DAVID, R. M.Sc.

Confidence intervals for inverse regression

CONFIDENCE INTERVALS FOR INVERSE REGRESSION WITH APPLICATIONS

By

Richard David

Department of Mathematics
McGill University
Montreal

May 1974

© Richard David 1974

CONFIDENCE INTERVALS FOR INVERSE REGRESSION WITH APPLICATIONS TO BLOOD HORMONE ANALYSIS

By

Richard David

Department of Mathematics
McGill University
Montreal

May 1974

CONFIDENCE INTERVALS FOR INVERSE REGRESSION WITH APPLICATIONS TO BLOOD HORMONE ANALYSIS

By

Richard David

A thesis submitted to the Faculty of Graduate Studies and Research, in partial fulfillment of the requirements for the degree of Master of Science.

Department of Mathematics
McGill University
Montreal

May 1974

CONFIDENCE INTERVALS FOR INVERSE REGRESSION WITH APPLICATIONS TO BLOOD HORMONE ANALYSIS

By

Richard David

Abstract

In classical regression analysis, one studies the dependence of a random variable y on a controlled variable x. The classical approach minimizes the sum of squares of differences between y and its expectation, which is assumed dependent on x. Our principal concern in this thesis is the estimation of x given a further value of y. Inverse regression solves this problem by a regression of x on y. In Chapter I we survey the literature on inverse regression which is highlighted by the recent work of R.G. Krutchkoff and make comparisons with the classical approach. The problem of confidence and tolerance intervals for x given y is studied in Chapter II. A new algorithm for such intervals is developed for polynomial regression in Chapter III. These methods are applied in detail to a blood hormone analysis in poultry science. A computer program is included in the Appendix.

Department of Mathematics
McGill University
Montreal

M.Sc.
May 1974

CONFIDENCE INTERVALS FOR INVERSE REGRESSION WITH APPLICATIONS TO BLOOD HORMONE ANALYSIS

By

Richard David

Résumé

Dans les problèmes concernant l'analyse de régression, on est appelé à étudier la dépendance entre une variable aléatorée y et une variable de contrôle x. Si on procède par la méthode classique, on minimise la somme des carrés des différences entre la variable y et son espérance, laquelle dépend de la variable x. Dans cette thèse, nous sommes avant tout préoccupés par l'estimation de x, connaissant la valeur y. En régression inverse, on solutionne ce problème en utilisant une régression de x sur y. Dans la première partie de cette thèse, nous faisons un bref historique de la littérature entourant le problème de la régression inverse; et c'est en considérant les récents travaux de R.G. Krutchkoff sur le problème de la régression inverse que nous allons comparer l'approche classique et l'approche inverse.

Dans le deuxième chapitre, nous allons étudier le problème des intervalles de confiance pour la valeur espérée x étant donné y et pour la valeur unique x étant donné y.

Dans le troisième chapitre, nous allons élaborer un algorithme capable de trouver ces intervalles de confiance lorsque nous avons une régression polynomiale. Nous allons finalement appliquer en détail ces méthodes à un problème concernant les sciences animales, plus précisément le domaine de l'aviculture; ce problème consiste en une analyse du niveau d'hormone des cellules sanguines.

En appendice, nous trouverons un "programme informatique".

Departement of Mathematics
McGill University
Montreal

M.Sc.
May 1974

ACKNOWLEDGEMENTS

I would like to thank Professor G.P.H. Styan of the Department of Mathematics at McGill University for his advice and continuous encouragement in the completion of this thesis. Thanks are also due to Professor Roger Buckland of the Department of Animal Science at Macdonald College for drawing my attention to a problem in poultry science which provided the initial impetus for this thesis; I am also grateful to Professor Buckland for his expert advice and support.

Finally, I want to express my gratitude to Mrs. Louise Bergeron who typed the manuscript.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

CHAPTER I - Survey on Inverse Regression.

INTRODUCTION

The problem of finding a relationship between two variables is one of the common problems in statistics. The relationship between the two variables may be linear; for example, if the speed S is kept constant, the distance D varies with the elapsed time T, since D= ST. The relationship is more or less linear if one wants to verify it empirically; the measurements could easily be subject to errors. If we want to predict D for a particular T, considering S fixed, we would use a straight line.

It could happen that a linear relationship would not be exact. Chatfield (Statistics for Technology; page 186, example 5) looks at the following example: let us consider the temperature and thrust of a rocket engine while it is running under the same operating conditions. By plotting the pair (X,Y) = (temperature, thrust), a diagram like Figure 1.0.1 will result:

X	Y	X	Y	X	Y
19	1.2	33	2.1	45	2.2
15	1.5	30	2.5	39	2.2
35	1.5	57	3.2	25	1.9
52	3.3	49	2.8	40	1.8
35	2.5	26	1.5	40	2.8

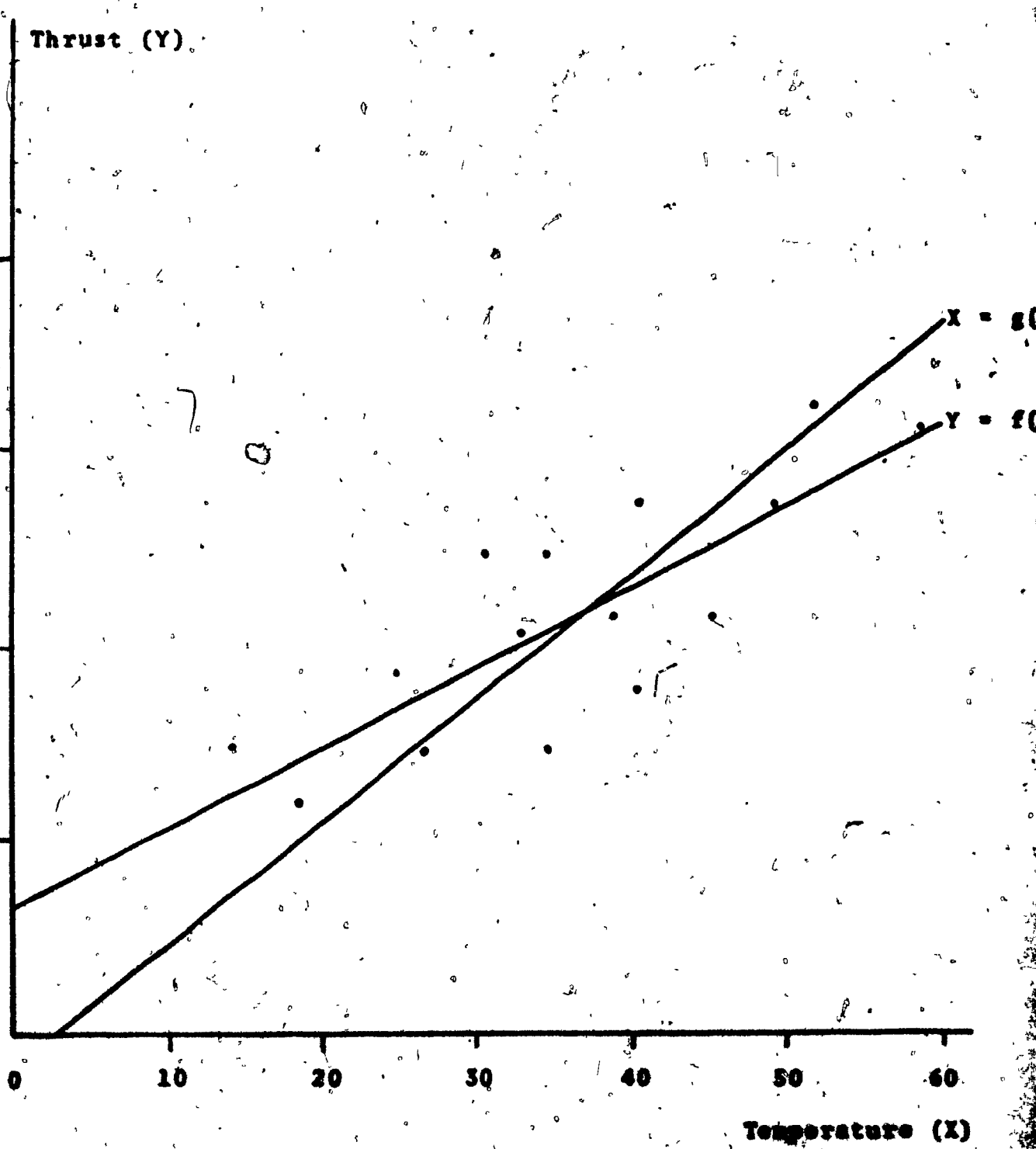


Figure 1.0.1 - Temperature and thrust of a rocket engine (Chatfield, Statistics for Technology, Figure 2A, page 193)

From the diagram, we see that for any given temperature there is a range of observed thrust. Let the variable X be the controlled variable and let us consider the variable Y as random. The average thrust for given observed temperature is called the regression curve of thrust on temperature. We can denote it by $Y = f(X)$. But there exists a regression curve of temperature on thrust; we can denote it by $X = g(Y)$. Suppose the two regression curves are straight lines (they are not always the same). Usually, we record a temperature but not the thrust, and we wish to estimate this thrust. In such problems of regression, the main interest always lay in the prediction of the variable Y (clearly, predicting Y from X is classical). Some people considered observing Y, without knowing X, and hence predicting that X.

The central topic of this thesis is the prediction of X when a value of Y is observed for which the X is not known. An important question is: "Should we use the regression of Y on X or the regression of X on Y in order to make the prediction of the variable X?"

From this follows the topic of confidence intervals. To derive a confidence interval for X given an observed

Y when we are dealing with linear regression is not difficult. But when the regression is quadratic or cubic, difficulties arise. We develop an algorithm which enables us to derive confidence intervals when the regression is polynomial.

The following sections of this chapter comprise a survey on the prediction questions raised above, in a historical perspective. We report on practically all the literature written on the subject, literature which covers more than thirty years.

1.1 C.Eisenhart (1939)

In Eisenhart's mind, the fundamental problem in the determination of a relationship between two variables, say X and Y, was to find, as accurately as possible, the simultaneous probability distribution of the observable quantities, say x and y, considered as random variables.

Eisenhart expresses the relationship between X and Y as a linear one, i.e.,

(1.1.1) $\alpha_0+\alpha_1 X+\alpha_2 Y = 0$,

which can be written as

(1.1.2) $Y = \alpha+\beta X$, when $\alpha = -\alpha_0/\alpha_2$, $\beta = -\alpha_1/\alpha_2$, $\alpha_2 \neq 0$

(1.1.3) $X = \gamma+\delta Y$, when $\gamma = -\alpha_0/\alpha_1$, $\delta = -\alpha_2/\alpha_1$, $\alpha_1 \neq 0$.

If one is interested in estimating Y from X, then one takes $\hat{Y} = a+bX$ as the estimate in (1.1.2); $\hat{Y}$ is the estimate of Y and a, b are, respectively, the least squares estimates of α and β. If one is interested in estimating X from Y, then one takes $\hat{X} = c+dY$ as the estimate in (1.1.3); $\hat{X}$ is the estimate of X and c, d are, respectively, the least squares estimates of γ and δ.

According to Eisenhart "...when the research worker selects the X values in advance, and holds x to these values without error, and then observes the corresponding y values, the errors are in the y values, so that even if he is interested in using observed values of Y to estimate X, he should nevertheless fit $\hat{Y} = a+bX$ and then use the inverse of this relation to estimate X ..."; this is in fact "inverse regression" since we are taking $X = (\hat{Y}-a)/b$.

By the theory of least squares, it is easy to show that $X = (\hat{Y}-a)/b$ is preferred to $\hat{X} = c+dY$. We select the values of X, and the values of Y are found by observation; if we minimize $\Sigma(y-\hat{Y})^2$ and $\Sigma(x-\hat{X})^2$, we obtain the two lines

(1.1.4) $\hat{Y} = a+bX$

(1.1.5) $\hat{X} = c+dY$;

a, b are, respectively, the least squares estimates of α and β, and c, d are, respectively, the least squares estimates of γ and δ.

The two fitted lines usually yield different estimates in (1.1.1), unless there exists a perfect correlation between X and Y. If we consider the analyses of variance for both lines, Eisenhart concludes the following: "...when the values of x have been selected by the research worker and the corresponding y values observed, the line obtained by minimizing $\Sigma(x-\hat{X})^2$ is meaningless and " $\hat{Y} = a+bX$ " is accordingly the only correct estimate of the postulated linear relationship between X and Y, wherefore, if it is desired to reason from Y to X this must be done by means of $X = (\hat{Y}-a)/b$ ..."

1.2 Charles P. Winsor (1946)

The problem of finding the most appropriate regression caused some confusion, despite the paper by Eisenhart (1939).

Let us consider the following case: we have a pair of variables u and v which has a bivariate normal distribution, with variances σ_u^2, σ_v^2 and

correlation ρ_{uv} . We have the measurements of u and v which are subject to error, and we assume that these errors are independent, unbiased for zero and normally distributed.

Suppose we have the following two measurements:

$$x = u+\delta \ , \ y = v+\epsilon \ ,$$

with δ, ϵ independent normal deviates with means zero and variances σ_δ^2, σ_ϵ^2. We can see that:

(1.2.1)

$$V(x) = \sigma_x^2 = \sigma_u^2+\sigma_\delta^2 \ ; \ V(y) = \sigma_y^2 = \sigma_v^2+\sigma_\epsilon^2 \ ; \ \text{corr}(x,y) = \frac{\sigma_u\sigma_v\ \rho_{uv}}{\sigma_x\sigma_y} \ .$$

From this, the regression slopes are:

(1.2.2)

$$\beta_{xy} = \frac{\sigma_u\sigma_v\ \rho_{uv}}{\sigma_x^2} \ ; \ \beta_{xy} = \frac{\sigma_u\sigma_v\ \rho_{uv}}{\sigma_y^2} \ .$$

Suppose there is a sample set of pairs of values of x and y. In Winsor's mind, the choice between two regressions, i.e., that of y on x or that of x on y, is meaningless; two things must be known in order to make the right choice:

(1) how were the (x, y) values obtained?

(2) "what are we going to use the regression for?"

The first question gives the following two situations:

(a) the pairs of (x, y) values are obtained as a random sample;

(b) the (x, y) values are obtained by selecting a set of values of one variable, say x, and subsequently measuring the corresponding values of y.

These two situations are indicated in Figure 1.2.1.

The second question gives the following possibilities:

(i) "We may want a relation from which we can estimate in the future the value of y, given a future measurement x."

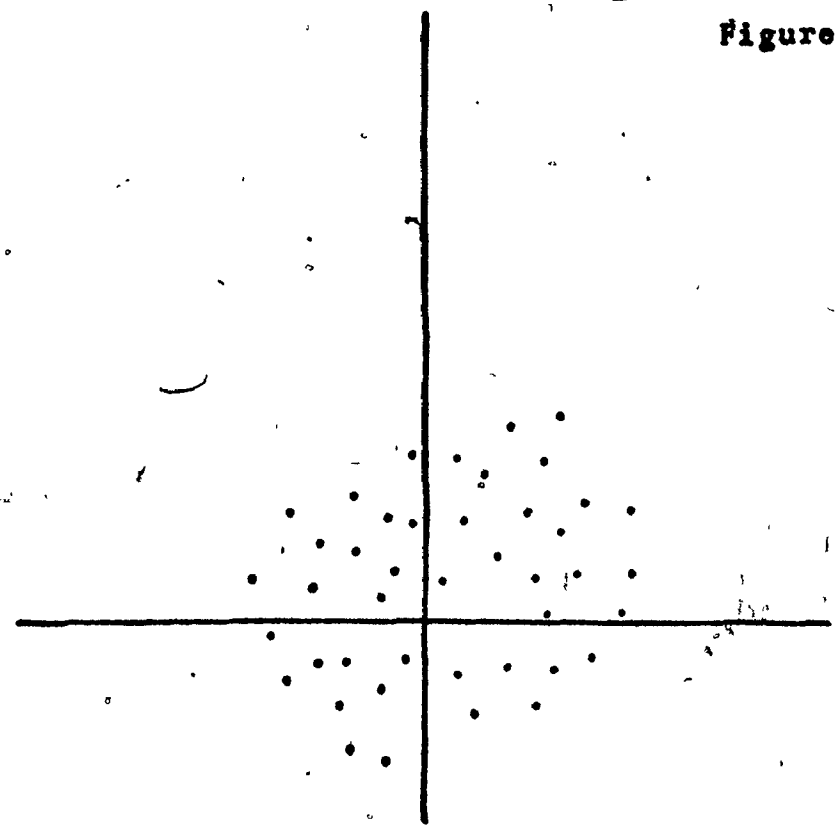
(ii) "We may want a relation from which to estimate x, given a future measurement y."

(iii) "We may want a relation for estimating the true value v, given a future measurement x (or, u given y)."

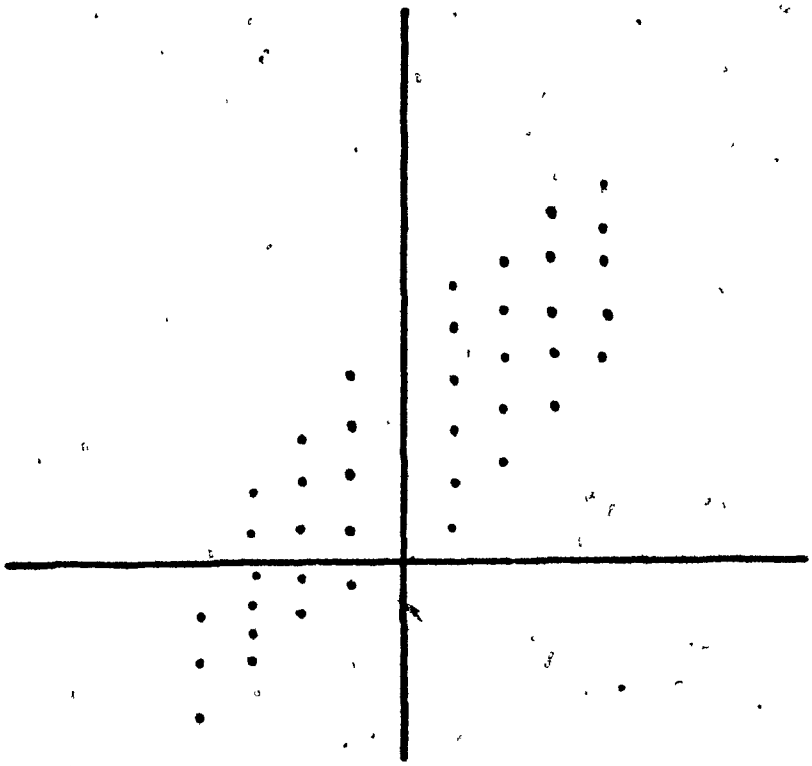
(iv) "We may wish to estimate the true relation between the true values u and v."

In situation (a), because the (x, y) pairs are a random sample, it is easy to estimate all the parameters of the (x,y) distribution. It is clear that for problems (i) and (ii), we need, respectively,

Figure 1.2.1 -



a - (x,y) pairs randomly sampled



b- x values arbitrarily chosen

the regression of y on x and the regression of x on y. In other words, this is the bivariate normal regression problem. In problem (iii), the regression of the true value v on the measurement x is the same as that of y on x, "since y differs from v by random and unbiased errors (y = v+ε)". In problem (iv), we need more information; if we want to estimate the true relationship between u and v, we need estimates of the error variances σ_δ^2, σ_ε^2.

In situation (b), the problem is different. Since we select a set of x values and measure the corresponding y, we can obtain an estimate of the regression of y on x, and of the variance $\sigma^2_{y\cdot x}$ of y around the regression curve. We cannot, however, estimate the population values of $\bar{x}$, $\bar{y}$, σ_x^2, nor can we estimate the population regression of x on y (cf. Winsor (1946), page 103).

From now on, the question is no longer: "Which regression should we use?" but "What can we do with the single regression we have?" Since it is impossible to estimate the population regression of x on y, it is important to establish a relation for predicting x given y. We will approach this problem using confidence intervals.

Let us introduce here the problem we considered earlier in Figure 1.0.1 (problem of temperature and thrust of a rocket engine). Let us assume that we know the regression line $E(y) = .62 - .044x$. Let us assume that we know the variance of y around the regression line: $\sigma^2_{y\cdot x} = 0.389$. Suppose we take a new pair (x, y) from the population; we are informed of its y value but we know nothing of its x value. What can we say about this x value?

We have the regression line:

(1.2.3) $E(y) = .62 + .044x$.

Since we assume normality, we know that the values of y are normally distributed with mean

$$E(y) = .62 + .044x$$

and variance $\sigma^2_{y\cdot x} = .389$.

Therefore, for any given x, the probability that y is in the interval:

(1.2.4) $(.62 + .044x - 1.96\sqrt{.389}\ ,\ .62 + .044x + 1.96\sqrt{.389})$

is .95; and by proper choice of the coefficient of $\sigma_{y\cdot x}$, we can make corresponding statements for any other level of probability. A diagram will be helpful to see what is happening; Figure 1.2.2 gives some explanations of the situation. We have the regression

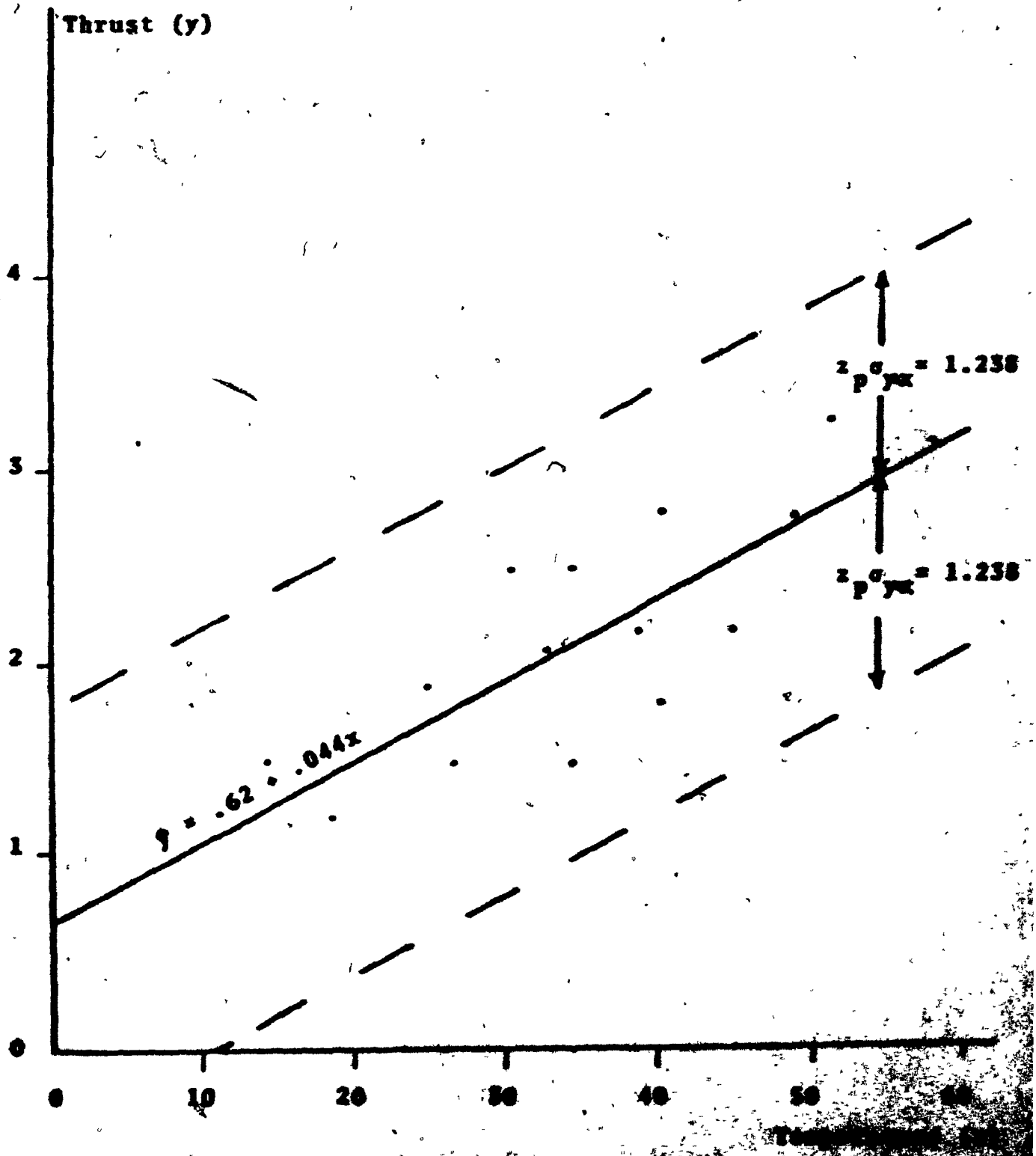


Figure 1.2.2 - Confidence limits for y given x.

line y on x, and above and below it we have lines at a vertical distance of $1.96\sigma_{yx}$. The probability that y falls between the two lines is .95 for every x value; therefore, the probability will be the same for all possible y values. If we take all possible pairs of (x, y) values, 95% of them should be inside the two lines.

Algebraically, we can express all this by the following inequality:

(1.2.5) $a+bx-z_p\sigma_{yx} < y < a+bx+z_p\sigma_{yx}$,

where σ_{yx} is the standard deviation of y around the regression curve and z_p is the critical value of N(0,1). If we apply inequality (1.2.5) to our example, we can set up the inequality:

$$.62 + .044x - 1.96\sigma_{yx} < y < .62 + .044x + 1.96\sigma_{yx} .$$

With z_p properly chosen, we assert that this inequality is satisfied with probability p = .95.

If we rearrange inequality (1.2.5), we obtain:

(1.2.6) $\frac{1}{b}(y-a-z_p\sigma_{yx}) < x < \frac{1}{b}(y-a\ z_p\sigma_{yx})$, with $b \neq 0$.

If we apply inequality (1.2.6) to our example, we obtain:

$$\frac{1}{.044}(y-.62-1.96\sigma_{yx}) < x < \frac{1}{.044}(y-.62+1.96\sigma_{yx}) .$$

Inequality (1.2.5) is a statement about the random variable y, which involves a fixed value x; it has probability p of being true for any arbitrary x and also for the aggregate of all values of x. In inequality (1.2.6), if the random varable is y, then statements (1.2.5) and (1.2.6) are equivalent in meaning. Since inequality (1.2.5) holds with propability p, this is true for all possible pairs (x, y). However, we cannot say that the above assertion holds for each particular y value; whenever y falls outside the limits $a-bx \pm z_p\sigma_{yx}$, the inequality (1.2.6) is false.

Let us consider the situation for the general case of the bivariate surface. Figure 1.2.3 gives the regression of y on x and that of x on y, and both sets of confidence limits.

We see that the two sets of confidence limits do not coincide and that the direct limits are narrower than the inverted limits, i.e. $t_p\sigma_{xy} < \frac{t_p\sigma_{yx}}{b_{yx}}$,

where t_p is the critical value of Student's t_{n-2}. We can easely see that the direct regression gives more advantages. Sometimes, the inversion of the

Inequality (1.2.5) is a statement about the random variable y, which involves a fixed value x; it has probability p of being true for any arbitrary x and also for the aggregate of all values of x. In inequality (1.2.6), if the random varable is y, then statements (1.2.5) and (1.2.6) are equivalent in meaning. Since inequality (1.2.5) holds with propability p, this is true for all possible pairs (x, y). However, we cannot say that the above assertion holds for each particular y value; whenever y falls outside the limits $a-bx \pm z_p\sigma_{yx}$, the inequality (1.2.6) is false.

Let us consider the situation for the general case of the bivariate surface. Figure 1.2.3 gives the regression of y on x and that of x on y, and both sets of confidence limits.

We see that the two sets of confidence limits do not coincide and that the direct limits are narrower than the inverted limits, i.e. $t_p\sigma_{xy} < \frac{t_p\sigma_{yx}}{b_{yx}}$, where t_p is the critical value of Student's t_{n-2}. We can easely see that the direct regression gives more advantages. Sometimes, the inversion of the

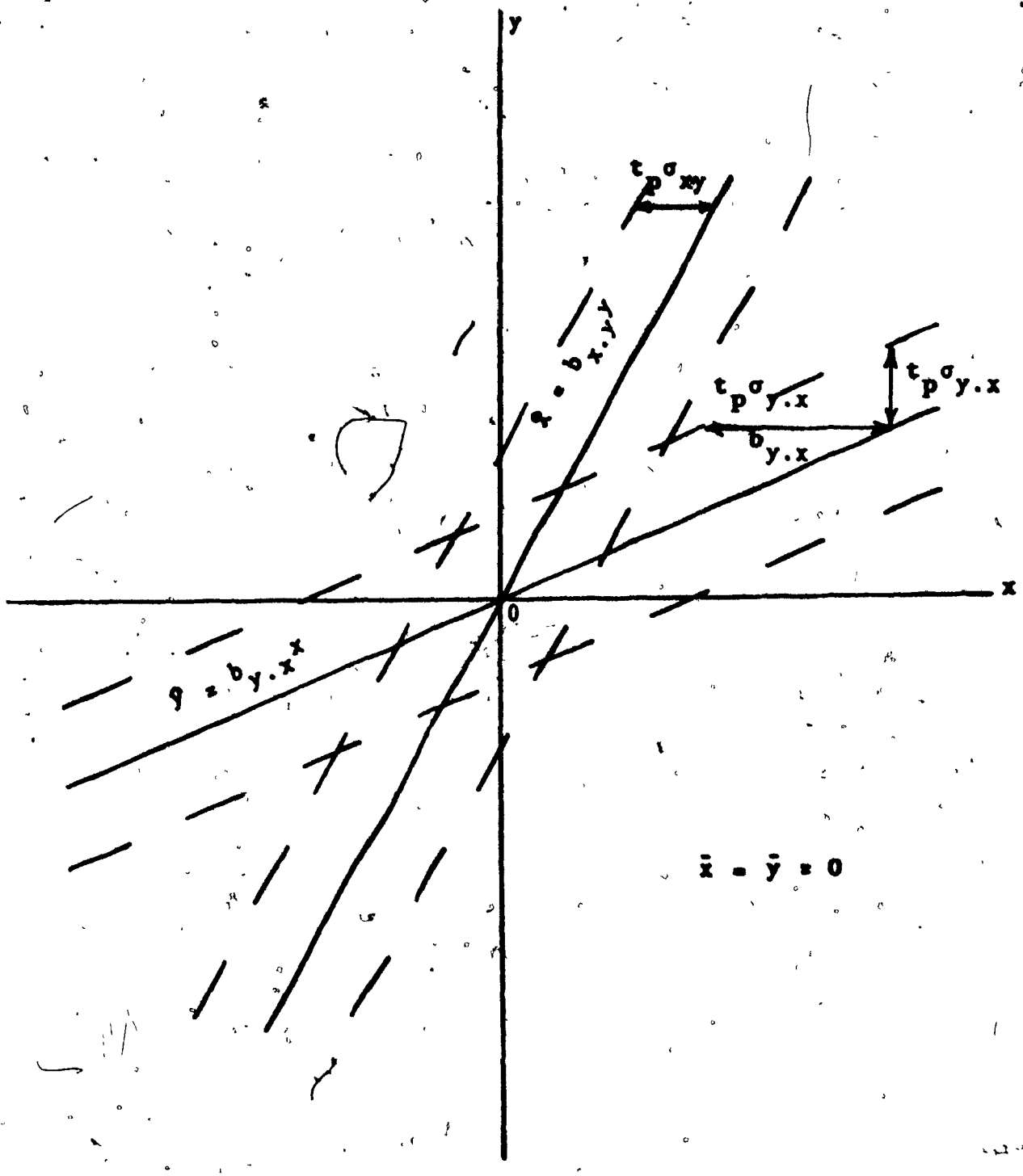


Figure 1.2.3 - Regression lines and confidence limits, bivariate normal, (x, y) pairs randomly sampled (Winsor (1946), page 108, Figure 5).

regression line and the confidence limits are the only solution to the regression and the estimation problem; for example, in biological assay, inversion of the regression line is used. We attempt to estimate the potency of an unknown drug in terms of a biological response, with the aid of a response curve based on known doses of a standard drug.

According to Winsor the general principle is: one should arrange the experiment so that the desired regression is determined directly, that is the variable from which prediction is to be made, should be taken as the independent variable. When this is not possible, one uses the inverted regression.

1.3 E.J. Williams (1959)

Suppose we have the following relationship between X and Y:

(1.3.1) $\alpha_0+\alpha_1 X+\alpha_2 Y = 0$.

We can write the linear relationship of equation (1.3.1) as

(1.3.2) $Y = \alpha+\beta X+\epsilon_1$, $\alpha = -\alpha_0/\alpha_2$, $\beta = -\alpha_1/\alpha_2$, $\alpha_2 \neq 0$,

ϵ_1= error term,

(1.3.3) $X = \gamma + \delta Y + \epsilon_2$, $\gamma = -\alpha_0/\alpha_1$, $\delta = -\alpha_2/\alpha_1$, $\alpha_1 \neq 0$,

ϵ_2 = error term.

A point which is often raised when regression equations are employed is: which variable in equation (1.3.1) is to be treated as the dependent variable? The dependent variable must be subject to random error in order that the usual theory of least squares may be applicable. If Y is the dependent variable, we use equation (1.3.2); otherwise, if X is the dependent variable, we use equation (1.3.3).

It is important to avoid choosing as dependent variable, one which is errorless, or which has been subject to selection. If both variables are subject to error, in simple regression, either regression equation could be used, depending on which variable is to be predicted.

Problems, in which it is not possible to do this, especially in calibration experiments, give rise to the device of inverse estimation from the regression equation. Suppose we are calibrating a pressure gauge; if x represents the controlled variable (the pressure) and y represents the measured variable (the gauge marking), then x and y are related by:

(1.3.4) $y = \alpha+\beta x+\varepsilon_1$, ε_1 = reading error.

The least squares line gives

(1.3.5) $y = a+bx$,

where a, b are the least squares estimates of α and β respectively. We suppose that the errors are independently and identically distributed with zero means. Therefore, if one wants to estimate an unknown pressure when we have a gauge reading, one should use the inverse estimate

(1.3.6) $x = (y-a)/b$, $b \neq 0$.

But if one uses the following equation

(1.3.7) $x = \gamma+\delta y+\varepsilon_2$

as the regression line, one will get the following least squares line

(1.3.8) $x = c+dy$,

where c, d are, respectively, the least squares estimates of γ and δ (here also we suppose that the errors are independently and identically distributed with zero means).

According to Williams, one should use equation (1.3.6) to estimate an unknown pressure; however, when equation (1.3.6) gives results of spurious accuracy, one should use equation (1.3.8).

1.4 R.G. Krutchkoff (1967)

In the preceding section, we introduced the problem of calibrating an instrument, say a pressure gauge. Williams (1959) considered only one method of calibration, that is, the classical method using $y = \alpha+\beta x+\varepsilon$, where x represents the controlled variable (the pressure) and y represents the measured variable (the gauge marking) and ε represents the reading error.

Related to the same problem, Krutchkoff considers two different methods of calibration.

Let x and y be related by

(1.4.1) $y = \alpha+\beta x+\varepsilon$, ε = reading error.

Method I: Classical Approach.

If there are N distinct observations, we can write model (1.4.1) as

$$y_i = \alpha+\beta x_i+\varepsilon_i,\ i = 1,2,\ldots,N .$$

The errors are independently and identically distributed with zero means. By minimizing the squared residuals of y, we obtain the predicting line

(1.4.2) $y = a+bx$,

where a, b are, respectively, the least square estimators of α and β. The corresponding calibration

equation becomes

$$x = (Y-a)/b \ , \ b \neq 0.$$

From this, the estimate for x is given by

(1.4.3) $\hat{X} = (Y-a)/b, \ b \neq 0.$

Method II: Inverse Approach.

We can rewrite model (1.4.1) as:

(1.4.4) $x = \gamma+\delta y+\varepsilon' \ , \ \gamma = -\alpha/\beta \ , \ \delta = 1/\beta \ , \ \varepsilon' = -\varepsilon/\beta \ ,$

$\beta \neq 0 \ .$

If we consider the same N observations as in the first method, we can write the model as:

(1.4.5) $x_i = \gamma+\delta y_i+\varepsilon_i' \ , \ i = 1,2,\ldots,N \ .$

By minimizing the squared residuals of x, we obtain the predicting line

(1.4.6) $x = c+dy \ ,$

where c, d are, respectively, the least squares estimates of γ and δ. If one reads Y on the gauge the estimate of the unknown pressure is

(1.4.7) $\hat{X} = c+dY \ .$

Using the criterion of mean squared error, Krutchkoff compares Method I and Method II, using Monte Carlo methods, and shows that Method II is better.

Using the model $y = \alpha+\beta x+\varepsilon$, with $\alpha = 0$ and $\beta = .5$, Krutchkoff makes the assumptions:

(i) the range of x is (0, 1)

(ii) the standard deviation of error is 10% of the range (σ = .1).

Taking three observations at each of the end points (x = 0, x = 1), Krutchkoff first finds

$$y = \alpha+\beta x+\epsilon ,$$

where ϵ is generated each time as a normal random number (the error ϵ is assumed to be normally distributed with zero mean and standard deviation = .1); then he finds the least squares estimates a, b, c and d.

Then for x = 0,.2,.4,.6,.8, 1, 1.2, 2, 5 and 10, Krutchkoff finds

$$y = \alpha+\beta x+\epsilon ,$$

where ϵ is generated each time as a normal random number.

Taking (1.4.3) and (1.4.7), he finds squared errors for the same x values; from (1.4.3), he has

$$\hat{X}_{c1} = (Y-a)/b , \quad b \neq 0 ,$$

and he obtains the squared errors by $\Sigma(\hat{X}_{c1}-x)^2$ for those x values.

From (1.4.7), he finds

$$\hat{X}_{in} = c+dY$$

and he obtains the squared errors by $\Sigma(\hat{X}_{in}-x)^2$.

He performs the same experiment 10,000 times. The mean squared errors are computed for the classical approach and the inverse approach; for each value of x, he computes

$$\Sigma(\hat{X}_{cl}-x)^2/10{,}000 \quad \text{and} \quad \Sigma(\hat{X}_{in}-x)^2/10{,}000.$$

The mean squared error of the classical approach is uniformly larger than the mean squared error of the inverse method (even for x values outside the range).

We performed the same experiment 10 times because of limited time on the computer. The results obtained empirically corroborate Krutchkoff's results. The following table gives the results we obtained.

TABLE 1.4.1 - Comparison of mean squared errors

	X=0	X=.2	X=.4	X=.6	X=.8
$Av(\hat{x}-x)^2_{cl}$	.05286	.05567	.05044	.05710	.04800
$Av(\hat{x}-x)^2_{in}$	.04470	.04366	.03860	.04461	.03837
Ratio(cl/in)	1.18249	1.27505	1.30680	1.27977	1.25194
	X=1	X=1.2	X=2	X=5	X=10
$Av(\hat{x}-x)^2_{cl}$	.05369	.09153	.11984	.69923	2.85713
$Av(\hat{x}-x)^2_{in}$	.04241	.05088	.09718	.54729	2.25144
Ratio(cl/in)	1.26612	1.79899	1.23321	1.27765	1.26902

Looking at the cases where the parameters and the design are different, Krutchkoff concludes that the effects due to: the intercept α, the slope β, the number of observations at each design point and the design, will not alter the conclusion that the mean squared error of Method II (the inverse approach) is uniformly smaller than the mean squared error of Method I (classical approach).

Suppose now we consider the situation where the errors are non-normal; the errors are obtained by generating random numbers from a Pearson distribution characterized by the measures of skewness:

$$\beta_1 = \frac{(\text{third central moment})^2}{(\text{variance})^3}$$

and kurtosis:

$$\beta_2 = \frac{(\text{fourth central moment})}{(\text{variance})^2} .$$

The effects of skewness and kurtosis do not change the conclusion already obtained.

If we finally look at the effect of a quadratic, i.e.,

$$y = \alpha+\beta x+\theta x^2+\epsilon ,$$

we get a still smaller mean squared error.

The overall conclusion is that the inverse

approach to the calibration problem has a uniformly smaller mean squared error than the classical approach.

1.5 Critiques of Krutchkoff's method of calibration.

1.5.1 E.J. Williams (1969a)

Suppose we have the following regression model:

$$E(Y|x) = \alpha+\beta x ,$$

where we assume that Y is normally distributed about $\alpha+\beta x$ with variance σ^2.

The least-squares estimated regression based on a sample of N observations is

$$\hat{y} = a+bx ,$$

where $b = \sum_{i=1}^{N} y_i (x_i-\bar{x}) / \sum_{i=1}^{N} (x_i-\bar{x})^2$, and $a = \bar{y}-b\bar{x}$ are, respectively, the least squares estimators of β and α. In Krutchkoff's paper (1967), the classical estimator of the unknown variable x, given an observation Y, is

$$\hat{x} = \bar{x} + (Y-\bar{y})/b , \quad b \neq 0 .$$

If the errors are normally distributed, b is normally distributed independently of Y and $\bar{y}$, and its reciprocal has infinite variance. The classical estimator $\hat{x}$ will have undefined expectation and infinite variance and also infinite mean squared deviation (m.s.d.).

The inverse estimator, as defined by Krutchkoff, is

$$\tilde{x} = \bar{x} + (Y-\bar{y})c \ ,$$

where $c = \sum_{i=1}^{N} y_i(x_i-\bar{x}) / \sum_{i=1}^{N}(y_i-\bar{y})^2$.

As long as $N > 4$, it may be shown that c and $\tilde{x}$ have finite m.s.d..

According to Williams, there is no need for computer simulation experiments to conclude that from the m.s.d. point of view, the inverse estimator is better than the conventional estimator. The fact that the m.s.d. of an estimator is less than infinity is not proving very much!

We know (Blackwell, 1947) that the minimum variance unbiased (m.v.u) estimator may be obtained if there exists an unbiased estimator of a parametric function and if there also exist sufficient statistics for the parameters in the distribution. This estimator is in fact the expected value of the unbiased estimator, conditional on the sufficient statistics. Because of the sufficiency proprety, the m.v.u. statistic is distributed independently of the parameters.

Let

(1.5.1.1) $\quad E(Y|X) = \hat{y} = \alpha+\beta x$

where x is a parametric function.

From (1.5.1.1), we have

$$x = (\hat{y}-\alpha)/\beta, \quad \beta \neq 0 .$$

For the parameters $\hat{y}$, α and β, $\bar{y}$, b and Y are collectively the sufficient statistics. ($\bar{y}$, b and Y are normally distributed and mutually independent.)

Since

$$E(Y) = \alpha+\beta x$$

and

$$E(\bar{y}) = \alpha+\beta\bar{x}$$

we have

$$t = \bar{x} + \frac{Y-\bar{y}}{\beta} , \quad \beta \neq 0$$

which is unbiased for x

$$E(t) = E(\bar{x}) + \frac{E(Y) - E(\bar{y})}{\beta} = \bar{x} + \frac{\alpha+\beta x-\alpha-\beta\bar{x}}{\beta} = x.$$

The statistic

(1.5.1.2) $\quad \bar{x} + (Y-\bar{y})g(b)$

is an unbiased estimator of x, based on a sufficient statistic: g(b) is the unique unbiased estimator of β^{-1}

$$g(b) = \frac{e^{b^2/\phi^2}}{\phi^2\sqrt{\pi}} \int_b^\infty e^{\frac{-t^2}{2\phi^2}} dt \quad (\beta>0), \quad \phi^2 = V(b) = \frac{\sigma^2}{\Sigma(x-\bar{x})^2}$$

$$g(b) = -\frac{e^{-b^2/2\psi^2}}{\psi^2} \int_{-\alpha}^{b} e^{-\frac{t^2}{2\psi^2}} dt \ , \ (\beta<0) \ .$$

The estimator defined in (1.5.1.2) has infinite variance, since g(b) has infinite variance. Therefore, we can say that no unbiased estimator will have finite variance.

We can apply the same statement to the m.s.d. criterion. We know that the m.s.d. of any statistic consists of two parts, the variance and the square of the bias. If we replace the statistics by their expectations conditional on sufficient statistics, we will reduce the variance and we will leave the bias unaltered; we therefore reduce the m.s.d..

According to Williams, the inverse estimator considered in Krutchkoff's paper (1967) was derived on the false assumptions that the errors were independent of the values of y; this estimator has finite variance. Since the classical and the unbiased estimators have infinite variance, the fact that Krutchkoff's estimator has finite variance is of little account. Any estimator which is a constant, will have finite variance; therefore from the m.s.d. point of view the inverse estimator will be preferable to the

classical estimator.

1.5.2 Joseph Berkson (1969)

For the classic least squares procedure of estimation, we minimize the sum of the residuals to obtain as estimators of β, α $(y = \alpha+\beta x+\varepsilon)$.

$$\hat{\beta} = \frac{\sum_{i=1}^{N}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{N}(x_i-\bar{x})^2},$$

$$\hat{\alpha} = \bar{y}-\hat{\beta}\bar{x}.$$

For the inverse least squares procedure of estimation (the one proposed by Krutchkoff, 1967), we minimize the x residuals to obtain as estimates of α', β' $(x = \alpha'+\beta' y+\varepsilon', \beta' = 1/\beta, \alpha' = \alpha/\beta, \beta \neq 0)$

$$\tilde{\beta} = \frac{\sum_{i=1}^{N}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{N}(y_i-\bar{y})^2},$$

$$\tilde{\alpha} = \bar{x}-\tilde{\beta}\bar{y}.$$

Furthermore, we get

$$\tilde{\beta} = 1/\hat{\beta}', \quad \tilde{\alpha} = \hat{\alpha}'/\hat{\beta}', \quad \hat{\beta}' \neq 0.$$

If we use the classical estimates, the calibration line for an estimate of an unknown value X is given by

(1.5.2.1) $\hat{X}_i = (Y_i - \hat{\alpha})/\hat{\beta}$, $\hat{\beta} \neq 0$.

If we use the inverse estimates, we will obtain

(1.5.2.2) $X_i = \tilde{\alpha}' - \tilde{\beta}' Y_i = (Y_i - \tilde{\alpha})/\tilde{\beta}$, $\tilde{\beta} \neq 0$.

Krutchkoff (1967) used only small values of N for all the experiments performed (N = 6 for most experiments, the largest being N = 20). However, it is important to use large values of N for estimating a calibration line.

For the classic unbiased estimates, we remark that

$$\hat{\beta} \xrightarrow{P} \beta \quad \text{and} \quad \hat{\alpha} \xrightarrow{P} \alpha \,, \quad N \to \infty \,.$$

The estimates $\hat{\alpha}$ and $\hat{\beta}$ are, therefore, consistent in probability. In other words, $\hat{\alpha}$ and $\hat{\beta}$ are consistent estimates if $V(\hat{\alpha}) \to 0$ and $V(\hat{\beta}) \to 0$ as $N \to \infty$.

For the inverse estimates, we note that

$$\tilde{\beta} \to \beta \frac{\sigma^2}{\beta \sigma_x^2} \quad \text{and} \quad \tilde{\alpha} \to \alpha - \frac{\sigma^2 \bar{x}}{\beta \sigma_x^2} \,, \quad \text{as } N \to \infty \,,$$

where σ is the standard error of the error of observation of y, and $\sigma_x^2 = \sum_{i=1}^{N} (x_i - \bar{x})^2 / N$ is the variance of the x used in the determination of the calibration line.

Because

$$\tilde{\beta} \to \beta \frac{\sigma^2}{\beta\sigma^2} \quad \text{and} \quad \tilde{\alpha} \to \alpha \frac{\sigma^2 \bar{x}}{\beta\sigma_x^2} \quad \text{as } N \to \infty$$

these inverse estimates are not consistent in probability.

Now let us estimate the values of X from readings of Y, with $N \to \infty$. If we look at the classic case, as $N \to \infty$, the expectation of $\hat{X}_i$ given by (1.5.2.1) tends in probability to X_i, that is, the estimate is asymptotically unbiased. The asymptotic variance is given by

(1.5.2.3) $$V(\hat{X}_i) = \left(\frac{\sigma}{\beta}\right)^2 , \quad \beta \neq 0 ,$$

since the asymptotic mean square error is given by

(1.5.2.4) $$\text{m.s.e.}\ (\hat{X}_i) \to \text{bias}^2\ (\hat{X}_i) = \left(\frac{\sigma}{\beta}\right)^2 ,$$

and the estimate is asymptotically unbiased since $\text{bias}^2\ (\hat{X}_i) \to 0$.

In the inverse case, as $N \to \infty$, the expectation of $\tilde{X}_i$ is given by

(1.5.2.5) $$\text{Asymptotic } E(\tilde{X}_i) = \frac{f}{\beta - f}\bar{X} + \frac{\beta}{\beta - f}X_i , \quad f = \frac{\sigma^2}{\beta\sigma_x^2} .$$

The asymptotic bias is given by

(1.5.2.6) $$\text{Asymptotic bias } (\tilde{X}_i) = \frac{\sigma^2}{\beta^2\sigma_x^2 - \sigma^2}(\bar{X} - X_i) = \frac{f}{\beta - f}(\bar{X} - X_i).$$

The asymptotic variance is given by

$$(1.5.2.7) \qquad V(\tilde{X}_i) = \frac{\sigma^2}{(\beta-f)^2} \quad .$$

and the asymptotic m.s.e. is given by

$$(1.5.2.8) \qquad \text{m.s.e.}\ (\tilde{X}_i) = \frac{\sigma^2 f^2(\bar{X}-X_i)^2}{(\beta-f)^2} \quad .$$

For both estimates, we notice that the m.s.e. is independent of α; we also notice that if β is large, $\tilde{\alpha}$, $\tilde{\beta}$ will approach $\hat{\alpha}$, $\hat{\beta}$ and $\tilde{X}_i$ will approach $\hat{X}_i$; we notice finally that the variance of $\tilde{X}_i$ is smaller than that of $\hat{X}_i$.

For $N \to \infty$, the direct estimate, $\hat{X}_i = (Y_i-\hat{\alpha})/\hat{\beta}$, is unbiased normal, and the inverse estimate, $\tilde{X}_i = (Y_i-\tilde{\alpha})/\tilde{\beta}$ is normal but biased, except when $X_i = \bar{X}$.

In Table 1.5.2.1, we compare the m.s.e.'s corresponding to the classic and inverse estimates of the calfbration line, for $N \to \infty$ and $N = 6$ for the experiment with the controlled positions of X as for Krutchkoff's Table 1 (1967).

TABLE 1.5.2.1 - Comparison[ω] of m.s.e's of estimates of X_i, for $N \to \infty$ and for $N = 6$.

	$N \to \infty$		Ratio m.s.e.	
X_i	m.s.e. $(\hat{X}_i)$	m.s.e. (X_i)	$N \to \infty$	$N=6$
0	.04	.034483	1.16	1.28
.2	.04	.031439	1.27	1.26
.4	.04	.029917	1.34	1.25
.6	.04	.029917	1.34	1.25
.8	.04	.031439	1.27	1.25
1.0	.04	.034483	1.16	1.25
1.2	.04	.039049	1.02	1.28
2.0	.04	.072533	.55	1.27
5.0	.04	.414983	.096	1.30
10.0	.04	1.746728	.023	1.28

For $X < 1.2$, we see that the ratio of m.s.e. is greater than unity, that is the m.s.e. of the classic estimate is larger, for $N \to \infty$ as well as for $N = 6$.

For $X > 1.2$, the m.s.e. of the classic estimate is smaller when $N \to \infty$.

If we use larger numbers for estimating the calibration line, the results would be essentially the same. In Table 1.5.2.2, we see that for values of $X < 1.2$, the m.s.e. is smaller for the inverse method, but for

---

(1) Berkson (1969), page 653.

values of X > 1.2, the m.s.e. is smaller for the classic method, which is essentially the same as for N → ∞.

TABLE 1.5.2.2 - Results[2] of Krutchkoff for N = 20, design: 10(X=0), 10(X=1).

m.s.e.			
X	Classic	Inverse	Ratio
0	.046	.037	1.22
.2	.043	.034	1.27
.4	.043	.033	1.32
.6	.044	.034	1.32
.8	.043	.034	1.27
1.0	.046	.038	1.20
1.2	.049	.043	1.14
2.0	.064	.077	.83
5.0	.063	.077	.81
10.0	.063	.077	.83

The fact that the m.s.e. is not smaller with the inverse estimate for all values of X, could mean that Krutchkoff's inverse method is not superior to the classic method.

In Krutchkoff's experiments, he considered an estimate of X_i from a single observation of Y_i;

(2) Berkson (1969), page 654.

when N is sufficiently large, the calibration line can be considered without appreciable error and $\hat{X}_i$ will be unbiased. Now let us consider an estimate of X_i from the mean of $n > 1$ observations of Y_i. As defined in (1.5.2.4), the asymptotic m.s.e. of $\hat{X}$ is given by

$$\text{m.s.e. } (\hat{X}) = \frac{\sigma^2}{n\beta^2} ,$$

and as defined in (1.5.2.8), the asymptotic m.s.e. of $\bar{\hat{X}}$ is given by

$$\text{m.s.e. } (\bar{\hat{X}}) = \frac{\sigma^2}{n(\beta-f)^2} + \frac{f^2}{(\beta-f)^2} (\bar{X}-X_i)^2 .$$

The first term of the r.h.s. of m.s.e.$(\bar{\hat{X}})$ will be less than m.s.e.$(\hat{X})$, but with increasing n, m.s.e.$(\hat{X})$ decreases more rapidly than m.s.e.$(\bar{\hat{X}})$, except at $X_i = \bar{X}$. Therefore, for some n_0, such that $n \geq n_0$, the m.s.e.$(\hat{X}_i)$ will be smaller than that of $\bar{\hat{X}}_i$ except in the neighbourhood of $X_i = \bar{X}$.

An overall conclusion is that the inverse procedure yields inconsistent estimates, that the m.s.e. of the inverse estimate in estimating X from a single observation of Y is smaller for a limited range of X, and finally that for N sufficiently large, if X is

estimated from the mean of $n > 1$ observations of Y, there is always an n_0 such that for values of $n \geq n_0$, the m.s.e. of the classic estimate will be smaller than that of the inverse estimate, except for $\hat{X}$.

Therefore, we cannot establish the superiority of the inverse method according to the principles stated by Krutchkoff.

1.5.3 Max Halperin (1970)

Krutchkoff (1967) has suggested that instead of using the estimated regression of y on x to estimate X, one should use the estimated regression of x on y to estimate X as by $\hat{X}_k = c+dY$, where c and d are, respectively, the least squares estimates of γ and δ ($x = \gamma+\delta y$).

For large samples, Krutchkoff's estimate is superior in the sense of closeness, that is the relative closeness of $\hat{X}$ and $\hat{X}_k$ to X ($\hat{X} = (Y-a)/b$ is the classic estimate). By definition, $\hat{X}$ is a superior estimate than is $\hat{X}_k$ if, for all X,

$$\Pr \{|\hat{X}-X| < |\hat{X}_k-X|\} > \tfrac{1}{2} ,$$

and if values of the independent variable are

restricted to a certain closed interval around the mean of the independent variates in the experiment and less than $\frac{1}{2}$ elsewhere. The width of the interval will vary inversely as the product of the absolute value of the standardized slope and the standard deviation of the independent variables used in the experiment. Since the parameter tends to be large, the interval where Krutchkoff's estimate is superior will be small.

Since the classical procedure allows an exact confidence interval while Krutchkoff's procedure does not, the classical estimate is to be preferred, using the above closeness criterion.

Now, according to the mean squared error criterion applied to the relevant asymptotic distribution, the classical procedure is to be preferred, except that the interval of superiority of Krutchkoff's estimate is no longer small even at best.

Finally, Krutchkoff's estimate is not consistent, while the classical estimate is; however, Krutchkoff's estimate can be modified so that it is consistent but will never be better in the sense of closeness.

Krutchkoff (1967) has demonstrated, by Monte

Carlo simulation, that the inverse method had a uniformly smaller mean squared error than did the classical method for all values of the unknown pressure within the calibration range.

In a subsequent paper, Krutchkoff (1969) has demonstrated that "if a sufficient number of observations were taken in the calibration phase then the classical method has a lower mean squared error outside the calibration range"; but inside the calibration range, the inverse method still has uniformly smaller mean squared error.

According to the concepts of sufficiency and consistency considered in articles written by Williams (1969a) and Berkson (1969), the classical procedure of estimation is superior to the inverse procedure of estimation. However, according to a recent paper written by Krutchkoff (1971), the results presented by Berkson and Williams do not alter the results obtained by Krutchkoff in earlier papers.

In response to Halperin (1969), who introduced the closeness criterion to compare the classical procedure and the inverse procedure, Krutchkoff uses the same procedure as in the earlier paper (1967). In the Halperin paper,

$$\Pr\{|\hat{X}_{cl}-X| > |\hat{X}_{in}-X|\}$$

was approximated by considering relevant asymptotic distributions. This probability was approximated for values of $X = \bar{x} \pm 2.50\sigma_x$ with

$$\sigma_x = \sqrt{\sum_{i=1}^{N}(x_i-\bar{x})^2/N} .$$

Krutchkoff arrives at the opposite conclusion to that of Halperin; the inverse method is, for the most part, superior or equivalent to the classical method; where the classical method is superior, it is only mildly superior.

1.6 Conclusions

Of the papers we have considered, all authors except Krutchkoff use the regression of y on x to make prediction on the unknown x; in fact there exists only one regression line.

If, like Krutchkoff, one is calibrating an instrument, say a pressure gauge, then the inverse procedure of estimation is the best method to use if one wants to make a better prediction on the unknown variable x (because of the mean squared error criterion which says that the inverse method has a uniformly smaller mean squared error than the classical method).

Other statisticians think that Krutchkoff's method is good only when we are working on a small range, that is the range covered by the data.

Shall we believe the majority of statisticians when they say that there is only one regression or shall we consider the method proposed by Krutchkoff as a valid method?

This will be answered in the second chapter when we apply both regressions to the problem of finding confidence intervals for the predicted x variable.

In chapter three, we present an algorithm used

in the prediction of the unknown x variable; this method, like the two others (classical and inverse), will be applied to some specific data. In fact, as was mentioned earlier in the introduction, we are interested in finding confidence intervals for the unknown variable x.

CHAPTER II - Classical and Inverse Methods for Confidence and Tolerance Intervals.

2.1 Confidence and tolerance intervals.

In linear regression, we have the model

(2.1.1) $y_i = \alpha+\beta x_i+u_i$, $i = 1, 2, \ldots, n$,

where α, β are unknown parameters and u_i is an error term.

We make the assumption that, in model (2.1.1),

i) u_i is a random variable with zero mean and variance σ^2;

ii) u_i and u_j are uncorrelated, $i \neq j$.

Therefore

$$E(y_i) = \alpha+\beta x_i \quad , \quad i = 1, 2, \ldots, n$$

and y_i and y_j, $i \neq j$, are uncorrelated.

A further assumption is that

iii) u_i is a normally distributed random variable with zero mean and variance σ^2, that is $u_i \sim N(0,\sigma^2)$.

From this last assumption, u_i and u_j are then necessarily independent. If we consider n observations, the model (2.1.1) may be written

(2.1.2) $\underset{n\times 1}{\underline{y}} = \alpha \underset{n\times 1}{\underline{e}} + \beta \underset{n\times 1}{\underline{x}} + \underset{n\times 1}{\underline{u}} = \underset{n\times 2}{X} \underset{2\times 1}{\underline{\gamma}} + \underset{n\times 1}{\underline{u}}$, $\underline{u} \sim N(\underline{0},\sigma^2 I_n)$,

where $X = (\underline{1},\underline{x})$, $\underline{1} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$, $\underline{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{pmatrix}$, $\underline{y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{pmatrix}$, and $\chi = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$.

From $\underline{y} = X\chi + \underline{\mu}$, we have $\underline{y} - X\chi = \underline{\mu}$; because $\underline{\mu} \sim N(0,\sigma^2 I_n)$, we have

(2.1.3) $\underline{y} \sim N(X\chi,\sigma^2 I_n)$.

From least squares, we minimize

$$\underline{\mu}'\underline{\mu} = \underline{y}'\underline{y} - 2\chi'X'\underline{y} + \chi'X'X\chi$$

to obtain the "normal equations"

$$X'X\hat{\chi} = X'\underline{y} .$$

Provided $(X'X)^{-1}$ exists,

$$\hat{\chi} = (X'X)^{-1}X'\underline{y} .$$

Since $\hat{\chi}$ is a linear transformation of $\underline{y}$, this implies that

(2.1.4) $\hat{\chi} \sim N(\chi,\sigma^2(X'X)^{-1})$.

We deduce that

(2.1.5) $\underline{g}'\hat{\chi} \sim N(\underline{g}'\chi,\sigma^2\underline{g}'(X'X)^{-1}\underline{g})$

with $\underline{g}$ a vector.

From equation (2.1.5)

(2.1.6) $(\underline{g}'\hat{\chi} - \underline{g}'\chi) / \sqrt{\sigma^2\underline{g}'(X'X)^{-1}\underline{g}} \sim N(0,1)$.

The residual sum of squares (SSE) is

(2.1.7) $SSE = \underline{y}'[I - X(X'X)^{-1}]\underline{y} = \underline{y}'M\underline{y} \sim \sigma^2\chi^2_{n-2}$

with 2 the rank of X (Searle (1971), Sec. 3.5).

Because SSE and $(\underset{\sim}{\ell}'\hat{\chi}-\underset{\sim}{\ell}'\chi)/\sqrt{\sigma^2\underset{\sim}{\ell}'(X'X)^{-1}\underset{\sim}{\ell}}$ are distributed independently,

(2.1.8) $(\underset{\sim}{\ell}'\hat{\chi}-\underset{\sim}{\ell}'\chi)/\sqrt{\underset{\sim}{\ell}'(X'X)^{-1}\ SSE/(n-2)} \sim t_{n-2}$, central, where t_{n-2} follows the t-distribution with n-2 degrees of freedom.

Let t^* be such that

$$P(-t^* < t < t^*) = 1-\epsilon \text{ , for } t \sim t_{n-2} .$$

Then

(2.1.9) $$P(-t^* < (\underset{\sim}{\ell}'\hat{\chi}-\underset{\sim}{\ell}'\chi)/\sqrt{\underset{\sim}{\ell}'(X'X)^{-1}\underset{\sim}{\ell}\ SSE/(n-2)} < t^*) = 1-\epsilon.$$

Rearranging probability statement (2.1.9), we obtain

$$P(\underset{\sim}{\ell}'\hat{\chi} - t^*\sqrt{\underset{\sim}{\ell}'(X'X)^{-1}\underset{\sim}{\ell}\ SSE/(n-2)} < \underset{\sim}{\ell}'\chi < \underset{\sim}{\ell}'\hat{\chi} + t^*\sqrt{\underset{\sim}{\ell}'(X'X)^{-1}\underset{\sim}{\ell}\ SSE/(n-2)}) = 1-\epsilon,$$

where $\hat{\chi}$ is considered random.

Hence $\underset{\sim}{\ell}'\hat{\chi} \pm t^*\sqrt{\underset{\sim}{\ell}'(X'X)^{-1}\underset{\sim}{\ell}\ SSE/(n-2)}$ provides a $100(1-\epsilon)\%$ confidence interval for $\underset{\sim}{\ell}'\chi$.

If we consider the model

$$\chi = \alpha\underset{\sim}{\ell}+\beta\underset{\sim}{x}+\underset{\sim}{\mu} \text{ , } \underset{\sim}{\mu} \sim N(\underset{\sim}{0},\sigma^2 I_n) ,$$

we have

$$E(\chi) = \alpha\underset{\sim}{\ell}+\beta\underset{\sim}{x} .$$

Confidence limits for $E(y_i)$ are given by

$$(2.1.10)\quad y_{\text{limits}} = \underset{\sim}{\hat{\beta}}'\underset{\sim}{\tilde{x}} \pm t^*\sqrt{\underset{\sim}{\hat{\beta}}'(X'X)^{-1}\underset{\sim}{\hat{\beta}}\ SSE/(n-2)}$$

$$= a + b(x_i-\bar{x}) \pm t^*\sqrt{[1/n(x_i-\bar{x})^2/\underset{\sim}{x}'C\underset{\sim}{x}]\ MSE},$$

where a, b are, respectively, the least squares estimates of α and β, and where

$$\underset{\sim}{\hat{\beta}}'\underset{\sim}{\tilde{x}} = a + b(x_i-\bar{x}),\quad (X'X)^{-1} = \begin{pmatrix} 1/n & 0 \\ 0 & 1/\underset{\sim}{x}'C\underset{\sim}{x} \end{pmatrix},\quad MSE = SSE/(n-2),$$

$$\underset{\sim}{x}'C\underset{\sim}{x} = \underset{\sim}{x}'[I - \underset{\sim}{e}\underset{\sim}{e}'/n]\underset{\sim}{x} = \underset{\sim}{x}'\underset{\sim}{x} - \underset{\sim}{x}'\underset{\sim}{e}\underset{\sim}{e}'\underset{\sim}{x}/n = \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2/n .$$

Confidence limits for $E(y_i)$ give limits within which $\alpha + \beta x_i$ is likely to lie.

But very often we are interested in the limits within actual values of y, corresponding to which a given value of x may lie.

We have the model $\underset{\sim}{y} = \alpha\underset{\sim}{e} + \beta\underset{\sim}{x} + \underset{\sim}{u}$, $\underset{\sim}{u} \sim N(\underset{\sim}{0}, \sigma^2 I_n)$, and we want to obtain the limits within which the actual value of y_i, or the limits within which an estimate of y_i will lie.

The variance of a single value y_i about the regression line is

$$V(y_i) = \sigma^2[1+(1/n)+(x_i-\bar{x})^2/\underset{\sim}{x}'C\underset{\sim}{x}] .$$

The limits within which y_i will lie are no longer called confidence limits, but tolerance limits (cf. Kendall and Buckland (1970), for the difference between tolerance and confidence limits).

The tolerance limits of y_i are given by

$$(2.1.11)\quad y_{\text{limits}} = \underline{\hat{\beta}}'\underline{x} \pm t^*\sqrt{[1+\underline{x}'(X'X)^{-1}\underline{x}]MSE}$$

$$= a + b(x_i-\bar{x}) \pm t^*\sqrt{[1+(1/n)+(x_i-\bar{x})^2/\underline{x}'C\underline{x}]MSE}$$

The difference between (2.1.10) and (2.1.11) is the value 1 under the square root sign. This can be explained in that for confidence limits of the predicted mean value of y_i for a given x_i, we use the variance of $\hat{y}_i = \sigma^2[(1/n)+(x_i-\bar{x})^2/\underline{x}'C\underline{x}]$; but in the case of the tolerance limits of the predicted value of an individual, the variance of $\hat{y}_i = \sigma^2[1+(1/n)+(x_i-\bar{x})^2/\underline{x}'C\underline{x}]$.

2.2. Tolerance intervals for x given y.

In the preceding section, we showed how to find tolerance intervals for the variable y. Suppose now we want to find tolerance intervals for the variable x. In linear regression, we can consider two methods. Let

$$(2.2.1)\quad y_i = \alpha+\beta x_i+\epsilon_i \ , \ i = 1, 2, \ldots, n \ .$$

If we consider Krutchkoff's inverse method, we rewrite model (2.2.1) as

$$(2.2.2)\quad x_i = \gamma+\delta y_i+\epsilon_i' \ , \ i = 1, 2, \ldots, n,$$

where $\gamma = -\alpha/\beta$, $\delta = -1/\beta$, $\varepsilon_i' = -\varepsilon_i/\beta$, $\beta \neq 0$.

If we minimize the squared residuals of x, we obtain the least squares line

$$x = c+dy ,$$

where c, d are, respectively, the least squares estimates of γ and δ.
The estimate of x becomes

$$\hat{x} = c+dy .$$

Using equation (2.1.11), the tolerance intervals for $\hat{x}$ will be of the form

$$(2.2.3) \quad \bar{x} + d(y_0-\bar{y}) \pm t^*\sqrt{[1+(1/n)+(y_0-\bar{y})^2/\sum_{i=1}^{n}(y_i-\bar{y})^2]MSE_x} .$$

If we use the classical method, we consider the model (2.2.1). Let a, b be, respectively, the least squares estimates of α and β.
The least squares line becomes

$$y = a+bx .$$

The estimate of x is

$$\hat{x} = (y_0-a)/b , \; b \neq 0 .$$

To obtain the tolerance intervals of $\hat{x}$, we use equation (2.1.11)

(2.2.4)

$$\underset{\sim}{g}'\hat{\underset{\sim}{\chi}} - t^* \sqrt{(\underset{\sim}{g}'(X'X)^{-1}\underset{\sim}{g}+1)MSE} \leq y \leq \underset{\sim}{g}'\hat{\underset{\sim}{\chi}} + t^* \sqrt{(\underset{\sim}{g}'(X'X)^{-1}\underset{\sim}{g}+1)MSE}.$$

Squaring and rearranging equation (2.2.4) gives

$$(2.2.5)\quad (y-\underset{\sim}{g}\hat{\underset{\sim}{\chi}})^2 \leq (t^*)\ (\underset{\sim}{g}'(X'X)^{-1}\underset{\sim}{g}+1)MSE\ .$$

Substituting the values of $\hat{\underset{\sim}{\chi}}$ and $\underset{\sim}{g}'(X'X)^{-1}\underset{\sim}{g}$ in equation (2.2.5), we obtain

$$(2.2.6)\quad \left[\left(\frac{\underset{\sim}{g}'C\underset{\sim}{\chi}}{\underset{\sim}{\chi}'C\underset{\sim}{\chi}}\right)^2 - t^{*2}\frac{MSE}{\underset{\sim}{\chi}'C\underset{\sim}{\chi}}\right](x-\bar{x})^2 + \left[-2(y-\bar{y})\frac{\underset{\sim}{g}'C\underset{\sim}{\chi}}{\underset{\sim}{\chi}'C\underset{\sim}{\chi}}\right](x-\bar{x})$$

$$+ [(y-\bar{y}) - t^{*2}MSE(1+1/n)] \leq 0\ .$$

Let $A = \left(\frac{\underset{\sim}{g}'C\underset{\sim}{\chi}}{\underset{\sim}{\chi}'C\underset{\sim}{\chi}}\right)^2 - t^{*2}\frac{MSE}{\underset{\sim}{\chi}'C\underset{\sim}{\chi}}$, $B = -2(y-\bar{y})\frac{\underset{\sim}{g}'C\underset{\sim}{\chi}}{\underset{\sim}{\chi}'C\underset{\sim}{\chi}}$

and $C = (y-\bar{y}) - t^{*2}MSE[1+1/n]$.

By solving the quadratic equation $A(x-\bar{x})^2 + B(x-\bar{x}) + C$, we can obtain tolerance intervals for x (Scheffé (1959), problem 2.3, page 51).

Let r_1 and r_2 be the two roots of $A(x-\bar{x})^2 + B(x-\bar{x}) + C$; the solution for r_1 and r_2 are

$$r_1 = \frac{-B - \sqrt{B^2 - 4AC}}{2A}$$

$$r_2 = \frac{-B + \sqrt{B^2 - 4AC}}{2A}\ .$$

There exist several possibilities when solving $A(x-\bar{x})^2 + B(x-\bar{x}) + C$, which we now investigate in detail.

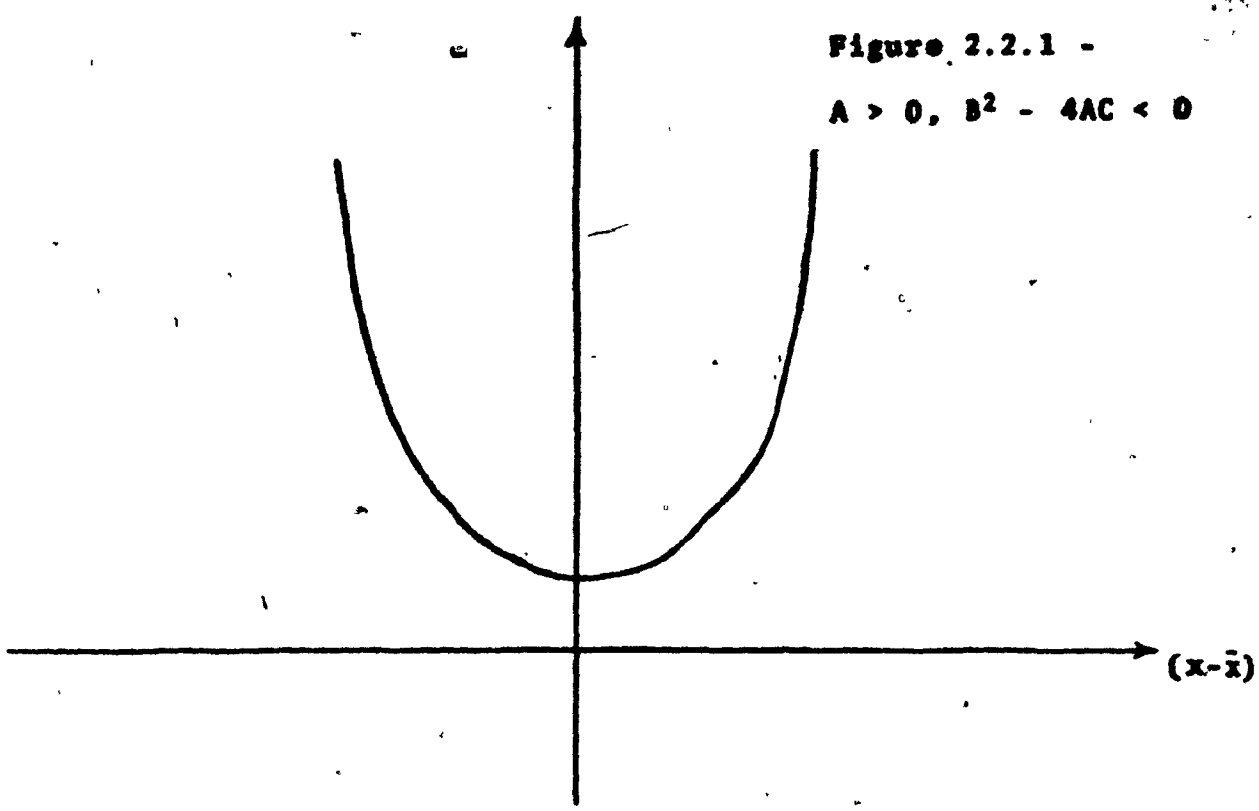
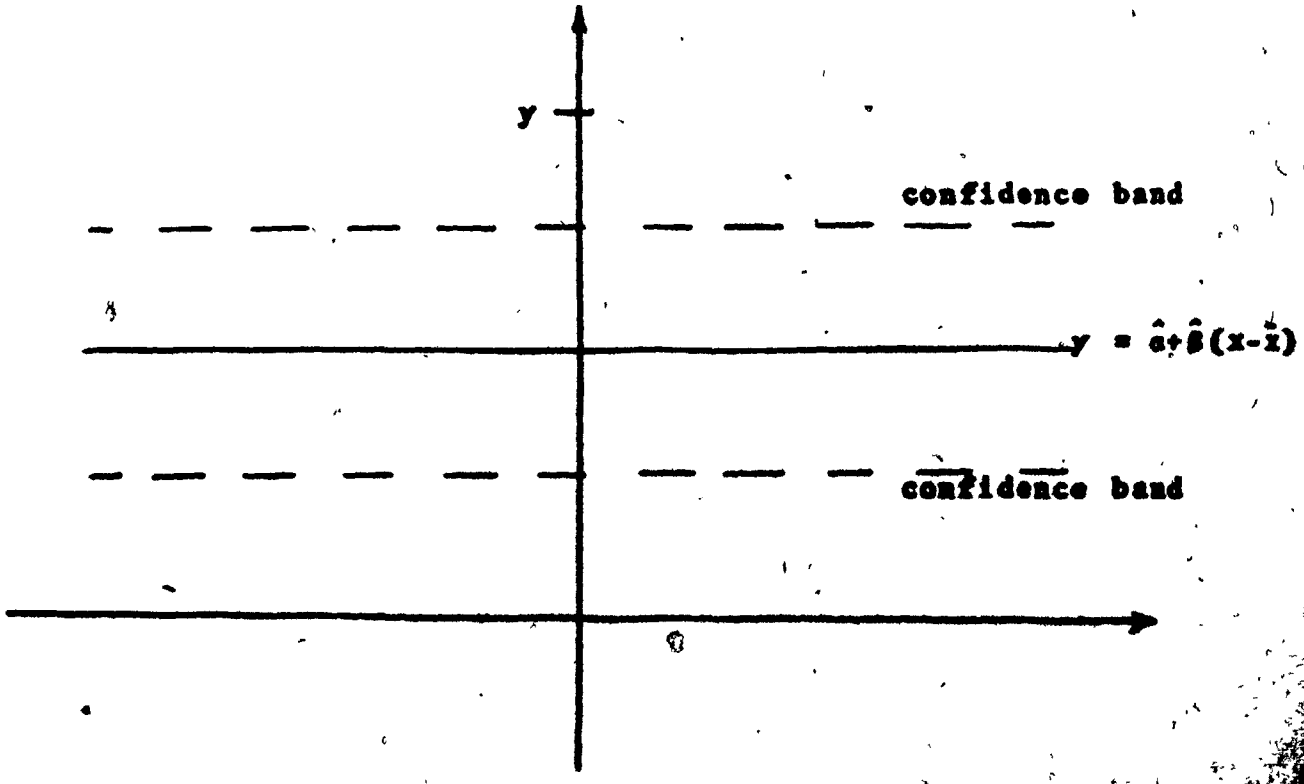


Figure 2.2.1 -

$A > 0,\ B^2 - 4AC < 0$

no point will satisfy the equation



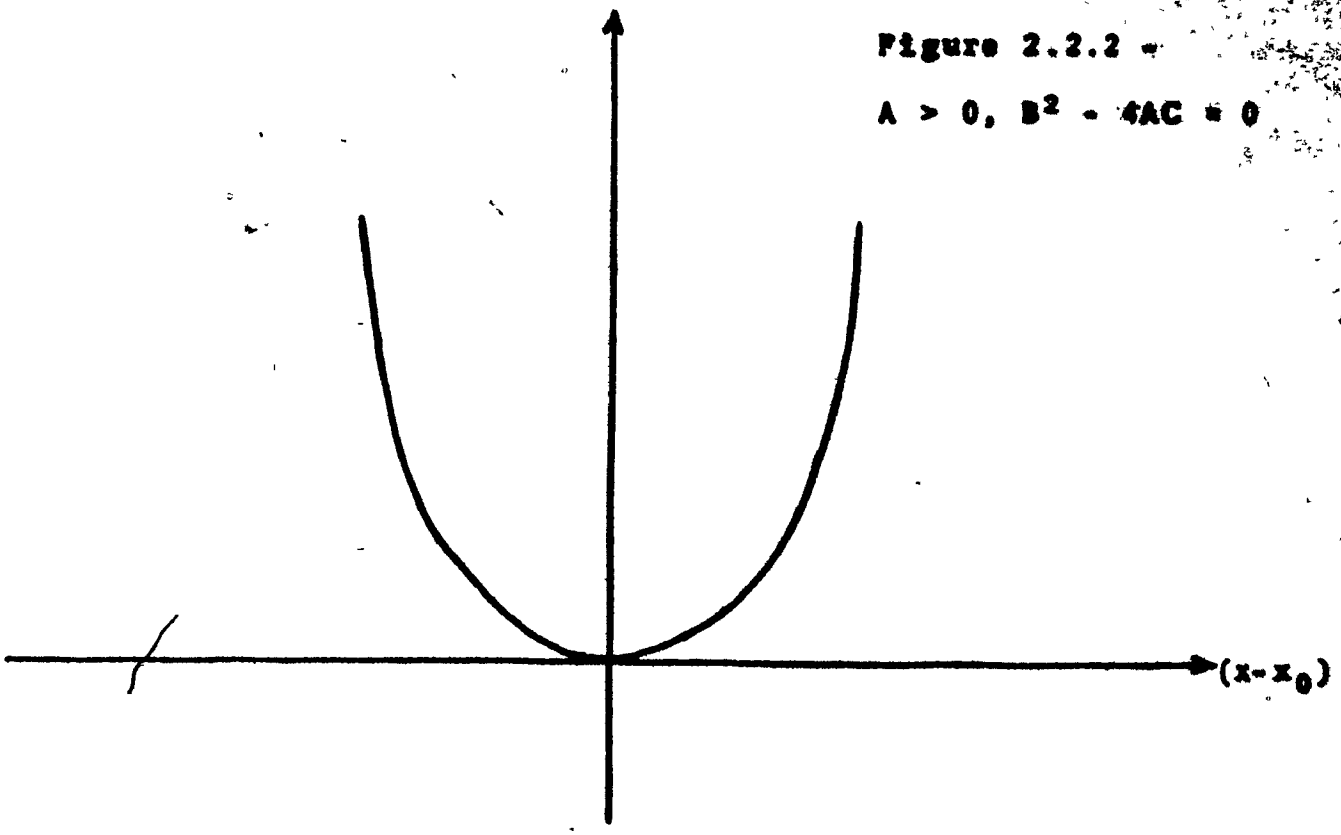
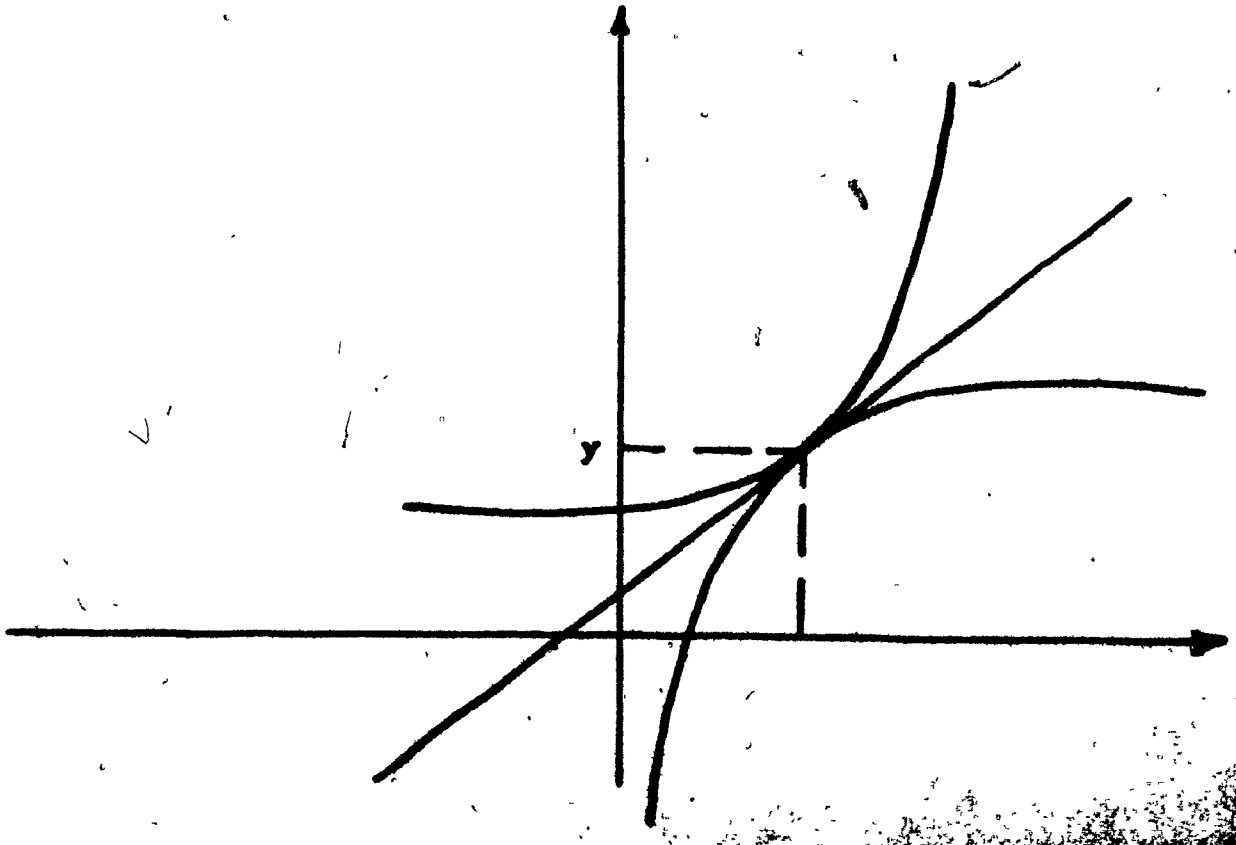


Figure 2.2.2 -

$A > 0,\ B^2 - 4AC = 0$

one point only satisfies the equation



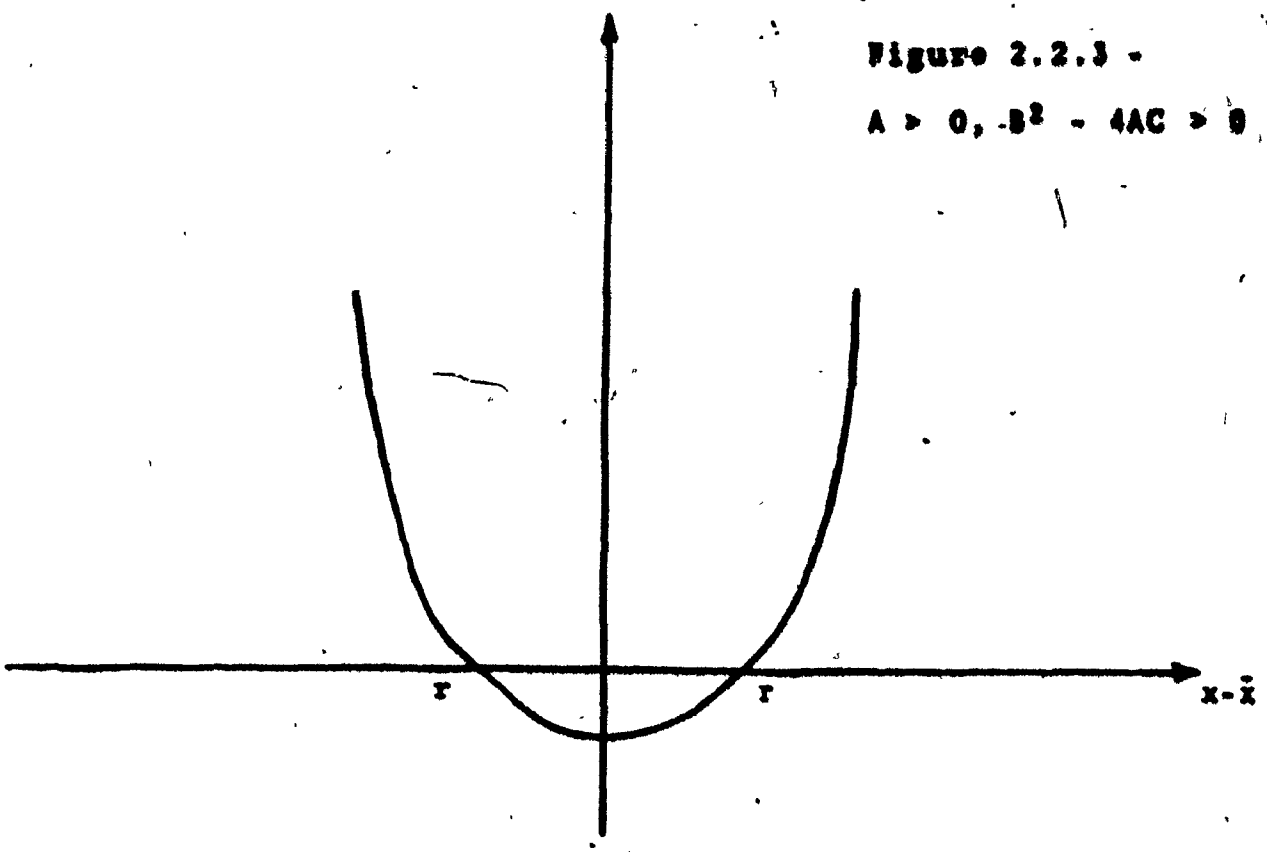
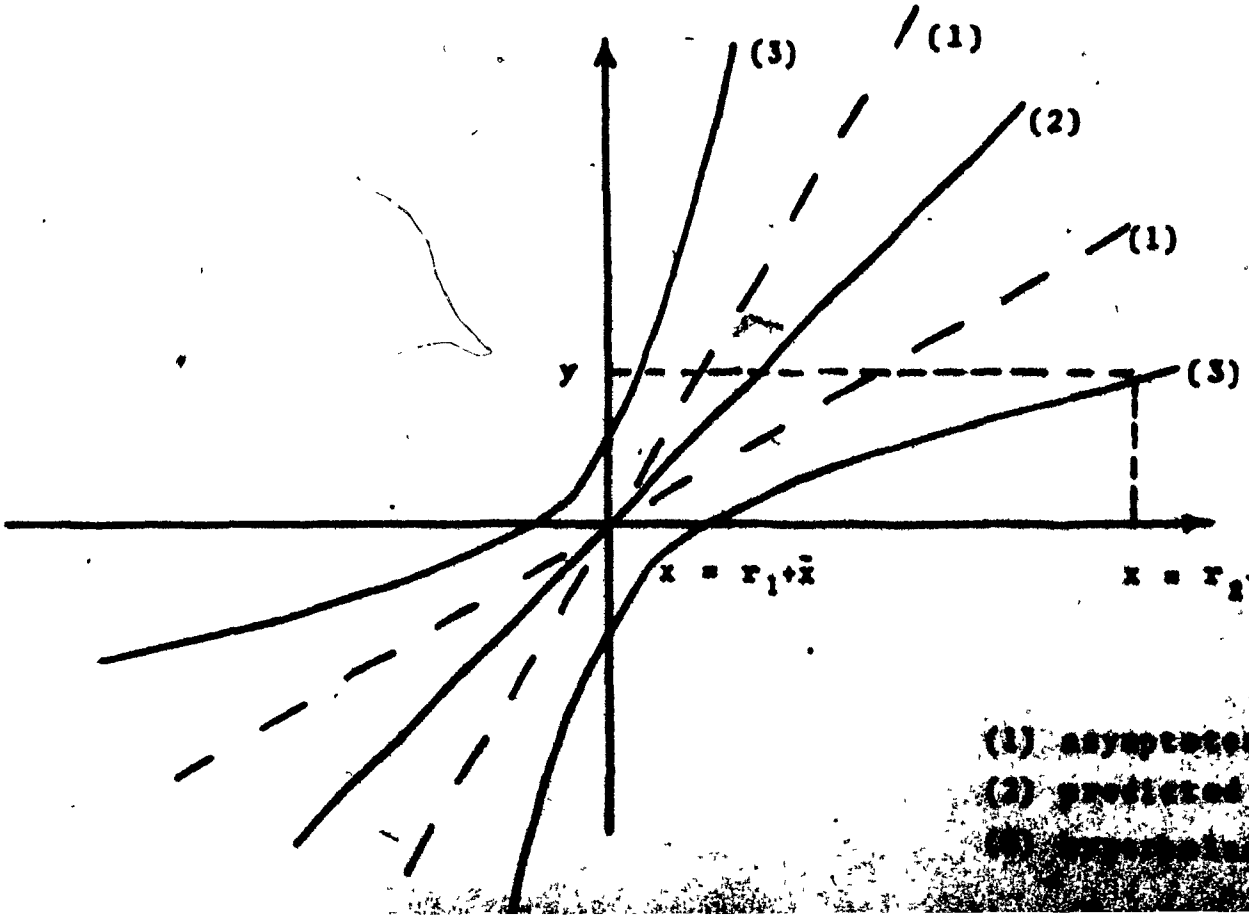


Figure 2.2.3 -
$A > 0, -B^2 - 4AC > 0$

the equation is satisfied when $r_1 < x-\bar{x} < r_2$



(1) asymptotes
(2) [illegible]
(3) [illegible]

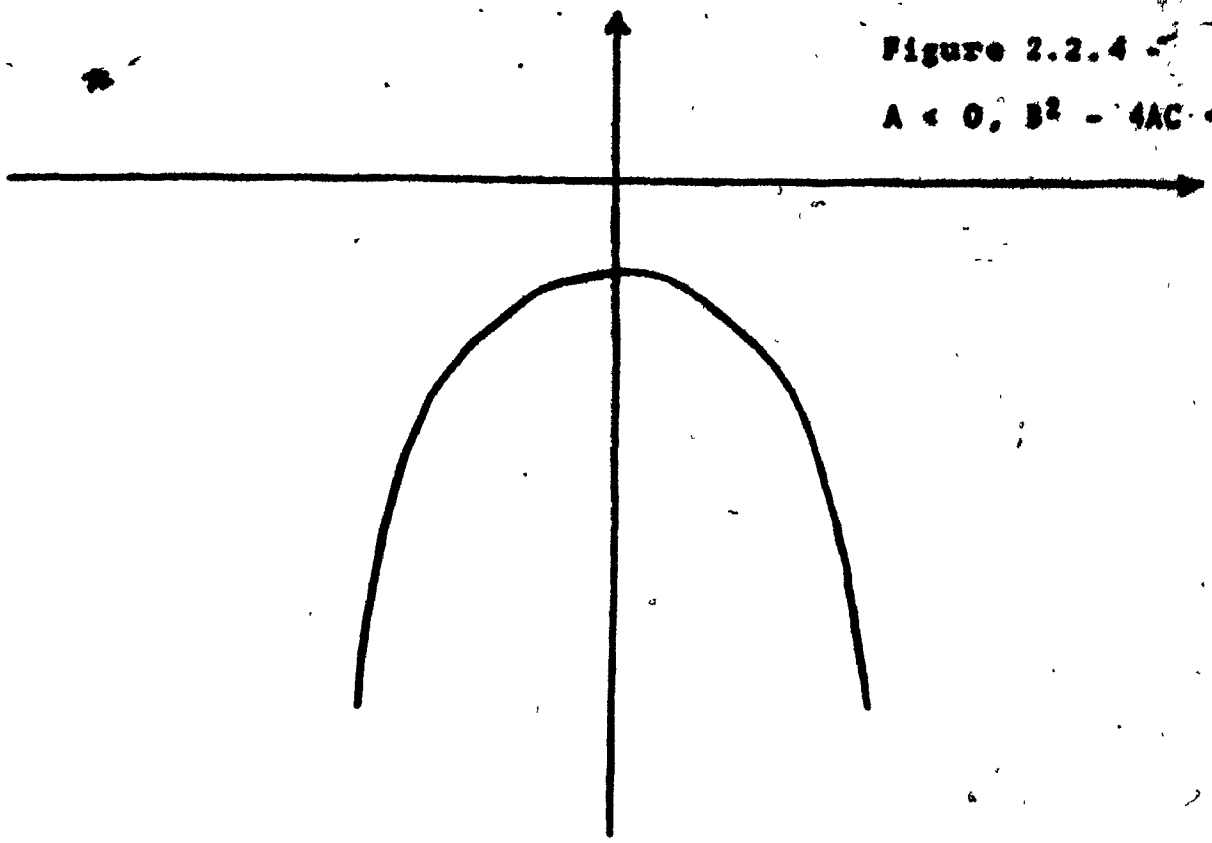
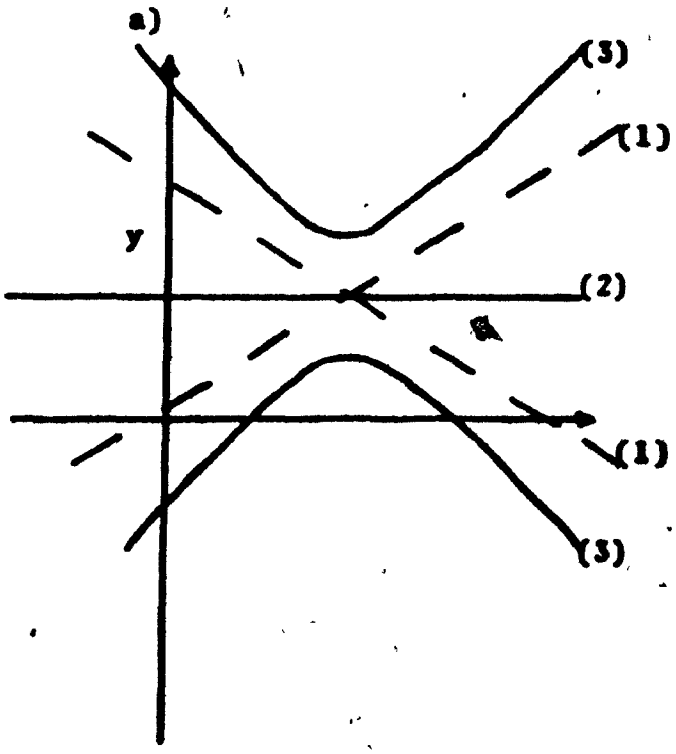


Figure 2.2.4 -
$A < 0,\ B^2 - 4AC < 0$

equation is satified for all x, $-\alpha < x < \infty$



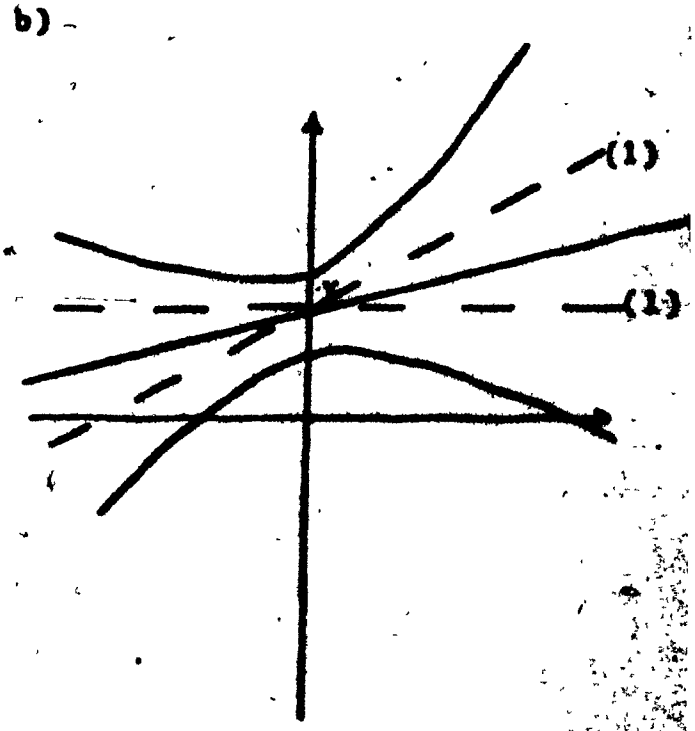
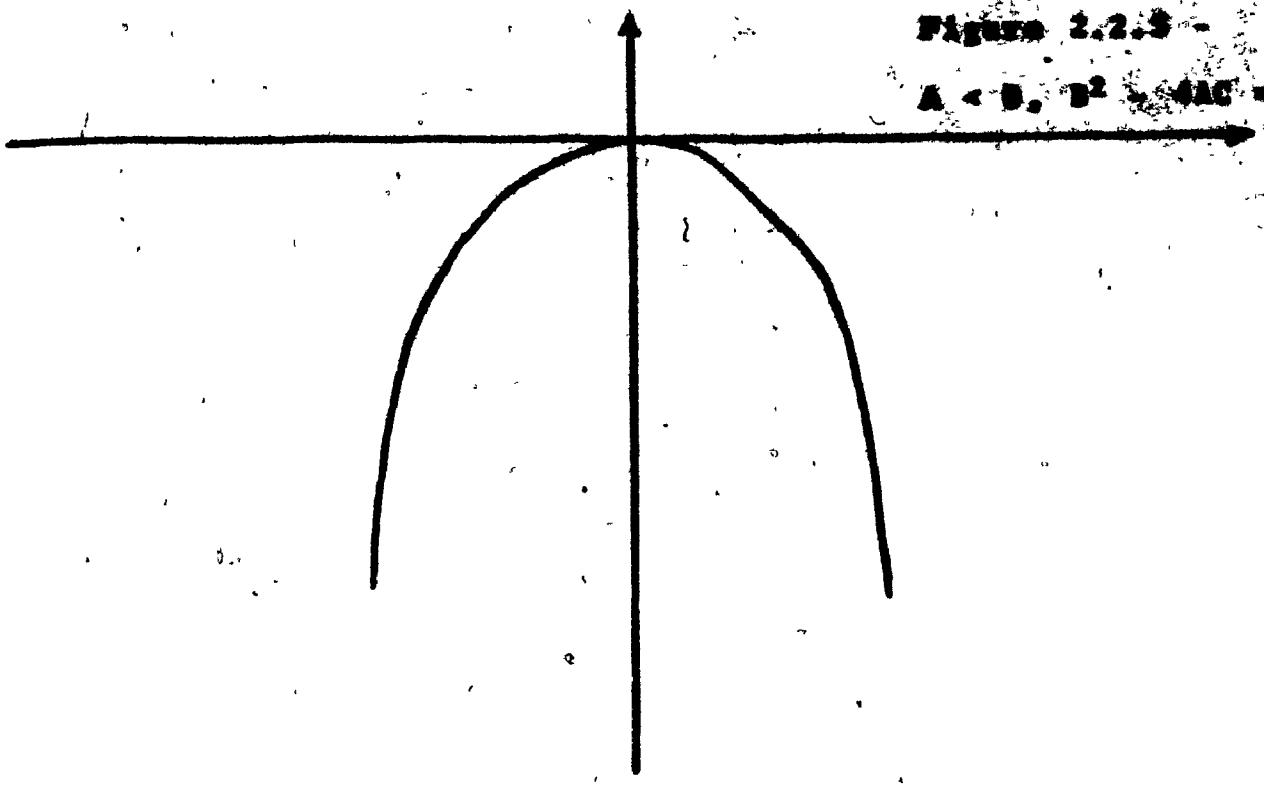
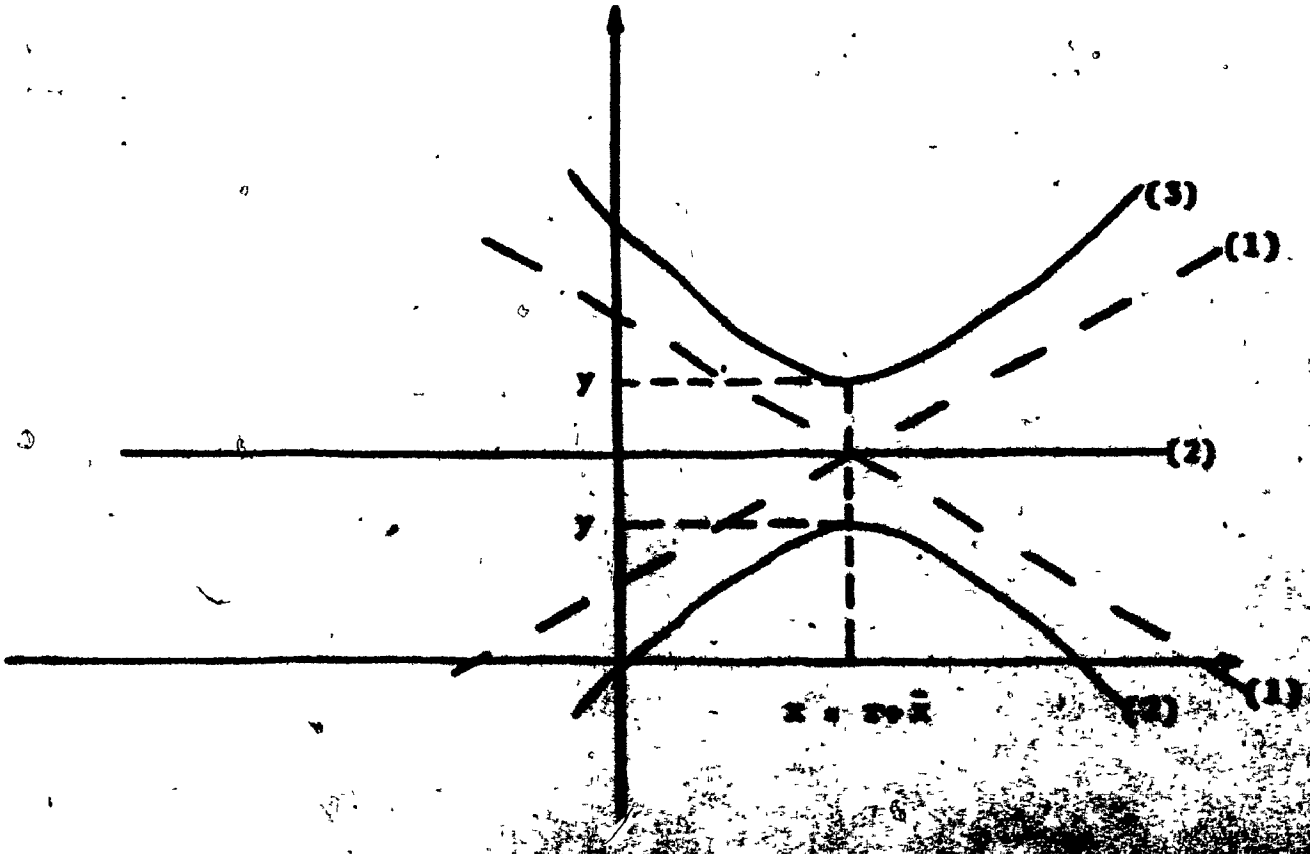


Figure 2.2.3 -

$A < 0, B^2 - 4AC = 0$



one double root at $r = x-\bar{x}$



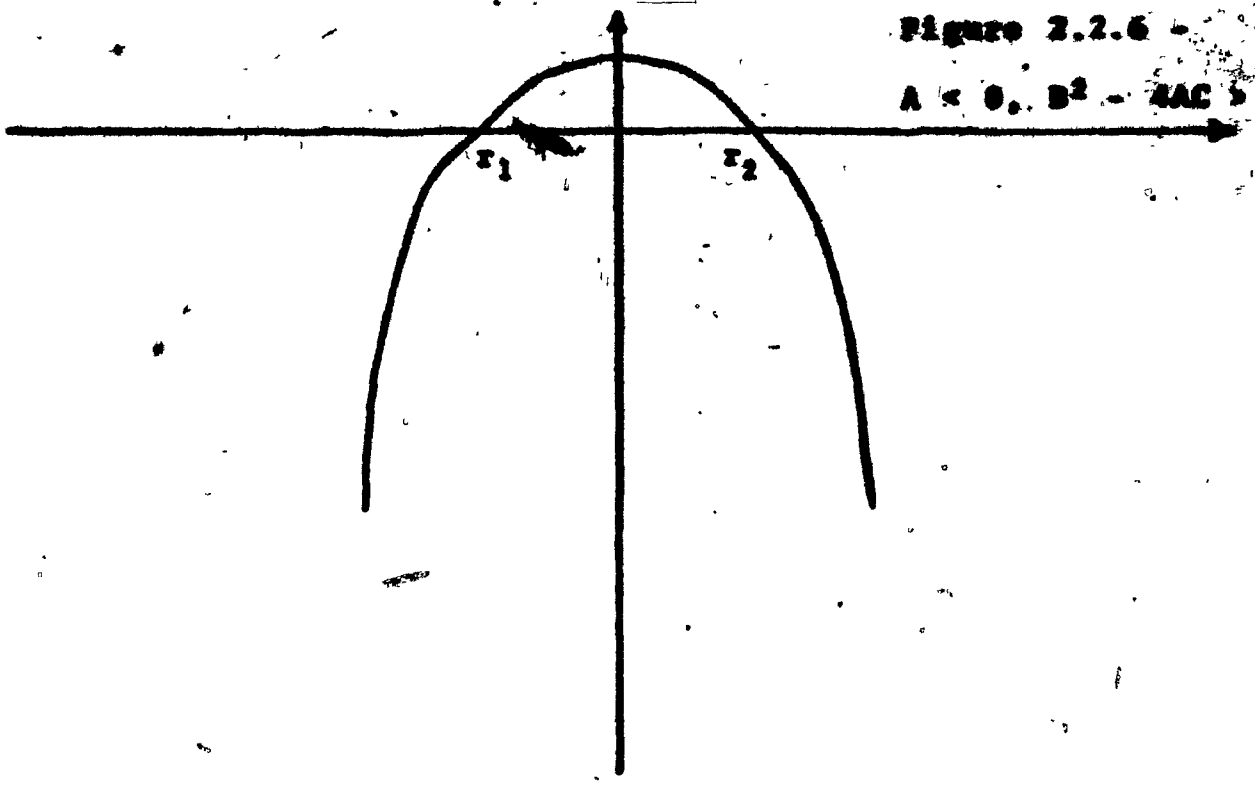
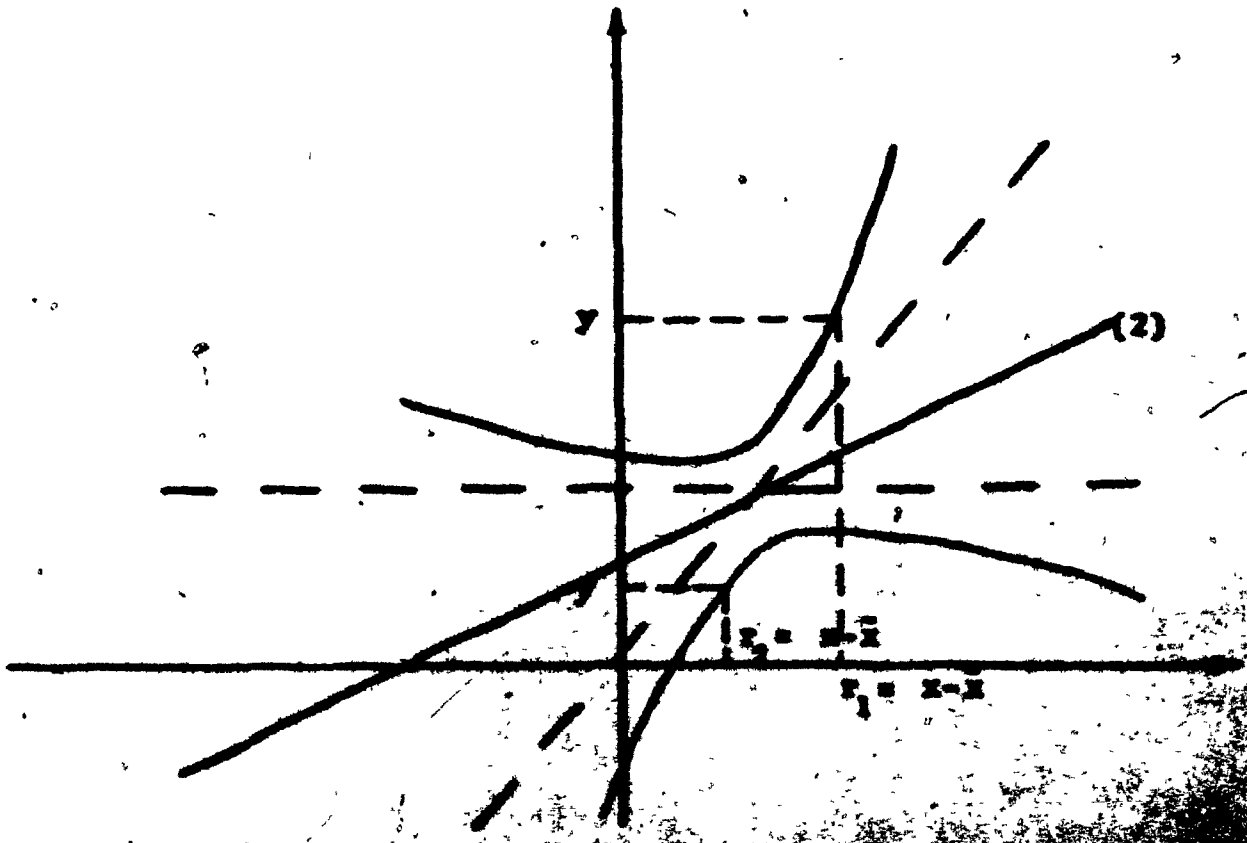


Figure 2.2.6 -
$A < 0,\ B^2 - 4AC > 0$

equation is satisfied for $r_1 < x-\bar{x}$ and $r_2 > x-\bar{x}$



We can, therefore, obtain a tolerance interval for x only when, as in Figure 2.2.3, $A > 0$ and $B^2 - 4AC > 0$; the equation for the tolerance interval for x is given by

(2.2.7) $$r_1 + \bar{x} < x < r_2 + \bar{x} .$$

2.3 Comparison between the inverse and classical methods.

Using the mean squared deviation criterion, Krutchkoff (1967) showed that the inverse method is better than the classical method, when one wants to predict some unknown x variable given the y variable. Many people have corroborated Krutchkoff's conclusion by simulation. Let us consider the data of problem 2.3 on page 51 of Scheffé (1959).

"In an experiment to calibrate an instrument for measuring the moisture of a certain material by an electrical method, the results obtained were, after ordering of x values, the following:

TABLE 2.3.1

x	6.0	6.3	6.5	6.8	7.0	7.1	7.5	7.5
y	39	58	49	53	80	86	115	124
x	7.6	7.8	8.0	8.2	8.4	8.4	8.9	
y	104	131	147	160	156	172	180	

where x is the moisture content in per cent, determined by an analytic method sufficiently accurate to be considered free of error, and y is the dial reading in the instrument..."

If we consider the model

(2.3.1) $y = \alpha+\beta x+\varepsilon$, ε = error term

the least squares line becomes

(2.3.2) $\hat{y} = a+bx$,

where a, b are, respectively, the least squares estimates of α and β.

Therefore, the estimate of x is

(2.3.3) $\hat{x}_{cl} = (\hat{y}-a)/b$, $b \neq 0$.

This x-estimate is obtained by the classical procedure.

Now, let us consider the model

(2.3.4) $x = \gamma+\delta y+\epsilon'$, $\gamma = -\alpha/\beta$, $\delta = 1/\beta$, $\epsilon' = -\epsilon/\beta$, $\beta \neq 0$.

Then, the least squares line is,

(2.3.5) $\hat{x} = c+dy$,

where c, d are, respectively, the least squares estimators of γ and δ; the x-estimate is

(2.3.6) $\hat{x}_{in} = c+dy$.

This x-estimate is obtained by the inverse procedure.

Now let us consider the squared errors about the regression line for both procedures; the squared errors of the classical estimate and the inverse estimate are respectively

(2.3.7) $(x-\hat{x}_{cl})^2$,

and

(2.3.8) $(x-\hat{x}_{in})^2$.

The following table contains the squared errors for the classical method and the inverse method and it also contains the ratios of both squared errors (those three values are computed for each x of Table 2.3.1).

TABLE 2.3.2 - Comparison of squared errors about the regression lines

	x=6.0	x=6.3	x=6.5	x=6.8	x=7.0
$(x-\hat{x})^2_{cl}$	.02865	.04629	.02211	.14128	.00711
$(x-\hat{x})^2_{in}$	.15259	.06719	.00942	.10732	.00346
Ratio(cl/in)	.54469	.68893	2.34764	1.31644	2.05562
	x=7.1	x=7.5	x=7.5	x=7.6	x=7.8
$(x-\hat{x})^2_{cl}$	.11564	.00279	.4695	.06122	.00195
$(x-\hat{x})^2_{in}$	.00299	.00239	.4207	.05863	.00071
Ratio(cl/in)	1.88874	1.17005	1.11608	1.04411	2.74272
	x=8.0	x=8.2	x=8.4	x=8.4	x=8.9
$(x-\hat{x})^2_{cl}$	.01833	.02961	.01015	.03630	.02684
$(x-\hat{x})^2_{in}$	.01091	.01694	.01941	.01918	.04956
Ratio(cl/in)	1.68097	1.74811	.52310	1.89283	.54161

From Table 2.3.2, we can see that the squared error of the classical method is larger than the squared error of the inverse method, except in four occasions where the squared error of the classical method is smaller. If we can judge the efficency of both approaches by looking at the squared error criterion, we can conclude that the classical approach is worse than the inverse approach.

To give more weight to this last statement, we will compute tolerance intervals for each x, using both methods.

If we use model (2.3.1)

$$y = \alpha+\beta x+\varepsilon ,$$

and if we look at the formulae in Section 2.2, the tolerance intervals for x will be obtained by solving the second degree inequality

$$\left[(\underline{x}'C\underline{\alpha}/\underline{x}'C\underline{x})^2 - \frac{SSEt^*}{(n-2)\underline{x}'C\underline{x}}\right](x_0-\bar{x})^2 + \left[-2(y-\bar{y})(\underline{x}'C\underline{\chi}/\underline{x}'C\underline{x})\right](x_0-\bar{x})$$

$$+ (y-\bar{y})^2 - \frac{t^{*2}SSE}{n-2}\left[1+\frac{1}{n}\right] \leq 0 ,$$

where y is known and x is the unknown value of x for which the tolerance interval is desired; t^* is the critical value of the t-distribution with n-2 degrees of freedom and

$$\underline{x}'C\underline{\chi} = \underline{x}'[I-\underline{e}\underline{e}'/n]\underline{\chi} = \sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i/n ,$$

$$\underline{x}'C\underline{x} = \underline{x}'[I-\underline{e}\underline{e}'/n]\underline{x} = \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2/n ,$$

$$X = (\underline{e}, C\underline{x}) \quad , \quad SSE = \underline{y}'[I-X(X'X)^{-1}X']\underline{y}$$

$$= \sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2/n - \left[\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i/n\right]^2 \Big/ \left[\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2/n\right].$$

Let $A = \left[(\underline{x}'C\underline{Y}/\underline{x}'C\underline{x}) - \frac{SSEt^{*2}}{(n-2)\underline{x}'C\underline{x}}\right]$,

$B = [-2(y-\bar{y})(\underline{x}'C\underline{Y}/\underline{x}'C\underline{x}]$,

and $C = (y-\bar{y})^2 - \frac{t^{*2}SSE}{n-2}\left[1+\frac{1}{n}\right]$.

Therefore, tolerance intervals for x will be computed by solving the second degree inequality

$$A(x_0-\bar{x})^2 + B(x_0-\bar{x}) + C \leq 0 .$$

The associated quadratic equation gives two roots

$$x_0-\bar{x} = \frac{-B \pm \sqrt{B^2 - 4AC}}{2A} .$$

Let r_1 and r_2 be the two roots ($r_1 < r_2$). Since $A > 0$ ($A = 3101.10$) and $B^2 - 4AC > 0$ ($B^2 - 4AC = 10860725.074$), then we have the situation as in Figure 2.3.1, where

$$r_1 < x_0-\bar{x} < r_2 ,$$

(2.3.7) $$r_1 + \bar{x} < x_0 < r_2 + \bar{x} .$$

Now, if we use model (2.3.4)

$$x = \gamma+\delta y+\varepsilon', \quad \gamma = -\alpha/\beta , \quad \delta = 1/\beta , \quad \varepsilon' = -\varepsilon/\beta , \quad \beta \neq 0 ,$$

the tolerance intervals for x are given by

(2.3.8) $$\underline{g}'\hat{\underline{\gamma}} - t^*\sqrt{(1+\underline{g}'(X'X)^{-1}\underline{g})MSE} < x < \underline{g}'\hat{\underline{\gamma}} + t^*\sqrt{(1+\underline{g}'(X'X)^{-1}\underline{g})MSE} .$$

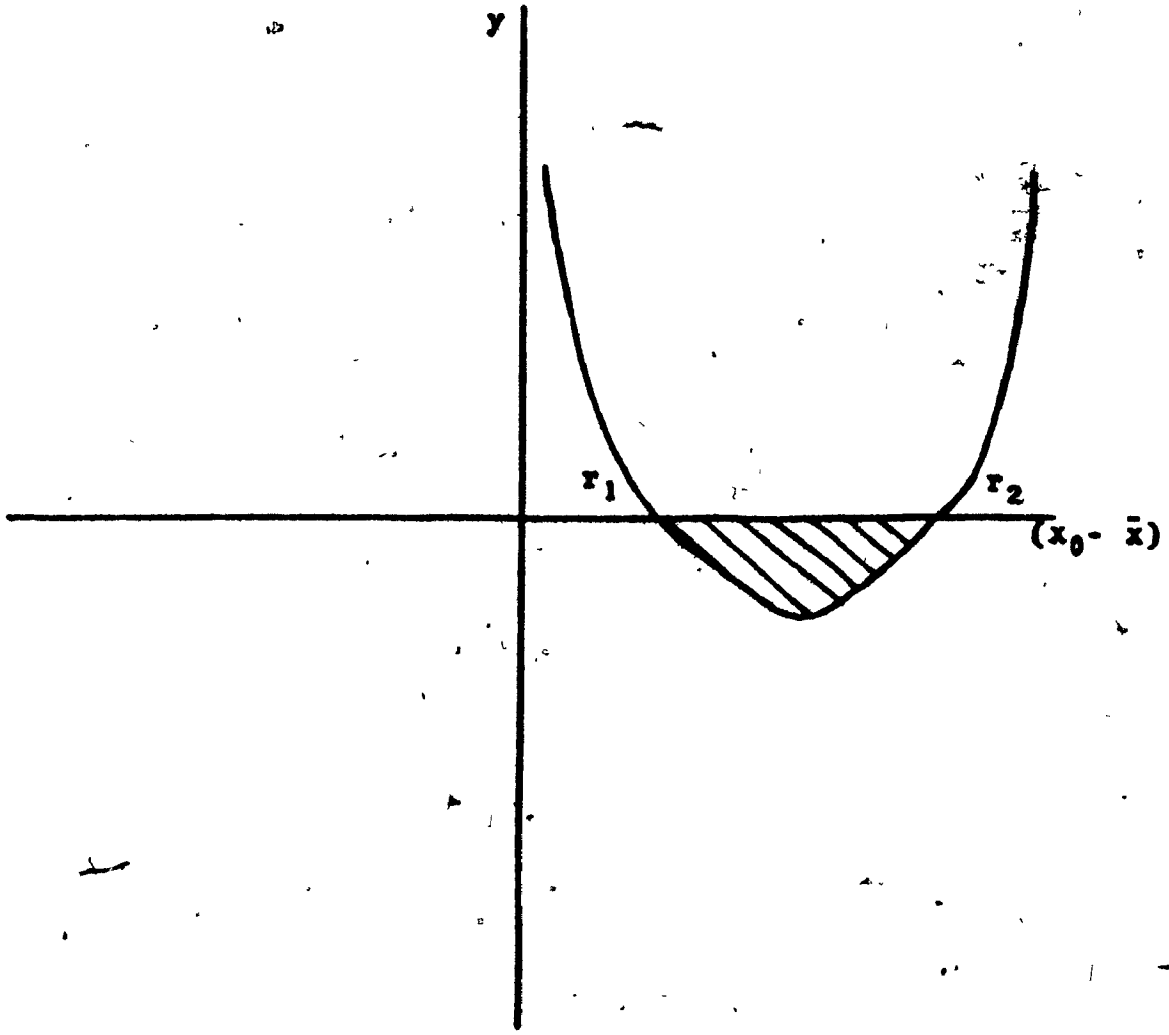


Figure 2.3.1

$$\bar{x} + (\underline{\ell}'C\underline{\chi}/\underline{\chi}'C\underline{\chi})(y_0-\bar{y}) - t^*\sqrt{(1+\frac{1}{n} + (y_0-\bar{y})^2/\underline{\chi}'C\underline{\chi})MSE_x} < x$$

$$< \bar{x} + (\underline{\ell}'C\underline{\chi}/\underline{\chi}'C\underline{\chi})\ (y_0-\bar{y}) + t^*\sqrt{(1+\frac{1}{n} + (y_0-\bar{y})^2/\underline{\chi}'C\underline{\chi})MSE_x}\ ,$$

where y_0 is the known variable, and where t^* is the critical value of the t-distribution with n-2 degrees of freedom and $\underline{X}_0 = (\underline{\ell},\underline{\chi})$

$$MSE_x = SSE_x/(n-2) = \underline{\ell}'[I-X(X'X)^{-1}X']\underline{\ell}/(n-2)$$

$$= \left[\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2/n \right.$$

$$\left. - \left(\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i/n\right)^2 \Big/ \left(\sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2/n\right)\right] \Big/ (n-2)\ .$$

Table 2.3.3 gives the tolerance intervals for x, using the classical method and the inverse method.

TABLE 2.3.3 - Tolerance intervals for x.

X	Y	CL. X_L	CL. X_U	IN. X_L	IN. X_U
6.0	39	5.51680	6.76157	5.71458	6.74409
6.3	58	5.88571	7.10048	6.06026	7.05816
6.5	49	5.71149	6.93942	5.89706	6.90885
6.8	53	5.78904	7.01088	5.96972	6.97508
7.0	80	6.30737	7.49841	6.45481	7.42754
7.1	86	6.42198	7.60801	6.56129	7.52941
7.5	115	6.96524	8.14443	7.06880	8.02890
7.5	124	7.13177	8.31318	7.22388	8.18633
7.6	104	6.76023	7.93965	6.87770	7.83803
7.8	131	7.26056	8.44517	7.34370	8.30958
8.0	147	7.55252	8.74927	7.61504	8.59383
8.2	160	7.78734	8.99876	7.83296	8.82731
8.4	156	7.71531	8.92177	7.76614	8.75524
8.4	172	8.00224	9.23090	8.03220	9.04476
8.9	180	8.14457	9.38661	8.16405	9.19070

We see that the intervals are wider for the classical method, for each x. Since the squared error is smaller for the inverse method, and since the inverse method gives narrower intervals, we may conclude that the inverse method seems better than the classical method in the prediction of the x-variable and in the construction of an associated tolerance interval.

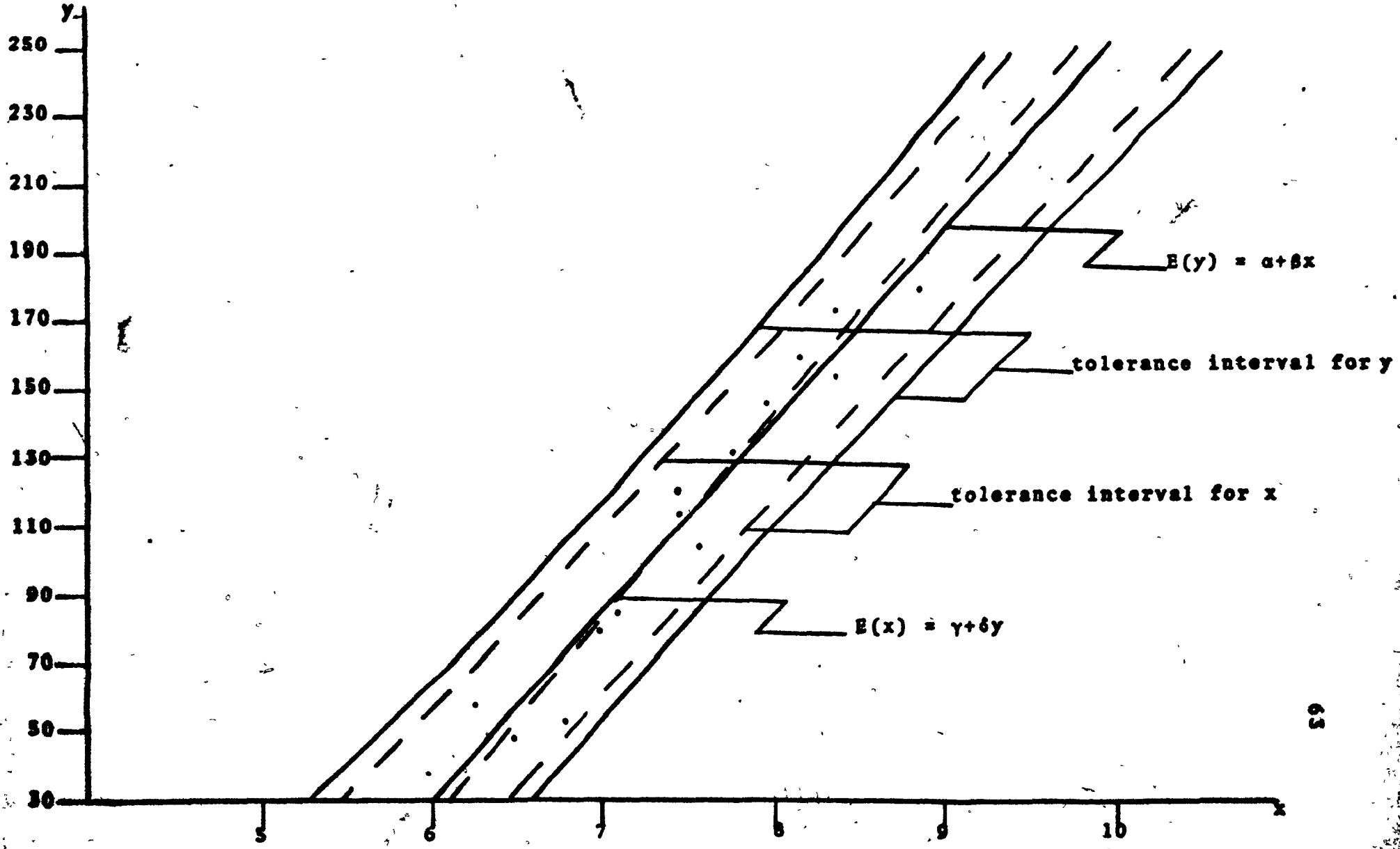


Figure 2.3.2 - Tolerance intervals for x using classical and inverse method.

2.4 Applications to blood hormone cell analysis.

In poultry production, the control of stressors, that is physical, chemical or emotional factors that cause body or mental tension, is well recognized by industry. In a paper in poultry science, Buckland, Blagrave and Lague (1973) developed and evaluated "a competitive protein-binding assay for corticoids in the peripheral plasma of the immature chicken which would measure their response to stressors". Four factors were considered in the evaluation of the protein-binding assay: specificity, accuracy, precision and sensitivity.

For their experiments, Buckland, Blagrave and Lague used single Comb White Leghorn chicks; these chicks were grown "in battery brooders with 14 hours of light and fed on an 18% protein medicated chick starter". 4 ml. of blood by heart puncture were taken at 39 days of age, and "Rhesus monkey plasma was chosen as the source of corticosteroid binding globulin". They extracted twice duplicate plasma, and "each duplicate was counted twice to an error of 15% in a Beckman model LS-235 scintillation counter".

Because corticosterone is the principal corticoid in chicken plasma, corticosterone was chosen for

the construction of "the standard curve". If P is the amount of corticosterone in nanograms, and A the count per minute, the usual regression problem would be to study the distribution of the response A-variable, when the P-variable is specified. An alternative problem exists when the values of P cannot be determined directly, but are to be analyzed by a study of the observed responses. This is widely used with a chemical stimulus and with responses of a biological system, and in this situation is called bioassay. The chemical is quantified or assayed by the biological response.

Let $y = \log A$ and $x = \log(P+1)$. We suppose that the logarithm of the count per minute has a polynomial regression on the logarithm of the amount of corticosterone, so that the regression model is

$$(2.4.1) \qquad y = \alpha + \sum_{i=1}^{p} \beta_i x_i + u \ , \ p \geq 1.$$

We will consider two cases: the linear case, that is $p = 1$, and the polynomial case, that is $p > 1$.

Let us consider the case where $p = 1$, that is

$$y = \alpha+\beta x+u \ .$$

If we consider the classical method, then the least squares line is given by

$$\hat{y} = a+bx \ ,$$

where a, b are, respectively, the least squares estimates of α and β. The estimate for x becomes

$$\hat{x}_{cl} = (\hat{y}-a)/b \ , \ b \neq 0 \ .$$

Now, if we look at the inverse approach

$$x = \gamma+\delta+\epsilon' , \ \gamma = -\alpha/\beta \ , \ \delta = 1/\beta \ , \ \epsilon' = -\epsilon/\beta \ , \ \beta \neq 0 \ ,$$

then the least squares line is

$$\hat{x}_{in} = c+dy \ ,$$

where c, d are, respectively, the least squares estimates of γ and δ.

As we have said earlier, there are four values in counts per minute for each amount of corticosterone. Therefore, the average squared errors of the classical estimate and the inverse estimate are respectively

$$\sum_{i=1}^{4} \frac{(x_j - \hat{x}_{cl(i,y)})^2}{4} \ , \text{ for each } x_j, \ x = 1, \ldots, 32 \ ,$$

$$\sum_{i=1}^{4} \frac{(x_j - \hat{x}_{in(i,y)})^2}{4} \ , \text{ for each } x_j, \ y = 1, \ldots, 32 \ .$$

The following table contains the average squared errors for the inverse method and for the classical method and it contains also their ratio (cl/in).

TABLE 2.4.1 - Comparison of the average squared error for the inverse and classical method for various levels of corticosterone, P.

	P=.5	P=1.	P=2.	P=3.
$Av(x-\hat{x})^2_{cl}$	.01007	.00261	.00179	.00048
$Av(x-\hat{x})^2_{in}$	.01074	.00225	.00376	.00177
Ratio(cl/in)	.93774	1.16119	1.08591	1.0079
	P=4.	P=6.	P=8.	P=10.
$Av(x-\hat{x})^2_{cl}$	.00048	.00026	.00016	.00356
$Av(x-\hat{x})^2_{in}$	.00044	.00024	.00011	.00356
Ratio(cl/in)	1.08911	1.08025	1.50491	1.0001

We see that the average squared error of the inverse method is smaller than the average squared error of the classical method, except for P=.5, where the average squared error of the classical method is smaller. Again, to give more weight to the above statement, we will compute the tolerance intervals of the x-variable for both methods. Using the classical approach, tolerance intervals of the x-variable are obtained by solving the second degree inequality:

$$\underbrace{\left[(\underline{x}'C\underline{y}/\underline{x}'C\underline{y})^2 - \frac{SSE(t^*)^2}{(n-2)\underline{x}'C\underline{x}}\right]}_{A}(x_0-\bar{x})^2 + \underbrace{\left[-2(y-\bar{y})(\underline{x}'C\underline{y}/\underline{x}'C\underline{x})\right]}_{B}(x_0-\bar{x})$$

$$+ \underbrace{(y-\bar{y})^2 - (t^*)^2\frac{SSE}{n-2}\left[1+\frac{1}{n}\right]}_{C} \leq 0 .$$

Let r_1 and r_2 be the roots of this quadratic equation; since $A > 0$ and $B^2 - 4AC > 0$, then we have a situation like in Figure 2.3.1. Using the inverse approach, the tolerance intervals of the x-variable are obtained by

$$\bar{x} + (\underset{\sim}{\chi}'C\chi/\chi'C\chi - t^*\sqrt{(1+\frac{1}{n} + (y_0-\bar{y})^2/\chi'C\chi)MSE_x} < x$$

$$< \bar{x} + \underset{\sim}{\chi}'C\chi/\chi'C\chi + t^*\sqrt{(1+\frac{1}{n} + (y_0-\bar{y})^2/\chi'C\chi)MSE_x} .$$

TABLE 2.4.2 - Tolerance intervals for P', using the classical and the inverse method.

		Classical		Inverse	
A	B	P_L	P_U	P_L	P_U
9080	.5	0.25859	0.67349	0.29885	0.64810
9048.6	.5	0.26579	0.68292	0.30623	0.65731
8177.1	.5	0.49523	0.98373	0.54141	0.95078
8178.7	.5	0.49475	0.98310	0.54091	0.95017
7784.4	1.0	0.62121	0.14891	0.67047	1.11178
7816.0	1.0	0.61045	1.13480	0.65941	1.09807
7422.5	1.0	0.75309	1.32184	0.80541	1.28024
7500.7	1.0	0.72317	1.28261	0.77479	1.24203
6175.0	2.0	1.37102	2.13270	1.43689	2.06904
6081.4	2.0	1.43113	2.21165	1.49824	2.14576
5958.6	2.0	1.51378	2.32023	1.58259	2.25125
5927.5	2.0	1.53543	2.34867	1.60467	2.27888
4770.7	3.0	2.70185	3.88418	2.79270	3.76855
4774.6	3.0	2.61241	3.76623	2.70173	3.65424
4992.5	3.0	2.35818	3.43114	2.44305	3.32932
4966.2	3.0	2.38732	3.46953	2.47271	3.36662
4220.2	4.0	3.41950	4.83198	3.52195	4.68654
4258.0	4.0	3.35562	4.14752	3.45708	4.60477
4193.5	4.0	3.46553	4.89286	3.56869	4.74546
4176.4	4.0	3.49542	4.93239	3.59903	4.78373
3480.1	6.0	5.05278	6.99766	5.17790	6.78100
3498.7	6.0	5.00044	6.92807	5.12490	6.71376
3434.1	6.0	5.18543	7.17405	5.31219	6.95144
3417.7	6.0	5.23385	7.23846	5.36121	7.01367
2983.1	8.0	6.77870	9.29834	6.92325	9.00254
2965.7	8.0	6.85308	9.39775	6.99837	9.09846
2939.3	8.0	6.96812	9.55156	7.11456	9.24687
2941.6	8.0	6.95799	9.53801	7.10433	9.23380
2705.4	10.0	8.11807	10.09162	8.27511	10.73224
2745.5	10.0	7.90263	10.80273	8.05779	10.45568
2532.5	10.0	9.15073	12.47859	9.31608	12.06909
2538.1	10.0	9.11439	12.42971	9.27946	12.02199

We see that the inverse approach gives smaller intervals, and this for each x-variable. Since the average squared error is smaller for the inverse approach (except for P=.5) and since the intervals are smaller using the inverse approach, we conclude that the method due to Krutchkoff (1967) seems quite appropriate in solving the problem of inverse regression. This contradicts, however, the conclusions given by Williams (1969a), Berkson (1969), Halperin (1970) and Martinelle (1970), who claimed that the classical method is better than the inverse method. Since the inverse method gives smaller tolerance intervals, it appears appropriate to reconsider Krutchkoff's method.

CHAPTER III - A New Algorithm to Tolerance Intervals for x given y.

3.1 Application to inverse regression.

Let

$$y = \alpha + \sum_{i=1}^{p} \beta_i x^i + u \ , \ p \geq 1 \ .$$

In the linear situation (p=1), we have considered two methods to compute tolerance intervals for the x-variable: the classical method and Krutchkoff's inverse method. We have concluded that Krutchkoff's inverse method was better because it gives shorter intervals.

Can we use the classical method and the inverse method when p > 1?

Suppose we take p=2; then the model will be

$$(3.1.1) \quad y = \alpha + \beta_1 x + \beta_2 x^2 + u \ .$$

First, let us consider the classical method. If we have n observations, model (3.1.1) becomes

$$(3.1.2) \quad \underset{n\times 1}{\underset{\sim}{Y}} = \alpha \underset{n\times 1}{\underset{\sim}{e}} + \beta_1 \underset{n\times 1}{\underset{\sim}{x}} + \beta_2 \underset{n\times 1}{\underset{\sim}{x}^{(2)}} + \underset{\sim}{u}, \ \underset{\sim}{u} \sim N(\underset{\sim}{0}, \sigma^2 I_n)$$

where $\underset{n\times 1}{\underset{\sim}{e}} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$, $\underset{\sim}{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{pmatrix}$, $\underset{\sim}{x}^{(2)} = \begin{pmatrix} x_1^2 \\ x_2^2 \\ x_3^2 \\ \vdots \\ x_n^2 \end{pmatrix}$, $\underset{\sim}{u} = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ \vdots \\ u_n \end{pmatrix}$.

From equation (3.1.2)

$$(3.1.3)\quad E(\underset{\sim}{Y}) = \alpha\underset{\sim}{\ell} + \beta_1\underset{\sim}{X} + \beta_2\underset{\sim}{X}^{(2)} = \gamma\underset{\sim}{\ell} + \delta_1 C_1\underset{\sim}{X} + \delta_2 C_2 C_1\underset{\sim}{X}^{(2)}$$

where $C_1 = [I - \underset{\sim}{\ell}\underset{\sim}{\ell}'/n]$, $C_2 = [I - C_1\underset{\sim}{X}\underset{\sim}{X}'C_1/\underset{\sim}{X}'C_1\underset{\sim}{X}]$.

Moreover,

$$\alpha\underset{\sim}{\ell}+\beta_1\underset{\sim}{X}+\beta_2\underset{\sim}{X}^{(2)} = \gamma\underset{\sim}{\ell} + \delta_1(I - \underset{\sim}{\ell}\underset{\sim}{\ell}'/n)\underset{\sim}{X} + \delta_2(C_1 - C_1\underset{\sim}{X}\underset{\sim}{X}'C_1/\underset{\sim}{X}'C_1\underset{\sim}{X})\underset{\sim}{X}^{(2)}$$

$$= \gamma\underset{\sim}{\ell} + \delta_1\underset{\sim}{X} - \delta_1\bar{x}\underset{\sim}{\ell} + \delta_2 C_1\underset{\sim}{X}^{(2)} - \delta_2 C_1\underset{\sim}{X}\underset{\sim}{X}'C_1\underset{\sim}{X}^{(2)}/\underset{\sim}{X}'C_1\underset{\sim}{X}$$

$$= \gamma\underset{\sim}{\ell} + \delta_1\underset{\sim}{X} - \delta_1\bar{x}\underset{\sim}{\ell} + \delta_2\underset{\sim}{X}^{(2)} - \delta_2\overline{x^2}\underset{\sim}{e} - \delta_2 f\underset{\sim}{X} + \delta_2 f\bar{x}\underset{\sim}{\ell}$$

where $f = \underset{\sim}{X}'C_1\underset{\sim}{X}^{(2)}/\underset{\sim}{X}'C_1\underset{\sim}{X}$.

Therefore,

$$\alpha = \gamma - \delta_1\bar{x} - \delta_2\overline{x^2} + \delta_2 f\bar{x} ,$$

$$\beta_1 = \delta_1 - f\delta_2 ,$$

$$\beta_2 = \delta_2 ,$$

or $\underset{n\times 1}{E(\underset{\sim}{Y})} = \underset{n\times 3}{X}\ \underset{3\times 1}{\underset{\sim}{\chi}}$

where $\underset{n\times 3}{X} = (\underset{\sim}{\ell},\ C_1\underset{\sim}{X},\ C_2C_1\underset{\sim}{X}^{(2)})$, $\underset{\sim}{\chi} = \begin{pmatrix}\gamma\\ \delta_1\\ \delta_2\end{pmatrix}$.

From the least squares procedure, we obtain

$$X'X\underset{\sim}{\hat{\chi}} = X'\underset{\sim}{Y} .$$

Then provided $(X'X)^{-1}$ exists,

$$\hat{\chi} = (X'X)^{-1}X'\chi .$$

But $(X'X)^{-1} = \begin{pmatrix} 1/n & 0 & 0 \\ 0 & 1/\underset{\sim}{x}'C_1\underset{\sim}{x} & 0 \\ 0 & 0 & 1/\underset{\sim}{x}^{(2)'}C_1C_2C_1\underset{\sim}{x}^{(2)} \end{pmatrix}$.

Then

$$\hat{\chi} = \begin{pmatrix} d \\ e_1 \\ e_2 \end{pmatrix} = \begin{pmatrix} \bar{y} \\ \underset{\sim}{x}'C_1\chi/\underset{\sim}{x}'C_1\underset{\sim}{x} \\ \underset{\sim}{x}^{(2)'}C_2C_1\chi/\underset{\sim}{x}^{(2)'}C_1C_2C_1\underset{\sim}{x}^{(2)} \end{pmatrix}.$$

The regression line becomes

$$(3.1.4)\quad y = a + b_1x + b_2x^2 = (d - e_1\bar{x} - e_2x^2 + e_2f\bar{x}) + (e_1 - fe_2)x + e_2x^2$$

where a, b_1, b_2 are, respectively, the least squares estimates of α, β_1 and β_2 and d, e_1, e_2 are, respectively, the least squares estimates of γ, δ_1 and δ_2.

From equation (2.1.11), tolerance intervals for y are given by

$$(3.1.5)\quad \underset{\sim}{\ell}'\hat{\chi} - t^*\sqrt{(1 + \underset{\sim}{\ell}'(X'X)^{-1}\underset{\sim}{\ell}\ SSE/(n-3)} < y < \underset{\sim}{\ell}'\hat{\chi} + t^*\sqrt{(1 + \underset{\sim}{\ell}'(X'X)^{-1}\underset{\sim}{\ell}\ SSE/(n-3)},$$

where t^* is the critical value of the t-distribution with n-3 degrees of freedom and where

$$\underset{\sim}{\ell}'\hat{\chi} = (1,\ x-\bar{x},\ x^2 - fx + f\bar{x} - \bar{x}^2)\begin{pmatrix} d \\ e_1 \\ e_2 \end{pmatrix},$$

$$SSE = \underset{\sim}{y}'(I - X(X'X)\ X')\underset{\sim}{y} = \underset{\sim}{y}'C_1\underset{\sim}{y} - (\underset{\sim}{\beta}'C_1\underset{\sim}{y})^2/\underset{\sim}{\beta}'C_1\underset{\sim}{\beta}$$

$$- (\underset{\sim}{y}'C_2C_1\underset{\sim}{\beta}^{(2)})^2/\underset{\sim}{\beta}^{(2)}{}'C_1C_2C_1\underset{\sim}{\beta}^{(2)} .$$

Squaring and rearranging equation (3.1.5), we obtain

$$(3.1.6)\quad (y - \underset{\sim}{\ell}'\hat{\underset{\sim}{y}})^2 \le t^{*2}(1 + \underset{\sim}{\ell}'(X'X)^{-1}\underset{\sim}{\ell})SSE/(n-3) .$$

Substituting the values of $\underset{\sim}{\ell}'\underset{\sim}{y}$ and $\underset{\sim}{\ell}'(X'X)^{-1}\underset{\sim}{\ell}$ in (3.1.6), we obtain a quartic inequality

$$(3.1.7)\quad Ax^4 + Bx^3 + Cx^2 + Dx + E \le 0$$

where

$$A = b_2 - [t^{*2}(SSE/(n-3))/(\underset{\sim}{\beta}^{(2)}{}'C_1C_2C_1\underset{\sim}{\beta}^{(2)})] ,$$

$$B = 2b_1b_2 - 2Af ,$$

$$C = b_1^2 - [t^{*2}(SSE/(n-3))/\underset{\sim}{\beta}'C_1\underset{\sim}{\beta}]\bar{x} + A(f^2 + 2f\bar{x} - 2\bar{x}^2) .$$

$$- 2b_2(y-\bar{y}) - 2b_1b_2(f + \bar{x}) ,$$

$$D = - 2[b_1^2 - t^{*2}(SSE/(n-3))/\underset{\sim}{\beta}'C_1\underset{\sim}{\beta}]\bar{x} - 2Af(f\bar{x} - \bar{x}^2)$$

$$- 2b_1(y-\bar{y}) + 2b_2(y-\bar{y}) + 2b_1b_2(2f\bar{x} - \bar{x}^2) ,$$

$$E = [b_1^2 - t^{*2}(SSE/(n-3))/\underset{\sim}{\beta}'C_1\underset{\sim}{\beta}]\bar{x}^2 + A(f\bar{x} - \bar{x}^2)^2 + 2b_1(y-\bar{y})\bar{x}$$

$$- 2b_1b_2(f\bar{x} - \bar{x}^2)\bar{x} + (y - \bar{y})^2 - t^{*2}(SSE/(n-3))(1 + 1/n) .$$

To find an estimate of x, one has to solve the quadratic equation

$$b_2x^2 + b_1x + (a-y) = 0 ;$$

this equation will give two values of x. One has to choose between the two values of x to obtain the better

estimate. In the case of the corticosterone data, a good estimate of x should fall in the range [.5, 10], or nearby.

To find tolerance intervals of x, one has to solve the corresponding quartic equation, and one has to choose between four values. Good limits are those which fall in the range [.5, 10] or will be in the neighbourhood of that range.

For the linear case, the construction of tolerance intervals of x involved a quadratic equation. Here, for the quadratic case, we have to solve a quartic equation to obtain tolerance intervals of x. If we consider the cubic case, one will have to solve an equation of order six to obtain tolerance intervals for x, and if we consider also that the algebraic manipulations willbe heavier and heavier each time, then we can conclude that the classical method is not a good method to apply in situations where p > 1.

Let us consider the inverse method. In the classical situation, the regression of y on x was quadratic. Does this imply that the regression of x on y is quadratic? Suppose we take the following observations. (Chatfield, Statistics for Technology, page 183, example 4):

x	0	1	2	3	4	5	6
y	6.3	5.7	6.3	7.3	9.9	12.5	18.1

We will first consider the classical method

$$E(y) = \alpha + \beta_1 x + \beta_2 x^2 .$$

We can rewrite this model

$$E(\chi) = \alpha \underset{\sim}{e} + \beta_1 \underset{\sim}{x} + \beta_2 \underset{\sim}{x}^{(2)} = a \underset{\sim}{e} + b_1 C_1 \underset{\sim}{x} + b_2 C_2 C_1 \underset{\sim}{x}^{(2)}$$

where

$$C_1 = (I - \underset{\sim}{e}\underset{\sim}{e}'/n) , \quad C_2 = (I - C_1 \underset{\sim}{x}\underset{\sim}{x}' C_1 / \underset{\sim}{x}' C_1 \underset{\sim}{x}) .$$

$$E(\chi) = X\chi = (\underset{\sim}{e}, C_1 \underset{\sim}{x}, C_2 C_1 \underset{\sim}{x}^{(2)}) \begin{pmatrix} a \\ b_1 \\ b_2 \end{pmatrix}.$$

By the least squares method, we obtain

$$\hat{\chi} = (X'X)^{-1} X'\chi = \begin{pmatrix} 1/n & & \\ & 1/\underset{\sim}{x}' C_1 \underset{\sim}{x} & \\ & & 1/\underset{\sim}{x}^{(2)\prime} C_1 C_2 C_1 \underset{\sim}{x}^{(2)} \end{pmatrix} \begin{pmatrix} \underset{\sim}{e}'\chi \\ \underset{\sim}{x}' C \chi \\ \underset{\sim}{x}^{(2)\prime} C_2 C_1 \chi \end{pmatrix}$$

$$= \begin{pmatrix} \bar{y} \\ \underset{\sim}{x}' C_1 \chi / \underset{\sim}{x}' C \underset{\sim}{x} \\ \underset{\sim}{x}^{(2)\prime} C_2 C_1 \chi / \underset{\sim}{x}^{(2)\prime} C_1 C_2 C_1 \underset{\sim}{x}^{(2)} \end{pmatrix} = \begin{pmatrix} 9.44 \\ 1.8786 \\ .5302 \end{pmatrix} .$$

We will test the significance of the parameter b , using a t-distribution with n-3 degrees of freedom, at 99%:

$$t = \frac{b_2}{\hat{\sigma}\sqrt{1/\underset{\sim}{x}^{(2)\prime} C_1 C_2 C_1 \underset{\sim}{x}^{(2)}}}$$

where

$$\hat{\sigma} = \sqrt{SSE/(n-3)} = \sqrt{(\chi'\chi - \chi'X\hat{\chi})/(n-3)} = \sqrt{.65/4} = .39$$

$$t = \frac{.5302}{.39\sqrt{1/84}} = 12.45$$

But $t_{4,.005} = 4.604$, and we reject the hypothesis of of significance of b_2, since $12.45 > 4.604$. Therefore, b_2 has an effect on the regression curve; so the regression curve appears to be at least quadratic.

Now let us consider the inverse method

$$E(\chi) = \delta\mathbf{1} + \gamma_1\chi + \gamma_2\chi^{(2)} = c\mathbf{1} + d_1C_1\chi + d_2C_2C_1\chi^{(2)}$$

where

$$C_1 = (I - \mathbf{1}\mathbf{1}'/n) \ , \quad C_2 = (I - C_1\chi\chi'C_1/\chi'C_1\chi) \ .$$

$$E(x) = Y\hat{\beta}_0 = (\mathbf{1}, C_1\chi, C_2C_1\chi^{(2)}) \begin{pmatrix} c \\ d_1 \\ d_2 \end{pmatrix} .$$

By the least squares method, we obtain

$$\hat{\chi} = (Y'Y)^{-1}Y'\chi = \begin{pmatrix} 1/n & & \\ & 1/\chi'C_1\chi & \\ & & 1/\chi^{(2)}{}'C_1C_2C_1\chi^{(2)} \end{pmatrix} \begin{pmatrix} \mathbf{1}'\chi \\ \chi'C_1\chi \\ \chi^{(2)}{}'C_2C_1\chi \end{pmatrix}$$

$$= \begin{pmatrix} \bar{x} \\ \chi'C_1\chi/\chi'C_1\chi \\ \chi^{(2)}{}'C_2C_1\chi/\chi^{(2)}{}'C_1C_2C_1\chi^{(2)} \end{pmatrix} = \begin{pmatrix} 3. \\ .4283 \\ .0417 \end{pmatrix} .$$

We will test the significance of the parameter d_2

using a t-distribution with n-3 degrees of freedom, at 99%:

$$t = \frac{d_2}{\hat{\sigma}\sqrt{1/\underset{\sim}{Y}^{(2)\prime}C_1C_2C_1\underset{\sim}{Y}^{(2)}}} \sim t_{n-3} ,$$

where

$$\hat{\sigma} = \sqrt{SSE/(n-4)} = \sqrt{(\underset{\sim}{Y}'\underset{\sim}{Y} - \underset{\sim}{X}'\underset{\sim}{Y}\hat{\underset{\sim}{Y}})/(n-3)} = \sqrt{2.96/4} = .8602$$

$$t = -\frac{.0417}{.8602\sqrt{1/1453.30}} = -1.84 .$$

But $-t_{4,.005} = -4.604$ and since $-1.84 > -4.604$ then we accept the hypothesis that d_2 is zero and thus the regression curve is probably not quadratic. Therefore, the quadratic regression of y on x does not necessarily mean a quadratic regression of x on y.

Since we cannot use the classical method and the inverse method, we develop a new method which is quite easy to apply. We will consider the quadratic case to illustrate the new algorithm.

Let

$$\underset{\sim}{Y} = \alpha\underset{\sim}{e} + \beta_1\underset{\sim}{X} + \beta_2\underset{\sim}{X}^{(2)} = \gamma\underset{\sim}{e} + \delta_1C_1\underset{\sim}{X} + \delta_2C_2\underset{\sim}{X}^{(2)} .$$

The least squares estimates d, e_1 and e_2 of γ, δ_1 and δ_2 are:

$$d = \bar{Y} ,$$

$$e_1 = \underset{\sim}{X}'C_1\underset{\sim}{Y}/\underset{\sim}{X}'C_1\underset{\sim}{X},$$

$$e_2 = \underset{\sim}{X}^{(2)\prime}C_2C_1\underset{\sim}{Y}/\underset{\sim}{X}^{(2)\prime}{}_1C_2C_1\underset{\sim}{X}^{(2)} .$$

Then the least squares estimates a, b_1 and b_2 of α, β_1 and β_2 are:

$$a = d - e_1\bar{x} - e_2\bar{x^2} + e_2 f\bar{x} \ , \quad f = \underline{X}'C_1\underline{X}^{(2)}/\underline{X}'C_1\underline{X} \ ,$$

$$b_1 = e_1 - fe_2 \ ,$$

$$b_2 = e_2 \ .$$

Those least squares estimates were derived earlier in this section. The least squares line becomes

$$\hat{y} = a + b_1 x + b_2 x^2 \ .$$

Now let

$$z = b_1 x + b_2 x^2 \ .$$

We may then compute

$$z_i = b_1 x_i + b_2 x_i^2 \ ; \ i = 1, \ldots, n \ .$$

The least squares line is, therefore,

$$\hat{y}_i = a + z_i \ .$$

Now, suppose we consider the following model

$$z_i = \upsilon + \chi y_i + u_i \ ; \ i = 1, \ldots, n \ ,$$

where $y_i = \hat{y}_i$ and $z_i = b_1 x_i + b_2 x_i^2$; the y_i's should not be considered as independent variables, but for the purpose of this inverse algorithm, we will implicity do so.

In vector notation, we have

$$t = \upsilon\underset{\sim}{\beta} + X\underset{\sim}{\chi} + \underset{\sim}{\mu} = \underset{\sim}{\mu}\underset{\sim}{\beta} + X C\underset{\sim}{\chi} + \underset{\sim}{\mu} = \underset{n\times 2}{Y}\underset{\sim}{\beta} + \underset{\sim}{\mu} ,$$

where $\underset{n\times 2}{Y} = (\underset{\sim}{\beta}, C\underset{\sim}{\chi})$, $\underset{\sim}{\beta} = \begin{pmatrix}\mu \\ \chi\end{pmatrix}$.

According to equation (2.1.11), the tolerance limits for z will have the form

$$\underset{\sim}{\ell}'\hat{\underset{\sim}{\beta}} \pm t^{*}\sqrt{1 + \underset{\sim}{\ell}'(Y'Y)^{-1}\underset{\sim}{\ell}\ SSE/(n-2)} .$$

where $\underset{\sim}{\ell}' = (1,\ y-\bar{y})$, $\hat{\underset{\sim}{\beta}} = \begin{pmatrix}\bar{z} \\ \underset{\sim}{z}'C\underset{\sim}{\chi}/\underset{\sim}{\chi}'C\underset{\sim}{\chi}\end{pmatrix}$, $(Y'Y)^{-1} = \begin{pmatrix}1/n & 0 \\ 0 & 1/\underset{\sim}{\chi}'C\underset{\sim}{\chi}\end{pmatrix}$,

$$SSE = [\underset{\sim}{z}'C\underset{\sim}{z} - (\underset{\sim}{z}'C\underset{\sim}{\chi})^{2}/\underset{\sim}{\chi}'C\underset{\sim}{\chi}] ,$$

where t^{*} takes a value of the t-distribution with n-2 degrees of freedom.

Now, let Z_L and Z_U be the lower bound and the upper bound for the tolerance intervals of z.
If we substitute for z the values Z_L and Z_U in equation

$$z = b_1 x + b_2 x^2 ,$$

and if we solve the quadratic equations

$$b_2 x^2 + b_1 x - Z_L = 0 ,$$

$$b_2 x^2 + b_1 x - Z_U = 0 ,$$

we obtain four roots, where two roots provide tolerance limits for x. The two bounds must fall in some restricted

range or must be very close to it. In the case of the corticosterone data, the restricted range is [.5, 10]; we will compute later the tolerance intervals of the corticosterone data and we will see how to choose the limits of the intervals. We can extend this algorithm to values of $p > 2$. One has only to compute the least squares estimates and proceed in the same way as in the quadratic case.

The main advantage of this algorithm is that one does not have to do heavy algebraic manipulations.

3.2 Tests for significance of regression.

Let us reconsider the corticosterone data. Buckland, Blagrave and Lague (1973) thought that the "general standard curve" was cubic, but after testing different sets of data, we suggest that the cubic curve is not "standard" throughout.

We have the following model

$$(3.2.1)\quad \underset{\sim}{Y} = \alpha \underset{\sim}{e} + \sum_{i=1}^{p} \beta_i \underset{\sim}{x}^{(i)} + \underset{\sim}{\mu}\ ,\ \underset{\sim}{\mu} \sim N(\underset{\sim}{0}, \sigma^2 I_n)\ ,\ p \geq 1\ .$$

To test the adequacy of the model, we may use a "lack of fit" test, and to test the presence or the absence of a parameter in the model, we may use the F-test.

Suppose we consider the general linear hypothesis

$$H : L'\underset{\sim}{Y} = \underset{\sim}{m}\ ,$$

where χ is the (p+1)-order vector of parameters of the model; L' is any matrix of r rows and (p+1) columns with rank r; finally, $\underline{m}$ is a vector of order r of specified constants.

We actually have the following from Section 2.1:

$$\underline{y} \sim N(X\chi, \sigma^2 I_n) ,$$

$$\hat{\chi} \sim N(\chi, \sigma^2 (X'X)^{-1}) .$$

Therefore,

$$L'\hat{\chi} - \underline{m} \sim N(L'\chi - \underline{m}, \sigma^2 L'(X'X)^{-1}L) .$$

If

$$q = (L'\hat{\chi} - \underline{m})'[L'(X'X)^{-1}L]^{-1}(L'\hat{\chi} - \underline{m})$$

then

$$q/\sigma^2 \sim \chi^2_r\{(L'\chi - \underline{m})'[L'(X'X)^{-1}L]^{-1}(L'\chi - \underline{m})/\sigma^2\}$$

(Searle (1971), Section 2.5, page 57).

Consider the error sum of squares

$$SSE = \underline{y}'[I - X(X'X)^{-1}X']\underline{y} ;$$

$SSE/\sigma^2 \sim \chi^2_{n-p-1}$, where p+1 is the rank of X (Searle (1971), Section 2.5, page 57). Since q/σ^2 and SSE/σ^2 are distributed independently (Searle (1971), Section 2.5, page 59), then

$$F(H) = \frac{(L'\hat{\chi} - \underline{m})'[L'(X'X)^{-1}L]^{-1}(L'\hat{\chi} - \underline{m})/r}{SSE/(n-p-1)}$$

$$\sim F(r, n-p-1, (L'\chi - \underline{m})'\frac{[L'(X'X)^{-1}L]^{-1}(L'\chi - \underline{m})}{\sigma^2}) .$$

Under the null hypothesis

$$H : L'\chi = \mathbb{B}$$

then

$$F(H) \sim F_{r,n-p-1} .$$

Now suppose we want to test the presence or absence or a parameter in the model

$$\chi = \alpha\underset{\sim}{1} + \sum_{i=1}^{p} \beta_i \underset{\sim}{x}^{(i)} ,$$

that is, we want to test the hpothesis

$$H : \underset{\sim}{\ell}'\chi = \beta_p = 0 ,$$

where

$$\underset{1x(p+1)}{\underset{\sim}{\ell}'} = (0,0,\ldots 1) , \quad \underset{(p+1)x1}{\chi} = \begin{pmatrix} \alpha \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} .$$

Here, $L' = \underset{\sim}{\ell}'$, $r = 1$, $\mathbb{B} = 0$.

Then

$$F(H) = \frac{(\underset{\sim}{\ell}'\hat{\chi})^2 / \underset{\sim}{\ell}'(X'X)^{-1}\underset{\sim}{\ell}}{SSE/(n-p-1)} .$$

Under the null hypothesis, $F(H)$ has the $F_{1,n-p-1}$ distribution.

We compute $F(H)$; if $F(H) > F_{1,n-p-1}$, that is if we have significance, then it is "more likely" that $\underset{\sim}{\ell}'\underset{\sim}{\beta} = \beta_p = 0$, that is, the parameter β_p has some

effect on the model

$$E(\underset{\sim}{y}) = \alpha \underset{\sim}{e} + \sum_{i=1}^{p} \beta_i \underset{\sim}{x}^{(i)} .$$

We could also consider a "lack of fit test". For the corticosterone data, we have

$y_{11}, y_{12}, y_{13}, y_{14}$ 4 repeat observations at x_1,

$y_{21}, y_{22}, y_{23}, y_{24}$ 4 repeat observations at x_2,

$\vdots$

$y_{k1}, y_{k2}, y_{k3}, y_{k4}$ 4 repeat observations at x_k.

Then the contribution to pure error sum of squares from the x readings is

$$\sum_{j=1}^{4} (y_{1j} - \bar{y}_1)^2,$$

where $\bar{y}_1 = (y_{11} + y_{12} + y_{13} + y_{14})/4$, and the sum of squares has 3 degrees of freedom. Similar quantities are evaluated for the other sets of y's. The total SS(pure error), $\sum_{i=1}^{k} \sum_{j=1}^{4} (y_{ij} - \bar{y}_i)^2$, has n-k degrees of freedom.

We can rewrite model (3.2.1) as

$$E(\underset{\sim}{y}) = X\underset{\sim}{\gamma} = X_1 X_0 \underset{\sim}{\gamma} ,$$

where

$$\underset{n\times(p+1)}{X} = \begin{pmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^p \\ 1 & x_2 & x_2^2 & \cdots & x_2^p \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_k & x_k^2 & \cdots & x_k^p \end{pmatrix} \begin{matrix} \\ \\ \\ \\ \end{matrix}$$

$$X_1 = \underset{n\times k}{\begin{pmatrix} \underset{\sim}{e}_1 & \underset{\sim}{0} & \cdots & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{e}_2 & & \vdots \\ \vdots & & \ddots & \underset{\sim}{0} \\ \underset{\sim}{0} & \cdots & & \underset{\sim}{e}_k \end{pmatrix}},$$

$$X_0 = \underset{k\times(p+1)}{\begin{pmatrix} 1 & x_1 & \cdots & x_1^p \\ 1 & x_2 & \cdots & x_2^p \\ \vdots & \vdots & & \vdots \\ 1 & x_k & \cdots & x_k^p \end{pmatrix}}.$$

We can rewrite SS(pure error)

$$SSPE = \chi'(I - X_1(X_1'X_1)^{-1}X_1')\chi .$$

The sum of squares due to "lack of fit" is defined as

$$SSLF = SSE - SSPE$$

$$= \chi'(X_1(X_1'X_1)^{-1}X_1' - X_1X_0(X_0'X_1'X_1X_0)^{-1}X_0'X_1')\chi ,$$

$$SSPE/\sigma^2 \sim \chi^2_{n-k}$$

$$SSLF/\sigma^2 \sim \chi^2_{k-p-1}(\delta^2)$$

(Searle (1971), Section 2.5, page 57), where δ_2, the non-centrality parameter is equal to

$$\delta^2 = [\chi'X'(X_1(X_1'X_1)^{-1}X_1' - X_1X_0(X_0'X_1'X_1X_0)^{-1}X_0'X_1')X\chi]/\sigma^2 .$$

Since SSPE and SSLF are distributed independently (Searle (1971), Section 2.5, page 59), then

$$F(LF) = \frac{SSLF/(k-p-1)}{SSPE/(n-k)} \sim F_{k-p-1,n-k}(\delta_2) .$$

Under the hypothesis $E(\underset{\sim}{Y}) = X\underset{\sim}{\beta}$,

$$F(LF) = \frac{SSLF/(k-p-1)}{SSPE/(n-k)} \sim F_{k-p-1,n-k,\ \text{central}} .$$

Significance of "lack of fit" will indicate inadequacy of the model

$$E(\underset{\sim}{Y}) = \alpha\underset{\sim}{e} + \sum_{i=1}^{p} \beta_i \underset{\sim}{x}^{(i)} .$$

In the computer program we wrote to compute tolerance intervals for the corticosterone levels, there is a part which looks for significance of the regression model. We consider the linear curve up to the quintic curve; a "lack of fit" test and a "significance" test were performed on the five cases. The first curve with no "lack of fit" was selected as the standard curve. If, for some reasons, there is "lack of fit" everywhere, then we will consider the curve to be the standard curve, since Buckland, Blagrave and Lague suggested that the cubic curve was the general standard curve.

3.3 Application to blood hormone cell analysis.

In Section 2.4, we introduced the corticosterone data; a goal of Buckland's research was to find an estimate and tolerance intervals for the corticosterone

level, where the corresponding count per minute was known.

The standard curve was constructed with P (amount of corticosterone) being the controlled variable and A (the count per minute) being the dependent variable. We suppose that the regression model was some polynomial regression model of order p ($p \geq 1$).

We applied the classical method and Krutchkoff's inverse method to corticosterone data when the regression fit was linear ($p=1$). For the case $p>1$, we developed a new algorithm. In this section, we will apply this algorithm to the corticosterone data when the regression is quadratic and when it is cubic.

As mentioned earlier, a computer program was written which studied five cases, that is, the linear case up to the quintic. The "lack of fit" test was used to identify an adequate fit; following a decision on the value of p, an estimate of x and tolerance intervals for x were computed for the corticosterone data.

For the following set of data

P corticosterone level	A count per minute	P corticosterone level	A count per minute
.5	10787.	4.	4388.3
.5	10764.5	4.	4445.3
.5	9591.3	4.	4515.3
1.	9572.2	4.	4563.5
1.	8923.5	6.	3668.3
1.	8976.2	6.	3708.5
1.	8300.0	6.	3567.5
1.	8257.2	6.	3628.6
2.	6835.8	8.	3290.6
2.	6922.8	8.	3297.8
2.	6363.8	8.	3083.2
2.	6440.1	8.	3088.5
3.	5015.0	10.	2939.0
3.	5089.0	10.	2964.4
3.	5080.9	10.	2996.8
3.	5055.3	10.	3024.4

we obtain the following values for the "significance" test and the "lack of fit" test, when the significance level $\alpha = .01$.

Case	F-test (Regression)	$F_{n_1,n_2,\alpha}$	F-test (lack of fit)	$F_{n_1,n_2,\alpha}$
linear	2282.4853	$7.56(n_1=1,n_2=30)$	5.5642	$3.67(n_1=6,n_2=24)$
quadratic	10.3979	$7.60(n_1=1,n_2=29)$	3.6480	$3.90(n_1=5,n_2=24)$
cubic	12.0815	$7.64(n_1=1,n_2=28)$	1.3770	$4.22(n_1=4,n_2=24)$
quartic	0.5545	$7.68(n_1=1,n_2=27)$	1.6380	$4.72(n_1=3,n_2=24)$
quintic	0.1218	$7.72(n_1=1,n_2=26)$	2.3902	$5.61(n_1=2\ n_2=24)$

The first time we meet no "lack of fit" is at the quadratic case since

$$F(LF) = 3.6480 < F_{5,24,.01} = 3.90;$$

then we use the quadratic curve to find the estimate for the corticosterone level and to construct a tolerance interval.

We will apply the algorithm described earlier; since a good fit is quadratic, we should obtain four values for tolerance intervals of x. Good tolerance limits for x are the two values which fall in the range [.5,10] or nearby; the computer program printed only two values of x instead of four values; the reason for this is that we were asking for printed values with format F9.4, i.e., XXXX.XXXX. Since two values were always outside this format, they were never printed out. The same thing happened with the estimate of x.

Table 3.3.1 gives estimates and tolerance intervals for the corticosterone levels when the counts per minute are known.

TABLE 3.3.1 - Estimates and tolerance intervals (for the quadratic case).

A count per minute	$\hat{P}$ estimate of corticosterone data	$P_{\text{Lower Bound}}$	$P_{\text{Upper Bound}}$
9743.2	.64342	.41066	.95498
9794.6	.63212	.40118	.94143
9410.3	.72041	.47516	1.04741
9360.2	.73261	.48536	1.06206
8646.5	.92633	.64700	1.29522
8702.6	.90961	.63308	1.27506
8426.8	.99460	.70377	1.37758
8458.2	.98456	.69543	1.36546
5908.8	2.30428	1.77531	2.97605
6014.7	2.21802	1.70567	2.86973
5478.4	2.70404	2.09644	3.47065
5386.8	2.80094	2.17390	3.59098
4332.9	4.37510	3.41360	5.56924
4258.5	4.53162	3.53508	5.76826
4411.8	4.21773	3.29115	5.36955
4425.2	4.19183	3.27097	5.33673
3554.4	6.54931	4.95619	8.16195
3834.0	5.60977	4.36399	7.14995
3862.7	5.52514	4.29941	7.04083
3092.7	8.75428	6.71236	11.28143
3099.6	8.71666	6.68481	11.23114

For the above unknown corticosterone levels, the estimates $\hat{P}$ (estimate of corticosterone levels) always occur between the bounds of the tolerance intervals. We see also that the algorithm gives quite reasonable values for the estimates and the tolerance intervals.

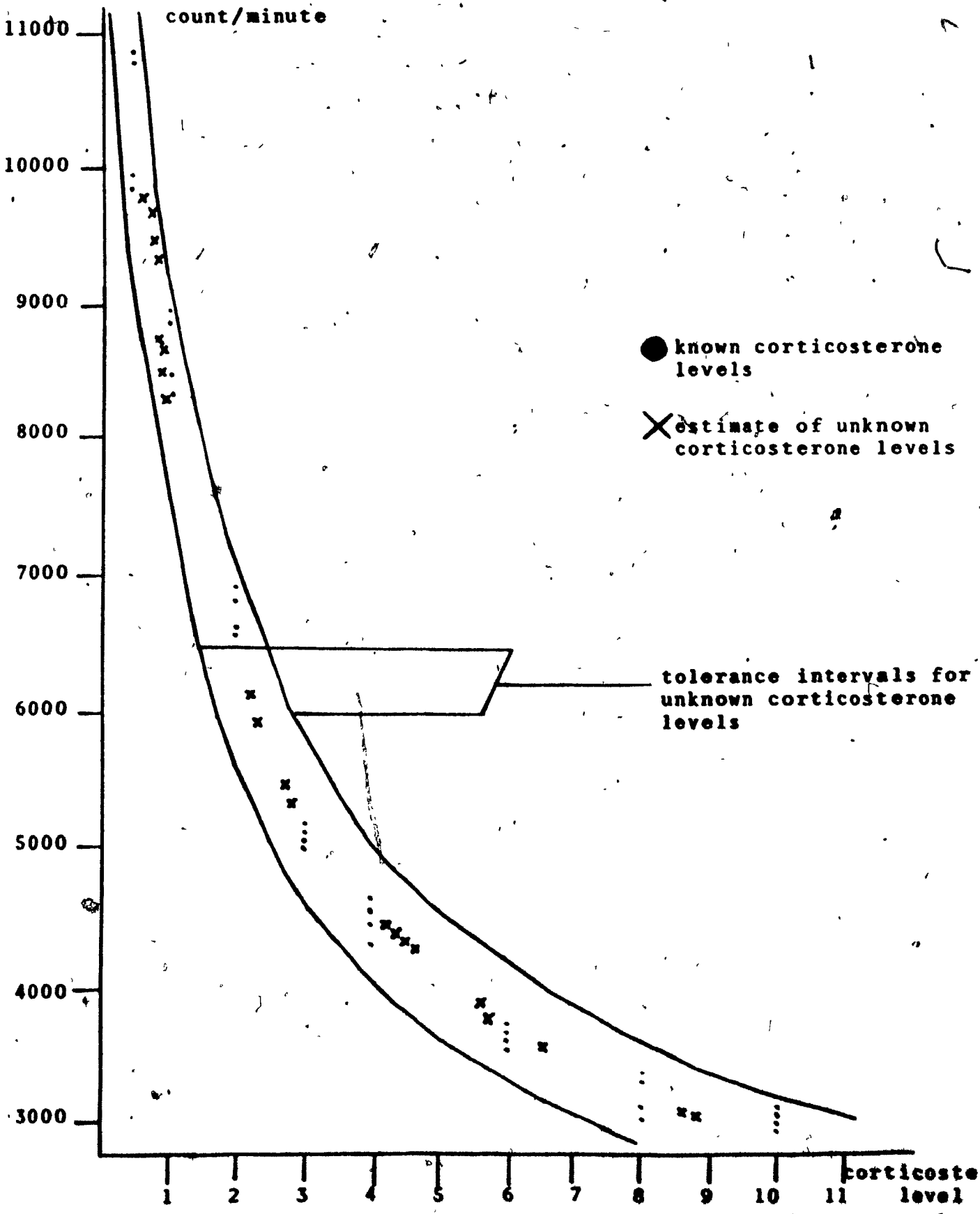


Figure 3.3.1 - Tolerance intervals for unknown corticosterone level.

Now for the following data

P corticosterone level	A count per minute	P corticosterone level	A count per minute
.5	11237.4	4.	5206.2
.5	11465.1	4.	5156.4
.5	11162.2	4.	5226.1
.5	11610.4	4.	5295.5
1.	10412.4	6.	4146.9
1.	10014.0	6.	4094.2
1.	10467.7	6.	4172.7
1.	10106.5	6.	4105.1
2.	7691.9	8.	3614.8
2.	7444.9	8.	3658.5
2.	7807.0	8.	3688.0
2.	7470.5	8.	3598.3
3.	6231.1	10.	3395.5
3.	6239.2	10.	3223.8
3.	6167.5	10.	3350.7
3.	6048.9	10.	3208.6

we obtain the following results for the "significance" test and the "lack of fit" test, when the significance level $\alpha = .01$

Case	F-test (Regression)	$F_{n_1,n_2,\alpha}$	F-test (lack of fit)	$F_{n_1,n_2,\alpha}$
linear	4950.6129	$7.56(n_1=1,n_2=30)$	13.8847	$3.67(n_1=6,n_2=24)$
quadratic	0.5369	$7.60(n_1=1,n_2=29)$	16.2716	$3.90(n_1=5,n_2=24)$
cubic	58.1399	$7.64(n_1=1,n_2=28)$	2.5617	$4.22(n_1=4,n_2=24)$
quartic	7.0346	$7.68(n_1=1,n_2=27)$	1.0561	$4.72(n_1=3,n_2=24)$
quintic	0.7858	$7.72(n_1=1,n_2=26)$	1.1857	$5.61(n_1=2,n_2=24)$

The first time we meet no "lack of fit" is at the cubic case, since

$$F(LF) = 2,5617 < F_{4,24} = 3.90 ;$$

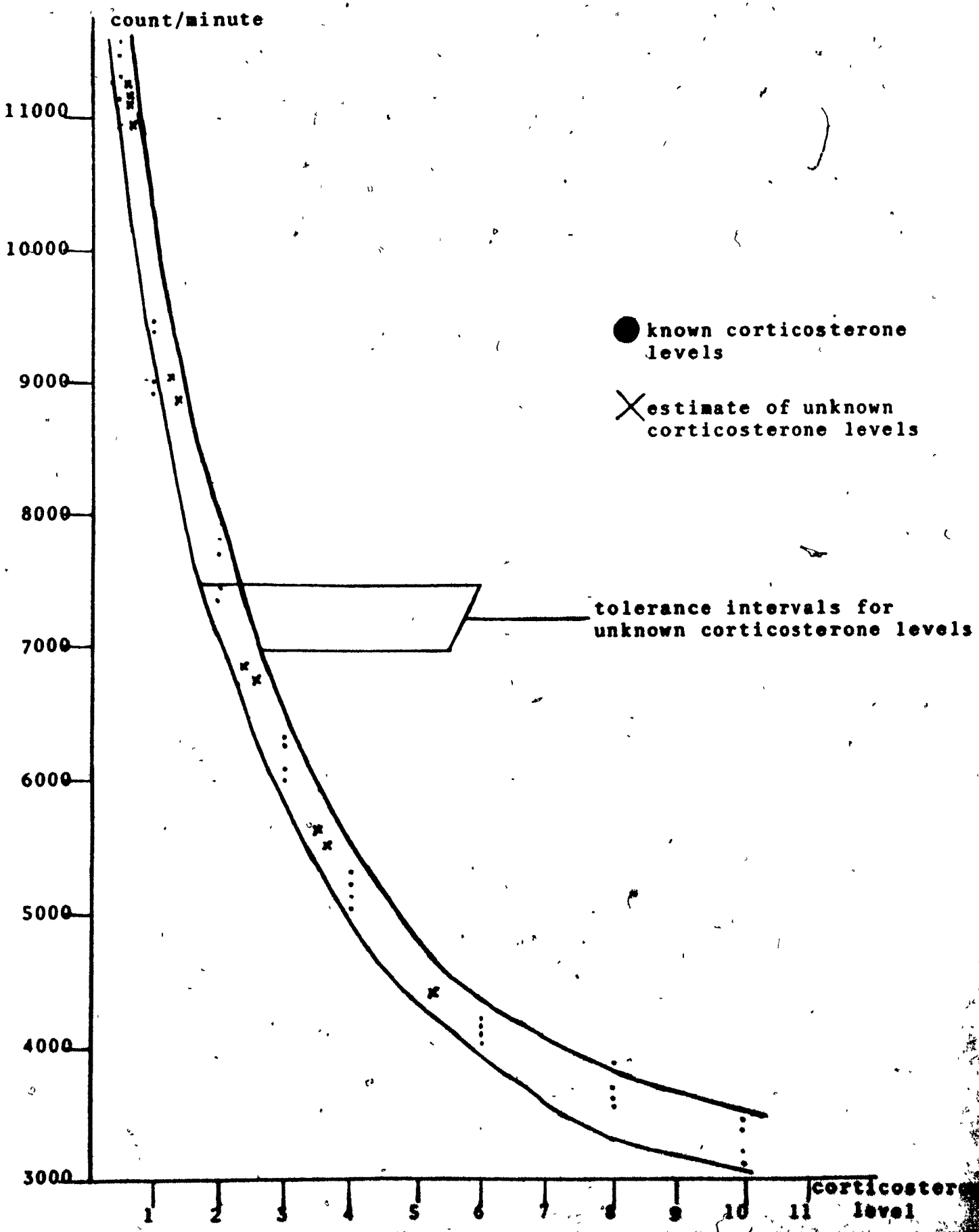
therefore, we use the cubic curve to estimate the corticosterone level and to construct a tolerance interval for it.

According to the algorithm, we will obtain six values for tolerance limits of x. Good limits are the two values which fall in the range [.5,10] or nearby.

Table 3.3.2 gives the six limits for tolerance intervals of x, and three values for the estimate of x.

TABLE 3.3.2 - Estimates and tolerance intervals (for the cubic case).

A count per minute	$\hat{P}$ estimate of corticosterone level	$P_{\text{Lower Bound}}$	$P_{\text{Upper Bound}}$
10995.6	.66951*	.44054*	.88344*
	-.28924	-.35291	-.19654
	51.59455	50.20801	52.92179
11263.4	.57513*	.31823*	.79672*
	-.25481	-.32917	-.13129
	52.16982	50.77980	53.49863
11199.4	.59796*	.34983*	.81739*
	-.26356	-.33506	-.14950
	52.03362	50.64444	53.36203
11232.4	.58622*	.33380*	.80673*
	-.25909	-.33204	-.14039
	52.10394	50.71433	53.43256
9147.9	1.33456*	1.13696*	1.54305*
	-.44518	-.47578	-.41008
	47.18312	45.81489	48.50615
8974.4	1.40424*	1.20549*	1.61455*
	-.45605	-.48505	-.42304
	46.72195	45.35470	48.04542
6979.4	2.41306*	2.17351*	2.67049*
	-.56045	-.57730	-.54211
	40.60062	39.22491	41.94918
6807.4	2.52680*	2.28078*	2.79138*
	-.56822	-.58437	-.55068
	39.98355	38.60434	41.33699
5617.0	3.58149*	3.21636*	3.87308*
	-.61884	-.63094	-.60576
	35.13258	33.70192	36.54500
5613.1	3.53563*	3.22017*	3.87597*
	-.61899	-.63109	-.60593
	35.11465	33.68371	36.52737
4477.1	5.18668*	4.70961*	5.71922*
	-.66391	-.67317	-.65391
	29.01486	27.41654	30.58322


count/minute
11000
10000
9000
8000
7000
6000
5000
4000
3000
known corticosterone levels
estimate of unknown corticosterone levels
tolerance intervals for unknown corticosterone levels
1
2
3
4
5
6
7
8
9
10
11
corticoster
level

For these unknown corticosterone levels, the asterisked values in Table 3.3.2 are respectively the estimates and tolerance limits for the corticosterone levels. The other values were not considered as estimates and limits, since they are "off-range".

For the last two sets of data, the algorithm gives fair values for the estimates and tolerance limits for corticosterone levels.

If we consider that the classical method is almost intractible for polynomial regression and if we consider that Krutchkoff's inverse method cannot be applied directly to polynomial regression, then we are left only with our new method. Also, the results obtained by our new method have proved satisfactory in practice (cf. Buckland, Blagrave and Lague (1973)). Thus we may conclude that the new method is useful when one is dealing with polynomial regression, and when one wants to predict the value of x from an observed value of y for which x is unknown and finally when one wants to construct tolerance intervals for such an x.

REFERENCES AND BIBLIOGRAPHY

Berkson, J., (1950). Are there two regressions? Journal of the American Statistical Association, 45, 164-180.

Berkson, J., (1969). Estimation of a linear function for a calibration line; consideration of a recent proposal. Technometrics, 11, 649-660.

Blackwell, D., (1947). Conditional expectation and unbiased sequential estimation. Annals of Mathematical Statistics, 18, 105.

Buckland, R.B., Blagrave, K., Lague, P.C., (1973). Competitive protein-binding assay for corticoids in the peripheral plasma of the immature chicken. Poultry Science, May 1974, Vol. 53 (in press).

Dixon, W.J., Massey, F.J., (1951). Introduction to Statistical Analysis, Third Edition, Mc Graw-Hill, New-York.

Draper, N., Smith, H., (1966). Applied Regression Analysis, Wiley, New-York.

Ehrenberg, A.S.C., (1963). Bivariate regression analysis is useless. Applied Statistics, Vol. 12, 3, 161-179.

Eisenhart, C., (1939). The interpretation of certain regression methods and their use in biological and industrial research. Annals of Mathematical Statistics, 10, 162-186.

Halperin, M., (1970). On inverse estimation in linear regression. Technometrics, 12, 727-736.

Kendall, M.G., Buckland, W.R., (1957). A Dictionary of Statistical Terms, Third Edition, Oliver and Boyd, Edinburgh.

Krutchkoff, R.G., (1967). Classical and inverse regression methods of calibration. Technometrics, 9, 425-440.

Krutchkoff, R.G., (1968). Letter to the editor, Technometrics, 10, 430-431.

Krutchkoff, R.G., (1969). Classical and inverse regression methods of calibration in extrapolation. Technometrics, 11, 605-608.

Krutchkoff, R.G., (1971). The calibration problem and closeness. Journal of Statistical Computer Simulation, 1, 87-95.

Martinelle, S., (1970). On the choice of regression in linear calibration. Comments on a paper by R.G. Krutchkoff. Technometrics, 12, 157, 161.

Odén, A., (1973). Simultaneous confidence intervals in inverse linear regression. Biometrika, 60, 339-343.

Pitman, E.J.G., (1937). The "closest" estimates of statistical parameters. Proc. Cambridge Philosophical Society, 33, 212-222.

Postelnicu, T., (1973). Hardness and tensile strenght of copper: simple linear regression. International Statistical Review, 41 (2), 253-258.

Scheffé, H., (1959). The Analysis of Variance. Wiley, New-York.

Scheffé, H., (1973). A statistical theory of calibration. Annals of Statistics, 1, 1-38

Searle, S.R., (1971). Linear Models. Wiley, New-York.

Williams, E.J., (1959). Regression Analysis. Wiley, New-York.

Williams, E.J., (1969a). A note on regression methods in calibration. Technometrics, 11, 189-192.

Williams, E.J., (1969b). Regression methods in calibration problems. Proc. 37th Session, Bulletin of the International Statistical Institute, 43 (1), 17-28.

Winsor, C.P.,(1946). Which regression? Biometrics Bulletin, 2, 101, 109.

APPENDIX

This Appendix contains the FORTRAN computer program which was used for the application fo the method described in Section 3.1.

```
      IMPLICIT REAL*8(A-H,P-Z)
      INTEGER O
      DIMENSION X(40),Y(40),XS(40),XC(40),XQ(40),XS(40),XL(40),YL(40)
      DIMENSION XL(72),YL(72),KK(4),K(4),Z(40)
      DIMENSION XCOF(6),COF(6),ROOTR(5),ROOTI(5),ROOTX(5),ROOTY(5)
      DIMENSION TOOT(5)
      READ(5,10)L
10    FORMAT(I2)
      READ(5,11)(K(I),I=1,L)
      READ(5,11)(KK(I),I=1,L)
11    FORMAT(4I3)
      DO 12 M=1,L
      N=K(M)
      J=KK(M)
      READ(5,1)(X(I),Y(I),I=1,N)
      READ(5,1)(XL(I),YL(I),I=1,J)
1     FORMAT(7X,F4.1,4X,F7.1)
      YSUM=0.
      DO 2 I=1,N
      X(I)=DLOG(X(I)+1.D0)
      XS(I)=X(I)*X(I)
      XC(I)=XS(I)*X(I)
      XQ(I)=XC(I)*X(I)
      XS(I)=XQ(I)*X(I)
      Y(I)=DLOG(Y(I))
2     YSUM=YSUM+Y(I)*Y(I)
      DO 20 I=1,J
20    YL(I)=DLOG(YL(I))
      XBAR=0.
      XSBAR=0.
      XCBAR=0.
      XQBAR=0.
      YBAR=0.
      XCX=0.
      XQCXQ=0.
      XQCY=0.
      XCXQ=0.
      XSCXQ=0.
      XCCXQ=0.
      XSCXS=0.
      XCCXC=0.
      XSCXS=0.
      XQCXS=0.
      XCCXS=0.
      XSCXS=0.
      XCXS=0.
      XSCY=0.
      XSBAR=0.
      YCY=0.
      XCY=0.
      XSCY=0.
      XCCY=0.
      XCXS=0.
      XCXC=0.
      XSCXC=0.
      O=1
      ZBAR=0.
      ZCZ=0.
      ZCY=0.
```

```
      TS=(2.750)**2
      CALL ESPE(XBAR,X,N)
      CALL ESPE(XSBAR,XS,N)
      CALL ESPE(XCBAR,XC,N)
      CALL ESPE(XQBAR,XQ,N)
      CALL ESPE(YBAR,Y,N)
      CALL ESPE(XSBAR,XS,N)
      CALL SS1(XCX,XBAR,X,N)
      CALL SS1(XSCXS,XSBAR,XS,N)
      CALL SS1(XCCXC,XCBAR,XC,N)
      CALL SS1(YCY,YBAR,Y,N)
      CALL SS1(XQCXQ,XQBAR,XQ,N)
      CALL SS1(XSCXS,XSBAR,XS,N)
      CALL SS2(XCY,X,XBAR,Y,YBAR,N)
      CALL SS2(XSCY,XS,XSBAR,Y,YBAR,N)
      CALL SS2(XCCY,XC,XCBAR,Y,YBAR,N)
      CALL SS2(XCXS,X,XBAR,XS,XSBAR,N)
      CALL SS2(XCXC,X,XBAR,XC,XCBAR,N)
      CALL SS2(XSCXC,XS,XSBAR,XC,XCBAR,N)
      CALL SS2(XQCY,XQ,XQBAR,Y,YBAR,N)
      CALL SS2(XCXQ,X,XBAR,XQ,XQBAR,N)
      CALL SS2(XSCXQ,XS,XSBAR,XQ,XQBAR,N)
      CALL SS2(XCCXQ,XC,XCBAR,XQ,XQBAR,N)
      CALL SS2(XQCXS,XQ,XQBAR,XS,XSBAR,N)
      CALL SS2(XCCXS,XC,XCBAR,XS,XSBAR,N)
      CALL SS2(XSCXS,XS,XSBAR,XS,XSBAR,N)
      CALL SS2(XCXS,X,XBAR,XS,XSBAR,N)
      CALL SS2(XSCY,XS,XSBAR,Y,YBAR,N)
      BH=XCY/XCX
      ALPHA=YBAR-BH*XBAR
      BETA=BH
      SH=(XCY**2)/XCX
      SE=YCY-SH
      GH=XCXS/XCX
      A1=XSCY-GH*XCY
      A2=XSCXS-GH*XCXS
      DH=A1/A2
      ALPHAS=ALPHA+DH*(GH*XBAR-XSBAR)
      BETAS=BETA-DH*GH
      TETAS=DH
      SHS=A1**2/A2
      SES=SE-SHS
      HK=XCXC/XCX
      HH=(XSCXC-GH*XCXC)/A2
      A3=XCCY-HK*XCY-HH*A1
      A4=XCCXC-(HK**2)*XCX-(HH**2)*A2
      FH=A3/A4
      ALPHAC=ALPHAS+FH*(HH*XSBAR-XCBAR+(HK-HH*GH)*XBAR)
      BETAC=BETAS+FH*(HH*GH-HK)
      TETAC=TETAS-FH*HH
      GAMMAC=FH
      SHC=A3**2/A4
      SEC=SES-SHC
      QH=XCXQ/XCX
      HH=(XSCXQ-QH*XCXS)/A2
      HL=(XCCXQ-QH*XCXC-HH*HK*A2)/A4
      AS=XQCY-(XCXQ*XCY)/XCX-(XCCXQ-(XCXQ*XCXC)/XCX)*(A3/A4)+((XSCXQ-(XC
     &XQ*XCXS)/XCX)+((XSCXC-(XCXS*XCXC)/XCX)*(A3/(A2*A4))-A1/A2))
      AS=XQCXQ-(XCXQ**2)/XCX-((XSCXQ-(XCXQ*XCXS)/XCX)**2)/A2-((XCCXQ-(XC
     &XC*XCXQ)/XCX)-((XSCXC-(XCXC*XCXS)/XCX)*(XSCXQ-(XCXS*XCXQ)/XCX))/A2
```

```
4)**2/A4
 MJ=A5/A6
 ALPHAQ=ALPHAC+MJ*(HL*XCBAR-XQBAR-(HH*HL-HM)*XSBAR-(HK*HL-DH-(HH*HL
&-HM)*GH)*XBAR)
 BETAQ=BETAC+MJ*(HK*HL-DH-(HH*HL-HM)*GH)
 TETAQ=TETAC+MJ*(HH*HL-HM)
 GAMMAQ=GAMMAC-MJ*HL
 OMEGAQ=MJ
 SHQ=A5**2/A6
 SEQ=SEC-SHQ
 YH=XCX5/XCX
 XH=(XSCX5-GH*XCX5)/A2
 VH=(XCCX5-YH*XCXC-XH*HM*A2)/A4
 UH=(XQCX5-DH*XCX5-HH*XH*A2-HL*VH*A4)/A6
 A7=XSCY-(XCX5*XCY)/XCX+(XSCX5-(XCX5*XCX5)/XCX)*((A5/(A2*A6))*(XSCX
&Q-(XCX5*XCXQ)/XCX)-A1/A2)-(A5/A6)*(XQCX5-(XCXQ*XCX5)/XCX)+(XCCX5-(
&XCXC*XCX5)/XCX-(XSCXC-(XCX5*XCXC)/XCX)*(XSCX5-(XCX5*XCX5)/XCX)/A2)
&*((A5/(A4*A6))*(XCCXQ-(XCXC*XCXQ)/XCX-(XSCXC-(XCX5*XCXC)/XCX)*(XSC
&XQ-(XCX5*XCXQ)/XCX)/A2)-(A5/A4))
 A8=XSCX5-(XCX5**2)/XCX-((XSCX5-(XCX5*XCX5)/XCX)**2)/A2-((XCCX5-(XC
&XC*XCX5)/XCX-(XSCXC-(XCXC*XCX5)/XCX)*(XSCX5-(XCX5*XCX5)/XCX)/A2)**
&2)/A4 -((XQCX5-(XCXQ*XCX5)/XCX-(XSCXQ-(XCXQ*XCX5)/XCX)*(XSCX5-(XCX
&5*XCX5)/XCX)/A2-((XCCXQ-(XCXQ*XCXC)/XCX-((XSCXQ-(XCXQ*XCX5)/XCX)*(
&XSCXC-(XCX5*XCXC)/XCX)/A2))*(XCCX5-(XCXC*XCX5)/XCX-((XSCXC-(XCXC*X
&CX5)/XCX)*(XSCX5-(XCX5*XCX5)/XCX)/A2)))/A6)**2)/A8
 TH=A7/A8
 ALPHA5=ALPHAQ+TH*(UH*XQBAR-XSBAR-(UH*DH-(UH*HL-VH)*HK-YH-GH*(UH*HM
&-XH-HH*UH*HL+HH*VH))*XBAR-(UH*HL-VH)*XCBAR-(UH*HM-HH*HL*UH+HH*VH-X
&H)*XSBAR)
 BETA5=BETAQ+TH*(UH*DH-(UH*HL-VH)*HK-YH-GH*(UH*HM-XH-HH*UH*HL+HH*VH
&))
 TETA5=TETAQ+TH*(UH*HM-XH-HH*UH*HL+HH*VH)
 GAMMA5=GAMMAQ+TH*(UH*HL-VH)
 OMEGA5=OMEGAQ-TH*UH
 PHI5=TH
 SH5=A7**2/A8
 SE5=SEQ-SH5
 DENUM=YSUM-(DFLOAT(N)*YBAR**2)
 R2=SH/DENUM
 RS2=RZ+SHS/DENUM
 RC2=RS2+SHC/DENUM
 RQ2=RC2+SHQ/DENUM
 RS2=RQ2+SH5/DENUM
 M=1.
 DF=DFLOAT(N-2)
 CALL GTEST(FTEST,SE,SH,N,DF)
 DFS=DF-1.
 CALL GTEST(FTESTS,SES,SHS,N,DFS)
 DFC=DFS-1.
 CALL GTEST(FTESTC,SEC,SHC,N,DFC)
 DFQ=DFC-1.
 CALL GTEST(FTESTQ,SEQ,SHQ,N,DFQ)
 DF55=DFQ-1.
 CALL GTEST(FTEST5,SE5,SH5,N,DF55)
 SUM=0.
 Y1BAR=0.
 YC1Y=0.
 L1=0
 N1=L1
 M1=N1+2
```

```
      J1=N1
      K1=1
      NN=N+1
  140 DO 110 I=N1,NN
      N1=N1+1
      K1=K1+1
      YIBAR=YIBAR+Y(I-1)
      IF(K1.EQ.NN)GO TO 120
      IF(X(I).NE.X(I-1))GO TO 120
  110 CONTINUE
  120 YIBAR=YIBAR/DFLOAT(N1)
      DO 130 JJ=J1,K1
  130 YC1Y=YC1Y+Y(JJ-1)**2
      J1=K1+1
      SUM=SUM+YC1Y-(DFLOAT(N1)*(YIBAR**2))
      IF(K1.EQ.N)GO TO 160
      N1=K1+1
      N1=0
      YC1Y=0.
      YIBAR=0.
  150 L1=L1+1
      IF(K1.NE.NN)GO TO 140
      FIT=SE-SUM
      FITS=SES-SUM
      FITC=SEC-SUM
      FITQ=SEQ-SUM
      FITS=SES-SUM
      DF1=DFLOAT(L1-2)
      DF2=DFLOAT(N-L1)
      DF3=DF1-1.
      DF4=DF3-1.
      DF5=DF4-1.
      DF6=DF5-1.
      TEST=(FIT/DF1)/(SUM/DF2)
      TESTS=(FITS/DF3)/(SUM/DF2)
      TESTC=(FITC/DF4)/(SUM/DF2)
      TESTQ=(FITQ/DF5)/(SUM/DF2)
      TESTS=(FITS/DF6)/(SUM/DF2)
      F1=FPROB(FTEST,M,DF)
      F2=FPROB(TEST,DF1,DF2)
      F1S=FPROB(FTESTS,M,DFS)
      F2S=FPROB(TESTS,DF3,DF2)
      F1C=FPROB(FTESTC,M,DFC)
      F2C=FPROB(TESTC,DF4,DF2)
      F1Q=FPROB(FTESTQ,M,DFQ)
      F2Q=FPROB(TESTQ,DF5,DF2)
      F1S=FPROB(FTESTS,M,DF55)
      F2S=FPROB(TESTS,DF6,DF2)
      DO 199 I=1,N
      X1(I)=DEXP(X(I))-1.
  199 Y1(I)=DEXP(Y(I))
      WRITE(6,200)
  200 FORMAT('1')
      WRITE(6,201)
  201 FORMAT(' ',39X,'X',8X,'Y')
      WRITE(6,202)(X1(I),Y1(I),I=1,N)
  202 FORMAT(' ',38X,F4.1,3X,F7.1)
      WRITE(6,200)
      WRITE(6,210)
  210 FORMAT('0','LINEAR CASE:')
```

```
      DISC=DSQRT(B**2-4.*A*C)
      ROOT1=-B+DISC/(2.*A)
      ROOT2=-B-DISC/(2.*A)
      UP=ROOT1+XBAR
      DOWN=ROOT2+XBAR
      CENT=(YL(I)-ALPHA)/BETA
      YL(I)=DEXP(YL(I))
  600 WRITE(6,28)XL(I),YL(I),DOWN,UP,CENT
   28 FORMAT(' ',2X,F4.1,3X,F7.1,18X,2F9.5,6X,F9.5)
      GO TO 12
  501 IF(F2S-.99)502,502,503
  502 WRITE(6,701)
  701 FORMAT('0','THE FIRST TIME WE MEET NO L-O-P IS AT THE QUADRATIC CA
     4SE')
      DO 601 I=1,N
  601 Z(I)=BETAS*X(I)+TETAS*XS(I)
      CALL ESPE(ZBAR,Z,N)
      CALL SS1(ZCZ,ZBAR,Z,N)
      CALL SS2(ZCY,Z,ZBAR,Y,YBAR,N)
      SEY=ZCZ-(ZCY**2)/YCY
      SEYM=SEY/DF
      D=2
      DO 602 I=1,J
      RAD=DSQRT((1.+1./DFLOAT(N)+((YL(I)-YBAR)**2)/YCY)*SEYM)
      TERM=ZBAR+(ZCY/YCY)*(YL(I)-YBAR)
      CALL CONF(ROOT1,ROOT2,TERM,T$,RAD)
      XCOF(1)=-ROOT1
      XCOF(2)=BETAS
      XCOF(3)=TETAS
      CALL POLRT(XCOF,COF,D,ROOTR,ROOTI,IER)
      DO 603 II=1,D
      IF(ROOTI(II).NE.0.)ROOTR(II)=0.
      ROOTX(II)=DEXP(ROOTR(II))-1.
  603 CONTINUE
      XCOF(1)=-ROOT2
      XCOF(2)=BETAS
      XCOF(3)=TETAS
      CALL POLRT(XCOF,COF,D,ROOTR,ROOTI,IER)
      DO 604 II=1,D
      IF(ROOTI(II).NE.0.)ROOTR(II)=0.
      TOOT(II)=DEXP(ROOTR(II))-1.
  604 CONTINUE
      XCOF(1)=ALPHAS-YL(I)
      XCOF(2)=BETAS
      XCOF(3)=TETAS
      CALL POLRT(XCOF,COF,D,ROOTR,ROOTI,IER)
      DO 605 II=1,D
      IF(ROOTI(II).NE.0.)ROOTR(II)=0.
      SOOT(II)=DEXP(ROOTR(II))-1.
  605 CONTINUE
      YL(I)=DEXP(YL(I))
      DO 798 II=1,D
      GRAND1=ROOTX(II)
      GRAND2=TOOT(II)
      GRAND3=SOOT(II)
  798 WRITE(6,28)XL(I),YL(I),GRAND1,GRAND2,GRAND3
  602 CONTINUE
      GO TO 12
  503 IF(F2C-.99)504,504,505
  504 WRITE(6,702)
```

```
      WRITE(6,220)ALPHA
  220 FORMAT('0','CONSTANT:',10X,F8.5)
      WRITE(6,230)BETA
  230 FORMAT(' ','COEFFICIENT OF X:',8X,F8.5)
      WRITE(6,240)
  240 FORMAT('0',15X,'DF(DENOM)',9X,'DF(NUM)',8X,'F-TEST(R)',4X,'F-PROB'
     6,6X,'DF(DENOM)'9X,'DF(NUM)',8X,'F-TEST(L-P)',4X,'F-PROB',6X,'R-S')
      WRITE(6,250)DF,W,FTEST,F1,DF2,DF1,TEST,F2,R2
  250 FORMAT('0',18X,F3.0,11X,F2.0,10X,F9.4,3X,F9.4,7X,F3.0,11X,F2.0,10X
     4,F9.4,3X,F9.4,4X,F7.5)
      WRITE(6,260)
  260 FORMAT('-','QUADRATIC CASE:')
      WRITE(6,220)ALPHAS
      WRITE(6,230)BETAS
      WRITE(6,270)TETAS
  270 FORMAT(' ','COEFFICIENT OF XS:',1X,F8.5)
      WRITE(6,240)
      WRITE(6,250)DFS,W,FTESTS,F1S,DF2,DF3,TESTS,F2S,RS2
      WRITE(6,280)
  280 FORMAT('-','CUBIC CASE:')
      WRITE(6,220)ALPHAC
      WRITE(6,230)BETAC
      WRITE(6,270)TETAC
      WRITE(6,290)GAMMAC
  290 FORMAT(' ','COEFFICIENT OF XC:',1X,F8.5)
      WRITE(6,240)
      WRITE(6,250)DFC,W,FTESTC,F1C,DF2,DF4,TESTC,F2C,RC2
      WRITE(6,300)
  300 FORMAT('-','QUARTIC CASE:')
      WRITE(6,220)ALPHAQ
      WRITE(6,230)BETAQ
      WRITE(6,270)TETAQ
      WRITE(6,290)GAMMAQ
      WRITE(6,310)OMEGAQ
  310 FORMAT(' ','COEFFICIENT OF XQ:',1X,F8.5)
      WRITE(6,240)
      WRITE(6,250)DFQ,W,FTESTQ,F1Q,DF2,DF5,TESTQ,F2Q,RQ2
      WRITE(6,330)
  330 FORMAT('-','QUINTIC CASE;')
      WRITE(6,220)ALPHAS
      WRITE(6,230)BETAS
      WRITE(6,270)TETAS
      WRITE(6,290)GAMMAS
      WRITE(6,310)OMEGAS
      WRITE(6,320)PHIS
  320 FORMAT(' ','COEFFICIENT OF XS:',1X,F8.5)
      WRITE(6,240)
      WRITE(6,250)DF55,W,FTESTS,F1S,DF2,DF6,TESTS,F2S,RS2
      WRITE(6,200)
      WRITE(6,27)
   27 FORMAT('0',1X,'ID',8X,'Y',21X,'CONFIDENCE INTERVAL')
      IF(P2-.99)500,501,501
  500 WRITE(6,700)
  700 FORMAT('0','THE FIRST TIME WE MEET NO L-O-P IS AT THE LINEAR CASE'
     4)
      A=(XCY/XCX)**2-(TS*SE/(XCX*DF))
      DO 600 I=1,J
      D=YL(I)-YBAR
      B=-2.*D*XCY/XCX
      C=D**2-(TS*SE*(DF+3.)/(DF*2.))
```

```
  702  FORMAT('0','THE FIRST TIME WE MEET NO L-O-P IS AT THE CUBIC CASE')
       DO 606 I=1,N
  606  Z(I)=BETAC*X(I)+TETAC*XS(I)+GAMMAC*XC(I)
       CALL ESPE(ZBAR,Z,N)
       CALL SS1(ZCZ,ZBAR,Z,N)
       CALL SS2(ZCY,Z,ZBAR,Y,YBAR,N)
       SEY=ZCZ-(ZCY**2)/YCY
       SEYM=SEY/DF
       DO 607 I=1,J
       RAD=DSQRT((1.+1./DFLOAT(N)+((YL(I)-YBAR)**2)/YCY)*SEYM)
       TERM=ZBAR+(ZCY/YCY)*(YL(I)-YBAR)
       CALL CONF(ROOT1,ROOT2,TERM,TS,RAD)
       D=3
       XCOF(1)=-ROOT1
       XCOF(2)=BETAC
       XCOF(3)=TETAC
       XCOF(4)=GAMMAC
       CALL PDLRT(XCOF,COF,D,ROOTR,ROOTI,IER)
       DO 608 II=1,D
       IF(ROOTI(II).NE.0.)ROOTR(II)=0.
       ROOTX(II)=DEXP(ROOTR(II))-1.
  608  CONTINUE
       XCOF(1)=-ROOT2
       XCOF(2)=BETAC
       XCOF(3)=TETAC
       XCOF(4)=GAMMAC
       CALL PDLRT(XCOF,COF,D,ROOTR,ROOTI,IER)
       DO 609 II=1,D
       IF(ROOTI(II).NE.0.)ROOTR(II)=0.
       TOOT(II)=DEXP(ROOTR(II))-1.
  609  CONTINUE
       XCOF(1)=ALPHAC-YL(I)
       XCOF(2)=BETAC
       XCOF(3)=TETAC
       XCOF(4)=GAMMAC
       CALL PDLRT(XCOF,COF,D,ROOTR,ROOTI,IER)
       DO 610 II=1,D
       IF(ROOTI(II).NE.0.)ROOTR(II)=0.
       SOOT(II)=DEXP(ROOTR(II))-1.
  610  CONTINUE
       YL(I)=DEXP(YL(I))
       DO 799 II=1,D
       GRAND1=ROOTX(II)
       GRAND2=TOOT(II)
       GRAND3=SOOT(II)
  799  WRITE(6,28)XL(I),YL(I),GRAND1,GRAND2,GRAND3
  607  CONTINUE
       GO TO 12
  505  IF(F20-.99)506,506,507
  506  WRITE(6,703)
  703  FORMAT('0','THE FIRST TIME WE MEET NO L-O-P IS AT THE QUARTIC CASE
      &')
       DO 611 I=1,N
  611  Z(I)=BETAQ*X(I)+TETAQ*XS(I)+GAMMAQ*XC(I)+OMEGAQ*XQ(I)
       CALL ESPE(ZBAR,Z,N)
       CALL SS1(ZCZ,ZBAR,Z,N)
       CALL SS2(ZCY,Z,ZBAR,Y,YBAR,N)
       SEY=ZCZ-(ZCY**2)/YCY
       SEYM=SEY/DF
       DO 612 I=1,J
```

```
      RAD=DSQRT((1.+1./DFLOAT(N)*((YL(I)-YBAR)**2)/YCV)*SEYN)
      TERM=ZBAR+(ZCV/YCV)*(YL(I)-YBAR)
      CALL CONF(ROOT1,ROOT2,TERM,TS,RAD)
      D=4
      XCOF(1)=-ROOT1
      XCOF(2)=BETAQ
      XCOF(3)=TETAQ
      XCOF(4)=GAMMAQ
      XCOF(5)=OMEGAQ
      CALL POLRT(XCOF,COF,D,ROOTR,ROOTI,IER)
      DO 613 II=1,D
      IF(ROOTI(II).NE.0.)ROOTR(II)=0.
      ROOTX(II)=DEXP(ROOTR(II))-1.
613   CONTINUE
      XCOF(1)=-ROOT2
      XCOF(2)=BETAQ
      XCOF(3)=TETAQ
      XCOF(4)=GAMMAQ
      XCOF(5)=GAMMAQ
      CALL POLRT(XCOF,COF,D,ROOTR,ROOTI,IER)
      DO 614 II=1,D
      IF(ROOTI(II).NE.0.)ROOTR(II)=0.
      TOOT(II)=DEXP(ROOTR(II))-1.
614   CONTINUE
      XCOF(1)=ALPHAQ-YL(I)
      XCOF(2)=BETAQ
      XCOF(3)=TETAQ
      XCOF(4)=GAMMAQ
      XCOF(5)=OMEGAQ
      CALL POLRT(XCOF,COF,D,ROOTR,ROOTI,IER)
      DO 615 II=1,D
      IF(ROOTI(II).NE.0.)ROOTR(II)=0.
      SOOT(II)=DEXP(ROOTR(II))-1.
615   CONTINUE
      YL(I)=DEXP(YL(I))
      DO 800 II=1,D
      GRAND1=ROOTX(II)
      GRAND2=TOOT(II)
      GRAND3=SOOT(II)
800   WRITE(6,28)XL(I),YL(I),GRAND1,GRAND2,GRAND3
612   CONTINUE
      GO TO 12
507   IF(F25-.99)508,508,509
508   WRITE(6,704)
704   FORMAT('0','THE FIRST TIME WE MEET NO L-O-F IS AT THE QUINTIC CASE
     &')
      DO 616 I=1,N
616   Z(I)=BETAS*X(I)+YETAS*XS(I)+GAMMAS*XC(I)+OMEGAS*RQ(I)+PHIS*XS(I)
      CALL ESPE(ZBAR,Z,N)
      CALL SS1(ZCZ,ZBAR,Z,N)
      CALL SS2(ZCV,Z,ZBAR,V,YBAR,N)
      SEY=ZCZ-(ZCV**2)/YCV
      SEYN=SEY/DF
      DO 617 I=1,J
      RAD=DSQRT((1.+1./DFLOAT(N)*((YL(I)-YBAR)**2)/YCV)*SEYN)
      TERM=ZBAR+(ZCV/YCV)*(YL(I)-YBAR)
      CALL CONF(ROOT1,ROOT2,TERM,TS,RAD)
      D=9
      XCOF(1)=-ROOT1
      XCOF(2)=BETAS
```

```
      XCOF(3)=TETAS
      XCOF(4)=GAMMAS
      XCOF(5)=OMEGAS
      XCOF(6)=PHIS
      CALL POLRT(XCOF,COF,O,ROOTR,ROOTI,IER)
      DO 618 II=1,O
      IF(ROOTI(II).NE.O.)ROOTR(II)=O.
      ROOTX(II)=DEXP(ROOTR(II))-1.
618   CONTINUE
      XCOF(1)=-ROOT2
      XCOF(2)=BETAS
      XCOF(3)=TETAS
      XCOF(4)=GAMMAS
      XCOF(5)=GAMMAS
      XCOF(6)=PHIS
      CALL POLRT(XCOF,COF,O,ROOTR,ROOTI,IER)
      DO 619 II=1,O
      IF(ROOTI(II).NE.O.)RCOTR(II)=O.
      TOOT(II)=DEXP(ROOTR(II))-1.
619   CONTINUE
      XCOF(1)=ALPHAS-YL(I)
      XCOF(2)=BETAS
      XCOF(3)=TETAS
      XCOF(4)=GAMMAS
      XCOF(5)=OMEGAS
      XCOF(6)=PHIS
      CALL POLRT(XCOF,COF,O,ROOTR,ROOTI,IER)
      DO 620 II=1,O
      IF(ROOTI(II).NE.O.)RCOTR(II)=O.
      SOOT(II)=DEXP(ROOTR(II))-1.
620   CONTINUE
      YL(I)=DEXP(YL(I))
      DO 601 II=1,O
      GRAND1=ROOTX(II)
      GRAND2=TOOT(II)
      GRAND3=SOOT(II)
601   WRITE(6,28)XL(I),YL(I),GRAND1,GRAND2,GRAND3
617   CONTINUE
      GO TO 12
539   WRITE(6,705)
705   FORMAT('0','WE CANNOT WORK THE CONFIDENCE INTERVALS BECAUSE THERE
     &IS L-O-F EVERYWHERE')
12    CONTINUE
      STOP
      END


      SUBROUTINE ESPE(XMBAR,XM,N)
      IMPLICIT REAL*8(A-H,O-Z)
      DIMENSION XM(1)
      DO 3 I=1,N
3     XMBAR=XMBAR+XM(I)
      XMBAR=XMBAR/DFLOAT(N)
      RETURN
      END


      SUBROUTINE SSI(XMCXM,XMBAR,XM,N)
      IMPLICIT REAL*8(A-H,O-Z)
      DIMENSION XM(1)
```

```
      DO 4 I=1,N
    4 XMCXM=XMCXM+(XM(I)-XMBAR)**2
      RETURN
      END


      SUBROUTINE SS2(XMCXM,XM,XMBAR,YM,YMBAR,N)
      IMPLICIT REAL*8(A-H,O-Z)
      DIMENSION XM(1),YM(1)
      DO 5 I=1,N
    5 XMCXM=XMCXM+(XM(I)-XMBAR)*(YM(I)-YMBAR)
      RETURN
      END


      SUBROUTINE GTEST(TEST,SSE,SSM,N,DF)
      IMPLICIT REAL*8(A-H,O-Z)
      TEST=SSM/(SSE/DF)
      RETURN
      END


      SUBROUTINE CONF(ROOT1,ROOT2,TERM,TS,RAD)
      IMPLICIT REAL*8(A-H,O-Z)
      ROOT1=TERM+DSQRT(TS)*RAD
      ROOT2=TERM-DSQRT(TS)*RAD
      RETURN
      END


      SUBROUTINE POLRT(XCOF,COF,M,ROOTR,ROOTI,IER)                    PLRT 450
      DIMENSION XCOF(1),COF(1),ROOTR(1),ROOTI(1)                      PLRT 460
      DOUBLE PRECISION XO,YO,X,Y,XPR,YPR,UX,UY,V,VT,XT,U,XT2,YT2,SUMSQ, PLRT 470
     1 DX,DY,TEMP,ALPHA                                               PLRT 480
      DOUBLE PRECISION XCOF,COF,ROOTR,ROOTI                           PLRT 500
      IFIT=0                                                          PLRT 680
      N=M                                                             PLRT 690
      IER=0                                                           PLRT 700
      IF(XCOF(N+1))10,25,10                                           PLRT 710
   10 IF(N) 15,15,32                                                  PLRT 720
C                                                                     PLRT 730
C        SET ERROR CODE TO 1                                          PLRT 740
C                                                                     PLRT 750
   15 IER=1                                                           PLRT 760
   20 RETURN                                                          PLRT 770
C                                                                     PLRT 780
C        SET ERROR CODE TO 4                                          PLRT 790
C                                                                     PLRT 800
   25 IER=4                                                           PLRT 810
      GO TO 20                                                        PLRT 820
C                                                                     PLRT 830
C        SET ERROR CODE TO 2                                          PLRT 840
C                                                                     PLRT 850
   30 IER=2                                                           PLRT 860
      GO TO 20                                                        PLRT 870
   32 IF(N-36) 35,35,30                                               PLRT 880
   35 NX=N                                                            PLRT 890
      NXX=N+1                                                         PLRT 900
      N2=1                                                            PLRT 910
      NJ1 = N+1                                                       PLRT 920
```

```
      DO 40 L=1,KJ1                                                    PLRT 930
      NT=KJ1-L+1                                                       PLRT 940
   40 COF(NT)=XCOF(L)                                                  PLRT 950
C                                                                      PLRT 960
C        SET INITIAL VALUES                                            PLRT 970
C                                                                      PLRT 980
   45 XO=.00500101                                                     PLRT 990
      YO=0.01000101                                                    PLRT1000
C                                                                      PLRT1010
C        ZERO INITIAL VALUE COUNTER                                    PLRT1020
C                                                                      PLRT1030
      IN=0                                                             PLRT1040
   50 X=XO                                                             PLRT1050
C                                                                      PLRT1060
C        INCREMENT INITIAL VALUES AND COUNTER                          PLRT1070
C                                                                      PLRT1080
      XO=-10.0*YO                                                      PLRT1090
      YO=-10.0*X                                                       PLRT1100
C                                                                      PLRT1110
C        SET X AND Y TO CURRENT VALUE                                  PLRT1120
C                                                                      PLRT1130
      X=XO                                                             PLRT1140
      Y=YO                                                             PLRT1150
      IN=IN+1                                                          PLRT1160
      GO TO 59                                                         PLRT1170
   55 IFIT=1                                                           PLRT1180
      XPR=X                                                            PLRT1190
      YPR=Y                                                            PLRT1200
C                                                                      PLRT1210
C        EVALUATE POLYNOMIAL AND DERIVATIVES                           PLRT1220
C                                                                      PLRT1230
   59 ICT=0                                                            PLRT1240
   60 UX=0.0                                                           PLRT1250
      UY=0.0                                                           PLRT1260
      V =0.0                                                           PLRT1270
      VT=0.0                                                           PLRT1280
      XT=1.0                                                           PLRT1290
      U=COF(N+1)                                                       PLRT1300
      IF(U) 65,130,65                                                  PLRT1310
   65 DO 70 I=1,N                                                      PLRT1320
      L =N-I+1                                                         PLRT1330
      TEMP=COF(L)                                                      PLRT1340
      XT2=X*XT-Y*YT                                                    PLRT1350
      YT2=X*YT+Y*XT                                                    PLRT1360
      U=U+TEMP*XT2                                                     PLRT1370
      V=V+TEMP*YT2                                                     PLRT1380
      FI=I                                                             PLRT1390
      UX=UX+FI*XT*TEMP                                                 PLRT1400
      UY=UY-FI*YT*TEMP                                                 PLRT1410
      XT=XT2                                                           PLRT1420
   70 YT=YT2                                                           PLRT1430
      SUMSQ=UX*UX+UY*UY                                                PLRT1440
      IF(SUMSQ) 75,110,75                                              PLRT1450
   75 DX=(V*UY-U*UX)/SUMSQ                                             PLRT1460
      X=X+DX                                                           PLRT1470
      DY=-(U*UY+V*UX)/SUMSQ                                            PLRT1480
      Y=Y+DY                                                           PLRT1490
   78 IF(DABS(DY)+DABS(DX)-1.00-05) 100,80,80                          PLRT1500
C                                                                      PLRT1510
C        STEP ITERATION COUNTER                                        PLRT1520
```

```
C                                                                       PLRT1580
   80 ICT=ICT+1                                                         PLRT1940
      IF(ICT-500) 60,85,85                                              PLRT1950
   85 IF(IFIT)100,90,100                                                PLRT1960
   90 IF(IN-3) 50,95,95                                                 PLRT1970
C                                                                       PLRT1980
C        SET ERROR CODE TO 3                                            PLRT1990
C                                                                       PLRT2000
   95 IER=3                                                             PLRT2010
      GO TO 20                                                          PLRT2020
  100 DO 105 L=1,NXX                                                    PLRT2030
      MT=NXX-L+1                                                        PLRT2040
      TEMP=XCOF(MT)                                                     PLRT2050
      XCOF(MT)=COF(L)                                                   PLRT2060
  105 COF(L)=TEMP                                                       PLRT2070
      ITEMP=N                                                           PLRT2080
      N=NX                                                              PLRT2090
      NX=ITEMP                                                          PLRT2100
      IF(IFIT) 120,55,120                                               PLRT2110
  110 IF(IFIT) 115,50,115                                               PLRT2120
  115 X=XPR                                                             PLRT2130
      Y=YPR                                                             PLRT2140
  120 IFIT=0                                                            PLRT2150
  122 IF(DABS(Y)-1.0D-4*DABS(X)) 135,125,125                            PLRT2160
  125 ALPHA=X+X                                                         PLRT2170
      SUMSQ=X*X+Y*Y                                                     PLRT2180
      N=N-2                                                             PLRT2190
      GO TO 140                                                         PLRT2200
  130 X=0.0                                                             PLRT2210
      NX=NX-1                                                           PLRT2220
      NXX=NXX-1                                                         PLRT2230
  135 Y=0.0                                                             PLRT2240
      SUMSQ=0.0                                                         PLRT2250
      ALPHA=X                                                           PLRT2260
      N=N-1                                                             PLRT2270
  140 COF(2)=COF(2)+ALPHA*COF(1)                                        PLRT2280
  145 NN=N                                                              
      IF(NN.LT.2) NN=2                                                  
      DO 150 L=2,NN                                                     
  150 COF(L+1)=COF(L+1)+ALPHA*COF(L)-SUMSQ*COF(L-1)                     PLRT1900
  155 ROOTI(N2)=Y                                                       PLRT1910
      ROOTR(N2)=X                                                       PLRT1920
      N2=N2+1                                                           PLRT1930
      IF(SUMSQ) 160,165,160                                             PLRT1940
  160 Y=-Y                                                              PLRT1950
      SUMSQ=0.0                                                         PLRT1960
      GO TO 155                                                         PLRT1970
  165 IF(N) 20,20,45                                                    PLRT1980
      END                                                               PLRT1990


      FUNCTION FPROB(F,A,B)
      REAL*8 A,B,FPROB,C,D,T,V,N1,N2,E,F
      E=A*F/B
      E=E/(1.+E)
      C=A/2.
      D=B/2.
      T=(C+D-1.)*(1.-E)
      IF(T.GE..8)GO TO 3
      V=T*(3.-E)-(1.-E)**(D-1.)
```

```
      V=(.5*V/D)**0.33333-(1.-1./(9.*D))/SQRT(1./(9.*D))
      PPROB=.5-.5*ERF(.707107*ANGL(V))
      RETURN
      W1=(D*E)**0.33333
      W2=(C*(1.-E))**0.33333
      V=3.*(W1*(1.-1./(9.*D))-W2*(1.-1./(9.*C)))/SQRT(W1**2/D+W2**2/C)
      PPROB=.5+.5*ERF(.707107*ANGL(V))
      RETURN
      END
```