PREDICTING URBAN WATER DISTRIBUTION MAINTENANCE STRATEGIES:

A CASE STUDY OF NEW HAVEN, CONNECTICUT

by

Lisa A. Jeffrey

SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS OF THE
DEGREE OF
MASTER OF SCIENCE
IN CIVIL ENGINEERING - WATER RESOURCES

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

January 1985

Signature of Author
Department of Civil Engineering - Water Resources
January 14,1985

Certified by
David H. Marks
Thesis Supervisor

Accepted by
Francois M. M. Morel, Chairman
Departmental Graduate Committee

PREDICTING URBAN WATER DISTRIBUTION MAINTENANCE STRATEGIES:

A CASE STUDY OF NEW HAVEN, CONNECTICUT

by

Lisa A. Jeffrey

Submitted to the Department of Civil Engineering - Water
Resources and Environmental Engineering on January 15, 1985
in partial fulfillment of the requirements for the Degree of
Master of Science in Civil Engineering

ABSTRACT

A probabilistic failure prediction model was developed using data from the water distribution system of New Haven, Connecticut. The model was developed to provide water distribution management with easy access to information on the projected status of a distribution system at various points of time in the future, the costs and risks associated with alternative maintenance programs, as well as information on the failure probability of each pipe.

The model development combined in depth statistical analysis with the mathematical regression techniques developed by Cox in 1972. The probability of failure was calculated for each pipe in the system from the following break causing factors: pressure, land development, length, date of installation, number of previous breaks, age at the time of second break.

It was concluded that this model could be applied by water distribution managers as a useful aid in the maintenance management of water distribution systems. Specifically, this would be accomplished by combining the identification of high risk pipes in a system with financial analysis on alternative maintenance solutions. Through the use of this technique, management could minimize both the risks involved with pipe failures and the associated maintenance costs.

Thesis Supervisor: Dr. David H. Marks

Title: Professor of Civil Engineering

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE

WATER DISTRIBUTION SYSTEM MAINTENANCE

INTRODUCTION

Mature water distribution systems are currently facing problems commonly associated with deteriorating infrastructure. A majority of the systems in the eastern United States have reached a critical age. as reflected by the high breakage rates, loss of carrying capacity, and high unaccountable water losses. Although the majority of management authorities acknowlege the increased unreliability of distribution systems at present, routine maintenance has been negligible in the past. As a result, many of the mature distribution systems in the U.S. have a backlog of maintenance and rehabilitation problems to solve in order to continue to meet water quality and supply requirements. The problem has become one of trying to maintain these neglected and deteriorated systems within the financial budgets available.

Due to limited municipal budgets, it is necessary for management to distinguish and focus immediate attention on the most critically deteriorated pipes and areas in the distribution system. This report presents a model which would help to isolate these critical pipes and areas of the distribution system. The results of this model could be used by water distribution management in order to make the best use of the limited funds and resources for maintenance.

Management decisions as to the proper maintenance alternative for each specific pipe should be based on realistic and up to date information on the distribution system, the economics of the alternatives, and the importance of each pipe in the system. Currently, management officials do not have easy access to extensive information on the past performance of the pipe from which to make such decisions. As a result, such decisions are based on immediate failures and flow requirements in a given area. A variety of emperical rules are also used to guide the decision making process.

The premise of this report is that to make the best use of the limited funds and resources available for maintenance, management officials need to be supplied with better, currently updated information which reflects the relative degeneration of each pipe, it's importance within the distribution network, as well as the economics of the maintenance alternatives. From this sort of information, management could systematically determine the pipes and areas which have the highest priority for attention and are most economical over time.

The model described in this report will predict the probability of experiencing a break in each specific pipe from the historical data on the system. The model can easily be updated so that the probability of a pipe failure can be determined at any point in time. Such failure probability could be used to rank pipes for replacement priority. Also, when flow tests show that a given area needs increased carrying capacity and is being scheduled for rehabilitation, the failure probability could be used to isolate pipes which should be replaced rather than rehabilitated. In this way, management could avoid expending additional funds to rehabilitate a pipe that would require replacement within the next five to ten years. This ability to predict the break rate should prove invaluable to water distribution management. It is hoped that by having such information readily available, management can make better use of the limited funds and resources for maintenance by concentrating on the truely critical areas and pipes within the system. It is also hoped that in the process of the development of the model, specific parameters which increase the failure probability in the system could be isolated and thereby add insight into the mechanisms of failure in the system.

Several methods have been used to approach the problem of predicting break rates. The model described in this report is based on the use of historical records of the system to give such predictions. Another approach is the use of the present physical parameters of the system to predict the failure rate. This is the sort of method which has been developed by O'day et al.(1984), using the predicted extent of interior and exterior corrosion to predict future failure probabilities. This method may prove to be quite useful for cities which have not kept sufficient historical records to provide accurate predictions through the use of our model. Both techniques deserve attention as the information which they provide on the system can be quite valuable to water distribution systems. Perhaps the future models will incorporate both types of information.

The objective of this analysis is to determine a typical maintenance plan using information developed from this model for a mature distribution system using New Haven, Connecticut as case study. It is hoped that this maintenance decision making plan could be translated to other cities and other maintenance problems. It must be stressed that many of the steps in the analysis must be specific to the system so that the actual predictive formulas and equations can not be used directly for other distribution systems. However, the same set of steps described in this report can be applied to other systems to provide similar predictive capabilities.

In the following section, the different maintenance alternatives will be quickly reviewed. This discussion is intended to provide the background necessary for the development of the model. In section three, the data on the New Haven Distribution System which has served as the case study for development of the model will be presented. An in depth description of our model as well as a comparison to other failure models will be described in chapter two. In chapter three, the model will be applied to the New Haven data set and the results discussed. Chapter four provides a description of how the results of the model could be of use in decision making.

OVERVIEW OF MAINTENANCE ALTERNATIVES

Conditions in many mature water distribution systems have reached the point where urgent upgrading is required to bring these systems to within an acceptable level of reliability and maintainability. A variety of maintenance requirements need to be addressed for immediate action including the increasing numbers of leaks and breaks, the high percentage of unaccountable water loss, and the rising energy costs required for additional pumping due to the loss of carrying capacity in aging mains.

There are three primary maintenance alternatives which combat these problems. The first is directed at decreasing the high percentage of unaccounted-for water losses. This program includes repair and replacement of the deteriorated metering systems as well as leak detection and repair. Although this is an important program, it will not be covered in this analysis. Rather, the focus of this analysis will be the two other maintenance programs, which address rehabilitation and replacement alternatives.

The Replacement Alternative

The replacement alternative is usually considered as the optimal strategy for reducing the number of leaks and breaks in water mains. It is also the strategy used in areas of high growth where the pipes are of inadequate size for the expansion requirements of the system. Replacement is generally recommended for pipes which have experienced extensive breaks or leaks in the past and is not recommended for those that have remained structurally sound.

There are several categories of problems which are thought to be responsible for the occurance of such leaks and breaks. These factors will be briefly discussed to provide sufficient background for the description of the model. The following section presents only a brief description of the many potential causes for main leaks and breaks. A more in depth presentation of these causes may be found in papers by Morris (1966) and by O'Day (1982).

The principle causes for main failure can be broken down into three categories: environmental factors, system specific factors, and pipe specific factors. Environmental factors which may influence the break rate of a specific pipe are the corrosivity of the soil, land development(excessive loads), stray electrical currents, soil stability, and temperature as it influences the degree of penetration of frost. System specific factors include the pH and corrosivity of the water being distributed. The pipe specific factors include the age, type, size, depth of the pipe, nearby construction, and length of the pipe, as well as the quality of workmanship and care used in laying the pipe. These factors all lead to common problems which result in water main failure, internal and external corrosion and/or structural instability. However, not all of these factors have been documented in the data bases available describing water distribution systems. The availabilty of accurate data for a system will greatly impact the accuracy of the model's prediction of failure rates.

The Rehabilitation Alternative

The rehabilitation alternative has been considered as the optimal maintenance strategy for aging pipes which have remained structurally stable but are suffering from problems associated with the loss of carrying capacity. In the large mains, where repair costs are minor due to the low break frequency, management may consider the alternative of rehabilitation. The replacement alternative for such large mains is extremely expensive. Thus, for large mains, rehabilitation may frequently be the prefered alternative.

The loss of carrying capacity is primarily due to head losses caused by the deterioration of the interior wall of the pipe. Substantial head losses may lead to additional costs associated with increased energy required for pumping as well as costs for new and more powerful pumps, once the head loss requirement can no longer be met by the specified capacity of the pumping facility.

The analysis of the rehabilitation alternative is based on predictions of loss of carrying capacity and forecasts of future operational costs such as additional energy and pumping requirements. One limiting factor in the analyses of the rehabilitation alternative is the lack of accurate data on the reliability of a pipe following rehabilitation. As the rehabilitaion of pipes is a recently used technique, data is not yet available on the history of pipes following rehabilitation. O'day et al. (1984) have suggested that the cleaning and lining processes of rehabilitation may create conditions which increase the frequency of breaks for some pipes following rehabilitation. However, current managers of the Boston and New Haven distribution systems feel that these rehabilitated pipes may remain as reliable as they were before. Records on rehabilitated pipes which are now five to ten years of age show that the majority of such pipes have remained structurally stable to date. The savings in costs for rehabilitation as compared to replacement would make this alternative economical if such pipes survived at least twenty years from the rehabilitation date.

There are two primary problems which are avoided through rehabilitation: 1) the internal corrosion and tuberculation which create water quality problems; and 2) the loss of head as it creates increased energy costs for pumping purposes.

These two problems have both system specific and pipe specific factors which influence them. The primary system specific factor is the Hazen-William C coefficient which changes with the age of the pipe in different distribution systems. The Hazen-William formula, which is recognized as the fundamental equation describing flow in pipes, helps to demonstrate how the loss of carrying capacity and the additional operating costs increase with the extent of internal corrosion and therefore with the age of the pipe. This equation relates the flow(Q) in a pipe to the diameter(D), the head loss due to friction(ΔH), the length(L) and the Hazen-William coefficient(C_{HW}) as follows:

$$\dot{Q} = KC_{HW} D^{2.63} (\frac{\Delta H}{L})^{N}$$

where N is the exponent of the hydraulic slope, usually taken as 0.54, and K is a constant dependent on the choice of units. The head losses vary inversly with the Hazen-William coefficients. McBean et.al.(1983) analysed the variation in Hazen-William coefficients of seven water distribution systems (Figure 1). The trends shown in the figure show that a drop in the value of the Hazen-William coefficient would lead to an increase in the head loss over the given length for a given flow rate.

This drop of the Hazen-William coefficients with time is system specific but can be described as being composed of an environmental and a human factor(Karaa,1984). The environmental factor is related to the water type. The human factor is related to the treatment methods and to the operational controls of the water treatment plants. The chemical corrosivity of the water is another important system specific factor for pipe rehabilitation. Corrosive water is high in total dissolved solids, low in pH, high in hardness, dissoved oxygen and carbon dioxide. Hudson (1966) noted that certain water treatment methods designed to control water quality often increase the corrosivity and thereby increase the loss of carrying capacity in the system.

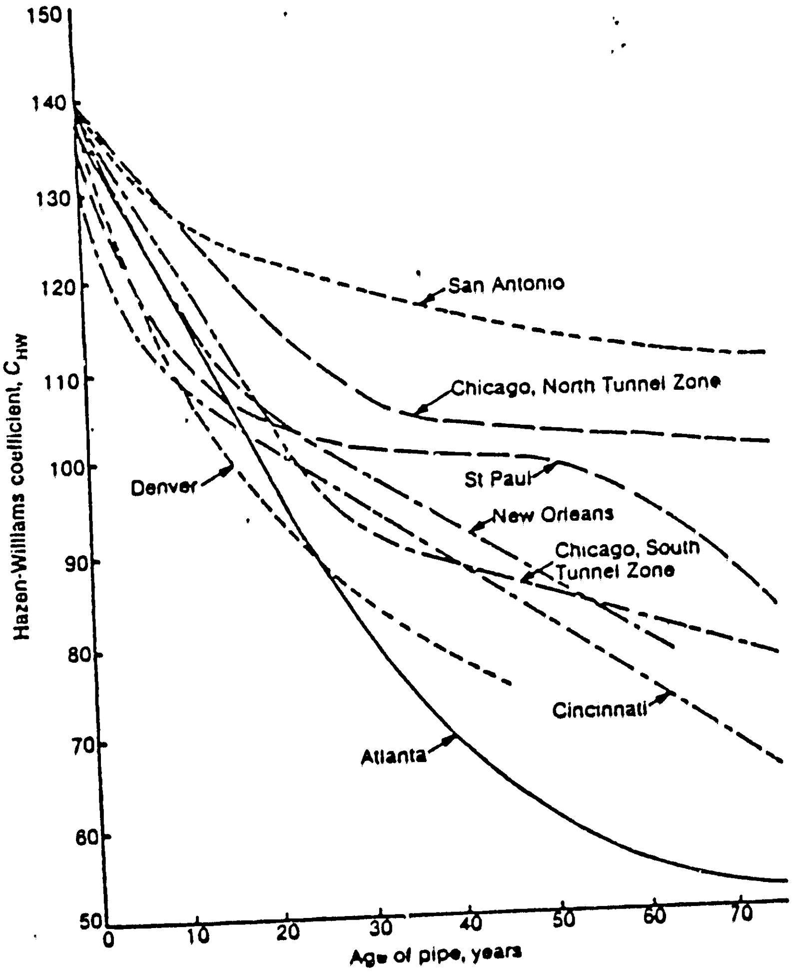


Figure 1: Trend Curves for Head Loss Tests (from McBean et al. (1983))

The primary pipe specific factor for head loss is the diameter. As a pipe's diameter is reduced over time due to tuberculation and formation of various internal deposits, the head loss is increased. McBean et al. suggest that this reduction in diameter commonly reaches 50% of the diameter.

The pipe specific factors which affect corrosion include the material, the diameter, and the interior lining of the pipe. Improper filtration and operation of treatment plants also affects the degree to which the interior surface is covered by tuberculation. The loss of carrying capacity varies with the initial physical layout and installation of the pipes. Poor alignment and large numbers of fittings and bends can reduce the carrying capacity significantly.

In summary, the system specific factors are reflected by the trend curves for the drop in the Hazen-William coeficients. This drop in the coefficients is due in part to the chemical balance of the conveyed water, an environmental factor, and in part to the operation of the treatment plants, a human factor. The pipe specific factors are the type of pipe and lining, and the physical layout and alignment of the pipe within the distribution network.

DESCRIPTION OF THE NEW HAVEN DISTRIBUTION SYSTEM

The model was tested on a sample data set from the water distribution system of New Haven, Connecticut. The data on the system were provided by the EPA. The data base was intended to combine all information presently available on each specific pipe in the system.

The New Haven water data consisted of 1391 individual pipe links whose diameters range from 6 to 48 inches. The oldest pipes in the data on the system date to 1900. The data set included the following covariates for each pipe link: diameter, length, pressure, type of pipe, soil corrosivity, soil stability, proportion of low, medium, and high land development, proportion of swamp land, date of installation, number of breaks, date of first break, date of second break.

General Observations:

Diameter:	11 discrete diameters, range 6-48"
Length:	Range 100-14,000 ft.
Pressure:	400 discrete values 12% had pressure values greater than 100psi.
Pipe type:	96% iron 1.4% concrete 2.2% other materials
Soil Corrosivity:	69% noncorrosive soil (corr=0) 31% corrosive soil (corr=1)
Soil Stability:	53% of pipes in unstable soil 23% moderately stable
Land Developement:	511 pipes (36.7%) covered by 100% low land developement 629 pipes (45.2%) covered by 100% medium land developement 115 pipes (8.3%) covered by 100% high land developement 136 pipes (9.8%) covered by a mixture of the above
Swamp:	95% of pipes not covered by swamp 1% completely covered by swamp

Date of Installation: 46% of pipes installed in 1930-35 small cluster in other years

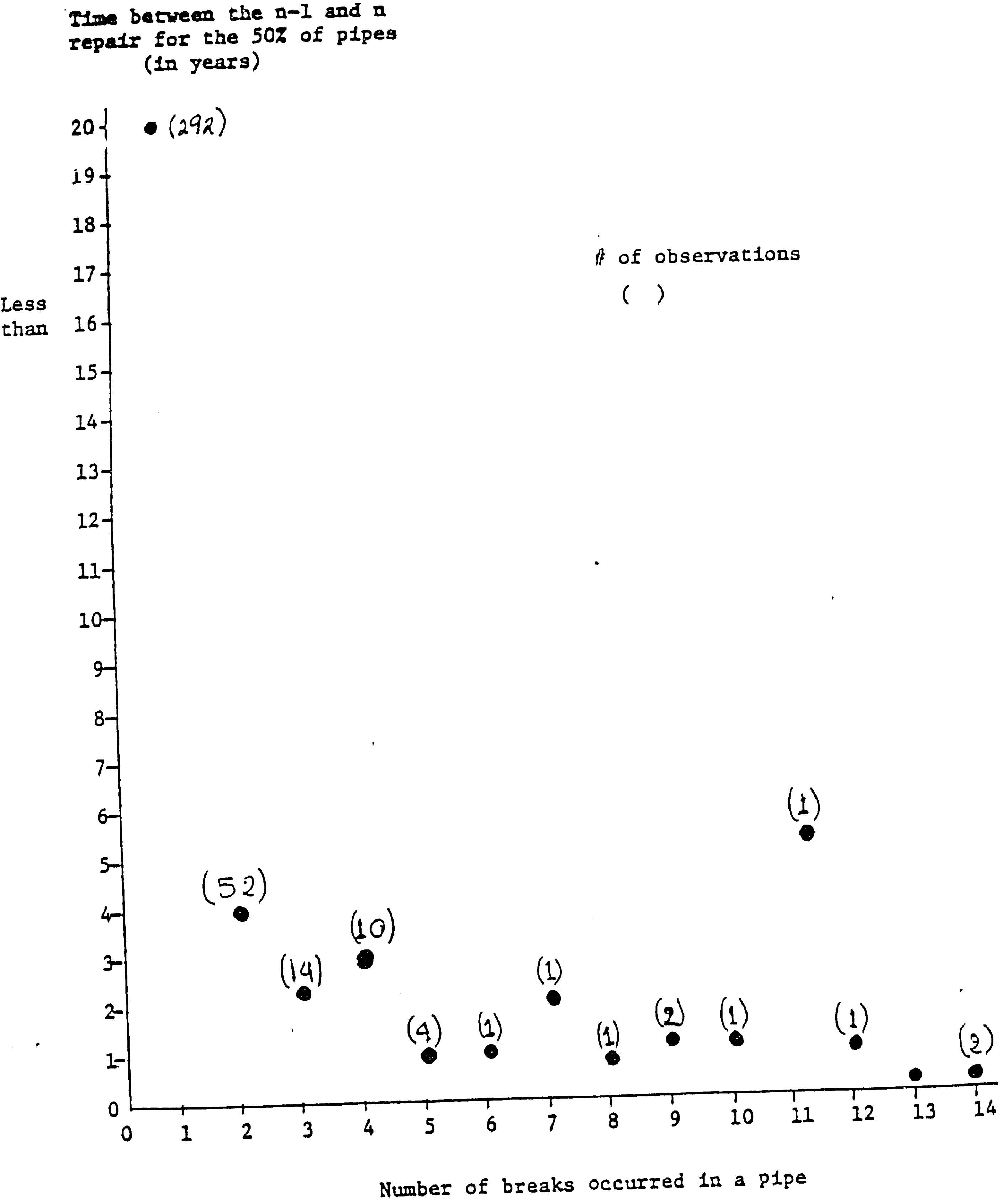
Breaks: 292 pipes (21%) had at least one break
51 pipes (3.7%) had two or more breaks

Time to First Repair: For 50% of the pipes which broke, the first break occurred after the 20th year.
For 4.8% of the pipes which broke, the break occured after the 1st year.
For 10% of the pipes which broke, the break occurred after the 2nd year.

The relation between the time from (n-1) to n repair versus the number of breaks that had already occurred for 50% of the pipes is shown in Figure 2. Although the time to subsequent repairs decreases rapidly with the number of breaks, we have too few observations after the third break to derive any statistically significant results after that point. As a result, the model as formulated for New Haven predicts breaks for pipes which have experienced less than four breaks.

There are several important points which need to be clarified regarding the New Haven data. The definition of pipe length and of break events was not clear from the data set. There are several pipes in the data set with lengths over 14,000 feet. These pipe lengths clearly do not correspond to single pipe units but rather must be related to single units from hydraulic analysis. This created some probems with regard to replacement recommendations. These extremely long pipe units would not be considered as single replacement units in reality. This problem could be avoided

Figure 2: Relation Between Number of Repair Events and Time Between Events



by a precise definition of pipe units as they would tend to be replaced.

There is also some confusion regarding the definition of a break event. The EPA did not clearly define this parameter in the data set. It has been assumed that a break reflects any sort of leak which requires maintenance attention. This difficulty could again have been easily clarified by more careful definition of the parameters as the data base was formed. It would also be useful if any information regarding the type of break or the time of such breaks were available in this data set.

From a meeting with the Regional Water Authority of Central Connecticut, further data was gained on both the month of the failure and the type of failure. A detailed record of break type, size of pipe, month/year of break, and date of installation for breaks which have occured after 1972 was obtained. Pipes in these data are defined by street names so there was little correspondence to the EPA data. This additional data also includes many pipes that do not exist in our data set (6 inch pipes and pipes installed before 1900). The data include a total of 676 breaks divided into four break types: ring crack, hole in pipe, split pipe, and joint leak. These additional data were analyzed for seasonality patterns in break events as well as for correlations to break type. These analyses were undertaken to provide additional information on the failure mechanism of water mains.

The pressure and land use variables were not recorded throughout the history of the system. As a result, the parameters reflect only the current status of the system rather than the temporal fluctuations. It would have been particularly useful to have recorded the pressure values at the time of the break events.

It is apparent that clarity in the documentation of the parameters for a data base is crucial to obtaining useful predictons from our model. The model validity is dependent upon the extent and accuracy of the data available on the distribution system.

CHAPTER TWO

MODEL DESCRIPTION

One of the primary objectives of this analysis is the development of an accurate predictive model for pipe failure. The methodology developed in this report can be easily distinguished from the techniques used by previous researchers which derive the number of breaks for the system in a given year by statistical regression techniques. In this model, the probability of a break in a specific pipe is obtained directly rather than as the expected number of breaks in the entire system. The model is unique in its ability to provide stratification of the data set and in its ability to work effectively with "censored data" sets. Data on pipes which have not yet failed are said to be "censored data". Special procedures must be used to analyze these sort of data.

These features create an advantage for this model as compared to the other regression-based models proposed in the literature. The New Haven Data Set contains primarily large diameter pipes with infrequent break events. In this case, knowing the probability of a break can be of much more use as the concept makes more intuitive sense than the determination of the small fractional number of expected breaks that would be obtained from the other models.

The probability calculation from this model is especially useful as it is dynamic in nature. The probability will be changing in time as the pipe ages since the last break. The probability also changes when a pipe experiences a new break. As a result, one could easily update the failure probability of a pipe at any time from this model.

Another advantage of this model is the usefulness of the results for the management of distribution systems. As the consequences of breaks in large diameter pipes are much more severe and the redundancies built into the system to accomodate possible disruption in service are much less than for small pipes, the management tends to be most concerned with reliable predictions of the break events for large diameter pipes. Similarly, the costs associated with the failure of small mains are relatively low and management is less interested in the prediction of small main failure. As a result, the predictions of this model, which address large water mains with infrequent break events,would be much more useful than the analysis of smaller mains with frequent break events.

A further advantage of the Cox Regression model is its ability to analyze censored data. A major problem in using field data sets to estimate parameters is that not all the pipes have failed. In the data set on New Haven, 80% of the pipes have had zero breaks. Special procedures must be used since the data set includes extensive data not only about the pipes that have failed but also about pipes that have not yet failed. The failure times of these pipes are said to be censored. It has been established in the literature that in order to account for the censoring of failure time data, specialized statistical models and methods are required(Kalbfleisch and Prentice,1980). This model is capable of analyzing data with such characteristics.

The model is extremely flexible regarding the stratification of the data according to various environmental and pipe characteristics. For the New Haven data, stratification groups by length, time of installation, pressure, soil stability, diameter, corrosivity, and land development were examined. By stratifying the data, many factors affecting the break rate were identified that would have otherwise been close to impossible to detect.

The primary predictive models available today will be quickly reviewed at this point to clarify the diffference between our model based on Cox's regression and the other regression based models.

REVIEW OF PREDICTIVE MODELS FOR PIPE FAILURE

Shamir and Howard's model which was published in 1979 was the first comprehensive predictive model developed for water main failure. Historical data on water main breaks and leaks was used as a basis for this model.

Pipe break data was plotted against time with regression techniques used to formulate equations to calculate the predicted number of breaks as a function of time. An exponential growth equation seemed to best fit the data. This regression equation was assumed to hold for each existing pipe in the future so that the number of breaks in any future year could be calculated.

The exponential growth of the number of breaks is thought to be a better approximation of reality for small diameter pipes where breaks are more frequent than for large diameter pipes. The regression equations were developed to determine break growth rates for pools of the data which were grouped with similar characteristics. Shamir and Howard noted that great care must be taken in this pooling process so that homogeneous aggregations are developed. However, it is important to note that this technique would be useful only in the case of small diameter pipes where many observation of breaks exist and thus the estimate of break growth rates can be meaningful and statistically significant.

Following the use of the regression equations to forecast the number of future breaks for each particular length of pipe, an economic analysis was suggested to determine the optimal timing for replacement. This analysis simply used the forecasted number of breaks in the future years, the forecasted number of breaks in a new pipe, the costs of repairing a break and of replacing the main, all discounted to present value for comparison. However, Shamir and Howard did not provide thorough data on the costs of the breaks or a description of how one might estimate these parameters in their model.

The simple regression model suggested by Shamir and Howard is only useful for small pipies where failure leads solely to the simple repair costs. For the larger water mains, the frequency of breaks are low and the impacts associated with break events are more serious due to the lack of redundancy of the larger feeder mains. The failure of such large mains would result in shortages in service as well as higher costs due to damages to the other utilites in the area. These social costs are often much greater than the simple repair costs. As a result, the model proposed by Shamir and Howard would be useful only in cases of small diameter pipes. However, as explained earlier such predictive information for the failure of small mains is much less useful to the management of water distribution systems. It is also obvious that by lumping many of the variables affecting breaks into only one coefficient, as was done in the Shamir and Howard model, insight into the interaction of those variables and their independant effect on the break rate becomes quite limited.

Another regression-based model was developed by Clark et.al.(1982). This model was based on a repair frequency analysis of data from two utilities, a large 260 mgd utility and a smaller 20 mgd utility. An important part of this model was survival analysis performed on repair data for all pipes in these distribution systems. The pipe records included data on the diameter, length, number of breaks, type of pipe, corrosivity of soil, pressure, and age of the pipe.

The data indicated that it was a minority of the pipes that were responsible for a majority of the maintenance events. Clark's model again formulated an exponential growth rate in the number of failures after the first break. The analysis suggested that a lag period occurred between the time the pipe was laid and the first maintenance event. The model also indicated that the time between failures becomes increasingly short as the number of failures increases. As a result, maintenance costs would quickly exceed the cost of replacement making a predictive model quite useful.

Two equations for both utilities were developed, one to estimate the time to the first repair event and the other to estimate the number of events occuring after the first event. The model used the economic analysis of Shamir and Howard to evaluate the optimal time for replacement from the predicted number of breaks in a given pipe at some time in the future.

Clark explicitly states that the equations should not be used for predictive analysis but rather only to indicate the variables which accelerate or retard the failure rate, yet such predictions are the basis of all the derived equations and of the optimal time analysis described in the report. The model also neglected to document the standard errors of the estimated coefficients of the variables.

Walski and Pellicia (1982) proposed another model for determining the optimal replacement time. This model was based on historical records from Binghamton, New York. An exponential function similar to that described by Shamir and Howard was fit to the data, which included the number of previous breaks, the diameter, the pipe type, and the age. An age break function was derived which could predict the break rate for each given pipe for any year in the future. The focus of their approach was the economics of the maintenance alternatives. Costs of water main repair and replacement were evaluated in detail and provided the basis for the decision of whether to replace or repair a main. Such decisions were to be determined from the present worth of the costs of repair and replacement. Following the development of the predictive equation, they applied the optimal time analysis described in the Shamir and Howard model. It seems that the main distinction between this model and that described by Shamir and Howard is the detailed analysis of the costs involved in water main repair and replacement.

The primary limitation of this model is the apparently arbitrary selection of correction factors for predicting the expected number of breaks such as the previous break factor and the pipe size factor. The way those factors act on the predictive model is arbitrarily chosen and no tests were performed to justify why they would act in the proposed way. Again the statistical validity of their predictive model is not given and consequently, it is impossible to determine whether their predictions have any kind of reliability. In this project, all assumptions made were tested and verified to be statistically valid.

These three models are the main models available today which are based on historical records. As stated earlier, another type of predictive model has been developed by O'Day et al. (1984), based on the physical status of the system at present rather than on the historical data on the system performance. It seems that both approaches are valid and that perhaps the next step should be incorporating this approach into the model described in this report.

In summary, the advantages of this methodolgy are it's ability to deal with censored and stratified data sets, it's dynamic nature, and the usefulness of the results to management.

Model Description

The actual model used in this analysis is based on in depth statistical analysis combined with the regression analysis developed by Cox in 1972. The objective was to derive a good predictive model for pipe breaks for the large water mains of the New Haven System. The methodology followed in this analysis can be broken into a series of steps. These steps will be outlined briefly at this point. A detailed description of the analysis and the results will follow in the subsequent chapter.

The first step was the development of descriptive statistics and histograms for all of the available variables. This step provided a more thorough understanding of the system. The ranges and variability of the pipe characteristics and of various environmental factors were demonstrated.

The next step involved bivariate analysis of the data and the calculation of cross correlation coefficients of the pairs of variables. Bivariate analysis was performed to determine possible correlation between pipe failures and other variables provided by the data set. Three types of correlation analysis were used; parametric, Spearman (non-parametric) and Kendall's Tau (non-parametric) correlation. Several types of correlation analysis were used due to the very skewed distributions of many of the variables and as data points were very often grouped in clusters and concentrated in particular regions, thus requiring more sophisticated weighting schemes. The SAS statistical package of the IBM computer system was used for this step in the analysis.

Survival analysis was then performed in order to reveal the survival pattern of the various categories of pipes. The survival time to the first break was investigated in detail by grouping the pipes according to the following predictor variables: period of installation, length, pressure, soil stability, diameter, corrosivity and land development. Survival curves were plotted for each case and interesting patterns were revealed. This preliminary statistical analysis strongly suggested that a proportional hazard model where variables could act multiplicatively on the hazard rate, would be appropriate for deriving a predictive model for pipe failure.

The proportional hazard model as developed by Cox in 1972 was used to develop the actual probabilistic results. The Cox model is a general failure prediction model which can be applied to pipe data to calculate the probability of a pipe experiencing a break or failure in any given time as a function of certain break causing factors. Cox's regression was applied to the data set in order to determine which variables were important in affecting the probability of failure of a particular pipe.

There are two basic assumptions underlying the use of this model. The first assumption implies that there are several break causing factors which have a multiplicative effect on the break rate. It seems reasonable to believe that there are numerous factors acting together to bring about the final deterioration of the pipe wall and the occurrence of a break. The second assumption is the presumption of a log-linear relation between the break rate and the break-causing factors. Evidence from various other models has indicated that once breaks start to occur in a pipe, they tend to increase exponentially (Shamir & Howard, 1979, Clark, 1982). This assumption of an exponential relation can be justified to a certain degree by the results from these other models. Both assumptions were tested and found to be statistically valid for the model proposed.

In order to accommodate the fact that one pipe can experience several breaks, a new data set was created which considered every pipe as a new observation after each break. Many preliminary computer runs were performed including all available variables that could affect the break rate. The application of a stepwise regression procedure indicated which variables were appropriate to use in the final model. The data set was also initially stratified according to pipe installation periods, pipe diameter, soil corrosivity, etc. Cox's regression was performed independantly for each stratification variable. The results of this analysis helped in choosing the appropriate covariates for the final model and also in creating new variables reflecting the various stratification patterns. Goodness-of-fit tests, log likelihood functions, p-values, and global chi-square statistics were computed for each run. Several runs were performed to test whether Cox's Regression assumptions were valid in our final model and also to investigate the existence of time-dependant covariates that might need to be included in the model.

Before an in depth mathematical description of the model can be formulated, several basic definitions must be established. Let $T \geq 0$ be a random variable representing life times of individuals in some population, and let

$$\underline{x} = (x_1, \ldots, x_o)^T$$

be the column vector of independant variables called "covariates" associated with each individual pipe, such as age, length, etc., in the data. Then the survivor function,

$$S(t:\underline{x}) = \text{Prob}(T \geq t \mid \underline{x})$$

$$h(t:\underline{x}) = \lim_{\Delta t \to 0} \frac{\text{Prob}(t \leq T \leq t+\Delta t \mid T \geq t, \underline{x})}{\Delta t}$$

$S(t:\underline{x})$ and the Hazard function, $h(t:\underline{x})$, given the covariate $\underline{x}$.

If $f(t:\underline{x})$ and $F(t:\underline{x})$ are the probability density function and the cumulative distribution function respectively of T given $\underline{x}$, then

$$S(t:\underline{x}) = 1-F(t:\underline{x})$$

$$h(t:\underline{x}) = \frac{f(t:\underline{x})}{S(t:\underline{x})}$$

The proportional hazard model assumes

$$h(t:\underline{x}) = h_o(t)\exp(\underline{\beta}^T\underline{x})$$

where $h_o(t)$, β are unknown, and where $\underline{\beta}^T\underline{x}$ is $\beta_1x_1 + \beta_2x_2 + \ldots + \beta_px_p$.

The function $h_o(t)$ is an arbitrary unspecified base line hazard function. The covariates act multiplicatively on the hazard function. $S(t:\underline{x}) = \exp[-\int_0^t h(u:\underline{x})du]$

$$= [S_0(t)]^{\exp(\underline{\beta}^T\underline{x})}$$

under the proportional hazard model, where

$$S_o(t) = \exp[-\int_0^t h_o(u)du]$$

represents a baseline survival curve.

Stratification of the data set:

The fact that the function $h_o(t)$ can be arbitrary gives the model great flexibility. The following common generalization was useful for the present data set: h (t) can vary in specific subsets of the data. If the set of pipes is divided into r strata and the hazard $h_j(t:x)$ in the jth stratum depends on an arbitrary shaped function $h_{oj}(t)$, then we can write:

$$h_j(t;x) = h_{oj}(t) \exp(\underline{\beta}^T x), \; j=1, \ldots, J.$$

This procedure is useful if, for example, an explanatory variable does not have a multiplicative effect on the hazard function. The range of this variable can then be divided into strata. In our data set such strata could correspond to various installation periods, pipe types, diameters, etc. The stratification variable is then removed from the set of covariates. Although the baseline hazard rate varies by stratum, the covariate coefficients are assumed constant across strata. Cox's regression estimates in the above expression using the method of maximum partial likelihood estimation.

Model Summary

The Cox regression model seems to be a better way to obtain accurate predictions of failure probabilities from large data sets on water distribution systems. This method provides the ability to work effectively with censored data sets so that information can be obtained from the large part of the data sets which does not yet have a break history. The model provides the ability to effectively stratify the data thereby exposing trends related to different age groups of pipes, different pipe characteristics and different environmental factors. The model derives a hazard rate as the probability of a break at a particular time given that the pipe has survived up to that time since its last break. This probability calculation is much more useful in the estimation of reliability at the network level than the previous attempts at estimating the expected number of breaks.

Following the regression analysis, several other analyses were performed. An excercise examining the effect of left censoring of data was examined by considering only the data for break events recorded after 1960. This analysis demonstrated the ability of our model to make accurate predictions from more limited pipe data. The data obtained from the Regional Water Authority of Central Connecticut was analysed for seasonality patterns and correlation to the additional break type information.

CHAPTER THREE

APPLICATIOM OF THE COX REGRESSION MODEL TO NEW HAVEN

PRELIMINARY STATISTICS

As described in Chapter One, the original data on the New Haven distribution system contained the following parameters: diameter, length, pressure, pipe type, soil corrosivity, soil stability, land development, swamp, date of installation, number of previous breaks, and time to first repair. Histograms are shown in Appendix A.

Bivariate Analysis:

The next step was the development of bivariate scatter plots of pairs of variables believed to be important in the analysis. These plots are shown in Appendix B where the number of breaks is plotted against the following variables: AGE83, Diameter(DIA), Pressure, Diameter/length(DIAL), Pressure/length(PREL), Diameter/pressure(DIAP), and Diameter * length(SUR). No clear trend was apparent from such plots.

Correlation Analysis:

1) Parametric correlation

The calculated correlation coefficients are shown in Appendix C for most of the pairs of variables of interest. The values of correlation coefficients were very low for all the variables that could be used as potential predictors of failure (calculated values did not exceed the range of 0.20 for such variables). Low correlation indicates the independence of the variables in the data set.

2)Spearman (non-parametric) Correlation

A non-parametric correlation analysis seemed appropriate because almost all the variables of interest were not normally distributed. The results from the Spearman correlation are shown in pages 8-10 of Appendix C. Again, no statistically significant correlations were found.

3) Kendall's Tau (non-parametric) Correlation

The correlation coefficients calculated from Kendall's non-parametric correlation are shown in Appendix C (page 11). No statistically significant correlations were found if we take into account that the standard error for this type of correlation is given by: s.e. $\approx\sqrt{\frac{4}{9N}}$ where N is the number of observations.

4) Other Observations:

A set of 29 pipes broke within the first two years of their installation. Since these pipes might have been defective for different reasons not related to our analysis, they were excluded from the data set. However, it was not expected that the results would be affected significantly, since these pipes are very few in number. Their characteristics are shown in page 12 of Appendix C. No apparent differences from the set of the other pipes are observed. 24% of them were installed before 1931 while 72%

were installed after 1957.

From the correlation analysis, it was evident that the variables more strongly correlated with the previous number of breaks and the time to first break also vary according to soil type, and the historical period during which the pipes were installed. A summary of this work which shows the degree of correlation between the variables BREAKS(number of previous breaks) and YR1RPRS(time to first break) with some other variables of interest is shown in table 1. (Low correlation implies $r<0.3$, medium: $0.3<r<0.7$, high: $r>0.7$)

Survival Analysis:

Detailed results of the survival analysis are shown in Appendix D. First, a survival curve for time to first break (for all pipes) is shown in figure 1 of the appendix. We observe that the percentage of pipes without a break drops sharply after about the 55th year of installation (this group represents pipes installed before 1930).

The survival time to the first break was then examined by grouping the pipes according to predictor variables. Pipes without breaks were censored at the age they survived without a break. The results can be summarized as follows:

Date of Installation:

Pipes were grouped according to the historical period of their installation as follows: Installed before 1930,1930-35, 1936-50, 1951-65, 1965-present. Significant differences

TABLE 1

Period of Installation	Type of Soil	No. of Observ.	Degree of correlation with # of BREAKS from calculated Pearson correlation coefficients						
			DIA	LENGTH	PRESSURE	DIAL	DIAP	PREL	SUR
1900-30	1=Corr.	24	Low	High	Low		Low	Medium	High
	0=Non-Corr.	112	Medium	Low	Low	Low	Low	Low	Low
1930-35	1	165	Medium	Low	Low	Low	Low	Low	Medium
	0	474	Low	Low	Low	Low	Low	Low	Low
1935-45	1	54	Low	Medium	Medium	Low	Low	Low	Medium
	0	91	Low	Low	Low	Low	Medium	Low	Low
1945-58	1	61	Low	Low	Medium	Low	Medium	Low	Low
	0	91	Low	Low	Low	Low	Low	Low	Low
1958-65	1	71	Low	Low	Low	Low	Low	Low	Low
	0	107	Low	Low	Low	Low	Low	Low	Low
1965-83	1	69	Low	Low	Low	Low	Low	Low	Low
	0	119	Low	Low	Low	Low	Low	Low	Low

TABLE 1 (continued)

Period of Installation	Type of Soil	No. of Observ.	Degree of correlation with YR1PRS from calculated Pearson correlation coefficients							
			DIA	LENGTH	PRESSURE	DIAL	DIAP	PREL	SUR	BREAKS
1900-30	1=Corr.	15	Low	Low	Low	Low	Low	Low	Low	Medium
	0=Non-Corr.	61	Low	Low	Low	Low	Low	Low	Low	Low
1930-35	1	24	Medium	Low	Low	Low	Medium	Low	Low	Low
	0	75	Low	Low	Low	Low	Low	Low	Low	Low
1935-45	1	9	Low	Medium	Low	Medium	Low	Medium	Medium	Medium
	0	25	Low	Low	Medium	Low	Medium	Medium	Medium	Low
1945-58	1	11	High	Low	Low	Low	Low	Low	Low	Medium
	0	17	Low	Medium	Low	Medium	Low	Medium	Medium	Low
1958-65	1	8	Low	Low	Low	Medium	Low	Low	Low	Low
	0	23	Medium	Low	Low	Low	Medium	Low	Medium	Medium
1965-83	1	6	Medium	Medium	Low	High	Medium	High	Low	Medium
	0	23	Low	Medium	Low	Low	Low	Low	Medium	Low

between groups were observed (in Gen Wilcox p=.0000). It appears that pipes installed after 1965 performed worst, and pipes installed between 1930-35 performed best. Of course this analyasis is limited by the fact that the various groups of pipes correspond to pipes of different ages, thus no good comparison can be made between old and young clusters of pipes.

Length:

Pipes were grouped as follows: Length less than 500 ft., 500-1000, 1000-1500,1500-2500, greater than 2500 ft. We observe that longest pipes have earliest breaks. "Dose effect" present (Gen Wilcox in p=.000).

Pressure:

Pipes were grouped as follows: Pressure less than 60, 60-80, 80-100, greater than 100. We observe that pipes with high pressure had earlier breaks, but there is a crossover for older pipes (p=0.0002).

Soil Stability:

We observe that unstable soil initially causes earlier breaks but as pipes are getting older, the opposite happens (p=.02)

Diameter:

Pipes were grouped as follows: diameter less than 8, 9-10, 11-12, 13-16, greater than 16 inches. No clear pattern can be observed, although for very old pipes it seems that pipes with diameters less tham 8" are starting to break earlier.

Corrosivity:

It appears that pipes in noncorrosive soil had slightly earlier breaks than those in corrosive soil.

Land Development:

Pipes were grouped as follows: N = no land development, S = some or full land development. No clear pattern was identified.

COX'S REGRESSION

In order to look at multiple covariates associated with survival time, Cox's regression was performed with various models. After several trials in various settings of stepwise regression, the following variables were identified as the initial "basic" varibles in describing the proportional hazard model:

LNLENGTH = the natural logarithm of pipe length in feet.

PRESSURE in 10 PSI.

LOW = the proportion of low land development

C35 = 1 if 1930<= Installation Date <= 1935
0 otherwise.

C50 = 1 if Installation Date >= 1950
0 otherwise

AGE = the age of the pipe at the last break, in years

PREVBRK = the number of previous breaks experienced by a pipe at a given time.

TABLE 2

DESCRIPTIVE STATISTICS FOR PREVBRK = 0, 1, 2

<u>PREVBRK = 0</u> (N=1391)

	MIN	MAX	MEAN	ST.DEV.
LNLENGTH	4.61	9.55	7.23	0.81
PRESSURE	0.4	17.3	7.93	2.02
LOW	0	1	0.40	0.48
C 35	0	1	0.46	0.50
C 50	0	1	0.30	0.46

<u>PREVBRK = 1</u> (N=292)

	MIN	MAX	MEAN	ST.DEV.
LNLENGTH	5.52	9.55	7.55	0.82
PRESSURE	3.0	17.3	8.40	2.07
LOW	0.0	1	0.33	0.45
C35	0	1	0.34	0.47
C50	0	1	0.24	0.43
AGE	0.0	80.0	22.79	17.94

<u>PREVBRK = 2</u> (N=90)

	MIN	MAX	MEAN	ST.DEV.
LNLENGTH	6.21	9.11	7.68	0.74
PRESSURE	3.0	15.5	8.19	1.95
LOW	0.0	1.0	0.26	0.42
C35	0	1	0.28	0.45
C50	0	1	0.19	0.39
AGE	2.0	77.0	28.66	17.85

The descriptive statistics for PREVBRK = 0,1,2 are listed in table 2. The results from the separate Cox regressions according to the previous number of breaks are shown in Table 3. (RR = the relative risk incurred by doubling the length, by increasing the pressure by 10 PSI, by changing LOW = 0 to LOW = 1, or by adding 10 years to the age at the last break.)

TABLE 3: COEFFICIENT ESTIMATES, STANDARD ERRORS AND RELATIVE RISKS

	(1391 OBS) PREVBRK=0		(292 OBS) PREVBRK=1		(90 OBS) PREVBRK=2	
GLOBAL χ^2	134.4(5D.F.) p=0.0		14.3(6D.F.) p=0.0263		16.3(6D.F.) p=0.012	
	COEFF(St.Error)	RR	COEFF(St.Error)	RR	COEFF(St.Error)	R
LNLENGTH	0.581 (0.079)	1.50	0.345 (0.141)	1.27	0.289 (0.273)	1.
PRESSURE	0.089 (0.028)	1.09	-0.048 (0.054)	0.95	-0.217 (0.098)	0.
LOW	-0.58 (0.14)	0.56	-0.46 (0.29)	0.63	-0.42 (0.45)	0.
C35	-0.688 (0.146)	0.50	-0.563 (0.253)	0.57	-0.802 (0.491)	0.
C50	0.491 (0.173)	1.63	-0.144 (0.342)	0.87	-0.316 (0.517)	0.
AGE	--	—	-0.006 (0.008)	0.94	-0.028 (0.013)	0.

This table reflects the differences in the coefficients for different numbers of previous breaks, especially in those of PRESSURE and AGE. For example, if PREVBRK = 0, the coefficient of PRESSURE is positive, while if PREVBRK = 1 or 2, it is negative. As a result, the final model used the following modified covariates:

LNLENGTH = natural logarithm of pipe length in feet

PRESBRK = (PRESSURE/10),if PREVBRK = 0
0 otherwise

LOW = Percentage of low land development covering the pipe, measured from 0 to 1.

C35 (dummy variable) = 1 if installed between 1930-35
0 otherwise

C50 (dummy variable) = 1 if installed after 1950
0 otherwise

AGEBRK = AGE of pipe at time of last break, if PREVBRK is greater or equal to 2
0 otherwise

PREVBRK = number of previous breaks occurred in the pipe.

The coefficients, standard errors, coef/se, and relative risk factors calculated for the final model are shown in Table 4. All of the independent variables in this final formulation of the model are statistically significant. The relative risk values listed in the table infer that if one pipe is twice as long as another, its probability of breaking is 45% greater, other things being equal. Similarly, increasing the pressure by 10 PSI will increase the hazard rate by 7%; increasing the proportion of low land development from 0 to 1 will decrease the probability of break by 58% of the baseline value. If a pipe was installed between 1930 and 1935, the hazard rate would drop to 51% of the baseline value. Experiencing a

TABLE 4: COEFFICIENTS, STANDARD ERRORS, COEF/SE, AND RELATIVE RISK FACTORS OF FINAL MODEL

VARIABLE	COEFFICIENT	STANDARD ERROR	COEF./S.E.*	RR***
LNLENGTH	0.537	0.066	8.069	1.45
PRESBRK	0.066	0.023	2.835	1.07
LOW	0.533	0.122	-4.538	0.58
C35	-0.680	0.121	-5.631	0.51
C50**	0.271	0.141	1.924	1.31
AGEBRK	-0.022	0.009	-2.433	0.80
PREVBRK	1.186	0.1830	6.480	3.27

LOG LIKELIHOOD = -2805.0324

GLOBAL CHI-SQAURE = 279.27, P-VALUE = 0.0

*This is similar to a t-statistic - general limit statistics says values >2 are significant estimates.

**C50 was included although it is marginally statistically significant, because its standard error was not so large as to be ignored completely and also shows the contrast to the pipes installed in the 1930-35 period.

***RR defined as in Table

TABLE 5: DESCRIPTIVE STATISTICS OF VARIABLES IN THE FINAL MODEL

PREVBRK≤2 (N=1773)

	MIN	MAX	MEAN	ST.DEV.
LNLENGTH	4.61	9.55	7.31	0.82
PRESBRK	0.0	17.3	6.23	3.72
LOW	0.0	1.0	0.38	0.47
C35	0	1	0.43	0.50
C50	0	1	0.28	0.45
AGEBRK	0.0	77.0	1.45	7.46
PREVBRK	0	2	0.27	0.55

second break at an age 10 years older than others reduces the baseline hazard by 80%. Each number of previous breaks will multiply the hazard rate by 3.27.

Descriptive statistics of the variables in the final model are listed in Table 5. In order to explain the development of the model, all variables found to be significant will be carefully evaluated at this point.

Pressure

Pressure appears to be important as a predictor of breaks only if a pipe has not experienced any breaks in the past. It is reasonable to believe that with time, the stresses imposed on the pipe wall due to high pressure would contribute to the occurrence of a break. On the other hand, if the pipe has already experienced a break, pressure is not an important predictor because the other variables such as the past number of breaks become much more important in predicting the hazard rate.

Land Development

Land development is directly related to external loads transmitted on pipes from buildings, trucks, etc. The effect of frost penetration also becomes much more important with the presence of high land development because the dynamic loads from trucks or other moving objects transmitted on the pipes can become very high. Thus, land development can be seen as a surrogate for all of these forces. The fact that land development has great explanatory power even in the case where breaks in the pipe have already occurred could imply that the type of dynamic loading present with high land development is more important in contributing to stresses related to breaks than is pressure.

Age At Time of Second Break

The stastical analysis has shown that the age of a pipe at the time of the second break is a very important predictor of the hazard rate. The less the age of the pipe at the second break, the higher the chances for the pipe to break in the future. A possible physical explanation for this may be that a pipe experiencing a second break early in its life is somehow more defective than a pipe which had a second break late in its life.

Installation Period

The fact that pipes installed during the period 1930-35 performed much better than pipes installed after 1950 could be explained by the construction practices and the materials used to make pipes during the various periods. Similar findings are reported by O'Day el al.(1984), where it is pointed out that changes of technology have not always improved water main reliability. In the city of Philadelphia, it was found that mains laid in the 1948-1952 period experienced a high rate of failure due to leadite joint related problems. This lead substitute material caused split bell failures even though the main wall was in relatively good condition. It was also mentioned in the same study that increase in cast iron unit strength has occurred after the 1930 period, thus mains laid in the 1930-1980 period are inherantly stronger than mains laid before 1920. The Regional Water Authority of Central Connecticut also correlated this pattern to the pipe material. It is believed that pit cast iron pipes which were primarily installed in the 1930's had higher safety factors than the more recent sand spun pipes. All of these findings support the variables related to construction periods that we found to be statistically significant in our study.

Number of Previous Breaks

The number of previous breaks which occured in a pipe was found to be a very important predictor of future failures. Each previous break in a pipe tripled the probability of the next break. Since the break causing mechanisms and their interaction are not yet well understood, occurrence of breaks in a pipe simply reveals the fact that those break mechanisms are present in such a pipe and will thus increase the hazard rate for that pipe in the future relatively more than that of pipes which have experienced less or no breaks in the past.

Length of Pipe

In the model, the hazard rate for an individual pipe varies proportionately to exp(.537*ln(Length)) or approximately to the square root of length. This implies that the per foot hazard rate in any pipe is inversely proportional to the square root of the length. Thus, longer pipes tend to have fewer breaks per foot of pipe than shorter pipes. It is not possible, with the available information, to give a good physical explanation for this fact. It can be argued that shorter pipe lengths are expected to be observed in more densely populated areas with higher land development. Pipe length could thus serve as a surrogate for forces acting on pipes because of land development, and it might be the case that it supplements the variable for land development that appears in the model.

Age of Pipes

In order to evaluate the effect of aging on the pipes, the baseline hazard function, h (t), which evaluates the the effect of time in the survival process must be estimated. In our model, the baseline hazard function was approximated by a second degree polynomial. The BMDP program plots a nonparametric estimate of the baseline hazard rate. This was approximated by simply fitting a 3-parameter curve to the plot, resulting in a smoothed estimate of $h_o(t)$ as:

$$h_o(t) = 0.000210476 - 0.000011171t + 0.000000199\ t^2$$

where t is the survival time since last break measured in years.

A hazard function of this type indicates that the hazard rate initially decreases as the pipe does not experience any breaks in the early years of installation, or early after a previous break. The effects of the aging process appear after approximately 28 years have passed, as the hazard rate starts increasing after that time.

This finding makes sense intuitively as it is very reasonable to believe that if a pipe had no break early after installation, the pipe was laid in relatively favorable conditions. That is, the pipe is not as defective and/or the break causing factors are so severe for that pipe as compared to the others. Also, it is reasonable to believe that age contributes to pipe failure much later after installation when internal corrosion has developed to a greater extent. Thus, in general the probablility of a break initially decreased if the pipe does not experience early breaks, and starts increasing again when the effect of age becomes important.

In order to interpret the results of our analysis, it is helpful to define a "reference pipe",describe its chance of breaking as a function of age, and then describe the risk of other pipes breaking relative to that of the "reference" pipe. We arbitrarily define a reference pipe as one in which:

Length = 1500 ft (~average in N.H.)
Pressure = 80 PSI
LOW = 38%
PREVBRK = 0
DATE = before 1930

For such a pipe, the rate of breaks at age t is

$$h(t) = h_o(t) \cdot e^{4.19}$$

where

$$h\;(t) = 0.000210476 - 0.0000111171(t) + 0.000000199(t)^2$$

$$Bz = 4.19.$$

Thus, we see that yearly risk of a break for the reference pipe changes with age from about 0.00021 at age 0 to 0.00005 at age 30, to 0.00059 at age 80.

Figures 3 and 4 present the histograms of the covariates involved in the model and two different types of hazard rates. RH is the estimated relative hazard of each pipe as a whole, while RHPF is the relative hazard per foot of length of each pipe where: $RH = \exp\{\beta^T(x-x_0)\}$

$$RHPF = RH \div (length/1500)$$

B is the estimated coefficient
x is the covariate vector of any particular pipe
x is the reference covariate vector corresponding to
LNLENGTH = 7.31, PRESBRK = 80, LOW = 0.38, C35 = 0
C50 = 0, AGEBRK = 0, and PREVBRK = 0.

The two histograms are displayed on a log scale of relative risk and represent N = 1771 observations.

Note that both the hazard rate per pipe and the hazard rate per foot show great variation. Many pipes are 10 or 20 times as likely to break as are the other pipes. This display

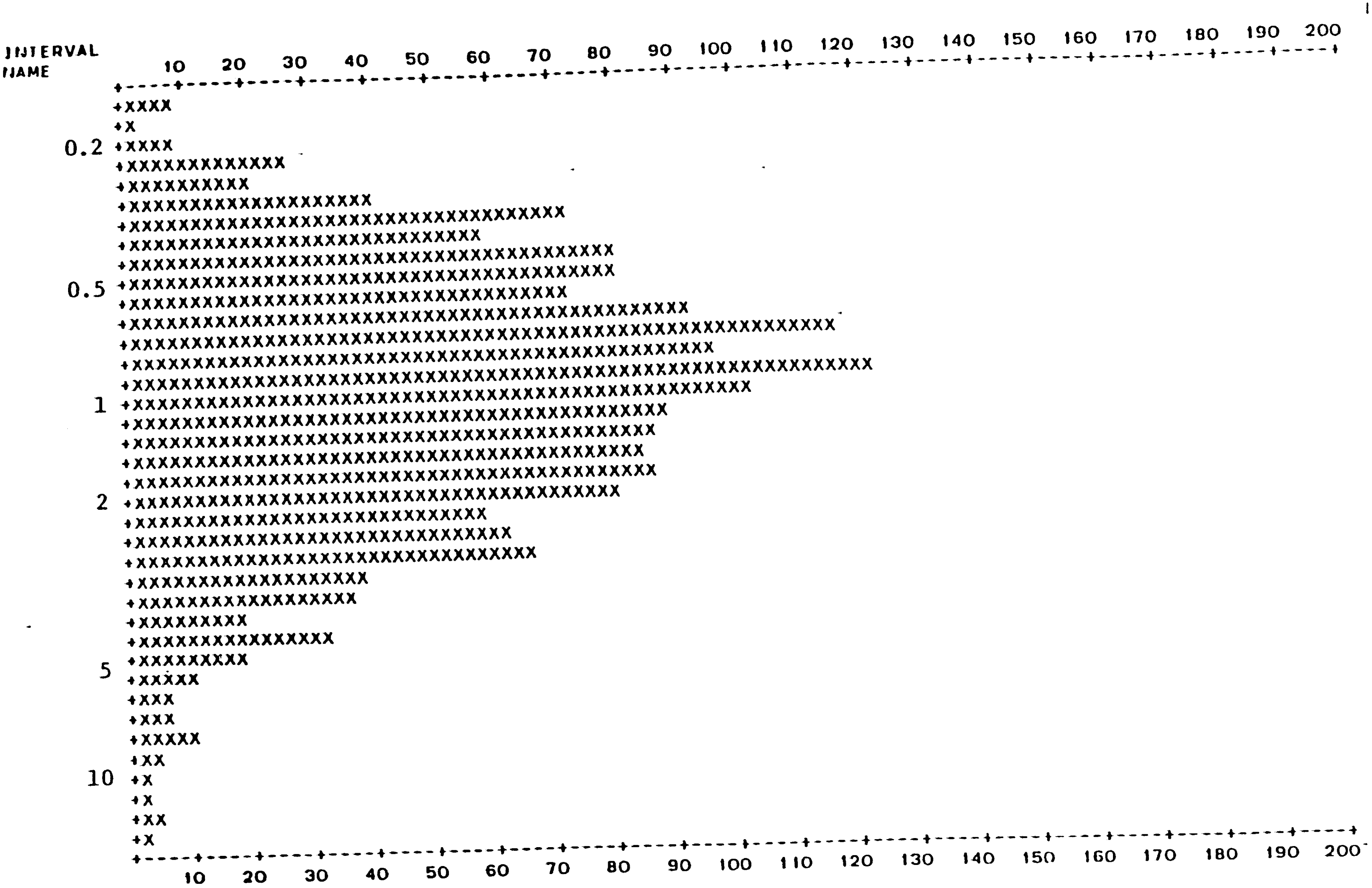


FIGURE 3: HISTOGRAM OF RELATIVE HAZARD RATE (RH)

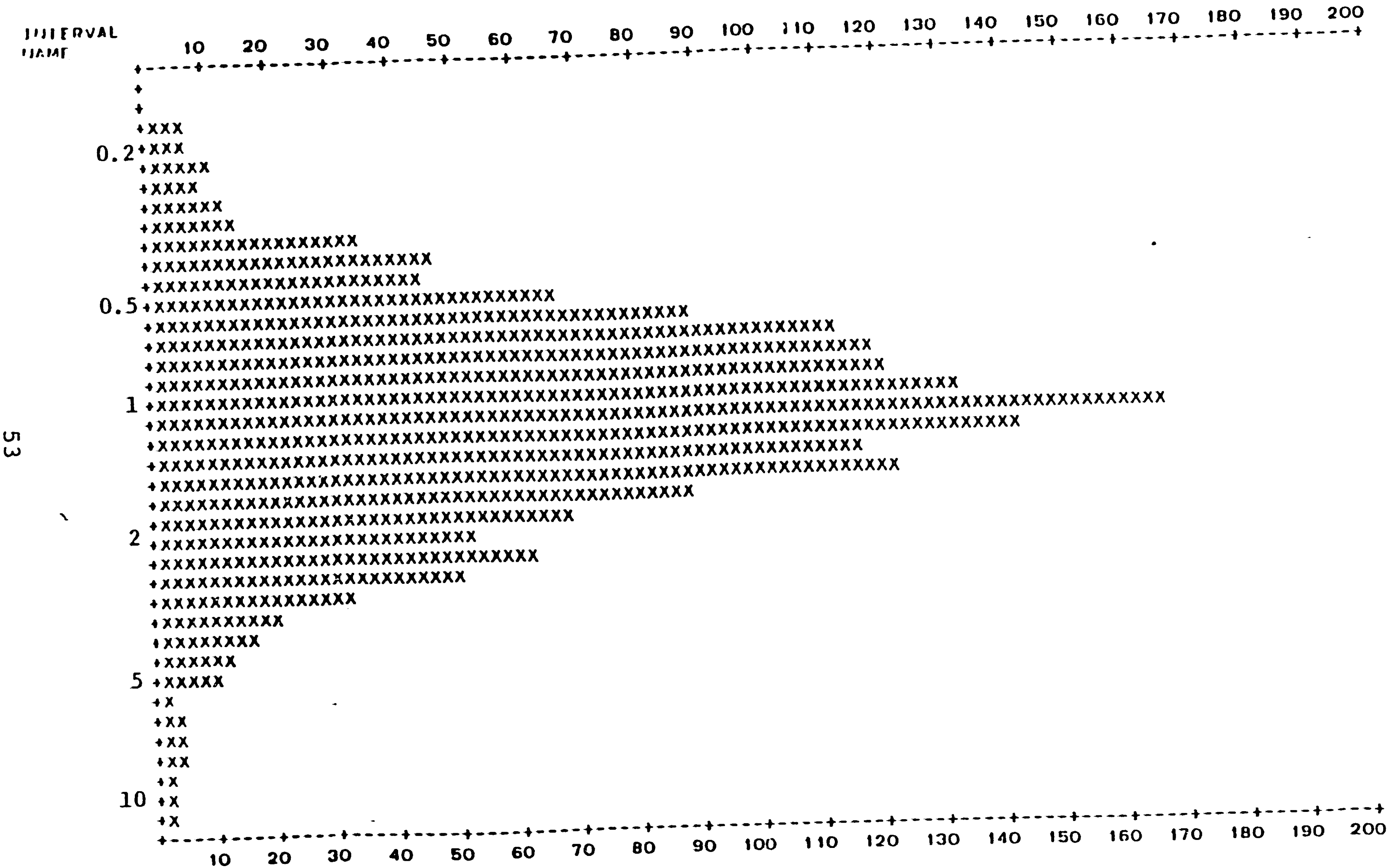
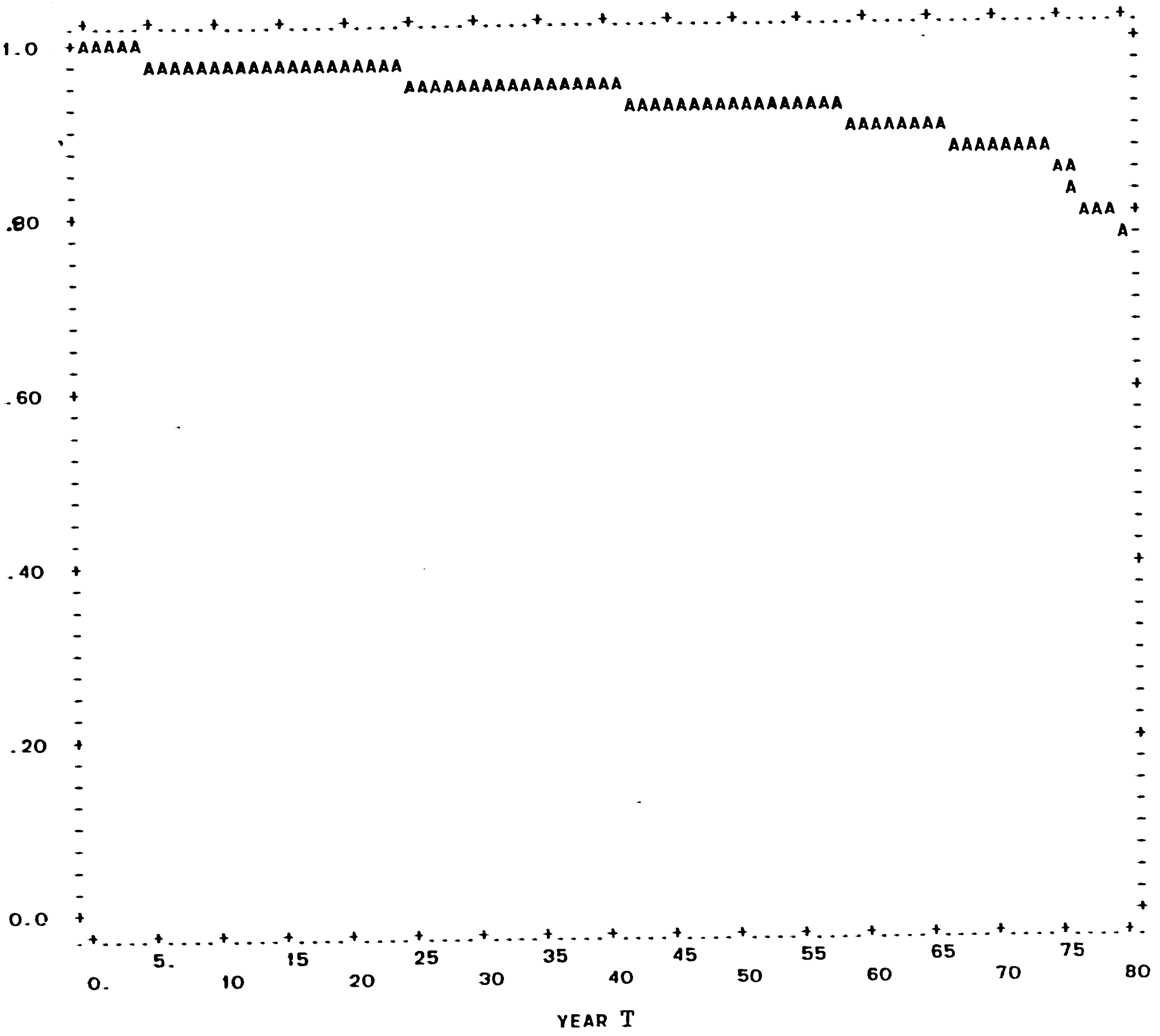


FIGURE 4: HISTOGRAM OF RELATIVE PER FOOT HAZARD RATE (RHPF)

shows that there is potentially a large advantage to be gained by identifying pipes with high hazard rates and by designing maintenance policies accordingly.

Examples of Pipe Survival Functions

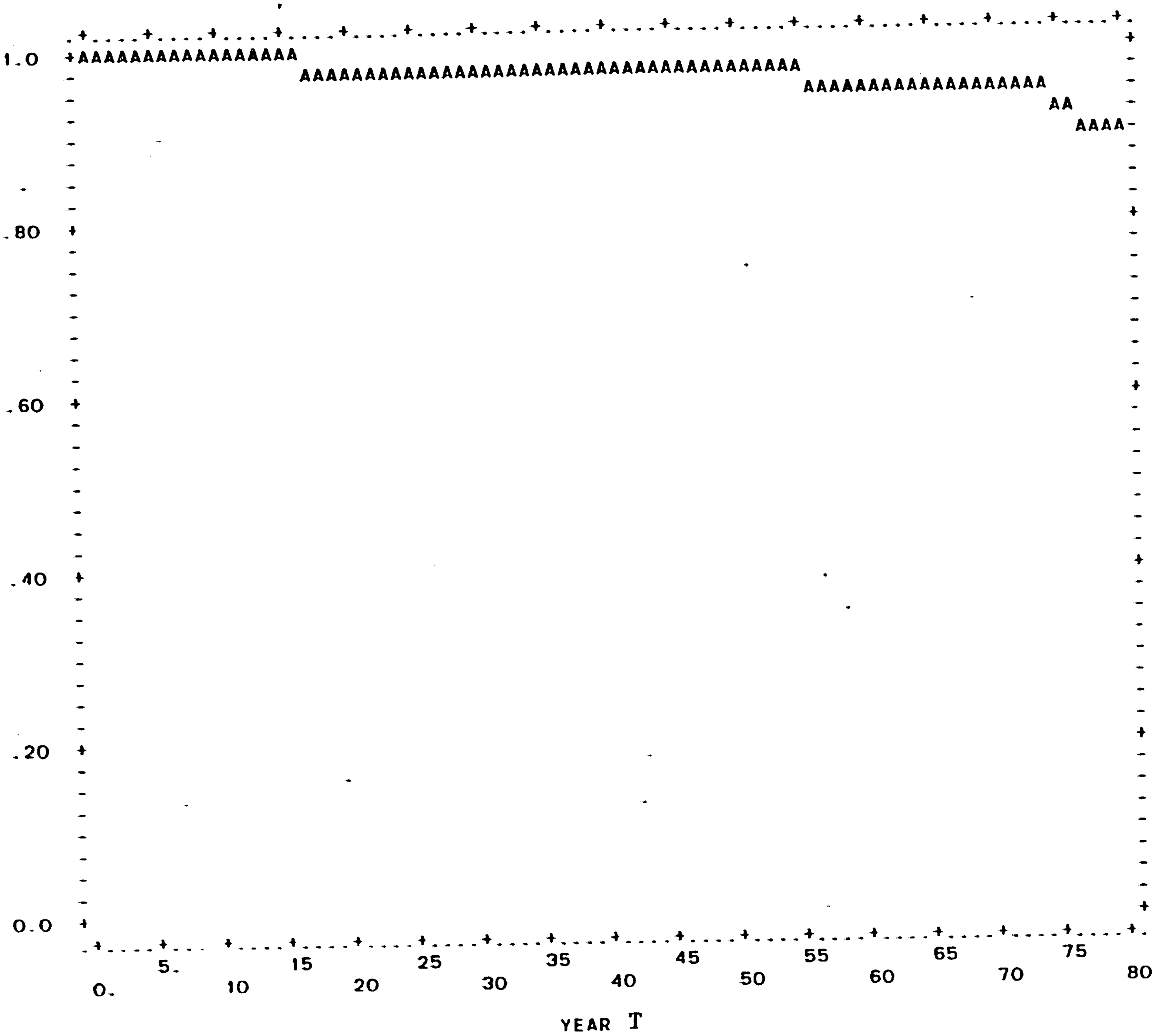
Figure 5 shows the estimated survival function of 100 ft pipe installed during the 1930-35 period which has experienced only one previous break and is covered by moderate (50%) land development. We observe that the probability of survival decreases with time since the last break, but the decrease is not so dramatic. Figure 6 shows the estimated survival function of a 100 ft pipe installed during 1930-35, covered totally with minimum land development which has already experienced 2 breaks with the last break occurring after 77 years from time of installation. We observe a more significant increase of the probability of failure than for the pipe in Figure 5. Figure 7 shows the survival function of a 100 ft pipe with no previous breaks installed after 1950, covered with maximum land development and subject to very high internal pressures of 173 PSI. The slope of the survival curve indicates the importance of internal pressure in decreasing the probability of survival for a pipe with no previous breaks. Figure 8 shows the dramatic drop of the probability of survival for a "worst case" pipe 14,000 ft long, with two previous breaks and which had the second break only 4 years after installation. Figure 9 shows the survival curve of a 100 ft pipe with 2 previous breaks installed after 1950, covered with maximum land



LENGTH = 100 ft.
PRESBRK = 0
D1 = 1
D3 = 0
AGEBRK = 0
PREVBRK = 1

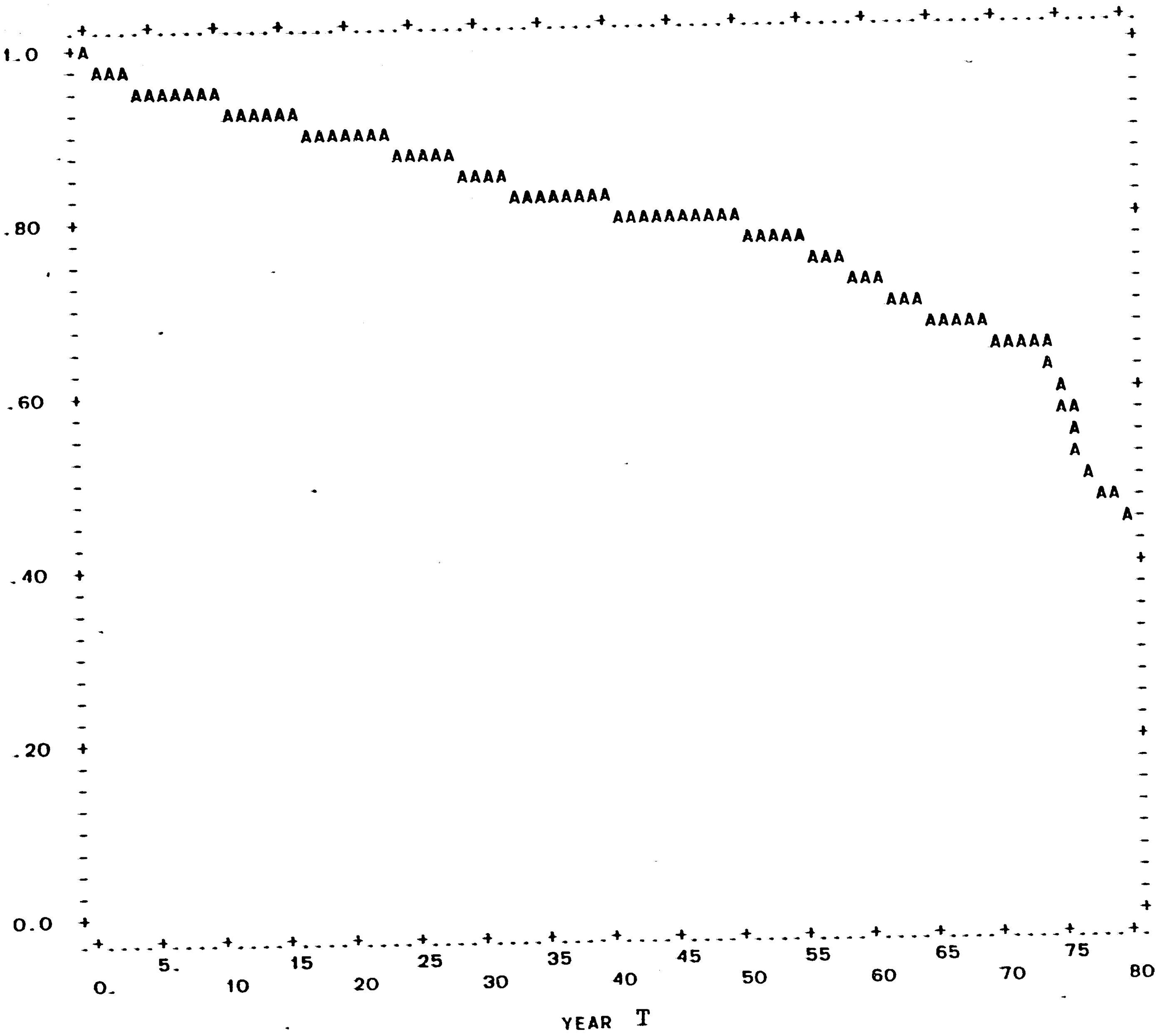
FIGURE 5:

PROBABILITY THAT THE PIPE WILL NOT EXPERIENCE A BREAK BEFORE YEAR T SINCE THE LAST BREAK



LENGTH = 100 ft
PRESBRK = 0
LOW = 100%
D1 = 1
D3 = 0
AGEBRK = 77
PREVBRK = 2

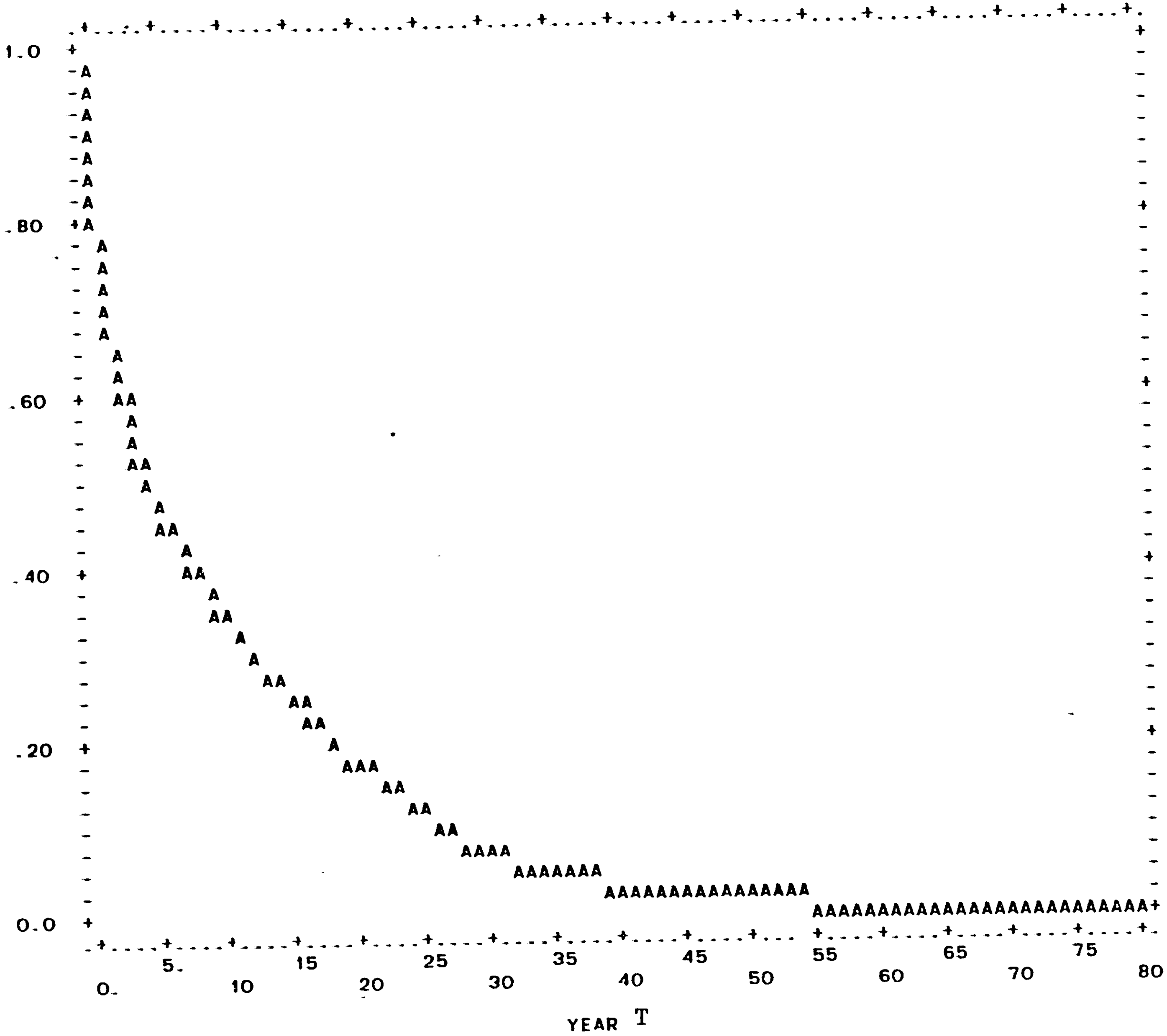
FIGURE 6:
PROBABILITY THAT THE PIPE WILL NOT EXPERIENCE A BREAK BEFORE YEAR T SINCE THE LAST BREAK



LENGTH = 100 Ft.
PRSBRK = 17.3 (i.e. PRES
= 173

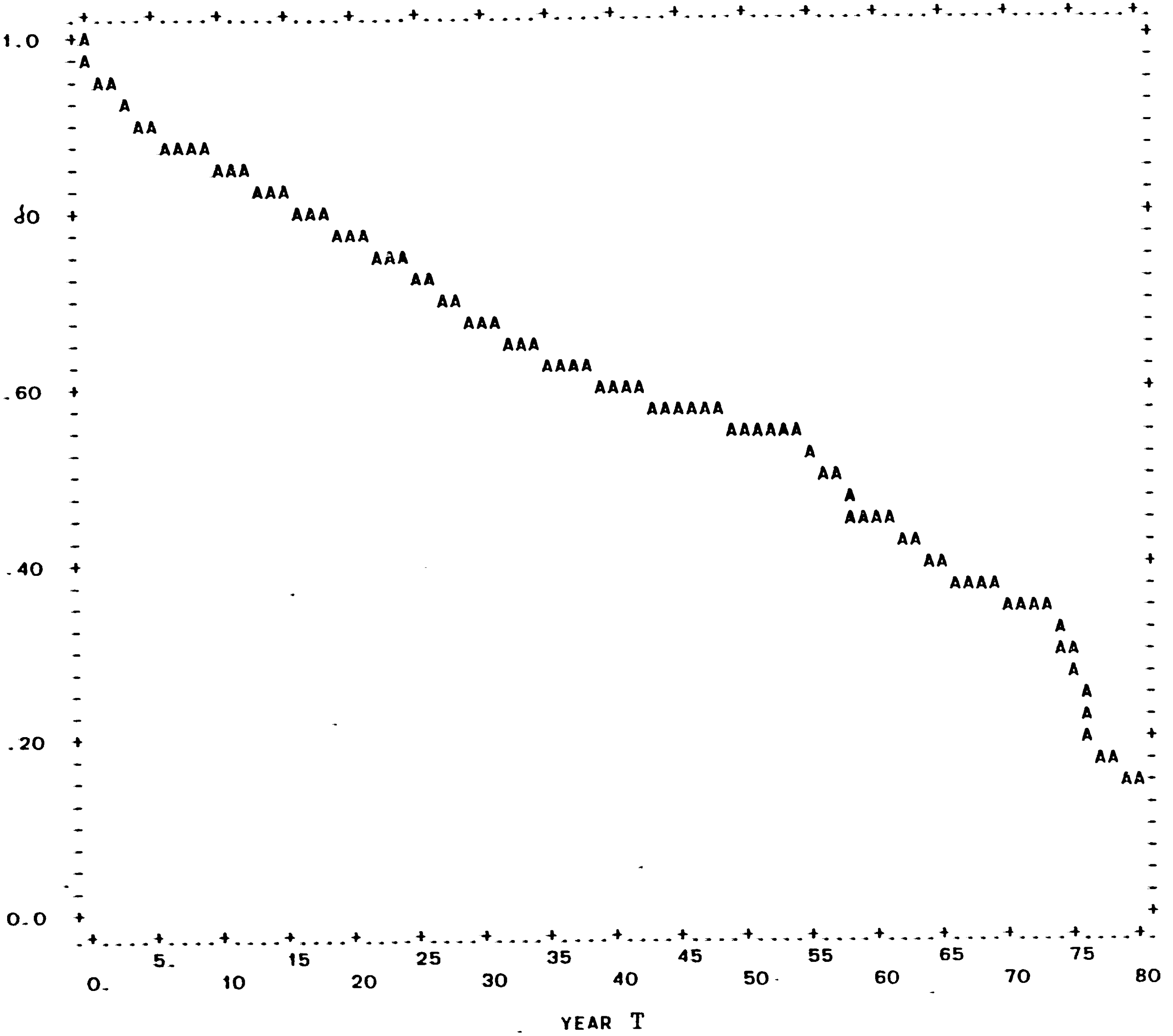
LOW = 0%
D1 = 0
D3 = 1
AGEBRK = 0
PREVBRK = 0

FIGURE 7:
PROBABILITY THAT THE PIPE WILL NOT EXPERIENCE A BREAK BEFORE YEAR T
SINCE THE LAST BREAK



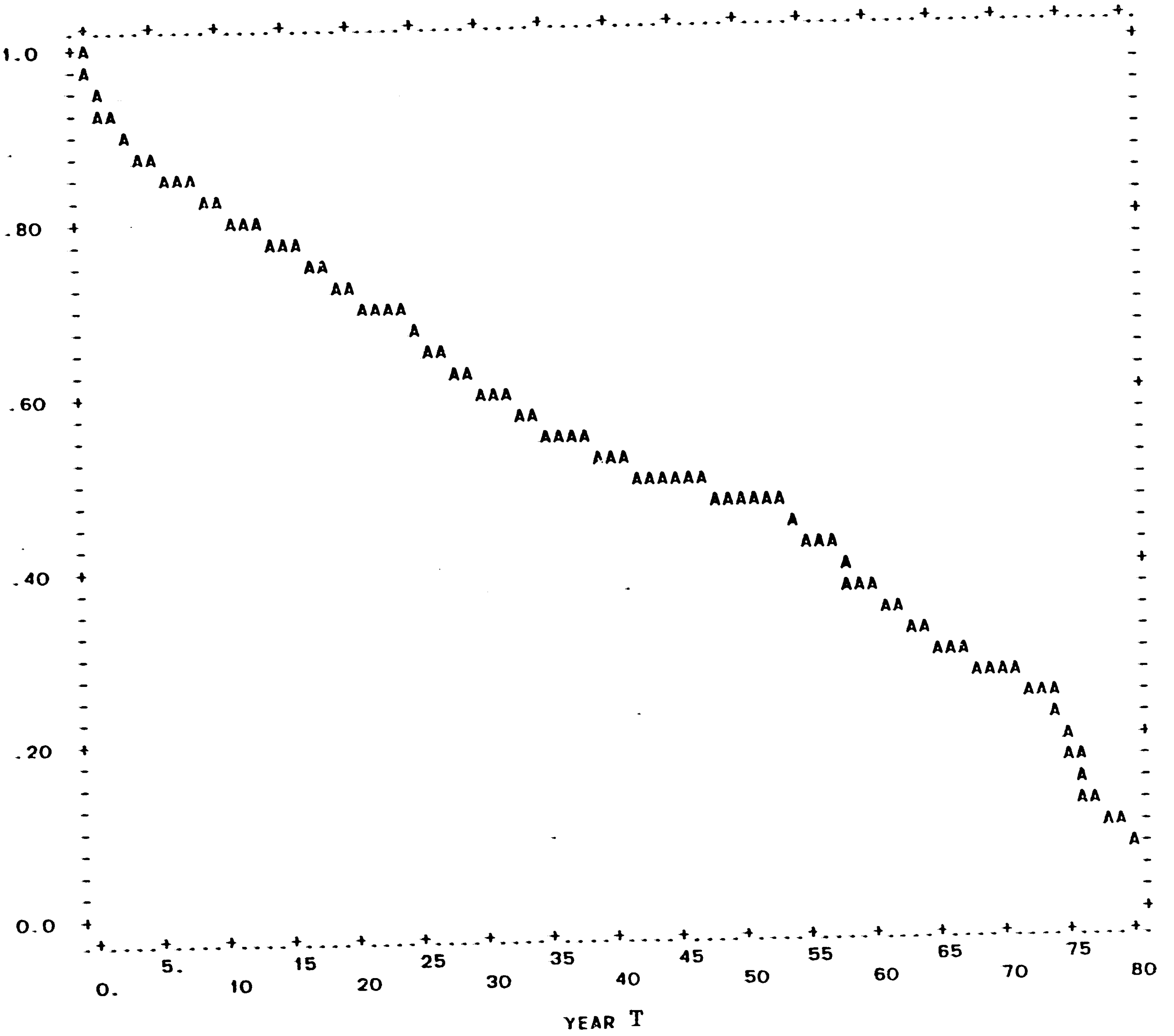
LENGTH = 14,000 ft
PRESBRK = 0
LOW = 0
D1 = 1
D3 = 0
AGEBRK = 4
PREVBRK = 2

FIGURE 8:
PROBABILITY THAT THE PIPE WILL NOT EXPERIENCE A BREAK BEFORE YEAR T SINCE THE LAST BREAK



LENGTH = 100 ft
PRESBRK = 0
LOW = 0
D1 = 0
D3 = 1
AGEBRK = 15
PREVBRK = 2

FIGURE 9:
PROBABILITY THAT THE PIPE WILL NOT EXPERIENCE A BREAK BEFORE YEAR T SINCE THE LAST BREAK



LENGTH = 100 ft
PRESBRK = 0
LOW = 0%
D1 = 0
D3 = 1
AGEBRK = 4
PREVBRK = 2

FIGURE 10:
PROBABILITY THAT THE PIPE WILL NOT EXPERIENCE A BREAK BEFORE YEAR T SINCE THE LAST BREAK

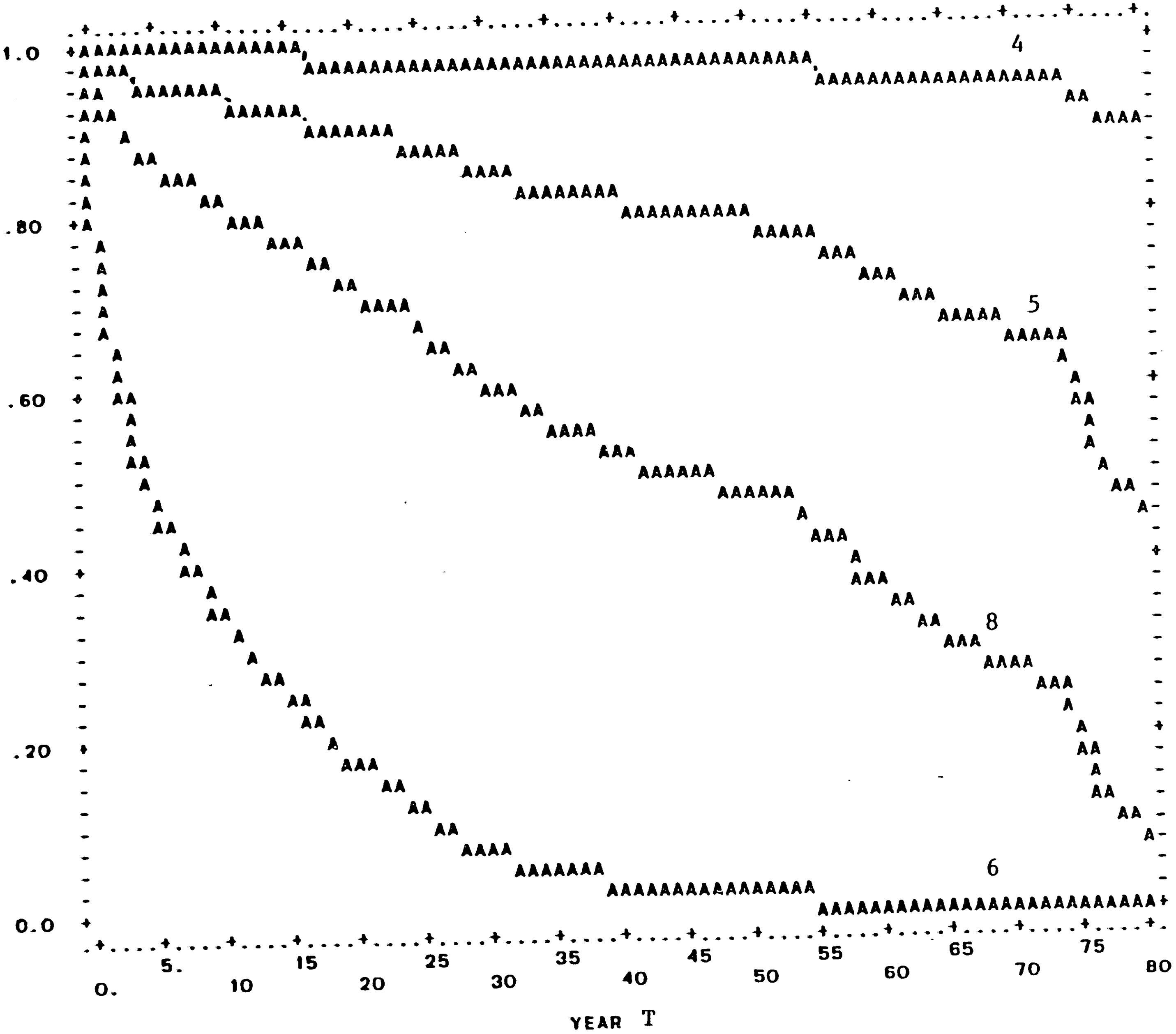


FIGURE 11:
PROBABILITY THAT THE PIPE WILL NOT EXPERIENCE A BREAK BEFORE YEAR T SINCE THE LAST BREAK

	PATTERN		
	4	5	8
LNLENGTH	4.6	4.6	4.6
PRESBRK	0	17.3	0
LOW	1	0	0
D1	1	0	0
D3	0	1	1
AGEBRK	77	0	4
PREVBRK	2	0	2

development and which experienced the second break 15 years after installation. Figure 10 shows the survival curve of a 100 ft pipe with 2 previous breaks, totally covered with land development installed after 1950 and which experienced the second break 4 years after installation. Comparing figures 6, 9, and 10, we observe how the age of the pipe at the second break and the installation period of the pipe can affect the probability of survival for a pipe which has already experienced 2 breaks. That is, the decrease of the probability of survival is not so dramatic in cases where other factors, such as age at the second break and installation period are very favorable.

Figure 11 shows the combined survival curves corresponding to figures 6, 7, and 8, respectively. The very different survival patterns that can exist in a distribution system are revealed in this figure.

Seasonality Pattern in Break Occurrences

The preliminary analysis of the seasonal variation of breaks for the New Haven data set including all major breaks since 1900 indicated a weak seasonal pattern with slightly higher break rate during the winter months (Table 6). It must be noted that the original data set did not include most of the 6 inch pipes in the system. It might be the case that the seasonality which is observed in many systems is mostly related to failure of small diameter mains.

A more detailed statistical analysis was performed using the data obtained from the Regional Water Authority of Central Connecticut after our September meeting. This data set included information covering the period from 1972 to 1984 on the break type, break date, location, size, and date of installation. No correspondence with the original data set existed because the link numbers, as defined in the original data set, were not available and also the 6" pipes had been included in the new data.

The results of the analysis are shown in table 7. Four major categories of main failures were considered: ring-crack, split-pipe, hole in pipe, and joint leak. A clear seasonal pattern was observable for "ring-crack" breaks which demonstrated higher break rates during the colder months of the year. No such pattern exists for the other three types of breaks. Since it is generally believed (personal communication with the Regional Water Authority of Central Connecticut) that ring-cracks are associated with the previous work of sewer contractors, it could be the case that soil disturbances caused by excavations for sewers creates favorable condition for a weather induced structural failure.

Break Type Analysis

Table 8 shows the number of breaks associated with each break type and diameter size. About 56% of the breaks in 6" and 8" pipes were of the ring-crack type while about 25% were of the split-type. Holes in pipes are much more frequent in small diameter pipes as well. Ring-cracks and holes in pipes were practically absent in pipes with

TABLE 6: SEASONALITY OF BREAK OCCURENCES

	Jan	Feb	Mar	Apr	May	June	July	Aug	Sept	Oct	Nov	Dec
Percentage of breaks	11%	8%	5%	6%	8%	8%	9%	8%	10%	7%	9%	9%

SEASON	Percentage of breaks %
Winter (Dec., Jan., Feb.,)	28%
Spring (Mar., Apr., May)	20%
Summer (June, July, Aug.)	25%
Fall (Sept., Oct., Nov.)	26%

TABLE 7: CORRELATION BETWEEN BREAK TYPE AND SEASON OF OCCURENCE

Month / Break Type	JAN.	FEB.	MAR.	APR.	MAY	JUNE	JULY	AUG.	SEPT.	OCT.	NOV.	DEC.
RING CRACK	65	33	12	10	6	8	7	8	8	24	28	38
SPLIT PIPE (all types of split)	14	18	11	13	12	12	14	18	19	12	14	9
HOLE IN PIPE	7	5	4	4	3	5	7	7	1	4	8	3
JOINT LEAK	10	10	3	0	9	7	3	7	3	7	4	3

TABLE 8: RELATION OF BREAK TYPE AND DIAMETER

Diameter / Type of Break	6"	8"	10"	12"	16"	20"	24"	30"	36"	48"
RING CRACK	95	123	6	9	-	-	-	-	-	-
SPLIT PIPE (all types of split)	41	57	7	33	7	3	1	1	-	1
HOLE IN PIPE	26	14	-	4	-	-	-	-	-	-
JOINT LEAK	8	26	2	14	9	-	1	2	1	1

diameters of 16" and larger. Large diameter pipes (16" and above) primarily experience splits and joint leaks. Thus only split pipe failures and joint leaks can be associated with every pipe size.

A split in a pipe can be caused by external loads transfered on the pipe from traffic, frost, soil cover etc., improper bedding conditions and improper handling of the pipe during installation. Clearly such conditions can be present in any pipe size. Joint leaks are also observed in any pipe size since joint materials are similar for every diameter. The fact that holes in the wall were found to be associated with only smaller diameter pipes (12" and below) could be explained by the fact that the pipe wall is also thinner for smaller size pipes. Since corrosion is the cause of those holes, it could be argued that smaller pipes will be affected faster by corrosion than will larger pipes. It has also been observed the tuberculous produced from corrosion of the interior pipe wall are more frequent in smaller diameter pipes thus indicating a faster progression of corrosion in those pipes. Ring-cracks are associated only with small diameter pipes and are clearly related to weather conditions. It could be argued that winter temperatures will basically affect pipes of 12" diameter and below, and will lead to ring crack failure.

Implications of Left Censoring

In order to investigate the effect of missing break records before a particular time in the past on the final configuration of this predictive model, the following experiment was performed on the New Haven data set. Break records were assumed to exist only after the 1960 period and pipes which had broken before 1960 were considered as having no breaks until 1960. The lack of records before a certain date is very common in the data records of many water utilties. The results of the application of Cox's Regression on left-censored data from New Haven was compared to the results from the application of the same model to the data which was uncensored from the left to demonstrate the sensitivity of the model to missing records on the system. The results of this analysis are shown in Table 9.

A general observation can be made that the variables found to be statistically significant in this analysis were also found to be statistically significant when data which were uncensored from the left were used. Most of the coefficients of the independent variables are not very different from those obtained with uncensored data. The only important discrepancy between the two models concerned the relative importance of the pipe installation periods. In the model where the data were censored from the left, pipes installed after 1960 were more prone to breaks than pipes installed before 1960. In the model where uncensored data were used, the same relation was found but with considerably less impact on the probability of failure. It is obvious that such results should have been expected; considering that pipes had no breaks before 1960, we Table 9

TABLE 9

PREDICTIVE MODEL USING DATA CENSORED FROM THE LEFT

(Break Records Only After 1960)

```
INDEPENDENT VARIABLES
  32 LNLENGTH    31 PRESBRK     11 LOW          23 D1          25 D3
  33 AGEBRK      20 PREVBRK

  LOG LIKELIHOOD =   -1557.3961
GLOBAL CHI-SQUARE =     239.98   D.F.=    7  P-VALUE =0.0000

                                     STANDARD
   VARIABLE       COEFFICIENT         ERROR     COEFF./S.E.    EXP(COEFF.)
   --------       -----------       --------    -----------    -----------
32 LNLENGTH            0.6323         0.0870         7.2716         1.8820
31 PRESBRK             0.0789         0.0291         2.7138         1.0821
11 LOW                -0.3744         0.1512        -2.4768         0.6877
23 D1                 -0.5274         0.1800        -2.9305         0.5902
25 D3                  1.3058         0.1982         6.5884         3.6908
33 AGEBRK             -0.0271         0.0171        -1.5831         0.9733
20 PREVBRK             1.2490         0.2336         5.3469         3.4869
```

PREDICTIVE MODEL USING UNCENSORED FROM THE LEFT DATA

```
INDEPENDENT VARIABLES
  32 LNLENGTH    31 PRESBRK     11 LOW          23 D1          25 D3
  33 AGEBRK      20 PREVBRK

  LOG LIKELIHOOD =   -2805.0324
GLOBAL CHI-SQUARE =     279.27   D.F.=    7  P-VALUE =0.0000

                                     STANDARD
   VARIABLE       COEFFICIENT         ERROR     COEFF./S.E.    EXP(COEFF.)
   --------       -----------       --------    -----------    -----------
32 LNLENGTH            0.5364         0.0665         8.0694         1.709
31 PRESBRK             0.0663         0.0234         2.8346         1.068
11 LOW                -0.5528         0.1218        -4.5383         0.575
23 D1                 -0.6798         0.1207        -5.6313         0.506
25 D3                  0.2708         0.1408         1.9239         1.31
33 AGEBRK             -0.0216         0.0089        -2.4334         0.97
20 PREVBRK             1.1858         0.1830         6.4799         3.27
```

implicitly penalize the pipes that were installed after 1960. However, this preliminary analysis shows that the model is fairly robust when applied to data censored from the left.

CHAPTER FOUR

THE APPLICATION OF THE MODEL RESULTS IN DECISION MAKING

There are several ways in which managers of water distribution systems could use the results from this model. The first way the analysis is useful is in establishing the general trends of the system. Such trends can be used to establish guidelines for making maintenance decisions. As an example, these analyses have shown that the probability of a failure in a pipe is highly correlated to the number of breaks which have already occured in that specific pipe. In other words, a pipe which has already experienced several breaks can be expected to break with increased frequency as time progresses. This trend could be used to suggest a guideline that when planning to rehabilitate a given area of the system, that pipes which had experienced more than two breaks should be replaced rather than cleaned and lined.

The data on the probability of failure of a given pipe would be useful in maintenance decision-making. The probability data could be used to identify high risk pipes and areas of the distribution system. The results showed that a small number of pipes could be identified as high risk. By arbitrarily defining a high risk pipe as one for which the probability of break is greater than 5%, 93 pipes were identified. The estimated probabilities for these pipes are shown in Appendix E. It must be pointed out that the break probabilities after ten years were obtained under the assumption of no failures during that period and also no replacement of pipes. Thus, such probabilities would be very sensitive to the replacement/repair strategies of the utility and to the actual break events that occur during the ten year period.

Appendix E contains probabilites of failure calculated using the entire length of the pipe as defined in the original data set. Also pipes with zero breaks and pipes with three or more breaks were included in the table. In order to further assist the replacement vs. repair decision a new list of the high risk pipes was generated taking into account the following considerations:

1. Pipes with zero previous breaks that were defined as high risk pipes in our model were not included as it would not make sense to replace them even though they had a high probability of failure.

2. The probabilty of a break per 100 ft. of pipe has also been calculated and shown in Table 10. The reason for doing so was that it is more convenient for replacement decisions to know the per 100 ft. probability of a break.

3. Information about land development and pipe diameter is also provided in Table 10 since those variables could be used as preliminary surrogates for determining the relative importantce of a pipe in the

distribution sysem and the potential consequences of a break.

4. Pipes with four or more breaks were not included since it is well known that their probability of failure is quite high.

Table 10 demonstrates the first step towards developing a ranking system for pipe replacement decisions through the use of a predictive model for pipe failures.

This information was coordinated using the maps obtained from the Regional Water Authority of New Haven, to classify about 60% of the high risk pipe with one or more previous breaks into groups of neighboring pipes. Thus, high risk areas were identified in the system. Those groups are shown in Table 11. The information obtained from this table could help identify further underlying causes of breaks and also implement bundling strategies for contracting projects in parts of the system.

Groups 1, 10, and 12 of Table 11 have a significantly higher number of breaks than the other groups. This observation could lead to specific measures taken for those very high risk areas in terms of future repair versus replacement decisions. Also by focusing attention to those areas, additional factors related to the occurrence of breaks might be revealed.

This ranking would encorporate information on the physical condition of the pipe only. The analysis could be taken one step further through the combination of economic

TABLE 10

PIPES WITH 1, 2 or 3 PREVIOUS BREAKS AND PROBABILITY OF BREAK > 5%

LINK NO.	PROBABILITY OF BREAK	LAND DEV.	DIAMETER	PROB. OF BREAK/100ft.
8	22%	Low	8	0.63%
24	9.8%	High	8	0.70%
47	12%	High	10	0.60%
136	5.6%	High	20	0.37%
140	6.2%	High	16	0.31%
191	9.0%	High	16	0.33%
218	26%	High	16	0.69%
290	5.3%	High	12	0.42%
425	6.3%	Low	12	0.28%
452	33%	Med.	8	1.32%
468	12%	Low	24	0.18%
509	6.3%	Low	12	0.21%
601	7.8%	High	8	0.16%
603	9.5%	Low	8	0.13%
623	7.7%	Low	12	0.31%
642	8.5%	Low	8	0.29%
644	50%	Low	8	1.72%
649	9%	Low	8	0.20%
683	12%	High	16	0.23%
684	11%	High	24	0.21%
693	7.4%	High	6	0.25%
774	7.9%	Low	24	0.23%

Continued

TABLE 10: Continued

LINK NO.	PROBABILITY OF BREAK	LAND DEV.	DIAMETER	PROF. OF BREAK/100ft.
809	8.3%	Low	12	0.14%
828	5.7%	Low	8	0.30%
833	6.0%	Low	16	0.13%
843	8.1%	Low	16	0.10%
861	13%	Low	12	0.17%
876	6.9%	Low	16	0.29%
878	6.6%	Low	12	0.33%
904	8.2%	Low	12	0.22%
940	5.9%	Low	10	0.13%
948	11%	Low	12	0.16%
978	10%	High	8	0.67%
1013	8.2%	Med.	8	0.14%
1020	17.5%	Low	8	0.58%
1021	11%	High	12	0.18%
1027	6.2%	Low	12	0.20%
1044	9.9%	Med.	48	0.07%
1069	7.2%	High	16	0.22%
1109	11%	High	12	0.18%
1156	42%	Low	12	0.74%
1162	8.8%	Low	12	0.20%
1187	10%	High	8	0.40%
1209	9.2%	Low	12	0.31%
1252	7.6%	Low	12	0.26%
1259	7.3%	Low	10	0.30%
1269	7.8%	Low	12	0.43%
1274	18.6%	Low	12	0.37%
1280	8.7%	Low	20	0.19%
1286	8.8%	Low	16	0.25%
1344	12.7%	Low	12	0.35%
1358	8%	Low	12	0.20%
1373	11.2%	Low	8	0.16%
1385	9.9%	Low	12	0.13%

TABLE 11

GROUP NUMBER	LINK NUMBER OF PIPES IN THE GROUP	INFORMATION ABOUT THE GEOGRAPHICAL LOCATION	TOTAL NUMBER OF PREVIOUS BREAKS FOR PIPES IN THAT GROUP
1	1069, 1026, 1078, 1080, 1087, 1023, 21, 1107, 1100, 1145, 1109, 38, 63	Close to the Shoreline of the Quinnipiak River	35
2	642, 644, 649, 468	Orange	6
3	1013, 1020, 1021	North Haven, Sackett Point Rd.	6
4	1280, 1286	North Branford	2
5	1156, 1162	East Haven	4
6	876, 878, 904	North Haven, Ridge Road	3
7	1259, 1269	Branford	3
8	107, 108, 1027 290	New Haven (Longwharf, Brewery, Spring, Union)	10
9	865, 843, 861	Hamden (Shepard Ave., Still Hill Rd.	6
10	723, 729, 726	New Haven Litchfield Ave., Whallay Ave.	21
11	683, 684	West Haven (Forest St.)	6
12	1132, 1133	New Haven (Cosey Beach Ave.)	24
13	136, 140, 191	New Haven (Whitney, Church, Elm, Prospect, Canal)	5

information on the cost of replacement and repair with probability of failure data. In such a way, one could develop an equation to predict the optimal time for replacement by combining economic priorities with the probability of failure. The optimal time calculation could be used to develop a more comprehensive list of the "critical" pipes in the system and to rank them for replacement priorities.

The optimal time analysis was done for New Haven using a method similar to that used by Shamir & Howard (1979). The hazard function obtained from Cox's Regression was used to determine the probability of failure per year for a given pipe link.

$$H(t) = H\ (t)\ \exp(BZ)$$

$$\text{where } H\ (t) = 0.000210476 - 0.000011171t + 0.000000199t^2$$

Estimates of costs of repair and of replacement were taken from Walski et.al. (1982) as actual figures were not available on the New Haven System. The optimal time for replacement was found by minimizing the future maintenance costs discounted to present value.

$$\min \int_{t_p}^{t_r} C(b) * H(t)\ e^{-r(t_r-t_p)}\,dt \quad + \quad C(r)\ e^{-r(t_r-t_p)}$$

where C(b) = cost of repair
C(r) = cost of replacement

This equation was solved by differentiation, then upon setting the differential to zero and substituting for H (t), the following solution was found.

$$t^* = 28 + \sqrt{\frac{5{,}025{,}126 * rC(r)}{C(b) * e^{\underline{x}}} - 270}$$

This equation was used to calculate the optimal time for pipes with less than four breaks in the New Haven data set. A list of "Critical" pipes was created which included the pipes that had optimal times for replacement within 25 years of the present.(Table 12) This list encorporates the economics of the maintenance alternatives in the development of maintenance priorities. Files of this nature could serve as excellant organizational tools for preparing budgets and maintenance forcasts for the system.

The results from the regression analysis which suggest the probability of a break and the results from the optimal time analysis which encorporate economic factors are particularly suited for aiding managers in replacement decisions. Such analyses are useful for exposing pipes which may be just reaching a critical point in their history where frequent breaks can be expected. The early destinction of such pipes leaves the manager with additional time before this critical level is reached during which the manager can look for the best time for the replacement and thereby can be more likely to coordinate with the other utilites and the city on the repaving costs.

The optimal time trends could also be used to aid in budget forcasting. For example, the New Haven list shows relatively few pipes in urgent need of replacement at the present date. Yet looking just five to ten years in the

TABLE 12: OPTIMAL TIME ANALYSIS

LINK	DIA	LEN	TYPE	BRK	DATE	AGE84	PRES	H(t)	OPT TIM
218	16	3750	1	3	16	50	0	0.293035	-18
47	10	2000	1	2	0	64	0	0.127046	-7
1021	12	6200	1	1	6	65	0	0.115785	-5
693	6	3000	1	1	0	64	0	0.07491	-3
1013	8	6000	1	2	40	36	0	0.083049	-1
683	16	5300	1	3	21	34	0	0.1252	-1
684	24	5300	1	3	21	32	0	0.112062	0
191	16	2750	1	3	25	35	0	0.092046	2
1338	16	600	1	0	0	84	11	0.073919	9
140	16	2000	1	1	0	65	0	0.063065	11
1209	12	3000	1	2	57	22	0	0.09645	11
978	8	1500	1	3	50	23	0	0.107859	12
1349	16	1250	1	0	8	76	8	0.065741	12
1391	10	1600	1	0	16	68	16	0.054237	13
1375	8	4250	1	0	36	48	10	0.037859	14
290	12	1250	1	1	0	67	0	0.053548	14
272	10	1000	1	3	31	38	0	0.038273	14
1123	12	5500	1	0	25	59	10	0.046916	15
1109	12	6000	1	3	28	17	0	0.124166	16
1152	8	5250	1	0	30	54	12	0.035791	17
1335	8	1250	1	2	41	30	0	0.034389	17
669	20	7000	1	1	30	50	0	0.046487	17
1329	12	8000	1	0	30	54	11	0.041733	17
898	8	2750	1	0	40	44	13	0.029833	17
136	20	1500	1	1	0	66	0	0.056514	17
787	16	2000	1	2	0	48	0	0.038187	18
1382	8	4500	1	1	50	28	0	0.035977	18
1020	8	3000	1	3	46	21	0	0.192861	19
653	12	4800	1	1	40	34	0	0.031263	19
604	8	7000	1	0	52	32	9	0.025983	20
886	16	4300	1	0	40	44	14	0.031464	21
1390	10	1000	1	2	16	45	0	0.029225	21
655	8	4000	1	0	40	44	8	0.025044	21
922	10	800	1	1	10	66	0	0.040323	21
654	8	5400	1	0	40	44	7	0.025759	21
623	12	2500	1	2	50	19	0	0.080575	21
1151	8	5500	1	0	30	54	8	0.028556	22

TABLE 12: OPTIMAL TIME ANALYSIS
CONTINUED

LINK	DIA	LEN	TYPE	BRK	DATE	AGE84	PRES	H(t)	OPT TIM
888	8	6250	1	0	40	44	10	0.023854	22
351	10	3250	1	0	22	62	8	0.036151	23
35	8	2500	1	0	21	63	8	0.03249	23
1180	8	5000	1	0	30	54	9	0.028227	23
974	8	4500	1	0	50	34	8	0.020785	24
624	12	4000	1	0	50	34	11	0.024581	24
1126	10	1100	1	0	25	59	10	0.033465	24
445	8	5500	1	0	30	54	7	0.026732	24
1121	12	3500	1	0	25	59	9	0.034228	24
1269	12	1800	1	2	56	17	0	0.082194	24
1370	10	600	1	0	8	76	7	0.041495	25
1176	8	2750	1	0	40	44	8	0.021447	25
944	12	3400	1	0	40	44	9	0.025339	25
781	24	1600	1	0	0	84	7	0.05604	25
350	10	3500	1	0	25	59	8	0.032468	25
236	12	600	1	1	0	69	0	0.039335	25

future one sees that the number of critical pipes per year seem to grow almost exponentially. Early awareness of this problem could allow management to plan ahead and possibly replace pipes earlier than the "critical date"at the present thereby maintaining a feasible number of replacements per year.

It must be stressed that pipes should not be replaced solely on the basis of their ranking by the optimal time analysis or by the probability of failure ranking. These predictions were intended to be used as an aid to support the decision making process. There are many factors outside the realm of such models which must be coordinated in such a decision process.

In order to determine which other factors might be important in determining the best maintenance plan, several water utility managers were interviewed (John Sullivan Jr. of Boston Water & Sewer Authority as well as several officials of the Regional Water Authority of Central Connecticut). Their suggestions on how they actually make such decisions provided us with a great deal of insight in this area.

One major area of concern to a utility is the balancing of the limited budget. Every potential for reduction in costs is explored. One of the areas of greatest potential for cost reduction is the cost due to repaving. By coordinating with the other utilities and road work projects water utiltiies can practically eliminate the expence involved in repaving roads. Repaving costs represent a substantial amount of the overall costs for replacement projects. Many pipes which are not yet "critical" based on our "critical pipe" list may be replaced or rehabilitated when repaving costs can be shared by other work in the area. Also, if work planned by the other utilities is to be unusually close to old water mains and it seems likely that the bedding of the pipe might be disturbed, the old pipe might be replaced at that time to avoid replacement in a recently repaved area. This model which allows early identification of the high risk pipes, would aid in the coordination between utilities by allowing more advanced notice of the problem.

Another important factor in determining actual maintenance schedules is the priority of a given pipe in the system. Priority ratings can be divided into two categories; hydraulic priorities and location priorities. Hydraulic priorities are determined by the position of the pipe in the distribution network and the extent of redundancies in the area. An example of a pipe of high hydraulic priority would be a major transmission line which if broken would take down the service to a large area of the system. Another type of priority is based on the physical location of the pipe. For example, pipes in downtown or highly populated areas receive much higher priority than pipes in the more rural districts. High priority pipes of both types should be replaced before they reach a "critical" level as the costs associated with failure might be quite extensive (especially as compared to the average pipe repair costs). A priority ranking can easily be added into this analysis by assigning relative priority factors to the pipe links in the system.

Rehabilitation is generally done on entire areas of the system at one time. The decision to rehabilitate an area is based on capacity requirements, especially fire flow requirements. When the flow in a given area is no longer sufficient to meet these fire flow requirements, rehabilitation through cleaning and cement lining of the pipes is recommended. Water quality problems associated with the tuberculation also create the need for rehabilitation of a given area. The results of our analysis would be useful in determining which pipes in a rehabilitation project ought to be replaced rather than rehabilitated. In this way, the utility could avoid the extra expense of rehabilitating a pipe which would need to be replaced in the next five or ten years.

Replacement decisions are usually based on high break rates, the thickness of the wall as determined by inspection of the main, the liability and costs associated with past break events, and inconvienence to the customers. Replacement is more commonly addressed on a street by street or pipe by pipe basis rather than on an areal basis as for rehabilitation. Priorities of particular pipes or particular areas of the system are usually determined from past cost information on the cost of damages and on the importance of the pipe in maintaining flow to customers.

LIMITATIONS TO THE APPLICATION OF RESULTS:

One major limitation in the application of the model proposed in this research is the necessity for accurate and fairly extensive historical records. As extensive data is rarely available on water distribution systems, it is our goal to see how well this model can work for systems with less extensive data.

Several problems were evident in the New Haven data set that were not related to the lack of data but to carelessness in the development of the data set. These problems could be avoided through more precise work when putting the data set together. Although the New Haven data set was ideal for developing this sort of analysis due to the length of the historical records, the data set was not a good representation of the actual New Haven System. Thus, the results of this model are of limited use for the New Haven Water Authority. These limitations are discussed in this section in order to suggest better methods of data collection for other systems before attempting this sort of methodology. It should be stressed that these limitations are only in the application of our results to the New Haven System and not to the methodology itself.

One of the most critically limiting factors to the use of these results was the way in which the pipe links were initially defined. The data set was collected by EPA and the

MIT research team had little access to the original data. It was only through a meeting with the Regional Water Authority of Central Connecticut that it was found that there was no obvious correlation between the pipe links listed in the data set and the maps of the actual distribution system. It seems as though the link numbers were arbitrarily defined and that the correlation to the actual system was not recorded. As a result, it was not possible to graphically display and analyse the network to determine the high priority areas. The method for the definition of pipe length also presents problems for our analysis. There are several pipes in the data set with lengths of over 14000 feet. These pipe lengths clearly do not correspond to single pipe units but rather must be related to single units from the hydraulic analyses (perhaps the length of a pipe with a constant diameter in a given direction). More careful thought should be given to the use of a constant method for defining pipes in the system and to coordinating this definition with the maps and other data as they become available. If the links in the New Haven data set were correlated to the actual pipes of New Haven, the results of our analysis could have been more useful to the managers of that utility.

Another significant limitation to the data set is that pipes which have been replaced or abandoned remain in the data set as intact and functional pipes. As the link numbers in the data set do not correspond to the maps and data available from the Regional Water Authority, there is no clear way to determine which pipes may have been replaced.

Consequently, it is possible that the pipes which appear in the model as the pipes with the highest probability of breaking may no longer be in use. This factor again limits the usefulness of the predictions from this model based on this data set to the managers of the New Haven distribution system. In a more realistic situation, this problem would not be likely to occur as the majority of such data bases are continually updated by the utility.

Another limitation to the New Haven data was the lack of information describing the type of breaks. It has been suggested that the break type is useful in determining a probable cause for the break. It is likely that many of the breaks are due to external factors such as the excavation by sewer contractors This sort of factor does not fit into our model. If this were true, these breaks should be removed from the regression analysis to yield more accurate predictions.

There are several categories into which break types can be divided: ring cracks, horizontal and vertical fractures, holes and joint failures. Data on the break types from 1972 on was aquired from the Regional Water Authority of Central Connecticut. This data showed that 40% of the breaks were ring cracks. Ring cracks are thought to be associated with excavation in the area which disturbs the bedding of the pipe immediatly prior to the break event. Such data infers that perhaps 40% of the breaks in our data set should be excluded from the regression analysis in developing the probabilistic model. Although such data is currently available on break types, it can not be correlated with the EPA data set due to the inconsistency between the link numbers and the pipes in the actual system. Again, this factor may limit the validity of the predictions of this model for New Haven. One potential way to alleviate this deficiency in other data sets might be to look at the repair records kept by sewer contractors to see if the records coordinate. If the records are kept with similar identification systems ie) name of street or intersections, a data file of the sewer repair records could be compared to the file of the water main breaks. In such a way, it might be possible to isolate the breaks caused by this external variable.

The data set was also imprecise regarding the actual definition of a break. Breaks have been assumed to be defined as "repair events". Further destinction of the time of year of a break, the size or type of a break would have proven to be quite useful in the development of the model. However, these limitations will be found in most of the historical records on most water distribution systems.

Another limitation was evident in the definition of several other variables: pipe type, land development, and pressure. The type of pipe was listed as either class 1 or class 2. More in depth classification of pipes by the material such as pit cast versus sandspun or ductile iron and of the initial wall thickness would have been useful. The variables reflecting land development and pressure do not reflect their change over time. The land development data was probably taken from a single time record such as the 1972 census report. Such a value does not reflect the changes in population, traffic, and land development which have occured over time for the system. During such times when the change in pressure is raised intentionally, a number of breaks are often experienced. The pressure variable listed in the data set does not reflect these extreme variations. The record of these changes in pressure which are related to a break event would have been useful. Yet as such intentional changes in pressure are related to decisions made by the water authority, such breaks would not be able to be forcasted. The actual pressure at the time of a break would be most useful in eliminating such pipes from the regression analysis and in the development of another predictive model to determine which pipes would be most likely to break given a particular change in the internal pressure.

It must be stressed again that these difficulties do not destract from the general validity of the model but rather merely restrict the application of the results in New Haven. It seems as though most of these problems would be avoided if the data set were developed by people working within the water authority with full access to maps and records and by careful and precise documentation of how the data set was formulated.

CHAPTER FIVE

MODEL SUMMARY AND CONCLUSION

The majority of water distribution systems in the eastern United States have reached a critical age as reflected by their increased breakage rates and maintenance expenses. It has become essential that the management of such utilities have sufficient information available to identify the pipes and areas of a system in most desparate need of maintenance. It is our contention that management decisions as to the proper selection and timing of such maintenance alternatives should be based on up to date information on the status of the system, the economics of the alternatives and the priority of the maintenance action within the framework of the system.

Currently, management must make such decisions from either emperical rules for maintenance of the system or as required by the immediate problems faced by the utility. In general, management officials are not provided with convenient access to information on the specific pipe or area of the water distribution system.

This report presents a model which would supply this information to management in a format useful for decision making. The model was designed to function as a decision support tool for water utility management. It provides the manager with easy access to current information on the status of the distribution system, prioritized maintenance lists, as well as information on the future failure probability for each pipe.

New Haven, Connecticut was used as a case study for the development of the model as well in the discussion on the application of the model. A data set on the water distribution system of New Haven was used to develop the statistical and regressional methodology presented in this report. The data set consisted of 1391 pipe links with the following descriptive covariates: diameter, length, pressure, pipe type, soil corrosivity, soil stability, land development, swamp, date of installation, number of previous breaks, and the time to the first repair.

The model combines in depth statistical analysis with the regression analysis developed by Cox in 1972. The statistical analysis was used to determine the appropriate variables for use in the regression analysis and was broken into three distinct steps. The first step provided descriptive statistics such as the range and variability of the parameters in the data base. This step helped to define any apparent trends in the data. The next step involved bivariate analysis and the calculation of cross correlation coefficients. Bivariate analysis was used to determine whether there was any correlation between pipe failures and the other variables in the data set. Due to the highly skewed distribution of the data, three types of correlation analysis were performed. The final step to these statistical procedures was survival analysis. Survival analysis was performed to reveal the survival pattern for various categories of pipes. Several stratification groups were analysed for correlation to failure.

The results of these statistical techniques demonstrated that there was a clear exponential increase in main failures as the number of previous breaks increased. This pattern strongly suggested that a proportional hazard model where variables act multiplicatively on the hazard rate would be appropriate for deriving a predictive model for pipe failure. As a result, the regression analysis which was developed by Cox for this exponential failure relation was adopted.

The Cox regression model is a general failure prediction model which was applied to New Haven pipe data to calculate the probability of a pipe experiencing a break or failure at any given point in time as a function of certain break causing factors. The purpose of this regression analysis was to define the variables affecting the probability of failure of a particular pipe and to establish the relationship between them.

The results of the application of the Cox regression model to the New Haven data set can be summarized as follows:

- Increased pressure in a pipe is a good indicator of an increased hazard rate if the pipe does not have a history of previous breaks.
- High land development is an important factor which raises the probability of breaks in a pipe. Land development can be seen as a surrogate factor for external loads transmitted to pipes from buildings and trucks etc. as well as from the increased frost penetration in such areas.
- A low age at the time of second break indicates a high probability of further breaks.
- Installation period was found to be very important in the determination of the probability of failure of a pipe. The fact that pipes installed during the 1930-35 period performed much better while the pipes installed after 1950 performed worse was explainable by the pipe material and construction practices used in the various time periods.
- Each previous break in a pipe triples the probability of the next break.
- Longer pipes tend to have fewer breaks per foot of pipe than shorter pipes.
- The hazard rate for breaks first decreases with the age of the pipe until approximately age 30, then it begins to rise sharply.

Other Observations:

- Seasonal breaks are associated with small diameter pipes (8" and below). No seasonality in breaks exists for break types other than ring cracks.
- Ring-cracks are experienced primarily by small diameter pipes which show an apparent increase during the winter.
- Holes in pipes are associated with small diameter pipes (12" and below).
- Large diameter pipes (16" and above) experience only split type breaks and joint leaks.

The model presented in this report has several advantages in comparison to other regression based models. This model derrives the probability of experiencing a break in each specific pipe rather than the expected number of breaks in a given year. The model is thus particularly well suited to the prediction of break probabilities for systems with relatively infrequent break events. Large mains tend to experience fewer breaks yet these breaks are of more extreme consequences. It appears that the prediction of these more consequential breaks is of greater interest to management. This model is designed to function on such pipes with relatively low probabilities of failure while several of the other regression based models would not be appropiate.

The Cox regression model is well suited to the prediction of failure probabilities from large data sets on water distribution systems. This method provides the ability to work effectively with censored data sets so that information can be obtained from the large part of data sets which do not yet have a failure history. The model provides the ability to effectively stratify the data thereby exposing trends related to different age groups of pipes, different pipe characteristics and different environmental factors. The model derives a hazard rate as the probability of a break at a particular time given that the pipe has survived up to that time since its last break. This probability calculation is much more useful in the estimation of reliability at the network level than the previous attempts at estimating the expected number of breaks.

The steps in this analysis can be helpful to the water distribution manager in several ways. The first step in the analysis, which provides the statistical description of the system, is useful in establishing general guidelines for maintenance proceedures for the system. The probability results from the regression analysis can be directly used to prioritize maintenance actions. These results could also be combined with economic information on the costs associated with the different maintenance alternatives so that an optimal time can be calculated. This optimal time could also be used to rank the pipes for maintenance priorities. The optimal time ranking would encorporate the economics of the alternatives with the probability information. From such information, a table could easily be prepared to show the probability of failure, general characteristics, optimal replacement time and the economics of the alternatives for each pipe in the system. This sort of information would serve as an excellant organizational tool for preparing maintenance plans and budgets. The early identification of the high risk pipes in a system allows the management more time to plan for maintenance in order to make more efficient contracts and to coordinate with other utilities on repaving costs. This sort of information could also be used to aid in the identification of high risk pipes in need of replacement within areas proposed for rehabilitation.

It must be stressed that maintenance decisions should not be made solely on an optimal time ranking or on a probability ranking. Such lists were only intended to be used as an aid to support the decision making process. Immediate factors such as break events, problems of insufficient fire flow, or problems associated with dirty water must clearly alter such priorities. The model was designed to be used as a tool by management to provide easy access to complete information on the status of the system as a whole as well as on each individual pipe in the system.

REFERENCES

Clark, R. M., C. L. Stafford, J. A. Goodrich, "Water Distribution Systems: A Spatial and Cost Evaluation," Journal of the ASCE, WP3, October 1982.

Cox, D. R., Regression Models and Life-tables, Royal Statistical Society, March, 1972.

Hudson, William D., "Studies of Distribution System Capacity in Seven Cities," JAWWA, pp. 157-164, February, 1966.

Kalbfleisch, J. D. and R. L., Prentice, The Statistical Analysis of Failure Time Data, John Wiley & Sons, 1980.

Karaa, F., "A Decision Support Model for the Investment Planning and Rehabilitation of Mature Water Distribution Systems", Doctoral Thesis, Massachusetts Institute of Technology, 1984.

McBean, E. A., S., Al-Nassri, and D. Clarke, "Some Probabilistic Elements of Field Testing in Water Distribution Systems," Proc. Instn. Civ. Engrs., Part 2,Vol. 75, pp.143-153, June, 1983.

Morris, R. E., Jr., "Principal Causes and Remedies of Water Main Breaks," JAWWA, pp. 782-798, July, 1966.

O'Day, D. K., "Organizing and Analyzing Leak and Break Data for Making Main Replacement Decisions," Journal AAWA, pp.589-594, November, 1982.

O'Day, D. K., R. Weiss, D. Blair, P. Cairo, "Philadelphia Water Supply Infrastructure Study, ASCE Urban Water Conference, Baltimore, MD, May 29, 1984.

Shamir, U., and C. D. Howard, "An Analytic Approach to Scheduling Pipe Replacement," Journal AAWA, pp. 248-258, May 1979.

Sullivan, J.,Jr., "Maintaining Aging Systems - Boston's Approach," Journal AAWA, pp.555-559, November, 1982.

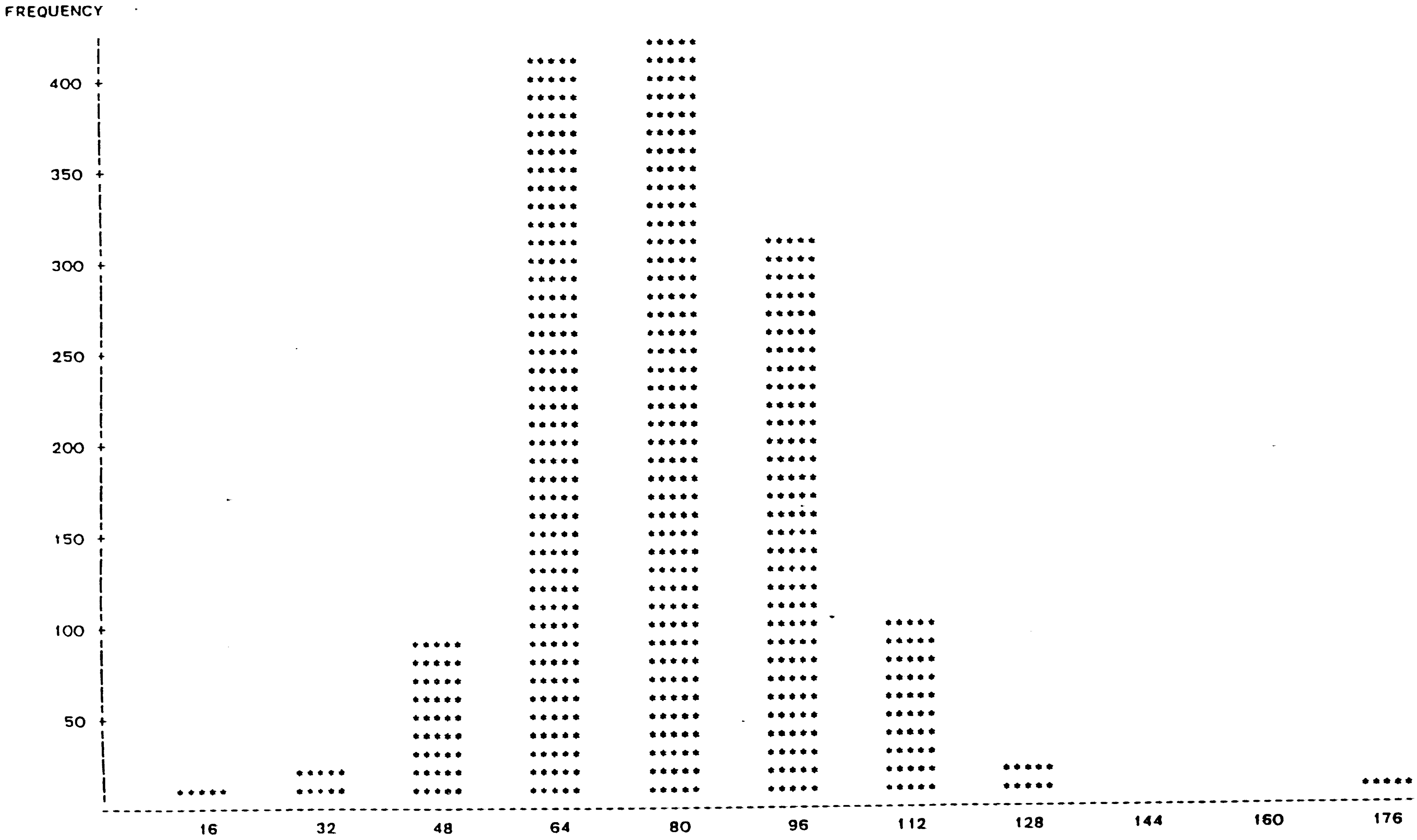
Walski, T. M., "Economic Analysis of Rehabilitation of Water Mains," Journal ASCE, WR3, pp. 296-308, October, 1982.

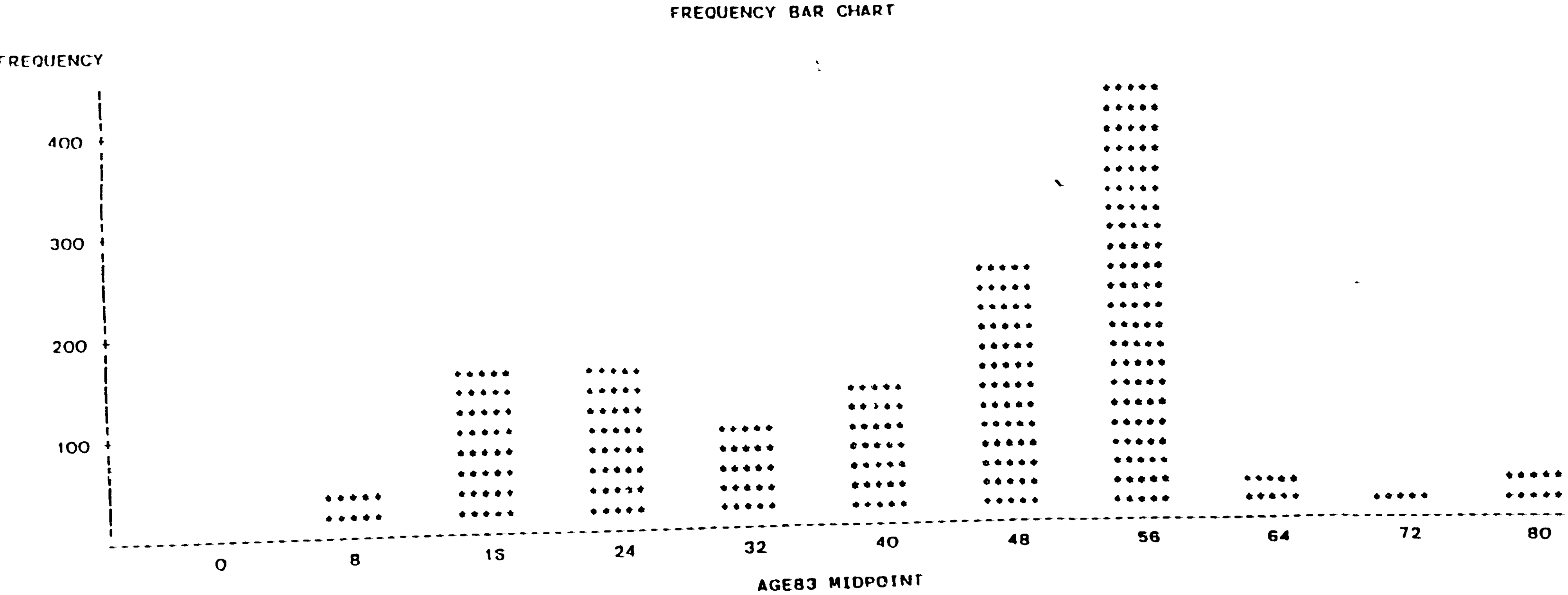
Walski, T. M., A. Pellicia, "Economic Analysis of Water Main Breaks," Journal AWWA, pp. 140-147, March, 1982.

APPENDIX A

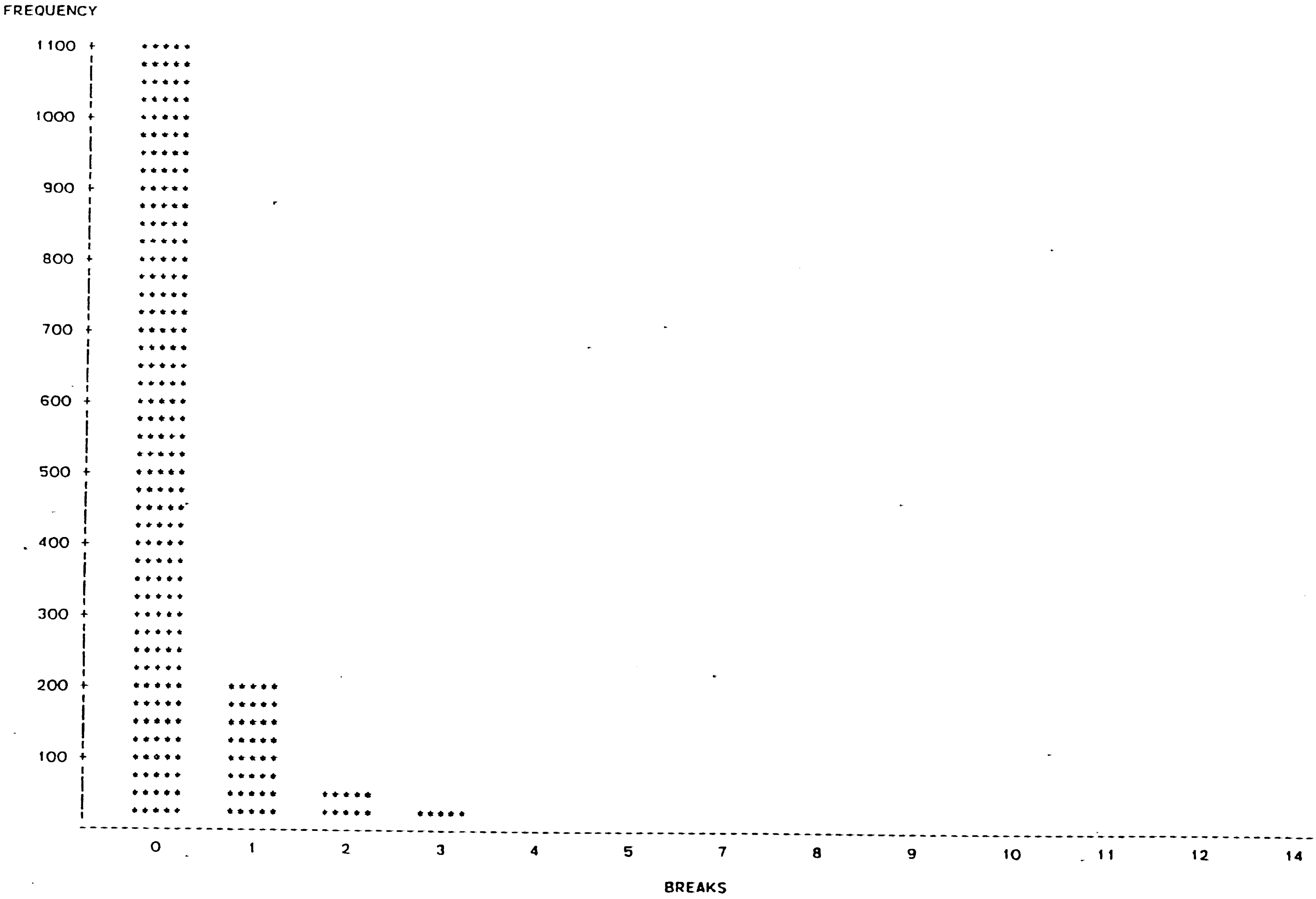
FREQUENCY HISTOGRAMS

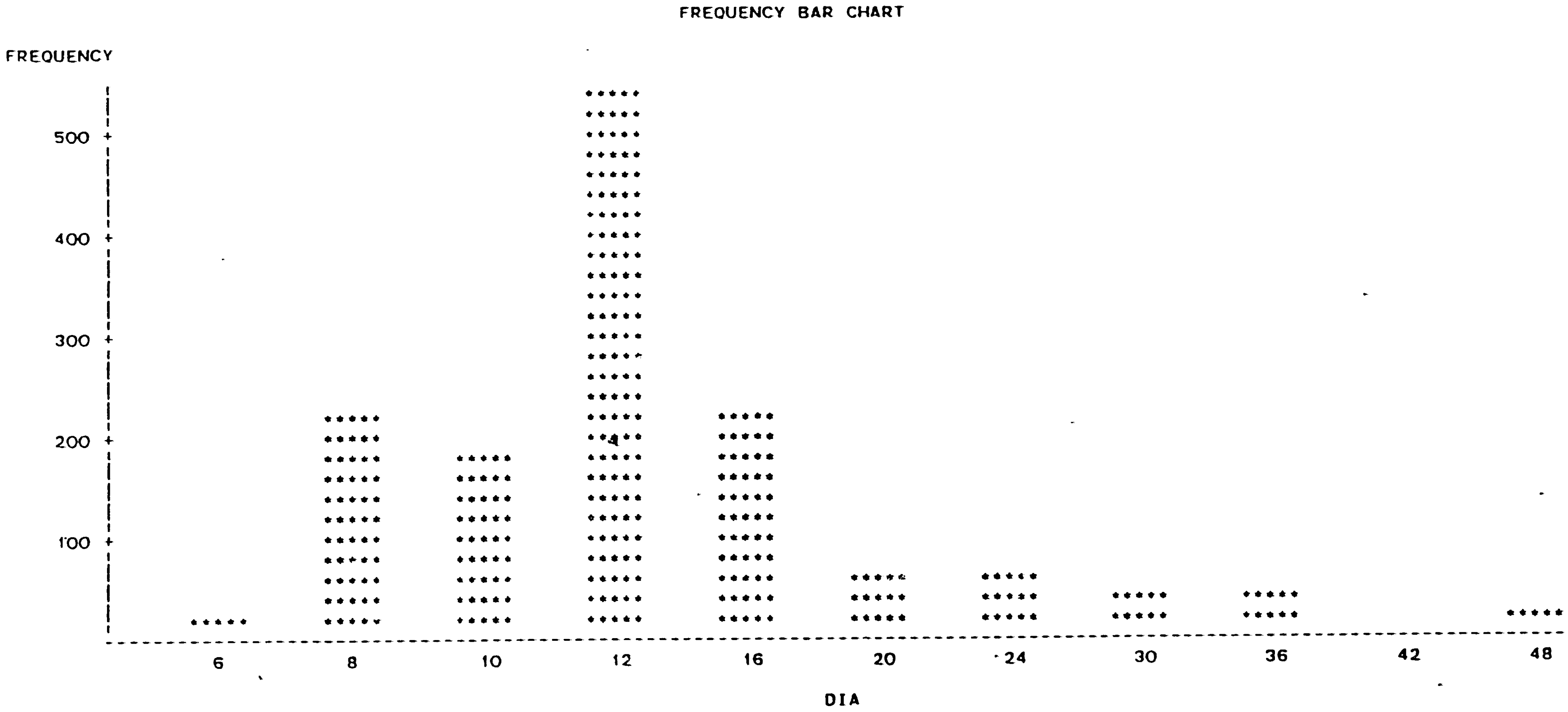
FREQUENCY BAR CHART




FREQUENCY BAR CHART
FREQUENCY
400
300
200
100
0
8
16
24
32
40
48
56
64
72
80
AGE83 MIDPOINT

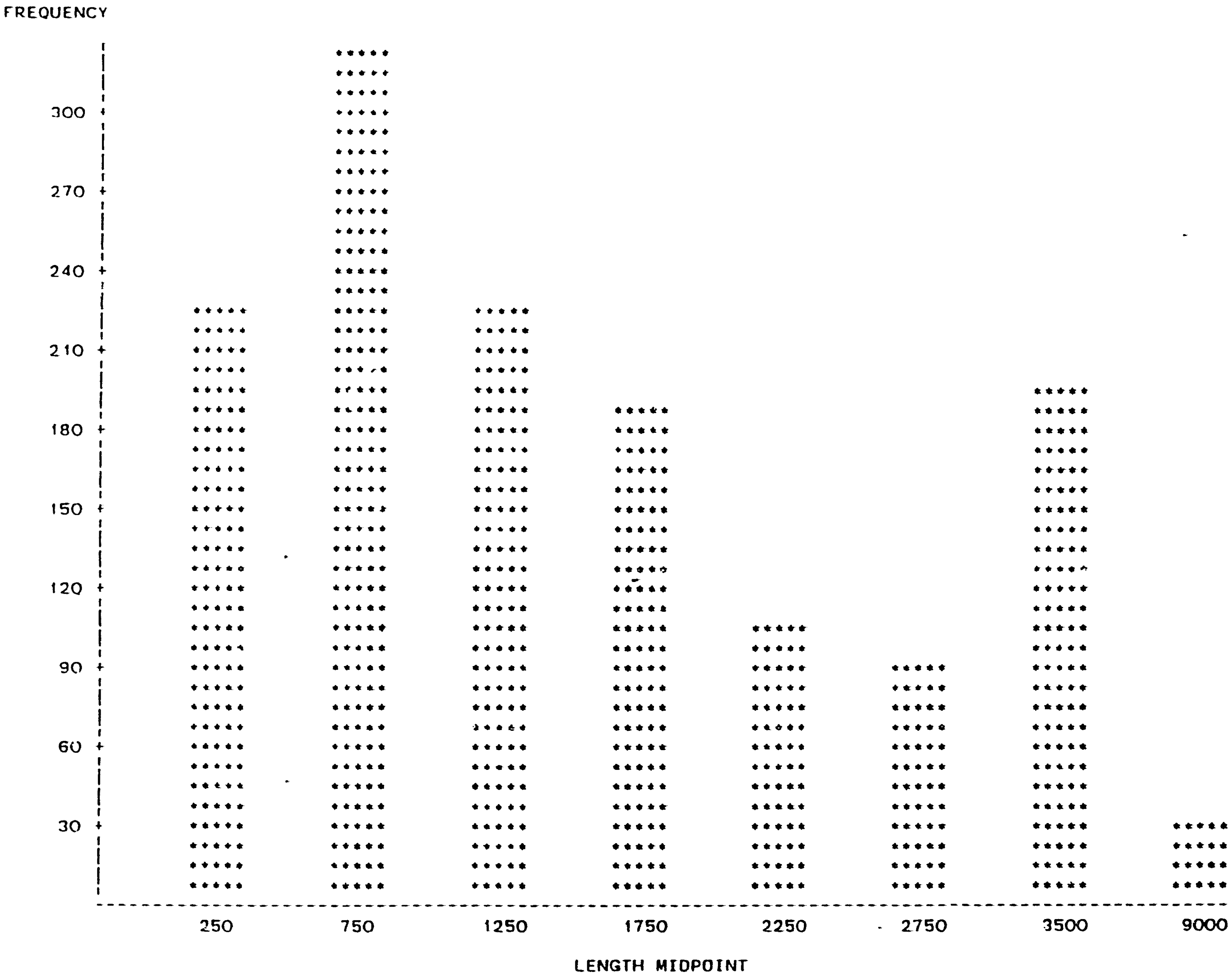
FREQUENCY BAR CHART




FREQUENCY BAR CHART
FREQUENCY
500
400
300
200
100
6
8
10
12
16
20
24
30
36
42
48
DIA

A 5

FREQUENCY BAR CHART



APPENDIX B

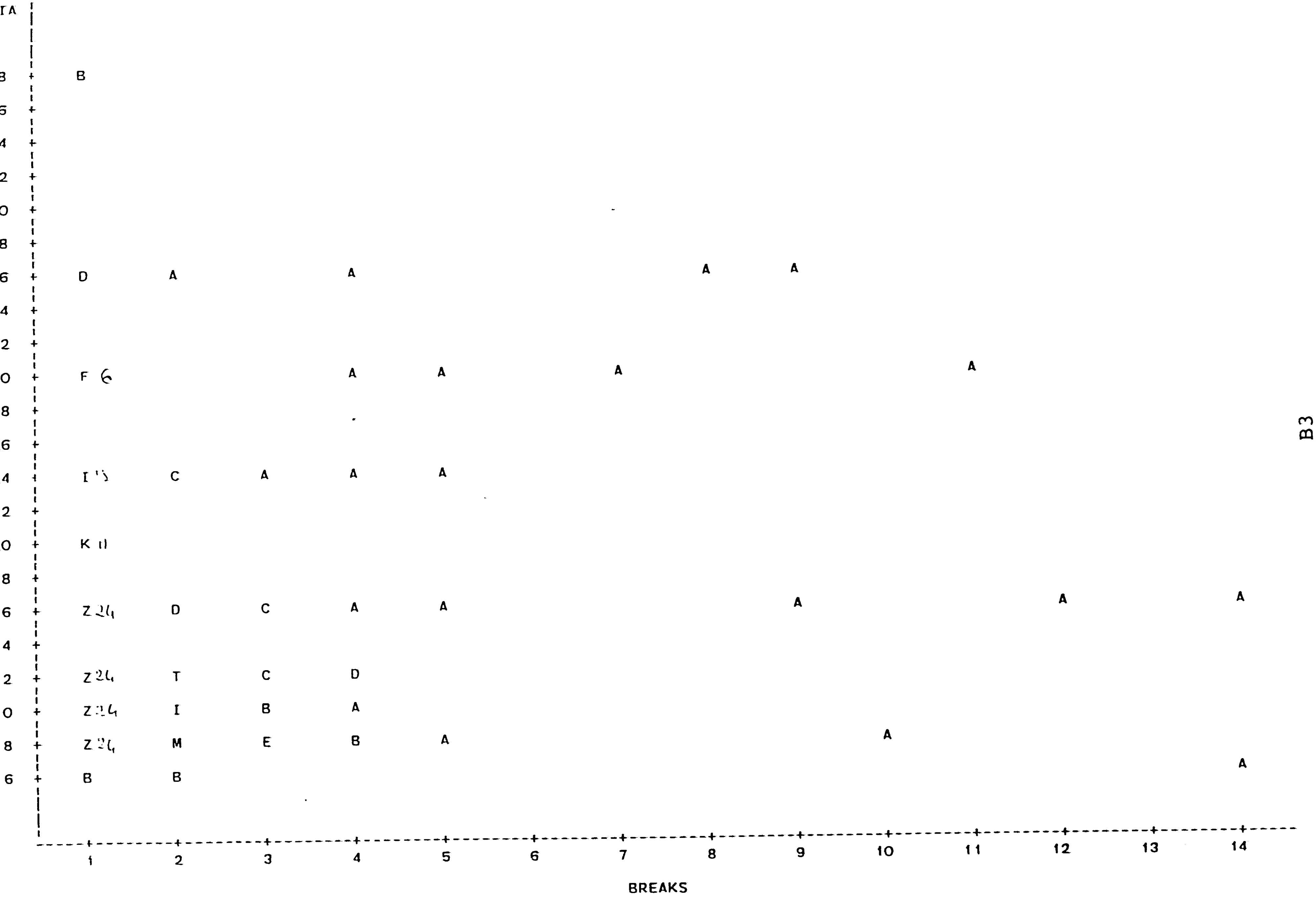
BIVARIATE SCATTER PLOTS FOR PAIRS OF VARIABLES

```
                                                     NUMBER OF BREAKS VS.AGE

                                    PLOT OF AGE83*BREAKS    LEGEND: A = 1 OBS. B = 2 OBS. ETC.

.GE83 |
      |
      |
      |
00    +
      |
      |
      |   M 13     E                 C        A                          A        A
      |            A
0     +
      |   B                                   A                                   A                          A                 A
      |   B
      |   A        B
      |   A        A
0     +   D                 A
      |   B        A        A
      |   B
      |   A
      |   C        B        B
0     +
      |   B        B        A
      |   H 8      A        A        B
      |   Z 26     I
      |   X 24     E        B        A        A                 A                                   A
0     +   O 15     F
      |   G 7
      |   B
      |   T 20     F        A
      |   B        B        A
0     +
      |   A                 A
      |
      |   G 7      A        A        A
      |   A
0     +   B
      |   D        B
      |   I        A                          A
      |   D        A        A        A                                                      A                                   A
      |   C                          A
0     +   G 7      A        A
      |   F 6
      |   C
      |   C        A                 A
      |   J 10     B        A
0     +   A
      |   A
      |
      |
      |
0     +
      ----+--------+--------+--------+--------+--------+--------+--------+--------+--------+--------+--------+--------+--------+-----
          1        2        3        4        5        6        7        8        9       10       11       12       13       14

                                                            BREAKS
```

B2

NUMBER OF BREAKS VS.DIAMETER

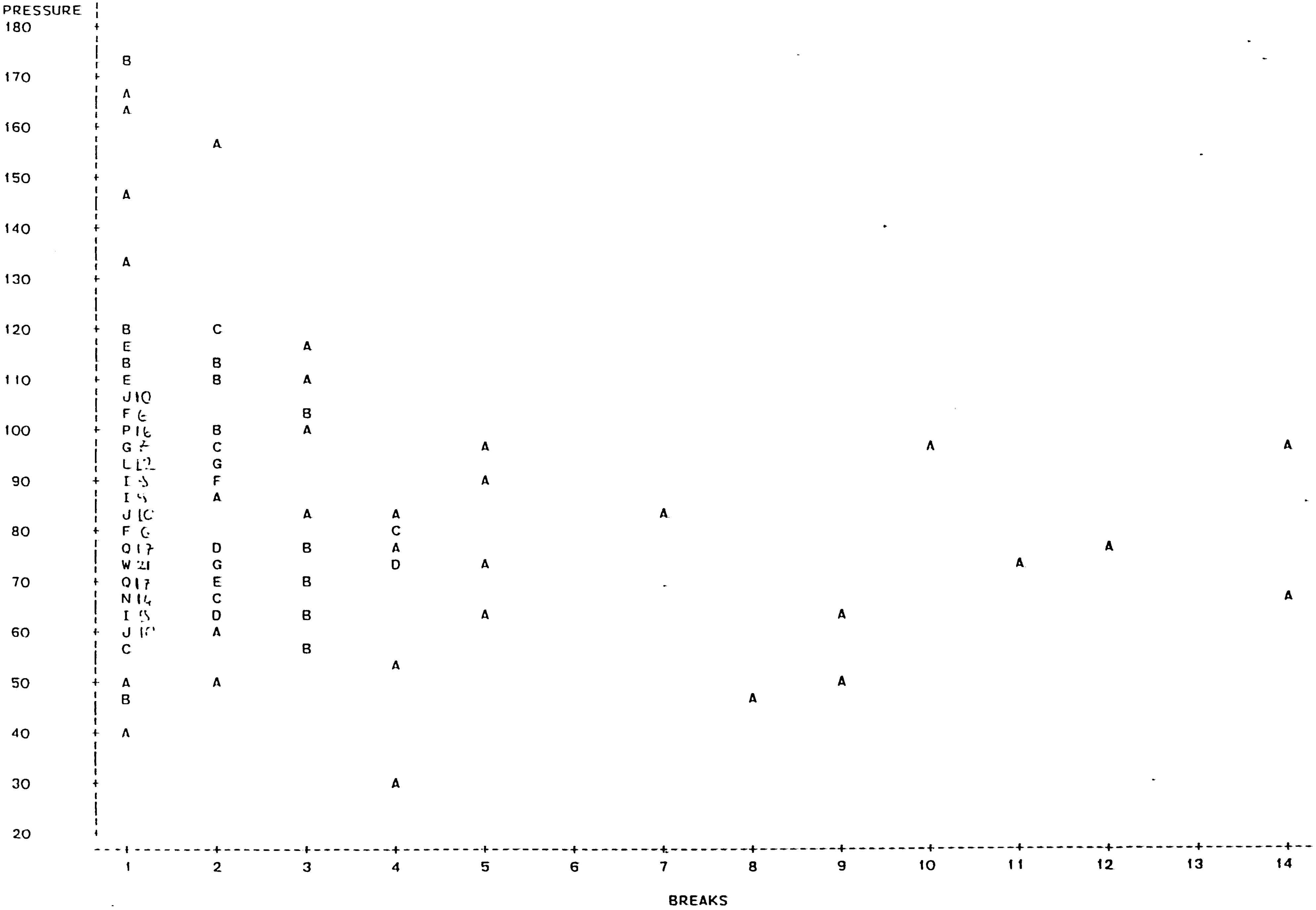
PLOT OF DIA*BREAKS   LEGEND: A = 1 OBS. B = 2 OBS. ETC.



B3

NUMBER OF BREAKS VS. PRESSURE

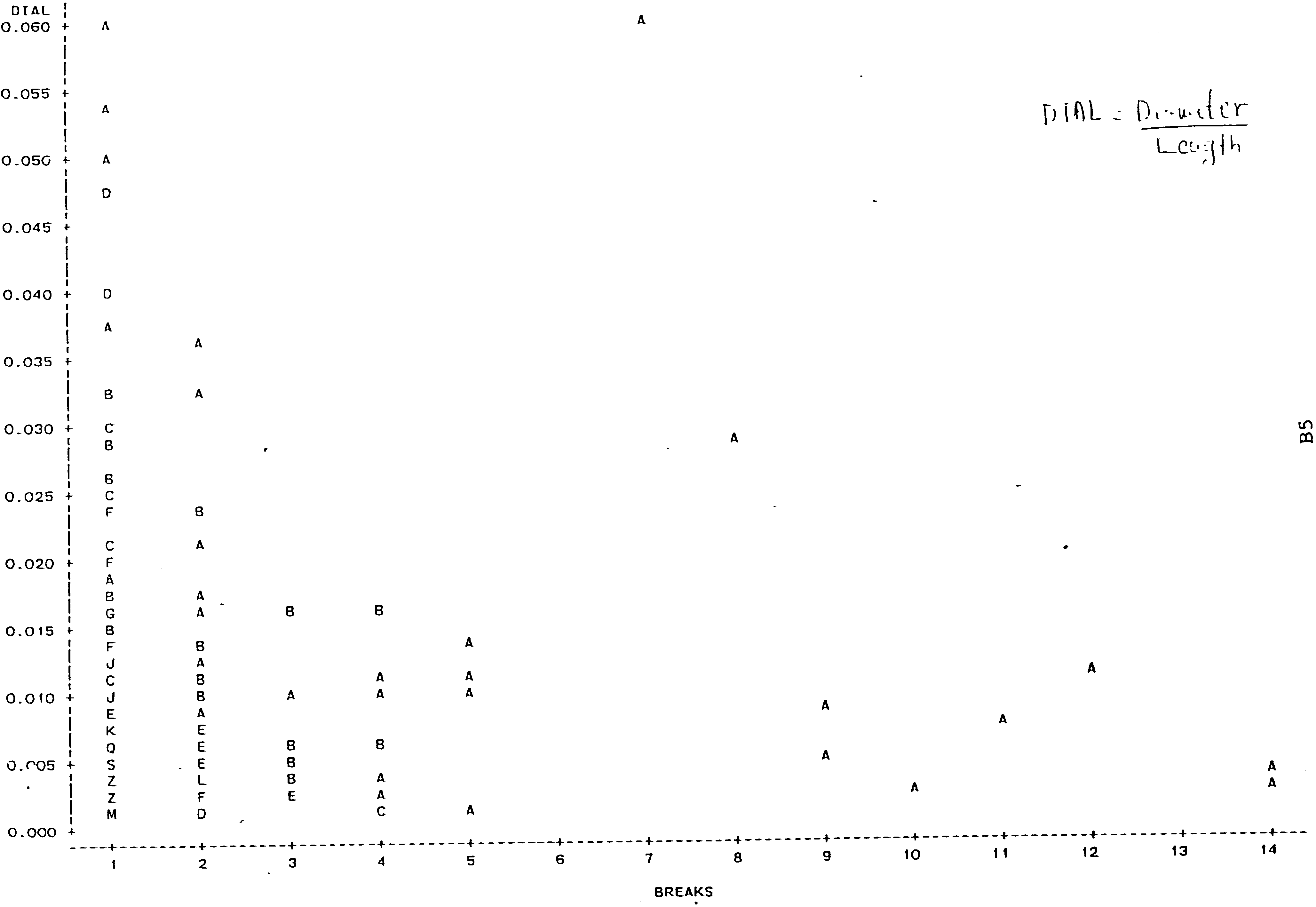
PLOT OF PRESSURE*BREAKS   LEGEND: A = 1 OBS, B = 2 OBS, ETC.



B4

NUMBER OF BREAKS VS.DIAL

PLOT OF DIAL*BREAKS   LEGEND: A = 1 OBS, B = 2 OBS, ETC.

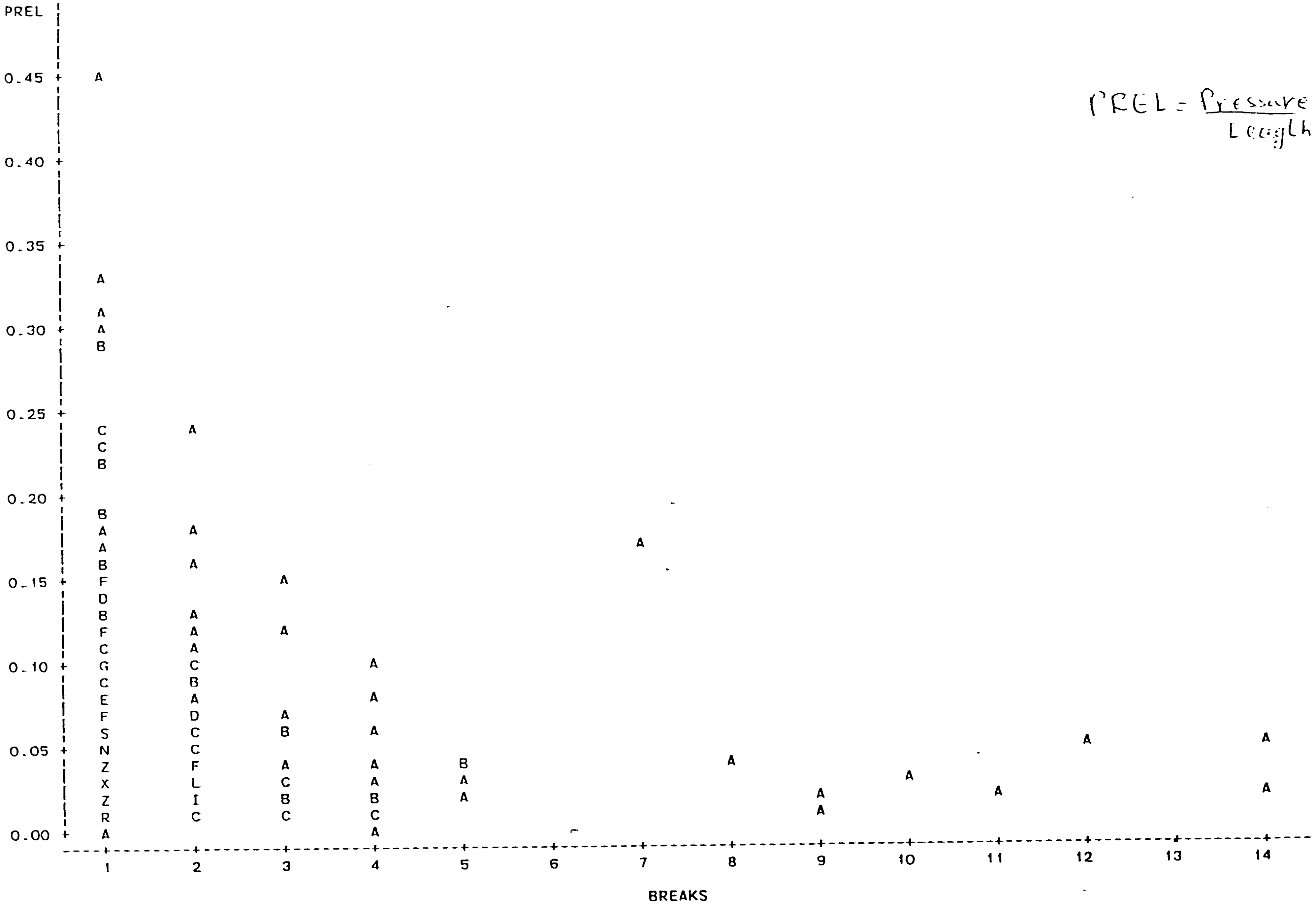


DIAL = Diameter / Length

B5

NUMBER OF BREAKS VS.PREL

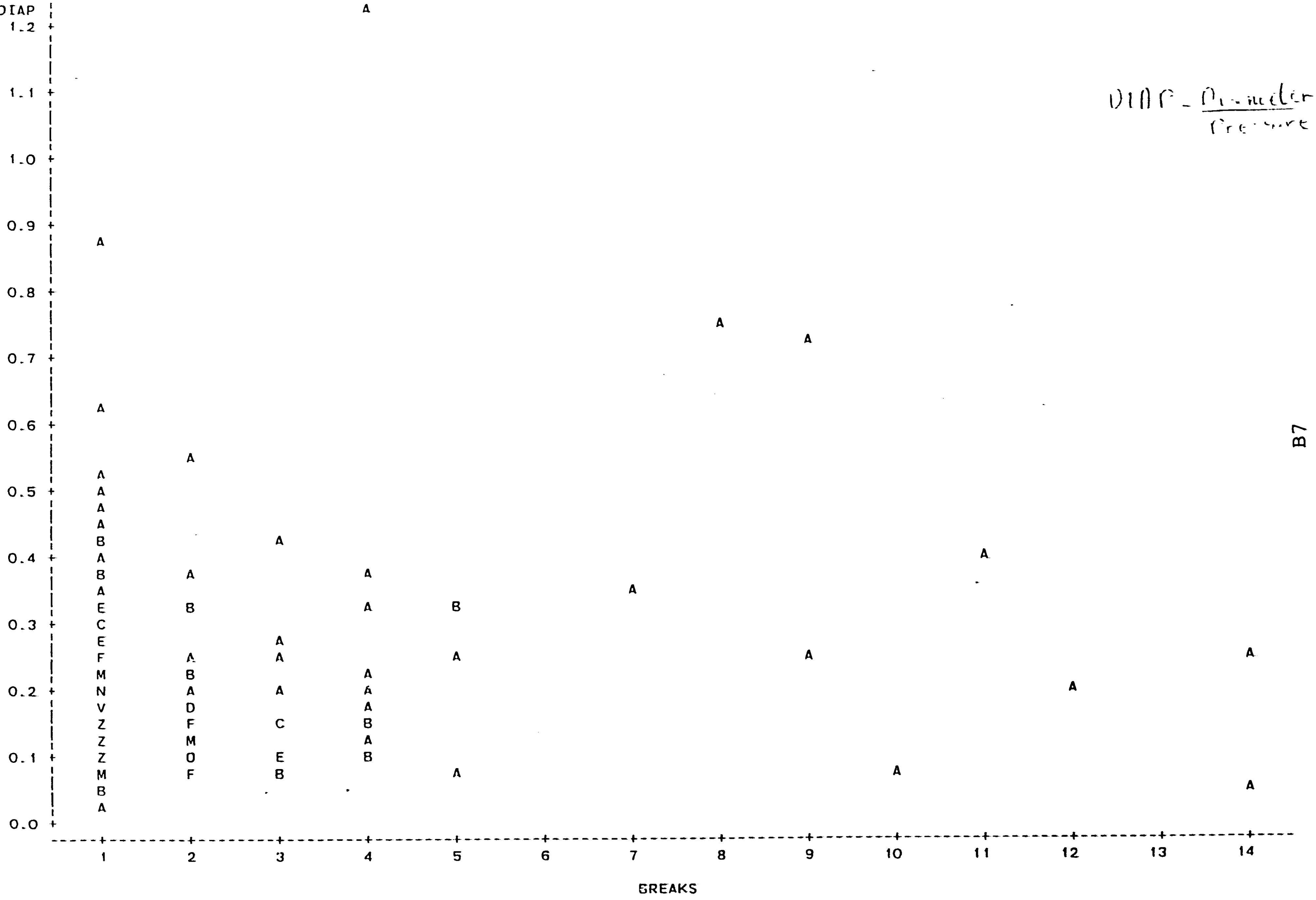
PLOT OF PREL*BREAKS   LEGEND: A = 1 OBS. B = 2 OBS. ETC.



PREL = Pressure / Length

NUMBER OF BREAKS VS.DIAP

PLOT OF DIAP*BREAKS   LEGEND: A = 1 OBS, B = 2 OBS, ETC.

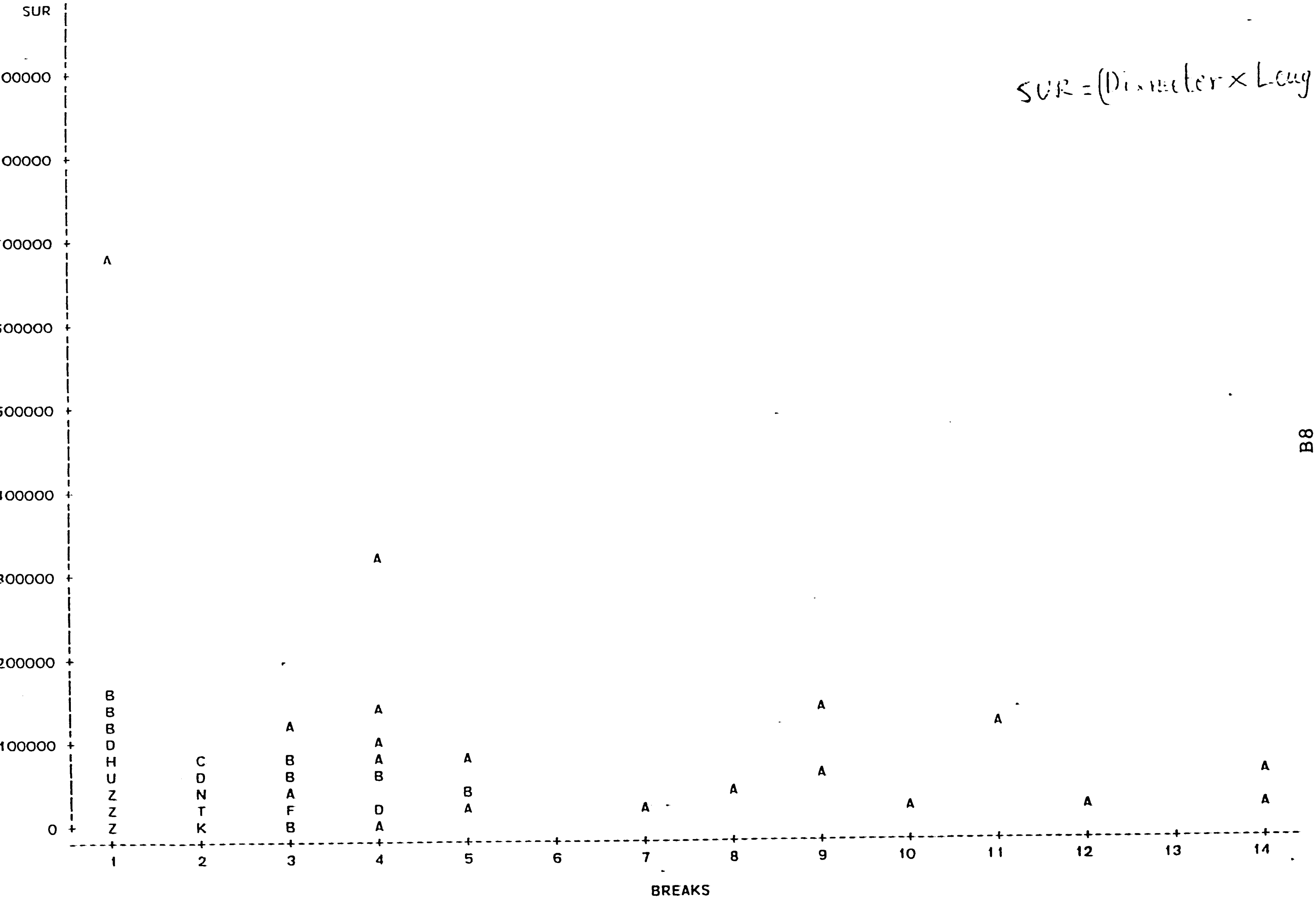


DIAP - Diameter / Pressure

B7

NUMBER OF BREAKS VS.SUR

PLOT OF SUR*BREAKS   LEGEND: A = 1 OBS. B = 2 OBS. ETC.



SUR = (Diameter × Length)

B8

APPENDIX C

CORRELATION ANALYSIS

1. Parametric Correlation Analysis

```
                                              BIVARIATE ANALYSIS

VARIABLE      N              MEAN            STD DEV                SUM            MINIMUM             MAXIMUM


LENGTH     1391     1901.54565061      1602.94309508      2645050.0000000       100.00000000      14000.00000000
BREAKS     1391        0.36376707         1.09279322          506.0000000                  0         14.00000000
CORR       1391        0.31056794         0.46289257          432.0000000                  0          1.00000000
AGE83      1391       41.79583034        17.09188069        58138.0000000         2.00000000         83.00000000
YR1RPRS     292       22.78164384        17.94002536         6652.2400000         0.08000000         80.00000000
YR2RPR       90        9.04033333        10.13074800          813.6300000         0.08000000         43.00000000
YR3RPR       38        3.78973684         3.84543147          144.0100000         0.01000000         18.00000000
YR4RPR       24        8.85625000        11.19440264          212.5500000         0.01000000         40.00000000
YR5RPR       13        2.47923077         4.29452260           32.2300000         0.01000000         15.00000000
YR6RPR        9        1.61444444         1.73823985           14.5300000         0.01000000          5.24000000
YR7RPR        9        3.92111111         3.25536267           35.2900000         0.01000000          9.00000000
YR8RPR        8        0.93625000         0.93812788            7.4900000         0.01000000          3.00000000
YR9RPR        7        2.12857143         3.19082240           14.9000000         0.07000000          9.00000000
YR10RPR       5        2.81800000         3.16028796           14.0900000         0.17000000          7.00000000
YR11RPR       4        5.89750000         2.02840455           23.5900000         3.00000000          7.42000000
YR12RPR       3        1.80333333         1.69441239            5.4100000         0.66000000          3.75000000
YR13RPR       2        1.08500000         1.29400541            2.1700000         0.17000000          2.00000000
YR14RPR       2        1.04000000         1.35764502            2.0800000         0.08000000          2.00000000
```

```
                                      BIVARIATE ANALYSIS

                CORRELATION COEFFICIENTS / PROB > |R| UNDER HO:RHO=0 / NUMBER OF OBSERVATIONS

            LENGTH    BREAKS      CORR     AGE83   YR1RPRS    YR2RPR    YR3RPR    YR4RPR    YR5RPR    YR6RPR    YR7RPR    YR8RPR    YR9RPR

LENGTH     1.00000   0.16733   0.08434  -0.11669  -0.02079   0.04081   0.09412  -0.08167  -0.00663  -0.44816   0.25478  -0.32556   0.30699
            0.0000    0.0001    0.0016    0.0001    0.7235    0.7025    0.5741    0.7044    0.9829    0.2264    0.5082    0.4313    0.5030
              1391      1391      1391      1391       292        90        38        24        13         9         9         8         7

BREAKS     0.16733   1.00000  -0.05710   0.16040  -0.11163  -0.06647  -0.17276  -0.37807  -0.22728  -0.25911   0.22613  -0.58145  -0.36076
            0.0001    0.0000    0.0332    0.0001    0.0567    0.5336    0.2996    0.0685    0.4552    0.5008    0.5585    0.1306    0.4266
              1391      1391      1391      1391       292        90        38        24        13         9         9         8         7

CORR       0.08434  -0.05710   1.00000  -0.16184  -0.06925  -0.01161   0.10097  -0.23348   0.14889   0.32960  -0.45126  -0.39895  -0.28449
            0.0016    0.0332    0.0000    0.0001    0.2381    0.9135    0.5464    0.2722    0.6273    0.3864    0.2228    0.3276    0.5363
              1391      1391      1391      1391       292        90        38        24        13         9         9         8         7

AGE83     -0.11669   0.16040  -0.16184   1.00000   0.61043   0.37287   0.12488   0.41616   0.25929   0.30290   0.37180   0.35792   0.40936
            0.0001    0.0001    0.0001    0.0000    0.0001    0.0003    0.4551    0.0431    0.3923    0.4282    0.3245    0.3840    0.3618
              1391      1391      1391      1391       292        90        38        24        13         9         9         8         7

YR1RPRS   -0.02079  -0.11163  -0.06925   0.61043   1.00000  -0.15821  -0.16024  -0.02053  -0.20845   0.22993   0.14347   0.63083   0.52399
            0.7235    0.0567    0.2381    0.0001    0.0000    0.1364    0.3365    0.9241    0.4943    0.5517    0.7127    0.0935    0.2274
               292       292       292       292       292        90        38        24        13         9         9         8         7

YR2RPR     0.04081  -0.06647  -0.01161   0.37287  -0.15821   1.00000   0.28113   0.17656   0.29113   0.14736  -0.21530  -0.31881   0.59986
            0.7025    0.5336    0.9135    0.0003    0.1364    0.0000    0.0873    0.4092    0.3345    0.7052    0.5780    0.4415    0.1545
                90        90        90        90        90        90        38        24        13         9         9         8         7

YR3RPR     0.09412  -0.17276   0.10097   0.12488  -0.16024   0.28113   1.00000  -0.00663  -0.16827   0.28270   0.40821  -0.04778   0.75276
            0.5741    0.2996    0.5464    0.4551    0.3365    0.0873    0.0000    0.9755    0.5827    0.4611    0.2754    0.9105    0.0508
                38        38        38        38        38        38        38        24        13         9         9         8         7

YR4RPR    -0.08167  -0.37807  -0.23348   0.41616  -0.02053   0.17656  -0.00663   1.00000   0.77069   0.08262   0.21083  -0.12939  -0.25342
            0.7044    0.0685    0.2722    0.0431    0.9241    0.4092    0.9755    0.0000    0.0020    0.8326    0.5861    0.7601    0.5835
                24        24        24        24        24        24        24        24        13         9         9         8         7

YR5RPR    -0.00663  -0.22728   0.14889   0.25929  -0.20845   0.29113  -0.16827   0.77069   1.00000  -0.25434  -0.11796  -0.45188  -0.26894
            0.9829    0.4552    0.6273    0.3923    0.4943    0.3345    0.5827    0.0020    0.0000    0.5090    0.7625    0.2610    0.5598
                13        13        13        13        13        13        13        13        13         9         9         8         7

YR6RPR    -0.44816  -0.25911   0.32960   0.30290   0.22993   0.14736   0.28270   0.08262  -0.25434   1.00000  -0.04193   0.02811  -0.02030
            0.2264    0.5008    0.3864    0.4282    0.5517    0.7052    0.4611    0.8326    0.5090    0.0000    0.9147    0.9473    0.9656
                 9         9         9         9         9         9         9         9         9         9         9         8         7

YR7RPR     0.25478   0.22613  -0.45126   0.37180   0.14347  -0.21530   0.40821   0.21083  -0.11796  -0.04193   1.00000   0.62541   0.13666
            0.5082    0.5585    0.2228    0.3245    0.7127    0.5780    0.2754    0.5861    0.7625    0.9147    0.0000    0.0973    0.7702
                 9         9         9         9         9         9         9         9         9         9         9         8         7

YR8RPR    -0.32556  -0.58145  -0.39895   0.35792   0.63083  -0.31881  -0.04778  -0.12939  -0.45188   0.02811   0.62541   1.00000   0.52442
            0.4313    0.1306    0.3276    0.3840    0.0935    0.4415    0.9105    0.7601    0.2610    0.9473    0.0973    0.0000    0.2269
                 8         8         8         8         8         8         8         8         8         8         8         8         7

YR9RPR     0.30699  -0.36076  -0.28449   0.40936   0.52399   0.59986   0.75276  -0.25342  -0.26894  -0.02030   0.13666   0.52442   1.00000
            0.5030    0.4266    0.5363    0.3618    0.2274    0.1545    0.0508    0.5835    0.5598    0.9656    0.7702    0.2269    0.0000
                 7         7         7         7         7         7         7         7         7         7         7         7         7
```

```
                                              BIVARIATE ANALYSIS

                          CORRELATION COEFFICIENTS / PROB > |R| UNDER HO:RHO=0 / NUMBER OF OBSERVATIONS

            LENGTH    BREAKS      CORR     AGE83   YR1RPRS    YR2RPR    YR3RPR    YR4RPR    YR5RPR    YR6RPR    YR7RPR    YR8RPR    YR9RPR

YR10RPR   -0.25722  -0.39349   0.46026   0.51571   0.75347   0.55981  -0.50800   0.67245   0.48085   0.13092  -0.70809  -0.80304  -0.47443
            0.6761    0.5122    0.4354    0.3738    0.1414    0.3264    0.3822    0.2136    0.4123    0.8338    0.1808    0.1018    0.4194
                 5         5         5         5         5         5         5         5         5         5         5         5         5

YR11RPR    0.31316  -0.31251   0.03369   0.98427   0.68280   0.59718   0.22931   0.81264   0.23845   0.76194  -0.04981   0.15373   0.53544
            0.6868    0.6875    0.9663    0.0157    0.3172    0.4028    0.7707    0.1874    0.7616    0.2381    0.9502    0.8463    0.4646
                 4         4         4         4         4         4         4         4         4         4         4         4         4

YR12RPR   -0.21328   0.58437   0.00000  -0.99495  -0.81780  -0.83021  -0.14343  -0.99495  -0.80979  -0.95592   0.18965  -0.05345  -0.48203
            0.8632    0.6027    1.0000    0.0640    0.3904    0.3764    0.9084    0.0640    0.3992    0.1897    0.8785    0.9660    0.6798
                 3         3         3         3         3         3         3         3         3         3         3         3         3

YR13RPR    1.00000   0.00000   0.00000   1.00000   1.00000   1.00000   1.00000   1.00000   1.00000   1.00000   1.00000   1.00000   1.00000

                 2         2         2         2         2         2         2         2         2         2         2         2         2

YR14RPR    1.00000   0.00000   0.00000   1.00000   1.00000   1.00000   1.00000   1.00000   1.00000   1.00000   1.00000   1.00000   1.00000

                 2         2         2         2         2         2         2         2         2         2         2         2         2

           YR10RPR   YR11RPR   YR12RPR   YR13RPR   YR14RPR


LENGTH    -0.25722   0.31316  -0.21328   1.00000   1.00000
            0.6761    0.6868    0.8632
                 5         4         3         2         2

BREAKS    -0.39349  -0.31251   0.58437   0.00000   0.00000
            0.5122    0.6875    0.6027
                 5         4         3         2         2

CORR       0.46026   0.03369   0.00000   0.00000   0.00000
            0.4354    0.9663    1.0000
                 5         4         3         2         2

AGE83      0.51571   0.98427  -0.99495   1.00000   1.00000
            0.3738    0.0157    0.0640
                 5         4         3         2         2

YR1RPRS    0.75347   0.68280  -0.81780   1.00000   1.00000
            0.1414    0.3172    0.3904
                 5         4         3         2         2

YR2RPR     0.55981   0.59718  -0.83021   1.00000   1.00000
            0.3264    0.4028    0.3764
                 5         4         3         2         2

YR3RPR    -0.50800   0.22931  -0.14343   1.00000   1.00000
            0.3822    0.7707    0.9084
                 5         4         3         2         2

YR4RPR     0.67245   0.81264  -0.99495   1.00000   1.00000
            0.2136    0.1874    0.0640
                 5         4         3         2         2
```

```
                                  BIVARIATE ANALYSIS

              CORRELATION COEFFICIENTS / PROB > |R| UNDER HO:RHO=0 / NUMBER OF OBSERVATIONS

            YR10RPR   YR11RPR   YR12RPR   YR13RPR   YR14RPR

YR5RPR      0.48085   0.23845  -0.80979   1.00000   1.00000
             0.4123    0.7616    0.3992
                  5         4         3         2         2

YR6RPR      0.13092   0.76194  -0.95592   1.00000   1.00000
             0.8338    0.2381    0.1897
                  5         4         3         2         2

YR7RPR     -0.70809  -0.04981   0.18965.  1.00000   1.00000
             0.1808    0.9502    0.8785
                  5         4         3         2         2

YR8RPR     -0.80304   0.15373  -0.05345   1.00000   1.00000
             0.1018    0.8463    0.9660
                  5         4         3         2         2

YR9RPR     -0.47443   0.53544  -0.48203   1.00000   1.00000
             0.4194    0.4646    0.6798
                  5         4         3         2         2

YR10RPR     1.00000   0.46283  -0.61859   1.00000   1.00000
             0.0000    0.5372    0.5754
                  5         4         3         2         2

YR11RPR     0.46283   1.00000  -0.98864   1.00000   1.00000
             0.5372    0.0000    0.0960
                  4         4         3         2         2

YR12RPR    -0.61859  -0.98864   1.00000  -1.00000  -1.00000
             0.5754    0.0960    0.0000
                  3         3         3         2         2

YR13RPR     1.00000   1.00000  -1.00000   1.00000   1.00000

                  2         2         2         2         2

YR14RPR     1.00000   1.00000  -1.00000   1.00000   1.00000

                  2         2         2         2         2
```

```
                                              BIVARIATE ANALYSIS

VARIABLE       N                 MEAN             STD DEV                  SUM              MINIMUM               MAXIMUM


DIA         1391          14.50179727          7.51773661        20172.0000000           6.00000000           48.00000000

LENGTH      1391        1901.54565061       1602.94309508      2645050.0000000         100.00000000        14000.00000000

BREAKS      1391           0.36376707          1.09279322          506.0000000                    0           14.00000000

CORR        1391           0.31056794          0.46289257          432.0000000                    0            1.00000000

PRESSURE    1391          79.35011503         20.23135705       110376.0100000           4.33000000          173.24000000

AGE83       1391          41.79583034         17.09188069        58138.0000000           2.00000000           83.00000000


                            CORRELATION COEFFICIENTS / PROB > |R| UNDER HO:RHO=0 / N = 1391

                                        DIA    LENGTH    BREAKS      CORR  PRESSURE     AGE83

                          DIA       1.00000   0.03149   0.01840  -0.00223  -0.15306  -0.07870
                                     0.0000    0.2405    0.4930    0.9338    0.0001    0.0033

                          LENGTH    0.03149   1.00000   0.16733   0.08434   0.05933  -0.11669
                                     0.2405    0.0000    0.0001    0.0016    0.0269    0.0001

                          BREAKS    0.01840   0.16733   1.00000  -0.05710   0.03441   0.16040
                                     0.4930    0.0001    0.0000    0.0332    0.1996    0.0001

                          CORROSIVITY
                                   -0.00223   0.08434  -0.05710   1.00000  -0.13397  -0.16184
                                     0.9338    0.0016    0.0332    0.0000    0.0001    0.0001

                          PRESSURE -0.15306   0.05933   0.03441  -0.13397   1.00000  -0.04371
                                     0.0001    0.0269    0.1996    0.0001    0.0000    0.1032

                          AGE83    -0.07870  -0.11669   0.16040  -0.16184  -0.04371   1.00000
                                     0.0033    0.0001    0.0001    0.0001    0.1032    0.0000
```

```
VARIABLE      N               MEAN              STD DEV                 SUM             MINIMUM               MAXIMUM

DIAL       1378         0.01425802           0.01529237           19.647556          0.00100000            0.16000000
DIAP       1378         0.20505897           0.21376044          282.571266          0.03485130            5.54272517
PREL       1378         0.07734012           0.06824856          106.574691          0.00144333            0.77960000
SUR        1378     27717.12626996       36011.85449761     38194200.000000       1000.00000000       672000.00000000
BREAKS     1378         0.28447025           0.66036656          392.000000                   0            4.00000000
YR1RPRS     279        22.94050179          17.98222437         6400.400000          0.08000000           80.00000000


           CORRELATION COEFFICIENTS / PROB > |R| UNDER HO:RHO=0 / NUMBER OF OBSERVATIONS

                        DIAL      DIAP      PREL       SUR    BREAKS   YR1RPRS

         DIAL        1.00000   0.23368   0.71647  -0.23808  -0.13837  -0.01057
                      0.0000    0.0001    0.0001    0.0001    0.0001    0.8605
                        1378      1378      1378      1378      1378       279

         DIAP        0.23368   1.00000  -0.10984   0.37653  -0.05321   0.01948
                      0.0001    0.0000    0.0001    0.0001    0.0483    0.7460
                        1378      1378      1378      1378      1378       279

         PREL        0.71647  -0.10984   1.00000  -0.41509  -0.11092   0.00228
                      0.0001    0.0001    0.0000    0.0001    0.0001    0.9697
                        1378      1378      1378      1378      1378       279

         SUR        -0.23808   0.37653  -0.41509   1.00000   0.12876  -0.01519
                      0.0001    0.0001    0.0001    0.0000    0.0001    0.8006
                        1378      1378      1378      1378      1378       279

         BREAKS     -0.13837  -0.05321  -0.11092   0.12876   1.00000  -0.13850
                      0.0001    0.0483    0.0001    0.0001    0.0000    0.0207
                        1378      1378      1378      1378      1378       279

         YR1RPRS    -0.01057   0.01948   0.00228  -0.01519  -0.13850   1.00000
                      0.8605    0.7460    0.9697    0.8006    0.0207    0.0000
                         279       279       279       279       279       279
```

```
VARIABLE     N                MEAN              STD DEV                  SUM              MINIMUM              MAXIMUM

DIAL       292          0.01033339           0.01091482             3.017351           0.00106667           0.06000000
DIAP       292          0.17974742           0.12586824            52.486246           0.03621876           1.21334681
PREL       292          0.06356447           0.06454249            18.560824           0.00329667           0.44896667
SUR        292      36665.75342466       50087.79500059      10706400.000000        2500.00000000      672000.00000000
BREAKS     292          1.73287671           1.82307130           506.000000           1.00000000          14.00000000
YR1RPRS    292         22.78164384          17.94002536          6652.240000           0.08000000          80.00000000


                    CORRELATION COEFFICIENTS / PROB > |R| UNDER HO:RHO=0 / N = 292

                             DIAL      DIAP      PREL       SUR    BREAKS   YR1RPRS

                  DIAL    1.00000   0.27154   0.77123  -0.25099  -0.04687   0.00479
                           0.0000    0.0001    0.0001    0.0001    0.4249    0.9351

                  DIAP    0.27154   1.00000  -0.13490   0.61750   0.20112   0.03253
                           0.0001    0.0000    0.0211    0.0001    0.0005    0.5798

                  PREL    0.77123  -0.13490   1.00000  -0.36741  -0.12485   0.00785
                           0.0001    0.0211    0.0000    0.0001    0.0330    0.8937

                  SUR    -0.25099   0.61750  -0.36741   1.00000   0.09700  -0.02456
                           0.0001    0.0001    0.0001    0.0000    0.0981    0.6759

                  BREAKS -0.04687   0.20112  -0.12485   0.09700   1.00000  -0.11163
                           0.4249    0.0005    0.0330    0.0981    0.0000    0.0567

                  YR1RPRS 0.00479   0.03253   0.00785  -0.02456  -0.11163   1.00000
                           0.9351    0.5798    0.8937    0.6759    0.0567    0.0000
```

2. Spearman (non-parametric) correlation

```
                                   SPEARMAN BIVARIATE ANALYSIS

VARIABLE       N                MEAN             STD DEV              MEDIAN              MINIMUM              MAXIMUM


LENGTH      1391       1901.54565061       1602.94309508       1500.00000000         100.00000000       14000.00000000
BREAKS      1391          0.36376707          1.09279322                   C                    0          14.00000000
YR1RPRS      292         22.78164384         17.94002536         20.00000000           0.08000000          80.00000000
YR2RPR        90          9.04033333         10.13074800          4.00000000           0.08000000          43.00000000
YR3RPR        38          3.78973684          3.84543147          2.37499905           0.01000000          18.00000000
YR4RPR        24          8.85625000         11.19440264          3.51999950           0.01000000          40.00000000
AGE83       1391         41.79583034         17.09188069         49.00000000           2.00000000          83.00000000


            SPEARMAN CORRELATION COEFFICIENTS / PROB > |R| UNDER HO:RHO=0 / NUMBER OF OBSERVATIONS

                           LENGTH    BREAKS   YR1RPRS    YR2RPR    YR3RPR    YR4RPR     AGE83

              LENGTH      1.00000   0.20046  -0.04375   0.09363   0.20387   0.10074  -0.13973
                           0.0000    0.0001    0.4564    0.3801    0.2196    0.6395    0.0001
                             1391      1391       292        90        38        24      1391

              BREAKS      0.20046   1.00000  -0.13256  -0.07554  -0.06275  -0.42033   0.13873
                           0.0001    0.0000    0.0235    0.4791    0.7082    0.0408    0.0001
                             1391      1391       292        90        38        24      1391

              YR1RPRS    -0.04375  -0.13256   1.00000  -0.13272  -0.10560   0.03296   0.57437
                           0.4564    0.0235    0.0000    0.2124    0.5281    0.8785    0.0001
                              292       292       292        90        38        24       292

              YR2RPR      0.09363  -0.07554  -0.13272   1.00000   0.27700   0.27016   0.33322
                           0.3801    0.4791    0.2124    0.0000    0.0922    0.2017    0.0013
                               90        90        90        90        38        24        90

              YR3RPR      0.20387  -0.06275  -0.10560   0.27700   1.00000   0.16030   0.11046
                           0.2196    0.7082    0.5281    0.0922    0.0000    0.4543    0.5091
                               38        38        38        38        38        24        38

              YR4RPR      0.10074  -0.42033   0.03296   0.27016   0.16030   1.00000   0.40073
                           0.6395    0.0408    0.8785    0.2017    0.4543    0.0000    0.0523
                               24        24        24        24        24        24        24

              AGE83      -0.13973   0.13873   0.57437   0.33322   0.11046   0.40073   1.00000
                           0.0001    0.0001    0.0001    0.0013    0.5091    0.0523    0.0000
                             1391      1391       292        90        38        24      1391
```

```
                                         SPEARMAN CORRELATION ANALYSIS

VARIABLE        N                     MEAN                 STD DEV                  MEDIAN                 MINIMUM                 MAXIMUM


DIAL         1378               0.01425802              0.01529237              0.00960000              0.00100000              0.16000000

DIAP         1378               0.20505897              0.21376044              0.15833223              0.03485130              5.54272517

PREL         1378               0.07734012              0.06824856              0.05505082              0.00144333              0.77960000

SUR          1378           27717.12626996          36011.85449761          18000.00000000           1000.00000000         672000.00000000

BREAKS       1378               0.28447025              0.66036656                       0                       0              4.00000000

YR1RPRS       279              22.94050179             17.98222437             20.00000000              0.08000000             80.00000000


                  SPEARMAN CORRELATION COEFFICIENTS / PROB > |R| UNDER HO     =0 / NUMBER OF OBSERVATIONS

                                        DIAL       DIAP       PREL        SUR     BREAKS    YR1RPRS

                         DIAL        1.00000    0.41467    0.84212   -0.61878   -0.20182   -0.00301
                                      0.0000     0.0001     0.0001     0.0001     0.0001     0.9601
                                        1378       1378       1378       1378       1378        279

                         DIAP        0.41467    1.00000   -0.07764    0.25229   -0.12344   -0.01202
                                      0.0001     0.0000     0.0039     0.0001     0.0001     0.8416
                                        1378       1378       1378       1378       1378        279

                         PREL        0.84212   -0.07764    1.00000   -0.86207   -0.15249    0.00510
                                      0.0001     0.0039     0.0000     0.0001     0.0001     0.9324
                                        1378       1378       1378       1378       1378        279

                         SUR        -0.61878    0.25229   -0.86207    1.00000    0.14215   -0.04683
                                      0.0001     0.0001     0.0001     0.0000     0.0001     0.4359
                                        1378       1378       1378       1378       1378        279

                         BREAKS     -0.20182   -0.12344   -0.15249    0.14215    1.00000   -0.12582
                                      0.0001     0.0001     0.0001     0.0001     0.0000     0.0357
                                        1378       1378       1378       1378       1378        279

                         YR1RPRS    -0.00301   -0.01202    0.00510   -0.04683   -0.12582    1.00000
                                      0.9601     0.8416     0.9324     0.4359     0.0357     0.0000
                                         279        279        279        279        279        279
```

```
                                       SPEARMAN CORRELATION ANALYSIS

VARIABLE      N                 MEAN              STD DEV               MEDIAN              MINIMUM              MAXIMUM


DIAL        292           0.01033339           0.01091482           0.00600000           0.00106667           0.06000000
DIAP        292           0.17974742           0.12586824           0.14197868           0.03621876           1.21334681
PREL        292           0.06356447           0.06454249           0.04113799           0.00329667           0.44896667
SUR         292       36665.75342466       50087.79500059       25400.00000000        2500.00000000      672000.00000000
BREAKS      292           1.73287671           1.82307130           1.00000000           1.00000000          14.00000000
YR1RPRS     292          22.78164384          17.94002536          20.00000000           0.08000000          80.00000000


                     SPEARMAN CORRELATION COEFFICIENTS / PROB > |R| UNDER HO:RHO=0 / N = 292

                                  DIAL      DIAP      PREL       SUR    BREAKS   YR1RPRS


                     DIAL      1.00000   0.40776   0.86197  -0.61533  -0.10306   0.01748
                                0.0000    0.0001    0.0001    0.0001    0.0787    0.7661

                     DIAP      0.40776   1.00000  -0.06969   0.30111  -0.01841   0.01076
                                0.0001    0.0000    0.2351    0.0001    0.7540    0.8547

                     PREL      0.86197  -0.06969   1.00000  -0.86368  -0.12939   0.01446
                                0.0001    0.2351    0.0000    0.0001    0.0270    0.8056

                     SUR      -0.61533   0.30111  -0.86368   1.00000   0.09472  -0.05152
                                0.0001    0.0001    0.0001    0.0000    0.1062    0.3804

                     BREAKS   -0.10306  -0.01841  -0.12939   0.09472   1.00000  -0.13256
                                0.0787    0.7540    0.0270    0.1062    0.0000    0.0235

                     YR1RPRS   0.01748   0.01076   0.01446  -0.05152  -0.13256   1.00000
                                0.7661    0.8547    0.8056    0.3804    0.0235    0.0000
```

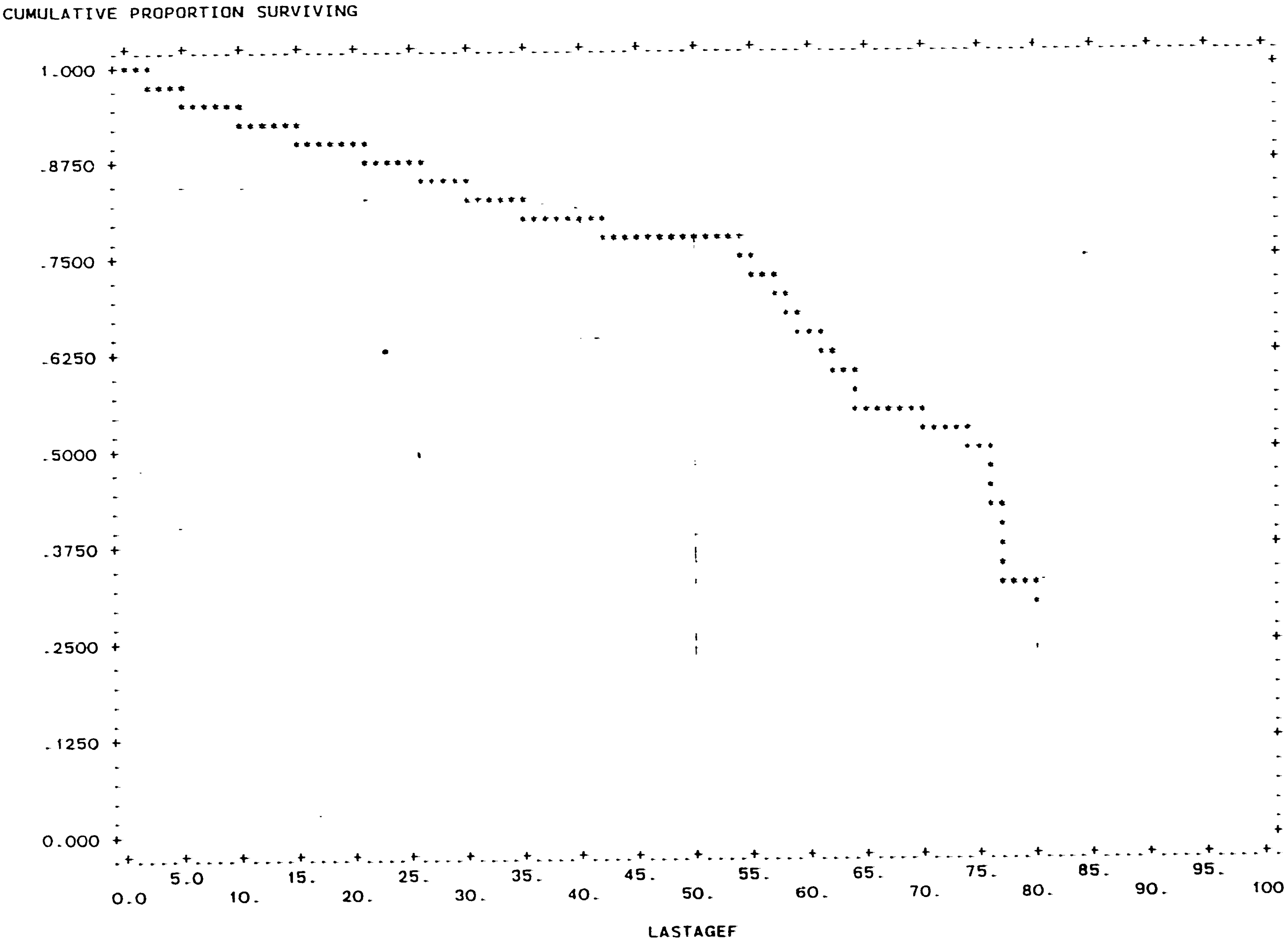
3. <u>Kendall (non-parametric) correlation</u>

DATA FOR PIPES WITH FIRST REPAIR WITHIN 2 YRS OF INSTALLATION

OBS	LINK	USNODE	DSNODE	YR1RPRS	YR2RPR	YR3RPR	YR4RPR	YR5RPR	YR6RPR	DIA	LENGTH	BREAKS	DATE	PRESSURE	TYPE1	TYPE2
1	8	10	11	0.08	5.92	.	.	.	.	8	3500	2	72	110.87	0	1
2	107	91	96	2.00	5.00	4.00	40.00	.	.	12	1200	4	28	74.71	0	1
3	138	72	71	2.00	.	.	.	.	.	20	800	1	31	66.70	0	1
4	248	202	201	2.00	.	.	.	.	.	16	2500	1	31	66.91	0	1
5	390	304	305	2.00	.	.	.	.	.	12	1500	1	60	82.59	0	1
6	392	303	302	2.00	.	.	.	.	.	16	2000	1	60	85.75	0	1
7	499	395	368	1.00	.	.	.	.	.	24	4000	1	58	94.42	0	1
8	637	498	510	1.00	.	.	.	.	.	16	2000	1	65	100.05	0	1
9	704	212	213	2.00	.	.	.	.	.	30	500	1	30	73.84	0	1
10	804	923	922	1.00	.	.	.	.	.	12	2250	1	71	77.31	0	1
11	840	935	937	2.00	.	.	.	.	.	12	1750	1	70	88.79	0	1
12	901	994	995	2.00	.	.	.	.	.	12	2100	1	71	120.19	0	1
13	903	995	997	2.00	.	.	.	.	.	12	1250	1	71	86.19	0	1
14	964	860	54	2.00	.	.	.	.	.	16	2000	1	62	65.40	0	1
15	1047	764	765	2.00	39.00	.	.	.	.	12	2250	2	26	74.71	0	1
16	1069	528	529	1.00	3.00	.	.	.	.	16	3250	2	63	68.43	0	1
17	1078	529	535	1.67	0.08	5.25	7.00	2.00	.	30	2750	5	58	90.95	0	1
18	1107	550	577	2.00	26.00	4.00	7.00	.	.	30	4500	4	31	80.99	0	1
19	1132	566	567	1.00	1.17	0.25	0.08	0.01	0.16	8	3000	10	60	96.36	0	1
20	1133	566	568	1.00	3.00	2.42	0.33	0.01	0.20	6	2000	14	60	96.36	0	1
21	1156	594	595	2.00	5.00	2.00	.	.	.	12	5700	3	63	116.29	0	1
22	1206	643	644	1.00	.	.	.	.	.	12	1000	1	57	100.70	0	1
23	1209	643	645	1.00	4.00	.	.	.	.	12	3000	2	57	92.47	0	1
24	1211	646	647	1.00	.	.	.	.	.	12	2800	1	57	68.86	0	1
25	1236	675	677	1.00	.	.	.	.	.	10	750	1	57	87.05	0	1
26	1326	705	706	1.00	0.50	1.08	1.42	.	.	12	6500	4	59	54.57	0	1
27	1344	1036	1037	1.00	4.00	.	.	.	.	12	1500	2	70	92.25	0	0
28	1380	836	850	1.00	.	.	.	.	.	8	1400	1	40	171.72	0	1
29	1390	909	911	2.00	21.00	.	.	.	.	10	1000	2	16	155.48	0	1

APPENDIX D

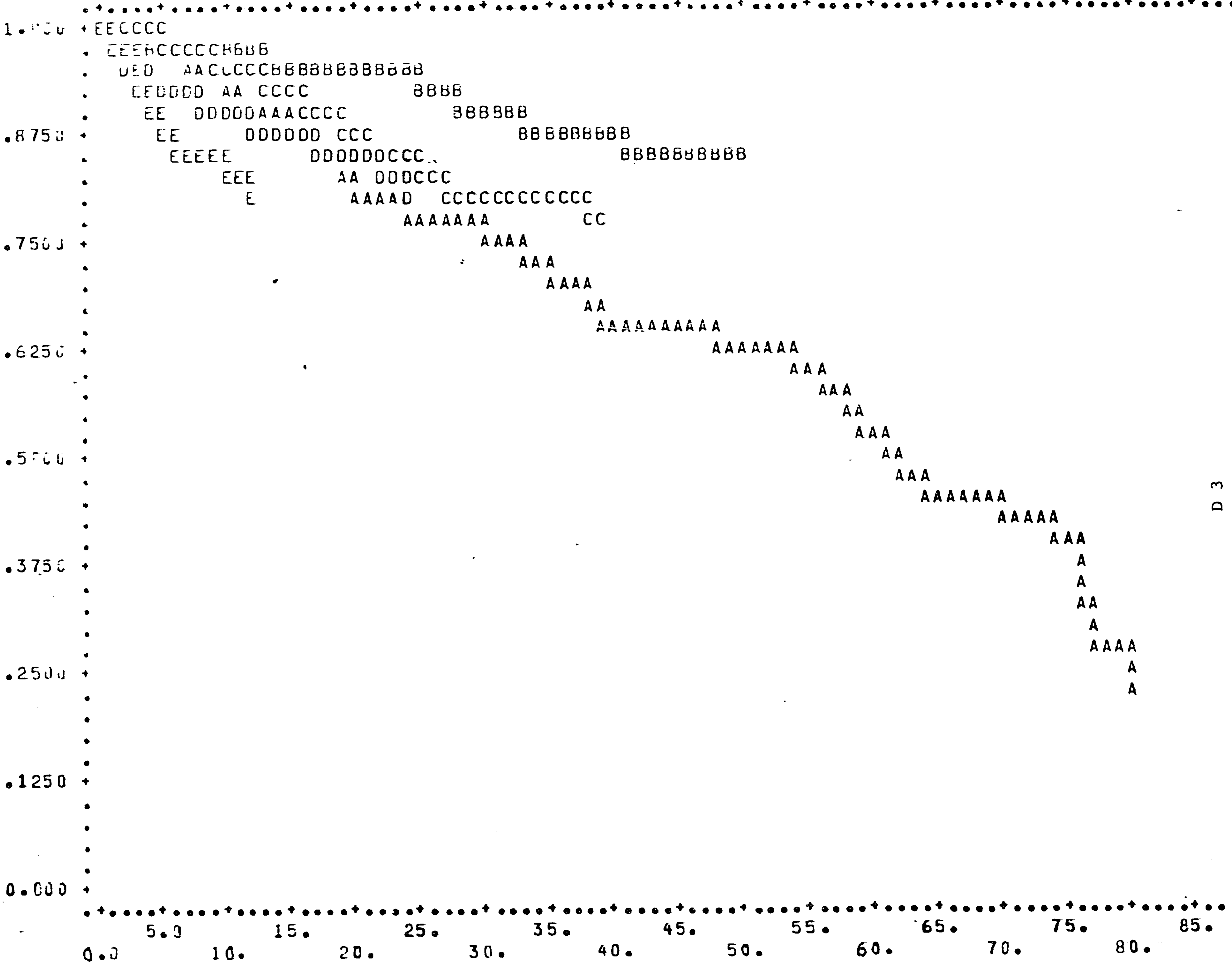
SURVIVAL ANALYSIS


CUMULATIVE PROPORTION SURVIVING
1.000
.8750
.7500
.6250
.5000
.3750
.2500
.1250
0.000
0.0
5.0
10.
15.
20.
25.
30.
35.
40.
45.
50.
55.
60.
65.
70.
75.
80.
85.
90.
95.
100
LASTAGEF

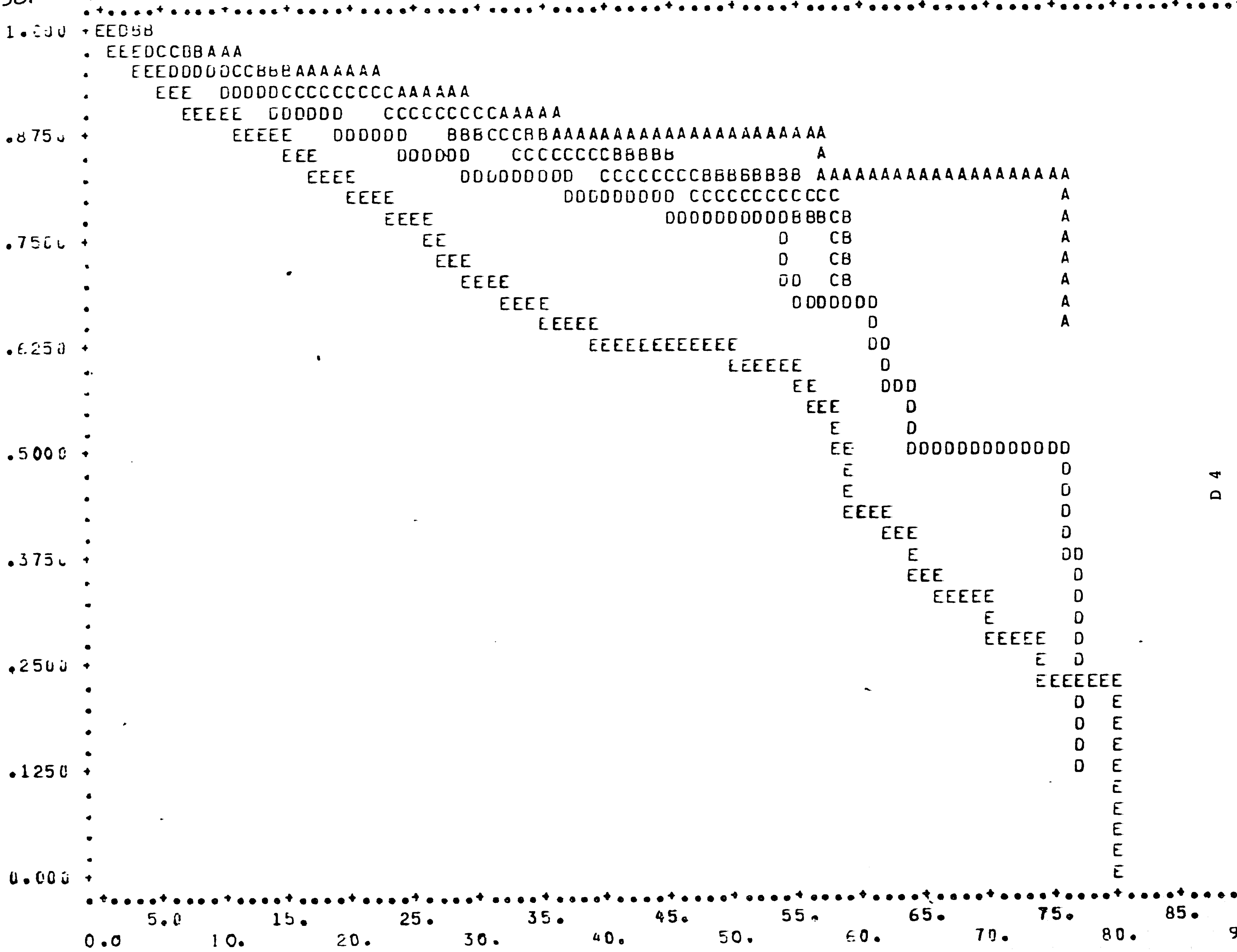
D 2

D = 51-65

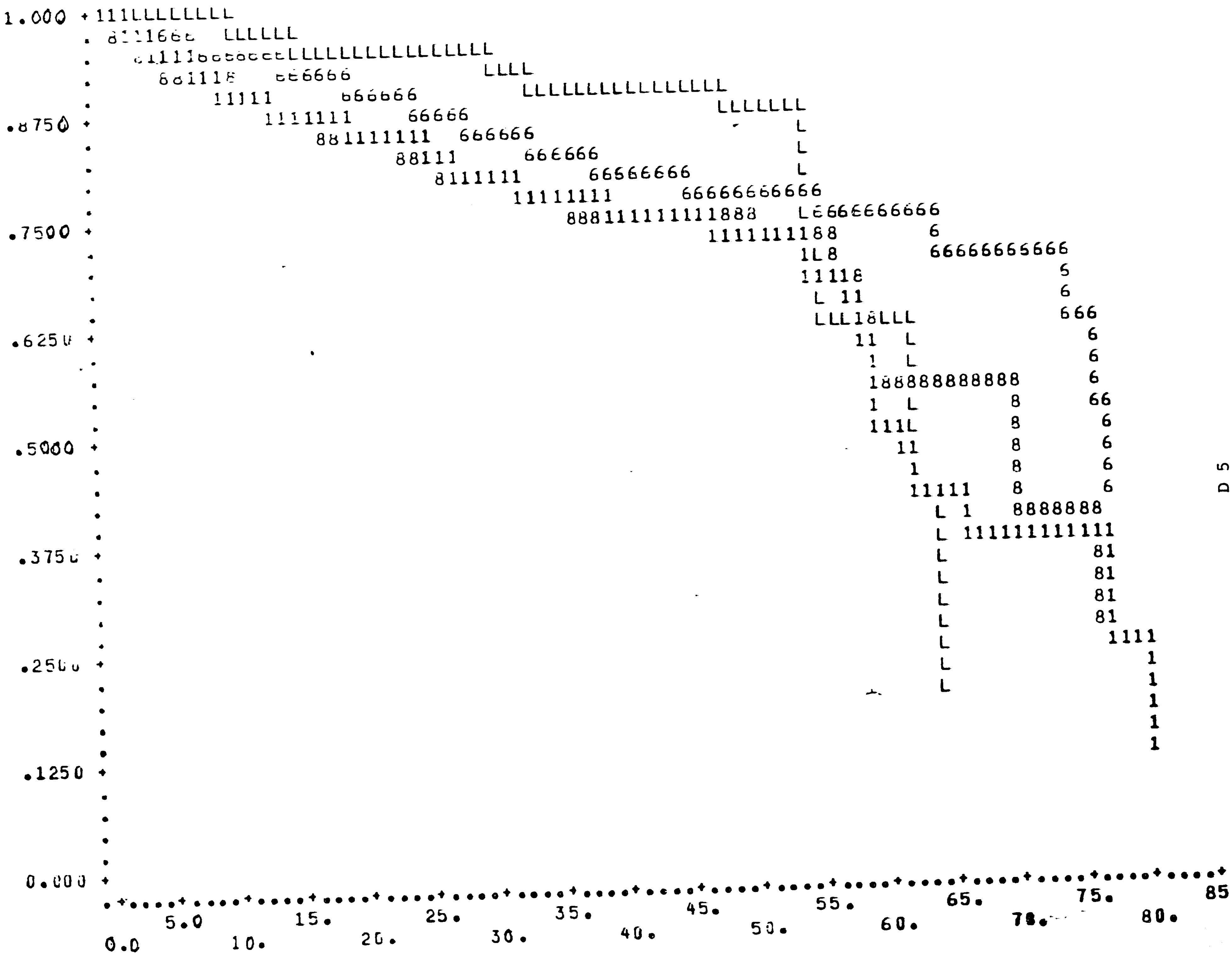
E = 65+


1.000
.8750
.7500
.6250
.5000
.3750
.2500
.1250
0.000
0.0
5.0
10.
15.
20.
25.
30.
35.
40.
45.
50.
55.
60.
65.
70.
75.
80.
85.
A
B
C
D
E
D3

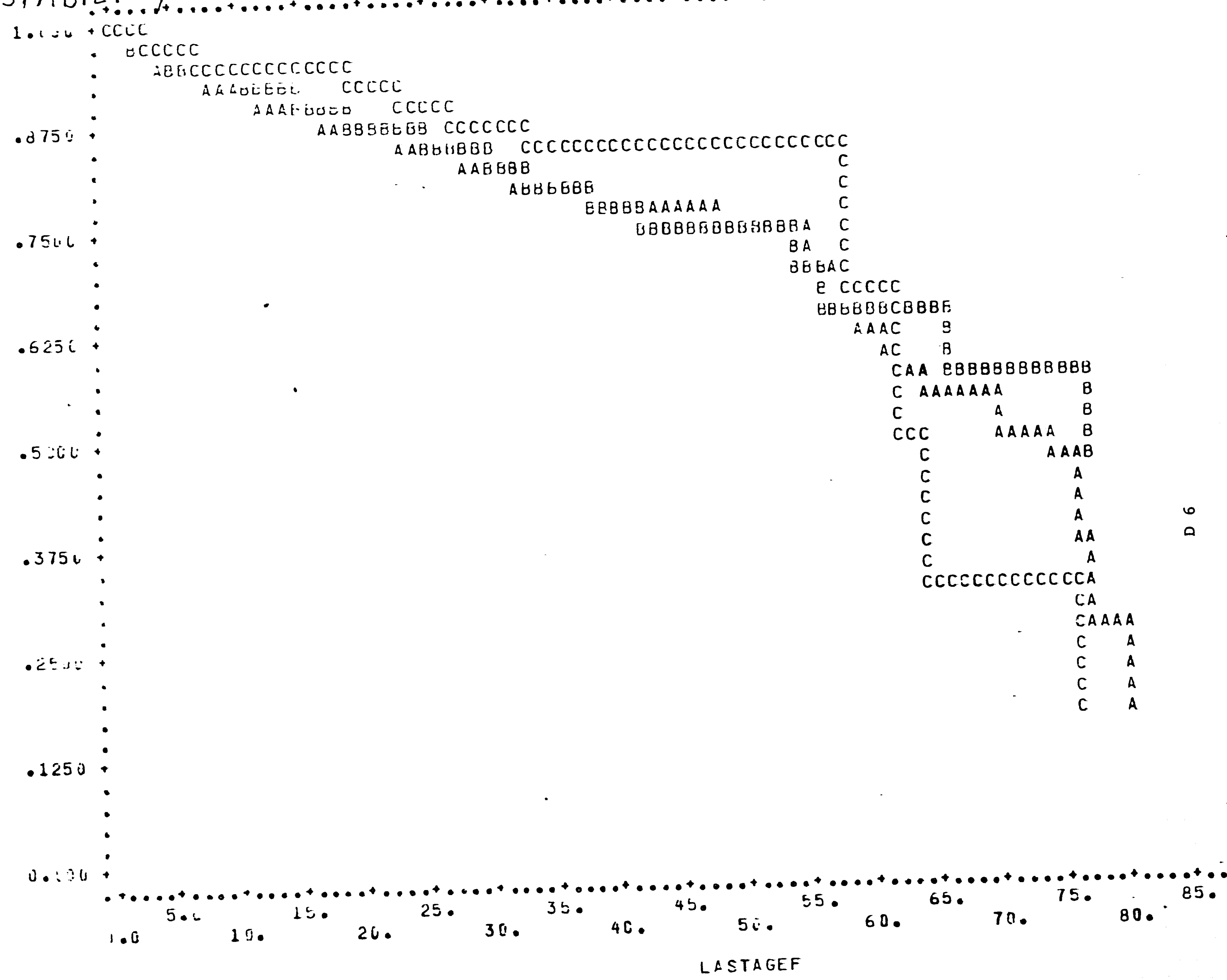
D= 1501 - 2500 E = 2500+



1 = 100 +

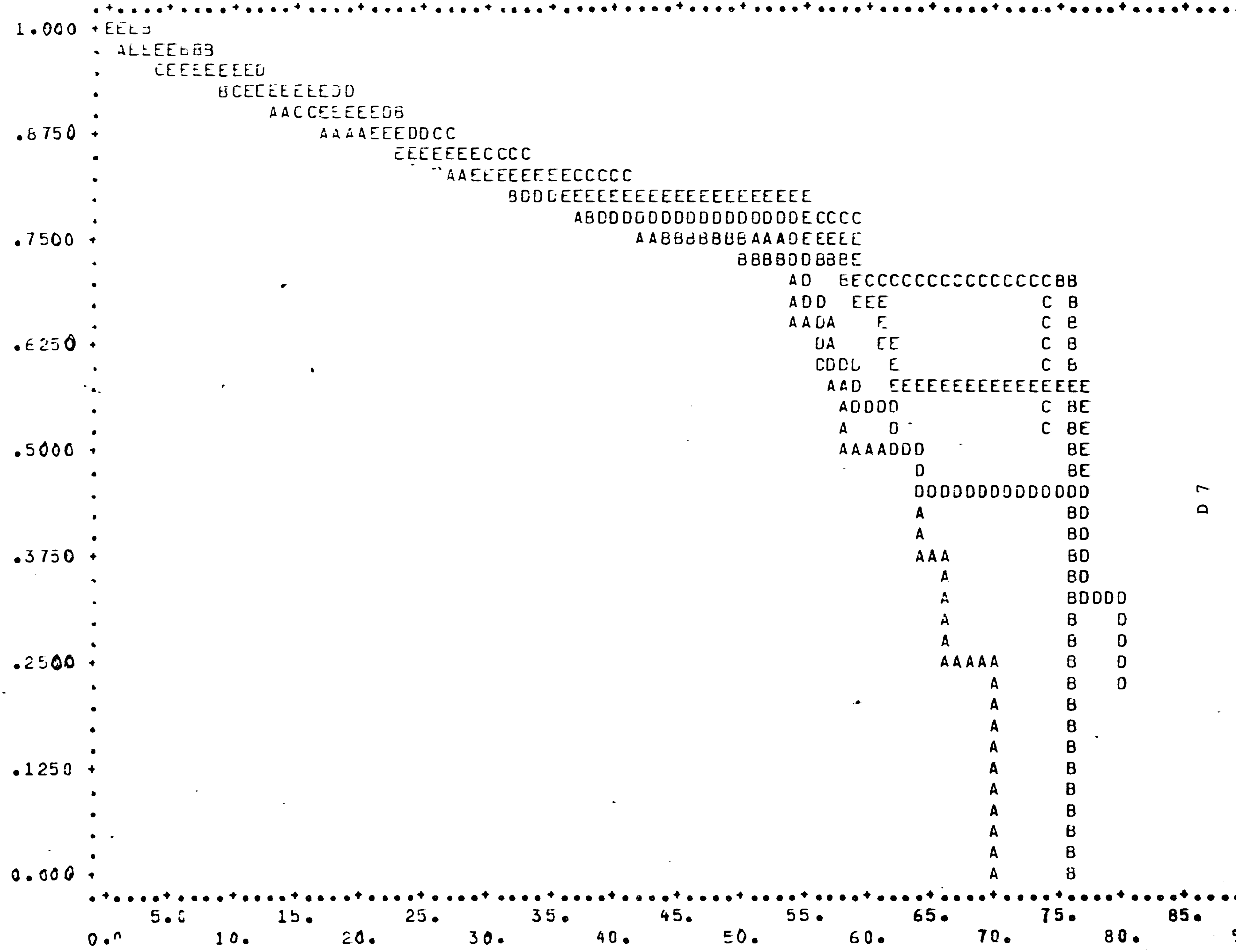


1.000
.8750
.7500
.6250
.5000
.3750
.2500
.1250
0.000
0.0
5.0
10.
15.
20.
25.
30.
35.
40.
45.
50.
55.
60.
65.
70.
75.
80.
85
D 5


1.000
.8750
.7500
.6250
.5000
.3750
.2500
.1250
0.000
1.0
5.0
10.
15.
20.
25.
30.
35.
40.
45.
50.
55.
60.
65.
70.
75.
80.
85.
LASTAGEF

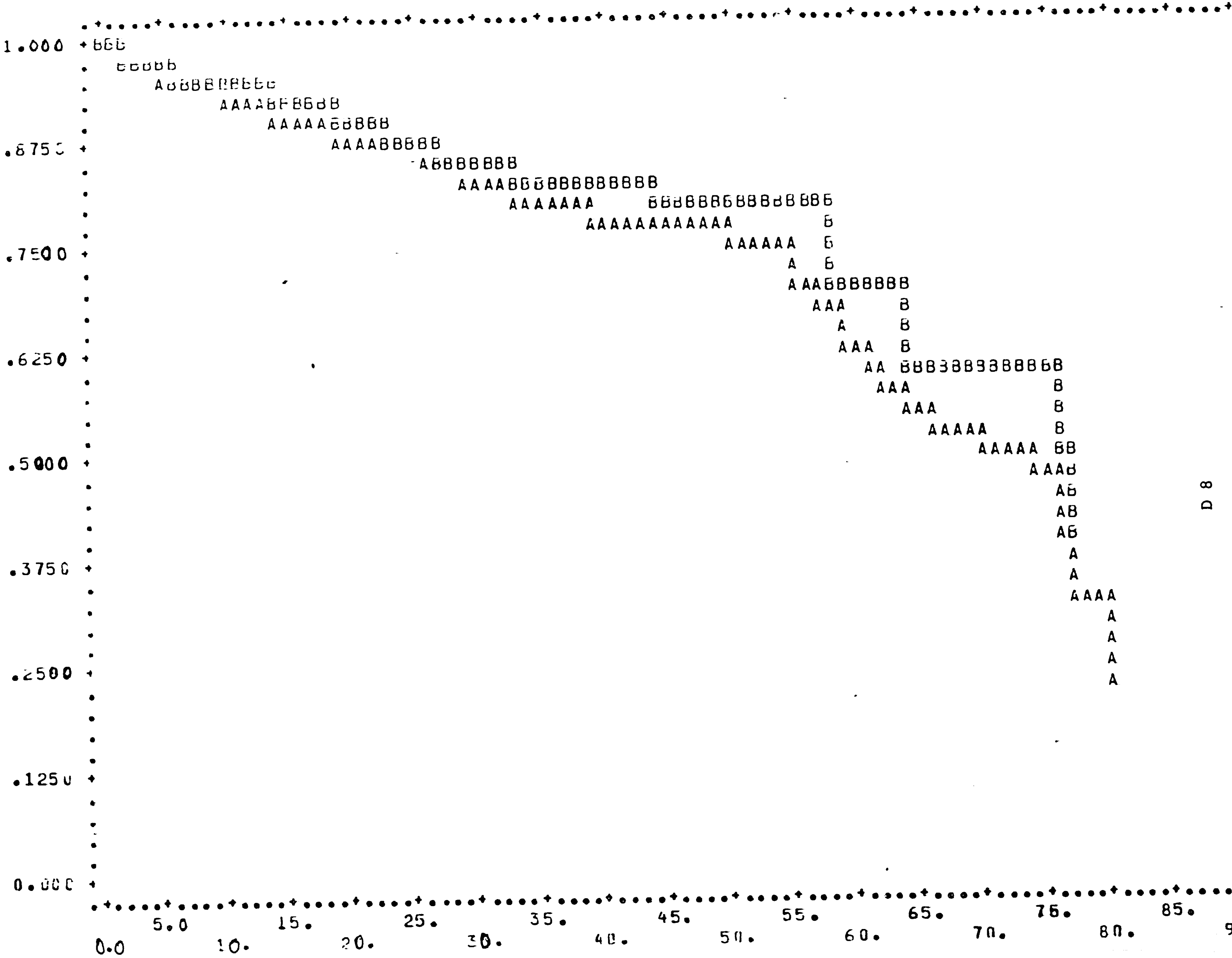
D 6

D= 13-16


1.000
.8750
.7500
.6250
.5000
.3750
.2500
.1250
0.000
0.0
5.0
10.
15.
20.
25.
30.
35.
40.
45.
50.
55.
60.
65.
70.
75.
80.
85.
A
B
C
D
E
D 7

CUMULATIVE PROPORTION SURVIVING

NOT CORROSIVE CORROSIVE



```
      .+....+....+....+....+....+....+....+....+....+....+....+....+....+....+....+....+....+....+
1.000 +SSN
      . SSSSS
      .     SSSSSSN
      .          SSSSSSN
      .               SSSSSSNN
.8750 +                    SSSSSNNN
      .                        SSSSSNNNN
      .                            SSSSSSNNN
      .                                 SSSSSSSSSSSSSN
      .                                            SSSSSSSNNNN
.7500 +                                                  SSSSSS
      .                                                       SSNN
      .                                                        S N
      .                                                        SSSSSSS
      .                                                           N  SSS
.6250 +                                                           NNN  S
      .                                                             NN SSSSSSSSSSSSS
      .                                                              NNN           S
      .                                                                NNN         S
      .                                                                  N         S
.5000 +                                                                  NNNNN     S
      .                                                                      N     S
      .                                                                      NNNNN S
      .                                                                          N S
      .                                                                          NNSS
.3750 +                                                                            NS
      .                                                                             S
      .                                                                             SNNN
      .                                                                             S  N
      .                                                                             S  N
.2500 +                                                                             S
      .                                                                             S
      .                                                                             S
      .
      .
.1250 +
      .
      .
      .
      .
0.000 +
      .+....+....+....+....+....+....+....+....+....+....+....+....+....+....+....+....+....+....+
           5.0       15.       25.       35.       45.       55.       65.       75.       85.
      0.0       10.       20.       30.       40.       50.       60.       70.       80.       9
```

D 9.

APPENDIX E

ESTIMATED PROBABILITIES FOR HIGH RISK PIPES

VARIABLES

HOT = $h_o(t)$ = baseline hazard function

BZ = $\bar{z}$ = covariate vector

HT = $h(t)$ = hazard rate

PRD = probability of break during the current year

PRB10 = probability of break after 10 years given that the pipe had no breaks until that time.

OBS	LINK	DIA	LENGTH	BREAKS	DATE	PRESSURE	LOW	FIRST	SECOND	T	PRESBRK	AGEBRK	HOT	BZ	HT	PRB	PRB10
1	8	8	3500	2	72	11.1	1.00	72	78	6	0.0	6	0.000150620	7.4462	0.26	0.22170	0.12861
2	21	16	1500	12	6	7.6	0.00	37	57	27	0.0	51	0.000053957	17.0372	1352.72	1.00000	1.00000
3	24	8	1400	3	41	7.8	0.00	74	76	8	0.0	35	0.000133852	6.6782	0.11	0.09811	0.05557
4	38	24	2250	4	0	7.5	0.00	18	35	49	0.0	35	0.000140945	8.1189	0.47	0.38569	0.56849
5	47	10	2000	2	0	6.9	0.00	19	20	64	0.0	20	0.000310700	6.0137	0.13	0.12190	0.17917
6	63	24	1750	5	0	7.4	0.00	14	34	50	0.0	34	0.000149476	9.1920	1.47	0.77922	0.92442
7	90	12	750	4	0	7.2	0.00	20	21	63	0.0	21	0.000296597	7.8370	0.75	0.53641	0.69158
8	107	12	1200	4	28	7.5	0.00	30	35	49	0.0	7	0.000140945	8.3974	0.63	0.47465	0.67054
9	108	10	1750	4	28	7.5	0.00	39	52	32	0.0	24	0.000056812	8.2260	0.21	0.19374	0.29926
10	133	24	1500	0	0	6.3	0.00	.	.	84	6.3	0	0.000676340	4.3430	0.05	0.05152	0.06919
11	136	20	1500	1	0	6.3	0.00	18	.	66	0.0	0	0.000340100	5.1132	0.06	0.05614	0.08287
12	140	16	2000	1	0	7.1	0.00	19	.	65	0.0	0	0.000325201	5.2677	0.06	0.06247	0.09268
13	191	16	2750	3	25	6.5	0.00	47	49	35	0.0	24	0.000063301	7.2827	0.09	0.08985	0.15307
14	218	16	3750	3	16	10.4	0.00	21	34	50	0.0	18	0.000149476	7.5812	0.29	0.26046	0.40300
15	290	12	1250	1	0	7.6	0.00	17	.	67	0.0	0	0.000355397	5.0153	0.05	0.05326	0.07815
16	345	8	7000	4	14	7.9	0.00	49	58	26	0.0	44	0.000054580	8.5304	0.28	0.24016	0.29097
17	425	12	2250	1	63	5.1	0.90	78	.	6	0.0	0	0.000150620	6.0996	0.07	0.06312	0.03518
18	452	8	2500	3	59	10.3	0.60	76	77	7	0.0	18	0.000142037	7.9663	0.41	0.32798	0.19659
19	468	24	6500	1	58	8.4	1.00	79	.	5	0.0	0	0.000159601	6.7246	0.13	0.12113	0.06845
20	480	16	6000	0	77	9.9	1.00	.	.	7	9.9	0	0.000142037	6.1490	0.07	0.06254	0.03494
21	509	12	3000	1	49	10.1	1.00	79	.	5	0.0	0	0.000159601	6.0384	0.07	0.06294	0.03507
22	601	8	5000	2	40	6.9	0.25	70	80	4	0.0	40	0.000168980	6.2040	0.08	0.07804	0.04371
23	603	8	7500	1	40	8.2	1.00	78	.	6	0.0	0	0.000150620	6.5305	0.10	0.09545	0.05361
24	623	12	2500	2	50	11.8	1.00	64	65	19	0.0	15	0.000070085	7.0675	0.08	0.07704	0.06153
25	642	8	2900	1	57	7.0	1.00	80	.	4	0.0	0	0.000168980	6.2912	0.09	0.08485	0.04760
26	644	8	2900	3	72	7.6	1.00	75	76	8	0.0	4	0.000133852	8.5752	0.71	0.49762	0.31694
27	649	8	4500	1	50	8.9	1.00	77	.	7	0.0	0	0.000142037	6.5272	0.10	0.08995	0.05058
28	683	16	5300	3	21	5.6	0.20	36	50	34	0.0	29	0.000060740	7.6356	0.13	0.12040	0.19948
29	684	24	5300	3	21	5.6	0.20	51	52	32	0.0	31	0.000056812	7.5916	0.11	0.10791	0.17187
30	693	6	3000	1	0	6.1	0.00	20	.	64	0.0	0	0.000310700	5.4854	0.07	0.07379	0.10990
31	723	36	4000	9	0	5.0	0.00	16	25	59	0.0	25	0.000244165	14.5779	523.34	1.00000	1.00000
32	726	36	1250	8	0	4.8	1.00	48	49	35	0.0	49	0.000063301	12.7923	22.75	1.00000	1.00000
33	729	36	9000	4	0	3.0	1.00	55	59	25	0.0	59	0.000055601	8.8884	0.40	0.32889	0.37436
34	774	24	3500	1	65	10.6	1.00	77	.	7	0.0	0	0.000142037	6.3922	0.08	0.07905	0.04434
35	781	24	1600	0	0	6.9	0.00	.	.	84	6.9	0	0.000676340	4.4173	0.06	0.05538	0.07432
36	809	12	6000	1	64	11.8	1.00	73	.	11	0.0	0	0.000111685	6.6816	0.09	0.08279	0.04848
37	828	8	1900	1	53	7.3	1.00	77	.	7	0.0	0	0.000142037	6.0641	0.06	0.05759	0.03214
38	833	16	4500	1	65	7.3	1.00	70	.	14	0.0	0	0.000093100	6.5272	0.06	0.05988	0.03772
39	843	16	8200	1	55	4.2	1.00	70	.	14	0.0	0	0.000093100	6.8494	0.09	0.08170	0.05168
40	861	12	7600	1	67	9.2	1.00	79	.	5	0.0	0	0.000159601	6.8086	0.14	0.13102	0.07421
41	865	12	7500	4	70	8.0	1.00	73	77	7	0.0	7	0.000142037	10.2055	3.84	0.97602	0.87182
42	876	16	2400	1	63	6.7	1.00	78	.	6	0.0	0	0.000150620	6.1896	0.07	0.06885	0.03842
43	878	12	2000	1	69	6.1	1.00	79	.	5	0.0	0	0.000159601	6.0917	0.07	0.06627	0.03695
44	904	12	3750	1	71	8.9	1.00	77	.	7	0.0	0	0.000142037	6.4292	0.09	0.08191	0.04597
45	940	10	4600	1	46	10.1	1.00	74	.	10	0.0	0	0.000118676	6.2680	0.06	0.05892	0.03376
46	948	12	6750	2	40	9.3	1.00	62	74	10	0.0	34	0.000118676	6.9119	0.12	0.10918	0.06329
47	950	8	5600	5	31	9.6	0.00	46	65	19	0.0	34	0.000070085	9.1366	0.65	0.46991	0.39516
48	978	8	1500	3	50	8.4	0.00	57	61	23	0.0	11	0.000058837	7.5142	0.11	0.10069	0.10353
49	1013	8	6000	2	40	9.1	0.50	45	48	36	0.0	8	0.000066260	7.1441	0.08	0.08238	0.14246
50	1020	8	3000	3	46	9.9	1.00	62	63	21	0.0	17	0.000063665	8.0354	0.20	0.17520	0.15913
51	1021	12	6200	1	6	8.4	0.00	19	.	65	0.0	0	0.000325201	5.8752	0.12	0.11168	0.16353
52	1023	16	3500	9	6	6.2	0.00	35	55	29	0.0	49	0.000053905	13.7782	63.43	1.00000	1.00000
53	1026	16	1700	5	6	6.4	0.00	61	62	22	0.0	56	0.000061052	8.3924	0.36	0.30008	0.29059
54	1027	12	3100	1	54	7.8	1.00	74	.	10	0.0	0	0.000118676	6.3270	0.07	0.06239	0.03578
55	1032	8	3500	4	50	7.5	1.00	60	63	21	0.0	13	0.000063665	9.6642	1.00	0.62502	0.58632

```
OBS LINK DIA LENGTH BREAKS DATE PRESSURE  LOW  FIRST SECOND  T  PRESBRK AGEBRK      HOT           BZ        HT     PRB     PRB10

 57  1045 16   3800    14     6     6.6   0.00   17     26    58    0.0     20    0.000232052 20.5904 203173 1.00000 1.00000
 58  1069 16   3250     2    63     6.8   0.00   64     67    17    0.0      4    0.000078097  6.8974      0 0.07241 0.05184
 59  1078 30   2750     5    58     9.1   0.00   60     60    24    0.0      2    0.000057020 10.4097      2 0.84533 0.87198
 60  1080 30    500     7    31     8.3   0.00   53     57    27    0.0     26    0.000053957 10.3872      2 0.82534 0.90145
 61  1087 30   3800    11    31     7.3   0.00   38     39    45    0.0      8    0.000110801 16.6164   1824 1.00000 1.00000
 62  1100 16   1000     4    62     8.3   0.00   67     67    17    0.0      5    0.000078097  8.6145      0 0.34197 0.25650
 63  1107 30   4500     4    31     8.1   0.00   33     59    25    0.0     28    0.000055601  7.9652      0 0.14652 0.16997
 64  1109 12   6000     3    28     6.3   0.00   50     67    17    0.0     39    0.000078097  7.3716      0 0.11376 0.08197
 65  1132  8   3000    10    60     9.6   0.00   61     62    22    0.0      2    0.000061052 16.3864    798 1.00000 1.00000
 66  1133  6   2000    14    60     9.6   0.00   61     64    20    0.0      4    0.000066676 20.8687  77112 1.00000 1.00000
 67  1156 12   5700     3    63    11.6   0.70   65     70    14    0.0      7    0.000093100  8.7062      1 0.42059 0.28807
 68  1162 12   4500     1    57     6.6   0.75   79      .     5    0.0      0    0.000159601  6.3889      0 0.08817 0.04942
 69  1187  8   2500     2    71     9.2   0.00   74     75     9    0.0      4    0.000126065  6.7565      0 0.09982 0.05705
 70  1209 12   3000     2    57     9.2   1.00   58     62    22    0.0      5    0.000061052  7.3854      0 0.09204 0.08873
 71  1252 12   2900     1    55     6.1   1.00   78      .     6    0.0      0    0.000150620  6.2912      0 0.07593 0.04245
 72  1259 10   2400     1    56     6.1   1.00   79      .     5    0.0      0    0.000159601  6.1896      0 0.07283 0.04067
 73  1269 12   1800     2    56     7.3   1.00   66     67    17    0.0     11    0.000078097  6.9791      0 0.07832 0.05613
 74  1274 12   5000     2    56     9.6   1.00   60     77     7    0.0     21    0.000142037  7.3077      0 0.18594 0.10711
 75  1280 20   4700     1    73     8.7   1.00   76      .     8    0.0      0    0.000133852  6.5505      0 0.08688 0.04908
 76  1286 16   3500     1    62     7.7   1.00   79      .     5    0.0      0    0.000159601  6.3922      0 0.08845 0.04958
 77  1326 12   6500     4    59     5.5   1.00   60     61    23    0.0      2    0.000058837 10.2386      2 0.80175 0.81106
 78  1332 12   8600     0     8    10.2   1.00    .      .    76   10.2      0    0.000510980  6.0912      0 0.20549 0.27676
 79  1337 12   1500     0     8    11.0   1.00    .      .    76   11.0      0    0.000510980  5.2062      0 0.09057 0.12517
 80  1338 16    600     0     0    11.0   1.00    .      .    84   11.0      0    0.000676340  4.7142      0 0.07380 0.09871
 81  1339 12   4000     0     8    10.8   1.00    .      .    76   10.8      0    0.000510980  5.7197      0 0.14671 0.20027
 82  1343 12   3400     0    80     9.8   1.00    .      .     4    9.8      0    0.000168980  5.8374      0 0.05477 0.03051
 83  1344 12   1500     2    70     9.2   1.00   71     75     9    0.0      5    0.000126065  7.0132      0 0.12710 0.07312
 84  1345 12   5000     0    78     8.5   1.00    .      .     6    8.5      0    0.000150620  5.9587      0 0.05506 0.03063
 85  1349 16   1250     0     8     7.5   1.00    .      .    76    7.5      0    0.000510980  4.8773      0 0.06605 0.09176
 86  1350 16   4600     0     8     2.4   1.00    .      .    76    2.4      0    0.000510980  5.2404      0 0.09357 0.12923
 87  1354 16   2500     0     8    10.1   1.00    .      .    76   10.1      0    0.000510980  5.4211      0 0.11104 0.15278
 88  1358 12   4000     1    65     8.3   1.00   76      .     8    0.0      0    0.000133852  6.4639      0 0.07997 0.04510
 89  1372 10   5750     0     8     7.6   1.00    .      .    76    7.6      0    0.000510980  5.7034      0 0.14452 0.19737
 90  1373  8   7000     1    50     7.7   1.00   77      .     7    0.0      0    0.000142037  6.7644      0 0.11263 0.06368
 91  1374  8   4700     0    14     7.5   1.00    .      .    70    7.5      0    0.000403676  5.5885      0 0.10430 0.14838
 92  1385 12   5500     1    71     9.9   1.00   77      .     7    0.0      0    0.000142037  6.6349      0 0.09965 0.05617
 93  1391 10   1600     0    16    15.5   0.00    .      .    68   15.5      0    0.000371092  4.9849      0 0.05390 0.07858
```