March, 1982

ARSONS IN NEWARK: A STUDY IN PREDICTION AND PATTERNS

BY

Hari Shiledar Baxi, Ph.D.
Consultant in Statistics & Data Processing
87-70 173 St., Jamaica, N.Y
(212) 657-0759

91028

tudy was supported by funds from the Law Enforcement
ance Administration through Grant No. 80-DF-AX-0006
d to City of Newark Fire Department.

ACKNOWLEDGMENTS

The author wishes to express his sincere thanks to the Law Enforcement Assistance Administration (LEAA) of U.S. Justice Department, New Jersey State Law Enforcement Planning Agency (SLEPA), and Newark Office of Criminal Justice Planning who provided the funds for this study.

The author deeply appreciates and expresses his most sincere thanks to the Hon. John P. Caufield, Director, City of Newark Fire Department, Mr. Stanley J. Kossup, Fire Chief, and Mr. Carl E. Stoffers, Battalion Chief, who provided all the cooperation necessary to complete this study. I would also like to express similar appreciation and thanks to the following persons for allowing me and the Fire Department staff to collect data from their respective departments: Mr. Joseph McGinley, Head of Code Enforcement Department, Mr. Joseph Paradise, Head of Water Department, Mr. Joseph Frisina, Tax Assessor of City of Newark, Mr. Kennth Joseph, Tax Collector of the City of Newark, Mr. James Morgan. Deputy Chief, Fire Department, Mr. Thomas Martin, Deputy Chief, Police Department, and Mr. Thomas DeMaio, Chief Identification Officer, Police Department. Finally, I would like to express my sincere appreciation and thanks to Mr. Francis Rudden, President, New Jersey Insurance Underwriters Association who took his valuable time to explain the insurance application procedure and the record keeping procedures of NJIUA.

Several persons of the Fire Department, including Chief Stoffers, helped me in collecting data from the above agencies and also in finding a suitable match for each structure in the arson sample. The author is very much impressed by the dedication of the following persons to this work, and appreciates their help and expresses his sincere thanks to each one of them for their help: Capt. Peter Romano, Mr. Richard Freeman, Mr. Joe Daly, and Mr. Chris Larson. Also, I would like to thank Mr. Robert Lefferts, and Mr. John Di Blasi for explaining the Fire Department records to me. Finally, I express my sincere thanks to Mr. Gerrit DeVries and Ms. Betty Williams for their help in making available the Newark computer facility to me.

Finally, the author would like to express his sincere thanks to Mrs. Conrad and Mrs. Iannucci of the Fire Department for their help in typing the report. Last but not least my most sincere thanks to Mrs. Helen Treilib of the NYS Division of Criminal Justice Services for her invaluable help in typing this report.

SUMMARY

As the title implies, the aims of this study are two-fold: The first is to develop a statistical formula, called a discriminant function, to predict structures that are likely to have arsons in the future, and the second is to determine if any pattern exists in the commission of structural arsons. Both these objectives have the same final goal, viz., to reduce or prevent structural arsons in Newark. The first of these objectives is designed to prevent arson of a specific structure while the latter is more general in nature.

A population of 897 structures in Newark was developed to study the above objectives. The population consisted of all structures in Newark that had arson for the first time during the study period January 1, 1980 to April 30, 1981. Each structure in the population may have had a subsequent arson during the study period. However, no structure had any arson, according to records, prior to its first arson during the study period. A simple random sample of 150 structures, called the arson sample, was drawn from the population. It has been established that this is a representative sample of the population with respect to the distribution of arson incidences by month of the year, by day of the week, by hour of the day, by fire districts, and the frequency of arson incidences.

A match for each structure in the arson sample was obtained, taking into account the shape and size of the structure in the arson sample, its assessed value, material used to build the structure, location (corner or non-corner) and neighborhood of the structure. In this way, 127 structures were selected that matched 127 structures of the arson sample. The remaining 23 structures, for each of which a suitable match could not be obtained, were deleted.

Data from various Newark city departments were collected for 127 cases (i.e., structures) in the arson and the match samples. On the basis of these data, an additional 25 cases were dropped from each of these samples because most of them were extreme cases (See Table 3.1).

A procedure has been suggested in Ch. II of this report to identify serious violations from non-serious violations of various types, viz., building code violations, health code violations, electrical code violations and fire code violations. Similarly, an insurance score for a structure has been developed, using information about the changes in the amount of insurance, the number of owners, and the amount of insurance loss claimed. A procedure to compute such an insurance score is given in Ch. III of this report. For developing a discriminant function, it has been observed that the insurance score is preferable to several other variables considered in this study.

Statistical analysis of the data for the arson and the match samples suggests that:

a) The distribution of "the total amount of all taxes due" differs significantly for the arson sample from that for the match sample. Moreover, the mean amount of all taxes due is significantly higher for the arson sample than the match sample. For the arson sample the mean amount is $992 and that for the match sample it is $371.

b) The ratio of the total number of non-serious violations in the arson sample to that in the match sample is 3.7 to 1

c) The ratio of the total number of non-serious violations to the similar statistic for serious violations for the arson sample is 2.1 to 1

d) The ratio of the total number of Part I crimes in a structure in the arson sample to a structure in the match sample is about 1.8 to 1

e) The ratio of the total number of Part II crimes in a structure in the arson sample to a structure in the match sample is about 1.4 to 1

f) The ratio of the total number of Part I crimes in a structure to the total number of Part II crimes for that structure in the arson sample is about 2.2 to 1

g) The mean insurance score is significantly lower for the arson sample than for the match sample. (Note that the lower the insurance score, the more likely it is that the structure will have an arson.)

From the above findings, it seems that the variables "the total amount of all taxes due," "the total number of non-serious violations," "the total number of Part I crimes in a structure," and "the insurance score" are more useful in discriminating an arson structure from a non-arson structure than many other variables. The discriminant function recommended for use of the Fire Department involves the following variables: The total amount of all taxes due, the total number of non-serious building code violations, the total number of non-serious fire code violations, the total number of Part I crimes committed in a structure, and the insurance score for a structure. This discriminant function is recommended after comparing the minimum misclassification probabilities or estimated values of these probabilities, and its relative cost efficience (RCE) with other possible discriminant functions.

We have seen above that the non-serious violations occur more frequently than the serious violations for the arson sample. Moreover, non-serious violations also occur more frequently in the arson sample than in the match sample. Because of these characteristics of the samples, the

non-serious violations have appeared in the formula for the discriminant function. In practice, however, it may happen that a structure has a serious building code or fire code violation prior to any non-serious violations. The proper use of the discriminant function in such a situation will be to use these serious violations in the formula. Since the non-serious violations occur more frequently than the serious violations the formula, in general, will alert the authorities to possible arson for a structure much in advance by considering the nonserious violations rather than the serious violations.

Using the RCE formula given in this report, comparison has been made of the recommended discriminant function to those reported for Boston and New York City. It has been shown that the Newark discriminant function, and the associated rules (DFAR), is at least 34% more cost efficient than either Boston or New York City DFAR when the ratio of the cost of arson damage to a structure to the cost of preventing it is 2. This means that for every $100 cost (due to wrong decisions) incurred by Newark, Boston will spend $137.80 and New York City will spend $134.40 for the similar decisions, assuming that the structures and all other conditions are identical for all these cities.

The patterns of arson have been studied using the available data for the entire population of 897 structures (which experienced 1123 arson incidences during the study period). Since 87% of the structures had only one arson during the study period, arson appears to be widespread in Newark. Moreover, 13% of the structures experienced 30% of the arsons during the same period which suggests that there are large proportions of arson incidences that are repeated arsons of the same structure.

There were 690 structures that experienced arson during 1980 or on an average about 58 new structures were torched every month of 1980. However, a slightly lower number of structures, viz., 50, were ignited each month of 1981 up to the end of April 1981. These 690 structures had 834 arson incidences during 1980 out of the total of 1123 arson incidences during the entire study period. This works out to about 70 arson incidences per month during 1980 and 72 arson incidences per month during the first four months of 1981. Thus, it appears that the number of arson incidences was not reduced during the year 1981.

The frequency of occurrence of the arson incidences is about the same for each month of the year and also for each day of the week. Similar patterns are also observed for repeated arsons of the same structures. However, when we group the number of arson incidences by hours, viz., 3-hour, 6-hour, 9-hour, or 12-hour periods, the difference between the number of arsons committed during the morning hours and the afternoon-evening hours is significant. In fact, there are more than twice the number of arsons committed during afternoon-evening hours than during the morning hours. This is also true when we consider these distributions either by month of the year or by days of the week. Similar patterns were also noted for repeated arsons of the same structures.

During 1980 one-third of all census tracts experienced 11 or more arsons. Furthermore, one out of every eight census tracts had more than twenty arsons each. These statistics also reflect the magnitude of structural arsons in Newark.

The five fire districts (FDs), each of which is composed of several census tracts, differ significantly in the occurrence of arson incidences. It may be remarked here that the arson incidences within each fire district are a function of the number of existing structures and the population size. The FD V has the least mean number of arson incidences per month (3.25) and FD IV has the highest number of these incidences (25.25). The mean number of arson incidences for FD V differs significantly from those for the remaining four fire districts. Similarly, the mean number of arsons incidences (23.25) for FD IV also differs significantly from those for FD I, FD II, and FD III. However, there is no significant difference between the mean number of arsons occurring per month in FD I and FD III.

A procedure for utilizing the available resources to combat arson has been given in the report using information about the number of arsons occurring within each fire district during each month of 1980. Moreover, analysis of the arson incidences by month for census tracts that had 20 or more arsons is presented, indicating how such information can be used to combat arsons.

Statistical analysis of arson incidences for lunar trends shows that 60% of the arsons take place during the lunar phases for the year 1980. However, there is practically no difference in the occurrence of arsons per day for the period of lunar phases and for the period of non-lunar phases. Thus, the data for Newark does not support the generally held view that more arsons take place during the lunar period than during the non-lunar period Moreover, although there are more full-moon days in 1980 with arson incidences than the new-moon days with such incidences, the mean number of arson incidences for these two days do not differ significantly for the year 1980.

We have noted above that the patterns for repeated arsons of the same structures are more or less similar to those for all arsons (i.e., one-time and repeated arsons combined) for the variables "month," "day," "hour," and "census tracts." We have also analyzed the data for repeated arsons by considering subsequent arsons. Analysis of subsequent arsons is done in this study in order to determine the probabilities of the occurrence of future arsons. For this purpose a table has been developed empirically, to compute the probability of a structure experiencing an arson for certain selected time periods of a year. Similarly, the limiting (ultimate) probability for the occurrence of an arson on any day of the week has been obtained empirically.

RECOMMENDATIONS

This report presents a study of structural arsons with respect to their predictability and also their patterns of occurrence. It has been shown that using the information for a structure about the amount of all taxes due, the number of non-serious building code violations, the number of non-serious fire code violations, the number of Part I (Index) crimes committed in the structure, and the insurance information, one can predict with 70% probability of correct classification, whether or not that structure will have an arson. For this purpose we use the discriminant function and the associated rules given in this report. Although this discriminant function has a slightly lower probability of predicting arson for a structure than those developed by New York City and Boston, its use is strongly recommended since the cost incurred due to wrong classification, using the recommeded discriminant function and the associated rules, is at least one-third less than the discriminant function and the associated rules given by either city. The recommended discriminant function and the associated rules should be reviewed at least once a year to ascertain that the discriminant function is appropriate for the then existing conditions and, if necessary, should be modified accordingly.

It is possible to predict more accurately an arson for a structure by:

(1) Considering, in developing a new discriminant function, variables not included in the above recommended discriminant function, such as percent of arsoned structures belonging to the owner of a structure in question;

(2) Developing separate discriminant functions for residential and non-residential structures;

(3) Developing a separate discriminant function for structures that are burnt for monetary profits only.

The methodology to be used for such improvements is similar to that described in this report.

Predictions of structural arsons on the lines suggested by Icove and Keith[1] need to be undertaken. These prediction models should prove useful in combating arson systematically. Due to non-availability of computer software packages for statistical analysis of this type and also because of the limited time available for this project, these models could not be developed.

The use of the discriminant function requires up-to-date information regarding the variables in the formula. We, therefore, recommend developing a manual information system, as was done by the New Haven (Conn.) Fire Department, or

1. Icove, D.J., and Keith, P.E. (1981): Principles of Incendiary Analysis: The arson pattern recognition system (APRS) approach to arson information management.

a limited computerized information system, as was done by the Knoxville (Tenn.) Fire Department. After establishing the usefulness of such a system for Newark, it should be expanded to include all relevant information for arson as was done by the New Haven Fire Department, which has found this system very helpful in preventing arson and in booking the arsonists.

Although there is a need to coordinate the data collection process and to develop a mechanism to collect such data on a regular basis for the above information system, there seems to be no difficulty in collecting data from all concerned agencies, except the insurance companies. Since insurance information is vital for developing an early warning system to prevent arson, it is strongly recommended that legislative measures be undertaken to collect such information from all insurance companies that are not required by the existing laws to provide the requisite information.

Since more arsons occur during the afternoon-evening hours of the day than during the morning hours, we recommend that the resources for preventing structural arsons be distributed proportionately for these hours of the day. Similarly, those census tracts with 20 or more arsons, that have different patterns of arson incidences by month than their respective fire districts, should be given due consideration in allocating resources to combat structural arsons. For this purpose it is recommended that the objective procedure suggested in this report to identify census tracts that have patterns of arsons different from their respective districts,should be used.

In this report arsons have been studied in relation to structures only. However, the arsonist is an important actor in torching a structure. Hence it is vital to know his background and be able to predict whether he is going to set fire to some other structures. It is recommended that such a prediction instrument be developed.

Statistics for the year 1980 suggest that the number of motor vehicle fires was about one-half of the structural fires. Since some of these motor vehicle fires can be for monetary profits, a study of these fire is also recommended.

TABLE OF CONTENTS

LIST OF TABLES

List of Tables (cont'd.)

List of Tables (cont'd.)

List of Tables (cont'd.)

List of Tables (cont'd.)

LIST OF FIGURES

CHAPTER I

POPULATION AND THE SAMPLES

1.1 INTRODUCTION

One of the costliest acts of violence except war is arson and it is a major and rapidly growing problem throughout the country. For example, in 1974 there were an estimated $1.3 billion in property losses, 1,000 deaths (including 45 firefighters), and 10,000 injuries from arson. This represented an increase of about 27% over the preceding decade 1964-74. The estimated property loss and the increase in number over the preceding decade exceeds those of all seven serious offenses called Index Crimes.[1] In Newark, this national trend is somewhat replicated. In 1975 there were 6,232 fires and in 1979 this figure rose to 6,603, an increase of about 6%.[2]. Of these fires, about one-fifth were arsons. The acute problem of arson in Newark becomes evident from the fact that while 41% of Essex County's population is in Newark but 95% of all the arsons in the county occur in Newark. Similarly, 5% of New Jersey State's population lives in Newark but 45% of all arsons in the state occur in Newark.

To combat arson in Newark, the Law Enforcement Assistance Administration (LEAA) provided a grant in 1979 of $200,000 to Newark Fire Department. These funds were utilized in order to increase the investigative capabilities of the Newark Arson Squad, to train investigators in detecting arsons, to improve information system capabilities so as to prevent future arsons and to increase capabilities of arson investigators in arresting and convicting arsonists.

1.2 THE STUDY

The chief aim of the present study is to develop an instrument to predict structures that are arson-prone. The definition of arson used here is a widely accepted definition viz, arson is the willful and malicious burning of another's property or the burning of one's own property for some improper purpose such as defrauding an insurer.[1]

The study period is from January 1, 1980 to April 30, 1981. This time period was selected after considering the availability of as much recent data as possible from all concerned Newark departments.

---

1. J.F. Boudreau, Q.Y. Kwan, W.F. Faragher, and G.C. Denault (1977): Arson and Arson Investigation - Survey and Assessment, NILECJ, Law Enforcement Assistance, U.S. Department of Justice.

2. Alan Zalkind (1980): The Arson problem in Newark. Newark Arson Squad Newsletter.

1.3. THE POPULATION

The study population included only structures and not motor vehicles, or arsons outside any structure. These structures may be used for dwelling,commercial, or industrial purposes. Although a garage may not be physically attached to a house, it is considered a part of the house for the study. Similarly, a duplex or a structure with multiple dwelling units divided by a wall from the basement to the roof of the building, is considered to be composed of two or more separete buildings, depending upon the number of walls separating them. As with biological twins such buildings are most ideal as a sample match for the present study.

Newark Fire Department's "Summary of Daily Fires" (SDF) files were used to generate the population of arsoned structures. As a first step, structures experiencing arson for the first time during the study period January 1, 1980 to April 30, 1981, were identified. Such structures were identified through the B.I. Signal Codes given for each incidence of fires in SDF files. (For arson incidence the B.I. Signal Codes used are 100, 102, or 200). This procedure generated a population of buildings, some of which had arson for the first time during the study period with no arson before this period, and the rest that had arsons both before and during the study period. The next step in the procedure, therefore. involved identifying structures that had arsons before the study period and deleting those structures from our study. This then generated the population of 897 structures that experienced arsons for the first time during the study period and none before that period.

1.3.1. CHARACTERISTICS OF THE POPULATION

Within the population of 897 structures which experienced arson for the first time during the priod of January 1, 1980 to April 30, 1981, 141 or 15.7% of the structures experienced two or more arsons during the study period. The distribution of 897 structures by the number of arsons for these structures, during the study period, is given in the following table.

TABLE 1.0: Distribution of the population of structures by number of arsons experienced during the study period.

Number of Arsons	Structures			Arson Incidences		
	Number	Percent	Cumulative Percent	Total Number	Percent	Cumulative Percent
1	756	84.3	84.3	756	67.3	67.3
2	102	11.4	95.7	204	18.2	85.5
3	19	2.1	97.8	57	5.1	90.6
4	11	1.2	99.0	44	3.9	94.5
5	4	0.5	99.5	20	1.8	96.3
6	2	0.2	99.7	12	1.1	97.4
7	1	0.1	99.8	7	0.6	98.0
10	1	0.1	99.9	10	0.8	98.8
13	1	0.1	100.0	13	1.2	100.0
Total	897	100.0	-	1,123	100.0	-

We notice from the above table that four out of five structures experienced only one arson during the sixteen month study period and one out of nine had two arsons during the same period. If we look at the distribution of 1,123 arsons, we find that two-thirds of the arsons are accounted for by 756 structures that experienced only one arson each. The remaining 33% of the arsons are shared by 16% of the structures in the population that had two or more arsons each during the study period.

The population of 897 structures is distributed among 85 of the Newark City's 98 census tracts. Thus, none of the structures in about 13% of the census tracks experienced any arson during the study period. An additional 48% of the census tracts had occurrence of 1 to 9 arsons and 10 or more arsons occurred in each of the remaining 39% of the census tracts. The highest number of arsons, viz, 38, occurred in census tract #92. The following table shows the distribution of the number of census tracts by the number of arsons which occurred during the study period.

TABLE 1.1: Distribution of the number of census tracts by number of arson incidences which occurred during the period January 1, 1980 to April 30, 1981.

Number of Arsons	Census Tracts Number	Census Tracts Cumulative	Number of Arsons	Census Tracts Number	Census Tracts Cumulative
0	13	13	13	5	73
1	4	17	14	3	76
2	2	19	15	1	77
3	8	27	16	2	79
4	5	32	17	5	84
5	8	40	18	1	85
6	5	45	19	2	87
7	7	52	21	2	89
8	4	56	22	1	90
9	5	61	24	1	91
10	1	62	25	4	95
11	3	65	29	1	96
12	3	68	33	1	97
			38		98

TABLE 1.2: Distribution of census tracts, grouped by number of arson incidences.

Number of Arsons	Census Tract Number	Census Tract Cumulative	Percent to the Total	Cumulative Percent
0	13	13	13.3	13.3
1- 5	27	40	27.6	40.9
6-10	22	62	22.4	63.3
11-15	15	77	15.3	78.6
16-20	11	88	11.2	89.8
21-34	10	98	10.2	100.0
--	98	-	100.0	-

Table 1.2 is derived from Table 1.1 and groups the number of census tracts by intervals of five for number of arson incidences. We notice that in more than one-fifth of the census tracts the number of arson incidences was more than 15 with almost 10% of the census tracts experiencing more than 20 arsons each during the study period.

Besides analyzing the information by census tracts it is useful to do such analysis by fire districts for Newark. The following table shows the composition of each of the five districts by census tracts. (It may be noted that census tract nos. 48 and 75 are further subdivided into 48.01, 48.02 and 75.01 and 75.02, resulting in the total of 100 census tracts.)

TABLE 1.3: Composition of fire districts by census tracts.

Fire District Number	Census Tract Number	Total Number of Census Tracts
I	9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 29, 64, 82, 83, and 84	23
II	48.02, 57, 59, 67, 80, 81, and 85	7
III	1, 2, 3, 4, 5, 6, 7, 8, 86, 87, 88, 89, 90 91, 92, 93, 94, 95, 96, and 97	20
IV	27, 28, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48.01, 49, 50, 51, 52, 53, 54, 55, 56, 58, 60, 61, 62, 63, 65, and 66	36
V	68, 69, 70, 71, 72, 73, 74, 75.01, 75.02, 76, 77, 78, 79, and 98	14

We notice that census tracts are unevenly distributed to the fire districts. This is due to the fact that the census tracts themselves are of uneven size. The area covered by each fire district, however, is not markedly dissimilar as the above distribution suggests.

The frequency distribution of the arsoned structures in the population by fire districts is given in Table 1.4. We notice that there are significant differences in the number of arsoned structures in these districts. It should be recognized, however, that the number of structures arsoned in a district is a function of the number of existing structures. Thus, District IV which is more crowded with structures , is likely to have more arsons than District V which has relatively fewer structures.

TABLE 1.4: Distribution of arsoned structures in the population by fire districts.

Fire District	Frequency	Percent	Cumulative Percent
I	190	21.2	21.2
II	71	7.9	29.1
III	210	23.4	52.5
IV	364	40.6	93.1
V	62	6.9	100.0
Total	897	100.0	-

We now consider the distribution of arson incidences by month of the year. Since the study period is from January 1, 1980 to April 30, 1981, the distribution of the arson incidences is presented separately for the years 1980 and 1981 in Table 1.5 below.

TABLE 1.5: Distribution of arson incidences in the population by month for the study period.

Month	Frequency	Percent	Cumulative Percent
Year 1980			
January	40	3.6	3.6
February	72	6.4	10.0
March	82	7.3	17.3
April	69	6.1	23.4
May	88	7.8	31.2
June	77	6.8	38.0
July	67	6.0	44.0
August	62	5.5	49.5
September	66	5.9	55.4
October	59	5.2	60.6
November	76	6.8	67.4
December	76	6.8	74.2
Year 1981			
January	64	5.7	79.9
February	60	5.4	85.3
March	81	7.2	92.5
April	84	7.5	100.0
Total	1,123	100.0	-

The lowest number of arson incidences occurred during the month of January, 1980. However, the month of January did not have this distinction for 1981. Of the 834 arson incidences during 1980, about 56.5% took place during the months of February, March, May, June, November and December. Assuming that this is a sample from a super-population (see the discussion in the next paragraph) we find that there are significant differences in the number of arsons during these months of 1980 and the remaining months of 1980 (χ^2=13.986; 1 d.f.)[1]

For the present study, we will assume that the population of structures defined here, is itself a sample from a super-population. This assumption is necessary to apply any statistical test to the data. The purpose of applying a statistical test is to derive information and conclusions from the data which otherwise may not be apparent. Since it is costly to collect data, it is evident that the most economical use calls for extracting as much information as possible from the data. One of the ways to do this is to make the above type of assumption for the population whenever a super-population comprising the population under study exists. The super-population in this case is the accumulation of all structures that experienced arson for the first time till the end of April 30, 1981. Note that this super-population includes the following structures that experienced arson for the first time before the study period and

(I) did not experience any fire or arson at all;
(II) did experience fire but no arson;
(III) did experience an arson after the study period; or
(IV) were demolished after the first arson.

In addition, the super-population includes the population of 897 structures defined above. Thus, in our case, we see that a super-population does exist.

We now consider the distribution of arson incidences by month and by hour of the day which is given in the following table, Table 1.6. At a first glance, it becomes evident that the arson incidences are almost evenly distributed by hour. There is almost no correlation between the month and the hour of incidence, (r=0.001). We, therefore, group the data by three-hour periods as shown in Table 1.7.

1 "d.f." means degrees of freedom.

TABLE 1.6: Distribution of arson incidences in the population by hour of the day and by month for the study period.

Hour	Month and Year																Total
	Year 1980												Year 1981				
	Jan	Feb	Mar	Apr	May	June	July	Aug	Sept	Oct	Nov	Dec	Jan	Feb	Mar	Apr	
0- 1	2	3	0	3	3	3	2	5	2	3	3	5	5	5	7	5	56
1- 2	1	3	1	3	5	3	3	4	0	2	4	3	1	3	4	2	42
2- 3	1	6	7	6	1	1	3	3	2	1	1	2	6	2	1	2	45
3- 4	0	2	5	1	2	7	2	1	2	3	2	5	2	3	0	5	42
4- 5	0	3	1	1	2	2	4	2	1	2	2	3	2	2	2	3	32
5- 6	2	0	3	1	2	1	1	0	1	2	1	2	2	2	2	1	23
6- 7	0	1	0	1	1	1	3	2	0	0	1	0	2	1	1	3	17
7- 8	1	1	0	0	1	0	1	0	0	1	2	1	1	1	1	2	13
8- 9	0	1	2	1	1	0	1	1	1	2	0	3	1	1	1	3	19
9-10	1	0	4	0	3	4	1	2	3	3	2	4	4	2	2	0	35
10-11	4	3	1	1	0	1	0	0	3	4	3	1	0	0	4	3	28
11-12	2	1	1	2	4	1	5	1	0	2	2	3	2	2	1	2	31
12-13	2	1	2	3	2	3	1	1	3	4	4	1	2	3	2	2	36
13-14	4	3	7	5	4	0	3	2	0	4	6	7	3	5	2	1	56
14-15	1	4	6	1	7	3	3	1	3	2	6	3	3	1	5	4	53
15-16	1	2	6	1	5	10	5	4	7	1	3	3	4	2	6	4	64
16-17	1	6	3	8	8	0	4	2	3	3	3	7	2	2	6	3	61
17-18	3	1	7	7	7	5	1	1	4	2	4	3	1	4	3	0	53
18-19	1	5	5	3	7	4	6	7	5	3	6	4	2	5	8	6	77
19-20	2	4	5	5	3	4	5	3	6	4	3	3	3	2	4	6	62
20-21	0	5	2	2	5	8	3	8	1	3	4	4	5	4	5	6	65
21-22	5	5	5	7	5	6	3	7	4	3	4	4	4	4	7	8	81
22-23	5	9	4	5	4	4	4	4	8	3	7	3	5	3	5	10	83
23-24	1	3	5	2	6	6	3	1	7	2	3	2	2	1	2	3	49
Total	40	72	82	69	88	77	67	62	66	59	76	76	64	60	81	84	1,123

TABLE 1.7: Distribution of arson incidences in the population by three-hour periods and by month for the study period.

Month	Hours of the Day								Total
	0-3	3-6	6-9	9-12	12-15	15-18	18-21	21-24	
Year 1980									
January	4	2	1	7	7	5	3	11	40
February	12	5	3	4	8	9	14	17	72
March	8	9	2	6	15	16	12	14	82
April	12	3	2	3	9	16	10	14	69
May	9	6	3	7	13	20	15	15	88
June	7	10	1	6	6	15	16	16	77
July	8	7	5	6	7	10	14	10	67
August	12	3	3	3	4	7	18	12	62
September	4	4	1	6	6	14	12	19	66
October	6	7	3	9	10	6	10	8	59
November	8	5	3	7	16	10	13	14	76
December	10	10	4	8	11	13	11	9	76
Year 1981									
January	12	6	4	6	8	7	10	11	64
February	10	7	3	4	9	8	11	8	60
March	12	4	3	7	9	15	17	14	81
April	9	9	8	5	7	7	18	21	84
Total	143	97	49	94	145	178	204	213	1,123

We notice from the above table that arson activity is low during the nine hour period 3-12 consistently throughout the year. Moreover, throughout the year much higher number of arsons take place during the hours 12-24 than during the time period 0-12 hours. The following table shows that for most of the months this increase far exceeds 50%. It is also evident that for the first four months of 1981, the trend somewhat changed with more arsons occurring during the hours 0-12 than the corresponding period for 1980. We will analyze this data further in Chapter IV.

TABLE 1.8: Monthly distribution of arson incidences for the 12-hour time periods 0-12 and 12-24.

Month	Number of Arson Incidences: Hours 0-12	Hours 12-24	Total	Percent of Col. (3) to Col. (2)
1	2	3	4	5
Year 1980				
January	14	26	40	185.7
February	24	48	72	200.0
March	25	57	82	228.0
April	20	49	69	245.0
May	25	63	88	252.0
June	24	53	77	220.8
July	26	41	67	157.7
August	21	41	62	195.2
September	15	51	66	340.0
October	25	34	59	136.0
November	23	53	76	230.4
December	32	44	76	137.5
Year 1981				
January	28	36	64	128.6
February	24	36	60	150.0
March	26	55	81	211.5
April	31	53	84	171.0
Total	383	740	1,123	-

Consider now the distribution of arson incidences by day of the week for the sixteen month study period as shown in the following table.

TABLE 1.9: Distribution of arson incidences in the population by day of the week.

Day	Frequency	Percent	Cumulative Percent
Sunday	158	14.1	14.1
Monday	182	16.2	30.3
Tuesday	155	13.8	44.1
Wednesday	141	12.6	56.7
Thursday	170	15.1	71.8
Friday	163	14.5	86.3
Saturday	154	13.7	100.0
Total	1,123	100.0	--

It is clear from the above table that the arson incidences are distributed more or less evenly throughout the week with a maximum of 16% taking place on Mondays. The Chi-Square test also confirms this observation ($\chi^2 = 6.302$ on 6 d.f.). Also, we note from this table that on an average 16% of the arsons occurred on Mondays and about 13% on Wednesdays. For other days of the week the percent of arson incidences were in the range of 13-16%. In the sample population, more arsons occurred on Mondays, Thursdays and Fridays in that order than on any other day of the week. Furthermore, there is only a marginal difference in the occurrence of arson incidences between Thursdays and Fridays on the one hand, and Saturdays and Sundays on the other. We observe somewhat the same pattern when we compare the number of arson incidences occurring on Mondays and Thursdays on the one hand and on Saturdays and Sundays on the other. Thus the data do not suggest that more arsons occur on weekends than on any other day of the week.

The 1,123 incidences in the sample population are, however, not distributed evenly for the 24 hours of the day. As can be seen from Table 1.10 below, more incidences take place during the hours 15-23 about 48.6%) than during any other time period.

TABLE 1.10: Distribution of arson incidences in the population by hour of the day.

Hour	Frequency	Percent	Cumulative Percent	Hour	Frequency	Percent	Cumulative Percent
0- 1	56	5.0	5.0	12-13	36	3.2	37.3
1- 2	42	3.7	8.7	13-14	56	5.0	42.3
2- 3	45	4.0	12.7	14-15	53	4.7	47.0
3- 4	42	3.7	16.4	15-16	64	5.7	52.7
4- 5	32	2.9	19.2	16-17	61	5.4	58.1
5- 6	23	2.0	21.3	17-18	53	4.7	62.8
6- 7	17	1.5	22.8	18-19	77	6.9	69.7
7- 8	13	1.2	24.0	19-20	62	5.5	75.2
8- 9	19	1.7	25.7	20-21	65	5.8	81.0
9-10	35	3.1	28.8	21-22	81	7.2	88.2
10-11	28	2.5	31.3	22-23	83	7.4	95.6
11-12	31	2.8	34.1	23-24	49	4.4	100.0
Total	-	-	-	-	1,123	100.0	-

TABLE 1.11: Distribution of arson incidences in the population by three-hour time periods of the day.

Time Interval	Frequency	Percent	Cumulative Percent
0- 3	143	12.7	12.7
3- 6	97	8.6	21.3
6- 9	49	4.4	25.7
9-12	94	8.4	34.1
12-15	145	12.9	47.0
15-18	178	15.8	62.8
18-21	204	18.2	81.0
21-24	213	19.0	100.0
Total	1,123	100.0	-

A look at Table 1.11 shows that the number of arson incidences taking place decreases from midnight till 9 AM and goes on increasing from 9 AM till the midnight of the next day. Moreover, almost two and a-half times arsons take place in the 9 hour period 15-24 hours than during the period 3-12 hours.

TABLE 1.12: Distribution of arson incidences in the population by hour of the day and by day of the week for the study period.

Hour	Day of the Week							Total
	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	
0- 1	10	9	8	8	7	5	9	56
1- 2	3	8	5	3	5	7	11	42
2- 3	2	6	11	5	3	9	9	45
3- 4	7	4	6	8	3	8	6	42
4- 5	6	3	4	3	7	6	3	32
5- 6	6	1	5	1	3	1	6	23
6- 7	3	2	4	1	2	2	3	17
7- 8	1	2	0	2	5	2	1	13
8- 9	6	2	2	2	5	2	0	19
9-10	3	5	6	5	6	7	3	35
10-11	5	4	2	3	6	6	2	28
11-12	4	6	7	4	7	0	3	31
12-13	2	10	2	7	8	3	4	36
13-14	6	12	8	5	7	10	8	56
14-15	9	10	6	6	6	13	3	53
15-16	8	8	3	12	11	15	7	64
16-17	7	8	7	5	12	9	13	61
17-18	9	8	10	8	4	7	7	53
18-19	11	14	10	12	14	6	10	77
19-20	14	6	10	7	10	8	7	62
20-21	4	13	11	10	12	8	7	65
21-22	14	13	10	11	12	12	9	81
22-23	10	18	12	7	8	11	17	83
23-24	8	10	6	6	7	6	6	49
Total	158	182	155	141	170	163	154	1,123

From the above table, Table 1.12, we notice the afternoon and evening hours of Mondays, Tuesdays, Thursdays and Fridays experience more arsons generally than the remaining three days of the week. Also, Mondays experience almost three times arson during the hours 12-24 than during the hours 0-12.

TABLE 1.13: Distribution of arson incidences in the population by three-hour time period of the day and by day of the week.

Time Period	Day of the Week							Total
	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	
0- 3	15	23	24	16	15	21	29	143
3- 6	19	8	15	12	13	15	15	97
6- 9	10	6	6	5	12	6	4	49
9-12	12	15	15	12	19	13	8	94
12-15	17	32	16	18	21	26	15	145
15-18	24	24	20	25	27	31	27	178
18-21	29	33	31	29	36	22	34	204
21-24	32	41	28	24	27	29	32	213
Total	158	182	155	141	170	163	154	1,123

What we have observed in Table 1.11 for the three-hour period about the occurrence of arson incidences is also true for days of the week. (See Table 1.13.) That is, except for Sunday, the number of arsons, decrease from midnight till 9 AM and then increase from 9 AM till midnight of the next day. Furthermore, there are almost three times more arsons during the 6-hour period 6 PM - 12 Midnight than the 6-hour period 6 AM - 12 Noon, as shown in the following table, Table 1.14.

TABLE 1.14: Distribution of arson incidences during the 6-hour time periods 6-12 and 18-24 and by day of the week.

Time Period	Day of the Week							Total
	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	
6-12	22	21	21	17	31	19	12	143
18-24	61	74	59	53	63	51	56	417
Percent*	272.3	352.4	281.0	311.8	203.2	268.4	466.7	291.6

*Percent of arson incidences during 18-24 hours as compared to 6-12 hours.

Although the arson incidences increase sharply each day during the period 6 PM - 12 Midnight over the 6 AM - 12 Noon period such increases are still more sharp on Saturdays, Mondays and Wednesdays, in that order.

Consider now the 9-hour periods 3-12 and 15-24 hours. Then the distribution looks as follows:

TABLE 1.15: Distribution of arson incidences for the 9-hour periods and by day of the week.

Time Period	Day of the Week							Total
	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	
3-12	41	29	36	29	44	34	27	240
15-24	85	98	79	78	90	82	83	595
Percent*	207.3	337.9	219.4	269.0	204.5	241.2	307.4	247.9

*Percent of arson incidences during 15-24 hours as compared to 3-12 hours

We find that the sharpest increase still occurs on Mondays, Saturdays, and Wednesdays, in that order. Thus, we conclude that during the 6-hour or the 9-hour periods considered above, at least twice the number of arsons occur during the late hour periods than during the earlier time periods of the day. Moreover, such increases range from two and a-half to three and a-half times as great fpr Mondays, Wednesdays, and Saturdays.

1.4. SAMPLE

From the population of 897 structures, a sample of 150 structures (16.7%) was selected using a simple random sampling without replacement scheme. If any of the 150 addresses initially selected in the sample represented vacant lands, as per information collected from the Newark City Department of Tax Collection and Assessment, these were replaced by the next unsampled member of the population. We will study below characteristics of this sample to establish its representativeness of the population. As will be discussed later, a sample of slightly reduced size is used to develop an instrument (i.e., an index) to predict future arson.

1.4.1. CHARACTERISTICS OF THE SAMPLE

The characteristics of the sample that will be examined here include the month, the day, and the hour of each arson incident and the fire districts in which the selected structures are located.

As we have seen earlier, there are 1,123 arson incidences during the study period for 897 structures in the population or about 1.25 arson incidences per structure in the population. In the sample, a slightly higher number of arson incidences, 1.36, took place than in the population, the total number of arson incidences for the sample being 204.

The distribution of 150 structures in the samples by the number of arsons is given in Table 1.16. Also, shown in the table is the distribution of 204 arson incidences for these 150 sampled structures.

TABLE 1.16: Distribution of the sampled structures by the number of arsons experienced during the study period.

Number of Arsons	Arsoned Structures			Arson Incidences		
	Total Number	Percent	Cumulative Percent	Total Number	Percent	Cumulative Percent
1	120	80.0	80.3	120	58.8	58.8
2	17	11.3	91.3	34	16.7	75.5
3	6	4.0	95.3	18	8.8	84.3
4	4	2.7	98.0	16	7.8	92.1
5	2	1.3	99.3	10	4.9	97.0
6	1	0.7	100.0	6	3.0	100.0
Total	150	100.0	-	204	100.0	-

From Tables 1.0 and 1.16 we find, using the Kolmogrov-Smirnov test, that there is no significant difference between the percent of arsoned structures in the sample and the population. However, the sample and the population distribution differ significantly, at the 5% level of significance, for the number of arson incidences. Thus, although the sample is representative of the population for the distribution of arsoned structures, as above, it is not so for the arson incidences. This is to be expected since the variable "number of arson incidences" was not used to select the sample. In fact, neither of those variables was used to select the sample which makes the representativeness of the sample for the distribution of arsoned structure more encouraging. It may be remarked here that generally the representativeness of the sample on the basis of the distribution of arsoned structure is more pertinent and sought for than the distribution for the number of arson incidences.

In what follows, we compare the sample distributions with the corresponding population distributions for the variables month, day, hour, and the five districts of arson incidence.

TABLE 1.17: Distribution of arson incidences in the sample, by month for the period.

Month	Frequency	Percent	Cumulative Percent
Year 1980			
January	6	2.94	2.94
February	12	5.88	8.82
March	20	9.80	18.62
April	14	6.87	25.49
May	23	11.27	36.76
June	13	6.37	43.13
July	14	6.87	50.00
August	11	5.39	55.39
September	11	5.39	60.78
October	16	7.85	68.63
November	10	4.90	73.53
December	13	6.37	79.90
Year 1981			
January	12	5.88	85.78
Februray	3	1.48	87.26
March	13	6.37	93.62
April	13	6.38	100.00
Total	204	100.00	-

These 204 arson incidences are distributed by month as shown in the above table. As in the case of the population, we have chosen to present for all the sixteen months of the study period rather than a twelve month period, to test the hypothesis that the above distribution for the arson incidences in the sample is not different from the corresponding distribution for the population. Using the Kolmogrov-Smirnov test, we find that the observed value of D, the Kolmogrov-Smirnov statistic, is D = 0.0803 and the table (theoretical) value is D = 0.0952 at the 5% level of significance. Thus, we conclude that there is no significant difference between the two distributions.

We now consider the distribution of arson incidences in the sample by day of the week, and it is as shown in the following table.

TABLE 1.18: Distribution of arson incidences in the sample by day of the week.

Day	Frequency	Percent	Cumulative Percent
Sunday	24	11.76	11.76
Monday	29	14.22	25.98
Tuesday	30	14.71	40.69
Wednesday	29	14.22	54.90
Thursday	26	12.75	67.65
Friday	34	16.67	84.31
Saturday	32	15.69	100.00
Total	204	100.00	-

Using the Kolmogrov-Smirnov test, we find that the observed value of the test statistics $D = 0.0432$ is much less than the table value of $D_{table} = 0.0952$. Thus, we conclude that there is no significant difference between the population and the sample so far as the distribution of arson incidences by day is concerned.

Comparing the hourly distribution of arson incidences in the sample (shown in the following table) with the corresponding distribution for the population, we find once again that there is no significant difference between the population and the sample distributions for these characteristics. (Here $D = 0.0603$.)

TABLE 1.19: Distribution of arson incidences in the sample by hour of the day.

Hour	Frequency	Percent	Cumulative Percent	Hour	Frequency	Percent	Cumulative Percent
0-1	9	4.41	4.41	12-13	7	3.43	33.33
1-2	7	3.43	7.84	13-14	6	2.94	36.27
2-3	11	5.39	13.24	14-15	13	6.37	42.65
3-4	5	2.45	15.67	15-16	14	6.86	49.51
4-5	4	1.96	17.65	16-17	10	4.90	54.41
5-6	2	0.98	18.63	17-18	10	4.90	59.31
6-7	3	1.47	20.10	18-19	16	7.84	67.16
7-8	5	2.45	22.55	19-20	13	6.37	73.53
8-9	0	0.0	22.55	20-21	10	4.90	78.43
9-10	7	3.43	25.98	21-22	13	6.37	84.80
10-11	4	1.96	27.94	22-23	16	7.84	92.65
11-12	4	1.96	29.90	23-24	15	7.35	100.00
Total	-	-	-	-	204	100.00	-

Finally, we consider the distribution of arson incidences by the five fire districts (FD) and it is as shown in Table 1.20. In the sample

TABLE 1.20: Distribution of arson incidences in the sample by fire districts.

Fire Districts	Number of Arsons	Percent	Cumulative Percent
I	43	21.08	21.08
II	30	14.70	35.78
III	42	20.59	56.37
IV	78	38.24	94.61
V	11	5.39	100.00
Total	204	100.00	-

the largest number of arson incidences are for FD IV and the least number of these incidences for FD V. Visual comparison of the distribution of arson incidences in the sample and the population (see Table 1.4), shows that these distribution are quite similar. The Kilomgrove-Smirnov test also confirms this.

Thus, we have shown that the distribution of the sample does not differ significantly, in the statistical sense, for the variables, "number of arsons", "number of arson incidences by month", "number of arson incidences by day", "number of arson incidences by hour", and "the number of arson incidences by fire districts". Hence the selected sample is representative of the population for these variables.

1.5 MATCH SAMPLE

In many sociological studies such as the present one, it is difficult to generate a control group in the strict sense. What a researcher usually accomplishes is to identify a group of subjects (in our case structures) with characteristics that resemble, individually or collectively, some known characteristics of those in the experimental group and presumable differ with the experimental group on variables under study. Since these units are purposely selected to match the units in the experimental group, we prefer here to label this group as a matched sample rather than a control group to emphasize that the matched group does not meet the strict conditions required for a control group in a laboratory condition.

For each arsoned structure in the sample a possible match was developed using the following building characteristics: its assessed value, building use, vicinity of the structure to be matched to the one that is in the arson sample, and corner location or not. This information is available from the Essex County Real Estate Directory for 1980-81. Each possible match obtained from the directory was later verified through site visits. Such visits helped to determine more accurately, both for arson and match structures, location of the structure, purpose for which the structure was actually used, building material used, and structure's size and shape. With this additional and accurate information, a match for an arson sample structure was accepted, provided it had no previous arson history. The "previous period" for a match is the same as that for the corresponding arsoned structure and includes all the period before the occurrence of the first arson.

In a few cases the assessed value was ignored for matching purposes because these values, as given in the Essex County Real Estate Directory differed significantly for both the structures even though they were similar in all other respects. Besides this, the structure use mentioned in the Essex County Real Estate Directory for 1980-81 also differed considerably from the actual use of the structure.

Information for 150 arson cases shows that in 69 cases (about 46%), the information given in the directory did not reflect the reality. The following table gives a breakdown of 150 arson sample buildings by their actual use and as given in the directory.

TABLE 1.21: Actual building use and the use as given in Essex County Real Estate Directory for 1980-81 for sampled arson structures.

Actual Building Use	Building use as In Real Estate Directory 1980-81							Total
	General*	Family Res.@	Apt. Res.@	Commercial	Industrial	Blank	Vacant	
General*	-	-	-	-	-	-	-	3
Family Res.@	-	26	2	6	-	3	-	41
Apt. Res.@	-	35	25	6	-	15	-	81
Commercial	-	1	-	9	-	4	-	14
Industrial	-	1	-	1	4	3	1	10
Blank	-	-	-	-	-	-	1	1
Total	-	63	27	22	4	32	2	150

* General use is educational, or public assembly building
@ Res. means residential.

In one-fifth of the cases there was no information about the building use in the Real Estate Directory. This is largely because these represented properties that are tax exempt. Of the remaining 118 buildings, information for only 64 cases is accurate. Thus ignoring the blanks, building use information for about 46% of the cases is inaccurate. We must be careful, therefore in accepting the information about the building use from the directory. Because of this, a match was obtained for each arson building through site visits. The procedure of matching described earlier resulted in the deletion of 23 arson structures for want of a suitable match for them. A proper match could not be found for these buildings either because of the peculiar shape and/or use of the building; or no other non-arsoned building was available for a match in its neighborhood.

CHAPTER II

DATA COLLECTION

2.1 INTRODUCTION

The data for the study have been collected from some of the Newark City departments. These departments are: the Fire Department, the Code Enforcement Department, the Electrical Department, the Water Department, Department of Tax Collection and Assessment and the Police Department. Also data have been collected from the New Jersey Insurance Underwriters' Association (NJIUA). In what follows we discuss the procedures used to collect information from each one of these agencies except the Electrical Department, which is part of the Code Enforcement Department.

To collect data from these sources it was necessary to determine the time period, called previous time period, for which data should be collected. The previous time period is defined here as the length of time which has as its most recent point in time the date of first arson (for an arson case and its match) and which has as its earliest point in time a date prior to the date of first arson. In order to develop a discriminant function, information just sufficient to develop the function was collected from each department. This required defining different previous time periods for collecting information from various departments.

2.2 FIRE DEPARTMENT

Information about the total number of fires before the date of first arson (and the same date for respective matched case), and the total number of fire code violations was collected from the Fire Department. The period since the beginning of 1978 till the date of first arson was defined as the previous time period for collecting information about the number of previous fires.

For each structure experiencing a fire, date and time of each fire incidence, B.I. Signal, name of the inspector, and structure use and size is given on a card, called a "White Card". This information is available on white card for each structure since November 1977. However, since information about fires due to unattended cooking incinerators, compactors and short-circuits is not available on white cards, it was collected from the Daily Summary of Fire for previous time periods between January 1, 1978 and April 30, 1981.

The information regarding fire code violations was obtained from registers maintained separately by the Fire Department for this purpose. One of these registers is only for smoke detectors while the other includes all types of fire code violations. Information regarding fire code violations has been collected for the previous time period between January 1, 1979 and April 30, 1981. Along with the information for fire code violations, information about building code, health code and electrical code violations, if any, was also collected. To avoid counting

the same violation more than once, the procedure adopted was as follows : When notices of such violations from the Fire Department appeared in the files of other departments, then these violations were excluded from considerations. This procedure of counting each violation once and only once was adopted uniformly for all departments.

The information from the Daily Summary of Fires is transcribed on "White Cards" each day, so that information about the number of previous fires appears to be complete. However, no check has been made of this since it involves going through the Daily Summary of Fires for the previous time period between January 1, 1978 - April 30, 1981, a task which is both time consuming and costly. Furthermore, an index of predictions developed with the available data will be more conservative than one obtained using the most accurate information about the total number of fires, because this number will invariably be greater than the one actually used.

The information about fire code violations is maintained by the Fire Department in a register that is in alphabetical order by street addresses. A similar register is also maintained by the Fire Department for smoke detector violations. Both these registers appear to be complete and accurate since these violations are registered only after the inspection of a structure is made.

2.3.. CODE ENFORCEMENT DEPARTMENT

This department maintains information about building code violations, health code violations, and electrical code violations, through notices issued to the owners of the structures. While all building code and health code violations are registered by this department, most electrical code violations and fire code violations observed during inspection are also recorded. For some of the properties included in the study electrical code violations have been notified to the Electrical Department. However, there appears to be no consistent procedure used by the Code Enforcement Department in this regard. Similar observations were made about notifying fire code violations to the Fire Department. Since only few cases have been observed where electrical or fire code violations were reported to the respective department, multiple counting of these violations posed no serious problems. Thus, from the Code Enforcement Department information about building code violations, health code violations, electrical code violations, and fire code violations was collected. Since information from the Fire Department was collected before collecting information from the Code Enforcement Department, those violations for which notices were issued to the Fire Department by the Code Enforcement Department were excluded from consideration. This information pertained to both abated and unabated violations.

In the Code Enforcement Department files are kept in alphabetical order by address of each structure that has any of the above type of violations. In many cases the history of violations goes back to 1974 , and all these violations were considered because the information could be collected without much additional cost. Roughly about half of the structures studied (both arsoned and matched) had no record of any kind of violations. A careful search of the files showed that this was indeed the case. The files appear to be remarkably well kept and up-to-date.

As stated above, information for both abated and unabated violations was collected. A more detailed information about these violations was obtained by considering separately serious and non-serious violations. In collecting information separately for serious and non-serious violations, it is assumed that structures experiencing arson are likely to have, statistically speaking, a significantly large number of serious violations than those structures that did not experience arson. We will test this assumption later in Chapter III.

Since a multitude and varied types of serious and non-serious violations can occur in practice, it is impossible to prepare a list of such violations. In what follows, we discuss in some detail a procedure used to determine a serious violation. To describe the procedure, we have selected a few violations reported in Code Enforcement files, which are given in Table 2.1 below.

In example (1), the violation of a broken window is coded serious as a building code violation and also as a health code violation. This is because the violation took place in the winter season. On the other hand, when the violation occurs in summer, as in example (23), the building code violation is treated as non-serious.

Although providing a painted apartment is essential, it cannot be treated by itself as a serious violation. However, since old paint on the walls is often a hazard to the health, especially to small children, it has been treated as a serious health code violation in example (2).

In example (3), since the lights have been provided for the apartment, and since the nature of the defect is not stated as in example (8), it is treated as a non-serious building code violation. A similar justification can be made about treating the violation in example (9) as non-serious.

Thus, to treat a violation as serious or non-serious, we examine the nature of the violation. Moreover, where violation of a health code is involved, the season of the year should also be taken into account to classify the violation in question as serious or non-serious. Once all such violations are properly classified for a structure then we add them to arrive at the total number of serious and non-serious violations for the structure. It is possible to classify, with the help of the Code Enforcement Department, some of the commonly occurring violations as serious violations or non-serious violations so as to increase uniform classification of these violations.

TABLE 2.1: Examples of violations treated as serious or non-serious for each type of cods violations.

Description of Violation	Building Code Serious	Building Code Non-Serious	Health Code Serious	Health Code Non-Serious	Fire Code Serious	Fire Code Non-Serious	Electrical Code Serious	Electrical Code Non-Serious
1. Reinforce all windows in the apartment to prevent air from coming in.	X		X					
2. Scrape and paint entire apartment sidewalls and ceilings.		X	X					
3. Repair defective light in front bedroom.		X			X		X	
4. Cover exposed wires on sidewalls in rear bedroom.	X				X		X	
5. Remove all trash, rubbish, and refuse in rear of the dwelling.		X	X		X			
6. Gutters: repair on right and rear side of the dwelling.		X		X				
7. Repair or replace defective sidewalls and ceiling back porch 1st throughout 3rd floors, also scrape and paint.	X		X					
8. Provide proper covering for lights in ceiling to protect electrical fixtures.	X				X		X	
9. First bedroom: repair defective switch for the ceiling light.		X				X		X

(continued)

TABLE 2.1 (cont'd.): Examples of violations treated as serious or non-serious for each type of code violation.

Description of Violation	Building Code Serious	Building Code Non-Serious	Health Code Serious	Health Code Non-Serious	Fire Code Serious	Fire Code Non-Serious	Electrical Code Serious	Electrical Code Non-Serious
10. Bathroom: repair defective ceiling, scrape and paint.		X	X					
11. Repair defective door knobs to front entrance door.		X						
12. Repair defective frame to secure door for side entrance.		X						
13. Repair or replace defective heater.		X	X		X			
14. Repair defective toilet flush box to prevent water from leaking.		X	X					
15. Exterminate entire building to prevent infestation.	X		X					
16. Remove all trash, rubbish and refuse from basement.	X		X		X			
17. Repair defective staircase leading to basement.		X				X		
18. Provide a proper light in ceiling for illumination.	X				X			X
19. Paint exterior wood and wood trim of building.		X						
20. Remove automobiles from rear yard.		X						

(continued)

TABLE 2.1 (cont'd.): Examples of violations treated as serious or non-serious for each type of code violation.

Description of Violation	Building Code Serious	Building Code Non-Serious	Health Code Serious	Health Code Non-Serious	Fire Code Serious	Fire Code Non-Serious	Electrical Code Serious	Electrical Code Non-Serious
21. Provide cover for garbage containers.		X		X				
22. Remove obstruction from bathtub drain and stagnant water from it.		X	X					
23. Replace glass of broken window.		X	X					
24. Provide water to wash basin.	X		X					
25. Provide clean and sanitary mattress for rented apartment.		X	X					
26. Provide fire extinguishers for all floors.	X				X			
27. Provide illuminated exit signs for all floors.	X				X			
28. Post rooming house license.		X						
29. Supply window screens for entire apartment.		X	X					
30. Relocate electrical wall switch.		X						X
31. Repair or replace door bells for each apartment.		X						
32. Repair or replace mailboxes for each tenant.		X						

(continued)

TABLE 2.1 (cont'd): Examples of violations treated as serious or non-serious for each type of code violation.

Description of Violation	Type of code violation: Building Code		Health Code		Fire Code		Electrical Code	
	Serious	Non-Serious	Serious	Non-Serious	Serious	Non-Serious	Serious	Non-Serious
33. Repair or replace ceiling of apartment with one hour fire retarding material.	X				X			
34. Repair defective sink in the bathroom and provide running hot water.	X		X					
35. Repair window sashes for free movement of window.		X						
36. Add an additional electrical outlet in kitchen.		X						
37. Provide window cords to kitchen window.		X						

2.4. WATER DEPARTMENT

The data collection procedure at the Water Department is not so straight forward as in the above named departments. Here the street addresses are used to get the account numbers and the book numbers which contains information for that account. There are in all 13 such books called Trial Balance Books for each year. Once the desired account number and the book in which this account number appears is known, then in the next step we look for that account in the Trial Balance Book bearing that number. Often we need the Trial Balance Book for the year in question as well as a similar book for the preceding year to determine the amount of tax due. Subtracting the amount of the bills paid during the period in question from the total amount due for that period, we can easily determine the amount of tax due. If the tax information is needed for any intermediate period other than that given in the books, one can estimate the amount of tax due for that period on the pro-rated basis. It has been observed that the billing dates for all accounts are not the same, nor is the periodicity of such dates uniform even for the same account. This, therefore, requires studying each account separately to compute the desired tax amount.

2.5. DEPARTMENT OF TAX COLLECTION AND ASSESSMENT

Information about the amount of taxes due has been collected from the Department of Tax Collection and Assessment. This information is readily available on microfiche and can be retrieved for any structure in Newark with correct information about its block and lot numbers. The property tax for a structure becomes due on January 1 of each year and a taxpayer is expected to make payment every quarter. Tax bills are sent twice a year, once in January and the next one in July. If the taxes are not paid within the ten days after the dispatch of the tax bills, the taxes become legally due. So, if the taxes are not paid within this ten day period, they are shown as balances due. Thus, to determine the current amount of tax due for a structure we add the tax amounts due for all the previous years and the current year. This procedure, however, is not complicated as in the Water Department.

It should be remembered that certain individuals and businesses pay their water as well as tax bills regularly. However, they may be somewhat late in their payments every time. For such cases, the balance due, if any, is the amount due after the payment of such bills.

2.6. POLICE DEPARTMENT

From the Newark Police Department, information regarding criminal activities taking place at the selected structures has been collected. This information was collected from the beginning of 1977 till the time of first arson for each structure. In the Police Department, the information is kept on cards and filed by addresses of the location for each year. When a criminal activity takes place at a location, and if the police are called, then a report is filed for that incident. Information regarding address, date and time of incident and each offense classification is then transposed from this report to a card which, as stated before, is filed by location for each year. For the present study, data regarding total number of index (Part I) offenses and non-index (Part II) offenses is collected for each location. (However, information regarding police services for elderly, sick persons and similar community services, which is collected by the Newark Police Department, is ignored for the present study since it does not seem to have any direct bearing on activities leading to arson for that structure). The information is up-to-date and very well kept. Hence the desired information could be easily collected.

2.7. NEW JERSEY INSURANCE UNDERWRITERS' ASSOCIATION

Information regarding insurance for a structure was collected from the New Jersey Insurance Underwriters' Association (NJIUA). The Association, through its application form for insurance, collects considerable information from the applicants. This information is available by location. However, since NJIUA provides insurance for only those structures that are usually denied insurance in the open market, the information is not available for all arson and match samples. Because of the confidentiality of the information, no information was collected for those addresses for which information was not available from NJIUA.

For those structures for which information was available, information regarding the amount of insurance for each year since 1976 till the date of first arson for an arsoned structure, and the corresponding date for the matched structure, was collected. Also, information about the names and addresses of all (past and present) owners, and the amount of loss was collected. The records are well kept and there is no difficulty in collecting such information.

CHAPTER III

ARSON AND MATCH SAMPLES

3.1 INTRODUCTION

In Chapter I, we have shown how the arson sample of 150 structures was selected from the population and we also established that the arson sample is representative of the population under study for the five variables stated earlier. Of these 150 sampled structures, a corresponding match could be developed for only 127 cases due to the peculiarity of the sampled structure, such as size and shape, or due to the non-availability of similar structures in its neighborhood. Out of these 127 cases, 25 cases had to be deleted for discriminant analysis for the reasons indicated below in Table 3.1.

TABLE 3.1 Reasons for deleting 25 cases from arson sample.

Reasons for Deleting	Number of Cases Deleted
1. Too much taxes due for arson cases.	1
2. No fires, no violations, and no taxes for arson and/or match cases.	12
3. Too many violations for match sample.	8
4. Too many fires for arson sample.	2
5. No violations and no taxes.	2
Total	25

The reasons mentioned in the above table, except the last one, represent the extreme cases which often distort the true picture of the population. It is, therefore, advisable to eliminate such cases from statistical analysis to avoid wrong conclusions about the population or to avoid developing an erroneous prediction model. The two cases that had no violations and no taxes (reason No. 5 in the above table) were deleted on the basis of the pattern for other cases in the sample. This judgement was partly subjective. However, as it turned out, the best prediction formula developed later in this chapter did not include the variable "number of pervious fires" so that the deletion seems to be justified.

3.2. ANALYSES OF DATA FOR ARSON AND MATCH SAMPLES

In this section we will analyze data collected from various departments for arson and match samples. The chief aim of this analysis is to determine if there exist any disparities between the arson and match samples in their characteristics considered here and, if so, for which characteristics. Such analysis will help us to gauge the accuracy of the discriminant function to be developed later in this chapter.

For the most part, the analyses employs simple statistical test such as the Kolmogrov-Smirnov test and χ^2 test (Chi-Square test). In many instances the former test is preferred to the latter test because it is more powerful.

3.2.1. ANALYSES OF FIRE AND TAX DATA

If we look at the frequency distributions of the "total number of previous fires" for the arson and match samples, given in the table below, we find that these distributions appear to be significantly different. However, when we apply the Kolomogrov-Smirnov test (two-tailed) to these frequencies we find that there is no significant difference between these frequencies.

TABLE 3.2: Frequency distribution for the total number of previous fires for arson and match samples.

Number of Previous Fires	Arson Sample		Match Sample	
	Number	Percent	Number	Percent
0	83	81.4	98	96.1
1	13	12.8	3	2.9
2	4	3.9	1	1.0
3	2	1.9	0	0
Total	102	100.0	102	100.0

The frequency distribution of all taxes due, for arson and match samples , is as follows. Here "all taxes" means water and property taxes combined together.

[1] See S. Siegel (1956): Non-parametric statistics for the Behavioral Sciences, McGraw-Hill Book Co., Inc., New York pp. 51 and 136

TABLE 3.3: Frequency distribution of all taxes due for arson and match samples.

All Taxes Due (in $)	Arson Sample			Match Sample		
	Frequency	Percent	Cumulative Percent	frequency	Percent	Cumulative Percent
0-1000	68	66.67	66.67	91	89.22	89.22
1001-2000	24	23.53	90.20	9	8.82	98.04
2001-3000	3	2.94	93.14	1	0.98	99.02
3001-4000	2	1.96	95.10	1	0.98	100.00
4001-5000	2	1.96	97.06	0	0.00	100.00
5001-6000	1	0.98	98.04	0	0.00	100.00
6001-7000	2	1.96	100.00	0	0.00	100.00
Total	102	100.00	-	102	100.00	-

When we test the hypothesis of no significant difference in the above distributions using the Kolmogrov-Smirnov test, we do not accept this hypothesis for the above data, at the 5% level of confidence. Thus, the distribution of all taxes for the arson sample differs significantly from that for the match sample. The mean amount of taxes due (which is $992.18) for arson cases is about three times the amount due (which is $371.18) for match cases. Using the t-test for these samples[1] we find $t = 4.319$ (101 d:f), which is highly significant even at 1% level of confidence.

3.2.2. ANALYSES OF DATA FOR CODE VIOLATIONS

As stated in the previous chapter, information regarding code violations is collected separately for serious and non-serious violations for each type of code violation viz, building code, health code, electrical code, and fire code violations. We will examine the disparities, if any, between arson and match samples for each code violation. However, to facilitate quick visual inspection of the data, details about the sample distributions viz; percentages and cumulative percentages for each distributions are not given. Since there are only 102 cases (N = 102), it may be noted that the percentages are very close to the figures given in the body of the tables.

[1] Snedecor G. W., and Cochman W. G., Statistical Methods , 1967, 6th edition. The Iowa State University Press, Ames, Iowa; pps. 92 - 94.

TABLE 3.4: Distribution of each type of non-serious violation for arson and match sample.

Number of Violations	Type of Violation: Building Code		Health Code		Electrical Code		Fire Code	
	Arson	Match	Arson	Match	Arson	Match	Arson	Match
0	52	77	59	79	87	95	85	95
1	11	5	8	6	6	4	10	4
2	4	3	2	5	4	3	3	3
3	4	1	4	3	2		3	
4	0	2	2	1	1		1	
5	1	2	2	2	1			
6	0	2	3	1	0			
7	2	1	1	1	1			
8	2	0	2	0				
9	1	2	0	1				
10	0	0	1	0				
11	0	1	2	0				
12	3	0	0	1				
13	0	0	1	0				
14	0	1	2	0				
15	2	0	3	0				
16	1	1	1	1				
17	1	0	0	0				
18	0	0	0	0				
19	3	0	1	0				
20	1	1	0	0				
21	0	0	2	0				
22	0	0	0	0				
23	1	1	0	0				
24	1	0	0	1				
25	1	0	1	0				
26 or more	11	2	5	0				
N*	102	102	102	102	102	102	102	102
Total	758	222	486	113	36	10	29	10

*N=Total number of cases.

A glance at Table 3.4 shows that there are almost three times more violations of each type for arson sample than for the match sample. The actual ratios are as shown in Table 3.5 given below:

TABLE 3.5: Ratio of each type of non-serious violations in the arson sample to the match sample.

Type of Violation	Building Code	Health Code	Electrical Code	Fire Code	All
Ratio	3.41	4.30	3.60	2.90	3.69

The individual ratios are consistently higher for the arson sample than for the match, but since they are not too large, the total number of violations of each type shows no significant difference. Here $\chi^2 = 3.628$ on 3 d.f (degrees of freedom) which is less than the table (theoretical) value of $\chi^2 = 7.82$ at the 5% level of significance. Applying the Mann-Whitney U-test to the data in Table 3.4 for totals of each type of violation, we find that $U = 4$ and the associated probability is 0.171 which is far greater than the 5% level of significance. This confirms the conclusion reached using the t - test viz, that there is no significant difference between the arson and match samples in the total number of violations of each type.

The test of the hypothesis regarding the distributions of various types of violations for the arson and match samples indicate that these distributions are significantly different, statistically speaking, only for the building code and health code violations. To some extent this is apparent from the fact that there are a larger number of cases in the match sample with no building code and health code violations than the arson sample. The Mann-Whitney U-test [1] a test more powerful than the Kolmogrov-Smirnov test, also confirmed that there are no significant differences in the distribution for arson and match samples for electrical code violations and fire code violations.

We now consider the distributions for serious violations given in Table 3.6 below.

[1] See S. Siegel (1956) op. cit., pp. 120 - 125

TABLE 3.6: Distribution of each type of serious violations for the arson and match samples.

Number of Violations	Type of Violation							
	Building Code		Health Code		Electrical Code		Fire Code	
	Arson	Match	Arson	Match	Arson	Match	Arson	Match
0	57	61	57	67	79	86	63	66
1	11	11	11	10	11	8	13	15
2	10	8	7	10	3	2	8	5
3	7	6	5	3	3	3	4	9
4	0	4	4	3	4	2	6	3
5	4	2	4	2	0	1	1	2
6	0	1	3	1	1		3	1
7	2	1	1	0	0		0	0
8	2	2	1	1	0		0	0
9	1	1	3	0	0		1	0
10	1	1	0	2	1		1	1
11	2	0	0	0			0	
12	1	2	1	1			1	
13	0	0	0	1			0	
14	0	0	0	0			1	
15	3	0	1	0				
16	0	1	2	0				
17	0	0	1	0				
18	0	0	0	0				
19	0	0	0	0				
20	0	0	0	1				
21	0	0	0					
22	1	1	1					
N*	102	102	102	102	102	102	102	102
Total	222	182	234	140	58	30	133	90

*N=Total number of cases.

The inspection of the distribution as well as the application of the Kolmogrov-Smirnov test show that there is no significant difference between the arson and the match sample for each type of serious violations.

In this study we have considered separately serious and non-serious violations of each type because they are qualitatively different. We have seen above that there are almost three times more non-serious violations in the arson sample than in the match sample and less than twice the serious violations in the arson sample than in the match sample. However, we only looked at the inter-sample differences. We will now examine the intra-sample differences between serious and non-serious violations.

From Tables 3.4 and 3.6, it is evident that the distributions of non-serious and serious violations, for each type of code violation do not differ significantly for the arson sample and also for the match sample. Consider now the total number of serious and non-serious violations of each type for the arson sample which are given in Table 3.7 below.

TABLE 3.7: Distribution of serious and non-serious violations by type of violations for arson sample.

Type of Violation	Type of Code				Total
	Building	Health	Electrical	Fire	
Serious	222	234	58	133	647
Non-serious	758	486	36	29	1,309
Total	980	720	94	162	1,956

The observed value of χ^2 = 298.891 on 3 d.f. is highly significant even at at the 1% level of significance so we conclude that there are significant differences between the total number of serious and non-serious violations for the arson sample. Further comparison of the total number of serious and non-serious violations, considering these values for each structure separately for the sample, by t - test (observed t = 3.828 on 101 d.f.) reveals that there is indeed significant difference between the serious and non-serious violations for the arson sample. Thus, we conclude that for the present data not only qualitative differences exist between serious and non-serious violations but also that the differences in the occurrence of these violations for each structure are statistically significant. Moreover, the total of all types of non-serious violations occur more frequently than the corresponding figure for serious violations.

We now consider the distribution of serious and non-serious violations for the match sample.

TABLE 3.8: Distribution of serious and non-serious violations by type of violations for match sample.

Type of Violation	Type of Code				Total
	Building	Health	Electrical	Fire	
Serious	181	140	30	90	441
Non-serious	222	113	10	10	355
Total	403	253	40	100	796

From the above table we find that $\chi^2 = 72.615$ on 3 d.f. is highly significant at the 5% level of confidence. As before, when we consider structure-wise difference between serious and non-serious violations, we find that t = 1.3936, on 101 d.f. which is not significant at the 5% level of significance. Hence we conclude that when structures are considered individually there is no significant difference between the serious and the non-serious violations for the match sample.

We now test the hypothesis of no significant difference between the total number of serious and non-serious violations for the arson and match samples. The 2 x 2 contingency table for this purpose is as follows.

TABLE 3.9: Total number of serious and non-serious violations for arson and match samples.

Violations	Sample		Total
	Arson	Match	
Serious	647	441	1,088
Non-serious	1,309	355	1,664
Total	1,956	796	2,752

The observed value of $\chi^2 = 117.02$ on 1 d.f. being far greater than the table (theoretical) value of $\chi^2 = 3.84$ on 1 d.f. at the 5% level of significance, we conclude that there are significant differences between serious and non-serious violations for the arson and match structures.

Similarly, when we test the hypothesis of no significant difference between the total number of serious violations for the arson and the match samples, we find that $\chi^2 = 39.004$ on 1 d.f., is also highly significant at the 5% level of significance. This is also true when we test a similar hypothesis for non-serious violations. (Here $\chi^2 = 546.945$ on 1 d.f.)

Within the arson sample, if we test the hypothesis of no significant difference between serious and non-serious violations, then we find that $\chi^2 = 224.051$ of 1 d.f., which is highly significant. Thus, we condlude that there are significant differences between the number of serious and non-serious violations occurring for the arson structures. The same conclusion holds true for the match sample. (Here $\chi^2 = 9.291$ on 1 d.f.)

We have thus shown that for the arson sample the serious and non-serious violations differ significantly, whether we consider them by the total number of such violations, or by the type of such violations, or by individual structures. This is, however, not the case with the match sample.

The above results for the arson and the match samples appear to be consistent with the reality. It is expected that those structures that experience arson have more frequent violations than those structures that do not experience any arson. Moreover, probably to avoid attention of the authority and neighbors, fewer serious code violations take place in such structures than the non-serious code violations.

Lastly, we consider the distribution of serious and non-serious violations combined for the arson and the match samples by the number of violations as shown in Table 3.10 below. We notice that there are no significant differences between the distributions of the arson and the match samples for serious violations. Similarly, we find, using the Kolmogrov-Smirnov test, that there are no significant differences between the distribution of serious and non-serious violations for the arson as well as the match samples. However, there are statistically significant differences in the arson and match sample distributions for non-serious violations. The differences between the arson and match samples are more pronouned when we consider the paired comparisons. The value of t = 3.829 on 101 d.f., and the associated probability is much less than 0.0001 which implies that the observed value of t is highly significant even at the 0.1% level of significance.

TABLE 3.10: Distribution of serious and non-serious violations for arson and match samples by the number of violation.

Number of Violations	Type of Violation: Serious		Non-Serious	
	Arson	Match	Arson	Match
0- 3	59	64	61	82
3- 6	9	13	8	4
6- 9	9	8	2	4
9-12	7	4	2	2
12-15	3	4	2	2
15-18	2	2	1	1
18-21	2	1	1	2
21-24	0	1	4	1
24-27	3	0	0	0
27-30	3	3	5	0
30-33	3	0	1	1
33-36	0	0	2	1
36-39	0	1	0	0
39-42	0	0	1	0
42-45	0	0	1	1
45-48	1	0	2	0
48-51	1	0	0	0
51-54		0	1	0
54-57		0	1	0
57-60		1	0	0
Over 60			7	1
N*	102	102	102	102
Total	647	441	1,309	355

*N = Total number of cases.

3.2.3. ANALYSES OF CRIME DATA

From Table 3.11. we find using the Kolmogrov-Smirnov test, that the distribution of the arson and match samples differ significantly for Part I and all crimes categories and not so for Part II crimes. We further find that for the arson sample, Part I and Part II crime categories differ significantly at the 5% level of significance since observed D = 0.2157 and the table value of D, at the 5% level of significance, is D = 0.1904, where D is the Kolmogrov-Smirnov Statistic.

TABLE 3.11: Distribution of Part I, Part II, and all crimes for arson and match samples by the number of such crimes.

Number of Crimes	Part I Crimes		Part II Crimes		All Crimes	
	Arson	Match	Arson	Match	Arson	Match
0	22	43	44	58	14	37
1	22	21	27	23	18	17
2	19	17	14	10	14	15
3	7	8	6	6	8	11
4	14	3	5	0	11	7
5	3	3	3	1	7	2
6	1	1	0	0	5	4
7	2	1	1	2	5	1
8	3	1	0	0	4	1
9	1	1	0	1	3	1
10	1	0	1	0	2	1
11	1	1	0	0	2	0
12	0	1	0	1	0	0
13	2	0	0		1	1
14	1	1	0		0	2
15	1		0		1	1
16	1		0		1	1
17	0		0		0	0
18	1		0		0	0
19			0		2	0
20			1		1	0
Above 20					2	1
N*	102	102	102	102	102	102
Total	315	173	145	101	460	274

#N=Total numbers of cases.

Considering the total number of Part I and Part II Crimes for the arson and the match samples, we test the hypothesis of no significant difference for these two categories of crimes using the 2 x 2 contingency table which is as follows:

TABLE 3.12: Total Part I and Part II crimes for arson and match samples.

Crime Category	Sample		Total
	Arson	Match	
Part I	315	173	488
Part II	145	101	246
Total	460	274	734

Since the observed value of $\chi^2 = 1.964$ on 1 d.f. is less than the table value of $\chi^2 = 3.84$ on 1 d.f., at the 5% level of significance we do not reject the preceding hypothesis. Thus, the data do not support the hypothesis that there are significant differences in the total number of Part I and Part II crimes committed in the arson and the match structures. However, the picture is entirely different when we test the hypothesis of no significant differences between the arson and the match samples for Part I crimes alone. Here $\chi^2 = 41.320$ on 1 d.f., which is highly significant at the 5% level of significance. Thus, we conclude that there are significant differences between the total number of Part I crimes committed for the arson and the match samples. We make a similar conclusion for Part II crimes ($\chi^2 = 7.87$ on 1 d.f).Like the preceding analysis for Part I and Part II crimes, when we test similar hypotheses for the arson and the match samples separately, we reach similar conclusions. (For the arson sample $\chi^2 = 62.826$, on 1 d.f. and for the match sample $\chi^2 = 18.92$ on 1 d.f.).

The above results suggest that although there are no significant differences between the two crime categories for the arson and the match samples when these categories are considered separately they do show significant differences between the two samples within each crime category. Precisely the same inferences can be drawn when we consider the two samples instead of the two crime categories.

When we test the hypothesis of no significant difference between the two crime categories for the arson sample by considering each structure individually rather than the total for all structures as above, we find that there are significant differences, (t = 3.774 on 101 d.f.) between Part I and Part II Crimes. The same conclusion is drawn when we test the match sample similarly. (Here t = 3.513 on 101 d.f.).

3.2.4. ANALYSES OF INSURANCE DATA

We will now examine the data about the insurance. As stated in the previous chapter, information about three variables was collected, viz, the amount of insurance for each year for which information is available, the name and addresses of the owners, and the insurance loss claimed. From this information three variables were defined, which are as follows:

(i) <u>The percent change in the amount of insurance (PCI)</u>:

The PCI for a year is defined by

$$PCI = \frac{AI - AIP}{AIP} \times 100$$

Where AI = Amount of insurance for that year
AIP = Amount of insurance for the previous year.

(ii) <u>The number of owners</u>:

This is the number of persons who owned the property. For our purpose, any joint ownership is treated as a single ownership. That is, two or more persons owning the same property at the same time is considered as one owner for that property.

(iii) <u>Percent loss claimed</u>:

The percent loss (PL) claimed is defined as

$$PL = \frac{AL}{AI} \times 100$$

Where AL = Total amount of loss for all the previous years
AI = Total amount of insurance for these years.

The total loss includes the loss due to the damage to the structure and also any loss to its contents.

Although insurance taken by a tenant of a structure was not included for the present study, it can be considered with slight modifications in the above definitions.

The above three variables were then combined in a single variable labelled as "Insurance Score". To obtain the insurance score we use the following procedure.

(a) Change the sign of PCI, i.e., if PCI is positive, change it to negative and if it is negative, change it to a positive sign. This simply means that if there is a decrease in the amount of an insurance policy, the structure is less likely to experience an arson than if there is an increase in the amount for a policy. This score is called the amount of insurance score (AS).

(b) We obtain the ownership score (OS) by dividing the number 10 by the number of owners as defined earlier in (ii). The number 10 is an arbitrary number so that any other number can be chosen instead of 10. Here this number is chosen because

(1) it is simple to compute and add such a score to other scores, when the number of owners is one or two, which was the case for the present data, and

(2) to keep the variability of the insurance score (IS) within managable limits for calculations on a calculator.

(c) The loss score (LS) is obtained by changing the sign of PL obtained in (iii) since such a loss for a structure indicates that the structure is more likely to have an arson than a structure that has not shown such a loss in the past.

(d) We compute the the insurance score, IS, by the formula:

$$IS = AS + OS + LS$$

Note that the higher the insurance score, the less likely it is to have an arson.

The comparison of the mean insurance score for the arson and the match samples show a value of $t = -2.411$ highly significant at the 5% level of significance. Thus, these two samples differ significantly in their mean insurance score.

3.2.5. SUMMARY OF RESULTS

Before we proceed with the discriminant analysis, it may be helpful to summarize the results we have so far obtained from the analyses of the data for arson and match samples:

(1) The distribution of "total number of previous fires" does not differ significantly for the arson and match samples.

(2) For the variable "total amount of all taxes due" we observe that

(a) The distribution differs significantly for the two samples

(b) The mean amount of taxes is significantly greater for the arson sample than for the match sample.

(3) For the variables serious and non-serious violations we find that

(a) Non-serious violations of each type occur more frequently in the arson sample than in the match sample and the ratio of all violations of this type for the arson sample to the match sample is 3.7:1.

(b) The distributions for non-serious building code and health code violations differ significantly for the two samples

(c) Serious violations of each type occur more frequently in the arson sample than in the match sample and the ratio of all serious violations for the arson sample to the match sample is 2:1

(d) The distribution for each type of serious violation is the same for the arson and the match samples. That is, the parent population for both the samples is the same for each type of serious violation

(e) The non-serious violations occur more frequently than the serious violations in the arson sample. The ratio of such occurrence is about 2:1.

(f) For the arson sample, the ratio of non-serious building and health code violations to similar serious violations is about 2.7:1 and the ratio of serious electrical and fire code violations to non-serious similar violations 2.9:1

(g) The serious and non-serious violations of all types differ significantly for the arson as well as the match samples, with these violations occurring more frequently in the arson sample. The ratio of occurrence of all types of violations in the arson sample to those in the match sample is 2.5:1

(h) The serious violations of all types occur more frequently than the non-serious violations of all types in the match sample. The ratio of the serious violations to non-serious violations is 1.24:1. Note that for the arson sample, the reverse is true, See (e) above.

(i) For the match sample, the ratio of non-serious building and health code violations to similar serious violations is 1:1 and the ratio of serious electrical and fire code violations to similar non-serious violations is 6:1.

(j) The distribution of serious violations by number of violations do not differ significantly for the arson and the match samples. The total number of such violations however, do differ significantly. This is also true when paired comparison is made by t - test of arson and match samples.

(4) For the variables Part I and Part II Crimes we find that

(a) The distribution of Part I crimes and all crimes categories differ significantly for the arson and the match samples. However, that is not the case for the Part II category of crimes.

(b) Within Part I and Part II categories, the total numbers of crimes committed differ significantly for the arson and match samples.

(c) Within the arson and the match samples, the total number of Part I Part II crimes committed differ significantly.

(d) For the arson sample, the ratio of Part I crimes committed to Part II crimes committed is 2.2:1.

(e) For the match sample, the ratio of Part I crimes committed to Part II crimes committed is 1.7:1.

(5) There is a significant difference between the means for insurance score for the arson sample, with a less than zero mean score for the arson sample. (Note that the lower the insurance score the more likely that the structure in question will have an arson.)

From (1) - (5) above, we can conclude that the following variables are more important in discriminating between the arson and match samples: Total amount of tax due, non-serious violations of all types, Part I crimes committed at the structure, and the insurance score.

3.3 DISCRIMINANT ANALYSIS

3.3.1 INTRODUCTION

To develop a discriminant function, it is necessary to consider

(i) the probability of misclassification (P_M) of a structure, under the assumption that each variable appearing in the discriminant function is independent, and

(ii) the estimated probability (P_E) of misclassification obtained empirically.

For the case of two groups, the formulae for P_M and P_E involved are simple and can be useful in selecting the variables to be included in a discriminant function. Let d_i denote the difference between the means of the arson and the match groups for the variable $i = 1, 2, \ldots, p$. Let S_i^2 denote the mean error sum of squares (MESS) obtained from the analysis of variance, for the i^{th} variable, or from the matrix for error sum of squares. Let

$$A_1 = \sum_{i=1}^{p} \frac{d_i^2}{S_i^2} \quad , \quad A_2 = 1/2\sqrt{A_1} \tag{3.1}$$

then $P_M = 1 - \{ \Pr. (X \geq A_2) \}$

It may be noted that P_M is the minimum probability attainable through the selected set of variables when all these variables are independent of one another.

Let the discriminant function be given by

$$Y = \sum_{i=1}^{p} a_i X_i \tag{3.2}$$

where a_i is the estimated coefficients, estimated by maximum likelihood estimation method, for the i^{th} variable and X_i is the value for the i^{th} variable. Let

$$B_1 = (n_1 + n_2 - 2) \sum_{i=1}^{p} a_i d_i \quad , \quad B_2 = 1/2\sqrt{B_1} \tag{3.3}$$

then $P_E = 1 - \{ \Pr. (X \geq B_2) \}$ (3.4)

where n_1 = number of structures in the arson group and

n_2 = number of structures in the match group.

The formula for B_1 in (3.3) suggests that we should exclude those variables from the discriminant function for each of which $a_i d_i$ is less than zero. Also, it can be easily seen from (3.4) that the estimated probability of misclassification increases if we include in computing a variable for which $a_i d_i$ is less than zero. In developing discriminant function below, this fact has been taken into account.

3.3.2. THE DISCRIMINANT FUNCTION

The following table, Table 3.13, gives the values of d_i, S_i and d_i/s_i for all the variables considered for discriminant analysis.

TABLE 3.13: Variables used for discriminant analysis and their respective values of di/si.

Variable	Abbreviations Used	d_i	s_i	d_i/s_i	Rank of d_i/s_i*
Number of previous fires	F	0.2157	0.4803	0.4490	4
Non-Serious Violations	-	9.3529	17.2618	0.4805	-
Building code	BC	5.2549	10.3200	0.5092	3
Health code	HC	3.6569	6.9350	0.5273	2
Electrical code	EC	0.2549	0.8050	0.3168	8
Fire code	FC	0.1863	0.5930	0.3136	9
Serious Violations	-	2.0196	9.6172	0.2090	-
Building code	BCS	0.4020	3.8310	0.1049	14
Health code	HCS	0.9216	3.7210	0.2478	10
Electrical code	ECS	0.2745	1.1960	0.2299	11
Fire code	FCS	0.4216	2.1330	0.1978	13
All violations	V	11.3431	25.4696	0.4454	5
All Taxes	T	0.6099	0.9830	0.6200	1
Water	-	0.4091	0.9021	0.4536	-
Property	-	0.2119	0.6743	0.3142	-
Crimes	-	1.8235	5.6762	0.3282	-
Part I	PI	1.3922	3.2516	0.4281	6
Part II	PII	0.4412	2.2250	0.1983	12
Insurance	INS	-11.7970	34.9407	-0.3376	7

* Rank is of the absolute (positive) value of di/s_i

There are in all 19 variables, some of which have been used as domain variables, e.g., non-serious violations, serious violations, and all taxes. There are variables which were purposly not used to develop discriminant function e.g., water taxes and property taxes. We have combined these two variables in a single domain variable labelled as "all taxes."

The last column of the above table ranks these variables by the absolute value of d_i/s_i. The rank denotes the relative importance of each variable in minimizing the probability of misclassification under the assumption that the variable included in the discriminant function is independent of any other variable appearing in that function. We find that the variable "Insurance" ranks seventh and ranks higher than most of the code violations and Part II crimes. Thus, the insurance score developed earlier seems to be useful in discriminating the structures that are arson prone from those that are not arson prone. This will now be established in what follows.

With 19 variables given in the above table, several discriminant functions can be tried. Once we obtain a few initial discriminant functions and corresponding values of P_E, it is not difficult to eliminate other discriminant functions, using different sets of variables, by looking at the respective values of P_M. Thus, if P_M is less than the maximum P_E obtained for a discriminant function developed so far, then there is no need to develop a discriminant function for that set of variables. Such considerations are helpful not only when calculations are to be done manually, as was the case for this project, but also when they are carried out using a computer. Using the preceding criterion, five discriminant functions were developed and one is selected (the manual computation of P_M and the time constraints did not permit considering all possible combinations). Three of these discriminant functions that are of interest to us here are given in the following table, Table 3.14. It is possible to improve upon these discriminant functions since all variables have not been considered simultaneously in any one of these functions.

TABLE 3.14: Results of discriminant analysis for three sets of variables.

DF* No.	Variables Entered in the DF**	Discriminant Coefficients for each variable in Col. 2	Variables Accepted for DF**	P_M	P_E
(1)	(2)	(3)	(4)	(5)	(6)
I	Fires (F), Taxes (T), Violations (V)	C_F=0.003812, C_T=0.003086, C_V=0.000096	Fires, Taxes Violations	0.3300	0.3310
II	Fires (F), Taxes (T), BCS, HCS, ECS, BC, HC, EC, FC, PI, PIII	C_F=-0.003142, C_T=0.003396, C_{BCS}=-0.001507, C_{HCS}=0.000850, C_{ECS}=-0.000831, C_{FCS}=-0.000056, C_{BC}=0.000188, C_{HC}=0.000351, C_{EC}=0.000933, C_{FC}=0.001542, C_{PI}=-0.000233, C_{PII}=0.000360	Taxes, HCS, BC, HC, EC, FC, PII	0.2565	0.2940
III	Fires (F), Taxes (T), BC, HC, EC, FC, PI, PII, INS	C_F=-0.000818, C_T=0.000118, C_{BC}=0.000035, C_{HC}=-0.000168, C_{EC}=-0.000368, C_{FC}=0.003318 C_{PI}=0.002963, C_{PII}=-0.001358, C_{INS}=-0.000048	Taxes, BC, FC, PI, INS	0.2598	0.2977

* DF No. denotes the discriminant function (DF) numbers referred to in the text.

** DF means discriminant function.

3.3.3. ON SELECTING A DISCRIMINANT FUNCTION

Among these three discriminant functions (which will be called DF I, DF II and DF III respectively), the most preferred is DF II. However, there is a difference of about 0.4% in the probabilities P_E for DF II and DF III. Since we need to gather information for two more variables for DF II than for DF III, the cost of employing DF II instead of DF III is higher. Thus, comparing the reduction in misclassification probability for DF II against the possible cost of collecting information for it, to similar statistics for DF III, it seems more preferable to use DF III.

Similarly, we notice that there is a difference of about 3.3% in the misclassification probabilities for DF I and DF III, with a lesser probability P_E for DF III. Is this difference large enough to choose DF III over DF I? The answer to this question can be found by comparing the relative cost of wrong decisions discussed below.

Associated with every discriminant function is a set of decision rules which classify a structure as likely to have an arson or it is not likely to have an arson. By classifying known cases of arson and non-arson, we can estimate how useful the designed set of rules for the discriminant function is in predicting future arson. Let us suppose that for a set of decision rules R, A number of known arson cases and M number of known match cases are classified as follows:

TABLE 3.15: Classification for A arson and M match cases.

Cases	Arson	Non-Arson	Total
Arson	A_A	A_N	A
Match	M_A	M_N	M
Total	$A_A + M_A$	$A_N + M_N$	A + M

The probability of wrong classification P_W is

$$P_W = \frac{A_N + M_A}{A + M} \tag{3.5}$$

Let d be the average cost of making a wrong decision that a would-be non-arson structure is likely to have an arson. This cost includes administrative and field costs in preventing an arson. Similarly, let Kd, $K > 1$, be the average cost of making a wrong decision that a would-be arson structure is not likely to have an arson. Since on an average, the value of the damages to a structure and its contents is greater due to arson than the cost of preventing an arson for a structure, we have assumed here that $K > 1$. Then from the above table it follows that the total cost of making a wrong decision is

$$C = d\,K\,A_N + d\,M_A \qquad (3.6)$$

and the average cost of making a wrong decision per structure is

$$C_a = \frac{d\,(M_A + K\,A_N)}{M_A + A_N} \qquad (3.7)$$

Let $P_{A/W}$ = Probability of classifying a structure as an arson (structure), given that it is a wrong classification

and $P_{M/W}$ = Probability of classifying a structure as non-arson (structure), given that it is a wrong classification.

From Table 3.15, we see that

$$P_{A/W} = \frac{M_A}{M_A + A_N} \qquad P_{M/W} = \frac{A_N}{M_A + A_N} \qquad (3.8)$$

From (3.8), the average cost of making a wrong decision in (3.7) can be written as

$$C = d\,P_{A/W} + d\,K\,P_{M/W} \qquad (3.9)$$

Let $C_{a\,I}$ and $C_{a\,II}$ denote the average cost of wrong decisions for two sets of rules. Then we define the relative cost efficiency (RCE) of Rule I, relative to Rule II, as

$$RCE = \frac{C_{a\,II}}{C_{a\,I}} \times 100 \qquad (3.10)$$

Returning to the problem of determining which of the discriminant functions DF I and DF III to accept, we consider the classifications obtained by the associated decision rules Rule I and Rule III for each function. The results of this classification are given in Table 3:16 below.

TABLE 3.16: Classification of known cases using Rule I and Rule III.

	Rule I			Rule III		
Cases	Arson	Non-Arson	Total	Arson	Non-Arson	Total
Arson	76	26	102	85	17	102
Match	36	66	102	46	56	102

Using formula (3.9) for $K = 2$, we find that

$$C_{aI} = \frac{36}{62}d + \frac{52}{62}d = \frac{88}{62}d$$

$$C_{aIII} = \frac{46}{63}d + \frac{34}{63}d = \frac{80}{63}d$$

Hence RCE of DF III, relative to DF I, is

$$RCE = \frac{88d}{62} \times \frac{63}{80d} \times 100 = 111.8\%$$

Similarly, when K = 5, we find that RCE = 128.8%. Hence we conclude that if the cost of damage due to arson to a structure is twice the cost of preventing such an arson, i.e., K = 2, then on an average, it will cost 11.8% more if we use DF I rather than DF III. Similarly, if such a cost is five times, i.e., K = 5, then the cost of using DF I is 28.8% more than if we use DF III. We, therefore, accept DF III as our formula for predicting future arsons.

3.3.4. ANALYSIS OF THE SELECTED DISCRIMINANT FUNCTION

As seen in the preceding subsection, DF III is more cost effective than DF I and hence we will use this function. The formula for the discriminant function, from Table 3.14, is:

$$Y = 0.000118X_1 + 0.003318X_2 + 0.002963X_3 - 0.000048X_4 + 0.000035X_5 \quad (3.11)$$

or, equivalently,

$$Y = 3.37X_1 + 94.80X_2 + 84.66X_3 - 1.37X_4 + X_5 \quad (3.12)$$

Where

Y = Discriminant Score

X_1 = Total amount of unpaid water and property taxes (in thousand dollars)

X_2 = Total number of non-serious fire code violations

X_3 = Total number of Part I (Index) crimes

X_4 = Insurance Score for the structure

X_5 = Total number of non-serious building code violations.

By applying the above formula in (3.12), one can compute the discriminant score Y for a structure and then classify that structure as likely or not likely to have an arson on the basis of the following decision rules:

a) If $Y \geq 50$, then the structure is likely to have an arson

b) If $Y < 50$, then the structure is not likely to have an arson.

In the last column of Table 3.14, we have already given the estimated probability of misclassification as $P_E = 0.2977$. We now test for significance the discriminant function. The analysis of variance table for this purpose is shown in Table 3.17 below.

TABLE 3.17: Analysis of variance of the discriminant function.

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Sum of Squares	F
Between variables	5	0.0015804	0.0003161	11.245
Within variables	198	0.0055667	0.0000281	
Total	203	0.0071471		

The value of $F = 11.245$ obtained in the above table is highly significant even at 1% level of significance. Thus, the discriminant function developed here is useful in classifying arson-prone structures from non-arson-prone structures with an estimated probability of correct classification as 0.7023 or 70.23%. (The probability of correct classification is $1 - P_E = 1 - 0.2977$.)

In applying the formula in (3.12), we note that we consider non-serious fire code and building code violations. Since such violations generally precede any serious violation, the formula suggests to use the non-serious violation. If in case no non-serious violation has occurred before a serious violation takes place, then we use this serious building or fire code violation in our formula. The reason for using a serious or a non-serious violation that occurs first in the above formula is to determine the possibility of an arson for a structure, in early stages of the code violations for it.

3.5.5. RELATIVE COST EFFICIENCY OF NEWARK RELATIVE TO BOSTON AND NEW YORK CITY

The formula for RCE in (3.10) can be used to compare the cost efficiency of any two discriminant functions used for discriminating arson from non-arson structure or any other phenomenon. Hence it can be used to measure the cost efficiency of discriminant functions developed by various localities. Here, as an illustration, we have chosen Boston and New York City for this purpose. The results of their classifications given in their respective reports are summarized in the following table.[1]

1 The City of New York Arson Strike Force (1980): Predicting arson in New York City.
Urban Educational Systems, Inc. (1980): The Research Manual: A manual of property research and arson analysis.

TABLE 3.18: Classification of known cases for Boston and New York City.

Cases	Boston			New York City		
	Arson	Non-Arson	Total	Arson	Non-Arson	Total
Arson	48	30	78	6,166	3,376	9,542
Match	10	68	78	1,404	10,819	12,223

Using formula (3.5), we find that the probabilities of wrong classification for Boston, New York, and Newark are respectively given by

$$P_W\text{ (Boston)} = 25.6\%,\quad P_W\text{ (NYC)} = 22.0\% \text{ and}$$

$$P_W\text{ (Newark)} = 30.9\%$$

We see that Newark has the highest probability of misclassification followed by Boston and then New York City. Using the formula in (3.10) we now compare RCE of Newark relative to Boston and New York City. In the computation of RCE it is assumed that the cost ratio (K) of arson to the non-arson structures is the same for the two cities being compared. This assumption seems necessary for comparison of the two classification rules. The following table gives the values of RCE for Newark, relative to Boston and New York City for different values of K.

TABLE 3.19: RCE for Newark, relative to Boston and New York for different values of K

K		2	3	4	5	10
RCE for Newark (in percent) relative to	Boston	137.8	162.4	179.6	192.4	226.0
	New York	134.4	156.7	172.4	184.0	214.6

We notice from the above table that it costs much less to use DF III and the associated rules developed by Newark than that developed either by Boston or by New York City. The result is no surprise because although both cities have lower probabilities of misclassification (P_W's) than Newark, each has a much higher percent of arsons misclassified as non-arsons as compared to Newark. In the case of Boston among the misclassifieds, about 75% are misclassified as non-arson and for New York it is 70.6%. On the other hand for Newark it is 27%. Thus, under the assumption that the cost ratio of arson structure to non-arson structure is the same for each city, the discriminant function and the associated rules developed by Newark are more cost efficient than those suggested by either Boston or New York City.

CHAPTER IV

PATTERNS OF ARSON

4.1 INTRODUCTION

In Chapter I we have observed some of the population characteristics. For example, we have seen that there is a significant difference between the number of arson incidences in the first four months of 1981 as compared to the corresponding period for the year 1980. We also observed that the arson incidences occur more frequently on Mondays, Thursdays, and Fridays than on any other days of the week. However, percentage wise, the arson incidences are distributed more or less evenly throughout the week with the minimum of 12.6% arson incidences on Wednesdays and the maximum of 16.2% arson incidences on Mondays, see Table 1.9. We have also observed in that chapter how the hourly distribution of arson incidences exhibit a clear pattern when grouped in three-hour time interval. Hence the aim of this chapter is to study the available data and try to identify if there exists any patterns of arson or any functional relationships between the variables such as month, day, and time. To make use of all the available data, we will assume that the population of arson incidences is itself a sample of all arson incidences from the superpopulation of structures described in Chapter I. However, before proceeding with the analysis of the data to study patterns of arson for different variables, we discuss below the distribution of arson incidences for the year 1980.

There were in all 834 arson incidences during 1980 out of the total of 1,123 incidences for the sixteen months study period. Thus, there were about 70 arson incidences per month. These arson incidences occurred in 690 new structures that were ignited during the year. Thus, on an average about 58 new structures were ignited every month of the year 1980. Of the 897 structures that experienced arson for the first time during the sixteen month study period, the remaining 203 were burned during the first four months of 1981. Thus, on an average 50 new structures were ignited during the first four months of 1981. We see, therefore, that on an average seven less new buildings were torched during each month of the period January 1, 1981 to April 30, 1981. However, the average number of arsons per month during this period, which was 72, remained almost the same as the preceding year.

In Table 4.1, we present the distribution of structures by the number of arsons for these structures during 1980. The distribution of the number of structures in this table is similar to the one for the entire population given in Table 1.0. We observe that this is also true for the distribution of the total number of arsons.

TABLE 4.1: Distribution of the population of structures by the number of arsons experienced during the year 1980.

Number of Arsons	Number of Structures	Percent of Structures	Cumulative Percent of Structures	Total No.* of Arsons	Percent of All Arsons	Cumulative Percent of All Arsons
1	600	87.0	87.0	600	71.9	71.9
2	66	9.6	96.6	132	15.8	87.7
3	11	1.6	98.2	33	4.0	91.7
4	6	0.9	99.1	24	2.9	94.6
5	4	0.6	99.7	20	2.4	97.0
6	1	0.1	99.8	6	0.7	97.7
7	1	0.1	99.9	7	0.8	98.5
12	1	0.1	100.0	12	1.5	100.0
Total	690	100.0	-	834	100.0	-

*Total number of arsons = number of arsons in Column (1) x number of structures in column (2).

Single arsons, which affected 87% of the structures, accounted for about 72% of all arsons. Thus, 13% of the structures accounted for 28% of all arsons. It should be noted that these 13% of the structures had their first arson during 1980 and were subjected to additional 144 (17%) arsons during 1980.

4.2: ANALYSIS OF ARSONS BY MONTH AND BY DAY.

In Chapter 1 the entire sixteen-months study period was considered since the object there was to present the population characteristics. However, for studying patters of arsons, it is more meaningful to use the 12-month period from January1, 1980 to December 31, 1980. Wherever required, we will make use of the information for the remaining four months of the study period, viz., January 1, 1981 to April 30, 1981.

For the monthly comparison of arson incidences, (See Table 1.5), we test the hypothesis of no significant difference in their distribution for the months of January to December 1980. Since the observed value of $\chi^2 = 24.475$ is greater than the table value of $\chi^2 = 19.68$ on 11 d.f., (degrees of freedom), we do not accept the preceding hypothesis. Thus, we conclude that there is a significant variation month to month in the occurrence of the arson incidences. We notice that the lowest number of arson incidences took place in the month of January followed by October. On the other hand, the highest number of arson incidences took place in the month of May followed by March, (See figure 4.1). This trend, however, does not seem to continue for the year 1981.

When we consider the four seasons, viz., winter (January-March), spring (April-June), summer (July-September), and fall (October-December), we do not find any significant differences in the distribution of arson incidences ($\chi^2 = 5.03$ on 3 d.f.) Thus, although there is a monthly variation in the occurrence of the arson incidences, there is no seasonal variation. Similarly, even if the total number of arson incidences increased from 263 in the first four months of 1980, to 289 during the corresponding period for 1981, (an increase of about 9.9%) monthly comparisons for these periods incidcate no significant changes in the occurrence of arsons.

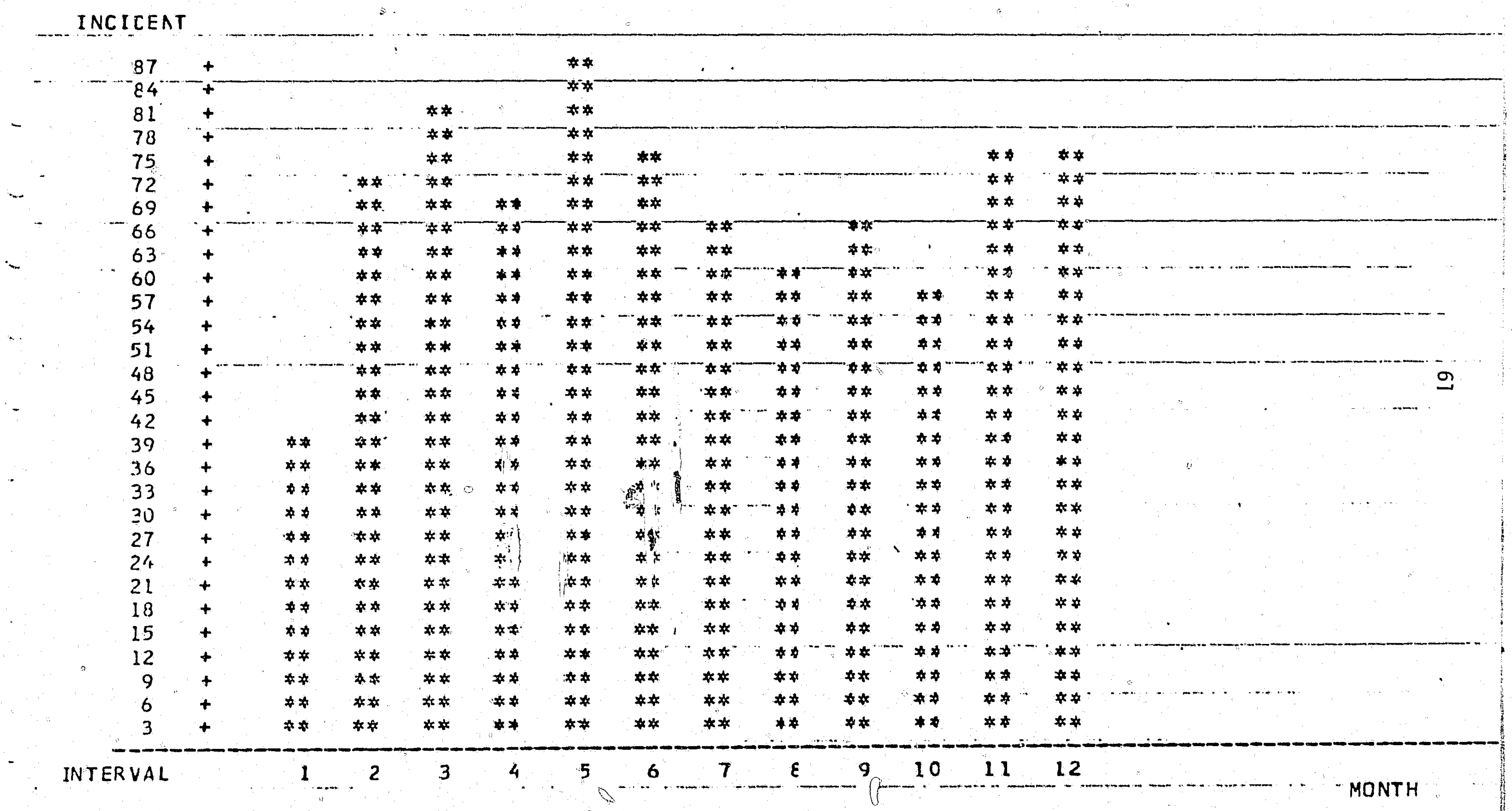
Tabe 4.2 below shows the distribution of arson incidences by day of the week for the year 1980. We notice that the highest number of incidences occur on Mondays followed by Thursdays and Sundays. The lowest number of arsons took place on Wednesdays (See Figure 4.2). We also find that these variations are not significant at the 5% level of confidence ($\chi^2 = 4.506$ on 6 d.f.) so that the arsons are more or less evenly distributed for each day of the week.

TABLE 4.2: Distribution of arson incidences by days of the week for the year 1980

Day	Arson Incidence		
	Number	Percent	Cumulative Percent
Sunday	122	14.63	14.63
Monday	136	16.31	30.94
Tuesday	116	13.91	44.84
Wednesday	107	12.83	57.67
Thursday	125	14.99	72.65
Friday	114	13.67	86.33
Saturday	114	13.67	100.00
Total	834	100.00	–

FIGURE 4.1: BAR GRAPH OF INCIDENT BY MONTH

EACH ** EQUALS 3 INCIDENT (NOTE: "INCIDENT" means number of arson incidences.)



WHERE INTERVAL

1	IS	MONTH	1	WITH INCIDENT	40		2	IS	MONTH	2	WITH INCIDENT	72
3	IS	MONTH	3	WITH INCIDENT	82		4	IS	MONTH	4	WITH INCIDENT	69
5	IS	MONTH	5	WITH INCIDENT	88		6	IS	MONTH	6	WITH INCIDENT	77
7	IS	MONTH	7	WITH INCIDENT	67		8	IS	MONTH	8	WITH INCIDENT	62
9	IS	MONTH	9	WITH INCIDENT	66		10	IS	MONTH	10	WITH INCIDENT	59
11	IS	MONTH	11	WITH INCIDENT	76		12	IS	MONTH	12	WITH INCIDENT	76

FIGURE 4.2: BAR GRAPH OF INCIDENT BY DAY

```
EACH ** EQUALS        4 INCIDENT     (Note: "INCIDENT" means number of arsons incidences.)

 INCIDENT

    136   +            **
    132   +            **
    128   +            **
    124   +            **                **
    120   +      **    **                **
    116   +      **    **    **          **
    112   +      **    **    **          **    **    **
    108   +      **    **    **          **    **    **
    104   +      **    **    **    **    **    **    **
    100   +      **    **    **    **    **    **    **
     96   +      **    **    **    **    **    **    **
     92   +      **    **    **    **    **    **    **
     88   +      **    **    **    **    **    **    **
     84   +      **    **    **    **    **    **    **
     80   +      **    **    **    **    **    **    **
     76   +      **    **    **    **    **    **    **
     72   +      **    **    **    **    **    **    **
     68   +      **    **    **    **    **    **    **
     64   +      **    **    **    **    **    **    **
     60   +      **    **    **    **    **    **    **
     56   +      **    **    **    **    **    **    **
     52   +      **    **    **    **    **    **    **
     48   +      **    **    **    **    **    **    **
     44   +      **    **    **    **    **    **    **
     40   +      **    **    **    **    **    **    **
     36   +      **    **    **    **    **    **    **
     32   +      **    **    **    **    **    **    **
     28   +      **    **    **    **    **    **    **
     24   +      **    **    **    **    **    **    **
     20   +      **    **    **    **    **    **    **
     16   +      **    **    **    **    **    **    **
     12   +      **    **    **    **    **    **    **
      8   +      **    **    **    **    **    **    **
      4   +      **    **    **    **    **    **    **
-------------------------------------------------------------------------------------
INTERVAL         1     2     3     4     5     6     7
                                                                               DAY
WHERE INTERVAL
  1 IS DAY 1  WITH INCIDENT      122          2 IS DAY 2  WITH INCIDENT      135
  3 IS DAY 3  WITH INCIDENT      116          4 IS DAY 4  WITH INCIDENT      107
  5 IS DAY 5  WITH INCIDENT      125          6 IS DAY 6  WITH INCIDENT      114
  7 IS DAY 7  WITH INCIDENT      114
```

The following table, Table 4.3, gives the distribution of arson incidences by month and by day of the week. When we test the hypothesis of no significant differences in the distribution of arson incidences by day, using the Friedman two-way analysis of variance method, we find that $\chi^2 = 5.554$ on 6 d.f., which is less than the table value . Hence we conclude that there are

TABLE 4.3: Distribution of arson incidences by month of the year and by days of the week for the year 1980.

Month	Day of the Week							Total
	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	
January	8	7	3	1	9	6	6	40
February	8	11	9	9	16	12	7	72
March	9	16	13	11	12	8	13	82
April	10	10	10	7	12	10	10	69
May	13	12	11	10	14	13	15	88
June	14	10	7	16	9	7	14	77
July	11	15	14	4	9	7	7	67
August	6	9	4	12	8	13	10	62
September	11	9	13	7	12	8	6	66
October	6	10	6	12	7	14	4	59
November	13	11	10	12	10	8	12	76
December	13	16	16	6	7	8	10	76
Total	122	136	116	107	125	114	114	834

no significant differences in the distribution of arson incidences by month for each day of the week. Again when we test a similar hypothesis for month we find that $\chi^2 = 21.912$ on 6 d.f., which is significant at the 5% level of significance. The preceding results confirm what we have already concluded from the distribution of arson incidences by day and a similar distribution by month. In addition, we conclude from the above test of hypothesis for month that not only significant differences exist in the distribution by month but also by days among these months.

The analysis of variance for the data in Table 4.3 is shown in Table 4.4. We see from the table that the F-value is significant for months while it is not

TABLE 4.4: Analysis of variance of the distribution of arson incidences by months of the year and by days of the week

Source	Degrees of Freedom	Sum of Squares	Mean Sum of Squares	F-Value
Months	11	243.000	23.000	2.481*
Days	6	44.738	7.456	1.243
Error	66	611.833	9.270	
Total	83	899.571		

*Significant at the 5% level of significance.

significant for days. These are the results we already obtained through the χ^2- test. Moreover, as shown below, the statistical test based on studentized range (for testing the differences among means) also shows significant differences in the mean number of arson incidences by month.

In Table 4.5, the mean numbers of arson incidences are arranged in ascending order. Using the method of Newman and Keuls[1] we find that the mean for the month of January only differs significantly from the mean for November, December, June, March and May. However, as can be seen from Table 1.5, this does not seem to be the case for the months of January and March of 1981. Thus we may not find a similar pattern for 1981.

TABLE 4.5: Mean number of arsons per day for the twelve months of 1980, arranged in ascending order.

Month	January	October	August	September	July	April	February	November	December	June	March	May
Mean	5.71	8.44	8.86	9.43	9.57	9.86	10.29	10.86	10.86	11.00	11.71	12.57

1 Snedecor, G.W., and Cochran, W.G., "Statistical Methods", 1967, 6th edition The Iowa State University Press, Ames, Iowa.

4.3 ANALYSIS OF ARSONS BY HOUR

We have seen in Chapter I, see Table 1.11, that the arson incidences decrease from midnight till 9:00 a.m., and increase again till midnight of the next day. Similarly, we observed that there are roughly two and a half times more arsons during the nine hour period 3:00 p.m. - midnight than during the period 3:00 a.m.-12 noon In this section we examine these and similar patterns for the twelve month period of 1980.

TABLE 4.6: Distribution of arson incidences by 3-hour time periods of the day for the year 1980.

Time Period	Frequency	Percent	Cumulative Percent
0-3	100	11.99	11.99
3-6	71	8.51	20.50
6-9	31	3.72	24.22
9-12	72	8.63	32.85
12-15	112	13.43	46.28
15-18	141	16.92	63.20
18-21	148	17.74	80.94
21-24	159	19.06	100.00
Total	834	100.00	-

We notice that the distribution of arson incidences in Table 4.6 above is somewhat similar to the one we observed earlier in Table 1.11. There are 2.3 times more arsons during the hours 12 - 21 than during the hours 3 -12 noon. Similarly, there are about two and a half times more arsons during the hours 15-24 than during the hours 3-12 noon. Furthermore, we notice that there are on an average 58 arsons occurring in any three-hour period from 3:00 a.m. to 12 noon while there are 140 such incidences (an increase of 141.4%) taking place during a three-hour period in the time interval 12 noon through midnight. Application of the Kolmogrov-Smirnov test to the distribution of arson incidences in Table 4.6 shows that it is significantly different from the theoretical distribution, i.e., when the arson incidences are assumed to be uniformly distributed throughout these time periods.

We now consider the distribution of arson incidences by 3-hour time period for each month of 1980 (See Table 4.7 below.) We find that

TABLE 4.7: Distribution of arson incidences by months of the year and by 3-hour time periods of the day for the year 1980.

Month	Arson Incidences for the Hours								Total
	0-3	3-6	6-9	9-12	12-15	15-18	18-21	21-24	
January	4	2	1	7	7	5	3	11	40
February	12	5	3	4	8	9	14	17	72
March	8	9	2	6	15	16	12	14	82
April	12	3	2	3	9	16	10	14	69
May	9	6	3	7	13	20	15	15	88
June	7	10	1	6	6	15	16	16	77
July	8	7	5	6	7	10	14	10	67
August	12	3	3	3	4	7	18	12	62
September	4	4	1	6	6	14	12	19	66
October	6	7	3	9	10	6	10	8	59
November	8	5	3	7	16	10	13	14	76
December	10	10	4	8	11	13	11	9	76
Total	100	71	31	72	112	141	148	159	834

in general there is more arson activity during the time period 3 p.m. to 12 midnight for each month of the year. The analysis of variance for the above data is shown in the following table.

TABLE 4.8: Analysis of variance of the distribution of arson incidences by months of the year and by 3-hour time periods of the day.

Source	Degrees of Freedom	Sum of Squares	Mean Sum of Squares	F
Month	11	212.625	19.330	2.164*
Hour	7	1154.292	164.899	18.464**
Error	77	687.708	8.931	
Total	95	2054.625	—	—

*Significant at the 5% level of significance.
**Significant at the 1% level of significance.

We notice that the F-value is significant for both months and hours. For the variable "month" we have already seen in Sec.4.2 that the mean number of arsons for the month of January differ significantly from those for the months of March, May, June, November and December. We will now make similar comparisons for the variable "hour."

Arranging the mean number of arson incidences per month by hour as in the following table, we apply the Newman and Keul's test. Using

TABLE 4.9: Mean number of arsons per month for the 3-hour time periods of a day, arranged in ascending order.

Hour	6-9	3-6	9-12	0-3	12-15	15-18	18-21	21-24
Mean	2.58	5.92	6.00	8.33	9.33	11.74	12.33	13.25

this test we conclude that the mean number of arson incidences per month differ significantly, at the 5% level of significance, for the following three-hour periods:

(a) For the time period 6-9, from rest of the three hour periods for the day.
(b) For the time period 3-6, from the hours of 12-15, 15-18, 18-21, and 21-24.
(c) For the time period 9-12, from the hours of 12-15, 15-18, 18-21, and 21-24 (same as in (b) above).
(d) For the time period 0-3, from the hours of 15-18, 18-21, and 21-24.
(e) For the time period 12-15, from the hours of 18-21, and 21-24.

We conclude from (a) - (d) above that the number of arsons occurring during the time period midnight to noon is significantly lower than those occurring during the time period noon to midnight. In fact, from Table 4.7 we observe that there are only 274 arsons reported during 0-12 hours and 560 arsons during 12-24 hours. Thus, there are slightly over twice the number of arsons during 12-24 hours as compared to the period 0-12 hours.

We now consider the distribution of arson incidences by day of the week and by hour of the day. This distribution is shown in Table 4.10.

TABLE 4.10: Distribution of arson incidences by days of the week and by 3-hour time periods of the day.

Day	Arson incidences for the hours								Total
	0-3	3-6	6-9	9-12	12-15	15-18	18-21	21-24	
Sunday	9	12	7	11	15	22	19	27	122
Monday	16	5	2	12	28	18	26	29	136
Tuesday	17	10	5	11	13	17	25	18	116
Wednesday	15	9	3	9	13	20	21	17	107
Thursday	10	12	7	16	13	20	26	21	125
Friday	18	11	4	8	19	23	12	19	114
Saturday	15	12	3	5	11	21	19	28	114
Total	100	71	31	72	112	141	148	159	834

We have already seen, through previous analysis, that the days do not differ significantly (see Table 4.4) and that the hours differ significantly among themselves. (see Table 4.8). This is further confirmed in the analysis of variance of the above data shown in Table 4.11 below.

TABLE 4.11 Analysis of variance of the distribution of arson incidences by days of the week and by 3-hour time periods of the day.

Source	Degrees of Freedom	Sum of Squares	Mean Sum of Squares	F
Days	6	67.107	11.184	1.455
Hours	7	1978.786	282.684	17.371 **
Error	42	683.464	16.273	
Total	55	2729.357		

**Significant at 1% level of confidence.

We now determine for which time periods do the mean number of arsons per day differ significantly. We once again use the Newman and Keuls test here. For this purpose the mean number of arsons per day for the three-hour period of a day are shown in ascending order in Table 4.12 below.

TABLE 4.12: Mean numbers of arsons per day for the 3-hour time periods of the day, arranged in ascending order.

Hour	6-9	3-6	9-12	0-3	12-15	15-18	18-21	21-24
Mean	4.43	10.14	10.28	14.28	16.00	20.14	21.24	22.71

Since the conclusions for the variable "day" are exactly similar to those obtained in (a) - (e) for the variable "month," we do not repeat them here. Thus, we conclude that the average number of arsons per day or average number of arsons per month follow the same pattern viz., about twice the arsons occur during the time period 12 noon to midnight than during the period midnight to 12 noon.

4.4: ANALYSIS OF ARSONS BY CENSUS TRACTS

We now examine patterns of arsons, if any, by census tract. Since there are large number of census tracts for the City of Newark, numbering about 100, we will analyze the data in some detail for those census tracts that experienced more arsons. The following table, Table 4.13, shows the number of census tracts experiencing one or more arsons. Since repeated arsons of the same structure is excluded from the number of arsons in this table, the number of arsons for a census tract is equal to the number of structures arsoned in that census tract. For sake of simplicity

TABLE 4.13: Distribution of census tracts (CT) by the number of structures arsoned for the first time during the year 1980

Number of Arsons	Number of CTs	Cumulative Number of CTs	Number of Arsons	Number of CT	Cumulative Number of CTs
0	11	11	11	4	78
1	9	20	12	4	82
2	7	27	13	2	84
3	10	37	14	2	86
4	5	42	16	1	87
5	6	48	17	5	92
6	6	54	18	1	93
7	6	60	19	1	94
8	4	64	20	1	95
9	3	67	22	1	96
10	7	74	25	1	97
			29	1	98

we have also omitted in this table percentages and cumulative percentages because the total number of census tracts 98 is close to 100. Thus, almost 25% of the census tracts in Newark had eleven or more structures experiencing arsons during 1980. On the other hand, there were about 25% of the census tracts with fewer than three structures experiencing arsons in each one of them, with about 11% of the census tracts experiencing no arson at all.

Table 4.14 gives the distribution of census tracts by intervals of five for arsons and is similar to Table 1.2 in presentation. Using the Kolmogrov-Smirnov test, we find that there are no significant differences in the distribution of census tracts as given is Tables 1.2 and 4.14.

TABLE 4.14: Distribution of census tracts (CT), grouped by number of arson incidences.

Number of Structures Arsoned	Number of CT	Cumulative Number of CT	Percent to the Total	Cumulative Percent
0	11	11	11.2	11.2
1-5	37	48	37.8	49.0
6-10	26	74	26.5	75.5
11-15	12	86	12.2	87.7
16-20	9	95	9.2	96.9
21-29	3	98	3.1	100.0
Total	98	-	100.0	-

Up until now we considered the number of structures arsoned by census tracts and the distribution of arson incidences by the specific range of arsoned structures. We consider in Table 4.15 the number of arsons that took place during 1980 in all 98 census tracts.

TABLE 4.15: Number of arson incidences by census tracts (CT) for the year 1980.

Number of Arsons	Number of census tracts	Percent of census tracts	Cumulative Percent
0	11	11.2	11.2
1-5	35	35.7	46.9
6-10	17	17.3	64.2
11-15	16	16.3	80.5
16-20	8	8.2	88.7
21-25	8	8.2	96.9
26-38	3	3.1	100.0
Total	98	100.0	-

From the above table we notice that almost one-third of the census tracts experienced 11 or more arsons. Moreover, one out of every eight census tracts had more than twenty arsons. We will analyze these census tracts in some detail in this section after analyzing the information for fire districts.

The distribution of the arson incidences for the five fire districts (FDs) by months is given in Table 4.16. We notice that fire districts II and V have fewer arsons each month as compared to the rest of the fire districts. This is largely due to the smaller number of structures and a lesser population in these districts. We also notice that the number of arson incidences vary considerably from fire district to fire district. The analysis of variance

TABLE 4.16: Distribution of arson incidences, by months of the year and by fire districts, for the year 1980.

Month	Fire District					Total
	I	II	III	IV	V	
January	10	3	8	17	2	40
February	18	14	21	15	4	72
March	17	11	19	31	4	82
April	17	7	14	26	5	69
May	18	9	18	37	6	88
June	16	12	22	24	3	77
July	17	8	19	19	4	67
August	16	8	14	23	1	62
September	17	8	21	18	2	66
October	16	4	17	22	0	59
November	27	8	14	25	2	76
December	23	10	15	22	6	76
Total	212	102	202	279	39	834

presented in Table 4.17 also brings out this fact because even at the 0.5% level of significance the F-value for fire districts is highly significant.

TABLE 4.17: Analysis of variance of arson incidences, by months of the year and by fire districts.

Source	Degrees of Freedom	Sum of Squares	Mean Sum of Squares	F-Value
Month	11	340.2	30.927	4.339*
Fire Districts	4	3303.567	825.892	115.866**
Error	44	313.633	7.128	
Total	59	3957.4		

*F-value significant at the 5% level of significance.
**F-value significant at the 1% level of significance.

When we test the mean number of arsons for fire districts, we find that (See Table 4.18) there is no significant difference in the mean number of arsons for fire districts I and III. However, fire districts II and V differ significantly, at the 5% level of significance, from fire districts I, III, and IV. Similarly, fire districts I and III differ significantly from fire district IV, and fire districts II differs significantly from fire district V.

TABLE 4.18: Mean number of arsons for fire districts, arranged in ascending order.

Fire District Mean	V 3.25	II 8.5	II 16.833	I 17.667	IV 23.25

We now examine how the arson incidences occur by month within each fire district. That is, for which months of the year 1980 there are more arsons in a fire district. Such information is useful for planning and utilizing manpower and it seems appropriate to consider it here.

Table 4.19 gives the months for each fire district ranked (within each fire district) in the descending order of magnitudes of the number of arsons for each month. Thus, rank 1 means the highest number of arsons for the month shown against it for a fire district. Then for rank 2, the month shown against it in the same fire district will have fewer arsons for that month than the month ranked 1 but greater than any of the remaining of the months and so on. For convenience, months are numbered 1-12 using the standard convention for months i.e., January = 1, February = 2 and so on.

TABLE 4.19: Arson incidences ranked in descending order of within each fire district, for twelve the twelve months of 1980.
(Note: Numbers in the body of table are months.)

Rank	Fire District I	II	III	IV	V
1	11	2	6	5	5
2	12	6	2	3	12
3	2	3	9	4	4
4	5	12	3	11	2
5	3	5	7	6	3
6	4	7	5	8	7
7	7	8	10	10	6
8	9	9	12	12	1
9	6	11	4	7	9
10	8	4	8	9	11
11	10	10	11	1	8
12	1	1	1	2	10

} - Symbol means the number of arson incidences for these months are the same.

As stated earlier, the information in Table 4.19 can be employed to plan and use available resources. Since the ranks are assigned within each fire district, they provide no clue to the relative magnitude of ranks for any two fire districts. For example, the month of May experienced the highest number of arsons for fire districts IV and V, and these are respectively 37 and 6. Therefore, it will not be advisable to allocate equal manpower to these fire districts for May on the basis of the ranks. Thus, for inter-district allocation of resources for each month, it will be necessary to use Table 4.16.

To illustrate how information in Tables 4.16 and 4.19 can be used, consider fire districts I and III. We notice from Table 4.19 that the highest number of arson incidences occurred in the months of November and June for fire districts I and III respectively. Thus, it may be possible to use the resources of fire district I for fire district III during the month of June. Assuming that all is the same for both these fire districts, except for the occurrence of the arson incidences, we notice from Table 4.16 that fire district I needs about one and a half times more resources during November than that of in the month of June (because there are one and a half times more arsons in November than in June). Similarly, in the case of fire district III, we need about one and a half times more resources during June than in the month of November. Thus, it may be possible to use some resources from fire district I for fire district III during June and similarly use some resources from fire district III for fire district I during November.

We now consider the distribution of census tracts which experienced more than 20 arsons during 1980. There were 11 such census tracts with 278 arson incidences, which was about one-third of all the arson incidences during 1980. Table 4.20 gives their distributions by month. The months of January, April, October, November, and December had 73 arson incidences which is about one-fourth of all arson incidences for these census tracts. The average number of arsons for these five months is 15 and for the rest of the seven months it is 29, which is significantly higher than the preceding statistic. The months of February, March, and June experienced the most arsons, followed by May, July, August, and September.

TABLE 4.20: Distribution of arson incidences for the census tracts (CT) with more than 20 arsons during the year 1980.

Month	FD-I Census Tract Number			FD-II Census Tract Number		FD-III Census Tract Number				FD-IV Census Tract Number		Total
	9	14	24	57	67	86	92	96	97	35	49	
January	1	1	1	-	1	1	1	1	1	3	-	11
February	-	3	2	4	2	1	9	3	3	5	2	34
March	4	-	2	4	2	6	2	3	3	3	1	30
April	2	2	-	1	1	-	3	-	1	5	-	15
May	3	2	2	2	4	-	1	2	6	2	5	29
June	2	3	5	4	5	4	4	2	-	-	4	33
July	2	2	5	1	2	4	4	1	2	-	2	25
August	-	4	1	4	1	5	1	2	1	7	1	27
September	1	3	-	-	4	7	3	3	2	3	1	27
October	4	-	1	-	-	-	4	-	-	2	2	13
November	-	2	1	2	1	-	4	1	1	-	4	16
December	2	1	1	1	2	1	2	3	2	3	-	18
Total	21	23	21	23	25	29	38	21	22	33	22	278

Note: " -" indicates no arsons.

In the following table, Table 4.21 we give the total number of arson incidences for the census tracts in Table 4.20 by fire districts and for each month of 1980. The second part of this table gives months ranked by the number of arsons in descending order of magnitude with rank 1 assigned within a fire district to the month for which the highest number of arsons occurred and so on. We notice from Tables 4.19 and 4.21 that the ranks for all arsons and those for selected census tracts differ. We will now examine the correlations

TABLE 4.21: Distribution of arson incidences and their ranks by fire districts and by months for census tracts (CT) experiencing more than 20 arsons during the year 1980.

Month	Number of arsons for F.D.					Rank	Months ranked by arsons for F.D.			
	I	II	III	IV	Total		I	II	III	IV
January	3	1	4	3	11	1	6	6	2	8
February	5	6	16	7	34	2	7	2	9	2
March	6	6	14	4	30	3	5	3	3	5
April	4	2	4	5	15	4	3	5	7	4
May	7	6	9	7	29	5	2	8	6	3
June	10	9	10	4	33	6	8	9	5	6
July	9	3	11	2	25	7	10	7	8	9
August	5	5	9	8	27	8	4	11	12	10
September	4	4	15	4	27	9	9	12	11	11
October	5	-	4	4	13	10	12	4	1	1
November	3	3	6	4	16	11	1	1	4	12
December	4	3	8	3	18	12	11	10	10	7
Total	65	48	110	55	278					

} - Symbol means the number of arson incidences for these months are the same.

of the these ranks. For fire districts I and IV the Speerman rank correlations are $r_s = 0.112$ and $r_s = 0.357$ respectively. Since these correlations are not significant, we conclude that the arsons taking place throughout these districts are not necessarily affected by arsons occurring in these census tracts. Furthermore, in the context of planning and utilizing the manpower within these districts, the above results can also be interpreted as follows: Any efforts to reduce arson throughout these districts will have little effect on the number of arsons occurring in the selected census tracts. Hence it may be helpful to make separate efforts, assuming these census tracts as separate entities. This is, however, not the case for fire districts II and III. Here we have $r_s = 0.874$ and $r_s = 0.816$ for fire districts II and III respectively. Moreover, both of these correlations are highly significant at the 5% level of significance. Hence if we make efforts to reduce arson for the district as a whole we will also be reducing arson for the selected census tracts and vice versa. These results seem to be consistent with the fact that out of the total number of 102 arsons in fire district II, (See Table 4.16), 48 nave occurred in the selected census tracts. Similarly, for fire district III the corresponding figures are 202 and 110.

Here we have considered the distribution of arson incidences by month and census tracts and shown how we can profit from the analysis of this data in planning to reduce arson. Similar analysis can be done using the variables day and hour instead of month. However, it seems that the analysis presented in Sec. 4.2 and Sec. 4.3 for these variables should prove helpful along with the above analysis for planning purposes.

4.5 ANALYSIS OF ARSONS BY LUNAR TRENDS:

Lunar cycle often affects the human mental behavior and especially so for those mentally ill. Icove and Keith[1] reported that these lunar cycles also affect arson activities. For example, they report that 69% of incendiary fires took place within a ±4 day period surrounding the full moon phases. Also, 55% of all the incendiary fires were during the full moon phase. In this section we will examine whether similar trends exist for Newark.

In Table 4.22 we present the total number of arson incidences for the four day period before and after the new/full moon phase, for the new/full moon day, and for the remainder of the month. Since the full moon was on January 2, 1980, only one day before it is considered instead of the usual four day period for the month of January. Similarly, for the month of March 1980 there were two full moon days and both are considered. Also, for comparisons all four days before or after the new/full moon day were included in the month in which the new/full moon day occurred even though portions of these days belonged to the preceding or the following month.

In reading Table 4.22 it should be noted that the totals for before and after new/full moon days are for four days and hence they are generally much larger than the corresponding new/full day. We find from this table that except for the month of June, the percent of arson incidences during the lunar phases exceeds 50% and reaches up to 70%. (Here we have excluded the percentage for the month of March which had 14 days of lunar phase instead of the usual nine days.) Overall, 60% of the arsons occur during 18 days of a calendar month. That is, six-tenth of the arsons occur during the six-tenths of the month that represents lunar phases.

We note from Table 4.22 that 340 arson incidences occurred during the non-lunar phases days totalling 148 days. Similarly, 502 arson incidences occurred during 218 (218=366-148) days of lunar-phases in 1980. Thus, about 2.3 arsons occurred per day during the lunar-phases as well as non-lunar-phases days. Hence there is no difference in the occurrence of arson incidences per day during the lunar-phases and the non-lunar-phases days for the year 1980. A closer look at these distributions is taken later in Table 4.23.

[1] D.J. Icove, and P.E.Keith (1981): Principles of Incendiary Crime Analysis: Approach to arson information management.

CONTINUED

1 OF 2

TABLE 4.22: Distribution of arson incidences by lunar phases and by months for the year 1980.

Month	Remainder of the Month		No. of Arsons for the New Moon Phase				No. of Arsons for Full Moon Phase				Total No.of Arsons for Moon Phases	Percent to total for the month
	No.of Arsons	No.of Days	Four Days before New Moon	New Moon Day	Four days after New Moon	Total	Four Days before Full Moon	Full Moon Day	Four Days After Full Moon	Total		
January	12	12	3	0	12	15	2	0	11	13	28	70.0
February	27	11	13	0	7	20	2	4	11	17	37	51.4
							10	2	6	18		
							10	3	18	31	76	85.4*
March	22	12	14	4	9	27	7	1	6	14	36	52.2
April	33	12	8	2	12	22	14	7	9	30	50	56.8
May	43	15	11	2	7	20	9	3	3	15	37	48.0
June	36	12	12	1	9	22	8	2	8	18	44	65.7
July	22	11	7	1	18	26	8	2	7	17	32	51.6
August	30	13	7	0	8	15	5	2	14	21	42	63.6
September	24	12	10	3	8	21	12	0	10	22	32	54.2
October	27	13	4	0	6	10	11	2	8	21	42	55.3
November	34	12	9	3	9	21	5	4	12	21	46	60.5
December	30	13	7	5	13	25						60.2
Total	340	148	105	21	118	244	103	32	123	258	502	

*Does not include incidences that occurred in April.

Although, Kolmogrov-Smirnov test shows that the distribution by month for new moon and full moon days differs significantly at the 5% level of significance, the mean number of arson incidences for these days do not differ significantly ($t = 0.994$; d.f. = 23).

To compare the arson activities before or after new/full moon days with the corresponding new/full moon days, we consider the mean number of arson incidences for these days as shown in Table 4.23. Here we test the hypothesis of no significant difference in the sampling distribution of the number of arsons for time periods being compared.

The Kolmogrov-Smirnov test shows significant differences between the distributions of the mean number of arsons for the four day period before new moon and that on the day of the new moon. A similar conclusion is also drawn when we compare the distribution of the mean number of arsons for four-day period after the new moon. When we compare the mean number of arsons for the rest of the month and the new moon, we find that the respective distributions also differ significantly. Hence all the preceding distributions

TABLE 4.23: Mean number of arson incidences for four days before and after new/full moon day, and for the remainder of the month.

Month	Remainder of the Month	Mean Number of Arson for New Moon Phase				Mean Number of Arsons for Full Moon Phase				Mean for all Days of Moon Phases
		Four Days Before New Moon	New Moon Day	Four Days After New Moon	All Nine Days of New Moon Phase	Four Days Before Full Moon	Full Moon Day	Four Days After Full Moon	All Nine Days of Full Moon Phase	
January	1.00	0.75	0	3.00	1.67	0.50	0	2.75	1.44	1.56
February	2.45	3.25	0	1.75	2.22	0.50	4.00	2.75	1.89	2.06
March	1.83	3.50	4.00	2.25	3.00	2.50	2.50	3.00	2.72	2.81
April	2.75	2.00	2.00	3.00	2.44	1.75	1.00	1.50	1.56	2.00
May	2.87	2.75	2.00	1.75	2.22	3.50	7.00	2.25	3.33	2.78
June	3.00	3.00	1.00	2.25	2.44	2.25	3.00	0.75	1.67	2.06
July	2.00	1.75	1.00	4.50	2.89	2.00	2.00	2.00	2.00	2.44
August	2.31	1.75	0	2.00	1.67	2.00	2.00	1.75	1.89	1.78
September	2.00	2.50	3.00	2.00	2.33	1.25	2.00	3.50	2.33	2.33
October	2.08	1.00	0	1.50	1.11	3.00	0	2.50	2.44	1.78
November	2.83	2.25	3.00	2.25	2.33	2.75	2.00	2.00	2.33	2.33
December	2.31	1.75	5.00	3.25	2.78	1.25	4.00	3.00	2.33	2.56
Mean*	2.30	2.19	1.75	2.46	2.26	1.98	2.46	2.36	2.20	2.14

* Mean = (Total no. of arsons)/(Total no. of days) and is obtained from Table 4.22.

i.e., those corresponding to mean number of arsons before and after new moon day, and for the remainder of the month, are coming from the same population and each one of these distributions differs significantly from the distribution of the number of arson incidences on the new moon day.

We also notice from Table 4.23 that at the most for five months the number of arsons on the new moon day exceed the mean number of arsons before or after that day. Hence the data for the four day period before and after new moon does not support the general findings that the peak of arson activities occurs on the new moon day, (See Icove and Keith.)[1] In fact, the Kolmogrov-Smirnov test shows that there are more arson activities on days before and after the new moon than on the new moon day. The overall mean for the four day period before or after the new moon does not differ significantly from the overall mean for the new moon day. Thus, the mean number of arson incidences for the four day period before or after the full moon day is somewhat similar to the number of arsons occurring on the new moon day.

[1]D.J. Icove and P.E. Keith (1981), op.cit.

Comparison of the distribution of the mean number of arson incidences for the four-day period before the full moon to a similar distribution for the full moon day shows significant differences between these two distributions. However, there exists no significant difference between the distribution for the number of arsons on the full moon day and the mean number of arsons for a four-day period after the full moon. Similar conclusion can be drawn when we compare the distributions of arsons for the full moon day to that of the average number of arsons for the remainder of the month. Finally, we conclude from the Kolmogrov-Smirnov test that there is more arson activity on the full moon day than the average for the preceding four days only. Thus, the data for Newark does not support that there are more arson activities on the full moon day than the average for the four days following it.

4.6 SUMMARY

In this chapter we have analyzed the data for the variables "month," "day," "hour," and "census tracts" to find if there exist any specific patterns for committing arsons in Newark. We found that, although the month of January experienced the lowest arson activities during 1980, such was not the case during 1981. This suggests that there are no significant differences in the arson activities by month during 1980. Similarly, we find that there are no significant seasonal variations or day to day variations in the arson activities. Also, the data do not support that there is an increased arson activity on weekends than on the week days. On the other hand, more arsons take place on Mondays, Thursdays and Fridays than on Saturdays and Sundays. However, significant differences do exist in the commission of arsons for different hours of each day of the week. For example, there are twice the number of arsons during the twelve-hour time period from twelve noon to midnight than during the time period midnight to twelve noon. There are also substantial and significant differences in arson activities for shorter time intervals than the said twelve-hour periods.

The analysis of arsons by structure shows that about fifty new structures are torched every month. About 72% of all arsons are one time arsons during the year,affecting 87% of all the arsoned structures. This clearly suggests that arson is wide spread in Newark. Had the case been otherwise, 50% or fewer structures would have been affected by a single arson during this period. Another indication of the widespread arson is the fact that almost 90% of the census tracts are affected by it.

Almost 83% of the arsons occur in fire districts I, III, and IV, with one-third (33.4%) occurring in fire district IV. Thus, fire district IV experienced significantly high proportions of arsons during 1980.

In our earlier analysis we observed that there are no significant differences in the distribution of arsons by month. However, analysis of the data for census tracts experiencing more than twenty arsons during 1980 shows that there are significant differences in the distribution of the number of arson incidences by month. During the year, the months of February, March, and June experienced the most arsons for these census tracts.

Finally, the analysis of lunar trends shows that six-tenths of the arsons occur during six-tenths of the month which is the period corresponding to the full and new moon phases. Further, there is more arson activity on the full-moon day than on the new-moon day, but the mean number of arsons for each month are not significantly different for these two days.

There are higher mean numbers of arsons during the remainder of the month, the four-day period before the new moon, or four-day period after the new moon than on the new moon-day itself. This trend for Newark is quite contrary to the generally observed trend. For the full-moon phase we found that there are significant differences in the distributions of the mean number of arsons for the full-moon day as compared to a four-day period before the full moon. In this case, the data does support the general finding that there are increased arson activities on the full-moon day than on the four-day period before the full-moon day.

CHAPTER V

REPEATED ARSONS

5.1 INTRODUCTION

In the last chapter we studied the patterns of arson for structures experiencing a single arson during 1980 as well as those structures that were arsoned more than once during the same period (called repeated arsons here). We have also seen in Table 4.1 that 13% of the population of structures experienced 28% of the arsons during 1980. Each structure among these experienced two or more arsons during this period. It may be interesting to study the characteristics of these arsons relative to the variables of month, day, hour, and census tracts of occurrence of these arsons. Also, to study patterns for these arsons, we consider only those repeated arsons that occurred during the year 1980 and not the entire study period. The phrase "repeated arsons" is used here to mean arsons of the same structure more than once. Also, "all arsons" or "all arson incidences" means arsons of all structures be it once or more than once.

There were in all 261 repeated arsons during 1980 and accounted for about 31% of the total number of 834 arsons during the same period. These 261 arsons occurred in 90 new structures that were ignited during 1980. Thus, every new structure that experienced more than one arson during 1980 was ignited on an average of three times during 1980. From Table 4.1 we see that about 73% of the structures among those having two or more arsons experienced two arsons and about 12% experienced three arsons during 1980. This also indicates that arson is widespread in Newark. Otherwise, we would have expected much higher frequency of repeated arsons for a small number of structures.

5.2 ANALYSIS OF REPEATED ARSONS BY MONTH AND BY DAY.

In Table 5.1 the distribution of repeated arsons by month is given. The month of January has the lowest number of arson incidences in the case of repeated arsons (Table 5.1) as well as in the case of all arsons (Table 1.5). However, for other months the order is quite different. For example, the highest number of arsons, 88 occur in the month of May in the case of all arsons while November experienced the highest number of repeated arsons. (However, the Kolmogrov-Smirnov test shows that there is no significant difference between the two distributions.) Moreover, since the Spearman's rank coorelation coefficient r_s = 0.547 is significant at the 5% level of significance, we conclude that there is a significant correlation between the number of all arson incidences occurring every month and the number of repeated arsons occurring in that month. Thus, any efforts to reduce all arsons will also help in reducing the incidences of repeated arsons. Finally, we notice that there is no significant

difference in the occurrence of the number of repeated arsons for the four seasons of a year, i.e., winter, spring, summer and fall (χ^2= 3.077, on 3 d.f.).

TABLE 5.1:- Distribution of repeated arson by months for the year 1980.

Month	Repeated Arsons		
	Number	Percent	Cumulative percent
January	8	3.06	3.06
February	24	9.20	12.26
March	21	8.04	20.30
April	12	4.60	24.90
May	26	9.96	34.86
June	31	11.88	46.74
July	23	8.81	55.55
August	22	8.43	63.98
September	25	9.58	73.56
October	15	5.75	79.31
November	34	13.03	92.34
December	20	7.66	100.00
Total	261	100.00	--

Similarly, the largest number of repeated arsons occur on Sundays, (See Table 5.2) while the largest number of all arsons occurs on Mondays. Thus, we find that the pattern of arson incidences is somewhat different for the days of a week. The Kolmogrov-Smirnov test does not indicate any significant difference between the distribution by day of repeated arsons and all arsons. Also, we find that the Spearman's rank correlation coefficient r_s = 0.5804 is not significant at the 5% level of significance. This indicates that the number of incidences of repeated arsons is not significantly related to the number of incidences of all arsons. In other words, a measure to reduce all arsons by day of the week will not bring about a significant reduction in the number of repeated arsons by day of the week.

TABLE 5.2:- Distribution of repeated arsons by day of the week for the year 1980.

Day	Repeated Arson		
	Number	Percent	Cumulative percent
Sunday	49	18.77	18.77
Monday	41	15.71	34.48
Tuesday	37	14.18	48.66
Wednesday	32	12.26	60.92
Thursday	34	13.02	73.95
Friday	30	11.49	85.44
Saturday	38	14.56	100.00
Total	261	100.00	--

Let us now consider the distribution of repeated arsons by month and by day as shown in Table 5.3. Using the Friedman's two-way analysis of variance technique, we find that χ^2 = 22.005 (on 6 d.f.) is significant at the 5% level of significance. Thus, we reject the hypothesis that there are no significant differences in the occurrence of repeated arsons every month when their distributions are considered by each day of the week. In other words, there are significant differences in the occurrence of the repeated arsons by each day of the week for the twelve months of the year 1980. For example, there were more arsons on Sundays, Mondays or Tuesdays than on any other day of the week during December. Similarly, Saturdays or Sundays experienced more arsons during June than any other day of the week. When we test a similar hypothesis for the days of the week, viz., that there are no significant differences in the distribution of repeated arsons for days of the week when these distributions are considered monthwise, then since the observed value of χ^2= 6.768, (on 1 d.f.) is not significant at the 5% level of significance, we conclude that there is no significant difference in the distribution of repeated arsons by days of the week for each month of the year. It may be recalled that we reached exactly similar conclusions when we tested the preceding type of hypotheses for all arsons.

TABLE 5.3:- Distribution of repeated arsons by months of the year and by days of the week for the year 1980.

Month	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Total
January	4	1	0	0	2	1	0	8
February	4	1	4	5	4	3	3	24
March	2	4	1	3	2	3	6	21
April	3	2	1	1	2	1	2	12
May	3	4	3	2	3	5	6	26
June	9	3	4	3	2	2	8	31
July	2	6	8	3	2	1	1	23
August	4	3	2	3	2	6	2	22
September	5	2	4	3	6	4	1	25
October	1	3	1	5	3	1	1	15
November	7	7	4	4	4	2	6	34
December	5	5	5	0	2	1	2	20
Total	49	41	37	32	34	30	38	261

The analysis of variance of the above data is shown in the following table. It confirms that there is a significant difference in the number of repeated arsons occurring each month.

TABLE 5.4:- Analysis of variance of the distribution of repeated arsons by months of the year and by days of the week.

Source	Degrees of Freedom	Sum of Squares	Mean Sum of Squares	F-value
Months	11	86.321	7.847	2.360*
Days	6	20.286	3.281	1.017
Error	66	219.429	3.325	
Total	83	326.036		

* Significant at the 5% level of significance.

The following table , Table 5.5, gives the mean number of repeated arsons per day by months, arranged in the ascending order of magnitude. Using the Newman and Keuls test we find that the mean number of repeated arsons for the months of June and November are significantly different from the corresponding figure for the month of January. This result is different from what we observed for all arsons. There the mean number of arsons for January differed significantly from the months of March, May, June, November and December. Since the preceding result did not hold true for the year 1981 (See Sec. 4.2) it is expected that in the case of repeated arsons, the number of arson incidences will not be significantly different from the rest of the study months of 1981.

TABLE 5.5:- Mean number of repeated arsons per day, for the twelve months of 1980, arranged in ascending order.

Month	Jan.	Feb.	Oct.	Dec.	Mar.	Aug.	July	Feb.	Sept.	May	June	Nov
Mean	1.14	1.71	2.14	2.86	3.00	3.14	3.29	3.43	3.57	3.71	4.43	4.86

5.3 ANALYSIS OF REPEATED ARSONS BY HOUR:

We now consider the distribution of repeated arsons by three-hour periods, which is shown in Table 5.6. Using the Kolmogrov-Smirnov test we find that there is no significant difference between the distribution of all arsons by three-hour period of the day (See Table 4.6) and a similar distribution for repeated arsons during 1980, given in Table 5.6. There is also a highly significant correlation between all arsons and repeated arsons for the three-hour periods of the day (Spearman's rank correlation coefficient $r_s = 0.9524$). Furthermore, we notice from Table 5.6 that about 2.7 times repeated arsons occur during the hours of 12 noon to 9 p.m. as compared to the hours 3 a.m. - 12 noon. We recall that a similar statistic for all arsons is 2.3. Thus, we observe that the distribution of all arsons and repeated arsons are very similar for the three-hour time periods.

TABLE 5.6:- Distribution of repeated arson incidences by 3-hour time periods of the day for the year 1980.

Time Period	Repeated Arsons Frequency	Percent	Cumulative Percent
0-3	24	9.20	9.20
3-6	22	8.43	17.63
6-9	12	4.60	22.23
9-12	17	6.51	28.74
12-15	37	14.18	42.92
15-18	46	17.62	60.54
18-21	54	20.69	81.23
21-24	49	18.77	100.00
Total	261	100.00	--

We now consider the distribution of repeated arsons for the three-hour time periods by each month of the year 1980, as shown in Table 5.7 below.

TABLE 5.7:- Distribution of repeated arsons by month and by 3-hour time periods of the day for the year 1980.

	Repeated Arsons for the Hours								
Month	0-3	3-6	6-9	9-12	12-15	15-18	18-21	21-24	Total
January	0	0	1	2	4	0	0	1	8
February	3	3	2	0	3	3	7	3	24
March	1	2	0	1	5	5	2	5	21
April	2	0	0	1	1	3	2	3	12
May	3	1	1	3	2	6	5	5	26
June	3	5	1	2	2	8	5	5	31
July	2	1	1	1	5	5	4	4	23
August	2	2	1	0	3	2	9	3	22
September	2	2	1	0	3	6	4	7	25
October	4	1	3	1	1	0	3	2	15
November	2	3	1	2	7	4	8	7	34
December	0	2	0	4	1	4	5	4	20
Total	24	22	12	17	37	46	54	49	261

Using Friedman's two-way analysis of variance method, we find that the distributions of repeated arsons by months differ significantly even at the 1% level of significance (χ^2 = 32.521 on 7 d.f.) for the three-hour periods. Also, using the same method we find that the distribution of repeated arsons by three-hour time periods differ significantly, at the 5% level of confidence, for the twelve months of 1980.

TABLE 5.8:- Analysis of variance of the repeated arsons by months and by 3-hour time periods of the day.

Source	Degrees of Freedom	Sum of Squares	Mean Sum of Squares	F-value
Months	11	75.531	6.866	2.623***
Hours	7	148.323	21.189	8.094**
Error	79	201.552	2.618	
Total	95	425.406		

** Significant at the 1% level of significance.

We have already seen in Sec. 4.2 that months differ significantly in the distribution of repeated arsons and also we determined months for which the mean number of repeated arsons differ significantly from other months of the year 1980. In the analysis of variance above, Table 5.8, we again find that the mean numbers of repeated arsons are significantly different since the F-value is significant at the 1% level of significance. We also observed, as should be expected, that the mean number of repeated arsons by three-hour period are significantly different. We now find out for which time periods these mean values differ significantly. Using the Newman and Keuls test we arrive at the following conclusions regarding the mean number of repeated arsons for the three-hour periods (See Table 5.9):

a) The mean for the hours 6-9 differs significantly from the corresponding values for the hours 12-15, 15-18, 21-24, and 18-21.

b) The mean for the time period 9-12 differs significantly from the corresponding values for the hours 15-18, 21-24, and 18-21.

c) The mean for the time period 3-6 differs significantly from the corresponding values for the hours 15-18, 21-24 and 18-21.

d) The mean for the time period 0-3 differs significantly from the corresponding values for the hours 15-18, 21-24 and 18-21.

We notice that these results are somewhat similar to what we observed for all arsons, (See Ch.IV, pp.67).

TABLE 5.9:- Mean numbers of repeated arsons by 3-hour time periods of the day arranged in ascending order.

Time Period	6-9	9-12	3.6	0-3	12-15	15-18	21-24	18-21
Mean	1.00	1.42	1.83	2.00	3.08	3.83	4.08	4.50

5.4 ANALYSIS OF REPEATED ARSONS BY CENSUS TRACTS

We now consider analysis of the repeated arsons by census tracts. As shown in Table 5.10, 43 census tracts had no such arsons. Nine census tracts that experienced only one arson during 1980 were also arsoned sometime in the first four months of 1981. There are six census tracts that had 10 or more arsons during 1980. These will be considered later in this section.

TABLE 5.10:- Distribution of census tracts (CTs) by number of repeated arsons

No.of Arsons	No.of CTs.	Cumulative No.of CTs	No.of Arsons	No.of CTs	Cumulative No.of CTs	No.of Arsons	No.of CTs	Cumulative No.of CTs
0	43	43	4	3	78	8	6	89
1	9	52	5	1	79	9	3	92
2	18	70	6	3	82	10/over	6	98
3	65	75	7	1	83	—	—	—
Total	—	—	—	—	—	—	98	—

As should be expected, the distribution of census tracts by number of repeated arsons is not similar to that for all arsons (See Table 4.15). A very large proportion of census tracts did not have any repeated arsons during 1980. Out of the 43 such census tracts in the above table, 11 census tracts had no arsons at all as shown in Table 4.15. The remaining 32 census tracts had only one arson during 1980 and hence, by definition of repeated arsons, are shown to have no arsons in the above table.

Consider now the distribution of all arsons and repeated arsons by fire districts. Roughly about one-third of the arsons are repeated arsons for FDs I-IV combined. There are significant differences in the occurrence of repeated arsons for these four district. (χ^2 =19.98 on 3 d.f.), with FD II having almost half the number of repeated arsons than each of the remaining three fire districts. Similarly, the number of repeated arsons for FD V is significantly lower than the rest of the FDs.

TABLE 5.11:- Distribution of all arsons and repeated arsons by fire districts (FDs)

Fire District	Number of All Arsons	Number of Repeat Arsons	Percent*
I	212	63	29.7
II	102	36	35.3
III	202	75	37.1
IV	279	83	29.7
V	39	4	10.2
Total	834	261	31.3

* Percent of repeated arsons to all arsons.

In Table 5.12 below, we consider census tracts with 10 or more repeated arsons. We notice from this table that for most of these census tracts repeated arsons accounted for a major portion of the arsons. Thus, using the approach suggested in Sec. 4.4, we can plan the resources to reduce such repeated arsons.

TABLE 5.12:- Census tracts (CT) with 10 or more repeated arsons.

Fire District	Census Tract	Number of all arsons	Number of Repeated Arsons	Percent*
I	82	19	12	83.2
II	57	23	10	43.5
	67	25	14	56.0
Total FD II	--	48	24	50.0
III	86	29	16	55.2
	92	38	22	57.9
Total FD III	--	67	38	56.7
IV	31	18	10	55.6
Total FD I-IV	--	152	84	55.3

* Percent of repeated arsons to all arsons for the census tract.

5.5 ANALYSIS OF SUBSEQUENT ARSONS

Up until now we analyzed patterns of repeated arsons in terms of months, days, hours of occurrence and by census tracts. However, often it is useful to know when the next arson for the same structure will occur. In what follows we ascertain the probability of occurrence of the next arson within a specified time in a year. Later we also

ascertain the probability of occurrence of a subsequent arson on a specified day.

Table 5.13 gives the frequency distribution of the number of subsequent arsons, from the day of the preceding arson, by selected time periods of the year 1980. To clarify the table, consider three arsons that occurred for a structure on, say, 5/11/80, 6/1/80, and 8/9/80. The arson on 6/1/80 occurred exactly on the 21st day from the occurrence of the arson on 5/11/80. Similarly, the arson on 8/9/80 occurred in the period 1-2 months from the occurrence of the previous arson of 6/1/80. This procedure generated a total of 147 arson incidences and their distribution is shown in the following table.

TABLE 5.13:- Distribution of subsequent arsons for a structure by selected time periods for the year 1980.

Time Period	No. of Arsons	Percent of total Arsons	Cumulative Percent	Time Period	No. Of Arsons	Percent of total Arsons	Cumulative Percent
same day	11	7.483	7.483	3-4 weeks	5	3.401	47.619
1 Day	10	6.803	14.286	1-2 months	27	18.367	67.986
2 Days	3	2.041	16.327	2-3 months	15	10.204	76.190
3 Days	7	4.762	21.089	3-4 months	14	9.524	85.714
4 Days	1	0.680	21.769	4-6 months	15	10.204	95.918
5 Days	4	2.721	24.490	6-9 months	5	3,401	99.319
6 Days	7	4.762	29.252	9-12 months	1	0.681	100.000
1-2 Weeks	16	10.884	40.136	----	--	---	---
2-3 Weeks	6	4.082	44.218	----	--	---	---
Total	--	---	---	----	147	100.00	---

From the above table we observe that the probability of the same structure being arsoned within 15 days of the occurrence of the last arson is about 40%. Similarly, the probability that the same structure will be arsoned within two months of the occurrence of the last arson is about 66%. Since this table is not based on a large number of arson incidences, some caution is called for in using this table.

In Table 5.14, distribution of arsons by day of occurrence of the subsequent arson is given together with the day of occurrence of the preceding arson. Thus, for example, if three arsons for a structure take place on Monday, Saturday, and Wednesday, then the first subsequent arson occurred on Saturday, after a first arson on Monday. The second one occurred on Wednesday following the occurrence of the arson on Saturday. This is how the following table has been generated.

TABLE 5.14:- Distribution of subsequent arsons by days of the week.

Day of the preceding arson	Day of the subsequent arson							Total
	Sun.	Mon.	Tues.	Wed.	Thurs.	Fri.	Sat.	
Sunday	5	6	3	3	1	4	5	27
Monday	3	4	9	2	2	1	3	24
Tuesday	2	5	4	0	3	2	3	19
Wednesday	2	6	2	6	3	0	0	19
Thursday	4	1	4	2	4	2	3	20
Friday	3	1	1	3	1	2	3	14
Saturday	2	5	4	1	2	4	6	24
Total	21	28	27	17	16	15	23	147

We note from the above table that of the 24 arsons that took place on Mondays, 9 arsons subsequently occurred on Tuesdays and only one on a Friday. It should be noted that these Tuesdays and Friday are not necessarily of the same or the following week but are the Tuesdays and Friday of the remainder of the year following the occurrence of the arson on a Monday.

The following table gives the probabilities of occurrence of the subsequent arson for each day of the week.

TABLE 5.15:- Probability of occurrence of the subsequent arsons by days of the week.

Day of the preceding arson	Day of the subsequent arson							Total
	Sun.	Mon.	Tues.	Wed.	Thurs.	Fri.	Sat.	
Sunday	0.185	0.222	0.111	0.111	0.037	0.148	0.185	1.0
Monday	0.125	0.167	0.375	0.083	0.083	0.042	0.125	1.0
Tuesday	0.105	0.263	0.211	0.0	0.158	0.105	0.158	1.0
Wednesday	0.105	0.316	0.105	0.316	0.158	0.0	0.0	1.0
Thursday	0.200	0.050	0.200	0.100	0.200	0.100	0.150	1.0
Friday	0.214	0.071	0.072	0.214	0.072	0.143	0.214	1.0
Saturday	0.083	0.208	0.167	0.042	0.083	0.167	0.250	1.0

From the above table we find that there is a 22.2% probability that a structure experiencing an arson on Sunday will have the next arson on a Monday. It may be recalled once again that this Monday is not necessarily immediately following the Sunday on which the arson took place but any Monday following that Sunday.

Table 5.13 and 5.15 together can be used to answer such questions as: What is the probability that the next arson for a structure will be on a Wednesday within two months of the occurrence of an arson on a Wednesday? From Table 5.13 we know that the probability of a next arson within two months is about 0.66. Multiplying this probability by the probability of being arsoned on Wednesday, when the preceding arson has occurred on Wednesday, which from Table 5.15, is 0.316, we find that the required probability is about 0.21, or equivalently, 21%.

One of the limitations of the above table, Table 5.15, is that it is specific for the year 1980 and must be used with caution for any other year. More useful in answering the above type of question will be the matrix of limiting (ultimate) probabilities.[1] These limiting probabilities are given in the following table.

TABLE 5·16: Limiting probabilities for the occurrence of the subsequent arson by days of the week.

Day	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
Probability	0.138	0.193	0.196	0.103	0.111	0.101	0.158

We notice from the above table that there are much greater probabilities of an arson being committed on Mondays and Tuesdays than on any other days of the week. It may be emphasized that these probabilities are for the occurrence of arsons in the long run, assuming that the present conditions remain the same in the future as well. Thus any change in these conditions such as commitment of a larger field staff than the present one by the Fire Department to fight arson, may alter the above probabilities.

As an illustration of how to use the above table, we return to the question of finding the probability of experiencing (by a structure) a subsequent arson on a Wednesday within two months of the occurrence of an arson on Wednesday for that structure. We observe from Table 5-16 that no matter on which day the arson in question is committed, the probability of having the subsequent arson remains the same. That is, whether the arson in question is committed on Sunday, Wednesday or Friday, say, the probability that the same structure will have its subsequent arson on Wednesday remain 0.103 (See Table 5-16). Since the probability of an arson within two months is 0.66 (See Table 5.13) the probability of having a subsequent arson on a Wednesday within two months is 0.103 x 0.66 = 0.068, or equivalently, about 6.8%

1 S.M. Ross (1972): Introduction to probability models. Academic Press New York.

5.6 SUMMARY:

There were about one-third repeated arsons during 1980 that occurred in 90 new structures. Moreover, these new structures experienced on an average about three arsons each during 1980. Similar to all arsons, the repeated arsons are also wide-spread. This further strengthens the conclusion of the last chapter that arson is widespread in Newark.

The monthly distribution of repeated arsons is similar to all arsons. We also observed a similarity of distributions for the days of the week for all arsons and repeated arsons. We have also seen that the distribution of repeated arsons by days of the week do not differ significantly from month to month. Furthermore, we observed that the distributions of repeated arsons and all arson are similar when we consider these distributions by three-hour time periods of a day. Thus, we conclude that the distribution of repeated arsons for the variables "month," "day," and "3-hour time period," are similar to those for all arsons. It, therefore, seems reasonable to conclude that any reduction in all arsons will bring about a corresponding reduction in repeated arsons. Analyses of the data for repeated arsons for fire districts and for census tracts with 10 or more arsons show similar kinds of patterns as were observed in the case of all arsons. This further strengthens the preceding conclusion that there is no need to treat repeated arsons separately from all arsons when planning to reduce all arsons.

Finally, we have presented the probability of a subsequent arson within a specific time ranging from none (same day) to one year. This table, Table 5.13, should prove helpful in predicting subsequent arson for a structure. Another table , Table 5.16,gives the limiting (ultimate) probabilities of subsequent arsons for a structure which experienced an arson in the past.

END