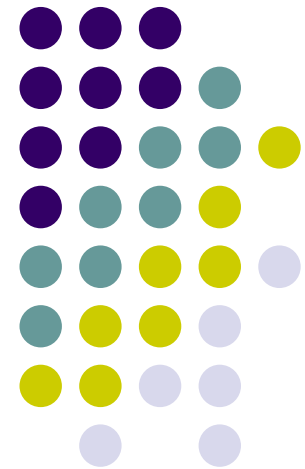


共分散分析

データ解析演習 2008/1/16

教育学研究科 M2 上野泰治





発表の流れ

- 心理学実験における変動
- 共分散分析のメリット
- 共分散分析の視覚的理解
- 共分散分析を n 要因に拡張する。
 - 理論的な話や気をつけることは、昨年の発表資料を参照してください。



心理学における分散

- 例えば、コーヒーを飲んだら成績が上がる、ということの研究をしたい。
- 理想的な結果は、右のようになる。

表1 両群のテスト得点(n=10)

飲まない群	飲む群
90	100
90	100
90	100
90	100
90	100

群内でバラつき(分散)がない！

群間でバラつき(分散)がある！



心理学における分散

- しかし現実は次のようになる。
- なぜバラつきがある？
 - 元々の成績・やる気・コーヒーアレルギー・腹痛
- 交絡(confounding)
- 剰余変数(extraneous variables)考慮の必要性。

表2 両群のテスト得点(n=10)

飲まない群	飲む群
90	100
95	100
92	80
85	90
98	95

群内でバラつき(分散)がある！
群間でバラつき(分散)が少ない！



剰余変数の統制

- 例えば、“やる気”という変数を考慮する。
- 元々やる気の高い人ほど成績が高い、という正の相関があったとしたら、次の状況では困る。

表3 両群のやる気度(n=10)

飲まない群	飲む群
1	5
1	5
1	5
1	5
1	5





剰余変数の統制方法

- 除去化 : コーヒーアレルギーの治療
- 独立変数化 : すでに分散分析にて学習済みです。
- 恒常化(一定化) : やる気ゼロの学生だけを集めよう!
- 無作為化(確率化) : 偏る事のないように偶然に任せよう!
- ブロック化 : やる気の低い人同士・高い人同士で比べよう!
- 統計的調整化 : 統計量に及ぼす剰余変数の影響を計算しよう!



統計的調整法 ~ 共分散分析

- 剰余変数が従属変数に与える影響は、誤差として扱われる。
- 剰余変数がどれだけ誤差に影響を与えているかが見積もり可能ならば、誤差を積極的に調整できる。
- 言い換えると、誤差を剰余変数がどの程度説明(予測)出来るか? という問いと等しい。



分散説明率

- 誤差とは、実験者が意図していない要因による分散のことである。
- では、その分散が、剰余変数の分散によってどの程度説明できるか？
- 変数 X の値から変数 Y の値を何%予測できるかという数値を、分散説明率と呼ぶ。
- 分散説明率は、変数 X と変数 Y の相関係数の2乗に等しい。



相関係数と分散説明率

- 剰余変数と従属変数の相関係数が測定できるなら、誤差変動の分散を相関係数の2乗分説明することができる。
- 言い換えると、相関係数の2乗分は、もはや未知の効果ではない。説明できる効果であり、剰余変数として放って置かず、積極的に誤差変動の中から除去すべき。
- 剰余変数と呼ばれていたものは、共変量と呼ばれる。



共分散分析

- 共分散分析は、共変量によって説明できる変動を、誤差変動の中から除去することにより、分析の精度を上げる手法である。
 - よって、統計的調整化と呼ばれる。



メリット・デメリット

- 共変量が複数あってブロック化が困難な時にも使用できる。
- データ取得後でも交絡の統制が可能。
- 交絡の防止のみならず、誤差変動を縮小し、分析の精度を上げるという積極的な意義をもつ。
- 結果の一般化可能性は、「共変量が一定の値の時」という範囲に収まる。



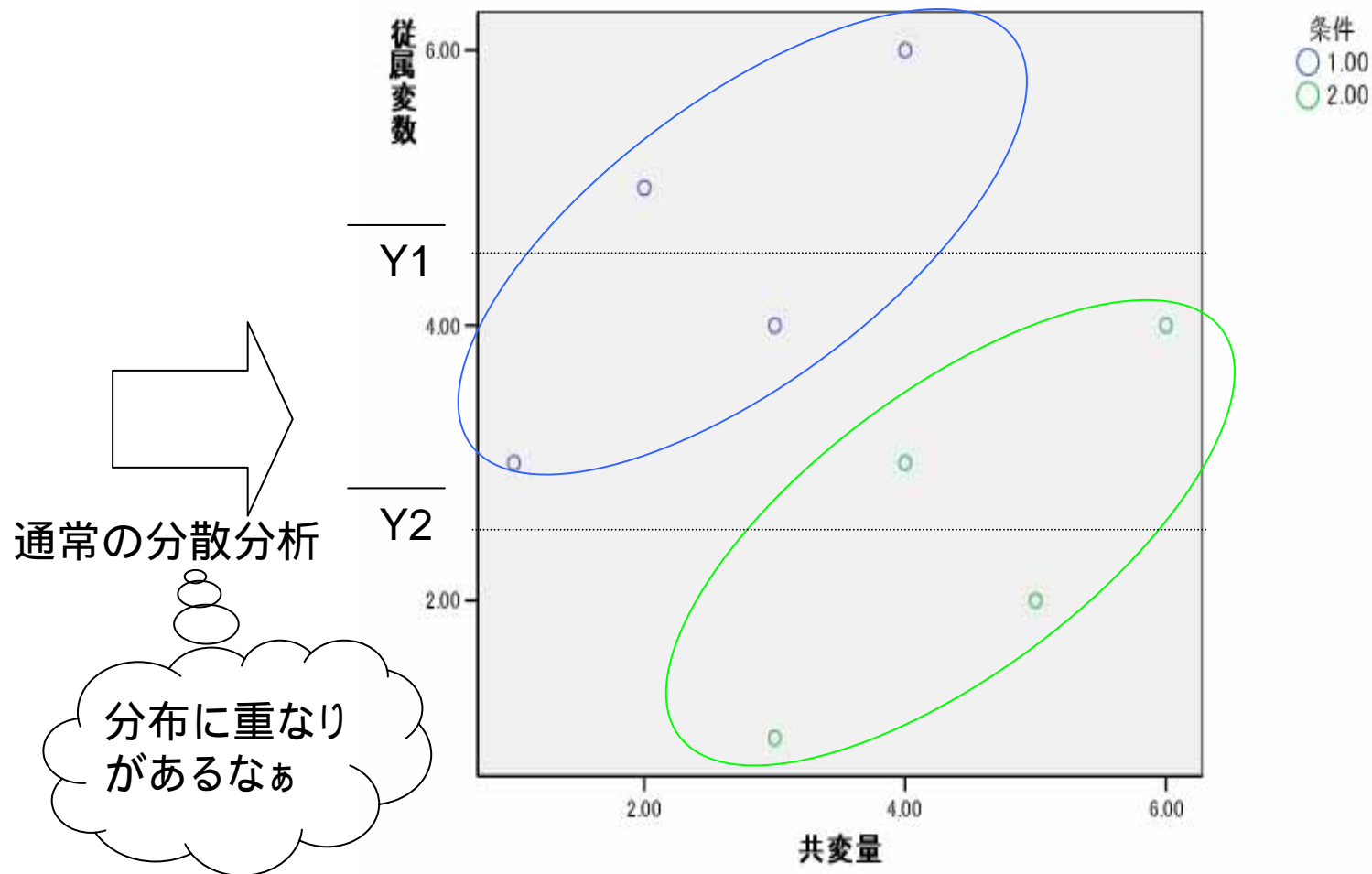
共分散分析の数学的理解

- 共分散分析は、共変量によって説明できる変動を誤差の中から取り除く手法。
- 全ての被験者の共変量が同一の値だったら、という状況を(統計により)想定して分散分析を行う手法ともいえる。
- 全ての被験者の共変量がある値の時、各被験者の従属変数の値はいくつになるかを求める。
- “ある値”には、各条件の共変量の平均値を求め、それらの平均値を採択する。

通常分散分析の視覚的理解



共変量と従属変数の散布図

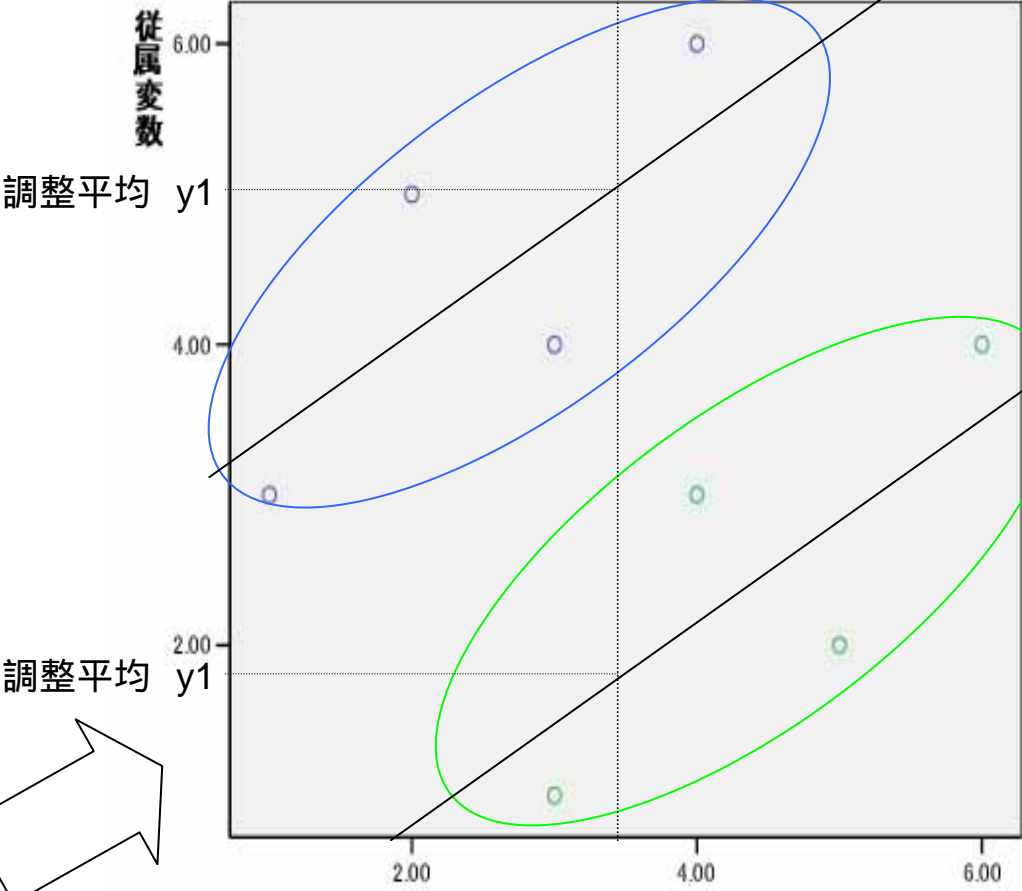


共分散分析の視覚的理解



共変量と従属変数の散布図

Y1のZ1への回帰直線



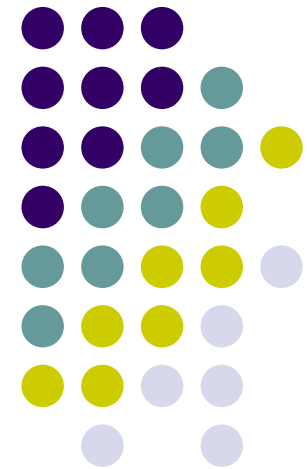
Y2のZ2への回帰直線

斜めから見ると、分布に重なりがない！

共分散分析は斜めから見る！

$$\frac{Z1+Z2}{2}$$

共分散分析を
n要因に拡張する。





よく受ける質問

- 2要因の時の共分散分析って、どうやるんですか？
 - 質問の意味がよくわからない....。
- 共分散分析の質問は、以下のようなものであるべきである。
 - 独立変数はいくつか。
 - 何が被験者間で、何が被験者内か。
 - 交絡している要因は、「何の要因と交絡しているのか」
 - 交絡している要因の水準間では調整平均の差が変わる。
 - 交絡していない要因の効果の結果は
 - 調整平均の差は変わらない
 - しかし、標準誤差が変わる 共変量と従属変数の相関が高ければ、結果はかなりかわる。



1 要因2水準被験者間分散分析

- まず、共分散分析とは関係なく、データ行列を書いてみる。(シート1)
 - 縦に、データを並べる。
 - 要因の水準1の下に水準2のデータを書く。
 - 決して横に並べることはない。
- 共分散分析をするには？
 - 共変量は、「国の要因と交絡している」
 - よって、国の要因の結果が変わるかもしれない。
 - 各々の被験者の従属変数と同じ行に共変量の値をかく。



同じ行にデータを書くことの意味1

- これらの値の間には、「何らかの関係がある」ことを意味している。
 - 例えば、被験者内要因の場合。(シート2)
 - 要因の各水準における従属変数の値を横に並べる。
 - なぜかというと、これらの値の間には、「同じ人から採られたデータですよ、という情報が含まれており、それによって説明できる変動(被験者の主効果)を誤差変動から除くことができる。

同じ行にデータを書くことの意味2



- 共変量も、同じ行にデータを書く。
 - この要因のこの水準における、Aさんの従属変数はである。だけどね、このAさんの共変量の値は××だったんだよ、と「対応付け」ないといけない。



何と交絡しているのか？

- さて、1要因2水準被験者間で、この2水準で共変量の値が交絡していなかったら？ (シート3)
- 当然、全く同じ値を投入することになる。
- 結果は標準誤差が変わりうる。



一要因3水準の場合は？

- 当然、データは縦にならべる。(シート4)
 - 1水準の下に2水準のデータがくる。
 - そしてその次に3水準のデータがくる。
- 共分散分析は？
 - この3水準ある要因と、交絡している共変量の値を、横に並べる。

1要因3水準：交絡していなかったら？



- 当然、交絡していない条件間では、共分散分析をしようが調整平均は変わらない。(同じくシート4)
- 例では、日本人と 人との間では、共変量の値は変わらない。
 - 共分散分析をしても、調整平均は変わらない。
 - 標準誤差は変わりうる。

2要因被験者間(どちらも被験者間)



- 当然、通常の分散分析のようにデータは縦に並べる。(シート5)
 - まず、A要因a1水準のデータを並べる
 - a1b1のデータ
 - a1b2のデータ
 - 次にA要因a2水準のデータを並べる
 - a2b1のデータ
 - a2b2のデータ



共分散分析の場合は？

- 当然、横に共変量の値を並べる。
 - そして、どの条件間で交絡しているのかをよくみる。
- デモデータでは、
 - 性別の要因と交絡している。
 - 男性より女性の方が共変量の値が大きい。
 - よって、性の主効果は変わる。
 - 国の要因とは？交絡していない！
 - よって、国の主効果については調整平均は変わらない。
 - 交互作用効果とは？交絡していない。
 - 日本人の女性における共変量の値と、アメリカ人の女性における共変量の値は同じ。
 - 日本人の男性における共変量の値と、アメリカ人の男性における共変量の値は同じ。



ここまでは被験者間要因

- 結局、何要因になろうが、何水準になろうが、やること(データの並べ方)は変わらない。
- どの要因と共変量が交絡していようが、あるいはどの水準間で交絡していようが、やることは変わらない。
 - どこで交絡していて、どこでしていなかったか、という情報が結果を見る際に助けになるだけ。



被験者内要因を含む場合。

- 被験者内要因と共変量が交絡しているのか、あるいは被験者間要因とのみ交絡しているのか、が問題。
- 後者なら、これまでと全く同じ。
- 例えば2要因混合計画(被験者間1要因と被験者内1要因)で、被験者間の要因と共変量が交絡している場合。(シート6)



1要因2水準被験者内要因の場合

- 独立変数: コーヒーを飲むか否か。(シート7)
- 従属変数: テストの成績
- あっと悲しいことに、条件間で、実験の実施時間帯が異なっていた。(時間帯とコーヒー要因の交絡)
 - コーヒーを飲む条件では、みんな朝の方に実験をやっていた。
 - コーヒーを飲まない条件では、みんな夜の方に実験をやっていた。
- そして、このテストは、朝の方にやればやるほど、いい成績が取れる。(そんなテストあるのか?)



確かに交絡している。

- では、データ行列はどのように書くのか？
- 右隣に、二つの値を書くのか？
 - デモのように。(シート7)
- そんなことをしてもだめ。飲んだ条件の実施時間帯が、「飲まなかった条件の従属変数」と対応付けられてしまう。
- 飲んだ条件の実施時間帯は、飲んだ条件の従属変数とのみ、飲まなかった条件の実施時間帯は、飲まなかった従属変数の実施時間帯と関連付けたい。

被験者内の要因 被験者間の要因に



- 例えば方法としては、被験者内の条件だったものを被験者間として扱ってみる(データを縦にならべる) (シート8)
- つまり、これまで、被験者内要因の各水準間では、「対応づけられている」としていたものを、「共変量の値が異なる」ため、「対応付けられていなかった」と考える。
- そして、共変量の値に関してコントロールを行う(=共分散分析を行う。)



これでOKか？

- 共変量の値を投入することによって、誤差分散がどの程度縮小されるのか？
 - 共変量と従属変数の相関係数
- 独立変数の要因を被験者内要因にすることによって、どの程度誤差分散が縮小されるのか (いわゆる、被験者の主効果)？
 - 独立変数の各水準における、従属変数の相関係数
- これらのどちらの相関係数が高いのかに注目。どちらによる変動が大きく、どちらを統制することがメリットがあるのかを考えていくことが大事？
-

いや、被験者の効果を変量因子としてモデリングする。



- 被験者の主効果は計算できないのか？
- いや、できる。被験者の要因を、ブロック要因(変量因子)であるかの様に扱う
 - (この表現は、分散分析を熟知しているひとにとっては、非常に変な表現。当たり前のことだから。)
- まず、共分散分析の前に、被験者内要因をブロック要因として扱うことを実践してみる。

被験者の要因を変量因子に投入する



- 1要因被験者内分散分析における要因 (ここではトリートメント)の主効果 (シート9)
- 1要因被験者間で、被験者の要因を変量因子としてモデリングする。その際のトリートメントの主効果
- これらは完全に一致する。
 - 当たり前。被験者ごとにブロック化し、トリートメントの効果を評価するのが被験者内分析であり、かつ、これは変量因子としてモデリングする、ということと同義。



被験者内要因と共変量の交絡

- ということは、あたかも被験者間のように扱って共変量の値を指定した後に、
- 被験者の要因を、変量因子としてモデリングすればいい。
 - シート(10)
- というわけで、被験者内の要因と交絡していたら、あたかも被験者間の要因のようにデータを並べ、そして、被験者の要因を変量因子として扱えば、被験者間の時と同じように、共変量を指定して共分散分析ができる。



参考文献

- 森敏昭, 吉田寿夫(1990) 『心理学のためのデータ解析テクニカルブック』, pp.279-282., 北大路書房
- 吉田寿夫(2003) 京都大学教育学部心理統計実習A・B 講義プリント