

2023/9/

## 分析ソフトSPSSの操作 —まずは命令文（シンタックス）を覚える—

村瀬 洋一  
(立教大学社会学部)

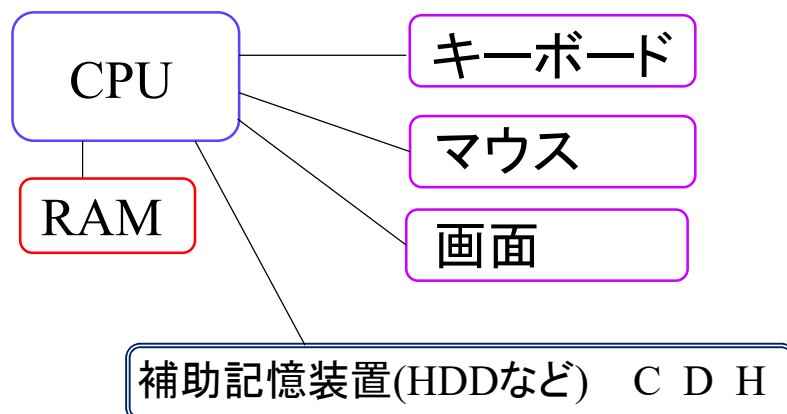
### ◆ 主な内容

- SPSSの3画面
- クロス集計
- カテゴリー合併などデータ加工をする
- 関連とは何か

まず、練習用データファイルを保存 →ゼミホームページ左上

1

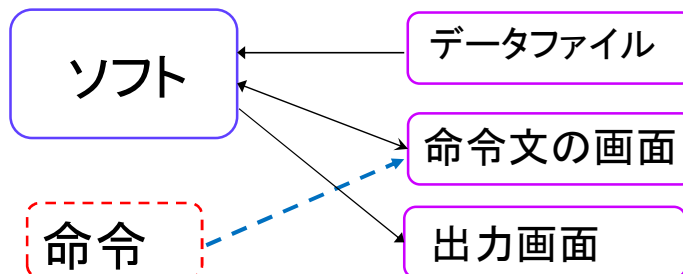
## 1. はじめに 1.1.PCの仕組み



記憶装置の中をフォルダで区切る

2

## 1.2.SPSSの仕組み



- 人間が命令を実行すると、出力画面に文字や表が出る
- データには**テキスト形式データ**と、SPSS形式データsavがある

3

## 自分でメモを書く、集中して聴く！

- ◆ 細かい部分でも、間違っているとシンタックスは動かないので注意。
- ◆ データを開いた状態で、シンタックスを実行する。  
ctrl+A, ctrl+R
- ◆ 存在する変数名を正確に書く  
データウィンドウで確認する。画面下のタブを切り替える。性別の問はQ52SEXなど。

4

## パソコン使用の注意点

- たくさんウィンドウを開かない。メモリを消費して動きが遅くなる。**開くのは最小限に！**
- ファイル名やフォルダ名、ユーザー名は、全角文字を使わない。分析ソフトを使った時にフリーズしやすい。
- 手元のパソコンにファイルを保存する。ネット上に置いたまま開かない。
- ZIPファイルから、CDドライブなど手元のハードディスクにまずデータファイルを出す。その後に、SPSSなどで開く。
- エクセルの「**セルの書式設定**」を使いこなす。
- エディターソフトを使う。作業効率がよくなる。

5

## 見本のデータ

SPSS形式データファイルは、**拡張子sav**

何か調査データを保存して、開いてみる。

村瀬研ホームページ資料

2022年仙台市調査、あるいは他でもよい

400人のデータは400行。

パスとは何かを理解する。データの場所。

C:**~~¥~~data15¥**15fukudata.txt

6

## 拡張子とは

### ◆ファイル名の最後の3文字。

- xls エクセルのシート
- docx ワードの文書
- txt テキスト形式ファイル
- dat 中身は数字のみのテキストデータ
- sav SPSS形式データファイル
- sps SPSSシンタックス(命令文)

7

## 見本のシンタックス

福島市調査 mihon1.spsなどを保存して、SPSSで開く

データをSPSSで開いた状態で、

→シンタックスをすべて選択して、実行。

シンタックスとは、分析やデータ加工の命令文

- ピリオドをつけすぎると動かない。
- 全角空白があるとエラーになる。絶対に入れない。
- 何らかのエディターソフトを自分のPCに入れると良い。  
。各種の仕事を素早くできとても便利。秀丸やさくらエディターなど。

8

# データウィンドウ

15FUKU11mihon.sav [データセット1] - IBM SPSS Statistics データエディタ

ファイル(F) 編集(E) 表示(V) データ(D) 変換(T) 分析(A) グラフ(G) ユーティリティ(U) 拡張機能(X) ウィンドウ(W) ヘルプ(H)

表示: 147 個 (147 変数中)

	ID	area	Q1 A	Q2 B	Q2 C	Q2 D	Q2 E	Q3 A	Q4 B	Q4 C	Q4 C	Q5E L	Q5W A	Q5G A	
1	1611	16	40	2	2	4	2	1	4	2	4	0	10	0	
2	1625	16	46	2	3	4	3	1	4	4	2	10	40	40	
3	101	1	3	1	2	3	3	4	1	4	4	3	0	7	0
4	109	1	78	2	2	2	2	2	1	4	4	2	1	10	0
5	110	1	65	1	2	2	2	2	1	3	4	1	1	7	1
6	111	1	60	1	1	3	3	2	1	3	4	3	1	3	0
7	112	1	39	2	2	2	2	9	1	4	4	2	0	14	0
8	113	1	58	1	2	2	3	1	1	4	2	1	1	14	0
9	114	1	2	2	1	3	3	2	3	4	4	2	0	0	0
10	120	1	40	3	1	3	3	2	1	2	2	1	1	15	0
11	124	1	25	1	1	3	2	1	1	4	3	3	0	8	1
12	125	1	72	2	2	3	3	2	1	4	4	1	2	7	0
13	128	1	40	3	3	4	4	4	3	4	4	1	0	0	0
14	130	1	50	2	2	4	3	2	1	4	4	4	1	5	0
15	201	2	60	1	2	3	3	2	1	4	4	3	2	14	0
16	206	2	40	2	2	4	3	2	1	4	4	1	5	7	3
17	222	2	70	1	1	2	2	2	1	4	3	1	1	7	1
18	224	2	3	4	3	4	3	4	3	4	4	1	0	0	0

データ ビュー 変数 ビュー

IBM SPSS Statistics プロセッサは使用可能です Unicode:ON クラシック

9

# シンタックス ウィンドウ

\*シンタックス1 - IBM SPSS Statistics シンタックスエディタ

ファイル(E) 編集(E) 表示(V) データ(D) 変換(T) 分析(A) グラフ(G) ユーティリティ(U) 実行(R) ツール(S) 拡張機能(X) ウィン

アクティブ データセット: データセット1

CRO

1	CRO
2	/TAB=Q2B BY Q10
3	/CEL=COL .
4	

IBM SPSS Statistics プロセッサは使用可能です Unicode:ON クラシック

シンタックスを書き換えて、名前をつけて保存すること。

10

## 命令文の実行

- ◆ シンタックスのQ2Aの部分をQ3にするなど、変数名を変えて実行してみる。  
Q2A をQ02A など変数名を間違えると動かない  
Q3 をQ3A など変数名を間違えると動かない
- ◆ 変数名が合っていれば動く。実行したい文をマウスで囲んでから▲ボタンで実行する。エラーが出た場合は変数名などを確認する。
- ◆ 書き換えたシンタックスは名前をつけて保存すること。保存場所をUSBメモリやホームディレクトリなどにする。

11

## シンタックスの利点

- シンタックスを書かなくても、画面上の「分析」をクリックし、記述統計 → クロス集計でも分析はできる。
- が、シンタックスでカテゴリー合併をしてから、クロス集計命令文を書いた方が分かりやすく、一度に大量に分析できる。
- recode文でのカテゴリー合併などを覚えることが大切。

12

## 新変数を作る例

資料図3の最後

新変数**JIEI**を作り、まず、全員を0にしている。  
新変数名はなんでもよい。

IF (**Q33**=6 OR **Q33**=7) **JIEI**=1.

スペルミスや変数名に注意

recode

record

13

## クロス集計シンタックスの例

CRO

/TAB = **Q18g** BY Q46SEX

/STA = PHI COR

/CEL =col .

ピリオドをつけすぎると動かない

**全角空白**を入れるとエラーになる

ファイル名やフォルダ名も半角文字のみがよい

14

### 1.3.データ行列とは何か

- ◆ 行が個人、列が変数となる数字の行列
- ◆ 具体例 3人分のデータ行列の例

```
00101 21412508 2421111111  
00102 21611402 1221213132  
00103 12714806 1222212121
```

- 5カラム目までがサンプル番号

- ◆ データ分析 ーデータ行列の分析  
これを分析して各種の表を作る

15

#### ◆ テキスト形式データ

SPSSシンタックスを使って読み込む。桁指定の  
命令文を書いて実行。

データがHドライブのsagyoフォルダにある場合は、  
シンタックスにて、以下のようにデータファイルの場所  
を書く。

H:¥sagyo¥15fukushima0714.txt

- ◆ SPSS形式データ、エクセル形式データ  
読み込みにはシンタックス不要。

16



## 作業手順

パソコンは拡張子が表示される  
設定にしておくこと

### ◆ SPS

### ◆ TXT

ZIPからこれら2つを、CやHドライブに出す。



### ◆ SAV :SPSS形式データファイル

ZIPから出す時、ドキュメントなどに出さない。Cドライブの  
どこかのフォルダなどへ。ZIPに入ったままにしない。

SPS, TXTの2ファイルがあればSAVファイルを作ることができる。

★シンタックスウィンドウ内に命令文を書けばよい<sup>7</sup>

## 2. データの場所(path)の書き方

### ◆ Cドライブのdap25フォルダ内にデータがある場合

シンタックスにて、

C:¥dap25¥ のあとにデータファイル名を書く。

例 C:¥dap25¥15fukushima0624.txt

¥ マークで区切ってフォルダの場所などを書く。  
大文字、小文字も区別するので注意。

### ◆ Windowsのドキュメントフォルダ内にある場合は、

C:¥Users¥murayo¥Documents¥ のあとにデータファイル名を書く。

ユーザー名murayoがの場合

### ◆ マックの場合は /Users/ 以降などフォルダ名を情報ウィンドウで 出す

Macにおいて、ファイルやフォルダの場所(フルパス)を取得するには

#### 1. ファイル情報を見る

txtファイルなどを選択し、コンテキストメニュー(右クリック、あるいはctrl+クリック)、またはメニューの情報を見る(Cmd+i)で、情報ウィンドウを表示させる。

その中の一般情報 > 場所がフォルダのパス。  
/Users/以降すべてが場所。

#### 2. テキストエディタにドロップ

テキストエディタを開き、txtのファイルなどを、そこに引っ張ってドロップすると、場所が出る。

19

パスとは何かを理解する。ファイルやフォルダがある場所(住所)のこと。Windowsでは ¥で区切って書く。

◆ 分からなければ「**ファイル** **パスとは** **MacOS**」などでグーグル検索すること。

・Mac OS のFinderでパスを表示する場合は、Finderを開いて、画面上「表示」→パスバーを表示 を選択。

・ファイルを右クリックした後にoptionを押すと「パス名をコピー」メニューが出る。

◆ **Documents**フォルダ内の**data15.txt**ファイルのパスは

¥Users¥murayo¥Documents¥data15.txt

murayoの部分はユーザー名、¥はバックスラッシュ

20

### 3. データ分析法の種類

- ◆ 単純集計とクロス集計
- ◆ 平均値や分散を見る(基本統計量)
- ◆ 多変量解析(3つ以上の変数を用いて分析)
  - 重回帰分析  $Y = X_1 + X_2 + e$
  - 分散分析
  - 因子分析、クラスター分析、MDS
  - 共分散構造分析(重回帰分析と因子分析の発展版) SEM ソフト名はAMOS

21

#### ◆ 分散分析

**カテゴリー間**の平均値の比較

F検定を行う。2つの平均値の比較であるt検定と結果は一致する。

Yは量的意味がある変数、Xは質的変数(カテゴリー)

#### ◆ 重回帰分析

ある連続変数Yの違いを、他のXにより予測(説明)する。

YもXも連続変数(**量的意味がある**)。

#### ◆ 質的変数とは何か

22

#### ◆ 因子分析

変数の背後にある要因(因子)を探るためのもの。  
すべて連続変数。

とくにYやXを設定しない。複数のXを、いくつか  
にまとめるためにやる、と言ってもよい。

これら3つの分析法が、よく使われる。

23

### 3.1. クロス集計とは

例 男女別の原発賛成率を出す

#### ◆ 変数は2つ

- 回答者性別 2段階回答
- 原発への態度 4段階回答

#### ◆ 2変数の関連とは

これらはカテゴリー変数

例 Q18a 原発への態度  $2 \times 4$  の表を作る

以下、変数名は2015年福島市調査の例

24

## シンタックス例

CRO

/TAB = Q9a BY Q46sex

/STA = PHI COR

/cells =COL .

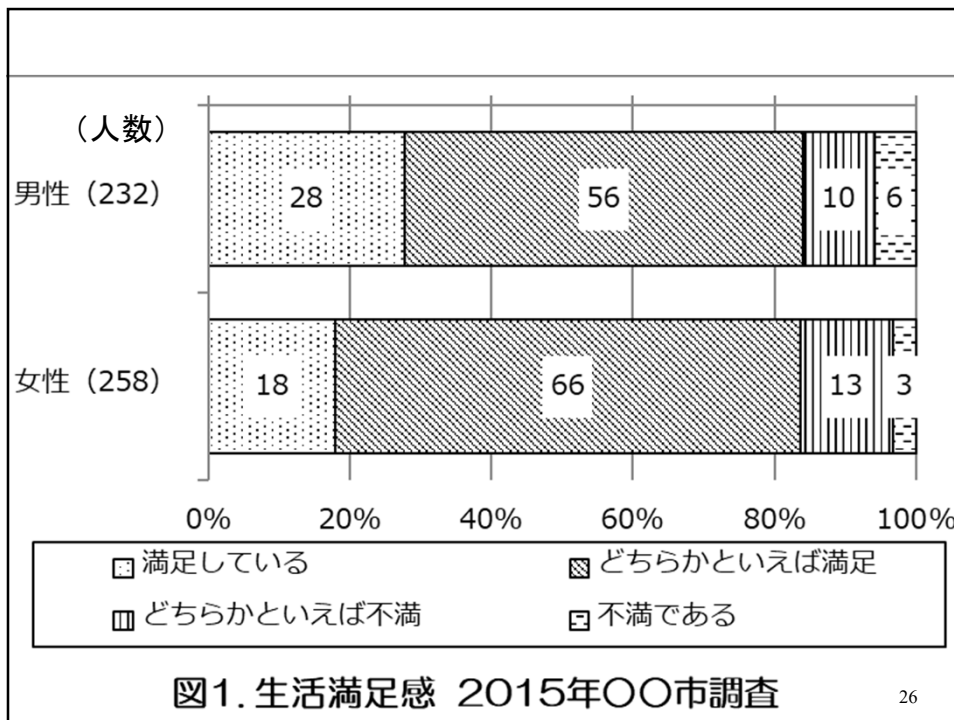
/\*\* カテゴリー合併した新変数 N10 の作成 \*\*/

COMPUTE N10 = Q10.

RECODE N10 (1,2=1)(3,4=2) .

Q10が元の変数(データファイル内にある)、N10が新しく作ったもの

25



26

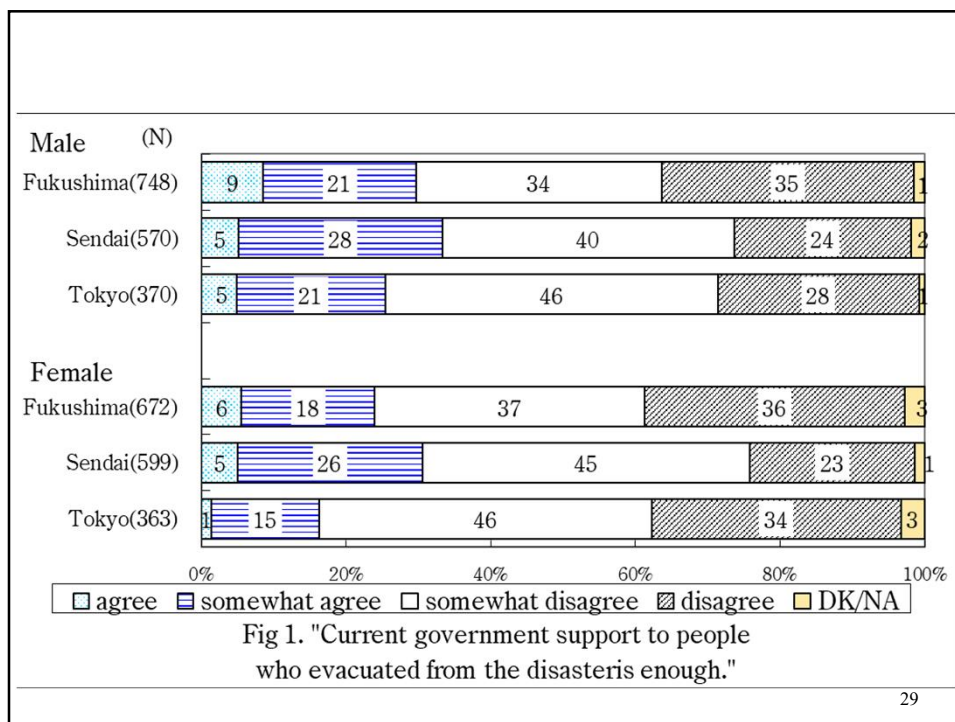
## 3.2. グラフ形式

- 調査名など適切なタイトルを図下につける。  
質問文が具体的に分かるよう詳しく書く。  
タイトルに図1 など番号を付けること。
- データラベルをつける。足して100%になるか注意。
- 100%の人数を書く。セルに男性(232)などとかく。
- 白黒印刷の時に分かる模様がよい。コピーして模様が  
つぶると、データラベルが見えない、などに注意。
- 3次元グラフは分かりにくいので使わないこと。
- この形式だと中央値がどこかが一目で分かる。

27

- グラフの中のデータラベルは、28%でなく28が見やすい。
- エクセル上で、セルの中を28にする。28のセルの近くで、100倍すれば良い。例えば、d51セルに%がある場合、半角文字で  $=d51*100$  と書く。
- 図タイトルは図下、表タイトルは表の上につける。
- 以下はよいグラフの例。

28



### 3.3.クロス集計の分析手順について

- ◆ まずYを決めて、因果関係について考える  
何が原因Xとなるか、候補を考える。
- ◆ 調査の原データファイルを入手して、分析。
- ◆ 必要ならばカテゴリー合併をして新変数を作り分析。例えば、回答者の年齢3段階の変数など。
- ◆ 資料や『社会調査演習』エラボレーションを見て、適切な形式の表を作ること。

- ◆ 表面的な関連が、因果関係とは限らない
- ◆ 分析の結果、出てきた関連の背後に、どのような要因がありうるか、考えてみるのが大切。
- ◆ 第三の要因の候補は、無限にありうる。

31

## 今日のポイント 6/13

- シンタックスを使って新変数を作る
  - 自分が作ったシンタックスを保存すること
  - 適切なカテゴリー合併をする方法とは  
→ まず分布を確認
- ・自分が作った新変数を使って分析する
- ・学歴や職業カテゴリーを作る
- 無関連とは何かを理解する

32



### 3.4. 社会調査について

- ◆ 調査票を作る工夫
- ◆ 複数回答
- ◆ 量的意味がない変数(カテゴリー変数)とは  
    関連係数 → クラマーV  
    **前回の配付資料を確認**

33

### 3.5. 適切な形式のグラフを作る

- ◆ まず、SPSS出力をエクセル形式で保存  
    出力画面で ファイル→名前をつけて保存  
    ファイル→エクスポート「**エクセル形式保存**」
- ◆ エクセル上にて、グラフを作成するとよい
  - セルの書式設定
  - セル結合はしない

34

### 3.6. カテゴリー合併をする練習

Occupational Status

p.38, 138

経営者、正規雇用、非正規、自営、無職に分ける

```
compute status = 0.  
recode (sv =1) status =1.  
recode (sv =2) status =2.  
recode (sv =3) status =3.  
recode (sv =4) status =3.  
recode (sv =5) status =3.  
...
```

sv が元の変数、 status が新変数。

svの部分に、適切な変数名を入れる。

従業上の地位の変数は、調査により異なる

35

## シンタックスのこつ

- カテゴリー合併をする練習  
recode文、if文を使いこなす
- 元の変数の分布をみる  
→どう分けるか自分で判断する  
年齢2カテゴリーにしても、どこで分けるかなど
- 新変数が正確にできているか調べるのが大切！
- 元の変数と新変数のクロス集計をする

36

## 4. 無関連とは何か

### ◆例 リサイクル商品への意識

男 賛成70, 反対30

女 賛成70, 反対30

→ 無関連

- 現実には、上記の横棒グラフのように、分布に偏りがあることが多い。
- 無関連と、現実の表の、2つの間の距離をもとに計算したものが、カイ二乗値

37

## カイ二乗値とは何か



現実と無関連との距離



2つの表を作り、差を出す

38

## クロス集計表の例

	北部	南部	計
白人			140
黒人			60
計	80	120	200

### 例 現実の分布

	北部	南部	計
白人	68	72	140
黒人	12	48	60
計	80	120	200

39

## 練習問題

### 無関連状態

	北部	南部	計
白人			140
黒人			60
計	80	120	200

自由度は1

### 最大関連状態

	北部	南部	計
白人			140
黒人			60
計	80	120	200

40

## 回答

無関連状態			期待度数 $F_{ij}$
	北部	南部	計
白人	56	84	140
黒人	24	36	60
計	80	120	200

→横にみると2:3 になっている

$n_{11}$ について 140人の40%  $140 \times 80 / 200 = 56$

最大関連状態			
	北部	南部	計
白人	80	60	140
黒人	0	60	60
計	80	120	200

41

## 回答 %に直すと

無関連状態			行%
	北部	南部	計 (人数)
白人	40	60	100 (140)
黒人	40	60	100 (60)
計	40	60	100 (200)

→ 2:3 になっている

---



---



---

42

## カイ二乗値の計算

無関連状態 $F_{ij}$		
	北部	南部
白人	56	84
黒人	24	36
計	80	120

→ 2:3 になっている

例 現実の分布 $n_{ij}$		
	北部	南部
白人	68	72
黒人	12	48
計	80	120

$n_{11}$ について  $68-56=12$

2つの表の内容を比べて差を出す → 差を二乗して合計 <sup>43</sup>

## 分析の基本とは 9/27

- ◆ ばらつき具合を把握すること
  - 平均値からの距離dを理解すること
- ◆ 例 身長の平均値160
  - 175センチの個人は $d=15$
- ◆ 標準偏差 standard deviation
  - d の標準的な値 (平均値)
  - 全員分のdを二乗して合計しn-1で割る

## 偏差値は平均=50に直した値 平均から1標準偏差離れた人は40または60

例 百万人が受けたテストの場合

- ・ 偏差値60（あるいは40）以上は15.87%  
標準偏差の1倍以上の人は 158,700人
- ・ 偏差値70（あるいは30）以上は2.28%  
標準偏差の2倍の人は 22,800人

◆ 『社会調査演習』 p. 32の図

真ん中の平均値付近は人数が多い

標準偏差2倍以上は全体の2.28%しかいない

両端を合わせると4.56%

◆ SDの1.959倍だと全人口の5%になる p. 74図の面積の両端

45

## 2×2表の関連係数

無関連は0、最大値が1か-1

- 四分点相関係数 $r$
- Q係数

2×2表ならば、基本的に $r$ を使えば良い。

46

## 2×3表以上の関連係数

- クラマーのV
- タウb
- タウc

数値に量的意味がない場合はクラマーのVを使う。

47

## 5.カテゴリー合併とは 『SPSSによる多変量解析』p.38

◆ 4カテゴリー → 2カテゴリー

◆ 新変数を作る     Q08 → N08など

Q8a Q08a Q8 などデータにより変数名は異なる。新変数名N08などは、好きな名前が良い。

◆ まず、欠損値処理を忘れずに

48



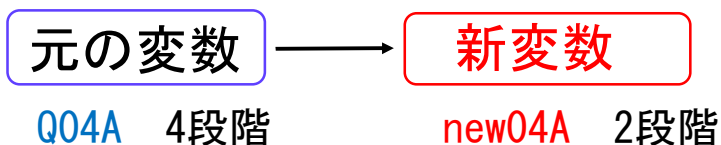
## SPSSによる新変数の作成

SPSSでデータを開いた状態で、シンタックスを  
実行

- 適切なカテゴリー合併をする。
- シンタックスに全角文字は入れない。
- まず、縦か横%のみを出す。
- 全体%や人数の表は意味が分からない。
- グラフは、図のタイトルは詳しく。質問文など。
- 100%の人数が分かるように。
- 新変数を作り、元の変数とのクロス集計を出す

49

### 5.1.シンタックスで新変数を作る ノートを書く！



- 元の変数の名前を正確に書く。
- RECODE文 ()内はカンマ,
- **RECORD**などだと動かない
- ピリオド. は最後に一つだけ書く。

50

## RECODE文の例

## 紙の資料 例1

- ◆ シンタックスウィンドウに以下を書いて実行。
- ◆ ピリオドをつける、全角空白は入れない。
- ◆ イコールの右側が元の変数

```
COMPUTE      NEN4 = Q46age.  
RECODE      NEN4 (10 THRU 39 =3) (40 THRU 49 =4)  
              (50 THRU 59 =5) (60 THRU 98 =6).  
MISSING VALUES NEN4 (99).
```

CRO

```
/TAB= Q9C by NEN4  
/STA=CHISQ PHI CORR BTAU CTAU  
/CEL= COL.
```

かっこが全角、ピリオド付け忘れ、は動かない

51

## 5.2.年代の新変数作成

```
COMPUTE NENDA2=Q46AGE.  
RECODE NENDA2 (10 THRU 45 =1) (46 THRU 98 =2) .  
CRO
```

```
/TAB= N16 by NENDA2  
/STA=CHISQ PHI CORR BTAU CTAU  
/CEL= COL.
```

かっこが全角、ピリオド付け忘れ、は動かない

今日のポイント

パスとは何か、シンタックスの操作

52

## ラベルをつけるには

VALUE LABELS **NENDA2** 1 '45以下' 2 '46以上'  
' 9 'わからない' .

' ' で囲む、全角かっこなど付けない。

## 横棒グラフのデータラベル

32.6% でなく33

エクセルでセルを右クリックし、セルの書式設定

→ 表示形式タブ → ユーザー定義 → 0

有効桁を考えると小数点以下は無駄。

53

## 新変数を作った時

元の変数とのクロス集計、問題ないか確認！

### ◆今日の課題

- ・平均値の折れ線グラフ
- ・適切な三重クロス集計表を作る
- ・逆転した変数や、教育年数の変数を3つ以上作り、相関correlationを出す

54

## 関連とは何か

- ◆ クロス集計表(離散変数、カテゴリー)の関連 → %の差を見る、離散変数の関連係数を使う(前述)。
- ◆ 量的変数の関連 → 散布図を見る、相関係数の考え方を理解する。

55

## ピアソンの積率相関係数 $r$

レンジは0～±1

点の散らばり具合をあわらす

テキストp.142 散布図

COR

/VAR=Q2A Q2B Q2C age.

年代でなく、できるだけ細かい変数、年齢などを使う

量的変数は、細かい変数を使うこと！

56

## 資料3. 分析命令文 6/20

- 学歴を教育年数(量的変数)に直す 例3
  - 変数逆転 例6
- 平均値を出す p.4 一元配置分散分析  
テキストp.99

```
ONE    Q2B EDU BY    nendai  
      /PLOT MEANS  
      /STA DES.
```

BYの後は、説明変数Xになるものを1つだけ書く。前は、複数の変数を書いても良い。

被説明変数Yは、量的意味があるものを使う

57

## 平均値の比較

### 一元配置分散分析

```
ONE    YARITAI BY    sin8job  
      /PLOT MEANS  
      /STA DES.
```

BYの後は、説明変数Xになるものを1つだけ書く。前は、複数の変数を書いても良い。 EGP6c

被説明変数Yは、量的意味があるものを使う

58

- ◆ 性別でデータを分割 資料 11.
- ◆ 折線グラフの縦軸 → 軸の書式設定  
適切な縦幅にする  
タイトルに数字の意味をかく
- ◆ 資料 8. クロス集計とエラボレイション  
3つの変数の因果関係を理解する  
表形式は別紙資料参照

59

### 5.3.階層帰属意識のカテゴリー合併

- ◆ 「上」と「中の上」を合併
- ◆ ほかの合併は自分で考える

60

## 5.4. 学歴カテゴリー作成

= の右側が元の変数

```
COMPUTE  EDUCT4 =Q43.  
RECODE  EDUCT4 (1, 2, 3=1) (4=2) (5=3) (6=4) (7, 9=9).  
MISSING VALUES  EDUCT4 (9).  
VALUE LABELS EDUCT4 1 '高卒以下' 2 '専門学校'  
                3 '短大高専' 4 '大卒以上' 9 'わからない' .  
  
CRO  
/TAB= N16 by EDUCT4  
/STA=CHISQ PHI CORR BTAU CTAU  
/CEL= COL.
```

かっこが全角、ピリオド付け忘れ、は動かない

61

```
COMPUTE  EDU3 =Q43.  
RECODE  EDU3 (1, 2, 3=1) (4=2) (5=2) (6=3) (7, 9=9).  
MISSING VALUES  EDU3 (9).  
  
VALUE LABELS EDU3 1 '高卒以下' 2 '専門短大高専'  
                3 '大卒以上' 9 'わからない' .
```

’ ’ で日本語を囲む、全角かっこなど付けない。

旧制の教育制度については原・海野(2004)p.204参照。

62

## 5.5.クロス集計のシンタックス例

CRO

/TAB= Q2A by EDU3

/STA=CHISQ PHI CORR BTAU CTAU

/CEL= COL.

- ◆ /CEL行を =row にすると横%が出る。
- ◆ =COU にすると実数が出る。

63

## 三重クロス集計のシンタックス例

CRO

/TAB= Q30 by JIEI by Q46sex .

CRO

/TAB= Q30 by EDU3 by Q46sex .

CRO

/TAB= Q30 by EDU3 by Q46sex

/STA=CHISQ PHI CORR BTAU CTAU

/CEL= COL.

- CEL行がないと人数が出る。
- カテゴリー合併しないと、各セルの人数が少なくなるので注意する。まず合併した新変数を作る。

64



- ◆ 学歴や性別によって、問30の健康状態は異なる。
- ◆ しかし、背後に年齢など、別の要因が存在するかもしれないので、注意する。
- ◆ 結果を、自分で解釈することが大切。表面的な関連が、真の因果関係ではない。

65

## 5.6.職業大分類の作成

- ◆ JGSSやSSP調査の見本シンタックスを参照
- ◆ 各自で見本シンタックスを保存すること。

66

## 5.7.職業威信スコアとは

- ・狭義の職業(本人の仕事内容)を評価した結果の平均値。
- ・95年調査では5段階評価。全員が真ん中に答えた場合は50になる。

### 問題点

- ・職業を評価できるか
- ・第一次産業は少なくなっている、農業の評価は上がってくる、などがありうるか。稼いでいる農家もある。
- ・規制緩和、銀行員の給料が減る、などの影響はありうるか

67

学歴カテゴリーごとの威信スコア **平均値**  
現職、初職、父職、配偶者職  
Yは量的変数。

ONEWAY **GEN** BY **EDUCT**  
/STA DESCRIPTIVES EFFECTS  
/PLOT MEANS.

ONEWAY **GEN** BY **NENDAI**  
/STA DESCRIPTIVES EFFECTS  
/PLOT MEANS.

COR /VAR = **GEN SYO JOBF JOBS.**

68

## 平均値や相関係数を出す前に

- ◆ 欠損値処理をしているか確認！
- ◆ 量的変数でないと、平均値を出しても意味がないことに注意。
- ◆ 分散分析のXはカテゴリー変数なので、あまり細かいと良くない。ageでなくNENDAI.
- ◆ Yはできるだけ細かい変数にする。

69

## 6. If文による変数加工 6/16

例 自営業なら1, そうでないなら0とした変数

- ◆ まず 新変数を作り、値は全員が0とする。
- ◆ ある問の値が6の場合は、新変数の値を1とする。
- ◆ ()内に、元の変数に関する条件を書く。

```
COMPUTE      JIEI  =0.  
IF (Q33=6)   JIEI  =1.  
IF (Q33=7)   JIEI  =1.
```

70

```

COMPUTE  HISEIKI =0.
IF (Q33=3 OR Q33= 4) HISEIKI =1.
IF (Q33=5 OR Q33= 8) HISEIKI =1.

IF ((Q33=3 OR Q33= 4) OR
(Q33=5 OR Q33= 8) ) HISEIKI =1.

CRO
  /TAB= Q30 by HISEIKI by Q46sex .
CRO
  /TAB= Q30 by JIEI by Q46sex .

```

71

## 例2 自営業ダミー変数作成 ISSP調査

```

COMPUTE          JIEI  =0.
IF (EMPREL > 1)  JIEI  =1.
MISSING VALUES  JIEI (8,9).

```

```

CRO
  /TAB =EMPREL BY JIEI.

```

課題 何らかのダミー変数を使って三重クロス集計をする

72

## 例 既婚ダミー変数作成

```
COMPUTE          MADMY =0.
IF (Q28 =1)      MADMY =1.
MISSING VALUES MADMY (9).
CRO
  /TAB = Q28 BY MADMY.

CRO
  /TAB= Q30 by MADMY by Q46sex
  /STA=CHISQ PHI CORR BTAU CTAU
  /CEL= COL.
```

73

## Occupational Status p.38

```
Compute jobstats= q33.
Recode jobstats (1=1)
(2=3)(3,4,5=4)(6,7=5)(else=9).
IF ((q34>5) and (q36 = 4)) jobstats=2.
IF ((q34>5) and (q36 = 5)) jobstats=2.
IF ((q34<5) and (q33 = 1)) jobstats=5.

VALUE LABELS jobstats 1 '経営者' 2 '管理
職' 3 '常時雇用' 4 '非正規雇用'
5 '自営・家族従業者' 9 '無職' .
```

74

## 三重クロス集計のシンタックス例

CRO

/TAB= Q16 by MADMY by Q46sex .

CRO

/TAB= V2 by JIEI by Q46sex .

CRO

/TAB= Q10 by MADMY by Q46sex

/STA=CHISQ PHI CORR BTAU CTAU

/CEL= COL.

- /CEL= COL COU .
- だと縦% と 人数 が出る。ROWは横%

75

## 基礎的分析

- ◆ 年齢別や都市規模別のクロス集計をすること。
- ◆ クロス集計の前に、適切なカテゴリー合併を行う。必ずシンタックスを使う。
- ◆ グラフ形式は、形式見本ファイルをよく見る。
  - ・データラベルは足して100になる%を書く
  - ・グラフタイトルを詳しく、具体的に。質問文などが分かるように。
  - ・100%の人数を書く。

76

## クロス集計の一般形

『SPSSによる多変量解析』 p.72

### ◆ 統計的独立とは何か p.74

無関連状態とは何かを理解する。

現実の表と、**独立の表**を比べて計算したものがカイ二乗値 (4.6)式

$\hat{F}_{ij}$  えふはつとアイジェイと読む

### ◆ 完全関連 p.77

### ◆ 独立性の検定 p.81 $H_0$ は無関連。

無関連と**現実との距離**について考える。現実のクロス表は、無関連状態から、どのくらい遠いか

77

## 7. 変数の計算

### ◆ 足し算やかけ算など計算できる。**+** **-** **\*** **/** を用いる。

例6 問4Aの変数の回答を逆転。N4Aが逆転した新変数だということを理解する。

Compute N1C=5-V3.

Compute N4A=5-Q4A.

1, 2, 3, 4 → 4, 3, 2, 1となる。

FRE var = V3 N1C.

例7 問7Aと問7Bの回答内容を足して、合計得点の新変数NEW7を作成する。

Compute NEW7=Q7A+Q7B.

★計算の前に、欠損値処理をしておくこと。

78

変数の逆転、連続量の相関係数 p.142  
量的変数でないと相関は出せない。

Compute N4=5-V4.

Compute KUNINO = 5-Q2C.

Compute N9B = 5-Q9B.

CRO /TAB= N9B by Q9B.

COR

/VAR = AGE KUNINO N9B

/STA DESCRIPTIVES .

79

復習 年代の新変数作成 6/23  
量的変数について平均値をだす。

COMPUTE NEND4=Q46AGE.

RECODE NEND4 (10 THRU 39 =3)

(40 THRU 49 =4) (50 THRU 59 =5) (60 THRU  
98 =6) .

ONEWAY KOFUKU N9B BY NEND4

/PLOT MEANS

/STA DES.

一元配置分散分析。Xが一つの平均値比較。 BYの左がY。 か  
っこが全角、ピリオド付け忘れ、は動かない

80



## 例7 合計得点の新変数

Compute **N9B** = 5-Q9B.

Compute **N9C** = 5-Q9C.

Compute **KAKUSU** = N9B + N9C.

ONEWAY **KAKUSU** BY **NEND4**

/PLOT MEANS

/STA DES.

一元配置分散分析。Xが一つの平均値比較。BYの左がY。 か  
っこが全角、ピリオド付け忘れ、は動かない

81

## シンタックスの流れ

- ・データファイル読み込み
- ・欠損値処理
- ・カテゴリー合併やIF文など
- ・クロス集計など**分析命令文**

見本シンタックスをしてみることに。

すべて実行すれば、一度にたくさん結果が出る  
シンタックスを書けば、大量に何度も処理できる。

82

## 復習 学歴カテゴリー作成 6/23

COMPUTE ED2 =Q43.

RECODE ED2 (1, 2, 3, 4=1) (5, 6=2) (7, 9=9).

MISSING VALUES ED2 (9).

VALUE LABELS ED2 1 'high' 2 'college'  
3 'DKNA' .

’ ’ で日本語を囲む、全角かっこなど付けない。

旧制の教育制度については原・海野(2004)p.204参照。

83

## 学歴カテゴリーごとの平均値 量的変数でないと平均値は出せない

Compute N4A=5-Q4A.

ONEWAY N4A BY ED2

/STA DESCRIPTIVES EFFECTS

/PLOT MEANS.

ONEWAY N4A BY NENDA2

/STA DESCRIPTIVES EFFECTS

/PLOT MEANS.

84

## 8. 各種の関連係数

- ◆ 量的変数の相関 $r$ とは
  - $r$ の定義とは
    - $x$ と $y$ の、平均値からの距離2つをかけている。
- ◆ 相関と回帰の違いとは
  - 回帰係数は、線の傾き
    - 別紙の回帰分析解説を参照
- ◆ 質的変数の関連 ... 別紙資料 安田・海野『社会統計学』

85

### 2×2表の関連 $0 \sim 1, -1$

- 四分点相関係数 $r$
- Q係数

2×2表ならば、基本的に $r$ を使えば良い。

### ◆ 2×3以上の表の関連 p.78

- クラマーのV 例えぱりんご1 みかん2 ばなな3  
値に量的意味がない場合。  $\phi$ 係数の絶対値と一致
- タウ $b$ (3×3表など)、タウ $c$ (3×4など対象でない表)  
値に量的意味がある場合 4段階回答など

86

## 分析結果のまとめ方 p.333

- ◆ (1)～(4)
- ◆ この後に、重要な変数について、さらにクロス集計やエラボレーション結果を載せる。

87

## 移動表について

8×8表を、3×3に直す

- ◆ 専門と管理
- ◆ 事務と販売
- ◆ それ以外

その上で、粗移動率を計算する

88

## 無関連＝独立とは

### ◆ 循環移動量の計算

計の数字の少ない方－非移動数

- ◆ 完全移動 ＝独立状態、理想状態  
循環移動量が最大

$$\text{開放性係数} = \frac{\text{現実の循環移動量}}{\text{独立状態の循環移動量}}$$

89

## オッズとは

- 分子 事象が起きない度数
- 分母 事象が起きる度数
- 例 馬が勝つと1割の人が考えて馬券買った  
→ 9/1 オッズは9

90

- 階層A出身者でAになった40 それ以外60  
→  $40/60 = 0.66$
- 階層B出身者だがAになった20 それ以外80  
→  $20/80 = 0.25$
- オッズ比 2つのオッズの比  
 $0.66 / 0.25 = 2.64$

階層が3つ以上の場合、2カテゴリーに合併し計算。

- 問 以下の例でオッズ比を計算せよ

別紙 内藤 表1 パネル

91

問  
以下のオッズ比は

		本人		
		A	B	計
親	A	16	24	40
	B	24	36	60
	計	40	60	

92

◆ 表1

- 職業Aのオッズ 16/24
- 職業Bのオッズ 24/36

→  $2/3$  と  $2/3$  の比

93

問 以下のオッズ比は

表2 閉鎖性が高い社会

		本人		
		A	B	計
親	A	10	20	30
	B	10	60	70
	計	20	80	

94

### ◆表 閉鎖性が高い社会

- 職業Aのオッズ 10/20
- 職業Bのオッズ 10/60

→ 1/2 と 1/6 の比

- 職業Aの子は、上層に入れる確率が3倍

95

## 無関連の例

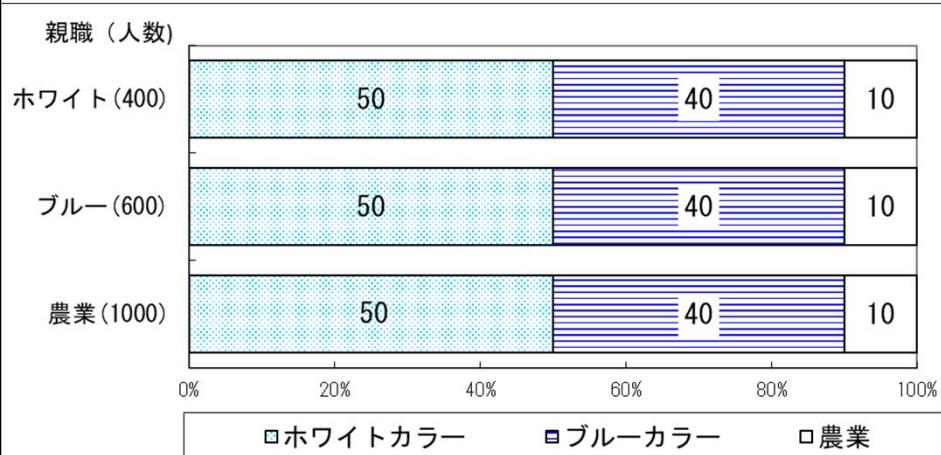


図3 社会移動表の例

96



## 信頼性(reliability)

- ◆ 尺度の安定性や一貫性。繰り返し測定しても同じ結果になるという意味。
- ◆  $\alpha$ 係数  
同じ尺度の内的一貫性を見ている。
- ◆ 再テスト法 一定の期間をおいて再度測定し、それらのスコアの相関を見るもの。
- ◆ 折半法 変数を2分割して、2つのテストと見なして信頼性を検討する。

97

## 9. エラボレイションとは

- ◆ 三重クロス集計による因果関係の検討。二重クロス集計をしてから、第三変数Zを導入する。
- ◆ 分割後の係数を含め、5つの関連係数を出す。
- ◆ シンタックスcro  
Yになるものを最初に書く Y by X by Z

98

# 因果関係についてよく考える

Yの原因として何があるか。

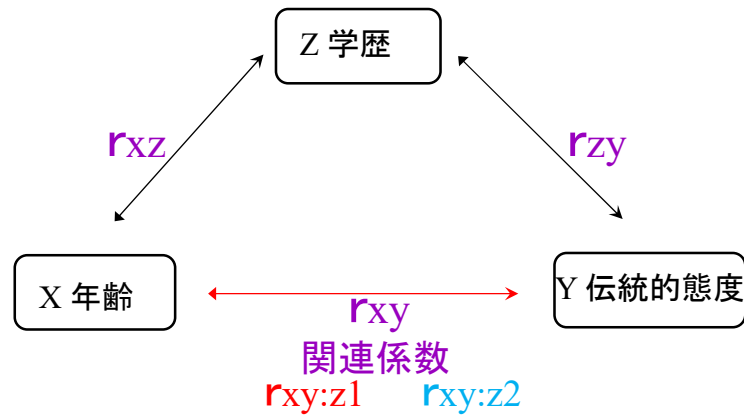


図1. 三重クロス集計による因果関係の検討

もし、最初にあった関連 $r_{xy}$ が、Zで分割した時になくなった場合は、XとYの間に直接効果なし、という結論になる。

99

## クロス集計とエラボレイション p.83

例えば、二重クロス集計の結果、XとYに関連があったとする。

例 婚姻Xと食べ物の好みY

初めは、この部分の関連しか見えない

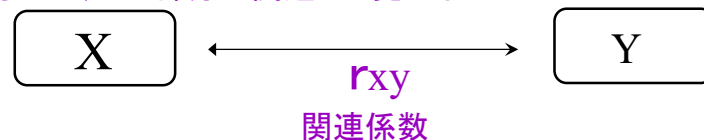


図2. 二重クロス集計における関連

この最初に出た関連 $r$ が、他の変数の効果を考慮しても、ほんとうに関連があるかは分からない。

100

まず、二重クロス集計を3つやる。  
 第三変数Z(X2)は年齢とする。

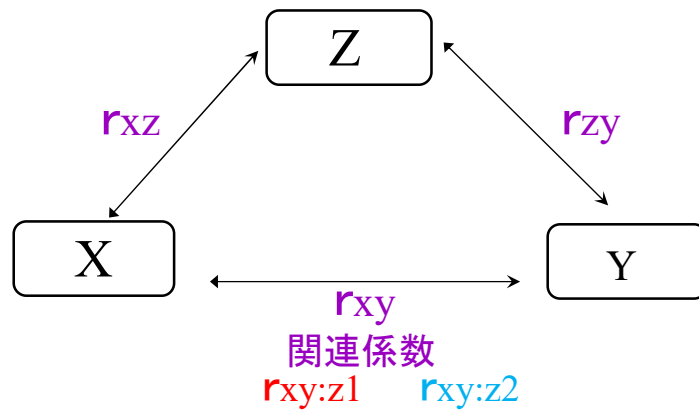


図3. エラボレイションにおける関連

次に、Z=1の時のr、Z=2の時のrを出す。  
 元の関連係数を含め5個の関連係数がある。

101

まず、二重クロス集計を3つやる。  
 第三変数Zは性別とする。

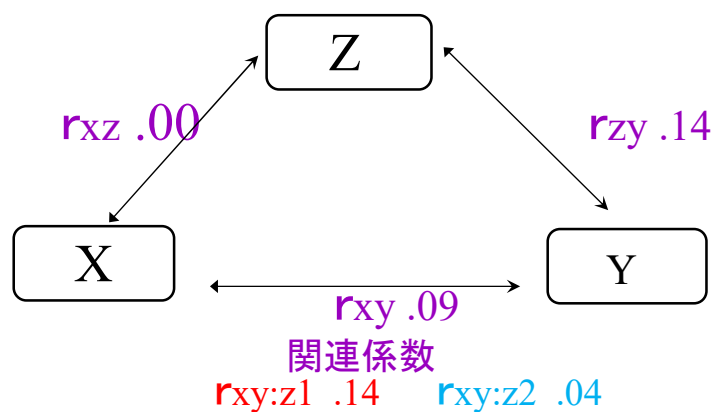


図4. エラボレイションにおける関連  
 『社会調査演習』のタイプ4の例

102

まず、二重クロス集計を3つやる。

第三変数Zは性別とする。

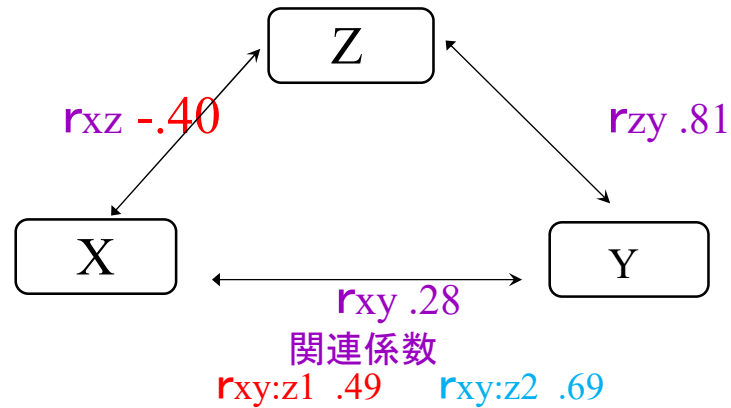


図5. エラボレーションにおける関連  
『社会調査演習』のタイプ5の例  
初めに見えていた関連.28が、第三変数Z導入後はより大きくなった例

103

## 例えば、第三変数Zが年齢の場合

- 高年齢に絞った場合のXYの関連係数と、
- 低年齢に絞った場合のXYの関連係数の、2つの関連を出せばよい。
- 元々のXYの関連と、違いがあるかについて検討する。それを元に、因果関係について考える。

104

### 第三変数Zの導入による因果関係の探索

つまり、Zが低年齢、Zが高年齢の時の、XYのクロス集計も行う。3つの関連係数を出す。

$r_{xy}$  ,  $r_{xy:z1}$  ,  $r_{xy:z2}$

例  $z1$ 低年齢、 $z2$ 高年齢

これら3つが同じ値の時 → Zの効果がない

$r_{xy:z1}$  ,  $r_{xy:z2}$

これら2つがゼロの時 → Zの効果が大きい

(『社会調査演習』2.4のキャンディーの例)

何をZにすると良いか、複数を挙げて、試行錯誤するのがコツ。

105

### 他の例

◆ 例 Xが回答者年代、Yが生活満足感  
Zとして、回答者学歴を導入

- Z=1 低学歴
- Z=2 高学歴

Z=1の時のみ、XとYの関連が大きかった、という場合、Zを使った意味があった、ということになる。

Zの値によって、XとYの関連が変わらない場合は、他の変数Zの影響力を考慮してもなお、XとYは関連がある、ということを確認したという意味がある。

106

## クロス集計のシンタックス例 二重クロスは3回やる。

三重クロス集計だと1セルの人数が減るので、まず合併した方がよい。その後、分析

CRO

/TAB=N10 by N2A

/STA=CHI PHI COR .

CROSSTABS

/TABLES= N10 by N2A by N13

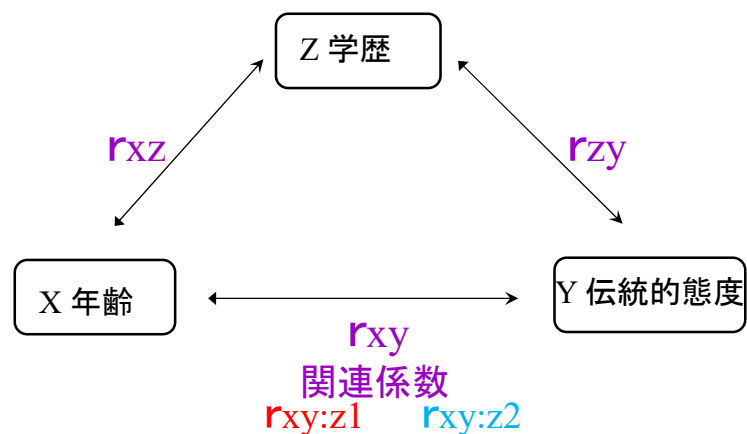
/STATISTICS=CHISQ PHI CORR

/CELLS=COL.

107

## 因果関係についてよく考える

Yの原因として何があるか。



もし、最初にあった $r_{xy}$ が、分割時になくなった場合は、年齢XとYの間に直接効果なし、という結論になる。最初の関連は疑似相関だった。

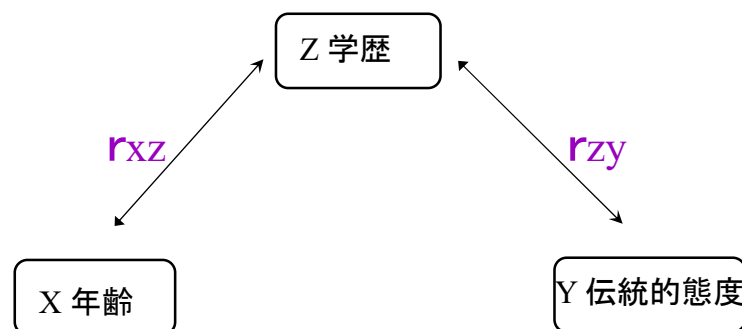
108

## エラボレイションの結果をどう解釈するか

- ◆ Z(X2)を分析に投入した結果として、XYの関連がなくなった場合、Zが真の原因と考えることができる。
- ◆ しかし、XYの関連がなくならなかったとしても、因果関係の確認という意味がある。
- ◆ 詳しくは原・海野『社会調査演習』のエラボレイションの解説をよく読む。p.88のような表を自分で作る。関連係数を5個出すことを理解すること。

109

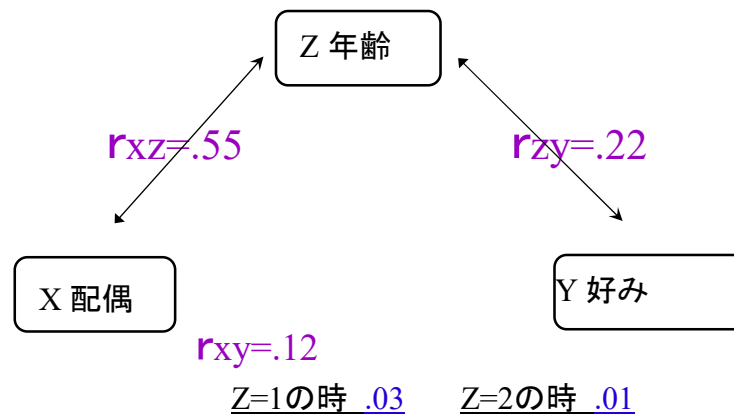
Yの原因とは何か。因果関係を考える。



もし、最初にあった $r_{xy}$ が、Zを投入した時になくなった場合は、ZがYの真の原因、という結論になる。

110

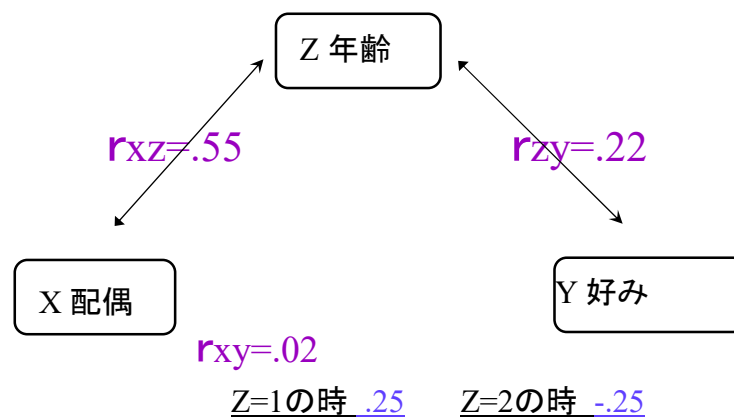
『社会調査演習』 p.88



もし、最初にあった $r_{xy}$ が、Zを投入した時に0に近くなった場合は、**ZがYの真の原因**、と解釈できる。

111

疑似無相関とは 『社会調査演習』 p.90



最初の $r_{xy}$ がゼロだが、Zを投入した時は、1つがプラス、もう1つがマイナスだった。

112



## エラボレーションの結果

- ◆ 3重クロス集計の結果をもとに、因果関係を考えるので、他の変数の効果を統制したわけではない。その意味で、因果関係を証明したとは言えない。
- ◆ ただし、現実の比率の差について、クロス集計表の形で明確にして、因果関係を確認する手がかりを得るために、分かりやすいし意味がある。
- ◆ もちろん、第三変数 $Z(X_2)$ の候補は、無限にあるので、そこを考えるのが難しい点ではある。

113

## シンタックス例 カテゴリー合併

4段階の変数を2段階にする。

compute N2 = Q2.

recode N2 (1,2 = 1) (3,4 = 2).

CRO /TAB= N2 by Q2.

CRO /TAB= N2 by N8

/STA=CHISQ PHI CORR BTAU CTAU

/CEL= COL.

114

## 10. 変数の逆転、連続量の相関係数 6/20

Compute KUNINO = 5-Q2C.

Compute N9B = 5-Q9B.

CRO /TAB= N9B by Q9B.

COR

/VAR = AGE EDU KUNINO N9B

/STA DESCRIPTIVES .

115

## 国を絞る ISSP18の例

select文によりデータ人数が減るので注意

/\*\* 国を3国に絞る \*/

compute datac1 =0.

IF (country=158) datac1 =1.

IF (country=392) datac1 =1.

IF (country=410) datac1 =1.

select (if datac1) =1.

/\*\* クロス集計 \*/

CROSSTABS

/TABLES=country BY v5

/CELLS=COUNT .

/\*\* データを分割 \*/

SORT CASES BY country.

SPLIT FILE SEPARATE BY country.

116

```
FRE VAR=MARITAL .
```

```
COMPUTE      JIEI =0.
```

```
IF (EMPREL > 1) JIEI =1.
```

```
MISSING VALUES  JIEI (8,9).
```

```
CRO
```

```
/TAB= V2 by JIEI by sex .
```

```
/** 三重クロス集計 **/
```

```
CRO
```

```
/TAB= V2 by JIEI by sex
```

```
/STA=CHISQ PHI CORR BTAU CTAU
```

```
/CEL= COL.
```

117

## 11. 重回帰分析の考え方

- ◆ YもXも量的な意味がある変数(できるだけ細かい変数)

- ◆ まず、YとXの関連を考える  
例

父親の体重 → 本人の体重

1000人分の散布図を作ったらどうなるか

118

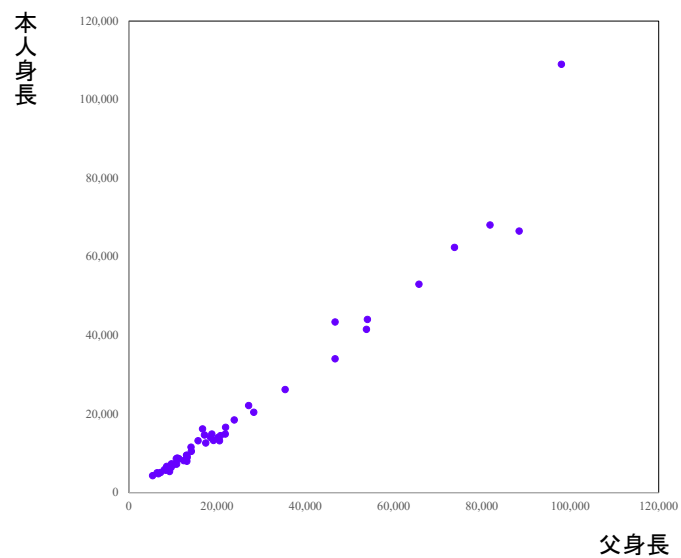


図5. 二変数の関連についての散布図

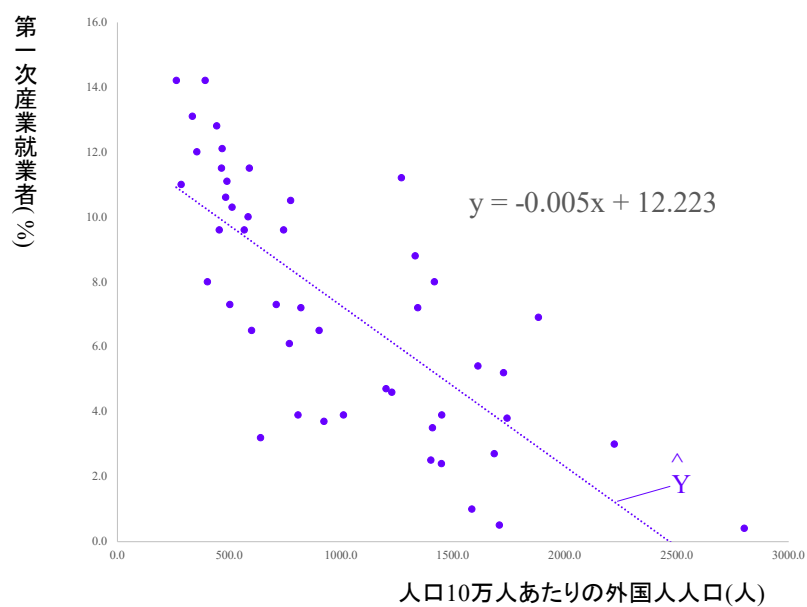


図6. 第一次産業割合と外国人の関連

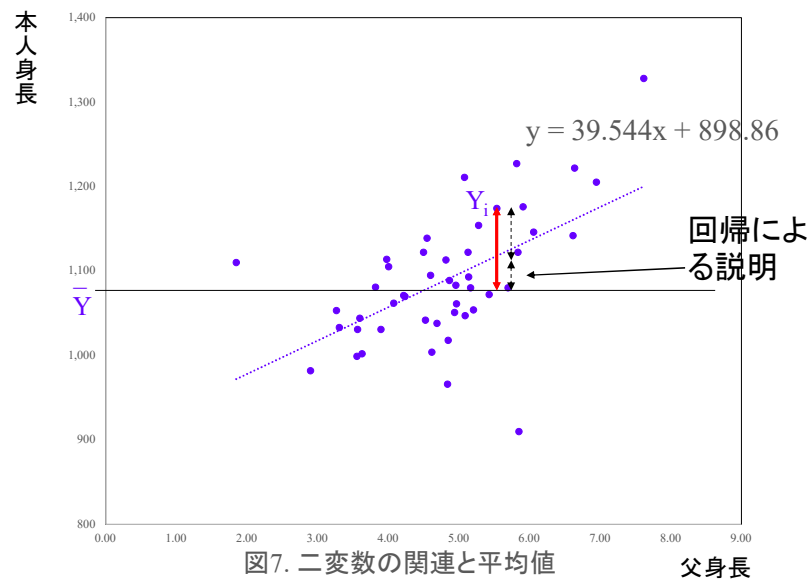


図7. 二変数の関連と平均値  
誤差とは何かを理解すること  
p.125の図で、**残差部分が小さくなればよい説明**

121

## 最小二乗法の意味とは

- ◆ p.125の図の意味を理解する。
- ◆ 偏差は二分割できる。
- ◆ **残差**(回帰で説明できなかった部分)の二乗の合計値を最小にするように計算する。
- ◆ **回帰部分**(モデルにより説明できる分散)が大きければ良い説明。

122

## F値とは

- ◆ 分母が残差部分、分子が回帰部分

F値の最大値は $\infty$

- ◆ ちなみに、0で割ると $\infty$

- ◆ なお、標本のばらつき具合は母集団より小さめになる。真のばらつきは大きい。
- ◆ 標本平均は母平均と一致するが、標本分散と母分散は一致しないのである！

123

## 多重共線性とは

$$X_1 = 2X_2$$

$$Y = 2 \times 2X_2 + 3X_2$$

$$Y = 7X_2$$

124

## 相関と回帰

- ◆ 相関は、点の散らばり具合について見ている。
- ◆ 回帰は、直線を当てはめた時の、線の傾きについて見る。

125

## 重回帰分析 例

```
/****** 逆転した新変数の作成 *****/  
COMPUTE      SYAYOKU  =5-Q14A.
```

```
COMPUTE      YARITAI  =5-Q15A.  
COMPUTE      DANZY0   =5-Q15C.  
compute slstrat = 6-Q16.
```

```
/****** 震災前の持ち家 *****/  
COMPUTE      Honer    =0.  
IF (Q22<3)    Honer    =1.  
compute HEALTHCN  = 6-Q30.
```

```
REG  
/DEP SYAYOKU  
/METHOD=ENTER sexdmy AGE EDU livrate FAMILYNO  
slstrat Honer HEALTHCN danzyo yaritai .
```

126

## ダミー変数を使った重回帰分析

### ◆ダミー変数作成の例

```
COMPUTE      SEIKI  =0.  
IF (Q33=2)    SEIKI  =1.
```

```
COMPUTE      HISEI  =0.  
IF (Q33=3)    HISEI  =1.  
IF (Q33=4)    HISEI  =1.  
IF (Q33=5)    HISEI  =1.  
CRO /TAB=    HISEI  by Q33.
```

Xの中に、自営ダミー変数など入れてみる。

127

### ◆例 職業（従業上の地位）が**正規、非正規、自営、無職**の4カテゴリーの場合

→ 正規以外の3つのダミー変数を入れる。  
つまり、正規が基準カテゴリーとなる。

### ◆Yが2値変数の場合は、ロジスティック回帰分析を行う。

128



## その他

- ◆ 男女別にデータを分割してから分析。
- ◆ **国別も、データ分割。**
- ◆ 階層帰属意識など入れてみる。
- ◆ その他、人間関係の問を入れる。

- ◆ 合計得点の作成

財産項目などの01の変数

Missing values Q17A to Q17E (9).

Compute NEW17=Q17A+Q17B+Q17C+Q17D+Q17E.

129

## 分析結果の解釈

- ◆ そもそも、社会階層とYが関連があるのか
  - 経済的資源 収入や資産
  - 情動的資源 学歴や資格や教養
  - 関係的資源 付き合いがあるか、交際、  
Social Capital
- ◆ 都市度、自営かどうか、なども分析するとよい
- ◆ **何をYにするか**、よく考える
- ◆ 仮説を明確に提示する

130

## 注意点

- ベータやR二乗に有意水準をつける
- 年齢を入れる。年代ではない。
- カテゴリー変数を入れることはできない。
- モデルの作り方 → 資料参照
- 社会的変数とは何か
- 年齢、教育年数、階層帰属意識、基本的な社会意識はXとして入れておく。

131

## 12.重回帰分析の注意点

- 結果の因果関係について、よく考える。自分の解釈を豊富に書くこと。
- 変数の方向を、分かりやすいように逆転するが、合併しない。XもYも、できる限り細かい変数を使うこと。
- エクセルで「セルの書式設定」
- p.167～ 偏回帰係数の解釈

132

- 例 性別無回答をデータから削除。

select if q46sex ne 9.

- p.182 国別に分析
- p.183～ 結果の解釈、モデルの作り方
- 例 ISSP2015データを日本だけに絞る。  
国の変数名は年度により異なる。  
select if c\_sample = 392.

133

## 13. 一元配置分散分析 9/20

シンタックス例

Xはカテゴリー変数を1つだけ入れる

ONEWAY

N9A N9B BY nendai

/PLOT MEANS

/STA DES.

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。○<sup>134</sup>

- ◆ 分散分析とは、カテゴリーごとの平均値の差の検定。
- ◆ **分散分析のY**や、重回帰分析のYとXは、量的な変数なので、できるだけ細かくし、カテゴリー合併しない。
- ◆ **分散分析のX**や、クロス集計で使う変数はあまり細かくしない。

135

## 一般線型モデル(GLM)

- ◆ 重回帰分析も分散分析も、線型モデルの一種
- ◆ 分散分析の例
  1. 専門職、管理職、事務職、その他職で生活満足度の平均値の差を検定(4群)
  2. 実験データ      攻撃行動の平均値の差を検定
    - 暴力映画を見た群(刺激1)
    - 外でたくさん遊んだ群(刺激2)
    - 何もしない群(統制群)

136

## 平均値からの距離

### ◆ 身長平均値の例

- ・全体平均が160
- ・男の平均値170  $j = 1$  グループ1
- ・女の平均値150  $J = 2$

ある男の平均値が175だとする

群平均からの距離は5  $y_{i1} - \bar{y}_1$

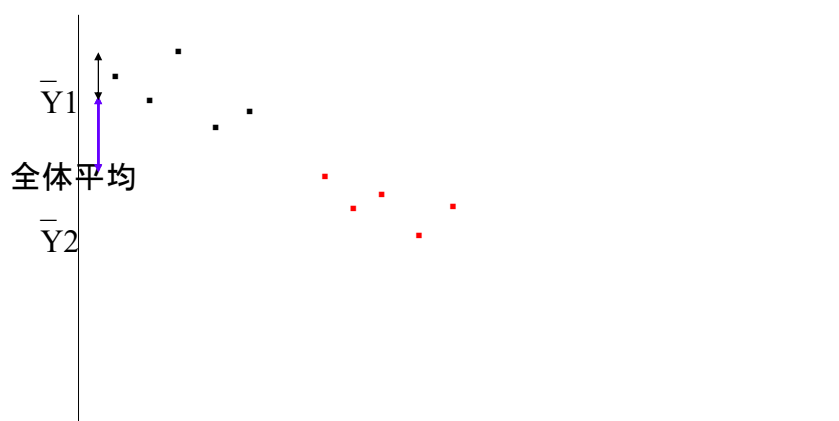
全体平均からの距離は15

人によって身長は異なるし、郡内でもばらつきがある。

137

## 平均からの距離は偏差D

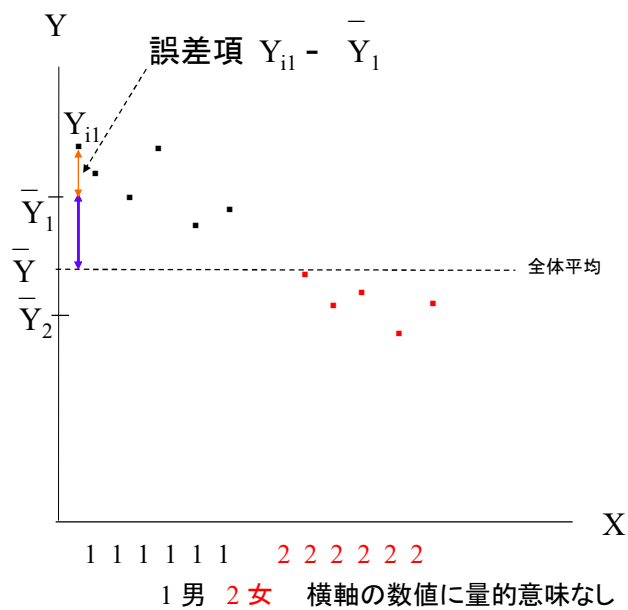
### ◆ Dを2分割する



横軸は意味が無い、縦軸は身長  
男女の身長平均値が異なる

138

## 平均からの距離は偏差D



139

## 平均からの距離は偏差D

### ◆ 郡内のばらつきがない例



140

## 14. 二元配置分散分析 9/20

シンタックス例

```
UNIANOVA    EDU  BY  nendai q52sex  
  /DES =  nendai q52sex  nendai*q52sex  
  /POS =  nendai ( SCHEFFE )  
  /PLO = PROFILE( nendai*q52sex )  
  /PRI = DES    .
```

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。

141

- ◆ デザイン行に、直接効果と交互作用効果を含む、自分のモデルを書く。
- ◆ 交互作用項はなくても良い。現実的には、Xの数は、あまり多くはできない。
- ◆ 分散分析とは、2つ以上の平均値の差の検定である。
- ◆ 分散分析は、100人以下のデータで使うこともあるため、有意かどうかの検定はするが、効果の大きさについて検討しないことが多い。

142

◆ 分散分析の被説明変数Yは、量的意味があるものを使う。

◆ 例 学歴変数を教育年数に直す  
中卒9、高卒12、短大卒14、大卒16

◆ 例 職業 管理職1、正社員2、非正規雇用3、  
自営業4、無職5  
数字に量的意味がない

143

◆ Xはカテゴリー合併する

- age → nendai
- JGSSシンタックス例  
XXJOB → JOB8

◆ F値とは モデルによる説明と誤差分散の比  
大きいほど有意

◆ 課題 3元配置以上で行う  
職業分類は、各種見本シンタックスを参考に、自分で作って見ること。

144



## SRモデル

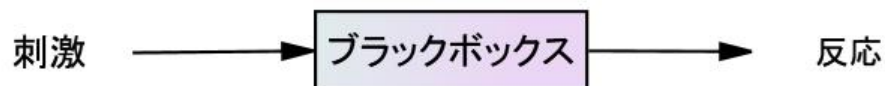


図 刺激 - 反応モデル

145

## 職業分類 7/18

自分で大分類を作り分散分析を行う

社会調査の職業とは → 地位と役割

産業と職業は違う

### ◆ 資料

- 2015SSPデータとシンタックス
- 01coding 501-689の小分類
- 原純輔盛山和夫『社会階層』p.143, 付録表1

146

## 職業の4次元

- 産業は組織の分野
- 従業先の規模
- Occupational Status  
パート、アルバイト、派遣、嘱託などはすべて  
非正規雇用
- 個人がやっている仕事内容が狭義の職業  
⇒ SSM職業小分類 501-689

Occupation, Job の違いとは

147

- ◆ まず本人の仕事内容の変数をよく見てみる  
こと(SSM調査職業小分類)  
SSP2015 変数SSM(1995SSM職業小分類コード)
- ◆ 国勢調査分類、日本標準職業分類、国際職業分類ISCOなどがある。
- ◆ F値とは、平均値からの距離を2分割したものの比である。どのような意味か、資料とテキストを読んで復習し理解すること。

148

◆ 分散分析表の作り方を理解する

- 関連の大きさ
- 背後にある要因を考えてみる
- 有意水準    5%   \*    1%   \*\*

◆ 課題 職業分類を用いた分散分析

年代や学歴もXとしていれる

◆ 産業分類      →    各自でみておくこと

ブラックボード教材coding

149

## 15. 社会的資源とは

人々の欲求の対象だが、充分にはないもの

- ・ものやお金
- ・学歴、情報、資格、教養
- ・関係的資源

資源分配の不平等な構造

→ 社会階層、地位の構造

地位の指標 → 実際には職業分類使う

150

## 16.職業分類とは

- ◆ 職業と産業の違い

- ◆ 職業の4次元とは

- ◆ 狭義の職業の分類

日本標準分類、国勢調査分類、その他  
威信スコアは、小分類をもとに作成

151

## 社会調査でよく使われる 職業大分類

- ◆ SSM旧8分類

専門、管理、事務、販売、熟練、半熟練、非熟練、農業

- ◆ SSM新8分類

1 '専門' 2 '大W' 3 '中小W' 4 '自営W' 5 '大B' 6 '中小B' 7 '自営B' 8 '農業' 96 '無職' 98 '学生'

- ◆ 安田・原の総合職業分類(社会調査ハンドブック参照)

152

## 職業分類をXとした分散分析

- ◆ まず、自分で5分類くらいを作る
- ◆ Yは連続量として意味があるものにする

153

## 政党支持のカテゴリー合併例

```
value labels Q24 01 '自民党' 02 '民主党' 03 '公明党' 04 '共産党'  
05 '維新の党' 06 '次世代の党' 07 '社民党' 08 '生活の党'  
09 '新党改革' 10 'その他' 11 '支持政党なし' 99 'DK/NA'.
```

```
COMPUTE N24 = Q24.  
RECODE N24 ( 6=5).  
RECODE N24 ( 7=5).  
RECODE N24 ( 8=5).  
RECODE N24 ( 9=5).  
RECODE N24 (10=5).  
RECODE N24 (11=6).
```

```
value labels N24 01 '自民党' 02 '民主党' 03 '公明党' 04 '共産党'  
05 'その他' 06 '支持政党なし' 99 'DK/NA'.
```

**新変数と元の変数**のクロス集計表を出す。

154

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。

#### ANOVAシンタックス例

```
UNI    SEIMAN BY  nendai q1sex  
      /DES = nendai q1sex  nendai*q1sex  
      /POS = nendai ( SCHEFFE )  
      /PLO = PROFILE( nendai*q1sex )  
      /PRI = DESCRIPTIVE  
      /EMMEANS = tables (nendai) .
```

BYの前に被説明変数(量的変数Y)を書く。

155

## 17.因子分析 10/18

- ◆探索的モデル
- ◆確証的モデル

156

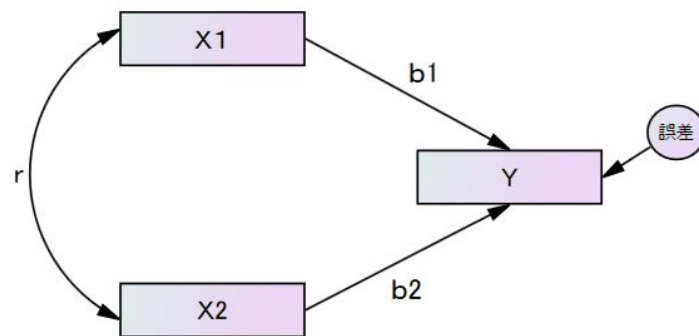


図1. 基本的な重回帰分析モデル

157

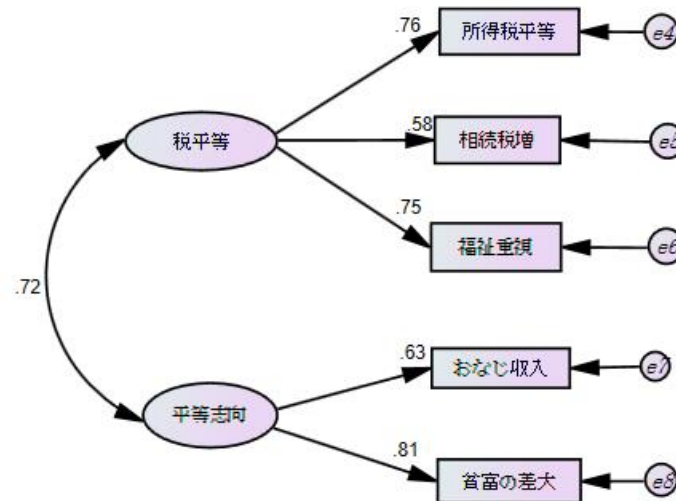


図2. 平等意識に関する因子分析 2因子モデル 標準化係数 1997年仙台調査男性  
変数下の数字はR-square 誤差項は略 Chi-square=11.06 df=4 p=.03  
AIC=33.06 GFI=.99 AGFI=.95 CFI=.98 RMR=.03 RMSEA=.07

158

## SPSSでの分析

データを開く

- 自分が使いたい変数の欠損値処理を確認。
- 必要なら変数逆転する。
- 因子分析のシンタックスを書いて実行。
- 分析結果を自分で解釈する。
- 解釈しにくい場合は、因子数を変える、などする。

159

## シンタックス例

FAC

/VAR KAKUSA SHISANKA OTOKOSO OTTOKAJI

/FORMAT SORT

/ROTATION VARIMAX

/PLOT ROTATION .

回転後の共通性は出力にない。

エクセルの二乗和          ()内のセル範囲はコロン

=SUMSQ(B5:D5)

160



## モデルの作り方

- ◆ 4変数を4因子に要約しても意味がない。
- ◆ どの変数を使うか、分析結果を見てからやり直すなど、試行錯誤する。
- ◆ 抽出法に決定版はない。因子分析結果は砂上の楼閣。
- ◆ 量的な変数しか使えない。

161

## 18. AMOSによる因子分析 11/8

- ◆ まず、SPSS側で、シンタックスにより欠損値を削除したデータを作る。  
select文を使う → 名前をつけてデータを保存(sav)
- ◆ AMOSでサンプルモデルを開き、「ファイル」をクリックして、自分が使うデータファイル(sav)を指定する。
- ◆ AMOSでモデルをかいてから、分析実行。

因子分析で使うことができるのは量的変数なので、できるだけ細かい変数を使う。

162

## シンタックス例

### SELECT文を使うとデータ人数が減る

/\*\*\*\*\* 欠損値を除く処理 \*\*\*\*\*/

SELECT IF Q6A <9.

SELECT IF Q6B <9.

SELECT IF Q6C <9.

SELECT IF Q29A <9.

SELECT IF Q29B <9.

/\*\*\*\*\* 方向を逆転した新変数作成 \*\*\*\*\*/

COMPUTE TYONAI =5-Q6A.

COMPUTE YAKUNIN=5-Q6B.

COMPUTE GIIN =5-Q6C.

COMPUTE axlife =5-Q29A.

COMPUTE axincom=5-Q29B.

**FRE** VAR= TYONAI YAKUNIN GIIN axlife axincom .

削りすぎて人数がゼロになっていることがある。度数分布をみて確認する。 <sup>163</sup>

## 自分でモデルを書く場合

◆ パスのどこに1をつけるか注意する。

◆ 誤差項からの矢印は必ず1がつく。

◆ 矢印の方向を間違えないこと。

因子から観測変数へ、という方向になる。

因子の名前は自分で解釈したものを、ラベルとして書けばよい。

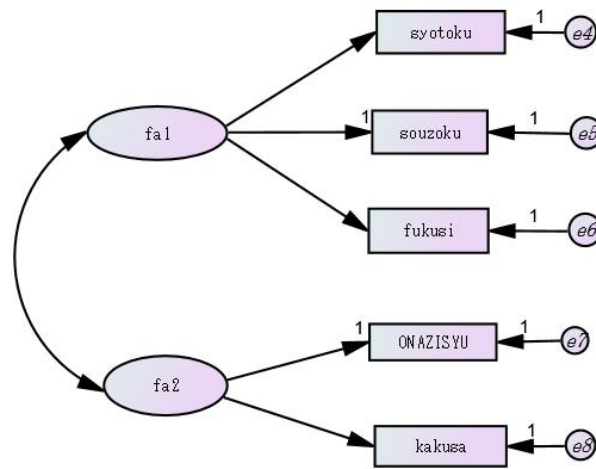


図11.平等意識に関する因子分析 2因子モデル 標準化係数  
2015年福島市調査男性 Chi-square=χ<sup>2</sup>min df=νdf p=χ<sup>2</sup>p  
AIC=χ<sup>2</sup>aic GFI=χ<sup>2</sup>gfi AGFI=χ<sup>2</sup>agfi CFI=χ<sup>2</sup>cfi RMR=χ<sup>2</sup>rmr RMSEA=χ<sup>2</sup>rmsea

165

## 19.AMOSによる重回帰分析

- ◆ 通常型の重回帰分析と変わらない。Xが複数ある。どこかを因子にすると、変わってくる。
- ◆ 左端にX、右端にYがくるようなモデルを自分で描いて、分析実行する。
- ◆ 因果関係についてよく考えることが重要。
- ◆ Yについての変数が複数あるならば、因子にしてみる。

166

誤差項がつくのは矢印が刺さっている変数

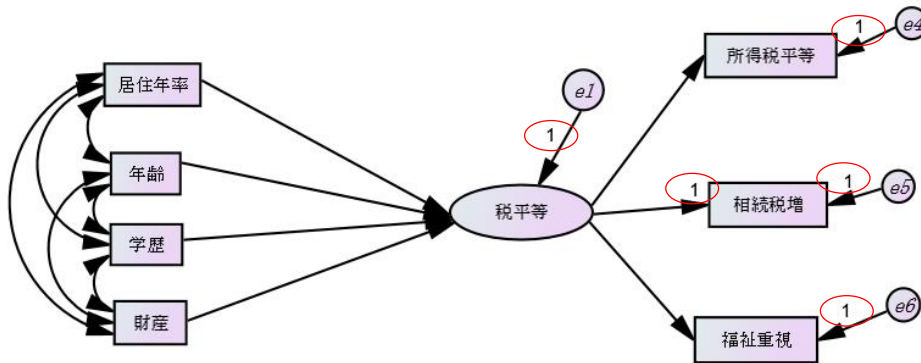


図2. 被説明変数を因子にした例 標準化係数 1997年仙台調査男性  
変数下の数字はR-square Chi-square=χ<sup>2</sup>min df=νdf p=χ<sup>2</sup>p  
AIC=χ<sup>2</sup>aic GFI=χ<sup>2</sup>gfi AGFI=χ<sup>2</sup>agfi CFI=χ<sup>2</sup>cfi RMR=χ<sup>2</sup>rmr RMSEA=χ<sup>2</sup>rmsea

167

## モデルの適合度が悪い場合

- ◆ 有意でないパスが残っている。
- ◆ 有意なパスを削除した。
- ◆ 無理な因子を作った。
- ◆ 左端に統制変数としておいた方がよい変数(基本属性など)がない。
- ◆ 有意でないパスは削除して、もう一度分析する。
- ◆ 男女別に分析した方が当てはまりがよくなることが多い。

168

## 因子を使ったパス解析の例 左端に統制変数をおくとよい

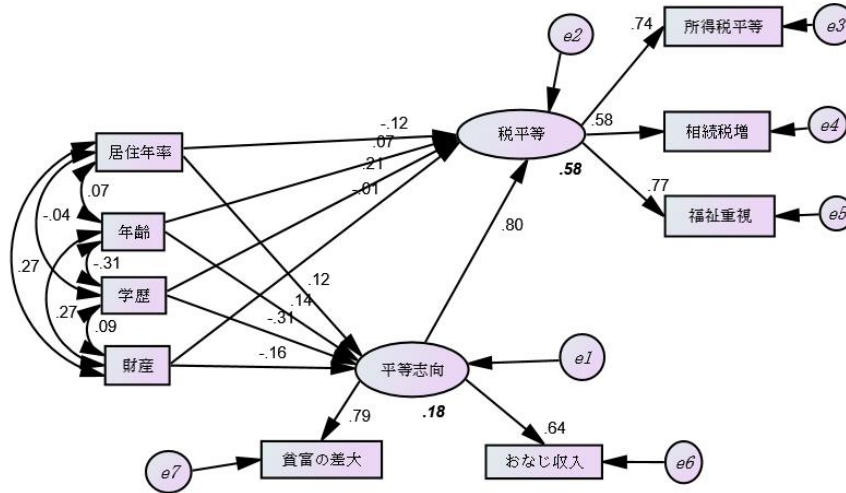


図3. 因子を用いたパス解析の例 標準化係数 1987年仙台調査男性  
変数下の数字はR-square Chi-square=28.41 df=18 p=.05  
AIC=84.41 CFI=.98 AGFI=.95 GFI=.98 RMSEA=.04

169

## 統制変数の例

\*居住年数率 livrate作成.

COMPUTE living= q1.

If (Q1 > age) Q1 = age.

COMPUTE livrate= living/age.

If (livrate > 1) livrate = 1.

\*家族人数の変数.

COMPUTE FAMILYNO=Q27.

IF (Q27=0) FAMILYNO=1.

MISSING VALUES FAMILYNO (9).

170

## 20.AMOSによるパス解析 11/17

- ◆ XからYへの直接効果だけでなく、他のXを通じた間接効果があるかどうかについて分析。
- ◆ 左端に基本属性、右端にYがくるようなモデルを自分で描いて、分析実行する。
- ◆ 因果関係についてよく考えることが重要。
- ◆ サンプルモデルを保存してから、自分で書き替えればよい。

171

## パス解析の例

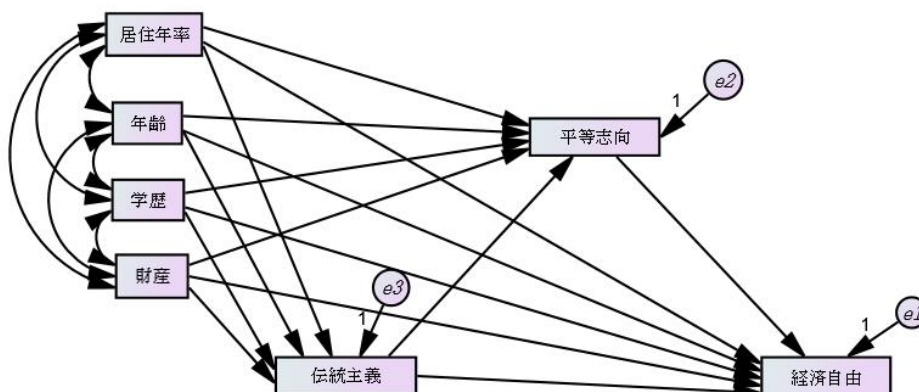


図5.経済自由志向のパス解析 標準化係数 2012年○×調査男性.  
変数下の数字はR-square. Chi-square=χ<sup>2</sup>min df=νdf p=χ<sup>2</sup>p  
AIC=χ<sup>2</sup>aic GFI=χ<sup>2</sup>gfi AGFI=χ<sup>2</sup>agfi CFI=χ<sup>2</sup>cfi RMR=χ<sup>2</sup>rmr RMSEA=χ<sup>2</sup>rmsea

172

## Path Analysis モデルの作り方

- ◆ 左端に年齢や教育年数(学歴)など基本属性に関する変数、右端にY3として最終的な被説明変数(目的変数)を置く。
- ◆ Y3が生活満足感や、政治的無力感など、表面的な意識であった場合、Y2を他の意識や、何らかの無意識や深層心理に近い価値観、
- ◆ Y1を人間関係や社会階層などの、社会学的な変数を使って、モデルを作ってみると良い。
- ◆ 途中のY1のみを因子にする、などのモデルもありうる。

173

## 注意点

- ◆ 左端に基本属性を置く。  
パスとは、Yへの通り道のこと。
- ◆ 因果関係を考えて、中間に何かを置く。
- ◆ 量的変数のみを使う。できるだけ細かい変数を使うこと。
- ◆ ダミー変数を置いてもよい。自営業かどうかの0,1など。

174

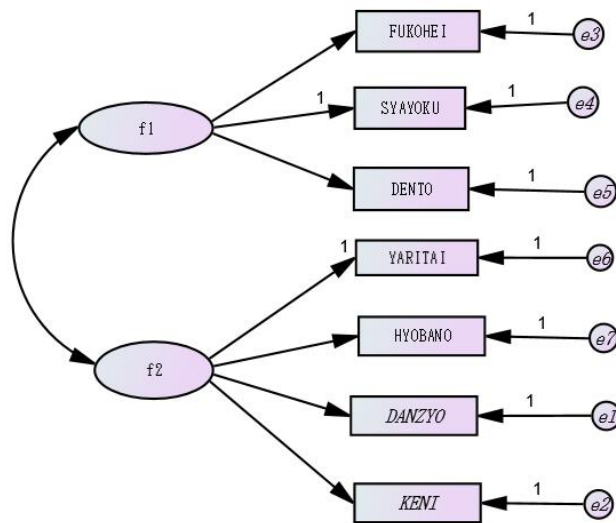


図11.平等意識に関する因子分析 2因子モデル 標準化係数  
1997年仙台調査 Chi-square=¥mmin df=¥df p=¥p  
AIC=¥aic GFI=¥gfi AGFI=¥agfi CFI=¥cfi RMR=¥rmr RMSEA=¥rmsea

175

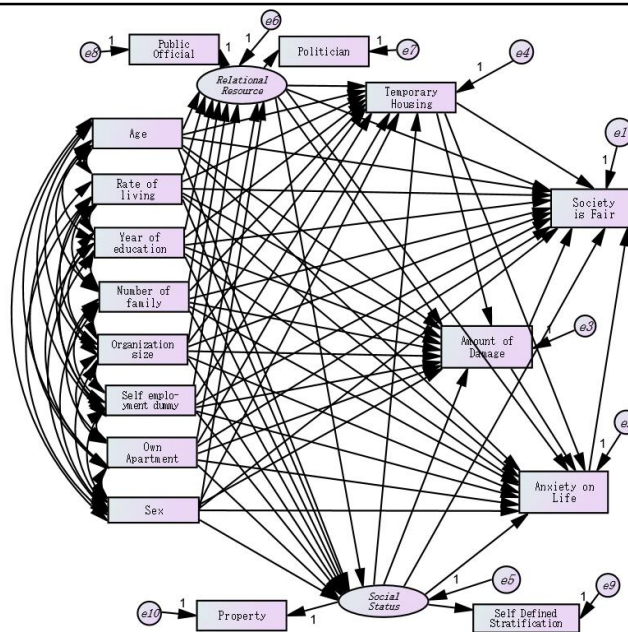


Fig 7. SEW on Fairness and Social Status. 2015 Fukushima Survey.  
Standardized coefficients. Bold numbers below variables are  
R-square, Chi-square=¥mmin df=¥df p=¥p AIC=¥aic  
GFI=¥gfi AGFI=¥agfi CFI=¥cfi RMR=¥rmr RMSEA=¥rmsea

176



## モデルの作り方 11/8

- ◆ 右端に最終的な被説明変数
- ◆ 左端に多くの基本属性
  
- ◆ その中間に
  - ・ 社会階層
  - ・ 人間関係保有
  - ・ 無意識に近い態度やパーソナリティ変数
  - ・ 生活満足感など基本的な社会意識

177

## データについて

- ◆ 抽出法や回収法など注意する。
  
- ◆ インターネット調査は無作為抽出になっておらず偏りが大きい。社会学的な分析をする研究に使わない方が良い。
- ◆ 実際のところ、社会心理学的な研究は無作為抽出をしていないものが多い。
  
- ◆ 態度や基本的な意識として何を入れるか。
- ◆ XとYが似すぎていると良くない。

178

## AMOSの操作

- ◆ まず、最終的な被説明変数Yを1つ決める。
- ◆ SPSS上で変数逆転などしてから、  
欠損値を削除したデータを作る。  
資料のselect文を理解する。
- ◆ 人数が減ったデータ(savファイル)が出来たことを確認し、データを「名前をつけて保存」。
- ◆ AMOSでモデルを描き、ファイル→データファイルの名前(savファイル)を指定。その後分析実行。
- ◆ 男女別に分析するときはグループ化変数、グループ値ボタンを使う。

179

### シンタックス例

```
/****** 欠損値を除く処理 ★実行すると人数が減るので注意 *****/  
SELECT IF sexdmy ne 9.  
SELECT IF ageb < 99.  
SELECT IF edu ne 99.  
SELECT IF GEN ne 99.  
SELECT IF st5job <9.  
SELECT IF fq7ffdnr <9.  
SELECT IF st5areay <9.  
SELECT IF op5happz <9.  
  
COMPUTE SIGOMAN =6-st5job.  
COMPUTE FADINNER =8-fq7ffdnr.  
COMPUTE TIKUMAN =6-st5areay.  
COMPUTE KOFUKU =6-op5happz.  
  
/****** 分析命令文 *****/  
FRE var = AGEB EDU sigoman fadinner tikuman kofuku .
```

180

## モデルの作り方の注意点

- ◆ 初めは、単なる重回帰か、因子を使わないパス解析など、単純なモデルの方が簡単。
- ◆ サンプルモデルを、名前をつけて保存し、書き換える。
- ◆ 因子が成立するか注意。似ていない複数の観測変数をもとに一つの因子を作ろうとしても無理。
- ◆ 誤差項の付け忘れに注意。

181

## 21.AMOSでの分析結果

- ◆ Yについて因果関係を考える
  - 規定因Xとして何がありうるか
  - 連続変数を入れること
- ◆ 段階的なモデル
  - 左端は基本属性、右端はY
- ◆ 分析プロパティ → 平均値と切片を推定
- ◆ モデルのカイ二乗値とは何か
- ◆ モデル全体のp値は大きい方がよい  
モデルと現実の距離についての検定

182

- ◆ AICもモデルと現実の距離
- ◆ 多重共線性に注意する
- ◆ ダミー変数を左端に入れることは問題ない
- ◆ 左から右へ、原因から結果の流れが分かりやすいように、モデルをかく

183

## AMOS実行前に

- まず、SPSSでシンタックス実行  
select文、変数逆転、度数分布
- データ画面で、人数が減ったデータを、名前をつけて保存 →savファイルを新たに作る
- AMOSで、savデータファイル選択
- データ選択時に、性別を絞ることができる →配布資料「グループ化変数」

184

- 因子からのパスは、+とーがあるが、とくに問題はない。
- テキスト出力をみて、パスが有意かどうかを確認 → 有意でないパスを削除して再度分析。
- タイトルは自分で編集する。  
サンプルを参考に書くと適合度係数が自動で出る。データ人数Nは自分の手で書く。

Yについての男女別横棒グラフ → エクセルファイル見本をよく見る。

185

## 卒論について 11/7

- 基本的な分析の後に、重回帰、分散分析などする。その前に、基本統計量、変数間の相関の表や、因子分析を載せる。その後にはパス解析やエラボレイション。
- まず、シンタックスを書くこと。自分に必要な新変数を作る。
- 正規と非正規や、ダミー変数など、作ってみる。
- 基本的な分布を確認する。
- 年齢別、都市規模別など確認する。
- 階層帰属意識とのクロス集計などやってみる。

186

## 実際の調査は

- ◆ 質問量が多すぎると現場の調査員はとても大変
- ◆ 現場での工夫
  - お願い状、訪問票の作り方
  - 大学だけでやっている調査だということを説明する
  - 手書きでお願い文を書き込む。

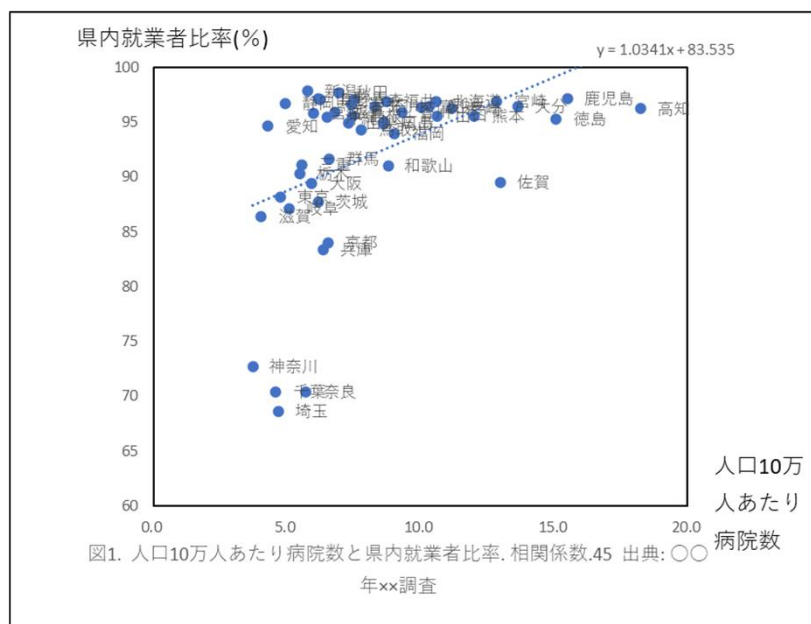
187

188

## 基礎演習 グラフについて

- 縦軸と横軸の単位が分かるように
- データの年度や出典は、タイトルに詳しく書く。
- 口頭や文章で補足しなくても、グラフだけを見て分かることが大原則！
- 散布図は正方形に。
- 相関係数は別途出す → タイトルに書く
- グラフ内の文字が小さいと見えないので注意！  
グラフ全体を選択した状態で右クリック → フォント

189



190

## スライド作成の注意点

スライドは大きく4構成に。

冒頭部で目的を明確に書く（ゼミホームページの卒論構成を参照）。 → 見本ファイルを見る。

箇条書きにする！長い文章を書かない。

文字は20ポイント以上に

- ・ 図をスライドに大きくはる、図の中の文字も大きくする。  
相関係数のケタ数は、あまり細かくしない  
× 0.48576 ○ 0.49
- ・ マクロデータは、疑似相関が多く出る（生態学的相関）。  
背後にある因果関係を考えることが大切。

191

## 相関関係と因果関係

Yの原因として何がありうるか。



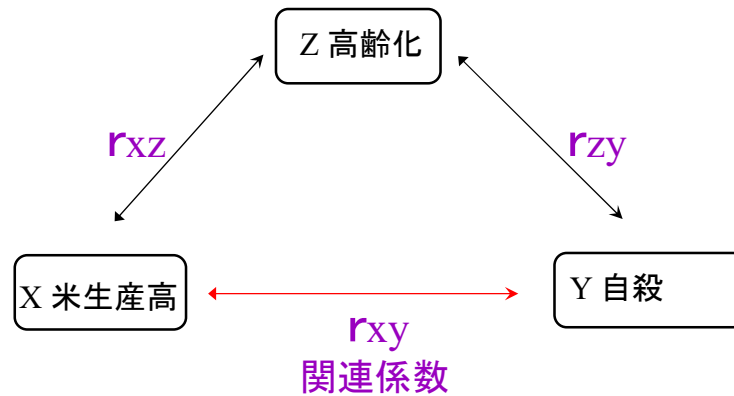
最初に見えた関連 $r_{xy}$ は、真の因果関係か

192



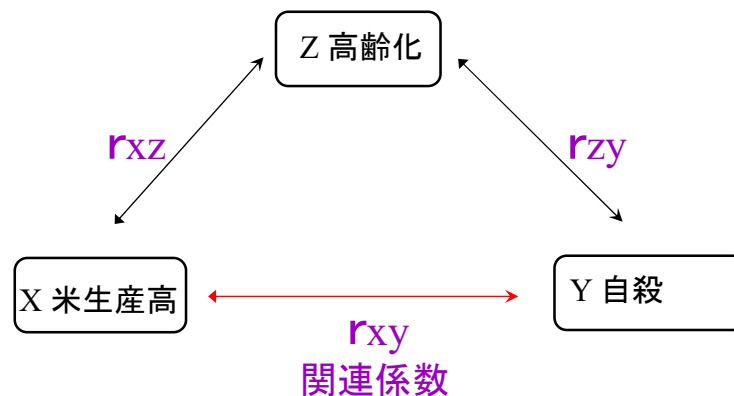
## 因果関係とについてよく考える

Yの真の原因とは



第三変数Zこそが、真の関連かもしれない。ただしZの候補は無限にある。 $r_{xy}$ はなぜ関連があったのか。 193

## 統計的統制とは



もし $r_{xz}$ や $r_{zy}$ 有意な場合、最初に見た関連 $r_{xy}$ も、ある程度の大きさがあるように見える。

- ◆ 相関係数は、あくまでも2変数の関連を見ただけ。
- ◆ 第三変数の効果を取り除く、統計的統制は、次年度以降に学ぶ。
- ◆ なお相関係数は、0.35など、2桁で良い。
- ◆ 目的の前に、問題の所在を書く

195

## 発表時の注意点

- 大きな声で、上を向いて話す  
下を向いて読み上げ型にならないように！
- 話しにメリハリをつける。強弱が大切。
- 身振り手振りを入れる。
- メモやスマホは机に置く。指し棒(ポインター)を持って話す
- パワーポイントは、F5キーで画面が大きくなる、↓キーで進む。
- MacOSで画面が出ない場合  
アップルメニュー → システム環境設定  
ディスプレイ → 「配置」タブ ミラーリングにチェック

196