

分散分析と F 検定

新 田 裕 史* 中 井 里 史** 高 木 廣 文***

はじめに

データが二群ある場合の比較の方法について前回解説した。今回は比較する群が多数ある場合の方法を解説することにする。

多群の比較の場合には、二群の比較の場合以上にさまざまな研究デザインが考えられ、また想定される仮説もいくつか考えられる。すべての場合についてここで述べることはできないので、基本的なものだけを示すことにする。

I. 一元配置の分散分析

例えば、血液中のある酵素濃度を測定する方法が4種類提案されている。5人の被験者から血液を採取し、4通りの方法で測定を行って、表1のようなデータを得た。

測定法毎に平均値を求めると、

A1	6.48
A2	6.20
A3	4.92
A4	5.32

となり、測定法A1が最も高い値を示し、測定法A3が最も低い値を示している。この成績から、測定法により差があるといえるかどうかが問題となる。

統計的検定の考え方に基づいた表現をすると、測定法により差がないと仮定したときに、このよ

表 1 血液中のある酵素濃度を測定する四方法

測定法	測 定 値				
A1	6.4	5.8	6.4	6.8	7.0
A2	6.2	5.2	6.8	7.0	5.8
A3	5.6	6.4	4.2	4.8	3.6
A4	5.2	5.0	6.4	6.4	3.6

うなデータが得られる確率はどの程度になるかを調べればよい。

一般に、 I 群のデータがそれぞれ n_i 個ずつ、全部で $N (= \sum_{i=1}^I n_i)$ 個の測定値が得られているとし、それを x_{ij} , ($i=1, \dots, I; j=1, \dots, n_i$) とする。ここでは N 個のデータが無作為に得られているとする。これは、一元配置の統計モデルと呼ばれ、次のように表される。

$$x_{ij} = \mu_i + \varepsilon_{ij}, \quad i=1, \dots, I; j=1, \dots, n_i$$

と書ける。上記の例でいえば、 I は 4, n_i は各群等しく 5 である。 x_{ij} がそれぞれの測定値に対応する。これは、各群の平均に誤差が加わったものが、測定値となっているというモデルである。 μ_i は i 番目の母集団の平均、 ε_{ij} は測定誤差である。いま、 ε_{ij} に不偏性、等分散性、無相関、および正規性を仮定する。すなわち、 ε_{ij} は互いに独立に正規分布 $N(0, \sigma^2)$ に従うと仮定する。

1. 一般的な検定

目的とすることは、母平均がすべて等しい " $\mu_1 = \mu_2 = \dots = \mu_I (= \mu)$ " という仮説から、データがどの程度ずれているかということを判断することである。以下ではすべてこのズレを平方和を

* 国立環境研究所

[〒305 つくば市小野川 16-2]

** 東京大学医学部疫学

*** 文部省統計数理研究所

表 2 分散分析表

要 因	平方和	自由度	分 散	F 値
処 理	T_A	$I-1$	V_A	V_A/V_e
誤 差	T_e	$N-I$	V_e	
全 体	T	$N-1$		

尺度として測る。必ずしも正確な定義ではないが簡単にいうと、ズレの方向を何も規定しない、つまり対立仮説が“少なくとも1組が $\mu_i \neq \mu_j$ ” の場合の一般的な検定が一元配置の分散分析の F 検定である。

言い換えれば、

$$x_{ij} = \mu_i + \varepsilon_{ij} \dots\dots\dots (1)$$

という統計モデル、すなわち群毎に平均値が異なるモデルと、

$$x_{ij} = \mu + \varepsilon_{ij} \dots\dots\dots (2)$$

という統計モデル(帰無仮説)、すなわち群を通して平均が一定というモデルのどちらが標本データと合致するかを調べることである。モデルと実際のデータとのズレを普通、差の平方和で測る。すなわち、 μ_i を $\bar{x}_i$ ($= \sum_{j=1}^{n_i} x_{ij}/n_i$) で推定すると、モデル (1) の下でのズレは、

$$T_e = \sum_{i=1}^I \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2$$

これに対して、モデル (2) すなわち 帰無仮説の下でのズレは、 μ を $\bar{x}_{..}$ ($= \sum_{i=1}^I \sum_{j=1}^{n_i} x_{ij}/N$) で推定すると、

$$T = \sum_{i=1}^I \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_{..})^2$$

である。したがって、モデル (1) とモデル (2) とのズレの大きさの差は、

$$\begin{aligned} T_A &= T - T_e \\ &= \sum_{i=1}^I \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_{..})^2 - \sum_{i=1}^I \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2 \\ &= \sum_{i=1}^I n_i (\bar{x}_i - \bar{x}_{..})^2 \end{aligned}$$

ここでつぎのような統計量を考える。

$$F_A = \frac{T_A/(I-1)}{T_e/(N-I)}$$

$T_e/(N-I)$ は誤差分散 σ^2 の不偏推定値となることがわかっている。したがって F_A はモデルか

表 3 分散分析表の実例

要 因	平方和	自由度	分 散	F 値
処 理	8.038	3	2.679	3.213
誤 差	13.344	16	0.834	
全 体	21.382	19		

らのズレの差が誤差に比べてどの程度の大きさになっているかを評価している。仮説の下で、 F_A は自由度 ($I-1$, $N-I$) の F 分布に従う。したがって、観測された F 値よりも大きくなる確率を計算できる。 F 値が大きければ大きいほど“各群の平均値に差がない”という仮説からズレていることになる。

普通、以上のことをまとめて表 2 のような分散分析表を作る。表中の分散は平均平方とも呼ばれ、平方和を自由度で割ったものである。

処理の平方和を級間変動、誤差の平方和を級内変動と呼ぶこともある。

先に示した例について、分散分析表を作成してみよう。実際にはもっと簡便に計算できる方法があるが、ここでは定義式に従って計算してみよう。

$$\begin{aligned} T_A &= 5 \times (6.48 - 5.73)^2 + 5 \times (6.20 - 5.73)^2 + \\ &\quad 5 \times (4.92 - 5.73)^2 + 5 \times (5.32 - 5.73)^2 \\ &= 8.038 \end{aligned}$$

$$\begin{aligned} T_e &= (6.4 - 6.48)^2 + (5.8 - 6.48)^2 + (6.4 - \\ &\quad 6.48)^2 + (6.8 - 6.48)^2 + (7.0 - 6.48)^2 + \\ &\quad (6.2 - 6.20)^2 + (5.2 - 6.20)^2 + \dots + \\ &\quad (6.4 - 5.32)^2 + (3.6 - 5.32)^2 \\ &= 13.344 \end{aligned}$$

$$V_A = 8.038/(4-1) = 2.679$$

$$V_e = 13.344/(20-4) = 0.834$$

$$\begin{aligned} F &= V_A/V_e \\ &= 3.213 \end{aligned}$$

これより分散分析表を作ると、表 3 のごとくとなる。適当な F 分布表から第 1 自由度が 3、第 2 自由度が 16 の 5% 点を求めると、3.239 である。この値よりも今回求めた F 値 3.213 は若干小さいので 5% 水準で有意ではないと判断される。ただし、この場合の有意確率すなわち p 値を求めてみると、 $p=0.051$ でぎりぎり 5% を越えている。

したがって、実際上は平均値に差があるかどうかの判断は保留すべきであろう。

以上の検定では、各群の分散が等しいと仮定している。二群の比較の場合と同様に、分散が等しいかどうかを検討するための方法が Bartlett の検定などいくつか知られている。詳しくは他の参考書を参照してほしい。ただし、各群の分散が極端に違っていなければ、2群の平均値の差の検定の場合のように区別してF検定を行う必要はない。

2. スチューデント化された範囲による方法

まず I 群のうち、最大の平均値 $\bar{x}_{max}$ と最小の平均値 $\bar{x}_{min}$ を用いて、平均値の範囲

$$R = \bar{x}_{max} - \bar{x}_{min}$$

を考える。各群のデータ数は等しく n とする。先の例では、 $\bar{x}_{max}=6.48$ (A1), $\bar{x}_{min}=4.92$ (A3) であるから、 $R=1.56$ となる。 σ^2 は先に示した分散分析表での誤差分散の推定値、すなわち、

$$\sigma^2 = T_e / (N - I)$$

である。

このとき、つぎの統計量、

$$\frac{R}{\sigma / \sqrt{n}}$$

をスチューデント化された範囲と呼んでいる。このスチューデント化された範囲の上側パーセント点は比較する群の数と誤差分散の自由度によって決まり、普通数表の形で与えられている(表4)。仮説からのズレが大きいほどこの統計量も大きい値をとる。

II. 同時信頼区間と多重比較

母平均がすべて等しいという仮説が棄却された場合には、実際どの群とどの群の差が統計的に有意か、また各群の平均値の差の大きさを推定することが次の仕事である。

多数の群がある時に、その中の二つの群の平均値の差の大きさを推定するために、前回学んだ t 分布に基づく2群の平均値の差の信頼区間を使うことも考えられる。その組合せ一つずつについてこの方法を適用すると、その全部が同時に真である確率はずっと小さくなる。信頼区間とは、真の母数(ここでは、母平均値の差)がその区間に

含まれる確率をあらかじめ定めた値に等しくなるように決めるものである。

例えば、いま5群のデータがあるとするすべての二つずつの平均値の差の組合せは10通りである。一つの組合せについて、信頼係数を95%にとるとする。それぞれの平均値の差に関する仮説が独立ならば、10通りの比較においてそれがすべて真である確率は、

$$0.95^{10} \approx 0.6$$

となり、95%よりずいぶん小さくなる。すべての組合せについて一定の信頼率を確保するためには信頼区間をより広くとる必要がある。このような信頼区間を特に同時信頼区間と呼んでいる。

二つの平均値の差の検定の立場から考えると、5%水準で検定を10通りも行えば一組ぐらい有意になる確率が5%を越えてしまうことになる。この点を考慮した検定方式を考える必要が出てくる。これは様々な比較の組合せから導かれることから、多重比較と呼ばれている。

1. Tukey 法

Tukey 法は群平均値の最大値と最小値の差に関する分布に基づいて信頼区間を構成するものである。すべての二つずつの平均値の差はこの中に含まれることになる。Tukey 法は基本的には各群のデータ数がすべて等しくなければならないが、等しくない場合にも Tukey-Kramer の近似によって適用できる。すべての平均値の差に対する信頼係数 $1-\alpha$ の同時信頼区間は次のように与えられる。

$$\bar{x}_i - \bar{x}_j \pm \sigma / \sqrt{n} \cdot q(I, N-I, \alpha)$$

また、

$$|\bar{x}_i - \bar{x}_j| / (\sigma / \sqrt{n})$$

が、 $q(I, N-I, \alpha)$ を越えたときに α 水準で i 群と j 群の平均値の差を有意と判断する。ただし、 $q(I, N-I, \alpha)$ はスチューデント化した範囲の上側 α パーセント点である。

先の例について、同時信頼区間を求めてみよう。 $q(4, 16, 0.05) = 4.046$ であるから、例えば、A1とA2の平均値の差の同時信頼区間は、

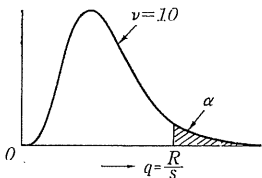
$$6.48 - 6.2 \pm \sqrt{0.834/5} \times 4.046$$

から、

$$-1.372 < \mu_1 - \mu_2 < 1.932$$

表 4 スチューデント化した範囲のパーセント点 (日本規格協会『統計数値表, 1972』より)

$q_{\alpha}(n, \nu) : \Pr[R/s \geq q_{\alpha}(n, \nu)] = \alpha$



$\alpha = 0.05$

$\nu \backslash n$	2	3	4	5	6	8	10	15	20	30
1	17.9693	26.9755	32.8187	37.0815	40.4076	45.3973	49.0710	55.3607	59.5576	65.1490
2	6.0849	8.3308	9.7980	10.8811	11.7343	13.0273	13.9885	15.6503	16.7688	18.2690
3	4.5007	5.9096	6.8245	7.5017	8.0371	8.8525	9.4620	10.5222	11.2400	12.2073
4	3.9265	5.0402	5.7571	6.2870	6.7064	7.3465	7.8263	8.6640	9.2334	10.0034
5	3.6354	4.6017	5.2183	5.6731	6.0329	6.5823	6.9947	7.7163	8.2080	8.8747
6	3.4605	4.3392	4.8956	5.3049	5.6284	6.1222	6.4931	7.1428	7.5864	8.1889
7	3.3441	4.1649	4.6813	5.0601	5.3591	5.8153	6.1579	6.7586	7.1691	7.7275
8	3.2612	4.0410	4.5288	4.8858	5.1672	5.5962	5.9183	6.4831	6.8694	7.3953
9	3.1992	3.9485	4.4149	4.7554	5.0235	5.4319	5.7384	6.2758	6.6435	7.1444
10	3.1511	3.8768	4.3266	4.6543	4.9120	5.3042	5.5984	6.1141	6.4670	6.9480
12	3.0813	3.7729	4.1987	4.5077	4.7502	5.1187	5.3946	5.8780	6.2089	6.6600
14	3.0332	3.7014	4.1105	4.4066	4.6385	4.9903	5.2534	5.7139	6.0290	6.4586
16	2.9980	3.6491	4.0461	4.3327	4.5568	4.8962	5.1498	5.5932	5.8963	6.3097
18	2.9712	3.6093	3.9970	4.2763	4.4944	4.8243	5.0705	5.5006	5.7944	6.1950
20	2.9500	3.5779	3.9583	4.2319	4.4452	4.7676	5.0079	5.4273	5.7136	6.1039
24	2.9188	3.5317	3.9013	4.1663	4.3727	4.6838	4.9152	5.3186	5.5936	5.9682
30	2.8882	3.4864	3.8454	4.1021	4.3015	4.6014	4.8241	5.2114	5.4750	5.8335
40	2.8582	3.4421	3.7907	4.0391	4.2316	4.5205	4.7345	5.1056	5.3575	5.6996
60	2.8288	3.3987	3.7371	3.9774	4.1632	4.4411	4.6463	5.0011	5.2412	5.5663
120	2.8000	3.3561	3.6846	3.9169	4.0960	4.3630	4.5595	4.8979	5.1259	5.4336
∞	2.7718	3.3145	3.6332	3.8577	4.0301	4.2863	4.4741	4.7959	5.0117	5.3013

$\alpha = 0.05$

$\nu \backslash n$	2	3	4	5	6	8	10	15	20	30
1	90.0242	135.0407	164.2577	185.5753	202.2097	227.1663	245.5416	277.0034	297.9972	325.9682
2	14.0358	19.0189	22.2937	24.7172	26.6290	29.5301	31.6894	35.4261	37.9435	41.3221
3	8.2603	10.6185	12.1695	13.3243	14.2407	15.6410	16.6908	18.5219	19.7648	21.4429
4	6.5112	8.1198	9.1729	9.9583	10.5832	11.5418	12.2637	13.5298	14.3939	15.5662
5	5.7023	6.9757	7.8042	8.4215	8.9131	9.6687	10.2393	11.2436	11.9318	12.8688
6	5.2431	6.3305	7.0333	7.5560	7.9723	8.6125	9.0966	9.9508	10.5378	11.3393
7	4.9490	5.9193	6.5424	7.0050	7.3730	7.9390	8.3674	9.1242	9.6454	10.3586
8	4.7452	5.6354	6.2038	6.6248	6.9594	7.4738	7.8632	8.5517	9.0265	9.6773
9	4.5960	5.4280	5.9567	6.3473	6.6574	7.1339	7.4945	8.1323	8.5726	9.1767
10	4.4820	5.2702	5.7686	6.1361	6.4275	6.8749	7.2133	7.8121	8.2256	8.7936
12	4.3198	5.0459	5.5016	5.8363	6.1011	6.5069	6.8136	7.3558	7.7305	8.2456
14	4.2099	4.8945	5.3215	5.6340	5.8808	6.2583	6.5432	7.0466	7.3943	7.8726
16	4.1306	4.7855	5.1919	5.4885	5.7223	6.0793	6.3483	6.8233	7.1512	7.6023
18	4.0707	4.7034	5.0942	5.3788	5.6028	5.9443	6.2013	6.6546	6.9673	7.3973
20	4.0239	4.6392	5.0180	5.2933	5.5095	5.8389	6.0865	6.5226	6.8232	7.2366
24	3.9555	4.5456	4.9068	5.1684	5.3735	5.6850	5.9187	6.3296	6.6123	7.0008
30	3.8891	4.4549	4.7992	5.0476	5.2418	5.5361	5.7563	6.1423	6.4074	6.7710
40	3.8247	4.3672	4.6951	4.9308	5.1145	5.3920	5.5989	5.9606	6.2083	6.5471
60	3.7622	4.2822	4.5944	4.8178	4.9913	5.2525	5.4466	5.7845	6.0149	6.3290
120	3.7016	4.1999	4.4970	4.7085	4.8722	5.1176	5.2992	5.6138	5.8272	6.1168
∞	3.6428	4.1203	4.4028	4.6028	4.7570	4.9872	5.1566	5.4485	5.6452	5.9106

大きな n の標本の範囲 R とこれと独立に得られた自由度 ν の分散 s^2 の平方根との比 $q = R/s$ のパーセント点を与えて
いる。1元配置のデータ (k 個の級、各級の繰返し数 r とする) で、 R を級平均の範囲、誤差分散を V_e とし $s = \sqrt{V_e/r}$
とすると、この表で $n = k$ 、 $\nu = k(r-1)$ として平均値の一様性の検定ができる。

例: $n = 8$ 、 $\nu = 24$ のとき $\Pr[R/s \geq 4.6838] = 0.05$ である。すなわち $q_{0.05}(8, 24) = 4.6838$ 。

〔補間法〕 n に関しては $1/n$ について補間するとよい。Neville の補間法 (解説 K5 参照) によれば 4 項とることによ
って 4 けた目で 2 以内の誤差の値を得る。 ν に関しても $1/\nu$ もしくは $120/\nu$ について補間するとよい。Neville
の補間法によれば 3 項で十分である。

と求められる。

2. その他の方法

多重比較の方法には Tukey 法のほかにも、Bonferroni の不等式による方法や Scheffé 法など多くの方法がある。また、各群の単純な比較ではなく、一つの対照群に対して他の群を比較するような場合もある。この場合には、Dunnett 法と呼ばれる方法を用いることが適当である。

種々の方法のうちどの方法を用いることが妥当かはデータの取り方、解析の目的によっても異なり、統計学の初心者のみならず専門家にとっても

重要な課題となっている。ここで示した方法は最も基本的なものであり、詳しくは成書を参照してほしい。

文 献

- 1) 広津千尋：統計的データ解析，日本規格協会，1983.
- 2) 吉村 功：毒性・薬効データの統計解析，サイエンス社，1987.
- 3) 高橋行雄，大橋靖雄，芳賀敏郎：SAS による実験データの解析，東京大学出版会，1989.

*

*

*