Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Experimental Design part 2

Advanced Systems Lab – 263-0007-00

G. Alonso, D. Kossmann, T. Roscoe

Systems Group, ETH Zürich, Switzerland

27 October 2009

Outline

For this week

Introduction

Testing the Assumptions

Confidence Intervals

2^{k-p} Fractional Factorial Design

One-factor experiments

Recap

What does it mean to design an experiment?

Recap

What does it mean to design an experiment?

We are interested in:

- Which factors are the most important?
- Which factors are related?

Goal: get the most information with least effort (minimum number of experiments)

Recap

What does it mean to design an experiment?

We are interested in:

- Which factors are the most important?
- Which factors are related?

Goal: get the most information with least effort (minimum number of experiments)

Goal for today: how to reduce the number of experiments you need for *k* factors.

Recall our assumptions

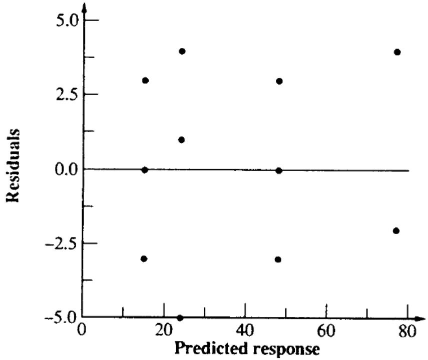
In particular:

- Model errors are IID (independent, identically distributed)
- Model errors are normally distributed

Use visual tests to validate these assumptions...

Are the errors IID?

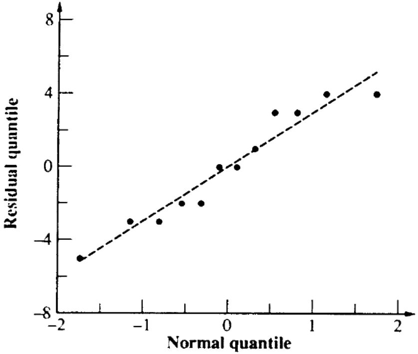
Plot each error (residual) against predicted response $\hat{y}_i$:



✔ Any trend ⇒ dependence of errors on factors

Are the errors normal?

Plot error quantile vs. normal quantile:



✔ Normal distribution ⇒ approximately linear plot

Confidence intervals for effects in $2^k r$ designs

Effects (the q values) are computed from a sample
$\Rightarrow$ they are random variables
$\Rightarrow$ can compute confidence intervals if:

- We know the variance of the sample estimates
- We assume errors are normally distributed, with zero mean

Let error variance be σ_e^2.
Then y_i's are are also normally distributed with variance σ_e^2

Confidence intervals for effects in $2^k r$ designs

Consider an effect like q_0:

$$q_0 = \frac{1}{2^2 r} \sum_{i,j} y_{ij}$$

q_0 is a linear combination of normally distributed vars
$\Rightarrow$ it is normally distributed with variance = $\sigma_e^2/(2^2 r)$
Estimate of variance of errors is therefore:

$$s_e^2 = \frac{SSE}{2^2(r-1)} = MSE$$

(MSE = Mean Square of Errors)

Confidence intervals for effects in $2^k r$ designs

$$MSE = s_e^2 = \frac{SSE}{2^2(r-1)}$$

The SSE has $2^2(r-1)$ independent terms. Why?

Confidence intervals for effects in $2^k r$ designs

$$MSE = s_e^2 = \frac{SSE}{2^2(r-1)}$$

The SSE has $2^2(r-1)$ independent terms. Why?

A.: In r replications, r error terms must add up to zero
$\Rightarrow$ only $r-1$ can be independently chosen

Estimated s.d. of q_0, and (in fact all the others) is:

$$s_{q_0} = s_{q_A} = s_{q_B} = s_{q_A B} = \frac{s_e}{\sqrt{2^2 r}}$$

Confidence intervals for effects in $2^k r$ designs

Confidence intervals for effects are:

$$q_i \mp t_{[1-\alpha/2;2^2(r-1)]} s_{q_i}$$

If one of these crosses zero, it's significant. For our example:
Standard deviation of effects is:

$$s_e = \sqrt{\frac{SSE}{2^2(r-1)}} = \sqrt{\frac{102}{8}} = 3.57$$

Standard deviation of errors is:

$$s_{q_i} = s_e / \sqrt{(2^2 r)} = 3.57/\sqrt{12} = 1.03$$

Confidence intervals for effects in $2^k r$ designs

t-value for 8 $(= 2^2(3-1))$ degrees of freedom and 90% confidence is 1.86.

Confidence intervals for effects are therefore:

$$q_i \mp (1.86)(1.03)$$

Giving:

q_0	(39.98, 42.91)
q_A	(19.58, 23.41)
q_B	(7.58, 11.41)
q_{AB}	(3.08, 6.91)

Conclusion: they are all significant at 90% confidence.

2^{k-p} Fractional Factorial Design

Consider k=7 factors, 2 levels per factor.

- 2^k Factorial design requires 128 experiments. This could be a lot
- Can we get away with 2^{k-p} instead?
  - If $p = 2$, only need one-quarter the number of experiments.
  - If $p = 4$, only need one-sixteenth (8 experiments in our case).

Recap: Sign tables (from last week)

I	A	B	AB	y
1	-1	-1	1	15
1	1	-1	-1	45
1	-1	1	-1	25
1	1	1	1	75
160	80	40	20	Total
40	20	10	5	Total/4

Notice that this is the design of the experiment: each row is an experiment with A and B set to different levels.

Why do sign tables work?

I	A	B	AB
1	-1	-1	1
1	1	-1	-1
1	-1	1	-1
1	1	1	1

How do 4 (in this case) experiments allow us to isolate the effect of each factor (and their interactions)? We can explain this in terms of vector algebra.

Orthogonality

1. The columns of the sign table are mutually orthogonal vectors:

$$\sum_j x_{ij} x_{ik} = 0, \forall j \neq k$$

2. The sum of each column is zero:

$$\sum_i x_{ij} = 0, \forall j$$

3. The magnitude of the vector is constant (here, 2^k):

$$\sum_i x_{ij}^2 = 4, \forall j$$

Sign table for 2^{7-4} design

Experiment	A	B	C	D	E	F	G
1	-1	-1	-1	1	1	1	-1
2	1	-1	-1	-1	-1	1	1
3	-1	1	-1	-1	1	-1	1
4	1	1	-1	1	-1	-1	-1
5	-1	-1	1	1	-1	-1	1
6	1	-1	1	-1	1	-1	-1
7	-1	1	1	-1	-1	1	-1
8	1	1	1	1	1	1	1

Fitting models

We can use this table fit the following model:

$$y = q_0 + q_A x_A + q_B x_B + q_C x_C + q_D x_D + q_E x_E + q_F x_F + q_G x_G$$

Fitting models

We can use this table fit the following model:

$$y = q_0 + q_A x_A + q_B x_B + q_C x_C + q_D x_D + q_E x_E + q_F x_F + q_G x_G$$

Since the vectors are orthogonal, we have 8 linearly independent equations in q_i, giving:

$$q_A = \sum_i y_i x_{Ai} = \frac{-y_1 + y_2 - y_3 + y_4 - y_5 + y_6 - y_7 + y_8}{8}$$

$$q_B = \sum_i y_i x_{Bi} = \frac{-y_1 - y_2 + y_3 + y_4 - y_5 - y_6 + y_7 + y_8}{8}$$

etc.

You may have noticed...

- Calculation is essentially the same as for 2^k designs
- Generalizes to $2^{k-p}r$ designs
- Same techniques can be used for:
  - allocation of variation
  - confidence intervals
  - standard deviations of effects
  - error estimation

Some example results

I	A	B	C	D	E	F	G	y
1	-1	-1	-1	1	1	1	-1	20
1	1	-1	-1	-1	-1	1	1	35
1	-1	1	-1	-1	1	-1	1	7
1	1	1	-1	1	-1	-1	-1	42
1	-1	-1	1	1	-1	-1	1	36
1	1	-1	1	-1	1	-1	-1	50
1	-1	1	1	-1	-1	1	-1	45
1	1	1	1	1	1	1	1	82
317	101	35	109	43	1	47	3	Total
39.62	12.62	4.37	13.62	5.37	0.125	5.87	0.37	Total/8

Some example results

I	A	B	C	D	E	F	G	y
1	-1	-1	-1	1	1	1	-1	20
1	1	-1	-1	-1	-1	1	1	35
1	-1	1	-1	-1	1	-1	1	7
1	1	1	-1	1	-1	-1	-1	42
1	-1	-1	1	1	-1	-1	1	36
1	1	-1	1	-1	1	-1	-1	50
1	-1	1	1	-1	-1	1	-1	45
1	1	1	1	1	1	1	1	82
317	101	35	109	43	1	47	3	Total
39.62	12.62	4.37	13.62	5.37	0.125	5.87	0.37	Total/8
37.26	4.74	43.40	6.75	0	8.06	0.03		% variation

You may also have noticed...

- We've lost the interactions.
- Works well if we are sure the interactions are small...
- More generally: 2^{k-p} designs trade off interaction analysis for reduced experimental overhead.
  - Intuition: our model assumes the experiment space is spanned by basis vectors corresponding to the predictors.
  - Before, it was the predictors and their subproducts.
- This will become clearer when we look at how the sign table is designed...

Where does this sign table come from?

For a design with 8 experiments, start with a 2^3 table:

Experiment	A	B	C	AB	AC	BC	ABC
1	-1	-1	-1	1	1	1	-1
2	1	-1	-1	-1	-1	1	1
3	-1	1	-1	-1	1	-1	1
4	1	1	-1	1	-1	-1	-1
5	-1	-1	1	1	-1	-1	1
6	1	-1	1	-1	1	-1	-1
7	-1	1	1	-1	-1	1	-1
8	1	1	1	1	1	1	1

Where does this sign table come from?

…and replace the interactions with extra factors

Experiment	A	B	C	D	E	F	G
1	-1	-1	-1	1	1	1	-1
2	1	-1	-1	-1	-1	1	1
3	-1	1	-1	-1	1	-1	1
4	1	1	-1	1	-1	-1	-1
5	-1	-1	1	1	-1	-1	1
6	1	-1	1	-1	1	-1	-1
7	-1	1	1	-1	-1	1	-1
8	1	1	1	1	1	1	1

Another example: 2^{4-1} design

Key point: the columns are orthonormal.

Experiment	A	B	C	AB	AC	BC	D
1	-1	-1	-1	1	1	1	-1
2	1	-1	-1	-1	-1	1	1
3	-1	1	-1	-1	1	-1	1
4	1	1	-1	1	-1	-1	-1
5	-1	-1	1	1	-1	-1	1
6	1	-1	1	-1	1	-1	-1
7	-1	1	1	-1	-1	1	-1
8	1	1	1	1	1	1	1

Confounding

- In this design, one can't separate the effects of D and ABC.
- In fact, they are the same:

$$q_D = \sum_i y_i x_{Di} = \frac{-y_1 + y_2 - y_3 + y_4 - y_5 + y_6 - y_7 + y_8}{8}$$

$$q_{ABC} = \sum_i y_i x_{Ai} x_{Bi} x_{Ci} = \frac{-y_1 + y_2 - y_3 + y_4 - y_5 + y_6 - y_7 + y_8}{8}$$

- What we really doing is calculating their total effect:

$$q_D + q_{ABC} = \sum_i y_i x_{Ai} x_{Bi} x_{Ci} = \frac{-y_1 + y_2 - y_3 + y_4 - y_5 + y_6 - y_7 + y_8}{8}$$

Confounding

We say that *D* and *ABC* are confounded in this experiment.

From Webster:

1. **a** archaic : to bring to ruin : destroy; **b** : baffle, frustrate ("conferences…are not for accomplishment but to confound knavish tricks" – J. K. Galbraith)
2. obsolete : consume, waste
3. **a** : to put to shame : discomfit ("a performance that confounded the critics"); **b** : refute ("sought to confound his arguments")
4. damn
5. to throw (a person) into confusion or perplexity
6. **a** : to fail to discern differences between: mix up; **b** : to increase the confusion of

Confounding

We write $D = ABC$.

Confounding

We write $D = ABC$.

Actually, every column in the sign table is a sum of two effects:

A = BCD	B = ACD	C = ABD	AB = CD
AC = BD	BC = AD	ABC = D	I = ABCD

How bad is this? Well, it could be worse…

Alternative 2^{4-1} design

Experiment	A	B	C	D	AC	BC	ABC
1	-1	-1	-1	1	1	1	-1
2	1	-1	-1	-1	-1	1	1
3	-1	1	-1	-1	1	-1	1
4	1	1	-1	1	-1	-1	-1
5	-1	-1	1	1	-1	-1	1
6	1	-1	1	-1	1	-1	-1
7	-1	1	1	-1	-1	1	-1
8	1	1	1	1	1	1	1

This one *is* worse

It has the following confoundings:

I = ABD	A = BD	B = AD	C = ABCD
D = AB	AC = BCD	BC = ACD	ABC = CD

As opposed to:

A = BCD	B = ACD	C = ABD	AB = CD
AC = BD	BC = AD	ABC = D	I = ABCD

This one *is* worse

It has the following confoundings:

I = ABD	A = BD	B = AD	C = ABCD
D = AB	AC = BCD	BC = ACD	ABC = CD

As opposed to:

A = BCD	B = ACD	C = ABD	AB = CD
AC = BD	BC = AD	ABC = D	I = ABCD

In general, we expect higher-order effects to be smaller than lower-order ones.

Confoundings form an Abelian group

One confounding for an experiment uniquely implies all the others:

- $Ix = x; \forall x$ (e.g. $IA = A$)
- $x^2 = I; \forall x$ (e.g. $A^2 = I$)
- Multiplication

Given $I = ABCD$, can generate all other confoundings by multiplying by A, B, C, or D and then reducing.

Similarly for $I = ABD$.

Design resolution

The order of an effect is the number of factors included in it.

- AB is of order 2
- I is of order 0

Design resolution

The order of an effect is the number of factors included in it.

- AB is of order 2
- I is of order 0

The order of a confounding is $i + j$ if an i th-order effect is confounded with a j th order term.

- $A = BCD$ is of order 4.
- $I = ABC$ is of order 3.

Design resolution

The resolution of an experimental design is the minimum order of effects confounded in it.

- The $I = ABCD$ design has resolution 4 (it is an R_{IV} experiment).
- The $I = ABD$ design is R_{III}.

One-factor experiments

Scenario: only one factor, but many levels of that factor. E.g.

- Comparing different systems
- Different cache sizes
- Different schemas

As before, we start with a model...

One-factor experimental model

We use the following model:

$$y_{ij} = \mu + \alpha_j + e_{ij}$$

where:

- y_{ij} is the ith response with the factor at level j.
- μ is the mean response
- α_j is the effect of level (or alternative) j
- e_{ij} is the error term

And we require:

$$\sum \alpha_j = 0$$

Computation of effects

Assume *r* observations for each of *a* alternatives (levels).
From the model:

$$\sum_{i=1}^{r}\sum_{j=1}^{a} y_{ij} = ar\mu + r\sum_{j=1}^{a}\alpha_j + \sum_{i=1}^{r}\sum_{j=1}^{a} e_{ij}$$

But:

- The effects (α_j) add up to zero
- Mean error should be zero

So:

$$\sum_{i=1}^{r}\sum_{j=1}^{a} y_{ij} = ar\mu$$

Computation of effects

Or:

$$\mu = \frac{1}{ar} \sum_{i=1}^{r} \sum_{j=1}^{a} y_{ij}$$

This is the grand mean of all responses.