Analysing Data from Participatory On-farm Trials

Richard Coe [1]

Summary:
Researchers carrying out participatory on-farm trials, particularly variety selection trials, often have difficulty analyzing the resulting data. The irregularity of designs means that some of the standard tools based on analysis of variance are not appropriate. In this paper some simple extensions to analysis of variance, using general linear models and linear mixed models, are shown to facilitate insightful analysis of these awkward designs.

Key words: *Variety selection trials, analysis of variance*

Résumé:
Les chercheurs effectuant des essais participatifs en milieu réels, en particulier des essais de sélection variétale, éprouvent souvent des difficultés à analyser les données qui en résultent. Le manque de régularité dans les dispositifs expérimentaux fait que certains des outils standards basés sur l'analyse de variance ne sont pas appropriés. Dans cet article, il est présenté des prolongements simples de l'analyse de variance, en utilisant les modèles linéaires simples et des modèles mixtes pour permettre une analyse plus appropriée des dispositifs expérimentaux compliqués.

Mots clés: *Essais de sélection variétale, analyse de la variance*

1. Introduction

Data from on-farm trials is of many types, from crop yields measured on individual plots to the reported consensus of participants at a group meeting. Any set of data that includes multiple observations which are not all identical will require some sort of statistical analysis in order to summarise the common patterns. Choice of appropriate analysis methods depends on:

1. The objectives of the analysis.
2. The design (who compared what treatments or varieties under which conditions).
3. The type of measurements taken.

[1]ICRAF, Nairobi, Email: r.coe@cgiar.org

In Section 2 of this paper I discuss differing styles and objectives of analysis. A formal approach, similar to that commonly carried out, for example, on crop yields measured in a classical variety trial using analysis of variance and reporting variety means, has a role in analysis of some participatory trials. The irregularity of designs often means that the well-known methods may be inappropriate. In Section 4 I show how some extensions to the usual methods can be used. Many researchers report that results from on-farm trials are highly variable. Section 5 shows how some of this variation may be interpreted to gain further insights, particularly into differing responses in different situations, or G x E interaction. Examples used to illustrate the methods are introduced in Section 3. The methods described in this paper are appropriate for responses measured on a continuous scale, such as a crop yields. Analysis of responses recorded as scores or ranks are the subject of a companion paper (Coe 2007).

The methods described in this paper are not new nor described in depth. Technical descriptions can be found in numerous publications including Kempton and Fox (1997) and Hildebrand and Russell (1998).

2. Approaches to Analysis

An assumption of this paper is that participation and the systematic collection, analysis and interpretation of data are not contradictory activities. Among some practitioners there is a belief that adoption of a participatory paradigm removes the need for, or even makes it impossible for, researchers to collect and analyse data. The purpose of participation is seen as empowerment of local people, which is inconsistent with researchers conducting activities that meet their own objectives. However many researchers recognise that generalisable conclusions, of relevance beyond the immediate participants, are still necessary, and that part of this research must be the collection and interpretation of data. Coe and Franzel (2000) summarise the research design principles that must still be followed if the research is to lead to valid inferences.

A participatory approach does, however, have implications for the collection, analysis and presentation of data. Data collection is discussed elsewhere. Analysis can be for and, to some extent by, different participants, each of whom will have their own interests and objectives. In the case of participatory crop breeding trials, participants might include farmers, researchers, extension staff and regional planners. While a farmer is interested in making decisions about varieties to select for their own farm, a regional planner might be interested in average performances and a researcher in

reasons for heterogeneous responses. Each will require a different type of analysis. As the researcher is often also the facilitator of the whole process it is the researcher's responsibility to ensure each participant has the data they need in a format that is useful.

It is particularly important for a researcher to make data and results available to farmers. There are at least 3 reasons

1. Farmers are supposed to be beneficiaries of the activities, and can only benefit if information gets back to them.
2. Giving farmers results is a courtesy as they have made the research possible though their involvement.
3. Farmers can provide considerable insights into the analysis and results. It is very common to hear the complaint that 'data from on-farm trials are very variable'. This variation is part of the real world, and understanding its causes should be an objective of the research. Such an understanding will eventually lead to improved farmer decision making. Farmers understand some of the reasons for the variation and their insights can often provide a framework or hypotheses for analysis.

When plant breeders did classical, on station experiments, the analysis done often followed a standard pattern – for example analysis of variance followed by tabulation of means and application of 'means separation procedures'. Often little attention was paid to exploratory analysis, designed to detect the main patterns and surprising observations. Nor was much effort made at imaginative presentation of results – researchers knew how to read the tables and they were the intended audience. When participatory approaches gained popularity, analysts made attempts to find interesting and informative presentations of data, but tended to forget about formal analysis and hence sometimes reached invalid conclusions.

Of course both approaches to analysis are needed and reinforce each other. Graphical and exploratory method show the important results, reveal odd observations and unexpected patterns. Formal methods allow measures of precision to be attached to results and allow extraction of estimates from complex data structures. We can not say that either approach is better - both are needed in differing roles. In this document I have concentrated on formal analysis as requested. It is also easier to find general methods and approaches that can be described and applied in many situations.

Presentation and analysis are not the same. The method of presenting results will depend on the nature of the result and the story that is to be told with them, as well as the audience. I am not aware of any work that shows

that literate farmers find it easier to interpret graphs than numerical information. Indeed it seems likely that a simple numerical table may be more familiar than a quantitative graph.

The steps in analysis of any data set can be summarised as:
1. Define the analysis objectives. These drive the rest of the analysis. It is impossible to do a good analysis of data without clear objectives. Often the key graphs and tables can be defined at this stage, even without having results yet to fill them in.
2. Prepare the data. Data sets will have to be entered and checked, suitable transformations made (eg to dry weight per unit area), relevant information from different sources (eg farm household data and plot level yields) extracted to the same file, and so on.
3. Exploratory and descriptive analysis. The aim is to summarise the main patterns and notice further patterns that may be relevant.
4. Formal statistical analysis. The aim is to add measures of precision and provide estimates from complex situations
5. Interpretation and presentation.

Iteration between the steps will be necessary. Training materials by Coe et al (2001) provide much more information on analysis of experiments.

A spreadsheet package such as Excel is useful for much of the descriptive analysis. Flexible facilities for data selection and transformation, tabulation and graphics are useful. However dedicated statistical software is needed for the analyses described here. They can not be done in Excel. There are several packages with roughly equivalent facilities. All the examples given use Genstat (2000) as I find it often the easiest to understand, particularly as methods for different problems can be addressed with a similar set of commands and most convenient. SPSS is widely used by social scientists but is not particularly useful for the analyses described here. There are further comments on software at the end of this paper.

The example results in this paper have been copied directly from the outpt produced by the software. These would normally need some editing and improved presentation for publication.

3. Examples

1. *Agroforestry of soil fertility in Malawi*
This is not a breeding trial but is included as the design is typical of many participatory on farm trials. Three soil fertility strategies are compared

over a number of years:
g – mixed intercropping of maize and gliricidia
s – relay planting of maize and sesbania
c – the control of continuous maize.

41 farmers each compared the control with one or both of the other treatments. Crop yield is the response of interest. A number of covariates were measured at the plot or farm level to help understand the reasons for variation across farms. In the analyses below, the data structure 'name' identifies the farmers, 'trt' the treatments to compare and 'yield98' is the yields response of interest.

2. *Maize varieties in Zimbabwe*
This was a 'baby' trial. 12 maize varieties were compared. 146 farmers in 25 different sites took part, each one testing 4 of the 12 varieties. The varieties for each farmer to test were chosen by the researcher. Some household and field covariates were recorded.

The actual crop yields obtained were not available for analysis, so the examples here use simulated yield data but the original field design. In the analyses below, the data structure 'FARM' identifies farmers, 'ENTRY' the varieties to compare and 'simyield' the response of interest.

4. Average treatment effects

Example 1

	Data	
trt	**Average of yield 98**	**Count of yield 98**
c	1.73	31
g	2.47	39
s	2.50	24
Grand Total	**2.23**	**94**

The starting point for the analysis should be simple explorations, such as the table of means above that gives the mean yield in the 98 season for each treatment, together with the number of observations.

The formal analysis aims generally to do two things:
1. Improve the estimates. In this case we know that the treatments do not all occur on each farm so some adjustment for farm effects may be needed (see Example 2 for more on this).

2. Provide measures of precision – standard errors and confidence intervals.

This is the role of analysis of variance and associated procedures in 'regular' designs. Exactly the same ideas can be used here.

Genstat commands to complete the analysis are:

```
model yield98
fit [p=a;fprob=y] name+trt
predict trt


***** Regression Analysis *****

*** Accumulated analysis of variance ***

Change                          d.f.         s.s.         m.s.      v.r.  F pr.
+ name                            38     168.6518       4.4382     13.39  <.001
+ trt                              2      15.9187       7.9594     24.01  <.001
Residual                          53      17.5691       0.3315

Total                             93     202.1396       2.1735


 Response variate: yield98
              Prediction        S.e.
          trt
            c      1.6386      0.1066
            g      2.6235      0.0952
            s      2.3677      0.1240
```

Standard errors of differences (sed) can also be found. They are:

	sed
g - c	0.145
s - c	0.166
g - s	0.160

While this analysis is correct and technically efficient it is possibly a little opaque! An alternative which is easier to understand follows.

The researcher is interested in comparison of the treatments and in the change in performance (eg yield) realizable by changing from one treatment to another. Farmers are also interested in this comparison, though the criteria for comparison may be different from those of researchers. Experiments are designed to assess this change. It is therefore natural to approach analysis of the data by focusing on these changes. The steps are:

1. Choose a treatment pair the comparison of which is of interest – eg (maize intercropped with gliricidia) and c (sole maize).
2. For each farm on which this pair occurs calculate the difference in response g – c.
3. Summarise this set of differences.

In this trial 31 farms have yield data for this pair of treatments in 1998. The column of differences is y98g_c.

```
Summary statistics for y98g_c

      Number of observations = 31
    Number of missing values = 10
                        Mean = 1.008
                      Median = 0.841
                     Minimum = -0.739
                     Maximum = 2.712
               Lower quartile = 0.400
               Upper quartile = 1.766
                    Variance = 0.791
          Standard deviation = 0.889
```

The mean difference of 1.008 has a standard error of $\sqrt{(0.791/31)} = 0.16$. A 95% confidence interval for the mean difference is thus $1.01 \pm 2 \times 0.16 = (0.69, 1.33)$. A statistical test of the hypothesis of no difference in mean yield from the two treatments would use the t statistic t=difference/se(difference) = 1.01/0.16 = 6.3. This mean, together with its standard error, is almost identical to that produced by the modelling analysis above.

Differences are due to:
(1) the modelling analysis uses part of the information from 3 farmers with sesbania and gliricidia but not the control treatment [If we can estimate g-s and s-c within farms then we also estimate g-c = (g-s)-(s-c)] and
(2) All the data is used to estimate the residual variance, not just part of it.

The summary statistics above emphasise the fact that looking at the mean difference is only the start of the analysis. There is considerable variation in the difference across different farms that needs understanding and interpreting. This is the subject of Section 5.

Example 2
The first step must be to check the data for errors and oddities. This is not illustrated.

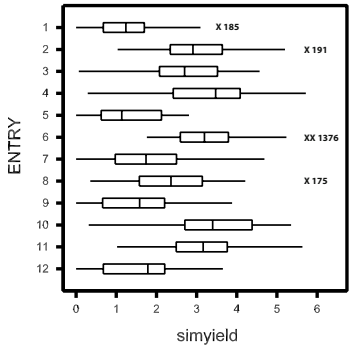
Next simple summaries - numerical and graphical – are needed. The following table gives the mean, 25% , 50% and 75% points for each entry

together with the number of plots from which it is calculated. Note Excel is very good for this type of tabulation but can not give the % points.

```
tabulate [class= ENTRY; p=nobs,means,quant; percent=!(25,50,75)] data=simyield
```

ENTRY	Nobservd	Mean	_25.0%	Median	_75.0%
1	50	1.276	0.679	1.238	1.699
2	47	3.077	2.344	2.909	3.639
3	47	2.713	2.076	2.699	3.521
4	50	3.305	2.416	3.473	4.083
5	49	1.323	0.624	1.138	2.124
6	49	3.371	2.594	3.195	3.792
7	50	1.760	0.973	1.742	2.499
8	49	2.429	1.573	2.362	3.143
9	42	1.436	0.659	1.584	2.202
10	51	3.448	2.708	3.401	4.380
11	50	3.099	2.494	3.165	3.761
12	50	1.597	0.677	1.788	2.206

Similar information is presented graphically in a boxplot:



This particular boxplot has highlighted some outlying observations which should be checked for possible errors.

These overall summaries are unlikely to be of interest to farmers in any one location, but the data from their neighbourhood should be very relevant. A simple table of farm by entry for each site may be a useful tool for discussion with this group of 8 farmers, as it highlights both variation between entries and the variation between farmers who test the same things. It is likely that farmers can tell you something about the reasons for variation. This may help direct formal analysis. For example, if they identify some of the low yields as coming from plots known to be infertile, some measures of fertility should be built into the formal analysis. They may also be able to tell you something about the tradeoffs between different assessment criteria – for example, expressing satisfaction with a variety which is not the highest yielder because of some other desirable property. The data may need converting to units that farmers use and understand.

SITE	1

Average of simyield	FARM								
ENTRY	1	2	3	4	5	6	7	8	Grand Total
1			2.03			1.70			**1.86**
2			3.39	2.43			2.63		**2.82**
3	1.51				2.66		2.11	1.81	**2.02**
4			4.97		3.36	4.01			**4.11**
5		0.28		0.29		1.55		0.74	**0.72**
6	3.06				2.35			1.96	**2.45**
7		0.45					1.82		**1.13**
8		2.00							**2.00**
9		0.00		1.77					**0.89**
10			4.47			3.15			**3.81**
11	2.06			2.40			1.02		**1.83**
12	1.40				1.79			0.40	**1.20**
Grand Total	**2.01**	**0.68**	**3.72**	**1.72**	**2.54**	**2.60**	**1.89**	**1.23**	**2.05**

Any of these tables can be rearranged to clarify important information – for example, sorting by mean may make the table easier to read:

SITE	1

Average of simyield	FARM								
ENTRY	1	2	3	4	5	6	7	8	Average
4			4.97		3.36	4.01			**4.11**
10			4.47			3.15			**3.81**
2			3.39	2.43			2.63		**2.82**
6	3.06				2.35			1.96	**2.45**
3	1.51				2.66		2.11	1.81	**2.02**
8		2.00							**2.00**
1			2.03			1.70			**1.86**
11	2.06			2.40			1.02		**1.83**
12	1.40				1.79			0.40	**1.20**
7		0.45					1.82		**1.13**
9		0.00		1.77					**0.89**
5		0.28		0.29		1.55		0.74	**0.72**
Average	**2.01**	**0.68**	**3.72**	**1.72**	**2.54**	**2.60**	**1.89**	**1.23**	**2.05**

The formal analysis of this data is needed to give means corrected for site and farm effects, together with correct standard errors of differences. The usual starting point would be an analysis of variance. However the analysis has to account for expected variation due to differences between farms and sites, and the design used ended up with a rather irregular distribution of varieties across farms and sites. For example, in Site 1 (table above) entries occur between 1 and 5 times. The design is described as 'unbalanced' (differing amounts of information about each treatment comparison) and treatments are 'non-orthogonal' to farms and sites. The later implies that treatment means adjusted for site and farm effects are more realistic summaries of treatment differences than raw means.

The need for some sort of adjustment can be seen by looking at the table above for Site 1. Entries 5, 7 and 9 have low means. However they all occur on farm 2, which may be a poor farm, hence depressing the means for these entries. Calculation of these adjusted means is described below. The results taking just the data for Site 1 show that the ranking of the Entries is changed considerably, but the logic of the changes is visible if compared with the data. For example, Entry 1 has the lowest adjusted mean. Looking at the raw data shows that this entry appeared on just 2 farms, both of which seem (from looking at the performance of other entries) to be good ones.

Entry	Raw mean	Adjusted mean
4	4.11	3.48
10	3.81	2.94
2	2.82	2.46
6	2.45	2.57
3	2.02	2.16
8	2.00	3.20
1	1.86	1.00
11	1.83	1.88
12	1.20	1.32
7	1.13	1.83
9	0.89	1.48
5	0.72	1.03

The adjusted means are found by fitting a model with farm and entry effects. This model can be used to predict the performance of each entry on

each farm, and the adjusted mean is then the average of these predictions across all the farms. The commands to do this in Genstat are simple, the last one being needed to get the standard errors of differences between adjusted means. The results below are now for the whole dataset, not just Site 1.

```
model simyield
fit [p=*] FARM+ENTRY
predict ENTRY
rpair !P(ENTRY)
```

```
 Response variate: simyield
               Prediction        S.e.
        ENTRY
            1        1.234       0.107
            2        2.878       0.111
            3        2.612       0.111
            4        3.328       0.107
            5        1.483       0.108
            6        3.305       0.108
            7        1.834       0.107
            8        2.423       0.108
            9        1.488       0.118
           10        3.409       0.107
           11        3.167       0.107
           12        1.667       0.107

 796  rpair !P(ENTRY)

***** Pairwise differences *****

***** Regression Analysis *****

 Response variate: simyield
     Fitted terms: Constant + FARM + ENTRY

Standard errors of pairwise differences
            1            *
            2       0.1549            *
            3       0.1560       0.1568            *
            4       0.1534       0.1569       0.1561            *
            5       0.1560       0.1561       0.1574       0.1528            *
            6       0.1543       0.1548       0.1564       0.1547       0.1535
            7       0.1535       0.1574       0.1561       0.1511       0.1562
            8       0.1533       0.1587       0.1570       0.1542       0.1565
            9       0.1565       0.1613       0.1608       0.1618       0.1617
           10       0.1524       0.1565       0.1599       0.1490       0.1486
           11       0.1557       0.1531       0.1518       0.1506       0.1518
           12       0.1494       0.1544       0.1541       0.1548       0.1544
```

```
                1          2          3          4          5

 6              *
 7         0.1538          *
 8         0.1550     0.1512          *
 9         0.1621     0.1600     0.1612          *
10         0.1536     0.1500     0.1494     0.1639          *
11         0.1516     0.1531     0.1550     0.1643     0.1549
12         0.1524     0.1523     0.1525     0.1605     0.1543

                6          7          8          9         10

11              *
12         0.1562          *

               11         12
```

Note that the seds are not all the same, due to the irregularity in the design. However they are close enough that it would be sense to quote a single sed of 0.16.

If these adjusted means are compared with the raw means the differences are not as great as when we analysed just one site. The means are averages over more farms, so the effects of 'good' and 'bad' farms on individual means tend to cancel out.

Entry	Raw mean	Adjusted mean
10	3.45	3.41
6	3.37	3.31
4	3.30	3.33
11	3.10	3.17
2	3.08	2.88
3	2.71	2.61
8	2.43	2.42
7	1.76	1.83
12	1.60	1.67
9	1.44	1.49
5	1.32	1.48
1	1.28	1.23

In this case the model could also have been fitted as

```
model simyield
fit [p=a] SITE/FARM+ENTRY
```

```
***** Regression Analysis *****

*** Accumulated analysis of variance ***

Change                          d.f.         s.s.         m.s.      v.r.
+ SITE                            24     189.0435       7.8768     16.57
+ SITE.FARM                      121     327.6509       2.7079      5.70
+ ENTRY                           11     289.1360      26.2851     55.28
Residual                         427     203.0184       0.4755

Total                            583    1008.8488       1.7304
```

This analysis of variance can be interpreted in the usual way, and shows that some of the between farm variation actually occurs between sites. That is, farms within a site tend to be more similar than farms in different sites, as expected.

The analysis presented above is valid. However it does not capture all the information in the data and hides some of the structure. An alternative approach is to treat sites and farms within sites as if there were a random selection from those available and use a model that describes this. REML procedures handle these problems and are easy to use in Genstat.

```
VCOMPONENTS [FIXED=ENTRY]  RANDOM=SITE/FARM
REML [PRINT=model,components,waldTests,means; PSE=differences] simyield
```

The option `FIXED=ENTRY` specifies that we want to estimate separate means for each of the entries. The parameter `RANDOM=SITE/FARM` tells Genstat that there are sites which are expected to vary and there are farms within each site which also vary. Genstat automatically adds the plot level or residual variance, but this could be put in explicitly if the dataset had another factor labelled PLOT by specifying `RANDOM=SITE/FARM/PLOT`. The output is shown below.

Note that the trial was originally planned with a 'replicate' being a set of all the varieties (spread across 3 farms) with 3 replicates per site. However due to lack of available land and some mistakes this is not how the design was implemented. 'Replicates' therefore do not correspond to any physical source of variation in the experiment so it does not make much sense to include them in the analysis. On the other hand both sites and farms correspond to physical layout factors that it is reasonable to expect will influence results, so these must be allowed for.

```
***** REML Variance Components Analysis *****

Response Variate: simyield

Fixed model     : Constant+ENTRY
Random model    : SITE+SITE.FARM

Number of units : 584
* Residual term has been added to model

*** Estimated Variance Components ***

Random term                Component        S.e.

SITE                          0.2516      0.0992
SITE.FARM                     0.3535      0.0616


*** Residual variance model ***

Term              Factor        Model(order)  Parameter         Estimate       S.e.

Residual                        Identity      Sigma2               0.475     0.0325


*** Wald tests for fixed effects ***


   Fixed term               Wald statistic      d.f.    Wald/d.f.    Chi-sq prob

* Sequentially adding terms to fixed model

   ENTRY                         663.07           11        60.28      <0.001


* Message: chi-square distribution for Wald tests is an asymptotic approximation
  (i.e. for large samples) and underestimates the probabilities in other cases.


*** Table of predicted means for Constant ***

            2.455    Standard error:  0.1165


*** Table of predicted means for ENTRY ***

   ENTRY        1        2        3        4        5        6        7        8
           1.308    2.984    2.681    3.369    1.495    3.377    1.858    2.478


   ENTRY        9       10       11       12
           1.528    3.469    3.205    1.704

Standard error of differences:     Average       0.1510
                                   Maximum       0.1585
                                   Minimum       0.1457

Average variance of differences:                 0.02281
```

The first part of the output reports variance components, interpreted in the next section.

The Wald test is equivalent to the F-test for treatment effect in the usual anova. The 'highly significant' effect says that there are real differences somewhere between these 12 variety means.

The table of predicted means gives means for each entry adjusted for farm and site effects. In this case most of the means are close to the unadjusted means. This will not always be the case. The adjustments allow for the fact that some farms are better (give higher average yields) than others. Entries that are tested on 'good' farms will have their means biased upwards compared with entries tested on 'bad' farms. In this design each entry is tested on about 50 farms, so the 'good' and 'bad' farms tend to cancel out. However if there were fewer farms this would not be the case. These predicted means are the ones that should be reported and interpreted, not the raw means presented earlier.

The sed values for comparing predicted means are not all equal, so Genstat reports the minimum, maximum and average. They are not equal because different pairs of means are compared with different precision. For example, counting shows that entries 1 and 2 occur together in the same farm 14 times. Treatments 9 and 10 only occur together on the same farm only 5 times. We would therefore expect the sed for comparing 1 and 2 to be lower than that for comparing 9 and 10. In this case the range in seds is not large, so we do not go far wrong if the average (or, more conservatively, the maximum) is used.

The output does not contain information on which entries are different for which other ones, just that overall there are some variety differences. We have not put any information about possible differences between entries into the analysis, so the only possibility would be an analysis based on ignorance, for example one that attached letters to varieties deemed to be 'not significantly different' from each other. There are both technical and philosophical problems with this approach and it should be avoided.

Suppose that the entries came from 3 groups, depending on pedigree, as follows.

Group	a	b	c
Entry	1, 5, 7, 9, 12	4, 10, 11	2, 3, 6, 8

Then we can look for differences between and within groups by replacing the fixed model by FIXED=GROUP/ENTRY.

```
*** Wald tests for fixed effects ***

   Fixed term              Wald statistic      d.f.    Wald/d.f.    Chi-sq prob

* Sequentially adding terms to fixed model

   GROUP                        602.80            2       301.40      <0.001
   GROUP.ENTRY                   60.27            9         6.70      <0.001


* Message: chi-square distribution for Wald tests is an asymptotic approximation
  (i.e. for large samples) and underestimates the probabilities in other cases.


*** Table of effects for GROUP ***

   GROUP         a         b         c
           0.000     2.061     1.676

Standard error of differences:     Average       0.1506
                                   Maximum       0.1521
                                   Minimum       0.1490
```

The Wald tests show there is considerable variation between groups of entries, but still some remaining variation between entries within a group. The table of effects for GROUP summarises the difference between groups – entries in Group b have means yields 2.06 higher than those in Group a.

Comparing approaches

In Example 1 we based an analysis of the difference between yields of two treatments either on a linear model or on the set of difference within each farm. The two methods gave almost identical results. Why not use the difference method in the Example 2? Some of the reasons are:

1. With the 3 treatments of Example 1 there are 3 pairs of treatments that could be used to form differences. Hence we might repeat the analysis 3 times. These analyses are not independent but that does not matter. However with the 12 treatments of Example 2 there are $12 \times 11/2 = 66$ pairs that we could choose to make differences. Analysis of all these would not only be tedious, it would involve a lot of repetition of the same information (there are only 11 df in 12 treatments). But which subset of pairs should be chosen?
2. The set of treatments of any farm is small – only 4 out of 12. Thus, for example, treatment 1 occurs on 50 farms and treatment 2 on 47, yet they occur together on only 14. So if we work with the Entry 1- Entry 2 difference we would use data from just 14 farms. However there is much more information about the two treatments. This is reflected in the

differing seds from the two approaches. Modelling gave an sed of 0.155 for Entry 1 – Entry 2 and the difference method gives 0.180. This may seem a small change but is equivalent to a 42% increase in trial size.

3. Other limitations of the difference methods will be described later.

The difference between the analysis that takes farms and sites as 'fixed' and the REML analysis that takes them as random lies in what it is reasonable to assume about farm and site differences. If they are different, but we can make no realistic assumptions about the nature of those differences then they should be taken as 'fixed'. This means each site or farm has its own characteristic mean, unconnected with any other, and these have to be estimated. The information on treatment differences then comes from differences within each farm. However if sites or farms can be thought of as a random sample from the set of possible sites or farms, and have effects which roughly follow a normal distribution, then we estimate the variance of that normal distribution. This changes the estimates of the treatments effects as between-farm and between-site information is recovered. The source of this information can be understood as follows. If all farms that had treatment 1 had a high mean, and all those that had treatment 2 had a low mean, it is evidence that treatment 1 is better than treatment 2. If farms really are a random sample then treatment 1 is unlikely to end up on all the best farms by chance. Hence there is some information from the farm effects to add to our evidence for treatment 1 having a higher mean than treatment 2. The REML method combines this information with the within-farm information. The estimates of treatment effects, and sed values are therefore modified compared with the earlier fixed effect analysis. If the assumptions of the random site and farm effects are realistic then this analysis will always be more efficient.

5. Understanding variation and GxE interaction

Example 2
The analysis above has produced estimates of variance components as follows:

component	estimate	standard error
SITE	0.2516	0.0992
FARM	0.3535	0.0616
PLOT or residual	0.4750	0.0325

What do these tell you?

The model used to analyse the data, as specified in the VCOMPONENTS command, is
yield = mean + site effect + farm effect + variety effect + residual
The residual is thus the deviation of an individual plot yield from the average for that site, farm and variety. It is all the unexplained variation from plot to plot, due to local environmental effects (soil, pests), management, measurement error and so on. The variance of 0.475 says the standard deviation of this plot to plot variation is $\sqrt{0.475} = 0.698$. If the data have roughly a normal distribution then most observations lie within 2 standard deviations of the mean. Thus the plot to plot variation represents variation of about ±1.4 about the mean for a farm growing a uniform variety. This is a typical level of variation in such trials.

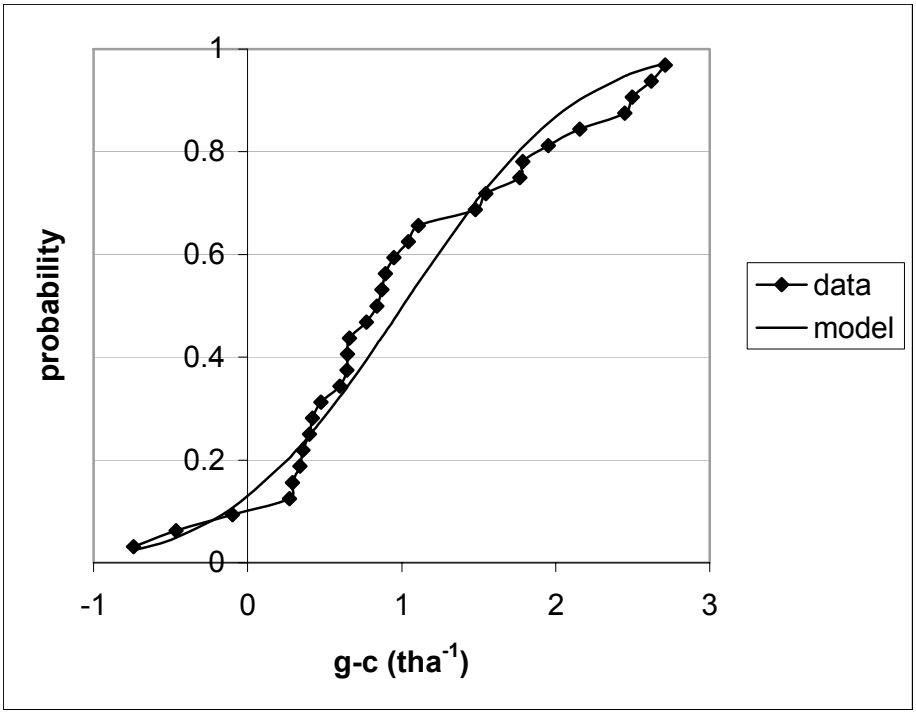
The farm variance can similarly be interpreted. It shows how much the average yield for a very large number of plots varies between farms within the same site.

Explaining variation – interaction and risk

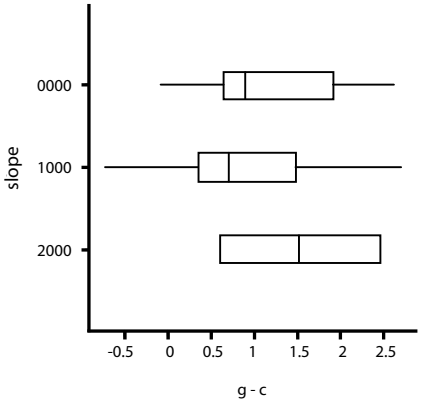
Example 1
In Section 4 we analysed Example 1 by taking the 31 differences in yield for g-c and looking at their mean and variation. Here I want to take this analysis further.

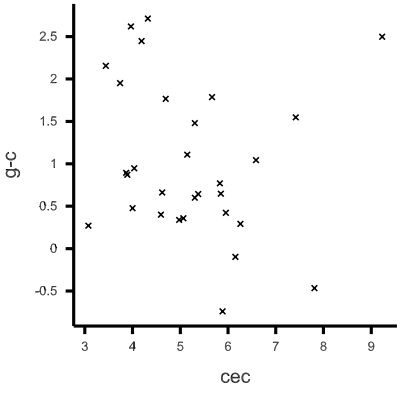
The mean difference of 1.01 tha-1 is of course of interest in some analyses and this is the quantity most often reported, together with a proud statement that it is 'significantly' greater than zero. However for an individual farmer it is uninteresting. A farmer's decision on whether to use g rather than c will depend on many things, but the yield component of the decision will be based on the yield increase they might achieve *on their farm*. In the absence of any other information the mean is the best guess for what this might be, but there is of course a lot of variation around this mean. This variation is an indication of the level of risk of a decision based on the mean. In the figure below the risk of obtaining a yield increase less than any specified amount is plotted. There is roughly a 10% change that a farmer will achieve lower yield with g than c and 55% chance of and increase less than 1 tha-1. However 20% of farmers achieve an increase of more than 2 tha-1. A simple model for the variation is obtained by assuming a normal distribution, also shown on the graph. It is not a particularly good fit but still has some value, explained later. Note that if there were many more than 31 farmers in the study we would expect a better (more precise) estimate of the mean difference between g and c but would not

expect a reduction in the variation in this difference across farms. More farms would give a better estimate of the chance of getting less yield with g than c, but it would still be around 10%.
Knowing what distinguishes a +2 tonne farmer from a +0 tonne farmer is important, both for the farmer's decision making and the researcher's understanding.



An approach to the problem should be clear. We have a set of 31 differences and want to know what determines them. Hypotheses of possible causes may come from farmers or researchers. The hypotheses are then tested by collecting suitable data and statistical analysis. Suppose that the slope of the plot and cec of the soil are hypothesised causes of the variation in this case so we can explore evidence for that in the data. Slope is in this case a categorical variable. A boxplot may be useful – it shows little evidence of a consistent difference in the size of g-c for different slope categories. Likewise we could use a scatter plot of g-c against cec. This does not show a clear relationship but does show some 'outlying' points which could be followed up. The farm in the top right of the scatter, for example, used fertilizer, suggesting further ideas for investigation.





Note that what we are doing here is identifying GxE interaction. The Gs are just the 3 treatments and the Es are characterised by slope and cec.

A formal statistical analysis would now use usual regression modelling approaches to quantify any effects. If y_{ij} is the yield on farm i under treatment j then the differences being analysed are

$d_i = y_{ig} - y_{ic}$ with variance σ_d^2. This is the variation reflected in the graph above and in the simple 'risk 'model.

A regression model to look at the effect of a farm level covariate x would then be:

$$d_i = c + b_{gc} x_i + e_i$$

Here b_{gc} is the regression effect when considering the g-c difference and e_i the residual. The variance of the residual is σ_r^2. This measures the still unexplained variation in d, or the risk still remaining with knowledge of the covariate. Again, if a normal distribution model is acceptable then the parameters of the regression model, with σ_r^2, allow predictions of the risk of changes in yield in switching from c to g conditional on the value of the covariate.

The usual analysis of variance model for this data, with treatments and farms in the design, would be:

$y_{ij} = c + f_i + t_j + e_{ij}$ with the variance of these residuals σ^2. Then the g-c differences are

$$d_i = t_g - t_c + e_{ig} - e_{ic}$$

The connection between the 'analysis of variance' approach and the analysis of plotwise differences then becomes clear. The variance of the differences $\sigma_d^2 = 2\sigma^2$. The effect of the covariate could be included in the analysis of variance model as

$$y_{ij} = c + f_i + t_j + b_j x_i + e_{ij}$$

The term $b_j x_i$ describes how the treatment effect is modified on farms of different type (ie with different values of the covariate x). It is thus a treatment by farm interaction, and is often the basis of the most useful results from an on farm trial. With information on such interactions we can refine predictions and recommendations and reduce the risk associated with decisions based on the data. The covariates useful for this may be social variables (gender, household size etc), biophysical variables (soil type, slope etc) or management variables (weeding, planting time etc).

A common misunderstanding in experimental design is that farm x treatment interaction can not be detected if only a single replicate is placed on each farm. The types of farm x treatment interaction that are important are those that are structured in some way to show consistent patterns across farms. These can be explained and predicted in terms of explanatory variables, and can be estimated from designs with no more than 1 replicate per farm, as shown here. This does not mean design is unimportant. If is known which covariates will be of interest before the trial starts then more effective designs can be used.

The analysis above identifies and describes what has always been known by breeders as GxE (genotype x environment) interaction. The classical approach to this has been a 'complete' trial in each of a number of locations, these locations representing different environments. Once a variety x location interaction is detected an attempt is made to find which aspects of the environmental variation is responsible for the interaction. The approach used here allows GxE interaction to be detected and described when only a subset of the genotypes are tested in a large number of locations, each in an unreplicated trial. It does require that the locations be characterised by measurement of appropriate covariates. One reason for doing participatory breeding trials is that critical GxE interaction is due to varying social or economic environment. For example, it is often hypothesized that men and women will favour different varieties, or farmers assessment of genotypes will depend on level of market integration. These types of interaction can be detected and described as long as the design covers sufficient variation and suitable indicators of the social or economic variables are recorded.

6. Summary

The key points from this paper are:

- Analysis of data from participatory trials can and should use a combination of exploratory/descriptive methods and formal statistical modelling.
- The analysis may be complicated by the irregular layout of the experiment and multiple layers of variation introduced by the hierarchical design.
- Approaching the analysis by calculation of treatment contrasts on each farm can simplify many complex problems and lead to new insights into the data. However it can be inefficient or too repetitive if there are many treatments.
- Approaching the analysis by fitting regression models or their equivalent with multiple error terms allows many designs to be analysed within a common framework. However the analysis can be opaque and estimates non-intuitive.
- The two approaches can often be made to be equivalent
- The most useful analysis is often one that concentrates on finding explanation for variation in treatment effects across farms.
- Variation (at any level in the design) can be interpreted as risk, not just as unexplained noise.

7. Note on software

Since the original version of this paper was published 5 years ago there have been important changes in statistical software suitable for this type of analysis.

Genstat, used here to illustrate methods, has been up dated to the 9th Edition. The basic commands needed to do the analyses illustrated have not changed. All are available to users through simple menus and dialogue boxes. Some details have changed. For example, the PREDICT command will now give standard errors of differences of predicted means. More importantly, VSNi, the company that produces Genstat, have made the Discovery Edition available free to researchers and educators in the developing world. Details are available from http://www.vsn-intl.com/products/discovery/ or http://www.worldagroforestrycentre.org/rmg/GDE/index.html.

A second source of high quality statistical software free to all is R. Development of this open source software has been by a consortium with many contributors. Details are available and the software can be downloaded from http://cran.r-project.org/

It can take new users a while to learn the basics of R, but the effort is repaid by giving access to a very wide range of statistical tools, often including the very latest developments in statistics methods. As a starter, the following commands will give the analyses of Example 2 from this paper.

```
#Read the data, in this case from the clipboard after copying in Excel
baby<-read.table("clipboard", header=TRUE, na.strings="*")
attach(baby)

#Change SITE, FARM and ENTRY to factors as this is not the default
FARM<-as.factor(FARM.)
ENTRY<-as.factor(ENTRY.)
SITE<-as.factor(SITE.)

#Fit the fixed effect model and look at results
fixedmodel<-lm(simyield~FARM+ENTRY)
summary(fixedmodel)
anova(fixedmodel)

#Calculate and plot the adjusted ENTRY means
library(effects)
ENTRY.means<-effect("ENTRY", fixedmodel)
plot(ENTRY.means)

#Fit the random effects model
library(lme4)
randommodel<-lmer(simyield~ENTRY-1+(1|SITE)+(1|SITE:FARM))
summary(randommodel)
```

Acknowledgments

This paper was originally published as Coe R (2002) Analyzing data from participatory on-farm trails. 18-34 in Bellon M. R. and J. Reeves (eds). Quantitative analysis of data from participatory methods in plant breeding. Mexico, DF: CIMMYT. We thank CIMMYT for permission to republish in this journal.

8. References

Coe R (2007) Analysing ranking and rating data from participatory on-farm trials. (This volume)

Coe R and Franzel S. (2000) Keeping research in participatory research. Paper presented at Third International PRGA Conference, Nairobi, 6-11 November.

Coe R, Stern R and Allen E (2001) Analysis of data from experiments. Training materials. SSC/ICRAF, 250pp.

Genstat (2000) Genstat for Windows (5th Edition). Oxford: VSN International.

Hildebrand PE and Russell JT (1998). Adaptability analysis. Iowa State U.P.

Kempton RA and Fox PN (1997). Statistical methods for plant variety evaluation. London; Chapman and Hall.