

分散分析 野菜ジュース飲料製造業者の コロナ期の株価変動

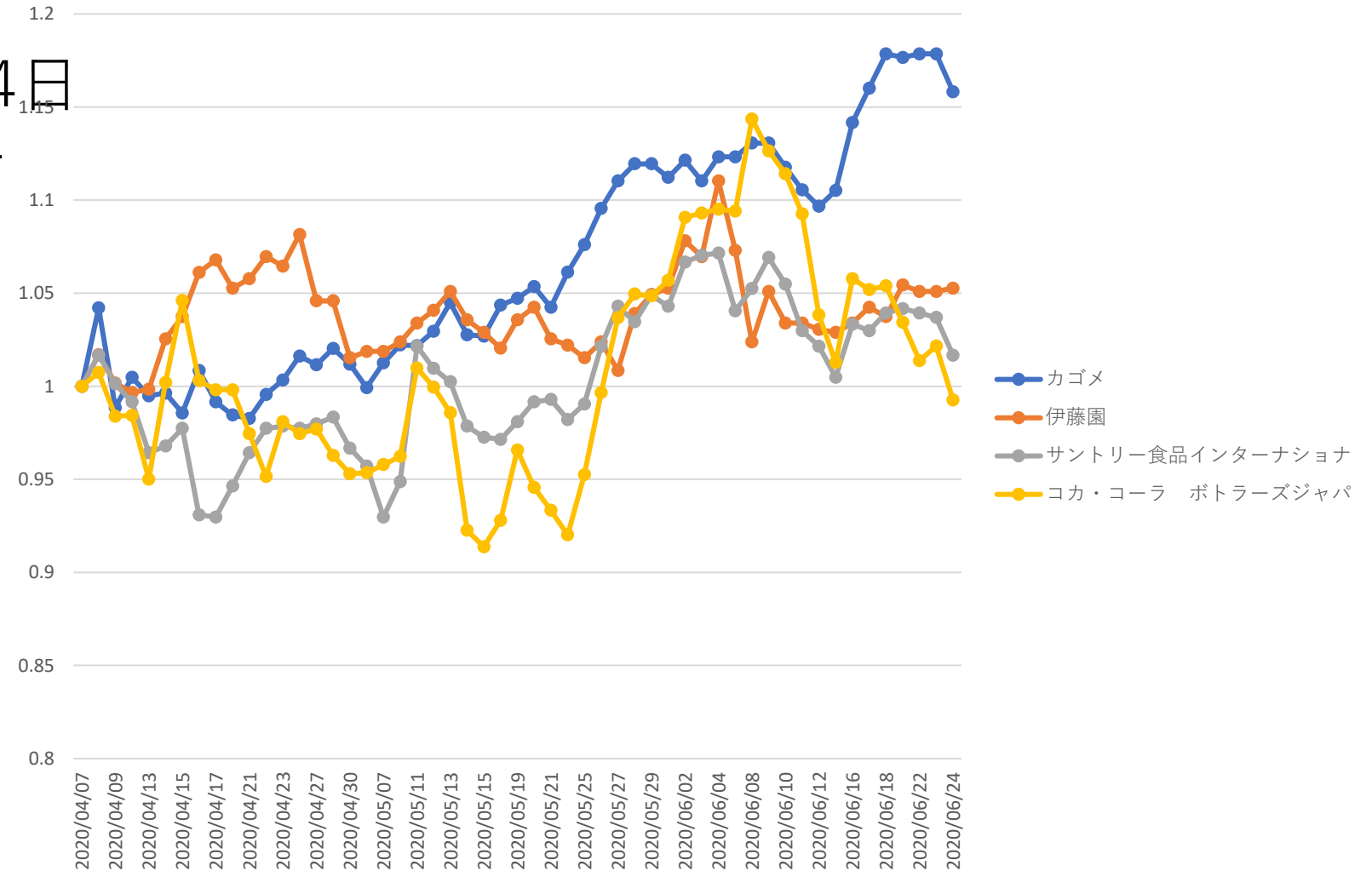
2020年7月14日

学習院大学経済学部経営学科

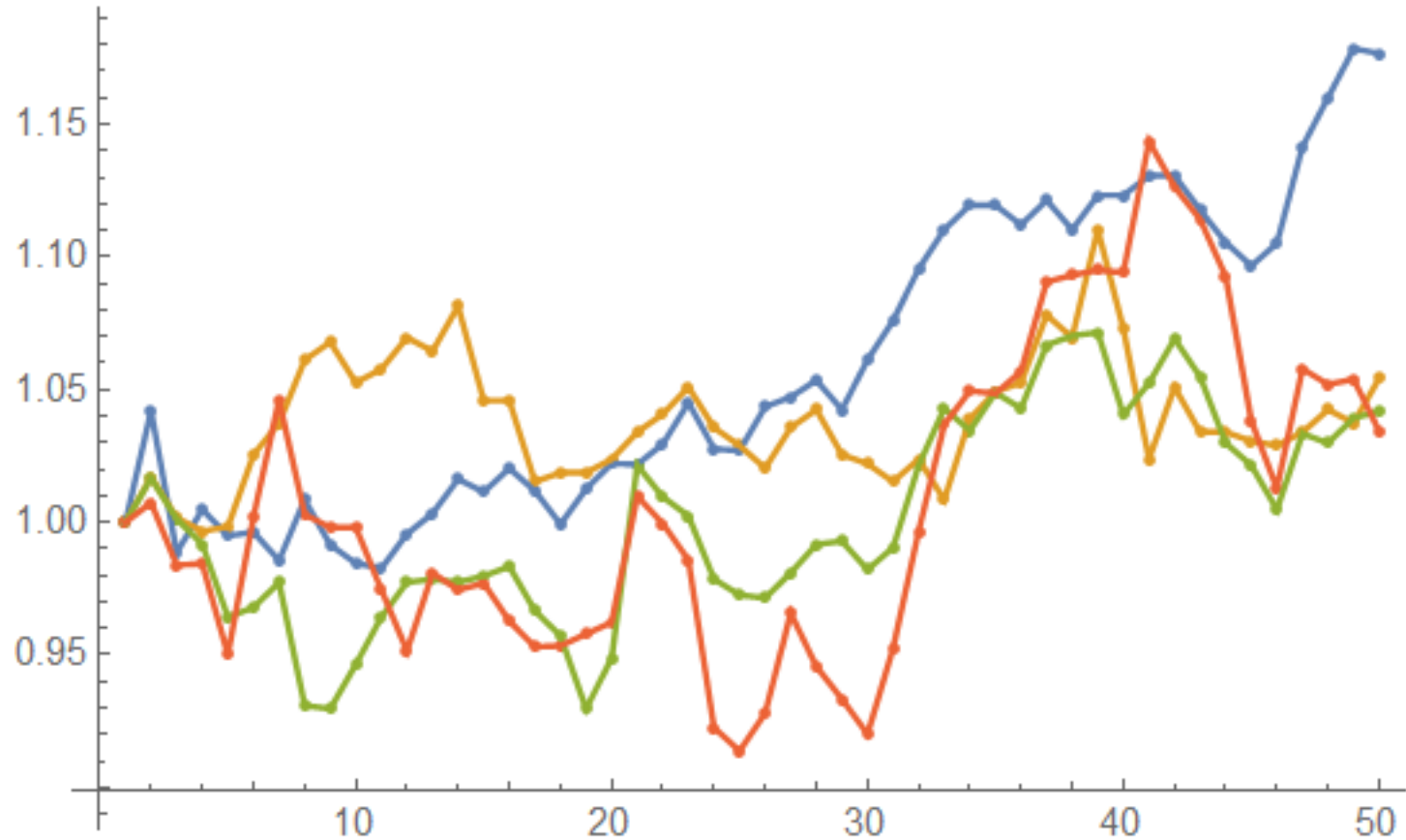
白田由香利

飲料メーカー4社の株価変動

- 4月7日から6月24日
- 4月7日を1として指標化



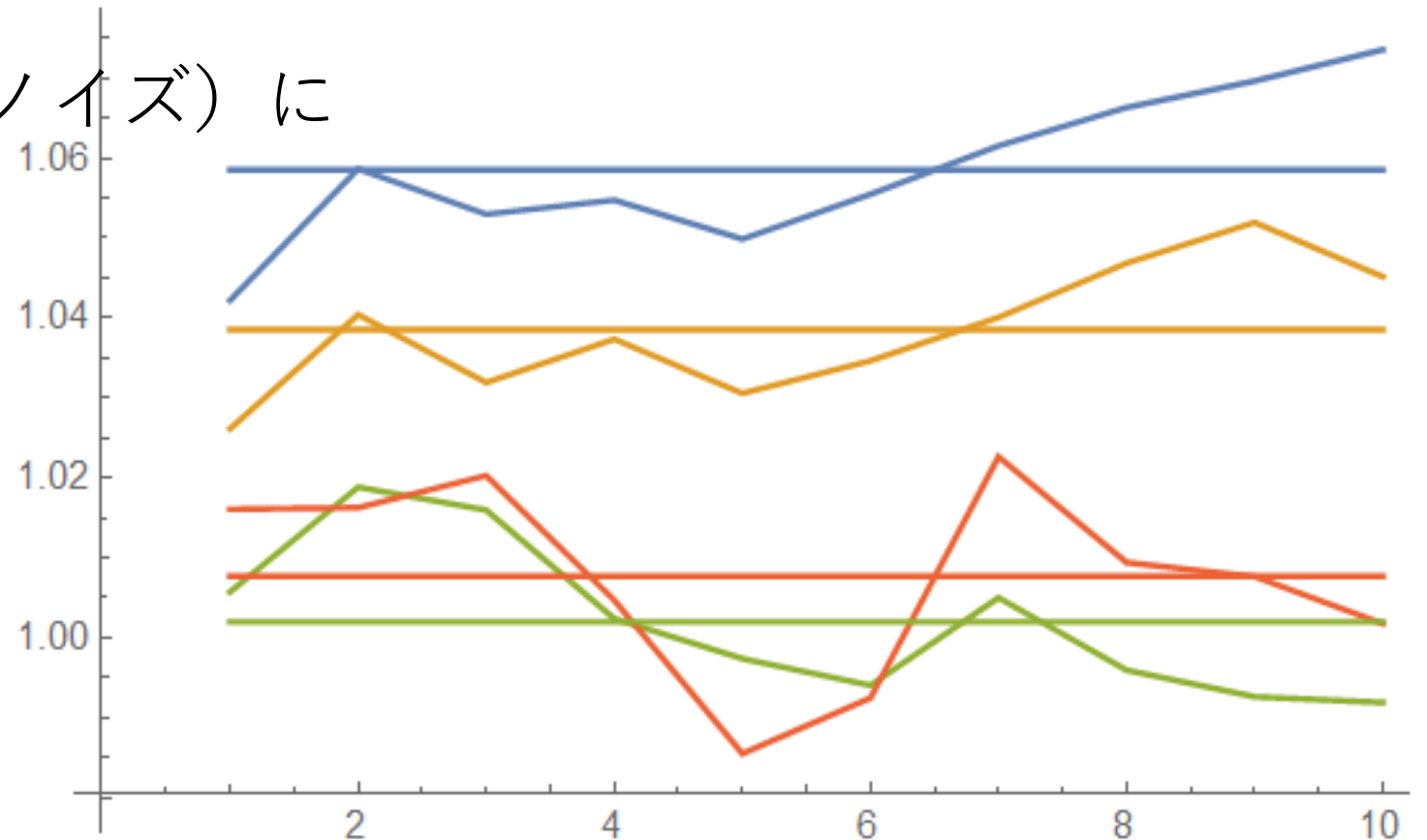
4月7日を1とした, 50営業日分の株価変動
50日の4群の平均 {1.0585,1.03851,1.00198,1.00767}



1元配置：分散分析 (ANOVA)

- 帰無仮説 $\mu_1 = \mu_2 = \mu_3 = \mu_4$ (全てのメーカーで株価母平均は同じ)
- 対立仮説 母平均は全てのメーカーで同じではない

群間変動が群内変動（ノイズ）に
比較して大きければ
群間で母平均に違いが
あると言える

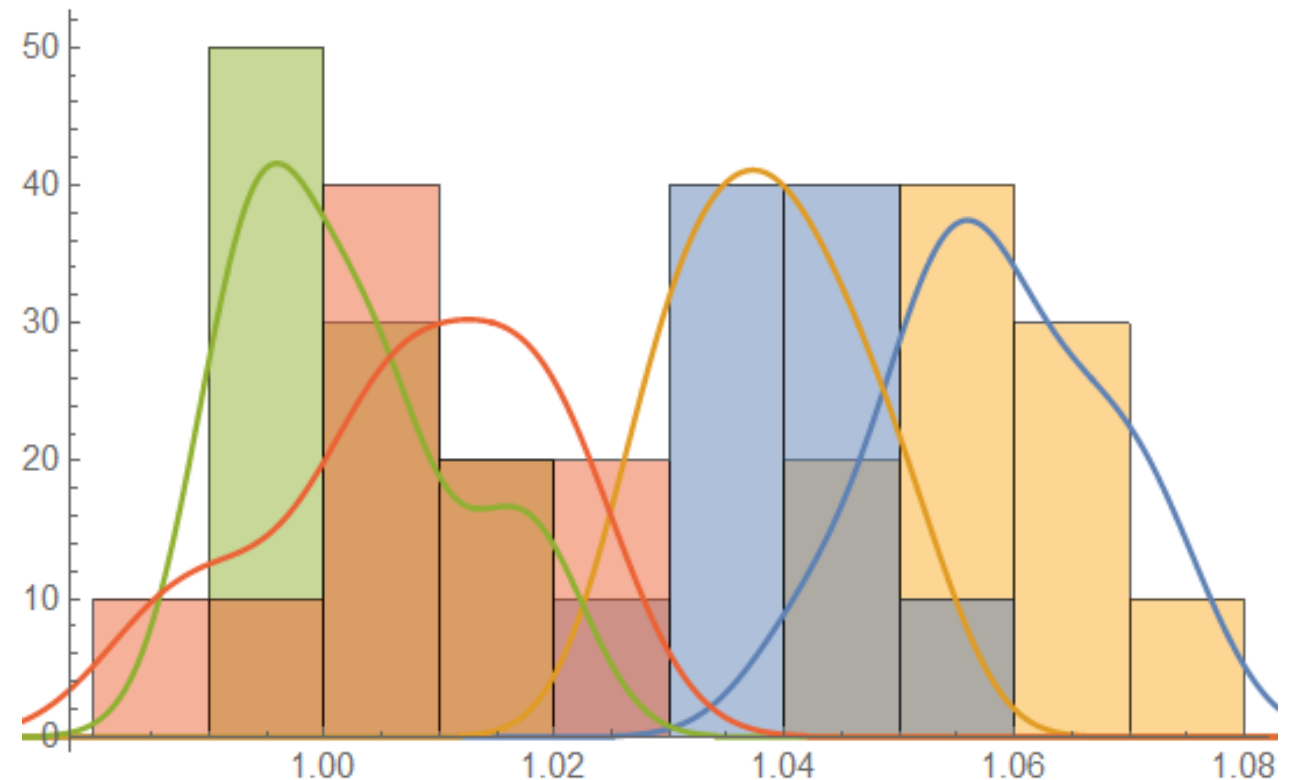


ANOVAは等分散性を仮定する これが言えなければANOVAは使えない

- 4 群の標本分散 {0.0000905727, 0.0000640607, 0.0000903517, 0.000144156}
- 等分散性の検定：Bartlett法
- 帰無仮説： $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2$

	Statistic	P-Value
Conover	0.912411	0.822432
Bartlett	1.4623	0.690998
Levene	0.46035	0.711695
Brown-Forsythe	0.432124	0.731269

- **Bartlett**
0.69 で帰無仮説は棄却されない
(有意水準 5 % の場合)
- 等分散性は否定されなかった



1元配置ANOVAをEXCELで行う

デー→データ分析⇒分散分析：一元配置

本澤飲料水データ - Excel

Shirota Yukari SY

ファイル ホーム 挿入 描画 ページレイアウト 数式 データ 校閲 表示 ヘルプ 希望する作業を入力

データの取得と変換 クエリと接続 データの種類 並べ替えとフィルター データツール 予測 アウトライン 分析

	E	F	G	H	I
1	カゴメ		カゴメ	伊藤園	サントリー食品インター
2	2711	2020/04/07	1	1	1
3	2825	2020/04/08	1.0420509	1.016977929	1.016686532
4	2680	2020/04/09	0.98856511	1.001697793	1.001191895
5	2724	2020/04/10	1.00479528	0.996604414	0.991656734
6	2697	2020/04/13	0.99483585	0.998302207	0.964243147
7	2701	2020/04/14	0.99631132	1.025466893	0.967818832
8	2672	2020/04/15	0.98561416	1.037351443	0.977353993
9	2734	2020/04/16	1.00848395	1.061120543	0.930870083
10	2688	2020/04/17	0.99151605	1.067911715	0.929678188
11	2669	2020/04/20	0.98450756	1.052631579	0.94636472
12	2664	2020/04/21	0.98266322	1.057724958	0.964243147
13	2699	2020/04/22	0.99557359	1.069609508	0.977353993
14	2720	2020/04/23	1.00331981	1.064516129	0.978545888
15	2755	2020/04/24	1.01623017	1.081494058	0.977353993
16	2742	2020/04/27	1.01143489	1.045840407	0.979737783
17	2766	2020/04/28	1.02028772	1.045840407	0.983313468
18	2743	2020/04/30	1.01180376	1.015280136	0.966626937

分散分析: 一元配置

入力元
入力範囲(W): \$A\$1:\$D\$51

データ方向:
 列(C)
 行(R)

先頭行をラベルとして使用(L)

α (A): 0.05

出力オプション
 出力先(Q):
 新規ワークシート(P):
 新規ブック(W)

OK
キャンセル
ヘルプ(H)

1元配置ANOVAをEXCELで行う

- F値 **16.0**
- 有意水準5%の境界値**2.65**
- 今回のF値が帰無仮説棄却領域に落ちる ⇒ 帰無仮説棄却
- 結論
母平均は全てのメーカーで同じではない

分散分析: 一元配置						
概要						
グループ	データの 個数	合計	平均	分散		
カゴメ	50	52.92475	1.058495	0.003357		
伊藤園	50	51.9253	1.038506	0.000545		
サントリー食品	50	50.09893	1.001979	0.001513		
コカ・コーラ	50	50.38364	1.007673	0.003388		
分散分析表						
変動要因	変動	自由度	分散	観測された分散比	P-値	F 境界値
グループ間	0.106174	3	0.035391	16.08212	2.19E-09	2.650677
グループ内	0.431331	196	0.002201			
合計	0.537506	199				

時系列の変化に対する違いはないか？
⇒ 10営業日ごとの平均の違いを見る

• **2元配置：分散分析**

• 因子 1：メーカーの違い

- 帰無仮説 $\mu_1 = \mu_2 = \mu_3 = \mu_4$ (全てのメーカーで母平均は同じ)
- 対立仮説 母平均は全てのメーカーで同じではない

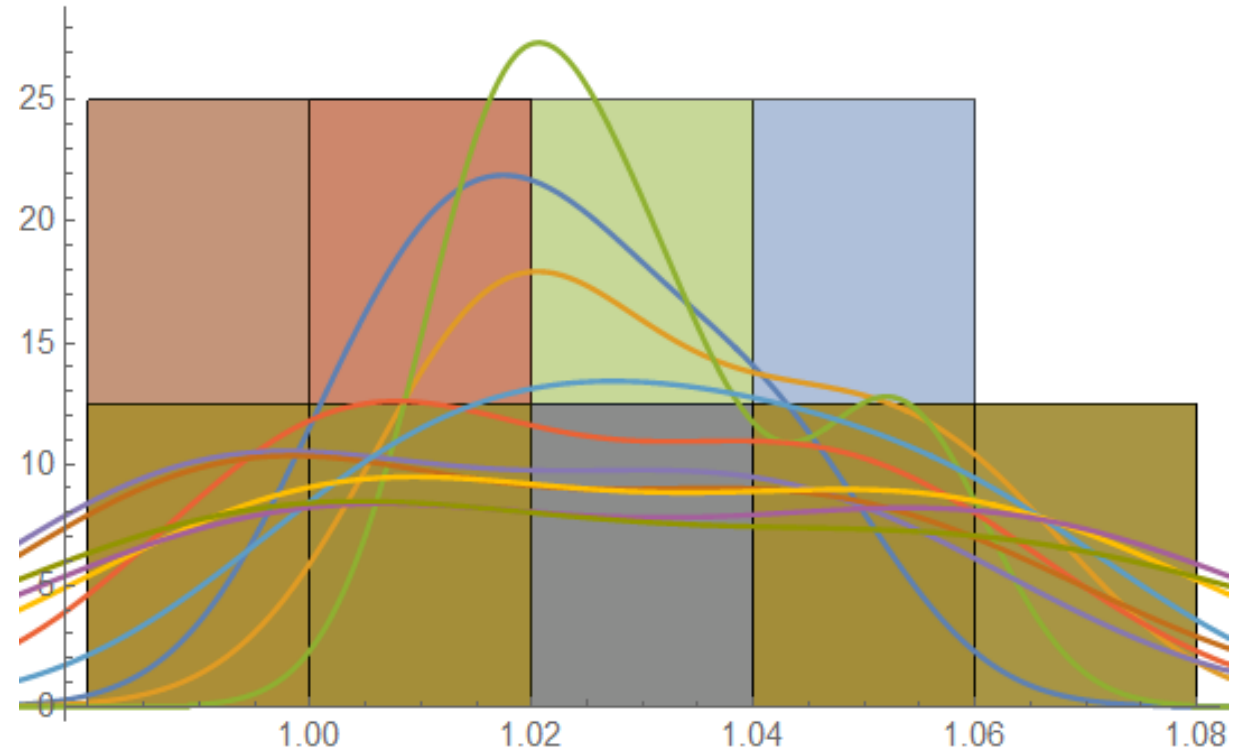
• 因子 2：時系列の違い

- 帰無仮説 $\mu'_1 = \mu'_2 = \mu'_3 = \mu'_4 = \mu'_5$ (全ての期間で母平均は同じ)
- 対立仮説 母平均は全ての期間で同じではない

期間の異なる 5 群に関する等分散性の検定

- 前提条件として，期間に関する 5 つの群に対しても等分散性が保証されないといけませんが，Bartlett法による検定で保証されたので，検定を行う。

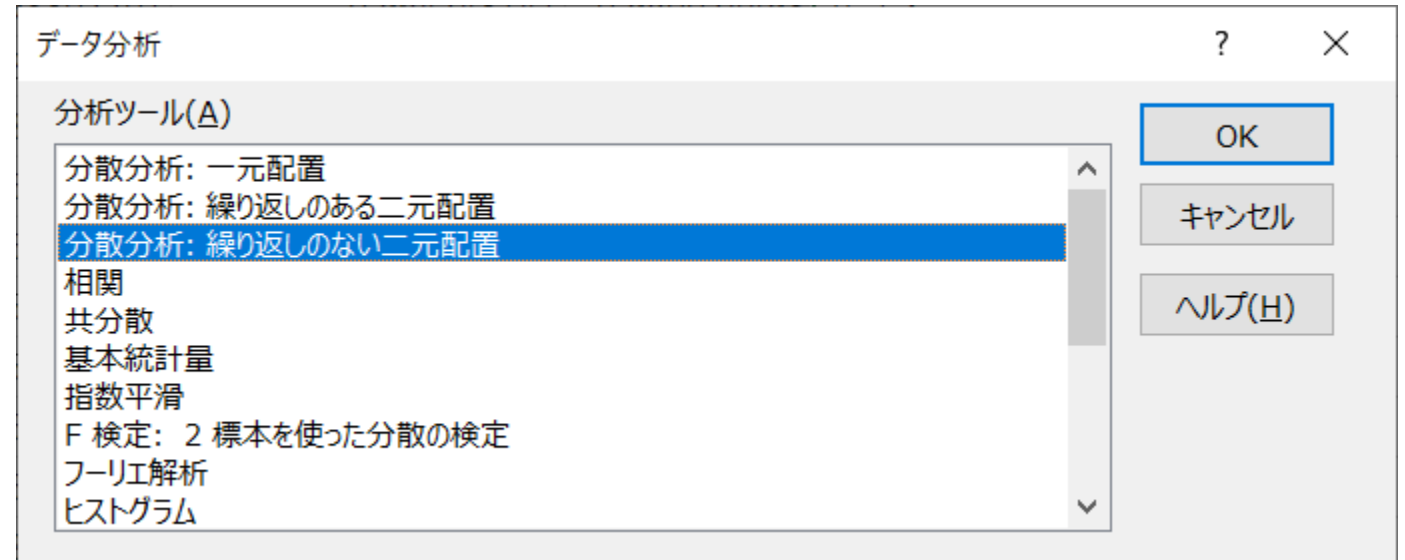
	Statistic	P-Value
Conover	13.0333	0.161106
Bartlett	4.29657	0.890832
Levene	2.32792	0.0399562
Brown-Forsythe	2.01536	0.0727156



課題提出：EXCEL

分散分析：繰り返しのない二元配置

- 1日目から50日目まで、10日ごとの平均を計算する。
- そのデータに対して2元配置分散分析を行うこと。



Mathematicaで行った結果を載せる

```
In[ ]:= Needs["ANOVA`"]
```

[存在しない場合に読み込む]

```
ANOVA[twowaydata, {factor1, factor2}, {factor1, factor2}, PostTests → {Tukey},  
SignificanceLevel → .05, CellMeans → False]
```

[有意水準]

[偽]

```
Out[ ]:= {ANOVA →
```

	DF	SumOfSq	MeanSq	FRatio	PValue
factor1	3	0.0212349	0.00707829	84.7744	7.38249×10^{-14}
factor2	9	0.00124789	0.000138654	1.66062	0.148093
Error	27	0.00225438	0.0000834956		
Total	39	0.0247371			

```
,
```

```
PostTests → {factor1 → Tukey {{1, 2}, {1, 3}, {2, 3}, {1, 4}, {2, 4}}, factor2 → Tukey {}}
```

- 因子 1 については有意水準 5 % で，帰無仮説棄却
- 因子 2 については有意水準 5 % で，帰無仮説は棄却されない
- ANOVAでは，もっと詳しく比較がしたい⇒ 多重比較法

多重比較 Tukey法による結果

```
In[*]:= Needs["ANOVA`"]
```

[存在しない場合に読み込む]

```
ANOVA[twowaydata, {factor1, factor2}, {factor1, factor2}, PostTests -> {Tukey},  
SignificanceLevel -> .05, CellMeans -> False]
```

[有意水準]

[偽]

```
Out[*]= { ANOVA -> 

|         | DF | SumOfSq    | MeanSq       | FRatio  | PValue                    |
|---------|----|------------|--------------|---------|---------------------------|
| factor1 | 3  | 0.0212349  | 0.00707829   | 84.7744 | $7.38249 \times 10^{-14}$ |
| factor2 | 9  | 0.00124789 | 0.000138654  | 1.66062 | 0.148093                  |
| Error   | 27 | 0.00225438 | 0.0000834956 |         |                           |
| Total   | 39 | 0.0247371  |              |         |                           |

 ,
```

```
PostTests -> {factor1 -> Tukey {{1, 2}, {1, 3}, {2, 3}, {1, 4}, {2, 4}}, factor2 -> Tukey {}}
```

- 因子 1 については有意水準 5 % で，帰無仮説棄却
- $1 > 2$, $1 > 3$, $1 > 4$: カゴメは他 3 社に有意差有
- $2 > 3$, $2 > 4$: その他，伊藤園はサントリーとコカコーラに有意差有

多重比較 Tukey法による結果

- $1 > 2, 1 > 3, 1 > 4$: カゴメは他3社に有意差有
- $2 > 3, 2 > 4$: その他, 伊藤園はサントリーとコカコーラに有意差有

4社飲料水種類	野菜飲料	全種類	野菜比率
カゴメ	53	80	0.66
伊藤園	21	129	0.16
サントリー食品インターナショナル	2	335	0.01
コカ・コーラ ボトラーズジャパ	0	141	0.00

白田ゼミ3年本澤君がカウントしてくれたデータ

多重比較

- 3群以上あるとき
- どの会社がどの会社と比較して有意差があるか，を検定するとき，全ての組合せについて調べていく．
- しかし，個々の検定をt検定で行なうということを繰り返していくと，有意差がないのに，有意差がある間違いが起こる可能性が高くなる．対策として有意水準を小さくしていく．
- 結論：多重比較をする際は，**多重比較法(Tukey法など)** を使わなくてはならない．