

第10講 分散分析

田中重人 (東北大学文学部准教授)

[テーマ] 分散分析 (ANOVA) の考えかたと計算方法を理解する

1 分散分析の考えかた

グループ別の平均値を当てはめて仮想の分散を求める分析法を「分散分析」(ANOVA: ANalysis Of VAriance) という。

- 従属変数 (dependent variable) と独立変数 (independent variable)

相関比 (イータ) の性質:

- 最小値:
- 最大値:

大きさの評価基準は、Cramer の連関係数 V と同様。

なぜ相関比を求めると、平均値を比較していることになるのか?

2 課題

次のデータ (11人) について、分散分析を行なう

男性: 1, 2, 3, 3, 4, 4

女性: 2, 3, 4, 4, 5

まず手計算で考えてみて、そのあと、SPSSにデータを入力して検算する。

3 相関比とエフェクトサイズの関係

相関比 η とエフェクトサイズ ES の間にはつぎの関係がある (n_1, n_2 は各グループの度数、 $N = n_1 + n_2$ は全体の度数)。

$$ES^2 = \frac{\eta^2}{1 - \eta^2} \times \frac{N^2}{n_1 n_2} \quad (1)$$

特に、2グループの度数が等しい ($n_1 = n_2$) なら、この式は次のようになる。

$$ES^2 = \frac{4\eta^2}{1-\eta^2} \quad (2)$$

(グループの度数が違えば、ESはこれより大きくなる)

さらに、 η があまり大きくない ($\eta < 0.4$ 程度) 場合であれば、次のような単純な式で近似できる：

$$ES = 2 \eta$$

4 モデルとデータの乖離

相関比も、モデルとデータの乖離を表した値と解釈できる

- 「モデル」は何か？
- データとの乖離はどうやって計算しているか？
- 係数の取りうる値の範囲は？

5 表の書きかた

- 各層と全体の平均値と標準偏差 (測定水準の2桁下まで)
- 各層と全体の人数
- 相関比またはエフェクトサイズ (小数第3位まで)
- 欠損数とその原因

6 グラフの書きかた

平均値をプロットし、上下にSDを表示する。誤差範囲 (error bar; 別名「ヒゲ」) にはSD以外を書く場合もあるので、必ず「±標準偏差」であることを明記する。

Excel では

- SPSS 出力をシートにはりつける
- 折れ線グラフを描く
- メニューの「レイアウト」から「誤差範囲」→「その他の誤差範囲オプション」をえらぶ
- 「ユーザ設定」→「値の指定」
- 「正の誤差の値」「負の誤差の値」にSDが入っているセル範囲を指定 (おなじものでよい)

より詳細に分布の違いを検討したいときは、グループ別に度数ポリゴン (または折れ線グラフ) を描いてもよい。