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

	Page
INTRODUCTION	1
IMPLEMENTATION	2
RESULTS	6
BIBLIOGRAPHY	13
APPENDIX A	14
ACKNOWLEDGMENTS	16

LIST OF TABLES

Table Page

1. Data for Figure 1 8
2. Data for Figure 2 10
3. Data for Figure 3 12

LIST OF FIGURES

Figure		Page
1.	$CI_{95\%}$ About $\mu_{Y \cdot X}$	7
2.	$CI_{95\%}$ About a Single Future Observation	9
3.	$CI_{95\%}$ About the Mean of Five Future Observations .	11

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 number of observations in the input data.[2] S^2 is the residual mean square from a regression analysis. The 1/q term resolves to zero for confidence bands about $\mu_{Y \cdot X}$, to one for bands on a single future observation, and to 1/q for bands about the mean of q future observations. The term 1/n is the reciprocal of the number of observations. X_i is the i^{th} observation of X; $\bar{X}$ is the sample mean of X; and $\Sigma(X_j-\bar{X})^2$ is the X sum of squares. The above equation is evaluated for each observed X.

The existing program provides $\hat{Y}_i$, X_i, n, ΣX_i, and $\Sigma(X_i-\bar{X})^2$. The control card gives $t_{\alpha/2}(n-2)$, S^2, and q. To insure smooth confidence band curves, twenty equidistant X and $\hat{Y}$ coordinates are generated beginning with the minimum X value and ending with the maximum X value. Previous study by Dr. Kenneth Kemp determined that twenty points give smooth curves for the nine by nine inch grid on which the data are plotted. The constants $A=t_{\alpha/2}(n-2)\sqrt{S^2}$, $B=1/q+1/n$, $C=\Sigma X_i/n$, and $D=1/\Sigma(X_i-\bar{X})^2$ are computed to avoid needless repetitive computation. As each $\hat{Y}_i$ is computed, the constants $E=X_i-C$ and $F=A\sqrt{B+E \cdot E \cdot D}$ are created, and the upper and lower y coordinates are evaluated as $U_i=\hat{Y}_i+F$ and $L_i=\hat{Y}_i-F$, respectively. Thus, the computations of the confidence band are completed. One may verify that the steps outlined above reproduce the equation in the preceding paragraph.

---

[2] It should be noted that the same data used to compute the regression model given on the MODEL card and the information on the CB card must be used to plot the confidence bands. Several of the terms used to compute the confidence points are recalculated by PLOTTER, so using different data will produce incorrect plots.

Using the procedure outlined above gives confidence bands that have 1-α% confidence at each X point. Such bands are often referred to as one-at-a-time confidence bands. To plot simultaneous confidence bands that have at least 1-α% confidence over all X points, substitute the appropriate value for the argument T on the CB control card. For example, Scheffé confidence bands are obtained by using the value of $\sqrt{2[F_{\alpha}(2,n-2)]}$, where $F_{\alpha}(2,n-2)$ is the Fisher's F value at the specified α level with 2 and n-2 degrees of freedom (n is the number of observations). Further discussion of the different types of simultaneous confidence bands can be found in Graybill (1976).

Plotting the confidence bands requires two steps in addition to those usually taken by PLOTTER. First, calls to the CalComp SCALE subroutine are made to obtain the minimum point and increment value used to scale the Y axis. Then calls to the CalComp LINE subroutine are used to draw the confidence bands.

RESULTS

The data used to illustrate the confidence band plots was taken from Table 1.1 of Applied Regression Analysis by Draper and Smith (1966). The predictor equation for the 25 X,Y pairs was $\hat{Y}_i=13.62-.08X_i$ with a residual mean square of .79. Figure 1 shows 95% confidence bands about $\mu_{Y\cdot X}$. The data for Figure 1 is given in Table 1. The 95% confidence bands for a single future observation are given in Figure 2, with corresponding data points given in Table 2. Figure 3 and Table 3 present the 95% confidence bands for the mean of five future observations and the data points, respectively. An example of the control card deck to produce all three figures is given in Appendix A. Listings of the modified PLOTTER program are available from the Kansas State University Statistical Laboratory.

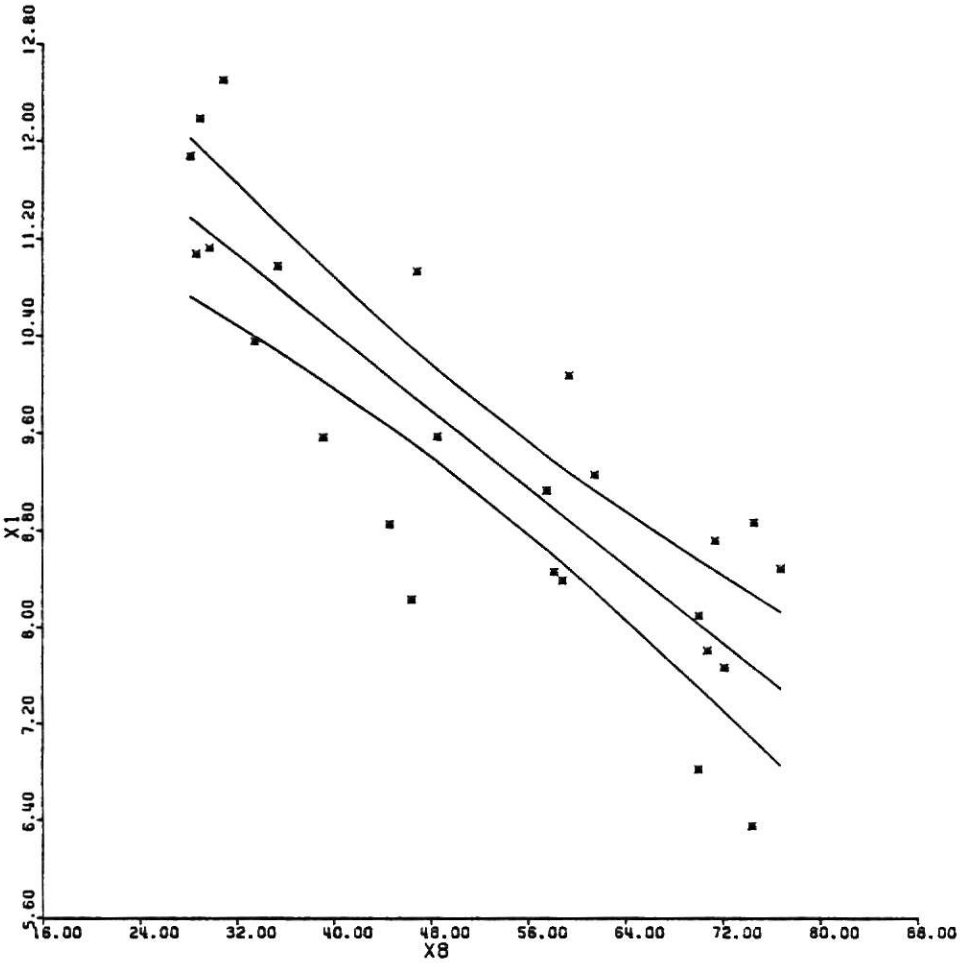

Figure 1. $CI_{95\%}$ About $\mu_{Y \cdot X}$

TABLE 1

DATA FOR FIGURE 1

Observations		Predicted Y Value	Confidence Points	
X	Y		Lower	Upper
28.1	11.88	11.38	10.73	12.03
28.6	11.08	11.34	10.70	11.98
28.9	12.19	11.32	10.68	11.95
29.7	11.13	11.25	10.63	11.87
30.8	12.51	11.17	10.56	11.76
33.4	10.36	10.96	10.40	11.51
35.3	10.98	10.81	10.28	11.33
39.1	9.57	10.50	10.03	10.97
44.6	8.86	10.06	9.66	10.47
46.4	8.24	9.92	9.53	10.31
46.8	10.94	9.89	9.50	10.28
48.5	9.58	9.75	9.37	10.13
57.5	9.14	9.03	8.65	9.42
58.1	8.47	8.98	8.60	9.37
58.8	8.40	8.93	8.54	9.32
59.3	10.09	8.89	8.50	9.28
61.4	9.27	8.72	8.31	9.14
70.0	8.11	8.04	7.51	8.56
70.0	6.83	8.04	7.51	8.56
70.7	7.82	7.98	7.44	8.52
71.3	8.73	7.93	7.38	8.48
72.1	7.68	7.87	7.31	8.43
74.4	6.36	7.68	7.08	8.28
74.5	8.88	7.68	7.08	8.28
76.7	8.50	7.50	6.86	8.14

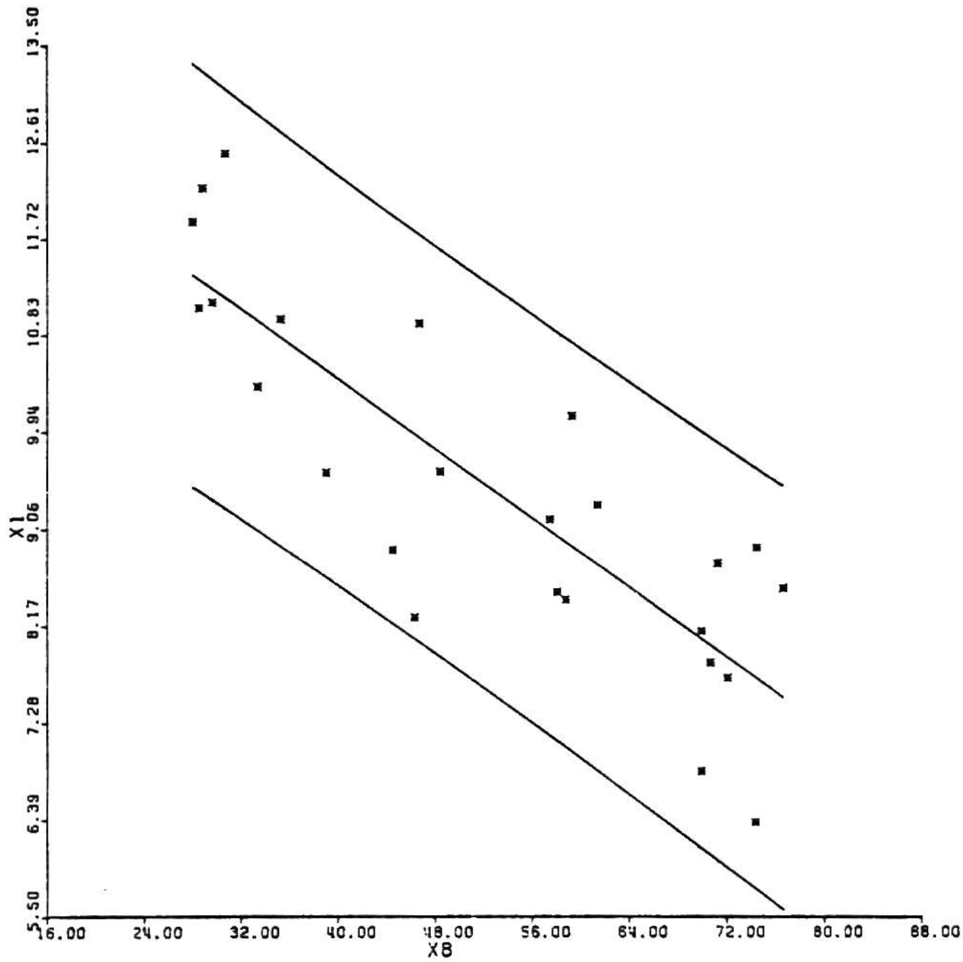

Figure 2. $CI_{95\%}$ About a Single Future Observation

TABLE 2

DATA FOR FIGURE 2

Observations		Predicted Y Value	Confidence Points	
X	Y		Lower	Upper
28.1	11.88	11.38	9.43	13.33
28.6	11.08	11.34	9.40	13.29
28.9	12.19	11.32	9.37	13.26
29.7	11.13	11.25	9.31	13.19
30.8	12.51	11.17	9.23	13.10
33.4	10.36	10.96	9.04	12.88
35.3	10.98	10.81	8.89	12.72
39.1	9.57	10.50	8.61	12.40
44.6	8.86	10.06	8.18	11.94
46.4	8.24	9.92	8.04	11.80
46.8	10.94	9.89	8.01	11.76
48.5	9.58	9.75	7.88	11.63
57.5	9.14	9.03	7.16	10.91
58.1	8.47	8.98	7.11	10.86
58.8	8.40	8.93	7.05	10.81
59.3	10.09	8.89	7.01	10.77
61.4	9.27	8.72	6.84	10.61
70.0	8.11	8.04	6.12	9.95
70.0	6.83	8.04	6.12	9.95
70.7	7.82	7.98	6.06	9.89
71.3	8.73	7.93	6.01	9.85
72.1	7.68	7.87	5.95	9.79
74.4	6.36	7.68	5.75	9.62
74.5	8.88	7.68	5.74	9.61
76.7	8.50	7.50	5.55	9.45

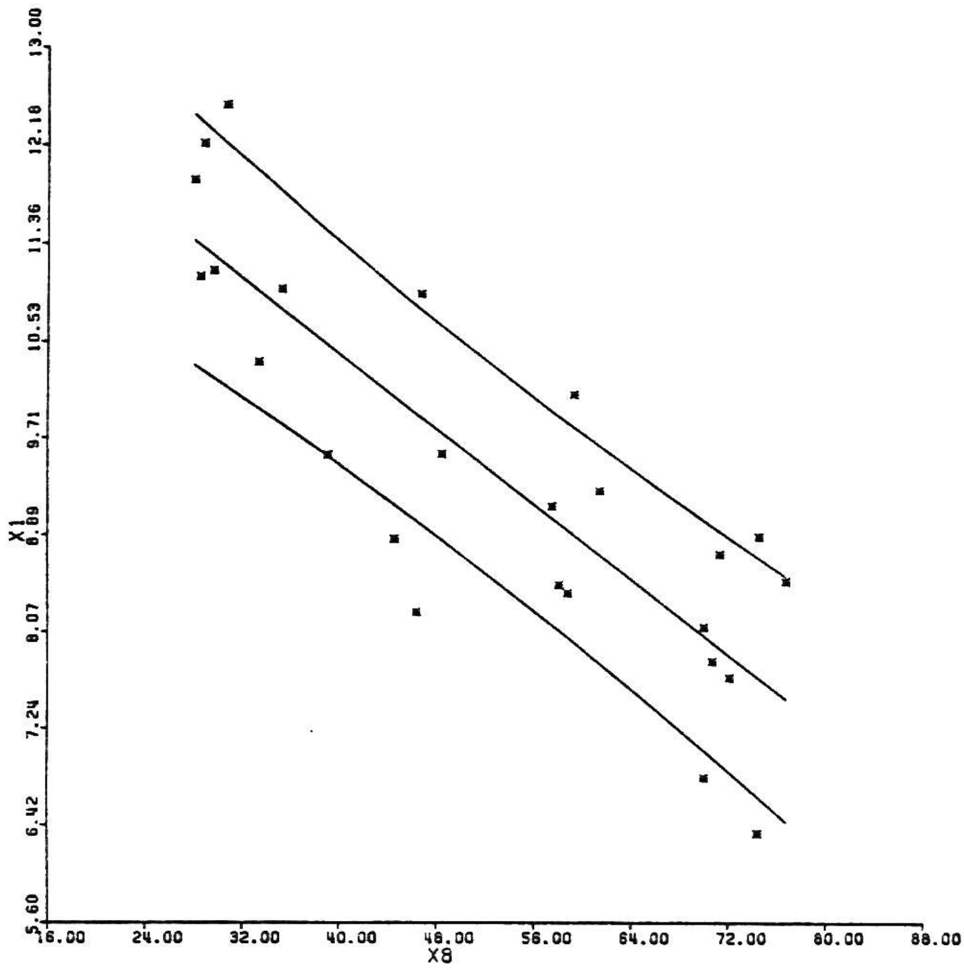

Figure 3. $CI_{95\%}$ About the Mean of Five Future Observations

TABLE 3

DATA FOR FIGURE 3

Observations		Predicted Y Value	Confidence Points	
X	Y		Lower	Upper
28.1	11.88	11.38	10.33	12.43
28.6	11.08	11.34	10.30	12.38
28.9	12.19	11.32	10.28	12.35
29.7	11.13	11.25	10.22	12.28
30.8	12.51	11.17	10.15	12.18
33.4	10.36	10.96	9.96	11.95
35.3	10.98	10.81	9.83	11.78
39.1	9.57	10.50	9.56	11.45
44.6	8.86	10.06	9.15	10.98
46.4	8.24	9.92	9.01	10.83
46.8	10.94	9.89	8.98	10.80
48.5	9.58	9.75	8.85	10.66
57.5	9.14	9.03	8.13	9.94
58.1	8.47	8.98	8.08	9.89
58.8	8.40	8.93	8.02	9.84
59.3	10.09	8.89	7.98	9.80
61.4	9.27	8.72	7.80	9.64
70.0	8.11	8.04	7.06	9.01
70.0	6.83	8.04	7.06	9.01
70.7	7.82	7.98	7.00	8.96
71.3	8.73	7.93	6.94	8.92
72.1	7.68	7.87	6.87	8.86
74.4	6.36	7.68	6.67	8.70
74.5	8.88	7.68	6.66	8.69
76.7	8.50	7.50	6.46	8.54

BIBLIOGRAPHY

Buckley, Adrian C., Huff, Karen E., and Vore, Richard L. Plotter Guide. Version 2.1. Manhattan, Kansas: Kansas State University Computing Center, 1976.

Draper, N. R. and Smith, H. Applied Regression Analysis. New York: John Wiley and Sons, Inc., 1966.

Graybill, Franklin A. Theory and Application of the Linear Model. North Scituate, Massachusetts: Duxbury Press, 1976.

Kemp, K. E., et al. A User's Guide for a Plotter Program Package for the IBM 360-70 (PLOTTER). Manhattan, Kansas: Department of Statistics and Agricultural Experiment Station, Kansas State University, n.d.

Snedecor, George W. and Cochran, William G. Statistical Methods. 6th edition. Ames, Iowa: The Iowa State University Press, 1967.

APPENDIX A

Program to Produce Figures 1-3

```
//CI95      JOB      (xxxxxxxxx,xxxxxxxx,,1),ARHEART,TIME=(,5)
/*TAPE9
//STEP1     EXEC     STATPROG,ROUTINE=PLOTTER
//STATPROG.PLOTTAPE DD UNIT=TAPE9,VOL=PRIVATE,DISP=(,KEEP)
//STATPROG.SYSIN DD *
DATA,2,,1.
LABEL(1)=X1,X8
MODEL,Y1=13.623005-.079829X2
CB,T=2.064,EMS=.7926
END
(2X,F4.2,F3.1)
011098353
021113297
031251308
040840588
050927614
060873713
070636744
080850767
090782707
100914575
110824464
121219289
131188281
140957391
151094468
160958485
171009593
180811700
190683700
200888745
210768721
220847581
230886446
241036334
251108286
END OF DATA
DATA,2,,2.
MODEL,Y1=13.623005-.079829X2
CB,T=2.064,EMS=.7926,FN=1
```

```
END
DATA,2,,2.
MODEL,Y1=13.623005-.079829X2
CB,T=2.064,EMS=.7926,FN=5
END
/*
```

ACKNOWLEDGMENTS

I would like to thank my advisory committee, Dr. Kenneth Kemp, chairman, Dr. Dallas Johnson, and Dr. Raja Nassar for their guidance. Special thanks are given to Dr. Kemp for suggesting the problem and giving freely of his time. I also appreciate the assistance and support given by my colleagues at the Computing Center. My family has my gratitude for their support and encouragement, especially my wife, Linda, and our parents.

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

AN ABSTRACT OF 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

ABSTRACT

This report details the installation of the ability to plot a set of confidence bands for simple linear regression in the Statistical Laboratory PLOTTER program. This paper covers the implementation of the control card, the computations, and the plot. Options exist to plot confidence bands about $\mu_{Y \cdot X}$, about a single future observation, or about the mean of q future observations.