

二元配置分散分析（繰り返しのないデータ）

1. 目的

二元配置分散分析は2つの要因に関する平均の差を検定するものです。

ここでは、高齢男女合計6人に記憶試験を施し、年齢（年代）の違い、男女の違いで差があるものを調査するものとします。つまり、

- ・要因Aは 年代（60歳代、70歳代、80歳代）（3択）
- ・要因Bは 男女（2択）

とします。

計測値は、試験結果です。

ここで検討するのは、以下です。

- ・要因Aの年代の違いは、結果に違いをもたらすのか？
- ・要因Bの男女の違いは、結果に違いをもたらすのか？

データは以下の通りで、すべての要因A、要因Bの組み合わせにおけるデータ数は1でなければなりません。

		要因A	要因B	値
No	試験 ID	年代	性別	計測値
1	60歳代 男性	60歳代	男性	40
2	70歳代 男性	70歳代	男性	35
3	80歳代 男性	80歳代	男性	20
4	60歳代 女性	60歳代	女性	30
5	70歳代 女性	70歳代	女性	28
6	80歳代 女性	80歳代	女性	15

帰無仮説は以下です。

- ・要因Aの年代（60代、70代、80代）の母平均は等しい。（差はない）
- ・要因Bの男女の母平均は等しい。（差はない）

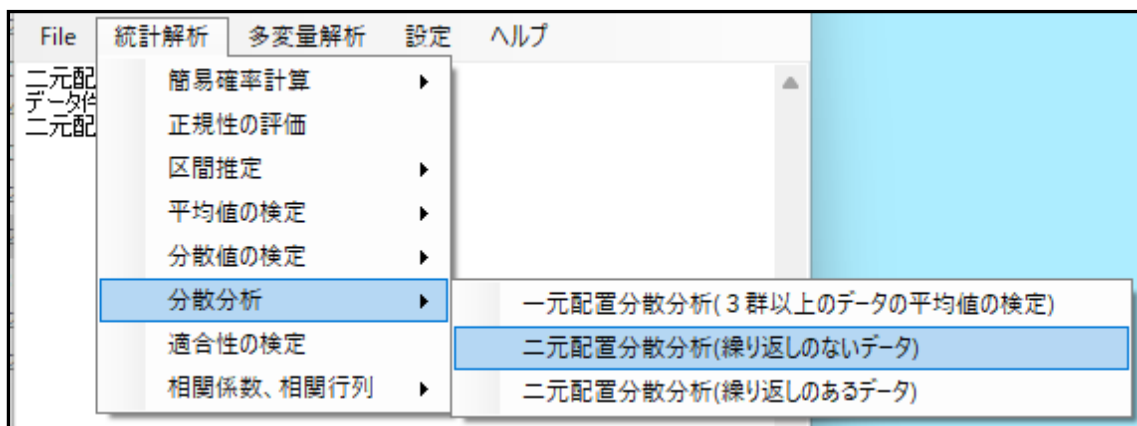
対立仮説は以下です。

- ・要因Aの年代(60, 70, 80代)の母平均は異なる。（年代による差はある）
- ・要因Bの男女の母平均は異なる。（男女による差はある）

2. 使用法

(1) メニューの選択

メニューの「統計解析→分散分析→二元配置分散分析(繰り返しのないデータ)」を選択します。



(2) パネルが表示されます。

棄却域の確率を示します。
通常 5%を利用するので、
デフォルトで5が指定されて
いる。変更可能。

The screenshot shows the '二元配置分散分析 (繰り返しのない)' (Two-way ANOVA (no repetition)) panel. The panel includes a '有意水準 α (%)' (Significance level α (%)) field set to 5, a '計算実行' (Calculate) button, and a table for the results of the ANOVA. A red circle highlights the significance level field, and a red arrow points from the text box above to it. Another red arrow points from the text box to the right to the '結果' (Results) column of the ANOVA table.

変動因	平方和	自由度	平均平方	F 値	P 値 (%)	結果	帰無仮説の採択域
要因 A							(,) 分布関数
要因 B							(,) 分布関数
誤差							
全体							

	Factor A	Factor B	個数	合計	平均	(不偏)分散

表データを貼り付け ☐ 先頭行をラベルとして使用 ☐ クリア ☐ データ分布

	NO	ID	Factor A	Factor B	Value
*					

計算結果が
表示される部分

(3) データの入力

パネルのグリッド（下の部分）にデータを入力します。

表データを貼り付け

☐ 先頭行をラベルとして使用

クリア

	NO	ID	FactorA	FactorB	Value
▶*					

データは表計算ソフトのデータをコピーして貼り付けます。
表データの形式は（ID、要因1、要因2、値）です。
要因A、要因Bの列には、データがどこに所属するかを示します。

例えば、今回の例では、
要因Aは年代ですので、“60歳代”～“80歳代”などの文字を入れます。
要因Bは性別ですので、例えば“男性”、“女性”などの文字が入ります。

		要因A	要因B	値
No	試験 ID	年代	性別	計測値
1	60歳代 男性	60歳代	男性	40
2	70歳代 男性	70歳代	男性	35
3	80歳代 男性	80歳代	男性	20
4	60歳代 女性	60歳代	女性	30
5	70歳代 女性	70歳代	女性	28
6	80歳代 女性	80歳代	女性	15

上記のように表計算ソフトで作成したデータの赤線部分をコピーします。

表題部分も取り込みますので、「先頭行をラベルとして使用」にチェックを入れます。

表データを貼り付け

をクリックすると、グリッド部分にコピーされます。

同時に、各要因 毎にデータの集計、平均、分散が表形式で表示されます。

	年代	性別	個数	合計	平均	(不偏)分散
▶	60歳代	***	2	70	35	50
	70歳代	***	2	63	31.5	24.5
	80歳代	***	2	35	17.5	12.5

上記のグリッドは表示が小さい為、スクロールして見ることになりますが、グリッドの左上隅のセルをクリックして、Ctrl-C を押すと、表形式のデータがコピーされます。その際、グリッドが 以下のように青くなるのを確認してください。

ここをクリックして Ctrl-C を押す

	年代	性別	個数	合計	平均	(不偏)分散
▶	60歳代	***	2	70	35	50
	70歳代	***	2	63	31.5	24.5
	80歳代	***	2	35	17.5	12.5

そのまま、表計算ソフトのシート上で Ctrl-V を押すと、表データがペーストされます。

年代	性別	個数	合計	平均	(不偏)分散
60歳代	***	2	70	35	50
70歳代	***	2	63	31.5	24.5
80歳代	***	2	35	17.5	12.5
***	男性	3	95	31.66667	108.3333
***	女性	3	73	24.33333	66.33334

- 1) 要因Aのカテゴリは年代であり、
 - ・60歳代の件数は2、平均 35、分散 50
 - ・70歳代の件数は2、平均 31.5、分散 24.5
 - ・80歳代の件数は2、平均 17.5、分散 12.5
- 2) 要因Bのカテゴリは“男性”、“女性”であり、
 - ・男性の件数は3、平均 31.67、分散 108.3
 - ・女性の件数は3、平均 24.33、分散 66.33

(4) 計算条件の指定

有意水準 α (%):

“有意水準”には デフォルトで 5 が指定されています。変更できます。

(5) 計算実行

計算実行 ボタンを押すと計算されます。

変動因	平方和	自由度	平均平方	F 値	P 値 (%)	結果	帰無仮説の採択域
要因 A	343	2	171.5	54.157894	1.813	有意	(0 , 19) 分布関数
要因 B	80.666666	1	80.666666	25.473684	3.709	有意	(0 , 18.512820) 分布関数
誤差	6.3333333	2	3.1666666				
全体	430	5					

平方和、自由度、平方平均、F 値等 が 入力された標本データをもとに計算されます。

要因 A (“60 歳代”、“70 歳代”、“80 歳代”) については：

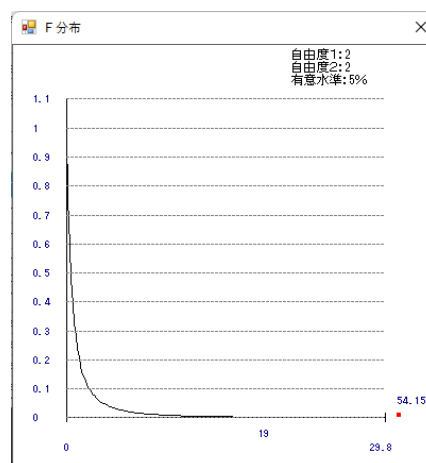
F 値が 54.16、P 値が 1.8 % となり、“有意”と判断されました。
つまり 年代別についての有意差が認められました。

要因 B (“男性”、“女性”) については：

F 値が 25.47、P 値が 3.7 % となり、“有意”と判断されました。
つまり 男女別についての有意差が認められました。

この様子は、 **分布関数** ボタンを押すことで確認できます。

例えば要因 A の **分布関数** ボタンを押すと、F 分布のグラフが描画されています。
計算された F 値は 54.15 なので、グラフのはるか右側に位置することになります。図では、図からはみ出さない位置に 赤い点で描画されています。



(6) 分布のグラフ描画

要因の効果を直感的に理解するためのグラフです。

要因の効果とは、この場合

- ① 年代の違いに差があるか？
- ② 男女の違いに差があるか？

となります。

各測定点 (X) から全データの平均 (μ) との差は、以下で表現されます。

$$X - \mu = a + b + \varepsilon$$

- ・ a は 要因A による効果
- ・ b は 要因B による効果
- ・ ε は 誤差

以降の各図は、

要因Aについての図は a の大きさを表示
要因Bについての図は b の大きさを表示
誤差についての図は 誤差の大きさを表示

となります。

少し正確に書くと、上記については 以下のようになります。

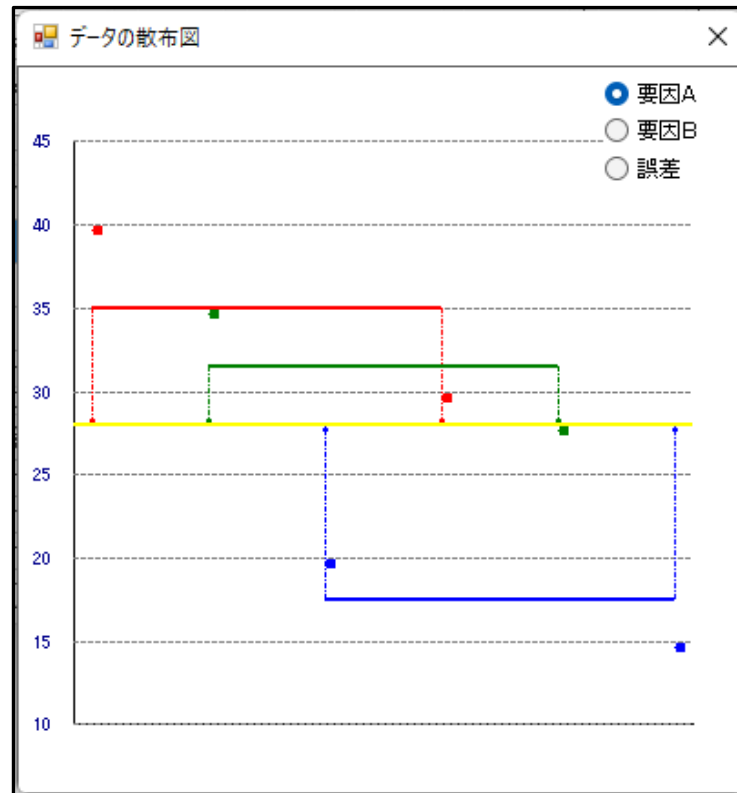
$$\begin{aligned} X_{i,j} - \mu &= a_i + b_j + \varepsilon_{i,j} \\ a_i &\equiv (\mu_i^A - \mu) \\ b_j &\equiv (\mu_j^B - \mu) \\ \varepsilon_{i,j} &\equiv X_{i,j} - \mu - (\mu_i^A - \mu) - (\mu_j^B - \mu) \end{aligned}$$

$$\mu = \frac{\sum_{i=1}^{n_A} \sum_{j=1}^{n_B} X_{i,j}}{n_A * n_B}, \mu_i^A = \frac{\sum_{j=1}^{n_B} X_{i,j}}{n_B}, \mu_j^B = \frac{\sum_{i=1}^{n_A} X_{i,j}}{n_A}$$

詳細は「考え方」のドキュメントを参照してください。

データ分布を直観的に理解する為に、 データ分布 ボタンを押します。

1) 最初に 要因A についてのグラフを表示します。



図の 赤い点、緑の点、青の点は 入力された測定点です。

この図は 要因Aを選んでますので、

- ・赤点は 要因Aの“60歳代”のデータ
- ・緑点は 要因Aの“70歳代”のデータ
- ・青点は 要因Aの“80歳代”のデータ

を示します。

また 水平の線は、

- ・赤線は 要因Aの“60歳代”の平均を示す線
- ・緑線は 要因Aの“70歳代”の平均を示す線
- ・青線は 要因Aの“80歳代”の平均を示す線

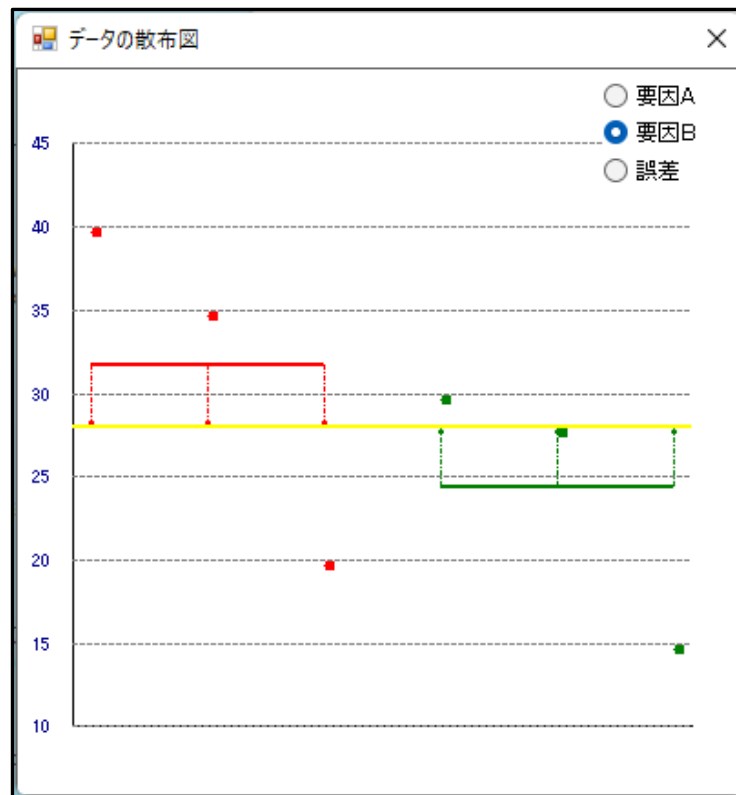
を示し、

- ・黄色い線は、全体の平均を示す線

です。

赤、緑、青の各線から黄色の線に垂直に矢印線が引かれていますが、この矢印が大きく表示されていると、要因Aについての群間差（年齢による差）が大きいことを示し、逆に小さいと群間差は小さいことを示します。

2) 今度は 要因B を指定してグラフを表示します。



この図は 要因Bを選んでますので、

- ・赤点は 要因Bの“男性”のデータ
- ・緑点は 要因Bの“女性”のデータ

を示します。

先と同様に 水平の線は、

- ・赤線は 要因Bの“男性”の平均を示す線
- ・緑線は 要因Bの“女性”の平均を示す線

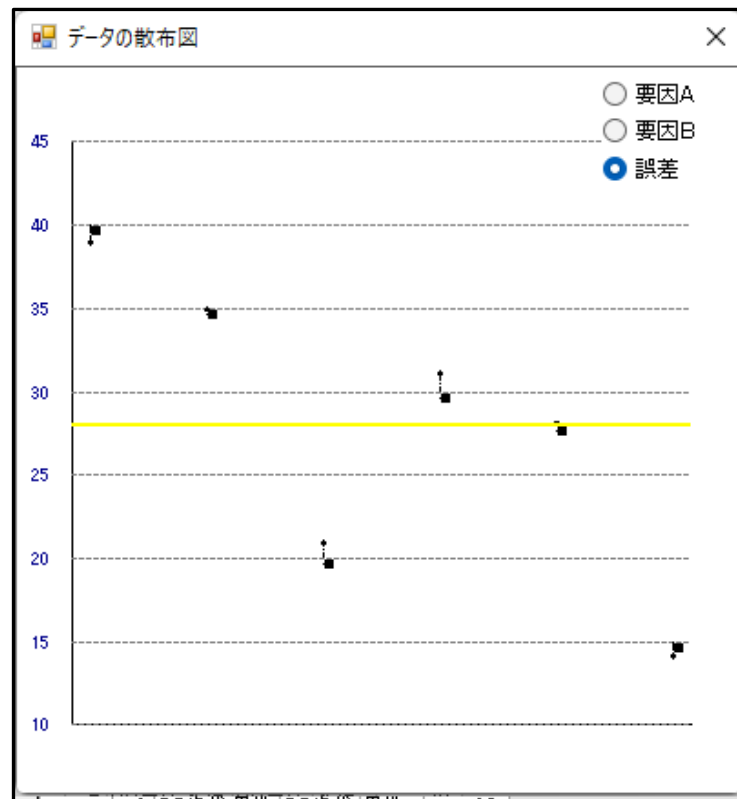
を示し、

- ・黄色い線は、全体の平均を示す線

です。

赤、緑の各線から黄色の線に垂直に矢印線が引かれていますが、この矢印が大きく表示されていると、要因Bについての群間差（男女別による差）が大きいことを示し、逆に小さいと群間差は小さいことを示します。

3) 続いて 誤差を指定してグラフを表示します。



黄色い線は、全体の平均点を示す線です。

誤差は $\varepsilon = X - \mu - a - b$ であり、測定点と全体の平均点と差異 ($X - \mu$) から、要因A (a) と要因B (b) との影響を除去したものです。

この誤差が小さければ、要因A，要因Bの効果が大きい、逆にこの誤差が大きければ、要因A，要因Bの効果は小さいということが直感的に理解できます。