

第7講 符号検定

田中重人 (東北大学文学部准教授)

[テーマ] 対応のある項目間の比較

1 変数間の関連と比較

- 類似性・因果関係 相関係数行列
- どの変数がより高い/低い? 変数間の比較 (対応のある分析、被験者内要因)
 - 例: 10歳から20歳間の身長の変化
 - 例: 授業の前後での知識の変化
 - 例: 「見れる」と「起きれる」ではどちらのほうが受容度が高いか?
 - 例: 問27の8項目のうち、最も「重要」と評価されているもの/されていないものはどれか?

とりあえず、平均値を並べて比較するには:

PSPP では「分析」 「記述統計量」 「記述統計量」で多くの変数の平均値 (と標準偏差) を並べて出力できる。

2 「対応」とは

同一ケースが複数の項目に回答している場合、項目間に「対応」があるという。

このような項目の比較には、対応を考慮した分析法を使う必要がある。

例題: つぎのような 3 ケース × 2 変数 のデータについて、どちらの変数が大きくなる傾向にあるかを考えてみよう:

	変数 A		変数 B	差
ケース X	4	-	5	=
ケース Y	3	-	4	=
ケース Z	5	-	1	=

平均値

多数決をとると?

このように、ケース間の異質性が大きい場合は、対応を考慮して分析しないと、データの特徴をつかみそこねる可能性がある。

3 ふたつの変数での大小の比較

3.1 例題

問27について、「B. 高い収入」 vs. 「D. 家族からの信頼・尊敬」 どちらが大切?

B>D の人と B<D の人のどちらが多いか?

クロス表を出力して数えてみよう(セルに「合計」のパーセントを入れるとよい)。

これら2つ以外に B=D の人がいるから、これをどうあつかうかが問題になる。

3.2 符号検定

「符号検定」(sign test) とは...

1. $A = B$ のケースを除外
2. 帰無仮説: 「母集団では同数」 (=50%)
3. 正規分布を利用して検定をおこなう

この考えかたは、比率の区間推定とおなじものなので、 $\text{比率} \pm \text{臨界値} \times \text{標準誤差}$ で信頼区間を求めていると考えてもよい

3.3 PSPP コマンド

- 「ノンパラメトリック検定」 「2 related samples」
- 比較したい変数をペアで 指定する
- 「符号検定」をチェック

4 課題

適当な2変数について、符号検定を行う。クロス表(または相関図)も出力して、結果を解釈すること。