Project Title: Wheat-field Variation and Nitrogen Fertilization Practices

Project Leader: Michael E. Tarter, *Professor of Biostatistics,*
University of California, Berkeley
(510) 642-4601
tarter@berkeley.edu

Cooperator: Steve Orloff, *Farm Advisor/ County Director,*
UCCE Siskiyou County, 1655 S. Main St. Yreka, CA
*(530) 842-2711, sborloff@ucdavis.edu***:**

Abstract/Summary of Results and Conclusions:
Based on simulated data as well as natural data supplied by the Cooperator an Invited Paper was presented this August at the Montreal, Canada, International Joint Statistical Meetings (JSM). The following findings were presented:

> A specialized experimental approach had been designed. Its goal was to take advantage of the cooperator's prior experience with the dependent variate, N rate in lbs/acre nitrogen. It was asserted that 80 different rates should be selected, from 0 to 400 lbs/acre clustered around 225 to 275 lbs/acre. The idea is to concentrate the spacing between adjacent selected N rates closest together at the value 250 lbs/acre. Both to the left and to the right of the value 250 lbs/acre the spacings between adjacent rates increase. When based on prior cooperator experience the new procedure for spacing optimization increased the precision of measured information near the value, here 250 lbs/acre (the value deemed optimal by the cooperator).

The JSM paper described how the 80 different rates were studied using model-free curve estimation software. This software showed that the suggested choice of rates assured that the assumptions which underlie both regression and correlation methodology were valid.

Introduction and Objectives:
Our investigations concerned the interrelationships between measurements such as flag leaf N and yield protein. Contrasts and comparisons between subgroups were taken to be of secondary importance. (Interrelationships are studied by means of multivariate analyses procedures while contrasts and comparisons are usually assessed with the help of analyses of variance and covariance methodology.) Besides the two variates, Y = needed N, and Z = yield % protein, our studies are designed to base decisions about optimal N on some third variate, X, such as midseason flag-leaf measured N. The figures shown below illustrate how growers and industry representatives can determine how much N is needed in order to achieve yield and protein goals. The question is, given that a grower goal is a z_0 % protein yield, for a measured value x, what quantity of N, here denoted by the letter y, best assures that this goal will be reached?

The significance of this project stems from connections between representational flexibility and farm individuality. Here the term *flexibility* denotes the capacity to change from a positive slope to a horizontal line beyond the point of diminishing returns. One definition of the term *flat lining*

is to be in a state of no progress or advancement which, for the purposes of this project, implies a change, beyond a single N rate, from a line with positive slope to a horizontal line.

Budget (describe how the Commission funding was spent): The cost of travel to and accommodations at the Montreal Joint Statistical Meetings (JSM) totaled $2,669. It is anticipated that by this year's grant expiration date, $1,331 dollars will have been expended on supplies and services, such as computer repairs, software and hardware such as the replacement of outdated computer hard drives as well as RAM memory, etc.

Results: Conventional regression methods often are based on the assumption that change can be represented with the help of a single straight line. For example, consider the data points and model-based regression line that are shown in Figure 1, as well as the same points and both model-based and nonparametric regression curves shown in Figure 2.

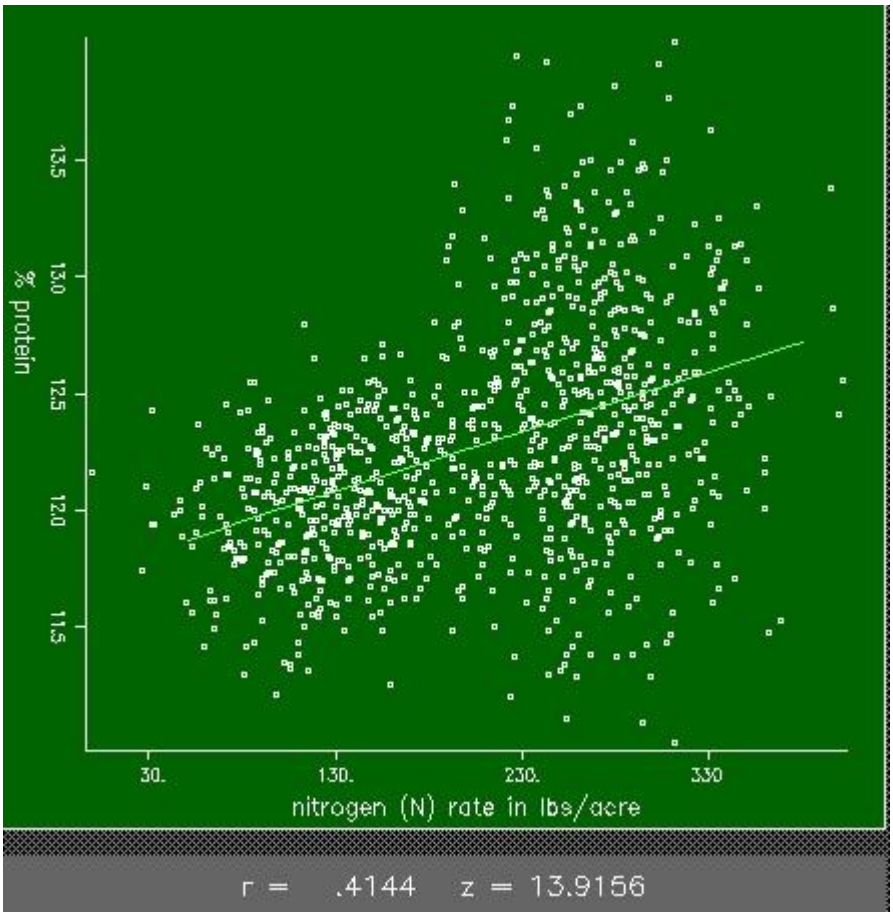

Figure 1. A Conventional Simple Linear Regression Curve obtained from data pertaining to nitrogen (N) rate in lbs/acre and % Protein.

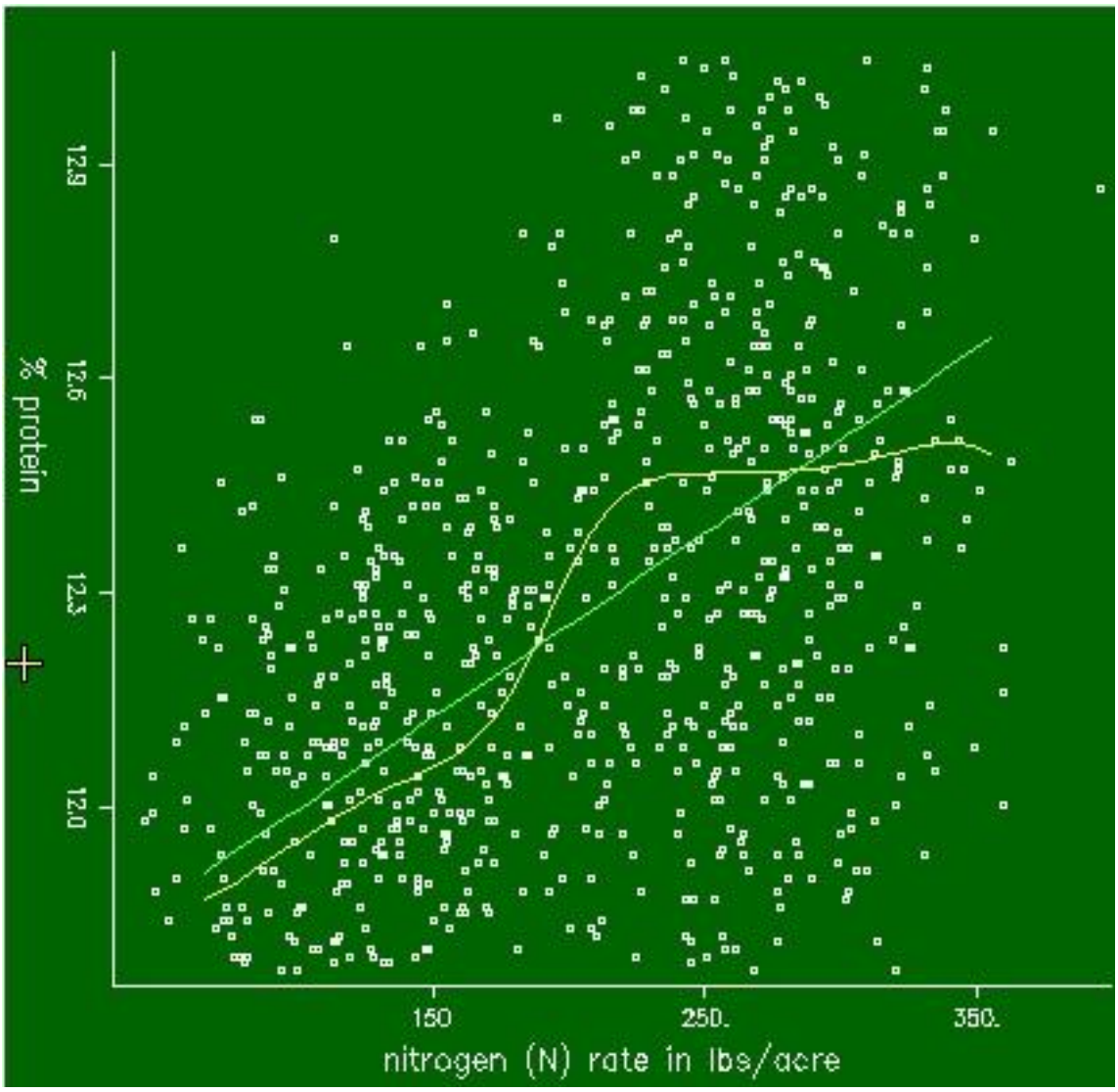

Figure 2. A Conventional Simple Linear Regression Curve (the straight line) and a Nonparametric Regression Curve capable of representing a flat lining of nitrogen (N) rate benefit beyond the value 250 lbs/acre.

The PC selected to obtain this curve was purposely chosen to be approximately five years old. Despite its venerable age, thanks to the speed and accuracy of modern digital computation the curves shown in Figure 2 were obtained in a few seconds. Indeed, there are some operations performed with the help of software used to construct these curves that, in an important sense, occur too quickly. Specifically, on a more modern PC or MAC computer certain graphical aids that had once helped to visualize the processes used to obtain important curves flash by so quickly that they are, today, unperceived by the viewer.

At one time the slash sign | that follows the letter *y* within, $f(x, y \mid z_0)$ could be seen interpreted as a vertical line quickly superimposed on a scatter diagram like that shown by Figures 1 and 2. On any modern PC or MAC this line is displayed so quickly that the estimate of the conditional density is shown a mere fraction of a second after the point z_0 through which the curve is sliced is selected.

To obtain the nonparametric regression curve shown in Figure 2, fifty equally spaced curve slices are selected automatically. Following the isolation of the fifty slices each individual slice is analyzed and from this slice a single point on the estimated nonparametric regression curve is determined. There is no need to assume that the curve constructed from a sequence of individual slice characteristics be a straight line or assume any other shape. However, nonparametric curves can be designed with specific applications in mind. For example, the yellow curve shown in Figure 2 was designed so that, were a single straight line actually appropriate, the yellow curve would approach a conventional least squares fitted line as the sample size increases. (The technical details that underlie the construction of the yellow curve are described by an attached paper.)

The representation $f(x, y \mid z_0)$ can be thought of as the symbolic counterpart of an important biological technique. Among the slides shown at the California Wheat Commission-sponsored JSM presentation this summer the slide that elicited the most audience response is shown in Figure 3.

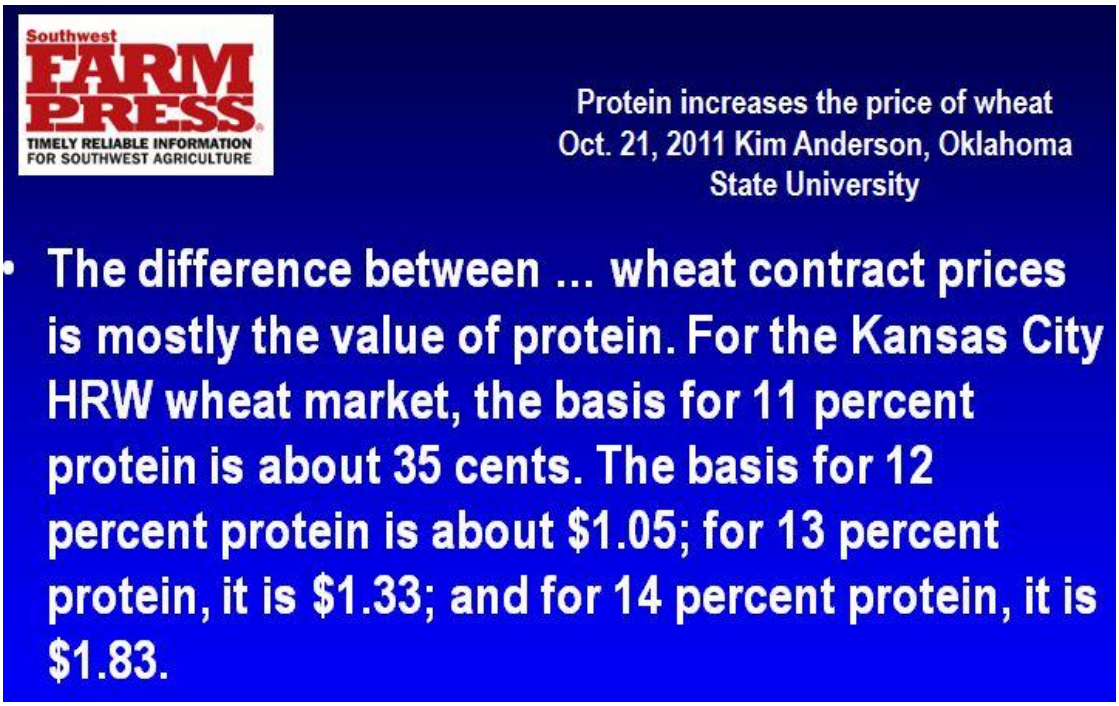

Figure 3. Four potential values of z_0 :11%; 12%; 13%; and 14%.

Figure 4 shows how a slice through a value close to 13% (specifically, though 13.056) was selected. Here $f(x, y \mid z_0 = 13.056)$ discloses how X and Y are distributed when 13.056% protein wheat is grown and provides, based on a measured value *x*, a suggested value of applied N, *y*.

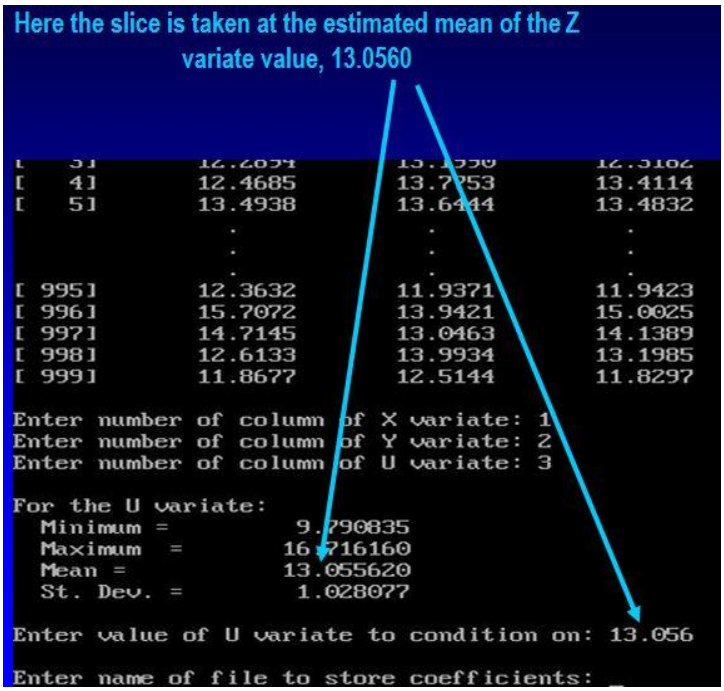

Figure 4. A slice taken through the mean value, here z_0 chosen to be 13.056%.

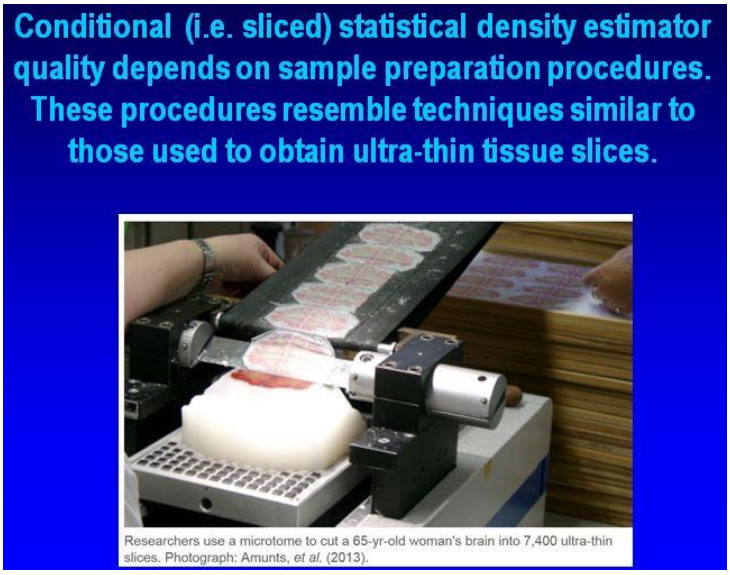

Figure 5. A biological technique that performs (with the help of a microtome), physically, what the interactive computer frame shown in Figure 4 is designed to perform, mathematically.

In summary, during the current year the problems addressed by the project concerned the gathering and preparation of measurements. This increased the scope and statistical power of studies of wheat-related interrelationships, such as those that characterize the curve $f(x, y \mid z_0)$. Here what the letter *f* represents is analogous to the color shown at a single microtome section's X and Y coordinates (as shown in Figure 5), while z_0 designates that the initial slice is being considered.

Regarding the Y variate, the quantity of applied N, one of the most important of this year's findings was that by using project-related methodology what (in technical terms) is called a marginal density, here, for example, *f*(*y*), can be assumed to have a symmetric and bell-shaped outline, as illustrated by Figure 6.

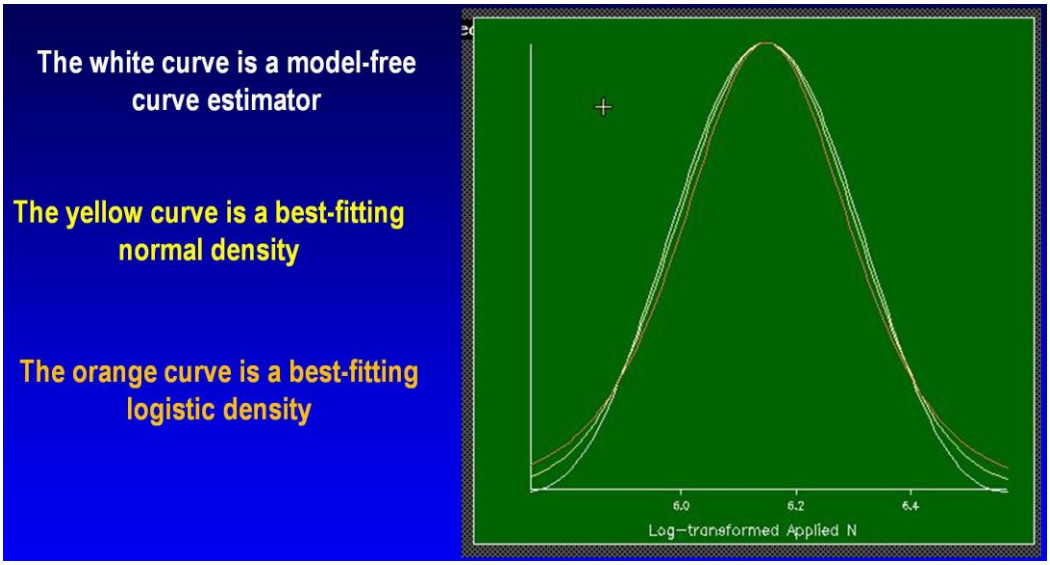

Figure 6. Three curves that are estimated by using the 80 Applied N determinations as data.

Discussion, Conclusions and Recommendations: Project implications and future research are discussed in a proposal to the California Wheat Commission dated Oct. 1. It is shown by Figure 6 that *f*(*y*) can be modified so that its modified form is bell-shaped and symmetric. By taking advantage of this finding, procedures for optimally determining applied N values *y* can be implemented. It is recommended that these determinations be implemented in a way that makes it easy for a grower to make use of findings obtained with the help of modern procedures (such the procedure that yielded the nonparametric regression curve shown in Figure 2).