

第12講 分散分析・2 (7/17)

田中重人 (東北大学文学部准教授)

[テーマ] 平均値を比較した表とグラフの書きかた

1 前回課題について

男性: 1, 2, 3, 3, 4, 4 → 平均 2.833

女性: 2, 3, 4, 4, 5 → 平均 3.6

全体の平均: 3.182

グループ別平均値を当てはめた「仮想」データの平方和は、つぎのようになる。下線部に注意。

$$\text{グループ間平方和} = 6(\underline{2.833 - 3.182})^2 + 5(\underline{3.6 - 3.182})^2 = 1.603 \quad (1)$$

これを $n (=11)$ で割って平方根をとると標準偏差が得られる。

$$\text{仮想SD} = \sqrt{\frac{1.603}{11}} = 0.382 \quad (2)$$

$$\eta = \frac{\text{仮想SD}}{\text{実際のSD}} = \frac{0.382}{1.113} = 0.343 \quad (3)$$

ただし、SPSS では平方和を $n - 1 (=10)$ で割って「標準偏差」を求めているので、注意。度数がある程度大きくなれば (およそ $n > 200$ の場合)、このことによる違いは気にしなくてよい。

2 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)。

$$ES^2 = \frac{\eta^2}{1 - \eta^2} \times \frac{N^2}{n_1 n_2} \quad (4)$$

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。

$$ES^2 = \frac{4\eta^2}{1 - \eta^2} \quad (5)$$

(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) であれば、次のような単純な式で近似できる：

$$ES = 2 \eta$$

3 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？

4 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

5 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 (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を描いてもよい。