301

IMPLEMENTING CONFIDENCE BANDS FOR SIMPLE LINEAR REGRESSION IN THE STATISTICAL LABORATORY PLOTTER PROGRAM

by

KRISTOPHER LEE ARHEART

B.S., Kansas State University, 1970
M.S., Kansas State University, 1973

A MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Statistics

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1978

Approved by:

K. E. Kemp
Major Professor

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Computing the y coordinates for upper and lower confidence bands for a simple linear regression by hand can be time consuming if there are many observations. These computations can be done easily by a simple computer program if one has basic programming skills, but graphically representing the confidence bands is another matter.

Drawing the confidence bands by hand is a tedious task if there are many observations. Additionally, confidence bands produced by hand are not always accurate. Using the CalComp model 663 incremental drum plotter, available at the Kansas State University Computing Center, requires moderate programming skills. Plotting confidence bands in a form acceptable for inclusion in a report demands even greater programming skill and effort. Considering these facts, how does a researcher with basic programming skills produce a good pictorial representation of confidence bands for a simple linear regression?

Heretofore, no computer package existed which would plot confidence bands for simple linear regression; now, the Statistical Laboratory PLOTTER program will. This paper reports the addition of this feature.

IMPLEMENTATION

The Statistical Laboratory PLOTTER program was chosen to provide confidence bands for two reasons. First, little programming skill is needed to use the program, making it fairly accessible to all researchers. Second, the program already plots the regression line and observations, so the addition of confidence bands is relatively inexpensive. Only simple linear regression confidence bands are considered because plotting more complex models requires more than two dimensions. PLOTTER handles only two dimensional drawings.[1]

Adding the confidence bands was a three-fold problem. A control card was added to request confidence bands and to specify options. The appropriate computations were incorporated to compute the points on the confidence bands. Plotting the confidence bands was the last task.

A control card of the form:

```
CB,T=t,EMS=r,FN=q
```

was added to specify confidence band plots. Including a valid CB card produces a plot of data points, the regression line,

[1] Routines are available to draw isometric representations of three dimensional objects, so it would be possible to handle a regression with two independent variables. However, the existing program would have to be modified extensively, or a new one written, which would be expensive.

and the desired confidence bands. The argument t provides the appropriate Student's t value for computing the desired confidence bands. The residual mean square from the regression analysis is represented by the argument r. Argument q controls the type of confidence bands that are plotted. If q has the value one, confidence bands about a single future observation are plotted. When q is a positive integer greater than one, prediction bands about the mean of q future observations are plotted. If q is omitted or given the value of zero, confidence bands are plotted about $\mu_{Y \cdot X}$. The CB card is valid only when a MODEL card specifying a simple linear regression has been included; otherwise, the CB card is ignored. Dr. Kenneth Kemp of the Kansas State University Statistical Laboratory implemented the control card.

To compute the y coordinates for the confidence bands, the following equation must be evaluated:

$$P_i = \hat{Y}_i \pm t_{\alpha/2}(n-2)\sqrt{s^2\{1/q+1/n+[(X_i-\bar{X})^2/\Sigma(X_j-\bar{X})^2]\}} \ .$$

P_i is either the upper or lower y coordinate corresponding to the i^{th} x coordinate. The upper y value is produced by adding, the lower by subtracting. The predicted y value at the i^{th} observation is $\hat{Y}_i$. The appropriate Student's t value for a confidence band of width α is $t_{\alpha/2}(n-2)$, where n is the