

現代日本論演習 / 比較現代日本論研究演習 III 「実践的統計分析」

第8講 対応のある平均値の差の検定

田中重人 (東北大学文学部准教授)

[テーマ] 二項検定; 対応のあるサンプルの平均値の差の分析

1 前回課題について

- 順序尺度以上の変数であって、同一の測定方式の変数の組み合わせを選ぶこと
- 回答選択肢の値の順序によっては、解釈を勘違いしやすいので、クロス表を見てよく考えること
- 検定統計量 (Z) は、正規分布を用いた比率の検定
- この種の分析の場合は、クロス集計表のセルに「全体」のパーセントを入れるとよい

2 二項検定

「符号検定」は、比率を使った検定の一種である

- $x = y$ のケースを除外しないで比率を求めることもできる
- 帰無仮説は、比率 = 0.5 でなくても、任意の比率を設定できる

このように、さまざまな比率の計算方法と帰無仮説の設定で統計的検定をおこなう一群の方法を「二項検定」と呼ぶ。

SPSS では、あらかじめ2つの変数の差を求め (下記参照)、「変数値の再割り当て」で2値変数を作って「分析」「ノンパラメトリック検定」「過去のダイアログ」「2項」を使う。

```
compute DIFF = X - Y.  
recode DIFF (lowest thru 0 = -1) (1 thru highest = 1) (missing=sysmis) into SIGN.
```

3 差の平均値の統計的推測

二つの変数の差について新たな変数を作ってみる：

- 「変換」「計算」
- 「目標変数」に適切な名前を
- 数式を作成 (compute 新変数 = 変数 x - 変数 y)
- シンタックス貼付、実行
- 度数分布 (「統計」オプションで平均、分散、SD、標準誤差を出力)

上で求めた「ふたつの変数間の差」の平均値について、信頼区間を求めるにはどうすればよいか?

4 平均値の差の統計的推測

二つの変数 x と y の平均値は母集団においてはどちらのほうが高いか？

差の変数 $x - y$ を作って、その平均値について区間推定するのと同じことになる

実際には、 x と y の平均値、標準偏差とそれらの間の Pearson の積率相関係数 r を使って計算できるので、そのやりかたがふつつかわれる。

[課題] 教科書 p. 192–197 の説明を読み、「対応」の有無によって計算方法がどのように変わるかを考える

5 SPSS コマンド

- 「平均値の比較」 「対応のあるサンプルの t 検定」
- 2変数の組を選択してパレットに入れる

6 宿題

適当な変数について、SPSS で次の二つの分析を行い、結果が同じになることを確かめる (ISTU で水曜正午まで)

- 差の変数を作って、その平均値の区間推定をおこなう
- 「対応のある」 t 検定をおこなう

7 今後の予定

次回以降は「多変量解析」に入ります。とりあげてほしい多変量解析手法がある場合は、田中まで。(希望がなければ、重回帰分析をとりあげます。)