Interpretation of statistical data

The results of replicated trials are presented as either the predicted or average (mean) for each of the replicates within a treatment. When analysing data statistically, authors generally use a REML spatial model (Hart method) or ANOVA, respectively.

Here, we describe the internal process used to analyse Hart data in GenStat.

A spatial analysis is conducted using restricted maximum likelihood (REML), incorporating row and column effects to account for spatial variation across the field. Treatment is fitted as a fixed effect. When $p \leq 0.05$, a multiple comparisons test is performed using the Bonferroni test at the 5% significance level. The Bonferroni critical difference (CD) is calculated using the average standard error of difference (SED) from the REML output and the appropriate Bonferroni-adjusted t critical value (calculated in excel using GenStat output). Differences between treatment means greater than the calculated Bonferroni CD are considered statistically significant. Not significant (NS) indicates that there is no difference between the treatments ($p > 0.05$).

At a 95% confidence interval if $p \leq 0.05$, we are 95% confident that observed differences in a trial are due to the treatments, and not by chance (5%).

Interpretation of replicated results: an example only

Below we use an example of a replicated wheat variety trial containing both grain yield and quality data (Table 1). Statistically significant differences were found between varieties for both grain yield and protein. The Bonferroni CD for grain yield is 0.40, meaning there must be more than 0.40 t/ha difference between yields before that variety's performance is significantly different to another. In this example Calibre is significantly different to all other varieties as it is the only variety followed by a superscript (a). Scepter, Vixen and Ballista are not significantly different from each other and are all followed by a superscript (b) as they all yielded within 0.4 t/ha of each other.

Similarly, for grain protein, variety performance was only significant from another if there was more than 0.9% difference in protein. In the example, Catapult contained a higher protein level compared to all other varieties which were not different to one another.

Where there are no significant differences between treatments, NS will be displayed as seen in the screenings column below (Table 1).

Table 1. Wheat variety grain yield, protein and screenings from a hypothetical example to illustrate interpretation of p-value and Bonferroni CD. Columns with shaded values show the best performing treatments.

Variety	Grain yield (t/ha)	Protein (%)	Screenings (%)
Catapult[b]	3.50[c]	10.3[a]	0.2
Ballista[b]	3.98[b]	8.4[b]	1.0
Vixen[b]	3.75[bc]	9.1[b]	0.5
Scepter[b]	4.05[b]	8.9[b]	0.9
Calibre[b]	4.77[a]	8.4[b]	0.4
P-value	**0.002**	**<0.001**	**NS**
Bonferroni CD	**0.40**	**0.9**	

Disclaimer

While all due care has been taken in compiling the information within this manual the Hart Field-Site Group Inc and researchers involved take no liability resulting from the interpretation or use of these results.

We do not endorse or recommend the products of any manufacturers referred to. Other products may perform as well or better than those specifically referred to.

Any research with un-registered products and rates in the manual does not constitute a recommendation for that particular use by the researchers or the Hart Field-Site Group Inc.