データ分析実習資料　　2026/07/

社会調査における分散分析ANOVAと実験計画
―平均値の差の検定―

村瀬　洋一

１．分散分析とは何か

1.1.分析の目的と具体例

◆目的　－　説明変数（独立変数）Ｘを複数設定し、被説明変数（従属変数）Ｙとの関連が強いのが、どの変数なのかを解明する。

ただし、Ｙは連続変数、Ｘは**離散変数（カテゴリー）**。

重回帰分析　－　ＸもＹも連続変数（量的）。両者は似ているがこの点が異なる。

数学的には、この2つは同じモデルである。

分析法の考え方　－　他の説明変数Ｘの影響を統計的に取り除いた上でも（統計的統制をした上でも、統計的コントロール後でも）、ある説明変数Ｘと被説明変数Ｙが関連しているか（Ｙに対するＸの効果があるか、つまりＸ内のグループ間で、Ｙの平均値に有意な差があるか）を解明する。具体的には、２つ以上の平均値の差の検定を行う。

・帰無仮説　各グループ（class, 級、組、群、カテゴリー）の平均値は全て等しい。
・対立仮説　各平均値の間に差がある。

◆具体例　－　ある調査データの意識に関する項目や所得等、連続変数として扱えるものをＹとし、性別、職業、学歴カテゴリー、年齢カテゴリーなどをＸとして分散分析を行う。例えば、テストの点数Ｙが、性別や、学歴３グループ（中、高、大）の間で異なるか分析する。別の例としては、1)ホワイトカラー、2)ブルーカラー、3)農業の３カテゴリー（職業Ｘ）で、Ｙの値（点数や身長や収入、ある意識の回答平均など）が異なるかについて分析する。つまり３つの平均値の間に違いがあるかを分析する。違いがあれば、ＸはＹと関連があるといえる。この例ではＸの数は１つで、Ｘ内のカテゴリーが３つある。職業の1, 2, 3に量的意味はない。男と女で、身長平均値に差があるかを検定した場合は、カテゴリー数は２つ。普通、平均値に差があるが、男女の各カテゴリー内でも、値にある程度のばらつきが存在する、ということになる。

分散分析は実験と深く結びついた分析法だが、調査データについて使用することもある。実験データは人数が少ないため、関連の強さは検討せず、差が有意かどうかのみを検討することが多い。また交互作用について検討することが特徴である。ただし、実験でも調査データでも、ゼロセルの問題を考えると、分散分析でのＸの数は３つか４つくらいが限界である。例えば、調査で四重クロス集計をすると、普通、ゼロに近いセルがかなり増えるだろう。重回帰分析ではＸが10個以上あることも珍しくないが、この点が異なる。

なお**t検定(統計学者StudentによるT-test)**は２つの平均値の差の検定。**分散分析のF検定**(Fisherが考えたもの)は３つ以上も可。両者は必ず同じ結果になる。

分散（ばらつき）とは何か、平均からの距離とは何か、ということが理解のこつである。

1.2.実験計画法（Design of Experiments; DOE）の考え方

実験：現象間の因果関係を解明するため、原因と思われる現象に意図的に操作を加え、結果と思われる現象の変化の有無を観察する手続き。

例：子どもの攻撃性の実験　　―　子供を２グループに分ける

暴力場面の多い映画を見せる群　→　攻撃行動が出た　　（実験群）
暴力場面のない映画を見せる群　→　攻撃行動はなかった（統制群）

∴　子どもの攻撃性（結果）を育てる原因（要因）は暴力場面である

操作された要因　―　**説明変数**　X　（**要因、独立変数**: independent variable）

観察された要因　―　**被説明変数** Y　（**従属変数**: dependent variable）

実験では、刺激（暴力場面）と反応の関連を見ることになる。サンプルを２つ以上のグループに分けるのが実験の特徴。したがって、２以上の平均値について、差の検定を行う。実験は、実験群の他に必ず統制群を作る。片方しかないと実験にならない。この例ではＸは１つだけ（暴力的映画）なので簡単だが、Ｘが複数あると、複雑な実験計画となる。

被説明変数Ｙにおいて観察された変化が、Ｘによって確実に引き起こされたと考えるためには、Ｘ以外のすべての条件が等しいことが必要である。しかし、実際には無数の条件の違いが混入してくる恐れがある。例えば、たまたま片方のグループに、活発な子供がたくさんいた、などの事情である。その中でＹに影響を及ぼすと仮定できるものを**剰余変数**(extraneous variable)**あるいは誤差要因**と呼ぶ。剰余変数の有力候補は以下である。

被験者変数．．．性、年齢、性格、能力や知能、本人のやる気や動機づけ
それ以外．．．．実験環境、課題条件、実験者のふるまい、教示

剰余変数を無効にする方法には、次の４種類がある。

（例　Ｘ：新しい教授法、Ｙ：学習成績、剰余変数：被験者の知能）

1 均衡化　　より多数の被験者を、実験群と統制群に無作為に割り付ける
2 相殺化　　同じ知能水準の被験者が実験群と統制群で同数づつになるよう配置する
3 恒常化　　一定の知能水準の被験者だけを集めて実験する
4 説明変数化
　被験者を高知能群・（中知能群）・低知能群に分け、もう一つの説明変数にする

現実には、大学の心理学実験の多くは、大学生のみからデータを取る。そこで、全員の年齢や学歴や能力は、ほぼ等しい（すでに恒常化した）と考えた上で実験を行うことが多い。また、心理学的な実験では、何らかの実験結果の得点や、因子得点をYにして、刺激を与えた群とそうでない群の差(2グループの平均値の差)について、分散分析を行うことが多い。つまり、**刺激－反応モデル**(Stimulus-Response model、**SRモデル**)に基づいた分析が多い傾向がある。刺激が何種類かある場合は、3つ以上の平均値の差の検定をする。行動主義的心理学や、実験心理学というものは、基本的な考え方としてはSRモデルに基づいている。SとRの間に人間がいるが、人間の頭の中というのはブラックボックスであり、中身で何が起きているのかは不明、というのが基本的な考え方である。

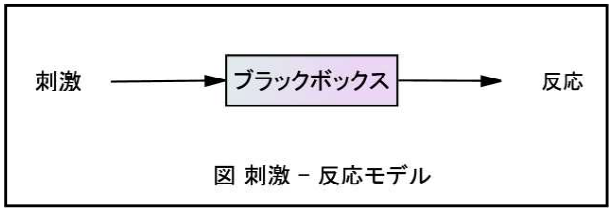


図 刺激－反応モデル

1.3.分散分析の考え方

実験を行なった結果、被説明変数に違いが見られる場合、その違いが偶然によって生じたのか、それとも実験処理によって生じたと考えた方がよいのかについて検討するため、統計的な推定を行なう。男女など２つの平均値の比較にはｔ検定を用いるが、処理水準（level; X内のカテゴリー数）が３つ以上の場合や、Ｘが２つ以上存在する場合の平均値の差の検定には、**分散分析**（Analysis of Variance:ANOVA）を用いる（なお、説明変数Xが１つで、カテゴリー数２ならば、**１元配置２水準**の分析と呼ぶ）。

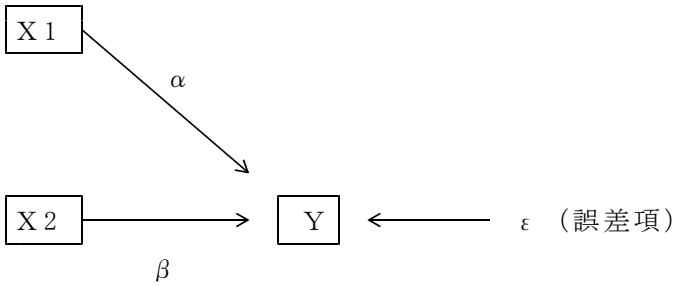


図1.　２元配置分散分析の基本モデル（グラフによる表現）

簡略化のため、Ｘが１つの場合を数式で表すと次のようになる。

分散分析の基本モデル（完全無作為法、水準数がm個で要因数１の場合）。

$$y_{ij} = \mu + \alpha_j + \varepsilon_{ij} \quad (i=1,...n;\ j=1,...m) \quad \cdots\cdots\cdots\cdots (1)$$

y_{ij}：j番目の水準のｉ番目の個体の値（各個人の測定値）
μ　：母集団の平均
α_j ：**母集団におけるj番目の水準の効果（グループに属する効果）**
ε_{ij}：j番目の水準のｉ番目の値に関する誤差

$\mu = \overline{y}$（全体平均）、$\alpha_j = \overline{y}_j - \overline{y}$（全体平均と各水準の平均値の距離）と置き換えると、

$$y_{ij} - \overline{y} = (\overline{y}j - \overline{y}) + (yij - \overline{y}j) \quad \cdots\cdots\cdots\cdots (2)$$

（２）式の左辺は、各個人の値と平均値との距離（ばらつき）である。

両辺を２乗してiとjについて総和して整理すると、

$$\sum_{1}^{n}\sum_{1}^{m} (yij - \overline{y})^2 = n\sum_{1}^{m} (\overline{y}j - \overline{y})^2 + \sum_{1}^{m}\sum_{1}^{n} (yij - \overline{y}j)^2 \quad \cdots (3)$$

すなわち（３）式は、ＳＳtotal＝ＳＳbetween＋ＳＳwithin　となる。

全体平方和が級間平方和と級内平方和（残差平方和）に分解されることがわかる。

ＳＳtotal　（全体平方和）：全体平均の回りの個々の測定値のばらつき
ＳＳbetween（級間平方和）：全体平均と各水準の平均値との距離（ばらつき）
ＳＳwithin　（級内平方和）：各水準の平均値の回りの個々の測定値のばらつき
ＳＳ： Sum of Square 二乗和

また偏差平方和（ＳＳ）をその自由度で割った商が平均平方（ＭＳ）である。

このMS(Mean Square)の期待値は、各々の自由度でχ^2分布することが知られている。

分散分析のモデルの基本的な考え方は、SSやMStotalの部分の長さを２つに分割して考えるということである。ある個人の値(以下の図の y_{i1})と全体平均($\bar{y}$)との距離は、次の図のように２分割して考えることができる。グループ平均(y_1)と全体平均の距離が大きければよいモデル（グループに属する効果が大きい）といってよい。

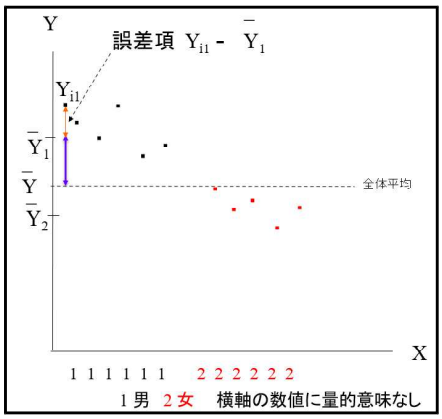


図2. SStotalの分割　村瀬他(2007:100)より

分散分析の具体的目的は、あるXの効果が有意であるか、すなわち各水準の平均値の差が「偶然のものではない」と見なしてよいか、の検定を行なうことである。これはすなわち、級間平均平方（分散）が級内平均平方（分散）より十分に大きいかどうかについて、検討しているのである（各水準の分散に関する図は**村瀬他(2007:101)**を参照）。

1.4. Ｆ値（Ｆ比）の考え方

$$F = \frac{MS_{BETWEEN}}{MS_{WITHIN}} = \frac{\text{級間平均平方}}{\text{級内平均平方}} = \frac{\text{モデルによって説明できる分散}}{\text{モデルによって説明できない分散}} \quad \cdots (4)$$

このＦ値は、χ^2分布する２つの変数の比であり、自由度（m-1, n-m）のＦ分布する。Ｆ分布表で、Ｆ値の出現する確率ｐが、帰無仮説（各平均値に差はないという仮説、つまり各平均値の差は偶然という仮説）が採用される確率（危険率、有意水準）となる。Ｆ値が大きければ、モデルによって説明できる部分が大きいといえる。

1.5. 実験計画の様々なデザイン

実験計画における説明変数の設定の仕方には、処理条件ごとに異なった被験者群を割り付ける完全無作為法(completely randomized design)とすべての同一の被験者群にすべての処理条件での測定を実施する乱塊法(randomized block design) がある。乱塊法は完全無作為法よりも被験者数を節約できる利点がある反面、先の条件での測定が後の条件での測定に影響を与える可能性があるため、その影響を相殺化などの操作で取り除くことができると考えられる場合以外には用いられない。またこれらの両方の種類の説明変数を用いた混合法(mixed design)と呼ばれる計画もある。

1.6. 交互作用（interatcion）

複数の説明変数を組み合わせた要因計画（factorial design）では、それぞれの要因Ｘの単独の影響（主効果）を検出できるほか、一方の要因の効果が他方の要因の水準によって異なるかどうか（交互作用）についても検討することが可能である。

以下の図は、要因Ｘ1（性別）と要因Ｘ2（職種）の組合せごとに、Ｙの平均値をグラフ化し、主効果と交互作用の関係を模式的に示したものである。２つの線が平行なときは交互作用はない。重回帰分析と違って、分散分析は交互作用を容易に検討できる。

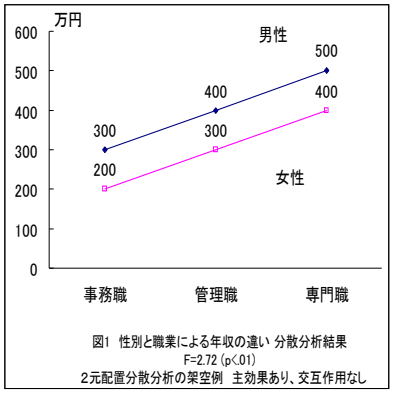


図1 性別と職業による年収の違い 分散分析結果
F=2.72 (p<.01)
2元配置分散分析の架空例 主効果あり、交互作用なし

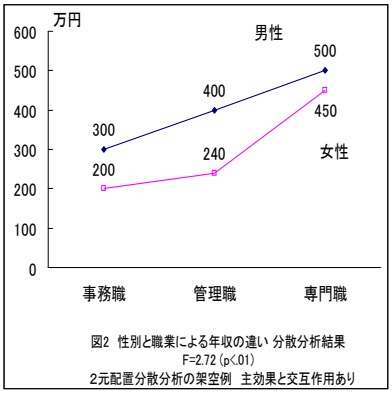


図2 性別と職業による年収の違い 分散分析結果
F=2.72 (p<.01)
2元配置分散分析の架空例 主効果と交互作用あり

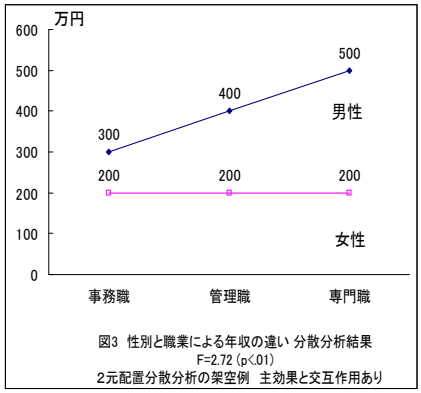


図3 性別と職業による年収の違い 分散分析結果
F=2.72 (p<.01)
2元配置分散分析の架空例 主効果と交互作用あり

1.7. 多重比較(multiple comparison, Post-hoc test)

分散分析において（交互作用が有意でなく）有意な主効果があったということは、「その要因の各水準の平均がすべて等しい」という帰無仮説が棄却されたことを意味するにすぎない。説明変数が「１０ｍ，２０ｍ，３０ｍ」のような比例尺度であったとしても、主効果が有意であることは、説明変数と被説明変数の間に線形関係があることを意味しない。両変数の関係がＵ字形または逆Ｕ字型の場合にも主効果は有意になる。さらに、Ｘの水準が３つ以上ある場合、どの水準とどの水準の間に統計的に有意な違いがあるかを知ることが必要になる。ここで通常のｔ検定を繰り返して実施すると、第１種の誤差に関して「甘い」検定になってしまう（検定の多重性）。

これを避けるため、全体としての有意水準を一定に保つための多重比較の方法が各種考案されている。統計学者Tukeyのスチューデント化した範囲の検定、Scheffeの対比による方法、Bonferroniのｔ検定などがよく用いられる。なおSPSSのメニューでは「**その後の検定（posthoc test)**」というボタンが多重比較のこと（翻訳ミスで変な名前になっている）。

1.8. 分散分析の適用条件

分散分析では1)標本が正規分布にしたがう母集団から抽出されたものであること（正規性）、2)各水準にはいる標本が独立であること（独立性）、3)各水準にはいる標本の分散が等しいこと（等分散性）、の３つが前提とされる。正規性と等分散性については、完全に満足されなくても分析結果にたいして影響しない（頑健性がある）ことが知られているが、あまりにも前提条件からはずれている場合、分散分析は行うべきでない。

測定の性質上の理由で、正規性や等分散性が崩れる場合には、測定値を適当に変換することにより、条件を満たすことができる場合がある。テストなど一定数の項目中の正答率を角変換したり、反応時間を対数変換することはよく行なわれる。ただし、標本数があまりに少ない場合や順序尺度である場合は、分散分析でなく、より制約の緩いノンパラメトリック検定を用いる方がよい。なお分散分析も重回帰分析も、一般線形モデル(General Linear Model: GLM)というより広い概念に含まれるものであり、数学的には同じものである。

1.9.非実験データの分析

上記のように、分散分析は本来、実験計画法と密接に結びついた分析法である。しかし社会調査データでも、カテゴリー別の平均に差があるかを分析するために、分散分析をよく用いる。ある変数Ｙ（例えば所得）についてのデータを、別の変数Ｘ（例えば人種や学歴カテゴリーや職業）など一定の基準で分割してグループをつくり、グループを説明変数Ｘとして平均値の差の検定を行う。これは個々人のＹの値がばらつく（人によって違う）内、グループに属することによる違いとして説明できる分が統計的に有意かどうかを、検討することになる。非実験データの場合、説明変数Ｘは操作されたものではなく、剰余変数の統制が困難であるので、ＸとＹの間に因果的な関係を推論しようとする際は慎重にすべきである。第２、第３のＸが存在する可能性は常にある。ただし実験データだからといって、関連が単なる相関でなく因果であると断定はできない。他の要因Ｘによる見せかけの相関である可能性も常にある。ただ調査データの方が、多くの要因を含むことが多い。実験データは大半が大学生のみのデータであり、均質であるが、多くの学歴や職業の人を含むわけではない。その意味で、実験データは代表性が欠けているデータである。

社会調査データなど、比較的人数が多い大標本では、差や関連の大きさについて、検討することを目的とする。実験など小標本データは、差の大きさや関連の強さは検討できないので、差が有意かどうかを解明することに目的を絞って、分析することが多い。

２．ＳＰＳＳによる分析

2.1.ＳＰＳＳの操作

分散分析は、計算方法が比較的単純であるため電卓や表計算ソフトなどで行うことが可能である。統計パッケージのＳＡＳやＳＰＳＳには分散分析を行うＡＮＯＶＡという名前のプログラムが用意されている。ＳＰＳＳでは、説明変数が１つだけの分散分析の場合、画面上の「分析」をクリックして、平均値の比較を選び、１元配置分散分析（シンタックスはoneway）を選ぶ。

あるいは、以下のようにシンタックスを書いても良い。太字の所に用いる変数を入れる。この例では、教育年数eduが被説明変数、年代nendaiが説明変数である。

重回帰分析の場合、説明変数は量的変数なので、年齢はなるべく細かい方が良い。しかし分散分析では、説明変数はカテゴリー変数なので、あまり細かい変数だと良くない。この例では、説明変数は年代という２０代から６０代まで５段階の変数を用いている。

シンタックス例　１元配置分散分析

```
ONEWAY    edu BY nendai          ←ONE だけでも動く　この例ではEDUがY
  /POSTHOC = TUKEY ALPHA( .05)     ←　テューキー法で多重比較をする命令文
  /STATISTICS DESCRIPTIVES         ←　なくてもよい。平均値などを出す命令文
  /PLOT MEANS  .                   ←　なくてもよい。平均値折線グラフを出す命令文
                                        (PLOTはPSPPでは使えない)
```

解説　１行目ONEWAY　のあとに　　Y by X の順にモデルで使う変数を書く

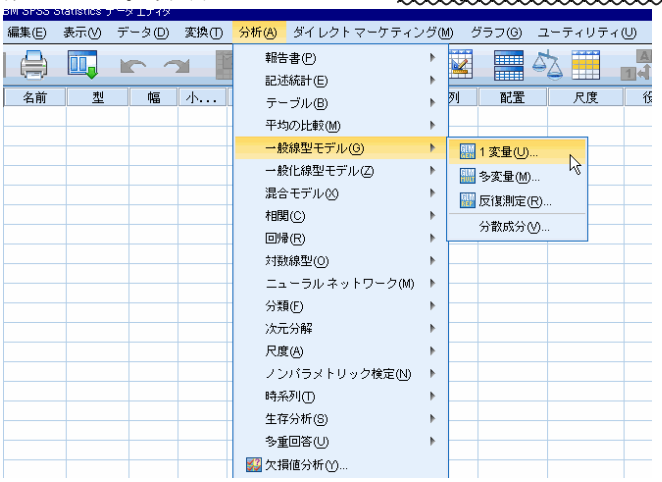
```
  /POSTHOC = TUKEY ALPHA( .05)   多重比較法にテューキー法を用い有意水準を5%とする
  /PLOT MEANS  .                 各カテゴリー別平均値の図を出す
```

出力として、F値や有意水準、各年代グループの平均値等が出る。plot行を書くと折れ線グラフも出る。また多重比較の表には、平均値間で有意な差があるところに*マークがつく。

２元配置以上の時は、「分析」をクリックして一般線形モデルを選び、「１変量」を選ぶ（シンタックスはunianova）。「多変量」は複数のYを１度に分析するがあまり使わない。

被説明変数として使いたいものを１つ選び、自分の好きなモデルを作ればよい。説明変数Ｘは普通、質的変数なので、**固定因子のボックス**に入れる（それ以外のボックスはあまり使わない）。**共変量のボックス**に入れるのは量的変数のＸである。「モデル」ボタンを押して、分析に用いる説明変数を選ぶ（モデルは「すべての因子による」を選び、とくに指定しなくてもよい）。平方和の計算法に何種類かあるが、タイプⅢがよく使われる。交互作用は、まずは２次までを入れれば良い。

「オプション」ボタンを押して、記述統計を出すと、各カテゴリーごとの平均値の数字が出る。また「作図」ボタンを押して、横軸と線の定義変数を指定すると、各カテゴリーごとの平均値のグラフが出る。多重比較を行いたい時は「その後の検定」ボタンを押し変数を選ぶ。分析目的に応じてTukeyやScheffeなどを選ぶとよい。



あるいは、以下のようにシンタックスを書くこと。太字の所に用いる変数を入れる。この例では、教育年数eduが被説明変数、年代nendaiと性別q52sexが説明変数である。説明変数Xは2つともカテゴリー数である。

/DESIGN行にモデルを書く。この例では主効果として nendai q52sex 、交互作用としてnendai*q52sex を入れたモデルとなっている。この例では２次の交互作用のみを入れているが、説明変数が３つ以上ある時は、３次の交互作用を入れることもできる。多重比較をする時は/POSTHOC行を書く。以下の例ではSCHEFFE法で多重比較をしている。

シンタックス例　２元配置分散分析

例１　簡略な例

```
UNI          EDU  BY  nendai q51sex          ← この例ではEDUが被説明変数Y
  /DES    = nendai q51sex  nendai*q51sex     ←自分の作ったモデルを書く
  /POS    = nendai  ( SCHEFFE )              ←多重比較をする場合に手法を書く
  /PRI    = DES                              ←平均など記述統計量を出す場合に書く
  /PLO    = PROFILE ( nendai*q51sex )  .     ←平均値の図を出す場合に書く
```

例２　詳しい例

```
UNIANOVA     EDU  BY  nendai q51sex          ←UNI だけでも動く　この例ではEDUがY
  /METHOD  =  SSTYPE(3)
  /DESIGN  =  nendai q51sex  nendai*q51sex    ←自分の作ったモデルを書く
  /POSTHOC =  nendai ( SCHEFFE )              ←多重比較をする場合に手法を書く
  /CRITERIA = ALPHA( .05)                     ←多重比較の有意水準を書く（省略可）
  /EMMEANS = tables (nendai)                  ←推定周辺平均を出す
  /PRINT = DESCRIPTIVE                        ←記述統計量を出す場合に書く
  /PLOT = PROFILE( nendai*q51sex )  .         ←平均値の図を出す場合に書く
```

以下が出力例。「被験者間効果の検定」という表が、いわゆる分散分析表。この例では、１行目がモデル全体のF値であり、19.776で有意であることが示されている。主効果は

nendai 38.383、q52sex 39.053であり、どちらも有意。交互作用のF値は有意ではない。

「被験者間因子」という表が別に出るが、これは各グループの人数Nが出ているだけである。記述統計量はグループごとに平均値等が出るだけ。シンタックスでplotオプションを書くと、平均値の折れ線グラフが出るので分かりやすい。また「その後の検定」という表は多重比較の結果である。平均値間に有意な差があるところに*マークがついている。

被験者間効果の検定(分散分析表の例)

従属変数 本人学歴

ソース	ﾀｲﾌﾟ Ⅲ 平方和	自由度	平均平方	F 値	有意確率
修正ﾓﾃﾞﾙ	1416.723	9	157.414	19.776	0.000
切片	203105.226	1	203105.226	25515.805	0.000
NENDAI	1222.108	4	305.527	38.383	0.000
Q1SEX	310.858	1	310.858	39.053	0.000
NENDAI * Q1SEX	54.265	4	13.566	1.704	0.147
誤差	8198.776	1030	7.960		
総和	226799.000	1040			
修正総和	9615.499	1039			

a R2乗 = .147 (調整済みR2乗 = .140)

なお、以下のようなEMMEANSコマンドを使うと、Yの**推定周辺平均**(Estimated Marginal Means)を出すことができる。これは、他の変数の効果や交互作用の効果を取り除いて調整した上での、Yの各カテゴリーにおける平均値である。他の変数の効果がある場合、記述統計量で出た平均値とは異なる値が出る。詳しい操作法は岸(2012:pp.187-)参照（なおSASのLSMEANと一致する）。

```
/EMMEANS = tables (nendai)
```

無料ソフトPSPPでは**UNIANOVA**が使えないので、以下のGLMコマンドを使う。

```
GLM EDU BY  NENDAI Q51SEX.
```

2.2.結果のまとめ方

分散分析の結果は変動因、平方和（ＳＳ）、自由度（ｄｆ）、平均平方（ＭＳ）、Ｆ値、有意水準（ｐ）などが一覧可能な、分散分析表という形式で表示する。上記のSPSS出力をもとに、村瀬他(2007, p.115)のような表にまとめること。**ただし、有意水準は5%として、表を作ればよい。0.1%はあまり使わない。**また、F値や平方和は小数点以下1ケタを書けばよい。なお字数に制限のある雑誌論文などでは**「F(2,37)=9.67 (p<.01)」**と省略して報告することもある（Ｆの後の括弧内の２つの数字は要因の自由度と誤差の自由度を示す）。

調査結果について、カテゴリーごとの平均値を提示する場合、上記の図のように平均値を折線グラフにすると分かりやすい。図タイトル下の部分に注として、Ｆ値や自由度DF、有意水準pを書くこと。平均値は、単純な平均値か、推定周辺平均なのか、明示すること。

３．分析時の注意点

3.1.分析の前に必ず欠損値処理をすること

多くの場合、欠損値は９か99。SPSSの場合、missing valuesコマンドを用いる。

回答が２桁の場合、欠損値99である。まず単純集計をとって確認するとよい。

3.2.分析の前に変数の向きを必要に応じて逆転し、わかりやすく設定する

分析を行う前に、すべての変数を逆転するなどして、数字が大きいほど肯定になるように直すこと。数字が小さいほど肯定、などの変数が混ざっていると、とても分かりにくい。

3.3.用いる変数について

用いる変数は、Yは連続変数（量的変数）、Xは**離散変数（カテゴリー変数）**であることに注意。各カテゴリー内の人数が少ない場合はカテゴリー合併を行うとよい。あるいは、人数が少なく異質すぎるカテゴリーは、はずれ値だと考えて分析から削除しても良い。なお、各カテゴリーの度数が異なる（例えば男女で人数が違う）場合は、アンバランスデータというが、データ人数が多ければ、分析上とくに問題はない。しかし、当然ながら、ある群の人数が数名しかいないような場合は、分析結果が安定しない。

４.モデルの考え方とデータ人数

重回帰分析と分散分析は、数学的にはほぼ同じモデルである。しかし、重回帰分析は線形モデル（直線）を前提としているが、分散分析は、そのような前提がない。

重回帰分析は、関連の大きさを見るため、普通、数百人以上の大きなデータでないと使わない。それに対して分散分析は、もともと少人数データの有意差検定のために作られたものである。有意かどうかを見るためには、数十人のデータでも良い。といっても、各群にある程度の人数がないと分析は難しいが、結局、分散分析とは、グループ間の平均値の差を検定しているだけである。平均値に差があるかどうか（有意かどうか）を分析しているのみである。100人以下の実験データなど、人数が少ないデータでは、関連の大きさについては、通常、考慮しない。大きさを示す**イータ二乗（相関比）**という係数を出すことができるが、少人数データではあまり使われない。データ人数が少ない場合、係数の値は安定しないのであまり意味がない。

重回帰分析や因子分析は、経験的には、最低でも数百人以上のデータが必要である。人数が少ない場合は、分析結果は不安定になるので信用できない。しかし分散分析で有意差をみるだけならば、ある程度少人数でも可能であり、数十人規模の実験データを分析することもある。本来は、少人数のデータで無理に多変量解析をやるべきではない。あまり少人数だと統計的検定はできないし、クロス集計等で、傾向を目で見るくらいで十分である。

５．課題

調査項目の中から、自分が分析に用いたい変数を決める（被説明変数Ｙは、連続変数を１つ決め、Ｘはカテゴリカル変数を３つ以上決める）。そしてプログラムを動かし分析を行う。何を説明したいか（Ｙ）を決め、それを説明するのにふさわしい変数Ｘが何かを考え、自由にモデルを作ること。Ｘとして年齢カテゴリーと学歴カテゴリーを使うなど、何種類かのモデルを自分で考え、試行錯誤すると良い。まずは、Ｘに性別と年齢カテゴリーの２つを入れ、主効果と、２次の交互作用のみを入れたモデルで分析すると良いだろう。

分析時には、上記の注意点に気をつけること。分析結果が出たら、Ｆ値や有意水準ｐを見て、上記の図１～３のように、各グループの平均値を、折れ線グラフにして結果をまとめる。Ｙの値を縦軸としてグラフにすればよい。SPSS出力でいくつかの平均値を出し、エクセルでグラフを作ること。図の中にタイトルとして、用いたデータやF値などを書く。

そして、各Ｘの主効果が有意か、交互作用は有意か、モデル全体の説明力はどうか等についてグラフ下に文章を書く。また、結果の解釈や自分なりの考察を書くこと。解釈とし

て自分独自の意見を書くことが重要。重回帰分析の課題と同様、データを男女に分割した後で、分析してもよい。データを男女に分割してからONEWAYコマンド使うと、STAで男女別の平均値を出すことができる。それをもとにエクセルで平均値の折れ線グラフを作ること。

６．発展版

職業カテゴリーを自分で作成する。５～１０カテゴリーなど変数を作り、年齢、学歴、職業をXとして、何らかのYを設定して、分析を行う。職業カテゴリーは、ＳＳＭ調査の旧８分類や新８分類、安田・原の総合職業分類などを参考に、自分で作成する。原・盛山『社会階層』の巻末用語解説などを読む。職業とは何か、産業とどう違うかなど、よく理解すること。社会調査は、職業や人間関係について詳しく分析することが１つの特徴である。職業とは、地位と役割を表す。職業は地位の総合的指標であり、社会階層の指標でもある。各種の見本シンタックスを参考に、自分で職業カテゴリーを作成すること。なお、職業の４次元については、安田・原(1982)参照。

SSM職業旧８分類　専門、管理、事務、販売、熟練、半熟練、非熟練、農業

SSM職業新８分類　専門、大ホワイト、中小W、自営W、大ブルー、中小B、自営B、農業

日本のSSM調査やJGSS調査においては、本人職業は、調査後に自由回答をアフターコードして、501から689までの1995SSM職業小分類コードを入力する。これを、8分類のような大分類に分けて、新変数を作ると良い。ISSP調査では国際標準職業分類ISCOが使われている。現実には、最近は農業が数%以下など、分析に使うためには問題がある。農業をブルーカラー（熟練労務的職業）と合併してカテゴリー数を減らす（値の再割り当てをする）など、自分の分析目的に応じて工夫すると良い。職業コーディングについては実習ホームページにある資料を見ること。

参考文献

ボーンシュテット・ノーキ. 1990. 『社会統計学　－社会調査のためのデータ分析入門』. ハーベスト社。

南風原朝和. 2002. 『心理統計学の基礎 ―統合的理解のために』有斐閣.

原純輔・盛山和夫. 1999. 『社会階層 -豊かさの中の不平等』東京大学出版会.

★巻末に職業カテゴリーの説明がある。

市川伸一・大橋靖雄・岸本淳司・浜田知久馬. 1993. 『ＳＡＳによるデータ解析入門　第２版』東京大学出版会.

石村貞夫. 1992. 『分散分析のはなし』東京図書.

石村貞夫. 2002. 『SPSS による分散分析と多重比較の手順　第２版』東京図書.

石村貞夫. 2004. 『SPSSによる統計処理の手順　第４版』東京図書

海保博之編. 1985. 『心理・教育データの解析法10講　基礎編』福村出版.

岸学. 2012. 『SPSSによるやさしい統計学　第2版』オーム社.

★分散分析について解説が詳しい

小牧純爾. 1995. 『データ分析法要説　－分散分析法を中心に』ナカニシヤ出版.

三宅一郎他編. 1991. 『新版SPSS X第３巻　解析編２』東洋経済新報社.

村瀬洋一他編. 2007. 『SPSSによる多変量解析』オーム社.

★各カテゴリーごとの分散（ばらつき）についての図はこれを参照

小野寺孝義・山本嘉一郎編. 2004. 『SPSS事典 BASE編』ナカニシヤ出版.

高橋行雄・大橋靖雄・芳賀敏郎. 1989.『ＳＡＳによる実験データの解析』東京大学出版会.

豊田秀樹. 1994. 『違いを見抜く統計学　－実験計画と分散分析入門』講談社ブルーバックス.

安田三郎・原純輔. 1982. 『社会調査ハンドブック（第３版）』有斐閣.

★職業とは何かについては、これを読むこと。

ANOVAシンタックス例

```
/*****    変数の方向を逆転 *****/
  MISSING VALUES   Q7A (9).
  COMPUTE          N7A =5-Q7A.

/*****  分散分析     *****/
  UNIANOVA     N7A  BY  nendai q51sex
       /DES =  nendai q51sex  nendai*q51sex
       /POS =  nendai ( SCHEFFE )
       /PLO = PROFILE( nendai*q51sex )
       /PRI = DESCRIPTIVE    .
```

シンタックス解説

BYの前に被説明変数(量的変数Y)を書く。

BYの後ろに説明変数(カテゴリー変数X)を書く。

```
/***** 一元配置分散分析で平均値を出す *****/
  ONEWAY     N7A BY nendai
      /STA DES  .
```

有意ということの考え方について

例えば、新薬の臨床試験等においても、分散分析はよく用いられる。臨床試験というと言葉は良いが、大人数を用いた人体実験である。新薬を使ったグループと、使わないグループ（統制郡）だけなら、２つの平均値の差の検定（t検定）を用いれば良いが、実際には、新薬を使い、かつ新しい治療法を用いたグループと、新薬を使い、かつ新治療法は用いなかったグループなど、３つ以上のグループとなることが多い。３つ以上の平均値の差の検定ならば分散分析を行う。

統計的に意味がある結果を得るためには、統制郡を含め、各グループに、ある程度の人数が必要である。人数が少なくて有意でなかった場合は、新薬の効果（例えば実験群において、平熱になった日数が短い）は有意ではないということになる。しかしこれは、新薬の効果がないということを意味するだろうか。もしかしたら、さらに実験を続け実験参加人数を増やせば、有意になる可能性もある。また、統制群が、たまたま若くて体力がある人が多かったり、恵まれた住宅地にある病院で、もともと健康状態が良かった等の場合は、やはり、有意な結果は出にくい。しかし、新薬を用いないグループ（**統制郡**）を、大人数揃えることが、実際にできるだろうか。症状が重い場合は、治療現場としては、やはり新薬を使いたくなるし、あえて使わないというのは人権上問題があるから、用いなかった場合の人々（**統制郡**）は、たまたま重症化していない人や、体力がある人が多めになってしまう可能性がある。つまり、統制郡は、そもそもあまり重症化していないような人が多めに入ることもありうる。すると、実験として適切ではない。

そもそも**剰余変数**として、一部の人々が健康状態が良い、などの事が、自明でない場合は、どのような偏りがあるのか、実験開始前には、よく分からないのである。日本では、都心の古い住宅地には貧しい黒人が多く、郊外の恵まれた住宅地に豊かな白人中流階層が多い、などの分かりやすい社会階層構造が少ない。しかし、日本にも格差があり、恵まれた生活をして、もともと栄養状態が良い人々人々は存在する。実際のところ、例えば私の知人が横浜市青葉区の恵まれた住宅地から、東京都足立区に、数年前に引っ越したところ、住んでいる人々の様子が大きく違ったということである。私も日本のやや貧しい地域にて、小さい子ども連れの家庭の夕食を見たところ、ラーメンと炒飯のみで、野菜や栄養のバランスをあまり気にしない食事であり、驚いたことがある。ただ日本において、貧困地区やスラム街は目立たないし、日頃の栄養の偏りや健康状態に、諸外国で見られるような大きな格差があるかは分からない。

以下は指導メモ

```
ONEWAY                ←  一元配置分散分析をせよ、という命令文
     NEW17 BY NENDAI2
      /PLOT MEANS
      /STA DES    .
```

PLOTはPSPPでは動かない。

分散分析の被説明変数Ｙは、量的意味があるものを使う
　説明変数Ｘはカテゴリー変数。あまり細かい変数は不可。

重回帰分析の被説明変数Ｙは、量的意味があるものを使う
　　説明変数Ｘも量的変数、つまり、できるだけ細かい変数
　　ただしダミー変数をＸとしてもよい。
　　必ず事前に欠損値処理をすること。

　　数字が大きいほど肯定、などの変数に、方向を変える。

SPSSシンタックスの並べ方

1 データ読み込み
2 欠損値処理
3 量的変数に直す。教育年数の新変数などを作る。
4 変数逆転

最後に、重回帰や分散分析

ANOVAシンタックス例

```
 UNI    SEIMAN  BY  nendai q52sex
       /DES =  nendai q52sex  nendai*q52sex
       /POS =  nendai ( SCHEFFE )
       /PLO = PROFILE( nendai*q52sex )
       /PRI = DESCRIPTIVE    .
```

BYの前に被説明変数(量的変数Y)を書く。
BYの後ろに説明変数(カテゴリー変数X)を書く。
/DES行に、主効果2つと交互作用1つなど、モデルをかく。

結果は村瀬他(2007) p.115のような表にまとめること。
　ただし有意水準は5%にする。0.1%はあまり使わない。