Lecture 9: Factorial Design

Montgomery: Chapter 5

Example I: Battery Life Experiment

An engineer is studying the effective life of a certain type of battery. Two factors, plate material and temperature, are involved. There are three types of plate materials (1, 2, 3) and three temperature levels (15, 70, 125). Four batteries are tested at each combination of plate material and temperature, and all 36 tests are run in random order. The experiment and the resulting observed battery life data are given below.

	temperature		
material	15	70	125
1	130,155,74,180	34,40,80,75	20,70,82,58
2	150,188,159,126	136,122,106,115	25,70,58,45
3	138,110,168,160	174,120,150,139	96,104,82,60

Example II: Bottling Experiment

A soft drink bottler is interested in obtaining more uniform fill heights in the bottles produced by his manufacturing process. An experiment is conducted to study three factors of the process, which are

the percent carbonation (A):	10, 12, 14 percent
the operating pressure (B):	25, 30 psi
the line speed (C):	200, 250 bpm

The response is the deviation from the target fill height. Each combination of the three factors has two replicates and all 24 runs are performed in a random order. The experiment and data are shown below.

	pressure(B)			
	25 psi		30 psi	
	LineSpeed(C)		LineSpeed(C)	
Carbonation(A)	200	250	200	250
10	-3,-1	-1,0	-1,0	1, 1
12	0, 1	2,1	2,3	6,5
14	5,4	7,6	7,9	10,11

Factorial Design

- a number of factors: F_1, F_2, . . ., F_r.
- each with a number of levels: l_1, l_2, . . ., l_r
- interested in (main) effects, 2-factor interactions (2fi), 3-factor interactions (3fi), etc.
- number of all possible level combinations (treatments): $l_1 \times l_2 \ldots \times l_r$

One-factor-a-time design as the opposite of factorial design.

Advantages of factorial over one-factor-a-time

- more efficient (runsize and estimation precision)
- able to accommodate interactions
- results are valid over a wider range of experimental conditions

Statistical Model (Two Factors: A and B)

- Statistical model is

$$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \epsilon_{ijk} \quad \begin{cases} i = 1, 2, \ldots, a \\ j = 1, 2, \ldots, b \\ k = 1, 2, \ldots, n \end{cases}$$

μ - grand mean

τ_i - ith level effect of factor A (ignores B) (main effects of A)

β_j - jth level effect of factor B (ignores A) (main effects of B)

$(\tau\beta)_{ij}$ - interaction effect of combination ij (Explain variation not explained by main effects)

$\epsilon_{ijk} \sim \mathrm{N}(0, \sigma^2)$

- Over-parameterized model: must include certain parameter constraints. Typically

$$\textstyle\sum_i \tau_i = 0 \qquad \sum_j \beta_j = 0 \qquad \sum_i (\tau\beta)_{ij} = 0 \qquad \sum_j (\tau\beta)_{ij} = 0$$

Estimates

- Rewrite observation as:

$$y_{ijk} = \overline{y}_{...} + (\overline{y}_{i..} - \overline{y}_{...}) + (\overline{y}_{.j.} - \overline{y}_{...}) + (\overline{y}_{ij.} - \overline{y}_{i..} - \overline{y}_{.j.} + \overline{y}_{...}) + (y_{ijk} - \overline{y}_{ij.})$$

- result in estimates

$$\widehat{\mu} = \overline{y}_{...}$$

$$\widehat{\tau}_i = \overline{y}_{i..} - \overline{y}_{...}$$

$$\widehat{\beta}_j = \overline{y}_{.j.} - \overline{y}_{...}$$

$$\widehat{(\tau\beta)}_{ij} = \overline{y}_{ij.} - \overline{y}_{i..} - \overline{y}_{.j.} + \overline{y}_{...}$$

- predicted value at level combination ij is

$$\widehat{y}_{ijk} = \overline{y}_{ij.}$$

- Residuals are

$$\hat{\epsilon}_{ijk} = y_{ijk} - \overline{y}_{ij.}$$

Partitioning the Sum of Squares

- Based on

$$y_{ijk} = \overline{y}_{...} + (\overline{y}_{i..} - \overline{y}_{...}) + (\overline{y}_{.j.} - \overline{y}_{...}) + (\overline{y}_{ij.} - \overline{y}_{i..} - \overline{y}_{.j.} + \overline{y}_{...}) + (y_{ijk} - \overline{y}_{ij.})$$

- Calculate $\mathrm{SS_T} = \sum (y_{ijk} - \overline{y}_{...})^2$
- Right hand side simplifies to

$$\begin{array}{rcl}
\mathrm{SS_A}: & bn\sum_i (\overline{y}_{i..} - \overline{y}_{...})^2 + & df = a-1 \\
\mathrm{SS_B}: & an\sum_j (\overline{y}_{.j.} - \overline{y}_{...})^2 + & df = b-1 \\
\mathrm{SS_{AB}}: & n\sum_i \sum_j (\overline{y}_{ij.} - \overline{y}_{i.} - \overline{y}_{.j} + \overline{y}_{..})^2 + & df = (a-1)(b-1) \\
\mathrm{SS_E}: & \sum_i \sum_j \sum_k (y_{ijk} - \overline{y}_{ij.})^2 & df = ab(n-1)
\end{array}$$

- $\mathrm{SS_T}$=SS$_\mathrm{A}$ + SS$_\mathrm{B}$ + SS$_\mathrm{AB}$ + SS$_\mathrm{E}$
- Using $\mathrm{SS/df}$ leads to $\mathrm{MS_A}$,$\mathrm{MS_B}$, $\mathrm{MS_{AB}}$ and $\mathrm{MS_E}$.

Testing Hypotheses

1 Main effects of A: $H_0 : \tau_1 = \ldots = \tau_a = 0$ vs H_1 : at least one $\tau_i \neq 0$.

2 Main effects of B: $H_0 : \beta_1 = \ldots = \beta_b = 0$ vs H_1 : at least one $\beta_j \neq 0$.

3 Interaction effects of AB:

$H_0 : (\tau\beta)_{ij} = 0$ for all i, j vs H_1 : at least one $(\tau\beta)_{ij} \neq 0$.

- Expected Mean Squares (EMS)

  E(MS$_{\mathrm{E}}$)=σ^2

  E(MS$_{\mathrm{A}}$) = $\sigma^2 + bn\sum\tau_i^2/(a-1)$

  E(MS$_{\mathrm{B}}$) = $\sigma^2 + an\sum\beta_j^2/(b-1)$

  E(MS$_{\mathrm{AB}}$) = $\sigma^2 + n\sum(\tau\beta)_{ij}^2/(a-1)(b-1)$

- Use F-statistics for testing the hypotheses above:

1: $F_0 = \frac{\mathrm{SS_A}/(a-1)}{\mathrm{SS_E}/(ab(n-1))}$ 2: $F_0 = \frac{\mathrm{SS_B}/(b-1)}{\mathrm{SS_E}/(ab(n-1))}$ 3: $F_0 = \frac{\mathrm{SS_{AB}}/(a-1)(b-1)}{\mathrm{SS_E}/(ab(n-1))}$

Analysis of Variance Table

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0
Factor A	SS_{A}	$a-1$	MS_{A}	$F_0 = \text{MS}_{\text{A}}/\text{MS}_{\text{E}}$
Factor B	SS_{B}	$b-1$	MS_{B}	$F_0 = \text{MS}_{\text{B}}/\text{MS}_{\text{E}}$
Interaction	SS_{AB}	$(a-1)(b-1)$	MS_{AB}	$F_0 = \text{MS}_{\text{AB}}/\text{MS}_{\text{E}}$
Error	SS_{E}	$ab(n-1)$	MS_{E}	
Total	SS_{T}	$abn-1$		

$\text{SS}_{\text{T}} = \sum\sum y_{ijk}^2 - y_{\ldots}^2/abn$; $\text{SS}_{\text{A}} = \frac{1}{bn}\sum y_{i..}^2 - y_{\ldots}^2/abn$

$\text{SS}_{\text{B}} = \frac{1}{an}\sum y_{.j.}^2 - y_{\ldots}^2/abn$; $\text{SS}_{\text{subtotal}} = \frac{1}{n}\sum\sum y_{ij.}^2 - y_{\ldots}^2/abn$

$\text{SS}_{\text{AB}} = \text{SS}_{\text{subtotal}} - \text{SS}_{\text{A}} - \text{SS}_{\text{B}}$; SS_{E}=Subtraction

$\text{df}_{\text{E}} > 0$ only if $n > 1$. When $n = 1$, no SS_E is available so we cannot test the effects. If we can assume that the interactions are negligible ($(\tau\beta)_{ij} = 0$), MS_{AB} becomes a good estimate of σ^2 and it can be used as MS_E. Caution: if the assumption is wrong, then error and interaction are confounded and testing results can go wrong.

Battery Life Example

```
data battery;
  input mat temp life;
  datalines;
1 1 130
1 1 155
1 1 74
: : :
: : :
3 3 104
3 3 82
3 3 60
;

proc glm;
  class mat temp;
  model life=mat temp mat*temp;
  output out=batnew r=res p=pred;
  run; quit;
```

SAS Output

```
Dependent Variable: life

                          Sum of
Source          DF       Squares    Mean Square     F Value     Pr > F
Model            8   59416.22222     7427.02778       11.00     <.0001
Error           27   18230.75000      675.21296
Cor Total       35   77646.97222

R-Square      Coeff Var      Root MSE     life Mean
0.765210       24.62372      25.98486      105.5278

Source          DF     Type I SS     Mean Square   F Value     Pr > F
mat             2    10683.72222      5341.86111    7.91       0.0020
temp            2    39118.72222     19559.36111   28.97       <.0001
mat*temp        4     9613.77778      2403.44444    3.56       0.0186
```

Checking Assumptions

1 Errors are normally distributed

Histogram or QQplot of residuals

2 Constant variance

Residuals vs $\hat{y}_{ij.}$ plot, Residuals vs factor A plot and Residuals vs factor B

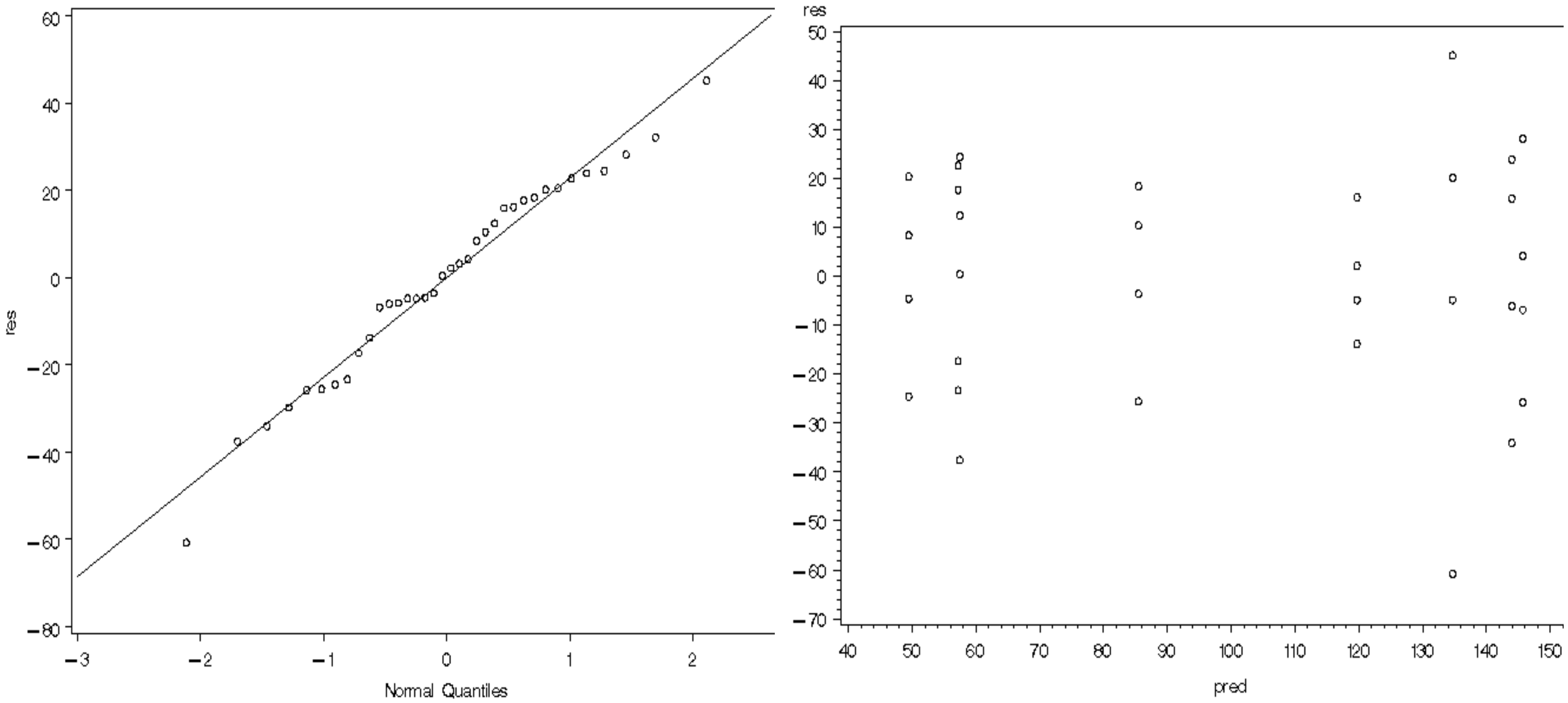
3 If n=1, no interaction.

Tukey's Test of Nonadditivity Assume $(\tau\beta)_{ij} = \gamma\tau_i\beta_j$. $H_0 : \gamma = 0$.

$$\mathrm{SS}_N = \frac{[\sum\sum y_{ij}y_{i.}y_{.j} - y_{..}(\mathrm{SS}_A + \mathrm{SS}_B + y_{..}^2/ab)]^2}{ab\mathrm{SS}_A\mathrm{SS}_B}$$

$$F_0 = \frac{\mathrm{SS}_N/1}{(\mathrm{SS}_E - \mathrm{SS}_N)/((a-1)(b-1)-1)} \sim F_{1,(a-1)(b-a)-1}$$

– the convenient procedure used for RCBD can be employed.


60
40
20
0
−20
−40
−60
−80
res
−3
−2
−1
0
1
2
Normal Quantiles
res
50
40
30
20
10
0
−10
−20
−30
−40
−50
−60
−70
40
50
60
70
80
90
100
110
120
130
140
150
pred

Effects Estimation (Battery Experiment)

0. $\hat{\mu}$=$\bar{y}_{...}$=105.5278

1.Treatment mean response, or cell mean, or predicted value,

$$\hat{y}_{ij} = \hat{\mu}_{ij} = \bar{y}_{ij.} = \hat{\mu} + \hat{\tau}_i + \hat{\beta}_j + \widehat{(\tau\beta)}_{ij}$$

	temperature		
material	1	2	3
1	134.75	57.25	57.50
2	155.75	119.75	49.50
3	144.00	145.75	85.50

2. Factor level means

row means $\bar{y}_{i..}$ for A; column means $\bar{y}_{.j.}$ for B.

material : $\bar{y}_{1..} = 83.166,\ \bar{y}_{2..} = 108.3333,\ \bar{y}_{3..} = 125.0833$

temperature : $\bar{y}_{.1.} = 144.8333,\ \bar{y}_{.2.} = 107.5833,\ \bar{y}_{.3.} = 64.1666$

3. Main effects estimates

$$\hat{\tau}_1 = -22.3612, \hat{\tau}_2 = 2.8055, \hat{\tau}_3 = 19.555$$

$$\hat{\beta}_1 = 39.3055, \hat{\beta}_2 = 2.0555, \hat{\beta}_3 = -41.3611$$

4. Interactions ($(\widehat{\tau\beta})_{ij}$)

	temperature		
material	1	2	3
1	12.2779	-27.9721	15.6946
2	8.1112	9.3612	-17.4722
3	-20.3888	18.6112	1.7779

Understanding Interactions

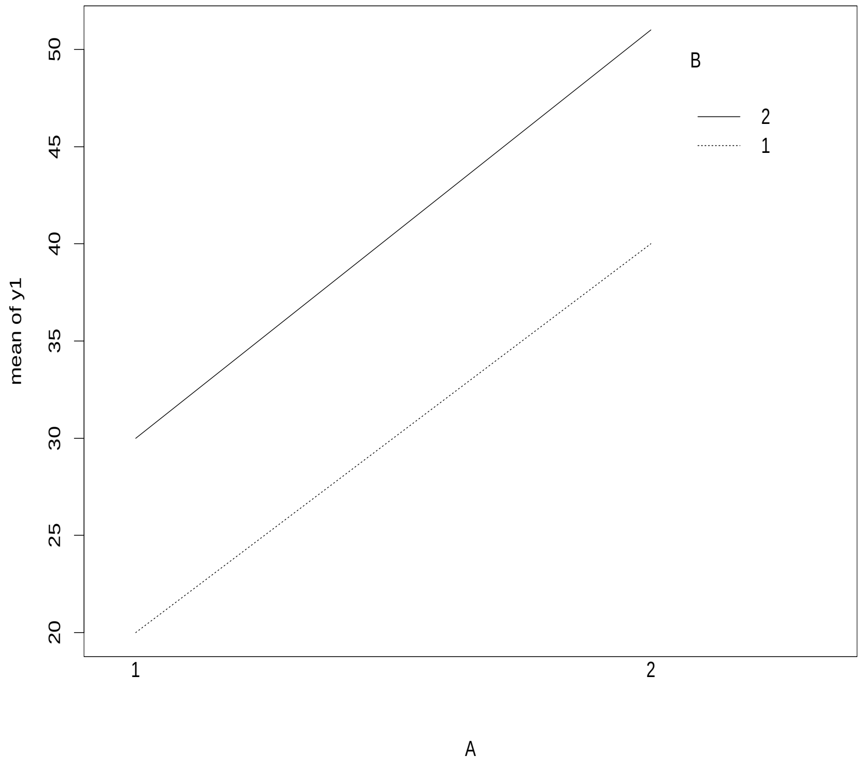
Example I Data 1:

```
A B resp;
1 1 18
1 1 22
1 2 27
1 2 33
2 1 39
2 1 41
2 2 51
2 2 51
```

Dependent Variable: resp

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
A	1	840.5000000	840.5000000	120.07	0.0004
B	1	220.5000000	220.5000000	31.50	0.0050
A*B	1	0.5000000	0.5000000	0.07	0.8025
Error	4	28.000000	7.000000		
Cor Total	7	1089.500000			

Interaction plot for A and B (No Interaction)



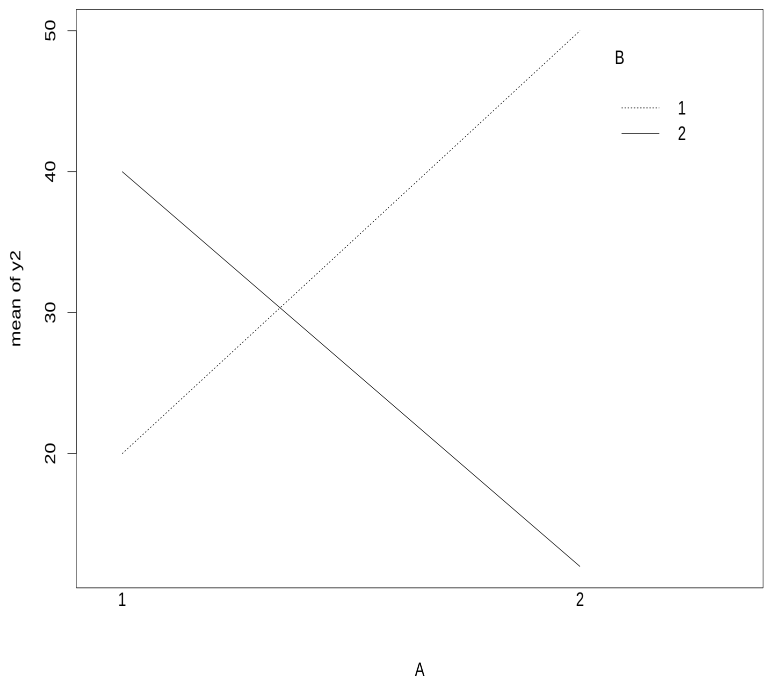
Difference between level means of B (with A fixed at a level) does not depend on the level of A; demonstrated by two parallel lines.

Example I Data 2:

A	B	resp
1	1	19
1	1	21
1	2	38
1	2	42
2	1	53
2	1	47
2	2	10
2	2	14

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
A	1	2.000000	2.000000	0.22	0.6619
B	1	162.000000	162.000000	18.00	0.0132
A*B	1	1682.000000	1682.000000	186.89	0.0002
Error	4	36.000000	9.000000		
Cor Total	7	1882.000000			

Antagonistic Interaction from B to A



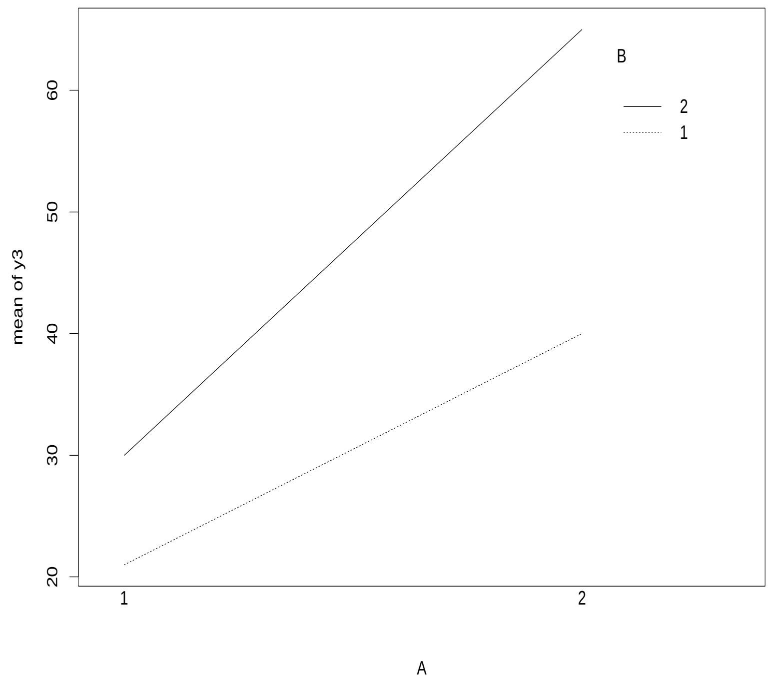
Difference between level means of B (with A fixed at a level) depends on the level of A. If the trend of mean response over A reverses itself when B changes from one level to another, the interaction is said to be antagonistic from B to A. Demonstrated by two lines with slopes of opposite signs.

Example I Data 3:

A	B	resp
1	1	21
1	1	21
1	2	27
1	2	33
2	1	62
2	1	67
2	2	38
2	2	42

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
A	1	1431.125000	1431.125000	148.69	0.0003
B	1	120.125000	120.125000	12.48	0.0242
A*B	1	561.125000	561.125000	58.30	0.0016
Error	4	38.500000	9.625000		
Co Total	7	2150.875000			

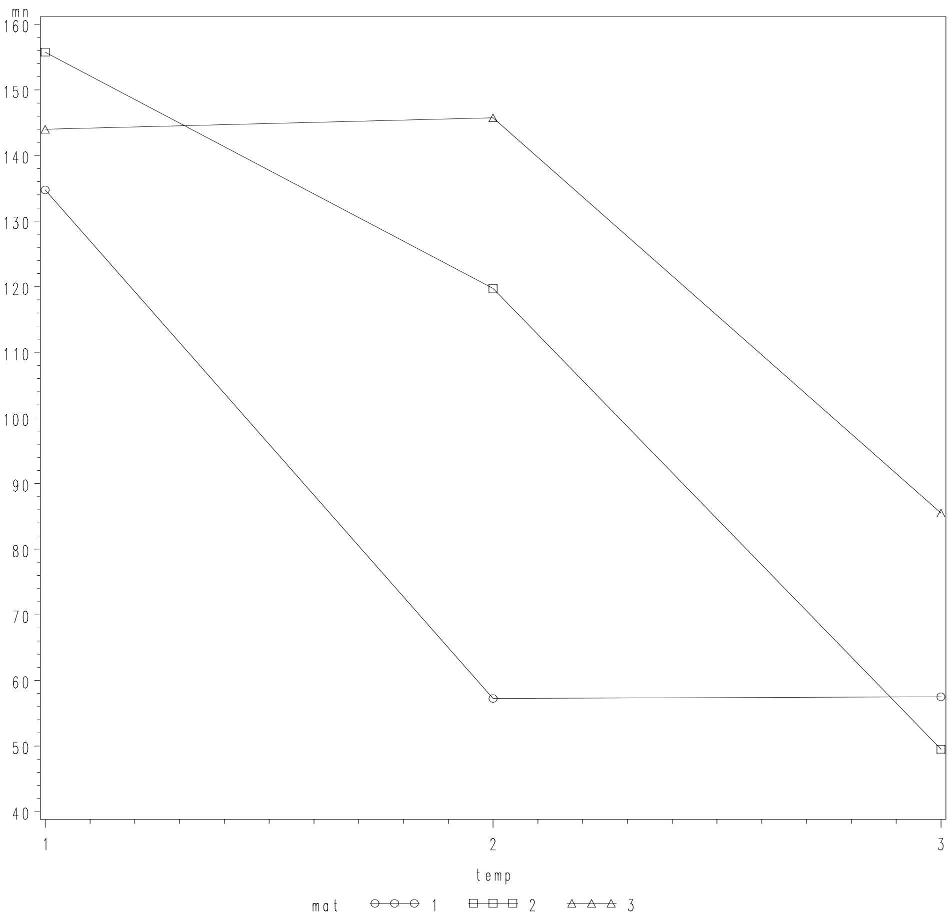
Synergistic Interaction from B to A



Difference between level means of B (with A fixed at a level) depends on the level of A. If the trend of mean response over A do not change when B changes from one level to another, the interaction is said to be synergistic; demonstrated by two unparalleled lines with slopes of the same sign.

Interaction Plot: Battery Experiment

```
data battery;
  input mat temp life;
  datalines;
1 1 130
.......
.......
3 3 60;

proc means noprint;
  var life;
  by mat temp;
  output out=batterymean mean=mn;

symbol1 v=circle  i=join;
symbol2 v=square  i=join;
symbol3 v=triangle i=join;
proc gplot;
  plot mn*temp=mat;
  run;
```


mn
160
150
140
130
120
110
100
90
80
70
60
50
40
1
2
3
temp
mat 1 2 3

Multiple Comparison When Factors Don't Interact

When factors don't interact, i.e., the F test for interaction is not significant in the ANOVA, factor level means can be compared to draw conclusions regarding their effects on response.

- Var$(\bar{y}_{i..}) = \frac{\sigma^2}{nb}$, Var$(\bar{y}_{.j.}) = \frac{\sigma^2}{na}$
- $$\text{For } A \text{ or rows : Var}(\bar{y}_{i..} - \bar{y}_{i'..}) = \frac{2\sigma^2}{nb}; \quad \text{For } B \text{ or columns : Var}(\bar{y}_{.j.} - \bar{y}_{.j'.}) = \frac{2\sigma^2}{na}$$
- Tukey's method
  For rows: CD $= \frac{q_\alpha(a, ab(n-1))}{\sqrt{2}}\sqrt{\text{MSE}\frac{2}{nb}}$
  For columns: CD $= \frac{q_\alpha(b, ab(n-1))}{\sqrt{2}}\sqrt{\text{MSE}\frac{2}{na}}$
- Bonferroni method: CD $= t_{\alpha/2m, ab(n-1)}$S.E., where S.E. depends on whether for rows or columns.

Level Mean Comparison When A and B Interact: An Example

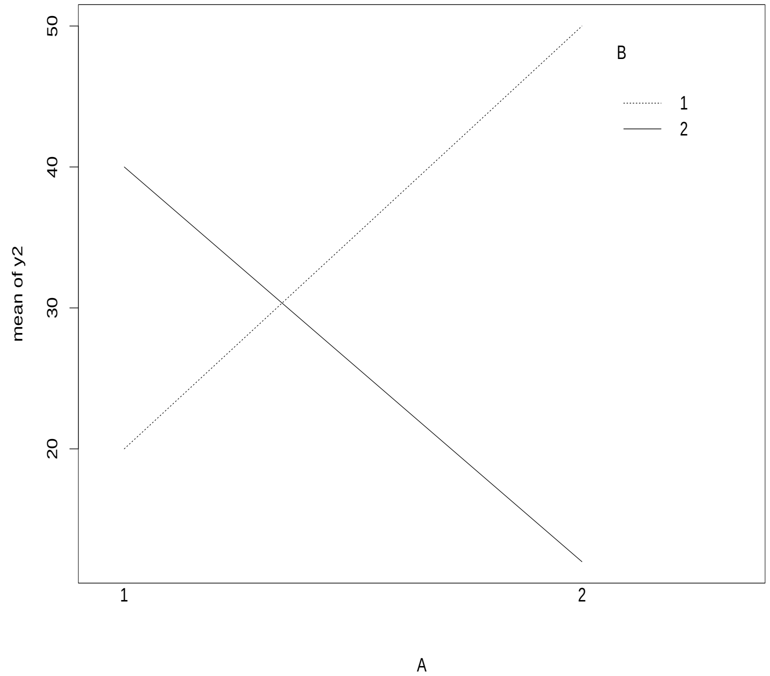
	B	
A	1	2
1	19,21	38,42
2	53,47	10,14

Compare factor level means of A:

$$\bar{y}_{1..} = (19 + 21 + 38 + 42)/4 = 30$$

$$\bar{y}_{2..} = (53 + 47 + 10 + 14)/4 = 31 \approx \bar{y}_{1..}$$

Does Factor A have effect on the response?



When interactions are present, be careful interpreting factor level means (row or column) comparisons because it can be misleading. Usually, we will directly compare treatment means (or cell means) instead.

Multiple Comparisons When Factors Interact

When factors interact, multiple comparison is usually directly applied to treatment means

$$\mu_{ij} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} \text{ vs } \mu_{i'j'} = \mu + \tau_{i'} + \beta_{j'} + (\tau\beta)_{i'j'}$$

- $\hat{\mu}_{ij} = \bar{y}_{ij.}$ and $\hat{\mu}_{i'j'} = \bar{y}_{i'j'.}$
- Var$(\bar{y}_{ij.} - \bar{y}_{i'j'.})$=$\frac{2\sigma^2}{n}$
- there are ab treatment means and $m_0 = \frac{ab(ab-1)}{2}$ pairs.
- Tukey's method:

$$\text{CD} = \frac{q_\alpha(ab, ab(n-1))}{\sqrt{2}}\sqrt{\text{MSE}\frac{2}{n}}$$

- Bonferroni's method.

$$\text{CD} = t_{\alpha/2m, ab(n-1)}\sqrt{\text{MSE}\frac{2}{n}}$$

Using SAS

```
proc glm data=battery;
  class mat temp;
  model life=mat temp mat*temp;
  means mat|temp /tukey lines;
  lsmeans mat|temp/tdiff adjust=tukey;
  run;
```

Source	DF	Type I SS	Mean Square	F Value	Pr > F
mat	2	10683.72222	5341.86111	7.91	0.0020
temp	2	39118.72222	19559.36111	28.97	<.0001
mat*temp	4	9613.77778	2403.44444	3.56	0.0186

Source	DF	Type III SS	Mean Square	F Value	Pr > F
mat	2	10683.72222	5341.86111	7.91	0.0020
temp	2	39118.72222	19559.36111	28.97	<.0001
mat*temp	4	9613.77778	2403.44444	3.56	0.0186

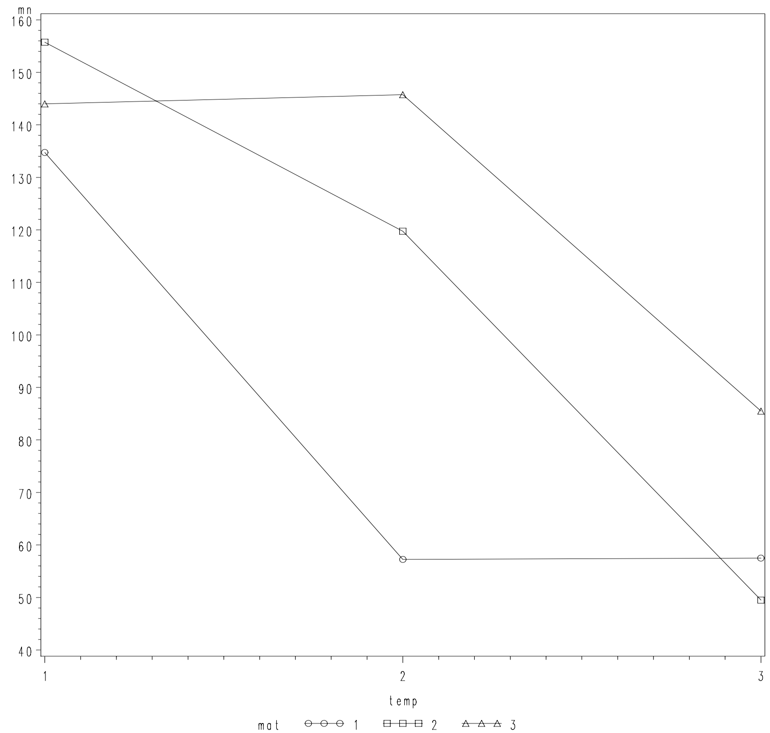
```
Least Squares Means
Adjustment for Multiple Comparisons: Tukey

                                LSMEAN
mat       life LSMEAN           Number
1           83.166667                1
2          108.333333                2
3          125.083333                3

      Least Squares Means for Effect mat
   t for H0: LSMean(i)=LSMean(j) / Pr > |t|

           Dependent Variable: life
i/j                 1              2              3
   1                        -2.37236       -3.95132
                              0.0628         0.0014
   2         2.372362                      -1.57896
               0.0628                        0.2718
   3         3.951318       1.578956
               0.0014         0.2718
```

```
Least Squares Means
Adjustment for Multiple Comparisons: Tukey

                              LSMEAN
temp      life LSMEAN         Number
1          144.833333              1
2          107.583333              2
3           64.166667              3

     Least Squares Means for Effect temp
   t for H0: LSMean(i)=LSMean(j) / Pr > |t|

           Dependent Variable: life
i/j                 1              2              3
   1                         3.51141       7.604127
                              0.0044         <.0001
   2      -3.51141                         4.092717
            0.0044                           0.0010
   3      -7.60413        -4.09272
            <.0001          0.0010
```

```
Least Squares Means
Adjustment for Multiple Comparisons: Tukey

                                        LSMEAN
  mat    temp      life LSMEAN          Number
  1      1          134.750000               1
  1      2           57.250000               2
  1      3           57.500000               3
  2      1          155.750000               4
  2      2          119.750000               5
  2      3           49.500000               6
  3      1          144.000000               7
  3      2          145.750000               8
  3      3           85.500000               9
```

i/j	1	2	3	4	5	6	7	8	9
1		4.2179	4.204294	-1.14291	0.816368	4.63969	-0.50343	-0.59867	2.680408
		0.0065	0.0067	0.9616	0.9953	0.0022	0.9999	0.9995	0.2017
2	-4.2179		-0.01361	-5.36082	-3.40153	0.42179	-4.72133	-4.81657	-1.53749
	0.0065		1.0000	0.0003	0.0460	1.0000	0.0018	0.0014	0.8282
3	-4.20429	0.013606		-5.34721	-3.38793	0.435396	-4.70772	-4.80296	-1.52389
	0.0067	1.0000		0.0004	0.0475	1.0000	0.0019	0.0015	0.8347
4	1.142915	5.360815	5.347209		1.959283	5.782605	0.639488	0.544245	3.823323
	0.9616	0.0003	0.0004		0.5819	0.0001	0.9991	0.9997	0.0172
5	-0.81637	3.401533	3.387926	-1.95928		3.823323	-1.31979	-1.41504	1.86404
	0.9953	0.0460	0.0475	0.5819		0.0172	0.9165	0.8823	0.6420
6	-4.63969	-0.42179	-0.4354	-5.78261	-3.82332		-5.14312	-5.23836	-1.95928
	0.0022	1.0000	1.0000	0.0001	0.0172		0.0006	0.0005	0.5819
7	0.503427	4.721327	4.707721	-0.63949	1.319795	5.143117		-0.09524	3.183834
	0.9999	0.0018	0.0019	0.9991	0.9165	0.0006		1.0000	0.0743
8	0.59867	4.81657	4.802964	-0.54425	1.415038	5.23836	0.095243		3.279077
	0.9995	0.0014	0.0015	0.9997	0.8823	0.0005	1.0000		0.0604
9	-2.68041	1.537493	1.523887	-3.82332	-1.86404	1.959283	-3.18383	-3.27908	
	0.2017	0.8282	0.8347	0.0172	0.6420	0.5819	0.0743	0.0604	

Fitting Response Curves/Surfaces: Battery Experiment



Goal: Model the functional relationship between lifetime and temperature at every material level.

- Material is qualitative while temperature is quantitative
- Want to fit the response using effects of material, temperature and their interactions
- Temperature has quadratic effect. Could use orthogonal polynomials as before. Here we will simply t and t^2.
- Levels of material need to be converted to dummy variables denoted by x_1 and x_2 as follows.

mat	x_1	x_2
1	1	0
2	0	1
3	-1	-1

- For convenience, convert temperature to -1,0 and 1 using

$$t = \frac{\text{temperature} - 70}{55}$$

Model Matrix

mat	temp	==>	x1	x2	t	t^2	x1*t	x2*t	x1*t^2	x2*t^2
1	15		1	0	−1	1	−1	0	1	0
1	70		1	0	0	0	0	0	0	0
1	125		1	0	1	1	1	0	1	0
2	15		0	1	−1	1	0	1	0	1
2	70		0	1	0	0	0	0	0	0
2	125		0	1	1	1	0	−1	0	1
3	15		−1	−1	−1	1	1	−1	−1	−1
3	70		−1	−1	0	0	0	0	0	0
3	125		−1	−1	1	1	−1	1	−1	−1

The following model is used:

$$y_{ijk} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 t + \beta_4 x_1 t + \beta_5 x_2 t + \beta_6 t^2 + \beta_7 x_1 t^2 + \beta_8 x_2 t^2 + \epsilon_{ijk}$$

Want to estimate the coefficients: $\beta_0, \beta_1, \beta_2, \ldots, \beta_8$ using regression

Using SAS

```
data life;
  input mat temp y @@;
  if mat=1 then x1=1;
  if mat=1 then x2=0;
  if mat=2 then x1=0;
  if mat=2 then x2=1;
  if mat=3 then x1=-1;
  if mat=3 then x2=-1;
  t=(temp-70)/55;
  t2=t*t; x1t=x1*t;  x2t=x2*t;
  x1t2=x1*t2;  x2t2=x2*t2;
  datalines;
1 15 130 1 15 155 1 70  34 1 70  40 1 125  20 1 125  70
1 15  74 1 15 180 1 70  80 1 70  75 1 125  82 1 125  58
2 15 150 2 15 188 2 70 136 2 70 122 2 125  25 2 125  70
2 15 159 2 15 126 2 70 106 2 70 115 2 125  58 2 125  45
3 15 138 3 15 110 3 70 174 3 70 120 3 125  96 3 125 104
3 15 168 3 15 160 3 70 150 3 70 139 3 125  82 3 125  60
;

proc reg;
  model y=x1 x2 t x1t x2t t2 x1t2 x2t2;
  run; quit;
```

SAS Output

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	8	59416	7427.02778	11.00	<.0001
Error	27	18231	675.21296		
CorrectedTotal	35	77647			

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	107.58333	7.50118	14.34	<.0001
x1	1	-50.33333	10.60827	-4.74	<.0001
x2	1	12.16667	10.60827	1.15	0.2615
t	1	-40.33333	5.30414	-7.60	<.0001
x1t	1	1.70833	7.50118	0.23	0.8216
x2t	1	-12.79167	7.50118	-1.71	0.0996
t2	1	-3.08333	9.18704	-0.34	0.7398
x1t2	1	41.95833	12.99243	3.23	0.0033
x2t2	1	-14.04167	12.99243	-1.08	0.2894

From the SAS output, the fitted model is

$$
\begin{aligned}
\hat{y} \quad &= 107.58 - 50.33x_1 - 12.17x_2 - 40.33t + 1.71x_1t - 12.79x_2t \\
&-3.08t^2 + 41.96x_1t^2 - 14.04x_2t^2
\end{aligned}
$$

Note that terms with insignificant coefficients are still kept in the fitted model here. In practice, model selection may be employed to remove unimportant terms and choose the best fitted model. But we will not pursue it in this course.

The model above are in terms of both x_1, x_2 and t. We can specify the level of material, that is, the values of dummy variable x_1 and x_2, to derive fitted response curves for material at different levels.

Fitted Response Curves

- Three response curves, with $t = (\text{temperature} - 70)/55$,

  — Material at level 1 ($x_1 = 1, x_2 = 0$)

  $$\hat{y}_{1t} = 57.25 - 38.62t + 38.88t^2$$

  — Material at level 2 ($x_1 = 0, x_2 = 1$)

  $$\hat{y}_{2t} = 119.75 - 53.12t - 17.12t^2$$

  — Material at level 3 ($x_1 = -1, x_2 = -1$)

  $$\hat{y}_{3t} = 145.74 - 29.25t - 31t^2$$

- These curves can be used to predict lifetime of battery at any temperature between 15 and 125 degree. But one needs to be careful about extrapolation.

  — The fitted curve at Material level 1 suggests that lifetime of a battery can be infinity when temperature goes to infinity, which is clearly false.

  — Question: What about the battery life with temperature at 97.5 (i.e., t=0.5)?

General Factorial Design and Model

- Factorial Design - including all possible level combinations
- a levels of Factor A, b levels of Factor $B, \ldots$
- (Straightforward ANOVA if all **fixed effects**)
- In 3 factor model $\rightarrow nabc$ observations
- Need $n > 1$ to test for all possible interactions
- Statistical Model (3 factor)

$$y_{ijkl} = \mu + \tau_i + \beta_j + \gamma_k + (\tau\beta)_{ij} + (\beta\gamma)_{jk} + (\tau\gamma)_{ik} + (\tau\beta\gamma)_{ijk} + \epsilon_{ijkl}$$

$$\begin{cases} i = 1, 2, \ldots, a \\ j = 1, 2, \ldots, b \\ k = 1, 2, \ldots, c \\ l = 1, 2, \ldots, n \end{cases}$$

— Parameter constraints and model assumptions?

ANOVA Table

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0
Factor A	SS_{A}	$a-1$	MS_{A}	F_0
Factor B	SS_{B}	$b-1$	MS_{B}	F_0
Factor C	SS_{C}	$c-1$	MS_{C}	F_0
AB	SS_{AB}	$(a-1)(b-1)$	MS_{AB}	F_0
AC	SS_{AC}	$(a-1)(c-1)$	MS_{AC}	F_0
BC	SS_{BC}	$(b-1)(c-1)$	MS_{BC}	F_0
ABC	SS_{ABC}	$(a-1)(b-1)(c-1)$	MS_{ABC}	F_0
Error	SS_{E}	$abc(n-1)$	MS_{E}	
Total	SS_{T}	$abcn-1$		

- What are those F_0?

Bottling Experiment: SAS Code

```
option nocenter
data bottling;
  input carb pres spee devi;
  datalines;
1  1  1  -3
1  1  1  -1
1  1  2  -1
1  1  2   0
:  :  :   :
3  2  1   9
3  2  2  10
3  2  2  11
;

proc glm;
  class carb pres spee;
  model devi=carb|pres|spee;
  run; quit;
```

Bottling Experiment: SAS Output

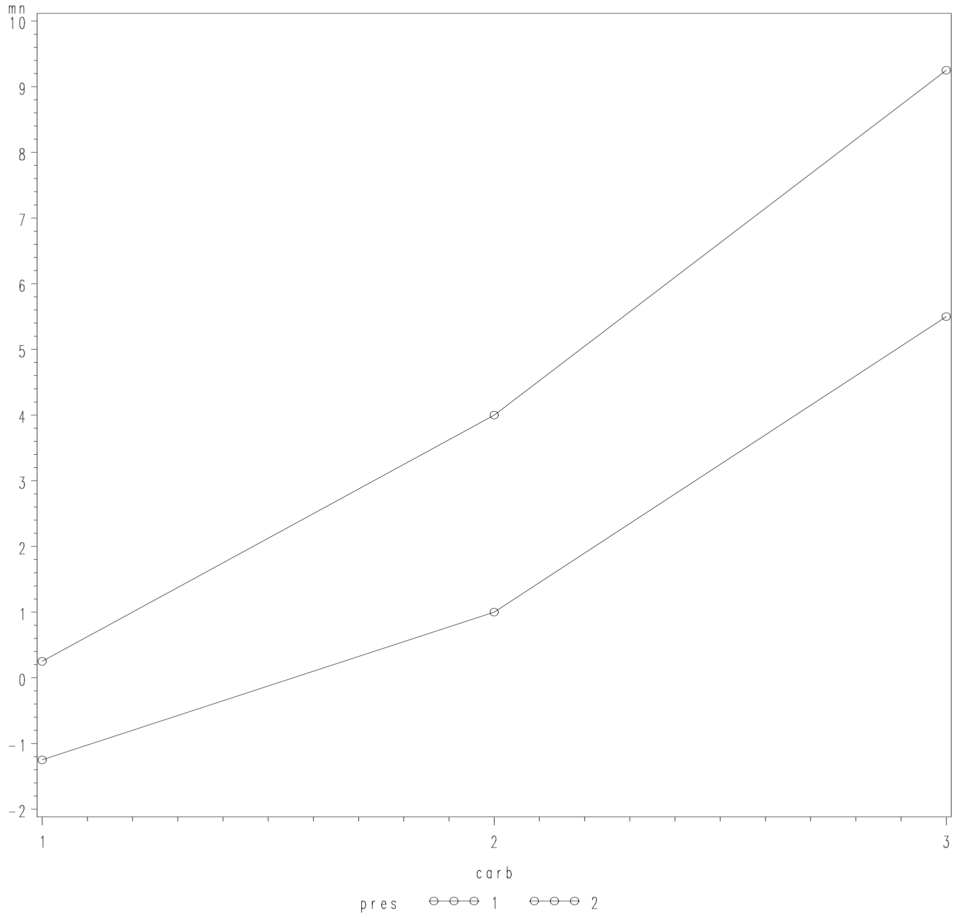
Dependent Variable: devi

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	11	328.1250000	29.8295455	42.11	<.0001
Error	12	8.5000000	0.7083333		
Co Total	23	336.6250000			

R-Square	Coeff Var	Root MSE	devi Mean
0.974749	26.93201	0.841625	3.125000

Source	DF	Type I SS	Mean Square	F Value	Pr > F
carb	2	252.7500000	126.3750000	178.41	<.0001
pres	1	45.3750000	45.3750000	64.06	<.0001
carb*pres	2	5.2500000	2.6250000	3.71	0.0558
spee	1	22.0416667	22.0416667	31.12	0.0001
carb*spee	2	0.5833333	0.2916667	0.41	0.6715
pres*spee	1	1.0416667	1.0416667	1.47	0.2486
carb*pres*spee	2	1.0833333	0.5416667	0.76	0.4869

Interaction Plot for Carb and Pressure



General Factorial Model

- Usual assumptions and diagnostics
- Multiple comparisons: simple extensions of the two-factor case
- Often higher order interactions are negligible.
- Beyond three-way interactions difficult to picture.
- Pooled together with error (increase df_{E})

Blocking in Factorial Design

Battery Lifetime Example: An engineer is studying the effective lifetime of some battery. Two factors, plate material and temperature, are involved. There are three types of plate materials (1, 2, 3) and three temperature levels (15, 70, 125). Four batteries are tested at each combination of plate material and temperature, and all 36 tests are run in random order. The experiment and the resulting observed battery life data are given below.

	temperature		
material	15	70	125
1	130,155,74,180	34,40,80,75	20,70,82,58
2	150,188,159,126	136,122,106,115	25,70,58,45
3	138,110,168,160	174,120,150,139	96,104,82,60

- If we assume further that four operators (1,2,3,4) were hired to conduct the experiment.
  - It is known that different operators can cause systematic difference in battery lifetime.
  - Hence operators should be treated as blocks
- The blocking scheme is that every operator conduct a single replicate of the full factorial design
- For each treatment (treatment combination), the observations were in the order of the operators 1, 2, 3, and 4.
- This is a **blocked factorial design**, indeed, a factorial within RCBD.

Statistical Model for Factorial Within RCBD Experiment

$$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \delta_k + \epsilon_{ijk}$$

$i = 1, 2, \ldots, a$, $j = 1, 2, \ldots, b$ and $k = 1, 2, \ldots, n$; δ_k is the effect of the kth block.

- Randomization restriction is imposed. (Factorial within RCBD).
- Interactions between blocks and treatment effects are assumed to be negligible.
- The previous ANOVA table (Page 9) for the experiment should be modified as follows:

  Add: Block Sum of Square

$$SS_{Blocks} = \frac{1}{ab}\sum_k y_{..k}^2 - \frac{y_{...}^2}{abn}, \qquad d.f. = n - 1$$

  Modify: Error Sum of Squares:

$$\text{(new)}SS_E = \text{(old)}SS_E - SS_{\text{Blocks}}, \qquad d.f. = (ab-1)(n-1)$$

- Other inferences should be modified accordingly.

ANOVA Table of Factorial Within RCBD

Source of Variation	Sum of Squares	Degrees of Freedom	Expected Mean Square	F_0
Blocks	SS_{Block}	$n-1$	$\sigma^2 + ab\sigma_\delta^2$	
A	SS_A	$a-1$	$\sigma^2 + bn\frac{\sum \tau_i^2}{a-1}$	MS_A/MS_E
B	SS_B	$b-1$	$\sigma^2 + an\frac{\sum \beta_j^2}{b-1}$	MS_B/MS_E
AB	SS_AB	$(a-1)(b-1)$	$\sigma^2 + n\frac{\sum\sum(\tau\beta)_{ij}^2}{(a-1)(b-1)}$	MS_{AB}/MS_E
Error	SS_E	$(ab-1)(n-1)$	σ^2	
Total	SS_T	$abn-1$		

$\text{SS}_\text{T} = \sum_i \sum_j \sum_k y_{ijk}^2 - abn\bar{y}_{...}^2$ $\quad$ $\text{SS}_\text{Block} = ab\sum_k \bar{y}_{..k}^2 - abn\bar{y}_{...}^2$

$\text{SS}_\text{A} = bn\sum_i \bar{y}_{i..}^2 - abn\bar{y}_{...}^2$ $\quad$ $\text{SS}_\text{B} = an\sum_j \bar{y}_{.j.}^2 - abn\bar{y}_{...}^2$

$\text{SS}_\text{AB} = n\sum_{i,j} \bar{y}_{ij.}^2 - abn\bar{y}_{...}^2 - SS_\text{A} - SS_\text{B}$

SS_E=SS_T-SS_Block-SS_A-SS_B-SS_AB

Using SAS

```
data battery;
  input mat temp oper life;
  dataline;
1 1 1 130
.........
.........
3 3 4  60
;

proc glm;
  class mat temp oper;
  model life=oper mat|temp;
  output out=new1 r=resi p=pred;
  run; quit;
```

SAS Output

```
Dependent Variable: life
                          Sum of
Source        DF         Squares      Mean Square  F Value   Pr > F
Model         11     59771.19444       5433.74495     7.30   <.0001
Error         24     17875.77778        744.82407
CorTotal      35     77646.97222

Source        DF      Type I SS       Mean Square  F Value   Pr > F
oper           3       354.97222        118.32407     0.16   0.9229
mat            2     10683.72222       5341.86111     7.17   0.0036
temp           2     39118.72222      19559.36111    26.26   <.0001
mat*temp       4      9613.77778       2403.44444     3.23   0.0297
```

Blocking Using Latin Squares: Factorial Within Latin Square Design

There are two blocking factors.

1. Suppose the experimental factors are F1 and F2. F1 has three levels (1,2, 3) and F2 has 2 levels. There are 3*2=6 treatment combinations. These treatments can be represented by Latin letters

F1	F2	Treatment
1	1	A
1	2	B
2	1	C
2	2	D
3	1	E
3	2	F

Two blocking factors are Block1 and Block2, each with 6 blocks.

2. A 6×6 Latin square can be used as the blocking scheme:

	Block1					
Block2	1	2	3	4	5	6
1	A	B	C	D	E	F
2	B	C	D	E	F	A
3	C	D	E	F	A	B
4	D	E	F	A	B	C
5	E	F	A	B	C	D
6	F	A	B	C	D	E

3. Statistical Model

$$y_{ijkl} = \mu + \alpha_i + \tau_j + \beta_k + (\tau\beta)_{jk} + \theta_l + \epsilon_{ijkl}$$

where, α_i and θ_l are blocking effects, τ_j, β_k and $(\tau\beta)_{jk}$ are the treatment main effects and interactions

Q: ANOVA Table? Expected Mean Squares?

Sample Size/Power Calculation

- Consider both factors are fixed
- Compute power for specific contrast or F test
- For F test, have three different Φ^2
  - Factor A: $\Phi^2 = nb\sum_i \tau_i^2/(a\sigma^2) = \delta_A/a$
  - Factor B: $\Phi^2 = na\sum_j \beta_j^2/(b\sigma^2) = \delta_B/b$
  - Interaction: $\Phi^2 = \frac{n\sum_{i,j}(\tau\beta)_{ij}^2}{\{(a-1)(b-1)+1\}\sigma^2} = \frac{\delta_{AB}}{(a-1)(b-1)+1}$
  - Determine sum of squared effects: minimum difference approach, $\cdots$ (Lecture 5)
- Completely Randomized Factorial Design: $\mathrm{df}_{\mathrm{E}} = ab(n-1)$
- Factorial within RCBD: $\mathrm{df}_{\mathrm{E}} = (ab-1)(n-1)$
- $\mathrm{df}_{\mathrm{Num}}$ for F test depends on test factor, see Page 9 & Page 49
- For contrast, need to determine its standard error
- If multiple tests, choose largest n

Example: Battery Life Experiment

- Recall that $a = 3, b = 3, \hat{\sigma}^2 = 675.21$
- Want to find any difference larger than 25 with 80% probability ($\alpha = .05$)

$$\Phi_A^2 = nbD_A^2/(2a\sigma^2) \approx 0.4628n, \qquad \Phi_B^2 = naD_B^2/(2b\sigma^2) \approx 0.4628n,$$
$$\Phi_{AB}^2 = nD_{AB}^2/\{2((a-1)(b-1)+1)\sigma^2\} = 0.0926n$$

- Completely Randomized Design: $\text{df}_{\text{E}} = ab(n-1) = 9(n-1)$
  - Main effects of A ($df_A = 2$) $\rightarrow n = 8$ with power$> 80\%$.
  - Main effects of B ($df_B = 2$) $\rightarrow n = 8$ with power$> 80\%$.
  - Interaction effects AB ($df_{AB} = 4$) $\rightarrow n = 27$ with power$> 80\%$
- Factorial within RCBD: $\text{df}_{\text{E}} = (ab-1)(n-1) = 8(n-1)$
  - Main effects of A ($df_A = 2$) $\rightarrow n = 8$ with power$> 80\%$.
  - Main effects of B ($df_B = 2$) $\rightarrow n = 8$ with power$> 80\%$.
  - Interaction effects AB ($df_{AB} = 4$) $\rightarrow n = 27$ with power$> 80\%$

Using SAS: Completely Randomized Factorial Design

```
data new; a=3; b=3; alpha=0.05; d=25; var=675.21;
  do n=7 to 27;
    df = a*b*(n-1);  /* Completely Randomized Factorial Design */
    nc_a = n*b*d*d/(2*var);  nc_b = n*a*d*d/(2*var);  nc_ab = n*d*d/(2*var);
    fcut_a  = finv(1-alpha,a-1,df);  fcut_b  = finv(1-alpha,b-1,df);
    fcut_ab  = finv(1-alpha,(a-1)*(b-1),df);
    beta_a=probf(fcut_a,a-1,df,nc_a);  beta_b=probf(fcut_b,b-1,df,nc_b);
    beta_ab=probf(fcut_ab,(a-1)*(b-1),df,nc_ab);
    power_a = 1- beta_a;  power_b = 1- beta_b;  power_ab = 1- beta_ab;  output;
  end;

proc print data=new;
  var n df nc_a nc_b nc_ab power_a power_b power_ab;  run;
```

Obs	n	df	nc_a	nc_b	nc_ab	power_a	power_b	power_ab
1	7	54	9.7192	9.7192	3.2397	0.77980	0.77980	0.24209
2	8	63	11.1077	11.1077	3.7026	0.83811	0.83811	0.27721
3	9	72	12.4961	12.4961	4.1654	0.88267	0.88267	0.31267
4	10	81	13.8846	13.8846	4.6282	0.91605	0.91605	0.34818
.	.	. .	. .	. .	. .	. . .	. . .	. . .
20	26	225	36.0999	36.0999	12.0333	0.99987	0.99987	0.79418
21	27	234	37.4883	37.4883	12.4961	0.99992	0.99992	0.81142

Using SAS: PROC GLMPOWER

```
***Need to create a data set of means (following Page 14)
***Notice no blocking effect in this approach;

data batterymns;
  input mat temp life @@;
  datalines;
1 1 134.75  1 2  57.25  1 3  57.50  2 1 155.75  2 2 119.75
2 3  49.50  3 1 144.00  3 2 145.75  3 3  85.50
;

proc glmpower data=batterymns;
  class mat temp;
  model life = mat temp mat*temp;
  power stddev=25.98  alpha=0.05
        ntotal=.      power=0.8;
  run;
```

Computed N Total

Index	Source	Test DF	Error DF	Actual Power	N Total
1	mat	2	18	0.815	27
2	temp	2	9	0.985	18
3	mat*temp	4	27	0.801	36

Using SAS: Factorial Within RCBD

```
data new; a=3; b=3; alpha=0.05; d=25; var=675.21;
  do n=7 to 27;
    df = (a*b-1)*(n-1); /* Factorial within RCBD */
    nc_a = n*b*d*d/(2*var);  nc_b = n*a*d*d/(2*var);  nc_ab = n*d*d/(2*var);
    fcut_a  = finv(1-alpha,a-1,df);  fcut_b  = finv(1-alpha,b-1,df);
    fcut_ab  = finv(1-alpha,(a-1)*(b-1),df);
    beta_a=probf(fcut_a,a-1,df,nc_a);  beta_b=probf(fcut_b,b-1,df,nc_b);
    beta_ab=probf(fcut_ab,(a-1)*(b-1),df,nc_ab);
    power_a = 1- beta_a;  power_b = 1- beta_b;  power_ab = 1- beta_ab;  output;
  end;

proc print data=new;
  var n df nc_a nc_b nc_ab power_a power_b power_ab;  run;
```

Obs	n	df	nc_a	nc_b	nc_ab	power_a	power_b	power_ab
1	7	48	9.7192	9.7192	3.2397	0.77673	0.77673	0.23977
2	8	56	11.1077	11.1077	3.7026	0.83576	0.83576	0.27485
3	9	64	12.4961	12.4961	4.1654	0.88091	0.88091	0.31029
4	10	72	13.8846	13.8846	4.6282	0.91475	0.91475	0.34580
. .	. .	. .	. .	. .	. .	. . .	. . .	. . .
20	26	200	36.0999	36.0999	12.0333	0.99987	0.99987	0.79298
21	27	208	37.4883	37.4883	12.4961	0.99991	0.99991	0.81030