SOME STATISTICAL INVESTIGATIONS

IN GENERAL INSURANCE

Roberto Westenberger

A Thesis Submitted for the Degree
of Doctor of Philosophy

The City University
Mathematics Department
(Actuarial Science)
London

June 1983

TABLE OF CONTENTS

List of Tables

List of Figures

ACKNOWLEDGEMENTS

I would like to express my deep gratitude to Professor Bernard Benjamin for his helpful supervision and constant encouragement throughout this research.

I am grateful also to Mr. Stewart Coutts for providing me with the data without which this study would not have been possible.

Thanks are also due to the staff of The City University for the library and computer facilities offered to me.

I am also indebted to the library staff of The Institute of Actuaries for their sympathetic assistance during this research.

I acknowledge also the financial support I received from Conselho Nacional de Desenvolvimento Científico e Tecnológico, who sponsored my research.

Finally, I would like to thank Ms. Mary Williams who carefully typed the final manuscript.

ABSTRACT

This research is concerned with the application of statistical methodology to general insurance, in particular to the motor insurance branch.

The research consists of two independent investigations in which two important aspects of motor insurance claims were studied. The aim of the first phase was to measure the influence of a given set of rating factors on both average claim size and claim frequency, and the second phase aimed at estimating an average pattern of settlement delays for motor insurance claims.

In the first investigation, five rating factors were considered : mileage, zone of garage, no claim bonus, vehicle make and vehicle age. No strong evidence was found that these five rating factors had a significant influence on the average claim, but the same was not true for the claim frequency, in which such an influence was not only detected, but also measured for a limited set of the observations.

An average pattern of settlement delays for motor insurance was estimated in the second investigation, based on two years of claims experience. It was shown that such a delay could give rise to a marginal profit for an insurance company, provided that the corresponding reserves for outstanding claims were appropriately invested.

Due to the difficulty in getting data from the insurance industry, the two phases of the research had to be carried out using data

from different sources. Thus, the first investigation was based on data from the Swedish claims experience during the year of 1977 and the second was based on data from a medium sized British insurance company relating to claims payments associated with accidents which took place in the course of the years 1972 and 1973.

KEY TO SYMBOLS AND ABBREVIATIONS

AC	Average claim (SP/NCLA)
AD	Accidental damage
B	No claim bonus
BMDP	Biomedical Computer Programs
ESP	Number of insured years (exposure)
EUNP	Estimated amount to be paid for unsettled claims
J	Vehicle age
LAC	Naperian logarithm of AC (ℓn (AC))
LMN	Naperian logarithm of MN (ℓn (MN))
M	Vehicle make
MN	Claim frequency ((NCLA × 100)/ESP)
$\overline{MN}$	Average claim frequency
NCLA	Number of claims
S	Mileage
SP	Sum of payments
SPN	Thousands of SP (SP/1000)
SPSS	Statistical Package for the Social Sciences
PAD	Payment for accidental damage claim
PTPBI	Payment for third party bodily injury claim
PTPPD	Payment for third party property damage claim
TPBI	Third party bodily injury
TPPD	Third party property damage
TSYR	Time of settlement (in years)
YACC	Year of accident
Z	Zone of garage

CHAPTER ONE

INTRODUCTION

The reliable estimation of future claims expenditure is one of the most difficult tasks faced by an insurance company, especially when this company deals with non-life business. Unlike life insurance, general insurance business is subject to claims of a varying size and which can happen more than once for the same policy.

It is of crucial importance for an insurance company to predict its future liabilities as accurately as possible, because financial reserves have to be set up in order to meet those liabilities. If the reserves are underestimated, the company will face problems in the future in terms of being unable fully to honour its promises. Alternatively, an overestimation of reserves will retard the emergence of profits for the company, and may reduce the growth prospects of the company.

Many attempts have been made to develop mathematical models aiming at forecasting future claims expenditure; however, one cannot say that there is a particular model which has gained general acceptance in practical terms to date. Even if one considers only a particular branch of general insurance, as for example motor insurance, no standard mathematical methodology can be regarded as being widely used by insurance companies to predict their claims expenditures.

One could divide the development of claims models in general insurance into two phases : before and after the foundations of Risk Theory had been set up by Filip Lundberg in the early part of this century (Jewell, 1980).

During the first phase, many attempts had been made to fit a great number of statistical distributions to past claims data. Thus binomial, Poisson and negative binomial distributions were successfully fitted to claim frequencies and normal, exponential, log-normal and Pareto distributions were fitted to claim amounts (Puzey, 1973).

It is worth noting that the mathematical models developed during this period concentrated only on particular branches of non-life insurance. This lack of generality was possibly the main reason why these models had been neglected by insurance companies.

During the second phase, a general approach to the claims process in non-life insurance was achieved with the development of the so-called Risk Theory. The importance of this theory lies in the fact that its compound Poisson model, although quite sophisticated in mathematical terms, is applicable to whatever kind of homogeneous portfolio i.e. it lacks specificity (Beard et al., 1977).

A great deal of research effort was necessary in order to formalize Risk Theory. The early developments of the theory were summarized by Dubourdieu (1952), and later on, Bühlmann (1970) gave a complete account of its mathematical framework, which made researchers aware of the difficulties to be overcome regarding its complex formulation.

It is thought that this complexity was the main reason why the theory did not achieve a practical appeal. Indeed, despite the efforts of Seal (1969), Beard et al., (1977), and more recently Gerber (1979), in giving an account of the whole theory with regard to practical applications, one can by no means say that Risk Theory achieved widespread use by insurance companies.

The mathematical complexity of the theory has, to a certain extent, prevented it from developing further for some time. With the advent of fast and relatively cheap computers, some hitherto intractable problems could be solved (for example, the evaluation of convolutions in Lundberg's integral formula) thus broadening the scope of the theory.

The horizons of Risk Theory have been further enlarged with the recent introduction of concepts from other disciplines. Thus Borch (1974) introduced the utility concept to insurance, Pentikäinen (1975) studied the claims process from a dynamic programming viewpoint and Balzer and Benjamin (1981) proposed the use of Control Theory concepts in non-life insurance. A complete account of the recent developments in modelling general insurance activity can be found in Jewell (1980).

1.1 Purpose of the study

Given a portfolio in non-life insurance, each policy can give rise to $(2\kappa+1)$ random variables, where κ itself is a random variable which represents the number of incurred claims for that policy. The other (2κ) random variables are :

X_i $(i = 1, \kappa)$: claim amount for the i^{th} claim.

T_i $(i = 1, \kappa)$: time elapsed between notification to the insurer and settlement of the i^{th} claim.

One criticism that can be made of Risk Theory is that it provides a framework for dealing with the distribution of a random sum of X's without, however, considering at all the T's random variables. Furthermore, the statistical approach to the distribution of the aggregate claims is frequency-based, that is to say, this distribution is estimated by taking into consideration solely past values of X's themselves, without accounting for external factors which could be interfering with the variation of the aggregate sum.

The approach which was used in this research was to study separately the influence of a given set of factors on motor insurance claims and how soon those claims are settled after their notification to the company.

To this end, two independent investigations were carried out : the first aiming at studying the influence of five rating factors on both the average claim and claim frequency, and the second one aiming at estimating an average pattern of settlement delays for motor insurance claims.

1.2 Outline of the study

The research consists of five chapters in addition to this introductory chapter :

Chapter 2 : discussion of methodology for the first investigation.

Chapter 3 : results of the first investigation.

Chapter 4 : discussion of methodology for the second investigation.

Chapter 5 : results of the second investigation.

Chapter 6 : conclusions and suggestions for further research.

CHAPTER TWO

USING REGRESSION MODELS TO MEASURE THE INFLUENCE OF RATING FACTORS ON CLAIMS

2.1 Introduction

When non-life insurance companies calculate their tariffs, they take into account some factors which are thought to influence the amount of claims. These factors are generally known as "rating factors".

The aim of this chapter is to discuss the ways in which statistical methodology, in particular regression analysis, can be used in order to measure the influence of rating factors on motor insurance claims amounts and frequencies.

2.2 Multiple regression models

Multiple regression is a general statistical technique by means of which one can analyse the relationship between a dependent variable and a set of independent or predictor variables. This technique can be regarded either as a descriptive tool by which the linear dependence of one variable on others is summarized and decomposed, or as an inferential tool by which observed relationships in sample data are extended to the underlying population from which the sample was drawn.

2.2.1 Simple bivariate regression

In simple regression analysis, values of the dependent variable are predicted from a linear function of the form :

$$Y' = A + BX \tag{1}$$

where Y' is the estimated value of the dependent variable, X is the independent variable and A and B are constants which are to be estimated.

It is worth noting that B represents the slope of the straight line given by (1) and A represents the intercept of this straight line with the vertical axis.

The difference between the actual and the estimated value of the dependent variable for each observation is called the residual, and may be represented by :

$$R = Y - Y' \tag{2}$$

The constants A and B can be estimated by the least-squares method, which consists of finding A' and B' such that the expression below is minimized.

$$SS_{res} = \sum(Y - Y')^2 \tag{3}$$

The summation above is performed throughout all the observations.

It can be shown that the least-squares estimates of A and B can be written as :

$$B' = \frac{\sum(X - \bar{X})(Y - \bar{Y})}{\sum(X - \bar{X})^2} = \frac{SP_{xy}}{SS_x} \tag{4}$$

$$A' = \bar{Y} - B'\bar{X} \tag{5}$$

The total sum of squares in Y is defined as :

$$SS_y = \sum(Y - \bar{Y})^2 \tag{6}$$

This sum of squares can be partitioned as :

$$SS_y = SS_{reg} + SS_{res}$$

$$\sum(Y - \bar{Y})^2 = \sum(Y' - \bar{Y})^2 + \sum(Y - Y')^2 \tag{7}$$

where SS_{reg} is the part of the total sum of squares due to the regression line (or explained by the regression line) and SS_{res} corresponds to the residual portion of the total sum of squares (not explained by the regression line).

A measure of prediction accuracy of the regression equation can be evaluated by the expression :

$$R^2 = \frac{SS_{reg}}{SS_y} \tag{8}$$

Indeed, the square root of R^2 is the Pearson product-moment correlation between the variables X and Y. If the regression equation fits the observed values SS_{res} will be approximately zero, in which case R^2 will be close to one. On the other hand, if SS_{res} is large in relation to SS_{reg}, R^2 will be close to zero, in which case the fit is poor.

If one rewrites equation (1) in terms of the actual observed value of the dependent variable, an additional error term must be added as below :

$$Y = A + BX + e \tag{9}$$

If this error term is distributed as normal with mean zero and constant variance, and if the errors are themselves uncorrelated, statistical tests can be derived in order to test the significance of the coefficients in the regression. Thus, the significance of the B coefficient can be tested by using the F statistic with one and (N-2) degrees of freedom below :

$$F = \frac{SS_{reg}}{SS_{res}/(N-2)} \tag{10}$$

where N is the number of observations.

Furthermore, the estimate B' of the B coefficient will be itself normally distributed with mean B and variance :

$$Var(B') = \frac{SS_{res}/(N-2)}{SS_x} \tag{11}$$

2.2.2 Multiple regression

The basic principles of regression analysis used in the bivariate case may be extended to situations involving two or more independent variables. The general form of the regression equation is :

$$Y' = A + B_1X_1 + B_2X_2 + \dots + B_kX_k \tag{12}$$

or

$$Y = A + B_1X_1 + B_2X_2 + \dots + B_kX_k + e \tag{13}$$

where Y' represents the estimated value of the dependent variable, Y represents the actual observed value of the dependent variable, $X_1, X_2, \dots, X_k$ are the independent variables, A , $B_1, B_2, \dots, B_k$ are coefficients to be estimated and e is the error term.

The $A, B_1, B_2, \dots, B_k$ coefficients are estimated in such a way that the sum of squared residuals is again minimized.

A convenient way to represent equation (13) is to write it in a matrix form as below :

$$Y = Xb + e \tag{14}$$

where :

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_N \end{bmatrix} \qquad b = \begin{bmatrix} A \\ B_1 \\ B_2 \\ \vdots \\ B_k \end{bmatrix} \qquad e = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_N \end{bmatrix}$$

$$X = \begin{bmatrix} 1 & X_{11} & X_{21} & \cdots & X_{k1} \\ 1 & X_{12} & X_{22} & \cdots & X_{k2} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & X_{1N} & X_{2N} & & X_{kN} \end{bmatrix}$$

In the matrices above, Y_i represents the i^{th} observation of the dependent variable and X_{ji} represents the i^{th} observation of the j^{th} independent variable.

It can be shown that the least-square estimate for the vector of coefficients is :

$$\hat{b} = (X'X)^{-1}X'Y \tag{15}$$

where X' is the transpose matrix of X.

Obviously, expression (15) can only be evaluated if the matrix (X'X) is nonsingular, therefore unique least-squares estimates of the regression coefficients exist only if the inverse matrix of (X'X) exists.

As in the bivariate case, the total sum of squares in Y can be partitioned into two components; therefore the goodness of fit of the regression equation can be evaulated by examining the R^2 value :

$$R^2 = \frac{SS_{reg}}{SS_y} \tag{16}$$

If the same assumptions are made for the error term as in the bivariate case, it can be shown that the significance of all coefficients in the regression can be tested with the following F statistic with k and (N-k-1) degrees of freedom :

$$F = \frac{SS_{reg}/k}{SS_{res}/(N-k-1)} \tag{17}$$

where k is the number of independent variables and N is the number of observations.

It can also be shown that the distribution of the vector of estimates $\hat{b}$ is multinormal with mean b and covariance matrix $(X'X)^{-1}\sigma^2$, where σ^2 represents the variance of each error term.

2.2.3. Regression with dummy variables

Dummy variables are used in regression when an insertion of nominal-scale variables is needed. As most of the rating factors in motor insurance are measured in nominal scales, that is to say, their values are represented by categories rather than by physical quantities, one will consider in this section how their effect can be measured in a quantitative variable.

Any nominal variable taking k values can be represented by a set of (k-1) dummy variables, each one taking only two possible values,zero or one. For example, the rating factor vehicle make with say four categories : "make A", "make B", "make C" and "other" , can be represented by three dummy variables D1, D2 and D3 as below :

	D1	D2	D3
Make A	1	0	0
Make B	0	1	0
Make C	0	0	1
Other	0	0	0

One of the categories must always be taken as a reference. In the above example, the fourth category was arbitrarily chosen as reference, which means that all the three dummy variables take the value zero for that category.

The regression equation can be written in terms of the dummy variables as

$$Y' = A + B_1D_1 + B_2D_2 + B_3D_3$$

For the observations belonging to the fourth category, the predicted value of the dependent variable would be given by

$$Y' = A$$

since all three dummy variables are equal to zero for such observations.

For "make A", the predicted value of Y would be :

$$Y' = A + B_1$$

since $D_1 = 1$, $D_2 = 0$ and $D_3 = 0$ for this category.

From the above expression, it can be noted that the regression coefficient B_1 is the difference in predicted Y for observations which are classified as "make A" as compared to those which are classified as "other".

Similarly, the predicted value of Y for "make B" is :

$$Y' = A + B_2$$

and for "make C" is :

$$Y' = A + B_3$$

The regression coefficients may be evaluated by the least-squares method as shown in the previous sections.

CHAPTER THREE

THE INFLUENCE OF RATING FACTORS ON AVERAGE CLAIM AND CLAIM FREQUENCY

3.1 Introduction

This chapter aims at studying the influence of certain rating factors on average claim and claim frequency in third party motor insurance.

To this end, a set of data from Sweden was analysed with the aid of two well-known statistical computer softwares : BMDP and SPSS.

Before any attempt was made to fit statistical models to the data, an exploratory analysis was performed for each of the two variables of interest, with the following objectives :

1. Disclosing possible errors and unusual values.
2. Studying the distribution of the variable.
3. Searching for evidence of the existence of a relationship between the rating factors and the variable.

When significant evidence was found that a relationship existed, a confirmatory analysis followed, in which a linear model was fitted to the observations, in order to quantify the effects of the rating factors on the variable of interest.

3.2 Description of the data

The data which will be analysed consists of third party liability claims incurred during the year of 1977 in Sweden (the whole country). The payments also include deposits for future settlements of claims incurred but not closed in that year.

Each observation is a combination of five rating factors, for which three variables are observed :

Name of the variable	Notation (for computer use)
Number of claims	NCLA
Number of insured years (exposure)	ESP
Sum of payments (in Swedish Crowns)	SP

The rating factors are :

1. Mileage - the average number of kilometers the insured drives in a year

Notation		Codes	Meaning
S	=	1	less than 10,000 km/year
		2	from 10,000 to 15,000 km/year
		3	from 15,000 to 20,000 km/year
		4	from 20,000 to 25,000 km/year
		5	more than 25,000 km/year

2. Zone of garage - According to the insured's home address.

Notation		Codes	Meaning
Z	=	1	Stockholm, Göteborg, Malmö with surroundings
		2	Other bigger cities with surroundings
		3	Smaller cities with surroundings in southern Sweden
		4	Rural areas in southern Sweden
		5	Smaller cities with surroundings in northern Sweden
		6	Rural areas in northern Sweden
		7	Gotland

3. No claim bonus - The insured starts at B = 1. Every year with no claim he is moved one class, with one exception : to be moved from 6 to 7 requires six years of no claims. The premium is p = a+b.c, where :

Notation		Codes	Meaning
B	=	1	b = 1.0
		2	b = 0.8
		3	b = 0.7
		4	b = 0.6
		5	b = 0.5
		6	b = 0.4
		7	b = 0.25

4. Vehicle make - Each code corresponds to one model, with the exception of code 9 which aggregates all models different from models 1 to 8.

Notation		Codes	Meaning	Relative Engine Size (Ref : M=4)
M	=	1	Volvo 142 - 144	2.20
		2	Volvo 145	2.40
		3	Volvo 242 - 244	2.43
		4	VW 1200	1.00
		5	Opel Rekord 1900	2.23
		6	Saab 96 V4	1.60
		7	Saab 99	2.17
		8	Mercedes Benz 220 0/8	1.43
		9	All others	

5. Vehicle Age - the age of the insured's vehicle in years.

Notation		Codes	Meaning
J	=	1	Less than 3 years old
		2	From 3 to 8 years old
		3	More than 8 years old

The total number of combinations of the above rating factors is $5 \times 7 \times 7 \times 9 \times 3 = 6615$; however, only 5413 observations are present in the data file due to the fact that the records with null exposure were excluded.

To clarify the structure of the data, an example of some observations from the actual file is given below :

S	Z	B	M	J	NCLA	ESP	SP
1	1	1	2	2	11	52.95936	22047
1	1	1	5	3	26	117.76117	96387
1	1	1	6	1	1	12.83534	3663
1	1	2	1	3	20	133.90332	57468
1	1	2	4	2	0	7.26768	0

For reference purposes, each combination of values of the five rating factors will be called a cell and the associated values of the variables will be called observations for that cell. Thus the fourth row above corresponds to a cell in which the rating factors are S = 1 (mileage less than 10,000 km/year), Z = 1 (insured's vehicle garaged in Stockholm, Göteborg or Malmö, including surroundings), B = 2 (level of no claim bonus corresponding to b = 0.8), M = 1 (model of the insured's vehicle : Volvo 142 - 144) and J = 3 (insured's vehicle more than 8 years old). The associated observations for this cell are : number of claims (NCLA) = 20, exposure (ESP) = 133.90332 and sum of payments (SP) = 57468.

3.3 The distribution of the average claim

The first variable to be analysed will be the average claim, which is defined as follows :

Notation		Meaning	
AC	=	SP/NCLA	; average claim

Using the procedure 2D of BMDP, the condensed distribution of AC was produced and the result is shown in Fig. (3.1). For convenience, the values of AC were rounded to the third digit.

There are 1974 values which were not counted, which means that there are many observations with NCLA = 0 and therefore with SP = 0 (zero claims).

The shape of the distribution resembles the lognormal, its mean being 5156.15 (only positive claims considered) and standard deviation 4982.61. There is a noticeable discontinuity in the tail of the distribution, where the value 31000 occurs 71 times, well apart from the next greatest value which is 23000 .

Using the procedure LIST CASES of SPSS, one can look more closely at these extreme observations in Figs. (3.2) and (3.3), and it becomes clear that in reality all the claims in that sample have exactly the same value, 31442. This seems to be no coincidence and the most likely explanation is that this particular value was assigned as an estimate for future settlement, rather than a value which has actually occurred as a real payment.

Even if these claims had not been estimated, their abnormal magnitude would not justify their analysis in a time basis as short as one year, and therefore they will be ignored throughout the study.

Using the procedure CONDESCRIPTIVE of SPSS successively for the whole file and for the sample in question, one can evaluate the percentage of payments to be ignored. To this end, a new variable SPN will be defined by dividing the values of SP by 1000

FIG (3.1) .

CONDENSED DISTRIBUTION OF AC

```
************
* AC       *
************
                                              MAXIMUM      31000.0000000     H
VARIABLE NUMBER . . . . . .          13       MINIMUM          0.            HHHH
NUMBER OF DISTINCT VALUES .          24       RANGE        31000.0000000     HHHH
NUMBER OF VALUES COUNTED. .        3439       VARIANCE 24826395.5000000     HHHHH                                         EACH 'H'
NUMBER OF VALUES NOT COUNTED       1974       ST.DEV.       4982.6093140    HHHHHH                                       REPRESENTS
***VALUES ARE ROUNDED TO. .1000.0000          (Q3-Q1)/2     2000.0000000    HHHHHH                                           58
                                                                            HHHHHH                                        COUNT(S)
LOCATION ESTIMATES                                 ST.ERROR                 HHHHHHHH
                   MEAN          5156.1500244      84.9651337               HHHHHHHHH                   H
                   MEDIAN        4000.0000000       0.                      HHHHHHHHHHHHHHHHHHH HHH     H
                   MODE          2000.0000000                              L-----------------------------------U

                                                                                   EACH '-' ABOVE =     1000.0000
                                                                                                 L=        0.
 521 CASES TOO MANY TO COMPUTE THE OPTIONAL ESTIMATES                                            U=    37000.0000

                                                                                   EACH '.' BELOW =      250.0000
S       Q         M   M        Q              S
M       M         M   M        3              +                                                                      M
I.......O.......E...E..................................................................................................A
N       D         D   A                                                                                                X
        E         I   N
```

VALUE	COUNT	PERCENTS CELL	PERCENTS CUM	VALUE	COUNT	PERCENTS CELL	PERCENTS CUM	VALUE	COUNT	PERCENTS CELL	PERCENTS CUM	VALUE	COUNT	PERCENTS CELL	PERCENTS CUM
0.	32	0.9	0.9	6000.	304	8.8	80.0	12000.	46	1.3	94.6	18000.	10	0.3	97.5
1000.	331	9.6	10.6	7000.	169	4.9	84.9	13000.	12	0.3	94.9	19000.	5	0.1	97.6
2000.	573	16.7	27.2	8000.	119	3.5	88.4	14000.	22	0.6	95.6	21000.	4	0.1	97.7
3000.	522	15.2	42.4	9000.	78	2.3	90.6	15000.	6	0.2	95.8	22000.	5	0.1	97.9
4000.	485	14.1	56.5	10000.	50	1.5	92.1	16000.	27	0.8	96.5	23000.	2	0.1	97.9
5000.	504	14.7	71.2	11000.	40	1.2	93.3	17000.	22	0.6	97.2	31000.	71	2.1	100.0

-NUMBER OF WORDS OF STORAGE USED 6051

FIG (3.2)

UNUSUAL VALUES OF AC

CASE-N	S	Z	B	M	J	ESP	NCLA	SP	AC
1	1.	1.	6.	6.	1.	65.604	1.	31442.	31442.000
2	1.	3.	3.	8.	2.	7.971	1.	31442.	31442.000
3	1.	5.	2.	6.	1.	9.649	1.	31442.	31442.000
4	1.	5.	3.	8.	2.	2.995	1.	31442.	31442.000
5	1.	6.	1.	6.	1.	23.752	1.	31442.	31442.000
6	1.	6.	1.	8.	3.	9.014	1.	31442.	31442.000
7	1.	6.	2.	7.	3.	21.487	1.	31442.	31442.000
8	1.	6.	5.	4.	3.	258.547	1.	31442.	31442.000
9	1.	7.	7.	6.	1.	36.141	1.	31442.	31442.000
10	2.	1.	2.	8.	2.	9.187	1.	31442.	31442.000
11	2.	1.	3.	6.	1.	26.096	2.	62884.	31442.000
12	2.	2.	1.	5.	1.	11.033	1.	31442.	31442.000
13	2.	3.	2.	3.	1.	44.368	1.	31442.	31442.000
14	2.	3.	2.	8.	2.	17.422	1.	31442.	31442.000
15	2.	4.	5.	5.	1.	58.978	1.	31442.	31442.000
16	2.	4.	7.	3.	2.	9.897	1.	31442.	31442.000
17	2.	5.	5.	7.	3.	8.814	1.	31442.	31442.000
18	2.	6.	2.	5.	1.	3.892	1.	31442.	31442.000
19	2.	6.	6.	8.	1.	11.543	1.	31442.	31442.000
20	2.	7.	2.	5.	2.	4.225	1.	31442.	31442.000
21	2.	7.	3.	6.	3.	7.919	1.	31442.	31442.000
22	2.	7.	7.	6.	2.	101.533	1.	31442.	31442.000
23	3.	1.	1.	2.	3.	6.978	1.	31442.	31442.000
24	3.	1.	6.	8.	2.	45.745	2.	62884.	31442.000
25	3.	2.	7.	8.	1.	71.209	1.	31442.	31442.000
26	3.	3.	1.	8.	2.	8.568	1.	31442.	31442.000
27	3.	3.	2.	2.	3.	6.658	1.	31442.	31442.000
28	3.	3.	3.	2.	3.	12.832	1.	31442.	31442.000
29	3.	3.	3.	6.	1.	26.755	1.	31442.	31442.000
30	3.	3.	4.	8.	3.	10.602	1.	31442.	31442.000
31	3.	3.	7.	2.	1.	0.442	1.	31442.	31442.000
32	3.	4.	2.	6.	1.	41.401	2.	62884.	31442.000
33	3.	5.	1.	4.	3.	12.545	1.	31442.	31442.000
34	3.	5.	5.	1.	3.	53.252	2.	62884.	31442.000
35	3.	5.	5.	8.	2.	22.970	1.	31442.	31442.000
36	3.	5.	7.	8.	1.	40.450	1.	31442.	31442.000
37	3.	6.	2.	6.	1.	20.634	1.	31442.	31442.000
38	3.	6.	2.	6.	3.	73.419	1.	31442.	31442.000
39	3.	7.	1.	9.	3.	44.118	1.	31442.	31442.000
40	3.	7.	2.	4.	3.	5.705	1.	31442.	31442.000
41	3.	7.	2.	5.	3.	2.177	1.	31442.	31442.000
42	3.	7.	4.	9.	3.	43.804	1.	31442.	31442.000
43	3.	7.	7.	3.	1.	29.986	1.	31442.	31442.000
44	4.	1.	2.	6.	3.	11.114	1.	31442.	31442.000
45	4.	1.	5.	5.	1.	10.234	1.	31442.	31442.000
46	4.	2.	4.	7.	2.	30.964	1.	31442.	31442.000

FIG (3.3)

UNUSUAL VALUES OF AC (CONTINUED) AND DESCRIPTIVE STATISTICS OF SPN

CASE-N	S	Z	B	M	J	ESP	NCLA	SP	AC
47	4.	3.	2.	8.	3.	3.683	1.	31442.	31442.000
48	4.	3.	3.	3.	1.	23.839	1.	31442.	31442.000
49	4.	4.	3.	8.	1.	3.074	1.	31442.	31442.000
50	4.	4.	6.	2.	1.	1.550	1.	31442.	31442.000
51	4.	5.	4.	2.	2.	7.258	1.	31442.	31442.000
52	4.	5.	5.	3.	1.	9.496	1.	31442.	31442.000
53	4.	5.	6.	7.	2.	22.672	1.	31442.	31442.000
54	4.	5.	6.	8.	3.	3.785	1.	31442.	31442.000
55	4.	6.	2.	7.	2.	9.293	1.	31442.	31442.000
56	4.	6.	7.	4.	3.	35.586	1.	31442.	31442.000
57	5.	1.	4.	1.	3.	12.323	1.	31442.	31442.000
58	5.	1.	5.	7.	3.	2.836	1.	31442.	31442.000
59	5.	1.	7.	5.	2.	58.835	1.	31442.	31442.000
60	5.	2.	4.	6.	1.	9.099	1.	31442.	31442.000
61	5.	2.	5.	5.	2.	8.319	1.	31442.	31442.000
62	5.	3.	1.	5.	3.	2.880	1.	31442.	31442.000
63	5.	4.	4.	3.	1.	47.932	1.	31442.	31442.000
64	5.	5.	1.	8.	2.	1.824	1.	31442.	31442.000
65	5.	5.	7.	6.	1.	45.376	1.	31442.	31442.000
66	5.	6.	1.	3.	1.	3.216	1.	31442.	31442.000
67	5.	6.	3.	5.	3.	2.261	1.	31442.	31442.000
68	5.	6.	4.	5.	2.	1.370	1.	31442.	31442.000
69	5.	6.	5.	5.	2.	6.809	1.	31442.	31442.000
70	5.	7.	6.	9.	3.	12.009	1.	31442.	31442.000
71	5.	7.	7.	1.	2.	37.254	1.	31442.	31442.000

VARIABLE SPN

MEAN	103.626	STD ERROR	5.496	STD DEV	404.359
VARIANCE	163505.941	KURTOSIS	96.991	SKEWNESS	8.701
RANGE	7331.117	MINIMUM	0.	MAXIMUM	7331.117
SUM	560925.524				

VALID OBSERVATIONS - 5413 MISSING OBSERVATIONS - 0

VARIABLE SPN

MEAN	33.213	STD ERROR	0.867	STD DEV	7.301
VARIANCE	53.309	KURTOSIS	13.849	SKEWNESS	3.932
RANGE	31.442	MINIMUM	31.442	MAXIMUM	62.884
SUM	2358.150				

VALID OBSERVATIONS - 71 MISSING OBSERVATIONS - 0

(otherwise the desired results will not be printed by the software due to space limitations), and the result of the procedure is given in Fig. (3.3)(below). The ratio is :

$$\frac{\text{SPN (sample)}}{\text{SPN (whole file)}} \times 100 = \frac{2358.150}{560925.524} \times 100 = 0.42\% ,$$

which is reasonably small and thus justifies the exclusion.

For the next step of the analysis, one will consider a reduced set of observations, that is to say, the above mentioned observations will be removed and also, for the time being, the observations with SP = 0 (zero claims), which will be considered later in the development of the study.

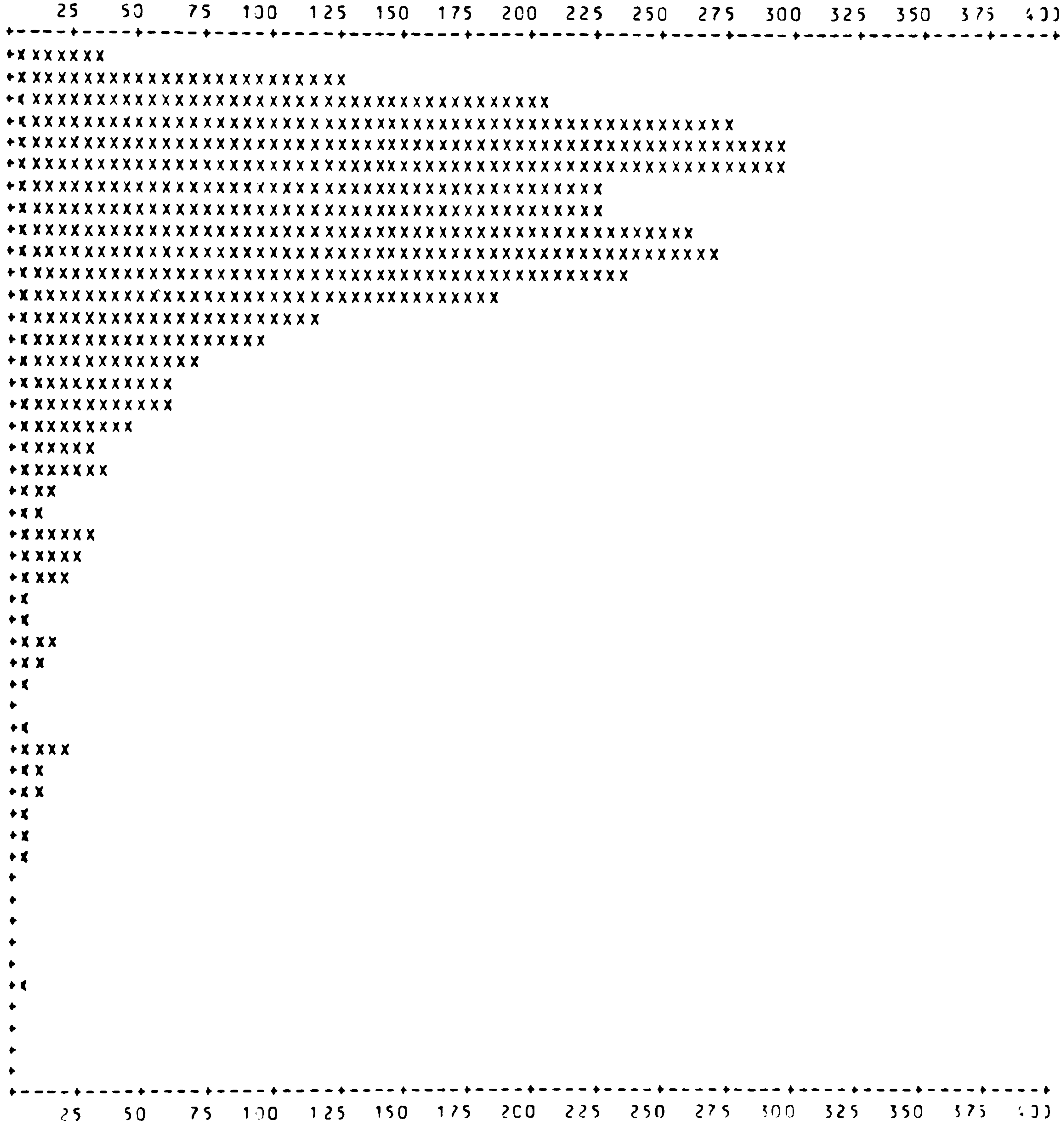
Using the procedure 5D of BMDP, the histogram of AC may be obtained and is shown in Fig. (3.4). The intervals were made 500 units wide and their upper limits are printed on the base of the histogram. A scale factor of 1 : 5 (each "X" = 5 counts) was used for convenience. The number of observations has been reduced to 3368 (71 abnormal observations plus 1974 zero claims were removed from the 5413 original number), and the mean and standard deviation are now 4609.24 and 3300.72 respectively, a reasonable reduction from the previous values of the original distribution. The lognormality shape appears to be stressed, as the picture of the distribution seems to suggest.

In order to check the hypothesis of AC being lognormally distributed, one will perform a naperian logarithmic transformation on AC, as defined below :

FIG (3.4)

HISTOGRAM OF AC

SYMBOL	COUNT	MEAN	ST.DEV.
X	3368	4609.240	3300.719



INTERVAL NAME	FREQUENCY INT.	FREQUENCY CUM.	PERCENTAGE INT.	PERCENTAGE CUM.
500.00	33	33	1.0	1.0
1000.0	127	160	3.8	4.8
1500.0	206	366	6.1	10.9
2000.0	276	642	8.2	19.1
2500.0	295	937	8.8	27.8
3000.0	294	1231	8.7	36.5
3500.0	227	1458	6.7	43.3
4000.0	224	1682	6.7	49.9
4500.0	261	1943	7.7	57.7
5000.0	268	2211	8.0	65.6
5500.0	237	2448	7.0	72.7
6000.0	186	2634	5.5	78.2
6500.0	117	2751	3.5	81.7
7000.0	97	2848	2.9	84.6
7500.0	72	2920	2.1	86.7
8000.0	58	2978	1.7	88.4
8500.0	61	3039	1.8	90.2
9000.0	46	3085	1.4	91.6
9500.0	32	3117	1.0	92.5
10000.	35	3152	1.0	93.6
10500.	15	3167	0.4	94.0
11000.	11	3178	0.3	94.4
11500.	29	3207	0.9	95.2
12000.	27	3234	0.8	96.0
12500.	19	3253	0.6	96.6
13000.	6	3259	0.2	96.8
13500.	6	3265	0.2	96.9
14000.	14	3279	0.4	97.4
14500.	8	3287	0.2	97.6
15000.	4	3291	0.1	97.7
15500.	2	3293	0.1	97.8
16000.	6	3299	0.2	98.0
16500.	21	3320	0.6	98.6
17000.	10	3330	0.3	98.9
17500.	12	3342	0.4	99.2
18000.	6	3348	0.2	99.4
18500.	4	3352	0.1	99.5
19000.	3	3355	0.1	99.6
19500.	2	3357	0.1	99.7
20000.	0	3357	0.	99.7
20500.	0	3357	0.	99.7
21000.	2	3359	0.1	99.7
21500.	2	3361	0.1	99.8
22000.	5	3366	0.1	99.9
22500.	0	3366	0.	99.9
23000.	1	3367	0.0	100.0
23500.	1	3368	0.0	100.0
LAST	0	3368	0.	100.0

Notation		Meaning
LAC	=	$\ell n(AC)$; Naperian logarithm of AC,

and its condensed distribution will be requested using the procedure 2D of BMDP once more. The result is shown in Fig. (3.5), where the values were rounded to units of 0.25 for convenience.

Apart from a quite long left tail, the distribution can be considered reasonably symmetric. The mean, median and mode are close to each other (8.20, 8.25 and 8.50 respectively) and the heavier density in the left hand side may well be admitted as due to random fluctuation.

To check the normality assumption in a more appropriate way, the procedure 5D of BMDP will be used in order to produce a normal plot for LAC. The result is shown in Fig. (3.6), where a rough linear trend appears to exist. This is, of course, no definitive confirmation of normality, and a more detailed histogram for LAC is needed to provide for a more careful visual inspection.

This will be done by using the procedure 5D of BMDP in its histogram version and the result is shown in Fig. (3.7) . The chosen intervals are 0.125 wide and the values which appear at the base of the histogram are upper limits of those intervals. Each "X" represents five counts, thus the left tail is not represented in the diagram, although the correct frequencies appear in the frequency column opposite to the base of the histogram. In Fig. (3.8) the same histogram is reproduced but now with a scale

FIG (3.5)

CONDENSED DISTRIBUTION OF LAC

```
************
* LAC      *
************
                                                MAXIMUM          10.0000000
VARIABLE NUMBER . . . . . .          12         MINIMUM           4.0000000
NUMBER OF DISTINCT VALUES .          24         RANGE             6.0000000
NUMBER OF VALUES COUNTED. .        3368         VARIANCE          0.5251357
NUMBER OF VALUES NOT COUNTED          0         ST.DEV.           0.7246625
***VALUES ARE ROUNDED TO. .      0.2500         (Q3-Q1)/2         0.5000000

LOCATION ESTIMATES                                   ST.ERROR
                  MEAN              8.1995991        0.0124868
                  MEDIAN            8.2500000        0.
                  MODE              8.5000000

                                                H
                                                H
                                                H
                                              H HHH            EACH 'H'
                                             HH HHH            REPRESENTS
                                            HHH HHH               64
                                            HHH HHHH           COUNT(S)
                                           HHHH HHHH
                                         H HHHH HHHH HHH
                      H H HH HHHH HHHH HHHH HHHH HHHH.
                     L-----------------------------------U

                                 EACH '-' ABOVE =        0.2000
                                              L=         3.6000
                                              U=        10.6000
                                                     Q1=      7.7500000
                            VALUE   VALUE/S.E.       Q3=      8.7500000
             SKEWNESS       -0.54       -12.69       S-=      7.4749357
             KURTOSIS        1.16        13.76       S+=      8.9242617

                                 EACH '.' BELOW =        0.0500
             S    Q                    Q S
 M           -    1     M M    M      3 +                         M
 I.......................................E.E....O......................A
 N                      A D    D                                  X
                        N I    E

                PERCENTS                    PERCENTS                    PERCENTS                    PERCENTS
 VALUE   COUNT  CELL  CUM   VALUE   COUNT  CELL  CUM   VALUE   COUNT  CELL  CUM   VALUE    COUNT  CELL  CUM
 4.00000     1  0.0   0.0   5.75000     7  0.2   0.5   7.25000   161   4.8  12.1   8.75000    395  11.7  85.3
 4.25000     1  0.0   0.1   6.00000    13  0.4   0.9   7.50000   259   7.7  19.7   9.00000    228   6.8  92.1
 4.75000     1  0.0   0.1   6.25000    14  0.4   1.3   7.75000   354  10.5  30.3   9.25000    121   3.6  95.7
 5.00000     1  0.0   0.1   6.50000    41  1.2   2.5   8.00000   390  11.6  41.8   9.50000     69   2.0  97.7
 5.25000     2  0.1   0.2   6.75000    60  1.8   4.3   8.25000   432  12.8  54.7   9.75000     65   1.9  99.7
 5.50000     3  0.1   0.3   7.00000   101  3.0   7.3   8.50000   638  18.9  73.6  10.00000     11   0.3 100.0
------------------------------------------------------------------------------------------------------------------
-NUMBER OF WORDS OF STORAGE USED      3841
```

FIG (3.6)

NORMAL PLOT FOR LAC

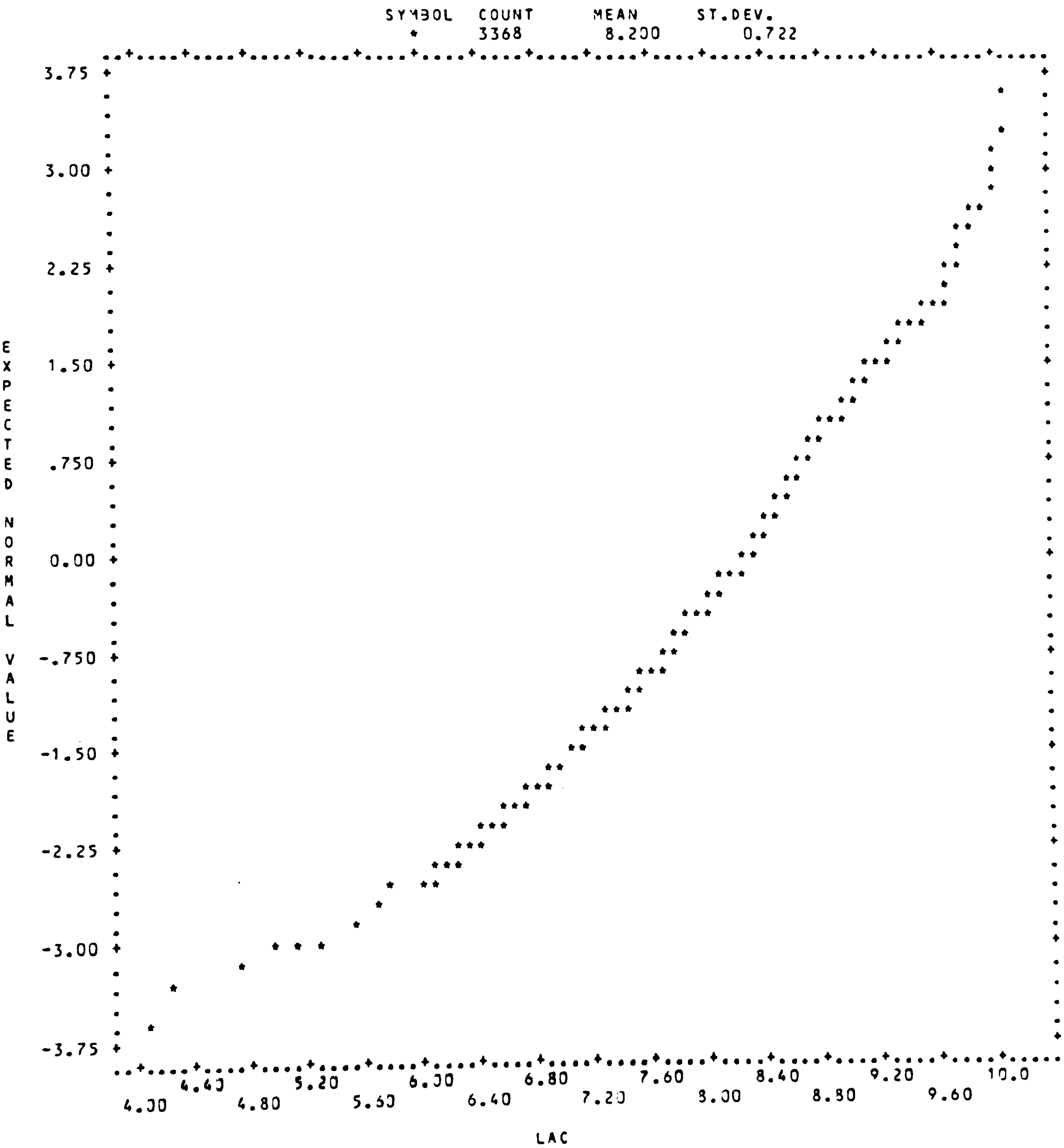


FIG (3.7)

HISTOGRAM OF LAC (SCALE 1 : 5)

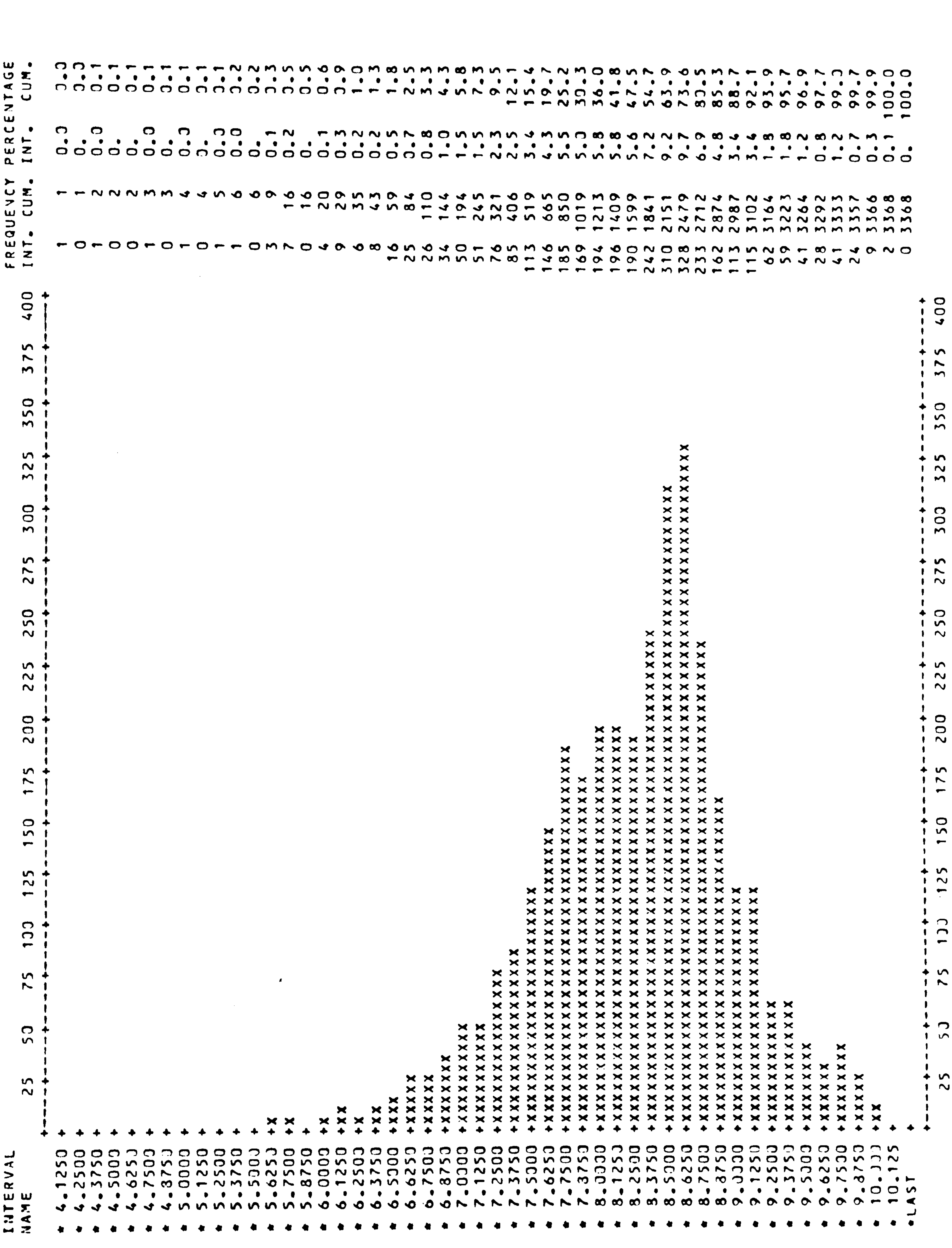
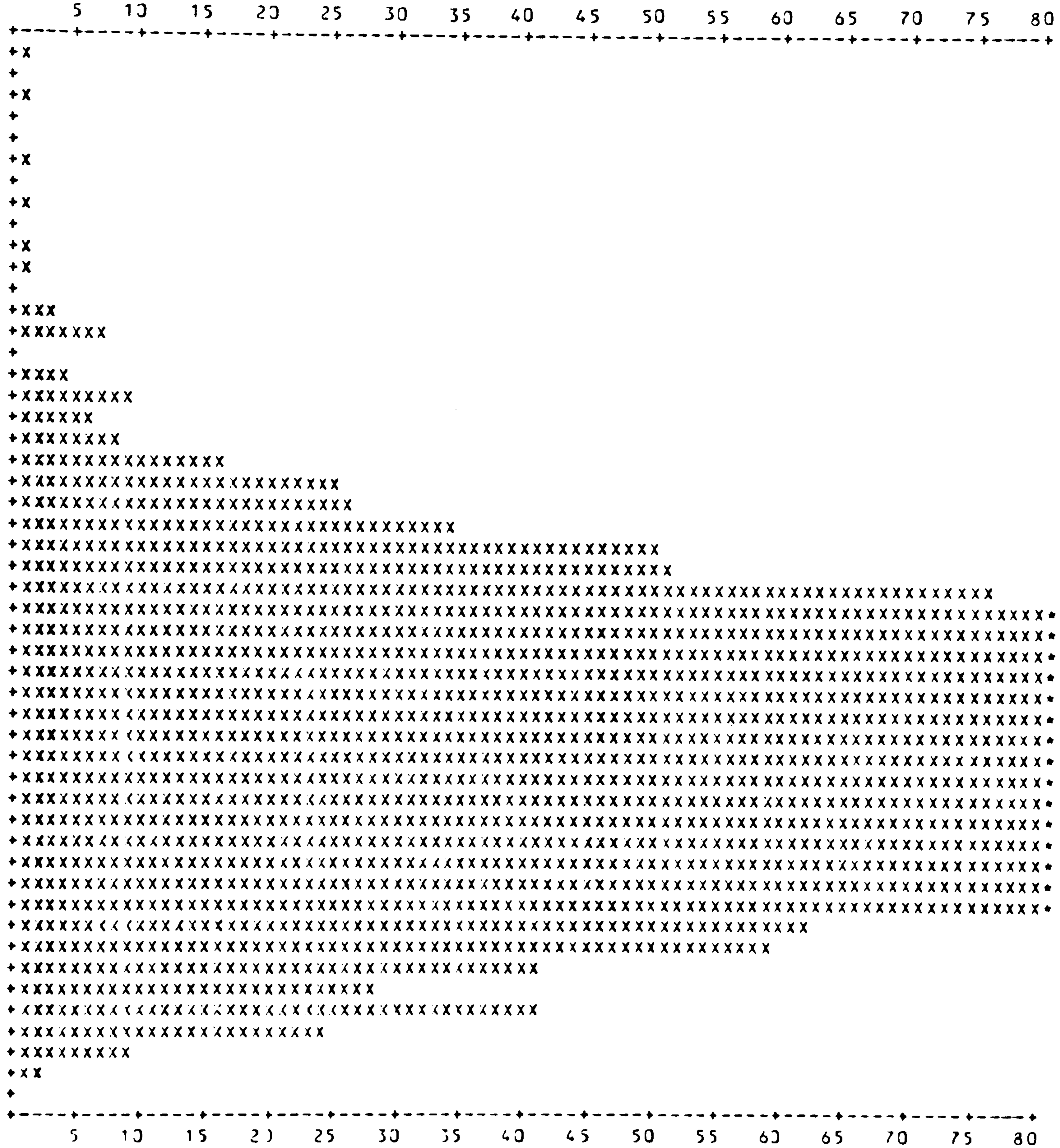


FIG (3.8)

HISTOGRAM OF LAC (SCALE 1 : 1)

SYMBOL	COUNT	MEAN	ST.DEV.
X	3363	8.200	0.722



INTERVAL NAME	FREQUENCY INT.	FREQUENCY CUM.	PERCENTAGE INT.	PERCENTAGE CUM.
* 4.1250	1	1	0.0	0.0
* 4.2500	0	1	0.	0.0
* 4.3750	1	2	0.0	0.1
* 4.5000	0	2	0.	0.1
* 4.6250	0	2	0.	0.1
* 4.7500	1	3	0.0	0.1
* 4.8750	0	3	0.	0.1
* 5.0000	1	4	0.0	0.1
* 5.1250	0	4	0.	0.1
* 5.2500	1	5	0.0	0.1
* 5.3750	1	6	0.0	0.2
* 5.5000	0	6	0.	0.2
* 5.6250	3	9	0.1	0.3
* 5.7500	7	16	0.2	0.5
* 5.8750	0	16	0.	0.5
* 6.0000	4	20	0.1	0.6
* 6.1250	9	29	0.3	0.9
* 6.2500	6	35	0.2	1.0
* 6.3750	8	43	0.2	1.3
* 6.5000	16	59	0.5	1.8
* 6.6250	25	84	0.7	2.5
* 6.7500	26	110	0.8	3.3
* 6.8750	34	144	1.0	4.3
* 7.0000	50	194	1.5	5.8
* 7.1250	51	245	1.5	7.3
* 7.2500	76	321	2.3	9.5
* 7.3750	85	406	2.5	12.1
* 7.5000	113	519	3.4	15.4
* 7.6250	146	665	4.3	19.7
* 7.7500	185	850	5.5	25.2
* 7.8750	169	1019	5.0	30.3
* 8.0000	194	1213	5.8	36.0
* 8.1250	196	1409	5.8	41.8
* 8.2500	190	1599	5.6	47.5
* 8.3750	242	1841	7.2	54.7
* 8.5000	310	2151	9.2	63.9
* 8.6250	328	2479	9.7	73.6
* 8.7500	233	2712	6.9	80.5
* 8.8750	162	2874	4.8	85.3
* 9.0000	113	2987	3.4	88.7
* 9.1250	115	3102	3.4	92.1
* 9.2500	62	3164	1.8	93.9
* 9.3750	59	3223	1.8	95.7
* 9.5000	41	3264	1.2	96.9
* 9.6250	28	3292	0.8	97.7
* 9.7500	41	3333	1.2	99.0
* 9.8750	24	3357	0.7	99.7
* 10.000	9	3366	0.3	99.9
* 10.125	2	3368	0.1	100.0
*LAST	0	3368	0.	100.0

factor of one (each "X" represents one observation), so that one can have a correct picture of the tails of the distribution.

Judging by the shape of the histograms, it does not seem unrealistic to accept the normality assumption for LAC (and therefore the lognormality assumption for AC), even though the left hand side of the distribution is a little fatter and the left tail longer than the right one. At least there is no strong evidence to reject this hypothesis, and one will stick to the previous conclusion that these abnormalities are due to random fluctuation.

The traditional one sample non-parametric tests (chi-square and Kolmogorov-Smirnov) will not be used in this context due to the fact that the sample size being very large (3368), will distort the associated probability values (p-values) leading to almost certain rejection of any hypothesis to be tested.

3.4 Relationship between LAC and NCLA

It will be necessary to observe the behaviour of the distribution of LAC when NCLA varies. To this end, the procedure BREAKDOWN of SPSS will be requested in order to produce the means and standard deviations of LAC within groups defined by different values of NCLA, ranging from 1 to 40. The result is given in Fig. (3.9), where it can be observed that there is a mean of 7.62 for NCLA = 1, contrasting with the means within the remaining groups, ranging from 8.07 to 8.59.

FIG (3.9)

MEANS AND STANDARD DEVIATIONS OF LAC

FOR THE FIRST 40 VALUES OF NCLA

- D E S C R I P T I O N   O F   S U B P O P U L A T I O N S - - - - - - - - - - - - - - - - - -

CRITERION VARIABLE LAC
BROKEN DOWN BY NCLA

VARIABLE	CODE	VALUE LABEL	SUM	MEAN	STD DEV	VARIANCE	N
FOR ENTIRE POPULATION			23085.0685	8.1458	0.7713	0.5950	(2834)
NCLA	1.		4754.8117	7.6199	0.8174	0.6682	(624)
NCLA	2.		3146.7310	8.0687	0.8612	0.7416	(390)
NCLA	3.		2195.8752	8.2243	0.8887	0.7898	(267)
NCLA	4.		1702.6375	8.2652	0.7705	0.5937	(206)
NCLA	5.		1301.5230	8.2900	0.7118	0.5066	(157)
NCLA	6.		1129.2411	8.3032	0.6523	0.4255	(136)
NCLA	7.		966.6987	8.4061	0.6329	0.4005	(115)
NCLA	8.		741.0530	8.3264	0.6148	0.3780	(89)
NCLA	9.		647.2150	8.2976	0.5867	0.3442	(78)
NCLA	10.		512.1996	8.3967	0.5791	0.3354	(61)
NCLA	11.		577.2445	8.3659	0.5712	0.3263	(69)
NCLA	12.		413.3438	8.4357	0.5294	0.2802	(49)
NCLA	13.		357.5825	8.5139	0.5040	0.2540	(42)
NCLA	14.		415.9626	8.3193	0.5371	0.2885	(50)
NCLA	15.		352.5981	8.3952	0.4929	0.2429	(42)
NCLA	16.		194.4308	8.4535	0.3279	0.1075	(23)
NCLA	17.		292.0910	8.3455	0.4636	0.2149	(35)
NCLA	18.		175.4240	8.3535	0.4214	0.1776	(21)
NCLA	19.		185.0499	8.4114	0.4141	0.1715	(22)
NCLA	20.		252.9213	8.4307	0.3939	0.1552	(30)
NCLA	21.		219.5959	8.4460	0.3763	0.1416	(26)
NCLA	22.		246.1804	8.4890	0.3202	0.1025	(29)
NCLA	23.		177.1521	8.4358	0.3827	0.1464	(21)
NCLA	24.		174.3792	8.3038	0.4355	0.1896	(21)
NCLA	25.		176.1918	8.3901	0.4331	0.1875	(21)
NCLA	26.		169.4234	8.4712	0.3919	0.1536	(20)
NCLA	27.		194.7081	8.4656	0.2802	0.0785	(23)
NCLA	28.		144.6810	8.5106	0.2971	0.0883	(17)
NCLA	29.		170.8752	8.5438	0.2979	0.0888	(20)
NCLA	30.		118.4599	8.4614	0.3202	0.1025	(14)
NCLA	31.		145.5964	8.5645	0.2959	0.0876	(17)
NCLA	32.		67.2623	8.4078	0.3666	0.1344	(8)
NCLA	33.		74.9891	8.3321	0.3090	0.0955	(9)
NCLA	34.		92.3976	8.4000	0.2595	0.0673	(11)
NCLA	35.		117.3257	8.3804	0.2390	0.0571	(14)
NCLA	36.		109.9458	8.4574	0.2946	0.0868	(13)
NCLA	37.		24.3431	8.1144	0.3018	0.0911	(3)
NCLA	38.		127.1809	8.4787	0.3861	0.1491	(15)
NCLA	39.		90.8089	8.2554	0.2728	0.0744	(11)
NCLA	40.		128.3788	8.5919	0.2395	0.0574	(15)

It is worth remembering that the observations for LAC with NCLA = 1 are also observations of the distribution of a single claim, which cannot be fully obtained due to the aggregated form in which the data was collected.

Regarding the standard deviations, it can be noticed that they decrease when NCLA increases, as expected. Indeed, it is reasonable that an average for LAC based on a large number of observations will be less influenced by unusual extreme values, as shown by the data.

It is expected that for large values of NCLA, a stabilization on both the mean and the standard deviation might happen. To check this assumption, the observations will be grouped into three categories :

A. small NCLA $(0 < \text{NCLA} \leq 10)$

B. moderate NCLA $(11 < \text{NCLA} \leq 100)$

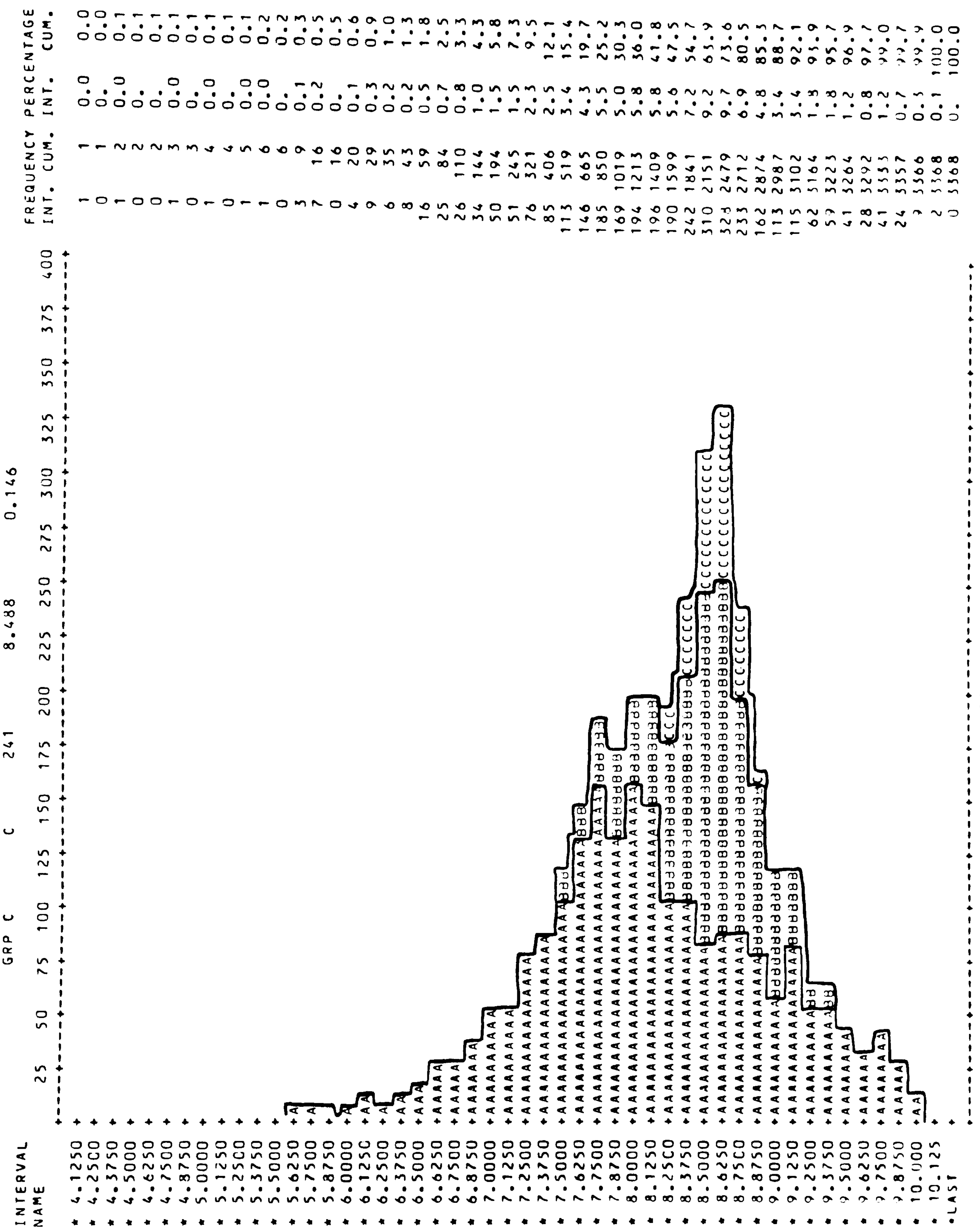
C. large NCLA $(\text{NCLA} > 100)$,

and the histogram of LAC will be re-examined. In Fig. (3.10), one can see the overall histogram in which each group above is represented by the corresponding letters A, B and C. The concentration of the observations in the middle of the diagram as NCLA increases is clear. Dividing lines were drawn on the picture in order to make clearer the separation of the groups. The scale which was used is 1 : 5 (each symbol = 5 counts).

FIG (3.10)

HISTOGRAM OF LAC FOR SMALL (A), MODERATE (B)

AND LARGE (C) VALUES OF NCLA



3.5 Influence of rating factors on LAC

The assessment of the influence of the rating factors on LAC (and therefore on AC),will be made by analysing and comparing the histograms of LAC for each level of each of the five rating factors.

To this end the procedure 7D of BMDP will be used and the result for the rating factor S is shown in Fig. (3.11). Each histogram corresponds to one level of S and the identification of these levels are printed on the top line, above each histogram. The vertical axis is the common base of the histograms and the intervals for LAC are identified by their midpoints.

Each asterisk represents one observation. There are lines however with too many observations in which case the program prints the actual number of observations in that interval.

The mean, standard deviation, standard error of the mean, maximum, minimum and number of observations (sample size) are evaluated for each level of S and are printed just below the corresponding histogram. These statistics are also evaluated for the whole sample and the results appear on the bottom left hand corner of the printout.

Although the program prints an analysis of variance table (ANOVA), one will not attempt to fully interpret its results due to the fact that a substantial number of cases have been omitted previously. This means that a large number of missing observations would be considered in the analysis, which would distort its results.

FIG (3.11)

HISTOGRAMS OF LAC FOR EACH LEVEL OF S

```
          * 1.C000            * 2.0000            * 3.0000            * 4.0000            * 5.0000
          ..................+...................+...................+...................+...................+
MIDPOINTS
  11.400)
  11.100)
  10.800)
  10.500)
  10.200)                                       *
   9.900)**********          ******              ******              *****               *******
   9.600)*************       ******************  *****************23 ********            ****************
   9.300)*****************28 *****************31 *****************25 *****************33 *****************22
   9.000)*****************59 *****************63 *****************60 *****************38 *****************55
   8.700)****************113 ****************145 ****************128 *****************94 *****************66
   8.400)****************163 ****************194 ****************158 *****************94 *****************87
   8.100)M***************112 M****************92 M****************98 M****************86 M****************77
   7.800)*****************97 ****************105 *****************97 *****************75 *****************61
   7.500)*****************57 *****************74 *****************63 *****************63 *****************55
   7.200)*****************38 *****************32 *****************25 *****************41 *****************32
   6.900)*************       *****************22 *****************28 *****************23 *****************22
   6.600)*********           ********            ************        ***************     ************
   6.300)****                *****               *****               ******              ****
   6.000)**                  ****                *                   *****               **
   5.700)                                        **                  ***                 ***
   5.400)                    **                                      *
   5.100)                                                            *                   *
   4.800)                    *
   4.500)
   4.200)                                                            *                   *
   3.900)
   3.600)
   3.300)
   3.000)
   2.700)
GROUP MEANS ARE DENOTED BY M'S IF THEY COINCIDE WITH *'S, N'S OTHERWISE

MEAN           8.245               8.237               8.233               8.095               8.152
STD.DEV.       0.649               0.683               0.698               0.796               0.804
S. E. M.       0.024               0.024               0.026               0.033               0.035
MAXIMUM        9.983               9.984              10.058               9.823               9.997
MINIMUM        5.999               4.745               5.714               4.277               4.094
SAMPLE SIZE      718                 802                 732                 593                 523

      ALL GROUPS COMBINED          ******************************  ANALYSIS OF VARIANCE TABLE  ******************************
(EXCEPT CASES WITH UNUSED VALUES   *
  FOR S        )                   *   SOURCE              SUM OF SQUARES     DF      MEAN SQUARE    F VALUE    TAIL PROBABILITY
                                   *
MEAN              8.200            *   BETWEEN GROUPS           11.0660        4       2.7665        5.33         0.0003
STD.DEV.          0.722            *   WITHIN  GROUPS         1745.3317     3363       0.5190
S. E. M.          0.012            *
MAXIMUM          10.358            *   TOTAL                  1756.3977     3367
MINIMUM           4.094            ******************************************************************************************
SAMPLE SIZE        3368            *   LEVENE'S TEST FOR EQUAL VARIANCES     4,3363                   9.80         0.0000
                                   ******************************************************************************************
                                   *   ONE-WAY ANALYSIS OF VARIANCE
                                   *   TEST STATISTICS FOR WITHIN-GROUP
                                   *   VARIANCES NOT ASSUMED TO BE EQUAL
                                   *      WELCH                          4,1597                   4.78         0.0008
                                   *      BROWN-FORSYTHE                 4,2961                   5.18         0.0004
```

Judging by what is shown in Fig. (3.11) there is not much evidence that the variation of LAC could be explained by S. Indeed, the cell means are almost the same, and more important than that, the five distributions seem to have approximately the same shape.

The standard deviations for S = 4 and S = 5 are larger than those for the remaining cells. However, this can be explained by the presence of extreme low values of LAC in both cells.

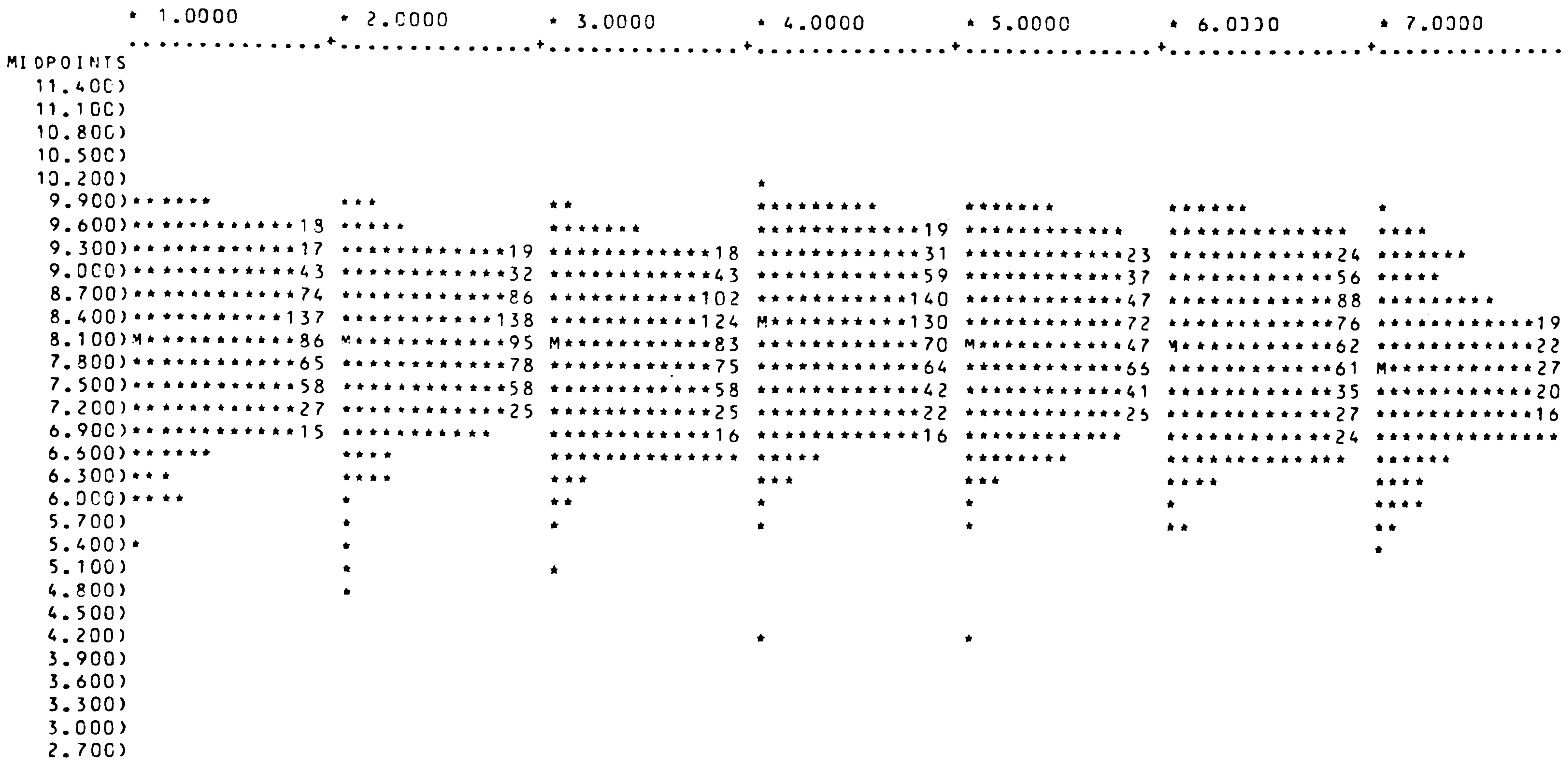
The disproportional distribution of the sum of squares between and within groups in the analysis of variance table can be interpreted as an indication that no relationship exists between LAC and S, although no definitive statements concerning this matter can be made in terms of the results in the table, as mentioned earlier.

If a significant relationship existed between LAC and S one would expect to find definite trends in the histograms according to the type of the relationship. If it is supposed, for instance, that LAC increases with S the observations in the histogram for S = 5 should be concentrated mostly on the top end of the vertical axis whereas those for S = 1 should lie mostly on the bottom end of the scale.

The histograms for the remaining rating factors are shown in Figs. (3.12) to (3.15). No sharp differences appear to exist regarding the shapes of the histograms throughout the levels of each rating factor, and therefore no significant relationship seems to exist between LAC and any one of those factors. However

FIG (3.12)

HISTOGRAMS OF LAC FOR EACH LEVEL OF Z



GROUP MEANS ARE DENOTED BY M'S IF THEY COINCIDE WITH *'S, N'S OTHERWISE

MEAN	8.211	8.163	8.168	8.349	8.178	8.228	7.792
STD.DEV.	0.695	0.649	0.678	0.695	0.776	0.766	0.856
S. E. M.	0.029	0.027	0.028	0.028	0.039	0.035	0.067
MAXIMUM	9.835	9.951	9.800	10.058	10.041	9.997	9.759
MINIMUM	5.541	4.745	5.153	4.094	4.277	5.656	5.521
SAMPLE SIZE	560	563	574	614	404	492	151

ALL GROUPS COMBINED
(EXCEPT CASES WITH UNUSED VALUES FOR Z)

MEAN	8.200
STD.DEV.	0.722
S. E. M.	0.012
MAXIMUM	10.058
MINIMUM	4.094
SAMPLE SIZE	3368

ANALYSIS OF VARIANCE TABLE

SOURCE	SUM OF SQUARES	DF	MEAN SQUARE	F VALUE	TAIL PROBABILITY
BETWEEN GROUPS	42.4981	6	7.0830	13.89	0.
WITHIN GROUPS	1713.8997	3361	0.5099		
TOTAL	1756.3978	3367			
LEVENE'S TEST FOR EQUAL VARIANCES		6,3361		6.59	0.0000
ONE-WAY ANALYSIS OF VARIANCE TEST STATISTICS FOR WITHIN-GROUP VARIANCES NOT ASSUMED TO BE EQUAL					
WELCH		6,1141		11.46	0.
BROWN-FORSYTHE		6,2017		13.05	0.

FIG (3.13)

HISTOGRAMS OF LAC FOR EACH LEVEL OF B

```
         * 1.0000        * 2.0000        * 3.0000        * 4.0000        * 5.0000        * 6.0000        * 7.0000
         ..............+...............+...............+...............+...............+...............+..............+
MIDPOINTS
  11.400)
  11.100)
  10.800)
  10.500)
  10.200)                                *
   9.900)*****           ***             ****            ****            *****           ******          *******
   9.600)******          ************    *******         *************15 **********      *************16 ************
   9.300)*************18 *************19 *************17 *************19 *************20 *************29 *************17
   9.000)*************30 *************34 *************35 *************35 *************32 *************37 *************72
   8.700)*************51 *************64 *************62 *************52 *************61 *************85 ************171
   8.400)************109 *************89 *************85 *************86 *************70 ************107 M***********150
   8.100)M************64 M************65 M************62 M************55 M************78 M************71 *************70
   7.800)*************61 *************58 *************60 *************50 *************63 *************74 *************70
   7.500)*************47 *************42 *************37 *************49 *************54 *************48 *************35
   7.200)*************19 *************22 *************28 *************23 *************28 *************31 *************17
   6.900)*************18 *************15 *************19 *************20 *************15 **********      ***********
   6.600)***********     *****           **********      ***********     ********        *********       **
   6.300)*******         ***             *               ***             **              ****            ****
   6.000)***             ****            ***             *               *               **
   5.700)**              **              *                                               **              *
   5.400)**                                                                                              *
   5.100)*               *
   4.800)                                *
   4.500)
   4.200)*               *
   3.900)
   3.600)
   3.300)
   3.000)
   2.700)
```

GROUP MEANS ARE DENOTED BY M'S IF THEY COINCIDE WITH *'S, N'S OTHERWISE

MEAN	8.094	8.163	8.139	8.172	8.146	8.230	8.370
STD.DEV.	0.792	0.760	0.734	0.741	0.699	0.722	0.603
S. E. M.	0.037	0.036	0.035	0.036	0.033	0.031	0.024
MAXIMUM	10.041	9.984	10.058	9.807	9.997	9.978	9.835
MINIMUM	4.094	4.277	4.745	6.082	5.991	5.557	5.521
SAMPLE SIZE	455	439	433	423	447	531	640

ALL GROUPS COMBINED
(EXCEPT CASES WITH UNUSED VALUES
FOR B)

MEAN	8.200
STD.DEV.	0.722
S. E. M.	0.012
MAXIMUM	10.058
MINIMUM	4.094
SAMPLE SIZE	3368

ANALYSIS OF VARIANCE TABLE

SOURCE	SUM OF SQUARES	DF	MEAN SQUARE	F VALUE	TAIL PROBABILITY
BETWEEN GROUPS	27.9955	6	4.6659	9.07	0.
WITHIN GROUPS	1728.4023	3361	0.5143		
TOTAL	1756.3978	3367			
LEVENE'S TEST FOR EQUAL VARIANCES		6, 3361		7.23	0.0000
ONE-WAY ANALYSIS OF VARIANCE TEST STATISTICS FOR WITHIN-GROUP VARIANCES NOT ASSUMED TO BE EQUAL					
WELCH		6, 1441		10.94	0.
BROWN-FORSYTHE		6, 3121		8.88	0.

FIG (3.14)

HISTOGRAMS OF LAC FOR EACH LEVEL OF M

```
           * 1.0000    * 2.0000    * 3.0000    * 4.0000    * 5.0000    * 6.0000    * 7.0000    * 8.0000    * 9.0000
           ...........+...........+...........+...........+...........+...........+...........+...........+...........+
MIDPOINTS
  11.400)
  11.100)
  10.800)
  10.500)
  10.200)                                      *
   9.900)****        **          *                         *****       *******     *****       **********
   9.600)*****       *********** *******     ****        **********  *********16 **********  *********12 ***
   9.300)*********19 *********18 *********** ****        *********18 *********21 *********13 *********19 *********16
   9.000)*********44 *********30 *********19 *********   *********31 *********49 *********22 *********27 *********44
   8.700)*********96 *********39 *********33 *********23 *********48 *********63 *********25 *********30 ********189
   8.400)M*******101 *********53 M********22 *********19 *********60 *********65 *********52 *********38 M*******286
   8.100)*********48 M********39 *********26 M********32 M********63 M********68 M********59 M********34 *********96
   7.800)*********70 *********57 *********16 *********26 *********78 *********71 *********55 *********34 *********29
   7.500)*********28 *********37 *********18 *********19 *********56 *********66 *********37 *********29 *********22
   7.200)*********12 *********21 *********** *****       *********32 *********35 *********28 *********15 *********
   6.900)***         *********** *****       *********   *********22 *********20 *********19 *********13 ******
   6.600)******      ******      ***         ***         ********    *********13 *****       ********    ****
   6.300)**          ***                                 ****        *****       *****       *****
   6.000)                        *           *           ***         *           ***         **          ***
   5.700)**                                                          **                      ***         *
   5.400)                                                            *           *           *
   5.100)                        *                       *
   4.800)                                                                                    *
   4.500)
   4.200)                                                *                       *
   3.900)
   3.600)
   3.300)
   3.000)
   2.700)
GROUP MEANS ARE DENOTED BY M'S IF THEY COINCIDE WITH *'S, N'S OTHERWISE

MEAN           8.322       8.167       8.254       8.100       8.048       8.114       8.025       8.163       8.401
STD.DEV.       0.615       0.727       0.783       0.701       0.771       0.783       0.797       0.939       0.450
S. E. M.       0.029       0.040       0.059       0.056       0.037       0.035       0.043       0.056       0.017
MAXIMUM       10.041       9.978       9.759      10.058       9.951       9.997       9.984       9.974       9.724
MINIMUM        5.656       6.238       4.997       5.966       4.094       5.521       4.277       4.745       5.694
SAMPLE SIZE      440         327         174         155         440         503         340         281         708

       ALL GROUPS COMBINED          ******************************  ANALYSIS OF VARIANCE TABLE  ******************************
(EXCEPT CASES WITH UNUSED VALUES    *
  FOR M        )                    *   SOURCE              SUM OF SQUARES     DF      MEAN SQUARE   F VALUE   TAIL PROBABILITY
                                    *
MEAN                 8.200          *   BETWEEN GROUPS           62.1335        8         7.7667      15.40          0.
STD.DEV.             0.722          *   WITHIN  GROUPS         1694.2643     3359         0.5044
S. E. M.             0.012          *
MAXIMUM             10.058          *   TOTAL                  1756.3978     3367
MINIMUM              4.094          ******************************************************************************************
SAMPLE SIZE          3368           *   LEVENE'S TEST FOR EQUAL VARIANCES        8,3359                   41.70          0.
                                    ******************************************************************************************
                                    *   ONE-WAY ANALYSIS OF VARIANCE
                                    *   TEST STATISTICS FOR WITHIN-GROUP
                                    *   VARIANCES NOT ASSUMED TO BE EQUAL
                                    *      WELCH                             8,1043                   21.21          0.
                                    *      BROWN-FORSYTHE                    8,2232                   14.01          0.
```

FIG (3.15)

HISTOGRAMS OF LAC FOR EACH LEVEL OF J

```
       * 1.0000                     * 2.0000                     * 3.0000
       ............................+............................+............................+
MIDPOINTS
11.400)
11.100)
10.800)
10.500)
10.200)                                                          *
 9.900)**********                   **************               **********
 9.600)****************             **************************35 ***************************
 9.300)**************************30 **************************63 **************************46
 9.000)**************************62 *************************127 **************************86
 8.700)*************************147 *************************226 *************************173
 8.400)*************************144 M************************313 *************************239
 8.100)M************************104 *************************176 M************************185
 7.800)**************************75 *************************179 *************************182
 7.500)**************************55 *************************108 *************************149
 7.200)**************************36 **************************58 **************************74
 6.900)**************************30 **************************34 **************************44
 6.600)***********************      ***************              ******************
 6.300)****                         ********                     ************
 6.000)***                          *******                      ****
 5.700)***                          ***                          **
 5.400)                             **                           *
 5.100)*                            *
 4.800)                             *
 4.500)
 4.200)                             *                            *
 3.900)
 3.600)
 3.300)
 3.000)
 2.700)
```

GROUP MEANS ARE DENOTED BY M'S IF THEY COINCIDE WITH *'S, N'S OTHERWISE

MEAN	8.216	8.254	8.130
STD.DEV.	0.752	0.712	0.710
S. E. M.	0.028	0.019	0.020
MAXIMUM	9.975	9.997	10.058
MINIMUM	4.997	4.094	4.277
SAMPLE SIZE	743	1371	1254

ALL GROUPS COMBINED
(EXCEPT CASES WITH UNUSED VALUES FOR J)

MEAN	8.200
STD.DEV.	0.722
S. E. M.	0.012
MAXIMUM	10.058
MINIMUM	4.094
SAMPLE SIZE	3368

ANALYSIS OF VARIANCE TABLE

SOURCE	SUM OF SQUARES	DF	MEAN SQUARE	F VALUE	TAIL PROBABILITY
BETWEEN GROUPS	10.4051	2	5.2025	10.03	0.0000
WITHIN GROUPS	1745.9926	3365	0.5189		
TOTAL	1756.3977	3367			
LEVENE'S TEST FOR EQUAL VARIANCES		2, 3365		2.24	0.1066
ONE-WAY ANALYSIS OF VARIANCE TEST STATISTICS FOR WITHIN-GROUP VARIANCES NOT ASSUMED TO BE EQUAL					
WELCH		2, 1888		10.27	0.0000
BROWN-FORSYTHE		2, 2744		9.83	0.0001

if one considers only the cell means some appreciable differences can be noticed, particularly regarding the extreme ones. Thus the smallest cell mean is 7.792, which corresponds to Z = 7 and the largest one is 8.401 which corresponds to M = 9.

One way to assess the relative contribution of the levels of the rating factors in explaining the variation in LAC is to perform a stepwise regression with dummy variables corresponding to these factor levels as the independent variables and LAC as the dependent one.

Each factor with n levels calls for n-1 dummy variables, as one of the levels is taken as reference. For each factor the reference level chosen will be the one which has its cell mean closest to the overall mean (8.200). Thus S = 3, Z = 1, B = 4, M = 2 and J = 1 will be taken as reference levels and the notation RI will be used to represent each dummy variable, where R stands for the rating factor identification (S, Z, B, M or J) and I the level of that factor. Thus, for instance, B6 means the dummy variable associated with B = 6.

Using the subprogram REGRESSION from SPSS and having previously created the required dummy variables as defined above, the result can be seen in Fig. (3.16). As expected, the squared multiple correlation coefficient (R SQUARE) is very small which means that a very poor fit was achieved by the model, and therefore it will be meaningless to try to interpret the coefficients in the model (the notation for the coefficients in the printout is B). However, if one looks at Fig. (3.17) where the step by step variations in R SQUARE are shown, it can be noticed that the

FIG (3.16)

REGRESSION OF LAC ON THE LEVELS OF THE RATING FACTORS

* M U L T I P L E   R E G R E S S I O N * * * * * * * * * * * * * *   VARIABLE LIST 1
REGRESSION LIST 1

DEPENDENT VARIABLE..   LAC

VARIABLE(S) ENTERED ON STEP NUMBER 25..   B2

MULTIPLE R	0.33831		ANALYSIS OF VARIANCE	DF	SUM OF SQUARES	MEAN SQUARE	F
R SQUARE	0.11445		REGRESSION	25.	201.02189	8.04088	17.27724
ADJUSTED R SQUARE	0.10783		RESIDUAL	3342.	1555.37586	0.46540	
STANDARD ERROR	0.68220						

------------------ VARIABLES IN THE EQUATION ------------------

VARIABLE	B	BETA	STD ERROR B	F
M9	0.3613262	0.20387	0.04735	58.229
Z7	-0.5920822	-0.17492	0.05875	101.576
B7	0.2753254	0.14958	0.04308	40.842
M1	0.2148024	0.10024	0.04999	18.466
Z4	0.1592733	0.08516	0.03471	21.060
S4	-0.1787222	-0.09426	0.03778	22.383
J2	0.1459565	0.09930	0.03569	16.720
S5	-0.1512619	-0.07586	0.03928	14.831
M3	0.1804980	0.05532	0.07171	6.336
B6	0.0986948	0.04981	0.04459	4.899
M7	-0.1476381	-0.06159	0.05296	7.771
M5	-0.0898186	-0.04192	0.05038	3.178
Z5	-0.0899342	-0.04046	0.04006	5.041
B1	-0.0715499	-0.03387	0.04610	2.409
Z2	-0.0444077	-0.02294	0.03565	1.552
Z3	-0.0356300	-0.01855	0.03543	1.011
M4	0.0320143	0.00929	0.06871	0.217
S2	0.0314736	0.01856	0.03493	0.812
S1	0.0237078	0.01345	0.03592	0.436
M8	-0.0364693	-0.01397	0.05586	0.426
B3	-0.0227524	-0.01055	0.04667	0.238
B5	-0.0156807	-0.00737	0.04629	0.115
J3	-0.0180612	-0.01209	0.03643	0.246
M6	-0.0172069	-0.00849	0.04927	0.122
B2	0.0063893	0.00298	0.04653	0.019
(CONSTANT)	8.0761731			

------------- VARIABLES NOT IN THE EQUATION -------------

VARIABLE	BETA IN	PARTIAL	TOLERANCE	F
Z6	-0.00185	-0.00155	0.62135	0.008

F-LEVEL OR TOLERANCE-LEVEL INSUFFICIENT FOR FURTHER COMPUTATION

STATISTICS WHICH CANNOT BE COMPUTED ARE PRINTED AS ALL NINES.

FIG (3.17)

REGRESSION OF LAC ON THE LEVELS OF THE RATING

FACTORS (CONTINUED)

* M U L T I P L E   R E G R E S S I O N * * * * * * * * * * * * *   VARIABLE LIST 1
REGRESSION LIST 1

DEPENDENT VARIABLE..   LAC

SUMMARY TABLE

VARIABLE	MULTIPLE R	R SQUARE	RSQ CHANGE	SIMPLE R	B	BETA
M9	0.14366	0.02064	0.02064	0.14366	0.3613262	0.20387
Z7	0.20800	0.04326	0.02262	-0.12646	-0.5920822	-0.17492
B7	0.24547	0.06026	0.01699	0.11467	0.2753254	0.14958
M1	0.26856	0.07212	0.01187	0.06576	0.2148024	0.10024
Z4	0.28438	0.08087	0.00875	0.09797	0.1592733	0.08516
S4	0.29602	0.08763	0.00675	-0.06707	-0.1787222	-0.09426
J2	0.30709	0.09430	0.00667	0.06307	0.1459565	0.09930
S5	0.31624	0.10001	0.00571	-0.02798	-0.1512619	-0.07586
M3	0.32392	0.10492	0.00491	0.01747	0.1804980	0.05532
B6	0.32901	0.10825	0.00333	0.01808	0.0986948	0.04981
M7	0.33227	0.11040	0.00215	-0.08075	-0.1476381	-0.06159
M5	0.33400	0.11156	0.00116	-0.08133	-0.0898186	-0.04192
Z5	0.33535	0.11246	0.00090	-0.01100	-0.0899342	-0.04046
B1	0.33656	0.11327	0.00082	-0.05775	-0.0715499	-0.03387
Z2	0.33691	0.11351	0.00023	-0.02247	-0.0444077	-0.02294
Z3	0.33730	0.11377	0.00027	-0.01961	-0.0356300	-0.01855
M4	0.33752	0.11392	0.00015	-0.03036	0.0320143	0.00929
S2	0.33769	0.11403	0.00011	0.02873	0.0314736	0.01856
S1	0.33786	0.11415	0.00011	0.03257	0.0237078	0.01345
M8	0.33798	0.11423	0.00008	-0.01541	-0.0364693	-0.01397
B3	0.33809	0.11430	0.00007	-0.03211	-0.0227524	-0.01055
B5	0.33817	0.11436	0.00006	-0.02881	-0.0156807	-0.00737
J3	0.33825	0.11441	0.00005	-0.07421	-0.0180612	-0.01209
M6	0.33830	0.11445	0.00003	-0.04985	-0.0172069	-0.00849
B2	0.33831	0.11445	0.00000	-0.01944	0.0063893	0.00298
(CONSTANT)					8.0761731	

entrance of the first three dummy variables in the model provokes considerable relative change in R SQUARE, whereas the others do not. So the little power of explanation the rating factors have in the variation in LAC is possibly concentrated on three particular levels of them, namely :

- M = 9, which corresponds to all other models of vehicles not accounted for in the previous levels of M. This category can possibly include more powerful vehicles which can cause greater average damages when they collide.

- Z = 7, which corresponds to a big city (Gotland), but not rated in the same class as the other big cities (Z = 1). This is possibly the reason why Gotland is rated separately (the negative coefficient in the equation means that the claims in that class are smaller on average).

- B = 7, which corresponds to the highest level of bonus discount, and therefore small claims tend to be absorbed by the policyholder to avoid losing the bonus.

3.6 Distribution of LAC when NCLA is small

In the previous section, an attempt was made to explain the variation in LAC in terms of the five rating factors, and no strong evidence was found. If one refers back to Fig. (3.10) it can be noticed that values of LAC based on a large number of claims suffer less variation than those based on small values of NCLA. Indeed, the values in the region denoted by A (from 1 to

10 claims) vary from 5.625 to 10.000 whereas those in the region denoted by C (more than 100 claims) vary in a quite narrower interval from 8.250 to 8.875 - this fact being easily explained by the central limit theorem.

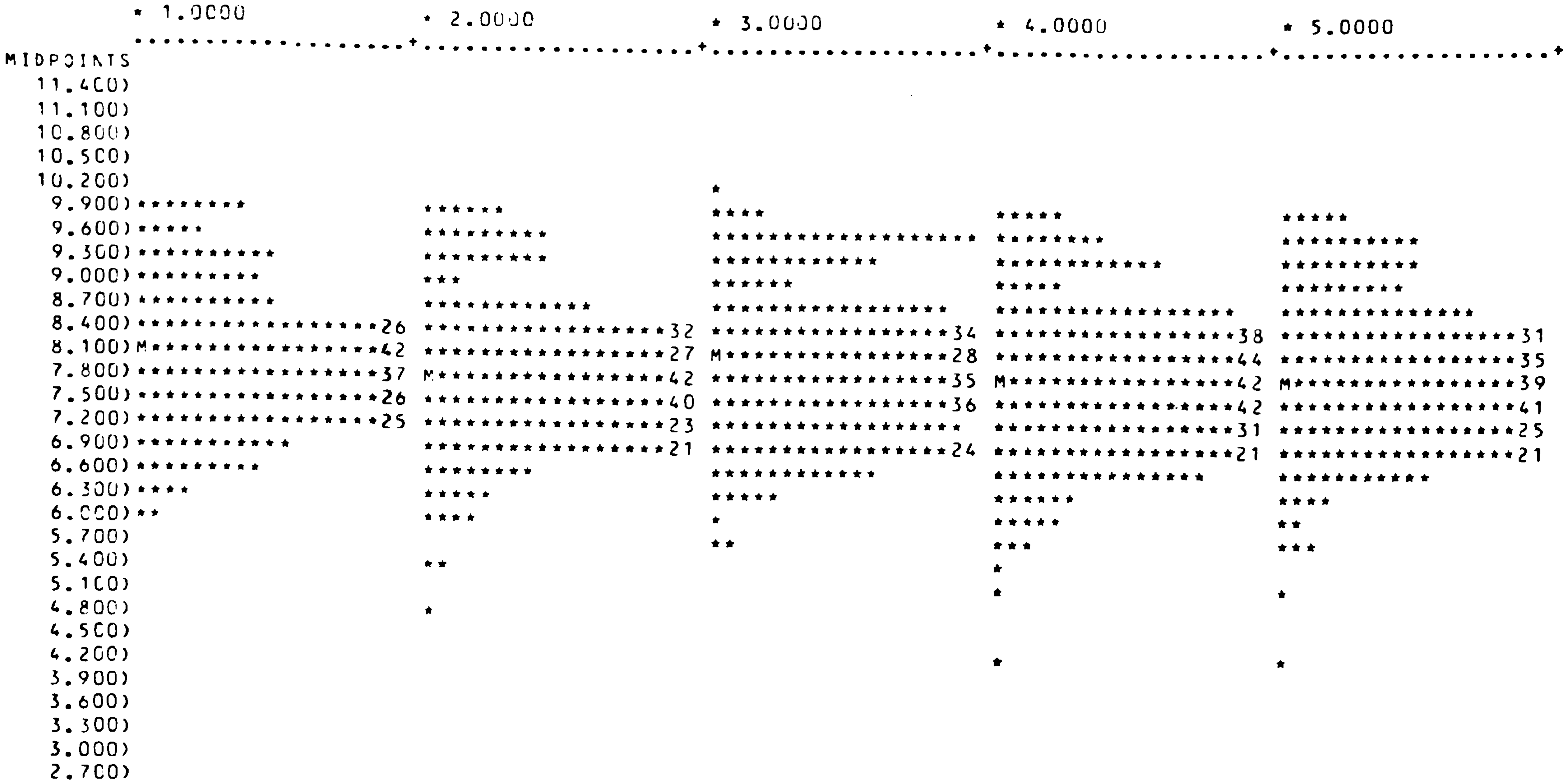
Based on the above considerations, it seems sensible to compare the histograms described in the previous section, considering only observations with a small number of claims , where most of the variation in LAC is contained. The result is shown in Figs. (3.18) to (3.22), where the observations were reduced to those with number of claims less than 4.

Once more, no sharp differences seem to exist regarding the shapes of the histograms throughout the levels of each rating factor. It is worth noticing that the largest cell mean (8.074) now corresponds to Z = 4, which was ranked in the fifth place in the stepwise analysis of the previous section. The second largest cell mean (8.015) corresponds to B = 7 and the smallest one (7.645) corresponds to Z = 7, in a reasonable agreement with the previous analysis. The cell mean for M = 9 has drastically decreased to 7.815 and the cause for that can well be the equally drastic reduction in the number of observations (sample size = 53) in that cell.

A rather curious feature can be observed in almost all histograms, namely the existence of two peaks (and sometimes more than two) in the majority of them. In order to further investigate this fact, one will request detailed histograms of LAC successively for observations with NCLA equal to 1, 2, 3, 4 and 5, using the procedure 5D of BMDP. The results can be seen in Figs. (3.23) to (3.27)

FIG (3.18)

HISTOGRAMS OF LAC FOR EACH LEVEL OF S (SMALL NCLA)



GROUP MEANS ARE DENOTED BY M'S IF THEY COINCIDE WITH *'S, N'S OTHERWISE

MEAN	7.965	7.834	7.973	7.804	7.858
STD.DEV.	0.814	0.875	0.913	0.898	0.903
S. E. M.	0.054	0.056	0.057	0.052	0.056
MAXIMUM	9.983	9.984	10.058	9.823	9.997
MINIMUM	5.999	4.745	5.714	4.277	4.094
SAMPLE SIZE	224	244	254	297	262

ALL GROUPS COMBINED
(EXCEPT CASES WITH UNUSED VALUES
FOR S)

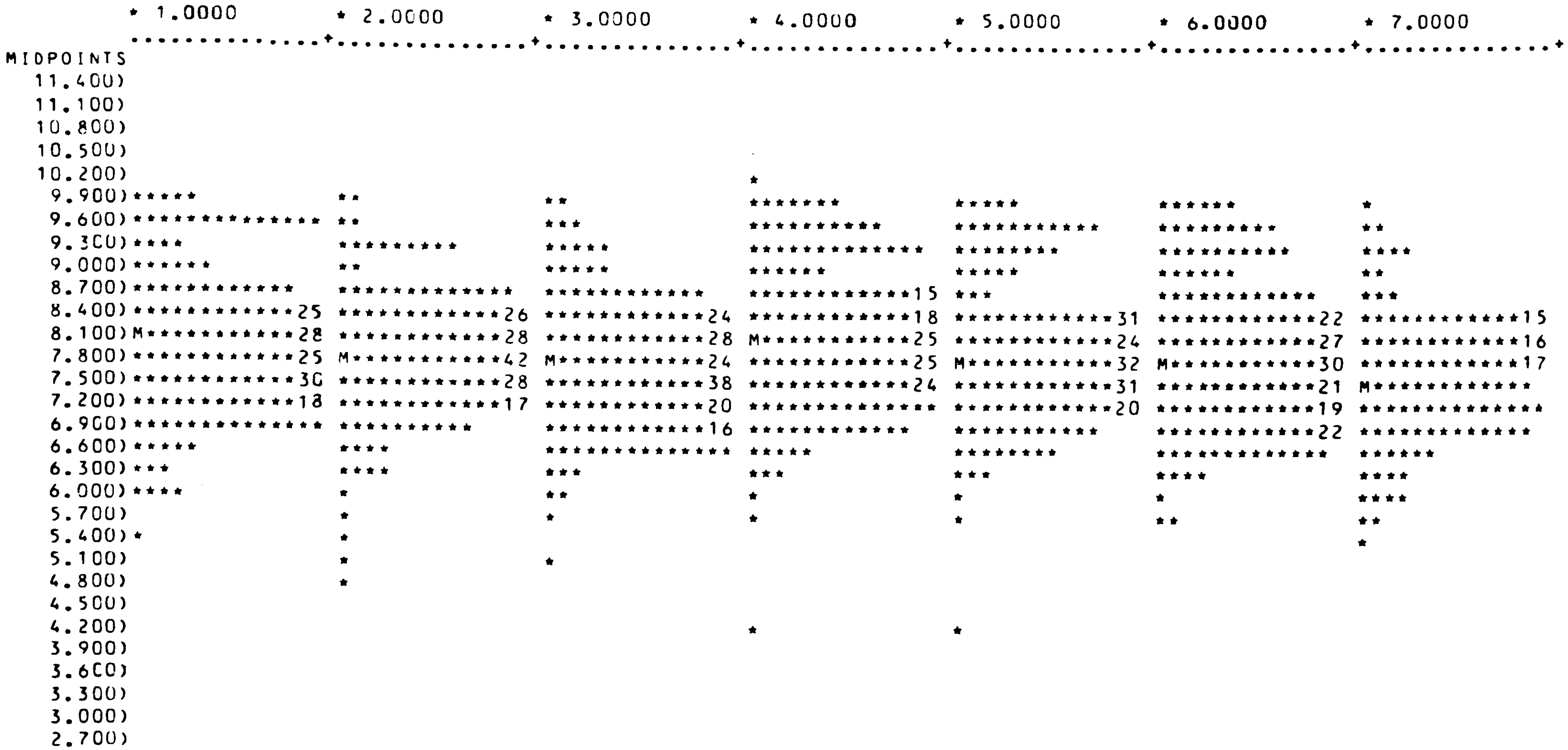
MEAN	7.882
STD.DEV.	0.885
S. E. M.	0.025
MAXIMUM	10.058
MINIMUM	4.094
SAMPLE SIZE	1281

ANALYSIS OF VARIANCE TABLE

SOURCE	SUM OF SQUARES	DF	MEAN SQUARE	F VALUE	TAIL PROBABILITY
BETWEEN GROUPS	6.2023	4	1.5506	1.99	0.0944
WITHIN GROUPS	996.3809	1276	0.7809		
TOTAL	1002.5832	1280			
LEVENE'S TEST FOR EQUAL VARIANCES		4,1276		1.34	0.2516
ONE-WAY ANALYSIS OF VARIANCE TEST STATISTICS FOR WITHIN-GROUP VARIANCES NOT ASSUMED TO BE EQUAL					
WELCH		4, 631		2.01	0.0922
BROWN-FORSYTHE		4,1268		2.00	0.0925

FIG (3.19)

HISTOGRAMS OF LAC FOR EACH LEVEL OF Z (SMALL NCLA)



GROUP MEANS ARE DENOTED BY M'S IF THEY COINCIDE WITH *'S, N'S OTHERWISE

	1	2	3	4	5	6	7
MEAN	7.959	7.846	7.745	8.074	7.922	7.906	7.645
STD.DEV.	0.897	0.801	0.795	0.963	0.888	0.926	0.867
S. E. M.	0.064	0.058	0.057	0.072	0.064	0.065	0.080
MAXIMUM	9.835	9.769	9.800	10.058	10.041	9.997	9.769
MINIMUM	5.541	4.745	5.153	4.094	4.277	5.656	5.521
SAMPLE SIZE	194	192	198	181	195	204	117

ALL GROUPS COMBINED
(EXCEPT CASES WITH UNUSED VALUES FOR Z)

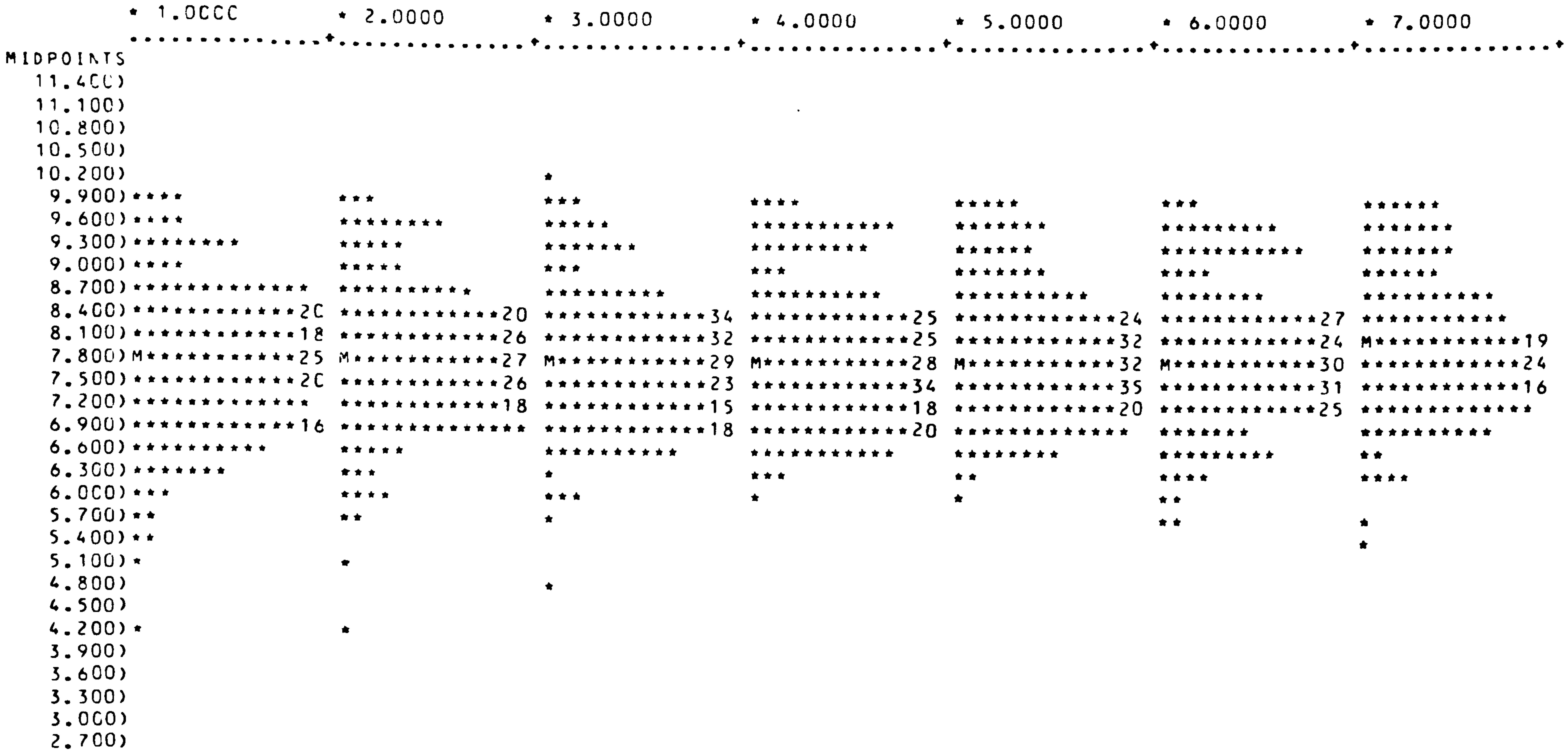
MEAN	7.882
STD.DEV.	0.885
S. E. M.	0.025
MAXIMUM	10.058
MINIMUM	4.094
SAMPLE SIZE	1281

ANALYSIS OF VARIANCE TABLE

SOURCE	SUM OF SQUARES	DF	MEAN SQUARE	F VALUE	TAIL PROBABILITY
BETWEEN GROUPS	18.8645	6	3.1441	4.07	0.0005
WITHIN GROUPS	983.7187	1274	0.7721		
TOTAL	1002.5832	1280			
LEVENE'S TEST FOR EQUAL VARIANCES		6,1274		2.20	0.0411
ONE-WAY ANALYSIS OF VARIANCE TEST STATISTICS FOR WITHIN-GROUP VARIANCES NOT ASSUMED TO BE EQUAL					
WELCH		6, 541		3.98	0.0007
BROWN-FORSYTHE		6,1203		4.07	0.0005

FIG (3.20)

HISTOGRAMS OF LAC FOR EACH LEVEL OF B (SMALL NCLA)



GROUP MEANS ARE DENOTED BY M'S IF THEY COINCIDE WITH *'S, N'S OTHERWISE

MEAN	7.744	7.822	7.876	7.902	7.926	7.906	8.015
STD.DEV.	1.008	0.926	0.836	0.854	0.791	0.870	0.923
S. E. M.	0.077	0.069	0.060	0.060	0.056	0.062	0.079
MAXIMUM	10.041	9.984	10.058	9.807	9.997	9.978	9.835
MINIMUM	4.094	4.277	4.745	6.082	5.991	5.557	5.521
SAMPLE SIZE	171	178	195	202	202	196	137

ALL GROUPS COMBINED
(EXCEPT CASES WITH UNUSED VALUES
FOR B)

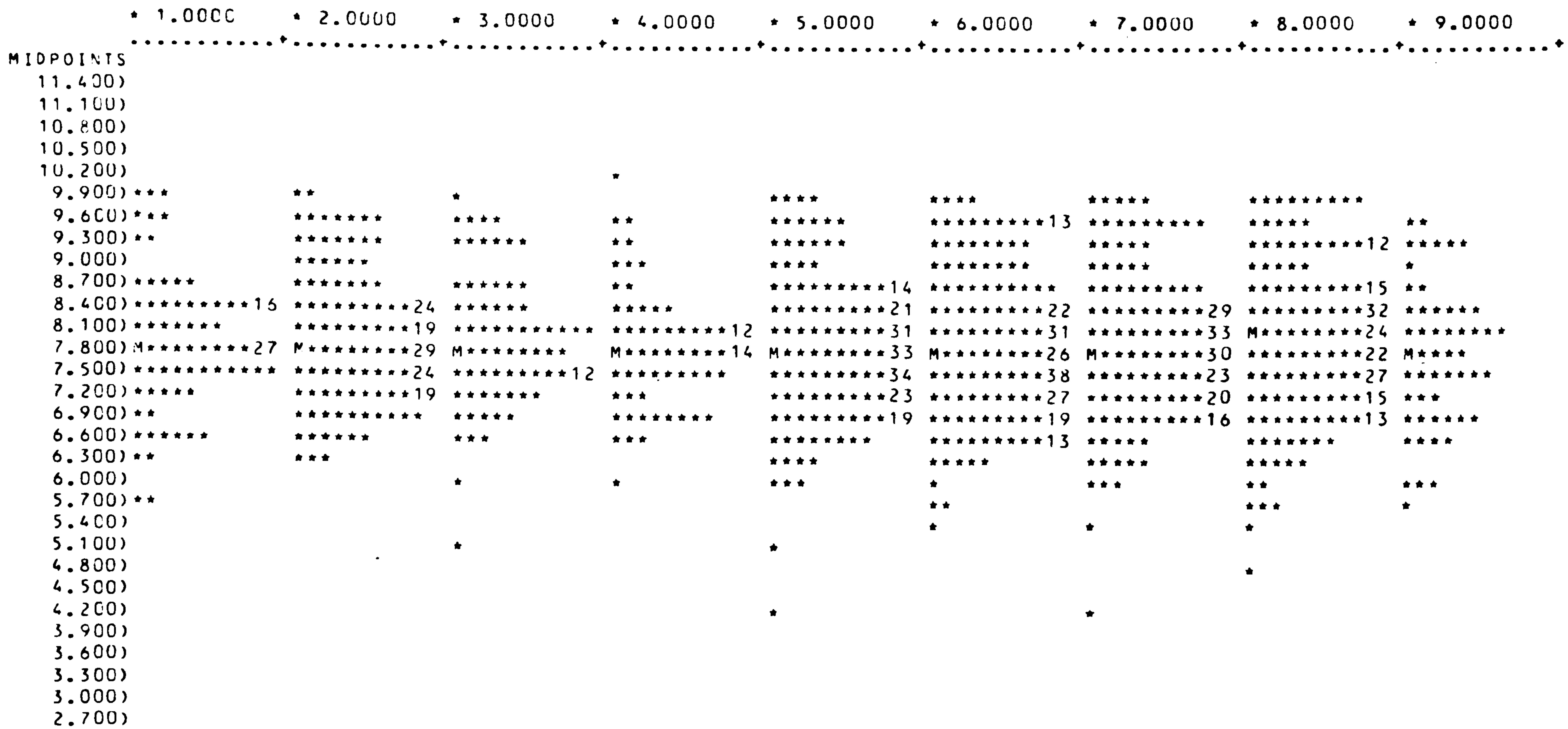
MEAN	7.882
STD.DEV.	0.885
S. E. M.	0.025
MAXIMUM	10.058
MINIMUM	4.094
SAMPLE SIZE	1281

ANALYSIS OF VARIANCE TABLE

SOURCE	SUM OF SQUARES	DF	MEAN SQUARE	F VALUE	TAIL PROBABILITY
BETWEEN GROUPS	6.9103	6	1.1517	1.47	0.1836
WITHIN GROUPS	995.6729	1274	0.7815		
TOTAL	1002.5832	1280			
LEVENE'S TEST FOR EQUAL VARIANCES		6,1274		1.70	0.1180
ONE-WAY ANALYSIS OF VARIANCE TEST STATISTICS FOR WITHIN-GROUP VARIANCES NOT ASSUMED TO BE EQUAL					
WELCH		6, 551		1.26	0.2720
BROWN-FORSYTHE		6,1185		1.45	0.1908

FIG (3.21)

HISTOGRAMS OF LAC FOR EACH LEVEL OF M (SMALL NCLA)



GROUP MEANS ARE DENOTED BY M'S IF THEY COINCIDE WITH *'S, N'S OTHERWISE

MEAN	7.902	7.940	7.940	7.863	7.802	7.837	7.888	7.962	7.815
STD.DEV.	0.836	0.793	0.933	0.805	0.855	0.903	0.883	0.972	0.986
S. E. M.	0.086	0.062	0.110	0.100	0.059	0.060	0.063	0.069	0.135
MAXIMUM	10.041	9.978	9.759	10.058	9.777	9.997	9.984	9.974	9.724
MINIMUM	5.656	6.238	4.997	5.966	4.094	5.521	4.277	4.745	5.694
SAMPLE SIZE	91	163	72	65	212	228	199	198	53

ALL GROUPS COMBINED
(EXCEPT CASES WITH UNUSED VALUES
FOR M)

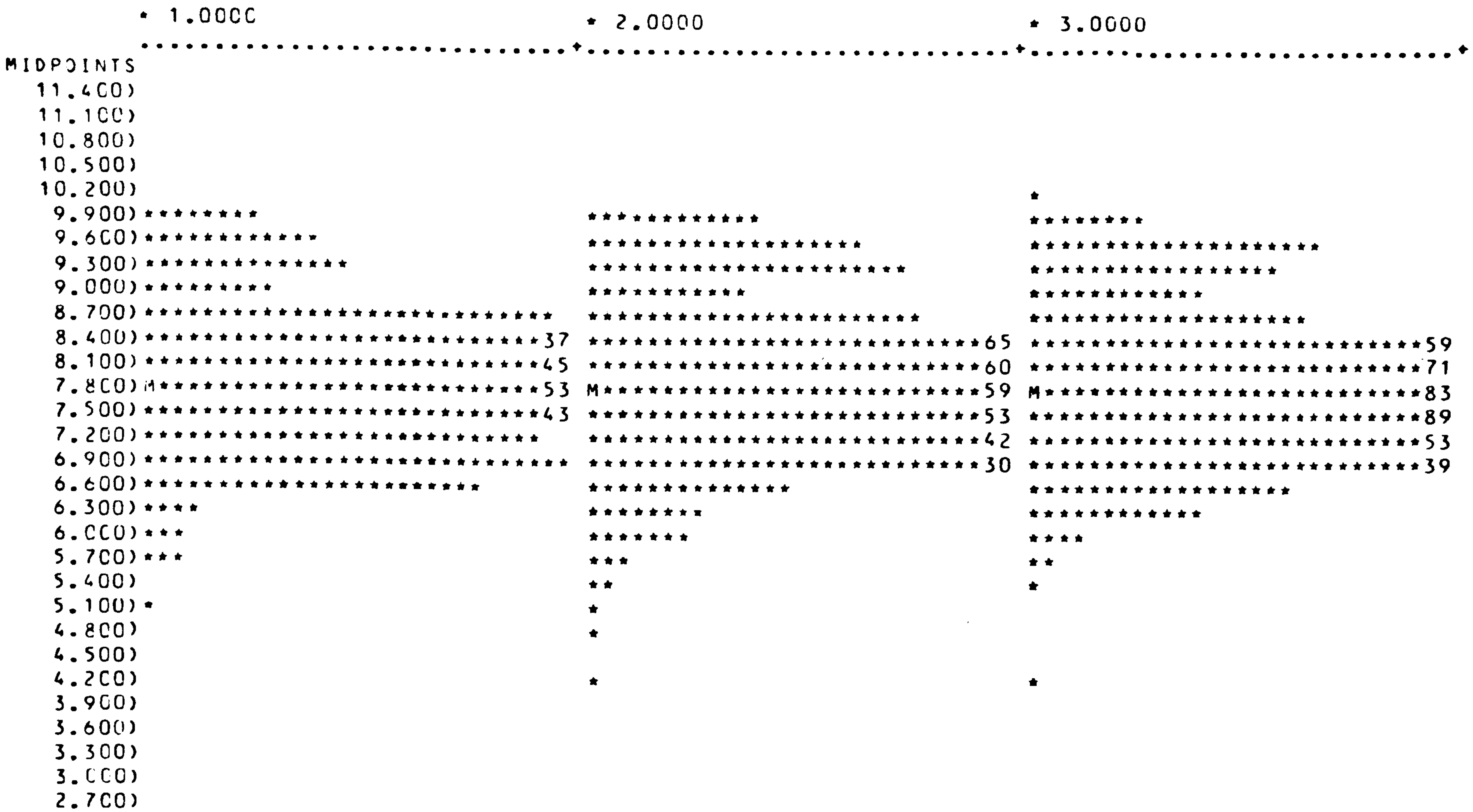
MEAN	7.892
STD.DEV.	0.885
S. E. M.	0.025
MAXIMUM	10.058
MINIMUM	4.094
SAMPLE SIZE	1281

ANALYSIS OF VARIANCE TABLE

SOURCE	SUM OF SQUARES	DF	MEAN SQUARE	F VALUE	TAIL PROBABILITY
BETWEEN GROUPS	4.1593	8	0.5199	0.66	0.7250
WITHIN GROUPS	998.4239	1272	0.7849		
TOTAL	1002.5832	1280			
LEVENE'S TEST FOR EQUAL VARIANCES		8,1272		1.72	0.0885
ONE-WAY ANALYSIS OF VARIANCE TEST STATISTICS FOR WITHIN-GROUP VARIANCES NOT ASSUMED TO BE EQUAL					
WELCH		8, 370		0.65	0.7385
BROWN-FORSYTHE		8, 831		0.66	0.7273

FIG (3.22)

HISTOGRAMS OF LAC FOR EACH LEVEL OF J (SMALL NCLA)

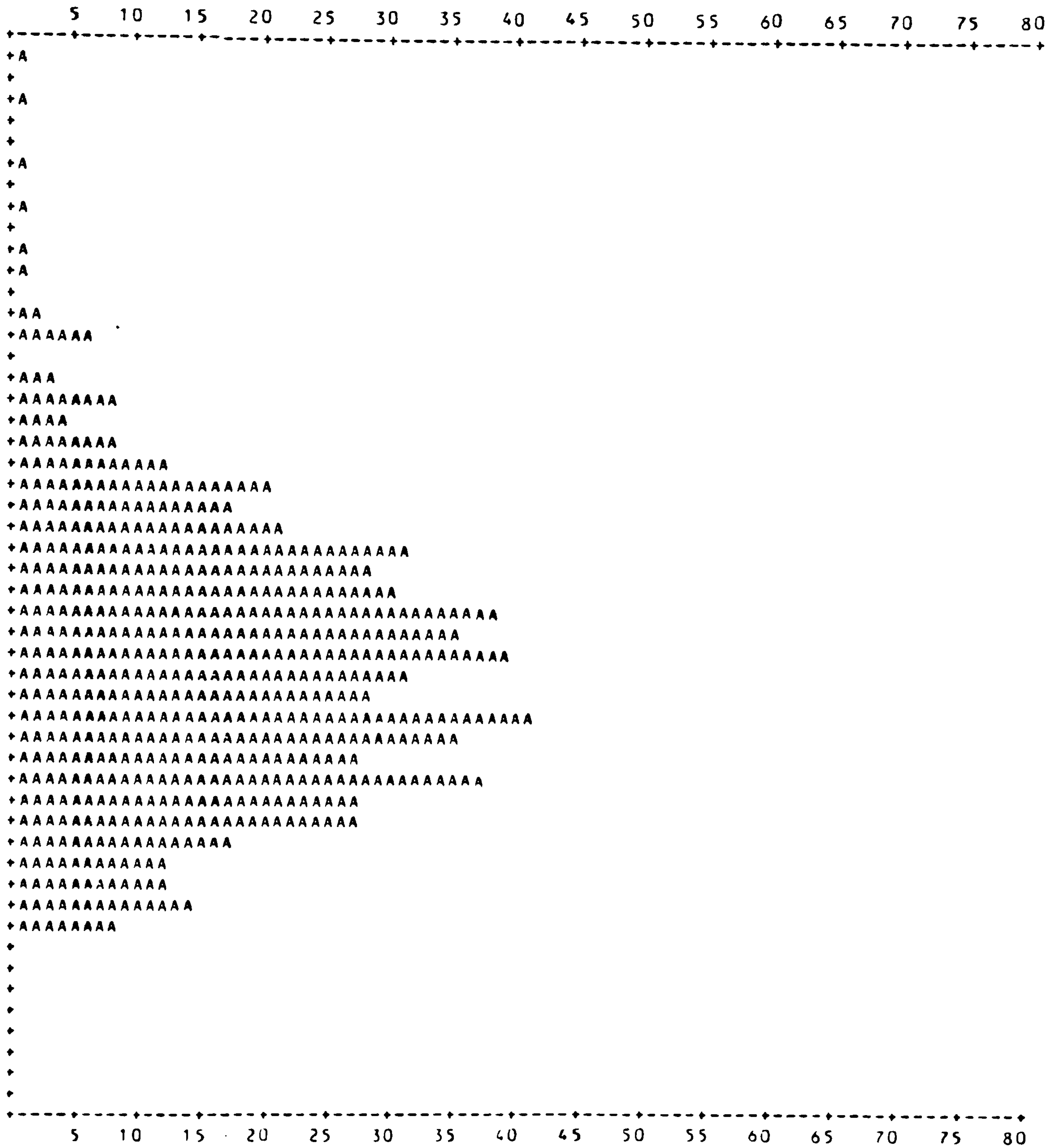


```
GROUP MEANS ARE DENOTED BY M'S IF THEY COINCIDE WITH *'S, N'S OTHERWISE

MEAN            7.889              7.929              7.839
STD.DEV.        0.889              0.935              0.837
S. E. M.        0.048              0.045              0.037
MAXIMUM         9.975              9.997             10.058
MINIMUM         4.997              4.094              4.277
SAMPLE SIZE       339                433                509

      ALL GROUPS COMBINED         ******************************  ANALYSIS OF VARIANCE TABLE  ******************************
(EXCEPT CASES WITH UNUSED VALUES  *
  FOR J       )                   *   SOURCE              SUM OF SQUARES     DF     MEAN SQUARE   F VALUE   TAIL PROBABILITY
                                  *
MEAN              7.882           *   BETWEEN GROUPS          1.8902          2       0.9451       1.21          0.2994
STD.DEV.          0.885           *   WITHIN  GROUPS       1000.6930       1278       0.7830
S. E. M.          0.025           *
MAXIMUM          10.058           *   TOTAL                1002.5832       1280
MINIMUM           4.094           ******************************************************************************************
SAMPLE SIZE        1281           *   LEVENE'S TEST FOR EQUAL VARIANCES      2,1278                    2.79          0.0616
                                  ******************************************************************************************
                                  *   ONE-WAY ANALYSIS OF VARIANCE
                                  *   TEST STATISTICS FOR WITHIN-GROUP
                                  *   VARIANCES NOT ASSUMED TO BE EQUAL
                                  *      WELCH                           2, 791                    1.21          0.3002
                                  *      BROWN-FORSYTHE                  2,1178                    1.19          0.3031
```

FIG (3.23)

HISTOGRAM OF LAC GIVEN ONE CLAIM

	SYMBOL	COUNT	MEAN	ST.DEV.
1	A	624	7.620	0.817



INTERVAL NAME	FREQUENCY INT.	FREQUENCY CUM.	PERCENTAGE INT.	PERCENTAGE CUM.
4.1250	1	1	0.2	0.2
4.25C0	0	1	0.	0.2
4.3750	1	2	0.2	0.3
4.5000	0	2	0.	0.3
4.6250	0	2	0.	0.3
4.7500	1	3	0.2	0.5
4.8750	0	3	0.	0.5
5.0000	1	4	0.2	0.6
5.1250	0	4	0.	0.6
5.2500	1	5	0.2	0.8
5.3750	1	6	0.2	1.0
5.5C00	0	6	0.	1.0
5.6250	2	8	0.3	1.3
5.7500	6	14	1.0	2.2
5.8750	0	14	0.	2.2
6.0000	3	17	0.5	2.7
6.1250	8	25	1.3	4.0
6.25C0	4	29	0.6	4.6
6.3750	8	37	1.3	5.9
6.5000	12	49	1.9	7.9
6.6250	20	69	3.2	11.1
6.7500	17	86	2.7	13.8
6.8750	21	107	3.4	17.1
7.00C0	31	138	5.0	22.1
7.1250	28	166	4.5	26.6
7.2500	30	196	4.8	31.4
7.3750	38	234	6.1	37.5
7.5000	35	269	5.6	43.1
7.6250	39	308	6.3	49.4
7.75C0	31	339	5.0	54.3
7.8750	28	367	4.5	58.8
8.0000	41	408	6.6	65.4
8.1250	35	443	5.6	71.0
8.2500	27	470	4.3	75.3
8.3750	37	507	5.9	81.3
8.5000	27	534	4.3	85.6
8.6250	27	561	4.3	89.9
8.75C0	17	578	2.7	92.6
8.8750	12	590	1.9	94.6
9.0000	12	602	1.9	96.5
9.1250	14	616	2.2	98.7
9.25C0	8	624	1.3	100.0
9.3750	0	624	0.	100.0
9.5000	0	624	0.	100.0
9.6250	0	624	0.	100.0
9.75C0	0	624	0.	100.0
9.8750	0	624	0.	100.0
1C.CC0	0	624	0.	100.0
10.125	0	624	0.	100.0
LAST	0	624	0.	100.0

FIG (3.24)

HISTOGRAM OF LAC GIVEN TWO CLAIMS

	SYMBOL	COUNT	MEAN	ST.DEV.
2	B	390	8.069	0.861

INTERVAL NAME	FREQUENCY INT.	FREQUENCY CUM.	PERCENTAGE INT.	PERCENTAGE CUM.
4.1250	0	0	0.	0.
4.2500	0	0	0.	0.
4.3750	0	0	0.	0.
4.5000	0	0	0.	0.
4.6250	0	0	0.	0.
4.7500	0	0	0.	0.
4.8750	0	0	0.	0.
5.0000	0	0	0.	0.
5.1250	0	0	0.	0.
5.2500	0	0	0.	0.
5.3750	0	0	0.	0.
5.5000	0	0	0.	0.
5.6250	0	0	0.	0.
5.7500	1	1	0.3	0.3
5.8750	0	1	0.	0.3
6.0000	0	1	0.	0.3
6.1250	1	2	0.3	0.5
6.2500	2	4	0.5	1.0
6.3750	0	4	0.	1.0
6.5000	4	8	1.0	2.1
6.6250	4	12	1.0	3.1
6.7500	6	18	1.5	4.6
6.8750	8	26	2.1	6.7
7.0000	9	35	2.3	9.0
7.1250	8	43	2.1	11.0
7.2500	14	57	3.6	14.6
7.3750	14	71	3.6	18.2
7.5000	21	92	5.4	23.6
7.6250	28	120	7.2	30.8
7.7500	33	153	8.5	39.2
7.8750	21	174	5.4	44.6
8.0000	27	201	6.9	51.5
8.1250	25	226	6.4	57.9
8.2500	25	251	6.4	64.4
8.3750	24	275	6.2	70.5
8.5000	22	297	5.6	76.2
8.6250	17	314	4.4	80.5
8.7500	15	329	3.8	84.4
8.8750	6	335	1.5	85.9
9.0000	0	335	0.	85.9
9.1250	2	337	0.5	86.4
9.2500	0	337	0.	86.4
9.3750	0	337	0.	86.4
9.5000	0	337	0.	86.4
9.6250	0	337	0.	86.4
9.7500	33	370	8.5	94.9
9.8750	19	389	4.9	99.7
10.000	1	390	0.3	100.0
10.125	0	390	0.	100.0
LAST	0	390	0.	100.0

FIG (3.25)

HISTOGRAM OF LAC GIVEN THREE CLAIMS

3	SYMBOL	COUNT	MEAN	ST.DEV.
	C	267	8.224	0.889

INTERVAL NAME	5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80	FREQUENCY INT.	FREQUENCY CUM.	PERCENTAGE INT.	PERCENTAGE CUM.
* 4.1250	+	0	0	0.	0.
* 4.2500	+	0	0	0.	0.
* 4.3750	+	0	0	0.	0.
* 4.5000	+	0	0	0.	0.
* 4.6250	+	0	0	0.	0.
* 4.7500	+	0	0	0.	0.
* 4.8750	+	0	0	0.	0.
* 5.0000	+	0	0	0.	0.
* 5.1250	+	0	0	0.	0.
* 5.2500	+	0	0	0.	0.
* 5.3750	+	0	0	0.	0.
* 5.5000	+	0	0	0.	0.
* 5.6250	+C	1	1	0.4	0.4
* 5.7500	+	0	1	0.	0.4
* 5.8750	+	0	1	0.	0.4
* 6.0000	+C	1	2	0.4	0.7
* 6.1250	+	0	2	0.	0.7
* 6.2500	+	0	2	0.	0.7
* 6.3750	+	0	2	0.	0.7
* 6.5000	+	0	2	0.	0.7
* 6.6250	+	0	2	0.	0.7
* 6.7500	+CCC	3	5	1.1	1.9
* 6.8750	+CC	2	7	0.7	2.6
* 7.0000	+CCCC	4	11	1.5	4.1
* 7.1250	+CCCCCCCC	8	19	3.0	7.1
* 7.2500	+CCCCCCCCC	9	28	3.4	10.5
* 7.3750	+CCCCCCCCCC	10	38	3.7	14.2
* 7.5000	+CCCCCCCCCCCCCCCC	16	54	6.0	20.2
* 7.6250	+CCCCCCCCCCCCCCCCC	17	71	6.4	26.6
* 7.7500	+CCCCCCCCCCCCCCCCCCCC	20	91	7.5	34.1
* 7.8750	+CCCCCCCCCCCCCCCCCCCCCC	22	113	8.2	42.3
* 8.0000	+CCCCCCCCCCCCCCCCCCC	19	132	7.1	49.4
* 8.1250	+CCCCCCCCCCCCCCCCCCCC	20	152	7.5	56.9
* 8.2500	+CCCCCCCCCCCC	12	164	4.5	61.4
* 8.3750	+CCCCCCCCC	9	173	3.4	64.8
* 8.5000	+CCCCCCCCCCC	11	184	4.1	68.9
* 8.6250	+CCCCCCCC	8	192	3.0	71.9
* 8.7500	+C	1	193	0.4	72.3
* 8.8750	+	0	193	0.	72.3
* 9.0000	+	0	193	0.	72.3
* 9.1250	+	0	193	0.	72.3
* 9.2500	+	0	193	0.	72.3
* 9.3750	+CCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	30	223	11.2	83.5
* 9.5000	+CCCCCCCCCCCCCCCCCCCCCCCC	24	247	9.0	92.5
* 9.6250	+CCCCCCCCC	9	256	3.4	95.9
* 9.7500	+CC	2	258	0.7	96.6
* 9.8750	+	0	258	0.	96.6
* 10.000	+CCCCCCC	7	265	2.6	99.3
* 10.125	+CC	2	267	0.7	100.0
*LAST	+	0	267	0.	100.0
	5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80				

FIG (3.26)

HISTOGRAM OF LAC GIVEN FOUR CLAIMS

```
                                SYMBOL  COUNT     MEAN     ST.DEV.
                       4           D      206      8.265      0.771
INTERVAL                                                                                 FREQUENCY PERCENTAGE
NAME           5    10    15    20    25    30    35    40    45    50    55    60    65    70    75    80  INT. CUM.  INT.  CUM.
            +----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+
* 4.1250 +                                                                                   0    0    0.     0.
* 4.2500 +                                                                                   0    0    0.     0.
* 4.3750 +                                                                                   0    0    0.     0.
* 4.5000 +                                                                                   0    0    0.     0.
* 4.6250 +                                                                                   0    0    0.     0.
* 4.7500 +                                                                                   0    0    0.     0.
* 4.8750 +                                                                                   0    0    0.     0.
* 5.0000 +                                                                                   0    0    0.     0.
* 5.1250 +                                                                                   0    0    0.     0.
* 5.2500 +                                                                                   0    0    0.     0.
* 5.3750 +                                                                                   0    0    0.     0.
* 5.5000 +                                                                                   0    0    0.     0.
* 5.6250 +                                                                                   0    0    0.     0.
* 5.7500 +                                                                                   0    0    0.     0.
* 5.8750 +                                                                                   0    0    0.     0.
* 6.0000 +                                                                                   0    0    0.     0.
* 6.1250 +                                                                                   0    0    0.     0.
* 6.2500 +                                                                                   0    0    0.     0.
* 6.3750 +                                                                                   0    0    0.     0.
* 6.5000 +                                                                                   0    0    0.     0.
* 6.6250 +D                                                                                  1    1    0.5    0.5
* 6.7500 +                                                                                   0    1    0.     0.5
* 6.8750 +D                                                                                  1    2    0.5    1.0
* 7.0000 +DDDD                                                                               4    6    1.9    2.9
* 7.1250 +DDDD                                                                               4   10    1.9    4.9
* 7.2500 +DDDDDDDDD                                                                          9   19    4.4    9.2
* 7.3750 +DDD                                                                                3   22    1.5   10.7
* 7.5000 +DDDDDDDDDD                                                                        10   32    4.9   15.5
* 7.6250 +DDDDDDDDDDDDDD                                                                    14   46    6.8   22.3
* 7.7500 +DDDDDDDDDDDDDDDD                                                                  16   62    7.8   30.1
* 7.8750 +DDDDDDDDDDDDD                                                                     13   75    6.3   36.4
* 8.0000 +DDDDDDDDDDDDDDDDDDDD                                                              20   95    9.7   46.1
* 8.1250 +DDDDDDDDDDDD                                                                      12  107    5.8   51.9
* 8.2500 +DDDDDDDDDD                                                                        10  117    4.9   56.8
* 8.3750 +DDDDDDDD                                                                           8  125    3.9   60.7
* 8.5000 +DDDDDD                                                                             6  131    2.9   63.6
* 8.6250 +DDDDDD                                                                             6  137    2.9   66.5
* 8.7500 +DDDDDD                                                                             6  143    2.9   69.4
* 8.8750 +                                                                                   0  143    0.    69.4
* 9.0000 +                                                                                   0  143    0.    69.4
* 9.1250 +DDDDDDDDDDDDDD                                                                    14  157    6.8   76.2
* 9.2500 +DDDDDDDDDDDDDDDDDDDDDDDD                                                          24  181   11.7   87.9
* 9.3750 +DDDDDDDDDDDDD                                                                     13  194    6.3   94.2
* 9.5000 +DDDDDD                                                                             6  200    2.9   97.1
* 9.6250 +D                                                                                  1  201    0.5   97.6
* 9.7500 +DDD                                                                                3  204    1.5   99.0
* 9.8750 +DD                                                                                 2  206    1.0  100.0
* 10.000 +                                                                                   0  206    0.   100.0
* 10.125 +                                                                                   0  206    0.   100.0
*LAST    +                                                                                   0  206    0.   100.0
            +----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+----+
               5    10    15    20    25    30    35    40    45    50    55    60    65    70    75    80
```

FIG (3.27)

HISTOGRAM OF LAC GIVEN FIVE CLAIMS

	SYMBOL	COUNT	MEAN	ST.DEV.
5	E	157	8.290	0.712

INTERVAL NAME	5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80	FREQUENCY INT.	FREQUENCY CUM.	PERCENTAGE INT.	PERCENTAGE CUM.
4.1250		0	0	0.	0.
4.2500		0	0	0.	0.
4.3750		0	0	0.	0.
4.5000		0	0	0.	0.
4.6250		0	0	0.	0.
4.7500		0	0	0.	0.
4.8750		0	0	0.	0.
5.0000		0	0	0.	0.
5.1250		0	0	0.	0.
5.2500		0	0	0.	0.
5.3750		0	0	0.	0.
5.5000		0	0	0.	0.
5.6250		0	0	0.	0.
5.7500		0	0	0.	0.
5.8750		0	0	0.	0.
6.0000		0	0	0.	0.
6.1250		0	0	0.	0.
6.2500		0	0	0.	0.
6.3750		0	0	0.	0.
6.5000		0	0	0.	0.
6.6250		0	0	0.	0.
6.7500		0	0	0.	0.
6.8750	E	1	1	0.6	0.6
7.0000		0	1	0.	0.6
7.1250		0	1	0.	0.6
7.2500	EEEEE	5	6	3.2	3.8
7.3750	EEEEEE	6	12	3.8	7.6
7.5000	EEEEEE	6	18	3.8	11.5
7.6250	EEEEEEEEEEE	11	29	7.0	18.5
7.7500	EEEEEEEEEEEEEEEEE	17	46	10.8	29.3
7.8750	EEEEEEEEE	9	55	5.7	35.0
8.0000	EEEEEEEEE	9	64	5.7	40.8
8.1250	EEEEEEEEEEEE	12	76	7.6	48.4
8.2500	EEEEEEEEEE	10	86	6.4	54.8
8.3750	EEEEEEEEEEEEE	13	99	8.3	63.1
8.5000	EEEEE	5	104	3.2	66.2
8.6250	EE	2	106	1.3	67.5
8.7500	EE	2	108	1.3	68.8
8.8750		0	108	0.	68.8
9.0000	EEEEEEEEEE	10	118	6.4	75.2
9.1250	EEEEEEEEEEEEEEEEEEE	19	137	12.1	87.3
9.2500	EEEEE	5	142	3.2	90.4
9.3750		0	142	0.	90.4
9.5000	EE	2	144	1.3	91.7
9.6250	EEEEEEEEEEE	11	155	7.0	98.7
9.7500		0	155	0.	98.7
9.8750	E	1	156	0.6	99.4
10.000	E	1	157	0.6	100.0
10.125		0	157	0.	100.0
LAST		0	157	0.	100.0

5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80

respectively, where the letters A, B, C, D and E were used to represent the distributions of LAC given that NCLA = k, for k = 1 to 5, respectively.

With the exception of Fig. (3.23), that is, the distribution of LAC given NCLA = 1, all histograms present clear gaps on their right hand side. As NCLA increases, the number of gaps seems to increase too, which suggests the existence of different populations (in a statistical sense) in the data.

The most likely explanation for this fact is the existence of bodily injury claims (which lead to large payments in general) mixed with third party material damage claims in the data which is being analysed. The average levels of these types of payments seem to be so different that even a logarithmic transformation was not able to cluster them around one single average value.

The existence of several gaps as NCLA increases can be explained in the light of the duality large-small claims since an average of three of such claims will be large if the three individual claims are large, moderate if there is a combination of small and large individual claims, and small if the three individual claims are all small. However, this will be only noticeable if the gap between large and small claims is large enough, which seems to be the case with the data under analysis.

The extension of this gap should have appeared in the distribution of LAC given that NCLA = 1 and the reason why it did not was the exclusion of abnormally large claims made in the beginning of the analysis (71 observations).

If one refers back to Figs. (3.2) and (3.3), it will become clear that most of those claims correspond to NCLA = 1. Although it seems to be true that they are estimated claims, when they are settled they will probably form the missing cluster of large claims quite apart from the others in Fig. (3.23), provided the estimation which was made (AC = 31,442 then LAC = 10.356) is reliable.

It is worth mentioning at this point that logarithmic transformations have been widely used to analyse third party motor insurance claims in the belief that by doing so, the effect of abnormally large claims could be studied together with all others. The Swedish data seems to contradict this belief and a separation of severe bodily injury claims seems to be necessary. As far as the data under analysis is concerned, such a separation is impossible as there is no identification of the type of the claim in the file.

So far no strong evidence was found that the rating factors significantly influence the average claim; therefore, most of the variation in AC is to be regarded as due to random fluctuation rather than to the rating factors themselves.

3.7 Analysis of zero claims

The observations with NCLA = 0 and therefore with SP = 0 (zero claims) were not considered in the previous analysis for the logarithms of the corresponding average claims could not be evaluated as they do not exist.

These observations will be studied separately now and one will start the analysis by requesting descriptive statistics of ESP for the sample consisting of zero claims (1974 observations) as well as for the whole observations.

The result was produced by the procedure CONDESCRIPTIVE of SPSS and the result is shown in Fig. (3.28).

The average exposure for zero claims is 6.697 whereas for the whole observations it is 440.268, this fact suggesting that zero claims are mainly due to low exposures rather than the influence of the rating factors themselves.

To check this assumption in a more appropriate way, the number of observations for each level of each rating factor in the sample of zero claims and in the whole file will be compared. To this end, the procedure FREQUENCIES of SPSS will be requested and the results are set out in Table (3.1), where an asterisk was used to point out extreme deviations from the overall proportion of observations in the sample and in the whole file (1974/5413 = 0.36).

The largest deviations occur for Z = 7 and M = 9, meaning that zero claims were much more frequent in the former class and much less in the latter. This fact is in reasonable agreement with previous results and suggests that average claims for Z = 7 are closer to zero than those for M = 9.

To a lesser degree, abnormal deviations were found for M = 4, M = 8, J = 1 (above the overall proportion) and for B = 7, J = 2 (below the overall proportion); however, definitive conclusions cannot be made due to possible random fluctuations as well as small average exposure for those observations.

FIG (3.28)

DESCRIPTIVE STATISTICS OF ESP

VARIABLE ESP

MEAN	6.697	STD ERROR	0.247	STD DEV	10.975
VARIANCE	120.458	KURTOSIS	21.265	SKEWNESS	3.836
RANGE	123.991	MINIMUM	0.008	MAXIMUM	124.000
SUM	13219.164				

VALID OBSERVATIONS - 1974 MISSING OBSERVATIONS - 0

VARIABLE ESP

MEAN	440.268	STD ERROR	29.849	STD DEV	2196.047
VARIANCE	**********	KURTOSIS	269.603	SKEWNESS	13.775
RANGE	66376.306	MINIMUM	0.008	MAXIMUM	66376.314
SUM	**********				

VALID OBSERVATIONS - 5413 MISSING OBSERVATIONS - 0

TABLE (3.1)

Frequencies of zero claims per rating factor

RATING FACTOR	LEVELS	NUMBER OF OBSERVATIONS		PROPORTION
		SAMPLE OF ZERO CLAIMS (I)	WHOLE FILE (J)	(I/J)
S	1	400	1127	.35
	2	325	1140	.29
	3	353	1106	.32
	4	434	1040	.42
	5	462	1000	.46
Z	1	229	799	.29
	2	248	816	.30
	3	227	813	.28
	4	226	846	.27
	5	344	761	.45
	6	278	784	.35
	7	422	594	.71 (*)
B	1	267	732	.36
	2	292	747	.39
	3	310	752	.41
	4	324	755	.43
	5	310	767	.40
	6	263	801	.33
	7	208	859	.24 (*)

TABLE (3.1)

continued

RATING	LEVELS	NUMBER OF OBSERVATIONS SAMPLE OF ZERO CLAIMS (I)	WHOLE FILE (J)	PROPORTION (I/J)
M	1	199	642	.31
	2	266	599	.44
	3	124	305	.41
	4	218	377	.58 (*)
	5	258	710	.36
	6	211	728	.29
	7	299	645	.46
	8	375	672	.56 (*)
	9	24	735	.03 (*)
J	1	885	1654	.54 (*)
	2	458	1849	.25 (*)
	3	631	1910	.33
TOTAL PER RATING FACTOR	:	1974	5413	.36

3.8 The distribution of the claim frequency

The purpose of this part of the investigation is to study the distribution of claim frequencies and to establish whether or not they are significantly influenced by the rating factors.

To this end, a new variable will be defined as follows :

Notation		Meaning
MN	=	(NCLA/ESP) × 100 ; claim frequency (in percentages of units of exposure)

The procedure 2D of BMDP was used in order to produce the condensed distribution of MN and the result is shown in Fig. (3.29). For convenience, the values of MN were rounded to the first digit. It can be noticed that the distribution is very long tailed and that almost all (99.6%) values of MN lie within the interval ranging from 0 to 100. The range of the remaining values is 900 (1000 - 100), nine times the range of the great majority of the observations. Considering that those abnormal observations will certainly distort the results of any further analysis based on average values (means), one will exclude those observations from the file, and will consider them separately.

A list of individual observations with MN > 100 was produced by the procedure LIST CASES of SPSS and is shown in Fig. (3.30), along with the descriptive statistics of ESP for that sample, produced by the procedure CONDESCRIPTIVE of SPSS.

There appears not to be any noticeable trend regarding the

FIG (3.29)

CONDENSED DISTRIBUTION OF MN

```
************
* MN       *
************
                                                 MAXIMUM        1000.0000000        H
VARIABLE NUMBER . . . . . .          12          MINIMUM           0.               H
NUMBER OF DISTINCT VALUES .          26          RANGE          1000.0000000        H
NUMBER OF VALUES COUNTED. .        5413          VARIANCE        487.2244301        H                                   EACH 'H'
NUMBER OF VALUES NOT COUNTED          0          ST.DEV.          22.0731609        H                                  REPRESENTS
***VALUES ARE ROUNDED TO. .     10.0000          (Q3-Q1)/2         5.0000000        H                                     530
                                                                                    H                                   COUNT(S)
LOCATION ESTIMATES                                    ST.ERROR                      H
                 MEAN              6.7651949          0.3000168                     H
                 MEDIAN            0.                 0.                            HHHHHHHH     HH                  H
                 MODE              0.                                              L-----------------------------------U

                                                                                               EACH '-' ABOVE =      30.0000
                                                                                                             L=       0.
                                                                                                             U=    1110.0000
 2649 CASES TOO MANY TO COMPUTE THE OPTIONAL ESTIMATES

                                                                                               EACH '.' BELOW =      10.0000
            SQQ S
            -MM +                                                                                                  M
            .IE..................................................................................................A
             NA                                                                                                    X
              N
```

VALUE	COUNT	PERCENTS CELL	PERCENTS CUM	VALUE	COUNT	PERCENTS CELL	PERCENTS CUM	VALUE	COUNT	PERCENTS CELL	PERCENTS CUM	VALUE	COUNT	PERCENTS CELL	PERCENTS CUM
0.	3213	59.4	59.4	70.	7	0.1	99.4	140.	2	0.0	99.7	230.	1	0.0	99.9
10.	1665	30.8	90.1	80.	6	0.1	99.5	150.	2	0.0	99.8	400.	1	0.0	99.9
20.	312	5.8	95.9	90.	5	0.1	99.6	160.	2	0.0	99.8	440.	1	0.0	99.9
30.	104	1.9	97.8	100.	1	0.0	99.6	170.	1	0.0	99.8	450.	2	0.0	100.0
40.	37	0.7	98.5	110.	3	0.1	99.6	180.	1	0.0	99.8	1000.	1	0.0	100.0
50.	30	0.6	99.0	120.	2	0.0	99.7	190.	1	0.0	99.9				
60.	10	0.2	99.2	130.	1	0.0	99.7	200.	2	0.0	99.9				

FIG (3.30)

UNUSUAL VALUES OF MN AND DESCRIPTIVE

STATISTICS OF ESP

CASE-N	S	Z	B	M	J	ESP	NCLA	SP	MN
1	1.	1.	6.	7.	1.	0.582	1.	3867.	171.70
2	1.	3.	1.	1.	1.	0.490	1.	1813.	203.91
3	1.	5.	2.	5.	1.	0.890	1.	4831.	112.31
4	1.	7.	4.	2.	2.	0.917	1.	5000.	109.06
5	2.	1.	1.	1.	1.	0.456	2.	5154.	438.44
6	2.	1.	3.	8.	1.	2.188	3.	13092.	137.11
7	2.	4.	1.	8.	1.	0.688	1.	5960.	145.27
8	2.	4.	5.	2.	1.	0.615	1.	9969.	162.58
9	2.	5.	1.	8.	3.	1.800	2.	6656.	111.10
10	2.	6.	4.	7.	1.	0.838	1.	4603.	119.34
11	2.	6.	5.	3.	3.	0.222	1.	3521.	450.00
12	3.	1.	1.	1.	1.	0.100	1.	551.	1000.00
13	3.	3.	7.	2.	1.	0.442	1.	31442.	226.42
14	3.	6.	4.	7.	1.	0.670	1.	1019.	149.18
15	3.	7.	1.	5.	3.	0.545	1.	906.	183.42
16	3.	7.	4.	7.	2.	1.492	3.	11762.	201.10
17	3.	7.	6.	5.	1.	0.780	1.	575.	128.29
18	4.	2.	1.	6.	1.	0.825	1.	1367.	121.24
19	4.	4.	1.	8.	1.	0.522	1.	684.	191.54
20	4.	6.	3.	7.	1.	0.250	1.	3535.	400.00
21	4.	7.	6.	5.	1.	0.696	1.	5000.	143.70
22	5.	3.	1.	8.	3.	1.250	2.	7451.	160.02
23	5.	7.	4.	6.	2.	0.958	1.	1231.	104.44
24	5.	7.	5.	7.	2.	0.225	1.	2935.	445.04

AFTER READING 5413 CASES FROM SUBFILE NONAME , END OF DATA WAS ENCOUNTERED ON LOGICAL UNIT # 8

VARIABLE ESP

MEAN	0.768	STD ERROR	0.102	STD DEV	0.498
VARIANCE	0.248	KURTOSIS	2.094	SKEWNESS	1.376
RANGE	2.088	MINIMUM	0.100	MAXIMUM	2.188
SUM	18.441				

VALID OBSERVATIONS - 24 MISSING OBSERVATIONS - 0

values of the rating factors, but the same cannot be said for the individual exposure values, which are abnormally small, even if one considers that the average number of claims for these observations is small. In fact, if one looks at Fig. (3.31), where the distributions of exposures were obtained for small values of NCLA (ranging from 0 to 9), it will become clear that the average exposure for the observations with MN > 100, which is 0.768, is far below the average exposures for observations with small numbers of claims, which are in the range from 6.697 to 196.844.

So it does not seem unreasonable to treat the observations with MN > 100 as outliers, as they represent instances of claims occuring for very short periods of time from the inception of the corresponding policies. The data show that these claims, although not impossible, are quite unusual, and therefore their exclusion is justifiable on the grounds that what is desired is a representation of the average behaviour of MN.

Considering then only the observations with MN $\leqslant$ 100, a more detailed histogram was produced for MN with the aid of the procedure 5D of BMDP. The result is shown in Fig. (3.32) where the intervals were made 2 units wide and a scale factor of 1 : 5 (each "X" = 5 counts) was chosen. Some intervals contain more observations than the selected limit (400), so an asterix was printed by the program and the actual number of observations appear in the first column on the right hand side of the picture.

The distribution of MN is discontinuous, as there is a concentration of 1974 observations for MN = 0 (zero claims), and a probabilistic

FIG (3.31)

HISTOGRAMS OF ESP FOR SMALL VALUES OF NCLA

```
         * 0.0000   * 1.0000   * 2.0000   * 3.0000   * 4.0000   * 5.0000   * 6.0000   * 7.0000   * 8.0000   * 9.0000
         ..........+..........+..........+..........+..........+..........+..........+..........+..........+..........+
MIDPOINTS
 750.000)
 720.000)
 690.000)
 660.000)
 630.000)                                                                                        *
 600.000)
 570.000)                                                                             *                     *
 540.000)                                                                                                   *
 510.000)                                                                                        *          ***
 480.000)                                                                                                   **
 450.000)                                                                                        *
 420.000)                                                                                                   **
 390.000)                                                                             *          *          **
 360.000)                                 **                                          *          *
 330.000)                                                                  *          *                     *
 300.000)                                                                  **         *          *          ****
 270.000)           *                     *                     *          *****      ***        ******     ***
 240.000)           *                                           **         ***        ********   ****       *******
 210.000)                                 *          ****       ****       *****      ***        ******     M**
 180.000)                      *          *******    ********11 ***        ********   ********11 *******    ********
 150.000)           **         ****       ****       ********13 ********15 ********19 ********15 M*******   ********
 120.000)*          ***        ********** ********16 ********15 ********20 M********  M*******18 *****      **********
  90.000)********   ********13 ********23 ********25 ********39 M*******32 ********24 ********17 ********20 ********11
  60.000)********23 ********49 ********74 M*******57 M*******46 ********38 ********29 ********24 ********17 **********
  30.000)*******202 M******235 M******173 *******124 ********68 ********41 ********29 ********11 *********  **
   0.   )M*****1740 *******387 *******109 ********30 ********** *          **
 -30.000)
 -60.000)
 -90.000)
-120.000)
GROUP MEANS ARE DENOTED BY M'S IF THEY COINCIDE WITH *'S, N'S OTHERWISE
```

MEAN	6.697	20.412	35.449	53.728	72.065	82.990	105.488	130.346	143.260	196.844
STD.DEV.	10.975	24.093	29.441	49.023	49.460	51.360	71.174	86.044	110.448	135.578
S. E. M.	0.247	0.917	1.483	3.000	3.446	4.099	6.103	8.024	11.707	15.351
MAXIMUM	124.000	258.547	176.817	367.454	214.424	255.989	336.016	563.205	628.560	583.864
MINIMUM	0.008	0.100	0.456	1.492	5.816	10.772	7.823	18.463	16.182	34.007
SAMPLE SIZE	1974	691	394	267	206	157	136	115	89	78

ALL GROUPS COMBINED
(EXCEPT CASES WITH UNUSED VALUES
FOR NCLA)

MEAN	34.320
STD.DEV.	58.016
S. E. M.	0.905
MAXIMUM	628.560
MINIMUM	0.008
SAMPLE SIZE	4107

ANALYSIS OF VARIANCE TABLE

SOURCE	SUM OF SQUARES	DF	MEAN SQUARE	F VALUE	TAIL PROBABILITY
BETWEEN GROUPS	7272169.0625	9	808018.7813	505.58	0.
WITHIN GROUPS	6547819.7363	4097	1598.1986		
TOTAL	************	4106			
LEVENE'S TEST FOR EQUAL VARIANCES		9,4097		296.21	0.

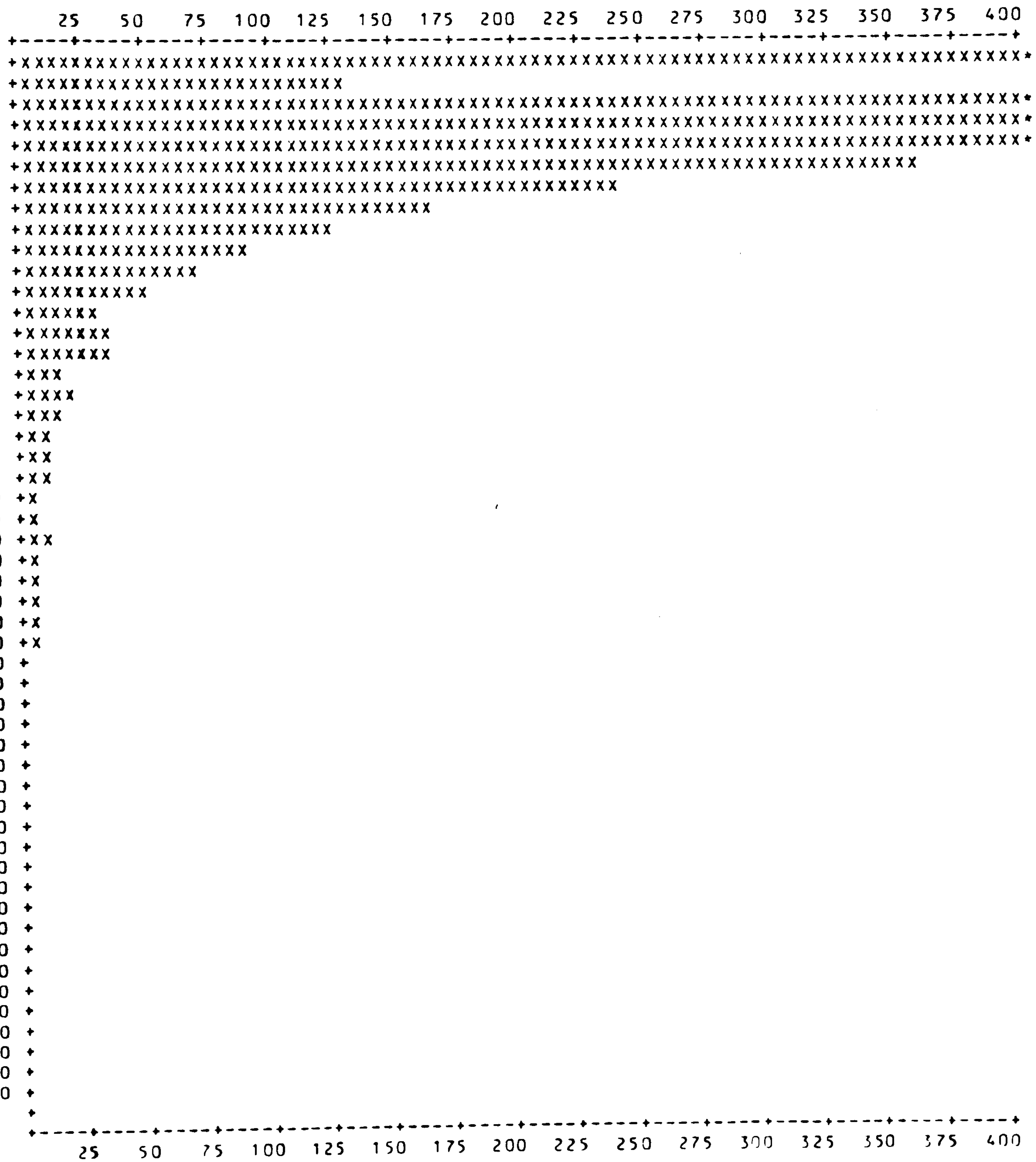
ONE-WAY ANALYSIS OF VARIANCE
TEST STATISTICS FOR WITHIN-GROUP
VARIANCES NOT ASSUMED TO BE EQUAL

	DF	F VALUE	TAIL PROBABILITY
WELCH	9, 558	241.99	0.
BROWN-FORSYTHE	9, 389	144.01	0.

FIG (3.32)

HISTOGRAM OF MN

SYMBOL X COUNT 5389 MEAN 6.006 ST.DEV. 8.981



INTERVAL NAME	FREQUENCY INT.	FREQUENCY CUM.	PERCENTAGE INT.	PERCENTAGE CUM.
0.0000	1974	1974	36.6	36.6
2.0000	131	2105	2.4	39.1
4.0000	677	2782	12.6	51.6
6.0000	771	3553	14.3	65.9
8.0000	486	4039	9.0	74.9
10.000	361	4400	6.7	81.6
12.000	240	4640	4.5	86.1
14.000	166	4806	3.1	89.2
16.000	126	4932	2.3	91.5
18.000	92	5024	1.7	93.2
20.000	69	5093	1.3	94.5
22.000	49	5142	0.9	95.4
24.000	31	5173	0.6	96.0
26.000	34	5207	0.6	96.6
28.000	33	5240	0.6	97.2
30.000	14	5254	0.3	97.5
32.000	19	5273	0.4	97.8
34.000	14	5287	0.3	98.1
36.000	12	5299	0.2	98.3
38.000	8	5307	0.1	98.5
40.000	11	5318	0.2	98.7
42.000	6	5324	0.1	98.8
44.000	3	5327	0.1	98.8
46.000	11	5338	0.2	99.1
48.000	4	5342	0.1	99.1
50.000	7	5349	0.1	99.3
52.000	5	5354	0.1	99.4
54.000	6	5360	0.1	99.5
56.000	4	5364	0.1	99.5
58.000	2	5366	0.0	99.6
60.000	0	5366	0.	99.6
62.000	2	5368	0.0	99.6
64.000	2	5370	0.0	99.6
66.000	1	5371	0.0	99.7
68.000	2	5373	0.0	99.7
70.000	1	5374	0.0	99.7
72.000	1	5375	0.0	99.7
74.000	2	5377	0.0	99.8
76.000	2	5379	0.0	99.8
78.000	2	5381	0.0	99.9
80.000	1	5382	0.0	99.9
82.000	0	5382	0.	99.9
84.000	1	5383	0.0	99.9
86.000	2	5385	0.0	99.9
88.000	1	5386	0.0	99.9
90.000	2	5388	0.0	100.0
92.000	1	5389	0.0	100.0
94.000	0	5389	0.	100.0
96.000	0	5389	0.	100.0
98.000	0	5389	0.	100.0
100.00	0	5389	0.	100.0
LAST	0	5389	0.	100.0

representation of it should be of the mixed type (a continuous function unless for a finite number of points). However, no attempt will be made to fit any distribution for MN before a preliminary assessment of the influence of the rating factors on this variable has been made.

3.9 Influence of rating factors on MN

The procedure 7D of BMDP was used in order to produce side by side histograms of MN for each level of each of the five rating factors.

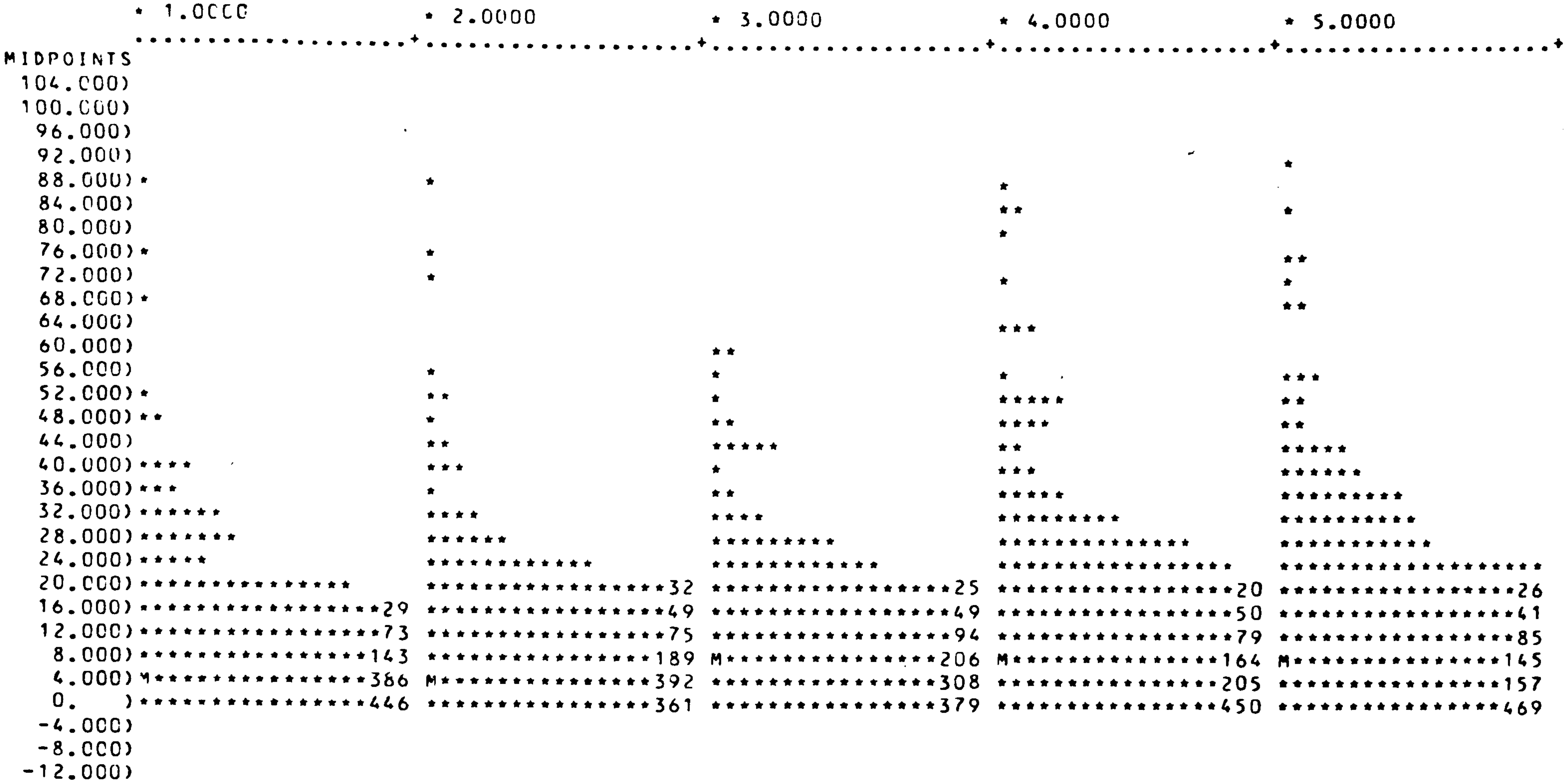
The result for the rating factor S is shown in Fig. (3.33), where it can be noticed that the average value of MN increases as the values representing the levels of S increase. This trend confirms the intuitive fact that the proportion of incurred claims is higher for policy-holders who use their vehicles more than for those who use them less.

No attempt will be made to interpret the remaining statistics shown in Fig. (3.33) as the chief interest at this stage of the analysis is just to establish whether or not a relationship is likely to exist between the claim frequency and the rating factors, a fact which seems to be true with regard to S.

The histograms for Z are shown in Fig. (3.34), where four distinct levels can be observed for the average claim frequency ($\overline{MN}$), namely :

FIG (3.33)

HISTOGRAMS OF MN FOR EACH LEVEL OF S



GROUP MEANS ARE DENOTED BY M'S IF THEY COINCIDE WITH *'S, N'S OTHERWISE

MEAN	4.822	5.890	6.014	6.663	6.778
STD.DEV.	7.468	7.851	7.587	10.595	10.986
S. E. M.	0.223	0.233	0.229	0.329	0.348
MAXIMUM	89.553	87.161	60.220	89.407	90.986
MINIMUM	0.	0.	0.	0.	0.
SAMPLE SIZE	1123	1133	1100	1036	997

ALL GROUPS COMBINED
(EXCEPT CASES WITH UNUSED VALUES
FOR S)

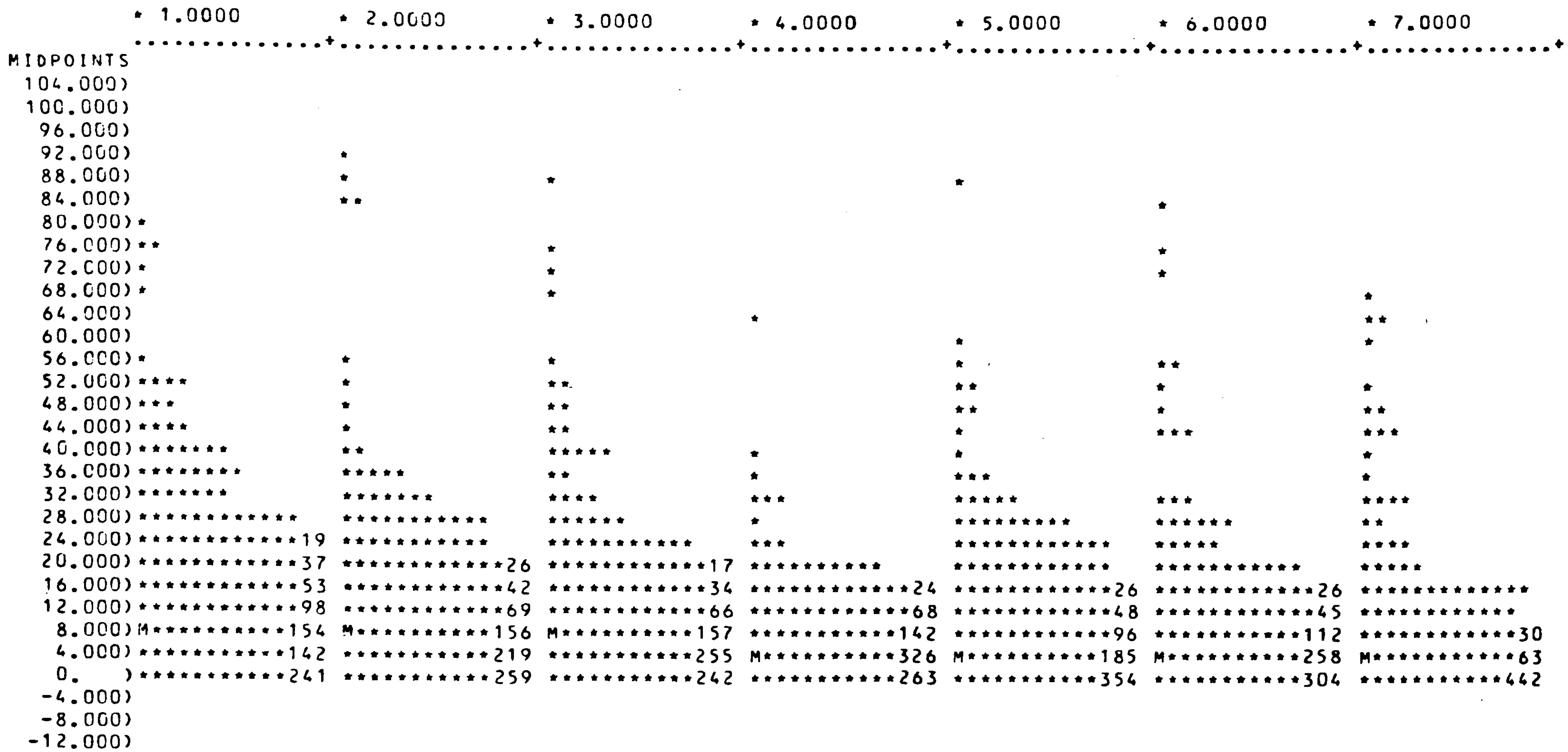
MEAN	6.006
STD.DEV.	8.981
S. E. M.	0.122
MAXIMUM	90.986
MINIMUM	0.
SAMPLE SIZE	5389

ANALYSIS OF VARIANCE TABLE

SOURCE	SUM OF SQUARES	DF	MEAN SQUARE	F VALUE	TAIL PROBABILITY
BETWEEN GROUPS	2630.0473	4	657.5118	8.19	0.0000
WITHIN GROUPS	431994.1758	5384	80.2367		
TOTAL	434624.2227	5388			
LEVENE'S TEST FOR EQUAL VARIANCES		4,5384	--	30.61	0.
ONE-WAY ANALYSIS OF VARIANCE TEST STATISTICS FOR WITHIN-GROUP VARIANCES NOT ASSUMED TO BE EQUAL					
WELCH		4,2644		8.74	0.0000
BROWN-FORSYTHE		4,4589		8.03	0.0000

FIG (3.34)

HISTOGRAMS OF MN FOR EACH LEVEL OF Z



GROUP MEANS ARE DENOTED BY M'S IF THEY COINCIDE WITH *'S, N'S OTHERWISE

	1	2	3	4	5	6	7
MEAN	9.004	7.102	6.674	4.890	5.297	5.192	3.101
STD.DEV.	11.033	9.658	9.192	5.475	8.554	8.289	8.674
S. E. M.	0.391	0.338	0.323	0.189	0.310	0.297	0.358
MAXIMUM	79.562	90.986	89.553	64.516	87.161	83.527	67.112
MINIMUM	0.	0.	0.	0.	0.	0.	0.
SAMPLE SIZE	795	815	810	843	759	780	587

ALL GROUPS COMBINED
(EXCEPT CASES WITH UNUSED VALUES FOR Z)

MEAN	6.006
STD.DEV.	8.981
S. E. M.	0.122
MAXIMUM	90.986
MINIMUM	0.
SAMPLE SIZE	5389

ANALYSIS OF VARIANCE TABLE

SOURCE	SUM OF SQUARES	DF	MEAN SQUARE	F VALUE	TAIL PROBABILITY
BETWEEN GROUPS	15388.8074	6	2564.8012	32.93	0.
WITHIN GROUPS	419235.4304	5382	77.8958		
TOTAL	434624.2383	5388			
LEVENE'S TEST FOR EQUAL VARIANCES		6,5382		23.87	0.
ONE-WAY ANALYSIS OF VARIANCE TEST STATISTICS FOR WITHIN-GROUP VARIANCES NOT ASSUMED TO BE EQUAL					
WELCH		6,2321		28.65	0.
BROWN-FORSYTHE		6,4806		32.83	0.

$\overline{MN} \cong 9$, for the three biggest cities in Sweden (Z = 1).

$\overline{MN} \cong 7$, for other big cities with exception of Gotland and small cities in southern Sweden (Z = 2 and Z = 3).

$\overline{MN} \cong 5$, for small cities in northern Sweden and rural areas (Z = 4, Z = 5, Z = 6).

$\overline{MN} \cong 3$, for Gotland (Z = 7)

The above figures show that the average claim frequency increases with the populational density in a given region. It is worth mentioning that the southern part of Sweden is more populated than the northern part, and this is possibly the reason why their small cities were ranked in different classes above. As seen before, Gotland appears as a special city, with a remarkably low average claim frequency, and the reason for this can possibly be the peculiar way in which the city itself was planned.

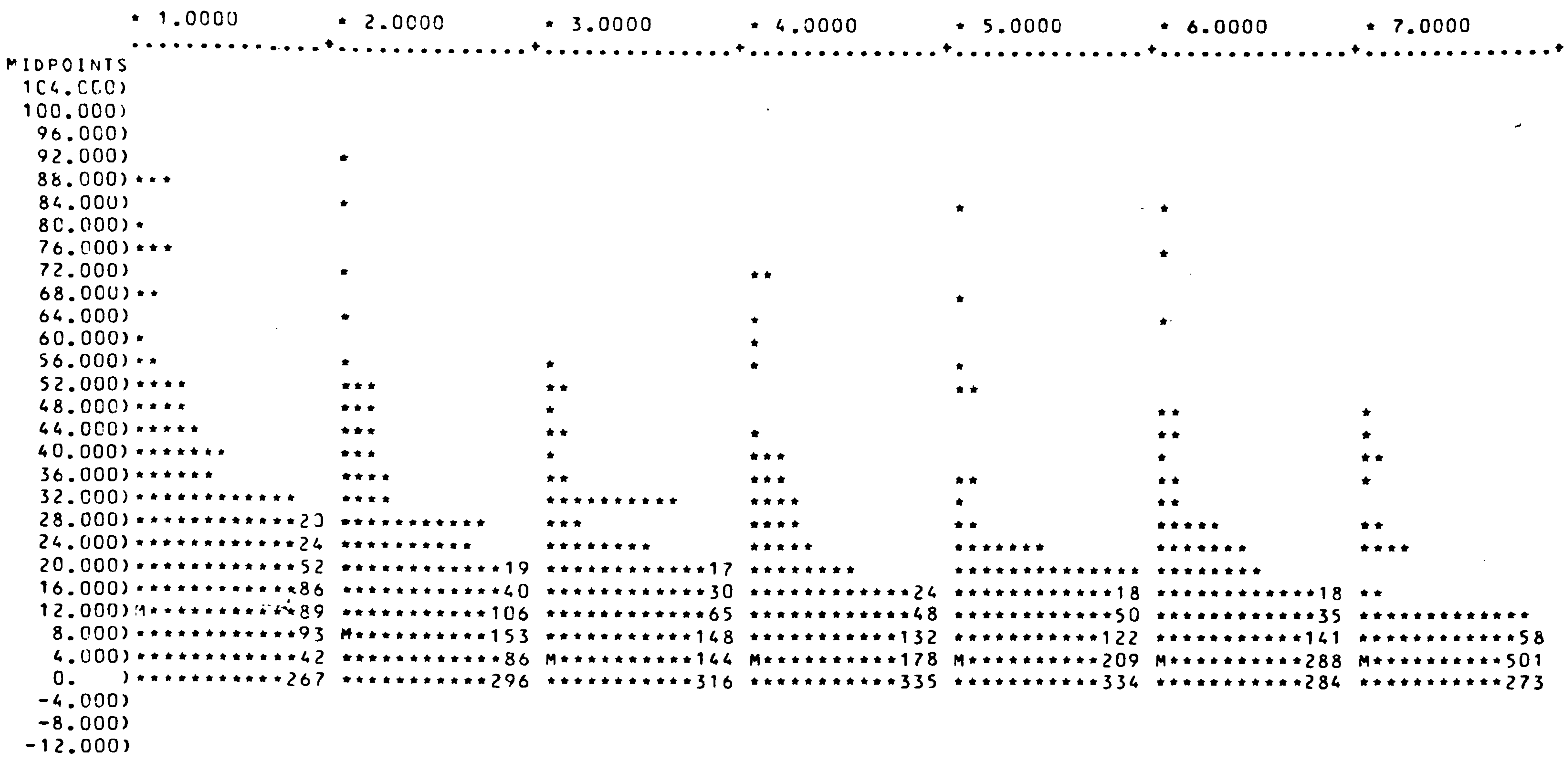
The histograms for B are displayed in Fig. (3.35), where a clear trend can be observed in the averages of MN, namely they decrease as the bonus discount increases (there is one exception regarding the average for B = 6 which is slightly greater than that for B = 5. However, this can be attributed as being due to random fluctuation).

The above mentioned trend agrees with beliefs derived from practical experience, in the sense that policyholders with high bonus discounts tend not to report small claims in order not to lose their future discounts.

The results for the rating factor M are presented in Fig. (3.36), where considerable differences in the averages of MN can be once more observed. The averages are ranked below along with the

FIG (3.35)

HISTOGRAMS OF MN FOR EACH LEVEL OF B



GROUP MEANS ARE DENOTED BY M'S IF THEY COINCIDE WITH *'S, N'S OTHERWISE

MEAN	10.953	7.529	5.816	5.129	4.793	5.034	3.427
STD.DEV.	13.537	10.323	7.836	8.056	7.249	7.519	4.260
S. E. M.	0.503	0.378	0.286	0.294	0.262	0.266	0.145
MAXIMUM	89.553	90.986	55.075	72.995	83.527	85.713	48.193
MINIMUM	0.	0.	0.	0.	0.	0.	0.
SAMPLE SIZE	723	746	750	750	764	798	858

ALL GROUPS COMBINED
(EXCEPT CASES WITH UNUSED VALUES
FOR B)

MEAN	6.006
STD.DEV.	8.981
S. E. M.	0.122
MAXIMUM	90.986
MINIMUM	0.
SAMPLE SIZE	5389

ANALYSIS OF VARIANCE TABLE

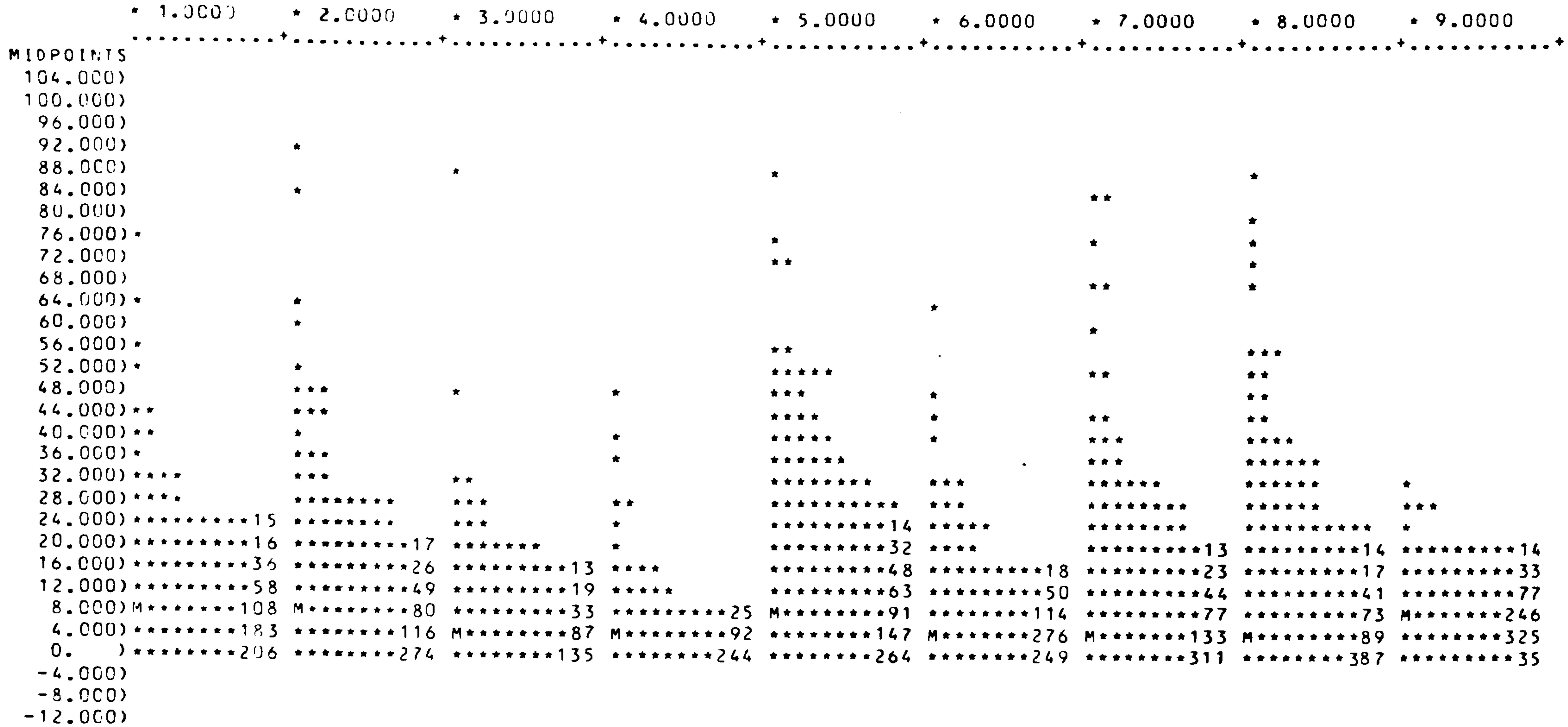
SOURCE	SUM OF SQUARES	DF	MEAN SQUARE	F VALUE	TAIL PROBABILITY
BETWEEN GROUPS	27616.3538	6	4602.7256	60.86	0.
WITHIN GROUPS	407007.8859	5382	75.6239		
TOTAL	434624.2383	5388			
LEVENE'S TEST FOR EQUAL VARIANCES		6, 5382		93.94	0.

ONE-WAY ANALYSIS OF VARIANCE
TEST STATISTICS FOR WITHIN-GROUP
VARIANCES NOT ASSUMED TO BE EQUAL

WELCH	6, 2312	50.81	0.
BROWN-FORSYTHE	6, 3651	59.04	0.

FIG (3.36)

HISTOGRAMS OF MN FOR EACH LEVEL OF M



GROUP MEANS ARE DENOTED BY M'S IF THEY COINCIDE WITH *'S, N'S OTHERWISE

	1.0000	2.0000	3.0000	4.0000	5.0000	6.0000	7.0000	8.0000	9.0000
MEAN	6.648	6.296	5.064	2.435	8.179	4.697	5.906	5.709	6.994
STD.DEV.	8.355	10.183	8.232	5.224	11.514	5.989	10.471	11.231	4.278
S. E. M.	0.331	0.417	0.472	0.269	0.433	0.222	0.414	0.435	0.158
MAXIMUM	74.809	90.986	89.553	48.022	87.161	62.767	85.713	89.407	31.978
MINIMUM	0.	0.	0.	0.	0.	0.	0.	0.	0.
SAMPLE SIZE	639	596	304	377	706	726	639	667	735

ALL GROUPS COMBINED
(EXCEPT CASES WITH UNUSED VALUES FOR M)

MEAN	6.006
STD.DEV.	8.981
S. E. M.	0.122
MAXIMUM	90.986
MINIMUM	0.
SAMPLE SIZE	5389

ANALYSIS OF VARIANCE TABLE

SOURCE	SUM OF SQUARES	DF	MEAN SQUARE	F VALUE	TAIL PROBABILITY
BETWEEN GROUPS	10750.9878	8	1343.8735	17.06	0.
WITHIN GROUPS	423873.2657	5380	78.7869		
TOTAL	434624.2539	5388			
LEVENE'S TEST FOR EQUAL VARIANCES		8,5380		42.43	0.
ONE-WAY ANALYSIS OF VARIANCE TEST STATISTICS FOR WITHIN-GROUP VARIANCES NOT ASSUMED TO BE EQUAL					
WELCH		8,1988		34.80	0.
BROWN-FORSYTHE		8,4217		17.56	0.

corresponding relative engine sizes :

average MN	level of M	relative engine size
2.435	4	1.00
4.697	6	1.60
5.064	3	2.43
5.709	8	1.43
5.906	7	2.17
6.296	2	2.40
6.648	1	2.20
6.994	9	-
8.179	5	2.23

The figures above suggest a relationship between the power of the vehicle and average claims frequency in the sense that the more powerful vehicles tend to produce higher claims frequencies. However, the relationship is not absolutely clear as the most powerful vehicle (M = 3) was ranked in the third place. Other characteristics of the vehicles might be interferring - for example, design features such as braking reliability, windscreen visibility, etc.

As regards the rating factor J, their associated histograms are shown in Fig. (3.37), where two levels of the average claims frequencies can be clearly distinguished :

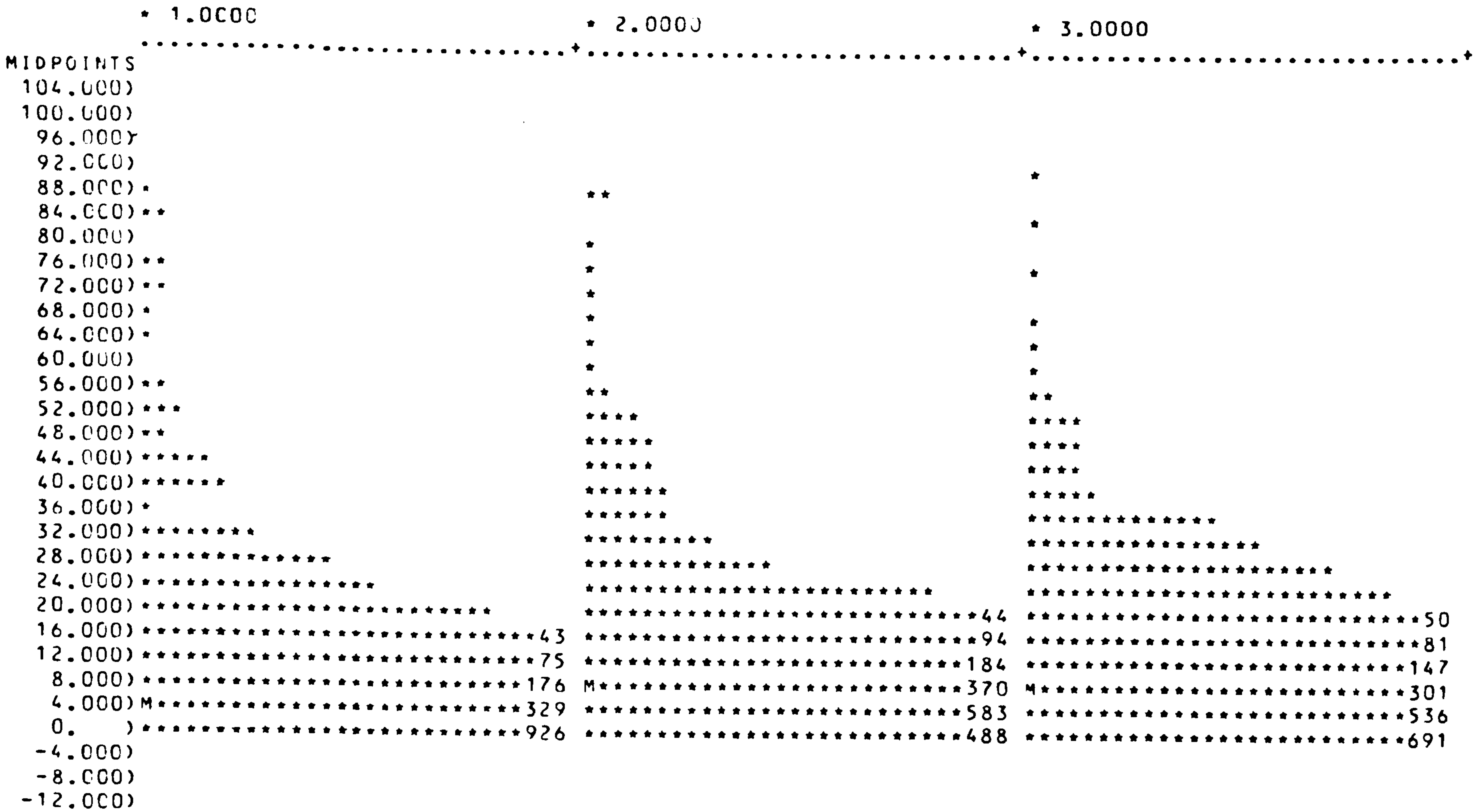
$\overline{MN} \cong 4$, for new (less than 3 years old) vehicles (J = 1)

$\overline{MN} \cong 6.5$, for old (more than 3 years old) vehicles (J = 2 and J = 3).

As expected, the average claim frequency is smaller for new

FIG (3.37)

HISTOGRAMS OF MN FOR EACH LEVEL OF J



GROUP MEANS ARE DENOTED BY M'S IF THEY COINCIDE WITH *'S, N'S OTHERWISE

	1.0000	2.0000	3.0000
MEAN	4.465	7.027	6.342
STD.DEV.	9.035	8.837	8.902
S. E. M.	0.223	0.206	0.204
MAXIMUM	87.161	89.553	90.986
MINIMUM	0.	0.	0.
SAMPLE SIZE	1638	1845	1906

ALL GROUPS COMBINED
(EXCEPT CASES WITH UNUSED VALUES FOR J)

MEAN	6.006
STD.DEV.	8.981
S. E. M.	0.122
MAXIMUM	90.986
MINIMUM	0.
SAMPLE SIZE	5389

ANALYSIS OF VARIANCE TABLE

SOURCE	SUM OF SQUARES	DF	MEAN SQUARE	F VALUE	TAIL PROBABILITY
BETWEEN GROUPS	6027.1321	2	3013.5660	37.87	0.
WITHIN GROUPS	428597.1094	5386	79.5761		
TOTAL	434624.2422	5388			
LEVENE'S TEST FOR EQUAL VARIANCES		2,5386		2.74	0.0647
ONE-WAY ANALYSIS OF VARIANCE TEST STATISTICS FOR WITHIN-GROUP VARIANCES NOT ASSUMED TO BE EQUAL					
WELCH		2,3551		37.47	0.
BROWN-FORSYTHE		2,5316		37.81	0.

vehicles than for old ones. The explanation for this is likely to be that policy holders tend to be more careful when driving new cars either because they know that repair costs are much higher for new vehicles, or because they do not want to lose the status of possessing a car in a perfect condition.

3.10 Further considerations about the distribution of the claim frequency

It was shown in the previous section that the rating factors do influence the frequency of claims and the aim of the next sections will be to quantify those influences. One possible statistical approach to measuring how a particular set of discrete variables affects a given continuous one is to perform a regression analysis in which the levels of the discrete variables are treated as dummy variables. However, traditional regression analysis requires normality of the dependent variable (amongst other assumptions), which is not the case with MN. Even the generalized approach to regression, in which the distribution of the dependent variable can be a member of an exponential family (binomial, Poisson, gamma etc.), is not applicable due to the fact that the distribution of MN presents a single discontinuity in the origin, as shown in section 3.8.

One possible way around this problem is to remove zero claims from the analysis, remembering that a separate study of them has already been made in Section 3.7. Therefore, the resulting distribution under analysis will be a conditional one, namely the distribution of MN given that NCLA > 0.

If one refers back to Fig. (3.32), it can be noticed that if zero claims are removed from the distribution, its shape will resemble the lognormal. This fact justifies a naperian logarithmic transformation on MN (given NCLA > 0) as defined below :

Notation		Meaning
LMN	=	ℓn(MN) ; Naperian logarithm of MN,

and its histogram was obtained by using the procedure 5D of BMDP. The result is shown in Fig. (3.38), where the intervals were made 0.1 units wide and a scale factor of 1 : 3 (each "X" represents 3 counts) was chosen. The resulting shape of the distribution strongly suggests normality, and this assumption will be checked with the aid of a normal plot.

Using the procedure 5D of BMDP once more, a normal plot for LMN was produced and the result is shown in Fig. (3.39). A clear linear trend can be observed, which confirms that the hypothesis of normality is a plausible one for LMN (given NCLA > 0).

3.11 Quantifying the influence of the rating factors on LMN

Under the assumption of normality for LMN (given NCLA > 0), the effects of the levels of the rating factors in explaining the variation of LMN can be assessed by fitting a linear model in which the dependent variable is LMN itself and the independent ones are dummy variables representing the levels of the rating factors.

The same notation as in Section 3.5 will be used for the dummy

FIG (3.38)

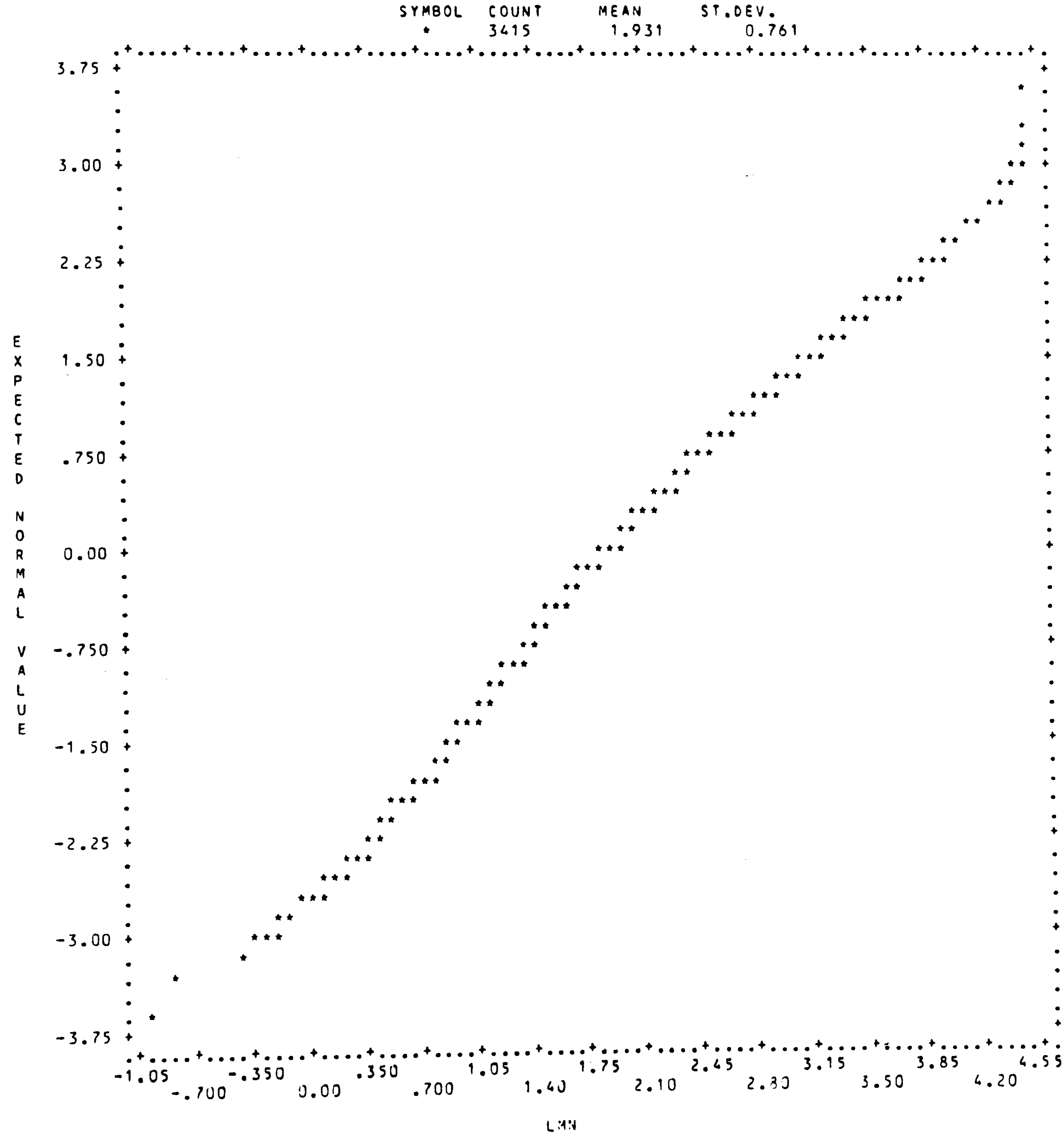
HISTOGRAM OF LMN (GIVEN POSITIVE CLAIMS)

SYMBOL	COUNT	MEAN	ST.DEV.
X	3415	1.931	0.761

INTERVAL NAME	FREQUENCY INT.	FREQUENCY CUM.	PERCENTAGE INT.	PERCENTAGE CUM.
.10000	16	16	0.5	0.5
.20000	4	20	0.1	0.6
.30000	9	29	0.3	0.8
.40000	13	42	0.4	1.2
.50000	22	64	0.6	1.9
.60000	36	100	1.1	2.9
.70000	33	133	1.0	3.9
.80000	48	181	1.4	5.3
.90000	61	242	1.8	7.1
1.0000	69	311	2.0	9.1
1.1000	101	412	3.0	12.1
1.2000	122	534	3.6	15.6
1.3000	126	660	3.7	19.3
1.4000	175	835	5.1	24.5
1.5000	189	1024	5.5	30.0
1.6000	195	1219	5.7	35.7
1.7000	178	1397	5.2	40.9
1.8000	191	1588	5.6	46.5
1.9000	168	1756	4.9	51.4
2.0000	178	1934	5.2	56.6
2.1000	177	2111	5.2	61.8
2.2000	156	2267	4.6	66.4
2.3000	155	2422	4.5	70.9
2.4000	144	2566	4.2	75.1
2.5000	120	2686	3.5	78.7
2.6000	104	2790	3.0	81.7
2.7000	104	2894	3.0	84.7
2.8000	79	2973	2.3	87.1
2.9000	85	3058	2.5	89.5
3.0000	64	3122	1.9	91.4
3.1000	49	3171	1.4	92.9
3.2000	35	3206	1.0	93.9
3.3000	44	3250	1.3	95.2
3.4000	30	3280	0.9	96.0
3.5000	27	3307	0.8	96.8
3.6000	19	3326	0.6	97.4
3.7000	20	3346	0.6	98.0
3.8000	10	3356	0.3	98.3
3.9000	16	3372	0.5	98.7
4.0000	14	3386	0.4	99.2
4.1000	8	3394	0.2	99.4
4.2000	4	3398	0.1	99.5
4.3000	5	3403	0.1	99.6
4.4000	5	3408	0.1	99.8
4.5000	6	3414	0.2	100.0
4.6000	1	3415	0.0	100.0
4.7000	0	3415	0.	100.0
4.8000	0	3415	0.	100.0
4.9000	0	3415	0.	100.0
5.0000	0	3415	0.	100.0
LAST	0	3415	0.	100.0

FIG (3.39)

NORMAL PLOT FOR LMN (GIVEN POSITIVE CLAIMS)



variables, and the reference levels will be chosen as the ones with smallest average claim frequency (see Figs. (3.33) to (3.37)) which are : S = 1, Z = 7, B = 7, M = 4 and J = 1.

The procedure REGRESSION of SPSS was used to fit the desired linear model and the result is shown in Fig. (3.40). It can be noticed that the squared multiple correlation coefficient (R SQUARE) has a considerable value of 0.52687 which means that the rating factors account for 52.7% of the variation in LMN, leaving the remaining proportion of variation attributable to random fluctuation (not explained by the model). This degree of explanation can be considered as representing a good fit, especially if one takes into consideration the number of observations which is quite large (3415 observations).

The plot of residuals in Fig. (3.41) shows no abnormal trend ; that is, the spreading of the observations throughout the plot seems to be uniform. This is a further indication that the model fits reasonably with the data.

Referring back to Fig. (3.40), one can notice in the analysis of variance table that the overall F is quite large (145.10882) which confers statistical significance to the coefficients of the model, which are printed in the column identified by "B" (not to be confused with the rating factor B).

The coefficients entered the equation in a stepwise form in order to give an idea of their relative importance in explaining the variation in LMN. Thus the rating factor B is to be regarded as the most important in the above sense.

FIG (3.40)

REGRESSION OF LMN (GIVEN POSITIVE CLAIMS) ON

THE LEVELS OF THE RATING FACTORS

* M U L T I P L E   R E G R E S S I O N * * * * * * * * * * * * * * VARIABLE LIST 1
REGRESSION LIST 1

DEPENDENT VARIABLE.. LMN

VARIABLE(S) ENTERED ON STEP NUMBER 26.. Z5

MULTIPLE R	0.72586
R SQUARE	0.52687
ADJUSTED R SQUARE	0.52324
STANDARD ERROR	0.52529

ANALYSIS OF VARIANCE	DF	SUM OF SQUARES	MEAN SQUARE	F
REGRESSION	26.	1041.01996	40.03923	145.10882
RESIDUAL	3388.	934.83569	0.27593	

------------------ VARIABLES IN THE EQUATION ------------------

VARIABLE	B	BETA	STD ERROR B	F
B1	1.4077912	0.62954	0.03237	1891.433
B2	1.0445749	0.46624	0.03238	1040.492
Z1	0.1547398	0.07565	0.04738	10.665
S5	0.5778750	0.27614	0.03026	364.664
B3	0.8175078	0.36007	0.03272	624.167
M6	0.2161416	0.10169	0.04946	19.097
S4	0.4590998	0.22999	0.02915	248.104
B4	0.6919884	0.30060	0.03303	438.912
Z4	-0.4489538	-0.22709	0.04710	90.864
M9	0.3184129	0.16996	0.04825	43.550
B5	0.5607976	0.25031	0.03235	300.496
B6	0.4442241	0.21228	0.03074	208.854
M5	0.8153509	0.36188	0.04999	266.037
M8	0.8718164	0.32050	0.05383	262.328
Z6	-0.3220317	-0.14992	0.04792	45.166
S3	0.2337129	0.12702	0.02747	72.398
M2	0.7451220	0.28943	0.05236	202.502
M7	0.7068271	0.27823	0.05211	183.987
M1	0.5758219	0.25362	0.04986	133.390
J2	-0.1369381	-0.12070	0.02743	46.431
M3	0.2542808	0.07470	0.06354	16.017
Z3	-0.1849916	-0.09151	0.04728	15.308
S2	0.1008019	0.05632	0.02693	14.010
J3	-0.0930884	-0.05919	0.02801	11.046
Z2	-0.0599648	-0.02933	0.04738	1.602
Z5	-0.0454448	-0.01952	0.04887	0.865
(CONSTANT)	0.7720739			

------------- VARIABLES NOT IN THE EQUATION -------------

VARIABLE	BETA IN	PARTIAL	TOLERANCE	F

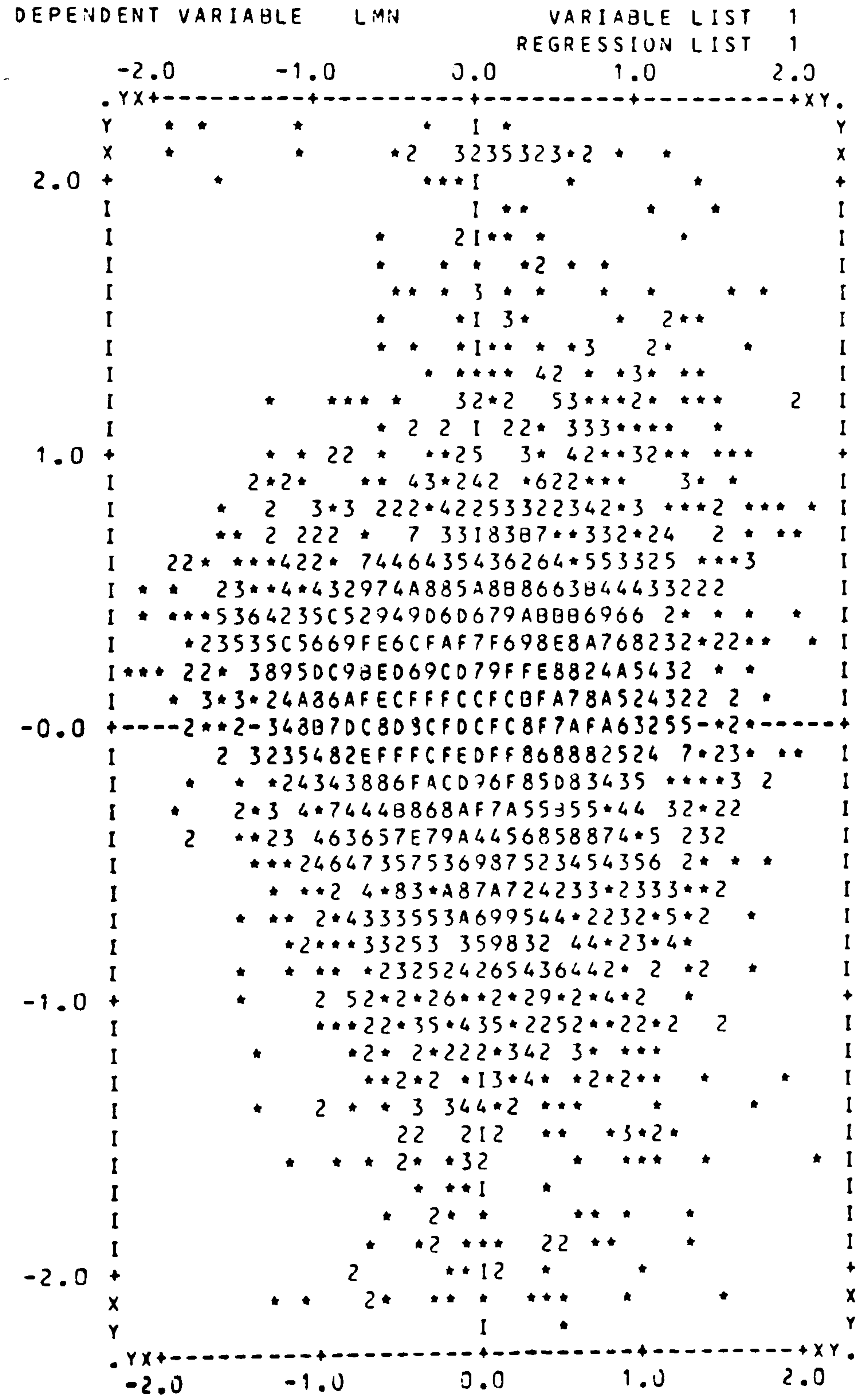
MAXIMUM STEP REACHED

STATISTICS WHICH CANNOT BE COMPUTED ARE PRINTED AS ALL NINES.

FIG (3.41)

REGRESSION OF LMN (GIVEN POSITIVE CLAIMS) ON

THE LEVELS OF THE RATING FACTORS (CONTINUED)



The last two dummy variables to enter the equation (Z2 and Z5) have standard errors in their coefficients too large if compared with the coefficients themselves. This means that the estimates of those coefficients are not reliable, and this fact must be taken into consideration when interpreting the regression equation.

In order to interpret the estimated coefficients, the model equation will be written as :

$$LMN = a + \sum_i k_i D_i \quad ,$$

where : "LMN" stands for the expected value of the naperian logarithm of the claim frequency,

"a" represents the constant term of the regression,

"k_i" is the coefficient of the dummy variable "D_i".

It is worth remembering that "a" is the value taken by "LMN" when all dummy variables are set to zero, therefore "a" is the expected value of "LMN" when all rating factors are set to their reference levels.

The above equation can be written as :

$$MN = \exp(a).\exp(\sum_i k_i D_i) \quad ,$$

which corresponds to an equation of a multiplicative model in which the claim frequency can be evaluated for any combination of levels of the rating factors. When a particular combination is chosen, their associated dummy variables are set to 1 and all others to zero, which means that the sum in the second exponent will have five non-zero terms at most. If there is one or more reference levels in the chosen combination, the number of non-zero terms in the sum reduces accordingly, as their effect has already been accounted for in the first exponent (constant term).

So, the equation of the model can be finally written as :

$$MN = \kappa . f_S . f_Z . f_B . f_M . f_J$$

where $\kappa = \exp(a) = \exp(0.772) = 2.16$

and the factors are obtained by exponentiating the corresponding coefficients in the regression equation. The results are displayed in the table below :

level	f_S	f_Z	f_B	f_M	f_J
1	1.00	1.17	4.09	1.78	(1.00)
2	1.11	0.94 *	2.84	2.11	(0.83)
3	1.26	0.83	2.26	1.29	0.91
4	1.58	0.64	2.00	(1.00)	
5	1.78	0.96 *	1.75	2.26	
6		0.72	1.56	1.24	
7		(1.00)	(1.00)	2.03	
8				(2.39)	
9				(1.37)	

Care must be taken in interpreting the above values as some of them are subject to unacceptable error (denoted by an asterix) and others (within brackets) correspond to levels with an abnormal frequency in the zero claims analysis performed in Section 3.7. As the distribution under consideration is a conditional one, that is, given that NCLA > 0, the values represented within brackets are certainly distorted, and therefore no attempt will be made to interpret them.

If one compares the remaining values in the Table above, some conclusions may be drawn regarding the way the rating factors influence the claim frequency.

The values of f_S increase with the levels of S; therefore it is confirmed that the claim frequency increases with the use of the vehicle.

In order to aid the interpretation of f_Z, their values (excluding the suspected ones) will be ranked along with the corresponding levels of Z :

f_Z	Z level	meaning
1.17	1	three biggest cities
0.83	3	small cities
0.72	6	rural areas in the north
0.64	4	rural areas in the south

There is some evidence that the claim frequency increases with populational density, judging by the results in the above Table.

Looking now at the values of f_B, it becomes clear that the claim frequency decreases as the bonus discount increases.

The values of f_M will be ranked in the same way as those of f_Z.

f_M	M level	relative engine size
2.26	5	2.23
2.11	2	2.40
2.03	7	2.17
1.78	1	2.20
1.29	3	2.43
1.24	6	1.60

Judging by the results above, it does not seem clear that the claim frequency increases with the power of the vehicle. Perhaps security aspects of the vehicles overcome the importance of the power in explaining the variation in the claim frequency.

Regarding the rating factor J, one cannot interpret the f_J values because just one of them can be regarded as reliable. However, the results derived in Sections 3.7 and 3.9 give some indication that the claim frequency increases with the age of the vehicle.

3.12 Conclusions

No strong evidence was found that the five rating factors have a significant influence on the average claim, that is, there was no indication that the variation in the average claim could be explained as due to the different levels of the rating factors, with the possible exceptions of the levels M = 9, Z = 7 and B = 7.

The same was not found to be true as regards the claim frequency, in which the influence of the rating factors was not only detected but also quantified with the aid of a multiplicative model. The measurement of the effects of the rating factors on the claim frequency was only possible to be made on a restricted set of the observations, namely those with at least one claim, otherwise one would be violating fundamental theoretical assumptions.

One possible and indeed reasonable interpretation of the above findings is that the occurrence of a claim in third party insurance does depend on the rating factors; however after such a claim has happened, its amount does not depend so much on the rating factors, but on random fluctuation or perhaps on other factors which were not considered in the analysis.

A lognormal distribution was found suitable to represent claim frequencies, given that NCLA > 0, but not for average claims based on a small number of claims. This means that a separation of severe bodily injury claims seems to be necessary when dealing with the distribution of a single claim in third party insurance.

CHAPTER FOUR

SETTLEMENT DELAYS IN GENERAL INSURANCE

4.1 Introduction

The settlement of claims in non-life insurance is subject to delays particularly when legal liabilities are involved. Other factors may influence these delays as, for example, administrative procedures within certain insurance companies.

For an insurance company, the ability to estimate its future liabilities is of great importance. Indeed, if the company manages to estimate its future payments with accuracy, this means that both its loss reserves and future premiums will be likely to be correct.

Most of the methods used by insurance companies to deal with settlement delays and ultimately with the estimation of outstanding claims amounts are based on the so-called run-off triangle, which will be discussed in the next section.

4.2 The run-off triangle

A run-off triangle is an array in which the payment history of several consecutive years may be summarized.

The figures in the cells of the run-off triangle may represent different quantities : claim numbers, total payments, average payments, etc. The figures may appear in cumulative or in non-cumulative form. Of course, one can be easily transformed into the other and vice versa.

An example of a run-off triangle is shown in Table (4.1), in which C_{ij} represents the cumulative claim amount paid by the end of year j in respect of claims incurred in year i. In other words, C_{ij} is the total amount paid in year of origin i and the following j years.

Such a triangle can be used for the calculation of provisions for outstanding claims by the so-called "chain ladder" method (Taylor, 1977). This method is based on the assumption that in the absence of exogeneous factors such as inflation, changing mix of risks, changing size of portfolio etc., the distribution of delays between the incident giving rise to a claim and the payments made in respect of that claim remain relatively stable over time (Benjamin, 1977).

It follows from the above assumption that the columns of the run-off triangle are proportional to one another, apart from random fluctuation.

The method consists of calculating the ratios (Taylor, 1977) :

$$\hat{M}_j = \left(\prod_{h=j}^{k-1} \hat{m}_h \right) \hat{M}_k \qquad (1)$$

where $\hat{M}_j$ is an estimate of $C_{i\infty}/C_{ij}$ and $\hat{m}_h$ is an estimate of $C_{i(h+1)}/C_{ih}$, which is calculated as :

TABLE (4.1)

THE RUN-OFF TRIANGLE

year of origin	development year							
	0	1	2	.	.	.	k-1	k
0	C_{00}	C_{01}	C_{02}	.	.	.	C_{0k-1}	C_{0k}
1	C_{10}	C_{11}	C_{12}	.	.	.	C_{1k-1}	
2	C_{20}	C_{21}	C_{22}	.	.	.		
.	.	.	.	.	.			
.	.	.	.	.				
.	.	.	.					
k-1	C_{k-10}	C_{k-11}						
k	C_{k0}							

$$\hat{m}_h = \frac{\sum_{i=0}^{k-j-1} C_{i(h+1)}}{\sum_{i=0}^{k-j-1} C_{ih}} \qquad (2)$$

In expression (1), $\hat{M}_k$ is obtained from an estimate of the outstanding liability as at the end of development year k (for year of origin 0).

The factors $\hat{M}_j$ evaluated in expression (1) can now be used to calculate outstanding claims provisions in respect to each of the years of origin. For year of origin i, the outstanding claims provision is :

$$C_{i(k-i)} (\hat{M}_{k-i} - 1) \qquad (3)$$

If the exogeneous factors referred to above are not negligible, the "chain ladder" method can give misleading results. In this case other methods must be used as, for example, the "separation method" (Benjamin, 1977), which is however based on a run-off triangle consisting of non-cumulative payments.

CHAPTER FIVE

SETTLEMENT DELAYS OF MOTOR INSURANCE CLAIMS

5.1 Introduction

This chapter aims at studying the speed of settlement of motor insurance claims. To this end, a set of data from a medium sized British insurance company was analysed with the aid of the statistical computer software SPSS.

The data consisted of individual payments made by the company referring to motor accidents which took place in the course of the years 1972 and 1973.

The reason for the choice of what may appear to be remote years of accident was the need to avoid the possibility of dealing with a large proportion of outstanding claims, so that the study could be carried out based as much as possible on real payments rather than on estimates.

The objective of the study was to establish the exact distributions of payments along the settlement years for each of the two years of accident so that an average pattern of settlement could be obtained.

5.2 Description of the data

The data was made available to this research in the form of computer records compiled by the company, each record containing the following information :

Notation (for computer use)		Description
1. YACC	-	the year of the accident which gave rise to the claim (either 72 or 73)
2. PAD	-	payment for accidental damage on the policyholder's own vehicle (units: £)
3. PTPBI	-	payment for third party bodily injury (units: £)
4. PTPPD	-	payment for third party property damage (units: £)
5. TSYR	-	duration of time elapsed from the notification of the claim to final settlement (units: years) If the claim is unsettled, this variable takes the arbitrarily chosen value of 24.
6. EUNP	-	estimated amount to be paid for unsettled claims (units: £) For settled claims, this variable takes the value zero.

The year taken as a reference for unsettled claims was 1980 when the computer file was made available by the company for this research.

The portfolio considered throughout the study was composed of private vehicles under a comprehensive cover policy.

5.3 Analysis of the pattern of settlement for YACC = 72

This section aims at evaluating the distribution of payments of each kind throughout the settlement years, for claims related to accidents which took place in the course of the year 1972.

For convenience the three kinds of payments will be denoted from now on as :

AD : accidental damage

TPBI : third party bodily injury

TPPD : third party property damage

One difficulty arises from the fact that if one claim gave rise to more than one kind of payment and if those payments were made at different times, the recorded time of settlement refers to the latest payment. Therefore, it is impossible to assign the correct individual time of settlement for composite payments.

The way around this problem was to define an auxiliary variable for each kind of payment so that it could be identified whether the corresponding payment was composite or not.

For AD payments, the auxiliary variable was defined as below :

Notation		Meaning
I1	=	1, if PAD $\neq$ 0 and PTPBI = 0 and PTPPD = 0
	=	2, if PAD $\neq$ 0 and PTPBI $\neq$ 0 and PTPPD = 0
	=	3, if PAD $\neq$ 0 and PTPBI = 0 and PTPPD $\neq$ 0
	=	4, if PAD $\neq$ 0 and PTPBI $\neq$ 0 and PTPPD $\neq$ 0

So, I1 = 1 means that only an AD payment was made,

I1 = 2 means a composite AD and TPBI payment,

I1 = 3 means a composite AD and TPPD payment,

I1 = 4 means a composite AD, TPBI and TPPD payment

A preliminary exploration of the data showed an insignificant small number of negative payments, the reason being either error in the process of codification of the data or possible recoveries made from other insurance companies. As it was impossible to distinguish the former from the latter, it was decided to treat all negative values as errors.

The distribution of AD payments throughout the years of settlement was produced by the procedure BREAKDOWN of SPSS. The restriction PAD > 0 was imposed to prevent negative AD payments from entering the analysis. The variable PAD was broken down by the auxiliary variable I1 and then by the variable TSYR and the results are shown in Fig. (5.1).

FIG (5.1)

DISTRIBUTION OF PAD ALONG THE YEARS OF

SETTLEMENT (YEAR OF ACCIDENT : 72)

- - - - - - - - - - - - - - - - - D E S C R I P T I O N O F S U B P O P U L A T I O N S - - - - - - - - - - - - - - - - -

CRITERION VARIABLE PAD
BROKEN DOWN BY I1
BY TSYR

VARIABLE	CODE	VALUE LABEL	SUM	MEAN	STD DEV	VARIANCE	N
FOR ENTIRE POPULATION			1167933.0000	98.4186	132.1996	17476.7263	(11867)
I1	1.	AD ONLY	751652.0000	79.0298	106.0111	11238.3500	(9511)
TSYR	0.		585927.0000	73.2592	95.3364	9089.0376	(7998)
TSYR	1.		140683.0000	110.0806	151.7464	23026.9653	(1278)
TSYR	2.		20929.0000	111.9198	121.8171	14839.4075	(187)
TSYR	3.		2852.0000	92.0000	79.5973	6335.7333	(31)
TSYR	4.		732.0000	61.0000	75.2547	5663.2727	(12)
TSYR	5.		132.0000	66.0000	76.3675	5832.0000	(2)
TSYR	6.		90.0000	45.0000	57.9828	3362.0000	(2)
TSYR	24.		307.0000	307.0000	0.	0.	(1)
I1	2.	AD + TPBI	108142.0000	280.8883	275.6735	75995.9016	(385)
TSYR	0.		27880.0000	224.8387	196.5707	38640.0226	(124)
TSYR	1.		40590.0000	281.8750	256.7249	65907.6906	(144)
TSYR	2.		16498.0000	299.9636	261.4284	68344.8135	(55)
TSYR	3.		15632.0000	390.8000	452.0539	204352.7282	(40)
TSYR	4.		3703.0000	284.8462	263.8579	69620.9744	(13)
TSYR	5.		193.0000	96.5000	99.7021	9940.5000	(2)
TSYR	6.		1925.0000	481.2500	596.4268	355724.9167	(4)
TSYR	7.		418.0000	418.0000	0.	0.	(1)
TSYR	24.		1303.0000	651.5000	129.4005	16744.5000	(2)
I1	3.	AD + TPPD	239406.0000	138.1454	139.3736	19424.9996	(1733)
TSYR	0.		162990.0000	129.4599	123.8656	15342.6921	(1259)
TSYR	1.		63392.0000	160.4861	169.8599	28852.3926	(395)
TSYR	2.		10556.0000	167.5556	187.1436	35022.7348	(63)
TSYR	3.		1070.0000	118.8889	126.0788	15895.8611	(9)
TSYR	4.		496.0000	165.3333	246.7739	60897.3333	(3)
TSYR	5.		569.0000	569.0000	0.	0.	(1)
TSYR	6.		178.0000	89.0000	108.8944	11858.0000	(2)
TSYR	24.		155.0000	155.0000	0.	0.	(1)
I1	4.	AD + TPBI + TPPD	68733.0000	288.7941	204.5955	41859.3034	(238)
TSYR	0.		18614.0000	235.6203	186.8768	34922.9309	(79)
TSYR	1.		30664.0000	286.5794	190.4799	36282.5856	(107)
TSYR	2.		7463.0000	355.3810	260.8167	68025.3476	(21)
TSYR	3.		5654.0000	314.1111	207.0189	42856.8105	(18)
TSYR	4.		3625.0000	604.1667	202.2458	40903.3667	(6)
TSYR	5.		2362.0000	393.6667	111.3224	12392.6667	(6)
TSYR	8.		351.0000	351.0000	0.	0.	(1)

TOTAL CASES = 11867

The column identified as SUM contains the sum of AD payments for each combination of values of the variables I1 and TSYR. It is worth remembering that TSYR = 0 corresponds to claims settled within one year from the notification to the insurer, TSYR = 1 corresponds to claims settled within the second year from notification and so on. TSYR = 24 corresponds to unsettled claims.

The results were rearranged as in Table (5.1) in which the sums of payments were rounded to £1000 units for the sake of clarity. It is worth noticing that the columns in the table corresponding to composite payments are AD payments which were made with other kinds of payments possibly at different times.

Similar procedures were adopted for TPBI and TPPD payments, that is, auxiliary variables were defined in order to identify composite payments and the procedure BREAKDOWN of SPSS was used to produce the distribution of these payments along the years of settlement. The results are shown in Fig. (5.2) for TPBI and in Fig. (5.3) for TPPD. In each case only positive payments entered the analysis in the same way as was done for AD payments.

The procedure CONDESCRIPTIVE of SPSS was used for the variable EUNP in order to produce the sum of estimated payments for outstanding claims and the result is shown on the bottom of Fig. (5.3). As expected, the sum of those payments is very small (£3250) and therefore will not be considered in the study.

The columns denoted by "N" in Figs. (5.1) to (5.3) represent the number of observations for each combination of the auxiliary variables and TSYR.

Table (5.1)

Settlement delays for AD payments (year of accident : 72)

time of settlement (years)	AD payments (1000£)			
	AD only	composite payments		
		with TPBI	with TPPD	with TPBI and TPPD
0	586	28	163	19
1	141	41	63	31
2	21	16	11	7
3	3	16	1	6
4	1	4	0	4
5		0	1	2
6		2		
7				
unsettled		1		

Total AD : £1168000

FIG (5.2)

DISTRIBUTION OF PTPBI ALONG THE YEARS OF SETTLEMENT (YEAR OF ACCIDENT : 72)

D E S C R I P T I O N   O F   S U B P O P U L A T I O N S

CRITERION VARIABLE PTPBI
BROKEN DOWN BY I2
BY TSYR

VARIABLE	CODE	VALUE LABEL	SUM	MEAN	STD DEV	VARIANCE	N
FOR ENTIRE POPULATION			452528.0000	539.3659	1742.0933	3034889.0628	(839)
I2	1.	TPBI ONLY	91330.0000	449.9015	1866.6527	3484392.2774	(203)
TSYR	0.		5140.0000	54.6809	120.9228	14622.3272	(94)
TSYR	1.		8562.0000	125.9118	287.3863	82590.8578	(68)
TSYR	2.		23754.0000	989.7500	1675.8926	2808615.8478	(24)
TSYR	3.		15990.0000	1453.6364	1908.2446	3641397.4545	(11)
TSYR	4.		32831.0000	6566.2000	9324.8342	86952533.7000	(5)
TSYR	5.		5053.0000	5053.0000	0.	0.	(1)
I2	2.	TPBI + AD	247805.0000	648.7042	1936.7697	3751077.0540	(382)
TSYR	0.		11163.0000	92.2562	288.7198	83359.1255	(121)
TSYR	1.		31239.0000	218.4545	363.7706	132329.0384	(143)
TSYR	2.		52305.0000	951.0000	1986.3997	3945783.8148	(55)
TSYR	3.		98987.0000	2414.3171	4130.0913	17057653.8220	(41)
TSYR	4.		38605.0000	2969.6154	3906.8344	15263355.2564	(13)
TSYR	5.		6309.0000	3154.5000	3570.1821	12746200.5000	(2)
TSYR	6.		7675.0000	2558.3333	2900.1952	8411132.3333	(3)
TSYR	7.		1501.0000	750.5000	1059.9531	1123500.5000	(2)
TSYR	24.		21.0000	10.5000	9.1924	84.5000	(2)
I2	3.	TPBI + TPPD	3869.0000	227.5882	378.8237	143507.3824	(17)
TSYR	0.		951.0000	105.6667	163.0905	26598.5000	(9)
TSYR	1.		1440.0000	240.0000	259.3260	67250.0000	(6)
TSYR	2.		1477.0000	1477.0000	0.	0.	(1)
TSYR	3.		1.0000	1.0000	0.	0.	(1)
I2	4.	TPBI + AD +TPPD	109524.0000	462.1266	1299.7317	1689302.3822	(237)
TSYR	0.		7484.0000	95.9487	168.8310	28503.9194	(78)
TSYR	1.		31122.0000	296.4000	886.4436	785782.2038	(105)
TSYR	2.		9023.0000	429.6667	477.3465	227859.6333	(21)
TSYR	3.		19725.0000	986.2500	1107.7199	1227043.4605	(20)
TSYR	4.		17469.0000	2911.5000	2866.7419	8218209.1000	(6)
TSYR	5.		24700.0000	4116.6667	4703.0869	22119026.6667	(6)
TSYR	8.		1.0000	1.0000	0.	0.	(1)

TOTAL CASES = 839

FIG (5.3)

DISTRIBUTION OF PTPPD ALONG THE YEARS OF SETTLEMENT (YEAR OF ACCIDENT : 72)

D E S C R I P T I O N   O F   S U B P O P U L A T I O N S

CRITERION VARIABLE PTPPD
BROKEN DOWN BY I3
BY TSYR

VARIABLE	CODE	VALUE LABEL	SUM	MEAN	STD DEV	VARIANCE	N
FOR ENTIRE POPULATION			154187.0000	63.0880	114.3385	13073.2935	(2444)
I3	1.	TPPD ONLY	19260.0000	41.5983	52.8311	2791.1283	(463)
TSYR	0.		12364.0000	36.4720	44.0606	1941.3387	(339)
TSYR	1.		4277.0000	44.0928	46.1375	2128.6684	(97)
TSYR	2.		1909.0000	86.7727	108.6869	11812.8506	(22)
TSYR	3.		710.0000	142.0000	137.9275	19024.0000	(5)
I3	2.	TPPD + AD	108474.0000	62.6293	93.4799	8738.4853	(1732)
TSYR	0.		72748.0000	57.5994	75.3463	5677.0723	(1263)
TSYR	1.		30974.0000	78.6142	134.3729	18056.0849	(394)
TSYR	2.		3361.0000	54.2097	68.1638	4646.2996	(62)
TSYR	3.		573.0000	71.6250	94.8411	8994.8393	(8)
TSYR	4.		145.0000	145.0000	0.	0.	(1)
TSYR	5.		1.0000	1.0000	0.	0.	(1)
TSYR	6.		2.0000	2.0000	0.	0.	(1)
TSYR	7.		4.0000	4.0000	0.	0.	(1)
TSYR	24.		666.0000	666.0000	0.	0.	(1)
I3	3.	TPPD + TPBI	603.0000	37.6875	37.1299	1378.6292	(16)
TSYR	0.		379.0000	47.3750	40.8479	1668.5536	(8)
TSYR	1.		214.0000	42.8000	33.7891	1141.7000	(5)
TSYR	2.		8.0000	4.0000	2.8284	8.0000	(2)
TSYR	3.		2.0000	2.0000	0.	0.	(1)
I3	4.	TPPD + AD + TPBI	25850.0000	110.9442	252.0898	63549.2770	(233)
TSYR	0.		6967.0000	90.4805	109.4315	11975.2529	(77)
TSYR	1.		13093.0000	124.6952	347.0165	120420.4832	(105)
TSYR	2.		2748.0000	137.4000	214.8888	46177.2000	(20)
TSYR	3.		1698.0000	89.3684	103.6126	10735.5789	(19)
TSYR	4.		795.0000	159.0000	97.2343	9454.5000	(5)
TSYR	5.		519.0000	86.5000	144.5901	20906.3000	(6)
TSYR	8.		30.0000	30.0000	0.	0.	(1)

TOTAL CASES = 2444

VARIABLE EUNP

MEAN	0.214	STD ERROR	0.171	STD DEV	21.156
VARIANCE	447.586	KURTOSIS	12913.512	SKEWNESS	111.334
RANGE	2500.000	MINIMUM	0.	MAXIMUM	2500.000
SUM	3250.000				

VALID OBSERVATIONS - 15220 MISSING OBSERVATIONS - 0

If one compares the number of observations for the same kind of composite payments, slight differences can be noticed. For example, in Fig. (5.1), the number of observations for composite AD payments with TPBI is 385 whereas that for composite TPBI payments with AD in Fig. (5.2) is 382. The reason for that lies in the way negative payments were dealt with in the analysis. If one particular record contained say a positive AD payment and a negative TPBI one, this record was included in the computation of AD payments but excluded in the TPBI ones, as only positive payments were allowed to enter the respective computations.

The results for TPBI and TPPD were rearranged in the same way as was done for AD payments and the results are shown in Tables (5.2) and (5.3).

Some conclusions can be drawn if one compares the values in Tables (5.1) to (5.3). The second column in each table contains sums of single payments only; therefore, their associated settlement delays can be regarded as correct. A comparison of those values shows that TPBI claims were those which took longer to settle as expected. Surprisingly though, TPPD claims were settled more rapidly than AD ones, although the payments involved in the latter type were much higher. It can also be noticed that longer settlement delays for composite payments occur when a TPBI payment is included.

Based on the above facts, it seems reasonable to assume that in a composite TPBI payment, the latest payment is likely to have been the TPBI one itself. With this assumption in mind, the values in each row of Table (5.2) can be added giving as a result

Table (5.2)

Settlement delays for TPBI payments (year of accident : 72)

time of settlement (years)	TPBI payments (1000£)				
	TPBI only	composite payments			
		with AD	with TPPD	with AD and TPPD
0	5	11	1	7
1	9	31	1	31
2	24	52	1	9
3	16	99		20
4	33	39		17
5	5	6		25
6		8		
7		2		
unsettled				

Total TPBI : £452000

Table (5.3)

Settlement delays for TPPD payments (year of accident : 72)

time of settlement (years)	TPPD payments (1000£)			
	TPPD only	composite payments		
		with AD	with TPBI	with AD and TPBI
0	12	73	0	7
1	4	31		13
2	2	3		3
3	1	1		2
4				1
5				1
6				
7				
unsettled		1		

Total TPPD : £155000

the total distribution of TPBI payments along the years of settlement. This was done and the result is displayed in the third column of Table (5.4), which shows the settlement delays for the three types of payments.

The total AD payments are shown in the second column of Table (5.4) and the criterion which was adopted to deal with composite AD payments was to assume that they were distributed along the years of settlement in the same proportion as the "AD only" payments. The same assumption was made for total TPPD payments and the results are shown in the fourth column of Table (5.4).

A comparative analysis of the results in Table (5.4) shows that AD payments corresponded to 65.8% of the total claims expenditure of the company whereas the percentages for TPBI and TPPD claims were 25.5% and 8.7% respectively.

AD claims are generally regarded as taking very little time on average to settle. Table (5.4) shows that a non-negligible amount of AD payments (£219000), corresponding to 12.3% of the total expenditure was settled within the second year from notification to the company, which is by no means a short delay as far as damage to the owner's vehicle is concerned.

5.4 Analysis of the pattern of settlement for YACC = 73

The same procedures of Section (5.3) were used to analyse the claims records corresponding to accidents which took place in the course of the year 1973.

Table (5.4)

Settlement delays for the three types of payments (year of accident : 72)

time of settlement (years)	type of payment (1000£)			Total
	AD	TPBI	TPPD	
0	910	24	98	1032
1	219	72	33	324
2	33	86	16	135
3	5	135	8	148
4	1	89		90
5		36		36
6		8		8
7		2		2
Total	1168	452	155	1775

The outputs from the procedure BREAKDOWN of SPSS are shown in Figs. (5.4), (5.5) and (5.6), respectively for AD, TPBI and TPPD payments. On the bottom of Fig. (5.6) the sum of estimated payments for outstanding claims is given and its value (£17584), although much higher than that of 1972, still can be considered small if compared with the real payments.

The results for the three types of payments were rearranged in the same way as in the previous section and are given in Tables (5.5) to (5.7). The same trends observed in the settlement delays for 1972 can be noticed for 1973 - namely, that TPBI claims take longer to settle than the other types and that TPPD claims settle more rapidly than AD claims.

The same criteria for dealing with composite payments in the previous section were adopted for 1973 claims, resulting in the distribution of the three types of payments along the settlement years, which is shown in Table (5.8).

It can be noticed that AD payments corresponded to 70.0% of the total claims expenditure contrasting with much smaller percentages for TPBI and TPPD which were 20.0% and 10.0% respectively.

Once more a substantial amount of AD payments (£247000) corresponding to 11.4% of the total expenditure was settled within the second year from notification to the company which confirms that AD claims are not settled as immediately as has commonly been thought by those working in the industry. Impressions can obviously be misleading.

FIG (5.4)

DISTRIBUTION OF PAD ALONG THE YEARS OF SETTLEMENT (YEAR OF ACCIDENT : 73)

- - - - - - - - - - DESCRIPTION OF SUBPOPULATIONS - - - - - - - - - -

CRITERION VARIABLE PAD
BROKEN DOWN BY I1
BY TSYR

VARIABLE	CODE	VALUE LABEL	SUM	MEAN	STD DEV	VARIANCE	N
FOR ENTIRE POPULATION			1527521.0000	114.7132	151.9866	23099.9328	(13316)
I1	1.	AD ONLY	1013859.0000	93.9454	125.6997	15800.4108	(10792)
TSYR	0.		814836.0000	87.3163	114.5811	13128.8220	(9332)
TSYR	1.		164123.0000	133.6507	171.0043	29242.4687	(1228)
TSYR	2.		26760.0000	143.1016	189.1376	35773.0380	(187)
TSYR	3.		7021.0000	189.7568	245.4090	60225.5781	(37)
TSYR	4.		1018.0000	169.6667	152.3741	23217.8667	(6)
TSYR	6.		89.0000	89.0000	0.	0.	(1)
TSYR	7.		12.0000	12.0000	0.	0.	(1)
I1	2.	AD + TPBI	108377.0000	301.0472	259.9324	67564.8641	(360)
TSYR	0.		25331.0000	238.9717	186.6751	34847.6087	(106)
TSYR	1.		36481.0000	287.2520	248.0783	61542.8249	(127)
TSYR	2.		25810.0000	339.6053	261.7438	68509.8154	(76)
TSYR	3.		12295.0000	423.9655	406.5027	165244.4631	(29)
TSYR	4.		5500.0000	423.0769	313.2075	98098.9103	(13)
TSYR	5.		1062.0000	265.5000	246.4366	60731.0000	(4)
TSYR	6.		481.0000	160.3333	151.5795	22976.3333	(3)
TSYR	7.		487.0000	487.0000	0.	0.	(1)
TSYR	24.		930.0000	930.0000	0.	0.	(1)
I1	3.	AD + TPPD	318883.0000	166.9545	174.8740	30580.8994	(1910)
TSYR	0.		215836.0000	151.8902	151.8943	23071.8710	(1421)
TSYR	1.		85768.0000	211.2512	225.1346	50685.5960	(406)
TSYR	2.		13340.0000	208.4375	225.8188	50994.1230	(64)
TSYR	3.		3227.0000	230.5000	210.2873	44220.7308	(14)
TSYR	4.		540.0000	135.0000	62.7535	3938.0000	(4)
TSYR	24.		172.0000	172.0000	0.	0.	(1)
I1	4.	AD + TPBI + TPPD	86402.0000	340.1654	275.2043	75737.4113	(254)
TSYR	0.		27706.0000	304.4615	235.5923	55503.7179	(91)
TSYR	1.		26934.0000	316.8706	242.9042	59002.4473	(85)
TSYR	2.		19234.0000	418.1304	354.6415	125770.6048	(46)
TSYR	3.		6439.0000	402.4375	248.4380	61721.4625	(16)
TSYR	4.		2012.0000	335.3333	380.9276	145105.8667	(6)
TSYR	5.		3007.0000	601.4000	423.7503	179564.3000	(5)
TSYR	6.		323.0000	107.6667	89.6679	8040.3333	(3)
TSYR	7.		652.0000	652.0000	0.	0.	(1)
TSYR	24.		95.0000	95.0000	0.	0.	(1)

TOTAL CASES = 13316

FIG (5.5)

DISTRIBUTION OF PTPBI ALONG THE YEARS OF SETTLEMENT (YEAR OF ACCIDENT : 73)

- - - - - - - - - - D E S C R I P T I O N   O F   S U B P O P U L A T I O N S - - - - - - - - - -

CRITERION VARIABLE PTPBI
BROKEN DOWN BY I2
BY TSYR

VARIABLE	CODE	VALUE LABEL	SUM	MEAN	STD DEV	VARIANCE	N
FOR ENTIRE POPULATION			433279.0000	511.5455	1780.0013	3168404.5390	(847)
I2	1.	TPBI ONLY	63208.0000	293.9907	1338.9068	1792671.4672	(215)
TSYR	0.		9313.0000	93.1300	204.9019	41984.7809	(100)
TSYR	1.		12712.0000	169.4933	402.1090	161691.6858	(75)
TSYR	2.		14921.0000	573.8846	1936.9342	3751713.9462	(26)
TSYR	3.		802.0000	80.2000	243.4506	59268.1778	(10)
TSYR	4.		3344.0000	1672.0000	1329.3607	1767200.0000	(2)
TSYR	5.		7303.0000	7303.0000	0.	0.	(1)
TSYR	6.		14813.0000	14813.0000	0.	0.	(1)
I2	2.	TPBI + AD	164797.0000	459.0446	1121.0380	1256726.2718	(359)
TSYR	0.		9169.0000	87.3238	192.1836	36934.5288	(105)
TSYR	1.		24528.0000	196.2240	314.4462	98876.4333	(125)
TSYR	2.		42729.0000	562.2237	893.3931	798151.2960	(76)
TSYR	3.		39712.0000	1323.7333	2267.3357	5140811.3747	(30)
TSYR	4.		32368.0000	2489.8462	2710.0777	7344521.3077	(13)
TSYR	5.		5773.0000	1443.2500	1205.8420	1454054.9167	(4)
TSYR	6.		8123.0000	2030.7500	2406.2544	5790060.2500	(4)
TSYR	7.		1473.0000	1473.0000	0.	0.	(1)
TSYR	24.		922.0000	922.0000	0.	0.	(1)
I2	3.	TPBI + TPPD	8522.0000	355.0833	527.1752	277913.6449	(24)
TSYR	0.		700.0000	77.7778	94.9918	9023.4444	(9)
TSYR	1.		3746.0000	416.2222	404.1373	163326.9444	(9)
TSYR	2.		4076.0000	679.3333	855.1790	731331.0667	(6)
I2	4.	TPBI + AD +TPPD	196752.0000	790.1687	2701.1463	7296191.1730	(249)
TSYR	0.		9045.0000	100.5000	174.9469	30606.4326	(90)
TSYR	1.		25713.0000	295.5517	442.7633	196039.3665	(87)
TSYR	2.		43780.0000	972.8889	1921.1944	3690988.0556	(45)
TSYR	3.		24584.0000	1891.0769	3644.3948	13281613.7436	(13)
TSYR	4.		12460.0000	2492.0000	2525.7346	6379335.5000	(5)
TSYR	5.		27886.0000	6971.5000	5393.3194	29087893.6667	(4)
TSYR	6.		13800.0000	4600.0000	7931.0834	62902084.0000	(3)
TSYR	7.		9279.0000	9279.0000	0.	0.	(1)
TSYR	24.		30205.0000	30205.0000	0.	0.	(1)

TOTAL CASES = 847

FIG (5.6)

DISTRIBUTION OF PTPPD ALONG THE YEARS OF

SETTLEMENT (YEAR OF ACCIDENT : 73)

D E S C R I P T I O N   O F   S U B P O P U L A T I O N S

CRITERION VARIABLE PTPPD
BROKEN DOWN BY I3
BY TSYR

VARIABLE	CODE	VALUE LABEL	SUM	MEAN	STD DEV	VARIANCE	N
FOR ENTIRE POPULATION			214695.0000	78.8450	184.2421	33945.1354	(2723)
I3	1.	TPPD ONLY	25873.0000	47.2135	55.7179	3104.4827	(548)
TSYR	0.		18598.0000	43.2512	48.1762	2320.9438	(430)
TSYR	1.		5717.0000	58.3367	70.5333	4974.9473	(98)
TSYR	2.		1332.0000	88.8000	109.4037	11969.1714	(15)
TSYR	3.		79.0000	26.3333	18.4481	340.3333	(3)
TSYR	4.		147.0000	73.5000	64.3467	4140.5000	(2)
I3	2.	TPPD + AD	148717.0000	77.9848	181.2386	32847.4158	(1907)
TSYR	0.		96631.0000	68.0500	177.9710	31673.6783	(1420)
TSYR	1.		39413.0000	97.3160	170.8204	29179.6127	(405)
TSYR	2.		7401.0000	117.4762	180.6073	32618.9954	(63)
TSYR	3.		5028.0000	359.1429	424.1057	179865.6703	(14)
TSYR	4.		242.0000	60.5000	56.5125	3193.6667	(4)
TSYR	24.		2.0000	2.0000	0.	0.	(1)
I3	3.	TPPD + TPBI	1791.0000	77.8696	171.5767	29438.5731	(23)
TSYR	0.		216.0000	24.0000	32.6343	1065.0000	(9)
TSYR	1.		1243.0000	155.3750	269.9344	72864.5536	(8)
TSYR	2.		332.0000	55.3333	101.8384	10371.0667	(6)
I3	4.	TPPD + AD + TPBI	38314.0000	156.3837	322.7885	104192.4096	(245)
TSYR	0.		10232.0000	113.6889	166.3336	27666.8684	(90)
TSYR	1.		9487.0000	118.5875	203.4019	41372.3467	(80)
TSYR	2.		9033.0000	205.2955	357.4340	127759.0502	(44)
TSYR	3.		4073.0000	254.5625	497.9731	247977.1958	(16)
TSYR	4.		3828.0000	765.6000	1373.7681	1887238.8000	(5)
TSYR	5.		1543.0000	308.6000	286.8925	82307.3000	(5)
TSYR	6.		112.0000	37.3333	29.3655	862.3333	(3)
TSYR	7.		4.0000	4.0000	0.	0.	(1)
TSYR	24.		2.0000	2.0000	0.	0.	(1)

TOTAL CASES = 2723

VARIABLE EUNP

MEAN	1.051	STD ERROR	0.730	STD DEV	94.466
VARIANCE	8923.838	KURTOSIS	9302.540	SKEWNESS	95.200
RANGE	10000.000	MINIMUM	0.	MAXIMUM	10000.000
SUM	17584.000				

VALID OBSERVATIONS - 16734 MISSING OBSERVATIONS - 0

Table (5.5)

Settlement delays for AD payments (year of accident : 73)

time of settlement (years)	AD payments (1000£)			
	AD only	composite payments		
		with TPBI	with TPPD	with TPBI and TPPD
0	815	25	216	28
1	164	36	86	27
2	27	26	13	19
3	7	12	3	6
4	1	6	1	2
5		1		3
6				0
7				1
unsettled		1		

Total AD : £1526000

Table (5.6)

Settlement delays for TPBI payments (year of accident : 73)

time of settlement (years)	TPBI payments (1000£)			
	TPBI only	composite payments		
		with AD	with TPPD	with AD and TPPD
0	9	9	1	9
1	13	25	4	26
2	15	43	4	44
3	1	40		25
4	3	32		12
5	7	6		28
6	15	8		14
7		1		9
unsettled		1		30

Total TPBI : £434000

Table (5.7)

Settlement delays for TPPD payments (year of accident : 73)

time of settlement (years)	TPPD payments (1000£)			
	TPPD only	composite payments		
		with AD	with TPBI	with AD and TPBI
0	19	97	0	10
1	6	39	1	9
2	1	7		9
3		5		4
4				4
5				2
6				
7				
unsettled				

Total TPPD : £213000

Table (5.8)

Settlement delays for the three types of payments

(year of accident : 73)

time of settlement (years)	type of payment (1000£)			Total
	AD	TPBI	TPPD	
0	1226	28	156	1410
1	247	68	49	364
2	41	106	8	155
3	11	66		77
4	1	47		48
5		41		41
6		37		37
7		10		10
unsettled		31		31
Total	1526	434	213	2173

5.5 Average pattern of settlement

The total payments in the fifth column of Tables (5.4) and (5.8) were expressed as percentages of the respective column totals so that an average pattern of settlement could be obtained.

Table (5.9) shows the average distribution of payments (in percentages) along the settlement years, based on the patterns of settlement for the two years of accident 1972 and 1973 evaluated as above. It is worth noticing that on average 61.5% of the claims expenditure is paid within the first year from notification to the company, whereas the remaining 38.5% is spread over seven years from notification according to the percentages in Table (5.9). This means that a marginal profit can be made by the company under favourable investment conditions.

To illustrate the above point numerically, the results for the year of accident 1972 will be re-examined. Assuming a constant rate of inflation of say 5%, the true risk premium in monetary units of 1972 can be evaluated from Table (5.4) as :

$$\underline{P} = 1032 + 324 \times 1.05^{-1} + 135 \times 1.05^{-2} + 148 \times 1.05^{-3} + 90 \times 1.05^{-4} + $$
$$+ 36 \times 1.05^{-5} + 8 \times 1.05^{-6} + 2 \times 1.05^{-7} = 1700.51$$

After deducting the claims expenses within the first year from notification, the available capital for investment will be :

$$C = 1700.51 - 1032.00 = 668.51$$

Assuming that this capital is invested at a rate of interest

Table (5.9)

Average settlement delays

time of settlement (years)	percentage of total payments		
	1972	1973	average
0	58.1	64.9	61.50
1	18.3	16.8	17.55
2	7.6	7.1	7.35
3	8.3	3.5	5.90
4	5.1	2.2	3.65
5	2.0	1.9	1.95
6	0.5	1.7	1.10
7	0.1	0.5	0.30
unsettled	0.0	1.4	0.70
Total	100.0	100.0	100.00

say 1% above inflation, that is 6% per year, the marginal profit after all claims have been settled can be evaluated as follows :

year of settlement	capital	interest	claims expenses	balance
1	668.51	40.11	324.00	384.62
2	384.62	23.08	135.00	272.70
3	272.70	16.36	148.00	141.06
4	141.06	8.46	90.00	59.52
5	59.52	3.57	36.00	27.09
6	27.09	1.63	8.00	20.72
7	20.72	1.24	2.00	19.96

So the marginal profit is obtained by expressing the final balance of £19960 in monetary units of 1972, that is :

$$\text{Marginal Profit} = 19960 \times 1.05^{-7} = £14185.20.$$

5.6 Conclusions

It was shown in this Chapter that long settlement delays in motor insurance claims are generally associated with third party bodily injury claims. Although AD and TPPD claims are settled more rapidly on average than TPBI claims, their settlement delays cannot be considered as negligible. Indeed, it was shown that some of these claims may take as long as four years to settle.

An average pattern of settlement was obtained by expressing the payments for the two years of accident as percentages of the respective total claims expenditure. It was seen that settlement delays can constitute a source of marginal profits for an insurance company provided that its average rate of return in investment exceeds the inflation rate.

CHAPTER SIX

SUMMARY OF CONCLUSIONS

6.1 Conclusions

The first investigation in this research aimed at studying the influence of a given set of rating factors on the average claim and claim frequency in third party motor insurance.

Five rating factors were considered and no strong evidence was found that they had a significant influence on the average claim, with the possible exception of three particular levels M = 9, Z = 7 and B = 7, for which an abnormal variation of the average claim was detected.

As regards the claim frequency, such an influence was not only detected, but also quantified for a limited set of the observations.

It was also found that a lognormal distribution was suitable to represent both claim frequencies given a positive number of claims and average claims based on a large number of claims. For a small number of claims, it was found that a separation of severe bodily injury claims was needed so that the average claim could be represented by a lognormal distribution.

The aim of the second investigation was to study the speed of settlement of motor insurance claims. An average pattern of settlement was obtained based on data from a medium sized British insurance company. It was shown that delays in settlement could give rise to marginal profits provided that the company's average rate of return in investment exceeded the inflation rate.

It was also confirmed that long settlement delays in motor insurance are generally due to third party bodily injuries, although the settlement delays for accidental damage and third party property damage claims were not found to be negligible.

6.2 Suggestions for further research

Due to the form in which the data was made available to this research, no attempt was made in the first investigation to study the effect of interactions among the factors. Indeed, the inclusion of interactions in linear models based on large-scale survey-type data when several factors are under consideration and just one observation per cell is available, is not recommended (Searle, 1971, Chap. 8).

One way in which further research could be carried out in this subject would be to examine the effect of interactions among the rating factors in explaining the variation of the average claim and claim frequency. To this end, data should be collected under a careful experimental design with replications in each cell in order to provide an adequate number of degrees of freedom to test the significance of such interactions.

BIBLIOGRAPHY

ABBOTT, W. M., CLARKE, T. G., HEY, G. B., REYNOLDS, D. I. W. and TREEN, W. R. (1974).
Some thoughts on technical reserves and statutory returns in general insurance.
The Journal of the Institute of Actuaries 101 : 217-265.

AITCHISON, J. and BROWN, J. A. (1957).
The Lognormal Distribution.
(Cambridge : Cambridge University Press).

BALZER, L. A. and BENJAMIN, S. (1981).
Dynamic response of insurance systems with delayed profit/loss-sharing feedback to isolated unpredicted claims.
The Journal of the Institute of Actuaries 108 : 513-528.

BEARD, R. E., PENTIKÄINEN, T. and PESONEN, E. (1977).
Risk Theory - 2nd Edition.
(London : Chapman and Hall).

BENCKERT, L. (1962).
The lognormal model for the distribution of one claim.
Astin Bulletin 2 : 9-23.

BENJAMIN, B. (1977).
General Insurance
(London : Heinemann).

BENJAMIN, S. (1976)
Profit and other financial concepts in insurance.
The Journal of the Institute of Actuaries 103 : 233-281.

BERLINER, B. (1980).
Deriving the Pareto and Exponential distribution function from classified observation data : a new approach.
Transactions of the 21st International Congress of Actuaries, Vol. 2.

BORCH, K. (1974).
Mathematical models in insurance.
Astin Bulletin 7 : 192-202.

BOX, G. E. P., HUNTER, W. G. and HUNTER, J. S. (1978).
Statistics for Experimenters.
(New York : Wiley).

BÜHLMANN, H. (1970).
Mathematical methods in Risk Theory.
(New York : Springer-Verlag).

CANNAR, K. (1979).
Motor Insurance - Theory and Practice.
(London : Witherby).

DELAPORTE, P. (1962).
Sur l'efficacité des critéres de tarification de l'assurance contre les accidents d'automobiles.
Astin Bulletin 2 : 84-95.

DIXON, W. J. and BROWN, M. B. (1979).
Biomedical Computer Programs - BMDP.
(Berkeley : University of California Press).

DRAPER, N. and SMITH, H. (1981).
Applied regression analysis - 2nd edition.
(New York : Wiley).

DUBOURDIEU, J. (1952).
Théorie mathématique du risque dans les assurances de répartition.
(Paris : Gauthier-Villars).

GERBER, H. U. and JONES, D. A. (1976).
Some practical considerations in connection with the calculation of stop-loss premiums.
Transactions of the Society of Actuaries 28 : 215-235.

GERBER, H. U. (1979).
An introduction to mathematical risk theory.
(Philadelphia : Huebner Foundation of Insurance Education).

GILLESPIE, R. G. (1973).
Risk Theory.
General Insurance studies Sub-Committee, Subject no. 11.
The Institute of Actuaries, London.

HALLIN, M. and INGENBLEEK, J. F. (1981).
Etude statistique des facteurs influençant le risque automobile :
Le montant cumulé des sinistres dans le portefeuille suédois en 1979.
Université Libre de Bruxelles, Institute de Statistique - Série actuarielle no. 10.

JEWELL, W. S. (1980).
Models in insurance : paradigms, puzzles, communications and revolutions.
Transactions of the 21st International Congress of Actuaries, Vol. 1.

JOHANSEN, P. (1978).
Mathematical models regarding fire and consequential loss.
Scandinavian Actuarial Journal : 229-234.

JOHNSON, P. D. and HEY, G. B. (1971).
Statistical studies in motor insurance.
The Journal of the Institute of Actuaries 97 : 199-232.

KAHANE, Y. and LEVY, H. (1975).
Regulation in the insurance industry : determination of premiums in automobile insurance.
Journal of Risk and Insurance 62 (1) : 117-132.

KAHANE, Y. (1979).
The theory of insurance risk premiums - a re-examination in the light of recent developments in capital marketing theory.
Astin Bulletin 10 : 223-239.

KARLSSON, J. E. (1976).
The expected value of IBNR claims.
Scandinavian Actuarial Journal : 108-110.

KENDALL, M. and STUART, A. (1976).
The Advanced Theory of Statistics. Vol. 3, 3rd Edition.
(London : Charles Griffin).

L'ASSOCIATION GÉNÉRALE DES SOCIÉTÉS D'ASSURANCES CONTRE LES ACCIDENTS (1977).
Exploitation du sondage automobile 1971 en France par une méthode d'analyse multidimensionnelle.
Astin Bulletin 9 : 10-25.

LITTLE, R. J. A. (1978).
Generalized linear models for cross-classified data from the WFS.
World Fertility Survey, Technical Bulletin no. 5.

LOIMARANTA, K. (1971).
Some asymptotic properties of bonus systems.
Astin Bulletin 6 : 233-245.

MORRISON, D. F. (1976).
Mutivariate Statistical Methods. 2nd Edition.
(New York : McGraw-Hill).

MUFF, M. (1971).
The influence of the franchise on the number of claims in motor insurance.
Astin Bulletin 6 : 191-194.

NIE, N. H., HULL, C. H.,JENKINS, J. G., STEINBRENNER, K. and BENT, D. H. (1975).
Statistical Package for the Social Sciences - SPSS. 2nd Edition
(New York : McGraw-Hill).

NORBERG, R. (1979).
The credibility approach to experience rating.
Scandinavian Actuarial JOurnal : 181-221.

PECHLIVANIDES, P. M. (1978).
Optimal reinsurance and dividend payment strategies.
Astin Bulletin 10 : 34-46.

PENTIKÄINEN, T. (1975).
A model of stochastic-dynamic prognosis.
Scandinavian Actuarial Journal : 29-53.

PITKÄNEN, P. (1974).
Tariff theory.
Astin Bulletin 8 : 204-228.

PUZEY, A. S. (1973).
Distributions fitted to non-life claim events and amounts.
General Insurance Studies Sub-Committee, Subject no. 5.
The Institute of Actuaries, London.

RAO, C. R. (1973).
Linear Statistical Inference and Its Applications. 2nd Edition.
(New York : Wiley).

REID, D. H. (1978)
Claim Reserves in General Insurance.
The Journal of the Institute of Actuaries 105 : 211-296.

ROSS, S. M. (1972).
Introduction to Probability Models.
(New York : Academic Press).

ROSSMAN, G. (1938).
Ajustement des écarts en assurance-grêle.
Bulletin Trimestral de le Institute Actuariel Français 44 : 75-78.

ROUSSAS, G. G. (1973).
A First Course in Mathematical Statistics.
(Reading : Addison-Wesley).

SAWKINS, R. W. (1973).
Solvency in non-life insurance.
Transactions of the Institute of Actuaries of Australia and New Zealand.

SCHEFFÉ, H. (1959).
The Analysis of Variance.
(New York : Wiley).

SEAL, H. L. (1969).
Stochastic theory of a risk business.
(New York : Wiley).

SEARLE, S. R. (1971).
Linear Models.
(New York : Wiley).

STRAUB, E. (1980).
What is an adequate dividing line between normal claims and large claims ?
Transactions of the 21st International Congress of Actuaries.

TAYLOR, G. C. (1975).
A Survey of Principal Results from the Theory of Risk.
Occasional Actuarial Research Discussion Papers - Paper no. 3.
The Institute of Actuaries, London.

TAYLOR, G. C. (1977).
Separation of inflation and other effects from the distribution of non-life insurance claim delays.
Astin Bulletin 9 : 219-230.

WELTEN, C. P. (1963).
Estimation of stop loss premium in fire insurance.
Astin Bulletin 2 : 356-361.

WOLD, H. O. A. (1937)
A technical study on reinsurance.
Transactions of the 11th International Congress of Actuaries, Vol. 1.

ZEHNWIRTH, B. (1979).
A hierarchical model for the estimation of claims rates in a motor car insurance portfolio.
Scandinavian Actuarial Journal : 75-82.