

# 分散分析

キーワード

$$\text{平方和} = \sum (x - \bar{x})^2$$

$$\text{分散} V = \sum (x - \bar{x})^2 / n$$

# 分散分析とは？

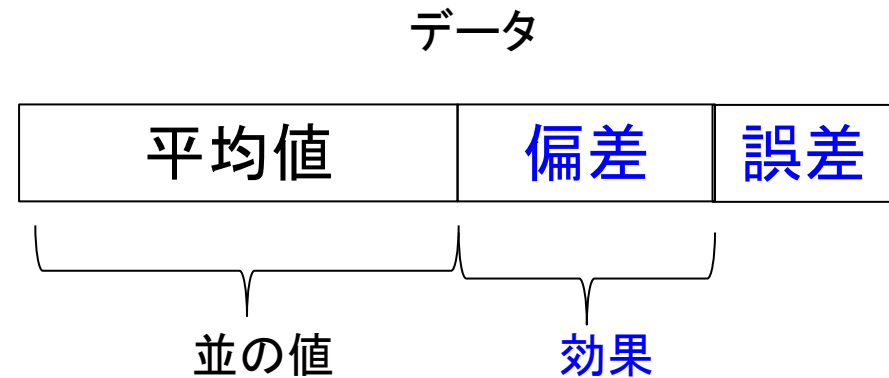
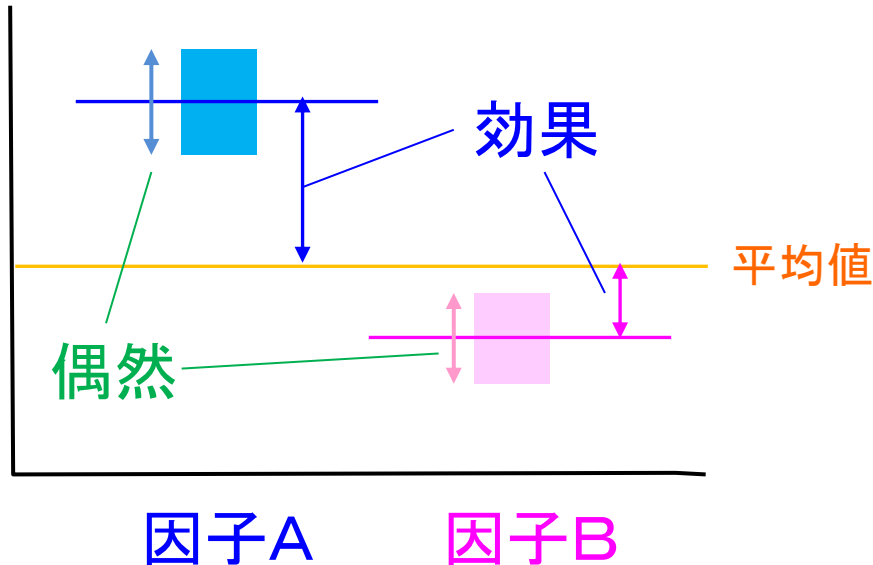
「因子の**効果**なのか」と「**偶然**なのか」を比較して判定する手法

誤差

分散(ばらつき)を算出



F検定



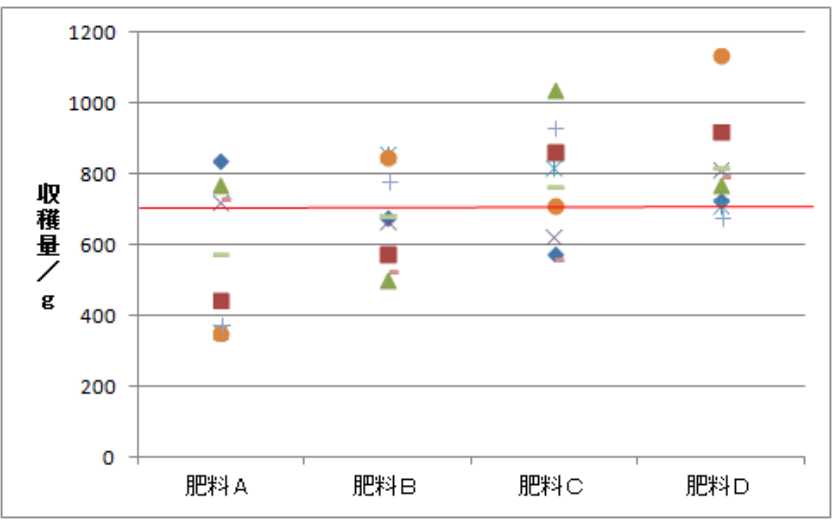
一元配置分散分析(繰返しあり)

事例1

肥料A、B、C及びDを使用した際の収穫量について8個のデータを得た  
肥料の違いが収穫量に効果があるか否かを検定する。

収穫量のデータ 単位: g

	肥料A	肥料B	肥料C	肥料D	
1	840	679	579	730	
2	444	574	865	920	
3	771	502	1038	771	
4	723	671	625	814	
5	357	857	822	714	
6	353	849	715	1135	
7	377	780	932	677	
8	731	529	562	797	
平均	575	680	767	820	全平均 710

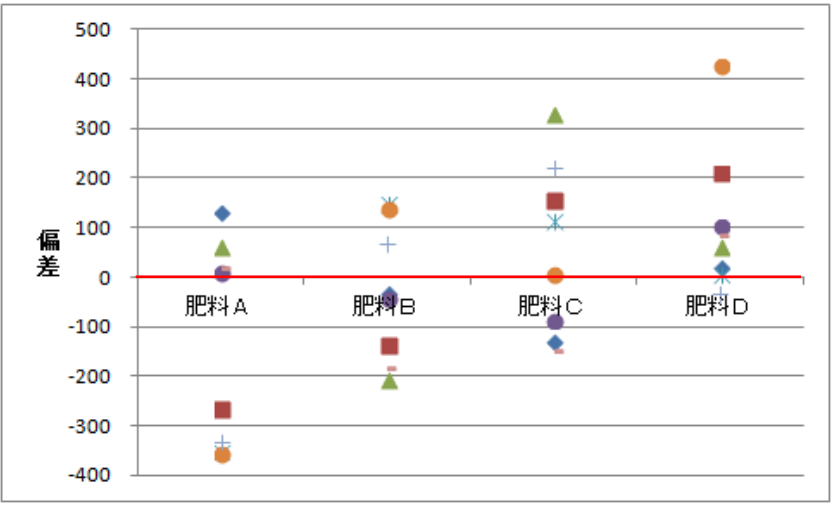


効果は偏差で表される



偏差 = 各データ - 全平均

	肥料A	肥料B	肥料C	肥料D
1	130	-31	-131	20
2	-266	-136	155	210
3	61	-208	328	61
4	13	-39	-85	104
5	-353	147	112	4
6	-357	139	5	425
7	-333	70	222	-33
8	21	-181	-148	87



収穫量のデータ 単位: g

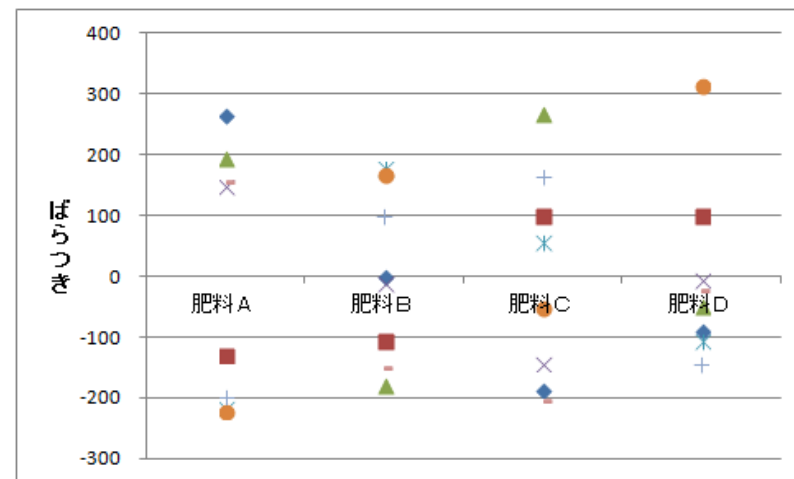
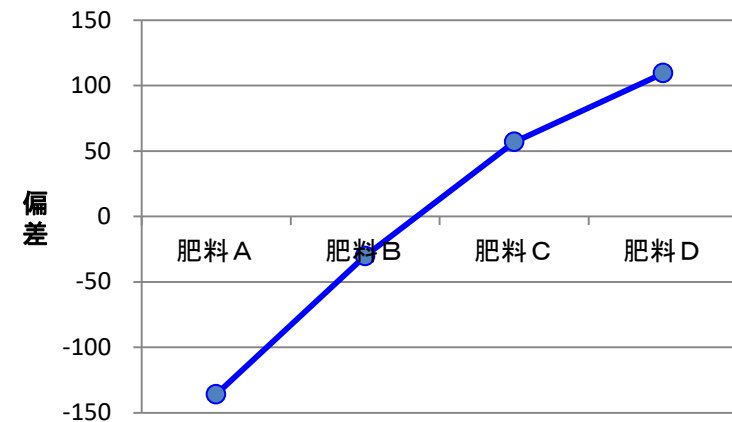
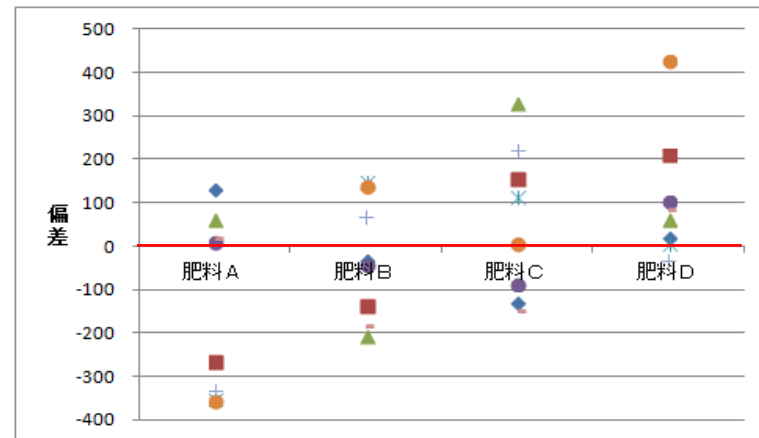
	肥料A	肥料B	肥料C	肥料D	
1	840	679	579	730	
2	444	574	865	920	
3	771	502	1038	771	
4	723	671	625	814	
5	357	857	822	714	
6	353	849	715	1135	
7	377	780	932	677	
8	731	529	562	797	
平均	575	680	767	820	全平均 710

肥料間偏差 = 各肥料データの平均 - 全平均

	肥料A	肥料B	肥料C	肥料D	
1	-136	-30	57	109	
2	-136	-30	57	109	
3	-136	-30	57	109	
4	-136	-30	57	109	
5	-136	-30	57	109	
6	-136	-30	57	109	
7	-136	-30	57	109	
8	-136	-30	57	109	肥料間偏差の平方和
平方和	147764	7336	25850	95648	276598

肥料内ばらつき = 各肥料データ - 各肥料データの平均

	肥料A	肥料B	肥料C	肥料D	
1	266	-1	-188	-90	
2	-131	-106	98	100	
3	197	-178	271	-49	
4	149	-9	-142	-6	
5	-218	177	55	-106	
6	-222	169	-52	315	
7	-198	100	165	-143	
8	157	-151	-205	-23	肥料内偏差の平方和
平方和	308052	135693	213532	151976	809252



	肥料A	肥料B	肥料C	肥料D	平方和	自由度	分散
肥料間偏差	147764	7336	25850	95648	276598	3	92199
肥料内ばらつき	308052	135693	213532	151976	809252	28	28902

肥料4種 データ8個  
 $= 4 - 1 = 3$   
 $= (8 - 1) \times 4 = 28$

	群間の自由度				
群内の自由度	1	2	3	4	5
10	4.96	4.10	3.71	3.48	3.33
20	4.35	3.49	3.10	2.87	2.71
28	4.20	3.34	2.95	2.71	2.56
30	4.17	3.32	2.92	2.69	2.53
40	4.08	3.23	2.84	2.61	2.45
50	4.03	3.18	2.79	2.56	2.40
57	4.01	3.16	2.77	2.53	2.38
60	4.00	3.15	2.76	2.53	2.37
70	3.98	3.13	2.74	2.50	2.35
80	3.96	3.11	2.72	2.49	2.33
90	3.95	3.10	2.71	2.47	2.32
100	3.94	3.09	2.70	2.46	2.31

$F = 92199 / 28902 = 3.19$

$2.95 < 3.19$

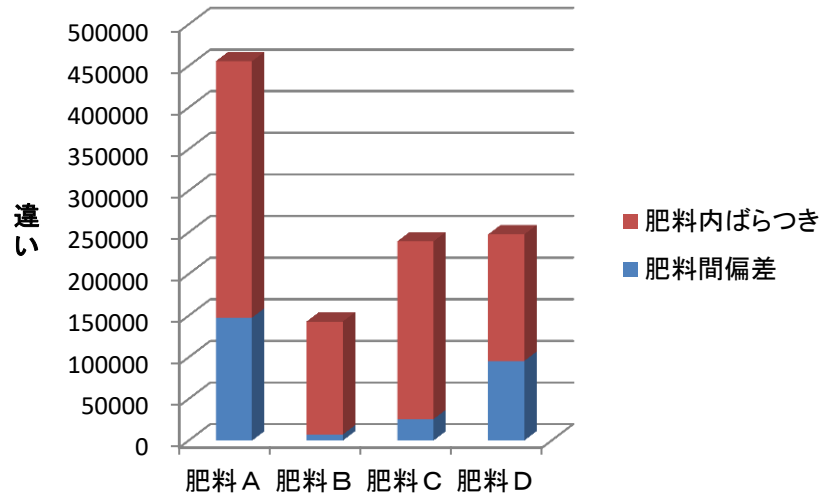


肥料の違い効果あり

Excelの結果

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	276597.84	3	92199.28	3.190082043	0.038891807	2.946685
グループ内	809251.88	28	28901.85			
合計	1085849.7	31				

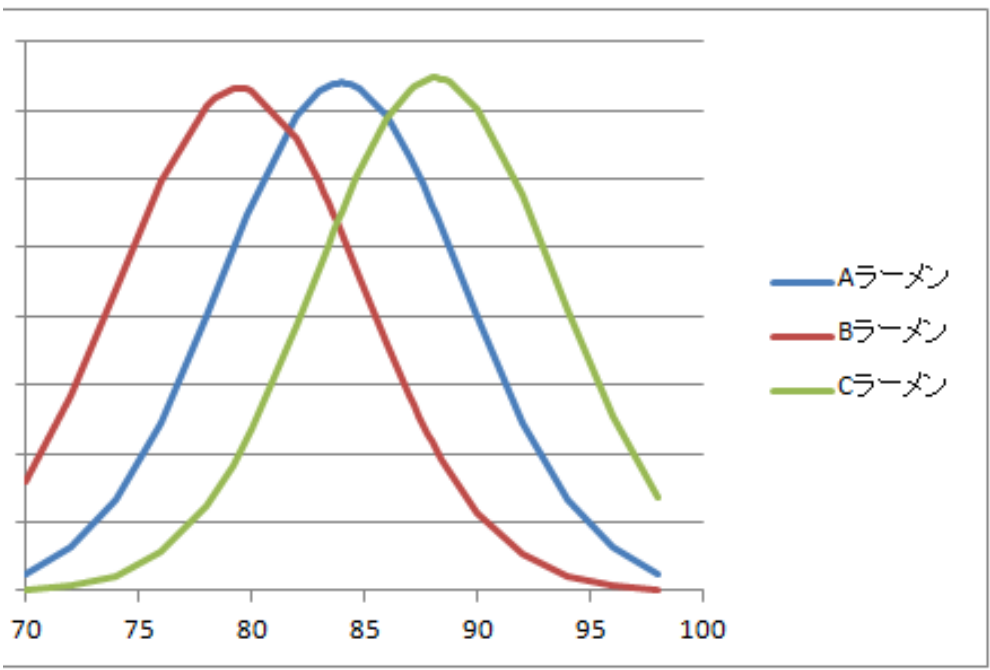


# 一元配置分散分析(繰返しあり)

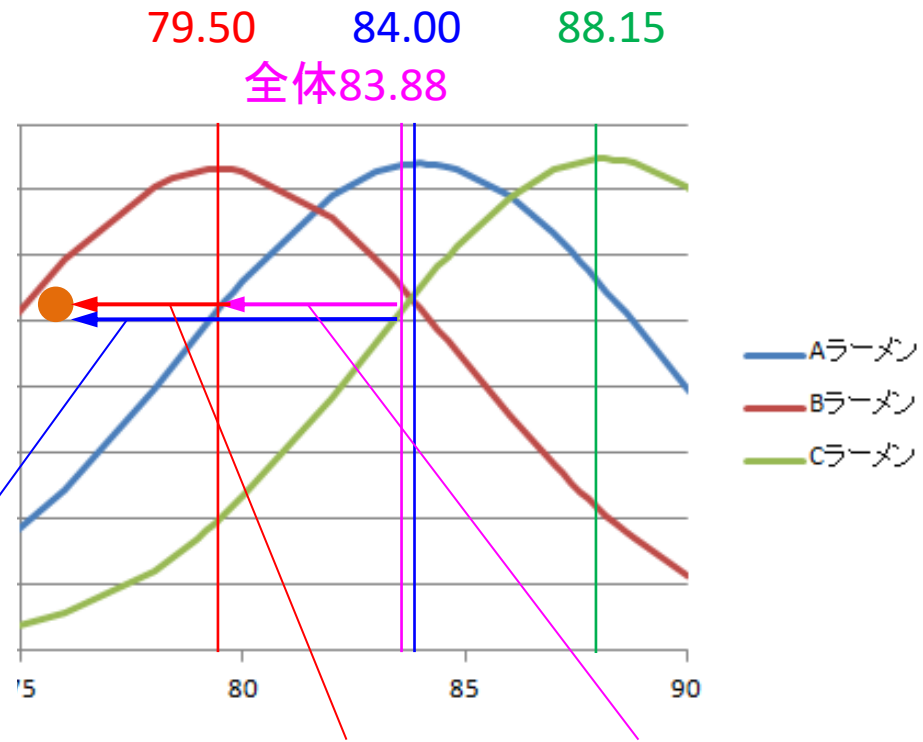
## 事例2

A、B及びCラーメン店の味について20名に100点満点で評価してもらったデータ

	Aラーメン	Bラーメン	Cラーメン
1	80	75	80
2	75	70	80
3	80	80	80
4	75	70	98
5	90	85	95
6	80	75	85
7	80	75	90
8	90	80	95
9	80	75	85
10	85	80	85
11	85	75	85
12	95	90	95
13	90	80	90
14	90	80	90
15	85	90	85
16	85	80	90
17	80	80	85
18	85	80	85
19	90	85	90
20	80	85	95
サンプルサイズ	20	20	20
平均	84.00	79.50	88.15
標準偏差	5.39	5.45	5.34



**帰無仮説:**  
3つの店の評価の平均に差はない



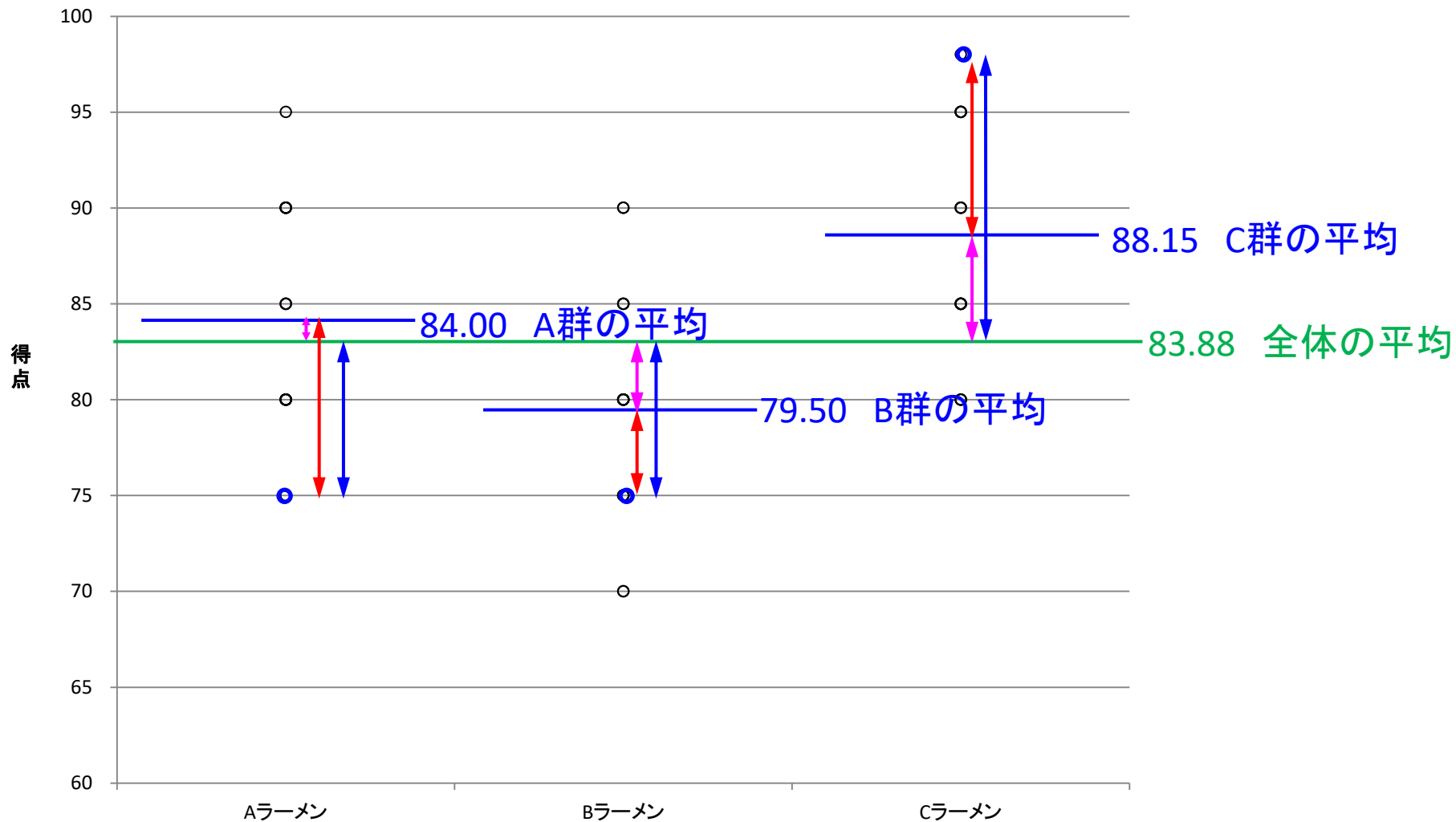
全体の平均からのズレ = 群内のズレ + 群間のズレ



全体の平方和 = 群内の平方和 + 群間の平方和

誤差                      効果





全体の平方和 = 群内の平方和 + 群間の平方和

誤差                      効果

要因	平方和	自由度	平均平方和	F
効果→ 群間	748.63	2	374.32	12.22
誤差→ 群内	1745.55	57	30.62	
全体	2494.18	59		

群間の自由度 = 3 - 1 = 2

群内の自由度 = (20 - 1) + (20 - 1) + (20 - 1) = 57

全体の自由度 = 60 - 1 = 59

$$\frac{\text{効果}}{\text{誤差}} = \frac{374.32}{30.62} = 12.22$$

F = 群間の平均平方和 / 群内の平均平方和

大小関係と比較する → 帰無仮説棄却

5%の有意水準

12.22 > 3.16 → 少なくとも1つの組み合わせには差がある

群内の自由度	群間の自由度				
	1	2	3	4	5
10	4.96	4.10	3.71	3.48	3.33
20	4.35	3.49	3.10	2.87	2.71
30	4.17	3.32	2.92	2.69	2.53
40	4.08	3.23	2.84	2.61	2.45
50	4.03	3.18	2.79	2.56	2.40
57	4.01	3.16	2.77	2.53	2.38
60	4.00	3.15	2.76	2.53	2.37
70	3.98	3.13	2.74	2.50	2.35
80	3.96	3.11	2.72	2.49	2.33
90	3.95	3.10	2.71	2.47	2.32
100	3.94	3.09	2.70	2.46	2.31
200	3.89	3.04	2.65	2.42	2.26
300	3.87	3.03	2.63	2.40	2.24

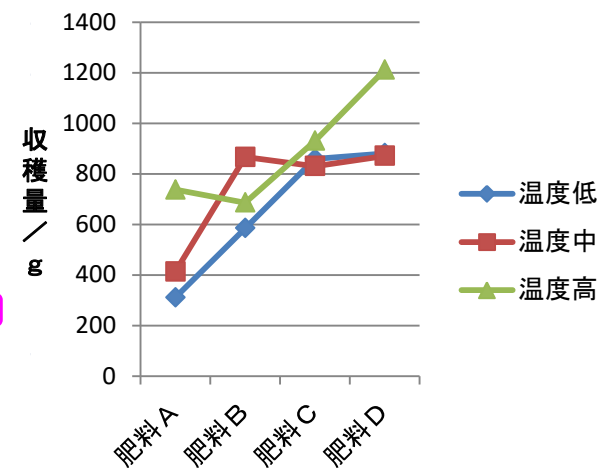
# 二元配置の分散分析(繰り返しなし)

## 事例3

肥料A、B、C及びDを使用し、ハウスの温度を低、中及び高で栽培した際の収穫量を計測した。この結果は、肥料の効果があるか、温度の効果があるかについて検定したい。

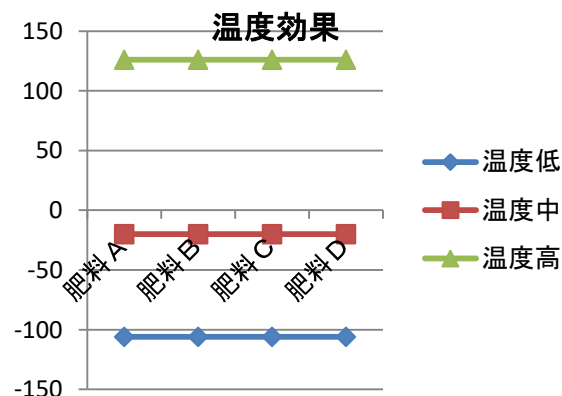
収穫量のデータ 単位: g

	肥料A	肥料B	肥料C	肥料D	温度の水準平均
温度低	311	585	859	881	659
温度中	413	866	830	871	745
温度高	737	686	930	1211	891
肥料の水準平均	487	712	873	988	765



温度の効果 = 温度の水準平均 - 全平均

	肥料A	肥料B	肥料C	肥料D	水準間変動
温度低	-106	-106	-106	-106	110048
温度中	-20	-20	-20	-20	
温度高	126	126	126	126	
平方和	27512	27512	27512	27512	

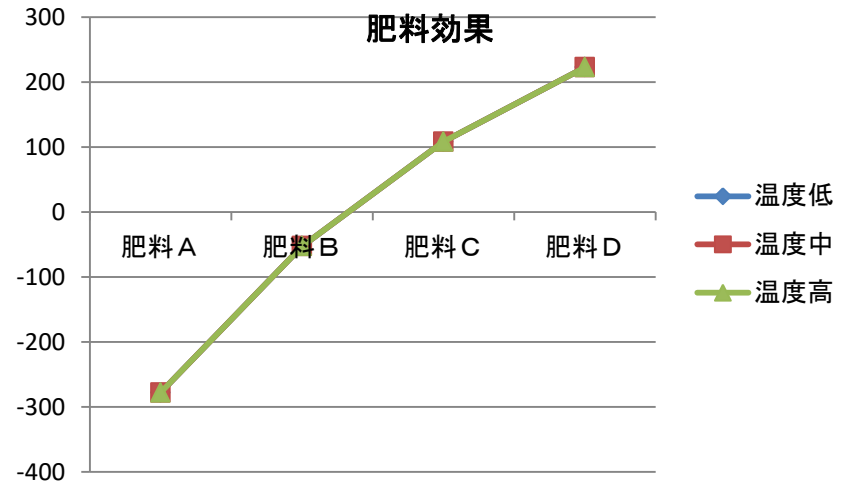


## 収穫量のデータ 単位: g

	肥料A	肥料B	肥料C	肥料D	温度の水準平均
温度低	311	585	859	881	659
温度中	413	866	830	871	745
温度高	737	686	930	1211	891
肥料の水準平均	487	712	873	988	765

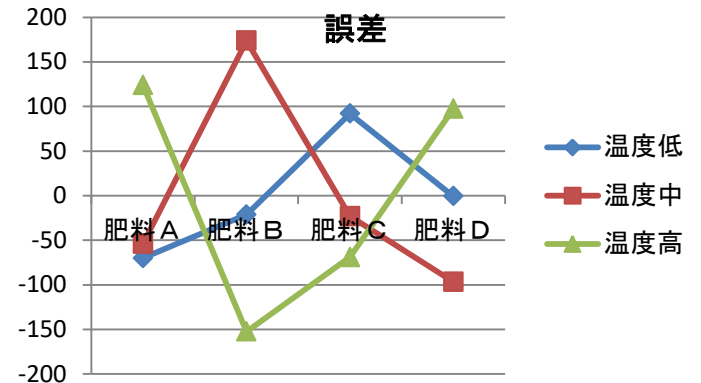
## 肥料の効果 = 肥料の水準平均 - 全平均

	肥料A	肥料B	肥料C	肥料D	
温度低	-278	-53	108	223	141302
温度中	-278	-53	108	223	141302
温度高	-278	-53	108	223	141302
	水準間変動				423907



収穫量のデータ 単位: g

	肥料A	肥料B	肥料C	肥料D	温度の水準平均
温度低	311	585	859	881	659
温度中	413	866	830	871	745
温度高	737	686	930	1211	891
肥料の水準平均	487	712	873	988	765



温度の効果 = 温度の水準平均 - 全平均

	肥料A	肥料B	肥料C	肥料D	水準間変動
温度低	-106	-106	-106	-106	
温度中	-20	-20	-20	-20	
温度高	126	126	126	126	
平方和	27512	27512	27512	27512	110048

肥料の効果 = 肥料の水準平均 - 全平均

	肥料A	肥料B	肥料C	肥料D	水準間変動
温度低	-278	-53	108	223	141302
温度中	-278	-53	108	223	141302
温度高	-278	-53	108	223	141302
平方和					423907

統計誤差 = データの偏差 - 肥料の効果 - 温度の効果

	肥料A	肥料B	肥料C	肥料D	誤差変動
温度低	-70	-21	92	-1	
温度中	-54	174	-23	-97	
温度高	124	-152	-69	97	
平方和					109585

	平方和	自由度	分散	分散比	5%の棄却値
温度効果変動	110048	2	55024	3.01	5.14
肥料効果変動	423907	3	141302	7.74	4.76
誤差変動	109585	6	18264		

	平方和	自由度	分散	分散比	5%の棄却値
温度効果変動	110048	2	55024	3.01	5.14
肥料効果変動	423907	3	141302	7.74	4.76
誤差変動	109585	6	18264		

$$F_{11} = 55024 / 18264 = 3.01$$

$$F_{12} = 141302 / 18264 = 7.74$$

3.01 < 5.14 → 棄却できない

7.74 > 4.76 → 棄却される

温度効果はあると言えない

肥料の違いによる効果はある

## Excelの結果

### 分散分析表

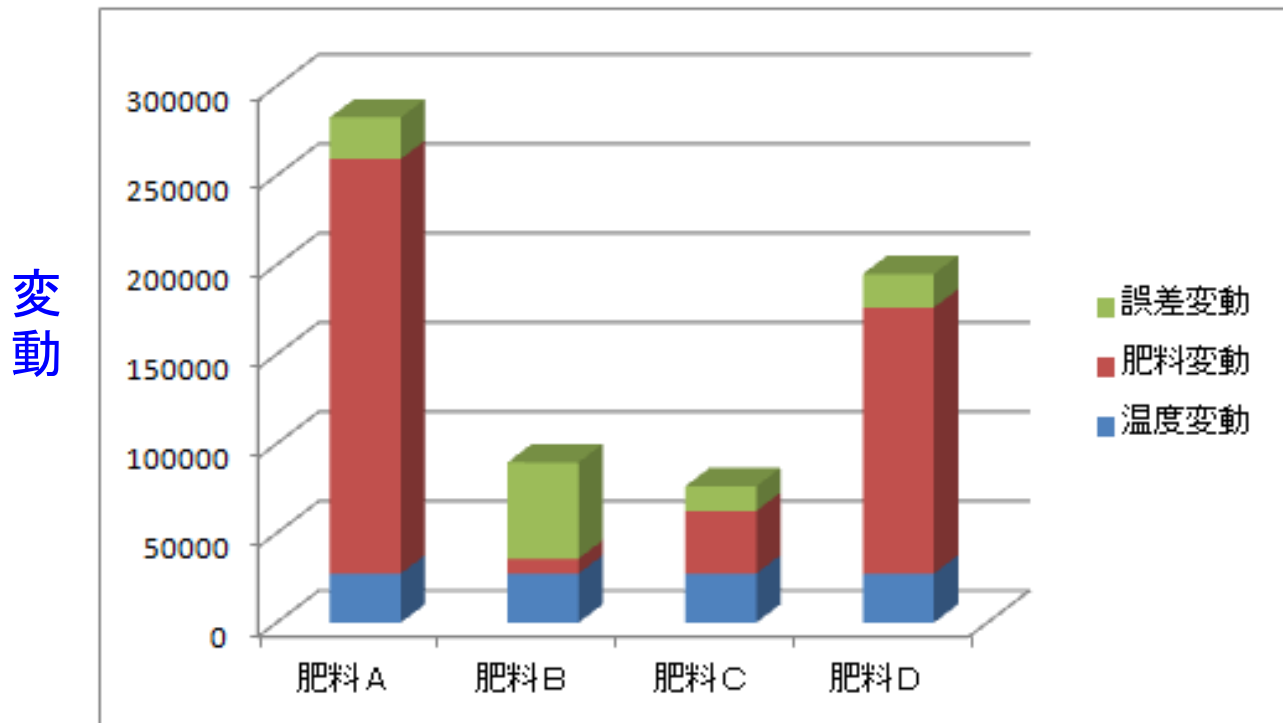
変動要因	変動	自由度	分散	観測された分散比	P-値	F 境界値
行	110048	2	55024	3.012665929	0.12421171	5.143253
列	423906.7	3	141302.2	7.73655842	0.017433072	4.757063
誤差	109585.3	6	18264.22			
合計	643540	11				

	肥料A	肥料B	肥料C	肥料D
温度変動	27512	27512	27512	27512
肥料変動	231852	8321	34992	148741
誤差変動	23192	53821	13754	18819
合計	282556	89654	76258	195072



変動を平方和で計算

分散分析結果は、下図からもイメージできる



# 二元配置の分散分析(繰り返しあり)

## 事例4

喜多方と札幌ラーメンの味噌味と醤油味について15名に100点満点で評価してもらったデータ

→ 地域性と味に関する2つのパラメータがあるため、二元となる

	喜多方ラーメン		札幌ラーメン	
	味噌味	醤油味	味噌味	醤油味
	75	80	85	85
	85	75	80	80
	75	70	75	65
	90	65	75	65
	75	70	60	85
	85	85	85	80
	85	75	70	75
	80	70	65	70
	85	80	80	80
	85	70	65	70
	80	60	65	75
	90	75	75	85
	65	60	65	60
	75	65	75	70
	65	65	70	70
サンプルサイズ	15	15	15	15
標本平均	79.67	71.00	72.67	74.33
標準偏差	7.63	7.12	7.50	7.72



要因	平方和	自由度	平均平方和	F
地域 → 要因1	50.42	1	50.42	0.84
→ 要因2	183.75	1	183.75	3.05
交互作用	400.42	1	400.42	6.65
残差	3370.00	56	60.18	
全体	4004.58	59		

= 要因1の平方和 / 残差の平方和

= 要因2の平方和 / 残差の平方和

= 交互作用 / 残差の平方和

大小関係を比較する

→ 要因1と2は、帰無仮説成立 (< 4.01)

→ 地域及び味については点数差なし

→ 交互作用による点数の差はある (> 4.01)

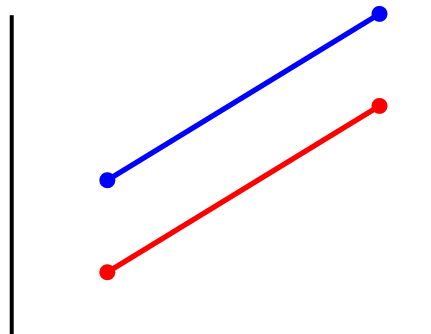
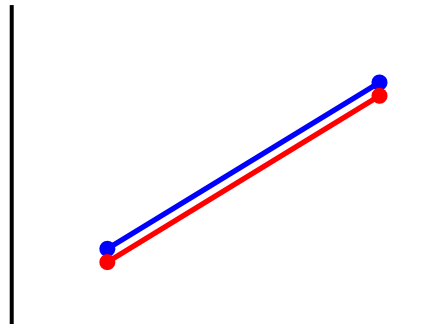
群内の自由度	群間の自由度				
	1	2	3	4	5
10	4.96	4.10	3.71	3.48	3.33
20	4.35	3.49	3.10	2.87	2.71
30	4.17	3.32	2.92	2.69	2.53
40	4.08	3.23	2.84	2.61	2.45
50	4.03	3.18	2.79	2.56	2.40
56	4.01	3.16	2.77	2.54	2.38
60	4.00	3.15	2.76	2.53	2.37
70	3.98	3.13	2.74	2.50	2.35
80	3.96	3.11	2.72	2.49	2.33
90	3.95	3.10	2.71	2.47	2.32
100	3.94	3.09	2.70	2.46	2.31
200	3.89	3.04	2.65	2.42	2.26
300	3.87	3.03	2.63	2.40	2.24

# 交互作用とは？

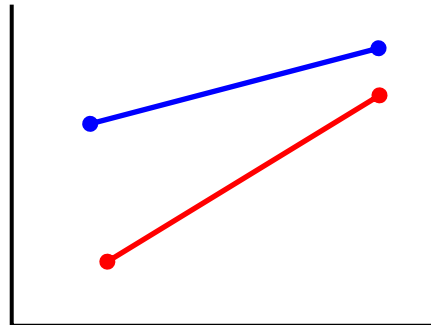
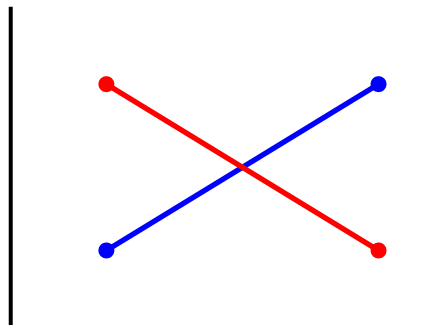
平行



交互作用なし



交互作用あり



	喜多方	札幌
味噌	79.67	72.67
醤油	71	74.33

