

分散分析

B3 60171018 片桐直哉

参考文献



- 「Rによるやさしい統計学」

- 山田剛史
- 杉澤武俊
- 村井潤一郎[共著]

- オーム社出版

- 第7章 分散分析

分散分析とは？

- 二つの平均値の比較を行う → t検定
- 三つ以上は？ → 分散分析
 - 学年全体でテストを受ける際、担当の教員によって結果に差がでるのか

準備

A	B	C	D
15	13	10	10
9	8	6	7
18	8	11	3
14	12	7	5
18	7	12	7

準備

- 帰無仮説 H_0
 - すべての群の母平均は等しい
- 対立仮説 H_1
 - すべての群の母平均は等しいとはいえない

分析手順(1)

● 要因にもたらされた差と、誤差による差の分解

- ① 全体の平均値 μ を求める
- ② 各要因ごとの平均 x_i (i は要因の数)を求める
- ③ x_i と μ の差をとる(この差を要因の**効果**という)
- ④ 各群の平均と、群内のデータの差をとる

これにより、すべてのデータは、
観測値 = μ + 効果 + 誤差

データの分解

● 観測値 = μ + 効果 + 誤差

A	B	C	D
15	13	10	10
9	8	6	7
18	8	11	3
14	12	7	5
18	7	12	7

$$\begin{bmatrix} 15 & 13 & 10 & 10 \\ 9 & 8 & 6 & 7 \\ 18 & 8 & 11 & 3 \\ 14 & 12 & 7 & 5 \\ 18 & 7 & 12 & 7 \end{bmatrix} = \begin{bmatrix} 10 & 10 & 10 & 10 \\ 10 & 10 & 10 & 10 \\ 10 & 10 & 10 & 10 \\ 10 & 10 & 10 & 10 \\ 10 & 10 & 10 & 10 \end{bmatrix} + \begin{bmatrix} 4.8 & -0.4 & -0.8 & -3.6 \\ 4.8 & -0.4 & -0.8 & -3.6 \\ 4.8 & -0.4 & -0.8 & -3.6 \\ 4.8 & -0.4 & -0.8 & -3.6 \\ 4.8 & -0.4 & -0.8 & -3.6 \end{bmatrix} + \begin{bmatrix} 0.2 & 3.4 & 0.8 & 3.6 \\ -5.8 & -1.6 & -3.2 & 0.6 \\ 3.2 & -1.6 & 1.8 & -3.4 \\ -0.8 & 2.4 & -2.2 & -1.4 \\ 3.2 & -2.6 & 2.8 & 0.6 \end{bmatrix}$$

分析手順(2)

- F値による判定

- 求めたF値とF分布表を参照して比較することで帰無仮説を棄却するか決める

- $$F = \frac{\text{群間平方和} / \text{群間の自由度}}{\text{群内平方和} / \text{群内の自由度}}$$

- 群間の自由度 = 群の数 - 1
- 群内の自由度 = (各群のデータ数 - 1)を全ての群について合計したもの

F値の算出

$$\begin{bmatrix} 15 & 13 & 10 & 10 \\ 9 & 8 & 6 & 7 \\ 18 & 8 & 11 & 3 \\ 14 & 12 & 7 & 5 \\ 18 & 7 & 12 & 7 \end{bmatrix} = \begin{bmatrix} 10 & 10 & 10 & 10 \\ 10 & 10 & 10 & 10 \\ 10 & 10 & 10 & 10 \\ 10 & 10 & 10 & 10 \\ 10 & 10 & 10 & 10 \end{bmatrix} + \begin{bmatrix} 4.8 & -0.4 & -0.8 & -3.6 \\ 4.8 & -0.4 & -0.8 & -3.6 \\ 4.8 & -0.4 & -0.8 & -3.6 \\ 4.8 & -0.4 & -0.8 & -3.6 \\ 4.8 & -0.4 & -0.8 & -3.6 \end{bmatrix} + \begin{bmatrix} 0.2 & 3.4 & 0.8 & 3.6 \\ -5.8 & -1.6 & -3.2 & 0.6 \\ 3.2 & -1.6 & 1.8 & -3.4 \\ -0.8 & 2.4 & -2.2 & -1.4 \\ 3.2 & -2.6 & 2.8 & 0.6 \end{bmatrix}$$

● 群間平方和 = 184

● 群内平方和 = 138

● 群間の自由度 = 3

● 群内の自由度 = 16

● F値 = 7.111111

$$F = \frac{\text{群間平方和} / \text{群間の自由度}}{\text{群内平方和} / \text{群内の自由度}}$$

分散分析表

	自由度 Df	偏差平方和 Sum Sq	平均平方和 Mean Sq	分散比 F値
与えた要因	3	184.000	61.333	7.1111
誤差	16	138.000	8.625	
全体	19	322.000		

分析手順(3)

- 求めたF値と、F分布表から比較
 - 分子の自由度3, 分母の自由度16のときのF値は3.24
 - 求めたF値のほうが大きければ、帰無仮説を棄却
- 要因には優位性があるといえる

Tukeyの方法

$$\bullet q = \frac{|\text{比較する群の平均偏差}|}{\sqrt{\frac{\text{群内の平均平方}}{\text{各群のデータ数}}}}$$

まとめ

- 2つ以上の標本について、平均値に有意差があるか、判定するために分散分析を使う
- 帰無仮説、対立仮説を決定し、観測値を分析・分解して、F値を求めることで優位性があるかを判定できる
- 要因間の優位性を確認する際は、多重比較を用いるとよい