

# 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—

これは簡略版 自分でメモをとること。

村瀬 洋一

（立教大学社会学部）

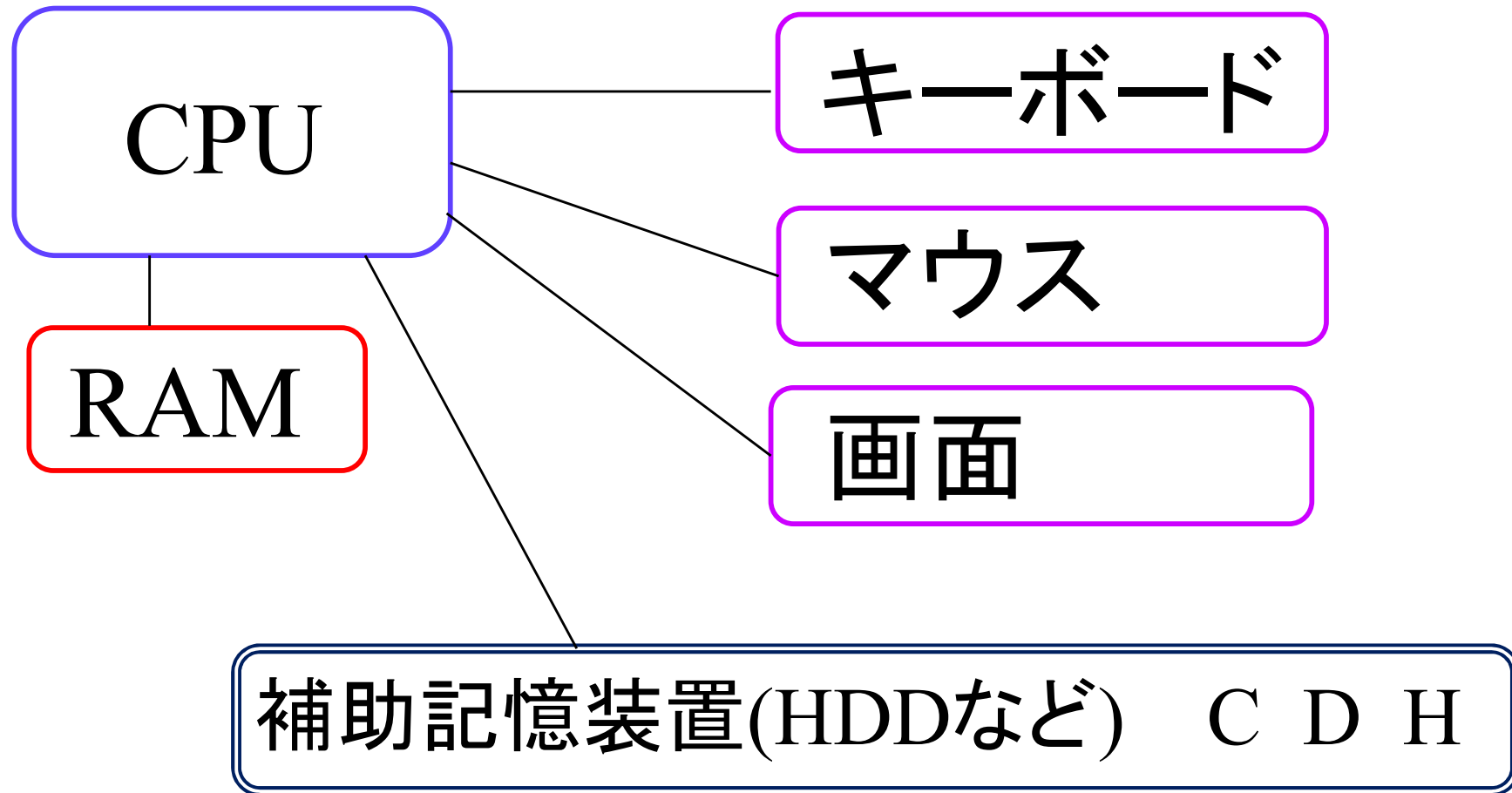
## ◆ 主な内容

- 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すること

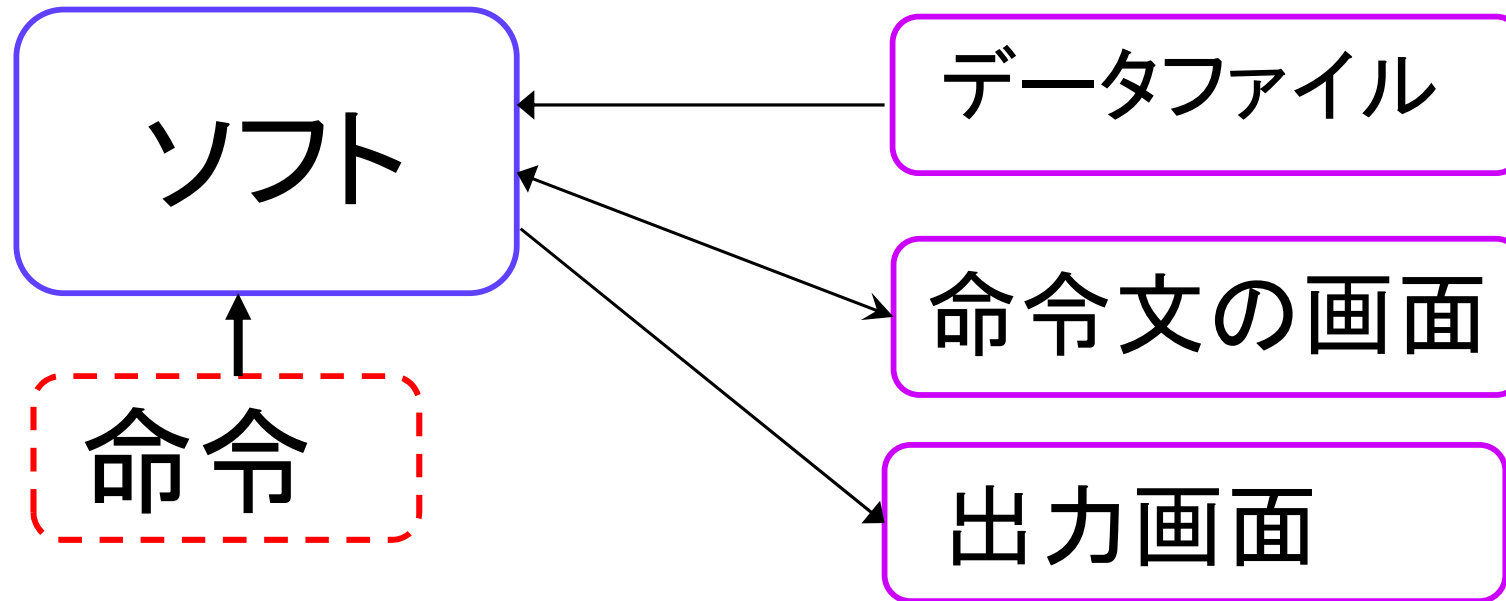
# 1. はじめに

## 1.1.PCの仕組み



記憶装置の中をフォルダで区切る

## 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

# 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

2015年調査、あるいは他のデータでもよい  
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

# 見本のシンタックス

15fmihon1.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。

# 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(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) 拡張機能(X) ウィンドウ(W) ヘルプ(H)

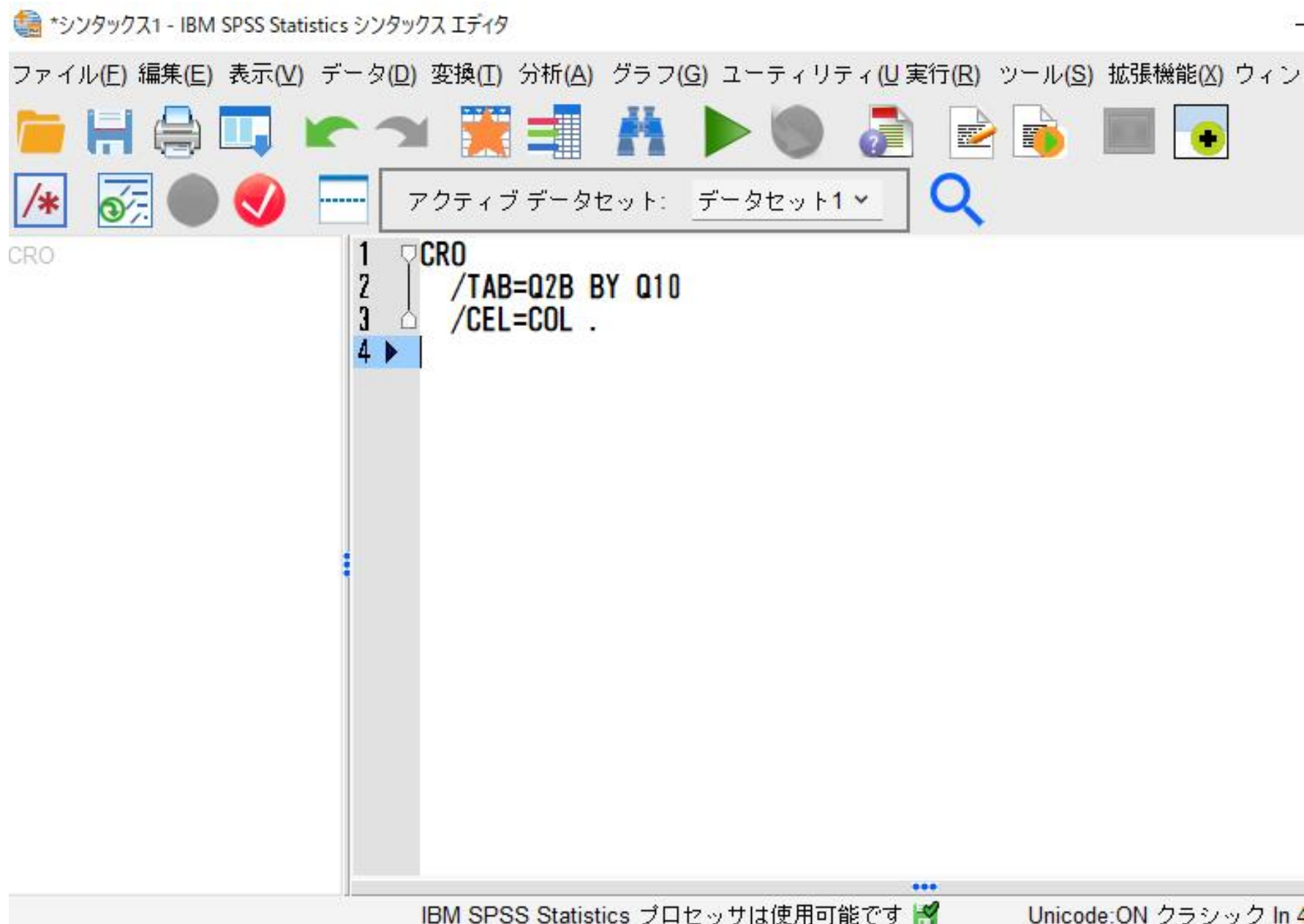
表示: 147 個 (147 変数中)

	ID	area	Q1	Q2 A	Q2 B	Q2 C	Q2 D	Q2 E	Q3	Q4 A	Q4 B	Q4 C	Q5E L	Q5W A	Q5G A
1	1611	16	40	2	2	4	4	2	1	4	2	4	0	10	0
2	1625	16	46	2	3	4	3	1	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	4	1	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

# シンタックス ウィンドウ



シンタックスを書き換えて、名前をつけて保存すること。

# 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。
- ◆ 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。



# 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。

# 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     — データ行列の分析  
これを分析して各種の表を作る

## ◆ 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。

## ◆ 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。

## 2.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か、Hの中へ。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

# データの場所の書き方

- ◆ Cドライブのda7フォルダ内にデータがある場合

シンタックスにて、

C:¥da7¥ のあとにデータファイル名を書く。

- ◆ Windowsのドキュメントフォルダ内だと、

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

- ◆ マックの場合は/Users/ 以降などフォルダ名を情報で出す

# 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。

# 3.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

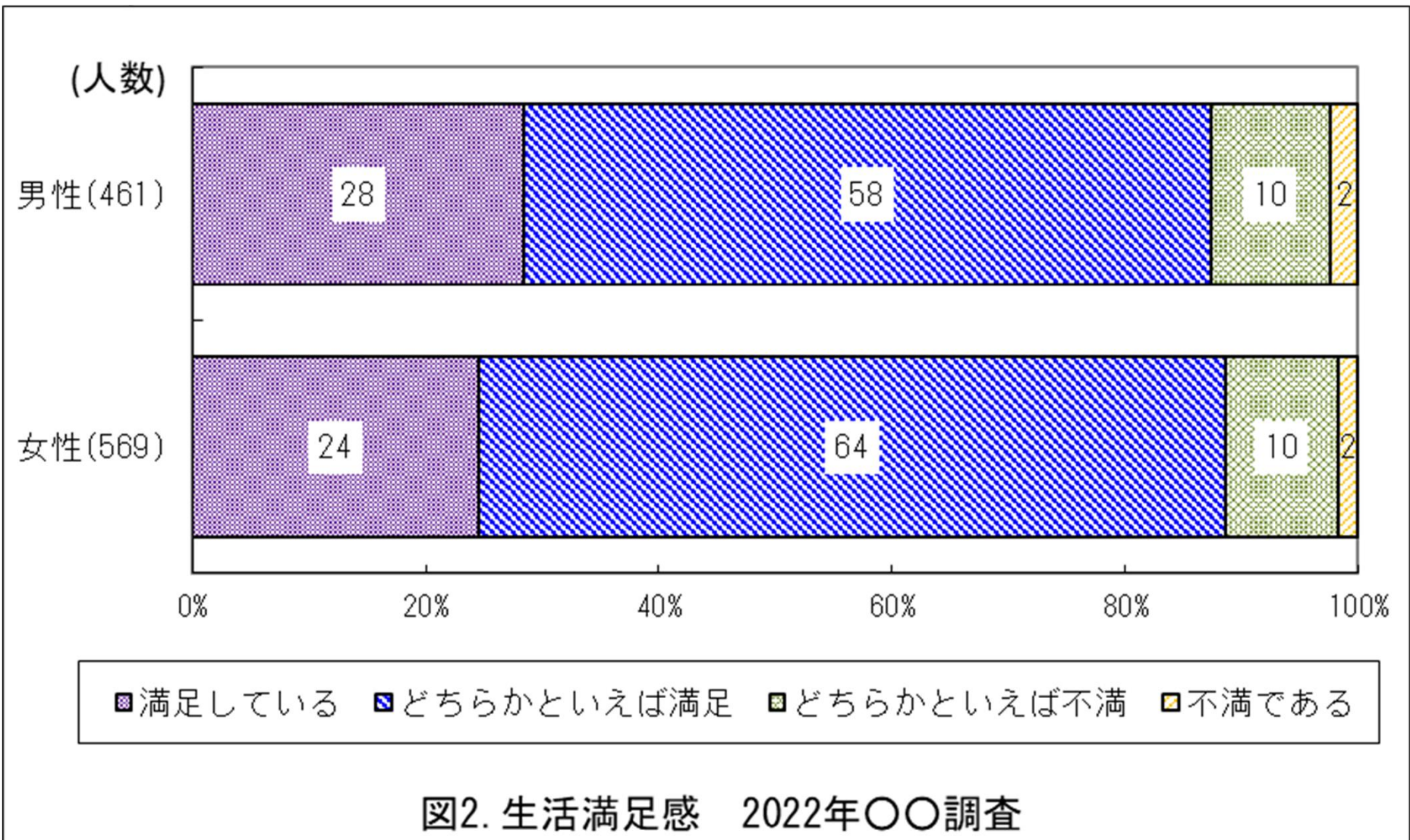
これらはカテゴリー変数

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

# クロス集計結果をグラフ化した例

- 男女別なら横棒が2つある。
- データラベルは横に合計して100になる%を書く





# 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

## 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

- 現実には分布に偏りがある

# 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

# 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

# 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

# 回答

## 無関連状態

	北部	南部	計
白人	56	84	140
黒人	24	36	60
計	80	120	200

→ 2:3 になっている

## 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

# 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。

# 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は、好きな名前が良い。

◆ まず、欠損値処理を忘れずに



# シンタックスで新変数を作る

元の変数



新変数

Q04A 4段階

new04A 2段階

- 元の変数の名前を正確に書く。
- 新変数の名前は何でもよい。
- RECODE文 ()内はカンマ,
- RECORDなどだと動かない

# RECODE文の例

- ◆ 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      NEN6  =  Q46age.
```

```
RECODE      NEN6 (10 THRU 29 =2) (30 THRU 39 =3)  
            (40 THRU 49 =4) (50 THRU 59 =5) (60 THRU 69 =6)  
            (70 THRU 98 =7).
```

```
MISSING VALUES NEN6 (99).
```

```
variable labels NEN6 '年齢カテゴリ' .
```

```
value labels   NEN6 2 '29以下' 3 '30dai'  
               4 '40dai' 5 '50dai' 6 '60dai' 7 '70以上' .
```

MISSING VALUES N10 (9) .

value labels N10 1 'happy' 2 'unhappy' .

## 5.1.年代の新変数作成

COMPUTE NENDA2=Q46AGE.

RECODE NENDA2 (10 THRU 45 =1) (46 THRU 98 =2) .

CRO

/TAB= Q7A 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

## 5.2.学歴カテゴリーの新変数作成

```
/******      GAKUREKI KATEGORI SAKUSEI      *****/  
COMPUTE      EDUC3 =Q43.  
RECODE      EDUC3 (1, 2=1) (3, 4=2) (5, 6=3) (7, 9=9) .  
  
FRE /var = EDUC3.  
  
VALUE LABELS  EDUC3  1  '中卒以下'  2  '高卒'  
                3  '大卒以上'  9  'わからない' .
```

## 5.3.クロス集計のシンタックス例

CRO

/TAB= Q2A by ED4

/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。

## 6. If文による変数加工 6/9

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.

COMPUTE HISEIKI =0.

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

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

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

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



# 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

## 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  $N4A = 5 - Q4A$ .

1,2,3,4  $\rightarrow$  4,3,2,1となる。

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。

## 8. 各種の関連係数

### ◆ 量的変数の相関とは

$r$ 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』

## 2×2表の関連

0 ~ 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

- クラマーの $V$       例えばりんご1    みかん2    ばなな3  
値に量的意味がない場合
- タウ $b$ (3×3表など)、タウ $c$ (3×4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

## 9. クロス集計とエラボレイション p.83

三重クロス集計による因果関係の検討のこと。

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



図3. 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。

二重クロス集計を3つやる。  
第三変数Zは年齢とする。

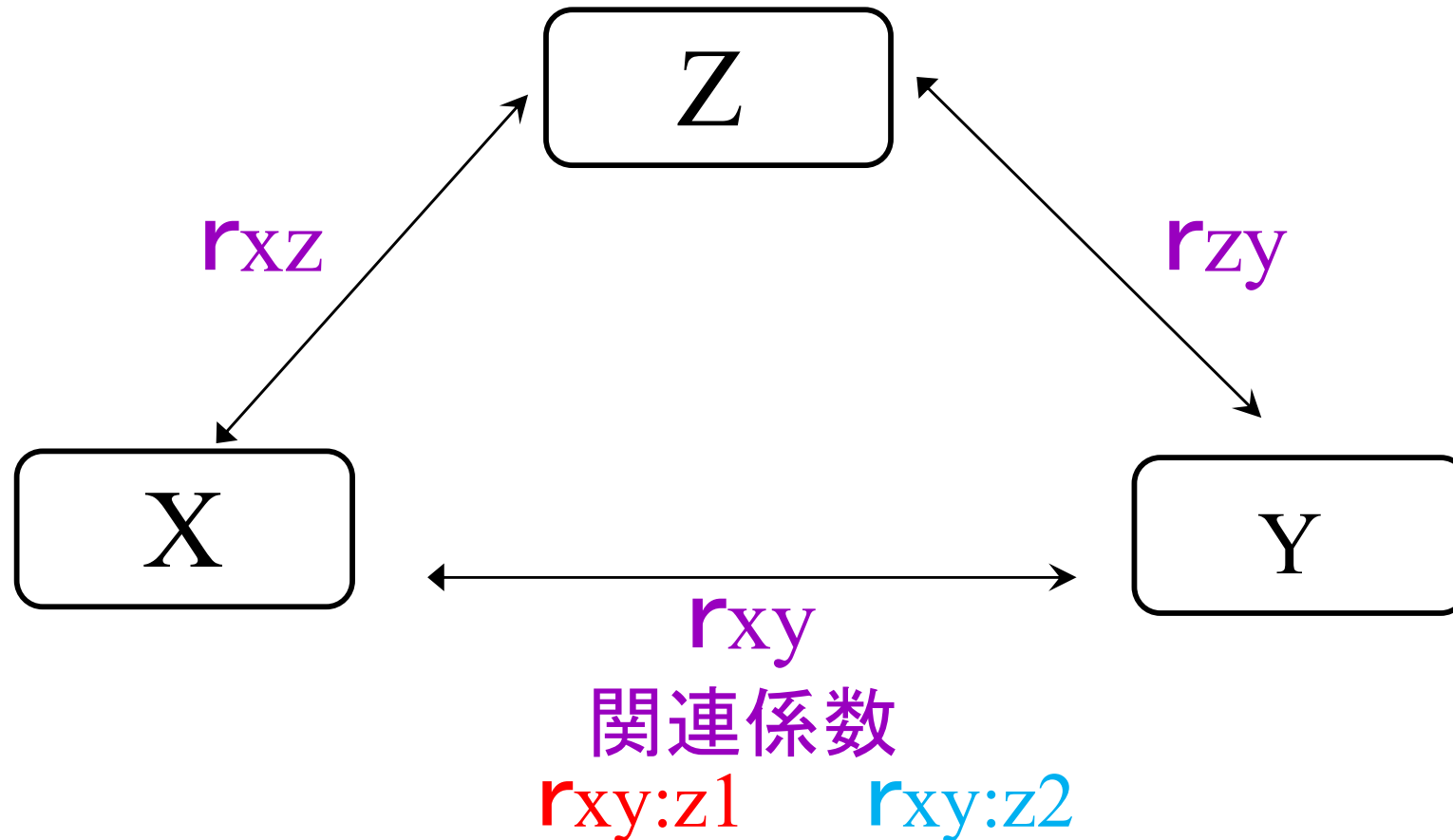


図4. エラボレイションにおける関連

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。

## 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。

# 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

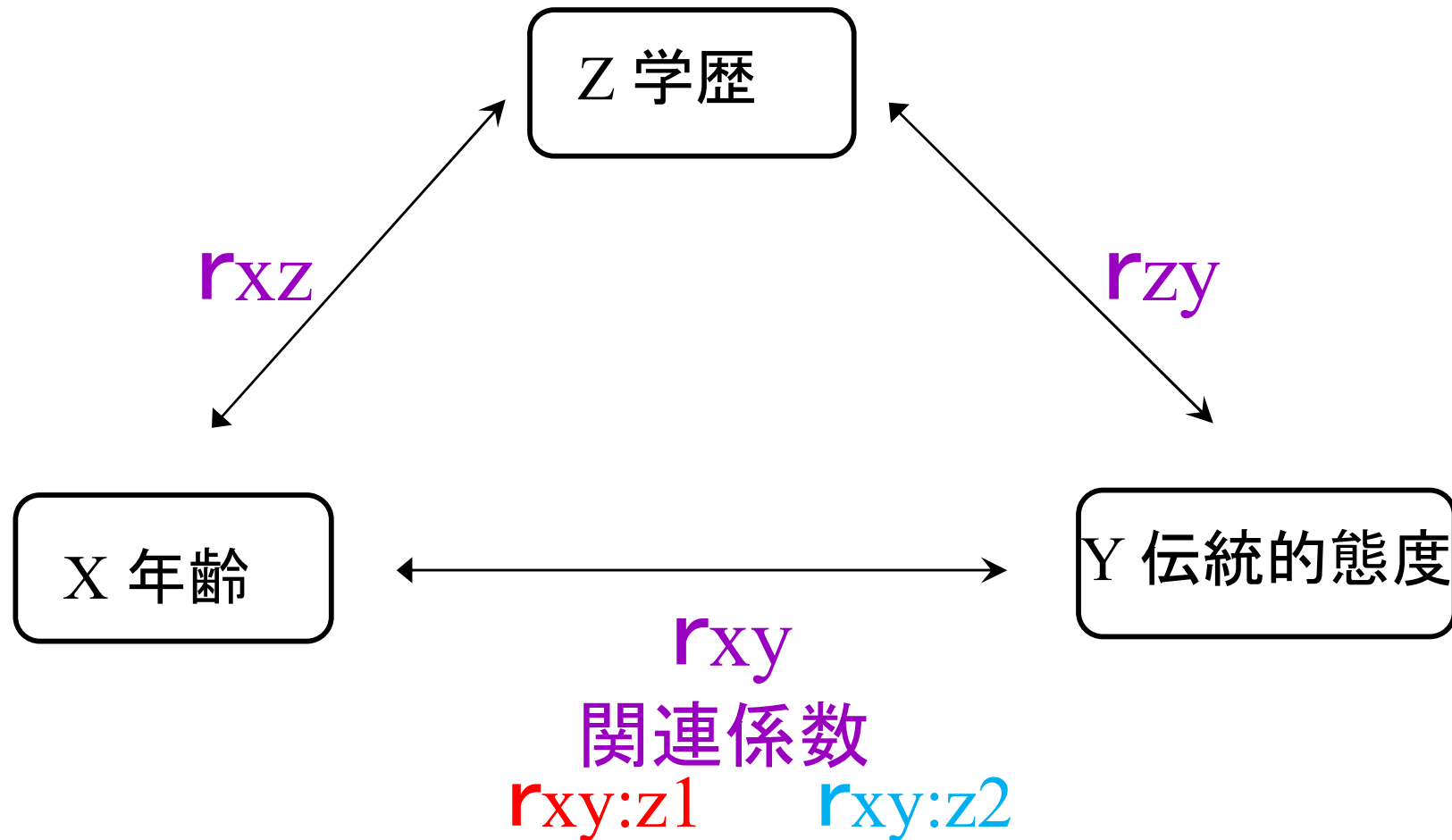
/STATISTICS=CHISQ PHI CORR

/CELLS=COL.



# 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。

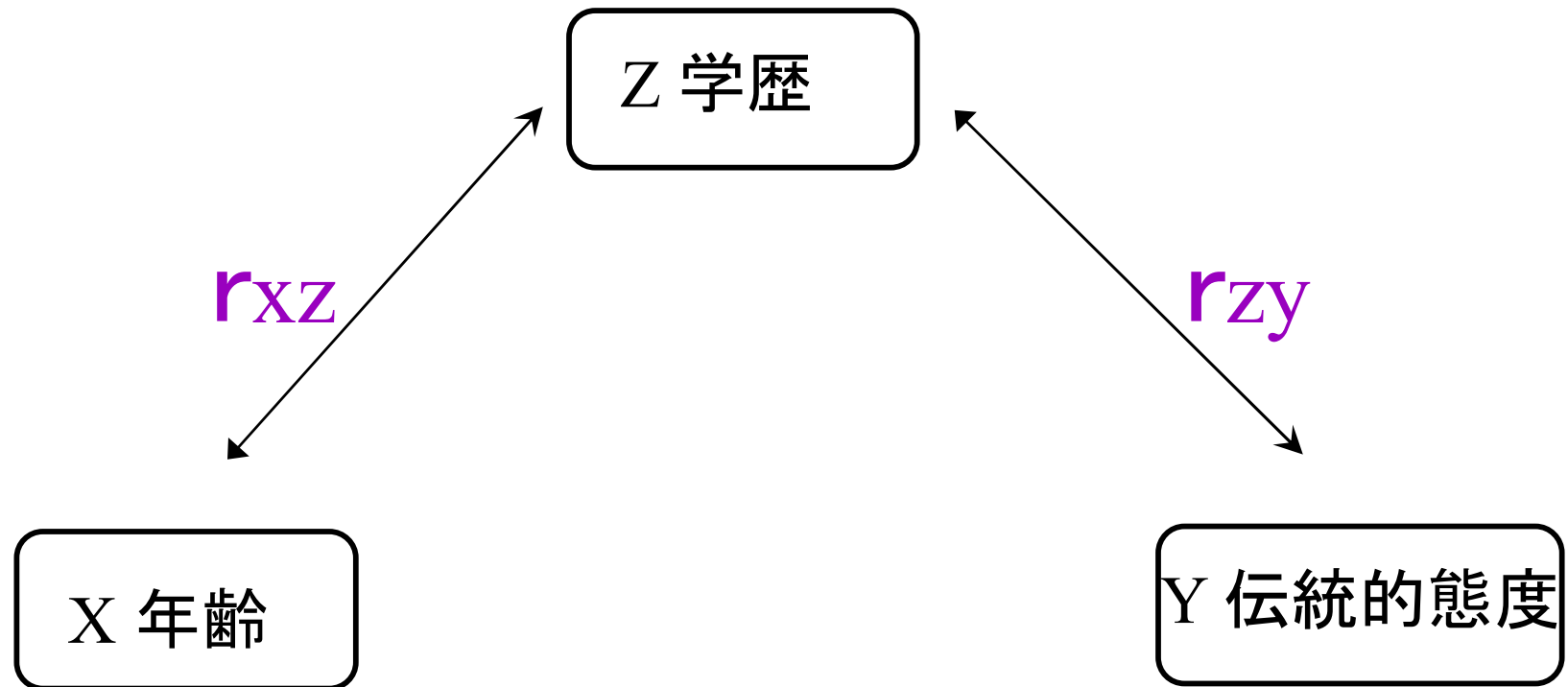


もし、最初にあった $r_{xy}$ が、分割時になくなった場合は、年齢XとYの間に直接効果なし、という結論になる。最初の関連は疑似相関だった。

# エラボレイションの結果をどう解釈するか

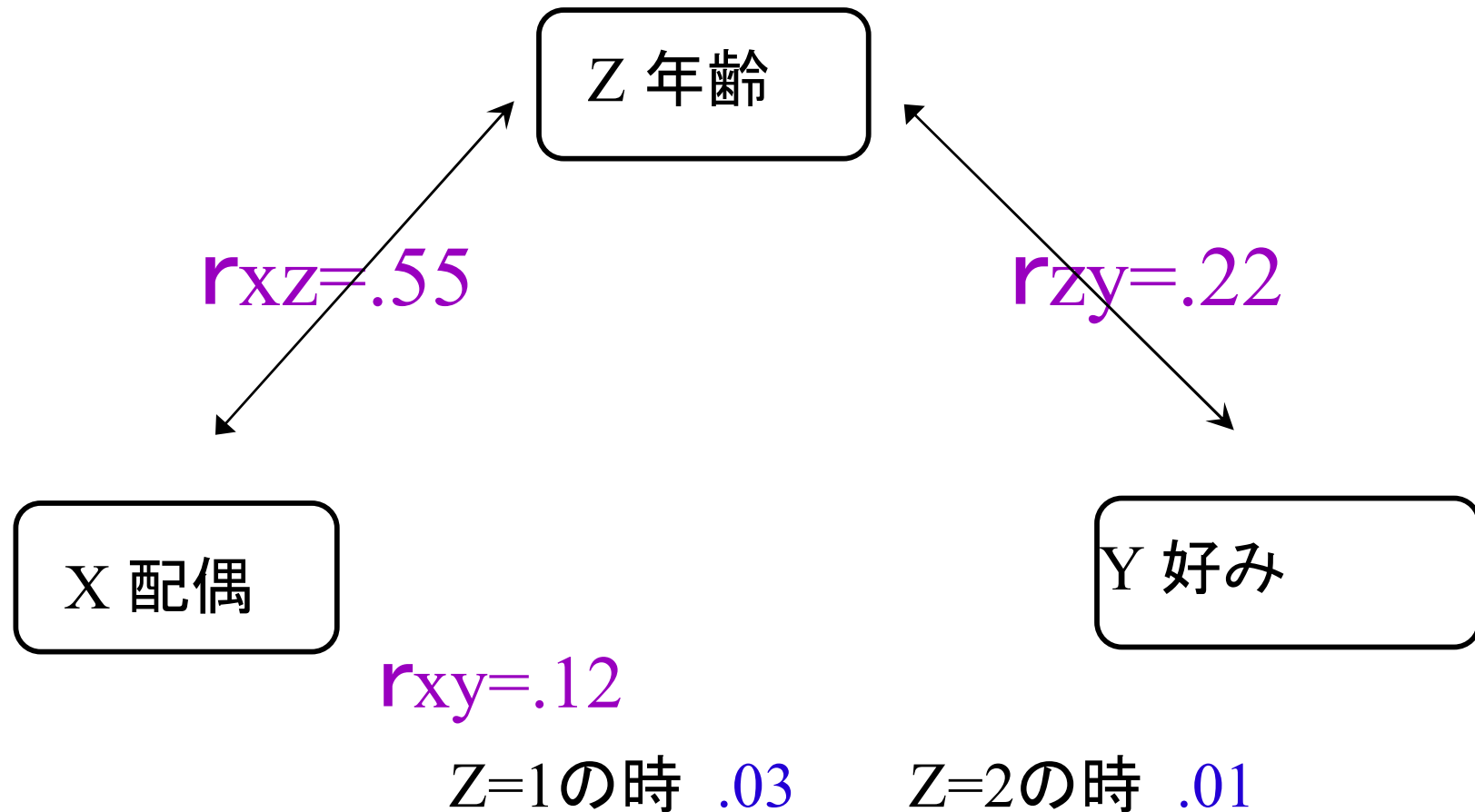
- ◆ Zを分析に投入した結果として、XYの関連がなくなった場合、Zが真の原因と考えることができる。
- ◆ しかし、XYの関連がなくならなかったとしても、因果関係の確認という意味がある。
- ◆ 詳しくは原・海野『社会調査演習』のエラボレイションの解説をよく読む。p.88のような表を自分で作る。関連係数を5個出すことを理解すること。

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。

## 『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。

# 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。



# 卒業研究 最終課題までに

- ◆ 主な仮説、分析の流れについて考える。
- ◆ 結果の解釈をもとに、さらに、仮説を作り直すことが大切！
- ◆ エラボレイションについては、3重クロス集計表と、関連係数を載せればよい。
- ◆ エラボレイション表の他に、2重クロス集計表を最低5つは載せること。10個以上を作って、よりよいものを載せるとよい。





# 10. 相関関係と因果関係

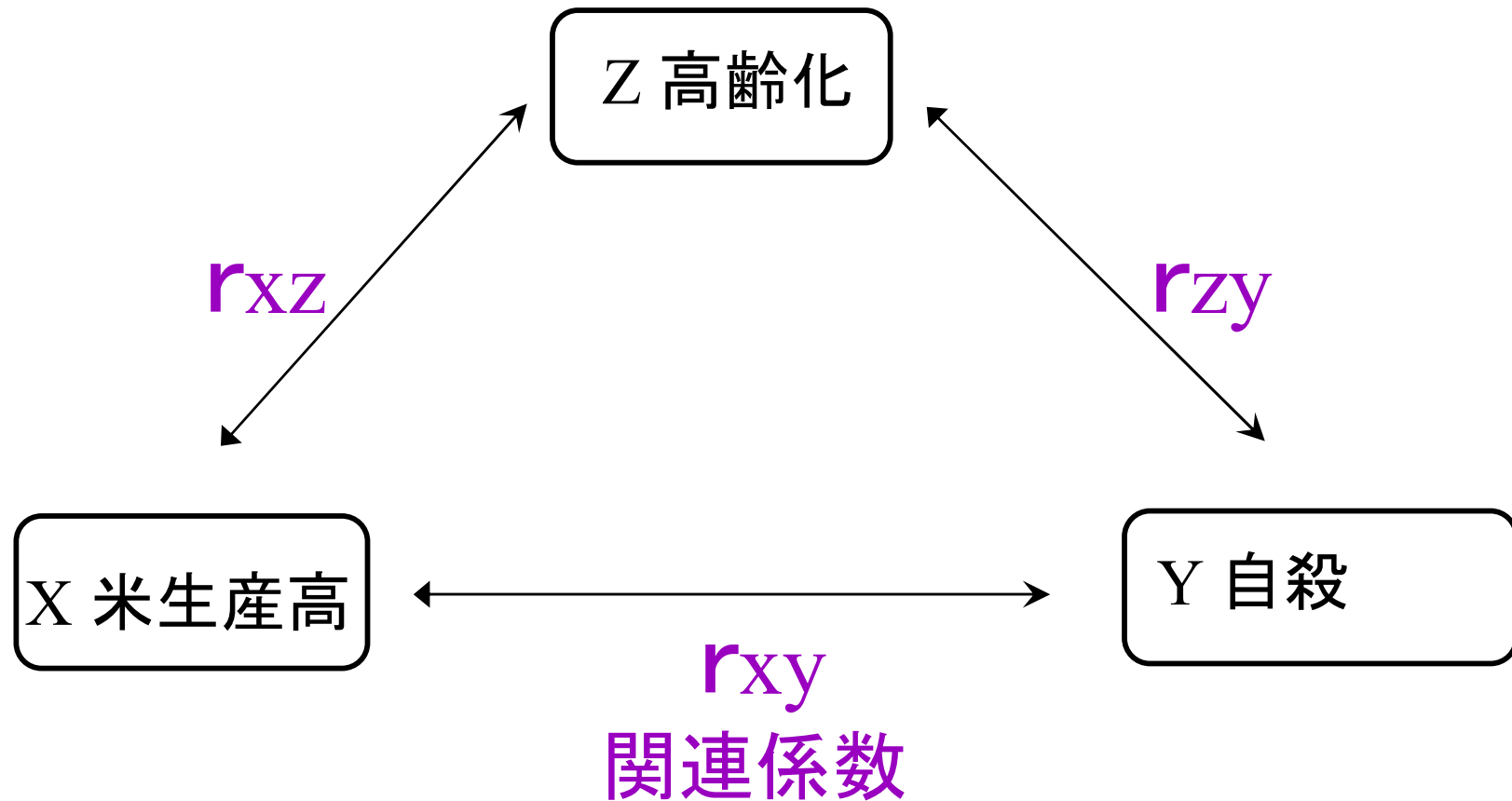
Yの原因として何がありうるか。



最初に見えた関連 $r_{xy}$ は、真の因果関係か

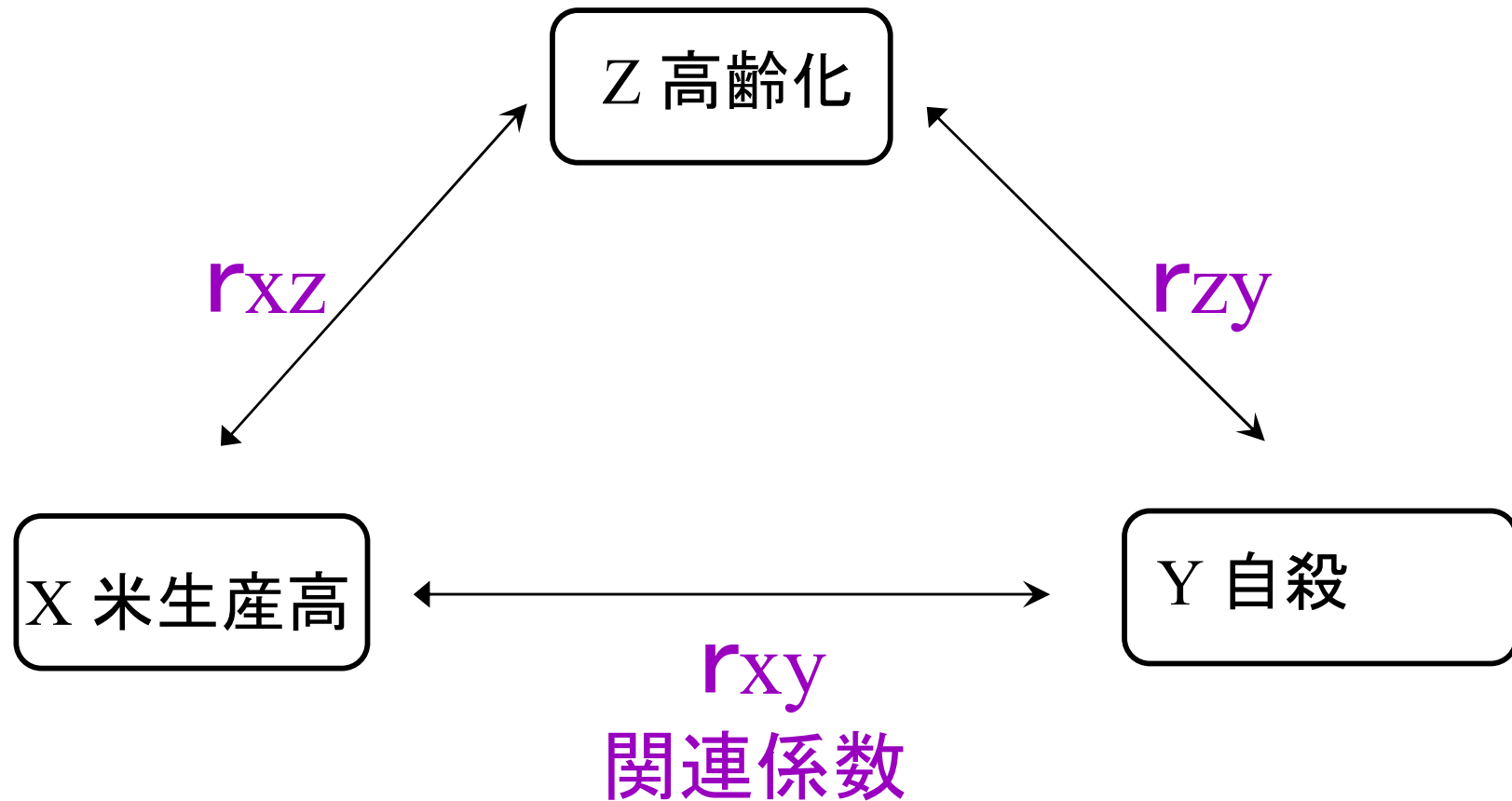
# 因果関係とについてよく考える

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。

# 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。

- ◆ 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