

第4回 クロス表と χ^2 乗検定(実習)

SPSS を使ってクロス表と χ^2 乗検定結果および Cramer's V を計算し、その結果を解釈する。

クロス表分析に必要なリコード (recode) の技法も習得する。

クロス表分析の結果をレポートにまとめる。

(1) システムの立ち上げから SPSS シンタックスエディタの読み込みまで (復習) *

*第2回の授業に欠席した人は、まず第2回のプリントにしたがってファイルを準備してください。第2回のプリントは、松本のWeb site (<http://www.rikkyo.ne.jp/web/ymatsumoto/index.html>) の「Office」→「社会学データ実習」からダウンロードできます。

- ①電源を入れる。
→ Windows XP と Vista の選択画面が出る。
- ② XP を選択する。
→ ID とパスワードの入力画面が出る。
- ③ ID とパスワードを入力して Windows にログオンする。
→ Windows が立ち上がる
- ④デスクトップにある「ホームディレクトリ」フォルダを開く。
→フォルダの中身が見える
- ⑤「tgss2000 授業用」フォルダをデスクトップにドラッグする。
→デスクトップに「tgss2000 授業用」フォルダがコピーされる。
- ⑥スタート→プログラム→プログラム・統計→ spss for windows → spss15.0J for windows をクリック
→初期画面が立ち上がる
- ⑦立ち上がっている spss 初期画面をキャンセル
→「無題 1」というデータファイル画面が残る。
- ⑧「無題 1」の F(ファイル) →開く→シンタックスをクリック
→シンタックスの読み取り画面が立ち上がる。
- ⑨ファイルの場所 デスクトップ→tgss2000 授業用をクリック
- ⑩ tgss00npr をクリック
→ファイル名欄に tgss00npr が入力される
- ⑪「開く」をクリック
→ spss シンタックスエディタが読み込まれる。

(2) クロス表の作成

【クロス表のコマンド】

```
cross 変数名1 by 変数名2 /cells=count row /statistics chisq phi.
```

正式のコマンドは `crosstabs tables` だが、`spss` は (百人一首と同じで) `cross` だけで通る。[変数名 1] は表側に、[変数名 2] は表頭に来る。

スラッシュ (/) 以下はオプション。 `cells` はクロス表のセル出力についての指定で、`count` は度数 (件数)、`row` は行 100 % の表示。 `column` とすると列 100 % となる。

変数名 1 を独立変数、変数名 2 を従属変数とすれば、行 100 % となるので、`row` を指定。逆に変数名 1 が従属変数、変数名 2 が独立変数なら、列 100 % となるので、`column` を指定。

つぎのスラッシュ以下は、統計の指定。 `chisq` は χ^2 乗検定結果 (`chi-squared`)、`phi` は ϕ 係数で、 2×2 以上のクロス表では `Cramer's V` が出力される。

【実習課題 1】 学歴と所得 (8 カテゴリー) との関連

本人の学歴と現在の仕事による収入とのあいだに関連があるかどうかを検討する。

独立変数 学歴 q43

従属変数 仕事による年収 q52

① `statistical command` の `freq` の行を次のように書き換える。

```
cross q43 by q52 /cells=count row /statistics chisq phi.
```

② (最初なので) 実行 → すべて

→ 出力結果が `spss` ビューアに現れる。

クロス表そのもの、 χ^2 乗検定結果、クラメールの `V` について確認する (結果は、ワードに貼り付けなくてよい)。

【実習課題 2】 学歴と所得 (4 カテゴリー) との関連

実習課題 1 では、現在の仕事収入のカテゴリー数が 8 であったため、クロス表が大きくなりすぎて読みにくくなっていた。この問題を解消するために、カテゴリーを 4 つにまとめる。新しい変数名を (まだ使っていない変数名であれば何でもよいが、ここでは説明の都合上、統一して) `income` とする。

入力場所は、読み取り書式の後ろであればどこでもよいが、とりあえず最終行で作業をすることにする。

① `compute` (計算) コマンドを使って、新変数 `income` に `q52` をそのままひき移す。

```
compute income=q52.
```

② `recode` (再コード) コマンドを使って、`income` のカテゴリーをふたつずつまとめる。

```
recode income (1,2=1) (3,4=2) (5,6=3) (7,8=4).
```

③変数ラベルをつける。出力の表が見やすくなるうえに、あとから見てもわかる。

`variable labels income "仕事収入".`

④値ラベルをつける。

`value labels income 1 "-300" 2 "300-700" 3 "700-1100" 4 "1100-".`

⑤確認のために度数分布表を出力するコマンドをつける。

`freq income.`

⑥カーソルを `compute` コマンドの行に戻して、「実行」→「最後まで」

→ `income` の度数分布表が出力される。

INCOME 仕事収入						
		度数	パーセント	有効パーセント	累積パーセント	
有効	1.00	-300	356	36.1	45.0	45.0
	2.00	300-700	261	26.4	33.0	78.0
	3.00	700-1100	129	13.1	16.3	94.3
	4.00	1100-	45	4.6	5.7	100.0
	合計	791	80.1	100.0		
欠損値	システム欠損値		196	19.9		
合計		987	100.0			

チェックポイント：4つのカテゴリーにうまく分けられていますか？

変数ラベルと値ラベルは出力されていますか？

欠損値の数は合っていますか？

⑦つぎに、学歴 `q43` と新変数 `income` とのクロス表を出力する（実習課題1の応用です）。

うまくいったら、ワードに、クロス表、 χ^2 自乗検定結果、Cramer's Vの計算結果を貼り付けておきましょう。

【実習課題3】年齢5カテゴリーと所得4カテゴリーとのクロス表を出力する。

①本人年齢 `q42b` を10歳刻みの新変数 `age10` に再コードして使用する。

`compute age10=q42b.`

`recode age10 (1,2=1) (3,4=2) (5,6=3) (7,8=4) (9,10,11=5).`

`value labels age10 1 "20-29" 2 "30-39" 3 "40-49" 4 "50-59" 5 "60-69".`

ここから先は、応用ですので自分でやってください。

② `age10` の度数分布表を出力し、うまく再コードされているかどうかを確認する。

③年齢5カテゴリーと所得4カテゴリーとのクロス表を出力する。 χ^2 自乗検定結果とクラメールのVもあわせて出力する。

④結果を解釈する（表を読んで考えましょう）。

【レポート課題】 実習課題 2、3 の結果をレポートにまとめる。

テーマ：「学歴と年齢が仕事の収入に及ぼす効果について」

仮説（必須）：

分析方法（必須）

分析結果（必須）：

考察（必須）：

授業についての感想・要望（自由）：

【チェックポイント】

- ・ 仮説：仮説が文章のかたちで明示されているかどうか。理論的仮説と操作的仮説が区別され、操作的仮説が理論的な仮説から導き出されているかどうか。
- ・ 分析方法：変数の定義が文章のかたちで明示されているかどうか。
- ・ 分析結果：クロス表が正しく作成されているかどうか、とくに%計算の方向が正しいかどうか。
- ・ 考察：クロス表の解釈が文章のかたちで適切に述べられているかどうか。理論的仮説との関連が述べられているかどうか。

【提出先】 レポートができたなら、「クロス表分析」という文書名をつけて、自分のフォルダに保存したうえで、V-Campus CHORUSの「社会学データ実習」→「レポート課題」→「クロス表分析」をひらき、ファイルを転送してください。締切は5月18日(月)24:00まで。

【注意】 PC上で作業を終えたあと、tgss2000npr.SPSのファイルを確実に保存し、デスクトップ上にあるフォルダをホームディレクトリに格納すること。今回行った変数変換は、次回以降も使いますので消さないようにしてください。