

**The General Structural Relation through SAS/AF windows:
an IML Macro you can't see.**

**Robert J. Zambarano
Statistical/Mathematical Services
Computation Center
University of Texas at Austin
Austin, TX**

Estimation the slope of the relationship between two variables is a common statistical problem, and there are many estimation methods available. Perhaps the most frequently used method for slope estimation is least squares regression. When the necessary assumptions are met, this approach is simple and powerful. When the necessary assumptions are violated, considerably biased estimates are produced. One crucial assumption of least squares regression is that the X axis variables (independent variables) are measured without error (Hays, 1988). Least square confidence intervals are estimated only in terms of the Y axis (the dependent variable). (See overhead #1.) When a variable represented along the X axis is measured with error, confidence zones should actually be ellipses (Kendall & Stuart, 1979). Data with elliptical confidence zones will have a different line-of-best-fit than data with Y axis confidence intervals. (See overhead #2.) When both Y and X are measured with error, least square estimates can be considerably biased, and this approach should not be used (Raynor, 1985).

General Structural Relation estimation

An alternative method is the General Structural Relation (GSR), which takes into account the error of any variable in a relationship (Kendall & Stuart, 1979). This approach has been developed in steps over the last 40 years, so it is relatively new. As a demonstration of the relative effectiveness of this approach, consider the following Monte Carlo study. One hundred data points were selected at random from a normal distribution with a mean of zero and a standard deviation of fifty. These values represent X. Y will be

defined as the following function of X: $Y=3X + 10$. To each value of X and Y, a random value will be added. For X this value will be randomly selected from a normal distribution with a mean of zero and a standard deviation of ten. For Y this value will be randomly selected from a normal distribution with a mean of zero and a standard deviation of two. These values represent error of measurement for X and Y. Five hundred such data sets were produced, and the parameter estimates were produced for each. The first overhead shows the sampling distributions for the estimates of the slope - first for the General Structural Relation, and then for least squares regression. (See overhead #3.) Recall that the parameter values are slope=3 and intercept=10. As you can see, the least square method produces a sampling distribution which is biased. The second overhead shows the sampling distributions for the estimates of the intercept - again, General Structural Relation on top, least square estimates on the bottom. (See overhead # 4.) For this parameter, both methods work about as well.

SAS and the General Structural Relation

The method is gaining acceptance, yet it cannot be performed on any widely available statistical software. For the simplest case of two variables, the calculations are fairly simple and can be readily performed by PROC IML. Unfortunately, IML is not a simple procedure to learn. There is an entire reference manual published for it (SAS Inst., 1989). While IML is very powerful, it is also fairly complex. Thus, this sort of programming does not attract many SAS users. To make the process as simple as possible for the user without actually doing the user's work, a generic IML program can be written using SAS MACRO language. This code can be embedded in a series of SAS AF windows, allowing the user to perform the desired statistics and produce customized output.

The Program

Using SAS version 6.04 for DOS, such a program was produced. This program works for the case of two variables, although the General Structural Relation is not limited in this way. A DOS exec file

and an AUTOEXEC.SAS file accompany the AF catalog, so that the AF program set can be used either directly from the DOS prompt, or from the SAS display manager. (The AUTOEXEC.SAS file must reflect the correct path to the directory where the AF catalog is stored.) To initialize from DOS, use the command "GSR". To initialize from the display manager, first define the path to the AF catalog using a LIBNAME statement, then from the command line of the display manager issue the command "AF c=struc.beta.screen1.program".

Either of these methods will bring the user to the first screen. (See overhead #5.) This screen is there simply to provide the initial user some information about the method and the relationship that is being modeled. When the user is ready to continue, the Enter/Return key is pressed.

The second screen gives instructions as to what the user must have to move through the program. If confused, the user can go back to the first screen, or if the necessary information does not exist to produce the statistics, the user can quit the program. (See overhead #6.)

The next screen prompts the user for the path to the data set, and the name of the data set. When done, pressing Enter/Return produces the next screen. This screen is programed so that the user cannot move to the next screen without first entering a value in both fields. (See overhead #7.)

If the user hits the enter key in error, or realizes that the wrong information was input on the last screen, this screen provides an option to return and make changes. Primarily, this screen prompts the user for the names of the two variables. Again, the user cannot get to the next screen without entering something in both fields. (See overhead #8.)

Again the user can move back to the previous screen and make changes. Assuming all is well, the user will now specify what is

known about the measurement error of the X and Y variables. Selection is made by tabing to the appropriate item number, and then pressing the Enter/Return key. (See overhead #9).

Depending on the user's selection, one of four screens appears. All are essentially the same. The user is prompted for the appropriate information and told this is the last screen to allow changes. For example, if we had selected option #4 on the last screen, this is the screen we would see next. (See overhead #10.)

The next screen asks the user to select the desired output. Possible output is either statistics, an overlay plot of the data and the estimated slopes, or both. (See overhead #11.)

If only statistics were selected, the program is complete; the results are sent to the Display manager Output window. If the plot is selected, there is one more screen to see. The user is asked to indicate the type of device, high resolution or standard. Also, if the device is high resolution the user can decide to have the plot sent immediately, or sent to a graphics stream file for processing at a later time. (See overhead #12.)

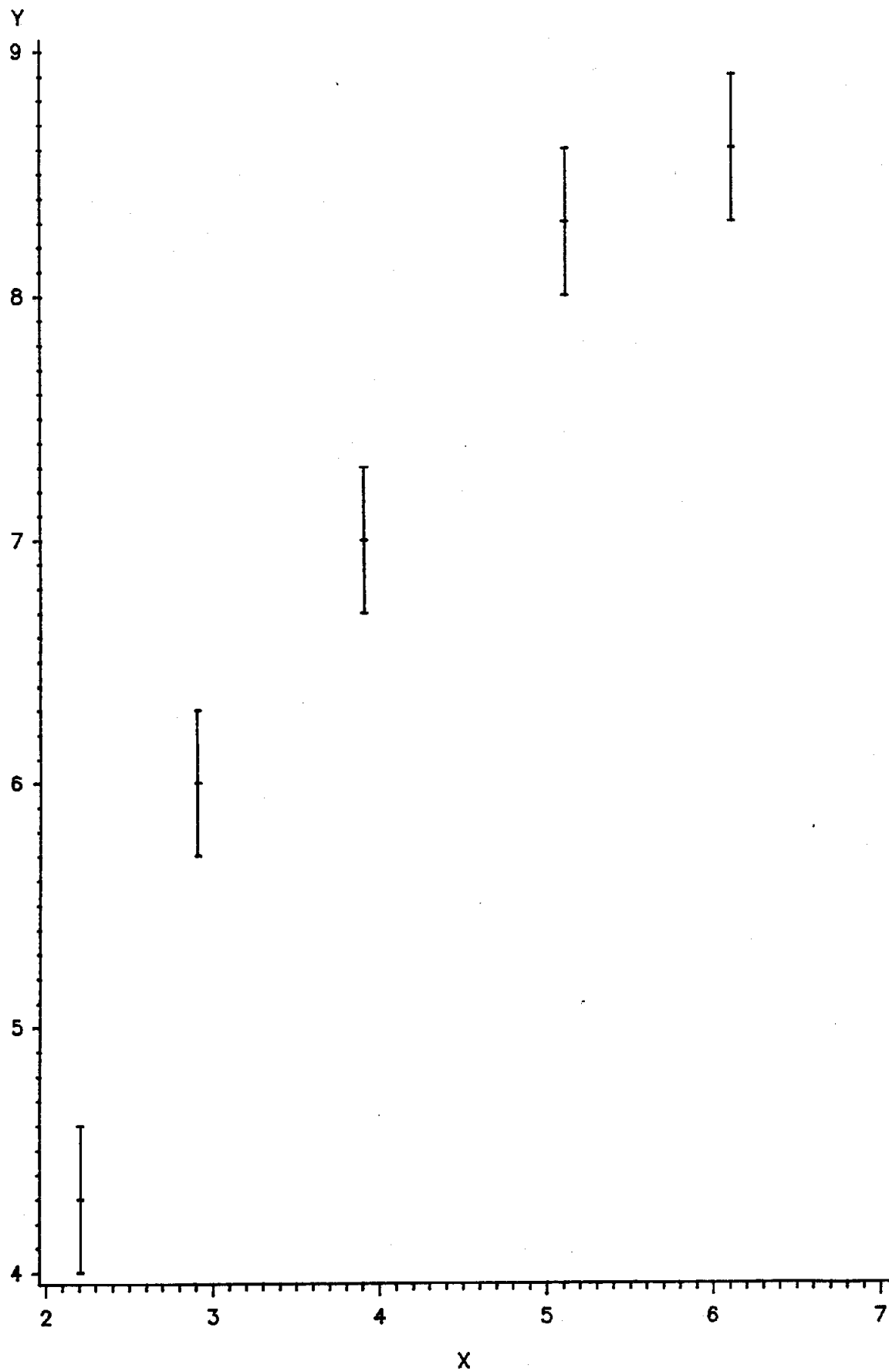
Here are the statistics from a artificial data set with slope=3 and intercept=10. (See overhead #13.) This is the plot of the data. (See overhead #14.)

This project is still in development. In the future I plan to increase the options for output, and also increase the number of variables that can be included in the model. I would be glad to provide copies of the program to anyone who sends me a blank disk and the stamps to cover return shipping charges.

References

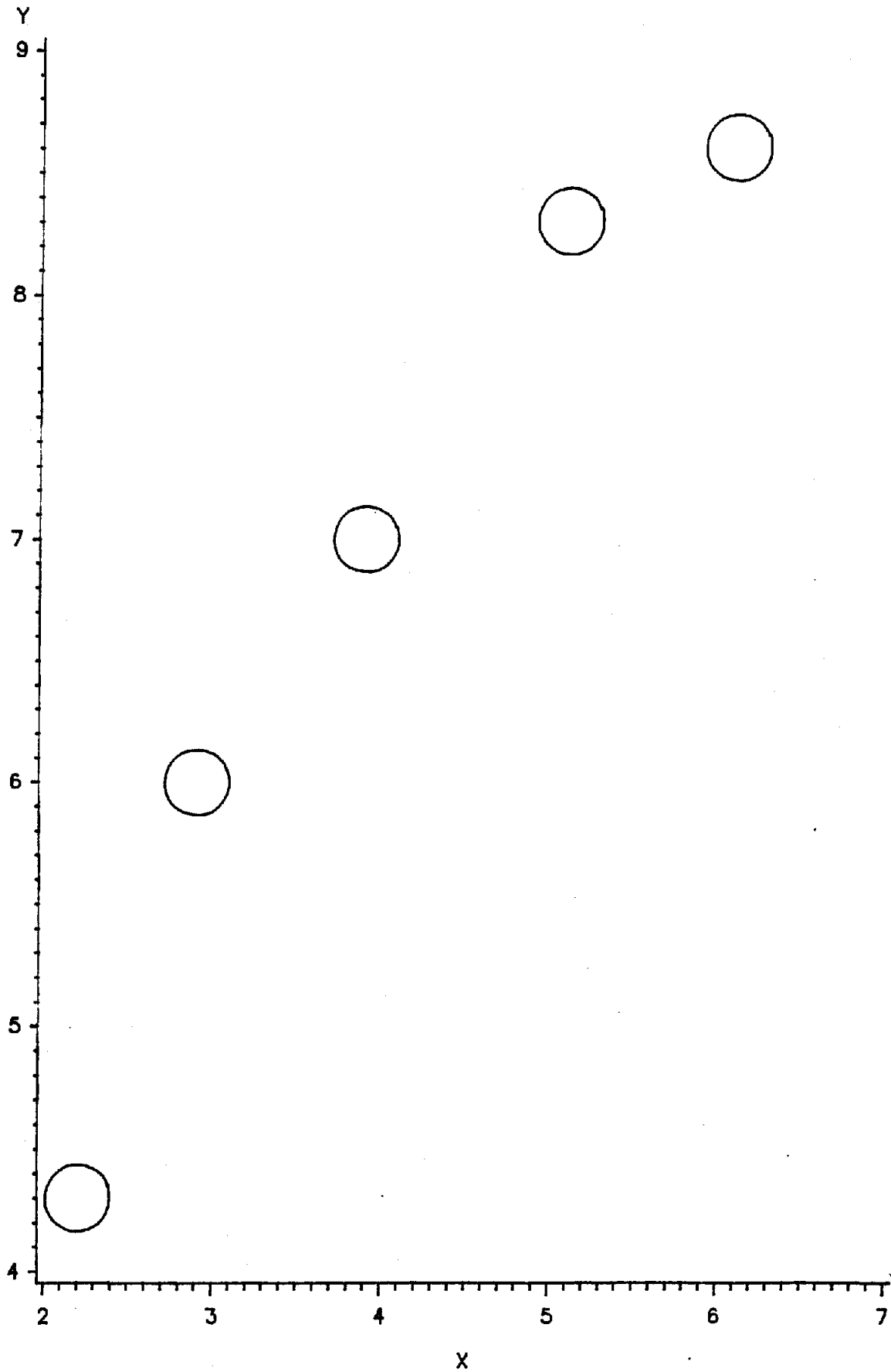
- Hays, W.L. (1988). Statistics (4th ed.), New York: Holt, Reinhart & Winston.
- Kendall, M.G. & Stuart, A. (1979), The Advanced Theory of Statistics, Vol. 2, New York, London: Oxford University Press.
- Rayner, J.M.V. (1985). Linear relations in biomechanics: the statistics of scaling functions. Journal of Zoology, London (A), (206), 415-439.
- SAS Institute Inc., (1989). SAS/AF Software: Usage and Reference, Version 6, First Edition, Cary, NC: SAS Institute Inc.
- SAS Institute Inc., (1990). SAS Guide to Macro Processing, Version 6, Second Edition, Cary, NC: SAS Institute Inc.
- SAS Institute Inc., (1989). SAS/IML Software: Usage and Reference, Version 6, First Edition, Cary, NC: SAS Institute Inc.
- SAS Institute Inc., (1990). SAS Screen Control Language: Usage and Reference, Version 6, First Edition, Cary, NC: SAS Institute Inc.

OVERHEAD #1: CONFIDENCE INTERVALS

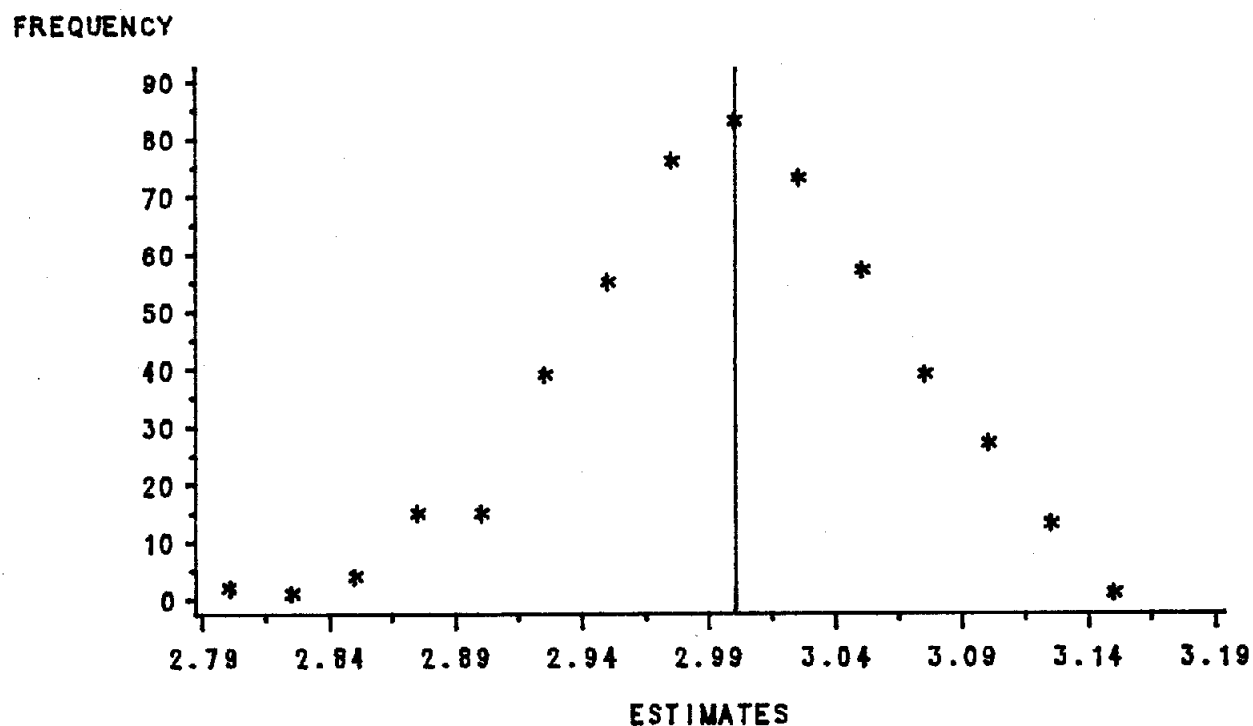


X

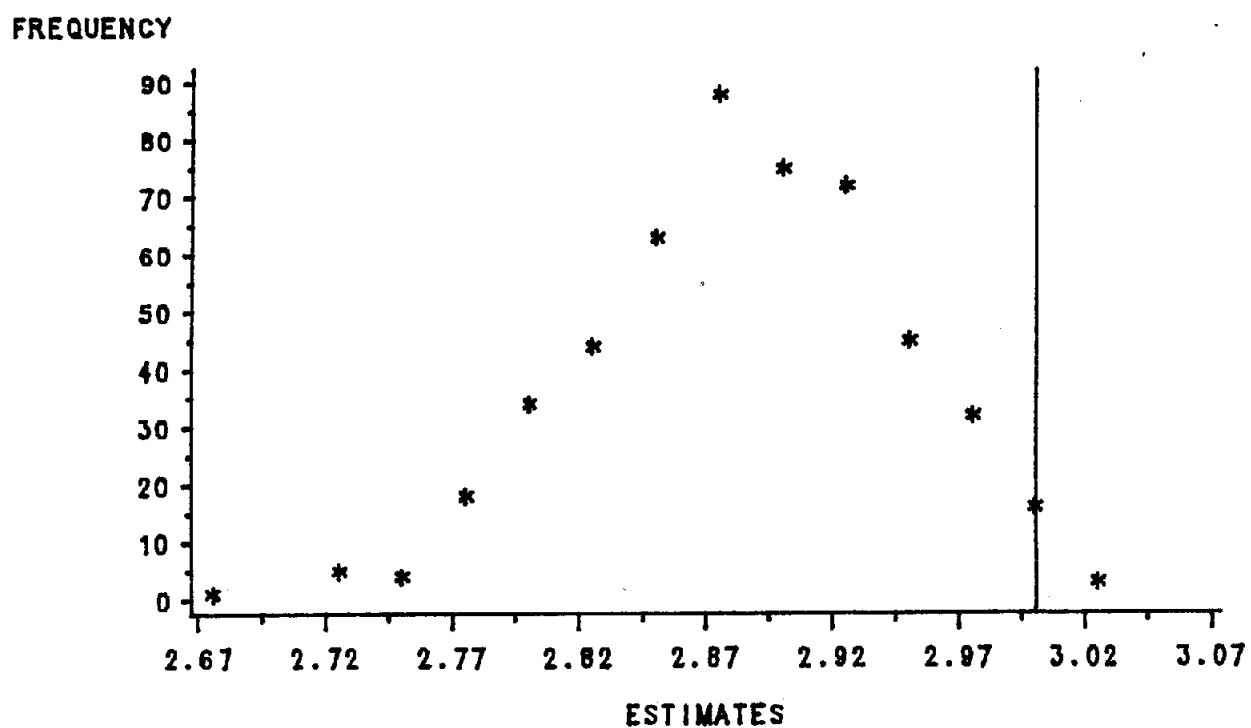
OVERHEAD #2: CONFIDENCE ELLIPSES



OVERHEAD #3: GSR ESTIMATES OF THE SLOPE

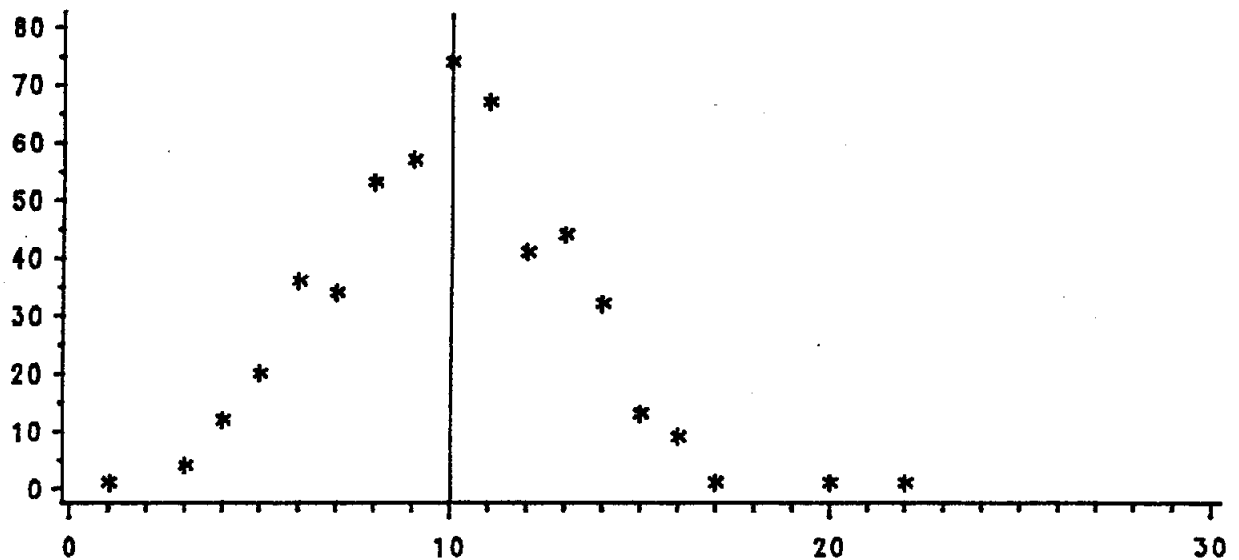


LSR ESTIMATES OF THE SLOPE



OVERHEAD #4: GSR ESTIMATES OF THE INTECEPT

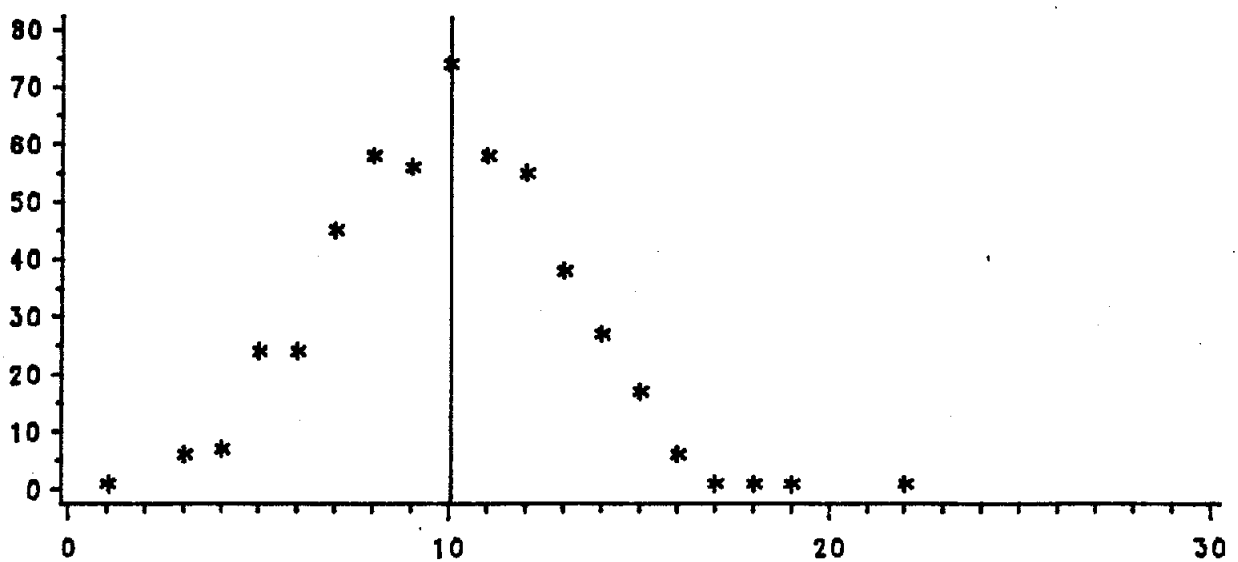
FREQUENCY



ESTIMATES

LSR ESTIMATES OF THE INTERCEPT

FREQUENCY



ESTIMATES

Overhead #5 - Initial screen

__AF-----

| Command ==>

| WHAT IT IS AND WHAT IT DOES

| The program you are using calculates a linear slope of best fit
| between two variables. The method is the General Structural Relation.
| (See Kendall and Stuart, 1979, THE ADVANCED THEORY OF STATISTICS, VOL.
| 2, chapter 29. New York, London: Oxford University Press.)

| In contrast to least square procedures, the General Structural
| Relation does not assume the error-free measurement of X axis
| variables, thus producing unbiased estimates of the (X,Y) relationship
| when both variables are measured with error. This condition is
| described as follows:

$$Y_{\text{observed}} = Y_{\text{actual}} + Y_{\text{error}}$$

$$X_{\text{observed}} = X_{\text{actual}} + X_{\text{error}}$$

| When you're ready,
| hit the Enter/Return key to continue....

Overhead #6 - Screen two

...AF-----

|Command ==>

|
|WHAT YOU MUST HAVE

- |1) A permanent SAS dataset including the two variables of interest.
- |2) One of the following:
- |a. The variance of Y_error, (see NOTE!),
 - |b. The variance of X_error, (see NOTE!),
 - |c. Both of the above, (see NOTE!),
 - |d. The ratio of variance Y_error over variance X_error, (see NOTE!).

|NOTE!: these values are not equivalent to the variance of Y_observed
|and/or X_observed. See the previous screen.

|Select one of the following - Tab key moves, Enter/Return selects.

|Continue

|See previous screen

|QUIT

Overhead #7 - Screen 3

...AF-----

| Command ===>

|
| WHERE'S YOUR DATA SET, AND WHAT'S IT CALLED?
|

| Use the Tab key to move to each field.

| There are two fields on this screen - fill in both.

| Enter the DOS path to the SAS dataset including the variables of
| interest.

| DOS path?

| (example: C:\data\sasdata)

| Enter the two part SAS name of the SAS dataset.

| SAS dataset?

| (example: struc.XY_error)

Hit Enter/Return when you are done.

Overhead #8 - Screen four

...AF-----

| Command ==>

|
| WHAT ARE THE VARIABLES?

| Use the Tab key to move to each field.

| There are two fields on this screen - fill in both.

| What is the name of the SAS variable that corresponds to Y_observed?

| Y_observed?

| What is the name of the SAS variable that corresponds to X_observed?

| X_observed?

| Hit Enter/Return when you are done

| *****
| * To make changes to the last screen, move the *
| * cursor to the word "CHANGES" and hit Enter/Return.*
| * *
| * CHANGES *
| *****

...AF-----

|Command ==>

|WHAT DO YOU KNOW ABOUT THE VARIANCE OF MEASUREMENT ERRORS?

|You may only select one of the following.

|Use the Tab key to move to the value you have, then hit enter.

- |1. Variance of measurement error in y
- |2. Variance of measurement error in x
- |3. Variance of measurement error in both y and x
- |4. Ratio of the measurement error variances of y/x

|*****

|* To make changes to the last screen, move the *

|* cursor to the word "CHANGES" and hit Enter/Return.*

|* *

|* CHANGES *

|*****

Overhead #10 - Screen 6

```
...AF-----
|Command ===>
|
|                ENTER THE VALUE
|
|What is the value of the Ratio of measurement error variance in y
|over measurement error variance in x?
|
| Value?
|
|When you hit Enter/Return, all statistics will be computed - be certain
|that you have correctly typed the above value
|
|
|                *****
|                *      To make changes to the last screen, move the      *
|                * cursor to the word "CHANGES" and hit Enter/Return.*
|                *
|                *                CHANGES                *
|                *****
```

Overhead #11 - Screen 7

...AF-----

| Command ===>

|
| WHAT DO YOU WANT TO SEE?
|

| Select only one of the following. Use the Tab key to move to the
| output you wish to see, then press Enter/Return.

- | 1. The estimates from the General Structural Relation.
| (Including, for comparison, estimates from standard
| least squares regression).
- | 2. An overlay plot of: the original data points;
| the line-of-best-fit from the General Structural Relation;
| the line-of-best-fit from standard least squares regression.
- | 3. Both of the above

Overhead #12 - Screen 8

..AF-----

| Command ==>

| HOW DO YOU WANT YOUR PLOT?

| Do you have standard or high resolution graphics capabilities?

| If you want a high resolution plot, do you want it sent directly
| to display, or would you prefer a graphics stream file?

| Make one selection from the following - Tab key moves, Enter/Return
| selects.

- | 1. Standard graphics (plot sent to Display Manager Output Window).
- | 2. High resolution (plot sent to display - You will be prompted for
| the device name).
- | 3. High resolution (plot sent to graphics stream file - You will be
| prompted for the device name).

Overhead #13 - output

..OUTPUT-----

|Command ==>

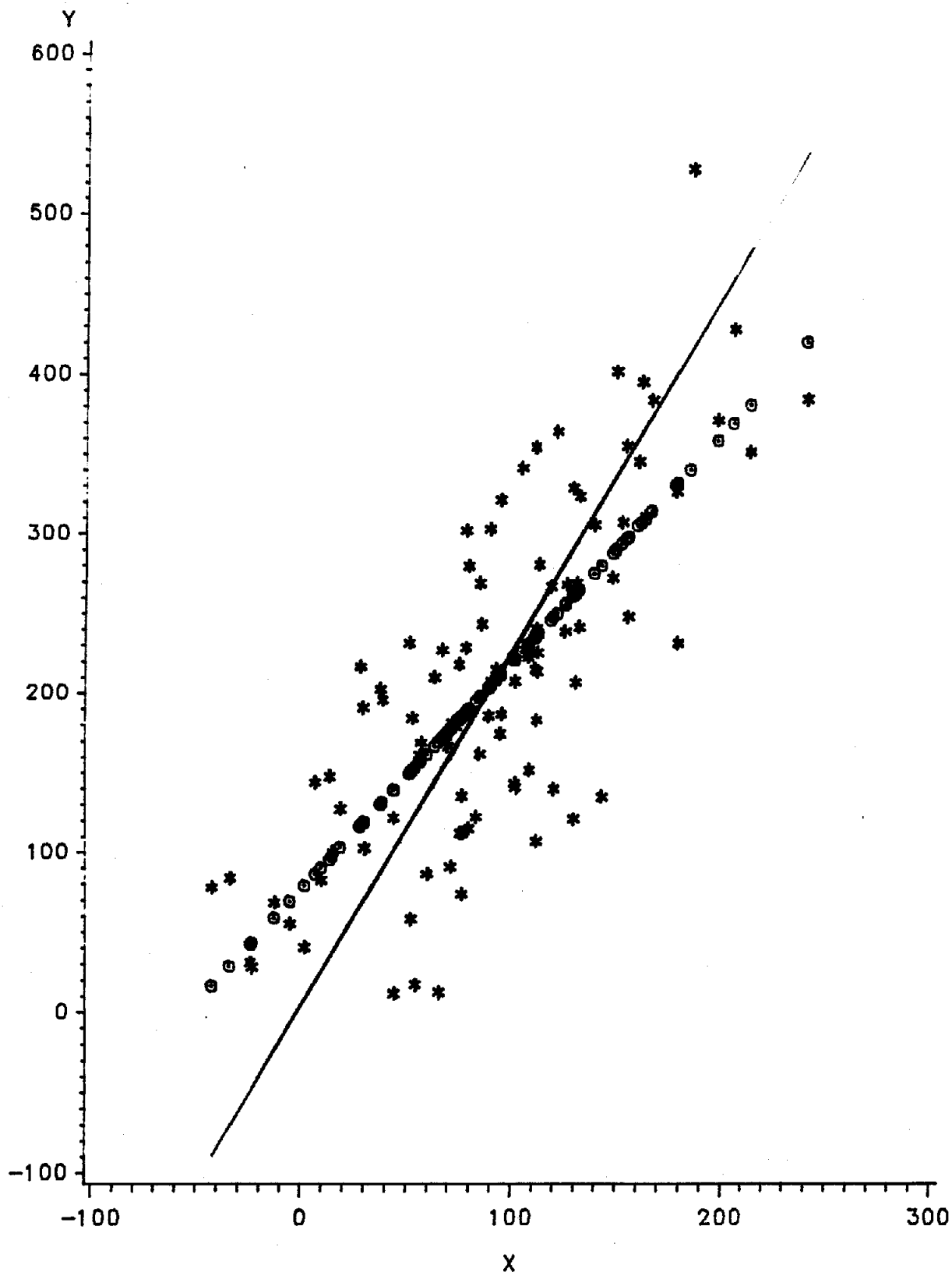
1

OBS	COL1	COL2
1	Structural relation slope =	2.2146030536
2	Intercept =	4.8420583617
3	Least squares regression slope =	1.4192143803
4	Intercept =	77.271713436

-----ZOOM-----I-----

OVERHEAD #14

SCATTER-PLOT OF THE DATA WITH OVERLAY OF LINES-OF-BEST-FIT



Stars are original data points.
Solid line estimated by General Structural Relation.
Dotted line estimated by least squares regression.