

第9講 分散分析

田中重人 (東北大学文学部教授)

[テーマ] 分散分析 (ANOVA) の考えかたと計算方法を理解する

1 前回課題について

- 間隔尺度とみなせる変数を使うこと
- グラフの縦軸は % にすること (クロス表を使ったほうがいいのかも)

2 PSPP 「グループの平均」の表示桁数

PSPP の「平均の比較」「グループの平均」では、データセットの「Decimal」の設定値で小数値が丸められてしまう。Decimal の桁数を変えれば、その桁数で平均値と SD が表示される。

「一元配置分散分析」なら、Decimal の設定にかかわらず、小数第 2 位まで表示される：

- メニューの「分析」「平均の比較」「一元配置分散分析」を選択
- 「従属変数リスト」と「因子」を指定
- 「記述統計量」をチェック

3 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？

4 課題

次のデータ (10 人) について、分散分析を行なう

男性: 1, 2, 3, 3, 4

女性: 2, 3, 4, 4, 5

まず手計算 (またはスプレッドシート) で考えてみて、そのあと、PSPP にデータを入力して検算する。

- (1) 全体の平均値と SD を求める
- (2) 男女別の平均値を求める
- (3) 男性の平均値 × 5 人と女性の平均値 × 5 人からなる仮想データを考えて SD を求める
- (4) (3) の SD を (1) の SD で割ったものが相関比

この相関比がなぜ「平均値の比較」の指標になるかを考えること。

PSPP 出力の「分散分析」表では、「群間」が (3) に、「合計」が (1) に対応しているので、割り算して平方根をとると、相関比 が求められる。

5 相関比とエフェクトサイズの関係

相関比 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 (< 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:

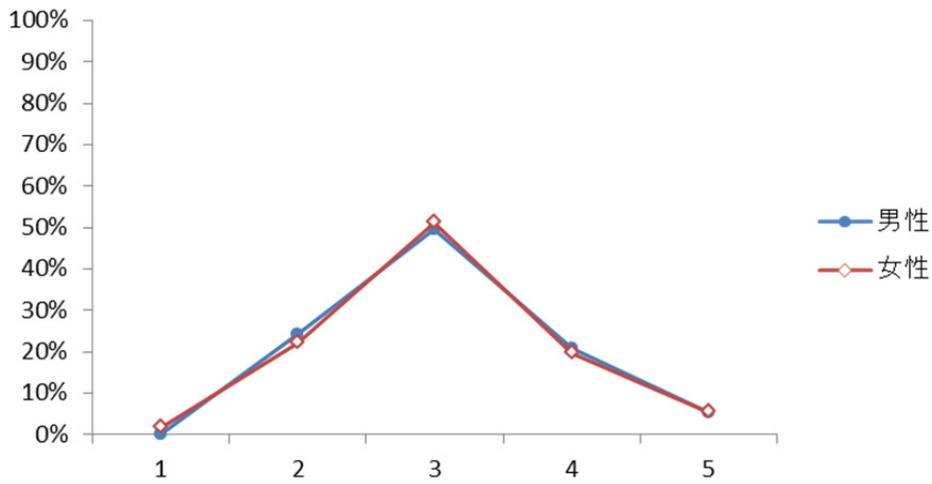
$$ES = 2\eta \quad (3)$$

6 モデルとデータの乖離

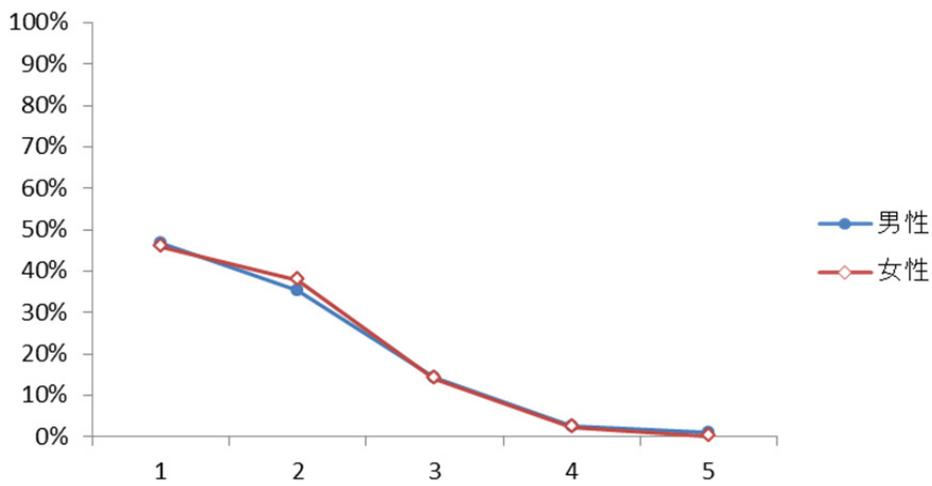
相関比 は、モデルとデータの乖離を表した値と解釈できる

- 「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?

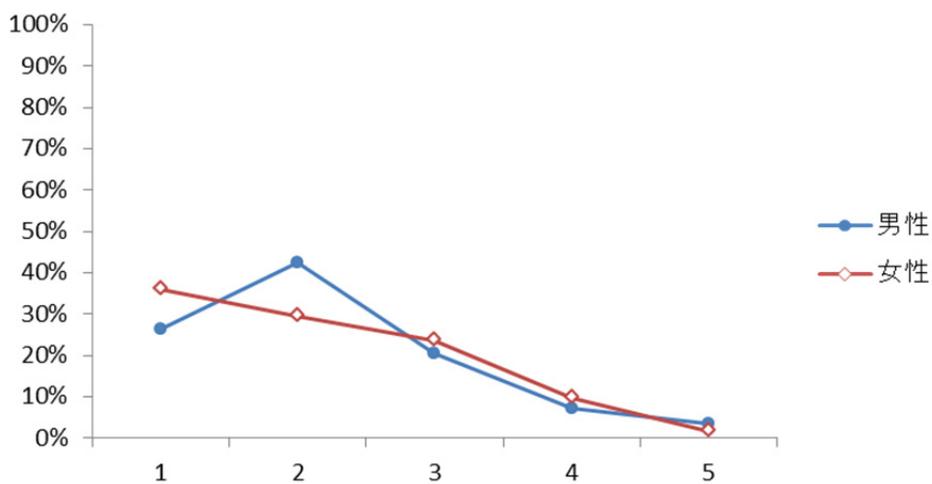
階層帰属意識5



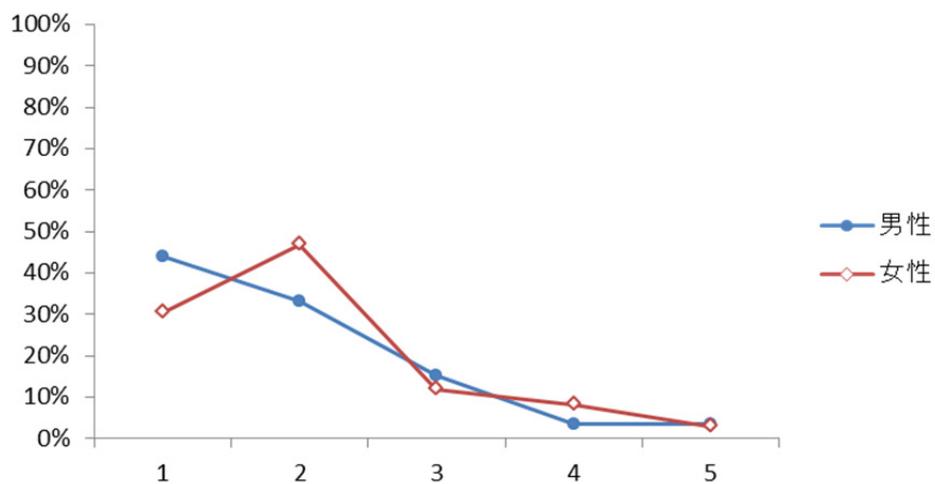
心の豊かさ



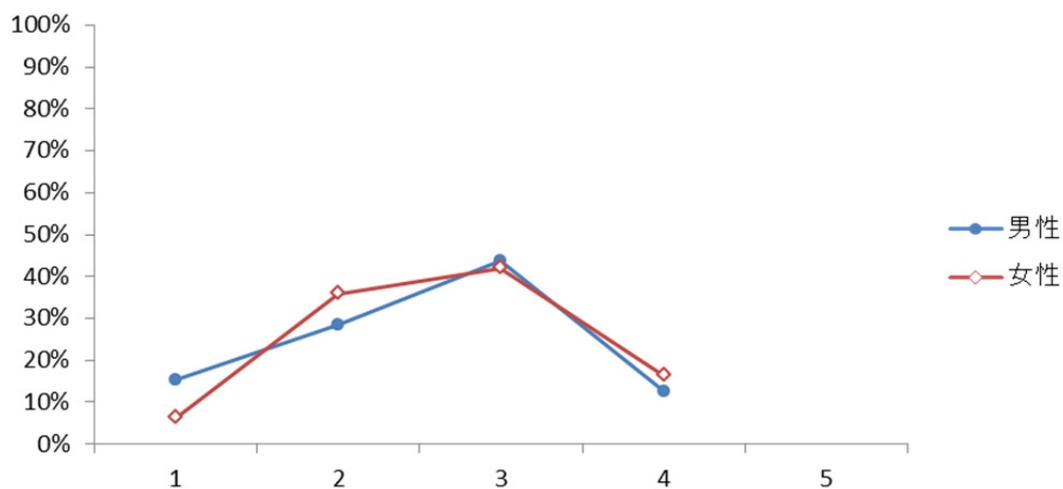
がんばっている



女は評価されない



評価高い職業



獲得したものを維持

