

Before you analyze data...



Analysis relies on data accuracy

- Review rating **immediately**, at trial site
- Outliers can exist, **typos** should not!

This video demonstrates the tools and functions ARM uses for data confirmation. You should complete this important step before analyzing data. The goal is to catch outliers and any inaccuracies with the assessment.

To ensure accuracy, review the data at the trial site immediately after the assessment. The researcher can then investigate and resolve any irregular values. Recording data on paper and entering it later may result in lost statistical power. Any issues found “after the fact” cannot be verified, and the only option is to exclude the irregular data points.

The issues we are looking for may sometimes just be a typo from data entry. Or it could be an outlier, which is an observation outside the normal or expected distribution. But just because a value is ‘unusual’, does not mean it is ‘not usable’! Some discernment is needed to determine if the data should be kept or thrown out.

Assessment Data - Line 35			
+	Column Number	1	
-	Reporting Basis		1
	Number of Subsamples	1	1
	Crop	TRZAW	TR
	Pest		
	- Pest Type	Disease	
	- Pest Code	LL11FY	
	- Pest Scientific Name	Complex of known fungi	
	- Pest Name	Complex, dif. fungi of cert. t	
	- Pest Stage Scale	DESC	
	- Pest Density	1.75 %	
	Timing	4	5
	Miscellaneous		
	- Data Reliability		
	- ARM Action Codes		
	Number of Decimals	2	2

Here is an example trial with some data to review. The assessment in column one measures disease coverage on the plant leaves, giving a rough estimate of the area where our disease covers the crop.

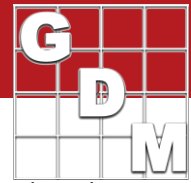
First, decide if this assessment is realistic or viable to analyze. Otherwise, all other data confirmation and review steps are irrelevant to this assessment. For example, if the pest infestation level is too low, you cannot make conclusions about the product's ability to control the pest.

Assessment Data - Line 35			
+	Column Number	1	
-	Reporting Basis		1
	- Pest Name	Complex, dif. fungi of cert. t	
	- Pest Stage Scale	DESC	
	- Pest Density	1.75 %	
	Timing	4	5
	Miscellaneous		
	- Data Reliability	1 Exclude	
	- ARM Action Codes		
	Number of Decimals	2	2
+	Sub	Rep	Blk
	Col	Plot ~	Tit
			1
			★ ✓
	1	1	1
	101	3	10.00
	1	1	2
	102	1	15.00
	1	1	3
	103	4	20.00

Use the Data Reliability shortcut buttons in ARM to indicate if conclusions should not be made from this assessment. Then add the NM Action Code to prevent statistics from running for this column.

Alternatively, when data confirmation is complete then select the green checkmark to note that this column has been reviewed for inaccuracies and is ready for analysis.

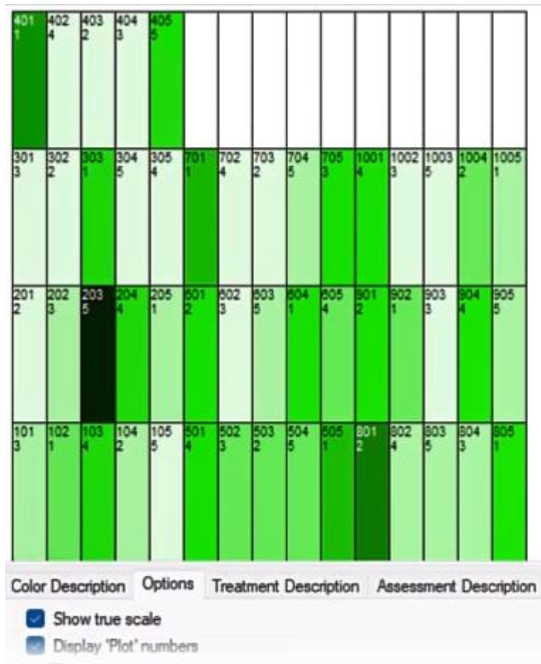
# Data Confirmation



The primary tool for confirming data is the Assessment Map. Find it on the Properties Panel, in the Tools section.

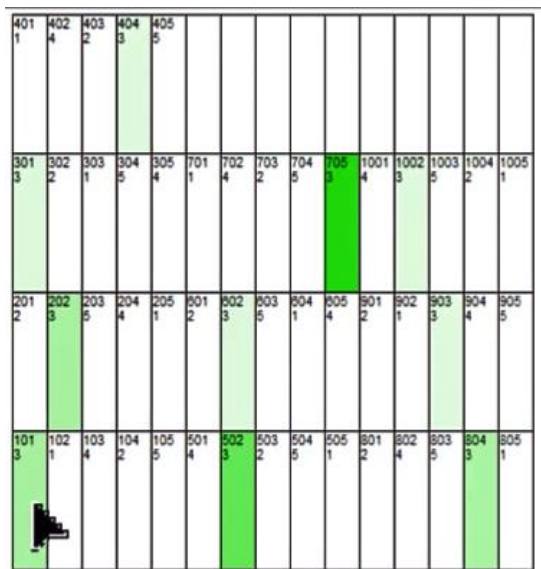


This is a heat map to visualize the differences in data values. The lightest color represents the smallest observation. The darkest color corresponds to the highest value in the assessment. A gradient of colors represents all values in between. Use this visual to quickly identify extreme values to investigate. For example, plot 904 has a large value.



Given that the assessment is percent area, a value of 120 must be a typo. If still in the field, the researcher can return to the plot and confirm the actual value. If this review occurs after leaving the field, the researcher will need to exclude the value, because there is no way to know if the observation is a 10 or 20.

After adjusting, return to the assessment map to continue the data review. Select "Show true scale" in the Options tab. ARM draws the plots according to the dimensions in the Trial Settings. This gives us an accurate spatial visualization of the assessment.

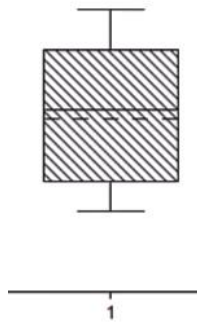


Click on a plot to color the map by the current treatment. All other treatments turn white, providing a visual representation of the consistency of the treatment across replicates. Use the first replicate to examine each treatment individually. To turn off this feature, uncheck the "Color by current treatment" option in the bottom right corner.

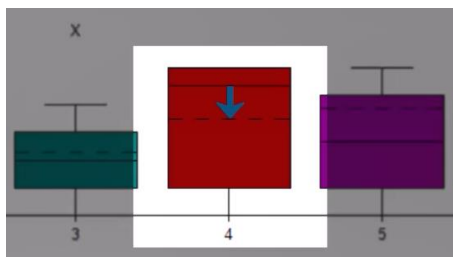
Finally, use the Next Column and Previous Column buttons to navigate between assessments.



Another tool available is the Box-Whisker graph. Click the button on the Properties Panel to create a graph for the selected assessment column. Select "Show X axis labels as legend" to make the graph larger.

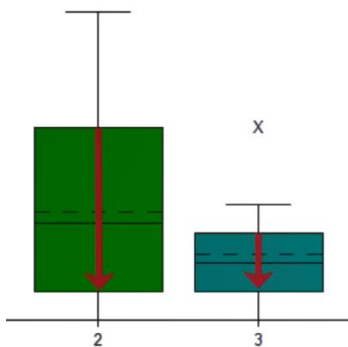


Each box represents a treatment. The box extends from the 25th to the 75th percentile. A solid line divides the box, representing the median. The whiskers extend from the ends of the box to the largest and smallest observation, not including outliers. An outlier is a point that lies outside the box by more than 1.5 times the height of the box. The mean is drawn as a dashed line, and can be toggled in Options.

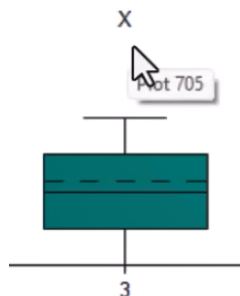


The following are things to look for on the box-whisker graph:

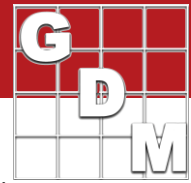
Comparing the mean and median gives you a sense of the skewness of the data. Look at Treatment 4 in this example. The mean is quite a bit lower than the median. This indicates a low value that is "pulling down" the mean. Treatment 3 shows a more balanced set of observations.



The height of the box is an estimate for the amount of variability in each treatment. Treatment 2 has the most variability in this example, while Treatment 3 has the least.



Check outliers by hovering over the "X." For example, plot 705 appears as an outlier on our box-whisker graph. It didn't stand out on the assessment map, but it is a good idea to take a second look and then document the rating with a comment or a photo.



### Outliers

- Un-**usual** data does not mean un-**usable** data!
- Don't throw out all calculated outliers automatically
  - Does this data fairly represent the trial and its objectives?
  - Did an external force affect this plot disproportionately?
  - Is this value within a reasonable range of outcomes?

A note on outliers: remember that *unusual* data does not mean it is *unusable*. When reviewing extreme values, ask the following before deciding whether to exclude them:

- Does this data fairly represent the trial and its objectives?
- Did an external force affect this plot disproportionately?
- Is the recorded value within a reasonable range of outcomes?

Remember, even a normally distributed data set with a bell-shaped curve will still have a few values outside the two standard deviations.

2	2
1	2
30.00	
5.00	
5.00	
10.00	
<del>25.00</del>	I
40.00	
10.00	

If you do decide to remove the value, use the Exclude checkbox on the Properties panel to the right. This preserves the documented value, but in the analysis ARM ignores the value and treats the plot as "missing" instead. If there is a missing value, do not leave it blank. Enter a period to make it clear that there is no value to enter.

In summary, take time to confirm the data entered in ARM. Tools like the Assessment Map and Box-Whisker visualize the assessment to identify inaccurate or unusual observations.