統計的仮説検定　その4

1　2元配置分散分析

前回までの分散分析では要因が一つでその中にいくつかの水準がある場合を扱った．しかし，現実の問題では要因が複数あることも多い．その場合にどのように分散分析を行うかについて，今回は紹介する．ただし，扱いが複雑になりすぎると理解も難しいので，要因が 2 つまでに限定する．

1.1　繰り返し

2 要因を扱う分散分析では**繰り返し**の有無とデータの対応でいくつかに場合分けされる．その中で繰り返しについて簡単に説明しておく．例えば，複数人のグループで 2 つの条件で何らかの実験を行うとするとき，各被験者は実験を 1 回ずつ行うとする．その場合には，表 1 のような実験データが集まってくる．

表 1: 繰り返しの無い 2 要因実験の例

		要因 B			
		実験 1	実験 2	$\cdots$	実験 q
要因 A	被験者 1	x_{11}	x_{12}	$\cdots$	x_{1q}
	被験者 2	x_{21}	x_{22}	$\cdots$	x_{2q}
	$\cdots$	$\cdots$			
	被験者 p	x_{p1}	x_{p2}	$\cdots$	x_{pq}

このような場合には，分散分析により調べたいのは被験者ごとに異なっているのか，実験条件によっても異なってくるのか，ということである．このような分散分析もあるのではあるが，今回は省略し，繰り返しのある場合についてのみ扱うことにする．すなわち，表 2 のような場合である．

表 2: 繰り返しのある 2 要因実験の例

		要因 B			
		実験 1	実験 2	$\cdots$	実験 q
要因 A	被験者 1	x_{111} $\cdots$ x_{11r}	x_{121} $\cdots$ x_{12r}	$\cdots$ $\cdots$ $\cdots$	x_{1q1} $\cdots$ x_{1qr}
	被験者 2	x_{211} $\cdots$ x_{21r}	x_{221} $\cdots$ x_{22r}	$\cdots$ $\cdots$ $\cdots$	x_{2q1} $\cdots$ x_{2qr}
	$\cdots$	$\cdots$			
	被験者 p	x_{p11} $\cdots$ x_{p1r}	x_{p21} $\cdots$ x_{p2r}	$\cdots$ $\cdots$ $\cdots$	x_{pq1} $\cdots$ x_{pqr}

1.2　変動の求め方

表 2 を見るとわかるようにデータの構造が 3 次元となっており，かなり複雑な状態である．各種変動の求め方にはよく注意してほしい．また，今回は繰り返しの回数は簡単化のために r 回ですべての組に共通としている．

まず，表 2 で被験者ごとに横にすべての x_{ijk} を足した和を $T_{i\cdot\cdot}$ と表す．添え字 i は被験者 1〜p までである．一方，縦に実験ごとに加えていく和を $T_{\cdot j\cdot}$ とする．添え字 j は実験 1〜q までである．また，各ブロックのデータの和を $T_{ij\cdot}$ と表す．各被験者が各実験で行った結果の和である．データの総個数 $N = pqr$ と，

$$T_{\cdots} = \sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{k=1}^{r} x_{ijk} \tag{1}$$

を用いて，全変動 S_T，行間変動（A の平方和）S_A，列間変動（B の平方和）S_B，群間変動（AB の平方和）S_{AB} をまず表すと以下のようになる．

$$\begin{aligned}
S_T &= \sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{k=1}^{r} x_{ijk}^2 - \frac{T_{\cdots}{}^2}{N} && (2)\\
S_A &= \sum_{i=1}^{p} \frac{T_{i\cdot\cdot}{}^2}{qr} - \frac{T_{\cdots}{}^2}{N} && (3)\\
S_B &= \sum_{j=1}^{q} \frac{T_{\cdot j\cdot}{}^2}{pr} - \frac{T_{\cdots}{}^2}{N} && (4)\\
S_{AB} &= \sum_{i=1}^{p}\sum_{j=1}^{q} \frac{T_{ij\cdot}{}^2}{r} - \frac{T_{\cdots}{}^2}{N} && (5)
\end{aligned}$$

2 元配置分散分析に特徴的な量は以下に示す交互作用変動（A×B の平方和）$S_{A\times B}$ と誤差変動 S_E である．

$$\begin{aligned}
S_{A\times B} &= S_{AB} - S_A - S_B && (6)\\
S_E &= S_T - S_{AB} && (7)
\end{aligned}$$

自由度も以下のように求められる．

$$\begin{aligned}
\phi_T &= N - 1 = pqr - 1 && (8)\\
\phi_A &= p - 1 && (9)\\
\phi_B &= q - 1 && (10)\\
\phi_{AB} &= pq - 1 && (11)\\
\phi_{A\times B} &= \phi_{AB} - \phi_A - \phi_B && (12)\\
\phi_E &= \phi_T - \phi_{AB} = pq(r-1) && (13)
\end{aligned}$$

これらの値からそれぞれの不偏分散は以下のように求められる．

$$\begin{aligned}
\sigma_A^2 &= \frac{S_A}{\phi_A} && (14)\\
\sigma_B^2 &= \frac{S_B}{\phi_B} && (15)\\
\sigma_{A\times B}^2 &= \frac{S_{A\times B}}{\phi_{A\times B}} && (16)\\
\sigma_E^2 &= \frac{S_E}{\phi_E} && (17)
\end{aligned}$$

誤差変動に対する各変動を F 値を計算することで議論するのがここからの作業となり，結局表 3 のような分散分析表を作成することになる．

4 つの不偏分散比について，それぞれ F 分布表を参照して有意性を議論するのではあるが，2 元配置分散分析の特徴は先ほども述べた交互作用である．考えている 2 つの要因が重なって平均値に大きく影響を与えることを意味しているが，これについては具体的な例が無いとわかりづらいであろう．また，要因 A と B の単独の効果を主効果と呼ぶ．ただし，交互作用が有意にあるときには主効果の検討は注意を要する．

表 3: 2 元配置の分散分析表

要因	変動	自由度	不変分散	不偏分散比
Ai 間変動	S_A	$p-1$	$\sigma_A^2 = \frac{S_A}{p-1}$	$\frac{\sigma_A^2}{\sigma_E^2}$
Bi 間変動	S_B	$q-1$	$\sigma_B^2 = \frac{S_E}{q-1}$	$\frac{\sigma_B^2}{\sigma_E^2}$
交互作用	$S_{A\times B}$	$(p-1)(q-1)$	$\sigma_{A\times B}^2 = \frac{S_{A\times B}}{(p-1)(q-1)}$	$\frac{\sigma_{A\times B}^2}{\sigma_E^2}$
水準内変動	S_E	$pq(r-1)$	$\sigma_E^2 = \frac{S_E}{(pq(r-1)}$	

1.3　交互作用の意味

今，表 4 のようなデータが得られたとする．このとき，どのパラメータを横軸にグラフを描くかで見た目は異なるが，例えば，図 1(a) のように描いたとする．このとき，１つめの条件が A1 から A2 になると値は B1 条件と B2 条件の両方で増加しており，条件 A も条件 B もともに結果に影響を与えているように見える．このような 2 本の線がほぼ平行に見える単純な違いの場合は主効果がメインであり，交互作用が無いといえる．ところが，図 1(b) や (c) のようなグラフの場合，条件の組み合わせによって結果が複雑に変化しており，このような場合に交互作用が介在している．(b) の場合は相殺効果，(c) は相乗効果ということがある．

表 4: 簡単のデータの例

条件	B1	B2
A1	4	7
A2	6	9

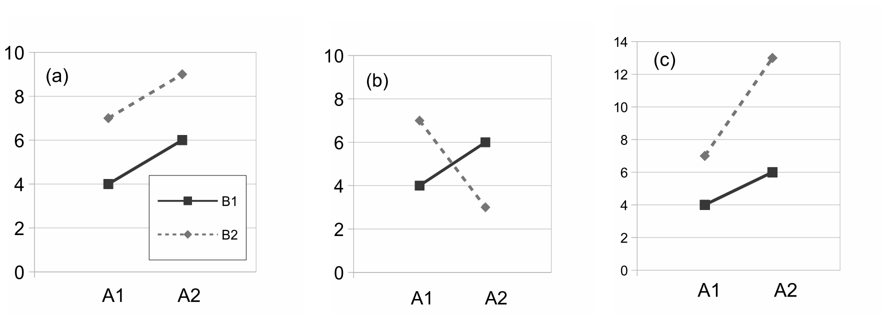

図 1: 2 要因実験のグラフの例

1.4　2 元配置分散分析の例

演習としてある実験の例を通して実習を行ってみる．資料も配布してあるので，そちらも参考にしてほしい．まず，実験であるが，ある化合物を合成し，その後熱処理を行うことで特性（透過率）を変える，ということを行うとする．熱処理温度は 100 ℃，200 ℃，および 300 ℃とし，熱処理時間は 1 時間，2 時間，4 時間，そして 8 時間の 4 通りとする．そして，表 5 のような結果が得られた．測定結果は各条件で 3 回ずつ合成した化合物の透過率測定の例とする．また，熱処理の目的は光の透過性を上げることであったとする．このとき，各実験条件の平均値をプロットすると，図 2 のようになる．このとき，これらの実験条件について 2 元配置分散分析を行ってみよう．

式 (1)〜(17) までを使って分散分析表を作成すると，表 6 のようになる．実際に F 分布表より F の値を求めて，このデータについて分析を行ってみよう．なお，2 元配置分散分析においても多重比較は行えるが，処理が煩雑なので今回は省略する．有意差が条件に対して存在するかの議論を行ってみよう．

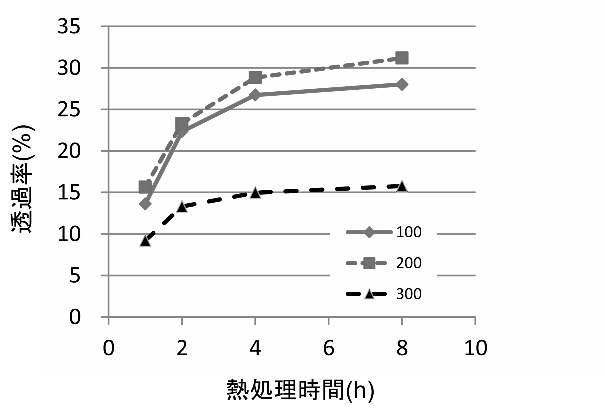

図 2: 熱処理条件による透過率の変化

表 5: 熱処理条件による透過率の変化のデータ

	熱処理温度 (°C)		
熱処理時間 (h)	100	200	300
1	13.2	16.1	9.1
	15.7	15.7	10.3
	11.9	15.1	8.2
2	22.8	24.5	11.9
	25.7	21.2	14.3
	18.5	24.2	13.7
4	21.8	26.9	15.1
	26.3	31.3	13.6
	32.1	28.3	16.2
8	25.7	30.1	15.2
	28.8	33.8	17.3
	29.5	29.6	14.8

表 6: 熱処理実験例の分散分析表

要因	変動	自由度	不変分散	不偏分散比
熱処理時間変動	798.2	3	266.1	47.17
熱処理温度変動	889.5	2	444.8	78.85
交互作用	89.89	6	14.98	2.656
水準内変動	135.4	24	5.640	