INTERNAL REPORT

Statistical aspects of experimental designs implemented in designs

Kristian Kristensen
Biometry Research Unit, Department of Animal Breeding and Genetics,
Danish Institute of Agricultural Sciences

September 2003

Introduction

In the project BAR-OF many trials had to be performed under special circumstances. Either there were many varieties to be compared in a single trial, trial to be laid out in existing trials or multifactorial trials where the effects were given different priority. This means that traditional randomised complete block designs would be inefficient in most cases. Therefore, special trials were designed or found in the literature for most of the trials.

Types of design

WP1: α-designs

The trials in WP1 are all laid out as α-designs. If traditional randomised complete block designs were used then each of the blocks would be very large and it is feared that the random variation between plot could be very large. Hence α-designs was used. As the α-designs are resolvable a traditional analysis as for complete blocks are still valid, but an analysis taking the incomplete block effects into account is expected to be more efficient. If the block effects can be regarded as random it is recommended to use a model were the block effects within replicates are assumed random. The recommended model is:

$$Y_{vrb} = \mu + \alpha_v + \beta_r + C_{rb} + E_{vrb}$$

where

Y_{vrb} is the response recorded for variety *v* i block *b* of replicate *r*

E_{vrb} and C_{rb} asumed independent and normal distributed with constant varians, σ^2_E and σ^2_C

In SAS the model to use can be written as:

```
Proc Mixed;
class r b v;
Model y=r v/ddfm=satterth outp=out;
Random b(r);
Lsmeans v;
Run;quit;
```

Pair wise comparisons between all varieties can be carried out by adding "/diff" just before the semicolon in Lsmeans statement.

Note that the simple variety means cannot be used in connection with this model. In stead the values $\hat{\mu}+\hat{\alpha}_v$ should be used. These values can be regarded as means corrected for block effects, i.e. if the variety happened to be in blocks with high fertility then the simple means are too high and must be corrected downwards. Similarly if the variety happens do be en blocks with low fertility then the simple means are too high and must be corrected upwards. Only if the variety happens to be in blocks with average fertility will the simple means be identical to the estimates using the expression $\hat{\mu}+\hat{\alpha}_v$.

WP2: Split-plots in combination with α-design

In WP2 2003 the trial should contain all 64 combinations of 2 nitrogen levels, ± herbicide, ±Mechanical weed control and 8 "varieties" (4 pure varieties, 3 mixtures of 2 pure varieties and 1 mixture using 3 pure varieties) in 3 replicates, i.e. a trial with 192 plots. The design constructed was a split-plot design with the 8 combinations of N levels, herbicides and mechanical weed control forming the whole plots. Each whole plot consists of 2 neighbouring rows with 4 plots in each. The sub-plots contained the "varieties". However, as the 8 plots in each whole plot had to be located in two rows an α-design were used to optimise the comparisons between varieties within whole plots. An α-design with r=8, v=8, k=4 and s=2 were used in each replicate. As each replicate were formed by 8 columns a further restrictions were added on the design in order to ensure that "variety" comparisons were independent of column effect, i.e. ensuring that all "varieties" occurred once within each column of each replicate.
The statistical model for the design is:

$$Y_{grcnhmv} = \mu + \alpha_n + \beta_h + \gamma_m + \delta_v +$$

[all 2 factor interactions]+[all 3 factor interactions]+[the 4 factor interaction]

$$E_g + F_{gnhm} + G_{gr} + H_{gc} + I_{grcnhmv}$$

where

$Y_{grcnhmv}$ is the response recorded for variety v in replicate g and treated with N-level n, herbicide h and mechanical weed control m (and located in row r and column c of the replicate)

E_g is the effect of replicate g

F_{gnhm} is the random effect of the whole plot with treatment combination nhm in replicate g

G_{gr} is the random effect of incomplete block (in teh α-design) r in replicate g

H_{gc} is the random effect of column c in replicate g

$E_{grcnhmv}$ is the random effect of the plot with variety v in replicate g and treated with N-level n, herbicide h and mechanical weed control m (and located in row r and column c of the replicate)

All random effects are assumed to be normal distributed with mean zero and constant variances:

$\sigma^2_{\mathrm{E}}, \sigma^2_F, \sigma^2_H$ and σ^2_I, respectively

The greeck letters identify systematic effects.

In SAS the model may be written as:

```
proc mixed;
class g r c n h m v;
model y=n|h|m|v/ddfm=satterth outp=out;
random g g*n*h*m r(g) c(g);
lsmeans n|h|m|v@2;
run;quit;
```

Pair wise comparisons between the levels can be carried out by adding "/diff" just before the semicolon in Lsmeans statement.

Note that the simple variety means and other means for different varieties cannot be used in connection with this model. If no observations are missing then means over varieties are unbiased. The design is constructed such that comparisons between "varieties" and interactions with "varieties" should be as effective as possible. The main effects of Nitrogen level, Herbicide, Mechanical weed control and their interactions is expected to be less effective, partly because less degree of freedom is available for testing these effects and partly because we expect a larger variation between whole plots that between subplots.

WP3: Randomised complete block design

In this trial there were 27 treatments, consisting of 9 disease applications in combination wits 3 varieties laid out I 3 replicates. It was decided to use a traditional randomised complete block design. The statistical model is:

$$Y_{rvt} = \mu + \alpha_r + \beta_v + \gamma_r + (\beta\gamma)_{vt} + E_{rvt}$$

where

Y_{rvt} is the response recorded for treatmnet t to variety v in replicare r

μ is the overall mean of the response in the trial

$\alpha_r, \beta_v, \gamma_r$ and $(\beta\gamma)_{vt}$ is the fixed effect of replicate, varieties, treatemnts and the interaction between varieties and treatments

E_{rvt} is the random effect of plots. E_{rvt} is assumed to be normal distributed withe the mean zero and variance σ^2

In SAS the model may be written as:

```
proc mixed;
class r v t;
model y=r v t v*t/outp=out;
lsmeans v|t;
run;quit;
```

Pair wise comparisons between the levels can be carried out by adding "/diff" just before the semicolon in Lsmeans statement.

WP4: Split-plot in combination with an unresolvable incomplete block design

In WP4 the trial had to be laid out in an existing long-term experiment. The trials were performed in 2 of the available fields. I both fields 4 long term treatments were available in 4 replicates. There was a strong wish to have 8 "varieties" (6 pure varieties and 2 mixtures of 3 varieties each) included in the experiment. However, the size and shape of each plot in the existing long-term experiment could not afford more than 6 subplots in each. The total number of subplots for each treatment in each field was thus 24, which made it possible to have 3 replicates of each combination of field, treatment and "variety". Therefore it was decided to construct an incomplete block design that could allow the varieties to be compared even though they could not all be accommodated in all whole plots (the original 4 treatments in 4 replicates in each field). The model for the design could be written as:

$$Y_{fbtv} = \mu + \alpha_f + \beta_t + \gamma_v + (\alpha\beta)_{ft} + (\alpha\gamma)_{fv} + (\beta\gamma)_{tv} + (\alpha\beta\gamma)_{ftv} + C_{fb} + D_{fbt} + E_{fbtv}$$

where

Y_{fbtv} is the response recorded for variety v in treatment t in block b of field f

$\alpha_f, \beta_t, \gamma_v, (\alpha\beta)_{ft}, (\alpha\gamma)_{fv}, (\beta\gamma)_{tv}$ and $(\alpha\beta\gamma)_{ftv}$ is the main and interaction effects of field, treatment and "variety"

C_{fb} is the random effect of replicates for the original treatments within field

D_{fbt} is the random effect of plots in the original experiment (becomes now whole plots)

E_{bftv} is the random effect of subplots

All random effects are assumed to be normal distributed with mean zero and constant variances:

σ_C^2, σ_D^2 and σ_E^2, respectively

The greeck letters identify systematic effects.

In SAS the model may be written as:

```
proc mixed;
class f b t v;
Model y=f|t|v/ddfm=satterth outp=out;
Random b(f) b*t(f);
Lsmeans f|t|v@2;
```

Pair wise comparisons between the levels can be carried out by adding "/diff" just before the semicolon in the Lsmeans statement.

Note that the simple variety means and other means for differences between varieties should not be used in connection with this model, as they are not the most efficient estimates. It can be tested weather the difference between fields is larger than what can be expected from the random variation within field; but this is not a test for weather the difference found is caused by the history of those two fields. However, it is possible to test the interaction between fields and e.g. varieties and to compare the varieties within each field. The effects of varieties and interactions with varieties are the effects that are expected to be estimated most efficiently.