森林標本調査法の基礎

大友栄松

昭和38年2月

農林省林業試験場経営部

目　次

まえがき

本稿は　これまでの国有林の森林資源調査の講義をもとにして，今回の研修にあたりこの数日間で急場作成したので，一部前年の原稿もあり，本年新に書きなおしたものもあり，前後の統一を欠く感があり，記号も一貫性がないかも知れない。印刷前に一度もよみ返す時間が全くなく，それに講義の全部にわたっておらず一部は章名にのみヒダめ，講義の際に補なわざるを得ない有様で恐縮している次第である。

この稿の中で〔付〕とか〔注〕をつけてあるのは，興味のある人達は読んで欲しいが，読まなくとも全体を理解するには差し支えない所で，むしろ初心者は飛ばした方がよい所である。

サンプリングは難しいということをいう人もあるが，決して難しいものではなく，割れいばむしろやさしいものである。そのためにできるだけ，証明などの数学的事項はさけるか〔付〕や〔注〕にまわした。しかしまだ不徹底な嫌いがあるので，将来かき直す積りで未定稿としたわけである。

1. 森林資源調査法

2. 調査における種々の誤差

どのような調査でも，色々の原因による誤差は必ずあるものである。誤差のない調査というものはあり得ない。それがために，すべての調査は無価値とはいえないのである，問題は誤差の種類とその大きさである，たとえば，ランダムな誤差は，互いに打ち消し合う性格をもち，偏よりは，その偏よりを把握できれば，取り去ることができる，ある樹の高さを測定するとき，正確な測高器を用いても，測定の都度，若干の異る値が得られる。しかしこれは偶然的なもので，真値に対する誤差は互いに打ち消しあうものである，しかし測高器が不正確で常に一定量だけ過大に見るというならば，この偏よりは，調査して，その分だけ差し引けばよいであろう，また単に傾向だけを見たい場合は，偏りがあっても差し支えない，

したがって調査にあたっては，誤差の種類や大きさ，原因等に十分留意して，調査の結果が目的に合致したものになるよう努力しなければならない，徒うに調査の正確度をあげようとすると莫大な費用と時間を要するから，常に調査の目的に最適な正確度をもつよう調査の設計を行はなければならない，

誤差には次のものが考えられる

2/ 母集団，調査対象を明確に定義しなかった場合

調査対象地の境界が明確に定められていないときは，全林毎木調査の場合でも測りすぎ，測定もれを生ずる，令級別に面積

蓄積を知りたいときは、調査簿や林相図にあやまりがあった場合にも、得た値に誤差を生ずる。このような誤差は　もとの図簿を修正しないと消すことは困難である。

2.2　調査説明書や定義のあいまいさに基く誤差

説明書や定義があいまいだと、各人各様の判断を下し、例えば、全国の森林資源調査の場合林地における直径 6cm の針葉樹とか直径 4cm の広葉樹のみにする場合、はっきりとその旨を述べておかないと、庭園樹をも調査したり 2cm 位いのものでも木だからといって測定したりする人も出てくる。このような定義や説明書の曖昧さから来る誤差は、説明書を作る際に注意すればさけられる。

2.3　測定誤差

この場合は　2種類の誤差が入ることが考えられる。1は、測定技術の拙劣等による入るもので、偶然的なものである。これは熟練や入念等により、その大きさを減ずるものであり　一般に正負両方の誤差を生じ相互に打消すものと考えられるからさほど問題にならない。もう一は、正なら常に正、負なら常に負という風に、例えば下駄をはかって身長をはかる場合、如何に入念に測っても、常に下駄の高さ丈、過大に測定するというような場合の誤差である。これを一般に偏り（偏倚）というが林業の場合は輪尺測定のとき、或は巻尺で直径を測るときに見られる。このような誤差は systematic error というが、こ

の error を生じないように常に器材を点検しておかなければならない。普通、標本調査理論では　偶然的な誤差を取扱い、systematic error は研究が十分ではない。

2.4　調査対象の変動による誤差

調査期間が長く、例えば　苗畑や幼令造林地等を、春から秋にわたって調査したような場合は、幼樹や苗木自身の生長の変化により変動を生ずる。とくに苗木は　生長期間内では1月といわずに多大の生長をするので、調査には慎重を要する。

2.5　調査洩れや重複調査による誤差

全林毎林調査でも　崖地などの林木は調査をしなかったり、あるいはいい加減に目測したりすることもあるし、境界の場合に隣接小班の木まで調査したり　あるいは境界木を落したり、また、標本調査では　通常の不便の地を故意に調査しなかったりすることがある。このようなときは、必ず系統的な誤差を生ずる。

2.6　誤記、誤算による誤差

毎木調査の場合でも標本調査の場合でも、注意しないと誤記や、誤算を生ずることがある。これらは、注意しさえすれば、除去することができる。

2.7　推定方法による偏より

取りまとめのとき、分類をあやまったり、あるいは、括約する巾が大きすぎたりしたためとか、1a当りの平均を算出すると

き加重平均すべきものを単に算術平均したりするときには偏よりを生ずる．特に　令林の平均林令などを出すときなどは，目的により　色々の方法があり，あやまった方法だと偏よりを生ずる．

2.8．標本調査に伴う誤差

以上の誤差の外に標本調査では，有意であろうとランダムであろうと必ず標本であるための誤差を伴う．ランダムな誤差はこれを評価し，コントロールする方法はあるが，有意標本の場合は，それができない．

有意標本では　特に標準地や標本木の選定の偏りというものを生ずる．これは，調査者の性格やその日の気分にも支配される．ある人は常に過少となるように標準地を選定し，ある人はその反対でもある．それが，必ずしも一定だけ過少，過大でなく　その日の気分に支配されることもある．従ってその誤差を評価し　コントロールする手段はない．もちろん　ランダムなサンプリングでも，調査地を測定し易いように　ずらしたりすると偏りを生ずることもあるし，調査困難地を省いたりするために偏りを生ずることもある．

次に両者とも推定方式による偏りを生ずることがある．有意標本から全林材積を推定するとき，常に一定だけ割引したりするとき，任意標本で誤まった公式を用いたりするとき　偏りが入るし割合を推定して　それから母集団の特性値を推定する場合は偏りがはいってくる場合もある．

しかし，あとの場合は　実用上問題にならぬことが多い．

3　全数調査と標本調査

標本調査と全数（悉皆）調査と何れがよいかということは，屢，論議されることであるが，これについては　直ちにどちらがよいとは断定できない．寧ろ　行なおうとする調査の目的によって定まって来るものである．今どのような場合に標本調査がセンサス（悉皆調査）にまさるか，あるいは得策であるかについて述べて見よう．

3.1．結果を出すのに日時，費用の制限がある場合

例えば，一定期間に，事業区全体の林木蓄積を知りたい場合，林業政策樹立のためや，未開発林の利用のためにその国内または一地方の森林の蓄積を知りたい場合は　日時，費用，労力の制限のため，到底全数調査は不可能であり　一方これらの目的達成のためには何万何千何百何十何立方米までも知らなくとも概数を知り，その誤差の程度を知り得れば十分目的を達成できる．もちろん　このような場合，全数調査を行うことができればよいが，莫大な費用と莫大な労力を要し，更にもし日数に制限なく行うとしても　その間に林木は成長し，その調査結果を集計しても，成長による変動が入り，正しい結果は得られないだろう．このようなわけで　林業でも標準地調査法が　必然的に発達したのであるが，標本調査理論は，確率論の力をかりて

別途に発展し合理的信頼性のある知識を与えてくれるものである。尤も林業においては標出抽出調査のひこばえは一般統計分野より早く導入され、既に昭和ノ2年糸井省吾氏により紹介されている。（雑誌御料林昭和ノ2年　月号）。いづれにしても、上に述べたように、日時、費用、労力に制限ある場合、調査結果がそれほど正確なものを要求されない場合、調査対象が時と共に変化するような場合等は、サンプリングによらないといけない。

調査というものは常に、その目的、調査対象に応じ、常に無駄のない、合理的なものでなければならない。

3.2 悉皆調査より事実上、正確な、偏りのない結果を得たい場合

われわれは、個々のものをすべて調べあげて、寄せ集めれば必ず全体に関して正確な知識を得られるとよく思う。この考えの中には個々のものの測定が只一つのものの測定と同様な正確さで行われ、かつ測定者が、皆同じように熟練しているという前提が内在している。しかし、全体の数が非常に大きくなった時、果してそうであろうか。同じような仕事を何百回と続けて行く場合は疲労、あるいは飽きる等の心理作用で測定の正確さが減じ、誤差が入ってくるものである。一方、更に同じように測定技術の練達の人が多数得るということは困難である。従って悉皆調査の場合はどうしても未熟な調査員にまかせなければならない。これらの場合の誤差は次々に累積されて行き非常な

大きなものになり見逃すことはできなくなって来る。特に悉皆調査の時は、測定そのものの拙劣から来る誤差の外に意識的無意識的な調査洩れもある。このように考えて来ると、悉皆調査必ずしも最良とはいえなくなる。むしろ、サンプリングにより少数の熟練者により正確に測定し、合理的な信頼性のある結果を出した方が、経済的にも、調査結果の早急な利用という点から見てもサンプリングの方が有利になって来る。サンプリングでは、調査数も少いので熟練者を多数要することもなく、あきてくるという心理作用も遥かに少くなる。また、調査結果の集計作業も、悉皆調査では厖大なため誤りの入る可能性が、多いが、サンプリングでは、比較的少数だから誤りの入る可能性も少ない。

しかし、サンプリングでは、少い調査結果を全体に拡大するから、調査は入念に正確に行なわなければならない。この点は特に注意すべきで、従来の全林毎木調査のような態度で行なってはならない。要するに、全部を調べれば、全体に関する正確な知識が得られるという考え方は改むべきである。

3.3. 悉皆調査が不可能なとき

工業では電球の寿命検査をしたいときは、全数を調査することは、不可能である。何んとなれば、全部を試験すれば、皆、電球は切れてしまって売るものはなくなってしまう。林木でも成長量を知りたいとき、全林木を樹幹解析を行なうことは不可

能で少なくとも成長錐調査ですら全林木に行なうことは不可能になって来る。このようなときはサンプリングによらなければならない。

3.4 以上述べたことから、直ちに標本調査が全数調査に勝ると判断するのも早計である。場合によっては、全数調査がよい場合もある。前に述べたと同様に調査の目的、調査の対象、経済的な制約、社会的な制約等を十分考慮して、それらに応じた調査方法を取るべきである。

数年前、ドイツ等の中欧諸国で、全林調査か、標本調査か何れがよいかを Prodan や Wech 等により論議されたが、当時標本調査に批判的であった Prodan は最近はむしろ賛成者の立場に立って来ているが、まず著者の見解をのべ更に当時の彼らの論ずる所を次に述べて見る。

3.4.1 前林毎木調査は、費用、労力、日時等の制約のため、事業区全部の森林について行なうことができない。従ってせいぜい第一分期指定林小班において行われるにすぎない。日本のように皆伐作業の多い林ではこの調査は余り有用ではない、何となれば、第一分期指定林小班は、伐採前に必ず収穫調査を行なうために、同一林小班が、僅かな期間内に二度も調査されることになり、経済的にも有利でない。その上、林木は皆伐されるので、次期の全林毎木調査個所は異る林地で行われるので照査法のような成長量の計算もできない。

この点、何かしら無駄な調査を行なっているような気がする。

3.4.2 上のように全林毎木調査はよいとはわかっても費用、労務、日時の関係から事業区の一部において実行されるにすぎずその残りの林分に対しては、目測や標準地による調査を行なわなければならず、従って事業区全体の蓄積成長量がどの位真値と異なっているか皆目見当がつかずに、諸種の計画を立案しなければならない。このようなことでは、保続の計算、伐採量の決定も極めて不安なものとならざるを得ない。

3.4.3 Prodan の見解

Prodan は森林調査法は森林経理に結びついているものであるから、採用する森林経理法によってきまってくると述べ、まず中欧諸国の森林経理法の特色を説明している。中欧諸国の森林経理はますます集約化をたどり、施業の単位はあくまでも林小班単位をとり、それも分班などというものも作られ、次第に小面積に移りつゝある。一方最近盛んになって来た単木施業は、やはり林分（林木）という枠内で考えられるべきで、林小班単位の施業に変りはない。北欧や北米のサンプリングは、大面積に対する概況な結果が必要で、小面積のそれは無意義だから採用されているのである。このような観点からは、中欧の森林経理ではこの方法を導入する何らの動機もないわけである。

森林調査法では、林分と目的に応じて調査の集約度（調査労力）を変えるべきだが、これについては、サンプリングでは及

化という方法を取っている．在来の調査法でも才一分期林分では毎木調査．幼壮令林分では目測調査や収穫表による材積査定あるいは代表的調査法によって調査しているから，この点については何ら問題はないわけである．たゞ，大きな欠点と思われるのは，令級の移動により経理期間ごとに，同一の林分は毎木調査を行なうことはなく，常に別の林分を調査しなければならないため，調査結果の比較が困難となり，収益性の計算に支障を来すことである．さらに中欧においては新しい林業の課題として，Standortkartierung, Betriebstypenfeststellung, dynamische, Bonitierung, Waldbewertung 等が起って来ているので，これらを考慮すると Lestungskontrolle が最適であろう．しかし，サンプリングについては，従来のドイツの研究を検討して見ると，ドイツの森林経理の特色である林小班単位の施業を生かすため，従来 10 〜 30 %行なわれた全林毎木調査を 10 %位にとゞめ，その残りに対してサンプリングを行なうことにすれば，森林経理上必要な要件はすべてみたされるから経費も 40 〜 70 %減少して来る．特にこの際ビッテルリッヒ法も併用すれば有利である．

3.4.4．以上の主旨を Prodan は 1955 年発表しているが，さらに彼は 1958 年追加補足した意見を発表している．それを次に紹介して見よう．サンプリングは北欧，北米の特有のものではない．Zetsche が既に 1770 年に発表し，Hennabke

が 1790 年寺院有林の経理業務に関連して Württemberg の J.F. Schultheiss に提案している．しかし，これらはすべて実行されずに主観的な方法のみが盛んに行なわれるようになった．才二次大戦後は，サンプリングの導入のため，研究には 3 の進み方がドイツでは見られる．どの方法でも統計数理を応用する点では共通だが，実際の調査方法と森林経理の計画についての根本的な考え方が異るのである．

i　Lötsch の質的量的調査法

全森林あるいは作業級全体が一定の精度で調査されなければならぬとし，各林小班は，作業級内のブロックと考える．全体の数値を才一義とし，林小班は才二義と考える．この考え方は，北欧三国の大面積施業法や考え方に則している．総括的な方法でもあり，その証拠として北欧三国のすばらしい成果があげられる．この方法は，特に統計数理的方法と容易に結びつけられるのが妙味である．しかし，中欧の施業は，個々の林分の考察をもとにし，個々のものから全体に及ぼすというような計画だから，この方法では中欧の林業は逆行する．

ii　Richter の東独の方法

中欧の従来の方法と大面積的考察法の中間の方法である．この方法では，時には実行者により個々の林分のデーターは別個に査定される．

iii　Arbeitkreis für forstliche Biometrie の方法

林小班は造林法や施業法の対象であって，これらは森林経理を通じて林班のような個々のものを土台として全体に融合する調査法は全林毎木，サンプリング，目測を併用する．

この方法は進歩した中欧の森林経理に最適なものである．大面積的な調査法によると個々が見失なわれ，造林的施業的要求と矛盾してくる．従ってこれらの要求にサンプリングをどの程度に行い，如何に結びつけるかは今后の研究課題である．何れにしてもドイツの森林経理にサンプリングの導入は，多くの場合，同じ時間，同じ費用では精度の向上になるといっている．

3.4.5. *Weck* の見解

これに対し *Reinbeck* のドイツの林業試験場の J.Weck は1957年に次のような主旨の論文を発表している．戦后，林業の集約化に伴い，森林調査では従来の材積だけの調査の外に，生産手段としての立木蓄積の解明を一層切実に要求するようになったが，労銀の高騰のため森林調査は当面の目的にのみ十分な精度で計画と照査に必要なデーターの調査のみに限定せざるを得ない．従来の森林経理は採用する経理法により造林法を制約して来た．すなわち，令級面積の調整が主体で，それに1920年来全林毎木調査を附加して行なって来た．しかし労銀の高騰のため，全林毎木調査を断念して，以前の面積法に逆戻りして経営経済的に最大効果を発揮するような蓄積の造成を断念するか，良く設計されたサンプリングに移るか（このサンプリングは計画と照査に不可欠のデータのみに調査を限定し，必要精度で有効な比較を行なえるような方法）の何れかを選ばなければならない．

造林技術と無関係なあるいは制約しない森林調査の要件としては次のようなことが考えられる。

i 一定精度で施業単位について蓄積成長量を樹種別，径級別，材積級ごとに確立できるもの

ii 調査の精度をあとで調べることができるもの（客観的調査法が前提）

iii 精度と費用の関係を考えあわせることができるもの（施業に実際必要なデータのみに限定してむやみにくわしい精度を得ても損失である。）

このような観点から，全林毎木調査法，照査法，サンプリングについて比較検討して見よう．

全林毎木調査法は費用の点から全面的に行なうことは不可能で，一部は収穫表，一部は目測によらなければならない．従って，蓄積にせよ成長量にせよ，その誤差の範囲は不明である．全林毎木調査の場合は抽出誤差はないけれども *Knuchel* によれば 0～2%程度の測定誤差等もあるといわれている。その誤差は面積に比例して大きく，疲労による誤差（記帳者の誤り，調査もれ，測定の不正確）はサンプリングの場合よりも大きい．また，各調査班の精度のコントロールも困難であり，場合によ

っては，区域の不確定の時もあり，そのために誤差を生ずることもある．

照査法は全林毎木調査法の1種だが，16 cm 以上の木のみ測定して費用の低減をはかっている．この方法にも成長量の査定については次のような問題が残されている．

i 択伐された木は比較のためには，立木になおし，Silv. の単位で測定しなければならない．これには時間がかかり 困難な上，費用もかかる．

ii 枯損木と進界木の調査にはさけられない誤差を伴なう．このような誤差を考えないでも全林毎木調査の誤差は±2%もあり，照査期間内に10%も蓄積がふえると査定される成長量は約30%の誤差となる．

また，経済の観点から林分ごとの全林毎木調査の費用は，引きあうものかどうかを吟味する必要がある．

しかし，択伐林で正確な林分ごとの蓄積のを必要とすれば照査法以外に方法はない．ただし今般林ではこの方法は不可能である．

従って照査法は疑いなく一定の施業方法と結びついた方法である．

標本調査法（略してサンプリングとする）最大の効果を保続的に最小費用で求めようとする現代の林業では，施業方法の効果を見極めるためには，決定的な最大資本構成をなし，生産手段である所の蓄積の量的質的構造を深く観察しなければならない。全林毎木調査は費用の点から不可能であるし，ドイツの従来の方法も誤差評価の面から見て好ましくない。Löttche は，客観的に全蓄積を把握し，あわせて形質級，樹高測定，成長錐による成長量測定を行う方法を採用した．一方林小班の調査は多二義的になるが必要に応じて，ビッテルリッヒ法やサンプリングの抽出率を高めて要求に応じている．

サンプリングの特色をあげると次のようになる．

i 施業の単位の構造や Patent（成長の潜在能力等）のあらゆる必要と思われる客観的観察が得られる．

ii 得られた Index Wert は 森林容体の実際の大きさと関係なく，役立てることができる．

iii 全林毎木調査より費用が少く，また誤差の範囲がいつも確保できる．

iv あらゆる任意の問題について，技術的に安価な計算により解答が得られる．（例えば Hallerith 等の使用）

v 成長量調査が客観的に経済的に一定の誤差範囲で行うことができる．

vi 調査は常にあらゆる場所で，希望する精度が得られるよう補充測定を行ない補完できる．

従って，照査法は良い方法であるが，次の3点から，将来はサンプリングの時代となるであろう．

ハ) 造林技術や、これにより造成された林分構造を拘束することなく、有効に利用できる。従って造林法の選択については全く自由に決定できる。

ニ) 全体に対して一定の精度が明かで、確信をもって色んなことが言うことができる

ホ) この方法のみが調査作業の経営経済的計算を行うことができる。就中精度の改善に要する費用の評定ができることにより、無駄な調査、不必要な精度のための莫大な費用の浪費がさけられる。

以上は西独の学者の見解であるが、東独ではすでにノタ5ヲ年 各大学、研究所より、第2期5ヶ年計画の基礎としてサンプリングを採用するよう提案され、ノタ5ム年の春に到り、予算化され、国民有林通じて実行されるようになっている。

この外の諸国のサンプリングの実行状態は Spurr著 Forest Inventary (西沢正久訳 新しい森林調査法に略述)を見られたい。同書にない例としては、メキシコ、カナダ、ニュージランド、タイ等の森林調査がある。

4 確率標本と有意標本

有意標本の欠点としては、客観性がないことである。これは致命的なもので、人により中庸、平均というどのの見方が変ってくるので、各人各様の結果を生み 何れが正しいか判定を下し難い

従って、何れの場合でも偏りを生じるが、客観的に誤差なり偏りを知る手段がない。

これに反し、任意標本では、客観性をもち、与えられた費用、時間を考慮し、誤差をコントロールする手段をもつが、標本抽出のための予備作業が大きく、また標本に代替性を許されないのでどんな場所にでも行き調査しなければならない。もちろん、予備作業は、副次抽出の場合は、調査対象をリストする必要がなく、抽出された単位内だけリストすればよい。また調査困難地が調査を省略されたとき偏りを生ずるが、その偏りの上限は、対象についての知識に基いて定めることができる。したがって誤差は測定可能となるからこのような場合でも、依然、確率標本の条件を満足していると言えよう。

このように任意標本は、有意標本よりは、一般にまさるが、有意標本は確率論の対象にならないから、有用な結果を出し得ないとは言えない。ただ、有用な結果をなぜ出すか、また何時出すかがよくわからないということである。この長所短所を十分知りつくして用いれば、有用性はますだろうと思われるし、現に試験調査では今もって広く用いられている。調査の説明書を作ったり費用の見当つけたりするには、有意標本で十分だし、また調査の性質により任意標本による調査が不可能なものである(枚護表調査)。従って、調査の目的、性格を十分見極めて、何れの標本調査を行うか決定すべきであろう。

5. 母集団

通俗的には母集団とは調査対象の全体を指していう人が多い。その場合、対象が有限な場合は、有限母集団といい、無限な場合は無限集団というが、これは正しくない。例えば1本の木がありそれを測定した値いは無限に行えば、値は無数に出てくる。したがって、この値の集合を無限である。このようなときは木は有限でも、母集団としては無限母集団と考えなければならない。

一般に調査しようとするものは調査対象といわれる、母集団とは調査対象を抽象して、n次元（調査項目）の点として、更にその各点にある確率をあたえたものと考えたときの、点の全体と考えたならばよいであろう。したがって、母集団というものは、われわれの抽出行為によって調査対象から作りあげられるものである無限であるか有限であるかは、当然抽出行為に関係するものである。

6 標本調査の設計

調査の設計は次の手順を経て行なうとよい。

6.1 母集団を明確にすること。

これは何を調査すべきかを明確に規定することである。例えばある地方の立木の総材積を知りたいとき 農家の屋敷林はその対象となるのか、また、胸高直径何糎以上のものが対象になるか、林木とはどういう樹種か（例えば木といっても果樹は含まれないとか）等をはっきりしておかなければならない。また、収穫調査の場合は主伐では一回限りの調査で調査対象区域となる林班小班は明確なことが多いが、森林計画の調査では5年ごとに同一地域の調査の繰返しを行ない蓄積成長量の対比も行なったりするのでできるだけ区域を一定にしておかなければならない。しかし、調査対象区域内に未立木地や伐採跡地があると、これを除いて、母集団を考え勝ちであるが このようにすると次期調査との対応がつかないことも起るのでこのような林地も母集団の中に取りこみ、例えば一つの列として区分し、蓄積0とするような措置をとるとよい。このことはまた完全な枠を作成することと通ずる、枠は統計用語で、例えば事業区から小班を抽出して調査を行なうとき、各小班は抽出単位となるが、この小班全体を残れなく記載している台帳が枠となる。また基本図に縦横に線を引いて長方形の地域をつくり、これらを抽出単位とするときは、この地図が枠となる。もし抽出単位が明確に定義できるものであれば、標本抽出を行なう前に対象全体を細分することは必ずしも必要でない 抽出単位が地図上で長方形であれば、これらの地域のすべての境界を定める必要はなく、例えばこれらの地域を座標を用いてあらわしておき、抽出後に、抽出された地域だけの境界を定めればよい。

標本は元来母集団から抽出されるべきだが現実には必ず枠を通じて得られるものであるので枠は母集団と一致し、重複したり、

洩れたりすることのないように作成すべきである．それでないと枠から作られた標本からはもはや正確な母集団に関する知識は得られない。枠については次のような欠陥が入らないよう注意すべきである．

i 不正確 （枠のなかに記入され．または それにより規定される単位についての情報が不正確なこと）

ii 不完全 （対象のある単位が全く欠けていること）

iii 重 複 （対象のある単位が2度以上含まれること）

iv 不適当 （調査に含まれるべき対象のすべての種類が枠に含まれていないこと）

v 古 い （古い枠だと (1)(2)(3) の誤りがはいり易い）

なお．枠は多段抽出調査の場合は最終単位について完全なものを作る必要がない．例えば一事業区の総蓄積を知りたいときまず．小班を抽出し．さらにその中を細く区画していくつかを抽出し．調査するときは，全小班についての区画（最小単位）の完全なリストは不要である．

6.2．調査の目的を明確に定めること．

調査の目的は何か．例えば経営計画区あるいは事業区の施業団ごとの令級別に．または径級別に蓄積．成長量を目標精度どのくらいで知りたいというようなことを明らかにしておく。これは調査の設計には是非知っておかなければならない．

6.3．利用すべき資料の収集

設計のときは．利用できる過去の情報は．できるだけ利用することが必要である．そのためには．森林調査簿．以前の計画の説明書．各種図面．その他の資料を収集し．利用できるようにしておく．

6.4 調査項目の決定

調査．収集する項目について．例えば規程に定めてあるものはもちろん規程に準拠して収集することにするが．それ以外に必要な項目があれば．それを定めておく．なお．この際．測定の方法も統一し規定して調査の各組ごとに．異った基準を用いないようにする．またあいまいな概念は明確にしておかなければならない，例えば．地位．立木度．疎密度などについては．最近の各書ともあいまいで．誤解があるものがある．

地位については．子幡氏P51～52，小沢氏143～146頁の説明には問題がある．すなわち．地位は土地の材積生産能力を示すということは正しいが．この材積にこだわり．材積を用いることが．正しいと述べている．このことは．現実林の立木度が1.0ならばよいが．実際はそうではない林が多いので．一般には用いられず各国とも地位は主林木の平均樹高または上層高によって査定している。ただし．両書とも立木度で材積を修正しておき地位を査定する方法を提示しているが．この方法によれば正しい。その他の方法も述べているが．各国とも用いていないのが現状である．林床型や伐期における総平均生長量

を用いて地位区分をしている場合とあるが（フィンランド、ドイツ、など）、地位を判定するときは、この場合でも主林木の平均樹高または上方高によっているのに注意する必要がある。この理由は、林分材積は、たとい、同年同地位でも取扱い（間伐など）により変り、立木度と地位の相乗効果としてあらわされるものである。一方主木林の平均樹高は、極端な場合を除き立木度には殆んど影響されず、地位にのみ関係することが、経験的に示されて来た。

このようなことから、現在はどこでも地位は樹高により判定することが行なわれている。

以上の考察から、立木度を材積であらわすことが正しくないことがわかる。今度は逆に材積であらわした立木度には地位の効果がはいってくる。この点はドイツの森林整理学者の *W. Mantel* *W. von Laer* や *H. Speidel* も誤解しているが測樹学者の *Prodan* は正しく立木度とは $\frac{\text{現実林の断面積合計}}{\text{収穫表断面積合計}}$ と理解している。したがって国有林の経営規程細目ノ8（々）は訂正すべきであろう。（子幡氏67～72頁、小沢氏ノ54～ノ55頁も訂正の要がある）子幡氏は本数をもって立木度をあらわすことを述べているが、立木度を材積、断面積、本数であらわしたもののうち、材積の方は比較的断面積立木度に近いが本数は非常に異ることがアメリカの研究で発表されており、一般には不適当である。

疎密度については、地位と同様細目は正しい。また子幡氏（P62～63）の説明も正しいが、小沢氏（ノ54～ノ56）は説明が明確でない。ドイツの森林経理ではすべて、林地が樹冠によりおおわれる程度を疎密度（うっぺい度、へいさ度）といっており、*W. Mantel* は $\frac{\text{本数} \times \text{平均樹冠断面積}}{\text{林地面積}}$ としており、簡単にあらわすには、$\frac{\text{本数}}{\text{平均木直径}}$ により区別するのがよいとのべている。

いずれにしても、調査にあたっては、各項目の定義を明確にして、その調査方法も一定にしておかないと、比較検討する場合、まちがった結論を導くかも知れない。

6.5. 分布および母数の想定

母集団の分布型や母数を知っておくと設計上便利である。母集団の構造すなわち分布の型をきめるのは母数であるが、大抵の分布は平均値や標準偏差をきまってくる。一般に、分布型は正規分布に近いもの（樹高や直径など）もあるが失業率や、死亡率は二項分布や *Poisson* 分布をし、枯損率と恐らく同じ分布をなすものと考えられる。所得の分布は *Gibra* 分布をするということは顕著なことである。母集団の分布型を知っておれば標本をくりかえし、任意にとったときの母数の推定値の分布を知ることができる。これを知りまた母数の大体の範囲がわかれば、与えられた費用のもとでどの位の精度の調査ができるかあるいは所要精度をうるためには、どれだけの標本をとれば

よいかが決定できる．もし，これらが不明のときは，予備調査の資料は，本調査の資料と合算できるから決して無駄にはならない．

6.6. 標本抽出単位の決定

森林調査では，ビッテルリッヒ法を除いては，標本林分をとって毎木調査を行なうのが普通だが，この場合，その大きさすなわち面積，形状をどうしたらよいかをきめておかなければならない．

このための研究は多くの人によって行なわれたが，形状については円形，正方形よりは一般に矩形の方が誤差が小さく，とくに費用を考察すると辺の長さの比が 3：1 が最も良いことを Johusen と Nixan は研究して発表している．これは，常識的にもうなずかれることである．何故なれば，正方形や円の中では矩形に比して，林木の大きさの変動は少ないことが考えられしたがって，標本地間の分散は大きくなるからである．しかし一般に外国では，ほとんど円形が用いられており，矩形，正方形は極めて少ない．

面積は余り小さいと過大な推定値をあたえると従来林業では いわれていたが，筆者の研究では，形状，面積では偏りは生じなかった．これは従来区劃の境界が正しく調査されず，また，境界にあった木の処置が適宜だったことなどによるものと思われる．一般に外国では，面積は小さく，筆者の「森林調査上の

問題点とその解明」（林野庁 1957）にあげた例の通りで 0.01〜0.1 ha である．Johnsen と Nixan のダグラスファアの老令天然林の研究では 0.12 ha が最良で大体 0.8〜0.14 ha が適当で円形は一般に悪いがその他は面積 0.1 ha が良いとなっており，Richter と Grossman の研究では，従来の方法

Ⅰ　半径，2.82 m 面積 $\frac{1}{4}$ アール

　　直径　7.0〜9.9 cm の木

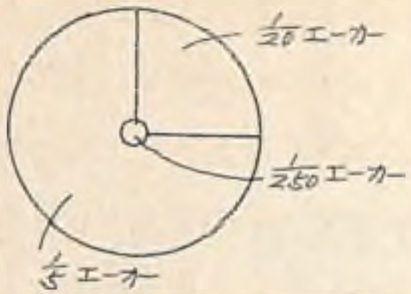
Ⅱ　半径　5.64 m 面積 1 アール

　　直径　10.0〜24.9 cm の木

Ⅲ　半径　11.28 m　面積 4 アール

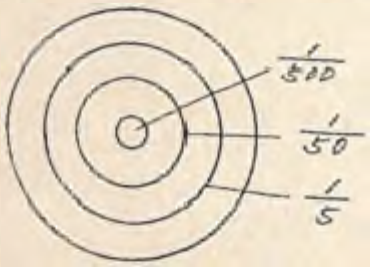
　　直径　25 cm　以上の木

とⅠ，Ⅱ，Ⅲの面積を2倍にした方法，（Ⅰ＝$\frac{1}{2}$アール，Ⅱ＝20 アール，Ⅲ＝8 アール）と第3の方法（Ⅰ　半径＝3.99，面積＝$\frac{1}{2}$アール，7〜14.9 cm の木　Ⅱ　半径＝9.77，面積＝3 a　15 cm 以上の木）を系統的抽出法で比較して第3の方法が最も良かったと発表している．上の書にない最近のアメリカの調査では，Nebraska では $\frac{1}{5}$ エーカー 0.08 ha 円形 Florida，South Carolina では断面積点数 10 の可変プロットを用いており，伐採調査のみ $\frac{1}{5}$ エーカー円形を用いている．老令天然林の多いアメリカの太平洋岸北方では，その北西部では図のように $\frac{1}{250}$ エーカー（＝0.0016 2 ha）の円径内では



直径12.45 cm以下のすべての木，$\frac{1}{20}$エーカー（＝0.02023 ha）の$\frac{1}{5}$円形内では12.70 cm〜27.70 cmの直径の木，$\frac{1}{5}$エーカー（＝0.08 ha）の円形内では，27.95 cm以上の木を測定することにしている。

ロッキー山脈地帯では，図のように$\frac{1}{500}$エーカー円形（＝



0.00081 ha内では樹高0.5呎（＝15 cm）以上直径4.99呎（＝12.45 cm）以下の木，$\frac{1}{50}$エーカー（＝0.008 ha）円形地内では，5.0〜10.99呎（12.70〜27.70 cm）の直径の木，$\frac{1}{5}$エーカー（＝0.08 ha）円形内では11呎（27.95 cm）以上の直径の木をすべて測定する．なお$\frac{1}{3}$エーカー（0.167 ha）円形内で5呎（12.70 cm）以上の過去5ヶ年内における枯死木を調査することになっている．ドイツの現行法もアメリカのこれらの方法はドイツの戦争前に行なわれ始めたKrutsch－Loetschの方法を基としている．これらの方法に比し，日本を含めた外国の方法は林木の直径に応じて変化する標本地を採用しないで，標本地面積は一定にしている．日本の国有林では，材積調査面積は0.1 ha（25×40 m，20×50 m）幼令林では，0.05 ha（20×25 m）

成長錐調査面積は，各，その$\frac{1}{10}$の面積（20×25 m）．（5×10 m）となっている．なお円形プロットは材積では0.04 ha，成長錐では0.004 haとなっており，山の傾斜面で円形に水平面に投影したとき0.04 ha，または0.004 haの長円になるようにとる。したがって山の傾斜に応じて標本地の円の半径は次のように異なってくる．

度	CoS θ	半径の長さ
0	1.00000	11.28 m
5	0.9961947	11.30
10	0.9848078	11.36
15	0.9659258	11.48
20	0.9396926	11.64
25	0.9063078	11.85
30	0.8660254	12.12
35	0.8191520	12.46
40	0.7660444	12.89
45	0.7071068	13.42
50	0.6427876	14.07

半径の長さをaとし，面積をAとすれば，θを傾斜角とすれば，

$\log a = \frac{1}{2}\log A - \frac{1}{2}\log \pi - \frac{1}{2}$

$\log \cos\theta = \frac{1}{2}\{\log A - \log\cos\theta - 0.4971499\}$

6.7 標本の抽出

標本の抽出単位の大きさ形状がきまり、母数の想定を終えれば、最も有効な抽出法を決定する。その上で調査すべき標本地の個数（標本の大きさ）を決定する。標本の大きさがきまれば具体的地図または航空写真上抽出作業を行なう。すなわち、普通は、地図上にはじめランダムに基線を引き、例えばドイツでは1万分の1の地図の場合で1kmまたは100m間隔に、格子線を引き、その交点を抽出する。なお、面積推察を行なう場合は、普通航空写真上で、5,000～10,000のplotを抽出するのが普通である。抽出法は等確率に単純無作為と系統的抽出を行なう場合が多いが、不等確率で抽出する場合もある。日本では等確率、単純無作為抽出を行なうのが殆んどであるが、外国では系統的（等間隔）抽出が大部分である。さらにこれらの抽出は推定する目的の因子の大きさが似通ったものをあつめて階層を作って、その階層ごとに上の抽出をする層別（層化）抽出を行なう場合や、前述の多段抽出を行う場合などがあるが、ともかく調査精度の効率を高める方法を選んでから抽出作業をおえる。

6.8. 現地調査

抽出作業が終れば、各調査員が調査項目の定義、測定方法などについて統一をはかり、必要に応じては調査説明書などを作り、調査員の訓練を行なう。さらに具体的な外業計画を作成し、

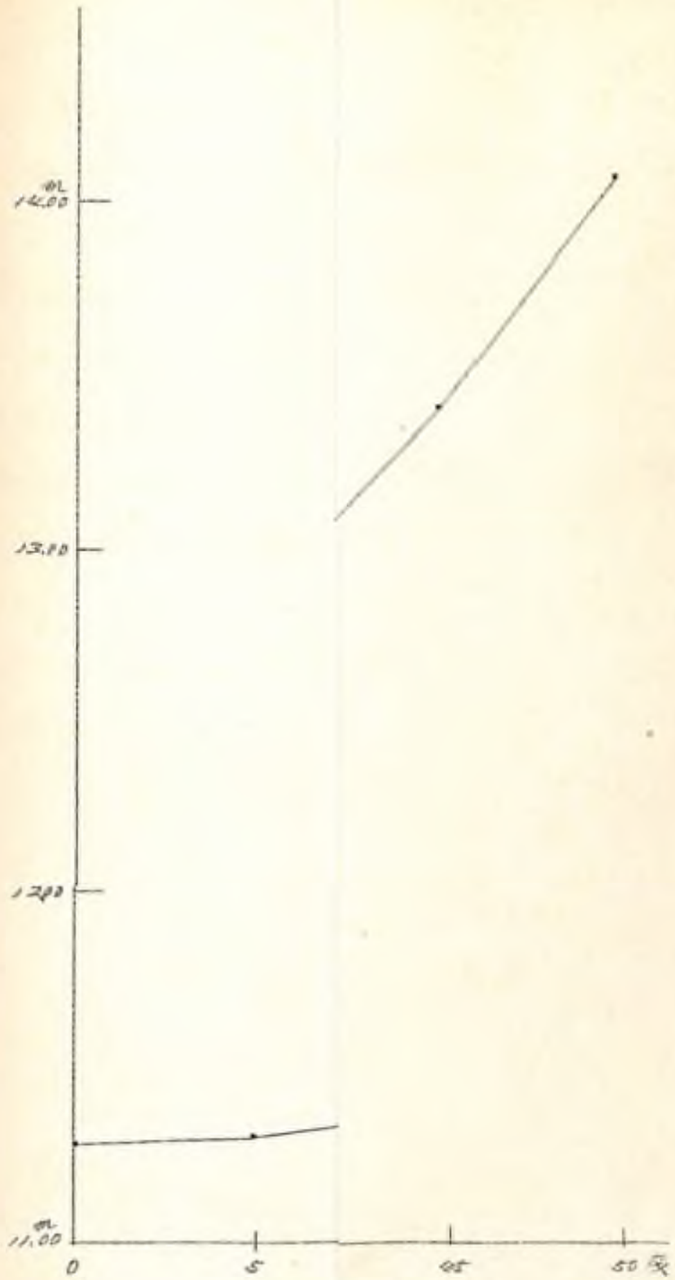
班員の数、外業日数、調査の手順などを決定する。

6.9 調査結果のとりまとめと分析

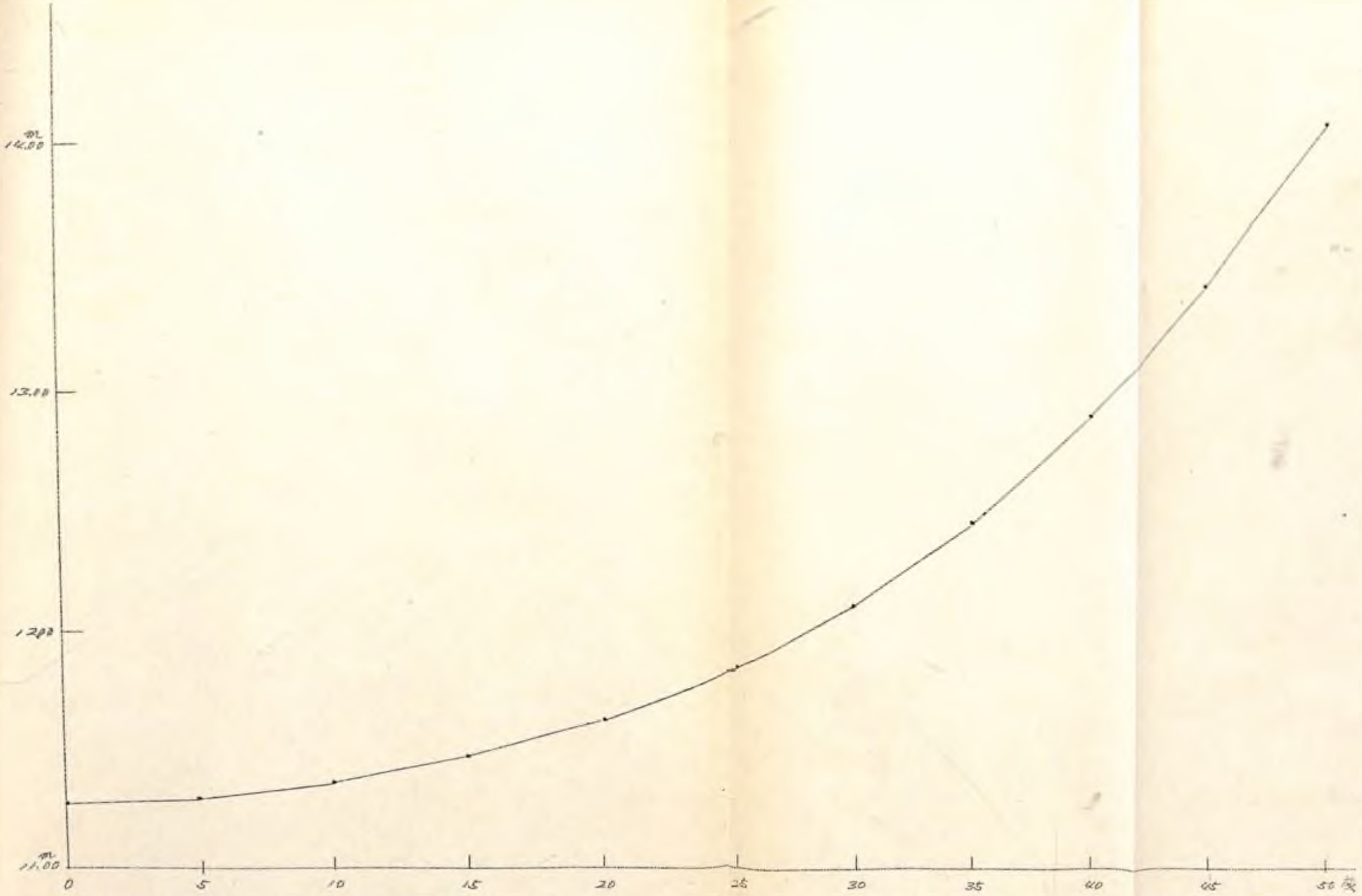
調査が終了したら結果や調査所要時間をとりまとめ、所要の推定値や、表、図などを作成する。また推定値の精度を示すものとして、信頼区間、またはC_vを計算して、将来の調査の用意をしておく。計算はかなりな大であるから、電子計算機を利用すればよい。

なお、この調査の結果をあらゆる面から分析し、将来における調査法の改善に資することが望ましい。

(33 ~ 34)



(33～90)



7. サンプリングの基礎

7.1. 標本の抽出

母集団から標本を抽出する場合、主観を入れず、確率的に抽出する手段として、サイコロ、トランプ、乱数表、乱数器などを用いる。サイコロには正六面体や正二〇面体のものがあるが、サイコロでもトランプでも数が少ないので、不便である。また割引用の機械の玉に番号をかいておけば利用できる。しかし一般には、これらを利用して乱数表を作っておき、これを利用するのが便利である。例えば割引器の中に 00, 01, 02, ……, 99 とかいた玉を100ヶ入れて、器械をよくまわして一つづつ取り出して、出た順に数字をならべて行くと、次のような表ができる。

59391	58030	……
99567	76364	……
10363	97518	……
86859	19558	……
11258	24591	……
95068	88628	……
54463	07237	……
16874	62677	……
92494	63157	……
15669	56689	……
……	……	……
……	……	……

この使用法は K 93～94 頁にある。表のよみ方は縦横どの方向に進んでもよい。この時、例えば 3/0 の単位からなる母集

団より抽出するとき，3桁の数をとり，進んで行くが，それでは3桁の乱数の中約$\frac{2}{3}$が無駄になるから，その3倍の990まで利用できるようにする。それには，211，422，633，の数字がでたら，これらはすべて211番目の個体を指すものとすればよい。

しかし，991から999まではどうしても利用できない。格子線を引いて抽出する場合は，縦横の線に番号をつけてその番号を併列1とよんで行く，今縦0～50　横0～60までの線のあれば4桁の数を乱数表からよんで行く，（0123）という数字が出れば1の横線と23の縦線の交点を抽出する，（0165）はないから省く，この場合，縦横いづれも49以下であれば上述の2倍にして乱数表を利用する方法が便える。

乱数表については吉田，西平著母論調査を参照するとよい。

7.2. 非復元抽出

乱数表で一度あたったものはその後あたっても省いて行く方法を非復元抽出といい，標本調査では普通用いられる方法である。これに反し，省かない方法を復元抽出という。

復元抽出は理論は比較的簡単であるが，非復元抽出は独立でないから面倒である。独立は44頁に説明してあるが，引きつづいて行なう抽出が，全く無関係なことで，抽出される確率は相互に影響されないことである。2つの事象が独立な場合は，

それが同時に起る確率は，両者が別々に起る確率の積になる。なお非復元抽出の場合1つの標本全体について考えれば他の標本とは独立になる。

7.3　抽出実験

今，a，b，c，d，e，fの6本の木があり，その材積が夫々，1，2，4，6，7，16 m^3とする。この中から3本の木を抽出調査して，6本の平均材積（または総材積）を推定したいとする。ランダムに抽出すると，次表の20組の標本の一つが同じ確率で抽出されるだろう。この母集団の平均は6 m^3であり，散布の程度を示す分散（0.20頁～28頁）は，

$$\left\{1^2+2^2+4^2+6^2+7^2+16^2-\frac{(1+2+\cdots\cdots+16)^2}{6}\right\}\div(6-1)$$

$$=\frac{146}{5}=29.2\ m^3$$

である。この20組の標本の平均は夫々$\frac{7}{3}$，3……9，$\frac{29}{3}$　で

この平均の平均は，$\frac{360}{3}\times\frac{1}{20}=6$　となり母集団の平均と一致する。

このように，可能な標本をとってその平均が母集団の値と一致するとき，その推定方式は，偏りがないという。サンプリングでは偏りと誤差の分散をできるだけ最小にするのが目的である。

上の母集団の分散は $\frac{\sum_{i=1}^{N}(y_i-\overline{Y}_N)^2}{N}$ ではなく，分母は $N-1$ にしてある。これを平均平方といっている人もある。今，各標本について，通常のように不偏分散を計算すると $\frac{14}{6}$，$\frac{42}{6}$ ……，$\frac{182}{6}$ となり，その平均は 28.2 となり，母集団の S^2 と一致する。

したがって標本について $s^2=\frac{\sum(y_i-\overline{y}_n)^2}{n-1}$ として計算した s^2 は母集団の S^2 の偏りのないすなわち不偏推定値であることがわかる。このことを $E(s^2)=S^2$ とかく。前の平均については $E\left\{\frac{\sum_{i=1}^{n}y_i}{n}\right\}=\overline{Y}_N$ となる。E は Expectation（期望値，期待値）の略である。可能な限りの標本を取り，

$s^2=\frac{\sum_{i=1}^{n}(y_i-\overline{y}_n)^2}{n-1}$，平均 $m=\frac{\sum_{i=1}^{n}y_i}{n}$ のようにして計算した s^2，m の平均はそれぞれ母集団の S^2，M（母集団の平均に一致するということを表わし，この推定方式が不偏であることをも表わしている。

7.4 平均値の分散

無限母集団の場合は，平均値の分散は各個体の分散を，標本の大きさすなわち測定個体の数でわったものであった。

$\left(\sigma_{\overline{y}}^2=\frac{\sigma^2}{N}\right)$（0.35頁） これは簡単にわかる。

今 $X_i=1\ 2\ 3$，$Y_i=2\ 4\ 6$ とし，Y_i の各値が X_i の各値の2倍ならば分散は $2^2=4$ となる。$Y_i=aX_i$ とすると平均は $\overline{Y}=a\overline{X}$，分散は

$$\frac{\sum(Y_i-\overline{Y})^2}{N}=\frac{\sum(aX_i-a\overline{X})^2}{N}=a^2\frac{\sum(X_i-\overline{X})^2}{N}$$

すなわち次の記号を用いると $V(Y)=a^2V(X)$。この関係は標本からの分散の推定値についても同様であることがわかろう。平均値は個々の値が n 個あったとすると，個々の値は平均値に夫々 $\frac{1}{n}$ だけ寄与している。$\overline{y}=\frac{1}{n}\times y_1+\frac{1}{n}y_2+\cdots+\frac{1}{n}y_n$ したがって $\overline{y}$ の分散は y_i の係数が $\frac{1}{n}$ だから，その自乗の $\frac{1}{n^2}$ が係数としてあらわれ，それが y_1 から y_n までの n 個の和があるから $\frac{1}{n^2}\times n=\frac{1}{n}$ が個体の分散にかけあわしたものである。

今あつかっている有限母集団の場合は，このように簡単には行かない。有限母集団は，無限母集団から抽出されたものと考えると，標本は有限母集団から取られた場合は，標本と有限母集団の関係は無限母集団よりはより近く，標本の平均値の変動範囲は有限母集団では無限母集団におけるより小さいこと が想像されよう。どれだけ小さいかというと，前の平均値の分散から有限母集団の分散 S^2 の $\frac{1}{N}$ 位ではなかろうかと考えられる。実際 $\frac{S^2_{\overline{y}}}{n}$ から $\frac{S^2_{\overline{y}}}{N}$ を引けばよい。

$a=1$　$c=4$　$e=7$
$b=2$　$d=6$　$f=16$　　$36\div=6\rightarrow$ 平均

$N=6$　$(1-\frac{3}{6})\cdot\frac{1}{3}\times 29.2$
$n=3$　$=\frac{292}{6}\doteqdot 4.9\doteqdot(2.2)^2$

$\sigma^2=\frac{25+16+4+1+100}{5}=\frac{146}{5}=29.2$

	組合せ	Sample	和 Σy_n	平均 $\bar{y}_n$	推定の誤差 $\bar{y}_n-\bar{y}_n$	平方和 Σy_n^2	平方和の3倍	補正項の(和の二	自由度 $n-1$	$s^2=\frac{\Sigma x^2-\frac{(\Sigma x)^2}{n}}{n-1}$	$(\bar{y}_e-\bar{y}_n)^2$	s^2-	s^2	$(s^2-s^2)^2$
(1)	a b. c	1 2 4	7	$\frac{7}{3}=2.3$	$-\frac{11}{3}$	21	63	49	2	$\frac{14}{2\times3}=2.3$	$\frac{121}{9}$	$\frac{76-876}{30}$	$-\frac{806}{30}$	
(2)	a b. d	1. 2. 6	9	3	$-\frac{9}{3}$	41	123	81	2	$\frac{42}{2\times3}=7.0$	$\frac{81}{9}$	$\frac{210-876}{30}$	$-\frac{666}{30}$	
(3)	a b. e	1. 2. 7	10	$\frac{10}{3}=3.3$	$-\frac{8}{3}$	54	162	100	2	$\frac{62}{2\times3}=10.3$	$\frac{169}{9}$	$\frac{310-876}{30}$	$-\frac{566}{30}$	
(4)	a. b. f	1. 2. 16	19	$\frac{19}{3}=6.3$	$-\frac{1}{3}$	261	783	361	2	$\frac{422}{2\times3}=70.3$	$\frac{1}{9}$	$\frac{2110-876}{30}$	$\frac{1234}{30}$	
(5)	a. c. d	1. 4 6	11	$\frac{11}{3}=3.7$	$-\frac{7}{3}$	53	159	121	2	$\frac{38}{2\times3}=6.3$	$\frac{49}{9}$	$\frac{190-876}{30}$	$-\frac{686}{30}$	
(6)	a. c. e	1. 4. 7	12	4	$-\frac{6}{3}$	66	198	144	2	$\frac{54}{2\times3}=9.0$	$\frac{36}{9}$	$\frac{270-876}{30}$	$-\frac{606}{30}$	
(7)	a c. f	1. 4. 16	21	7	$-\frac{3}{3}$	273	819	441	2	$\frac{378}{2\times3}=63.0$	$\frac{9}{9}$	$\frac{1890-876}{30}$	$\frac{1014}{30}$	
(8)	a d. e	1. 6. 7	14	$\frac{14}{3}=4.7$	$-\frac{4}{3}$	86	258	196	2	$\frac{62}{2\times3}=10.3$	$\frac{16}{9}$	$\frac{310-876}{30}$	$-\frac{566}{30}$	
(9)	a. d. f	1. 6 16	23	$\frac{23}{3}=7.7$	$\frac{5}{3}$	293	879	529	2	$\frac{350}{2\times3}=5.8$	$\frac{25}{9}$	$\frac{1750-876}{30}$	$\frac{874}{30}$	
(10)	a e f	1. 7. 16	24	8	$\frac{6}{3}$	306	918	576	2	$\frac{342}{2\times3}=57.0$	$\frac{36}{9}$	$\frac{1710-876}{30}$	$\frac{834}{30}$	
(11)	b. c. d	2. 4. 6	12	4	$-\frac{6}{3}$	56	168	144	2	$\frac{24}{2\times3}=4.0$	$\frac{36}{9}$	$\frac{120-876}{30}$	$-\frac{756}{30}$	
(12)	b. c. e	2. 4. 7	13	$\frac{13}{3}=4.3$	$-\frac{5}{3}$	69	207	169	2	$\frac{38}{2\times3}=6.3$	$\frac{25}{9}$	$\frac{190-876}{30}$	$-\frac{686}{30}$	
(13)	b. c. f	2. 4. 16	22	$\frac{22}{3}=7.3$	$\frac{4}{3}$	276	828	484	2	$\frac{344}{2\times3}=57.3$	$\frac{16}{9}$	$\frac{1720-876}{30}$	$\frac{844}{30}$	
(14)	b. d. e	2. 6. 7	15	5	$-\frac{3}{3}$	89	267	225	2	$\frac{42}{2\times3}=7.0$	$\frac{1}{9}$	$\frac{210-876}{30}$	$-\frac{666}{30}$	
(15)	b. d. f	2. 6. 16	24	8	$\frac{6}{3}$	296	888	576	2	$\frac{312}{2\times3}=52.0$	$\frac{36}{9}$	$\frac{1560-876}{30}$	$\frac{684}{30}$	
(16)	b. e. f	2. 7. 16	25	$\frac{25}{3}=8.3$	$\frac{7}{3}$	309	927	625	2	$\frac{302}{2\times3}=50.3$	$\frac{64}{9}$	$\frac{1510-876}{30}$	$\frac{634}{30}$	
(17)	c. d. e	4. 6. 7	17	$\frac{17}{3}=5.6$	$-\frac{1}{3}$	101	303	289	2	$\frac{14}{2\times3}=2.3$	$\frac{1}{9}$	$\frac{70-876}{30}$	$-\frac{806}{30}$	
(18)	c. d. f	4. 6. 16	26	$\frac{26}{3}=8.7$	$\frac{8}{3}$	308	924	676	2	$\frac{248}{2\times3}=41.3$	$\frac{64}{9}$	$\frac{1240-876}{30}$	$\frac{364}{30}$	
(19)	c. e. f	4. 7. 16	27	9	$\frac{9}{3}$	321	963	729	2	$\frac{234}{2\times3}=39.0$	$\frac{81}{9}$	$\frac{1170-876}{30}$	$\frac{294}{30}$	
(20)	d. e. f	6. 7. 16	29	$\frac{29}{3}=9.7$	$\frac{11}{3}$	341	1023	841	2	$\frac{182}{2\times3}=30.3$	$\frac{121}{9}$	$\frac{910-876}{30}$	$\frac{34}{30}$	
	Total			$\frac{360}{3}=120$	0					$\frac{3504}{2\times3}=584$	$\frac{876}{9}=\frac{292}{3}$		$\frac{8810-8810}{30}=0$	$\frac{10,519,680}{900}=11,688.5333\cdots$
	平均			$\frac{120}{20}=6$	0					平均 $\frac{584}{20}=29.2$	$\frac{292}{3\times20}=\frac{292}{6}$			$11,688.5333/20=584.4267$

この $\left(\frac{1}{n}-\frac{1}{N}\right)S^2$ の係数，$\left(1-\frac{n}{N}\right)\frac{1}{n}=\frac{N-n}{N}\cdot\frac{1}{n}$ のうち後者は無限母集団のときの平均値の分散に関するもので後者は有限母集団なるために生じたもので，これを有限母集団修正（f.P.Cと略す）という。

果して実際はどうだろうか。この例では，$S^2=29.2$ だから $\frac{N-n}{N}\cdot\frac{S^2}{n}=\frac{6-3}{6}\cdot\frac{29.2}{3}=\frac{1}{2}\cdot\frac{29.2}{3}=\frac{14.4}{3}$ と計算される。実際に各標本の平均値の真の平均値から誤差の自乗の平均を求めると

$$\frac{\frac{121}{9}+\frac{81}{9}+\cdots\cdots+\frac{121}{9}}{20}=\frac{\frac{876}{9}}{20}=\frac{29.2}{6}$$ となり，

両者は一致する。$E(s^2)=S^2$ だから，$\frac{N-n}{N}\quad\frac{s^2}{n}$ は平均値の誤差分散の不偏推定値をあらわすことがわかる。なお f.P.C の等き方について，従来の証明方法は各書に見られるから，今迄の方法と違った上の考え方に基いた理論的証明方法を参考としてかかげておく。

なおこの分散の平方根である標準偏差を，この場合通例，標準誤差といっている。

(注) 不偏でない推定値としては，この標準偏差が好例である。

分散の推定値は不偏だから

$s^2=S^2+\epsilon$ とおくと　$E(\epsilon)=0$ となる。

$E(\epsilon^2)=V(s^2)$ である。

$s=(S^2+\epsilon)^{\frac{1}{2}}=S\left(1+\frac{\epsilon}{S^2}\right)^{\frac{1}{2}}$　ϵ は n が大きくなるにつれて 1 に近い確率をもって S^2 よりも少さくなるから 3 次以上の ϵ を無視すると

$$s=S\left\{1+\frac{1}{2}\cdot\frac{\epsilon}{S^2}+\frac{\frac{1}{2}\left(-\frac{1}{2}\right)}{1\cdot2}\cdot\left(\frac{\epsilon}{S^2}\right)^2+\cdots\cdots\right\}$$

両辺の期待値をとると

$$E(s)=S\left\{1-\frac{1}{8}\cdot\frac{V(s^2)}{S^4}\right\}$$

これから s は S を過小評価することがわかる。

たゞし，n が大きいと無視できる。

なお s の分散は

$$V(s)\fallingdotseq\frac{V(s^2)}{4S^2}$$

(注)

有限母集団の N ケの値は無限母集団からの標本と考える。有限母集団の分散や，平均値と無限母集団におけるそれらは一致しないであろう。有限母集団平均を m，無限母集団の平均を M，分散はそれぞれ σ^2 と σ'^2 とすれば，有限母集団の平均は，M とは $\frac{\sigma'}{\sqrt{N}}$ だけちがうと考えられる。σ'^2 の値は推定しなければならないが，差し当り既知のものとしよう。標本の平均は $\bar{x}$ であらわすと，

$$(\bar{y}-m)=(\bar{y}-\mu)-(m-\mu)$$

$$(\bar{y}-m)^2=(\bar{y}-\mu)^2-2(\bar{y}-\mu)(m-\mu)+(m-\mu)^2$$

無限母集団からの標本に対しては，

$$E(\bar{y}-\mu)^2=\frac{\sigma^2}{n}$$

$$E(m-\mu)^2=\frac{\sigma^2}{N}$$

$$E(\bar{y}-\mu)(m-\mu)=E\left(\frac{e_1+e_2+\cdots+e_n}{n}\right)\left(\frac{e_1+e_2+\cdots+e_N}{N}\right)$$

$$=\frac{1}{nN}\left[E(e_1)^2+E(e_2^2)+\cdots\cdots\cdots+E(e_n^2)\right]$$

$$=\frac{1}{nN}(n\sigma^2)=\frac{\sigma^2}{N}$$

ゆえに

$$E(\bar{y}-m)^2=\frac{\sigma^2}{n}-\frac{2\sigma^2}{N}+\frac{\sigma^2}{N}=\frac{\sigma^2}{n}\left(\frac{N-n}{N}\right)$$

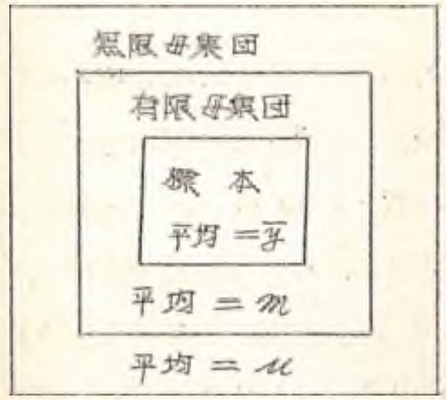
有限母集団で求められる σ^2 の最良不偏推定値は，

$$S^2=\frac{1}{N-1}\sum_{i=1}^{N}(y_i-m)^2=\sigma^2$$

ただし $\sigma^2=\frac{1}{N-1}\sum_{i=1}^{N}(y_i-m)^2$ と定義する．

故に $E(\bar{y}-m)^2=\frac{\sigma^2}{n}\left(\frac{N-n}{N}\right)$ または $\frac{\sigma^2}{n}\left(1-\frac{n}{N}\right)$

σ^2 の不偏推定は $s^2=\frac{\sum_{i}^{n}(y_i-\bar{y})^2}{n-1}$



7.5 有効性

推定方法は色々なもの水考えられる．母分散を推定する式でも色々考えられる．比や回帰などを利用する推定などはそれである．普通分散の逆数を精度といい，抽出方法や推定方法の比較のとき，その比すなわち目対効果をとって行なう。今2種の抽出推定方法により得られた分散を V_1, V_2 とし，$\frac{V_1}{V_2}=1.5$ になったとすると，この両者が同じ精度を得るには，V_2 が V_1 より 150 % の有効性があるという．すなわち，両者が同じ精度を得るに V_1 が 150 の標本数 V_2 が 100 の標本数を要することを意味している．

7.6. 理論分布

サンプリングではほとんどが大きい個体数から作られた平均値を扱うので、正規分布、二項分布の理論について知っておれば十分である。(029頁～36頁、K 53頁～57頁、79頁～83頁参照)、すなわち中心極限定理(K. 102頁)「母平均 μ、母分散 σ^2 の(任意の分布の)母集団からの大きさ n の標本の標本平均 $\bar{Y}$ に対し基準化された確率変数 $(\bar{Y}-\mu)/\frac{\sigma}{\sqrt{n}}$ の確率分布は、$n \to \infty$ のとき、基準化正規分布 $N(0,1)$ に収束する」を利用し、すべて正規分布の理論で片付ける傾向がある。正規分布の場合、標準偏差の未知のときは、その推定値 S を代用するときはいわゆる Studentの t 分布を生ずることがよく知られている。(K. 115 頁参照)

7.7 平均値の信頼区間

推定値は、偏度がちがうから、推定値については、推定値により推定しようとする特性値にどの位近いかを示すために、その信頼性を述べておくこと、必要である。(ここでは簡単のため無限母集団での正規分布を例にして説明する。)

$\bar{Y}$ の標本分布からその母平均は μ、分散は $\frac{\sigma^2}{n}$ で正規分布をすることを知っている。(ただしもとの分布も正規分布とする) 正規分布の表から μ からある距離にある標本平均の出現する割合がわかる。例えば、$(\mu-\bar{Y})$ が $-1.960/\sqrt{n}$ と

$1.960/\sqrt{n}$ の間に現われる回数は、全体の 95 %である。これは次の不等式であらわすことができる。

$$\frac{-1.960\sigma}{\sqrt{n}} < \mu - \bar{x} < \frac{1.96\sigma}{\sqrt{n}}$$

この不等式の各項に $\bar{x}$ を加えると、

$$\bar{x} - \frac{1.960\sigma}{\sqrt{N}} < \mu < \bar{x} + \frac{1.960\sigma}{\sqrt{N}}$$

となる。

もし 95 %の代りに 99 %を用いるときは、1.96 を 2.58 にかえればよい。しかしこの 99 %に相当する区間は、些か大きすぎる。

例えばσが既知で、母平均を推定するために 16 この測定値を求めたとする。この測定値の平均は、15.70 であったとする。σは、2.80 で既知とする。この値を上式に代入すれば、

$$15.70 - \frac{1.96(2.80)}{\sqrt{16}} < \mu < 15.70 + \frac{1.96(2.80)}{\sqrt{16}}$$

$$15.70 - 1.37 < \mu < 15.70 + 1.37$$

$$14.33 < \mu < 17.07$$

この結果から、このような抽出方法で求めた区間の全体の内の 95 %は μ を含むということができる。この場合の区間は 14.33 から 17.07 であるが、これを 95 %信頼区間という。区間の端の値を信頼限界という。もし標本平均が 15.20 だ

つたとしたら、信頼区間は 13.83 から 16.57 となる。もし平均を 17.0 であるならば区間は 15.63 から 18.37 となる。母集団の平均は、この場合未知だがある一定の値をもっていることに注目してほしい。われわれは、標本平均に基いて、母集団平均を含む機会が 95 % ある区間を計算しているのである。

もちろん、平均は実際に含まれているということは、どんな特定の場合にもわからない。わかるのは、このような方法で μ の推定を繰返して行なえば、このような沢山の区間の内に μ は全体の 95 % に含まれるだろうということだけである。

100α 百分率をあらわすために正規分布の表から求めた係数を Z_α であらわすことにすると、$Z_{.025} = -1.96$、$Z_{.975} = 1.96$ である。

$100(1-\alpha)$ % 信頼区間は

$$\bar{x} + Z_{\frac{1}{2}\alpha}\sigma/\sqrt{n} < \mu < \bar{x} + Z_{1-\frac{1}{2}\alpha}\sigma/\sqrt{n}$$

であらわすことができる。

上に示したように μ を推定するには母集団の σ を知る必要がある。標本から計算した標準偏差 s は σ の推定値であるから σ の代りに用いることができるであろう。しかし、係数 1.96 はもはや 95 % 信頼限界を与えてくれない。

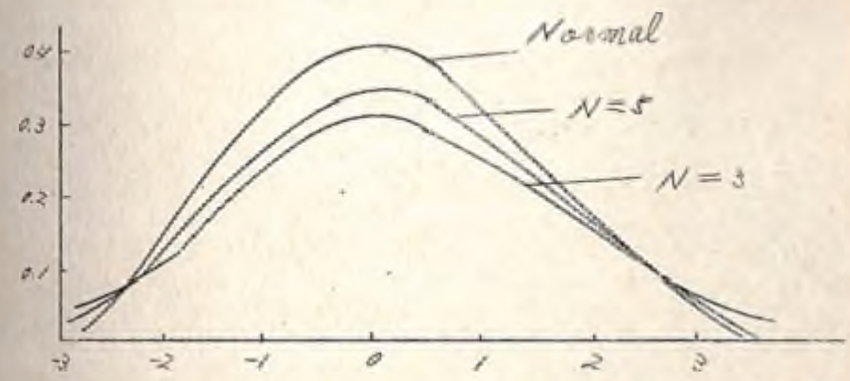
$\frac{\bar{x}-\mu}{s/\sqrt{n}}$ の標本分布は $\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$ の標本分布とは異り前述の *Student* の t 分布に従がう。

$s^2 = \frac{\sum(x_i - \bar{x})}{n-1}$

s^2 は平均すると σ^2 に等しいが、全体の個数の半分以上は σ^2 より小である。この事実から統計量 $t = \frac{\bar{x}-\mu}{s/\sqrt{n}}$ の標本分布は正規分布より多少ひろがっているのではないかと感じさせる。

もし n が非常に大きいと s^2 は σ^2 に非常に近い値を示し、大きい n に対しては $t = \frac{\bar{x}-\mu}{s/\sqrt{N}}$ の標本分布は $z = \frac{\bar{x}-\mu}{\sigma/\sqrt{N}}$ の分布と殆んどちがはなってくるだろう。

正規分布する母集団（正規母集団）を考えよう。この母集団から $N=3$ の多くの標本をとりだす、次に $N=5$、次に $N=10$ の多くの標本をとり出して、$t = \frac{\bar{x}-\mu}{s/\sqrt{n}}$ の値を計算して見る、N が大きくなるにつれて、測定した分布は分散が 1、平均が 0 の正規分布に近ずくことがわかる。



図は $N=3, 5$ と正規曲線を示す。（$\mu=0$ $\sigma=1$）

この曲線を見ると，曲線下の面積の一定 % を含むには，小さい標本程，区間は長くならなければならない．例えば，面積を 95 % 含むには，区間は $n=3$ の曲線に対しては，-4.30 から $+4.30$ までとなるし，$n=5$ のときは -2.78 から，$+2.78$ までとなる．正規分布では -1.960 から $+1.960$ までである．

t 表は t 分布の各 % の値を示している．百分率は表の上に示され，各 t 分布は表の左端の df の下に示されている．自由度の数は分散を推定するに必要な自由度の数と同じである．すなわち独立変数の数に等しい．

t 表の用法の若干を下に示す．

i. $N=10$ のとき $n-1=9$. $t_{.95}=1.83$ $\frac{\bar{x}-\mu}{s/\sqrt{n}}$ が 1.83 より小である観測回数は，全体の 95 % である．

0.50 より小さい α に対する t_α は負で表の下端に示された % を用いると得られる．$100-\alpha=\alpha$ とすればよい．そして表の値に一の符号をつけるとよい．

ii $N=10$. $\frac{\bar{x}-\mu}{s/\sqrt{N}} \leq t_{.05}=-1.83$ なのは全体の回数の 5 %

iii $N=30$ $\left|\frac{\bar{x}-\mu}{s/\sqrt{N}}\right| = 2.76$ は 全体の 99 %

$t=29$. $t_{.005}$ と $t_{.995}$ から見る．

iv. $N=3$. t の観測回数の 5 % が $t=4.30$ をこすか．

-4.30 より小

ふつう $t_\alpha^{(df)}$ の記号は 示された自由度で t 分布の，α の割合より大きい値を示すのに使う．これにより上の値を表わすと，

i $t_{.95}(9)=1.83$

ii $t_{.05}(9)=-1.83$

iii $t_{.995}(29)=2.76$; $t_{.005}(29)=-2.76$

iv $t_{.975}(2)=4.30$; $t_{.025}(2)=-4.30$

〔付〕

仮説 母分散 σ^2 は不明だが母平均はある特定の値に等しい，この検定法

1. H: $\mu=\mu_0$

ii α を選ぶ

iii 統計量 $t=\frac{\bar{x}-\mu_0}{s/\sqrt{n}}$ を計算する．

iv 棄却域は $t<t_{\frac{1}{2}\alpha}(n-1)$ と $t>t_{\frac{1}{2}\alpha}(n-1)$ でこれは t 表からよむ

標本から t を計算して，もし $t_{\frac{1}{2}\alpha}(N-1)$ より小か $t_{1-\frac{1}{2}\alpha}(N-1)$ より大ならば仮説を すてる．

さもなければ仮説を受容する。

例．ある種の鼠が生れて3ヶ月間に 65 gram の平均体重増があった。12匹の鼠が生れてから3ヶ月特定の食事をうけて次の体重増があった．55. 62. 54. 58. 65. 64. 60. 62. 59.

67. 62. 61. 5 %の有意水準で食物は体重増の原因となったか.

i Hi $\mu = 65$

ii $\alpha = .05$

iii $t = \frac{\bar{X}-65}{S/\sqrt{12}}$

iv この統計量は df 11 の t 分分をすると仮定する.

v $t < -2.20$ か $t > 2.20$ なら仮説をすてる。

vi $t = \frac{\bar{X}-65}{S/\sqrt{12}} = \frac{\sqrt{12}}{3.5406}(60.75-65) = -3.83$

$t < -2.20$

仮説はすてられる. 体重増は食物の原因だとわかる

仮説 σ^2 は未知. 平均 μ は特定の数より大きくない.

手続き

i. Hi $\mu \leq \mu_0$

ii α をきめる

iii 仮説検定の統計量として $t = \frac{\bar{x}-\mu_0}{s/\sqrt{N}}$ を用いる.

iv 母集団は正規分布すると仮定すると. s の t は $t(N-1)$ 分布する.

v. 臨界域は $t > t_{1-\alpha}$ とする. 有意水準 α を与える. $t_{1-\alpha}$ よりも大きいときのみ. 仮説をすてる.

vi. その値を計算する. もしも $t_{1-\alpha}$ より小なときは仮説

をうけ入れ. 大きいときは仮説をすてる.

例2. あるホルモンをめんどりに注射すると. 卵の目方を 0.3 オンス増加するといわれる.

30ケの卵の標本は. 注射前の平均を 24 オンス上まわった。また S は 0.20 に等しかった。これは平均増加が 0.3 オンスあったということの十分な理由となりうるか

i Hi $\mu \leq 0.3$

ii $\alpha = 0.05$ とする

iii $t = \frac{\bar{X}-0.3}{S/\sqrt{N}}$

iv このとき $df = 30-1 = 29$ の t 分布すると仮定する

v $t > t_{.95}(29) = 1.70$ なら. 仮説 H をすてる.

vi $t = \frac{.4-.3}{.20/\sqrt{30}} = 0.5\sqrt{30} = 2.75 > 1.70$

ゆえに $\mu \leq 0.3$ の仮説はすてる. 平均増は 0.3 オンス以上あるといってもよかろう. ここで注目すべきは. ホルモンが異なるめんどりに異なる反応を与えるというようなことは. 考慮されていないことである。もしそのような可能性があれば、ホルモンの効果を検定できるような適当な実験を行なうべきであろう。このような例は分散分析の際扱うことになる.

仮説. 母平均は. 特定の数より小ではない. σ^2 は未知.

これは信頼区間が t_{α} となる外は前例と同様

前例において信頼区間を付けて見ると次のようになる。

標本番号	信頼区間 Sに基くもの (Z_{α}=2とする)	(正規分布を仮定) $1-\alpha$=.95 Sに基くもの (t分布を使用)
(1)	−2.1〜6.7	
(2)	−1.4〜7.4	−0.5〜5.0
(3)	−1.1〜7.7	−1.6〜9.6
(4)	1.9〜10.7	
(5)	−0.7〜8.1	−6.7〜8.1
(6)	1.4〜8.8	
(7)	3.6〜11.4	
(8)	0.3〜9.1	
(9)	3.3〜12.1	
(10)	3.6〜12.4	
(11)	−0.4〜8.4	
(12)	−0.1〜8.7	
(13)	2.9〜11.7	
(14)	0.6〜9.9	
(15)	3.6〜12.4	
(16)	3.9〜12.7	
(17)	2.6〜11.0	
(18)	4.3〜13.1	
(19)	4.6〜13.4	
(20)	5.3〜14.1	

$$t_{0.05} = 4.3$$

Sに基くものは $t_{0.05}$=4.3として計算して見ること、計算結果は20組の標本中信頼区間に真値を含まないものが第1標本1ケだけであることを知るだろう。すなわち、この抽出では20回に1回(5%)は誤う可能性があることがわかる。但し、S^2 がわかっていると全部の標本の信頼区間に真値は含まれる。

地形概略図

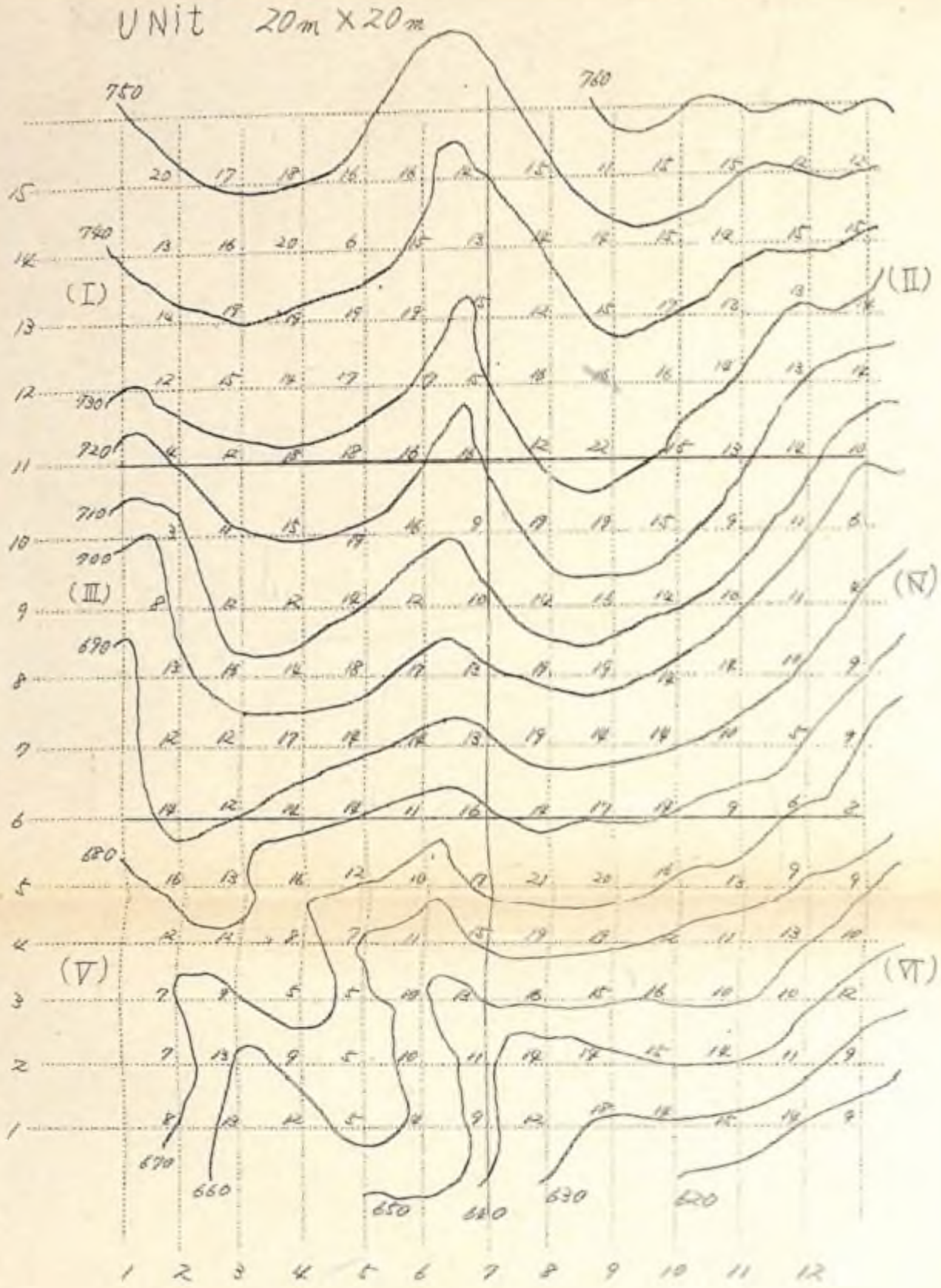
UNit 20m×20m 単位 m³

	全体	BLOCK I	BLOCK II	BLOCK III	BLOCK IV	BLOCK V	BLOCK VI
N	180	30	30	30	30	30	30
$\bar{Y}=\frac{1}{N}\sum y_i$	13.22	15.77	16.20	13.17	12.73	10.13	13.69
$S^2=\frac{1}{N-1}\sum(y_i-\bar{y})^2$	13.8665	10.0091	8.7172	10.1437	17.7964	12.8122	11.8161

$S = 3.72$　　C.V = 0.28

地形概略図

単位 m^3



平均、分散、平均平方の計算表

	全体	BLOCK I	BLOCK II	BLOCK III	BLOCK IV	BLOCK V	BLOCK VI
N	180	30	30	30	30	30	30
$\bar{Y}=\frac{1}{N}\sum y_i$	13.22	15.77	14.20	13.17	12.23	10.13	13.67
$S^2=\frac{1}{N-1}\sum(x_i-\bar{x})^2$	13.8665	10.0491	14.7172	10.1437	17.7954	12.8782	11.8161

$S = 3.72$　　$C.V = 0.28$

8. 単純無作為抽出法

単純無作為抽出法は母集団全体からランダムに抽出する方法で、母集団について何らの予備知識のないときはこの方法による以外にはないが、抽出法としては最も効率の悪い方法であるため普通は用いられないが、他の抽出法の基礎となるため十分理解しておく必要がある。別図は東京営林局天城営林署の太谷国有林のすぎ 61 年生の林分で縦 300 m 横 240 m、面積 9.2 ha 総蓄積 2390.4 m³、ha 平均 332 m³ で 林業試験場で研究するため入念な毎木調査を行なった地区である。この森林を 20 × 20 m² の小林分に区劃して、格子線に図のように番号を付けて格子点の右上の区劃を 6 個を抽出する。乱数表で 4 桁の数字をとり横座標は 15 × 6 = 90 以下縦座標は 12 × 6 = 72 以下とすると、表から次の数の組が得られた。

(0307) (1496) (1256) (5559) (1622) (6301)

これは次のものを抽出することになる。

(0314) (2401) (1208) (0211) (0406) (0301)

を小区劃の材積は夫々 20, 9, 12, 14, 12, 5 m³ となる。

この平均は $\bar{x} = \frac{20+9+12+14+5}{6} = \frac{72}{6} = 12\ m^3$

分散の推定値 $s^2 = \frac{(20-12)^2+\cdots\cdots+(5-12)^2}{6-1} = \frac{126}{5} = 25.2$

$f.P.C = \frac{180-6}{80} = \frac{174}{180} = 0.97$ 　　通例 $\frac{n}{N}$ を抽出率(f)

といい、これが0.1以下のときは f.P.C を省略する。

平均値の分散は $\frac{25.2}{6} = 4.3$

したがって標準誤差は $\sqrt{4.3} \fallingdotseq 2.1$

母集団が正規分布するとすれば自由度 $6-1=5$ で $\frac{\bar{x}-M}{s/\sqrt{n}}$ は Student の t 分布するから 95% の信頼区間をつくるには、表より $t=2.6$ を得る。

信頼区間は $(12-2.6\times2.1,\ 12+2.6\times2.1)=(6.5,\ 17.5)$ となる。その百分率誤差は $100\times\frac{5.5}{12} \fallingdotseq 46\%$ になる。

(このまの 真の値は $13.9\,m^3$ でこの場合の誤差は $-1.9\,m^3$ で誤差の 5% は実際は極めて小さい。) 標準誤差が大きかったため百分率誤差が大きくなったものである。なお、一般にサンプリングでは標本の大きさが大きいから $t=2$ とするのが普通である。

誤差の百分率をおさえられて、それ以内で推定値を求めるためには、標本の大きさは、次のようにして計算する。信頼区間から

$$\bar{x} \pm \frac{2s}{\sqrt{n}}$$

$L=\frac{2s}{\sqrt{n}}$ から必要な標本の大きさが $n=\frac{4s^2}{L^2}$ となり

また $E=\frac{L}{\bar{x}}$ とすれば $E=\frac{2s}{\bar{x}\sqrt{n}}$ $\therefore n=\frac{1}{E^2}\cdot\frac{4s^2}{\bar{x}^2}$ となる。これから n をきめればよい。

母集団のまで、s をわったものを変動(異)係数 (C.V と略す) という。

今誤差を すなわち0.10、10% におさえるには上例では

C.V=0.28 だから、

$$n=\frac{4\times(0.28)^2}{(0.1)^2} \fallingdotseq 31.4$$

したがって標本の大きさを32にすればよい。

また、s が大体予測され、$s=4\,m^3$、誤差 $=2\,m^3$ とするときは

$$n=\frac{4\times16}{4}=16$$

でよいことになる。この場合 16 は 180 の 10% に達しないから f.P.C を考えないでよいが、前例の32の場合は母集団総数の10% 以上だから f.P.C を考えなければならない。そのときは $\frac{n}{N}=\varphi$ とすると、修正された標本の大きさ $n'=\frac{n}{1+\varphi}$ により計算される。前例では $n'=\frac{32}{1+\frac{32}{180}} \fallingdotseq 28$ となり、大体28ケで十分であることがわかる。

属性に関する単純無作為抽出の場合は、例えばカラマツ先枯病にかかった苗の比率や枯損木の比率を推定したいときは、その比率 p は抽出木総数に対してその属性をもったものの割合を計算して推定値とすればよい。p の標準偏差は $S_p=\sqrt{\frac{pq}{n}}\sqrt{1-\varphi}$ となる。 $(q=1-p)$

今、苗畑で450本のカラマツ苗木から50本の苗木を抽出調査したところ 10本だけ先枯病にかかっていたとする。このとき $p=0.2$ $q=0.8$ であり、

$$S_p = \sqrt{\frac{(0.2)(0.8)}{50}}\sqrt{1-\frac{50}{450}} \fallingdotseq 0.053$$

もし，f.P.Cを無視すれば 0.057 となる．

上の S_p の式が成立するのは，各単位が2つの級どちらかに区別できる場合に限る．

もし集落抽出を用いて，各集落の各要素を分類するときは，S_p の公式は異なってくる．

前畑を数多の区（集落）にわけ，区を抽出し各区ごとに P を調査する場合は P に対し，普通の式を用いればよい．（ただし区内の個体数は同じとする．）

$$S = \sqrt{\frac{\Sigma(P_i - P)^2}{n-1}}$$

n は区の数である．これから $S_p = \frac{S}{\sqrt{n}}\sqrt{1-f}$ となる．

各区（集落）の個体数が異なる場合は，m_i は i 番目の区の個体数，a_i は目的の属性を有する個体の数とすれば，$P_i = \frac{a_i}{m_i}$ 標本全体の比率 $P = \frac{\Sigma a_i}{\Sigma m_i}$，$S$ すなわち P_i の標準偏差の推定の近似式は m_i による加重平均平方を用いる．

$$S = \sqrt{\frac{1}{(n-1)}\Sigma\left\{\left(\frac{m_i}{\bar{m}}\right)^2 (P_i - P)^2\right.}$$

ただし $\bar{m} = \frac{\Sigma m_i}{n}$ で標本の区（集落）の平均の大きさである．

なお，この式は S の正確な式ではない。計算の便のためには，

$$S = \frac{1}{\bar{m}}\sqrt{\frac{1}{(n-1)}\left\{\Sigma a_i^2 - 2P\Sigma a_i m_i + P^2\Sigma m_i^2\right\}}$$

としておいた方がよい。

なお，2項分布をする比率を推定する場合，95% 信頼確率に対する許容誤差 $L = 2\sqrt{\frac{PQ}{n}}$ から $n = \frac{4PQ}{L^2}$ となる．この場合 P の概略の値を予め知っておかなければならないが，P が 0.35～0.65 ならば，ごく大雑把でよい。$P \times Q$ の値は，この間では余り変らないからである。ただし P がこの区間を離れるに従い正確に予測しなければならない．またもとめた n が母集団の大きさ N の 1% 以上ならば，前の場合のように

$$n' = \frac{n}{1+f}$$

を求める必要がある．

調査項目が多くあり，そのそれぞれ計算した n が大差ないときは，最大の n を用いるとよいが，余りちがうときは，少なくても項目（n の小さいもの）については，抽出した n 個の内の一部についてのみ行うようにするとよい。

9. 系統的抽出法

前の目で左上から 1, 2, 3 ……, 180 まで番号をつけ, 10ケ( )の抽出するとする. 1 …… 10 の間の中の任意の数を選んで 5 だったとする. そこで次に 18 の間隔を選び, 5, 23, 41, 59, 77, 95, 113, 131, 149, 167 番目のものを抽出すると次の数が選ばれる。

4, 16, 16, 5, 13, 12, 18, 16, 14, 15.

ただしこの場合, 番号を次のようにつけてある。

1	16	……	165
2			
3			
15	30	……	180

これをジグザグ型につけた方がよい場合もある。

1	30	31 ……	165
2			
3			
15	16	45 ……	180

なお抽出の場合, 全部の数 1 から 180 までの中からランダムにえらび, 例えば, 50 番目があたったとしたら, その前後に 18 の間隔でえらんで行く, この方2の抽出法は前の方1の抽出法よりは 方1法が全母集団の大きさがわからなくとも行なえるのに比べると多少不便だが, 平均値の推定が不偏であるという長所がある。

このような抽出法を系統的抽出法という, これは次のように考えると集落抽出法で標本の大きさが1の場合だとも考えられる.

1	2	3	4	5	6 ……	18
19	20 ……			23 ……		36
37 ……				41		
143 ……				151 ……		160
165 ……				169	……	180

両方法とも18ケの標本の組からなるから, 同じことのように考えられるが, 実は異なる.

今, 母集団の大きさは 169 としたとき, 方1法では, 抽出される確率は18ケの標本はすべて $\frac{1}{18}$ である。しかし, 方2法では, 方5標本があらわれるのは, 5でも 23 で 41 でも …… 169でもその中の1つがあらわれれば, よいから抽出の確率は $\frac{10}{180}$ となる. すなわち方1～方5は $\frac{10}{180}$, 方6 …… 18 は $\frac{9}{180}$ となる. これらの値を y_1 …… y_{169} とすれば, その平均の期待値は

方1法では $E_1(\bar{y}_n) = \frac{10}{180}\left(\frac{y_1+y_{19}+\cdots\cdots+y_{163}}{10}\right) + \cdots\cdots$

$$\cdots + \frac{10}{180}\left(\frac{y_{21}+\cdots\cdots+y_{200}}{10}\right) + \frac{10}{18}\left(\frac{y_{25}+\cdots\cdots+y_{152}}{9}\right) + \cdots$$

$$\cdots\cdots + \frac{10}{180}\left(\frac{y_{18}+\cdots\cdots+y_{146}}{9}\right)$$

$\neq$ 全体の平均値 $(\bar{y}_N)$

オ2法では $E_2(\bar{y}_n) = \frac{10}{180}\left(\frac{y_1+\cdots\cdots+y_{200}}{10}\right) + \cdots\cdots$

$$\cdots + \frac{10}{180}\left(\frac{y_5+\cdots\cdots+y_{195}}{10}\right) + \frac{9}{180}\left(\frac{y_2+\cdots\cdots+y_{152}}{9}\right) + \cdots$$

$$\cdots + \frac{9}{180}\left(\frac{y_{18}+\cdots\cdots+y_{158}}{9}\right)$$

$=$ 全体の平均値 $(\bar{y}_N)$

となり，オ1法は不偏ではないが，オ2法では不偏となる．ただし抽出率が小さい場合はこの偏りは殆んど無視できる．このように抽出方法により，推定値が偏りを生ずる場合があることに注意されたい。

全体の平均を推定する場合は，何れの場合も

$$\bar{y}_n = \frac{\sum_{i=1}^{n} y_i}{n}$$ の式により計算すればよい。

この例では $\bar{y}_n = \frac{4+16+\cdots\cdots+14+15}{10} = 13$ で真値13.2に比し 0.2 m^3 だけ小さい。

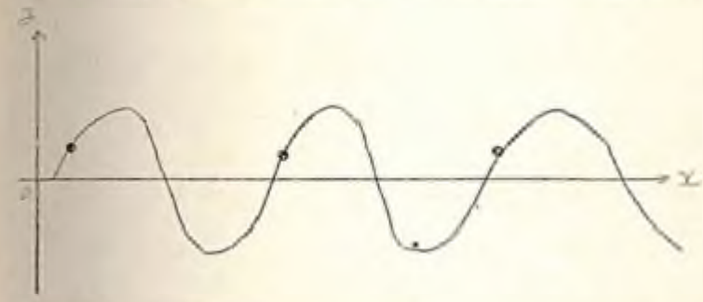
次にオ1法とオ2法の比較をのべておく．

オ1法は母集団全体の要素の数がわからなくとも抽出比がわか

れば適用できる。オ2法は全数がわからないといけないから，小さい母集団でないと中々実行が面倒である。オ1法だと未熟練者でもできる。

系統的抽出法の長所は，標本の要素が母集団内に一様にちらばる結果，単純無作為抽出法より一般に正確な結果をあたえ，作業も容易であり，未熟練の調査者がいるとき監督に行ったり，またあとから調査地点を見出したりする場合に便利であるので，日本以外の各国では，殆んどこの方式によっている。

しかし，母集団の要素に周期的な変動がある場合，その波が抽出間隔またその倍数と一致すると偏った標本になる。たとえば図のように標本がとられると，標本内の変動は小さいが，すべて



この方に偏っている。また，職員録で，男女性比，給与などをしらべるとき，偶々，各頁の一番目の人があたれば，大抵，役所などでは……長という人が多いのでほとんど男で 給与の平均も高いものになるおそれがある。したがって系統的抽出を行なうときは十分母集団の型を知っておく必要がある。しかし無秩序の場合

には一般にこのような心配はないといわれている.

系統的抽出法での次の欠点は誤差の不偏推定値を計算する適当な公式がないことである。

したがって、その誤差計算の場合は普通の単純抽出の公式を代用して行なっている。

前の 4, 16, 18, 5, 13, 12, 17, 16, 14, 15 の場合では

$$V(\bar{y}_n) = \frac{N-n}{N} \cdot \frac{\Sigma(y_i-\bar{y})^2}{n(n-1)} = \frac{180-10}{180} \cdot \frac{\Sigma(y_i-\bar{y})^2}{10 \cdot 9} \fallingdotseq \frac{1}{90}\Sigma(y_i-\bar{y})^2$$

$$\fallingdotseq \frac{1904 - 1690}{90} = \frac{214}{90} = 2.38$$

これは直接比較するのは無理だがランダムに取った場合の真の分散 1.89 に比べ、ずっと大きくなっている。一般に系統的抽出の場合このようにして計算した分散の推定値は過大な値を示すものと言われ、したがって安全だと言われている。

系統的抽出は普通単独に用いられることはなく、他のランダムな抽出法とともに用いられるのが ふつうである。殊に二段抽出法（副次抽出法）の場合は、最初の段階でランダムに取り、二段目は系統的抽出を行なうことにすれば、普通の二段抽出法の分散の計算を行なってよい。二段抽出法の場合は、二段目の誤差分散の寄与は小さく、抽出比が小さいときは、オ一段目の分散を計算すると自動的に含まれてくるからである。（これをイエーツの規則という）

10. 分散分析

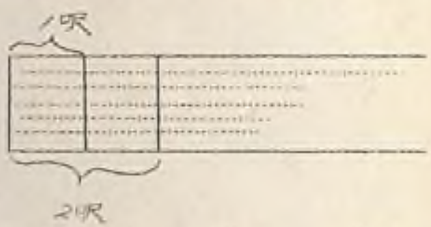
11. 抽出単位と集落抽出法

母集団は抽出単位から構成されるが、この単位は一意的ではなく、かなり便宜的なものである。森林調査では、調査の対象は林木だが、実際は、個々の林木の適当な集りからなる林分を区劃してこれを抽出の単位とする。この大きさは対象林木の大きさや調査の都合や目的に応じて、種々であり、任意的なものである。精度だけから言うとできるだけ小さな単位を広く母集団内に散在させた方がよい。費用や労力の面からは限界がある。目的は最小の費用で最大の精度をうることであるから、その適当の大きさのものが最小のものよりは効率が高い。今、例を次に示そう。

マツの苗畑、6列で1列の長さ 43.4呎

マツの全本数の推定で4つの抽出単位の型を使用する。

抽出単位の型
(全数調査の結果)



	(1) 列長1呎	(2) 列長2呎	(3) 束巾1呎	(4) 束巾2呎
N_i : 単位の総数	2.604	1302	434	217
S_i^2 : 単位あたり母分散(平均平方)	2.587	6.746	23.094	69.558
15分に算えられることができる列の長さ呎	44呎	62	78	108
推定される全体の分散	$=N_i^2\frac{S_i^2}{n_i}$			

抽出単位の各型で同じ分散を得るには、最小の単位を基準とすれば

$$N_1^2\frac{S_1^2}{n_1}=N_i^2\frac{S_i^2}{n_i} \qquad \therefore \quad n_i=n_1\frac{N_i^2}{N_1^2}\cdot\frac{S_i^2}{S_1^2}$$

例えば

$$n_2=n_1\frac{1}{4}\cdot\frac{6.746}{2.587}=.665\,n_1$$

	(1)	(2)	(3)	(4)
n_i の相対値	n_1	$.665n_1$	$.253n_1$	$.188n_1$
列の長さ	n_1	$1.330n_1$	$1.518n_1$	$2.256n_1$
15分を単位とした時の相対調査時間	$\frac{n_1P}{44}$	$\frac{1.330n_1}{62}$	$\frac{1.518n_1}{78}$	$\frac{2.256n_1}{108}$
(相対)費用の比較	C_1	$.944C_1$	$.856C_1$	$.919C_1$
	$\left(C_1\times\frac{1.330n_1}{62}\times\frac{44}{n_1}\right)$			
相対効率	100	106	117	109
(抽出単位を100とし)たときの費用の比較				

母分散(平均平方)のわからないときは、大きい標本によればよい。

2. 集落抽出

上例列長1呎は最終単位、または要素と考えられる。大きいものは集落(クラスター)と見なしてよい。

農業調査で家族は個人のクラスター，村は畑のクラスター

（たゞし実地では畑のクラスターは村より小）

クラスターサンプリングの利点

1．外業の費用減．（旅行の費用を減ずる）

2．集落の平均大同じになるようにつくると精度も向上する

（イエーツ，3.13）

クラスターSの効率

母集団はNケのクラスター，1ケのクラスターはMケの単位，これからnケのクラスターが抽出された。

$$S_b^2 = \frac{\sum^{N}(\bar{y}_i - \bar{y}_N)^2}{N-1} \quad \text{とする。}$$

nのクラスターの平均値の分散 $= \frac{N-n}{Nn} S_b^2$

シンプルランダムで nMケが抽出されたとき

$$S^2 = \frac{\sum^{NM}(y_i - \bar{y}_N)^2}{NM-1} \quad \text{とすれば，}$$

平均値の分散 $= \frac{NM-nM}{NM} \cdot \frac{S^2}{nM}$

クラスターサンプリングの効率

$$\frac{N^2M^2 \frac{NM-nM}{NM^2} \cdot \frac{S^2}{nM}}{N^2M^2 \frac{N-n}{Nn} \frac{S_b^2}{n}} = \frac{(N-n)MS^2}{S_b^2 M}$$

あるいは

$$= \frac{\frac{NM-nM}{NM} \cdot \frac{S^2}{nM}}{\frac{N-n}{Nn} S_b^2} = \frac{S^2}{MS_b^2}$$

分散分析表

	自由度	平均平方（分散）
クラスター間	$N-1$	$\frac{M}{N-1}\sum^{N}(\bar{y}_i - \bar{y}_{N.})^2 = MS_b^2 = B$
クラスター内	$N(M-1)$	$\frac{1}{N(M-1)}\sum^{N}\sum^{M}(y_{ij} - \bar{y}_i)^2 = S_w^2 = W$
計	$NM-1$	$\frac{1}{NM-1}\sum^{N}\sum^{M}(y_{ij} - \bar{y}_N)^2 = S^2 = T$

クラスターサンプリングの効率

$$= \frac{\text{要素間の平均平方}}{\text{クラスター間の平均平方}} = \frac{T}{B}$$

例 大きさ，2，4，8，16，の畑次1ケ村で小麦の作付面積の推定に用いた。

MS_b^2	2	4	8	16	
クラスター間平均平方	138.6	180.7	245.1	333.9	村内の畑間の平均平方 $=142.3$ S^2
効率	.78	.60	.42	.32	

実際は標本からの実際のデータの場合，次の分散分析表を用いる。

原因	自由度	平均平方
クラスター間	$\frac{n-1}{n}$	$\frac{1}{n-1}\sum^{n} M(\bar{y}_i - \bar{y}_n)^2 = MS_b^2$
		MS_b^2 の不偏推定値
クラスター内	$n(M-1)$	$\frac{1}{n(M-1)}\sum^{n}\sum^{M}(y_{ij} - \bar{y}_i)^2 = S_w^2$
		S_w^2 の不偏推定値
計	$nM-1$	$\frac{1}{nM-1}\sum^{n}\sum^{M}(y_{ij} - \bar{y}_n)^2 = S^2$
		S^2 の不偏推定ではない。

S^2 は S^2 の不偏推定値でないのは特に標本が小さいときそうだが標本内ではクラスター間，クラスター内の平均平方は，母集団内におけると同じ割合で S^2 に貢与しないからである。S^2 は分散分析表を母集団まで拡大すれば推定できる。

$$S^2 = \frac{(N-1)M_n S_b^2 + N(M-1)S_w^2}{NM-1} = \frac{(1-\frac{1}{N})MS_b^2 + (M-1)S_w^2}{M-\frac{1}{N}}$$

N が大きいときは $= \frac{MS_b^2 + (M-1)S_w^2}{M} = S_b^2 + \frac{M-1}{M}S_w^2$

相対精度は

$$\frac{S^2}{MS_b^2} = \frac{S_b^2 + \frac{M-1}{M}S_w^2}{MS_b^2} = \frac{1}{M} + \frac{(M-1)}{M}\cdot\frac{S_w^2}{MS_b^2}$$

例　11の標本の村で各村で4つの要素からなる4つのクラスターの例で，小麦の作付面積に対して次の分散分析表が得られた。

原因	自由度	平均平方
村間	10	290.1
クラスター間 村内	$33 = (4-1)\times 11$	$254.4 = MS_b^2$
クラスター内 調査単位要素	$308 = (8-1)\times 11\times 4$	$112.8 = S_w^2$
計	$351 = 11\times 4\times 8-1$	

S^2の推定値 $= \frac{254.4}{8} + \frac{7}{8}(112.8) = 130.1$

相対効率の推定 $= \frac{130.1}{254.4} = 0.52$

大きさの異なるクラスターの相対（精度）効率を研究する問題は，要素とクラスターを比較する問題より一般的である。ある大きさのクラスター内と間の分散は，一定の大きさのクラスターに対する分散から推定できる。この計算式も比較的やさしい。

1つのクラスターの各要素は適当正の相関があるから，クラスターの平均値の分散は $\frac{S^2}{M}$ でなくそれより大きい。Fairfield Smith は $S_b^2 = \frac{S^2}{M^2}$

（上例では $254.4 \div 8 = 31.8025 = S_b^2$　$\frac{S^2}{M} = \frac{130.1}{8} = 16.25$）

であらわされることを示した。すくい となる。

この関係から

$$S_w^2 = \frac{(NM-1)S^2 - (N-1)M\frac{S^2}{M^2}}{(N(M-1)} \fallingdotseq \frac{(M-\frac{1}{N})S^2 - (1-\frac{1}{N})M\frac{S^2}{M^2}}{M-1}$$

$$\fallingdotseq \frac{MS^2}{M-1}\left(1-\frac{1}{M^2}\right)$$

N を大にすると

ただし、$(NM-1)S^2=(N-1)MS_b^2+N(M-1)S_w^2$ だから

$$S_b^2=\frac{(NM-1)S^2-N(M-1)S_w^2}{M(N-1)}=\frac{(NM-1)S^2-NMS^2(1-M^{-b})}{M(N-1)}$$

$$=\frac{S^2\{NM-1-NM+NMM^{-b}\}}{M(N-1)}$$

$$=\frac{S^2(NMM^{-b}-1)}{M(N-1)}=\frac{S^2}{M(N-1)}\left(\frac{NM}{M^b}-1\right)$$

Jessen によれば農場では $S_w^2=aM^b$

ゆえに $S_b^2=\frac{(NM-1)S^2-N(M-1)aM^b}{M(N-1)}$

S^2 と a、b は標本データから推定を要する。

母集団全体を NM 要素からなるノケの最大のクラスターと考えれば、$S^2=a(NM)^b$ S_b^2 は、a b のみに依存する。これはあるクラスターの大きさと母集団から推定する。

(付) 集落抽出法の効果

標本抽出法は、母集団を有限ケの明白に区別できる、抽出単位に分割できることを前提としている。母集団の分割された最小単位を要素といい、要素のあつまりがクラスターである。抽出単位がクラスターの時は、クラスターサンプリングという。母集団を含む全地域を小面積に分割して、母集団の各個の要素が只一つの小面積と結びつけたとき、この抽出法は地域抽出法という。

要素のリストは必ずしも利用できないときはクラスターやエーリヤサンプリングは便宜である。クラスター次小さい程、標本内の要素の一定数に対し母集団特性値の推定値は正確になるの次通例であろう。散在した標本を調査することは、クラスターの同等の標本を調査するより金次かいる。最適のクラスターは、抽出される母集団の所定調査歩合や与えられた費用と対して最小の標準誤差をもつ推定値を与えるものである。

下に例をあげる。

N = 母集団の集落の数　　M = 各集落の要素の数

ρ = 級内相関　　S_b^2 = 集落平均間の平均平方

S_w^2 = 集落内の要素間の平均平方

S^2 = 母集団内の要素の平均平方

$$S_b^2=\frac{S^2}{M}[1+(M-1)\rho]$$

$$S_w^2=S^2(1-\rho)$$

$$(NM-1)S^2=(N-1)MS_b^2+N(M-1)S_w^2$$

$$=[(N-1)\{1+(M-1)\rho\}+(NM-N)(1-\rho)]S^2$$

$$=S^2\{N-1+NM\rho-N\rho-M\rho+\rho+NM-NM\rho-N+N\rho\}$$

$$=S^2\{NM-1-\rho(M-1)\}$$

$$S^2=S^2\left\{1-\rho\frac{(M-1)}{NM-1}\right\}$$

あるいは.

$$S^2\left\{\frac{1}{M}\left(1+\overline{M-1}\,\rho\right)+\frac{(M-1)(1-\rho)}{M}\right\}=S^2$$

$$\downarrow \qquad\qquad \downarrow$$

$$S_b^2 \qquad + \quad \frac{(M-1)S_w^2}{M} \longrightarrow S^2$$

nからなる標本による平均値の推定値の分散は.

$$\frac{S_b^2}{n} \quad \text{or} \quad \frac{S}{nM}\left[1+(M-1)\rho\right]$$

NMからnばれたnMケの要素の標本では.

$$\frac{S^2}{nM}$$

$$\frac{S_b^2}{m}=\frac{S^2}{nM}\left[1+(M-1)\rho\right] \text{ と } \frac{S^2}{nM} \text{ を比べると.}$$

$(M-1)\rho$は要素ではなく、集落を抽出するために生じた分散の相対的変化をあらわすことがわかる.

逆例　$\rho>0$ で、$(M-1)\rho$ が Mと共に増加するようにMが増すにつれて、ρは減少する。

クラスターの大きさが増すことは、推定値の抽出分散をますようになる。

上の理論を実験計画に応用して見よう. S^2をブロックを考慮せず plots 間の誤差の平均平方とすれば.

$S^2(1-\rho)$ は乱塊法での plot あたりの誤差の平均平方となり

$S^2(1+\overline{M-1}\,\rho)$ は Mplots からなるブロック間の平均平

方となる.

乱塊法の相対精度は $\frac{100}{1-\rho}$ となる.

枝分れ試験での main plot の処理に対し、sub plot の相対効率は、$1+\overline{M-1}\,\rho$ 乃至 $1-\rho$ となる.

例　10ケのブロックで棉の2品種の収量分散分析乱塊法と無制限ランダム法との配列の効果を比べる.

	d.F	M.S
ブロック	9	407796 $=2S_b^2$
処理	1	423.2
誤差	9	54280 S_w^2

$S_b^2 = S^2(1-\rho) = 54280$

$MS_b^2 = S^2[1+(M-1)\rho] = 407796$

$\therefore S^2 = 23104.0$

$\rho = .7650$

相対効率は $\frac{100}{1-\rho} = 425.6\%$ に増加した.

12. 層化（別）抽出法

12.1 抽出の原理と母数の推定

母集団をいくつかの階層にわけ，各層ごとに無作為抽出する方法を層化（別）抽出法という。

この層わけは i) 調査の目的が層ごとの情報を要求するとき ii) 調査の精度を高めたいときに行なわれるもので，実際にはほとんどこの2目的を備えており，層化抽出法は大抵の場合用いられている。

今前述の模型実験を繰返して見る。

a, b, c, d, e, f をそれぞれ 1, 2, 4, 6, 7, 10 とし，この6ケの中から4ケを単純無作為に抽出するときは 15 組の標本の1つが得られる。

$$\left(\binom{N}{n}=\binom{6}{4}=\frac{6\cdot5\cdot4\cdot3}{1\cdot2\cdot3\cdot4}=15\right)$$

この平均 $\bar{y}$ で分散は公式から

$$S_{\bar{y}}^2=\frac{N-n}{N}\cdot\frac{\sum_{N}^{N}(y_i-\bar{Y})^2}{n\cdot(N-1)}=\frac{2}{6}\cdot\frac{56}{4\cdot5}=\frac{1}{3}\ \frac{14}{5}=\frac{14}{15}$$

となり，表で実際に行なって見たものと一致する。

単純無作為抽出の例

$a=1,\ b=2,\ c=4,\ d=6,\ e=7,\ f=10$

		和	平均	平均$-\frac{20}{4}$	$(\text{平均}-\frac{20}{4})^2$
2 b, c, d	1, 2, 4, 6	13	$\frac{13}{4}$	$-\frac{7}{4}$	$\frac{49}{16}$
1 b c e	1, 2, 4, 7	14	$\frac{14}{4}$	$-\frac{6}{4}$	$\frac{36}{16}$
2 b e f	1, 2, 4, 10	17	$\frac{17}{4}$	$-\frac{3}{4}$	$\frac{9}{16}$
2 c d e	1, 2, 6, 7	16	4	-1	1
2 b d f	1, 2, 6, 10	19	$\frac{19}{4}$	$-\frac{1}{4}$	$\frac{1}{16}$
a b e f	1, 2, 7, 10	20	5	0	0
a c d e	1, 4, 6, 7	18	$\frac{18}{4}$	$-\frac{2}{4}$	$\frac{4}{16}$
2 c d f	1, 4, 6, 10	21	$\frac{21}{4}$	$+\frac{1}{4}$	$\frac{1}{16}$
a c e f	1, 4, 7, 10	22	$\frac{22}{4}$	$+\frac{2}{4}$	$\frac{4}{16}$
a d e f	1, 6, 7, 10	24	6	$+1$	1
b c d e	2, 4, 6, 7	19	$\frac{19}{4}$	$-\frac{1}{4}$	$\frac{1}{16}$
b c d f	2, 4, 6, 10	22	$\frac{22}{4}$	$+\frac{2}{4}$	$\frac{4}{16}$
b c e f	2, 4, 7, 10	23	$\frac{23}{4}$	$+\frac{3}{4}$	$\frac{9}{16}$
b d e f	2, 6, 7, 10	25	$\frac{25}{4}$	$+\frac{5}{4}$	$\frac{25}{16}$
c d e f	4, 6, 7, 10	27	$\frac{27}{4}$	$\frac{7}{4}$	$\frac{49}{16}$
計		300	$\frac{300}{4}$	0	$\frac{224}{16}=14$
平均		20	5	0	$\frac{14}{15}$

今度は、(a.b.c)(d.e.f) の二戸にわけて、各戸から2ケづつ抽出すると、次のような9組の標本が得られる。

標本	和	平均	真値からの差	差の自乗
a b d e	16	4	−1	1
a b d f	19	$\frac{19}{4}$	$-\frac{1}{4}$	$\frac{1}{16}$
a b e f	20	5	0	0
a c d e	18	$\frac{18}{4}$	$-\frac{2}{4}$	$\frac{4}{16}$
a c d f	21	$\frac{21}{4}$	$\frac{1}{4}$	$\frac{1}{16}$
a c e f	22	$\frac{22}{4}$	$\frac{2}{4}$	$\frac{4}{16}$
b c d e	19	$\frac{19}{4}$	$-\frac{1}{4}$	$\frac{1}{16}$
b c e f	22	$\frac{22}{4}$	$\frac{2}{4}$	$\frac{4}{16}$
b c e f	23	$\frac{23}{4}$	$\frac{3}{4}$	$\frac{9}{16}$
計		$\frac{180}{4}=45$	0	$\frac{80}{16}=\frac{5}{2}$
平均		$\frac{45}{9}=5$		$\frac{5}{18}$

この表からわかるように、平均を求めるときは

$\frac{N_1\bar{y}_1+N_2\bar{y}_2}{N}$ $\left(\bar{y}_1=\frac{y_{11}+y_{12}}{2},\ \bar{y}_2=\frac{\bar{y}_{21}+\bar{y}_{22}}{2}\right.$,

この一番上の標本の例では、$\bar{y}_1=\frac{1+2}{2}=\frac{3}{2}$, $\bar{y}_2=\frac{6+7}{2}=\frac{13}{2}$

したがって $\bar{y}=\frac{3\times\frac{3}{2}+3\times\frac{13}{2}}{6}=\frac{16}{4}=4$ となる$\left.\right)$

＝平均が真の平均5の不偏推定になることがわかる。また分散は $\frac{5}{18}$ となる。サンプリングの有効性を見るには、分散の逆数の比をとればわかるので、単純無作為と戸化抽出のこの場合の比較は、

$$\frac{\frac{18}{5}}{\frac{15}{4}}=\frac{72}{25}=3.35$$

となり、戸化抽出は3倍以上も能率がよいことがわかる。すなわち、戸化抽出では単純無作為の場合の $\frac{1}{3}$ 以下の抽出個数でたりることになる。(この場合は個数は4だから無理であるが) なお上の目からも戸化抽出の能率がよいことがわかろう。

戸化抽出の場合の分散推定値の計算はどうしたらよいか。まず平均値 $\bar{y}=\frac{N_1\bar{y}_1+N_2\bar{y}_2}{N}$ として推定した。下の添字は1層 2層を表わす。$n_1+n_2=n$。上式は

$\bar{y}=\frac{N_1}{N}\bar{y}_1+\frac{N_2}{N}\bar{y}_2$ となっている。平均値の分散でのべたように n_1 と n_2 はこの場合独立だから

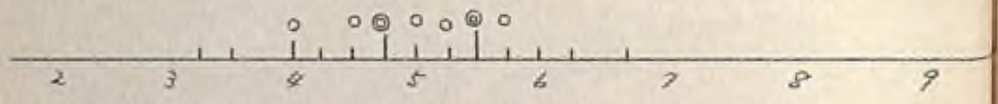
$$v(\bar{y})=\left(\frac{N_1}{N}\right)^2 v(\bar{y}_1)+\left(\frac{N_2}{N}\right)^2 v(\bar{y}_2)\quad\text{となる。}$$

$$v(\bar{y}_i)=\frac{N_i-n_i}{N_i}\frac{S_i^2}{n_i}\qquad\left(\begin{array}{l}\text{ただし } N_i \text{ は } i \text{ 戸の総単位}\\ S_i^2 \text{ は } i \text{ 戸の分散の推定値}\end{array}\right)$$

から $v(\bar{y})=\frac{N_1-n_1}{N_1}\left(\frac{N_1}{N}\right)^2\frac{S_1^2}{n_1}+\frac{N_2-n_2}{N_2}\left(\frac{N_2}{N}\right)^2\frac{S_2^2}{n_2}$

$$N^2 = \frac{1}{n^2}\left[\frac{N_1(N_1-n_1)S_1^2}{n_1} + \frac{N_2(N_2-n_2)S_2^2}{n_2}\right]$$

となる．f.P.C を省略できるときは



$$V(\bar{x}) = \frac{1}{N^2}\left[\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right] \quad \text{となる．}$$

なお，全体の推定値は $N\bar{x}$ だから，その分散の推定値は $N^2V(\bar{x})$ となる．したがって，

$$V(n\bar{x}) = \left[\frac{(N_1-n_1)S_1^2}{N_1 n_1} + \frac{(N_2-n_2)S_2^2}{N_2 n_2}\right] \quad \text{となる．}$$

上例を上の式で計算すると

$$V(s) = \left(\frac{1}{6}\right)^2\left[(3-2)\cdot\frac{3\cdot 21}{2\cdot 9} + (3-2)\frac{3\cdot 39}{2\cdot 9}\right]$$

$$= \frac{1}{36}\cdot\frac{3}{2}\cdot\frac{1}{9}\cdot(21+39) = \frac{60}{12\times 18} = \frac{5}{18} \qquad \left(S_1^2 = \frac{21}{9},\ S_2^2 = \frac{39}{9}\right)$$

となり，表で計算したものと一致するので，このようにして計算すれば分散の不偏推定が求められることがわかる．

前述のように，戸化抽出は単純抽出より分散は大抵の場合小さい．これは戸平均値間の差は，各戸から抽出されるので，推定値 $\bar{x}_{st}$ の抽出誤差には関係なく，たゞ戸内の抽出誤差のみは後者に関係するもので，結局，戸内の変動だけが $\bar{x}_{st}$ に関係

することになる．従って階 戸内が均質になるよりは 戸別けをすれば，当然 $\bar{x}_{st}$ の抽出誤差は小さくなるわけである．戸は母集団より小さいから変動が小さいのは当然で有限修正を無視できれば常に戸化抽出の方が単純抽出より抽出誤差は小になることがわかるだろう．

戸化抽出法の特殊な例として比例抽出法がある．これは各階戸から同一比例で標本を抽出する場合である．すなわち

$$\frac{n_1}{N_1} = \frac{n_2}{N_2} = \cdots\cdots = \frac{n_k}{N_k} = \frac{n}{N}$$

これから

$$W_k = \frac{N_k}{N} = \frac{n_k}{n}$$

したがって一般公式

$$\bar{x}_{st} = \frac{\sum N_k \bar{x}_k}{N} = \sum W_k \bar{x}_k = \frac{\sum n_k \bar{x}_k}{n} = \bar{x}$$

すなわち $\frac{\text{標本和}}{n} = \bar{x}_{st}$ となり，計算が容易になる．

分散も

$$V(\bar{x}_{st}) = \frac{1}{N^2}\sum N_k(N_k-n_k)\frac{S_k^2}{n_k} \qquad \left(S_k^2 = \frac{\sum(x_{ki}-\bar{x}_k)^2}{n_k-1}\right)$$

$$= \sum W_k^2(1-f_k)\frac{S_k^2}{n_k} \qquad \left(\frac{n_k}{N_k} = f_k = \frac{n}{N} = f\right)$$

$$= \sum W_k^2(1-f)\frac{S_k^2}{n_k}$$

$$= (1-f)\sum W_k^2\cdot\frac{S_k^2}{n_k} = \frac{(1-f)}{n}\sum W_k S_k^2 \qquad (n_k = nW_k)$$

$$= \left(\frac{1}{n}-\frac{1}{N}\right)\sum W_k S_k^2$$

層内母分散が等しいときは

$S_w^2 = \Sigma W_h S_h^2$ とおくと

$= \frac{S_w^2}{n}(1-\varphi)$ となる.

例 1

番号	1	2	3	4	平均	S^2
農場の大きさ	2	3	4	7	4	$14/3$
家畜の頭数	6	10	13	19	12	30

シンプルランダムでえらぶ

$\frac{N-n}{N} \cdot \frac{S^2}{n} = \frac{4-2}{4} \cdot \frac{30}{2} = 7.5$

農場の大きさで母集団を農場を (1,2) (3,4) と層化する.

標本	平均 $\bar{Y}_W$	$\bar{Y}_W - \bar{Y}_N$	$(\bar{Y}_W - \bar{Y}_N)^2$
1, 3	9.5	−2.5	6.25
1, 4	12.5	.5	.25
2, 3	11.5	−.5	.25
2, 4	14.5	2.5	6.25

平均 3.25

公式からは

$\left(\frac{1}{1}-\frac{1}{2}\right)\times\frac{1}{4}\times 8 + \left(\frac{1}{1}-\frac{1}{2}\right)\times\frac{1}{4}\times 18$

$= 1 + 2.25 = 3.25$

例. 2. 畑を等面積の3階層にわけて. 各層から任意に長さ 1 m のうねの 10 標本を抽出して. 収量をはかった. その分散分析表は次のとおりである.

変動因	自由度	平方和	平均平方
全体	29	8540	295.3
層間	2	2073	1036.5
層内	27	6491	240.4

$S_w = \sqrt{240.4} = 15.5$ $n = 30$ $\frac{n}{N}$ は無視できるものとすると

$$S(\bar{y}_{st}) = \frac{S_w}{\sqrt{n}} = \frac{15.5}{\sqrt{30}} = 2.83\ g/m$$

近似的に今どの位層化の効果が上ったかを見ると. 単純任意抽出による平均値の標準誤差の推定値としては.

$$S(\bar{y}) = \frac{\sqrt{295.3}}{\sqrt{30}} = 3.14\ g/m$$

で 2.83 g/m と対比すれば標準誤差は約 10 % 小となる. これから層平均に差が大きいほど層化抽出法が有利なことがわかる.

(注) $\sum_h^L \sum^{N_h} (Y_{hi} - \bar{Y}_N)^2 = \sum_h^L \sum^{N_h} (Y_{hi} - \bar{Y}_{Nh})^2 + \sum_h^L N_h (\bar{Y}_{Nh} - \bar{Y}_N)^2$

この式は $(N-1)S^2 = \sum (N_h - 1) S_h^2 + \sum N_h (\bar{Y}_{Nh} - \bar{Y}_N)^2$

近似的に $S^2 = \sum W_h S_h^2 + \sum W_h (\bar{Y}_{Nh} - \bar{Y}_N)^2$

これから単純と層化標本の差は.

$$\sum_h^k \left(\frac{1}{n}-\frac{W_h}{n_h}\right) W_h S_h^2 + \left(\frac{1}{n}-\frac{1}{N}\right) \sum_h^k W_h \left(\overline{Y}_{Nh}-\overline{Y}_N\right)^2$$

あるいは層化抽出と無層化抽出の比較は次のようにしてもできる.

単純　　層化

$$\frac{(N-n)}{Nn} S^2 = \left(\frac{1}{n}-\frac{1}{N}\right) S^2, \qquad V(\overline{y}_w) = \sum_h^k \left(\frac{1}{n_h}-\frac{1}{N_h}\right) W_h^2 S_h^2$$

$$\sum_h^k \left(\frac{1}{n}-\frac{1}{N}\right) W_h S_h^2 + \sum \left(\frac{1}{n}-\frac{1}{N}\right) W_h \left(\overline{Y}_{Nh}-\overline{Y}_N\right)^2 - \sum_h^k \left(\frac{1}{n_h}-\frac{1}{N_h}\right) W_h^2 S_h^2$$

$$= \sum \left(\frac{1}{n}-\frac{1}{N}-\frac{N_h}{N\,n_h}+\frac{n_h}{N\,N_h}\right) W_h S_h^2 + \sum \left(\frac{N-n}{N}\right) W_h \left(\overline{Y}_{Nh}-\overline{Y}_N\right)$$

$$W_h = \frac{N_h}{N}$$

$$= \sum \left(\frac{1}{n}-\frac{W_h}{n_h}\right) W_h S_h^2 + \left(\frac{1}{n}-\frac{1}{N}\right) \sum W_h \left(\overline{Y}_{Nh}-\overline{Y}_N\right) \quad \leftarrow \text{単純と層の差}$$

12.2 標本の大きさのきめ方

標本の大きさは、与えられた費用に対して、

$$V(\overline{y}_{st}) = \sum_h^k \left(\frac{1}{n_h}-\frac{1}{N_h}\right) W_h^2 S_h^2 \quad \text{が最小になるように定}$$

める. 最も簡単な場合として、費用が標本の大きさに比例するとすると

$$n_h = n \frac{W_h S_h}{\sum W_h S_h} \qquad (Neyman\ \text{割当})$$

のように定めるとよい.

上式はまた　$n_h = n \cdot \frac{N_h S_h}{\sum N_h S_h}$　ともなる.

この式は　$\frac{n_h}{N_h S_h} = \frac{n}{\sum N_h S_h}$　より来り、各層の $\frac{n_h}{N_h S_h}$ が

皆等しいことを意味している. したがって比例抽出の場合、$\frac{n_h}{N_h}$ が皆等しいから　この場合は、各層の S_h が等しいことを前提としていることがわかる. 上の n_h の公式より大きな階層が変動の大きな階層では多く抽出すべきことがわかる. さらに費用が各層ごとに異なるときは、$n_h = N_h S_h / \sqrt{C_h} \times n$ (C_h は h 層における1単位の費用)　この式から費用がかかる程 n_h は小さくなり、われわれの常識と一致する. この割当を、最適割当 (Neyman-Deming 割当) という.

しかし、S_h か C_h が不明なときは、比例割当を用いるのが普通である. (あとの計算が容易であるため)

Neyman 割当の計算は次のように行なう.

例1.

層	層面積割合 W_h	平均面積	標準偏差 S_h	変動係数 C.V	$W_h S_h$
I	0.064	0	0	0	0
II	0.484	122	77	63.0 %	37.0
III	0.456	435	220	44.5 %	100.0

$\sum W_h S_h = 137.0$

層化後の S_h の平均は 137.0 m^3 と考えられるから、これに対応する C.V は 50 % となる。5 % の百分率誤差で推定値を求めたいとすれば、

$$n = \frac{50^2 \cdot 4}{25} = 400$$

各層の n_h は $\frac{n \times W_h S_h}{\Sigma W_h S_h}$ により、まず

$$n' = \frac{400 \times 1.0}{137} = 2.92$$

を計算し、$W_h S_h$ にかけると、

I		0
II	37 × 2.92	108
III	100 × 2.92	292
		400

となる。

例2　（Cochran より）

4年制の1大学あたり総在籍者数のデータ

階層（大学あたりの学生数）	大学の数 N_h	在籍者数	大学当り平均 $\bar{Y}_h$	大学あたりの標準偏差 S_h
I 1,000未満	661	292,691	443	234
II 1,000～3,000	205	375,302	1,834	625
III 3,000～10,000	122	672,728	5,514	2,008
IV 10,000以上	31	572,683	18,506	10,023
計	1,019	1,913,394		

最適な標本の大きさのための計算

層	N_h	$N_h S_h$	$\frac{N_h S_h}{\Sigma N_h S_h}$	実際の標本の大きさ	抽出率
I	661	155,996	.1857	65	10
II	205	128,125	.1526	53	26
III	122	244,976	.2917	101	83
IV	31	310,713	.3700	31	100
計	1,019	839,810	1.000	250	

IV層は抽出率が 37 % になり、標本の大きさがこの場合 250 とすると　$839 \times 250 = 92$　となるが、実際は 31 しかない。このようなときは　この層については、100 % 調査することとし、残りを　最適配分の計算にもとづいて割あてればよい。したがって

$$\text{I} \quad (250 - 31)\left\{\frac{.1857}{.1857 + .1526 + .2917}\right\} = 65$$

となる。

問題　天城のスギ林分の6層に $n = 92$ として、Neyman 割当で割当てて見、更に各層に4ヶ宛比例割当した場合の誤差を計算して比較する。

12.3 属性についての標本抽出

属性の層化標本抽出の場合は、

$$P_{st} = \sum W_h p_h$$

として推定すればよい。ただし p_h は、h 層の標本比率である。P_{st} の分散は前の式の S_h^2 の代りに $p_h q_h$ を用いればよい。

12.4 要　約

層化抽出では平均値、層の標準偏差値に共に大きな相違があるときは Neyman 割当が有効であるが、単に層平均値には差があるだけで、標準偏差に大きい差のないときは労力其の他費用を考えるときは比例割当の方が適かによい。

なお、層化は直接調査したい変量例えば林積により層化することはもちろんよいが、それと相関のある因子、例えば 令級平均経級、樹高、密度などの一層つかみ易い因子で層化を行なってもよい。また層化は調査の目的が施業図別、あるいは事業区別に 人、天 別、針広別、令級別に材積を知りたいなどのために行なうことも必要な場合もあるが、これらは、精度向上という点をもかね備えていることが多い。

層化抽出の場合の計算の様式を次に示しておく。

i) 平均値の推定値の計算

層	N_h	$W_h = \frac{N_h}{N}$	n_h	$\sum_{(i)}^{n_h} y_{hi}$	$\bar{y}_h$	$W_h \bar{y}_h$
I	—	—	—	—	—	—
II	—	—	—	—	—	—
⋮						
計	○	○	○	○		○

ii) 平均値の分散の推定値の計算

層	N_h	n_h	$\sum_{(i)}^{n_h} y_{hi}^2$	s_h^2	W_h^2	$W_h^2 s_h^2$	$\left(\frac{1}{n_h} - \frac{1}{N_h}\right)$	$\left\{\frac{1}{n_h} - \frac{1}{N_h}\right\} W_h^2 s_h^2$
I								
II								
⋮								
計	○	○	○					○

計は ○ 以外は不要

比例抽出の場合は次のように簡単になる。

i) 平均の推定値の計算

戸	n_h	$\sum_{i=1}^{n_h} n_{hi}$	$\bar{y}_i$	
I				
II				
.....				
計	○ ①	○ ②		○

$\bar{y}_{st} = \frac{②}{①}$ となる

ii) 平均分散の推定値の計算

戸	N_h	$W_h = \frac{N_h}{N}$	$\sum_{i=1}^{n_h} y^2_{hi}$	S^2_h	$W_h S^2_h$
I					
II					
.....					
計	$N=\Sigma N_h$	1.000	$\sum_h \sum_i y^2_{hi}$ ①		$\Sigma W_h S^2_h$ ②

分散の推定値 $S^2_{\bar{y}st} = \left(\frac{1}{n} - \frac{1}{N}\right) \times ②$

最後に信頼区間について述べる．信頼区間は各戸の標本数が等いときは，

$\bar{y}_{st} \pm t S_{\bar{y}st}$

を用いるが，Banerjee の式をもととすれば，次のようになる．

$\bar{y}_{st} \pm \sum_{h=1}^{k} t^2_h S^2_{\bar{y}h} W^2_h (1 - g_h)$

ただし，t_h は Student の t 分布（自由度 $n_h - 1$）の値である．（もちろん母集団の型は正規分布として）

また Satterthwaite によれば $\bar{y}_{st} \pm t S(\bar{y}_{st})$ の t は近似的に次の自由度における t であると述べている．

$\frac{(\Sigma f_h S^2_h)^2}{\Sigma \frac{f_h^2 S^4_h}{n_h - 1}}$ ただし $f_h = \frac{N_h (N_h - n_h)}{n_h}$

$= \frac{N_h (1 - g_h)}{g_h}$

13 副次抽出法と多段抽出法

13.1 副次抽出法（二段抽出法）

この方法は、図のように、母集団、または階層の中から、まず第一次抽出単位を抽出し、さらにこの単位の中から、第二次抽出単位を抽出し、その全要素について調査をするものである。

第二次単位からさらに第三次単位を抽出し、全数調査を行うときは、いわゆる三段抽出法となる。これを拡張して行けば、多段抽出法となる。

副次抽出の利点はまず、各集落を全数調査する費用、労力でより多くの集落にわたり、その一部を調査する方が、標本が母集団を広くカバーするということのため、精度があがることである。つぎには、調査の内容から全数調査ができないものであることである。たとえば、事業区の全蓄積を知るため、小班を抽出する場合、各小班の全林毎木調査は困難なため、その中から、さらに小林分を選んで調査する。また、国有林の資源調査では $\frac{1}{20,000}$ の地図上で5mm内至1cmに格子線を切ることになっている。すなわち1区画が1ha　4haになっている。この全林毎木調査は不利なので、そのうちの0.04haの林分を1ヶ所あるいは2ヶ所抽出調査することになっている。これも副次調査の一種と見ることもできる。このほか、酪農工場の生産品であるバタの含有率、ある地方における小麦の蛋白

含有量、林木の葉の病虫害、人の赤血球の数などの調査では、どうしても副次抽出によらないと不可能である。

この副次抽出の利点は単純抽出に比べ、リストを作るのが簡単なことである。すなわち第一次抽出単位のリストは比較的簡単に作れ　第二次抽出単位のリストは抽出された集落についてのみ作ればよい。

13.2 母平均とその分散の推定

N = 母集団の集落の数

M = 集落内の要素の数（集落内の要素の数は同じとする）

y_{ij} = i 番目の集落の j 番目の要素の値

$\bar{y}_{..}$ = 母集団の要素の平均値

$$= \frac{1}{NM} \sum_{i}^{N} \sum_{j}^{M} y_{ij}$$

n = 標本内の集落の数

m = 抽出された集落からさらに抽出される要素の数

$$\bar{y}_{nm} = \frac{1}{nm} \sum_{i}^{n} \sum_{j}^{m} y_{ij}$$

$\bar{y}_{nm}$ は $\bar{y}_{..}$ の不偏推定値である。

$\bar{y}_{nm}$ の分散は

$$V(\bar{y}_{nm}) = \left(\frac{1}{n} - \frac{1}{N}\right) S_b^2 + \left(\frac{1}{m} - \frac{1}{M}\right) \frac{S_w^2}{n}$$

ただし S_b^2 = 母集団の集落間の平均平方

$$= \frac{\sum (\bar{y}_{i.} - \bar{y}_{..})^2}{N-1}$$

S_w^2 = 母集団の集落内の平均平方

$$= \frac{1}{N(M-1)} \sum_{i}^{N} \sum_{j}^{M} (y_{ij} - \bar{y}_{i.})^2$$

分散は2成分からなることがわかる。

$n = M$ のときは、分散は第1成分だけからなる。したがってこの第2項は、副次抽出から生じた分散への影響である

N、M が n、m に比して大きいとき（普通比率 10 % 以下）

$$V(\bar{y}_{nm}) = \frac{S_b^2}{n} + \frac{S_w^2}{nm}$$

となる。

したがって、分散を推定するには、S_b^2 と、S_w^2 を推定しなければならない。S_b^2 と S_w^2 を対応する標本の量とすれば

S_b^2 は $s_b^2 - \left(\frac{1}{m} - \frac{1}{M}\right) s_w^2$ により推定され

S_w^2 は、s_w^2 により推定される。

これを上式に代入すると

$V(\bar{y}_{nm})$ は $\frac{N-n}{nN} S_b^2 + \frac{M-m}{mMN} \cdot S_w^2$

$$= \frac{1}{mn} \left\{ \frac{N-n}{N} m S_b^2 + \frac{M-m}{MN} \cdot n S_w^2 \right\}$$ から推定できる

N、M が大きいと

$\frac{S_b^2}{n}$ だけでよいことがわかる。

すなわち

$$S^2_{\bar{y}_{nm}} \frac{S_b^2}{n} = \frac{\sum (\bar{y}_{i.} - \bar{y}_{nm})^2}{n(n-1)} = \frac{\sum (y_{i.} - y_{..})^2}{m^2 n (n-1)}$$

$$\left[\begin{array}{l} y_{i.} = m \bar{y}_{i.} \\ y_{..} = m \bar{y}_{nm} \end{array} \right) \text{とおく}$$

この式は、三段抽出の公式に容易に拡張できる。

N、M、Pを大とすれば

$$V(\bar{y}_{nmp}) = \frac{S_b^2}{n} + \frac{S_w^2}{nm} + \frac{S_p^2}{nmp}$$

S_w^2 は $S_b^2 - \frac{S_w^2}{m}$

S_w^2 は $S_w^2 - \frac{S_p^2}{p}$

S_p^2 は S_p^2

} によって推定されるから

$V(\bar{y}_{nmp})$ は $\frac{S_b^2}{n}$ により推定できる。

このことから二段抽出以上の多段抽出では、各集落のNMPが大きいときは、その分散は、第1段の集落の要素の平均値の分散の推定値により推定できることがわかる。

このことは、副次抽出や多段抽出における計算を極めて容易ならしめる。

例、昭和37年2～3月天城国有林での概本調査実習で（副次抽出を行なう）

第四層（30～39年の令級）で次のような結果を得た。この調査では地図を1/5に区劃し、それから任意に1/5の区劃を第Ⅲ、第Ⅳ層では、それぞれ6ケヒタケを抽出し、その中で0.04haの円形林地を2ケ系統的に抽出調査した。

樹種は、すぎ、ひのき、広葉樹である。

第	四	層	第三層
第1次単位	第2次単位	材積	材積
1	1	3.85	7.55
	2	4.73	8.46
	計	8.58	16.01
2	1	0.86	2.47
	2	0.83	4.46
	計	1.69	6.93
3	1	0.93	8.02
	2	1.71	9.93
	計	2.64	17.95
4	1	13.73	5.02
	2	10.51	5.87
	計	24.24	10.89
5	1	4.48	6.30
	2	9.96	4.15
	計	14.44	10.45
6	1	3.33	5.36
	2	9.92	8.50
	計	13.25	13.86
7	1	2.88	
	2	1.32	
	計	4.20	
8	1	3.91	
	2	1.57	
	計	5.48	
9	1	7.04	
	2	10.88	
	計	17.92	
	計	92.44	76.07

	第四層	第三層
層面積	129.11 ha	96.64 ha
層の全要素数	$N = \frac{129.11}{0.04}$	$\frac{96.64}{0.04} = 241$
割当個数	$n = 9 \quad m = 2$	$n = 6. \ m = 2$
調査面積	$0.04 \times 9 \times 2 = 0.72$ ha	0.48
拡大常数	$\frac{N}{n} = \frac{129.11}{0.72} = 179.32$	201.33
標本内総蓄積	$\sum_{i=1}^{9} y_i = 92.44...$	76.07
総蓄積	$y = 92.44 \times 179.32 = 16576.34 m^3$	15,315.17
平均蓄積	$\bar{y} = \frac{92.44}{9 \times 2} = 5.14$	6.34
分算推定値の計算		
二乗和	$\sum_{i=1}^{9} y_i^2 = 1423.8926$	1046.0065
補正項	$\frac{(\sum y_i)^2}{9} = 949.4615$	964.4483
平方和	$\sum (y_i - \bar{y})^2 = 474.4311$	81.5656
平均平方	$\frac{\sum (y_i - \bar{y})^2}{9-1} = 59.3039$	16.3131
要素の平均蓄積の平均平方	$S_{\bar{y}_i}^2 = \frac{1}{40} \times 59.3039 = 1.48260$	0.4078
標準誤差	$S_{\bar{y}_i} = 3.85 m^3$	2.02
CV	$\frac{3.85}{5.14} \times 100 = 0.749...$	0.319
要素の平均蓄積の平均の分散	$S_{\bar{y}}^2 = 1.48260/9 = 1.6473$	$\frac{0.7829}{6} = 0.1$
合上・標準誤差	1.283	0.825
百分率誤差（95% 信頼区間で）	約50%	約2.60%

上の計算は N. M が大きいので f. P. C を省略して計算をしたがもしf. P. C が省略できないときは次のように分散分析

表を作って、S_6^2 や S_{10}^2 を推定しなければならない。

第三層

データ

7.55	2.47	9.93	5.10	6.30	5.36	
8.46	4.46	8.02	5.97	4.15	8.50	
計16.01	6.93	17.95	10.87	10.45	13.86	76.07

計算

1. $\sum Y_{ij} = 76.07$
2. 補正項 $C = \frac{(76.07)^2}{12} = 482.2204$
3. 全体 $\sum y_{ij}^2 = (7.55)^2 + (8.46)^2 + \cdots + (5.36)^2 + (8.50)^2 - C$
   $= 534.3409 - 482.2204$
   $= 52.6205$
4. 各ノ n_i 区劃の要素平均 $= \frac{(16.01)^2 + (6.93)^2 + \cdots + (13.86)^2}{2} - C$
   $= \frac{1046.0065}{2} - 482.2204 = 523.0033 - 482.2204$
   $= 40.7829$

分散分析表

変動因	自由度	平方和	平均平方
全体	11	52.6205	
第一次単位	5	40.7829	8.1565 ($=2S_{\bar{y}}^2$)
第二次単位	6	11.8376	1.9729 ($=S_w^2$)

これら母集団の S_b^2 の推定値は

$$\frac{8.1565 - 1.9729}{2} = 3.0918$$

$S_w^2 = 1.9729$

$V(\bar{y}_{sm}) = \frac{S_b^2}{n} + \frac{S_w^2}{nm}$ より

$= \frac{1}{6}\left(\frac{8.1565-1.9729}{2} + \frac{1.9729}{2}\right)$

$= \frac{1}{6} \cdot \frac{8.1565}{2}$ ← $\frac{S^2}{n}$ に等しい

$= 0.6797$

この場合 f、P、C を省略しなければ

$V(\bar{y}_{sm}) = \frac{1}{mn}\left[\frac{N-n}{N} m S_b^2 + \frac{M-m}{MN} n S_w^2\right]$ より

$= \frac{1}{6\times 2}\left(\frac{241600-6}{241600} \times 8.1565 + \frac{25-2}{25\times 241600} \times 6 \times 1.9729\right)$

$= \frac{1}{12}\left(\frac{241594}{241600} \times 8.1565 + \frac{23\times 6\times 1.9729}{6042000}\right) = \frac{1}{12}$

$(8.1563 + 0) = 0.6797$

この値は $\frac{S^2}{n}$ に差がないことがわかろう。

今この母集団を

i) 第1次単位を12にして、第2次単位を1ヶ抽出すると分散は、

$V_i = \frac{3.0918}{12} + \frac{1.9729}{12} = \frac{5.0647}{12} = 0.4220$

ii) 第1次単位を3にして、第2次単位を4抽出すると

$V_{ii} = \frac{3.0918}{3} + \frac{1.9729}{12} = 1.1783$

この結果から第1次単位を多くとり、第2次単位を減じた方が、精度がよくなることがわかるが、そのために調査個所が分

散し、労力費用が多くかかる。したがって次節にのべるように費用を考慮して、各段階の割当をきめなければならない。なお上の i) の方法は、国有林の資源調査で（層化）単純無作為抽出といっているものに相応する。

前に示した天城国有林を例にとると、図のように全体を4ケの要素よりなる20ケの集落にわけて、その分散分析表を作って見ると次の通りになる。

変動因	自由度	平方和	平均平方
全体	119	2542.1111	
ブロック	19	1348.4444	70.9707 ($=7S_b^2$)
プロット内	100	1206.6667	7.6917 ($=S_w^2$)

1	6	11	16
2	7	12	17
3	8	13	18
4	9	14	19
5	10	15	20

1	4	7
2	5	8
3	6	9

母集団の $S_b^2 = \frac{70.5097 - 7.5417}{7} = \frac{62.9680}{9} = 7.1409$

となる $S_w^2 = S_b^2$ である。

この値を用いて次の i)、ii) の調査を行なうとすれば、そのときの平均値の分散は次のとおりである。

i) 全体の20ブロックから、7ブロックを抽出し、さらに各ブロックから4ケの $(20\times20)m^2$ の林分を抽出調査するとその全体の平均値の分散は、

$\frac{1}{28}\left(\frac{20-7}{20} \times 70.5097 + \frac{9-4}{20\times 9} \times 7 \times 7.5417\right)$

$= 1.6901$

ii) 全体から4ブロックを抽出し、さらに各ブロックよりクケの（20×20）m^2 の林分を抽出調査する。

平均値の分散は

$$\frac{1}{28}\left(\frac{20-4}{20}\times 70.5497+\frac{9-7}{20\times 9}\times 4\times 7.5417\right)$$

$$= 2.0227$$

この両法ともに抽出率は、$\frac{28}{180}\times 100 = 15.6\%$ であるが、その相対的効果は、$\frac{2.0227}{1.6901} \fallingdotseq 1.20$ で ii) 法が、2割も能率がよいことがわかる。

以上の2例や公式からわかるように、n は S_e^2、と S_w^2 の何れにも関係するのに、m は S_w^2 に関係し しかと S_w^2 は S_e^2 よりずっと小さなため、n を大きくすること、すなわち第一次単位を多くした方が精度が向上することがわかる。

次に3段抽出の例をあげる。

例3. 印度の農業調査で *Panse* は綿のサンプリングについて次の分散分析表を発表している。調査対象の地方を6の階層に別け、各層から10ケ村を選び、各村から2の畑を、さらに各畑から2プロットを抽出した。N、M、Pを、大きいものとして平均の推定値の分散を計算しよう。

変動因	自由度	平均平方
村	54	1595 $(=4S_e^2)$
畑	60	1150 $(=2S_w^2)$
プロット	120	154 $(=S_p^2)$

また i) 畑の数を2から4に増し、畑ごとに、1プロットをとる。

ii) 各村から2の畑、各畑から4プロットを取る。

iii) 畑の数は変えないで すなわち1村2の畑をとりその全数調査を行う。

$S_{\bar{y}_{nmp}}^2 = \frac{S_e^2}{n}$ から $\frac{1595}{4\times 60} = 6.65$

また、S_b^2、S_w^2、S_p^2 の推定値を求めると、

$S_e^2 \longrightarrow \frac{1595}{4} - \frac{1150}{4} = \frac{445}{4} = 111.25$

$S_w^2 \longrightarrow \frac{1150}{2} - \frac{154}{2} = 498$

$S_p^2 \longrightarrow 154$

$S_{\bar{y}_{nmp}}^2$ は $\frac{S_e^2}{n}+\frac{S_w^2}{nm}+\frac{S_p^2}{nmp}$ で推定されるから

$V(\bar{y}_{nmp})S_{\bar{y}_{nmp}}^2 = V_0 = \frac{111.25}{60}+\frac{498}{120}+\frac{154}{240} = 6.65$

i) の分散は $n=60$、$m=4$、$p=1$ として上式に代入すれば、

$$V_i = \frac{111.25}{60} + \frac{498}{240} + \frac{154}{240} = 4.57$$

同様

$$V_{ii} = \frac{111.25}{60} + \frac{498}{60} + \frac{154}{240} = 10.79$$

$$V_{iii} = \frac{111.25}{60} + \frac{498}{120} + 0 = 6.00$$

この例でも例1．2と全く同じことが言える．すなわち．V_0 と V_{iii} を比較するとわかるが．全く同じような抽出を行い．畑を全数調査した場合と2プロットサンプリングした場合の分散の相違はわずかである．これは畑内のプロット間の変動が小さいことによる．またV_iとV_0を比べると．畑の数の増加により分散が小さくなることが明らかである．これは全変動の成分としては畑の間の変動が大きいためである。

13.3 各段階における標本の割当

上述の $\frac{S_a^2}{n} + \frac{S_w^2}{nm}$ の平均値の分散推定式より2段抽出では．各段階におけるnとmの割当のいかんにより、サンプリングの精度がきまってくることはわかるがこれはまた、前節の例からも知られたところである。逆に誤差分散を一定におさえた場合には．nとmの色々の組合せが生じ、その判定が困難である。その場合．判断の基礎となるのが費用である．前節の例から明らかなように費用が標本の大きさに正比例するときは、$m=1$とするのが最も良い。3段抽出の場合の最適割合は$m=1$，$p=1$ となる。

例.1 前節の例3の資料により

iv) $n=120$，$m=1$，$p=2$.

v) $n=120$，$m=2$ $p=1$.

vi) $n=240$，$m=1$，$p=1$

の場合の分散を計算しよう。

$$V_{iv} = \frac{111.25}{250} + \frac{498}{120} + \frac{154}{240} = 5.71$$

$$V_v = \frac{111.25}{120} + \frac{498}{240} + \frac{154}{240} = 3.64$$

$$V_{vi} = \frac{111.25}{240} + \frac{498}{240} + \frac{154}{240} = 3.18.$$

種々のn,m,pに対し，次のような分散が得られる

		$n=60$			$n=120$			$n=240$	
	$p=$	1	2	4	1	2	4	1	2	4
m	1	12.7	11.4	10.8	6.4	5.7		3.2		
	2	7.3	6.6		3.6					
	4	4.6								

$m=1$，$P=1$のとき村の数が 60，120，240 と増加するにつれ分散は、.127，6.4，3.2と減少する。

$n=60$，$n=120$ の場合の畑の数が1から2，2から4に増えても，分散の減少の程度は上ほどでない。Pプロット数についても同じことが言える。分散は，$n=240$，$m=1$，$P=1$の場合が最小である。

このように費用が標本の大きさに比例することはほとんどないだろう。上例でも他村の畑を調査するより，同じ村の畑を調査する方は費用が少くてすむだろう。国有林の資源調査における層化単純抽出よりは層化副次抽出の方が労力費用が少なくてすむだろう。（共に層化副次抽出法で前者は$m=1$，後者は$m=2$の場合に相当する）

今全費用をCとし、第一次単位の調査費用をC_1，第二次単位の調査費用をC_2としよう。共通の費用は一応考慮の外におけば近似的に$C=C_1n+C_2nm$であらわされる。誤差分散をVとする。（$V=\frac{S_b^2}{n}+\frac{S_w^2}{nm}$）

この場合次の2が考慮される。

i) 費用が一定な場合できるだけ分散を小さくする。

ii) 分散の目標が定められており、費用をできるだけ小さくする。

この2は、根本的には、同じことになる。いずれの場合も，積

$$VC=\left(\frac{S_b^2}{n}+\frac{S_w^2}{nm}\right)(C_1n+C_2nm)$$

を最小にすればよい

$$V=(S_b^2C_1+S_w^2C_2)+mS_b^2C_2+\frac{S_w^2C_1}{m}$$
$$=\left(\sqrt{C_1S_b^2}-\sqrt{C_2S_w^2}\right)^2+\left(\sqrt{mC_2S_b^2}-\sqrt{\frac{C_1S_w^2}{m_2}}\right)^2$$

したがって $m=\sqrt{\frac{C_1S_w^2}{C_2S_b^2}}$

$$=\sqrt{\frac{C_1}{C_2}}\times\frac{S_w}{S_b}=\sqrt{\frac{C_1}{C_2}\left(\frac{1}{P}-1\right)}$$

nは費用が定められていれば，費用の式から$\hat{n}=\frac{C}{C_1+mC_2}$として求められ，分散に限界($V_0$)が設けられていれば，分散の式から，

$\hat{n}=\frac{S_b^2+S_w^2/m}{V_0}$ として求められる。

上のmの式から，mはC_1，C_2 に関係し、級内相関係数Pにも関係することがわかる。一般に

i) 一次単位間の旅行の費用や、C_1を構成する他の費用が安い場合

ii) 一次単位内の二次単位の調査費用が大きい場合

iii) 級内相関が高い場合

は，mは小さくなる，

最適な副次抽出の割合は，同一母集団でも調査項目ごとに変ることはもちろんであるが，Pが余り変らないような互いに関連ある項目については、明らかに効率を失わずに十分最適なmを推量することができる。

もしPが変れば，各項目にわたり，最適な割当が不可能になる。これは一般的な目的の調査についてよくあてはまることで、このよ

うなときは、関連ある調査項目を一まとめにして、群を作り、各群ごとにちがった抽出計画を作るとよいであろう。

mの式で気の付くことは、C_1, C_2の絶対値がわからなくともその比がわかればmが計算できることである。S_δ, S_w についても同様である。しかも一般にS_δ はS_w に比べて大きいのでC_1, C_2 の比がかなり大きくないと m が1以上にならない。

これらのことを次の若干の例について見よう。

例2. 前の印度の例で$C = 7n + 2nm$であるとき、その地方の平均収量を抽出分散δで、推定するにはn, mをどのようにきめるべきか。

$n = \dfrac{S_\delta^2 + S_w^2/m}{V_0}$ の式に $m = 1, 2, 4, 6$ を代入して、nの値を求める。これを費用の式に代入して費用を計算すると次のようになる。

m	$\frac{S_w^2}{m}$	$S_\delta^2 + \frac{S_w^2}{m}$	$n = \frac{S_\delta^2 + \frac{S_w^2}{m}}{\delta}$	C
1	498	609	101	909
2	249	360	60	606
4	124	235	39	585
6	83	194	32	608

費用は、$m = 4$, $n = 39$のときが最低になる。

なお直接、式から計算すると

$$m = \sqrt{\frac{7}{2} \times \frac{498}{111.25}} = \sqrt{15.6} \fallingdotseq 4$$

$$n = \frac{111 + 124}{6} = \frac{235}{6} \fallingdotseq 39$$

例3 一つのかぶからランダムに選ばれた4枚の葉の各について4回にわたりカルシウムの含有量を測定した。そのデータおよび分散分析表は次のとおりである。

葉	カルシウムの百分率				和	平均
1	3.28	3.09	3.03	3.03	12.43	3.111
2	3.52	3.42	3.38	3.38	13.76	3.44
3	2.88	2.80	2.91	2.76	11.25	2.81
4	3.34	3.38	3.23	3.26	13.21	3.30

分散分析表

反動因	自由度	平均平方	推定されるパラメーター
葉	3	0.2961	$S_w^2 + 4S_\delta^2$
測定	12	0.0066	S_w^2

$$S_\delta^2 = \frac{0.2961 - 0.0066}{4} = 0.0724, \quad S_w^2 = 0.0066$$

もし経済的な1葉当りの測定数は

$$m = \sqrt{\frac{C_1}{C_2} \frac{S_w^2}{S_\delta^2}} = \sqrt{\frac{0.0066}{0.0724}} \sqrt{\frac{C_1}{C_2}} = 0.30\sqrt{\frac{C_1}{C_2}}$$

mは整数でなければならないからC_1はC_2より相当大きくない限りは$m = 1$でよい。しかしC_2は化学的測定を含むから、逆にC_2の方がC_1より大きいだろう。従って大抵の場合は$m = 1$でよいだろう。この

場合は費用の比についての詳細な情報がない場合でも m の決定が上式からできることがわかる．なお n の値は標本誤差の影響をうけ

三段抽出の場合は標本平均の分散は

$$V = S^2_{\bar{y}_{nmp}} = \frac{S_b^2}{n} + \frac{S_w^2}{nm} + \frac{S_e^2}{nmp}$$

となり，費用は，

$$C = c_1 n + c_2 nm + c_3 nmp$$

となる．前と同様の計算で VC を最小にすれば

$$m = \sqrt{\frac{c_1}{c_2}\frac{S_w^2}{S_b^2}} \quad , \quad p = \sqrt{\frac{c_2}{c_3}\frac{S_e^2}{S_w^2}}$$

となる．これを費用が与えられておれば費用の式に代入して，n をきめる．（分散が与えられておれば分散の式によりきめる）。

13.4 層化副次抽出法

母集団が L 層にわけられ，各層ごとに1次単位の n_h 個が抽出され，さらに一次単位から m_h 個を抽出したとする。ただし h 番目の層は，1次単位は N_h 個，2次単位は M_h から構成されているものとする。母平均は

$$\bar{y}_{n_h m_h} = \frac{\sum_{i=1}^{n_h}\sum_{j=1}^{m_h} y_{hij}}{m_h n_h} \text{ とすれば，} \bar{\bar{y}}_{st} = \frac{\sum_h M_h N_h \bar{y}_{n_h m_h}}{\sum M_h N_h} \text{ で}$$

不偏に推定される．

分散の不偏推定値は

$$v(\bar{\bar{y}}_{st}) = \frac{\sum_h \frac{(M_h N_h)^2}{m_h n_h}\left\{\frac{(N_h - n_h)}{N_h} m_h S_{eh}^2 + \frac{(M_h - m_h)}{M_h}\frac{n_h}{N_h} S_{wh}^2\right\}}{(\sum M_h N_h)^2}$$

$\frac{n_h}{N_h}$，$\frac{m_h}{M_h}$ が小さいときは，（0.10以下）上式は，次のように簡単になる。

$$v(\bar{\bar{y}}_{st}) = \sum_h \frac{(M_h N_h)^2}{n_h} S_{bh}^2 \Big/ (\sum M_h N_h)^2$$

国有林の資源調査におけるように1次単位の大きさ M_h と2次単位の抽出数が等しい場合は

$$v(\bar{\bar{y}}_{st}) = \sum_h \frac{N_h^2}{n_h} S_{bh}^2 \Big/ (\sum_h M_h)^2$$

$$= \frac{1}{N^2}\sum_h \frac{N_h^2}{n_h^2} S_{eh}^2$$

したがって総蓄積の分散は

$$v(Y) = \sum_h \frac{N_h^2}{n_h^2} S_{eh}^2$$

天城国有林において昭和35年行なった実例を次に示す。

調査対象は第2種林地の第2営用施業団中の人工林で10年生を超した面積は次のとおりである。

層	林令	面積
I	11～20年	189.75
II	21～30	202.62
III	31～40	90.79
IV	41～50	153.76
V	51～60	246.54
VI	61～	225.38
	計	1122.17

この地域を$\frac{1}{20,000}$の地図上を5 mmの巾の格子網でおゝい，その交点をランダムに抽出した．現地調査では，抽出された個所と，その北方50 mの距離の個所とで，0.04 haの円形地を設定して，地内の毎木の直径と，0.04 haの同正円地内の毎木の樹高，直径，生長錐調査を実行した。な　各層の標本の割当はNeyman法を利用したがどの層でも2次単位が10％以下にならぬよう修正した。

その割当は次のとおりである．

	1次単位	2次単位
I	5	10
II	7	14
III	5	10
IV	12	24
V	10	20
VI	20	40
計	59	118

各層ごとの分散の計算の方法は，前に示したとおりである。全体の推定についての計算過程および取りまとめは次のとおりである。

推定の結果は（223,778±26,993）m^3で百分率誤差は12 %である。

総蓄積の推定（層化副次抽出法）

(1) 階層記号と手順	(2) 面積 A_k ha	(3) 割当個数 n_k	(4) 調査面積 2×(3)×0.04 $2n_k$×0.04	(5) 拡大定数 (2)/(4) $A_k/0.08n_k$	(6) 各層面積の割合 $W_k=A_k/A$	(7) 標本蓄積合計 $\sum\sum x_{kij}$ m^3	(8) 層蓄積 (5)×(7) $\frac{A_k}{0.08n_k}\sum\sum x_{kij}$	(9) 平均蓄積 (5) $\bar{x}_k=\sum\sum x_{kij}/2n_k$
I	184.72	5	0.40	474.410	0.16412	9.192	4,361,444	0.919
II	212.12	7	0.56	366.821	0.18856	54.611	19,757,467	3.401
III	45.59	5	0.40	284.926	0.04503	72.992	12,461,236	2.678
IV	158.76	12	0.96	165.375	0.14148	169.531	28,036,124	7.254
V	246.84	10	0.80	308.550	0.21997	278.415	85,905,[illegible]	13.924
VI	225.38	20	1.60	140.863	0.20084	484.435	68,240,578	12.111
計	1122.17	59					223,778,[?]	

(1) 階層記号と手順	(10) 各層プロット蓄積間分散 (C) S_{xk}^2	(11) 各層プロット平均蓄積の分散 (E) $S_{\bar{x}k}^2$	(12) (6)の2乗 $(6)^2$ W_k^2	(13) 総平均蓄積の分散の計算 (11)×(12) $W_k^2\times S_{\bar{x}k}^2$	(14) 標準偏差 (D) S_{xk}	(15) 変動係数 (F) C %
I	0.736436	0.157387	0.02360157	0.00159754	0.9652	105.3
II	7.852286	1.122715	0.03260191	0.03660396	2.4634	71.9
III	8.733447	1.746649	0.00724729	0.01553558	2.9552	41.5
IV	8.805215	0.733745	0.02001659	0.01468694	2.9674	42.1
V	26.484077	2.648408	0.04838656	0.12814744	5.1460	37.0
VI	15.913216	0.795661	0.04033571	0.03129747	3.9726	32.8
計				0.23031150		

総平均蓄積の分散：$S_{\bar{x}}^2=\sum w_i^2 S_{\bar{x}_i}^2=0.23031150$

プロット当たり（0.04 ha）の平均蓄積　$\bar{x}=(8)の計/\frac{A}{0.04}=7.977$

標準誤差：$S_{\bar{x}}$　　$=0.4799$

プロット当り蓄積と信頼度 95%の信頼区間

$$= 7.977 \pm 2 \times 0.480 = 7.977 \pm 0.960\ (7.017 \sim 8.937)$$

総蓄積と　〃　〃

$$= 223{,}777.90 \pm 26{,}932.08\ (196{,}845.82 \sim 250{,}709.98)$$

百分率誤差 = 12.0 %

14 不等確率抽出法

14.1 復元抽出

ある施業団の小班を抽出して，全体の蓄積合計を推定したとき，各小班を等確率に抽出するこれまでの方法のほかに，小班の面積または平方根に比例した確率を与えて抽出する方法がある。

また事業区の総蓄積を推定したい場合に林班を抽出単位とするとき，林班内の小班の数に比例した確率を与えて林班を抽出する方法も考えられる。この方法はたとえ林分が似ていても林小班の面積に従って蓄積の変化がはげしいわが国では，等確率に抽出する単純無作為抽出法よりも有効な場合が多い。

この抽出法では，復元抽出を行なうと，計算が簡単になる。すなわち，一度抽出されたものが再度抽出された場合でも採用して，計算の中に取り入れるものである。その抽出法は，次のように上述の2番目の場合各林班を通じて小班の累加を行ない，各林班に割当番号の範囲をつけ，乱数表で，この範囲内の数がでたら，その林班を抽出調査する。

林班番号	小班数	累加和	割当番号
1	10	10	1 − 10
2	20	30	11 − 30
3	15	45	31 − 45
4	12	57	46 − 57
5	18	75	57 − 75
6	30	105	76 − 105
7	5	110	106 − 110
8	24	134	111 − 134

乱数表で 43 とでたら4林班 120 とでたら8林班を抽出する。

この方法は手数がかゝるが D. B. Lanisi の方法は，簡単である。すなわち，a，b 2数をまずえらぶ，a は集落全体の数，等しいかより大きい数で，b は各集落内の単位の数の最も大きいものに等しいかより大きい数である。この a，b をこさない数を抽出する方法である。上例で3桁の数を抽出してはじめ 109 とでたら 1番の4番となるので，これを採用する。次に，550 とでたら5番は 121 が小班はないからこの乱数表はすてる。こうして必要数だけ抽出する方法である。

i 番目の個体が抽出される確率 P_i はこの場合 $\dfrac{X_i}{\sum X_i} = \dfrac{X_i}{X}$ となる。

$Z_i = \dfrac{y_i}{NP_i}$ とすると，$E(Z_i) = \bar{y}_N$（母平均）

となる。したがって $E(\bar{Z}_n) = \bar{y}_N$ となる。母平均の推定値は

$\bar{z} = \frac{1}{n}\sum_{i=1}^{n} z_i = \frac{X}{Nn}\sum_{i=1}^{n} \frac{y_i}{X_i}$，$Y_i = \dfrac{y_i}{X_i}$ とすれば

$\bar{Z} = \frac{X}{nN} r_i = \frac{\bar{X}}{n} \Sigma r_i$ となる。

$\bar{Z}$ の分散の推定値は

$$S^2(\bar{Z}) = \frac{1}{n(n-1)} \sum_{i=1}^{n} (Z_i - \bar{Z})^2$$

$$= \frac{1}{n(n-1)} \left\{ \sum_{i=1}^{n} Z_i - n\bar{Z} \right\}$$

$$= \frac{\bar{X}^2}{n(n-1)} \left\{ \Sigma r_i^2 - \frac{(\Sigma r_i)^2}{n} \right\}$$

例1．35の農場をその面積に比例にして抽出してその収穫量を調査した．その結果は次のとおりである．

農場番号	x(面積)	y(収穫量)	r
3	62	10	0.1423
18	110	24	0.2182
28	300	59	0.1967
34	410	72	0.1756
35	340	103	0.2395

$\Sigma r_i = 1.0023$

$\bar{r} = 0.2045$

$\Sigma r_i^2 = 0.21143703$

$\Sigma (r_i - \bar{r})^2 = 0.21143703 - (1.0023)^2 / 5$

$= 0.00245752$

$$S_r^2 = \frac{\Sigma (r_i - \bar{r})^2}{n-1} = \frac{0.00245752}{4} = 0.00061438$$

$$v(\bar{r}) = \frac{\Sigma (r_i - \bar{r})^2}{n(n-1)} = \frac{0.00061438}{5} = 0.00012288$$

母集団では

$N = 35$　$\sum_N x = 5759$　$\bar{X} = 164.54$

したがって，$\bar{y} = \bar{z} = \frac{\bar{X}}{n} \Sigma r_i = 0.2045 \times 164.54$

全体の $Y = 0.2045 \times 5759 = 1178$

分散の推定値は

$v(\bar{y}) = (164.54)^2 \times 0.00012288 = 33.268$

$v(Y) = (5759)^2 \times 0.00012288 = 4075$

$\bar{y}$ の95%信頼帯は

$33.65 \pm 2.78 \times \sqrt{33.268} = 33.65 \pm 5.07$

$= 28.58 \sim 38.72$

この場合 x と y の相関が高いほど効率は高くなる。抽出確率が y の大きさに比例するときは推定値の分散は0となる。

この例は，既述のように抽出単位をもとに戻す方法を前提としている。もし抽出された単位がもとに戻されなければ，2番目以降に抽出される単位の抽出確率は P_i と変ってくる。このような平均値の分散の計算は2単位の標本以上は難しくなる。また実際的にはこの抽出は集落抽出や多段抽出の1次単位の抽出に普通用いられる。

もし林地などが地目上に，明確に区別されている場合，地目上にランダム点をおとして，おちた点の周囲の林地をえらべば林地の不等確立標本となる。もし同じ場所が2度以上あたれば，あたった回数だけ標本に算入されなければならない。またもし $\frac{n}{N}$ が0.10以下の

ときは、この抽出のとき非復元抽出を行なっても偏りは小さい。

なお、相関の高い変数に比例した確率でこの方法を行なおうと決定する前に、層化抽出を行えるかどうか吟味した方がよい。とくに相関の高い変数を用いて層化することができるかどうかの検討が望ましい。ことに、そのような層化が各層の平均値が非常にちがってくるときは層化する方が有利である。これは大きさに比例した確率抽出法は、単位の大きさが小さいときは、層化抽出より手続きが面倒だからである。

例1はxは計量であったが、計数であってもよい。N個の集落からなる母集団からn個の標本を確率 $P_i=\frac{M_i}{\sum_{i=1}^{N} M_i}$ で抽出し、各集落を全数調査する場合も前と同様になる。すなわち、

平均値の推定 $\bar{y}=\frac{1}{n}\sum_{i=1}^{n}\frac{y_i}{M_i}=\frac{1}{n}\sum_{i=1}^{n}\bar{y}_i$

平均値の分散の推定 $S^2(\bar{y})=\frac{1}{n(n-1)}\sum_{i=1}^{n}(\bar{y}_i-\bar{y})^2$

例2　53郡の農場の家畜の総数を知りたい、各郡の農場数は M_i 個で異なるので郡の農場数に比例した確率で14郡をさらんだ。その資料は次表のとおりである。

標本郡 i	農場数 M_i	家畜数 $Y_i=M_i\bar{y}_i$	郡あたりの平均家畜数 $\bar{y}_i$
1	19	66	3.47
2	28	326	11.64
3	28	392	14.00
4	29	350	12.07
5	31	331	10.68
6	31	331	10.68
7	46	697	15.15 …
8	51	586	11.49
9	53	739	13.94
10	55	914	16.62
11	61	619	10.15
12	64	784	12.25
13	83	906	10.92
14	83	946	11.42
計	662	7947	163.98

平均値の推定 $\bar{y}=\frac{163.98}{14}=11.7129$

$\bar{y}$ の分散の推定値

$$S^2(\bar{y})=\frac{1}{14\times13}\left\{(3.47-11.7129)^2+\cdots+(11.42-11.7129)^2\right\}$$

$$=\frac{120.3341}{182}=0.6612$$

推定値は例1と同じく不偏である。

この方式は母集団のすべての単位の大きさ x_i が既知であるかまたは x の全体の値が既知である場合に適用できるが、x_i や全体の X が

既知でない場合は、$\bar{Y}$を標本から推定しなければならない。したがって後述する比推定の形をとる。例えば、地図上にランダムにまたは系統的に点をおとして行き、蓄積調査を行なうような場合を考えよう。

Aを総面積とし、n_0を標本点の総数nが調査目的物におちた点の数とすると、

母集団の$r=\Sigma\frac{Y}{X}/n$、X、Yの推定値は標本の$\bar{r}$、$A\frac{n}{n_0}$、$\bar{r}A\frac{n}{n_0}$より推定される。もしAが厳密に知られていないときは、この式の代りに単位面積あたりの密度を用いるとよい。その場合$A=\frac{n_0}{d}$、$X=n/d$となる。

例3　ある事業区のすぎ人工林の面積と蓄積を調査するため、地目（または集成航空写真）上で1haに4点の密度で、系統的に点を配置し、点がおちた所で0.1haのプロットで現地調査された。全部の点数は10,000点で、そのうちすぎ林におちた点数は2000点であり現地調査されたすぎ林の0.1haあたり20m^3であった。すぎ人工林の総面積と蓄積を推定せよ。

面積$=X=\frac{2000}{4}=500$ ha

蓄積$=Y=200^{m^3}\times 500=100{,}000^{m^3}$

1haあたりの蓄積の標準偏差を10m^3と仮定して点の分布がランダムとしてこの場合の抽出誤差を推定すれば

$V(\frac{n}{n_0})=\frac{n}{n_0}(1-\frac{n}{n_0})/n_0$

Aが正確に判っていれば、

$V(X)=A^2\frac{n}{n_0}(1-\frac{n}{n_0})/n_0=\frac{X^2}{n}\cdot\frac{n_0-n}{n_0}$

$V(Y)=X^2V(\bar{r})+\bar{r}^2V(X)=\frac{X^2}{n}(S_r^2+\frac{n_0-n}{n_0}\bar{r}^2)$

て、Aが正確にわかっていない場合でも一般に分散の増加はそれほどでなく、無視してもよい。この場合$A=\frac{n_0}{d}$とすれば

$V(X)=\frac{n(n_0-n)}{d^2n_0}$

この例で0.1ha あたりの平均蓄積の標準誤差は

$SE(\bar{r})=\frac{1}{\sqrt{2000}}=0.00224$

$V(X)=\frac{2000\times 8000}{4^2\times 10.000}=100\,(ha)^2$

$SE(X)=10\,ha$

$V(Y)=\frac{(500)^2}{2000}\left(100+\frac{8000}{10.000}\times(200)^2\right)$

$=\frac{250}{2}(100+32.000)$

$=125\times 32100$

$=4012.500(m^3)^2$

または$V(Y)=(500)^2\times\frac{100}{2000}+(200)^2\times 100=400125$

として計算してもよい

$SE(Y)=2003\,m^3$

95%の信頼区間は$(10.000\pm 4006)m^3$となる。

例4、例3の場合、さらに1haあたり16点の密度となるようにしたところ、全体で20080点追加され、すぎ人工林には6007点追加された。例3に対する修正面積、蓄積を求めよ。

面積　$X = 8007/16 \fallingdotseq 501$ ha

蓄積　$Y = 501 \times 200 = 100,200\ m^3$

$$V(X) = \frac{1}{16^2} \frac{200^2 \times (39990 - 8007)}{39990} = 25.01\ (ha)^2$$

$SE(X) \fallingdotseq 5$ ha

$$V(Y) = \frac{(5 \times 10)^2 \times 1}{2000} + (20)^2 \times 2501$$
$$= 1250 + 1,000,400$$
$$= 1,012,550\ (m^3)^2$$

$SE(X) \fallingdotseq 1024\ m^3$

故に95%信頼区間は $(10200 \pm 2048)\ m^3$ で二相抽出の結果誤差は前の半分になっている。これは面積の誤差が半分になったことに基く。

14.2　二段抽出法における大きさに比例した確率を与えて抽出する方法。

各集落の大きさが異なる場合1次単位をその大きさ（この場合、単位の個数とする）に比例した確率で n 個の復元抽出を行ない、さらに2次単位をランダムに m_i 個抽出する二段抽出の場合を考えよう。

この場合は、$\bar{y}_p = \frac{1}{n} \sum_{i=1}^{n} \bar{y}_i$ が母平均の不偏推定量であって、その分散は

$$V(\bar{y}_p) = \frac{S_{Pb}^2}{n} + \frac{1}{n} \frac{1}{N} \sum_{i=1}^{N} \frac{M_i}{M} \left(\frac{1}{m_i} - \frac{1}{M_i} \right) S_w^2$$

でこれは

$$\hat{V}(\bar{y}_p) = \frac{S_{Pb}^2}{n} \quad \text{ただし} \quad S_{Pb}^2 = \frac{n}{n-1} \sum_{i=1}^{n} (\bar{y}_i - \bar{y}_p)^2$$

により不偏に推定される。ここで費用函数を想定して最適な m_i を求めて見ると、m_i は第 i 集落内の標準偏差 σ_i に比例しなければならないことがわかる。しかし、一般に $m_i = m\ (i = 1, \cdots, n)$ として行なうことが便利である。すなわち、

$$\bar{y}_p = \frac{1}{m} \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m} y_{ij}$$

例1．5郡のすべての農場の牛の1農場あたり平均頭数を推定し、その抽出誤差を推定したい。各郡には M_i 個の農場があり $P_i = \frac{M_i}{\sum_N M_i}$ の確率で、5郡より4郡が抽出され、各郡からはランダムに $\frac{1}{4}$ の抽出比で農場が抽出され、牛の頭数が調査された。その結果は次の表のとおりである。

標本郡	農場数 M_i	標本内の農場数 m_i	標本内牛の頭数 $\sum_{j=1}^{m} y_{ij}$	標本農場あたりの牛の平均頭数 $\bar{y}_i$
1	13	3	36	12.0000
2	15	3	58	19.3333
3	19	5	74	14.8000
4	28	7	93	13.2857
5	39	10	162	16.2000
6	41	11	88	8.0000
7	41	12	102	8.5000
8	46	12	102	8.5000
9	48	12	203	16.9167
10	51	13	134	10.3077
11	58	14	185	13.2857
12	74	19	272	14.3158
13	83	20	242	12.1000
14	83	20	242	12.1000
計	645	161	1917	163.4307

計算

平均値の推定 $\bar{y} = \frac{163.4307}{14} = 11.6736$

$\bar{y}$ の分散の推定 $S^2(\bar{y}) = \frac{1}{14\times13}\left\{(10.0000)^2+\cdots+(12.1000)^2-14\times(11.6736)^2\right\}$

$= \frac{2.326753}{182} = 1.30159$

この例は14.1の例2に対応するもので 例2 とちがう所は2段抽出で2次単位の平均から1次単位の平均を推定していることである。しかも全体の平均を出す場合に重みをつけていたが不偏推定である。多段抽出は集落抽出に対し、抽出する集落の数が同数のときは精度はおちるが、抽出する要素の数が同一のときは、精度が高いことは自明のことであろう。しかも次2次単位が少ない程精度が高いから、コクランは2次抽出単位は2個がよいと言っている。なお既述したように、同数の要素を含む単純無作為抽出に対しては、その精度は $\frac{1}{1+(M_0-1)\rho}$ である。(ρは級内相関係数) ρは一般に正であることが多いから、精度はより低くなることが多い。

なおこの方法の欠点は各集落の大きさを知らなければならないことだが、これがわかれば計算が簡単である。したがって2段抽出法ではこの方法が他の方法よりすすめられている。なお、とくに他の2段抽出の推定式とは全く異なることは、各推定式が模元抽出だから、f.P.Cと全く考える必要もなく、Nにも関係しないことである。

例2 集落の大きさがちがう場合の2段抽出だが100の林班からなる1施業団で、小班の平均材積を知りたいようなとき 例1のよう

な方法もあるが $P_i = \frac{X_i}{\sum_{i=1}^{N} X_i}$ (X_iは林班の総面積)の確率でM_0個の小班からなる14の林班を抽出し、さらにこの林班からm_i個の小班をランダムに$\frac{1}{4}$の抽出比で抽出調査する。y_{ij} をi林班のj小班の蓄積とする。

ただし1次単位の林班の抽出は模元抽出とする。

$Z_{ij} = \frac{M_i}{\sum_{i=1}^{N} M_i} \cdot \frac{1}{P_i} y_{ij}$ とすれば

$\bar{Z}_i = \frac{M_i}{\sum_{N} M_i} \cdot \frac{1}{P_i} \bar{y}_i$ ($\bar{y}_i = \frac{1}{m_i}\sum_{j=1}^{m_i} y_{ij}$)

平均の推定

$\bar{Z} = \frac{1}{n}\sum \bar{Z}_i$ で平均を推定する

$\bar{Z}$の分散の推定

$S^2(\bar{Z}) = \frac{1}{n(n-1)}\sum_{i=1}^{n}(\bar{Z}_i - \bar{Z})^2$ により推定

標本林班	小班の数 M_i	標本内の小班数 m_i	標本内の蓄積 $\sum_{j=1}^{m_i} y_{ij}$ 単位 千m³	標本小班の平均蓄積 $\bar{y}_i$ 千m³	林班面積 X_i ha	$P_i = \frac{X_i}{\sum_{i=1}^{N} X_i}$	$\bar{Z}_i = \frac{M_i}{\sum_{N} M_i}\cdot\frac{1}{P_i}\cdot\bar{y}_i$ 千m³
1	19	5	14	2.8000	162	0.0024	9.9078
2	23	5	82	16.4000	620	0.0093	18.4763
3	31	8	207	25.8750	928	0.0140	26.6940
4	40	10	124	12.4000	1059	0.0161	14.5307
5	54	13	113	8.6923	1187	0.0180	12.1317
6	54	13	113	8.6923	1187	0.0180	12.1317
7	39	10	114	11.4000	1397	0.0220	9.7413
8	56	14	242	17.2857	1579	0.0249	18.3972
9	46	12	203	16.9167	1636	0.0258	14.5568
10	83	20	256	12.8000	1871	0.0295	17.2237
11	74	19	272	14.3158	2297	0.0362	14.1182
12	70	17	131	7.7059	2346	0.0370	9.0393
13	60	15	208	13.8667	2746	0.0465	11.6486
14	60	15	208	13.8667	2946	0.0465	11.6486
計	708	176	2237		22180		192.1363

計算

平均の推定 $\bar{z} = \frac{192.0363}{14} = 13.71169$ 千m³

分散の推定

$$S^2(\bar{z}) = \frac{1}{14\times13}\left\{(9.9078)^2 + \cdots + (8.6486)^2 - 14\times(13.71169)^2\right\}$$

$$= \frac{35.9263}{182} = 1.9756$$

この推定も例1と同じく不偏である．また例1はこの例2の特別な例でXがMになっただけにすぎない。すなわち $P_i = M_i / \Sigma M_i$ により $z_{ij} = y_{ij}$ となり $\bar{z}_i = \bar{y}_i$ となっている。

層化多段抽出の場合は $\bar{y} = \Sigma w_k \bar{y}_k$，$V(\bar{y}) = \Sigma w_k^2 V(\bar{y}_k)$ により推定する。

14.3 大きさ(X)に反比例した確率による抽出法

これまで述べた方法は，予め定めたn個の単位を大きさに比例した確率で抽出する方法である。もしnがNに比べてかなり大きいときは，標本にはn個がすべてちがった単位は含まれないだろう。しかし，同じ単位が抽出されればその分だけ調査する必要はなく，只単にその結果を2度かくだけにすぎない．始めにn個分の調査費用がわりあてられているときは，その分だけ余ることになる。したがって全部違うn個の単位が抽出されるまで標本抽出を行えば，より多くの情報が得られ，精度もますことになるので，この方法を述べておく．丁度n番目の異なる単位を抽出すると同時に抽出をやめれば偏りがはいるから，標本がn+1番目の異なる単位が含まれるまで抽出を続ける．これがa+1番目の抽出としよう。
この最後の a+1 番目の単位は除いてしまって、計算は

a個の単位について行なう。したがって標本の大きさはaとなる。

したがって $r_i = \frac{y_i}{x_i}$ とすれば

$$\bar{r}_a = \frac{\sum_a^a r_i}{a} = \frac{\sum_a^a y_i}{a}$$

$$\bar{y} - \bar{r}_a \bar{x}_N = \frac{1}{aN} \sum_a \frac{y_i}{x_i} \sum_N X_i \quad (\text{不偏})$$

$$v(\bar{r}_a) = \frac{S_r^2}{a} \quad (S_r^2 \text{が不偏ならば不偏})$$

$$S_r^2 = \frac{1}{a-1} \sum_a (r_i - \bar{r}_a)^2 \quad (\text{偏りがあり。})$$

$\bar{r}_a$ の分散の正確な式は極めて複雑だから次の近似式を用いるとよい．

$$V(\bar{r}_a) = \frac{\tilde{\sigma}^2}{n}\left(1 - \frac{n}{2N}\right) = V(\bar{r}_n)\left(1 - \frac{1}{2}f\right)$$

$\bar{y}$ の分散の推定は，$v(\bar{y}) = \bar{x}^2 v(\bar{r}_a)$ により行なえばよい。

[付]

系統的抽出，抽出間隔 10，標本数 18

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
20	13	14	12	4	3	8	13	12	14	16	12	7	7	8	17	16	19
15	12	11	12	16	12	12	13	12	9	13	13	18	20	19	14	18	15
12	14	17	14	16	8	5	9	12	16	16	19	17	18	19	14	18	14
14	12	7	5	5	5	16	15	19	17	16	16	12	17	14	11	10	11
12	10	4	14	13	15	15	16	9	10	13	13	16	17	15	13	11	9
15	14	12	16	12	17	14	19	19	14	21	19	16	14	12	11	14	15
16	22	19	16	19	14	17	20	19	15	14	18	15	15	17	16	15	15
14	14	14	14	16	12	16	15	14	15	14	13	14	13	9	10	14	10
7	13	11	10	14	15	12	15	13	13	14	11	11	10	5	6	9	13
10	11	14	12	15	14	14	10	6	4	9	9	8	9	10	12	9	9

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	GT
総和	135	135	123	125	130	117	127	145	135	127	146	143	134	140	128	124	134	130	2339
平均	13.5	13.5	12.3	12.5	13.0	11.7	12.9	14.5	13.5	12.7	14.6	14.3	13.4	14.0	12.8	12.4	13.4	13.0	
自乗和	1929	1919	1689	1657	1904	1549	1775	2211	1997	1713	2216	2155	1924	2122	1846	1628	1904	1784	
補正項	1822.5	1822.5	1512.9	1562.5	1690.0	1368.9	1664.1	2102.5	1822.5	1612.9	2131.6	2044.9	1795.6	1960.0	1638.4	1537.6	1795.6	1690.0	
平方和	106.5	96.5	176.1	94.5	214.0	180.1	110.9	108.5	174.5	100.1	84.4	110.1	128.4	162.0	207.6	90.4	108.4	94.0	
分散 ($σ^2$)	11.166	10.722	19.56	10.499	23.7	24.45	14.54	12.05	19.38	11.56	9.37	12.23	14.26	17.99	23.06	10.04	12.04	10.44	
σ	3.3417	3.2745	4.4234	3.2404	4.8762	4.9413	3.8137	3.4721	4.4037	3.9455	3.0623	3.4976	3.7771	4.2426	4.8028	3.1699	3.4705	3.1673	
0.094×分散	1.0502	1.0122	1.8471	0.9912	2.2446	2.3086	1.3730	1.1321	1.8303	1.4665	0.8813	1.1548	1.3468	1.6992	2.1775	0.9482	1.1390	0.9857	
0.3022×σ	1.0216	1.0019	1.3587	0.9920	1.4850	1.5192	1.1716	1.0666	1.3427	1.2121	0.9407	1.0745	1.1603	1.3033	1.4754	0.9726	1.0661	0.9736	

$\frac{N-n}{Nn} = \frac{180-10}{180}\cdot\frac{1}{10} = \frac{17}{180} = 0.094$，$\sqrt{0.094} = 0.3072$

$\Sigma(平均)^2 = 3159.58$
補 3146.3889
$17)\ 10.6111$
0.624182 分散
$σ = 0.7901$

15 比推定法と回帰推定法

15.1 比推定における考え方

今ある事業区で，VI令級の林分の総材積を推定したいと思って，小班単位に標本抽出を行うことにした．この場合，全小班の平均材積は抽出されたn個の小班の材積を平均して求められ，それに全小班数NをかければVI令級の全材積が求められる．しかし，一般に小班の面積には大小の変動がはげしいのでこの場合の分散はかなり大きいことが想像され，サンプリングの精度も低下しよう．

この場合に精度をあげるためには，小班面積の類似したものを集めて階層を作り，層別抽出を行なうことも考えられるが，ここでは別の方法により精度を向上させる方法を考えて見よう．すなわち，1haあたりの各小班の材積は，面積の異なる小班の材積よりは変動が少ないことは容易に考えられるので

$$r = \frac{材積}{面積} = \frac{y}{x}$$ という比を考える．

rは上の考え方から，その変動係数（C.V.）は，yのC.V.より小さいことになる．VI令級の総材積はrの平均の値を標本を用いて推定して，総面積をかければよいことがわかろう．各小班のhaあたり材積が等しいと，標本からのrの推定値は真の値と一致するであろう．しかし，実際はrは小班ごとに異なる

のが常であるが，その変動係数が y の C.V より明かに低ければ，この推定法による y の推定値の誤差は普通の場合より小さくなるに違いない。この推定法を比推定法という。この補助情報を提供する x を補助変量という。

一般に比推定法は、補助変量 x が，母集団全部について知られており $\frac{y}{x}$ の C.V が y の C.V より小さいと思われるとき用いる。この2つの要件がみたされれば，x は何でもよいので r も学問的あるいは経済的な意味をもっている必要もない。たとえば上例で，各小班について面積でなく、胸高断面合計しか知っておらず，それも5年前の胸高断面積合計しかわかっていないとすると，r の実際的な意味は何もない。しかし，この場合でも結果に大きな偏りを生ずることはないであろう。

15.2 母平均と母集団総計の比推定の式

抽出は無作為に非復元で行なわれたものとする。したがってどの抽出単位も等確率で抽出されている。

母集団の比率 R は次により定義する。

$$R = \frac{\sum_{i=1}^{N} y}{\sum_{i=1}^{N} x} = \frac{\bar{y}_N}{\bar{x}_N} = \frac{y\text{の母集団総計(平均)}}{x\text{の母集団総計(平均)}}$$

上例では $R = \frac{\text{林令級の総材積}}{\text{林令級の総面積}}$

上式から $\bar{y}_N = R\bar{x}_N$

$\bar{x}_N$ は既知とし，R は標本から推定することにする。R は次の標本の r で推定される。

$$\bar{r}_w = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i} = \frac{\bar{y}_n}{\bar{x}_n} = \frac{y\text{の標本和(平均)}}{x\text{の標本和(平均)}}$$

各単位では $r_i = \frac{y_i}{x_i}$ だから

$\bar{r}_w = \frac{\sum^n r_i x_i}{\sum^n x_i}$ となり，$\bar{r}_w$ は r_i の x による重み付平均となっていることがわかる。なお r_i の重みなしの平均は $\bar{r}_n = \frac{\sum r_i}{n}$ とする。

$\bar{r}_w$ は R の不偏推定値ではないが，偏りは一般に大きくない。しかし $\bar{r}_n$ の方は，r が x に伴い一定の傾向で変化する場合は偏りが大きくなるであろう。

母集団の平均の推定値は，

$\bar{y}_r = \bar{r}_w \bar{x}_N$ となる。

R が偏りがあるから，$\bar{y}_r$ も $\bar{y}_N$ に対して偏りがある。

母集団総計の推定は $\bar{y}_r$ に N をかければよい。

$$\hat{Y}_r = N\bar{r}_w\bar{x}_N$$

例1 10 ha の小班でビッテルリッヒ法を行ない，次の13の断面積と材積の値を得た。小班全体の平均材積，総材積を推定せよ。なお小班の平均断面積は 40 m² であった。

断面積 (x) m^2	材積 (y) m^3	x m^2	y m^3
30	350	30	270
40	490	40	370
40	400	30	280
20	180	計 500	4910
50	440		
40	380	$\bar{r}_w = 9.82$	
50	480	$\bar{y}_r = 9.82 \times 40 = 392.8$	
40	410		
40	400	$Y_r = 392.8\,m^3 \times 10 = 3928\,m^3$	
50	460		

y の値だけから計算された平均値は

$\bar{y}_n = \frac{4910}{13} \fallingdotseq 377.7\ m^3$　　総計は

$\hat{Y}_n = 377.7 \times 10 = 3777\ m^3$

15.3 比推定の分散

$\bar{y}_n$ と $\hat{Y}$ の分散の推定式は

$$v(\bar{y}_n) = \frac{S_y^2}{n}\left(1 - \frac{n}{N}\right)$$

$$v(\hat{Y}) = N^2 v(\bar{y}_n)$$

で，S_y^2 は平均値からの個々の y の値のちらばり方，散勢を示すものであったが，比推定では $\bar{r}_w$ から $r_i = \frac{y_i}{x_i}$ という個々の

値のちらばり方により散勢が示される。すなわち $\bar{r}_w$ とからの y_i のちらばり方，散勢によって示される。したがって上式の S_y^2 の代りに次の $S_{y\cdot x}^2$ を用いないといけない。

$$S_{y\cdot x}^2 = \frac{\Sigma(y - \bar{r}_w x)^2}{n-1}$$

ゆえに

$$v(\bar{y}_r) = \frac{S_{y\cdot x}^2}{n}\left(1 - \frac{n}{N}\right)$$

$$v(\hat{Y}_r) = N^2 v(\bar{y}_r) = \frac{N^2 S_{y\cdot x}^2}{n}\left(1 - \frac{n}{N}\right)$$

比々の分散の推定値は

$$v(\bar{r}_w) = \frac{v(\bar{y}_r)}{\bar{x}^2} = \frac{S_{y\cdot x}^2}{n\bar{x}^2}\left(1 - \frac{n}{N}\right)$$

なお $v(\bar{y}_r)$ は

$$v(\bar{y}_r) = \frac{S_{y\cdot x}^2}{n}\left(1 - \frac{n}{N}\right) \qquad \left(S_{y\cdot x}^2 = \frac{\sum_{i=1}^{N}(y - Rx)^2}{N-1} \text{ とする}\right)$$

の推定値であるが，不偏ではない．したがって真の分散の近似値にすぎなく，標本の大きさが，$\frac{S_{y\cdot x}^2}{n^2}$ が $\frac{S_{y\cdot x}^2}{n}$ に比べて偏りを無視できるほど大きいことを仮定している．この偏りも近似の程度も小標本の場合以外は重要でない．

なお上の分散の推定式は計算の場合，次式による方が便利である．

$$v(\bar{y}_r) = \frac{(N-n)}{N}\frac{\Sigma(y_i - \bar{r}_w x_i)^2}{n(n-1)} = \left(1 - \frac{n}{N}\right)\frac{(\Sigma y_i^2 + \bar{r}_w^2 \Sigma x_i^2 - 2\bar{r}_w \Sigma y_i x_i)}{n(n-1)}$$

同様に比の分散は

$$V(\bar{r}_w)=\left(1-\frac{n}{N}\right)\frac{(\Sigma y_i^2+\bar{r}_w^2\Sigma x_i^2-2\bar{r}_w\Sigma y_i x_i)}{n(n-1)\bar{x}^2}$$

例2. 例1の分散の推定値を計算してみよう ビッテルリッヒ法の場合は、無限母集団と考えてよいので、f.p.cは無視する。

x^2	y^2	xy
900	122,500	10,500
1600	240,100	19,600
1600	160,000	16,000
400	32,400	3,600
2500	193,600	22,000
1600	144,400	14,200
2500	230,400	24000
1600	168,100	16,400
1600	160,000	16,000
2500	211,600	23,000
900	72,900	8,100
1600	136,900	14,800
900	78,400	8,400
$\Sigma y_i^2=$		1,951,300

$+\bar{r}_w^2\Sigma x_i^2=(9.82)^2\times 20200=96432.4\times 202=1947934$

$-2\bar{r}_w\Sigma x_i y_i=-19.64\times 196,600=3861224$

$\Sigma(y_i^2+\bar{r}_w^2\bar{x}_i^2-2\bar{r}_w x_i y_i)=38010$

$S_{yx}^2=\frac{38010}{12}=3167.5$

$V(\bar{y}_r)=\frac{3167.5}{13}=243.65$

$S.E(\bar{y}_r)=15.6$

$SE(\bar{r}_r)=15.6$

$SE(\bar{r}_w)=\frac{SE(\bar{y}_r)}{\bar{x}_N}=\frac{15.6}{40}=0.39$

S_{yx}^2 は自由度12だから、95%水準の t の値は 2.18 となる。 $\bar{y}_N$ の信頼区間は

$392.8\pm(2.18\times 15.6)=392.8\pm 34.0$

$=(358.8\sim 426.8)$

母集団全体では10倍して (3588～4268) となる。

今、比推定によらないで、y だけの値で推定したときの値 $3927m^3$ の分散を求めてみよう。

$\Sigma(y-\bar{y})^2=\Sigma y^2-\frac{(\Sigma y)^2}{n}=1951300-\frac{(4910)^2}{13}$

$=1951300-\frac{24108100}{13}=1951300-1854469$

$=96831$

$S_{\bar{y}_n}^2=\frac{96831}{12}=8069$

$$v(\bar{y}_R) = \frac{8069}{13} = 620.69$$

$$SE(\bar{y}_R) = 24.9$$

比推定の単純推定に精度を見ると.

$$\frac{620.69}{243.65} = 2.54$$

となり、比推定を用いたために精度が2倍半向上したことがわかる。ただし この場合，推定値を比較しているので 両法の妥当な比較には、母集団の分散から計算した理出分散を比較しなければならない．なおYates 有藤、茂井は比推定の分散式に $\frac{\bar{x}_N^2}{\bar{x}_n^2}$ の係数をかけたものを用いているが，これによって分散の推定は余り良らないので、不必要である．とくに分散を小さくする場合には用いない方がよい．

$\bar{y}_r$ の分散としては上の式により計算するのが便利だが，他の計算とも関連して，次式によることが都合のよいこともある。

$$v(\bar{y}_r) = \frac{(N-n)}{Nn}\left\{S_y^2 - 2S_yS_x\rho\frac{\bar{y}_n}{\bar{x}_n} + \frac{S_x^2\bar{y}_n^2}{\bar{x}_n^2}\right\}$$

$$= \frac{(N-n)}{Nn}\left\{S_y^2 - 2S_yS_x\rho\,\bar{r}_w + S_x^2\,\bar{r}_w^2\right\}$$

S_y, S_xは y, x の標準偏差，ρ は x と y の相関係数である。

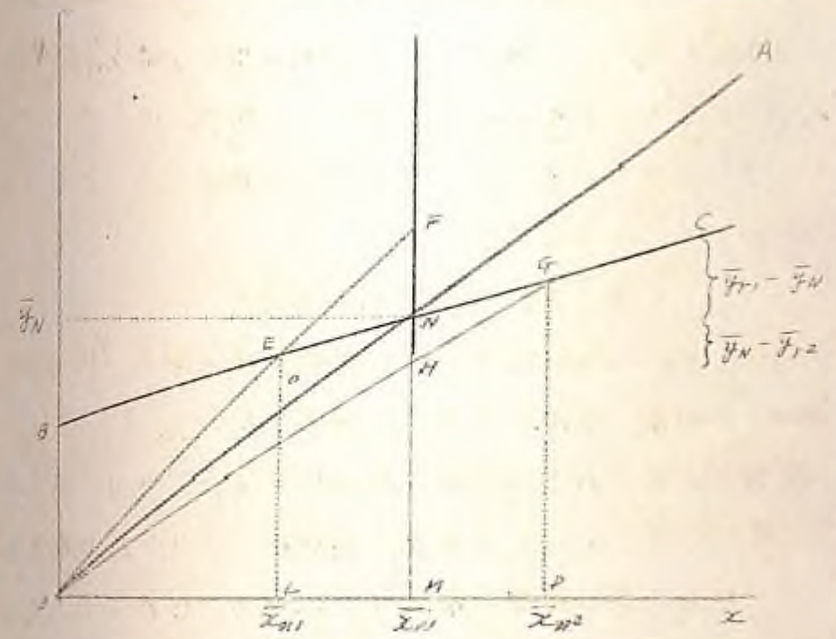
したがって

$$(n-1)\rho\,S_yS_x = \sum_{i=1}^{n}(y_i - \bar{y}_n)(x_i - \bar{x}_n) = \sum_{i=1}^{n} y_ix_i - n\bar{y}_n\bar{x}_n$$

上の例1.2 からわかるように、比推定に母集団の合計の x がわかればよいので，各単位の x を何にすべて知る必要はない．

15.4 比推定法の幾何学的な説明

比推定は一般に偏りを伴なうが，この関係を図によって見てみよう．下図で横軸に x 縦軸に y を目盛っている。



今，標本を得たとしてそれが，図のE点に対応したとする．その場合 x の平均は $\bar{x}_{n1}$ となり，比 $\bar{r}_w$ は図の直線OEの傾斜の値に一致する．$\bar{x}_N$ に相当する点Mから垂線を立てて，OEとの交点をFとすれば，MFがこの時の y の比推定値 $\bar{y}_r$ に一致する．

$$FM = \frac{OM \times EL}{OL} = \frac{\bar{x}_N \times \bar{y}_{n1}}{\bar{x}_{n1}} = \bar{x}_N \bar{r}_{n1} = \bar{y}_{r1}$$

比 $\frac{y}{x}$ がどの単位でも正確に同じであるならば母集団の比 R に一致し、各単位とも $y = Rx$ となり、どの標本も $\bar{y}_n = R\bar{x}_n$ となる。このような標本をあらわす点は、原点をとおり 傾斜が R であるような図の直線 OA 上になければならない。

OA は MF と N 点で交わるから、そのような標本では正確に母集団の $\bar{y}_N$ に一致する。この例の標本では $\bar{x}_{n1}$ に対応する点は D となることもわかるであろう。実際には、すべての点は OA 上にはないだろうが、OA からの偏差はランダムであろう。したがって、任意の単位に対しては、

$$y = Rx + \text{ランダムな偏差}$$

となり $\bar{y}_N = R\bar{x}_N$ だから、母集団全体の偏差の和は 0 となろう。その結果、OA から $\bar{y}_n$ の偏差の和と $\bar{y}_N$ からの $\bar{y}_N$ ($\bar{y}_n$ から計算されたもの) の偏差の和は何れも 0 になるであろう。

したがって x に対する y の関係が、ランダムな偏差は別として、原点を通る直線で表わすことができるならば、$\bar{y}_r$ は $\bar{y}_N$ の不偏推定子である。しかし、もちろん、これは、直線が原点を通らなければ成立たない。例えば、母集団のすべての x, y が図の直線 BC 上にあるとしよう。 BC は $y = \alpha + \beta x$ によって表わされるとする。

点 $(\bar{x}_n, \bar{y}_n)$ はこの直線上にある。今 $\bar{x}_N$ より $\bar{x}_{n1}$ と $\bar{x}_{n2}$

は $\bar{x}_N$ の反対側に等距離にあるとする。$(\bar{x}_{n1}, \bar{y}_{n1})$、$(\bar{x}_{n2}, \bar{y}_{n2})$ に対応する点を E、G とする。E 点による比推定値は、前述のように OE と MF の交点 F の値ですなわち、MF の長さに等しいが、G に対するものは、OG と MF の交点 H で、その値は、MH になる。したがって、$\bar{y}_{r1}$ と $\bar{y}_{r2}$ の $\bar{y}_N$ との差は、それぞれ、NF、HN となる。図から明らかなように、NF は HN より大きい。すなわち $(\bar{y}_{r1} - \bar{y}_N) > (\bar{y}_N - \bar{y}_{r2})$ である。これは α が正ならば常に成立つ関係であり、また α が正ということは、標本点が原点より上を通る直線上にあることを意味するので、当然一般的に言えよう。

$\bar{x}_N$ に対して、$\bar{x}_n$ が負の偏差をもつために生じた $\bar{y}_r$ の偏差は、それと同一ではあるが正の偏差をもつ $\bar{x}_n$ により計算された $\bar{y}_r$ の偏差よりは大きく、しかも正負相互に反対である。

無作為抽出では、$\bar{x}_n$ の偏差はすべての可能な標本については、その和が 0 になるが、$\bar{y}_r$ の偏差は平均しても 0 にならず、正に方に偏することがわかる。

もし、直線 BC が原点より下を通れば、上の議論に丁度反対になろう。

なお、現実には点は直線上になく、$y = \alpha + \beta x$ + ランダム偏差となっておれば、ランダムな偏差は $(\bar{y}_r - \bar{y}_N)$ の偏差を増す傾向をあるかも知れないが、偏りには影響しないだろう。

x, と y の関係は $y = \alpha + \beta x$ + ランダムな偏差で大抵の母集団

ではあらわされるが、この関係は必ずしも一次でなくともよい。その場合は、点は直線でなく、曲線からランダムに散らばっているように見えるだろう。このような場合は、偏りの性質は複雑になってくるが、偏りの大部分は、曲線に近似させた直線について述べることができよう。この点については、回帰推定にゆずる。

15.5 比推定における偏り

比推定における偏りを述べる前に回帰について若干ふれておく。

xに対するyの関係は一般に

$y = f(x)$ + ランダムな偏差

で表わされる。$f(x)$は、xの数学的な関数をあらわすが、$f(x)$は一次式でも曲線式でも常数でもよい。$y = \alpha + \beta x$は、この特別な例で、$y = \alpha$のときはαは普通の平均$\bar{y}_N$になる。上の式が母集団において成立つとき、$y_x = f(x)$はxに対するyの回帰式と呼ばれている。この式は、xが与えられたときy_xのまわりに現実のyがランダムに分布しているという前提で、yを推定するのに用いられる。$f(x) = \alpha + \beta x$のときは線型回帰と呼ばれ、それ以外の場合は、曲線回帰と呼ばれる。

真の回帰が一次であろうがなかろうが、どの母集団でも線型回帰$y_{Lx} = A + Bx$のA, Bは次のように定義される。

$$B = \frac{\sum_{i=1}^{N}(x_i - \bar{x}_N)(y_i - \bar{y}_N)}{\sum_{i=1}^{N}(x_i - \bar{x}_N)^2}$$

$$A = \bar{y}_N - B\bar{x}_N$$

もし真の回帰が線型ならばA, Bはαとβに等しい。曲線回帰の場合は、上の式は直線による近似では最良のものを与えるが、推定値y_{Lx}とyの真の値との偏差は全くランダムなものとはならず、系統的な誤差$(y_x - y_{Lx})$を含み、これはxのある範囲では正、他の範囲では負となる。

回帰はyがxに統計的に変化する場合だが、これと似た概念に相関がある。相関はxとyとが相互に関連する場合であり、その計算上の扱いは殆んど変らない。その関係の程度を示すものとして、相関係数(ρ)があるが、ρは次のように定義されている。

$$\rho = \frac{\sum_{i=1}^{N}(x - \bar{x}_N)(y - \bar{y}_N)}{N\sigma_x\sigma_y} = \frac{B\sigma_x}{\sigma_y}$$

$\frac{\sum_{i=1}^{N}(x - \bar{x}_N)(y - \bar{y}_N)}{N} = \sigma_{xy}$ とすれば

$$\rho = \frac{\sigma_{xy}}{\sigma_x\sigma_y} \qquad B = \frac{\sigma_{xy}}{\sigma_x^2}$$

ρ は＋1から－1の値をとり、$\rho=\pm1$ のときは、(x,y) 点は全部直線上に来る。点が直線をはなれて、ばらっくに従い、ρ の絶対値は小となる。ρ の正負の符号は、直線の傾斜が正か負かによって定まる。もし点が完全にランダムに分布し、y が x に無関係のときや、y の傾向が x の傾向と一致しないときは $\rho=0$ となるわけだし、$\rho=0$ でも y と x は独立であるということは断言できない。$\rho=0$ のようなときは、一次回帰は推定の目的には役に立たないだろう。

比推定の期待値は正確には決定できないが、近似的には、

$$E(\bar{y}_r)=\bar{y}_N+\frac{(N-n)}{n\bar{x}_N}(R\sigma_x^2-\rho\sigma_y\sigma_x)$$

$$=\bar{y}_N+\left(\frac{N-n}{N-1}\right)\frac{1}{n\bar{x}_N}(R\sigma_x^2-\rho\sigma_y\sigma_x)$$

または、

$$E(\bar{y}_r)=\bar{y}_N+\frac{A}{n}\left(\frac{\sigma_x}{\bar{x}_N}\right)^2\left(\frac{N-n}{N-1}\right)$$

$$=\bar{y}_N+\frac{A}{n}\left(\frac{S_x}{\bar{x}_N}\right)^2\left(\frac{N-n}{N}\right)$$

A は回帰直線が y 軸を切る点の y の値である。この式は、x に対する y の回帰が直線であってもなくても、成立つ。この式の第2項は比推定値の偏りを与える。この式かられが大きくなり、N に一致すると、偏りはなくなることがわかる。また、偏りの符号は A の符号と一致する。したがって、前述のように、回帰

線は原点より上を通れば、偏りは正で、下になれば負になる。$A=0$ のときは、直線は原点をとり、偏りは近似的に0になる。もし真の回帰が線型で原点を通れば、比推定は不偏となり、$A=0$ だが、真の回帰は曲線型のときは上の偏りで無視したものがのでなくなるが、実際的にはあまり重要でない。一般に A は小さいときは、偏りは小さくなる。また式から、x の変動係数 $\frac{\sigma_x}{\bar{x}_N}$ が大きいと、偏りは大きくなることがわかる。かつ、標本の大きされが大きいと、偏りが小さくなることもわかるだろう。$\bar{y}_r$ の標準誤差が $\frac{1}{\sqrt{n}}$ に比例し、偏りは $\frac{1}{n}$ に比例するので、れが大きくなると標準誤差より減少の程度が早いので れがかなり大きいときは、偏りの大きさは、抽出誤差に比して無視してもよくなるだろう。

すなわち

$$\frac{\text{偏り}}{\text{抽出誤差}}=\left\{\sqrt{\frac{N-n}{N}}\quad\frac{S_x}{\sqrt{n}\,\bar{x}}\right\}\left\{\frac{(RS_x-\rho S_y)}{\sqrt{(S_y^2+R^2S_x^2-2R\rho S_xS_y}}\right\}$$

である。

x との括弧は自乗して見るとわかるとおり、高々1にしかならない。y の括弧内は x の変動係数に $\sqrt{\frac{1}{n}-\frac{1}{N}}=\sqrt{1-\frac{n}{N}}\sqrt{\frac{1}{n}}$ をかけたものである。それゆえ、$\frac{|\text{偏り}|}{\text{抽出誤差}}\leq x$ の変動係数となる。

母集団が x の大きさにより増加した場合 x の層内の変動係数

C.Vは母集団のC.Vより平均的に小さくなるだろうが，層の標本の大きさは小さくなり，これらが偏りに対し，相反する効果を与えるので 層別により，偏りは大きくなる場合もでてこよう．

例3. 例1と例2を母集団の全体を示す値として再び考察してみよう．

$\bar{x}_n = 500/13 = 38.432$　　$\bar{y}_n = 4910/13 = 377.692$

$\Sigma(x_i-\bar{x})^2 = \Sigma x_i^2 - \frac{(\Sigma x_i)^2}{n}$　　$\Sigma(y_i-\bar{y})^2 = \Sigma y_i^2 - \frac{(\Sigma y)^2}{n}$

$= 20200 - \frac{250000}{13}$　　$= 1951300 - \frac{24108100}{13}$

$= 20200 - 19231$　　$= 1951300 - 1854476$

$= 914$　　$= 96.324$

$S_x^2 = \frac{914}{12} = 76.16$　　$S_x = 8.73$

$S_y^2 = \frac{96.324}{12} = 2027$　　$S_y = 295$

$\Sigma(x_i-\bar{x})\Sigma(y_i-\bar{y}) = \Sigma x_i y_i - \frac{\Sigma x_i \Sigma y_i}{n}$

$= 196600 - \frac{500\times4910}{13}$

$= 196.600 - 188.847$

$= 7753$

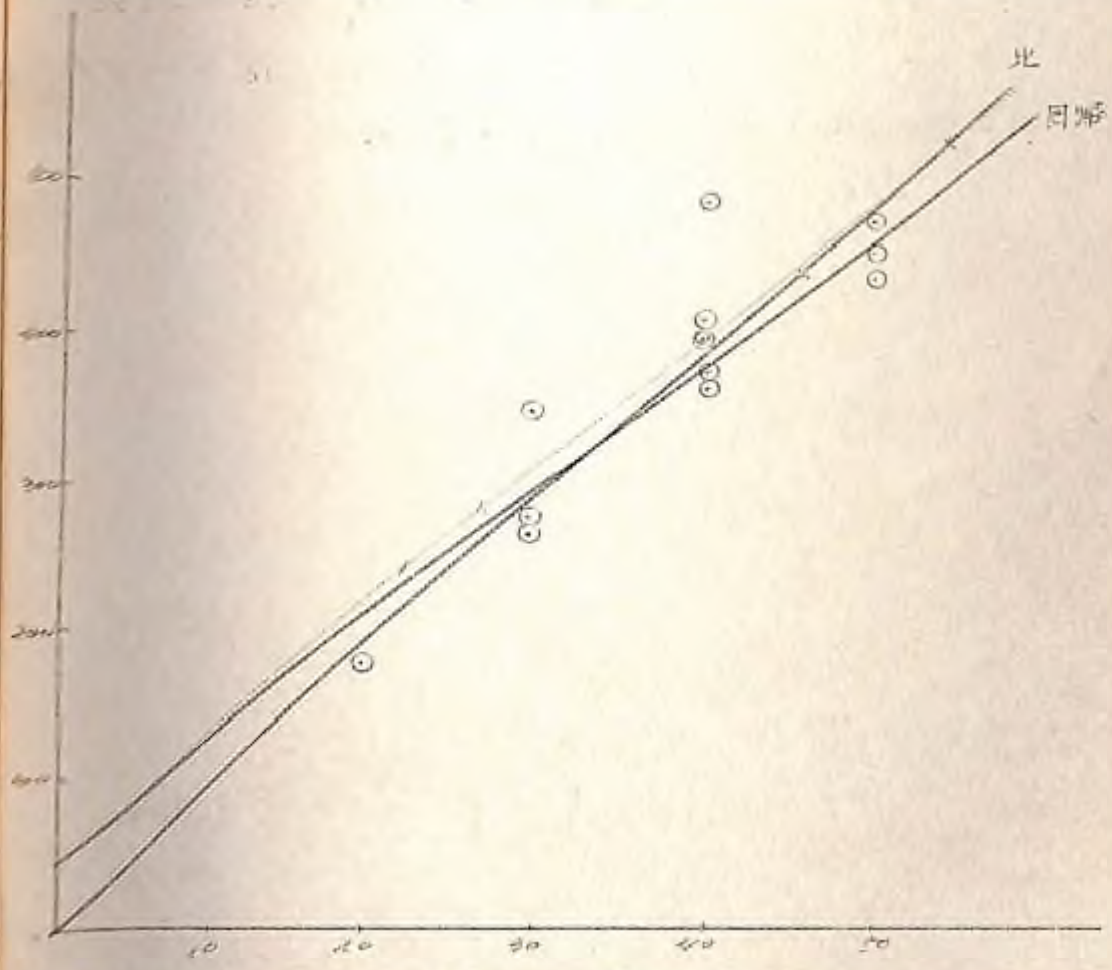
$S_{xy} = \frac{\Sigma(x_i-\bar{x})\Sigma(y_i-\bar{y})}{n-1} = \frac{7753}{12} = 646.1$

$B = \frac{7753}{914} = 8.4836$

$A = 377.7 - (8.4836)\times(38.432) = 377.7 - 326$

$= 456$

この直線式は $\hat{y}_{lx} = 456 + 8.4836x$ となる．



図からわかるように、回帰線は原点の近くを通り　真の回帰線も直線のように見える。したがって比推定の偏りも小さいように思われる。この偏りは、f, P, C を無視すれば

$$\frac{456}{n}\left(\frac{76.17}{32.43^2}\right) = \frac{1}{n} \times \frac{3473.352}{1476.8409} = \frac{2352}{n}$$

で、材積が200～500 m^3 もあるのに比べて見ても小さいことがわかる。ただし、この場合母集団の S_x, その他の値は標本の値と同じものと仮定している。現実には、S_x その他の値は母集団では異なるからこのとおりではない。

15.6 比推定の精度

比推定は単純推定に比べ、どのような場合に精度が高いだろうか。単純推定では、普通の平均値の抽出分散は

$$V(\bar{y}_n) = \frac{S_y^2}{n} \cdot \frac{(N-n)}{N} \qquad \text{となり}$$

比推定では、

$$V(\bar{y}_r) = \frac{S_{y\cdot x}^2}{n} \cdot \frac{(N-n)}{N} \qquad \text{となる。}$$

ただし $S_{y\cdot x}^2 = S_y^2 - 2RPS_xS_y + R^2S_x^2$

その精度を比較すれば、

$$\frac{V(\bar{y}_r)}{V(\bar{y}_n)} = 1 - \frac{2RPS_x}{S_y} + \frac{R^2S_x^2}{S_y^2}$$

$$= 1 - \frac{2RS_x^2}{S_y^2}\left(\frac{PS_y}{S_x} - \frac{R}{2}\right)$$

森林調査では、R はほとんど正である。（ただし、成長量の場合はそうでないこともあるかも知れない）。R が正とすれば、この比較は括弧内が正か負かで決定する。括弧内が正であれば、$\bar{y}_r$ の抽出分散は、

$$\frac{PS_y}{S_x} > \frac{R}{2} \qquad \text{ならば}$$

$\bar{y}_n$ のそれより小となる。すなわち、$P > \frac{1}{2}\frac{C.V(x)}{C.V(y)}$ か、$B > \frac{1}{2}R$ のときは、比推定が有利である。

上例では、頁に示したように、$\frac{1}{2.54}$ となり、また $B=0.48$

で $\frac{892}{2} = 446$ に比べ，はるかに大きく，βも

$$\frac{646.1}{8.73 \times 89.5} = \frac{646.1}{771.3} = 0.838$$ で正である。

したがって，このような母集団では，単純推定より，比推定の方が有利である。

なお比推定は，今まで等確率抽出として述べて来たが，気がついたことと思うが，既述のXに比例した確率で抽出する方法も，一種の比推定法を用いている。この2法について今比較して見よう。等確率法は一般に，とくにXが大きい場合にYが変動があるときは，確率比例法に比べ，分散が小さい場合が多い。しかし，確率比例抽出は不偏であるが，これは偏りをもっている。したがって，偏りが小さいと思われるときや，Xが大きくなるにつれYに変動あるようなときは，普通の比推定法がよい。なお確率比例抽出法で非復元抽出を行なうときは，偏りがはいるが，この偏りも比推定ほどは大きくなく，かつ分散は復元抽出の場合よりも小さくなる場合があると言われている。

比推定では，さらに，比RとXが相関のない場合下ではRの不偏推定になることを付け加えておく。

15.7 層別標本における比推定法

比推定法は層別抽出法で層の平均や母集団の総計の推定にも利用できることは当然で平均値は

$$\bar{y}_{st} = \sum_{h=1}^{L} \left(\frac{N_h}{N}\right) \bar{y}_{rh}$$ として推定すればよく，分散も，

$$V(\bar{y}_{st}) = \sum_{h=1}^{L} \left(\frac{N_h}{N}\right)^2 \frac{S^2_{y \cdot x_h}}{n_h} (1 - f_h)$$

で推定すればよいことは 層化抽出法の理論から当然考えられるとおりである。

標本の割当においても Neyman 割当を行なうときは S_{yh} の代りに $S_{y \cdot xh}$ を各層に用いるだけであとは同様である。ただし，このとき確率比例抽出を用いるとすれば

$$n_h = n \frac{N_h \bar{X}_{Nh} \sigma_{rh}}{\sum_{h=1}^{L} N_h \bar{X}_{Nh} \sigma_{rh}}$$ を用いる。

ただし，$\sigma^2_{rh} = \sum_{i=1}^{N_h} \left(\frac{X}{\sum X}\right)(r_h - R_h)^2 = \frac{\sum x r^2 h}{\sum x} - R_h^2$

で定義している。 （前章参照）

上記の層別標本の比推定の分散推定式は，もちろん近似式で，不偏でないので，各層ごとの標本の大きさが小さいと妥当なものでなくなってくる。n_h が小なときの信頼すべき公式はない。

また，n_hが小さく，層の数Kが大きいときは，Yの比推定

値 $\bar{y}_R$ の偏りは標準偏差に比べて大きくなり，無視できなくなるであろう。すなわち，一つの厅では，

$$\frac{|\bar{y}_{Rh}\text{の偏り}|}{\sigma(\bar{y}_{Rh})} \leqq \bar{x}_h \text{ の変動係数}$$

だから，偶々，偏りがどの厅でも同一符号をもてば，$\bar{y}_R$ の偏りは各 $\bar{y}_{Rh}$ の場合のK倍になろう。一方分母のσの方は，$\sqrt{K}$倍になるから，

$$\frac{|\bar{y}_R\text{の偏り}|}{\sigma(\bar{y}_R)} \text{ は } \bar{x}\text{の変動係数の}\sqrt{K}\text{倍になろう。}$$

例えば，厅の数が3hあり，各厅の$\bar{x}_h$ の C.V. が α ノヒすれば，$\bar{y}_R$の偏りは，標準誤差の$\alpha\delta$倍にもなろう。一般に，偏りは標準誤差のα_2以上の場合は比推定は用いてはいけない。したがって，$\sqrt{K}\times$（$\bar{x}_h$の変動係数）次の2以上のときは比推定は望ましくないか，この値は，上の比の上限を示すから，各厅ともx,yの直線が原点近くを通るときは，差支えない。

一般に厅別と比推定の総合効果を述べることは困難である。xで厅別すれば，比推定により除かうとしているyの変動も一部は，除かれ，比推定の効果は，無厅別のときほどあがらないだろう。しかし，xが厅内でかなりまだ変動がある場合は，厅内での比推定の利用は有効であろう。ただし，比推定を行なうとき xの値で厅別することは依然疑問である。もし，xに対するyの回帰が原点を通るならば，厅別けを行なっても比推定

の精度をまさない。しかし，厅別けが，xとは全く無関係なものを基準にして行なうと，比推定は，無厅化母集団の場合と同じ位有効になる。

確率比例抽出では，偏りはないが，厅ごとにその大きさ，xの平均や比の変動が著しく異なるときは，この方法は単純抽出による普通の比推定よりは能率がずっと低いであろう。これは前掲の分数の推定式で，N_h，$\bar{x}N_h$，σ_{Fh} の大きな厅では，$(1-f_h)$が大きくひびき，また$\bar{x}N_h$，σ_{Fh}も関係して，差がでてくるものと思われる。

15.8 複合比推定

厅別標本では，各厅ごとに比推定値を計算し，それらをあわせて母集団全体の推定値を求めた。しかし、標本の大きさが小さく，xとRが相関ある場合は，各厅の推定値の偏りが大きくなり，妥当な推定を与えないだろう。このような場合で比Rが各厅で一定であると仮定できるような場合は，Rの込みにした推定値を用いることがある。このような推定法を複合比推定法という。

標本から

$\bar{y}_{st} = \sum_h N_h \bar{y}_h$ ， $\bar{x}_{st} = \sum_h N_h \bar{x}_h$

を計算しYの複合比推定値 $\bar{y}_{RC}$ は，次により計算する。

$$\hat{R}_{RC} = \frac{\hat{R}_{st}}{\hat{X}_{st}} \times = \frac{\bar{y}_{st}}{\bar{x}_{st}} \times$$

$$\bar{y}_{RC} = \frac{\hat{R}_{st}}{\hat{X}_{st}} \bar{X}_N = \frac{\bar{y}_{st}}{\bar{x}_{st}} \bar{X}_N$$

この推定にも，全体のxの値は必要なく，その合計，平均値がわかればよい。

$\hat{R}_{RC}$ の分散は，標本の大きさ n が大きければ

$$V(\hat{R}_{RC}) = \sum_h \frac{N_h(N_h - n_h)}{n_h} \left[S^2_{yh} + R^2 S^2_{xh} - 2R S_h S_{yh} \right]$$

で近似される。前節の場合は分離比推定というが，両者の分散式において異なるのは，Rだけであることがわかるであろう。

なお，この推定式は，それぞれの R, S^2, ρ に対応する推定値を代入すればよい。

両推定値の比較

$$V(\hat{R}_{RC}) - V(\hat{R}_{RS})$$

$$= \sum_h \frac{N_h(N_h - n_h)}{n_h} \left\{ (R^2 - R^2_h) S^2_{xh} - 2(R - R_h) \rho_h S_{yh} S_{xh} \right\}$$

$$= \sum_h \frac{N_h(N_h - n_h)}{n_h} \left[(R - R_h)^2 + 2(R_h - R)(\rho_h S_{yh} S_{xh} - R_h S^2_{xh}) \right]$$

$(\rho_h S_{yh} S_{xh} - R_h S^2_{xh}) = S^2_{xh}(B_h - R_h)$ は各戸で回帰線が原点を通るときは0になるし，比推定が妥当な場合は一般に小さい。

15.9 二重抽出法

今迄の説明ではxの母集団総計や平均は既知であることを仮定して来た。しかし，一般には知られないことが多い。もし，xの測定が安価に行なわれ，かなり大きな標本がとることができるとした場合は，それから $\bar{X}_N$ をかなり精度高く推定できる。その中の副次標本をとり，yを測定してRを推定できる。例えば，xを胸高直径，yを樹高または材積を想定したならば，納得できるであろう。すなわち，小班材積を推定するために，全林の中の10%の立木の胸高直径をはかり，その内からさらに10%の木についてはさらに樹高を測り，材積を計算する。このような方法を二重（相）抽出法という。

第一相標本が n_1 個の単位で，その平均は $\bar{x}_1$，第二相標本は n_2 個の単位からなり，その時のx，yの平均は $\bar{x}_2$，$\bar{y}_2$ とする。両段階の標本抽出は単純無為とすると，Rは次式から推定できる。

$$\hat{r}_w = \frac{\bar{y}_2}{\bar{x}_2}$$

yの母平均は

$$\bar{y}_{1r} = \hat{r}_w \bar{x}_1$$

より推定する。$\bar{y}_{1r}$ は，実は第一相標本の n_1 単位のyの平均 $\bar{y}_1$ の比推定値である。しかし，第一相標本は無作為だから，もし $\bar{y}_1$ がわかっているならば $\bar{y}_1$ を用いて母平均 $\bar{y}_N$ を推定

するのが当然であろう。したがって，この $\bar{y}_1$ の推定値である $\bar{y}_{1r}$ を用いて，$\bar{y}_N$ を推定するのである。この点，$\bar{y}_{1r}$ は推定値の推定値ということができる。分散の式もその二段推定の性格を備えている。すなわち，

$$V(\bar{y}_{1r}) = \frac{S^2_{y\cdot x}}{n_2}\left(1-\frac{n_2}{n_1}\right) + \frac{S^2_y}{n_1}\left(1-\frac{n_1}{N}\right)$$

S^2_y は従来通りの y の母分散（σ^2_y）の $\frac{N}{N-1}$ 倍で，$S^2_{y\cdot x}$ は，$\frac{\sum_{i=1}^{N}(y-Rx)^2}{N-1}$ である。上の式の第1項は $\bar{y}_1$ の推定値と考えられた $\bar{y}_{1r}$ の分散で，第2項は，$\bar{y}_N$ の推定値と考えられた $\bar{y}_1$ の分散である。実際は S^2_x，S^2_y ともに n_2 の大きさの副次標本から推定され，S^2_y は $\frac{S^2_y}{n}$，$S^2_{y\cdot x}$ は

$$S^2_{y\cdot x} = \frac{\sum_{i=1}^{n_2}(y-\bar{r}_w x)^2}{n_2-1}$$ によって推定される。

この式の第1項には $\left(\frac{\bar{x}_1}{\bar{x}^2}\right)^2$ をつける場合がある。

（このことについては比推定の場合を参照）

二重抽出の目的は，費用のかかる y の標本を小さくして，その代り費用のかからない x の標本を大きくし，x，y の比を利用して，n_2 個の単位のみより得られる情報よりもより広い情報を得て推定の精度をあげようとするものである。

したがって二重抽出の効果は，x，y の調査に要する相対的費用の大小，x の変動により説明できる y の変動の予察により

きまってくる。

もし費用が，x，y ともに調査するのに，x だけ調査するよりも k 倍かかるならば，$\frac{n_2}{n_1}$ の最適値（P）は，

$$P = \sqrt{\frac{S^2_y \cdot x}{(k-1)(S^2_y - S^2_{y\cdot x})}}$$

により定まる。

もし，n_2 の最適値が n_1 に等しいならば，第一相標本の全単位の y を調査しなければならないことになり，$\bar{y}_1$ の真の値を計算することになる。そのときは x の値は何ら必要でなくなるので，xについては調査せず，できるだけ多くの単位について y を調査すればよい。

それ故，上式の P の値より小さいときだけ二重抽出法が効果があるわけである。

$$\frac{S^2_y x}{(k-1)(S^2_y - S^2_{y\cdot x})} < 1$$

$$\frac{S^2_y - S^2_{y\cdot x}}{S^2_{y\cdot x}} > \frac{1}{k-1}$$

$$\frac{S^2_y}{S^2_{y\cdot x}} > \frac{1}{k-1} + 1 = \frac{k}{k-1}$$

$$\therefore \frac{S^2_{y\cdot x}}{S^2_y} < \frac{k-1}{k}$$

k が大きいと，$\frac{k-1}{k}$ はほとんど1に近くなるから，$S^2_{y\cdot x}$ が S^2_y より僅か小でも，二重抽出は有利になろう。反対に k が

小さいと $S^2_{y\cdot x}$ は S^2_y よりかなり小さくないと，二重抽出は有効にはならない。

15.10 回帰推定

y の x に対する真の回帰が $y = \alpha + \beta x + \varepsilon$ （ε はランダムな誤差，$\neq 0$）のときは，比推定は偏りをもつから，標本から，R でなく

$$y_{Lx} = A + Bx$$

を推定して，$x = \bar{x}_N$ のときの $\bar{y}_N$ を y_{Lx} により推定したい。母数 A，B の推定値を a，b とするときは，

$$a = \bar{y}_n - b\bar{x}_n$$

$$b = \frac{\sum_{i=1}^{n}(x_i - \bar{x}_n)(y_i - \bar{y}_n)}{\sum_{i=1}^{n}(x - \bar{x}_n)^2}$$

直線回帰推定値 $\bar{y}_{lg}$ は次のようになる。

$$\bar{y}_{lg} = \bar{y}_n - b\bar{x}_n + b\bar{x}_N$$
$$= \bar{y}_n + b(\bar{x}_N - \bar{x}_n)$$

比推定の場合の断面積に対する材積の回帰では

$$\bar{y}_{lg} = 3クメク + \frac{7753}{914}(\bar{x}_N - 38.43)$$
$$= 3クメク + 8.48(\bar{x}_N - 38.43) = 45.6 + 8.48\bar{x}_N$$

したがって，$\bar{x}_N = 40$ に対しては $\bar{y}_{lg} = 385\ m^3$ となる。（比推定の場合は，378 m^3 であった。）

回帰推定は，完全に不偏というわけには行かない。しかし，真の回帰が直線ならば，偏りは微々たるもので，真の回帰直線が原点または原点の近くを通るときの以外は，比推定の偏りよりははるかに小さい。真の回帰が曲線の場合は，回帰推定は偏りがあるが，比推定によるよりは小さい。すなわち $\frac{1}{n}$ のオーダーまで書けば，$\bar{y}_{lg}$ の偏りは近似的に

$$\frac{-(N-n)}{(n-1)NS^2_x}\left\{\frac{\sum \varepsilon_i (x_i - \bar{x})^2}{(N-1)}\right\}$$

ε_i は前に記したものである。括弧内は ε_i と $(x_i - \bar{x})^2$ の共分散である。これは x に対する y の二次の回帰に起因するもので，y と x の関係が一次ならば 0 となる。$\bar{y}_{lg}$ の偏りは $\frac{1}{n}$ のオーダーで，標準偏差は $\frac{1}{\sqrt{n}}$ のオーダーだから，大標本では，この偏りは無視できるほど小さくなる。

回帰推定値の分散は近似的に次式により与えられる。

$$V(\bar{y}_{lg}) = \frac{S^2_e}{n}\left(1 + \frac{1}{n}\right)(1-f)$$

S^2_e は，直線回帰からの y の偏差から計算される。

$$S^2_e = \frac{\sum_{i=1}^{N}[y - (A + Bx)]^2}{N-1}$$

回帰直線からの y の偏差平方和は他のどの直線からのそれよりも小さい。もちろん $y = Rx$ よりも小である。それゆえ，回帰直線が原点を通らない限りは S^2_e は $S^2_{y\cdot x}$ より小さい。

$V(\bar{y}_r) = \dfrac{S_{y \cdot x}^2}{n}\left(1 - \dfrac{n}{N}\right)$ が比推定の分散式として有効に用いられるほど n が大きいときは，$V(\bar{y}_{rg})$ の式の $\left(1+\dfrac{1}{n}\right)$ はほとんど 1 に等しくなる。したがって，両式の差は $S_{y \cdot x}^2$ と S_e^2 だけになってくる。一般に，母集団の回帰直線が原点を通らない限りは回帰推定値の分散は，比推定値の分散より小さい。

S_e^2 は標本の $S_e^2 = \dfrac{\sum_{i=1}^{n}[y_i - (a + bx_i)]^2}{n-2}$ から推定できる。

この場合の自由度は $(n-2)$ である。（比推定の場合は $(n-1)$ であった。）

S_e^2 は S_{y}^2x よりは大きいことはないが，推定値の S_e^2 は S_y^2x より大きいことはあり得る。

$$S_{y \cdot x}^2 = \frac{\Sigma(y - \bar{r}_w x)^2}{n-1} \quad と \quad S_e^2 = \frac{\sum_{i=1}^{n}[y - (a + bx)]^2}{n-2}$$

の両式で $a = 0$ となると $b = \bar{r}_w$ となり，分子は同じになり，比 $\dfrac{S_e^2}{S_{y \cdot x}^2}$ のとり得る最大の値は $\dfrac{n-1}{n-2}$ となる。a が正確に 0 となることは，ありそうもないが，原点 0 に近い場合があろう。

そのときは，S_e^2 は S_y^2x より僅か大きくなる。

すなわち $V(\bar{y}_r) = \dfrac{(N-n)}{N-n}\left(S_y^2 + R^2S_x^2 - 2R\rho S_yS_x\right)$

$V(\bar{y}_{rg}) = S_y^2(1-\rho^2)$ となるから，$-\rho^2S_y^2 < R^2S_x^2 - 2R\rho S_yS_x$ の時は比推定より回帰推定の方が有効である。これから

$(\rho S_y - RS_x)^2$ で常に正だから $\rho S_y = RS_x$ すなわち

$\rho = \dfrac{RS_x}{S_y} = \dfrac{x の CV}{y の CV}$ 以外は回帰推定の方が有効であることがわかる。ただし回帰推定の方は比推定より計算が厄介なのは欠点である。

S_e^2 の計算には，次式によるとよい。

$$S_e^2 = \frac{1}{n-2}\left\{\sum_{i=1}^{n}(y - \bar{y}_n)^2 - \frac{\left[\sum_{i=1}^{n}(x - \bar{x}_n)(y - \bar{y}_n)\right]^2}{\Sigma^n(x - \bar{x})^2}\right\}$$

$$= \frac{\Sigma(y - \bar{y}_n)^2}{n-2}(1 - r^2) \qquad r は標本相関係数$$

前の例1〜例3の問題において，

$\bar{y}_{rg} = 3717 + 8.4836(\bar{X}_N - 38.432)$ 　$\bar{X}_N = 40$ に対し $\hat{y}_N = 385\ m^3$ だったが その分散の推定値は

$$S_e^2 = \frac{1}{13-2}\{96.324 - 8.4836 \times 7.753\}$$

$$= \frac{1}{11}\{96.324 - 65.424\} = \frac{1}{11} \times 30.900$$

$$= 2,809$$

$$V(\bar{y}_{rg}) = \frac{28.09}{13}\left(1 + \frac{1}{13}\right)$$

$$= \frac{28.09 \times 14}{13^2} = \frac{393.26}{169} = 2.327$$

$SE(\bar{y}_{rg}) = 1.53$ 　（比推定の場合は $V(\bar{y}_{rg}) = 2.437$

$SE(\bar{y}_{lrg}) = 1.6$)

この場合 $\bar{X}_N$ がわからず、推定しなければならないときは、回帰による2相推定法が行なわなければならない。そのときは、分散の推定式は、次のようになる。

$$v(\bar{y}_{lrg}) = \frac{S_e^2}{n_2}\left(1+\frac{1}{n_2}\right)\left(1-\frac{n_2}{n_1}\right) + \frac{S_y^2}{n_1}\left(1-\frac{n_1}{N}\right)$$

ただし、$S_e^2 = \frac{\sum[y-(a+bx)]^2}{n-2} = \frac{\{\sum(y_i-\bar{y})^2(1-r^2)\}}{n-2}$

$$S_y^2 = \frac{\sum(y_i-\bar{y})^2}{n-1}$$

なお、Tekkiwal によれば S_e^2 の分母の $(n-2)$ は $(n-3)$ が正しい。上式は、第一相の大標本の中から第2相の小標本を抽出して、しかも x が正規分布していることを条件にしている。

$\frac{1}{n_1}$ が小さいときは上式は近似的に、

$$\frac{S_e^2}{n^2}\left(1-\frac{n_2}{n_1}\right) + \frac{S_y^2}{n}$$

$$= \frac{S_y^2(1-r^2)}{n^2} - \frac{S_y^2(1-r^2)}{n_1} + \frac{S_y^2}{n_1}$$

$$= \frac{S_y^2(1-r^2)}{n_2} + \frac{r^2 S_y^2}{n_1}$$

また一般に有限修正 $\left(1-\frac{n}{N}\right)$ は二重抽出では省かれる。

標本の割当は、費用を $C = C_1 n_1 + C_2 n_2$ とすれば、

$$n_{1opt} = C \Big/ \left[C_1 + C_2\sqrt{\frac{1-r^2}{r^2}\cdot\frac{C_1}{C_2}}\right]$$

$$n_{2opt} = n_{1opt} \times \sqrt{\frac{1-r^2}{r^2}\cdot\frac{C_1}{C_2}}$$

$$v(\bar{y}_{drgopt}) = \frac{S_y^2}{C}\left[r\sqrt{C_1} + \sqrt{(1-r^2)C_2}\right]^2$$

比推定を用いる二重抽出法では $C = C_1 n_1 + C_2 n_2$ ならば、

$$n_{1opt} = \frac{C}{C_1 + C_2 Z}$$

$$n_{2opt} = n_{1opt} \times Z$$

$$S^2_{\bar{y}dr(opt)} = \frac{1}{C}\left[\sqrt{C_1(2RS_{xy}-R^2S_x^2)} + \sqrt{C_2(S_y^2+R^2S_x^2-2RS_{xy})}\right]$$

ただし $Z = \sqrt{\frac{S_y^2+R^2S_x^2-2RS_{xy}}{2RS_{xy}-R^2S_x^2}\cdot\frac{C_1}{C_2}}$

$$S_{xy} = \rho S_x S_y$$

となる。

比推定と回帰推定の比較

一般に回帰の方が直線が原点を通るときを除いては精度が高い。しかし計算が厄介。y が x に比例するときは直線が原点近くを通ると思われるから比推定の方がよい。

比推定、回帰推定を用いるときは (x, y) の図を入れて見ることが重要である。

層別抽出における比推定　（分離比推定）の計算様式

比および全体の推定

層	N_i	$\sum_{j}^{N_i} x_{ij}$	n_i	$\sum_{j=1}^{n_i} x_{ij}$	$\sum_{j=1}^{n_i} y_{ij}$	$\bar{r}w_i$	$\hat{Y}_r = \bar{r}w_i \sum_{j=1}^{n_i} x_{ij}$
1	○	○	○	○	○	○	○
2	○	○	○	○	○	○	○
⋮	⋮						
計	○						○

分散の推定

層	$\sum_{j=1}^{n_i}(y_{ij} - \bar{r}w_i x_{ij})^2$	$S_{y \cdot x_i}^2$	$V_i(\hat{Y}_r)$
1	○	○	○
2	○	○	○
⋮			
計			○

層別抽出における回帰推定

回帰係数の推定　（分離）

層	$\sum_{j=1}^{n_i} x_{ij}^2$	$\sum_{j=1}^{n_i} x_{ij} y_{ij}$	$\sum_{j=1}^{n_i} y_{ij}^2$	$\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_{ni})^2$	$\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_{ni})(y_{ij} - \bar{y}_{ni})$	b_i	a_i
1							
2							
⋮							

平均および全体の推定

層	$\bar{x}_{Ni}$	$\bar{x}_{ni}$	$\bar{x}_{Ni} - \bar{x}_{ni}$	$b_i(\bar{x}_{Ni} - \bar{x}_{ni})$	$\bar{y}_{ni}$	$\bar{y}_{rgi}$	$Y_{rgi} = N_i \bar{y}_{rgi}$
1	○	○	○	○	○	○	○
2	○	○	○	○	○	○	○
⋮							
計							○

分散の推定

層	$\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_{ni})^2$	$\sum_{j=1}^{n_i}[y_{ij} - (a_i + b_i x_{ij})]^2$	自由度 $n_i - 2$	$S_{e \cdot i}^2$	$V_i(\bar{y}_{rgi})$
1	○	○	○	○	○
2	○	○	○	○	○
⋮					
計					○

いづれの方法も各層の全体の値を推定して加えあげると，母集団全体の値となる。

[付] 複合回帰推定

複合回帰推定については、比推定と同様なことが言える。次にその場合の推定式のみをあげておく。

$$\bar{y}_{regc} = \bar{y}_{st} + b(\bar{X}_N - \bar{x}_{st})$$

ただし、$\bar{y}_{st} = \frac{\sum N_h \bar{y}_h}{N}$, $\bar{x}_{st} = \frac{\sum N_h \bar{x}_h}{N}$

$$b = \frac{\sum_{h=1}^{L} \sum_{i=1}^{n_h} (y_{hi} - \bar{y}_h)(x_{hi} - \bar{x}_h)}{\sum_{h=1}^{L} \sum_{i=1}^{n_h} (x_{hi} - \bar{x}_h)^2}$$

分散の推定値は

$$s^2_{y \cdot x_h} = \frac{1}{(n_h - 1)} \sum_i \{(y_{hi} - \bar{y}_h) - b(x_{hi} - \bar{x}_h)\}^2$$

$$= \frac{1}{n_h - 1} \left[\sum_i (y_{hi} - \bar{y}_h)^2 - 2b \sum_i (y_{hi} - \bar{y}_h)(x_{hi} - \bar{x}_h) \right.$$

$$\left. + b^2 \sum_i (x_{hi} - \bar{x}_h)^2 \right]$$

16 余論

16.1 小標本の場合の母集団比率の信頼区間

母集団の比が0や1に余りに近くない場合や抽出した個数nが余り小さくない場合は、正規近似の理論で信頼区間を設定してもよいが、この2つの条件を満足しないときで、無限母集団の場合は二項分布の理論によって構成しなければならない。

今、大きな母集団（従って、抽出比fは0とみなしてもよい）から、抽出された10単位からなる一標本を考えよう。目的の属性をそなえたものが1単位しかないとしよう。

それゆえ　　$p = 0.1$

$$v(p) = \frac{0.1 \times 0.9}{9}$$

$$= 0.01$$

$$s_p = 0.10$$

自由度9で、5%水準（95%信頼限界）表の$A = 2.26$だから上のpの95%信頼区間は$(-0.13 \sim 0.33)$となる。この限界は明らかにおかしい。pは決して負になり得ない。したがって下界を0におきかえればよいが、負の下限が生じたことから当然、上限の正確性についても疑問が生じる。もし、一方が明らかに低すぎれば他方も低すぎるのではないかという疑問が生ずる。

正確な限界はrの真の確率分布を用いれば計算できる。Nが非常に大きいとして、目的の属性を有するr個の個体を含む標本の大きさをnとすれば、

$$P_2(r) = \frac{n!}{r!(n-r)!} P^r Q^{n-r}$$

（$0! = 1$ と定義する）

となる．$n = 10$，$r = 1$ とすれば $r = 0$ か 1 しか起らないときの確率は

$$P_r(0) = \frac{10!}{0!\,10!} \times P^0 Q^{10} = Q^{10} = (1-p)^{10}$$

$$P_r(1) = \frac{10!}{1!\,9!} PQ^9 = 10PQ^9 = 10p(1-p)^9$$

p の95%限界の下の方の値を p_L とすれば r の値を得る確率を0.025以上にするものである．それゆえ $r = 1$ 以上のものを得る確率は1から0を得る確率を引いたものである．

したがって，

$$1 - (1-p_L)^{10} = 0.025$$

上限についても同様で p の95%の上限 p_u は r の値を得る確率を0.025に等しいか，より小なものにするものである．この例では p_u は次式の解である．

$$(1-p_u)^{10} + 10p_u(1-p_u)^9 = 0.025$$

この結果，p_L は0.0025 p_u は0.4455となり，上記の（−0.13, 0.33）とは大分変ってくる．

Fischer Yates の表から $n = 10$, $r = 1$ 5%限界は，0.0253と4.45として，よみとられる．

もし母集団が大きくない場合は二項分布は不適当である．しかし上表は よい近似を与えてくれる（Cochran，34）．

なお スネデッカーの訳書（4頁）の表だと p の95%信頼区間は 0.0～0.445 となっている。

16.2 極めて小さい p に対する標本単位の割当個数のきめ方

今，誤差を，95% の信頼水準で5%，25%におさえた場合の抽出個数を計算してみると，次の表のようになる。

ただし，母集団の大きさは極めて大きいとする。

p	誤差 $\pm$ 5%	誤差 $\pm$ 25%
0.5	1,600	64
0.4	2,400	96
0.3	3733	149
0.2	6400	256
0.1	14,400	576
0.05	30,400	1216

この表から見ると p が減少すると n の大きさは 急速に増加し p の予測を相当正確にしないと，ヒンでもないことになる．（ただしこの計算は前に述べた方法，誤差 $= \frac{4pq}{n}$ より算出した）

この方法に対して，非常に大きい母集団では，近似的に次式が成立つ

$$m = 1 + \left(\frac{3}{2}\right)^2$$

m は抽出した n 個の中で目的の属性を有するものの個数，d は誤差の真値に対する比。

今 $\lambda = 0.4$ とすれば

$m = 26$ となる

この式は p とは無関係に計算できるので 便利である。種々の λ に対して m の値を次に示しておく。

λ	m
0.05	1601
0.10	401
0.20	101
0.30	46
0.40	26

16.3 各種の林相の林分面積の比率を推定する方法

各種の林相の林分面積の総面積に対する比率を推定するには既述のようにその地域内にランダムに点をおとして、全体の点に対する各種林相内におちた点の比により推定すればよいが、その方法としては次の3が考えられる。

(1) 信頼すべき大縮尺の地図があれば、その中に現地調査を行なって得られた林相区分を記入した地図

(2) 修正した航空写真

(3) 現地におとした点

(1)、(2)の方法は 地図や写真上でプラニメーターをまわしてもよいが、小面積の林相区分を行なった場合は、ランダムに点をおとす方法よりは時間がかかるだろう。この時得られた p の分散は $\frac{pq}{n-1}$ で推定される。なお、無限母集団だから有限修正は不要である。

この外に格子点網で、地図をおおって、各林相におちた点を数え全林地の点数との比を求める方法であるが、偏りをさけるために、まず格子の方向を予めきめておく。例えば、北南、東北というようにする。格子の定点に対して 地図上に座標をランダムにきめて、格子網の位置にランダム性を与える。この方法は系統的抽出法にあたるので正しい誤差計算はできない もし あらい目の格子網だったら、数回ランダムにおいて、p_i を計算して その平均値 $\bar{p}$ の分散を計算すればよい。

16.4 層別を利用する二重抽出法

母集団全単位を N とし、まず単純無作為抽出により n の大きさの第一標本を抽出し、x_i について測定し、この x_i にもとづいて 層別する。各層から n_h 個の単位からなる第二標本を抽出して $\bar{y}_i$ を測定する。第二標本は第一標本から抽出してもよいし全く無関係に抽出してもよい。

今 h 層に含まれる母集団の割合 $\frac{N_h}{N} = W_h$ とし、h 層にはいる第一標本の割合 $\frac{n_h}{n} = w_h$ とすれば w_h は W_h の推定値となる。

母平均 $\bar{Y} = \sum_{h=1}^{K} W_h \bar{Y}_h$

その推定値は $\bar{y}_{st} = \sum_{h=1}^{K} w_h \bar{y}_h$

この推定式で w_h と $\bar{y}_h$ は共に確率変量と見なすことができる

この式で、$\bar{y}_{st}$ は不偏推定値である。

この分散の推定式は、

$$v(\bar{y}_{st}) = \frac{n_1}{(n_1-1)} \sum_h \left[\left\{ w_h^2 - \frac{w_h}{n_1} \right\} \frac{s_h^2}{n_h} + \frac{w_h (\bar{y}_h - \bar{y}_{st})^2}{n_1} \right]$$

である n_1 は通常 n_h に比べ大きいから

$$v(\bar{y}_{st}) \fallingdotseq \sum_h w_h^2 \frac{s_h^2}{n_h}$$

により近似してもよい。

また、h 層において ある林相やある属性をもつものの割合（標本での）P_h とすれば、母集団全体の P_{st} の推定値は $\frac{\sum w_h P_h}{\sum w_h}$ で推定され P_{st} の分散の推定値は、近似的に

$$\sum_h \left[\frac{w_h^2 P_h q_h}{n_h - 1} + \frac{w_h (P_h - P_{st})^2}{n_1} \right]$$

となる。

なお、全費用を既述のように $C = n_2 C_2 \sum n_h + n_1 C_1$

$$= C_2 n_2 + n_1 C_1$$

とすれば、最適割当は 有限修正を無視すれば

$\frac{n_2}{\sqrt{V_2 C_1}} = \frac{n_1}{\sqrt{V_1 C_2}}$ と $C = n_2 C_2 + n_1 C_1$ とにより定める。

ただし、$V_1 = \sum w_h (\bar{Y}_h - \bar{Y})^2$, $V_2 (\sum w_h S_h)^2$ で、何れも母集団の値だが、想定して代入する。

この方法は x と y との関係が、比や回帰ではっきり表わされないと思われるとき利用するとよい。次に、この方法による例を挙げる。

例 1.

アメリカの北西部地方のある郡で航空写真による森林の蓄積を推定するため、森林調査を行うことになった。そのため 写真上で 多数の標本地プロットを抽出して 材積級に応じて階層区分し、さらに、そのうちのいくつかを抽出して地上調査を行なった その過程は次の通りである。

ハ 必要調査標本地の個数の決定

上の式から、地上調査プロットの数 n_2 は

$$n_2 = \frac{C \alpha}{\alpha C_2 + \beta \sqrt{C C_2}}$$

C は全費用、C_1 は写真標本地の費用

C_2 は地上標本地の費用

$\alpha = \sum w_h S_h$, $\beta^2 = \sum w_h (y_h - \bar{y}_{st})^2$

w_h はその層の全体に対する比、S_h はその層の標準偏差

X_h はその層の平均材積、又は全体の平均材積

計算した n_2 は整数に丸める。

写真プロットの数 n_1 は 上式から

$$n_1 = \frac{C - n_2 C_2}{C_1}$$

n_1 は整数に丸めるが これらの n_1, n_2 をきめるときは、$(n_2 - 1)$, n_2, $(n_2 + 1)$ による場合の精度を計算してきめる

この郡の全土地面積は 334,600 エーカーあって、194,000 エーカーが林地だった。林地ノエーカーあたりの森林調査費用は、

0.0039ドル．したがって全予算は405.60ドルとなる．従来の経験から，この地方の森林調査費は平均して $C_2=17.27$ ドル，写真上での林地プロット調査費 $B_1=0.21$ ドル，非林地調査費 $B_2=0.02$ ドル，$\Sigma W_h S_h=394$，$\Sigma W_h(X_h-\bar{X})^2=355721$ であることがわかっている．さらに $C_1=P_f B_1+(1-P_f)B_2$ である．P_f は林地の割合で，$P_f=104000\div 336600=0.309$

$C_1=0.07871$ ドル　$\sqrt{W_h(X_h-\bar{X})^2}=597$，$(\Sigma W_h S_h)^2=155236$

$\sqrt{C_1 C_2}=1.1659$ だから

$$n_2=\frac{(405.60)(394)}{394(17.27)+597(11659)}=21.3$$

$$n_1=\frac{(405.60)(394)}{0.07871}=545.4$$

n_2 が20，22のときは n_1 は765.326となる．これらの数字を前に示した式 Neyman 割当を行なうのでそのように変形した $S^2=\frac{\Sigma W_h S_h}{n_2}=\frac{\Sigma W_h(y_h-\bar{y}_{st})^2}{n_1}$ に代入すると

n_2	S_2
20	822706
21	8045.26
22	8147.96

となる．したがって，$n_2=21$，$n_1=546$ が最良となった．写真の枚数は185あったので，これを各写真に平均に割付けると，$\frac{546}{185}=2.94$ となったので，一枚あたり3点を写真プロ

ットとして割当てることになる．写真判読の結果は，I，II，III，IV，V層にそれぞれ13，40，45，81，376点はいったので，これを層ごとに分配するため Neyman 割当を用いて次のような表を作った．

層	n_i	W_h	S_h	$\Sigma W_h S_h$	$n_2 h=\frac{W_h S_h}{\Sigma W_h S_h}$
I	13	0.024	1083	26	4
II	40	0.072	797	57	8
III	45	0.081	443	36	5
IV	81	0.146	138	20	3
V	376	0.677	6	4	1
計	555	1.000		143	21

抽出間隔をきめるため，写真プロットの数を n_2 で割り，例えばI層では $13\div 4=3\frac{1}{4}$ ゆえに抽出間隔は3にして，1から4までの乱数をえらび，そのあとの間隔で抽出した．同様にして，II.5，III.9，IV.27，V.376 ときめた．抽出誤差は上の S^2 の式から計算したが，誤差計算の資料を得るために特別な調査を行なった．誤差計算の表をまとめると次のようになる．

層	W_h	S_h	$W_h S_h$	$\bar{y}_h$	$W_h S_h$	$W_h(\bar{y}_h-\bar{y})^2$
Ⅰ	0.024	1.132	27.2	1432	34.4	40.873
Ⅱ	72	702	50.5	727	52.3	25.920
Ⅲ	81	397	32.1	331	26.8	8.874
Ⅳ	146	161	23.5	87	12.7	234
Ⅴ	677	2	1.4	1	0.7	10.748
計	1.000		1347			26.649

この結果から

$$S^2=\frac{134.7^2}{21}+\frac{26649}{255}=864+156=1020$$

$$S=31.9$$

となった。

なお写真プロットの判読は、兵のまわりノエーカーの林分状態の目測によりきめている。地上の標本地さは、大径林には6エーカー、小径林には3エーカーの面積を採用した。

総材積 $V=A\Sigma W_h \bar{y}_h$ により推定した。

例 2.

イランの森林調査

写真兵を1/24,146兵抽出して、林地、無林地面積を定め、その内から、林地について、25.4/1の兵を抽出し、0.1ha プロットについて、樹冠密度と、最も高い3本の木の高さを測定し、平均樹高を求める。さらに、地上プロット 72ヶを抽出し、通常の

森林調査を行なうことにした。得られた材積と 写真から求めた樹高と樹冠密度の回帰を利用して平均材積を求め、これに推定した林地面積の全林地面積の比をかけて、全土地面積の 1ha あたり平均材積を算出し、同時に信頼区間を計算する。

なお 土地全体の面積には誤差はないものである。

1ha あたり林積の推定式

$$\bar{V}_L=P_f(a+b_1\bar{X}_1+b_2\bar{X}_2)$$

$\bar{V}_L$ =土地全体に対する 1ha あたり平均材積

P_f =写真兵から判読した土地面積に対する林地面積の比

a =定数、最小自乗法より計算

b_1 =樹高に対する回帰定数、最小自乗計算

b_2 =樹冠密度に対する回帰係数 最小自乗計算

$\bar{X}_1$ =写真プロット上の3本の木の平均樹高

$\bar{X}_2$ =同上樹冠密度、

1ha あたり平均材積の分散の推定式、

$$S_{\bar{V}_L}^2=\frac{P_f^2 S_{x.1,2}^2}{N_f}+\frac{P_f^2(b_1^2 S_{x_1}^2+b_2^2 S_{x_2}^2+2b_1 b_2 C_{ov}(x_1x_2))}{N_P}$$

$$+\frac{\bar{V}_L^2 P_f Q}{N}+P_f^2\left\{S_{x.1,2}^2(C_{11}\bar{d}_1^2+C_{22}\bar{d}_2^2+2C_{12}\bar{d}_1\bar{d}_2)\right\}$$

$$=\frac{A}{N_f}+\frac{B}{N_P}+\frac{C}{N}+D \quad \text{とする}$$

ただし

$S_{\bar{V}_L}^2$ =土地全体の平均材積の分散、

$S_{x.1,2}$ =回帰により推定した林地における平均材積の回帰とは独

立木分散

$S_{x_1}^2$ ＝3本の木の平均樹高の分散

$S_{x_2}^2$ ＝樹冠密度の分散

$Cov x_1 x_2 = x_1$ と x_2 の共分散

$\bar{v}_f$ ＝林地の $1ha$ あたり平均材積

Q ＝土地全体に対する無林地の割合

C_{11}, C_{22}, C_{12} ＝C係数

d_1 ＝N_p プロットと N_g プロット（全部写真上で測定したもの）の樹高の諸平均値間の差

d_2 ＝同様樹冠密度間の差

N_g ＝地上プロットの個数

N_p ＝写真プロットの個数

N ＝写真の抽出点の全数

また抽出個数を定めるには次式によった。

1. 地上プロット数

$$N_g = \frac{\sqrt{A}}{S_{\bar{v}_2}^2 \sqrt{C_g}} \left(\sqrt{A C_g} + \sqrt{B C_p} + \sqrt{C C_f} \right)$$

2. 写真プロット数

$$N_p = N_g \sqrt{\frac{B C_g}{A C_p}}$$

3. 写真点数

$$N = N_g \sqrt{\frac{C C_g}{A C_f}}$$

たゞし C_g ＝地上プロットの費用

C_p ＝地上プロットの費用

C_f ＝写真点の費用

実際の計算

1. 過去の資料から定められた各因子の値。

項目	記号	予想値
1	$\bar{v}_t$	69 m³
2	$\bar{v}_f$	135 m³
3	P_f	0.60
4	Q	0.50
5	Q	1-0.76
6	d_1	6.46
7	d_2	2.84
8	$\bar{x}_1$	20 m.
9	$\bar{x}_2$	65
10	$S_{\bar{v}_t}^2$	1
11	$S_{d_1 d_2}^2$	2564
12	$S_{x_1}^2$	36
13	$S_{x_2}^2$	770
14	$Cov x_1 x_2$	2
15	C_g	3.040.000リアル
16	C_p	7.60 リアル
17	C_f	0.76 リアル

2．分散式の各項の計算．

符号	記号	値
A	$P_f S_{\bar{y}1.2}^2$	$.25(2,564)=641$
B1	$b_1^2 S_{x_1}^2$	$(646)^2 36 \;=1,502$
B2	$b_2^2 S_{x_2}^2$	$(2.84)^2 770=6.214$
B3	$2b_1 b_2 Cov(x_1 x_2)$	$2(6.46)(2.84)^2= \;73$
	$B_1+B_2+B_3$	$=7.789$
B	$P_f^2(B_1+B_2+B_3)$	$.25(7.789)= \;1.947$
C	$V_f^2 P_f Q$	$138^2(.25)=4.761$
D	$P_f^2 S_{\bar{y}1.2}^2(C_{11}d_1^2+C_{22}d_2^2+2C_{12}d_1 d_2)= \;0$	

3　地上プロット．写真プロット，写真点の個数の決定．

i) 地上プロット

$$N_g=\sqrt{\frac{641}{1.0\sqrt{3.040}}}\left[\sqrt{641(3.040)}+\sqrt{1947(7.60)}+\sqrt{4.761(0.76)}\right]$$

$$=\frac{2.53}{10.(55)}\left[1.396+122+66=726\right]$$

ii) 写真プロット

$$N_p=726\sqrt{\frac{1.947(3.040)}{641(7.60)}}=726(35)=25.410$$

iii) 写真点

$$N=726\sqrt{\frac{4.761(3.040)}{641(0.76)}}=726(171)=124.126$$

iv) 費用の計算

	数	プロット当費用	計
地上プロット	726	3.040.00	2207.040
写真プロット	25.410	7.60	193.116
写　真　点	124.146	0.76	94.351
合　　計			2.494.507

このような設計の下に現在実行中とのことである．なお対象面積250,000エーカーである由。

この中にでて来たC乗数については　次項を参照されたい．

例2は回帰を利用する二重抽出法である．

15.5　C　乗　数

回帰を利用した計算では　C乗数を用いると便利である．

回帰直線は，変数が多かろうと少なかろうと　実際のデーターから　最小自乗法を用いて計算すると　必ず，全変数の平均点を通るから　回帰式は次のようにかける．

$$Y-\bar{y}=b_1(X_1-\bar{X}_1)+b_2(X_2-\bar{X}_2)+\cdots\cdots+b_r(X_r-\bar{X}_r)$$

$Y-\bar{y}=y$,　$X_i-\bar{X}=x$　$(i=1,\cdots,r)$とおき，さらに

$\sum_{j=1}^{n} y_j^2=S_y^2$，$\sum_{j=1}^{n} x_{ij}^2=S_{x_i}^2$，$\sum_{j=1}^{n} x_{ij}y_j=S_{x_i y}$ とおくと．

上の式の最小二乗解を得るための $b_i\,(i=1,\cdots\cdots,r)$ の r 個の連立方程式は次のようになる．

$$\begin{cases} b_1 S_{x_1}^2+b_2 S_{x_1x_2}+\cdots\cdots+b_i S_{x_1x_i}+\cdots\cdots+b_r S_{x_1x_r}=S_{x_1y} \\ b_1 S_{x_2x_1}+b_2 S_{x_2}^2+\cdots\cdots+b_i S_{x_2x_i}+\cdots\cdots+b_r S_{x_2x_r}=S_{x_2y} \end{cases}$$

- - - - - - - - - - - - - - - -

$$\begin{cases} b_1 S_{x_1 x_1} + b_2 S_{x_1 x_2} + \cdots\cdots + b_i S_{x_1 x_i} + \cdots\cdots + b_r S_{x_1 x_r} = S_{x_1 y} \\ \qquad\qquad \text{----------} \\ b_1 S_{x_r x_1} + b_2 S_{x_r x_2} + \cdots\cdots + b_i S_{x_r x_i} + \cdots\cdots + b_r S_{x_r^2} = S_{x_r y} \end{cases}$$

これを簡単に表わせば

$$b_1 S_{x_i x_1} + b_2 S_{x_i x_2} + \cdots\cdots + b_i S_{x_i^2} + \cdots + b_r S_{x_i x_r} = S_{x_i y}$$

$$(i = 1, \cdots\cdots, r)$$

とはる。

これを機械的にとくには いわゆる Doolittle 法を用いればよい。(スネデッカー統計的方法、林試発行 材積表調整説明書参照)

上式の第1式の $S_{x_1 y} = 1$ 、他の $S_{x_i y} = 0$ $(i = 2, \cdots, r)$

$b_1 = C_{11}$、$b_2 = C_{21}, \cdots\cdots b_r = C_{r1}$ とおくと、

$$\begin{cases} C_{11} S_{x_1^2} + C_{21} S_{x_1 x_2} + \cdots\cdots + C_{r1} S_{x_1 x_r} = 1 \\ C_{11} S_{x_1 x_2} + C_{21} S_{x_2^2} + \cdots\cdots + C_{r1} S_{x_2 x_r} = 0 \\ \qquad\qquad \text{----------} \\ C_{11} S_{x_1 x_r} + C_{21} S_{x_2 x_r} + \cdots\cdots + C_{r1} S_{x_r^2} = 0 \end{cases}$$

r 個の C_{i1} $(i = 1, 2, \cdots\cdots r)$ の解が求められる。

次に同様にして $(r-1)$ 個の組の r 変数の連立方程式を作る。

S と S_x

$$C_{1i} S_{x_1 x_i} + C_{2i} S_{x_2 x_i} + \cdots\cdots + C_{ji} S_{x_j x_i} + \cdots + C_{ri} S_{x_r x_i}$$

$$= \begin{cases} 1 \\ \text{または} \\ 0 \end{cases}$$

ただし $i = 2, \cdots\cdots, r$ とする。

$i = 2$ とした場合、C_{11}、$C_{21}, \cdots\cdots C_{r1}$ の代りに $C_{12}, C_{22}, \cdots C_{r2}$ を上の連立方程式に入れて、$C_{22} S_{x_2^2}$ の項のある式をしておく以外の式は0とする。すなわち

$$\begin{cases} C_{12} S_{x_1^2} + C_{22} S_{x_1 x_2} + \cdots\cdots + C_{r2} S_{x_1 x_r} = 0 \\ C_{12} S_{x_1 x_2} + C_{22} S_{x_2^2} + \cdots\cdots + C_{r2} S_{x_2 x_r} = 1 \\ \qquad\qquad \text{----------} \\ C_{12} S_{x_1 x_r} + C_{22} S_{x_r x_2} + \cdots\cdots + C_{r2} S_{x_r^2} = 0 \end{cases}$$

このような組を r 個作り、その C の解を求めると、全部で r^2 の C の個が求まる。この C を C 乗数という。

この組合せは、次のようになり、その対角線に対称になっている C_{ij} は C_{ji} と同じ値をとる。

$$\begin{bmatrix} C_{11} & C_{12} & \cdots\cdots & C_{1r} \\ C_{21} & C_{22} & \cdots\cdots & C_{2r} \\ & & \cdots\cdots & \\ C_{r1} & C_{r2} & \cdots\cdots & C_{rr} \end{bmatrix}$$

この解から次の b、b の分散の推定値が計算できる。すなわち

$$b_k = C_{k1} S_{x_1 y} + C_{k2} S_{x_2 y} + \cdots\cdots + C_{kr} S_{x_r y}$$

また

$\Sigma \hat{y}^2 = \Sigma b_i \Sigma x_i y$　　$\hat{y}$ は最小二乗推定値

$r^2 = \Sigma \hat{y}^2 / \Sigma y^2$　　r = 相関係数

$S^2_{x_1 x_2 \cdots} = \Sigma d^2 / (n - m)$　m は x 変数の数、n は標本の大きさ

$S_{\beta i} = S_{y\cdot 12\cdots x}\sqrt{C_{ii}}$　　$S_{\beta i}$はβ_iの標準偏差

$S_{\beta_i - \beta_j} = S_{y\cdot 12\cdots x}(\Sigma C_{ii} - 2\Sigma C_{ij})$　$i<j$

$S_{\hat{y}}^2 = S_{y\cdot 12\cdots x}^2(\frac{1}{n} + \{C_{ii}x_i^2 + 2\Sigma C_{ij}x_i x_j)$, $i<j$

$\hat{y}$は最小二乗推定値

$S_{\bar{y}}^2 = S_{y\cdot 12\cdots x}^2(1 + \frac{1}{n} + \Sigma C_{ii}x_i^2 + 2\Sigma C_{ij}x_i x_j)$, $i<j$

$\bar{y}$は個々の値

C乗数の応用例はスネデッカー統計的方法を参照されたい。なおこの項でいうx, yはx−$\bar{x}$, y−$\bar{y}$の略であることを銘記されること。

16.6. 写真判読林相が誤まっていることを現地調査で発見した場合の面積の比の推定およびその誤差の推定

写真判読プロットの内、何%かを抽出して、現地調査を行なうが、その時、誤まって判読したことを発見した場合どうするか今、2つの層があり、5,600の写真プロットを判読して、第I層に1650プロット　第II層に3920プロットと分けられたとする。第I層の全面積に対する比率は、$P_1 = \frac{n_1}{n} = \frac{1650}{5600} = 0.300$となる。$S_{P_1}$は$\sqrt{\frac{0.3\times 0.7}{5600}}$だから95%の確率の水準では0.012（4%の誤差）となる。

今、現地調査で第1層から150プロット、第II層から50プロットを踏査した結果、第I層に3個、第II層に2個のプロットを誤判していたことがわかった。第I層の修正面積歩合(adj i)P_1は、写真判読の層の割合（P_j）と第I層として現地で決定された第j個の現地検討プロットの割合p_{j1}とをかけた$P_i p_{i1}$を各層について計算し、それを加えると求められる。すなわち

$(adj\ i)P_1 = \sum_j P_j p_{j1}$　（$p_{11} = \frac{n_{11}}{n_1}$, Mは層の個数）

この修正面積歩合の標準誤差は写真判読と地上照査の標準誤差の複合誤差で次のようにして計算する。

$$S^2(adj\ i)P_1 = \sum_j^M \frac{P_j^2 p_{j1}(1-p_{j1})}{n_j} + \frac{1}{n}\left(\sum_j P_j p_{j1} - (\sum_j P_j p_{j1})^2\right)$$

次表は、基礎資料と修正のための計算を示している。

層	写真プロット		照査プロット			修正		誤差計算		
	n_j	P_j	n_j	n_{j1}	n_{j2}	p_{j1}	$P_j p_{j1}$	$P_j p_{j1}^2$	$\frac{P_j^2 p_{j1}}{n_j}$	$\frac{P_j^2 p_{j1}}{n_j}$
j=1	1650	0.29	150	147	3	0.98	0.2842	0.278812	0.000588	0.000576
2	3920	0.70	50	2	48	0.04	0.028	0.00112	0.000392	0.000016
計	5600	1.00	200	—	—	—	0.3122	0.278924	0.000980	0.000592

誤判読による誤差を修正した面積歩合は、

$(adj\ i)P_i = 0.3122$

この推定誤差は、

$S^2(adj\ i)P_1 = 0.000989 - 0.000592 + \frac{0.278924 - 0.1952^2}{5600}$

$S(adj\ i)P_1 = \pm 0.02052$

この修正層面積歩合の標準誤差は、95%水準で、0.04104(=12.75%)僅かな誤判読の標準誤差は 4%からほとんど13%に増大した。このように、僅かな誤判読はかなりの影響をサンプリングに与えることが示される。

2項分布についての95%信頼区間(百分率)

観察数 f	標本の大きさ n										観察比率 f/n	標本の大きさ			
	10	15	20	30	50		100					250		1000	
41					67	91	37	51			41	34	47	38	44
42					71	93	32	52			42	36	48	39	45
43					73	94	33	53			43	37	49	40	46
44					76	95	34	54			44	38	50	41	47
45					78	97	35	55			45	39	51	42	48
46					81	98	36	56			46	40	52	43	49
47					83	99	37	57			47	41	53	44	50
48					86	100	38	58			48	42	54	45	51
49					89	100	39	59			49	43	55	46	52
50					93	100	40	*60			50	44	+56	47	+53

* f が50を超えるときは 100−f=観察数として表を読み、その各信頼限界を100から引け

+. f/n が50を超えるときは 100−f/n=観察比率として表を読み、その各信頼限界を100から引け

2項分布についての 95% 信頼区間（百分率）

観察数 S	標本の大きさ n						観察比率 f/n	標本の大きさ	
	10	15	20	30	50	100		250	1000
0	0 31	0 22	0 17	0 12	0 07	0 4	00	0 1	0 0
1	0 45	0 32	0 25	0 17	0 11	0 5	01	0 2	0 2
2	3 56	2 40	1 31	1 22	0 14	0 7	02	1 5	1 3
3	7 65	4 48	3 38	2 27	1 17	1 8	03	1 6	2 4
4	12 74	8 55	6 44	4 31	2 19	1 10	04	2 7	3 6
5	19 81	12 62	9 49	6 35	3 22	2 11	05	3 9	4 7
6	26 88	16 68	12 54	8 39	4 24	2 12	06	3 10	5 8
7	35 93	21 73	15 59	10 43	6 27	3 14	07	4 11	6 9
8	44 97	27 79	19 64	12 46	7 29	4 15	08	5 12	6 10
9	55 100	32 84	23 68	15 50	9 31	4 16	09	6 13	7 11
10	69 100	38 88	27 73	17 53	10 34	5 18	10	7 14	8 12
11		45 92	32 77	20 56	12 36	5 19	11	7 16	9 13
12		52 96	36 81	23 60	13 38	6 20	12	8 17	10 14
13		60 98	41 85	25 63	15 41	7 21	13	9 18	11 15
14		68 100	46 88	28 66	16 43	8 22	14	10 19	12 16
15		78 100	51 91	31 69	18 46	9 24	15	10 20	13 17
16			56 94	34 72	20 46	9 25	16	11 21	14 18
17			62 97	37 75	21 48	10 26	17	12 22	15 19
18			69 99	40 77	23 50	11 27	18	13 23	16 21
19			75 100	44 80	26 53	12 28	19	14 24	17 22
20			83 100	47 83	27 54	13 29	20	15 26	18 23
21				50 85	28 57	14 30	21	16 27	19 24
22				54 88	31 59	14 31	22	17 28	19 25
23				57 90	32 61	15 32	23	18 29	20 26
24				61 92	34 63	16 33	24	19 30	21 27
25				65 94	36 64	17 35	25	20 31	22 28
26				69 96	37 66	18 36	26	20 32	23 29
27				73 98	39 68	19 37	27	21 33	24 30
28				78 99	41 70	19 38	28	22 34	25 31
29				83 100	43 72	20 39	29	23 35	26 32
30				88 100	45 73	21 40	30	24 36	27 33
31					47 75	22 41	31	25 37	28 34
32					50 77	23 42	32	26 38	29 35
33					52 79	24 43	33	27 39	30 36
34					54 80	25 44	34	28 40	31 37
35					56 82	26 45	35	29 41	32 38
36					57 84	27 46	36	30 42	33 39
37					59 85	28 47	37	31 43	34 40
38					62 87	29 48	38	32 44	35 41
39					64 88	29 49	39	33 45	36 42
40					66 90	30 50	40	34 46	37 43
41					69 91	31 51	41	35 47	38 44
42					71 93	32 52	42	36 48	39 45
43					73 94	33 53	43	37 49	40 46
44					76 95	34 54	44	38 50	41 47
45					78 97	34 55	45	39 51	42 48
46					81 98	36 56	46	40 52	43 49
47					83 99	37 57	47	41 53	44 50
48					86 100	38 58	48	42 54	45 51
49					89 100	39 59	49	43 55	46 52
50					93 100	40 * 60	50	44 + 56	47 + 53

* f が 50 を超えるときは、100 − f = 観察数として表を読み、その各信頼限界を 100 から引け

+ f/n が 50 を超えるときは、100 − f/n = 観察比率として表を読み、その各信頼限界を 100 から引け

(1971/1988)