Factorial Designs III

Design of Experiments - Montgomery
Sections 5-4 - 5-7

General Factorial Model

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

$$y_{ijkl} = \mu + \tau_i + \beta_j + \gamma_k + (\tau\beta)_{ij} + (\beta\gamma)_{jk} + (\tau\gamma)_{ik} + (\tau\beta\gamma)_{ijk} + \epsilon_{ijkl}$$
$$\begin{cases} i = 1, 2, \ldots, a \\ j = 1, 2, \ldots, b \\ k = 1, 2, \ldots, c \\ l = 1, 2, \ldots, n \end{cases}$$

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
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	$abcn-1$		

General Factorial Model

- Usual assumptions and diagnostics
- Multiple comparisons simple extension of two factor

- Often higher order interactions ignored
- Beyond three-way interactions difficult to picture
- Pooled together with error (increase df_E)

- If all factors random

  For example a two factor model

  $E(MS_E) = \sigma^2$

  $E(MS_A) = \sigma^2 + bn\sigma^2_\tau + n\sigma^2_{\tau\beta}$

  $E(MS_B) = \sigma^2 + an\sigma^2_\beta + n\sigma^2_{\tau\beta}$

  $E(MS_{AB}) = \sigma^2 + n\sigma^2_{\tau\beta}$
- Don't necessarily test over MSE
- Mixture of fixed/random effects $\rightarrow$ mixed models
- Will be discussed in Chapter 12

Response Curves and Surfaces

When factors are quantitative (distinct ordering of levels of factor) may want to look at response curve or response surface to interpolate (i.e., predict the response for levels between those actually used). This is frequently done when the goal of the experiment is to find the "best" combination of factors (highest average response).

Example 5-4: An experimenter is interested in the effective life of a battery made of $a = 3$ different type of material and stored at $b = 3$ different temperatures. Since temperature is a quantitative factor, the experimenter is interested in not only seeing if either of the factors (or combination) affect battery life but also predicting what the battery life would be for other temperature values.

One can get crude representations of the response curves by creating an interaction plot (Figure 5-9). While there is an overall decrease in the battery life as temperature increases, the rate at which it decreases is quite different across materials. Notice that the life using Material 1 dramatically drops between the lowest and middle temperatures. This suggests an interaction and this interaction is found significant in Table 5.5 ($P = .0186$).

Response Curves and Surfaces

Montgomery presents a response surface analysis using Design-Expert. This response surface analysis is a combination of ANOVA and regression. Orthogonal polynomial contrasts (Table X) separate the appropriate SS into linear and quadratic components (both for the main effect and interaction). From this analysis, you see the difference in response curves is primarily due to a differing quadratic component ($P = .0106$). This is pretty obvious from the interaction plot.

Below are the commands to do this same analysis using SAS. The temperature variable is treated as if it were a covariate (i.e., considered continuous). The parameter estimates can then be used to estimate the response curve for each material.

```
options nocenter;

data machine;
 input mat temp y @@;
 temp2 = temp*temp;
cards;
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 glm;
 class mat;
 model y=temp mat temp2 mat*temp mat*temp2/ solution;
run;
```


Dependent Variable: y

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

Source	DF	Type I SS	Mean Square	F Value	Pr > F
temp	1	39042.66667	39042.66667	57.82	<.0001
mat	2	10683.72222	5341.86111	7.91	0.0020
temp2	1	76.05556	76.05556	0.11	0.7398
temp*mat	2	2315.08333	1157.54167	1.71	0.1991
temp2*mat	2	7298.69444	3649.34722	5.40	0.0106

Source	DF	Type III SS	Mean Square	F Value	Pr > F
temp	1	1239.179404	1239.179404	1.84	0.1867
mat	2	1147.938218	573.969109	0.85	0.4385
temp2	1	76.055556	76.055556	0.11	0.7398
temp*mat	2	7170.660365	3585.330183	5.31	0.0114
temp2*mat	2	7298.694444	3649.347222	5.40	0.0106

Parameter		Estimate		Standard Error	t Value	Pr > |t|
Intercept		132.7623967	B	20.56765571	6.45	<.0001
temp		0.9028926	B	0.75514800	1.20	0.2422
mat	1	36.6177686	B	29.08705766	1.26	0.2188
mat	2	26.8615702	B	29.08705766	0.92	0.3639
mat	3	0.0000000	B	.	.	.
temp2		-0.0102479	B	0.00526030	-1.95	0.0619
temp*mat	1	-3.4043388	B	1.06794054	-3.19	0.0036
temp*mat	2	-1.0762397	B	1.06794054	-1.01	0.3225
temp*mat	3	0.0000000	B	.	.	.
temp2*mat	1	0.0230992	B	0.00743919	3.11	0.0044
temp2*mat	2	0.0045868	B	0.00743919	0.62	0.5427
temp2*mat	3	0.0000000	B	.	.	.

Response Curves and Surfaces

The following model is used

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

where

$$x_1 = \begin{cases} 1 & \text{if mat=1} \\ 0 & \text{if mat=2} \\ -1 & \text{if mat=3} \end{cases} \qquad x_2 = \begin{cases} 0 & \text{if mat=1} \\ 1 & \text{if mat=2} \\ -1 & \text{if mat=3} \end{cases}$$

and results in the following three response curves.

$$E(y_{1t}) = 169.38 - 2.5014t + 0.0129t^2$$
$$E(y_{2t}) = 159.62 - 0.1733t - 0.0057t^2$$
$$E(y_{3t}) = 132.76 + 0.9029t - 0.0102t^2$$

These curves could then be used to predict other lives for different temperatures. Interestingly, material 1 has a curve that suggests the best temperature is plus or minus infinity. One needs to be careful with interpolation.

Blocking in Factorial Model

- For randomized complete block
  - Each treatment appears in each block
  - CRD replicate in each block
- For factorial design
  - All combinations appear in each block
  - Factorial replicate in each block
- Blocking on two factors - Latin Square
  - Need p treatments run in $p \times p$ square
  - If $ab.. = p$, combinations can be run in Latin Square
- Additional additivity assumptions
- Interaction with blocks reduces df for error

Using SAS

EXAMPLE 5-6

```
options nocenter ls=75;

data new;
input oper filter clutter resp @@;
cards;
1 1 1 90 1 1 2 102 1 1 3 114 1 2 1 86
1 2 2 87 1 2 3 93 2 1 1 96 2 1 2 106
2 1 3 112 2 2 1 84 2 2 2 90 2 2 3 91
3 1 1 100 3 1 2 105 3 1 3 108 3 2 1 92
3 2 2 97 3 2 3 95 4 1 1 92 4 1 2 96
4 1 3 98 4 2 1 81 4 2 2 80 4 2 3 83
;

proc glm;
class oper filter clutter;
model resp = oper filter|clutter;
output out=new1 r=res p=pred;
run;
```



```
Dependent Variable: resp

                                    Sum of
Source                      DF     Squares    Mean Square  F Value  Pr > F

Model                        8  1881.500000   235.187500    21.21  <.0001
Error                       15   166.333333    11.088889
Corrected Total             23  2047.833333


R-Square     Coeff Var      Root MSE     resp Mean
0.918776      3.508339      3.329998      94.91667


Source                      DF    Type I SS   Mean Square  F Value  Pr > F

oper                         3   402.166667   134.055556    12.09  0.0003
filter                       1  1066.666667  1066.666667    96.19  <.0001
clutter                      2   335.583333   167.791667    15.13  0.0003
filter*clutter               2    77.083333    38.541667     3.48  0.0575

Source                      DF  Type III SS   Mean Square  F Value  Pr > F

oper                         3   402.166667   134.055556    12.09  0.0003
filter                       1  1066.666667  1066.666667    96.19  <.0001
clutter                      2   335.583333   167.791667    15.13  0.0003
filter*clutter               2    77.083333    38.541667     3.48  0.0575
```


Unbalanced Factorial Design

- Sometime n_i varies for each combination
- Could be result of missing observations
- Could be designed to be unbalanced

  Certain combinations of more importance

- Use same approaches as before
  - Regression Approach
  - Estimate missing value(s)
  - Remove data so balanced
- SAS Type I-Type IV SS
  - Type I - fit in order of model statement
  - Type II - fits only unrelated terms or lower order first
  - Type III - fits all other terms first
  - Type IV - used when some $n_i = 0$

Example

A factorial experiment is performed to investigate the effects of genetic strain and environment on the number of errors a rat makes in a maze. Due to a shortage of rats, the number of replicates per combination varied.

	Genetic Strain		
Environment	1	2	3
Free	92 89 100	85 76 72 92	51 61 47
Controlled	98 106	80 72 92	73 82 69 77

From SAS

```
Level of   Level of          -------------NUM-------------
ENV        TRAIT      N       Mean              SD

1          1          3      93.666667         5.6862407
1          2          4      81.250000         8.9953692
1          3          3      53.000000         7.2111026
2          1          2     102.000000         5.6568542
2          2          3      81.333333        10.0664459
2          3          4      75.250000         5.5602758
```

SS_E is the sum of the squared differences between the cell observations and the cell mean

```
options nocenter ls=75;

data rat;
 input env trait num @@;
cards;
1 1 92 1 1 100 1 1 89
2 1 106 2 1 98
1 2 85 1 2 76 1 2 72 1 2 92
1 3 51 1 3 61 1 3 47
2 2 80 2 2 72 2 2 92
2 3 73 2 3 82 2 3 77 2 3 69
;

proc glm;
class env trait;
model num=env|trait / ss1 ss2 ss3 ss4;
means env|trait;
lsmeans env|trait / stderr;
output out=new r=res p=pred;

proc plot;
plot res*pred;
run;
```


Dependent Variable: num

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	3815.271930	763.054386	13.43	<.0001
Error	13	738.833333	56.833333		
Corrected Total	18	4554.105263			

R-Square	Coeff Var	Root MSE	num Mean
0.837765	9.460831	7.538789	79.68421

Source	DF	Type I SS	Mean Square	F Value	Pr > F
env	1	214.049708	214.049708	3.77	0.0743
trait	2	3172.919753	1586.459877	27.91	<.0001
env*trait	2	428.302469	214.151235	3.77	0.0512

Source	DF	Type II SS	Mean Square	F Value	Pr > F
env	1	503.721340	503.721340	8.86	0.0107
trait	2	3172.919753	1586.459877	27.91	<.0001
env*trait	2	428.302469	214.151235	3.77	0.0512

Source	DF	Type III SS	Mean Square	F Value	Pr > F
env	1	470.222222	470.222222	8.27	0.0130
trait	2	3254.880952	1627.440476	28.64	<.0001
env*trait	2	428.302469	214.151235	3.77	0.0512

Source	DF	Type IV SS	Mean Square	F Value	Pr > F
env	1	470.222222	470.222222	8.27	0.0130
trait	2	3254.880952	1627.440476	28.64	<.0001
env*trait	2	428.302469	214.151235	3.77	0.0512

Means

Level of env	N	num Mean	num Std Dev
1	10	76.5000000	18.3862872
2	9	83.2222222	12.7355583

Level of trait	N	num Mean	num Std Dev
1	5	97.0000000	6.7082039
2	7	81.2857143	8.6161532
3	7	65.7142857	13.1999278

Least Squares Means

env	num LSMEAN	Standard Error	Pr > |t|
1	75.9722222	2.4059469	<.0001
2	86.1944444	2.6155400	<.0001

trait	num LSMEAN	Standard Error	Pr > |t|
1	97.8333333	3.4409705	<.0001
2	81.2916667	2.8789224	<.0001
3	64.1250000	2.8789224	<.0001

env	trait	num LSMEAN	Standard Error	Pr > |t|
1	1	93.666667	4.352522	<.0001
1	2	81.250000	3.769394	<.0001
1	3	53.000000	4.352522	<.0001
2	1	102.000000	5.330729	<.0001
2	2	81.333333	4.352522	<.0001
2	3	75.250000	3.769394	<.0001