

分散分析の概説【第58回生物統計学】

1 一元配置分散分析の原理

三つ以上の群について母平均の比較を行う際、二群間の母平均の差の比較を繰り返してしまうと多重性の問題が生じてしまうことを第14回生物統計学「多重性の問題が生じる要因」において言及しました。これを回避するための手法の一つに分散分析があります。分散分析においては  $A_1, A_2, \dots, A_k$  の  $k$  個の各群の平均値  $\mu_1, \mu_2, \dots, \mu_k$  について各変動の比(分散比)を検定統計量としてF分布で帰無仮説  $H_0: \mu_1 = \mu_2 = \dots = \mu_k$  を受容するかどうか一度に検定します。

被験者を分類する基準として一つの因子だけを考慮するもっとも単純なケースを例に分析の流れを示します。

表1は、3つの群に割り付けられた各群10名ずつの試験参加者について、あるバイオマーカーを測定した実測値と、全体の平均値からの偏差を示したものです。個々の試験参加者の全体の平均値からの偏差は、群の違いによって生じる差(群間偏差)と、所属している群の平均値と参加者各自の間に生じる差(群内偏差)の二つに分けることができます。各試験参加者の群間偏差(=各群の平均値-全体の平均値)と群内偏差(=各試験参加者の実測値-各群の平均値)を表2に示しました。

全試験参加者の平均値との偏差の二乗の総和を「全変動」、群間偏差の二乗の総和を「群間変動」、群内偏差の二乗の総和を「群内変動」と呼び、全変動は群間変動と群内変動の和に等しいという性質があります。

表1. 各試験参加者の実測値と全体の平均値からの偏差

プラセボ群		低容量投与群		高容量投与群	
実測値	偏差	実測値	偏差	実測値	偏差
8.90	2.66	7.33	1.10	6.69	0.45
1.20	-5.04	3.69	-2.55	12.70	6.47
4.12	-2.12	5.60	-0.63	4.54	-1.70
5.23	-1.01	5.03	-1.21	6.64	0.40
0.50	-5.73	7.86	1.63	9.59	3.36
8.23	1.99	3.29	-2.95	5.34	-0.90
5.85	-0.39	7.67	1.43	7.31	1.07
6.02	-0.22	4.43	-1.81	9.04	2.80
0.22	-6.02	4.75	-1.49	11.16	4.92
4.83	-1.41	8.74	2.51	10.62	4.38



表 2. 各試験参加者の群間偏差と群内偏差

プラセボ群		低容量投与群		高容量投与群	
群間偏差	群内偏差	群間偏差	群内偏差	群間偏差	群内偏差
-1.73	4.39	-0.40	1.50	2.13	-1.67
	-3.31		-2.15		4.34
	-0.39		-0.24		-3.82
	0.72		-0.81		-1.73
	-4.01		2.02		1.23
	3.72		-2.55		-3.03
	1.34		1.83		-1.05
	1.51		-1.41		0.68
	-4.29		-1.09		2.80
	0.32		2.90		2.25

一元配置の分散分析においては、群内変動を試験参加者数-群数の自由度で割った値（=群以外が原因で生じた不偏分散）に対する、群間変動を群数-1の自由度で割った値（=群が原因で生じた不偏分散）の比（F値）を検定統計量とし、自由度が（群数-1，試験参加者数-群数）のF分布に従うことを以て、全ての群の母平均に差はないとする帰無仮説を採択します。言い換えると、一元配置分散分析とは群が原因で生じた分散が群と関係ない試験参加者間のばらつきと比べて十分に大きいと言えるかどうか検定していると言えます。例とした試験結果からは、全変動  $S_T \div 257.50$ 、群間変動  $S_B \div 76.58$ 、群内変動  $S_W \div 180.92$  が算出できます。従って検定統計量は  $S_B / (3-1) / S_W / (30-3) \div 76.6 * 27 / (180.9 * 2) \div 5.7141$  となり、自由度（2,27）のF分布の両側95%信頼区間[0.0253,4.2421]の上限を上回るため、帰無仮説は棄却され、三群全ての母平均が等しいわけではないことが明らかになります。この時注意したい点として、一元配置分散分析で導かれた有意差はあくまでも母平均が等しくないことを示しているだけであり、標本平均で高容量投与群や低容量投与群がプラセボ群を上回っているからといって、母平均もプラセボ群より高いという大小関係を示す根拠になるわけではないという点があります。

## 2 分散分析表の読み方

統計解析ソフトを用いて分散分析を行うと、過程で必要となったパラメータを示した分散分析表を出力してくれます。表3に示した通り、各変動と自由度から群間及び群内の不偏分散とその比、その場合の棄却域を算出し示すことができます。

## 3 参考文献

1. 石井俊全. 統計学大百科事典-仕事で使う公式・定理・ルール 113-. 株式会社翔泳社. 2020; 204-9



表 3. 分散分析表

	変動	自由度	分散 (変動/自由度)	F 値 (群間の分散/ 群内の分散)	上側 2.5%点
群間	76.58	2 (群数-1)	38.29	5.7141	4.2421
群内	180.92	27 (試験参加者数-群数)	6.701		
全	257.50	29 (試験参加者数-1)			