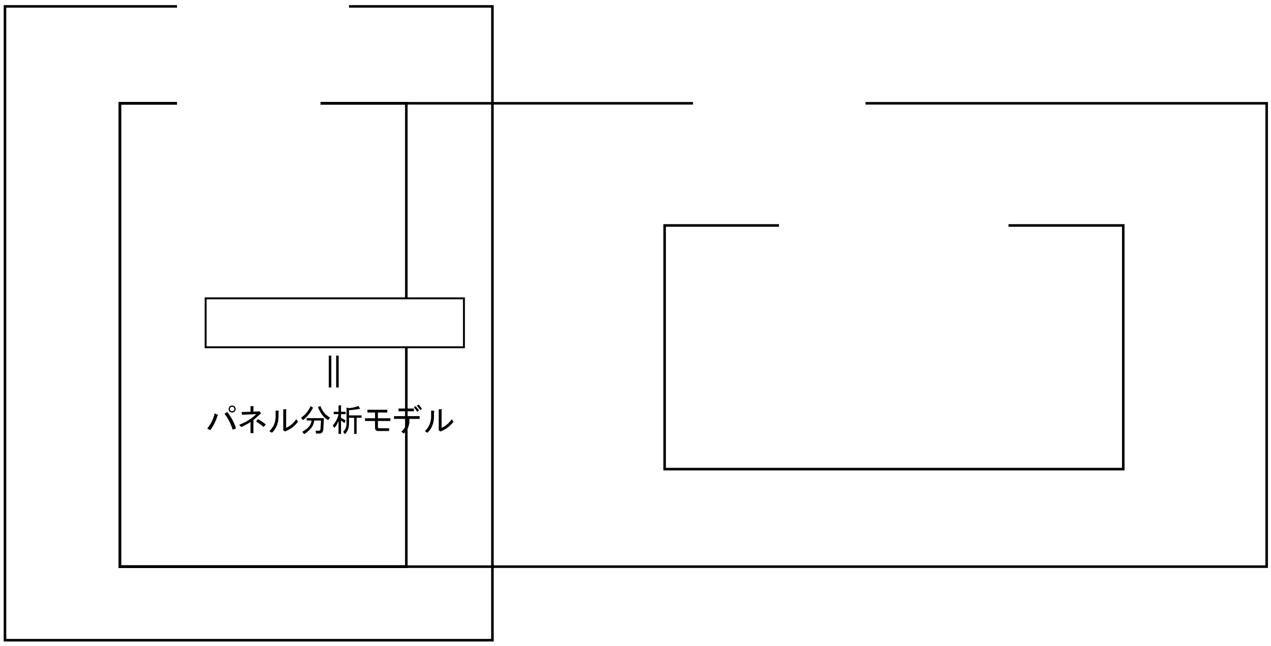
# 比較統計システム論 講義録（2000 年度）

## 分散分析と回帰モデルの関係




## 共分散分析の考え方

共分散分析は回帰分析と分散分析を同時に行う手法である。構造モデルの一般形

$$Y_{ij}=\mu+a_j+x_{ij}b+x_{ij}(ab)_j+e_{ij}$$

となる。i∈Iは生徒のセット（I＝4）、j∈Jは教え方のセット（J＝2）とし、$Y_{ij}$はj群に振り分けられたi番目の生徒のテストの得点、$a_j$はj群の教え方と以前のテストの交互作成を示すとしよう。これをマトリックスで示すと、

$$\begin{pmatrix} Y_{11} \\ Y_{21} \\ Y_{31} \\ Y_{41} \\ Y_{12} \\ Y_{22} \\ Y_{23} \\ Y_{24} \end{pmatrix} = \begin{pmatrix} 1 & 1 & & x_{11} & x_{11} & \\ 1 & 1 & & x_{21} & x_{21} & \\ 1 & 1 & & x_{31} & x_{31} & \\ 1 & 1 & & x_{41} & x_{41} & \\ 1 & & 1 & x_{12} & & x_{12} \\ 1 & & 1 & x_{22} & & x_{22} \\ 1 & & 1 & x_{32} & & x_{32} \\ 1 & & 1 & x_{42} & & x_{42} \end{pmatrix} \begin{pmatrix} \mu \\ a_1 \\ a_2 \\ b \\ (ab)_1 \\ (ab)_2 \end{pmatrix} + \begin{pmatrix} e_{11} \\ e_{21} \\ e_{31} \\ e_{41} \\ e_{12} \\ e_{22} \\ e_{23} \\ e_{24} \end{pmatrix}$$

計画行列の左側は分散分析、真ん中は回帰分析、右側は両方の交互作用を表現している。また、構造モデルを

とみなすと、b＋(ab)$_j$と$\mu$＋a$_j$は、それぞれ第j群の回帰係数と切片と解釈することができる。つまり、群ごとに別々の回帰直線を当てはめていることになる。

　ただし、独立した回帰分析を群ごとに実施した解に一致するわけではない。構造模型には誤差項が 1 つしかないので、その分散は群間で共通という制約の下で解を求めているからである。

　分散分析表は次のようになる。

| 変動要因 | 自由度(df) | EMS | 統計量(F) |
|---|---|---|---|
| A | a－1 | | |
| X | 1 | | |
| AX | a－1 | $\sigma^2_e+n\sigma^2_{ab}$ | $MS_{AX}$／$MS_e$ |
| e | (n－2)a | $\sigma^2_e$ | |
| 全体(T) | na－1 | | |

特徴的なことは、交互作用の項しか検定していないことである。もちろん帰無仮説は(ab)$_j$＝0である。もし棄却されたら群ごとに回帰分析を行う。

　棄却されない場合、あるいは(ab)$_j$の推定値がすべて0に近い場合は、とりあえず(ab)$_j$＝0を仮定し、構造モデルを

のように制約して分析を進める。

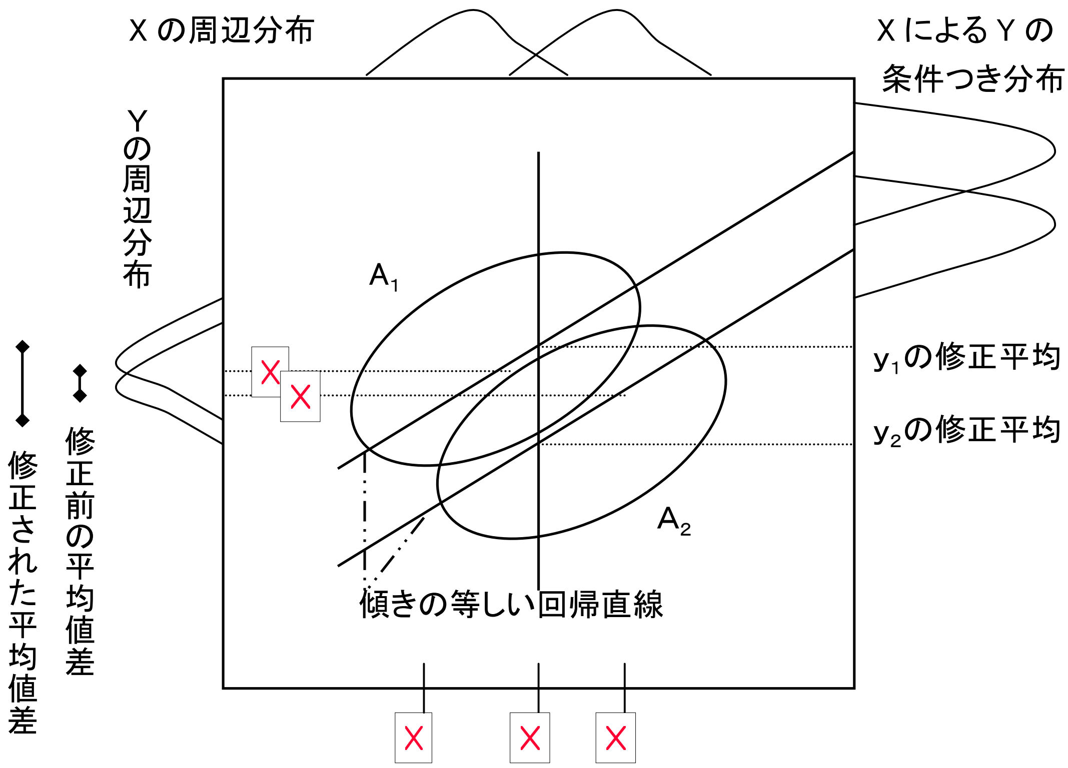
　bと$\mu$＋a$_j$を回帰係数と切片と解釈できる。

$$\begin{pmatrix} Y_{11} \\ Y_{21} \\ Y_{31} \\ Y_{41} \\ Y_{12} \\ Y_{22} \\ Y_{23} \\ Y_{24} \end{pmatrix} = \begin{pmatrix} 1 & 1 & & x_{11} \\ 1 & 1 & & x_{21} \\ 1 & 1 & & x_{31} \\ 1 & 1 & & x_{41} \\ 1 & & 1 & x_{12} \\ 1 & & 1 & x_{22} \\ 1 & & 1 & x_{32} \\ 1 & & 1 & x_{42} \end{pmatrix} \begin{pmatrix} \mu \\ a_1 \\ a_2 \\ b \end{pmatrix} + \begin{pmatrix} e_{11} \\ e_{21} \\ e_{31} \\ e_{41} \\ e_{12} \\ e_{22} \\ e_{23} \\ e_{24} \end{pmatrix}$$

回帰直線の傾きが群間ですべて等しいという仮定は強いものであり、一見、不自然な仮定のようにも感じられる。しかし回帰分析と分散分析を効果的に同時に提供するためには必要な仮定である。

# 共分散分析のモデル

2群の集団$A_1$と$A_2$について変量xと変量yの測定値を表現する（同時分布）。




Yの周辺分布を見ると、$A_1$の方が平均値が大きい。$A_1$と$A_2$の平均値の差はそれ程大きくはなく、2 つの分布の分離は明瞭ではないから検定力は大きくない。

Xの周辺分布は$A_2$の方が平均値が大きい。$A_1$と$A_2$に傾きの等しい回帰直線を引き、その上で決まってくる$\bar{x}$におけるYの平均値を見る。これを<u>修正平均</u>というが、図表の右側に破線で示したように、予測値の差は元の平均値の差よりも大きくなっていることがわかる。また$\bar{x}$が与えられた場合のyの分布を図の右上に示したが、これを<u>xによるyの条件つき分布</u>という。

条件つき分布を考察すると、yの（周辺）分布と比較して平均値の差も大きくなっているが、そればかりではなく郡内の分散も小さくなっている。その結果、条件つき分布の分離は明瞭になり、この分布を利用した方が周辺分布を利用するよりも検定力は高くなりそうである。

変量yの群別の平均値の差がそれほど大きくなく、単独で分散分析を行った場合に、仮に有意差が見出されなくても、変量xを組み込んで共分散分析を行った場合に有意差が見出される可能性はある。このような変量xを<u>補助変数</u>とか<u>補助因数</u>という。

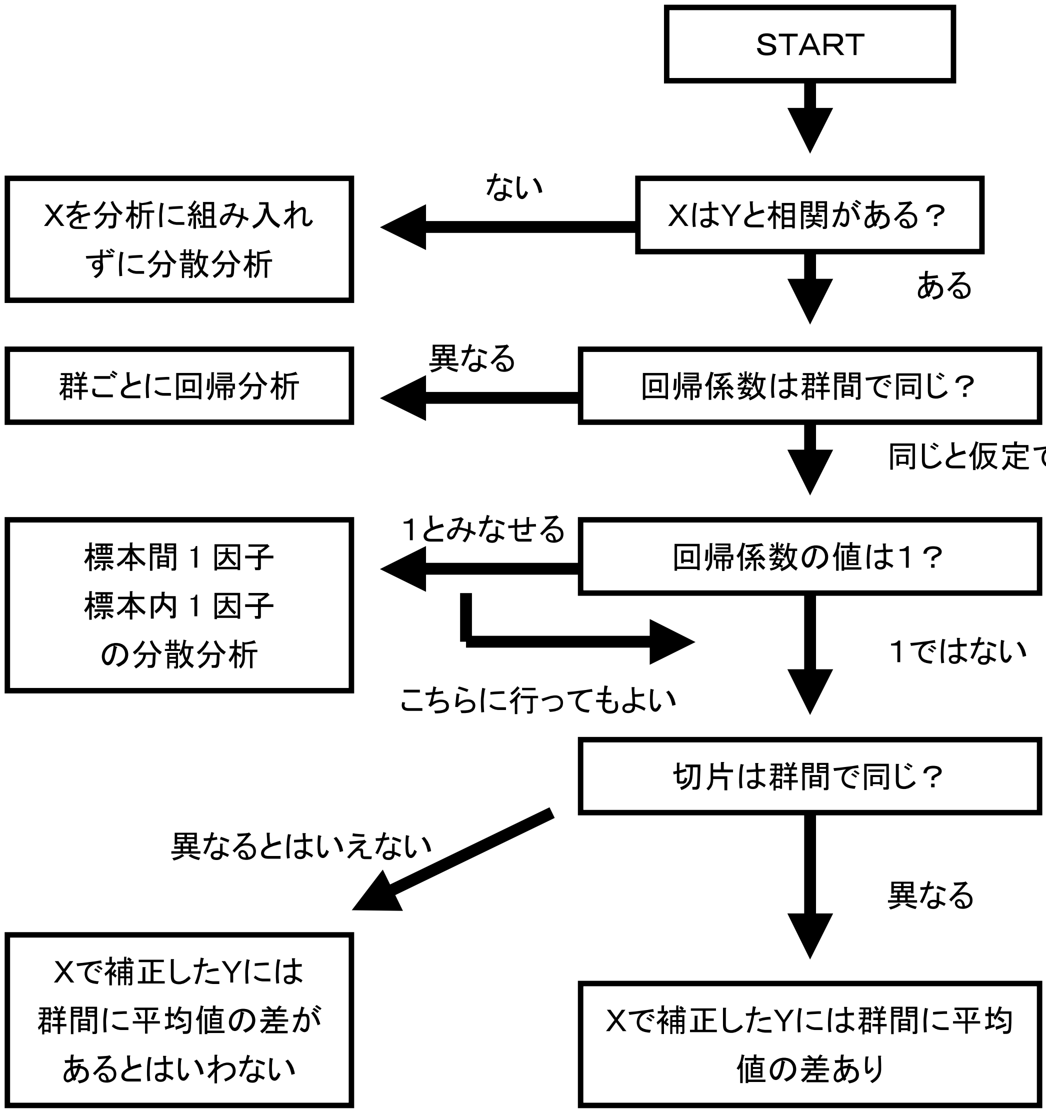
回帰直線の傾きが群間ですべて等しいという仮定が必要になる理由は何だろうか。[1]回

---

[1] これは比較静学で用いられる「他の条件を一定として（ceteris paribus）」という条件を用い

帰係数が共通であれば、その差はxの値によらずに一定である。つまり修正平均の差は補助変数xの値と関係なくなる。これがxとAの交互作用が有意でないということの意味である。簡単に言えば、xがyに与える影響は全ての群に対して同じであるという仮定である。すなわち、yの変動の要因からxの影響をはずすという意味がある。というわけで「回帰直線の傾きが群間ですべて等しい」という仮定は、群間の差を検定したいという目的上はむしろ自然な仮定なのである。

共分散分析の分析手続き




（出典）『違いを見ぬく統計学』、豊田秀樹（1994）、講談社ブルーバックス、図表８－４。

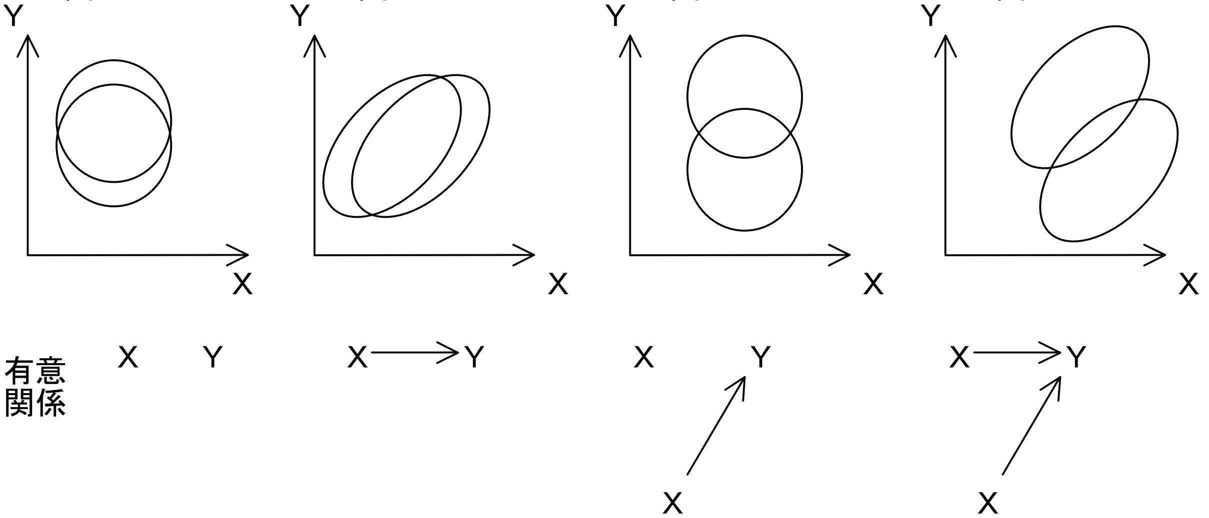
---

ていると観ることができる。計量経済学では、ある説明変数を加えることで被説明変数の変動に対するその説明変数の影響をコントロールするという。これもある変数の影響を取り除くという意味で類似した考え方である。

共分散分析には特性値Y、因子A、補助変数Xの 3 種類の変数が登場するが、分析結果を解釈するときには、3 つの変数の意味的な関係に留意する必要がある。特に因子と補助変数の関係は重要であり、結果の解釈を容易にするためには因子が補助変数に影響を及ぼさないことが望ましい。

（1）因子が補助変数に影響を与えない場合

図1　　図 2　　図3　　図4



有意関係

図 1 は、Y, X, Aがいずえrも関係していない場合であり、補助変数も因子も特性値に影響しなければ、相関のない 2 つの同時分布がほぼ重なり合う。図 1 ではXとYが相関を持たないから共分散分析は実行されず、1 因子の分散分析の結果も有意ではなくなる。

図 2 は補助変数のみが特性値に影響する場合であり、共分散分析を実行しても有意差は見出されない。

図 3 は、図 1 とは逆に、因子のみが特性値に影響している場合である。これは特性値に影響すると考えられた補助変数が無効であった場合に相当する。

図 4 は、共分散分析の典型例であり、特性値が因子と補助変数の両方から影響を受けている。

因子が補助変数に影響しないということは、X軸に関して 2 つの同時分布がほぼ同じであることを意味しており、それはまた補助変数に関して因子の水準とは無関係に一つの母集団を仮定できることを意味している。

（2）因子が補助変数に影響を与える場合

図5

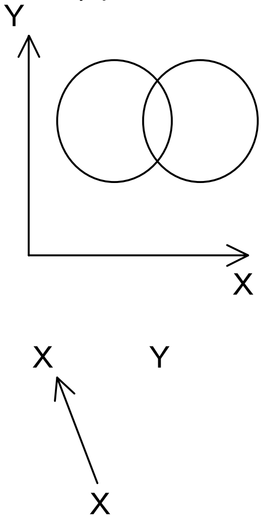



図 6

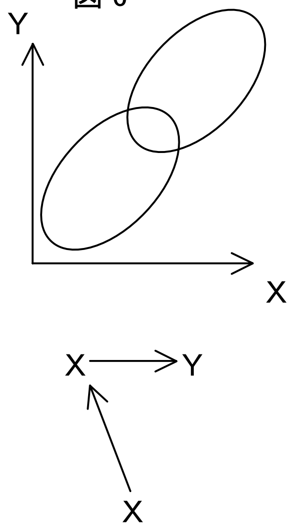



図 7

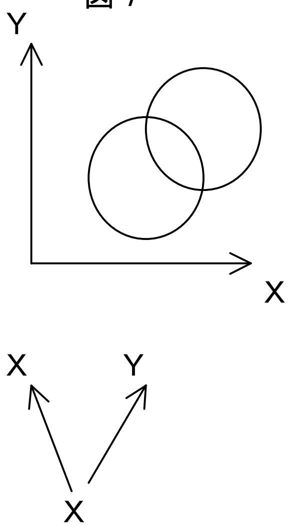



図 8

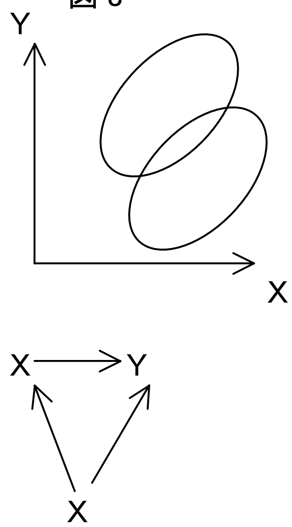



図 1～4 と図 5～8 の明らかな違いは、Xの群別の標本平均の値が明らかに離れていることである。

図 5～8 の中で共分散分析が実行できないのは図 5 である。図 6 では、特性値Yだけで分散分析を行うと群間に有意差が見出されるが、共分散分析を行うと有意差は見出されなくなる。これはA→Yの関係がないということではなく、Aの効果はXを経由してYに表れると考えられるということである。図 8 は特性値だけで分散分析を行っても、共分散分析を行っても群間に有意差が見出される。この場合は、Yの値Iに対して因子Aだけではなく、補助変数Xも影響を与えていることを意味する。図7は特性値だけで分散分析を行っても、共分散分析を行っても群間に有意差が見出されるが、共分散分析では傾きがなくなってしまう。これはYとXが直接関係しているわけではなく、背後に存在する因子のため見かけ上の相関が生じているためであると解釈できる。これを<u>疑相関</u>といい、YとXからAの影響を取り除いたときの偏相関がないという自体に相当する。