1

土木技術者のための 新数学講座

土木計画のための確率・統計

その1／吉川和広*

1. はじめに

土木工学のための数学講座を開設するにあたり，私に与えられた課題は今月号と来月号の2回にわたって確率・統計を紹介することである。

土木学会誌今月号においては，土木計画という立場から確率・統計を眺め直したときの問題点を指摘することにより，計画における数理統計学の位置づけを試みる。次いで統計学の基礎概念を明らかにする。さらに土木計画のための統計調査の考え方を述べ，最後に統計調査の手法を紹介する。来月号においては，土木計画のための統計的予測と統計的決定について，その考え方と手法の概要を紹介する予定である。

一口に統計学といっても，人によってその内容の理解の仕方にかなりの差異がある。このことは，その学問の成立史に負うところが大きく，統計的研究自体が統計学の内容をなすものか，あるいは研究方法がその内容をなすものかについて，統計学の学問的性質に議論が分かれてきた。社会統計学派と方法論派がすなわちこれである。社会統計学派を組織づけ体系化したのはマイヤ（G.v. Mayr）であり，方法論派，なかんずく数理統計学の定礎者はピアソン（K. Pearson）であった。これは19世紀後半から20世紀前半にかけて近代統計学のたどった2つの道であるが，以後，統計学は概念の変革により，機能概念を確立した後者の方向に走ったのである。

われわれが土木計画面にあらわれる現象の解明を確率・統計によって行なおうとするとき，まず第一に土木計画における統計学の位置づけから初めなければならないだろう。すなわち，計画科学の方法論としての統計的研究法と，他の実質科学としての計画科学との関係を明らかにしていくことが重要である。この場合，因果的法則の確立のための統計的推理を展開したカゥフマン（A. Kaufmann）の理論は多くの示唆を含んでいると考えられる。

カゥフマンは，すべての統計理論を数学的確率論の流出物（Ausfluss）とみなすという立場にたち，実質科学としての統計学を主張したマイヤに対し，方法論すなわち補助科学としての統計学を主張した。カゥフマンは数理統計学の提案する公式の形式的・機械的な適用については，これを深く戒めるとともに，実質科学との密接な関連において統計方法を展開することの必要性を強調している。この意味において，カゥフマンの理論は，純数理的解析を説く数理統計学の抽象性・形式性とは区別されるべきであると考える。

周知の通り，土木計画における計画科学は経験科学として，自然および社会現象に関する法則性を明らかにすること，および計画のための方法選択に至る過程を科学的に組織化することであるといえよう。したがって，計画科学の一研究方法としての統計方法もその任務を計画科学と共有するものでなければならず，またそれの定立する統計的法則は単なる関数関係を示すものであってはならないだろう。土木計画面に現われる現象の混沌とした偶発的事象のなかに合法則性を発見するための研究方法としての統計方法は，そのために現象の因果関係を究明することを課題としなければならないだろう。

このようにみてくると，以下に紹介する数理統計の核心をなしているものは「現象解析のための方法論」であるということができよう。したがって統計理論は非常に実際的なものでなければならない。現実に有用なものでない限り，全く無用である。理論と実際とが，一貫した軸の上にのって，実際が理論を発展させ，理論が実際を解明していくという一体的な関係が考えられねばならない。

このように考えてくると，統計学の根本思想は林 知巳夫によってまとめられているとおり

① 常に現象解析ということを第一義と考え，常に根源にさかのぼって本質的に考えを進めることを重視する。実質科学との関連を密にする。

② 現象解析の妥当性，有効性を根本にして考えを進める。

③ 逐次近似の考えをとる。

④ 常に問題発見の形で考えを進める。

これに基づいて，発展される統計学の内容を具体化すれば「いろいろの現象をフォーミュレートし，それに対して，調査・実験を加え，現象を表現し，数量化し，解析し，推論し，予測する。こういったことをいかにすればよいか」という方法を含んでいるとみることができよう。

* 正会員 工博 京都大学教授，工学部土木工学科

2. 統計学の基礎概念

(1) 集　　団

統計学は集団を取り扱い，また集団を背景としてものを考える。もちろん，ここでは，いろいろの現象を「集団の現象とみなせるようにして取り扱う」ということも含んでいるのを忘れてはならない。集団とは，即物的には，2つ以上の要素からなるものということができる。そして，どのような要素のとらえ方をするかが，問題のフォーミュレーションにあたって重要である。さて，背景としての論理的集団とは，確率という考えを積極的に利用し，母集団という概念によって表現されるものである。

(2) 操作的な考え方について

統計学の大切な考え方である。測定なくして，標識(mark) はないのである。「われわれの測定」によって「標識」をわれわれが与えるのである。即物的に測定され得ぬものは取り扱わないし,測定操作を媒介せずして，標識は存在しないと考えるのである。この操作という意識がないと，統計の嘘といわれるものがでてくるのである。いわゆるオペレーショナリズムの考え方との関係はきわめて深いというべきであろう。

(3) 統計における二つの立場

統計における一つの立場は，取り上げられたいくつかの計測値群の似たものをあつめて分類することである。この場合似ているということを統計的にきちんと表現しておかなければならない。似ているということを観念的にいうことはできるが，これを操作的に一義的に定義づけることはできない。操作的にはいくつかの定義の仕方があろう。このうちで，「よさそう」なものを決断をもって選ぶのである。よさそうとは，その定義が可能な限り客観性をもち，かつ統計分析の処理が妥当性を示しつつ，容易であることを意味するのである。この定義にしたがって，計測値の同じ傾向を示すものを集めて分類を行なう。

もう一つは，われわれの知りたいこと，それを知ることが妥当性のあることを端的に指し示すもの，目的とするものを直接的に指し示すものであるといえよう。このためには，分類のわかっていない似た現象に対して，外的基準がないとして，ある仮説にしたがって類似度を設定し，計測値パターンの同じものを集めるような統計的操作を行なってみる。そして，これによる分類がすでに得られている分類に一致しているとすれば，この類似度の定義,統計的操作には妥当性があるということになる。

以上二つの立場の特色と使いわけを述べてきたが，統計的方法はわれわれの目的と相対的な関係にあるので，この二つの立場を峻別しつつ，統計的方法をつくり，かつ用いなくてはならないのである。

(4) 標識づけについて

集団を取り扱うときに，標識をつけることが第一に重要であり，これは広義の測定によってなされる。標識は目的に応じて相対的に与えられるべきものであり，目的が変わればたとえ同一のものであっても，与えられるべき標識は変わってくる。標識は，そのものに内在する「与えられているもの」ではなく，われわれが与えるものなのである。

(5) 標識に差異のあるものの取り扱い

標識を与える，つまり測定をすることの目的は，集団の要素を弁別するためである。すなわち，要素の個々が弁別されるように測定を行なうのであるが，その結果同一の標識を持つもの，持たぬものが混在してあらわれるのが一般である。この不均一性をあらわすのに，いろいろの物差しが考えられているが，もし標識が数量である場合には分散という量 σ^2 が考えられる。

$$\sigma^2=\sum_{i=1}^{N}(X_i-\overline{X})^2/N \quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots(2.1)$$

ここに X_i は i の標識であり，$\overline{X}$ は X_i の算術平均，Nは要素の総数である。こうした平均値と分散との関係は非常に大切な意味を持ち，これを結びつけるものとして，チェビシェフの不等式がある。

さて，集団の特性をあらわすのにいろいろな量が工夫されているが，統計量といわれるものがこれである。こうしたものも，確率論を媒介として，母集団の，あるいは標本の特性をあらわすものとして，推論の世界へと転化されていく。この転化させるところに統計学の興味ある考え方の進展があるわけである。

(6) データの取り方について

統計の基本的な考えの一つに，データをいかに能率よくとるかということがある。標本調査と実験計画法がこれに相当する。単にデータをとるのではなく，最後の解析の仕方，結論の妥当性まで見通した上で，どのようにしてデータをとったらよいかを考えなければならない。

このデータ獲得においては，偏見のない見方，現象の本質を見とおす活眼が必要であって，ここを軸として統計理論が発展しているのである。

(7) 数量化について

測定によって得られる標識には，数量のものと数量で

ないものとがある。これらは，測定方法の性格からいってそれぞれ特色をもつものであるが，これらをともに，あるアイテム・カテゴリー反応として把えることが重要である。このためには，測定誤差を考慮し，妥当なカテゴリーを設定する。このようにしてできたアイテム・カテゴリーに対して，目的に沿った数量化を考えることになる。つまり，アイテム・カテゴリーに対してメトリックを導入し，新しい空間をつくり出すことになる。こうすることによって，妥当性のある結論がみえてくるようにするのである。

3. 土木計画のための統計調査

土木計画のための調査の目的は，計画の方針なり方法を決めるのに，正確な判断をくだすのに役立つだろう情報を得るためであるということができる。

調査を企画するには，まずどのような情報を集めることが必要かということ，すなわち情報の質的な面の検討から始めなければならない。このためには，対象となっている計画の目的および立場の認識を通して調査目的を明確にすること，土木計画のプロセスを究明するのに必要な情報の質の検討を通して，調査対象と調査項目を明らかにすることが重要である。

次に，実際の調査を実施するにあたっては，ことごとく調査するか，対象の一部を抽出調査するかが問題となる。前者の調査を全数調査（census）といい，後者を標本調査（sampling）という。標本調査は次の場合に行なわれる。

① 全数調査が無意味なとき
② 日時・費用が制限されているとき
③ 全数調査が不可能なとき
④ 調査内容を豊かにしたいとき

標本調査にあたっては，われわれは次のことを考えねばならない。

① どのような方法で母数を推定するか。
② 最良の推定を行なったとき，どれほどの誤差があるか。
③ 許容誤差範囲で標本をどのように抽出し，標本を何個抽出するか。
④ 標本抽出調査にあたって費用・人員・時間がどれほど必要か。

などを総合的に考えなければならない。このため近来は誤差論を背景とした無作為抽出法(random sampling)が用いられている。

以上明らかにした，標本抽出法とならんで重要な統計調査の手法は実験計画法である。実験計画法の目的は，どのようにデータをとり，どのようにデータを分析したら間違った判断をする危険が少なくなるか，すなわち情報量の獲得効率をいかにして高めるかということである。またデータをとる場合，経費や判断の遅れによる損失があるが，どれくらいデータに経費や時間をかけたら，一番有利かなどについて研究することである。

実験計画法がいままでの調査・実験のやり方と異なる点は，次の4項目に大別される。

① コントロールできないいろいろな環境条件や，未知の原因による実験データに対するかく乱や影響の程度を，客観的に定量的に評価することが可能である。
② 直交表の発見と，その使い方の著しい進歩により，実験に必要な資材・日数などをあまり増さないで，多くの因子を同時にまたは互いに関連づけて研究することが可能となった。
③ 実験計画法では，いろいろな要因の影響度を実験データに基づいて定量的・実証的に評価することができる。すなわち分散分析・F検定や要因の寄与率の評価を行なうことができる。
④ 実験計画法では，計量値以外のあらゆる種類のデータに対しても，計量値の場合と同様なデータ解析が可能である。

以上で土木計画のための調査を企画するにあたっては，情報の質的な面での検討および量的な面での情報処理が重要であるということを明らかにしたが，次節以下においては，確率論および数理統計学を基礎とした土木計画のための調査の手法についてその概要を述べることとする。

4. 統計調査の手法

(1) 統計資料の整理法

一つの確立変数 X の分布を観察によって知ろうとするには，一般には可能なかぎり多くの試みを行なって，Xの出現値を見るのがよい。一つの変数 X に対して考えられるあらゆる試みに現われる値の全体を X の母集団(population)といい，その個々の値を個体という。Xの集団から取りだした個体列 $x_1, x_2, \cdots\cdots, x_n$ を標本(sample) といい，nをその大きさという。

表—1 は，都市内街路上の自動車の走行速度xの分布をみるため，500 台の自動車を調査した結果である。

表—1 では，変数の値をいくつかのクラスに分けて，各クラスに属する個体数が示されている。これらの個体数を各クラスの度数（frequency）という。各クラスにおける変数の代表的数値をクラスの標識（class mark）といい，通常クラスの中央の値をその標識とみなす。

表—1 自動車走行速度の度数分布表

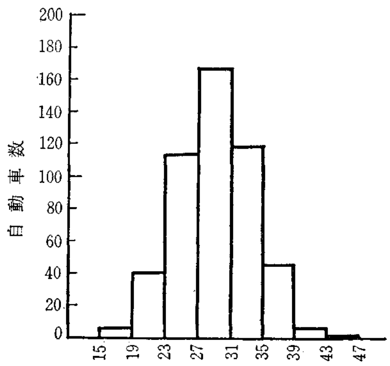
速 度 (km/h) x_i	自動車数 f_i	速 度 (km/h) x_i	自動車数 f_i
15～19	6	31～35	119
19～23	40	35～39	46
23～27	114	39～43	7
27～31	167	43～47	1

注：N=500

$x_1, x_2, \cdots\cdots, x_n$ をクラスの標識 $f_1, f_2, \cdots\cdots, f_n$ をクラスの度数，N を度数の合計とするとき，変数の分布状態を知るためには通常度数多角形，ヒストグラム，度数曲線等の度数分布図が用いられる。

ヒストグラム (histogram) とは，横軸上で各クラスに対応する区間上に，それぞれの度数に比例した高さの長方形を画いたもので，変数の分布状態の概略を知るために最もよく用いられている。図—1 は，表—1 からつくったヒストグラムである。

図—1 自動車走行速度のヒストグラム



統計資料の分布状態は度数分布図によって大体知ることができるが，この集団の特徴を表わす統計量として，分布の位置を示す代表値 (average)，分布の散らばりの程度をはかる変動 (variation) などが考えられている。

a) 代 表 値

代表値として最も普通に用いられるのは算術平均 (arithmetic mean)，幾何平均 (geometric mean)，中央値 (median)，および最頻値 (mode) の4つである。

算術平均： $\bar{x}=\dfrac{1}{N}\sum_{i=1}^{n} x_i f_i$ ························ (4.1)

幾何平均：$\log G=\dfrac{1}{N}\sum_{i=1}^{n} f_i \log x_i$ ·············· (4.2)

中央値：変数 X の数値のうちで，一つの値 m_e より大きいものと小さいものの現われる確率が相等しいとき，m_e を中央値という。

最頻値：変数 X が一つの値 m_0 の近くに最も頻繁に結集して現われるとき，m_0 を最頻値という。$f_1, f_2, \cdots\cdots, f_n$ のうちで最大な f_m に対応する変数の値 x_m.

b) 変 動

統計資料が代表値のまわりにどのようにちらばっているかの大体は，度数分布図などでわかるが，これを量的に表現するとき最も普通に用いられるのが標準偏差 (standard deviation) とその平方の分散 (variance) である。標準偏差 S_x は

$$S_x^2=\frac{1}{N}\sum_{i=1}^{n}(x_i-\bar{x})^2 f_i=\frac{1}{N}\sum_{i=1}^{n}x_i^2 f_i-\bar{x}^2 \quad \cdots\cdots(4.3)$$

以上，統計資料の整理法について述べたが，ここでいろいろな形の理論分布形について紹介しておこう。

1) 2項分布 (binomial distribution)

平均値 np，標準偏差 $\sqrt{npq}$ の2項分布は

$$P(x)=\binom{n}{x}p^x q^{n-x} \quad q=1-p \quad x=0, 1, 2, \cdots\cdots, n \quad \cdots\cdots(4.4)$$

2) ポアソン分布 (Poisson's distribution)

平均値 m，分散 m のポアソン分布は

$$P(x)=\frac{m^x e^{-m}}{x\,!} \quad x=0, 1, 2, \cdots\cdots(4.5)$$

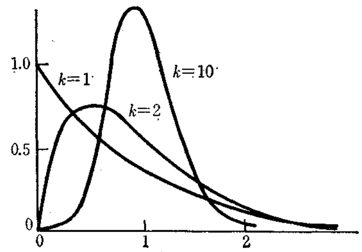
3) アーラン分布 (Erlang distribution)

位相 (phase) k のアーラン分布は

$$\left.\begin{array}{ll} P(x)=\dfrac{(\lambda k)^k x^{k-1}}{(k-1)\,!}e^{-\lambda k x} & x>0 \\ \quad =0 & x\leq 0 \end{array}\right\} \cdots\cdots(4.6)$$

平均値は $1/\lambda$，分散は $1/k\lambda^2$ である。

図—2 平均値1のアーラン分布



4) 正規分布 (normal distribution)

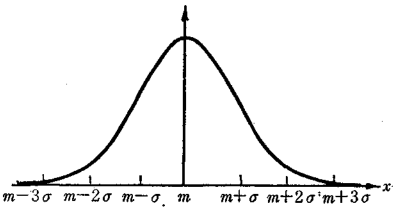
平均値 m，標準偏差 σ の正規分布は，

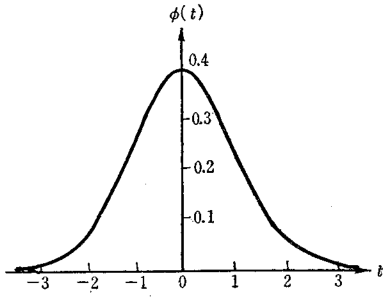
$$P(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-(x-m)^2/2\sigma^2} \cdots\cdots(4.7)$$

m, σ^2 の最尤推定値 $\hat{m}, \hat{\sigma}^2$ は

$$\left.\begin{array}{l} \hat{m}=\dfrac{1}{N}\sum_{i=1}^{N}x_i=\bar{x} \\ \hat{\sigma}^2=\dfrac{1}{N}\sum_{i=1}^{N}(x_i-\bar{x})^2=S_x^2 \end{array}\right\} \cdots\cdots(4.8)$$

図—3 正 規 分 布





いま

$$t=\frac{x-m}{\sigma}\quad\cdots\cdots(4.9)$$

とおくとき，変数 t の確率密度 $\phi(t)$ は

$$\phi(t)=\frac{1}{\sqrt{2\pi}}e^{-t^2/2}\quad\cdots\cdots(4.10)$$

で与えられる。この分布 $N(0, 1)$ を標準正規分布という。

5) 対数正規分布(logarithmic normal distribution)

対数正規分布は，正規分布の変数 x を対数変換したもので

$$P(x)=\frac{1}{\sqrt{2\pi}D(\log x)}\exp\left[-\frac{\{\log x-m(\log x)\}^2}{2\{D(\log x)\}^2}\right]\quad\cdots\cdots(4.11)$$

平均値 $m(\log x)$，標準偏差 $D(\log x)$ の最尤推定値は，

$$\left.\begin{aligned}m(\log x)&=\frac{1}{N}\sum_{i=1}^{N}\log x_i\\ D(\log x)&=\left[\frac{1}{N}\sum_{i=1}^{N}\{\log x_i-m(\log x)\}^2\right]^{1/2}\end{aligned}\right\}(4.12)$$

土木計画面にあらわれる代表的な理論分布形の概要は以上の通りであるが，ここで **表—1** に示した都市内街路上の自動車走行速度の度数分布をもとにして，これから標本平均および標本分散を求め，正規分布のあてはめを行なってみよう。理論度数 F_i を求めるための計算手順は **表—2** に示す通りである。

標本から母集団の形を想定し，その未知の母数についてある仮説をたて，その仮説のもとにこの標本の実現値が，どの程度の難易で出現するかによって，その仮説の採否を決定することを統計的仮説検定(test of statistical hypothesis) という。統計的仮説検定の手順は

① 仮説 H をたてる。

② 仮説 H のもとに事象 E の起こる確率 P を求める。

③ $P\leqq\alpha$ ならば仮説 H を捨てる。

④ $P>\alpha$ ならば仮説 H を捨てることができない。

ここに α は有意水準である。

いま実際の調査に基づくクラスの度数を $f_1, f_2, \cdots\cdots, f_n$ とし，これに基づいて仮定される母集団の分布法則から理論度数 $F_1, F_2, \cdots\cdots, F_n$ を求め

$$\chi^2=\sum_{i=1}^{n}\frac{(f_i-F_i)^2}{F_i}\quad\cdots\cdots(4.13)$$

を計算すると，この値は近似的に自由度 ν の χ^2 分布 (chi square distribution) に従う。ここに，

$$\nu=n-q\quad\cdots\cdots(4.14)$$

q は理論度数決定のための使用統計量の個数である。正規分布の場合には，統計量 N, m, σ を使用しているから $q=3$ となる。

さきに **表—2** で自動車走行速度分布への正規分布のあ

表—2 自動車走行速度分布への正規分布のあてはめ

速度 (km/h)	① クラスの標識 x_i	② 実現度数 f_i	③ x_if_i	④ $x_i^2f_i$	⑤ $t=\frac{x_i-m}{\sigma}$	⑥ $\Phi(t)$	⑦ $\phi(t)=\Delta\Phi(t)$	⑧ 理論度数 $F_i=N\phi(t)$
15～19	17	6	102	1 734	−2.141	−0.4838	0.0162*	8.1
19～23	21	40	840	17 640	−1.300	−0.4032	0.0806	40.3
23～27	25	114	2 850	71 250	−0.460	−0.1772	0.2260	113.0
27～31	29	167	4 843	140 447	0.380	0.1480	0.3252	162.6
31～35	33	119	3 927	129 591	1.221	0.3890	0.2410	120.5
35～39	37	46	1 702	62 974	2.061	0.4804	0.0914	45.7
39～43	41	7	287	11 767	2.901	0.4981	0.0177	8.9
43～47	45	1	45	2 025	3.742	0.4999	0.0018	0.9
計		N=500	14 596	437 428			0.9999	500

注：

平均値 $m=\frac{\Sigma x_if_i}{N}=\frac{14\,596}{500}=29.19$

標準偏差 $\sigma=\sqrt{\frac{\Sigma x_i^2f_i}{N}-m^2}=\sqrt{\frac{437\,428}{500}-(29.19)^2}=4.77$

$\Phi(t)=\int_0^t\phi(t)dt=\frac{1}{\sqrt{2\pi}}\int_0^t e^{-t^2/2}dt$

この値は正規分布の確率積分表から求める。

* 0.0162＝−0.4838−(−0.5000)

てはめを行なったが，その適合度を検定するための計算は表—3に示す通りである。$\chi^2=1.026$ となり，自由度 $\nu=7-3=4$

したがって，χ^2 分布表より $0.90<P<0.95$

よって走行速度が正規分布に従うという仮説は，有意水準10%をもって捨て去ることができない。したがって，自動車の走行速度分布は正規分布をなすといってよい。

このほか，一般に，統計的母数の推定や仮説の検定には F 分布や t 分布が用いられるが，ここでは紙数の関係上その説明を省略することとする。

表—3 自動車走行速度分布が正規分布に従うとしたときの適合度検定

クラスの標識 x_i	実現度数 f_i	理論度数 F_i	f_i-F_i	$\frac{(f_i-F_i)^2}{F_i}$
17	6	8.1	−2.1	0.544
21	40	40.3	−0.3	0.002
25	114	113.0	1.0	0.009
29	167	162.6	4.4	0.119
33	119	120.5	−1.5	0.019
37	46	45.7	0.3	0.002
41	7} 8	8.9} 9.8	−1.8	0.331
45	1	0.9		
計	500	500		$\chi_0^2=1.026$

（2） 標本抽出の方法

土木計画の問題においては，ある与えられた精度を得るためには，最小限どれだけの標本数が必要かということがしばしば問題となる。したがって，ここではいろいろな標本抽出のための統計手法を紹介することとする。

a） 無作為抽出法（random sampling）

変数 X の N 個の有限母集団から n 個の標本 x をとって調査したときの標本平均・不偏分散・母平均・母分散をそれぞれ $\bar{x}, S^2, m, \sigma^2$ とおけば

$$\left.\begin{array}{l}\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i \quad E(\bar{x})=m \\ S^2=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2 \quad E(S^2)=\frac{N}{N-1}\sigma^2\end{array}\right\}\cdots(4.15)$$

変動係数を $C=\sigma/\bar{x}$ とおけば，精度 α をもって母平均を推定するには

$$n\geqq\frac{N}{1+N(\alpha^2/C^2)}\cdots\cdots(4.16)$$

でなければならない。

b） 層別無作為抽出法（stratified sampling）

母集団が均一でなく，差異のあるいくつかのものに分かれているとき，いろいろの対象について等質の部分を一つの階層とし，異質の部分を他の階層になるように階層分けをする。

母集団 $X_1, X_2, \cdots\cdots, X_N$ を r 個の階層に分け，第 i 層の個体を $X_{i1}, X_{i2}, \cdots\cdots, X_{iN_i}$ とし，その個数を N_i 個，母平均を m_i，母分散を σ_i^2，この階層からの標本数を n_i 個，標本を $x_{i1}, x_{i2}, \cdots\cdots, x_{in_i}$，標本平均を $\bar{x}_i$，不偏分散を S_i^2 とおくと，

$$\left.\begin{array}{l}m_i=\frac{1}{N_i}\sum_{j}^{N_i}X_{ij} \quad \bar{x}_i=\frac{1}{n_i}\sum_{j}^{n_i}x_{ij} \\ \sigma_i^2=\frac{1}{N_i}\sum_{j}^{N_i}(X_{ij}-m_i)^2 \quad S_i^2=\frac{1}{n_i-1}\sum_{j}^{n_i}(x_{ij}-\bar{x}_i)^2\end{array}\right\}\cdots\cdots(4.17)$$

次に N 個の個体からなる母集団を r 個の階層にわけて，これから n 個の標本を抽出するのに各層からどのような割合で抽出すべきかが問題となる。この抽出法には比例抽出法と最適割当法がよく用いられる。いま抽出比率を k とおけば

① 比例抽出法の場合

$$k=\frac{n_i}{N_i}=\frac{n}{N} \quad i=1, 2, \cdots\cdots, r \cdots\cdots(4.18)$$

② 最適割当法の場合

$$k=\frac{n}{\sum_{i}^{r}N_iS_i}\cdots\cdots(4.19)$$

c） 集絡抽出法（cluster sampling）

N 個からなる母集団を M 個の集絡に分割し，第 i 集絡の合計を y_i として，抽出集絡 m 個より母平均 $\bar{x}$ を推定しようとするのが集絡抽出法である。いま，

$$\bar{Y}=\frac{1}{M}\sum_{i}^{M}y_i \quad \bar{y}=\frac{1}{m}\sum_{j}^{m}y_j\cdots\cdots(4.20)$$

とおけば，$\bar{x}=M\bar{Y}/N$ となり

$$S_y^2=\frac{1}{m-1}\sum_{j}^{m}(y_j-\bar{y})^2\cdots\cdots(4.21)$$

であるから，母平均 $\bar{x}$ を信頼度係数 u，誤差率 ε で推定するには

$$m\geqq\frac{M}{1+(\varepsilon/uC_y)^2M}\cdots\cdots(4.22)$$

でなければならない。ここに C_y は変動係数であり，$C_y=S_y/\bar{y}$ として求められる。

（3） 実験計画法

土木計画の問題においては，どのようにデータをとり，どのようにデータを分析したら間違った判断をする危険が少なくなるかということに常に留意しなければならない。このような問題を解決するための統計手法が分散分析を骨子とする実験計画法である。

a） 1元配置法（one way layout）

一つの因子 A が r 種類に分かれているとし，各 A_i に対する実験値を x_{ij} とする。そして m, a_i を定数として

$$x_{ij}=m+a_i+z_{ij} \quad i=1, \cdots\cdots, r;\ j=1, \cdots\cdots, n_i \cdots\cdots(4.23)$$

とおくとき z_{ij} は実験誤差とみなされるとする。すなわ

ち，z_{ij} は $\sum_{i=1}^{r} n_i$ 個の独立な変数で，いずれも正規分布 $N(0, \sigma^2)$ に従うとする。各 i, j に対して x_{ij} は $N(m+a_i, \sigma^2)$ に従うことになるが，この a_i が因子 A の要因としての A_i による効果を示しているわけである。ここで

$$\sum_{i=1}^{r} n_i a_i = 0 \quad \cdots\cdots(4.24)$$

と仮定してよい。また次の記号を導入する。

$$\left.\begin{array}{l} T_i = \sum_{j=1}^{n_i} x_{ij} \\ \bar{x}_i = \dfrac{T_i}{n_i} \quad i=1, \cdots\cdots, r \\ T = \sum_{i=1}^{r} T_i = \sum_{i=1}^{r} \sum_{j=1}^{n_i} x_{ij} \quad \bar{x} = \dfrac{T}{n} \end{array}\right\} \cdots\cdots(4.25)$$

このとき

$$\sum_{i=1}^{r} \sum_{j=1}^{n_i} (x_{ij} - \bar{x})^2 = \sum_{i=1}^{r} \sum_{j=1}^{n_i} (x_{ij} - \bar{x})^2 + \sum_{i=1}^{r} n_i (\bar{x}_i - \bar{x})^2 \quad \cdots\cdots(4.26)$$

上式の左辺および右辺の第1，第2項にある平方和を，それぞれ全変動・級内変動・級間変動という。

さてわれわれは要因 A_i に差別がないという仮説

$$H : a_1 = \cdots\cdots = a_r = 0$$

を検定する方法を見出したいのであるが，n_i を一定にたもつと，仮説 H のもとでは

$$F = \frac{\sum_i n_i (\bar{x}_1 - \bar{x})^2}{r-1} \Big/ \frac{\sum_i \sum_j (x_{ij} - \bar{x}_i)^2}{n-r} \quad \cdots\cdots(4.27)$$

は自由度 $n_1 = r-1$, $n_2 = n-r$ の F 分布に従う。

以上を分散分析表にまとめると，**表—4** に示す通りである。

表—4

要因	平方和	自由度	平均平方
級間変動	$\sum_i n_i (\bar{x}_i - \bar{x})^2$	$r-1$	$\frac{1}{r-1} \sum_i n_i (\bar{x}_i - \bar{x})^2$
級内変動(誤差)	$\sum_i \sum_j (x_{ij} - \bar{x}_i)^2$	$n-r$	$\frac{1}{n-r} \sum_i \sum_j (x_{ij} - \bar{x})^2$
全変動	$\sum_i \sum_j (x_{ij} - \bar{x})^2$	$n-1$	

b) 2元配置法 (two way layout)

以下紙数の都合上分散分析表のみ示すにとどめ，くわしくは参考文献にゆずることとする

表—5

要因	平方和	自由度	平均平方
行間変動 (A)	$s \sum_{i=1}^{r} (\bar{x}_{i\cdot} - \bar{x})^2$	$r-1$	$\frac{s}{r-1} \sum_{i=1}^{r} (\bar{x}_{i\cdot} - \bar{x})^2$
列間変動 (B)	$r \sum_{j=1}^{s} (\bar{x}_{\cdot j} - \bar{x})^2$	$s-1$	$\frac{r}{s-1} \sum_{j=1}^{s} (\bar{x}_{\cdot j} - \bar{x})^2$
誤差変動	$\sum_{i=1}^{r} \sum_{j=1}^{s} (x_{ij} - \bar{x}_{i\cdot} - \bar{x}_{\cdot j} + \bar{x})^2$	$(r-1)\ (s-1)$	$\frac{1}{(r-1)(s-1)} \sum_{i=1}^{r} \sum_{j=1}^{s} (x_{ij} - \bar{x}_{i\cdot} - \bar{x}_{\cdot j} + \bar{x})^2$
全変動	$\sum_{i=1}^{r} \sum_{j=1}^{s} (x_{ij} - \bar{x})^2$	$rs-1$	

c) ラテン方格法 (Latin square)

表—6

要因	平方和	自由度	平均平方	分散比
行間変動 (R)	$r \sum_{i=1}^{r} (\bar{x}_{i\cdot} - \bar{x})^2 = \frac{1}{r} \sum_{i=1}^{r} T_{i\cdot}{}^2 - \frac{T^2}{r^2}$	$r-1$	$\frac{r \sum_{i=1}^{r} (\bar{x}_{i\cdot} - \bar{x})^2}{r-1} = V_R$	$\frac{V_R}{V_E}$
列間変動 (C)	$r \sum_{j=1}^{r} (\bar{x}_{\cdot j} - \bar{x})^2 = \frac{1}{r} \sum_{j=1}^{r} T_{\cdot j}{}^2 - \frac{T^2}{r^2}$	$r-1$	$\frac{r \sum_{j=1}^{r} (\bar{x}_{\cdot j} - \bar{x})^2}{r-1} = V_C$	$\frac{V_C}{V_E}$
処理間変動 (M)	$r \sum_{k=1}^{r} (\bar{x}_{(k)} - \bar{x})^2 = \frac{1}{r} \sum_{k=1}^{r} T_{(k)}{}^2 - \frac{T^2}{r^2}$	$r-1$	$\frac{r \sum_{k=1}^{r} (\bar{x}_{(k)} - \bar{x})}{r-1} = V_M$	$\frac{V_M}{V_E}$
誤差変動	$\sum_{i=1}^{r} \sum_{j=1}^{r} (x_{ij(k)} - \bar{x}_{i\cdot} - \bar{x}_{\cdot j} - \bar{x}_{(k)} + 2\bar{x})^2$	$(r-1)\ (r-2)$	$\frac{\sum_{i=1}^{r} \sum_{j=1}^{r} (x_{ij(k)} - \bar{x}_{i\cdot} - \bar{x}_{\cdot j} - \bar{x}_{(k)} + 2\bar{x})^2}{(r-1)(r-2)} = V_E$	
全変動	$\sum_{i=1}^{r} \sum_{j=1}^{r} (x_{ij(k)} - \bar{x})^2 = \sum_{i=1}^{r} \sum_{j=1}^{r} x_{ij(k)}{}^2 - \frac{T^2}{r^2}$	r^2-1		

図—4 ラテン方格割付け表

	C_1	C_2	C_3	C_4
R_1	M_1	M_2	M_3	M_4
R_2	M_2	M_1	M_4	M_3
R_3	M_3	M_4	M_1	M_2
R_4	M_4	M_3	M_2	M_1

表—7

要　　因	平　　方　　和	自　由　度
行間変動　(A)	$ns\sum_{i=1}^{r}(\bar{x}_{i\cdot\cdot}-\bar{x})^2$	$r-1$
列間変動　(B)	$nr\sum_{j=1}^{s}(\bar{x}_{\cdot j\cdot}-\bar{x})^2$	$s-1$
交互作用による変動 ($A-B$)	$n\sum_{i=1}^{r}\sum_{j=1}^{s}(\bar{x}_{ij\cdot}-\bar{x}_{i\cdot\cdot}-\bar{x}_{\cdot j\cdot}+\bar{x})^2$	$(r-1)\ (s-1)$
誤　差　変　動	$\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{n}(x_{ijk}-\bar{x}_{ij\cdot})^2$	$rs(n-1)$
全　　変　　動	$\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{n}(x_{ijk}-\bar{x})^2$	$rsn-1$

d)　交互作用のある場合の実験計画法

2つの因子 A, B があって要因 (A_i, B_j) に対して n 個ずつの独立な実験値を求める。それを

$$x_{ijk}=m+a_i+b_j+c_{ij}+z_{ijk}$$

$$i=1,\ \cdots\cdots,\ r\ ;\ j=1,\ \cdots\cdots,\ s\ ;\ k=1,\ \cdots\cdots,\ n \qquad \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots(4.28)$$

とおくとき，rsn 個の z_{ijk} が独立にいずれも $N(0,\sigma^2)$ に従うと仮定する。ここに a_i は因子 A の主効果，b_j は因子Bの主効果，c_{ij} は因子 (A, B) の間の交互作用 (interaction) を示している。

われわれは3種の仮説

$$H_1: a_1=\cdots\cdots=a_r=0$$

$$H_2: b_1=\cdots\cdots=b_s=0$$

$$H_3: c_{ij}=0$$

の検定法を考えるのであるが，分散分析表をあげるにとどめることとする。

参　考　文　献

1) 足利末男：統計学と社会，ミネルヴァ書房，昭43.7
2) A. Kaufmann : Theorie und Methoden der Statistik, Tübingen, 1913
3) G.v. Mayer : Statistik und Gesellschaftslehre, Erster Band, Zweite umgearbeitete und Vermehrte Auflage, Tübingen, 1914
4) 林知己夫：統計学の基礎にあるもの，思想 No. 513, 岩波書店，昭42.3, pp. 69〜82
5) 吉川和広：土木計画と OR, 丸善，昭 44.7
6) 奥川光太郎：数理統計学概説，学術図書出版社，昭 37.2
7) 森村英典・大前義次：待ち行列の理論と実際，日科技連，昭40.7
8) 岩井重久：Slade 型分布の非対象性の吟味およびその2，3の新解法，土木学会論文集第4号，昭 24.6
9) 米谷栄二・定井喜明：交通工学のための推計学，国民科学社，昭 41.4
10) ウィルクス・小河原正巳訳：数理統計学，春日出版社，昭 31.4
11) 石川栄助：実用近代統計学，槙書店，昭 36.4
12) 田口玄一：実験計画法，丸善，昭 31.8
13) 増山元三郎：実験計画法，岩波書店，昭 38.6
14) 北川敏男・増山元三郎：新編統計数値表，河出書房，昭 27