

一元配置分散分析

高木 廣文
東邦大学医学部看護学科
国際保健看護学研究室

1

多グループの母平均値の比較

- 3つ以上のグループでの母平均値の差の検定の方法を解説する。

- ・ 一元配置分散分析
パラメトリック検定
- ・ Kruskal-Wallis検定
ノンパラメトリック検定

2

●一元配置分散分析

表から疾患により尿中細菌数が異なるといえるだろうか。

急性膀胱炎	急性腎盂腎炎	急性副睾丸炎
5.5×10^6	4.6×10^5	3.1×10^5
6.3×10^6	8.3×10^6	0
5.7×10^7	6.7×10^4	5.3×10^6
8.1×10^7	2.9×10^5	3.6×10^4
3.2×10^5	7.4×10^5	1.2×10^3
1.5×10^7	7.9×10^5	4.7×10^4
7.7×10^7	8.9×10^4	0
8.8×10^6	1.3×10^6	0
4.4×10^5 (個/ml)		

3

●一元配置分散分析の簡単な理論

- ・ 変数Xについて、k個のグループでそれぞれ独立に測定。
- ・ 第i群での標本数を n_i 、各データを x_{ij} ($i=1, 2, \dots, k, j=1, 2, \dots, n_i$) .

- Q. 各群間の母平均値に差があるか？
帰無仮説：
「各群の母平均値はすべて等しい」

4

●各グループ平均と総平均

- ・ 第iグループ平均値 x_{Mi} :

$$x_{Mi} = \frac{1}{n_i} \sum x_{ij}$$

- ・ 全体の総平均値 x_T 、Nは総標本数 :

$$x_T = \frac{1}{N} \sum_i \sum_j x_{ij} = \frac{1}{N} \sum_i n_i x_{Mi}$$

5

●全変動

- ・ データの全体的なばらつき具合：
総平均値からの偏差の2乗和T

$$T = \sum_i \sum_j (x_{ij} - x_T)^2$$

6

● 級間変動

- すべての変動は、各群の平均値にあると仮定した場合の変動—級間変動B :

$$B = \sum_i n_i (X_{mi} - X_T)^2$$

7

● 級内変動

- 各グループ内の変動の合計W :

$$W = \sum_i \sum_j (x_{ij} - x_{mi})^2$$

級内変動、残差変動、誤差変動

8

● 各変動間の関係

全変動 T = 級間変動 B + 級内変動 W

$$T = B + W$$

9

● 級間変動と級内(残差)変動

- 級間変動 :
各群の平均値と総平均値の間のばらつき
- 残差変動 :
グループ内のばらつき

BとWを用いて、群間の差と残差のどちらが大きいかを調べる。

10

● 級間分散と級内(残差)分散

- 各不偏分散 (平均平方和) :
級間変動と残差変動をその自由度で割る。
- ・ 級間分散 V_B :

$$V_B = B / (k-1)$$

- ・ 残差分散 V_W :

$$V_W = W / (N-k)$$

F検定の方法

- 帰無仮説 : 「各群の母平均値は等しい」

$$F_0 = V_B / V_W$$

F_0 : 自由度 (k-1, N-k) の F 分布に従う。

12

F検定の方法2

- 自由度 $(k-1, N-k)$ のF分布の上側 α 点の値 $F_\alpha(k-1, N-k)$:

$$F_0 > F_\alpha(k-1, N-k) \Rightarrow \text{有意水準 } \alpha \text{ (5\%か1\%など) で仮説を棄却}$$

- ・有意の場合 :
少なくとも1つのグループの平均値は他と異なるものと考えてよい。
- ・有意でない場合 : 消極的に仮説を採択し、差があるとはいえない。

13

●例題の計算1

- ・急性膀胱炎をA群, 急性腎盂腎炎をB群, 急性副睾丸炎をC群。
- ・各群の平均値 x_A, x_B, x_C , および全体の平均値 x_T :

$$\begin{aligned}x_A &= 2.79289 \times 10^7 \\x_B &= 1.50450 \times 10^6 \\x_C &= 7.11775 \times 10^5\end{aligned}$$

および

$$x_T = 1.07636 \times 10^7$$

14

●例題の計算2

- ・全変動 T :

$$T = 1.33155 \times 10^{16}$$

- ・級間変動 B :

$$\begin{aligned}B &= 9 \times (x_A - x_T)^2 + 8 \times (x_B - x_T)^2 \\&\quad + 8 \times (x_C - x_T)^2 = 4.14599 \times 10^{15}\end{aligned}$$

- ・残差変動 W :

$$W = T - B = 9.16956 \times 10^{15}$$

15

●例題の計算3

- ・級間分散 V_B :

$$V_B = B / (3-1) = 2.07299 \times 10^{15}$$

- ・残差分散 V_W :

$$V_W = W / (25-3) = 4.16798 \times 10^{14}$$

16

●例題の計算4

帰無仮説 :

「各群の母平均値は等しい」の検定

$$F_0 = V_B / V_W = 4.973$$

F_0 は自由度 $(2, 22)$ のF分布に従う。

$$p = 0.017$$

したがって, 5%水準で有意差あり。

17

●例題の計算5

【結論】

5%水準で仮説は棄却され, 少なくとも1つは母平均値の異なる群があるものと考えてよい。

18

●多重比較の方法

- 一元配置分散分析の結果が有意の場合

一元配置分散分析の結果が有意である場合、どの群間の母平均値に有意差があるのか？

- 各2群間の母平均値の差について検定を行う。⇒ 対比較 (対比)

19

●多重比較の問題点

- グループ数がk個ある場合：
各2群間のすべての組合せ数＝
k個から2つを選ぶ組合せの個数＝

$$k(k-1)/2 \text{ 個}$$

20

●グループ数が5の場合

- 組合せ数は10通り (${}_5C_2=10$)

各検定の有意水準が5%の場合：
全体の有意水準の設定：

$$1 - (1 - 0.05)^{10} = 0.40 \text{ (40\%)}$$

極めて大きな有意水準の設定と同等

21

●Bonferroniの方法

- 全体の検定水準を α ：

$$\text{各検定的水準} = 2\alpha / [k(k-1)]$$

- グループ数が大きくなるにつれ、個々の検定の有意水準は極めて小さくなる。
(有意水準の設定としては厳し過ぎ)

22

●シェッフエ(Scheffe)の方法1

- k個のグループ
- 変数Xの母平均値を $\mu_1, \mu_2, \dots, \mu_k$
- 各平均値の線形結合を考える
- $w_i (i=1, 2, \dots, k)$ ：重み

$$\mu = w_1 \mu_1 + w_2 \mu_2 + \dots + w_k \mu_k$$

$$w_1 + w_2 + \dots + w_k = 0$$

23

●シェッフエ(Scheffe)の方法2

- μ の期待値 $E(\mu)$ と分散 $V(\mu)$

$$E(\mu) = w_1 \times m_1 + w_2 \times m_2 + \dots + w_k \times m_k = \bar{x}_M$$

$$V(\mu) = \sigma^2 \left[\frac{w_1^2}{n_1} + \dots + \frac{w_k^2}{n_k} \right]$$

n_1, \dots, n_k ：各群の標本数，

σ^2 ：誤差分散

24

●シェッフェ(Scheffe)の方法3

- σ^2 の推定: V_W (級内分散, 残差分散)
- 分散 S_M^2 :

$$S_M^2 = V_W [\sum_i w_i^2 / n_i]$$

25

●シェッフェ(Scheffe)の方法4

- 母平均 μ の100(1- α)信頼区間
- 自由度(k-1, N-k)のF分布の上側 α 点を $F_\alpha(k-1, N-k)$:
- μ の信頼区間の上側点 μ_U , 下側点 μ_L

$$d = \sqrt{[(k-1) \cdot F_\alpha(k-1, N-k)]}$$
$$\mu_U = X_M + d \cdot S_M$$
$$\mu_L = X_M - d \cdot S_M$$

- 信頼区間(μ_L, μ_U)の間に0が含まれなければ水準 α でグループ間に有意差あり。 26

●シェッフェ(Scheffe)の方法5

◎実際の計算:
一方のグループに含まれる群の個数をK, 他方をLとした場合, 比較に用いない群の平均値にかける重みは0。

- 前者の重み=1/K, 後者の重み=-1/L
- 全体での重みの合計が0になるようする
- 2群の母平均値の検定: 1と-1の重み

27

●シェッフェ(Scheffe)の方法6

検定すべき群の組合せの個数が多い場合, やや面倒。

$$|X_M| > \Delta = d \cdot S_M \Rightarrow \text{有意差あり!}$$

28

【例題の計算】

• 先に表に示した尿中細菌数について, 「シェッフェの方法」により各群の平均値の差を検定してみよう。

- 検定全体の有意水準は0.05(5%)。
- 自由度(2, 22)のF分布の上側5%点3.443

$$d = \sqrt{[(3-1) \times 3.443]} = 2.642$$

29

計算1

(1) A群とB群の母平均値の差の検定:

$$\mu_{A-B} = |x_A - x_B| = 2.642 \times 10^7$$

• 検定のための差の有意限界:

$$\Delta_{A-B} = d \times \sqrt{[V_W \times \{1^2/9 + (-1)^2/8\}]}$$
$$= 2.612 \times 10^7$$

$$\Rightarrow \mu_{A-B} > \Delta_{A-B} \rightarrow \text{有意差あり}$$

30

計算2

(2) A群とC群の母平均値の差の検定:

$$\mu_{A-C} = |x_A - x_C| = 2.722 \times 10^7$$

・検定のための差の有意限界:

$$\Delta_{A-C} = d \times \sqrt{[V_W \times \{1^2/9 + (-1)^2/8\}]} \\ = 2.612 \times 10^7$$

$$\mu_{A-C} > \Delta_{A-C} \rightarrow \text{有意差あり}$$

31

計算3

(3) B群とC群の母平均値の差の検定:

$$\mu_{B-C} = |X_B - X_C| = 7.927 \times 10^5$$

$$\Delta_{B-C} = d \times \sqrt{[V_W \times \{1^2/8 + (-1)^2/8\}]} \\ = 2.697 \times 10^7$$

$$\Rightarrow \mu_{B-C} < \Delta_{B-C} \rightarrow \text{有意差なし}$$

32

計算4

(4) A群とB, C群(2群):

- ・A群 : 1の重み
- B, C群 : 各-1/2の重み

$$\mu_{A-BC} = |X_A - (X_B + X_C)/2| = 2.682 \times 10^7$$

$$\Delta_{A-BC} = d \times \sqrt{[V_W \times \{1^2/9 + (-1/2)^2 \times (1/8 + 1/8)\}]} \\ = 2.247 \times 10^7$$

$$\Rightarrow \mu_{A-BC} > \Delta_{A-BC} \rightarrow \text{有意差あり}$$

33

計算5

(5) B群とA, C群(2群)の差:

$$\mu_{B-AC} = |X_B - (X_A + X_C)/2| = 1.282 \times 10^7$$

$$\Delta_{B-AC} = d \times \sqrt{[V_W \times \{1^2/8 + (-1/2)^2 \times (1/9 + 1/8)\}]} \\ = 2.314 \times 10^7$$

$$\Rightarrow \mu_{B-AC} < \Delta_{B-AC} \rightarrow \text{有意差なし}$$

34

計算6

(6) C群とA, B群(2群)の差:

$$\mu_{C-AB} = |X_C - (X_A + X_B)/2| = 1.400 \times 10^7$$

$$\Delta_{C-AB} = d \times \sqrt{[V_W \times \{1^2/8 + (-1/2)^2 \times (1/9 + 1/8)\}]} \\ = 2.314 \times 10^7$$

$$\Rightarrow \mu_{C-AB} < \Delta_{C-AB} \rightarrow \text{有意差なし}$$

35

例題の結論

上記の結果から、尿中細菌数について急性膀胱炎と他の2群、すなわち急性腎盂腎炎と急性副睾丸炎の各群、および一緒にして平均値を比べた場合に有意差が認められることが判った。

36