5.　分散分析

概要:　我々はしばしば目的変数を様々な方法で向上させようと試みる。例えば農業収穫量を向上させるために様々な肥料を試す。我々は収穫量から肥料の有効性を評価する。ある肥料の場合、平均値は他の肥料を使ったものに比べて向上しても、その中には他の肥料を使ったものより収穫量の低いものも含まれる。この平均値の向上とそのグループ内のバラツキから、その肥料が有効であるかどうかを定量的に評価するのが分散分析である。

キーワード s:分散分析；相関比；一元配置；二元配置。

5.1.　序

農業においては、我々は穀物の収量を多くしたい。したがって、それに対する様々な取り組みの効果を評価したい。一つの例として肥料がある。このように収穫に寄与すると思われるものを因子と呼ぶ。ここでは、因子として肥料のみを考えているから、1 因子となる。肥料 A にはいろいろの種類のもの A_j があるとする。我々は様々な肥料の効果を知りたい。いろいろな肥料のことをレベルとよぶ。各肥料、すなわち各レベルにおける平均値は一般に肥料によって異なる。しかし、肥料によっては、平均値は多くても、個々のデータの中には値が逆転しているものもある。この内部のバラツキと平均値の違いを比較してある肥料が有効なのかを定量的に判断したい。

5.2.　相関比（一元配置の分散分析）

ここではまず 1 因子のみを考える。このように 1 因子のみを考える場合を一元配置と呼ぶ。

三つの種類の肥料 A_1，A_2 および A_3 の収穫量に対する効果を検討する。我々は 1kg の種をまき、それに対する収穫量を評価した。その結果を表 1 に示す。

表 1　収穫量に対する各肥料の効果

	A:Fertilizer		
	A_1	A_2	A_3
	5.44	7.33	9.44
	7.66	6.32	7.33
	7.55	6.76	7.66
	4.33	5.33	6.43
	5.33	5.89	7.45
	4.55	6.33	6.86
	6.33	5.44	9.78
	5.98	8.54	6.34
Average	5.90	6.49	7.66

それぞれの肥料の場合のデータ数は以下である。

$$n_{A_1} = 8, n_{A_2} = 8, n_{A_3} = 8 \tag{1}$$

したがって、トータルのデータ数 n は

$$n = n_{A_1} + n_{A_2} + n_{A_3} = 24 \tag{2}$$

となる。

対応する各水準の平均値は以下である。

$$\mu_{A_1} = \frac{1}{n_{A_1}} \sum_{i=1}^{n_{A_1}} x_{iA_1} = 5.90 \tag{3}$$

$$\mu_{A_2} = \frac{1}{n_{A_2}} \sum_{i=1}^{n_{A_2}} x_{iA_2} = 6.49 \tag{4}$$

$$\mu_{A_3} = \frac{1}{n_{A_3}} \sum_{i=1}^{n_{A_3}} x_{iA_3} = 7.66 \tag{5}$$

全体の平均値は以下である。

$$\mu = \frac{1}{n_{A_1} + n_{A_2} + n_{A_3}} \left(\sum_{i=1}^{n_{A_1}} x_{iA_1} + \sum_{i=1}^{n_{A_2}} x_{iA_2} + \sum_{i=1}^{n_{A_3}} x_{iA_3} \right) = 6.68 \tag{6}$$

各肥料の平均値は、A_1 で 5.90 kg、A_2 で 6.49 kg 、A_3 で 7.66 kg である。したがって、平均値から判断すれば、肥料 A_3 が最も有効で、肥料 A_1 は有効でないと思える。しかしながら、肥料 A_1 を使って収量が 7.55 のものがあり、一方、肥料 A_3 を使って収量が 6.34 のものがある。つまり、個々のデータを見ると肥料 A_1 と使った収穫量のほうが肥料 A_3 を使った収穫量よりも多いものも存在する。この状態で統計的にどのように定量的に判断できるのであろうか？

そのためには、肥料の効果とその肥料を使った場合の収量のバラツキを評価して、肥料の効果のほうが大きいのかを評価する必要がある。

ある水準のデータは以下のように評価できる。

$$x_{ij} - \mu = (\mu_i - \mu) + (x_{ij} - \mu_i) \tag{7}$$

つまり、各データの偏差はその水準の平均からの全体平均からの偏差とそのデータの水準平均からの偏差の和となる。右辺の第項目は肥料の効果、第 2 項目はその肥料内でのバラツキと考えることができる。

肥料の種類に関連する分散をS_{ex}^2とすると、それは各水準平均と全体の平均の差の 2 乗と考えられるから、以下で与えられる。

$$S_{ex}^2 = \frac{n_{A_1}\left(\mu_{A_1} - \mu\right)^2 + n_{A_2}\left(\mu_{A_2} - \mu\right)^2 + n_{A_3}\left(\mu_{A_3} - \mu\right)^2}{n_{A_1} + n_{A_2} + n_{A_3}} = 0.53 \tag{8}$$

それぞれの肥料を利用した場合の、その内部での分散の和を S_{in}^2とすると、それは以下で与えられる。

$$S_{in}^2 = \frac{\sum_{i=1}^{n_{A_1}}\left(x_{iA_1} - \mu_{A_1}\right)^2 + \sum_{i=1}^{n_{A_2}}\left(x_{iA_2} - \mu_{A_2}\right)^2 + \sum_{i=1}^{n_{A_3}}\left(x_{iA_3} - \mu_{A_3}\right)^2}{n_{A_1} + n_{A_2} + n_{A_3}} = 1.27 \tag{9}$$

この二つの分散の比η^2 を以下のように定義する。

$$\eta^2 = \frac{S_{ex}^2}{S_{in}^2 + S_{ex}^2} = 0.30 \tag{10}$$

もしも、肥料の依存性が大きければ、このη^2は 1 に近づき、小さければ 0 に近づく。

ここで、η^2についてさらに議論する。

トータルの分散S_{tot}^2は以下のように定義される。

$$S_{tot}^2 = \frac{\sum_{i=1}^{n_{A_1}}\left(x_{iA_1} - \mu\right)^2 + \sum_{i=1}^{n_{A_2}}\left(x_{iA_2} - \mu\right)^2 + \sum_{i=1}^{n_{A_3}}\left(x_{iA_3} - \mu\right)^2}{n} \tag{11}$$

この Eq. (11)の分子の最初の項は以下のように変形される。

$$\begin{aligned}
\sum_{i=1}^{n_{A}}\left(x_{iA_1} - \mu\right)^2 &= \sum_{i=1}^{n_{A_1}}\left(x_{iA_1} - \mu_{A_1} + \mu_{A_1} - \mu\right)^2 \\
&= \sum_{i=1}^{n_{A_1}}\left(x_{iA_1} - \mu_{A_1}\right)^2 + 2\sum_{i=1}^{n_{A_1}}\left(x_{iA_1} - \mu_{A_1}\right)\left(\mu_{A_1} - \mu\right) + \sum_{i=1}^{n_{A_1}}\left(\mu_{A_1} - \mu\right)^2 \\
&= \sum_{i=1}^{n_{A_1}}\left(x_{iA_1} - \mu_{A_1}\right)^2 + n_{A_1}\left(\mu_{A_1} - \mu\right)^2
\end{aligned} \tag{12}$$

同じような解析を他の項にも適用できて、以下を得る。

$$
\begin{aligned}
S_{tot}^2 &= \frac{\sum_{i=1}^{n_{A_1}}\left(x_{iA_1}-\mu_{A_1}\right)^2+\sum_{i=1}^{n_{A_2}}\left(x_{iA_2}-\mu_{A_2}\right)^2+\sum_{i=1}^{n_{A_3}}\left(x_{iA_3}-\mu_{A_3}\right)^2}{n} \\
&+\frac{n_{A_1}\left(\mu_{A_1}-\mu\right)^2+n_{A_2}\left(\mu_{A_2}-\mu\right)^2+n_{A_3}\left(\mu_{A_3}-\mu\right)^2}{n} \\
&= S_{ex}^2+S_{in}^2
\end{aligned}
\tag{13}
$$

したがって、η^2 はトータルの分散を用いて以下のようになる。

$$
\eta^2=\frac{S_{ex}^2}{S_{in}^2+S_{ex}^2}=\frac{S_{ex}^2}{S_{tot}^2} \tag{14}
$$

この η^2 のことを相関比と呼ぶ。しかしながら、我々はこの相関比を最終的に判断する比としては用いない。そのことは次の節で議論する。

5.3. 依存性の判定

S_{ex}^2 の自由度を考える。それを ϕ_{ex} と置く。この場合の変数は $\mu_{A_1},\mu_{A_2},\mu_{A_3}$ である。しかしながら、それらは Eq. (6)を満足しなければならない。したがって、自由度は変数の数から 1 を引いたものになる。つまり、ϕ_{ex} は以下となる。

$$
\phi_{ex}=p-1=3-1=2 \tag{15}
$$

ここで、p は肥料の種類の数であり、ここでは 3 である。したがって、この自由度を考慮した分散を不偏分散を s_{ex}^2 とすると、それは以下となる。

$$
s_{ex}^2=\frac{n}{\phi_{ex}}S_{ex}^2=6.45 \tag{16}
$$

S_{in}^2 の自由度を考える。それを ϕ_{in} と置く。それぞれのグループのデータ数は n_{A_1},n_{A_2},n_{A_3} である。そして、それぞれのグループには平均がある。したがって、その自由度は各データにおいて、それぞれのデータ数から 1 を引いたものになる。すなわち、以下となる。

$$
\begin{aligned}
\phi_{in} &= \left(n_A-1\right)+\left(n_B-1\right)+\left(n_C-1\right) \\
&= n-p \\
&= 24-3 \\
&= 21
\end{aligned}
\tag{17}
$$

したがって、対応する不偏分散 s_{in}^2 は以下となる。

$$
s_{in}^2=\frac{n}{\phi_{in}}S_{in}^2=1.44 \tag{18}
$$

最終的にはこれらの不偏分散の比をFとすると、それは以下となる。

$$F = \frac{s_{ex}^2}{s_{in}^2} = 4.46 \tag{19}$$

これは自由度 $\left(\phi_{ex}, \phi_{in}\right)$ のF分布になる。F分布のP点を推定確率 95% で評価したものを

$F_C\left(\phi_{ex}, \phi_{in}, 0.95\right)$ と置くと、それは以下になる。

$$F_C\left(\phi_{ex}, \phi_{in}, 0.95\right) = 3.47 \tag{20}$$

ここで、$F > F_C$ であるから、この依存性は有ると判断される。

以上の描像を図 1 に示す。

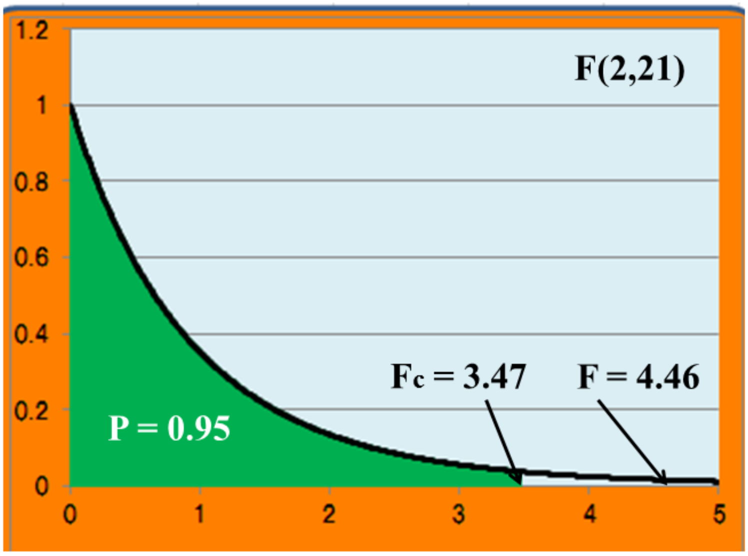

図 1　F 分布による判定

5.4.　繰返しのない二元配置の分散分析

これまでは、収穫量に対して肥料の効果があるかどうかをみてきた。つまり、収穫に関しては肥料のみ、つまり 1 因子で考えてきた。

ここでは二つの因子Aおよび因子Bを考える。つまり、前の例でいえば、収穫量に寄与する因子は肥料の 1 種類ではなく、その他にもう一つあり、全部で 2 種類あると考える。例えば、Aはこれまでの解析で扱ってきた肥料であり、Bは温度に相当する。 A としては A_1, A_2, A_3, A_4, を考え、その数を n_A とする。 Bとしては B_1, B_2, B_3 とし、データ数としてはn_Bとする。つまり、ここでは

$n_A = 4$ および $n_B = 3$ である。ここでは、一つの水準 $B_i A_j$ に対しては一つのデータがあるとする。それを**エラー！ 参照元が見つかりません。**に示す。総データ数 n は以下で与えられる。

$$n = n_A \times n_B \tag{21}$$

ここで水準 $A_j B_i$ に対応するデータを x_{ij} とする。x_{ij} を以下のように表現する。

$$x_{ij} - \mu = \left(\mu_{A_j} - \mu\right) + \left(\mu_{B_i} - \mu\right) + e_{ij} \tag{22}$$

つまり、x_{ij} が全体の平均と同じであれば、e_{ij} は 0 になる。つまり、e_{ij} の絶対値が大きければ大きいほど依存性があるということになる。

一元配置の場合は以下であった。

$$x_{ij} - \mu = \left(\mu_i - \mu\right) + \left(x_{ij} - \mu_i\right) \tag{23}$$

この場合は、i に関するデータが複数あったが、繰り返しのない二元配置では、その値は 1 個しかないから、上のような表記になる。

2 元配置に対応する自由度を考える。その具体的なデータを表 2 に示す。自由度に関する方程式は以下となる。

$$\phi_{tot} = \phi_A + \phi_B + \phi_e \tag{24}$$

ここで、トータル、因子 A 、 B に対する自由度は以下となる。

$$\begin{aligned} \phi_{tot} &= n - 1 \\ &= 12 - 1 \\ &= 11 \end{aligned} \tag{25}$$

$$\begin{aligned} \phi_A &= n_A - 1 \\ &= 4 - 1 \\ &= 3 \end{aligned} \tag{26}$$

$$\begin{aligned} \phi_B &= n_B - 1 \\ &= 3 - 1 \\ &= 2 \end{aligned} \tag{27}$$

したがって、誤差に関する自由度は以下となる。

$$\begin{aligned} \phi_e &= \phi_{tot} - \left(\phi_A + \phi_B\right) \\ &= \left(n - 1\right) - \left(n_A - 1 + n_B - 1\right) \\ &= n - \left(n_A + n_B\right) + 1 \\ &= 12 - \left(4 + 3\right) + 1 \\ &= 6 \end{aligned} \tag{28}$$

表 2　繰返しのない 2 因子の場合のデータ

		A					
		A_1	A_2	A_3	A_4	μ_{Bi}	$D\mu_{Bi}$
B	B_1	7.12	8.33	9.12	11.21	8.95	-1.20
	B_2	8.63	8.15	10.98	15.43	10.80	0.65
	B_3	7.99	8.55	11.79	14.45	10.70	0.55
	μ_{Aj}	7.91	8.34	10.63	13.70		
	$D\mu_{Aj}$	-2.23	-1.80	0.48	3.55		

Total average	10.15

議論は 1 元配置の分散分析と同様である。

トータルの平均 μ を求める。次に、因子 A の各水準に対する平均 μ_{A_j} を求める。ここで、$j=1,2,3,4$ である。最後の因子 B の各水準に対する平均 μ_{B_i} を求める。ここで $i=1,2,3$ である。これらは**エラー! 参照元が見つかりません。**に示してある。

まず因子 A の効果を検討する。ここで、因子 A には 4 水準ある。関連する分散は以下である。

$$\begin{aligned}S_A^2 &= \frac{n_B \sum_{j=1}^{n_A} \left(\mu_{A_j} - \mu\right)^2}{n} \\ &= \frac{1}{n_A} \sum_{j=1}^{n_A} \left(\mu_{A_j} - \mu\right)^2 \\ &= \frac{1}{4}\left[\left(-2.23\right)^2 + \left(-1.80\right)^2 + \left(0.48\right)^2 + \left(3.55\right)^2\right] \\ &= 5.26\end{aligned} \tag{29}$$

これに対する自由度は 4 -1 =3 であるから、対応する不偏分散は以下となる。

$$\begin{aligned}s_A^2 &= \frac{n}{\phi_A} S_A^2 \\ &= \frac{12}{3} \times 5.26 \\ &= 21.08\end{aligned} \tag{30}$$

次に、因子 B に関する分散を評価する。ここで、因子 B には 3 水準ある。関連する分散は以下である。

$$
\begin{aligned}
S_B^2 &= \frac{n_A \sum_{i=1}^{n_{B_i}} \left(\mu_{B_i} - \mu\right)^2}{n} \\
&= \frac{1}{n_B} \sum_{i=1}^{n_{B_i}} \left(\mu_{B_i} - \mu\right)^2 \\
&= \frac{1}{3}\left[\left(-1.20\right)^2 + \left(0.65\right)^2 + \left(0.55\right)^2\right] \\
&= 0.72
\end{aligned}
\tag{31}
$$

これに対する自由度は 3 -1 =2 であるから、対応する不偏分散 s_B^2 は以下となる。

$$
\begin{aligned}
s_B^2 &= \frac{n}{\phi_B} S_B^2 \\
&= \frac{12}{2} \times 0.72 \\
&= 4.34
\end{aligned}
\tag{32}
$$

誤差は以下で評価できる。

$$
e_{ij} = x_{ij} - \mu - \left[\left(\mu_{A_j} - \mu\right) + \left(\mu_{B_i} - \mu\right)\right]
\tag{33}
$$

これを表 3 に示す。

表 3　誤差のデータ

		A			
		A_1	A_2	A_3	A_4
B	B_1	0.41	1.19	-0.31	-1.29
	B_2	0.06	-0.85	-0.30	1.08
	B_3	-0.47	-0.34	0.61	0.20

誤差に対応する分散 S_e^2 は以下で与えられる。

$$
\begin{aligned}
S_e^2 &= \frac{1}{n} \sum_{i=1}^{3} \sum_{j=1}^{4} e_{ij}^2 \\
&= 0.50
\end{aligned}
\tag{34}
$$

したがって、誤差に関する不偏分散は以下で与えられる。

$$
\begin{aligned}
s_e^2 &= \frac{n}{\phi_e} S_e^2 \\
&= \frac{12}{6} \times 0.50 \\
&= 1.01
\end{aligned}
\tag{35}
$$

F 因子 A に関する F 値は以下となる。

$$F_A = \frac{s_A^2}{s_e^2} = \frac{21.08}{1.01} = 20.87 \tag{36}$$

因子 A に関する臨界 F 値を F_{Ac} とすると、それは以下で与えられる。

$$F_{Ac}\left(3,6,0.95\right) = 4.76 \tag{37}$$

したがって、F_A は F_{Ac} より大きく、したがって、因子 A には依存すると言える。

因子 B に関する F 値は以下で与えられる。

$$F_B = \frac{s_B^2}{s_e^2} = \frac{4.34}{1.01} = 4.29 \tag{38}$$

因子 B に関する臨界 F 値を F_{Bc} とすると、それは以下で与えられる。

$$F_{Bc}\left(2,6,0.95\right) = 5.14 \tag{39}$$

したがって、F_B は F_{Bc} より小さく、したがって、因子 B には依存しているとは言えない。

5.5. 繰返しのある二元配置の分散分析

前節と同じように二つの因子 A と B を考える。因子 A は 4 つのレベル A_1, A_2, A_3, A_4 があるとし、それを n_A とする。因子 B には 3 つのレベル B_1, B_2, B_3 あるとし、それを n_B とする。つまり、ここでは $n_A = 4$ 、$n_B = 3$ とする。前の節では、一つの水準 $A_j B_i$ には一つのデータしかないと仮定していた。しかし、同じ水準に複数のデータが存在するとする。ここでは、二つのデータセットがあるとする。それを s で表す。つまり、データは x_{ij} ではなく、x_{ij_s} と表す。このようなデータを表 4 に示す。

水準 $A_j B_i$ には $s = 1,2$ 二つのデータが存在する。

総データ数 n は以下で与えられる。

$$\begin{aligned} n &= n_A \times n_B \times n_s \\ &= 3 \times 4 \times 2 \\ &= 24 \end{aligned} \tag{40}$$

トータルの平均 μ は以下で与えられる。

$$\mu = \frac{1}{n}\sum_{s=1}^{n_s}\sum_{j=1}^{n_A}\sum_{i=1}^{n_B} x_{ij\_s} = 10.07 \tag{41}$$

それぞれのデータの総平均からのずれ Δx_{ij_s} は以下で評価される。

$$\Delta x_{ij\_s} = x_{ij\_s} - \mu \tag{42}$$

この結果を表 5 に示す。

表 4　2 因子で 2 セットのデータがある場合。 データ 1 データ 2 とする

Data 1		A			
		A_1	A_2	A_3	A_4
B	B_1	7.55	8.98	8.55	11.02
	B_2	8.12	6.96	11.89	11.89
	B_3	8.97	10.95	12.78	16.55

Total mean	10.07

Data 2		A			
		A_1	A_2	A_3	A_4
B	B_1	6.11	8.78	8.44	9.89
	B_2	9.01	9.65	9.86	13.22
	B_3	8.79	9.21	12.01	12.47

表 5　トータルの平均からの偏差

Data 1 deviation		A			
		A_1	A_2	A_3	A_4
B	B_1	-2.52	-1.09	-1.52	0.95
	B_2	-1.95	-3.11	1.82	1.82
	B_3	-1.10	0.88	2.71	6.48

Data 2 deviation		A			
		A_1	A_2	A_3	A_4
B	B_1	-3.96	-1.29	-1.63	-0.18
	B_2	-1.06	-0.42	-0.21	3.15
	B_3	-1.28	-0.86	1.94	2.40

二つの表からそれぞれの水準の平均を以下のように評価する。

$$\bar{x}_{ij} = \frac{1}{2}\left(x_{ij\_1} + x_{ij\_2}\right) \tag{43}$$

この結果を表 6 に示す。各水準には 2 個のデータがあるから、このデータを 2 セット得たとみなす。

表 6　データ 1 の 2 の平均とトータルの平均からの偏差

Average of data 1,2		A					
		A_1	A_2	A_3	A_4	μ_{Bi}	$\Delta\mu_{Bi}$
B	B_1	6.83	8.88	8.50	10.46	8.67	-1.40
	B_2	8.57	8.31	10.88	12.56	10.08	0.01
	B_3	8.88	10.08	12.40	14.51	11.47	1.40
	μ_{Aj}	8.09	9.09	10.59	12.51		
	$D\mu_{Aj}$	-1.98	-0.98	0.52	2.44		

各因子の平均値は以下となる。

$$\mu_{Aj} = \frac{1}{n_B}\sum_{i=1}^{n_B}\bar{x}_{ij} \qquad (44)$$

$$\mu_{Bi} = \frac{1}{n_A}\sum_{j=1}^{n_A}\bar{x}_{ij} \qquad (45)$$

その平均値の全体平均からの偏差は以下となる。

$$\Delta\mu_{Aj} = \mu_{Aj} - \mu \qquad (46)$$

$$\Delta\mu_{Bi} = \mu_{Bi} - \mu \qquad (47)$$

因子 A の効果を判定する。関連する分散 S_{Aex}^2 は以下で与えられる。

$$\begin{aligned} S_{Aex}^2 &= \frac{\left[\sum_{i=1}^{n_A}\left(\mu_{A_i} - \mu\right)^2\right] n_B \times 2}{n} \\ &= \frac{\left[\sum_{i=1}^{n_A}\left(\mu_{A_i} - \mu\right)^2\right] n_B \times 2}{2n_A n_B} \\ &= \frac{\sum_{i=1}^{n_A}\left(\mu_{A_i} - \mu\right)^2}{n_A} \\ &= \frac{\left(-1.98\right)^2 + \left(-0.98\right)^2 + \left(0.52\right)^2 + \left(2.44\right)^2}{4} \\ &= 2.77 \end{aligned} \qquad (48)$$

関連する自由度 ϕ_{Aex} は水準数から 1 を引いたもので、以下になる。

$$\begin{aligned} \phi_{Aex} &= n_A - 1 \\ &= 4 - 1 \\ &= 3 \end{aligned} \qquad (49)$$

従って、対応する不偏分散は s_{Aex}^2 は以下のように与えられる。

$$\begin{aligned} s_{Aex}^2 &= \frac{n}{\phi_{Aex}} S_{Aex}^2 \\ &= \frac{24}{3} \times 2.77 \\ &= 22.17 \end{aligned} \qquad (50)$$

次に因子 B の効果と関連するパラメータを計算する。因子 B と関連する分散 S_{Bex}^2 は以下のように与えられる。

$$
\begin{aligned}
S_{Bex}^2 &= \frac{\left[\sum_{i=1}^{n_B}\left(\mu_{B_i}-\mu\right)^2\right] n_A \times 2}{n} \\
&= \frac{\left[\sum_{i=1}^{n_B}\left(\mu_{B_i}-\mu\right)^2\right] n_A \times 2}{2 n_A n_B} \\
&= \frac{\sum_{i=1}^{n_B}\left(\mu_{B_i}-\mu\right)^2}{n_B} \\
&= \frac{(-1.40)^2+(0.01)^2+(1.40)^2}{3} \\
&= 1.31
\end{aligned}
\tag{51}
$$

関連する自由度 ϕ_{Bex} はレベル数 3 から 1 を引いて以下で与えられる。

$$
\begin{aligned}
\phi_{Bex} &= n_B - 1 \\
&= 3 - 1 \\
&= 2
\end{aligned}
\tag{52}
$$

したがって、対応する不偏分散 s_{Bex}^2 は以下で与えられる。

$$
\begin{aligned}
s_{Bex}^2 &= \frac{n}{\phi_{bLex}} S_{Bex}^2 \\
&= \frac{24}{2} \times 1.31 \\
&= 15.69
\end{aligned}
\tag{53}
$$

表 7　二つのデータの平均からの偏差

Data 1		A			
		A_1	A_2	A_3	A_4
B	B_1	0.72	0.10	0.05	0.57
	B_2	-0.45	-1.35	1.02	-0.66
	B_3	0.09	0.87	0.39	2.04

Data 2		A			
		A_1	A_2	A_3	A_4
B	B_1	-0.72	-0.10	-0.06	-0.57
	B_2	0.45	1.35	-1.02	0.67
	B_3	-0.09	-0.87	-0.39	-2.04

我々は一つの水準 $A_j B_i$ に二つのデータを持っている。この二つのデータは一般に異なり、その異なる程度は二つのデータの平均からの偏差で表すことができる。それは以下で与えられる。

$$
e_{pure\_ij\_s} = x_{ij\_s} - \bar{x}_{ij} \tag{54}
$$

今は二つのデータしかないから、以下の関係となる。

$$
e_{pure\_ij\_2} = -e_{pure\_ij\_1} \tag{55}
$$

これを表 7 に示す。

この差に関連する分散 $S_{e\_pure}^2$ は以下で与えられる。

$$\begin{aligned} S_{e\_pure}^2 &= \frac{1}{n}\left[\sum_{j}^{n_B}\sum_{i}^{n_A}\left(e_{pure\_ij\_1}\right)^2\right]\times 2 \\ &= \frac{1}{24}\left\{\left[(0.72)^2+(0.10)^2+(0.05)^2+(0.57)^2\right]\right. \\ &\quad +\left[(-0.45)^2+(-1.35)^2+(1.02)^2+(-0.66)^2\right] \\ &\quad \left.+\left[(0.09)^2+(0.87)^2+(0.39)^2+(2.04)^2\right]\right\}\times 2 \\ &= 0.78 \end{aligned} \tag{56}$$

この誤差に関連する自由度 ϕ_{e_pure} を評価する。

総データ数 n は以下で与えられる。

$$n = n_a \times n_b \times 2 \tag{57}$$

平均値としては $n_a \times n_b$ 種類のものを用いている。したがって、自由度は以下となる。

$$\begin{aligned} \phi_{e\_pure} &= n_a \times n_b \times 2 - n_a \times n_b \\ &= n_a \times n_b \times (2-1) \\ &= 4\times 3 \\ &= 12 \end{aligned} \tag{58}$$

したがって、この誤差の不偏分散 $s_{e\_pure}^2$ は以下で与えられる。

$$\begin{aligned} s_{e\_pure}^2 &= \frac{n}{\phi_{e\_pure}} S_{e\_pure}^2 \\ &= \frac{24}{12}\times 0.78 \\ &= 1.57 \end{aligned} \tag{59}$$

それぞれのデータのトータルの平均からのずれ e_{ij_s} は以下のように与えられる。

$$e_{ij\_s} = x_{ij\_s} - \mu - \left[\left(\mu_{A_j} - \mu\right) + \left(\mu_{B_i} - \mu\right)\right] \tag{60}$$

このずれと、2 セットある場合のずれの関係はどうなっているのであろうか? もし、因子 A と B の効果が独立であれば、この二つのエラーは同じになる。したがって、この違いは二つの因子間の相互作用を表す。この二つの誤差の差を相互作用誤差と呼び、それは以下で与えられる。

$$
\begin{aligned}
e_{\text{interact}\_ij\_s} &= e_{ij\_s} - e_{pure\_ij\_s} \\
&= \left\{ x_{ij\_s} - \mu - \left[\left(\mu_{A_j} - \mu \right) + \left(\mu_{B_i} - \mu \right) \right] \right\} - \left(x_{ij\_s} - \bar{x}_{ij} \right) \\
&= \bar{x}_{ij} - \mu - \left[\left(\mu_{A_j} - \mu \right) + \left(\mu_{B_i} - \mu \right) \right]
\end{aligned}
\tag{61}
$$

これは s に依存しないことに注意しよう。これを具体的に評価したものが**エラー！ 参照元が見つかりません。**である。 対応する分散 S_{interact}^2 は以下となる。

$$
S_{\text{interact}}^2 = \frac{2 \times \left\{ (0.14)^2 + (1.20)^2 + \cdots + (0.41)^2 + (0.61)^2 \right\}}{24} = 0.36 \tag{62}
$$

これに対応する自由度 ϕ_{interact} を評価する。自由度に関する方程式は以下である。

$$
\phi_{tot} = \phi_{aLex} + \phi_{bLex} + \phi_{\text{interact}} + \phi_{e\_pure} \tag{63}
$$

したがって、以下を得る。

$$
\begin{aligned}
\phi_{\text{interact}} &= \phi_{tot} - \left(\phi_{Aex} + \phi_{Bex} + \phi_{e\_pure} \right) \\
&= n - 1 - \left(n_A - 1 + n_B - 1 + n_A \times n_B \times \left(n_s - 1 \right) \right) \\
&= 23 - \left(3 + 2 + 12 \right) \\
&= 6
\end{aligned}
\tag{64}
$$

よって対応する不偏分散 s_{interact}^2 は以下となる。

$$
\begin{aligned}
s_{\text{interact}}^2 &= \frac{n}{\phi_{\text{interact}}} S_{\text{interact}}^2 \\
&= \frac{24}{6} \times 0.36 \\
&= 1.45
\end{aligned}
\tag{65}
$$

表 8　相互作用.

Data 1		A			
		A_1	A_2	A_3	A_4
B	B_1	0.14	1.20	-0.69	-0.65
	B_2	0.47	-0.79	0.28	0.04
	B_3	-0.61	-0.41	0.41	0.61

因子 A の効果は以下のように評価される。

$$
F_A = \frac{s_{Aex}^2}{s_{e\_pure}^2} = 14.14 \tag{66}
$$

対応する F の臨界値 F_{ac} は以下で与えられる。

$$
\begin{aligned}
F_{Ac} &= F\left(\phi_{Aex}, \phi_e, P \right) \\
&= F\left(3, 12, 0.95 \right) \\
&= 3.49
\end{aligned}
\tag{67}
$$

したがって、因子 A は収穫量に対して有効である。

因子 B の効果は以下のように評価される

$$F_B = \frac{s_{Bex}^2}{s_{e\_pure}^2} = 10.01 \tag{68}$$

対応する F の臨界値 F_{Bc} は以下で与えられる。

$$\begin{aligned} F_{Bc} &= F\left(\phi_{Bex}, \phi_e, P\right) \\ &= F\left(2, 12, 0.95\right) \\ &= 3.89 \end{aligned} \tag{69}$$

したがって、因子 B も有効である。

相互作用の効果をみる。

$$F_{\text{interact}} = \frac{s_{\text{interact}}^2}{s_{e\_pure}^2} = 0.92 \tag{70}$$

対応する F の臨界値 $F_{\text{interactc}}$ は以下で与えられる。

$$\begin{aligned} F_{\text{interactc}} &= F\left(\phi_{\text{interact}}, \phi_e, P\right) \\ &= F\left(6, 6, 0.95\right) \\ &= 3.00 \end{aligned} \tag{71}$$

したがって、相互作用はないと言える。

以上を表 8 に示す。

5.6. まとめ

この章のまとめを行う。

- ◆ 一元配置の場合の分散分析

1 因子の効果と関連する分散 S_{ex}^2 は以下で与えられる。

$$S_{ex}^2 = \frac{n_{A_1}\left(\mu_{A_1} - \mu\right)^2 + n_{A_2}\left(\mu_{A_2} - \mu\right)^2 + n_{A_3}\left(\mu_{A_3} - \mu\right)^2}{n_{A_1} + n_{A_2} + n_{A_3}}$$

各水準内の分散の和 S_{in}^2 は以下で与えられる。

$$S_{in}^2 = \frac{\sum_{i=1}^{n_{A_1}}\left(x_{iA_1} - \mu_{A_1}\right)^2 + \sum_{i=1}^{n_{A_2}}\left(x_{iA_2} - \mu_{A_2}\right)^2 + \sum_{i=1}^{n_{A_3}}\left(x_{iA_3} - \mu_{A_3}\right)^2}{n_{A_1} + n_{A_2} + n_{A_3}}$$

対応する相関比 η^2 は以下となる。

$$\eta^2 = \frac{S_{ex}^2}{S_{in}^2 + S_{ex}^2}$$

これは 0 から 1 の間の値をとり、1 に近いほど効果ありとなる。しかし、最終的にはこれを判定に使わない。.

因子と関連する不偏分散は以下で与えられる。

$$s_{ex}^2 = \frac{n}{\phi_{ex}} S_{ex}^2$$

ここで、

$$\phi_{ex} = p$$

p はレベルの数である。

また、水準内の不偏分散は以下で与えられる。

$$s_{in}^2 = \frac{n}{\phi_{in}} S_{in}^2$$

ここで、自由度は以下である。

$$\phi_{in} = n - p$$

したがって、対応する F 値は以下となる。

$$F = \frac{s_{ex}^2}{s_{in}^2}$$

これは自由度 $\left(\phi_{ex}, \phi_{in}\right)$ の F 分布に従う確率変数となることが知られてりう。したがって、

推定確率を P とすると対応する F の臨界値を $F_C\left(\phi_{ex}, \phi_{in}, P\right)$ として評価する。

もし、$F > F_c$ であれば、この因子は有効であり、そうでなければ有効でない、と判断される。

◆ 繰返しのない 2 元配置の場合の分散分析

二つの因子 A およびを考える。それぞれのデータのレベル数は n_A および n_B とする。各水準 $A_j B_i$ にはデータが 1 個 x_{ij} しかないとする。総データ数は n は以下となる。

$$n = n_A \times n_B$$

この場合、各データ x_{ij} を以下のように表現する。

$$x_{ij} = \mu + \left(\mu_{A_j} - \mu\right) + \left(\mu_{B_i} - \mu\right) + e_{ij}$$

関連する分散は以下のようになる。

$$S_{Aex}^2 = \frac{\sum_{i=1}^{n_A} \left(\mu_{A_i} - \mu\right)^2}{n_A}$$

$$S_{Bex}^2 = \frac{\sum_{i=1}^{n_B} \left(\mu_{B_i} - \mu\right)^2}{n_B}$$

$$S_e^2 = \frac{1}{n} \sum_{i=1}^{n_B} \sum_{j=1}^{n_A} e_{ij}^2$$

それぞれの自由度は以下になる。

$$\phi_{tot} = n - 1$$

$$\phi_A = n_A - 1$$

$$\phi_B = n_B - 1$$

誤差と関連する自由度は以下となる。

$$\begin{aligned} \phi_e &= \phi_{tot} - \left(\phi_A + \phi_B\right) \\ &= \left(n-1\right) - \left(n_A - 1 + n_B - 1\right) \\ &= n - \left(n_A + n_B\right) + 1 \end{aligned}$$

したがって、種々の不偏分散は以下となる。

$$s_A^2 = \frac{n}{\phi_A} S_A^2$$

$$s_B^2 = \frac{n}{\phi_B} S_B^2$$

$$s_e^2 = \frac{n}{\phi_e} S_e^2$$

因子 A と関連する F 値は以下となる。

$$F_A = \frac{s_A^2}{s_e^2}$$

これと関連する臨界の F 値を $F_{Ac}\left(\phi_A, \phi_c, P\right)$ とする。F 値がこの臨界の $F_{Ac}\left(\phi_A, \phi_c, P\right)$ より大きければ、因子 A は有効であると判断され、小さければ無効を判断される。

因子 B と関連する F 値は以下となる。

$$F_B = \frac{s_B^2}{s_e^2}$$

F 値がこの臨界の $F_{Bc}\left(\phi_B, \phi_c, P\right)$ より大きければ、因子 B は有効であると判断され、小さけれ

ば無効を判断される。

◆　繰返しのある 2 元配置の分散分析

二つの因子 A およびを考える。それぞれのデータのレベル数は n_A および n_B とする。各水準 $A_j B_i$ にはデータ x_{ij} が n_s 個あるとする。総データ数は n は以下となる。

$$n = n_A \times n_B \times n_s$$

総平均は以下となる。

$$\mu = \frac{1}{n} \sum_{s=1}^{n_s} \sum_{j=1}^{n_A} \sum_{i=1}^{n_B} x_{ij\_s}$$

それぞれのデータの全体平均からの偏差 Δx_{ij_s} は以下のように与えられる。

$$\Delta x_{ij\_s} = x_{ij\_s} - \mu$$

それぞれの水準における平均は以下である。

$$\bar{x}_{ij} = \frac{1}{2}\left(x_{ij\_1} + x_{ij\_1} \right)$$

各因子のレベル平均は以下で与えられる。

$$\mu_{Aj} = \frac{1}{n_B} \sum_{i=1}^{n_B} \bar{x}_{ij}$$

$$\mu_{Bi} = \frac{1}{n_A} \sum_{j=1}^{n_A} \bar{x}_{ij}$$

各水準平均の全体平均からのずれは以下で与えられる。

$$\Delta \mu_{Aj} = \mu_{Aj} - \mu$$

$$\Delta \mu_{Bi} = \mu_{Bi} - \mu$$

したがって、関連する分散はおよび不偏分散は自由度を考慮して以下となる。

$$S_{Aex}^2 = \frac{\sum_{i=1}^{n_A} \left(\mu_{A_i} - \mu \right)^2}{n_A}$$

$$\phi_{Aex} = n_A - 1$$

$$s_{Aex}^2 = \frac{n}{\phi_{Aex}} S_{Aex}^2$$

$$S_{Bex}^2 = \frac{\sum_{i=1}^{n_B}\left(\mu_{B_i} - \mu\right)^2}{n_B}$$

$$\phi_{Bex} = n_B - 1$$

$$s_{Bex}^{(2)} = \frac{n}{\phi_{bLex}} S_{Bex}^{(2)}$$

各水準の誤差およびそれに関連する分散は以下で評価される。

$$e_{pure\_ij\_s} = x_{ij\_s} - \overline{x}_{ij}$$

$$S_{e\_pure}^2 = \frac{1}{n_A n_B n_s}\left[\sum_{j}^{n_B}\sum_{i}^{n_A}\sum_{s}^{n_s}\left(e_{pure\_ij\_s}\right)^2\right]$$

各データの全体に対する誤差は以下で与えられる。

$$e_{ij\_s} = x_{ij\_s} - \mu - \left[\left(\mu_{A_j} - \mu\right) + \left(\mu_{B_i} - \mu\right)\right]$$

この二つの誤差の差は相互作用と関連し、その量および関連する分散は以下で与えられる。

$$\begin{aligned} e_{\mathrm{interact}\_ij\_s} &= e_{ij\_s} - e_{pure\_ij\_s} \\ &= \overline{x}_{ij} - \mu - \left[\left(\mu_{A_j} - \mu\right) + \left(\mu_{B_i} - \mu\right)\right] \end{aligned}$$

$$\begin{aligned} S_{\mathrm{interact}}^2 &= \frac{1}{n_A n_B n_s}\sum_{j}^{n_A}\sum_{i}^{n_B}\sum_{s}^{n_s}\left(e_{\mathrm{interact}\_ij\_s}\right)^2 \\ &= \frac{1}{n_A n_B}\sum_{j}^{n_A}\sum_{i}^{n_B}\left\{\overline{x}_{ij} - \mu - \left[\left(\mu_{A_j} - \mu\right) + \left(\mu_{B_i} - \mu\right)\right]\right\}^2 \end{aligned}$$

$$\begin{aligned} \phi_{\mathrm{interact}} &= \phi_{tot} - \left(\phi_{Aex} + \phi_{Bex} + \phi_{e\_pure}\right) \\ &= n - 1 - \left(n_A - 1 + n_B - 1 + n_A \times n_B \times \left(n_s - 1\right)\right) \end{aligned}$$

$$s_{\mathrm{interact}}^2 = \frac{n}{\phi_{\mathrm{interact}}} S_{\mathrm{interact}}^2$$

因子 A と関連する F 値は以下となる。

$$F_A = \frac{s_{Aex}^2}{s_{e\_pure}^2}$$

対応する臨界の F 値を F_{ac} とする。$F_A \geq F_{Ac}$ であれば、因子 A は効果があるとみなされ、その逆であれば効果があるとは言えないということになる。ただし、

$$F_{Ac} = F\left(\phi_{Aex}, \phi_e, P\right)$$

である。

因子 B と関連する F 値は以下となる。

$$F_B = \frac{s_{Bex}^2}{s_{e\_pure}^2}$$

対応する臨界の F 値を F_{Bc} とする。$F_B \geq F_{Bc}$ であれば、因子 B は効果があるとみなされ、その逆であれば効果があるとは言えないということになる。ただし、

$$F_{Bc} = F\left(\phi_{Bex}, \phi_e, P\right)$$

である。

因子 A と B の相互作用と関連する F 値は以下のように評価される。

$$F_{\mathrm{interact}} = \frac{s_{\mathrm{interact}}^2}{s_{e\_pure}^2}$$

対応する臨界の F 値を $F_{\mathrm{interactc}}$ とする。$F_{\mathrm{interact}} \geq F_{\mathrm{interactc}}$ であれば、相互作用効果があるとみなされ、その逆であれば効果があるとは言えないということになる。ただし、

$$F_{\mathrm{interactc}} = F\left(\phi_{\mathrm{interact}}, \phi_e, P\right)$$

である。