SIMPLE STATISTICS FOR SCIENCE FAIR WEATHER PROJECTS

William P. Roeder * and Dewey E. Harms
45th Weather Squadron
Patrick AFB, FL

1. INTRODUCTION

Science fair weather projects can often be improved through better application of statistics. After more than 20 years of judging science fairs by the authors, the same statistical shortfalls continue to be observed. While some projects have excellent statistics, most do not. The most frequent shortfall is the total absence of any statistics. This is surprising, since the competitors and teachers know that statistics will be a part of the judging criteria. The second major shortfall observed in science fair projects is the weak application of statistics when they are used. The explanation for these shortfalls is likely the general lack of statistical instruction in America's secondary education. This same problem is also seen in meteorology programs in higher education (Brown, et al., 1999).

2. GENERALIZED STATISTICAL PROCESS

A simple generalized statistical process can be applied to many science fair projects. This process consists of five sequential steps, as listed in Table-1. All the steps in the process are simple enough to be done on standard computer spreadsheets, i.e. expensive statistical analysis software is not needed.

TABLE 1. Generalized Statistical Process

Step	Action
1	Collect Adequate Sample Size
2	Graph Raw Data and Inspect Visually
3	Apply Data Quality Control
4	Calculate Average And Standard Deviation of Average. Graph and Inspect Visually.
5	Perform Statistical Tests

Before using this process, the student should be taught basic statistics. Otherwise, the student is just mindlessly applying a rote checklist and hasn't learned any science. The student must absolutely understand the basics of statistics: natural variation, statistically insignificant differences, the advantage of using averages over actual observations, and the advantages of large sample size in getting a representative average and standard deviation and in reducing the standard deviation of the average. Some useful resources for teaching statistics are listed in Table-2.

* *Corresponding author address*: William P. Roeder, 45 WS/SYR, 1201 Edward H. White II, Patrick AFB, FL 32925-3238; e-mail: william.roeder@patrick.af.mil
website: www.patrick.af.mil/45og/45ws/ws1.htm

This five-step generalized statistical process can also be used as an aid to experiment design, which must be done before beginning the experiment. The student should ensure that all required resources, data sample sizes, and analysis techniques will be available. This is especially needed if a specific level of accuracy or level of confidence in the final decision is required – an appropriate sample size is absolutely vital to provide the statistical tests sufficient power in those cases.

2.1 *Step-1, Adequate Sample Size*

The first step in the general process is to collect an adequate sample size. Normally a sample of at least 25-30 independent events is recommended for representative statistics. But in science fairs, samples of just 5 or less are typical. It is important to recognize that 25-30 or more independent events are required per statistical stratification. If the data are to be analyzed into various categories, or a regression equation with various predictor variables is to be created, then at least 25-30 independent events for each category or each predictor variable is needed.

2.2 *Step-2, Graph Raw Data and Visual Inspection*

The second step is graphing the raw data for visual inspection. While graphing is done in many science fair projects, it is far from universal. When graphing is done, often the best type of graph is not used. Scatter diagrams are useful for examining relationships between variables and detecting outliers. Histograms are good for comparing the numbers of members of various categories. Pie charts are best used to compare the relative proportions between various categories in a group. Time-series show data collected in sequence over time. Other types of graphs exist, along with variations on the ones listed here. But these graphs cover the needs for most science fair projects.

Visual inspection of the graph can help identify outliers to be eliminated from subsequent analysis. Visual inspection can also identify patterns in the data, which can guide selection of future statistical tests. Plotting the standard deviation of the raw data can be very helpful, but this is virtually never done in science fair projects. The standard deviation of the raw data is a function available in virtually all computer spreadsheets, and its formula is not provided here. Students interested in the formula may find them in the references in Table-2, or in the 'Help' function in their spreadsheets.

TABLE 2. Useful References

Title	Comments
Statistical Methods In The Atmospheric Sciences (Wilks, 1995)	Best of the new meteorology statistics books
Essential Statistics, 4th Edition (Rees, 1999)	No nonsense "how to" guide
Some Applications Of Statistics To Meteorology (Panofsky and Brier)	Old, but still useful. Exceptional clarity of instruction.
Electronic Statistics Textbook (StatSoft, Inc., 2001)	Broad but brief. Good survey of all statistical techniques. (www.statsoft.com/textbook/stathome.html)

2.3 *Step-3, Apply Data Quality Control*

The visual inspection of the graph of the raw data is part of data quality control, especially in removing outliers.   While caution must be used in removing outliers, lest valid data is thrown away, it is just as important to remove true outliers, to avoid misleading results.   Clear evidence is needed before removing data from the analysis. Just because a data point looks extreme is insufficient reason to remove it as an outlier. Independent evidence that the datum is suspect is needed.   Highly detailed experiment logbooks are important for this. Any unusual occurrences during a trial, as recorded in the logbook, can be used to help justify removing candidate outliers.

Removing even a single outlier can make a large difference in the data analysis. Figure 1 shows an example where removing just one outlier causes a huge change. When the single outlier is not removed, the linear regression has a slope of 0.96 and a correlation constant of 0.94 (90% of the variance explained by the linear regression). At first glance, the high correlation constant implies the linear regression is an excellent fit to the data. But removing the outlier results in a slope of 0.34 and a correlation constant of 0.30 (9% of the variance explained). A vitally important difference would have been missed without the data quality control.

Figure 2 shows how visual inspection can suggest future statistical analyses.   The scatter diagram indicates two clusters of data. This suggests the two clusters should be analyzed separately. For example, an average and standard deviation might be calculated for each cluster, rather than for the entire group as a whole. Or a regression analysis might be done on each separate cluster.   However, as with outliers, just because some data look like a separate groups is insufficient reason to treat them separately. Independent evidence is needed first. There should be a physical explanation for the separate clusters. Also, hypothesis testing can be used with the average and standard deviation of the average from each cluster to test if they are statistically significantly different. More advanced statistical techniques like cluster analysis and discriminate analysis could be used for this test, and for classifying future data. But this exceeds the scope of simple statistics for science fairs. Students need to understand the simpler techniques, before advancing to more advanced procedures.   Science fair students interested in the more advanced techniques are referred to Wilks (1995).

As can be seen, graphing and visual inspection of the data as a step towards data quality control can be vitally important to obtaining realistic results.
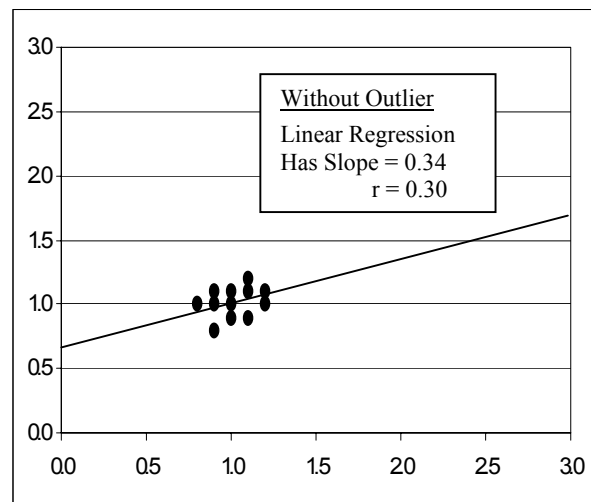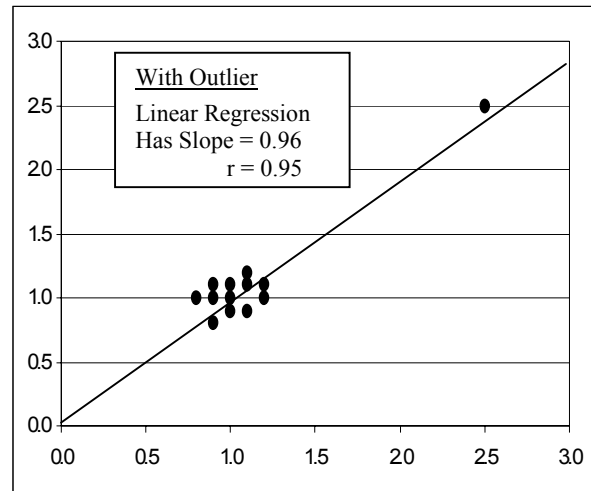

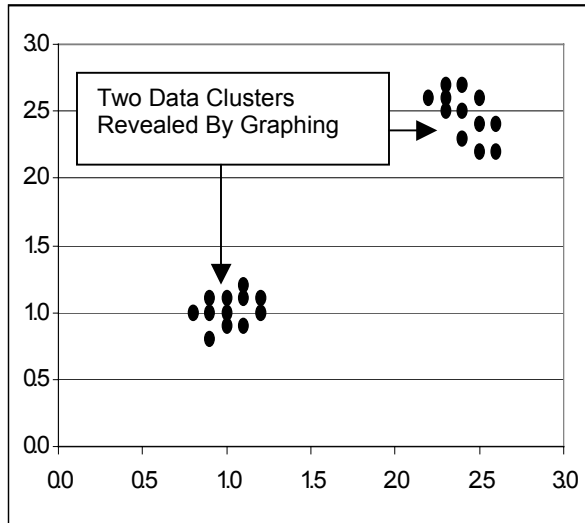**Figure 1.** Removing even just one outlier can cause a large change.

Figure 2. Visual inspection of graphed data can reveal distinct clusters that may need to be analyzed separately.

2.4 *Step-4, Calculate Average and Standard Deviation Of Average, Graph And Inspect*

The fourth step is to calculate and graph the average(s) and standard deviation of the average(s) for the overall data and its various stratifications. It is important to note that this is the average and standard deviation of the <u>average</u>, as opposed to the standard deviation of the raw data, which was discussed in step-2.

The average and standard deviation of the raw data are functions available in virtually all computer spreadsheets, and their formulas are not provided here. Students interested in the formulas may find them in the references in Table-2, or in the 'Help' function in their spreadsheets. However, the standard deviation of the average is usually not provided as a spreadsheet function. But it is easily calculated by the following equation:

$$s_{avg\ x} = s_x / \sqrt{n_x}$$

where $s_{avg\ x}$ is the standard deviation of the average of the generic variable x, s_x is the standard deviation of the variable x, and n_x is the sample size of variable x.

It is important to note that the standard deviation of the average decreases as the inverse square root of the sample size. Thus, as the sample size increases, the uncertainty in the average becomes smaller, and the power to detect statistically significant differences increases. This is one of the main advantages in larger sample sizes.

Graphing the averages and their standard deviations presents another opportunity to inspect the data points for outliers. Depending on which statistical tests are selected, the averages and their standard deviations will be needed in step-5.

2.5 *Step-5, Perform Statistical Tests*

The fifth step is the final statistical testing, such as hypothesis testing and confidence intervals, or regression and correlation analysis, or performance evaluation, depending on the experiment goals.

Hypothesis tests and confidence intervals are useful in determining if statistically significant changes were introduced by different experimental conditions, or if different categories are statistically significantly different. Regression analysis finds the best-fit line to the data and is useful in creating forecast models, while the correlation coefficient measures how well the regression line fits the data. These tests are usually taught in introductory statistics courses, which presumably research science teachers have taken. The tests are also well described in the resources in Table-2. As discussed previously, many of calculations required by these statistical tests are provided as functions or are easily calculated on computer spreadsheets. Functions for linear regression best-fit slope and intercept are also available on spreadsheets, and their formulas are not provided here. Students interested in the formulas may find them in the references in Table-2, or in the 'Help' function in their spreadsheets.

Weather science fair projects sometimes evaluate the performance of weather forecasts. Unfortunately, most introductory statistics courses do not teach forecast verification. A 2 x 2 contingency table is usually used to evaluate the performance of binary yes/no forecasts (Figure 3). Even this simplest of all possible forecasts requires three independent metrics to fully describe the forecast's performance. But in science fair evaluation of weather forecasts, even these most basic of verifications usually aren't done.

The three most commonly used metrics are Probability Of Detection (POD), False Alarm Rate (FAR), and Critical Success Index (CSI). Formulas for calculating POD, FAR, and CSI are in Figure 3. The POD measures how well the forecast technique predicts the event, when it actually occurs – 100% is perfect and 0% is the worst possible score. The FAR measures how poorly the forecast technique predicts the event, when it actually doesn't occur – 0% is perfect and 100% is the worst. A perfect POD can easily be obtained by a mindless no-skill forecast technique; always forecast the event to occur. But then the FAR will be degraded. A perfect FAR can also be easily obtained by a no-skill forecast technique; never forecast the event to occur. But then the POD will suffer. Thus, a third metric that measures skill and represents an optimal balance between POD and FAR will be degraded. In the forecast verification sense, skill means performance as compared to some baseline forecast technique. The CSI is the most frequently used metric for skill and measures performance relative to random forecasting – 100% is perfect and 0% is the worst. The point of zero skill versus random forecasting lies between 0% and 100%, but its exact value is unknown, since it varies with the frequency of occurrence of the event. Thus CSI can not be easily interpreted and care must be taken in

comparing skill scores from different time periods or different locations. If the frequency of occurrence of the event has changed, CSI can rise or fall in value, without necessarily meaning the skill has increased or decreased, respectively.

POD and FAR are good performance metrics and are easily understood. But other measures of skill, such as Heidke Skill Score or Kuiper Skill Score are superior to CSI. However, CSI remains the most frequently used skill metric and is the only skill metric presented here. The other skill scores are discussed in Wilks (1995).

Figure 3. 2x2 contingency table and formulas for forecast verification metrics.

3. SUMMARY

Many science fair weather projects can be improved with better statistics. A simple 5-step statistical process was presented as a general guideline for most science fair projects. A wealth of other more advanced statistics tests is obviously available. But students need to be able to perform these basic procedures before moving on to more advanced topics. Given the generally low level of statistics observed in most science fair projects, the application of even just this basic five-step procedure will significantly improve the quality of science fair projects.

The root problem with low level of statistics in science fairs is the general absence of statistics in America's secondary education. A long-term fix to this problem would be integrating statistics into secondary school science courses.

REFERENCES

Brown, T. J., L. M. Berliner, D. S. Wilks, M. B. Richman, and C. K. Wikle, 1999: Statistics Education in the Atmospheric Sciences, Bulletin of the American Meteorological Society, Vol. 80, No. 10, Oct 99, 2087-2097

Panofsky, H. A., G. W. Brier, 1958: Some Applications Of Statistics To Meteorology, *Pennsylvania State University Press*, pp. 224

Rees, D. G., 1999: Essential Statistics, 4th Edition*, Chapman and Hall/CRC Press*, pp. 224

StatSoft, Inc., 2001: Electronic Statistics Textbook. Tulsa, OK. WEB: www.statsoft.com/textbook/stathome.html

Wilks, D. S., 1995: Statistical Methods In The Atmospheric Sciences, *Academic Press*, pp. 467