



When things go wrong...

Handling unexpected situations:

1. Data Outliers
2. “Failed” Experiments

Random Normal Statistics

- In a Science Fair project, most often you use Random Normal Statistics, like average
- Commonly used statistics have certain underlying assumptions
 - Data samples are randomly selected from a population (random, Normal)
 - The data represent the population
 - The population members (data) follows the Normal distribution



1. Data Outliers

Data Outliers

- If your data does not fit Random Normal Statistics then you may have Data Outliers:
 - Some of your data points may, when graphed, look suspiciously like they don't really belong in the data set
 - These data points may be Data Outliers
 - Let's look at how we can test for them.....

Data Outliers

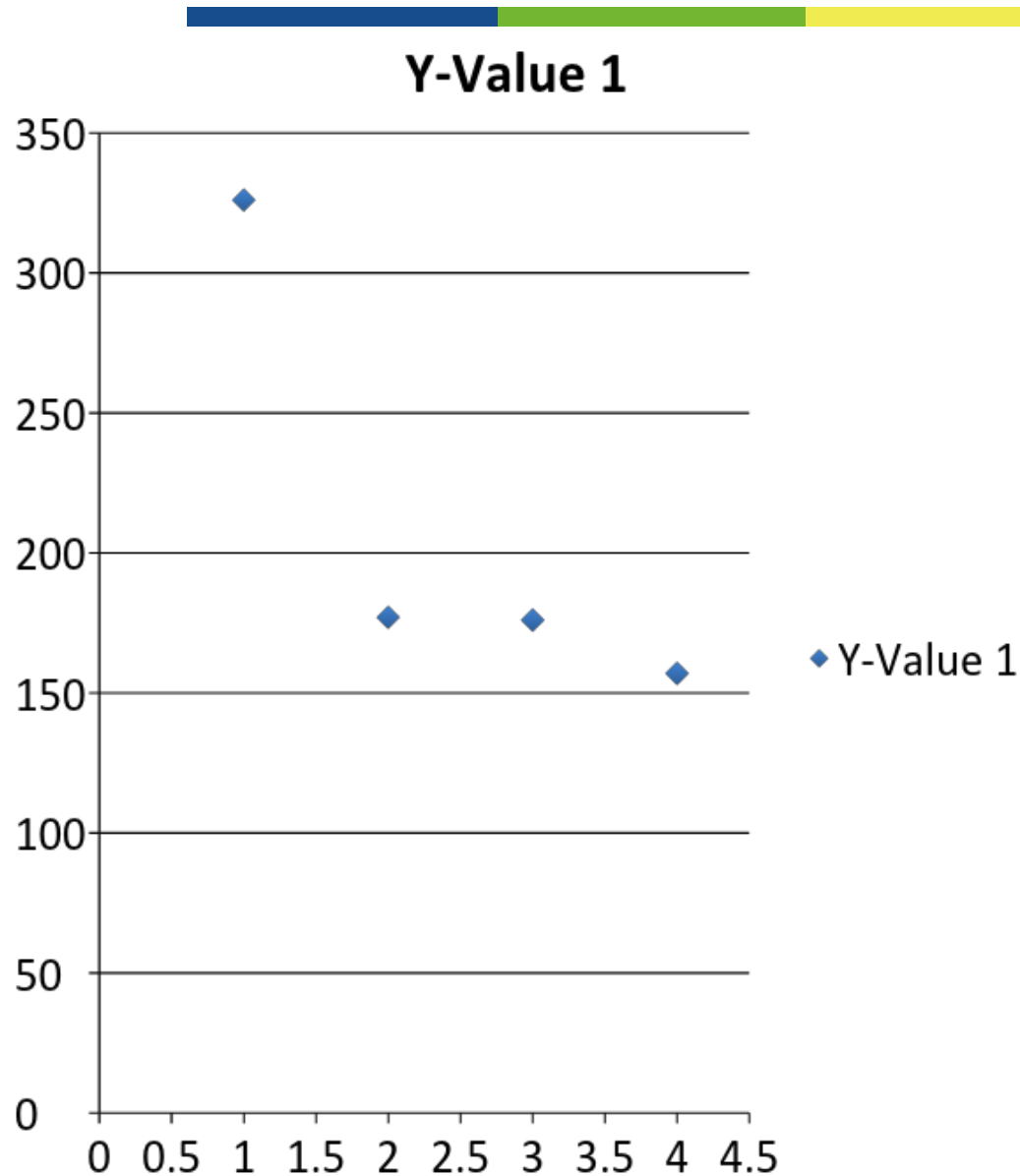
- How do data outliers occur?
 - When we set up a project experiment, we may assume that all of our sample data comes from the same population of data
 - This may not be true!
 - A weight or count is not recorded correctly
 - The calibration of a scale used to weight objects might change with temperature or time, or we don't recalibrate or check the scale over several days
 - Our assumptions may be incorrect

Data Outliers

- How do I know if I have data outliers?
 - Data outliers are data points that don't belong with the rest of the data we collect
 - To obtain valid results, we need to check to see if all of our data samples make sense
 - Often a plot of our data will show some data points that seem to not be part of the data set
 - We make 4 observations of weight (in grams) are: 326, 177, 176, 157
 - The 326 weight looks suspicious and might be an outlier (we read the scale incorrectly or wrote the wrong number down)

Data Outliers

From the previous slide, the data points plot shows that the value 326 g. doesn't seem to fit the other three points.



Data Outliers

- Detecting data outliers
 - Graph the data (scatter diagram) and look at the scatter of the data points
 - Do they all look logical?
 - Do any look suspicious, that is, far away from all the other data?
 - If you perform a linear regression, is the r-squared value far away from ± 1 (for example, 0.5) or fails the test of significance at the 0.95 level?
 - These data points may be the outlier data

Outlier Data

- How do I test for Outliers?
 - Order your data by size from largest to smallest
 - 326, 177, 176, 157
 - Subtract the largest value from the next largest
 - $177 - 326 = -149$
 - Divide by the range of values
 - $-149 / (157 - 326) = 0.882$
 - From the table in slide 10 we expect a value of 0.765 or less; 0.882 is not, so the 326 is an outlier

Test for Outliers

Number of Observations (data points)	Critical Value at alpha = 0.05
3	0.941
4	0.765
5	0.642
6	0.560
7	0.507
8	0.554
9	0.512
10	0.477
Source: Dixon and Massey, Introduction to Statistical Analysis, McGraw-Hill, 1957 Pages 276 and 412 (Table 8e).	Note: for larger sample sizes or other values of alpha, consult the reference.

Data Outliers

- What to do if you find outliers:
 - Remove data that are outliers and recalculate statistics used in your project with the remaining data points.
 - It may be useful to identify the reason (assignable cause) that an outlier is not part of the sample
 - E.g., our definition of the population includes cases where our assumption is incorrect
 - This might cause us to redefine how we analyze the data that we collected



2. The Experiment Appears to Have “Failed”

The conclusions do not support my
Hypotheses.

“Failed” Experiment

- Sometimes the data does not support our hypothesis
 - This is what science experiments are all about!
 - We have proven our (null) hypothesis is incorrect and we may have to redefine our experiment or project
 - However, this does not prove the alternative hypothesis is correct, and that is what we have to then test

Gone wrong?

- As we encounter other examples, this module will be updated
- Any suggestions for additional content?
Contact us at: sciencefair@asq1010.org