IV-2. データの構造と取り扱い

IV-2-1. 分散の分離

実際のデータは、様々な要因が関係して出来た結果です。ですから、それらの影響を切り分けて論ずる必要があります。具体的には、データの分散（バラつき）をその原因（要因）ごとに切り分けるということです。ある原因と他の原因が原因と結果のような関係性を持っているという場合は、積の分散（共分散）もありますが、それは相関分析になります。ここではまず、お互いに関係を持たない、独立した原因がある場合を考えます。

様々な説明の仕方がありますが、ここでは、解析が前提としているモデル、論理構造を鮮明にするために、いくつかの要因によって出来ているデータセットを、自分たちで作ってみることにします。例としては、和の分散で取り上げたデータの組み合わせの例を使います。

まず、要因Ａによる以下のデータがあるとします。

データ

1、5、 6

ここのデータをA_iと表すことにすると

$$A_1 = 1,\ A_2 = 5,\ A_3 = 6$$

たとえば、三つの植木鉢に、それぞれ窒素量の違う肥料を入れて、それぞれの鉢に植えた草の丈が、１ｃｍ、５ｃｍ　６ｃｍだったというようなことを考えてください

この情報を整理すると

サンプルサイズ（データ数）

$$n_A = 3$$

自由度

$$\mathrm{df}_A = n_A - 1 = 3 - 1 = 2$$

平均値

$$M_A = \frac{\text{データーの総和}}{\text{サンプルサイズ}} = \frac{Sum_A}{n_A} = \frac{\sum_{i=1}^{n_A} A_i}{n_A} = \frac{1+5+6}{3} = 4$$

平方和

$$SS_A = (1-4)^2 + (5-4)^2 + (6-4)^2 = 14$$

分散

$${\sigma_A}^2 = \frac{SS_A}{df_A} = \frac{(1-4)^2 + (5-4)^2 + (6-4)^2}{3-1} = 7$$

標準偏差

$$\sigma_A = \sqrt{{\sigma_A}^2} = \sqrt{7}$$

データがあるとき、そのデータが正規分布しているという前提があれば、それらのデー

タ群は、平均値と分散で代表させせることができます。データは、様々な要因によって異なる値を取りますが、それぞれの要因について偏りなくデータがとられていれば、一つの要因内の、平均値と分散（データの散らばり方）で代表させることもできます。

次に別に要因 B によるデータがあって、以下のようなデーターだとします。
データ

1、5、６、8

ここのデータをB_jと表すことにすると

$$B_1 = 1,\ B_2 = 5,\ B_3 = 6,\ B_4 = 8$$

たとえば、４つの植木鉢があって、そこに入れる肥料のリンの量を４段階に変えて、草丈を比較したというようなことを考えてください。
要因 A によるデータと同様に、このデーターを代表する統計量を計算すると
サンプルサイズ

$$n_B = 4$$

自由度

$$\mathrm{df}_B = 4 - 1 = 3$$

平均値

$$M_B = \frac{1+5+6+8}{4} = 5$$

平方和

$$SS_B = (1-5)^2 + (5-5)^2 + (6-5)^2 + (8-5)^2 = 26$$

分散

$$\sigma_B{}^2 = \frac{26}{3} = 8.66667$$

標準偏差

$$\sigma_B = \sqrt{8.66667}$$

2 つの要因によるデータの統計量をまとめると、以下の通りです。
A：平均 M_{A}＝4　サンプルサイズ　$n_A = 3$ 平方和 SS_{A}　＝14　分散 σ^2_{A}＝７
B：平均 M_{B}＝5　サンプルサイズ　$n_A = 4$ 平方和 SS_{B}　＝26　分散 σ^2_{B}＝8.66667

この２つの要因によるデータの広がり方（バラつき）を比較するには、その比を計算しますね。

$$F = \frac{\sigma_A{}^2}{\sigma_B{}^2} = \frac{7}{8.66667} = 0.807692$$

この値を見ると、A の要因によるデータの広がりは、B の要因による広がりよりも少し小さいけれど、その違いはあまりないなと、感じるでしょう。多分、人は直感的にこういう比

較をしています。
例えば、データの広がりが、B の 10 分の 1 になっている、次のようなデータ群 C を考えます。
データ

0.1、0.5、0. 6 、0.8

ここのデータをC_jと表すことにすると

$$C_1 = 0.1,\ C_2 = 0.5,\ C_3 = 0.6,\ C_4 = 0.8$$

たとえば、4 つの植木鉢に、異なる人の写真を張り付けて、その人の顔が植木鉢の中の草の成長に与える影響を比較するというようなことを考えてください。そんなことが成長に影響するはずがないと思うかもしれませんが、決めつけてはいけません。世の中何があるかわかりません。

根データの統計量は以下の通りです
サンプルサイズ

$$n_C = 4$$

自由度

$$df_C = 4 - 1 = 3$$

平均値

$$M_C = \frac{0.1 + 0.5 + 0.6 + 0.8}{4} = 0.5$$

平方和

$$SS_C = (0.1 - 0.5)^2 + (0.5 - 0.5)^2 + (0.6 - 0.5)^2 + (0.8 - 0.5)^2 = 0.26$$

分散

$$\sigma_C{}^2 = \frac{0.26}{3} = 0.0866667$$

標準偏差

$$\sigma_C = \sqrt{0.0866667}$$

これも A の統計量と並べて見比べてみましょう。
2 つの要因によるデータの統計量をまとめると、以下の通りです。
A : 平均 M_{A}＝4　サンプルサイズ　$n_A = 3$　平方和 SS_{A} ＝14　分散 $\sigma^2{}_{\mathrm{A}}$＝ 7
C: 平均 M_{C}＝0.5　サンプルサイズ　$n_C = 4$　平方和 SS_{C} ＝0.26　分散 $\sigma^2{}_{\mathrm{C}}$＝0.0866667
この分散の大きさを比べると、

$$F = \frac{\sigma_A{}^2}{\sigma_C{}^2} = \frac{7}{0.0866667} = 80.7692$$

つまり、A の要因によるデータの広がりは、C の要因のデータの広がりの１００倍近い大き

さを持って広がっているのです。
A の要因と C の広がりの大きさが、同じだという人はいないでしょう。もちろん、これは感覚的なものですが、A と C のデータの広がりが同じであるにもかかわらず、たまたま、母集団から取り出したデータの取り出しかたによって、この様な比になることが、どのくらいの確率で起こるのかを計算しておけば数量的な根拠をもってそのような判断が出来るでしょう。たとえば、5%以下の確率でしか起こらないという境界になる F の値を数学的モデルから計算しておき、実際のデータの値と数学的なモデルから計算した値と大きさを比較すれば、実際のデータから計算した値が、データの取り出し方によってたまたま出た値であるかどうか判断することができます。この F のことを分散比とよび、こういう分析の仕方を分散分析（F 検定）といいます。
この議論ができたのは、要因ごとのデータの統計量を私たちが知っていたからです。実際のデータはこういう形では与えられません。様々な要因が加わって出来た結果だけをデータとして与えられます。そこで、２つの要因が重なって出来た（相加的な要因という言い方が良いかもしれません。）データの値がどのようになるかを考えてみます。要因 A と要因 B が相加的に関与してできるデータで、要因同士は独立しています。この場合、A の特定のレベルと B の特定のレベルだけを選び出してたしあわせることはできませんから、すべてのデータを総当りでたしあわせることにします。

表 1. 二つの要因によるデータの総当たりの和

	1	5	6	Sum	Mean
1	1＋1 = 2	5＋1 =6	6＋1 = 7	15	5
5	1＋5=6	5＋5 =10	6＋5 = 11	27	9
6	1＋6 =7	5＋6 = 11	6＋6 = 12	30	10
8	1＋8 = 9	5＋8 = 13	6＋8 = 14	36	12
Sum	24	40	44	108	9
Mean	6	10	11	9	

赤字が A の要因で決まるデータ、青字が B の要因で決まるデータです。

A の平均 M_A＝4　SS_A　＝14　分散 σ^2_A＝ 7

B の平均 M_B＝5　SS_B　＝26　分散 σ^2_B＝8.66667

全体の平均は 9 ですね。$M=M_A+M_B$ になっています。データをたし合わせているのだから、その平均値も平均値の和になるというのは当然です。
この表で示された、データの和によって出来たデータ集が、この様な計算によって意図的に作られたものだと知らない人は、表中の 2,6,7.6,10,11,7,11,12,9.13.14 というデータから、すべての合計を

$$Sum_{total} = 2 + 6 + 7 + 6 + 10 + 11 + 7 + 11 + 12 + 9 + 13 + 14 = 108$$

平均を

$$M_{total}=\frac{Sum_{\ total}}{n_{total}}=\frac{Sum_{\ total}}{n_A n_B}=\frac{108}{12}=9$$

全体の平方和を

$$SS_{\ total}=(2-9)^2+(6-9)^2+(7-9)^2+(6-9)^2+(10-9)^2+(11-9)^2+(7-9)^2 \\ +(11-9)^2+(12-9)^2+(9-9)^2+(13-9)^2+(14-9)^2 \\ =49+9+4+9+1+4+4+4+9+0+16+25=134$$

と計算するでしょう。

この全体の自由度は、サンプルサイズ引く１で、$(3\times 4)-1$＝１１、だから、

全体の分散は、

$$\sigma_{total}{}^2=\frac{SS_{total}}{\mathrm{df}_{total}}=\frac{134}{11}=12.18182$$

この結果に「あれ変だな」と思う人がいると思います。変だと思わない人もいると思います。この結果に対する違和感は後で説明しますが、その前に、この計算は正直で丁寧な計算ですが少し面倒ですね

$SS_{\ total}$の個々のデータの平均値からの距離の２乗の計算値を、表に入れてみると

表 11. SS の計算

	1	5	6	Sum
1	49*	9	4	62
5	9	1	4	14
6	4	4	9	17
8	0	16	25	41
Sum	62	30	42	134

こんな風になっていて、この表の、行ごとに計算したものを、縦にたしあわせるか、列ごとに計算したものを横にたしあわせるかで、134 という値が求まります。

ところで、＊印のところの計算ですが、$(1+1-9)^2=49$と計算しています。この計算は、このデータを作った元の A と B２の要因に戻って考えれば、$\{(1-4)+(1-5)\}^2$と言う計算です。

このように考えると、１行目の各列をたしあわせた総和の計算は

$$\{(1-4)+(1-5)\}^2+\{(5-4)+(1-5)\}^2+\{(6-4)+(1-5)\}^2=62$$

と言う計算でもあります。この式の外側のカッコを外して展開してみましょう。

$$\{(1-4)+(1-5)\}^2+\{(5-4)+(1-5)\}^2+\{(6-4)+(1-5)\}^2 \\ =(1-4)^2+(5-4)^2+(6-4)^2+2(1-5)\{(1-4)+(5-4)+(6-4)\}+3(1-5)^2$$

ところで、

$$SS_A=(1-4)^2+(5-4)^2+(6-4)^2=14$$

$$(1-4)+(5-4)+(6-4)=0$$

ですね、$(1-4)$、$(5-4)$、$(6-4)$は平均値からの距離なのだからその総和は０です。
つまり１行目の平方和は

$$SS_A+n_A(1-5)^2=14+3\times16=62$$

ということです。
これは２行目についても同じで
２行目は

$$SS_A+n_A(5-5)^2=14+3\times0=14$$

以下同様に
３行目

$$SS_A+n_A(6-5)^2=14+3\times1=17$$

４行目

$$SS_A+n_A(8-5)^2=14+3\times9=41$$

です。これらをたしあわせたものが、$SS_{\ total}$ですから

$$SS_{\ total}=n_BSS_A+n_A\{(1-5)^2+(5-5)^2+(6-5)^2+(8-5)^2\}$$

ですが

$$SS_B=(1-5)^2+(5-5)^2+(6-5)^2+(8-5)^2$$

ですから

$$SS_{\ total}=n_BSS_A+n_ASS_B=4\times14+3\times26=134$$

となります。

$$SS_{\ total}=n_BSS_A+n_ASS_B$$

この式は、全平方和$SS_{\ total}$がどのような要因で構成されているのかを示しています。２つの要因で説明されるデータで、それぞれの要因の組み合わせの中に繰り返しがない時の分析を、２要因分散分析言います。何かの要因と何かの要因を組み合わせて実験するような場合のことです。たとえば、植木鉢を１２個用意して、それを４ずつ３組に分けて、それぞれの組の肥料の窒素濃度を３段階、（A1，A2,A3）に設定する。そのそれぞれの組について、４段階のリンの濃度（B1, B2, B3, B4）を設定して、合計１２組の肥料濃度の異なる、植木鉢を作って、それぞれの植木鉢に草を１本だけ植えて、その成長を比較するというような場合です。実験条件としては、A1B1, A1B2, と言うような組み合わせになります。でも、これは少し不自然ですね。普通、こういう場合は、草を数本植えて繰り返しを作るでしょうから、あまり現実的な想定ではありませんが、絶対にあり得ないことではないし、初めはできるだけ単純な方が考えやすいので、この形のデータについて考えます。
とにかく次のような表３のようなデータが得られたとします。このデータは、今まで検討してきた合成したデータそのものです。しかし、分析者は合成データだということを知りません。この場合、それぞれの行ごと列ごとに平均値を求めるでしょう。それらの平均値を使って平均値の平均値を計算し、A、B の要因によう る平方和（SS）を計算します。

表 3. 表 1 と同じ

	A1	A2	A3	Sum	Mean
B1	2	6	7	15	5
B2	6	10	11	27	9
B3	7	11	12	30	10
B4	9	13	14	36	12
Sum	24	40	44	108	9
Mean	6	10	11	9	

A の要因間の平均値

$$\frac{6+10+11}{3}=9$$

要因 A による SS、

$$(6-9)^2-(10-9)^2-(11-9)^2=14$$

B の要因間の平均

$$\frac{5+9+10+12}{4}=9$$

要因 B による SS、

$$(5-9)^2-(9-9)^2-(10-9)^2+(12-9)^2=26$$

確かに我々が合成した、元のデータの要因ごとの分散になっています。どうしてそうなるのかと言うことは、多分わかると思いますが、念のために説明します。

A の要因間の平均値の計算のために各行ごとの平均値を求めるために、各行のデータの和を求めます。合成した元のデータまでさかのぼって 1 行ずつ計算すると

$$\{1+1\}+\{5+1\}+\{6+1\}=(1+5+6)+3\times(1)$$

となっています。その平均を求めて、行の平均としているのだから。

$$\frac{Sum_A+n_A\times 1}{n_A}=M_A+1$$

という計算をしていることになります。

以下同様に

２行目については

$$\frac{Sum_A+n_A\times 5}{n_A}=M_A+5$$

3 行目については

$$\frac{Sum_A+n_A\times 6}{n_A}=M_A+6$$

4 行目については

$$\frac{Sum_A + n_A \times 8}{n_A} = M_A + 8$$

つまり、その行の元のデータの値に、A の要因（列データ）によるデータの平均値をたしたものが、行の平均なのです。だから、当然、行の平均データをすべての行についてたしあわせたものは

$$n_B M_A + 1 + 5 + 6 + 8$$

になります。また

$$M_B = \frac{1 + 5 + 6 + 8}{n_B}$$

ですから、

$$n_B M_A + 1 + 5 + 6 + 8 = n_B M_A + n_B M_B = n_B(M_A + M_B)$$

となり、行の平均値の平均は

$$\frac{n_B M_A + n_B M_B}{n_B} = M_A + M_B$$

となります。列についても同様に、列の平均値の平均は

$$\frac{n_A M_A + n_A M_B}{n_A} = M_A + M_B$$

となります。次に平方和について考えると

１行目について

$$\{(M_A + 1) - (M_A + M_B)\}^2$$

という計算をしているのだから、

$$\{(M_A + 1) - (M_A + M_B)\}^2 = (1 - M_B)^2$$

以下

２行目

$$(5 - M_B)^2$$

3 行目

$$(6 - M_B)^2$$

4 行目

$$(8 - M_B)^2$$

となり、その合計は

$$(1 - M_B)^2 + (5 - M_B)^2 + (6 - M_B)^2 + (8 - M_B)^2$$

となりますが、これはSS_Bですね。

このことは列についても同様で、列の平均値の分散はSS_Aです。

行の要因（この場合は要因 B）による平方和は、個々の行ごとの平均値の平方和で、列の要因（この場合は要因 A）による平方和は、個々の列の平均の平方和です。合成した元のデータの分散を知らなくても、データからそれを構成するデータの平方和と分散を求めること

ができるということです。
この場合は、

$$\sigma_A{}^2 = \frac{SS_A}{\mathrm{df}_A} = \frac{14}{2} = 7$$

$$\sigma_B{}^2 = \frac{SS_B}{\mathrm{df}_B} = \frac{26}{3} = 8.66667$$

$$\sigma_{total}{}^2 = \frac{SS_{total}}{\mathrm{df}_{total}} = \frac{134}{11} = 12.18182$$

さて、ここで、前に取り上げた、一部の気のまわる人が感じる違和感の話です。

$$\sigma_{total}{}^2 \neq \sigma_A{}^2 + \sigma_B{}^2$$

だということです。もちろん、計算のプロセスの詳細な解説を読んだ後ならば、この結果は当然のことですが、この計算の詳細な解説を聞く前は

$$\sigma_{total}{}^2 = \sigma_A{}^2 + \sigma_B{}^2$$

となることを期待したのではないでしょうか。
分散の合計を

$$\sigma_{A+B}{}^2 = \sigma_A{}^2 + \sigma_B{}^2$$

と表すと

$$\sigma_{total}{}^2 \neq \sigma_{A+B}{}^2$$

です。
注意深くデータを見ると、自由度についても、平方和についても同じことが言えます。

$$\mathrm{df}_A = 2$$

$$\mathrm{df}_B = 3$$

$$\mathrm{df}_{total} = 11$$

$$\mathrm{df}_{total} \neq \mathrm{df}_A + \mathrm{df}_B$$

$$SS\ _{total} = n_B SS_A + n_A SS_B$$

ですから、当然、

$$SS\ _{total \neq} SS\ _{A+B} = SS_A + SS_B$$

$$\sigma_A{}^2 = \frac{SS_A}{\mathrm{df}_A} = \frac{14}{2} = 7$$

$$\sigma_B{}^2 = \frac{SS_B}{\mathrm{df}_B} == \frac{26}{3} = 8.6667$$

これで、全体のデータの広がりを構成する個々の要因の分散が計算できたことになります。
整理すると、

$$\sigma_{total}{}^2 = \frac{SS_{total}}{\mathrm{df}_{total}} = \frac{134}{11} = 12.18182$$

自由度１１

$$\sigma_A{}^2 = \frac{SS_A}{\mathrm{df}_A} = 7$$

自由度２

$$\sigma_B{}^2 = \frac{SS_B}{\mathrm{df}_B} = 8.6667$$

自由度３

ここで分散の大きさを比較して、その相対的な大きさを判断することになりますが、

$$F_{B-A} = \frac{\sigma_B{}^2}{\sigma_A{}^2} = \frac{8.6667}{7} = 1.2381$$

分子の自由度３、分母の自由度２

この内、F_{B-A}と言う比較は、あまり意味のある比較ではないでしょう。実は比較すべき分散は他にあるのですが、この分析では、その分散は原理的に０になってしまうので、それができません。実際のデータは要因と要因が完全に独立していなかったり、何か不確定の要因で変動したり、特定の要因の組合わせの時に、何かに対する強い影響が観察されたりします。このデータはそういうことが起こらないように、元のデータを総当りにして、それぞれの要因のレベルの間で、分散が異ならないようにしているのです。しかし、そうしたことを検討するには、データの繰り返し（ランダムな変動）をどのように扱えば良いのか理解する必要があります。そこで、これまでの検討を拡張して、データの繰り返しがある場合について考えます。具体的には、分布の広がりに差があるデータの組み合わせの例として出した、要因ＡとＣの場合について考えます。データが、0.1、0.5, 0.6、0.8 と言う例です。

サンプルサイズ

$$n_C = 4$$

自由度

$$df_C = 4 - 1 = 3$$

平均値

$$M_C = \frac{0.1 + 0.5 + 0.6 + 0.8}{4} = 0.5$$

平方和

$$SS_{\hat{C}} = (0.1 - 0.5)^2 + (0.5 - 0.5)^2 + (0.6 - 0.5)^2 + (0.8 - 0.5)^2 = 0.26$$

分散

$$\sigma_{\hat{C}}{}^2 = \frac{26}{3} = 0.0866667$$

これを総当り表で表すと表４のようになります。

表 4. データのたしあわせの総当たり表

	A1	A2	A3	Sum	Mean
C1	1.1	5.1	6.1	12.3	4.1
C2	1.5	5.5	6.5	13.5	4.5
C3	1.6	5.6	6.6	13.8	4.6
C4	1.8	5.8	6.8	14.4	4.8
Sum	6.0	22.0	26.0	54.0	4.5
Mean	1.5	5.5	6.5	4.5	

2 つの要因によるデータの統計量をまとめると、以下の通りです。

A : 平均 M_A＝4　サンプルサイズ　$n_A = 3$ 平方和$SS_{\hat{A}} = 14$　分散$\sigma_{\hat{A}}{}^2 = 7$

C: 平均 M_C＝0.5　サンプルサイズ　$n_C = 4$ 平方和$SS_{\hat{C}} = 0.26$　分散$\sigma_{\hat{C}}{}^2 = 0.0866667$

次のように考えてみます。私たちが、A が有意ではないかと考えるのは、C という要因を考えたときに、F の値が大きく、A のデータの広がりが C に比べて、十分に大きかったからです。つまり、要因 C をランダムにおこる変動のように捉えて、それに比べて十分大きいと考えたのです。それならば、その感覚に近いモデルを考えるべきでしょう。そもそも、植木鉢に張った人の顔写真などというものが、成長に影響を与えるはずがありません。もしろ、それは、実験の際に、いくつかのグループに分けてその植木鉢の管理をする人を決めて、わかりやすいように、その人の写真を張ったという程度のことでしょう。C を単なる繰り返しに過ぎないと考え見ます、やることは、

	A1	A2	A3	Sum	Mean
C1	1.1	5.1	6.1	12.3	4.1
C2	1.5	5.5	6.5	13.5	4.5
C3	1.6	5.6	6.6	13.8	4.6
C4	1.8	5.8	6.8	14.4	4.8
Sum	6.0	22.0	26.0	54.0	4.5
Mean	1.5	5.5	6.5	4.5	

という表を表 5 の様に書き換えるということです。と書き換えるということです。行がなくなるので、行の平均という概念がありません。

表５のように書き換えても

$$\mathrm{M}_{total} = 4.5$$

$$SS_{total} = 56.78$$

が変わらないことを確かめておいてください。

表 5. 1 要因分散分析

	A1	A2	A3
	1.1	5.1	6.1
	1.5	5.5	6.5
	1.6	5.6	6.6
	1.8	5.8	6.8
Sum	6.0	22.0	26.0
Mean	1.5	5.5	6.5

ここで要因 A による分散と、繰り返しによるランダムな分散(これを残差分散 residual : $\sigma_{residual}{}^2$と言います。)を計算します。

まず自由度ですが、

$$\mathrm{df}_{total} = \mathrm{df}_A + \mathrm{df}_{residual}$$

$$\mathrm{df}_{residual} = \mathrm{df}_{total} - \mathrm{df}_A = (n_A n_B - 1) - (n_A - 1) = n_A(n_B - 1)$$

次に平方和の構成は、

$$SS_{total} = SS_A + SS_{residual}$$

$SS_{residual}$は A の各レベルの繰り返しの中の平方和を全レベルについてたしあわせたものだから

$$\begin{aligned} SS_{residual} &= \{(1.1-1.5)^2 + (1.5-1.5)^2 + (1.6-1.5)^2 + (1.8-1.5)^2\} \\ &\quad + \{(5.1-5.5)^2 + (5.5-5.5)^2 + (5.6-5.5)^2 + (5.8-5.5)^2\} \\ &\quad + \{(6.1-6.5)^2 + (6.5-6.5)^2 + (6.6-6.6)^2 + (6.8-6.5)^2\} \\ &= 3 \times 0.26 = 0.78 \end{aligned}$$

この値が、$n_A SS_C$であることは、もちろん、当然のことです。

残りの部分が、要因 A に起因するデータの広がりだから、

$$SS_A = SS_{total} - SS_{residual} = 56.78 - 0.78 = 56$$

$$\sigma_{residual} = \frac{SS_{residual}}{\mathrm{df}_{residual}} = \frac{0.78}{3(4-1)} = \frac{0.78}{9} = 0.086667$$

$$\sigma_A = \frac{SS_A}{\mathrm{df}_A} = \frac{14}{(3-1)} = 7$$

$$F_{A-residual} = \frac{\sigma_A{}^2}{\sigma_{residual}{}^2} = \frac{7}{0.086667} = 80.76892$$

これは、かなり大きな値で、観測値 80.76892、分母の自由度9、分子の自由度2で F 表を引くまでもなく、A の変動はランダムな変動に比べて有意に大きいと言えるでしょう。このような検定を、1要因分散分析による検定といいます。計算の途中で、気が付かれたかと思いますが、14という値は、$\frac{56}{4} = n_B SS_A/4$になっています。行の平均値から、$SS_A$を求めて、

n_B（繰り返しの数）を乗じて、56 という値を求めるという方法がないわけではありませんが、繰り返し数は条件間で単一になるとは限らないので、残差を先に求めて要因間の分散を全分散との間の差し引きで求める方が一般的で安全でしょう。