \bar{X}_2)]^2}{n-1} = \frac{\sum [(X_1 - \bar{X}_1) - (X_2 - \bar{X}_2)]^2}{n-1} \\ = \frac{\sum (X_1 - X_2)^2}{n-1} \quad (3)$$



これを展開すると

$$S_d^2 = \frac{\sum X_1^2}{n-1} + \frac{\sum X_2^2}{n-1} - 2 \frac{\sum (X_1 X_2)}{n-1} \quad (4)$$

$$= S_1^2 + S_2^2 - 2 \frac{\sum (X_1 X_2)}{n-1} \quad (5)$$

これをみると  $S_d^2$  は  $X_1$  と  $X_2$  の分数の和から2つの標本での平均からの偏差の積和  $\sum (X_1 X_2)$  を差し引いたものである。この量に関するより多くの知識は回帰および相関の研究のところで与えられる。しかし今は  $X_1$   $X_2$  なる対の値をとるとき両方が共に正又は共に負であれば積和は正の量になるということを指摘しておけば十分である。一方一つの変量が正で他の変量が負であれば、積和は負となる。 $X_1$  が正の値であるものの中ほぼ半分が  $X_2$  の負の値をとり、他の半分では  $X_2$  の正の値をとるとするのが中間的な場合である。 $X_1$  の負の値は同様に分けられ、その結果の積和は0となる。後者の場合、2変量は独立であるという。

我々の例では、場処の系列での対となつたプロットから収量が得られる。そして我々が各収量の平均からの偏差を決定すれば正および負の偏差は一緒におけるといふことが認められる。これは積和  $\sum (X_1 X_2)$  は一定の正の量であるから方程式(5)を考慮にいれねばならないということの意味する。 $S_d^2$  は方程式(4)から計算できるが、方程式(3)から  $S_d^2$  を計算する場合、分子をきめるために次の関係を利用すると便利である。

$$\sum (X_1 - X_2)^2 = \sum (X_1 - X_2)^2 - \frac{[\sum (X_1 - X_2)]^2}{n} \quad (6)$$

差  $X_1 - X_2$  は

6.9 12.8 6.6 1.6 1.5 9.2 0.8 0.9 1.0 6.5 3.0 6.3 および  
総計即ち  $\sum X_1 - \sum X_2 = 57.1$

したがつて

$$\sum (X_1 - X_2)^2 = 437.85 - \frac{57.1^2}{12} = 166.15$$

そして

$$S_d^2 = \frac{166.15}{11} = 15.105$$

$$S_{md}^2 = \frac{S_d^2}{n} = \frac{15.105}{12} = 1.2588$$

$$\therefore S_{md} = \sqrt{1.2588} = 1.12$$

$S_{md}$  を得ると比  $(\bar{X}_1 - \bar{X}_2) / S_{md}$  が計算できる。これは統計量  $t$  として知られている。ところで問題となるのはこの比を確率計算にたいして、 $(\bar{X}_1 - \bar{X}_2) / \sigma_{md}$  の代用として用

いてよいかどうかということである。これは容易に答えることができる。 $\sigma_{md}$  は母数でありしたがつて標本が変つても一定である。一方  $S_{md}$  は標本から推定されるものであるから変動する。それは一つの統計量であり、通常の抽出変動を受け、すべてその固有の分布をもっている。もしも  $n$  が大きければ  $S_{md}$  の値は全く一貫性があり、一方  $n$  が小さければ  $S_{md}$  の変動はそれに比例して大となることを特に注意しておく。これは  $n$  が大きければ  $S_{md}$  を  $\sigma_{md}$  に代用したときに生ずる。誤差は無視できるが、 $n$  が小さければ、その誤差が問題となつてくることを示す。実際にこの問題を解くとき必要なものは、 $n$  の色々な値にたいする  $t$  の抽出分布である。もしこれが分れば、仮説が真であるときには観察したものと同じか又はそれより大なる  $t$  の値をうる確率を問題にたいしてきめることができるであろう。

この問題は最初1908年に student によつて解かれた。彼はそれを用いて、有意性の検定を行うことができる  $t$  と密接な関係をもつ Z 分布の表を作成した。Fisher は  $t$  分布を与え、Z 分布についての公式的な証明をつけ加えた。

$t$  分布に基いた確率表に關係して注意すべき第一の点は、理論において  $\sigma^2$  推定に利用できる自由度  $(n-1)$  の各々の値にたいして、色々な分布があるということである。全範囲にわたるように  $(n-1)$  の表が用意される。ここに  $(n-1)$  は1から $\infty$ 迄の値をとる。これは明らかに不可能であるが、分布の検定に際して  $(n-1) = 30$  以上では変化は非常に小さく、その見地から  $(n-1) = \infty$  に基づく  $t$  表を用いればよいことが分る。

しかし、全範囲にわたるためには、各々正規曲線の確率積分の表と同じ位の大きさの30枚の表が必要となる。これは亦非常に努力のかかる仕事である。 $p$  の与えられた値にたいして、 $n-1$  が1から30迄それから40, 60, 120 および  $\infty$  の範囲にわたるものの値を与える表が Fisher および Fisher & Yates により簡単にまとめられた。 $p$  の値は主に0.01, 0.05 および0.10のような有意性検定に用いられる値である。

さて数値例にもどつてみると次のことが分る。

$$t = \frac{\bar{X}_1 - \bar{X}_2}{S_{md}} = \frac{4.76}{1.12} = 4.25$$

表A-3をみると  $n-1=11$  にたいして  $p=0.01$  の  $t$  の値は3.11である。そして我々の  $t$  の値はこれより大きいから  $p$  は0.01以下であることが分る。2品種の収穫能力には本当に差があることが立証された。

対の変量系列の平均値の差にも検定を適用する方法をまとめると

$$S_d^2 = \frac{\sum (X_1 - X_2)^2}{n-1} \quad (7)$$

$$\text{ここに } \sum (X_1 - X_2)^2 = \sum (X_1 - X_2)^2 - \frac{(\sum X_1 - \sum X_2)^2}{n} \quad (8)$$



$$S_{md}^2 = \frac{S_d^2}{n} \quad (9)$$

$$t = \frac{\bar{X}_1 - \bar{X}_2}{S_{md}} = (\bar{X}_1 - \bar{X}_2) \sqrt{\frac{n(n-1)}{\sum (x_1 - x_2)^2}} \quad (10)$$

t表は自由度 (n-1) についてよまれる。

#### 4. 変量が対でない場合のt検定の応用

上に与えた例はプロットの対の系列にたいするものであつた。そして  $S_{md}^2$  を推定する際に我々は方程式にはいつてくる  $\sum (x_1 - x_2)^2$  の値を考慮しなければならぬことを注意した。変量が対でないとき、この項は考えなくてもよい。何故ならば変量  $X_1$  および  $X_2$  が独立に抽出されたとすれば正および負の積  $x_1 x_2$  は互に打消すからである。そのとき次の簡単な方程式を得る。

$$S_d^2 = S_1^2 + S_2^2 \quad (11)$$

これには二つの場合がある。(1) 2つの標本において変数の数が等しい場合、即ち  $n_1 = n_2$  の場合。(2)  $n_1 \neq n_2$  なる場合。これらの場合にたいし  $S_{md}^2$  を推定する公式は多少違つてくる。

(1) 変量が対でなく、 $n_1 = n_2 = n$  の場合。

$$S_d^2 = \frac{\sum x_1^2 + \sum x_2^2}{2(n-1)} \quad (12)$$

$$S_{md}^2 = \frac{S_d^2}{n} + \frac{S_d^2}{n} = S_d^2 \left( \frac{2}{n} \right) \quad (13)$$

$$t = \frac{\bar{X}_1 - \bar{X}_2}{S_{md}} = (\bar{X}_1 - \bar{X}_2) \sqrt{\frac{n(n-1)}{\sum x_1^2 + \sum x_2^2}} \quad (14)$$

そして計算には次を用いる

$$\sum x_1^2 = \sum md_1^2 - \frac{(\sum md_1)^2}{n} \quad \text{及び} \quad \sum x_2^2 = \sum X_2^2 - \frac{(\sum X_2)^2}{n} \quad (15)$$

t表をみる際自由度は  $2(n-1)$  であることに注意せよ

(2) 変量が対でなく  $n_1 \neq n_2$  の場合

$$S_d^2 = \frac{\sum x_1^2 + \sum x_2^2}{n_1 + n_2 - 2} \quad (16)$$

$$S_{md}^2 = S_d^2 \left( \frac{n_1 + n_2}{n_1 n_2} \right) \quad (17)$$

$$t = \frac{\bar{X}_1 - \bar{X}_2}{S_{md}} = (\bar{X}_1 - \bar{X}_2) \frac{n_1 n_2 (n_1 + n_2 - 2)}{(n_1 + n_2) (\sum x_1^2 + \sum x_2^2)} \quad (18)$$

t表をみる際、自由度は  $n_1 + n_2 - 2$

対でない変量の組に適用する検定は、母集団からとられた標本は同じ分数をもち、 $S_1^2$  と  $S_2^2$  の間の観測された差は無作為抽出のみによるという仮定に基いている。別ないい方をすると我々が実際に検定したいものは平均値間の差であるが我々が実際に検定しているのは平均値の差とこみにした標準偏差との比即ちtの有義性である。最密にはこの検定は標準偏差が抽出さるべき母集団で同一であると仮定できたときのみ平均値の差に適用できる。標準偏差が違つていられる場合の方法は、Fisher & Yates によつて与えられた表、Behrens - Fisher 検定として知られているものを適用することである。下に示す近似検定を行えば非常に多くの例において計算が幾分簡単になる。

(a)  $n_1 = n_2 = n$  の場合にはtを推定するため (n-1) なる自由度が利用できる場合のみ変量が対にされていると仮定しても通例十分である。有義性に対して必要なtの値はそれにしたがつて増加する。

(b)  $n_1 \neq n_2$  の場合には Cochran & Cox が述べた方法で十分である。次を計算する。

$$S_1^2 = \frac{\sum x_1^2}{n_1 - 1} \quad S_2^2 = \frac{\sum x_2^2}{n_2 - 1} \quad (19)$$

$$S_{X_1}^2 = \frac{S_1^2}{n_1} = \omega_1 \quad S_{X_2}^2 = \frac{S_2^2}{n_2} = \omega_2 \quad (20)$$

そのときもし  $\bar{X}_1 - \bar{X}_2 = d$  を平均値間の差とすれば

$$t = \frac{d}{\sqrt{\omega_1 + \omega_2}} \quad (21)$$

与えられた水準での有義性に対して、tの値は次によつて与えられる。

$$t' = \frac{t_1 \omega_1 + t_2 \omega_2}{\omega_1 + \omega_2} \quad (22)$$

ここに  $t_1$  は自由度 ( $n_1 - 1$ ) に対する5%点での値で  $t_2$  は自由度 ( $n_2 - 1$ ) に対する5%点での値である。この方法は例4-2に示す。

#### 5. 例4-1. 変量が対でない場合のt検定の応用

3節の資料を計算方法の例に用い得る。まづ第一に

$$\sum x_1^2 = 24.4^2 + 27.9^2 + \dots + 31.5^2 - \frac{321.0^2}{12} = 8806.30 - 8586.75 = 219.55$$

$$\sum x_2^2 = 17.5^2 + 15.1^2 + \dots + 25.2^2 - \frac{263.9^2}{12} = 6168.91 - 5803.60 = 365.31$$

$$S_d^2 = \frac{219.55 + 365.31}{22} = 26.5845$$

$$S_{md}^2 = 26.5845 \left( \frac{2}{12} \right) = 4.4308$$



$$t = \frac{25.75 - 21.99}{\sqrt{4.4308}} = 2.26$$

ここで $t$ の推定には自由度22を用いる；したがって自由度22なる5%点の $t$ をみる。これは2.08である。したがってこの検定によれば観察された差は有意である。4節で行った検定と比較すると方程式(4)の右辺の第三項 $S_d^2$ の計算を考慮にいれていない点が違っている。ここでは $t$ が有意となる点まで $S_d^2$ が増した。結果は実験の正しい計画および分析の値を示している。まづ第一に、誤差を減少する方法としては出来る場合には比較の対を用いなければならない。第二に比較を対にしたとき、正しい分析方法にしたがうことが本質的である。更に比較を対にしても差の減少が期待されない場合には、 $t$ の推定に用いられる自由度は減少するから対からの正確度に損失があり又与えられた有意の程度にたいしてより大きい $t$ の値が必要となる。

6. 例4-2.  $n_1 \neq n_2$  の場合二つの母集団の分散が違っていると仮定したときの $t$ 検定の応用

次の資料は J. A. Anderson の Protein Survey of Western Canadian Wheat 1950 Crop からとつたものである。第一行は Saskatchewan における個々の場毎にたいして、非常に多くの蛋白質含有量から無作為にとつた8つの値を示す(標本1)。第二行は Manitoba における同様な系列を表わす(標本2)

標本1 15.1 14.3 11.5 14.5 15.4 12.5 14.6 16.6

標本2 12.2 12.5 11.2 12.6 11.0 11.6 12.0 12.5 11.8 12.4 11.5 12.0  
11.6 12.7

$$\Sigma X_1 = 114.5 \quad \bar{X}_1 = 14.31 \quad d = 14.31 - 11.97 = 2.34$$

$$\Sigma X_2 = 167.6 \quad \bar{X}_2 = 11.97$$

$$\Sigma X_1^2 = 1657.13 - \frac{(114.5)^2}{8} = 18.35$$

$$S_1^2 = \frac{18.35}{7} = 2.6214 \quad S_{\bar{X}_1}^2 = \frac{2.6214}{8} = 0.328$$

$$\Sigma X_2^2 = 2010.20 - \frac{(167.6)^2}{10} = 3.79$$

$$S_2^2 = \frac{3.79}{9} = 0.2915 \quad S_{\bar{X}_2}^2 = \frac{0.2915}{10} = 0.021$$

$$t = \frac{2.34}{\sqrt{0.328 + 0.021}} = 3.97$$

$$t_1 = 2.35 \quad t_2 = 2.16$$

$$t' = \frac{2.35 \times 0.328 + 2.16 \times 0.021}{0.328 + 0.021} = 2.35$$

$t$ は2.35なる近似的な5%点より大きいので、蛋白質含有量での差は有意と考えることが

できる。

この例では2つの標本に対する平均平方には非常にはつきりした差がある。比は2.6214/0.2915=0.9である。分散分析を取扱う第5章で、この比を有意性の検定に用い得ることを示す。この比に対する5%点は2.84で1%点は4.44である。したがってここでは抽出した二つの母集団の分散は明らかに差があり、もつと精密な検定法を適用すべきであろう。通常の方法では我々は、自由度20にたいする $t$ の5%点をとつた。これは2.09であるからより正確な検定法では与えられた有意水準にたいして平均値間のより大なる差を必要とすることは明らかである。

#### 7. 帰無仮説

3節の $t$ 検定では、品種の収量間の差は平均値0のまわりに正規に分布するという仮説をたてた。R. A. Fisher 教授はこれを帰無仮説と名づけた。これは非常に簡単な概念であり、混乱のおそれはない。貨幣投げにおいては、帰無仮説は貨幣には偏倚がないということである。不偏な貨幣での表の分布は $n$ 回投げた場合には $(\frac{1}{2} + \frac{1}{2})^n$ なる二項展開で与えられる。したがって帰無仮説は理論分布を誘導するための基礎である。理論分布は実際分布と一致するものでなく、帰無仮説から直接生ずるということを明らかに心にためておかねばならない。

すべての有意性検定においては仮説をたてる必要がある。第二の考え方は、仮説から得られる理論分布は既知であるということである。もしも分布が分からなくてもそれを数学的方法によつて導くことは可能である。第三段階は得た結果と仮説にしたがって得た結果の理論分布とを比較することで、これからその統計的有意性に関する何等かの結果を得ることである。

#### 8. 信頼限界

帰無仮説は設定し得る無数の仮説の中ただ一つのものである。例えば3節で行った検定では品種間の差は平均値3.00のまわりに正規に分布するという仮説を設けることもできるであろう。 $\bar{X}_1 - \bar{X}_2 = d$ とおくと帰無仮説は次の公式からの $t$ の計算を行うことになる。

$$\pm t = \frac{d - 0}{S_{md}} \quad (23)$$

平均値が3であるという仮説は次の公式によつて $t$ の計算を行うことになる。

$$\pm t = \frac{d - 3.00}{S_{md}}$$

いいかえれば $t$ にたいする一般公式は

$$\pm t = \frac{d - m}{S_{md}} \quad (24)$$

ここに $m$ は必要な仮説にしたがってどのようにでも与えることができる\*。

以上の一般式をどのように利用する事が出来るかというきもんが生ずる。

R. A. Fisher 教授は更に帰無仮説の検定の他に計算した $t$ が $t$ 表の与えられた百分率点での $t$ と等しいような2つの $m$ を公式(24)を用いて定めることによつて更に進んだ知識



を得ることができるという非常に価値ある示指を行つた。それらの $m$ の値は許容限界として知られているものである。例えば3節の資料を用いると

$$\pm t = \frac{4.76 - m}{1.12}$$

したがつて

$$m_1 = 4.76 + 1.12t$$

$$m_2 = 4.76 - 1.12t$$

そのとき自由度11なる5%での $t$ の値は2.20であるから

$$m_1 = 4.76 + 2.20 \times 1.12 = 7.22$$

$$m_2 = 4.76 - 2.20 \times 1.12 = 2.30$$

これらの信頼限界の意味は次によつて明らかとなる。即ちもし我々が $m$ にたいして7.22に等しいか又はそれ以上の値を選べば、 $t$ 検定から得た $p$ の値は $\leq 0.05$ となるであろう。同様に $m$ が2.30に等しいかそれ以下であれば $p \leq 0.05$ である。ほかのいい方でこの立場をのべると2.30と7.22との間の $m$ の値は5%水準では4.76と有意の差はない。2.30と7.22の範囲以外のいかなる値も5%水準で4.76と有意の差がある。したがつて我々は品種収量間の平均差は2.30と7.22との間にあるという合理的に良好な保証を得る。

信頼限界の意味に関連して2品種の検定と同様に非常に多くの標本の抽出及び各標本からの信頼限界の決定を心に描くことがのぞましい。勿論信頼限界は標本が変れば違つてくることが期待される。このような決定の95%において、計算した信頼限界は真の平均値を含むと期待される。真の平均値を表わす垂直線を書くことは価値あることである。そのとき1つの標本にたいする信頼限界は水平線で表わすことができる。各標本にたいして限界をプロットし5%点で限界を計算するとその中の95%は真の平均値を表わす垂直線を横切ることが期待される筈である。

研究者は上の問題に関して真の平均値が2.30と7.22との間にある確率を0.05であるようにきめたと論述をすることは用心しなければならない。真の平均値は母数であり、したがつて一定値である。そして我々は真の平均値がある場処を正確な確率をもつて述べることはできない。

\* 注)  $t$ にたいする一般公式は次のようにするのが正しい。

$$\pm t = \frac{\bar{X} - M}{S_{\bar{X}}}$$

(24) では $d$ は $\bar{X}$ に依する変量で  $S_{md}$  は  $S_{\bar{X}}$  に依する変量であるから  
公式は同じものである。

## 第5章 分散分析

### 1. 定義

基礎的には分散分析は与えられた結果の組においてその変動を成分に分割する代数的方法である。これを理解するためには変動の成分とは何かということを知る必要がある。

変動が不斉である場合には一つ以上の成分が存在する。同じ種族である母集団から抽出された男性の一群を考え、研究しようとする変量がインチで表わした身長であると仮定しよう。度数分布を画くと近似的に正規であることが分つた。したがつてこの群の変動はほぼ斉であると仮定してもよい。同様なことが女性の一群にもいえるであろう。しかしもしもこれらの2群の資料を一緒にして新しい群を作れば変動の第二の成分が入ってくる。即ちそれは2群の平均値間の差である。組合せた群にたいする度数分布はこの差によつて、おそらく二つの尖り又はモードを示すほど変動が大となるであろう。更に同様な論法で第三の群は13~15才の少年で、第四の群は同年の少女であるとしよう。これらの4つの群を全部組合せたときには度数分布はあたかも正規分布であるかのように見えるが、変動の二つの成分、即ち一つは群内の変動を表わし他は群間の変動を表わす二つの成分が実際には存在することが分る。分散分析の代数的な方法を用いるとこのような混合母集団の変動の成分を分割し、解明することが可能である。

完全な分散分析には、実際には二つの役割がある。まづ第一には分散成分を分割し推定する。第二には有意性の検定を与える。分割方法は単に機械的であつて、すべての場合に適用できるが、そのようにして得た分散の推定量の信頼度は資料をあつめた方法の如何による。これは分析方法の詳細を明らかにする前には説明することが困難であるから後の節にゆづる。しかし有意性検定の考え方は、分析方法及びそれが作られた理論を述べる際に同時にとりあげよう。

### 2. 基礎原理

第2章4節において、母集団の標準偏差と、大きさ $n$ なる標本平均の標準偏差との間の関係を述べた。分散の項で表わすと次のようである。

$$n\sigma_{\bar{x}}^2 = \sigma^2 \quad (1)$$

ここに $n$ は各標本における数、 $\sigma_{\bar{x}}^2$ は平均値の分散、 $\sigma^2$ は母集団の分散である。この式から非常に簡単ではあるが重要な推論を行うことができる。大きさが $n$ の標本の系列を抽出すると一つの標本にたいして平均平方 $V_1$ をどんな場合でも計算できる。更に我々は標本平均から $V_{\bar{x}}$ を計算できる。そして上の式(1)から $nV_{\bar{x}}$ を $\sigma^2$ の推定量としてとることができることは明らかである。

上の考え方をすつきりさせるためには、次の記号で標本の系列を表わそう。



| 標 本 | 変 量      |          |       |          |              | 平 均 値       |
|-----|----------|----------|-------|----------|--------------|-------------|
| 1   | $X_{11}$ | $X_{12}$ | ..... | $X_{1j}$ | ... $X_{1n}$ | $\bar{X}_1$ |
| 2   | $X_{21}$ | $X_{22}$ | ..... | $X_{2j}$ | ... $X_{2n}$ | $\bar{X}_2$ |
| ⋮   | ⋮        | ⋮        |       | ⋮        |              | ⋮           |
| i   | $X_{i1}$ | $X_{i2}$ | ..... | $X_{ij}$ | ... $X_{in}$ | $\bar{X}_i$ |
| ⋮   | ⋮        | ⋮        |       | ⋮        |              | ⋮           |
| k   | $X_{k1}$ | $X_{k2}$ | ..... | $X_{kj}$ | ... $X_{kn}$ | $\bar{X}_k$ |

こゝに各々n個の変量をもつk個の標本がある。記号で具体的な意味を与えるためにkn匹のネズミがk個の標本に無作為に分けられたと考えよう。それから $X_{11}$ は第一の標本の第一番目のネズミの重量である。また $X_{ij}$ はi番目の標本の第j番目のネズミの重量である。

各標本にたいして次が計算できる。

$$V_1 = \frac{\sum_{j=1}^n (X_{1j})^2}{n-1} \quad (2)$$

こゝに $X_{1j}$ は標本平均からの単一偏差( $X_{1j} - \bar{X}_1$ )を表わす。

これらの変量の各々は $\sigma^2$ の不偏推定量であるから系列は単一推定量を求めるため平均できる。即ち

$$V = \frac{\sum_{i=1}^k (V_i)}{k} = \frac{\sum_{i=1}^k \sum_{j=1}^n (X_{ij})^2}{k(n-1)} \quad (3)$$

これは各標本平均からのすべての偏差の平方和を標本内で利用できる自由度の総数で割つたものである。

$nV_{\bar{x}}$ を求めるには

$$nV_{\bar{x}} = \frac{\sum_{i=1}^k (\bar{X}_i - \bar{X})^2}{k-1} \quad (4)$$

こゝに $\bar{X}$ は総平均即ちすべての標本平均の平均値である。

もしすべての標本における変量を無作為に抽出するとすればV及び $nV_{\bar{x}}$ 共に $\sigma^2$ の不偏推定量である。勿論Vは $k(n-1)$ の自由度に基いたものであり、 $nV_{\bar{x}}$ は $(k-1)$ 個のみの自由度に基いたものであるから同様に信頼できない。無作為に配列すると一匹のネズミが等しい確率でどの標本にも含まれるということは確実である。したがってV及び $nV_{\bar{x}}$ 双方の値に影響するただ一つの要因はそれから標本が抽出された母集団の変動性であるということとは

明らかである。

実際的な数字をこのような場合に適用すると面白い。下に示す数字は5つの変量の5群に無作為におかれたものである。

| 群 | 変 量 |    |    |    |    | 平 均 値 |
|---|-----|----|----|----|----|-------|
| 1 | 29  | 45 | 14 | 25 | 11 | 24.8  |
| 2 | 44  | 38 | 48 | 31 | 31 | 38.4  |
| 3 | 16  | 29 | 29 | 18 | 46 | 27.6  |
| 4 | 37  | 11 | 31 | 28 | 44 | 30.2  |
| 5 | 50  | 12 | 19 | 20 | 28 | 25.8  |

Vは公式3から計算されるが、後節で述べる簡略法により

$$V = \frac{3006.80}{20} = 150.34$$

又

$$nV_{\bar{x}} = \frac{594.96}{4} = 148.74$$

通常期待するよりもこの場合はよく一致している。計算目的は実際的な数字で代数的な記号の意味を示すことである。しかしこのような標本の大きな系列ではV及び $nV_{\bar{x}}$ は非常によく一致すると考えられる。

我々は上に概要を述べた論題の下では、平均値による分散は群間の根本的な差によつて影響されるという。第1節に述べに例のように、母集団が斉一でない場合にとるべき立場に主として関心がある。ネズミの群では各群に別個の糧食が与えられたかも知れない。群内に変動が生ずるのは母集団がもともと持っている変動によるものであるうか標本平均は後から与えられた糧食の差によつても変動するであろう。この場合記号で次のように表わすことができる。

| 標 本 | 変 量            |                |       |                |  | 標本平均              | 平方平均  |
|-----|----------------|----------------|-------|----------------|--|-------------------|-------|
| 1   | $X_{11} + Y_1$ | $X_{12} + Y_1$ | ..... | $X_{1n} + Y_1$ |  | $\bar{X}_1 + Y_1$ | $V_1$ |
| 2   | $X_{21} + Y_2$ | $X_{22} + Y_2$ | ..... | $X_{2n} + Y_2$ |  | $\bar{X}_2 + Y_2$ | $V_2$ |
| ⋮   | ⋮              | ⋮              |       | ⋮              |  | ⋮                 | ⋮     |
| k   | $X_{k1} + Y_k$ | $X_{k2} + Y_k$ | ..... | $X_{kn} + Y_k$ |  | $\bar{X}_k + Y_k$ | $V_k$ |

標本の各変量の一つの値は、すべての変量を無作為に抽出した場合のXなる部分と各変量にたいして一定であるような部分Yとの和からなつていことに注意せよ。したがって標本1の平均は $\frac{\sum X_{1j}}{n} + \frac{nY_1}{n} = \bar{X}_1 + Y_1$ である。Yは一つの標本では一定であるから $V_1$ の値には影響しない。標本1にたいして代数的に次を得る



$$V_1 = \frac{\sum_1^n [(X_{11} + Y_1) - (\bar{X}_1 + \bar{Y}_1)]^2}{n - 1} = \frac{\sum_1^n (X_{11} - \bar{X}_1)^2}{n - 1} = \frac{\sum_1^n (x_{11}^2)}{n - 1} \quad (5)$$

したがってすべての標本に基くVはYなる要因が存在していない場合と同じである。

標本平均にたいする新しい平均平方を検定するとYの値の間の差が影響することが明らかとなる。それを $V'_x$ で表わしてみよう。

$$\begin{aligned} V'_x &= \frac{\sum_1^k [(\bar{X}_1 + Y_1) - (\bar{X} + \bar{Y})]^2}{k - 1} = \frac{\sum_1^k [(\bar{X}_1 - \bar{X}) + (Y_1 - \bar{Y})]^2}{k - 1} \\ &= \frac{\sum_1^k (\bar{X}_1 - \bar{X})^2}{k - 1} + \frac{\sum_1^k (Y_1 - \bar{Y})^2}{k - 1} + \frac{2 \sum_1^k (\bar{X}_1 - \bar{X})(Y_1 - \bar{Y})}{k - 1} \\ &= V_{\bar{X}} + V_{\bar{Y}} + \frac{2 \sum_1^k (\bar{X}_1 - \bar{X})(Y_1 - \bar{Y})}{k - 1} \quad (6) \end{aligned}$$

$V'_x$ は $V_{\bar{X}} + V_{\bar{Y}}$ に偏差の積和の項を加えたものである。

さて $(\bar{X}_1 - \bar{X})$ で表わされる標本平均と総平均との差の部分は全く無作為抽出から生じ、 $(Y_1 - \bar{Y})$ のそれに対応する値とは無関係である。したがって大きな標本系列ではこの積和は0と仮定することができる。

そのとき次のものが残る。

$$V'_x = V_{\bar{X}} + V_{\bar{Y}} \quad (7)$$

又は  $nV'_x = nV_{\bar{X}} + nV_{\bar{Y}}$

このような標本系列にたいしては $nV'_x$ は $nV_{\bar{X}}$ よりも $nV_{\bar{Y}}$ だけ大であることがわかる。

上の数値例は今迄の代数的な論述を具体的にはどのようにして計算すればよいかを示すため、次のように変換することができる。例えば  $Y_1 = 0, Y_2 = +2, Y_3 = 0, Y_4 = 0$  及び  $Y_5 = -2$  としよう。そのときの数字は

| 群 | 重  | 量  | 平均値 |
|---|----|----|-----|
| 1 | 29 | 45 | 14  |
| 2 | 46 | 40 | 50  |
| 3 | 16 | 29 | 29  |
| 4 | 37 | 11 | 31  |
| 5 | 48 | 10 | 17  |
|   |    |    | 25  |
|   |    |    | 11  |
|   |    |    | 33  |
|   |    |    | 33  |
|   |    |    | 46  |
|   |    |    | 28  |
|   |    |    | 44  |
|   |    |    | 26  |

これから次を得る。

$$V = \frac{3006.80}{20} = 150.34 \quad \text{これは前と同じである。}$$

$$nV'_x = \frac{886.97}{4} = 221.74$$

後者は前の数字148.74よりかなり大となり、Yの影響を示している。

実際的な実験では無作為変量のみを示す数字の組がなければ  $V'_x$ の値は分らない。我々がもっているのはXとYを一緒にした最終的な数字である。しかしVは $nV'_x$ の良好な推定量であることは分っている。したがって(7)からの

$$nV_{\bar{Y}} = nV'_x - V$$

は全分散の中標本に適用した処理から生ずるYの影響にわりあてられるべき平均値による部分の良好な推定量を与える。こゝでは

$$221.74 - 150.34 = 71.40$$

この方法は処理による分散成分の推定として知られており分散分析の重要な役目の一つである。

同様な一般原理は資料が処理の系列の簡単な反復実験のように二つの方向に分類されているときにも成立する。資料は次のように配列される。

|   |     | 処 理         |             |       |             |
|---|-----|-------------|-------------|-------|-------------|
| 反 | 1   | $X_{11}$    | $X_{12}$    | ..... | $X_{1n}$    |
|   | 2   | $X_{21}$    | $X_{22}$    | ..... | $X_{2n}$    |
|   | ... | ...         | ...         | ..... | ...         |
|   | r   | $X_{r1}$    | $X_{r2}$    | ..... | $X_{rn}$    |
|   |     | $\bar{T}_1$ | $\bar{T}_2$ | ..... | $\bar{T}_n$ |

こゝに $X_{11}$ は反復1の処理1での収量を示す。この表は各々n個からなるr個の標本又は各々r個からなるn個の標本と考えられる。したがってもし分散が $\sigma^2$ である母集団から変量をすべて無作為に抽出し上例のように列及び行に配列したとすれば分散は次の三通りに分けることができる。(1) 列平均からの分散 (2) 行平均からの分散、(3) 列及び行によつて影響されない残差平均平方からの分散。

これをはつきりさせるためには各収量をXだけでなく  $X+Y+Z$  で表わすと便利である。そのときYは反復を表わす部分、Zは処理を表わす部分である。そのとき各n個からなるk個の標本に例を適用する場合と同じ理由で、反復平均からの母分散の推定量は



$$nV_{\bar{R}}' = nV_{\bar{R}} + nV_Y \quad (9)$$

又処理平均からの母分散の推定量は

$$rV_{\bar{T}}' = rV_{\bar{T}} + rV_Z \quad (10)$$

で与えられる。

(9)式の $nV_Y$ は反復間の真の変動による平均平方の部分でプロット毎の無作為変動の部分ではない。又(10)式の $rV_Z$ は処理のみによる平均平方の成分である。

各変量を $(X+Y+Z)$ で表わすとき、 $Y_1$ を反復1による特性、 $Z_1$ を処理1による特性とした場合に第1の変量を $(X_{11}+Y_1+Z_1)$ で表はすことができる。そのとき無作為変動のみの推定量は

$$(X_{11}+Y_1+Z_1-M)-[(\bar{X}_{R1}+Y_1+\bar{Z})-M] \\ -[(\bar{X}_{T1}+Y+Z_1)-M] \quad (11)$$

となることに注意しなければならぬ。こゝに $M$ は総平均 $\bar{X}_{R1}$ は $X_{11}$ を含む反復平均、 $\bar{X}_{T1}$ は $X_{11}$ を含む処理平均で

$$M=\bar{X}+\bar{Y}+\bar{Z} \quad \text{であるから}$$

$$(X_{11}-\bar{X})-(\bar{X}_{R1}-\bar{X})-(\bar{X}_{T1}-\bar{X}) \quad (12)$$

のように簡単になる。これは総平均からの個々の偏差から反復及び処理平均による偏差を差引くとすれば残りの部分が無作為変動による部分のみであることを示すのである。

反復及び処理によるすべての影響が取除かれて $\sigma^2$ の推定量である平均平方がどのようにして計算できるかを示す。実際に計算をする方法は後節で示そう。

### 3. 有意性の検定

第5章2節においては、処理の影響のために処理又は標本平均から推定した平均平方 $V_{\bar{X}}'$ は無作為抽出変動による分散の推定量 $V$ より大となる傾向があることを注意した。 $V$ は誤差分散、 $nV_{\bar{X}}'$ は処理分散と呼ばれる。

このような場合我々は得た結果に関してある種の確率的論述ができるようにすることが必要である。このことは $nV_{\bar{X}}'/V$ のような此の理論分布がきまれば可能となろう。問題は R. A. Fisherが $Z=\frac{1}{2}\log_e(V_1/V_2)$ の理論分布を与えたときに解決された。こゝで $V_1$ と $V_2$ とは同一の分散のそれぞれ自由度 $n_1$ 及び $n_2$ なる独立な推定量である。

この場合には $nV_{\bar{X}}'$ は自由度 $(k-1)$ 、 $V$ は自由度 $k(n-1)$ に基いている。比は

$$\frac{nV_{\bar{X}}'}{V} = \frac{nV_{\bar{X}}}{V} + \frac{nV_Y}{V} \quad (13)$$

であるが $nV_{\bar{X}}'$ 又は $V$ に等しくなる傾向があるので次のようになる。\*

$$\frac{nV_{\bar{X}}'}{V} = 1 + \frac{nV_Y}{V} \quad (14)$$

したがって事実上、有意性の検定は $nV_Y/V$ が0と違っている程度をきめることである。

有意性の検定をするための完全な表は、 $n_1$ 及び $n_2$ の各対にたいする分布が違ふから非常に面倒となろう。しかし Fisher 及び Fisher & Yates が作った $n_1$ 及び $n_2$ の特定の値の系列及び各確率水準にたいしての完全な表を利用すればこの困難はなくなる。Snedecor は Fisher の表を分散比とよばれる $V_1/V_2$ の形にした。これは記号 $F$ で表わす。Fisher & Yatesは $Z$ 及び $F$ の両方にたいする表を与えた。選ばれた確率水準は0.01、0.05及び0.10のような普通用いられる有意水準である。上に与えた数値例では

$$F = \frac{221.74}{150.34} = 1.47$$

分子の平均平方は自由度4で、分母の平均平方は自由度20である。A-6表で $n_1=4$ 、 $n_2=20$ をみると5%点での $F$ の値は2.87である。分散成分 $V_Y$ は0と有意な差がないと結論されるであろう。

### 4. 直交性

我々は2節から処理平均間 $(\bar{X}_1+Y_1)$ の変動は $V$ に対して何等の影響も与えないということを知った。何故なら平均平方 $V$ は標本内の変位間の変動から出てくるからである。これが二つの平均平方間の関係であるときにはそれらは直交しているという。直交性と独立性とはほとんど同じ意味をもつ数学的な言葉である。実験において、別個に推定でき、決して入りまじらない二つの効果はお互に直交である。厳密にいえば二つの効果は数学的な分析をしなければ必ずしも直交性があるとはいえないが、複雑な場合でもそれほど深く考えずに適用しても間違いない一般的な規則がある。このことは次の簡単な例で明らかである。実験が3反復の4処理から成立つていとする。そして実験を行つた後、次のような結果をまとめた表を作ることができる。

\*注) 二つの平方平均の比の平均値は実際には $n_1/(n_2-2)$ であつて1より少し大である。一般の場合には殆んど1に近いから(14)は近似的に真であるとしてもよい。



|      |   | 処 理         |             |             |             | 反復平均        |
|------|---|-------------|-------------|-------------|-------------|-------------|
|      |   | A           | B           | C           | D           |             |
| 反 復  | 1 | $X_{A1}$    | $X_{B1}$    | $X_{C1}$    | $X_{D1}$    | $\bar{R}_1$ |
|      | 2 | $X_{A2}$    | $X_{B2}$    | $X_{C2}$    | $X_{D2}$    | $\bar{R}_2$ |
|      | 3 | $X_{A3}$    | $X_{B3}$    | $X_{C3}$    | $X_{D3}$    | $\bar{R}_j$ |
| 品種平均 |   | $\bar{T}_A$ | $\bar{T}_B$ | $\bar{T}_C$ | $\bar{T}_D$ | G           |

この表を検討すると、各品種の一つのプロットは各反復をうけていることは明らかである。処理Aは他のものよりも多い収量を与えると仮定しよう。これは処理Aは各反復に含まれているから反復間変動には影響しない。同様にもしも反復1が他のものよりも多い収量を与えても処理間変動には影響しない。したがって我々は処理及び反復を表す平均平方は直交性があると断定できる。それと反対の立場を考えてみると、Aが処理の中最大の収量であり、反復1におけるAのプロットはないものと仮定しよう。これは $\bar{R}_2$ 及び $\bar{R}_3$ と比較して $\bar{R}_1$ が減少する傾向であり、 $\bar{T}_A$ ,  $\bar{T}_B$ ,  $\bar{T}_C$ ,  $\bar{T}_D$  及び $\bar{R}_1$ ,  $\bar{R}_2$ ,  $\bar{R}_3$  で表わした処理及び反復の効果はもはや直交ではないであろう。

## 5. 平方和及び自由度の割当

次のような結果をまとめた表があるとしよう。

|     |      | 処 理      |          |          |       |          | 反復総計  |
|-----|------|----------|----------|----------|-------|----------|-------|
|     |      | 1        | 2        | 3        | ..... | n        |       |
| 反 復 | 1    | $X_{11}$ | $X_{12}$ | $X_{13}$ | ..... | $X_{1n}$ | $R_1$ |
|     | 2    | $X_{21}$ | $X_{22}$ | $X_{23}$ | ..... | $X_{2n}$ | $R_2$ |
|     | ...  | ...      | ...      | ...      | ...   | ...      | ...   |
| 処 理 | r    | $X_{r1}$ | $X_{r2}$ | $X_{r3}$ | ..... | $X_{rn}$ | $R_r$ |
|     | 処理総計 | $C_1$    | $C_2$    | $C_3$    | ..... | $C_n$    | G=全総計 |

一つの観測例えは $X_{11}$ にたいして次のように書ける。

$$(X_{11} - \bar{X}) = (X_{11} - \bar{R}_1) + (\bar{R}_1 - \bar{X})$$

こゝに $\bar{X}$ は総平均で、 $\bar{R}_1$ は反復1の平均である。そのとき

$$(X_{11} - \bar{X})^2 = (X_{11} - \bar{R}_1)^2 + (\bar{R}_1 - \bar{X})^2 + 2(X_{11} - \bar{R}_1)(\bar{R}_1 - \bar{X})$$

で、反復1について加えると

$$\sum_1^n (X - \bar{X})^2 = \sum_1^n (X - \bar{R}_1)^2 + n(\bar{R}_1 - \bar{X})^2 + 2(\bar{R}_1 - \bar{X}) \sum_1^n (X - \bar{R}_1)$$

最後の項は平均からの偏差の和を含むから0である。 $\left[ \sum_1^n (X - \bar{R}_1) = 0 \right]$ 。第二項は、 $(\bar{R}_1 - \bar{X})$ は一定であるから加えるとn倍になる。前後の項を除くと

$$\sum_1^n (X - \bar{X})^2 = \sum_1^n (X - \bar{R}_1)^2 + n(\bar{R}_1 - \bar{X})^2$$

さて各反復についてこれをくりかえし、全実験で加えあげると

$$\sum_1^{rn} (X - \bar{X})^2 = \sum_1^r \sum_1^n (X - \bar{R}_r)^2 + n \sum_1^r (\bar{R}_r - \bar{X})^2 \quad (15)$$

これは全平方和は反復内平均からの偏差を表わす部分と、総平均からの反復平均の偏差を表わす部分とに分けることができるという非常に重要な方程式である。簡単に総平均からの $X_{11}$ の偏差を代数的に展開して

$$\begin{aligned} (X_{11} - \bar{X}) &= (\bar{R}_1 - \bar{X}) + (\bar{C}_1 - \bar{X}) + [(X_{11} - \bar{X}) - (\bar{R}_1 - \bar{X}) - (\bar{C}_1 - \bar{X})] \\ &= (\bar{R}_1 - \bar{X}) + (\bar{C}_1 - \bar{X}) + (X_{11} - \bar{R}_1 - \bar{C}_1 + \bar{X}) \end{aligned}$$

とかけば、全平方和は

$$\sum_1^{rn} (X - \bar{X})^2 = n \sum_1^r (\bar{R} - \bar{X})^2 + r \sum_1^n (\bar{C} - \bar{X})^2 + \sum_1^{rn} (X - \bar{R}_1 - \bar{C}_1 + \bar{X})^2 \quad (16)$$

$$\text{総和} = \text{反復} + \text{処理} + \text{無作為変動}$$

に分けることができる。こゝに、1からrまで、jは1からnまでである。

自由度も同様に割当てることができる。平均にたいする補正は自由度が1である。したがって総計 $(rn-1)$ の自由度が割当てに利用できる。反復にたいして $(r-1)$ 、処理にたいして $(n-1)$ の自由度をとれば、無作為変動にたいして

$$(rn-1) - (r-1) - (n-1) = (r-1)(n-1)$$

が残る。したがって (16) に応ずる自由度にたいしては

$$rn-1 = (r-1) + (n-1) + (r-1)(n-1)$$

$$\text{総計} = \text{反復} + \text{処理} + \text{無作為変動}$$

完全な分散分析は無作為変動を誤差項として表わして、下に示したような表の形に作ると最も便利である。

|     | 平方和                                        | 自由度          | 平均平方           |
|-----|--------------------------------------------|--------------|----------------|
| 反 復 | $n \sum (\bar{R} - \bar{X})^2$             | $r-1$        | $nV_{\bar{R}}$ |
| 処 理 | $r \sum (\bar{C} - \bar{X})^2$             | $n-1$        | $rV_{\bar{C}}$ |
| 誤 差 | $\sum (X - \bar{R} - \bar{C} + \bar{X})^2$ | $(r-1)(n-1)$ | V              |
| 計   | $\sum (X - \bar{X})^2$                     | $rn-1$       |                |



この形は表としても有意性検定についても便利である。反復及び処理にたいしてFは次のように計算できる。

$$F_r = \frac{nV_{\bar{R}}}{V} \quad F_c = \frac{nV_{\bar{C}}}{V}$$

$F_r$  にたいする自由度は  $(r-1)$  及び  $(r-1)(n-1)$  であり、 $F_c$  にたいする自由度は  $(n-1)$  及び  $(r-1)(n-1)$  である。

実際の計算には次の公式を用いると容易である。

$$\text{全体} \quad \sum_1^{rn} (X - \bar{X})^2 = \sum_1^{nr} (X^2) - \frac{G^2}{rn} \quad (17)$$

$$\text{反復} \quad \sum_1^r n(\bar{R} - \bar{X})^2 = \frac{\sum_1^r (R^2)}{n} - \frac{G^2}{rn} \quad (18)$$

$$\text{処理} \quad r \sum_1^n (\bar{C} - \bar{X})^2 = \frac{\sum_1^n (C^2)}{r} - \frac{G^2}{rn} \quad (19)$$

$$\text{誤差} = \text{全体} - \text{反復} - \text{処理}$$

有意性の検定を含む計算例を例5-1に示す。

#### 6. 例5-1. 変量の二重分類.

豚の飼料実験において Dunlop は5-1表のような結果を得た。三つの飼料A, B, Cはビタミン含有量が違っていた。動物は各々3匹の4群であり、一腹の仔及び当初の体重を基として組分けされた。我々の目的には群わけは単に反復の問題と仮定してもよいであろう。

5-1表

飼料A, B, Cで飼育した場合の豚の体重増

| 飼料 | I    | II   | III  | IV   | 計     |
|----|------|------|------|------|-------|
| A  | 7.0  | 16.0 | 10.5 | 13.5 | 47.0  |
| B  | 14.0 | 15.5 | 15.0 | 21.0 | 65.5  |
| C  | 8.5  | 16.5 | 9.5  | 13.5 | 48.0  |
| 計  | 29.5 | 48.0 | 35.0 | 48.0 | 160.5 |

分析の型は

| 変動因 (S. V) | 自由度 (D. F) |
|------------|------------|
| 飼料         | 2          |
| 群          | 3          |
| 誤差         | 6          |
| 計          | 11         |

平方和を計算すると

$$\text{全体} = 2316.75 - \frac{(160.5)^2}{12} = 2316.75 - 2146.69 = 170.06$$

$$\text{飼料} = \frac{(47.0^2 + 65.5^2 + 48.0^2)}{4} - 2146.69 = 54.12$$

$$\text{群} = \frac{(29.5^2 + 48.0^2 + 35.0^2 + 48.0^2)}{3} - 2146.69 = 87.73$$

$$\text{誤差} = 170.06 - 54.12 - 87.73 = 28.21$$

分散分析を行えば次の通りである。

|    | S. S.  | D. F. | M. S. | F.   | 5%点  |
|----|--------|-------|-------|------|------|
| 飼料 | 54.12  | 2     | 27.06 | 5.75 | 5.14 |
| 群  | 87.73  | 3     | 29.24 | 6.22 | 4.76 |
| 誤差 | 28.21  | 6     | 4.702 |      |      |
| 全体 | 170.06 | 11    |       |      |      |

飼料にたいする平均平方は有意である。群にたいする平均平方が有意であるという意味は群を分類した方法如何によつてきまる。我々はこゝで群を単に反復であると仮定した。この場合においては誤差の平均平方は飼料によらない群内の変動によるものである。

したがつてお互に比較できる誤差としてこの平均平方を考えることは妥当である。群の平均平方は実験の計画如何によつて支配され、且誤差の管理がどのように行われているかを示すものである。もし配列が群以外のものであれば簡単に飼料間及び飼料内の分け方をしたであろう。飼料内平均平方は今の配列によるものよりも大きくなる。したがつて実験の正確度は減少したであろう。

#### 7. 実験誤差

前の論議では、変量を  $X_{11} + Y_1 + Z_1$  のような記号で表わした。こゝに  $X_{11}$  は無作為変動のみによる部分であり、分散分析における我々の目的は処理、反復等の効果より生ずる他の平均平方とこの無作為変動から生ずる平均平方とを分離することであつた。この無作為変動から生ずる平均平方を普通、実験誤差の平均平方と呼ぶ。いゝかえればそれは全変動の中でコントロール出来ない部分を表わす。もし前に述べたことに注意深く従うとすれば、実験誤差は他の平均平方の有意性の検定をするための論理的な測定尺度であることは明らかである。なぜならば、



実験の真の効果によつて、検定しようとする平均平方が増大しなければ、その平均平方は実験誤差に等しくなるような傾向がなければならぬからである。

実験資料を分散分析することは非常に容易であるけれども有意性の検定にあたつて正しく誤差を指示することが困難な場合が多い。いゝかえれば何が与えられた検定に対する妥当な誤差を構成するかということは必らずしも明らかではない。

## B. 妥当な誤差の選択

有意とは相対的なもので絶対的な言葉ではない。結果をどのように解釈するかにしたがつて任意に選んだ一つの原因から生ずる変動に関して、その差が有意か有意でないかが分る。これに関係して、実験者は与えられた仮説の検定に基いた有意性の検定の概念を明らかに理解しなければならぬ。検定にたいして選んだ仮説によつて何が妥当な誤差であるかがきまる。この点を明らかにするために、2種類の資料を比較して化学的な決定をする実験を行つたと仮定しよう。各資料から各々20の標本を抽出した。実験室で各標本を分配し、各々二回反復して分析した。そのとき分散分析の型は次のようになる。

|             | D. F. | M. S. の期待値   | 実際の実験の M. S. |
|-------------|-------|--------------|--------------|
| 資料 (A 及び B) | 1     | $\sigma_t^2$ | $s_t^2$      |
| A 標本間変動     | 19    | $\sigma_a^2$ | $s_a^2$      |
| B 標本間変動     | 19    | $\sigma_b^2$ |              |
| 反 復 間       | 40    | $\sigma_o^2$ | $s_o^2$      |
| 計           | 79    |              |              |

この種の大抵の研究の場合と同じように、 $\sigma_a^2$  及び  $\sigma_b^2$  は同じ大きさであると仮定でき、一緒にして抽出による変動を表わす単一分散  $\sigma_g^2$  とする。この例で検定している仮説は資料は同一であるということである。この仮説が真であるとき  $\sigma_t^2$  の期待値は  $\sigma_g^2$  である。いゝかえれば、標本による変動は二つの平均値間の無作為変動による要因のみである。 $\sigma_g^2$  をみると、もしも標本に変動がなければ、期待値は  $\sigma_o^2$  に等しい。したがつて標本及び資料が全変動の原因となり、その原因が資料間の真の差によるものが  $\sigma_p^2$  であり、標本間の真の差によるものが  $\sigma_q^2$  であるような場合の真の実験では

$$\sigma_t^2 = \sigma_p^2 + \sigma_q^2 + \sigma_o^2$$

$$\sigma_g^2 = \sigma_q^2 + \sigma_o^2$$

$$\sigma_o^2 = \sigma_o^2$$

この基礎的な関係は、 $s_g^2$  又は  $s_o^2$  が  $s_t^2$  の有意性の検定をする際の誤差項でなければならぬかどうかの問題を解決するのに応用できる。A 及び B 間の差の存在はおそらくある工業的な過程では重要であろう。そして A は

B よりもかなりよい結果を与えると仮定しよう。さてもしこれら二つの資料のいずれかを用い

るとすれば、得た結果はこの抽出誤差は資料が全く一樣でないという事実を反映するから、抽出誤差をうける。更に研究室での技術の熟練の程度が高ければ  $\sigma_o^2$  は非常に小さくなる。したがつて誤差としての  $\sigma_g^2$  の推定量は資料のいずれかの結果が実際にうける無作為変動の尺度を与えないから誤差をおそれる。勿論、これは実際には  $\sigma_g^2$  が  $\sigma_o^2$  より大となる結果を与える。実験計画により  $\sigma_g^2$  の期待値は  $\sigma_o^2$  に等しいか大きいものであるが例外的な場合には実験で得た推定値は  $s_g^2 > s_o^2$  を与える。そのとき誤差として  $s_g^2$  をとつてよいであろう。

この種の実験に関連して、しばしば問題となる質問は次の通りである。「もしも結果を有意性検定に用いなければ、二回反復までして何故決定するのか？」この答にたいして再び

$$F = \frac{s_t^2}{s_g^2} = \frac{s_p^2 + s_q^2 + s_o^2}{s_g^2}$$

に注意しよう。こゝで  $s_g^2$  を小さくすれば、それだけ資料間の差の検定はますます鋭敏となる。したがつてこの誤差の原因の尺度をもつことは大切であろう。 $s_g^2$  で表わした誤差の部分はそれは資料の変動性を表わすからコントロールできない。しかし  $s_g^2$  は研究室の技術の改善及び検定の鋭敏性をますますによりかなり小さくできることは想像される。一般的な規則として、この型の実験では、 $s_g^2$  は  $s_o^2$  より大きい。これは通常抽出による真の変動性が存在するということと同じである。これをきめるためには、次の有意性検定をする。

$$F = \frac{s_g^2}{s_o^2}$$

勿論もし  $\sigma_o^2$  のある尺度がなければ検定は不可能である。

## 9. 交 互 作 用

分散分析法は処理及び反復のような二通りに分類される資料にたいして特に用いられることを指摘した。実際には、この方法は同時に数通りに分類した資料にも適用できる。例えば、各々2水準での3処理（例えば各々2つの割合での3つの肥料）があり、全実験を4回反復する場合である。資料は各反復にたいして3重分類を与える  $3 \times 2$  の4つの表に作ることができる。しかし、この場合に我々は6つの処理総計を作り得る  $2 \times 3$  表に関心がある。表を次のように仮定しよう。

|     |   | 処 理 |   |   |       |
|-----|---|-----|---|---|-------|
|     |   | A   | B | C |       |
| 水 準 | 1 |     |   |   | $T_1$ |
|     | 2 |     |   |   |       |

自由度の立場から、この表をみると、5つの自由度総計の中、水準にたいしてはわづか1、処



理にたいしては2であることがわかる。残りの2つの自由度は処理による効果を表わさねばならぬが、辺の総計には表われてこない。したがって、処理及び水準間の関係によるものであることがわかる。この効果は通常、交互作用と呼ばれるものである。交互作用の特性を示すためには、 $2 \times 2$ 表を調べると最も便利である。次のような結果をまとめた表があり、3反復、2水準、2処理があると仮定しよう。

|   | 処 理     |         |            |
|---|---------|---------|------------|
|   | A       | B       |            |
| 1 | 4       | 8       | $12 = T_1$ |
| 2 | 8       | 12      | $20 = T_2$ |
|   | 12      | 20      |            |
|   | $= T_A$ | $= T_B$ |            |

(17), (18)及び(19)なる方程式に与えた一般の分析法にしたがえばよいが、平方和にたいする2つの新しい公式を用いると簡単となる。

1.  $T_1$  及び  $T_2$  のような2つの総計にたいする平方和。

$$\frac{T_1^2 + T_2^2}{n} - \frac{(T_1 + T_2)^2}{2n} = \frac{(T_1 - T_2)^2}{2n} \quad (20)$$

こゝに  $n$  は各総計における個々の測定の数である。

2.  $2 \times 2$ 表の交互作用にたいする平方和、表の副総計を次のように表わそう。

|   | A        | B        |
|---|----------|----------|
| 1 | $t_{1A}$ | $t_{1B}$ |
| 2 | $t_{2A}$ | $t_{2B}$ |

交互作用平方和は次式で計算できる。

$$\frac{[(t_{1A} + t_{2B}) - (t_{2A} + t_{1B})]^2}{2n} \quad (21)$$

こゝに分子は斜めの総計間の差の平方にすぎない。上の例では次のようになる。

$$\text{処 理 (S. S)} = \frac{(12 - 20)^2}{12} = 5.33$$

$$\text{水 準 (S. S)} = \frac{(12 - 20)^2}{12} = 5.33$$

$$\text{交互作用 (S. S)} = \frac{(16 - 16)^2}{12} = 0$$

そして副総計の性質を調べると、交互作用平方和はもしも処理の反応が両水準共同一であれば0となることは明らかである。記号で表わすと、

|   | A     | B     |
|---|-------|-------|
| 1 | a     | b     |
| 2 | $a+x$ | $b+x$ |

こゝに  $x$  は水準1と比較したときの水準2による反応の増加量である。 $(a+b+x) - (a+x+b) = 0$  であるから交互作用の平方和=0となる。

二方向の反応を考えると、代数的に次の表で表わすことができよう。

|   | A     | B       |
|---|-------|---------|
| 1 | a     | $a+x$   |
| 2 | $a+y$ | $a+x+y$ |

こゝに  $x$  はAと比較したときのBによる反応を表わし、 $y$  は1と比較したときの2の反応を表わす。したがって

$$a + (a+x+y) - [(a+y) + (a+x)] = 0$$

で交互作用の平方和=0である。

反応が等しくない場合を考えよう。例えば次の結果を得たとしよう。

|   | A | B  |
|---|---|----|
| 1 | 6 | 4  |
| 2 | 8 | 12 |

これではBはAよりも処理の増加量にたいして大なる反応を与える。

こゝにおいて次のものを得る。

$$\text{処 理 (S. S)} = \frac{(14 - 16)^2}{12} = 0.33$$

$$\text{水 準 (S. S)} = \frac{(10 - 20)^2}{12} = 8.33$$

$$\text{交互作用 (S. S)} = \frac{(18 - 12)^2}{12} = 3.00$$

こゝでは交互作用が評価できる。

交互作用が0であるも一つの反応の型を示そう。ここでは比例的な反応はない。例えば関係が

|   | A                 | B                 |
|---|-------------------|-------------------|
| 1 | a                 | b                 |
| 2 | $a + \frac{a}{x}$ | $b + \frac{b}{x}$ |

であるとき、交互作用の平方和  $= \frac{(b-a)^2}{x^2} \cdot \frac{1}{2n}$  で  $a=b$  でなければ0とならないことは明らかである。しかしもし対数に変換したら交互作用は0となることは容易に示すことができる。

## 10. 高次の交互作用

実験において、より多くの要因が導入されると、その結果は高次の交互作用となる。例えば、



2水準、2処理に加えて、各組が2つの違った日に適用されたとする。そのとき、反復と共に4重分類となるであろう。処理総計は次の三つの表で作られる。

|     |   | 日 1 |   | 日 2 |   | 日 1 及び 日 2 |   |
|-----|---|-----|---|-----|---|------------|---|
|     |   | A   | B | A   | B | A          | B |
| 水 準 | 1 |     |   |     |   |            |   |
|     | 2 |     |   |     |   |            |   |

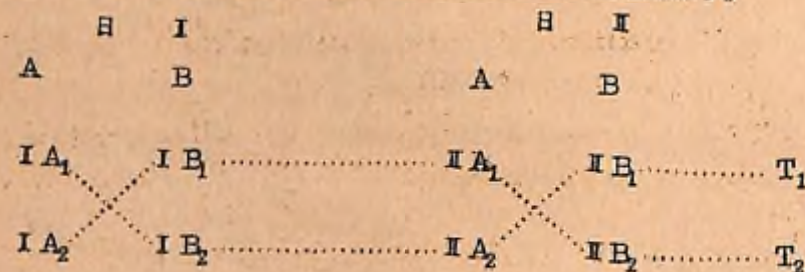
各日にたいして

|          |   | D. F. |                                                                       |
|----------|---|-------|-----------------------------------------------------------------------|
| 処 理      | 1 |       | で、日を考慮にいと、自由度は日にたいして1、日毎の上の三つの効果の交互作用にたいして各々1で次のような自由度が7である完全な配分ができる。 |
| 水 準      | 1 |       |                                                                       |
| 2要因の交互作用 | 1 |       |                                                                       |

|                  | D. F. |
|------------------|-------|
| 日                | 1     |
| 処 理              | 1     |
| 水 準              | 1     |
| 2因子交互作用(処理×水準)   | 1     |
| 2因子交互作用(日×処理)    | 1     |
| 2因子交互作用(日×水準)    | 1     |
| 3因子交互作用(日×処理×水準) | 1     |

3因子交互作用の意味は簡単な交互作用と関係させると最も容易に解釈できる。もしも2因子交互作用(処理×水準)が2つの日にたいして同じであれば、3因子交互作用は0である。これは数字をいれて平方和を計算するとすぐ証明できる。したがって、3因子交互作用は、どの2因子にたいしても一つの因子交互作用が3番目の因子により変化する程度を測るものである。

2×2表にたいする3因子交互作用は簡単な交互作用に用いた幾何的な方法を簡単に拡張して計算できる。下に示す図で、点線は必要な総計を得る方法を示す。



記号で表わすと

$$T_1 = I A_1 + I B_1 + II A_2 + II B_1$$

$$T_2 = I A_2 + I B_1 + II A_1 + II B_2$$

したがって、平方和は

$$\frac{(T_1 - T_2)^2}{2n}$$

である。ここにnは再びT<sub>1</sub>のような各総計が含む個々の変数の数である。

3重分類から3因子交互作用をうることが分つたから、四番目の因子があると4因子交互作用を得、次々に因子が加わると交互作用も増すことは分るであろう。しかし、少なくとも3因子以上の高次の交互作用は実際的にはあまり大切ではない。それらは、まづ第一に統計的にめつたに有意でない。第二に実際の解釈にたいして複雑すぎる。我々が実験計画の研究をすすめてゆくにつれて、主効果及び簡単な交互作用の研究をもつと正確にするためには、実験計画において高次の交互作用を犠牲にすることが望ましいということが分つてくるであろう。

### 11. 平方和の計算法

分散分析の一般原則を知つた後には、平方和の計算法の知識が必要である。次に大抵の実験にたいする平方和を組織的な形で計算する方法をいくつか示そう。

a. n個の変数X<sub>1</sub>, X<sub>2</sub>, X<sub>3</sub>, ..., X<sub>1</sub>, ..., X<sub>n</sub>の組にたいする総計。

$$\sum_{i=1}^n (x_i^2) = \sum_{i=1}^n (X_i^2) - \frac{\left(\sum_{i=1}^n (X_i)\right)^2}{n} \quad (22)$$

$\sum_{i=1}^n (X_i^2)$ なる項はなまの平方和として知られており、 $\left(\sum_{i=1}^n (X_i)\right)^2 / n$ を補正項と呼ぶ。

(略してG. T. で表わす)。 $\sum_{i=1}^n (X_i)$ は単なる全総計であるから、Gで表わすと便利である。

b. 各群がn個の変数からなるk個の群にたいする総計。

$$\sum_{i=1}^k (\bar{X}_i - \bar{X})^2 = \frac{\sum_{i=1}^k (\sum (X_i))^2}{n} - \frac{\left(\sum_{i=1}^k (\sum (X_i))\right)^2}{kn} \quad (23)$$

c. 変数の数が各群同一でないk個の群にたいする総計。

総計の系列及び各々の数を次のように表わすとして。

$$\begin{array}{l} \text{総 計} \quad \sum X_1 \quad \sum X_2 \quad \sum X_3 \quad \dots \quad \sum X_j \quad \dots \quad \sum X_k, \text{ 合計} = G \\ \text{数} \quad n_1 \quad n_2 \quad n_3 \quad \dots \quad n_j \quad \dots \quad n_k, \text{ 合計} = N \end{array}$$

そのとき平方和は



$$\frac{(\sum X_1)^2}{n_1} + \frac{(\sum X_2)^2}{n_2} + \dots + \frac{(\sum X_j)^2}{n_j} + \dots + \frac{(\sum X_k)^2}{n_k} - \frac{G^2}{N} \quad (24)$$

である。これは一般公式であり、各総計の数が等しければ (b) で与えたものと同じになることに注意せよ。

d. (2×n) 重分類表の平方和。

表を次のように仮定する。

|     |    | 処 理                            |                                |                                                                                                      |                                                                           |
|-----|----|--------------------------------|--------------------------------|------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------|
|     |    | 1                              | 2                              | 3..... i ..... n                                                                                     |                                                                           |
| 水 準 | I  | X <sub>1</sub>                 | X <sub>2</sub>                 | X <sub>3</sub> ..... X <sub>i</sub> ..... X <sub>n</sub>                                             | ΣX                                                                        |
|     | II | Y <sub>1</sub>                 | Y <sub>2</sub>                 | Y <sub>3</sub> ..... Y <sub>i</sub> ..... Y <sub>n</sub>                                             | ΣY                                                                        |
|     |    | X <sub>1</sub> +Y <sub>1</sub> | X <sub>2</sub> +Y <sub>2</sub> | X <sub>3</sub> +Y <sub>3</sub> ... X <sub>i</sub> +Y <sub>i</sub> ... X <sub>n</sub> +Y <sub>n</sub> | G=Σ(X <sub>1</sub> +Y <sub>1</sub> )                                      |
|     |    | X <sub>1</sub> -Y <sub>1</sub> | X <sub>2</sub> -Y <sub>2</sub> | X <sub>3</sub> -Y <sub>3</sub> ... X <sub>i</sub> -Y <sub>i</sub> ... X <sub>n</sub> -Y <sub>n</sub> | S=Σ(X <sub>1</sub> -Y <sub>1</sub> )<br>=ΣX <sub>1</sub> -ΣY <sub>1</sub> |

必要な平方和を作ると

|      | DF   | 2水準の代りに、I, IIを反復とすれば、その場合には第三項は誤差となる。      |
|------|------|--------------------------------------------|
| 処 理  | n-1  |                                            |
| 水 準  | 1    | 全平方和は次から得る。                                |
| 交互作用 | n-1  |                                            |
| 合計   | 2n-1 | $\Sigma X^2 + \Sigma Y^2 - \frac{G^2}{2n}$ |

これは単に上の (a) の応用にすぎない。

$$\text{水準にたいして} \quad \frac{(\Sigma X - \Sigma Y)^2}{2n} \quad (25)$$

$$\text{処理にたいして} \quad \frac{\Sigma (X_1 + Y_1)^2}{2} - \frac{G^2}{2n} \quad (26)$$

そのとき、交互作用即ち誤差項にたいしては次のやり方にしたがうことができる。まづ第一に、総計から処理及び水準を差引いて、差として誤差を得ることができる。第二に次式から直接に三番目の項が計算できる。

$$\frac{\Sigma (X_1 - Y_1)^2}{2} - \frac{(\Sigma X - \Sigma Y)^2}{2n} \quad (27)$$

こゝに右の項は (26) と同じである。

これらの公式は各枠内に単一変量しかないということに基いていることに注意しなければならぬ。もし上の実験をr反復行つたとすれば、rを各項の分母にいれねばならない。

即ち、最後の公式は次の通りになる。

$$\frac{\Sigma (X_1 - Y_1)^2}{2r} - \frac{(\Sigma X - \Sigma Y)^2}{2rn} \quad (28)$$

また、上に与えた全平方和は表の (2n-1) なる自由度にたいする総計である。全平方和

$$\text{は、} \quad \Sigma (X^2) + \Sigma (Y^2) - \frac{G^2}{2rn} \quad (29)$$

となろう。ここにX及びYは個々の変量、副総計ではない。

e. (m×n) 重分類表の平方和

表にはn列、m行あると仮定しよう。

|   |          | 列        |          |       |          |          |
|---|----------|----------|----------|-------|----------|----------|
|   |          | 1        | 2        | ..... | n        |          |
| 行 | 1        | $X_{11}$ | $X_{12}$ | ..... | $X_{1n}$ | $T_{R1}$ |
|   | 2        | $X_{21}$ | $X_{22}$ | ..... | $X_{2n}$ | $T_{R2}$ |
|   | $\vdots$ | $\vdots$ | $\vdots$ | ..... | $\vdots$ | $\vdots$ |
|   | m        | $X_{m1}$ | $X_{m2}$ | ..... | $X_{mn}$ | $T_{Rm}$ |
|   |          | $T_{C1}$ | $T_{C2}$ | ..... | $T_{Cn}$ | G        |

平方和及び自由度の概略を作ると

|            | SS. | DF.        |
|------------|-----|------------|
| 列          |     | n-1        |
| 行          |     | m-1        |
| 交互作用 (C×R) |     | (n-1)(m-1) |
| 全 体        |     | nm-1       |

そのとき

$$\text{全平方和} = \Sigma X^2 - \frac{G^2}{mn}$$

$$\text{列} = \frac{\Sigma (T_c^2)}{m} - \frac{G^2}{mn}$$

$$\text{行} = \frac{\Sigma (T_r^2)}{n} - \frac{G^2}{mn}$$

$$\text{交互作用} = \Sigma X^2 - \frac{\Sigma T_c^2}{m} - \frac{\Sigma T_r^2}{n} + \frac{G^2}{mn}$$



最後のものは、全体から行及び列の平方和を差し引いたものである。実験が $r$ 回反復であれば、上のすべての表示の分母に $r$ をつける。また上に与えた全平方和は自由度 $m n - 1$ を表わす表の副総計にたいする総計である。全平方和は

$$\sum X^2 - \frac{G^2}{r m n} \quad \text{であり、} X \text{は個々の変数で、副総計ではない。}$$

#### f. 3因子交互作用

1水準以上で3因子A, B, Cがあると仮定すると、3因子交互作用を計算する一般的な方法は三番目の各水準にたいして、各2因子にたいする2因子交互作用をうることである。組合せた三番目の因子のすべての水準にたいする2因子交互作用は、これらの和から差し引くと3因子交互作用をうる即ち、いつも三つの表を作る。例えばA, B, Cが各々2水準であれば、次をうる。

|       |       |       |       |       |       |       |             |       |
|-------|-------|-------|-------|-------|-------|-------|-------------|-------|
|       | $C_1$ |       |       | $C_2$ |       |       | $C_1 + C_2$ |       |
|       | $A_1$ | $A_2$ |       | $A_1$ | $A_2$ |       | $A_1$       | $A_2$ |
| $B_1$ |       |       | $B_1$ |       |       | $B_1$ |             |       |
| $B_2$ |       |       | $B_2$ |       |       | $B_2$ |             |       |

$A \times B \times C = C_1$  にたいする  $(A \times B) + C_2$  にたいする  $(A \times B) - (C_1 + C_2)$  にたいする  $(A \times B)$

ある場合には簡単となる。例えば、 $2 \times 2 \times 2$ の実験では、3因子交互作用は下図に示すように2つの総計をうることににより直接決定できる。

|       |       |       |       |       |       |       |
|-------|-------|-------|-------|-------|-------|-------|
|       | $C_1$ |       |       | $C_2$ |       |       |
|       | $A_1$ | $A_2$ |       | $A_1$ | $A_2$ |       |
| $B_1$ |       |       | $B_1$ |       |       | $T_1$ |
| $B_2$ |       |       | $B_2$ |       |       | $T_2$ |

この方法は66頁に与えたが、一つの場合にすべての計算法に便利であるから再びくりかえして説明する。総計を $T_1$ 及び $T_2$ とすれば、平方和は

$$\frac{(T_1 - T_2)^2}{2n} \quad (31)$$

である、こゝに $n$ は各総計に含まれる変数の数である。

また、 $3 \times 3 \times 3$ の実験にたいする3因子交互作用の計算にも、この簡易化を用い得る。因子をN, P及びKとして各々3水準と仮定すれば次のような三つの表をうる。

|   |  |       |     |     |   |       |     |     |   |       |     |     |
|---|--|-------|-----|-----|---|-------|-----|-----|---|-------|-----|-----|
|   |  | $K_1$ |     |     |   | $K_2$ |     |     |   | $K_3$ |     |     |
|   |  | 1     | 2   | 3   |   | 1     | 2   | 3   |   | 1     | 2   | 3   |
| 1 |  | 111   | 121 | 131 | 1 | 112   | 122 | 132 | 1 | 113   | 123 | 133 |
| 2 |  | 211   | 221 | 231 | 2 | 212   | 222 | 232 | 2 | 213   | 223 | 233 |
| 3 |  | 311   | 321 | 331 | 3 | 312   | 322 | 332 | 3 | 313   | 323 | 333 |

各表を用いると、Kの1水準にたいして交互作用N×Pを計算できる。

まづ第一に $3 \times 3$ 表は下に示す方形において斜めの総計の2組に分けることができるということに注意せよ。

3×3表

|     |     |     |
|-----|-----|-----|
| 1 1 | 1 2 | 1 3 |
| 2 1 | 2 2 | 2 3 |
| 3 1 | 3 2 | 3 3 |

I斜線

$$\begin{aligned} 11 + 22 + 33 &= I_1 \\ 21 + 32 + 13 &= I_2 \\ 31 + 12 + 23 &= I_3 \end{aligned}$$

J斜線

$$\begin{aligned} 11 + 32 + 23 &= J_1 \\ 21 + 12 + 33 &= J_2 \\ 31 + 22 + 13 &= J_3 \end{aligned}$$

$3 \times 3$ 表における交互作用にたいする4つの自由度は、I総計にたいして2、J総計にたいして2に分けることができる。平方和は次のようになる。

$$\left. \begin{aligned} \text{I (総計)} & \frac{I_1^2 + I_2^2 + I_3^2}{3n} - \frac{G^2}{9n} \\ \text{J (総計)} & \frac{J_1^2 + J_2^2 + J_3^2}{3n} - \frac{G^2}{9n} \end{aligned} \right\} \quad (32)$$

こゝに、 $G$ は全総計であり、方形の各副総計は $n$ 個の測定からなっている。

三つの資料の表から、各々にたいしてI総計の1組をまづ書くことができる。例えば  
 $111 + 221 + 331 = A_1$     $112 + 222 + 332 = B_1$     $113 + 223 + 333 = C_1$   
 $211 + 321 + 131 = A_2$     $212 + 322 + 132 = B_2$     $213 + 323 + 133 = C_2$   
 $311 + 121 + 231 = A_3$     $312 + 122 + 232 = B_3$     $313 + 123 + 233 = C_3$   
I総計からも一度I組及びJ組をかくことができる。即ち



$$\begin{array}{ll} \text{I 組} & \text{J 組} \\ A_1 + B_2 + C_3 = I_1 & A_1 + B_3 + C_2 = J_1 \\ A_2 + B_3 + C_1 = I_2 & A_2 + B_1 + C_3 = J_2 \\ A_3 + B_1 + C_2 = I_3 & A_3 + B_2 + C_1 = J_3 \end{array}$$

更に始めから出発して、まず各級からJ総計をかくことができる。

$$\begin{aligned} 111 + 231 + 321 &= A'_1 \quad 112 + 232 + 322 = B'_1 \quad 113 + 233 + 323 = C'_1 \\ 211 + 121 + 331 &= A'_2 \quad 212 + 122 + 332 = B'_2 \quad 213 + 123 + 333 = C'_2 \\ 311 + 221 + 131 &= A'_3 \quad 312 + 222 + 132 = B'_3 \quad 313 + 223 + 133 = C'_3 \end{aligned}$$

そして、これらからまたI'及びJ'組が作れる。

$$\begin{array}{ll} \text{I' 組} & \text{J' 組} \\ A'_1 + B'_2 + C'_3 = I'_1 & A'_1 + B'_3 + C'_2 = J'_1 \\ A'_2 + B'_3 + C'_1 = I'_2 & A'_2 + B'_1 + C'_3 = J'_2 \\ A'_3 + B'_1 + C'_2 = I'_3 & A'_3 + B'_2 + C'_1 = J'_3 \end{array}$$

3総計の上の組の各々は、3因子交互作用にたいする2つの自由度に於ける平方和を与える。

そしてこれらを4組にたいして合計すると計算が完了する。

この原則は再び実験計画の論議において用いるであろう。こゝでは、総計の4組から生ずる平方和は直交であるということに注意しておくだけで十分である。したがって、3因子交互作用  $N \times P \times K$  の自由度8は相互に直交である4対に分割される。

#### 15. 個々の自由度に応じて偏差平方和を成分に分割する方法。

分散分析法の結果の一つは、実験者は処理間に有意差があるとそれで結論は十分であると假定しがちであるということである。実際には、これは処理の差及び著しい効果について更に研究をすすめるための予備的なものにすぎない。処理の平方和を個々の成分に分割するのに用いる方法は、各々が一つの自由度を表わす直交成分に平方和を分割する Fisher, Yates 及び Cochran & Cox が述べた方法である。

成分は処理総計の線型函数といわれるものから生ずる。例えば、5つの処理総計  $T_1, T_2, T_3, T_4$  及び  $T_5$  では、一つの線型函数は

$$Z = T_1 + T_2 - T_3 - T_4 + (0 \times T_5) \quad (33)$$

であろう。

$$\text{一般に } Z = l_1 T_1 + l_2 T_2 + \dots + l_k T_k \quad (34)$$

こゝにk個の総計があり、このような線型函数で我々の目的に必要な条件は

$$l_1 + l_2 + \dots + l_k = 0 \quad (35)$$

であるということである。例えば、方程式(33)では

$$1 + 1 - 1 - 1 + 0 = 0 \quad \text{である。}$$

$$\text{このような線型函数にたいして、それに対応する平方和は } Z^2/D \text{ である。ここに} \quad (36)$$

$$D = r(l_1^2 + l_2^2 + \dots + l_k^2)$$

で、rは反復の数である。

自由度1を表わすこのような線型函数のどれか一つを書いた後には、他の同様な函数が残り

の自由度にたいして存在しなければならぬことは明らかである。これらの函数は次に基いた規則によつてかくことができる。

$$\left. \begin{aligned} Z_1 &= l_{11} T_1 + l_{12} T_2 + \dots + l_{1k} T_k \\ Z_2 &= l_{21} T_1 + l_{22} T_2 + \dots + l_{2k} T_k \end{aligned} \right\} \quad (37)$$

のような2つの線型函数は

$$l_{11} l_{21} + l_{12} l_{22} + \dots + l_{1k} l_{2k} = 0 \quad (38)$$

であれば直交であるといわれる。

上の方程式(33)にこれを適用すると

$$Z_1 = T_1 + T_2 - T_3 - T_4 + (0 \times T_5)$$

及び

$$Z_2 = T_1 - T_2 + (0 \times T_3) + (0 \times T_4) + (0 \times T_5)$$

は係数の和=0で  $Z_1$  及び  $Z_2$  にたいする係数の積和は

$$(1 \times 1) + (1 \times -1) = 0$$

であるような方程式である。したがって二つの対比は直交しているということが証明された。

方程式(33)にもどると、次の4つの線型函数は直交であることが証明できる。積和を作る際に便利のように係数を各函数の横で書いておく。

$$(1) \quad \frac{1}{2} T_1 + \frac{1}{2} T_2 - T_3 \quad \frac{1}{2} + \frac{1}{2} - 1 + 0 + 0 = 0$$

$$(2) \quad T_4 - T_5 \quad 0 + 0 + 0 + 1 - 1 = 0$$

$$(3) \quad 2 T_1 + 2 T_2 + 2 T_3 - 3 T_4 - 3 T_5 \quad 2 + 2 + 2 - 3 - 3 = 0$$

$$(4) \quad T_1 - T_2 \quad 1 - 1 = 0$$

函数のすべての対の積和は0となることは分るであろう。したがって函数は直交であると假定できる。

対応する平方和は次の通りである。こゝに分母は方程式(36)で与えられ、4反復と假定している。

$$(1) \quad \frac{Z_1^2}{D_1} = \frac{\left[ \frac{1}{2}(T_1 + T_2) - T_3 \right]^2}{6}$$



$$(2) \frac{Z_2^2}{D_2} = \frac{(T_1 - T_5)^2}{8}$$

$$(3) \frac{Z_3^2}{D_3} = \frac{[2(T_1 + T_2 + T_3) - 3(T_4 + T_5)]^2}{120}$$

$$(4) \frac{Z_4^2}{D_4} = \frac{(T_1 - T_2)^2}{8}$$

実際の数字を用いて

$$T_1 = 1652 \quad T_2 = 1579 \quad T_3 = 347 \quad T_4 = 317 \quad T_5 = 151$$

としよう。そのとき

$$(1) = \frac{\left[\frac{1}{2}(1652 + 1579) - 347\right]^2}{6} = 268,182.0$$

$$(2) = \frac{(317 - 151)^2}{8} = 3,444.5$$

$$(3) = \frac{[2(1652 + 1579 + 347) - 3(317 + 151)]^2}{120} = 275,712.5$$

$$(4) = \frac{(1652 - 1579)^2}{8} = 666.1$$

$$\text{合計} = 548,005.1$$

通常の方法で処理にたいする平方和を計算すると

$$\frac{1652^2 + 1579^2 + 347^2 + 317^2 + 151^2}{4} - \frac{4046^2}{20} = 548,005.2$$

で、このような数学的分割法が正しいことを示す。

こゝに論じた比較は個々の自由度を表わすから、もし平均値に基いた比較の標準誤差を得ることができれば、もし検定を用いて有意性の検定ができる。標準誤差を計算する一般的な規則は

$$s_z = s \sqrt{\frac{1^2}{r_1} + \frac{1^2}{r_2} + \dots + \frac{1^2}{r_k}} \quad (39)$$

こゝに  $s$  は分散分析の誤差項の平均平方の平方根である。例えば検定を対比 (3) に対して適用してみよう。平均値を用いると

$$2(413.0 + 394.8 + 86.8) - 3(79.2 + 37.8) = 1438.2$$

を得る。

$$s = 54.7$$

としよう。そのとき

$$s_z = 54.7 \sqrt{\frac{2^2 + 2^2 + 2^2 + 3^2 + 3^2}{4}} = 150$$

そして

$$t = \frac{1438.2}{150} = 9.59$$

処理の各水準にたいして一つの対照がある場合の要因実験と同じようにすべての処理にたいして  $t$  は必ずしも同じでないことを注意しなければならぬ。

### 13. 例 5-2. 妥当な誤差の選択

4つの場処で生育している5品種の小麦の系列にたいして、それから作った粉のパン焼き試験を行った。各品種の標本を各場処からとり、粉にひいた。各標本から二つのパンを焼いた。したがって、二つのパンの容積間の差が測定誤差を与える。

分散分析を行う形をきめるには、まず第一に、場処の平均平方は自由度が3であり、品種の平均平方は自由度が5であることに注意しよう。また、違つた場処での品種の特異な反応と考えられる交互作用の影響もある筈である。交互作用にたいしての自由度を見出す規則は交互作用をする因子の自由度を乗することである。したがって交互作用の平均平方(MS)は自由度  $3 \times 4 = 12$  である。

5 - 2 表

4つの場処で生育した5品種の小麦から作った2対のパンの容積(ccでのパンの容積-500)  $\times 10$

|     | 場 処   |       |       |       |      |      |      |      | 計     |
|-----|-------|-------|-------|-------|------|------|------|------|-------|
|     | I     | II    | III   | IV    |      |      |      |      |       |
| 1   | 7.5   | 4.5   | 15.5  | 14.0  | 16.5 | 14.5 | 19.0 | 18.6 | 110.1 |
| 品 2 | 12.6  | 13.2  | 20.0  | 18.5  | 15.0 | 14.0 | 23.8 | 24.4 | 141.4 |
| 3   | 7.0   | 1.0   | 10.0  | 8.0   | 15.5 | 14.0 | 17.8 | 18.5 | 91.8  |
| 4   | 1.5   | 2.0   | 13.0  | 15.0  | 8.5  | 9.0  | 14.8 | 16.6 | 80.4  |
| 種 5 | 28.0  | 29.0  | 19.5  | 16.0  | 10.5 | 12.0 | 22.0 | 24.8 | 161.8 |
| 計   | 106.2 | 149.5 | 129.5 | 200.3 |      |      |      |      | 585.5 |

測定は全部で40であるから、自由度は全体で39である。残り20の自由度は二つの測定の誤差を表わす。我々は20対のパンがあるから、これをチェックした、そして各対は自由度1であるから全部で20でなければならぬ。分散分析の最終の表は



| M. S.       | D. F. |                                  |
|-------------|-------|----------------------------------|
| 場 処         | 3     | 平方和を得るには、他の表が必要である。              |
| 品 種         | 4     | 下に示すように、この表は $(X_1 - X_2)$       |
| 交互作用(品種×場処) | 12    | 及び $(X_1 + X_2)$ の値を示す。こゝに $X_2$ |
| 誤 差         | 20    | 及び $X_2$ は対の値を表わす。               |
| 全 体         | 39    |                                  |

|             | I    | II   | III  | IV   |
|-------------|------|------|------|------|
| $X_1 - X_2$ |      |      |      |      |
| 1           | 3.0  | 1.5  | 2.0  | 0.4  |
| 2           | -0.7 | 1.5  | 1.0  | -0.6 |
| 3           | 6.0  | 2.0  | 1.5  | -0.7 |
| 4           | -0.5 | -2.0 | -0.5 | -1.8 |
| 5           | -1.0 | 3.5  | -1.5 | -2.8 |

|             | I     | II    | III   | IV    | 計     |
|-------------|-------|-------|-------|-------|-------|
| $X_1 + X_2$ |       |       |       |       |       |
| 1           | 12.0  | 29.5  | 31.0  | 37.6  | 110.1 |
| 2           | 25.7  | 38.5  | 29.0  | 48.2  | 141.4 |
| 3           | 8.0   | 18.0  | 29.5  | 36.3  | 91.8  |
| 4           | 3.5   | 28.0  | 17.5  | 31.4  | 80.4  |
| 5           | 57.0  | 35.5  | 22.5  | 46.8  | 161.8 |
| 計           | 106.2 | 149.5 | 129.5 | 200.3 | 585.5 |

これらの資料を概略分析の形にするには、まず論理的に次のように分割する。

D. F.

|     |    |                                  |
|-----|----|----------------------------------|
| 対 間 | 19 | この最初のもは次のように $(X_1 + X_2)$ の表を用い |
| 対 内 | 20 | て計算し第二のもは $(X_1 - X_2)$ の表を用いて計算 |
|     |    | する。                              |

$$\text{対 間} = \frac{1}{2} \sum (X_1 + X_2)^2 - \frac{[\sum (X)]^2}{N}$$

こゝに、Nは観測の総数40である。

$$\text{対 内} = \frac{1}{2} \sum (X_1 - X_2)^2$$

これは分析の誤差項である。

さて、対間の平方和及び自由度は次のように分割されるべきである。

|             | D. F. |
|-------------|-------|
| 場 処         | 3     |
| 品 種         | 4     |
| 交互作用(品種×場処) | 12    |

計算は次の通りである。

$$\text{計} = 10,329.73 - \frac{585.5^2}{40} = 1759.47$$

$$\text{誤 差} = \frac{1}{2} (93.33) = 46.67$$

$$\text{対 間} = \frac{20,566.13}{2} - \frac{585.5^2}{40} = 10,283.06 - 8,570.26 = 1,712.80$$

$$\text{場 処} = \frac{90,519.03}{10} - 8,570.26 = 481.64$$

$$\text{品 種} = \frac{73,186.61}{8} - 8,570.26 = 578.07$$

$$\text{交互作用} = 1,712.80 - 481.64 - 578.07 = 653.09$$

分散分析は次のとおりである。

|             | S. S.    | D. F. | M. S. |
|-------------|----------|-------|-------|
| 場 処         | 481.64   | 3     | 160.5 |
| 品 種         | 578.07   | 4     | 144.5 |
| 交互作用(品種×場処) | 653.09   | 12    | 54.42 |
| 誤 差         | 46.67    | 20    | 2.334 |
| 全 体         | 1,759.47 |       |       |

4つの成分の平方和を加えたものが全体の平方和になることをチェックとして用いることに注意せよ。

さて場処及び品種間の差の有意性検定に二つのパンの容積からの平均平方を用いるか、又は交互作用の平均平方を用いるかをきめねばならない。もし、抽出した場処全域で、どの品種が



最大のパン容積を与えるかをきめることが実験の目的とすれば、この場合には場処は単なる実験の反復にすぎないから、交互作用の平均平方と品種の平均平方とを比較することが必要である。二つのパン容積からの誤差は単に研究室での技術の正確度の指示を与えるにすぎない。もし、それが大であれば、交互作用の分散の値を増大させるから差の有意性が減少する。

品種の平均平方を比較すると、Fの値は2.66である。そして5%点では3.26であるから、抽出された全域を考えると、パン容積の差は有意でないと結論される。いゝかえれば、場処毎の品種の平均パン容積での変動は非常に大であるから全地域に対する平均値間の差はこの変動により容易に説明されるということである。

交互作用平均平方は二つのパン容積間の差の平均平方よりもかなり大である。これは、実験誤差は、この実験では結果の精度に影響するほど大きくないという意味である。

品種の検定を各場処での反復プロットで行うから、もしパン容積の測定を各プロットで行えば、他の誤差の尺度を得ることができるであろう。この誤差は土壌の不斉一性による変動を測るものであろう。そしてもし、全域に対しての品種平均平方が土壌の不斉一性による一括誤差と比較して有意であれば、一般に各場処で、品種の平均値間の差はこのような抽出変動により説明できるものより大であるということを示すであろう。しかし、これは交互作用が誤差であるとして行つた検定に基いた我々の結論を変えることはないであろう。

#### 14. 例 5-3. 変量の三重分類

実験的なパン焼きにおいて、パンをこねる機械を検定するのに、Geddes 等はA, B, Cと名付ける三つの補正を行つた機械を用い、5つの粉の系列を2つの方式にしたがつて焼いた実験をした。パンの容積の資料を5-3表に示す。

5 - 3 表

パンをこねる機械の検定でのパンの容積結果  
(CC でのパン容積 - 500) / 10

| 方 式      | 機械装置 | 粉    |      |      |      |       | 計     |
|----------|------|------|------|------|------|-------|-------|
|          |      | 1    | 2    | 3    | 4    | 5     |       |
| 簡 単 (S)  | A    | 9.4  | 2.6  | 12.3 | 4.6  | 13.5  | 42.4  |
|          | B    | 9.6  | 3.1  | 13.0 | 4.3  | 13.8  | 43.8  |
|          | C    | 9.6  | 2.7  | 12.4 | 1.8  | 13.0  | 39.5  |
|          | 小 計  | 28.6 | 8.4  | 37.7 | 10.7 | 40.3  | 125.7 |
| 臭素塩酸 (B) | A    | 13.7 | 21.6 | 19.4 | 13.5 | 24.5  | 92.7  |
|          | B    | 12.7 | 22.6 | 20.6 | 10.4 | 24.3  | 90.6  |
|          | C    | 12.6 | 21.8 | 20.9 | 6.8  | 23.2  | 85.3  |
|          | 小 計  | 39.0 | 66.0 | 60.9 | 30.7 | 72.0  | 268.6 |
|          | 総 計  | 67.6 | 74.4 | 98.6 | 41.4 | 112.3 | 394.3 |

分析の形を作つて、前に作つたものと比較すると、複雑さが増していることが分る。5-3表の6列は2分類を表わすが、この場合には簡単な二重分類を与える6級と考える。そのとき分析の形は

|            | D. F. |
|------------|-------|
| 粉          | 4     |
| 級          | 5     |
| 交互作用 (粉×級) | 20    |
| 全 体        | 29    |

しかし、級の自由度5は次のように分けられねばならない。

|               | D. F. |                                      |
|---------------|-------|--------------------------------------|
| 機械装置 (ABC)    | 2     | したがつて、最初の分析における交互作用は粉と上の3因子の交互作用である。 |
| 方 式 (SB)      | 1     | これを明らかにすると、次のような完全な分析の形にかくことができる。    |
| 交互作用 (ABC×SB) | 2     |                                      |

|                   | D. F. |
|-------------------|-------|
| 粉 (1……5)          | 4     |
| 機械装置 (ABC)        | 2     |
| 方 式 (SB)          | 1     |
| 交互作用 (ABC×SB)     | 2     |
| 交互作用 (1…5×ABC)    | 8     |
| 交互作用 (1…5×SB)     | 4     |
| 交互作用 (1…5×ABC×SB) | 8     |
| 全 体               | 29    |

最終の交互作用は3因子交互作用と呼ぶものである。この例では、それは(ABC×SB)となる交互作用が粉が違ふとどの位違ふかの程度を表わす。もし交互作用(ABC×SB)が各粉共同じであれば、3要因交互作用は0である。

上に定めた成分にたいして平方和を決定するには、下に示すような三つの計算表を作る必要がある。

| 機械装置 | 粉    |      |      |      |       | 計     |
|------|------|------|------|------|-------|-------|
|      | 1    | 2    | 3    | 4    | 5     |       |
| A    | 23.1 | 24.2 | 31.7 | 18.1 | 38.0  | 135.1 |
| B    | 22.3 | 25.7 | 33.6 | 14.7 | 38.1  | 134.4 |
| C    | 22.2 | 24.5 | 33.3 | 8.6  | 36.2  | 124.8 |
| 計    | 67.6 | 74.4 | 98.6 | 41.4 | 112.3 | 394.3 |



| 方 式 | 粉     |       |       |       |       | 計      |
|-----|-------|-------|-------|-------|-------|--------|
|     | 1     | 2     | 3     | 4     | 5     |        |
| S   | 28.6  | 8.4   | 37.7  | 10.7  | 40.3  | 125.7  |
| B   | 39.0  | 66.0  | 60.9  | 30.7  | 72.0  | 268.6  |
| S+B | 67.6  | 74.4  | 98.6  | 41.4  | 112.3 | 394.3  |
| S-B | -10.4 | -57.6 | -23.2 | -20.0 | -31.7 | -142.9 |

| 方 式 | 機 械 装 置 |       |       | 計      |
|-----|---------|-------|-------|--------|
|     | A       | B     | C     |        |
| S   | 42.4    | 43.8  | 39.5  | 125.7  |
| B   | 92.7    | 90.6  | 85.3  | 268.6  |
| S+B | 135.1   | 134.4 | 124.8 | 394.3  |
| S-B | -50.3   | -46.8 | -45.8 | -142.9 |

計算は次の通りである。

$$\begin{aligned}
 \text{全 体} &= 6618.43 - \frac{394.3^2}{30} = 6618.43 - 5182.42 = 1436.01 \\
 \text{粉}(1\cdots 5) &= (34,152.33/6) - 5182.42 = 509.64 \\
 \text{装置}(ABC) &= (51,890.41/10) - 5182.42 = 6.62 \\
 \text{方式}(SB) &= (268.6 - 125.7)^2/30 = 680.68 \\
 \text{交互作用}(ABC \times SB) &= [\sum (S-B)^2/10] - 680.68 = (6817.97/10) - 680.68 = 1.12 \\
 \text{交互作用}(1\cdots 5 \times ABC) &= \\
 \text{表の全計} &= (11,436.57/2) - 5182.42 = 535.86 \\
 \text{粉}(1\cdots 5) &= 509.64 \\
 \text{装置}(ABC) &= 6.62 \\
 \text{残り}(1\cdots 5 \times ABC) &= 19.60 \\
 \text{交互作用}(1\cdots 5 \times SB) &= [\sum (S-B)^2/6] - 680.68 = (5369.05/6) - 680.68 = 214.12 \\
 \text{交互作用}(1\cdots 5 \times ABC \times SB) &= \text{残り}
 \end{aligned}$$

分散分析を詳細に行えば次の通りである。

|                        | SS.     | DF. | MS.   | F      | 5% 臨 |
|------------------------|---------|-----|-------|--------|------|
| 粉(1...5)               | 509.64  | 4   | 127.4 | 243.1  | 3.84 |
| 方式(SB)                 | 680.68  | 1   | 680.7 | 1299.0 | 5.32 |
| 交互作用(1...5 × SB)       | 214.16  | 4   | 53.54 | 102.2  | 3.84 |
| 装置(ABC)                | 6.62    | 2   | 3.31  | 6.31   | 4.46 |
| 交互作用(ABC × SB)         | 1.12    | 2   | 0.560 | 1.07   | 4.46 |
| 交互作用(1...5 × ABC)      | 19.60   | 8   | 2.450 | 4.68   | 3.44 |
| 交互作用(1...5 × ABC × SB) | 4.19    | 8   | 0.524 |        |      |
| 全 体                    | 1436.01 | 29  |       |        |      |

妥当な誤差の見地から、例5-3を詳細に研究することは興味がある。まず第一に、測定は2回繰返して行われていないから、技術における誤差の真の尺度はないということに注意されたい。第二に、実験の第一の目的は色々な機械装置によるパン容積の差、及び同じ装置による色々な反応を研究することである。この理由で分散分析は2部分に分けられた。最初の群の三つの効果は、以前の経験でこのような結果が得られるであろうということを農芸化学者が予測できたから、特に興味はない。三つの効果の一つの群に分離することは、実験で得た資料の結果でなくて、前に予想していることであつた。そして実験を行う前にこうなるであろうということを決めていた。

装置による分散を考えると、質問するべき第一の問題は純実験誤差の分散力又は2因子交互作用の分散に対して検定するべきか否かということである。それは(ABC × SB)という交互作用に対して検定されるべきではない。何故なら実際にパン焼きをする場合には単純方式又は臭素塩酸方式の何れかが用いられるのであつて両者が同時に用いられることはないからである。実際に、で答えなければならぬ二つの問題があり、共に適切な解答がある。もし装置の効果が違つた粉を使つた場合の結果を関連して考えるならば、正確な誤差(粉 × 装置)の交互作用であろう。その結果は装置の効果に対して有意性を示さない。もし装置の効果を同じ粉からのパンの系列で判断しなければならぬとすれば、正しい誤差は、それがほとんど実験誤差である3因子交互作用であろう。

Fの値及び5%点は分析に与えてあり、これを用いると結果を非常に迅速にまとめることができる。粉及び方式間の交互作用と同様、粉及び方式の差は実験誤差と比較すると非常に大であり、論述を誤ることはない位である。実験において、第一に関心を抱くことは、機械の装置に関するもので、他の因子と装置の交互作用である。装置は実験誤差に比して有意である。そして全体を一べつすると、これはC装置がA又はBよりも小さいパン容積しか与えていないという事実に帰因している。方式S及びBとABCの交互作用は有意ではない。それは、装置間の差がパン焼きの二つの方法に対して全く一致しているということを示す。装置と粉の交互作用は有意である。そして我々は粉での結論として、機械装置によりある程度変化するということが分る。結果をよく調べてみると、これはB及びCの装置が他のものよりもこの粉に対してパン容積を大いに少なくするから、粉4に帰因するように思えるであろう。

## 15. 分散分析の限界

Gochran は「分散分析は処理及び環境の影響が加法的であり、実験誤差が確率の意味で独立であり、等分散で正規に分布しているという仮定による」と述べた。第15章2節において、個々の結果を、Xを無作為変動による部分、Yを反復による部分、Zを処理による部分とすると、記号  $X+Y+Z$  で表わしたことを思い出すであろう。F検定ができるためにはXは独立に、正規に、平均値0、分散 $\sigma^2$ をもつて分布するということが必要である。実験誤



差間の相関はある条件の下で存在するかも知れない。実際に、野外実験では、隣接のプロットの誤差は同じように変化する傾向があるので明らかに相関がある。この問題を後章で示す無作為化の道程によつて、野外実験ではなくすることが可能である。分散の異なる場合がしばしばある。第11章においては、この場合の一つの実験を述べる。そして分散分析を直接に適用する場合よりも正確な有意性検定を行う方法を例示する。しばしば、資料が正規でない場合がある。それで他の尺度に資料を変換して正規に近くする方法を下に述べよう。

## 16. 変換

変換を必要とする資料の普通の型に関して以下述べよう。

### a. 小さい整数

我々が昆虫が集まつている数が非常に少ない土壌の平方フット当り昆虫数を数えると仮定しよう。個々のフットの数を数えると0があり、最大は10位であろう。最もよく用いる変換は $\sqrt{X}$ でXは数えあげた数である。0の場合には変換変量として $\sqrt{\frac{1}{2} + X}$ をとることをすすめる。

### b. 分数又は百分率

百分率は特に正規分布からかけはなれているようである。事実、第3章から、期待分布は二項分布をすることが明らかとなつている。p≠q場合は分布は明らかに歪む。p=qか又はほとんどpがqに等しい場合には分布は対称となり、正規分布として取扱つて十分である。ほとんどの資料が30~70%の間になれば、百分率を変換するということが規則と考えてよい。使用する変換は、その数のsinが変換すべき百分率の平方根である角度にすることである。

(訳者註 角度 $=\sin^{-1}\sqrt{\text{百分率}}$ という変換をすることでブリスの変換といわれる)。これは Fisher & Yates, Bliss & Snedecor が作つた数で逆正弦変換と呼ばれている。実験の系列を単独に分析し、一つの分析にもちこむときには、範囲を考えないで、変換を全域で行わねばならぬことは明らかである。

### c. 広範囲での整数の数えあげ

再び土壌昆虫の数えあげを考えると、もし集まりが密であるとすれば、平方フット中の数は非常に多くなる。昆虫の検査に対しての土壌処理を、有効な処理の場合には5~20の範囲が数えられ、有効な処理でない場合には100~10,000の範囲が数えられると想像してみよう。このような場合、平均値及び標準偏差の間には通常比例的な関係が存在する。用いる変換は数の対数をとることである。更に0である場合は、対数をとる前に各数に1を加えることをすすめる。

一般に、資料を検定する際には、歪分布かどうか、分散が等しくないかどうか、他の非正規型分布かどうかをみきわめて、分析を行う前に、正しい補正又は変換を行うことが必要である。

この問題について、更に詳細なことは、1947年3月の Biometrics に述べてある。これには Churchill Eisenhardt, W. G. Cochran 及び M. S. Bartlett による論文が、分散分析の必要性、資料の非正規性の重要性及び使用する変換について説明を行つている。



## 第6章 直線回帰分析

### 1. 一般的な認識

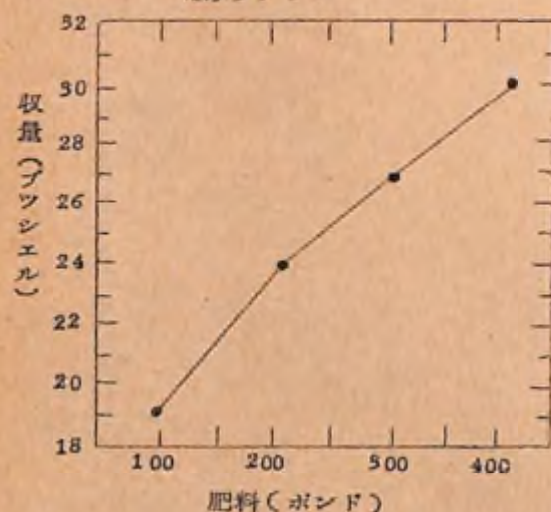
前の論議においては、小麦のプロットの収量、動物の体重、プロット当り消費にかつた植物の数又は一塊のパン容積などのような一つの変量に生ずる変動に重きを置いて述べた。本章においては、二つの変量の組合せ変動を主として取扱ひ、特に一つの変量の変動が他の変量の第二の変量の変動によつて影響されるような場合を取扱う。この様な場合一つを従属変数と云ひ、他を独立変数と云ふ。今二つの変量、雨量と収量があるとする。雨量は収量に影響し、したがつて、雨量は独立変数で、収量は従属変数であることはよく分つてゐることである。

第5章では、この型の問題にも一歩ふれた。例えば色々な水準の肥料を施したプロットの収量は、実際には回帰の問題を提供する。すなわち、こゝでは、収量は従属変数で、肥料の量は独立変数である。実際に試行して、次のような結果を得た。

|                 | 100 | 200 | 300 | 400 | 250 |
|-----------------|-----|-----|-----|-----|-----|
| 肥料の量(エーカー当りポンド) |     |     |     |     | 250 |
| 収量(エーカー当りブッシェル) | 19  | 24  | 27  | 30  | 25  |

この結果を6-1図に示す。

6-1図 収量と肥料の量との関係を示すグラフ



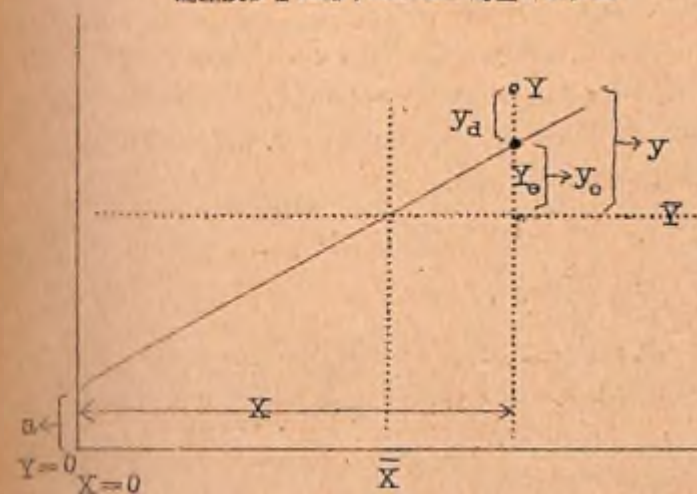
この結果について、まづ我々の注意をひく点は、肥料の量が一樣に増加すると、収量もかなり一樣に増加するということである。これが肥料の収量に対する効果を表わす実際的な方法である。しかし、これを統計用語でいえば、肥料の量にたいする小麦の収量の反応は直線的である。すなわち、直線により表わすことができるという。その次の考え方は、もし二つの変量の間の関係が直線であれば、 $Y = a + bX$  の型の方程式で表わすことができるということである。こゝにYは収量、Xは肥料の量である。項aは直線の原点を表わす、すなわち  $X=0$  なるときのYの値を表わし、bは直線の傾斜を表わす。したがつてbはXの単位増にたいするYの増加量を表わす。

したがつてbはXの単位増にたいするYの増加量を表わす。

### 2. 回帰方程式の誘導

第1節で概略述べた立場は、どのようにして必要な方程式を作ることができるかという問題を提供する。これには非常に多くのすぐれた理論が含まれてゐる。そして、すべての問題についてのすぐれた説明は、G. W. Snedecor の *Statistical methods* の第6章を参照されたい。こゝでは、普通の問題にたいして最も適当な方程式は、直線からの偏差の平方和が最小となるものであるということだけを述べるだけで満足しなければならない。今、 $Y_0 = a + bX$  を適合すべき回帰線としよう。こゝに添字0はYの推定量であることを示す符号である。この特別な型の直線方程式を導くには、あるきまつた理由がある。まづ第一に、常数aがきまると、Y軸上に直線が通るに必要な位置の値がきまり、それ以外のY軸上のどのような点も通つてはならない。したがつて、もし  $X=0$  であれば  $Y_0 = a$  である。もし  $Y_0 = bX$  とおくと、直線はYとXの原点を通らなければならない。常数a及びbの値は、推定の誤差の平方和が最小となるようであるべきである。6-2図は変量Y及びXの対にたいしての立場を示したものである。

6-2図 X, Yなる変量の一対にたいする回帰からの偏差及びYの標準平均からの偏差を示す図



$y_0$  を  $\bar{Y}$  からの  $Y_0$  の偏差とし、 $y_d$  を回帰線からのYの偏差とすれば、 $y = y_0 + y_d$  であることがわかる。適合すべき回帰線は、 $Y_0 = a + bX$  であるが、Y及びXを夫々の平均値からの偏差に変換してもよい。したがつて、回帰方程式は、もしそれが  $Y_0 = a + bX$  と同じであればa及びbにたいしてはこの段階ではまだ分らないから最初の記号を用いて  $Y_0 - \bar{Y} = a' + b'(X - \bar{X})$

又は  $y_0 = a' + b'x$  となる。最小にしなければならない平方和  
 $= \sum (Y - Y_0)^2 = \sum y_d^2$  6-2図から、 $y_d = (y - y_0)$ 、したがつて、 $y_0$  に代入すると、  
 $y_d = (y - a' - b'x)$ 。最小にしなければならない平方和は

$$\sum (y - a' - b'x)^2 = f$$

これを  $a'$  について微分すると

$$\frac{df}{da'} = -2 \sum (y - a' - b'x) = -2 \sum y + 2na' + 2b' \sum x$$

又は  $b'$  について微分すると、



$$\frac{df}{db'} = -2 \sum (y - a' - b'x) = -2 \sum xy + 2a' \sum x + 2b' \sum x^2$$

こゝに、 $n$ は変量の対の数を表わす。

両方の方程式を0とおいて次の二つの方程式を得る。

$$\sum y = a' n + b' \sum x \quad (1)$$

$$\sum xy = a' \sum x + b' \sum x^2 \quad (2)$$

これ等を最小自乗法によつて導かれた正規方程式と呼ぶ。したがつて、 $\sum y$ 及び $\sum x$ は0に等しいので

$$a' = 0 \quad b' = \frac{\sum (xy)}{\sum x^2} \quad (3)$$

統計量 $b'$ は通常、回帰係数と呼ぶ。回帰方程式は $y_o = b'x$ で、もし $y_o$ にたいして $y_o - \bar{y}$ を、 $x$ にたいして $x - \bar{x}$ を代入すれば

$$y_o = \bar{y} + b' (x - \bar{x}) \quad \text{又は} \quad = \bar{y} + b' x$$

$$\text{及び} \quad y_o = (\bar{y} - b' \bar{x}) + b' x \quad (4)$$

これは  $a = \bar{y} - b' \bar{x}$  とおくと  $y_o = a + b'x$  の一般的な型にすることができる。

小麦の収量及び肥料の量の例にたいして、回帰方程式を計算することは興味があることであるが、まず第一に、計算の手間をはぶくように $x$ についてある操作を行う。これは $x$ を100で割つて、ポンドの100の代りにポンドでの肥料の量を与える。これを資料の簡略と呼ぶ。そのとき、平均値からの偏差で結果を表わすと、

|     |      |      |     |     |
|-----|------|------|-----|-----|
| $x$ | -1.5 | -0.5 | 0.5 | 1.5 |
| $y$ | -6   | -1   | 2   | 5   |

そのとき

$$\sum x^2 = (-1.5)^2 + (-0.5)^2 + (0.5)^2 + (1.5)^2 = 5.0$$

$$\text{又} \quad \sum xy = (-1.5)(-6) + (-0.5)(-1) + (0.5)(2) + (1.5)(5) = 18.0$$

$$\text{で} \quad b = \frac{18}{5} = 3.6$$

このことは、回帰方程式により、肥料の100ポンドにたいしての収量の期待増加量は3.6ブツシエルであるということを示す。実際方程式は

$$y_o = 2.5 + 3.6(x - 2.5)$$

又は  $y_o = 2.5 - (3.6)(2.5) + 3.6x = 16.0 + 3.6x$   
グラフ上に直線を描くには、二点だけが必要である。

$$16.0 + 3.6 \times 1 = 19.4$$

$$16.0 + 3.6 \times 4 = 30.4$$

をとつて、グラフ上で直線を正しい位置におくことができる点を得る。回帰方程式は $\bar{x}, \bar{y}$

通り、このことは計算のチェックとなると云う事に注意せよ。

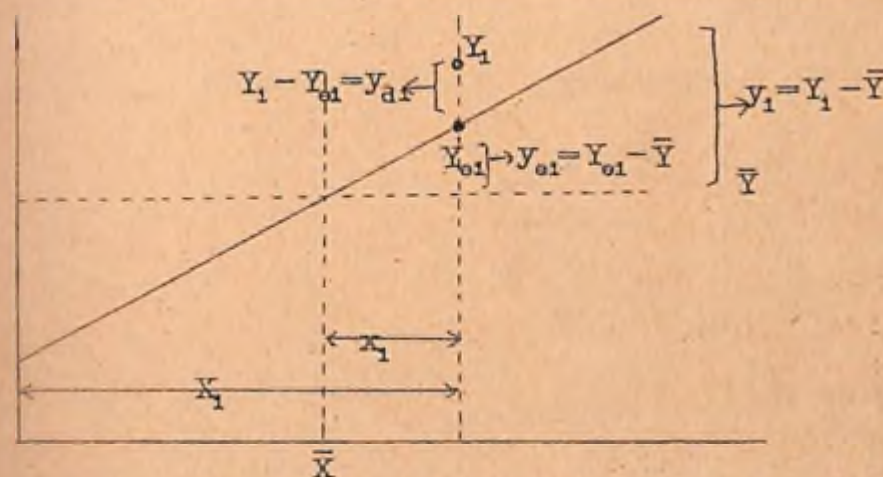
### 3. 回帰線の適合度

6-3図をみると、平均値 $\bar{y}$ からの個々の変量 $Y_1$ の偏差 $y_1$ は二つの部分 $y_{d1}$ 及び $y_{o1}$ とから成立っていることがわかる。すなわち

$$y_1 = y_{d1} + y_{o1}$$

$$\text{又} \quad \sum (y^2) = \sum (y_d + y_o)^2 = \sum y_d^2 + \sum y_o^2 + 2 \sum (y_d y_o) \quad (5)$$

6-3図 回帰からの偏差平方和及び総平均からの推定値の偏差の原理を示す図



回帰方程式  $y_o = \bar{y} + bx$  から  $y_o - \bar{y} = bx$  又は  $y_o = bx$  を得る。グラフから  $y_d = y - y_o = y - bx$  であることがわかる。したがつて

$$\sum (y_d y_o) = \sum (y - bx) bx = b \sum xy - b^2 \sum x^2$$

方程式 (3) から  $\sum xy$  にたいして、 $b \sum x^2$  を代入できることが分つていたので、明らかに

$$\sum (y_d y_o) = 0 \quad (6)$$

である。したがつて、(5)式の右辺の第三項はなくなる。故に次の非常に大切な一般式をうる。

$$\sum y^2 = \sum y_o^2 + \sum y_d^2 \quad (7)$$

これは $y$ の偏差の総和は、回帰を表わす第一の部分と、回帰からの偏差を表わす第二の部分に分けることができることを示す。

次の段階は各平方和に関連する自由度をきめることである。その手がかりは、 $\sum (y_o^2)$  をうるために標本から計算される統計量の数である。その数は順番に全く回帰方程式 $y_o = bx$ による。回帰係数は、この方程式をきめるに必要な一つの附加的な統計量である。したがつて  $\sum (y_o^2)$  は自由度1で表わされる。これは次のような分散分析を作る基礎を与える。



|         | SSi             | DF.   | MS.   |
|---------|-----------------|-------|-------|
| 全 体     | $\Sigma(Y^2)$   | $n-1$ |       |
| 回 帰     | $\Sigma(y_o^2)$ | 1     | $V_o$ |
| 回帰からの偏差 | $\Sigma(y_d^2)$ | $n-2$ | $V_d$ |

有意性の検定は  $F = V_o/V_d$  を計算し、自由度 1 及び  $n-2$  の 5% 点の  $F$  と比較すればよい。これは  $b$  が 0 と有意に違うかどうかの検定となる。

$b$  の有意性は  $t$  検定によっても検定できる。

$$s_o = \sqrt{\frac{\Sigma(y_o^2)}{n-2}} \quad (8)$$

としよう。こゝに  $s_o$  は推定量の標準誤差と呼ぶものである。そのとき、回帰係数の標準誤差は

$$s_b = \frac{s_o}{\sqrt{\Sigma(x^2)}} = \sqrt{\frac{\Sigma(y_o^2)}{(n-2)\Sigma(x^2)}} \quad (9)$$

であり

$$t = \frac{b}{s_b} \quad (10)$$

である。

二つの有意性の検定は、検定する平均平方を自由度 1 から推定するとき、 $F = t^2$  であるから同じものであるということは容易に証明される。

#### 4. 回 帰 係 数 の 信 頼 限 界

回帰係数を検定する際の  $t$  の一般式は

$$\pm t = \frac{(b - \beta)}{s_b} \quad (11)$$

である。こゝに  $\beta$  は我々が選んだ仮想母集団の回帰係数である。いゝかえれば、 $\beta = 0$  という母集団から無作為抽出して  $b$  を得たという仮定を含む通常の有意性検定では、 $\beta = 0$  とおくことができる。これにより  $b$  に対する信頼限界を定める事が出来る。すなわち

$$\beta = b \pm t s_b \quad (12)$$

そして、5% 点での信頼限界をうるためには、

$$l_1 = b - t_{0.05} s_b, \quad l_2 = b + t_{0.05} s_b \quad (13)$$

をとる。こゝに  $t_{0.05}$  は  $p = 0.05$  及び自由度  $(n-2)$  にたいする表での  $t$  の値である。

#### 5. 推 定 量 Y の 標 準 誤 差

推定量  $Y$  の標準誤差という言葉を開いて、それが  $Y$  の推定量の正確度に適用されると考えてはいけない。推定量の標準誤差は、特に我々が生検定を行つた際に分つたように、回帰係数の変動に適用する。しかし、有意性の検定を行うか又は  $Y_o$  にたいする信頼限界を作ることが必要な場合もある。回帰方程式  $y_o = \bar{Y} + bX$  から  $Y$  の平均平方は  $\bar{Y}$  及び  $bX$  の変動によるもので、それらは夫々別個な平均平方  $s_o^2/n$  及び  $(s_o^2/\Sigma X^2)X^2$  である。そして、これらを加えると  $Y$  にたいしての平均平方を得る。すなわち

$$s_{Y_o}^2 = \frac{s_o^2}{n} + \left(\frac{s_o^2}{\Sigma X^2}\right)X^2 = s_o^2 \left(\frac{1}{n} + \frac{X^2}{\Sigma X^2}\right) \quad (14)$$

計算目的には、次の形が便利である。

$$s_{Y_o}^2 = \frac{\Sigma(y_o^2)}{n-2} \left(\frac{1}{n} + \frac{X^2}{\Sigma X^2}\right) \quad (15)$$

この方程式から、 $X = (X - \bar{X}) = 0$  のとき  $s_{Y_o}^2$  が最小となることは明らかである。すなわち、 $Y_o$  を予測する  $X$  の値がその平均値に等しいときに最小である。  $X$  がその平均値から離れるにつれて、 $\bar{Y}_o$  の平均平方は大となる。

回帰直線は、抽出された母集団の  $X$  の与えられた値にたいする  $Y$  の値の推定量を表わし、又 (14) で与えられるような標準誤差は、回帰直線の両側での信頼限界を作るのに用いる。与えられた  $X$  を選び、それに応ずる  $Y$  を予測するために回帰方程式を用いることは全く別問題である。例えば、回帰問題は、豚の年令  $x$  の体重の回帰を研究することを含んでいる。与えられた年令  $X_1$  の豚にたいして、我々は平均体重が  $Y_1$  であろうということを回帰方程式で予測する。これは、母集団にたいする推定の問題で、信頼限界を作るために必要な標準誤差は、方程式 (14) のそれであろう。さて、もし我々が年令  $X_1$  の一匹の豚を選ば、その体重を推定したいとすれば、その答は再び  $Y_1$  であろうが、このような推定量の誤差は大となり、次式で与えられるものとなる。

$$s_{Y_o}^2 = s_o^2 \left(1 + \frac{1}{n} + \frac{X^2}{\Sigma X^2}\right) \quad (16)$$

#### 6. 両方の変量が共に誤差をうける場合の直線回帰

上に述べた例では、資料を肥料の色々な適用水準を含む一つの実験からとつたと仮定した。より一般的な場合は、恐らく二つの変量  $X$  及び  $Y$  の値の対の系列がある場合であろう。そして  $X$  から  $Y$  を予測する回帰直線を決定することが必要である。この場合に対しては、 $X$  は誤差なしに測定できるという仮定があることに注意しなければならぬ。  $X$  を病気の感染率、  $Y$  を収量とし、資料をプロットの系列からえたと仮定しよう。各プロットの感染率は正確に決定でき、これは一般に真のものに非常に近いので、回帰係数の値には影響しないと仮定できる。しかし、



もし我々が小標本により病気の感染率を推定すれば、 $X$ を決定する際には明らかに誤差があり、したがって回帰係数は偏倚を生ずるのである。

このことに関し、yates は次のように述べている。『独立変数の誤差は、仮りに0からの差だけを検定するとすれば、通常の有意性検定の妥当性には影響しない。そして通常のようにして評価した係数は類似の誤差をうける次々の観測からの予測の目的には常に最良の係数であろう』。彼は更に、独立変数の誤差は一般に回帰係数の絶対値を小さくする偏倚を導くことを指摘している。したがって、偏倚のある $b$ の値は、厳密にはもし従属変数に誤差があれば理論的母集団値 $\beta$ とは比較できない。これは、当然信頼限界を作る理論に影響してくるのである。独立変数が誤差をうける場合の有意性検定を取扱い、そしてそれを行う一般的な問題は、かなり複雑であるから、ここで詳細に述べることは容易ではない。完全な論議と、適した方法については、Deming の Statistical adjustment of data を参照されたい。

#### 7. 群にとりまとめていない資料に対する直線回帰の計算。

最も簡単な場合は $X$ 及び $Y$ の対の系列がある場合に起る。例えば

$$\begin{array}{ccccccc} X_1 & X_2 & X_3 & \dots & X_i & \dots & X_n \\ Y_1 & Y_2 & Y_3 & \dots & Y_i & \dots & Y_n \end{array}$$

があるとし、回帰方程式  $Y_p = a + bX$  が得たいとしよう。通常、我々は約50対以上でなければ資料を群分けしないであろう。適当な計算方法は対の値から直接に $\Sigma X$ ,  $\Sigma Y$ 及び $\Sigma XY$ を得ることである。それから次を計算する。

$$\Sigma x^2 = \Sigma X^2 - \frac{(\Sigma X)^2}{n}$$

$$\Sigma y^2 = \Sigma Y^2 - \frac{(\Sigma Y)^2}{n}$$

$$\Sigma xy = \Sigma XY - \frac{(\Sigma X)(\Sigma Y)}{n}$$

そして最後に

$$b = \frac{\Sigma xy}{\Sigma x^2}$$

を計算する。有意性の検定は、それに応ずる平方和及び自由度が夫々

$$\Sigma y^2 = \frac{(\Sigma xy)^2}{\Sigma x^2} + \Sigma (y_d^2) \quad (7)$$

$$n-1 = 1 + (n-2) \quad (8)$$

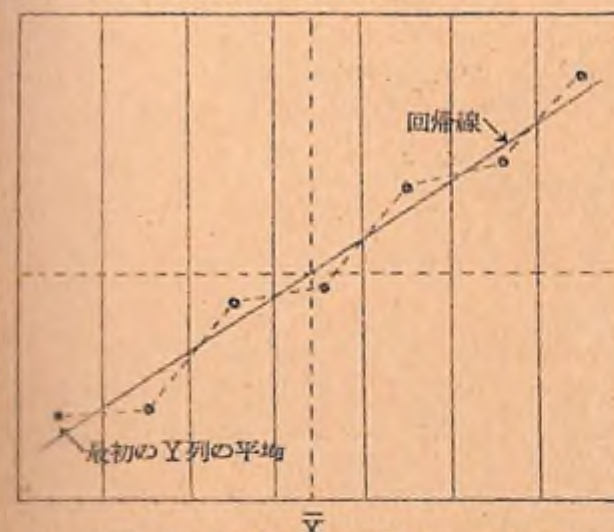
である。分散分析から行うことができる。

もし $b$ に対する信頼限界を作りたければ、公式 (8) 及び (9) のようにして $S_d$ を定め、4節のように計算を行えばよい。

#### 8. 回帰表に配列した資料に対する直線回帰

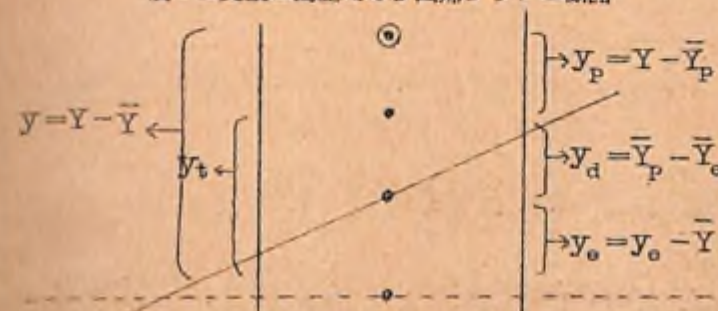
回帰分析は、7節で述べた方法によつていつも行うことができるが、50対以上ある場合には、6-1表のような表に資料を配列すると便利である。この種の表を作ることのもう一つの利益は、直線回帰からの偏差の有意性の検定が可能であるということである。

6-4図 回帰表からの列平均のグラフ



6-4図のように、回帰表をグラフで表わすと、我々は単一の観測と共に6-5図のように $Y$ の列の一つの拡大した断面を示すことができる。

6-5図 三つの部分にわけた $Y$ の平均値からの個々の変量の偏差を示す回帰グラフの断面



そのとき、 $y$ は  $y_p + y_d + y_o$  に分割することができる。ここに $y_p$ は列平均からの個々の観測値の偏差、 $y_d$ は列平均の回帰からの偏差

$y_o$ は $Y$ の平均からの回帰線の偏差である。平方和および自由度は次のように表わすことができることは容易に証明される。\*



$$\sum_1^N y^2 = \sum_1^q n_p y_p^2 + \sum_1^q n_p y_d^2 + \sum_1^q \sum_1^1 n_p y_p^2 \quad (19)$$

$$N-1 = 1 + (q-2) + (N-q) \quad (20)$$

全体 = 回帰 + 偏差 + 誤差

ここに、 $n_p$  は  $y$  列の度数、 $q$  は列の数、 $N$  は観測の対の全数である。これらの式から、有意性の検定を回帰に対してか又は回帰からの偏差に対して行うことができる。

註)\*

一つの例において

$$\sum_1^{n_p} y^2 = \sum_1^{n_p} (y_p + y_t)^2 = \sum_1^{n_p} y_p^2 + n_p y_t^2 + 2 \sum_1^{n_p} y_p y_t$$

又、 $y_t$  は列に対して一定  $\sum_1^{n_p} y_p = 0$  であるから

$$\sum_1^{n_p} y_p y_t = 0$$

$$\text{故に } \sum_1^{n_p} y^2 = \sum_1^{n_p} y_p^2 + n_p y_t^2$$

$\sum_1^q n_p = N$  である  $q$  列全体では

$$\sum_1^N y^2 = \sum_1^q \sum_1^{n_p} y_p^2 + \sum_1^q n_p y_t^2$$

これは列内と列間の平方和を表わす分析と同じ表示である。これに應ずる自由度は

$$N-1 = (N-q) + (q-1)$$

列間平方和は直線回帰を表わす部分と直線回帰からの偏差を表わす部分の二つの部分に分けることができる。すべての列上で

$$\sum_1^q n_p y_t^2 = \sum_1^q n_p (y_d + y_o)^2 = \sum_1^q n_p y_d^2 + \sum_1^q n_p y_o^2 + 2 \sum_1^q n_p (y_d y_o)$$

これは3箇の方程式 (6) と同じで、もし  $n_p$  がすべての列上で一定であれば、方程式 (6) になる。したがって、 $\sum_1^q n_p (y_d y_o) = 0$  で、これに應ずる自由度をもつ平方和に対しての完全な式は (19) および (20) で与えられる通りである。

回帰分析においては、概略を述べた有意性の検定がまづ行われるべきであろう。もし直線回帰の平均平方が有意でなければ、直線を適合する目的はなくなる。もし回帰からの偏差の平均平方が有意であれば、そのことは直線函数以外のある函数の方がよい適合を与えるということである。これはたとえ直線回帰の平均平方がまた有意であつても勿論言えることである。

## 9. 回帰表に対する計算

計算をどういふ風にやればよいかをおぼえる最も容易な方法は、例6-2の通りにやることである。この例はまた、表そのものの作り方も説明している。

多くの例において、資料を簡略にして計算労力を減らすとができる。これは、各変量の値から一定量を差引いたり、一定量で割つたり、ある場合には両方を組合せたりすることである。実際の値は左の端のような数字であると仮定しよう。右の値はそれを簡略にする方法を例示したものである。

| 簡略にしないもの |      | 簡略にしたもの |     |                   |
|----------|------|---------|-----|-------------------|
| X        | Y    | X       | Y   |                   |
| 2402     | 2785 | 240     | 278 | 10で割つて最後の数をまとめる。  |
|          |      | 40      | 78  | 200を引く            |
| 198      | 196  | 8       | 6   | 190を引く            |
| 195      | 193  | 5       | 3   |                   |
| 256      | 274  | 56      | 74  | 200を引く、しかし計算器を用いる |
| 229      | 198  | 29      | -2  | 場合は一般に負の値はさけない。   |

## 10. 例6-1 対の値の系列での回帰分析

仮想的な例として、10対の変量の値が下のように与えてあるとしよう。

|   |   |   |   |   |   |   |   |   |   |   |               |                 |
|---|---|---|---|---|---|---|---|---|---|---|---------------|-----------------|
| X | 9 | 8 | 7 | 7 | 6 | 4 | 3 | 3 | 1 | 2 | $\sum X = 50$ | $\bar{X} = 5.0$ |
| Y | 7 | 9 | 7 | 8 | 7 | 3 | 6 | 1 | 2 | 2 | $\sum Y = 52$ | $\bar{Y} = 5.2$ |

両方の変量の組の平方和および積和を見出すには、次のようにやればよい。

$$\sum XY = (9 \times 7) + (8 \times 9) + \dots + (2 \times 2) = 321$$

$$\frac{(\sum X)(\sum Y)}{N} = \frac{50 \times 52}{10} = 260$$

$$\text{差} = \sum XY = 61$$

$$\sum Y^2 = 7^2 + 9^2 + \dots + 2^2 = 346$$

$$\frac{(\sum Y)^2}{N} = \frac{(52)^2}{10} = 270.4$$

$$\text{差} = \sum Y^2 = 75.6$$



$$\begin{aligned}\Sigma X^2 &= 9^2 + 8^2 + \dots + 2^2 &= 318 \\ \frac{(\Sigma X)^2}{N} &= \frac{(50)^2}{10} &= 250 \\ \text{差} = \Sigma X^2 &= 68\end{aligned}$$

したがって、直線回帰にたいする平方和は

$$\frac{(\Sigma xy)^2}{\Sigma x^2} = \frac{(61)^2}{68} = 54.72$$

分散分析は次のように行うことができる。

|               | SS    | DF | MS    | F     |
|---------------|-------|----|-------|-------|
| 回 帰           | 54.72 | 1  | 54.72 | 20.97 |
| 誤差すなわち回帰からの偏差 | 20.88 | 8  | 2.610 |       |
| 全 体           | 75.60 | 9  |       |       |

回帰は非常に有意であるから、次によつて回帰係数を求める。

$$b_{yx} = \frac{\Sigma xy}{\Sigma x^2} = \frac{61}{68} = 0.8971$$

回帰方程式は

$$\begin{aligned}Y_c &= (\bar{Y} - b\bar{X}) + bX \\ &= (5.2 - 0.8971 \times 5.0) + 0.8971X \\ &= 0.7145 + 0.8971X\end{aligned}$$

5%点での  $b_{yx}$  の信頼限界は次で与えられる。

$$l_1 = b - t_{0.05} s_b \quad l_2 = b + t_{0.05} s_b$$

ここに  $s_b = \sqrt{20.88 / (8 \times 68)} = 0.1959$  で  $n_1 = 1$ ,  $n_2 = 8$  での 5%点の  $t$  の値は  $t_{0.05} = 2.31$  である。

$$\begin{aligned}\text{故に} \quad l_1 &= 0.8971 - 2.31 \times 0.1959 = 0.4446 \\ l_2 &= 0.8971 + 2.31 \times 0.1959 = 1.3496\end{aligned}$$

肥料の色々な水準に対する収量の推定量の標準誤差は、方程式 (5) から決定できる。一般に

$$s_{y_0}^2 = \frac{20.88}{8} \left( \frac{1}{10} + \frac{x^2}{68.00} \right) = 0.2610 + 0.0384x^2$$

したがって

$$\begin{aligned}x=0 \quad \text{にたいして} \quad s_{y_0} &= \sqrt{0.2610} = 0.511 \\ x=2 \quad \quad \quad s_{y_0} &= \sqrt{0.4146} = 0.644 \\ x=4 \quad \quad \quad s_{y_0} &= \sqrt{0.8754} = 0.936\end{aligned}$$

これらの計算から、回帰線の端附近での  $Y$  の値を推定する際には誤差の入る確率がかなり大きいことが分る。

## 11. 例6-2. 回帰表にした大きな対の系列に対する回帰分析

非常に多くの変量を取扱うときには、度数表を作つて、資料をとりまとめ、平均値及び標準偏差の計算労力を減すると便利である。同様に、回帰研究の場合にも、非常に多くの対の値があれば、二つの変量の度数表を組合せた表を作ることが望ましい。次章でまた、この表が相関係数の計算に役立つことが分るであろう。

表を作るには、取扱い易いカード上に対の値を書いておくことと便利である。すなわち、もし6-4表のカロチンの資料に対して表を作るとすれば、最初の三つの対にたいするカードは次の通りである。

| 第1のカード | 第2のカード | 第3のカード |
|--------|--------|--------|
| X 1.13 | X 2.13 | X 1.41 |
| Y 2.39 | Y 3.11 | Y 2.15 |

第2章9節で概略述べたような級の値をきめてから、変量の一つ、 $Y$ の方がよいが、このカードを分配し、ついで他の変量の札を分配する。各札を分配する時に、度数を表の適当な箇処に記載することができる。

6-1表は、本当の病人からとつた400の赤血球の透明度および面積を読んで、上述のようにして用意された。透明度の値は任意の単位で、光電管で操作したガルバノメーターのふれに基づいたものである。枠内において、実際の級の値を1から始まる自然数の系列でおきかえていることに注意せよ。これは計算の労力を減する簡略法の一つの型である。

回帰表を作つてから、後で用いる数字をのせる6-2表のような表を作ると便利である。欄頭に書いてある“ $Y$ 列総計”は次のようにして、各列での総計を計算して得ることができる。ここに  $Y$  の最初の列は  $X$  の最初の級に属する  $Y$  変量の分布である。

$$\begin{aligned}\text{第1列} \quad & (1 \times 7) + (2 \times 4) &= 15 \\ \text{第2列} \quad & (2 \times 8) + (2 \times 7) + (3 \times 6) + (1 \times 3) &= 51 \\ & \text{等}\end{aligned}$$

この欄の総計は明らかに  $\Sigma Y$  である。同様に、 $X$  列に対する総計および  $\Sigma X$  を得る。欄頭に  $\Sigma XY$  というのが二つあるが、これは二通りの方法で  $\Sigma XY$  を計算してチェックに用いる。最



初の $\Sigma XY$ の数字はY列の総計を、それに対応するXの級の値に乘じて得たものである。第二の $\Sigma XY$ の欄はX列の総計に、それに対応するYの級の値を乘じて得たものである。また、度数分布について第2章で概略のべたと同じ方法で、 $\Sigma X^2$ 、 $\Sigma Y^2$ を計算する。もし計算器が利用できれば、度数にそれに対応する級の値の自乗を乘じて加え合せると最も便利である。 $\Sigma XY$ もまた、計算器では連続的な乗法及び加法を行つて得られる。そしてその場合には6-2表の個々の欄は書かなくてもよい。

この例では

$$\begin{array}{lcl} \Sigma XY = 18,345.0000 & \Sigma Y^2 = 27,509.0000 & \Sigma X^2 = 13,337.0000 \\ \frac{\Sigma X \Sigma Y}{N} = 17,679.5325 & \frac{(\Sigma Y)^2}{N} = 25,808.4225 & \frac{(\Sigma X)^2}{N} = 12,111.0025 \end{array}$$

$$\Sigma xy = 665.4675 \quad \Sigma y^2 = 1,700.5775 \quad \Sigma x^2 = 1,225.9975$$

この段階で、我々は回帰からの偏差の有意性検定を導く計算を除く。この場合には、回帰分析を次によつて与えられる平方和を計算してから直接に行うことができる。

$$\text{回帰の平方和} = \frac{(\Sigma xy)^2}{\Sigma x^2} = \frac{(665.4675)^2}{1,226.00} = 361.22$$

分散分析

|     | SS      | DF  | MS     |
|-----|---------|-----|--------|
| 回 帰 | 361.22  | 1   | 361.22 |
| 誤 差 | 1339.36 | 398 | 3.3652 |
| 全 体 | 1700.58 | 399 |        |

Fの値は非常に大きいので、その計算はしなくてもよい。

回帰係数は次のように計算される。

$$b = \frac{\Sigma xy}{\Sigma x^2} = \frac{665.47}{1,226.00} = 0.5428$$

回帰方程式は

$$\begin{aligned} y_o &= (\bar{Y} - b_{yx} \bar{X}) + bX \\ &= (8.0325 - 0.5428 \times 5.5025) + 0.5428X \\ &= 5.0457 + 0.5428X \end{aligned}$$

もし図が必要であれば、 $y_o$ の二つの値、例えば  $X=1$  および  $X=12$  に対して計算し、それによつて回帰直線を図く。 $b_{yx}$ の標準誤差は次により与えられる。

$$s_b = \sqrt{\frac{3.3652}{1,226.00}} = \sqrt{0.002745} = 0.0524$$

5%点での $b_{yx}$ の信頼限界は

$$L_1 = 0.5428 - 1.97 \times 0.0524 = 0.4396$$

$$L_2 = 0.5428 + 1.97 \times 0.0524 = 0.6460$$

回帰からの偏差の検定をするには、6-2表の最後の欄の総計が必要である。この欄の各値はY列の総計の平方をそれに対応する度数で割つたものである。

$$\text{すなわち} \quad \frac{15^2}{75.00} = \frac{51^2}{325.12} = \text{等}$$

これは次のように列間の平方和を与える。方程式 (19) 参照

$$\Sigma n_p Y_d^2 = 26,181.23$$

$$\frac{(\Sigma Y)^2}{N} = 25,808.42$$

$$\Sigma n_p y_d^2 = 3,728.1$$

したがつて、完全な分析を行うことができる。

|         | SS      | DF  | MS     |
|---------|---------|-----|--------|
| 回 帰     | 361.22  | 1   | 361.22 |
| 回帰からの偏差 | 1159    | 10  | 115.9  |
| 誤 差     | 1327.77 | 387 | 3.431  |
| 全 体     | 1700.58 |     |        |

回帰からの偏差は期待したよりも小さいので、Fの値を計算する必要はない。

回帰からの偏差が小さいのは、実際には統計的に有意であることを示すことができる。

比  $3.431 / 115.9 = 2.96$  は自由度387および10での5%より大である。このような結果は、直線に列平均が異常に適合している理由を見出すために、実験者がもつと綿密な検定をしなければならないことを示すものである。この例では、なまの数字だけで、実験の詳細について分らないから、このように適合がよいのは、単に偶然にすぎないということが結論できるだけである。

もし回帰からの偏差が有意であつたならば、 $b$ に対して信頼限界を作る方法は、誤差項の平方和は回帰からの偏差を含むので、全々適当でないということとをここで指摘しなければならない。この分析では、 $b$ の信頼限界を得る方法は、誤差の平均平方3.431を用いるべきである。すなわち



6.2表 6-1表の資料の回帰分析の計算

したがって

$$L_1 = 0.5428 - 1.97 \times 0.0529 = 0.4386$$

$$I_2 = 0.5428 + 1.97 \times 0.0529 = 0.6740$$

回帰からの偏差は有意でないので、信頼限界は、前に得たものと非常によく一致している。

6.1表 400人の血液細胞に対する面積(X)への透明度(Y)の回帰の資料

|    | X             | 3 | 8 | 36 | 66 | 96 | 90 | 53 | 31 | 10 | 4 | 3 | 1 | 400 = n |
|----|---------------|---|---|----|----|----|----|----|----|----|---|---|---|---------|
| 14 | 0.691 - 0.720 |   |   |    |    |    |    |    |    |    | 1 |   |   | 1       |
| 13 | 0.661 - 0.690 |   |   |    |    |    |    |    |    |    |   |   |   | 0       |
| 12 | 0.631 - 0.660 |   |   |    | 1  | 1  | 3  | 1  | 1  |    |   |   | 1 | 8       |
| 11 | 0.601 - 0.630 |   |   | 1  | 3  | 3  | 6  | 5  | 4  | 2  | 2 | 3 |   | 29      |
| 10 | 0.571 - 0.600 |   |   | 4  | 5  | 13 | 16 | 11 | 12 | 3  | 1 |   |   | 66      |
| 9  | 0.541 - 0.570 |   |   | 4  | 8  | 14 | 21 | 15 | 6  | 3  |   |   |   | 71      |
| 8  | 0.511 - 0.540 |   | 2 | 4  | 3  | 29 | 21 | 12 | 3  | 2  |   |   |   | 86      |
| 7  | 0.481 - 0.510 | 1 | 2 | 6  | 10 | 17 | 10 | 4  | 4  |    |   |   |   | 54      |
| 6  | 0.451 - 0.480 |   | 3 | 5  | 8  | 8  | 8  | 4  |    |    |   |   |   | 36      |
| 5  | 0.421 - 0.450 |   |   | 5  | 12 | 7  | 2  |    |    |    |   |   |   | 26      |
| 4  | 0.391 - 0.420 | 2 |   | 3  | 3  | 3  | 3  | 1  |    |    |   |   |   | 15      |
| 3  | 0.361 - 0.390 |   | 1 | 3  | 1  | 1  |    |    | 1  |    |   |   |   | 7       |
| 2  | 0.331 - 0.360 |   |   |    |    |    |    |    |    |    |   |   |   | 0       |
| 1  | 0.301 - 0.330 |   |   | 1  | 1  |    |    |    |    |    |   |   |   | 2       |

|    |    |    |    |    |    |    |    |    |    |    |    |
|----|----|----|----|----|----|----|----|----|----|----|----|
| 23 | 23 | 23 | 23 | 41 | 44 | 47 | 50 | 53 | 56 | 59 | 62 |
| 24 | 24 | 24 | 24 | 42 | 45 | 48 | 51 | 54 | 57 | 60 |    |

Y

1

2

3

4

5

6

7

8

9

10

11

12

| Y   | Y  | X   | X          | Y           | $\Sigma(XY)$ | X           | $\Sigma(XY)$       | SS       |
|-----|----|-----|------------|-------------|--------------|-------------|--------------------|----------|
| 1   | 2  | 1   | 3          | 15          | 15           | 7           | 7                  | 7500     |
| 2   | 0  | 2   | 8          | 51          | 102          | 0           | 0                  | 32512    |
| 3   | 7  | 3   | 36         | 238         | 714          | 28          | 84                 | 1,573.44 |
| 4   | 15 | 4   | 65         | 465         | 1,860        | 63          | 252                | 3,326.54 |
| 5   | 26 | 5   | 96         | 750         | 3,750        | 110         | 550                | 5,859.38 |
| 6   | 36 | 6   | 90         | 759         | 4,554        | 169         | 1,014              | 6,400.90 |
| 7   | 54 | 7   | 53         | 464         | 3,248        | 268         | 1,876              | 4,062.19 |
| 8   | 86 | 8   | 31         | 285         | 2,280        | 465         | 3,720              | 2,620.16 |
| 9   | 71 | 9   | 10         | 95          | 855          | 420         | 3,780              | 902.50   |
| 10  | 65 | 10  | 4          | 46          | 460          | 403         | 4,030              | 529.00   |
| 11  | 29 | 11  | 3          | 33          | 363          | 204         | 2,244              | 363.00   |
| 12  | 8  | 12  | 1          | 12          | 144          | 54          | 648                | 144.00   |
| 13  | 0  |     |            |             |              | 0           |                    |          |
| 14  | 1  |     |            |             |              | 10          | 140                |          |
| 400 |    | 400 | 3,213      | 18,345      | 2,201        | 18,345      | 26,181.23          |          |
| n   |    | n   | $\Sigma Y$ | $\Sigma XY$ | $\Sigma X$   | $\Sigma XY$ | $\Sigma np(y_a^2)$ |          |



## 第 7 章 相 関

### 1. 共 変 動

この言葉は、これから相関法がでてくる二つの変量に関する基礎的な関係を表わすものである。前章では、二つの変量が論理的に他に関連あるものと考えられるようにして関係づけられている場合に、回帰法がこの分析に完全に適用できることを指摘した。しかし、二つの変量が従属および独立と考えられないときは、回帰法は全くは満足でない。兄弟姉妹の身長間の関係の研究をすると仮定しよう。対の一方の身長が他の身長に関係があると考えすることは論理的ではないであろうが、或る場合に二変量が関係ある事がはつきりしていてこの場合、その関係を測定したく、他の類似の関係とそれとをある論理的方法で比較できるようにしたい。二つのこのような変量に対してしばしば次のような質問がおこる。すなわち、兄妹の身長は一緒にどの程度変化するか？すなわち、我々は共分散という言葉を用い、共変動の測定に対しての旧来の統計量は相関係数である。

### 2. 相 関 の 定 義

7-1表は、 $X_1$  および  $X_2$  で表わした二つの変量の測定値としてとつた数字の三組を表わす。

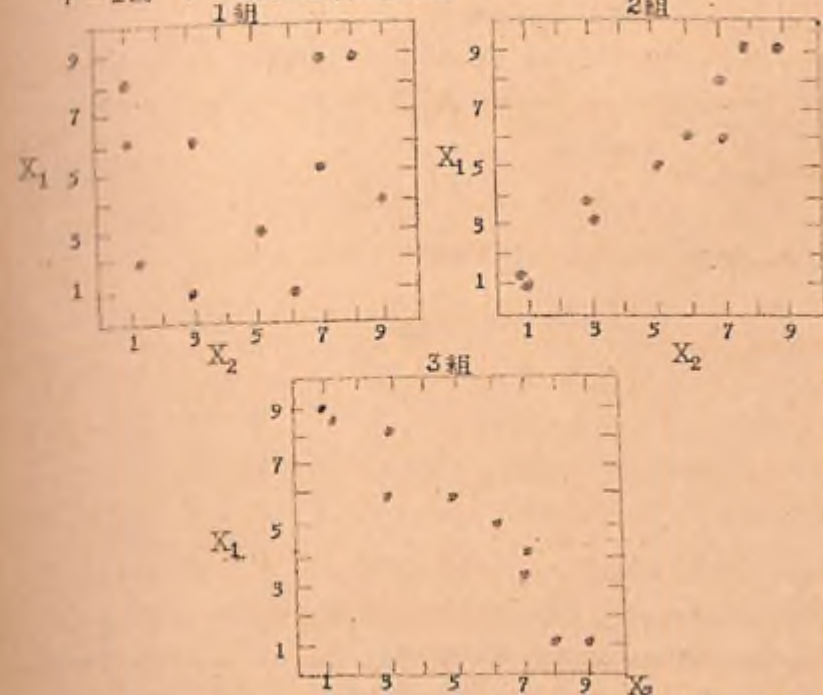
7-1表 相関現象を例示する対の変量の三つの標本

|    |       |   |   |   |   |   |   |   |   |   |   |         |
|----|-------|---|---|---|---|---|---|---|---|---|---|---------|
| 1組 | $X_1$ | 7 | 7 | 1 | 6 | 5 | 3 | 8 | 9 | 3 | 1 | 合計 = 50 |
|    | $X_2$ | 5 | 9 | 6 | 1 | 3 | 1 | 9 | 4 | 6 | 8 | 合計 = 52 |
| 2組 | $X_1$ | 9 | 8 | 7 | 7 | 6 | 5 | 3 | 3 | 1 | 1 | 合計 = 50 |
|    | $X_2$ | 9 | 9 | 8 | 6 | 6 | 5 | 4 | 3 | 1 | 1 | 合計 = 52 |
| 3組 | $X_1$ | 1 | 1 | 3 | 3 | 5 | 6 | 7 | 7 | 8 | 9 | 合計 = 50 |
|    | $X_2$ | 9 | 9 | 8 | 6 | 6 | 5 | 4 | 3 | 1 | 1 | 合計 = 52 |

これらを検定すると、 $X_1$  および  $X_2$  の間の関係が各々異つていことが分るであろう。第2組では、 $X_1$  の大きな値が  $X_2$  の大きな値と関係があり、第3組では、 $X_1$  の大きな値が  $X_2$  の小さな値と関係がある。二つの組共に明らかな関係があるが、一方は他の逆である。一方、第1組では、二つの変量間には関係がないようである。これらの組は、対の変量の無限母集団からの標本と考えられる。第2組が抽出された母集団では、変量の対を無作為に抽出するときにはいつでも、もし対が  $X_1$  の大きな値を含めば、 $X_2$  の大きな値がそれと関係するであろうという事が期待される。第3組の標本によつて表わされた母集団では、 $X_1$  の大きな値が  $X_2$  の小さな値と関係することが期待される。これら二つの反対の関係は、正および負の相関と呼

ばれるものである。第1組はなおそれ以外の関係を表わす。 $X_1$  の大きな値は  $X_2$  の大または小の値何れとも関係はないようである。いいかえれば、母集団では、二つの変量は独立に変動しているということが期待される。これら3つの標本の結果を図示すれば7-1図の通りである。各標本は点図表または散布図表として知られているものにより表わされる。 $X_1$  の値を縦軸に、 $X_2$  の値を横軸に表わすと、各対は図上に点で表わすことができる。最終結果は点のちらばりによつて、2変量間の関係を表わす図である。

7-1図 7-1表の資料の散布図



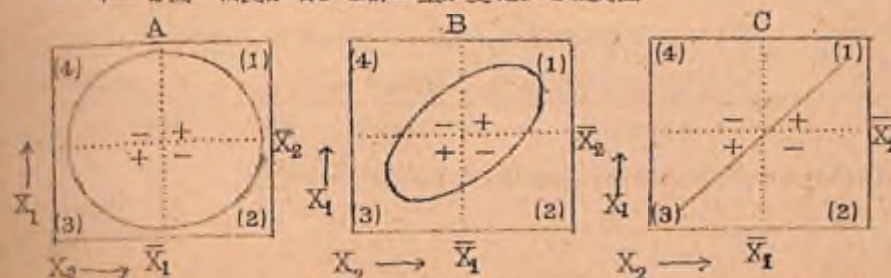
第1組は、点は多少とも全面に一樣に分布している。第2組および3組に対しては、点の傾向が正方形の対角線に沿つて直線に配列して、変量間には一定の関係がある。我々は、ここで前章の回帰図表を思い起す。その相違は、我々

は一つの変量の他の変量への影響を研究するのではなく、変量がおそらく両方に共通なある影響のために一緒に変化する程度を研究するということである。もしこのような測定が兄妹の身長であるとすれば、この共通の影響はその遺伝の類似性であるということとは明らかである。

### 3. 相 関 の 尺 度

7-2図は三つの相関の異つた程度に対する相関面での群の型を示す。

7-2図 相関が増すと群の型が変化する相関面





Aの円群は0相関を示す。Cでは群はすべて斜線上に属し、完全相関を表わしている。Bは群の二つの極端な場合の中間の状態を示す。今各面を平均値の位置で垂直にたてた線で4象限に分け、各象限で+と-の符号平均からの  $X_1$  および  $X_2$  の偏差の積の符号を表わしている。したがって、右上の象限1では  $X_1$  および  $X_2$  の偏差は共に正であるから偏差の積は正である。したがって、積は1象限と3象限では正で、2象限では負である。さて、Aでは正および負の積はお互に打消し、その和は0であることは明らかである。Cでは、すべての積は正となり、その和は最大となろう。Bでは、その条件はAおよびCの中間である。正の積は負の積よりも大きい。したがって、正ではあるが完全相関ではない。

さて、7-1表の数字の組を考えよう。各組に対して、積和  $\sum X_1 X_2$  を計算すると上に横軸述べた理論と一致することがわかる。これらの計算をするためには次の等式を用いる。

$$\sum X_1 X_2 = \sum X_1 X_2 - \frac{\sum X_1 \sum X_2}{n} \quad (1)$$

ここに、 $\sum X_1$  は  $X_1$  の値の和、 $\sum X_2$  は  $X_2$  の値の和で  $n$  は対の数である。計算は次のようにして実行する。

|     | $\sum X_1 X_2$ | $\frac{\sum X_1 \sum X_2}{n}$ | $\sum X_1 X_2$ |
|-----|----------------|-------------------------------|----------------|
| 1 組 | 262            | 260                           | 2              |
| 2 組 | 335            | 260                           | 75             |
| 3 組 | 186            | 260                           | -74            |

この結果は、積の和が相関の尺度であるという理論と一致する。

積和は相関の絶対的尺度であるが、相対的尺度としては役に立たない。なぜならば、それは個々の変量に対しての資料の特性値である因子に依存し、それらの間の相関と何等関係がないからである。それは、測定した変量の対の数、測定単位および両方の変量の変動性による。第一の難点は自由度で割って我々が共分散と呼んでいる、 $\sum X_1 X_2 / (n-1)$  を用いると除去される。これを記号で  $\mu_{12}$  で表わすことができる。他の点を除去するには標準尺度として知られている変量を思い出せばよい。もし変量を  $X$  で表わせば、標準尺度では

$$d = \frac{X - \bar{X}}{s_x} = \frac{x}{s_x}$$

で表わされる。ここに  $s_x$  は標本標準偏差である。したがって、最後に次のものを得る。

$$r_{12} = \frac{\sum (x_1/s_1)(x_2/s_2)}{n-1} = \frac{\sum (x_1 x_2) / (n-1)}{s_1 s_2} = \frac{\mu_{12}}{s_1 s_2} \quad (2)$$

標準偏差の代りに、 $X_1$  および  $X_2$  の分散を与える他の公式は

$$r_{12} = \frac{\sum (x_1 x_2) / (n-1)}{\sqrt{\nu_1 \nu_2}} \quad (3)$$

ここに、 $\nu_1$  は  $X_1$  の分散、 $\nu_2$  は  $X_2$  の分散である。回帰係数は

$$b_{12} = \frac{\sum x_1 x_2}{\sum x_2^2} = \frac{\sum x_1 x_2 / (n-1)}{\sum x_2^2 / (n-1)} = \frac{\mu_{12}}{\nu_2}$$

$$\text{および } b_{21} = \frac{\mu_{12}}{\nu_1}$$

であるから

$$r_{12}^2 = \frac{\mu_{12}}{\nu_2} \times \frac{\mu_{12}}{\nu_1} = b_{12} b_{21}$$

$$\text{また } r_{12} = \sqrt{b_{12} b_{21}} \quad (4)$$

したがって、相関係数は二つの回帰係数の幾何平均である。

#### 4. 相関係数の範囲および解釈

$$d_1 = \frac{x_1}{s_1} \quad \text{および} \quad d_2 = \frac{x_2}{s_2}$$

とおけば、相関係数は次のきれいな形に書ける。

$$r_{12} = \frac{\sum (d_1 d_2)}{n-1} \quad (5)$$

および次のようになることは容易に証明できる。\*

$$r_{12} = 1 - \frac{\sum (d_1 - d_2)^2}{2(n-1)} \quad (6)$$

註)\* (6) の右辺を展開すると

$$r_{12} = 1 - \left( \frac{\sum d_1^2}{2(n-1)} + \frac{\sum d_2^2}{2(n-1)} - \frac{\sum d_1 d_2}{n-1} \right) \quad \begin{matrix} s_1^2 = \sum x_1^2 / (n-1) \\ \text{であるから} \end{matrix}$$

$$\frac{\sum d_1^2}{2(n-1)} = \frac{\sum (x_1/s_1)^2}{2(n-1)} = \frac{\sum x_1^2}{2s_1^2(n-1)} = \frac{1}{2} \quad \text{同時に} \quad \frac{\sum d_2^2}{2(n-1)} = \frac{1}{2}$$

$$\therefore r_{12} = 1 - \left( \frac{1}{2} + \frac{1}{2} - \frac{\sum d_1 d_2}{n-1} \right) = \frac{\sum (d_1 d_2)}{n-1}$$



また

$$r_{12} = -1 + \frac{\sum (d_1 + d_2)^2}{2(n-1)} \quad (7)$$

方程式 (6) は  $\sum (d_1 - d_2)^2$  は 0 以下にはなり得ないから、 $r_{12}$  は +1 で最大となることを示している。方程式 (7) は  $r_{12}$  が -1 で最小となることを示している。したがって一般に

$$1 \geq r_{12} \geq -1 \quad (8)$$

これが相関係数がとり得る値の完全な範囲である。

相関と直線回帰との関係は、第 6 章 3 節で与えた従属変数の平方和に対しての方程式をふりかえつてみると第 7 章 3 節で与えたものから拡張できる。従属変数として  $X_1$  を考えると、基礎方程式は

$$\sum X_1^2 = \sum X_{1d}^2 + \sum X_{1o}^2 \quad (9)$$

ここに  $X_{1d}$  は  $X_{1o} = b_{12} X_2$  によつて定義された回帰線からの  $X_1$  の値の偏差を表わし、 $X_{1o}$  は回帰方程式によつて推定された  $X_1$  の値を表わす。回帰係数  $b_{12}$  は勿論、 $\sum X_1 X_2 / \sum X_2^2$  で与えられる。 $X_{1o} = b_{12} X_2$  であるから

$$\sum X_{1o}^2 = b_{12}^2 \sum X_2^2$$

$b_{12}$  に  $\sum X_1 X_2 / \sum X_2^2$  を代入すると、 $\sum X_{1o}^2 = \frac{(\sum X_1 X_2)^2}{\sum X_2^2}$

$$r_{12}^2 = (\sum X_1 X_2)^2 / \sum X_1^2 \sum X_2^2 \quad \text{であるから定義により} \quad \sum X_{1o}^2 = r_{12}^2 \sum X_1^2$$

そのとき平方和に対しての方程式は次のように書くことができる。

$$\sum X_1^2 = \sum X_{1d}^2 + r_{12}^2 \sum X_1^2 \quad (10)$$

$$= (1 - r_{12}^2) \sum X_1^2 + r_{12}^2 \sum X_1^2 \quad (11)$$

これに應ずる自由度は

$$n-1 = (n-2) + 1$$

これから  $r_{12}$  の有意性検定を与える分散分析がでてくる。明らかに

$$F = \frac{r_{12}^2 (n-2)}{1 - r_{12}^2} \quad (12)$$

であり、これは回帰係数を検定するために  $F$  を計算したときに得られる値と全く同じ値である。この分析は、本質的に相関係数は直線回帰の尺度であるということを示している。

相関と回帰の直線性との関係は、幾分違った方法で方程式 (9) を取扱うともつと明らかになることができる。 $\sum X_{1o}^2$  に  $r_{12}^2 \sum X_1^2$  を代入すると、

$$r_{12}^2 \sum X_1^2 = \sum X_1^2 - \sum X_{1d}^2 \quad (13)$$

$$\text{また} \quad r_{12}^2 = 1 - \frac{\sum X_{1d}^2}{\sum X_1^2} \quad (14)$$

同様に  $\sum X_{1d}^2$  に  $(1 - r_{12}^2) \sum X_1^2$  を代入すると、

$$r_{12}^2 = \frac{\sum X_{1o}^2}{\sum X_1^2} \quad (15)$$

我々はまづ (14) から  $r_{12}^2$  は回帰からの偏差平方和が 0 に近づくことと最大に近づくことを注意する。方程式 (15) は  $r_{12}^2$  は回帰に対する平方和と全体に対する比にすぎないということを示している。後者はどのような回帰の型をも表わす相関の一般概念の始めとして考えられるから非常に有用な概念である。したがって、もし曲線を表わすもつと複雑な方程式を適合すれば、曲線上にある  $X_1$  の推定値の系列をうるであろう。直線回帰に対しての  $r$  の概念を拡張すると次のようにおくことができよう。

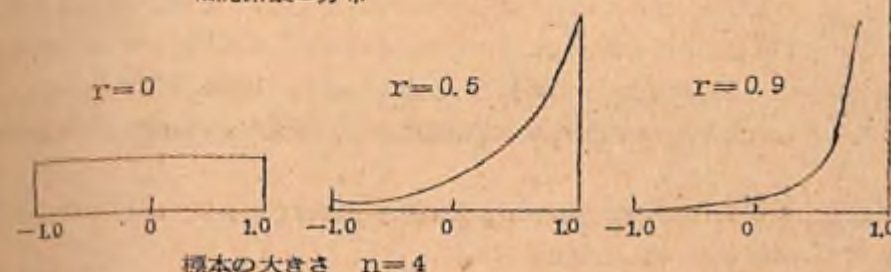
$$r_{12}^2 = \frac{\sum X_{1o}^2}{\sum X_1^2}$$

ここに  $\sum X_{1o}^2$  は曲線の回帰に対しての平方和を表わす。このような係数は通常、曲線相関指数 (Indices of curvilinear correlation) と呼ばれる。方程式 (11) は相関係数の値の解釈には特に重要である。従属変数の平方和は二つの部分にわかれる。一つの部分は  $1 - r_{12}^2$  に比例する回帰からの偏差を表わし、もう一つの部分は  $r_{12}^2$  に比例する回帰を表わす。近似的ではあるが価値ある概念は、 $r_{12}^2$  が相関の重要性を測るものであると仮定することである。例えば、収量と土壌水分の関係を測つて  $r_{12} = 0.90$  なる係数を得たが、プロットを処理する方法を変えることによつて、これを 0.95 にすることができるとしよう。 $r_{12}^2$  の増加量は  $0.95^2 - 0.90^2 = 0.0925$  である。もし最初の係数が 0.40 であり、それが 0.50 に増加すれば、その差もやはり 0.09 である。したがって、この水準での 0.1 の増加はより高次の水準での 0.05 の増加ほど重要ではない。

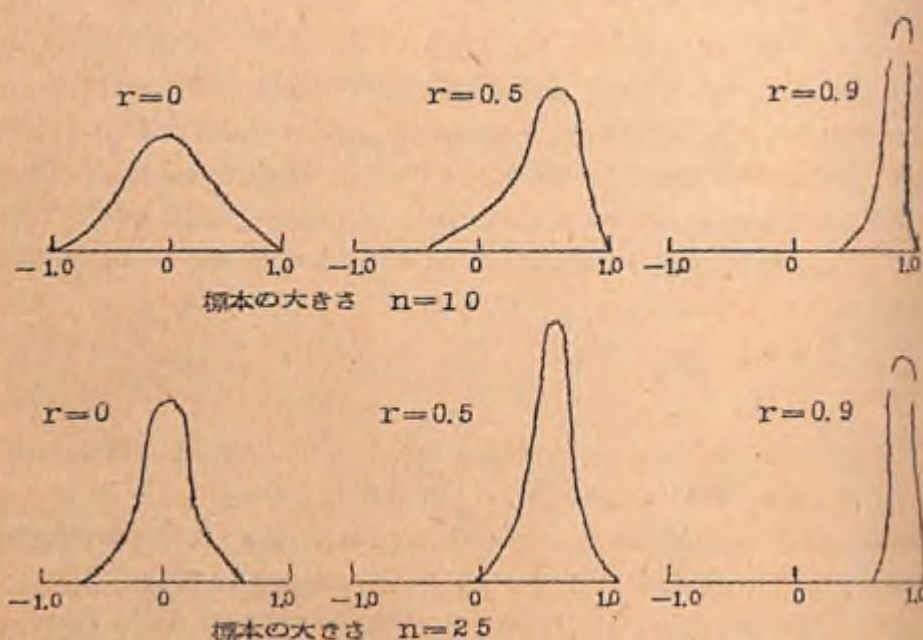
## 5. 相関係数の抽出分布

相関係数  $\rho_{12} = 0$  (記号  $\rho_{12}$  は  $r_{12}$  を推定量とする母数であるから、 $r_{12}$  の代りに用いる) という母集団があり、標本の多くの系列をとり、各々に対して  $r_{12}$  をきめたと仮定しよう。

7-3 図  $\rho$  の色々な値および色々な標本の大きさに対しての相関係数の分布







もし  $r_{12}$  の値を度数分布に作ると、標本が大きくなつたときのみ正規に近づくが、分布は対称であることがわかる。一方、 $\rho_{12}=0.8$ であれば、 $r_{12}$  の分布は全く歪んでしまう。7-3図では、これは Shewhart から引用したものであるが、この現象を色々な  $\rho$  の値ばかりでなく色々な標本の大きさに対して例示してある。

$r$  の分布の性質は有意性の検定および信頼限界の決定に重要である。t 分布に基いた通常の有意性普通検定法は、我々の帰無仮説が  $\rho_{12}$  は抽出された母集団では 0 であるから何等の影響も受ない。 $\rho_{12}=0$  という母集団を仮定すると、我々は観測したものと同じかまたは大きい  $r_{12}$  の値を得る確率を見出したい。我々は次のようにきめる。

$$t = \frac{r_{12} \sqrt{n-2}}{\sqrt{1-r_{12}^2}} \quad (16)$$

これは  $r_{12} s_{r_{12}}$  に等しい。ここに  $s_{r_{12}}$  は  $r_{12}$  の標準誤差である。分散分析からきめたように、 $t = \sqrt{F}$  であることをここに注意しておこう。そのとき、t 表から  $(n-2)$  なる自由度に対する  $p$  の相応する値を見出す。この検定は、 $\rho_{12}=0$  という母集団からとられた  $r_{12}$  の値の系列からきめられたところの t 分布は典型的なもので、理論とよく一致するから有効である。

しかし、もし  $\rho_{12}=0.8$  であれば、標本から計算した t 分布は  $r_{12}$  の分布の歪みのためにその表の基となつてゐる分布と一致しないであろう。

したがつて、 $\rho_{12}=0$  でなければ

$$\pm t = \frac{r_{12} - \rho_{12}}{s_{r_{12}}} \quad \text{に基いた検定に対しては、t 分布は満足ではない。また、t 分布は } \rho_{12} \text{ が 0 か 0 に近いものでなければ、} r_{12} \text{ の信頼限界を作るのに適當ではない。}$$

この困難を克服するため、R. A. Fisher は  $r$  を近似的に正規分布する統計量  $Z$  に変換する方法を発表した。これは次のようにすればよい。

$$Z = \frac{1}{2} \log_e \left( \frac{1+r}{1-r} \right) \quad (17)$$

$$\text{また } s_z = \frac{1}{\sqrt{n-3}} \quad (18)$$

$$\text{したがつて } \pm t = \frac{Z - \zeta}{s_z} \quad (19)$$

$$\text{これは } l_1 = Z - t_{0.05} s_z \quad l_2 = Z + t_{0.05} s_z$$

を与える。そして  $l_1$  および  $l_2$  を  $r$  の値にもどすと、5%点での信頼限界を得るものである。したがつて、正規偏差の表は  $t$  の表と同じ結果を与えるであろう。

変換はまた相関係数間の差の検定にも有用である。すなわち

$$s_{z_1 - z_2} = \sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}} \quad (20)$$

ここに  $n_1$  および  $n_2$  は二つの相関がそれからきめられた標本での変量の対の数である。そのとき

$$t = \frac{Z_1 - Z_2}{s_{z_1 - z_2}} \quad (21)$$

この値の検定を行う際に、相関係数の二つの値は、相互に相関があつてはならないということと記憶しておかねばならない。例えば、我々が、測定に対して各プロットから無作為に一部をとり、プロットの系列で昆虫の被害と収量との相関係数を計算したとしよう。それから、もし同様な相関係数が同じプロットの組を抽出して後日得られたとすれば、二つの相関係数は関係がありここで述べた検定法は用いられない。しかし、全く別々なプロットの組から相関を得たとすれば、係数は独立と考えることができ、検定は妥當である。

相関の対または直線回帰そのものが関係があり、差の有意性検定が必要であるときには、Yates が述べた方法が全く満足である。文献は、F. Yates: proc. Roy. Soc. Edinburgh, 59, 1938~39 参照のこと。簡単な場合すなわち、 $X_1$  の一つの値および独立変数  $X_2$  および  $X_3$  の値の二組があり、 $r_{12}$  と  $r_{13}$  を比較するような場



合には、新しい変量  $X_4 = (X_2 - X_3)$  をきめ、 $X_1$  と  $X_4$  の相関をきめるとよい。yates は  $X_1$  および  $X_2$  と同様、 $X_1'$  および  $X_2'$  の二つの変量の二組があり、そして相関または回帰があるということがわかっている場合のもつと困難な問題及びこの問題を上に引用した論文にこの問題を論じている。

## 6. 相 関 係 数 の 計 算

もし研究者が第 6 章で与えた回帰係数の計算方法がわかれば、この方法は相関係数の計算に適用できるということが明らかとなる。回帰係数に対しては、単に  $\sum X_1 Y$  および  $\sum X_2^2$  だけが必要であるが、相関係数に対しては  $\sum Y^2$  を加える必要がある。この章の記号を用いると

$$\sum X_1 X_2 = \sum X_1 X_2 - \frac{\sum X_1 \sum X_2}{n}$$

$$\sum X_1^2 = \sum X_1^2 - \frac{(\sum X_1)^2}{n}$$

$$\sum X_2^2 = \sum X_2^2 - \frac{(\sum X_2)^2}{n}$$

これらは、もし  $n \leq 50$  対以下であれば、対にした値から直接計算できるが、50 対以上であれば第 6 章のような回帰表を作った方がよい。

## 第 8 章 偏回帰と重回帰および 偏相関と重相関

### 1. 偏回帰および重回帰の問題

多くの回帰の問題において、研究者は他の変量への一つの変量の影響を取扱っている。穀物の蛋白質含有量を測定する新しい方法が見出されて、これと非常に正確な結果を与えることが分っている。以前の十分検討された方法とを比較する場合を仮定してみよう。この二つの方法は標本の非常に大きな系列で検討されて、新しい方法の結果に対する古い方法での結果の回帰が定まる。この目的は、新しい方法によつて得た結果から、標本の蛋白質含有量がどの位厳密に予測できたかを調べることである。この例では、新しい方法で実際に蛋白質を測定できるかどうか、またある特性が蛋白質と密接な関係があるかどうかというようなことは問題ではない。必要なものは、予測の際の高度の正確さである。

これと違つた種類の例では、我々はよく従属変数への一つ以上の独立変数の影響に関心をもつ場合がある。そして簡単な回帰係数で与えられるような全体の関係は、満足な情報を与えない。我々には、雨量への小麦の収量の回帰が、雨量の 10 分の 1 インチ毎に 1 ブツシエルの増収によつて変えられることが分つたとしよう。このことから、雨量の与えられた増加が常に収量の与えられた増加の原因となつていると結論できるであろうか？この答は否である。というのは収量が他の因子例えば温度などによつて影響されるからである。それらのものは異なる条件のもとで得られた資料では違つてくるのである。また、温度への雨量の有意な回帰が存在するかも知れないし、観測した収量の増加が単一の因子によるかまたは二つの因子の組合せによるものかをきめることは困難である。

この種の例で我々が求めるものは、他の変量が一定であるときの、二つの変量間の関係の尺度である。これが偏回帰および偏相関法の目的である。

### 2. 偏回帰法

単一すなわち全回帰は  $Y_0 = a + bx$

という形の回帰方程式を基としているから、偏回帰は

$$y_0 = a + b_1 X_1 + b_2 X_2 + \dots + b_p X_p \quad (1)$$

という形の回帰方程式に基いている。ここに  $p$  は独立変数の数で、回帰係数  $b_1, b_2, \dots, b_p$  は偏回帰係数と呼ばれるものである。方程式、我々が関心を持つている  $p$  個の変量の、個々の値から  $Y$  の値を予測する方法を表わすから、実際には一つの重回帰方程式である。

平均値からの偏差に変換すると、方程式 (1) は次のようになる。

$$y_0 = b_1 x_1 + b_2 x_2 + \dots + b_p x_p \quad (2)$$

また、 $Y$  の単一直を推定する際の誤差は



$$y - b_1 x_1 - b_2 x_2 - \dots - b_p x_p \quad (3)$$

である。

単純回帰と同じ方法に従うと、我々の問題は推定の誤差の平方和を最小にするような  $b_1, b_2, \dots, b_p$  の値を見出すことである。すなわち、

$$\sum (y - b_1 x_1 - b_2 x_2 - \dots - b_p x_p)^2 \quad (4)$$

を最小にする必要がある。そしてこれから最小自乗法によつて  $p$  個の連立方程式が導かれる。これは3つの独立変数に対して下に示してあるが必要な独立変数の数に対して容易に拡張できる。

$$\begin{aligned} b_1 \sum x_1^2 + b_2 \sum x_1 x_2 + b_3 \sum x_1 x_3 &= \sum x_1 y \\ b_1 \sum x_1 x_2 + b_2 \sum x_2^2 + b_3 \sum x_2 x_3 &= \sum x_2 y \\ b_1 \sum x_1 x_3 + b_2 \sum x_2 x_3 + b_3 \sum x_3^2 &= \sum x_3 y \end{aligned} \quad (5)$$

これらの方程式を解くと、 $b_1, b_2, b_3$  の値が求まり、次の予測方程式を得る。

$$y_e = b_1 x_1 + b_2 x_2 + b_3 x_3$$

偏回帰係数は、方程式の中では省略された記号で表わされているということに注意しておかなければならない。したがつて、詳細に  $b_{y1.23}$  を書くためには、 $b_{y1.23}$  と書かねばならず、これは変数  $x_2$  および  $x_3$  を一定にしたときの、 $Y$  の  $x_1$  への偏回帰係数であることを示す。

### 3. 偏回帰の意味

偏回帰係数  $b_{y1.23}$  は  $x_2$  と  $x_3$  を固定したときの  $Y$  の  $x_1$  への回帰を表わすと述べたが、この関係をもつと直接的な形にすることが可能である。2つの独立変数  $x_1, x_2$  があり、偏回帰係数  $b_{y1.2}$  がほしいと仮定しよう。我々は次の単純回帰方程式をきめることができる。

$$y_e = b_{y2} x_2 \quad \text{および} \quad x_{1e} = b_{12} x_2$$

最初の方程式から  $Y$  の推定値の系列を得ることができる。それから第2の方程式から  $x_1$  の推定系列を得ることができる。それぞれの回帰線からの  $Y$  および  $x_1$  の実測値の偏差は次で表わされる。

$$u = y - b_{y2} x_2 \quad \text{および} \quad v = x_1 - b_{12} x_2$$

これは共に  $x_2$  と独立である。したがつて、必要な回帰係数は  $b_{uv}$  となる。これを偏回帰係数の誘導の出発点とすることができる。それは正規方程式を解いて得た  $b$  の値に代入したり代数的操作をほどこしたりして得る。このようにして  $b_{uv} = b_{y1.2}$  となる。この方法は  $p$  個の独立変数をもつ回帰方程式の誘導にもおしよめることができる。

### 4. 偏回帰係数の計算

方程式 (5) は、未知数を次々に消去して解く初等的な方法を、直接に適用して解き得るとは明らかである。これ等は代数学の本にのつてゐる。ドーリットル法の (Doolittle

method) と呼ばれる正規方程式の解の表形式の方法は非常に便利で、最近 Dwyer がすべての利用できる方法を取りまとめて述べ、簡略ドーリットル法と呼ぶ方法を発表した。この方法は時間および労力を最小にすることができるから、ここで用いよう。やり方は例8-1に示す。

### 5. 例8-1 偏回帰係数の計算

次の資料は小麦の品質間の関係の研究で得られたものである。資料は小麦の139系統についてとつた。研究した特性はパンの容積 ( $Y$ )、蛋白質含有量 ( $x_1$ )、粉の収量 ( $x_2$ )、およびパン粉の色の記録 ( $x_3$ ) であつた。この例ではパン容積は従属変数として考えられている。これはパン粉の色との関係については完全に満足とは言えないというのは実際には、色に影響し相関があるのはパン容積の変動であり容積に影響するのは色の変動ではないからである。

$$\begin{aligned} \sum (x_1 y) &= 694.96 & \sum (y^2) &= 1142.45 & \sum (x_1 x_2) &= -20.834 \\ \sum (x_2 y) &= 79.648 & \sum (x_1^2) &= 947.17 & \sum (x_1 x_3) &= 27.276 \\ \sum (x_1 y) &= 407.00 & \sum (x_2^2) &= 414.23 & \sum (x_2 x_3) &= -35.400 \\ & & \sum (x_3^2) &= 603.00 & & \end{aligned}$$

8-1表 偏相関係数の計算

|    |         |           |          |         |    |             |             |             |          |
|----|---------|-----------|----------|---------|----|-------------|-------------|-------------|----------|
| 1  | 947.17  | -20.834   | 272.76   | 694.96  | 1  | $a_{11}$    | $a_{21}$    | $a_{31}$    | $a_{y1}$ |
| 2  |         | 414.23    | -354.00  | 79.648  | 2  |             | $a_{22}$    | $a_{32}$    | $a_{y2}$ |
| 3  |         |           | 630.00   | 407.00  | 3  |             |             | $a_{33}$    | $a_{y3}$ |
| 4  | 947.17  | -20.834   | 272.76   | 694.96  | 4  | $a_{11}$    | $a_{21}$    | $a_{31}$    | $a_{y1}$ |
| 5  | 1.0     | -0.021996 | 0.28797  | 0.73372 | 5  | 1.0         | $b_{21}$    | $b_{31}$    | $b_{y1}$ |
| 6  |         | 413.77    | -348.00  | 91.934  | 6  |             | $a_{22}$    | $a_{32}$    | $a_{y2}$ |
| 7  |         | 1.0       | -0.84105 | 0.22944 | 7  |             | 1.0         | $b_{32}$    | $b_{y2}$ |
| 8  |         |           | 258.77   | 286.72  | 8  |             |             | $a_{33}$    | $a_{y3}$ |
| 9  |         |           | 1.0      | 1.1080  | 9  |             |             | $b_{33}$    | $a_{y3}$ |
| 10 | 0.44019 | 1.1613    | 1.1080   |         | 10 | $b_{y1.23}$ | $b_{y2.13}$ | $b_{y3.12}$ |          |

$$\text{チェック } (947.17 \times 0.44019) + (-20.834 \times 1.1613) + (272.76 \times 1.1080) = 694.96$$

\* ドーリットルは最も簡略と思われる方法を発表している。それは附録に示す。



行

1~3 平方和および積和を記載

4 1行をそのまま移す。

5 4行の数字を $a_{11}$ で割る

6  $a_{22,1} = a_{22} - a_{21}b_{21}$   $a_{32,1} = a_{32} - a_{31}b_{21}$   $a_{y2,1} = a_{y2} - a_{y1}b_{21}$

7 6行の数字を $a_{22,1}$ で割る。

8  $a_{33,12} = a_{33} - a_{31}b_{31} - a_{32,1}b_{32,1}$   $a_{y3,12} = a_{y3} - a_{y1}b_{31} - a_{y2,1}b_{32,1}$

9 8行の数字を $a_{33,12}$ で割る。

10 逆 解

$b_{y3,12}$  は 9 行から移す

$b_{y2,13} = b_{y2,12} - b_{32,1}b_{y3,12}$

$b_{y1,23} = b_{y1} - b_{31}b_{y3,12} - b_{21}b_{y2,13}$

チェック  $a_{11}b_{y1,23} + a_{21}b_{y2,13} + a_{31}b_{y3,12} = a_{y1}$

回帰に対する平方和は

$b_1 \sum x_1 y + b_2 \sum x_2 y + b_3 \sum x_3 y$

である。ここに $b_{y1,23}$ は簡単に $b_2$ 等と書いてある数字で示すと

$$(0.44019 \times 694.96) + (1.1613 \times 79.648) + (1.1080 \times 407.00) \\ = 305.91 + 92.50 + 450.96 = 849.37$$

方程式 (5) に応ずる正規方程式は次の通りである。

$$b_1(947.17) + b_2(-20.834) + b_3(272.76) = 694.96$$

$$b_1(-20.834) + b_2(414.23) + b_3(-354.00) = 79.648$$

$$b_1(272.76) + b_2(-354.00) + b_3(630.00) = 407.00$$

これらの方程式を解くには、8-1表のように配列するとよい。表の右の部分は Dwyer と記号を同じにした記号の項で表わした既知の値である。表の下には計算を実行する行毎の説明がしてある。8-1表の計算を用いたら、重回帰の分散分析が可能である。

|         | SS      | DF        | MS     |
|---------|---------|-----------|--------|
| 回 帰     | 849.37  | 3 p       | 283.12 |
| 回帰からの偏差 | 293.08  | 135 n-p-1 | 2.1710 |
| 全 体     | 1142.45 | 138 n-1   |        |

Fの値は非常に大きいから、もうこれ以上検定をすすめる目標はない。重回帰係数の有意性およびおそらく信頼限界を研究することはより多く有益であろう。その方法は8節で述べる。

重回帰係数は次式で与えられ、その有意性検定は上の分散分析で行えることに注意しておこ5。

$$r_{y,123} = \sqrt{\frac{849.37}{1142.45}} = 0.8622$$

## 6. 従属変数の色々な組に対する重回帰係数の計算

研究者は一組以上の従属変数で重回帰分析を行うような場合がしばしばある。収量に対する高度、経度、緯度の影響について Fisher が与えた例では、これらは、与えられた組では同じであるが、収量は毎年変化するであろう。このような環境の下では、最小労力で重回帰係数の組を計算できるようにすることが望ましい。Fisher はこの計算をガウス乗数と呼ぶ統計量の組を用いてどのようにして行うことができるかを示した。後で述べるが、ガウス乗数には他にも大切な応用がある。

この方法は3つの独立変数の場合、下のように正規方程式を作つて計算例を示すことができる。研究者は3つ以上の独立変数の場合に方程式を書く基礎としてこれらを容易に用い得ることができよう。

$$\left. \begin{aligned} C_{11} \sum x_1^2 + C_{12} \sum x_1 x_2 + C_{13} \sum x_1 x_3 &= 0 \\ C_{11} \sum x_1 x_2 + C_{12} \sum x_2^2 + C_{13} \sum x_2 x_3 &= 0 \\ C_{11} \sum x_1 x_3 + C_{12} \sum x_2 x_3 + C_{13} \sum x_3^2 &= 0 \end{aligned} \right\} \quad (6)$$

$$\left. \begin{aligned} C_{21} \sum x_1^2 + C_{22} \sum x_1 x_2 + C_{23} \sum x_1 x_3 &= 0 \\ C_{21} \sum x_1 x_2 + C_{22} \sum x_2^2 + C_{23} \sum x_2 x_3 &= 0 \\ C_{21} \sum x_1 x_3 + C_{22} \sum x_2 x_3 + C_{23} \sum x_3^2 &= 0 \end{aligned} \right\} \quad (7)$$

$$\left. \begin{aligned} C_{31} \sum x_1^2 + C_{32} \sum x_1 x_2 + C_{33} \sum x_1 x_3 &= 0 \\ C_{31} \sum x_1 x_2 + C_{32} \sum x_2^2 + C_{33} \sum x_2 x_3 &= 0 \\ C_{31} \sum x_1 x_3 + C_{32} \sum x_2 x_3 + C_{33} \sum x_3^2 &= 0 \end{aligned} \right\} \quad (8)$$

これらの方程式の解は乗数のマトリックスを与える。

$$\begin{array}{ccc} C_{11} & C_{12} & C_{13} \\ C_{21} & C_{22} & C_{23} \\ C_{31} & C_{32} & C_{33} \end{array} \quad \text{ここに } C_{21} = C_{12} \text{ 等}$$

また、そのとき回帰係数は次から得ることができる。

$$\left. \begin{aligned} b_1 &= C_{11} \sum x_1 y + C_{12} \sum x_2 y + C_{13} \sum x_3 y \\ b_2 &= C_{12} \sum x_1 y + C_{22} \sum x_2 y + C_{23} \sum x_3 y \\ b_3 &= C_{13} \sum x_1 y + C_{23} \sum x_2 y + C_{33} \sum x_3 y \end{aligned} \right\} \quad (9)$$



これは、Yの一群に対して単に $\sum x_1y$ ,  $\sum x_2y$ ,  $\sum x_3y$ を計算すればよいことを示している。

例8-1と同じ回帰係数の組を求めるためにこの方法を例8-2に応用する。

#### 7. 例8-2 ガウス乗数を用いた偏回帰係数の計算

(6), (7)および(8)で与えた正規方程式の三組は、簡略ドウリツトル法の表の型で同時に解くことができる。その方法を8-2表に示す。8-1表と同じ計算規則を適用するが、参考のために説明付きの記号の表を与えておく。

回帰係数は方程式(9)を用いて得ることができる。

$$b_1 = (0.0013376 \times 694.96) + (-0.00082267 \times 79.648) + (-0.0010413 \times 407.00) = 0.4402$$

$$b_2 = (-0.00082267 \times 694.96) + (0.0051504 \times 79.648) + (0.0032502 \times 407.00) = 1.1613$$

$$b_3 = (-0.0010413 \times 694.96) + (0.0032502 \times 79.648) + (0.0038644 \times 407.00) = 1.1080$$

8 - 2 表

| ガウス乗数の計算 |        |           | 簡略ドウリツトル解   |             |            |
|----------|--------|-----------|-------------|-------------|------------|
| 1        | 947.17 | -20.834   | 272.76      | 1.0         | 0          |
| 2        | 414.23 | -354.00   | 0           | 0           | 1.0        |
| 3        |        | 630.00    | 0           | 0           | 1.0        |
| 4        | 947.17 | -20.834   | 272.76      | 1.0         | 0          |
| 5        | 1.0    | -0.021996 | 0.28797     | 0.0010558   | 0          |
| 6        | 413.77 | -348.00   | 0.021996    | 1.0         | 0          |
| 7        | 1.0    | -0.84105  | 0.000553160 | 0.0024168   | 0          |
| 8        |        | 258.77    | -0.26947    | 0.84105     | 1.0        |
| 9        |        | 1.0       | -0.0010413  | 0.0032502   | 0.0038644  |
| 10       |        |           | 0.0013376   | -0.00082267 | -0.0010413 |
| 11       |        |           |             | 0.0051504   | 0.0032502  |
| 12       |        |           |             |             | 0.0038644  |

チェック

$$(0.0013376 \times 947.17) + (-0.00082267 \times -20.834) + (-0.0010413 \times 272.76) = 1.1080$$

回帰係数の有意性は8節の方程式(10)で与えた標準誤差の公式を用いて、t検定で決定できる。

例8-1の分散分析から  $s^2 = 2.1710$  をうる。したがって  $S = 1.473$

$$s_{b1} = 1.473 \sqrt{0.0013376} = 0.0539$$

$$s_{b2} = 1.473 \sqrt{0.0051504} = 0.1057$$

$$s_{b3} = 1.473 \sqrt{0.0038644} = 0.0916$$

8 - 3 表

ガウス乗数の計算 - 簡略ドウリツトル解使用の説明

|    |          |          |          |            |            |            |
|----|----------|----------|----------|------------|------------|------------|
| 1  | $a_{11}$ | $a_{21}$ | $a_{31}$ | 1.0        | 0          | 0          |
| 2  |          | $a_{22}$ | $a_{32}$ | 0          | 1.0        | 0          |
| 3  |          |          | $a_{33}$ | 0          | 0          | 1.0        |
| 4  | $a_{11}$ | $a_{21}$ | $a_{31}$ | $d_{11}$   | 0          | 0          |
| 5  | 1        | $b_{21}$ | $b_{31}$ | $e_{11}$   | 0          | 0          |
| 6  |          | $a_{22}$ | $a_{32}$ | $d_{111}$  | $d_{121}$  | 0          |
| 7  |          | 1.0      | $b_{32}$ | $e_{111}$  | $e_{121}$  | 0          |
| 8  |          |          | $a_{33}$ | $d_{1112}$ | $d_{1212}$ | $d_{1312}$ |
| 9  |          |          | 1.0      | $e_{1112}$ | $e_{1212}$ | $e_{1312}$ |
| 10 |          |          |          | $C_{11}$   | $C_{12}$   | $C_{13}$   |
| 11 |          |          |          |            | $C_{22}$   | $C_{23}$   |
| 12 |          |          |          |            |            | $C_{33}$   |

行 説 明

1, 2, 3. 平方和および積和の記入

4 1行を移す

5 4行を $a_{11}$ で割る

6  $a_{22.1} = a_{22} - a_{21}b_{21}$ ,  $a_{32.1} = a_{32} - a_{31}b_{21}$ ,  $d_{11.1} = 0 - d_{11}b_{21}$  etc.

7 6行を $a_{22.1}$ で割る

8  $a_{33.12} = a_{33} - a_{31}b_{31} - a_{32}b_{32.1}$ ,  $d_{11.12} = 0 - d_{11}b_{31} - d_{11.1}b_{32.1}$  etc.  
 $d_{12.1}$ ,  $d_{13.12}$ に対しても同様

9 8行を $a_{33.12}$ で割る。

10, 11, 12.  $C_{11} = d_{11}e_{11} + d_{11.1}e_{11.1} + d_{11.12}e_{11.12}$

$C_{12} = d_{11.1}e_{12.1} + d_{11.12}e_{12.12}$

$C_{13} = d_{11.12}e_{13.12}$

$C_{22} = d_{12.1}e_{12.1} + d_{12.12}e_{12.12}$

$C_{23} = d_{12.12}e_{13.12}$

$C_{33} = d_{13.12}e_{13.12}$

#### 8. 偏回帰係数の有意性の検定

ガウス乗数を用いると、偏回帰係数の標準誤差の決定が容易となる。これらを次に示す。

$$s_{b1} = S \sqrt{C_{11}} \quad s_{b2} = S \sqrt{C_{22}} \quad s_{b3} = S \sqrt{C_{33}} \quad (10)$$



$$s = \sqrt{\frac{\sum (Y - Y_0)^2}{n - p - 1}} \quad (11)$$

pは独立変数の数で  $\sum (Y - Y_0)^2$  は次で容易に決定される。

$$\sum (Y - Y_0)^2 = \sum Y^2 - b_1 \sum x_1 Y - b_2 \sum x_2 Y - b_3 \sum x_3 Y \quad (12)$$

2つの回帰係数間の差の標準誤差は、またガウス乗数によつて表わすことができる。すなわち、 $b_{y12 \dots n}$  と  $y_{21 \dots n}$  との間の差に対する標準誤差は

$$s \sqrt{C_{11} + C_{22} - 2C_{12}} \quad (13)$$

もつと進んだ応用は、 $Y_{01}$  のようなYの推定値の標準誤差を決定することである。3つの独立変数に対して、 $Y_0$  の標準誤差は次で与えられる。

$$SE = s \sqrt{\frac{1}{n} + C_{11}x_1^2 + C_{22}x_2^2 + C_{33}x_3^2 + 2C_{12}x_1x_2 + 2C_{13}x_1x_3 + 2C_{23}x_2x_3} \quad (14)$$

ここに  $x_1, x_2$  および  $x_3$  は、それからYが推定される特別な値をもつ平均値からの偏差である。再び6章と同じようにして、推定値Yは  $x_1, x_2$  および  $x_3$  の与えられた相に対してYの推定量であることを指摘しなければならぬ。

## 9. 変量の削除

回帰または相関分析を終つてから、ある変量を除いたらどんな結果が得られるかを研究したい。ガウス乗数を用いると簡単な方法が利用できる。下に与える公式によつて新しいマトリックスを計算することが必要である。これらの公式は一般項で表わすとすこし面倒であるから、一般の場合に適用できる特別な場合を与えてある。 $rC_{11}$  は変量が減少したマトリックスでの値を表わす。

変量  $Y, X_1, X_2, X_3 : X_3$  を削除

$$rC_{11} = C_{11} - \frac{C_{13}^2}{C_{33}} \quad rC_{12} = C_{12} - \frac{C_{13}C_{23}}{C_{33}} \quad rC_{22} = C_{22} - \frac{C_{23}^2}{C_{33}} \quad (15)$$

変量  $Y, X_1, X_2, X_3 : X_2$  を削除

$$rC_{11} = C_{11} - \frac{C_{12}^2}{C_{22}} \quad rC_{13} = C_{13} - \frac{C_{12}C_{23}}{C_{22}} \quad rC_{33} = C_{33} - \frac{C_{23}^2}{C_{22}} \quad (16)$$

公式8-2から  $X_2$  を削除するに必要な計算は例8-3に与えてある。同じ公式は相関分析から得るガウス乗数のマトリックスにも、適用できる。

10. 例8-3  $Y, X_1, X_2$  および  $X_3$  に対する回帰分析からの  $X_2$  の削除

$$rC_{11} = 0.0013376 - \frac{(-0.00082267)^2}{0.0051504} = 0.0012062$$

$$rC_{13} = -0.0010413 - \frac{(-0.00082267 \times 0.0032502)}{0.0051504} = -0.00052215$$

$$rC_{33} = 0.0038644 - \frac{(0.0032502)^2}{0.0051504} = 0.0018133$$

チエツク

$$(0.0012062 \times 947.17) + (-0.00052215 \times 272.76) = 1.00005$$

$$b_1 = C_{11} \sum x_1 Y + C_{13} \sum x_3 Y \\ = 0.0012062 \times 694.96 + (-0.00052215 \times 407.00)$$

$$b_3 = C_{13} \sum x_1 Y + C_{33} \sum x_3 Y \\ = (-0.00052215 \times 694.96) + (0.0018133 \times 407.00) \\ = 0.37514$$

チエツク

$$(0.62574 \times 947.17) + (0.37514 \times 272.76) = 695.00$$

回帰の平方和  $= b_1 \sum x_1 Y + b_3 \sum x_3 Y$

$$= (0.62574 \times 694.96) + (0.37514 \times 407.00) = 587.55$$

$$r_{y,13} = \sqrt{\frac{587.55}{1142.45}} = 0.7171 \quad \text{等}$$

## 11. 偏相関と重相関

上の3節によれば、2つの独立変数  $X_1, X_2$  に対して、簡単な回帰方程式

$$Y_0 = b_{y2} X_2 \quad \text{および} \quad X_{10} = b_{12} X_2$$

は  $X_2$  の各値に対するYの推定値の系列および  $X_1$  の推定値の同様な系列を与える。各回帰線からのYおよび  $X_1$  の実際値の偏差は  $X_2$  と独立である。これを公式で表わすと、

$$u = Y - b_{y2} X_2, \quad v = X_1 - b_{12} X_2$$

で、 $r_{uv}$  は偏回帰係数であり、記号的には  $r_{y12}$  で表わすことができる。

一般に、相関分析では、回帰分析ほどは独立変数と従属変数の間の区別をしない。したがつて、我々は  $X_1, X_2, X_3$  等のような変量に関心をもつが、ここでの回帰問題との関連性は、 $X_1$  はYに依存する変量としてとられているということである。したがつて正規方程式に対して、4つの変量があり3つの未知数がある。

$$\begin{aligned} b'_{12,34} + r_{23} b'_{3,24} + r_{24} b'_{14,23} &= r_{12} \\ r_{23} b'_{12,34} + b'_{13,24} + r_{34} b'_{14,23} &= r_{13} \\ r_{23} b'_{12,34} + r_{34} b'_{13,24} + b'_{14,23} &= r_{14} \end{aligned} \quad (17)$$

ここに添字は  $1=X_1, 2=X_2, 3=X_3, 4=X_4$  を表わし、回帰分析とあわせれば  $1=Y, 2=X_1, 3=X_2, 4=X_3$  である。

これ等の方程式は (5) から導かれる。回帰係数は、実際には標準的回帰係数であるから、グラフシユをつけて区別している。いゝかえれば、含まれる変量の標準偏差で標準化されている。



$$\left. \begin{aligned} b'_{1234} &= b_{1234} \sqrt{\frac{\sum x_2^2}{\sum x_1^2}} \\ b'_{1324} &= b_{1324} \sqrt{\frac{\sum x_3^2}{\sum x_1^2}} \end{aligned} \right\}$$

(18)

偏回帰係数の計算法は回帰係数の計算法と似ている。基礎的な方法は標準の回帰係数に対する正規方程式を解き、次の関係から偏相関数を得ることである。

$$r_{1234 \dots p} = \sqrt{b'_{1234 \dots p} b'_{2134 \dots p}}$$

(17) の正規方程式から明らかであるが、 $b'_{1234}$ 、 $b'_{1324}$  および  $b'_{1423}$  を与える一組の正規方程式の解によつてはどの相関係数も計算はできないだろう。何故ならば  $r_{1234}$ 、 $r_{1324}$  および  $r_{1423}$  を求めるには上の偏回帰係数の外に、 $b'_{2134}$ 、 $b'_{3124}$  および  $b'_{4123}$  が必要であるから。これ等の偏回帰係数は、方程式を新しく配列しなおし、再び解いて、すべての標準の回帰係数の可能な値が得られるまでこのような再配列をつづけてゆけば求めることができるが、 $\Gamma$ マトリックスの逆行列である乗数のマトリックスを計算すると、すべての問題は極めて簡単になる。すなわち、下の左側の  $\Gamma$ マトリックスでの計算は右側に示す  $C'$  マトリックスを考へる。

$$\begin{array}{cccc|cccc} r_{11} & r_{12} & r_{13} & r_{14} & C'_{11} & C'_{12} & C'_{13} & C'_{14} \\ r_{12} & r_{22} & r_{23} & r_{24} & C'_{12} & C'_{22} & C'_{23} & C'_{24} \\ r_{13} & r_{23} & r_{33} & r_{34} & C'_{13} & C'_{23} & C'_{33} & C'_{34} \\ r_{14} & r_{24} & r_{34} & r_{44} & C'_{14} & C'_{24} & C'_{34} & C'_{44} \end{array}$$

(19)

$C'$  マトリックスからいかなる偏または相関係数も計算することができる。すなわち、次のことが示される。

$$b_{1234} = - \frac{C'_{12}}{C'_{11}} \quad b_{2134} = - \frac{C'_{12}}{C'_{22}} \quad \text{等}$$

これから次のものを得る。

$$r_{1234} = - \frac{C'_{12}}{\sqrt{C'_{11} C'_{22}}} \quad \text{等}$$

(20)

$$r_{2314} = \frac{C'_{23}}{\sqrt{C'_{22} C'_{33}}}$$

偏回帰係数の有意性は次に示す  $t$  によつて検定する。

$$t = \frac{r}{\sqrt{1-r^2}} \sqrt{n-p-1}$$

(21)

ここに  $r$  は検定する係数、 $n$  は標本数、 $p$  は独立変数の数である。

例 8-4 で、6 変量をもつ問題に対して、乗数のマトリックスを得るのに簡略ドーリツトル法を用いる。

## 12. 例 8-4. 偏相関係数および重相関係数の計算

8-4 表のすべての相関係数は粉の収量および品質の小麦の物理的性質の影響についての研究から引用したものである。

8-4 表 6 つの変量間の関係に対する単純相関係数

|   | 2       | 3       | 4       | 5       | 6       |
|---|---------|---------|---------|---------|---------|
| 1 | -0.4589 | -0.5612 | -0.3947 | -0.3123 | 0.6412  |
| 2 |         | 0.3114  | 0.0429  | 0.2861  | -0.3190 |
| 3 |         |         | -0.0655 | 0.1467  | -0.4462 |
| 4 |         |         |         | 0.1882  | -0.3511 |
| 5 |         |         |         |         | -0.3092 |

- |                   |               |
|-------------------|---------------|
| 1. 粉の収量           | 4. 未熟の割合      |
| 2. ふすまの割合         | 5. 緑の穀粒の割合    |
| 3. ひどく霜白しているものの割合 | 6. プツシエル当りの重量 |

8-5 表および 8-6 表に行毎の説明および詳細計算のやり方を示している。もつと簡単な方法は附録にのせてある。(8.5 表および 8.6 表次頁挿入)

## 13. 重相関係数の解釈と有意性

重回帰方程式は

$$y_o = b_1 x_1 + b_2 x_2 + \dots + b_p x_p$$

であることはすでに指摘したところである。ことに従属変数は  $Y$  で、独立変数は  $X_1, X_2, \dots, X_p$  である。この方程式の予測力が大きくなればなるほど (項数が増せば増すほど)  $y$  の実測値と予測値との間は一貫してくる。したがつて、方程式の予測力は簡単な相関  $r_{yyo}$  によつて測定しなければならない。実際には、これは重相関係数  $r_{Y.12}$  で、明らかにその最小値は 0 であり、最大値は +1 である。それは二つの変量間の単純相関の場合のように -1 から +1 までの範囲ではない。

$Y$  の平方和の分析においては、単純相関の場合のように、

$$\sum y^2 = \sum (Y - Y_o)^2 + \sum (Y_o - \bar{Y})^2$$

すなわち

$$\sum y^2 = \sum y_d^2 + \sum y_o^2$$

であるが、この場合に対応する自由度は

$$n-1 = (n-p-1) + p$$



8-6表 個および重相関係数の計算\*

|    |           |           |           |           |           |           |                                                                                                               |           |           |           |           |           |
|----|-----------|-----------|-----------|-----------|-----------|-----------|---------------------------------------------------------------------------------------------------------------|-----------|-----------|-----------|-----------|-----------|
| 1  | $a_{11}$  | $a_{21}$  | $a_{31}$  | $a_{41}$  | $a_{51}$  | $a_{61}$  | 1                                                                                                             | 0         | 0         | 0         | 0         | 0         |
| 2  |           | $a_{22}$  | $a_{32}$  | $a_{42}$  | $a_{52}$  | $a_{62}$  | 0                                                                                                             | 1         | 0         | 0         | 0         | 0         |
| 3  |           |           | $a_{33}$  | $a_{43}$  | $a_{53}$  | $a_{63}$  | 0                                                                                                             | 0         | 1         | 0         | 0         | 0         |
| 4  |           |           |           | $a_{44}$  | $a_{54}$  | $a_{64}$  | 0                                                                                                             | 0         | 0         | 1         | 0         | 0         |
| 5  |           |           |           |           | $a_{55}$  | $a_{65}$  | 0                                                                                                             | 0         | 0         | 0         | 1         | 0         |
| 6  |           |           |           |           |           | $a_{66}$  | 0                                                                                                             | 0         | 0         | 0         | 0         | 1         |
| 7  | $a_{11}$  | $a_{21}$  | $a_{31}$  | $a_{41}$  | $a_{51}$  | $a_{61}$  | $d_{11}$                                                                                                      | 0         |           |           |           |           |
| 8  | 1         | $b_{21}$  | $b_{31}$  | $b_{41}$  | $b_{51}$  | $b_{61}$  | $e_{11}$                                                                                                      | 0         |           |           |           |           |
| 9  |           | $a_{22}$  | $a_{32}$  | $a_{42}$  | $a_{52}$  | $a_{62}$  | $d_{11}$                                                                                                      | $d_{12}$  | 0         |           |           |           |
| 10 |           | 1         | $b_{32}$  | $b_{42}$  | $b_{52}$  | $b_{62}$  | $e_{11}$                                                                                                      | $e_{12}$  | 0         |           |           |           |
| 11 |           |           | $a_{33}$  | $a_{43}$  | $a_{53}$  | $a_{63}$  | $d_{11}$                                                                                                      | $d_{12}$  | $d_{13}$  | 0         |           |           |
| 12 |           |           | 1         | $b_{43}$  | $b_{53}$  | $b_{63}$  | $e_{11}$                                                                                                      | $e_{12}$  | $e_{13}$  | 0         |           |           |
| 13 |           |           |           | $a_{44}$  | $a_{54}$  | $a_{64}$  | $d_{11}$                                                                                                      | $d_{12}$  | $d_{13}$  | $d_{14}$  | 0         |           |
| 14 |           |           |           | 1         | $b_{54}$  | $b_{64}$  | $e_{11}$                                                                                                      | $e_{12}$  | $e_{13}$  | $e_{14}$  | 0         |           |
| 15 |           |           |           |           | $a_{55}$  | $a_{65}$  | $d_{11}$                                                                                                      | $d_{12}$  | $d_{13}$  | $d_{14}$  | $d_{15}$  | 0         |
| 16 |           |           |           |           | 1         | $b_{65}$  | $e_{11}$                                                                                                      | $e_{12}$  | $e_{13}$  | $e_{14}$  | $e_{15}$  | 0         |
| 17 |           |           |           |           |           | $a_{66}$  | $d_{11}$                                                                                                      | $d_{12}$  | $d_{13}$  | $d_{14}$  | $d_{15}$  | $d_{16}$  |
| 18 |           |           |           |           |           | $b_{66}$  | $e_{11}$                                                                                                      | $e_{12}$  | $e_{13}$  | $e_{14}$  | $e_{15}$  | $e_{16}$  |
| 19 | $b'_{61}$ | $b'_{62}$ | $b'_{63}$ | $b'_{64}$ | $b'_{65}$ |           | $C'_{11}$                                                                                                     | $C'_{12}$ | $C'_{13}$ | $C'_{14}$ | $C'_{15}$ | $C'_{16}$ |
| 20 | $b'_{51}$ | $b'_{52}$ | $b'_{53}$ | $b'_{54}$ |           | $b'_{56}$ |                                                                                                               | $C'_{22}$ | $C'_{23}$ | $C'_{24}$ | $C'_{25}$ | $C'_{26}$ |
| 21 | $b'_{41}$ | $b'_{42}$ | $b'_{43}$ |           | $b'_{45}$ | $b'_{46}$ |                                                                                                               |           | $C'_{33}$ | $C'_{34}$ | $C'_{35}$ | $C'_{36}$ |
| 22 | $b'_{31}$ | $b'_{32}$ |           | $b'_{34}$ | $b'_{35}$ | $b'_{36}$ |                                                                                                               |           |           | $C'_{44}$ | $C'_{45}$ | $C'_{46}$ |
| 23 | $b'_{21}$ |           | $b'_{23}$ | $b'_{24}$ | $b'_{25}$ | $b'_{26}$ |                                                                                                               |           |           |           | $C'_{55}$ | $C'_{56}$ |
| 24 |           | $b'_{12}$ | $b'_{13}$ | $b'_{14}$ | $b'_{15}$ | $b'_{16}$ |                                                                                                               |           |           |           |           | $C'_{66}$ |
| 25 |           |           |           |           |           |           | $r_{11}$                                                                                                      | $r_{12}$  | $r_{13}$  | $r_{14}$  | $r_{15}$  | $r_{16}$  |
| 26 |           |           |           |           |           |           |                                                                                                               | $r_{22}$  | $r_{23}$  | $r_{24}$  | $r_{25}$  | $r_{26}$  |
| 27 |           |           |           |           |           |           | $r_{1,2,3,4,5,6}^2 = a_{61}b_{61} + a_{62}b_{62} + a_{63}b_{63} + a_{64}b_{64} + a_{65}b_{65} + a_{66}b_{66}$ |           | $r_{33}$  | $r_{34}$  | $r_{35}$  | $r_{36}$  |
| 28 |           |           |           |           |           |           |                                                                                                               |           |           | $r_{44}$  | $r_{45}$  | $r_{46}$  |
| 29 |           |           |           |           |           |           |                                                                                                               |           |           |           | $r_{55}$  | $r_{56}$  |
| 30 |           |           |           |           |           |           |                                                                                                               |           |           |           |           | $r_{66}$  |

\* 添字は簡略にしている。すなわち  $b_{65,4} = b_{65,12,34}$

8-5表 個および重相関係数に対する簡略 F-ラット

|    |         |         |         |         |         |         |         |         |         |         |        |         |
|----|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|--------|---------|
| 1  | 1.0000  | -0.4589 | -0.5612 | -0.3947 | -0.3123 | 0.6412  | 1.0     | 0       | 0       | 0       | 0      | 0       |
| 2  |         | 1.0000  | 0.3114  | 0.0429  | 0.2861  | -0.3190 | 0       | 1.0     | 0       | 0       | 0      | 0       |
| 3  |         |         | 1.0000  | -0.0655 | 0.1467  | -0.4462 | 0       | 0       | 1.0     | 0       | 0      | 0       |
| 4  |         |         |         | 1.0000  | 0.1892  | -0.3511 | 0       | 0       | 0       | 1.0     | 0      | 0       |
| 5  |         |         |         |         | 1.0000  | -0.3032 | 0       | 0       | 0       | 0       | 1.0    | 0       |
| 6  |         |         |         |         |         | 1.0000  | 0       | 0       | 0       | 0       | 0      | 1.0     |
| 7  | 1.0     | -0.4589 | -0.5612 | -0.3947 | -0.3123 | 0.6412  | 1.0     | 0       | 0       | 0       | 0      | 0       |
| 8  | 1.0     | -0.4589 | -0.5612 | -0.3947 | -0.3123 | 0.6412  | 1.0     | 0       | 0       | 0       | 0      | 0       |
| 9  |         | 0.7894  | 0.0539  | -0.1332 | 0.1428  | -0.0248 | 0.4589  | 1.0     | 0       | 0       | 0      | 0       |
| 10 |         | 1.0     | 0.0683  | -0.1751 | 0.1809  | -0.0314 | 0.5813  | 1.2668  | 0       | 0       | 0      | 0       |
| 11 |         |         | 0.6814  | -0.2776 | -0.0383 | -0.0847 | 0.5299  | -0.0682 | 1.0     | 0       | 0      | 0       |
| 12 |         |         | 1.0     | -0.4074 | -0.0562 | -0.1243 | 0.7777  | -0.1001 | 1.4676  | 0       | 0      | 0       |
| 13 |         |         |         | 0.7069  | 0.0743  | -0.1369 | 0.6909  | 0.1473  | 0.4074  | 1.0     | 0      | 0       |
| 14 |         |         |         | 1.0     | 0.1051  | -0.1937 | 0.9774  | 0.3004  | 0.5763  | 1.4146  | 0      | 0       |
| 15 |         |         |         |         | 0.8667  | -0.0948 | 0.1865  | -0.3002 | 0.0134  | -0.1051 | 1.0    | 0       |
| 16 |         |         |         |         | 1.0     | -0.1094 | 0.2152  | -0.2310 | 0.0155  | -0.1213 | 1.1538 | 0       |
| 17 |         |         |         |         |         | 0.5407  | -0.4067 | 0.0296  | 0.2047  | 0.1822  | 0.1094 | 1.0     |
| 18 |         |         |         |         |         | 1.0     | -0.7522 | 0.0547  | 0.3766  | 0.3370  | 0.2023 | 1.8494  |
| 19 | 0.4067  | -0.0296 | -0.2047 | -0.1822 | -0.1094 | -0.1720 | 2.7002  | 0.6069  | -1.0248 | 0.8177  | 0.1323 | -0.7522 |
| 20 | -0.1130 | 0.1913  | -0.0494 | 0.0719  | 0.0567  | -0.2264 | 1.3522  | -0.0071 | 0.2435  | -0.2260 | 0.0547 | 0.0547  |
| 21 | -0.5192 | -0.1630 | -0.4324 | 0.0567  | -0.0330 | -0.2127 | 1.7801  | 0.6437  | 0.0569  | 0.0569  | 0.0569 | 0.0569  |
| 22 | -0.5757 | 0.0040  | -0.3516 | -0.0330 | -0.0330 | -0.2127 | 1.4988  | -0.0844 | 0.0569  | 0.0569  | 0.0569 | 0.0569  |
| 23 | -0.4488 | 0.0052  | -0.1794 | 0.1664  | -0.0404 | -0.2786 | 1.1759  | 0.0569  | 0.0569  | 0.0569  | 0.0569 | 0.0569  |
| 24 | -0.2248 | -0.3795 | -0.3038 | -0.0492 | 0.2786  |         |         |         |         |         |        | 1.8494  |
| 25 |         |         |         |         |         | -0.318  | -0.457  | -0.406  | -0.074  | 0.337   | 0.337  | 0.337   |
| 26 |         |         |         |         |         | 0.005   | -0.171  | 0.178   | -0.034  | -0.034  | -0.034 | -0.034  |
| 27 |         |         |         |         |         | -0.395  | -0.039  | -0.039  | -0.209  | -0.209  | -0.209 | -0.209  |
| 28 |         |         |         |         |         | 0.064   | -0.203  | -0.203  | -0.203  | -0.203  | -0.203 | -0.203  |
| 29 |         |         |         |         |         | -0.137  |         |         |         |         |        | -0.137  |
| 30 |         |         |         |         |         |         |         |         |         |         |        |         |

$$r_{1,2,3,4,5,6}^2 = 0.678$$



である。ここに  $p$  は独立変数の数である。重回方程式または重相関係数の有意性の検定は、次によつて計算すればよい。

$$F = \frac{\sum y_0^2}{p} \times \frac{n-p-1}{\sum y_d^2} \quad (25)$$

そして  $F$  表で自由度  $n-p-1$  のところを読む。

相関のみがわかっている場合には、 $F$  は次から計算できる。

$$F = \frac{r_{12 \dots p}^2}{1 - r_{12 \dots p}^2} \left( \frac{n-p-1}{p} \right) \quad (26)$$

これは

$$r_{12 \dots p}^2 = \frac{\sum y_0^2}{\sum y^2} \quad (27)$$

および

$$1 - r_{12 \dots p}^2 = \frac{\sum y_d^2}{\sum y^2} \quad (28)$$

であるからである。(25) 式を

$$r_{12 \dots p}^2 = 1 - \frac{\sum y_d^2}{\sum y^2} \quad (29)$$

の形にすると、重相関係数の解釈に役立つ。回帰からの偏差が 0 に近づくにつれて、係数は +1 に近づく。

#### 14. 特別な応用

分散分析を行うと、重相関係数計算の際に得た附加的な情報の、有意性の検定ができる。この原則は穀類化学の実験問題で、Geddes および Goulden が応用した。まづ第一に小麦粉のパン容積と蛋白質の百分率との間の相関を決定した。最近の研究において、蛋白質は消化性のものと、非消化性のものの二つに分けた。そしてこれら二つの部分を変数としてパン容積に対する相互効果の重相関を計算した。もしパン容積に対して、二種類の蛋白質の割合が重要な影響を与えるとすれば、重相関係数は、全蛋白質およびパン容積に対する単純相関よりも有意に大きくなる筈である。したがって、粉の品質予測の目的に対して、蛋白質の全含有量の他に消化性および非消化性の蛋白質の量を知ることの実際的な意味は二つの相関を比較する方法によつて定められる。

もし、 $X_1$  がパン容積、 $X_2$  が消化性蛋白質、 $X_3$  が非消化性蛋白質および  $X_p$  が全蛋白質を表わすとすれば、それに応ずる単純および重相関係数は  $r_{1p}$  および  $r_{123}$  である。勿論、全蛋白質は二つの部分の和 ( $X_2 + X_3$ ) である。

これらの相関が 20 対の値から決定されるとすれば、回帰函数からの偏差を表わす平方和は、 $\sum (1 - r_{1p}^2)$  および  $(1 - r_{123}^2)$  に比例し、それに応ずる自由度は 18 および 17 である。重回帰の場合と同様に、 $X_1$  を決定するために、より多くの変数を導入することによる影響は回帰函数からの偏差平方和が減少することであるが、導入した各附加的な変数に対して自由度 1 が失われるから、平方和の減少が自由度の損失に比例してそれ以上でなければ精度の利益はない。したがって分析は次のようにして行う。

|                                 | SS                                 | DF | MS  |
|---------------------------------|------------------------------------|----|-----|
| $X_1$ への $X_p$ の回帰、偏差           | $1 - r_{1p}^2$                     | 18 |     |
| $X_1$ への $X_2$ および $X_3$ の回帰、偏差 | $1 - r_{123}^2$                    | 17 | (1) |
| 附加的な自由度                         | $(1 - r_{1p}^2) - (1 - r_{123}^2)$ | 1  | (2) |

平均平方 (1) および (2) に  $F$  検定を適用すると、誤差として (1) をとつて、他の変数を加えることによる情報の利益の有意性を決定することができる。

1962 の北方小麦からの 20 の粉の系列に対する、ある実験的な実験において、 $r_{1p} = 0.511$  および  $r_{123} = 0.732$  であつた。分析は次の通りである。

| SS              | DF       | MS | F       | 1%点        |
|-----------------|----------|----|---------|------------|
| $1 - r_{1p}^2$  | 0.738879 | 18 |         |            |
| $1 - r_{123}^2$ | 0.464176 | 17 | 0.02730 |            |
| 差               | 0.274703 | 1  | 0.2747  | 10.06 8.40 |

#### 説明

標準の偏回帰係数および偏相関係数を見出す逆マトリックスの手引 (8-6 表の説明)

- 1-6 正規方程式の既知の値を書く
- 7 1 行をそのまま移す
- 8 各項を  $a_{11}$  で割つたもの
- 9  $a_{22.1} = a_{22} - a_{21}b_{21}$ ,  $a_{32.1} = a_{32} - a_{31}b_{21}$ ,  $a_{42.1} = a_{42} - a_{41}b_{21}$ , .....  
 $d_{12.1} = 1 - 0 \times b_{21}$
- 10 9 行を  $a_{22.1}$  で割つたもの
- 11  $a_{33.2} = a_{33} - a_{31}b_{31} - a_{32.1}b_{32.1}$ ,  $a_{43.2} = a_{43} - a_{41}b_{31} - a_{42.1}b_{32.1}$ , .....  
 $d_{13.2} = 1 - 0 \times b_{32.1}$
- 12 11 行を  $a_{33.2}$  で割つたもの



## 第 9 章 共分散分析

### 1. 共分散分析の原理。

単一変量の分散分析では、変動を色々な成分に分割することができる。二元配置の場合には分散成分を行、列および残差に分けることができる。乱塊法の実験では、プロット、処理および誤差に起因する分散成分に分けることができる。同じことが二つの変量の相関変動すなわち共変動にも成立つ。共分散の影響を分類するやり方は共分散分析法として知られているものである。

実際の値で考えるために、我々は変量が穀物プロットからの穀物およびわらの収量であると仮定しよう。r個の乱塊プロットで行ったn個の処理を仮定して、Yを収量、Xはわらの量とすると、次の値の組を得る。

|              |       | 処 理                         |                             |       |                             | ブロック平均            |
|--------------|-------|-----------------------------|-----------------------------|-------|-----------------------------|-------------------|
|              |       | 1                           | 2                           | ..... | n                           |                   |
| プロ<br>ッ<br>ク | 1     | $X_{11} Y_{11}$             | $X_{12} Y_{12}$             | ..... | $X_{1n} Y_{1n}$             | $X_{b1} Y_{b1}$   |
|              | 2     | $X_{21} Y_{21}$             | $X_{22} Y_{22}$             | ..... | $X_{2n} Y_{2n}$             | $X_{b2} Y_{b2}$   |
|              | ..... | .....                       | .....                       | ..... | .....                       | .....             |
|              | r     | $X_{r1} Y_{r1}$             | $X_{r2} Y_{r2}$             | ..... | $X_{rn} Y_{rn}$             | $X_{br} Y_{br}$   |
| 処理平均         |       | $\bar{X}_{t1} \bar{Y}_{t1}$ | $\bar{X}_{t2} \bar{Y}_{t2}$ | ..... | $\bar{X}_{tn} \bar{Y}_{tn}$ | $\bar{X} \bar{Y}$ |
|              |       |                             |                             |       |                             | 総平均               |

XおよびYのn個の処理平均があり、XおよびYのr個のブロック平均があるが、これらの列にされた値の各々の組はXへのYの回帰係数を決定できることに注意しよう。また、総平均からのXおよびYの個々の偏差を表すrn個の値の組があり、これからブロックおよび処理平均による効果を差引く。

明らかに、これらは次の通りである。

$$\begin{aligned} & (X_{11} - \bar{X}_{b1} - \bar{X}_{t1} + \bar{X})(Y_{11} - \bar{Y}_{b1} - \bar{Y}_{t1} + \bar{Y}) \\ & (X_{12} - \bar{X}_{b1} - \bar{X}_{t2} + \bar{X})(Y_{12} - \bar{Y}_{b1} - \bar{Y}_{t2} + \bar{Y}) \\ & \vdots \\ & (X_{rn} - \bar{X}_{br} - \bar{X}_{tn} + \bar{X})(Y_{rn} - \bar{Y}_{br} - \bar{Y}_{tn} + \bar{Y}) \end{aligned}$$

この組から、またXへのYの回帰係数が作れる。

回帰の効果进行分类する問題に関連して、まづ第一に問われることは次の通りである。プロッ

$$13 \quad a_{44} = a_{44} - a_{41}b_{41} - a_{42}b_{42} - a_{43}b_{43}, \quad a_{54} = a_{54} - a_{51}b_{41} - a_{52}b_{42} - a_{53}b_{43}, \quad \dots \quad d_{14} = 1 - 0 \times b_{53}$$

14~18 次に概略述べた組織的な方法に従って計算を完成する。

$$\begin{aligned} 19 \quad C'_{11} &= d_{11}e_{11} + d_{11.1}e_{11.1} + d_{11.2}e_{11.2} + \dots + d_{11.5}e_{11.5} \\ C'_{12} &= d_{11.1}e_{12.1} + d_{11.2}e_{12.2} + \dots + d_{11.5}e_{12.5} \\ C'_{13} &= d_{11.2}e_{13.2} + \dots + d_{11.5}e_{13.5} \\ &\text{等} \end{aligned}$$

(この段階では回帰係数の計算は除く)

$$\begin{aligned} 20 \quad C'_{22} &= d_{12.1}e_{12.1} + d_{12.2}e_{12.2} + d_{12.3}e_{12.3} + \dots + d_{12.5}e_{12.5} \\ C'_{23} &= d_{12.2}e_{13.2} + d_{12.3}e_{13.3} + \dots + d_{12.5}e_{13.5} \\ &\text{等} \end{aligned}$$

$$21 \quad C'_{33} = d_{13.2}e_{13.2} + d_{13.3}e_{13.3} + \dots + d_{13.5}e_{13.5} \\ \text{等}$$

22~24 乗数の計算を完成する。

$$25 \quad r_{12,3456} = \frac{-C'_{12}}{\sqrt{C'_{11} C'_{22}}} \quad r_{13,2456} = \frac{-C'_{13}}{\sqrt{C'_{11} C'_{33}}} \quad \text{等}$$

$$26 \quad r_{23,1456} = \frac{-C'_{23}}{C'_{22} C'_{33}} \quad r_{24,1356} = \frac{-C'_{24}}{C'_{22} C'_{44}} \quad \text{等}$$

27~30 個相関係数の計算の完成

$$19 \text{ 行にもどつて } b'_{61} = \frac{-C'_{16}}{C'_{66}}, \quad b'_{62} = \frac{-C'_{26}}{C'_{66}}, \quad \dots, \quad b'_{16} = \frac{-C'_{16}}{C'_{11}}$$

これらの計算に対しては、最初に  $C'_{11}, C'_{22}, \dots, C'_{66}$  の逆数を作つて、これ等を組織的に数値に乗じて必要な商を得る。  $r_{12,3456} = \sqrt{b'_{12} \cdot b'_{21}}$ , 等 から個相関係数の計算をくりかえすことによつて完全なチェックができる。



ク平均を $\bar{y}_0$ 、処理平均を $\bar{y}_i$ 、残差すなわち誤差項を $y_{ij}$ で表わした場合、回帰係数の期待値は何であるか？我々が標本を抽出する大母集団では、 $Y$ が $X$ の各水準に対して正規に分布し、 $X$ への $Y$ の回帰係数が $\beta_{yx}$ であるとしよう。そのとき、もし $X$ および $Y$ の値がこの母集団から無作為に抽出され、プロットおよび処理に対して二元配置を行ふとすれば、 $\bar{y}_0$ 、 $\bar{y}_i$ および $y_{ij}$ の期待値は $\beta_{yx}$ であることを示すことができる。すなわち、もし実際の実験での $\bar{y}_i$ がいかなる場合でも期待値 $\beta_{yx}$ をもつ $\bar{y}_0$ より、かなり大であるとすれば、結論としては、資料は回帰の効果に関して不斉である、すなわち、いゝかえれば、処理に対する回帰の効果は、誤差項で表わされた母集団回帰の単なる無作為抽出からは生じないということである。

したがって、回帰および相関の効果をはけることは明らかに共分散分析の応用である。全効果を測定する回帰係数の明確な解釈は不可能であるし、またもし不斉性が存在すれば、妥当な有意性の検定も提供することはできない。6つの場処の各々配置した20プロットに対して抽出した麦稈の量への稈長の収量の全回帰があると仮定しよう。平方和および自由度は次のように配列できる。

| SS      | DF  |
|---------|-----|
| 回 帰     | 1   |
| 回帰からの偏差 | 118 |
| 全 体     | 119 |

そして、回帰は非常に有意であるように見えるが、もし場処間および場処内の成分に分けると、次の二つの分析ができる。

| 場 処 間   | DF | 場 処 内   | DF  |
|---------|----|---------|-----|
| 回 帰     | 1  | 回 帰     | 1   |
| 回帰からの偏差 | 4  | 回帰からの偏差 | 113 |
| 全 体     | 5  | 全 体     | 114 |

そして、もし全回帰の大部分が場処平均によるものであるとすれば、場処間回帰は自由度がわずか4つだけに基いている誤差に対して検定しなければならぬから、いづれの場合においても有意性がないであろうということは余くありそうなことである。

上述の型の全共分散は、時にはみせかけの効果を含むといわれることがある。これは共分散分析においては考慮に入れてあつて、いわゆる、みせかけの効果が除かれるばかりでなく、また全体の中の明らかな成分として評価されるのである。

## 2. 誤差のコントロールへの共分散法の応用。

共分散分析の最も重要な応用の一つは、実験を通じて無作為に生じ、反復によつても注意することができない誤差をコントロールすることである。例えば、トウモロコシの穂大根のような

作物に対して、プロット当り植物の数は、実験を通じて無作為に生じ、その数が単一プロットの収量に影響する限り、実験誤差を増す。収量が直接植物の数に比例するということを基として、収量を補正することが実際にはよく行われるが、収量がプロット当りの植物の数にいつも比例しているようなことは稀であり、このような補正は、植物がないプロットの収量を過大に見つもらさうであるということを示すことはさして困難ではない。しかし、資料によつて示されるように、収量と植物本数との間の、正確な関係を基とした補正は、むしろ正当なものであり、このような補正をする方法が共分散の技術の、自然な展開である。同じやり方の多くの応用は、勿論他の分野の研究者にも必要となるであろう。

共分散法による誤差のコントロールを示すために、次のように共分散分析を代数的に示そう。この実験は、乱塊法の圃場試験であると仮定している。

|         | DF  | $\sum x^2$ | $\sum xy$ | $\sum y^2$ | $b_{yx}$        | $b_{yx} \sum xy$ | $\sum (y_d^2)$  | $\sum (y_d^2)$ の DF |
|---------|-----|------------|-----------|------------|-----------------|------------------|-----------------|---------------------|
| プロット(B) | p   | $A_0$      | $B_0$     | $C_0$      |                 |                  |                 |                     |
| 処 理(T)  | q   | $A_1$      | $B_1$     | $C_1$      | $b_1 = B_1/A_1$ | $b_1 B_1$        | $C_1 - b_1 B_1$ | q-1                 |
| 誤 差(E)  | n   | $A_2$      | $B_2$     | $C_2$      | $b_2 = B_2/A_2$ | $b_2 B_2$        | $C_2 - b_2 B_2$ | n-1                 |
| T+E     | n+q | $A_t$      | $B_t$     | $C_t$      | $b_t = B_t/A_t$ | $b_t B_t$        | $C_t - b_t B_t$ | n+q-1               |

欄頭の $\sum (y_d^2)$ は、同じ行の回帰係数により補正した $Y$ に対する平方和を示す。

計算は表の各行毎に完成される。回帰係数は $B/A$ であり、 $Y$ に対する平方和についての補正は $bB$ すなわち $B^2/A$ である。最後の行では、 $A_t = A_1 + A_2$ 、 $B_t = B_1 + B_2$ 、 $C_t = C_1 + C_2$ のような処理および誤差のみを考えている。

方法の第二段階は次に示す通りである。

|                   | DF    | SS                            | MS      |
|-------------------|-------|-------------------------------|---------|
| T+E               | n+q-1 | $C_t - b_t B_t$               |         |
| E                 | n-1   | $C_2 - b_2 B_2$               | $\nu_2$ |
| T(差)              | q     | $C_1 + b_2 B_2 - b_t B_t$     | $\nu_1$ |
| T                 | q-1   | $C_1 - b_1 B_1$               | $\nu_3$ |
| $(b_1 - b_2)$ (差) | 1     | $b_1 B_1 + b_2 B_2 - b_t B_t$ | $\nu_4$ |

処理に対する最初の平方和は $C_1 + b_2 B_2 - b_t B_t$ に依する自由度に対しては $q+1-1$ であるから、その自由度は明らかに $q$ である。処理に対する第二の平方和は最初の表からそのまま移したもので、処理係数で補正されたから、自由度は $q-1$ である。第1の処理平方和から第二の処理平方和を引くと、平方和は $b_1 B_1 + b_2 B_2 - b_t B_t$ となり、簡単に次の等式が成立することが分る。



$$b_1B_1 + b_2B_2 - b_1B_1 = b_1^2A_1 + b_2^2A_2 - b_1^2A_1 = \frac{A_1A_2}{A_1 + A_2} (b_1 - b_2)^2$$

$b_1 = b_2$  のとき、この平方和は0となり、それに対応する分散 $\nu_2$ の有意性の検定は、誤差と処理の回帰係数間の差の有意性の検定となる。

処理の差の有意性の検定は、分散 $\nu_2$ と $\nu_1$ との比較である。 $\nu_1$ が $b_1 - b_2$ による著しい効果を含んでいるという事実は、このような効果は明らかに処理のある因子の特性によるのであるから検定の意味をそそぐことはない。プロット当り収量と植物の本数に関して、処理の回帰係数 $b_1$ は $b_2$ よりも大きい、そしてこれは $\nu_1$ の有意性に寄与しよう。しかし、 $b_2$ は処理内の植物本数に対する収量の回帰を表わし、収量への植物本数の効果の真の尺度として採用される。もし処理の回帰係数が $b_2$ よりも大であれば、これはおそらく附加的な遺伝的関係を反映するものであろうし、処理間の差の有意性に寄与するものであろう。しかし、 $\nu_2$ は更に進んだ検定が適用される、そして $\nu_3$ と $\nu_4$ の有意性を比較することにより、品種の効果を完全に描写することができる。例えば、もし根の数に収量に有意な効果をもつとすれば、このような分析の価値は、もし処理の差があるときには誤差分散が、減少し必然的に処理の差の有意性が増加するということである、もし補正した収量に有意差がなく補正をしない収量の分析が有意の差を示したとすれば、これは単に根本的な差は根の本数によるもので、根当り平均収量によつて測定したときの処理の効果によるものではないということしか意味しない。R. A. Fisherは、誤差を減する方法の有効性の尺度は分散の逆数であることを指摘した。これは時に逆分散(invariance)または情報の量(quantity of information)と呼ばれ、 $\frac{1}{\sigma^2}$ で表わす。共分散分析による誤差への減少を測る際には、この尺度は特に価値がある。最初の誤差分散が、関連した変量に対する平方和を補正して得た最終の誤差分散の大きさの2倍であれば、最初の型では、補正した結果と同じ正確度を与えるには約2倍の反復を必要とするであろう。このことから我々は、処理間の差の有意性がこれによつて常に増大すると結論してはならない。というのは同時に関連した変量による処理間の差がまた除かれていることを思いおこさねばならぬからである。

概略述べたように、有意性の検定を適用してから、次の段階は品種平均に対して補正を行うことである。誤差の欄での回帰係数は、収量に対する根の数の実際の影響を表わすものと考えられるから、この回帰係数は補正を行うため公式にいれなければならない。そのとき、補正平均は、もしそれらが根の本数の変動によつて影響されないならばどのような平均値の中においても最良可能な推定量であろう。回帰方程式は次の型のものとなる。

$$Y_i = \bar{Y}_{t1} + b_{yx} (\bar{X}_{t1} - \bar{X})$$

ここに、 $\bar{X}_{t1}$ は処理1に対するXの平均値、 $\bar{Y}_{t1}$ は同じ処理に対するYの平均値、 $b_{yx}$ は誤差欄でのXへのYの回帰で、 $\bar{Y}_{t1}$ は品種の推定平均である。

補正平均間の、差の標準誤差に対する公式はWishartが与えた。 $Y_p$ および $Y_q$ のような二つの補正平均値を比較するための二つの平均値間の差の標準誤差は

$$\sqrt{s^2 \frac{2}{r} + \frac{(\bar{X}_p - \bar{X}_q)^2}{A_2}} \quad (2)$$

である。ここに $s^2$ は共分散分析表の誤差欄の平均平方(例えば9-2表では $19.56/8 = 2.445$ )

であり、 $A_2$ は同じ行のXに対する平均平方で、rは反復数、 $(\bar{X}_p - \bar{X}_q)$ は方程式(1)から $Y_p$ および $Y_q$ を計算するための二式での二つの平均値を間の差である。したがって

$$Y_p = \bar{Y}_p - b_{yx} (\bar{X}_p - \bar{X}) \text{ および } Y_q = \bar{Y}_q - b_{yx} (\bar{X}_q - \bar{X})$$

二つの変量 $X_1$ および $X_2$ に対して補正した二つの平均値を比較する際に、次のようにして平均値の差の標準誤差を計算する。

$$\sqrt{s^2 \left( \frac{2}{r} + \frac{u^2B - 2u\nu P + \nu^2A}{AB - P^2} \right)} \quad (3)$$

ここにAおよびBは $X_1$ および $X_2$ に対する誤差の行の平方和で、Pは誤差の行の $X_1$ および $X_2$ に対する積和であり、 $u = (\bar{X}_{1p} - \bar{X}_{1q})$ は $X_1$ 平均値間の差で $\nu = (\bar{X}_{2p} - \bar{X}_{2q})$ は $X_2$ 平均値間の差である。Finneyは二つの補正平均値間の差の分散に対する簡単な公式を提案した。上に与えた正確な公式は各比較に対して異なっている。Finneyの提案は $(\bar{X}_p - \bar{X}_q)^2$ の平均の値を一括にして折衷することである

$$\frac{2A_1}{r(t-1)} \quad (4)$$

これはに等しいとおくことができる。ここに $A_1$ はXに対しての処理平方和、rは反復数およびtは処理の数である。差の分散に対する一般公式は簡単に次のようになる

$$\frac{2s^2}{r} \left( 1 + \frac{A_1}{A_2(t-1)} \right) \quad (5)$$

ここに $A_1$ は誤差の行でのXの平方和である。

3. 回帰係数の系列に対する不斉一性の検定。

共分散分析は、二つ以上の回帰係数間の差の有意性の検定に対してすばらしい方法である。前節と同じ記号を用いて、その方法を次に示す。最後の平方和は次に示すものである。

$$\sum_i (b_i B_i) - b_{0B} = \sum_i \left( \frac{A_i A_i (b_i - b_0)^2}{A_1 + A_2 + \dots + A_p} \right) \quad (6)$$

ここに $b_0$ および $b_i$ は回帰係数のすべての可能な対を表わす。また $A_i$ および $A_j$ はそれに対応するXの平方和のすべての可能な対を表わす。

したがって、F検定による $\nu_1$ および $\nu_2$ の比較は回帰係数の不斉一性の必要な検定を与える。



予 備 計 算

| 群   | DF  | $\Sigma(x^2)$ | $\Sigma(xy)$ | $\Sigma(y^2)$ | $b_{yx}$        | $b_{xy}$  | $\Sigma(xy)$    | $\Sigma(y^2)$ | $\Sigma(y^2)$ |
|-----|-----|---------------|--------------|---------------|-----------------|-----------|-----------------|---------------|---------------|
| 1   | q   | $A_1$         | $B_1$        | $C_1$         | $b_1 = B_1/A_1$ | $b_1 B_1$ | $C_1 - b_1 B_1$ | $q - 1$       | $q - 1$       |
| 2   | q   | $A_2$         | $B_2$        | $C_2$         | $b_2 = B_2/A_2$ | $b_2 B_2$ | $C_2 - b_2 B_2$ | $q - 1$       | $q - 1$       |
| ... | ... | ...           | ...          | ...           | ...             | ...       | ...             | ...           | ...           |
| P   | q   | $A_p$         | $B_p$        | $C_p$         | $b_p = B_p/A_p$ | $b_p B_p$ | $C_p - b_p B_p$ | $q - 1$       | $q - 1$       |
| 計   | Pq  | $A_t$         | $B_t$        | $C_t$         | $b_t = B_t/A_t$ | $b_t B_t$ | $C_t - b_t B_t$ | Pq - 1        | Pq - 1        |

有意性の検定

| 計  | DF     | $\Sigma(y_d^2)$        | MS                   |
|----|--------|------------------------|----------------------|
| 群内 | p(q-1) | $\Sigma(C - bB)$       | $\mu_1$ (誤差平均平方)     |
| 群間 | (p-1)  | $\Sigma(bB) - b_t B_t$ | $\mu_2$ (回帰係数間の差による) |

4. 誤差のコントロールに対する変量の選択。

誤差のコントロールに用いられるべき変量を選ぶ際には、いづれか注意を必要とする。簡単な規則は、それそのものが処理によつて影響されるような変量は選ばないことである。なぜならばこのような変量では、コントロールを行うと我々が測定したい処理の効果の部分を取り去る結果となるからである。次の年の実際の試行に用いられるプロットの系列で実行された一様な試行から得た収量は明らかに誤差のコントロールに対して理想的である。

果樹の実験では、処理をほどこす前に木当りの果樹の収量をとる。木の大きさも同様に誤差のコントロールに対しておそらく一年以上にわたつて便利な変量である、しかしこの変量は、処理の大きさに影響し得るようになってからは、適当ではなくなるのであろう。同様に、すべての誤差のコントロールのための変量では、処理の効果は除かれねばならない。

5. 積和の計算

共分散分析に於て、色々な成分に対する積和は、平方和の計算と全く類似の方法で計算できる。次表のようなr反復のn処理に対しての値の組を仮定しよう。

|     |     | 処 理                             |                                 |     | 反 復                             |                                 |
|-----|-----|---------------------------------|---------------------------------|-----|---------------------------------|---------------------------------|
|     |     | 1                               | 2                               | ... | n                               | 総 計                             |
| 反 復 | 1   | $X_{11}Y_{11}$                  | $X_{12}Y_{12}$                  | ... | $X_{1n}Y_{1n}$                  | $\Sigma X_{b1}$ $\Sigma Y_{b1}$ |
|     | 2   | $X_{21}Y_{21}$                  | $X_{22}Y_{22}$                  | ... | $X_{2n}Y_{2n}$                  | $\Sigma X_{b2}$ $\Sigma Y_{b2}$ |
|     | ... | ...                             | ...                             | ... | ...                             | ...                             |
| 処 理 | r   | $X_{r1}Y_{r1}$                  | $X_{r2}Y_{r2}$                  | ... | $X_{rn}Y_{rn}$                  | $\Sigma X_{br}$ $\Sigma Y_{br}$ |
|     | 総 計 | $\Sigma X_{b1}$ $\Sigma Y_{b1}$ | $\Sigma X_{b2}$ $\Sigma Y_{b2}$ | ... | $\Sigma X_{bn}$ $\Sigma Y_{bn}$ | $G_x$ $G_y$ 全総計                 |

積和は次の通りである

$$\text{全 体} \quad \Sigma'_{\text{r}}(xy) = \Sigma'_{\text{r}}(XY) - \frac{G_x G_y}{nr}$$

$$\text{処 理} \quad r \Sigma_1^n (\bar{x}_t \bar{y}_t) = \Sigma_1^n \left( \frac{\Sigma X_t \Sigma Y_t}{r} \right) - \frac{G_x G_y}{nr}$$

$$\text{反 復} \quad n \Sigma_1^r (\bar{x}_b \bar{y}_b) = \Sigma_1^r \left( \frac{\Sigma X_b \Sigma Y_b}{n} \right) - \frac{G_x G_y}{nr}$$

誤差 = 全体 - 処理 - 反復

$$\text{こゝに、} \bar{x}_t = \bar{X}_t - \bar{X}, \bar{y}_t = \bar{Y}_t - \bar{Y}, \bar{x}_b = \bar{X}_b - \bar{X}, \bar{y}_b = \bar{Y}_b - \bar{Y}$$

6. 例9-1. 共分散分析

この一つの例は共分散分析の重要な応用のいくつかを示すものである。資料はBartlettのものからとつた。資料はもともと欠測値の推定に対する共分散法の応用を示すものに用いられたものである。得た結果は9-1表に示してある。Bartlettの分析では、飼料3の反復1の結果は誤りであることが分り、推定値が代用された。こゝでは、もとの値を用いて分析を行う。

この実験では、牛を調べて、コントロールした期間にその可能なミルクの収量を得た。このコントロールした期間についての結果は、共分散分析による誤差のコントロールに対する基礎を与える。

共分散分析は9-2表にしてある。不斉一性の検定においてこの型のものが必要になると同じく処理によつて得られる平方和、積和および総計が推奨される。

積和の計算は次の通りである。

全 体

|     |   |      |      |                                      |
|-----|---|------|------|--------------------------------------|
| 飼 料 | 1 | 1937 |      |                                      |
| 飼 料 | 2 | 1405 |      |                                      |
| 飼 料 | 3 | 1861 | 7271 | $\frac{388 \times 293}{16} = 165.75$ |
| 飼 料 | 4 | 2068 |      |                                      |

$$\text{反 復} \quad \frac{(111 \times 81) + (93 \times 72) + \dots + 388 \times 293}{4} - \frac{16}{16} = 92.50$$

$$\text{飼 料} \quad \frac{(97 \times 79) + (85 \times 66) + \dots + 388 \times 293}{4} - \frac{16}{16} = 24.00$$



誤差 165.75-92.50-24.00-49.25

また、平方和を計算してから後、共分散分析を9-2表のように行うことができる。

9-1表

日々の牛の飼料の試行、コントロールおよび実験期間に対する  
(週当りポンド)でのミルクの平均収量

10

コントロールした期間の収量=X 実験期間の収量=Y

|    |   | 飼料 |    |    |    |     |    |     |    | 総計  |     |
|----|---|----|----|----|----|-----|----|-----|----|-----|-----|
|    |   | 1  |    | 2  |    | 3   |    | 4   |    | 5   |     |
|    |   | X  | Y  | X  | Y  | X   | Y  | X   | Y  | X   | Y   |
| 反復 | B | 28 | 21 | 22 | 17 | 27  | 18 | 34  | 25 | 111 | 81  |
|    | G | 24 | 21 | 22 | 16 | 26  | 20 | 21  | 15 | 93  | 72  |
|    | D | 20 | 16 | 18 | 16 | 19  | 14 | 21  | 16 | 78  | 62  |
|    | E | 25 | 21 | 23 | 17 | 31  | 19 | 27  | 21 | 106 | 78  |
| 総計 |   | 97 | 79 | 85 | 66 | 103 | 71 | 103 | 77 | 388 | 293 |

9-2表

共分散分析 9-1表の資料

|       | DF | $\sum x^2$ | $\sum xy$ | $\sum y^2$ | $b_{yx}$ | $b_{yx} \sum xy$ | $\sum y_d^2$ | DF |
|-------|----|------------|-----------|------------|----------|------------------|--------------|----|
| 反復    | 3  | 163.50     | 92.50     | 52.69      |          |                  |              |    |
| 処理    | 3  | 54.00      | 24.00     | 26.19      | 0.44444  | 10.67            | 15.52        | 2  |
| 誤差    | 9  | 73.50      | 49.25     | 52.56      | 0.67007  | 33.00            | 19.56        | 8  |
| 処理+誤差 | 12 | 127.50     | 73.25     | 78.75      | 0.57451  | 42.08            | 36.67        | 11 |

最後の欄は処理および誤差の平方和および積和を加えて得られることを注意しておこう。欄頭の  $\sum y_d^2$  は Y に対する平方和から前の欄(回帰に対する平方和)を差し引いて得られる。

分析の次の段階では誤差の行での回帰の有意性を決定しなければならない。これは補正F値値あらしめるためには有意でなければならない。次の簡単な分析を得る。

|    | SS    | DF | MS    | F    |
|----|-------|----|-------|------|
| 回帰 | 33.00 | 1  | 33.00 | 13.5 |
| 誤差 | 19.56 | 8  | 2.445 |      |

Fは1%点よりもはるかに大きいから、コントロールした収量への実験収量の回帰の有意性が確認され、処理平均に対しての補正は価値がある。

なお、我々は処理平均間の差の有意性を検定しなければならない。検定は9-3表のように実行する。

9-3表

処理効果の有意性の検定

|                 | DF | SS    | MS    | F    | 5%点  |
|-----------------|----|-------|-------|------|------|
| 処理+誤差           | 11 | 36.67 |       |      |      |
| 誤差              | 8  | 19.56 | 2.445 |      |      |
| 差=処理            | 3  | 17.11 | 5.703 | 2.33 | 4.07 |
| 処理(処理回帰により補正した) | 2  | 15.22 |       |      |      |
| $b_0 - b_1$     | 1  | 1.59  | 1.59  |      |      |

処理+誤差に対して補正した平方和は、誤差に対して補正した平方和と共に、9-2表からうつされる。差は補正した誤差に対して検定することができる処理に対する平方和である。最後の2行で行ったもう一つ進んだ段階は処理および誤差の回帰間の差の有意性の検定を得ることである。

補正した処理平均の対間の差の有意性の検定に対しては、Finneyが提案した公式(5)を適用できる。こゝでは二つの平均値間の差の分散は

$$\frac{2 \times 2.445}{4} \left( 1 + \frac{54.00}{73.50 \times 3} \right) = 1.522$$

で、平均値の差の標準誤差  $= \sqrt{1.522} = 1.23$

この実験の分析はある興味ある問題点を導く。第一に、コントロールした期間中のミルク収量で補正したことによる誤差の減少は全く著しい。補正前後の結果の相対的正確度は近似的に

$$\frac{52.56}{9} \div \frac{19.56}{8} = 2.39$$

で、情報に非常に決定的な利益があることを示す。

補正前の飼料の効果のF検定は1.49なるFの値であるのに、補正後のFの値は2.33である。これらのいずれも5%点に等しくないが、補正後に十分改良されて方法が健全であることを示している。正確度に非常に利益があるため、我々は飼料の有意性への大なる効果を期待する。しかしこの関係では、補正しない資料における飼料の効果の割合は真の飼料の効果に等しい。これは補正によつて取り除かれるということを再び思いおこさねばならない。例9-2 回帰の不均一性の検定。

もし平方和および積和を各飼料毎に決定すれば、9-4表のような分析を行うことができる。



| 飼料 | DF | $\Sigma x^2$ | $\Sigma xy$ | $\Sigma y^2$ | $b_{yx}$ | $b_{xy}$ | $\Sigma xy$ | $\Sigma y_d^2$ | DF |
|----|----|--------------|-------------|--------------|----------|----------|-------------|----------------|----|
| 1  | 3  | 32.75        | 21.25       | 18.75        | 0.6488   | 13.79    | 4.96        | 13.36          | 2  |
| 2  | 3  | 14.75        | 2.50        | 1.00         | 0.1695   | 0.42     | 0.58        |                | 2  |
| 3  | 3  | 74.75        | 32.75       | 20.75        | 0.4381   | 14.35    | 6.40        |                | 2  |
| 4  | 3  | 11.75        | 85.25       | 64.75        | 0.7429   | 63.33    | 1.42        |                | 2  |
| 合計 | 12 | 237.00       | 141.75      | 105.25       | 0.5981   | 84.78    | 20.47       |                | 11 |

すべての補正した平方和は自由度が11であるが、飼料毎に対する補正した平方和は4×2=8という自由度にすぎないということを注意しておこう。残り3つの自由度は4つの飼料間の差を表わす。これは我々に次の分析を可能ならしめる。

|    | DF | SS    | MS    | F    |
|----|----|-------|-------|------|
| 全体 | 11 | 20.47 |       |      |
| 飼料 | 8  | 13.36 | 1.670 |      |
| 差  | 3  | 7.11  | 2.370 | 1.42 |

不斉一性が著しいという証拠は殆んどない。

## 第10章 非直線回帰

### 1. 直線および非直線回帰成分の分析

第5章12節で、一組の平均値を表わす平方和は、個々の自由度を表わす平方和に分割できることを指摘した。簡単な規則にしたがつて、一つの与えられた比較を決定すると、他の比較を、すべての比較が直交的であるように選ぶことができることを示した。直線効果を表わす比較が必要な場合に、特に興味のある問題が生じてくる。例えば、処理の4水準に対する反応を表わす次の結果を得たと仮定しよう。

|        |     |      |     |     |
|--------|-----|------|-----|-----|
| 水準 (X) | 1   | 2    | 3   | 4   |
| 反応 (Y) | 9.4 | 12.6 | 8.7 | 6.3 |

結果をみると、第2の水準まで反応は増大し、それ以後は漸次減少している。直線的な反応は、全体のうちの  $y = a + bX$  という直線回帰方程式によつて簡単に表わすことができる部分であろう。我々の回帰法の知識から、適合さるべき方程式は、次の形で表わされることは明らかである。

$$Y = (\bar{Y} - b\bar{X}) + bX$$

ここに、 $b$ は回帰係数で次で与えられる。

$$b = \frac{\Sigma xy}{\Sigma x^2}$$

ここに  $x = X - \bar{X}$ 、 $y = Y - \bar{Y}$  である。我々はこの計算をやってみよう。そして後でもっと容易な方法でどのようにしてやれるかを示そう。

$$\Sigma(xy) = 85.9 - \frac{10.0 \times 37.0}{4} = -6.60$$

$$\Sigma x^2 = 30.0 - \frac{10.0^2}{4} = 5.00$$

それから、直線回帰の平方和は

$$\frac{(\Sigma xy)^2}{\Sigma x^2} = \frac{6.6^2}{5} = 8.71$$

である。Y平方和は勿論

$$9.4^2 + 12.6^2 + 8.7^2 + 6.3^2 - \frac{37.0^2}{4} = 20.25$$

である。そのとき分析は次のように行うことができる。



|         | SS    | DF | MS   |
|---------|-------|----|------|
| 直線回帰    | 8.71  | 1  | 8.71 |
| 回帰からの偏差 | 11.54 | 2  | 5.77 |
| 全 体     | 20.25 | 3  |      |

直線回帰の平方和は、全体の中あまり大きな割合を占めていない。これは明らかに直線でない反応においては、上り下りがあるため期待されることである。この結果は  $Y = a + bX + cX^2$  のようなもつと複雑な函数を適合する必要があることを示している。これは二次式であつて拋物線を与えるものである。この曲線は方程式  $Y = a + bX$  と同じような方法で適合できるが、その方法はすこし複雑で、Fisher および Yates が与えた表を応用するともつと容易な方法が利用できる。我々は実際には、これらの表で平方和を直接計算し、特別な目的に対して必要でない限り方程式を決定するまでは進まない。更に、表から  $Y = a + bX + cX^2 + dX^3$  のような高次の方程式に対する平方和をきめることは全く容易である。

我々の問題に対して必要な Fisher および Yates の表からの抜粋を Y の値と共に下にのせておく。

| Y   |      |     |     |
|-----|------|-----|-----|
| 9.4 | 12.6 | 8.7 | 6.3 |
| -3  | -1   | +1  | +3  |
| +1  | -1   | -1  | +1  |
| -1  | +3   | -3  | +1  |

必要な平方和は、次のようにして計算される。

直 線 反 応

$$\frac{(-3 \times 9.4 - 1 \times 12.6 + 1 \times 8.7 + 3 \times 6.3)^2}{20} = 8.71$$

二 次 反 応

$$\frac{(1 \times 9.4 - 1 \times 12.6 - 1 \times 8.7 + 1 \times 6.3)^2}{4} = 7.84$$

三 次 反 応

$$\frac{(-1 \times 9.4 + 3 \times 12.6 - 3 \times 8.7 + 1 \times 6.3)^2}{20} = 3.70$$

こゝで、各反応の分母は Fisher および Yates の表からとつた係数の平方和である。すなわち、最初の分母は  $3^2 + 1^2 + 1^2 + 3^2 = 20$  である。

平方和を加えると、総計は 20.25 となり、成分はすべて計算されたことになる。ここで二次の反応は、 $Y = a + bX + cX^2$  という形の二次式を適合したことによる平方和を表わすものではないということを注意することが大切である。それはただ直線方程式を適合した上に、更に二次式を適合して得た附加的な平方和を表わすに過ぎない。実際に、二次式によつて説明できる全体の平方和は  $8.71 + 7.84 = 16.55$  である。同様に、平方和 3.70 は三次式を適合することによる附加的な平方和を表わす。また、これから平方和は自由度 1 を表わすことになる。

完全な分析は次の通りである。

|         | SS   | DF | MS   |
|---------|------|----|------|
| 一 次 反 応 | 8.71 | 1  | 8.71 |
| 二 次 反 応 | 7.84 | 1  | 7.84 |
| 三 次 反 応 | 3.70 | 1  | 3.70 |

もし、各測定を二重に行えば 3 つの効果の有意性検定に対する誤差として適当な 3 つの自由度に対応するある附加的な平方和が得られたであらう。

また、測定した効果が直交的であることを注意しなければならない。これは、各行の係数の和が 0 であり、二つの行の対応する係数の積和が 0 であるから真である。

## 2. 回帰成分の分析の実例

17 品種の小麦の粉の系列を用いてパンを焼く際に臭素酸カリウムの色々な水準を加えた場合の効果についての実験において、Larmour は 10-1 表のような結果を得た。資料は 100 で割つて小数一位で表るめた立方で表わしたパン容積である。

10-1 表

パン焼きの際に加えた臭素酸カリウムの 4 水準で検定した

小麦の 17 品種の粉に対するパン容積

| 品 種 | 臭 素 酸 塩 の 水 準 |      |     |     | 品 種<br>総 計 | 回 帰 成 分 |      |      |
|-----|---------------|------|-----|-----|------------|---------|------|------|
|     | 1             | 2    | 3   | 4   |            | 一次      | 二次   | 三次   |
| 1   | 10.8          | 10.6 | 9.8 | 8.8 | 40.0       | -6.8    | -0.8 | 0.4  |
| 2   | 9.8           | 9.6  | 8.6 | 8.2 | 36.2       | -5.8    | -0.2 | 1.4  |
| 3   | 8.5           | 8.2  | 7.7 | 7.4 | 31.8       | -3.8    | 0    | 0.4  |
| 4   | 8.2           | 7.6  | 7.2 | 7.0 | 30.0       | -4.0    | 0.4  | 0    |
| 5   | 10.4          | 10.6 | 9.8 | 9.4 | 40.2       | -3.8    | -0.6 | 1.4  |
| 6   | 9.6           | 9.8  | 9.2 | 8.4 | 37.0       | -4.2    | -1.0 | 0.6  |
| 7   | 9.0           | 9.0  | 8.8 | 7.8 | 34.6       | -3.8    | -1.0 | -0.6 |
| 8   | 8.6           | 8.7  | 8.5 | 8.5 | 34.3       | -0.5    | -0.1 | 0.5  |



|    |       |       |       |       |       |       |      |      |
|----|-------|-------|-------|-------|-------|-------|------|------|
| 9  | 9.4   | 10.0  | 9.6   | 9.6   | 38.6  | 0.2   | -0.6 | 1.4  |
| 10 | 10.0  | 10.2  | 9.6   | 9.0   | 38.8  | -3.6  | -0.8 | 0.8  |
| 11 | 9.4   | 9.6   | 9.5   | 9.2   | 37.7  | -0.7  | -0.5 | 0.1  |
| 12 | 8.4   | 8.7   | 8.8   | 8.8   | 34.7  | 1.3   | -0.3 | 0.1  |
| 13 | 6.6   | 6.5   | 6.8   | 6.6   | 26.5  | 0.3   | -0.1 | -0.9 |
| 14 | 9.1   | 8.9   | 8.4   | 7.8   | 34.2  | -4.4  | -0.4 | 0.2  |
| 15 | 10.8  | 10.7  | 10.2  | 10.0  | 41.7  | -2.9  | -0.1 | 0.7  |
| 16 | 7.5   | 7.4   | 7.2   | 7.2   | 29.3  | -1.1  | 0.1  | 0.3  |
| 17 | 8.4   | 8.6   | 8.2   | 8.2   | 33.4  | -1.0  | -0.2 | 1.0  |
| 総計 | 154.5 | 154.7 | 147.9 | 141.9 | 599.0 | -44.6 | -6.2 | 7.8  |

10-2表

分散分析 10-1表の資料

|         | SS      | DF | MS     | F    | 5%点  |
|---------|---------|----|--------|------|------|
| 品 種     | 70.0097 | 16 | 4.376  | 40.1 | 1.86 |
| 処 理     | 6.5947  | 3  | 2.198  | 20.2 | 2.80 |
| 交 互 作 用 | 5.2303  | 48 | 0.1090 |      |      |

これらの資料に対する簡単な分散分析は10-2表に与えてある。各水準での各品種の値は一つの標本しかないから、誤差としては交互作用をさる。臭素酸塩の効果は全く有意であるが、処理総計を検討すると、第1の水準から第2の水準へはすこししか増加していず、第2の水準から第3の水準へ、第3の水準から第4の水準へは共に減少していることを注意しておこう。これは二次の効果を示しているから、臭素酸塩水準への平均ペン容積の回帰の一次、二次および三次の成分を決定することは興味がある。計算は次の通りである。

$$\text{一次成分} \quad \frac{(-3 \times 154.5 - 1 \times 154.7 + 1 \times 147.9 + 3 \times 141.9)^2}{17 \times 20} = 5.8505$$

$$\text{二次成分} \quad \frac{(1 \times 154.5 - 1 \times 154.7 - 1 \times 147.9 + 1 \times 141.9)^2}{17 \times 4} = 0.5653$$

$$\text{三次成分} \quad \frac{(-1 \times 154.5 + 3 \times 154.7 - 3 \times 147.9 + 1 \times 141.9)^2}{17 \times 20} = 0.1789$$

総 計 6.5947

総計がこれらの計算に入ってくると、分母には各総計に含まれる観測数を乗することを注意し

ておこう。

10-3表

一次、二次および三次成分への処理の効果の分析

|      | SS     | DF | MS     | F    | 5%点  |
|------|--------|----|--------|------|------|
| 一次成分 | 5.8505 | 1  | 5.850  | 53.7 | 4.04 |
| 二次成分 | 0.5653 | 1  | 0.5653 | 5.19 | 4.04 |
| 三次成分 | 0.1789 | 1  | 0.1789 | 1.64 | 4.04 |
| 交互作用 | 5.2303 | 48 | 0.1090 |      |      |

こうなるとこれ以上完全な分析は10-3表に与えることができる。一次および二次の効果は有意であるが、三次の効果は有意でない。

これ以上の情報は資料から色々な品種に対する一次および二次の効果の一致性の検討によつて得ることができる。いゝかえれば、この研究では品種での一次の効果および品種での二次の効果の交互作用が必要である。交互作用の成分は10-1表の7および8欄で与えた数字から計算される。これらの数字は各欄で一次、二次および三次の成分を計算することによつて得られる。すなわち

$$-6.8 = -3 \times 10.8 - 1 \times 10.6 + 1 \times 9.8 + 3 \times 8.8$$

$$-0.8 = 1 \times 10.8 - 1 \times 10.6 - 1 \times 9.8 + 1 \times 8.8$$

$$0.4 = -3 \times 10.8 + 3 \times 10.6 - 3 \times 9.8 + 1 \times 8.8$$

その時、平方和は次の通りである。

$$\text{一次効果} \times \text{品種} \quad \frac{(6.8^2 + 5.8^2 + \dots + 1.0^2)}{20} = 5.8505 = 4.2665$$

$$\text{二次効果} \times \text{品種} \quad \frac{(0.8^2 + 0.2^2 + \dots + 0.2^2)}{4} = 0.5655 = 0.6297$$

さて、平方和および自由度の完全な分割は次の通りである。

|      | SS      | DF |
|------|---------|----|
| 品 種  | 70.0097 | 16 |
| 処 理  |         |    |
| 一次成分 | 5.8505  | 1  |
| 二次成分 | 0.5653  | 1  |



|       |         |    |
|-------|---------|----|
| 三次成分  | 0.1789  | 1  |
| 交互作用  |         |    |
| 一次×品種 | 4.2665  | 16 |
| 二次×品種 | 0.6297  | 16 |
| 三次×品種 | 0.3341  | 16 |
| 総計    | 81.8347 | 67 |

実験の真の反復内の変動から計算される実験誤差がない場合には、交互作用の有意性の検定は、適切だと思われる残差変動に対して行わなければならない。例えば、一次×品種の成分は最後の2行の平方和および自由度を加えて作られた誤差に対して検定される。二次の交互作用は三次の交互作用の効果に対してのみ検定できるであろう。

### 3. 非直線回帰における基礎因子。

1節と2節では実験資料に適用した分散分析法の、簡単な拡張としての非直線回帰を紹介した。本章の残りでは、非直線的な性質をもつ資料を適合する一般問題および方法を述べることにしよう。第6章で、我々は、独立変数である $X$ と従属変数である $Y$ の値が対になつていような資料に、直線を適合する方法を研究した。一次の関係が本質的に成立つと考えられる場合が多いことから一次方程式を強調するのであるが、我々の経験が色々な種類の資料の研究によつて広がるに従つてすぐ多くの例で、直線が不適当であり、もし数学的な函数を適合すれば簡単な一次型よりも多少更に複雑となるに違いないということが分つてくる。勿論、一般に研究している変数が、二つの原因による変動を示すことも認めなければならない。一つの原因は、実際の測定にあり、他の原因は“有機的”なものである。いいかえれば、数および強度(intensity)が変る汎山の作用因子がある為である。このような変数では、わずかに曲線的な関係にむかうというような傾向は非常に大きな数の研究による以外には現れてこないであろう。心にどめておくべき第二の点は普通は限られた変動範囲上での二つの変数の現れ方を取扱っているということである。例えば、温度と穀物の収量間の関係の研究において、自然の条件の下では、実際には温度の範囲は非常に限られている。これ等の条件の下では、温度への収量の回帰は正であると考えことにする、そのとき我々は人工的な条件の下で同様な型の実験を計画することに決める。もし温度がある点以上になれば、明らかに、収量は減るであろう。そうしてこれは、最初は直線的にのぼり、それから曲線は下に下り、最後に0値に落ちるような曲線で完全にあらわされるであろう。肥料実験では、普通収量は適用した肥料の量と共に増加することが経験的にみられるが、このような関係には非常に明らかな限度がある。ある点では、これ以上、収量は増加せず、また十分肥料を施用すれば最後には効果は有徳にな

り収量は減ってくる。また生物的な変数では背後に存在する根本的な曲線関係のために、本質的に曲線であるようなある種の関係がある。成長曲線は非常によく知られた例である。プレート上のバクテリアの数の時間的増加を考えよう。時間の単位はバクテリアが分裂して2つのバクテリアになるに要する時間とする。明らかに理論的な増加は単一の単位から出発して、1, 2, 4, 8, 16, 32等で表わされ、ついには条件が飽和点に達して、割合がゆるやかとなつてくる。方程式は $Y = 2^X$ である、ここに $Y$ はバクテリアの数、 $X$ は時間の単位の数である。一般に、このような曲線は、 $Y = ab^X$ または $Y = ae^{kX}$ という型の方程式で表わされる、ここに $a$ および $b$ は常数で、 $e$ は近似的に2.71828でNaperの自然対数の底である。これらは指数函数と呼ばれる。

一般に、適合に用い得る函数は次の二つの群にわけることができる。(1)  $Y = a + bX + cX^2$ のような代数函数、(2) 変数 $X$ が指数として現われる、 $Y = ce^{kX}$ のような指数函数。

### 4. 曲線を適合するための前提。

我々が取扱う資料の型、実験目的および他の因子は曲線を適合するための前提となるものである。すべての回帰研究の下にある基礎的な前提は、よく一つの予測方程式を作ることであるといわれているのが、これは生物学的資料の研究には必ずしも重要な特性ではなく、おそらくあまり強調されすぎてきたきらいがある。二つの変数の対になつた値の組の分析に出会った場合、結果をとりまとめ、有意性の検定を適用できる統計量に、これらの資料を整理することが明らかに必要である。このことそれ自身が曲線を適合することに対する十分な理由なのである。

統計量を求め、有意性の検定を行うことの必要性は、また適合すべき方程式の型の選択の問題を簡単にするものである。我々が必要とするすべてのものは、合理的であり良好な適合を与えるようなものである。良好な適合を与える二三の方程式があるが、我々はこれらの中から最も簡単なもので最も容易に適合できるものを選ばなければならない。

この理由で、生ずる問題の二つの型の間にはかなりはっきりした区別がある。まず第一に背後にある基礎的な法則に基いた一定の数学的函数がある場合、また第二に、目的が単に資料をより簡単な形になおす場合である。第1の問題は、ある数学的な法則から、すべての変動は測定誤差による、という仮定がしばしば行なわれるような物理または化学の実験においてあらわれるものである。このような実験が成功するのは、良好な適合を与える曲線を得た場合ではなくて、基礎的な法則を表わす方程式に良好な適合を得る場合が多い。

### 5. 傾向が非直線的である場合の回帰方程式の選択。

これはその中から選ぶ方程式の数が非常に多いので必ずしも簡単な問題ではない。また各々に対して我々はそれに含まれる計算、適合度および妥当な有意性の検定ができるかどうかを考慮しなければならない。これらは二つの一般型をもっている。(1) 多項式型、(2) 指数式

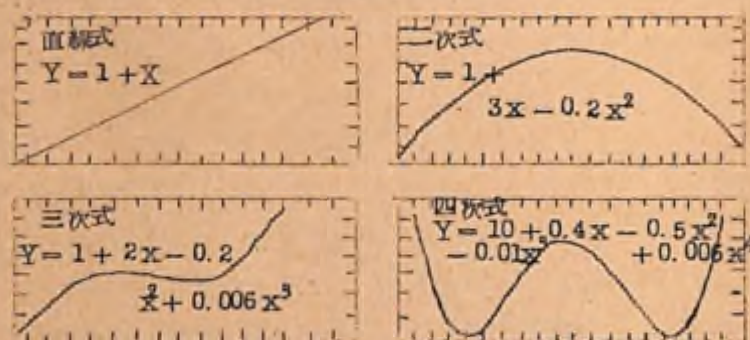


型である。指数式型は適合前によく対数に変換されるから対数曲線と呼ばれることがある。典型的な例を次にかけておく。

| 多項式型                     | 対数式型                |
|--------------------------|---------------------|
| $Y=a+bx$ (一次式)           | $Y=a+b \log X$      |
| $Y=a+bx+cx^2$ (二次式)      | $\log Y=a+bx$       |
| $Y=a+bx+cx^2+dx^3$ (三次式) | $\log Y=a+b \log x$ |
| 等                        | 等                   |

10-1図を見ればどのような資料に多項曲線がよく適合するかがわかるであろう。色々な次数の多項式の一般型を注意しよう。色々な次数の曲線は変曲点の数によつて区別する。第一の変曲点は三次式の場合にあらわれ適合の次数が高くなるごとに一つづつ点加わつてゆく。おなじみの幾つかある曲線が必要な資料には、高次の多項式を適合しなければならない。多項方程式の望ましい性質の一つはそれらが各適合の次数に対する適合度の検定を容易にするものである。このことは必要な曲線を選ぶ問題を簡易化する。

10-1図 4つの多項函数のグラフ



対数曲線には範囲の一端には平坦になつてゐるという特長がある。これは1, 2, 3, 4, 5, 6, の数字の対数をプロットすると簡単に観察できる。対数間の間隔は各段階で小さくならつて、ついには非常に小さくなるということに気がつくであらう。上の三つの型の特長は、ある仮想的な方程式に対してYの値を計算し方眼紙上にプロットすると観察できる。対数方程式が、どうすれば適当な対数方眼紙に書かれ、一次方程式に変換されるかを調べるために、特別な目盛をした紙が研究されなければならない。

#### 6. 多項方程式のあてはめ

もし

$$Y=a+bx+cx^2+dx^3 \quad (1)$$

という型の多項方程式が必要であれば、結局これは推定の誤差の平方和を最小にするようにa, b, c, およびd, を決定するという問題に帰着する。最小にしなければならない式は

$$\sum_{i=1}^n (y_i - a - bx_i - cx_i^2 - dx_i^3)^2 \quad (2)$$

であり、これは次のような4つの同時方程式の組を導く。

$$\left. \begin{aligned} an+b\sum X+c\sum X^2+d\sum X^3 &= \sum Y \\ a\sum X+b\sum X^2+c\sum X^3+d\sum X^4 &= \sum XY \\ a\sum X^2+b\sum X^3+c\sum X^4+d\sum X^5 &= \sum X^2Y \\ a\sum X^3+b\sum X^4+c\sum X^5+d\sum X^6 &= \sum X^3Y \end{aligned} \right\} \quad (3)$$

この連立方程式は必要な多項式の次数にしたがつて、縮むように拡張もできまた縮小もできる。もし適合の次数をrで表わすと決定されるべき各係数に対して、方程式1個が必要だから、方程式はr+1個となり、左辺の項の数はr+1である。

同次方程式によればどのような次数の多項式についても適合の一般方法が与えられる。方法は $\sum Y$ ,  $\sum XY$ ,  $\sum X^2Y$ , 等およびXの巾和を決定することである。方程式は我々が最もよく知つてゐる方法で、表の形にして解くことができる。こゝではドゥリットルの表示法が与えられるが、ドゥリットル法と同じ消去法および代入法で直接解くのを好む人もある。附録にのせてあるクラウト(CROUT)法のような行列式で解く人もある。

解法を詳しく研究する前に、ある種の簡単化が可能であることに注意しなければならない。まづ第1に、大きな数を避けるためXおよびYの値をできるだけ小さくなるように簡略化することが殆んど本質的なことである。10-4表のような表から出発して、多項式を適合する場合、まづ単位の級間隔を作り、ついでXおよびYの実際の平均に近い任意原点を作るのがよい。

ある型の資料では、別な簡略法がある、そして可能な場合には次節で簡単に述べる直交多項式と呼ばれるものを適合すると有益である。

#### 7. 直交多項式

与えられた資料の組に $Y=a_1+b_1X+c_1X^2$ の型の多項式を適合し、それから別の $Y=a_2+b_2X+c_2X^2+d_2X^3$ の型を適合したとすれば、 $a_1+a_2$ ,  $b_1+b_2$ 等ということが示されるのが分るであらう。これを他の面で考えると二次に適合した後、三次に適合する方法はただ単に直交方程式に $X^3$ の項を加えるという丈の問題ではないということである。いいかえれば、我々が各段階で同次方程式を解く方法を全体として進めてゆくのでなければ、次々との段階で適合度を検定しながら、適合をすゝめることが不可能であるということである。

直交多項式では、適合した方程式の型は次のようなものである。



ここに  $\xi$  は  $X$  の直交関数である。例えば

$\hat{\varepsilon}_1 = k_1 + x_1$ ,  $\hat{\varepsilon}_2 = k_2 + k_3 x + x^2$  等であると、これらの関数は次のようにして決定される。

$$\sum \xi_1 = 0, \sum \xi_2 = 0, \dots, \sum \xi_1 \xi_2 = 0, \sum \xi_1 \xi_3 = 0, \dots \dots \dots (5)$$

$X$ を自然数列としてとることができて $Y$ を等荷重であれば、 $\xi$ の値の代数的偏差はかなり簡単である。 $X$ の代りに $X$ をとると、次のようになる。

$$\xi_2 = (x^2 - \frac{n}{12}) \quad (6)$$

二次の直交多項式に対して最小にさるべき誤差は

$$\Sigma (Y - A - B \xi_1 - C \xi_2)^2$$

で、次の同次方程式を導く。

$$\begin{aligned} nA + B \sum \xi_1 + C \sum \xi_2 &= \sum Y \\ \sum \xi_1 A + B \sum \xi_1^2 + C \sum \xi_1 \xi_2 &= \sum \xi_1 Y \\ \sum \xi_2 A + B \sum \xi_1 \xi_2 + C \sum \xi_2^2 &= \sum \xi_2 Y \end{aligned} \quad (7)$$

そして、定義により  $\sum \xi_1 = 0$ ,  $\sum \xi_2 = 0$ ,  $\sum \xi_1 \xi_2 = 0$  であるから、次のように簡単になる。

$$nA = \sum Y, B \sum \xi_1^2 = \sum \xi_1 Y, C \sum \xi_2^2 = \sum \xi_2 Y \quad (8)$$

すなわち

$$A = \frac{\sum Y}{n}, \quad B = \frac{\sum \xi_1 Y}{\sum \xi_1^2}, \quad C = \frac{\sum \xi_2 Y}{\sum \xi_2^2} \quad (9)$$

三次の直交多項式を最小にするとき、 $A, B, C$ についても上と同じ結果が得られ、また

$$D = \frac{\sum \xi_j Y}{\sum \xi_j^2}$$

を得ることも明らかであろう。

回歸の平方和は

|      |                                                 |      |
|------|-------------------------------------------------|------|
| 一次回歸 | $B\Sigma\xi_1Y$                                 |      |
| 二次回歸 | $B\Sigma\xi_1Y + C\Sigma\xi_2Y$                 | (10) |
| 三次回歸 | $B\Sigma\xi_1Y + C\Sigma\xi_2Y + D\Sigma\xi_3Y$ |      |

で与えられるから、各自由度に対する回州の平方和は適合の各段階毎に得ることができる。

8 例10-1、Yの値が等荷重でない資料への多項式のあてはめ。

この型の例では、直交多項式を適合することは便利でないので、我々は直接に  $\sum X_i \sum X_i^2$  .....  $\sum Y_i$ ,  $\sum X_i Y_i$  ..... の値を計算して、(3)式で示したような連立正規方程式を解いてゆく。

10-4表

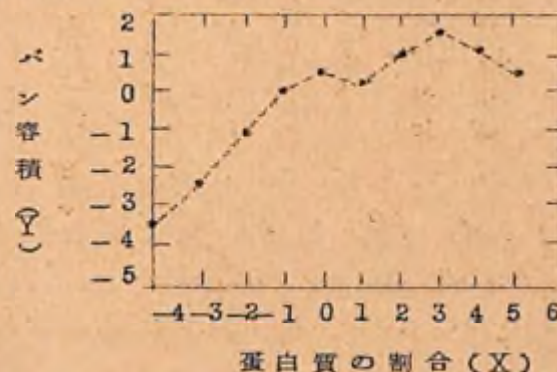
多項回帰方程式の計算のためにまとめた資料

|    |    |    |    |    |    |    |    |    |   |   |     |
|----|----|----|----|----|----|----|----|----|---|---|-----|
|    | 5  | 13 | 23 | 23 | 35 | 15 | 16 | 26 | 6 | 2 | 164 |
| 4  |    |    |    |    |    |    |    | 1  |   |   | 1   |
| 3  |    |    |    |    |    |    | 2  | 5  |   |   | 7   |
| 2  |    |    |    |    | 5  | 2  | 5  | 6  | 3 | 1 | 22  |
| 1  |    |    |    | 6  | 15 | 7  | 3  | 12 | 2 |   | 45  |
| 0  |    |    | 9  | 12 | 14 | 5  | 6  | 2  | 1 | 1 | 50  |
| -1 |    | 1  | 7  | 3  | 1  | 1  |    |    |   |   | 13  |
| -2 |    | 4  | 5  | 2  |    |    |    |    |   |   | 11  |
| -3 | 4  | 6  | 2  |    |    |    |    |    |   |   | 12  |
| -4 |    | 2  |    |    |    |    |    |    |   |   | 2   |
| -5 | 1  |    |    |    |    |    |    |    |   |   | 1   |
|    | -4 | -3 | -2 | -1 | 0  | 1  | 2  | 3  | 4 | 5 |     |

蛋白質の割合 (X)

10-24

10-1表からのパン客積平均のグラフ



10-4表は普通の相関表であるが、適合のためにYとXの実際の間隔を単位の間隔に変え、原点はYについては6、Xについては5とした。これは計算の手数を省く簡単な簡略法である。第一段階は適合すべき多項式をきめることである。そして適宜10-2図のようなグラフをY列の平均から作る。直線はあまりよくあてはまらないが、三次以上をすくめてゆく必要は至すないことは明らかである。実際には二次曲線が良好な適合を与えるが、三次曲線をあてはめてみよう、そしてもし三次を適合する必要がなければ、二次に変換することは簡単なことであ



るから。

次に正規方程式の解に必要な値を得るために下に示すような10-5表を作らなければならない。

$$\begin{aligned} an + b \sum x + c \sum x^2 + d \sum x^3 &= \sum y \\ a \sum x + b \sum x^2 + c \sum x^3 + d \sum x^4 &= \sum xy \\ a \sum x^2 + b \sum x^3 + c \sum x^4 + d \sum x^5 &= \sum x^2 y \\ a \sum x^3 + b \sum x^4 + c \sum x^5 + d \sum x^6 &= \sum x^3 y \end{aligned}$$

ここに  $x = X - 5$   $y = Y - 6$

10-5表

三次の多項式を適合するために必要な資料の計算

| 1       | 2   | 3   | 4                | 5                | 6              | 7                           | 8                           | 9                           | 10                          | 11                          | 12     |
|---------|-----|-----|------------------|------------------|----------------|-----------------------------|-----------------------------|-----------------------------|-----------------------------|-----------------------------|--------|
| 統計<br>x | y   | xy  | x <sup>2</sup> y | x <sup>3</sup> y | f <sub>x</sub> | f <sub>x</sub> <sup>2</sup> | f <sub>x</sub> <sup>3</sup> | f <sub>x</sub> <sup>4</sup> | f <sub>x</sub> <sup>5</sup> | f <sub>x</sub> <sup>6</sup> |        |
| -4      | -17 | 68  | -272             | 1088             | 5              | -20                         | 80                          | -320                        | 1,280                       | -5,120                      | 20,480 |
| -3      | -35 | 105 | -315             | 945              | 13             | -39                         | 117                         | -351                        | 1,053                       | -3,159                      | 9,477  |
| -2      | -23 | 46  | -92              | 184              | 23             | -46                         | 92                          | -184                        | 368                         | -736                        | 1,472  |
| -1      | -1  | 1   | 1                | 1                | 23             | -23                         | 23                          | -23                         | 23                          | -23                         | 23     |
| 0       | 24  |     |                  |                  | 35             |                             |                             |                             |                             |                             |        |
| 1       | 10  | 10  | 10               | 10               | 15             | 15                          | 15                          | 15                          | 15                          | 15                          | 15     |
| 2       | 19  | 38  | 76               | 152              | 16             | 32                          | 64                          | 128                         | 256                         | 512                         | 1,024  |
| 3       | 43  | 129 | 387              | 1,161            | 26             | 78                          | 234                         | 702                         | 2,106                       | 6,318                       | 18,954 |
| 4       | 8   | 32  | 128              | 512              | 6              | 24                          | 96                          | 384                         | 1,536                       | 6,144                       | 24,576 |
| 5       | 2   | 10  | 50               | 250              | 2              | 10                          | 50                          | 250                         | 1,250                       | 6,250                       | 31,250 |

30 439 -29 4,303 164 31 771 601 7,887 10,201 107,271

$\sum y \sum xy \sum x^2 y \sum x^3 y n \sum x \sum x^2 \sum x^3 \sum x^4 \sum x^5 \sum x^6$

1, 2および6欄はもとの相関表から直接に与つたもので、1および6はそのまま移し、2欄は普通のようにして計算したことに注意しよう。それから次のように左から右に向つて計算を進めてゆくことができる。

$$\begin{aligned} -4 \times 17 &= 68 & -4 \times 68 &= -272 & -4 \times -272 &= 1,088 \\ -4 \times 5 &= -20 & -4 \times -20 &= 80 & -4 \times 80 &= -320 \end{aligned}$$

等

欄を加えると最後の行に示したような値となる。解かなければならない方程式は次のようにして作ることができる。

$$164a + 31b + 771c + 601d = 30$$

$$31a + 771b + 601c + 7,887d = 439$$

$$771a + 601b + 7,887c + 10,201d = -29$$

$$601a + 7,887b + 10,201c + 107,271d = 4,303$$

それから、10-6表を用いるとド-リットル法により方程式を解くことができる。やりたければ簡略ド-リットル法でやつてもよいが、こゝでは完全解がより便利である。この表で行つた方法は簡略ド-リットル解について第8章第5節で述べたものと同じであるが、すべての項が書き加れている。すべての方程式でa, b, c, d, の決定した値を書き加えてチェックしてみよ。

多項式の値はいま次の方程式から作ることができる。

$$y = 0.5728 + 0.5957x - 0.1096x^2 + 0.003533x^3$$

ここに  $y = Y - 6$ ,  $x = X - 5$ 。X=1に対して  $X = 1 - 5 = -4$  であることに注意しよう。したがつて、上の方程式に-4を代入することは  $X = 1$  を代入することと同じである。X=-4のとき  $Y = -3.79$  を得る。そのとき  $Y = 6 - 3.79 = 2.21$  等を得る。したがつて、必ずしもyに対して  $(Y - 6)$ 、xに対して  $(X - 5)$  において多項式の値を計算する前に簡単にする必要はない。

この方法で最も興味のある部分は回帰曲線の適合度の検定である。10-6表は三次曲線ばかりでなく、二次および一次の曲線の適合度の検定に対する資料をも与える。これによつて我々は必要な適合の次数を決定することができる。我々はまづ10-4表からYの全平方和を計算する。これは

$$434 - \frac{(30)^2}{164} = 428.51$$

である。そのとき、10-6表から直線方程式  $y_0 = a_0 + b_0x$  を適合したことによる回帰の平方和は  $(5.2) \times (6.5)^2$  で与えられる、ここで数は夫々行および列に関するものである。数字で示せば

$$765.1402 \times (0.5663397)^2 = 245.41$$

である。  $y_1 = a_1 + b_1x + c_1x^2$  なる型の二次式を適合した場合、回帰に対する平方和は  $245.41 + (10.3) \times (11.5)^2 = 245.41 + 3991.4762 \times (0.1071958)^2 = 291.27$

最後に三次式  $y_2 = a_2 + b_2x + c_2x^2 + d_2x^3$  を適合した場合の平方和は

$$\begin{aligned} 245.41 + 45.86 + (16.4) \times (17.5)^2 \\ = 245.41 + 45.86 + 24,200.021 \times (0.00353334)^2 = 291.57 \end{aligned}$$



となる。

次の段階は回帰に対する平方和が適合の各次数によつてどれ位増加したかを検定するため、分散分析を行なうことである。すなわ、次を得る。

|          | SS     | DF  | MS     |
|----------|--------|-----|--------|
| 一次回帰     | 245.41 | 1   | 245.41 |
| 二次による超過分 | 45.86  | 1   | 45.86  |
| 三次による超過分 | 0.30   | 1   | 0.30   |
| 残差       | 136.94 | 160 | 0.86   |
| 全体       | 428.51 | 163 |        |

明らかに一次回帰は有意であるが、二次の適合によつて非常に有効な改善が行なわれている。

三次の適合は大して価値はない。

#### 10-6表

三次の多項式のあてはめに対する同次方程式の解

|    | 1             | 2             | 3              | 4              | 5           |
|----|---------------|---------------|----------------|----------------|-------------|
| 1  | 164           | 31            | 771            | 601            | 30          |
| 2  | 1.0           | 0.1890244     | 4.701220       | 3.664631       | 0.1829268   |
| 3  |               | 771           | 601            | 7.887          | 439         |
| 4  |               | -5.859756     | -145.7378      | -113.6037      | -5.670732   |
| 5  |               | 765.1402      | 455.2622       | 7,773.3963     | 433,3293    |
| 6  |               | 1.0           | 0.5950049      | 10.15944       | 0.5663397   |
| 7  |               |               | 7.887          | 10,201         | -29         |
| 8  |               |               | -3,624.6406    | -2,825.4332    | -141.0366   |
| 9  |               |               | -270.8832      | -4,625.2089    | -257.8331   |
| 10 |               |               | 3,991.4762     | 2,750.3759     | -427.8697   |
| 11 |               |               | 1.0            | 0.6890623      | -0.1071968  |
| 12 |               |               |                | 107.271        | 4,303       |
| 13 |               |               |                | -2,202.4450    | -109.9390   |
| 14 |               |               |                | -78,973.3633   | -4,402.3830 |
| 15 |               |               |                | -1,895.1803    | 294.8289    |
| 16 |               |               |                | 24,200.021     | 85.5089     |
| 17 |               |               |                | 1.0            | 0.00353334  |
| 18 |               |               |                | d = 0.00353334 | 0.00353334  |
| 19 |               |               | c = -0.1096305 | -0.00243469    | -0.1071968  |
| 20 |               | b = 0.5956736 | 0.0652307      | -0.0358968     | 0.5663397   |
| 21 | a = 0.5727787 | -0.1125968    | 0.5153971      | -0.0129484     | 0.1829268   |

9. 直交多項式あてはめのためのフィッシャーの加算法。

Yの値が等荷重であるかまたはそう仮定できる場合、Xが等間隔であるときにはR. A. Fisherが展開した直交多項式をあてはめる方法が、実際の多項式の値から理論的な多項式の値に到る決定的な近路を与える。等荷重的な努力もその大部分は計算器で行うことができる連続的な加算から成立っているから同様に容易である。同じ数学的な背景から生じたものと最近の進展はFisherおよびYatesの与えたあてはめである。これは更に早い方法ではあるが、しかし計算器および彼等が用意した表が必要である。例10-3で詳細に述べるこの方法は1節、2節で用いられたものである。

フィッシャーの加算法は例によつて述べるのが最も都合がよい。

10. 例10-2, 直交多項式のあてはめ—加算法。

10-7表の資料は, Larmourが報告したパン焼き実験における、パンの容積に関するものである。この実験では、ねり粉を1分から4分まで時間を変えて、高速度のミキサーでこね合せた。

#### 10-7表

ねり粉とかきまぜる時間のパン容積への効果についての実験結果

| かきまぜる時間<br>(分) | (X) | パン容積<br>550を差引いた立方 $\alpha$ (Y) |
|----------------|-----|---------------------------------|
| 1.0            |     | 235                             |
| 1.5            |     | 280                             |
| 2.0            |     | 255                             |
| 2.5            |     | 190                             |
| 3.0            |     | 120                             |
| 3.5            |     | 55                              |
| 4.0            |     | 20                              |

適合法は下に示す段階の系列に分けられる。

1.  $S_0, S_1, S_2, \dots, S_r$  の計算。ここにrは実行さるべきあてはめ次数を表わす。

Xの実際の値を自然数1, 2, 3, ..., 7でおきかえてから、10-8表のように資料をまとめるこの段階では三次曲線が十分と仮定しているから、 $S_0, S_1, S_2, S_3$ をうるための連続的な加算の手順が例示される。

#### 10-8表

$S_0, S_1, S_2, S_3$  の計算

| X | Y   |     |     |     |     |
|---|-----|-----|-----|-----|-----|
| 1 | 255 | 235 | 235 | 235 | 235 |



|       |       |       |        |        |             |
|-------|-------|-------|--------|--------|-------------|
| 2     | 280   | 515   | 750    | 985    |             |
| 3     | 255   | 770   | 1,520  | 2,505  | 3欄では        |
| 4     | 190   | 960   | 2,480  | 4,985  | 235+280=515 |
| 5     | 120   | 1,080 | 3,560  | 8,545  | 515+255=770 |
| 6     | 55    | 1,135 | 4,695  | 13,240 | 等           |
| 7     | 20    | 1,155 | 5,850  | 19,090 | 4欄および5欄も同じ  |
| <hr/> |       |       |        |        |             |
|       | 1,155 | 5,850 | 19,090 | 49,585 |             |

$S_0 \quad S_1 \quad S_2 \quad S_3$

2. a, b, c, d, e, ……および  $a', b', c', d', e', \dots$  の計算。ここに  $r$  次をあてはめるためにはこれらの中の  $r+1$  個を計算しなければならない

適合の次数

$$\begin{aligned}
 0 \quad a &= \frac{1}{n} S_0 & a' &= a \\
 1 \quad b &= \frac{2}{n(n+1)} S_1 & b' &= a - b \\
 2 \quad c &= \frac{6}{n(n+1)(n+2)} S_2 & c' &= a - 3b + 2c \\
 3 \quad d &= \frac{24}{n(n+1)(n+2)(n+3)} S_3 & d' &= a - 6b + 10c - 5d \\
 4 \quad e &= \frac{120}{n(n+1)\dots(n+4)} S_4 & e' &= a - 10b + 30c - 35d + 14e \\
 5 \quad f &= \frac{720}{n(n+1)\dots(n+5)} S_5 & f' &= 15b + 70c - 140d + 126e - 42a \\
 \vdots & & & \\
 r &= \frac{(r+1)!}{n(n+1)\dots(n+r)} S_r & &
 \end{aligned}$$

上の右辺の方程式の係数は、系列がなくなるまで、次々に  $\frac{r(r+1)}{1 \times 2}, \frac{(r-1)(r+2)}{2 \times 3}, \dots$  等に乗じて作る。例えば、 $d'$  の計算に対する係数を得るには、 $r=3$  のとき  $\frac{3 \times 4}{2} = 6, 6 \times \frac{2 \times 5}{2 \times 3} = 10, 10 \times \frac{1 \times 6}{3 \times 4} = 5$  である。

したがって

$$\begin{aligned}
 a &= \frac{1}{7}(1155) = 165.00000 & a' &= 165.00000 \\
 b &= \frac{1}{28}(5850) = 208.92857 & b' &= (165.00000 - 208.92857) = -43.92857 \\
 c &= \frac{1}{84}(19,090) = 227.26190 & c' &= 165.00000 - (3 \times 208.92857) + (2 \times 227.26190) = -7.26191 \\
 d &= \frac{1}{210}(49,585) = 236.11905 & d' &= 3.45233
 \end{aligned}$$

3. あてはめの次数に対する平方和の計算および有意性の検定。

ここで、必要なあてはめの次数を決定するために有意性の検定を行うことができる。あてはめの各次数に対する平方和は次で与えられる。

| 自由度      | 平方和の公式                                                               | 例                                    |
|----------|----------------------------------------------------------------------|--------------------------------------|
| 0        | $n d'^2$ (平均の適合を表す)                                                  | $7 \times (165.0)^2 = 190,575.00$    |
| 1        | $3 \frac{n(n+1)}{n-1} b'^2$                                          | $28 \times (43.92857)^2 = 54,032.14$ |
| 2        | $5 \frac{n(n+1)(n+2)}{(n-1)(n-2)} c'^2$                              | $84 \times (7.26191)^2 = 4,429.77$   |
| 3        | $7 \frac{n(n+1)\dots(n+3)}{(n-1)(n-2)(n-3)} d'^2$                    | $294 \times (3.45233)^2 = 3,504.06$  |
| 4        | $9 \frac{n(n+1)\dots(n+4)}{(n-1)(n-2)\dots(n-4)} e'^2$               |                                      |
| 5        | $11 \frac{n(n+1)\dots(n+5)}{(n-1)(n-2)\dots(n-5)} f'^2$              |                                      |
| $\vdots$ |                                                                      |                                      |
| $r$      | $(2r+1) \frac{n(n+1)\dots(n+r)}{(n-1)(n-2)\dots(n-r)} (\text{常数})^2$ |                                      |

$\Sigma Y^2 = 252,575, \Sigma y^2 = 252,575 - 190,575.00 = 62,000.00$  で、次のように分散分析を行うことができる。



|          | SS        | DF | MS     |
|----------|-----------|----|--------|
| 一次回帰     | 54,032.14 | 1  | 54,032 |
| 二次による超過分 | 4,429.77  | 1  | 4,430  |
| 三次による超過分 | 3,504.06  | 1  | 3,504  |
| 残差       | 34.03     | 3  | 11.35  |
| 全体       | 62,000.00 | 6  |        |

回帰の平方和はすべて非常に有意である。我々がこれ以上の段階に進まねばならないかどうかの問題は残差の平方和がどれ位小さいかによつて十分合理的に答えられる。これ以上を増しても平方和が多少共減じさうには思えない。

#### 4. 多項式の値の計算。

あてはめた曲線のグラフを作るためには、Xの各値に対する多項式の値を決定する必要がある。その方法はまづ、五次迄のあてはめに対する次の一般公式によつて与えられる量 $Y$ ,  $\Delta Y_1$ ,  $\Delta Y_2$ ,  $\Delta Y_3$ , ...,  $\Delta Y_r$ を得ることである。これは好きなだけ拡大も縮小もできるものである。

五次のあてはめに対する一般公式

例

$$\begin{aligned}
 Y_1 &= a' + 3b' + 5c' + 7d' + 9e' + 11f' & Y_1 &= 165.0 + (3 \times -43.92857) \\
 & & & + (5 \times -7.26191) + (7 \\
 & & & \times 3.45233) = 21.07105 \\
 \Delta Y_1 &= (-1)^{\frac{1}{2}} \frac{6}{(n-1)} (b' + 5c' + 14d' + 30e' & \Delta Y_1 &= -\left(\frac{6}{6}\right) [(-43.92857 \\
 & + 55f')] & & 7 + 15 \times -7.26191) + \\
 & & & (14 \times 3.45233) = \\
 & & & 31.90550 \\
 \Delta Y_2 &= (-1)^{\frac{2}{2}} \frac{6 \cdot 0}{(n-1)(n-2)} (c' + 7d' + 27e' + & \Delta Y_2 &= \frac{6 \cdot 0}{3 \cdot 0} \times (-7.261 \\
 & 77f') & & 91) + (7 \times 3.452 \\
 & & & 33) = 33.80880 \\
 \Delta Y_3 &= (-1)^{\frac{3}{2}} \frac{840}{(n-1)(n-2)(n-3)} (d' + & \Delta Y_3 &= \frac{840}{120} \times 3.45233 = \\
 & 9e' + 44f') & & -24.16631 \\
 \Delta Y_4 &= (-1)^{\frac{4}{2}} \frac{15,120}{(n-1)(n-2) \dots (n-4)} (e' + 11f') \\
 \Delta Y_5 &= (-1)^{\frac{5}{2}} \frac{332,640}{(n-1)(n-2) \dots (n-5)} (f') & (13)
 \end{aligned}$$

多項式の値は、結局、 $Y_1$ ,  $\Delta Y_1$ ,  $\Delta Y_2$ , ...,  $\Delta Y_r$ の値から出発して、連続的な加算法の手順によつて作ることができる。この過程を例の資料で示そう。

| X | Y      |         |                    |
|---|--------|---------|--------------------|
| 1 | 236.31 |         |                    |
| 2 | 277.02 | -40.714 |                    |
| 3 | 254.88 | 22.143  |                    |
| 4 | 194.05 | 60.833  | -62.8564           |
| 5 | 118.69 | 75.357  | -38.6901           |
| 6 | 52.98  | 65.714  | -14.5238           |
| 7 | 21.071 | 31.9055 | 9.6425             |
|   |        |         | 33.80880 -24.16631 |

加算の出発点は右下隅である。すなわち、 $33.80880 - 24.16631 = 9.6425$

$9.6425 - 24.16631 = -14.5238$ 等である。ここで計算は小数以下の全桁数について行なうが書いてあるのは小数以下4位だけである。

次の欄では、 $31.9055 + 33.8088 = 65.714$ ,  $65.714 + 9.6425 = 75.357$ 等である。この場合にももう一桁多く計算しなければならない。

11. フイツシャーおよびイエーツの $\xi$ の表を用いた直交多項式のあてはめ。

利用できれば、フィツシャーおよびイエーツの表を用いると直交多項式のあてはめは最も迅速にできる。それにはYの値が等荷重でXが等間隔であることが必要である。

あてはめられる方程式は次の形のものである。

$$Y_0 = A + B\xi_1 + C\xi_2 + D\xi_3 + \dots \quad (14)$$

こゝに、 $\xi$ はそれぞれのものがXの多項式である。(7節参照)。あてはめを簡単にするため、実際の方程式は

$$Y_0 = A + B'\xi'_1 + C'\xi'_2 + D'\xi'_3 + \dots \quad (15)$$

とする、こゝに $\xi'_1 = \text{入}\xi$ で、入の値は表に与えてある。

7点に5次の多項式をあてはめるため $\xi$ の表の一部をこゝに再録しておこう。

| X                | Y     | $\xi'_1$ | $\xi'_2$ | $\xi'_3$      | $\xi'_4$       | $\xi'_5$       |
|------------------|-------|----------|----------|---------------|----------------|----------------|
| 1                | $Y_1$ | -3       | 5        | -1            | 3              | -1             |
| 2                | $Y_2$ | -2       | 0        | 1             | -7             | 4              |
| 3                | $Y_3$ | -1       | -3       | 1             | 1              | -5             |
| 4                | $Y_4$ | 0        | -4       | 0             | 6              | 0              |
| 5                | $Y_5$ | 1        | -3       | -1            | 1              | 5              |
| 6                | $Y_6$ | 2        | 0        | -1            | -7             | -4             |
| 7                | $Y_7$ | 3        | 5        | 1             | 3              | 1              |
| $\Sigma(\xi')^2$ |       | 28       | 84       | 6             | 154            | 84             |
| 入                |       | 1        | 1        | $\frac{1}{6}$ | $\frac{7}{12}$ | $\frac{7}{20}$ |



直線をあてはめるためには、 $\xi_1'$  の値だけが必要である。二次の場合には  $\xi_1'$ 、 $\xi_2'$  が必要である以下同じ。

第一段階は、あてはめの各次数に対して平方和を決定することである、そして分散分析を行って必要な多項式を決定することである。平方和は次で与えられる。

$$\begin{aligned} \text{一 次} & \quad \frac{(\sum Y \xi_1')^2}{\sum \xi_1'^2} \\ \text{二 次} & \quad \frac{(\sum Y \xi_2')^2}{\sum \xi_2'^2} \\ & \quad \text{等} \end{aligned} \quad (16)$$

必要な多項式の次数をきめると、次のものがきまる。

$$\begin{aligned} A' &= \frac{\sum Y \xi_1'}{\sum (\xi_1')^2} \\ B' &= \frac{\sum Y \xi_2'}{\sum (\xi_2')^2} \\ & \quad \text{等} \end{aligned} \quad (17)$$

そのとき、方程式に  $\xi_1'$ 、 $\xi_2'$  等の表の値を代入して、多項式の値が決定できる。例えば、7 点の資料に四次曲線をあてはめること、

$$Y_1 = A' + B'(-3) + C'(5) + D'(-1) + E'(3)$$

$$Y_2 = A' + B'(-2) + C'(0) + D'(1) + E'(-7)$$

等

必要ならば、方程式(15)はXの項でYを与える方程式に変換できる。まづXに変換し、つぎにXに変換する二段階でこれを行うと最もよい。これは五次までの方程式に対して下に与えた等式で行うことができる。

$$\xi_1' = \xi_1 = \lambda X$$

$$\xi_2' = \xi_2 = \lambda \left( X^2 - \frac{n^2-1}{12} \right)$$

$$\xi_3' = \xi_3 = \lambda \left( X^3 - \frac{3n^2-7}{20} X \right)$$

$$\xi_4' = \xi_4 = \lambda \left( X^4 - \frac{3n^2-13}{14} X^2 + \frac{3(n^2-1)(n^2-9)}{560} \right) \quad (18)$$

$$\xi_5' = \xi_5 = \lambda \left( X^5 - \frac{5(n^2-7)}{18} X^3 + \frac{15n^4-230n^2+407}{1008} X \right)$$

こゝに、入の値は各行で異なる、そしてフィッシャーおよびイエーツの表に与えてある。

上の公式でnを含む式の数値を見出してからこれらを(15)に代入して簡単にする。もし方程式がXの代りにXについて必要であれば、 $(X - \bar{X})$ を代入する必要がある。それにおいて $\bar{X}$ は $\bar{X}$ の実際の数値で更に簡単にされる。

12. 例10-3。フィッシャーおよびイエーツの $\xi$ の表による多項式のあてはめ。

フィッシャーおよびイエーツの $\xi$ の表を用い次々の段階で多項式をあてはめるこのやり方を10-9表の資料で例示しよう。こゝではGeddesおよびHunterのPHとenzyme asparaginase(一種の酵素)の活力間の関係がのせてある。フィッシャーおよびイエーツの表からとつた $\xi$ の値も同じく表に与えてある。

10-9表

5次の直交多項式をあてはめるのに必要な $\xi'$ の値とPH(X)およびenzyme asparaginaseの活力(Y)に対する資料

| X  | Y   | $\xi_1'$ | $\xi_2'$ | $\xi_3'$ | $\xi_4'$ | $\xi_5'$ | 四次方程式に対するY |
|----|-----|----------|----------|----------|----------|----------|------------|
| 1  | 0.2 | -13      | 13       | -143     | 143      | -143     | 0.27       |
| 2  | 0.4 | -11      | 7        | -11      | -77      | 187      | 0.17       |
| 3  | 1.4 | -9       | 2        | 66       | -132     | 132      | 1.69       |
| 4  | 4.1 | -7       | -2       | 98       | -92      | -28      | 4.02       |
| 5  | 6.6 | -5       | -5       | 96       | -13      | -139     | 0.48       |
| 6  | 8.7 | -3       | -7       | 67       | 63       | -145     | 8.55       |
| 7  | 9.8 | -1       | -8       | 24       | 108      | -60      | 9.89       |
| 8  | 9.9 | 1        | -8       | -24      | 108      | 60       | 10.26      |
| 9  | 9.5 | 3        | -7       | -67      | 53       | 145      | 9.62       |
| 10 | 8.2 | 5        | -5       | -96      | -13      | 139      | 8.05       |
| 11 | 6.4 | 7        | -2       | -98      | -92      | 28       | 5.81       |
| 12 | 3.3 | 9        | 2        | -66      | -132     | -132     | 3.29       |
| 13 | 0.3 | 11       | 7        | 11       | -77      | -187     | 1.04       |
| 14 | 0.1 | 13       | 13       | 143      | 143      | 143      | -0.24      |

|    |      |       |                   |               |               |                |                |         |       |
|----|------|-------|-------------------|---------------|---------------|----------------|----------------|---------|-------|
| 和  | 105  | 68.9  | $\sum (\xi_i'^2)$ | 910           | 728           | 97,240         | 136,136        | 235,144 | 68.90 |
| 平均 | 27.5 | 4.924 | $\lambda^2$       | $\frac{1}{2}$ | $\frac{5}{3}$ | $\frac{7}{12}$ | $\frac{7}{30}$ |         |       |

A', B', C', D', E'およびF'の計算は各あてはめの附加的な次数を表わす平方和と一緒に下に与えてある。

$$\sum Y^2 = 541.7100 - 339.0864$$

$$= 202.6236$$

$$\sum (\xi_i' Y) = 41.3 \quad B' = 0.045385$$

$$SS = \frac{(41.3)^2}{910} = 1.8744$$



$$\Sigma(\xi_2'Y) = -361.8 \quad C' = -0.49698 \quad SS = \frac{(361.8)^2}{728} = 179.8066$$

$$\Sigma(\xi_3'Y) = -574.2 \quad D' = -0.0059050 \quad SS = \frac{(574.2)^2}{97240} = 3.3006$$

$$\Sigma(\xi_4'Y) = 1484.4 \quad E' = 0.010904 \quad SS = \frac{(1484.4)^2}{136136} = 16.1856$$

$$\Sigma(\xi_5'Y) = 162.4 \quad F' = 0.00069064 \quad SS = \frac{(162.4)^2}{235144} = 0.1122$$

$$\text{残差} = 1.2542$$

分散分析は10-10表で与えられた形で行う。

10-10表

あてはめの個々の次数に対する分散分析

|    | SS       | DF | MS       | F     | 5%点  |
|----|----------|----|----------|-------|------|
| 一次 | 1874.4   | 1  | 1874.4   | 12.0  | 5.32 |
| 二次 | 1798.066 | 1  | 1798.066 | 114.7 | 5.32 |
| 三次 | 3390.6   | 1  | 3390.6   | 21.6  | 5.32 |
| 四次 | 16.1856  | 1  | 16.1856  | 10.3  | 5.32 |
| 五次 | 0.1122   | 1  | 0.1122   |       |      |
| 残差 | 1.2542   | 8  | 0.1568   |       |      |

この分析から四次の多項式があてはめられなければならないことは明らかである。多項式の値を与える最も簡単な方程式は

$$Y = A + B\xi_1' + C\xi_2' + D\xi_3' + E\xi_4' \quad (19)$$

したがって

$$Y_F = 4.921 + 0.04538(-13) - 0.4970(13) - 0.005905(-143) + 0.01090(143) = 0.273$$

そして、 $\xi_1'$   $\xi_2'$   $\xi_3'$  および  $\xi_4'$  の値を更に代入して、多項式の値の完全な組を最後の欄に与えたようにして得る。

多項方程式はXの中の項で表わすことができるが、すこし計算の労力を必要とする。まず第一に  $\xi_1'$   $\xi_2'$   $\xi_3'$  および  $\xi_4'$  はXの中であらなければならない、こゝにXはいつものように  $(X - \bar{X})$  を表わす。変換の方程式は(18)に与えてある。これらを用いて、 $\xi_1' = \xi_1$  ということに注意しなければならない、そして  $\xi_2'$  の他の値は次のようである。

$$\xi_1' = 2\xi_1 = 2X$$

$$\xi_2' = \frac{1}{2}\xi_2 = \frac{1}{2}(X^2 - 16.25) = \frac{1}{2}X^2 - 8.125$$

$$\xi_3' = \frac{5}{3}\xi_3 = \frac{5}{3}(X^3 - 29.05) = \frac{5}{3}X^3 - 48.4167X$$

$$\xi_4' = \frac{7}{12}\xi_4 = \frac{7}{12}(X^4 - 41.0714X + 195.3482)$$

$$= \frac{7}{12}X^4 - 23.9583X^2 + 113.9531$$

(19)に代入すると

$$\begin{aligned} Y &= 4.9214 + 0.04538(2X) - 0.4970\left(\frac{X^2}{2} - 8.125\right) - \\ &\quad 0.005905\left(\frac{5}{3}X^3 - 48.4167X\right) + 0.01090\left(\frac{7}{12}X^4 - \right. \\ &\quad \left. 23.9583X^2 + 113.9531\right) = 10.2016 + 0.3767X - \\ &\quad 0.5096X^2 - 0.009842X^3 + 0.006358X^4 \end{aligned}$$

この方程式はXの平均値からの偏差を代入してYの値の推定量を得る。すなわち

$$\begin{aligned} Y_F &= 10.2016 + 0.3767(-6.5) - 0.5096(42.25) \\ &\quad - 0.009842(-274.625) + 0.006358 \\ &\quad (1785.0625) = 0.27 \end{aligned}$$

必要ならば、我々はXの代りに  $(X - 7.5)$  を代入できる、そして簡単にXが直接代入できる方程式を得る。