FACTORS THAT INFLUENCED THE EXTENT OF DAMAGES TO ABORTING WOMEN AT THE GANDHI MEMORIAL HOSPITAL IN ADDIS ABABA

A THESIS
PRESENTED TO THE
SCHOOL OF GRADUATE STUDIES
ADDIS ABABA UNIVERSITY

IN PARTIAL FULFILMENT OF THE
REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE
IN STATISTICS

BY
ZELEKE BEKELE WORKU
JUNE 1992

ADDIS ABABA UNIVERSITY

SCHOOL OF GRADUATE STUDIES

TITLE OF RESEARCH Factors that Influenced the extent of damages to aborting women at the Ghandhi Hospital.

NAME OF CANDIDATE Zeleke Bekele

DEPARTMENT Statistics

FACULTY Science

Approved By

	Advisor & Chairman	External Examiner	Internal Examiner
ame	Prof. Ayenew Ejigou	Dr. Desta Hamito	Dr. Alemayehu Melaku
ign.			

Date June 5, 1992

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

ACKNOWLEDGEMENT

I feel very much honoured to express my deepest and most sincere gratitude and appreciation to my advisor, **Professor Ayenew Ejigou**, for every indispensable professional assistance and guidance that he generously gave me all the way from the begining to the end of this project work. Little would have been achieved had it not been for his relentless efforts throughout the course of this study.

Further, my profound thanks go to the following:

Dr. Molla Tsega, the medical director of the Gandhi Memorial Hospital, **Dr. Getinet Abebe** and **Dr. Selamawit Ashagrie**, both resident gynaecologist-obstetricians at the Tikur Anbessa Hospital, for collecting clinical data on all 3151 cases of abortion that have been studied here.

The **Ministry of Health** and the **Gandhi Memorial Hospital**, for granting permission to have the clinical data of this study collected and analysed.

Ato Solomon Zewdie of the ILCA and **Woizerit Serkalem Demissie** of the Ministry of Health, for making available relevant statistical packages along with their manuals for analytical procedures involved in this project work.

The **School of Graduate Studies** of Addis Ababa University, for financial and administrative assistances it kindly offered.

ABSTRACT

Out of a total of 9817 cases of abortion that were observed at the Gandhi Memorial Hospital (GMH) from September 1990 to September 1991 (1983 Ethiopian Calendar), 3151 cases were studied in an attempt to determine specific factors that strongly influenced the extent of damage caused on women who were cases of abortion. Analyses were done using two different approaches: the log-linear and the logistic models. The two methods of analyses have led to more or less similar findings.

Seven explanatory variables (condition on arrival at hospital, the order of abortion, the type of abortion, the mechanism used to induce abortion, the duration of pregnancy, gravidity, and already-existing medical complications) were used to study the factors influencing the eventual health condition of patients at discharge from the Gandhi Memorial Hospital.

The eventual health condition of the patient at discharge from the hospital is found to be strongly influenced by three explanatory variables : the patient's health condition on arrival to the hospital, the type of abortion, and already-existing medical complications.

Recommendations are made to legalize or at least liberalize abortions at the Gandhi Memorial Hospital and other government health facilities, and to promote the provision of sex-education and family-planning methods.

1. INTRODUCTION

1.1 OBJECTIVES OF STUDY

It is believed that most abortions rarely reach the Gandhi Memorial Hospital (GMH), the only maternity hospital in Ethiopia, before they are induced elsewhere. Existing laws prohibit the physician in charge from practising abortions at the hospital, with the result that people with no medical expertise make a highly attractive business by routinely inducing bleeding which is followed by the disintegration of the fetus.

When such cases finally find their way to the emergency room of the GMH, the physician on duty is left with no choice other than saving the life of the troubled woman by way of carrying out the otherwise prohibited procedure of medical curretage and evacuation of the uterus. Without such measures, there is little doubt that the patient will almost certainly expire.

There exist several reasons to believe that criminal abortions will continue to be made if they have to. Perhaps, the most important is that, abortion seekers practice criminal abortion in an attempt to get rid of unwanted pregnancies even when such an attempt may cost them their lives or is likely to render them permanently disabled.

The study attempts to answer the following questions:

(i). What factors influence the extent of damages caused to aborting women, as observed at the point of discharge from the hospital ?

(ii). Which of the factors in (i) above are most influential ? i.e. Which of them have significant interdependence with the eventual health condition of the patient at discharge from the hospital ?

(iii). What feasible measures can be taken to minimize the problems in (ii) above ?

The aim of this research is to identify factors the control of which may help to reduce existing potential risks to the lives and health of women who happen to be cases of abortion

1.2 LIMITATIONS OF STUDY

The research admits the following limitations:

From among a total of 9817 cases of abortions (spontaneous and illegal) reported to the GMH from September 1990 to September 1991, only 3151 of them have been considered because in only 3151 hospital cards, all variables of study were correctly recorded and consequently confirmed clinically. It, therefore, follows that this is an observational study whose validity rests on the assumption that:

(i) Those cases with incomplete records are not basically any different from those with complete records in overall characteristics. This opinion is held by practising physicians at the Gandhi Memorial Hospital.

(ii) The selected cases of abortion constitute a sample from similar cases in the period immediately before and after September 1990 to September 1991.

2. DATA COLLECTION AND ANALYSES

2.1 VARIABLE SELECTION AND DATA COLLECTION

2.1.1 VARIABLE SELECTION

The following ten variables were originally selected, but two of them, X6 and X8, were not used for analyses :

I. Condition of patient at discharge (Y)

The condition of patient at discharge from the hospital is the dependent variable of the research, which is denoted by Y.
A patient at discharge is classified as dead, significantly damaged or insignificantly damaged as follows:

(i) Dead

In this case, death has been confirmed clinically.

(ii) Significantly damaged

A case is classified as ' significantly damaged ' if one or more of the following disorders has taken place:
total infertility, renal failure, sepsis, lacerated cervix, post abortive bleedings or shock, perforated uterus due to excessive or rigorous manipulation of the lower genital tract by unskilled workers, post-abortive pelvic inflamatory disease, heart failure, peritnotis (pus), post abortive absus, traumatic genital such as the uterus, vulva or cervix, or any other major disorder in vital reproductive organs as a consequence of the practice of abortion.

(iii) Insignificantly damaged

This includes cases with slight problems involving morbidity, anaemia or minor damages managed at the OPD (Out Patients Department) level of the hospital.

II. CONDITION ON ARRIVAL AT HOSPITAL (X1)

If there is profuse bleeding or the patient is acutely sick-looking on arrival at the OPD of the hospital, her condition on arrival is coded as 'bad'. Otherwise, it is regarded as being 'fair' to ' moderate ' .

III. THE ORDER OF ABORTION (X2)

The more frequent the practice of abortion, the more confident the woman becomes to practice it, while at the same time, her capacity to sustain subsequent pregnancies and end up with live-births diminishes, because the cervix (entrance to the uterus) gets loosened. The fact is that, there is an upper limit restricting the number of times illegal abortions can be made without running the risk of being significantly damaged or dead. Hence, X2 is also expected to influence Y.

IV. THE TYPE OF ABORTION (X3)

Abortions are said to occur spontaneously if there's no medical history of interference made to interrupt pregnancy. A pregnancy may be a wanted one, but can be abruptly interrupted due to an accident, a strenuous physical exercise, an excessive excitement, terrible shock, and other extraordinary reasons.

Such abortions are called spontaneous. If on the other hand, an unwanted pregnancy gets interrupted by the insertion of a catheter or by way of drinking poisonous liquors or drugs with inducing side effects, we say an illegal abortion has been made. It goes without saying that illegal abortions are daring or even suicidal, and far more risky than spontaneous ones, given the circumstances under which they usually occur. Hence, the inclusion of X3 was considered reasonable.

V. THE MECHANISM USED TO INDUCE ABORTION (X4)

The extent to which damages are caused is expected to depend on the mechanism used to induce bleeding (which is profuse in most cases), and the opening of the cervix so as to allow the admission of the forefinger. According to the GMH, medical curretage can only be administered if the cervix admits the forefinger, and there is profuse bleeding.
Now, illegal abortion-seekers appear to be aware of these rules, and desperately seek mechanisms that are sure to result in profuse bleeding and the wide opening of the cervix so as to secure the otherwise unavailable privileges of medical curretage and evacuation of the uterus. However, not all of them settle their daring experiences with a clean bill. A sizeable number of them actually get rid of unwanted pregnancies at the expense of major obstructions of vital reproductive organs or even death. Hence, X4 has also been taken as a potential explanatory variable to account for the variation in Y.

VI. THE DURATION OF PREGNANCY (X5)

Up until three months time, abortion is a fairly simple matter to manage clinically, but from then on, it becomes quite complicated and extremely risky. This is because three months after conception, all organs of the fetus are fully formed.
When and if abortions are made under these circumstances, medical curretage and evacuation of the uterus are administered using 'pitocin' and 'general anaesthesia', in which case the patient may be severely damaged, particularly depending on the mechanism used to induce abortion and her condition on arrival at the hospital. Blood transfusion is required along with meticulous follow-ups in cases where, for example, a five-month pregnant woman, with already existing medical complications, is rushed to the emergency room of the hospital with profuse bleeding induced by the insertion of a metal catheter.

The World Health Organization(**Martin, 1978**) defines abortion as a process of the expulsion of the fetus when the expelled fetus is less than or equal to 20 to 22 weeks old. In Ethiopia, this figure is as high as 28 weeks because of the absence of facilities such as incubators without which prematurely delivered babies can't be sustained. Hence, X5 is also taken as a possible candidate for explaining the variation in Y.

VII. THE AGE OF PATIENT (X6)

The age group, 13 to 20, is widely speculated to be the 'high-risk' age group because teenagers in this age group are exposed to lack of sex-education and family-planning methods as a result of which they often become victims of unwanted pregnancies.

Early exposure to sex is likely to result in unwanted pregnancy which, in turn, leads to the need to resort to potentially risky procedures for illegal abortions.

The variable X6 was recorded in this research as a categorical variable with two levels, when in fact, it should have been recorded as a continous variable, or at least with multiple categories. This failure has led to the decision to discard X6.

VIII. GRAVIDITY (X7)

Poor women with high gravidity(total number of children born plus abortions) may tend to seek criminal abortions. When women practice illegal abortion frequently, they may feel that they can get rid of unwanted pregnancies easily, but then the risk of being severely damaged or dead may increase as well.

IX. PARITY (X8)

The number of deliveries after a gestational period of 28 weeks, regardless of wether the new-born is alive or dead is called parity. If each new-born is alive, parity becomes the number of children. Most poverty-stricken mothers with a large number of children may seek criminal abortion because of the need to feed more mouths than they actually can. Hence, we take X8 as a potential explanatory variable for Y.

Neverthless, because X8 (parity) and X7 (gravidity) are very much similar and closely associated with one another, and since the proliferation of such a categorical variable had a tendency to lead to data sparsity and to problems of estimation, it was decided to discard X8.

X. PRE-EXISTING MEDICAL COMPLICATIONS (X9)

Major pre-existing medical complications such as hyper-tension, diabetis, heart failure, severe renal disorder, and other known chronic obstructions of vital reproductive organs may worsen the eventual condition of the patient. It was therefore, considered worthwhile to take X9 as another potential explanatory variable for Y.

2.1.2 THE CODING OF VARIABLES

The research uses the following 8 variables : Y, X1, X2, X3, X4, X5, X7, X9, where Y is the dependent variable to be studied and X1, X2, X3, X4, X5, X7 and X9 are independent or explanatory variables. Using the definitions given above, the following coding was adopted.

Table 2.1 : Levels of the eight variables used for analyses

VARIABLE	DESCRIPTION	CODE	
		LEVEL 1	LEVEL 2
Y	Condition at discharge	Dead or damaged significantly	Insignificantly damaged
X1	Condition on arrival	Bad	Fair to moderate
X2	Order of abortion	2, 3, ...	First time
X3	Type of abortion	Illegal	Spontaneous
X4	Mechanism of abortion	Insertion	Others
X5	Duration of pregnancy	≥ 3 months	< 3 months
X7	Gravidity	5, 6, 7, ...	1, 2, 3, 4.
X9	Pre-existing medical complications	Major	Minor or none

The two explanatory variables **X6 (age)** and **X8 (parity)** were discarded from all analyses for the reasons given earlier.

2.1.3. DATA COLLECTION

The task of collecting data on the eight variables for each of the 3151 cases of abortion was accomplished by Dr. Molla Tsega, the medical director of the Gandhi Memorial hospital, Dr. Getinet Abebe and Dr. Selamawit Ashagrie, both resident gynaecologist-obstetricians at the Tikur Anbessa Hospital, Addis Ababa.

2.2 METHODS OF ANALYSES

2.2.1 THE LOG-LINEAR MODEL

It is well known that collapsing multidimensional data in to a set of two-way tables could lead to an obscuring of the structure of the data and hence possibly to incorrect conclusions about relationships between variables(Upton, 1978). It is this desire to study the structure of multidimensional data that led Goodman (Goodman, 1973) to develop the log-linear model.

The log-linear model enables researchers to investigate the possible presence of interdependence between two or more variables that are suspected to be highly interactive with one another. As the number of variables and their respective levels rises, the interactions that exist become extremely complex. The log-linear model gives us a mathematically sound method of testing hypotheses about such relationships. The simplest form of multidimensional data is that given by the simple 2x2 table, and its analysis involves the application of the usual chisquare test of independence.

2.2.1.1 ANALYSIS OF THE BASIC 2X2 TABLE

In real life, the basic 2x2 contingency table is of little help to researchers as it is restricted to the capacity of analysing and testing the presence of significant interaction between only two factors at a time, on condition that all other variables can be ignored. However, studies based on the basic 2x2 table may be of little or no help for all practical purposes because they ignore too much vital information about the very subject of investigation.

The traditional practice of the data analyst faced with cross-tabulated data is to compute a chisquare test of independence for each 2x2 subtable. This strategy may not result in a systematic evaluation of the relationships between variables. Simpson's paradox(**Simpson, 1951**) is one such example whereby an interaction of order two may be both significant and insignificant at the same time(**Upton, 1978**).

Further, the classical chisquare approach does not give estimates of the effects of the variables. Regression analysis and the analysis of variance can't be used here, because we have categorical data. The log-linear model was therefore developed by Goodman (**Goodman, 1973**) to enable the analysis of such multivariate data. Log-linear models enable us to handle to handle any number of variables simultaneously and interpret the results appropriately.

To see how we can formulate such a model, let us consider the basic 2x2 table involving two categorical variables A and B each at two levels, henceforth denoted by 1 and 2. Such a table involves the use of a model which has as its subject the natural logarithms, V_{ij} of the cell probability P_{ij} .
The model consists of an average term, μ, which is the average of the logarithm of the cell probabilities, and three additive terms:

the two main effects, α_i^A , α_i^B , say, and one

interaction effect α_{ij}^{AB} corresponding to the two main effects.

Such a model is given by:

$$V_{ij} = \mu + \alpha_i^A + \alpha_j^B + \alpha_{ij}^{AB}$$

for $i, j = 1,2$

... (2.1)

where

$$\sum_i \alpha_i^{A} = \sum_j \alpha_j^{B} = \sum_i \alpha_{ij}^{AB} = \sum_j \alpha_{ij}^{AB} = 0 .$$

The four parameters for the model given in (2.1) could be listed as

μ , α_1^{A} , α_1^{B} and α_{11}^{AB} , and then from the restrictions given above on the α's, we get the following relations:

$$\alpha_2^{A} = - \alpha_1^{A} , \qquad \alpha_2^{B} = - \alpha_1^{B}$$

... (2.2a)

and,

$$\alpha_{22}^{AB} = - \alpha_{12}^{AB} = - \alpha_{21}^{AB} = + \alpha_{11}^{AB}$$

... (2.2b)

To get the defining relations for the α's, we need to define the following. Let

$$V_{i.} = \left(\sum_j V_{ij} \right) / 2 , \qquad V_{.j} = \left(\sum_i V_{ij} \right) / 2$$

$$V.. = \left(\sum\sum_{ij} V_{ij} \right) / 4$$

where the number of categories of each of the two variables, A and B, is 2 . Here,

$V_{i.}$ is the average of log probabilities of all cells in the ith row,

$V_{.j}$ is the average of log probabilities of all cells in the jth column, and

V.. is the average of all 4 log-probabilities of the table, for i = j = 1, 2.

Hence, we find that $\mu = V..$, and

$$\alpha_i^{A} = V_{i.} - V.. \quad , \qquad \alpha_j^{B} = V_{.j} - V.. \quad ,$$

$$\alpha_{ij}^{AB} = V_{ij} - V_{i.} - V_{.j} + V..$$

(2.3)

One must observe that the number of parameters in (2.1) is equal to the number of cells and on this account, the model in (2.1) is known as a **saturated log-linear model**. Further, the re-expression of the α's in terms of the cell probabilities shows that

α_i^{A} and α_i^{B} are proportional to the averages of the log odds of outcomes A_i and B_i, respectively, and α_{ii}^{AB} can be regarded as being proportional to the logarithm of the odds ratios of factor A computed at the two levels of B, or it can similarly be interpreted as the logarithm of the odds ratios of factor B computed at the two levels of factor A.

From a consideration of the hypothesis of independence in a 2x2 table, we can show that if A and B are independent, then

$$\alpha_{ij}^{AB} = 0 \quad \text{for} \quad i = j = 1,2.$$

Then, (2.1) reduces to the log-linear model of independence:

$$V_{ij} = \mu + \alpha_i^{A} + \alpha_j^{B}$$

$$i, j = 1, 2$$

... (2.4)

The goodness of fit of such a model can be assessed using the **LIKELIHOOD RATIO** test, which is a large sample test.

We shall now extend our results for the basic 2x2 table to the case of the categorical three-way, 2x2x2, table.
Let the three categorical variables A, B, C be given, and suppose that each of them has 2 categories. Let P_{ijk} be the unknown probability of a randomly selected observation falling in cell (i, j, k), and let $V_{ijk} = \text{Ln}(P_{ijk})$,

for i, j, k = 1, 2.

Then, following the method of reasoning used in the 2x2 table, the full(saturated) model is given by:

$$V_{ijk} = \mu + \alpha_i^{A} + \alpha_j^{B} + \alpha_k^{C} + \alpha_{ij}^{AB} + \alpha_{ik}^{AC} + \alpha_{jk}^{BC} + \alpha_{ijk}^{ABC}$$

for i, j, k = 1, 2

... (2.5)

In the model given in (2.5), the following restrictions hold.

$$\sum_i \alpha_i^{A} = \sum_j \alpha_j^{B} = \sum_i \alpha_{ij}^{AB} = \sum_j \alpha_{ij}^{AB} = \sum_i \alpha_{ik}^{AC}$$

$$= \sum_k \alpha_{ik}^{AC} = \sum_j \alpha_{jk}^{BC} = \sum_k \alpha_{jk}^{BC} = \sum_k \alpha_{ijk}^{ABC} = 0$$

Here, α_i^{A} relates to the log odds, α_{12}^{AB} to the log of the odds ratios and α_{ijk}^{ABC} to differences between the logarithms of odds ratios between factors A and B at the levels of C, or between A and C at the levels of B, or between B and C at the levels of A.

The definitions of the parameters in the saturated model in (2.5) follow the same pattern as for the 2x2 table.

Thus, $\alpha_i^{A} = V_{i..} - V...$

for i = 1, 2

... (2.5a)

Therefore, α_i^{A} is a measure of how much more (or less) likely category A is than the average A category.

Also,

$V_{ij.} = (1/K) \sum_k V_{ijk}$; $V_{i.k} = (1/J) \sum_j V_{ijk}$, etc

for i, j = 1, 2

Then, $\alpha_{ij}^{A} = V_{ij.} - V_{i..} - V_{.j.} + V...$

for i,j = 1, 2

... (2.5b)

$$\alpha_{ijk}^{ABC} = V_{ijk} - V_{ij.} - V_{i.k} - V_{.jk} + V_{i..} + V_{.j.} + V_{..k} - V...$$

... (2.5c)

The relation in (2.5.b) implies that α_{ij}^{AB} measures the extent to which the joint occurence of the categories A_i and B_j is more or less likely than would have been expected if variables A and B had been independent. Similarly,

α_{ijk}^{ABC} measures the extent to which the interdependence of variables A and B is itself dependent on the category of C. We can further extend the approach that we had used so far for the 2x2x2 table to any number of higher dimensions, and all analogous results remain essentially true.

As the number of variables and the number of their categories increases, so does the number of individual cells in the table. Since, more often than not, the number of cells exceeds the number of observations, there are large numbers of cells with zero cell frequencies. To avoid problems in such cases, the constant 0.5 is added to every cell frequency, so that no empty cells shall be encountered.

Indeed, the statistical package that has been used here(**Nie et al, 1989**) also provides for the addition of **0.5** when fitting the saturated model, and we have used this procedure for our analyses.

2.2.1.2 THE HIERARCHICAL LOG-LINEAR MODEL

A hierarchical log-linear model obeys the following rule, which is framed in the general multidimensional setting as follows:

Suppose that the parameter involving a set of variables S is included in the model. Then, the model must also contain or include all the parameters involving any subset of S.

For example, let α_{ijk}^{ABC} be included in a multidimensional model.

Then, α_i^{A} , α_j^{B} , α_k^{C} , α_{ij}^{AB} , α_{ik}^{AC} , α_{jk}^{BC} , α_{ijk}^{ABC}

must also appear in the model in order for the model to be hierarchical. Hence, for a 2x2 table with variables A and B, we could hypothesize that, if A and B are independent,

$$V_{ij} = \mu + \alpha_i^{A} + \alpha_j^{B} \qquad \ldots (2.6)$$

If the A categories are equally probable,

$$V_{ij} = \mu + \alpha_j^{B} \qquad \ldots (2.7)$$

If all categories are equally probable, we have :

$$V_{ij} = \mu \qquad \ldots(2.8)$$

The estimated cell frequency e_{ij} , in cell (i,j) , in such a 2x2 table are then given by :

$$e_{ij} = \left(\sum_{i=1}^{2} \sum_{j=1}^{2} f_{ij} \right) / 4 \qquad \ldots(2.9)$$

for i, j = 1, 2 , where f_{ij} is the observed frequency in cell (i,j).

Likewise, for the model involving dependence on factor A alone, the model is :

$$V_{ij} = \mu + \alpha_i^{A} \qquad \text{for } i = 1, 2$$

The estimated cell frequencies in row i are obtained by dividing the observed row total f_{io} by 2. Consequently, the sum of the estimated cell frequencies, e_{io} in row i is equal to the sum of the observed cell frequencies, f_{io} .

Likewise, it follows that $e_{oo} = f_{oo}$, for this model. Thus, for a log-linear model involving only μ , $e_{oo} = f_{oo}$,

and for a model involving $\alpha_i{}^A$, $e_{io} = f_{io}$, and $e_{oo} = f_{oo}$, where e_{oo} is the total of estimated frequencies, and f_{oo} is that for the observed frequencies.

Such models are examples of unsaturated log-linear models. The correspondence between the relevant estimated and observed marginal totals and the parameters being fitted in the model is not restricted to these simple situations, but is quite general(**Upton, 1978**).

For example, suppose that A, B, C and D are four dichotomous variables, with f_{ijkl} being a typical cell frequency. Let the only three-variable interaction in our unsaturated model be $\alpha_{ijk}{}^{ABC}$. If we denote the corresponding estimated cell frequency by e_{ijkl}, then

$$\sum_l e_{ijkl} = \sum_l f_{ijkl} \qquad \ldots(2.10)$$

since $\alpha_{ijk}{}^{ABC}$ is in the model, for all i, j, k = 1, 2.

Let us now sum the left and right hand side of (2.10) over all categories of variable D. Then, we get the following:

$$e_{ijk1} + e_{ijk2} = f_{ijk1} + f_{ijk2}$$

i.e., $e_{ijko} = f_{ijko}$

$$i, j, k = 1,2 \qquad \ldots(2.11)$$

In general, we get the identity between marginal totals of estimated and observed frequencies when the corresponding α is included in the model. This gives rise to the hierarchical log-linear model where, once a higher-order interaction is included in our model, all other lower-order interactions are automatically included into the same model under consideration.

2.2.1.3 THE SATURATED MODEL

If a log-linear model has as many parameters as there are cells, then it is called a **saturated model**. Because a saturated log-linear model contains an equal number of parameters and cells, it is a perfect fit which does not commit an error while estimating parameters. However, for practical purposes, such a saturated model is of little use. A useful model by contrast is one which provides an adequate fit to a given set of data with as few parameters as possible.

There exist methods that help us eliminate less important interactions from a saturated model so as to make the model much more realistic and easier to manage and interpret.

We recall the saturated model for the basic 2x2 table in (2.1). Since there are only four cells in a 2x2 table, a log-linear model can have at most four distinct parameters. On this account, the relation in (2.1) is an example of a saturated model.

The α's in a log-linear model can be estimated using the restrictions imposed on them earlier. For example, the data for the general IxJxK situation consist of observed frequencies $\{ f_{ijk} \}$ in the various cells $\{ (ijk) \}$ of the three-way classification.

If we write :

$Y_{ijk} = Ln(f_{ijk})$, then the parameter estimates can be easily derived. It can be shown that the estimator of

α_{ij}^{AB} is $(Y_{ij.} - Y_{i..} - Y_{.j.} + Y...)$

where $Y_{ij.} = (1/K) \sum_{k} Y_{ijk}$ and $Y_{i..} = (1/JK) \sum\sum_{j\ k} Y_{ijk}$, etc

This result can be generalized to more variables.
The **SPSS/PC + Version 4.0** package (**Nie et al, 1989**) has been used to find parameter estimates in this research.

The estimated variance of the natural logarithm of a poisson frequency is approximately the reciprocal of that frequency:

$$\hat{Var}(Y_{ijk}) \approx \frac{1}{f_{ijk}}$$

...(2.12)

We can write the estimator of α_{ijk} , $\hat{\alpha}_{ijk}$, as a linear combination of the cell frequencies :

$$\left(\hat{\alpha}_{ijk} \right) = \sum\sum\sum_{ijk} a_{ijk} Y_{ijk}$$ where the $\{ a_{ijk} \}$ are suitably chosen constants. Using the above relations, the estimated variance of

$\left(\hat{\alpha}_{ijk} \right)$ is approximately given by the following:

$$\mathrm{Var}\left[\hat{\alpha}_{ijk} \right] = \sum\sum\sum_{ijk} \left(a_{ijk} \right)^2 \cdot \frac{1}{f_{ijk}} \qquad \ldots (2.13)$$

Within a particular saturated model, the estimated variances of parameters need not all be the same,

and to put the $\hat{\alpha}$'s on an equal footing, we standardize them, so that the standardized value, $Z(\hat{\alpha}_{ijk})$, has variance 1.

Hence, the standardized value $Z(\hat{\alpha}_{ijk})$ is given below in (2.14) .

$$Z(\hat{\alpha}_{ijk}) = \frac{\hat{\alpha}_{ijk}}{[\mathrm{Var}\{\hat{\alpha}_{ijk}\}]^{1/2}} \qquad \ldots (2.14)$$

When the sample size is large, these standardized values are approximately distributed with mean equal to the corresponding α, and variance 1.

Important α's are selected(at the 0.05 level of significance), depending on wether their corresponding observed standardized values lie outside the range (-1.96, +1.96). This criterion of selection can be used as a rough guide only.

2.2.1.4 THE UNSATURATED MODEL

If a log-linear model contains a smaller number of parameters than there are cells, it is called an unsaturated model.

In unsaturated hierarchical log-linear models, the number of parameters is less than the number of cells so that the parameters in such a model may then be easier to interpret than in a saturated log-linear model.

Model reduction proceeds by first testing the family of pooled highest order interactions, and if this is not significant, the next lowest family of pooled interactions is tested simultaneously. The reduction continues until a family of pooled effects is significant(**Whittaker and Aitkin, 1978**).

The parameters for any fitted model are estimated by the **MLE (maximum likelihood estimation)** method on the basis of the multinomial sampling model.

In view of the considerable simplicity of the task of estimation and interpretation offered by an unsaturated model, it is worthwhile to review a mechanism with which only the most important effects can be selected while the least important effects are discarded from the saturated model.

This can be done by following the **" BACKWARD ELIMINATION "** procedure that is commonly used in regression analysis. This procedure is implemented on many modern statistical software packages, including **SPSS, BMDP, SAS** and **GENSTAT**. The SPSS option **(Nie et al, 1989)** of the **" BACKWARD ELIMINATION "** procedure has been used here.

2.2.1.5 MODEL DIAGNOSTICS

The null hypothesis that the optimum unsaturated log-linear model is a satisfactory fit is tested using the **PEARSON** or **LIKELIHOOD RATIO (LR)** chisquare test statistics. A **Normality** (p-p) plot was also used to detect radical departures.

(i). THE **PEARSON CHISQUARE** GOODNESS OF FIT TEST

The test of the null hypothesis that the fitted log-linear model fits the observed data in a 2x2 table, for example, can be based on the familiar **PEARSON-CHISQUARE** statistic given below:

$$X^2 = \frac{\sum_{ij} [f_{ij} - e_{ij}]^2}{e_{ij}} \qquad \ldots(2.15)$$

The degrees of freedom are given by the number of cells in the table minus the number of independent parameters in the model. We accept the null hypothesis that the fit is adequate, at the **0.05 level of significance**, if the value of the **PEARSON-CHISQUARE** statistic is less than the critical chisquare value.

(ii). THE **LIKELIHOOD RATIO** (**LR**) CHISQUARE STATISTIC

An alternative statistic for use in 2x2 tables is the **LIKELIHOOD RATIO** chisquare statistic given below:

$$X^2_{LR} = \sum_{ij} f_{ij} \; Ln[f_{ij} / e_{ij}] \qquad \ldots(2.16)$$

For large sample sizes, the two statistics are equivalent. The advantage of the **LIKELIHOOD RATIO** chisquare statistic is that, it, like the total sums of squares in the analysis of variance, can be split into interpretable parts that add up to the total. This property is useful in the **"BACKWARD ELIMINATION"** procedure of model selection.

2.2.2 THE LOGISTIC REGRESSION MODEL

2.2.2.1 THE LOGISTIC MODEL

Suppose that $Y' = (y_1, y_2, \ldots , y_n)$ is a sample of n independent random variables such that y_i is distributed as Bernoulli with parameters 1 and π_i, where π_i is unknown, and

$$\pi_i = \frac{\text{Exp} [X'_i \beta]}{1 + \text{Exp} [X'_i \beta]} \qquad \text{for } i = 1,2, \ldots , n$$

...(2.17)

In (2.17), X_i is a px1 vector of explanatory variables, mostly often continous, and β is an unknown parameter vector of the same dimension. Then, from (2.17), we find that :

$$\text{Logit}(\pi_i) = \text{Log} (\pi_i / 1-\pi_i) = X'_i \beta = \sum_{j=1}^{p} \beta_j X_{ij}$$

...(2.18)

Relation (2.18) represents the general logistic regression model. Given the vector of responses $Y' = (y_1, .., y_n)$, and the explanatory vector X_i , $i = 1, ..., p$, the objective of logistic regression analysis is to seek a parsimonious logistic regression model that provides a good fit to the data and estimates of β that can be used for further analysis and interpretation.

In linear regression, we estimate the parameters of the model using the method of least squares. That is, we select regression coefficients that result in the smallest sum of squared distances between the observed and the predicted values of the dependent variable.

In logistic regression, the parameters of the model are estimated using the maximum likelihood method. That is, the coefficients that make our observed results most likely are selected(**Hosmer and Lemeshaw, 1989**). Since the logistic regression model is non-linear, an iterative algorithm, the **NEWTON-RAPHSON** method of iteration, is used for parameter estimation.

2.2.2.2 INTERPRETING THE ODDS OF AN EVENT

We have seen that $Pr [Y_i = 1] = 1 / [1 + Exp(-Z_i)]$, where

$Z_i = X'_i \beta$.

If we now define φ as the odds of occurence of an event, then
φ = Pr [the event occurs] / Pr [the event does not occur]

Using this for the event, $\{ Y_i = 1 \}$, we have:

$$\text{Log}(\varphi_i) = \beta_o + \beta_1 X_{i1} + \ldots + \beta_p X_{ip} ,$$

where the regression line does not pass through the origin.

...(2.19)

From (2.19), we see that a logistic coefficient can be interpreted as the change in the log-odds associated with a one-unit change in the corresponding independent variable.

Since it is easier to think in terms of odds, rather than log-odds, the logistic model is rewritten in terms of odds as follows:

$$\Pr[Y_i = 1] \,/\, \Pr[Y_i = 0] = \text{Exp}[\, \beta_o + \beta_1 X_{i1} + \ldots + \beta_p X_{ip} \,]$$

... (2.20)

In (2.20), $\text{Exp}(-2\beta_j)$, $j = 1, \ldots , p$ is the factor by which the odds change when the jth independent variable increases by one unit, when the explanatory variables are in indicator mode.

If $-2\beta_j$ is positive, this factor will be greater than 1, which means that the odds are increased, but if $-2\beta_j$ is negative, the factor will be less than 1, which means that the odds are decreased. If $-2\beta_j$ is zero, the factor equals 1, which leaves the odds unchanged.

2.2.2.3 THE HUNT FOR AN OPTIMUM LOGISTIC MODEL

The " **STEPWISE BACKWARD ELIMINATION** " method is one of the procedures used to obtain an optimum logistic regression model. At each step, variables are evaluated for removal and entry.

The **LIKELIHOOD RATIO** test for the null hypothesis that the coefficients of the terms removed are zero is obtained by dividing the likelihood for the reduced model by the likelihood for the complete model. If the null hypothesis is true, and the sample size is sufficiently large, the quantity
-2 times the LIKELIHOOD RATIO statistic has a chisquare distribution with r degrees of freedom, where r is the difference between the number of parameters in the full model and the reduced model.

In general,the distributional properties of the likelihood ratio tests can be summarized as follows:

If $\beta_{max} = (\beta_1, \ldots , \beta_n)'$ where n is the number of observations, $\beta = (\beta_1, \ldots , \beta_p)'$, $p < n$, and the likelihood function is denoted by $L(\beta;Y)$, the model describes the data well if $L(\beta;Y) \approx L(\beta_{max};Y)$,

or poorly if $L(\beta;Y) << L(\beta_{max};Y)$. This suggests the use of the generalized **LIKELIHOOD RATIO (LR)** statistic as a measure of goodness-of-fit,

LR = $L[b_{max};Y] / L[b;Y]$, or $Log(LR) = l(b_{max};Y)-l(b;Y)$ where $l(b;Y)$ is the log-likelihood function evaluated at the maximum likelihood estimator b of β. Large values of Log(LR) provide evidence that β is a poor model for the data.

It can also be shown that the statistic $2[\, l(b;Y) - l(\beta;Y) \,]$ has the central chisquare distribution with p degrees of freedom, if the model is good.

We now use a test statistic based on this last result to assess the fit of a model and to compare alternative models. We define:

$$\textbf{DEVIANCE} = D = 2Log(LR) = 2[\, l(b_{max};Y) - l(b;Y) \,] \qquad \ldots(\ 2.21\ )$$

D can be rewritten as follows:

$$D = 2\ \{\ [\, l(b_{max};Y)-l(\beta_{max};Y)\,] - [\, l(b;Y)-l(\beta;Y)\,] + [\, l(\beta_{max};Y)-l(\beta;Y)\,]\ \} \qquad \ldots\ (\ 2.22\ )$$

The first expression on the right hand side of (2.22) has the central chisquare distribution with n degrees of freedom; the second term has the central chisquare distribution with p degrees of freedom, and the third term is a constant which is positive, but near zero if the model based on ß describes the data nearly as well as the maximal model does.

Thus, the statistic, **D**, has the central chisquare distribution with **n - p** degrees of freedom, if the model is good; if the model is poor, the third term on the right hand side of (2.22) will be large, and so **D** will tend to be larger than X^2 (n-p). In fact, **D** has the non-central chisquare distribution in this case.

2.2.2.4 DIAGNOSTIC PROCEDURES FOR THE LOGISTIC MODEL

The goodness-of-fit of the logistic regression model may be assessed using two different methods: the **CLASSIFICATION TABLE FOR Y** and the **LIKELIHOOD RATIO** statistic.

The first method involves the use of the classification table for the response vector Y. The classification table for Y gives four readings in a 2x2 table, for predicted and observed values of the binary response. The off-diagonal entries of the table tell us the number of cases that have been incorrectly classified.

The second method is based on the **LIKELIHOOD RATIO** statistic. A reliable method of assessing the goodness of fit of the model is to actually examine how likely the sample results are, given the parameter estimates. This is reasonable because we have chosen parameter estimates which would make our observed results as likely as possible, as reflected in the likelihood.

Since the likelihood is a small number less than 1, it's customary to use
-2[LOGLIKELIHOOD] as a measure of how well the estimated model fits the data.

A good model is one which results in a high likelihood of the observed results. This translates to a small value of **-2[LOGLIKELIHOOD]**. For a model that fits perfectly, the **LIKELIHOOD** is 1, so that **-2[LOGLIKELIHOOD]** is zero.

Hence, to test the null hypothesis that the observed likelihood does not differ from 1 (the value for which the model fits perfectly), we can use the statistic **-2[LOGLIKELIHOOD]** , which has a chisquare distribution with $n - p$ degrees of freedom, under the null hypothesis that the model fits almost perfectly with p parameters.

We reject the null hypothesis that the model fits well if the value of **-2[LOGLIKELIHOOD]** is greater than $X^2 (n - p)$ at the 0.05 level of significance, where p is the number of parameters in the model, and n is the sample size.

3. DATA ANALYSES AND RESULTS

3.1 ANALYSIS BASED ON 2x2 TABLES

A preliminary analysis was done, to survey which of the explanatory variables X1, X2, X3, X4, X5, X7 and X9 appear to have a strong association with the dependent variable Y.

The analysis was made using the classical independent test for 2x2 contingency tables, essentially after collapsing the multidimensional data that we have at hand. Major results are given in **Table 3.1** below.

Table 3.1 : Results of seven 2x2 tables

INTERACTION	PEARSON CHISQUARE	PHI COEFFICIENT	APPARENT RANK ORDER OF ASSOCIATION
Y by X1	107.4	0.185	2
Y by X2	1.8	0.024	6
Y by X3	34.4	0.104	3
Y by X4	22.9	0.085	4
Y by X5	8.9	0.053	5
Y by X7	0.0	0.002	7
Y by X9	332.2	0.325	1

Table 3.1 indicates that, from among the seven interactions of order two, Y by x1, ... , Y by X9 , the variables X9, X1, X3, X4 and X5 (in decreasing order of the strength of association with Y) appear to have significant associations with Y, at the 0.05 level of significance. Here, it is no surprise that the rank order is identical when based on the **PEARSON CHISQUARE** and the **PHI COEFFICIENT.**

Because of the several potential limitations of this method of analysis(collapsing multidimensional data to 2x2 contingeny tables), results shown above in Table 3.1 are not to be regarded as fully reliable; we will compare these results with the up-coming findings using the log-linear model, which is by far the most powerful technique for the analysis of such multidimensional data.

3.2 ANALYSIS BASED ON THE LOG-LINEAR MODEL

3.2.1 THE SATURATED MODEL

As a first step in the analysis using the log-linear model, parameters of the saturated model were estimated, and the following 11 significant interactions were observed on the basis of the criterion discussed in Section 2.2.1.3.

Table 3.2 : Significant interactions

#	INTERACTION	Z-VALUE
1.	y*x3*x4*x9	4.46
2.	y*x3*x4	-4.93
3.	x1*x3*x4	2.95
4.	y*x4*x9	-2.06
5.	x3*x4*x9	-6.33
6.	y*x1	1.95
7.	y*x4	2.53
8.	x3*x4	10.07
9.	x3*x7	-2.09
10.	y*x9	7.46
11.	x4*x9	3.09

Next, an attempt was made to arrive at an unsaturated model that would provide a good fit to the data with the smallest number of parameters in the log-linear model. This was attempted using hierarchical log-linear modelling. In the process, three of the 3151 cases had to be deleted since they appeared to be very unusual as reflected by their respective standardized residuals. The sample size for the log-linear analysis was therefore reduced from 3151 to 3148.

3.2.2 THE UNSATURATED MODEL

It is reasonable and desirable to find an unsaturated log-linear model containing a smaller number of significant lower-order interactions, without losing on gooness of fit. To this effect, the **"BACKWARD ELIMINATION"** procedure was used.

Using such a procedure, the final and optimal log-linear model came out to include the following effects :

```
 y by x1
 y by x3
x1 by x3
x3 by x7
 y by x9
x1 by x9
 y
x1
x3
x7
x9.
```

...(3.1)

The optimum log-linear model specified by the effects given above in (3.1) was finally estimated as shown in **Table 3.3.** Standard errors of estimated parameters and the corresponding 95% confidence intervals have also been included.

Table 3.3 : Parameter estimates for the optimum model

#	EFFECT	COEFF.	STD ERR.	Z-VALUE	95% C.I.
1.	y by x1	0.660	0.098	6.731	(.468, .852)
2.	y by x3	0.221	0.044	5.008	(.134, .307)
3.	x1 by x3	0.056	0.019	2.917	(.018, .095)
4.	x3 by x7	-0.240	0.022	-10.621	(-.284, -.196)
5.	y by x9	0.708	0.069	10.154	(.572, .847)
6.	x1 by x9	0.841	0.254	3.310	(.343, 1.339)
7.	y	-1.266	0.114	-11.035	(-1.490, -1.041)
8.	x1	1.527	0.269	5.667	(.999, 2.055)
9.	x3	-0.264	0.046	-5.737	(-.354, -.173)
10.	x7	-0.476	0.022	-21.055	(-.520, -.432)
11.	x9	-2.137	0.254	-8.397	(-2.636, -1.638)

Before accepting the results of the analysis based on the optimum log-linear model, we need to assess the adequacy of the model. We do this using the **LIKELIHOOD-RATIO** or the **PEARSON CHISQUARE** values arising from the model.

These are readily available from the output providing the estimates of all parameters of the unsaturated model. These statistics for goodness-of-fit are as follows:

LIKELIHOOD RATIO CHISQUARE = 9.43646 **DF** = 20 **P** = 0.977

PEARSON CHISQUARE = 10.48331 **DF** = 20 **P** = 0.959

Hence, the optimum log-linear model appears to fit the data well. The **NORMAL PROBABILITY PLOT** of the residuals shown below in **Fig 1** also suggests that there is no serious violation.

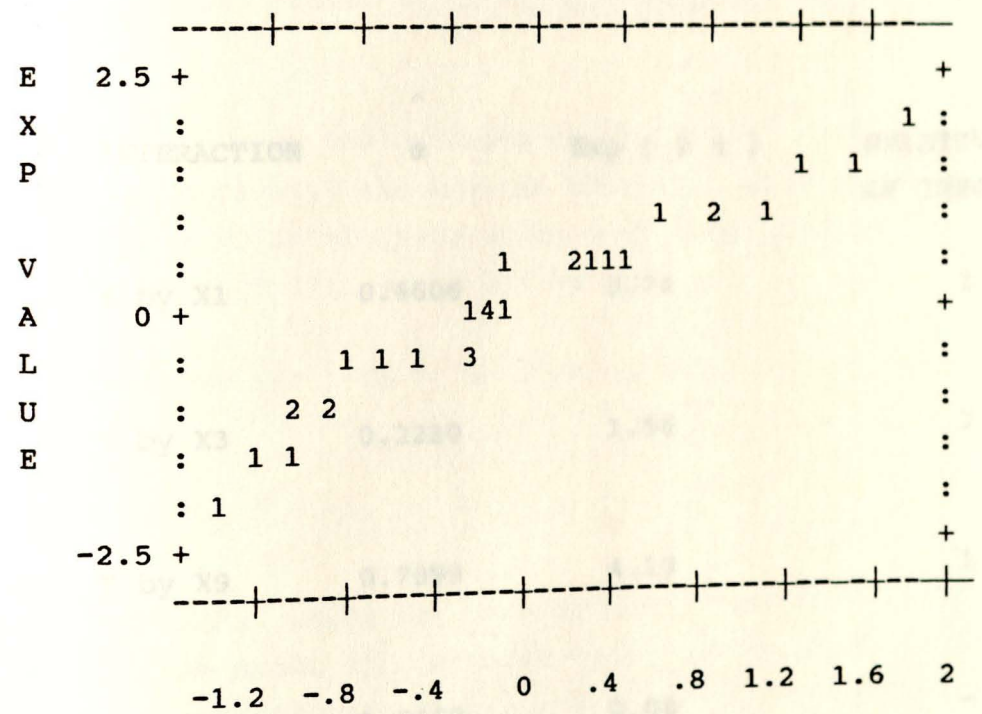

Fig 1. Normal probability plot

Now that the adequacy of the optimum log-linear model used above has been endorsed, it remains to discuss the results obtained using this same model.

All 11 effects given in Table 3.3 above are significant at the 0.05 level of significance. Out of these 11 significant interactions, 4 of them involve Y. These four are effects that were the objective of this research.

These significant interactions involving Y shall be further discussed, since these are the ones of primary interest in this study. Table 3.4 provides results of further analysis :

Table 3.4 : Significant effects involving Y

INTERACTION	$\hat{\alpha}$	Exp (2 $\hat{\alpha}$)	RELATIVE ORDER OF IMPORTANCE
Y by X1	0.6606	3.74	2
Y by X3	0.2210	1.56	3
Y by X9	0.7099	4.13	1
Y	-1.2660	0.08	-

Table 3.4 shows that all interactions are positive and significant so that each factor appears to raise the probability of significant damage or death. We shall first examine the effect of each variable at a time, and then consider their joint effect on the odds of serious damage to an aborting woman.

(i) Y by X1

The significance of the interaction between Y and X1 shows that serious damage(significant damage or death) in aborting women is positively related to the bad condition of women on arrival at the Gandhi Memorial Hospital.

We note that the estimate of α_{11}^{YX1} is 0.660. Alternatively, the odds of the event of a serious damage to an aborting woman with a bad condition on arrival at the hospital works out as :

Exp(2x0.660) to 1, or 3.74 to 1.

(ii) Y by X3

Likewise, using the corresponding results for X3, the odds of the event of a serious damage to an aborting woman that has experienced illegal abortion gets to be 1.56 to 1.

(iii) Y by X9

> The odds of the event of a serious damage to an aborting woman with major pre-existing medical complications is estimated at 4.13 to 1.

Hence, the odds of serious damage are most seriously affected by X9, followed by X1, and then by X3. All other variables in our study, excluding X6 and X8, which were dropped out early, appear to have no significant effect on the odds of experiencing a serious damage.

Lastly, the joint effect of these factors is clearly most pronounced if each of X1, X3 and X9 is at level 1. Aborting women with X1, X3 and X9 at level 1 have an estimated odds of a serious damage equal to [3.74x1.56x4.13x0.08] to 1, or 1.93 to 1.

3.3 ANALYSIS BASED ON THE LOGISTIC MODEL

3.3.1 ESTIMATES AND GOODNESS OF FIT

To start with, the logistic regression of Y on the explanatory variables X1, X2, X3, X4, X5, X7, X9 , using the **"BACKWARD STEPWISE ELIMINATION"** method was run. This procedure revealed that only X1, X3, X5, X7, X9 are important. The final(optimal) logistic regression model was therefore constituted by the five variables X1, X3, X5, X7, X9 only.

All explanatory variables were used under a deviation scheme whereby the value of X_i at level 1 is -1 and, at level 2, it is 1, i = 1, 2, ... ,n. Analysis was done using 3150 cases after removing one case which appeared to be an outlier.

The adequacy of the final model was then evaluated using two alternative methods.

(i). USING THE CLASSIFICATION TABLE FOR Y

The classification based on the logistic regression model was compared to that observed. The following table provides the results obtained.

Table 3.5 : The classification table for Y

	PREDICTED		
			PERCENT CORRECT
OBSERVED	0	1	
0	2979	29	99.04%
1	110	32	22.54%
		OVERALL	95.59%

As shown in **Table 3.5**, the adopted model has an overall predictive accuracy of **95.59%** , and this may ordinarily be considered adequate. However, whereas the model does well with the prediction of those that are not seriously damaged, its performance on the other category does not appear to be satisfactory, and this is a well-known (**Nie et al, 1989 , Page B-88**) experience with this model.

(ii) USING THE STATISTIC **-2[LOGLIKELIHOOD]**

The **DEVIANCE** is the second approach for assessing the goodness-of-fit of the logistic regression model. Results are given below in **Table 3.6.**

Table 3.6 : Measures of goodness-of-fit

STATISTIC	CHISQUARE	DF	SIGNIFICANCE
-2[LOGLIKELIHOOD]	1049.953	3145	1.0000
MODEL CHISQUARE	3316.874	5	.0000
GOODNESS-OF-FIT	2982.349	3145	.0000

The statistic **-2[LOGLIKELIHOOD]** compares the present model to a perfect model. The large observed significance level indicates that this model does not differ significantly from the perfect model.

The **MODEL CHISQUARE** statistic is the difference between **-2[LOGLIKELIHOOD]** for the model with only a constant and **-2[LOGLIKELIHOOD]** for the current model. The **MODEL CHISQUARE** tests the null hypothesis that the coefficients for all of the terms in the current model, except the constant, are zero. This is comparable to the overall **F-TEST** for regression. The degrees of freedom for the **MODEL CHISQUARE** are the difference between the degrees of freedom for the two models being compared.

It can be noted that the goodness-of-fit of the logistic regression model, as suggested by the classification table for Y appears to be fairly satisfactory, while the statistical test in Table 3.6 points otherwise. Even so, the output obtained using the logistic regression model does not seriously contradict earlier results obtained from the log-linear model. **Table 3.7** gives estimates arrived at using **"BACKWARD ELIMINATION"** ; all variables are in the **deviation** mode.

Table 3.7 : Estimates for the final logistic regression model

VARIABLE	B	S.E.	EXP(-2B)
x1(1)	-0.5464	0.1012	2.9826137
x3(1)	-0.5762	0.0913	3.1657817
x5(1)	-0.2588	0.0890	1.6779956
x7(1)	-0.3044	0.0927	1.8382242
x9(1)	-3.0777	0.1031	471.25531

The variables X1, X3 and X9 are still considered important factors, in the same sense and direction as that seen in the log-linear model. In addition, X5 and X7 appear to have some effect, but this is perhaps only apparent than real.

3.3.2 INTERPRETATION OF RESULTS

The results given above for the logistic regression model in **Table 3.7** can be used for further discussion and interpretation.

(i) The fitted regression equation
Using results in **Table 3.7**, the estimated logistic model is written as follows :

$\hat{}$

Prob[Y = 1] = 1 / [1 + Exp(-Z)], where

Z = -0.5464x1 - 0.5762x3 - 0.2588x5 - 0.3044x7 - 3.0777x9

(ii) The rate of change of the odds

We recall that Exp($-2\beta_i$) , i = 1, ... , p is the factor by which the odds change when the ith independent variable increases by one unit in the indicator mode. Using results in Table 3.7, the following five interpretations are apparent :

Exp($-2\beta_1$) = 2.9826 indicates that when x1 changes from 0 to 1 (condition on arrival, from good to bad), the odds that the patient gets discharged with serious damage are increased by a factor of 2.98.

The corresponding figures for X3, X5 and X7 are 3.17, 1.68, and 1.84, respectively.

The odds are raised tremendously when x9 changes from 0 to 1. In all cases, the binary values of 0 and 1 for the explanatory variables are such that 1 corresponds to level 1, while 0 corresponds to level 2, as has been defined in Section 2.1.2.

4. DISCUSSION, CONCLUSION AND RECOMMENDATIONS

First, we note that, from analyses of the 2x2 tables, the relative order of importance of the explanatory variables x1, ... , x9 , in decreasing order of the strength of association with Y, is given by **x9, x1, x3, x4 and x5.**

The corresponding order, for the log-linear model, taking the three variables involved in the optimal log-linear model, is given as : **x9, x1, x3.**

Thirdly, for the optimum logistic regression model, we have the order as : **x9, x3, x1, x5, x7.** Results obtained using the logistic regression model are not identical to results obtained using the log-linear model, and this is understandable since the analysis of such categorical data is best handled with the log-linear model, a model that has been explicitly designed to handle the analysis of categorical data. However, results from the logistic regression model have also been included here so that they may serve a comparative role as the percentage of correct classification is as high as **95.59%** .

It is worth-while to note that the classical 2x2 contingency table has given rise to results that are not too different from those obtained using the log-linear and the logistic regression methods of analyses.

The results from the log-linear analysis are the most reliable and informative since there is a good fit to the optimum model, and since the effects of factors are estimated taking that of others into consideration.

This was not the case for the 2x2 table, and the goodness-of-fit were not entirely satisfactory, as expected(**Nie et al, 1989**) in the case of the logistic regression method.

Based on the results of these analyses, one is led to conclude that the possibility of a serious damage to aborting women is further increased by **the condition on arrival, the type of abortion and the nature of pre-existing medical complications.** The control of these three factors, therefore, appears to be a reasonable approach to the reduction of serious damage.

In order to give appropriate recommendations in this area, we need to consider the nature of the following major facts that are directly relevant to the issue under consideration.

Sex education is hardly given at the schools, with the result that most school children are unaware of even simple methods which may be helpful to prevent unwanted pregnancies.

Criminal abortions appear to be made even when the penality for practising such abortions may be death or the permanent obstruction of vital reproductive organs.

The **Gandhi Memorial Hospital**, and hence the **Ministry of Health**, loses considerable resources (reagents, chemicals and drugs, labour, time, equipment, bed-occupation, money for the treatment of " free patients " , etc) in the daily routine of administering medical curretage and evacuation of the uterus; all this loss is incurred essentially to save troubled lives.

There exist regulations prohibiting practising gynaecologists and obstetricians from carrying out abortion procedures at government hospitals. The fact that 33% of those in the study have experienced illegal abortion indicates that a sizeable number of women have practised illegal abortion.

This prompts one to ask if there should be laws that are not obeyed. Should business men with no clinical expertise continue to indefinitely induce illegal abortions ? Regardless of the presence of prohibitive laws, the ultimate task of treating cases of criminal abortion rests on government hospitals. These cases are eventually rushed to the emergency rooms of hospitals at a point of time when little can be done to save their lives and the attempt to help the troubled patient is very costly to the hospitals.

Consideration of the account given above would lead to the following recommendations which may help identify workable remedial actions:

1. The provision of sex education and family planning methods must be available to school children.

2. Abortions need to be legalized or at least liberalized, even if that would mean a huge challenge to the government and a major attitudinal change to the society.

REFERENCES

Goodman, L. (1973). Simple methods for analysing three-factor interactions in contingency tables. **Journal of the American Statistical Association** 59, 319-352.

Hosmer, D.W. and Lemeshaw, S. (1989). **Applied Logistic Regression.** New York: Wiley.

Martin, B. (1978). Spontaneous and induced abortions, and the resulting mortality and morbidity. **WHO Technical Report Series** 461, 52-65.

Nie, N.H., Hull, C., Jenkins, J., Steinbrenner, K., and Bent, D. (1989). **SPSS : Statistical Package for the Social Sciences.** New York: McGraw-Hill.

Simpson, E.H. (1951). The interpretation of interaction in contingency tables. **Journal of Royal Statistical Society** B, 238-241.

Upton, G.J. (1978). **The Analysis of Cross Tabulated Data.** New York: Wiley.

Whittaker, J. and Aitkin, M. (1978). A flexible strategy for fitting complex log-linear models. **Biometrics** 34, 487-495.

DECLARATION

I, the undersigned, hereby declare that this thesis is my original work, and has not been presented for a degree in any university other than **Addis Ababa University.** I also certify that all sorts of utilized material have been dully acknowledged.

Name : Zeleke Bekele Worku
Signature : Zeleke
Place : The Department of Statistics
Science Faculty
Addis Ababa University
Addis Ababa, Ethiopia

I certify that this thesis has been submitted for examination with my approval as University advisor.

Professor Ayenew Ejigou
Chairman, The Department of Statistics
Addis ababa University.