

# BCG RESEARCH METHODOLOGY

## TRIAL SITES

BCG managed 22 research and demonstration sites across the Wimmera and Mallee in 2019. The sites are repositioned each year in order to generate results representative of a range of soil types, paddock history and climatic conditions. BCG research trial sites included replicated experiments investigating a range of issues including weed control, herbicide resistance and efficacy, varietal assessments, crop nutrition, disease management, barley agronomy, crop sequencing, frost management and season-specific issues.

## REPLICATED TRIALS AND EXPERIMENTS

Replicated trials and experiments are those in which the treatments are repeated more than once (usually four times). This allows for the use of statistical tests which can determine whether differences observed in average (mean) results are likely to be due to the treatments or whether they occurred purely by chance.

## INTERPRETING BCG TRIAL RESULTS

Interpreting and understanding trial results is not always easy. In this publication, BCG attempts to present the results in a standard format. Results from replicated trials may be used to assist in making confident on-farm decisions.

### **Statistical tests**

The results of replicated trials are presented as the average (mean) for each of the replicated treatments.

BCG usually conducts statistical tests at a 95 per cent (%) confidence level. This means that if a 'significant' result is reported, the author is 95% confident that there is a difference between treatments. Statistical tests such as t-tests and analysis of variance (ANOVA) give a 'P' value which signifies the probability of two results being different due to chance. A P-value less than 0.05 ( $P < 0.05$ ) indicates that there is at least a 95% chance that the results are different and that the difference is considered to be significant (S). A P-value greater than 0.05 ( $P > 0.05$ ) indicates that there is a less than 95% probability that the results are different and that the difference is considered to be not significant (NS).

In the case of ANOVA, in which the means of more than one treatment are compared, those which are significantly different from each other are shown using the Least Significant Difference (LSD), which is calculated at the 95% confidence level. Average results that lie within the mean of another result ( $\pm$  the LSD) are considered to be not significant.

## Demonstrations

Each year, BCG conducts a number of single treatment demonstrations. Without the need for replication, demonstrations allow the exhibition of a large number of treatments. These, however, cannot be statistically analysed. Observed differences cannot be validated and may be the result of paddock variability or chance. Demonstrations are usually conducted in conjunction with a replicated trial in order to ensure some scientific rigour in the findings.

## Coefficient of variation

The coefficient of variation (CV) is a measure of the variability of data used to calculate a treatment average. It is expressed as a percentage. A low CV (<10%) suggests that a treatment performed similarly across all replicates within a trial. A high CV (>20%) suggests that a treatment performed inconsistently across the replicates within a trial. This can be due to differences in environmental conditions such as soil type or rainfall, variations in weed density and disease pressure or in measurement error.

## Replicated results: an example

In this publication, statistical results will be displayed either as a table or as a chart. Below is an example of the same set of results for a replicated trial containing wheat yield and quality data, with statistical interpretation in both table and chart form. *Please note these figures are hypothetical and are not based on any replicated trials performed by BCG.*

**Table 1. Wheat variety yield, protein and screenings results from a hypothetical example using a table to illustrate LSD.**

Variety	Yield (t/ha)	Protein (%)	Screenings (%)
Lincoln	4.6 <sup>d</sup>	9.5 <sup>b</sup>	3.7
Axe	4.0 <sup>c</sup>	10.7 <sup>cd</sup>	3.6
Correll	3.9 <sup>bc</sup>	10.6 <sup>c</sup>	3.5
Gregory	3.5 <sup>ab</sup>	8.5 <sup>a</sup>	4.0
Mace	3.3 <sup>a</sup>	11.4 <sup>d</sup>	4.1
Yitpi	3.2 <sup>a</sup>	11.3 <sup>c</sup>	3.4
<b>P-value</b>	<b>0.001</b>	<b>&lt;0.001</b>	<b>0.06</b>
<b>LSD (P=0.05)</b>	<b>0.4</b>	<b>0.9</b>	<b>NS</b>
<b>CV%</b>	<b>5.2</b>	<b>2.0</b>	

In the example in Table 1, mean yields, protein and screenings of the different varieties were analysed. Significant differences were found between varieties in yield and protein, but not in screenings. The LSD of 0.4t/ha for yield shows that there must be more than 0.4t/ha difference between yields before that variety's performance is significantly different from another. In this case, the yields of Yitpi and Mace were not significantly different from each other. Both varieties are labelled with an identical superscript letter (*a*). This illustrates that there was no significant difference between the yields of the two varieties.

However, Yitpi (*a*) yielded significantly lower than Correll (<sup>bc</sup>), Axe (<sup>c</sup>) and Lincoln (<sup>d</sup>), as there was more than 0.4t/ha difference between varieties. The superscript letters which are different from those on other varieties denote significant differences. If the letters between varieties are the same, they are not significantly different from one another. Where there are no significant differences between treatments, NS, meaning not significant, will be displayed.

In Table 2, the mean yields of all varieties and both time of sowing (TOS) were analysed, and the interaction between variety and TOS. This means three sets of statistics are produced, two main effects of either variety, or TOS, and then also the interaction effect (variety x TOS).

**Table 2. Wheat variety yield from two TOS results from a hypothetical example using a table to illustrate a two way interaction.**

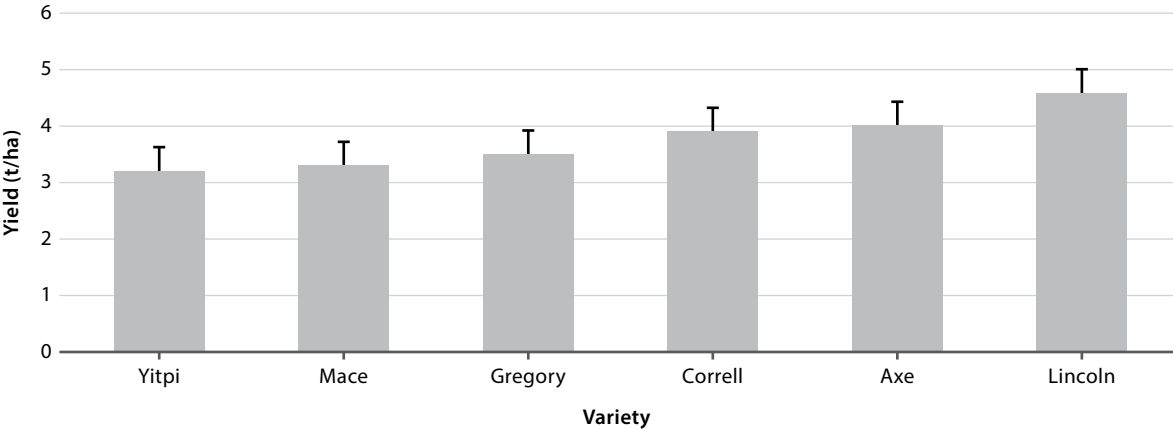
Variety	TOS 1 yield (t/ha)	TOS 2 yield (t/ha)
Trojan	7.4 <sup>a</sup>	6.8 <sup>bcd</sup>
Cosmick	7.2 <sup>ab</sup>	6.5 <sup>defg</sup>
Scepter	7.1 <sup>abc</sup>	6.7 <sup>de</sup>
Derrimut	6.8 <sup>bcd</sup>	6.1 <sup>fgh</sup>
Corack	6.6 <sup>de</sup>	6.6 <sup>def</sup>
Mace	6.4 <sup>defg</sup>	6.1 <sup>ghi</sup>
Wallup	6.2 <sup>fg</sup>	6.7 <sup>cd</sup>
Kord CL Plus	6.1 <sup>ghi</sup>	5.6 <sup>ij</sup>
Scout	6.1 <sup>ghi</sup>	5.7 <sup>hj</sup>
DB Aurora	5.5 <sup>j</sup>	6.2 <sup>efg</sup>
<b>Sig. diff. Variety</b>	<b>&lt;0.001</b>	
TOS	<b>0.536</b>	
Variety x TOS	<b>&lt;0.001</b>	
<b>LSD (P=0.05) Variety</b>	<b>0.3</b>	
TOS	<b>NS</b>	
Variety x TOS	<b>0.4</b>	
<b>CV%</b>	<b>4.3</b>	

Variety is the main effect and it is significant ( $P < 0.001$ ) meaning the average yield of variety (from both TOS) show differences. The LSD is the number that needs to be met to represent a significant difference, which in this case is 0.3t/ha. So if the average yield of a variety is more than 0.3t/ha different to another, they are significantly different. The second main effect, TOS is not significant ( $P = 0.536$ ) meaning the average overall yield of TOS1 is equal to TOS2. The LSD that would have been required for TOS to be significant was 3.4t/ha.

The interaction of variety and TOS was significant ( $P < 0.001$ ) and the LSD was 0.4t/ha. This means if a variety in TOS1 is more than 0.4t/ha different to a variety in TOS2 they are statistically different. For example, Trojan in TOS1 yielded 7.4t/ha, Trojan in TOS2 yielded 6.8t/ha meaning they are significantly different. Trojan TOS1 is also significantly different to Cosmick in TOS2 (6.5t/ha). On the other hand, Corack in TOS1, is not significantly different to Mace in TOS2 as they do not meet the LSD of 0.4 t/ha between means.

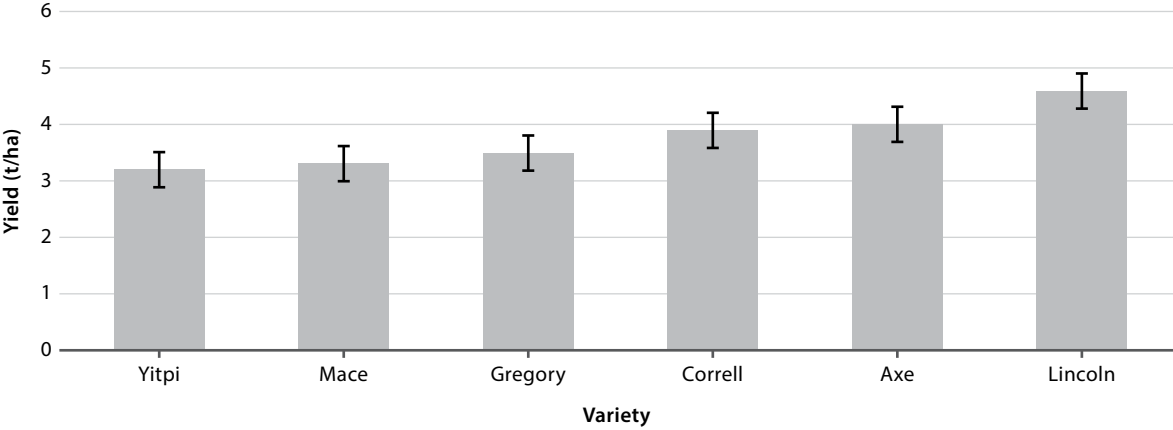
Throughout this publication, statistical results may also be presented as graphs. Error bars at the top of each solid column within graphs can represent the LSD or Standard Error.

As the error bars in Figure 1 extend only upwards, they are expressing the LSD between the yields of wheat varieties. Whether a variety (or treatment) is significantly different from another can be determined by looking at the LSD error bars. When the LSD bar does not reach the top of a solid column, a significant difference is evident. In Figure 1, there is no significant difference between the grain yields of Yitpi, Mace and Gregory. The yield of Gregory is, however, significantly different from those of Axe and Lincoln, but not of Correll.



**Figure 1. Wheat variety yields from a hypothetical example using error bars to illustrate LSD.**

Error bars that express the standard deviation extend both up and down from the top of each solid column. A standard deviation is a statistical measurement used to show how much variability exists in a set of data around the average, or expected value. A long standard deviations bar indicates a board range of possible values relative to the expected value. A short standard deviation bar means the data points are considered close to the expected value.



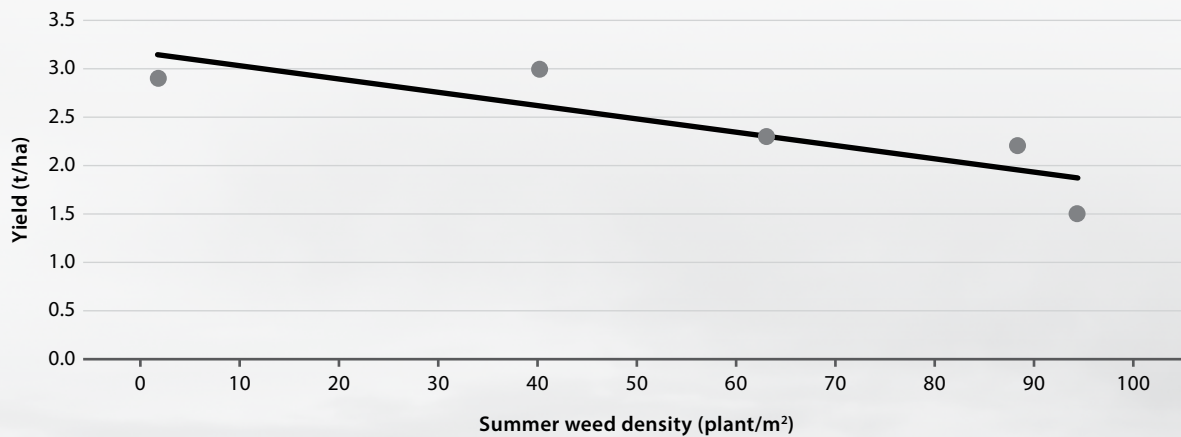
**Figure 2. Wheat variety yields from a hypothetical example using error bars to illustrate standard deviation.**

### Regression analysis

Regression analysis is a statistical technique used to assess the relationships between two variables: a dependent variable (such as grain yield) and an independent variable (such as weed density). Regression analysis helps to explain what will happen to a dependent variable as the independent variable changes. For example, what happens to yield as weed density increases?

Regression analysis allows a set of data points to be simplified into a line of best fit; that is, a line which passes through the set of data points and indicates the relationship between the two variables. When looking at the relationship between two variables, we need to understand how close the relationship is. To achieve this, a statistical measure known as  $R^2$  is used.  $R^2$  is the variation between the data points and the line of best fit. It indicates how well a group of points on a graph adheres to a straight line. Essentially, the closer the points are together, and the more closely they resemble a straight line, the stronger the relationship between the data points.

Figure 3 is an example of regression analysis. This graph shows how summer weed density (plants/ $m^2$ ) influences yield (t/ha). It is clear from this example that, as weed density increases, yield decreases. Figure 3 shows an  $R^2$  value of 0.72 or 72%. This means that 72% of the variation in yield can be explained by changes in summer weed density. The higher the  $R^2$  value, the stronger the relationship; in this instance, we can see the relationship is quite strong. It is dependent on the variables under consideration.



**Figure 3. Using regression analysis to determine the effect of summer weed density on yield.  $R^2=0.72$ .**

### **Gross margins or economic analysis**

Where appropriate, gross margins have been performed to compare treatments. No statistical analysis has been performed on the gross margins unless stated. In some instances, statistical analyses have been performed on the economics. While gross margins can be a good guide to show the expected return of different treatments, many variables are involved. If there are no significant differences realised at one of these levels, gross margins may become misleading. Gross margin outputs should be treated with caution. Accompanying comments which outline the parameters employed in the development of gross margin outputs must be taken into consideration.

Gross margins do not take into account fixed or capital cost, which is one of the primary reasons they can be ambiguous. They should be treated with caution.

### **Normalised Difference Vegetation Index (NDVI)/GreenSeeker®**

In trials conducted by BCG, NDVI (normalised difference vegetation index) is used as a substitute measure of crop biomass and greenness. The NDVI sensor is directed over the trial plots and emits brief bursts of red and infrared light. Optical sensors measure the amount of each type of light reflected back from the target area, which is a measure of the amount and intensity of greenness in the target. The sensor records the measured value in terms of an NDVI reading ranging from 0.00 to 0.99, where high values indicate intensely green covering 100 per cent of the ground, and lower values may indicate less ground cover, or a less green crop.

