NON-NORMAL ANALYSIS OF VARIANCE AND REGRESSION PROCEDURES BASED ON MODIFIED MAXIMUM LIKELIHOOD ESTIMATORS

By

Sean M. Milosevic-Hill, B.Sc.(Hons.), M.Sc.

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfilment of the Requirements

for the Degree

Doctor of Philosophy

McMaster University

NON-NORMAL ANOVA AND REGRESSION

DOCTOR OF PHILOSOPHY (1995)
(Mathematics)

MCMASTER UNIVERSITY
Hamilton, Ontario

TITLE: Non-normal analysis of variance and regression procedures based on modified maximum likelihood estimators.

AUTHOR: Sean M. Milosevic-Hill, B.Sc. (Hons.), M.Sc.
(McMaster University, Hamilton, Ontario)

SUPERVISOR: Dr. M.L. Tiku, Ph.D., D.Sc. (Aberdeen)
Department of Mathematics and Statistics

NUMBER OF PAGES: x, 167

ABSTRACT

The assumption of normality appears prevalently in well-known statistical procedures. This may be a drawback, since it is very often the case that the data under study is not normally distributed. It is of interest to relax this normality assumption, and extend many commonly used statistical procedures to include non-normal situations.

With this in mind, we have proposed non-normal Regression and Analysis of Variance schemes based on Tiku and Suresh's (1992) Modified Maximum Likelihood procedure, for a symmetric non-normal family of distributions. The results have been derived both for complete and censored samples. The resulting non-normal procedures are exactly similar in form to the classical results, and are no more difficult to implement. They are also asymptotically fully efficient. Simulation studies have shown that the new methods are extremely efficient, even for distributions far from normal, and for small samples.

It is hoped that these new techniques present viable alternatives for data analysis when the normality assumption may not be justified.

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my supervisor, Dr. M.L. Tiku. His knowledge, guidance and encouragement was greatly appreciated throughout the course of my studies. I am also grateful for the monetary support he provided for me through the department.

I would also like to thank the members of my supervisory committee, Dr. H. Heinig and Dr. A. El-Shaarawi, for their support and assistance. A special thanks to Dr. D.C. Vaughan of Wilfrid Laurier University, for his assistance, encouragement, and monetary support. Thanks also to Dr. R. Viveros for his assistance in times of need.

A very special and heartfelt thanks goes out to my wife, Vicky. Her unbounded love, energy and encouragement of me throughout the course of my studies really kept me going. She helped me realize my dreams, and together we will help realize hers.

I would also like to send my thanks out to my parents and brother Chris for their support and encouragement, throughout all my post-secondary studies. A special thanks also to Vicky's parents, for all the help they have also given throughout the past few years.

Finally, I would like to thank all my friends in the Department of Mathematics and Statistics: Al Tyson, Tarit Saha, Ben Chan, Rob Stamicar, Franco Guerriero, and especially Michael Pfau and Spiro Daoussis, for making my time at McMaster most enjoyable and very interesting.

TABLE OF CONTENTS

Page No.

Page No.

Chapter 3: Non-Normal Analysis of Variance

Chapter 4: Non-Normal Regression

Page No.

LIST OF TABLES AND FIGURES

Page No.

Chapter 1

1.1 Introduction: Normality in Statistics

Of all the concepts and tools used in statistical theory, none is unilaterally more important and pervasive than the Normal distribution. In fact, many of the statistical procedures and results that the majority of scientists and researchers from other fields (other than statistics) are exposed to during their course of study, and utilize in their research and statistical software packages, are based on the normality assumption. Consider this whimsically-shaped quotation, attributed to the statistician W.J. Youden:

THE
NORMAL
LAW OF ERROR
STANDS OUT IN THE
EXPERIENCE OF MANKIND
AS ONE OF THE BROADEST
GENERALIZATIONS OF NATURAL
PHILOSOPHY ♦ IT SERVES AS THE
GUIDING INSTRUMENT IN RESEARCHES
IN THE PHYSICAL AND SOCIAL SCIENCES AND
IN MEDICINE, AGRICULTURE, AND ENGINEERING ♦
IT IS AN INDISPENSABLE TOOL FOR THE ANALYSIS AND THE
INTERPRETATION OF THE BASIC DATA OBTAINED BY OBSERVATION AND EXPERIMENT

\- W.J. Youden

From this quotation it seems clear that the concept of normality is of utmost importance to modern statistical theory, and in fact to many is the be-all and end-all of statistics, and its relevance to scientific research. In contrast, the view that:

"NORMALITY IS A MYTH; THERE NEVER WAS,
AND NEVER WILL BE, A NORMAL DISTRIBUTION"

- R.C. Geary (Biometrika, 1947)

illustrates that normality is not essential. In fact, in many situations, assuming normality of the data set would be detrimental to the calculation, evaluation, and conclusions. The assumption of normality for every data set is a common over-generalization seen in many research papers. This generalization may lead to incorrect interpretations by the researcher.

It is this disparity regarding the normality assumption for all data that we wish to address in this dissertation. We will begin by presenting a brief discussion of how the normality assumption appears prevalently in commonly used statistical procedures, such as *Regression* and *Analysis of Variance* (ANOVA). Following this, a brief discussion of some "Robust" procedures, developed to deal with departures from normality in experimental situations, will be presented. Finally, we will give a brief introduction to Modified Maximum Likelihood (MML) estimation. MML estimation is a modification of the classical maximum likelihood procedure. [For a complete account of MML estimation, the reader is referred to the volume Robust Inference by Tiku, Tan, and Balakrishnan (1989), and, importantly, the paper by Tiku and Suresh (1992)].

Following this introduction, the major results of this thesis will be presented. Specifically, the Modified Maximum Likelihood method of Tiku (1967, 1967a, 1968), with the form extended to symmetric non-normal families of location-scale distributions (Tiku and Suresh, 1992) will be examined. This method will be used to create ANOVA and Regression procedures that do not rely on the traditional normality assumption. Instead, this format will require one only to assume a symmetric non-normal family of distributions (indexed by a shape parameter p). It will be shown that these new non-normal procedures are asymptotically fully efficient; for small samples, they are almost 100% efficient. Further, it will be seen that these new procedures are no more complicated to implement than the usual (classical) procedures. Thus the new procedures present a viable alternative for the researcher who needs to analyze data in less than ideal settings.

Looking at the majority of statistical procedures that students learn in the average Introductory Statistics course, it is evident that these procedures are based on the following (often hidden) assumptions:

(1) The data points under investigation are independent and identically distributed.

(2) Further, the data is usually assumed to be <u>Normally Distributed</u>.

Point (1) is not particularly worrisome, since it can often be made to be true by setting up and running the experiment properly (eg. by completely randomizing the subjects, etc.). Unfortunately for the practitioner of statistics, point (2) is very often <u>not</u> true. Thus blindly following statistical procedures without understanding the underlying assumptions may result in misleading or incorrect inferences from the statistical analysis. Based on this point, it is evident that it is desirable to extend or modify classical statistical procedures based on normality to include non-normal situations, and even to create entirely new approaches not related to the classical procedures. The need for this is clearly illustrated in the next section.

1.2 Some Common Procedures and Their Assumptions

Statistical literature is filled with examples of well known and commonly used procedures that include normality as a tacit assumption. The following are just a few, taken from Introductory Statistics by Wonnacott and Wonnacott (1977).

(1) Confidence Intervals:

Say one wishes to determine a $100(1-\alpha)\%$ confidence interval for a location parameter μ, based on a sample $x_1, x_2, \ldots, x_n$. The appropriate (small sample) confidence interval is given by:

$$C.I. = \left(\bar{x} - t_{\alpha/2}(n-1)\frac{s}{\sqrt{n}},\ \ \bar{x} + t_{\alpha/2}(n-1)\frac{s}{\sqrt{n}} \right) \qquad (1.2.1)$$

where $\bar{x}$ and s are, respectively, the usual sample mean and standard deviation, and $t_{\alpha/2}(n-1)$ is the $100(1-\alpha)\%$ point of a Student's-t distribution with n-1 degrees of freedom. In this case the confidence interval **(1.2.1)** follows from the fact that the statistic $t = \sqrt{n}\,(\bar{x}-\mu)/s$ has a Student's-t distribution with n-1 degrees of freedom, only if the original sample is Normally Distributed. Similar confidence intervals for scale parameters (σ^2), etc. are also based on the assumption of normality.

(2) Classical t-test:

Consider two samples $x_1, x_2, \ldots, x_{n_1}$ and $y_1, y_2, \ldots, y_{n_2}$. Based on these samples, it is of interest to test the hypothesis:

$$\begin{aligned} &H_0: \mu_x = \mu_y \\ \text{vs.}\quad &H_1: \mu_x \neq \mu_y. \end{aligned}$$

The appropriate test statistic is:

$$t = \frac{\bar{x} - \bar{y}}{s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

$$\text{where}\qquad s_p^2 = \frac{(n_1-1)\,s_x^2 + (n_2-1)\,s_y^2}{(n_1+n_2-2)}$$

is the 'pooled' sample variance, and $\bar{x}$, $\bar{y}$, s_x^2, and s_y^2 are the respective sample means and variances. Once again, (assuming of course $\sigma_x^2 = \sigma_y^2$) the above test statistic t has a Student's-t distribution with n_1+n_2-2 degrees of freedom only if the x_i's and y_i's are Normally Distributed.

(3) Analysis of Variance (ANOVA), Regression:

Consider the following k samples:

$$\begin{array}{cccc} x_{11} & x_{21} & \cdots & x_{k1} \\ x_{12} & x_{22} & \cdots & x_{k2} \\ \vdots & \vdots & & \vdots \\ x_{1n_1} & x_{2n_2} & \cdots & x_{kn_k} \\ \hline \bar{x}_1 & \bar{x}_2 & \cdots & \bar{x}_k \end{array}$$

where we wish to test:

$$\begin{aligned} &H_0: \mu_1 = \mu_2 = \cdots = \mu_k \\ \text{vs.}\quad &H_1: \text{at least one } \mu_i \text{ is different.} \end{aligned}$$

This is an extension of the t-test to more than two populations, and is carried out using a 1-way ANOVA model:

$$x_{ij} = \mu + \tau_i + \varepsilon_{ij} \qquad (i = 1,\ldots k;\ j = 1,\ldots,n_i)$$

in which the τ_i's represent the effect of the i^{th} group. Estimates of the model parameters, and inferences based on these estimates usually assume that the errors ε_{ij}'s are independent and identically distributed as Normal with mean 0 and common variance σ^2.

This normal-error assumption also appears in the evaluation of the simple linear regression model:

$$y_i = \mu + \beta x_i + \varepsilon_i \qquad (i = 1,\ldots,n)$$

where μ and β are the intercept and slope of a fitted line where the sample $x_1, x_2, \ldots, x_n$ represents an independent variable, and $y_1, y_2, \ldots, y_n$ represents a dependent (or response) variable.

There are many more examples of this nature. All involve procedures which are widely used by (non-statistician) scientists and researchers. Often the implied normality assumption is never even checked; clearly, this could lead to inferences that are misleading (eg. show effects which are not there or mask effects which are there), or completely wrong, and hence useless. For this reason, many alternate schemes have been implemented in an attempt to render the normality assumption less crucial.

1.3 Minimizing Use of the Normality Assumption

(1) Apply the Central Limit Theorem:

If $x_1, x_2, \ldots, x_n$ is a sample of size n from any distribution having finite mean μ and variance σ^2, then

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \sim N(\mu, \frac{\sigma^2}{n}) \quad \text{for } n \text{ large.}$$

Basically, this means that the normality assumption can be treated as true, providing one has a large enough sample. Unfortunately, in most experimental situations one does not have the luxury of obtaining very large samples. In fact it is often the case that obtaining data is very expensive, and one would want to minimize the amount of data needed to reach a reliable conclusion.

(2) Non-Parametric Methods:

These are procedures which base computation and inference on the ranks of the ordered observations, rather than the values of the observations, and require no distributional assumptions on the data. Examples include the Rank Test, Spearman's Rank correlation coefficient, etc. It may be argued, however, that the data under study follows (or at least approximates) some distribution; therefore there may be some information inherent in the data that non-parametric procedures do not take into consideration.

(3) Employ a ROBUST Procedure:

Many procedures have been implemented which attempt to relax the normality assumption. These procedures are called "Robust"; in this context Robust refers to any procedure that is relatively insensitive to small deviations from the assumptions (Huber, 1981, p. 1). Many such procedures have been proposed and investigated. According to Huber (1981), a robust procedure should:

(a) "have a reasonably good (optimal or nearly optimal) efficiency at the assumed model.

(b) be robust in the sense that small deviations from the model assumptions should impair the performance only slightly ... say in terms of the asymptotic variance of an estimate or the level and power of a test.

(c) somewhat larger deviations from the model should not cause a catastrophe."

(Huber, 1981, p. 5)

The paper "An Introduction to Robust Estimation" by R.V. Hogg (in Launer & Wilkinson, 1979) presents a brief discussion of some of the more commonly used robust methods. For the purposes of our exposition a few of them will be highlighted here.

(1) Trimmed (Windsorized) Mean:

From an ordered sample of size n, censor (remove) r observations from the left and right, then compute:

$$\bar{x}_T = \frac{1}{n-2r} \sum_{i=r+1}^{n-r} x_{(i)} \ .$$

'Bad' data points (or 'outliers') are indicative of either measurement errors, or an underlying distribution with heavier tails than the normal (Hogg, 1979, p. 1, as cited in Launer & Wilkinson, 1979). The trimmed mean seeks to remove the influence of these bad data points by eliminating them from computation. However, depending on the choice of r, 'good' data points may be removed along with the 'bad' ones.

(2) Huber (or M-) Estimators:

In usual practice, to find the least-squares estimate of a parameter (eg. normal mean), one minimizes:

$$\sum_{i=1}^{n} (x_i - \mu)^2$$

$$\Rightarrow 2\sum_{i=1}^{n} (x_i - \mu) = 0$$

$$\Rightarrow \hat{\mu} = \bar{x} .$$

If we define $z_i = x_i - u$, then this is equivalent to minimizing $\sum z_i^2$, i.e.setting $\sum z_i = 0$. Huber calls these two functions of z_i as $\rho(z)$ and $\psi(z)$; thus in this case $\rho(z) = z^2$ and $\psi(z) = z$. Huber estimation entails replacing these functions with others that give less weight in the tails of the distribution, for example:

$$\rho(z) = \begin{cases} z^2 & |z| \leq a \\ a|z| - \frac{a^2}{2} & |z| > a \end{cases}$$

and $$\psi(z) = \begin{cases} -a & z < -a \\ z & |z| \leq a \\ a & z > a \end{cases}$$

for some choice of a. In practice, it is not clear what an appropriate choice should be for a. Further, for some choices of a, the resulting equations may have to be solved iteratively.

(3) Best Linear Unbiased Estimation (BLUE's):

To obtain estimates of μ and σ, find the best coefficients a_i and b_i (in the sense that the results will be unbiased and have minimum variance) and compute:

$$\mu^* = a_1 x_{(1)} + a_2 x_{(2)} + \cdots + a_n x_{(n)}$$

$$\text{and} \quad \sigma^* = b_1 x_{(1)} + b_2 x_{(2)} + \cdots + b_n x_{(n)} .$$

a_i and b_i are called the BLUE coefficients, and are computed using the means and variances of order statistics. Unfortunately, these coefficients are different (and must be computed separately) for each sample size under consideration.

Many of these robust procedures are more complex than the traditional procedures, so it would seem that there would be some resistance to adopt them by the researcher who uses statistical procedures regularly. For this reason, it is of interest to come up with methods that closely resemble the classical procedures. It this way, it is hoped that these more powerful techniques will make their way into the mainstream of statistical practice.

1.4 MML Estimation For Normal Populations

Modified Maximum Likelihood estimators (MMLE's) were introduced by Tiku (1967). The likelihood function is approximated in such a way as to facilitate analytic solutions of the likelihood equations. Here, we outline the derivation of the MML estimators for the normal distribution, for the special case of symmetric censoring. For a more complete discussion, see Sections 2.6 and 2.7 in Tiku, Tan, and Balakrishnan (1986).

Consider the Type-II symmetrically censored sample:

$$x_{(r+1)} \le x_{(r+2)} \le \cdots \le x_{(n-r)} \tag{1.3.1}$$

(where r denotes the number of observations censored at the left and right) from a distribution of the form $(1/\sigma)f[(x-\mu)/\sigma]$ where μ and σ are, respectively, the location and scale parameters. Then the likelihood function based on (1.3.1) is given by:

$$L \propto \sigma^{-(n-2r)} \left\{ \prod_{i=r+1}^{n-r} f(z_{(i)}) \right\} [F(z_{(r+1)})]^{r} [1 - F(z_{(n-r)})]^{r}$$

where $F(z) = \int_{-\infty}^{z} f(z)\, dz, \quad z = \dfrac{x-\mu}{\sigma}, \quad z_{(i)} = \dfrac{x_{(i)} - \mu}{\sigma}.$

The regular MLE's are known to be the solutions of $\partial \ln L/\partial\mu = 0$ and $\partial \ln L/\partial\sigma = 0$. In the case of the normal distribution, we have:

$$\frac{\partial \ln L}{\partial \mu} = \frac{n}{\sigma} \left\{ \frac{1}{n} \sum_{i=r+1}^{n-r} z_{(i)} - q\,[\,g_1(z_{(r+1)}) + g_2(z_{(n-r)})\,] \right\} = 0 \tag{1.3.2}$$

$$\frac{\partial \ln L}{\partial \sigma} = \frac{n}{\sigma} \left\{ -(1-2q) + \frac{1}{n} \sum_{i=r+1}^{n-r} z_{(i)}^2 - q\,[\,z_{(r+1)}\, g_1(z_{(r+1)}) + z_{(n-r)}\, g_2(z_{(n-r)})\,] \right\} = 0 \tag{1.3.3}$$

where $q = r/n$, $g_1(z) = f(z)/F(z)$, and $g_2(z) = f(z)/[1-F(z)]$. Equations (1.3.2) and (1.3.3) do not admit explicit solutions and are usually solved by iterative methods.

To obtain explicit solutions, Tiku (1967) replaces the likelihood function L by a modified likelihood function L^*, in the following manner:

Consider the linear approximations

$$g_1(z_{(r+1)}) \approx \alpha_1 - \beta_1 z_{(r+1)} \quad \text{and} \quad g_2(z_{(n-r)}) \approx \alpha_2 + \beta_2 z_{(n-r)}$$

$$\text{where} \quad \beta_1 = -\frac{g_1(k_1) - g_1(h_1)}{k_1 - h_1}; \quad \alpha_1 = g_1(h_1) + h_1 \beta_1$$

$$\text{and} \quad \beta_2 = \frac{g_2(h_2) - g_1(h_1)}{k_2 - h_2}; \quad \alpha_2 = g_2(h_2) + h_2 \beta_2 .$$

In the case of symmetric censoring, $\beta_1 = \beta_2 = \beta$ and $\alpha_1 = \alpha_2 = \alpha$, ie.

$$\beta = -\frac{g_1(k) - g_1(h)}{k - h}; \quad \alpha = g_1(h) + h\beta$$

which implies

$$g_1(z_{(r+1)}) \approx \alpha - \beta z_{(r+1)} \quad \text{and} \quad g_2(z_{(n-r)}) \approx \alpha + \beta z_{(n-r)} .$$

The values h and k are chosen so that the interval (h, k) contains $(t, z_{(r+1)})$, where $t = E(z_{(r+1)}) = -E(z_{(n-r)})$. We compute h and k from:

$$1 - F(h) = q + \sqrt{\frac{q(1-q)}{n}}$$

$$\text{and} \quad 1 - F(k) = q - \sqrt{\frac{q(1-q)}{n}} .$$

As n tends to ∞, h and k tend to t; hence we may compute

$$\beta = -f(t)\,[\,t - f(t)/q]/q$$

$$\alpha = [f(t)/q] - \beta t$$

for n sufficiently large [$n \geq 10$ is shown to be sufficient; see Tiku (1980)].

Using these linear approximations, we obtain the MML equations:

$$\frac{\partial \ln L^*}{\partial \mu} = \frac{1}{\sigma}\left\{\sum_{i=r+1}^{n-r} z_{(i)} + r\beta\,(z_{(r+1)} + z_{(n-r)})\right\}$$

$$= (m/\sigma^2)(K - \mu) = 0$$

$$\frac{\partial \ln L^*}{\partial \sigma} = \frac{1}{\sigma}\left\{-(n-2r) + \sum_{i=r+1}^{n-r} z_{(i)}^2 + r\alpha\,(z_{(n-r)} - z_{(r+1)}) + r\beta\,(z_{(r+1)}^2 + z_{(n-r)}^2)\right\}$$

$$= -\sigma^{-3}[(A\sigma^2 - B - C) - m(K - \mu)^2] = 0$$

where

$$A = n - 2r \quad , \quad m = A + 2r\beta$$

$$K = \frac{1}{m}\left\{\sum_{i=r+1}^{n-r} z_{(i)} + r\beta\,(z_{(r+1)} + z_{(n-r)})\right\}$$

$$B = r\alpha\,(z_{(n-r)} - z_{(r+1)})$$

$$C = \sum_{i=r+1}^{n-r} z_{(i)}^2 + r\beta\,(z_{(r+1)}^2 + z_{(n-r)}^2) - mK^2 .$$

The MML estimators of μ and σ are then:

$$\hat{\mu} = K$$

and $\hat{\sigma} = \dfrac{B + \sqrt{B^2 + 4AC}}{2\sqrt{A(A-1)}}$, corrected for bias.

As pointed out in Tiku et al. (1986), $\hat{\mu}$ and $\hat{\sigma}$ are a robust pair of estimators of the location and scale parameters, and can be used to compute more efficient estimates of these parameters under non-normal situations.

1.5 Main Thesis Aims: Non-Normal Estimation Procedures

All of the procedures that have been discussed so far, including the MML procedure, first base their procedure on the normal distribution, and then show that they still perform well, even when the true distribution deviates from normal (ie. they are robust with respect to departures from normality). Alternatively, one could dispense with the normality assumption entirely, and come up with an estimation procedure which is valid for, say, a large family of symmetric location-scale distributions.

This is what will be presented in this dissertation. In Chapter 2, we will introduce the non-normal MML procedure of Tiku and Suresh (1992) based on a symmetric t-family of distributions. We will discuss its properties and efficiency, and will present results regarding estimation for small values of the shape parameter, as discussed by Vaughan (1992). In Chapter 3, this method will be extended to one and two-way Analysis of Variance models, both for complete and censored samples. It will be seen that the resulting analyses are exactly similar to the classical results, and are no more difficult to implement.

In Chapter 4, the same procedure will be used to come up with non-normal simple linear and multiple linear regression models (for complete and censored samples). Once again, these new procedures will mirror the classical ones directly. Finally, Chapter 5 will deal with the MML estimation procedure in the bivariate regression setting. Throughout the exposition, it will be shown (through use of simulations), that these new methods are highly efficient, even when the underlying distribution is far from normal.

Chapter 2

2.1 Introduction

Statistical analysis of data continues to be of increasing importance in all areas of scientific inquiry. Even areas that traditionally have not employed advanced statistical techniques, such as psychology, sociology, and other areas of the social sciences are turning to statistics in order to give credence to their theories and conclusions. In the biological sciences, especially genetics, statistical methodology plays a key role in helping the scientist uncover important genetic relationships, through studying the distribution of genes in a population.

Statistical analysis involves making various assumptions regarding the data being examined. Often, some underlying distribution is assumed, usually including unknown parameters which are estimated (in some mathematically optimal way) from the data. In some areas, such as regression analysis and experimental design, a model representing the presumed relationship between the variables of interest is assumed, along with a (usually parametric) distributional assumption on the underlying error in the data. Both the parameters in the model and the parameters in the postulated error distribution must then be estimated from the given data.

As with any assumption, it is often the case that it is not fully justified. For example, one generally assumes that in a regression problem, the errors are normally distributed with mean 0 and (unknown) variance σ^2. In many cases the distribution is not normal, but rather is skewed or has a sharper peak than the normal distribution. In other instances, the postulated model may not be completely correct. One may assume a linear relationship exists between two variables, when in fact a quadratic or exponential relationship may be more appropriate.

It is, therefore, desirable to relax some (or all) of these assumptions so that statistical methodology can be used in a wider variety of situations. The objective of this thesis is to extend numerous available statistical methods, which are based on the assumption of normality, to non-normal distributions, particularly the family of Student's-t distributions. This has been possible through applications of the MML (modified maximum likelihood) method of estimation developed over the years by Tiku (1967-1990) and Tiku and Suresh (1992); see also Lee et al. (1980), Bhattacharya (1985) and Vaughan (1992). Before presenting the main results of this thesis, we begin with an introduction to these estimation procedures.

2.2 The MML Estimation Procedure

Let

$$x_1,\ x_2,\ \ldots,\ x_n$$

be a random sample of size n from the $N(\mu,\sigma^2)$ distribution. It is well known that the maximum likelihood estimators of μ and σ^2 are given by $\bar{x}$ and s^2 respectively. However, in cases where a complete sample is not available (eg. censoring) or the sample is from some distribution other than normal, the derivatives of the log-likelihood function may not admit explicit solutions for the unknown parameters, as seen in Tiku (1967-1990), and pointed out in Tiku and Suresh (1992). Tiku and Suresh, therefore, have extended Tiku's modified maximum likelihood method to location-scale distributions of the form $(1/\sigma)f[(x-\mu)/\sigma]$; more specifically they have considered the Student's-t family of distributions.

2.3 MML Estimation for the t-Family

Let x_1, x_2, ..., x_n be a random sample of size n from the symmetric distribution having p.d.f.

$$f(x) = \frac{1}{\sqrt{k}\,\sigma\,\beta(1/2, p-1/2)}\left\{1 + \frac{(x-\mu)^2}{k\sigma^2}\right\}^{-p}, \quad -\infty < x < \infty,\ p \geq 2 \tag{2.3.1}$$

where $k = 2p - 3$; p is a known shape parameter. It may be noted that $\sqrt{\nu}(x-\mu)/\sigma\sqrt{k}$ has a Student's-t distribution with $\nu = 2p - 1$ d.f., for $p \geq 2$.

From (2.3.1), the log-likelihood is given by

$$\ln L = \text{const.} - n\ln\sigma - p\sum_{i=1}^{n} \ln\left\{1 + \frac{(x_i-\mu)^2}{k\sigma^2}\right\} \tag{2.3.2}$$

Define $z_i = (x_i-\mu)/\sigma$, so that $\partial z_i/\partial\mu = -1/\sigma$ and $\partial z_i/\partial\sigma = -z_i/\sigma$. Thus (2.3.2) becomes

$$\ln L = \text{const.} - n\ln\sigma - p\sum_{i=1}^{n}\left\{1 + \frac{z_i^2}{k}\right\} \tag{2.3.3}$$

Differentiating (2.3.3) with respect to μ and σ yields

$$\frac{\partial \ln L}{\partial \mu} = \frac{2p}{k\sigma} \sum_{i=1}^{n} \frac{z_i}{1 + z_i^2/k} = 0 \qquad (2.3.4)$$

and $$\frac{\partial \ln L}{\partial \sigma} = -\frac{n}{\sigma} + \frac{2p}{k\sigma} \sum_{i=1}^{n} \frac{z_i^2}{1 + z_i^2/k} = 0 \qquad (2.3.5)$$

Now define $g(z) = z/[1 + z^2/k]$. Hence (2.3.4) and (2.3.5) become

$$\frac{\partial \ln L}{\partial \mu} = \frac{2p}{k\sigma} \sum_{i=1}^{n} g(z_i) = 0 \qquad (2.3.6)$$

and $$\frac{\partial \ln L}{\partial \sigma} = -\frac{n}{\sigma} + \frac{2p}{k\sigma} \sum_{i=1}^{n} z_i g(z_i) = 0 \qquad (2.3.7)$$

Equations (2.3.6) and (2.3.7) do not yield explicit solutions for μ and σ due to the non-linearity of the function g. It is possible to solve these equations iteratively, but as pointed out by Vaughan (1992), Barnett (1966) and Lee et al. (1980) have shown that iterative methods often experience difficulties such as slow convergence, multiple roots, convergence to incorrect values, or even divergence (see specifically Lee et al., 1980). Tiku and Suresh (1992) have overcome these problems by approximating the likelihood function in such a way as to allow explicit estimators for μ and σ to be computed. Specifically, let

$$x_{(1)} \le x_{(2)} \le \cdots \le x_{(n)}$$

be the order statistics obtained by arranging the original sample in ascending order. Now (2.3.6) and (2.3.7) can be rewritten as

$$\frac{\partial \ln L}{\partial \mu} = \frac{2p}{k\sigma} \sum_{i=1}^{n} g(z_{(i)}) = 0 \qquad (2.3.8)$$

and $$\frac{\partial \ln L}{\partial \sigma} = -\frac{n}{\sigma} + \frac{2p}{k\sigma} \sum_{i=1}^{n} z_{(i)} g(z_{(i)}) = 0 \qquad (2.3.9)$$

where $z_{(i)} = (x_{(i)} - \mu)/\sigma$. In order to allow explicit solutions for (2.3.8) and (2.3.9), Tiku and Suresh (1992) expand the function g in a Taylor series around $t_{(i)} = E(z_{(i)})$ as follows:

$$\begin{aligned} g(z_{(i)}) &\approx g(t_{(i)}) + [z_{(i)} - t_{(i)}]\, g'(t_{(i)}) \\ &= \frac{t_{(i)}}{1 + t_{(i)}^2/k} + [z_{(i)} - t_{(i)}]\frac{1 - t_{(i)}^2/k}{[1 + t_{(i)}^2/k]^2} \\ &= \frac{t_{(i)} + t_{(i)}^3/k - t_{(i)} + t_{(i)}^3/k}{[1 + t_{(i)}^2/k]^2} + z_{(i)}\frac{1 - t_{(i)}^2/k}{[1 + t_{(i)}^2/k]^2} \\ &= \frac{2t_{(i)}^3/k}{[1 + t_{(i)}^2/k]^2} + \left\{\frac{1 - t_{(i)}^2/k}{[1 + t_{(i)}^2/k]^2}\right\} z_{(i)} \end{aligned}$$

One assumes, of course, that the first two derivatives of $g(z)$ exist. Now define

$$\alpha_i = \frac{2t_{(i)}^3/k}{[1 + t_{(i)}^2/k]^2} \quad \text{and} \quad \beta_i = \frac{1 - t_{(i)}^2/k}{[1 + t_{(i)}^2/k]^2}$$

Hence we have $g(z_{(i)}) \approx \alpha_i + \beta_i z_{(i)}$. A numerical justification for such linear approximations is given by Tiku (1967), who showed that over small intervals $a \le z \le b$ the approximated function is essentially linear. Note that $\alpha_i = -\alpha_{n-i+1}$ and $\beta_i = \beta_{n-i+1}$. This follows from the fact that the t-family is symmetric; hence $t_{(i)} = -t_{(n-i+1)}$ and $t_{(i)}^2 = t_{(n-i+1)}^2$. From this result it is then clear that

$$\sum_{i=1}^{n} \alpha_i = 0 \qquad (2.3.10)$$

We now substitute these linear approximations into (2.3.8) and (2.3.9) to obtain:

$$\frac{\partial \ln L^*}{\partial \mu} = \frac{2p}{k\sigma} \sum_{i=1}^{n} [\alpha_i + \beta_i z_{(i)}] = 0 \qquad (2.3.11)$$

and

$$\frac{\partial \ln L^*}{\partial \sigma} = -\frac{n}{\sigma} + \frac{2p}{k\sigma} \sum_{i=1}^{n} z_{(i)} [\alpha_i + \beta_i z_{(i)}] = 0 \qquad (2.3.12)$$

Direct algebraic manipulation of (2.3.11) and (2.3.12) along with an application of (2.3.10) now gives explicit solutions for μ and σ, namely:

$$\hat{\mu} = \frac{1}{m} \sum_{i=1}^{n} \beta_i x_{(i)} \qquad (2.3.13)$$

$$\hat{\sigma} = \frac{B + \sqrt{B^2 + 4nC}}{2\sqrt{n(n-1)}}, \quad \text{corrected for bias,} \qquad (2.3.14)$$

(i.e. the divisor n was replaced by $\sqrt{n(n-1)}$),

where

$$m = \sum_{i=1}^{n} \beta_i \qquad (2.3.15)$$

$$B = \frac{2p}{k} \sum_{i=1}^{n} \alpha_i x_{(i)} \qquad (2.3.16)$$

$$C = \frac{2p}{k} \sum_{i=1}^{n} \beta_i \left[x_{(i)} - \hat{\mu} \right]^2$$

$$= \frac{2p}{k} \left\{ \sum_{i=1}^{n} \beta_i x_{(i)}^2 - m \hat{\mu}^2 \right\} \qquad (2.3.17)$$

For small values of p, however, $\hat{\sigma}$ may cease to be real. Tiku and Suresh (1992) and particularly Vaughan (1992) have suggested methods to remedy this drawback; see Section 2.11 of this chapter. Note that in the case of the normal distribution (i.e. $p = \infty$), $\alpha_i = 0$ and $\beta_i = 1$. In that case $g(z_i) = z_i$, and (2.3.13) and (2.3.14) reduce to $\hat{\mu} = \bar{x}$ and $\hat{\sigma}^2 = s^2$. Therefore this method reproduces the classical result if the underlying distribution is normal. It may also be noted that if we equate $g(z_i)$ to z_i, then the MML method reproduces exactly the least-squared estimators of μ and σ^2, for complete samples.

As Tiku (1967, 1970), Bhattacharya (1985), Tiku and Suresh (1992) and Vaughan (1992) point out, the modified likelihood equations are (under some very general regularity conditions) asymptotically equivalent to the likelihood equations; i.e. $(\ln L - \ln L^*)/n$ converges to zero as $n \to \infty$. As a result, the MML's $\hat{\mu}$ and $\hat{\sigma}$ are asymptotically equivalent to the corresponding ML estimators and are fully efficient; see also Bhattacharya (1985) and Vaughan (1992). Consequently, $\hat{\mu}$ and $\hat{\sigma}$ are asymptotically MVB estimators; see also Section 2.6.

2.4 Asymptotic Variances and Covariances

Straightforward differentiation of (2.3.11) and (2.3.12) gives:

$$\frac{\partial^2 \ln L^*}{\partial \mu^2} = -\frac{2p}{k\sigma^2}\sum_{i=1}^{n}\beta_i = -\frac{2pm}{k\sigma^2} \tag{2.4.1}$$

$$\frac{\partial^2 \ln L^*}{\partial \mu \partial \sigma} = -\frac{2p}{k\sigma^2}\sum_{i=1}^{n}\beta_i z_{(i)} \tag{2.4.2}$$

$$\frac{\partial^2 \ln L^*}{\partial \sigma^2} = \frac{n}{\sigma^2} - \frac{2p}{k\sigma^2}\sum_{i=1}^{n}[2\alpha_i z_{(i)} + 3\beta_i z_{(i)}^2] \tag{2.4.3}$$

The elements of the information matrix are then

$$-E\left(\frac{\partial^2 \ln L^*}{\partial \mu^2}\right) = \frac{2pm}{k\sigma^2}$$

$$-E\left(\frac{\partial^2 \ln L^*}{\partial \mu \partial \sigma}\right) = \frac{2p}{k\sigma^2}\sum_{i=1}^{n}\beta_i t_{(i)} = 0 \qquad \text{(by symmetry)}$$

and

$$-E\left(\frac{\partial^2 \ln L^*}{\partial \sigma^2}\right) = -\frac{n}{\sigma^2} + \frac{2p}{k\sigma^2}\sum_{i=1}^{n}[2\alpha_i t_{(i)} + 3\beta_i E(z_{(i)}^2)]$$

$$= -\frac{n}{\sigma^2} + \frac{2p}{k\sigma^2}\sum_{i=1}^{n}[2\alpha_i t_{(i)} + 3\beta_i(t_{(i)}^2 + \sigma_{i,i:n})] = \Omega/\sigma^2$$

where $\sigma_{i,i:n} = \mathrm{Var}(z_{(i)})$. Thus the information matrix is:

$$\underset{\sim}{I}(\mu, \sigma) = \begin{bmatrix} \dfrac{2pm}{k\sigma^2} & 0 \\ & \\ 0 & \dfrac{\Omega}{\sigma^2} \end{bmatrix}$$

and

$$\underset{\sim}{V}(\mu, \sigma) = \underset{\sim}{I}^{-1}(\mu, \sigma) = \begin{bmatrix} \dfrac{k\sigma^2}{2pm} & 0 \\ & \\ 0 & \dfrac{\sigma^2}{\Omega} \end{bmatrix}.$$

The expected values and variances of the order statistics of random samples from (2.3.1) required to calculate Ω are tabulated for $p = 2(.5)10$ and $n \le 20$ in Tiku and Kumra (1981), in Barnett (1966) for $p = 1$, and in Vaughan (1992) for $p = 1.5$. For $n > 20$, Tiku and Suresh (1992) give in their Appendix a method by David and Johnson (1954) for computing the necessary values. However, we will see in later chapters that the minimum variance bound for σ provides an adequate approximation for $V(\hat{\sigma})$; it will therefore be used in most cases in place of σ^2/Ω, which will eliminate the need for tables of variances of order statistics. Expected values will, however, be needed to compute the estimators. This will be discussed further in subsequent sections.

2.5 Efficiency of the MML Estimators

From Tiku and Kumra (1981), we know that the minimum variance bounds for μ and σ are, respectively:

$$MVB(\mu) = \frac{(p-\frac{3}{2})(p+1)}{np(p-\frac{1}{2})}\sigma^2 \qquad (2.5.1)$$

and $$MVB(\sigma) = \frac{p+1}{2n(p-\frac{1}{2})}\sigma^2 \qquad (2.5.2)$$

From the asymptotic variance-covariance matrix, we have

$$Var(\hat{\mu}) \simeq \frac{k}{2pm}\sigma^2 \qquad (2.5.3)$$

and $$Var(\hat{\sigma}) \simeq \left(\frac{n}{n-1}\right)\frac{\sigma^2}{\Omega} \qquad (2.5.4)$$

(n/(n-1) appears due to the bias correction for $\hat{\sigma}$.)

To evaluate the efficiency of the MML estimators $\hat{\mu}$ and $\hat{\sigma}$, Tiku and Suresh (1992) have tabulated the MVB's and the variances (scaled by σ^2) for various values of p, and various sample sizes.

TABLE I: **Comparison of the MVB and the Exact and Approximate Variances of $\hat{\mu}$ and $\hat{\sigma}$**

n	p	$MVB(\mu)/\sigma^2$	Exact $Var(\hat{\mu})/\sigma^2$	Approx.: $k/2pm$	$MVB(\sigma)/\sigma^2$	Approx. $Var(\hat{\sigma})/\sigma^2$
5	2	0.1000	0.1201	0.1056		
	4	0.1786	0.1898	0.1691		
	7	0.1934	0.1975	0.1837		
	10	0.1968	0.1989	0.1888		
10	2	0.0500	0.0551	0.0516		
	4	0.0893	0.0923	0.0876		
	7	0.0967	0.0979	0.0944	0.061	0.056
	10	0.0984	0.0990	0.0964	0.058	0.056
15	2	0.0333	0.0357	0.0341		
	4	0.0595	0.0607	0.0589		
	7	0.0645	0.0650	0.0635	0.041	0.038
	10	0.0656	0.0659	0.0647	0.039	0.037
20	2	0.0250	0.0264	0.0254		
	4	0.0446	0.0453	0.0444		
	7	0.0484	0.0487	0.0478	0.031	0.029
	10	0.0492	0.0494	0.0487	0.029	0.028

From Table I it can be seen that $\mathrm{Var}(\hat{\mu})/\sigma^2$ and $\mathrm{Var}(\hat{\sigma})/\sigma^2$ attain values remarkably close to the MVB's, especially as the sample size n increases. This clearly illustrates that $\hat{\mu}$ and $\hat{\sigma}$ are highly efficient estimators; see also Section 2.12. Moreover, it is seen that $\mathrm{Var}(\hat{\sigma})/\sigma^2$ is so close to $\mathrm{MVB}(\sigma)/\sigma^2$ that the latter may be used as a very good approximation for $\mathrm{Var}(\hat{\sigma})$, thus eliminating the need for tables of the variances of order statistics for the t-family (2.3.1); see also Vaughan (1992) for many more such comparisons. It may also be noted that the true variance of $\hat{\mu}$ is also very close to $\mathrm{MVB}(\mu)$, and the latter may be used for $V(\hat{\mu})$.

2.6 MML Estimators for Censored Samples

Tiku and Suresh (1992) consider the estimation procedure of Section 2.2 in the case of Type-II censored samples. Arrange the original sample in ascending order and censor the r_1 smallest and r_2 largest observations, yielding the following Type-II censored sample:

$$x_{(r_1+1)} \le x_{(r_1+2)} \le \cdots \le x_{(n-r_2)}$$

Here, the log-likelihood equation is

$$\ln L = \text{const.} - A\ln\sigma - p\sum_{i=r_1+1}^{n-r_2} \ln\left\{1 + \frac{z_{(i)}^2}{k}\right\}$$
$$+ r_1 \ln F(z_{(r_1+1)}) + r_2 \ln[1 - F(z_{(n-r_2)})] \qquad (2.6.1)$$

where $F(z) = \int_{-\infty}^{z} f(t)\,dt$ is the c.d.f. of the family (2.3.1).

In a manner similar to the complete sample case, we have:

$$\frac{\partial \ln L}{\partial \mu} = \frac{2p}{k\sigma} \sum_{i=r_1+1}^{n-r_2} g(z_{(i)}) + \frac{1}{\sigma} \{ r_2 h_2(z_{(n-r_2)}) - r_1 h_1(z_{(r_1+1)}) \} = 0 \tag{2.6.2}$$

and

$$\frac{\partial \ln L}{\partial \sigma} = -\frac{A}{\sigma} + \frac{2p}{k\sigma} \sum_{i=r_1+1}^{n-r_2} z_{(i)} g(z_{(i)}) + \frac{1}{\sigma} \{ r_2 z_{(n-r_2)} h_2(z_{(n-r_2)}) - r_1 z_{(r_1+1)} h_1(z_{(r_1+1)}) \} = 0 \tag{2.6.3}$$

where $A = n-r_1-r_2$, $h_1(z) = f(z)/F(z)$, and $h_2(z) = f(z)/[1 - F(z)]$. As before, (2.6.2) and (2.6.3) do not yield explicit solutions due to the presence of the intractable functions g, h_1, and h_2. As in Section 2.2, Tiku and Suresh (1992) use the approximation $g(z_{(i)}) \approx \alpha_i + \beta_i z_{(i)}$, $i = r_1+1, \ldots, n-r_2$. Further, h_1 and h_2 are approximated by expanding them in Taylor series about the points $t_{(r_1+1)}$ and $t_{(n-r_2)}$ respectively; i.e.

$$\begin{aligned} h_1(z_{(r_1+1)}) &\approx a_1 - b_1 z_{(r_1+1)} \\ h_2(z_{(n-r_2)}) &\approx a_2 + b_2 z_{(n-r_2)} \end{aligned} \tag{2.6.4}$$

For large n, the coefficients in (2.6.4) are given by:

$$\begin{aligned} b_i &= -f(t_i) \left\{ \frac{2p}{k} g(t_i) - \frac{f(t_i)}{q_i} \right\} \Big/ q_i \\ a_i &= \frac{f(t_i)}{q_i} - b_i t_i \end{aligned} \tag{2.6.5}$$

for $i = 1,2$; $q_i = r_i/n$ and t_i is the solution of

$$\int_{-\infty}^{t_i} f(z)\,dz = 1 - q_i \ .$$

The exact values of a_i and b_i are, in fact, given by taking $t_i = E(z_{(i)})$ in (2.6.5), but the values above are close to the exact values; see Tiku and Suresh (1992) and Vaughan (1992). We will therefore use these values as has been done by Tiku and Suresh (1992).

Substituting the approximation for g and (2.6.4) into (2.6.2) and (2.6.3), we have:

$$\frac{\partial \ln L^*}{\partial \mu} = \frac{2p}{k\sigma} \sum_{i=r_1+1}^{n-r_2} [\alpha_i + \beta_i z_{(i)}]$$

$$+ \frac{1}{\sigma} \{ r_2 a_2 - r_1 a_1 + r_1 b_1 z_{(r_1+1)} + r_2 b_2 z_{(n-r_2)} \} = 0 \qquad (2.6.6)$$

$$\frac{\partial \ln L^*}{\partial \sigma} = -\frac{A}{\sigma} + \frac{2p}{k\sigma} \sum_{i=r_1+1}^{n-r_2} z_{(i)} [\alpha_i + \beta_i z_{(i)}]$$

$$+ \frac{1}{\sigma} \{ r_2 a_2 z_{(n-r_2)} - r_1 a_1 z_{(r_1+1)} + r_1 b_1 z^2_{(r_1+1)} + r_2 b_2 z^2_{(n-r_2)} \} = 0 \qquad (2.6.7)$$

Equations (2.6.6) and (2.6.7) now yield explicit solutions for μ and σ:

$$\hat{\mu}_c = K + D\hat{\sigma} \qquad (2.6.8)$$

and $$\hat{\sigma}_c = \frac{B + \sqrt{B^2 + 4AC}}{2\sqrt{A(A-1)}} \qquad \text{(corrected for bias)}, \qquad (2.6.9)$$

where

$$M = \frac{2p}{k} \sum_{i=r_1+1}^{n-r_2} \beta_i + r_1 b_1 + r_2 b_2 \qquad (2.6.10)$$

$$N = \frac{2p}{k} \sum_{i=r_1+1}^{n-r_2} \alpha_i + r_2 a_2 - r_1 a_1 \ , \qquad D = N/M \qquad (2.6.11)$$

$$K = \frac{1}{M} \left\{ \frac{2p}{k} \sum_{i=r_1+1}^{n-r_2} \beta_i x_{(i)} + r_1 b_1 x_{(r_1+1)} + r_2 b_2 x_{(n-r_2)} \right\} \qquad (2.6.12)$$

and

$$B = \frac{2p}{k}\sum_{i=r_1+1}^{n-r_2} \alpha_i x_{(i)} + r_2 a_2 x_{(n-r_2)} - r_1 a_1 x_{(r_1+1)} - NK \qquad (2.6.13)$$

$$C = \frac{2p}{k}\sum_{i=r_1+1}^{n-r_2} \beta_i \{x_{(i)} - K\}^2$$

$$+ r_1 b_1 \{x_{(r_1+1)} - K\}^2 + r_2 b_2 \{x_{(n-r_2)} - k\}^2$$

$$= \frac{2p}{k}\sum_{i=r_1+1}^{n-r_2} \beta_i x_{(i)}^2 + r_1 b_1 x_{(r_1+1)}^2 + r_2 b_2 x_{(n-r_2)}^2 - MK^2 \qquad (2.6.14)$$

Note that in the case of symmetric censoring (i.e. when $r_1 = r_2 = r$), $a_1 = a_2 = a$ and $b_1 = b_2 = b$, and the resultant expressions are greatly simplified. For complete samples, $r = 0$, $a = 0$ and $b = 1$, and the estimators reduce to those given in Section 2.2, as is expected. The estimates $\hat{\mu}$ and $\hat{\sigma}$ above are asymptotically MVB estimators; see the Appendix.

2.7 Asymptotic Variances and Covariances Based On Censored Samples

Further differentiation of (2.6.6) and (2.6.7) yield the following second-order partial derivatives:

$$\frac{\partial^2 \ln L^*}{\partial \mu^2} = -\frac{2p}{k\sigma^2}\sum_{i=r_1+1}^{n-r_2} \beta_i - \frac{1}{\sigma^2}\{r_1 b_1 + r_2 b_2\} = -\frac{M}{\sigma^2}$$

$$\frac{\partial^2 \ln L^*}{\partial \mu \partial \sigma} = -\frac{N}{\sigma^2} - \frac{2}{\sigma^2}\left\{\frac{2p}{k}\sum_{i=r_1+1}^{n-r_2} \beta_i z_{(i)} + r_1 b_1 z_{(r_1+1)} + r_2 b_2 z_{(n-r_2)}\right\}$$

$$\frac{\partial^2 \ln L^*}{\partial \sigma^2} = \frac{A}{\sigma^2} - \frac{2p}{k\sigma^2}\sum_{i=r_1+1}^{n-r_2} [2\alpha_i z_{(i)} + 3\beta_i z_{(i)}^2]$$

$$- \frac{1}{\sigma^2}\{2[r_2 a_2 z_{(n-r_2)} - r_1 a_1 z_{(r_1+1)}] + 3[r_1 b_1 z_{(r_1+1)}^2 + r_2 b_2 z_{(n-r_2)}^2]\}$$

The elements of the information matrix are then:

$$-E\left(\frac{\partial^2 \ln L^*}{\partial \mu^2}\right) = \frac{M}{\sigma^2}$$

$$-E\left(\frac{\partial^2 \ln L^*}{\partial \mu \partial \sigma}\right) = \frac{N}{\sigma^2} + \frac{2}{\sigma^2}\left\{\frac{2p}{k}\sum_{i=r_1+1}^{n-r_2} \beta_i t_{(i)} + r_1 b_1 t_{(r_1+1)} + r_2 b_2 t_{(n-r_2)}\right\} = \Lambda/\sigma^2$$

$$-E\left(\frac{\partial^2 \ln L^*}{\partial \sigma^2}\right) = -\frac{A}{\sigma^2} + \frac{2p}{k\sigma^2}\sum_{i=r_1+1}^{n-r_2} [\, 2\alpha_i t_{(i)} + 3\beta_i (t_{(i)}^2 + \sigma_{i,i:n})\,]$$

$$+ \frac{1}{\sigma^2}\{\, 2[\, r_2 a_2 t_{(n-r_2)} + r_1 a_1 t_{(r_1+1)}\,]$$

$$+ 3\beta_i [\, r_1 b_1 (t_{(r_1+1)} + \sigma_{r_1+1,r_1+1:n}) + r_2 b_2 (t_{(n-r_2)} + \sigma_{n-r_2,n-r_2:n})\,]\,\}$$

$$= \Omega/\sigma^2$$

Therefore the information matrix can be written as:

$$\underset{\sim}{I}(\mu_c, \sigma_c) = \frac{1}{\sigma^2}\begin{bmatrix} M & \Lambda \\ \Lambda & \Omega \end{bmatrix}$$

and

$$\underset{\sim}{V}(\mu_c, \sigma_c) = \underset{\sim}{I}^{-1}(\mu_c, \sigma_c) = \frac{\sigma^2}{\Delta}\begin{bmatrix} \Omega & -\Lambda \\ -\Lambda & M \end{bmatrix} \quad \text{where } \Delta = M\Omega - \Lambda^2 .$$

2.8 Efficiencies Based On Symmetrically Censored Samples

As shown in Tiku and Suresh (1992), for symmetric censoring (i.e. $r_1 = r_2 = r$):

$$-E\left(\frac{\partial^2 \ln L^*}{\partial \mu \partial \sigma}\right) = \frac{2}{\sigma^2}\left\{\frac{2p}{k}\sum_{i=r+1}^{n-r} \beta_i t_{(i)} + r b\,[\,t_{(r+1)} + t_{(n-r)}]\right\} \quad \text{since } N = 0$$

$$= 0 \quad \text{by symmetry.}$$

Therefore $V(\hat{\mu}_c) = \frac{\sigma^2}{M}$ and $V(\hat{\sigma}_c) = \frac{\sigma^2}{\Omega}$.

Tiku and Suresh (1992) compare $1/M$ with the exact variance $V(\hat{\mu})/\sigma^2$. These results are reported in Table II, for $p = 2$, $n = 10,20$, and various values of r. For $p > 2$, the two sets of values are even closer than for $p = 2$.

TABLE II: Comparison of exact variance $V(\hat{\mu})/\sigma^2$ and $1/M$ for n = 10, 20

n	r	$V(\hat{\mu}_c)/\sigma^2$	1/M
10	1	0.0550	0.0513
	2	0.0550	0.0508
20	1	0.0262	0.0255
	2	0.0263	0.0255
	3	0.0263	0.0255
	4	0.0263	0.0254

2.9 Modifications for Small Values of p

Tiku and Suresh (1992) note in the final section of their paper that, in the case $1 \le p < 2$, k is taken to be 1, and the family (2.3.1) reduces to:

$$f(x) \propto \frac{1}{\sigma}\left\{1 + \frac{(x-\mu)^2}{\sigma^2}\right\}^{-p}, \qquad -\infty < x < \infty \qquad (2.9.1)$$

In the specific case $p = 1$, (2.9.1) reduces to the Cauchy distribution, and Tiku and Suresh note that their estimators $\hat{\mu}$ and $\hat{\sigma}$ also work in this case upon setting $p = 1$ and $k = 1$ in the corresponding expressions. However, since the expected values of extreme order statistics from a Cauchy distribution are infinite, $\beta_1 = -\beta_n = 0$. Moreover, $\beta_2 = -\beta_{n-1}$ is very small since $E(z_{(2)}) = -E(z_{(n-1)})$ is very large (Barnett, 1966). For $p = 1$, therefore, the Tiku-Suresh estimator of μ essentially reduces to:

$$\hat{\mu}_0 = \sum_{i=3}^{n-2} \beta_i x_{(i)} \Big/ \sum_{i=3}^{n-2} \beta_i \, .$$

Tiku and Suresh (1992) and Vaughan (1992) show that $\hat{\mu}_0$ is highly efficient; see the next section. The estimation of σ for small values of p is, however, problematic. This is discussed in detail in Section 2.11.

2.10 Estimation of the Location Parameter for Small p

Vaughan (1992) has conducted a study for $p = 1$ and 1.5 to compare the efficiency of the MMLE $\hat{\mu}$ to the BLUE (Best Linear Unbiased Estimator) μ^* of Barnett (1966), at various levels of censoring. His results are reproduced in Table III.

Based on these values, Vaughan (1992) concludes that the MMLE $\hat{\mu}$ gains in efficiency (as compared to μ^*) as censoring increases, and is only slightly less efficient for complete samples. It may be noted that the computation of μ^* is very time consuming and, therefore, the MML estimators are preferred.

2.11 Estimation of the Scale Parameter for Small p

For small p ($1 \le p < 3.5$), Vaughan (1992) has shown that the Tiku-Suresh estimator $\hat{\sigma}$ can cease to be real. In order for $\hat{\sigma}$ to be real and positive, Vaughan argues that it is necessary that

$$\frac{2p}{k}\beta_{r+1} + rb \geq 0 \ . \qquad (2.11.1)$$

Expression (2.11.1) being satisfied in turn guarantees that

$$B + \sqrt{B^2 + 4AC} > 0,$$

i.e. $\hat{\sigma} > 0$. Vaughan (1992) has computed the necessary values of r satisfying (2.36). For various sample sizes and $p = 1$, for example, he gives r satisfying (2.11.1):

n	8	10	15	20	30	50	100
r	3	3	5	7	10	18	36

Based on this censoring scheme, it is now possible to construct MMLE's for σ in the range $1 \leq p < 3.5$. Once again, Vaughan's results are compared to Barnett's (1966) BLUE's. We summarize these results in Table IV; the values of r needed to give real and positive values of $\hat{\sigma}$ are also reproduced. The values calculated from the ad-hoc approximation for censored samples (obtained by replacing n by n-$2r$)

$$V(\hat{\sigma}_c) \approx MVB(\sigma) = \frac{(p+1)}{2(n-2r)(p-\frac{1}{2})}\sigma^2$$

are also given. For $r = 0$, this formula reduces to (2.5.1). It is shown that this approximation is quite good. Replacing n-$2r$ with n-$2r$-1 in the denominator gives a closer approximation for $p \geq 3.5$.

From Table IV it is evident that the Tiku-Suresh MMLe's $\hat{\mu}$ and $\hat{\sigma}$ are, on the whole, jointly more efficient than the BLUE's μ^* and σ^*; the former can, however, be calculated without tables of coefficients which are demanded by the BLUE's. Also, the amount of censoring to ensure that $\hat{\sigma}$ is positive does not adversely affect the efficiency.

Table III: Comparison of Exact Variances $V(\hat{\mu}_c)/\sigma^2$ and $V(\mu_c^*)/\sigma^2$ for $p = 1, 1.5$; $n = 10, 15, 20$

n	r/n	$V(\hat{\mu}_c)/\sigma^2$	$V(\mu_c^*)/\sigma^2$
$p = 1$			
10	2	0.357	0.326
	3	0.337	0.336
	4	0.336	0.336
15	2	0.207	0.182
	3	0.191	0.185
	4	0.191	0.191
	5	0.194	0.193
	6	0.194	0.194
20	2	0.138	0.126
	3	0.132	0.127
	4	0.131	0.129
	5	0.133	0.133
	6	0.138	0.135
	7	0.138	0.136
$p = 1.5$			
10	1	0.099	0.098
	2	0.100	0.099
	3	0.100	0.099
15	1	0.066	0.062
	2	0.063	0.063
	3	0.064	0.063
20	1	0.048	0.046
	2	0.046	0.046
	3	0.046	0.046
	4	0.046	0.046

Table IV: **Comparison of Efficiencies of the MMLE's and BLUE's based on Vaughan's (1992) modified censoring scheme.**

p	n	r	Exact $V(\mu_c^*)/\sigma^2$	Exact $V(\hat{\mu}_c)/\sigma^2$	Approx. $MVB(\mu)/\sigma^2$	Exact $V(\sigma_c^*)/\sigma^2$	Exact* $V(\hat{\sigma}_c)/\sigma^2$	Approx. $MVB(\sigma)/\sigma^2$
1	10	3	0.336	0.337		0.474	0.478	
	15	5	0.193	0.194		0.283	0.248	
	20	7	0.135	0.139		0.209	0.162	
1.5	10	2	0.099	0.100		0.196	0.188	
	15	3	0.063	0.064		0.118	0.115	
	20	4	0.046	0.046		0.085	0.082	
2.0	10	1	0.055	0.055	0.062	0.129	0.126	0.125
	15	2	0.036	0.036	0.045	0.085	0.084	0.077
	20	3	0.026	0.026	0.036	0.064	0.062	0.071
1.5	10	1	0.074	0.075	0.087	0.114	0.116	0.109
	15	1	0.049	0.049	0.054	0.068	0.066	0.067
	20	1	0.036	0.036	0.039	0.049	0.048	0.049
3.0	10	0	0.084	0.084	0.080	0.091	0.094	0.080
	15	1	0.055	0.055	0.061	0.063	0.060	0.062
	20	1	0.041	0.041	0.044	0.045	0.044	0.044
3.5	10	0	0.089	0.089	0.086	0.084	0.086	0.083 †
	15	0	0.059	0.059	0.057	0.054	0.056	0.056 †
	20	0	0.044	0.044	0.043	0.040	0.041	0.042 †

* These values are based on 10,000 Monte Carlo runs.

† These values are calculated from the approximation
$V(\hat{\sigma}) \approx (p+1)\sigma^2/[2(n-2r-1)(p-\frac{1}{2})]$

2.12 Asymptotic Distributions for Complete Samples

We note that

$$\frac{\partial lnL}{\partial \mu} \approx \frac{\partial lnL^*}{\partial \mu} = \frac{2pM}{k\sigma^2}\left[\hat{\mu} - \mu\right].$$

Since $\frac{1}{n}\left[\frac{\partial lnL}{\partial \mu} - \frac{\partial lnL^*}{\partial \mu}\right]$ is asymptotically equal to zero, and $\partial \ln L/\partial \mu$ is asymptotically distributed as normal (Bartlett, 1953) with mean zero and variance $-E(\partial^2 \ln L/\partial \mu^2)$, it immediately follows that $\sqrt{2p/k}\,\sqrt{M}\,(\hat{\mu} - \mu)$ is asymptotically distributed as normal N(0,1).

Again, we note that

$$\frac{\partial lnL}{\partial \sigma} \approx \frac{\partial lnL^*}{\partial \sigma} \approx -\frac{1}{\sigma^3}\left[(n\sigma^2 - C) - m(\hat{\mu} - \mu)^2\right]$$

which is exactly similar to the normal-theory equation

$$\frac{\partial lnL}{\partial \sigma} = -\frac{1}{\sigma^3}\left[(n\sigma^2 - S) - n(\bar{y} - \mu)^2\right]$$

$$\text{where} \quad S = \sum_{i=1}^{n}(y_i - \bar{y})^2 .$$

Therefore, $(n-1)\hat{\sigma}^2/\sigma^2 \approx C/\sigma^2$ is asymptotically distributed as chi-square with n-1 d.f; see Bartlett (1953). An improvement over this result can be achieved by noting that

$$E(\hat{\sigma}^2) \approx \sigma^2\left\{1 + \frac{p+1}{2(n-1)(p-½)}\right\}$$

Now let us regard $h\hat{\sigma}^2/\sigma^2$ as an approximate chi-square random variable with n-1 degrees of freedom, for some appropriate constant h. Equating expected values, we immediately obtain

$$h = \frac{n-1}{1 + \frac{p+1}{2(n-1)(p-\frac{1}{2})}}$$

The distribution of $h\hat{\sigma}^2/\sigma^2$ may then be referred to a chi-squared distribution with n-1 degrees of freedom.

2.13 Main Thesis Aims

It will be the aim of this thesis, then, to incorporate both the Tiku-Suresh (1992) estimation procedure and the modified censoring scheme for small p due to Vaughan (1992), and to extend the classical results (based on the assumption of normality) to non-normal distributions represented by (2.3.1). Specifically, we will be considering the following:

(1) One-Way Analysis of Variance based on complete and censored samples.

(2) Linear and Multiple Linear Regression based on complete and censored samples.

(3) Bivariate and Multivariate Regression with Concomitant Variables based on complete and censored samples.

In each of these cases it will be shown that the Tiku-Suresh estimation procedure produces results similar in form to the classical results (based on the assumption of normality). Moreover, it will be shown that the estimators so produced are remarkably efficient, while at the same time being easy to compute.

Appendix A1

The MML equation (2.6.6) can be written as

$$\frac{\partial \ln L^*}{\partial \mu} = \frac{M}{\sigma^2}(K + D\sigma - \mu) \tag{A1.1}$$

where M, D, and K are given by equations (2.6.10)-(2.6.12). For fixed $q_1 = r_1/n$ and $q_2 = r_2/n$ and large n, therefore, $K + D\sigma$ is conditionally (σ known) the MVB estimator, and

$$V(\hat{\mu}_c/\sigma) = \frac{\sigma^2}{M} . \tag{A1.2}$$

Moreover, the conditional distribution of $\hat{\mu}_c$ given σ is asymptotically (q_1 and q_2 are fixed and n tends to infinity) normal with mean μ and variance σ^2/M. For symmetric censoring, however, $D = 0$ in which case the results above are true unconditionally. These results follow from the fact that for fixed q_1 and q_2,

$$\lim_{n\to\infty} \frac{1}{n}\left\{\frac{\partial \ln L}{\partial \mu} - \frac{\partial \ln L^*}{\partial \mu}\right\} = 0$$

and the distribution of $\partial \ln L/\partial \mu$ is asymptotically (n tends to infinity) normal with mean zero and variance

$$-E\left(\frac{\partial^2 \ln L}{\partial \mu^2}\right) ; \tag{A1.3}$$

see Bartlett (1954) and Bhattacharya (1985) for details.

It can similarly be shown that for symmetric censoring, the asymptotic distribution of $(A-1)\hat{\sigma}_c^2/\sigma^2$ is a chi-square with A-1 d.f. A better approximation is obtained along the lines of Section 2.12, i.e. the distribution of

$$h\hat{\sigma}_c^2/\sigma^2 \tag{A1.4}$$

is referred to a chi-square distribution with A-1 d.f. Here

$$h = \frac{A-1}{1 + \frac{p+1}{2(A-1)(p-\frac{1}{2})}} \; . \tag{A1.5}$$

Chapter 3

3.1 Introduction: The Analysis of Variance

Statistical methodology is used most often in practice to analyze the results of scientific experiments. For example, an agricultural scientist might wish to examine the effects of different fertilizers on the growth of grain. An engineer may be interested in the tensile strength of different alloys used in bridge construction. A geneticist might wish to assess the effects of various mutagens on bacteria cells. In each of these cases, various treatments (fertilizer, type of alloy, mutagenic substance) are examined in a systematic way to see if their effects differ. For instance, in the fertilizer example, if the mean height of plants grown with Fertilizer A is significantly larger than for Fertilizer B or C, then this suggests that Fertilizer A is the best choice.

In light of a particular set of data, how then does one decide if an observed difference is significant, or merely a reflection of the random error in the experiment? Statistical techniques for answering this question are collectively termed *Experimental Design and Analysis*. Statistical methods are generally applied at three stages during a scientific experiment, as outlined in Hicks (1973), Chapter 1:

(1) The Experiment

When setting up an experiment, the experimenter must decide on the variable(s) of interest, and how to measure them with the desired accuracy. Both the independent variables and the response (or dependent) variable must be identified. In the fertilizer example, the independent variable is fertilizer type, and the response variable is plant height. All other sources of variation must be eliminated so that any differences in response may clearly be attributed to the independent variables.

(2) The Design

Statistics enters into the design phase in various ways. Decisions must be made regarding the amount of data to be collected (sample size determination) and how large an error in making conclusions will be tolerated (significance level). Ideally, these questions should be answered before the start of the experiment. Also, experimental items should be given the various treatments in a random fashion, so that an assumption about their independence of errors is justified.

(3) The Analysis

The analysis of experimental data includes data collection and reduction, and generally involves the calculation of *test statistics* from which conclusions may be drawn (in a mathematically precise manner) regarding the source of variation. In Experimental Design, this is usually accomplished by the creation of an **ANOVA** (**AN**alysis **O**f **VA**riance) table.

It is part (3) of this outline that we wish to address in this chapter. Specifically, we will begin by outlining and discussing the classical approach to Analysis of Variance, including the computation of the ANOVA table. In subsequent sections, we will extend this approach through the application of the MML method of estimation, and will show that the results obtained are exactly similar in form to the classical formulae, and are no more computationally involved. At the end of the chapter we will report some simulations based on complete and censored samples which will show that the proposed ANOVA procedure is more efficient than the classical normal-theory procedure for the family of distributions considered.

3.2 Classical One-Way Analysis of Variance

Consider the completely randomized single-factor design given by the model

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij} \qquad (i = 1,\dots,\ell;\; j = 1,\dots,n_i) \tag{3.2.1}$$

where: y_{ij} denotes the j^{th} observation in treatment group i

τ_i denotes the effect of the i^{th} treatment group

ε_{ij} represents the random error inherent in y_{ij}

and μ is the overall (grand) mean.

In the classical setting it is usually assumed that the ε_{ij}'s are independent and identically distributed as $N(0,\sigma^2)$, where σ^2 is the common treatment variance. The data matrix consists of ℓ independent random samples as follows:

$$\begin{array}{ccccc} \multicolumn{5}{c}{\text{Treatments}} \\ 1 & 2 \;\cdots & i & \cdots & \ell \\ \hline y_{11} & y_{21} \;\cdots & y_{i1} & \cdots & y_{\ell 1} \\ y_{12} & y_{22} \;\cdots & y_{i2} & \cdots & y_{\ell 2} \\ \vdots & \vdots & \vdots & & \vdots \\ y_{1n_1} & y_{2n_2} \;\cdots & y_{in_i} & \cdots & y_{\ell n_\ell} \end{array} \tag{3.2.2}$$

Using this data matrix one wishes to estimate the overall mean μ, the treatment effects τ_i, and the error variance σ^2. However, μ and τ_i's are not separately estimable; rather $\mu + \tau_i$ is estimable for each i. Generally what is done in this case is to define $\tau_i^* = \mu + \tau_i$ and re-write the model (3.2.1) as:

$$y_{ij} = \tau^*_i + \varepsilon_{ij} \tag{3.2.3}$$

from which the τ_i^{*}'s may be estimated (knowing that under the constraint $\sum_{i=1}^{\ell} n_i \tau_i = 0$, an estimate of μ is $\bar{\bar{y}}$, the sample grand mean). Assuming that the ε_{ij} are i.i.d. normal $N(0,\sigma^2)$, the likelihood function based on (3.2.3) is:

$$L \propto \frac{1}{\sigma^N} \prod_{i=1}^{\ell} \prod_{j=1}^{n_i} \exp\left\{\frac{-(y_{ij} - \mu - \tau_i)^2}{2\sigma^2}\right\}$$

$$= \frac{1}{\sigma^N} \prod_{i=1}^{\ell} \prod_{j=1}^{n_i} \exp\left\{\frac{-(y_{ij} - \tau_i^{*})^2}{2\sigma^2}\right\}$$

where $N = \sum n_i$. Now define $z_{ij} = (y_{ij} - \tau_i^{*})/\sigma$. Then we have

$$L \propto \frac{1}{\sigma^N} \prod_{i=1}^{\ell} \prod_{j=1}^{n_i} \exp\left\{-\frac{z_{ij}^2}{2}\right\}$$

$$\text{or} \qquad \ln L = \text{const.} - N \ln\sigma - \frac{1}{2} \sum_{i=1}^{\ell} \sum_{j=1}^{n_i} z_{ij}^2 \tag{3.2.4}$$

Differentiating (3.2.4) with respect to τ_i^{*} and σ gives:

$$\frac{\partial \ln L}{\partial \tau_i^{*}} = \frac{1}{\sigma} \sum_{j=1}^{n_i} z_{ij} = 0, \quad i = 1, \ldots, \ell \tag{3.2.5}$$

$$\text{and} \quad \frac{\partial \ln L}{\partial \sigma} = -\frac{N}{\sigma} - \frac{1}{\sigma} \sum_{i=1}^{\ell} \sum_{j=1}^{n_i} z_{ij}^2 = 0 \tag{3.2.6}$$

The solutions of (3.2.5) and (3.2.6) are the following ML estimators:

$$\hat{\tau}_i^{*} = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij} = \bar{y}_i, \quad (i = 1, \ldots, \ell) \tag{3.2.7}$$

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{i=1}^{\ell} \sum_{j=1}^{n_i} \{ y_{ij} - \bar{y}_i \}^2 \tag{3.2.8}$$

Now note that based on (3.2.8), $E(N\hat{\sigma}^2) = (N-\ell)\sigma^2$; we therefore correct for bias to obtain:

$$\hat{\sigma}^2 = \frac{1}{N-\ell}\sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\{y_{ij} - \bar{y}_i\}^2 .$$

3.3 Normal Theory ANOVA Table

Usually one wishes to test hypotheses regarding the treatment effects to determine if one or more treatments in the experiment differ in any way from the others. In other words, one wishes to test the null hypothesis H_0: $\tau_i = 0$ $(i = 1,\ldots,\ell)$; i.e. there are no differences between treatments. This test is performed through the use of the ANOVA table. Consider the partition of the total sum of squares as follows:

$$\begin{aligned} SST &= \sum_{i=1}^{\ell}\sum_{j=1}^{n_i}(y_{ij} - \bar{\bar{y}})^2 \\ &= \sum_{i=1}^{\ell}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_i)^2 + \sum_{i=1}^{\ell} n_i(\bar{y}_i - \bar{\bar{y}})^2 \\ &= \quad SSE \quad + \quad SS_{Treat} \end{aligned}$$

Dividing each of these expressions by their degrees of freedom, we obtain the corresponding mean sums of squares as:

$$MST = \frac{SST}{N-1}; \qquad MSE = \frac{SSE}{N-\ell}; \qquad MS_{Treat} = \frac{SS_{Treat}}{\ell-1};$$

MSE provides an unbiased estimator of σ^2 always, but MST provides an unbiased estimator of σ^2 only if H_0 is true. It may also be noted that MSE is the ML estimator (corrected for bias) for σ^2, and MST is the ML estimator (corrected for bias) for σ^2 under H_0.

We now form the ANOVA table, including the corresponding F-ratio (since each mean square is distributed as χ^2 with its corresponding degree of freedom):

Source	df	SS	MS	F
Between Treatments	ℓ-1	SS_{Treat}	$SS_{Treat}/(\ell-1)$	MS_{Treat}/MSE
Error	N-ℓ	SSE	SSE/(N-ℓ)	
Total	N-1	SST	SST/(N-1)	

To test H_0 at the α^{th} level of significance, we refer the F-value to the 100(1-α)% point of a central F distribution with d.f. (ℓ-1,N-ℓ). If the computed value exceeds this critical point, one rejects H_0; i.e. there is evidence that the treatments are different.

Under H_1: $\tau_i \neq 0$, the distribution of F is non-central F with (ℓ-1,N-ℓ) degrees of freedom and noncentrality parameter $(1/\sigma^2)\sum_{j=1}^{\ell} n_j\tau_j^2$. This distribution determines the power of the F-test above; see, for example, Tiku (1967). It may be noted in passing that the F-statistic mentioned above is essentially the likelihood-ratio statistic.

3.4 ANOVA Using the MML Estimators

The Analysis of Variance procedure outlined in Section 3.2 is well known and constitutes an optimal procedure if the underlying normality assumption is correct. In practice, however, one often finds situations in which normality does not hold; in such cases the classical ANOVA procedure above can lead to incorrect or misleading inferences. We therefore will employ the method of Tiku and Suresh (1992) discussed in Chapter 2, and will develop ANOVA procedures for models of the form (3.2.1).

Consider the same model as (3.2.1), with the ε_{ij}'s having the distribution (2.3.1):

$$y_{ij} = \tau^*_i + e_{ij} \qquad (i = 1,\dots,\ell;\ j = 1,\dots,n_i) \tag{3.4.1}$$

The likelihood function based on the model (3.4.1) is then (after ordering the observations in treatment i with respect to j):

$$L \propto \left(\frac{1}{\sigma\sqrt{k}}\right)^N \prod_{i=1}^{\ell}\prod_{j=1}^{n_i}\left\{1 + \frac{(y_{(ij)} - \tau^*_i)^2}{k\sigma^2}\right\}^{-p}$$

Now define $z_{(ij)} = [y_{(ij)} - \tau_i^*]/\sigma^2$. Hence, after taking log,

$$\ln L = \text{const.} - N\ln\sigma - p\sum_{i=1}^{\ell}\sum_{j=1}^{n_i} \ln\left\{1 + \frac{z^2_{(ij)}}{k}\right\} \tag{3.4.2}$$

where $N = \sum_i n_i$. Differentiating (3.4.2) with respect to τ_i^* and σ, we obtain:

$$\frac{\partial \ln L}{\partial \tau^*_i} = \frac{2p}{k\sigma}\sum_{j=1}^{n_i} g(z_{(ij)}) = 0, \qquad i = 1,\dots,\ell \tag{3.4.3}$$

$$\frac{\partial \ln L}{\partial \sigma} = -\frac{N}{\sigma} + \frac{2p}{k\sigma}\sum_{i=1}^{\ell}\sum_{j=1}^{n_i} z_{(ij)}\, g(z_{(ij)}) = 0 \tag{3.4.4}$$

where $g(z) = z/[1 + z^2/k]$. Equations (3.4.3) and (3.4.4) do not yield explicit solutions due to the presence of g. As in Chapter 1, we approximate g by a linear function as follows:

$$g(z_{(ij)}) = \gamma_{ij} + \delta_{ij} z_{(ij)} \tag{3.4.5}$$

Note that this approximation has been extended for each observation in each group i. Now, since the t-family is symmetric,

$$\sum_{j=1}^{n_i} \gamma_{ij} = 0 \qquad \text{for each } i \tag{3.4.6}$$

Substituting (3.4.5) into (3.4.3) and (3.4.4) gives:

$$\frac{\partial \ln L^*}{\partial \tau_i^*} = \frac{2p}{k\sigma} \sum_{j=1}^{n_i} [\gamma_{ij} + \delta_{ij} z_{(ij)}] = 0, \qquad i = 1, \ldots, \ell \tag{3.4.7}$$

$$\frac{\partial \ln L^*}{\partial \sigma} = -\frac{N}{\sigma} + \frac{2p}{k\sigma} \sum_{i=1}^{\ell} \sum_{j=1}^{n_i} z_{(ij)} [\gamma_{ij} + \delta_{ij} z_{(ij)}] = 0 \tag{3.4.8}$$

Algebraic manipulation of (3.4.7) and (3.4.8) along with an application of (3.4.6) produces the MML estimators:

$$\hat{\tau}_i^* = \frac{1}{m_i} \sum_{j=1}^{n_i} \delta_{ij} y_{(ij)} = \bar{y}_i \tag{3.4.9}$$

$$\hat{\sigma} = \frac{B + \sqrt{B^2 + 4NC}}{2\sqrt{N(N-\ell)}} \quad \text{(corrected for bias)} \tag{3.4.10}$$

$$\approx \sqrt{\frac{C}{N-\ell}} \quad \text{since } \frac{B}{\sqrt{NC}} \text{ will be negligibly small.}$$

where

$$m_i = \sum_{j=1}^{n_i} \delta_{ij} \tag{3.4.11}$$

$$B = \frac{2p}{k} \sum_{i=1}^{\ell} \sum_{j=1}^{n_i} \gamma_{ij} \{y_{(ij)} - \bar{y}_i\} \tag{3.4.12}$$

$$C = \frac{2p}{k} \sum_{i=1}^{\ell} \sum_{j=1}^{n_i} \delta_{ij} \{y_{(ij)} - \bar{y}_i\}^2$$

$$= \frac{2p}{k} \sum_{i=1}^{\ell} \left\{ \sum_{j=1}^{n_i} \delta_{ij} y_{(ij)}^2 - m_i \bar{y}_i^2 \right\} \tag{3.4.13}$$

It should be noted here that when $p = \infty$, the t-family reduces to normal, in which case $\gamma_{ij} = 0$ and $\delta_{ij} = 1$. Consequently, the estimators above revert back to the corresponding classical estimators (based on normality).

3.5 Variance of the MML Estimators

Differentiating the log-likelihood twice, we obtain:

$$\frac{\partial^2 \ln L^*}{\partial \tau_i^{*2}} = \frac{2p}{k\sigma} \sum_{j=1}^{n_i} \delta_{ij}\left(-\frac{1}{\sigma}\right) = -\frac{2p}{k\sigma^2} m_i \qquad i = 1, \ldots, \ell$$

$$\frac{\partial^2 \ln L^*}{\partial \tau_i^* \partial \tau_q^*} = 0 \qquad \text{for } i \neq q$$

$$\frac{\partial^2 \ln L^*}{\partial \tau_i^* \partial \sigma} = -\frac{2p}{k\sigma^2} \sum_{j=1}^{n_i} [\gamma_{ij} + 2\delta_{ij} z_{(ij)}]$$

$$\frac{\partial^2 \ln L^*}{\partial \sigma^2} = \frac{N}{\sigma^2} - \frac{2p}{k\sigma^2} \sum_{i=1}^{\ell} \sum_{j=1}^{n_i} [2\gamma_{ij} z_{(ij)} + 3\delta_{ij} z_{(ij)}^2]$$

Taking negative expectation gives:

$$-E\left(\frac{\partial^2 \ln L^*}{\partial \tau_i^{*2}}\right) = \frac{2p}{k\sigma^2} m_i \qquad i = 1, \ldots, \ell$$

$$-E\left(\frac{\partial^2 \ln L^*}{\partial \tau_i^* \partial \tau_q^*}\right) = 0 \quad \text{for } i \neq q; \qquad -E\left(\frac{\partial^2 \ln L^*}{\partial \tau_i^* \partial \sigma}\right) = 0 \quad \text{by symmetry}$$

$$-E\left(\frac{\partial^2 \ln L^*}{\partial \sigma^2}\right) = -\frac{N}{\sigma^2} + \frac{2p}{k\sigma^2} \sum_{i=1}^{\ell} \sum_{j=1}^{n_i} [2\gamma_{ij} t_{(ij)} + 3\delta_{ij}(t_{(ij)}^2 + \sigma_{j,j:n_i})]$$

Thus the information matrix is given by

$$\underline{I}(\tau_1^*,\ldots,\tau_\ell^*,\sigma) = \frac{2p}{k\sigma^2}\begin{bmatrix} m_1 & 0 & \cdots & \cdots & 0 \\ & m_2 & 0 & \cdots & 0 \\ & & \ddots & & \vdots \\ & & & m_\ell & 0 \\ & & & & \Omega \end{bmatrix}$$

$$\text{where} \quad \frac{2p}{k\sigma^2}\,\Omega = -E\left(\frac{\partial^2 \ln L^*}{\partial\sigma^2}\right)$$

Therefore the asymptotic variance-covariance matrix is

$$\underline{V}(\tau_1^*,\ldots,\tau_\ell^*,\sigma) = \frac{k\sigma^2}{2p}\begin{bmatrix} \frac{1}{m_1} & 0 & \cdots & \cdots & 0 \\ & \frac{1}{m_2} & 0 & \cdots & 0 \\ & & \ddots & & \vdots \\ & & & \frac{1}{m_\ell} & 0 \\ & & & & \frac{1}{\Omega} \end{bmatrix}$$

Also, $V(\hat{\mu}) = V(\bar{\bar{y}}) = (k\sigma^2/2p)M$. As mentioned before, $V(\hat{\sigma}^2)$ will be approximated by MVB(σ), thus eliminating the need for tables of variances of t-family order statistics:

$$\text{MVB}(\sigma) = \frac{p+1}{2(N-\ell)(p-\tfrac{1}{2})}\,\sigma^2 \qquad (3.5.1)$$

3.6 The ANOVA Table Based on MML Estimators

Consider (3.4.13) (ignoring $2p/k$ for the moment):

$$\sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\delta_{ij}\{y_{(ij)} - \bar{y}_i\}^2$$

$$= \sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\delta_{ij}\{y_{(ij)} - \bar{\bar{y}} - (\bar{y}_i - \bar{\bar{y}})\}^2$$

$$= \sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\delta_{ij}\{y_{(ij)} - \bar{\bar{y}}\}^2 - \sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\delta_{ij}\{\bar{y}_i - \bar{\bar{y}}\}^2 . \qquad (3.6.1)$$

Rearranging (3.6.1) gives:

$$\sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\delta_{ij}\{y_{(ij)}-\bar{\bar{y}}\}^2$$
$$=\sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\delta_{ij}\{y_{(ij)}-\bar{y}_i\}^2 + \sum_{i=1}^{\ell} m_i\{\bar{y}_i-\bar{\bar{y}}\}^2 .$$

As in Section 3.3, we may identify these expressions as the total, treatment, and error sums-of-squares, as follows:

$$\begin{aligned} \text{SST} &\propto \sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\delta_{ij}\{y_{(ij)}-\bar{\bar{y}}\}^2 \approx \frac{k}{2p}(N-1)\hat{\sigma}_0^2 \\ \text{SS}_{\text{Treat}} &\propto \sum_{i=1}^{\ell} m_i\{\bar{y}_i-\bar{\bar{y}}\}^2 \approx \frac{k}{2p}\left\{(N-1)\hat{\sigma}_0^2-(N-\ell)\hat{\sigma}^2\right\} \\ \text{SSE} &\propto \sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\delta_{ij}\{y_{(ij)}-\bar{y}_i\}^2 \approx \frac{k}{2p}(N-\ell)\hat{\sigma}^2 \end{aligned} \tag{3.6.2}$$

The expressions given in (3.6.2) follow from the fact that terms of the form $B/\sqrt{NC}$ are very small, especially for large n_i, as mentioned earlier. Note that $\hat{\sigma}_0^2$ is the MML estimator of σ^2 obtained by equating the τ_i's to zero in the model (3.4.1). Thus the mean sums-of-squares are:

$$\text{MST} \approx \frac{2p}{k}\frac{\text{SST}}{n-1}, \quad \text{MS}_{\text{Treat}} \approx \frac{2p}{k}\frac{\text{SS}_{\text{Treat}}}{\ell-1}, \quad \text{and} \quad \text{MSE} \approx \frac{2p}{k}\frac{\text{SSE}}{N-\ell} .$$

We may now proceed to form the ANOVA Table as in Section 3.3. The F-statistic for testing the null hypothesis H_0: $\tau_i=0$ $(i=1,\ldots,\ell)$ is given by:

$$F^* = \frac{2p}{k}\frac{\sum_{i=1}^{\ell} m_i\{\bar{y}_i-\bar{\bar{y}}\}^2}{(\ell-1)\hat{\sigma}^2} . \tag{3.6.3}$$

(A variance-corrected F-statistic is given in the next section.) The null distribution of F^* is referred to a central F-distribution with $(\ell-1, N-\ell)$ degrees of freedom. The non-null distribution of F^* is referred to a non-central F-distribution with $(\ell-1, N-\ell)$ degrees of freedom and noncentrality parameter $(2p/k)\sum_{i=1}^{\ell} m_i \tau_i^2/\sigma^2$. Since $2p/k > 1$ and $m_i > n_i$, the F^*-test is more powerful than the normal-theory F-test, at least for large n_i. For $p = \infty$, of course, F^* reduces to the normal-theory F.

3.7 Improvement to F^* Based on MVB(σ)

The statistic F^* above may be improved by noting that (as in Chapter 1),

$$V(\hat{\sigma}) = \frac{(p+1)}{2(N-\ell)(p-½)}\,\sigma^2$$

i.e. the MVB(σ) provides a close approximation to $V(\hat{\sigma})$. Since $E(\hat{\sigma}) \approx \sigma$,

$$E(\hat{\sigma}^2) = \sigma^2\left\{1 + \frac{p+1}{2(N-\ell)(p-½)}\right\} . \tag{3.7.1}$$

Let us now define

$$\chi^2_{(N-\ell)} = h\hat{\sigma}^2/\sigma^2 \text{ for some constant } h . \tag{3.7.2}$$

Taking expectation on both sides of (3.7.2), we have, by applying (3.7.1):

$$N-\ell = h\left\{1 + \frac{p+1}{2(N-\ell)(p-½)}\right\}$$

$$\Rightarrow \quad h = \frac{N-\ell}{1 + \frac{p+1}{2(N-\ell)(p-½)}}$$

Thus $h\hat{\sigma}^2/\sigma^2$ is approximately a chi-squared random variable with N-ℓ degrees of freedom. Further, based on using $V(\hat{\mu}) \approx MVB(\mu)$, we have:

$$\bar{y}_i \text{ is asymptotically } N\left(\mu,\ \frac{(p-\frac{3}{2})(p+1)}{p(p-\frac{1}{2})}\ \frac{\sigma^2}{n_i}\right)$$

and therefore

$$\chi^2 = \frac{p(p-\frac{1}{2})}{(p-\frac{3}{2})(p+1)}\ \frac{\sum_{i=1}^{\ell} n_i(\bar{y}_i - \bar{\bar{y}})^2}{\sigma^2} \tag{3.7.3}$$

is approximately a chi-squared random variable with ℓ-1 degrees of freedom. Based on $h\hat{\sigma}^2/\sigma^2$ and (3.7.3), we may form a modified F-statistic similar to F^*:

$$F_m^* = \frac{p(p-\frac{1}{2})}{(p-\frac{3}{2})(p+1)}\left\{1 + \frac{p+1}{2(N-\ell)(p-\frac{1}{2})}\right\}\frac{\sum_{i=1}^{\ell} n_i(\bar{y}_i - \bar{\bar{y}})^2}{(\ell-1)\,\hat{\sigma}^2}$$

The modified statistic F_m^* may be referred to a central F-distribution with (ℓ-1,N-ℓ) degrees of freedom. Some simulations have been performed to assess the performance of the MML estimators $\hat{\tau}_i^*$ and $\hat{\sigma}$ and the modified F-statistic F_m^* for various values of p, and for small n_i. The simulation results will be reported at the end of the chapter.

There are also situations in which one wishes to test hypotheses involving linear contrasts of the treatment effects:

$$H_0:\ T = \sum_{i=1}^{\ell} a_i \tau_i^* = 0$$
$$\text{vs. } H_1:\ T \neq 0$$
$$\text{where } \sum_{i=1}^{\ell} a_i = 0\ .$$

An unbiased estimator of T is given by:

$$\hat{T} = \sum_{i=1}^{\ell} a_i \hat{\tau}^{*}_i = \sum_{i=1}^{\ell} a_i \bar{y}_i \ ,$$

and since the $\bar{y}_i$'s are independently distributed,

$$\begin{aligned} V(\hat{T}) &= \sum_{i=1}^{\ell} a_i^2 V(\bar{y}_i) \\ &= \frac{(p-\frac{3}{2})(p+1)}{p(p-\frac{1}{2})} \left\{ \sum_{i=1}^{\ell} \left(\frac{a_i^2}{n_i} \right) \right\} \sigma^2 \end{aligned}$$

Now based on the approximate chi-squared variable defined above, we may define the following t-statistic:

$$t^{*} = \sqrt{\frac{p(p-\frac{1}{2})}{(p-\frac{3}{2})(p+1)}} \sqrt{1+\frac{p+1}{2(N-\ell)(p-\frac{1}{2})}} \sum_{i=1}^{\ell} a_i \bar{y}_i \Big/ \hat{\sigma} \sqrt{\sum_{i=1}^{\ell} \left(\frac{a_i^2}{n_i} \right)}$$

The statistic t^{*} may then be referred to a t-distribution with N-ℓ d.f. Large values of t^{*} indicate that the contrast is not zero. Once again, simulations have been done to assess the performance of the statistic t^{*} for small n_i, and these results will be reported at the end of the chapter.

3.8 The ANOVA Procedure Based on Type-II Symmetrically Censored Samples

Once again, let us consider the one-way model given by:

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij} \qquad (i = 1,\dots,\ell;\ j = 1,\dots,n_i)$$
$$= \tau_i^* + \varepsilon_{ij} \qquad \text{(for simplicity)} .$$

In some cases, complete samples may not be available in each treatment group; for example, the experiment may be terminated (due to cost considerations) before measurements have been obtained for each experimental unit. This naturally gives rise to Type-II censored samples in each treatment group. Therefore, let us in particular consider symmetrically censored samples, censoring r_i observations from the left and r_i from the right of the i^{th} sample, leaving:

$$\begin{array}{ccc} y_{(1,r_1+1)} & \le \cdots \le & y_{(1,n_1-r_1)} \\ \vdots & & \vdots \\ y_{(\ell,r_\ell+1)} & \le \cdots \le & y_{(\ell,n_\ell-r_\ell)} \end{array}$$

Let $a_i = r_i+1$ and $b_i = n_i-r_i$ $(i = 1,\dots,\ell)$. Then the log-likelihood function based on the ℓ censored samples is:

$$L \propto \prod_{i=1}^{\ell} \left\{ \prod_{j=a_i}^{b_i} \frac{1}{\sigma} \left[1 + \frac{[y_{(ij)} - \tau_i^*]^2}{k\sigma^2} \right]^{-p} \right.$$
$$\left. \times \left[F\left(\frac{y_{(i,a_i)} - \tau_i^*}{\sigma} \right) \right]^{r_i} \left[1 - F\left(\frac{y_{(i,b_i)} - \tau_i^*}{\sigma} \right) \right]^{r_i} \right\} \qquad (3.8.1)$$

Writing $z_{(ij)} = [y_{(ij)} - \tau_i^*]/\sigma$ and taking log, (3.8.1) becomes:

$$\ln L = \text{const.} - A\ln\sigma - p\sum_{i=1}^{\ell}\sum_{j=a_i}^{b_i}\ln(1 + z_{(ij)}^2/k)$$

$$+ \sum_{i=1}^{\ell} r_i\,\{\ln F(z_{(i,a_i)}) + \ln[1 - F(z_{(i,b_i)})]\,\} \qquad (3.8.2)$$

where $A = \sum_i (n_i - 2r_i)$ and F is the c.d.f of the t-family (2.3.1). To find maximum likelihood estimators for τ_i^*'s and σ, we would as usual differentiate (3.8.2) with respect to these parameters:

$$\frac{\partial \ln L}{\partial \tau_i^*} = \frac{2p}{k\sigma}\sum_{j=a_i}^{b_i} g(z_{(ij)})$$

$$+ \frac{1}{\sigma} r_i \left\{ \frac{f(z_{(i,b_i)})}{[1 - F(z_{(i,b_i)})]} - \frac{f(z_{(i,a_i)})}{F(z_{(i,a_i)})} \right\} = 0 \qquad (3.8.3)$$

$$\frac{\partial \ln L}{\partial \sigma} = -\frac{A}{\sigma} + \frac{2p}{k\sigma}\sum_{i=1}^{\ell}\sum_{j=a_i}^{b_i} z_{(ij)}\, g(z_{(ij)})$$

$$+ \frac{1}{\sigma}\sum_{i=1}^{\ell} r_i \left\{ z_{(i,b_i)} \frac{f(z_{(i,b_i)})}{[1 - F(z_{(i,b_i)})]} \right.$$

$$\left. - z_{(i,a_i)} \frac{f(z_{(i,a_i)})}{F(z_{(i,a_i)})} \right\} = 0 \qquad (3.8.4)$$

where g has been previously defined. As in Chapter 1, let us write $h_1(z) = f(z)/F(z)$ and $h_2(z) = f(z)/[1-F(z)]$. Then we may write (3.8.3) and (3.8.4) as:

$$\frac{\partial \ln L}{\partial \tau_i^*} = \frac{2p}{k\sigma}\sum_{j=a_i}^{b_i} g(z_{(ij)})$$

$$+ \frac{1}{\sigma} r_i\,\{ h_2(z_{(i,b_i)}) - h_1(z_{(i,a_i)}) \} = 0 \qquad (3.8.5)$$

and $$\frac{\partial \ln L}{\partial \sigma} = -\frac{A}{\sigma} + \frac{2p}{k\sigma}\sum_{i=1}^{\ell}\sum_{j=a_i}^{b_i} z_{(ij)}\, g(z_{(ij)})$$

$$+ \frac{1}{\sigma}\sum_{i=1}^{\ell} r_i\,\{ z_{(i,b_i)} h_2(z_{(i,b_i)}) - z_{(i,a_i)} h_1(z_{(i,a_i)}) \} = 0 \qquad (3.8.6)$$

Now, equations (3.8.5) and (3.8.6) do not admit explicit solutions for the parameters due to the presence of the functions g, h_1, and h_2, and are almost impossible to solve iteratively. Therefore we introduce (as before) the following approximations:

$$g(z_{(ij)}) \approx \gamma_{ij} + \delta_{ij} z_{(ij)}$$
$$h_1(z_{(ij)}) \approx \alpha_i - \beta_i z_{(ij)}$$
$$h_2(z_{(ij)}) \approx \alpha_i + \beta_i z_{(ij)} .$$

Substituting these approximations into (3.8.5) and (3.8.6), we arrive at the following modified likelihood equations:

$$\frac{\partial \ln L^*}{\partial \tau^*_i} \approx \frac{2p}{k\sigma} \sum_{j=a_i}^{b_i} [\gamma_{ij} + \delta_{ij} z_{(ij)}] + \frac{1}{\sigma} r_i \beta_i [z_{(i,a_i)} + z_{(i,b_i)}] = 0 \tag{3.8.7}$$

and
$$\frac{\partial \ln L^*}{\partial \sigma} \approx -\frac{A}{\sigma} + \frac{2p}{k\sigma} \sum_{i=1}^{t} \sum_{j=a_i}^{b_i} z_{(ij)} [\gamma_{ij} + \delta_{ij} z_{(ij)}] + \frac{1}{\sigma} \sum_{i=1}^{t} r_i \Big[\alpha_i (z_{(i,b_i)} - z_{(i,a_i)}) + \beta_i [z^2_{(i,a_i)} + z^2_{(i,b_i)}] \Big] = 0 \tag{3.8.8}$$

In order to simplify notation, we write:

$$\alpha_i = \frac{2p}{k} \alpha^*_i; \qquad \beta_i = \frac{2p}{k} \beta^*_i; \qquad A = \frac{2p}{k} A^* .$$

Now consider (3.8.7) (after dividing through by $2p/k$):

$$\frac{1}{\sigma}\sum_{j=a_i}^{b_i}\left[\gamma_{ij}+\delta_{ij}\left\{\frac{y_{(ij)}-\tau_i^*}{\sigma}\right\}\right]+\frac{1}{\sigma}r_i\beta_i^*\left[\frac{y_{(i,a_i)}-\tau_i^*}{\sigma}+\frac{y_{(i,b_i)}-\tau_i^*}{\sigma}\right]=0$$

$$\Rightarrow\quad \sigma\sum_{j=a_i}^{b_i}\gamma_{ij}+\sum_{j=a_i}^{b_i}\delta_{ij}y_{(ij)}-\tau_i^*\sum_{j=a_i}^{b_i}\delta_{ij}+r_i\beta_i^*[y_{(i,a_i)}+y_{(i,b_i)}]-2r_i\beta_i^*\tau_i^*=0$$

$$\Rightarrow\quad \tau_i^*[m_i+2r_i\beta_i^*]=\sum_{j=a_i}^{b_i}\delta_{ij}y_{(ij)}+r_i\beta_i^*[y_{(i,a_i)}+y_{(i,b_i)}]$$

Now define $m_i^*=m_i+2r_i\beta_i^*$ and $m_i\bar{y}_i=\sum_{j=a_i}^{b_i}\delta_{ij}y_{(ij)}$ and substitute:

$$m_i^*\tau_i^*=m_i\bar{y}_i+r_i\beta_i^*[y_{(i,a_i)}+y_{(i,b_i)}]$$

$$\Rightarrow\quad \hat{\tau}_i^*=\frac{1}{m_i^*}\left\{m_i\bar{y}_i+r_i\beta_i^*[y_{(i,a_i)}+y_{(i,b_i)}]\right\}$$

$$=\bar{y}_i^*,\qquad \text{for } i=1,\ldots,\ell\,. \tag{3.8.9}$$

Thus the MML estimate of the i^{th} treatment effect is the (modified) sample mean of the i^{th} censored sample. In a similar manner, we proceed to solve (3.8.8):

$$-\frac{A^*}{\sigma}+\frac{1}{\sigma}\sum_{i=1}^{\ell}\sum_{j=a_i}^{b_i}\left\{\frac{y_{(ij)}-\tau_i^*}{\sigma}\right\}\left[\gamma_{ij}+\delta_{ij}\left\{\frac{y_{(ij)}-\tau_i^*}{\sigma}\right\}\right]+$$

$$\frac{1}{\sigma}\sum_{i=1}^{\ell}r_i\left[\alpha_i^*\left(\frac{y_{(i,b_i)}-\tau_i^*}{\sigma}-\frac{y_{(i,a_i)}-\tau_i^*}{\sigma}\right)+\beta_i^*\left[\left\{\frac{y_{(i,a_i)}-\tau_i^*}{\sigma}\right\}^2+\left\{\frac{y_{(i,b_i)}-\tau_i^*}{\sigma}\right\}^2\right]\right]=0$$

$$A^*\sigma^2-\sigma\sum_{i=1}^{\ell}\left\{\sum_{j=a_i}^{b_i}\gamma_{ij}y_{(ij)}+r_i\alpha_i^*[y_{(i,b_i)}-y_{(i,a_i)}]\right\}$$

$$-\sum_{i=1}^{\ell}\left\{\sum_{j=a_i}^{b_i}\delta_{ij}[y_{(ij)}-\tau_i^*]^2+r_i\beta_i^*\left\{[y_{(i,a_i)}-\tau_i^*]^2+[y_{(i,b_i)}-\tau_i^*]^2\right\}\right\}=0$$

$$\Rightarrow\quad A^*\sigma^2-B^*\sigma-C^*=0$$

where, after substituting $\hat{\tau}_i^* = \bar{y}_i^*$,

$$B^* = \sum_{i=1}^{\ell} \left\{ \sum_{j=a_i}^{b_i} \gamma_{ij} y_{(ij)} + r_i \alpha_i^* [y_{(i,b_i)} - y_{(i,a_i)}] \right\}$$

$$C^* = \sum_{i=1}^{\ell} \left\{ \sum_{j=a_i}^{b_i} \delta_{ij} [y_{(ij)} - \bar{y}_i^*]^2 + r_i \beta_i^* \{ [y_{(i,a_i)} - \bar{y}_i^*]^2 + [y_{(i,b_i)} - \bar{y}_i^*]^2 \} \right\}$$

Hence

$$\hat{\sigma} = \frac{B^* + \sqrt{B^{*2} + 4A^*C^*}}{2A^*}$$

$$\text{or} \quad \hat{\sigma} = \frac{B^* + \sqrt{B^{*2} + 4A^*C^*}}{2\sqrt{A^*(A^* - \frac{k}{2p}\ell)}} \quad \text{(corrected for bias)}. \qquad (3.8.10)$$

3.9 Variance of the MML Estimators Based on Censored Samples

Differentiating (3.8.7) and (3.8.8) with respect to τ_i^* and σ gives:

$$\frac{\partial^2 \ln L^*}{\partial \tau_i^{*2}} = \frac{2p}{k\sigma} \sum_{j=a_i}^{b_i} \delta_{ij}\left(-\frac{1}{\sigma}\right) + \frac{1}{\sigma} r_i \beta_i \left(-\frac{2}{\sigma}\right)$$

$$= -\frac{2p}{k\sigma^2} \{ m_i + 2 r_i \beta_i^* \} = -\frac{2p}{k\sigma^2} m_i^* , \qquad i = 1, \ldots, \ell$$

$$\frac{\partial^2 \ln L^*}{\partial \tau_i^* \partial \tau_q^*} = 0 , \qquad i \neq q$$

$$\frac{\partial^2 \ln L^*}{\partial \tau_i^* \partial \sigma} = -\frac{2p}{k\sigma^2} \left\{ \sum_{j=a_i}^{b_i} [\gamma_{ij} + 2\delta_{ij} z_{(ij)}] - 2 r_i \beta_i^* [z_{(i,a_i)} + z_{(i,b_i)}] \right\}$$

$$\frac{\partial^2 \ln L^*}{\partial \sigma^2} = \frac{2p}{k\sigma^2} \left\{ A^* - \sum_{i=1}^{\ell} \sum_{j=a_i}^{b_i} [2\gamma_{ij} z_{(ij)} + 3\delta_{ij} z_{(ij)}^2] \right.$$

$$\left. + \sum_{i=1}^{\ell} r_i [2\alpha_i^* (z_{(i,b_i)} - z_{(i,a_i)}) + 3\beta_i^* (z_{(i,a_i)}^2 + z_{(i,b_i)}^2)] \right\}$$

Taking negative expectation gives:

$$-E\left(\frac{\partial^2 \ln L}{\partial \tau_i^{*2}}\right) = \frac{2p}{k\sigma^2} m_i^* , \qquad i = 1, \ldots, \ell$$

$$-E\left(\frac{\partial^2 \ln L}{\partial \tau_i^* \partial \tau_q^*}\right) = 0 , \qquad i \neq q$$

$$-E\left(\frac{\partial^2 \ln L}{\partial \tau_i^* \partial \sigma}\right) = \frac{2p}{k\sigma^2} \left\{ \sum_{j=a_i}^{b_i} [\gamma_{ij} + 2\delta_{ij} t_{(ij)}] - 2r_i \beta_i^* [t_{(i,a_i)} + t_{(i,b_i)}] \right\}$$

$$= 0 , \qquad \text{by symmetry}$$

$$-E\left(\frac{\partial^2 \ln L^*}{\partial \sigma^2}\right) = \frac{2p}{k\sigma^2} \left\{ -A^* + \sum_{i=1}^{\ell} \sum_{j=a_i}^{b_i} [2\gamma_{ij} t_{(ij)} + 3\delta_{ij}(t_{(ij)}^2 + \sigma_{j,j:n_i})] \right.$$

$$+ \sum_{i=1}^{\ell} r_i \Big[2\alpha_i^* (t_{(i,b_i)} - t_{(i,a_i)})$$

$$\left. + 3\beta_i^* [(t_{(i,a_i)}^2 + \sigma_{a_i,a_i:n_i}) + (t_{(i,b_i)} + \sigma_{b_i,b_i:n_i})] \Big] \right\}$$

$$= \Omega^*/\sigma^2 .$$

Thus the information matrix (based on censored samples) is given by:

$$\underset{\sim}{I}(\tau_1^*, \ldots, \tau_\ell^*, \sigma) = \frac{2p}{k\sigma^2} \begin{bmatrix} m_1^* & 0 & \cdots & \cdots & 0 \\ & m_2^* & 0 & \cdots & 0 \\ & & \ddots & & \vdots \\ & & & m_\ell^* & 0 \\ & & & & \Omega^* \end{bmatrix}$$

$$\text{where} \quad \frac{2p}{k\sigma^2} \Omega^* = -E\left(\frac{\partial^2 \ln L^*}{\partial \sigma^2}\right)$$

Therefore the asymptotic variance-covariance matrix is

$$\underline{V}(\tau_1^*, \dots, \tau_t^*, \sigma) = \frac{k\sigma^2}{2p} \begin{bmatrix} \frac{1}{m_1^*} & 0 & \cdots & \cdots & 0 \\ & \frac{1}{m_2^*} & 0 & \cdots & 0 \\ & & \ddots & & \vdots \\ & & & \frac{1}{m_t^*} & 0 \\ & & & & \frac{1}{\Omega^*} \end{bmatrix}.$$

Once again, $V(\hat{\mu}) = V(\bar{\bar{y}}^*) = k\sigma^2/2p\, M^*$; $V(\hat{\sigma})$ will be approximated by $MVB(\sigma)$, eliminating the need for tables of variances of order statistics from the t-family (2.3.1).

3.10 The ANOVA Table Based on Symmetrically Censored Samples

As in Section 3.6, in can be shown that C^* can be written as:

$$\sum_{i=1}^{t} \left\{ \sum_{j=a_i}^{b_i} \delta_{ij} [y_{(ij)} - \bar{y}_i^*]^2 + r_i \beta_i^* \{ [y_{(i,a_i)} - \bar{y}_i^*]^2 + [y_{(i,b_i)} - \bar{y}_i^*]^2 \} \right\}$$

$$= \sum_{i=1}^{t} \left\{ \sum_{j=a_i}^{b_i} \delta_{ij} [y_{(ij)} - \bar{\bar{y}}^*]^2 + r_i \beta_i^* \{ [y_{(i,a_i)} - \bar{\bar{y}}^*]^2 + [y_{(i,b_i)} - \bar{\bar{y}}^*]^2 \} \right\}$$

$$- \sum_{i=1}^{t} m_i^* [\bar{y}_i^* - \bar{\bar{y}}^*]^2$$

We may once again, after rearranging terms, identify these expressions as the total, treatment, and error sums-of-squares:

$$SST = \sum_{i=1}^{t} \left\{ \sum_{j=a_i}^{b_i} \delta_{ij} [y_{(ij)} - \bar{\bar{y}}^*]^2 + r_i \beta_i^* \{ [y_{(i,a_i)} - \bar{\bar{y}}^*]^2 - [y_{(i,b_i)} - \bar{\bar{y}}^*]^2 \} \right\}$$

$$SS_{Treat} = \sum_{i=1}^{t} m_i^* [\bar{y}_i^* - \bar{\bar{y}}^*]^2$$

$$SSE = \sum_{i=1}^{t} \left\{ \sum_{j=a_i}^{b_i} \delta_{ij} [y_{(ij)} - \bar{y}_i^*]^2 + r_i \beta_i^* \{ [y_{(i,a_i)} - \bar{y}_i^*]^2 + [y_{(i,b_i)} - \bar{y}_i^*]^2 \} \right\}$$

These expressions follow from the fact that terms of the form $B^*/\sqrt{(A^*C^*)}$ are very small for large $n_i\text{-}2r_i$. The mean squares are then:

$$\text{MST} = \frac{\text{SST}}{A-1}, \quad \text{MS}_{\text{Treat}} = \frac{\text{SS}_{\text{Treat}}}{\ell-1}, \quad \text{and} \quad \text{MSE} = \frac{\text{SSE}}{A-\ell} .$$

The ANOVA Table may now be constructed in the usual manner. Realize that the F-statistic based on censoring, say F_c^*, is to be referred to a central F-distribution with $(\ell\text{-}1, A\text{-}\ell)$ d.f., where A is the total censored sample size.

3.11 An Improved F-test Based on Censoring

In a similar manner to the procedure outlined in Section 3.7, we may improve the F-statistic F_c^*:

$$F_{cm}^* = \left\{ 1 + \frac{p+1}{2(A-\ell)(p-\tfrac{1}{2})} \right\} \frac{\sum_{i=1}^{\ell} M_i^* (\bar{y}_i - \bar{\bar{y}})^2}{(\ell-1)\,\hat{\sigma}^2} , \quad M_i^* = \frac{2p}{k} m_i^* .$$

Note that F_{cm}^* is similar to the modified F-statistic in Section 3.7, the main difference being that N is replaced be A, and the estimator $\hat{\sigma}$ is based on censored samples. Simulations of F_{cm}^* have been performed for samples with 10% censoring, and various values of p, and will be reported along with the complete sample results.

A similar test for the contrast:

$$\begin{aligned} \text{H}_0\text{:} \ & T = \sum_{i=1}^{\ell} a_i \tau_i^* = 0 \\ \text{vs. } \ \text{H}_1\text{:} \ & T \neq 0 \end{aligned}$$

may be constructed for censored samples by creating a modified t-statistic t_c^* as follows:

$$t_c^* = \sqrt{1+\frac{p+1}{2(A-\ell)(p-\frac{1}{2})}}\ \sum_{i=1}^{\ell} a_i \bar{y}_i \Big/ \hat{\sigma}\sqrt{\sum_{i=1}^{\ell}\left(\frac{a_i^2}{M_i^*}\right)}$$

The statistic t_c^* may now be referred to a t-distribution with A-ℓ d.f. Large values of t_c^* indicate that the contrast is not zero. Once again, simulations have been done to assess the performance of the statistic t_c^*, for 10% censoring, and various values of p.

3.12 Simulation Results and Discussion

In order to assess the small sample properties of the MML estimators $\hat{\mu}$, $\hat{\tau}_i^*$ (i = 1,...,ℓ), and $\hat{\sigma}$ (both for complete and censored samples), we have performed 10,000 Monte Carlo simulations to estimate the averages of $\hat{\mu}$, $\hat{\tau}_i^*$, and $\hat{\sigma}$ and their variances, and also the significance levels for the test statistics F_a^* and t^* (F_{ca}^* and t_c^* for censored samples).

The simulations were carried out on a VAX mainframe computer using FORTRAN-77 and IMSL subroutines. Random samples of size n_i from the t-family (2.3.1) were generated in the following manner:

(i) Generate independent $x_1, x_2, ..., x_{n_i}$ from N(0,1).

(ii) Generate independent z from χ^2_ν, where $\nu = 2p-1$.

(iii) Compute $y_j = \dfrac{\sqrt{k}\,x_j}{\sqrt{z}}$, $j = 1, ..., n_i$.

This results in the random sample

$$y_1, y_2, \ldots, y_{n_i}$$

from the distribution (2.3.1).

In Table V, we report the simulated means of $\hat{\mu}$, $\hat{\tau}_i^*$, and $\hat{\sigma}$, and their simulated and asymptotic variances (scaled by σ^2) based on complete samples. Three groups ($\ell = 3$) were used, with a sample of size $n_i = 10$ in each group. The actual parameter values were $\mu = 0$, $\tau_i^* = \mu + \tau_i = 0$, and $\sigma = 1$. Simulations were performed for $p = 4$, 5, and 10.

From Table V it is evident that the MML estimators perform remarkably well, even for p as small as 4. Realize that these estimators are based on asymptotic (large sample) approximations; the fact that these estimators perform so well for sample sizes of only $n_i = 10$ is very gratifying.

Comparing the simulated and asymptotic variances of the estimators, we see that the simulated variances are extremely close to the corresponding MVB's. This clearly shows that the MML estimators are highly efficient, even for small samples.

The estimators have also been simulated for 10% censored samples ($A_i = n_i - 2 = 8$) for $p = 2$, 3, 4, 5, and 10. Note that, as mentioned in Section 2.11, for $p < 3.5$ censoring must be employed to ensure that $\hat{\sigma}$ is positive. The results are reported in Table VI. As in the case of complete samples, it can be seen that the estimators $\hat{\mu}$, $\hat{\tau}_i^*$, and $\hat{\sigma}$ are quite good, and display remarkable efficiency, even for small (censored) sample sizes $A_i = 8$.

Based on the estimators discussed above, we have also simulated the significance level of the test statistics F_a^* and t^* (and F_{ca}^* and t_c^* for 10%

censoring), and have reported these values in Table VII. Compared to the actual significance level ($\alpha = 0.05$), it is clear that the test statistics F_m^* and t^* essentially attain the true significance level 0.05. The significance levels of the test statistics F_{cm}^* and t_c^* also are close to the presumed significance level $\alpha = 0.05$, even when p is very small. Moreover, for a given value of p, the test statistics based on censoring seem to hold the significance level better that those for complete samples. This is an indication that employing censoring will enhance robustness (Tiku, 1980). Also, it is gratifying indeed that the approximations presented in Sections 3.7 and 3.11 work very well, even though they are based on asymptotic considerations.

In summary, it is clear that the non-normal ANOVA procedure based on MML estimators is highly efficient for all values of p and even for small n_i, and should be employed. The ANOVA procedure above extends to higher-order designs in a sraightforward fashion, as shown in the next section.

Table V: $\hat{\mu}$, $\hat{\tau}_i^*$, and $\hat{\sigma}$ and their Asymptotic and Simulated Variances for ℓ=3, n_i=10, μ=0, τ_i=0, σ=1; p = 4, 5, 10. (Complete Samples)

p	Parameter	Estimate*	Simulated* Variance/σ^2	Asymptotic Variance/σ^2
4	μ	0.0010	0.0310	0.0293
	τ_1^*	-0.0061	0.0922	0.0879
	τ_2^*	0.0070	0.0928	0.0879
	τ_3^*	0.0019	0.0948	0.0879
	σ	1.0469	0.0311	0.0265
5	μ	-0.0019	0.0318	0.0304
	τ_1^*	-0.0036	0.0953	0.0911
	τ_2^*	-0.0016	0.0960	0.0911
	τ_3^*	-0.0004	0.0969	0.0911
	σ	1.0364	0.0273	0.0247
10	μ	0.0005	0.0324	0.0321
	τ_1^*	-0.0004	0.0988	0.0964
	τ_2^*	0.0002	0.0975	0.0964
	τ_3^*	0.0016	0.0987	0.0964
	σ	1.0129	0.0226	0.0214

* These values are based on 10,000 Monte Carlo simulations.

Table VI: $\hat{\mu}$, $\hat{\tau}_i^*$, and $\hat{\sigma}$ and their Asymptotic and Simulated Variances for $\ell=3$, $n_i=10$, $\mu=0$, $\tau_i=0$, $\sigma=1$; p = 2,3,4,5,10. (10% Censoring)

p	Parameter	Estimate*	Simulated* Variance/σ^2	Asymptotic Variance/σ^2
2	μ	-0.0008	0.0187	0.0171
	τ_1^*	-0.0021	0.0557	0.0513
	τ_2^*	0.0006	0.0562	0.0513
	τ_3^*	-0.0004	0.0567	0.0513
	σ	1.0739	0.0515	0.0476
3	μ	-0.0012	0.0277	0.0266
	τ_1^*	0.0003	0.0832	0.0798
	τ_2^*	-0.0024	0.0841	0.0798
	τ_3^*	-0.0015	0.0823	0.0798
	σ	1.0305	0.0388	0.0381
4	μ	-0.0004	0.0317	0.0297
	τ_1^*	0.0033	0.0941	0.0892
	τ_2^*	-0.0025	0.0940	0.0892
	τ_3^*	-0.0021	0.0958	0.0892
	σ	1.0183	0.0347	0.0340
5	μ	0.0013	0.0321	0.0310
	τ_1^*	-0.0016	0.0956	0.0930
	τ_2^*	0.0006	0.0958	0.0930
	τ_3^*	0.0005	0.0949	0.0930
	σ	1.0075	0.0320	0.0317
10	μ	-0.0011	0.0330	0.0331
	τ_1^*	0.0002	0.1015	0.0994
	τ_2^*	-0.0018	0.0995	0.0994
	τ_3^*	-0.0015	0.0992	0.0994
	σ	0.9911	0.0288	0.0276

* These values are based on 10,000 Monte Carlo simulations.

Table VII: Simulated Level of Significance of F_m^* and t^* for Complete Samples, and F_{cm}^* and t_c^* for 10% Censoring; $\ell=3$, $n_i=10$, $\mu=0$, $\tau_i=0$, $\sigma=1$; p = 2,3,4,5,10. (actual significance level α = 0.05)

p	$P(F_m^* > F_\alpha)$	$P(F_{cm}^* > F_\alpha)$	$P(t^* > t_\alpha)$	$P(t_c^* > t_\alpha)$
2		0.0524		0.0500
3		0.0456		0.0496
4	0.0435	0.0532	0.0479	0.0527
5	0.0449	0.0504	0.0458	0.0523
10	0.0454	0.0551	0.0490	0.0529

All values are based on 10,000 Monte Carlo simulations.

3.13 Two-Way ANOVA Using the MML Estimators

As in Section 3.4, one may develop non-normal procedures for analyzing fixed-effect two-way ANOVA models by employing the Tiku-Suresh MML estimators. To motivate this discussion, we will consider some well-known data from Box and Cox (1964), as was done by Tiku, Tan and Balakrishnan (1986, p. 184). The data gives the results of a 3×4 factorial experiment on survival times of animals exposed to 3 different poisons and 4 different treatments, with 4 animals in each cell. The results are summarized as follows:

	Treatment			
Poison	A	B	C	D
I	0.31	0.82	0.43	0.45
	0.45	1.10	0.45	0.71
	0.46	0.88	0.63	0.66
	0.43	0.72	0.76	0.62
II	0.36	0.92	0.44	0.56
	0.29	0.61	0.35	1.02
	0.40	0.49	0.31	0.71
	0.23	1.24	0.40	0.38
III	0.22	0.30	0.23	0.30
	0.21	0.37	0.25	0.36
	0.18	0.38	0.24	0.31
	0.23	0.29	0.22	0.33

We would like to develop a procedure to analyze data of this form, assuming that the underlying distribution is one of the members of the t-family (2.3.1). Specifically, let us consider the model

$$y_{ijd} = \mu + \tau_i + \phi_j + \psi_{ij} + e_{ijd} \qquad \begin{pmatrix} i = 1,\dots,\ell \\ j = 1,\dots,c \\ d = 1,\dots,n \end{pmatrix} \qquad (3.13.1)$$

where y_{ijd} is the d^{th} observation in the $(i,j)^{th}$ cell

ε_{ijd} is the error associated with y_{ijd}

τ_i represents the effect of the i^{th} level of Treatment A

ϕ_j represents the effect of the j^{th} level of Treatment B

ψ_{ij} represents the interaction between the i^{th} level of A and the j^{th} level of B

μ is the overall (grand) mean.

We note here that we are considering a balanced design (i.e. there are the same number of observations, n, in each cell). It is possible to extend this method to the case of unbalanced designs (n_{ij} observations in the $(i,j)^{th}$ cell), but this will not be considered here.

As is the case with the one-way ANOVA model, it is usually assumed that the ε_{ijd}'s are normally distributed with mean 0 and (unknown) variance σ^2. Here, we will assume **(2.3.1)** for the error distribution.

The log-likelihood function based on the two-way table is:

$$\ln L = \text{const.} - N\ln\sigma - p\sum_{i=1}^{\ell}\sum_{j=1}^{c}\sum_{d=1}^{n}\ln\left\{1 + \frac{z^2_{(ijd)}}{k}\right\} \qquad \textbf{(3.13.2)}$$

where $N = \ell cn$ and $z_{(ijd)} = [y_{(ijd)} - \mu - \tau_i - \phi_j - \psi_{ij}]/\sigma$ and the parentheses around subscripts denote ordering in the $(i,j)^{th}$ cell (i.e. with respect to index d).

Along the same lines as Section 3.4, we introduce the approximation

$$g(z_{(ijd)}) = \gamma_{ijd} + \delta_{ijd}z_{(ijd)} \qquad \textbf{(3.13.3)}$$

where $g(z) = z/[1 + z^2/k]$. Using (3.13.3), we obtain the (modified) log-likelihood derivatives:

$$\frac{\partial \ln L^*}{\partial \mu} = \frac{2p}{k\sigma} \sum_{i=1}^{\ell} \sum_{j=1}^{c} \sum_{d=1}^{n} [\gamma_{ijd} + \delta_{ijd} z_{(ijd)}] = 0 \tag{3.13.4}$$

$$\frac{\partial \ln L^*}{\partial \tau_i} = \frac{2p}{k\sigma} \sum_{j=1}^{c} \sum_{d=1}^{n} [\gamma_{ijd} + \delta_{ijd} z_{(ijd)}] = 0 \qquad (i = 1, \ldots, \ell) \tag{3.13.5}$$

$$\frac{\partial \ln L^*}{\partial \phi_j} = \frac{2p}{k\sigma} \sum_{i=1}^{\ell} \sum_{d=1}^{n} [\gamma_{ijd} + \delta_{ijd} z_{(ijd)}] = 0 \qquad (j = 1, \ldots, c) \tag{3.13.6}$$

$$\frac{\partial \ln L^*}{\partial \psi_{ij}} = \frac{2p}{k\sigma} \sum_{d=1}^{n} [\gamma_{ijd} + \delta_{ijd} z_{(ijd)}] = 0 \qquad \begin{pmatrix} i = 1, \ldots, \ell \\ j = 1, \ldots, c \end{pmatrix} \tag{3.13.7}$$

$$\frac{\partial \ln L^*}{\partial \sigma} = -\frac{N}{\sigma} + \frac{2p}{k\sigma} \sum_{i=1}^{\ell} \sum_{j=1}^{c} \sum_{d=1}^{n} z_{(ijd)} [\gamma_{ijd} + \delta_{ijd} z_{(ijd)}] = 0 \tag{3.13.8}$$

Let us now define the following expressions:

$$m = \sum_{d=1}^{n} \delta_{ijd} \tag{3.13.9}$$

$$M = \sum_{i=1}^{\ell} \sum_{j=1}^{c} \sum_{d=1}^{n} \delta_{ijd} = \ell c m \tag{3.13.10}$$

$$\bar{y}_{ij.} = \frac{1}{m} \sum_{d=1}^{n} \delta_{ijd} y_{(ijd)} \tag{3.13.11}$$

$$\bar{y}_{i..} = \frac{1}{cm} \sum_{j=1}^{c} m \bar{y}_{ij.} = \frac{1}{c} \sum_{j=1}^{c} \bar{y}_{ij.} \tag{3.13.12}$$

$$\bar{y}_{.j.} = \frac{1}{\ell m} \sum_{i=1}^{\ell} m \bar{y}_{ij.} = \frac{1}{\ell} \sum_{i=1}^{\ell} \bar{y}_{ij.} \tag{3.13.13}$$

$$\bar{y}_{...} = \frac{1}{\ell c} \sum_{i=1}^{\ell} \sum_{j=1}^{c} \bar{y}_{ij.} = \frac{1}{\ell} \sum_{i=1}^{\ell} \bar{y}_{i..} = \frac{1}{c} \sum_{j=1}^{c} \bar{y}_{.j.} \tag{3.13.14}$$

Again, without loss of generality, we assume:

$$\sum_{i=1}^{\ell} \tau_i = \sum_{j=1}^{c} \phi_j = \sum_{i=1}^{\ell} \psi_{ij} = \sum_{j=1}^{c} \psi_{ij} = 0 \tag{3.13.15}$$

Now using (3.13.15) and the fact that $\sum_{d=1}^{n} \gamma_{ijd} = 0$ for all i,j (by symmetry), we may proceed to solve equations (3.13.4)-(3.13.8) to obtain the MML estimators of the parameters in the model (3.13.1):

$$\hat{\mu} = \bar{y}_{...} \tag{3.13.16}$$

$$\hat{\tau}_i = \bar{y}_{i..} - \bar{y}_{...} \qquad (i = 1,\ldots,\ell) \tag{3.13.17}$$

$$\hat{\phi}_j = \bar{y}_{.j.} - \bar{y}_{...} \qquad (j = 1,\ldots,c) \tag{3.13.18}$$

$$\hat{\psi}_{ij} = \bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...} \qquad \begin{pmatrix} i = 1,\ldots,\ell \\ j = 1,\ldots,c \end{pmatrix} \tag{3.13.19}$$

and $$\hat{\sigma} = \frac{B + \sqrt{B^2 + 4NC}}{2\sqrt{N(N-\ell c)}}, \text{ corrected for bias} \approx \frac{\sqrt{C}}{\sqrt{N-\ell c}} \tag{3.13.20}$$

where

$$B = \frac{2p}{k} \sum_{i=1}^{\ell} \sum_{j=1}^{c} \sum_{d=1}^{n} \gamma_{ijd} [y_{(ijd)} - \bar{y}_{ij.}]$$

$$C = \frac{2p}{k} \sum_{i=1}^{\ell} \sum_{j=1}^{c} \sum_{d=1}^{n} \delta_{ijd} [y_{(ijd)} - \bar{y}_{ij.}]^2$$

It may be noted here that the MML estimators so produced are exactly similar in form to the results for the classical two-way ANOVA procedure based on normality; the only difference is that the row, column, cell, and grand means are replaced by their corresponding MML means (and $\hat{\sigma}$, of course, reduces to the normality-based estimate when $p \to \infty$).

3.14 The Two-Way ANOVA Table Based on MML Estimators

Since the MML parameter estimators mimic the form of the classical (normality-based) solution, it is then easy to write expressions for the sums-of-squares in the two-way ANOVA table, as follows:

$$SS_{Row} = \frac{2p}{k}\sum_{i=1}^{\ell}\sum_{j=1}^{c}\sum_{d=1}^{n}\delta_{ijd}\,[\bar{y}_{i..} - \bar{y}_{...}]^2$$

$$= cm\frac{2p}{k}\sum_{i=1}^{\ell}[\bar{y}_{i..} - \bar{y}_{...}]^2$$

$$SS_{Col} = \frac{2p}{k}\sum_{i=1}^{\ell}\sum_{j=1}^{c}\sum_{d=1}^{n}\delta_{ijd}\,[\bar{y}_{.j.} - \bar{y}_{...}]^2$$

$$= \ell m\frac{2p}{k}\sum_{j=1}^{c}[\bar{y}_{.j.} - \bar{y}_{...}]^2$$

$$SS_I = \frac{2p}{k}\sum_{i=1}^{\ell}\sum_{j=1}^{c}\sum_{d=1}^{n}\delta_{ijd}\,[\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...}]^2$$

$$= m\frac{2p}{k}\sum_{i=1}^{\ell}\sum_{j=1}^{c}[\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...}]^2$$

$$SSE = (N-\ell c)\,\hat{\sigma}^2$$

$$= C = \frac{2p}{k}\sum_{i=1}^{\ell}\sum_{j=1}^{c}\sum_{d=1}^{n}\delta_{ijd}\,[y_{(ijd)} - \bar{y}_{ij.}]^2$$

$$SST = \frac{2p}{k}\sum_{i=1}^{\ell}\sum_{j=1}^{c}\sum_{d=1}^{n}\delta_{ijd}\,[y_{(ijd)} - \bar{y}_{...}]^2 .$$

Dividing each sum-of-squares by its corresponding degrees of freedom, we obtain the mean squares as:

$$MS_{Row} = \frac{SS_{Row}}{\ell-1}; \qquad MS_{Col} = \frac{SS_{Col}}{c-1}$$

$$MS_I = \frac{SS_I}{(\ell-1)(c-1)}; \qquad MSE = \frac{SSE}{N-\ell c} .$$

We may now write the ANOVA able in its usual form:

Source	df	SS	MS	F
Treatment A (Rows)	ℓ-1	SS_{Row}	$SS_{Row}/(\ell-1)$	MS_{Row}/MSE
Treatment B (Cols)	c-1	SS_{Col}	$SS_{Col}/(c-1)$	MS_{Col}/MSE
Interaction (A×B)	$(\ell-1)(c-1)$	SS_I	$SS_I/[(\ell-1)(c-1)]$	MS_I/MSE
Error	N-ℓc	SSE	SSE/(N-ℓc)	
Total	N-1	SST	SST/(N-1)	

The three F-ratios produced have, respectively, $(\ell-1, N-\ell c)$, $(c-1, N-\ell c)$, and $((\ell-1)(c-1), N-\ell c)$ degrees of freedom and test the hypotheses that there are no Row, Column, or Interaction effects.

Now based on considerations discussed in Section 3.7, the F-ratios given above may be modified (using minimum variance bounds) to give more accurate results:

$$F_R = \frac{p(p-\frac{1}{2})}{(p-\frac{3}{2})(p+1)}\left\{1+\frac{p+1}{2(N-\ell c)(p-\frac{1}{2})}\right\}\frac{cn\sum_{i=1}^{\ell}[\bar{y}_{i..}-\bar{y}_{...}]^2}{(\ell-1)\,\hat{\sigma}^2}$$

$$F_C = \frac{p(p-\frac{1}{2})}{(p-\frac{3}{2})(p+1)}\left\{1+\frac{p+1}{2(N-\ell c)(p-\frac{1}{2})}\right\}\frac{\ell n\sum_{j=1}^{c}[\bar{y}_{.j.}-\bar{y}_{...}]^2}{(c-1)\,\hat{\sigma}^2}$$

$$F_I = \frac{p(p-\frac{1}{2})}{(p-\frac{3}{2})(p+1)}\left\{1+\frac{p+1}{2(N-\ell c)(p-\frac{1}{2})}\right\}\frac{n\sum_{i=1}^{\ell}\sum_{j=1}^{c}[\bar{y}_{ij.}-\bar{y}_{i..}-\bar{y}_{.j.}+\bar{y}_{...}]^2}{(\ell-1)(c-1)\,\hat{\sigma}^2}$$

The three F-statistics F_R, F_C, and F_I given above are then referred to central F-distributions with $(\ell-1, N-\ell c)$, $(c-1, N-\ell c)$, and $((\ell-1)(c-1), N-\ell c)$ degrees of freedom, as above.

We may note here that, as in Section 3.5, it is possible to twice differentiate the log-likelihood function with respect to the model parameters to determine the asymptotic variances of the MML estimators. For brevity, these results have not been included here.

3.15 The Two-Way ANOVA Based on MML Estimators for Censored Samples

Similar to the one-way ANOVA case, the results just mentioned may be extended to the case of symmetrically censored samples. Specifically, we consider censoring r observations from the left and right of each ordered sample, leaving

$$y_{(ij,a)} \le y_{(ij,a+1)} \le \cdots \le y_{(ij,b)}$$

in the $(i,j)^{th}$ cell. Here $a = r+1$ and $b = n-r$. The log-likelihood based on the ℓc censored samples is now:

$$\ln L = \text{const.} - A\ln\sigma - p\sum_{i=1}^{\ell}\sum_{j=1}^{c}\sum_{d=a}^{b}\ln\left\{1 + \frac{z^2_{(ijd)}}{k}\right\}$$
$$+ \sum_{i=1}^{\ell}\sum_{j=1}^{c} r\,\{\ln F(z_{(ija)}) + \ln[1 - F(z_{(ijb)})\}\qquad (3.15.1)$$

We now introduce the following approximations (as in Section 3.8):

$$\begin{aligned} g(z_{(ijd)}) &\simeq \gamma_{ijd} + \delta_{ijd} z_{(ijd)} \\ h_1(z_{(ijd)}) &\simeq \alpha - \beta z_{(ijd)} \\ h_2(z_{(ijd)}) &\simeq \alpha + \beta z_{(ijd)} \end{aligned}\qquad (3.15.2)$$

where the functions h_1 and h_2 have been defined previously. Now differentiating (3.15.1) with respect to the model parameters and substituting in the approximations (3.15.2) gives the following modified likelihood derivatives:

$$\frac{\partial \ln L^*}{\partial \mu} = \sum_{i=1}^{\ell}\sum_{j=1}^{c}\sum_{d=a}^{b}[\gamma_{ijd} + \delta_{ijd} z_{(ijd)}]$$

$$+ \frac{1}{\sigma}\sum_{i=1}^{\ell}\sum_{j=1}^{c} r\beta^* [z_{(ija)} + z_{(ijb)}] = 0 \qquad (3.15.3)$$

$$\frac{\partial \ln L^*}{\partial \tau_i} = \sum_{j=1}^{c}\sum_{d=a}^{b}[\gamma_{ijd} + \delta_{ijd} z_{(ijd)}]$$

$$+ \frac{1}{\sigma}\sum_{j=1}^{c} r\beta^* [z_{(ija)} + z_{(ijb)}] = 0 \qquad (i = 1,\ldots,\ell) \qquad (3.15.4)$$

$$\frac{\partial \ln L^*}{\partial \phi_j} = \sum_{i=1}^{\ell}\sum_{d=a}^{b}[\gamma_{ijd} + \delta_{ijd} z_{(ijd)}]$$

$$+ \frac{1}{\sigma}\sum_{i=1}^{\ell} r\beta^* [z_{(ija)} + z_{(ijb)}] = 0 \qquad (j = 1,\ldots,c) \qquad (3.15.5)$$

$$\frac{\partial \ln L^*}{\partial \psi_{ij}} = \sum_{i=1}^{\ell}\sum_{d=a}^{b}[\gamma_{ijd} + \delta_{ijd} z_{(ijd)}]$$

$$+ \frac{1}{\sigma} r\beta^* [z_{(ija)} + z_{(ijb)}] = 0 \qquad \begin{pmatrix} i = 1,\ldots,\ell \\ j = 1,\ldots,c \end{pmatrix} \qquad (3.15.6)$$

$$\frac{\partial \ln L^*}{\partial \sigma} = -\frac{A^*}{\sigma} + \sum_{i=1}^{\ell}\sum_{j=1}^{c}\sum_{d=a}^{b} z_{(ijd)}[\gamma_{ijd} + \delta_{ijd} z_{(ijd)}]$$

$$+ \frac{1}{\sigma}\sum_{i=1}^{\ell}\sum_{j=1}^{c} r\,[\alpha^*(z_{(ijb)} - z_{(ija)}) + \beta^*(z^2_{(ija)} + z^2_{(ijb)})] = 0 \qquad (3.15.7)$$

where we have written:

$$\alpha = \frac{2p}{k}\alpha^*; \qquad \beta = \frac{2p}{k}\beta^*; \qquad \text{and } A = \frac{2p}{k}A^* .$$

and $A = \ell c(n-2r)$. Let us now define the following expressions (based on censored samples) analogous to those in Section 3.13:

$$m = \sum_{d=a}^{b}\delta_{ijd}\,; \qquad m^* = m + 2r\beta^* \qquad (3.15.8)$$

$$M^* = \ell c m^* \qquad (3.15.9)$$

$$\bar{y}^*_{ij.} = \frac{1}{m^*}\left\{\sum_{d=a}^{b}\delta_{ijd} y_{(ijd)} + r\beta^* [y_{(ija)} + y_{(ijb)}]\right\} \qquad (3.15.10)$$

$$\bar{y}^{*}_{i..} = \frac{1}{c\,m^{*}} \sum_{j=1}^{c} m^{*}\,\bar{y}^{*}_{ij.} = \frac{1}{c}\sum_{j=1}^{c} \bar{y}^{*}_{ij.} \qquad (3.15.11)$$

$$\bar{y}^{*}_{.j.} = \frac{1}{\ell\,m^{*}} \sum_{i=1}^{\ell} m^{*}\,\bar{y}^{*}_{ij.} = \frac{1}{\ell}\sum_{i=1}^{\ell} \bar{y}^{*}_{ij.} \qquad (3.15.12)$$

$$\bar{y}^{*}_{...} = \frac{1}{\ell c} \sum_{i=1}^{\ell}\sum_{j=1}^{c} \bar{y}^{*}_{ij.} = \frac{1}{\ell}\sum_{i=1}^{\ell} \bar{y}^{*}_{i..} = \frac{1}{c}\sum_{j=1}^{c} \bar{y}^{*}_{.j.} \qquad (3.15.13)$$

Now, using the (3.13.15) and the fact that $\sum_{d=a}^{b} \gamma_{ijd} = 0$ for all i,j (by symmetry of censoring), we may solve (3.13.3)-(3.13.7) to obtain the MML estimators for censored samples:

$$\hat{\mu} = \bar{y}^{*}_{...} \qquad (3.15.14)$$

$$\hat{\tau}_{i} = \bar{y}^{*}_{i..} - \bar{y}^{*}_{...} \qquad (i = 1,\dots,\ell) \qquad (3.15.15)$$

$$\hat{\phi}_{j} = \bar{y}^{*}_{.j.} - \bar{y}^{*}_{...} \qquad (j = 1,\dots,c) \qquad (3.15.16)$$

$$\hat{\psi}_{ij} = \bar{y}^{*}_{ij.} - \bar{y}^{*}_{i..} - \bar{y}^{*}_{.j.} + \bar{y}^{*}_{...} \qquad \begin{pmatrix} i = 1,\dots,\ell \\ j = 1,\dots,c \end{pmatrix} \qquad (3.15.17)$$

and $$\hat{\sigma} = \frac{B^{*} + \sqrt{B^{*2} + 4A^{*}C^{*}}}{2\sqrt{A^{*}\left(A^{*} - \left(\frac{k}{2p}\right)\ell c\right)}}, \text{ corrected for bias} \qquad (3.15.18)$$

$$\approx \frac{\sqrt{C^{*}}}{\sqrt{A^{*} - \left(\frac{k}{2p}\right)\ell c}}$$

where

$$B^{*} = \sum_{i=1}^{\ell}\sum_{j=1}^{c}\sum_{d=a}^{b} \gamma_{ijd}\,[y_{(ijd)} - \bar{y}^{*}_{ij.}] + r\alpha^{*}\sum_{i=1}^{\ell}\sum_{j=1}^{c} [y_{(ijb)} - y_{(ija)}]$$

$$C^{*} = \sum_{i=1}^{\ell}\sum_{j=1}^{c}\sum_{d=a}^{b} \delta_{ijd}\,[y_{(ijd)} - \bar{y}^{*}_{ij.}]^{2} + r\beta^{*}\sum_{i=1}^{\ell}\sum_{j=1}^{c} \left\{ [y_{(ija)} - \bar{y}^{*}_{ij.}]^{2} + [y_{(ijb)} - \bar{y}^{*}_{ij.}]^{2} \right\} .$$

As in the case of complete samples, one can define the following sums-of-squares:

$$SS_{Row} = cm^* \sum_{i=1}^{\ell} [\bar{y}_{i..}^* - \bar{y}_{...}^*]^2$$

$$SS_{Col} = \ell m^* \sum_{j=1}^{c} [\bar{y}_{.j.}^* - \bar{y}_{...}^*]^2$$

$$SS_I = m^* \sum_{i=1}^{\ell} \sum_{j=1}^{c} [\bar{y}_{ij.}^* - \bar{y}_{i..}^* - \bar{y}_{.j.}^* + \bar{y}_{...}^*]^2$$

$$\begin{aligned} SSE &= (A^* - (\tfrac{k}{2p})\ell c)\,\hat{\sigma}^2 \\ &= C^* = \sum_{i=1}^{\ell} \sum_{j=1}^{c} \sum_{d=a}^{b} \delta_{ijd} [y_{(ijd)} - \bar{y}_{ij.}^*]^2 \\ &\quad + r\beta^* \sum_{i=1}^{\ell} \sum_{j=1}^{c} \{[y_{(ija)} - \bar{y}_{ij.}^*]^2 + [y_{(ijb)} - \bar{y}_{ij.}^*]^2\} \end{aligned}$$

$$\begin{aligned} SST &= \sum_{i=1}^{\ell} \sum_{j=1}^{c} \sum_{d=a}^{b} \delta_{ijd} [y_{(ijd)} - \bar{y}_{...}^*]^2 \\ &\quad + r\beta^* \sum_{i=1}^{\ell} \sum_{j=1}^{c} \{[y_{(ija)} - \bar{y}_{...}^*]^2 + [y_{(ijb)} - \bar{y}_{...}^*]^2\} \end{aligned}$$

The corresponding mean squares are:

$$MS_{Row} = \frac{SS_{Row}}{\ell - 1}; \qquad MS_{Col} = \frac{SS_{Col}}{c - 1}$$

$$MS_I = \frac{SS_I}{(\ell - 1)(c - 1)}; \qquad MSE = \frac{SSE}{A - \ell c} .$$

A similar ANOVA table may now be formed, noting that SST and SSE now have A-1 and A-ℓc degrees of freedom respectively. We also may form modified F-ratios, say F_R^*, F_C^*, and F_I^*, similar to those given in Section 3.14, the main difference being that N is replaced by A in the corresponding expressions, and $\hat{\sigma}^2$ is based on censored samples. Of course, when there is no censoring ($r = 0$), all results reduce to those obtained by the MML procedure for complete samples.

3.16 Analysis of the Box-Cox Data

To demonstrate the application of the MML-based Two-Way ANOVA procedure that has been outlined in Sections 3.13-3.15, we will analyze the aforementioned data from Box and Cox (1964). As Tiku, Tan, and Balakrishnan (1986) have noted, various authors including Brown (1975) have suggested employing the transformation $x \rightarrow 1/x$ to reduce the interaction. Following this lead, we will present analyses of both the original and transformed data using the procedure developed above, with the t-family **(2.3.1)** being assumed for the error distribution.

For purposes of comparison, we present the results of the classical Analysis of Variance performed on the original and transformed data, as well as results of the MML procedure based on normality performed by Tiku, Tan, and Balakrishnan (1986, p. 185-186). These are presented in Table VIII.

As noted by Tiku, Tan, and Balakrishnan (1986), these results support the conclusion that the main effects are highly significant as shown by other authors including Brown (1975) and Schrader and McKean (1977). It is also noted that the classical ANOVA procedure fails to detect the interaction in the original data, though it is known to exist (Schrader and McKean, 1977). The MML procedure based on normality rectifies this difficulty by producing an F-value of 3.45 for the interaction effect in the original data, which is significant at the 5% level.

Table VIII: Analysis of the Box-Cox Data using the classical ANOVA procedure and the MML ANOVA procedure based on normality (for the original and transformed data)

Classical ANOVA Procedure:

	Original Data				Transformed Data			
Source	SS	df	MS	F	SS	df	MS	F
Poison	1.033	2	0.517	23.3	34.88	2	17.44	72.7
Treatment	0.921	3	0.307	13.8	20.41	3	6.80	28.3
Interaction	0.250	6	0.042	1.9	1.57	6	0.26	1.1
Error	0.801	36	0.022	–	8.64	36	0.24	–
	$\hat{\sigma} = 0.149$				$\hat{\sigma} = 0.490$			

MML ANOVA Procedure (based on censored normal samples):

	Original Data				Transformed Data			
Source	SS	df	MS	F	SS	df	MS	F
Poison	0.90	2	0.45	62.5	32.882	2	16.44	151.8
Treatment	0.696	3	0.232	32.2	16.066	3	5.355	49.4
Interaction	0.153	6	0.0255	3.54	1.190	6	0.198	1.83
Error	0.0867	12	0.072	–	1.299	12	0.1083	–
	$\hat{\sigma} = 0.085$				$\hat{\sigma} = 0.329$			

We will now proceed to re-analyze the data using the above MML procedure (for complete and censored samples) based on the t-family (2.3.1). In order to employ the procedure, it is necessary to know the value of the shape parameter p in the distribution (2.3.1). To obtain an estimate of p for this data set, the above MML procedure was carried out for $p = 4.0(0.5)10.0$, and the optimal value of p was chosen which minimizes the MSE in the model. For example the MSE for various values of p for the complete, untransformed data set are:

p	4.0	6.0	8.0	9.0	10.0	∞
MSE	0.0268	0.0250	0.0242	0.0240	0.0238	0.0220

For all cases done (original and transformed data, complete and censored samples), $p = \infty$ minimized the MSE. For illustration, however, we take $p = 10$ as the optimal value. (For $p > 10$, tables of the expected values and variances of order statistics from (2.3.1) are not available, however when p becomes large the distribution (2.3.1) becomes very close to normal anyway.)

In Table IX we present, then, the results of the above procedure with $p = 10$ for the original and transformed data.

Table IX: **Analysis of the Box-Cox Data using the Tiku-Suresh ANOVA procedure for complete samples using $p = 10$. (for the original and transformed data)**

	Original Data				Transformed Data			
Source	SS	df	MS	F^*	SS	df	MS	F^*
Poison	0.936	2	0.468	22.36	31.82	2	15.91	69.21
Treatment	0.825	3	0.275	13.14	18.34	3	6.11	26.60
Interaction	0.220	6	0.037	1.75	1.40	6	0.234	1.02
Error	0.856	36	0.024	–	9.399	36	0.261	–
	$\hat{\sigma} = 0.154$				$\hat{\sigma} = 0.511$			

The first thing to be noted in Table IX is that the above MML procedure appears to be reproducing the classical ANOVA result. That was to be expected since the distribution (2.3.1) with $p = 10$ is almost indistinguishable from a normal distribution. The reason that Tiku's MML procedure is successful here in locating the interaction effects, and not the above procedure, seems to be due to the fact that the Box-Cox data possibly contains outliers. Consequently, the family (2.3.1) fails to provide an adequate model for this data.

Chapter 4

4.1 Introduction: Regression Analysis

Like the Analysis of Variance, Regression Analysis is often used to study data resulting from scientific experiments. However, unlike the ANOVA procedure which can only show if a relationship exists, Regression Analysis seeks to uncover the actual form of the relationship between the response variable and the independent variable(s). Returning to our fertilizer example of Chapter 2, for example, say an ANOVA procedure has been done and has shown that Fertilizer A has the most effect on the height of plants. Knowing this, the experimenter might wish to determine how the growth of plants changes with changes in the amount of fertilizer administered. In other words, we seek a *model* which relates the independent variable (amount of fertilizer) to the response variable (plant height). After a model has been fitted and shown to be a good description of the pattern in the data, it may be used to predict the height of plants for a new amount of fertilizer. Realize that this model could take many functional forms such as linear, quadratic, exponential, etc. The simplest and most often used model is the *Linear Regression Model*, which will be the focus of this and subsequent chapters.

It must be noted that the model resulting from a Regression Analysis only approximates the true relationship (assuming one exists), taking into account as much of the variability in the data as possible. The fact that this model is only an approximation is noted by Draper and Smith in their introduction to *Applied Regression Analysis* (1981):

> " In any system in which variable quantities change, it is of interest to examine the effects that some variables exert (or appear to exert) on others. There may in fact be a simple functional relationship between variables; in most physical processes this is the exception rather than the rule. Often there exists a functional relationship which is too complicated to grasp or to describe in simple terms. In this case we may wish to approximate this functional relationship by some simple mathematical function, such as a polynomial, which contains the appropriate variables and which graduates or approximates to the true function over some limited range of the variables involved. By examining such a graduating function we may be able to learn more about the true relationship and to appreciate the separate and joint effects produced by changes in certain important variables."

Regression Analysis is the study of mathematically optimal procedures for determining this approximating function. Usually, a model is proposed for the form of the approximating function (such as linear regression, multiple linear regression, quadratic regression, etc.) which includes unknown parameters to be estimated and a distributional assumption for the errors. Typically (as is the case with the classical ANOVA procedure), the errors are assumed to be normally distributed with mean 0 and (unknown) variance σ^2. As in Chapter 2, we wish to extend existing regression procedures to include cases where the error distribution comes from some class of symmetric non-normal distributions. Specifically, we will consider the t-family of distributions (2.3.1) and will extend the Tiku-Suresh MML method to include linear and multiple linear regression models. We will show, using asymptotic results and Monte Carlo simulations, that the regression procedure so obtained is highly efficient. Before presenting these results, let us first examine the classical regression problem with its usual (classical) solution.

4.2 Simple Linear Regression Model

Let us consider the simple linear regression model:

$$y_{ij} = \mu + \phi u_i + \varepsilon_{ij} \qquad (i = 1,\dots,\ell;\ j = 1,\dots,n_i) \qquad (4.2.1)$$

where:

- y_{ij} denotes the j^{th} observation on the response variable at the i^{th} level of the design variable
- u_i denotes the i^{th} level of the design variable
- ϕ is the regression coefficient
- ε_{ij} is the random error inherent in y_{ij}

and $\mu + \phi\bar{u}$ is the overall (grand) mean; $\bar{u} = \sum n_i u_i / N$.

In the context of regression analysis, ϕ is referred to as the *slope* of the regression line, and μ is the *y-intercept*. It is generally assumed that the ε_{ij}'s are independent and identically distributed as $N(0,\sigma^2)$. Let us suppose we have ℓ random samples of sizes n_i ($i = 1,\dots,\ell$) representing the observations on the response variable y at the ℓ levels u_i ($i = 1,\dots,\ell$) of the design variable u. Thus, we have the following data matrix:

	Level of u				
	u_1	u_2	$\cdots$ u_i	$\cdots$	u_ℓ
Response y:	y_{11}	y_{21}	$\cdots$ y_{i1}	$\cdots$	$y_{\ell 1}$
	y_{12}	y_{22}	$\cdots$ y_{i2}	$\cdots$	$y_{\ell 2}$
	$\vdots$	$\vdots$	$\vdots$		$\vdots$
	y_{1n_1}	y_{2n_2}	$\cdots$ y_{in_i}	$\cdots$	$y_{\ell n_\ell}$

Based on these random samples, we wish to obtain the best fitting line $\hat{y} = \hat{\mu} + \hat{\phi}u$ which captures as much of the linear trend evident in the data as possible. We may represent this graphically as:

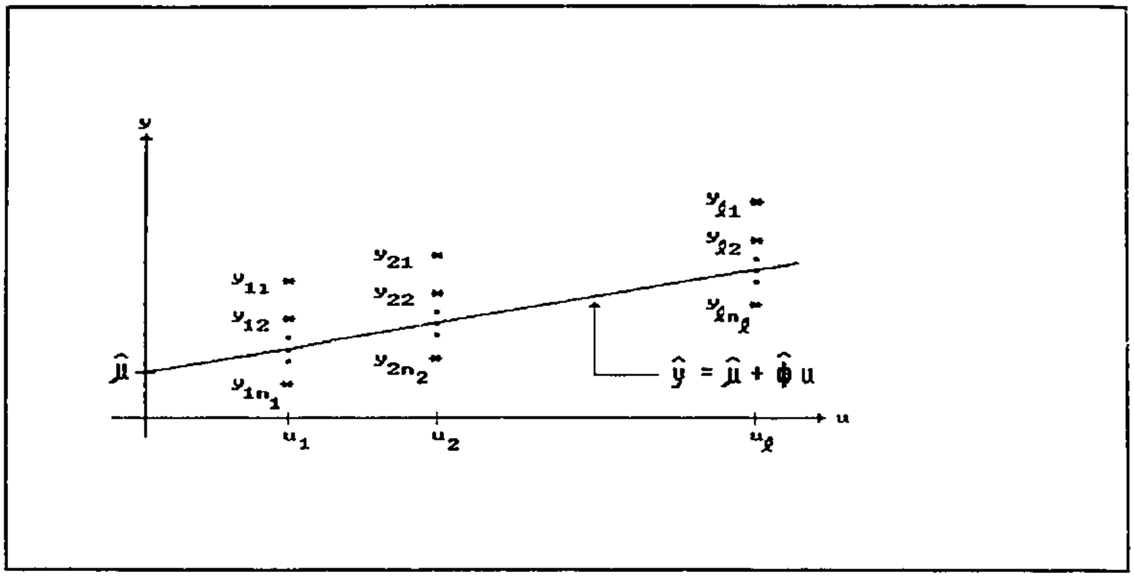

Figure 3.1: Example of a Best-Fit Regression Line

Also, we wish to estimate the unknown variance σ^2. Various methods are available to accomplish this, such as the method of least squares and the method of maximum likelihood. We will employ the maximum likelihood procedure, since it is recognized to be efficient if the underlying distribution is known and, in particular, is normal. (Of course, it is well known that under normality, the method of least squares produces estimators identical to the ML estimators.) With this in mind, let us form the likelihood function based on (4.2.1):

$$L \propto \frac{1}{\sigma^N} \prod_{i=1}^{\ell} \prod_{j=1}^{n_i} \exp\left\{\frac{-(y_{ij} - \mu - \phi u_i)^2}{2\sigma^2}\right\}$$

where $N = \Sigma n_i$. Now define $z_{ij} = (y_{ij} - \mu - \phi u_i)/\sigma$. Then we have:

$$L \propto \frac{1}{\sigma^N} \prod_{i=1}^{t} \prod_{j=1}^{n_i} \exp\left\{\frac{-z_{ij}^2}{2}\right\}$$

$$\text{or} \quad \ln L = \text{const.} - N\ln\sigma - \frac{1}{2}\sum_{i=1}^{t}\sum_{j=1}^{n_i} z_{ij}^2 \qquad (4.2.2)$$

We differentiate (4.2.2) with respect to μ, ϕ, and σ, knowing that $\partial z_{ij}/\partial\mu = -1/\sigma$, $\partial z_{ij}/\partial\phi = -u_i/\sigma$, and $\partial z_{ij}/\partial\sigma = -z_{ij}/\sigma$:

$$\frac{\partial \ln L}{\partial \mu} = \frac{1}{\sigma}\sum_{i=1}^{t}\sum_{j=1}^{n_i} z_{ij} = 0 \qquad (4.2.3)$$

$$\frac{\partial \ln L}{\partial \phi} = \frac{1}{\sigma}\sum_{i=1}^{t} u_i \sum_{j=1}^{n_i} z_{ij} = 0 \qquad (4.2.4)$$

$$\frac{\partial \ln L}{\partial \sigma} = -\frac{N}{\sigma} + \frac{1}{\sigma}\sum_{i=1}^{t}\sum_{j=1}^{n_i} z_{ij}^2 = 0 \qquad (4.2.5)$$

The solution of equations (4.2.3)-(4.2.5) produce the following ML estimators:

$$\hat{\phi} = \frac{\sum_{i=1}^{t} n_i (u_i - \bar{u})\bar{y}_i}{\sum_{i=1}^{t} n_i (u_i - \bar{u})^2} = \frac{S_{yu}}{S_{uu}} \qquad (4.2.6)$$

$$\hat{\mu} = \bar{\bar{y}} - \hat{\phi}\,\bar{u} \qquad (4.2.7)$$

$$\hat{\sigma}^2 = \frac{1}{N-2}\sum_{i=1}^{t}\sum_{j=1}^{n_i}\left\{y_{ij} - \bar{\bar{y}} - \hat{\phi}(u_i - \bar{u})\right\}^2 \qquad (4.2.8)$$

where

$$\bar{y}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} y_{ij}$$

$$\bar{\bar{y}} = \frac{1}{N}\sum_{i=1}^{t} n_i \bar{y}_i$$

$$\text{and} \quad \bar{u} = \frac{1}{N}\sum_{i=1}^{t} n_i u_i \,.$$

It may be noted here that the error SS $(N-2)\hat{\sigma}^2$ above has two orthogonal partitions, namely:

$$(N-2)\,\hat{\sigma}^2 = \sum_{i=1}^{\ell} n_i\left\{\bar{y}_i - \bar{\bar{y}} - \hat{\phi}\,(u_i - \bar{u})\right\}^2 + \sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\left\{y_{ij} - \bar{y}_i\right\}^2 \qquad (4.2.9)$$

$$= \qquad SS_{LOF} \qquad + \qquad SSE$$

The two SS on the right hand side of (4.2.9) are called the lack-of-fit SS and error SS, respectively. The mean error SS

$$\hat{\sigma}^2_e = \frac{1}{N-\ell}\sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\{y_{ij} - \bar{y}_i\}^2 \qquad (4.2.10)$$

provides an unbiased estimator of σ^2 always.

4.2.1 Asymptotic Variances and Covariances

Differentiating (4.2.2) twice with respect to μ, ϕ, and σ yields:

$$\frac{\partial^2 \ln L}{\partial\mu^2} = \frac{1}{\sigma}\sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\left(-\frac{1}{\sigma}\right) = -\frac{N}{\sigma^2}$$

$$\frac{\partial^2 \ln L}{\partial\mu\partial\phi} = \frac{1}{\sigma}\sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\left(-\frac{u_i}{\sigma}\right) = -\frac{1}{\sigma^2}\sum_{i=1}^{\ell} n_i u_i = -\frac{N\bar{u}}{\sigma^2}$$

$$\frac{\partial^2 \ln L}{\partial\mu\partial\sigma} = \frac{1}{\sigma}\sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\left(-\frac{z_{ij}}{\sigma}\right) = -\frac{1}{\sigma^2}\sum_{i=1}^{\ell}\sum_{j=1}^{n_i} z_{ij}$$

$$\frac{\partial^2 \ln L}{\partial\phi^2} = \frac{1}{\sigma}\sum_{i=1}^{\ell} u_i\sum_{j=1}^{n_i}\left(-\frac{u_i}{\sigma}\right) = -\frac{1}{\sigma^2}\sum_{i=1}^{\ell} n_i u_i^2$$

$$\frac{\partial^2 \ln L}{\partial\phi\partial\sigma} = -\frac{1}{\sigma}\sum_{i=1}^{\ell} u_i\sum_{j=1}^{n_i} z_{ij} + \frac{1}{\sigma}\sum_{i=1}^{\ell} u_i\sum_{j=1}^{n_i}\left(-\frac{z_{ij}}{\sigma}\right) = -\frac{2}{\sigma^2}\sum_{i=1}^{\ell} u_i\sum_{j=1}^{n_i} z_{ij}$$

$$\frac{\partial^2 \ln L}{\partial\sigma^2} = \frac{N}{\sigma^2} - \frac{1}{\sigma^2}\sum_{i=1}^{\ell}\sum_{j=1}^{n_i} z_{ij}^2 + \frac{1}{\sigma^2}\sum_{i=1}^{\ell}\sum_{j=1}^{n_i} 2\,z_{ij}\left(-\frac{z_{ij}}{\sigma}\right)$$

$$= \frac{N}{\sigma^2} - \frac{3}{\sigma^3}\sum_{i=1}^{\ell}\sum_{j=1}^{n_i} z_{ij}^2$$

Taking negative expectation gives :

$$-\mathrm{E}\left(\frac{\partial^2 \ln L}{\partial \mu^2}\right) = \frac{N}{\sigma^2}$$

$$-\mathrm{E}\left(\frac{\partial^2 \ln L}{\partial \mu \partial \phi}\right) = \frac{N\bar{u}}{\sigma^2}$$

$$-\mathrm{E}\left(\frac{\partial^2 \ln L}{\partial \mu \partial \sigma}\right) = -\frac{1}{\sigma^2}\sum_{i=1}^{t}\sum_{j=1}^{n_i}\mathrm{E}(z_{ij}) = 0 \qquad \text{since } \mathrm{E}(z_{ij}) = 0$$

$$-\mathrm{E}\left(\frac{\partial^2 \ln L}{\partial \phi^2}\right) = \frac{1}{\sigma^2}\sum_{i=1}^{t} n_i u_i^2$$

$$-\mathrm{E}\left(\frac{\partial^2 \ln L}{\partial \phi \partial \sigma}\right) = 0 \quad \text{since } \mathrm{E}(z_{ij}) = 0$$

$$-\mathrm{E}\left(\frac{\partial^2 \ln L}{\partial \sigma^2}\right) = -\frac{N}{\sigma^2} + \frac{3}{\sigma^3}\sum_{i=1}^{t}\sum_{j=1}^{n_i}\mathrm{E}(z_{ij}^2)$$

$$= \frac{2N}{\sigma^2} \quad \text{since } \mathrm{E}(z_{ij}^2) = \sigma^2 \ .$$

Thus the information matrix is:

$$\underset{\sim}{I}(\mu,\phi,\sigma) = \frac{1}{\sigma^2}\begin{bmatrix} N & N\bar{u} & 0 \\ N\bar{u} & \sum_{i=1}^{t} n_i u_i^2 & 0 \\ 0 & 0 & 2N \end{bmatrix}$$

and

$$\underset{\sim}{V}(\mu,\phi,\sigma) = \underset{\sim}{I}^{-1}(\mu,\phi,\sigma) = \sigma^2\begin{bmatrix} \frac{1}{N} + \frac{\bar{u}^2}{\Delta} & -\frac{\bar{u}^2}{\Delta} & 0 \\ -\frac{\bar{u}^2}{\Delta} & \frac{1}{\Delta} & 0 \\ 0 & 0 & \frac{1}{2N} \end{bmatrix}$$

$$\text{where} \qquad \Delta = N\sum_{i=1}^{t} n_i (u_i - \bar{u})^2$$

4.2.2 ANOVA for Linear Regression

One important aspect of statistical regression analysis is testing the suitability of the assumed model, and testing the null hypotheses H_0: $\phi = 0$. For this, we utilize the following ANOVA table, of course, under the assumption of normality:

ANOVA Table for Linear Regression

Source	SS	df	MS	F
Regression	$\hat{\phi}S_{yu}$	1	$\hat{\phi}S_{yu}$	MS_R/MSE
Lack of Fit	SS_{LOF}	ℓ-2	$SS_{LOF}/(\ell-2)$	MS_{LOF}/MSE
Error	SSE	N-ℓ	SSE/(N-ℓ)	
Total	SST	N-1	SST/(N-1)	

Under H_0: $\phi = 0$, the F-ratio MS_R/MSE is distributed as central-F with $(1, N-\ell)$ degrees of freedom. A significant value for this ratio indicates that the slope parameter ϕ is non-zero. The ratio MS_{LOF}/MSE is also distributed as central-F with $(\ell-2, N-\ell)$ degrees of freedom. A significant value for this F-statistic indicates that a linear fit is not adequate for the data under investigation.

As is the case with most classical statistical procedures, the regression analysis outlined above performs very well when the data are normally distributed. However, as mentioned before, when the assumption is not justified, the classical regression procedure can be highly inefficient and exhibit low power. Unfortunately, in the "real world" non-normality is usually the rule rather than the exception. Thus it is of interest to develop a regression procedure (similar to the classical procedure) that performs well even when the normality assumption is no longer true. As in Chapter 2, to accomplish this we will employ the Tiku-Suresh MML method discussed in the first two chapters to develop simple linear and multiple regression models for a family of non-normal error distributions. It will be shown that the modified method is very similar in form as compared to the classical method, but is much more efficient and powerful under non-normality, represented by the family (2.3.1).

4.3 Non-Normal Linear Regression Procedure for Complete Samples

Once again, let us consider the linear regression model:

$$y_{ij} = \mu + \phi u_i + e_{ij} \qquad (i = 1,\dots,\ell;\ \ j = 1,\dots,n_i) \qquad (4.3.1)$$

To introduce non-normality, let us assume that the ε_{ij}'s have the t-distribution (2.3.1). We may now form the likelihood function based on (4.3.1) and the t-family assumption, after ordering each of the i samples with respect to j:

$$L \propto \frac{1}{\sigma^N}\prod_{i=1}^{\ell}\prod_{j=1}^{n_i}\left\{1 + \frac{[y_{(ij)} - \mu - \phi u_i]^2}{k\sigma^2}\right\}^{-p}, \quad \text{or}$$

$$\ln L = \text{const.} - N\ln\sigma - p\sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\ln\left\{1 + \frac{[y_{(ij)} - \mu - \phi u_i]^2}{k\sigma^2}\right\} \quad (4.3.2)$$

where $N = \sum n_i$ and parentheses around subscripts denotes ordering with respect to j. Writing $z_{(ij)} = [y_{(ij)} - \mu - \phi u_i]/\sigma$, (4.3.2) becomes:

$$\ln L = \text{const.} - N\ln\sigma - p\sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\ln\left\{1 + \frac{z_{(ij)}^2}{k}\right\} \qquad (4.3.3)$$

We now differentiate (4.3.3) with respect to μ, ϕ, and σ, noting that $\partial z_{(ij)}/\partial\mu = -1/\sigma$, $\partial z_{(ij)}/\partial\phi = -u_i/\sigma$, and $\partial z_{(ij)}/\partial\sigma = -z_{(ij)}/\sigma$:

$$\frac{\partial \ln L}{\partial \mu} = \frac{2p}{k\sigma}\sum_{i=1}^{\ell}\sum_{j=1}^{n_i} g(z_{(ij)}) = 0 \tag{4.3.4}$$

$$\frac{\partial \ln L}{\partial \phi} = \frac{2p}{k\sigma}\sum_{i=1}^{\ell} u_i \sum_{j=1}^{n_i} g(z_{(ij)}) = 0 \tag{4.3.5}$$

$$\frac{\partial \ln L}{\partial \sigma} = -\frac{N}{\sigma} + \frac{2p}{k\sigma}\sum_{i=1}^{\ell}\sum_{j=1}^{n_i} z_{(ij)}\, g(z_{(ij)}) = 0 \tag{4.3.6}$$

where $g(z) = z/[1 + z^2/k]$. The simultaneous solution of equations (4.3.4)-(4.3.6) would yield the maximum likelihood estimators of μ, ϕ, and σ. Unfortunately, as discussed in Chapters 1 and 2, these equations do not admit closed-form solutions (due to the presence of the function g), and must be solved iteratively. Procedures exist to accomplish this; however they may experience difficulty in converging or may converge to incorrect values. Thus we will approximate the function g by expanding it in a Taylor's Series around the point $t_{(ij)} = E(z_{(ij)})$, obtaining:

$$g(z_{(ij)}) \approx \gamma_{ij} + \delta_{ij} z_{(ij)} \tag{4.3.7}$$

$$\text{where } \gamma_{ij} = \frac{2\, t_{(ij)}^3/k}{[1 + t_{(ij)}^2/k]^2} \quad \text{and} \quad \delta_{ij} = \frac{1 - t_{(ij)}^2/k}{[1 + t_{(ij)}^2/k]^2}$$

Substituting (4.3.7) into (4.3.4)-(4.3.6) produces the following modified likelihood derivatives:

$$\frac{\partial \ln L^*}{\partial \mu} = \frac{2p}{k\sigma}\sum_{i=1}^{\ell}\sum_{j=1}^{n_i} [\gamma_{ij} + \delta_{ij} z_{(ij)}] = 0 \tag{4.3.8}$$

$$\frac{\partial \ln L^*}{\partial \phi} = \frac{2p}{k\sigma}\sum_{i=1}^{\ell} u_i \sum_{j=1}^{n_i} [\gamma_{ij} + \delta_{ij} z_{(ij)}] = 0 \tag{4.3.9}$$

$$\frac{\partial \ln L^*}{\partial \sigma} = -\frac{N}{\sigma} + \frac{2p}{k\sigma}\sum_{i=1}^{\ell}\sum_{j=1}^{n_i} z_{(ij)} [\gamma_{ij} + \delta_{ij} z_{(ij)}] = 0 \tag{4.3.10}$$

Equations (4.3.8)-(4.3.10) now admit explicit estimators for the parameters, namely:

$$\hat{\phi} = \frac{S_{yu}}{S_{uu}} \tag{4.3.11}$$

$$\hat{\mu} = \bar{\bar{y}} - \hat{\phi}\,\bar{u} \tag{4.3.12}$$

and $$\hat{\sigma} = \frac{B + \sqrt{B^2 + 4NC}}{2\sqrt{N(N-2)}}, \quad \text{corrected for bias} \tag{4.3.13}$$

where

$$m_i = \sum_{j=1}^{n_i} \delta_{ij}\ ; \qquad M = \sum_{i=1}^{\ell} m_i \tag{4.3.14}$$

$$\bar{y}_i = \frac{1}{m_i}\sum_{j=1}^{n_i} \delta_{ij}\, y_{(ij)}\ ; \qquad \bar{\bar{y}} = \frac{1}{M}\sum_{i=1}^{\ell} m_i \bar{y}_i \tag{4.3.15}$$

$$\bar{u} = \frac{1}{M}\sum_{i=1}^{\ell} m_i u_i \tag{4.3.16}$$

$$S_{yu} = \sum_{i=1}^{\ell} m_i (u_i - \bar{u})\bar{y}_i \tag{4.3.17}$$

$$S_{uu} = \sum_{i=1}^{\ell} m_i (u_i - \bar{u})^2 \tag{4.3.18}$$

$$B = \frac{2p}{k}\sum_{i=1}^{\ell}\sum_{j=1}^{n_i} \gamma_{ij}\,[\,y_{(ij)} - \bar{\bar{y}} - \hat{\phi}\,(u_i - \bar{u})] \tag{4.3.19}$$

$$C = \frac{2p}{k}\sum_{i=1}^{\ell}\sum_{j=1}^{n_i} \delta_{ij}\,[\,y_{(ij)} - \bar{\bar{y}} - \hat{\phi}\,(u_i - \bar{u})]^2 \tag{4.3.20}$$

Notice that for $p = \infty$, the t-family (2.3.1) reduces to normal. In that case, $\gamma_{ij} \to 0$, $\delta_{ij} \to 1$, and $2p/k \to 1$. Consequently, (4.3.11)-(4.3.13) reduce to the expressions based on normality, as was expected.

4.3.1 Asymptotic Variances and Covariances

As in Section 4.2.1, to obtain the asymptotic variances and covariances for $\hat{\mu}$, $\hat{\phi}$, and $\hat{\sigma}$, we differentiate **(4.3.8)-(4.3.10)** again to obtain:

$$\frac{\partial^2 \ln L^*}{\partial \mu^2} = -\frac{2p}{k\sigma^2} \sum_{i=1}^{\ell} \sum_{j=1}^{n_i} \delta_{ij} = -\frac{2p}{k\sigma^2} M$$

$$\frac{\partial^2 \ln L^*}{\partial \mu \partial \phi} = -\frac{2p}{k\sigma^2} \sum_{i=1}^{\ell} m_i u_i = -\frac{2p}{k\sigma^2} M\bar{u}$$

$$\frac{\partial^2 \ln L^*}{\partial \mu \partial \sigma} = -\frac{2p}{k\sigma^2} \sum_{i=1}^{\ell} \sum_{j=1}^{n_i} [\gamma_{ij} + 2\delta_{ij} z_{(ij)}]$$

$$\frac{\partial^2 \ln L^*}{\partial \phi^2} = -\frac{2p}{k\sigma^2} \sum_{i=1}^{\ell} m_i u_i^2$$

$$\frac{\partial^2 \ln L^*}{\partial \phi \partial \sigma} = -\frac{2p}{k\sigma^2} \sum_{i=1}^{\ell} u_i \sum_{j=1}^{n_i} [\gamma_{ij} + 2\delta_{ij} z_{(ij)}]$$

$$\frac{\partial^2 \ln L^*}{\partial \sigma^2} = \frac{N}{\sigma^2} - \frac{2p}{k\sigma^2} \sum_{i=1}^{\ell} \sum_{j=1}^{n_i} z_{(ij)} [2\gamma_{ij} + 3\delta_{ij} z_{(ij)}]$$

Now taking negative expectation, we have, where $t_{(ij)} = E(z_{(ij)})$:

$$-E\left(\frac{\partial^2 \ln L^*}{\partial \mu^2}\right) = \frac{2p}{k\sigma^2} M$$

$$-E\left(\frac{\partial^2 \ln L^*}{\partial \mu \partial \phi}\right) = \frac{2p}{k\sigma^2} M\bar{u}$$

$$-E\left(\frac{\partial^2 \ln L^*}{\partial \mu \partial \sigma}\right) = \frac{2p}{k\sigma^2} \sum_{i=1}^{\ell} \sum_{j=1}^{n_i} [\gamma_{ij} + 2\delta_{ij} t_{(ij)}] = 0$$

since $\sum_{j=1}^{n_i} \gamma_{ij} = 0$ and $\sum_{j=1}^{n_i} \delta_{ij} t_{(ij)} = 0$

$$-E\left(\frac{\partial^2 \ln L^*}{\partial \phi^2}\right) = \frac{2p}{k\sigma^2} \sum_{i=1}^{\ell} m_i u_i^2$$

$$-E\left(\frac{\partial^2 \ln L^*}{\partial\phi\partial\sigma}\right) = 0 \qquad \text{(by symmetry)}$$

$$-E\left(\frac{\partial^2 \ln L^*}{\partial\sigma^2}\right) = -\frac{N}{\sigma^2} + \frac{2p}{k\sigma^2}\sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\left[2\gamma_{ij}t_{(ij)} + 3\delta_{ij}\left(\sigma_{jj:n_i} + t_{(ij)}^2\right)\right]$$

where $\sigma_{jj:n_i} = V(z_{(ij)})$

Thus the information matrix is:

$$\underset{\sim}{I}(\mu,\phi,\sigma) = \frac{2p}{k\sigma^2}\begin{bmatrix} M & M\bar{u} & 0 \\ M\bar{u} & \sum_{i=1}^{\ell} m_i u_i^2 & 0 \\ 0 & 0 & \Omega \end{bmatrix}$$

where $\frac{2p}{k\sigma^2}\Omega = -E\left(\frac{\partial^2 \ln L^*}{\partial\sigma^2}\right)$

and hence the asymptotic variance-covariance matrix is:

$$\underset{\sim}{V}(\mu,\phi,\sigma) = \underset{\sim}{I}^{-1}(\mu,\phi,\sigma) = \frac{k\sigma^2}{2p}\begin{bmatrix} \frac{1}{M} + \frac{M\bar{u}^2}{\Delta} & -\frac{M\bar{u}}{\Delta} & 0 \\ -\frac{M\bar{u}}{\Delta} & \frac{M}{\Delta} & 0 \\ 0 & 0 & \frac{1}{\Omega} \end{bmatrix}$$

where $\Delta = M\sum_{i=1}^{\ell} m_i(u_i - \bar{u})^2$

4.3.2 ANOVA for Linear Regression

As in Section 4.2.2, we may partition the error sum-of-squares $(N-2)\hat{\sigma}^2$ in order to test the validity of the model and the hypothesis H_0: $\phi = 0$. First note that from **(4.3.13)**:

$$\hat{\sigma} = \frac{B + \sqrt{B^2 + 4NC}}{2\sqrt{N(N-2)}} \approx \frac{\sqrt{C}}{\sqrt{N-2}}$$

since terms involving B^2/NC are negligible for large N. Consequently,

$$(N-2)\hat{\sigma}^2 \approx C$$

$$= \frac{2p}{k} \sum_{i=1}^{t} \sum_{j=1}^{n_i} \delta_{ij} [y_{(ij)} - \bar{\bar{y}} - \hat{\phi}(u_i - \bar{u})]^2 \qquad \textbf{(4.3.2.1)}$$

Ignoring $2p/k$ for the moment, consider **(4.3.2.1)**:

$$\sum_{i=1}^{t} \sum_{j=1}^{n_i} \delta_{ij} [y_{(ij)} - \bar{\bar{y}} - \hat{\phi}(u_i - \bar{u})]^2$$

$$= \sum_{i=1}^{t} \sum_{j=1}^{n_i} \delta_{ij} \left\{ [y_{(ij)} - \bar{\bar{y}}]^2 - 2\,[y_{(ij)} - \bar{\bar{y}}]\,\hat{\phi}\,(u_i - \bar{u}) + \hat{\phi}^2 (u_i - \bar{u})^2 \right\}$$

$$= \sum_{i=1}^{t} \sum_{j=1}^{n_i} \delta_{ij} [y_{(ij)} - \bar{\bar{y}}]^2 - 2\hat{\phi} \sum_{i=1}^{t} m_i (u_i - \bar{u})\,\bar{y}_i + \hat{\phi}^2 S_{uu}$$

$$= SST - 2\hat{\phi} S_{yu} + \hat{\phi}^2 S_{uu} = SST - \hat{\phi} S_{yu}$$

Hence $\quad SST = \hat{\phi} S_{yu} + \frac{k}{2p}(N-2)\,\hat{\sigma}^2$

Also, notice that we may write:

$$\frac{k}{2p}(N-2)\,\hat{\sigma}^2 = \sum_{i=1}^{t} \sum_{j=1}^{n_i} \delta_{ij} [y_{(ij)} - \bar{y}_i + \bar{y}_i - \bar{\bar{y}} - \hat{\phi}\,(u_i - \bar{u})]^2$$

$$= \sum_{i=1}^{t} \sum_{j=1}^{n_i} \delta_{ij} [y_{(ij)} - \bar{y}_i]^2 + \sum_{i=1}^{t} m_i\,[\bar{y}_i - \bar{\bar{y}} - \hat{\phi}\,(u_i - \bar{u})]^2$$

$$= \qquad SSE \qquad + \qquad SS_{LOF}$$

Therefore, we have the following break-down of the total SS:

$$SST = \hat{\phi} S_{yu} + SS_{LOF} + SSE \qquad (4.3.2.2)$$

where the partitions correspond to the regression, lack-of-fit, and pure error terms respectively. (Of course each term in (4.3.2.2) should be multiplied by $2p/k$, but since the goal is to form F-ratio statistics, these constants would divide out, so they do not need to be included.) Once again, we may form the ANOVA Table for Linear Regression as in Section 4.2.2. This can be done since the sums-of-squares $\hat{\phi}S_{yu}$, SS_{LOF}, and SSE will have (when divided by σ^2) χ^2-distributions with 1, ℓ-2, and N-ℓ degrees of freedom respectively, using the result from Appendix A1 that $(N-2)\hat{\sigma}^2/\sigma^2$ is (approximately) χ^2 with N-2 degrees of freedom.

Hence the F-ratios MS_R/MSE and MS_{LOF}/MSE will have F-distributions with (1, N-ℓ) and (ℓ-2, N-ℓ) degrees of freedom as before. The testing of the hypotheses then continues in the usual manner. Later on, we will present some simulations of the efficiencies of the above-mentioned estimators, and will also demonstrate the efficacy of the MML Regression procedure.

4.4 Linear Regression Based on Type-II Symmetrically Censored Samples

As we have discussed in Chapter 2, it is often the case when performing an experiment that the collection of data must be terminated before all units have been subjected to the treatment(s). In the regression context, this means that we may have Type-II censored samples at each level of the independent variable. Let us then again consider the simple linear regression model:

$$y_{ij} = \mu + \phi u_i + \varepsilon_{ij} \qquad (i = 1,\ldots,\ell;\ j = 1,\ldots,n_i)$$

where now at each level i of the independent variable u we order the observations (with respect to j) and censor the r_i smallest and largest. The data matrix in this case is:

	Level of u						
	u_1	u_2	$\cdots$	u_i	$\cdots$	u_ℓ
	$y_{(1,a_1)}$	$y_{(2,a_2)}$	$\cdots$	$y_{(i,a_i)}$	$\cdots$	$y_{(\ell,a_\ell)}$
	$\wedge\!\!\!/$	$\wedge\!\!\!/$		$\wedge\!\!\!/$		$\wedge\!\!\!/$
Response y:	$y_{(1,a_1+1)}$	$y_{(2,a_2+1)}$	$\cdots$	$y_{(i,a_i+1)}$	$\cdots$	$y_{(\ell,a_\ell+1)}$
	$\wedge\!\!\!/$	$\wedge\!\!\!/$		$\wedge\!\!\!/$		$\wedge\!\!\!/$
	$\vdots$	$\vdots$		$\vdots$		$\vdots$
	$\wedge\!\!\!/$	$\wedge\!\!\!/$		$\wedge\!\!\!/$		$\wedge\!\!\!/$
	$y_{(1,b_1)}$	$y_{(2,b_2)}$	$\cdots$	$y_{(i,b_i)}$	$\cdots$	$y_{(\ell,b_\ell)}$

where $a_i = r_i+1$ and $b_i = n_i-r_i$. The likelihood function based on the ℓ censored samples is then:

$$L \propto \frac{1}{\sigma^A}\prod_{i=1}^{\ell}\prod_{j=a_i}^{b_i}\left\{1 + \frac{[y_{(ij)} - \mu - \phi u_i]^2}{k\sigma^2}\right\}^{-p}$$
$$\times\left[F\left(\frac{y_{(i,a_i)} - \mu - \phi u_i}{\sigma}\right)\right]^{r_i}\left[1 - F\left(\frac{y_{(i,b_i)} - \mu - \phi u_i}{\sigma}\right)\right]^{r_i} \qquad (4.4.1)$$

where $A = \sum(n_i-2r_i)$. Writing $z_{(ij)} = [y_{(ij)} - \mu - \phi u_i]/\sigma$ and taking log, (4.4.1) becomes:

$$\ln L = \text{const.} - A\ln\sigma - p\sum_{i=1}^{\ell}\sum_{j=a_i}^{b_i}\left\{1 + \frac{z_{(ij)}^2}{k}\right\}$$
$$+ \sum_{i=1}^{\ell} r_i\{\ln F(z_{(i,a_i)}) + \ln[1 - F(z_{(i,b_i)})]\} \qquad (4.4.2)$$

Differentiating (4.4.2), we have:

$$\frac{\partial \ln L}{\partial \mu} = \frac{2p}{k\sigma} \sum_{i=1}^{l} \sum_{j=a_i}^{b_i} g(z_{(ij)}) + \frac{1}{\sigma} \sum_{i=1}^{l} r_i \, [\, h_2(z_{(i,b_i)}) - h_1(z_{(i,a_i)})] = 0 \qquad (4.4.3)$$

$$\frac{\partial \ln L}{\partial \phi} = \frac{2p}{k\sigma} \sum_{i=1}^{l} u_i \sum_{j=a_i}^{b_i} g(z_{(ij)}) + \frac{1}{\sigma} \sum_{i=1}^{l} u_i r_i \, [\, h_2(z_{(i,b_i)}) - h_1(z_{(i,a_i)})] = 0 \qquad (4.4.4)$$

$$\frac{\partial \ln L}{\partial \sigma} = -\frac{A}{\sigma} + \frac{2p}{k\sigma} \sum_{i=1}^{l} \sum_{j=a_i}^{b_i} z_{(ij)}\, g(z_{(ij)})$$

$$+ \frac{1}{\sigma} \sum_{i=1}^{l} r_i \, [\, z_{(i,b_i)} h_2(z_{(i,b_i)}) - z_{(i,a_i)} h_1(z_{(i,a_i)})] = 0 \qquad (4.4.5)$$

where the functions g, h_1, and h_2 have been defined previously. As mentioned before, equations (4.4.3)-(4.4.5) do not yield explicit ML estimators due to the presence of g, h_1, and h_2. In order to avoid costly and possibly inaccurate numerical iterations, we will employ the Tiku-Suresh approximations:

$$g(z_{(ij)}) \approx \gamma_{ij} + \delta_{ij} z_{(ij)}$$
$$h_1(z_{(ij)}) \approx \alpha_i - \beta_i z_{(ij)}$$
$$h_2(z_{(ij)}) \approx \alpha_i + \beta_i z_{(ij)}$$

where γ_{ij}, δ_{ij}, α_i, and β_i are all defined in Section 2.6. Substituting these approximations into (4.4.3)-(4.4.5) and simplifying gives the following modified likelihood equations:

$$\frac{\partial \ln L^*}{\partial \mu} = \frac{1}{\sigma}\sum_{i=1}^{t}\sum_{j=a_i}^{b_i}[\gamma_{ij} + \delta_{ij} z_{(ij)}] + \frac{1}{\sigma}\sum_{i=1}^{t} r_i \beta^*_i [z_{(i,a_i)} + z_{(i,b_i)}] = 0 \qquad (4.4.6)$$

$$\frac{\partial \ln L^*}{\partial \phi} = \frac{1}{\sigma}\sum_{i=1}^{t} u_i \sum_{j=a_i}^{b_i}[\gamma_{ij} + \delta_{ij} z_{(ij)}] + \frac{1}{\sigma}\sum_{i=1}^{t} r_i \beta^*_i u_i [z_{(i,a_i)} + z_{(i,b_i)}] = 0 \qquad (4.4.7)$$

$$\frac{\partial \ln L^*}{\partial \sigma} = -\frac{A^*}{\sigma} + \frac{1}{\sigma}\sum_{i=1}^{t}\sum_{j=a_i}^{b_i} z_{(ij)}[\gamma_{ij} + \delta_{ij} z_{(ij)}]$$

$$+ \frac{1}{\sigma}\sum_{i=1}^{t} r_i \left\{ \alpha^*_i [z_{(i,b_i)} - z_{(i,a_i)}] + \beta^*_i [z^2_{(i,a_i)} + z^2_{(i,b_i)}] \right\} = 0 \qquad (4.4.8)$$

where we have divided out $2p/k$ in each expression and α_i^*, β_i^*, and A^* are as defined previously. We now introduce the following expressions for censored samples similar to those in Section 4.2:

$$m_i^* = m_i + 2 r_i \beta^*_i = \sum_{j=a_i}^{b_i} \delta_{ij} + 2 r_i \beta^*_i \ ; \quad M^* = \sum_{i=1}^{t} m_i^* \qquad (4.4.9)$$

$$\overline{y}_i = \frac{1}{m_i}\sum_{j=a_i}^{b_i} \delta_{ij} y_{(ij)} \qquad (4.4.10)$$

$$\overline{y}_i^* = \frac{1}{m_i^*}\left[m_i \overline{y}_i + r_i \beta^*_i (y_{(i,a_i)} + y_{(i,b_i)}) \right] \qquad (4.4.11)$$

$$\overline{\overline{y}}^* = \frac{1}{M^*}\sum_{i=1}^{t} m_i^* \overline{y}_i^* \qquad (4.4.12)$$

$$\overline{u}^* = \frac{1}{M^*}\sum_{i=1}^{t} m_i^* u_i \qquad (4.4.13)$$

$$S^*_{yu} = \sum_{i=1}^{t} m_i^* (u_i - \overline{u}^*) \overline{y}_i^* \qquad (4.4.14)$$

$$S^*_{uu} = \sum_{i=1}^{t} m_i^* (u_i - \overline{u}^*)^2 \qquad (4.4.15)$$

We may now proceed to solve equations **(4.4.6)-(4.4.8)** using **(4.4.9)-(4.4.15)** and the fact that $\sum_{j=a_i}^{b_i} \delta_{ij} = 0$ (since censoring is symmetric) to obtain explicit estimators as:

$$\hat{\phi} = \frac{S^{*}_{yu}}{S^{*}_{uu}} \tag{4.4.16}$$

$$\hat{\mu} = \bar{\bar{y}}^{*} - \hat{\phi}\,\bar{u}^{*} \tag{4.4.17}$$

$$\hat{\sigma} = \frac{B^{*} + \sqrt{B^{*2} + 4A^{*}C^{*}}}{2\sqrt{A^{*}\left(A^{*} - \left(\frac{k}{2p}\right)2\right)}}, \quad \text{(corrected for bias)} \tag{4.4.18}$$

where $B^{*} = \sum_{i=1}^{\ell}\sum_{j=a_i}^{b_i} \gamma_{ij}\{y_{(ij)} - \bar{\bar{y}}^{*} - \hat{\phi}(u_i - \bar{u}^{*})\} + \sum_{i=1}^{\ell} r_i\alpha^{*}_i\{y_{(i,b_i)} - y_{(i,a_i)}\}$

and $C^{*} = \sum_{i=1}^{\ell}\sum_{j=a_i}^{b_i} \delta_{ij}[y_{(ij)} - \bar{\bar{y}}^{*} - \hat{\phi}(u_i - \bar{u}^{*})]^2$

$$+ \sum_{i=1}^{\ell} r_i\beta^{*}_i\left\{[y_{(i,a_i)} - \bar{\bar{y}}^{*} - \hat{\phi}(u_i - \bar{u}^{*})]^2 + [y_{(i,b_i)} - \bar{\bar{y}}^{*} - \hat{\phi}(u_i - \bar{u}^{*})]^2\right\}$$

Equations (4.4.16)-(4.4.18) are the MML estimators for the simple linear regression model (4.2.1) based on symmetrically censored samples. Notice that their forms are identical to those obtained in the complete-sample case (and even the classical normally distributed case); the only difference is that each expression has been replaced with the corresponding expression based on censoring. Of course when there is no censoring ($r_i = 0$, $i = 1,\dots,\ell$), (4.4.16)-(4.4.18) reduce to the complete sample results given in Section 4.3.

4.4.1 Asymptotic Variances and Covariances for Symmetrically Censored Samples

Differentiating (4.4.6)-(4.4.8) with respect to μ, σ and ϕ gives:

$$\frac{\partial^2 \ln L^*}{\partial \mu^2} = -\frac{2p}{k\sigma^2}\sum_{i=1}^{t}[m_i + 2r_i\beta^*_i] = -\frac{2p}{k\sigma^2}M^*$$

$$\frac{\partial^2 \ln L^*}{\partial \mu \partial \phi} = -\frac{2p}{k\sigma^2}\sum_{i=1}^{t}[m_i + 2r_i\beta^*_i]\,u_i = -\frac{2p}{k\sigma^2}M^*\,\overline{u}^*$$

$$\frac{\partial^2 \ln L^*}{\partial \mu \partial \sigma} = -\frac{2p}{k\sigma^2}\left\{\sum_{i=1}^{t}\sum_{j=a_i}^{b_i}[\gamma_{ij} + 2\delta_{ij}z_{(ij)}] + 2\sum_{i=1}^{t}r_i\beta^*_i\,[z_{(i,a_i)} + z_{(i,b_i)}]\right\}$$

$$\frac{\partial^2 \ln L^*}{\partial \phi^2} = -\frac{2p}{k\sigma^2}\sum_{i=1}^{t}[m_i + 2r_i\beta^*_i]\,u_i^2 = -\frac{2p}{k\sigma^2}\sum_{i=1}^{t}m^*_i u_i^2$$

$$\frac{\partial^2 \ln L^*}{\partial \phi \partial \sigma} = -\frac{2p}{k\sigma^2}\left\{\sum_{i=1}^{t}u_i\sum_{j=a_i}^{b_i}[\gamma_{ij} + 2\delta_{ij}z_{(ij)}] + 2\sum_{i=1}^{t}r_i u_i\beta^*_i\,[z_{(i,a_i)} + z_{(i,b_i)}]\right\}$$

$$\frac{\partial^2 \ln L^*}{\partial \sigma^2} = -\frac{2p}{k\sigma^2}\left\{-A^* + \sum_{i=1}^{t}\sum_{j=a_i}^{b_i}z_{(ij)}\,[2\gamma_{ij} + 3\delta_{ij}z_{(ij)}]\right.$$

$$\left. + \sum_{i=1}^{t}r_i\,[2\alpha^*_i(z_{(i,b_i)} - z_{(i,a_i)}) + 3\beta^*_i(z^2_{(i,a_i)} + z^2_{(i,b_i)})]\right\}$$

Now taking negative expectation and noting that, under symmetric censoring (where $t_{(ij)} = \mathrm{E}(z_{(ij)})$),

$$\sum_{j=a_i}^{b_i}\gamma_{ij} = 0, \quad \sum_{j=a_i}^{b_i}\delta_{ij}t_{(ij)} = 0, \quad t_{(i,a_i)} = -t_{(i,b_i)},$$

we have:

$$-\mathrm{E}\left(\frac{\partial^2 \ln L^*}{\partial \mu^2}\right) = \frac{2p}{k\sigma^2}M^*$$

$$-\mathrm{E}\left(\frac{\partial^2 \ln L^*}{\partial \mu \partial \phi}\right) = \frac{2p}{k\sigma^2}M^*\,\overline{u}^*$$

$$-\mathrm{E}\left(\frac{\partial^2 \ln L^*}{\partial \mu \partial \sigma}\right) = 0$$

$$-\mathrm{E}\left(\frac{\partial^2 \ln L^*}{\partial \phi^2}\right) = \frac{2p}{k\sigma^2}\sum_{i=1}^{t}m^*_i u_i^2$$

$$-\mathrm{E}\left(\frac{\partial^2 \ln L^*}{\partial\phi\partial\sigma}\right) = 0$$

$$-\mathrm{E}\left(\frac{\partial^2 \ln L^*}{\partial\sigma^2}\right) = \frac{2p}{k\sigma^2}\left\{-A^* + \sum_{i=1}^{\ell}\sum_{j=a_i}^{b_i}\left[2\gamma_{ij}\, t_{(ij)} + 3\,\delta_{ij}\left(\sigma_{jj:n_i} + t_{(ij)}^2\right)\right]\right.$$

$$\left. + \sum_{i=1}^{\ell} r_i\left[4\,\alpha_i^*\, t_{(i,b_i)} + 6\,\beta_i^*\left(\sigma_{b_i,b_i:n_i} + t_{(i,b_i)}^2\right)\right]\right\}$$

Hence the information matrix based on symmetrically censored samples can be written as:

$$\underset{\sim}{I}(\mu,\phi,\sigma) = \frac{2p}{k\sigma^2}\begin{bmatrix} M^* & M^*\,\bar{u}^* & 0 \\ M^*\,\bar{u}^* & \sum_{i=1}^{\ell} m_i^*\, u_i^2 & 0 \\ 0 & 0 & \Omega^* \end{bmatrix}$$

$$\text{where} \quad \frac{2p}{k\sigma^2}\,\Omega^* = -\mathrm{E}\left(\frac{\partial^2 \ln L^*}{\partial\sigma^2}\right)$$

and the asymptotic variance-covariance matrix is therefore:

$$\underset{\sim}{V}(\mu,\phi,\sigma) = \frac{k\sigma^2}{2p}\begin{bmatrix} \dfrac{\sum_{i=1}^{\ell} m_i^*\, u_i^2}{\Delta^*} & -\dfrac{M^*\,\bar{u}^*}{\Delta^*} & 0 \\ -\dfrac{M^*\,\bar{u}^*}{\Delta^*} & \dfrac{M^*}{\Delta^*} & 0 \\ 0 & 0 & \dfrac{1}{\Omega^*} \end{bmatrix}$$

$$\text{where} \quad \Delta^* = M^* \sum_{i=1}^{\ell} m_i^*\left[u_i - \bar{u}^*\right]^2$$

It is very pleasing to notice that this matrix is identical in form to the one given in Section 4.3.1, with each expression being replaced with the corresponding expression for symmetrically censored samples, as is expected.

4.4.2 ANOVA for Linear Regression for Symmetrically Censored Samples

As in the complete sample case, note that we may write:

$$\hat{\sigma} = \frac{B^* + \sqrt{B^{*2} + 4A^*C^*}}{2\sqrt{A^*(A^* - \left(\frac{k}{2p}\right)2)}} \approx \frac{\sqrt{C^*}}{\sqrt{A^* - \left(\frac{k}{2p}\right)2}} = \frac{\sqrt{C}}{\sqrt{A-2}}$$

since $\frac{B^{*2}}{A^*C^*}$ will be negligible for large A.

Hence the error sum-of-squares $[A^*-(k/2p)2]\hat{\sigma}^2$ can be written as:

$$(A^* - \left(\tfrac{k}{2p}\right)2)\hat{\sigma}^2 \approx C^*$$

$$= \sum_{i=1}^{t}\sum_{j=a_i}^{b_i} \delta_{ij}\,[y_{(ij)} - \bar{\bar{y}}^* - \hat{\phi}(u_i - \bar{u}^*)]^2$$

$$+ \sum_{i=1}^{t} r_i\,\beta_i^* \left\{ [y_{(i,a_i)} - \bar{\bar{y}}^* - \hat{\phi}(u_i - \bar{u}^*)]^2 + [y_{(i,b_i)} - \bar{\bar{y}}^* - \hat{\phi}(u_i - \bar{u}^*)]^2 \right\}$$

$$= \sum_{i=1}^{t}\sum_{j=a_i}^{b_i} \delta_{ij} \left\{ [y_{(ij)} - \bar{\bar{y}}^*]^2 - 2\hat{\phi}\,[y_{(ij)} - \bar{\bar{y}}^*](u_i - \bar{u}^*) + \hat{\phi}^2 (u_i - \bar{u}^*)^2 \right\}$$

$$+ \sum_{i=1}^{t} r_i\,\beta_i^* \left\{ [y_{(i,a_i)} - \bar{\bar{y}}^*]^2 - 2\hat{\phi}\,[y_{(i,a_i)} - \bar{\bar{y}}^*](u_i - \bar{u}^*) + \hat{\phi}^2 (u_i - \bar{u}^*)^2 \right.$$

$$\left. + [y_{(i,b_i)} - \bar{\bar{y}}^*]^2 - 2\hat{\phi}\,[y_{(i,b_i)} - \bar{\bar{y}}^*](u_i - \bar{u}^*) + \hat{\phi}^2 (u_i - \bar{u}^*)^2 \right\}$$

(cont...)

$$= \sum_{i=1}^{t} \left\{ \sum_{j=a_i}^{b_i} \delta_{ij} [y_{(ij)} - \bar{\bar{y}}^*]^2 + r_i \beta_i^* \left\{ [y_{(i,a_i)} - \bar{\bar{y}}^*]^2 + [y_{(i,b_i)} - \bar{\bar{y}}^*]^2 \right\} \right\}$$

$$- 2 \hat{\phi} \sum_{i=1}^{t} m_i^* (u_i - \bar{u}^*) \bar{y}_i^* + \hat{\phi}^2 \sum_{i=1}^{t} m_i^* (u_i - \bar{u}^*)^2$$

$$= \text{SST} - 2 \hat{\phi} S_{yu}^* + \hat{\phi}^2 S_{uu}^* \quad = \quad \text{SST} - \hat{\phi} S_{yu}^*$$

where we have defined SST to be the first expression in the first of the above equations. Therefore we may write:

$$\text{SST} = \hat{\phi} S_{yu}^* + C^* \ .$$

As before, we may also partition C^* as:

$$C^* = \sum_{i=1}^{t} \sum_{j=a_i}^{b_i} \delta_{ij} [y_{(ij)} - \bar{y}_i^* + \bar{y}_i^* - \bar{\bar{y}}^* - \hat{\phi} (u_i - \bar{u}^*)]^2$$

$$+ \sum_{i=1}^{t} r_i \beta_i^* \left\{ [y_{(i,a_i)} - \bar{y}_i^* + \bar{y}_i^* - \bar{\bar{y}}^* - \hat{\phi} (u_i - \bar{u}^*)]^2 \right.$$

$$\left. + [y_{(i,b_i)} - \bar{y}_i^* + \bar{y}_i^* - \bar{\bar{y}}^* - \hat{\phi} (u_i - \bar{u}^*)]^2 \right\}$$

$$= \sum_{i=1}^{t} \left\{ \sum_{j=a_i}^{b_i} \delta_{ij} [y_{(ij)} - \bar{y}_i^*]^2 + r_i \beta_i^* \left\{ [y_{(i,a_i)} - \bar{y}_i^*]^2 + y_{(i,b_i)} - \bar{y}_i^*]^2 \right\} \right\}$$

$$+ \sum_{i=}^{t} m_i^* [\bar{y}_i^* - \bar{\bar{y}}^* - \hat{\phi} (u_i - \bar{u}^*)]^2$$

$$= \text{SSE} \quad + \quad \text{SS}_{\text{LOF}}$$

We thus have a similar partition of the total SS for the symmetrically censored case, namely:

$$\text{SST} = \hat{\phi} S_{yu}^* + \text{SS}_{\text{LOF}} + \text{SSE} \ ,$$

and the ANOVA table can be written as:

ANOVA Table for Linear Regression
(for Symmetrically Censored Samples)

Source	SS	df	MS	F
Regression	$\hat{\phi}^* S_{yu}$	1	$\hat{\phi}^* S_{yu}$	MS_R/MSE
Lack of Fit	SS_{LOF}	ℓ-2	$SS_{LOF}/(\ell-2)$	MS_{LOF}/MSE
Error	SSE	A-ℓ	SSE/(A-ℓ)	
Total	SST	A-1	SST/(A-1)	

Once again, since $[A^*-(k/2p)2]\hat{\sigma}^2/\sigma^2$ is approximately χ^2 with A-1 degrees of freedom (by the results in Appendix A1), each partition (when divided by σ^2) will be χ^2 with its corresponding degrees of freedom. Thus to test H_0: $\phi = 0$ we refer the F-ratio MS_R/MSE to a central F-distribution with (1, $A-\ell$) degrees of freedom, and to test for lack-of-fit in the model, we refer MS_{LOF}/MSE to a central F-distribution with (ℓ-2, $A-\ell$) degrees of freedom. To assess the performance of the estimators and test statistics given in Sections 4.4-4.4.2, we have performed Monte Carlo simulations (for various levels of censoring). These results will be reported in subsequent sections. Asymptotically, (n_i tends to infinity and $q_i = r_i/n_i$ are fixed), of course, the estimators above are fully efficient.

4.5 Multiple Regression

In many statistical experiments it is often the case that the investigator wishes to determine how a dependent variable is affected by changes in more than one independent variable. For example, an experiment might be conducted to determine how lung capacity is affected by smoking and exercise. One wishes to then regress the dependent variable (lung capacity) on the two independent variables (smoking and exercise). Regression models such as this in which more than one independent variable contributes to changes in the dependent variable are called *Multiple Regression Models*. As in the simple linear regression case, our aim in this and following sections is to develop non-normal multiple regression procedures by employing the Tiku-Suresh method of estimation. We will go on to show that these results are very powerful and efficient even when the error distribution is non-normal.

Let us consider the following multiple regression model:

$$y_{ij} = \mu + \phi_1 u_{1i} + \phi_2 u_{2i} + \cdots + \phi_c u_{ci} + \varepsilon_{ij} \quad \begin{pmatrix} i = 1,\ldots,\ell \\ j = 1,\ldots,n_i \end{pmatrix} \qquad (4.5.1)$$

where now:

y_{ij} denotes the j^{th} observation at the i^{th} level of the independent variables $u_1,\ldots,u_c$

u_{qi} denotes the i^{th} level of the q^{th} independent variable

ϕ_q is the regression coefficient corresponding to the q^{th} independent variable ($q = 1,\ldots,c$)

ε_{ij} is the random error inherent in y_{ij}.

As in the simple linear regression case, we seek the best-fit solution:

$$\hat{y} = \hat{\mu} + \hat{\phi}_1 u_1 + \hat{\phi}_2 u_2 + \cdots + \hat{\phi}_c u_c \qquad (4.5.2)$$

which takes into account as much of the variability in the data as possible. Note that, in the simple linear regression case, the best-fit solution is a best-fit line in 2-dimensional space. Correspondingly, (4.5.2) will be a best-fit plane in (c+1)-dimensional space.

As before we will employ the method of maximum likelihood to find the best-fit solution (4.5.2). Usually it is assumed that the errors ε_{ij} are i.i.d. $N(0,\sigma^2)$. In this case the solution is well-known and is similar in form to the simple linear regression solution given in Section 4.2, and thus will not be presented here. Instead let us assume that the ε_{ij}'s are distributed as the t-family (2.3.1). The log-likelihood equation based on (4.5.1) is exactly the same as (4.3.3), namely:

$$\ln L = \text{const.} - N\ln\sigma - p\sum_{i=1}^{t}\sum_{j=1}^{n_i}\ln\left\{1 + \frac{z_{(ij)}^2}{k}\right\} \qquad (4.5.3)$$

where we have now defined $z_{(ij)} = [y_{(ij)} - \mu - \sum_{d=1}^{c}\phi_d u_{di}]/\sigma$.

Now differentiating (4.5.3) with respect to μ, ϕ_q ($q = 1,\dots,c$), and σ, we have:

$$\frac{\partial \ln L}{\partial \mu} = \frac{2p}{k\sigma}\sum_{i=1}^{t}\sum_{j=1}^{n_i} g(z_{(ij)}) = 0 \qquad (4.5.4)$$

$$\frac{\partial \ln L}{\partial \phi_q} = \frac{2p}{k\sigma}\sum_{i=1}^{t} u_{qi}\sum_{j=1}^{n_i} g(z_{(ij)}) = 0 \qquad (q = 1,\dots,c) \qquad (4.5.5)$$

$$\frac{\partial \ln L}{\partial \sigma} = -\frac{N}{\sigma} + \frac{2p}{k\sigma}\sum_{i=1}^{t}\sum_{j=1}^{n_i} z_{(ij)}\, g(z_{(ij)}) = 0 \qquad (4.5.6)$$

and substituting the usual approximation $g(z_{(ij)}) \approx \gamma_{ij} + \delta_{ij} z_{(ij)}$ into (4.5.4)-(4.5.6) gives:

$$\frac{\partial \ln L^*}{\partial \mu} = \frac{2p}{k\sigma} \sum_{i=1}^{t} \sum_{j=1}^{n_i} [\gamma_{ij} + \delta_{ij} z_{(ij)}] = 0 \tag{4.5.7}$$

$$\frac{\partial \ln L^*}{\partial \phi_q} = \frac{2p}{k\sigma} \sum_{i=1}^{t} u_{qi} \sum_{j=1}^{n_i} [\gamma_{ij} + \delta_{ij} z_{(ij)}] = 0 \qquad (q = 1, \ldots, c) \tag{4.5.8}$$

$$\frac{\partial \ln L^*}{\partial \sigma} = -\frac{N}{\sigma} + \frac{2p}{k\sigma} \sum_{i=1}^{t} \sum_{j=1}^{n_i} z_{(ij)} [\gamma_{ij} + \delta_{ij} z_{(ij)}] = 0 \tag{4.5.9}$$

Solving (4.5.7) (noting that $\sum_{j=1}^{n_i} \gamma_{ij} = 0$) gives:

$$\hat{\mu} = \bar{\bar{y}} - \sum_{d=1}^{c} \hat{\phi}_d \bar{u}_d \tag{4.5.10}$$

where the $\hat{\phi}_d$'s ($d = 1, \ldots, c$) are to be determined, $M\bar{u}_d = \sum_{i=1}^{t} m_i u_{di}$ and $\bar{\bar{y}}$, m_i, and M are as defined in (4.3.14)-(4.3.15). Substituting (4.5.10) into (4.5.8) and simplifying results in the system of c equations in c unknowns:

$$\sum_{d=1}^{c} \phi_d \sum_{i=1}^{t} m_i (u_{qi} - \bar{u}_q) u_{di} = \sum_{i=1}^{t} m_i (u_{qi} - \bar{u}_q) \bar{y}_i \tag{4.5.11}$$

for $q = 1, \ldots, c$.

If we now define

$$S_{qd} = \sum_{i=1}^{t} m_i (u_{qi} - \bar{u}_q) u_{di} \qquad (q, d = 1, \ldots, c)$$

$$\text{and} \quad Q_q = \sum_{i=1}^{t} m_i (u_{qi} - \bar{u}_q) \bar{y}_i \qquad (q = 1, \ldots, c)$$

we may write (4.5.11) as:

$$\sum_{d=1}^{c} \phi_d S_{qd} = Q_q \qquad (q = 1, \ldots, c) \qquad (4.5.12)$$

In order to solve the system (4.5.12), it is easiest to define the following matrices:

$$\underset{\sim}{\phi} = \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_c \end{bmatrix}, \quad \underset{\sim}{S} = \begin{bmatrix} S_{11} & \cdots & S_{1c} \\ S_{21} & \cdots & S_{2c} \\ \vdots & \ddots & \vdots \\ S_{c1} & \cdots & S_{cc} \end{bmatrix}, \quad \text{and } \underset{\sim}{Q} = \begin{bmatrix} Q_1 \\ Q_2 \\ \vdots \\ Q_c \end{bmatrix}$$

using which we may write (4.5.12) in matrix form as:

$$\underset{\sim}{S}\,\underset{\sim}{\phi} = \underset{\sim}{Q}$$

from which the solution of the system is clearly:

$$\hat{\underset{\sim}{\phi}} = \underset{\sim}{S}^{-1}\underset{\sim}{Q} \qquad (4.5.13)$$

Note that each element of $\underset{\sim}{S}$ is of the same form as S_{uu} given in (4.3.17), and each element of $\underset{\sim}{Q}$ corresponds to S_{yu} in (4.3.18). Therefore (4.5.13) represents a generalization of the MML solution $\hat{\phi} = S_{yu}/S_{uu}$ given in (4.3.11) (for more than one ϕ). Obviously, then, if we set $c = 1$ in (4.5.10) and (4.5.13) we recover the simple linear regression solutions given in Section 4.3.

Now substituting (4.5.10) and (4.5.13) into (4.5.9) and solving gives the MML estimator for σ:

$$\hat{\sigma} = \frac{B + \sqrt{B^2 + 4NC}}{2\sqrt{N(N-(c+1))}}, \quad \text{corrected for bias} \qquad (4.5.14)$$

where

$$B = \frac{2p}{k}\sum_{i=1}^{t}\sum_{j=1}^{n_i}\gamma_{ij}\,[\,y_{(ij)} - \bar{\bar{y}} - \sum_{d=1}^{c}\hat{\phi}_d\,(\,u_{di} - \bar{u}_d)\,]$$

$$\text{and}\quad C = \frac{2p}{k}\sum_{i=1}^{t}\sum_{j=1}^{n_i}\delta_{ij}\,[\,y_{(ij)} - \bar{\bar{y}} - \sum_{d=1}^{c}\hat{\phi}_d\,(\,u_{di} - \bar{u}_d)\,]^2$$

where the $\hat{\phi}_d$'s are the elements of the solution matrix $\hat{\underline{\phi}}$ given in (4.5.13). Once again, we see that when we set c = 1 in the above expressions, (4.5.14) reverts back to $\hat{\sigma}$ for simple linear regression.

4.5.1 Asymptotic Variances and Covariances for Multiple Regression

Differentiating (4.5.7)-(4.5.9) again with respect to μ, ϕ_d (d = 1,...,c), and σ, simplifying and taking negative expectation proceeds in the same manner as the procedure for the simple linear regression model given in Section 4.3.1, and will not be shown here. This results in the following information matrix, which is clearly a generalization of the matrix given in Section 4.3.1:

$$\underline{I}(\mu,\underline{\phi},\sigma) = \frac{2p}{k\sigma^2}\begin{bmatrix} M & M\bar{u}_1 & M\bar{u}_2 & \cdots & M\bar{u}_c & 0 \\ & \sum_{i=1}^{t} m_i u_{1i}^2 & \sum_{i=1}^{t} m_i u_{1i} u_{2i} & \cdots & \sum_{i=1}^{t} m_i u_{1i} u_{ci} & 0 \\ & & \sum_{i=1}^{t} m_i u_{2i}^2 & \cdots & \sum_{i=1}^{t} m_i u_{2i} u_{ci} & 0 \\ & & & \ddots & \vdots & \vdots \\ & & & & \sum_{i=1}^{t} m_i u_{ci}^2 & 0 \\ & & & & & \Omega \end{bmatrix}$$

where $\frac{2p}{k\sigma^2}\,\Omega = -E\left(\frac{\partial^2 \ln L^*}{\partial\sigma^2}\right)$

from which the asymptotic variance-covariance matrix is $\underset{\sim}{V}(\mu,\underset{\sim}{\phi},\sigma) = \underset{\sim}{I}^{-1}(\mu,\underset{\sim}{\phi},\sigma)$. As mentioned, this matrix generalizes the one for simple linear regression, and reduces to that result when c = 1 (i.e. a single ϕ).

4.5.2 ANOVA for Multiple Regression

In the same manner as Section 4.3.2, we may write:

$$\frac{k}{2p}[N-(c+1)]\,\hat{\sigma}^2 = \frac{k}{2p}C$$

$$= \sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\delta_{ij}\,[y_{(ij)} - \bar{\bar{y}} - \sum_{d=1}^{c}\hat{\phi}_d\,(u_{di} - \bar{u}_d)]^2$$

$$= \sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\delta_{ij}\left\{[y_{(ij)} - \bar{\bar{y}}]^2 - 2\,[y_{(ij)} - \bar{\bar{y}}]\sum_{d=1}^{c}\hat{\phi}_d\,(u_{di} - \bar{u}_d) + \sum_{i=1}^{\ell}m_i\left[\sum_{d=1}^{c}\hat{\phi}_d\,(u_{di} - \bar{u}_d)\right]^2\right\}$$

$$= \sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\delta_{ij}\,[y_{(ij)} - \bar{\bar{y}}]^2 - 2\sum_{d=1}^{c}\hat{\phi}_d\sum_{i=1}^{\ell}m_i\,(u_{di} - \bar{u}_d)\,\bar{y}_i + \sum_{i=1}^{\ell}m_i\left[\sum_{d=1}^{c}\hat{\phi}_d\,(u_{di} - \bar{u}_d)\right]^2$$

$$= \text{SST} - 2\sum_{d=1}^{c}\hat{\phi}_d Q_d + \sum_{d=1}^{c}\hat{\phi}_d^2\sum_{i=1}^{\ell}m_i\,(u_{di} - \bar{u}_d)^2 + \sum_{\substack{d=1\\ d\neq e}}^{c}\sum_{e=1}^{c}\hat{\phi}_d\hat{\phi}_e\sum_{i=1}^{\ell}m_i\,(u_{di} - \bar{u}_d)\,(u_{ei} - \bar{u}_e)$$

The last line above may be simplified using the matrix notation introduced in Section 4.5:

$$\frac{k}{2p}C = \text{SST} - 2\,\underset{\sim}{\hat{\phi}}\,\underset{\sim}{Q} + \underset{\sim}{\hat{\phi}}'\,\underset{\sim}{S}\,\underset{\sim}{\hat{\phi}} \qquad (4.5.2.1)$$

Now, noting that $\underset{\sim}{S}\underset{\sim}{\hat{\phi}} = \underset{\sim}{Q}$, rearranging (4.5.2.1) gives:

$$\text{SST} = \underset{\sim}{\hat{\phi}}'\,\underset{\sim}{Q} + \frac{k}{2p}C\,. \qquad (4.5.2.2)$$

Also, we may further break down $(k/2p)C$ as:

$$\frac{k}{2p}C = \sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\delta_{ij}\,[y_{(ij)} - \bar{y}_i + \bar{y}_i - \bar{\bar{y}} - \sum_{d=1}^{c}\hat{\phi}_d\,(u_{di} - \bar{u}_d)]^2$$

$$= \sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\delta_{ij}\,[y_{(ij)} - \bar{y}_i]^2 + \sum_{i=1}^{\ell} m_i\,[\bar{y}_i - \bar{\bar{y}} - \sum_{d=1}^{c}\hat{\phi}_d\,(u_{di} - \bar{u}_d)]^2$$

$$= \qquad \text{SSE} \qquad + \qquad \text{SS}_{\text{LOF}}$$

using which we may write (4.5.2.2) as:

$$\text{SST} = \hat{\underset{\sim}{\phi}}'\underset{\sim}{Q} + \text{SS}_{\text{LOF}} + \text{SSE}\ .$$

Based on this partitioning, we may form the ANOVA Table in the usual manner:

ANOVA Table for Multiple Regression

Source	SS	df	MS	F
Regression	$\hat{\underset{\sim}{\phi}}'\underset{\sim}{Q}$	c	$\hat{\underset{\sim}{\phi}}'\underset{\sim}{Q}/c$	MS_R/MSE
Lack of Fit	SS_{LOF}	ℓ-(c+1)	$SS_{LOF}/[\ell-(c+1)]$	MS_{LOF}/MSE
Error	SSE	N-ℓ	SSE/(N-ℓ)	
Total	SST	N-1	SST/(N-1)	

To test the hypothesis H_0: $\phi_d = 0$ $(d = 1,\ldots,c)$ the F-ratio MS_R/MSE is referred to a central F-distribution with $(c, N-\ell)$ degrees of freedom. To test for lack-of-fit, MS_{LOF}/MSE is referred to a central F-distribution with $(\ell-(c+1), N-\ell)$ degrees of freedom. As for the simple linear regression model, we have performed Monte Carlo simulations to assess the efficiency of these estimators and test statistics for the multiple regression case. The results will be reported in a later section along with the simple linear regression simulations.

4.6 Multiple Regression Based on Type-II Symmetrically Censored Samples

As we have done with the simple linear regression model, the multiple regression results given in Sections 4.5-4.5.2 may be extended to situations in which complete samples are not available in each group i. Rather, let us consider the situation in which r_i observations have been censored from the left and right of the i^{th} ordered group, as discussed in Section 4.4. The log-likelihood function will then be identical to the simple linear regression case, i.e.:

$$\ln L \propto \text{const.} - A\ln\sigma - p\sum_{i=1}^{t}\sum_{j=a_i}^{b_i}\left\{1 + \frac{z_{(ij)}^2}{k}\right\}$$
$$+ \sum_{i=1}^{t} r_i\left\{\ln F(z_{(i,a_i)}) + \ln\left[1 - F(z_{(i,b_i)})\right]\right\} \qquad \textbf{(4.6.1)}$$

where $a_i = r_i + 1$, $b_i = n_i - r_i$, $A = \sum(n_i - 2r_i)$, and (as in Section 4.5) we have defined $z_{(ij)} = [y_{(ij)} - \mu - \sum_{d=1}^{c}\phi_d u_{di}]/\sigma$ . . Differentiating **(4.6.1)** with respect to the model parameters, and following the same procedure as for the simple linear regression model, we arrive at the following derivatives of the modified likelihood function L^*:

$$\frac{\partial \ln L^*}{\partial \mu} = \frac{1}{\sigma}\sum_{i=1}^{\ell}\sum_{j=a_i}^{b_i}[\gamma_{ij} + \delta_{ij} z_{(ij)} + \frac{1}{\sigma}\sum_{i=1}^{\ell} r_i \beta_i^* [z_{(i,a_i)} + z_{(i,b_i)}] = 0 \qquad (4.6.2)$$

$$\frac{\partial \ln L^*}{\partial \phi_q} = \frac{1}{\sigma}\sum_{i=1}^{\ell} u_{qi}\sum_{j=a_i}^{b_i}[\gamma_{ij} + \delta_{ij} z_{(ij)}]$$

$$+ \frac{1}{\sigma}\sum_{i=1}^{\ell} r_i \beta_i^* u_{qi} [z_{(i,a_i)} + z_{(i,b_i)}] = 0 \qquad (q = 1, \dots, c) \qquad (4.6.3)$$

$$\frac{\partial \ln L^*}{\partial \sigma} = -\frac{A^*}{\sigma} + \frac{1}{\sigma}\sum_{i=1}^{\ell}\sum_{j=a_i}^{b_i} z_{(ij)} [\gamma_{ij} + \delta_{ij} z_{(ij)}]$$

$$+ \frac{1}{\sigma}\sum_{i=1}^{\ell} r_i [\alpha_i^* (z_{(i,b_i)} - z_{(i,a_i)}) + \beta_i^* (z^2_{(i,a_i)} + z^2_{(i,b_i)})] = 0 \qquad (4.6.4)$$

To solve (4.6.2)-(4.6.4), we employ the notation defined in equations (4.4.9)-(4.4.12), as well as the following expressions:

$$\bar{u}_q^* = \frac{1}{M^*}\sum_{i=1}^{\ell} m_i^* u_{qi} \qquad (q = 1, \dots, c) \qquad (4.6.5)$$

$$S_{qd}^* = \sum_{i=1}^{\ell} m_i^* (u_{qi} - \bar{u}_q^*) u_{di} \qquad (q, d = 1, \dots, c) \qquad (4.6.6)$$

$$Q_q^* = \sum_{i=1}^{\ell} m_i^* (u_{qi} - \bar{u}_q^*) \bar{y}_i^* \qquad (q = 1, \dots, c) \qquad (4.6.7)$$

Using the above notation, the solution of (4.6.2) is:

$$\hat{\mu} = \bar{\bar{y}}^* - \sum_{d=1}^{c} \hat{\phi}_d \bar{u}_d^* \qquad (4.6.8)$$

where the $\hat{\phi}_d$'s are to be determined. Substituting (4.6.8) into (4.6.3) and simplifying, we have:

$$\sum_{d=1}^{c} \phi_d S_{qd}^* = Q_q^* \qquad (q = 1, \dots, c)$$

$$\text{or} \quad \underline{S}^* \underline{\hat{\phi}} = \underline{Q}^* \qquad (4.6.9)$$

where $\underset{\sim}{S}^*$, $\underset{\sim}{Q}^*$, and $\underset{\sim}{\phi}$ are identical to the matrices defined in Section 4.5, with the expressions replaced with those based on symmetrically censored samples. Therefore the solution of **(4.6.9)** is identical to **(4.5.13)**, namely:

$$\hat{\underset{\sim}{\phi}} = \underset{\sim}{S}^{*-1} \underset{\sim}{Q}^* . \qquad \textbf{(4.6.10)}$$

Now solving **(4.6.4)** gives:

$$\hat{\sigma} = \frac{B^* + \sqrt{B^{*2} + 4A^*C^*}}{2\sqrt{A^*(A^* - \frac{k}{2p}(c+1))}}, \quad \text{corrected for bias} \qquad \textbf{(4.6.11)}$$

where

$$B^* = \sum_{i=1}^{\ell} \sum_{j=a_i}^{b_i} \gamma_{ij} \, [y_{(ij)} - \bar{\bar{y}}^* - \sum_{d=1}^{c} \hat{\phi}_d (u_{di} - \bar{u}_d^*)]$$

$$+ \sum_{i=1}^{\ell} r_i \alpha^*_i \, [y_{(i,b_i)} - y_{(i,a_i)}]$$

and
$$C^* = \sum_{i=1}^{\ell} \sum_{j=a_i}^{b_i} \delta_{ij} \, [y_{(ij)} - \bar{\bar{y}}^* - \sum_{d=1}^{c} \hat{\phi}_d (u_{di} - \bar{u}_d^*)]^2$$

$$+ \sum_{i=1}^{\ell} r_i \beta^*_i \left\{ [y_{(i,a_i)} - \bar{\bar{y}}^* - \sum_{d=1}^{c} \hat{\phi}_d (u_{di} - \bar{u}_d^*)]^2 \right.$$

$$\left. + [y_{(i,b_i)} - \bar{\bar{y}}^* - \sum_{d=1}^{c} \hat{\phi}_d (u_{di} - \bar{u}_d^*)]^2 \right\} .$$

Equations **(4.6.8)**, **(4.6.10)**, and **(4.6.11)** are then the MML estimators for the model parameters based on Type-II symmetrically censored samples. It is quite clear that when there is no censoring ($r_i = 0$, $i = 1,\dots,\ell$) all results reduce to those given in Section 4.5.

4.6.1 Multiple Regression Variance-Covariance Matrix for Symmetrically Censored Samples

A similar argument as in Section 4.5 gives the following information matrix based on censoring:

$$\underset{\sim}{I}(\mu,\underset{\sim}{\phi},\sigma) = \frac{2p}{k\sigma^2}\begin{bmatrix} M^* & M^*\overline{u}_1^* & M^*\overline{u}_2^* & \cdots & M^*\overline{u}_c^* & 0 \\ & \sum_{i=1} m_i^* u_{1i}^2 & \sum_{i=1} m_i^* u_{1i}u_{2i} & \cdots & \sum_{i=1} m_i^* u_{1i}u_{ci} & 0 \\ & & \sum_{i=1} m_i^* u_{2i}^2 & \cdots & \sum_{i=1} m_i^* u_{2i}u_{ci} & 0 \\ & & & \ddots & \vdots & \vdots \\ & & & & \sum_{i=1} m_i^* u_{ci}^2 & 0 \\ & & & & & \Omega^* \end{bmatrix}$$

where $\frac{2p}{k\sigma^2}\Omega^* = -E\left(\frac{\partial \ln L^*}{\partial \sigma^2}\right)$

which gives the asymptotic variance covariance matrix as $\underset{\sim}{V}(\mu,\underset{\sim}{\phi},\sigma) = \underset{\sim}{I}^{-1}(\mu,\underset{\sim}{\phi},\sigma)$. Note that the matrix given above is identical to the one in Section 4.5.1, with each element simply being replaced by the corresponding expression based on censored samples.

4.6.2 Multiple Regression ANOVA Table For Censored Samples

Following the same procedure as in Section 4.5.2, we arrive at the following partitioning of the total sum-of-squares:

$$SST = \sum_{i=1}^{t} \left\{ \sum_{j=a_i}^{b_i} \delta_{ij} [y_{(ij)} - \bar{\bar{y}}^*]^2 + r_i \beta_i^* \left\{ [y_{(i,a_i)} - \bar{\bar{y}}^*]^2 + [y_{(i,b_i)} - \bar{\bar{y}}^*]^2 \right\} \right\}$$

$$SS_R = \hat{\underset{\sim}{\phi}}' \underset{\sim}{Q}^*$$

$$SS_{LOF} = \sum_{i=1}^{t} m_i^* [\bar{y}_i^* - \bar{\bar{y}}^* - \sum_{d=1}^{c} \hat{\phi}_d (u_{di} - \bar{u}_d^*)]^2$$

$$SSE = \sum_{i=1}^{t} \left\{ \sum_{j=a_i}^{b_i} \delta_{ij} [y_{(ij)} - \bar{y}_i^*]^2 + r_i \beta_i^* \left\{ [y_{(i,a_i)} - \bar{y}_i^*]^2 + [y_{(i,b_i)} - \bar{y}_i^*]^2 \right\} \right\}$$

from which the ANOVA table may be constructed in the same manner as in Section 4.5.2. The testing of hypotheses then proceeds in the usual manner.

4.6.3 <u>Summary</u>

In the first sections of this chapter we have developed non-normal regression procedures for both simple linear regression and multiple regression models incorporating the t-family (2.3.1) as the error distribution. The results are based on the Tiku-Suresh MML estimation procedure, and are derived both for complete and symmetrically censored samples.

Because the MML method is based essentially on asymptotic considerations, it is then necessary to perform small-sample Monte Carlo simulations to assess the efficacy of these methods in "real life" (i.e. small samples) situations. Some simulations have been performed for the linear regression model for complete and symmetrically censored samples. These results, as well a their interpretation, are presented in the next section. Following that, we discuss the problem of asymmetrical censoring. It will be shown that the results are essentially the same as what has been already been demonstrated for symmetric censoring, but does add a slightly increased level of computational complexity.

4.7 Simulation Results

As we have seen in this chapter, the derivation of these non-normal regression procedures relies on asymptotic (large-sample) considerations. It is then necessary to demonstrate that, while asymptotics were used, the resulting procedures still perform quite well for small samples. For purposes of this demonstration, we concentrate on the simple linear regression model, for small samples ($n_1 = 10$), and consider the complete sample case as well as 10% censoring. Various vales of p are used, namely $p = 2$, 3, 4, 5, and 10. (As discussed in Chapter 2, for $p < 3$, it is necessary to censor the sample to ensure that the estimator $\hat{\sigma}$ is tractable.) The results are reported in Tables X-XIV. All simulations were done on a VAX mainframe using FORTRAN-77 and IMSL subroutines, or a MIPS mini-system using FORTRAN-77 and NAG subroutines. Simulated values are based on 10,000 Monte Carlo runs.

From Tables X and XI, it is seen that the estimators for the model parameters perform quite well, even for small values of p. Also, the simulated variance for the estimators agree quite closely with the theoretical (asymptotic) result. This is a clear indication that the procedure works quite well for small samples, regardless of its asymptotic underpinnings. For censored samples, even $p = 2$ results are quite good, even though this represents highly non-normal samples.

Table XII gives the simulated level of significance for the regression F-statistic $F(\text{reg})$. Compared to the fixed level ($\alpha = 0.05$), it is seen that this test statistic performs extremely well, essentially maintaining the observed significance level at the theoretical value of 0.05, throughout the range of p values. The values agree, both for complete and 10% censored samples. Tables XIII and XIV show the results of simulated power for the same test statistic, as ϕ moves away from 0. (We note here that since the test statistic is a function of $\hat{\phi}^2$, it is sufficient to compute the power for positive non-zero values of ϕ.) It can bee seen that the test statistic F(reg) attains very high power quite quickly, especially for censored samples. Once again, this is a clear indication that this non-normal procedure is performing very well, even in the small sample case.

Table X: $\hat{\mu}$, $\hat{\phi}$, and $\hat{\sigma}$ and their Asymptotic and Simulated Variances for $\ell=3$, $n_i=10$, $\mu=0$, $\phi=0$, $\sigma=1$ with $u_i = -1, 0, 1$; p = 3,4,5,10 (Complete Samples)

p	Parameter	Estimate*	Simulated* Variance/σ^2	Asymptotic Variance/σ^2
3.0	μ	-0.0006	0.0281	0.0265
	ϕ	0.0035	0.0417	0.0398
	σ	1.0713	0.0361	0.0267 †
4.0	μ	-0.0027	0.0301	0.0293
	ϕ	0.0049	0.0476	0.0440
	σ	1.0508	0.0299	0.0185 †
5.0	μ	0.0031	0.0323	0.0304
	ϕ	-0.0011	0.0471	0.0456
	σ	1.0394	0.0273	0.0182 †
10.0	μ	-0.0010	0.0335	0.0321
	ϕ	0.0010	0.0490	0.0482
	σ	1.0125	0.0209	0.0175 †

* These values are based on 10,000 Monte Carlo Simulations.

† These values have been replaced by $MVB(\sigma)/\sigma^2 = (p+1)/[2N(p-\frac{1}{2})]$

Table XI: $\hat{\mu}$, $\hat{\phi}$, and $\hat{\sigma}$ and their Asymptotic and Simulated Variances for ℓ=3, n_i=10, μ=0, ϕ=0, σ=1 with u_i = -1, 0, 1; p = 2,3,4,5,10 (10% Censoring)

p	Parameter	Estimate*	Simulated* Variance/σ^2	Asymptotic Variance/σ^2
2.0	μ	0.0000	0.0181	0.0171
	ϕ	-0.0012	0.0280	0.0256
	σ	1.0408	0.0385	0.0417 †
3.0	μ	-0.0005	0.0280	0.0266
	ϕ	0.0036	0.0416	0.0399
	σ	1.0408	0.0385	0.0333 †
4.0	μ	-0.0028	0.0302	0.0297
	ϕ	0.0048	0.0475	0.0446
	σ	1.0231	0.0329	0.0298 †
5.0	μ	0.0033	0.0323	0.0310
	ϕ	-0.0011	0.0474	0.0465
	σ	1.0135	0.0314	0.0227 †
10.0	μ	-0.0011	0.0341	0.0331
	ϕ	0.0010	0.0501	0.0497
	σ	0.9963	0.0268	0.0218 †

* These values are based on 10,000 Monte Carlo Simulations.

† These values have been replaced by $MVB(\sigma)/\sigma^2 = (p+1)/[2A(p-\frac{1}{2})]$

Table XII: **Simulated Level of Significance of F(reg) for Complete Samples and F_c(reg) for 10% Censoring; ℓ=3, n_1=10, μ=0, ϕ=0, σ=1; p = 2, 3, 4, 5, 10 (actual significance level α = 0.05)**

p	P{F(reg) > F_α}	P{F_c(reg) > F_α
2.0		0.0438
3.0	0.0404	0.0451
4.0	0.0496	0.0522
5.0	0.0467	0.0532
10.0	0.0484	0.0522

All values are based on 10,000 Monte Carlo simulations.

Table XIII: Simulated power values for F(reg) for p = 3,4,5,10 (Complete Samples)

p	ϕ_0	P{F(reg) > F_α}
3.0	0.1	0.0995
	0.2	0.2919
	0.3	0.5545
	0.4	0.8006
	0.5	0.9311
	0.6	0.9842
	0.7	0.9966
	0.8	0.9992
	0.9	1.0000
	1.0	1.0000
4.0	0.1	0.1032
	0.2	0.2670
	0.3	0.5238
	0.4	0.7706
	0.5	0.9173
	0.6	0.9804
	0.7	0.9957
	0.8	0.9993
	0.9	1.0000
	1.0	1.0000
5.0	0.1	0.0959
	0.2	0.2672
	0.3	0.5235
	0.4	0.7676
	0.5	0.9104
	0.6	0.9776
	0.7	0.9950
	0.8	0.9987
	0.9	0.9999
	1.0	1.0000
10.0	0.1	0.0962
	0.2	0.2627
	0.3	0.5143
	0.4	0.7566
	0.5	0.9100
	0.6	0.9774
	0.7	0.9962
	0.8	0.9997
	0.9	0.9999
	1.0	1.0000

Table XIV: **Simulated power values for F(reg) for p = 2,3,4,5,10 (10% Censoring)**

p	ϕ_0	P{F(reg) > F_α}
2.0	0.1	0.1384
	0.2	0.4080
	0.3	0.7334
	0.4	0.9265
	0.5	0.9821
	0.6	0.9973
	0.7	0.9995
	0.8	0.9999
	0.9	1.0000
	1.0	1.0000
3.0	0.1	0.1044
	0.2	0.3034
	0.3	0.5675
	0.4	0.8073
	0.5	0.9343
	0.6	0.9865
	0.7	0.9965
	0.8	0.9995
	0.9	0.9998
	1.0	1.0000
4.0	0.1	0.1084
	0.2	0.2770
	0.3	0.5300
	0.4	0.7720
	0.5	0.9196
	0.6	0.9818
	0.7	0.9961
	0.8	0.9992
	0.9	1.0000
	1.0	1.0000
5.0	0.1	0.0991
	0.2	0.2716
	0.3	0.5279
	0.4	0.7656
	0.5	0.9113
	0.6	0.9761
	0.7	0.9944
	0.8	0.9988
	0.9	0.9998
	1.0	0.9999
10.0	0.1	0.0988
	0.2	0.2633
	0.3	0.5067
	0.4	0.7459
	0.5	0.9008
	0.6	0.9738
	0.7	0.9944
	0.8	0.9990
	0.9	1.0000
	1.0	1.0000

4.8 Simple Linear Regression for Asymmetrically Censored Samples

In Section 4.4 we have discussed the Tiku-Suresh MML solution when each of the ℓ groups of ordered observations have been symmetrically censored (i.e. r_i observations are removed from the left and right of the i^{th} group). However, it is often the case in practice that the ordered samples may exhibit asymmetrical censoring. For example, the data may only be left-censored or right-censored, or may have, say, the 3 smallest observations unavailable, and only one largest observation unavailable.

It is of interest, then, to extend the results of Section 4.4 to the asymmetrically censored case. We will see that this no longer allows us to simplify expressions using "symmetry" arguments, and hence requires a greater amount of computation. Still, it will be seen that the results obtained are of a familiar form, and are highly efficient as compared to the normality-based procedure.

Let us begin by considering the usual simple linear regression model (with the t-family (2.3.1) as the error distribution):

$$y_{ij} = \mu + \phi u_i + \varepsilon_{ij} \qquad (i = 1,\dots,\ell;\ \ j = 1,\dots,n_i)$$

where now at each level u_i we order the observations with respect to j and censor the r_i smallest and s_i largest observations, leaving:

$$\begin{array}{ccccc} y_{(1,s_1+1)} & \le y_{(1,s_1+2)} & \le \cdots \le & y_{(1,n_1-s_1)} \\ \vdots & & & \vdots \\ y_{(\ell,r_\ell+1)} & \le y_{(\ell,r_\ell+2)} & \le \cdots \le & y_{(\ell,n_\ell-s_\ell)} \end{array} .$$

Note here that to include left (or right) censored samples, one merely sets r_i (or s_i) equal to zero.

Let us now define $a_i = r_i + 1$ and $b_i = n_i - s_i$. Then the log-likelihood function can be written as:

$$\ln L \propto \text{const.} - A\ln\sigma - p\sum_{i=1}^{\ell}\sum_{j=a_i}^{b_i}\ln\left\{1+\frac{z_{(ij)}^2}{k}\right\}$$
$$+\sum_{i=1}^{\ell}\{\, r_i\ln F(z_{(i,a_i)}) + s_i\ln[1 - F(z_{(i,b_i)})]\,\} \tag{4.8.1}$$

where $A = \sum(n_i - r_i - s_i)$ and $z_{(ij)} = [y_{(ij)} - \mu - \phi u_i]/\sigma$. Differentiating (4.8.1) with respect to μ, ϕ, and σ, and employing our usual definitions of g, h_1, and h_2, we arrive at the following equations:

$$\frac{\partial\ln L}{\partial\mu} = \frac{2p}{k\sigma}\sum_{i=1}^{\ell}\sum_{j=a_i}^{b_i} g(z_{(ij)})$$
$$+\frac{1}{\sigma}\sum_{i=1}^{\ell}[\, s_i h_2(z_{(i,b_i)}) - r_i h_1(z_{(i,a_i)})\,] = 0 \tag{4.8.2}$$

$$\frac{\partial\ln L}{\partial\phi} = \frac{2p}{k\sigma}\sum_{i=1}^{\ell} u_i\sum_{j=a_i}^{b_i} g(z_{(ij)})$$
$$+\frac{1}{\sigma}\sum_{i=1}^{\ell} u_i[\, s_i h_2(z_{(i,b_i)}) - r_i h_1(z_{(i,a_i)})\,] = 0 \tag{4.8.3}$$

$$\frac{\partial\ln L}{\partial\sigma} = -\frac{A}{\sigma} + \frac{2p}{k\sigma}\sum_{i=1}^{\ell}\sum_{j=a_i}^{b_i} z_{(ij)}\, g(z_{(ij)})$$
$$+\frac{1}{\sigma}\sum_{i=1}^{\ell}[\, s_i z_{(i,b_i)} h_2(z_{(i,b_i)}) - r_i z_{(i,a_i)} h_1(z_{(i,a_i)})\,] = 0 \tag{4.8.4}$$

As mentioned previously, equations (4.8.2)-(4.8.4) do not admit explicit ML estimators due to the presence of the functions g, h_1, and h_2, and must be solved iteratively. To avoid numerical iterations, we will approximate these functions as done before. As usual let

$$g(z_{(ij)}) = \gamma_{ij} + \delta_{ij} z_{(ij)} \tag{4.8.5}$$

where γ_{ij} and δ_{ij} have been defined previously. Now, however, since the censoring is no longer symmetric, we approximate h_1 and h_2 as:

$$h_1(z_{(ij)}) \approx \alpha_{1i} - \beta_{1i} z_{(ij)} \qquad (4.8.6)$$

$$\text{and} \quad h_2(z_{(ij)}) \approx \alpha_{2i} + \beta_{2i} z_{(ij)} \qquad (4.8.7)$$

where now if we define $q_{1i} = r_i/n_i$ and $q_{2i} = s_i/n_i$, α_{1i}, α_{2i}, and β_{1i}, β_{2i} are:

$$\beta_{1i} = -f(t_{1i}) \left\{ \frac{2p}{k} g(t_{1i}) - \frac{f(t_{1i})}{q_{1i}} \right\} \Big/ q_{1i}$$

$$\alpha_{1i} = \frac{f(t_{1i})}{q_{1i}} - \beta_{1i} t_{1i}$$

$$\text{and} \quad \beta_{2i} = -f(t_{2i}) \left\{ \frac{2p}{k} g(t_{2i}) - \frac{f(t_{2i})}{q_{2i}} \right\} \Big/ q_{2i}$$

$$\alpha_{2i} = \frac{f(t_{2i})}{q_{2i}} - \beta_{2i} t_{2i}$$

for large n_i. The values t_{1i} and t_{2i} above are the solutions of the equations:

$$\int_{-\infty}^{t_{1i}} f(z)\,dz = 1 - q_{1i}$$

$$\text{and} \quad \int_{-\infty}^{t_{2i}} f(z)\,dz = q_{2i} \,.$$

Note that when the censoring is symmetric (i.e. $r_i = s_i$), $\alpha_{1i} = \alpha_{2i} = \alpha_i$, and $\beta_{1i} = \beta_{2i} = \beta_i$ given in the previous sections. Now substituting the approximations (4.8.5) and (4.8.6)-(4.8.7) into (4.8.2)-(4.8.4) and simplifying produces the following modified derivatives:

$$\frac{\partial \ln L^*}{\partial \mu} = \frac{1}{\sigma}\sum_{i=1}^{\ell}\sum_{j=a_i}^{b_i}[\gamma_{ij} + \delta_{ij} z_{(ij)}]$$

$$+ \frac{1}{\sigma}\sum_{i=1}^{\ell}[(s_i\alpha_{2i}^* - r_i\alpha_{1i}^*) + r_i\beta_{1i}^* z_{(i,a_i)} + s_i\beta_{2i}^* z_{(i,b_i)}] = 0 \qquad (4.8.8)$$

$$\frac{\partial \ln L^*}{\partial \phi} = \frac{1}{\sigma}\sum_{i=1}^{\ell} u_i \sum_{j=a_i}^{b_i}[\gamma_{ij} + \delta_{ij} + z_{(ij)}]$$

$$+ \frac{1}{\sigma}\sum_{i=1}^{\ell} u_i [(s_i\alpha_{2i}^* - r_i\alpha_{1i}^*) + r_i\beta_{1i}^* z_{(i,a_i)} + s_i\beta_{2i}^* z_{(i,b_i)}] = 0 \qquad (4.8.9)$$

$$\frac{\partial \ln L^*}{\partial \sigma} = -\frac{A^*}{\sigma} + \frac{1}{\sigma}\sum_{i=1}^{\ell}\sum_{j=a_i}^{b_i} z_{(ij)}[\gamma_{ij} + \delta_{ij} z_{(ij)}]$$

$$+ \frac{1}{\sigma}\sum_{i=1}^{\ell}[(s_i\alpha_{2i}^* z_{(i,b_i)} - r_i\alpha_{1i} z_{(i,a_i)})$$

$$+ r_i\beta_{1i}^* z_{(i,a_i)}^2 + s_i\beta_{2i}^* z_{(i,b_i)}^2] = 0 \qquad (4.8.10)$$

Note that we have divided out $2p/k$ throughout and defined:

$$A = \frac{2p}{k}A^*$$

$$\left.\begin{aligned} \alpha_{\cdot i} &= \frac{2p}{k}\alpha_{\cdot i}^* \\ \beta_{\cdot i} &= \frac{2p}{k}\beta_{\cdot i}^* \end{aligned}\right\} \quad (\cdot = 1,2)$$

Let us first consider **(4.8.8)**. Substituting for $z_{(ij)}$ and simplifying, we have:

$$\sigma\sum_{i=1}^{\ell}\left\{\sum_{j=a_i}^{b_i}\gamma_{ij} + (s_i\alpha_{2i}^* - r_i\alpha_{1i}^*)\right\}$$

$$+\sum_{i=1}^{\ell}\left\{\sum_{j=a_i}^{b_i}\delta_{ij} y_{(ij)} + r_i\beta_{1i}^* y_{(i,a_i)} + s_i\beta_{2i}^* y_{(i,b_i)}\right\}$$

$$-\mu\sum_{i=1}^{\ell}[m_i + r_i\beta_{1i}^* + s_i\beta_{2i}^*]$$

$$-\phi\sum_{i=1}^{\ell}[m_i + r_i\beta_{1i}^* + s_i\beta_{2i}^*]u_i = 0 \qquad (4.8.11)$$

where $m_i = \sum_{j=a_i}^{b_i} \delta_{ij}$. Notice that the term involving σ cannot be eliminated (as in the symmetric censoring case) because $\sum_{j=a_i}^{b_i} \gamma_{ij} \neq 0$ and $s_1\alpha_{21}^* \neq r_1\alpha_{11}^*$. To solve (4.8.11), then, let us define the following notation, some of which has been used before:

$$m_i^* = m_i + r_i\beta_{1i}^* + s_i\beta_{2i}^* \; ; \qquad M^* = \sum_{i=1}^{t} m_i^*$$

$$\Gamma_i = \sum_{j=a_i}^{b_i} \gamma_{ij}$$

$$\Gamma_i^* = \Gamma_i + (s_i\alpha_{2i}^* - r_i\alpha_{1i}^*) \; ; \qquad \overline{\Gamma}^* = \frac{1}{M^*}\sum_{i=1}^{t} \Gamma_i^*$$

$$\overline{y}_i^* = \frac{1}{m_i^*}\left\{m_i\overline{y}_i + r_i\beta_{1i}^* y_{(i,a_i)} + s_i\beta_{2i}^* y_{(i,b_i)}\right\}$$

$$\overline{\overline{y}}^* = \frac{1}{M^*}\sum_{i=1}^{t} m_i^*\overline{y}_i^*$$

$$\overline{u}^* = \frac{1}{M^*}\sum_{i=1}^{t} m_i^* u_i \quad .$$

Using this notation, we may write (4.8.11) as:

$$M^*\overline{\Gamma}^*\sigma + M^*\overline{\overline{y}}^* - M^*\mu - M^*\overline{u}^*\phi = 0$$

which gives the estimator of μ as:

$$\hat{\mu} = \overline{\overline{y}}^* - \hat{\phi}\,\overline{u}^* + \hat{\sigma}\,\overline{\Gamma}^* \qquad (4.8.12)$$

where $\hat{\phi}$ and $\hat{\sigma}$ are to be determined. We immediately see that (4.8.12) is of the same form as $\hat{\mu}$ based on symmetric censoring, except for the extra term $\hat{\sigma}\,\overline{\Gamma}^*$. In fact, when the censoring is symmetric, $\overline{\Gamma}^* = 0$ and (4.8.12) reduces to the symmetric result, as is expected.

Now we substitute (4.8.12) into (4.8.9), and simplify:

$$\sigma \sum_{i=1}^{t} \Gamma_i^* (u_i - \bar{u}^*) - \phi \sum_{i=1}^{t} m_i^* (u_i - \bar{u}^*)^2$$

$$+ \sum_{i=1}^{t} [m_i \bar{y}_i + r_i \beta_{1i}^* y_{(i,a_i)} + s_i \beta_{2i}^* y_{(i,b_i)}] (u_i - \bar{u}^*) = 0 \qquad (4.8.13)$$

Let us define

$$S_{uu}^* = \sum_{i=1}^{t} m_i^* (u_i - \bar{u}^*)^2$$

$$S_{yu}^* = \sum_{i=1}^{t} m_i^* (u_i - \bar{u}^*) \bar{y}_i^*$$

$$S_\Gamma^* = \sum_{i=1}^{t} \Gamma_i^* (u_i - \bar{u}^*)$$

using which the solution of (4.8.13) is:

$$\hat{\phi} = \frac{S_{yu}^*}{S_{uu}^*} + \hat{\sigma} \frac{S_\Gamma^*}{S_{uu}^*} , \qquad (4.8.14)$$

where $\hat{\sigma}$ is to be determined. Once again, when the censoring is symmetric, $S_\Gamma^* = 0$, and (4.8.14) reduces to the symmetric result.

In order to obtain $\hat{\sigma}$, we must substitute (4.8.12) and (4.8.14) into (4.8.10), simplify, and collect terms involving σ^2, σ, and the constant term, which gives:

$$\hat{\sigma} = \frac{B^* + \sqrt{B^{*2} + 4 A^* C^*}}{2\sqrt{A^* (A^* - \left(\frac{k}{2p}\right)2)}} , \quad \text{corrected for bias} \qquad (4.8.15)$$

where

$$B^* = \sum_{i=1}^{t} \left\{ \sum_{j=a_i}^{b_i} \gamma_{ij} [y_{(ij)} - \bar{\bar{y}}^* - \frac{S_{yu}^*}{S_{uu}^*} (u_i - \bar{u}^*)] \right.$$

$$\left. + s_i \alpha_{2i}^* [y_{(i,a_i)} - \bar{\bar{y}}^* - \frac{S_{yu}^*}{S_{uu}^*} (u_i - \bar{u}^*)] - r_i \alpha_{1i} [y_{(i,a_i)} - \bar{\bar{y}}^* - \frac{S_{yu}^*}{S_{uu}^*} (u_i - \bar{u}^*)] \right\}$$

$$C^* = \sum_{i=1}^{t} \left\{ \sum_{j=a_i}^{b_i} \delta_{ij} [y_{(ij)} - \bar{\bar{y}}^* - \frac{S^*_{yu}}{S_{uu}} (u_i - \bar{u}^*)]^2 \right.$$

$$\left. + r_i \beta^*_{1i} [y_{(i,a_i)} - \bar{\bar{y}}^* - \frac{S^*_{yu}}{S^*_{uu}} (u_i - \bar{u}^*)]^2 + s_i \beta^*_{2i} [y_{(i,b_i)} - \bar{\bar{y}}^* - \frac{S^*_{yu}}{S^*_{uu}} (u_i - \bar{u}^*)]^2 \right\}$$

Note that in the case of symmetric censoring, $r_i = s_i$, $\alpha_{1i}{}^* = \alpha_{2i}{}^*$, $\beta_{1i}{}^* = \beta_{2i}{}^*$, and $S_{yu}{}^*/S_{uu}{}^* = \hat{\phi}$; therefore B^* and C^* (and hence $\hat{\sigma}$) reduce to the results given in the symmetric censoring case.

Thus equations **(3.8.12)**, **(3.8.14)**, and **(3.8.18)** constitute the MML estimators for the simple linear regression model based on asymmetrically censored samples. Under symmetric censoring, it is easily seen that these estimators reduce to those given in Section 3.4, and are of a familiar form. It may be noted that for p equal to ∞ (i.e. when the error distribution is normal), all the results given in this chapter reduce to those of Tiku (1978).

4.8.1 Variance-Covariance Matrix for Asymmetrically Censored Samples

To obtain the asymptotic variance-covariance matrix in the case of asymmetric censoring, the usual procedure is followed. Differentiating (4.8.8)-(4.8.10) again with respect to μ, ϕ, and σ and taking negative expectation gives (where $t_{(ij)} = E(z_{(ij)})$ and $\sigma_{jj:n_i} = V(z_{(ij)})$):

$$-E\left(\frac{\partial^2 \ln L^*}{\partial\mu^2}\right) = \frac{2p}{k\sigma^2} M^*$$

$$-E\left(\frac{\partial^2 \ln L^*}{\partial\mu\partial\phi}\right) = \frac{2p}{k\sigma^2} M^* \overline{u}^*$$

$$-E\left(\frac{\partial^2 \ln L^*}{\partial\mu\partial\sigma}\right) = \frac{2p}{k\sigma^2}\left\{ M^*\overline{\Gamma}^* + 2\sum_{i=1}^{\ell}\left[\sum_{j=a_i}^{b_i}\delta_{ij}t_{(ij)} + r_i\beta_{1i}^* t_{(i,a_i)} + s_i\beta_{2i}^* t_{(i,b_i)}\right]\right\}$$

$$-E\left(\frac{\partial^2 \ln L^*}{\partial\phi^2}\right) = \frac{2p}{k\sigma^2}\sum_{i=1}^{\ell} m_i^* u_i^2$$

$$-E\left(\frac{\partial^2 \ln L^*}{\partial\phi\partial\sigma}\right) = \frac{2p}{k\sigma^2}\sum_{i=1}^{\ell} u_i\left\{\Gamma_i^* + 2\left[\sum_{j=a_i}^{b_i}\delta_{ij}t_{(ij)} + r_i\beta_{1i}^* t_{(i,a_i)} + s_i\beta_{2i}^* t_{(i,b_i)}\right]\right\}$$

$$\begin{aligned}-E\left(\frac{\partial^2 \ln L^*}{\partial\sigma^2}\right) = \frac{2p}{k\sigma^2}\Bigg\{ & -A^* + \sum_{i=1}^{\ell}\sum_{j=a_i}^{b_i}\left[2\gamma_{ij}t_{(ij)} + 3\delta_{ij}\left(t_{(ij)}^2 + \sigma_{jj:n_i}\right)\right] \\ & + \sum_{i=1}^{\ell}\Big\{ 2\left(s_i\alpha_{2i}^* t_{(i,b_i)} - r_i\alpha_{1i}^* t_{(i,a_i)}\right) \\ & \quad + 3\left[r_i\beta_{1i}^*\left(t_{(i,a_i)}^2 + \sigma_{a_i,a_i:n_i}\right) + s_i\beta_{2i}^*\left(t_{(i,b_i)}^2 + \sigma_{b_i,b_i:n_i}\right)\right]\Big\}\Bigg\}\end{aligned}$$

Now, writing:

$$t_i^* = \sum_{j=a_i}^{b_i}\delta_{ij}t_{(ij)} + r_i\beta_{1i}^* t_{(i,a_i)} + s_i\beta_{2i}^* t_{(i,b_i)} \qquad (i = 1,\ldots,\ell)$$

and $\frac{2p}{k\sigma^2}\Omega^* = -E\left(\frac{\partial^2 \ln L^*}{\partial\sigma^2}\right)$

the information matrix may be written as:

$$\underset{\sim}{I}(\mu,\phi,\sigma) = \frac{2p}{k\sigma^2}\begin{bmatrix} M^* & M^*\bar{u}^* & M^*\bar{\Gamma}^* + 2\sum_{i=1}^{t} t_i^* \\ & \sum_{i=1}^{t} m_i^* u_i^2 & \sum_{i=1}^{t} u_i(\Gamma_i^* + 2t_i^*) \\ // & & \Omega^* \end{bmatrix}$$

and hence the asymptotic variance-covariance matrix is obtained as $\underset{\sim}{V}(\mu,\phi,\sigma) = \underset{\sim}{I}^{-1}(\mu,\phi,\sigma)$. It is of interest here to note that in the complete sample and symmetric censoring cases the elements representing the covariance between μ & σ and ϕ & σ are 0; however in this case these entries are non-zero, indicating that the estimators of μ and ϕ are correlated with $\hat{\sigma}$, which is otherwise clear since the estimators $\hat{\mu}$ and $\hat{\phi}$ contain $\hat{\sigma}$, and therefore must be correlated. Of course, when the censoring is symmetric (or there is no censoring), $\bar{\Gamma}^* = 0$ and $t_i^* = 0$, and the results reduce to the earlier ones.

4.8.2 The ANOVA Table for Asymmetric Censoring

Once again, considering (4.8.15), we may write

$$\hat{\sigma} = \frac{\sqrt{C^*}}{\sqrt{A^* - \left(\frac{k}{2p}\right)2}} = \frac{\sqrt{C}}{\sqrt{A-2}}$$

by ignoring the negligible terms involving B^{*2}/A^*C^*. We may then proceed to partition the error sum-of-squares as:

$$(A^* - \left(\frac{k}{2p}\right)2)\,\hat{\sigma}^2 = C^*$$

$$= \sum_{i=1}^{t} \left\{ \sum_{j=a_i}^{b_i} \delta_{ij} [\, y_{(ij)} - \bar{\bar{y}}^*]^2 + r_i \beta_{1i}^{*} [\, y_{(i,a_i)} - \bar{\bar{y}}^*]^2 + s_i \beta_{2i}^{*} [\, y_{(i,b_i)} - \bar{\bar{y}}^*]^2 \right\}$$

$$- 2 \frac{S_{yu}^*}{S_{uu}^*} S_{yu}^* + \left(\frac{S_{yu}^*}{S_{uu}^*}\right)^2 \sum_{i=1}^{t} m_i^* (\, u_i - \bar{u}^*)^2$$

$$= \mathrm{SST} - \left(\frac{S_{yu}^*}{S_{uu}^*}\right)^2 S_{uu}^* \,.$$

Hence we may write

$$\mathrm{SST} = \frac{S_{yu}^{*2}}{S_{uu}^*} + C^*$$

and we may also partition C^* further to obtain:

$$C^* = \sum_{i=1}^{t} \left\{ \sum_{j=a_i}^{b_i} \delta_{ij} [\, y_{(ij)} - \bar{y}_i^*]^2 + r_i \beta_{1i}^{*} [\, y_{(i,a_i)} - \bar{y}_i^*]^2 + s_i \beta_{2i}^{*} [\, y_{(i,b_i)} - \bar{y}_i^*]^2 \right\}$$

$$+ \sum_{i=1}^{t} m_i^* [\, \bar{y}_i^* - \bar{\bar{y}}^* - \frac{S_{yu}^*}{S_{uu}^*} (\, u_i - \bar{u}^*)]^2$$

$$= \quad \mathrm{SSE} \quad + \quad \mathrm{SS_{LOF}}$$

which gives the final partition as:

$$\mathrm{SST} = \frac{S_{yu}^{*2}}{S_{uu}^*} + \mathrm{SS_{LOF}} + \mathrm{SSE} \,. \qquad (4.8.2.1)$$

Notice that, in (4.8.2.1), S_{yu}^{*2}/S_{uu}^* reduces to $\hat{\phi} S_{yu}^*$ when the censoring is symmetric, and the partitions become as given in Section 4.5.2. The ANOVA Table (and the tests of hypotheses) may then be formed in the usual way.

4.9 Multiple Regression with Asymmetric Censoring

Let us once again consider the multiple regression model discussed in Section 4.5:

$$y_{ij} = \mu + \phi_1 u_{1i} + \phi_2 u_{2i} + \cdots + \phi_c u_{ci} + \varepsilon_{ij} \qquad \begin{pmatrix} i = 1,\dots,\ell \\ j = 1,\dots,n_i \end{pmatrix} \quad (4.9.1)$$

where now we order each of the ℓ groups and censor the r_i smallest and s_i largest observations. The log-likelihood function is then similar to (4.6.1):

$$\ln L \propto \text{const.} - A\ln\sigma - p\sum_{i=1}^{\ell}\sum_{j=a_i}^{b_i}\left\{1 + \frac{z_{(ij)}^2}{k}\right\}$$
$$+ \sum_{i=1}^{\ell}\{\, r_i \ln F(z_{(i,a_i)}) + s_i \ln[1 - F(z_{(i,b_i)})]\,\} \qquad (4.9.2)$$

where, as before, $A = \sum(n_i - r_i - s_i)$, $a_i = r_i + 1$, and $b_i = n_i - s_i$. Differentiating (4.9.2) with respect to μ, ϕ_q ($q = 1,\dots,c$), and σ and employing the same approximations given in (4.8.5)-(4.8.7), the modified log-likelihood derivatives are similar to those given in equations (4.8.8)-(4.8.10):

$$\frac{\partial \ln L^*}{\partial \mu} = \frac{1}{\sigma}\sum_{i=1}^{\ell}\sum_{j=a_i}^{b_i}[\gamma_{ij} + \delta_{ij} z_{(ij)}]$$
$$+ \frac{1}{\sigma}\sum_{i=1}^{\ell}[(s_i\alpha_{2i}^* - r_i\alpha_{1i}^*) + r_i\beta_{1i}^* z_{(i,a_i)} + s_i\beta_{2i}^* z_{(i,b_i)}] = 0 \qquad (4.9.3)$$

$$\frac{\partial \ln L^*}{\partial \phi_q} = \frac{1}{\sigma}\sum_{i=1}^{\ell} u_{qi}\sum_{j=a_i}^{b_i}[\gamma_{ij} + \delta_{ij} + z_{(ij)}]$$
$$+ \frac{1}{\sigma}\sum_{i=1}^{\ell} u_{qi}[(s_i\alpha_{2i}^* - r_i\alpha_{1i}^*) + r_i\beta_{1i}^* z_{(i,a_i)} + s_i\beta_{2i}^* z_{(i,b_i)}] = 0 \qquad (4.9.4)$$

$$(q = 1,\dots,c)$$

$$\frac{\partial \ln L^*}{\partial \sigma} = -\frac{A^*}{\sigma} + \frac{1}{\sigma}\sum_{i=1}^{t}\sum_{j=a_i}^{b_i} z_{(ij)} [\gamma_{ij} + \delta_{ij} z_{(ij)}]$$

$$+ \frac{1}{\sigma}\sum_{i=1}^{t} [(s_i \alpha^*_{2i} z_{(i,b_i)} - r_i \alpha_{1i} z_{(i,a_i)})$$

$$+ r_i \beta^*_{1i} z^2_{(i,a_i)} + s_i \beta^*_{2i} z^2_{(i,b_i)}] = 0 \qquad (4.9.5)$$

where $2p/k$ has been divided out in the right-hand side of all equations. In the same manner as Section 4.5, we proceed to solve equations (4.9.3)-(4.9.5), noting that $\sum_{j=a_i}^{b_i} \gamma_{ij} \neq 0$. The solution of (4.9.3) gives:

$$\hat{\mu} = \bar{\bar{y}}^* - \sum_{d=1}^{c} \hat{\phi}_d \bar{u}^*_d + \bar{\Gamma}^* \hat{\sigma} \qquad (4.9.6)$$

where the $\hat{\phi}_d$'s and $\hat{\sigma}$ are to be determined. Substituting (4.9.6) into (4.9.4) and defining:

$$S^*_{qd} = \sum_{i=1}^{t} m^*_i (u_{qi} - \bar{u}^*_q) u_{di} \qquad (q, d = 1, \ldots, c)$$

$$S^*_{\Gamma_q} = \sum_{i=1}^{t} \Gamma^*_i (u_{qi} - \bar{u}^*_q) \qquad (q = 1, \ldots, c)$$

$$Q^*_q = \sum_{i=1}^{t} m^*_i (u_{qi} - \bar{u}^*_q) \bar{y}^*_i \qquad (q = 1, \ldots, c)$$

we arrive at the following system of equations:

$$\sum_{i=1}^{t} \phi_d S^*_{qd} = Q^*_q + S^*_{\Gamma_q} \sigma \qquad (q = 1, \ldots, c) \qquad (4.9.7)$$

If we employ the matrix notation of Section 4.5 and define

$$\underset{\sim}{S}^*_{\Gamma} = \begin{bmatrix} S^*_{\Gamma_1} \\ S^*_{\Gamma_2} \\ \vdots \\ S^*_{\Gamma_c} \end{bmatrix}$$

we may write (4.9.7) as:

$$\underset{\sim}{S}^{*}\underset{\sim}{\phi} = \underset{\sim}{Q}^{*} + \underset{\sim}{S}_{\Gamma}^{*}\sigma$$

and the solution is then:

$$\hat{\underset{\sim}{\phi}} = \underset{\sim}{S}^{*-1}\underset{\sim}{Q}^{*} + \underset{\sim}{S}^{*-1}\underset{\sim}{S}_{\Gamma}^{*}\hat{\sigma} \tag{4.9.8}$$

where $\hat{\sigma}$ is to be determined.

We may now proceed to solve (4.9.5). Let us first define the row vector:

$$\underset{\sim}{U}'_{i} = \left[\ (u_{1i} - \bar{u}_{1}^{*})\ \ (u_{2i} - \bar{u}_{2}^{*})\ \cdots\ (u_{ci} - \bar{u}_{c}^{*})\ \right] \qquad (i = 1,\ldots,\ell)$$

using which we may write

$$\sum_{d=1}^{c} \hat{\phi}_{d}\,(u_{di} - \bar{u}_{d}^{*}) = \underset{\sim}{U}'_{i}\,\hat{\underset{\sim}{\phi}} \qquad (i = 1,\ldots,\ell)\ .$$

Now substituting (4.9.6) and (4.9.8) into (4.9.5) and employing the above notation we have, after collecting terms and simplifying:

$$A^{*}\sigma^{2} - B^{*}\sigma - C^{*} = 0$$

which gives the solution:

$$\hat{\sigma} = \frac{B^{*} + \sqrt{B^{*2} + 4A^{*}C^{*}}}{2\sqrt{A^{*}(A^{*} - \frac{k}{2p}(c+1))}}\ , \quad \text{corrected for bias.} \tag{4.9.9}$$

where:

$$B^{*} = \sum_{i=1}^{\ell}\left\{\sum_{j=a_i}^{b_i}\gamma_{ij}\,[\,y_{(ij)} - \bar{\bar{y}}^{*} - \underset{\sim}{U}'_{i}\underset{\sim}{S}^{*-1}\underset{\sim}{Q}^{*}\,]\right.$$

$$\left. + s_{i}\alpha_{2i}^{*}\,[\,y_{(i,b_i)} - \bar{\bar{y}}^{*} - \underset{\sim}{U}'_{i}\underset{\sim}{S}^{*-1}\underset{\sim}{Q}^{*}\,] - r_{i}\alpha_{1i}^{*}\,[\,y_{(i,a_i)} - \bar{\bar{y}}^{*} - \underset{\sim}{U}'_{i}\underset{\sim}{S}^{*-1}\underset{\sim}{Q}^{*}\,]\right\}$$

and

$$C^* = \sum_{i=1}^{t} \left\{ \sum_{j=a_i}^{b_i} \delta_{ij} [y_{(ij)} - \bar{\bar{y}}^* - \underset{\sim}{U}'_i \underset{\sim}{S}^{*^{-1}} \underset{\sim}{Q}^*]^2 \right.$$

$$\left. + r_i \beta^*_{1i} [y_{(i,a_i)} - \bar{\bar{y}}^* - \underset{\sim}{U}'_i \underset{\sim}{S}^{*^{-1}} \underset{\sim}{Q}^*]^2 + s_i \beta^*_{2i} [y_{(i,b_i)} - \bar{\bar{y}}^* - \underset{\sim}{U}'_i \underset{\sim}{S}^{*^{-1}} \underset{\sim}{Q}^*]^2 \right\}$$

Thus equations **(4.9.6)**, **(4.9.8)**, and **(4.9.9)** are the MML estimators for the multiple regression model **(4.9.1)** based on asymmetrically censored samples. Note that these results are a complete generalization of the estimators given in Section 4.5, and clearly reduce to those expressions when the censoring is symmetric.

4.9.1 Multiple Regression Asymptotic Variance-Covariance Matrix for Asymmetric Censoring

Using the same procedure as in Section 4.6.1 gives the following information matrix based on asymmetrically censored samples, where t_i^* and Ω^* have already been defined:

$$\underset{\sim}{I}(\mu, \underset{\sim}{\phi}, \sigma) = \frac{2p}{k\sigma^2} \begin{bmatrix} M^* & M^* \overline{u_1}^* & M^* \overline{u_2}^* & \cdots & M^* \overline{u_c}^* & M^* \overline{\Gamma}^* + 2 \sum_{i=1}^{t} t_i^* \\ & \sum_{i=1}^{t} m_i^* u_{1i}^2 & \sum_{i=1}^{t} m_i^* u_{1i} u_{2i} & \cdots & \sum_{i=1}^{t} m_i^* u_{1i} u_{ci} & \sum_{i=1}^{t} u_{1i} (\Gamma_i^* + t_i^*) \\ & & \sum_{i=1}^{t} m_i^* u_{2i}^2 & \cdots & \sum_{i=1}^{t} m_i^* u_{2i} u_{ci} & \sum_{i=1}^{t} u_{2i} (\Gamma_i^* + t_i^*) \\ & & & \ddots & \vdots & \vdots \\ & & & & \sum_{i=1}^{t} m_i^* u_{ci}^2 & \sum_{i=1}^{t} u_{ci} (\Gamma_i^* + t_i^*) \\ & & & & & \Omega^* \end{bmatrix}$$

and, as usual, $\underset{\sim}{V}(\mu,\underset{\sim}{\phi},\sigma) = \underset{\sim}{I}^{-1}(\mu,\underset{\sim}{\phi},\sigma)$. Notice that the entries in the right-most column (and bottom row), except for Ω^{*}, become zero when the censoring is symmetric, as is expected.

4.9.2 Multiple Regression ANOVA Table for Asymmetric Censoring

Finally, as in section 4.6.2, we may partition the total sum of squares as follows:

$$SST = \sum_{i=1}^{t} \left\{ \sum_{j=a_i}^{b_i} \delta_{ij} [y_{(ij)} - \bar{\bar{y}}^{*}]^2 + r_i \beta^{*}_i \left\{ [y_{(i,a_i)} - \bar{\bar{y}}^{*}]^2 + [y_{(i,b_i)} - \bar{\bar{y}}^{*}]^2 \right\} \right\}$$

$$SS_R = \underset{\sim}{Q}' \underset{\sim}{S}^{*-1} \underset{\sim}{Q}^{*}$$

$$SS_{LOF} = \sum_{i=1}^{t} m_i^{*} [\bar{y}_i^{*} - \bar{\bar{y}}^{*} - \underset{\sim}{U}'_i \underset{\sim}{S}^{*-1} \underset{\sim}{Q}^{*}]^2$$

$$SSE = \sum_{i=1}^{t} \left\{ \sum_{j=a_i}^{b_i} \delta_{ij} [y_{(ij)} - \bar{y}_i^{*}]^2 + r_i \beta^{*}_i \left\{ [y_{(i,a_i)} - \bar{y}_i^{*}]^2 + [y_{(i,b_i)} - \bar{y}_i^{*}]^2 \right\} \right\}$$

from which the ANOVA Table and hypothesis testing proceeds in the usual manner.

For asymmetrically censored samples, the F-distributions do not provide accurate approximations for the F statistics for small samples. This is due to the fact that the numerators and denominators in these F statistics are correlated. For large samples, however, the denominator converges to σ^2 and the distributions of the F statistics are closely approximated by chi-square distributions. We do not, however give details for conciseness.

Chapter 5

5.1 Introduction

Up until now, the ANOVA and regression procedures which have been discussed have focused on univariate models. In all cases, there was only one design variable (y) being modelled. In many instances, however, experimenters measure several other variables of interest besides the main response variable, and often wish to include these in their model. For example, in the fertilizer problem we have used previously, one may wish to measure not only the height of the plants produced, but also the yield of fruit on each plant. This new variable (yield) is termed a *concomitant* variable. One, then, would wish to include this variable in the model to see how the design variable(s) affect it. In the context of regression analysis, this model would be a *Bivariate Regression Model*, and may be analyzed in a fashion similar to the procedure laid out in Section 4.2. Of course, there may be more than one concomitant variable associated with the response variable, in which case a *Multivariate Regression Model* would be appropriate for the analysis of the data obtained. In what follows, we will be discussing the bivariate regression case (one concomitant variable). All results discussed may easily be generalized to the multivariate case, h... for the purposes of brevity the multivariate model will not be discussed here.

As is the case with the univariate regression and ANOVA procedures, multivariate regression techniques based on normality are well known and have been in use for a long time. However, as we have been stressing, it is often the case that the assumption of multivariate (or bivariate) normality may not be justified. It is of great interest, therefore, to relax this assumption in multivariate situations, as in the univariate case. Tiku and Kambo (1992) have done this by applying the Tiku-Suresh method of estimation (which we have been using throughout these chapters) to a symmetric bivariate location-scale family which they introduced. Specifically, to obtain a symmetric non-normal bivariate family, Tiku and Kambo (1992) replace one of the normal marginal distributions with the t-family distribution (2.3.1). This results in the symmetric bivariate distribution:

$$f(x,y) = g(x|y)\,h(y) \tag{5.1.1}$$

where

$$g(x|y) \propto \frac{1}{\sqrt{\sigma_1^2(1-\rho^2)}} \exp\left\{ \frac{-1}{2\,\sigma_2(1-\rho^2)} \left[x - \mu_1 - \rho\left(\frac{\sigma_1}{\sigma_2}\right)(y - \mu_2) \right]^2 \right\}$$

$$\text{and} \quad h(y) \propto \frac{1}{\sqrt{K}\,\sigma_2} \left\{ 1 + \frac{(y - \mu_2)^2}{k\sigma_2^2} \right\}^{-p}$$

(Tiku & Kambo, 1992, p. 1684)

The parameters in the model (μ_1, μ_2, σ_1^2, σ_2^2, and ρ) represent, respectively, the means and variances of X and Y, and the correlation between X and Y, just as in the bivariate normal case.

For this non-normal bivariate location-scale distribution, Tiku and Kambo (1992) have computed the MML estimators of μ_1, μ_2, σ_1^2, σ_2^2, and ρ for complete and Type-II symmetrically censored samples. For a more complete account, see Tiku and Kambo (1992) (and also Tiku & Gill, 1989, and Tiku & Gill, 1990). Extending the Tiku and Kambo (1992) procedure to more complex models is of great value. It is with this in mind that the following chapter will develop this procedure for bivariate regression and multiple regression models.

5.2 Bivariate Regression Based on the Modified Bivariate Distribution

Consider the following bivariate regression model:

$$\begin{cases} x_{ij} = \mu_1 + \psi u_i + \varepsilon_{ij} \\ y_{ij} = \mu_2 + \phi u_i + \xi_{ij} \end{cases} \qquad (i = 1,\dots,\ell;\ j = 1,\dots,n_i)$$

Here, ψ and ϕ are the corresponding regression parameters. In the classical procedure, the error pair (ε,ξ) is usually assumed to be bivariate normally distributed (with some correlation ρ). To extend these results, we will consider (ε,ξ) to have the distribution BF($\mu_1 + \psi u$, $\mu_2 + \phi u$; σ_1^2, σ_2^2; $\rho|p$) (assuming p is known) where 'BF' refers to the distribution in **(5.1.1)**. To employ the Tiku-Suresh estimation procedure, we order the y-observations in each group i with respect to j. We denote the ordered y-observations as $y_{(ij)}$, and denote the x-observation concomitant to $y_{(ij)}$ as $x_{[ij]}$. Note that the x-observations are not necessarily ordered by this procedure. Then the log-likelihood function based on (5.1.1) is:

$$\begin{aligned} \ln L = \text{const.} &- \frac{N}{2} \ln [\sigma_1^2 \sigma_2^2 (1-\rho^2)] \\ &- \frac{1}{2\sigma_1^2(1-\rho^2)} \sum_{i=1}^{t} \sum_{j=1}^{n_i} \left\{ x_{[ij]} - \mu_1 - \psi u_i - \rho \left(\frac{\sigma_1}{\sigma_2}\right) [y_{(ij)} - \mu_2 - \phi u_i] \right\}^2 \\ &- p \sum_{i=1}^{t} \sum_{j=1}^{n_i} \ln \left\{ 1 + \frac{[y_{(ij)} - \mu_2 - \phi u_i]^2}{k\sigma_2^2} \right\} \end{aligned} \qquad (5.2.1)$$

For brevity of notation, let us write:

$$w_{[ij]} = \frac{x_{[ij]} - \mu_1 - \psi u_i}{\sigma_1}$$

$$\text{and} \quad z_{(ij)} = \frac{y_{(ij)} - \mu_2 - \phi u_i}{\sigma_2} .$$

Then (5.2.1) becomes:

$$\begin{aligned} \ln L = \text{const.} &- \frac{N}{2} [\sigma_1^2 \sigma_2^2 (1-\rho^2)] \\ &- \frac{1}{2\sigma_1(1-\rho^2)} \sum_{i=1}^{t} \sum_{j=1}^{n_i} [w_{[ij]} - \rho z_{(ij)}]^2 - p \sum_{i=1}^{t} \sum_{j=1}^{n_i} \ln \left\{ 1 + \frac{z_{(ij)}^2}{k} \right\} \end{aligned} \qquad (5.2.2)$$

Differentiating (5.2.2) with respect to the model parameters, we have:

$$\frac{\partial \ln L}{\partial \mu_1} = \frac{1}{\sigma_1(1-\rho^2)} \sum_{i=1}^{t}\sum_{j=1}^{n_i} [\, w_{[ij]} - \rho\, z_{(ij)}] = 0 \qquad (5.2.3)$$

$$\frac{\partial \ln L}{\partial \psi} = \frac{1}{\sigma_1(1-\rho^2)} \sum_{i=1}^{t} u_i \sum_{j=1}^{n_i} [\, w_{[ij]} - \rho\, z_{(ij)}] = 0 \qquad (5.2.4)$$

$$\frac{\partial \ln L}{\partial \sigma_1} = -\frac{N}{\sigma_1} + \frac{1}{\sigma_1(1-\rho^2)} \sum_{i=1}^{t}\sum_{j=1}^{n_i} z_{(ij)}\, [\, w_{[ij]} - \rho\, z_{(ij)}] = 0 \qquad (5.2.5)$$

$$\frac{\partial \ln L}{\partial \rho} = \frac{N\rho}{1-\rho^2} - \frac{\rho}{(1-\rho^2)^2} \sum_{i=1}^{t}\sum_{j=1}^{n_i} [\, w_{[ij]} - \rho\, z_{(ij)}]^2$$

$$+ \frac{1}{1-\rho^2} \sum_{i=1}^{t}\sum_{j=1}^{n_i} z_{(ij)}\, [\, w_{[ij]} - \rho\, z_{(ij)}] = 0 \qquad (5.2.6)$$

$$\frac{\partial \ln L}{\partial \mu_2} = \frac{-\rho}{1-\rho^2} \sum_{i=1}^{t}\sum_{j=1}^{n_i} [\, w_{[ij]} - \rho\, z_{(ij)}] + \frac{2p}{k\sigma_2} \sum_{i=1}^{t}\sum_{j=1}^{n_i} g(z_{(ij)}) = 0 \qquad (5.2.7)$$

$$\frac{\partial \ln L}{\partial \phi} = \frac{-\rho}{\sigma_2(1-\rho^2)} \sum_{i=1}^{t} u_i \sum_{j=1}^{n_i} [\, w_{[ij]} - \rho\, z_{(ij)}]$$

$$+ \frac{2p}{k\sigma_2} \sum_{i=1}^{t} u_i \sum_{j=1}^{n_i} g(z_{(ij)}) = 0 \qquad (5.2.8)$$

$$\frac{\partial \ln L}{\partial \sigma_2} = -\frac{N}{\sigma_2} - \frac{\rho}{\sigma_2(1-\rho^2)} \sum_{i=1}^{t}\sum_{j=1}^{n_i} z_{(ij)}\, [\, w_{[ij]} - \rho\, z_{(ij)}]$$

$$+ \frac{2p}{k\sigma_2} \sum_{i=1}^{t}\sum_{j=1}^{n_i} z_{(ij)}\, g(z_{(ij)}) = 0 \qquad (5.2.9)$$

where $N = \sum n_i$ and the function g has been defined previously.In order to solve the system of equations (5.2.3)-(5.2.9), we first note that simplifying the first four equations produces the following identities:

$$\sum_{i=1}^{t}\sum_{j=1}^{n_i} [\, w_{[ij]} - \rho\, z_{(ij)}] = 0$$

$$\sum_{i=1}^{t} u_i \sum_{j=1}^{n_i} [\, w_{[ij]} - \rho\, z_{(ij)}] = 0$$

and $$\sum_{i=1}^{t}\sum_{j=1}^{n_i} z_{(ij)}\, [\, w_{[ij]} - \rho\, z_{(ij)}] = 0 \,.$$

Using these results, we may eliminate the first terms from equations (5.2.7)-(5.2.9). Now employing the approximation $g(z_{(ij)}) \approx \gamma_{ij} + \delta_{ij}z_{(ij)}$ as discussed in Chapter 4, we arrive at the following three equations in three unknowns:

$$\frac{\partial \ln L^*}{\partial \mu_2} = \frac{2p}{k\sigma_2}\sum_{i=1}^{\ell}\sum_{j=1}^{n_i}[\gamma_{ij} + \delta_{ij}z_{(ij)}] = 0 \qquad (5.2.10)$$

$$\frac{\partial \ln L^*}{\partial \phi} = \frac{2p}{k\sigma_2}\sum_{i=1}^{\ell}u_i\sum_{j=1}^{n_i}[\gamma_{ij} + \delta_{ij}z_{(ij)}] = 0 \qquad (5.2.11)$$

$$\frac{\partial \ln L^*}{\partial \sigma_2} = -\frac{N}{\sigma_2} + \frac{2p}{k\sigma_2}\sum_{i=1}^{\ell}\sum_{j=1}^{n_i}z_{(ij)}[\gamma_{ij} + \delta_{ij}z_{(ij)}] = 0 \qquad (5.2.12)$$

where we have used the notation L^* to indicate the modified likelihood function. We now note here that these equations are identical to equations (4.3.8)-(4.3.10) for the simple linear regression model, with μ and σ being replaced here by μ_2 and σ_2. Therefore the Tiku-Suresh MML estimators $\hat{\mu}_2$, $\hat{\sigma}_2$, and $\hat{\phi}$ are the same as those given in Section 4.3, namely:

$$\hat{\mu}_2 = \bar{\bar{y}} - \hat{\phi}\,\bar{u}_m \qquad (5.2.13)$$

$$\hat{\phi} = \frac{S^m_{yu}}{S^m_{uu}} \qquad (5.2.14)$$

and $$\hat{\sigma}_2 = \frac{B + \sqrt{B^2 + 4NC}}{2\sqrt{N(N-2)}} \qquad (5.2.15)$$

where B and C are as defined in section 4.3, and $\bar{\bar{y}}$, $\bar{u}_m$, S_{uu}^m and S_{yu}^m are also as defined in Section 4.3 (with the 'm' denoting weighted sums with respect to the m_i's).

Using the results in (5.2.13)-(5.2.15), it is now possible to solve equations (5.2.3)-(5.2.6) for the remaining parameters. Let us first consider (5.2.3). After substituting for $w_{(ij)}$ and $z_{(ij)}$ and simplifying, we have:

$$\sum_{i=1}^{t}\sum_{j=1}^{n_i} x_{[ij]} - N\mu_1 - \psi\sum_{i=1}^{t} n_i u_i$$

$$= \rho\left(\frac{\sigma_1}{\theta_2}\right)\left\{\sum_{i=1}^{t}\sum_{j=1}^{n_i} y_{(ij)} - N\hat{\mu}_2 - \hat{\phi}\sum_{i=1}^{t} n_i u_i\right\} \tag{5.2.16}$$

Now define:

$$\bar{x}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} x_{[ij]} \tag{5.2.17}$$

$$\bar{\bar{x}} = \frac{1}{N}\sum_{i=1}^{t} n_i \bar{x}_i \tag{5.2.18}$$

$$\bar{Y}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} y_{(ij)} \tag{5.2.19}$$

$$\bar{\bar{Y}} = \frac{1}{N}\sum_{i=1}^{t} n_i \bar{Y}_i \tag{5.2.20}$$

$$\bar{u}_n = \frac{1}{N}\sum_{i=1}^{t} n_i u_i \tag{5.2.21}$$

Note here that we have used capital $\bar{Y}_i$ and $\bar{\bar{Y}}$ to denote the unweighted averages of the y-observations (as opposed to $\bar{y}_i$ and $\bar{\bar{y}}$), and we use the notation $\bar{u}_n$ to differentiate this expression from the previous $\bar{u}_m$ (average of the u_i's weighted with the m_i's). Using (5.2.17)-(5.2.21) then, (5.2.16) becomes:

$$N\bar{\bar{x}} - N\mu_1 - \psi N\bar{u}_n = \rho\left(\frac{\sigma_1}{\theta_2}\right)\left\{N\bar{\bar{Y}} - N\hat{\mu}_2 - \hat{\phi}N\bar{u}_n\right\}$$

$$\Rightarrow \quad \hat{\mu}_1 = \bar{\bar{x}} - \hat{\psi}\,\bar{u}_n - \hat{\rho}\left(\frac{\theta_1}{\theta_2}\right)\left\{\bar{\bar{Y}} - \hat{\mu}_2 - \hat{\phi}\,\bar{u}_n\right\} \tag{5.2.22}$$

where $\hat{\psi}$, $\hat{\sigma}_1$, and $\hat{\rho}$ are to be determined.

Considering (5.2.4) we have (after appropriate substitutions and simplification):

$$\sum_{i=1}^{t} n_i u_i \bar{x}_i - N\bar{u}_n \hat{\mu}_1 - \psi \sum_{i=1}^{t} n_i u_i^2$$

$$= \rho\left(\frac{\sigma_1}{\theta_2}\right) \left\{ \sum_{i=1}^{t} n_i u_i \bar{Y}_i - N\bar{u}_n \hat{\mu}_2 - \hat{\phi} \sum_{i=1}^{t} n_i u_i^2 \right\} .$$

Now substituting for $\hat{\mu}_1$ and simplifying further, we obtain:

$$\sum_{i=1}^{t} n_i (u_i - \bar{u}_n) x_i - \psi \sum_{i=1}^{t} n_i (u_i - \bar{u}_n)^2$$

$$= \rho\left(\frac{\sigma_1}{\theta_2}\right) \left\{ \sum_{i=1}^{t} n_i (u_i - \bar{u}_n) \bar{Y}_i - \hat{\phi} \sum_{i=1}^{t} n_i (u_i - \bar{u}_n)^2 \right\} \qquad (5.2.23)$$

Let us also define:

$$S_{xu} = \sum_{i=1}^{t} n_i (u_i - \bar{u}_n) x_i$$

$$S_{yu}^{n} = \sum_{i=1}^{t} n_i (u_i - \bar{u}_n) \bar{Y}_i$$

$$S_{uu}^{n} = \sum_{i=1}^{t} n_i (u_i - \bar{u}_n)^2$$

where the superscripted 'n' distinguishes these from S_{yu}^{m} and S_{uu}^{m} defined previously. Then (5.2.23) becomes:

$$S_{xu} - \psi S_{uu}^{n} = \rho\left(\frac{\sigma_1}{\theta_2}\right) \left\{ S_{yu}^{n} - \hat{\phi} S_{uu}^{n} \right\}$$

Hence:

$$\hat{\psi} = \frac{S_{xu}}{S_{uu}^{n}} - \hat{\rho}\left(\frac{\theta_1}{\theta_2}\right) \left\{ \frac{S_{yu}^{n}}{S_{uu}^{n}} - \hat{\phi} \right\}$$

$$= \frac{S_{xu}}{S_{uu}^{n}} - \hat{\rho}\left(\frac{\theta_1}{\theta_2}\right) \left\{ \frac{S_{yu}^{n}}{S_{uu}^{n}} - \frac{S_{yu}^{m}}{S_{uu}^{m}} \right\} \qquad (5.2.24)$$

where $\hat{\sigma}_1$ and $\hat{\rho}$ are to be determined.

Now (5.2.6) becomes, after substituting in (5.2.5):

$$\sum_{i=1}^{t}\sum_{j=1}^{n_i} w_{[ij]}\, z_{(ij)} - \rho \sum_{i=1}^{t}\sum_{j=1}^{n_i} z^2_{(ij)} = 0$$

or, after substituting for $w_{[ij]}$ and $z_{(ij)}$,

$$\hat{\rho} = \frac{\sum_{i=1}^{t}\sum_{j=1}^{n_i} w_{[ij]}\, z_{(ij)}}{\sum_{i=1}^{t}\sum_{j=1}^{n_i} z^2_{(ij)}}$$

$$= \frac{\hat{\sigma}_2}{\sigma_1} \frac{\sum_{i=1}^{t}\sum_{j=1}^{n_i} [x_{[ij]} - (\hat{\mu}_1 + \hat{\psi} u_i)]\,[y_{(ij)} - (\hat{\mu}_2 + \hat{\phi} u_i)]}{\sum_{i=1}^{t}\sum_{j=1}^{n_i} [y_{(ij)} - (\hat{\mu}_2 + \hat{\phi} u_i)]^2} \qquad (5.2.25)$$

The above expression is not the final estimator of ρ, since $\hat{\mu}_1$ and $\hat{\psi}$ are both functions of ρ. To simplify this expression further, let us write:

$$\mu_{1i} = \mu_1 + \psi u_i$$

$$\text{and} \quad \mu_{2i} = \mu_2 + \phi u_i \ .$$

Therefore (5.2.25) may be written as:

$$\hat{\rho} = \frac{\hat{\sigma}_2}{\sigma_1} \frac{\sum_{i=1}^{t}\sum_{j=1}^{n_i} [x_{[ij]} - \hat{\mu}_{1i}]\,[y_{(ij)} - \hat{\mu}_{2i}]}{\sum_{i=1}^{t}\sum_{j=1}^{n_i} [y_{(ij)} - \hat{\mu}_{2i}]^2} \qquad (5.2.26)$$

Now,

$$\hat{\mu}_{1i} = \hat{\mu}_1 + \hat{\psi}\, u_i$$

$$\Rightarrow \quad \sum_{i=1}^{\ell} n_i \hat{\mu}_{1i} = \sum_{i=1}^{\ell} n_i \,[\, \hat{\mu}_1 + \hat{\psi}\, u_i \,]$$

$$= N\hat{\mu}_1 + \hat{\psi} N \bar{u}_n$$

$$= N\bar{\bar{x}} - \hat{\psi} N\bar{u}_n - \hat{\rho}\left(\frac{\sigma_1}{\partial_2}\right)[\, N\bar{\bar{Y}} - N\hat{\mu}_2 - \hat{\phi} N \bar{u}_n \,] + \hat{\psi} N \bar{u}_n$$

$$= N\bar{\bar{x}} - \hat{\rho}\left(\frac{\sigma_1}{\partial_2}\right)[\, N\bar{\bar{Y}} - N\hat{\mu}_2 - \hat{\phi} N \bar{u}_n \,]$$

$$= \sum_{i=1}^{\ell} n_i \left\{ \bar{x}_i - \hat{\rho}\left(\frac{\sigma_1}{\partial_2}\right)[\, \bar{Y}_i - \hat{\mu}_2 - \hat{\phi} u_i \,] \right\}$$

Thus we have

$$\hat{\mu}_{1i} = \bar{x}_i - \hat{\rho}\left(\frac{\sigma_1}{\partial_2}\right)[\, \bar{Y}_i - \hat{\mu}_2 - \hat{\phi}\, u_i \,]$$

$$= \bar{x}_i - \hat{\rho}\left(\frac{\sigma_1}{\partial_2}\right)[\, \bar{Y}_i - \hat{\mu}_{2i} \,]$$

which we may substitute into (5.2.26), obtaining:

$$\hat{\rho} = \frac{\hat{\sigma}_2}{\sigma_1} \; \frac{\displaystyle\sum_{i=1}^{\ell}\sum_{j=1}^{n_i} \left\{ x_{[ij]} - \bar{x}_i - \hat{\rho}\left(\frac{\sigma_1}{\partial_2}\right)[\, \bar{Y}_i - \hat{\mu}_{2i} \,] \right\} [\, y_{(ij)} - \hat{\mu}_{2i} \,]}{\displaystyle\sum_{i=1}^{\ell}\sum_{j=1}^{n_i} [\, y_{(ij)} - \hat{\mu}_{2i} \,]^2}$$

$$= \frac{\hat{\sigma}_2}{\sigma_1} \; \frac{\displaystyle\sum_{i=1}^{\ell}\sum_{j=1}^{n_i} [\, x_{[ij]} - \bar{x}_i \,]\,[\, y_{(ij)} - \hat{\mu}_{2i} \,]}{\displaystyle\sum_{i=1}^{\ell}\sum_{j=1}^{n_i} [\, y_{(ij)} - \hat{\mu}_{2i} \,]^2} + \hat{\rho} \; \frac{\displaystyle\sum_{i=1}^{\ell}\sum_{j=1}^{n_i} [\, \bar{Y}_i - \hat{\mu}_{2i} \,]\,[\, y_{(ij)} - \hat{\mu}_{2i} \,]}{\displaystyle\sum_{i=1}^{\ell}\sum_{j=1}^{n_i} [\, y_{(ij)} - \hat{\mu}_{2i} \,]^2} \, .$$

Rearranging this expression for $\hat{\rho}$ and simplifying gives:

$$\hat{\rho} = \frac{\hat{\sigma}_2}{\sigma_1} \; \frac{\displaystyle\sum_{i=1}^{\ell}\sum_{j=1}^{n_i} [\, x_{[ij]} - \bar{x}_i \,]\,[\, y_{(ij)} - \hat{\mu}_{2i} \,]}{\displaystyle\sum_{i=1}^{\ell} \left\{ \sum_{j=1}^{n_i} [\, y_{(ij)} - \hat{\mu}_{2i} \,]^2 - n_i\,[\, \bar{Y}_i - \hat{\mu}_{2i} \,]^2 \right\}} \, .$$

This may be simplified further by noting that

$$\sum_{j=1}^{n_i} [x_{[ij]} - \bar{x}_i][y_{(ij)} - \hat{\mu}_{2i}]$$

$$= \sum_{j=1}^{n_i} [x_{[ij]} - \bar{x}_i][y_{(ij)} - \bar{Y}_i] + [\bar{Y}_i - \hat{\mu}_{2i}] \sum_{j=1}^{n_i} [x_{[ij]} - \bar{x}_i]$$

$$= \sum_{j=1}^{n_i} [x_{[ij]} - \bar{x}_i][y_{(ij)} - \bar{Y}_i]$$

and

$$\sum_{j=1}^{n_i} [y_{(ij)} - \hat{\mu}_{2i}]^2 - n_i [\bar{Y}_i - \hat{\mu}_{2i}]^2$$

$$= \sum_{j=1}^{n_i} [y_{(ij)} - \bar{Y}_i]^2 + 2[\bar{Y}_i - \hat{\mu}_{2i}] \sum_{j=1}^{n_i} [y_{(ij)} - \bar{Y}_i] + n_i [\bar{Y}_i - \hat{\mu}_{2i}]^2 - n_i [\bar{Y}_i - \hat{\mu}_{2i}]^2$$

$$= \sum_{j=1}^{n_i} [y_{(ij)} - \bar{Y}_i]^2 .$$

Therefore we may write $\hat{\rho}$ as:

$$\hat{\rho} = \frac{\hat{\sigma}_2}{\sigma_1} \frac{\sum_{i=1}^{t}\sum_{j=1}^{n_i} [x_{[ij]} - \bar{x}_i][y_{(ij)} - \bar{Y}_i]}{\sum_{i=1}^{t}\sum_{j=1}^{n_i} [y_{(ij)} - \bar{Y}_i]^2} .$$

If we introduce some familiar notation (where p denotes 'pooling'):

$$S^2_{1(p)} = \frac{1}{N}\sum_{i=1}^{t}\sum_{j=1}^{n_i} [x_{[ij]} - \bar{x}_i]^2$$

$$S^2_{2(p)} = \frac{1}{N}\sum_{i=1}^{t}\sum_{j=1}^{n_i} [y_{(ij)} - \bar{Y}_i]^2$$

and $$S_{12(p)} = \frac{1}{N}\sum_{i=1}^{t}\sum_{j=1}^{n_i} [x_{[ij]} - \bar{x}_i][y_{(ij)} - \bar{Y}_i]$$

then we may write

$$\hat{\rho} = \frac{\hat{\sigma}_2}{\hat{\sigma}_1} \frac{S_{12(p)}}{S^2_{2(p)}} \tag{5.2.27}$$

where $\hat{\sigma}_1$ is still to be determined. Note that, as mentioned in Tiku and Kambo (1992), under bivariate normality $\hat{\rho}$ reduces to S_{12}/S_1S_2.

Finally, to solve for σ_1, we consider (5.2.5). After substituting for $w_{[ij]}$ and $z_{(ij)}$ and simplifying, we obtain:

$$N\sigma_1^2 (1 - \hat{\rho}^2) - \sum_{i=1}^{l} \sum_{j=1}^{n_l} [x_{[ij]} - \bar{x}_l]^2 + \hat{\rho} \left(\frac{\sigma_1}{\partial_2} \right) \sum_{i=1}^{l} \sum_{j=1}^{n_l} [x_{[ij]} - \bar{x}_l] [y_{(ij)} - \hat{\mu}_{2l}] = 0$$

Now using the notation introduced for the computation of $\hat{\rho}$:

$$N\sigma_1^2 (1 - \hat{\rho}^2) - NS^2_{12(p)} + N\hat{\rho} \left(\frac{\sigma_1}{\partial_2} \right) S_{12(p)} = 0 \tag{5.2.28}$$

Now we note that:

$$N\sigma_1^2 (1 - \hat{\rho}^2) = N\sigma_1^2 \left\{ 1 - \frac{\partial_2^2}{\sigma_1^2} \frac{S^2_{12(p)}}{[S^2_{2(p)}]^2} \right\}$$

$$= N\sigma_1^2 - N\hat{\sigma}_2^2 \frac{S^2_{12(p)}}{[S^2_{2(p)}]^2}$$

$$\text{and} \quad \hat{\rho} \left(\frac{\sigma_1}{\partial_2} \right) S_{12(p)} = \frac{S^2_{12(p)}}{S^2_{2(p)}} .$$

Then (5.2.28) becomes:

$$N\sigma_1^2 - N\hat{\sigma}_2^2 \frac{S_{12(p)}}{[S^2_{2(p)}]^2} - NS^2_{1(p)} + N \frac{S_{12(p)}}{S^2_{2(p)}} = 0$$

$$\Rightarrow \quad \sigma_1^2 = S^2_{1(p)} + \hat{\sigma}_2^2 \frac{S_{12(p)}}{[S^2_{2(p)}]^2} - \frac{S^2_{12(p)}}{S^2_{2(p)}}$$

$$= S^2_{12(p)} + \frac{S^2_{12(p)}}{S^2_{2(p)}} \left[\frac{\partial_2^2}{S^2_{2(p)}} - 1 \right]$$

Hence

$$\hat{\sigma}_1 = \sqrt{S^2_{1(p)} + \frac{S^2_{12(p)}}{S^2_{2(p)}}\left[\frac{\theta^2_2}{S^2_{2(p)}} - 1\right]} \qquad (5.2.29)$$

where $\hat{\sigma}_2$ has been given previously.

For reasons given in Tiku and Kambo (1992), the estimators $\hat{\sigma}_1$ and $\hat{\sigma}_2$ are always real and positive, and $\hat{\rho}$ is always between -1 and 1. Also, it may be noted that all estimators produced in this section reduce to the estimators based on bivariate normality when $p = \infty$, which is to be expected.

5.3 Variance-Covariance Matrix for Modified Bivariate Regression Based on Complete Samples

To obtain the asymptotic variance-covariance matrix for the estimators derived in the previous section, it is necessary to obtain all the mixed second-order partial derivatives of the log-likelihood function (as in Section 4.2.1), take their (negative) expectation to form the information matrix, then invert to produce the variance-covariance matrix. We will not report all derivatives here (since there are so many). We will, rather, report all non-zero expectations, which will show that the information matrix may be split into two sub-matrices, one containing information on μ_1, μ_2, ϕ, and ψ, and the other containing information on σ_1, σ_2, and ρ.

Before performing the expectations, we need some notation and results. Let us define

$$v_{[ij]} = E(w_{[ij]})$$
$$\text{and} \quad t_{(ij)} = E(z_{(ij)}) \ .$$

Now since the t-family (2.3.1) is symmetric and the distribution of w is (conditionally) symmetric, we immediately have:

$$\sum_{j=1}^{n_i} v_{[ij]} = 0 \ , \quad \sum_{j=1}^{n_i} t_{(ij)} = 0 \ , \quad \text{and} \quad \sum_{j=1}^{n_i} \delta_{ij} t_{(ij)} = 0 \qquad (5.3.1)$$

for all $i = 1,\ldots,\ell$. Using (5.3.1) it can be shown that many of the second order derivatives have zero expectation. To compute the remaining expectations, we need to find the following:

$$\sum_{i=1}^{\ell}\sum_{j=1}^{n_i} E(w_{[ij]}^2) \qquad (5.3.2)$$

$$\text{and} \quad \sum_{i=1}^{\ell}\sum_{j=1}^{n_i} E(w_{[ij]} z_{(ij)}) \ . \qquad (5.3.3)$$

Let us first define

$$Q = \sum_{i=1}^{\ell}\sum_{j=1}^{n_i} E(z_{(ij)}^2)$$

Q may be found using the means and variances of order statistics from (2.3.1), i.e.:

$$Q = \sum_{i=1}^{\ell}\sum_{j=1}^{n_i} \{ t_{(ij)}^2 + \sigma_{jj:n_i} \} \ .$$

To find (5.3.2) and (5.3.3), then, we note that from (5.2.6):

$$\sum_{i=1}^{\ell}\sum_{j=1}^{n_i} w_{[ij]} z_{(ij)} = \rho \sum_{i=1}^{\ell}\sum_{j=1}^{n_i} z_{(ij)}^2$$

$$\text{i.e.} \quad \sum_{i=1}^{\ell}\sum_{j=1}^{n_i} E(w_{[ij]} z_{(ij)}) = \rho \sum_{i=1}^{\ell}\sum_{j=1}^{n_i} E(z_{(ij)}^2) = \rho Q \qquad (5.3.4)$$

and from (5.2.5)

$$\sum_{i}^{\ell}\sum_{j=1}^{n_i} E(w_{[ij]}^2) = \rho\sum_{i=1}^{\ell}\sum_{j=1}^{n_i} E(w_{[ij]}z_{(ij)}) + N(1-\rho^2)$$
$$= \rho^2 Q + N(1-\rho^2) \tag{5.3.5}$$

Now using (5.3.4) and (5.3.5), we may compute the remaining (negative) expectations:

$$-E\left(\frac{\partial^2 \ln L}{\partial\mu_1^2}\right) = \frac{N}{\sigma_1^2(1-\rho^2)} \; ; \qquad -E\left(\frac{\partial^2 \ln L}{\partial\mu_1\partial\psi}\right) = \frac{N\bar{u}_n}{\sigma_1^2(1-\rho^2)}$$

$$-E\left(\frac{\partial^2 \ln L}{\partial\mu_1\partial\mu_2}\right) = \frac{-N\rho}{\sigma_1\sigma_2(1-\rho^2)} \; ; \qquad -E\left(\frac{\partial^2 \ln L}{\partial\mu_1\partial\phi}\right) = \frac{-N\rho\,\bar{u}_n}{\sigma_1\sigma_2(1-\rho^2)}$$

$$-E\left(\frac{\partial^2 \ln L}{\partial\psi^2}\right) = \frac{\sum_{i=1}^{\ell} n_i u_i^2}{\sigma_1^2(1-\rho^2)} \; ; \qquad -E\left(\frac{\partial^2 \ln L}{\partial\psi\partial\mu_2}\right) = \frac{N\bar{u}_n}{\sigma_1\sigma_2(1-\rho^2)}$$

$$-E\left(\frac{\partial^2 \ln L}{\partial\psi\partial\phi}\right) = \frac{\sum_{i=1}^{\ell} n_i u_i^2}{\sigma_1\sigma_2(1-\rho^2)} \; ; \qquad -E\left(\frac{\partial^2 \ln L}{\partial\mu_2^2}\right) = \frac{N\rho^2}{\sigma_2^2(1-\rho^2)} + \frac{2pM}{k\sigma_2^2}$$

$$-E\left(\frac{\partial^2 \ln L}{\partial\phi^2}\right) = \frac{\rho^2}{\sigma_2(1-\rho^2)}\sum_{i=1}^{\ell} n_i u_i^2 + \frac{2p}{k\sigma_2^2}\sum_{i=1}^{n_i} m_i u_i^2$$

$$-E\left(\frac{\partial^2 \ln L}{\partial\mu_2\partial\phi}\right) = \frac{N\rho^2\bar{u}_n}{\sigma_2^2(1-\rho^2)} + \frac{2pM\bar{u}_m}{k\sigma_2^2}$$

and

$$-E\left(\frac{\partial^2 \ln L}{\partial\sigma_1^2}\right) = -\frac{4N}{\sigma_1^2} - \frac{\rho^2 Q}{\sigma_1^2(1-\rho^2)} \; ; \qquad -E\left(\frac{\partial^2 \ln L}{\partial\sigma_1\partial\sigma_2}\right) = \frac{-\rho^2 Q}{\sigma_1\sigma_2(1-\rho^2)}$$

$$-E\left(\frac{\partial^2 \ln L}{\partial\sigma_1\partial\rho}\right) = \frac{\rho}{\sigma_1(1-\rho^2)}[-2N + Q] \; ; \qquad -E\left(\frac{\partial^2 \ln L}{\partial\rho^2}\right) = \frac{1}{1-\rho^2}\left[\frac{2N\rho}{1-\rho^2} + Q\right]$$

$$-E\left(\frac{\partial^2 \ln L}{\partial\sigma_2\partial\rho}\right) = \frac{-\rho Q}{\sigma_2(1-\rho^2)}$$

$$-E\left(\frac{\partial^2 \ln L}{\partial\sigma_2^2}\right) = -\frac{N}{\sigma_2^2} - \frac{\rho Q}{\sigma_2^2(1-\rho^2)} + \frac{2p}{k\sigma_2^2}\sum_{i=1}^{\ell}\sum_{j=1}^{n_i}[\gamma_{ij}t_{(ij)} - 2\delta_{ij}E(z_{(ij)}^2)]$$

Thus we may write the information matrix in the form

$$\underset{\sim}{I} = \begin{bmatrix} I_1 & \underset{\sim}{0} \\ \underset{\sim}{0} & I_2 \end{bmatrix}$$

where I_1 involves μ_1, μ_2, ψ, and ϕ, and I_2 involves σ_1, σ_2, and ρ. Hence we have:

$$I_1 = \begin{bmatrix} \frac{N}{\sigma_1^2(1-\rho^2)} & \frac{-N\rho}{\sigma_1\sigma_2(1-\rho^2)} & \frac{N\overline{u}_n}{\sigma_1^2(1-\rho^2)} & \frac{-N\rho\,\overline{u}_n}{\sigma_1\sigma_2(1-\rho^2)} \\ & \frac{N\rho^2}{\sigma_2^2(1-\rho^2)} + \frac{2pM}{k\sigma_2^2} & \frac{N\overline{u}_n}{\sigma_1\sigma_2(1-\rho^2)} & \frac{N\rho^2\,\overline{u}_n}{\sigma_2^2(1-\rho^2)} + \frac{2pM\overline{u}_m}{k\sigma_2^2} \\ // & & \frac{\sum_{i=1}^{t} n_i u_i^2}{\sigma_2^2(1-\rho^2)} & \frac{\sum_{i=1}^{t} n_i u_i^2}{\sigma_1\sigma_2(1-\rho^2)} \\ & & & \frac{\rho^2 \sum_{i=1}^{t} n_i u_i^2}{\sigma_2^2(1-\rho^2)} + \frac{2p}{k\sigma_2^2}\sum_{i=1}^{t} m_i u_i^2 \end{bmatrix}$$

and

$$I_2 = \begin{bmatrix} -\frac{4N}{\sigma_1^2} + \frac{\rho^2 Q}{\sigma_1^2(1-\rho^2)} & \frac{-\rho^2 Q}{\sigma_1\sigma_2(1-\rho^2)} & \frac{\rho}{\sigma_1(1-\rho^2)}[-2N + Q] \\ & -\frac{N}{\sigma_2^2} - \frac{\rho Q}{\sigma_2^2(1-\rho^2)} + \Omega & \frac{-\rho Q}{\sigma_2(1-\rho^2)} \\ // & & \frac{1}{1-\rho^2}\left[\frac{2N\rho^2}{1-\rho^2} + Q\right] \end{bmatrix}$$

where $\Omega = \frac{2p}{k\sigma_2^2} \sum_{i=1}^{t} \sum_{j=1}^{n_i} [\gamma_{ij} t_{(ij)} - 2\,\delta_{ij} E(z_{(ij)})]$.

Thus the variance-covariance matrix $\underset{\sim}{V}$ is then:

$$\underset{\sim}{V} = \underset{\sim}{I}^{-1} = \begin{bmatrix} I_1^{-1} & \underset{\sim}{0} \\ \underset{\sim}{0} & I_2^{-1} \end{bmatrix} = \begin{bmatrix} V_1 & \underset{\sim}{0} \\ \underset{\sim}{0} & V_2 \end{bmatrix}.$$

5.4 The Bivariate Multiple Regression Model

As is the case with the simple linear regression model, the bivariate model discussed in the preceding sections of this chapter may be easily extended to the multiple regression setting. Let us consider an extension of the bivariate model given in Section 5.2:

$$\begin{cases} x_{ij} = \mu_1 + \psi_1 u_{1i} + \cdots + \psi_c u_{ci} + \varepsilon_{ij} \\ y_{ij} = \mu_2 + \phi_1 u_{1i} + \cdots + \phi_c u_{ci} + \xi_{ij} \end{cases} \qquad (i = 1,\ldots,\ell;\ j = 1,\ldots,n_i)$$

Here now, ψ_d and ϕ_d ($d = 1,\ldots,c$) represent the fact that both the dependent variable (y) and the concomitant variable (x) are regressed on more than one independent variable. If we define (after ordering the y_{ij}'s with respect to j):

$$w_{[ij]} = \frac{x_{[ij]} - \mu_1 - \sum_{d=1}^{c} \psi_d u_{di}}{\sigma_1}$$

$$z_{(ij)} = \frac{y_{(ij)} - \mu_2 - \sum_{d=1}^{c} \phi_d u_{di}}{\sigma_2}$$

then the log-likelihood function for this model is identical to (5.2.2). Differentiating with respect to all parameters produces a system of equations similar to (5.2.3)-(5.2.9), the only differences being the derivatives with respect to the regression parameters, i.e.:

$$\frac{\partial \ln L}{\partial \psi_q} = \frac{1}{\sigma_1(1-\rho^2)} \sum_{i=1}^{t} u_{qi} \sum_{j=1}^{n_i} [\, w_{[ij]} - \rho\, z_{(ij)}] = 0 \tag{5.4.1}$$

$$\frac{\partial \ln L}{\partial \phi_q} = \frac{-\rho}{\sigma_2(1-\rho^2)} \sum_{i=1}^{t} u_{qi} \sum_{j=1}^{n_i} [\, w_{[ij]} - \rho\, z_{(ij)}] + \frac{2p}{k\sigma_2} \sum_{i=1}^{t} u_{qi} \sum_{j=1}^{n_i} g(z_{(ij)}) = 0 \tag{5.4.2}$$

for $q = 1,\dots,c$. Using (5.3.1) to eliminate the first term in (5.4.2), and employing the usual approximation $g(z_{(ij)}) \approx \gamma_{ij} + \delta_{ij} z_{(ij)}$, we have:

$$\frac{\partial \ln L^{*}}{\partial \phi_q} = \frac{2p}{k\sigma_2} \sum_{i=1}^{t} u_{qi} \sum_{j=1}^{n_i} [\, \gamma_{ij} + \delta_{ij} z_{(ij)}] = 0 \tag{5.4.3}$$

where we have used L^{*} to denote the modified log-likelihood function. It is now readily apparent that equations (5.2.7), (5.2.9), and (5.4.3) are identical to those for the univariate multiple regression model given in Section 4.5 (the first terms in equations (5.2.7) and (5.2.9) have been eliminated as discussed in Section 5.2), where μ and σ are replaced by μ_2 and σ_2. The solution for these parameters is, as in Section 4.5, given by:

$$\hat{\mu}_2 = \bar{\bar{y}} - \sum_{d=1}^{c} \hat{\phi}_d \bar{u}_{d(m)} \tag{5.4.3}$$

$$\hat{\underline{\phi}} = \underline{S}_m^{-1} \underline{Q}_m \tag{5.4.4}$$

$$\text{and} \quad \hat{\sigma}_2 = \frac{B + \sqrt{B^2 + 4NC}}{2\sqrt{N(N-(c+1))}} \tag{5.4.5}$$

where B and C are defined in Section 4.5,

$$\bar{u}_{d(m)} = \frac{1}{M} \sum_{i=1}^{t} m_i u_{di} \qquad (d = 1,\dots,c)$$

and the elements of $\underset{\sim}{S}_m$ and $\underset{\sim}{Q}_m$ are given by:

$$S_{qd(m)} = \sum_{i=1}^{t} m_i \,(u_{qi} - \bar{u}_{q(m)}) \, u_{di} \qquad (q, d = 1, \ldots, c)$$

$$\text{and} \quad Q_{q(m)} = \sum_{i=1}^{t} m_i \,(u_{qi} - \bar{u}_{q(m)}) \, \bar{y}_i \qquad (q = 1, \ldots, c) \;.$$

The notation 'm' above is used to differentiate these expression from similar notation being introduced shortly.

Now that we have $\hat{\mu}_2$, $\hat{\phi}_d$ (d = 1,…,c) and $\hat{\sigma}_2$ computed, to find the MML estimators for the remaining parameters, we follow the same procedure which has been presented in detail in Section 5.2 to arrive at:

$$\hat{\mu}_1 = \bar{\bar{x}} - \sum_{d=1}^{c} \hat{\psi}_d \bar{u}_d(n) - \hat{\rho}\left(\frac{\partial_1}{\partial_2}\right)\left[\bar{\bar{Y}} - \hat{\mu}_2 - \sum_{d=1}^{c} \hat{\Phi}_d \bar{u}_{d(m)}\right] \qquad (5.4.6)$$

$$\hat{\underset{\sim}{\psi}} = \underset{\sim}{S}_n^{-1} \underset{\sim}{Q}_x - \hat{\rho}\left(\frac{\partial_1}{\partial_2}\right)\left[\underset{\sim}{S}_n^{-1} \underset{\sim}{Q}_n - \underset{\sim}{S}_m^{-1} \underset{\sim}{Q}_m\right] \qquad (5.4.7)$$

$$\hat{\rho} = \frac{\hat{\sigma}_2}{\hat{\sigma}_1} \, \frac{S_{12(p)}}{S^2_{2(p)}} \qquad (5.4.8)$$

$$\hat{\sigma}_1 = \sqrt{S^2_{1(p)} + \frac{S_{12(p)}}{S^2_{2(p)}}\left[\frac{\partial_2^2}{S^2_{2(p)}} - 1\right]} \qquad (5.4.9)$$

where we have defined the elements of the matrices $\underset{\sim}{Q}_x$, $\underset{\sim}{Q}_n$, and $\underset{\sim}{S}_n$ as:

$$S_{qd(n)} = \sum_{i=1}^{t} n_i \,(u_{qi} - \bar{u}_{q(n)}) \, u_{di}$$

$$Q_{q(n)} = \sum_{i=1}^{t} n_i \,(u_{qi} - \bar{u}_{q(n)}) \, \bar{Y}_i$$

$$Q_{q(x)} = \sum_{i=1}^{t} n_i \,(u_{qi} - \bar{u}_{q(n)}) \, \bar{x}_i$$

$$\text{with} \quad \bar{u}_{q(n)} = \frac{1}{N} \sum_{i=1}^{t} n_i \, u_{qi}$$

for $q,d = 1,\ldots,c$. Thus it is easily seen that (5.4.6)-(5.4.9) are generalizations of the results in Section 5.2, and reduce to those results when we set $c = 1$.

5.4.1 Variance-Covariance Matrix for Bivariate Multiple Regression

Computation of the expected values of the mixed second-order partial derivatives of the log-likelihood function proceeds in exactly the same manner as in Section 5.3, with modifications only due to the fact that there are more than one ϕ and ψ. Specifically, the additions are:

$$-E\left(\frac{\partial^2 \ln L}{\partial \psi_q^2}\right) = \frac{1}{\sigma_1^2(1-\rho^2)}\sum_{i=1}^{\ell} n_i u_{qi}^2 \; ; \qquad -E\left(\frac{\partial^2 \ln L}{\partial \psi_{q_1}\partial \psi_{q_2}}\right) = \frac{1}{\sigma_1^2(1-\rho^2)}\sum_{i=1}^{\ell} n_i u_{q_1 i} u_{q_2 i}$$

$$-E\left(\frac{\partial^2 \ln L}{\partial \psi_q \partial \phi_q}\right) = \frac{1}{\sigma_1\sigma_2(1-\rho^2)}\sum_{i=1}^{\ell} n_i u_{qi}^2 \; ; \qquad -E\left(\frac{\partial^2 \ln L}{\partial \psi_{q_1}\partial \phi_{q_2}}\right) = \frac{1}{\sigma_1\sigma_2(1-\rho^2)}\sum_{i=1}^{\ell} n_i u_{q_1 i} u_{q_2 i}$$

$$-E\left(\frac{\partial^2 \ln L}{\partial \phi_q^2}\right) = \frac{\rho^2}{\sigma_2^2(1-\rho^2)}\sum_{i=1}^{\ell} n_i u_{qi}^2 + \frac{2p}{k\sigma_2^2}\sum_{i=1}^{\ell} m_i u_{qi}^2$$

$$-E\left(\frac{\partial^2 \ln L}{\partial \phi_{q_1}\partial \phi_{q_2}}\right) = \frac{\rho^2}{\sigma_2^2(1-\rho^2)}\sum_{i=1}^{\ell} n_i u_{q_1 i} u_{q_2 i} + \frac{2p}{k\sigma_2^2}\sum_{i=1}^{\ell} m_i u_{q_1 i} u_{q_2 i}$$

for q, q_1, $q_2 = 1,\ldots,c$ (where $q_1 \neq q_2$). Also, the expectations involving ϕ_d (or ψ_d) and another parameter are expanded to include $d = 1,\ldots,c$ as well. In terms of the information matrix, the sub-matrix I_2 is identical to that given in Section 5.3. Based on what was just discussed above, I_1 may be generalized in a straightforward manner, and will not be presented here.

5.5 The Bivariate Regression Model Based on Type-II Symmetrically Censored Samples

As has been done with the previous models we have discussed, it is of interest to extend the bivariate regression procedure presented in Section 5.2 to include the case of censored samples. Specifically, we consider the case where the r_i smallest and largest y-observations at each level of u (and the corresponding concomitant x-observations) are unavailable. We would like to repeat the procedure of Sections 5.2-5.3 including this situation.

The log-likelihood function based on the above censored bivariate sample is then:

$$\ln L = \text{const.} - \frac{A}{2}\ln\left[\sigma_1\sigma_2(1-\rho^2)\right] - \frac{1}{2\,(1-\rho^2)}\sum_{i=1}^{t}\sum_{j=a_i}^{b_i}\{w_{[ij]} - \rho\, z_{(ij)}\}^2$$
$$- p\sum_{i=1}^{t}\sum_{j=a_i}^{b_i}\ln\left\{1 + \frac{z^2_{(ij)}}{k}\right\} + \sum_{i=1}^{t} r_i\{\ln F(z_{(i,a_i)}) + \ln[1 - F(z_{(i,b_i)})]\} \quad (5.5.1)$$

where $w_{[ij]}$ and $z_{(ij)}$ are defined in Section 5.2, $a_i = r_i+1$, $b_i = n_i-r_i$, $A = \sum(n_i-2r_i)$, and F is the cdf of the t-family (2.3.1). Now differentiating (5.5.1) with respect to the model parameters and employing the usual approximations

$$g(z_{(ij)}) = \gamma_{ij} + \delta_{ij} z_{(ij)}$$
$$h_1(z_{(ij)}) = \alpha_i - \beta_i z_{(ij)}$$
$$h_2(z_{(ij)}) = \alpha_i + \beta_i z_{(ij)}$$

where $h_1(z) = f(z)/F(z)$ and $h_2(z) = f(z)/[1-F(z)]$, we have the following system of equations:

$$\frac{\partial \ln L}{\partial \mu_1} = \frac{1}{\sigma_1(1-\rho^2)} \sum_{i=1}^{\ell} \sum_{j=a_i}^{b_i} [\, w_{[ij]} - \rho\, z_{(ij)}] = 0 \tag{5.5.2}$$

$$\frac{\partial \ln L}{\partial \psi} = \frac{1}{\sigma_1(1-\rho^2)} \sum_{i=1}^{\ell} u_i \sum_{j=a_i}^{b_i} [\, w_{[ij]} - \rho\, z_{(ij)}] = 0 \tag{5.5.3}$$

$$\frac{\partial \ln L}{\partial \sigma_1} = -\frac{A}{\sigma_1} + \frac{1}{\sigma_1(1-\rho^2)} \sum_{i=1}^{\ell} \sum_{j=a_i}^{b_i} z_{(ij)} [\, w_{[ij]} - \rho\, z_{(ij)}] = 0 \tag{5.5.4}$$

$$\frac{\partial \ln L}{\partial \rho} = \frac{A\rho}{1-\rho^2} - \frac{\rho}{(1-\rho^2)^2} \sum_{i=1}^{\ell} \sum_{j=a_i}^{b_i} [\, w_{[ij]} - \rho\, z_{(ij)}]^2 + \frac{1}{1-\rho^2} \sum_{i=1}^{\ell} \sum_{j=a_i}^{b_i} z_{(ij)} [\, w_{[ij]} - \rho\, z_{(ij)}] = 0 \tag{5.5.5}$$

and

$$\frac{\partial \ln L^*}{\partial \mu_2} = \frac{2p}{k\sigma_2} \sum_{i=1}^{\ell} \sum_{j=a_i}^{b_i} [\, \gamma_{ij} + \delta_{ij}\, z_{(ij)}] + \frac{1}{\sigma_2} \sum_{i=1}^{\ell} r_i \beta_i [\, z_{(i,a_i)} + z_{(i,b_i)}] = 0 \tag{5.5.6}$$

$$\frac{\partial \ln L}{\partial \phi} = \frac{2p}{k\sigma_2} \sum_{i=1}^{\ell} u_i \sum_{j=a_i}^{b_i} [\, \gamma_{ij} + \delta_{ij}\, z_{(ij)}] + \frac{1}{\sigma_2} \sum_{i=1}^{\ell} r_i \beta_i u_i [\, z_{(i,a_i)} + z_{(i,b_i)}] = 0 \tag{5.5.7}$$

$$\frac{\partial \ln L}{\partial \sigma_2} = -\frac{N}{\sigma_2} + \frac{2p}{k\sigma_2} \sum_{i=1}^{\ell} \sum_{j=a_i}^{b_i} z_{(ij)} [\, \gamma_{ij} + \delta_{ij}\, z_{(ij)}] + \frac{1}{\sigma_2} \sum_{i=1}^{\ell} r_i [\, \alpha_i (z_{(i,b_i)} - z_{(i,a_i)}) + \beta_i (z^2_{(i,a_i)} + z^2_{(i,b_i)})] = 0 \tag{5.5.8}$$

where we have used the information from (5.5.2)-(5.5.5) to eliminate terms from (5.5.6)-(5.5.8), as discussed in Section 5.2. Now, as before, equations (5.5.6)-(5.5.8) are identical to those equations derived for the univariate linear regression model for censored samples presented in Section 4.4, with μ and σ being replaced by μ_2 and σ_2. Accordingly, the MML estimators for these three parameters are:

$$\hat{\mu}_2 = \bar{\bar{y}}^* - \hat{\phi}\,\bar{u}_m^* \tag{5.5.9}$$

$$\hat{\phi} = \frac{S^*_{yu(m)}}{S^*_{uu(m)}} \tag{5.5.10}$$

$$\hat{\sigma}_2 = \frac{B^* + \sqrt{B^{*2} + 4A^*C^*}}{2\sqrt{A^*(A^* - 2)}} \tag{5.5.11}$$

where all notation is defined in Section 4.4, with the modifications:

$$\bar{u}_m^* = \frac{1}{M^*}\sum_{i=1}^{\ell} m_i^* u_i$$

$$S^*_{yu(m)} = \sum_{i=1}^{\ell} m_i^* (u_i - \bar{u}_m^*)\,\bar{y}_i^*$$

$$S^*_{uu(m)} = \sum_{i=1}^{\ell} m_i^* (u_i - \bar{u}_m^*)^2$$

where 'm' denotes weights with respect to the m_i's. Now using the results in (5.5.9)-(5.5.11), we may proceed to compute the MML estimators for the remaining parameters in the same fashion as Section 5.2. We first define the following notation (where $A_i = n_i - 2r_i$):

$$\overline{x}_i = \frac{1}{A_i}\sum_{j=a_i}^{b_i} x_{(ij)} \; ; \qquad \overline{\overline{x}} = \frac{1}{A}\sum_{i=1}^{\ell} A_i\,\overline{x}_i$$

$$\overline{Y}_i = \frac{1}{A_i}\sum_{j=a_i}^{b_i} y_{(ij)} \; ; \qquad \overline{\overline{Y}} = \frac{1}{A}\sum_{i=1}^{\ell} A_i\,\overline{Y}_i$$

$$\overline{u}_A = \frac{1}{A}\sum_{i=1}^{\ell} A_i\,u_i$$

$$S_{xu} = \sum_{i=1}^{\ell} A_i\,(u_i - \overline{u}_A)\,\overline{x}_i$$

$$S_{yu}^{A} = \sum_{i=1}^{\ell} A_i\,(u_i - \overline{u}_A)\,\overline{Y}_i$$

$$S_{uu}^{A} = \sum_{i=1}^{\ell} A_i\,(u_i - \overline{u}_A)^2$$

using which the remaining MML estimators become:

$$\hat{\mu}_1 = \overline{\overline{x}} - \hat{\psi}\,\overline{u}_A - \hat{\rho}\left(\frac{\vartheta_1}{\vartheta_2}\right)[\,\overline{\overline{Y}} - \hat{\mu}_2 - \hat{\phi}\,\overline{u}_A\,] \qquad (5.5.12)$$

$$\hat{\psi} = \frac{S_{xu}}{S_{uu}^{A}} - \hat{\rho}\left(\frac{\vartheta_1}{\vartheta_2}\right)\left[\frac{S_{yu}^{A}}{S_{uu}^{A}} - \frac{S_{yu(m)}^{*}}{S_{uu(m)}^{*}}\right] \qquad (5.5.13)$$

$$\hat{\rho} = \frac{\hat{\sigma}_2}{\hat{\sigma}_1}\,\frac{S_{12A(p)}}{S_{2A(p)}^2} \qquad (5.5.14)$$

$$\hat{\sigma}_1 = \sqrt{S_{1A(p)}^2 + \frac{S_{12A(p)}^2}{S_{2A(p)}^2}\left[\frac{\vartheta_2^2}{S_{2A(p)}^2} - 1\right]} \qquad (5.5.15)$$

where

$$A\,S_{1A(p)}^2 = \sum_{i=1}^{\ell}\sum_{j=a_i}^{b_i}[\,x_{(ij)} - \overline{x}_i\,]^2$$

$$A\,S_{2A(p)}^2 = \sum_{i=1}^{\ell}\sum_{j=a_i}^{b_i}[\,y_{(ij)} - \overline{Y}_i\,]^2$$

$$A\,S_{12A(p)} = \sum_{i=1}^{\ell}\sum_{j=a_i}^{b_i}[\,x_{(ij)} - \overline{x}_i\,]\,[\,y_{(ij)} - \overline{Y}_i\,]$$

denote the 'pooled' estimators for censored samples.

We note here that these estimators just derived are essentially identical to the bivariate regression results for complete samples, with each expression being replaced by the corresponding expression based on censored samples. Of course, when there is no censoring ($r_i = 0$), all results simplify to the complete sample results.

5.5.1 The Variance-Covariance Matrix for Censored Samples

Following the same procedure outlined in Section 5.3, we arrive at the following sub-matrices of the information matrix, similar to those given previously:

$$
I_1^* = \begin{bmatrix}
\dfrac{A}{\sigma_1^2(1-\rho^2)} & \dfrac{-A\rho}{\sigma_1\sigma_2(1-\rho^2)} & \dfrac{A\bar{u}_A}{\sigma_1^2(1-\rho^2)} & \dfrac{-A\rho\,\bar{u}_A}{\sigma_1\sigma_2(1-\rho^2)} \\
 & \dfrac{A\rho^2}{\sigma_2^2(1-\rho^2)} + \dfrac{2pM^*}{k\sigma_2^2} & \dfrac{A\bar{u}_A}{\sigma_1\sigma_2(1-\rho^2)} & \dfrac{A\rho^2\bar{u}_A}{\sigma_2^2(1-\rho^2)} + \dfrac{2pM^*\bar{u}_m}{k\sigma_2^2} \\
 & // & \dfrac{\sum_{i=1}^{k} n_i u_i^2}{\sigma_2^2(1-\rho^2)} & \dfrac{\sum_{i=1}^{k} n_i u_i^2}{\sigma_1\sigma_2(1-\rho^2)} \\
 & & & \dfrac{\rho^2\sum_{i=1}^{k} n_i u_i^2}{\sigma_2^2(1-\rho^2)} + \dfrac{2p}{k\sigma_2^2}\sum_{i=1}^{k} m_i^* u_i^2
\end{bmatrix}
$$

and

$$I_2^* = \begin{bmatrix} -\frac{4N}{\sigma_1^2} + \frac{\rho^2 Q^*}{\sigma_1^2(1-\rho^2)} & \frac{-\rho^2 Q^*}{\sigma_1\sigma_2(1-\rho^2)} & \frac{\rho}{\sigma_1(1-\rho^2)}[-2A + Q^*] \\ & -\frac{A}{\sigma_2^2} - \frac{\rho Q^*}{\sigma_2^2(1-\rho^2)} + \Omega^* & \frac{-\rho Q^*}{\sigma_2(1-\rho^2)} \\ // & & \frac{1}{1-\rho^2}\left[\frac{2A\rho^2}{1-\rho^2} + Q^*\right] \end{bmatrix}$$

where $Q^* = \sum_{i=1}^{t}\sum_{j=a_i}^{b_i} E(z_{(ij)}^2)$ and $\Omega^* = \frac{2p}{k\sigma_2^2}\sum_{i=1}^{t}\sum_{j=a_i}^{b_i}[\gamma_{ij}t_{(ij)} - 2\delta_{ij}E(z_{(ij)})]$.

Then the variance-covariance matrix $\underset{\sim}{V}$ for the censored sample case is:

$$\underset{\sim}{V} = \underset{\sim}{I}^{-1} = \begin{bmatrix} I_1^{*-1} & \underset{\sim}{0} \\ \underset{\sim}{0} & I_2^{*-1} \end{bmatrix} = \begin{bmatrix} V_1 & \underset{\sim}{0} \\ \underset{\sim}{0} & V_2 \end{bmatrix} .$$

Once again, it is easily seen that I_1^* and I_2^* are generalizations of I_1 and I_2, and clearly reduce to those matrices when $r_1 = 0$ (i.e. when there is no censoring).

As with the complete-sample bivariate regression case, all results given in the preceeding two sections may easily be generalizable to the multiple reegression setting. One merely replaces each estimator with the corresponding result based on symmetric censoring. For the purposes of conciseness, these results will not be presented here.

Summary and Conclusion

In summary it is evident that the procedures derived in this dissertation present a viable alternative for data analysis when the normality assumption may not be justified. From the simulations, it is seen that the MML technique is remarkably efficient and powerful, even for small samples from a decidedly non-normal distribution. The following are just a few points in favour of using these procedures:

(1) The MML estimators are asymptotically equivalent to the ordinary maximum likelihood estimators, yet numerical iteration is not required to compute them.

(2) The choice of the shape parameter p (based on minimizing the model MSE) "picks" the best distribution from the symmetric location-scale family, for a particular data set.

(3) The MML estimators produced are very similar in form to results from the classical procedures, and are not much more computationally intensive. As a result, the MML procedures are likely candidates for adoption by the typical user of statistical routines.

(4) As has been shown, this procedure can be extended to a variety of models (eg. regression, ANOVA, etc.).

(5) The same procedure can be carried out for any other symmetric (or even asymmetric) location-scale family of distributions.

(6) The procedure incorporates both the complete and censored sample cases.

In conclusion, then, the non-normal procedures laid out in this thesis can be used as powerful tools for analyzing data in a variety of settings. It is hoped that due to their familiarity of form, they may be easily integrated into mainstream statistical knowledge.

REFERENCES

Andrews, D.F., Bickel, P.J., Hampel, F.R., Huber, P.J., Rogers, W.H., and Tukey, J.W. (1972). Robust Estimates of Location. New Jersey: Princeton University Press.

Barnett, V. (1966). Order statistics estimators of the location of the Cauchy distribution. Journal of the American Statistical Association, 61, 1205-1218.

Bartlett, M.S. (1953). Approximate confidence intervals. Biometrika, 40, 12-19.

Bhattacharya, G.K. (1985). The asymptotics of maximum likelihood and related estimators based on Type II censored data. Journal of the American Statistical Association, 80, 398-404.

Box, G.E.P., and Cox, D.R. (1964). An analysis of transformations (with discussion). Journal of the Royal Statistical Society, B26, 211-252.

Brown, M.B. (1975). Exploring interaction effects in the ANOVA. Journal of the Royal Statistical Society, C24, 288-298.

David, H.A. (1981). Order Statistics (2nd Ed.). New York: John Wiley & Sons.

Draper, N.R., and Smith, H. (1981). Applied Regression Analysis. New York: John Wiley & Sons.

Geary, R.C. (1947). Testing for normality. Biometrika, 34, 209-220.

Hicks, C.R. (1973). Fundamental Concepts in the Design of Experiments. New York: Holt, Rinehart and Winston.

Hoaglin, D.C., Mosteller, F., and Tukey, J.W. (1983). Understanding Robust and Exploratory Data Analysis. New York: John Wiley & Sons, Inc.

Hogg, R.V., and Craig, A.T. (1978). Introduction to Mathematical Statistics. New York: Macmillan Publishing Co., Inc.

Hopkins, T., and Phillips, C. (1988). Numerical Methods in Practice: Using the NAG Library. Reading Mass.: Addison-Wesley.

Huber, P.J. (1981). Robust Statistics. New York: John Wiley & Sons.

John, P.W.M. (1971). Statistical Design and Analysis of Experiments. New York: The Macmillan Company.

Launer, R.L. and Wilkinson, G.N. (1979). Robustness in Statistics. New York: Academic Press, Inc.

Lee, K.R., Kapidia, C.H., and Dwight, B.B. (1980). On estimating the scale parameter of the Rayleigh distribution from double censored samples. Statistiche Hefte, 21, 14-29.

Page, R.L., and Didday, R.L. (1986). FORTRAN 77 for Humans. St. Paul: West Publishing Co.

Phillips, J. (1986). The NAG Library: A Beginner's Guide. New York: Oxford University Press.

Rey, W.J.J. (1970). Lecture Notes in Mathematics - Robust Statistical Methods. New York: Springer-Verlag.

Schrader, R.M., and McKean, J.W. (1977). Robust analysis of variance. Communications in Statistics - Theory and Methods, A6(9), 879-894.

Staudte, R.G., and Sheather, S.J. (1990). Robust Estimation and Testing. New York: John Wiley & Sons, Inc.

Tiku, M.L. (1967). Estimating the mean and standard deviation from a censored normal sample. Biometrika, 54, 155-165.

Tiku, M.L. (1968a). Estimating the parameters of log-normal distribution from censored samples. Journal of the American Statistical Association, 63, 134-140.

Tiku, M.L. (1968b). Estimating the mean and standard deviation from progressively censored normal samples. Journal of the Indian Society of Agricultural Statistics, 20, 20-25.

Tiku, M.L. (1970). Monte Carlo study of some simple estimators in censored normal samples. Biometrika, 57, 207-221.

Tiku, M.L. (1971). Student's t-distribution under non-normal situations. Australian Journal of Statistics, 13, 142-148.

Tiku, M.L. (1973). Testing group effects from Type II censored normal samples in experimental design. Biometrics, 29, 25-33.

Tiku, M.L. (1980). Robustness of MML estimators based on censored samples and robust test statistics. Journal of Statistical Planning and Inference, 4, 123-143.

Tiku, M.L. (1982). Testing linear contrasts of means in experimental design without assuming normality and homogeneity of variances. Biometrical Journal, 24, 613-627.

Tiku, M.L. (1988). Modified maximum likelihood estimation. In. S. Kotz and N.L. Johnson, Eds., Encyclopedia of Statistical Sciences, Supplemental Vol., John Wiley & Sons, New York, 98-103.

Tiku, M.L., and Gill, P.S. (1989). Modified maximum likelihood estimators for the bivariate normal based on Type II censored samples. Communications in Statistics - Theory and Methods, 18, 3503-3518.

Tiku, M.L., and Gill, P.S. (1990). Testing the mean vector and the correlation coefficient in life-testing under bivariate normality. Biometrical Journal, 32(8) (Invited Paper), 899-913.

Tiku, M.L. and Kambo, N.S. (1992). Estimation and hypothesis testing for a new family of bivariate nonnormal distributions. Communications in Statistics - Theory and Methods, 21(6), 1683-1705.

Tiku, M.L., and Kumra, S. (1981). Expected values and variances and covariances of order statistics for a family of symmetric distributions (Student's t). In: Selected Tables in Mathematical Statistics, Vol. 8. American Mathematical Society, Providence, RI; 1985, 141-270.

Tiku, M.L., and Suresh, R.P. (1992). A new method of estimation for location and scale parameters. Journal of Statistical Planning and Inference, 30, 281-292.

Tiku, M.L., Tan, W.Y., and Balakrishnan, N. (1986). Robust Inference. New York: Marcel Dekker.

Vaughan, D.C. (1992). On the Tiku-Suresh method of estimation. Communications in Statistics - Theory and Methods, 21.

Wonnacott, T.H., and Wonnacott, R.J. (1977). Introductory Statistics (3rd Ed.). New York: John Wiley & Sons, Inc.