The Theory of Designed Experiments

5. Optimal Choice of Treatments

UNESP-Botucatu, August 2010

How Much Replication?

Example A clinical trial was to be conducted to compare the effect of hawthorn berries on blood pressure with a control. The main question was how many patients should be given each treatment?

This is the question about design of experiments which statisticians are most frequently asked - how many replicates of each treatment are needed?

To answer it, we need to have clearly defined objectives - how many replicates are needed to do what?

Example

In the example, the objective is to estimate (and test) the difference between the two treatments. A 95% confidence interval (CI) for the difference is given by

$$\bar{y}_C - \bar{y}_T \pm t_{n-p;0.975} \times \sqrt{\frac{s^2}{n_C} + \frac{s^2}{n_T}},$$

where $n - p$ is the residual d.f. and n_C and n_T are the replicates of the treatments.

Randomization ensures that the estimates are unbiased and so the CI will be centred in the right place. How do we make the CI narrower?

Example

We can get more precise estimates by:

- decreasing $t_{n-p;0.975}$ by increasing residual d.f. - should always try to ensure 10-15 but little is gained by having > 20 (see t tables);
- decreasing s^2 by introducing more blocking to remove the larger sources of variation between units;
- increasing n_C and/or n_T.

Example

Clearly, unless costs are different, we should take $n_C = n_T$.

To make any further decisions we need a reliable prior estimate of s^2. We can usually get this from previous studies.

In the example, experience suggests $s^2 \approx 400$ over the whole population, but $s^2 \approx 100$ if we arrange people into pairs according to age (blocks of size 2).

We can calculate the width of the CI for various numbers of replicates, with and without blocking.

Example

	No Blocking ($s^2 \approx 400$)			
$n_C = n_T$	$n - p$	$t_{n-p;0.975}$	$\sqrt{\frac{2s^2}{n_T}}$	$\frac{1}{2}$ CI width
2	2	4.303	20.00	86.1
4	6	2.447	14.14	34.6
6	10	2.228	11.55	26.7
8	14	2.145	10.00	21.4
10	18	2.101	8.944	18.8
12	22	2.074	8.165	16.9
14	26	2.056	7.559	15.5
16	30	2.042	7.071	14.4
18	34	2.032	6.667	13.5
20	38	2.024	6.325	12.8

Example

	Blocking ($s^2 \approx 100$)			
$n_C = n_T$	$n - p$	$t_{n-p;0.975}$	$\sqrt{\frac{2s^2}{n_T}}$	$\frac{1}{2}$ CI width
2	1	12.71	10.00	127
4	3	3.182	7.071	22.5
6	5	2.571	5.774	14.8
8	7	2.365	5.000	11.8
10	9	2.262	4.472	10.1
12	11	2.201	4.082	8.98
14	13	2.160	3.780	8.16
16	15	2.131	3.536	7.54
18	17	2.110	3.333	7.03
20	19	2.093	3.162	6.62

Example

The same ideas extend to more complex treatment structures.

In an experiment to compare two new drugs (A and B) for controlling blood pressure with a control (C) and with each other, experience suggests that $s^2 \approx 100$. It is required to estimate the difference in effect between A and B to within ± 2 and the difference between the average effect of A and B and the effect of C to within ± 6.

Example

$$
\begin{aligned}
s.e.(\bar{Y}_B - \bar{Y}_A) &= \sqrt{\frac{2s^2}{n_A}} \leq 1 \\
&\Rightarrow \quad \frac{200}{n_A} \leq 1 \\
&\Rightarrow \quad n_A \geq 200.
\end{aligned}
$$

$$
\begin{aligned}
s.e.\left\{\bar{Y}_C - \frac{1}{2}(\bar{Y}_A + \bar{Y}_B)\right\} &= \sqrt{\frac{s^2}{n_C} + \frac{1}{4}\frac{2s^2}{200}} \leq 3 \\
&\Rightarrow \quad \frac{100}{n_C} + \frac{1}{4} \leq 9 \\
&\Rightarrow \quad n_C \geq 11.4.
\end{aligned}
$$

So use $n_A = n_B = 200$ and $n_C = 12$.

Quantitative Treatment Levels

Example: A new enzyme has been developed which has potential industrial uses. An experiment is going to be conducted to study the effect of pH on the activity of the enzyme. The main objectives are to find the optimum pH and to study how robust the enzyme is to changes in pH. Which pH levels should be studied and how many runs should be made at each?

Quantitative Treatment Levels

When the treatments are levels of a continuous variable, it is often useful to fit response curves. Possible objectives in fitting curves:

- To estimate the parameters of a known (from scientific theory) model.
- To discriminate between a number of theoretical models.
- To predict the response, or changes in response, over a range of levels of the stimulus variable.
- To identify the level of stimulus which maximizes the response.
- To obtain a simple description of the relationship between the response and the stimulus.

Quantitative Treatment Levels

The design implications of the different objectives must be considered. In particular, we should ask:

- Which model(s) will we fit?
- Which quantities (parameters, functions of parameters) do we want to estimate from the models?
- Is the model mechanistic or empirical?

Mechanistic models, suggested by scientific theory, are usually nonlinear.

Quantitative Treatment Levels

We often have to use *empirical* models, where the form of relationship is not known. In this case the response functions used are often (but not always) linear, e.g.

$$E(Y) = \mu + t_r = \beta_0 + \beta_1 x + \beta_{11} x^2;$$

$$E(Y) = \beta_0 + \beta_1 \sqrt{x} + \beta_{11} x;$$

$$E(Y) = \beta_0 + \beta_1 \log x.$$

In general, we will write the linear model in matrix form as

$$E(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta},$$

so that

$$\mathbf{V}(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}.$$

Simple Linear Model

Example: When studying the environmental impact of cattle grazing it may be assumed that there is a linear relationship between nitrogen content of feed and nitrogen content of faeces. Consider estimating a specific parameter from this model, namely the slope. In the simple linear regression $E(Y) = \beta_0 + \beta_1 x$ we get the best estimate of the slope by minimizing $V(\hat{\beta}_1)$.

$$V(\hat{\beta}_1) = \sigma^2 / \sum_{i=1}^{n} (x_i - \bar{x})^2$$

and so we maximize $\sum_{i=1}^{n} (x_i - \bar{x})^2$.

Simple Linear Model

We do this by

1. increasing the number of experimental units;
2. for a fixed n, increasing the range of x values;
3. for a fixed range, taking half of the units at each end of the range.

Step 3 is an application of *optimal design theory*.

Quadratic Model

If there is more than one parameter of interest, the design which is optimal for estimating one parameter will usually be sub-optimal for estimating another.

In our initial example, we may decide to fit a quadratic model,

$$E(Y) = \theta_0 + \theta_1(pH) + \theta_{11}(pH)^2,$$

between pH 4 and 5.5.

It is good practice, for numerical stability and ease of interpretation, to work with *coded* levels of continuous treatment factors, so that the coded levels are between –1 and 1. Here, we use $x = \frac{pH-4.75}{0.75}$ and fit the model

$$E(Y) = \beta_0 + \beta_1 x + \beta_{11} x^2.$$

Quadratic Model

Consider a number of alternative designs with 12 experimental units.

Design	$V(\hat{\beta}_1)/\sigma^2$	$V(\hat{\beta}_{11})/\sigma^2$
$6(-1, 1)$	0.083	∞
$3(-1), 6(0), 3(1)$	0.167	0.333
$4(-1, 0, 1)$	0.125	0.375
$5(-1), 2(0), 5(1)$	0.1	0.6
$3(-1, -\frac{1}{3}, \frac{1}{3}, 1)$	0.15	0.422
$(-1, -\frac{9}{11}, -\frac{7}{11}, -\frac{5}{11}, -\frac{3}{11}, -\frac{1}{11}, \frac{1}{11}, \frac{3}{11}, \frac{5}{11}, \frac{7}{11}, \frac{9}{11}, 1)$	0.212	0.576

Different designs minimize the variances of different parameters. We might also consider covariances. We need a *design efficiency criterion*, a scalar function of the design matrix $\mathbf{X}$ which can be used to compare designs.

Design Efficiency Criteria

The appropriate criterion for comparing designs for a specific experiment should be closely related to the objectives of that experiment.

Several criteria are related to "generally useful" objectives.

- Weighted-A-efficiency minimizes a weighted mean of the variances of the parameter estimates,

$$w_1 V(\hat{\beta}_1) + \cdots + w_p V(\hat{\beta}_p).$$

This is equivalent to minimizing

$$tr\{\mathbf{A}(\mathbf{X}'\mathbf{X})^{-1}\},$$

where $\mathbf{A}$ is a diagonal matrix with ith diagonal element w_i.

Design Efficiency Criteria

- A-efficiency, with equal weights, is sometimes defined, but is scale-dependent, e.g. designs have a different ordering if the factors are coded or uncoded. Hence, *there is no such thing as an A-optimal design*, only an A-optimal design with respect to a particular parameterization.
- c-efficiency minimizes the variance of the estimate of a linear function of parameters, $V(\mathbf{c}'\hat{\boldsymbol{\beta}})$. This is equivalent to minimizing $\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}$.

Design Efficiency Criteria

- L-efficiency minimizes a weighted mean of the variances of the estimates of several linear functions of parameters,

$$w_1 V(\mathbf{c_1}'\hat{\beta}) + \cdots + w_q V(\mathbf{c_q}'\hat{\beta}).$$

This is equivalent to minimizing

$$\begin{aligned} & w_1 \mathbf{c_1}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c_1} + \cdots + w_q \mathbf{c_q}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c_q} \\ = \; & tr\{\mathbf{C}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}\} \text{ where } \mathbf{C} = [\sqrt{w_1}\mathbf{c_1} \cdots \sqrt{w_q}\mathbf{c_q}] \\ = \; & tr\{\mathbf{L}(\mathbf{X}'\mathbf{X})^{-1}\} \text{ where } \mathbf{L} = \mathbf{C}\mathbf{C}' \end{aligned}$$

Design Efficiency Criteria

- E-efficiency minimizes the maximum variance of an estimate of a (scaled) linear function of the parameters

$$\max_{\mathbf{c}} \frac{V(\mathbf{c}'\hat{\boldsymbol{\beta}})}{\mathbf{c}'\mathbf{c}},$$

  which is equivalent to minimizing

$$\max_{\mathbf{c}} \frac{\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}}{\mathbf{c}'\mathbf{c}}.$$

Design Efficiency Criteria

- D-efficiency minimizes the volume of a joint confidence region of the parameters, which is equivalent to minimizing $|(\mathbf{X}'\mathbf{X})^{-1}|$. This is equivalent to maximizing $|\mathbf{X}'\mathbf{X}|$. This last equivalence makes the D criterion particularly easy to compute.
- D_s-efficiency minimizes the volume of a joint confidence region of a subset of the parameters, which is equivalent to minimizing $|[(\mathbf{X}'\mathbf{X})^{-1}]_{22}|$, the part of $|(\mathbf{X}'\mathbf{X})^{-1}|$ corresponding to the same subset of the parameters. This is equivalent to maximizing

$$\frac{|[\mathbf{X}'\mathbf{X}]|}{|[\mathbf{X}'\mathbf{X}]_{11}|}.$$

There is a large body of optimal design theory, especially with regard to D efficiency. We must always remember that its relevance to a particular experiment depends on the objectives of that experiment.