経　営	6 7
測　定	1 9

関連のある測定値の解折法

Quenouille

昭和３４年３月

農林省林業試験場経営部

序

近代統計学の林業への応用は、最近著しいものがある。特に、関連のある測定値の統計的解析法は、林業において、広い応用範囲を持つものであるが、この種の問題を扱った統計的方法は、従来、質的特性に関するものが主であって、われわれに関係の深い、量的測定値の解析法について、述べたものは少なかった。

本書は、数学的説明を最小限に留め、主として生物学の例を用いて、巧妙に、統計的解析法と、その実際的な目的とを、系統的に説明している。

内容は、解析法の形式に従って、4部に分けてある。才1部は、新旧両法による、遅関のグラフによる推定法と検定法について説明している。才2部は、数値解析法という標題で、分散分析、共分散分析、回帰分析法が簡単に述べてある。推定および解析の簡略法に関する才3部では、観測値の大まかな組分けによる回帰の簡略推定法や、さらに進んだ共分散分析の使用法について論じている。最後に、解析的計算法という表題の下に、特殊な問題と解析、時系列解析、多変量解析の3つの章が含まれている。時系列解析では、相関係数と共分散分析とを用いる近似的方法を主として扱っている。ただ本書の記述には、ところどころ不注意な点があり、又説明に用いている例は、特殊なものであり、その説明は、はっきりと理解できるにもかかわらず、充分な理論的背景を持たない読者は、著者が考えているものと幾分違った例に、本書に説明されている方法を適用する場合には、困難を感ずるかも知れない。しかし、統計的理論について若干の知識があれば、本書の解析法は、林業の諸問題の解決に、有力な武器となるであらう。

なお、この訳は、当研究室の栗屋技官の原稿に手を加えたものである。

昭和３４年３月

測定研究室長　大友栄松

目　　次

1　グラフによる研究　Graphical Analysis

1.1.　連関という問題　Problem of association

調査の際、二つ又はそれ以上の量の間の関係について調べたい場合が屢々ある。この様にして動物の体重の変化を摂取する食物の量と関係づけたり、時とともに推移する人口の変化を調べたり、経済的変量が他の変量におよぼす影響を研究したりする。

普通我々が出合う問題は3つの組に分類出来る。この3つの内どれか1つが必要なこともあれば、その全てが必要な場合もある。

1.　連関の有無、さらに連関があれば連関の程度の推定

2.　連関の型の推定

3.　さらに進んだ研究での推定された連関の利用

その個々の方法について考察しよう。

しかし最初に次の点をはつきりと理解しておかねばならない。即ち本書で述べている連関を調べる方法はこの3つの目的に限られている。（この方法は連関の原因についての情報を全然呈供しないし、調査範囲を越えるものについては連関状態についての情報も与えないであらう。）かゝる情報は連関の性質の仮定に基づいているか或はそれと結び付けた統計解析によつて求められるが、統計解析だけでは役に立たない。実際、この様なことが出来るならば妙なことになるだろう。もしそうであれば、例えば途中で起る事件にかゝわりなく2000年間にわたるGreat Britainの人口の予想が可能であり、発生する順序或はその他の理由にかゝわりなく観測値を〝原因〟と〝結果〟の2組に分けることが可能となるからである。

測定値の連関を調べていることや、このような研究は、連関の原因を調べたり、将来又は、拡張された連関の状態をさえ、予測するものでないと

いうことは、とかく解折の途中では忘れ勝ちである。なんらかの連関の起りそうな状態を知つて、原因の探求や予測などもできる場合が屢々ある。このような原因の探求や将来の予測は、研究者の仕事に属するものである。

前掲の問題の才1は連関の有無を検べることであつた。この段階は必要のないことが多いであらう。2変量間に関係のあることが分り、そして連関の型を推定するのが才1の要務である。連関の有無が分らなければ、2変量をグラフにプロツトすればその関係がはつきりと示され、連関の型の推定に直ちに取りかゝることが出来る。しかし、この様な推定に先だつてこの関係の適切な代数学的表示法を判定しなければならないので、どの様な研究でも、才1段階としてデータをグラフにプロツトして調べてみるのが普通である。グラフの使用法については本章および次の2章で考察しよう。

1.2. グラフと散布図

解折に先だつて、データをグラフで調べてみれば不必要な仕事ははぶけることが多い。

グラフは2つ以上の変量間の連関を調べ、その解折の仕方を示す簡易な方法である。短的に云えば、調査者が次の様な決定をするのに役立つ。

1. データの統計的解折により得られるものがあるだらうか、測定値の組が明らかに無相関であつたり、その関係が殆んどなく、実際には重要でないことが度々ある。或は2組の測定値間の関係が、全然解折の必要のない程はつきりしていることもあろう。

2. 観測値の組間にどの様な連関があるのか又、連関があればどの様な型をしているか。

例えば、何んらかの連関があれば、その関係を表す手段として直線で充

分であるかどうかということをグラフは示すであらう。

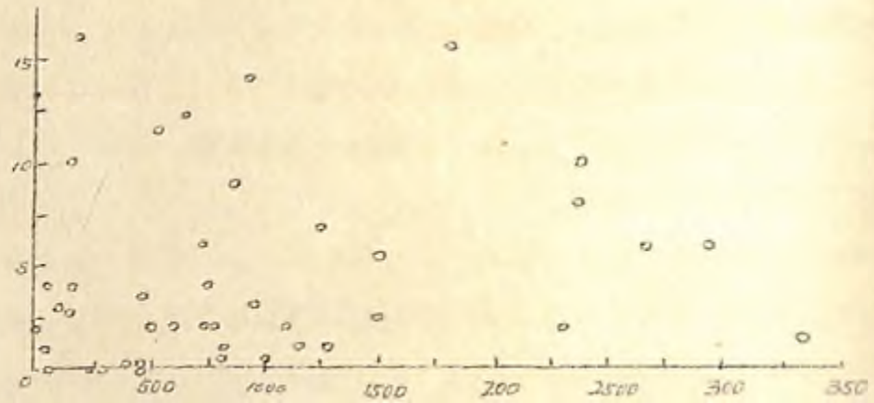
3. 解折を始める前に観測値の測定尺度を変えること。即ち変換又は観測値の重みずけが必要であるか。例えば図の或る部分では散ばりがかなり大きいとか、或る変量のとる範囲が、別の変量の値によつて変るという様なことがある。どちらの場合でも、解折を正しく行うには、変換又は、何んらかの重み付けが必要である。

4. どの様な観測値を研究又は棄却する必要があるか。グラフ上の或る点から離れておれば、この様な観測値の妥当性、さらに妥当とみなした場合に、解折に不当な影響をおよぼすかどうかを考察してみる必要がある。

データをグラフで調べてみることが色々な面で役立つている例を1図～4図は示している。

1図は一定の期間において漁船で得られた観測値をプロツトしたものである。これは補獲した時点におけるニシンの漁獲高とCalanus（ニシンの餌料）の数とを示している。これはニシンの漁獲高とCalanusの数との関係を調査した時A.C Hardy 等の集めたデータの一部である。両者の関係は明らかに有意でないからこの関係を推測或は検定するため観測値を解折してみても無駄であろう。しかしより広般な調査の一部と考えれば違つた条件におけるニシンの漁獲高の指標としてのCalanusの数の信頼度を示すため、この観測値は他のものと関連させて使用されるであろう。

測定値間に明らかに連関がある場合の2例を2図は示している。2a図には、干草の標本に含有されている可溶性硝酸塩の百分率がペプシン処理後の可溶性硝酸の百分率に対してプロツトしてある。（A.J.Barnettのデータ）この関係は殆んど直線性であり、直線で充分表示出来る。しかし、曲線によれば、この連関はもつとよく表示されることが暗示されている。若いネズミのヘモグロビンの百分率と総体重の百分率で表わした心臓



1　図　Calanus 数に対するニシンの漁穫高

の重量との関係を示している2b図では正しくこの通りである。(M.Richard.A.Greigのデータ)、この例では、直線は観測値間の関係を表わす形としては明らかに不適切なものである。

3図は、3～14才の男の子の体重に対してプロットした基本新陳代謝量を示している。この場合には、この関係は直線で表示出来るけれども体重の増すにつれて散ばりが大きくなつている。したがつて測定の尺度を変えるか、或はむしろ、その信頼度に従つて観測値に重みを付けることを、解析を始める前に考察してみるべきである。

4図は、1才児の頭と胸の周囲をプロットしたものである。中心の周りに点は巾広く散ばつているが、特に3つの点が中心から離れている。その1つは、31.8cmの胸囲、2番目は39.2cmの頭の周囲、3番目は、胸囲49.0cm、頭の周囲50.2cmの点である。このような場合には観測値のとり方が正しかつたかどうかを、まず考えてみる必要がある。この場合には異常な観測値をチエツクできたし、測定および写しには全然誤りはなかつ

た。そこで、これらの観測値を解析に使用すべきであるか否かについて考察する必要があつた。もし、これらをそのまゝにしておけば解析に重大な影響をおよぼすことは明らかである。測定値が1つだけで異常な値をとる初めの2点は特にそうであるし3番目の点は頭と胸の周囲が大きいだけであつて、他の点程散布図の中心から離れておらず、連関の傾向とかなり一致している。よつて初めの2点を解析の際棄却することにする。

どの観測値を棄却するかを決めるには、いろいろと考察してみなければならない。観測の施行中或は記録に疑があるとか、とられた値が異常なものと考えられ、正常な値の解析に関係なければ、棄却する。大低の場合、観測値を棄却し、次いでこれらが偶然変動の結果起つたものとしては生じそうもない程、極端な値が否かを検定出来るであらう。この方法については後で考えることにする。

1.3.　尺度の変換

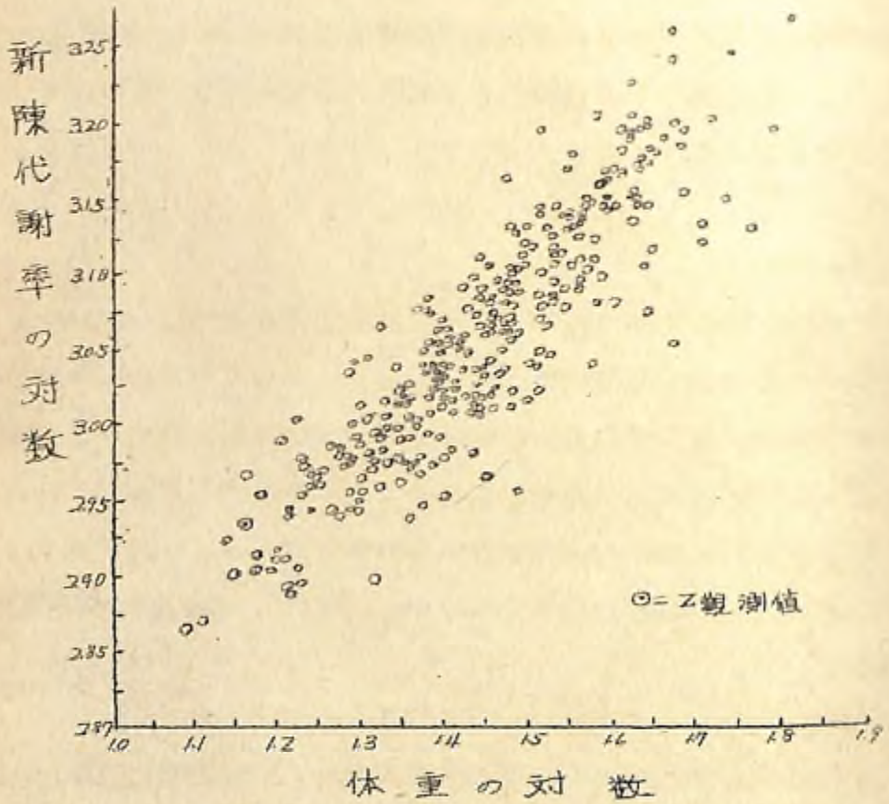
データの吟味、解析を簡単にするのに尺度の変換は役に立つ。一般に要求されることは散布図の各部分の点の散らばりが等しくなる様に尺度を変換することである。そうすれば次の作業で各観測値の精度の相違を考慮する必要はない。例えば3図では場所によつて点の散らばりが違つている。散布図を描くのに基本新陳代謝量と体重の対数を使えば、5図が得られる。この図では点の散はりは一様になつており、二変量間の関係は極めて直線に近くなつている。

したがつて解析はそう苦労しなくても行うことができる。

測定尺度の変換によつて得られる才2の効果は、二変量間の連関を表わすために使われる関係の形が単純になることである。例えば、2b図で測定値の対数を使えば、6図が得られる。その場合、曲線性は殆んど無くな

り、もつと詳細に調べてみないと曲線がこの関係を一層よく表示しているかどうか言明することは不可能であるが、直線が使えればこの関係を表わすことができるであろう。この問題は6章で考察することにする。

測定値目盛の決定は、連関の解折に欠くことができないものであるがかなりの欠点もある。目盛付けを考慮するのに失敗すれば解折の際偏りが生じ、又ある目的のために用いられた目盛は別の困難性の生ずる原因ともなる。目盛付けで起る問題については10章で考察することにする。



5 図　男児の体重の対数に対する新陳代謝率の対数

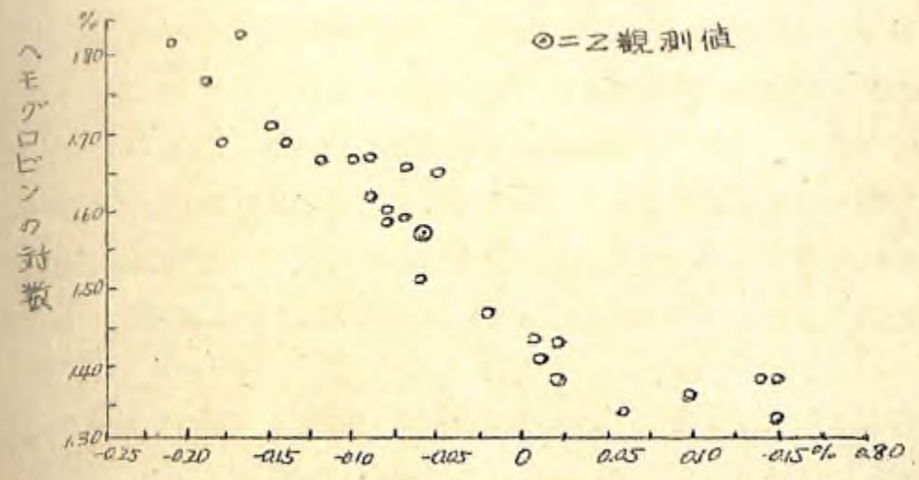
1.4 従属変量と独立変量　Dependent and independent variables

ある変量が別の変量に従属している程度、状態を決定する必要のある場合が多い。例えば身長に対する体重の関係を求め、いろいろな身長に対する体重を予測したい場合を考えてみる。この場合、体重は従属変量、身長は独立変量といわれる。

2つの変量の関係を図で調べてみたい時には独立変量が同じか、略同じ観測値の組の平均を使うと便利なことが多い。例えば4図に示してある観測値は、胸囲を42.0cm以下、42.0～42.9cm、43.0～43.9cm……49.0cm以上の組にまとめられ、頭の周囲と胸囲の関係を示すグラフをプロットするのに各組の平均を使うことが出来る。

7図はこの関係を示している。

観測値を頭の周囲により組分けすれば、導つた線が得られることに注意せよ。頭の周囲に対する胸囲の関係を示す線は7図に点線で示してある。



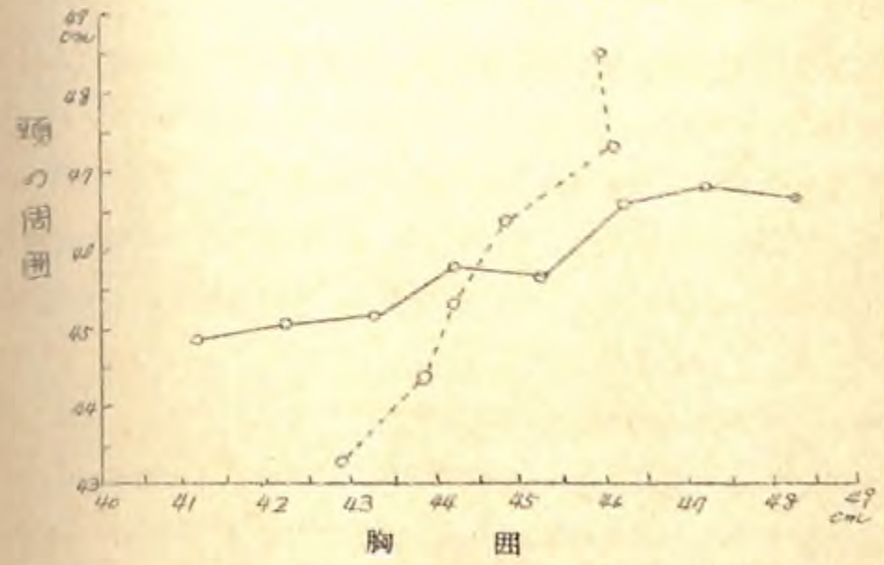
6 図　若いネズミのヘモグロビン百分率と心臓重量の百分率

この2つの線の区別は特に重要であり次章でもつと詳しく考察しよう。

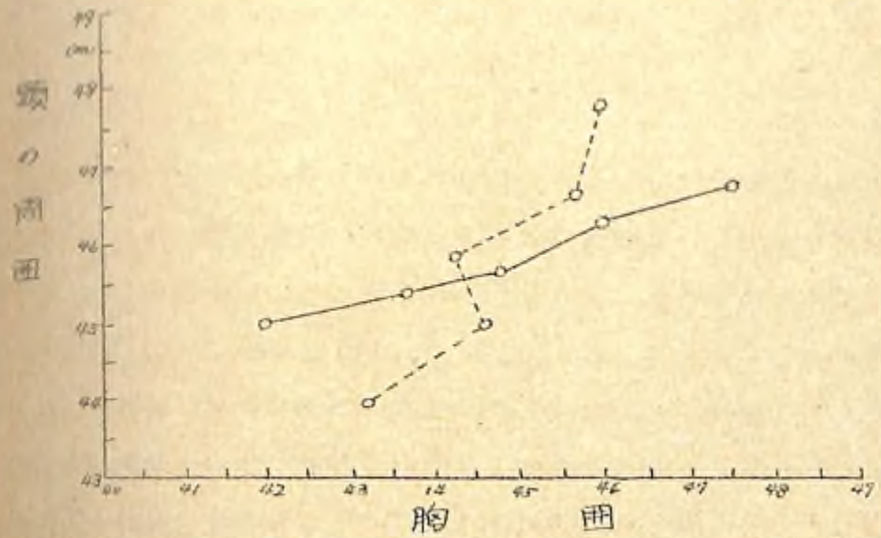
この方法で観測値を組にまとめた結果、2つの障害が起る。才1に組にまとめられる観測値の範囲が大きなものであり、2変量間の関係が極めて曲線性の強いものであると、平均した値はその関係の形について誤まつた考えを抱かせる。しかし、この様なことは実際には稀であり、通常平均することによつてその関係が著しく歪む様なことは起らない。才2に、平均値は特に観測数の違つたものから求められているので平均値の不正確性は簡単には評価できない。例えば、5図では点は5～34個の観測値から求められている。この結果各点の精度はまちまちであり、その線のもつ、はつきりした特性の信頼度を確めることは難しい。才2の障害を除く一つの方法は観測数の等しい組にまとめることである。例えば観測値を大きさの等しい5つの組に分割し、各組の平均を使う。この様にして得られた5点からなる散布図は余り複雑でない観測値間の連関の形態を判定する場合に特に役に立つ。8図は4図のデータから求めた5点からなる線を示している。この例では各点は約23個の観測値から求めたものである。

二変量間の直線的連関を速かに検定し推定しうるのが5点からなる線を使う長所である。例えば2組の測定値間に全然連関がなければ、一様に上昇或は下降することが5点からなる線で観られる機会は30回に1回より少い。したがつてこの様な上昇或は下降が観られゝば、測定値間には連関があると確信できる。8図にこのように胸囲と頭との内に連関があることを示している。

4図に関連してとられた7図、又は8図は2組の測定値間の関係が直線で充分表わせることを示している。これらの線の方程式は、4図或は8図から直接推定できるが、出来る限り正確にしたいならば、数値解析を4章で説明する方法で行う。



7　図　　胸囲に対する頭の周囲の関係（———）と
頭の周囲と　　胸囲の関係（- - - - -）を示す線



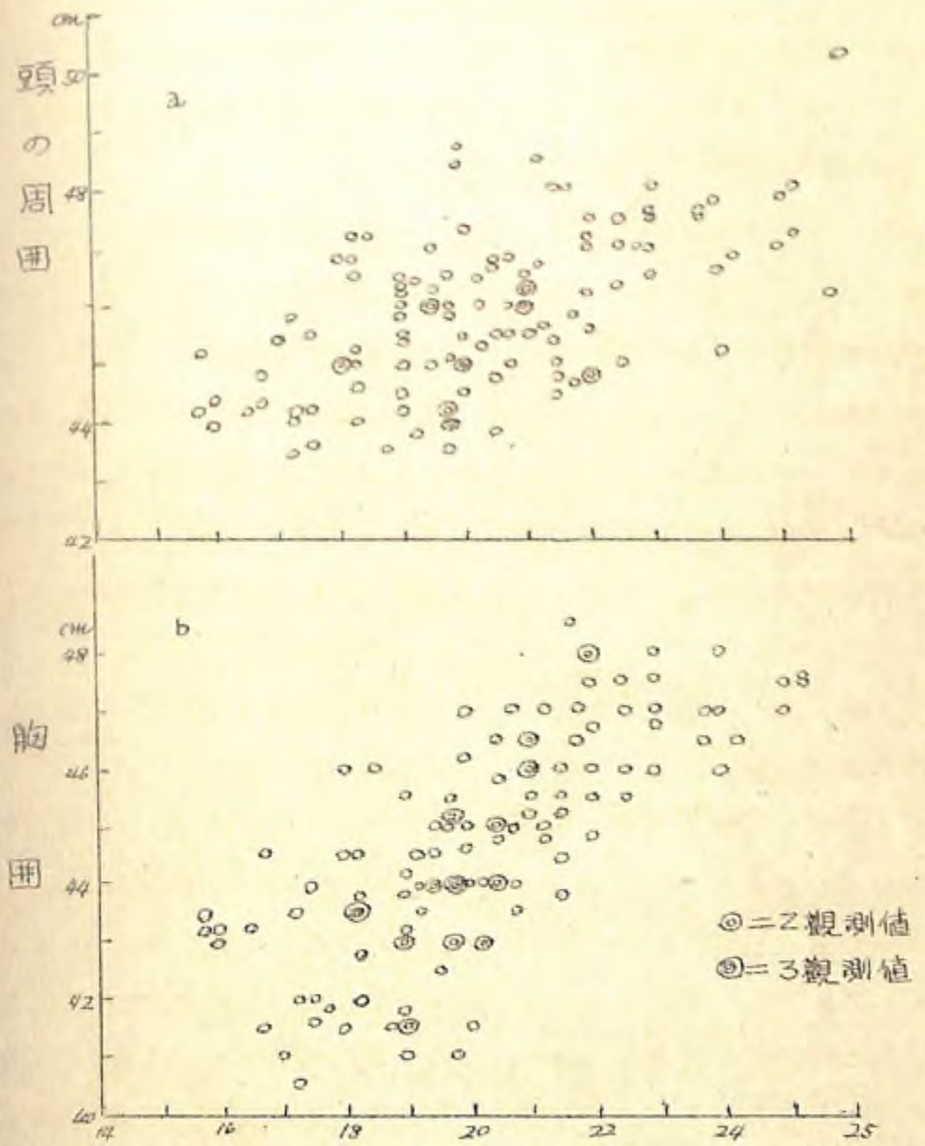
8　図　　胸囲に対する頭の周囲の関係（———）と頭の周囲
の関係（- - - - -）を示す　5－poimt line

1.5. 重連関を示すグラフ Graphs showing multiple association

時には1種類の測定値と数種類の測定値との連関を調べたいこともある。例として穀類の収穫前の気温と降水量との関係を調べたいとする。かゝる場合には、最初の測定値を他の各測定値とどの様な関係にあるか他の組の測定値がお互にどの様な関係にあるか、あらゆる組の観測値を解析に使う必要があるのか即ち、どの測定値が不要であるかを考察してみる必要がある。

第1段階として普通は観測値を対にして相互にプロットする。例として頭の周囲が体重と胸囲とどの様な関係にあるかを知りたいとする。4図の様に胸囲に対する頭の周囲、又9図の様に体重に対する頭の周囲、体重に対する胸囲をプロットする必要がある。これらのグラフを調べてみると一般に他と一致しておらず次の研究では除外しなければならぬ観測値が分る。前に4図に示したかけ離れた観測値と同じく、9図においても、体重の極く小さい処にかけ離れた観測値が1つある。これは9a図では特にはっきりしており、次の解析では除外した。

前にも指摘した様に、棄却すべき観測値を決めるにはいろいろと考察してみなければならない。数組の観測値を同時に考える場合にはある種の測定値のかけ離れた値が他種の測定値の対応する値との不一致性を示しているか否かを考察してみる必要がある。これに他の観測値の組のどれか一つのかけ離れた値で説明出来るかもしれぬということである。或は連関の形に関する特定の仮定と一致するかもしれない。例えばこのかけ離れた観測値は、体重は胸囲と、胸囲は頭の周囲と相関があるが体重は頭の周囲とは間接的にしか相関がないという仮定と一致している。この型のかけ離れた

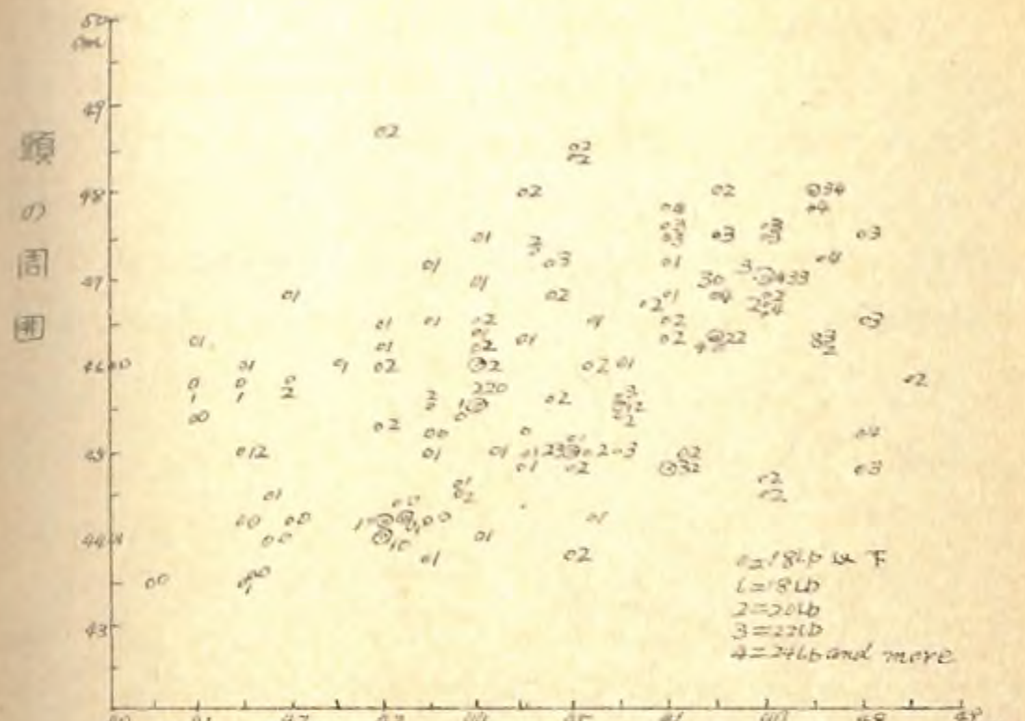


9 図　体重に対する頭の周囲と胸囲

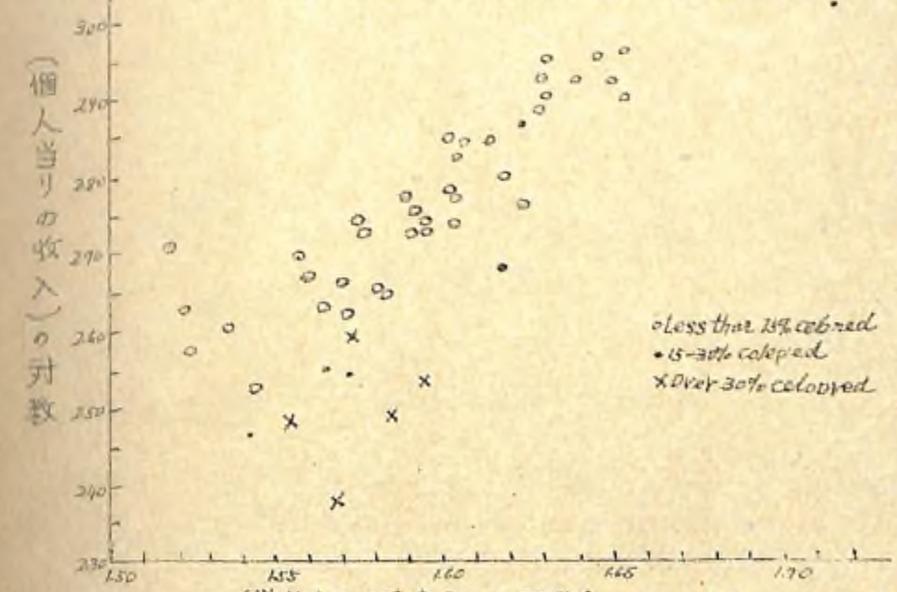
点は解析によつて得られる結論に明らかに重大な影響をおよぼす。間接的であるが胸囲と両測定値間の相関の頭の周囲と体重との間に相関があると認めれば、9図のかけ離れた点は解析に含ませてもよい。しかし、この場合は各測定値の関連状態を調べるのが解析の目的であるから、その影響を避けるため、解析の際にはかけ離れた観測値を棄却する方が良い。これは最初の解析が終つてからさらに調べてみる。

ある組の測定値と他の組のものとの関係を調べる時の次2段階としては2組以上のものを同じグラフ上にうまく表わせるかどうかをみてみる。例えば10図は各小児の体重を0～4なる数字で表わして、胸囲に対する頭の周囲をプロットしたものである。この図表には各小児の体重、頭および胸の周囲が示してある。重い体重の小児は頭および胸の周囲の大きな小児であり、頭の周囲は体重又は胸囲を使つて予測できるが、この両者には密接な連関があるから、両者を使つたとしても得られるものは少いことが分る。

11図には、3種の測定値を示す次2の散布図の例が示してある。この場合にはU，S，Aの48州にコロンビヤ地方を加えた1940年における値（全部で49点）を点は表わしている。プロットしてある値は個人当りの収入の対数（ドル単位）と各州の労仂人口の百分率であり違つた種類の点は有色人種の人口割合の表わすのに用いてある。この場合には、有色人種の割合の高い州の個人当りの収入は他の州に較べて低く、したがつて労仂人口の割合や有色人種の割合についての観測値は個人当り収入と同時相関のあることが分る。もちろん、この2つの測定値が個人当り収入の変動の原因となつていることを意味するのではなく、その指標となるに過ぎないことをこのことは意味しているのである。



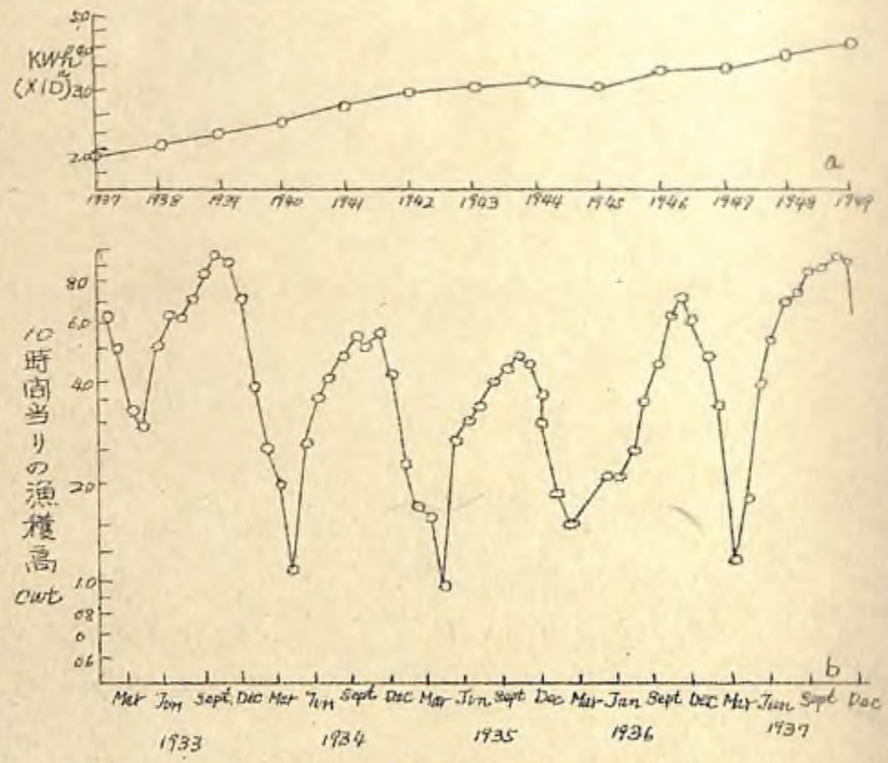
10図　体重で組にまとめた1才児の頭の周囲と胸囲



11図　(労働人口の百分率)の対数に対する(個人当り収入)の対数

1.6. 時　系　列　　Time series

一連の観測値が違つた時点でとられている時、これらの観測値は時系列を成すといわれる。２つの主目的のために時系列は使われる。将来の状態の予想を目的として、観測された変量の変化する傾向を考究するために用いられる。或は、ある時系列の観測値の変化が他の対応する変化とどの様な連関にあるかを調べたいこともある。例として、ある期間（年）観測した穀物収量と、これらの年における収穫前の平均気温と降水量との連関を調べたいとする。



１２図　a、Kingdom における発電量
b、North Sea の南部地区での１０時間当りニシン漁穫高（cwt）

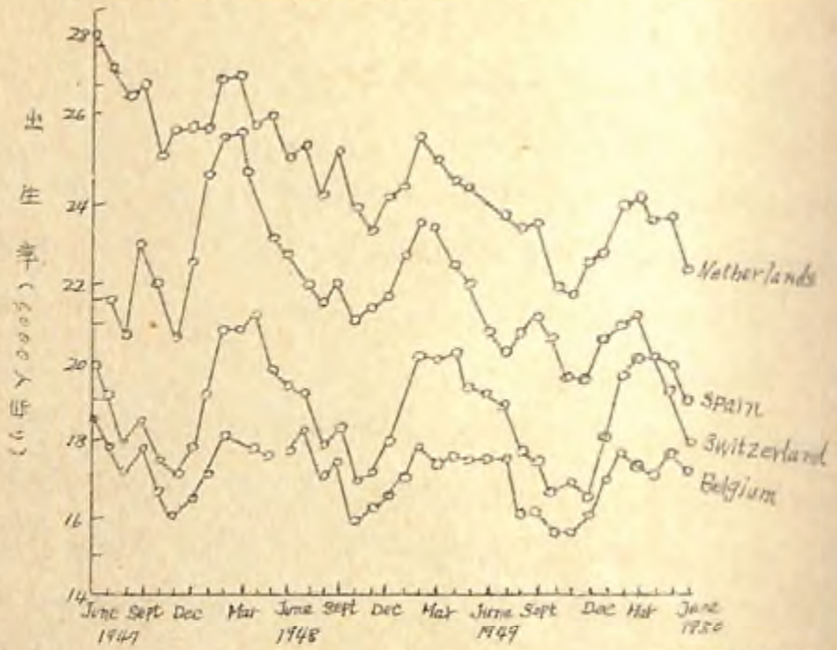
他の連関の場合と同じく、時系列を調べるには、まずグラフに描いて調べてみなければならない。一般に、横軸に時点をとり、各時点の観測値を順次直線で結ぶ様にする。１２図は対数目盛でプロットした２つの時系列の例である。

１２a図は大体一つの傾向からなつている時系列を示している。しかし、１２b図は一般的傾向に月による評価出来ない変動が加わつて、季節的変動が極めて大きな時系列を示している。どちらの場合でも、この最初にプロットしたものは時系列の特性を示している。

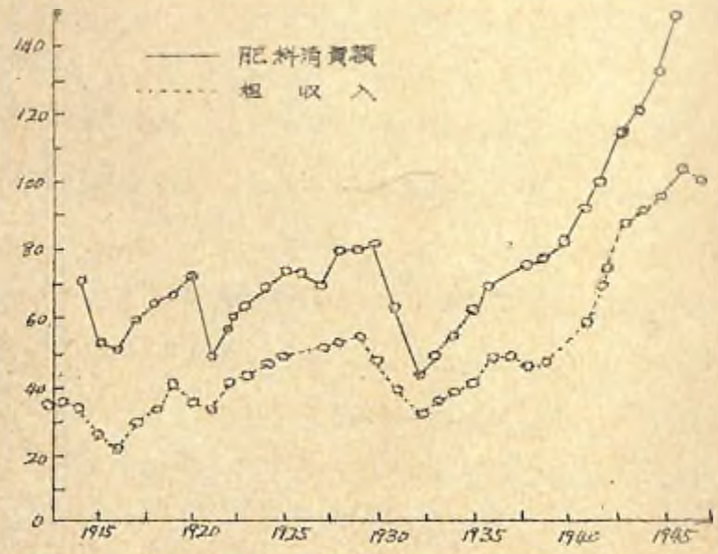
時系列をなす２組以上の観測値間の連関を調べたい場合には、時点による変化を研べることができる様に、まず同じグラフに時系列をプロットしてみる。この様な仕方よりてどの様な連関があるかが示される。例えば、１３図の時系列間の連関はいづれも主としてこれらの系列に共通な季節的変動のためであることが分る。

これとは逆にU.S.A.における肥料消費高と潜在的消費指数（農家の粗収入を化学製品の卸売価格指数で除して測つたもの）を示す１４図の時系列間の連関は主として共通な傾向のためと考えられる。この図を詳細に検討してみると肥料消費高は粗収入に時間的ずれていることが示される。両測定値の谷は一致しているが、消費高の頂点はいづれの場合でも収入の頂点の一年後に起つている。

したがつて同一年度の粗収入を使つて肥料消費高は予測できるが、前年度の粗収入を考慮すればこの予測は改良されることがこの図の検討によつて分る。もつとはつきりした結論に到達するには、さらに調査と解析を行なわねばならない。それにもかゝわらずグラフの助をかりてデータを慎重に予め吟味するだけでも得ることが多い。



１３図　オランダ、スペイン、スイス、ベルギーにおける出生率



１４図　U・S・A における、肥料消費額と農家粗収入（化学製品の価格で表わしてある）

2　グラフによる推定　Graphical Estimation

2.1.　グラフを使つての推定 Estimation using graphs

２つ又はそれ以上の変量間の連関の形の数学的即ち代数学的表現を求めたい場合が多々ある。これはグラフの助けを借りて行える。完全な統計的解析によつて推定される関係の形を計るのにグラフは使われる。しかし、時には２つ又はそれ以上の変量間の関係を直接グラフから推定できる場合もある。この方法を本章では考察する。

２つ以上の変量間の連関の適当な表現をグラフを使つて推定することは完全な統計解析よりは一般に精度は低いが、大低の目的にとつて充分間に合う精度に達し得る場合が多い。グラフを使つて求めた予備的推定値はさらに精しい調査や或は完全な統計的解析の有効度を計るために用いられる場合が多い。

グラフによる推定値の主な欠点はその主観的性格にある。人が違つたり、人は同じでも場所が違うと同じデータから異なつた推定値がひきだされる。通常、この種の不正確性のあることを知り、最終的結論を下す際なんらかの斟酌をすれば、このことは問題とはならないであらう。

実際にはグラフによる推定は完全な統計的解析に較べて遙かに迅速で、精度も等しい推定方法であるといえる。しかし、関係を推定する確実で迅速なやり方を決めることは不可能である。

したがつて本章は図形的推定方法におけるグラフの用法について２、３のヒントを与えるに過ぎない。

2.2.　直線的関係　Straight line dependences

直線的関係を推定する時には、従属変量と独立変量の区別を念頭におか

ねばならない。1.4節に例示してある様に、どの様な変量の組についても、相互の従属関係を示す2本の線が存在する。互に他の変量を正確に決定できる様な完全な対応関係が2変量間に存在しない限り、この2本の線は一致しないであらう。

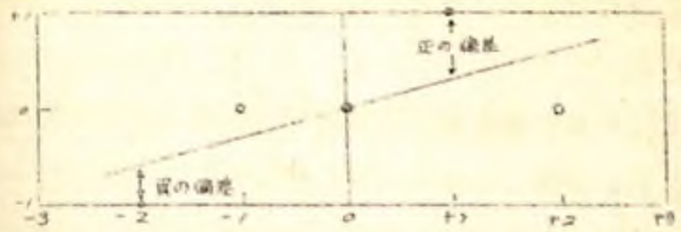
この様な場合変量間の連関は完全であるといわれる。しかし変量間に完全な連関があれば、従属関係の推定には全然問題は起らず、通例2つの従属関係の区別が必要とならう。

ある変量Xに対する別の変量Yの従属関係を表わす線はXに関するYの回帰線と呼ばれている。Xのどの様な観測値に対しても回帰線上にYの対応する期待値が存在するであらう。勿論、Yの個々の観測値は回帰線上の期待値とは一致していないかもしれない、この差を回帰線からの偏差といわれる。15図は直線回帰とこの線からの偏差の例である。線より上にある観測値は正の偏差、下にある観測値は負の偏差をもつといわれる。

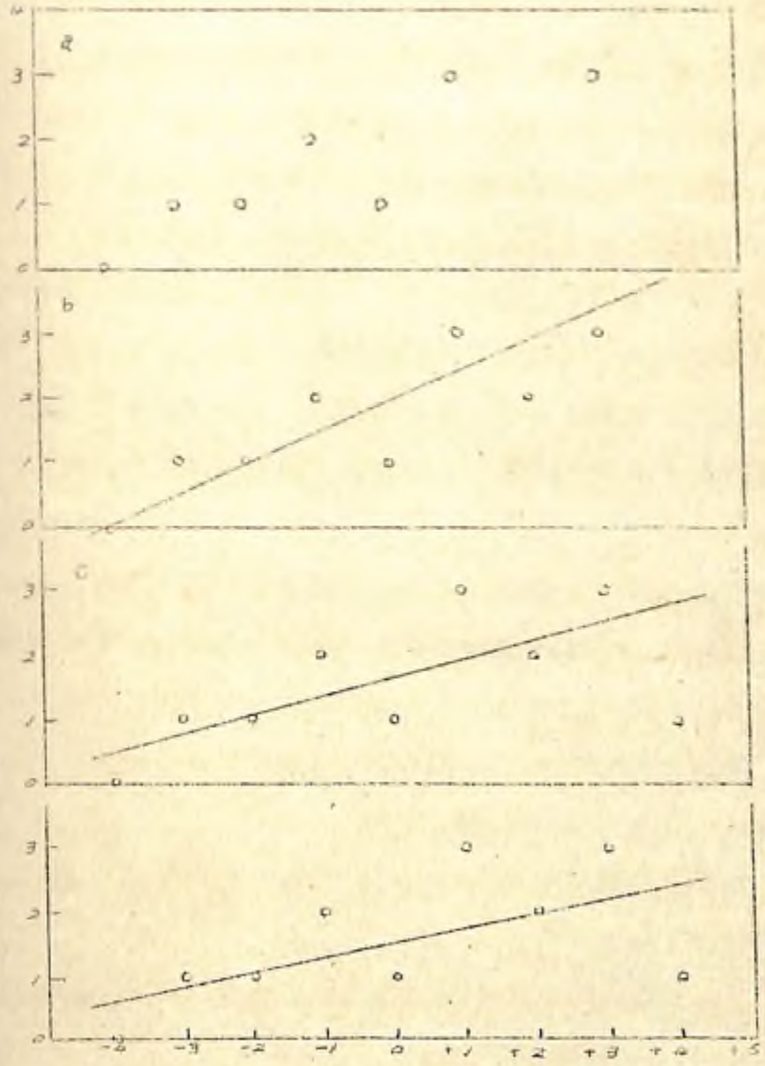
さて図形的に回帰線を推定する場合の規約について考察しよう。広く使われている3つの規約は、1）回帰線からの偏差合計は0である。2）回帰線からの偏差の各象限における和は0になるべきである。3）回帰線は偏差が最小になる様に、特に大きな偏差を減す様に定めるべきである。直線回帰、曲線回帰のいづれを推定する場合にもこの規約は用いられる。

この規約の応用例として16図を考えてみる。

直観的が16a図に示してある。才1回の当てはめが16b図に示してある。この大略の当てはめには多くの不備の点がある。線の右端にある大きな負の偏差は他のものと釣合がとれておらず、又正の偏差は線の左端に存在するだけである。この不備を除くには160図の様に線の右端を下げ左端を上げる必要がある。この線はかなり良く点と一致しているが、線の右端にある大きな負の偏差を減すことにはならなかつた。



15図　回帰線と偏差



16図　回帰直線のあてはめ

回帰線の最終的近似が１６ａ図に示してある。この時には上記の規約は全て考慮されている。

勿論、実際には何本もの線を描く必要はなく、普通良好な近似は上記の規約が満されるまで、用紙上で物指又は糸を動かして求められる。

直線が多数の観測値に適合している時には個々の観測値は重要でなくなる。普通、当てはめは２以上の段階をふんで行なうのが最も良い。

観測値の組の中心点が初めに求まれば、この線は比較的に小数の点にも当てはまる。特に１.４で述べた様な五点図表が、線型関係を推定するために使われる。勿論この様な図を作る時正確な平均を計算する必要はなく、この平均は簡単に判定でき、図上にはつきりと示されている。

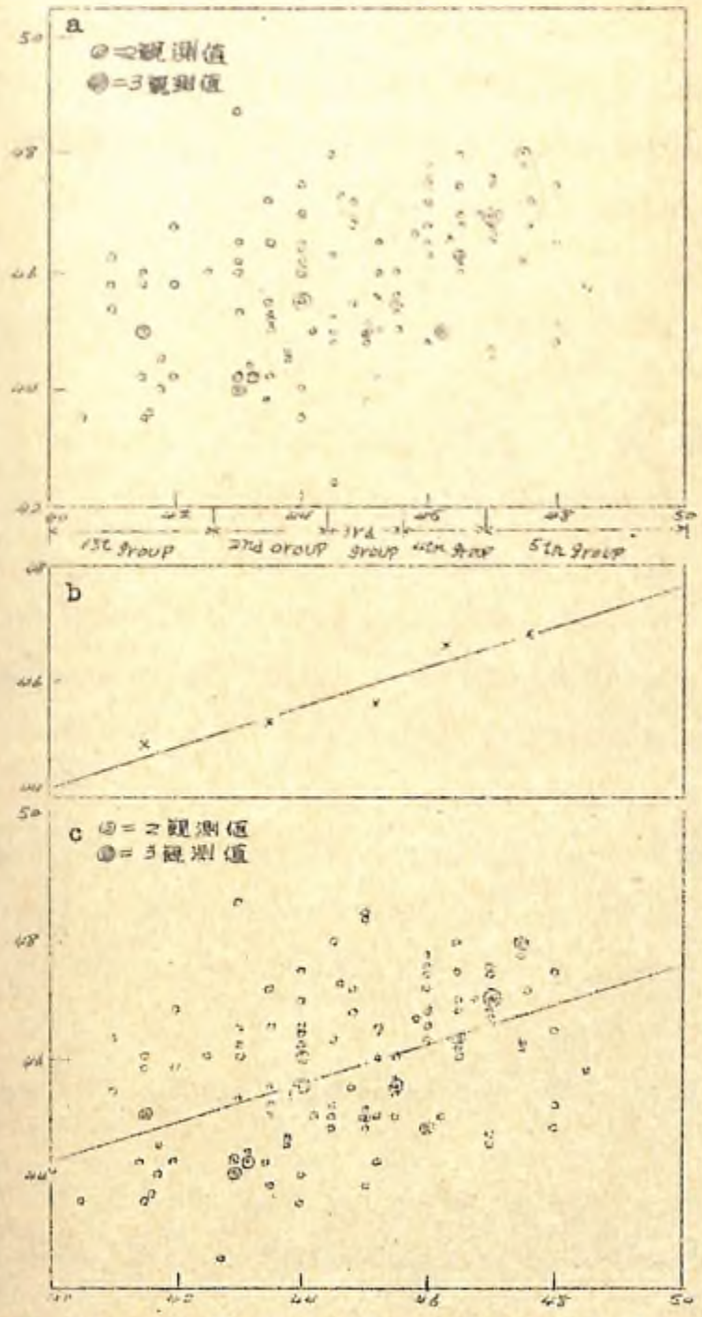
１７図は４図の観測値に、胸囲に対する頭の周囲の従属関係を示す直線を当てはめた例である。前に示した様に胸囲で点を組にしたものの大体の中心を表はす５点と観測値とを１７ａ図は示している。この平均点に大体適合させた直線を１７ｂ図は示しており，１７ｃ図は元の観測値に対する同一直線を示す。

この様な当てはめを行う時には、大きな偏差を決め、これを考慮することによつてどの様に当てはめが改善されたかをみるため、常に元の観測値について当てはめた線を調べてみる様にする。１７ｃ図において、３cm以上の偏差は、僅か２個であり、これは相互に相殺される傾向がある。当てはめた線を再考してみる必要は全然ない。

普通にはこの全過程は同一の図で行なわれることに注意せよ。平均点は薄く書き、後で消す様にする。

直線を当てはめた場合図からその方程式を求めると便利である。これは次式で決定するのが最も便利である。

$$Y - Y_0 = b (X - X_0)$$



１７図　観測値の数が多い場合の回帰直線のあてはめ

bは直線の傾き、X_0 , Y_0 , は直線上の任意の点であるが、中心に近い点が望ましい。例えば１７ｂ図、１７ｃ図の線は、胸囲の１０cm増加に対し、頭の周囲が3.4cm増すことを示している。故にｂ＝0.34 である。この線は点（ 44.00 , 45.55 ）を通る。したがつてその方程式は、

$$Y = 45.55 = 0.34(X-44.00)$$

即ち

$$Y = 0.34X + 30.59$$

2.3. 曲線関係 Curved line dependences

直線関係の推定について前節に示した規約は曲線関係の推定にも同様に適用される。しかし、従属関係が曲線の場合にはさらに考察を要する。

第1に当てはめられた曲線における望ましい或は可能な複雑性の程度を常に考える必要がある。例えば、2.3.4.図では、当てはめた線は直線ではないとしても僅かに変曲した線であつて、これらのいづれにおいても最大或は最小点をもつ曲線は必要でない。

第2に考察することは当てはめた曲線の代数式を求める必要があるか否かである。もし必要であれば当てはめてからその方程式が簡単に求められる様な曲線を当てはめるべきである。方程式を求める必要がなければ前節の方法で当てはめを行う。

例として、１６ａ図に示す点は最大点をもつ曲線で当てはめるべきであると決めたとする。

第1段階として最大点の落ちる大略の位置を決める。この場合、この点は明らかに図の上部に近い二点の間に存在している。１８図の様に、曲線の略図を描く。

曲線を当てはめ、その方程式を決めるには予め色々な方程式に対応する

曲線の型を知つておらねばならない。したがつて曲線は試行錯誤法で適合せしめられる。例えば最もよく使用される方程式は二次式である。

$$Y = c + bX + aX^2$$

即ち

$$Y = e + a(X-d)^2$$

この曲線は点（ｄ，ｅ）において唯一の最大又は最小値をもつている。曲線はこの最大又は最小値を中心として左右対象である。即ち一方の側の上昇と同じ速度で他の側では下降している。

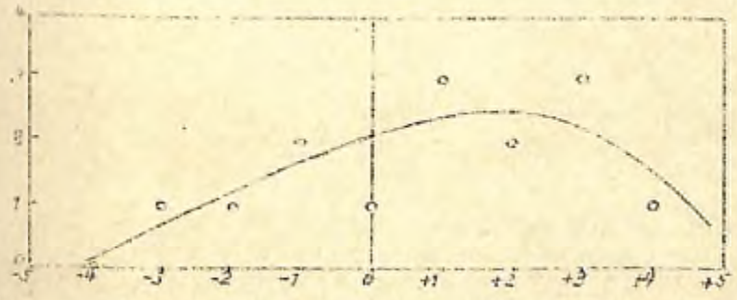
量ａはこの上昇又は下降の割合を決めａの大きい程曲線の変曲性は大である。ａの正の値は最小点をもつ曲線を表わし、負の値は最大点をもつ曲線を表はしている。１６ａ図の観測値に二次式を適合させたい。これは１９ａ図にも示してある。

第1段階として最大点の落ちる位置を推定する。

このために、概略の線を描く必要がある。１９ｂ図はこの様にして書いた線の例を示す。この場合最大点は（ 5.0　2.5 ）の近くにある。したがつて方程式は概略

$$Y = 2.5 + a(X - 5.0)^2$$

である。



１８図　回帰曲線のあてはめ

さらに曲線は（ 0.0 ）点又はその近くを通ると考えられるのでａの近似値は次式から求められる。

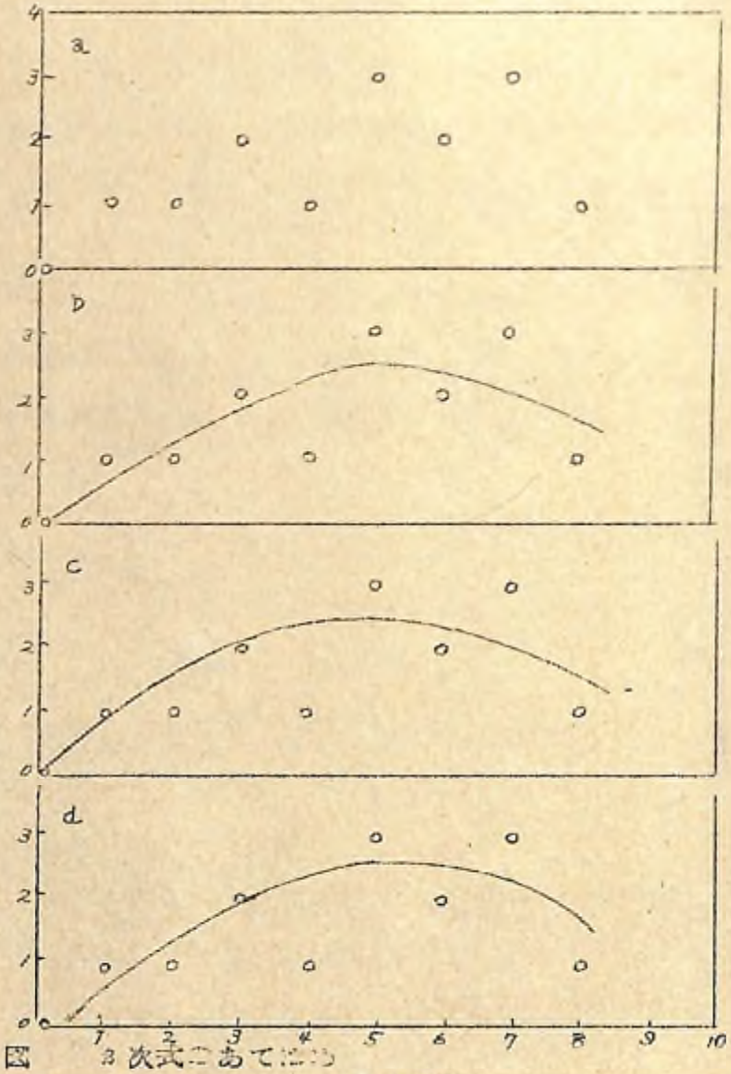
$$0 = 2.5 + a\,(\,0 - 5.0\,)^2$$

即ち

$$a = -1$$

したがつて曲線は近似的に次式で与えられる。

$$Y = 2.5 - 0.1\,(\,X - 5.0\,)^2$$



１０図　３次式にあてはめ

この曲線は１９０図に示してある。負の偏差が正の偏差より大きくなつているとはいえど、この曲線は良く当てはまつている。この場合、最大値を少しにずらして－右方へ 0.4 －簡単に修正される。したがつて方程式は

$$Y = 2.5 + 0.1\,(\,X - 5.4\,)^2$$

となる。

これは１９ａ図に示してある。

勿論上記の方法で行つた当てはめは、もちろん当てはめる曲線が複雑であれば、時間がかゝる。

当てはめる曲線の方程式およびその特性が分つていると当てはめは非常に楽になる。当てはめるべき曲線の型を決める場合に役立たせるため、曲線の特性ついての説明をつけて２０図にその例が示してある。この図を参考すれば、当てはめるべき曲線、当てはめ方を決めるのに役立つであらう。

2.4 多変量関係　Multiple dependences

多量関係の図による推定法は複雑であり、個々の従属関係を推定する必要がある。しかし、従属変量と個々の独立変量との関係を遂次推定して行くだけでは充分でない。というはこれは全体にわたる関係を示していないからである。これは独立変量が相互に連関しているためである。例えば７～１０才の小児は１年に平均５ポンドの割で体重が増し又、この年令の小児は身長の増加１cm当り１ポンドの割で体重が増加することが分つている。しかし、１年間に身長が６cmのびた小児は体重が５＋６＝１１ポンド増すと考えるのは正しくないであらう。この場合身長と年令には連関があるから、これらの２因子は体重に対して独立ではない。身長と年令を使つて体重を推定したいならば、新しい従属関係を計算すべきである。

一般に多変量関係を推定するには次に説明する二つの方法のいづれかで

a

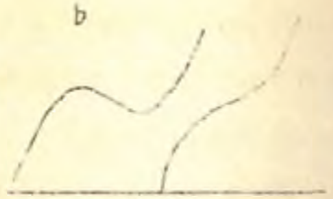
一般の二次式

$y = e + a(x-d)^2$

曲線は(d, e)に関して対象

aは曲線の勾配を示す

b



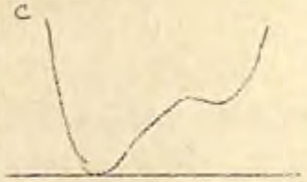
二次式

$y = e + a(x-d)^3 + bx$

曲線は変曲点(d, e)に関して対象

$b = (d, e)$における勾配

c



4次式

$y = e + a(x-a)^4 + b(x-d)^3 + C(x-d)$

d = 3つの最大又は最小横軸の値の平均

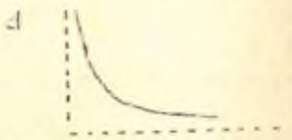
e = この横軸に対応する縦軸の値

C = (d, e)における勾配

b = 中心部で曲線が下降する割合

a = 両端部において曲線が下降する割合

d



双曲線

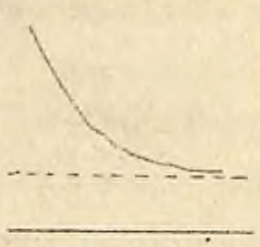
$y = ex + b/(x-a)$

$x=a$は垂直方向の漸近線

$y=e$は水平方向の漸近線

b = 漸近線から曲線までの距離

e



指数曲線

$y = e + d_2^{-x/e}$

$y=e$は水平方向の漸近線

曲線と漸近線間の距離はxのaだけの増加に対し半減する。

軸$x=0$における曲線と漸近線間の距離はdである。

f

対数曲線

$y = eb \log(x-a)$

$x=a$は垂直方向の漸近線

yは$x-a$が10倍される毎にebだけ減小する。

$x = a + e$の時 $y = e$

20図　曲線の性質

行う必要がある。

Yが従属変量、X_1、X_2、X_3が3つの独立変量を表わすものとし、次の型の方程式を当てはめたいとする。

$$Y = f_1(X_1) + f_2(X_2) + f_3(X_3)$$

$f_1(X_1)$, $f_2(X_2)$, $f_3(X_3)$ は夫々X_1, X_2, X_3 の函数である。*

*注　例えばX_1 の効果がX_2 等の値によつて変る時にはX_1X_2, $X_1^2X_2$, $X_1X_2X_3$ の様な積の形が必要なこともある。この方法については6.4節で説明することにする。

こゝに、例として3つの独立変量をとつたのであつて、3以上でも、又以下であつても同様に行えるのである。次の2つの方法のいづれか一法を用いて当てはめ行うことにする。オ1の方法は一般的に使用できるが、オ2の方法は多変量の線型関係の推定に限られている。

1.　まづ変量Yを一つの独立変量X_1と関係付け、従属関係を示す方程式

$$Y = \phi_1(X_1)$$

を推定する。

回帰線からの偏差$Y - \phi(X_1)$をX_2 と関係付け従属関係を示す方程式

$$Y - \phi_1(X_1) = \phi_2(X_2)$$

を推定する。

次に偏差$Y - \phi_1(X_1) - \phi_2(X_2)$を$X_2$ と関係付け、最後に、X_1, X_2, X_3 と関係のない量$Y - f_1(X_1) - f_2(X_2) - f_3(X_3)$ が得られるまで、X_1, X_2, X_3 等と順次関係付けた偏差を計算する。この様にすればYとX_1, X_2, X_3 との従属関係を表わす方程式

$$Y = f_1(X_1) + f_2(X_2) + f_3(X_3)$$

が得られる。

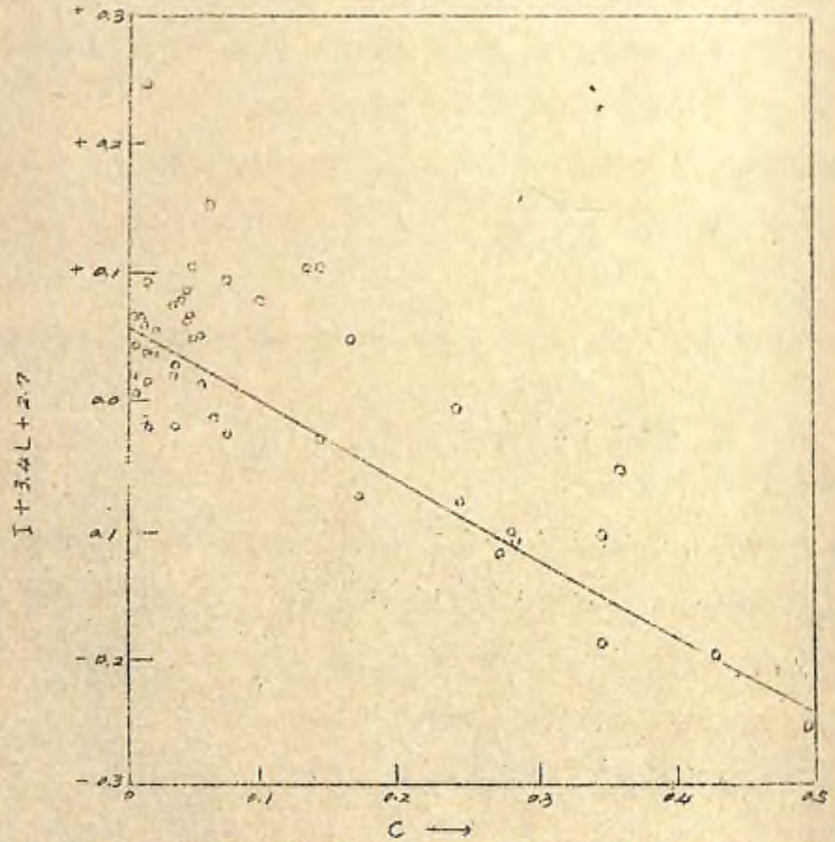
2. 変量Y,X2,X3を独立変量の一つX1と直線で関連させる。従属関係を表わす方程式は

$$Y = a_1X_1+b_1$$

$$X2= a_2X_1+b_2$$

$$X3= a_3X_1+b_2$$

偏差$Y-a_1X_1-b_1$,$X_3-a_3X_1-b_3$を$X_2-a_2X_1-b_2$と関連させ、回帰線からの偏差$Y-c_1X_1-d_1X_2-e_1$,$X_3-c_2X_1-d_2X_2-e_2$を推定する。X1,X2,X3に対するYと従属関係を表わす方程式を求めるため、この偏差の第1の組を第2の組と関連させる。



21図　Cに対してI－3.4 L＋2.7をプロットしたもの

この第1の方法は曲線で表わされる従属関係に順次近ずけてゆくものであり、第2の方法は線型関係を直接推定するものであることが分る。従属関係が線型でなければ、当然第1の方法を使うべきである。しかし、従属関係が線型であれば、どちらの方法も使える。一般に後者の方法は迅速であり、独立変量が相互に連関しておれば信頼度は高いものである。

この方法の応用例を示すため、11図に示した関係を考察しよう。（人頭割りの収入）の対数Iを（労働力人口の百分率）の対数Lと、黒人の百分率Cに対して次の線型方程式で関連させたいとする。

$$I = a_1L + a_2C + a_3$$

いづれの方法によつたとしても、第1段階として、IとLを関連させた方程式を推定する。これは2.2に述べた方法で11図から求まる方程式は近似的に

$$I = 3.4L - 2.7$$

となる。

第1の方法を使つたとすると、I－3.4 L＋2.7　をCに対してプロットする。このプロットした点が21図に示してある。この図から次の関係が推定される。

$$I - 3.4\ L + 2.7 = 0.06 - 0.6C$$

22図に示してある様にI－3.4 L＋0.6 C＋2.64　をLに対してプロットし、これから次の関係が求められる。

$$I - 3.4\ L + 0.6\ C + 2.64 = 1.37 - 0.85\ L$$

最後にCに対してI－2.55 L＋0.6 C＋1.27をプロットしたものは全然連関を示していない。即ちIと他の変量との推定された関係は

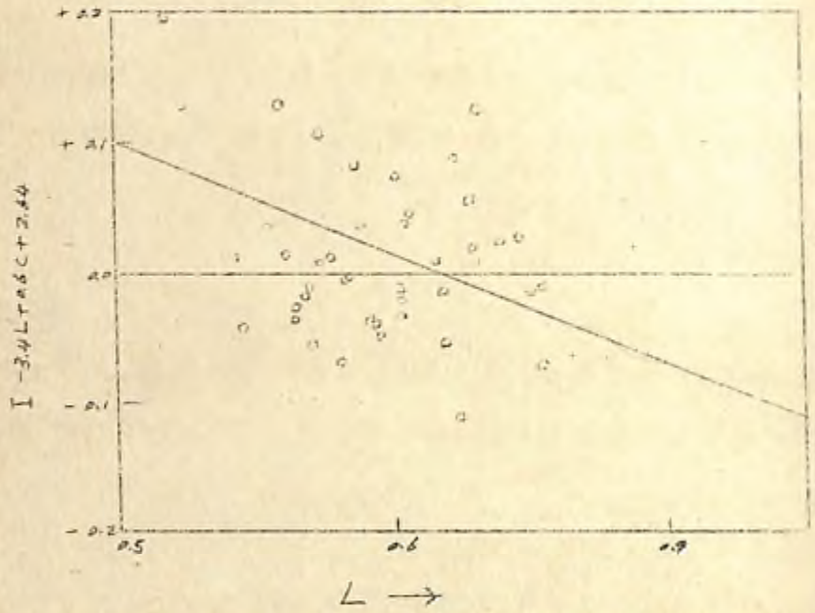
$$I = 2.55\ L - 0.6\ C - 1.27$$

変量の数が多く特にそれらが相互に密接な関連をもつていると、この全

過程を行うには非常に時間がかゝる、しかし、従属変量と独立変量間の関係に才1近似を使つて簡略化しうる場合もある。才1近似からの偏差を上記の方法で調べるのである。

上記の才2の方法が使われるなら、Lに対してcをプロットする必要がある。この様にすると、次の近似的従属関係が求まる。

$$c = 1.06 - 0.6\ L$$



22図　Lに対してI－3,4 L＋0,6 C＋2,6 4をプロットしたもの

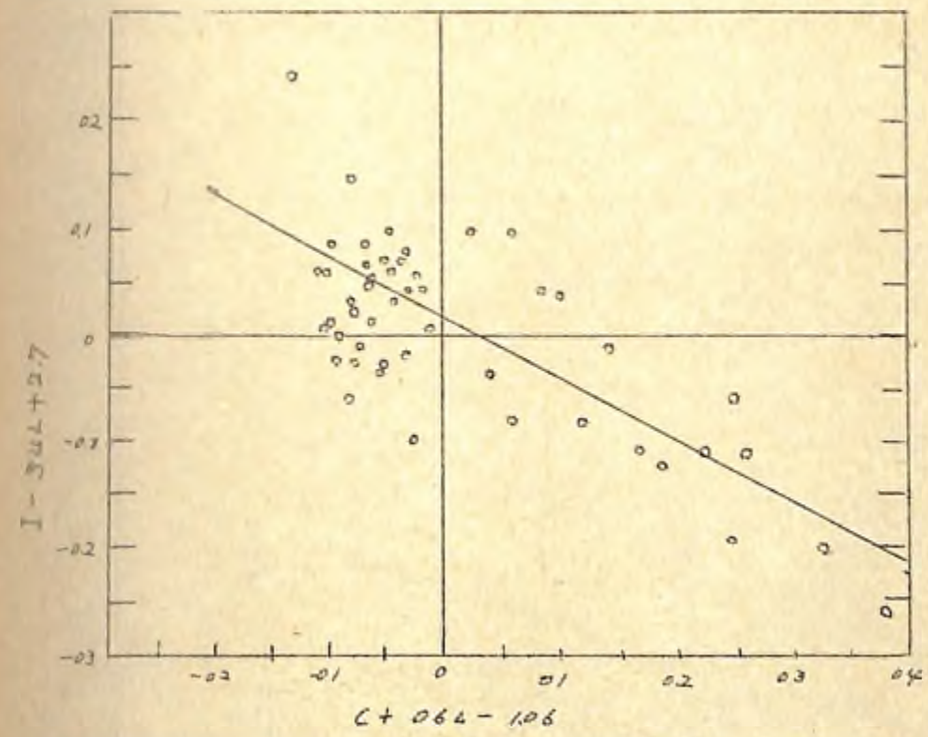
23図の様にI－3.4 L＋2.7　をc＋0.6 L－1.06に対してプロットする。これから次式が求められる。

$$I - 3.4\ L + 2.7 = 0.02 - 0.6\ (c + 0.6\ L - 1.06)$$

即ち

$$I = 3.04\ L - 0.6\ c - 2.044$$

Lの回帰係数に0.5の差はあるが、これは才1法で推定した関係とよく一致している。この場合、前者の方法の4段階に対して後者の方法は3段階でよい。不正確なものがあると精度の低い結果となるから、図による推定はいづれも出来るだけ正確でなければならないことが才2法の主な欠点である。才1法では、勿論不正確なものがあれば必然的に何段もの近似を行なわねはならないが、前段階における不正確さは次の段階で除去される。



23図　C＋0,6 L－1,06に対してI－3,4 L＋2,7をプロットしたもの

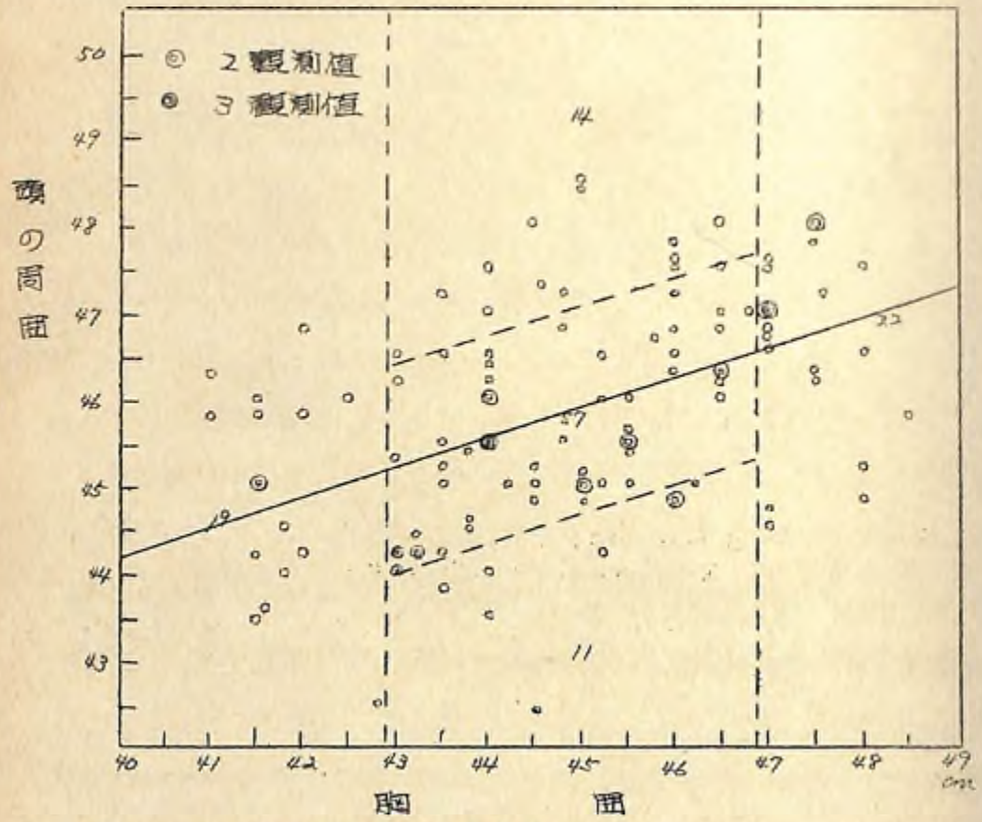
2.5. 誤差の推定　Estimation of error

数値調査における誤差の推定は観測値に含まれている説明できない変動を示すのに役立つ。

観測値数が多いとか、誤差の概略の推定値を必要とする場合を除いて数値解析を行うことが望ましい。それにもかゝわらず図による推定値の誤差

が推定できると便利なことが多い。この目的で次の〝７０％方式〟が示してある。

個々の観測値の誤差を推定するには独立変量に対して従属変量をプロットしたものが用いられ、多変量の場合には従属関係を推定する際使われた最後の図表が必要である。点の約７０％が中央部に入り、残りの点が他の２つの部分に略等分される様に、独立変量の目盛を三部に分割する。中央部にある点の７０％がその中に含まれる様に回帰線に平行で等距離にある線を描く。この線と推定回帰線との距離が個々の観測値の標準偏差の推定値である。



２４図　７０％方式を適用した場合

２４図はこの方式の使用法の例である。この場合、中央部には１２３点の中８２点即ち６７％がある。推定した回帰線に平行な線内にはこの８２点の中５７点即ち７０％が含まれている。この線は推定した回帰線から1.2cm離れている。したがつて標準偏差の推定値は 1.2cmである。

観測数の多い時には標準偏差を推定するには〝７／８方式〟を使うと便利である。この方式では点の中央郡の７／８を含む様に平行線を描く。平行線間の隔りは標準偏差の約３倍である。上例では７／８即ち７２点が間隔3.3～3.6cmの線間に含まれている。したがつて標準偏差は略 1.15cmである。

平行線の位置は錯誤法で決定すべきである。変曲した線では図上に近似的〝平行〟曲線を描くとよい。これに対する一寸した調整値は簡単に求まる。

この方法で求めた標準偏差は、普通偶然によるものと考えられる以上に観測値が離れているか否かを検定するために用いられる。事実、観測値が回帰線の周りに正規分布をしておれば、標準偏差の３倍以上線から隔つているのは、３００の内１より少く標準偏差の４倍以上隔つているのは 14,000 の内１以下である。したがつて標準偏差の３倍も回帰線から隔つている観測値は異常なものと考えられ、標準偏差の４倍以上も離つているものは極めて異常なものである。勿論これ以外の考察も適用されるが、これは異常観測値を棄却するための一つの統計的基準である。２４図では回帰線からの最大偏差は標準偏差の 2.5 倍であることに気付くであらう。しかし、４図で棄却した２点については、その偏差は標準偏差の 3.8 , 5.5 倍である。

2.6. 基本的な関係　Underlying relationships

回帰線は実際には二変量間の関係を表わしていないことを指摘した。別の変量を使つてある変量を予測しうる方法を示すことが回帰線の主目的なのである。回帰線が通常必要とされるのはある連関が決定された時別の変量の値を代入して変量の値を決定できるからである。しかし、一般法則を公式化するため測定値のもつている関係を推定したい場合がある。例えば、1図の様なデータでは大低の場合、与えられたCalanusの数に対するニシンの平均漁獲高の予測が要求される。これはCalanusの数が一定水準に達した時、期待しうる漁獲高を示すであらう。しかし、ニシンの漁獲高とCalanusの数との基本的関係を推定するのは、生物学的にも興味のあることであらう。一般にこの様な関係は予測には使えないが生物学自然法則を求める時に用いられる。

具合の悪いことには、さらに仮定を設けるか、求めんとするものを明確にしなければ基本的関係を推定することは殆んど不可能である。オ1の場合では、基本的関係を求める時には当然測定誤差、抽出誤差は除かねばならない。したがつて、例えばニシンの漁獲高とCalanusの数との関係を決める場合には、この観測値はいづれも標本に過ぎないということを考慮に入れておらねばならない。同様に4図の頭と胸の周囲の測定の際起る様な測定誤差も考慮する必要がある。しかし、これらの誤差が除かれても、やはり各測定値には変動性が存在し、それらの関係は、代数的関係でなくて統計的なものである。したがつて、観測値を図又は数値的に当てはめる前に統計的な分布型を指定する必要がある。

この問題は10章で扱うことにするが、こゝで概略の説明をすることは適切であらう。

最も関心をひき、広く取扱かわれる分布型は観測値、X，Yが次の形の関係で決定される時に起る。

$$X = f(t) + \varepsilon_1$$

$$Y = g(t) + \varepsilon_2$$

t，ε1，ε2 は独立な確率変量であつて、tと関係させることによりX，Yが結び付けられる連関を示す線はXとYを関連させた曲線を決定する次式であたえられる。

$$X = f(t)$$

$$Y = g(t)$$

この曲線の周りの散ばりはX，Yに含まれる附加的な独立な変動ε_1，ε_2によるものである。2変量間のこの様な関係を図で推定するには次の様な規約を用いて行なわれる。

1. 横および縦方向の点の散ばりを出来るだけ等しくする様にグラフの目盛を選ぶ。即ち変動ε_1，ε_2が比較出来る様に目盛を選ぶ。
2. 当てはめた線からの垂直偏差の合計は0としなければならない。
3. 線の各領域における垂直偏差の合計は0としなければならない。この領域は横軸、縦軸とは無関係に当てはめた線を垂直に切る線でグラフを分割して作る。
4. 推定する関係直線はそれからの垂直偏差を最小にする様に定め、特に大きな偏差が最小になる様に定める。

しかし、考察している変量の相対的変動は未知のことが多いから、1の内容には疑問があるということに気付かねばならない。したがつてε_1，ε_2の相対的大きさについて仮定を設けることが出来なければ基本的関係の推定で先に進むことは多くの場合不可能である。ε_1が0即ちXからYが決定されることが分つている場合には基本的関係はXに関するYの回帰から推定される。逆にε_2が0であることが分つていれば、Yに関するXの回帰は基本的関係を与える。しかし、この様な知識のないことが多く、変動

$\varepsilon_1, \varepsilon_2$ の相対的大きさは情報から推定しなければならない。

実例について云えば、本章および前章で使つている頭と胸の周囲の観測値は、さし当つては測定誤差はないものと仮定されている。この２つの観測値間の基本的関係を推定するには観測値相互の関連の仕方について多くのことを知らねばならない。これらはいづれも共通因子ｔ－〃成長〃といわれるーと直線的関係にあると仮定しうる。即ち

胸　　囲 $= C = a_1 t + b_1 + \varepsilon_1$

頭の周囲 $= H = a_2 t + b_2 + \varepsilon_2$

こゝで、$\varepsilon_1, \varepsilon_2$ なる量は一方の測定値にだけ影響をおよぼす未知の因子を表わしている。

基本的関係は次式で与えられる。

$$C = a_1 t + b_1 , \quad H = a_2 t + b_2$$

したがつて

$$\frac{C - b_1}{a_1} = \frac{H - b_1}{a_2}$$

この線ー直線ーを推定するにはＣにおける原因不明の変動とＨにおけるそれとの比率即ち ε_1 に帰因する変動と ε_2 に帰因するそれとの比率を知る必要がある。ε_1 が０であれば、胸囲は〃生長〃と完全に連関しており、胸囲に対する頭の周囲の回帰は、この２測定値間の基本的関係を表はす。逆に、ε_2 が０であれば頭の周囲に対する胸囲の回帰が必要である。この２つの極端な場合は、いづれも正しくなさそうである。真の関係はその間に存在するであらう。

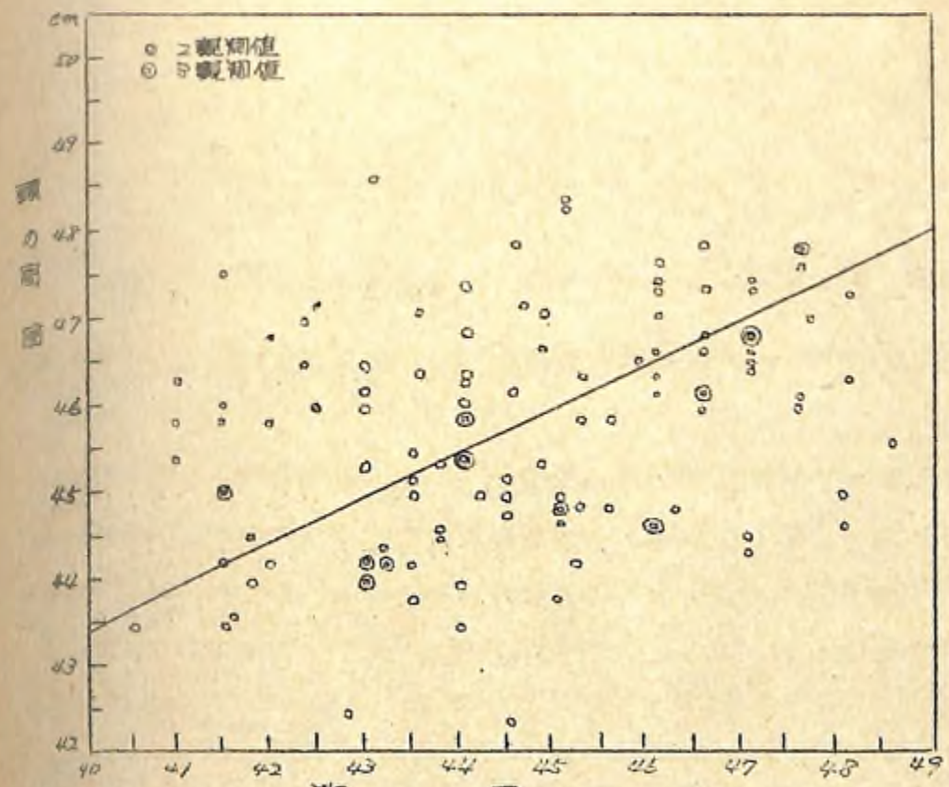
この２つの測定値に含まれる原因不明の変動が比較できると仮定すれば、上の規約の２，３，４が適用され、２５図の様に直線が当てはめられる。或は頭の周囲に含まれる原因不明の変動が胸囲のそれの半分に過ぎないと仮

定すれば、図で当てはめる前に、軸の目盛を½し違つた線が得られる。実際に胸囲の原因不明の変動と頭の周囲のそれとの比をＲで表わせば、Ｒの値毎に違つた線が得られる。これは２６図に示してある。

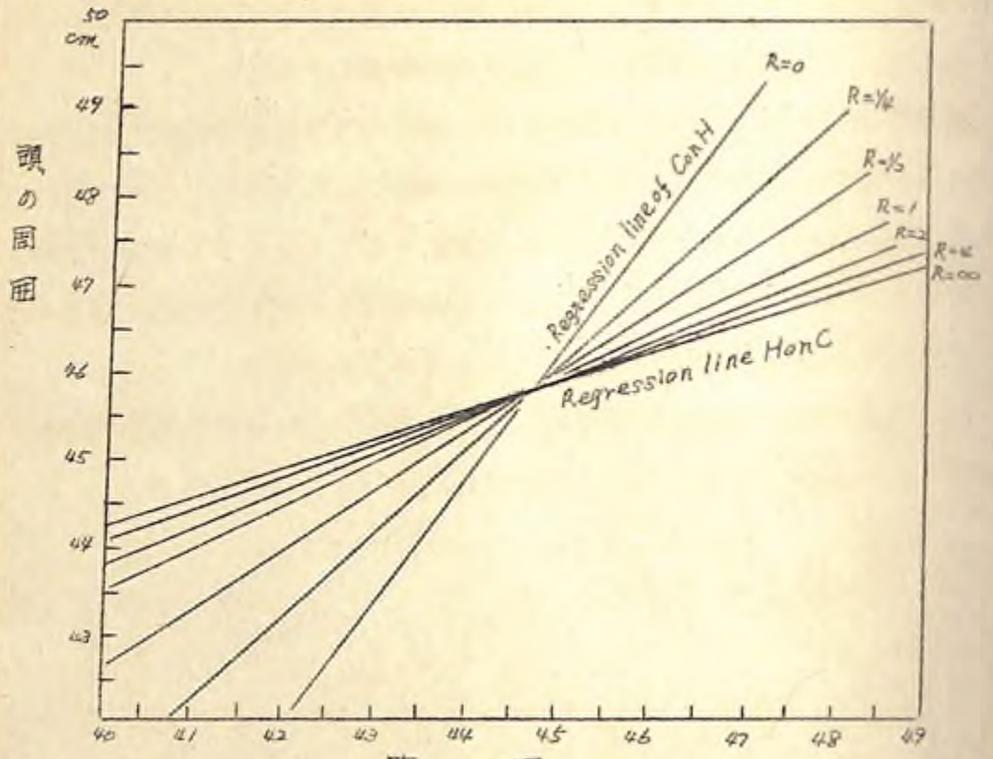
極端な場合として、Ｒ＝∞の時には、胸囲に対する頭の周囲の回帰線、Ｒ＝０の時には頭の周囲に対する胸囲の回帰線となる。

この場合にはＲは½と１の間にあると仮定してもさしつかえないから、その結果、真の関係はこの２線の間に存在する。しかし、これは仮定に過ぎず実際には別の情報から誤つていることが証明される。

$\varepsilon_1, \varepsilon_2$ は原因不明の変動を表わしているから測定誤差、抽出誤差はいづれもこれらの内に含まれていることに注目しなければならない。



２５図　誤差が等しいと仮定してあてはめた頭の周囲と胸囲との関係



26図　誤差の比をRとしてあてはめた頭の周囲と胸囲との関係

この種の誤差があつても時には基本的関係の推定にこれ以上の困難は起らない。

しかしながら、抽出誤差の影響を除外して、回帰線を推定したい時には相当慎重な考慮が必要がある。この場合には一方変量の原因不明の変動の合計と他の変量の抽出誤差の比の推定が必要である。測定値に対する適切な目盛を決めるため、この比が使われ、前に説明した方法で線を推定するのである。

3　図による検定　Graphical Testing

3.1.　検定の目的　Purposes of testing

2つ以上の変量間にどの様な連関があるかを決める場合にグラフが使えることを指摘した。普通、観測値をグラフ上にプロットすることにより、即座にこの問題に対する解答の得られることもあるが、時には明確な解答をグラフが与えないこともある。外観に表われた連関が偶然によるものか否かを決めるため検定を行つてみる必要がある。

真の連関をみつける機会を最大にするには完全な数値解析および検定を行う必要がある。

しかし、グラフ上の点を数えることによつて直接検定することも可能である。この検定法は数値による検定法程鋭敏ではない。したがつて、この方法で有効な連関をみつけそこなつたとしても、この失敗は、これ以上解析しても有効連関は検出出来ないということを意味しているのではなく、点勘定による検定を使つた有効連関の検出は屢々これ以上解析を必要としないことを意味している。この様な検定は簡易であるから、完全な解析を始める前にやつてみる価値のあることが多い。図形による検定方法を説明することが本章の目的である。

連関の検定方法は他の統計的検定に使われるものと同じである。まず帰無仮設が設けられる。この場合には観測値間には連関はないということである。この様にして求めたものと同じ位極端な観測値を得る確率は、帰無仮設を真として求められる。この確率が非常に小さければ、帰無仮設は棄却される。この場合には有意な連関が在ると結論してもよいであらう。

有意な連関のあることを明言する時には通常その確率の値を使う。即ち次の様な説明、〃この2組の測定値には1%水準で有意な連関がある〃

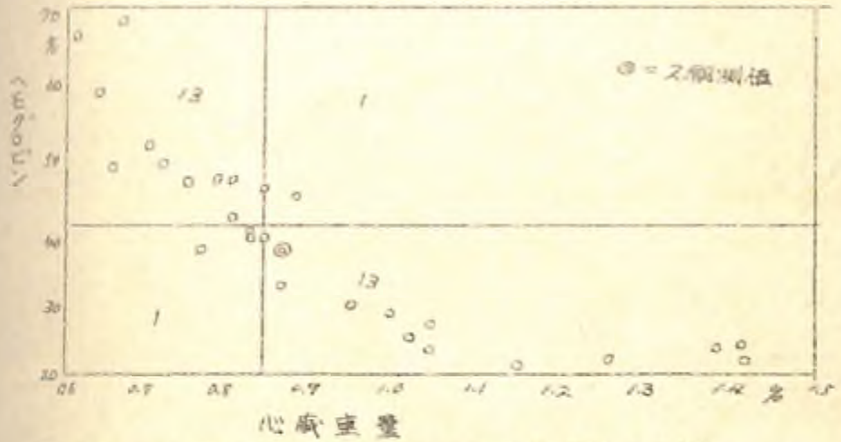
"この2組の測定値には有意な連関がある（P＜0.0.1）"は、この様に極端な観測値の組は１００回に１回以下しか起らないことを示している。この実際のやり方は本書に載せてある。

3.2. 中央値による検定 Medial test

medial testの基となる考え方は極めて単純である。まずグラフを等数の点を含む2つの部分に垂直線でもつて二分する。これを垂直中央線vertical medial lineと呼ぶことにする。点の数が奇数であれば、この線は次の段階では勘定に入れない一点を通る。グラフを4象限に分割する様に水平中央線horizontal medial lineを描く。各象限の点数を数え上げ、それが著しく多いか或は少ないならば連関があると推定する。

附録の１表を使つてどの象限の点数が過大或は過小であるかを決める。例としてグラフ上の２５点の内、一つの象限に入る点は３点であるとする１表によればこの様に極端な値が得られる確率は0.05以下であることが分る。したがつてこの測定値には連関のあることを強く示している。真く同様に一つの象限に入る点が２５点の内2点以下であれば、有意な連関がある（P＜0.01）と結論する。

２７図は２b図の観測値にこの検定法は適用したものを表わしている。全体で２９の観測値がある。したがつて垂直中央線はその中の１つを通る。各象限は１～１３点を含んでおり、１表によればこれらがP＜0.01の値（夫々3・11）を越えていることが分る。したがつてこれらの観測値間には有意な連関がある（P＜0.01）と結論する。対角線方向の象限の点数は等しく、一方の象限に落ちる点数によつて他の象限に落ちる点数が決まつてしまうから、象限以外に落ちる点数を数えたとしても何も得られない



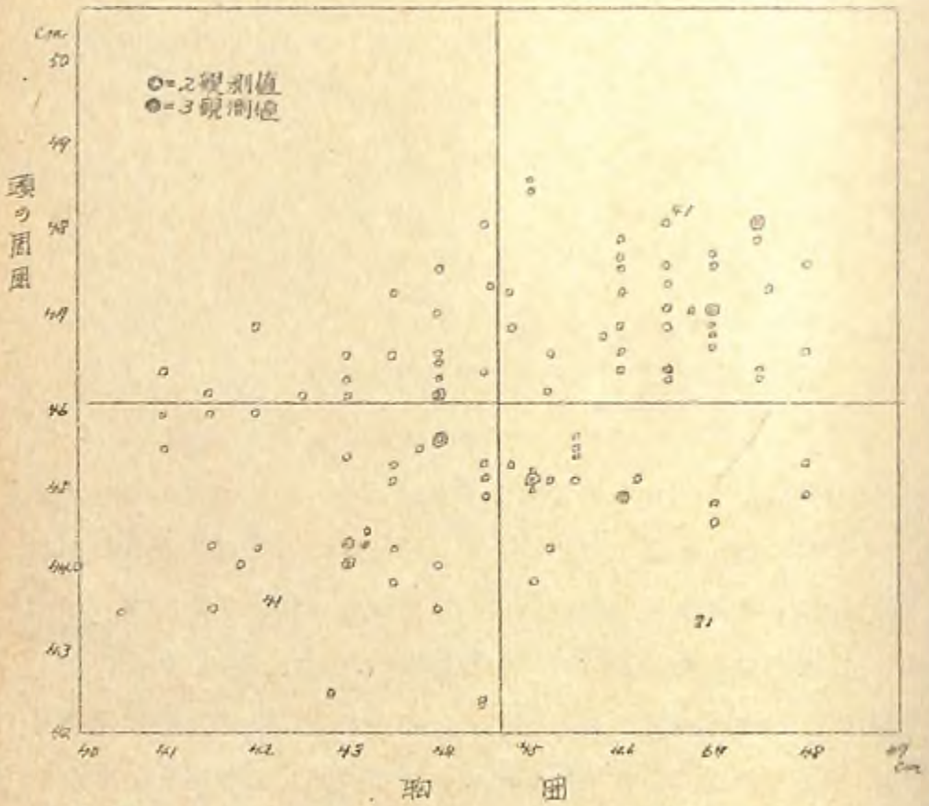
２７図　2.b図にmedial testを適用した場合

ことに注意せよ。しかし、何個かの観測値が同じ値をとる場合には、中央線が一個以上の観測値を通ることもありうる。したがつて上記の検定法を適用するには、観測値を出来るだけ等しい組に分割する様に線を描く必要がある。線上に落ちる点は上記の検定法を適用する時には無視し、１表には、２つの対角象限に落ちる点数の平均を使うべきであらう。

中央線が何個かの点を通る場合の検定の仕方を示すため４図の観測値を再び考察してみよう。この場合には観測値を２つの組に正確に分割する垂直線は描けない。２８図に示してある線は観測値を６１と６２の組に分割している。又水平線を引いて６１と６２の組に観測値を分けることもできる。したがつて、この場合には、各象限は夫々20,41,41,21点を含んでいる。表１では２０½，４０を使わねばならない。

この図には総計１２３の点がある。１２０－１２１点に対して、任意の象限に落ちる点が２４以下になるのは１００回に１回以下しか起らないこ

とを1表は示している。したがつて総計１２３点の中任意の象限に落ちる点が２０又は２１以下となることは１００回に１回以下である。胸囲と頭の周囲とには有意な連関（Ｐ＜0.０１）があると結論される。

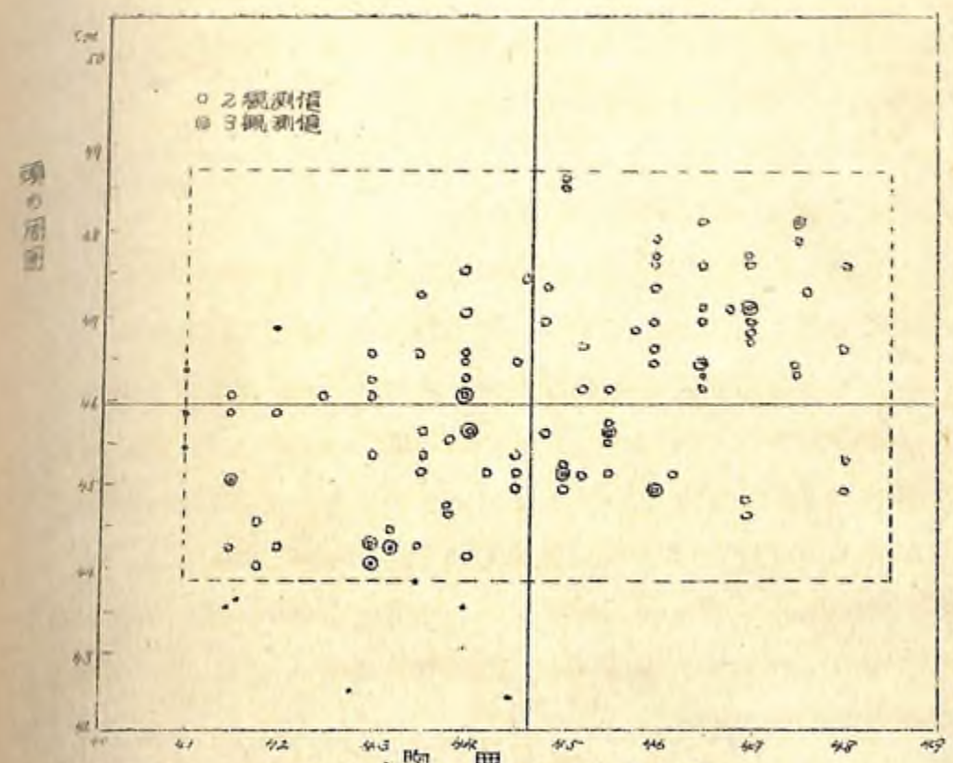


２８図　４図の関係にmedial testを適用した場合

3.3. Tukey のコーナーテスト Tukey's corner test

Tukey's corner test は応用面では前説でのべた中央値による検定法と同じである。しかし、最も極端な観測値だけを考えるという点が違つている。その方法は次の通りである。

前説と同じ方法で散布図表を４象限に分割する。正号が右上と左下の象限に、負号が右下と左上の象限に附けてあるかの様に、各象を考察する。図の上部から始め観測数を数え上げる。この時には垂直中央線越してはならない。適当な符号をつけて、この数を記載する。これを図の右から左、左から右、下から上へと繰返す。この様にして得られた４つの数の代数和は Tukeyにより象限和と名付けられた。これは２表を使つて検定される。例えば、象和が１１となるか、これを越える確率は0.０５以下であり、１５以上となるのは１００回に１回以下である。以上のことから分るように、この検定法の特長は、事業上観測数に独立であること即ち全体にわた



２９図　４図にTukeyo corner test

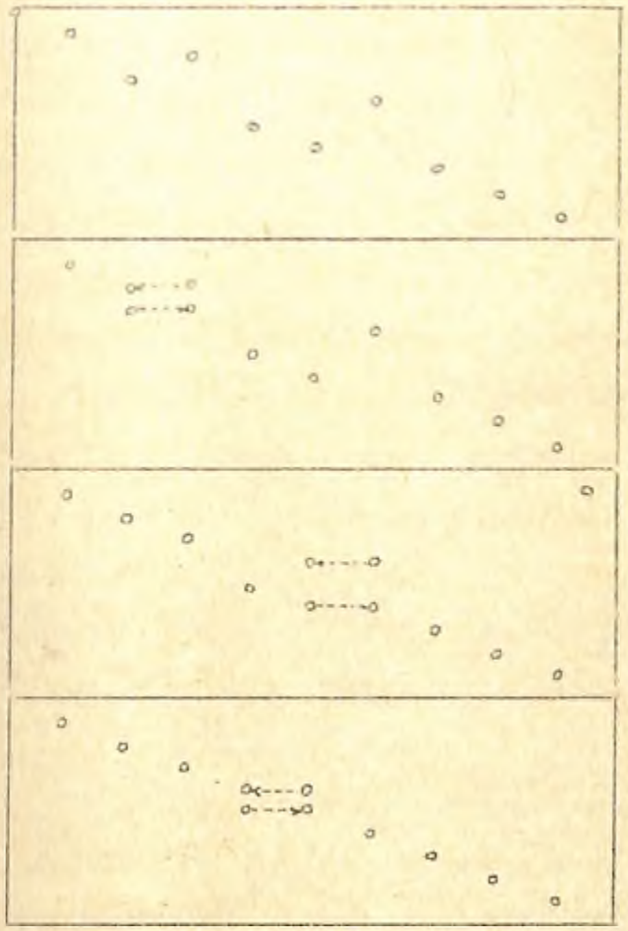
つて観測数を数える必要のないことである。

応用例として、4図のデータを考察しよう。これは29図に示してある。まず上から下に動く時には負の点に出合う前にたゞ1ケの正の点に出合う。同様に右から左へ動けば負の点に出合う前にたゞ1ケの正の点に出合う。下から上に動けば、1ケ正の点と1ケ負の点が同時に出合う前に6ケの正の点に出合う。この場合、点数は＋6½である。最後に2ケの正の点と1ケの負の点が同時に出合う前に2ケの正の点に出合う。したがつてこの場合の計算は2＋（2×½）＝3である。象限和は1＋1＋6½＋3＝11½であり、連関（P＜0.05）のあることがこの検定から結論される。

この例のよう数個の点に同時に達した場合は1／（1＋達した負の点の数）の得点を、到達したの各点に与えるべきである。ここでは到達した負の点はいづれも1ケだけであるから、各正の点には½の得点が与えられる。例えば、2ケの正の点と2ケの負の点に同時に達したとすれば、各正の点には¼の得点が与えられるであらう。

3.4. 順位による検定　Ordering test

順位による検定法は観測数が少く、それらに多小なりとも順序がある時にも最も有効である。この方法を使つた例は既でに1.4節に示してある。この時には、2つの5点からなる線のいづれにおいても、一様に上昇或は下降することが観測される機会は30回に1回以下であると述べた。したがつて5点以上の系列で一様に上昇又は下降することが観測されゝば有意な連関（P＜0.05）が存在する。上昇又は下降が中断しておれば、有意性を確実にするためより多くの観測値をとらねばならない。この場合、観測値が順序に配列していない程度は、完全な順列をうるため、隣接横座標点とのの交換の最小数で測る。



30図　完全な順序を得るに要する交換数の推定

30図は交換数を推定する方法の例である。この場合には、3回の交換が必要である。各点の左側でその点より上にある（連関線の傾向が上向であれば下位）にある点数を数えればこの値は直ちに求まる。これらはそれから加え上げられる。この場合夫々1回するものと2回するものがあつて、

その結果交換数が前のように3回となる。

観測された傾向の有意性を検定するには、交換数を計算し、この数が有意に小さいか、否かを調べるためⅢ表を使う必要がある。したがつて、10点の内6回交換したとすると、Ⅲ表によれば、8回以下の交換は１００回に1回以下しか起らないことが分る。よつて有意な連関（Ｐ＜0.０１）があると結論される。二つ以上の点が同じ縦座標又は横座標をもつ時には上記の方法は若干の修正が必要である。この時には同じ横座標をもつ点の平均を使えばよい。同一縦座標をもつ上の点に対しては推定交換数の推定には½を加える必要がある。同様に同一縦座標をもつ、3或は4点があれば、推定数に⅔又は2を加える。この様な補正値は近似値に過ぎない。

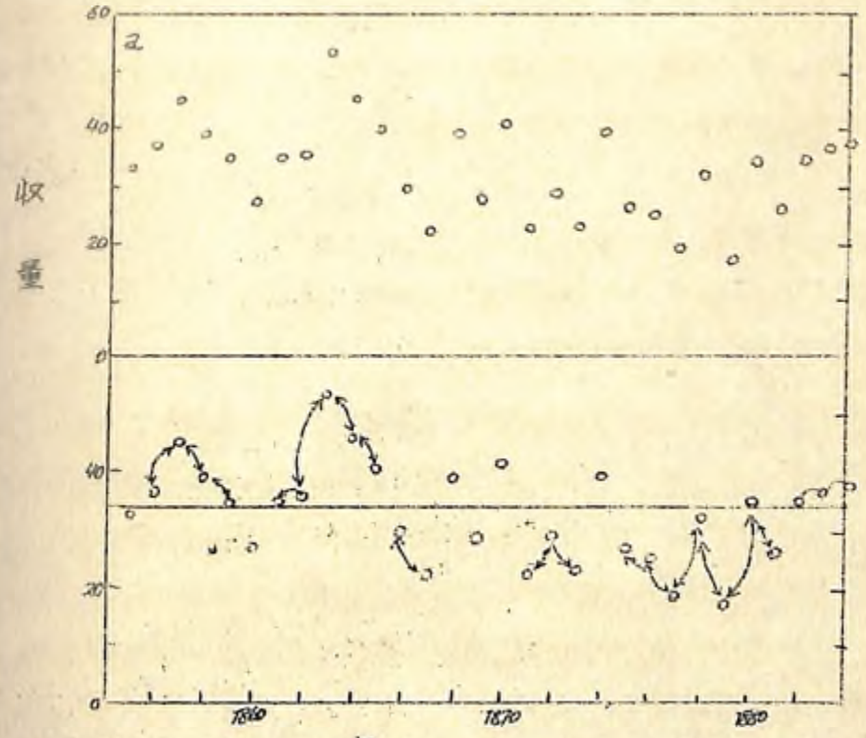
一般に、同一横座標は縦座標をもつ点が二つあることが殆んどない様に観測数が少い時には、順位検定は最も有効である。観測数が多いか、同じ値の観測値を生ずる傾向があれば、普通、1.4節で述べた様に、比較的に数の少い信頼度の高い組とするため観測値をまとめることが望ましい。

3.5. ポイントペヤー法　Point－pair test

二変量間の連関が、一方の変量が上昇するのに対し他方が下降するとか、両変量が共に上昇又は下降する傾向のものである場合に、前の3節で説明した検定法は最も有効なものである。実際に、直線的連関はこれらの方法で簡単に検定できるが、２ｂ図の様な曲線的連関も同様にして検定できる。観測値の大部分が反対の象限にある様な連関の形をとることが、これらの検定法を効果的にするため最も必要なことである。

しかし、連関の形が非常に曲りくねつたものであれば、これらの検定法は充分なものではなく別の検定法を探求せねばならない。最も有効な方法は Point pair test である。

Point pair test は、一方の変量は順序に配列しており、他方の変量はそれと非線型的関係にある様な時に最も有効である。特に、時系列は継時観測値間の一般的傾向即ち連関について検定できるのである。３１図はばつきりした傾向のある時系列の例である。Rothamsted の Broad-balk 小麦試験地の７ｂ号標準地の肥料（アンモニヤ塩）を施した穀類の収穫量が１８５５－８４年にわたつて示してある。この場合、収穫量は明らかに最初は減少し続いて僅か増加している。この様な傾向の有意性を検定するため Point－pair検定法が使われる。



３１図　１８５５－８４年における収量

この検定法の第一段階は水平中央線を描くことである。この線上にある点又は点群を除き、この線の同じ側にある継時観測値の各点を point－pair と呼ぶ。point－pair の数を数え、Ⅳ表を使つて連関の有意性を判定する。例えば、４０個の観測値がとられ、２８個の point－pair が数えられたとする。Ⅳ表によれば連関のないことが１００回に１回以下であれば、２７以上の point－pair の生ずることが示されている。したがつて有意な連関（ P＜0.０１ ）があると結論する。

３１ｂ図は、３１ａ図の観測値にこの検定法を適用したものである。図に示してある通り３０観測値に対して１８個の point－pair がある。Ⅳ表によれば、5％水準で有意であるためには１９個の point－pain が必要である。したがつて、この検定の結果、傾向は示されてはいるが、有意ではない。この傾向が偶然の結果であるか、否かを検定するには、より広範囲にわたる解析が必要である。

3.6. 中央相関係数、順位相関係数、逐次相関係数
Coefficient of medial, rank and sequential correlation

前説で説明した検定法は２組の観測値間にどの様な連関があるかを推定するために用いられる。しかし、この検定で得られた有意水準は連関の程度を示すのではなくて、その存在のたしからしさを示に過ぎないのである。連関の程度を測るにはこの検定から相関係数を求める必要がある。３つの相関係数－中央相関係数、順位相関係数、逐次相関係数－が使われる。これらは次の様に定義されている。

$$\text{中央相関係数} = \frac{2\,(\text{右上と左下の象限の点の総数})}{N} - 1$$

$$\text{順位相関係数} = 1 - \frac{2\,(\text{最小交換数})}{\frac{1}{2}N(N-1)}$$

$$\text{逐次相関係数} = \frac{2\,(\text{point－pair 数})}{N-2} - 1$$

こゝでNは中央線上に落ちない観測数を表わす。一般に点を正確に等しい二組に分ける様に中央線が描けたならば、右上と左下の象限の点数は等しいであらう。この様な時には中央相関係数は

$$\text{中央相関係数} = \frac{4\,(\text{右上象限の点数})}{\text{点の総数}} - 1$$

中央相関係数、順位相関係数、逐次相関係数を夫々 ρ, ζ, ψ で表わすことにする。どの係数でも、とりうる最大の値は１である。両変量が同時に増加する完全な連関がある時にはこれらは全て値１となる。これらの係数のこの外の特性については 3.6ａ表に示してある。

前節の検定法でみた通り、逐次相関係は曲線的関係に対して最も有効である。曲線関係以外については、中央相関係数又は順位相関係数を用いねばならない。

3.6ａ表一　相関係数　の特性

関係の形態	相関係数 中央	順位	係数
完全、両変量同時に増加	＋1	＋1	＋1
完全、一変量の減小に対し他変量は増加	－1	－1	＋1
無相関観測値	0	0	0
完全、たゞし曲線	小	小	1に近い

したがつてこれらの係数のいづれでも、大きな値は2変量間に良好な連関のあることを示している。しかし、変量間の高い連関は1に近い係数を与えるが、その逆は必ずしも正しくはない。これは特に中央相関係数と逐次相関係数については、勘定した点数だけに依存しており中央線に関係のない観測値の大きさによらないからである。順位相関係数は観測値の順位だけで、その大きさを考慮していないが完全な連関からの隔りに対しては最も鋭敏である。

関係の方向を示す順位相関係数には普通符号がつけてある。正号は両変量が同時に増加する関係を示し、負号は一方変量が減少する時他の変量が増加する関係を示している。これらの値はいづれも＋1と－1との間の値をとるに過きないことを意味するからこのことは便利な仕方である。

例として、これらの係数を31図に示す収穫量のデータについて計算してみる。垂直中央線を描けば、右上象限には5個点が落ちる。

したがつて

$$\phi = \frac{4(5)}{30} - 1 = -0.33$$

を求めるには、まず交換数を推定する必要がある。各観測値の左側でその観測値より小さい観測値数を数え上げれば、

交換数＝1＋2＋2＋1＋0＋3＋4＋8＋8＋7＋1＋0＋9＋2
＋12＋1＋4＋2＋14＋3＋3＋0＋10＋0＋13
＋6＋16＋19＋21
＝172

$$= -\left(1 - \frac{2 \times 172}{\frac{1}{2} \times 30 \times 29}\right)$$

＝－0.21

最後に逐次相関係数は次の様にして求められる。

$$4 = \frac{2 \times 18}{28} - 1$$

＝0.29

これらの係数が3つ共小さいという事は、時間と標準地収量との傾向は余り大きくないことを示している。ψ は ϕ, ζ と略同じ大きさであるから、明らかに、この関係は曲線的なものではない。これらの係数の有意性を検定する場合、Ⅰ，Ⅲ，Ⅳ表の使用を避けるため、別の表が直接この有意性を検定するために調製された。

附録のⅤ表は、5％および1％有意水準の中央相関係数を示している。例えば、Ⅴ表によれば、観測数30では5％水準で有意であるためには、中央相関係数は絶対値で0.425を越える必要がある。したがつて上記の係数－0.33は有意ではない。順位相関係数および逐次相関係数の値を検定するため同様な表が調製されるであらうが、これらは余り使用されておらず、したがつてこゝには示してない。

3.7. 重中央相関係数と偏中央相関係数

Multiple and partial medial correlation corffi-cients.

これまでに考察してきた検定法は、2変量間の連関を取扱う場合限られている。2変量以上の場合にはこれらの検定法を拡張しなければならない。これを前節で述べた、中央相関係数を用いて行なつてみよう。2つの変量X，Y間の連関を第三の変量に帰因させうるか、即ち別の形で云えば、ZによるXの予測はYも考慮することにより改良できるかという問題によく出合う。第3の変量Zを考慮した時の二変量X，Y間の連関の程度を測る

ための偏相関係数を計算してみればこの問題の解答が得られる。例として $\phi_{xy}, \phi_{xz}, \phi_{yz}$ は夫々 x,y, x,z, y,z 間の中央相関係数を表わすものとする。Zの効果を除いた時のx，y間の相関を測る中央相関係数は次式から求められる。

$$\phi_{xy\cdot z} = \frac{\phi_{xy} - \phi_{xy}\phi_{yz}}{\sqrt{(1-\phi_{xz}^2)\phi_{yz}^2}}$$

これは多変量関係における個々の変量の効果を決めるために使われる。

今度は4，9図のデータを頭の周囲H、胸囲C、体重Wが同時に、推定される方法を推論するために使うことにしよう。これを行うには、散布図毎に、中央線を描き、線上に落ちる点を全部除去しなければならない。この場合には線上に落ちる点はなく、したがつて中央相関係数を計算する時には総計123個の点を使う。

$$\phi_{HC} = \frac{2(82)}{123} - 1 = 0.333$$

$$\phi_{HW} = \frac{2(84)}{123} - 1 = 0.366$$

$$\phi_{CW} = \frac{2(108)}{123} - 1 = 0.675$$

この係数は4，9図で示されたこと、即ち胸囲と体重とは最も密接な関係があり、頭の周と他のものとの相関は、胸囲との相関より、体重との相関の方が幾分大きいが、略同じであることを裏書きしている。胸囲と同様に体重を考慮することにより、頭の周囲の予測がどの様に改良されるかを決めるために、偏相関係数 $\phi_{HW\cdot C}$ を計算した。これは

$$\frac{0.366 - 0.333 \times 0.675}{\sqrt{(1-(0.333)^2(0.675)^2}} = 0.149$$

非常に大きいわけではないが体重を考慮することにより幾分改良されることとを、この値は示している。

この値の有意性を検定するには、中央相関係数の表－V表を使う必要がある。この時には観測値の対数より1つ少い数を使う。したがつて、この場合使う観測数は122である。V表によれば、5%水準で有意であるためには約0.19という値が必要である。即ち胸囲から頭の周囲を予測する方程式に体重を入れても有意な改良とはならないことを示している。

しかし上記の偏相関係数は大きなものであり、完全な解析を行えば、体重を入れた時には、有意な改良がなされるであらう。

この逆も考察してみる必要がある。即ちすでに体重が考慮されている場合の頭の周囲の予測を胸囲はどの様な改良するか。

これを検定するには $\phi_{HC\cdot W}$ を計算する。即ち

$$\frac{0.333 - 0.366 \times 0.675}{\sqrt{(1-(0.366)^2(0.675)^2}} = 0.089$$

これは前の値より小さい、したがつて胸囲は体重が考慮された場合の頭の周囲の予測に大した貢献はしないことが分る。

次の様なオ2の検定が時には必要である。ある変量Xは有意な範囲で2つ以上の変量YZ …… に同時にと関係するか否か、多くの場合これは重中央相関係数を使つて計ることができる。

3変量に対しては、$\phi_{X\cdot YZ}$ は次の様に定義する。

$$\phi_{X\cdot YZ} = [\phi_{XY} \quad \phi_{XZ}] \begin{bmatrix} 1 & \phi_{YZ} \\ \phi_{YZ} & 1 \end{bmatrix}^{-1} \begin{bmatrix} \phi_{XY} \\ \phi_{XZ} \end{bmatrix}$$

$$= \frac{\phi_{XY}^2 + \phi_{XZ}^2 - 2\phi_{YZ}\phi_{XY}\phi_{XZ}}{1 - \phi_{YZ}^2}$$

こゝで［］は普通のマトリックス記号である。

この数の大きさはY，Zに関するXの依存の程度を示すものである。その有意性はVI表を使つて検定する。

例えば、胸囲と体重に対する頭の周囲の重中央相関係数は

$$\sqrt{\left\{\frac{0.333^2+0.366^2-2\times0.675\times0.333\times0.366}{1-0.675^2}\right\}}=0.384$$

VI表によればこれは1%水準（0.30）で有意である。この係数の値とφ_{HC}，φ_{HW}とを比較してみると、体重と胸囲を使つたとしても得られるものは極く僅であることをこの場合にも示している。

同様に数ケの変量の偏中央相関係数、重中央相関係数を定義することも可能である。しかし変量の数が増すと共に複雑となるから、3変量以上については、簡単な図形的解析法を使うか、完全な数値解析に依存することが望ましい。

数値解析

Numerical Analysis

4 直線関係 Linear Association

4.1. 数値解析における仮定 Assumption in numerical analysis

2つ以上の変量間の連関の数値解析を行うには、色々な仮定に基づいて予備的解析を行つてみなければならない。これらの仮定は慎重に吟味し、次いで検定されるべきであらうが、この様な仮定が常に存在するものであることを知つておらねばならない。一般に、これらの仮定は合理的なものであり、観測値と予盾しないことが示されるかもしれないが、一度たてた仮定を証明することは不可能である。この様な理由から、解析を始める前に、たてられる仮定について若干の注意を与えておくのが至当と考えられる。そうしないと目的もなしに、長たらしい意味のない解析をすることになる。

普く一般に、連関を推定する時には3つの仮定がなされる。即ち

1. 観測値はそれらがとられた母体の状態を代表している。
2. 観測値は互に独立である。
3. 回帰線からの偏差は同じ分散で正規分布している。

このことは、たてられた仮定だけではなく、一般に解析から得られた結論の妥当性に関係している。これらの仮定について順次考察してみよう。

オ1の仮定は普通、科学的調査の必要条件であり、図形的研究のいづれにも適用される。しかし、連関を調べる際には、推定された関係の範囲を誤つて解釈し、それが実際に持つている以上に広い一般性があるものとして使用しがちであることに注意せねばならない。例えば1930年のスコットランド北東部の1才児の胸囲と頭の周囲との関係に関する調査は、

この時期におけるこの地方の、この年令の小児に適用出来る結果を与えるに過ぎない。その場合でも、観測値がこの地方の小児を代表するものでない限り、この結果は使用できない。しかし、この観測値は代表的なものと仮定しても差支えのない、予防策を講ずれば、その結果は同様にその地方時期、年令の代表であると考えてもよい、もちろん、この関係を別の地方或は時期に適用するか否かを決めるには、これらの条件で同じ結果が得られると仮定して差支えないかどうかによる。調査の分野で制約があるために、経済的変量間の関係の解釈に困難をきたすことが多い。通常、関係は限られた期間でとられた観測値の間で推定しなければならない。この結果その期間内に起つてることを正しく代表するが、この関係が変らないと仮定されない限り、他の時期に起ると思われることについては、何等の説明も与えないであらう。解析の際関係のある変量を全て考慮に入れゝば、この仮定は合理的なものとなる。しかし、重要な変量を落したり、効果があるといつて新しい変量が影響を生じ始めた場合は、この関係の全体の型は、その期間の後では変つてしまう。したがつて経済的変量の調査で一番考えねばならないことは、推定された関係が、他の場合にも適切なものであると仮定できる様に、出来るだけ多くの関係のある変量を考慮に入れることである。上記の才2の仮定ー観測値は独立である。ーも数値解析では一般的のものである。観測値が独立でなければ、その散ばりは推定された関係の信頼性に誤まつた印象を与え、さらに重要なのは、その関係の偏りのある推定値がえられるであろう。したがつて大低の解析ではこの様な仮定が設けられ、その否定は恐らく誤まつた結論を導びくであらうということを知つておらねばならない。

一般に観測値は互に独立であると仮定しても差支えないが、場合によつてはこのことは正しくないであらう。例えば、ある関係を推定するのに、

いくつかの方面から得られた観測値を使う時には、同じ処から求めた観測値はかたまる傾向がおうおうにしてみられる。この様な場合には資料源間の変動性は同一資料源からとられた観測値間の変動性よりも大きな影響を推定された関係の精度におよぼす。この結果、関係の精度が観測値の散ばりから計られるならば、推定された関係は見かけは現実のものより正確にるであらう。各資料源から数の違う観測値がとられるとさらに困難となる。観測値全体がその状態を代表しているかどうかを考察してみるか、もしそうでなければ後の推定で各資料源がそれにふさわしいように表わされている様にしておく必要がある。

観測値の非独立性は経済的データでは極く普通である。時系列ではある観測値に関する知識は次の観測値の大略の予測値となるから、継時観測値が独立と見做されることはほとんどあり得ない。一般に各観測値毎に、前後の観測値間にある傾向があれば連続した観測値は従属関係があると云える。

この様な理由から、時系列における観測値の解析には、１１章で考察される様な形の特殊な方法が必要である。

回帰線からの偏差は正規分布しているという才３の仮定は、原因不明の変動は多数の独立な原因に帰因するものであり、その一つ一つがその結果として生じた偏差の小部分を構成していると仮定することと同じである。このことは単に偏差が正規分布する原因となるばかりでなく、この仮定を考察する最も簡単な方法である。正規性の概念はさらに次の様な仮定を含んでいる、大きな正の偏差は大きな負の偏差と同じ頻度で起り、小さい偏差は大きな偏差より数多く起る。したがつて偏差の相対的頻度は、３２図に示してある様な正規分布に従う。

観測値をプロットすることにより、偏差が同じ分散をもつて正規分布し

ているか否かを計ることが出来る。例えば、3図の偏差は明らかに、図のあらゆる部分で同じ変動性を示してはいない、又大きな負の偏差は大きな正の偏差より数多く起る傾向があるため。図の右手の観測値は比較的にばらついている。この傾向は5図では余り顕著ではなく、この図の正規性からの隔りは、それが僅かであるから分析には殆んど影響がない程である。

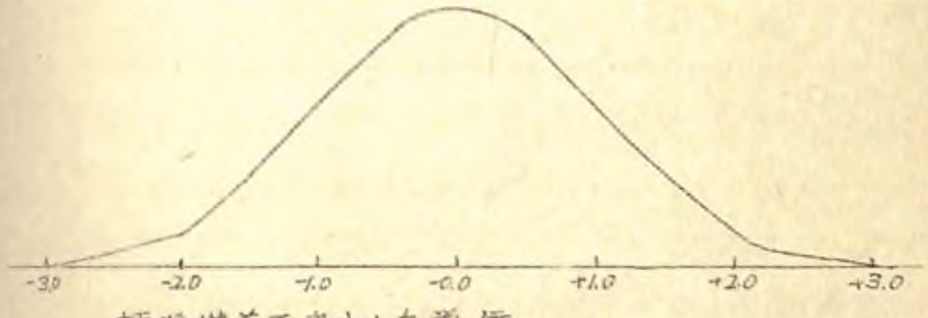
正規性からの重大な隔りは散布図でかなりはっきりするだらうということは一般に云える。しかし10章で説明する方法で非正規性を検定し、必要があればこれを修正するため、一段と精しい解析をすることも出来る。この様な解析には通常数多くの余分な仕事が含まれており、しかも多くの場合正規性の仮定に基ずくものと著しく違つた結果となることは少い。

さらに、どの様な解析の時でも一般に次の様な仮定が立てられる。即ち当てはめた関係の形は観測値を合理的表現を与える。どの様な形の関係が観測値に最も良く当てはまるか検定出来るが、勿論これが次の観測値に、特にそれが原観測値の範囲外にある時にやはり当てはまるとは云えない。多くの場合、この関係の形は原観測値の範囲外の値に対してはつきりした相違はないであらうと仮定する必要がある。この様な"補外" extra-polation は統計的作業では屢々必要とされるが、継時観測値の規則性に関する仮定によつて始めてこの様なことが行なえるのである。

4.2. 分散および分散分析 Variance and analysis of variance

分散および共分散は普通の統計用語であり、これについては大低の統計の書物に詳しく説明している。次の2節では読者の他日の用に資するため、分散および共分散の主な性質と計算法について簡単に考察しよう。しかし、完全な説明については、もつと基本的な書物を参照されたい。*

*例えば、Introductory Statistics, M.H. Quenouille,



3 2図　　正規分布に従う偏差の相対的頻度

London, 1950

分散は一組の観測値の散ばりの測度である。

これは総平均からの観測値の平均平方偏差である。したがつて分散の平方根は観測値がその総平均から隔つている程度についてある指標となる。この平均平方偏差の平方根は標準偏差 standard deviation といわれている。

正規分布では分散又は標準偏差により平均から任意の距離内にある観測値の割合を充分に決定できる。例えば観測値の68%は平均から1標準偏差の範囲にあり、95%は平均から2標準偏差の範囲内にある。4.2a表は正規分布の場合平均から任意の距離にある観測値の百分率を示している。

4.2a表　正規分布の場合、平均から任意の距離にある観測値の百分率

標準偏差に掛ける乗数	0.0	0.2	0.4	0.6	0.8	1.0	1.2	1.4
観測値の百分率	0.0	7.9	15.5	22.6	28.8	34.1	38.5	41.9
標準偏差に掛ける乗数	1.6	1.8	2.0	2.2	2.4	2.6	2.8	3.0
観測値の百分率	44.5	46.4	47.7	48.6	49.2	49.5	49.74	49.87

この百分率は厳密には正規分布に適用されるだけであるが、近似的には殆んどの分布に対しても使える。

一組の観測値 $x_1\ x_2\ x_3$…… の分散は一般に　　或は Var(x) で表わされる。標準偏差は s で表わされる。

通例、一組の観測値の分散又は標準偏差は未知ではあるが、観測値から推定できる。このことは観測値が同じ分布からとられた時ですら、いつでも同じ値が得られるとは限らないことを意味している。推定された関係を検定したり、その精確度を求める時には推定された分散の不正確さについて何んらかの考慮をしなければならない。

一組の観測値 x_1, x_2, x_3…… の分散の推定値は一般に s^2 又は s.var (x) で表わされる。これに対応して標準誤差の推定値は s で表わされる。n個の観測値の分散は次式から推定される。

$$s^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1} = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - \frac{(\sum_{i=1}^{n} Xi)^2}{n}\right]$$

こゝで Σ は和を表わす記号である。例えば8個の観測値 6, 5, 3, 4, 7, 1, 2, 3, の分散の推定値は

$$s^2 = \frac{1}{n}\left[36+25+9+16+49+1+4+9-\frac{(31)^2}{n}\right] = 4.125$$

標準偏差の推定値は

$$s = \sqrt{(4.125)} = 2.031$$

n−1は推定値の自由度といわれている。全体の変動を異なる原因に帰因する変動を表わす成分に分けるため一組の観測値について分散分析が行なわれる。例として4.2b表に示してある20個の観測値について考えてみる。

4.2b表　観測値の二元表

16	21	23	32	27	119
24	31	19	42	36	152
12	23	25	29	30	119
16	16	22	27	21	102
計 68	91	89	130	114	492

観測値の分散の推定値は

$$\frac{1}{19}\left[16^2+21^2+\cdots\cdots+21^2-\frac{492^2}{20}\right] = \frac{1039}{19}$$

$$= 54.7$$

自由度は19である。総平方和1039は3つの部分に分割される。即ち行間の変動を表わす部分、列間の変動を表わす部分、行および列内の変動を表わす残差の部分。行間の変動を表わす部分は、この場合には

$$\frac{119^2+152^2+119^2+102^2}{5} - \frac{492^2}{20} = 263$$

列間の変動を表わす部分は

$$\frac{68^2+91^2+89^2+130^2+114^2}{4}-\frac{492^2}{20}=577$$

行および列内の変動を表わす部分は

$$1039-263-577=199$$

初めの2つの平方和の際数5，4はこの場合にはそれぞれの計に含まれている観測数である。

この値は次の分散分析の表にまとめられる。

表4.2c　　分散分析

変動因	自由度	平方和	分散の推定値
行間	3	263	87.7
列間	4	577	144.2
行および列内	12	199	16.6
計	19	1039	54.7

この場合〃行間〃および〃列間〃の成分に対する自由度はそれぞれ行および列の数より1少い。

分散の推定値は平方和を対応する自由度で割ることにより求められる。

観測値の変動の大部分は行間および列間の変動により説明できることが分散の推定値から分る。実際、行間および列間の分散を考慮することによつて分散は54.7から16.6に減小している。

分散分析により一組の観測値の変動を考究し、その原因を調査できるのである。例えばこの20個観観測値は一日の5の異つた時刻に4つの気温で得られたものであるとする。即ち行は気温、列は時刻を表わすとする。

測定値の変動の大部分は温度と時刻に帰因するものであり、温度、時刻に関する知識により測定値の予測はもつと正確になることを上の分析は示している。

分散又は標準偏差を比較すればどの位正確になつたかが分る。

上例は分散分析の使用法のほんの一例に過ぎない。次章には幾多の例が示されるであらう。その例の中で、ある変動因が全体変動に顕著な役割を演じているか否かを検定する必要のあることが分るであらう。例えば、行間変動が全変動に有意な関係があるか否を検定する必要がある。このことは分散比ー検定を使つて行なわれる。

まず分散の推定値の比を計算する。この場合には 87.7／16.6 = 5.28 附録のⅦ，Ⅷ表の様な分散比の表を使う。分子、分母の自由度に対する表の値を使う。この場合、自由度はそれぞれ3，7でありⅦ表およびⅧ表によれば5％および1％有意水準の値はそれぞれ3.49と5.95である。5.28はこの最初の値を越しているから行は全分散に対して有意な関係がある（$P<0.05$）と結論された。同様に分散比144.2／16.6 = 8.69は行が観測値の分散に有意な関係をしていることを示している。

次に一組の観測値の分散を特定の原因に帰因する分散と原因不明の分散とに分割する方法の例を示さう。特定の原因の有意性は分散比検定をつかつて検定される。

4.3　共分散および共分散分析　Covariance and analysis of covariance

二変量XYを関係付けたい時、各変量の分散はそれぞれの散ばりを示しているが、一方の変量の上昇又は下降が他の変量の上昇又は下降におよぼす影響を測るためには、〃共分散〃を使う必要がある。これは全体の平均

からの各変量の偏差の積の平均である。

符号はある変量の上昇に対し他の変量が上昇するか下降するかを示している。i正の符号は2変量が共に上昇する傾向であることを示し、負の符号は一方の上昇に対して他方が下降する傾向を示す。しかし共分散の大きさは、連関の程度を示すものではない。このためには、2変量の分散に関連させて共分散の大きさを考察しなければならない。共分散の意義については4.8節で充分に考察することにする。

2変量X，Y間の共分散を表わすには一般に記号cov （X，Y）が使われている。

2変量の共分散は通常未知であり、したがつて観測値から推定しなければならない。

X_1Y_1 i X_2Y_2 i …… X_nY_n を観測値の組とすればX，Yの共分散は次の公式から推定される。

$$e.cov(X,Y) = \frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{n-1}$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n}X_iY_i - \frac{(\sum_{i=1}^{n}X_i)(\sum_{i=1}^{n}Y_i)}{n}\right]$$

例えば、次の様な観測値の対の共分散の推定値は

6,9 5,6 3,3 4,4 7,13 1,0 2,1 3,4

$$e.cov(X,Y) = \frac{1}{7}\left[54+30+9+16+91+0+2+12-\frac{(31)(40)}{8}\right]$$

$$= 8.429$$

n－1がやはりこの推定値の自由度として使われる。

共分散分析は2組の観測値についてその全共変動を異なる原因に帰因する成分に分割するために行なわれる。例として4.3a表に示してある観測値について考えてみよう。

4.3表　観測値の二元表

	変量X					
	16	21	23	32	27	119
	24	31	19	42	36	152
	12	23	25	29	36	119
	16	16	22	27	21	102
計	68	91	89	130	114	

	変量Y					
	6	7	12	3	17	45
	4	3	9	2	15	33
	9	12	19	5	24	69
	10	15	22	7	20	74
計	29	37	62	17	76	221

分散分析表は前節に示した方法で作ることができる。

4.3b表　分散分布

変動因	自由度	X 平方和	X 分散の推定値	Y 平方和	Y 分散の推定値
行間	3	263	87.7	228	76.0
列間	4	577	144.2	588	147.0
行および列内	12	199	10.6	49	[illegible]
計	19	1039	54.7	865	45.5

この2組の観測値の変動の大部分は行間および列間の変動に帰因させうることをこの分析は示している。

このような分析における平方和に対して積和が計算される。全体の積和は

$$16\times6+21\times7+\cdots\cdots+21\times20-\frac{(492)(221)}{20}=-162$$

行間の積和は

$$\frac{119\times45+152\times33+119\times69+102\times74}{5}-\frac{(492)(221)}{20}=-211$$

列間の積和は

$$\frac{68\times29+91\times37+\cdots\cdots\cdots+114\times76}{4}-\frac{(492)(221)}{20}=-4$$

行および列内のX，Yの同時変動を表わす部分は

$-162-(-211)-(-4)=53$

この様に分割されたものは次の共分散分析表に纒められる。

4.3c　共分散分析

変動因	自由度	平方和	共分散の推定値
行間	3	−211	−70.3
列間	4	−4	−1.0
行および列内	12	53	4.4
計	19	−162	−8.5

観測値には負の連関があるがこれは主として行和間に負の連関があるためであることをこの分析の符号は示している。行効果を除けば、この観測値間には正の連関が存在する。

上例において行は気温を、列は時刻を表はしているとしよう。XとYの

変動の大部分は、気温と時刻に帰因させうることをこの分析は示している。さらにXとYには負の連関があるがこれは主として気温の効果のためであることを示している。即ちこれは一方が上ると同時に他方は下るためである。最後に行および列内の共分散は与えられた気温と時刻においてこの両変量間には正の連関のあることを示している。

上記の例は共分散分析の計算法および解釈の一例に過ぎない。連関の推定および検定におけるそれの使用法については次の数章で詳しく考察しよう。

4.4　回帰方程式の推定および検定

Estimation and testing of regression equations

2.2で与えられた直線の方程式は

$$Y-Y_0=B(X-X_0)$$

Bは線の傾き、$(X_0\ Y_0)$ はその線上の任意の点、回帰線を推定するには、その傾きおよびその線上の任意の一点を推定しなければならない。一方の測定値の平均値は他方の測定値の平均値に対応するから線上に定められる点は平均点である。したがつて2組のn個の観測値の算術平均がこの点を推定するために用いられるであらう。

$$\bar{X}=\frac{\sum_{i=1}^{n}X_i}{n}\qquad\bar{Y}=\frac{\sum_{i=1}^{n}Y_i}{n}$$

であれば、$(\bar{X},\bar{Y})$ が直線方程式を推定する場合に用いられる。

回帰線の傾きを推定するには $B=\frac{cov(X,Y)}{var(X)}$ であることに注目する。

したがつて傾きの推定値bは次式から求められる。

$$b = \frac{e.cov(X,Y)}{e.var(X)} = \frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{\sum_{i=1}^{n}(X_i-\bar{X})^2}$$

この公式の使用法の例として、4.4a表に示してある頭の周囲と胸囲に関する123個の観測値を考察してみよう。

これらの観測値の和、平方和、積和は次の様に計算された。

$\Sigma H = 5635.2$　　$\Sigma C = 5497.9$

$\Sigma H^2 = 258408.30$　$\Sigma HC = 252048.25$　$\Sigma C^2 = 246229.33$

この値から次の様に計算された。

$\bar{H} = 45.8$　　$\bar{C} = 44.7$

$\Sigma(H-\bar{H})^2 = 233.67$　$\Sigma(H-\bar{H})(C-\bar{C}) = 163.97$　$\Sigma(C-\bar{C})^2 = 432.14$

$e.var(H) = 1.9153$　$e.cov(H.C) = 1.3440$　$e.var(C) = 3.9520$

$$b = \frac{1.3440}{3.9520} = 0.3401$$

したがつてCに対するHの回帰線の方程式は次の様に推定される。

$$H - 45.8 = 0.3401(C - 44.7)$$

故に

$$H = 0.3401C + 30.60$$

2.2節で求めた図型による推定値はこの値と良く一致している。さらにHに関するCの回帰を与える才2の線は次式から求められる。

$$C - 44.7 = \frac{1.3440}{1.9153}(H - 45.8)$$

$$= 0.7017\,(H - 45.8)$$

推定した回帰線のあてはまりの良否を検べるには分散分析を使わねばならない。これによつて従属変量Yの全変動はXに帰因する部分

$$b^2\sum_{i=1}^{n}(X_i-\bar{X})^2 = b\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y}) = \frac{\left[\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})\right]^2}{\sum_{i=1}^{n}(X_i-\bar{X})^2}$$

と原因不明の部分

$$\sum_{i=1}^{n}(Y_i-\bar{Y})^2 - b\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y}) = \sum_{i=1}^{n}\left[(Y_i-\bar{Y})-b(X_i-\bar{X})\right]^2$$

完成した分散分析表は4.4b表の様になる

4.4b表　直線回帰を検定するための分散分析表

変動因	自由度	平方和	分散の推定値
Xに帰因するもの	1	$b\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})$	$b\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})$
原因不明	n−2	$\sum_{i=1}^{n}(Y_i-\bar{Y})^2 - b\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y}) = S_1$	$S_1/n-2$
計	n−1	$\sum_{i=1}^{n}(Y_i-\bar{Y})^2 = S2$	$S2/n-1$

この様な分散分析における各成分を比較すれば、XがYの変動に有意な関与をしているかどうかが示されるであらう。さらに分散の推定値の検討によつて全体の散ばりおよび回帰線の周りの散ばりについての概念を得ることも可能である。

4.4a　胸囲と頭の周囲の観測値

	C	H		C	H		C	H		C	H
1	40.0	44.0	9	41.5	44.2	17	42.0	45.8	25	43.0	48.7
2	40.5	43.5	10	41.5	45.8	18	42.5	46.0	26	43.0	44.0
3	41.0	45.8	11	41.5	43.5	19	42.8	42.5	27	43.0	44.2
4	41.0	45.4	12	41.6	43.6	20	43.0	46.2	28	43.0	44.0
5	41.0	46.3	13	41.8	44.5	21	43.0	45.3	29	43.2	44.4
6	41.5	45.0	14	41.8	44.0	22	43.0	44.2	30	43.2	44.2
7	41.5	45.0	15	42.0	44.2	23	43.0	46.0	31	43.2	44.2
8	41.5	46.0	16	42.0	46.8	24	43.0	46.5	32	43.5	47.2

	C	H		C	H		C	H		C	H
33	43.5	44.2	56	44.5	48.0	79	45.5	45.0	102	47.0	47.0
34	43.5	46.5	57	44.5	42.4	80	45.5	45.5	103	47.0	47.6
35	43.5	45.5	58	44.5	45.0	81	45.5	45.6	104	47.0	44.5
36	43.5	45.0	59	44.5	44.8	82	45.8	46.7	105	47.0	44.7
37	43.5	45.2	60	44.5	45.2	83	46.0	46.3	106	47.0	47.5
38	43.5	43.8	61	44.6	47.3	84	46.0	47.8	107	47.0	46.7
39	43.8	44.5	62	44.8	45.6	85	46.0	47.6	108	47.0	47.0
40	43.8	44.6	63	44.8	46.8	86	46.0	47.5	109	47.0	46.8
41	43.8	45.4	64	44.8	47.2	87	46.0	46.5	110	47.0	46.6
42	44.0	46.5	65	45.0	44.8	88	46.0	44.8	111	47.0	47.0
43	44.0	46.0	66	45.0	45.0	89	46.0	44.8	112	47.5	46.3
44	44.0	47.0	67	45.0	48.5	90	46.0	46.8	113	47.5	46.2
45	44.0	45.5	68	45.0	45.1	91	46.0	47.2	114	47.5	48.0
46	44.0	45.5	69	45.0	43.8	92	46.2	45.0	115	47.5	47.8
47	44.0	45.5	70	45.0	45.0	93	46.5	46.3	116	47.5	48.0
48	44.0	47.5	71	45.0	48.4	94	46.5	46.3	117	47.6	47.2
49	44.0	46.0	72	45.2	44.2	95	46.5	48.0	118	48.0	47.5
50	44.0	44.0	73	45.2	46.5	96	46.5	46.0	119	48.0	45.2
51	44.0	46.4	74	45.2	46.0	97	46.5	47.5	120	48.0	46.5
52	44.0	46.2	75	45.2	45.0	98	46.5	47.0	121	48.0	44.8
53	44.0	43.5	76	45.5	45.5	99	46.5	46.2	122	48.5	45.8
54	44.2	45.0	77	45.5	46.0	100	46.5	46.8	123	49.0	50.2
55	44.5	46.3	78	45.5	45.4	101	46.8	47.0			

頭の周囲と胸囲に関する観測値の分散分析は 4.4C表の様に計算された。

4.4c表　胸囲（C）に関する頭の周囲（H）の回帰の分散分析

変動因	自由度	平方和	分散の推定値
Cに帰因するもの	1	55.77	55.77
原因不明	121	177.90	1.470
計	122	233.67	1.915

回帰の有意性を検定するには、分散比 55.77／1.470＝37.9 を計算しなければならない。自由度１，１２１のⅦ表をみれば、この様に高い比が生ずるのは１００回に１回以下であることが示されている。したがつて連関の有意性については疑問の余地はない。

頭の周囲の標準偏差は $\sqrt{(1.915)}=1.384$ であるが回帰線の周りの標準偏差は $\sqrt{(1.470)}=1.212$ であることに気付くであらう。この後者の推定値は 2.5 節で求めた図形による推定値とよく一致している。

4.5　回帰係数の推定値の誤差
Error of estimated regression coefficients

回帰係数の推定値が前に求めた値或は理論的な値と違つているか否かを検定するため、それの確率誤差を求めたい場合が多い。これは回帰係数の分散を推定し、これを使つて真の回帰係数の値に対する限界を定めることにより行なわれる。

xに無関係なyの分散を表わすために $var_x(y)$ を使えば n 対の観測値から推定された回帰係数の分散は

$$var(b)=\frac{var_x(y)}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

これの平方根は回帰係数の標準誤差を与える。

もちろん $var_x(y)$ を推定することが必要であり、この時には V(b)は次式から推定される。

$$e\ var(b)=\frac{e\ var_x(y)}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

真の回帰係数に対する限界を定める時にはこの推定値に対して多少考慮が払われる。このためにはK表－t表を使う。

この表の $e\ var_x(y)$ の自由度の処で、与えられた確率水準Pに相当する値tを求める。真の回帰係数に対する限界は次式で与えられる。

b ± t ×（b の標準誤差）

真の値がこの範囲を越える確率はPである。上の公式の応用例として、前節の回帰係数の推定値を考察してみよう。b の分散の推定値は

$$e \quad var(b) = \frac{1.4\,7\,0}{4\,8\,2.1\,4}$$

$$= 0.0\,0\,3\,0\,4\,9$$

$$b \text{ の標準誤差} = \sqrt{0.0\,0\,3\,0\,4\,9}$$

$$= \pm 0.0\,5\,5\,2\,2$$

b の限界を定めるには，K表で自由度１２１に相当する値を求める必要がある。P＝0.05　に対してはtの値は1.98　P＝0.01に対しては2.62　P＝0.001に対しては3.37である。したがつて９５%の確実性をもつて　b の真値は

0.3401 ±(1.98×0.05522)＝ 0.2308 と 0.4494

９９%の確実性で

0.3401 ±(2.62×0.05522)＝ 0.1954 と 0.4848

９９.９%の確実性で

0.3401 ±(3.37×0.05522)＝ 0.1540 と 0.5262

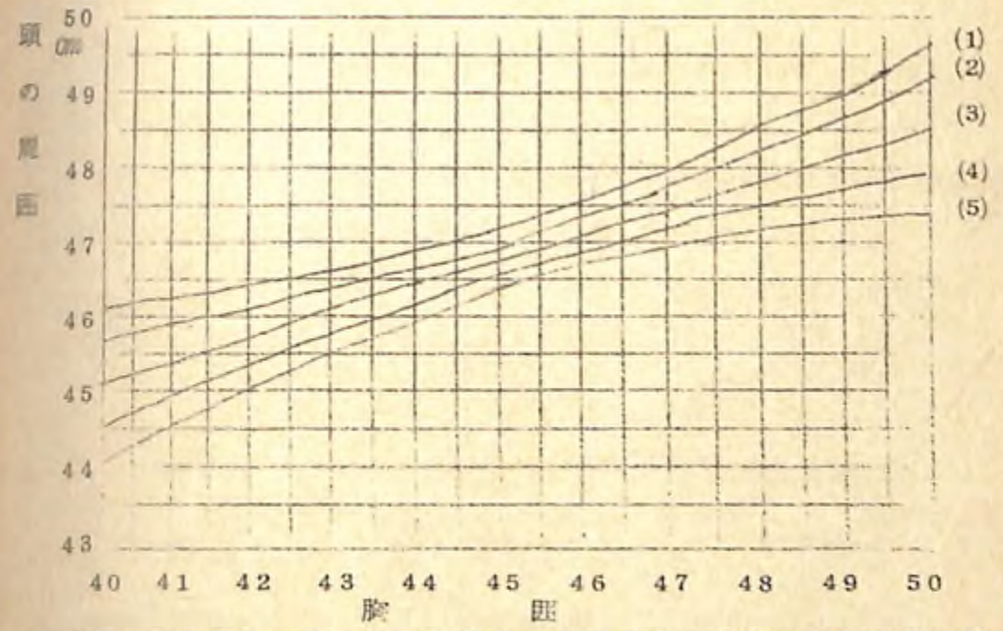
の範囲内にある。

この b の限界を定める方法は，真の値がほゞ確実に０でないことを示しているから推定された回帰の有意性の別の検定となつていることが分る。

回帰係数は，夫々の分散を使つて比較することも出来る。しかし，このためには，原因不明の分散のこみした推定値を求める必要があり，完全な比較方法は非常に長たらしいものになるであろう。このことに興味のある読者はIntroductory Statistics の第７章を参照されたい。5.4節の方法は，この目的にも使える。

4.6　推定値の誤差　Error of an estimated value

直線関係が推定されたならば，これは従属変量が与えられた独立変量の値に対してとる平均的な値を推定するため使われる。しかし推定された関



３３図　与えられたCの値に対応するHの推定値の９５%および99.9%信頼限界

(1) Upper 99.9%limit　(2) Upper 95%limit
(3) Regression line　(4) Lower 95%limit
(5) Lower 99.9%limit

係数には誤差が含まれる可能性があるから推定された値にも当然誤差がある。独立変量の値Xに対応する推定値が $\bar{y} + b(X - \bar{x})$ であることに注目すれば推定値の分散は次式から求められる。

$$var(\bar{y}) + (X-\bar{x})^2 \quad Var(b) = \frac{var_x(y)}{n} + \frac{(X-\bar{x})^2 \, var_x(y)}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

$$= var_x(y)\left[\frac{1}{n} + \frac{(X-\bar{x})^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right]$$

これは、推定値の限界を定めるのに使われる。

頭と胸の周囲に関するデータについては、4.4節の値を使う。任意の胸囲Cに対する頭の周囲の平均の推定値は

$$0.3401C+30.60$$

この推定値の標準誤差は

$$\sqrt{\left[1.470\left(\frac{1}{123}+\frac{(C-44.7)^2}{482.14}\right)\right]}$$

したがつて、例えば95%の確実さで真の値が含まれる範囲は

$$0.3401C+30.60 \pm 1.98\sqrt{\left[1.470\left(\frac{1}{123}+\frac{(C-44.7)^2}{482.14}\right)\right]}$$

ここで、1.98はP=0.05　自由度121　に対するtの値である。任意の正確度に対しても、同様な範囲を計算できる。33図にはいろいろな胸囲の値に対する頭の平均周囲の推定値の95%、および99.9%の限界が示してある。

さらにm個の観測値をとつたときその平均が、当てはめられた回帰線と有意に違つているか否かを検定するには、回帰線に含まれている誤差と同様に、観測値の変動も考慮しなければならない。当てはめた線からの観測値の偏差の分散は

$$var_x(y)\left[\frac{1}{m}+\frac{1}{m}+\frac{(X-\bar{x})^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right]$$

Xは観測値である。

この公式を使えば、与えられた百分率で観測値が落ちる範囲が定められる。例えば、頭と胸の周囲の追加観測値の99.9%は次式で与えられる範囲内に落ちるであろう。

$$H = 0.3401C + 30.60 \pm 3.37\sqrt{\left[1.470\left(1+\frac{1}{123}+\frac{(C-44.7)^2}{482.14}\right)\right]}$$

ある追加観測値 C = 45.0　H = 39.2　がこの観測値の組と一致しているかどうか検定したいとする。この場合の99.9%限界は

$$H = 0.3401\times45.0+30.60\pm3.37\sqrt{\left[1.470\left(1+\frac{1}{123}+\frac{(45.0-44.7)^2}{482.14}\right)\right]}$$

$$= 45.90 \pm 3.37 \times 1.217$$

$$= 41.80 \text{ と } 50.00$$

追加観測値はこの限界を遙かに越えており、したがつて前にとつた観測値とは有意に違つていると結論される。

4.7　異常観測値の検定　Testing extreme observations

前節の方法は、追加観測値が当てはめられた関係から有意に隔つているか否かを検べるのに使われた。しかし異常な観測値が別の観測値と一諸にとられた様な時には、違つた検定方法が採用される。

一般に、解析の頭初で、極めて異常な観測値だけを棄却する様にしなければならない。これらの値は前節の方法で後で吟味してもよい。この目的のためには、高い有意水準を使わねはならない。

解析の頭初には、異常性の少ない観測値即ち余り疑がわしくない観測値を含ませておく必要があり、それらの、当てはめた線からの偏差も亦高い有意水準で検定する。例えば、全体で100個の観測値については、異常値は0.1%又はそれ以上の有意水準で吟味する必要がある。

解析に含まれている観測値の観測された偏差の分散を計算するため、観測値の分散から推定値の分散を差し引かねばならない。即ち偏差の分散は

$$var_x(y)\left[1-\frac{1}{n}-\frac{(X-\bar{x})^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right]$$

Xは観測値である。

例えば胸囲に対する頭の周囲の回帰を推定する際使われた観測値 C=49.0 H=50.2 を検定する場合 99.9%限界はHに対して次の様に定められる。

$$H = 0.3401\times 49.0+30.60\pm 3.37\sqrt{\left[1.470\left(1-\frac{1}{123}-\frac{(49.0-44.7)^2}{482.14}\right)\right]}$$

$$= 47.26 \pm 3.37 \times 1.184$$

$$= 43.27 \text{ 又は } 51.25$$

この観測値は充分この限界内にあり、従つて棄却する必要はない。

ここで使用した $var_x(y)$ の推定値は隔りと独立ではないから、限界を定めるこの方法は、実際には近似的なものに過ぎない。この結果限界は、それが当然あるべきものより幾分広めに定められる。しかし幸にも多数の観測値がとられた場合には、この影響は僅かであり、選択の結果生ずる偏つた分散の推定を防ぐ傾向がある。

正規性の仮定からの隔りは有意水準に大きな影響を与え勝ちであるからどの様な場合でも異常観測値を調べる方法は決定に大略の基礎を与えるに過きないということを注意を要する。第1章で指摘した様に特定の観測値又は観測値の組を棄却するかを最終的に決めるには、いろいろ考察してみなければならないし、異常観測値の検定はある場合にたけ必要とされる。

4.8 積率相関係数

Product-moment correlation coefficient

共分散の意味およびその説明について考察しよう。共分散の符号は2変量が正の相関であるか負の相関であるかを示していることを前に指摘した。しかし、2変量の分散と関連させて考えない限り、共分散の大きさは何の意味も表わさない。即ち変量の測定尺度を変えることにより共分散はそれに対応して変るであろう。したがつて2変量間の連関の程度を表わす指数を求めるには、測定の尺度が何んらかの方法で標準化された場合の共分散を考察してみる必要がある。かゝる標準化は実際には2つの分散が等しくなる様に尺度を選ぶことによつて行なわれる。したがつて共分散は連関の程度を示している。元の分散、共分散から求めた新しい共分散は、

$$\frac{cov(x,y)}{\sqrt{\left[var(x)\ var(y)\right]}}$$

この量は積率相関係数と呼ばれ ρ で表わされる。これに相当する推定値は r で表わされる。即ち

$$r = \frac{s\ cov(x,y)}{\sqrt{\left[s\ var(x)\ s\ var(y)\right]}}$$

この相関係数は尺度に無関係であるから、別のものと比較することが出来、一般に2変量間の連関の程度を表示するのに使われる。これは一番良く使われる相関係数であり、多くの場合この様な理由から、全然修飾語を付けずに相関係数と呼ばれている。したがつて、我々が2変量間の相関係数の話をする時には、それは普通こゝで説明している積率相関係数のことである。

相関係数は1を越える値をとることは出来ず、又1の値は変量間に完全な正の直線的連関のある時に限られている。それと対照的にマイナス1以

下の値をとることも出来ず、又－1という値は完全な負の直線的連関のある時にとられるものである。この中間にある値では相関係数の平方はある変量に関する一次回帰で説明できるもう一方の変量に含まれる全変動の割合を示す。xとyとの間の相関係数が0.7である場合にはxの変動、即ちxの分散の49%はyに関する一次回帰に因るものと考えることが出来る。その逆も同じである。

頭と胸の周囲に関するデータでは、相関係数は次の様に推定された。

$$r = \frac{1.3440}{\sqrt{1.9153 \times 3.9520}}$$

$$= 0.489$$

これは一方測定値の全変動の24%が他方の変量に対する一次回帰により説明できることを示している。

観測された相関係数の有意性を検べるにはX表を使わねばならない。例えば観測数60では、絶対値0.254或は0.330を越える相関係数は、夫々5%および1%有意水準で有意である。

観測数120では、1%有意水準は0.234であり、したがつて、上で計算した値は明らかに高度に有意である。

相関係数の自由度として、観測値の対数より2減した値を使うと便利なこともある。それは、直線を推定するには2対の観測値が必要であり、少くとも3対の観測値を使わない限り、連関を検定することは出来ないからである。したがつて相関係数の自由度は連関の検定に使える観測値の有効対数を示す。

相関係数の検定は回帰係数の検定と真く同じ仮設、即ち2つの変量間に相関がないか、即ち、それらの間の一次関係は偶然の結果生じたものより大きいかという仮設を吟味するものである。もちろんこの2つの検定は同

意味のものであり、真く同じ結果が得られる。

この方法のいづれを使うかは一変量の他変量に対する一次的従属関係を推定し検定したいのか或は2変量間の一次的連関の程度を推定したいのかどうかにより変つてくる。

4.9 相関係数の比較と組合せ
Comparing and combining correlation coefficients

同じものであるかどうかを決めるため2つ以上の場合で観測された相関係数を比較しなければならない場合がある。それには連関の程度が別の考えにより影響されるかどうかを検べてみる必要がある。そのためには相関係数を変換し、もつと便利な量、Zを使う必要がある。この変換はX表を使つてなされる。例えば0.315に等しいrの値は、0.322に等しいZの値に変換される。この変換された値Zは分散$1/(n-3)$をもつて近似的に正規分布する。nは観測対の数である。したがつてn_1およびn_2対の観測値に基ずく2つのZの値が比較されるならば、その差の標準誤差は

$$\sqrt{\left(\frac{1}{n_1-3}+\frac{1}{n_2-3}\right)}$$

有意差があるか否かを検定するには次式を計算する必要があり、これを正規正規偏差として検定する。

$$(Z_1-Z_2)\Big/\sqrt{\frac{1}{n_1-3}\quad\frac{1}{n_2-3}}$$

この比の5%、1%、0.1%有意水準は夫々1.96、2.58、3.29である。

2つ以上のZの値に有意な差がなければ、それから計算された相関係数

は有意に違つておらず、全体の相関の推定をすることが出来る。これは次の式から計算され

$$Z = \frac{(n_1-3)Z_1 + (n_2-3)Z_2 + \cdots}{(n_1-3) + (n_2-3) + \cdots\cdots}$$

$\bar{r}$ の値を求めるため逆に変換する。この $\bar{r}$ の値は実際には $n_1+(n_2-3)+(n_3-3)+\cdots$ 対の観測値に基づくものである。

例として、頭と胸の周囲のデータを考察してみる。１２３観測値の中６５は少年、５８は少女から得られたものであつた。この２組のそれぞれの相関係数は0.４０１と0.５０５であつた。これは有意に違つているだろうかという質問が起る。その分析は、4.9 a表の様にして行う。

4.9 a表　相関係数の比較と組合せ

性別	n	r	Z	(n−3)	(n−3)Z
男	６５	0.４０１	0.４２５	５２	２６.３５０
女	５８	0.５０５	0.５５４	５５	３０.４７０
				１１７	５６.８２０

$$\frac{Z_1 - Z_2}{\sqrt{\frac{1}{n_1-3} + \frac{1}{n_2-3}}} = \frac{0.425 - 0.554}{\sqrt{\frac{1}{62} + \frac{1}{55}}}$$

$$= -0.70$$

$$\bar{Z} = \frac{56.820}{117}$$

$$= 0.486$$

$$\bar{r} = 0.451$$

この２つのZの値の差は有意でない。したかつて、こみにした相関係数の推定値が求められる。この場合この値は0.４５１で実際に１２０対の観測

値に基づいている。

この方法で求めたこみにした相関の推定値は全観測値を一諸にして求めた相関の推定値とは同じではないことに注意しなければならない。こみにした推定値は個々の組の平均、分散、および個々の回帰線の傾斜間の差を考慮していない。この結果、全観測値を一諸に使つて求めた推定値より観測数は少い。

観測値が色々な資料源からとられた様な場合には、Z変換を使つて求めた相関の推定値を考える方が望ましい。というのはこれを使えば資料源間の差違は除去されると考えられるからである。さらに幾組もの観測値の平均が相互に有意に違つているかどうかを検定することも有意差がなくても全観測値を一諸に使うことも困難な問題が生じてくるからである。

我々が時析当面するこれとは違つた型の問題は２組の観測値のいづれが第３の組とより密接な相関があるかということを決めることである。例えば２つの簡単な心理的テストのいづれが時間のかかる知能テストと密接な相関のあるかを知りたいとする。この検定が出来ればより相関の高い方が知能の測度として使われるであろう。

この様な場合の正確な方法は、相関係数を比較することである。それらの差の有意性は次の様にして検定する。

２変量 x，y のいづれが第３の変量 a とより密接な相関のあるかを決めねばならないとしよう。まず、対としてとつたこれら３変量間の相関係数を計算しなければならない。これにより３つの値 r_{xa}, r_{ya}, r_{xy} が求められる。r_{xa} と r_{ya} との比較は x 或いは y のいづれが a と密接な関係があるか否かを示し、この２つの相関係数の差の有意性は相関係数として次式を検定することにより決定される。

$$\frac{r_{xa} - r_{ya}}{2(1-r_{xy})}$$

この量は普通の係数と同数の観測値対に基づくものと考えられる。

この検定の例として、4.4 a表のデータを使つて、体重 W，頭の周囲 H のいづれが胸囲 C とより密接な相関があるかということを考察してみよう。この場合の相関係数の推定値は

$$r_{WC} = 0.755 \qquad r_{HC} = 0.489 \qquad r_{WH} = 0.564$$

（3.7節の値と比較せよ）

この値から体重が胸囲と密接な相関のあることは明らかである。しかしこれが有意性のある密接な相関であることを決めるには次の様な計算を行う必要がある。

$$\frac{r_{WC} - r_{HC}}{2(1-r_{WH})} = \frac{0.755 - 0.489}{2(1-0.564)} = 0.305$$

X表から1%有意水準の値は 0.2 3 4 であり、したがつてこの値は極めて有意である。

したがつて体重は頭の周囲よりも胸囲の有意なより良き指標であると結論される。もちろん体重と頭の周囲の両者に使つた測度は個々のものより当然良い結果が得られる。この種の組合した連関を含む問題は次章で考察することにする。しかし時間その他の原因により，一組の測定値しかとれない場合には，上記の検定法はどの測定値の組が最も有効であるかを決定する時に役立つであろう。

4.10 固有の関係の推定

Estimation of underlying relationships

両組の測定値に含まれる原因不明の変動の相対的大きさについて仮定を設けることか出来ない限り，固有の線型関係は推定できないことを，2.6節で指摘した。

$$x = a_1 t + b_1 + \varepsilon_1$$

$$y = a_2 t + b_2 + \varepsilon_2$$

であるとする。

x と y 間の固有の関係は

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

である。

しかしこの関係を推定するには比

$$R = \frac{var(\varepsilon_2)}{var(\varepsilon_1)}$$

が分つていなければならない。

この比が既知であれば、推定された関係は

$$y - \bar{y} = b(x - \bar{x})$$

$$b = \frac{\varepsilon\, var(y) - R\,\varepsilon\, var(x)}{2\,\varepsilon\, cov(x,y)} + \sqrt{\left[R + \left(\frac{\varepsilon\, var(y) - R\,\varepsilon\, var(x)}{2\,\varepsilon\, cov(x,y)}\right)^2\right]}$$

頭と胸の周囲のデータでRを 0.5 として 4.4 節に示してある値を代入すれば

$$b = \frac{1.9153 - 0.5 \times 3.9520}{2 \times 1.3440} + \sqrt{\left[0.5 + \left(\frac{1.9153 - 0.5 \times 3.9520}{2 \times 1.3440}\right)^2\right]}$$

$$= 0.685$$

$$H - 45.8 = 0.685\,(C - 44.7)$$

別のRの値に対して同様な線が定められ 2 6 図に示してある様にいろいろのRの値に対して一組の線か求められる。

残念ながら補足的情報がないとRを推定することは不可能であり，一般にはRについて大略の推定をして，色々なRの値に対する推定された関係の感度を確めるためある範囲の値を代入してみる必要がある。

別の推定方法が7.6および10.5節に示してあるが，いづれの場合でも ϵ_1 , ϵ_2 の特性についての仮定が必要である。Rの妥当な値について追加情報が得られるか，さらに仮定が設けられて，始めて固有の関係が推定できるのである。

2.6節で指摘した様に，この方法は独立変量に抽出誤差その他の誤差のある時にも使える。この様な誤差がなければどの様な回帰関係があるかを推定したい。この場合，この問題は"理想的"回帰関係を求めることである。

独立変量に含まれる抽出誤差と従属変量に含まれる全ての誤差との比が分つておれば，実際にこのことは行える。これを上記方程式のRの所に代入する。別法が8.7節に示してある。

5 多変量の連関 Multiple Association

5.1 重回帰方程式の推定

Estimation of multiple rgression equation

この章では，幾組かの測定値が同時にとられた時の関係の推定法を考えてみよう。

例えば，小麦収量，降水量，日照時間，降霜日数，平均気温が幾年にもわたつて，同一場所で観測され，収穫量がいろいろな気象状条によつてどの様な影響をうけるかを推定したい場合がある。

一般に数組の観測値 $x_1, x_2, \cdots x_n$; $y_1, y_2 \cdots y_n$; $t_1, t_2 \cdots t_n$ があり，次の型の重線型関係を推定したいとする。

$$y - y_0 = \beta_x(x - x_0) + \beta_t(t - t_0) + \cdots$$

この様な関係により他の変量と y の従属関係を測り，x , t ・・の対応する値が与えられた時 y の値を推定することができるのである。独立変量が一つの線型回帰の場合と同じく，回帰方程式を推定するには回帰を満足する任意の値を推定し，独立変量に対する従属変量の回帰係数を推定する必要がある。はつきりしている値は，前と同様観測値の算術平均である。

したかつて

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} \quad , \quad \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \quad , \quad t = \frac{\sum_{i=1}^{n} t_i}{n}$$

であれば，y_0, x_0, t_0 の代りに $\bar{y}$, $\bar{x}$, $\bar{t}$ が使われる。

係数 βx, βy は連立一次方程式から推定される。

$$b_x e\, var(x) + b_t e\, cov(x,t) + \cdots = e\, cov(x,y)$$

$$b_x e\, cov(x,t) + b_t e\, var(t) + \cdots = e\, cov(t,y)$$

或は次の方程式を使つて直接平方和、積和から簡単に推定することもできる。

$$b_x \sum_{i=1}^{n}(x_i-\bar{x})^2+b_t \sum_{i=1}^{n}(x_i-\bar{x})(t_i-\bar{t})^2+\cdots = \sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})$$

$$b_x \sum_{i=1}^{n}(x_i-\bar{x})(t_i-\bar{t})+b_t \sum_{i=1}^{n}(t_i-\bar{t})^2+\cdots = \sum_{i=1}^{n}(t_i-\bar{t})(y_i-\bar{y})$$

例として、１１，２１，２２，２３図で使つた収入に関する観測値について考えてみよう。

5.1a表　　１９４０年のアメリカ各州における

log(平均収入)，I，log(労働者百分率)L，

有色人種の割合　C，

州	I	L	C
Maine	2.729	1.591	0.003
New Hampshire	2.769	1.624	0.001
Vermont	2.742	1.595	0.001
Massachusetts	2.905	1.631	0.014
Rhode Island	2.903	1.654	0.016
Connecticut	2.962	1.654	0.020
New York	2.598	1.646	0.044
New Fersey	2.922	1.650	0.055
Pennsylvania	2.826	1.605	0.048
Ohio	2.851	1.602	0.049
Indiana	2.777	1.589	0.036
Illinois	2.888	1.629	0.050
Michigan	2.843	1.607	0.041
Wisconsin	2.756	1.592	0.008
Minnesota	2.732	1.596	0.008

州	I	L	C
Iowa	2.729	1.577	0.007
Missouri	2.740	1.604	0.065
North Dakota	2.634	1.565	0.016
Sowth Dakota	2.625	1.572	0.037
Nebraska	2.657	1.581	0.014
Kansas	2.666	1.570	0.037
Delaware	2.951	1.632	0.135
Maryland	2.869	1.624	0.166
Dist of Col	3.023	1.715	0.285
Virginia	2.686	1.586	0.247
West Virginia	2.629	1.523	0.062
N. Carolina	2.544	1.572	0.281
S. Carolina	2.491	1.585	0.429
Georgia	2.531	1.594	0.348
Florida	2.683	1.618	0.272
Kentucky	2.524	1.545	0.075
Tennessee	2.548	1.565	0.174
Alabama	2.483	1.555	0.347
Mississippi	2.377	1.568	0.493
Arkansas	2.465	1.542	0.248
Louisiana	2.596	1.573	0.360
Oklahoma	2.604	1.537	0.099
Texas	2.651	1.583	0.145
Montana	2.772	1.604	0.034
Idako	2.674	1.561	0.011
Wyoming	2.784	1.602	0.016
Colorado	2.745	1.574	0.015
New Mexico	2.575	1.524	0.074

州	I	L	C
Arizona	2.699	1.558	0.145
Utah	2.708	1.518	0.013
Nevada	2.923	1.639	0.056
Washington	2.849	1.616	0.022
Oregon	2.801	1.619	0.013
California	2.928	1.630	0.045

第1段階として、これらの値の和、平方和、積和を計算する。

$\Sigma L = 78.096$　　$\Sigma C = 5.180$　　$\Sigma I = 133.727$

$\Sigma L^2 = 124.541714$　　$\Sigma LC = 8.231496$　　$\Sigma LI = 213.360273$

$\Sigma C^2 = 1.329078$　　$\Sigma CI = 13.635331$

$\Sigma I^2 = 366.056775$

この値から4.2，4.3節の公式を使つて平均値および平均値の周りの平方和、積和を計算すれば

$\bar{L} = 1.594$　　$\bar{C} = 0.106$　　$\bar{I} = 2.729$

$\Sigma(L-\bar{L})^2 = 0.072628$, $\Sigma(L-\bar{L})(C-\bar{C}) = -0.024367$, $\Sigma(L-\bar{L})(I-\bar{I}) = 0.226726$

$\Sigma(C-\bar{C})^2 = 0.781478$　　$\Sigma(C-\bar{C})(I-\bar{I}) = -0.501523$

$\Sigma(I-\bar{I})^2 = 1.099417$

したかつて回帰係数を決めるための方程式は

$$0.072628\, b_L - 0.024367\, b_C = 0.226726$$

$$-0.024367\, b_L + 0.781478\, b_C = -0.501523$$

この方程式から

$$b_L = 2.9372$$

$$b_C = -5502$$

が得られる。したかつて推定された回帰線は

$$I - 2.729 = 2.9372(L-1.594) - 0.5502(C-0.106)$$

故に

$$I = 2.9372\,L - 0.5502\,C - 1.895$$

この関係は2.4節でグラフで求めたものと比較しよう。この3つの結果はよく一致していることが分る。

5.2　重回帰の分散分析　Analysis of variance for multiple regression

分散分析は多変量従属関係の有意性の検定および従属変量に含まれている原因不明の変動性を推定するために使われる。この方法は独立変量が1つの場合の回帰と同じであり、回帰に帰因する平方和は

$$b_x \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) + b_t \sum_{i=1}^{n}(t_i - \bar{t})(y_i - \bar{y}) + \cdots\cdots$$

これは独立変量の数だけの自由度をもつている。

全体の平方和は普通のようにして計算され、残差の平方和は引算で求められる。

例として前節で行つた推定の分散分析を考えよう。回帰に帰因する平方和は

$$2.9372 \times 0.226726 + (-0.5502)(-501523) = 0.941878$$

したかつて分散分析は5.2 a表の様になる。

5.2 a表　重回帰の分散分析

変動因	自由度	平方和	分散の推定値
LCに帰因	2	0.9419	0.4709
原因不明	46	0.1575	0.00342
計	48	1.0994	0.02290

分散比 0.4709／0.00342=137.7(自由度２，４６）は極めて有意である。Ｌ，Ｃを考慮したとき log（収入）の標準偏差が$\sqrt{0.00342}=\pm 0.0585$であることを、原因不明の分散は示している。Ｌ，Ｃを考慮した時収入の百分率で表わした変動性を表示するため、この値の真数が求められる。この値は大体１５％である。

この方法で行つている分散分析は重回全体の有意性を検定しているのであつて、回帰に含まれている特定の変量の有意性を検定するものでないことに注意を払わねはならない。これを行うには、さらに解析を行つてみる必要がある。

5.3　特定の変量に関する検定
Testing particular variables

ある変量が従属変量の予測に有意な貢献をしているか否かを検定するには、まず、その変量を含めた分散分析を行い、次いでその変量を除いた分析を行う必要がある。その方法ではその変量を含むための平方和を推定し分散比検定を使つて検べる。

前節の回帰にＣを含ませることの有意性を検定したいとする。Ｃを除いた時のＬに対するＩの回帰係数は 0.226726／0.072628=3.1217 であり分散分析における平方和は 3.1217×0.226726=0.7078 である。したがつてＣを含むため生する余分な部分は 0.9419−0.7078=0.2341 であり、分散分析は 5.3 a表の様になる。

5.3 a表　　重回帰の分散分析

変動因	自由度	平方和	分散の推定値
Ｌに帰因するもの	1	0.7078	0.7078
Ｃに帰因するもの	1	0.2341	0.2341
ＬとＣに帰因するもの	2	0.9419	−
原因不明	46	0.1575	0.00342
計	48	1.0994	0.02290

Ｃを含ませる効果を検べる分散比は 0.2341／0.00342=68.5（自由度 1, 46）であり、Ｆ表からこれは１％水準で有意である。したがつてＣは全体の回帰に有意な貢献をしている。

変量の数の如何を問わず、上記と同じ方法が使える。

5.4　分散分析を使つての回帰の比較
Comparison of regressions using the analysis of variance

有意差があるか否かを知るため一連の回帰を比較する必要のあることがある。これを行う方法を示すには次の例が役に立つであろう。3匹のねこの環状動脈流Ｆ，心耳部の血圧Ａ，肺動脈の血圧Ｐについての観測値を 5.4a 表は示している（E.W.H.Cruickshank 教授のデータ）

5.4 a表　　３匹のねこについての観測値

ねこ	Coronary flow F	Auricular pressure A	Pulmonary artery pressure P
I	8	7.0	27.5
	11	9.6	29.5
	14	12.2	32.0
	10	10.5	30.2
	9	8.0	27.0
	7	6.5	24.2
	4	5.0	22.5
	3	3.7	19.0
	5	6.8	24.8

	F	A	P
	3	6.8	22.5
	2	6.8	21.5
	0	6.8	22.0
Ⅱ	3	8.5	22.5
	9	9.8	24.0
	10	11.2	25.0
	14	11.2	25.5
	14	11.2	24.5
	14	11.2	24.0
	10	11.2	23.0
	7	11.2	22.0
	9	11.2	23.2
	13	13.5	25.0
	22	15.5	27.2
	10	11.8	23.5
	5	8.5	20.2
	5	5.8	17.0
	8	9.8	21.5
Ⅲ	17	11.2	29.5
	14	8.8	25.0
	15	10.5	27.5
	15	12.0	29.0
	17	13.8	31.5
	19	16.5	34.0
	12	12.2	30.0

AとPに対するFの従属関係が3匹全てについて同じであるかどうかを検べるには各動物について別々に回帰を求める必要がある。

これは次の3組の方程式で表わされる。

ねこⅠ；
$$59.64bA + 94.06bP = 85.9$$
$$94.06bA + 176.36bP = 171.3$$

ねこⅡ；
$$69.07bA + 66.29bP = 123.4$$
$$66.29bA + 83.13bP = 128.1$$

ねこⅢ；
$$36.52bA + 41.35bP = 22.0$$
$$41.35bA + 49.00bP = 25.0$$

この方程式から

	ねこⅠ	ねこⅡ	ねこⅢ
bA＝	−0.5765	1.3110	0.5555
bP＝	1.2788	0.4955	0.0414

3匹の分散分析が5.4b表に示してある。

5.4b表　3匹のねこに対する分散分析

変動因	ねこⅠ Df	Ss	Ev	ねこⅡ Df	Ss	Ev	ねこⅢ Df	Ss	Ev
AとPに帰因するもの	2	169.5	84.75	2	225.3	112.65	2	13.3	6.65
原因不明	9	23.2	2.58	12	89.1	7.43	4	18.4	4.60
計	11	192.7	17.52	14	314.4	22.46	6	31.7	5.28

Ⅶ表によれば、最も大きな比 7.43/2.58=2.88 は有意でないから*、この3つの方程式の原因不明の変動は比較しうる量であることが示されている。したがつてこの3つの解析は5.4c表の様に1つの解析にまとめることができる。

5.4c表　こみにした分散分析

変動因	Df	Ss	Ev
3つの回帰のAPに帰因するもの	6	408.1	68.02
原因不明	25	130.7	5.23
計	31	538.8	

一括した回帰に帰因する平方和を推定するためこの３組の方程式を加え合せこれを解く。

３組の方程式の和は

$$165.23\,bA + 201.70\,bP = 231.3$$

$$201.70\,bA + 308.49\,bP = 324.4$$

これから　$bA = 0.5756$

$bP = 0.6752$

したがつてこみにした回帰に帰因する平方和は

$0.5756\times231.3+0.6752\times324.4=352.2$　　自由度　2

これと回帰の和との差 408.1 − 352.2 = 55.9 は３匹のねこの回帰間の差を検定するのに使われる。したがつてこの解析は、5.4 d 表の様にして完成される。

5.4 d 表　　綜合分散分析

変　動　因	Df	Ss	Ev
こみした回帰に帰因するもの	2	352.2	176.10
回帰係数間の差に帰因するもの	4	55.9	13.98
回帰の和に帰因するもの	6	408.1	
原　因　不　明	25	130.7	5.23
計	31	538.8	

分散比 13.98/5.23=2.67 自由度４，２５　は５％有意水準 2.76 に達していない。したかつて回帰間の差は大きいけれど有意ではなく、偶然によるものであると結論される。

（＊注　疑問のある時には分散の一様性についての Bartlett 検定法を適用する必要がある。）

この解析は回帰係数間の差を検定するに過ぎず、一連の平行な回帰線の差は分らないことに注意せよ。この様な差の存在を検定するには、さらに解析を行う必要がある。観測値を全部一括して、ねこ間の差を無視して新しい回帰を求める。これは

$$282.45\,bA + 242.37\,bP = 428.1$$

$$242.37\,bA + 497.53\,bP = 485.0$$

$bA = 1.1670$

$bP = 0.4063$

分散分析は 5.4 e 表に示してある通りである。

5.4 e 表　　全観測値を使つた分散分析

変　動　因	自由度	平方和	分散の推定値
回帰に帰因するもの	2	696.6	348.30
回帰係数間の差（5.4 d 表より）	4	55.9	
原　因　不　明	27	165.4	13.98
計	33	917.9	

回帰線間の隔りを検定する平方和を求めるには、5.4 e 表の原因不明の平方和から 5.4 d 表の原因不明の平方和を引かねばならない。その値は

$$165.4 - 130.7 = 34.7$$

自由度は２である。したがつて最終的な分散分析は 5.4 f 表の通りである。

5.4 f 表　　最終的分散分析

変　動　因	自由度	平方和	分散の推定値
回帰に帰因するもの	2	696.6	348.30
回帰係数間の差に帰因するもの	4	55.9	13.98
回帰線間の隔りに帰因するもの	2	34.7	17.35
原　因　不　明	25	130.7	5.23
計	33	917.9	

回帰線間の隔りを検べる分散比 17.35／5.23=3.32 は5%水準（3.38）で有意でないが、その値はかなり大きいかさらに調べてみるのが妥当である。したかつてこの3匹のねこに関する観測値は一つの回帰式で表わせると結論される。

$$F - 9.94 = 1.1670(A-9.89)+0.4063(P-25.21)$$

$$F = 1.1670A + 0.4063P - 11.84$$

しかし、回帰線間の隔りを検べる平方和が大きければ3つの平行な回帰線を使わねはならない。

ねこⅠ $F-6.33 = 0.5756(A-7.46)+0.6752(P-25.22)$

$$F = 0.5756A + 0.6752P - 15.00$$

ねこⅡ $F-10.20 = 0.5756(A-10.77)+0.6752(P-23.21)$

$$F = 0.5756A + 0.6752P - 11.67$$

ねこⅢ $F-15.57 = 0.5756(A-12.14)+0.6752(P-29.50)$

$$F = 0.5756 + 0.6752 - 11.34$$

この場合には回帰係数は3組の回帰係数を加え合せてbA，bP について解いたものから計算されていることに注意せよ。

5.5 逆行列と回帰係数の標準誤差

The inverse matrix and standard error of regression coefficients

回帰係数の標準誤差を推定するためには、回帰係数の推定に用いられた方程式の係数の逆行列を計算する必要かある。その方法を述べる方程式を

$$a_{11}b_1+a_{12}b_2+a_{13}b_3+\cdots = A_1$$
$$a_{21}b_1+a_{22}b_2+a_{23}b_3+\cdots = A_2$$
$$a_{31}b_1+a_{32}b_2+a_{33}b_3+\cdots = A_3$$
$$\cdots\cdots\cdots\cdots$$

とする。

この場合 $a_{ij} = a_{ji}$

そこで、係数の行列は

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} & \cdots\cdots \\ a_{21} & a_{22} & a_{23} & \cdots\cdots \\ a_{31} & a_{32} & a_{33} & \cdots\cdots \\ \cdots\cdots & & & \end{pmatrix}$$

逆行列を求めるには、方程式の右辺の値を順次 1, 0, 0, 0, ····; 0, 1, 0, 0, ····; 0, 0, 1, 0, ····; でおきかえて考察する。どのような場合でもAは1個所を除いて0で置換され、この除外個所は1で置換される。

次の値を求めるため方程式を解く。

	第1回の解	第2回の解	第3回の解
$b_1 =$	c_{11}	c_{12}	c_{13}
$b_2 =$	c_{21}	c_{22}	c_{23}
$b_3 =$	c_{31}	c_{32}	c_{33}

逆行列はこの解の組から成つている。

$$\begin{pmatrix} c_{11} & c_{12} & c_{13} & \cdots\cdots \\ c_{21} & c_{22} & c_{23} & \cdots\cdots \\ c_{31} & c_{32} & c_{33} & \cdots\cdots \\ \vdots & \vdots & \vdots & \end{pmatrix}$$

$c_{ij} = c_{ji}$ 即ち行列は対称であるということを用いて計算を吟味する。

例として次の方程式を考えてみる。

$$
\begin{aligned}
b_1 + 2b_2 - 3b_3 - 2b_4 &= 4\\
2b_1 + 6b_2 - 2b_3 - 2b_4 &= 24\\
-3b_1 - 2b_2 + 18b_3 + 9b_4 &= 19\\
-2b_1 - 2b_2 + 9b_3 + 8b_4 &= 11
\end{aligned}
$$

係数の行列は

$$
\begin{pmatrix}
1 & 2 & -3 & -2\\
2 & 6 & -2 & -2\\
-3 & -2 & 18 & 9\\
-2 & -2 & 9 & 8
\end{pmatrix}
$$

逆行列を求めるため、4組の方程式を同時に解く。

	(1)	(2)	(3)	(4)
$b_1+2b_2-3b_3-2b_4 =$	1	0	0	0
$2b_1+6b_2-2b_3-2b_4 =$	0	1	0	0
$-3b_1-2b_2+18b_3-9b_4 =$	0	0	1	0
$-2b_1-2b_2+9b_3+8b_4 =$	0	0	0	1

1式を使つて2, 3, 4,式から b_1 を消去する。

	(1)	(2)	(3)	(4)
$2b_2+4b_3+2b_4 =$	−2	1	0	0
$4b_2+9b_3+3b_4 =$	3	0	1	0
$2b_2+3b_3+4b_4 =$	2	0	0	1

この方程式の1式を使つて2, 3式から b_2 を消去する。

	(1)	(2)	(3)	(4)
$b_3 - b_4 =$	7	−2	1	0
$-b_3+2b_4 =$	4	−1	0	1

最後に b_3 を消去すれば

	(1)	(2)	(3)	(4)
$b_4 =$	11	−3	1	1

逆解法により4つの解が得られる。

	(1)	(2)	(3)	(4)
$b_1 =$	173	−48	18	11
$b_2 =$	−48	13.5	−5	−3
$b_3 =$	18	−5	2	1
$b_4 =$	11	−3	1	1

したがつて逆行列は

$$
\begin{pmatrix}
173 & -48 & 18 & 11\\
-48 & 13.5 & -5 & -3\\
18 & -5 & 2 & 1\\
11 & -3 & 1 & 1
\end{pmatrix}
$$

方程式の左辺の係数だけがこの計算に用いられることに注意せよ。しかし、右辺の数値を全て含ませると逆行列が計算されるのと同時にこの方程式は解かれる。この解は $b_1 = 3$, $b_2 = 4$, $b_3 = 1$, $b_4 = 4$ であり これは自動的に逆行列の吟味となる。

$$
\begin{aligned}
b_1 &= 173\times4 - 48\times24 + 18\times19 + 11\times11 = 3\\
b_2 &= -48\times4 + 13.5\times24 - 5\times19 - 3\times11 = 4\\
b_3 &= 18\times4 - 5\times24 + 2\times19 + 1\times11 = 1\\
b_4 &= 11\times4 - 3\times24 + 1\times19 + 1\times11 = 2
\end{aligned}
$$

回帰方程式の係数の逆行列は回帰方程式の分散、共分散を推定するのに用いられる。

c_{ij} が i 行 j 列の要素を表わすものとすれば

$$
\begin{aligned}
\mathrm{cov}(b_i, b_j) &= c_{ij}\, var_x(y)\\
var(b_i) &= c_{ii}\, var_x(y)
\end{aligned}
$$

$var_x(y)$ は y に含まれる原因不明の変動を表わす。例えば、上の数値例では

$var(b_2) = 13.5\ var_x(y)$

$Cov(b_1\ b_2) = -48\ var_x(y)$

前節の観測値を使つてもつと実際的な例を示さう。前節で3匹のねこを一諸にして推定した回帰係数の誤差を求めるため，次の方程式を解く。

		(1)	(2)	(3)
282.45bA	+ 242.37bP	= 1	0	428.1
242.37bA	+ 497.53bP	= 0	1	485.0

これから次の値が得られる。

	(1)	(2)	(3)
bA =	0.006084	−0.002964	1.1670
bP =	−0.002964	0.003454	0.4063

即ち逆行列は

$$\begin{pmatrix} 0.006084 & -0.002964 \\ -0.002964 & 0.003454 \end{pmatrix}$$

これから

bA の標準誤差 $= \sqrt{(0.006084\times 5.23)}$

$= \pm\ 0.1784$

bP の標準誤差 $= \sqrt{(0.003454\times 5.23)}$

$= \pm\ 0.1344$

& $Cov(bA\ bP) = -0.002964\times 5.23$

$= -0.015501$

bA, bP の相関 $= \dfrac{-0.015501}{0.1784\times 0.1344}$

$= -0.646$

標準誤差は回帰係数の真値に対する範囲を定めるのに用いられる。b_A と b_P との間に高い負の相関のあることは，b_A が真値の過大推定値であれば b_P は過小推定値となることを示している。逆も亦真である。

5.6 推定値の誤差 Error of an estimated value

推定値に含まれる誤差を計る時には回帰係数を推定する方程式の係数の逆行列が必要である。一組の値（X，T，：：）に対応する y の値を推定したいとする。即ち

$$\bar{y} + b_x(X-\bar{x}) + b_t(T-\bar{t}) + \cdots\cdots$$

その分散は次式で示される

$$var_x(y)\left\{\frac{1}{n} + C_{xx}(X-\bar{x})^2 + C_{tt}(T-\bar{t})^2 + \cdots\text{（同様な平方の項）}\right.$$
$$\left. + 2C_{xt}(X-\bar{x})(T-\bar{t}) + \cdots\cdots\text{（同様な積の項）}\right\}$$

n は観測数であり，$var_x y$ は y の原因不明の分散である。

例として心耳の血圧8.5，肺動脈の血圧28.0に対応する穂状動脈流を推定したいとする。

この推定値は

$9.94 + 1.1670(8.5-9.89) + 0.4063(28.0-25.21)$

$= 9.94 - 1.1670\times 1.39 + 0.4063\times 2.79$

$= 9.45$

この値の推定分散は

$$5.23\left[\frac{1}{34} + 0.006084(1.39)^2 + 0.003454(2.79)^2 - 2\times 0.002964\right.$$
$$\left.(-1.39)(2.79)\right]$$

$= 5.23[0.02941 + 0.01175 + 0.02689 + 0.02299]$

$= 0.4761$

標準誤差は $\sqrt{(0.4761)} = \pm\ 0.69$ である。したかつて推定値の95%信

頼限界は

$$9.45 \pm 2.06(0.69) = 8.03 \text{ と } 10.87$$

追加観測値平均がこの期待値と一致しているか否かを検定するには上記の分散に、この観測値の平均値の分散を加える必要がある。例えば１０個の追加観測値が、Ｆ＝１２４，Ａ＝8.5，Ｐ＝２８.０であり、Ｆの期待値が上と同じく 9.45 であるとすれば、差 12.4−9.45=2.95 の分散は

$$\frac{5.23}{10} + 0.4761 = 0.9991$$

追加観測値の平均値の９９%限界は

$$9.45 + 2.79\sqrt{(0.9991)} = 6.66 \text{ と } 12.24$$

したがつて追加観測値は当てはめた回帰線とは有意に違つている。

5.7 異常観測値の検定 Testing extreme observations

異常観測値は前節の方法を使つて検定される。しかし独立変量が１つ以上の時でも 4.7 節の注意がそのまま適用される。異常観測値として疑がいうる先見的な理由かなければ、最も異常なものを除いて、全観測値を解析に使うべきである。

偏差の分散を計算するために、前と同じ様に観測値の分散から推定値の分散を引かねばならない。即ち偏差の分散は

$$var_{x,y}\left[1-\frac{1}{n}-C_{xx}(X-\bar{x})^2 - C_{tt}(T-\bar{t})^2 - \cdots - 2C_{xt}(X-\bar{x})(T-\bar{t})\cdots\right]$$

異常観測値を検定するには、一般に用いられる場合より高い有意水準を使うべきであるか、$var_x(y)$の推定値は偏差に独立でないから、限界を定める時に用いられるｔは近似的なものである。

したかつて、この方法による検定は、次の活動のための完全な過程というよりは、一つの指標に過きない。

例として、ねこＩの最後の観測値、Ｆ=0，Ａ=6.8，Ｐ=22.0 と考えてみよ

5。Ｆの期待値は

$$1.1670(6.8)+0.4063(22.0)-11.84 = 5.03$$

期待値からの偏差の分散は

$$5.23\left[1-\frac{1}{34}-0.006084(3.09)^2 - 0.003454(3.21)^2 - 2\times 0.002964(3.09)(3.21)\right] = 4.279$$

期待値の９９%限界は、大略

$$5.03 \pm 2.79\sqrt{(4.279)} = -0.74 \text{ と } 10.80$$

Ｆ＝０なる値はこの限界内にあり、従つてこの観測値は充分偶然変動の領域内にあることが分る。

5.8 重相関係数と偏相関係数

Multiple and partial crrelation coefficients

ある変量と他の数変量との一次関係の程度や他の変量を考慮した時その変量と任意の一つの変量との相関の程度を示すために重相関係数や偏相関係数を計算すると役に立つことがある。普通の相関係数は $r_{yx}{}^2 = x$ に帰因する y の全平方和の割合と定義されている。

これに対して、重相関および偏相関係数は次の様に定義される。

$R^2_{yxt} = x$，ｔ……に帰因する y の全平方和の割合

$r^2_{y\,tx} = x$ の効果を除いた時のｔに帰因する y の全平方和の割合

この量は分散分析表から直接計算される。例えば、5.3 a 表でＩ，Ｌ，Ｃについて計算した分散分析表から、次の係数が得られる。

$$r_{IL} = \sqrt{\left(\frac{0.7078}{1.0994}\right)} = 0.802$$

$$R_{ICL} = \sqrt{\left(\frac{0.9419}{1.0994}\right)} = 0.926$$

$$I_{IC\text{-}L} = \sqrt{\left(\frac{0.2341}{1.0994-0.7078}\right)} = 0.773$$

これは元の解析と分散比を使つて極めて便利に検定されるか，直接に検定を行う表を用いることも出来る。例えば観測数より1を減じた自由度のX表の値を用いて偏相関係数が検定され，同様に偏相関係数の組合せや比較のためZ変換が用いられる。

重相関係数および偏相関係数は原則的には連関の程度を示すために用いうべきである。大低の場合，同時に連関を推定するのが普通であるから，この係数は分散分析表から求めるのが最も便利である。しかし変量の各組合せ毎に求めた相関係数を使つて直接重相関，偏相関係数を計算することも出来る。このための公式は変量が3以下の場合を除いて，一般にその用法は複雑であり，時間のかかるものである。

この場合には

$$R_{y\,xt}{}^2 = \frac{\gamma_{yx}{}^2 + \gamma_{yt}{}^2 - 2\gamma_{xt}\gamma_{yx}\gamma_{yt}}{1 - \gamma_{xt}{}^2}$$

$$\gamma_{yt\,x} = \frac{\gamma_{yt} - \gamma_{yx}\gamma_{xt}}{\sqrt{(1-\gamma_{yx}{}^2)(1-\gamma_{xt}{}^2)}}$$

5.9 固有の関係の推定

Estimation of underlying relationships

数変量間の関係を推定する方法はかなり複雑で各測定値に含まれる相対的変動性を知る必要がある。したがつて，例えば3変量の時には

$$y = a_1\xi + b_1\eta + C_1 + \varepsilon_1$$

$$x = a_2\xi + b_2\eta + c_2 + \varepsilon_2$$

$$t = a_3\xi + b_3\eta + c_3 + \varepsilon_3$$

固有の関係は

$$(a_2b_3-a_3b_2)(y-c_1)+(a_3b_1-a_1b_3)(x-c_2)+(a_1b_2-a_2b_1)(t-c_3)=0$$

この関係を推定したいならば，$\varepsilon_1, \varepsilon_2, \varepsilon_3$ の相対的な大きさが分つていなければならない。

一般には

$$\lambda_1 = \frac{var(\varepsilon_2)}{var(\varepsilon_1)} \qquad \lambda_2 = \frac{var(\varepsilon_3)}{var(\varepsilon_1)} \quad \cdots\cdots$$

したがつて推定される関係は

$$y-\bar{y} = b_x(x-\bar{x})+b_t(t-\bar{t})+\cdots\cdots$$

この場合

$$b_x\sum_{i=1}^{n}(x_i-\bar{x})^2+b_t\sum_{i=1}^{n}(x_i-\bar{x})(t_i-\bar{t})+\cdots=\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})+\lambda_1 b_x S$$

$$b_x\sum_{i=1}^{n}(x_i-\bar{x})(t_i-\bar{t})+b_t\sum_{i=1}^{n}(t_i-\bar{t})^2+\cdots=\sum_{i=1}^{n}(t_i-\bar{t})(y_i-\bar{y})+\lambda_2 b_t S$$

$$S=\frac{\sum_{i=1}^{n}(y-\bar{y})-b_x\left[\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})-\lambda_1 b_x S\right]-b_t\left[\sum_{i=1}^{n}(t_i-\bar{t})(y_i-\bar{y})-\lambda_2 b_t S\right]}{1+\lambda_1 b_x{}^2+\lambda_2 b_t{}^2+\cdots}$$

この b_x，b_t ……S の方程式は逐次近似法で解くのが最もよい。b_xS, b_tS の大体の値をこの方程式の右辺に代入して b_x，b_t ……の値を求める。これはS 従つてb_x S，b_t S のさらに良い近似値を求め

るために使われ、これは順次この値は b_x, b_t の推定値を改善するのに使われる。確実な係数の値が得られるまでこの方法を繰返す。

この計算について次の3つの点に注意を要する。第1に係数の逐次近似値を求めるためにこの方程式の逆行列が使われる。第2にSの値に対する逐次近似は減小する。第3に b_x, b_t の最終的推定値は次の方程式を満足する。

$$S=\sum_{i=1}^{n}(y_i-\bar{y})^2-b_x\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})-b_t\sum_{i=1}^{n}(t_i-\bar{t})(y_i-\bar{y})-\cdots$$

これはこの計算全体の吟味にもなる。

例として、5.4節の生理学的観測値を考えてみる。$\lambda_1-\lambda_2=1$ の時、F，A，P間の関係を推定したいとする。係数 b_A, b_p は

$$282.45b_A + 242.37b_P = 428.1 + b_A\ S$$

$$242.37b_A + 497.53b_P = 485.0 + b_P\ S$$

から推定される。

ただし

$$S=\frac{917.9-b_A(428.1-b_AS)-b_P(485.0-b_P^S)}{1+b_A^2+b_P^2}$$

第1近似としてS＝b_A S＝b_P S＝0 とおく。その場合には、方程式は、

$$282.45\,b_A+242.37\,b_P=428.1$$

$$242.37\,b_A+497.53\,b_P=485.0$$

即ち

$$b_A = 1.1670$$

$$b_P = 0.4063$$

$$S=\frac{917.9-1.1670(428.1)-0.4063(485.0)}{1+(1.1670)^2+(0.4063)^2}$$

$$=87.56$$

したがつて第2近似値はS=87.56 b_AS=102.2 b_P S=35.6であり b_A と b_p を決める方程式は

$$282.45\,b_A+242.37\,b_P=530.3$$

$$242.37\,b_A+497.53\,b_P=520.6$$

となる。したがつて

$$b_A = 1.683$$

$$b_P = 0.226$$

$$S=\frac{917.9-1.683(325.9)-0.226(449.4)}{1+(1.683)^2+(0.226)^2}$$

$$=68.97$$

したがつて第3近似値はS＝68.97，b_A S＝116.1，b_P S＝15.6であり，方程式は

$$282.45\,b_A+242.37\,b_P=544.2$$

$$242.37\,b_A+497.53\,b_P=500.6$$

故に

$$b_A = 1.827$$

$$b_P = 0.116$$

$$S=\frac{917.9-1.827(312.0)-0.116(469.4)}{1+(1.827)^2+(0.116)^2}$$

$$=67.43$$

これ以上の近似値は5.9a表に示してある。

	4 th	5 th	6 th	7 th	8 th	9 th
S	67.43	67.19	67.11	67.06	67.06	67.03
b_A S	123.2	127.2	129.3	130.4	131.0	131.3
b_P S	7.8	4.6	3.0	2.2	1.8	1.6
b_A	1.893	1.927	1.945	1.945	1.959	1.961
b_P	0.068	0.045	0.033	0.027	0.024	0.023

第5近似値以降においてはSの減少は極めて緩慢であり、したがつてこの近似で求めた b_A , b_P の値が最終的な値の代りに用いられる。計算の吟味として

$$917.9-1.961(428.1)-0.023(4850) = 67.21$$

に注意する。この値はSの最終値とかなりよく一致している。したがつて推定された関係は

$$F - 9.94=1.961(A-9.89)+0.023(P-25.21)$$

即ち $$F- 1.961A-0.023P+10.03=0$$

Pは、AとPに対するFの回帰に有意な働きをおよぼすが、この関係の推定には極く僅かな貢献をするに過きないことに気付かねばならない。固有関係の推定では、この様なことは極く普通である。この様なことがどの様にして起るかを示すため、特殊な例を考えてみる。

$$y = 2\xi + 2\eta$$

$$x = \xi + \eta + \varepsilon$$

$$t = 2\eta$$

とする。

x と t に対する y の回帰は次の回帰方程式で示される。

$$y = t + b(x - \tfrac{1}{2}t)$$

bは2より小さく誤差 ε によつて変る。したがつて、$b = 1.2$ であり、

回帰方程式は

$$y = 1.2x + 0.4t$$

となる。

しかしこの場合固有の関係は

$$y = 2x$$

であり、tは表われていない。したがつて変量tは単に別の独立変量 x の変動 ε の結果として回帰方程式の内に入つている。しかし y の x との間の主な連関とは無関係である。

この方法で興味があるのは、解れるべき方程式が変量と同数の解をもつている点である。上記の方法はSが最小である解を取り上げたものである。例えば、上の方程式を満足するSの値は3つある、67，187，1444この内一番前のものは上記の解法で求まつたものである。

後に－12.3節－他の法が特定の意味をもつており、或る場合にはさらに精しく固有関係として用いられるであろう。たゞ一つの関係が必要な時には、最も明確に定められている上記の方法を選ぶ。

第6章　曲線的連関 Curvilinear Association

6.1　分散分析と非線型的連関

Analysis of variance and non－linear association

2つ以上の変量間の連関の検定は分散分析によつて行われる。観測値が独立変量の値によつて組に分けられておれば、従属変量の各組の平均値が有意に違つているか否かを検べるため、分散分析が行われる。

この様な差は従属変量と独立変量との連関の存在を示しており、この場合には連関の形について仮定を設ける必要はない。したがつて連関の一般的検定となる。

この型の検討の妥当性を確かめるには、従属変量の値とは無関係に組分けを行う必要がある。独立変量の各組毎に、数個の観測値がとられゝば、同じ独立変量の値をもつ観測値から成つている組を選び出すことが出来る。

非常に複雑な関係があると考えれない限り普通ほぼ同数の観測値からなる5～6組に分けるとよい。

6.1a　表に示してある観測値の解析で、この検定を使つた簡単な例を示そう。この表には11才の50人の小児の算術商arithmetic quotient（平均＝100）の値と5点法5－point scale で、測つた発育程度（0～4）が示してある。

6.1a表　発育程度に対する算術商　arithmetic quotient の値

総計 ＝ 5472

発育程度	0	1	2	3	4
算術商	103	120	112	125	132
	93	111	127	135	117
	91	106	130	121	121
	76	103	114	123	143
	84	78	105	100	123
	112	121	112	117	107
	90	114	118	136	135
	73	103	107	125	125
	95	79	108	96	125
	80		118		99
	90		94		
計	987	935	1245	1078	1227

この場合組は5つの発育程度であり、組間の差を検定するため分散分析が行われる。

組に対する平方和は、自由度4で

$$\frac{(987)^2}{11}+\frac{(935)^2}{9}+\frac{(1245)^2}{11}+\frac{(1078)^2}{9}+\frac{(1227)^2}{10}-\frac{(5472)^2}{50}=7,425.9$$

全体の平方和は自由度49で14,788.3で、その分散分析は6.1b表に示してある。

6.1b表　算術商の分散分析

変動因	自由度	平方和	分散の推定値
組間	4	7,425.9	1,856.5
原因不明	45	7,362.4	163.61
計	49	14,788.3	

分散比 1,8 5 6.5 ／ 1 6 3.6 1 ＝ 1 1.3 5 は 1％水準で有意であり、したがつて組間には有意差があり、発育と算術商とには連関があると結論された。

この検定は発育程度の順位を示すものではなく連関についての一般的検定であることに注意されたい。しかしこの検定は、より一般的な形が有意に一層よく表示しているか否かを決めるため、特定の形の連関に対する検定と組合して行うことが出来る。例えば、4章で説明した方法で、発育に対する算術商の線型回帰を計算すれば、回帰に帰因する平方和は 6,9 5 4.1 である。この値は 6.1 0 表に示してある様に、上記の解析と一緒に使われる。

6.1 0 表　次に行われた分散分析

変　動　因	自由度	平方和	分散の推定値
発育の線形効果に帰因するもの	1	6.9 5 4.1	6.9 5 4.1
発育の非線形効果に帰因するもの	3	4 7 1.8	1 5 7.2 7
組　　間	4	7,4 2 5.9	—
原 因 不 明	4 5	7,3 6 2.4	1 6 3.6 1
計	4 9	1 4,7 8 8.3	

組間分散の大部分は発育の線形効果に帰因させることが出き、発育効果の分散の推定値は原因不明の分散より大きくないことが分る。

実際非線型的効果の平方和は小さいから、直線以外の関係を当てはめる必要はない。

さらにこの検定法の応用例として 6 図にプロットしてある観測値を使ってみよう。この値は 6.1 d 表に、log (百分率で表わした心臓重量) で組分けして示してある。計算の便宜上、この表の W の値は 1 を減じて、また H は 1 を加えてある。

勿論この様なことをしたとしても分析の形は変らない。

6.1 d 表　若い鼠の百分率で表わした心臓重量とヘモグロビン含有率の対数値

log(百分率で表わした心臓量) W		log(ヘモグロビン含有率) H	
0,7 9		0,8 2	
0,8 1		0,7 7	
0,8 2	計 ＝ 4.9 6	0,6 9	計 ＝ 4,5 1
0,8 3		0,8 3	
0,8 5		0,7 1	
0,8 6		0,6 9	
0,8 8		0,6 7	
0,8 9		0,5 7	
0,9 0		0,6 7	
0,9 1	計 ＝ 6.3 3	0,6 7	計 ＝ 4,3 9
0,9 1		0,6 2	
0,9 2		0,6 0	
0,9 2		0,5 9	
0,9 3		0,6 6	
0,9 3		0,5 9	
0,9 4	計 ＝ 5,6 3	0,5 7	計 ＝ 3,5 5
0,9 4		0,5 7	
0,9 4		0,5 1	
0,9 5		0,6 5	
0,9 8		0,4 7	
1,0 0	計 ＝ 5,0 3	0,4 5	
1,0 1		0,4 1	計 ＝ 2,1 4
1,0 2		0,4 3	
1,0 2		0,3 8	
1,0 6		0,3 4	
1,1 0	計 ＝ 5,6 0	0,3 6	
1,1 4		0,3 8	計 ＝ 1,7 9
1,1 5		0,3 8	
1,1 5		0,3 3	
総計	2 7,5 5		1 6,3 8

組間の平方和は

$$\frac{(4.51)^2}{6}+\frac{(4.39)^2}{7}+\cdots\cdots+\frac{(1.79)^2}{5}-\frac{(16.39)^2}{29}=0.5484$$

全体の平方和は、0.6025である。

したがつて分散分析は6.1e表に示してある通りである。

6.1e表　log(ヘモグロビン含有率)Hの予備的分散分析

変動因	自由度	平方和	分散の推定値
組間	4	0.5484	0.1371
原因不明	24	0.0541	0.00225
計	28	0.6025	

組間には極めて有意な差があり、したがつて2つの測定値間には連関がある。しかし直線でこの連関が表わされるか否かを検定するには、回帰を計算する際各組の独立変量の値を、組平均でおきかえる必要がある。これは組間の変動を検べる時には同じ組の観測値は区別出来ないからである。

さらに検定を行うには、組間の平方和、積和を計算しなければならない。

$$\frac{(4.96)^2}{6}+\frac{(6.33)^2}{7}+\cdots\cdots+\frac{(5.60)^2}{5}-\frac{(27.55)^2}{29}=0.2669$$

$$\frac{(4.96)(4.51)}{6}+\frac{(6.33)(4.39)}{7}+\cdots\cdots+\frac{(5.60)(1.79)}{5}-\frac{(27.55)(16.38)}{29}=-0.3742$$

Wの線型効果に帰因する組間の平方和は、$(-0.3742)^2/0.2669=0.5246$である。したがつて分散分析は6.1f表に示してある様に拡張される。

6.1f表　拡張された分散分析

変動因	自由度	平方和	分散の推定値
Wの線型効果に帰因するもの	1	0.5246	0.5246
Wの非線型効果に帰因するもの	3	0.0238	0.00793
組間	4	0.5484	—
原因不明	24	0.0541	0.00225
計	28	0.6025	

この場合にはWの非線型効果を検定する分散比0.00793／0.00225＝3.52は5%水準で有意であり、WとHの従属関係を有意に一層良く表現するには曲線を用いるとよい。これらの観測値に曲線を当てはめる方法は次節で説明する。

上記の分析において次の2点は注目の価値がある。第1に、各組の全観測値がその独立変量に対して同じ値をもつている時には組間の差に帰因する平方和は、独立変量に関する回帰による平方和よりも大である。

組間の平方和と全体の平方和との比の平方根は相関比と云われている。例えば6.1b表から計算した相関比は、

$$\sqrt{\left(\frac{7.4259}{14.7883}\right)}=0.709$$

この相関比は必然的に同じ観測値から計算した相関係数或は重相関係数より大でなければならない。

全回帰は組内変動の一部を説明するに過ぎないから組内の観測値がその独立変量に対して正確に同じ値でない場合にはこの様な説明は正しくない。そこで相関比は曲線回帰で説明される全変動の割合の指標として役立つ。

6.2　曲線回帰式の推定

Estimation of curvilinear regression equations

曲線回帰式の推定は重回帰方程式の推定と全く同じである。

仮にXに対するyの三次の従属関係を推定したいならば、x, x^2, x^3に対するyの重回帰を計算する様にして、解析が行われる。

したがつて推定される回帰方程式は

$$y-\bar{y}=b_1(x-\bar{x})+b_2(x^2-\bar{x}^2)+b_3(x^3-\bar{x}^3)$$

$$b_1\Sigma(x-\bar{x})^2+b_2\Sigma(x-\bar{x})(x^2-\bar{x}^2)+b_3\Sigma(x-\bar{x})(x^3-\bar{x}^3)=\Sigma(x-\bar{x})(y-\bar{y})$$

$$b_1\Sigma(x-\bar{x})(x^2-\bar{x}^2)+b_2(x^2-\bar{x}^2)+b_3\Sigma(x^2-\bar{x}^2)(x^3-\bar{x}^3)=\Sigma(x^2-\bar{x}^2)(y-\bar{y})$$

$$b_1\Sigma(x-\bar{x})(x^3-\bar{x}^3)+b_2(x^2-\bar{x}^2)(x^3-\bar{x}^3)+b_3(x^3-\bar{x}^3)=\Sigma(x^3-\bar{x}^3)(y-\bar{y})$$

この方程式の係数は4章の簡略式を使つて計算される。例えば

$$\Sigma(x-\bar{x})(x^3-\bar{x}^3)=\Sigma x^4-\frac{(\Sigma x)(\Sigma x^5)}{N}$$

$$\Sigma(x^3-x^3)=\Sigma x^6-\frac{(\Sigma x^5)^2}{N}$$

6.1d表のデータから

$\Sigma W=27.55$　　$\Sigma W^2=26.4517$　　$\Sigma HW=15.1797$

$\Sigma W^3=25.67484$　　$\Sigma W^4=25.197434$　　$\Sigma HW^2=14.20300$

$\Sigma W^5=25.0048422$　　$\Sigma W^6=25.0889989$　　$\Sigma HW^3=13.423509$

$\Sigma W^7=25.4479189$　　$\Sigma W^8=26.0853665$　　$\Sigma HW^4=12.8204949$

この値は4次の曲線を適合させる回帰方程式を計算するのに用いられる。

$0.279200b_1+0.545720b_2+0.806341b_3+1.067280b_4=-0.381300$

$0.545720b_1+1.070108b_2+1.586117b_3+2.105725b_4=-0.737649$

$0.806341b_1+1.586117b_2+2.358062b_3+3.139645b_4=-1.078346$

$1.067280b_1+2.105725b_2+3.139645b_3+4.191895b_4=-1.411711$

この方程式から

$b_1=20.0913$

$b_2=0.2646$

$b_3=-25.0192$

$b_4=13.1538$

4次曲線に帰因する平方和は

$(20.0913)(-0.381300)+(0.2646)(-0.737649)$

$+(-25.0192)(-1.078346)+(13.1538)(-1.411711)=0.5539$

したがつて分散分析は6.2a表に示してある様になる。

6.2a表　4次曲線の適合検定のための分散分析

変動因	自由度	平方和	分散の推定値
Wの4次式に帰因するもの	4	0.5539	0.13848
原因不明	24	0.0486	0.00202
計	28	0.6025	−

4次式による平方和は前節で使つた組分けによるものより幾分大きくなつていることが分る。この値は明らかに極めて有意である。しかし、この有意性は3次又は5次式より4次式を使う方がよいことを示しているのではない。当てはめるべき方程式の適当な次数を決めるには、直線より2次式、2次式より3次式 ····· を当てはめることによりどの様に良くなつて行くかを決める必要がある。3次式の代りに4次式を当てはめた結果、有意な改良が得られなかつたならば、3次式を使うべきである。多項式の当ては

めは逐次法で行うのがよい。2次、3次の項の有意性は、解析を行いつつ検定される。

6.3 逐次検定 Step-by-step testing

逐次検定法を前節の例を使つて説明しよう。

当てはめるべき多項式の次数を決めるため逐次検定を行うには、必要と考えられる大体の次数を決め、高次の多項式の係数を推定するため、方程式を作る必要がある。したがつて、3次の多項式が必要であると考えられゝば、4次の多項式の係数を推定するための方程式が作られる。

$$0.279200b_1+0.545720b_2+0.806341b_3+1.067280b_4=-0.381300$$
$$0.545720b_1+1.070108b_2+1.586117b_3+2.105725b_4=-0.737649$$
$$0.806341b_1+2.105725b_2+2.358062b_3+3.139645b_4=-1.078346$$
$$1.067280b_1+2.105725b_2+3.139645b_3+4.191895b_4=-1.411711$$

Wの一次効果による平方和は $(-0.381300)^2/0.279200$ $=0.5207$ で、原因不明の平方和は $0.6025-0.5207=0.0818$ 即ち原因不明の分散は $0.0818/27=0.000303$ である。したがつて、一次効果は明らかに有意であり、2式から 1式×0.545720／0.279200、3式から1 × 0.806341／0.279200 を引くことにより、上記方程式から b_1 が消去される。

この様にして求めた方程式は

$$0.003452b_2+0.010055b_3+0.019636b_4 = 0.007633$$
$$0.010055b_2+0.029316b_3+0.057297b_4 = 0.022864$$
$$0.019636b_2+0.057297b_3+0.112072b_4 = 0.045860$$

Wの二次の効果による平方和は $(0.007633)^2/0.003452$ $=0.0169$ で、原因不明の平方和は $0.0818-0.0169=0.0649$

即ち原因不明の分散は $0.0649/26=0.000250$

したがつて二次の効果は有意であり、 1式に適当な乗数を掛けたものを引くことにより b_2 は2， 3式から消去される。

この様にして

$$0.000028b_3+0.000101b_4 = 0.000628$$
$$0.000101b_3+0.000377b_4 = 0.002436$$

Wの三次の効果による平方和は $(0.000628)^2/0.000028$ $=0.01411$ で原因不明の平方和 $0.0649-0.01411 = 0.0508$ 即ち原因不明の分散は $0.0508/25=0.000203$ であるる。したがつて三次の効果は有意であり、1）式に $0.000101/0.000028$ を掛け、 2式から引けば、この方程式から b_3 が消去される。

$$0.000013b_4 = 0.000171$$

したがつてWの4次の効果による平方和は、

$(0.000171)^2/0.000013 = 0.0022$

原因不明の平方和は

$0.0508-0.0022 = 0.0486$

即ち原因不明の分散は

$0.0486/24 = 0.000202$ である。したがつて4次の効果は原因不明の分散より僅かに大であるに過きない。従つて有意ではない。三次式は、観測値に充分適合しているといえる。

さて、解析は、$b_4=0$ とおき他の係数を推定するため最初の3つの式を使つて行われる。 この様にして

$$b_1 = 57.2303$$
$$b_2 = 63.1186$$
$$b_3 = 22.4286$$

が求められ、方程式は

$$H-0.565 = 57.2303(W-0.95000)-63.1186(W^2-0.91213)+22.4286(W^3-0.88534)$$

即ち

$$H=-16.088+57.23W-63.12W^2+22.43W^3$$

分散分析の最後の形は6.3a表の通りである。

6.3a表　最終的分散分析

変動因	自由度	平方和	分散の推定値
Wの一次の効果	1	0.5207	0.5207
Wの二次の効果	1	0.0169	0.0169
Wの三次の効果	1	0.0141	0.0141
Wの全効果	3	0.5517	—
原因不明	25	0.0508	0.00203
計	28	0.6025	—

Wの4次の項が有意であれば、方程式は5次の項を含む様に拡張され、これは同様な方法で検定されることに注意せよ。

6.4　一般的な回帰分析　General regression analysis

実際曲線回帰分析は独立変量を x , x^2, x^3-------- とした重回帰分析の特殊な場合に過ぎない。分析は従属変量の誤差に関係があるに過ぎないから、この様な独立変量の選択は可能である。この結果、一般に独立変量の選択には広い巾がある。

第1には独立変量の函数が独立変量として使つてもよい。したがつて y

の回帰は、$\log x, e^x \sin x$ 等の函数或は函数を組合せたものに対して求められる。特に既知の期間における周期性の解析は重回帰分析に独立変量として $\sin x, \cos x, \sin 2x, \cos 2x,$ ------を使つて行われる。

第2に、2つ以上の独立変量の函数は独立変量として使つてもよい。この場合には $x_1x_2, x_1x_2^2, x_1x_2x_3, e^{x_1}\log x_2$ の様な項が回帰分析に使われる。特にこうすれば他の変量に対する回帰係数の従属関係を検べることもできる。例えば x_1 に対する y の一次回帰係数が別の変量 x_2 と関係がある様に考えられるとする。

x_1 と x_1x_2 に関する y の回帰分析をおこなわねばならない。回帰方程式は

$$y = a+bx_1+cx_1x_2$$
$$= a+(b+cx_2)x_1$$

となる。

したがつて積の項 x_1x_2 の係数は回帰係数の不変性その他を示すのに役にたつ。

最後に任意の独立変量の組をそれと一次関係のある別の組で置き換え、その組の値を第2の組から類推してもよい。例えば、独立変量 $x_1x_2x_3$ に関する回帰分析は z_1, z_2, z_3 に関する知識が $x_1\ x_2\ x_3$ を決めるに充分なものであれば当然次の変量を使つて行なつてもよいであらう。

$$z_1 = a_1x_1+b_1x_2+c_1x_3+d_1$$
$$z_2 = a_2x_1+b_2x_2+c_2x_3+d_2$$
$$z_3 = a_3x_1+b_3x_3+c_3x_3+d_3$$

このことは分析に必要な計算を減らすのに極めて有用な場合が多い。

例えば x と x^2 に関する回帰分析の代りに $(x-a)$ と $(x-a)^2$ に対

する回帰分析が使われ、さらに一般的には、xに対する任意の一次および2次函数が使われる。この様にすれば、比較的に取扱いやすい量で作業するため計算は非常に減つてくる。この様な方法を利用した重要な実例を次節で示す。

6.5 直交多項式を使つた傾向線の当てはめ
Trend fitting using arthogonal polynomials

曲線回帰の独立変量が等間隔で配列されている時には、直交多項式を利用すれば多変式の当てはめおよび検定は非常に簡単になる。直交多項式の性質、内容について詳しく説明することは本書の範囲を逸脱するものであるからその使用方法について簡単に説明するに留めよう。

前節で指摘した様に、x, x^2, x^3に関する回帰は、その一般性を失なはすにxの一次、二次、三次の函数の組で置き換えられる。直交多変式はこの様なxの一次、二次、三次の函数を表わすものである。

一般にこの函数はギリシヤ文字ξ_1, ξ_2, ξ_3で表わされ、ξ_1はxの一次函数を、ξ_2は二次の函数……を表わす。

多項式のもつ特殊な性質は特定の観測数に対して

$$\Sigma \xi_i = 0 \qquad \Sigma \xi_i \xi_j = 0 \quad (i \neq j)$$

この意味は回帰係数の推定方程式に含まれる項は主対角線を除いて0であるということである。この結果各回帰係数および分散分析に含まれるこれに対応する平方和は直接かつ独立に推定される。

傾向の当てはめや検定の際の計算を楽にするため、平方和$\Sigma \xi_i^2$と一緒にいろいろなnの値に対するξ_iの表が作られている。Fisher, Yatesの統計数値表 Statistical Tables(Edinburgh, 1952) には

n=3~75　i=1~5に対する直交多項式の値が示してある。附録のXII表にはn=3~12　i=1~3に対する値が示してある。

この表の使用法を示すため、収穫量に関する観測値yを考えてみる。この解析は6.5a表に示してある。平均収量から、オ1段階として積和$\Sigma \xi_1 y$を計算する。このためにξ_1の値を直接XII表から求める。例えば

$$\Sigma \xi_2 y = 7(86.9) + 1(83.5) - 3(81.2) - \cdots\cdots + 7(131.6)$$
$$= 235.3$$

6.5a表

1866年~1945年におけるU.S.での馬鈴薯の平均収量

(エーカ当りブツセル)

期間	収量(y)	あてはめた値 $3.122\xi_1 + 1.401\xi_2 + 98.01$	i	$\Sigma \xi_i y$	$b = \Sigma \xi_i y / \Sigma \xi_i^2$	$\Sigma(\Sigma \xi_i y)^2 / \Sigma \xi_i^2$
1866-75	86.9	86.0	1	524.5	3.122	1637.5
76-85	83.5	83.8	2	235.3	1.401	329.6
86-95	81.2	84.4	3	-45	—	0.1
96-05	88.2	87.9				
1906-15	99.7	94.1				
16-25	101.7	103.2				
26-35	111.1	115.0				
36-45	131.6	129.7				
総平均	98.01	98.01				

変動因	自由度	平方和	分散の推定値
時間の一次の効果	1	1,637.5	1,637.5
時間の二次の効果	1	329.6	329.6
時間の三次の効果	1	0.1	0.1
残差	4	63.2	15.80
計	7	2,030.4	

$\Sigma \xi_2^2$ の値を使えば－この値もXII表にのせてある－回帰係数およびこれに対応する平方和が推定出来る。例えば

$$b=235.3/168 = 1.401$$

時間の二次の効果による平方和は

$$(235.3)^2/168 = 329.6$$

分散分析は6.5a表に示してある様に直接行われ、各項の有意性が検定される。

この場合、一次および二次の項は極めて有意であるが三次の項は無視出来る。したがつてこの傾向は二次曲線であてはめられる。当てはめた傾向は係数の推定値と総平均を使つて求められる。これは $3.122\xi_1+1.401\xi_2$ $+98.01$ であり、例えば4番目の数のあてはめた値は

$$3.122(-1)+1.401(-5)+98.01 = 87.9$$

6.5a表に示してあるあてはめた値はこの方法で求めたものである。

最後にx即ち期間数で傾向式を表わすには ξ_1, ξ_2 を置換する必要がある。この場合 $\xi_1=2x-9$, $\xi_2=x^2-9x+15$ とすれば、傾向式は

$$y=3.122(2x-9)+1.401(x^2-9x+15)+98.01$$
$$=1.401x^2-6.365x+90.93$$

回帰係数およびこれを組合せたものの標準誤差は、それらが互に独立であるから簡単に計算される。例えば次の10年間、1946～55年の馬鈴薯の平均収量の推定値は上式に $x=9$ を入れることによつて求められる。即ち推定値は

$$3.122(9)+1.401(15)+98.01 = 147.12$$

この推定値の分散は

$$9^2 \mathrm{var}(b_1)+15^2 \mathrm{var}(b_2)+\mathrm{var}(\bar{y})$$
$$=81\times\frac{15.80}{168}+225\times\frac{15.80}{168}+\frac{15.80}{8}$$
$$=30.75$$

この結果標準誤差は±5.55である。したがつて1946～55年における馬鈴薯の平均収量の推定値の99%信頼限界は

$$147.12\pm(4.60\times5.55) = 121.6 \text{ と } 172.7 \text{ である。}$$

しかし、これは平均収量の観測値の限界とはならない。これを求めるには標準誤差の計算で求めた推定値の分散に個々の観測値の分散を加えねばならない。したがつてこの値は $\sqrt{30.75+15.80}=\sqrt{46.55} = \pm 6.83$ 故に1946～55年における馬鈴薯の平均収量の観測値の99%信頼限界は

$$147.12\pm(4.60\times6.83) = 115.70 \text{ と } 178.44 \text{ である。}$$

この関係を補外する時には、当然この関係が連続であるという仮定が設けられるが、この場合には、次の10年間について補外したとしてもそう不正確な結果とはならないであらう。しかし条件の変化或は技術の革命的変化は上記の限界を正確な値というより一つの指標に変えてしまうであらう。

観測値の組分け Grouping of Observations

7.1 効率の悪い方法 Inefficient methods

ある関係を研究、推定するのに非常に多くの観測値を用いれば、完全な統計解析をするには非常に手間がかゝる。それにもかゝわらず、これが資料を最も有効に利用する方法であればこの様な解析が必要であらう。しかし、手間や費用をそうかけずに追加観測値が求められる場合には、効率は悪いが簡便な解析法を用いると好都合なことがある。

例えば１００個の観測値から完全解析でえられるのと同じ精度の推定値が２００個の観測値を使つて求められる簡略解析法を使用することはやつてみるだけの価値があらう。各観測値は完全解析で得られる情報の５０％の寄与をするに過ぎないからこの様な方法は効率５０％であるといつても良いであらう。それにもかゝわらず、効率が劣る代りに解析が迅速であるため、簡略法が推奨されている。

迅速ではあるが効率の悪い回帰分析法の簡単な例として一次回帰分析の上下四分法 upper-and lower quartile (u.l.q.)法を説明しよう，独立変量における大きい方の２５％と小さい方の２５％の観測値の平均を使つて一次回帰が推定される。この二点を結んだ線が回帰の推定値を与える。

この様な回帰係数の推定方法は完全解析の約８０％の効率である※。即ち完全解析を使えば８０個で達しえた精度にこの方法で達するには１００個の観測値が必要である。この結果効率の低下はかなり小さく、一方解析の時間が節約されるという観点からいえば極めて価値のある場合が多い。

標準誤差を推定するにはさらに解析をしなければならないことが

※注 ８０％もつと正確にいえば80.7%という値は独立変量が正規分布しているという仮定に基ずいて出されている。しかし他の条件の下でも、これは効率には良い指針となる。例えば短形分布の場合はこの方法の効率は８４.４％であるし、$f(x)=\frac{1}{2}\exp(-x)$ の分布に対しては７１.７％である。

上記の方法の主な欠点である。しかし大低の場合標準誤差の迅速であるが大略の推定を求めることができる。したがつて を標準偏差の推定値（２５節で説明した方法で求められる）nを各平均値の計算に使つた観測数、dを独立変量の平均値間の差とすれば

$$\frac{s}{d}\sqrt{\frac{2}{n}}$$

は回帰係数の標準誤差の推定値である。同様に、回帰線上にある両端の点の中間にある点をとつた時の誤差はその標準誤差の推定値を用いて計ることが出来る。:

$$\frac{s}{2}\sqrt{\frac{2}{n}}$$

sを推定する簡略法の一つは独立変量の値が同じ様な値をもつていい一対の観測値の差を使つて行なわれる。y_1，y_2 が同じ値の独立変量を持つているとすれば $(y_1-y_2)^2/2$ が Δ^2 の推定値である。この様な推定値の加え上げればかなり正確な Δ^2 の値が求められるのであらう。

例えば 44a表の頭と胸の周囲のデータでは最初の３１対と最後の３１対の観測値の平均は

才１組： $\bar{H}=44.96$　$\bar{C}=42.12$

才２組　$\bar{H}=46.77$　$\bar{C}=47.21$

Cに関するHの回帰線は

$$H-\frac{44.96+46.77}{2}=\frac{46.77-44.96}{47.21-42.12}\left(C-\frac{42.12+47.21}{2}\right)$$

即ち

$$H-45.87=0.3556\,(C-44.66)$$

$$H=0.3556C-2.999$$

より精確な推定値 $H=0.3401C+3.060$ と比較してみよ。

Δ^2を推定するには、同じ値の胸囲をもつ観測値の対を使えば

$$\Delta^2=\frac{1}{50}\left[(458-454)^2+(450-450)^2+(460-442)^2+\cdots\right.$$

$$\left.+(465-448)^2\right]=1.767$$

$$\Delta=1.329$$

この場合Δの自由度は５０であり、一般に必要とされる数より多い。普通１０～１２の自由度で充分である。

さてbの標準誤差の推定値は

$$\frac{1.329}{509}\sqrt{\frac{2}{31}}=\pm 0.006632$$

より精確な推定値±0.005522　と比較してみよ

平均胸囲４４６６に対応する頭の平均周囲の推定値４５８７の誤差は

$$\frac{1.329}{2}\sqrt{\left(\frac{2}{31}\right)}=\pm 0.1688$$

この２つの誤差を使えば推定値に対する限界が定められる。

一般に推定値の誤差が同時に計られる様な便利な解析方法を使うことが望まれている。

次節には簡単な回帰分析法として観測値の組分けが示されている。

７２　線型回帰分析の Quantile 法　Quantile method of linear regression analysis

ほぼ同数の観測値を含むいくつかの組にデータを分け、普通の回帰分析に各組の平均を用いることが多数のデータを使つて一次回帰分析を行う場合最も有効な方法である。これが Quantile 法と呼んでよいであらう。

多の組数をもつと多くすれば解析は当然効率がよくなるが解析を行う時には仕事の量も増加する。

いろいろな組数を使う時の効率の規準が、７２ａ表に示してある。独立変量は正規分布をしており、多数の観測値がとられるという仮定にこの表は基ずいている。しかし、この主要な性質は別の仮定の

もとでも正しい。

2組に分割した様な時でさえ63.7% の効率であり、5組に分割すれば求めうる全情報の90%がえられることが分る。

7.2 a表 Quantile 法の効率

群の数	効率 %
2	63.7
3	79.3
4	86.1
5	89.7
6	91.9
7	93.4
8	94.5
9	95.3
10	95.6

観測数が少い時には、功率は上に示したものより高くなる傾向がある。例として5組をとると1mを各群の平均観測数とすれば、効率は大体89.7 + 10.3/mである。観測値が25であれば各組に含まれる平均数は5であり、したがつて効率は略91.8%である。

かなり高い効率が得られ又回帰線をあてはめた後、推定値の誤差を求めるのに3の自由度が便えるので、一般には5がデータを分類するのに適切な組数である。しかしあてはめた線と別の線との有意差を決めるため効果的な検定をする必要があれば、組数をより多くすることが望ましい。推定値の精度は著しく変りはしないが、この様にすれば、推定値の誤差の推定により大きな自由度が使用できるであろう。

この方法の便用例として4.4 a表の頭と胸の周囲のデータに対する回帰線のあてはめを考察してみる。このデータの5点直線は既に8図に示されている。胸囲に 従つて5組に分けた頭と胸の周

囲の平均値は7.2 b表に示してある。

7.2 b表　胸囲に従つて組分けした頭と胸の周囲の平均値

観測数	平均胸囲 C	頭の平均周囲 H
28	42.00	45.04
25	43.74	45.48
22	44.82	45.72
25	46.04	46.32
23	47.45	46.78

5組の平均値の解析は7.2 C表に示してある様に行う。

7.2 C表　頭と胸の周囲の解析

$\bar{H} = 45.85$　　$\bar{C} = 44.81$

$\sum (H-\bar{H})^2 = 1.9891$　　$\sum (H-\bar{H})(C-\bar{C}) = 5.8217$

$\sum (C-\bar{C})^2 = 17.5236$

$b = 0.3322$

$H = 0.3322C + 30.96$

	自由度	平方和	分散の推定値
回帰	1	1.9341	
残差	3	0.0550	0.01833
計	4	1.9891	

$$\text{levar}(b) = \frac{0.01833}{17.5236} = 0.001046$$

bの標準誤差 = ± 0.03234

この様な回帰線の推定値はより正確な推定値 H=0.3401C+30.60 と余り違つていないことが分る。さらに自由度は少いけれど、標準誤差はこの様にしてあてはめた線の精度について大略の概念を与える。したがつて、例えば回帰係数の95%信頼限界は正確に

あてはめられた線の 95 %信頼限界 0.2305 と 0.4434 に対して

0.3322±3.18×0.03234=0.2294 と 0.4350

である。

自由度の減少、それに伴う標準誤差の精度の減少は一般に信頼限界の巾を広くするから、この場合の一致性は普通期待されるものに較べて非常に良い。こゝで自由度 3 と 121 に対する t の値 3.18 と 1.98 を較べてみると、くみわけした観測値は組分けしない観測値を用いた時に較べて 95 %信頼限界が少くとも 60 %広くなることが期待されるということが分る。

※注 これは 99 %の信頼限界に対しても真である。99 %信頼限界は 45 節で得られた 0.154 と 0.526 の値に較べれば 0.086 と 0.751 となる。

前節で説明した方法で標準誤差の推定値を改良するためには同じ様な値をもつ観測値が用いられる。これを行う時には組平均の分散を求めるためこの方法で求めた分散の推定値を各組の平均観測数は全体で 121、即ち、組当り 2.42 であるから、平均値の分散は大体 1.767/242=0.07302 で改良された回帰係数の標準誤差の推定値は

$$\sqrt{\frac{0.07302}{172.536}} = \pm 0.06455$$

この値は自由度 50 であるから、回帰係数の 95 %信頼限界は

0.3322±2.00×0.06455=0.2031 と 0.4613

73 各組の観測数を等しくした組分け

最もよく知られている組分けの方法は、両変量に対して組間隔を等しくすることであり、この様にすることにより分割表として知られている表が作られる。73a表は 44a 表の頭と胸の周囲の測定値についての分割表の例である。

73a表では、胸囲は 2 cm の組に、頭の周囲は 1 cm の組にわけられている。各組に含まれる観測値は全然差別されておらず、いづれも夫々の間隔の中心にあるものとして処理される。

分割表の解析を行うときには観測値を一定量だけ減すことが望まれる。この様にすれば一般に推定や有意性の検定に必要な計算量は減つてくる。

73a表 胸囲(C)と頭の周囲(H)の分割表

H \ C	40-	42-	44-	46-	48-50	計
42-		1	1			2
43-	3	1	2			6
44-	4	11	4	4	1	24
45-	5	6	14	1	2	30
46-	2	6	11	13	1	33
47-		1	4	14	1	20
48-		1	3	3		7
49-						0
50-51					1	1
計	14	27	41	35	6	

例えば、胸囲の値を 45 cm 減すと各組の中心は -4 -2 0 2 4 となる。同様に頭の周囲を 45.5 cm 減すと各組の中心は -3 -2 -1 0 1 2 3 4 5 となる。この値を用いた解析は 73b 表の通りに行なわれる。

73b表 分割表の解析

H \ C	40- -4	42- -2	44- 0	46- 2	48	計	積和
縮約した値							
42- -3		1	1			2	-2
43- -2	3	1	2			6	-14
44- -1	4	11	4	4	1	24	-26
45- 0	5	6	16	1	2	30	-22
46- 1	2	6	11	13	1	33	10
47- 2		1	4	14	1	20	30
48- 3		1	3	3		7	4
49- 4						0	0
50-54.5					1	1	4
計	14	27	41	35	6	123	-16
積和	-8	-5	17	46	7	57	162

各行および列の簡約値の積和および頻度はこの表で計算される。例えば

$$3\times(-4)+1\times(-2)+2\times0=-14$$

$$14\times(-4)+27\times(-2)+41\times0+35\times2+6\times4=-16$$

次に各測定値の平方和を計算しなければならない。この計算は真く簡単である。

$$\sum C^2=14\times(-4)^2+27\times(-2)^2+41\times0+35\times(2)^2+6\times(4)^2=568$$

$$\sum H^2=2\times(-3)^2+6\times(-2)^2+\cdots\cdots+1\times(6)^2=267$$

回帰分析或は相関分析に関係のある数値が計算される。

$$\bar{C}=45-16/123=44487$$

$$\bar{H}=455+57/123=45496$$

$$\sum(C-\bar{C})^2=568-(-16)^2/123=56572$$

$$evar(C)=46387$$

$$\sum(C-\bar{C})(H-\bar{H})=162-(-16)(57)/123=16241$$

$$evar(CH)=13886$$

$$\sum(H-\bar{H})^2=267-(57)^2/123=24057$$

$$evar(H)=19720$$

Cに対するHの回帰線の傾きは $b=13886/46387=02993$ から推定される。

具合の悪いことには、組分ける結果、分散の推定値は過大になる傾向があるため、この推定値は偏つている。これを除くには、この推定値に sheppard の補正を適用しなければならない。これは群間隔の平方の $1/12$ を引くことである。したがつて修正した分散の推定値は

$$evar(C)=46387-2^2/12=43054$$

$$evar(H)=19720-1^2/12=18887$$

修正した回帰係数の推定値は $13886/43054=03225$ である。

しかし、この修正は推定の時だけに使うべきであつて、有意性の検定には、無修正の分散および平方和を使わねばならない。このため、組分けによつて検定の感度は幾分落ちるけれども、計算上の長所を考慮すれば、大したものではない。 sheppard の補正を使つて求めた回帰係数の推定値に標準誤差

を結びつけるには直接無修正の推定値から求めたものとして考えるとよい。例えば上記の無修正の推定値は 02993±005266である。修正推定値 03225±005674 を求めるためこの値に

46387／43054＝10774を乗ずる。この標準誤差が直接計算した回帰係数の標準誤差 005522に対応する。

7.4 重回帰分析の場合の組分け
Grouping in multile regression analysis

重回帰分析の場合にも組分けは使われるが得られる利益は極く僅かなものである。変量の数が増すにつれて分析に使われる副次的な組の数が増す。例えば、4変量を夫々5間隔にまとめると $5^4=625$ の副次的な組が必要である。この様な数はもちろん無意味であるが副次的群の数が多いので解析で組分をおこなつた結果得られる利益は非常に僅かなものになつてしまう。さらに計算の比較的大きな部分は同時方程式の解で占められているから組分けの利益は僅かである。次の解析に使うため回帰係数の大体の推定値を求める必要のある場合には、これに対応する相関係数および標準偏差の大体の推定値を使つて求めることができる。

x, t, z, に関する y の回帰の推定回帰係数 bx, bt, bz, は次式で求められる。

$$\delta_x b_x + r_{xt}\delta_t b_t + r_{xz}\delta_z b_z = r_{xy}\delta_y$$

$$r_{xt}\delta_x b_x + \delta_t b_t + r_{tz}\delta_z b_z = r_{ty}\delta_y$$

$$r_{xz}\delta_x b_x + r_{tz}\delta_t b_t + \delta_z b_z = r_{zy}\delta_y$$

ここで Sx, St, Sz は標準偏差の相対的な値の推定値であり、いろいろな方法で求められる。例えば内部4分位数間の距離 inter-quaritile distance 即ち大きい方50%と小さい方の値50%の平均値間の距離が標準偏差の相対的な値を示すものとして使われている。

Vxt, Vxz… は相関係数の推定値であつて簡略法で求められる。特に中位相関係数medial correlation coefficients (3.6節参照) は次の公式を使えば積率相関係数を与える。

$$r = \sin 70\phi$$

この式は与えられたϕの値に対応するrの大体の大きさを計るための換算公式である。例えば8.6節で計算したϕの値は0.333、0.366、0.675であつた。上記の公式で計算すれば、これに相当するrの推定値は0.422 0.544、0.872であり、これは直接推定した値0.487、0.564、0.755に比敵する。

同様な公式が順位相関係数での変換にも成立する。

7.5 曲線回帰分析の場合の組分け Grouping in Curvilinear regression analysis

曲線回帰分析の場合の組分けは偏つた回帰係数の推定値となることが多い。各組の観測値を平均した時組分けの巾がその組の平均値が真の回帰曲線から離れてしまうような時は広くすると特に偏つてくる。事実この様なことは、各組内で直線によつて回帰をうまく表わすことが出来ない程曲線性が強い時に起る。これ以外の場合では組分による偏奇は僅かである。

7.3節で述べた分割表法を拡張すれば、この解析は都合よく行くこの方法の例として回帰が事実上直線であることを立証するため7.3b表の結果を使つてみよう。

オ1段階として、胸囲の平方を表わす評点をこの解析に導入し、この新評点に対応する積和を計算する。その方法は7.5a表に示してあり、さらにこの関係の曲線性を検定するため値が最後の列に示してある。

25a表　回帰の曲線性を検べるための解析

		C							
	簡約数	-4	-2	0	2	4	計	積和	平方した値の積和
簡約数	その平方	16	4	0	4	16			
-3			1	1			2	-2	4
-2		3	1	2			6	-14	52
-1		4	11	4	4	1	24	-26	140
0		5	6	16	1	2	30	-22	140
H 1		2	6	11	13	1	33	10	124
2			1	4	14	1	20	80	76
3			1	3	3		7	4	16
4							0	0	0
5						1	1	4	16
計		14	27	41	35	6	123	-16	568
積和		-8	-5	17	46	7	57	162	148

さらに $\Sigma C^3 \Sigma C^4$ を計算する必要がある。

$\Sigma C^3 = 14(-64)+27(-8)+35(8)+6(64) = -448$

$\Sigma C^4 = 14(256)+27(16)+35(16)+6(256) = 6112$

これを用いて、2次の回帰に必要な修正平方和積和が計算できる

$\Sigma(C-\bar{C})(C^2-\bar{C}^2) = -448-(-16)(568)/123$
$= -37411$

$\Sigma(C^2-\bar{C}^2) = 6112-(568)^2/123 = 348704$

$\Sigma(H-\bar{H})(C^2-\bar{C}^2) = 148-(57)(568)/123 = -11522$

1次および2次の回帰係数に関する方程式は

$$56592b_1 - 37411b_2 = 16241$$
$$-37411b_1 - 348704b_2 = -11522$$

1次の回帰に帰因する平方和は $(16241)^2/56592 = 5071$ である。2次の回帰による平方和を求めるには 1)式 を $37411/56592$ 倍して2式に加えることにより b_1 を消去しなければならない。

$$324173b_2 = -323$$

回帰に含まれる2次の項は平方和 $(-323)^2/324173$ を説明するものであり、したがつて分散分析は25b 表に示してある様になる。

この関係の曲線性は全然無視することができ、直線によつてこの回帰は充分に表現されることが分る。

25b 表　Cに関するHの回帰の直線性を検べるための分散分析

変動因	自由度	平方和	分散の推定値
Cに関する直線回帰に帰因	1	5071	5071
2次の項に帰因するもの	1	000	000
残差	120	18288	1582
計	122		

7 6　固有の一次関係の推定 Estimation of underlying linear relationships

通常、特殊な場合を除いて固有の関係の推定にはくみわけは使用できない。それは誤差を伴なわない組分けをすることが難しいからである。したがつて２６節で説明した様に次の様に仮定されている一組の観測値ｘ，ｙが測定されたとすると

$$x = f(t) + \varepsilon_1$$

$$y = g(t) + \varepsilon_2$$

（この場合ε_1，ε_2は独立な確率変量である。）

ｔの値によつて観測値の組分けを行うのが根本的問題である。

関係の推定を不偏なものとするには誤差を伴なはない組分けを行なわねばならない。

もちろんｔの値は未知であるからむしろその取扱いに特殊な条件が必要である。

３４図は通常誤差を伴なはないくみわけが行える場合の一つの形の例である。この場合観測値ははつきりと２組に分れている。このことは確率誤差ε_1，ε_2が小さくｔは異つた２組の値をとることを示している。（ε_1，ε_2の二頭分布 bimodal distribution を含むいろいろな可能な場合があるが通常実際上は考えなくてもよい。）

ｔの値が異つた組に分けられる才２の場合は一つの変量ｘの誤差が一定の大きさａより小さいことが分つており、ｘの値が少くとも２ａだけ違つている組に観測値が分けられる様な時である。比較的に誤差の小さい物理学や化学の測定値ではこの様なことがよくある。

いづれの場合でも次の形の関係を推定することが必要である。

$$y - y_0 = \beta(x - x_0)$$

これは次の様に行なわれる。

ｘ，ｙの平均値は相互に対応しているから線は

$$y - \bar{y} = \beta(x - \bar{x})$$

と書かれる。

Ｂを推定するにはｔの値が残りのものより大きいものと小さいものとの２つの組に観測値を分けることが必要である。これは誤差のない様にして行なわねばならない。$(\bar{x}_1, \bar{y}_1)$，$(\bar{x}_2, \bar{y}_2)$をこの組の平均とすればβは

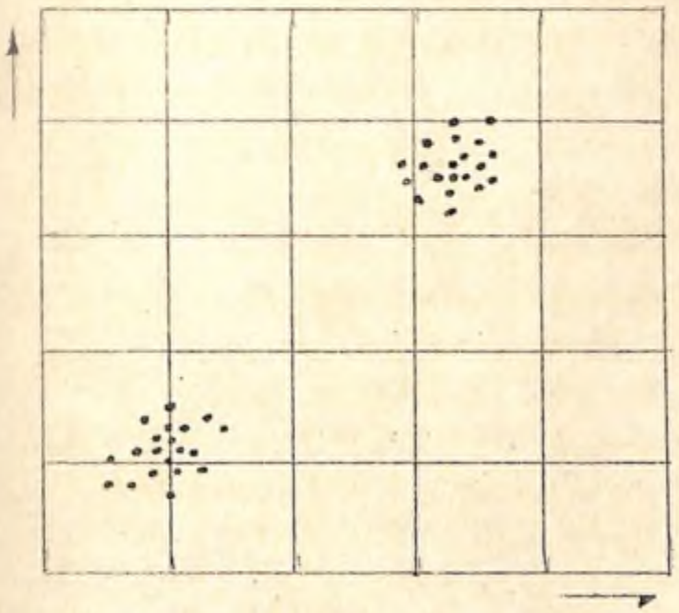
$$b = \frac{\bar{y}_2 - \bar{y}_1}{\bar{x}_2 - \bar{x}_1}$$

から推定される。

この方法の例として一組の通行量に関する観測値について考察してみよう。この観測値は７６ａ表に示してあり、Ｔは特定の日にある町のいろいろな地区とその中心で測られた通行者数（千単位）の対数、Ｒはその地区の登録者数（千人単位）の対数である。

ＲとＴの間には基本的に一次関係がある。即ち登録者数が比例的に変化すればこれに応じて通行者数が比例的に変化すると仮定する。

Ｒに全然誤差がないと仮定できれば、この関係はＲに関するＴの線形回帰と同じものである。Ｒに僅かな誤差があれば一次回帰は調べようとしている〝法則〟と同等とはいえないであらう。



３４図　誤差を伴なわずに組わけのできる場合

76a表　　交通量の観測値

R	T	R	T	R	T	R	T
041	039	019	014	006	006	−027	−046
028	027	013	026	000	009	−031	−021
022	047	013	025	−003	017	−054	−055
019	039	012	018	−007	004	−072	−074
019	038	011	012	−009	−003		
019	026	008	009				

$\bar{R}_1 = -0124$　$\bar{T}_1 = -0208$　　$\bar{R}_2 = --0460$　$\bar{T}_2 = --0490$

$\bar{R} = 0013$

$\bar{T} = 0075$

しかし登録者数の誤差は２０％以上ではない、即ちRの誤差は０．０８以上でないと仮定してもよい。したがってRの２つの値の間の差０．１６は登録者数に実際差を表わしていると仮定できる。

この仮定から７６ａ表の最後の４つの値は別の組を構成していることが分る。したがってこの２群の平均値は上表の様になり、βの推定値は

$$b = \frac{0208+0490}{0124+0460} = 1195$$

故にこの関係は次の様に推定される

$$T - 0075 = 1195\,(R - 0013)$$

即ち　$T = 1195R + 0059$

（比較：Rに関するTの回帰＝１．１２３Ｒ＋０．００６０）

推定された傾きの標準誤差は、観測値と真の関係との隔の標準偏差を含んでいるため、その評価は幾分難しい。これは次式で与えられる。　bの標準誤差 $= \frac{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}{\bar{x}_1 - \bar{x}_2}$

n_1, n_2 は２群に含まれる観測数、Sは観測値と真の関係を表わす線との隔りの標準偏差である。この後者の値は分散－共分散分析

を使つて求められるが、この場合、データを分割した２又は３つの組間の差は消去されている。次の例はこの方法を説明するのに役立つであらう。

76a表のデータの分散－共分散分析が76b表に示してある。

76b表 分散－共分散分析

変動因	D.f	R		T		R.T	
		S.	E.v	S.	E.	S.p	E.c
群間の差によるもの	1	1.104		1.578		1.320	
残差	19	0395	0.0208	0.468	0.0246	0347	0.0183
計	20	1.499		2046		1.667	

この解析から、線からの隔りの分散の推定値は

$$= 0.0246 - 2\,(0.0183)\,\beta + 0.0208\,\beta^2$$

$$= 0.0246 - 0.0366\,\beta + 0.0208\,\beta^2$$

オ１近似としてβを1.195とすればβ^2は0.0106となり、標準偏差は0.103である。bの標準誤差は

$$\text{bの標準誤差} = \frac{0.103\ \sqrt{\frac{1}{17}+\frac{1}{4}}}{0.584} = \pm 0.098$$

と推定される。

故にbの９５％信頼限界は1.195±(2.09×0.098)即ち0.990と1.400と推定される。

この方法は実用的には充分正確なものであり、βの推定の時βの代りにbを使つたとしても、信頼限界が近似的になるだけである。

もつと正確な９５％信頼限界を求めるにはβ＝1.400に対するSの値を用いて上限を、β＝0.990を用いて下限を再計算する必要がある。この様にして新しい限界1.009と1.431が得られる。必要があれば、信頼限界の適当な値が得られるまでこの方法を繰返す。しかし一回り計算すれば実用的には充分正確な値が得られる。

尚は　の方程式ではβをそのまゝにしておき、信頼限界を与える方

程式でβについて解けば信頼限界を定める全過程を正確に行うことができる。したがつて上記の例ではβの信頼限界は

$$1.195-\beta=\pm 2.09\times\frac{\frac{1}{17}+\frac{1}{4}}{0.584}$$

$$=1.989s$$

これは次式に等しい。

$$(1.195-\beta)^2=(1.989)^2(0.0246-0.0366\beta+0.0208\beta)^2$$

根1.007と1.439をもつβの二次方程式

この根はβの95%信頼限界を与える。

一般に　観測値の上群と下群とを作るいくつかの方法がある時には　傾きの出来るだけ正確な推定値を与える様な組を選ばねばならない。M.S. Bartlett※は観測値を組分けする問題について研究した。彼は最も正確な（※ Bartlett M.S., Biometrics, 5 (1946) 207）推定値を求めるには　両端の組のそれぞれに観測値のほぼ1／3が入る様に選ぶべきであると結論した。

7.7　重みを付けた時の解析 Weighting analysis

組分けを使つた解析では普通、組平均は個々の観測値より正確であらうということを確める必要がある。7.2節の様に平均値が同数の観測値（又はそれに近い）から求められている場合には必要はない。しかし平均値が異つた観測数から求められている場合にはどの様な解析においてもこの事実を確めることが必要である。この様なことは重みを付けることによつて行われる。

重みを付けた時の解析の簡単な例は6.1節で既に示した。重みを付けた時のもつと一般的な解析方法を示すためさらに精しくこれを考察してみる。この観測値の合計および平均が7.7a表に示してある。

7.7a表 組内個数	観測値の組合計および平均、log（心臓の重さの百分率） W 計	平均	log（ヘモグロビン含有率） H 計	平均
6	4.96	0.827	4.51	0.752
7	6.33	0.904	4.39	0.627
6	5.63	0.938	3.55	0.592
5	5.03	1.006	2.14	0.428
5	5.60	1.120	1.79	0.358
計29	27.55		16.38	

6.1節の例では組間の平方和、積和は組の計から直接計算された。これは疑いもなく最も簡単かつ最良の方法であるが、各組の平均値を用いる別の方法も亦注目する価値がある。

この場合には　各観測値はそれが含まれている組平均に等しい値をもつていると考えられる。したがつて　例えば第1組の6個の観測値はW=0.827、H=0.752の値をもつていると考えられ、組間の積和は次の様に計算する。

$$6(0.827)(0.752)+7(0.904)(0.627)+\cdots 5(1.120)(0.358)-\frac{(27.55)(16.38)}{29}=-0.3725$$

代数学的には次の様に表わされる

$$n_1\bar{x}_1\bar{y}_1+n_2\bar{x}_2\bar{y}_2+\cdots n_m\bar{x}_m\bar{y}_m-\frac{(\Sigma n_i\bar{x}_i)(\Sigma n_i\bar{y}_i)}{\Sigma n_i}$$

$$=\Sigma n_i\bar{x}_i y_i-\frac{(\Sigma n_i\bar{x}_i)(\Sigma n_i\bar{y}_i)}{\Sigma n_i}$$

および　$n_1\bar{x}_1^2+n_2\bar{x}_2^2+\cdots n_m\bar{x}_m^2-\frac{(\Sigma n_i x_i)^2}{\Sigma n_i}$

$$=\Sigma n_i\bar{x}_i^2-\frac{(\Sigma n_i x_i)^2}{\Sigma n_i}$$

nなる量が常に平均値の前についており　したがつてこれが平均値に付加されるべき相対的な値即ち重みを意味している以外は　普通の平方和、積和の表わし方に対応していることが分る。明らかに

価値の違う観測値がとられた時にも同じ様な解析が行なわれる。例えば観測値の内あるものが他に較べて2倍或は3倍正確であれば2倍或は3倍の重みを与えるべきである。

一般に，$x_1\ y_1$ ；$x_2\ y_2$ ；… $x_n\ y_n$ が相対的重み W_1，W_2 … W_n をもつn対の観測値の組であれば平方和、積和は次式から計算される。

$$\Sigma w_i x_i^2 - \frac{(\Sigma w_i x_i)^2}{\Sigma w_i},\ \Sigma w_i x_i y_i - \frac{(\Sigma w_i x_i)(\Sigma w_i y_i)}{\Sigma w_i} \text{ および } \Sigma w_i y_i^2$$

これらの平方和はいづれも 自由度 n－1 をもつている。

回帰分析を行う時には 平均値の推定にも異なる重みを考慮すべきである。これは次式から計算される。

＝ ――― ， ＝ ―――

この解析法を説明するため 3図の組分けしたデータについて行つた解析を考察してみる。77b表は体重で組分けしたこれらの小児の平均体重と基本代謝率を示している。

77b表 男児の平均体重と基本代謝率

体重による組	人数n	平均体重W	平均代謝率M	重み係数 $W=10^6 n/M^2$
200以上	52	1722	8789	67
200 －	73	2275	10259	69
250 －	65	2736	11099	53
300 －	51	3208	12276	34
350 －	35	3730	13547	19
400以上	47	4594	14847	21

組に含まれる人数は35～73まで変つている。さらに この場合には組内の変動性はかなり変化しており，平均体重の大きな組は散らばりも大きい。したがつて平均値の相対的な値を評価するには各平均値に関与している人数と同様に観測値の散らばりを考慮しなければならない。

平均値を重み付けるための適当な指標は n_i/σ_i^2 である。たゞし σ_i^2 は i 番目の組における回帰線の周りの分散である。この重みの係数は組による変動の違いを考慮に入れてはいるが その計算は甚しく時間がかゝりかつ複雑である。この場合には 3図から 標準偏差は代謝率にほぼ比例して増加することが容易に分るから近似的な重み係数を与えるため平均代謝率が標準偏差の代りに用いられる。77b表では $10^6 n/M^2$ なる量が重み係数として使われた。

この重み係数を使つて解析は速やかに行われる。まず次の計算をする必要がある。

$\Sigma w = 263$　$\Sigma_w W = 693773$　$\Sigma_w M = 2871545$

$\Sigma w W^2 = 2009983 7$　$\Sigma_w WM = 79652656$　$\Sigma_w M^2 = 3220632 95$

この値から

$\overline{W} = 2638$　$\overline{M} = 1091.8$

$\Sigma_w (W-\overline{W})^2 = 1798659$　$\Sigma_w (W-\overline{W})(M-\overline{M}) = 3903592$

$\Sigma_w (M-\overline{M})^2 = 8535892$

故に $b = 3903592/1798659$

$= 21.703$

したがつてWに関するMの回帰方程式は

$M = 1091.8 + 21.703 (W-2638)$

$= 5193 + 21.703W$

分散分析は普通の方法で行なわれ 77c表に示してある。

77c表 Wに関するMの回帰の分散分析

変動因	自由度	平方和	分散の推定値
Wの一次効果によるもの	1	8471884	8471884
残差	4	64008	16002
計	5	8535892	

観測値の総数としてNの代りに Σ_w，個々の観測値の重みとして $10^6/M^2$ を使う必要のある以外は標準誤差は普通の方法で求められ

る。

したがつて体重の測定値Wに対応する代謝率の測定値Mと推定値mとの差の標準誤差は

$$\sqrt{\left[1600\,2\left(\frac{m^2}{10^6}+\frac{1}{268}+\frac{(21.703)^2(W-2638)^2}{1798659}\right)\right]}$$

8

共分散分析の利用　Use of the Analysis of Covariance

8 1　統計的手段としての共分散分析
Analysis of covariance as a statistical tool

厳密に云えば　共分散分析は推定や解析の簡略法というよりはむしろ推定や解析の方式化の手段である。分散分析と同じく、これは計算を便利かつ簡潔に行うことのできる方法を提供する。この方法は簡単に学ぶことができすぐに応用できる。分散分析に関しては、これも亦通常でない比較的時間のかゝる解析にある根拠を与える。即ち連関のある測定値の解析を具体化する一つの形を提供する。したがつて　必要な全体の計算を減すことではなくて便利で簡単に理解できる形にそれを組立てることにより解析は短縮される。

規則通りに共分散分析を用いれば形にはまつた解析法と同じ様な非難をうける。そうすると解析のある面の重要性を見失なうことが多い。例えば推定の問題を等閑視して有意性の検定を強調する傾向を生ずるかもしれないし又これらの解析で当然考慮しなければならない分散の一様性および傾きの平行性に関する基本的仮定を解析者が見落すかもしれない。しかし　この非難は解析より寧ろ解析者に関するものである。共散分析はそれ自身が目的であるというよりはむしろ一つの道具であることを知つておれば　この様な非難は起らないであらう。

手軽に取扱いうる共分散分析を使つたいくつかの問題を次節で示そう。この問題は必ずしも共分散分析を用いて処理する必要があるというものでもなく又、この様な場合にのみ共分散分析が使用できるというものでもない。しかしこの問題は共分散分析を適要しうる問題の型を示すのに役立ち又この方法はこの種の問題を扱う場合最も有効であろう。

8 2　別の変量が除かれねばならない場合の連関の問題

別の因子による影響を除いた時の2つ以上の変量間の連関を考察する必要のある場合が多い。この様なことは共分散分析を使つて行なわれるであらう。

例えば各変量を観測した時点における一般的な時間的傾向を除いてから二変量を関係づけたいことが多い。例として関係があると信じられている或る因子の発生と病気の範囲－小児まひ－とを関係づけたいとする。この場合、小児まひの影響範囲と他の因子の発生は、いづれも考察している期間を通じて絶えず変つており、したがつてこれらの変量間の相関がこの様な普通の又は無関係の変化に帰因する可能性を除くには　分析を始める前に時間的傾向を除く必要がある。

これを行うには各変量に対して多項式をあてはめ、分散共分散分析で残差の分散および共分散を推定すればよい。この残差の分散、共分散は普通の回帰分析で用いられるだらう。多項式を当てはめのために行う積和の計算を除いてこれは全然面倒ではない。これは推定された多項式の回帰係数と各変量を推定するための方程式の右辺との積和から計算される。即ち式で表わせば

$$b_1 \Sigma x (t-\bar{t}) + b_2 \Sigma x (t^2-\bar{t}^2) + \cdots$$

両方法でこの積和を計算すれば　解析※ の吟味になる。

もちろん　この種の解析では　両変量が確実に同じ方法で処理されておらねばならない。

したがつて一つのこの様に高次の多項式が必要なのかこれらの変量の一方だけの場合でも　傾向を除くには両変量に対して同次の多項式をあてはめる必要がある。

※注　この問題に関するこれ以上の説明と本節の解析のためには
1 1 6と1 1 7節を参照されたい。

回帰分析に着手する前にある型の組又は級間の差を除く必要のある場合には、これと関連した問題が起る。例として§2表は1947～1950年におけるGreat Britainでの休業率incapacity rateと工業生産指数とを示している。工業生産の一般的傾向はさておき、休業率と工業生産のいづれにも明らかに季節的変動がある。したがつて工業生産が休業により受ける影響の状態をより詳しく知るには解析に着手する前に　四半期および年による変動を除くことが望ましい。この様にすれば無関係の変動の大部分を取除き、もつと鋭敏な解析を行うために効果がある。

§2a表　休業率I，工業生産指数、P（G．B）

休業率（100人当り日数）

四半期	1947	1948	1949	1950	計
1	158	114	136	128	536
2	84	82	90	89	345
3	65	84	78	73	300
4	100	126	121	108	455
計	407	406	425	398	1636

工業生産指数（全工業）

四半期	1947	1948	1949	1950	計
1	96	120	128	140	484
2	109	122	129	140	500
3	109	115	123	133	480
4	119	126	135	147	527
計	433	483	515	560	1991

この数値を使つて行つた分散、共分散分析が§2b表に示してある。

§2b表　分散、共分散分析

変動因	自由度	1の平方和	1とPの積和	Pの平方和
四半期	3	341	348	8556
年	3	2146	−59	98
残差	9	159	−428	1581
計	15	2646	−139	10235

この表を調べてみると生産と休業との相関は

$$r = \frac{-139}{\sqrt{2646 \times 10235}} = -0.03$$

であることが分るが　生産と休業の両者から生産の一般的傾向と正常な四半期変動を除けば　相関は

$$r = \frac{-428}{\sqrt{159 \times 1581}} = -0.85$$

となる。

後の相関は大きく、したがつて極めて有意であり、工業生産と休業の変動とは互に密接に連関していることを示している。

休業が一単位増す毎に工業生産は平均して428/1581減少する。このことは別の因子が（例えばWorks expression）両者に影響しているかもしれぬから休業の変化が生産の変化の原因となつているということを必ずしも示してはいない。

§3　回帰分析における別の変量の追加

前節では最初に曲線的傾向や組間の差を除いた分散、共分散分析を考察した。既に回帰分析で用いられた観測値の共分散を考察することももちろんできる。この様にすれば簡単に回帰分析に、さらに別の変量を導入し、これを検定することが許される。

$x_1\ x_2 \cdots x_n$ に関するyの回帰分析は既になされており　さらに変量 $z_1, z_2 \cdots z_m$ を導入することにしたとする。まず $x_1\ x_2 \cdots x_n$

に対する y_1 $z_1 \cdots z_m$ の回帰を計算する。即ち次の様な方程式が得られる。

$$y = a + b_1 x_1 + b_2 x_2 + \cdots b_n x_n$$
$$z_1 = a_1 + b_{11} x_1 + b_{12} x_2 + \cdots b_{1n} x_n$$
$$z_2 = a_2 + b_{21} x_1 + b_{22} x_2 + \cdots + b_{2n} x_n$$
$$- - - - - \quad - - - -$$

どの回帰の計算にも係数の同じ逆行列が使用できるからこの計算はさして面倒ではない。

回帰分析が行なわれたならば x_1, $x_2 \cdots x_n$ の効果が除かれた分散、共分散分析をすることができる。この分析で得られた残差が新変量の有意性を検べ、回帰係数 c_1, $c_2 \cdots c_m$ を推定するのに用いられる。こみにした回帰方程式は

$$y - a - b_1 x_1 \cdots - b_n x_n = c_1 (z_1 - a_1 - b_{11} x_1 - \cdots - b_{1n} x_n) + c_2 (z_2 - a_2 - b_{21} x_1 - \cdots - b_{2n} x_n) + \cdots$$

故に

$$y = a^* + b_1^* x_1 + b_2 x_2 + \cdots b_n^* x_n + c_1 z_1 + c_2 z_2 + \cdots c_m z_m$$

修正した回帰係数 b_1^*, $b_2^* \cdots b_n^*$ の有意性の検定が必要になつてくるが これは初めの回帰の逆行列と共分散回帰の逆行列を用いてなされる。これを行うには 次の事に注意する。

$$b^*i = bi - c_1 b_1 i - c_2 b_2 i - \cdots - c_m b_{mi}$$

したがつて

$$Var(b i^*) = Var(bi) + b_{1i}^2 Var(c_1) + b_{2i}^2 Var(c_2) + \cdots$$
$$+ 2 b_{1i} b_{2i} cov(c_1 . c_2) + 2 b_{1i} b_{3i} (c_1 . c_3) + \cdots$$

var (b_i), var (c_i), cov $(c \cdot cj)$ の値は2つの逆行列から求められる。

この方法は本来 M.H.Quenouille※ が行列 記号で説明したものであるが回帰分析に一個の追加変量を導入するため W.G. Cochran によつて与えられた結果を拡張したものである。

※ Quenouille, M.H. Econometrika 18 (1950) 27
Cochran, W.G. E.R. statist Soc. Supple 5 (1938) 171

8 4 回帰線間の隔り Distances between regression lines

共分散分析が回帰線間の隔りの検定に使用できることを示すため 8 4 a 表に示してある観測値を考察してみる。

8 4 a 表患者の血漿の容積（リットル）P，体重（kg）W，身長（in），H

男						女					
P.	W.	H.	P.	W.	H.	P.	W.	H.	P.	W.	H.
260	53	63	317	61	66	308	64	67	226	52	62
287	69	65	253	59	64	303	67	66	316	71	67
289	64	66	383	73	68	282	64	67	284	68	60
348	58	68	350	76	69	278	68	67	272	76	62
339	69	66	308	79	67	385	85	72	234	54	61
312	70	70	377	69	68	291	60	67	312	83	62

この表は24人の血漿容積、体重、身長の観測値を示しており、この表から、体重、身長を用いて血漿容積を豫想する式を求めたい。

この観測値は男女両性についてとられているので 両者同じ関係が成立するか否かをまず考察してみる必要がある。両性の回帰係数に差があつたとしても、そう大きなものとは思われない。この様に少数の観測値の場合でも 回帰係数間の有意な差は殆んど起りそうもない。したがつて 両性をこみにして解析が行われる。8 4 b 表は 8 4 a 表の観測値の分散、共分散分析である。

8 4 b 表 分散、共分散分析

変動因	自由度	P の平方和	W の平方和	H の平方和
性	1	0.7688	0	9800
差 残	22	3.5942	1807	10333
計	23	4.3630	1807	20133

変動因	自由度	PとWの積和	PとHの積和	WとHの積和
性	1	−041	8080	−4.7
残差	22	5335	12840	2294
計	23	5294	20920	2247

この表の値を使つて　回帰係数を推定するための方程式が求められる。

$$1807\ b_w + 2294\ b_h = 5235$$
$$2294\ b_w + 10333\ b_h = 12840$$

故に

$$b = 0.00191$$
$$b = 0.0818$$

WとHに帰因する平方和は

$$(0.00191)(5335) + (0.0818)(12840) = 20693$$

回帰係数の標準誤差を求めるにはPの分散分析を完成し回帰係数の逆行列、したがつて係数の共分散行列を求めなければならない。

これは84c表に示してあり、これからWとHは回帰に有意な寄与をしていることが分る。

84c表　Pの分散分析と共分散行列

変動因	自由度	平方和	分散の推定値
WとHに帰因するもの	2	20693	
新しい残差	20	15249	00762
もとの残差	22	35942	

逆行列 $= \begin{bmatrix} 0.0007706 & -0.001711 \\ -0.001711 & 0.01348 \end{bmatrix}$

共分散行列 $= \begin{bmatrix} 0.00005872 & -0.0001304 \\ -0.0001304 & 0.001027 \end{bmatrix}$　注 回帰係数の分散共分散行列

b_W の標準誤差 $= \sqrt{(0.00005872)} = \pm 0.00771$

b_H の標準誤差 $= \sqrt{(0.001027)} = \pm 0.0322$

男性および女性に対する回帰方程式は次の様にたる。男性に対しては

$$P = 3153 + 0.0191\,(W-671) + 0.0818\,(H-670)$$
$$= 0.0191W + 0.0818H - 3609$$

女性に対しては

$$P = 2.740 + 0.0191\,(W-673) + 0.0818\,(H-623)$$
$$= 0.0191W + 0.0818H - 3642$$

この二つの線は殆んど一致しているから　全体を通じて一つの関係が使えるかどうか考察してみる。これを行うには2つの線間の隔りを計算する必要がある。

$$0.413 + 0.0191\,(0.2) - 0.0818\,(4.7) = 0.032$$

その標準誤差は

$$\sqrt{\frac{\mathrm{Var}(P)}{18} + \frac{\mathrm{Var}(P)}{6} + (0.2)^2 \mathrm{Var}(b_w) + (4.7)^2 \mathrm{Var}(b_H) - 2\,(0.2)\,(4.7)\,\mathrm{cov}(b_w\ b_H)}$$

$$= \sqrt{\frac{0.0762}{18} + \frac{0.0762}{6} + (0.2)^2\,(0.00005872) + (4.7)\,(0.001027) - 2\,(0.2)\,(4.7)\,(0.0001304)} = \pm 0.200$$

線間の隔りは明らかに有意でない。したがつて　身長と体重に対する血漿の関係を説明するのに一本の線即ちP = 0.0191W + 0.0818H − 3616が使える。

数本の回帰線間の隔りを検定する場合には上記の方法で　対毎に比較をすればよいが　全体を通じて差の検定を行うには　5 4節で説明した様な解析をする必要がある。共分散分析を利用すればもっと簡単に行える。

この方法を説明するため、今一度　5 4 a 表に示してある3匹のネコの観測値を考察してみよう。上記の方法でネコの対を互に比較するのであるが　対の選び方には規則はないから　全体を通じての有意性の検定が必要である。これは次の様にしてなされる。

まず　このデータの分散共分散分析をまず計算する。この分析表を使えばネコ間の差が除かれている時も、そうでない時についてもA，Pに帰因する変動　したがつてFに含まれる原因不明の分散を推定することが可能である。前者の場合には原因不明の分散は1861後者の場合には2213である。

8 4 d 表　分散、共分散分析

変動因	自由度	Fの平方和	Aの平方和	Pの平方和
ネコ間	2	3791	11722	18904
原因不明	31	5388	16523	30849
計	33	9179	28245	49753

変動因	自由度	FとAの積和	FとPの積和	AとPの積和
ネコ間	2	1968	1606	4067
原因不明	31	2313	3244	20170
計	33	4281	4850	24237

AとPに帰因する変動を除いた場合のネコ間の差を検定する分散分析は8 4 e 表に示してある様になる。AとPの効果を除いたネコ間の差は、もちろん回帰線間の隔りである。

したがつて　この分析は回帰線間の隔りを検定することになる。

8 4 e 表　AとPの効果を除いたFの分散分析

変動因	自由度	平方和	平均平方
ネコ間	2	347	1735
ネコ内	29	1866	643
計	31	2213	

分散比2 7 0は5 %有意水準に達していないから全体の平方和、積和から計算した回帰方程式が　AとPに対するFの関係を説明するのに使用できる。

8 5 実験の時の共分散分析　Analysis of covariance in exp erimentation

実験の際　処理による影響は受けないが　実験の変動には影響をおよぼす量について補助的な観測値がとられる場合がある。例えば農作物の試験では　主な測定値は各種の処理をほどこした時の収量であるが　処理前の土壌条件を知るために、既往の収量、酸度等の測定が行われる。

この様な場合、処理効果を検定するには　通例補助測定値を利用して補助観測値に対する主変量の回帰が処理によつて違つているかどうか検定してみるとよい。しかし変動性と回帰係数は普通一定であるから一般的には平行線間の隔り、即ち補助変量の差に対して修正をほどこした時の主変量の平均値間の差を検定すれば充分である。したがつて農作物の試験では、各処理によつて得られる平均収量は標準地の頭初の肥沃度に応じて調整される。

共分散分析の方法は試験の際補助観測値を使つて測定された変動性を除くのに用いられることも多い。しかし回帰線間の隔りが何んらかの意味を持つておれば　補助観測値は処理の相違による影響をうけてはならないことに注意する必要がある。処理の相違により影響されるものが主観測値だけであれば　線間の隔りは当然、主観測値に対する処理効果と呼ばれる。そうでなければ線間の隔りは主変量又は附随変量のいづれかに帰因するものであらう。

実験に共分散分析を応用した例としてF.Yatos※ の示した例を考察してみよう。

(※　Yatos J.Agric Sci. 24 (1934) 511)

Yatosは6匹の豚の組の各々に4種の飼料を与えた試験研究した。豚を3つの小屋に分配した。豚の半分には湿性のえさを　半分には「ひきわり」を与え　この2つのグループの夫々半分には緑銅を加えた。体重は試験を始める前と6週間後に記録した。

分散分析の詳細は複雑であるがわれわれにはこの場合関係がない。食銅を混らした時緑銅の効果の変化に応じて　ある成分を除くことは注目する価値がある。この成分は交互作用緑銅X湿性の銅と名付けられる。

85a表は6週間後の体重Wと最初の体重Iとの分散、共分散分析を示している。

期待したとおり、最初の体重には処理による有意差はないことがわかつた。さらに最初の体重の効果を考慮すれば　回帰に帰因する平方和は $(854.51)^2/42378=172303$ である。これは極めて有意であり、残差分散の大部分は回帰に因るものである。この部分を除けば　新しい残差平方和290567−172303=118264 が得られる。

85a表　分散共分散分析

	自由度	Wの平方和	WとIの平方和	Iの平方和
飼料	3	900084	431958	218016
小屋	2	47308	24470	12658
湿性食銅の効果	1	98817	7058	504
緑銅の効果	1	29400	−4550	704
交互作用 緑銅X湿性食銅	1	2408	−1629	1102
性	1	38400	14800	5704
残差	14	290567	854.51	42378
計	23	1406984	557558	281096

最初の体重の影響を除いた時の　これとは別の比較 comparison に用いられる平方和を計算するには　その各々の平方和と残差平方和との組合せを考える必要がある。したがつて、例えば　最初の体重の影響を除いた時の小屋の平方和に　残差平方和を加えたものは

$$47308+290567-\frac{(24470+85451)^2}{12658+42378}=118335$$

小屋の平方和（最初の体重の影響は除かれている。）は118335−118264=071

この様な一連の計算により、最初の体重の影響を除いた時の分散分析表は構成される。

これは85b表に示してある。

85b表　分散分析（最初の体重による影響は除かれている）

	自由度	平方和	分散の推定値
飼料	3	44402	14800
小屋	2	071	036
湿性食銅の効果	1	71551	71551
緑銅の効果	1	49784	49784
交互作用 緑銅X湿性食銅	1	8301	8301
性	1	1680	1680
残差	13	118264	9097
計	22	294053	−−−

湿性食銅と緑銅の分散は有意である。これらは回帰係数854.51/42378=2016 を使つて推定できる。例えば　湿性食銅と緑銅による6週間目の平均体重の差は12831bであるが　初めの体重間の差は0921b であつた。最初の体重の効果を除いた時の湿性食銅の効果は

1283−2016X092−10981b

この推定値の標準誤差は

$$\sqrt{90.97\left\{\frac{1}{12}+\frac{1}{12}+\frac{(0.92)^2}{423.78}\right\}}$$

$$= \pm 3.92$$

8.6 数個の資料源からとられた観測値の解析
Analysis of observation from several sources

時には数個の資料源からとられた観測値について回帰分析或は相関分析を行う必要がある。この場合には若干複雑になる　第1にいろいろな場所でとられた観測値の精度は違つているかも知れない。第2に　回帰係数は場所によつて違つているかも知れない。第3に資料源間の原因不明の変動は　任意の資料源からとつた観測値の原因不明の変動より大きいかも知れない。最後に各資料源の平均値間の関係は　任意の資料源からとつた観測値から計算した関係と違つているかも知れない。

例として　個人のエネルギーの必要量Eと体重Wとの関係に調査を実行するように決められたとする。この目的でいろいろな国で数人の観測者が個人についてE，とWの値を報告するように要求されたとする　その結果得られたデータの解析には次の様な問題が含まれている。

1. どの観測者についてもエネルギー必要量の算定精度は等しいか（或はどの地方の個人の変動も等しいか）
2. EとWの関係はどの地方においても同じであるのか即ち
   2回帰係数は同じであるか？　もしそうであれば与えられたWの値に対するEの値にどの様な国別差違（即ち観測者の偏り）が在在するのか？
4. いろいろな資料源から求めたEとWの関係は任意の資料源の関係と同じであるか

任意の場所で高いEの値を生じ易い　気候の様な条件は　低いWの値を生じ易い傾向がおうおう生じるので　この最後の問題は特に

大切である。

したがつて場所間のEとWの連関は負となるが、一方ある場所では正である。

この問題の初めの3つは5.4節で説明した方法で扱かわれ、変動性と回帰係数がどの資料源についても同じであると仮定すれば、前節の方法で扱うことができる。各資料源からとつた観測値の精度の違つていることが分れば　7.7節で説明した様な重み付きの解析を使わねばならない。このためには　各資料源からのデータの残差分散を推定し　重み係数

$$w_i = \frac{f_i - 2}{f_i s_i^2}$$

を使つて全体の解析を行う必要がある。

ただしs_i^2はi番目の資料源の残差分散の推定値で　自由度f_iに基ずいている。

通例　最後の問題の扱い方（即ちいろいろな資料源から得られた平均値間の関係は任意の資料源の観測値間の関係と違つているか）は初めの3つの問題に対する解答により違つてくる。一般に　この問題は回帰係数がどの場所についても同じである場合にのみ重要である。

そうでない時には　どの様な場合でもその関係を説明するのに何本かの回帰線が必要である。

資料源間の変動が同じ資料源の観測値間の変動より有意に大きくない場合には簡単な有意性の検定法が用いられる。その方法を示すため　5つの資料源からとられた6個の観測値の組の（仮設的）解析を考察してみる。2つの変量yとxが測定され、その分散、共分散分析は8.6a表に示してある様になつたとする。

8.6a表　仮設的な分散、共分散分析

変動因	自由度	yの平方和	xとyの積和	xの平方和
資料源間	4	6	4	3

資料源内	25	12	5	10
計	29	18	9	13

xの効果を除いた時の資料源間の差を検定する分散分析は8 6 b表に示してあるとおりである。

8 6 b表　分散分析（xの効果を除去）

変動因	自由度	平方和	分散の推定値
資料源間	4	11.77−9.50=2.27	0.568
資料源内	24	$12-(5^2/10)=9.50$	0.396
計	28	$18-(9^2/13)=11.77$	

資料源間の差は有意でないから、各資料源の平均値の回帰が各資料源の観測値の回帰と違つているか否かを検定できる。同じ資料源からの観測値の回帰係数は5／10=0.5であり、一方資料源の平均値の回帰係数は4／3=1.3であるので　多分そうなるものと思われる。

各資料源の平均値を使つた回帰に帰因する平方和は$4^2/3=5.33$であり　したがつてこの回帰線の周りの残差平方和は6−5.33=0.67である。回帰の推定に1自由度が使われているので　この自由度は3である。

この最後の値と8 6 b表の値2.27との相違は　前者は資料源内のデータから計算した回帰を使つたものであり、後者は資料源間の回帰を使つたものであるという点である。そこで2つの回帰間の差を検定するのに　この2つの値の差が用いらる。したがつて8 6 b表の分析は8 6 c表の様に拡張される。

この分析では　回帰間の差を検定する成分は大きく、分散比4.04は5％水準で殆んど有意である。（4.28）このことはさらに詳しく分析すれば　平均値の回帰と資料源内の回帰との差が示されるであらうことを暗示しているが　この場合にはこの様な差について

の確証はない。

8 6 c表　回帰間の差を検定するための分散分析（xの効果を除去）

変動因	自由度	平方和	分散の推定値
回帰間	1	1.60	1.600
資料源の平均値の回帰	3	0.67	0.223
資料源間	4	2.27	0.568
資料源内	24	9.50	0.396
計	28	11.77	

この例の様に2つの回帰間に有意差がなければ、計の平方和を使つて全体の回帰を計算してさしつかえない。したがつてこの場合には　xに関するyの回帰係数として9／13=0.69が用いられる。

組平均値間の差が有意であれば　上に示したのと同じ方法が用いられるが有意性の検定は違つてくる。この場合には2つの回帰間の差を検定する分散の推定値は資料源内の分散と平均値の回帰線の周りの分散との重みを付けた組合せに対して検定される。

独立変量が一個の場合には　^{2}を資料源内分散の推定値、β_1^2　を回帰線の周りの平均値の分散の推定値、B, B_1 を資料源内及資料源間の独立変量の平方和とすれば、回帰間の差を検定するのに用いられる分散は $(B_1\beta^2+B\beta_1^2)/(B+B_1)$ である。回帰線間の差を検定する分散の推定値は逆に両方の他の分散に対して検定されるので通常はこの種のあらゆる組合せに対して計算する必要はない。どの場合についても同じ結論がえられゝばこの2つの分散の組合せを用いる必要はない。次の例は　その方法を示すのに役立つであらう。

8 6 d表は5つの資料源からとつた6個の観測値の（仮設的）分析を示している。変量y, xが測定され、8 6 d表に示してある様に分散、共分散分析がなされたとする。

8.6 d表　仮設的な分散共分散分析

変動因	自由度	yの平方和	x，とyの共分散	xの平方和
資料源間	4	12	4	3
資料源内	25	6	5	10
計	29	18	9	13

xの効果を除けば

変動因	自由度	平方和	分散の推定値
回帰間	1	1.60	1.600
資料源の値平均の回帰の周り	3	$12-(4^2/3)$ =6.67	2.223
資料源間	4	11.77−3.50 =8.27	2.068
資料源内	24	$6-(5^2/10)$ =3.50	0.146
計	28	$18-(9^2/13)$ =11.77	

この場合2つの回帰間の差を検定する分散の推定値1.600は資料源内分散0.146と比較した時には有意であるが、回帰線の周りの分散2.223と比較した時には有意でない。したがつてこれらを組合せたものに対して検定を行う。

$$\frac{10\times2.223+3\times0.146}{10+3}=1.744$$

分散の推定値はこの値より小さいから明らかに有意でない。

この場合には、各資料源の平均値間には差があるから　一括した回帰係数を推定するのに合計の平方和は使えない。この様な差の生じた原因により　なすべきことは変つてくるが、それが観測者、機械の相違による当然の変動と見なされるならば、資料源間の平方和と積和の分散を資料源内のそれに付け加えねばならない。この分散は2つの分散の推定値の比0.146／2.223=0.066により決定され

したがつてこの場合回帰係数の推定に使うべき平方和、積和は

10+3×0.066 = 10.20

5+4×0.066 = 5.26

であり回帰係数の推定値は5.26／10.20=0.52　となる。

同じ様な性質をもつている多くの問題が共分散分析を用いて処理される。例えば最初の分散、共分散分析でそれに相当する成分を除くことにより資料源毎に変る変量、例えば平均気温を扱うことができる。回帰係数の最終的推定の時には適当な分数を乗じたこれらの変量に対応する平方和、積和が再び導入される。

色々な資料源による結果を分析する際に生ずる複雑な問題を取り扱うことは本書の範囲外であるが　上記の例は共分散分析がこの様な結果の分析の基礎となることを示すのに役立つに違いない。

8.7　分散、共分散の成分の推定

Estimation of components of variance and covariance

一組の観測値の変動共変動をいろいろな原因に帰因する部分に分割する方法を分散、共分散分析は与える。この方法によれば大きな変動を生ずる原因の意義を推定することができ、これを除くことによつて残りの"原因不明"の変動に焦点を合すことができるのである。時には　これとは逆に原因不明の変動性を除いて既知の変動因に焦点を合すために分散、共分散分析を使う必要のある場合もある。あるいはまた、解析法を変えて、観測値の取り方の相違によつて起る事態を推定することもできる。いづれの場合にも　分散、共分散の成分を推定する必要がある。

この様な解析の使用法を示すため　まず K.Mather※ の行つた解析を考察してみよう。Mather は別の目的で　大麦の雑種 Spratt × Goldberg の第2世代からとつたデータを使つた。この世代の170ケの個体から夫々2本の穂をとり各穂についていろいろな測定を行つた。8.7 a表はこの2つの測定値の分散共分散

(※　Mather, K., Biometrical Genetics, London, 1949)

分析を示している。Bは（中心の6つの節間の）最大巾、Cはその長さである。

87a表　分散、共分散分析

	自由度	B 平方和	分散の推定値	BとC 積和	共分散の推定値	C 平方和	分散の推定値
個体間	169	14206	841	−33223	−1966	122861	7270
個体内	170	2300	139	−1395	−082	8565	504
計	339	16506	—	−34618	—	131426	—

個体間の変動は極めて有意であることが分る。即ち　どの測定値においても　個体内より個体間の変動の方が大きいことが示されている。分散、共分散の推定値の差を使えば分散（又は共分散）の増加する程度が測れる。841−139=702, −1966−(−082)=−1884, 7270−504=6766

さて今度は　個々の測定値の変動又は共変動が2つの部分から成つているものとして考察する；xを個体内変動、Yを個体間変動とする。したがつて各個体からとつた個々の観測値間の分散、共分散はx＋Yに等しい。各個体からn個の観測値がとられたとすると観測値の平均値間の分散共分散はx/n＋Yで表わされる。なぜならば各個体についてn個が測定されているから前の成分はnで割るが後の成分は個々の個体に関係するものであるからそのままにしてある。これに対応して　分散分析では個体間の分散、共分散の推定値は　n(x/n+Y)=x+nYである。この方法によれば各成分を別々に推定することができる。

上記の例では、nは2である。したがつて上の差は成分Yの2倍の推定量である。87b表は87a表の分析の結果から求めたx、Yの推定値である。

87b表　分散、共分散成分の推定値

	B	BとC	C
x	139	−082	504
Y	351	−942	3383
x＋Y	490	−1024	3887
1/10x+Y	365	−950	3433

この表から個々の観測値の分散、共分散x＋Yや組平均、即ち10個の観測値の分散、共分散　1/10x＋Y　が推定できる。これはどの様な場合においても相関係数や回帰数を推定するのに用いられる。例えば個々の穂についてBを測定し、Cの値を豫測するのにこれを用いるとすれば、回帰係数の推定値は−1024/490=−209　となる。10個の観測値からなる組に対しては　回帰係数は　−350/365=−260　となり　観測数が非常に大きくなれば−942/351=−268　となる。

この方法によれば抽出変動やその他の原因不明の変動を除いた時の回帰関係や相関係数を推定できることが分る。特に、この方法は環境による変動を除いた時の変動の遺伝的成分間の相関を推定するため遺伝学で広く用いられている。例えば同一の双子内および双子間の測定値を解析することにより遺伝上の測定値の相関の程度が推定できる。

分散成分の第2の例として5図に示してあるデータの解析を考えてみる。これは87c表に示してあり　この表により年単位の年令の組間の差を表わす分散と年令に対する直線回帰が推定される。

87c表　対数で表わした陳代率mと対数で表わした体重wの分散、共分散分析

	自由度	m 平方和	分散の推定値	mとw 積和	共分散の推定値	w 平方和	分散の推定値
年令に対する直線回帰	1	1.298	1.2980	2.456	2.4560	4.646	4.6460

年令組間の残差	11	0.022	0.0020	0.023	0.0021	0.052	0.0047
年令組間	12	1.320	0.1100	2.479	0.2066	4.698	0.3915
残差	310	0.724	0.0023	0.724	0.0023	1.517	0.0049
計	322	2.044	—	3.203		6.215	—

各年令級の観測数の平均 n=323/13 を使えば年令の組内および組間の分散、共分散の成分が推定される。これによつて87d表に示してある分散の成分が得られる。

87d表　分散、共分散の成分

	m	m と w	w
x	0.0023	0.0023	0.0049
T	0.0043	0.0082	0.0156

しかし　この場合にはこの方法はどちらかといえば不自然である。回帰分析は　年令に対するmとwの従属関係は、いづれも直線であることを示しているから、組間の変動の成分は回帰線による変動を表わしている。この事実を認めればもつとうまく表すことができる。したがつてtが年令を表わすとすればtに関するmの従属関係は m=a+b(t−t̄)+ε で与えられ

mの分散＝回帰線の周りの分散＋(t−t̄)²×回帰に帰因する分散

＝x＋(t−t̄)²T

その他の分散、共分散も同様な式で表わされる。

Tを推定するには、直線回帰の分散の推定値は X＋TΣ(t−t̄)² であることに注目しなければならない。残差分散と回帰による分散の推定値の差を Σ(t−t̄)² で割ればTが得られる。

この場合 Σ(t−t̄)²=3607 であるから分散、共分散成分の推定値は87c表に示してある様になる。

87c表　分散、共分散の成分

	m	m と w	w
X	0.0023	0.0023	0.0049
T	0.000359	0.000680	0.001287
X+20T	0.00948	0.01590	0.03064
X+2T	0.00302	0.0366	0.00747
$\frac{1}{10}$X+20T	0.00741	0.01383	0.02623

成分Tについてmとwとの相関が1.0004となり1より大きくなることに気付くであらう。回帰成分の推定には常にこの様なことが起る。

今や、色々な抽出条件の下で起ると思われることについて研究できる。例えば、tの分散が20になる様に標本がとられたとするとこれに相当するm，wの分散、共分散はX＋20Tから推定される。この値は87c表に示してあり　この場合wに対するmの回帰係数は0.01590/0.03064となることが分る。これに対してtの分散が2に等しければ回帰係数は 0.00366/0.00747=0.490　となる。

又　平均年令の分散が20である10人の小児の組平均が解析に使われたならば分散、共分散の推定値は　$\frac{1}{10}$X＋20T　となる。87c表から回帰係数は 0.01383/0.02623=0.527となることが分る。

最後に同じ歳の多数小児からなる　組を使えば回帰係数は 0.000680/0.001287　となる。

この方法によれば希望する回帰係数、相関係数の値を分散の成分から推定できる。

この成分および誘導された、統計量の誤差を求めるには幾分複雑であり、I. Bross※　の行つた研究を参考とする必要がある。

大きな自由度を使うことができれば　これは次に示す様に簡単に行なえる。　$S^2=a_1s_1^2+a_2s_2^2$　を誘導された成分、s_1^2，s_2^2　は

自由度 f_1 ・ f_2 に基ずく分散とすれば s^2 は自由度

$$F = \frac{s^4}{\frac{a_1^2 s_1^4}{f_1} + \frac{a_2^2 s_2^4}{f_2}}$$

に基ずくものとして検定される。

例えば87b表で各個体について10本の穂を測つた時のBに関するCの回帰を推定したいとする。したがつてBに対して

$$1/10X+r = 3.65$$

$$= \frac{1}{10} \times 1.39 + \frac{1}{2}(8.41-1.39)$$

$$= \frac{1}{2} \times 8.41 - \frac{2}{5} \times 1.39$$

$$= 4.205 \quad -0.556$$

この2つの分散はそれぞれ自由度169と170に基ずくものである。故にこの推定値の有効自由度は次式で与えられる。

$$F = \frac{(3.65)^2}{\frac{(4.205)^2}{169} + \frac{(0.556)^2}{170}} = 125$$

大さつぱな分散分析表が作られ、87f表に示してある様に回帰係数およびその標準誤差が推定される。

87f表　近似的な分散分析

	自由度	平方和	分散の推定値
回帰	1	$(-11875)^2/45625 = 3090.75$	
残差	124	$42912.5 - 30907.5 = 12005.0$	96.8
計	125	125×34.33	$= 4291.25$ 34.33

$b = -11875/45625 = -2.60$

b の標準誤差 $= \sqrt{\frac{96.8}{45625}} = \pm 0.146$

この方法は近似的なものに過ぎないが実用的には充分正確な結果を与える。

9
大規模調査　Large－Scale Investigation

9.1　観測値の収集　Taking the observations

連関の問題に関する大規模調査では出来るだけ作業量を減らし、精度の高い結果の得られることが望ましい。数個の変量についての回帰分析や相関分析は時間のかゝるものとなり、無効でしかも不必要に時間のかゝる解析がおこなわれた発表論文は費用がかさむと同時に徒らに煩雑なものである。したがつて不必要かつ無効の努力が払われることに対してどの様な保証がなされるか或は予防策がこうじられるかを予め考察しておくことが望ましい。

予防策の一つとしては　確実な結論を導び出すことができ　できれば解析を簡単にする様な方法で　観測値をとることである。多くの場合観測値の収集は殆んど　否　全然制御することはできないのでこの方法は常に実行可能とは限らないが、もし制御可能であれば観測値収集方法について考慮を払うべきである。

試験研究では　最大の正確度と最小の計算量で変量間の関係を調べることが出来るように設計がなされてきた。この種の設計の一つが要因配置法である。この場合各独立変量によつてとられる値は限定されているが　観測値は独立変量の値のあらゆる組合せの生じた順に並べられる。例えば　3の独立変量の効果を　全部で $27(=3\times3\times3)$ の独立変量の組合せを使つて推定したい。この場合　独立変量はいづれも3個の値をとることに限定されている。

この場合　独立変量は肥料の使用量、薬品の服用量、吸入量の水準

その他の処理であり一方これに対応する従属変量は収穫量、致死率体重の増加量その他反応の測定値であらう。

要因分析は各独立変量の回帰係数が従属変量の値から直接に推定されるという長所をもつている。さらに各変量が独立に働いているか、或は x_1 x_2 , $x_1^2 x_2$ ※ の様な積の項を含むように、回帰が拡張されるかどうかを簡単に推定できる。

この様な積の項は一般に独立変量の交互作用として知られているものを検定するのに用いられる。

要因配置法は正規回帰分析を使つて解析ができるが計算量を最小にするためその手順は公式化されている。この手順の詳細については実験計画の書物を参照されたい。

多くの場合 独立変量の値を限定した計画を樹てることは不可能であるし又好ましくもない。独立変量が囲の中の動物の数とか宇宙の微細物質の密度の様な量であれば これらの値を制御することはできない。選択された設計にかなつた独立変量の値だけを選び出すことは可能であるが この様なことは一般に好ましくない。

実験の場合のように独立変量の値が制御できれば その関係が推定された値の範囲内においてのみ、推定された関係は一つの意味をもつている。なおこの数値の範囲は任意にとれるであろう他方独立変量の値が選択によつて得られたものであれば、推定された関係が示す事態は実際に起らないこともありうる人為的なものである。したがつて この関係は誤解を招きやすい。

例えば全予算に対する娯楽費の百分率におよぼす所得と家族数の影響を推定したいとする。これを行うため幾つかの所得の値と可能な家族数とを選び、収入と家族数の組合せ毎に一個の観測値がとられる。正規母集団においては所得と家族数のあらゆる組合せが同じ頻度で起らないであろうという点、例えば大きな所得と大きな家族の組合せは比較的に稀であるという点でこの手順は人為的なものである。他の観測値と同じ重みで計算すれば所得と家族数の平均効果

(※ 積の項のあてはめの説明は6.4節を参照のこと)

の推定値は不正確なものとなるかもしれない。

独立変量が実験の場合のように制御出来ない時には 第一に考察しなければならぬことは収集された観測値は推定された関係が、この場合に適用できるようにある定められた状態を表わしていることを保証できるものでなければならない。このことは実験の場合と同様 観測値の選択には確率的要素が含まれており データを規則的に収集する傾向が、その解析で認められることを示している。したがつて 例えば各種の資料源からとつた観測値を解析するには同じ資料源からとられた観測値の類似性を考慮に入れねばならないが(8.7節)時系列解析では連続した観測値の類似性の影響を考慮する必要がある。(11章)

従属変量は一般に抽出方法を修正する基礎として用いてはならないという例外はあるが 観測値の収集の際 修正した抽出方法を採用することはもちろん可能である。したがつて層化抽出法によれば独立変量やその他の利用しない変量が正しい割合で標本の中に表われる様に独立変量に対してそれを用いることができる。例えば所得を独立変量とすれば各所得の組が比例的に表われる様に家族数が選ばれる。さらに この方法は観測値が得られた地域毎に行なわれる。この様にすれば地域間の関係に含まれている変動を調べることができ、その型についてもつと正確な概念がえられるであらう。

8.2 観測値の吟味 Scruting of observation

大規模な解析に着手する前に予防策として、極端に異常な観測値やとび離れた観測値がデータに含まれないようにする。いくつかの変量を含む解析では 誤つて報告された観測値は係数の推定値にかなりの影響をおよぼし、たとえその観測値が明らかに誤つたものであつても 解析をかなり進めてその誤りを発見することは殆んどできない。したがつてかけ離れた値を発見するには 観測値をさらに調べてみる必要がある。

極端な値はグラフにプロットしたものと一緒に調べてみると簡単に分ることを第1章で指摘した。数個の変量がある場合には あら

ゆる変量の対をグラフにプロットすることはかなり時間がかゝり、それと同時に極端な値が生じ易い時には不適当である。この場合大体の回帰方程式の推定が可能であれば、通例　この推定値からの偏差を調べることが　すすめられる。又　データの検討或はその関係の形に関する一般的な知識から　回帰を大略推定できれば　次の解析を簡略にしかつ吟味するためにこの値が用いられる。

例として　9 2 a 表は　W.L. Fons※ が報告 した密閉疎な薪炭林における森林火災の延焼に関する４９の実験結果が示してある。測定された変量は

$x_1 = \frac{3}{2}$ (m.p.h 単位の風速の対数)

$x_2 = \frac{1}{500}$ (°F 単位の薪炭材の温度)

$x_3 = \frac{\cdot}{60}$ (湿度百分率)

x_4 = in 単位の薪炭材の表面積と材積との比の対数

x_5 = in 単位の薪炭材間隔面の対数

y = ft/hr 単位の延焼速度の対数

9 2 a 表　火災の延焼に関する実験結果

x_1	x_2	x_3	x_4	x_5	x_6	y
0.90	0.16	0.21	1.32	0.18	2.35	2.29
0.90	0.17	0.15	1.32	0.18	2.42	2.35
0.90	0.17	0.17	1.32	0.18	2.40	2.32
1.16	0.14	0.20	1.20	0.18	2.48	2.35
1.16	0.18	0.18	1.32	0.18	2.66	2.42
1.16	0.16	0.20	1.32	0.18	2.62	2.45
1.16	0.17	0.16	1.32	0.18	2.67	2.53

(※ Fons, W.L.J. Agric. Res. 72 (1946) 93

x_1	x_2	x_3	x_4	x_5	x_6	y
1.16	0.14	0.19	1.44	0.10	2.65	2.56
1.35	0.14	0.21	1.20	0.24	2.72	2.35
1.35	0.15	0.18	1.20	0.24	2.76	2.42
1.35	0.12	0.19	1.20	0.21	2.69	2.41
1.35	0.14	0.18	1.20	0.18	2.69	2.47
1.35	0.14	0.20	1.20	0.18	2.67	2.38
1.35	0.14	0.18	1.20	0.14	2.65	2.49
1.35	0.12	0.19	1.20	0.10	2.58	2.39
1.35	0.12	0.16	1.32	0.10	2.73	2.56
1.35	0.12	0.20	1.32	0.05	2.64	2.45
1.35	0.17	0.16	1.32	0.18	2.86	2.69
1.35	0.15	0.17	1.32	0.18	2.83	2.55
1.35	0.16	0.21	1.32	0.18	2.80	2.60
1.35	0.14	0.18	1.32	0.14	2.77	2.63
1.35	0.14	0.19	1.32	0.10	2.72	2.53
1.35	0.12	0.18	1.32	0.10	2.71	2.52
1.35	0.15	0.19	1.44	0.14	2.89	2.66
1.35	0.12	0.18	1.44	0.14	2.87	2.57
1.35	0.12	0.15	1.44	0.14	2.90	2.70
1.35	0.15	0.17	1.44	0.10	2.87	2.79
1.35	0.12	0.19	1.44	0.05	2.77	2.70
1.35	0.13	0.22	1.44	0.00	2.70	2.60
1.35	0.15	0.18	1.44	0.00	2.76	2.67
1.15	0.14	0.19	1.32	0.14	2.56	2.35
1.15	0.14	0.20	1.32	0.10	2.51	2.37
1.15	0.14	0.18	1.20	0.10	2.41	2.23
1.15	0.14	0.19	1.20	0.18	2.48	2.17

x_1	x_2	x_3	x_4	x_5	x_6	x
1.15	0.18	0.08	1.44	0.10	2.79	2.66
1.15	0.18	0.09	1.44	0.05	2.73	2.64
1.15	0.18	0.09	1.32	0.18	2.74	2.51
1.15	0.19	0.08	1.20	0.24	2.70	2.37
1.15	0.18	0.08	1.32	0.05	2.62	2.47
1.15	0.18	0.10	1.44	0.14	2.81	2.60
1.15	0.18	0.09	1.20	0.14	2.58	2.35
1.15	0.18	0.08	1.20	0.21	2.66	2.38
1.15	0.17	0.12	1.44	0.00	2.64	2.56
1.15	0.16	0.12	1.32	0.14	2.65	2.44
1.34	0.18	0.09	1.20	0.24	2.87	2.59
1.34	0.18	0.09	1.32	0.14	2.89	2.60
1.34	0.17	0.11	1.20	0.14	2.74	2.40
1.34	0.17	0.11	1.20	0.21	2.81	2.39
1.34	0.17	0.12	1.20	0.18	2.77	2.41

W．L．Fon は問題の理論的な面を追及して、y と他の変量との間の理論的（基礎的）関係を誘導した。

この関係は近似的に次式で表わされる。

$$y = x_1 + x_2 - x_3 + x_4 + x_5 + \text{常数}$$

したがつて y と $x_1 \sim x_6$ との回帰関係を推定するためにこの観測値を使用したい場合　極端な観測値を調べるには y と

$x_6 - x_1 + x_2 - x_3 + x_4 + x_5$ を比較すればよい。

x_6 の値は 9 2 a 表に示してあり、極端な値を検定するにはグラフその他の方法が使用できる。例えば　9 2 b 表に示してあるように偏差 $y - x_6$ の分布を作ることができる。

9 2 b 表　偏差 $y - x_6$ の頻度分布

偏差	頻度
−0.05−	8
−0.10−	7
−0.15−	10
−0.20−	10
−0.25−	6
−0.30−	5
−0.35−	2
−0.40−−0.44	1
計	49

この表をみれば最大の偏差は −0.40 と −0.44 の間にあり（9 2 a 表の下から二番目の観測値）この表は残りの観測値とよく一致していることが分る※ −0.60 或は +0.20 という偏差があれば　これに対応する観測値は当然棄却される。しかしこの場合は全観測値が偶然変動の範囲内にあり　したがつて全部を完全な数値解析に使用すべきである。

9 3　解析法の計画　Planning the analysis

解析法を予め計画できると迅速かつ適切な作表、表示方法を工夫することが可能であるばかりでなく、解析がなされる形式について考察すれば、データ収集法の修正が暗示される場合が多いから大規模調査の解析法は調査に着手する前に考察する必要がある。

例えば　解析に 7 章で考察した形式の組分けを使用することが決められたとすると、独立変量のとる値の順序又はそれに近い順序でデータを作表するとよい。このようにすると比較的簡単に組分けを

（※　観測値の中心部の 2/3 は −0.10 と −0.30 の間にある。したがつて標準偏差は大体 0.10 であり、観測値の 95% は 0.00 と −0.40 の間にあることが期待される。この範囲の外にある 1 個の観測値は棄却の必要はない。）

行うことができる。或は　この場合には　簡単に分類や配列のできるカードにデータを書き込むことも考えられる。
余分な仕事の大部分は非常に桁数が多いか非常に少い数値を解析で取扱うために生ずるものであるから報告する価値のある有効数字の桁数は調査の初期の段階で考察すべきである。

独立変量に桁数の多いもの或は少い数値を用うるかについての判断は　最後の桁数が変つた時従属変量の値に著しい影響をおよぼすかどうかを考察することにより決定される。従属変量については、その残差標準偏差の1/4 以内にある値は報告すれば充分である。

例えば前節の例では従属変量は約0.10の残差標準偏差を持つている。したがつて　0.025 以内又は小数点以下2桁のyを報告すれば充分である。この例の独立変量の変動はyに同じ増大をもたらすことが期待されるからである桁数はyと同じでよい。

解析を行うに当つて主として考えねばならない点はどの様な型の計算機が使用可能か或は使用すべきかということである。計算設備に制約があれば効率の悪い簡略解析法を用い、これを補うためにより多くの観測値をとればよい。逆に計算設備に制約がなく、データから出来るだけ多くの情報を引き出すのが重要であれば　完全解析が行なわれる。

完全解析を行う時生ずる計算に関する主な問題の一つは平方和、積和の計算である。

最新式の机上計算器を使用すれば　この種の計算に要する仕事量は大いに減小する。（計算機および計算者の経験によつて変るが、2～3桁の100個の数字の平方和、積和は大体10分以内で計算できる。）　その場合でも変量および観測数が共に大であれば、仕事の量は莫大なものとなる。変量が10で観測値が1000組あると平方和、積和を求めるのに55000個の平方および積を計算しなければならない。（連続計算して約90時間）
この場合吟味は全然行われていない。
このような場合については当然便利な計算の解析法が考案されねばならない。その一つは英国の Hollerith, U. S の I. B. M. のようなパンチカード式の計算機を使うことである。これらの機械は80列からなり、各列には10 position のあるカードを使用する（外の分類を行うため各列には餘分に2個の position が付け加えてある。）各列、各 position に穴をあける。即ち　例えば3桁の数字は3つの列で表わされる。例えば329は定められたらんでそれぞれ3，2，9番目の position に穴をあけて表わされる。この様にすれば　同じカードに多数の数字　例えば40個の2桁の数字、26個の3桁の数字を表わすことができる。

パンチされた情報にしたがつてこれらのカードを分類したり並べたりする機械が使用され、この機械は又　このカードの数字の平方和、積和を求めるために使用できる。大規模調査の解析にこの種の機械を利用することは考察してみる価値がある。

組分けした解析が行われる場合には　この種の機械は観測値を組に分け各組の総計を求めることができる。　又　完全解析が行われる場合には観測値の平方和、積和は組毎に決定される。例えば変量10で1000組の観測値の平方和、積和を計算するには10変量の各組をカードにパンチし機械によつて各変量とカードにパンチした全変量との積和が順次　計算される。故に10組の積は全ての平方和、積和を与える。2～3桁の場合には　各操作は一時間当り10000枚の割で行われる。　したがつて計算の全過程は一時間で完了する。

これにはカードにパンチしたり　機械を調整するに要する時間―さらに2時間かゝる―は含まれていないが　机上計算機より遙かに早く結果を求めることができる。

こゝでパンチカード式機械の詳しい説明をすることは適当でないと思われる。詳細については　計算法を取り扱つている文献を参照されたい。

特に F. Yates※ は数節にわたつてパンチカードによる方法およびパンチカードに用いられる機械について説明している。

パンチカードを使用するか否かを決めるには計算の正確さと同時にどの様な保証がえられるかを考察してみる必要がある。繰返し解析することができれば誤りを見落す機械は減るであらう。しかし各解析について完全に繰返しをすると時間がかゝるから一般に計算を吟味する手段を解析に加えることが望ましい。

これを行うには解析の際　特別の変量として任意の変量の組合を使用すればよい。この特別の変量の平方和、積和は直接計算され他の変量の平方和、積和との適当な組合せに対して吟味される。特に前節に示した方法で観測値を調べるために回帰の概略の推定値が使用されているならば　平方和、積和を吟味する場合にこれが使用される。

例えば前節の $x_1 \sim x_6$ に対する変量 y の回帰係数を推定する方程式は次の様に計算される。

$$0.8072b_1 - 0.0625b_2 + 0.0540b_3 - 0.0343b_4 - 0.0335b_5 + 0.6229b_6 = 0.3980$$
$$-0.0625b_1 + 0.0225b_2 - 0.0343b_3 - 0.0064b_4 + 0.0172b_5 + 0.0048b_6 = -0.0059$$
$$0.0540b_1 - 0.0340b_2 + 0.0937b_3 + 0.0083b_4 - 0.0096b_5 - 0.0750b_6 = -0.0306$$
$$-0.0343b_1 - 0.0064b_2 - 0.0083b_3 + 0.4103b_4 - 0.1731b_5 + 0.1882b_6 = 0.4291$$
$$-0.0335b_1 + 0.0172b_2 - 0.0096b_3 - 0.1731b_4 + 0.1841b_5 + 0.0043b_6 = -0.1601$$
$$0.6229b_1 + 0.0048b_2 - 0.0750b_3 + 0.1882b_4 + 0.0043b_5 + 0.8952b_6 = 0.6917$$

しかし　$x_6 = x_1 + x_2 - x_3 + x_4 + x_5$　であるから　この方程式は独立でない。したがつてb式は1，2，4，5式の和から3式を引いたものに等しいということから　この計算は吟味される。例えば $-0.0343-0.0064-0.0083+0.4103-0.1731=0.1882$ である。

これらの方程式は対称であるため計算式の係数の　誤りはある特定の平方和、積和に含まれている誤差に帰因していることが多い。例えば b_4 の係数だけがあやまつている時には　恐らく　x_4 の平方和に誤りがあるものと考えられる。しかし　b_4 と b_5 の係数の双方に誤りがある時には恐らく x_4 と x_5 の積和に誤りがあるものと考えられるが　この場合には x_4，x_5 の平方和にも誤差があるものと考えてよい。

3.4　変量に関する考察 Examination of variables

解析に多数の変量が含まれておれば　従属変量の推定に有意かつ評価しうる貢献をなす変量だけを取上げることが望まれる。これは曲線回帰の逐次検定法で説明したのと同じ解析法で行なえる。（6.3節）次の例にその方法を示している。

ある試験で　学生の成績 y，を予測する目的で58人の学生について5つのテスト $t_1 \sim t_5$ が行なわれた。y の総平方和は 94.91。で、回帰係数 $b_1 \sim b_5$ を推定する方程式は

$$96643b_1 + 56914b_2 + 43921b_3 + 53707b_4 + 7328.8b_5 = 38285 \quad \text{平方和} \ 151.67$$
$$56914b_1 + 631122b_2 + 411959b_3 + 301136b_4 + 569274b_5 = 79040 \quad \text{平方和} \ 9899$$
$$43921b_1 + 411959b_2 + 642388b_3 + 325829b_4 + 488636b_5 = 88917 \quad \text{平方和} \ 12308$$
$$53707b_1 + 301136b_2 + 325829b_3 + 464943b_4 + 351362b_5 = 46652 \quad \text{平方和} \ 4681$$
$$73288b_1 + 569274b_2 + 488636b_3 + 351362b_4 + 698741b_5 = 94663 \quad \text{平方和} \ 12825$$

各変量が単独に説明する平方和が　この方程式の右端に示してある。　例えば t_4 は平方和 $(46652)^2/464943=4681$ を説明している。

t_1 に帰因するものが全平方和の大部分を占めており、これを除いた残りは $94.911-15.167=79.444$ であることが分る。したがつて

1)式に適当な乗数を掛けて他の方程式から引けばその効果は除かれる。例として1式を53707/96643＝ 55573 倍して4式から引く。このようにすれば 次の方程式が得られる。

$$597605\ b_2 + 386093\ b_3 + 269507\ b_4 + 526114\ b_5 = 56496$$
平方和 53.41

$$386093\ b_2 + 622427\ b_3 + 301421\ b_4 + 455329\ b_5 = 71518$$
平方和 8218

$$269507\ b_2 + 301421\ b_3 + 435097\ b_4 + 310634\ b_5 = 25376$$
平方和 1430

$$526114\ b_2 + 455329\ b_3 + 310634\ b_4 + 643164\ b_5 = 65630$$
平方和 6697

この残りの変量が説明している平方和は方程式の次に示してある。

今度は変量t_3が残りの平方和の大部分を説明しており、したがつてこれを除いた残りは79744−8217−71526である。したがつて、2)式に適当な乗数を掛けて他の式から引いて これらの式からb_3を消去することにより その効果は除かれる。このようにすれば次の新しい方程式がえられる。

$$358111\ b_2 + 82535\ b_4 + 243672\ b_5 = 12131$$ 平方和 411

$$82535\ b_2 + 289129\ b_4 + 90133\ b_5 = -9258$$ 296

$$243672\ b_2 + 90133\ b_4 + 310074\ b_5 = 13312$$ 572

残りの変量によつて除かれる平方和は極く僅かであるから この手順はこゝで終る。

$b_2 = b_4 = b_5 = 0$ とおけば

$b_3 = 71518/622427$

$= 0.0115$

$b_1 = (38285 - 43921 \times 0.0115)/96643$

$= 0.3439$

分散分析は 94a表 のようになり、推定された係数の標準誤差が計算される。

94a表 回帰に対する分散分析

	自由度	平方和	分散の推定値
t_1によるもの	1	15167	15167
t_3によるもの	1	8218	8218
t_1とt_3によるもの	2	23385	
残差	55	71526	1300
計	57	94911	

この方法で解析を行うには次の3点に注意しなければならない。第1に 必要がなければ独立変量の積和を計算する必要のないことが分る。この方法によれば次の解析で重要でなくなる独立変量の積和の計算を避けることができる。したがつて 例えばt_2 t_4 t_5の積和は上記の解析では不必要であり使用されないし積和の逐次計算法が採用されゝば、この積和は決して計算されなかつたであろう。

前の3節で示したような回帰方程式に対する第1近似が得られるときには この事は特に役に立つ。上記の解析によれば この近似に必要な修正が推定されるであらう。これが良好な近似を示しておれば、他の変量および変量の組合せは回帰に明らかに関与しないものと思われる。

これが最良又はそれに近い関係を表わしていない時には 他の変量は回帰に対して有意な寄与をしている。

例えば 前節の例で x_6 を回帰方程式から削除すれば 次の方程式がえられる。

$$0.373772\ b_1 - 0.065840\ b_2 + 0.106187\ b_3 - 0.165254\ b_4 - 0.036492\ b_5 = -0.083300$$
平方和 0.0186

$$-0.065840\ b_1 + 0.022474\ b_2 - 0.033598\ b_3 - 0.007409\ b_4 + 0.017177\ b_5 = -0.009609$$
平方和 0.0041

$$0.106187\ b_1 - 0.033598\ b_2 + 0.087416\ b_3 + 0.024067\ b_4 - 0.009240\ b_5 = -0.027351$$
平方和 0.0086

$-0.165254b_1 -0.007409b_2 +0.024067b_3 +0.370734b_4 -0.174004b_5$

平方和 $= 0.283682$　0.2171

$-0.036492b_1 +0.017177b_2 -0.006240b_3 -0.174004b_4 +0.184079b_5$

平方和 $= -0.163423$　0.1450

1），2），4），5）式の和から3）式を引いたものは0であるということにより　この削除は吟味される。

x_6 の効果を除けばyの平方和は0.8824から0.3479に減小する。しかし　これでは他の変量を無視することはできないが x_4 は0.2171という平方和を説明している。故に b_4 を削除すれば次の方程式がえられる。

		平方和
$0.300110b_1 -0.069143b_2 +0.116915b_3 -0.114054b_5 = 0.043151$		0.0062
$-0.069143b_1 +0.022326b_2 -0.033117b_3 +0.013700b_5 = -0.003940$		0.0007
$0.116915b_1 -0.033117b_2 +0.085854b_3 +0.002056b_5 = 0.008935$		0.0009
$-0.114054b_1 +0.013700b_2 +0.002056b_3 +0.102410b_5 = -0.030277$		0.0090

これは1），2）4）式の和から3）式をひくことにより吟味される。

今度は他の変量で説明される平方和は極めて小さいから（即ちこみにして0.0120を説明している）x_4，x_6 に関するyの回帰方程式は次のようになる。

$$y = 0.765x_4 + 0.612x_6 - 0.159$$

$$= 0.612x_1 + 0.612x_2 - 0.612x_3 + 1.377x_4 + 0.612x_5 - 0.159$$

比較：全変量を用いて推定された回帰方程式は

$$y = 0.675x_1 + 1.185x_2 - 0.366x_3 + 1.257x_4 + 0.305x_5 - 0.164$$

この回帰は理論的関係とはかなり違つている。これは仮定した基礎的）関係が不正確であることを示しているのではなく　予測を目的としたときには　上記の方程式にあてはめて使用すればよりよき予測値がえられることを示しているに過ぎない。しかし極めて異常な観測値のないことを確めるため、あてはめた回帰からの偏差を再び調べてみたい。これを行つてみると最大の偏差は－0.129即ち残差標準偏差の2.42倍であることが分つた。

これはこの観測値の棄却の理由となる程大きなものではない。

さらに、上記の手順を使用した時、変動の最大の量を説明する変量を繰返し選んだとしても変動の大部分を説明する成分の組合せが得られるとは限らない点に注意を要する。一般にはその通りであるが変量間に高い相関関係があればそうでない場合もある。例えば上例の t_2 と t_5 の積和が充分大であると t_2 も t_5 も別々では非常に大きな平方和を説明していないにもかゝわらず、t_2 と t_5 をこみにしたものに帰因する平方和は t_1，t_3 の両者によるものより大となるであろう。

変量のあらゆる組合せに対する回帰を求める以外に、この形の発生について保証することはできない。しかし、普通には、独立変量の積和を検べてみると、このようなことの起り易いことが分る。この可能性を防ぐ必要があれば、積和を求めたければならない。

この様な場合には、解析の段階で任意の変量の平方和は、これと高い相関関係のある変量を削除することにより増大するとにも注意しなければならない。したがつてどこで解析を止めるかを決めるには、高い相関のある変量の対を調べてみるとよい。上例では t_2，t_4，t_5 をこみにしたものに帰因する平方和は僅かに1258であり、したがつて、この組合せは重要でない。最後に上の解析では有意性については全然説明されなかつたことに注意せよ。平方和の大きさ即ち変量の予測された値だけに関心の払われる場合が多いが　場合によつては変量の有意性が大切なこともある。これは平方和と残差との平均平方とを比較することにより、各段階で検定される。しかし推定値に含まれるものが有意水準に達した変量だけであれば、検定される変量の数によつて有意水準を調整しなければならない。多数の変量が検定されるならば、最大の分散比は20回に1回以上即ち5%有意水準に達するであらうということを考慮すべきである。このことは有意水準を変えることによつて大体補なわれる。※

一般には5％又は1％の有意水準が使用される場合には、変量の

※ Hartley, H.O.J.R.statist Soc., B.5 (1938) 80

数で有意水準の値を割れば十分である。例えば　上例でも t_1 を削除した時の分散比は１０６５である。５％水準でその有意性を検定するには５つの変量が検定されるということを考慮に入れておく必要がある。したがつてこの値の有意性を検定するため、Ⅷ表（１％水準）が使用されねばならない。その値は有意水準に達している。

同様に t_3 の分散比632　は5／4％－１２５％の有意水準を用いて検定される。この水準で有意であるために必要な値（補間によつて求める）は669　であるから t_3 は有意の分れ目にある。

任意の段階で、ある変量に帰因する平方和が他に較べて有意に大であるかどうかを決めるには4.9節の公式を用いねばならない。例えば x_2 と x_1 を用いたものが x_3 と x_1 を用いたものより有意に劣るかどうかを決めたいとする。

第２段の計算を使用すれば x_1 の効果を除いた時の偏相関係数がえられる。

$$r\ (x_2\ ,\ y) = \frac{56494}{\sqrt{597605 \times\ 79744}} = 0.2588$$

$$r\ (x_3, y) = \frac{71518}{\sqrt{622427 \times 79744}} = 0.3210$$

$$r\ (x_2\ ,\ x_3\ ) = \frac{386093}{\sqrt{597605 \times 622427}} = 0.6331$$

最初の２つの係数間の差を検定するため次式を計算する。

$$\frac{0.3210 - 0.2588}{2\ (1 - 0.6331)} = 0.085$$

これは５７対の観測値に基ずく相関係数として検定される。

x_2 の代りに x_3 を使用したことにより得られるものは餘り意味がない。したがつて分析には、これ以上の観測値が含まれている様な錯覚を起すものであることが分る。

特殊な問題とその解析 Special Problems and Analysis

10.1　回帰分析における仮定の検討

Examination of assumptions in regression analysis

4.1節で指摘したように、数値解析は観測値の性質に関する次の様な仮定に基ずいて行なわれる。

最も一般的な３つの仮定は

1. 観測値はそれがとられた状態を代表している。
2. 観測値は互に独立である。
3. 回帰線からの偏差は同一の変動性をもつて正規分布している。

これらの仮定の意味をそれぞれの節で調べられた。

解析を行う際には、この様な仮定或はそれ以外のものが合理的に正しいことを確めたいであろう。

そうでないと誤つた結論が下されることになるであろう。そこで、この節では仮定を検定する方法を考察してみる。

9.1節で指摘したように、上の第１期の仮定は観測値をとる前に考えねばならないことである。したがつて観測値がその状態を代表していることを保証するように調査計画が樹てられる。

通常この第１の仮設の検定では余り多くのことはできないが、従属変量と独立変量の値の分布が期待した分布と一致しているか否かを検べることのできる場合が多い。例えば、家族の収入に関する研究が行なわれる時にはこれらの収入の分布は、もし既知であれば彼等が住んでいる地区の分布と比較される。この場合差が大きいことは、この観測値が代表的なものでないことを示しているであろう。

一組の観測値の分布が、期待される分布と有意に違つているか否かを検定するために有意性の検定が行なわれるが一般にはその必要はない。というのは普通は差が推定された回帰に影響およぼ

す即ち偏りを生せしめるか否かを決めることの方が重要であるからである。もし偏りを生せしめるものであれば若干のデータを棄却するか、或はその状態を代表する観測値とするためにさらに多くのデータを収集しなければならない。しかし、このようにすれば解析に余分な因子が入る危険がある。一般には、後でこのようなことをするよりも全ての組を正しく代表するように調査計画を樹てることが望ましい。

観測値は互に独立であるという仮定は次の2つの場合には無用である。数組の観測値が違つた場所又は条件でとられたとすれば、異なる組の観測値間より同じ組の観測値間の方が一層類似しているであろう。これを検定し修正するには8.6節で概説した解析法を行う必要がある。

又時系列の場合の連続した観測値は似たような傾向をとるであろう。これを検定し修正するには特殊な方法によらねはならない。11章にこのことが概説してある。

上記の最後の仮定は残差の分布型に関するものである。この場合には幾分面倒になる。変動性が回帰線の部分によつて異るので必然的に重み付の解析（7.7節）又は従属変量の変換のいづれかを利用しなければならない。又、異常な観測値が解析に含まれているためか、分布型における一般的な差のため偏差は正規分布をしていないかもしれない。

異常観測値は2.5、4.7、5.7、9.2節で示したようにして調べられる。しかし回帰からの偏差が非正規であるか否かを検定することは難しい。

これを検定するには偏差の尖度、歪度を計算すればよいが、正規分布するとした時与えられた領域に入ると期待される観測数を推定する方が簡単かつ効果的である。これをその領域内にある数と比較する。

例えば、4図に示してあり、4.4節で解析を行つた頭と胸の周囲に関するデータでは、推定された回帰線は $H = 0.3401C + 30.60$

標準偏差の推定値は1.212であつた。観測値の90％は標準偏差の1.645倍、即ち回帰線から1.99の範囲内にあることが期待される，即ち観測値の90％は直線 $H = 0.3401 + 28.61$ と $H = 0.3401C + 52.59$ の間にあることが期待される。実際に123点の内112点即ち91％がこの線内にあり、5点が線の上に6点が下にあつた。この値は期待値に近いものであり、したがつて、実用的には偏差の正規性については殆んど疑問はない。

データの完全解析が不可能な場合であつても、簡便法による正規性の近似的な検定は可能である。例えば、歪度は推定された回帰線から上に、ある距離以上離れている観測数 n_1 と、下に同じ距離以上離れている観測数 n_2 とを比較すれば検定できる。この数の差は自由度1の χ^2 として $(n_1-n_2)^2/(n_1+n_2)$ を使つて検定できる。

例えば、3図のデータは7.7節で回帰線 $M = 519.3 + 21.703W$ を推定するために使用された。

これは観測値の組平均を使つて推定されたものである。個々の観測値が正規分布していなくても、その平均はそうなるであろうからこの推定値は観測値の正規性に関する仮定とは関係がない。※

※注　厳密に云えば、多数の観測値の平均値が正規分布するということは常に正しいとは限らない。しかし個々の観測値の分布が極度に異常でないかぎり実用的には正しい。

正規性を検定するために、例えば10％以上回帰線から離れている観測数を考察しよう。☆

☆注　この例では標準偏差は代謝率に比例して増すと仮定されているので百分率偏差が使われている。これには直線 $M = 0.90(519.3 + 21.703W)$ 以下にある観測数と直線 $M = 1.10(519.3 + 21.703W)$ 以上の観測数を数えればよい。その値はそれぞれ31と28である。

したがつてこの分布の歪度は

$$\chi^2 = \frac{(51-28)^2}{31+28} = 0.15 \quad \text{自由度 } 1$$

この値は有意でない、従つてこの対象性の検定に関する限りにおいて、正規性の仮定の正当なことを示している。同じ様な検定を行えばグラフの異なる領域でこの結論の正しいことが立証されるであろう。

非正規性の問題が重要になつてくるのは回帰分析だけであることに注意しなければならない。相関分析では、たとえ分布の型に疑問があつても、3章で説明した検定を行うか、係数の計算によりこの問題は解決される。これらの検定は正規性の仮定とは無関係である。

特に、観測値を順序に並べ、順位をつけることができれば、順位相関係数が計算され、連関を検定するのに用いられる。ある型の主観的、選択的測定値を取り扱うときにはこの方法は特に役に立つであろう。

10.2 回帰分析に用いられる変換
Use of trans Formation in regression analysis

前節で指摘したように、従属変量の変換はあてはめた回帰線からの偏差を正規分布に従せるために用いられ、さらに変換を用いれば回帰線の各部分の変動は等しくなる。

変換を利用する才2の理由は、その関係を単純化し、測定値を無理のない尺度にするためである。

例として、タラの体長Lに対する重量Wの従属関係が調査されたとする。重量は容積に略比例して増加し、容積は体長の3乗に比例して大となる。したがつてこの関係は、$W = aL^3$ 即ち $\log W = A + 3\log L$ で近似的に表わされる。これはLの3次式以上の多項式があてはめられるか、より簡便な方法としてはWとLが対数に変換されることを示している。

対数変換は最も一般的なものである。観測値の散ばりが従属変量の平均値に比例して増大する時には特に有効である。3図はこ

のような場合の例であり、5図は対数変換により、この難点がどのように除かれるかを示している。

対数変換により正規性、又は準正規性が満され、大きな正の偏差は大きな負の偏差以上に減少する傾向がある。しかし観測値の比例的変動が相当大きくない限り、普通目盛の分布と対数目盛の分布との差は実際に問題となる程大きくはない。

例えば3図のデータの変動係数！ 標準偏差を平均値で除したものは約8%である。観測値の対数が正規分布をしておれば、回帰線から上15.4%と下14.6%の間にその95%が入つていることが期待される。これは殆んど対象的であり、正規性からの差が生ずるのは数千回の観測の内2.3回以下であるから統計的に有意でなさそうである。

よく用いられるこれ以外の変換法は、平方根、$\sin^{-1} P$
$\frac{1}{B} \sinh^{-1} \beta x$ である。その使用法の詳細については
Introductory Statistics の8章を参照されたい。

どの様な変換を行う時でも、変換されない元の変量を予測するためには、推定された関係に含まれる偏りを除くため
Introductory Statistics に説明してある型の修正を適用しなければならない。これは変換変量の平均値は変換されない元の変量の平均値と正確には一致していないからである。

例えば、対数変換では、推定された関係に残差分散（回帰による平方和だけを除いたもの）の1.15（自然対数の½）を加える必要がある。

例えば、log（代謝率）とlog（体重）との回帰の推定式は $m = 0.5154W + 2.3117$ で残差の平均平方は0.0012である。この関係は任意の体重に対応するlog（代謝率）の平均値を推定するのに用いられる。しかし、任意の体重に対応する平均代謝率が必要とされゝば、方程式に $1.15 \times 0.0012 = 0.0014$ を加える必要がある。

$m = 0.5154w + 2.3131$

即ち $M = 2.056 W^{0.5154}$

この場合には偏りに対する補正は僅かであるが（約0.3％）、比例的変動が大きくなるにつれて偏りは無視出来なくなる。変動係数が５０％の時は偏りの補正は8.5％となるが、１００％であれば２７.１％となる。

10.3　プロビット変換 Probit transformation

変量間の関係を単純化したり、変量の測定尺度を改めるために、回帰分析に変換が使用されることがある。例えば、直接尺度で測つた指数は５０から１００の変化を１００から２００の変化と比較できるという錯覚を起しやすい。どちらの変化も１００％増加している。このことはある時点で指数５０の変化を生みだす条件は、別の時点で１００の変化を生ぜしめるということを意味している。この難点を克服するには指数の対数を使うと便利である。

この様にすれば尺度の各部分における比例的変化は数値的に比較できるであろう。

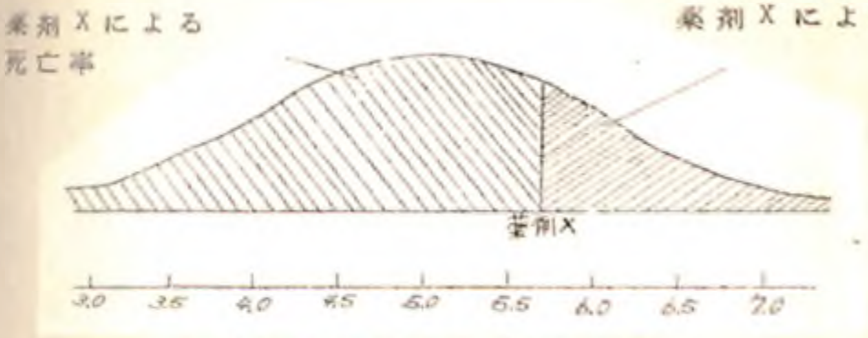
割合や比率を解析する必要のある場合にも同じ様な困難性が起る。

この場合には５０％～５１％の変化が、例えば９９％～１００％の変化と比較できることは稀である。最初の％の如何を問わず、百分率の変化を比較するには変換を用いる必要がある。プロビット変換が一般に用いられている。

プロビット変換は正規分布をする耐毒値 tolerance value の概念に基ずいている。次の例は、この概念を明確にするのに役立つであろう。動物にいろいろな量の毒薬を与え、量別に死亡した動物の百分率を記載する実験が行なわれたとする。

全ての動物が正確に同じ状態であれば、或る量以下では全部生き残り、それを越えれば全部死亡する。実際には動物が耐え得る量はいろいろであり、ある動物はある量で死亡し、他の動物は別の量で死亡する。

したがつて、各動物が毒等に耐え、生き残ることのできる最大量を示す耐薬力の分布を考察することができる。この様な分布の面積は与えられた量で生残るか、或は死亡する動物の百分率を与える。３５図はこの様な場合の例である。



第３５図　プロビット変換の例

投薬量（あらゆる独立変量が使われても）の目盛を適当に変えることで、耐薬力の分布が常に正規となる様に即ち３５図に示してある形の様に並べることができる。

プロビット変換を適用すれば、分布の横軸に対するこの分布の両端部の百分率が変つてくる。この様な場合横軸は平均が５、標準偏差が１の正規分布に対応する様に選ばれるのが普通であるがこれは便宜的なことである。真の結果は百分率のプロビットと投薬量（又は変換した投薬量）とが相互に直接的に対応しており、線型回帰分析によつて関係づけられるであろう。

１０.３ａ表はプロビット変換を適用した例である。これは体重で組分けした１９４７年における Aberdeen の新生児の死亡率を示しており、考えられている体重の範囲では、体重と死亡率のプロビットは一次関係であることを示している。

一般に百分率の取扱いを簡単にし、死亡率と生れた時の体重との関係の場合のような曲線的関係を直線化するのにプロビット変換は役に立つであろう。しかしこのようにすれば、分散の一様性

の問題は無視される。したがつてプロビットを使うときには2.7節で説明した型の重み付き解析を行なわねばならない。

さらに面倒なことには解析の過程で重みを推定しなければならない。しかし全体の手順は計算を流す様に形式化されている。プロビット変換の応用に関する詳しい説明については，読者はD.J.Finneyの Probit Analysis(Cambridge, 1947)やC.W.Emmensの Principles of Biological Assay(London, 1948) を参照されたい。

10.3表　新生児の死亡率と生れた時の体重

生れた時の体重 lb	新生児の死亡率	死亡率のプロビット
2—	50%	5.00
3—	28%	4.42
4—	15%	3.96
5—	4%	3.25
6—7	1.2%	2.74

10.4　基礎関係の説明 Specification of underlying relationships

基礎関係に関する問題については前の数節で説明を行つた。2.6節では基本母数tを使つてこの問題を扱う方法と，その解法の難かしさのいくつかを示した。4.10と5.4では2つの測定値に含まれる原因不明の変動の割合が既知の場合の関係を推定する方法を示し，7.6節では母数tに関して組わけを行つた時の推定方法を示した。

これらの節はいづれも，2.6節で指摘したように，この関係の真の型は統計的なものであるにもかゝわらず，統計的関係よりむしろ数学的関係を推定することにのみ重点をおいた。したがつて場合によつては統計的関係を定めることが望ましいこともあろう。今度はその方法を考察してみよう。

それを定める1つの方法は観測値のいろいろな組合せが表われる。相対的頻度を与える多変量分布函数に関するものである。この様な函数を推定することができれば，必要と思われる全ての情報がそれによつて求められるが，一変量以上の分布函数の応用および推定はいづれも不可能な場合が多い。

最もよく知られている多変量分布函数は二変量正規分布である。

$$f(x,y)dxdy=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{(1-P^2)}}\exp\left[-\frac{1}{2(1-P^2)}\left\{\frac{(x-\mu_1)^2}{\sigma_1{}^2}+\frac{(y-\mu_2)^2}{\sigma_2{}^2}-\frac{2P(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}\right\}\right]$$

変量x,yはいづれも正規分布しており，ある変量の任意の値に対して，他の変量によつて与えられる値は正規分布をする。μ_1, μ_2は2つの変量の総平均でありσ_1,σ_2は標準偏差，Pは相関係数である。

円

$$\frac{(x-\mu_1)^2}{\sigma_1{}^2}\ \frac{(y-\mu_2)^2}{\sigma_2{}^2}-\frac{2P(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}=\text{一定}$$

はこの二変量分布の等頻度の線を表わしている。これらの楕円は共通の長軸と短軸をもち，或る条件においては前者は変量間の基礎関係を表わす線と考えられる。

二変量正規分布はいろいろな方法で生ずるものと考えられる。主としてyを決定するものとしてxを考えれば，分布は次の関係に起因している。

$x=t$，$y=at+b+\varepsilon$

又逆にxを生ずるものとしてyを考えれば

$x=at+b+\varepsilon$，$y=t$

これは2.6節の公式の極端な場合であり，この関係を表わす回帰線を生ずる。

又xとyは二組の因子により決定されるものと考えることもできる。この因子の一つは真の基礎関係に従つており，もう一つの因子は真の関係から離反する原因に従つている。したがつてこの関係を表わす方法は

$x = \cos\theta . t_1 + \sin\theta . t_2$

$y = \sin\theta . t_1 - \cos\theta . t_2$

である。

母数 t_1, t_2 の変動が分布を決定する。この場合 $var(t_1)$ は、$var(t_2)$ より大きいと仮定されている。

この場合には t_1 の変動は傾斜が $\tan\theta$ である直線基礎関係を説明するものと考えられる。t_2 の変動はこの直線から直角方向の偏差を作る。

この例では一次的基礎関係は二変量正規分布の主軸と同じである。これは 4.10、5.9 節の公式で R = 1 とおくことにより推定される。

しかしその傾斜は二変量の尺度によつて変る即ち両者が同一単位でない限り、その関係を表わしているという説明には疑問があるであろう。

推定を目的する時には、両変量の尺度が標準偏差を 1 にするように選ばれておれば、この難点は克服される。この表示方によれば次の方程式が得られる。

$x = S_1(t_1+t_2)/\sqrt{2}$

$y = S_2(t_1-t_2)/\sqrt{2}$

ただし、$e.var(t_1)=1+r_{xy}$, $e.var(t_2)=1-r_{xy}$, S_1, S_2 は標準偏差である。

多くの場合これはよりよき表示方であり又統計的関係の要約となる。しかし、使用される形は主としてその時の状態によつて変る。いづれも統計的関係を要約する方法を与える。したがつて相互の測定尺度の関係を考察しなければならない。この問題に関するこれ以上の説明は次節および 12.3 節でおこなうことにする。

10.5 基礎関係を推定するさらに詳しい方法
Further methods of estimating underlying relationships

8.7 節で説明した分散、共分散成分に関する方法は基礎関係を推定する問題にさらに詳しい方法を与えることに注目する。

観測値を研べたいと考えている特性に関して組分けすることができれば、分散、共分散の必要成分を計算することにより他の変量の効果は消去されるであろう。例えば家族内および家族間の変動成分を推定することにより、測定値間の基本的遺伝関係が推定できる。同様に 8.7 節の最初の例では植物間の分散、共分散成分の推定値は抽出変動を除いたときの植物間の基礎関係を生ずる。

この後者の関係は数学的と云うよりはむしろ統計的なものであり、推定された特性

$e.var(B)=3.51$, $e.cov(B,C)=-9.42$, $e.var(C)=33.83$, $r_{BC}=-0.864$ をもつている。これは二変量正規分布で表わされるが、さらに前節の母数 t を含んだ形式でもつと具合よく表わされる。

両測定値は同一単位であるから、オ 1 の母数を含んだ形式が使われる。これを求めるには

$$\frac{e.var(C)-e.var(B)}{2e.cov(B,C)} = -\frac{33.83-3.51}{2\times 9.42} = -1.6093$$

$$\tan\theta = -1.6093 - \sqrt{1+(1.6093)^2} = -3.504$$

$\sin\theta = 0.9616$, $\cos\theta = 0.2744$, $\sin 2\theta = 0.5278$

$$e.var(t_1)=\frac{1}{2}\left[33.83+3.51+\frac{2(9.42)}{0.5278}\right]=36.52$$

$$e.var(t_2)=\frac{1}{2}\left[33.83+3.51-\frac{2(9.42)}{0.5278}\right]=0.82$$

したがつて各植物の測定値 B, C の固有の値は次の統計的関係で表わされる。

$B = -0.2744 t_1 + 0.9616 t_2$

$C = 0.9616 t_1 + 0.2744 t_2$

たゞし、$e.var(t_1) = 36.52$　　$e.var(t_2) = 0.82$

このことは、この関係が傾き -3.504 の直線で表わされるが、完全な一次関係をこわしている線の周りに 0.82 の分散があることを示している。

植物の抽出分散を除いたときの、このものの理想的状態の推定値を表わしているこの関係は基本的なものである。しかしこれはただ一つの基礎関係でないことが分る。変動の遺伝的成分間の関係は別の基礎関係を与える。この場合には、解析を行なっているときに同じ雑種からえられた植物を一緒めにしなければならない。

一般に基礎関係を推定するのに分散、共分散の成分が用いられる時には調べたいと考えている型とその他のものとの差（例えば遺伝的差、植物間の差）が表われるように組分けを行うことが大切である。したがつて例えば上記の解析は、環境の影響が植物間の差に表われているので、遺伝的関係を推定するには不充分である。同じ理由から種族が同じ様に処理されていない限り種族内および種族間の解析を行うことはできない。

その結果環境による差はその比較に入つてくるはずはなかろう。

8.7節の才2の例では除きたいと考えている生理的変動を年令の差は反映しそうもないから上記の条件はかなり充されていると思われる。したがつて、組分けして計算した分散成分Yと回帰から計算した分散成分Tが、この場合に当てはめた関係を推定するのに用いられる。

この場合には、この関係を表わす才1の母数形式は

$m = 0.412t_1 + 0.911t_2$

$w = 0.911t_1 - 0.412t_2$

ただし、$e.var(t_1)=0.001646$ $e.var(t_2)=-0.0000002$ である。この小さな負の分散は8.7節で指摘した抽出変動のため生じたものである。

同様に、この関係を表わす才2の母数形式は

$m = 0.01340(t_1+t_2)$

$w = 0.528(t_1+t_2)$

たゞし、$e.var = 2.0004$　　$e.var(t_2) = -0.0004$

実用的にはいづれの場合でも、この関係は完全に直線で表わされ、wに関するmの傾きの推定値は0.528である。これはwに関するmの回帰から求めた値0.515とmに関するwの回帰からもめた値0.638と比較される。予期したようにmの測定誤差が大きいため、この傾きはwに関するmの回帰に一層近い。

10.6 推定値の比に関する問題

Problems invalving ratiss of estimates

より広般な問題では、2つの推定量の比を考察すること必要となつてくる。その典型的な問題を次に示す。

1. 与えられた従属変量の値に対応する独立変量の値が推定される。この場合、この値は

$$x - \bar{x} = \frac{y - \bar{y}}{b}$$

で与えられる。

特にy軸上の切片の推定値即ちプロビット解析で平均して50%の致死率を与える独立変量の値を求めたいことがある。

2. 二次曲線の最大値と最小値を推定する。

この場合推定された曲線は $y = a+b_1x+b_2x^2$ であり、その最大値、最小値は

$x = -b_1/2b_2$

の点で生ずるものとして推定される。

3. 2直線の交点を推定する。直線の方程式が、$y = a_1+b_1x$, $y = a_2+b_2x$であれば、交点は

$$x = -\frac{a_1-a_2}{b_1-b_2}$$

なる点で生ずる。

4. あてはめた2つの回帰線の相対的な傾きを推定する。$y = a_1+b_1x$, $y = a_2+b_2x$であれば、傾きの比は

$R = b_1/b_2$

である。

特に、傾きの比（shope ratios assays）を調べる場合にはこの※2種の薬品の相対力価は、それぞれの反応曲線の傾きの比から推定される。

※　その理由は普通 $a_1 = a_2$ であることは正しい。解析の時、曲数 $y = a_1+b_1x_1+b_2x_2$ はデータの全ての組にあてはまる。たゞし x_1, x_2 は2種の薬品の服用量を表わす。

この様な場合はいづれも、通常の解析によつて求めようとする比、$R = x/Y$ の推定値が得られ、同時に分子、Xと分母、Yの分散、共分散の推定値が得られる。これはRの標準誤差のおおよその値の推定又はRの正確な信頼限界を定めるために用いられるであろう。

Rの標準誤差の近似値は次式で与えられる。

$$R \quad \frac{\mathrm{var}(X)}{X^2}+\frac{\mathrm{var}(Y)}{Y^2}-\frac{2\,\mathrm{cou}(X,Y)}{XY}$$

しかし、Yの大になるにつれてその不正確さが大きければ、これは極めて正確なものとは云えない。又、Rは正規分布していないので、その使用はさらに制約をうける。

もつと正確な方法は $Z = X - \mu Y$ とおいて、Rの信頼限界を推定することである。たゞし

$$\mathrm{e.var}(Z) = \mathrm{e.var}(X)-2\mu\,\mathrm{e.cou}(X.Y)+\mu^2\mathrm{e.var}(Y)$$

しかしながら $Z^2/\mathrm{e.var}(Z)$ は分散比の分布に従うから、この量を例えば分散比の5%有意水準に等しくすることにより、Rの95%信頼限界を与える二次方程式が得られる。

この方法の例として、小児の平均代謝率が1000カロリーである体重を推定してみよう。

log（体重）に対するlog（代謝率）の回帰方程式を

$m = a + bw$ とすれば、次の推定値が得られる。

$a = 2.3131$※　　$b = 0.5154$

（※　10.2節の修正値を使つた）

$\mathrm{e.var}(a)=0.00041283$, $\mathrm{e.cou}(a.b)=-0.00028537$, $\mathrm{e.var}(b)=0.000$

0.00019631　自由度 521

したがつてmの値3（$= \log 1000$）に相当するwの推定値は

$$w = \frac{3.0000-2.3131}{0.5154} = \frac{0.6869}{0.5154} = 1.3328$$

この値の標準誤差の近似値は

$$1.3328\ \frac{0.00041285}{(0.6869)^2}+\frac{0.00019631}{(0.5154)^2}-\frac{2\times0.00028537}{(0.6869)(0.5154)} = \pm\,0.004831$$

であり、95%信頼限界は大体 $1.3328 \pm 1.96\times0.004831$

$= 1.3233$ と 1.3423 である。

この限界を推定するのに正確な方法を使えば

$Z = 0.6869 - 0.5154\,\mu$

$\mathrm{e.var}(Z) = 0.00041283 - 2\mu\,(0.00028537)+\mu^2\,0.00019631$

R^2 の95%限界は

$$\frac{Z^2}{\mathrm{e.var}(Z)} = 3.84$$

この後の公式は平方根が 1.3229, 1.3419 となる二次方程式を与える。

この値は正確な95%信頼限界を与え、体重21.00と21.97Kgに相当する。平均代謝率が1000カロリーとなる体重に対するこの限界は、代謝率が1000カロリーである個々の人の平均体重に対する限界22.49と23.35Kgとは違うものであることに注意しなければならない。この区別は2本の回帰線間の区別と同じである。

11.

11 時系列の解析　Time Series Analysis

11.1 主要な問題　The main problem

一定の時間順序で測定された一組の観測値は、相関および回帰分析の特別な問題を提示する。一般に時系列と呼ばれているこのような系列は通常、互に独立でなく、その結果一般の統計解

析で行なわれている独立性の仮定を満足していない観測値からなつている。極めて普通のことであるが，時系列は連続した観測値の従属性と類似性によつて特徴づけられている。したがつて各観測値はある程度前の観測値と関係があつて，場合によつてはこれから次の観測値が決定されることもある。例としてWolferの太陽黒点数※で示した（1770～1869年にわたる）太陽黒点の活動が11.1a表に示してある。この表を調べてみると太陽黒点の激しい活動があれば，それに引続いて激しい黒点の活動があるか或はその逆のはつきりした傾向がわかる。この結果この表に示してある観測値は独立ではない。

※ Yule.G.U.,Philos.Trans.A.226.297 より引用

11.1a表 Walferの太陽黒点数（1770－1869）

1770	101	1787	132	1804	48	1821	7
1771	82	1788	131	1805	42	1822	4
1772	66	1789	118	1806	28	1823	2
1773	35	1790	90	1807	10	1824	8
1774	31	1791	67	1808	8	1825	17
1775	7	1792	60	1809	2	1826	36
1776	20	1793	47	1810	0	1827	50
1777	92	1794	41	1811	1	1828	62
1778	154	1795	21	1812	5	1829	67
1779	126	1796	16	1813	12	1830	71
1780	85	1797	6	1814	14	1831	48
1781	68	1798	4	1815	35	1832	28
1782	38	1799	7	1816	46	1833	8
1783	23	1800	14	1817	41	1834	13
1784	10	1801	34	1818	30	1835	57
1785	24	1802	45	1819	24	1836	122
1786	83	1803	43	1820	16	1837	138

1838	103	1846	62	1854	21	1862	59
1839	86	1847	98	1855	7	1863	44
1840	63	1848	124	1856	4	1864	47
1841	37	1849	96	1857	23	1865	30
1842	24	1850	66	1858	55	1866	16
1843	11	1851	64	1859	94	1867	7
1844	15	1852	54	1860	96	1868	37
1845	40	1853	39	1861	77	1869	74

これらの観測値を例えば日間隔で記載すればもつと極端な例が得られる。このような場合にはある日の活動は前日の活動から殆んど完全に決定されるであろう。当然これらのものは独立な観測値とみなすことはできず，殆んど結果の単なる繰返に過ぎない。この繰返しを無視して普通の検定法を適用すれば連関の有意性に関して誤まつた概念が得られるであろう。

ある期間の連続した観測値に相関の傾向があれば，この系列には系列相関があるという。

したがつて一般には，解析や検定を行う時には，この系列相関を考慮する必要がある。

1.2 系列相関係数 Serial correlation coefficients

任意の系列の系列相関の程度を測りかつ有意性を検定するため，系列相関係数を計算することができる。したがつて，連続した項に相関があるかないかを検定したいときには1次系列相関係数を計算する。これは2組の観測値を結びつけることで行なわれる。即ち1番目の値から原系列の最後から2番目の値まで，2番目から最後の値までを系列とする。

例えば11.1a表のデータでは，1次系列相関係数0.817を推定するため，770～1868年の観測値から成る系列が1771～1869年の系列と結び付けられた。

同様に必要ならば2番目、3番目 .... の項間の相関を推定するため、2次、3次 .... 系列相関係数が推定される。これは、連続した項が従属的関係にある程度を示すのに役立つ。

系列相関係数の有意性は、近似的に過ぎないが普通の相関係数に対する有意性を検定する表（附録10表）を用いて簡単に検べられる。系列相関係数の計算に用いられた2組の値は実際には一組の値に過ぎないということを考えれば、別の表を用いる必要がある。しかし、係数の推定値に $1/N$ を、観測値の対数に2を加えて普通の相関係数に対する表を用いれば同じ結果が得られる。これば概略の修正に過ぎないが、実用的には充分正確である。

例えば11.1a表のデータについて計算した相関係数 r_{12} を検定するには0.817+0.010=0.827 なる値の有意性を検定しなければならない。

この値は99対の観測値に基ずいているからX表は101対のものを使用する。1%有意水準は0.256であり、したがつて観測された係数は極めて有意である。

観測値の系列の偏系列相関係数を計算することもできる。例えば中間項の効果を除いた2つ離れた項の偏相関係数を計算することができる。各項が将来起ると考えられることについての適切な情報を全て含んでいるか、或は、過去に得られた知識が将来の事態に関するさらに詳しい情報を前もつて与えるか否かをこれは示すであろう。

昆虫数に関する毎年の母集団推定値や宇宙塵の密度の毎日の推定値のような型の観測値では各観測値は最後に用いられる観測値だけと関係がある。このことが分つておれば、以前のデータでは、将来起ると考えられるさらに詳しい情報は得られない。この場合には観測値は一重Markoff過程の型をしていると云われる。

又、任意の一観測値が将来の実現しそうな観測値に関する情報を含んでいないしこの時には一個以上の観測値を用いねばならない。

このような場合には2，3項以上前の項が必要であるか否かによつて、2重、3重のMarksff過程をしているといわれる。このようなことは値が週期性をなしている時とか、観測値とその変化の割合が将来の観測値を示すのに役立つ場合には特によくあることである。

したがつて偏相関係数は、ある系列の観測値が属しているMarkoff型のオーダーを推定するのに用いられる。系列は独立の観測値から成立していないので、これは近似的な指標に過ぎない。

しかし実用的にはこの方法で得られた指標で充分である。

例として11.1a表に示してある100個の観測値の系列が、夫々1－98、2－99、3－100の項からなる98個の3系列に分割されゝば、この系列間の相関係数を計算すると次のようになる。

r_{12}=0.8218，r_{23}=0.8156，　r_{13}=0.4360

この最初の2つは1次系列相関係数の推定値を表わし、才3のものは2次系列相関係数の推定値である。

中間観測値が既知の場合2項離れている観測値間の相関を検定するため偏相関が計算される。これは

$$r_{13.2}\ \frac{r_{13}-r_{12}r_{23}}{(1-r_{12}^{2})(1-r_{23}^{2})} = -0.711$$

優略の検定に過ぎないが、観測値100に対する1%有意水準の値0.257と比較すれば、この値は有意であり、したがつて観測値は一重のMarkoff過程では表わせないことをはつきりと示している。しかし計算された偏相関係数が－0.200と0.200の間にあれば観測値を表わすため一重のMarkoff過程を使つてよい。

同様な方法で中間観測値が既知の場合3つ離れた観測値間の相関が $r_{14.23}$=0.021と計算された。この値は有意ではない。したがつて11.1a表のデータを表わすのに二重Markoff過程が

使用できる。さらにこの可能性は、高次の偏相関係数 $r_{15.234}$ $r_{16.2345}$を計算することにより検定される。

11.3 時系列間の相関 Correlation between time Series

相関係数は正規の方法で時系列間について計算される。時系列に系列相関があれば、普常の有意性の検定ではもはや当てはまらない。

狭い時間間隔で観測値をとつた場合の効果を考えればこのことははつきりする。

例えば年の代りに日又は時間間隔で観測値をとれば何度も同じ観測値が表われるであろう。

間隔を変えても相関係数は不変であるが観測数は明らかに間隔を縮めることにより際限なしに増加する。したがつて、普通の有意性の検定が用いられゝば、充分狭い時間間隔を選ぶことによりどのような値では有意にすることができるであろう。

この障害を除くには相関係数の計算に入つてくる有効な独立観測値の数を推定する必要がある。これを行うには次に説明するBartlett※の公式を使えばよい。

(※ Bartlett,M.S. J.R.statist.Soc.1935.98)

N項からなる2つの系列の系列相関係数が $r_1, r_2, r_3 \ldots$, および $r'_1, r'_2, r'_3 \ldots$ であれば系列間の相関を検定するのに使われる有効な独立観測値数は略 $N/(1+2r_1r'_1+2r_2r'_2+\ldots)$☆である。

☆注 Bartlettの公式は系列間の相関の検定に用いられる自由度即ちN－2と関係がある。近似公式の次数に関しては自由度即ち観測数に補正をほどこすか否かは問題でない。

この公式は各系列の観測数が大きい場合の系列相関に対する近似的補正をあたえる。

どちらかの系列の系列相関係数が小さいか又は0であれば観測数を修正する必要はないことに注意すべきである。その理由はどちらかの系列が無作為又は無作為に近ければその系列の各観測値

は相関係数の値について全く新しい情報を与えているからである。

この公式の使用法を示すため11.3a表のデータと一緒に11.11a表のデータを考察してみよう。この後の表はオーロラの活動、N(S.Tromholt,Kristianiaの行つたノールウエの観測値の要覧からとつたもの)と地震活動Eの指数を示している。(Milneの行つたII、III型の強震の記録からとつたもの)

11.3a表 オーロラの活動Nと地震活動Eの指数

	N	E		N	E
1770	155	66	1791	0	92
1771	113	62	1792	12	70
1772	3	66	1793	0	46
1773	10	197	1794	37	96
1774	0	63	1795	14	78
1775	0	0	1796	11	110
1776	12	121	1797	28	79
1777	86	0	1798	19	85
1778	102	113	1799	30	113
1779	20	27	1800	11	59
1780	98	107	1801	26	86
1781	116	50	1802	0	199
1782	87	122	1803	29	53
1783	131	127	1804	47	81
1784	168	152	1805	36	81
1785	173	216	1806	35	156
1786	238	171	1807	17	27
1787	146	70	1808	0	81
1788	0	141	1809	3	107
1789	0	69	1810	6	152
1790	0	160	1811	18	99

1812	15	177	1816	64	22
1813	0	48	1817	126	45
1814	3	70	1818	38	102
1815	9	158	1819	33	111
	N	E		N	E
1820	71	90	1845	20	86
1821	24	86	1846	51	127
1822	20	119	1847	72	201
1823	22	82	1848	118	76
1824	13	79	1849	146	64
1825	35	111	1850	101	31
1826	84	60	1851	61	138
1827	119	118	1852	87	163
1828	86	206	1853	53	98
1829	71	122	1854	69	70
1830	115	134	1855	46	155
1831	91	131	1856	47	97
1832	43	84	1857	35	82
1833	67	100	1858	74	90
1834	60	99	1859	104	122
1835	49	99	1860	77	70
1836	100	69	1861	106	96
1837	150	67	1862	113	111
1838	178	26	1863	103	42
1839	187	106	1864	68	97
1840	76	108	1865	67	91
1841	75	155	1866	82	64
1842	100	40	1867	89	81
1843	68	75	1868	102	162
1844	93	99	1869	110	137

この三系列の特性を調べ、前節で説明した方法で要約する。これは11.36表に示してある。

11.36表　系列相関係数と偏系列相関係数

		太陽黒点の活動S	オーロラの活動N	地震活動E
系列相関係数	r_1	0.817	0.715	−0.015
	r_2	0.456	0.427	−0.025
	r_3	0.071	0.298	0.006
偏系列相関係数	$r_{13.2}$	−0.711	−0.164	−0.025
	$r_{14.23}$	0.021	0.081	−0.005

この表の値を調べてみると太陽黒点の活動は二重 Markoff 過程で表わされるがオーロラの活動は一重 Markoff 過程で表わされることが分る。

これに反して、地震活動に関する連続した観測値は事実上独立であることを表わしており、これらが無作為系列としてとられたものであることを示している。

三系列間の相関は

$r_{SN} = 0.430$，$r_{SE} = -0.063$，$r_{NE} = 0.060$

地震活動は明らかに無作為であるから、後の2つ相関係数が普通の相関係数の検定法（X表）を用いて検定される。

いづれも有意でなかつた。

この最初の相関を検定するには両系列の系列相関を即ち系列の途中に高い値と低い値とが生ずるという事に考慮する必要がある。これを行うには独立観測値に等しい観測数を計算すればよい。即ち、

$1+2r_1r_1'+2r_2r_2'+\cdots = 1+2(0.817)(0.715)+2(0.456)(0.4$

$=(0.427)+2(0.071)(0.298)+$

$=2.58+\cdots\cdots$

多数の系列相関係数を計算しなければ、ある正確さをもつてこの量を評価することは不可能である。しかしこの系列の3～4観

測値が一つの独立の観測値に相当する、したがってこの系列の１００観測値が無作為系列の２５と３３観測値の間に相当するということができる。したがってＸ表によれば、この観測値は５％水準では有意であるが１％水準では有意でないことがわかる。

この方法で検定すれば系列間の相関の有意性の指標がえられるが、系列に系列相関係数があれば、データを完全に利用することは普通出来ない。偏系列相関係数を使えばより効果のある方法が得られる。

１１.４　偏相関係数を用いる時系列間の相関の検定
Testing the correlation between time series using partial correlation coefficients

無作為系列相関のある系列、例えば前節の太陽黒点と地震系列間の相関を検定したいとする。このために普通の相関係数が用いられるが、偏相関係数の使用も考えられる。

太陽黒点の活動は２つの部分を含んでいると考えられる。即ち以前の太陽黒点の活動に関する知識から予測された部分と無作為の部分とである。しかし前者が地震活動と相関しておれば、実際にはそうでないが、これは地震活動と相関しておれば、実際にはそうでないが、これは地震活動と系列相関のあることを示している。したがって二系列間の相関を検定するときに使用しなければならないのは太陽黒点の活動の無作為部分である。実際、偏相関係数が計算されゝば、この様にすることにより以前の太陽黒点の活動に関する効果が除かれる。この系数は二系列の同時点の無作為部分に相関があるか否を検定する。

しかし二系列の同時的でない無作為部分の相関の有無を示すことはできない。これを検定するには、時差偏相関係数が計算されねばならない。例えば太陽黒点の活動の無作為変動が、一年後の地震活動に影響しておれば、これは同じ年の活動間の偏相関係数を用いたのでは検出できない。このためには予めあらゆる年の太

陽黒点の活動の効果を除いて、地震活動と前年の太陽黒点の活動とを結びつける必要があろう。

この手法を示すため、上記の系列間の偏相関の計算を考察してみよう。これらの系列は原系列の１－９８、２－９９、３－100なる項からなる三つの副系列に分割され、これを夫々添字１、２３で表わす。したがって　S_2　は１７７１～１８６８年の太陽黒点の活動を表わし、$r(S_3E_3)$ は１７７２～１８６９年の黒点活動と地震活動の相関を表わし、$r(S_3E_3/S_2)$は前年の黒点活動を考慮したときの地震活動と黒点活動の偏相関係数を表わす。

この記号を用いて次のように相関係数、偏相関係数が計算される。

$r(S_3E_3) = -0.046$

$r(S_3E_3/S_2) = 0.087$

$r(S_3E_3/S_2S_1) = 0.134$

これはいづれも黒点活動と地震活動との相関を示すが、２番目のものでは前年の黒点活動の効果が除かれており、最後のものでは前２年の黒点活動の効果が除かれている。

これらの値の有意性を検定するときには、普通の偏相関係数の検定が用いられる。例えばこの最後の係数は９８－２＝９６観測値に基ずくものとして検定される。Ｘ表を用いればこれらの値はいづれも有意でないことが分る。明らかに黒点活動と地震活動との間には直接の相関はない。しかしこの最後の係数が有意であれば地震活動と黒点活動の予測できない同時的変化とは相関があると結論できる。

相関に時差（ずれ）があるかどうか検べるには時差偏相関係数を計算しなければならない。したがって例えば$r(S_3E_2/S_2S_1)$は地震活動が黒点活動の予想出来ない変化の前に起るか否かを検定し、$r(S_2E_3/S_1)$は黒点活動の変化が地震活動の前に起るか否かを検定するものである。

系列相関のある二系列を相関させる場合にも同様な方法が用い

られる。この場合には偏相関係数はどちらか一方の系列又は双方の系列の既往の効果を除くために用いられる。両系列の既往の効果が除かれゝば二系列の予想できない変動に相関があるか否かが相関係数によつて検定できる。例えば黒点活動とオーロラ活動の系列では、相関係数および偏相関係数は

$r(S_3N_3) = 0.409$　　$r(S_3N_3/N_2) = 0.176$

$r(S_3N_3/S_2) = 0.305$　　$r(S_3N_3/S_2N_2) = 0.272$

$r(S_3N_3/S_2S_1) = 0.328$　　$r(S_3N_3/S_2S_1N_2) = 0.296$

既往の活動の効果が除かれゝば相関の大きさは小さくなるのが普通であるが、適当な既往の効果が除かれると0.296なる値が得られる。この値は普通の偏相関係数の検定を用いて検定されるが、どちらの系列にも系列相関があるため、この検定は近似的なものに過ぎない。

X表から、98－3＝95観測数の1%有意水準は0.264である。ある年における原因不明の黒点変動とオーロラ活動との相関は有意である。

上に述べたように、このことはある系列が他の系列に時差の効果をおよぼしていることを否定していない。したがつて$r(S_3N_2) = 0.417$であるからオーロラ活動は次の年の黒点活動と連関のあることが分る。したがつて、この相関が系列の系列相関によるもの、例えばある年の黒点活動が順次その年のオーロラ活動と、連関のある前年の黒点活動と連関しているという事実のためであるか否かを決定する必要がある。

このためには$r(S_3N_2/S_2S_1N_1) = 0.118$を計算する必要がある。これは有意でないから、ある年のオーロラ活動の原因不明の変動と翌年の黒点活動とには相関がないことを示している。

逆に$r(S_2N_3/S_1N_2) = 0.077$は別の面についても時差相関がない、例えばある年の黒点活動の変動は翌年のオーロラ活動の変動と連関していないことを示している。

それ故に黒点活動とオーロラ活動との連関を検定して、次のような結論に達した。

1. いづれの系列にも系列相関がある、即ち、どの項も、一部分は以前の項に関する知識から予測することができるであろう。又黒点活動は二重Marksff過程で表わすことができ、オーロラ活動は一重Marksff過程で表わされる即ち夫々2つおよび1つ前の項は実現しそうな値に関する情報の大部分を含んでいる。
2. ある年の黒点活動とオーロラ活動との予測できない同時的活動は互に相関があるがこの相関はその期間の前又は後に拡張されない。即ち予測できない変動の間には時差相関は存在しない

11.5 時系列間の回帰分析

Regression analysis between time series

通常時系列間の相関の検定は解析の才1段階に過ぎない。連関が存在すれば、回帰分析で連関の形を推定する必要がある。これは4～6章で述べた方法で行なわれるが、系列に系列相関があればこの章の方法では、最良かつ最も正確な推定方法はえられず、この方法によつて得られる標準誤差は不正確なものである。

したがつて回帰分析では系列相関の存在が推定を無効にし、特殊の解析をする必要があるか否かを考察しなければならない。

前節で指摘したようにどちらかの系列に系列相関がなければ普通の相関係数の検定が有効である。回帰分析についても同じことが云える。即ち従属変量か独立変量のいづれかに系列相関がなければ普通の回帰分析法が使用できる。したがつて例えば11.3節の地震系列は普通の回帰分析法で独立変量としても、従属変量としても使用できる。

これだけが普通の回帰分析が有効となる条件ではない。別の条件は回帰線からの偏差は系列として独立でなければならないということである。もしそうであれば、普通の解析が有効となる。この結果、系列相関のある変量間の回帰分析が有効かどうか検定するには回帰からの偏差が独立であるか或は系列相関があるかを検

定する必要がある。

あてはめた回帰線からの偏差の独立性を検定するために系列相関係数が使用される。しかしこの係数の有意性を検定する場合には、この回帰関係の当はめを考慮する必要がある。残念ながら、これは簡単にはできない。Durbin,Watson※は有意性の検定は、あてはめられた回帰線の独立変量の値によつて変り、したがつて実際には観測値の組毎に有意性の検定をしなければならないことを示した。この障害を除くためDurbin,Watsonはあらゆる条件における有意水準の限界を計算した。

したがつてこの上限を越える値は明らかに有意であり、下限以下の値は明らかに有意でない。この限界内にある値の有意性については若干疑問がある。

11.5a表は普通の相関係数の検定（X表）を用いて系列相関係数を検定する時の観測数の変化によるDurbin,Watsonの値を表わしている。

この表を使用するには2つの値が必要である。即ち回帰に含まれる独立変量の数と有意性を検定する水準である。

11.5a表　回帰からの偏差の系列相関の検定に用いられる観測値の近似的修正

独立変量の数	有意水準 P＝0.05	有意水準 P＝0.02
1	1～20	1～16
2	-5～35	-5～30
3	-10～60	-10～50
4	-15～100	-15～75

この表の修正値は多数の観測値が得られる場合にのみ使用すべきものである。（あてはめられる各独立変量に対して最小10個の観測数例えば3独立変量が当てはめられれば最小30個の観測数が必要である。）境界上にあるときとかもつと正確な検定を行うときにはDurbin,Watsonの論文を参照する必要がある。

例として2つの独立変量をあてはめた50個の観測値の偏差の系列相関を検定したいとする。11.5a表によれば5%有意水準は50－5＝45と50＋35＝85観測数に対する相関係数の5%有意水準の間（即ち0.295と0.213との間）にある。この前者の値を越える係数は5%水準で有意であり、後者の値以下のものは明らかに有意でない。この間にある値はP≈0.05と云えるが、完全な時間のかゝる解析をしなければ5%有意水準に達しているか否か云えない。

しかし通常、偏差の独立性の仮定が正しいかどうか定めれば充分である。それ故、一般的にはこの概略の検定で充分である。

誤差に系列相関があれば、これを認めた解析を行なわねばならない。このデータについては現在著者の知るところでは最適の形は分らないが二つの解決法が暗示されている。

1. 従属変数自身の前の値に対する従属変数の系列の回帰からの偏差を計算する。独立変換とその前の値に対する回帰分析にこの偏差を使用する。
2. 独立変数に対する従属変数の回帰と独立変数の前の値と独立変数に対する従属変数の回帰とを計算する。

この方法はいづれも必然的に最良の近似的解析法であり、その選択は研究者の要求にまつべきである。前者の方法は従属変数の無作変動を予測するのに用いられる（独立変数の範囲を推定する。後者の方法は独立変数と以前の観測値から従属変数を予測する最良の方程式を推定する。

時系列の回帰分析を示すため黒点活動Sに対するオーロラ活動Nの回帰を計算する。

直接計算した時には回帰係数の推定値は0.5938である。しかし回帰N－0.5938Sからの偏差の系列相関は$r_1 = 0.667$，$r_2 = 0.345$である。11.5a表を用いれば5%有意水準で前者は100－119観測値に基ずくものとして検定し、後者は99－

注　Durbin,J.,and Watson,GS.Biometrika.38(1951)159.

118観測値に基ずくものとして検定しなければならないことが分る。したがつて5%有意水準は0.197と0.179の間にある。この二つの一次の系列相関は明らかに有意である。しかし係数を推定するため普通の方法で標準誤差を決定することはできず、有効な推定値をうるにはさらに解析を行なわねばならない。

上記の才1の方法を用いれば、黒点系列は一重 Markoff 過程で表わされることが11.5b表は示している。一次の系列相関係数は0.715であるから偏差 $n_i = N_i - 0.715N_{i-1}$ は系列的に独立であり、したがつて回帰分析に使用できる。同様に $Si = Si - 1.41Si_{-1} + 0.71Si_{-2}$ は黒点系列の偏差を表わしており、系列的に独立である。

したがつて系列 S_3, S_2, S_1 に対する系列 n_3 について回帰分析が行なわれる。この様にすれば回帰方程式の推定式は

$$n_3 = 0.685S_3 - 0.834S_2 + 0.356S_1 + 8.9$$
$$= 0.685(S_3 - 1.22S_2 + 0.49S_1) + 8.9$$

この方程式の形は、n_3がS_3即ち黒点系列の無作為変動と関係の深いことを示している。これは11.5b表に示してある分散分析の形で検定される。

11.5b表　n_3の分散分析

	自由度	平方和	分散の推定値
Sに導入される変動S_3に対する回帰	1	10467	10467
Sの以前の変動に対する回帰	2	1018	509
Sに対する回帰$S_1 \cdot_2 S_3$	3	11485	3828
n_3 の残差分散	94	113501	1207
n_3 の全分散	97	124986	

オーロラ活動の無作為変動は（現在および過去の値を用いた）

黒点活動の指標と有意な相関があり、この相関は殆んど全てが黒点活動の同時的無作為変動に帰因することがわかる。この二つの無作為要素間の回帰関係は

$$n_3 = 0.686S_3 + 3.4$$

即ちN_3は

$$N_3 = 0.715N_2 + 0.686S_3 - 0.967S_2 + 0.487S_1 + 3.4$$

を使つて予測される。

上記の才2の方法を用いれば、重回帰分析が行なわれる。この場合にはN_2, S_3, S_2, S_1に対するN_3の従属関係が計算でき、次の関係が推定される。

$$N_3 = 0.667N_2 + 0.710S_3 - 0.837S_2 + 0.548S_1 + 10.5$$

残差平方和は115035である。

通常この型の回帰線の妥当性は残差を考慮に入れて吟味しなければならない。しかし、この場合には残差を考慮する余地がない程、才1の方法で得られた回帰とよく一致している。

両方法を一諸に考えれば二つの無作為要素に関する方程式は二系列間の連関の形を充分に要約していることが分る。したがつて方程式$n_3 = 0.686S_3 + 3.4$はこの系列間の連関を合理的に要約している。

しかし、場合によつては直接系列間の関係を推定するのが本質であると論ぜられることもある。したがつて以前の観測値が未知であれば他の変量の同時点の値だけを用いてある変量を予測したい場合がある。

この目的で直接計算した回帰係数が用いられるが、その分散は次の因子で修正する必要がある。

$$F = 1 + 2r_1r_1' + 2r_2r_2' + \cdots\cdots$$

$r_1r_2\cdots$および$r_1'r_2'\cdots$は二系列の系列相関である。この因子は系列相関のある二系列間の相関係数を吟味する時に、自由度を修正するために用いたものと同じである。

例えばSに関するNの回帰係数は0.594であり、普通の方法で

計算したその標準誤差は，±0.126であつた。しかし11.3節の結果からFは3と4の間に在るから，標準誤差は3と4の間にある因子で修正される。このようにすれば±0.218と±0.252の間にある修正された標準誤差がえられる。これはおゝよその範囲であるが，係数の推定値に含まれる誤差の大きさを示すのに役立つ。

あてはめた回帰線からの偏差が系列的に独立であるように独立変量と従属変量とが同時的かつ同じように変換できれば，係数およびその誤差の最良の推定値が得られる。このようなことが何時でも出来るとは限らず，残差の検討を行う試行錯誤法が用いられるが，これはせいぜい概略の近似的方法として考えられているに過ぎない。しかし回帰係数およびその誤差を推定するには別の根拠がある。

例えばあてはめた回帰線 $R = N - 0.594S$ の残差の系列的従属関係を推定するため回帰が計算されたならば $R_{i+2} = 0.8R_{i+1} - 0.2R_i$ なることが分る。

したがつて変換した系列 $N_3 - 0.8N_2 + 0.2N_1$ と $S_3 - 0.8S_2 + 0.2S_1$ が次の回帰の計算に用いられる。この場合の回帰係数の推定値は0.504±0.175であり，残差の系列相関は $r_1 = -0.006$，$r_2 = -0.033$，$r_3 = 0.202$，$r_4 = -0.082$ である。これらの値は推定値を合理的なものとするには低すぎる値である。しかし，残差Ri間の推定された関係に含まれそうな誤差の点から云えば，この係数につけられた標準誤差の推定値は真の標準誤差に対する下限であると考えねばならない。

11.6 既知の形の傾向の除去
Elimination of trend of known type

時系列は，観測値の平均水準が順序良く変化しているために起る一般的傾向の影響をうる場合が多い。例えば，12a図に示してあるUnited Kingdomの電気生産量の観測値は大略の一般的

傾向を示している。この傾向には幾分小さい変動があるかも知れないが，この傾向が観測値に差の生ずる主な原因となつている。11.6a表のTorontoにおける年平均気温の観測値は，平均気温が増加してゆく，緩慢であるがはつきりした傾向を示している。

11.6a表　Ontario州Torontoにおける，年平均気温Tと年降水量R

	T	R		T	R
1851	43.9	30.7	1876	43.9	32.4
52	42.9	40.9	77	46.1	25.6
53	43.7	28.9	78	47.0	48.5
54	44.1	32.7	79	43.8	29.4
55	43.4	41.5	80	45.3	35.3
56	41.3	28.1	81	46.2	26.9
57	42.0	40.6	82	45.6	24.8
58	44.7	32.6	83	42.2	34.1
59	44.2	39.8	84	44.1	28.6
60	43.7	28.0	85	41.7	32.9
61	44.2	34.5	86	43.9	35.1
62	44.2	34.1	87	44.4	25.8
63	44.7	32.8	88	43.0	26.3
64	45.0	36.9	89	45.8	31.2
65	45.5	32.9	90	45.4	37.8
66	44.0	39.4	91	46.0	31.5
67	44.4	30.1	92	44.7	29.5
68	43.4	34.3	93	43.8	39.7
69	43.7	39.6	94	46.9	29.6
70	46.1	46.2	95	44.4	28.0
71	44.3	32.7	96	45.4	29.1
72	43.5	25.3	97	46.0	32.5
73	42.2	31.6	98	47.2	30.9

1874	43.8	24.4	1899	45.8	29.0
75	40.5	29.7	1900	47.1	29.6.
平均	43.7	33.9		45.0.	31.4

	T	R		T	R
1901	45.7	32.3	1926	43.9	37.9
02	45.7	31.0	27	46.5	30.7
03	45.9	30.6	28	46.5	35.4
04	42.4	35.7	29	45.8	37.0
05	44.8	31.2	30	47.3	25.8
06	46.3	31.0	31	49.7	27.3
07	44.2	31.7	32	48.0	37.0
08	46.3	29.5	33	48.1	23.8
09	45.8	32.9	34	46.2	24.8
10	45.7	33.6	35	46.6	26.7
11	46.7	29.2	36	46.9	28.0
12	43.5	32.5	37	47.9	32.9
13	47.2	28.8	38	48.3	25.6
14	45.3	27.2	39	46.7	27.7
15	46.2	34.7	40	45.4	35.4
16	45.7	32.0	41	48.4	25.8
17	42.5	34.4	42	47.2	38.1
18	45.0	34.4	43	45.5	32.6
19	47.5	29.8	44	47.4	30.7
20	45.1	29.9	45	46.3	40.6
21	49.1	27.3	46	48.1	29.4
22	47.3	29.1	47	47.3	33.4
23	45.4	33.5	48	47.8	28.4
24	44.3	33.9	49	49.7	24.8
25	45.8	30.5	50	46.7	33.7
平均	45.6	31.5		47.1	30.9

時系列に傾向があれば、一般に解析に着手する前又はその途上でこれを除去しなければならない。これをしないと相関のある系列は同じような傾向をもつことになるため、誤まつた連関が決められてしまう。例えばU.Sのアルミニューム生産量と英国教会の会員数は過去50年を通じて一般に増加している。この二系列の観測値について相関係数を計算すれば、疑もなく極めて有意となるが、これは二系列に同じような傾向のあることを示すに過ぎない。この両系列の傾向を除き、傾向の周りの変動に相関があるか否かを調べることによつて初めて意味のある相関を確めることができるのである。

この点で傾向と系列相関との区別に注目することが大切である。傾向と系列相関はいづれも連続した観測値の類似性によつて現われるものであるが、傾向は平均水準の一般的変化を表わし、解析の際その影響を除去する必要があるが、系列相関は連続した平均水準の周りの傾向を表わし、その影響を解析の際修正するか、考慮に入れることが望ましい。

どのような観測値にも傾向と系列相関とがある。例えば長年月にわたる太陽黒点とオーロラの観測数は系列相関と同様に観測の集約度の変化のため傾向がある。このような場合の主要な問題の一つは傾向と系列相関とを区別することである。起ると考えられる傾向の形が分つておれば、前節の方法を用いて解析の行なわれることが多い。例で、この方法を示そう。

11.6a表の二組の観測値を相関させ、年降水量Rに対する平均気温Tの回帰を推定したいとする。平均気温の値を調べてみると、緩漫に一様に増加しており（気象学者や地質学者の観点からの）この様な比較的短期間についてはこの系列の傾向は直線で表わせると仮定してさしつかえないことが分る。同様にRの傾向は直線で表わせる。したがつて才3の一次変数Lが傾向の形を解析する際入つてくるものと考えられる。これらの変数は任意の一次集合の値であり、便宜上-99，-97，-95，……:-99即ち一次直交

多項式の値をとる。

気温と降水量の系列の性格を調べるため、TおよびRの値と一次傾向線の周りの偏差系列相関係数を計算する。実際には後者の系列相関はLを除いた時の偏相関係数から計算される。この値が11.6b表に示してある。

11.6b表　TとRの系列相関係数

	T	R	Lを除いた T	R
r_1	0.519	−0.071	0.119	−0.136
r_2	0.457	0.109	0.030	0.065
r_3	0.438	0.131	0.009	0.087

Lを除いた係数は11.5a表の近似的修正値を用いて検定しなければならない。この場合この修正により、5%水準で、X表で使用すべき観測数は100と119の間にある。この結果Lを除いた5%有意水準は0.180と0.197の間にある。係数はいづれもこの下限の値に達していない即ち系列には明らかに系列相関はなく、したがつて、附加変数Lを用いて普通の相関分析又は回帰分析が行なわれる。この場合相関係数は夫々 $r_{TR}=-0.251$, $r_{TR\cdot L}=-0.133$ である。これは傾向を除く前には気温と降水量の相関は有意であるが、除いてしまうと有意でなくなることを示している。降水量に関する気温の回帰係数即ち傾向を除く前は -0.097 ± 0.038、除いてしまうと -0.039 ± 0.030 を用いても同じ結論に達するであろう。

傾向線からの偏差に系列相関があれば、前節で説明した形のTR,Lに関する解析を行う必要がある。Lの効果を除いてから有意性の検定を行なわねばならない。

傾向の形が観測値の組平均で説明される場合がある。この様なことは特に解析を行う前に、周期変動又は季節変動が除去されねばならない時におこる。例えば、126図に示してある10時間の漁労当りの月平均漁穫高（1933－1937）が11.6c表に示

してある。

11.6c表　10時間の漁労当りの漁穫高の対数

	1933	1934	1935	1936	1937	平均
1月	0.79	0.58	0.36	0.28	0.67	0.536
2月	0.69	0.40	0.23	0.18	0.53	0.406
3月	0.52	0.36	0.20	−0.15	0.08	0.190
4月	0.46	0.04	0.00	−0.22	0.26	0.108
5月	0.68	0.41	0.43	0.32	0.60	0.488
6月	0.80	0.56	0.49	0.32	0.73	0.580
7月	0.80	0.62	0.53	0.40	0.85	0.640
8月	0.84	0.67	0.60	0.54	0.88	0.706
9月	0.92	0.75	0.63	0.65	0.94	0.778
10月	0.97	0.72	0.68	0.81	0.95	0.826
11月	0.96	0.76	0.66	0.86	0.98	0.844
12月	0.85	0.60	0.48	0.79	0.98	0.740
平均	0.773	0.534	0.441	0.398	0.704	0.570

この数値を解析するとき、偽相関が入らないように季節変動と年による傾向を除きたい。

これを行うには8.2節に示した分散共分散分析法を用いればよいが、これはこの方法で系列的相関のない残差がえられるだろうかという質問と関係がある。そうでなければ標準誤差さらに有意性の検定の妥当性が疑われる。この方法において年および月別計の差を除くことにより相関のない残差が得られるかどうかを検べるには、残差を計算する必要がある。これを求めるには各値から月および年平均を引き、それに総平均を加える。例えば、11.6c表で1933年1月の残差は 0.79−0.536−0.773+0.570=0.051 であり1933年2月の残差は 0.69−0.406−0.773+0.570=0.081 である。この様にして求めた残差の値は11.6d表に示してある。残差の間には明らかに系列相関のあることがわかる。その一次系

列相関係数は0.480である。

この形の残差から推定した系列相関係数の有意性を検定するには、組平均が除かれているということを認めねばならない。

充分な観測値が得られれば、これを行うには相関の検定に用いられた観測数から除去された比較の自由度を差し引けばよい。この例では自由度11および4が夫々月および年の比較を表わしている。したがつて一次系列相関係数は59+2-11-4=46対の観測値に基ずく相関係数として検定されるべきである。これは1%有意水準で有意である。

11.6d表　対数で表わした漁獲高の残差

	1933	1934	1935	1936	1937
1月	+0.051	+0.080	-0.047	-0.084	0.000
2月	+0.081	+0.030	-0.047	-0.054	-0.010
3月	+0.127	+0.146	+0.139	-0.168	-0.244
4月	+0.149	-0.032	+0.021	-0.156	+0.018
5月	-0.011	-0.042	+0.071	+0.004	-0.022
6月	+0.017	+0.016	+0.039	-0.088	+0.016
7月	-0.043	+0.016	+0.019	-0.068	+0.076
8月	-0.069	0.000	+0.023	+0.006	+0.040
9月	-0.061	+0.008	-0.019	+0.044	+0.028
10月	-0.059	-0.070	-0.017	+0.156	-0.010
11月	-0.087	-0.048	-0.055	+0.188	+0.002
12月	-0.093	-0.104	-0.131	+0.222	+0.106

各平均値の計算に4個以上の観測値が用いられゝばこの近似法はよく合う。観測値が少なければAndersonとAnderson※およびDurbinとWatson*のもつと正確な検定を用うべきである。

注※ Anderson,R.L.,and Anderson,T.W.Ann.Math.Stat.21(1950)59.
注* Durbin,J.and Watson,G.S.Bismetrika38(1951)159.

この場合のように残差に系列相関のあることが示されゝば、残差の系列的独立性を検べるため、さらに相関分析又は回帰分析を行はねばならず、必要あれば前節で説明した方法で偏回帰又は偏相関を用いねばならない。

11.7　一般的傾向の除去 Elimination of general trend

時系列に含まれる傾向の形が未知であれば、適当な表示法を推定する必要があるという問題がさらに起る。観測値を組分けして組間の差を除くことによつて、これが行なわれる場合もある。前節で説明した型の解析が使用できる。しかし14図の例の様に系列に強い傾向があれば、このようなデータの組分けは非現実的でありかつ不充分なものであろう。

現在のところ、傾向を除去する方法の問題に対する一般的解答は分つていない。この目的で移動平均差式がよく用いられるが、これは明らかに結論に影響を及ぼし易い。※多くの場合移動平均比を用いた方が良いが、その効果については余り知られていない。

一般に、傾向を除去する際の主要な問題は傾向と系列相関とを区別うることである。両者とも短い系列では同じ様子を示している。しかもなおこの問題に対する解答はえられない。解答の得られるような時期が来るまで次のようなやり方が提案されている。

多項式のあてはめにより系列から傾向を除くため解析を行う。系列内の何番前の項を使用すべきかを推定するため、回帰分析では傾向を除いた残差を用いる。この項がみつかれば、それを修正する必要の有無を決めるため、あてはめた傾向を再び調べてみる。修正の必要があれば、修正した傾向を用いて解析を始めからやりなおす。

この方法は傾向の形の逐次近似値を求めるのに役立つが、その応用に当つて、有意な傾向又は系列相関を決めるには主として個人的判断によらねばならない。次の例はこの方法をはつきりさせ

注※ Spencer-Smith,J.L.J.R.statist.Soc.,B9(1947)104

るであろう。

11.7a表はU.Sにおける肥料消費額Cと農家収入Iとを示す。これは14図にプロットしてあり、2組の観測値間には明らかに連関のあることを示している。

11.7a　U.Sにおける肥料消費額（1000トン）と農家粗収入の指数

年	消費額	収入	年	消費額	収入	年	消費額	収入
1911	5610	32.2	1924	6826	47.2	1936	6931	48.7
1912	5767	34.7	1925	7534	49.3	1937	8226	48.8
1913	6337	36.1	1926	7329	48.9	1938	7548	46.1
1914	7100	34.7	1927	6844	51.0	1939	7707	47.4
1915	5324	26.3	1928	7986	52.8	1940	8249	49.2
1916	5125	22.0	1929	8012	54.4	1941	9183	58.3
1917	5926	29.5	1930	8222	47.5	1942	9949	69.2
1918	6467	33.0	1931	6354	39.1	1943	11463	87.2
1919	6626	41.8	1932	4385	32.1	1944	12055	90.9
1920	7177	35.8	1933	4908	35.6	1945	13202	95.9
1921	4863	33.7	1934	5583	39.5	1946	14874	103.9
1922	5671	40.2	1935	6276	42.3	1947	15039	100.0
1923	6445	43.8						

1912～1947年までの収入に対する肥料消費額の関係を調べる時には傾向は直交多項式を用いてあてはめられるであろう。残差間の相関を推定するには、8.2節で説明した形の共分散分析が用いられる。これは11.7b表で行つている。

この解析を調べてみると、系列相関がないと仮定すればこの傾向を表わすのに5次の多項式が必要であることが分る。Iに関するCの回帰係数の推定値は108999.0／861.72＝126.5であり、この回帰は残差の有意な平方和13,787,288を説明している。

11.7b表　CとIの分散、共分散分析

	自由度	Cの平方和	CとIの共分散	Iの平方和
一次	1	126,755,029	10864649	93126.4
二次	1	59,597,872	36541148	22404.8
三次	1	27,294,581	21555486	17022.3
四次	1	5,551,774	56150.1	589.89
五次	1	5,485,571	51813.0	770.20
傾向	5	220,482,627	17555,391.4	144154.44
残差	30	24,082,861	1089990	861.72
計	35	244,565,488	18643,904	152771.6

今やIの効果を除いたCの残差を調べることができる。この一次系列相関係数は0.229である。しかし6つの常数があてはめられているから（傾向に対するもの5とIの1）これを検定する簡単な方法はない。

その大きさは系列相関の存在および解析に用うべき既往の値を示している。このことは14図を調べてみても分る。即ち1.6節でのべたように消費額と収入には時差相関のあることが暗示される。したがつて前年の収入と肥料消費額の効果が調べられた。

1911～1946年の肥料消費指数と収入を表わす新しい変量C_{-1}, I_{-1}が解析に導入された。11.7C表はこれらの変量の分散、共分散分析を示す。

11.7C表　時差収入I_{-1}と時差消費額C_{-1}の分散分析

	自由度	I_{-1}の平方和	I_{-1}とCの積和	I_{-1}とC_{-1}の積和	I_{-1}とIの積和
一次	1	7,501.20	975,090.0	837,409.8	8,357.99
二次	1	1,744.48	322,439.7	269,681.3	1,976.99
三次	1	1,665.55	213,213.7	209,333.1	1,683.79
四次	1	928.06	55,773.1	73,980.5	601.53
五次	1	204.54	26,700.7	11,410.9	396.91
傾向	5	12,043.83	1,593,217.2	1,401,815.6	13,017.21
残差	30	910.81	102,744.1	114,469.8	529.38
計	35	12,954.64	1,695,961.3	1,516,285.4	13,546.59

	自由度	C_{-1} の平方和	C_{-1} とCの積和	C_{-1} とIの積和
一次	1	93,485,674	108855,833	953.058.9
二次	1	41,690,333	49,846,315	305,624.7
三次	1	26,309,870	26,797,605	211,625.5
四次	1	5,897,383	4,445,975	47,951.6
五次	1	636,606	1,489,610	22,143.1
傾向	5	168,019,866	191,435,338	1,520,403.8
残差	30	24642,382	11,571,653	57,315.6
計	35	192,662,248	203,006,991	1,557,719.4

I, I_{-1}, C_{-1} の残差 e_I, e_{I-1}, e_{C-1} に対するCの残差 e_C の回帰分析により次の関係が推定される。

$$e_C = 96.6 e_{I-1} + 0.1448 e_{C-1}$$

これはCの平方和16,155,295を説明しており、Iだけに帰因するものより有意に大きい。したがつて前年度の値は消費額の予想に有意な寄与をしている。

新しい残差が計算され調べられる。一次系列相関係数は0.063となり、これはこの値が殆んど無作為に近いことを示している。

したがつてあてはめた傾向の正しいことが立証できれば、上記の関係の推定値が使用さる。

したがつて次の段階は5次の多項式を使用の要があるか否かを考察してみることである。この問題は共分散分析を用いて取り扱かわれる。

I, I_{-2}, C_{-1} の効果を除いてから各傾向の平方和を計算せねばならない。特別な項、即ち二次の項に対応する平方和の計算を除いて、これは普通の方法で計算され、高次の多項式即ち三次、四次五次の多項式に対応する項は全て残差に一括されている。このようにすれば11.7d表に示す分散分析表がえられる。

11.7d表　I, I_{-1}, C_{-1} の効果を除いた傾向の有意性を検定する分散分析

	自由度	平方和	分散の推定値
一次	1	452,595	
二次	1	1,641,568	
三次	1	85,187	
四次	1	671,768	
五次	1	1,147,626	
傾向	5	3,998,744	
残差	27	7,927,566	293.614
計	32	11,926,310	

この解析では二次の項だけが有意であつた。（五次の項も又有意であるが）この結果二次曲線だけで傾向を表わす可能性を考えてみる必要がある。このため三次、四次、五次の項を残差に含めてしまうと次の新しい関係が推定される。

$$e_C = 81.3 e_I + 8.2 e_{I-1} + 0.261 e_{C-1}$$

今度はこの関係からの残差を考察する。その系列相関は $r_1 = 0.040, r_2 = -0.028, r_3 = -0.131$ である。これは充分に小さいから残差の独立性の仮定は妥当である。したがつて完全な関係

$$C = 81.3 I + 8.2 I_{-1} + 0.261 C_{-1} + 3.21 t^2 - 122.17 t + 21.33$$

を定めることでこの解析は完成する。たゞしtを含む多項式は傾向を表わす。1911年に対して $t = 1$ ）

この解析は暫定的なものであり、使用されている方法は不完全なものであることを強張しなければならない。例えば11.7C表の解析では C_{-1} が除かれているから分散比の検定はもはや正確ではない。

しかしこの場合にはIと I_{-1} だけを除いても同様な結果が得られるからこれは結論に大きな影響をおよぼすとは思われない。同様に C_{-1} が除かれているから残差の系列相関の簡単な検定は得ら

れない。この結果上記の関係は傾向と系列相関の影響をうけていないことは明白であるが、適当な有意性の検定方法が知られていないので試験的に取り扱うべきである。

1.8 時系列解析における計算と検定

Computing and testing in time series analysis

本章の別の節で 時系列間の関係が推定できる方法を示した。この方法を行うにあたつては、色々な量を計算し、検定しなければならない。今までのところ、これらの量の計算方法については何も説明されていなかつた。今度はこの方法を考察してみよう。

計算の大部分は平方和、積和の計算で占められ、これを減すわけにはゆかない。系列が原系列の1～78番目、2～77番目(3～100番目)のような副次系列に分離されゝば、吟味を行うことができる。平方和、積和の多くは一項違うだけであり、相互に吟味できる。

系列からの残差の系列相関を検定するため幾時でもはつきりと残差を計算する必要があるとは限らない。残差の平方和、積和の代数展開は簡単な式で表わされることが多い。

例えば一組の残差 $xi-bxi-1$ の一次系列相関を計算するには

$$\sum_{i=2}^{n-1}(xi-bxi-1)^2=\sum_{i=2}^{n-1}x_i^2-2b\sum_{i=2}^{n-1}x_ix_{i-1}+b^2\sum_{i=2}^{n-1}w_{i-2}^2$$

$$\sum_{i=3}^{n}(xi-bx_{i-1})^2=\sum_{i=3}^{n}x_{i-1}^2-2b\sum_{i=3}^{n}xixi-1+b^2\sum_{i=3}^{n}xi-12$$

$$\sum_{i=2}^{n-1}(xi-bxi-1)(xi+1-bxi)=\sum_{i=2}^{n-1}xixi+1-b\left(\sum_{i=2}^{n-1}xi^2+\sum_{i=2}^{n-1}xi-1xi+1\right)+b^2\sum_{i=2}^{n-1}xi-1xi$$

この式の平方和、積和はいづれも直接計算できる。

その他の計算にも同様な式がある。

本章の系列相関の検定についての説明では正、負の系列相関が在ると仮定した。普通はその通りであるが、多くの場合には正の系列相関だけを考える必要がある。連続した項に負の相関のある可能性は排除できる。確信をもつて排除することができれば、正規の検定で与えられる確率水準は半分となる。系数のとりうる値の範囲にこの制約を認めればより鋭敏な検定がなされる。

この例では3.5節の point-pair 検定を用いて迅速ではあるが効率の悪い正の系列相関の検定がなされた。傾向が系列から除かれておれば系列相関はこの方法を用いて迅速に検定される。特に2～3離れた項間の相関は2～3離れた点の対を用いて検定できる。

1.9 時系列解析の最近における発展

Recent developments in time series analysis.

時系列に関する最近の研究は、主に学究的なものであり、多くの場合その応用は限定されている。単一および組分けした時の時系列の姿を表わす方程式の検定およびあてはめの問題について多くの注意が払われた。特に "自己回帰" auto-regressive と "移動平均" moving average 法の体系が広く研究された。本章に示してあるものより高次の系列相関係数についても幾多の注意が払われた；時系列の固有の構造を調べるため系列相関係数の組でつくられている "コレログラム" correlogram が用いられた。

この後者の方法で得られた結果はより正確な回帰方程式の推定値をうるために用いられるが多くの場合、小数の一次相関係数だけを用いて求めたものよりそれ程正確ではないであろう。この方法で得られる主な利点は基礎関係に関する見解が与えられる点にある。

基礎関係の推定の別の面も多くの研究特に経済統計家によつて与えられている。推定された基礎関係を有意義な関係と結び付ける問題は広く研究されており（動態経済模型 Dynamic Economic Models の統計的推論 New York, 1950 参照）この様な関係の推定方法が提案された（例えば12.6節参照）

時系列の発展の大部分はこゝで取り扱うには進歩が早くかつ複雑である。この分野の最近の研究については参考文献が参考になるがこれは全体の一部分に過ぎない。しかしこれを用いれば次の読書の手引となるであろう。

12 多変量解析 Multivariate Analysis

12.1 順位のある組間の判別 Discrimination between orderd groups

前章では主としてある変量と他の変量の系列との関係の推定およ

び検定に限られていた。これは連関に関する問題では最も一般的なものであるが，この形の問題だけに出合うとは限りない。本章では別の形の連関の問題を取り扱う。

この例は数変量の観測値が数組又は数組の違つた条件で取られたときに起る。測定値がとられた組わけ又は条件を最もよく反映する測定値の組合せを求めたい。例として同属の数種のものの特性に関する測定値の組がとられ，これらの測定値から種を簡単に区別できる指標を作り上げたいとする。或は，ある病気で死亡したものと生き残つたものについて耐久力（処理を受けたものを含む）が観察された。この病気で生き残るものを区別する特性又は特性の組合せを決定したい。この例はいずれも判別函数を推定する必要のあることを示している。

判別したいと考えている組の順位を付けることができれば，その解析は通常，回帰分析と同じ形に変えることができる。即ちある値は各組に割当てられ，回帰分析で従属変量として用いられる。解析でいろいろな種が道化論的尺度で順位が付けられておれば，その相対的位置を示すため，種に表点を与えるとよい。同様に患者の病状の程度をいろいろな初期の徴候に従つて順位が付けられれば病気の尺度で患者のグループの相対的位置を示すため表点が用いられる。いずれの場合でも，解析では組を区別する適当な判別函数を推定するため表点が用いられる。

回帰分散は指定グループの判別函数を推定する適切なる方法であり，これによつて誤りなくグループに分類できる。この仮定は普通の場合と正反対である。即ち誤差は全て独立変量にあるものと仮定され，従属変量には誤差はないものとされている。しかし解析の方法は同じである。

順位のある組間の判別函数を推定する例として栄養調査のデータを考察してみよう。この場合，個々の観測値は，家族が摂取した食物によつて3組に分類され，3つの物理的特性即ち体重w，身長hおよび全身長に対する胸骨の長さの比によつて測られた。

この例では，栄養物質の組は順位が付けられ，1～6の値が付けられた。この基本的仮定はこの組は栄養価の点で等間隔に配列されているというのであり，この例では合理的と思われる。

問題は物理的測定値に基ずいて最も上手に組を区別し，個体を組に分類することのできる指標を求めることである。この様な指標は個体又は同様の組の栄養状態を測るのに用いられる。

判別函数を推定するため（477人の）11～12才の小児が解析で用いられた。組に割り当てられた値の平方和は2699.87であり，w，h，cに関する回帰分析を用いて誘導された方程式は

$$35328.63\,b_w + 1892.657\,b_h + 16.955\,b_c = 2507.7$$
$$1892.657\,b_w + 1716.651\,b_h + 14.447\,b_c = 1614.6$$
$$16.955\,b_w + 14.447\,b_h + 0.1320\,b_c = 3.34$$

この方程式から求めた係数は $b_w=0.0353$，$b_h=0.0414$，$b_c=16.23$であり，3変量は同時に平方和202.68を説明している。

普通の分散分析の方法で有意性は検定されるが，余分な測定値を含むための効果も検定される。12.1a表は体重を含ませる効果を検定する分析である。

12.1a表　分散分析

	自由度	平方和	分散の推定値
cとhに帰因	2	184.33	92.16
wに帰因	1	18.35	18.35
c，h，wに帰因	3	202.68	67.56
残差	473	2497.19	5.28
計	476	2699.87	

判別函数は有意であるが体重は函数全体に有意な寄与をしていないことが分る。したがつてc，hだけを用いた函数を使用してもよい。

これは$b_w=0$とおき，上記方程式の後の2つを解くことで計算できる。

多くの場合は、正しいが、判別変量に対して選ばれた尺度について組の指標が正確か否かに多少疑があるならば、方程式の左辺に（全平方和、積和の代りに）組内の平方和、積和を用いれば、不正確となる影響は避けることができる。有意性の検定には補正をほどこす必要があるが、推定された判別函数は順位の不正確によつては影響されないであろう。

上例についてこれができると推定された判別函数は0.0391w+0.0446h+18.6c=1.12 (0.0549w+0.0416h+16.61c) である。この後の形によれば事実上、上記の函数の倍数になつていることが分る：いずれの解析でも3つの測定値には同じ重みが付けてある。したがつて前記の解析が認められる。

（注、Bartlett,M.S.J.R. statist,Soc.B,9 (1947) 183.

2つの組だけを区別しなければならない時には幾時でも前者の方法が使用できることに注意を要する。この場合組は常に順序付けられたものと考えられ、任意の値、例えば夫々+1，−1が判別函数を推定するために、それらに割り当てられ、組に割り当てられた値の如何を問わず測定値のある組合せ又はそれらの倍数が判別函数として推定されるであろう。

判別函数を推定する別の例として384本のマツの苗木を用いた実験の結果を考察してみよう。96ブロックに配列した苗木についていくつかの処理がほどこされ、最終的に各苗木について4つの測定値がとられた。

x_1，mm 単位の樹幹長

x_2，mm 単位の針葉の平均長

x_3，色の指標

x_4，尖りの指標

この例では特に苗木の3/4は特定の形のアイソレイトで処理されており、したがつて $x_1 - x_4$ を用いてアイソレイト処理を受けた植物と未処理のものとの差違を表わす指標を作りたいとする。

このために、未処理に対しては値3，アイソレイト処理を受けた植物に対しては−1をとる変量yが導入された。（即ちyの総計は0）ブロックおよび処理間の差を除いた $x_1 \sim x_4$ に対するyの分散、共分散分析が行なわれた。自由度368の残差の平方和、積和を用いて次の回帰方程式を作つた。

$$291.193b_1 + 29.655b_2 - 852b_3 - 79b_4 = -753$$
$$29.655b_1 + 52.169b_2 - 473b_3 - 220b_4 = -1453$$
$$-852b_1 - 473b_2 + 98b_3 - 6b_4 = 116$$
$$-79b_1 - 220b_2 - 6b_3 + 152b_4 = -6$$

この方程式に対するyの分散分析が12.16表に示してある。これからアイソレイト処理を受けた植物と未処理のものとの差は有意であるが、x_2，x_3だけに基ずく指標は適切な情報の大部分を含んでいると推論される。

x_1，x_4を削除して求めた回帰方程式は

$y = -0.0179x_2 + 1.0973x_3 + 0.2855$ であり、これはアイソレイトに処された苗木とそうでないものを判別するのに用いられる。或は上例のように x_2 と x_3 が同じ割合で起る別の指標が用いられる。例えば $x_2 + 61x_3$ か $0.016x_2 + x_3$ がアイソレイト処理と未処理とを判別するため用いられる。

	自由度	平方和	分散の推定値
x xに帰因するもの	2	153	76.5
x xに帰因するもの	2	2	1.0
x～xに帰因するもの	4	155	38.8
残差	364	997	2.74
計	368	152	

組の相対的位置が誤差を伴なわずに測定できない時の適当な手法についての問題が生ずる。例えば小麦の品種を区別するために測定された特性を用いたいとする。この品種は収穫物の品質による尺度で配列されている。

判別函数は、もし推定出来るならば、穂の長さや茎の高さ、巾

の様な測定された特性から収量の高い品種を選び出す手段として役立つ。しかしこの場合には各品種の相対的収量の推定に誤差を伴うという難点がある。独立変量，従属変量はいずれも誤差を伴つている。

H. Fairfield Smith[文] はこの問題を考究し，組の相対的位置を決定する際の誤差の修正方法を求めた。この方法を行うには各品種の収量と別の測定値の観測平均値との共分散を推定する必要がある。これを行うには品種内および品種間に変動を分割する共分散分析を行い，これから8.7節で説明した共分散の成分を推定する。（本節の記号では成分Y）

この成分が推定されたならば，普通の方法で解析が行なわれ，この成分は回帰方程式の右辺の要素を構成し，品種平均間の分散，共分散の推定値は左辺の係数を構成する。この後者の値は各品種の観測数で品種間の分散，共分散を割ることにより直接求められる。

この方法は，最初の例では植物を選ぶために用いられているけれども，測定された特性の誤差が推定できる場合の判別函数に関する問題にも同じ様に用いられる。例えば種々の個体のある組織の重さの差を示すような外部より測定できる特性値の組合を求めたいとする。この組織の重さは不正確な方法例えばX－線以外には推定できないとする。ある個体の組織の重さの推定と別の測定を繰返し，上記のような解析にこれを用いる必要がある。

この方法も本節の始めに説明した方法も新しい手法は用いていない。即ちいずれも共分散分析と回帰分析を用いて行なわれる。しかし先験的知識又は補足的測定値を用いても判別したいと考えている組に順位をつけることができなければ，判別函数の推定は一層困難となり，特殊な方法が必要となる。このことについては12.6節で考察することにする。その前に別の適当な方法および問題を考察してみよう。

[文注] Smith, H. Fairfield. Ann. Eugen. 7 (1936) 240

12.2 決定方程式の行列式，平方根およびベクトル
Determinants and the roots and vectors of determinantal equations

次の節では，一般に用いられている行列の誘導を考える必要があろう。その1つが行列式である。

行列式は行列の要素から計算される。

これを誘導する方法は，正確に各行および列から1つの項をとつて，あらゆる組の項の積を計算することである。次の規約に従つてこれらの積には符号が付けられる。まず1列，2列…… に項がある行に注目する必要がある。完全な順序を得るに必要な変換数を数える。これが偶数であれば積には正号を付け，奇数であれば負号を付ける。

この手順を示すため，行列の行列式を計算する。

$$\begin{pmatrix} 607.12 & 212.54 & 408.26 \\ 212.54 & 233.67 & 163.97 \\ 408.26 & 163.97 & 482.14 \end{pmatrix}$$

普通の記号を使えば，これは次の様に表わせる。

$$\begin{vmatrix} 607.12 & 212.54 & 408.26 \\ 212.54 & 233.67 & 163.97 \\ 408.26 & 163.97 & 482.14 \end{vmatrix}$$

括弧の代りに平行線が使われる。

この行列式の計算では6つの項が求められる。

	行順序	交換数	符号
607.12×233.67×482.14=68,399,143255056	123	0	＋
607.12×163.97×163.97=16,323,126.005608	132	1	－
212.54×212.54×482.14=21,779,831.526424	213	1	－
212.54×163.97×408.26=14,227,936.038188	231	2	＋
408.26×212.54×163.97=14,227,936.038188	312	2	＋
408.26×233.67×408.26=38,947,234.103292	321	3	－

各積に付ける符号はこの項の右辺に示してある。例えば，3番

目の項は２，１，３行から求めたものであり２１３から１２３に変えるのに要する交換数は１である。したがつてこの項には負号がつけられる。この符号を用いて，総計19.804.−823.696108が計算される。これが行列式の値である。

この方法が用いられるには，n行n列の行列はn項（半分は＋半分は−）を計算する必要がある。大きな行列式に対しては小行列式の値を使用する別の計算方法がある。この方法については，読者は行列式に関する書物を参考されたい。

（注　例えばAitkins,A.C, Determinants and Matricee, Edinburgh）

次の段階としては行列式を用いた方程式を解く必要がある。例えばこの方程式を満足するSの値を求めるためには次式を解かなければならない。

$$\begin{vmatrix} 607.12-S & 212.54 & 408.46 \\ 212.54 & 233.67-S & 163.97 \\ 408.26 & 163.97 & 482.14-S \end{vmatrix} = 0$$

そのためにはSの3次式を与える行列式を展開すればよいが，Sの色々な値は対する行列式を計算すればもつと簡単に解ける場合が多い。

例として，Sのいろいろな値に対する行列式の値が１２.２ａ表に示してある。

１２.２ａ表　行列式の値

S	行列式	1次差分商	2次差分商	3次差分商
0	19,804,823.696108			
123	12,009.215408	−160,917.1909		
124	− 15,492.655492	− 27,501.8709	1075.93	
153	− 10,097.811592	186.0291	922.93	−1.00
154	16,846.117508	26,943.9291	891.93	−1.00
1046	98,391.954708	, 91.4191	.− 30.07	−1.00
1047	−726,711.136192	−825,103.0909	−924.07	−1.00

この方程式の根はこの例では大体１２３.４，１５３.４，１０４６.１であり，この積が原行列の行列式に等しいという事で吟味される。

一般に差分商による方法を用いて多項式から計算した値で自動的に吟味できる。この計算が正確であれば多項式と同次の差分商は一定となり，Sの係数（この場合には−１）の行列式に等しくなるであろう。

１次差分商は行列式の連続した値間の差をそれに対応するSの値の差で割ることにより計算される。例えば

$$\frac{-10097.811592+15.492.655492}{153-124} = 186.0291$$

この値の差を２つ離れたこれに対応するSの値の差で割れば２次次差分商がえられる。以上同じ。

例えば，

$$\frac{186.0291+27501.8709}{153-123} = 922.93$$

$$\frac{922.93-1075.93}{153-0} = -1.00$$

高次の差分商に関する吟味は逆に使用することができる。即ち行列式はSの別の値を簡単に計算するのに用いられる。これによつて根の逐次近似値が計算できる方法がえられ，これを用いて行列方程式が解かれる。

例えば一次補間によれば上の方程式の最大根は約1046.12である。このSの値に対応する行列式の値が必要であれば，その解析は１２.２ｂ表のように行う。

この場合，３次差分商は−１.００であり，他の値は上の手順を逆に用いて求められる。

−924.07−1.00 (1046.12−154) =−1816.19

−825,103.0909−1816.19 (1046.12−1046) =−825,321.3337

−726711.136,192−825321.3337 (1046.12−1047) =−428.362536

12.2 b表　行列の値の推定

S　S	行　列	一次差分商	二次差分商	三次差分商
154	−	−	−	−
1046	−	−	−	−
1047	−726,711.136192	−825,103.0909	−924.07	−1.00
1046.12	−428.3.62536	−825,321.3337	−816.19	−1.00

この後の値ーかなり小さいがーはその根が仮定した値1,046.12に極く近いことを示している。1,046と1046.12の間の一次補間の値を用いればもつとよい近似値がえられる。この様にすれば根の推定値1046.11948がえられ，行列式の対応する値は−0.027である。

行列方程式を作る方法について説明しよう。

n＋1変量$x_1 x_2 \cdots\cdots x_n$を含むn組の方程式があり，この方程式が次の形をしているとする。

$$(a_{11}+b_{11}S)x_1+(a_{12}+b_{12}S)x_2+\cdots\cdots+(a_{1n}+b_{1n}S)x_n=0$$
$$(a_{21}+b_{21}S)x_1+(a_{22}+b_{22}S)x_2+\cdots\cdots+(a_{2n}+b_{2n}S)x_n=0$$
$$\cdots\cdots\cdots\cdots$$
$$(a_{n1}+b_{n1}S)x_1+(a_{n2}+b_{n2}S)x_2+\cdots\cdots+(a_{nn}+b_{nn}S)x_n=0$$

即ちこの式は変量x_1～x_nを含む線型同次式である。

Sの方程式を求めるためにこの方程式の組から変量x_1～x_nを削除する。

$$\begin{vmatrix} a_{11}+bS_{11} & a_{12}+b_{12}S & \cdots\cdots & a_{1n}+b_{1n}S \\ a_{21}+bS_{21} & a_{22}+b_{22}S & \cdots\cdots & a_{2n}+b_{2n}S \\ \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots \\ a_{n1}+bS_{n1} & a_{n2}+b_{n2}S & & a_{nn}+b_{nn}S \end{vmatrix} = 0$$

これはSのn次の行列式であり，Sに対するn個の値を求めるため上記の方法で解くことができる。

このようにして求めた値を原方程式に代入し，x_1～x_nの値を求めるため解くことができる。

しかし我々は方程式より一個多い変量でもつて出発しているからx個の解には1個の不確定常数を含んでいる。

x個のものが1に等しいとおくこともできるが，多くの場合にはその平方和が1に等しいように解は列べてある。この後の場合には解は行列方程式の固有ベクトルといわれている。

例えば上記の方程式で最大の根1,046.12に対応する固有ベクトルを求めるため次の方程式を解かねばならない。

$$-439.00x_1+212.54x_2+408.26x_3=0$$
$$212.54x_1-812.45x_2+163.97x_3=0$$
$$408.26x_1+163.97x_2-563.98x_3=0$$

この方程式は独立ではなく，たゞ一意の解は得られない。x_3を1とし，初めの二つの方程式を解けば$x_1=1.1767$，$x_2=0.5097$が得られる。（第3の方程式に代入することによつて吟味できる。これらの値を$(1.1767)^2+(0.5097)^2)+(1.0000)^2=1.626\,16$で割れば$x_1=0.7236$　$x_2=0.3134$　$x_3=0.6149$が得られこれが固有ベクトルである。

同様な方法で他の根に対応するベクトルは

S＝123.44に対して

$x_1=-0.6767$，　$x_2=0.4974$，　$x_3=0.5428$

S＝153.37に対して

$x_1=-0.1352$，$x_2=-0.8093$，$x_3=0.5716$

計算全体に対して丸めの誤差の範囲内でベクトルの任意の対の積和は0に等しいという自動的吟味が行なえる。例えば

$(0.7236)(-0.6767)+(0.3134)(0.4904)+(0.6149)(0.5428)=-0.000007\,24$

本節で説明した手法は次節で多変量に対する手法を説明する時に用いられる。この節では良く使われる行列や行列式を説明するためいろいろな項が用いられる。m組の変量を考えれば，観測値から計算されたm×m行列に対して一定の項が用いられ　のであろう。i行j列の要素がi番目とj番目の変量の補正された全積和から成つておれば，このような行列は全積行列と呼ばれる。これに対して要素が変量の共分散又は相関から成つている行列は共

分散行列或は相関行列と呼ばれる。このような量からなる行列式に対しても同じ様な項が用いられるであろう。

12.3 主成分 Princi Pal components

数変量が同時に観測されている場合には，その相互関係を表わす方法に関心がひかれることが多い。特に変量の意味する情報を元の変量の代りに使用できる小数の変量でうまく適当に表わせるかどうかの知ることが大切である。例えば人体測定学では一般に座高、身長、脚長、胴長等の様な多数の測定値がとられる。これは、これらの測定値の相互関係を表す方法および小数の測定値で必要な情報全でを求めることができないか否かという問題と関係がある。

この種の問題は主成分法を用いて取り扱かわれる。この方法はまず最大の分散を持つている観測値の函数を求める。それから第1の函数に独立な，最大分散を持つている函数を求める。次に初めの2つの函数に独立な，最大分散をもつている函数を求める。この様な函数が主成分である。

今m変量が観測されたとするとこのような函数のm個で原観測値の全ての変動が説明できることは明らかである。問題はより小数の函数が使用できるか否かということである。小数の函数を用いることができれば、主成分のあるものは変動を全然説明しておらず、単に他の主成分から若干の測定値を誘導するに使われるであろう。

主成分に解析する手法は前節で述べたものと同じである。実際この節で用いている行列は、4,7,8図および数値は第4章で用いた体重、頭の周囲、胸囲の測定の全積和行列である。

この行列から作られる行列方程式の根は他の各成分に帰因する平方和に対応しており、その最大のものは1,046.12であり、他のものは123.44と153.37である。

この3つの成分の形は固有ベクトルから求められ、これは原測

定値に付けられる係数を与える。

したがつてこの3つの成分は

$t_1 = 0.7236W + 0.3134H + 0.6149C$

$t_2 = -0.6767W + 0.4974H + 0.5428C$

$t_3 = -0.1352W - 0.8095H + 0.5716C$

第2，第3の成分は全変動の無視出来難いが小さい部分を説明している。したがつて3つの測定値の変動の大部分を説明するためめ2つの変量を用いてもよいように思われるが、これだけでは全体の変動を説明できないであろう。

逆に，この3つの測定値は次の方程式でこの主成分を用いて表わせる。

$W = 0.7236t_1 - 0.6767t_2 - 0.1352t_3$

$H = 0.3134t_1 + 0.4974t_2 - 0.8093t_3$

$C = 0.6149t_1 + 0.5428t_2 + 0.5716t_3$

この形では比較的小さな成分を無視又は消却した効果は簡単に調べられる。例えば、主成分に対応する軸は

$$\frac{W}{0.7236} = \frac{H}{0.3134} = \frac{C}{0.6149} = t_1$$

これは変動大部分を示す直線を与える。

この解析は事実上多変量分布の表示法および基礎関係の推定法を導びく時10.4節で用いたものと同じである。この場合に用いられた計算は違つているが、5.9節を参照すればそこで逐次近似で解かれた方程式は、こゝで前節の方法で解いたものと同じ型の方程式であることが分るであろう。

5.9節の結果の形とこの場合のものとの相違に注目してみよう。5.9節における基礎関係の推定方法は最小分散をもつ変量は0に等しいとすることに相当する。これは別の変量の値からある変量を決めるのに用いられる単一関係を求める最良の基礎を与える。

しかし実際には理論的解析を行うには2つ以上の関係が存在するか、それが必要である。この場合には最小の変量は0に等しい

とおくことによつてさらに別の関係が誘導される。例えば上式で $t_2=t_3=0$ とおけば、その関係が求まる。もちろん t_1 にだけ変動が存在すると考えることで得られる直線と同じものが得られる。

10.4節で用いた表示方法と同じ批判が主成分法についてもなされる。

これは誘導された成分は測定尺度によつて変るということであり、ポンドからキログラム、cmからインチに変換することにより、違つた成分の組が求まる。

この事実は重要ではあるが主成分の最も重要な特性は変らない。即ち全分散の0又は無視しうる部分を説明する各成分は、測定値間に値の等しいものが存在することを示している。

六つの測定値に含まれている全ての情報が事実上、例えば二つの主成分に含まれていることが解析によつて示されれば、このことは用いられる測定尺度の如何を問わず常に正しい。

10.5節の第2の例はこの簡単な例である。

この結果主成分解析を有効な独立変量の数の推定に限定すれば、困難は生じないであろう。

しかし主成分が実際的な意味をもつていると仮定すれば、この方法を詳しく調べてみる必要がある。

測定尺度の問題の処理方法の一つは分散が1になる様に観測値を標準化することである。実際には、これを行うには成分分析で全積和行列の代りに相関行列を用いる。例えば、上に用いたデータについては次の行列が用いられる。

$$\begin{pmatrix} 1 & 0.5643 & 0.7546 \\ 0.5643 & 1 & 0.4885 \\ 0.7546 & 0.4885 & 1 \end{pmatrix}$$

これより行列式の根 2.2123、0.5495、0.2382 がえられる。この根の数は行列の次数に等しい。これに対応するベクトルは (0.6124、0.5237、0.6123)(0.2629、−0.8414、0.4722) (−0.7457、0.1338、0.6527) である。標準偏差 2.231、1.384

1.988 を用いれば、次式が得られる。

$$W = 2.251(0.6124t_1 + 0.2629t_2 - 0.7457t_3)$$
$$H = 1.384(0.5237t_1 - 0.8414t_2 + 0.1338t_3)$$
$$C = 1.988(0.6123t_1 + 0.4722t_2 + 0.6527t_3)$$

この解析の主な性質は前と同じである。：一成分が分散の大部分を説明しており、その軸は

$$\frac{W}{0.6831} = \frac{H}{0.3624} = \frac{C}{0.6086} = 2t_1$$

である。

しかし、この形式の相違は十分大きいのでいづれかの形式が、簡便な表示法より深い意味があると解釈しないよう用心する必要がある。

主成分の最後の用法に言及しよう。多数の冗長な回帰分析を行う必要のある場合には、独立又は独立に近い変量を用いると便利なことが多い。初めの因子分析によつて一般に分析を簡単にし、(例えば。逐次近似法によつて)それによつて、正規の手順で行うことができる変数変換(積和行列の変換)が発見される場合が多い。

12.4 因子分析法 Factor analysis

主成分解析で取り扱かわれるものより難しい問題を解くため因子分析法が行なわれる。観測された変量は関係のある小数の因子で殆んど決定できるが、各測定値に含まれる誤差或各測定値に特有の因子の誤差は結果に影響すると仮定する。

互に関連して測定値を決定するのに役立つ因子を推定することが主要な問題である。その補助手段として観測値間の相関をうまく要約できる最小数の因子を推定する。

例えば5つの測定値 $X_1 \sim X_5$ がとられたとする。これらの間の関連を一つの因子即ち $x_i = a_i t + g_i$ $(i=1\sim5)$ を用いて表わすことができるか、二つの因子、即ち $X_i = a_i t_1 + b_i t_2 + g_i$ 又は三つの

因子を用いて表わさねばならぬかという問題生ずる。5測定値の場合には3因子を用いれば常に充分表示できることが示される。即ち3因子以上は考慮する必要はない。※

因子分析法の問題は本来心理学的測定の分野で発生したものであり、その場合には心理学的吟味の結果を説明しうる因子を分離することが望まれている。このような因子は"知能""数学的能力""記憶力""音学的能力"等の測度として役に立つ。しかしその応用は心理学の分野を越えて人類学，経済学さらには任意の事態に関係のある因子を推定する必要のある分野におよんでいる。

しかし一寸考えてみれば分ることであるが、因子分析法の問題が基礎関係を推定する問題と真く同じである。このことは一般に観測値の誤差（又は特定の因子）について仮定を設けない限り前に進めないことを意味している。

それらの相対的な大きさが既知であれば、5.9節の解析が用いられる。又それらが一定の量を越えないことが分つておれば7.6節で説明した形の組分けによる解析が用いられる。或は観測値をとる方法の計画又は制御により測定値の誤差を推定することも可能になる。8.7.10.4 10.5節で述べた方法で分散共分散の成分を用いれば、因子は直接推定できるであろう。

Tukey [+] はこの方法についての仮設的例を示している。

+ Tukey,J.W. Biometrics, 7(1951)61.

今5.9節で与えた解を12.2節の行列式の形に直すと、行列方程式の根およびベクトルの必要なことが分る。

$$\begin{vmatrix} \sum_{i=1}^{n}(x_i-\bar{x})^2-\lambda_1 S & \sum_{i=1}^{n}(x_i-\bar{x})(t_i-\bar{t}) \cdots \cdots \sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y}) \\ \sum_{i=1}^{n}(x_i-\bar{x})(t_i-\bar{t}) & \sum_{i=1}^{n}(t_i-\bar{t})^2-\lambda_2 S \cdots \cdots \sum_{i=1}^{n}(t_i-\bar{t})(y_i-\bar{y}) \\ \cdots\cdots\cdots\cdots & \end{vmatrix} = 0$$

※ 一般にm組の観測値に対しては、$\frac{1}{2}(2m+1-\sqrt{8m+1})$個以上の因子は不必要である。この値が整数でなければ切り上げた整整数を用いなければならない。

$$\begin{vmatrix} \sum_{i=1}^{n}(x_i-\bar{x})^2-\lambda_m S & \sum_{i=1}^{n}(x_i-\bar{x})(t_i-\bar{t}) \cdots \cdots \sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y}) \\ \sum_{i=1}^{n}(x_i-\bar{x})(t_i-\bar{t}) & \sum_{i=1}^{n}(t_i-\bar{t})^2-\lambda_m S \cdots \cdots \sum_{i=1}^{n}(t_i-\bar{t})(y_i-\bar{y}) \\ \cdots\cdots\cdots\cdots & \end{vmatrix}$$

ただし、$\lambda_1, \lambda_2 \cdots \lambda_m$ は、x,t,… yに含まれる変動の分散（又は平方和）である。

これは又次の形に直すこともできる。

$$\begin{vmatrix} 1-P_1 S & V_{xt} \cdots\cdots\cdots & V_{xy} \\ V_{xt} & 1-P_2 S \cdots\cdots\cdots & V_{ty} \\ V_{xy} & V_{ty} \cdots\cdots\cdots & 1-P_m S \end{vmatrix} = 0$$

ただし $P_1, P_2, \cdots P_m$ は変動（即ち測定値に特有の誤差又は因子）に帰因することのできる各変量における総変動の割合である。

割合 $P_1 \sim P_m$ が全て等しければ、その解析は主成分の推定の場合と同じである。等しくなければPの推定値を求めることから始めてその後の解析でこの値を考察する必要がある。この場合に因子分析法が用いられる。

こゝで割合 $P_1 \sim P_m$ の値を求めるために用いられる方法について充分に説明することは不可能である。この説明については、読者は因子分析に関する書物を参照されたい。※

※ 例えば Thomson,G.H. The Factorial Analysis of Human Ability, Edinburgh; 1939

こゝでやれることは、必要な因子数を推定する方法および、因子数が推定されたら割合 $P_1 \sim P_m$ の値を求める方法を示すことだけである。

要素が、P＋1個の測定値の一組と他の組との相関から成る全ての行列式が0に等しい時のみm個の観測値の系列はP個の因子を含む関係で表わせることを示すことができる。

例として6個の測定値から成るある組の相関行列を

$$\begin{pmatrix} 1.00 & -0.12 & 0.08 & -0.36 & -0.84 & 0.40 \\ -0.12 & 1.00 & 0.87 & 0.60 & 0.42 & 0.78 \\ 0.08 & 0.87 & 1.00 & 0.51 & 0.21 & 0.88 \\ -0.36 & 0.60 & 0.51 & 1.00 & 0.56 & 0.34 \end{pmatrix}$$

$$\begin{pmatrix} -0.84 & 0.42 & 0.21 & 0.56 & 1.00 & -0.14 \\ 0.40 & 0.78 & 0.88 & 0.34 & -0.14 & 1.00 \end{pmatrix}$$

初めの3つと後の3つの測定値との相関の行列式は

$$\begin{pmatrix} -0.36 & -0.84 & 0.40 \\ 0.60 & 0.42 & 0.78 \\ 0.51 & 0.21 & 0.88 \end{pmatrix} = 0.$$

又1、3、5番目の測定値と2、4、6番目の測定値との相関の行列式は

$$\begin{pmatrix} -0.12 & 0.87 & 0.42 \\ 0.36 & 0.51 & 0.56 \\ 0.40 & 0.88 & -0.14 \end{pmatrix} = 0$$

この両行列式、さらに同じようにして作られた行列式はいづれも0になる。

オーダー2の行列式については0とならないからこの場合の測定値間の相関を説明するには2因子が必要と思われる。

しかし2因子で充分であれば、主対角線の要素を適当な値で置換した時には、この行列の3次の行列式が全て0となるであろう。

例えば、xを推定さるべきある量で置換した時には

$$\begin{pmatrix} 0.08 & -0.20 & -0.84 \\ 0.87 & 0.60 & 0.42 \\ 0.x & 0.51 & 0.21 \end{pmatrix} = 0$$

とならねばならない。このために主対角線の要素にはいくつかの制約が加えられるが、このことが必ずしも満されるとは限らない。

これが満されなければ、高次の行列式を用いて満足できるか否かを考察してみる必要がある。

この場合主対角線の要素を夫々0.80、0.90、0.89、0.50、0.98、1.00で置換すれば、行列は

$$\begin{pmatrix} 0.80 & -0.12 & 0.08 & -0.20 & -0.84 & 0.40 \\ -0.12 & 0.90 & 0.87 & 0.60 & 0.42 & 0.78 \\ 0.08 & 0.87 & 0.89 & 0.51 & 0.21 & 0.88 \\ -0.20 & 0.60 & 0.51 & 0.50 & 0.56 & 0.34 \\ -0.84 & 0.42 & 0.21 & 0.56 & 0.98 & -0.14 \\ 0.40 & 0.78 & 0.88 & 0.34 & -0.14 & 1.00 \end{pmatrix}$$

となり、この行列に含まれる三次の行列式はいづれも0となる。

このことは、測定値間の相関を2因子を用いて説明できることを示している。その因子に帰因する測定値の全変動の割合は0.80、0.90、0.89、0.50、0.98、100となる。即ち次の推定値を得た。

$P_1=0.20$、$P_2=0.10$、$P_3=0.0.11$、$P_4=0.50$、$P_5=0.02$、$P_6=0.00$

これらの値は因子および6個の測定値に対する関係を推定することにより解析を完成するのに用いられる。これを行うには主対角線に1−0.20S、1−0.10S、1−0.11S、1−0.50S、1−0.02S、1を用いて上記の行列の根およびベクトルを求めればよい。

この例は全く説明のためのもので因子分析法に含まれている考え方を示しているに過ぎないことを憶えておかねばならない。実際には、行列式が厳密に0となることは稀であり、これは描出変動と一諸にして考えるべきである。さらに因子分析に用いられる正規の解析法は計算を最小にするため形式化されている。

一般にこの解析で最も難しいのは主対角線に挿入する要素一いわゆるCommunalitiesの推定である。このためには特に複雑になりこれらの要素に対する一意的な値が求まるとは限らない。期待した次数の行列式を0とするような値がない（かゝる時にはより高次の行列式を用いねばならない）か、或はCommunalitiesのため数組の値が用いられることが多い。このような場合には、なんらかの解を得るにはさらに仮定を設ける必要のあることが多い。

12.5 正準相関 Canonical correlations

今度は二相の変量間の関係を表わす方法を考えてみることになる。考察される問題の形は二組の変量間の相関を説明する因子をどの様にして求めるかということである。

例として頭と脳に関する測定値、即ち頭の長さ、幅、周囲、脳の重量と一連の知能その他の試験結果との関係を調べたい。この場合には頭に関する測定値の組合せと最も密接な相関のある試験の組合せ、もつと一般的には、いづれも観測値の変動を説明でき、

一対一の組合をした時に、試験も測定値との相関が説明できる試験の組合せと測定値の組合せとの組を求めることに関心がはらわれる。

このような解析では対応する対に列べることのできる知能と測定値の因子を生ずる。

牧野について測定が行われた時にも同じような問題が起る。その測定値一つは土壌の化学的成分を与えるものであり、他の一つは牧野の植物構成を与えるものである。この場合には対にした化学的条件と植物的条件とのどの組が化学的測定値と植物測定値間の相関を説明しうるかを知ることが大切である。このような組に関する知識はこれらの測定値の相互作用を説明するであろう。

この例はいづれも、標準変量が計算される場合の例である。特殊な場合には、二組のこの様なものが存在する。いづれの組も元の測定値の組の一つに対応しており、その組に含まれるもの と同数の変量から構成されている。このような正準変量には、いづれもそれ自身と同じ組の全ての変量および一つの組を除いて他の組の全ての変量と独立であるという性質がある。

この2組の変量は対になつているから、比較的大きな組の変量の或るものは、両組の全ての変量に対して完全に独立しており、二組の間の相関に全然関与しない。

例えば二組の測定値 $x_1, x_2 \cdots\cdots x_n$ と $y_1, y_2 \cdots\cdots y_{m+n}$ がとられたとすると、元の測定値の分散、共分散を説明するために、正準変量 $\xi_1 \xi_2 \cdots \xi_n$ と $\gamma_1, \gamma_2, \cdots\cdots_{m+n}$ が計算される。ξ_1 が γ_1 と ξ_2 が γ_2 ····· とが相関していること。を除いて、これらの変量は完全に無相関である。変量 $\gamma_{n+1} \cdots\cdots \gamma_{n+m}$ は他の如何なる正準変量とも相関していない。

対応する正準変量の対間の相関は2組の測定値が連関している程度を推定するのに役立つ。この相関は普通、正準相関といわれている。

正準相関を推定するにはまず、観測値の全積和、共分散又は相関行列を計算する必要がある。

この場合には、どちら を用いてもさしつかえないが、相関行列が用いられるとすると、測定値はその標準偏差で割つたものとみなされる。この行列が次式で表わされるとする。

$$\begin{pmatrix} a_{11} & a_{12}\cdots\cdots a_{1n} & b_{11} & b_{12}\cdots\cdots & b_{1m+n} \\ a_{21} & a_{22} \quad a_{2n} & b_{21} & b_{22} & b_{2m+n} \\ & \cdots\cdots\cdots & & & \\ a_{n1} & a_{n2} \quad a_{nm} & b_{n1} & b_{n2} & b_{nm+n} \\ b_{11} & b_{21} \quad b_{n1} & c_{11} & c_{12} & c_{1m+n} \\ b_{12} & b_{22} \quad b_{n2} & c_{21} & c_{22} & c_{2m+n} \\ & \vdots \quad \vdots & \vdots & \vdots & \\ b_{1m+n} & b_{2m+n} \quad b_{nm+n} & c_{m+n1} & c_{m+n2} & c_{m+n\ m+n} \end{pmatrix}$$

即ちaは第1組の測定値の積和（又は共分散、相関）を表わし、cは第2組の積和を、bは2組の間の積和を表わしている。

正準相関を求めるには、次式の根とベクトルを求めるため

$$\begin{pmatrix} -a_{11}S - a_{12}S \cdots\cdots -a_{1n}S & b_{11} \cdots\cdots & b_{1m+n} \\ -a_{21}S - a_{22}S \quad -a_{2n}S & b_{21} \cdots\cdots & b_{2m+n} \\ \cdots\cdots\cdots & \cdots\cdots\cdots & \\ -a_{n1}S - a_{n2}S \quad -a_{nn}S & -b_{n1} & b_{nm+n} \\ b_{11} \quad b_{21} \quad b_{n1} & -c_{11}S \cdots\cdots\cdots & c_{1m+n}S \\ b_{12} \quad b_{22} \quad b_{n2} & -c_{21}S \cdots\cdots\cdots & c_{2m+n}S \\ \cdots\cdots \quad \cdots & \cdots & \\ b_{1m+n} \quad b_{2m+n} \quad b_{n\ m+n} & -c_{m+n}S \quad -c_{m+n}S \quad -c_{m+n\ m+n}S & \end{pmatrix}$$

を解く必要がある。この方程式の根は正準相関を与え、そのベクトルは正準変量の対の係数を与える。

簡単に行列の記号で表わせば例えば積和の行列は

$$\begin{pmatrix} A & B \\ \cdots\cdots & \cdots\cdots \\ B & C \end{pmatrix}$$

解べき方程式は

$$\begin{vmatrix} AS & \vdots & B \\ \cdots & \cdots & \cdots \\ B & \vdots & CS \end{vmatrix} = 0$$

という具合にこれは表わされる。

この記号によれば、方程式はさらに次のように短くなる。

$$BC^{-1}B - AS^2 = 0$$

これはSの簡略推定法を与える。

この解析の例としてThurstone※が報告した286人についての一連の試験結果の一部をとり上げよう。知能、記憶力、理解力等の因子を推定するため因子分析に関する完全な結果を使用した。数学的能力テストと推理力のテスト間の正準相関を求めるため、この結果を考察してみよう。考察されたテストは

x_1 加法　　y_1 推理Ⅰ

x_2 代数　　y_2 推理Ⅱ

x_3 幾法　　y_3 推理Ⅲ

推理力のテストと数学的テストの組合で最も密接な連関のあるものを推定する。

これらのテストの相関行列は（Thurstoneの与えたもの）、

$$\begin{pmatrix} 1.000 & 0.284 & 0.684 & 0.070 & 0.064 & 0.049 \\ 0.284 & 1.000 & 0.368 & 0.447 & 0.412 & 0.391 \\ 0.684 & 0.368 & 1.000 & 0.108 & 0.049 & 0.076 \\ 0.070 & 0.447 & 0.108 & 1.000 & 0.368 & 0.355 \\ 0.064 & 0.412 & 0.049 & 0.368 & 1.000 & 0.391 \\ 0.049 & 0.391 & 0.076 & 0.355 & 0.391 & 1.000 \end{pmatrix}$$

即ち次式を解く必要がある。

$$\begin{pmatrix} -S & -0.284S & -0.684S & 0.070 & 0.064 & 0.049 \\ -0.284S & -S & -0.368S & 0.447 & 0.412 & 0.391 \\ -0.684S & -0.368S & -S & 0.108 & 0.049 & 0.076 \\ 0.070 & 0.447 & 0.108 & -S & -0.368S & 0.355S \\ 0.064 & 0.412 & 0.049 & -0.368S & -S & 0.391S \\ 0.049 & 0.391 & 0.076 & -0.355S & -0.391S & -S \end{pmatrix} = 0$$

或は

$$\begin{pmatrix} 1.000 & 0.368 & 0.355 \\ 0.368 & 1.000 & 0.391 \\ 0.355 & 0.391 & 1.000 \end{pmatrix}^{-1} = \begin{pmatrix} 1.231581 & -0.333214 & -0.306924 \\ -0.333214 & 1.270626 & 0.378523 \\ -0.306924 & 0.378523 & 1.256961 \end{pmatrix}$$

および

$$\begin{pmatrix} 0.070 & 0.064 & 0.049 \\ 0.447 & 0.412 & 0.391 \\ 0.108 & 0.049 & 0.076 \end{pmatrix} \begin{pmatrix} 1.231581 & -0.333214 & -0.306924 \\ -0.333214 & 1.270626 & 0.378523 \\ -0.306924 & -0.378523 & 1.256961 \end{pmatrix} = \begin{pmatrix} 0.00679 & 0.04474 & 0.00852 \\ 0.04474 & 0.30195 & 0.05784 \\ 0.00852 & 0.05784 & 0.01329 \end{pmatrix}$$

に注目すれば

この方程式は

$$\begin{vmatrix} 0.00674-S^2 & 0.04474-0.284S^2 & 0.00852-0.684S^2 \\ 0.04474-0.284S^2 & 0.30195-S^2 & 0.05784-0.368S^2 \\ 0.00852-0.684S^2 & 0.05784-0.368S^2 & 0.01329-S^2 \end{vmatrix} = 0$$

この方程式から求めたS^2の値は0.00020、0.00457、0.31307であり、これから正準相関±0.014、±0.068、±0.56がえられる。この正負の符号は正準変量につける符号を任意に選んでよいことを示している。

最初の2つの正準相関が非常に小さいことは、実質的に数学的能力と推理力間のすべての相関が2つの正準変量間の相関を用いて計れることを示している。この最初の値は行列式の最大根に対応する固有ベクトル（−0.029、0.985、−0.169）で示される。それは

$$\xi_1 = -0.029x_1 + 0.985x_2 - 0.169x_3$$

他の値を求めるには、根S＝0.560に対応する完全行列の固有ベクトルを求めねばならない。

これは（−0.029、0.985、−0.169、0.488、0.397、0.334）である

したがつて正準変量は

$y_1 = 0.488y_1 + 0.397y_2 + 0.334y_3$)

したがつて3つの推理力テストの平均は（前者には幾分大きな重みがつけてある。）乗算の能力に対して修正を施した算術テストと相関のあることが分る。これはテスト間の大部分の相関を説明している。

12.6 順位のつけられない組間の判別

判別しようとする組を既往の知識により順位を付けることができない場合に判別函数を求める方法を考察することにする。このためには組の相対的順位を推定する解析を行うと同時に、これらの組をうまく判別する観測値の函数を求めねばならない。これを行うには前節で示した正準相関法を用いればよい。その理由について若干の考察をしてみよう。

2組の場合には、それらは常に正確に順位がつけられているとみなしうるから、判別函数は回帰分析で推定できることを12.1節で指摘した。このために、任意の記号を各組に割当てる。

3組の場合には、任意の1組の記号では、組は正確な順位に配列されないが、2組の記号を適当に組合せればそうすることができる。

例えば2組の記号の適当な組合せ（1、0、−1）と（1、−2、1）により、3組の相対的順位は再現される。同様にP組の場合には、例えばP−1組の記号の適当な組合せにより、組は任意の方法で順位をつけることができる。

判別函数を推定する問題は、大分はつきりしてきた。即ち、どのような表点法 scoring system 即ち、表点のどの様な組合せが測定値のどの様な組合せと最も密接な相関があるかを求めねばならない。したがつてこれは正準相関および変量を推定する問題である。

P組の場合には、P−1組の表点法と測定値との正準相関は前

節の方法で推定される。これらの内最大のものに対応するベクトルは、適切な表点と判別函数を与える。

前節の解析法は次の行列式が得られるので簡単に示すことができる。

$$\left| \begin{pmatrix} \text{組間} \\ \text{積和行列} \end{pmatrix} - S^2 \begin{pmatrix} \text{全積和} \\ \text{行列} \end{pmatrix} \right| = 0$$

Sは前と同様正準相関を表わす。この公式はどのような表点法が組に使われたとしても正しい。この結果、推定された表点と特定の理論的表点（例えば、等間隔の組）との隔りの状態について特に配慮されていない限り一般に表点法を形式化する必要はない。この解析は表点法を用いなくても行うことができる。

この形の解析の例として12.1節で用いた栄養物質のデーターをもう一度考察して、完全判別解析を用いた栄養物質グループの相対的位置を推定してみよう。

この3つの測定値の組間の平方和、積和の行列は

$$\begin{pmatrix} 2947.04 & 1951.03 & 3.545 \\ 1951.03 & 1639.04 & 2.284 \\ 3.545 & 2.284 & 0.0111 \end{pmatrix}$$

即ち解くべき方程式は

$294704 - 353286S^2$　　$195103 - 189263.7S^2$　　$3545 - 1635S^2$

$195103 - 189263.7S^2$　　$163904 - 171665.1S^2$　　$2284 - 14447S^2$

$3545 - 16935S^2$　　$2284 - 14447S^2$　　$00111 - 01320S^2$

この方程式を満足するS^2の値は0.03238、0.06534、0.11107でありこの最大値が判別函数を与える。

この函数の正準相関は（0.11107）＝±0.333であり、判別函数の係数の推定値は0.001435、0.001726、1.0000である。この値は任意の係数即ち20を乗ずることにより、係数0.0287、0.0345、20.000がえられる。これは回帰分析により12.1節で求めた係数とよく一致している。したがつて組の相対的順位は直線から離れていないことが分る。

6組に対する判別函数の表点を推定することによりこのことは立証できる。これは公式

表点 = 0.0287W+0.0345H−20C−18

に組平均を代入すれば求まる。この場合常数18は表点を適当な大きさに減すため引かれる値である。この結果表点は夫々0.24、0.37、0.52、0.56、0.88、0.98となり、これは正確に順序に並んでおり殆んど直線である。

この解析で普通の表点の代りに二次の表点を使うことにより正確な順位が確保されることに注目すると面白い。適切な表点法は一次の表点一の組と二次の表点の組の適当な組合せから求められる。一次の組の表点法に対する観測値の2次の回帰を計算する必要があることを除いては、この解析は上記と同じ方法で行なわれるであろう。判別函数は

$$\left|\begin{pmatrix}\text{二次の回帰の}\\ \text{積和行列}\end{pmatrix} - S^2 \begin{pmatrix}\text{全積和行列}\end{pmatrix}\right| = 0$$

の最大根に対応するベクトルである。

この場合には、この方程式は0でない根を二つ持っているに過ぎない。

この方法で求めた判別函数は、全体にわたつて組間を最も良く判別するが、組の特定の対間を判別する方法としては不充分なものであることに注意しなければならない。例えば4組の表点法が+1、+1、−1、−1と推定されたとすると、判別函数は1組と2組、3組と4組とを区別することはできない。これを行うには組の各対毎に判別函数を計算するか、又は行列式の第2、第3、.. ..... の最大限に対応する正準変量を計算する必要がある。これらのものは補助的な有力でない判別函数を与え、この判別函数は主函数と一諸に用いられて組のあらゆる対を区別するであろう。

12.7 時系列における基礎関係

Underlying relationships in time series

M. S. Bartlett は時系列における基礎関係を推定するのに判別解析を利用することを暗示した。時間と表点を同一のものとして扱い、上記の二次の回帰の行列の代りに一般の曲線回帰の積和行列を用いることにより、観測値の正準変量を求めることができる。これは原系列の時間における変動を説明するのに用いられる。変動の有意でない部分を説明する正準変量はいづれも時間に独立である、したがつて"基礎"関係を表わすものとみなされる。

あてはめられた回帰の次数がなんであつてもこの関係にどのような変量が用いられようともこの方法が使用される。二つ以上の関係も恐らくこのような関係で推定されるであろう。

例として12.7a表には11.7節の肥料消費量と農家収入の解析から求めた最小の正準相関に対応する正準変量が示してある。この表は時差変量C_{-1}とI_{-1}が用いられるか否かによつて二つの場合が示してある　1と5の間のあてはめられた諭位の)正準変量(Cの係数は1に等しいように選ばれている。　が示してある

12.7a表　時系列解析のための正準変量

時差なしの解析

次数	S^2の値	対応する正準変量	S^2の次に大きな値
一次	0.000	C−116.7I	0.632
二次	0.225	C−127.2I	0.769
三次	0.225	C−127.0I	0.882
四次	0.250	C−125.9I	0.902
五次	0.395	C−121.6I	0.944

時差を入れた解析

次数	S^2の値	対応する正準変量	S^2の次に大きな値
一次	0.000		0.000
二次	0.000	不確定	0.000
三次	0.000	C−1.20C₋₁−1818I+206.1I₋₁	0.018
四次	0.010	C−1.20C₋₁−72.1I+85.9I₋₁	0.100
五次	0.010	C−1.18C₋₁−77.5I+88.0I₋₁	0.297

S^2の最低の値は傾向の自由度が変量の数より小さい時には0であり、少くともこの場合には対応する正準変量は正確には決定できないことが分る。

(☆ Bartlett M.S. Econometrica, 16(1948)323

この様な場合以外は推定された関係は明らかに一定である。時差を用いない解析では正準変量は略 C−127I でありこれは C = 127I＋常数を意味している。時差を入れた解析は恐らくより正確な関係を与えるであろう。(S2の値を比較せよ) これは略

$$C-1.20C_{-1}=72.1I-85.9I_{-1}+\text{常数}$$

でありもつとはつきりした形で表わせば

$$C-1.2C_{-1}=72(I-1.2I_{-1})+\text{常数}$$

この後の形は $C-1.2C_{-1}$ と $I-1.2I_{-1}$ の増加量が一次関係であることを意味している。

有意でない変量を構成するものを決めるため Bartlett は次節で説明する検定を用いることを暗示した。しかしこれらの検定は推定された関係からの残差が系列的に独立である場合にのみ正当であるに過ぎない※ したかつてこの方法は有意性についての概略の指標として用いられるだけであり一般に或る程度の許容範囲

※注 次節で用いられる X^2 の値が因子 $1+2r.r'+r_1r_2'+\ldots$ で修正されるであろうということ以外は前章と類似している。ここで $r_1\,r_2\cdots$ は残差の系列相関を表わし、$r_1'\,r_2'\cdots$ は対応する正準傾向函数の系列相関を表わす。

をもうけて用いねばならない。

12.8 多変量解析における有意性の検定

Tests of significance in multivariate analysis

今や多変量解析における有意性の検定を考察する段階に到達した。特に前節で論じた因子および成分の有意性の検定方法について概略述べることにしよう。このような方法一主として M.S. Bartlett によるものであるが一は近似的なものに過ぎないか、有意性を判断する基礎を与える。

まず多変量の観測値を扱うための分散比の検定の拡張を考察しよう。この問題はここでは数組の測定値の組間変動が偶然に帰因させることのできる変動より大きいかどうかを決めることである

これを行う手段として各測定値の組間、組内、全体の平方和、積和から成る行列を計算する。そして主要な因子の効果は、組間組内の平方和、積和を加えて全体の平方和、積和が得られるように予め除かれている。次いで組内の行列式と全体の行列の比が計算されねばならない。これを Λ で表わす。q と n−q が組間と組内の自由度を表わし、p個の測定値が用いられたとすると組間の差の有意性は自由度が pq の x^2 として

$-[n-\frac{1}{2}(p+q+1)]\log_e \Lambda$ を用いて検定される。

例えば 12.1 節の6つの栄養物の組が同時とられた3つの測定値 W, H, C で有意に違つているかどうかを検定したいとする。これを行うには次のように計算する必要がある。

$q=5,\ n-q=471,\ n=476,\ p=3$

組内の積和の行列式 $=$

3238.159	1697.534	13.390
1697.534	1552.747	12.163
13.390	12.163	0.1209

全体の積和行列式＝
$$\begin{vmatrix} 35328.63 & 18926.37 & 16.935 \\ 18926.37 & 17166.51 & 14.447 \\ 16.935 & 14.447 & 0.1320 \end{vmatrix}$$
$$=29,734,707$$

$$\Lambda=\frac{23.905,074}{29,734,707}=0.80395$$

$$-[n-\tfrac{1}{2}(p+q+1)]\log_c \Lambda=-471.5\ \log_c 0.80395$$
$$=102.83$$

この最後の量が自由度15のx^2として検定されその結果は極めて有意である。したがつて、この組は3つの測定値について有意差があると結論されるであろう。

次に起る問題は組を判別するために求めたどの函数が有意であるかということである。

したがつて組間の差は主判別函数で説明できるか、或は別の判別函数を考える必要があるのかという疑問が起る。

これを行うには正準相関を用いねばならない。これが大きさ$S_1^2, S_2^2 \cdots$の順に配列されておれば、次の関係の成立することが分る。

$\Lambda\Lambda=(1-S_1^2)(1-S_2^2)(1-S_3^2)\cdots\cdots$

したがつて

$-[n-\frac{1}{2}(p+q+1)]\log_c \Lambda=-[n-\frac{1}{2}(p+q+1)]\log_c (1-S_1^2)$
$-[n-\frac{1}{2}(p+q+1)]\log_c (1-S_2^2)\cdots$

この左辺のx^2を大まかに分割することを表わし各項の自由度は略 $p+q-1, p+q-3, p+q-5\ldots$ である。

この方法により各判別函数の有意性を計ることが可能である。

上記の例では$S_1^2=0.11107$, $S_2^2=0.06534$, $S_3^2=0.03238$であることが示されている。吟味として、次の計算を行う。

$(1-S_1^2)(1-S_2^2)(1-S_3^2)=(0.88893)(0.93466)(0.96762)$
$=0.803944=\Lambda$

したがつて全変動の分割は12.8a表に示してあるように行なわれる。

12.8a表　判別函数の有意性の検定

函数	自由度	x^2
1番目	7	$-471.5\log_c 0.88893=55.50$
2番目	5	$-471.5\log_c 0.93466=31.86$
3番目	3	$-471.5\log_c 0.96762=15.47$
計	15	$-471.5\log_c 0.80395=102.83$

この例では3つの成分はいづれも極めて有意であり、最初の判別函数は組間の差に関する全ての情報を含んでいないことを示している。完全な判別を行うには他の判別函数を用いる必要がある。

Bartlett によるこの解析の形は極めて近似的なものであり、その精度を改善するためいろいろな示唆が行なわれた。その内の一つは Marriott※によるものであり、彼は第1の判別函数は自由度 $p+q-1+\frac{1}{2}[(p-1)(q-1)]$ を用いてより正確に検定されることを示した。上例では、これは7の代りに $5+3-1+\frac{1}{2}[4\times2]^{3/3}=9$ なる自由度を与える。しかし第1の判別函数の有意性は依然として変らない。

もちろん同じ形の解析が正準関の有意性を検定するのに用いられる。例えば12.7b表は最大の正準成分だけが有意であるという12.5節で到達した結論を立証している。この場合に最大の正準根を検定するためのより正確な自由度の推定値は7であるがこれは、その有意性には影響しない。

※ Marriott.F.H.C. Biometrika.39(1952)58

12.8 b表 正準相関の有意性の検定（n=285，p=q=3）

相関	自由度	X^2
I	5	−281.5 $\log_e$ 0.68693=105.71
II	3	−281.5 $\log_e$ 0.99543= 1.29
III	1	−281.5 $\log_e$ 0.99980= 0.06
計	9	−281.5 $\log_e$ 0.68365=107.06

実際、主成分および因子の有意性の検定には同じ様な分割が行なわれ、Bartlett *は相関行列から計算された成分を検定するための適当な解析の形式を示した。

この検定の詳細については、この論文および同著者* の別の論文を参照すべきである。

現在では連関のある測定値を扱う方法についての研究の多くは多変数解析の方法および検定を拡張したものと関係がある。毎年、新しく、より正確な検定方法が発見され、採用されている。それらのものを完全に説明するには長いものとなり、すぐ旧式となつてしまうであろう。したかつて新しい方法および検定が完全に決定された、表が作られるまでは、この検定および関連測定に関する論議をこゝで結ぶのが適当である。

* Baretiitt M.S.Brit j.Psychol.statist Sect. 3(1950) 77と1(1948)73

I 表

任意の象限に入る点数の有意水準

点数	下限 0.05	下限 0.01	上限 0.05	上限 0.01
8− 9	0	−	4	−
10−11	0	0	5	5
12−13	0	0	6	6
14−15	1	0	6	7
16−17	1	0	7	8
18−19	1	1	8	8
20−21	2	1	8	9
22−23	2	2	9	9
24−25	3	2	9	10
26−27	3	2	10	11
28−29	3	3	11	11
30−31	4	3	11	12
32−33	4	3	12	13
34−35	5	4	12	13
36−37	5	4	13	14
38−39	6	5	13	14
40−41	6	5	14	15
42−43	6	5	15	16
44−45	7	6	15	16
46−47	7	6	16	17
48−49	8	7	16	17
50−51	8	7	17	18
52−53	8	7	18	19
54−55	9	8	18	19
56−57	9	8	19	20

点　数	下限 0.05	下限 0.01	上限 0.05	上限 0.01
58－ 59	10	9	19	20
60－ 61	10	9	20	21
62－ 63	11	9	20	22
64－ 65	11	10	21	22
66－ 67	12	10	21	23
68－ 69	12	11	22	23
70－ 71	12	11	23	24
72－ 73	13	12	23	24
74－ 75	13	12	24	25
76－ 77	14	12	24	26
78－ 79	14	13	25	26
80－ 81	15	13	25	27
82－ 83	15	14	26	27
84－ 85	16	14	26	28
86－ 87	16	15	27	28
88－ 89	16	15	28	29
90－ 91	17	15	28	30
92－ 93	17	16	29	30
94－ 95	18	16	29	31
96－ 97	18	17	30	31
98－ 99	19	17	30	32
100－101	19	18	31	32
110－111	21	20	34	35
120－121	24	22	36	38
130－131	26	24	39	41
140－141	28	26	42	44
150－151	31	29	44	46
160－161	33	31	47	49

点　数	下限 0.05	下限 0.01	上限 0.05	上限 0.01
170－171	35	33	50	52
180－181	37	35	53	55
200－201	42	40	58	60
220－221	47	44	63	66
240－241	51	49	69	71
260－261	56	54	74	76
280－281	61	58	79	82
300－301	66	63	84	87
320－321	70	67	90	93
340－341	75	72	95	98
360－361	80	77	100	103
380－381	84	81	106	109
400－401※	89	86	111	114

※ N が大きければ $\frac{N}{4} \pm \left(\frac{1}{2} + \frac{a\sqrt{N}}{4} \right)$ を使う。

ただし a ＝5%有意水準に対しては 1.96
1% 〃 〃 2.58

II 表

Tukey の Corner test の有意水準

P＝	有意となるために要する象限和
0.10	± 9
0.05	±11
0.02	±13
0.01	±15※

※ 観測値が 50 以上であれば ±14 を使う。

Ⅲ　表

順位検定．許容しうる最大交換数

N	P=0.05	P=0.01	N	P=0.05	P=0.01
5	0	—	15	32	26
6	1	0	16	37	31
7	3	1	17	43	36
8	5	3	18	50	42
9	8	5	19	57	48
10	11	8	20	64	54
11	14	10	21	72	61
12	18	13	22	80	69
13	22	17	23	88	77
14	27	21	24	97	85

Nが大きければ $\frac{(N-2)(N+1)}{4} - d\sqrt{\frac{N(N-1)(2N+5)}{72}}$ のすぐ下の整数を使う。

ただし $d = \begin{matrix} P=0.05 \text{に対しては } 1.96 \\ P=0.01 \quad 〃 \quad 2.58 \end{matrix}$

Ⅳ　表

中位線の同じ側に落ちる点の対数に対する有意水準

点数	有意対数 P=0.05	P=0.01	点数	有意対数 P=0.05	P=0.01
8— 9	6	--	56—57	34	37
10—11	7	8	58—59	35	38
12—13	9	10	60—61	36	39
14—15	10	11	62—63	37	40
16—17	11	12	64—65	39	41
18—19	12	14	66—67	40	42
20—21	14	15	68—69	41	44
22—23	15	16	70—71	42	45

点数	有意対数 P=0.05	P=0.01	点数	有意対数 P=0.05	P=0.01
24—25	16	17	72—73	43	46
26—27	17	19	74—75	44	47
28—29	18	20	76—77	45	48
30—31	19	21	78—79	46	49
32—33	21	22	80—81	47	50
34—35	22	24	82—83	48	51
36—37	23	25	84—85	49	53
38—39	24	26	86—87	51	54
40—41	25	27	88—89	52	55
42—43	26	28	90—91	53	56
44—45	27	30	92—93	54	57
46—47	29	31	94—95	55	58
48—49	30	32	96—97	56	59
50—51	31	33	98—99※	57	60
52—53	32	34	100—101	58	62
54—55	33	35			

※　Nが大きければ $\frac{N-1}{2} + d\sqrt{N-1}$ のすぐ上の整数を使う

ただし $d = \begin{matrix} P=0.05 \text{に対して } 0.823 \\ P=0.01 \quad 〃 \quad 1.163 \end{matrix}$

V 表

中央相関係数の有意水準

点数	P=0.05	P=0.01	点数	P=0.05	P=0.01
8— 9	0.943	—	74— 75	0.255	0.326
10—11	0.820	1.000	76— 77	0.251	0.322
12—13	0.732	0.910	78— 79	0.248	0.317
14—15	0.667	0.831	80— 81	0.244	0.313
16—17	0.615	0.769	82— 83	0.241	0.309
18—19	0.573	0.718	84— 85	0.238	0.305
20—21	0.538	0.676	86— 87	0.235	0.301
22—23	0.509	0.640	88— 89	0.232	0.297
24—25	0.483	0.609	90— 91	0.229	0.294
26—27	0.461	0.582	92— 93	0.226	0.290
28—29	0.442	0.558	94— 95	0.223	0.287
30—31	0.425	0.537	96— 97	0.221	0.284
32—33	0.409	0.518	98— 99	0.218	0.281
34—35	0.395	0.501	100—101	0.216	0.278
36—37	0.382	0.485	110—111	0.205	0.264
38—39	0.371	0.471	120—121	0.196	0.252
40—41	0.360	0.457	130—131	0.187	0.241
42—43	0.350	0.445	140—141	0.180	0.232
44—45	0.341	0.434	150—151	0.173	0.224
46—47	0.332	0.423	160—161	0.167	0.216
48—49	0.325	0.413	170—171	0.162	0.209
50—51	0.317	0.404	180—181	0.157	0.203
52—53	0.310	0.396	200—201	0.149	0.192
54—55	0.304	0.388	220—221	0.141	0.183
56—57	0.298	0.380	240—241	0.135	0.175
58—59	0.292	0.373	260—261	0.129	0.167
60—61	0.286	0.366	280—281	0.124	0.161

点数	P=0.05	P=0.01	点数	P=0.05	P=0.01
62—63	0.281	0.359	300—301	0.120	0.155
64—65	0.276	0.353	320—321	0.116	0.150
66—67	0.272	0.347	340—341	0.112	0.146
68—69	0.267	0.342	360—361	0.109	0.141
70—61	0.263	0.336	380—381	0.106	0.137
72—63	0.259	0.331	400—401※	0.103	0.134

※ N が大きいときには $\frac{2}{N}+\frac{d}{N}$ を使う

ただし $d=$ P=0.05の時は 1.96
P=0.01 〃 2.58

VI 表　重相関係数の有意水準（3変量）

点数	P=0.05	P=0.01
16	0.789	0.935
20	0.689	0.820
24	0.617	0.737
28	0.564	0.675
32	0.521	0.625
36	0.486	0.584
40	0.458	0.551
44	0.423	0.522
48	0.412	0.497
52	0.394	0.475
56	0.378	0.456
60	0.363	0.439
64	0.350	0.424
68	0.338	0.410
72	0.328	0.397

点 数	P=0.05	P=0.01
76	0.318	0.385
80	0.309	0.375
84	0.301	0.365
88	0.293	0.356
92	0.286	0.347
96	0.279	0.339
100	0.273	0.332
120	0.247	0.301
140	0.227	0.277
160	0.211	0.258
180	0.198	0.242
200	0.187	0.229
240	0.170	0.208
280	0.156	0.191
320	0.146	0.178
360	0.137	0.168
400	0.129	0.159

VII表 分散比の表※ 5%点

偶然だけの機会で試行の5%を越える比の値が示してある。

		分子の自由度										
		1	2	3	4	5	6	7	8	9	10	12
分	1	161	200	216	225	230	234	237	239	241	242	244
	2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41
	3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.84	8.81	8.78	8.74
	4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91
	5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68
母	16	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.00	4.00
	17	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57
	18	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28
	19	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07
の	20	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91
自	11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	3.95	2.90	2.85	2.79
	12	4.75	3.88	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69
	13	4.67	3.80	3.41	3.18	3.02	2.92	2.83	2.77	2.71	2.67	2.60
	14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53
由	15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48
度	16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42
	17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38
	18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34
	19	4.38	3.52	3.13	2.90	2.74	2.62	2.54	2.48	2.42	2.38	2.31
	20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28

21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18
25	4.24	3.38	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16
26	4.22	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.30	2.25	2.20	2.13
28	4.20	3.34	2.95	2.71	2.56	2.44	2.36	2.29	2.24	2.19	2.12
29	4.18	3.33	2.93	2.70	2.54	2.43	2.35	2.28	2.22	2.18	2.10
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07	2.02	1.95
60	4.00	3.15	2.76	2.52	2.37	2.25	2.17	2.10	2.04	1.99	1.92
70	3.98	3.13	2.74	2.50	2.35	2.23	2.14	2.07	2.01	1.97	1.89
80	3.96	3.11	2.72	2.48	2.33	2.21	2.12	2.05	1.99	1.95	1.88
100	3.94	3.09	2.70	2.46	2.30	2.19	2.10	2.03	1.97	1.92	1.85
150	3.91	3.06	2.67	2.43	2.27	2.16	2.07	2.00	1.94	1.89	1.82
200	3.89	3.04	2.65	2.41	2.26	2.14	2.05	1.98	1.92	1.87	1.80
∞	3.84	3.00	2.60	2.37	2.21	2.09	2.01	1.94	1.88	1.83	1.75

(268)

※ Pergamon Press の好意により Introductory Statistics より引用

分子の自由度										
	16	20	24	30	40	50	60	75	100	∞
1	246	248	249	250	251	252	252	253	253	254
2	19.43	19.45	19.45	19.46	19.47	19.47	19.48	19.48	19.49	19.50
3	8.69	8.66	8.64	8.62	8.59	858	8.57	8.57	856	8.53
4	584	580	577	574	572	5.70	5.69	5.68	566	563
5	4.60	456	453	450	446	444	443	442	440	436
6	392	387	384	381	377	375	374	372	371	367
7	349	344	341	338	334	332	330	329	328	323
8	320	315	312	308	304	303	300	300	298	293
9	298	294	290	286	282	280	279	277	276	271
10	282	277	274	270	266	264	262	261	259	254
11	270	265	261	257	253	250	249	247	245	240
12	260	254	250	247	242	240	238	236	235	230
13	251	246	242	238	234	232	230	228	226	221
14	244	239	235	231	227	224	222	221	219	213
15	239	233	229	225	220	218	216	215	212	207
16	233	228	224	219	215	213	210	209	207	201
17	229	223	219	215	210	208	206	204	202	196
18	225	219	215	211	206	204	202	200	198	192
19	221	216	211	207	203	200	198	196	194	188
20	218	212	208	204	199	196	195	192	190	184

(692)

21	2.15	2.10	2.05	2.01	1.96	1.93	1.92	1.89	1.87	1.81
22	2.13	2.07	2.03	1.98	1.94	1.91	1.89	1.87	1.84	1.78
23	2.10	2.05	2.01	1.96	1.91	1.88	1.86	1.84	1.82	1.76
24	2.09	2.03	1.98	1.94	1.89	1.86	1.84	1.82	1.80	1.73
25	2.06	2.01	1.96	1.92	1.87	1.84	1.82	1.80	1.77	1.71
26	2.05	1.99	1.95	1.90	1.85	1.82	1.80	1.78	1.76	1.69
27	2.03	1.97	1.93	1.88	1.84	1.80	1.78	1.76	1.74	1.67
28	2.02	1.96	1.91	1.87	1.82	1.78	1.77	1.75	1.72	1.65
29	2.00	1.94	1.90	1.85	1.80	1.77	1.75	1.73	1.71	1.64
30	1.99	1.93	1.89	1.84	1.79	1.76	1.74	1.72	1.69	1.62
40	1.90	1.84	1.79	1.74	1.69	1.66	1.64	1.61	1.59	1.51
50	1.85	1.78	1.74	1.69	1.63	1.60	1.58	1.55	1.52	1.44
60	1.81	1.75	1.70	1.65	1.59	1.56	1.53	1.50	1.48	1.39
70	1.79	1.72	1.67	1.62	1.56	1.53	1.50	1.47	1.45	1.35
80	1.77	1.70	1.65	1.60	1.54	1.51	1.48	1.45	1.42	1.32
100	1.75	1.68	1.63	1.57	1.51	1.48	1.45	1.42	1.39	1.28
150	1.71	1.64	1.59	1.54	1.47	1.44	1.41	1.37	1.34	1.22
200	1.69	1.62	1.57	1.52	1.45	1.42	1.39	1.35	1.32	1.19
∞	1.64	1.57	1.52	1.46	1.39	1.35	1.32	1.28	1.24	1.00

第V表　分散比の表　　1％点

偶然だけの機会で試行の1％を越える比の値が示してある。

	分子の自由度										
	1	2	3	4	5	6	7	8	9	10	12
1	4052	4999	5403	5625	5764	5859	5928	5982	6022	6056	6106
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.38	99.39	99.40	99.42
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.34	27.23	26.05
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55	14.37
5	16.26	13.27	12.06	11.39	11.97	10.67	10.46	10.29	10.16	10.05	9.89
6	13.74	10.92	9.78	9.15	9.75	8.47	8.26	8.10	7.98	7.87	7.72
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47
8	11.26	8.65	7.59	7.01	7.63	6.37	6.18	6.03	5.91	5.81	5.67
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	4.35	5.26	5.11
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.71
11	9.65	7.20	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.40
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16
13	9.07	6.70	5.74	5.20	5.86	4.62	4.44	4.30	4.19	4.10	3.96
14	8.86	6.51	5.56	5.04	5.69	4.46	4.28	4.14	4.03	3.94	3.80
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.67
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.55
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.45
18	8.28	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.37
19	8.18	5.93	5.01	4.50	4.17	3.94	3.76	3.63	3.52	3.43	3.30
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23

21	8.02	5.78	487	437	4.04	381	3.64	3.51	3.40	331	3.17
22	7.94	5.72	4.82	431	3.99	3.76	359	3.45	3.36	3.26	3.12
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.07
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.03
25	7.77	5.57	4.68	4.18	3.86	3.63	3.46	3.32	3.22	3.13	2.99
26	7.72	553	464	4.14	3.82	359	3.42	3.29	3.18	3.09	296
27	7.68	5.49	460	4.11	3.78	356	3.39	3.26	3.15	3.06	2.93
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03	2.90
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09	3.00	2.87
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	284
40	7.31	5.18	431	383	3.51	329	3.12	2.99	2.89	280	266
50	7.17	5.06	4.20	3.72	3.41	3.18	3.02	288	2.78	270	256
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	250
70	7.01	4.92	4.08	3.60	3.29	3.07	2.91	2.77	2.67	259	245
80	6.96	4.88	4.04	356	3.25	3.04	2.87	2.74	2.64	255	241
100	6.90	4.82	3.98	3.51	320	2.99	2.82	2.69	2.59	251	236
150	6.81	4.75	3.91	3.44	3.14	292	2.76	2.63	2.53	244	230
200	6.76	4.71	3.88	3.41	3.11	290	2.73	2.60	2.50	241	2.28
∞	6.63	4.60	3.78	332	3.02	280	2.64	251	2.41	232	2.18

※ Pergamon Press の好意により Introductory Statistics より引用

分子の自由度										
	16	20	24	30	40	50	60	75	100	∞
1	6169	6209	6235	6261	6287	6302	6313	6323	6334	6366
2	99.44	99.45	99.46	99.47	99.47	99.48	99.48	99.49	99.49	99.50
3	26.83	26.69	26.60	26.50	26.41	26.35	26.32	26.27	26.23	26.12
4	14.15	14.02	13.93	13.84	13.74	13.69	13.65	13.61	13.57	13.46
5	9.68	9.55	9.47	9.38	9.29	9.24	9.20	9.17	9.13	9.02
6	7.52	7.40	7.31	7.23	7.14	7.09	7.06	7.02	6.99	6.88
7	6.27	6.16	6.07	5.99	5.91	5.85	5.82	5.78	5.75	5.65
8	5.48	5.36	5.28	5.20	5.12	5.06	5.03	5.00	4.96	4.86
9	4.92	4.81	4.73	4.65	4.57	4.51	4.48	4.45	4.41	4.31
10	4.52	4.40	4.33	4.25	4.16	4.12	4.08	4.05	4.01	3.91
11	4.21	4.10	4.02	3.94	3.86	3.80	3.78	3.74	3.70	3.60
12	3.98	3.86	3.78	3.70	3.62	3.56	3.54	3.49	3.46	3.36
13	3.78	3.66	3.59	3.51	3.42	3.37	3.34	3.30	3.27	3.16
14	3.62	3.50	3.43	3.35	3.27	3.21	3.18	3.14	3.11	3.00
15	3.48	3.37	3.29	3.21	3.13	3.07	3.05	3.00	2.97	2.87
16	3.37	3.26	3.18	3.10	3.02	2.96	2.93	2.89	2.86	2.75
17	3.27	3.16	3.08	3.00	2.92	2.86	2.83	2.79	2.76	2.65
18	3.19	3.08	3.00	2.91	2.84	2.78	2.75	2.71	2.68	2.57
19	3.12	3.00	2.92	2.84	2.76	2.70	2.67	2.63	2.60	2.49
20	3.05	2.94	2.86	2.78	2.69	2.63	2.61	2.56	2.53	2.42

21	2.99	2.88	2.80	2.72	2.64	2.55	2.58	2.36
22	2.94	2.83	2.75	2.67	2.58	2.50	2.53	2.31
23	2.89	2.78	2.70	2.62	2.54	2.45	2.48	2.26
24	2.85	2.74	2.66	2.58	2.49	2.40	2.44	2.21
25	2.81	2.70	2.62	2.54	2.45	2.36	2.40	2.17
26	2.77	2.66	2.58	2.50	2.42	2.33	2.36	2.13
27	2.74	2.63	2.55	2.47	2.38	2.29	2.33	2.10
28	2.71	2.60	2.52	2.44	2.35	2.26	2.30	2.06
29	2.68	2.57	2.49	2.41	2.32	2.23	2.27	2.03
30	2.66	2.55	2.47	2.39	2.30	2.21	2.24	2.01
40	2.49	2.37	2.29	2.20	2.11	2.02	2.05	1.80
50	2.39	2.26	2.18	2.10	2.00	1.91	1.94	1.68
60	2.32	2.20	2.12	2.03	1.94	1.84	1.87	1.60
70	2.28	2.15	2.07	1.98	1.88	1.79	1.82	1.53
80	2.24	2.11	2.03	1.94	1.84	1.74	1.78	1.49
100	2.19	2.06	1.98	1.89	1.79	1.69	1.73	1.43
150	2.12	2.00	1.91	1.83	1.72	1.62	1.66	1.33
200	2.09	1.97	1.88	1.79	1.69	1.58	1.62	1.28
∞	1.99	1.88	1.79	1.70	1.59	1.47	1.52	1.00

IX 表※ 与えられた推定偏差 t を越える試行の百分率 を与える表

自由度	偏差が越える試行の百分率							
	50	25	10	5	25	1.0	0.5	0.1
1	1.00	2.41	6.31	12.71	22.45	63.66	127.32	636.62
2	0.82	1.60	2.92	4.30	6.20	9.92	14.09	31.60
3	0.76	1.42	2.35	3.18	4.18	5.84	7.45	12.94
4	0.74	1.34	2.13	2.78	3.50	4.60	5.60	8.61
5	0.73	1.30	2.01	2.57	3.16	4.03	4.77	6.86
6	0.72	1.27	1.94	2.45	2.97	3.71	4.32	5.96
7	0.71	1.25	1.89	2.36	2.84	3.50	4.03	5.40
8	0.71	1.24	1.86	2.30	2.75	3.35	3.83	5.04
9	0.70	1.23	1.83	2.26	2.68	3.25	3.69	4.78
10	0.70	1.22	1.81	2.23	2.63	3.17	3.58	4.59
11	0.70	1.21	1.80	2.20	2.59	3.11	3.50	4.44
12	0.70	1.21	1.78	2.18	2.56	3.05	3.43	4.32
13	0.69	1.20	1.77	2.16	2.53	3.01	3.37	4.22
14	0.69	1.20	1.76	2.14	2.51	2.98	3.32	4.14
15	0.69	1.20	1.75	2.13	2.49	2.95	3.29	4.07
16	0.69	1.19	1.74	2.12	2.47	2.92	3.25	4.01
17	0.69	1.19	1.74	2.11	2.46	2.90	3.22	3.96
18	0.69	1.19	1.73	2.10	2.44	2.88	3.20	3.92
19	0.69	1.19	1.73	2.09	2.43	2.86	3.17	3.88
20	0.69	1.18	1.72	2.09	2.42	2.84	3.15	3.85
22	0.69	1.18	1.72	2.07	2.40	2.82	3.12	3.79
24	0.68	1.18	1.71	2.06	2.39	2.80	3.09	3.75
26	0.68	1.17	1.71	2.06	2.38	2.78	3.07	3.71
28	0.68	1.17	1.70	2.05	2.37	2.76	3.05	3.67
30	0.68	1.17	1.70	2.04	2.36	2.75	3.03	3.65
40	0.68	1.16	1.68	2.02	2.33	2.70	2.97	3.55
50	0.68	1.16	1.68	2.01	2.31	2.68	2.93	3.50
60	0.68	1.16	1.67	2.00	2.30	2.66	2.91	3.46
∞	0.67	1.16	1.64	1.96	2.24	2.58	2.81	3.29

※ Pergamon Press の好意により Introductory Statistics より引用

X表　積率相関係数の有意水準

点数	P=0.05	P=0.01
3	0.9969	0.99988
4	0.950	0.990
5	0.878	0.959
6	0.811	0.917
7	0.754	0.875
8	0.707	0.834
9	0.666	0.798
10	0.632	0.765
11	0.602	0.735
12	0.576	0.708
13	0.553	0.684
14	0.532	0.661
15	0.514	0.641
16	0.497	0.623
17	0.482	0.606
18	0.468	0.590
19	0.456	0.575
20	0.444	0.561
22	0.423	0.537
24	0.404	0.515
26	0.388	0.496
28	0.374	0.478
30	0.361	0.463
35	0.334	0.430
40	0.312	0.403
50	0.279	0.361
60	0.254	0.330
70	0.235	0.306
80	0.220	0.286
90	0.207	0.270
100	0.197	0.257
120	0.179	0.234
150	0.160	0.210
200	0.139	0.182
300	0.113	0.149
400	0.098	0.129

XI表　Z変換表

r	z	z	r
0.00	0.000	0.00	0.000
0.05	0.050	0.05	0.050
0.10	0.100	0.10	0.100
0.15	0.151	0.15	0.149
0.20	0.203	0.20	0.197
0.25	0.255	0.25	0.245
0.30	0.310	0.30	0.291
0.35	0.365	0.35	0.336
0.40	0.424	0.40	0.380
0.45	0.485	0.45	0.422
0.50	0.549	0.50	0.462
0.55	0.618	0.55	0.501
0.60	0.693	0.60	0.537
0.64	0.758	0.65	0.572
0.68	0.829	0.70	0.604
0.72	0.908	0.75	0.635
0.76	0.996	0.80	0.664
0.80	1.099	0.85	0.691
0.82	1.157	0.90	0.716
0.84	1.221	0.95	0.740
0.86	1.293	1.00	0.762
0.88	1.376	1.10	0.800
0.90	1.472	1.20	0.834
0.91	1.528	1.30	0.862
0.92	1.589	1.40	0.885
0.93	1.658	1.50	0.905
0.94	1.738	1.60	0.922
0.95	1.832	1.80	0.947
0.96	1.946	2.00	0.964
0.97	2.092	2.20	0.976
0.98	2.298	2.40	0.984
0.985	2.443	2.60	0.989
0.990	2.647	2.80	0.993
0.995	2.994	3.00	0.995

XII 表　n=3～12 に対する 3 次までの直行多項式

3 ξ_1　ξ_2	4 ξ_1　ξ_2　ξ_3	5 ξ_1　ξ_2　ξ_3	6 ξ_1　ξ_2　ξ_3	7 ξ_1　ξ_2　ξ_3
−1　+1 0　−2 +1　+1	−3 +1 −1 −1 −1 +3 +1 −1 −3 +3 +1 +1	−2 +2 −2 −1 −1 +2 0 −2 0 +1 −1 −2 +2 +2 +1	−5 +5 −5 −3 −1 +7 −1 −4 +4 +1 −4 −4 +3 −1 −7 +5 +5 +5	−3 +5 −1 −2 0 +1 −1 −3 +1 0 −4 0 +1 −3 −1 +2 0 −1 +3 +5 +1
2　6 X − 2 $3X^2-12X$ +10	20　4　20 $2X-5$ X^2-5X+5 $\frac{1}{3}(10X^3-75X^2+167X-105$	10　14　13 X − 3 X^2-6X+7 $\frac{1}{6}(5X^3-45X^2+118X-84)$	70　84　180 $2X-7$ $\frac{1}{2}(3X^2-21X+28)$ $\frac{1}{6}(10X^3-105X^2+317X-252)$	28　84　6 X − 4 $X^2-8X+12$ $\frac{1}{6}(X^3-12X^2+103X-99)$

8 ξ_1　ξ_2　ξ_3	9 ξ_1　ξ_2　ξ_3	10 ξ_1　ξ_2　ξ_3	11 ξ_1　ξ_2　ξ_3	12 ξ_1　ξ_2　ξ_3
−7 +7 −7 −5 +1 +5 −3 −3 +7 −1 −5 +3 +1 −5 −3 +3 −3 −7 +5 +1 −5 +7 +7 +7	−4 +28 −14 −3 +7 +7 −2 −8 +13 −1 −17 +9 0 −20 0 +1 −17 −9 +2 −8 −13 +3 +7 −7 +4 +7 +14	−9 +6 −42 −7 +2 +14 −5 −1 +35 −3 −3 +31 −1 −4 +12 +1 −4 −12 +3 −3 −31 +5 −1 −35 +7 +2 −14 +9 +6 +42	−5 +15 −30 −4 +6 +6 −3 −1 +22 −2 −6 +23 −1 −9 +14 0 −10 0 +1 −9 −14 +2 −6 −23 +3 −1 −22 +4 +6 −6 +5 +15 +30	−11 +55 −33 −9 +25 +3 −7 +1 +21 −5 −17 +25 −3 −29 +19 −1 −35 +7 +1 −35 −7 +3 −29 −19 +5 −17 −25 +7 +1 −21 +9 +25 −3 +11 +55 +33
168　168　264 $2X-9$ $X^2-9X+15$ $\frac{1}{3}(2X^3-27X^2+103X-99)$	60　2772　990 X − 5 $3X^2-30X+55$ $\frac{1}{6}(5X^3-75X^2+316X-330$	330　132　8580 $2X-11$ $\frac{1}{3}(X^2-11X+22)$ $\frac{1}{6}(10X^3-165X^2+761X-858)$	110　858　4290 X − 6 $X^2-12X+26$ $\frac{1}{6}(5X^3-90X^2+451X-546)$	572　12012　6148 $2X-13$ $3X^2-39X+91$ $\frac{1}{3}(2X^3-39X^2+211X-273)$

XIII 表　x^2 分　布　表

自由度 n	x^2 が越える試行の百分率							
	50	25	10	5	2.5	1.0	0.5	0.1
1	0.45	1.32	2.70	3.84	5.02	6.63	7.88	10.83
2	1.39	2.77	4.60	5.99	7.38	9.21	10.60	13.82
3	2.36	4.11	6.25	7.81	9.35	11.34	12.84	16.27
4	3.36	5.38	7.78	9.49	11.14	13.28	14.86	18.46
5	4.35	6.62	9.24	11.07	12.83	15.09	16.75	20.52
6	5.35	7.84	10.64	12.59	14.45	16.81	18.55	22.46
7	6.34	9.04	12.02	14.07	16.01	18.48	20.28	24.32
8	7.34	10.22	13.36	15.51	17.53	20.09	21.96	26.12
9	8.34	11.39	14.68	16.92	19.02	21.67	23.59	27.88
10	9.34	12.55	15.99	18.31	20.48	23.21	25.19	29.59
11	10.34	13.70	17.28	19.68	21.92	24.72	26.76	31.26
12	11.34	14.84	18.55	21.03	23.34	26.22	28.30	32.91
13	12.34	15.98	19.81	22.36	24.74	27.69	29.82	34.53
14	13.34	17.12	21.06	23.68	26.12	29.14	31.32	36.12
15	14.34	18.24	22.31	25.00	27.49	30.58	32.80	37.70
16	15.34	19.37	23.54	26.30	28.84	32.00	34.27	39.25
17	16.34	20.49	24.77	27.59	30.19	33.41	35.72	40.79
18	17.34	21.60	25.99	28.87	31.53	34.80	37.16	42.31
19	18.34	22.72	27.20	30.14	32.85	36.19	38.58	43.82
20	19.34	23.83	28.41	31.41	34.17	37.57	40.00	45.32
22	21.34	26.04	30.81	33.92	36.78	40.29	42.80	48.27
24	23.34	28.24	33.20	36.42	39.36	42.98	45.56	51.18
26	25.34	30.43	35.56	38.88	41.92	45.64	48.29	54.05
28	27.34	32.62	37.92	41.34	44.46	48.28	50.99	56.89
30※	29.34	34.80	40.26	43.77	46.98	50.89	53.67	59.70

※　自由度が 30 より大きい時には x は x^2-2n-1 が近似的に正規分布するということを使う。

例：

自由度 45 の x^2 の値が 74.23 であるとする。$2\times74.23-2\times45-1$ は近似的に正規分布をする。IX表を使えばこのように高い値が現るのは、偶然だけの現象で六行の 1 %以下であることが分る。

「補　註」

Associated Measurment は、連関のノンパラメントリック検定法を含む図形的解析法の即から始まつている。こゝで行つている偏相関についての記述には批判の余地がある。　正の象限に含まれる観測値のメジアンの点で散分図を二分し、その偏台Pを求めたもの　二変量X,Y間のを著者は定義し、次のように置いている

$$\phi_{xy} = 4P - 1$$

無限母集団では ϕ_{xy} の値は、完全相関の時には±1、無相関の時には0である。又XとYが無相関の場合には、ϕ_{xy} の分布は母集団の形と関係がなく、したがつて独立性に関するノンパラメトリック検定を行うことができる。二変量正規母集団ではW.F.Sheppard の公式から、ϕ_{xy} は積率相関係数 Product moment coefficient P_{xy} と関係がある。

$$P_{xy} = \sin\left(\frac{\pi\phi_{xy}}{2}\right) \qquad (2)$$

次にQuenouille氏は偏中央相関係数を次式で定義している。

$$\phi_{xy.z} = \frac{\phi_{xy} - \phi_{xy}\phi_{yz}}{(1-\phi_{xy}^2\ \phi_{yz}^2)^{\frac{1}{2}}} \qquad (3)$$

この式の由来は示していないし、一貫してこの式が使われてはいるか、これは積率相関係数との類似性から導びかれた次式の誤植ではないかと思う。

$$\phi_{xy.z} = \frac{\phi_{xy} - \phi_{xz}\ \phi_{yz}}{(1-\phi_{xz}^2)^{\frac{1}{2}}(1-\phi_{yz}^2)^{\frac{1}{2}}} \qquad (4)$$

(3)(4)式はいづれも完全に誤つた結果を導びくこともありうる。特殊な三変量正規母集団を考えてみると、この点がはつきりする例として、$x=y+z$ を考えてみる。この場合 y z は夫々独立に正規分布をするものとする。したがつて、Zを固定すれば、Xはyと完全相関がある。しかし $\phi_{xyz}=\phi_{xy}=\frac{1}{4}$ $\phi_{xyz}=13$ したがつて偏相関 ϕ の値は、Zを固定させても相関は変らないか、変つたとしても僅かであることを示している。この場合、Zを固定させた時

XとYとが独立であれば、もつとはつきりする。この時には、ϕ_{xyz}, ϕ_{xyz} はかならずしも0とはならないが、明らかに ϕ_{xy} より常に小である。しかし ϕ_{xyz} ϕ_{xyz} の0になることは P_{xyz} か小であることを意味しているのではない。例えば $P_{xy}=0.696$ $P_{xy}=P_{yz}$ $z=0.891$ とすれば $\phi_{xyz}=\phi_{xyz}=0$ であるにもかゝわらず P_{xyz} は　$\phi_{xyz}=\phi_{xyz}=0$ となつて−1に近接させることができる。偏相関 ϕ を無批判に使うと誤りを犯し易いことをはつきりさせるために説明したのである。

同様にM.G.Kendall 教授が彼の順位相関法 Rank Correlation Methode で示している偏順位相関係数についても同様なことがえる。正規母集団の標本から導びかれる順位の理論は、一般的順位理論の稍く特殊な場合 Kendall の偏相関係数 T は連続母集団の形に誤つた概念を与えるということは注目に値する。正規母集団からの標本の T の期待は (2)　と同じ形で与えられるから、ϕ の場合と同じことが起る。

$$P = \sin\frac{\pi T}{2}$$

母集団が正規である場合には、偏相関を計算する前に(2)又は(5)式で ϕ 又は T を変形することにより、この困難性はもちろん避けることができる。一般的方法としてはこれに対して若干の反対もある。