Plant Physiol. (1988) 87, 797–798
0032-0889/88/87/0797/02/$01.00/0

Communication

Frequency Distribution Histograms for the Rapid Analysis of Data[1]

Received for publication January 12, 1988 and in revised form March 5, 1988

PATRICIA V. BURKE, BERTHA L. BULLEN, AND KENNETH L. POFF*
Michigan State University-Department of Energy Plant Research Laboratory, Michigan State University, East Lansing, Michigan, 48824

ABSTRACT

The mean and standard error are good representations for the response of a population to an experimental parameter and are frequently used for this purpose. Frequency distribution histograms show, in addition, responses of individuals in the population. Both the statistics and a visual display of the distribution of the responses can be obtained easily using a microcomputer and available programs. The type of distribution shown by the histogram may suggest different mechanisms to be tested.

Perhaps the most common method used for data analysis in plant biology is the comparison of the means of two populations. For example, the response of plants to some experimental treatment can be compared with the response of plants to a control condition with no treatment. Typically, a number of individuals are subjected to the experimental parameter, and a number of other individuals are subjected to the control condition. The response of each individual plant is measured, the responses of each population averaged, and the two means compared. Variability is typically indicated by the standard error of the mean. Thus, a graph may be based on a series of data points, each representing the mean for a particular population. Each data point is typically associated with a vertical bar representing ±1 SE of the mean giving the range into which the true mean probably falls.

While analyzing data from *Arabidopsis thaliana* mutants with altered phototropism or altered gravitropism, we noted that frequency distributions often contain information which may not be obvious from the values for mean and standard error. Hence, we recommend that frequency distribution histograms be plotted to visually check data for information that may not be evident from the usual statistics. We feel that the potential benefit of this technique is widespread and have chosen to demonstrate its power with model data rather than using actual data from a specific experiment. However, the theoretical distributions model distributions that we have measured for mutant strains with altered phototropism or gravitropism.

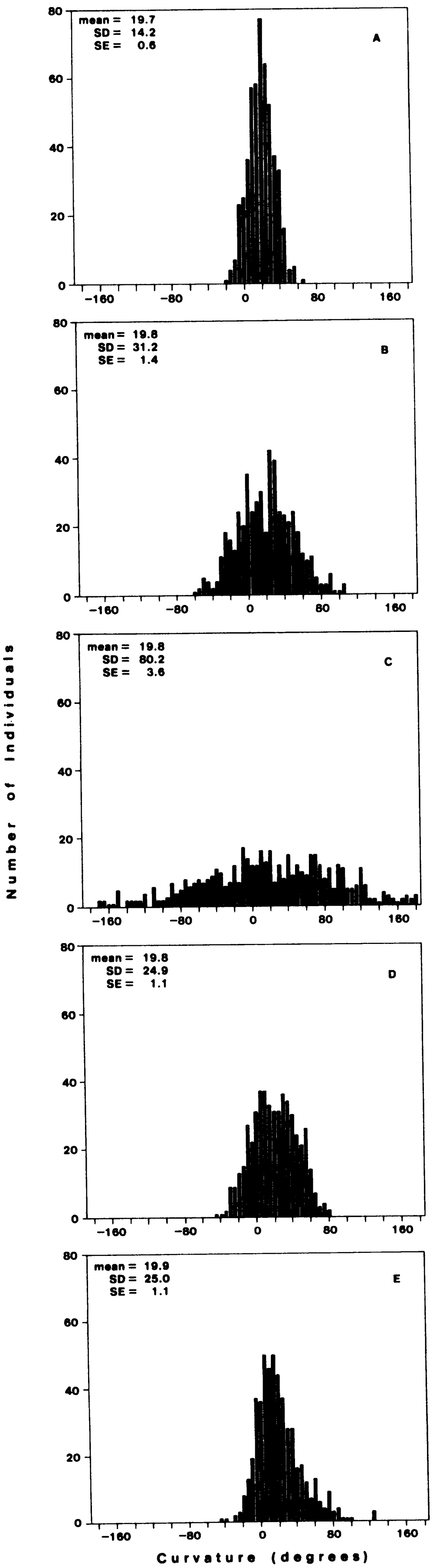
The data for the frequency distributions presented here were generated with an IBM PC computer using a program written in 'C.' The algorithm generates 'data' by randomly sampling Gaussian distributions with the prescribed means and standard deviations. The data have been assigned to the appropriate bin, and the frequency distributions have been constructed and printed using VP-planner (Stephenson Software Inc.). The abscissa for these frequency distributions has been set equal to degrees curvature to model the curvature of plant shoots exposed to a phototropic or gravitropic stimulus.

For each distribution (Fig. 1, A–E), a population of 500 individuals has been used. The mean, standard deviation, and standard error are given with each distribution. It would not be readily evident, based on the means and standard errors, that the five populations are as dissimilar as the distributions indicate. Although the standard error increases slowly from 1A to 1C, the standard deviation increases rapidly and individuals are found at all possible values in 1C. For a normally distributed population, the standard deviation is an estimate of the width of the distribution. The standard error is an estimation of the standard deviation of a number of randomly determined means of the population, not an estimation of the variability of individuals in the population. Moreover, the standard error is strongly dependent upon the number of individuals in the sample. This makes it difficult to compare the standard error of two samples with different numbers of individuals if the number is not given and a nuisance if the number is given. For example, the standard error for a distribution of type 1B with about 75 individuals would be the same as the standard error for type 1C with 500 individuals. Thus, the standard error by itself does not give any information about the underlying distribution, which may or may not be Gaussian. Note that the distributions in 1B, 1D, and 1E have the same mean and standard error. Although an error bar representing ±1 SE is appropriate as an indicator of the range within which the 'true' mean is expected to lie, this error bar should not be expected to be particularly informative as to the width or shape of the distribution of the population (3, 6), and as discussed below, the shape of the distribution may be important in supporting or discarding different underlying models.

Most plant biologists probably calculate means and standard errors for their data using an electronic calculator. Given the widespread availability of personal computers and relatively inexpensive statistical software (*e.g.* VP-planner, Lotus 1-2-3, Excel, Statworks), it takes little additional time to use a computer in place of a calculator and obtain a frequency distribution in addition to mean, standard deviation, and standard error.

In any biological system, we expect to deal with variability in our data, whether this arises from heterogeneities in the population, variability in the response of an individual, or variability in the stimulus itself. If there are random variations in one parameter, the distribution should be Gaussian. If there are random variations in the steps of a linear sequence (pathway), a logarithmic plot of the frequencies should be Gaussian (5). Thus, different distributions may support totally different models (1, 2, 4).

[1] Supported by the U.S. Department of Energy under contract DE-AC02-76ERO-1338, and by the U.S. National Aeronautics and Space Administration under grant NAGW-882.



Given the frequency distribution, it is easier to accept the variability as a potential source of information about the response of individuals in the population rather than solely as an irritation to be minimized.

Acknowledgments—We are indebted to T. Best, M. Gregg, K. Keith, and D. Lerner for their helpful comments.

LITERATURE CITED

1. Berg HC 1983 Random Walks in Biology. Princeton University Press, Princeton, NJ
2. Block SM, JE Segall, HC Berg 1983 Adaptation kinetics in bacterial chemotaxis. J Bacteriol 154: 312–323
3. Meyer SL 1975 Data Analysis for Scientists and Engineers. John Wiley & Sons, New York
4. Miller RH, RJ Lasek, MJ Katz 1987 Preferred microtubules for vesicle transport in lobster axons. Science 235: 220–222
5. Montroll EW, MF Shlesinger 1982 On 1/f noise and other distributions with long tails. Proc Natl Acad Sci USA 79: 3380–3383
6. Taylor JR 1982 An introduction to error analysis. University Science Books, Mill Valley, CA

Fig. 1. Frequency distributions for computer-generated sample populations chosen randomly from Gaussian distributions with prescribed means and standard deviations. Mean, standard deviation, and standard error are given on the figure for each population. The abscissa has been arbitrarily set as curvature in degrees from −180° to +180°. A to C, Each graph shows 500 individuals from a single Gaussian distribution. D to E, Each graph shows the sum of two populations taken from different Gaussian distributions; for D, one distribution of 250 'individuals' with mean of 0.3° and SD of 15.5°, and a second distribution of 250 individuals with mean of 39.3° and SD of 15.6°; for E, one distribution of 330 individuals with mean of 10.1° and SD of 15.1°, and a second distribution of 170 individuals with mean of 38.7° and SD of 29.2°.