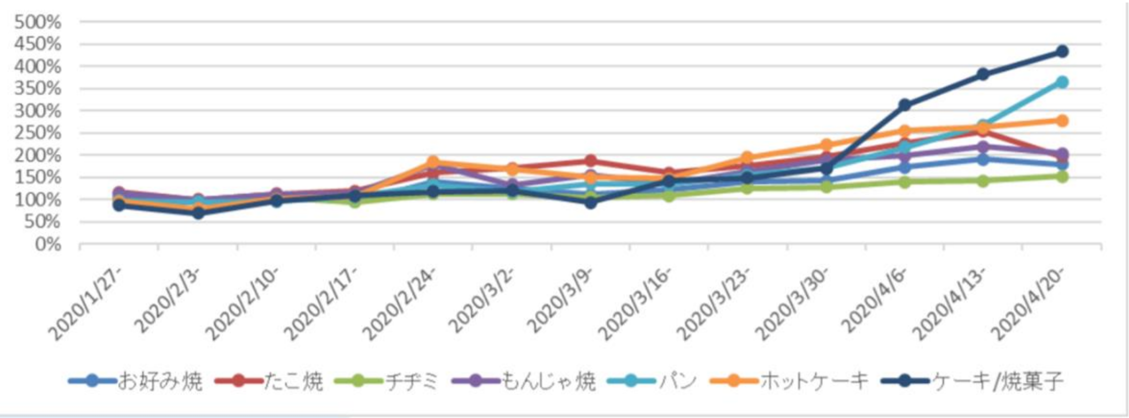
分散分析
コロナ期のホットケーキの売上

2020年7月15日

学習院大学経済学部経営学科
白田由香利

コロナ期の物品の売上の違いを見る



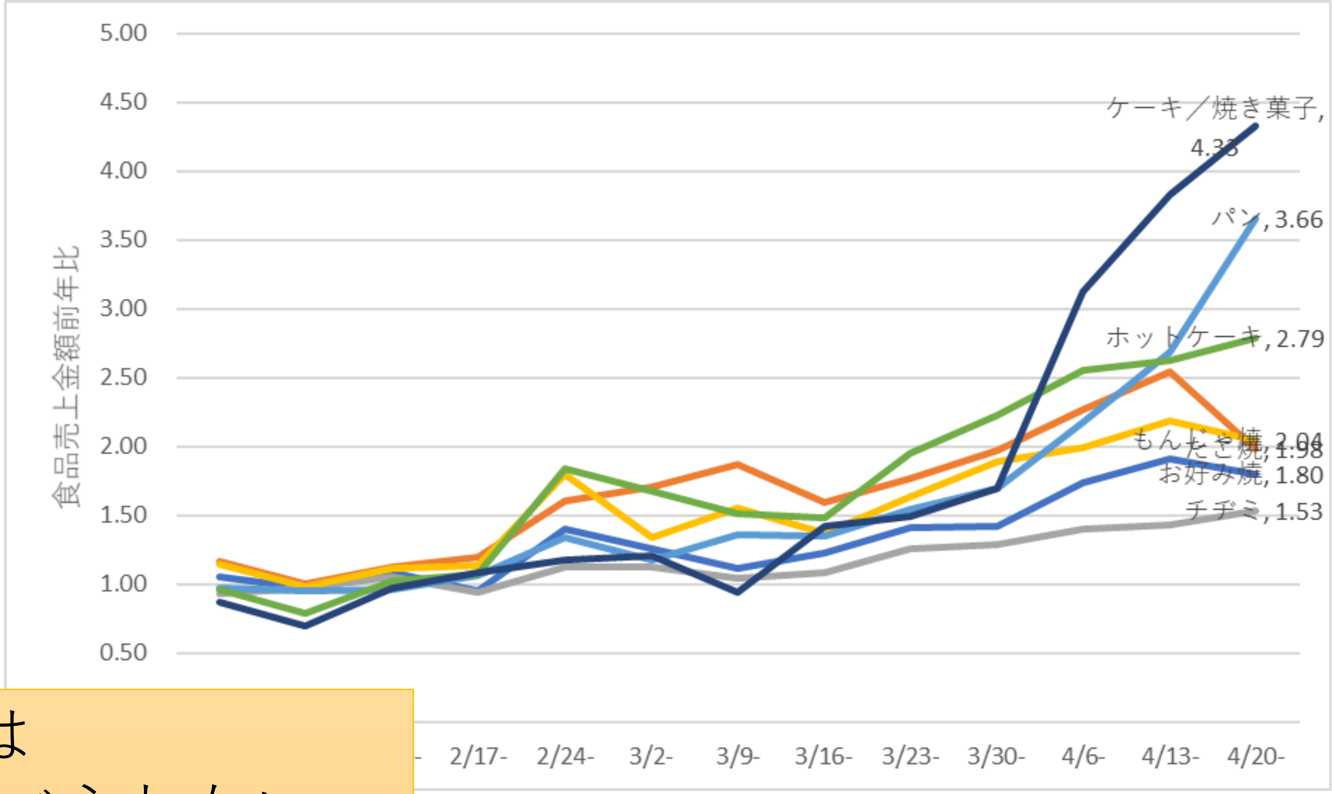
	1/27–	2/3–	2/10–	2/17–	2/24–	3/2–	3/9–	3/16–	3/23–	3/30–	4/6–	4/13–	4/20–
お好み焼	106%	97%	109%	95%	140%	126%	112%	123%	141%	142%	174%	191%	180%
たこ焼	117%	100%	113%	120%	161%	171%	187%	160%	177%	197%	227%	254%	198%
チヂミ	93%	97%	106%	94%	113%	113%	105%	109%	126%	129%	140%	143%	153%
もんじゃ焼	115%	98%	112%	114%	180%	134%	156%	137%	164%	189%	199%	219%	204%
パン	97%	95%	96%	107%	134%	118%	136%	135%	155%	170%	218%	269%	366%
ホットケーキ	96%	79%	102%	108%	184%	168%	151%	148%	195%	223%	256%	263%	279%
ケーキ/焼菓子	87%	70%	97%	109%	118%	121%	94%	142%	149%	170%	313%	383%	433%

▲KSP-POS食品SM（全国、週次2020年1月27日週～4月20日週）

- データ引用：識学総研, **「ホットケーキミックスはなぜ品薄になった？　コロナ禍での商機を支えた「攻め」のマーケティングとは」，2020年6月1日，** https://souken.shikigaku.jp/4258/
- オリジナル出典（図１-図表３、表１）：[2] KSP-SP（2020）「KSP-POSマーケットトレンドレポートvol.131.5：新型コロナウイルスの影響（３）」（2020年４月30日発行）　図１:p.1、表１:p.2、図２:p.3、図表３:p.4 https://www.ksp-sp.com/open_data/topics/2020/0430.pdf

品目による売上の違いはあるか？
日次による売上の違いはあるか？

- ケーキ⇒パン⇒ホットケーキの順で売上高大きい
- 日時が進むと増加傾向
- 因子が2種類なので
- **繰り返しの無い2元配列分散分析**



繰り返しの無い2元配列分散分析では
交互作用の無いことを仮説検定で調べられない.
よって，交互作用は無いと仮定できるデータに
対してのみ行うこと.

帰無仮説は２種類

$$y_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ij}$$

- 要因1：　品目による売上の違いはない
  $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \alpha_5 = \alpha_6 = \alpha_7$
- 要因2：　日次による売上の違いはない
  $\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7$

要因1：　品目による売上の母平均の違いはないか？
結果は帰無仮説棄却された

概要	データの個数	合計	平均	分散
お好み焼	13	17.36	1.335385	0.10066
たこ焼	13	21.82	1.678462	0.214464
チヂミ	13	15.21	1.17	0.038267
もんじゃ焼	13	20.21	1.554615	0.161977
パン	13	20.96	1.612308	0.637719
ホットケーキ	13	22.52	1.732308	0.454119
ケーキ／焼き菓子	13	22.86	1.758462	1.438231
1/27-	7	7.11	1.015714	0.012929
2/3-	7	6.36	0.908571	0.013381
2/10-	7	7.35	1.05	0.004733
2/17-	7	7.47	1.067143	0.008924
2/24-	7	10.3	1.471429	0.081148
3/2-	7	9.51	1.358571	0.057181
3/9-	7	9.41	1.344286	0.108495
3/16-	7	9.54	1.362857	0.02759
3/23-	7	11.07	1.581429	0.052814
3/30-	7	12.2	1.742857	0.103924
4/6-	7	15.27	2.181429	0.315514
4/13-	7	17.22	2.46	0.5669
4/20-	7	18.13	2.59	1.1058

分散分析表

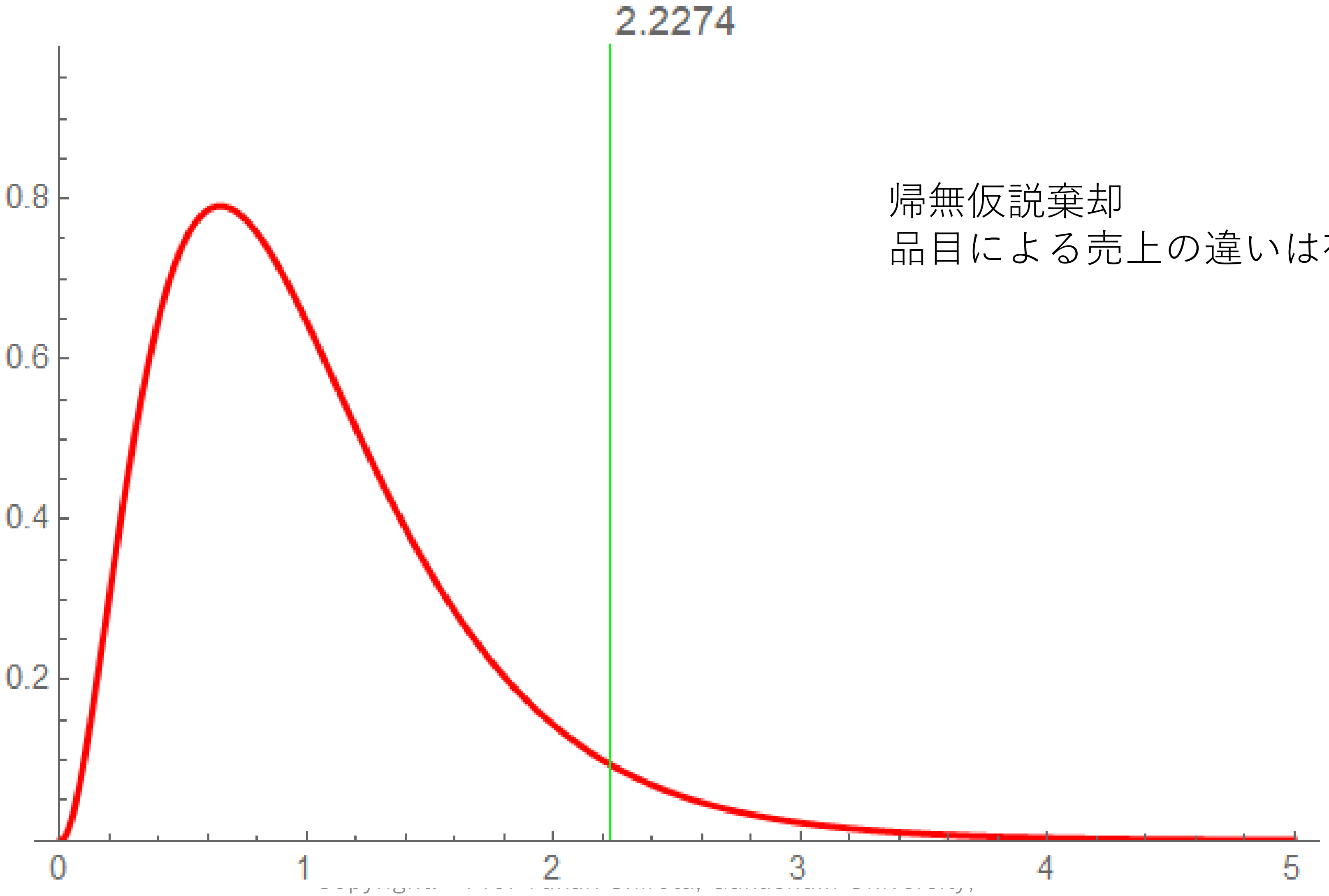
変動要因	変動	自由度	分散	観測された分散比	P-値	F 境界値
行	3.738121	6	0.62302	4.071333	0.001431496	2.227404
列	25.52737	12	2.127281	13.90142	2.37108E-14	1.889242
誤差	11.01788	72	0.153026			
合計	40.28337	90				

品目に関する自由度：　７－１＝６
誤差に関する自由度：　（１３－１）×（７－１）＝７２

F(６，７２）を使う．
５％の境界値は　2.2

後述のように，等分散性の検定が言えなかったので，結局このようなパラメトリック検定（1元配置分散分析など）は参考程度にみておくこと

F(６，　７２）を使う.
５％の境界値は　2.2



帰無仮説棄却
品目による売上の違いは存在する

等分散の仮説検定

概要	データの個数	合計	平均	分散
お好み焼	13	17.36	1.335385	0.10066
たこ焼	13	21.82	1.678462	0.214464
チヂミ	13	15.21	1.17	0.038267
もんじゃ焼	13	20.21	1.554615	0.161977
パン	13	20.96	1.612308	0.637719
ホットケーキ	13	22.52	1.732308	0.454119
ケーキ／焼き菓子	13	22.86	1.758462	1.438231

	Statistic	P-Value
Conover	14.2919	0.0265401
Bartlett	46.1634	2.74689×10^{-8}
Levene	5.68716	0.000053527

等分散性の検定を有意水準５％で行うと，いずれの方法でも，帰無仮説は棄却された．つまり，等分散性は言えなかった．

「等分散性の検定もサンプルサイズが小さいとあてにあてにならないので参考程度と考えたほうがよい」というものの，これだけ等分散であるという帰無仮説が強くリジェクトされると，7種類について2元配列分散分析を行うことは無理と考えた．ANOVAの結果は参考程度にみておくこと．

繰り返しの無い2元配列分散分析では
交互作用と実験誤差の大きさを分離して把握できない

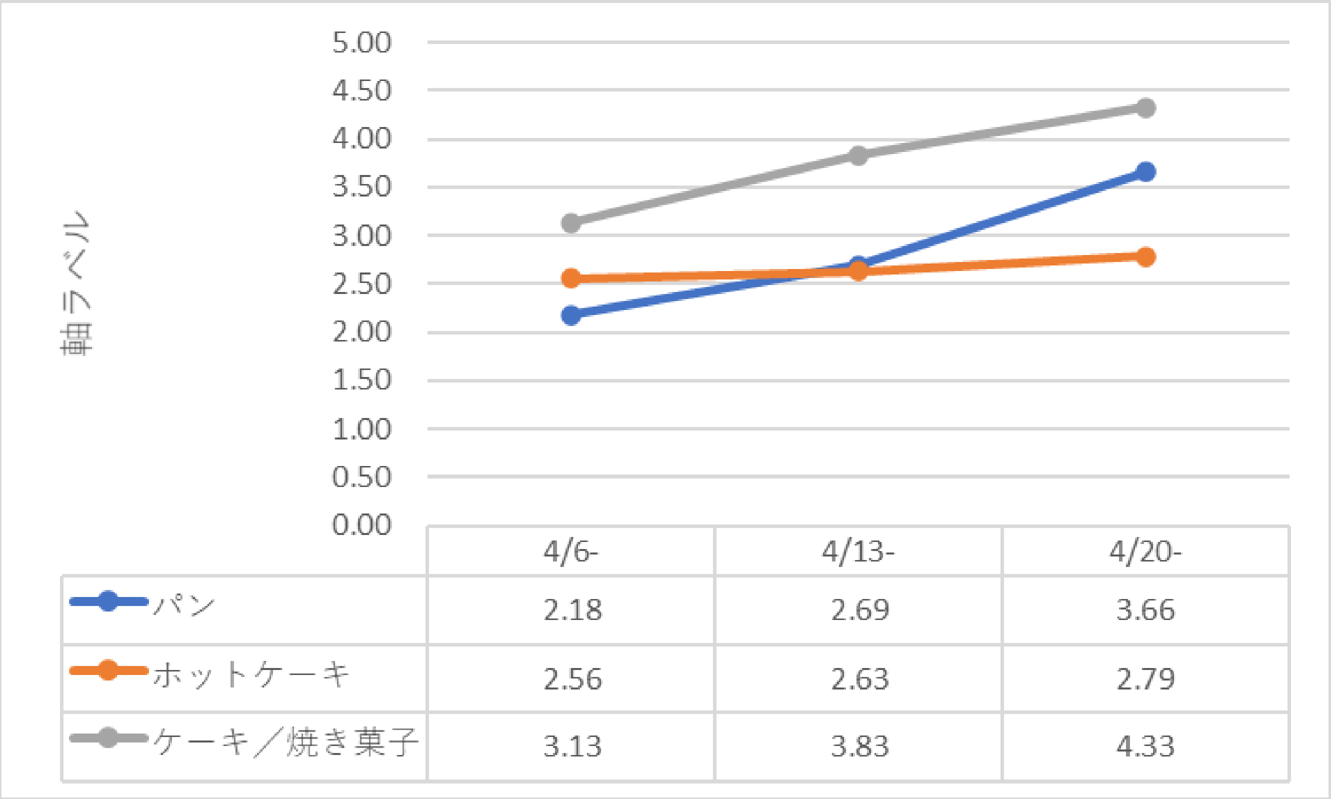
- 従来の知見などで，交互作用が存在しないことが分かっている場合にのみ，繰り返し無し2元配列分散分析を使うことができる(交互作用が存在する場合，結果の信頼性が低い)
- どうして，分離できないのか？
  - 繰り返しのない2元配列のデータの構造モデルは
  - $y_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ij}$
  - 交互作用と誤差が同じ添え字をもつので，誤差Seには両方の大きさの情報が交じり合って含まれるので，どちらが効いているのか分からないため.

それでも，何とか分析をしたい．分析対象を絞り込む（自分は何が発見したいのか？）
4月7日の緊急事態宣言以降で調査
ケーキ，パン，ホットケーキのみで分散分析

- 特別措置法第３２条に基づき、緊急事態宣言を発出することといたします。対象となる範囲は、関東の１都３県、東京都、神奈川県、千葉県、埼玉県、関西の大阪府と兵庫県、そして九州の福岡県であります。

- 引用
  首相官邸：**新型コロナウイルス感染症に関する安倍内閣総理大臣記者会見，2020年4月7日,**
  https://www.kantei.go.jp/jp/98_abe/statement/2020/0407kaiken.html

3週間のみ，3品目に限定して
2元配置ANOVA繰り返し無し



ホットケーキが横ばいは
品切れか？

従来の知見などで，交互作用が存在しないことが分かっている場合にのみ，
繰り返し無し2元配列分散分析を使うことができる(交互作用が存在する場合，
結果の信頼性が低い)

分散分析: 繰り返しのない二元配置				
概要	データの個数	合計	平均	分散
パン	3	8.53	2.843333	0.565233
ホットケーキ	3	7.98	2.66	0.0139
ケーキ／焼き菓子	3	11.29	3.763333	0.363333
4/6-	3	7.87	2.623333	0.228633
4/13-	3	9.15	3.05	0.4572
4/20-	3	10.78	3.593333	0.596233

等分散性の検定
帰無仮説は棄却されない
等分散性は否定されなかった

{0.565233, 0.0139, 0.363333}

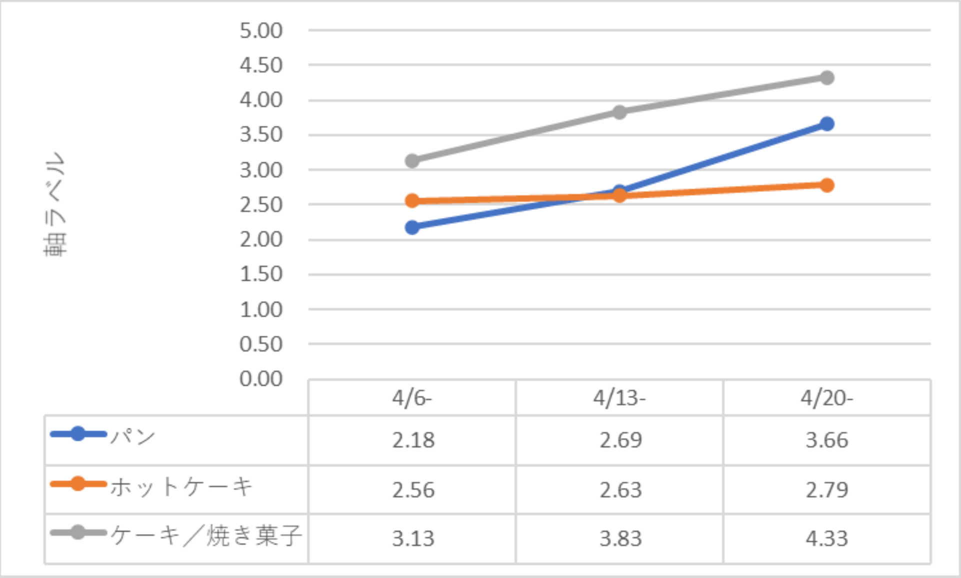
	Statistic	P-Value
Bartlett	3.90308	0.142055
Brown-Forsythe	1.15714	0.375821
Conover	1.71283	0.424681
Levene	2.30716	0.180625

{0.228633, 0.4572, 0.596233}

	Statistic	P-Value
Bartlett	0.368143	0.831876
Brown-Forsythe	0.124668	0.885018
Conover	0.592572	0.743575
Levene	0.395659	0.689591

多重比較　Tukey法,
有意水準7%

主因子１：品目
5：パン
6：ホットケーキ
7：ケーキ




```
In[•]:= ANOVA[twowaydata, {factor1, factor2}, {factor1, factor2}, PostTests → {Tukey},
         SignificanceLevel → .07, CellMeans → False]
         有意水準                               偽
```

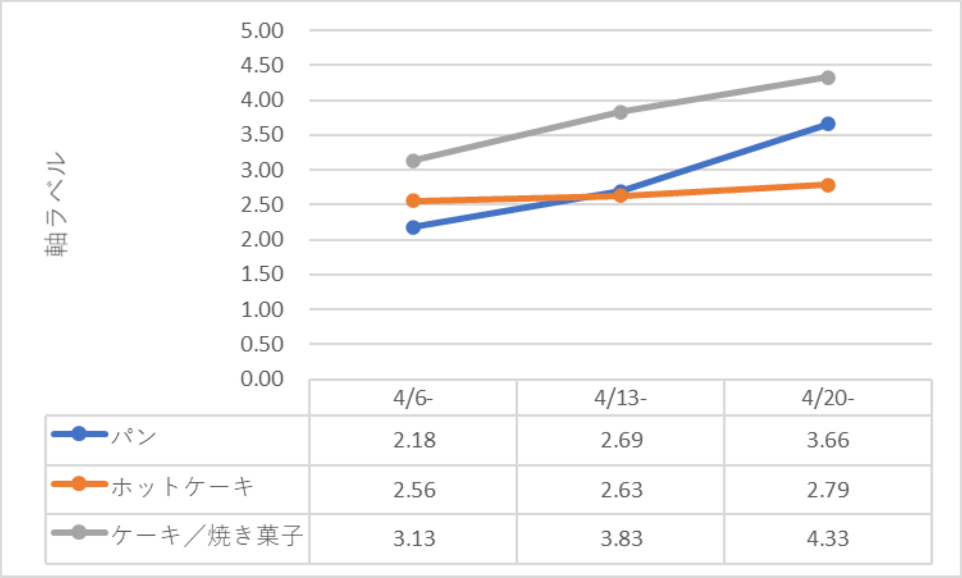
Out[•]= {ANOVA →

	DF	SumOfSq	MeanSq	FRatio	PValue
factor1	2	2.09736	1.04868	8.98653	0.033139
factor2	2	1.41816	0.709078	6.07636	0.0613237,
Error	4	0.466778	0.116694		
Total	8	3.98229			

PostTests → {factor1 → Tukey {{5, 7}, {6, 7}}, factor2 → Tukey {11, 13}}}

品目に関してのみ
帰無仮説棄却

多重比較　Tukey法，有意水準7%




```
factor1 → Tukey {{5, 7}, {6, 7}}, factor2 → Tukey {11, 13}}
```

主因子１：品目に関して
5：パンは，7：ケーキに対して母平均に有意差がある
6：ホットケーキは，7：ケーキに対して母平均に有意差がある

主因子２：期間に関して
１１：4月6日の週は，13：4月20日の週に対して母平均に有意差がある

多重比較　Bonferroni法, 有意水準7%

```
In[∘]:= ANOVA[twowaydata, {factor1, factor2}, {factor1, factor2}, PostTests → {Bonferroni},
          SignificanceLevel → .07, CellMeans → False]
          有意水準                                          偽
```

Out[∘]= {ANOVA →

	DF	SumOfSq	MeanSq	FRatio	PValue
factor1	2	2.09736	1.04868	8.98653	0.033139
factor2	2	1.41816	0.709078	6.07636	0.0613237,
Error	4	0.466778	0.116694		
Total	8	3.98229			

PostTests → {factor1 → Bonferroni $\boxed{\{6, 7\}}$, factor2 → Bonferroni {}}}

Bonferroni法の方が，Tukey法に比較して有意差がでにくい．有意水準の条件が厳しくなるため．