

実 務 と 統 計 (8)

——一元配置分散分析——

山内俊平[†] (豊橋市立看護専門学校 非常勤講師,
豊橋市食肉衛生検査所 元所長)



分散分析 (ANOVA: Analysis of Variance) を厳密に理解するには、統計学的専門知識が必要となり、多くの実務者が苦戦するところかと思えます。ここでは、「実務で使える」ことを目的に、“直感で伝わる”よう、数式を最小限にして書いてみたいと思います。

1 一元配置分散分析の考え方 (直感的イメージ)

一つの要因について、主に“3 群以上の平均値に差があるのか”を分析する手法です (理論上は 2 群でも可能ですが、この場合は一般に t 検定を用います)。

〈F 値〉

F 値 (分散比) = 「群間の分散 (各群平均の全体平均からのばらつき)」 ÷ 「群内の分散 (同一群内データのその群の平均からのばらつき)」 (厳密には、自由度で調整した分散 (平均平方) の比) です。

*「群間の分散」は“要因 + 偶然”による「群間差」,
「群内の分散」は「個体差 (測定誤差も含む)」のイメージです。ざっくり言うと、F 値は“個体差に対して群間差がどれだけ大きいのか”を示す指標です。

F 値が大きいほど、この違いが偶然だけで生じるとは考えにくくなり、P 値は小さくなります。一元配置分散分析とは“群同士の違い (平均差) が個体差を超えて十分大きいのか”を、F 値 (分散比) を指標として仮説検定により明らかにする手法です。

2 一元配置分散分析 (one-way ANOVA)

例：ある薬剤を投与した 3 群 (A, B および C 群) の平均収縮期血圧 (mmHg) を比較する。

手順 1：サンプルサイズ (n) の設計

群 数 3 群 (n の割付：1:1:1)
効 果 量 $f=0.5$
有意水準 0.05
検 出 力 0.70

この条件で、専用ソフトや AI などを使って n を計算すると、各群 $n=11$ となります。

*効果量 f とは、群間差 (分散ベース) を群内のばらつきで標準化したもので、差の効果 (群の違いが、個体差に対してどの程度大きい) を表す指標です。

・効果の目安 (絶対基準ではない)

$f=0.10$ (小さな効果)

$f=0.25$ (中程度の効果)

$f=0.40$ (大きな効果)

本例では、演習用として少数例でも群間差が検出できるように効果量を大きめに設定 ($f=0.5$) しましたが、実際の研究だと $f=0.25$ (中程度) ぐらいが妥当かと思えます (この場合だと各群 $n=53$ 程度が必要)。

手順 2：データ収集

次の仮想データ (各群 $n=11$) を EZR に取り込みます (実務と統計 (4) 79 巻 2 号参照)。

血圧 mmHg	群
128	A
118	A
122	A
129	A
130	A
122	A
124	A
121	A
126	A
119	A
125	A
125	B
123	B
133	B
124	B
127	B
129	B
118	B
132	B
126	B
129	B
126	B
128	C
132	C
127	C
130	C
141	C
137	C
128	C
132	C
129	C
136	C
131	C

[†] 連絡責任者：山内俊平 (豊橋市立看護専門学校 非常勤講師, 豊橋市食肉衛生検査所 元所長)

手順3：前提条件の確認

分散分析は、独立性（群間データ同士が影響しあっていない）、正規性（各群の母集団が概ね正規分布（大きなズレはない））、等分散性（各群のばらつきに大きな相違はない）を前提（特に独立性は重要）とするパラメトリック検定です。

収集した各群データ（仮想）は独立で、QQプロットなどから分布に大きな歪みはなく、SD（A群 4.00, B群 4.23, C群 4.39）に顕著な違いもない。正規性や等分散性は概ね保たれており、一元配置分散分析の前提条件は満たされている、と判断できます。

手順4：多重比較法の選択

一元配置分散分析は群全体として差があるかの検定であり、どの2群間で有意なのかは判別できません。そこで各2群を比較しますが、この場合、単純なt検定では多重性の問題が生じるため、一般に多重比較法（Tukey法など）を用います。

〈多重性の問題（ α エラーの増大）〉

$\alpha_{\text{エラー}} = 0.05$ （有意水準5%）におけるt検定（2群間比較）では、「ほんとは差がないのに、差がある」と誤判定する確率5%です。有意水準5%で、3群の各2群間t検定（3回）を行うと、少なくとも1回は差があると誤判定する確率は5%を超えます。仮に10回の独立した検定なら、その確率はおおよそ40%（ $1 - 0.95^{10} \approx 0.4$ ）にもなります。このように複数の検定を繰り返すことで、統計的に有意の結果が得られる可能性が高まること（ α エラーの増大）を、多重性の問題といいます。

〈多重性の調整〉

多重性を軽減するために有意水準などを調整した検定を多重比較（検定）といい、Tukey法やDunnnett法などが用いられます。

Tukey法：すべてのペア比較（ nC_2 ：3群だと ${}_3C_2=3$ 、4群だと ${}_4C_2=6$ ）を同時に行う場合に適した方法

Dunnnett法：基準群を置いた組合せ比較（3群だと2、4群だと3）に適した方法

*本例では、Tukey法で2群間比較することとします。

手順5：分析

一元配置分散分析（主解析）

統計解析

→連続変数の解析

→3群以上の間の平均値の比較

（一元配置分散分析 one-way ANOVA）

帰無仮説：すべての群の平均は同じ
対立仮説：少なくとも1つの群の平均が異なる
帰無仮説棄却（慣例）： $P < 0.05$

3群以上の間の平均値の比較

（一元配置分散分析 one-way ANOVA）

モデル名を入力

ketuatuhikaku

目的変数
(1つ選択)

血圧 mmHg

比較する群

(1つ以上選択)

群

血圧 mmHg

グラフ

☒ 箱ひげ

☐ 棒

☐ 折れ線

等分散と考えますか？

☒ はい（一元配置分散分析）

☐ いいえ（Welch 検定）

☐ 2群ずつの比較（Bonferroni の多重比較）

☐ 2群ずつの比較（Holm の多重比較）

☒ 2群ずつの比較（Tukey の多重比較）

☐ 2群ずつの比較（Games-Powell の多重比較）

☐ 2群ずつの比較（Dunnnett の多重比較）

（基準を置いた比較）

例えば、A群を基準にA-BとA-C群を比較する場合（検定数2）には、Dunnnettを☒する。

分散がかなり違う、等分散性が疑わしいようなケースでは、Welch ANOVA（等分散と考えますか？ ☒ → いいえ（Welch 検定））を選択し、2群ずつの比較にGames-Howellの多重比較を☒します（EZR画面上ではGames-Powellの表記となっていますが、Games-Howellです）。

summary, anova

	平均	標準偏差	P 値
群=A	124.0000	4.000000	0.000438
群=B	126.5455	4.227615	
群=C	131.9091	4.392142	

× TukeyHSD (AnovaModel.2, "factor (群)")
Tukey multiple comparisons of means
95% family-wise confidence level

\$factor (群) \	diff	lwr	upr	p adj
B-A	2.545455	-1.8797167	6.970626	0.3445942
C-A	7.909091	3.4839197	12.334262	0.0003550
C-B	5.363636	0.9384651	9.788808	0.0149039

結果

(ANOVA)

$P = 0.000438$ ($P < 0.05$) より統計的に有意
(Tukey の多重比較)

B-A 群間： $P = 0.34459$ （非有意）

C-A 群間： $P = 0.00036$ （有意）

C-B 群間： $P = 0.01490$ （有意）

結 論

ANOVA より、少なくとも 1 つの群の平均が他と異なる。

Tukey より、C 群は、A および B 群と比較して平均収縮期血圧が有意に高い ($P < 0.05$)。一方、A と B の群間には有意差は認められない ($P \geq 0.05$)。

(注意点 1)

因果関係を分析したわけではありません。「薬剤投与により C 群の血圧が有意に上昇した」と紐づけして書くと、“書きすぎ”となるので注意してください。

(注意点 2)

Tukey 法では、複数グループ間の全てのペア比較を行っても「全体としての $\alpha_{\text{エラー}}$ 」が 5% 以内に収まるように設計されています。よって、EZR で出力された P 値を追補正する必要はなく、有意水準 5% で統計的結論 (有意・非有意) を導きます。

Kruskal-Wallis 検定 (感度分析)

Kruskal-Wallis 検定は、分布 (特に中央値) の違いを検出するノンパラメトリック検定です。結果の頑健性 (前提 (正規性・等分散性) に依存しない方法でも同様の結果が得られるのか) を確認してみます。

統計解析

→ ノンパラメトリック検定

→ 3 群以上の間の比較 (Kruskal-Wallis 検定)

帰無仮説: すべての群の分布に差はない
対立仮説: 少なくとも 1 つの群の分布が他と異なる
帰無仮説棄却 (慣例): $P < 0.05$

3 群以上の間の比較 (Kruskal Wallis 検定)

目的変数 (1 つ選択)

血圧 mmHg

グループ (1 つ選択)

群

血圧 mmHg

2 群ずつの比較 (post-hoc 検定)

☐ 2 群ずつの比較 (Bonferroni の多重比較)

☐ 2 群ずつの比較 (Holm の多重比較)

☒ 2 群ずつの比較 (post-hoc 検定, Steel-Dwass の多重比較)

☐ 2 群ずつの比較 (post-hoc 検定, Steel の多重比較)

(基準を置いた比較)

例えば、A 群を基準に A-B と A-C 群を比較する場合 (検定数 2) は、こちらを ☒

3 群以上の間の比較 (Kruskal-Wallis 検定) P 値 = 0.00156

> Steel.Dwass (Dataset\$ 血圧 mmHg, Dataset\$ 群)

	t	p
A:B	1.383661	0.349461839
A:C	3.323598	0.002558737
B:C	2.503056	0.032969230

結 果

(Kruskal-Wallis)

$P = 0.00156$ ($P < 0.05$) より統計的に有意

(Steel-Dwass の多重比較)

A-B 群間: $P = 0.34946$ (非有意)

A-C 群間: $P = 0.00256$ (有意)

B-C 群間: $P = 0.03297$ (有意)

結 論

Kruskal-Wallis より、少なくとも 1 つの群の分布に違いがあり、Steel-Dwass (すべてのペア比較) より、C 群が A および B 群と異なる。

結果の総合評価

「C 群は、A および B 群と比較して平均収縮期血圧が有意に高い」という主解析の結果は、感度分析 (ノンパラメトリック解析) でも支持されたことから、比較的頑健な結果と考えられます。

これまでも繰り返し述べてきましたが、「統計的に有意」であっても、その差が、実務的に意味があるとは限りません。推定される実際の差の大きさや効果量などを併せて実務的価値を総合判断することが重要です。“実務あつての統計”です。