Analysis of a field trial set out in randomized blocks

Mick O'Neill, Statistical Advisory & Training Service Pty Ltd Australia
Curt Lee, Agro-Tech, Inc. USA

The experiment

This discussion, based on a randomised block field trial with 4 blocks and 12 treatments, is written for those with only basic stats. The results of the analysis surprised the crop scientists, as treatments expected to be superior were not.

We firstly explain the approach to analysis of variance (ANOVA), and show how to use the residual diagnostic tools to full advantage. We then introduce the concept of a linear mixed model with what we loosely refer to as a *REML* analysis - an analysis which gives identical answers to an ANOVA when the ANOVA assumptions hold, but is much more versatile when they do not. *In this experiment it appears that blocking in the field has not been 100% successful.*

Table 1. Plot yields in field position.

	Plot											
Block	1	2	3	4	5	6	7	8	9	10	11	12
1	26.7	32.8	30.4	36.8	35.5	36.3	38.0	39.9	39.5	37.3	35.5	34.2
2	28.0	27.6	30.1	30.7	35.8	33.9	34.6	35.3	37.5	36.4	33.3	35.6
3	34.6	27.9	32.2	34.5	30.0	35.1	34.6	41.0	35.9	30.9	37.5	39.5
4	30.6	23.4	33.1	29.4	33.9	32.9	34.1	32.0	33.8	29.7	28.9	27.2

Table 2. Randomisation of treatments into 4 blocks. Sample means are used to flag the position of the two treatments with highest means (in green, darkest is the higher) and the treatments with lowest means (red for lowest, orange for next three treatments which were roughly similar).

	Plot											
Block	1	2	3	4	5	6	7	8	9	10	11	12
1	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
2	**12**	**5**	**7**	**10**	**9**	**4**	**2**	**3**	**6**	**8**	**1**	**11**
3	**11**	**2**	**4**	**9**	**1**	**12**	**10**	**8**	**5**	**3**	**6**	**7**
4	**9**	**10**	**7**	**3**	**6**	**4**	**2**	**1**	**12**	**11**	**8**	**5**

Notice that the two highest yielding treatments are mainly in the right half of the field. That should not matter if the plots are alike prior to the randomization of treatments within each block. But if that is not the case, then the yields from those two treatments are (possibly) advantaged - or disadvantaged - by where they were grown. We can see if analysis of variance detects this as a problem.

What is ANOVA?

From the experiment we calculate that the overall plot mean is 33.43 bu/ac with the best treatment being treatment 8 (with a mean yield of 36.6 bu/ac) and the worst treatment the control (with a mean yield of 30.5 bu/ac). There is clearly variation among the treatments, but the question is, is this simple random variance, or indicative of real treatment differences?

The usual way to measure the plot to plot variation is to calculate the overall *standard deviation* (s.d.) of the yield data, which turns out to be 3.85 bu/ac. Note that this measure is on the same scale as the yield data.

An alternative measure the plot to plot variation is the *variance* of the yield data. This is a squared measure, and the definition is

$variance = (standard\ deviation)^2$

Hence the overall variance of the yield data is 14.81 squared units. It is like an average of the *squared distances* of every yield away from the overall mean yield of 30.5 bu/ac.

In conducting a randomized block field trial, there are two stages to complete.

Stage 1 is the construction of **blocks** in the field. You do this recognising that the site of the experiment varies in some way; maybe there is a fertility or water gradient in a particular direction. You therefore construct blocks in such a way that allows for different growing conditions *across* blocks, but similar conditions *within* each block.

Stage 2 is the construction of **plots** in each block. A randomized *complete* block (RCB) design simply has a complete set of treatments randomized into the plots of each block - so that in general you would have the same number of plots in each block as you have treatments. Often this is not feasible, so certain incomplete designs have been developed for that eventuality. *You can skip the next concept and return to it later.*

In GenStat's notation, these two stages correspond to *strata* in the field (blocks and plots within blocks) and hence to strata in the analysis. When we set up the blocking structure in GenStat, we basically mimic what is done in the field trial:

Block/Plot

is the way we indicate that we have set up blocks and then plots in each block. In fact, Block/Plot is a GenStat shortcut. We could equally write Block/Plot as

Block+Block.Plot

however, more of this later.

If the overall variance is exactly 0, *then there is no variation*, and hence every yield is the same. That's obviously not going to happen in practice.

- We anticipate that the mean yields of the plots in each block will differ across blocks, since we identified different growing areas in the field that were assumed different from each other. Thus, we anticipate that the block variance will be real and larger than chance variance, and need to remove it in the analysis of our yield data.

  The block means are 35.24, 33.23, 34.48 and 30.75 and *each of these is based on 12 individual plot yields*. The variance of these four means is 3.865.

- In Table 1 we saw that the overall means of the twelve treatments differ, and *each of these is based on 4 individual plot yields*, one from each of four blocks. The variance of these twelve means is 4.129.

So let's examine the ANOVA table for this RCD experiment. We expect to see a component that measures

- the *overall variance of the yield data*,
- the *variance of the block means*,
- the *variance of the treatment means*;
- and there should also be a *residual* - or *experimental* - *variance*.

This is the ANOVA table:

Analysis of variance

Variate: Yield

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Block stratum	3	139.14	46.38	4.08	
Block.Plot stratum					
Treatment	11	181.66	16.52	1.45	0.197
Residual	33	375.30	11.37		
Total	**47**	**696.11**			

The component labelled **Total** is where the overall variance of the yield data (14.81) should be found. Where is it? In fact, if we divide 696.11 by 47 we obtain 14.81.

Had we used a regression menu, we actually do have this value printed at the bottom of the analysis of variance table:

Change	d.f.	s.s.	m.s.	v.r.	F pr.
+ Block	3	139.14	46.38	4.08	0.014
+ Treatment	11	181.66	16.51	1.45	0.197
Residual	33	375.30	11.37		
Total	**47**	**696.11**	**14.81**		

Definition of the ANOVA terms

s.s.	Sum of Squares
d.f.	Degrees of freedom
m.s.	Mean Square, defined as **s.s./d.f.**
v.r.	Variance Ratio - often labelled F in some packages, defined for example as **Treatment m.s./Residual m.s.** (Check that 16.51/11.37 = 1.45)
F pr	Probability of obtaining the observed v.r. *or a larger one*; often labelled **P value** in packages

The sum of squares is an old fashioned hand-calculation column still included in the ANOVA table. Notice that the sums of squares of blocks, treatments and residual add to the total sum of squares (139.14+181.66+375.30 = 696.11). Notice also that the two components that were included because of the way the experiment was conducted - blocks and treatments - explain less than *half* of the Total s.s.: 139.14+181.66 = 320.80 as a percentage of 696.11 is only 46%. This leaves 54% as residual - is that an indication of something wrong with the experiment? We aim to have as small a residual component as possible (by accurately allowing for blocks,

by careful sowing, management, harvest and measuring) so that the treatments can be compared as accurately as possible. That brings us back to the concept of variances.

- As we have said, the sample variance of al 48 plot yields is 14.81. The degrees of freedom of a sample variance are simply *one less than the number of data values* whose variance is being calculated, in this case 47.

- If the experiment has been conducted carefully and accurately, the Residual m.s. is the best estimate to use for the variance of the random yield data. On that assumption, the best estimate of the individual plot *variance* is 11.37, and hence the best estimate of the individual plot *standard deviation* is $\sqrt{11.37}$ = 3.37 bu/ac.

- We constructed blocks in the field and found the variance of the block means (each of 12 plots) is 3.865. To compare this with the residual variance, we need to make the comparison "fair"; the residual variance is based on single plots, the block variance is based on means of 12 plots. Hence we need to scale up the block variance by 12, obtaining 12×3.865 = 46.38. This is the Block m.s. in the ANOVA table. GenStat then calculates the variance ratio, 46.38/11.37 = 4.08. There are 4 block means, so the numerator in this variance ratio has 4-1 = 3 degrees of freedom.

  GenStat does not calculate a P value for this variance ratio since (a) why test for block differences when you, the experimenter, thought there were block differences before the experiment even started, and (b) unless you regard the blocks used in this experiment as just a random choice then each block is just an unreplicated large growing area; only replicated factors can be compared statistically.

- In order to test whether some treatment means (each based on 4 plots, one from each block) are different, we compare the treatment variance to the residual variance. Again, to make the comparison "fair", we need to scale up the treatment variance, in this case by 4, obtaining 4×4.129 = 16.51. This is the Treatment m.s. in the ANOVA table. GenStat then calculates the variance ratio, 16.51/11.37 = 1.45. There are 12 treatment means, so the numerator in this variance ratio has 12-1 = 11 degrees of freedom.

Note that if there are *no* treatment mean differences, we would expect the Treatment m.s. to be *about the same* as the Residual m.s. (and hence the variance ratio to be about 1). If there ***are*** treatment differences, then mathematically we would expect the Treatment m.s. to be ***larger*** than the Residual m.s.. Thus, the P value (or F. pr.) for testing treatments is based on an F distribution (named after Fisher, the English statistician who developed much of the ANOVA approach) and is one sided - **we reject that the treatment means are all equal if the variance ratio is *significantly large***. While a P value (in this case) of 0.197 would indicate no treatment differences, there are times when individual treatments are compared they turn up significantly different. We will therefore examine the part of GenStat's output relating to means.

Summary.

Our expectations for the components of an ANOVA based on what was done in the field (with *b* blocks and *t* treatments) were:

Expected to see:	**ANOVA equivalent**
the *overall variance of the yield data*,	*Overall variance* is Total m.s.
the *variance of the block means*,	$t \times$ *Block variance* is Block m.s
the *variance of the treatment means*;	$b\times$ *Treatment variance* is Treatment m.s
and there should also be a *residual* - or *experimental - variance*.	Residual m.s used as the denominator of the F test of the treatment means

We now look at the remaining part of the ANOVA output from GenStat.

Tables of means

When there are no treatments missing in any block, a treatment mean is simply the sample mean from the plots that had that treatment applied.

Variate: Yield

Grand mean 33.43

Treatment	Treatment 01	Treatment 02	Treatment 03
	30.50	32.35	31.50
Treatment	Treatment 04	Treatment 05	Treatment 06
	33.85	33.95	36.55
Treatment	Treatment 07	Treatment 08	Treatment 09
	35.10	32.78	31.50
Treatment	Treatment 10	Treatment 11	Treatment 12
	31.55	36.30	35.18

Standard errors of means

Table	Treatment
rep.	4
d.f.	33
e.s.e.	1.686

e.s.e. stands for *effective standard error*. For this design, the standard error is estimated as $\sqrt{Residual\ m.s./4}$ $=\sqrt{11.37/4}$ = 1.686. We use 4 as a divisor since this is the number of plots in each mean.

Standard errors of differences of means

Table	Treatment
rep.	4
d.f.	33
s.e.d.	2.385

s.e.d. stands for the *standard error of the difference* between two means. For this design, the s.e.d. is estimated as $\sqrt{Residual\ m.s.\times\left(\frac{1}{4}+\frac{1}{4}\right)}=\sqrt{11.37\times\frac{2}{4}}=2.385$. In this example the number of plots in every treatment mean is 4.

Least significant differences of means (5% level)

Table	Treatment
rep.	4
d.f.	33
l.s.d.	4.852

l.s.d. stands for the *least significant difference* and is used either to compare any two means, or to calculate a confidence interval (C.I.) for the difference between two means.

Using the l.s.d.

Any two treatments can be compared using a t test, defined as $t=\frac{mean\ 1-mean\ 2}{s.e.d.}$. This value is then compared to say a 5% critical t value (t_{crit}) and the mean are said to be significantly different if the observed t value is larger than t_{crit} (ignoring the sign).

With several means comparisons to make, it is simpler to calculate $t_{crit} \times s.e.d.$. This is called the *l.s.d.* value. In this formula, the degrees of freedom to use when looking up t_{crit} are the Residual d.f. (in this case 33); t_{crit} here is 2.035, and *l.s.d.* = 2.035×2.385 = 4.852.

Then:

- Any two means that differ in magnitude by more than the *l.s.d.* value are declared significant at 5% (at least).

- If you add and subtract the value to the difference in means you obtain a (95%) confidence interval for the true mean difference. This is an interval that (95%) of the time should contain the true mean difference.

With an *l.s.d.* of 4.852 you can see that the *control* is different to the treatment labelled Treatment 10; there are four other differences significant at 5%.

The 95% C.I. for this difference is 5.80 ± 4.852. Rounding to one decimal, while the best estimate of this treatment difference is 5.8 bu/ac, it could be as low as 0.9 bu/ac or as high as 10.7 bu/ac.

Sometimes you want to obtain a (95%) C.I. for an individual mean. This is obtained by adding and subtracting $t_{crit} \times e.s.e.$ = 2.035×1.686 = 3.431 to the mean. For example, while the best estimate of the control mean is 30.5, it could be as low as 30.5-3.431 = 27.1 bu/ac, or as high as 33.9 bu/ac

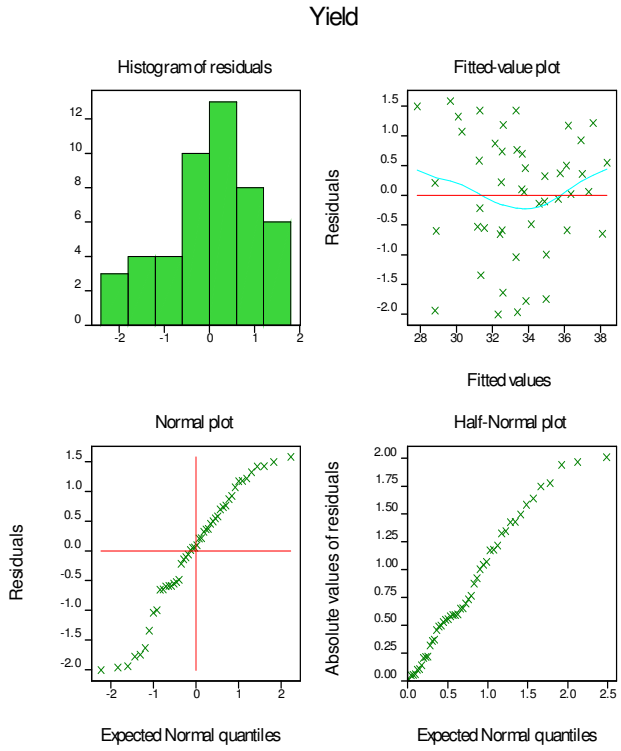
Assumptions underlying the ANOVA

The ANOVA is based on a linear model, and that gives rise to property that the various residual sums of squares add to the Total s.s. for designs like this (randomised block).

We assume that the yield data are normally distributed for the F test to apply. Data which are not normal can sometimes be analysed this way, but with care.

We assume that the variances are constant across every treatment and every block. This applies rarely to count data, and in the past transformations were used to overcome the problem. There are better modern analyses that can be used for count data. Experiments that include different spacings or different harvest times should also be checked carefully.

We assume that the residuals are just random noise - there should be no patterns in the residuals in time or in field position. GenStat allows you to plot the (standardised) residuals against fitted value (a valuable plot for detecting an increasing variance situation) or in field position (via a contour plot):



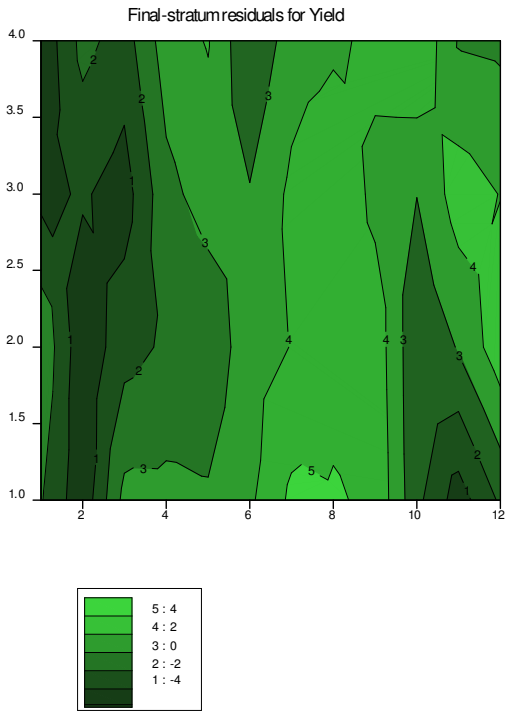
Top L: a *histogram* of the residuals. Useful when you have a large number of residuals; should be bell-shaped.

Top R: Residual versus fitted value plot. Should have no trend, should be a random swarm of points around the red line positioned at 0.0.

Bottom: Two versions of the Q-Q plot, with the residuals plotted against a "perfect" set of normally distributed residuals. The tighter the points are to the line the better.

There are no particular indications of a problem with the analysis based on these plots.

	Plot											
Block	1	2	3	4	5	6	7	8	9	10	11	12
1	-5.6	-1.4	-2.9	1.0	2.1	-1.8	1.0	1.5	2.6	4.0	-0.2	-0.4
2	-4.6	-3.8	-4.9	-0.6	0.9	0.1	2.4	4.0	1.4	0.0	3.0	1.9
3	-0.3	-5.5	-2.8	-1.7	-1.6	1.3	2.1	3.4	3.3	-1.7	0.2	3.3
4	-1.8	-5.4	0.6	0.6	0.3	1.6	4.4	4.2	3.7	-1.5	-5.0	-1.7



We also requested the residuals to be printed out in field position (above) and plotted as a contour plot.

There should be no pattern in field position, no preponderance of positive or negative residuals in any area of the field.

The contour plot suggests a trend exists among the residuals *across the field* and casts doubt about the assumptions made when forming blocks.

There is also an option to print out the stratum variances:

Stratum	variance	effective d.f.	**variance component**
Block	46.381	3.000	**2.917**
Block.Plot	11.373	33.000	**11.373**

The variation among block means is considerably smaller than the plot to plot variance within a block. Block 4 appears to have low yields on average:

Block	1	2	3	4
	35.24	33.23	34.47	30.75

Using the Residual m.s. as a basis for F tests is done on the assumption that everything we could control for is removed by the way we ran the experiment. The Residual m.s. should be an estimate only of random variation, and should not include variation in other factors not accounted for by us.

We can take the trend across the field into account using a **REML** analysis. REML stands for Residual Maximum Likelihood, is used in models which contain both fixed and random factors and is versatile in that there is no restriction about the independence of the residuals or the variance structure in the data.

The Fixed Model in the REML menu for this experiment is simply the treatment design of the ANOVA (so simply Treatment), and the Random Model is the **block structure** of the ANOVA, *plus* the **plot variation** that the contour plot suggests is still present (so Block+Plot).

Linear Mixed Model (REML) accounting for plot and block variation

Estimated variance components

Random term	**component**	s.e.
Block	**3.455**	3.158
Plot	**6.140**	3.400

Residual variance model

Term	Factor	Model(order)	Parameter	**Estimate**	s.e.
Residual		Identity	Sigma2	**4.923**	1.482

Tests for fixed effects

Sequentially adding terms to fixed model

Fixed term	Wald statistic	n.d.f.	F statistic	d.d.f.	**F pr**
Treatment	28.86	11	2.62	24.9	**0.022**

- This analysis has now detected a significant treatment effect (P=0.022). It appears to be much more sensitive an analysis for the orientation of plots and blocks in this experiment.

- The Residual variance has reduced from 11.37 in the ANOVA to **4.923** in the REML analysis.

- The variation left to right in the field (**6.140**) is almost *twice* that of the block variance (top to bottom in the field, **3.455**).

Because the treatments are no longer balanced with respect to plots and blocks, the means are adjusted by their field positions (see REML mean column below), and there is now a different standard error for each mean and for each mean differences.

Treatment	ANOVA mean	ANOVA e.s.e.	REML mean	REML e.s.e.
Treatment 01	30.5	1.69	29.9	1.70
Treatment 02	32.4	1.69	33.9	1.88
Treatment 03	31.5	1.69	32.2	1.71
Treatment 04	34.0	1.69	34.2	1.69
Treatment 05	31.6	1.69	34.5	1.76
Treatment 06	**36.3**	**1.69**	**34.5**	**1.77**
Treatment 07	35.2	1.69	35.0	1.69
Treatment 08	**36.6**	**1.69**	**32.6**	**1.70**
Treatment 09	**35.1**	**1.69**	**31.1**	**1.70**
Treatment 10	31.5	1.69	31.7	1.69
Treatment 11	33.9	1.69	35.4	1.70
Treatment 12	32.8	1.69	36.0	1.76

Notice that some REML means are adjusted downwards relative to the ANOVA means (e.g. treatment 6, 8 and 11) because it is judged that they received favourable conditions, whereas other treatment means are adjusted up.

The s.e.d. value from the REML analysis all vary from a minimum of 1.664 to a maximum of 2.094 with an average s.e.d. of 1.813 - much less than the constant value 2.385 from the ANOVA.

A more advanced Linear Mixed Model (REML) accounting for plot and block variation

A model with a random block term is equivalent to a model where the plots in each block are uniformly correlated - that is, the yields are correlated in exactly the same way irrespective of distance apart. Similarly, a model with a random plot term is equivalent to a model where the plots across the blocks in each position are uniformly correlated. Neither assumption is likely to be practical.

Agronomists have started to use spatial models in one or both directions in the field. To do that, they argue that plots close together are likely to be more highly correlated than plots further apart. A model that is useful in field trials is an **autoregressive correlation structure** (**AR**); an AR1 model simply says that the yield of a plot is directly affected by the yield of its neighbouring plot, but not *directly* by the yield of plots further apart. The mathematics show that this structure is equivalent to a correlation r say for two plots side by side, r^d for two plots at a distance d apart. An AR2 model says the direct influence is from the neighbouring plot *and* the next neighbour. In this case an AR1 in both block and plot directions is indicated.

- With this model, the treatments are now **strongly significant (P<0.001)**:

Tests for fixed effects

Sequentially adding terms to fixed model

Fixed term	Wald statistic	n.d.f.	F statistic	d.d.f.	**F pr**
Treatment	76.56	11	6.93	19.9	**<0.001**

- The correlation between two plots side by side in any block is strong, 0.77. The correlation between two neighbouring plots across blocks is 0.45:

Residual variance model

Term	Factor	Model(order)	Parameter	Estimate	s.e.
	Block.Plot		Sigma2	16.10	7.21
	Block	AR(1)	phi_1	**0.4540**	0.1784
	Plot	AR(1)	phi_1	**0.7672**	0.1034

The means are more sensitively adjusted, changing both standard errors of means and of mean differences:

	ANOVA mean	REML mean	AR1×AR1 mean	ANOVA ranking	REML ranking	AR1×AR1 ranking
Treatment 01	30.5	29.9	29.2	12	12	12
Treatment 02	32.4	33.9	31.9	8	7	9
Treatment 03	31.5	32.2	31.0	10	9	10
Treatment 04	34.0	34.2	33.2	5	6	6
Treatment 05	31.6	34.5	33.2	9	4	5
Treatment 06	36.3	34.5	34.0	2	5	4
Treatment 07	35.2	35.0	34.4	3	3	2
Treatment 08	36.6	32.6	33.2	1	8	7
Treatment 09	35.1	31.1	29.6	4	11	11
Treatment 10	31.5	31.7	32.1	10	10	8
Treatment 11	33.9	35.4	34.1	6	2	3
Treatment 12	32.8	36.0	34.9	7	1	1

Under this model, there is less of a range in both means and their s.e.d. values; the minimum s.e.d. is 1.094, the maximum 1.641 and the average 1.366. Recall that the ANOVA s.e.d was a constant 2.385 and the gain in precision becomes clear. GenStat can print out (or save to an Excel file) the whole matrix of s.e.d. values.

Conclusion:

- A classical ANOVA of the data produced a surprisingly non-significant overall treatment effect. This was explained when a close examination of the residuals in field position detected a strong trend across the field that had not been noticed when setting up the experiment.

- A REML analysis with an AR1 × AR1 structure (which means that plots are directly correlated with their closest neighbour both vertically and horizontally in the field) produces a strongly significant treatment effect with treatment rankings that better fit biological expectations based on past treatment performance.