University of Alberta

Statistical Models for Assessing Sewer Infrastructure Inspection Requirements

by

Yuqing Yang ©

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of Master of Science

in

Construction Engineering & Management

Department of Civil and Environmental Engineering

Edmonton, Alberta

Spring 1999

0-612-40127-8

University of Alberta

Library Release Form

Name of Author: Yuqing Yang

Title of Thesis: Statistical Models for Assessing Sewer Infrastructure Inspection Requirements

Degree: Master of Science

Year this Degree Granted: 1999

Yuqing Yang

Yuqing Yang
520 L, Michener Park
Edmonton, Alberta
T6H 4M5

Dated: Jan. 29, 1999

University of Alberta

Faculty of Graduate Studies and Research

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled **Statistical Models for Assessing Sewer Infrastructure Inspection Requirements** submitted by **Yuqing Yang** in partial fulfillment of the requirements for the degree of **Master of Science in Construction Engineering & Management**.

Dr. S.T. Ariaratnam

Dr. S.M. AbouRizk

Dr. N.G. Prasad

Approved on: January 29, 1999

To My Father, My Mother

Jianfei, Noah and Dongya

ABSTRACT

The sewer system is often considered to be the "life-line" for a city and is the most cost intensive infrastructure system. Due to their low visibility, rehabilitation of the sewer system is generally conducted on a "reactionary" basis, which usually results in difficult and costly rehabilitation. "Proactive" action for sewer rehabilitation is gaining more attention because it allows the decision-maker to schedule the rehabilitation of the sewer system prior to the occurrence of urgent situations.

Two statistical models are presented in this research: 1) Logit model; and 2) Logistic model. Both of these models are utilized to demonstrate an approach for predicting pipe deficiency probability, thus providing a feasible approach for ranking candidate sewers for inspection. Rankings, produced from the Logit model, list the deficiency probabilities for all possible sewer types. The magnitude of deficiency probability determines the priority for initiating sewer inspection actions. The second model predicts deficiency probability based on changes of pipe attributes. The outcome can greatly improve the proactive strategy and objectivity in sewer rehabilitation.

ACKNOWLEDGMENT

I would like to thank my supervisor, Dr. Samuel T. Ariaratnam, for his encouragement, support, and technical guidance during the course of this research.

Special appreciation is extended to Dr. Simaan M. AbouRizk, a member of my examining committee, for his invaluable contribution towards my degree.

Appreciation is also extended to Dr. N.G. Prasad for serving as a member on my examining committee and contributing to technical aspects of the statistical models.

Thanks to Mr. Ken Chua, Mr. Gordan Lam, and others at the City of Edmonton – Drainage Services, Asset Management and Public Works for their assistance in data collection and technical guidance throughout this research.

The author wishes to thank research team member Dr. Ashraf El-Assaly for his assistance. Many thanks to all members of the University of Alberta Construction Engineering and Management Group for their support and assistance throughout my graduate program.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1 INTRODUCTION

1.1 INTRODUCTION

Wastes originating from domestic, commercial, and industrial sources (often mixed with stormwater) are collected, treated, and discharged back into the environment. Protecting wastewater collection and treatment systems with the least risk to public health and safety in the most cost-effective manner is the goal of any sewer rehabilitation program.

A sewer system is the "life-line" for a municipality and is generally the most cost intensive of all infrastructure systems. These systems deteriorate due to a number of factors including excessive usage, aging, change of the surrounding soil, and mismanagement. Furthermore, due to their low visibility, rehabilitation of sewer systems is frequently neglected until a major failure occurs, often resulting in difficult and costly rehabilitation requirements.

Sewers are defined as conduits that collect and transport wastewater or drainage water from an area to a discharge location. There are three main types of sewer systems that fall under the jurisdiction of most municipalities. These sewer collection systems consist of sanitary, storm, and combined conduits.

A *sanitary sewer* system carries waterborne wastes containing minor quantities of inadvertent storm, surface, and groundwater from residences, commercial buildings, industrial plants, and institutions.

A *storm sewer* system carries storm runoff, along with street waste and wash wastewater or drainage. It excludes domestic and industrial wastewater.

A *combined sewer* system is typically found in older parts of cities and represents collection systems that carry a mixture of domestic and industrial wastewater along with storm runoff.

Insufficient structural and hydraulic standards are the main signs of sewer pipe deficiency. Lateral deflection, crown sag, and offset joints, along with deteriorated mortar and exposed reinforcing caused by hydrogen sulfide corrosion are common occurrences (Wirahadikusumah, 1998). Minor defects can lead to major structural inadequacies in specific soil conditions when a sewer is subjected to surcharge because of insufficient hydraulic capacity (ASCE manual, 1994).

Defective sewer systems are regarded as "time bombs". Leaking underground sanitary sewer systems can cause exfiltration of raw wastewater and industrial discharge through leaking pipes, polluting soil and groundwater. In other cases, infiltration of groundwater through leaky joints and cracks in the pipe system can lead to excess cost at the treatment plant or contribute to pipe collapse (ASCE manual, 1994).

Due to their low visibility to the public, rehabilitation of underground sewer systems is often neglected until a catastrophic failure occurs. This results in costly and difficult rehabilitation options because of the need for emergency response. In other words, the discovery of damage in the sewer system initiates immediate action, which results in high monetary costs due to their urgent nature. There are two main avenues to improving sewer rehabilitation planning. The first is the collection and storage of adequate inspection information regarding the current condition of the sewer system; the second is the ability to predict sewer deficiency prior to failure to facilitate sewer inspection and repair prior to collapse.

In the last decade, technological advances such as closed-circuit television cameras (CCTV) have made it possible to observe the condition of existing sewers. Various assessment methods have been developed to assess the performance of sewers for various defective conditions. Advancements in computer technologies make it possible to store large amounts of physical and assessment data. Today, historical data on sewer system condition is available. How to make use of these historical data to predict the likely chances of deficiency of a sewer system so that proactive measures can be taken is the key issue for an effective sewer rehabilitation strategy.

1.2 RESEARCH SCOPE

The physical attributes of sewer systems owned by the City of Edmonton were extracted from their existing database. The current sewer assessment system was adopted to quantify the sewers into deficient and non-deficient status. Statistical measures are used to develop a ranking and prediction model based on historical sewer data. The research scope is limited to:

- Determining which attribute(s) significantly contribute to the pipe deficiency and to what extent.
- Obtaining a deficiency probability ranking for all combinations of sewer types so that the rehabilitation priority may be determined.
- Establishing a prediction model to predict the likelihood of deficiency for each combination of sewer types.

1.3 RESEARCH OBJECTIVES

The main objective of this research is to use statistical measures to develop a methodology for determining the rehabilitation priority for the City of Edmonton's sewer system based on a deficiency probability ranking of all sewer types. The use of statistical measures would reduce the "subjectivity" of these rankings and enable the decision-maker to allocate budgets and schedule the inspection of the entire sewer system, so that proactive rehabilitation measures are taken to avoid emergency response situations. Accordingly, rehabilitation costs will be greatly decreased.

The second objective of this research is to determine the physical attribute(s) of a sewer pipe that contributes significantly to deficiency and to what extent. Sewer pipe attributes may be extracted from the historical data. Statistical measures are used in the analysis.

The third objective is to develop a prediction model to predict the likelihood of a sewer pipe being in a deficient state. This should assist the decision-maker in short-term and long-term planning for budget allocation.

1.4 THESIS ORGANIZATION

Chapter 2 provides a state-of-the-art literature review and background of current sewer assessment methods and techniques for inspecting sewer systems. Chapter 3 "Data Analysis" focuses on data acquisition, initial analysis, and evaluation of sample size based on confidence level and error margin. The section on Data Category explains the methodology for classifying data into categorical levels.

Chapter 4 presents the basic concepts and procedures of statistical measures with the variables used in the statistical analysis of the sewer system described. The development of contingency tables and chi-square test in the association analysis is described. Additionally, the development of sample distributions is presented.

Chapter 5 shows the steps used to develop the ranking model using the loglinear model. Model selection and diagnoses are presented in detail along with the methods and procedures for establishing the loglinear model for sewer system ranking. Goodness-of-fit statistics and residual analysis are employed to diagnose the designed model followed by an outcome analysis once the designed model is accepted.

Chapter 6 presents the development of the logistic regression model. The logistic regression model is used to predict the likelihood of the sewer system being in a deficient state. The main different between the logistic regression and the linear regression model is that the outcome variable for the logistic model is binary.

Chapter 7 details the development of a prototype computer model for assisting decision-makers in determining the probability of a sewer combination being in a deficient state. The program, **T**rends **A**nalysis **S**ewer **E**valuation **S**ystems (TASES), is a conceptual window-based program intended to assist City of Edmonton personnel in determining the projected level of deficiency of a given sewer based on objective criteria from historical records.

Chapter 8 contains conclusions and recommendations for future areas of related research in the area of sewer infrastructure management.

CHAPTER 2 LITERATURE REVIEW

2.1 INTRODUCTION

Traditionally, infrastructure rehabilitation has been driven mainly by response to failures rather than through prevention. A majority sewer repair projects are executed on a "reactionary" basis, rather than adopting a "proactive" approach. Differences between "reactionary" and "proactive" approaches of sewer rehabilitation are illustrated in Figures 2-1 and 2-2, respectively.

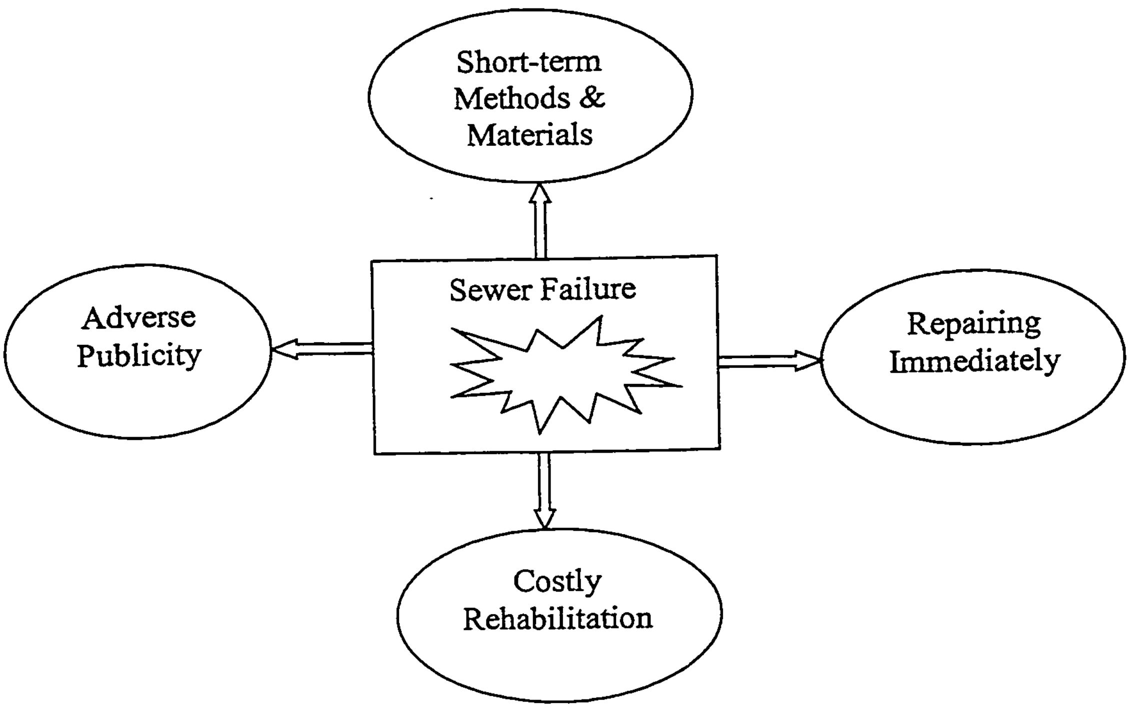

Figure 2-1 Reactionary Sewer Rehabilitation Approach

When a failure occurs, immediate inspection and repair are required due to its urgent nature. This emergency response, more often than not, results in adverse publicity due to traffic congestion, etc. Additionally, the methods and materials used to repair the deficient sewer primarily address immediate needs, rather than long term needs. This results in prohibitively high costs to the municipality.

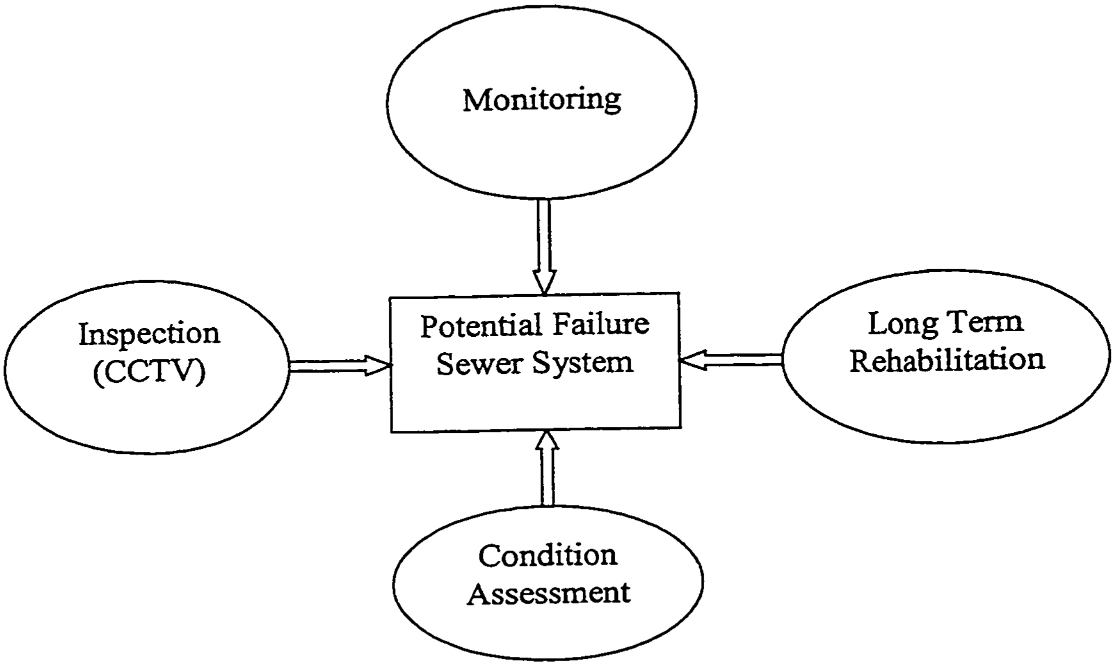

Figure 2-2 Proactive Sewer Rehabilitation Approach

Even if a sewer system is in service, it is possible that some deficiency already exists, which can deteriorate over time and result in total failure of the system. If the decision-maker has prior knowledge of the location of sewer to be inspected based on historical likelihood of failure, sewer inspection and assessment may be performed in an economical manner. The choice of rehabilitation approach and materials can then be based on long-term needs rather than a quick-fix solution. Additionally, measures can be taken to minimize the inconvenience to the public and other effects resulting from the rehabilitation resulting in overall cost savings.

The real question arising is how the decision-maker can determine the type and location of sewers requiring immediate inspection to prevent emergency responses to failure. The development of models using statistical measures to address this question is the main contribution of this reseach as outlined in Chapter 1.

2.2 CURRENT METHODS OF CONDITION ASSESSMENT AND SEWER EVALUATION

2.2.1 Sewer Condition Assessments

The choice of a proper, cost-effective rehabilitation procedure for a given sewer is best made through a thorough understanding of all possible methods along with knowledge of sewer condition. A complete evaluation is essential in determining whether it is more cost effective to rehabilitate or replace a particular sewer section or if merely monitoring is sufficient (ASCE manual, 1994).

In his paper "Assessment Technologies for Sewer System Rehabilitation," Wirahadikusumah presents an analysis of current sewer inspection technologies (Wirahadikusumah, 1998). He references current practices in sewer system assessment as mainly internal inspection performed through three methods: physical inspections; photographic inspection; and CCTV inspections. *Physical inspection* involves direct man-entry inspection of large sewers not in service. *Photographic inspection* utilizes a camera to take a series of color photographs along the inside of sewer lines. This technique is mainly used for analyzing the structural conditions of the sewer. *CCTV inspection* is primary an internal inspection technique. The camera is mounted in a casing and is subsequently pulled through the sewer with cables. Visual information is transmitted onto a TV monitor located in a control vehicle and stored via videotape. However, traditional CCTV inspection is frequently inadequate. The main reason mentioned in the paper (Wirahadikusumah, 1998) is that the quantity of information obtained by CCTV is dependent on the experience and skill of the technician and the reliability of the TV picture. The increasing need for more reliable data in condition assessment of sewer systems has prompted researchers to look for enhanced alternatives. Some promising non-destructive, remote-sensing diagnosis methods of infrastructure assessment have recently been developed and are presented Table 2-1.

2.2.2 Sewer Evaluation

Severe or catastrophic sewer collapses though rare are becoming more frequent. Disruption, adverse publicity, and public apprehension about potential reoccurrence normally follow sewer collapses. Failure is often associated with difficult ground conditions, large wastewater flows, adjacent utility impacts, traffic congestion, and deep excavation. Subsequent investigations often reveal incomplete records, infrequent inspections, or a failure to remedy known defects as factors contributing to the failure (ASCE manual, 1994).

Table 2-1 Evaluation of assessment technologies for sewer system rehabilitation (Wirahadikusumah, 1998)

Techniques	Advantages	Obstacles for implementation
1) CCTV	• Very commonly-used, well-known technology	• Dependent on skill and experience of technicians
	• New developments include production of higher quality images and portable inspection system.	• Dependent on the quality of TV picture • Difficult to estimate field production • Does not provide backfill condition information • Inaccuracy in detection of certain pipe defects
2) Infrared thermography system	• Large area inspection • Allows nighttime inspection • Detects pipe wall defects and provides backfill information • High field production	• Does not provide information on the depth of cracks (deep defects are difficult to detect) • Image interpretation depends on environment and surface conditions • Relies on single-sensor data collection
3) Sonic distance measurement method	• Describes cross sections of pipes • Measures pipe wall defection, corrosion loss, volume of debris • High field production	• Records only the part of the pipe that is not under water or below water, not simultaneously • Relies on single-sensor data collection
4) Ground penetrating radar	• Provide continuous cross-section profile of pipe walls • Identifies depth of cracks • High field production	• Data interpretation is very difficult, requires experience and training
5) Advanced system (KARO, PIRAT, SSET)	• Multi-sensory system (hence, provides more reliable data) • Provides continuos profile of pipe walls • Robot module • Higher benefit/cost ratio is anticipated	• In the prototypical or testing stage (requires further development for field implementation) • High initial cost

Structural condition assessment is a principal objective of any pipeline system inspection program. Closed-circuit television (CCTV) requires careful review and analysis to identify where structural rehabilitation or replacement is required. Field inspection provides information on the corrosion or deterioration. This also provides information about specific location conditions that affect the hydraulic performance of individual pipeline reaches, such as debris, roots, open joints and misaligned joints. The rating factors for internal condition evaluation suggested by the ASCE Manual of Existing Sewer Evaluation and Rehabilitation are presented in Table 2-2.

In addition to revealing opportunities for correcting capacity needs, sewer evaluation should identify structural and corrosion defects requiring correction, their severity and the potential consequence of failure. Sewer condition assessment helps establish priorities for rehabilitation or replacement. The likelihood of failure and the associated risk analysis are essential to the evaluation when budgetary constrains affect the work (ASCE Manual, 1994).

Table 2-2 Internal Condition Rating Factors

Description	Rating Factor
Collapse or collapse imminent	5
Collapse likely in foreseeable future	4
Collapse unlikely in near future, deterioration likely	3
Minimal collapse risk in short term but potential for future deterioration.	2

The goal of sewer rehabilitation is to arrest deterioration. Therefore, the choice of rehabilitation approach should be based on information available on the sewer system and on funds available. Based on the information initially developed, the following alternatives are proposed for evaluation (ASCE Manual, 1994):

- Level 1 – monitoring and information collection;
- Level 2 – stabilization of existing sewer;
- Level 3 – rehabilitation of existing sewer ; and
- Level 4 – replacement of existing sewer.

A routine inspection program should be established to monitor performance and develop needed design information that could help reduce rehabilitation costs. The main issue facing decision-makers is where and which kind of sewers should be inspected.

The main contribution of this research is the development of a ranking model that enables the decision-maker to determine more objectively which kind of sewers with high priorities should be monitored based on their high deficiency probability. The more objective decision stems from the statistical analysis for the historical sewer system rehabilitation data. This deficiency ranking varies from one city to another because of changing environment factors. Although the methodology to obtain the deficiency rankings is the same for different cities, the historical sewer system rehabilitation data for each city varies.

One problem in data analysis using statistical measures is how to define the pipe status, that is how to determine the severity of pipe deterioration in order to classify the pipe status into deficiency and non-deficiency states. Almost every city across North America has its own sewer evaluation system and rating systems vary across each municipality. Table 2-3 presents the structural condition rating system utilized by the City of Edmonton, Alberta. This system was adopted by the City of Edmonton in 1994. The assessment of pipe deficiency from the CCTV inspections are quantified into five condition levels from 1 to 5 increasing in severity of deficiency.

Table 2-3 Structural Condition Rating for the City of Edmonton, Alberta

Defect	Code	Rating
Rating of 1		
- Cracking Light	CL	1
- Corrosion Light	HM	1
- Sag Light	SL	1
- Open Joint Light	OL	1
Rating of 2		
- Sag Moderate	SM	2
- Joint Displacement Light	JL	2
- Open Joint Moderate	OM	2
Rating of 3		
- Deformed Pipe Light	DL	3
- Fracture Light	FL	3
- Crack Moderate	CM	3
- Corrosion Moderate	HM	3
Rating of 4		
- Deformed Pipe Moderate	DM	4
- Fracture Moderate	FM	4
- Joint Displacement Moderate	JM	4
- Crack Severe	CS	4
- Fracture Severe	FS	4
- Corrosion Severe	HS	4
- Open Joint Severe	OS	4
- Sag Severe	SS	4
Rating of 5		
- Collapse Pipe	DX	5
- Broken Pipe	FX	5
- Deformed Pipe Severe	DS	5
- Joint Displacement Severe	JS	5

Table 2-4 Pipe Status Evaluations

Structural Condition Rating	*Pipe Status Evaluation*
1	Non-deficiency
2	Non-deficiency
3	Deficiency
4	Deficiency
5	Deficiency

Pipe status mainly depends on the structural conditions of the sewer pipe. Pipes with structural rating 1-2 are defined as "non-deficient" due to the acceptable performance. Pipes with structural ratings between 3-5 are classified as "deficient" due to the severity of defects (see Table 2-4).

2.3 STATISTICAL BACKGROUND

The factors that are responsible for the deterioration of a sewer system such as waste type (i.e. sanitary, storm, combined) are referred to as categorical variables, also known as discrete variables. These variables are in a discrete state and in contrast to continuous variables such as time. It is more difficult to perform statistical analysis using discrete variables than continuous variables. Statistical methodology for categorical data has only recently reached the level of sophistication achieved early in this century by methodologies for analyzing continuous data (Agresti, 1990). Most methodologies that prevail nowadays were developed in the 1960's. The recent development of methods for categorical data was stimulated by the increasing methodological sophistication of the social and biomedical sciences.

Categorical data analysis utilized in infrastructure maintenance and rehabilitation is quite new for two main reasons. One reason is that the methodology for categorical data analysis has become more sophisticated, especially in its application. The other reason is that availability of fast computers makes it easier to store tremendous amounts of infrastructure information.

The purpose of utilizing statistical measures for infrastructure maintenance is to build a model to predict the failure probability based on historical data. The ranking of a sewer system is based on the failure probability with higher priority given to the higher failure probability.

Chouinard et al., 1996 used the condition indexing system, developed by Andersen and Torrey in 1996 to design a statistical model to rank the performance of concrete dams using two parameters – age and height. In this research, existing sewer performance assessment in City of Edmonton, Alberta is adopted to code the Pipe Status (deficient vs. non-deficient). Five main parameters (Pipe Age, Pipe Diameter, Pipe Material, Waste Type and the Average Depth of Cover) are considered to be responsible for the deterioration of the sewer system. Logit statistical model is utilized to obtain sewer pipe deficiency probability; Logistic statistical model is employed to predict the pipe deficiency probability with pipe age.

CHAPTER 3 DATA ANALYSIS

3.1 INTRODUCTION

3.1.1 Project Background

During the fall of 1997, the City of Edmonton commissioned the Construction Engineering and Management Group at the University of Alberta to assist them with their Local Sewer Rehabilitation Strategy initiative. The work scope presented in this research outlines the development of a sewer rehabilitation plan/strategy for the proactive implementation and scheduling of the Local Sewer Rehabilitation Program. Local sewers are defined as those with diameters from 150 mm to 1200 mm. It should be noted that services and trunk sewers are defined as having diameters less than 150 mm and greater than 1200mm, respectively (Ariaratnam, et al. 1998). The overview of the data analysis of this project is illustrated in the Figure 3-1.

Drainage Operations at the City of Edmonton is responsible for conducting inspection of the all drainage infrastructure and has inspected approximately 800 km of the local sewer network over the past 10 years as part of their yearly inspection program. Currently, an average of 220 km/year of local sewer is inspected using closed circuit television (CCTV).

The expected service life of a sewer network is variable and depends on several parameters that affect pipe conditions including: hydraulic loading; wastewater quality; soil characteristics; pipe material; location; construction; installation, etc. (Ariaratnam, et al. 1998).

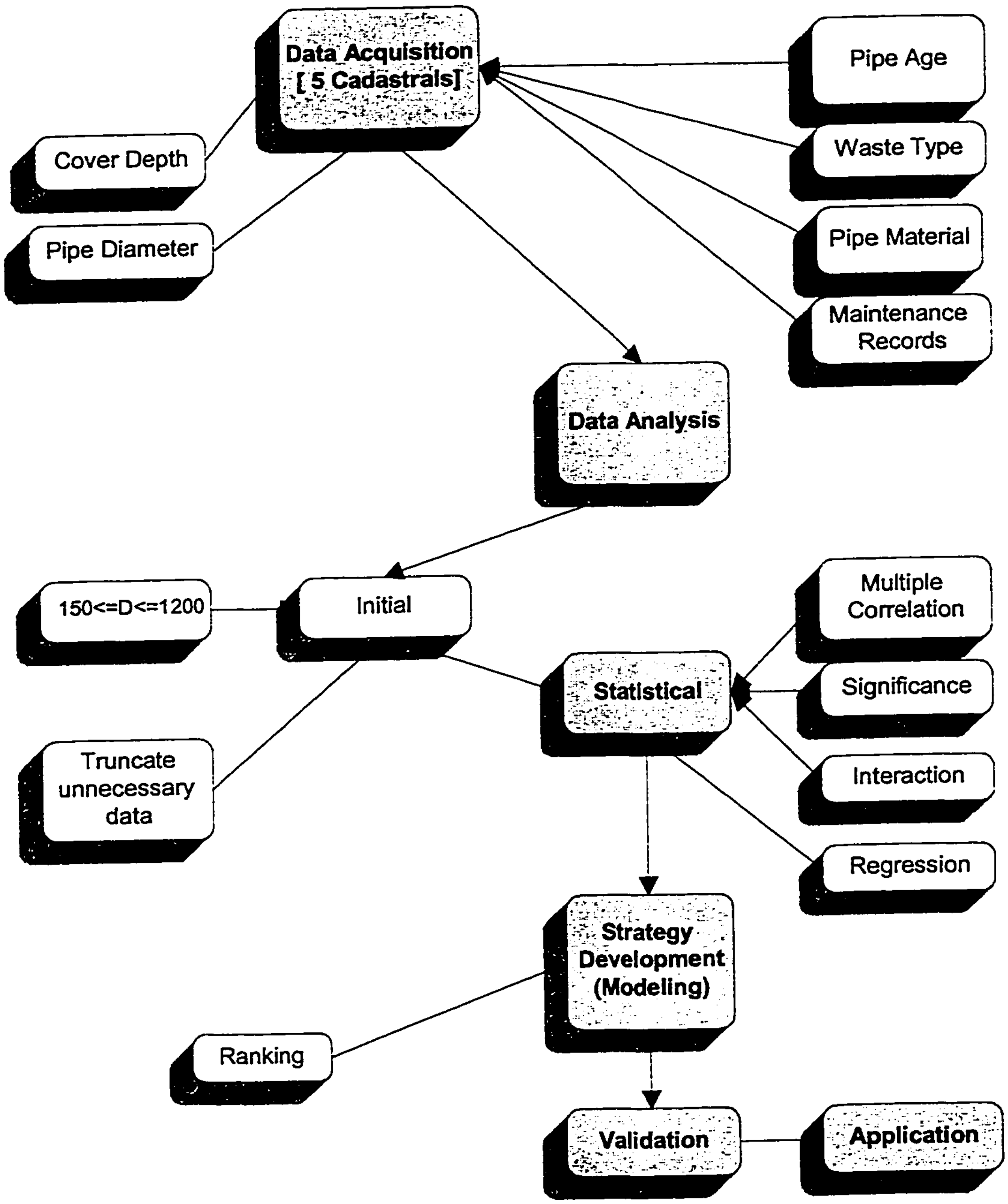

Figure 3-1 Project Data Analysis Overview

3.1.2 Data Sources

Sewer information was obtained from various databases, sewer rehabilitation records, CCTV reports, and analysis of sewer types based on condition assessment criteria. The following sources of historical sewer data were used in the research:

- Sewer maintenance records
- Sewer rehabilitation records
- Sewer deficiency records
- Sewer physical attributes
- CCTV report information
- Inner City Prioritization Study – Drainage System Assessment by CH2M Hill Engineering Ltd., 1993
- Sewer Physical Condition Classification Manual, 1996

The main source of historical sewer data came from DRAINS, a database containing historical sewer inspection records. Figure 3-2 illustrates a screen from the Oracle browser for DRAINS, which facilitated ease of obtaining information contained in the databases by providing the option of extracting information deemed relevant. Information contained in the DRAINS database include:

- Location of each sewer section;
- Starting and ending manhole number;
- Pipe diameter;
- Pipe material;
- Waste type;

- Cover depth;
- Year of construction;
- Deficiency records; and
- Maintenance records.

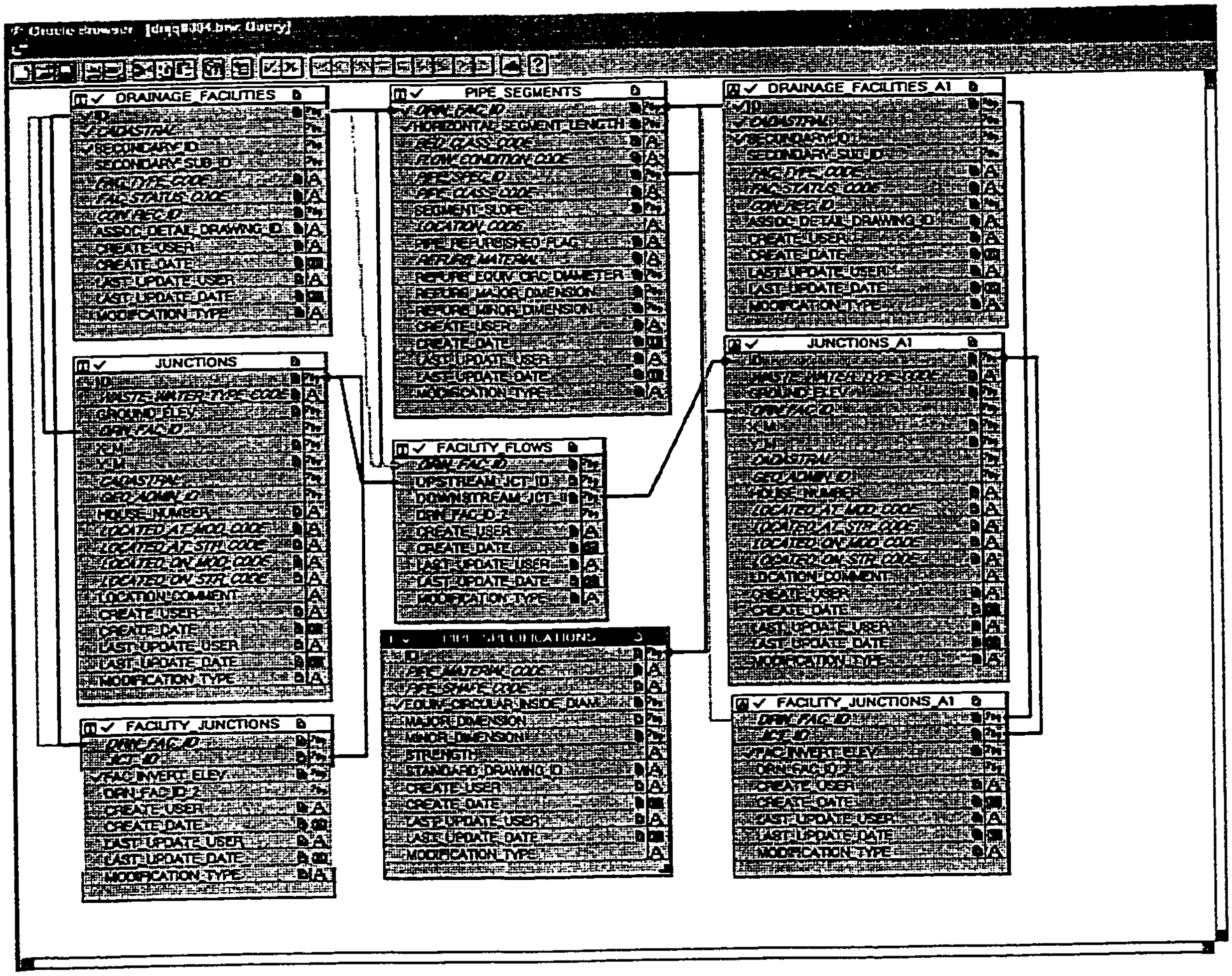

Figure 3-2 Oracle Browser Screen for DRAINS

3.2 DATA ACQUISITION

The data acquisition phase included gathering information regarding the sewer network governed by the City of Edmonton. Data samples from five cadastral maps were collected to serve as a representative sample of the sewer network. The City of Edmonton is divided into a number of areas (or cadastrals) each containing 25 smaller area maps. These cadastral maps were selected based on discussions with personnel at the Drainage Branch and cover both residential and industrial locations. The following cadastrals were chosen:

1) Riverdale, Boyle St./McCauley (934+36);
2) Richie (928+36);
3) Millwoods (922+40);
4) Millwoods (925+40); and
5) Millwoods (922+36).

Local sewer maintenance information is only available electronically in DRAINS from 1987. This created limitations in the analysis because historical information from the date of construction was unavailable; however, this issue is one often encountered by infrastructure systems. The acquired data was extracted from the City's database and presented to the research team in spreadsheets format.

One issue affecting the sample data taken randomly derives from the sewer system feature. A Pipe ID identifies each sewer section, which is determined by the length between subsequent manhole locations. Because of the cost, the inspection of a sewer system is measured as a section (rather than individual pipe), which may consist of a number of sewer pipes. Subsequently, there may be a number of pipes within one section with similar physical attributes, such as waste type, or similar pipe age, etc. As a result, although the sewer sections are selected randomly, this does not mean that the sewer pipes are also selected randomly. This issue reduces the reliability of any statistical analysis to some extend.

3.3 DATA INITIAL ANALYSIS

Data analysis was performed using the SPSS (Version 6.5) statistical analysis package. This medium was selected because of its flexibility and ease of importing the data from DRAINS. The data analysis process was divided into two phases: 1) Initial analysis; and 2) Frequency analysis. This section focuses primarily on the Initial analysis with the Frequency analysis detailed in the Data Categories section.

In this stage, the types of data required to achieve the objectives of the research were analyzed and insignificant information in the collected data files deleted. These data included those deemed to have no relation to the attributes of the pipe. For example, data for sewers with diameters less than 150 mm and greater than or equal to 1200 mm were eliminated as they are not contained in the local sewer system.

Based on experience and availability of sewer data, the following five attributes are selected as contributing to sewer system deterioration:

- *Pipe Material* including TP: Clay Tile Pipe; PVC: Polyvinylchloride Pipe (for detailed pipe material codes see Appendix B).
- *Pipe Age* is calculated by taking the difference between the current year, 1998, and the pipe built year. For instance, the pipe age = 1998 – 1958 = 40 years for any pipe built in 1958. Several records with built year = 9999, meaning that the pipe built year is not available, were eliminated from the data analysis.

- *Waste Type* are categorical into three types: SAN: Sanitary: STM: Storm and CMB: Combined Sewer System.
- *Pipe Diameters* were truncated for ranges less than 150 mm and greater than or equal to 1200 mm.
- *Average Cover Depth* refers to the depth from the ground to the upper profile of sewer pipe as illustrated in Figure 3-3 and calculated using Equation 3-1.

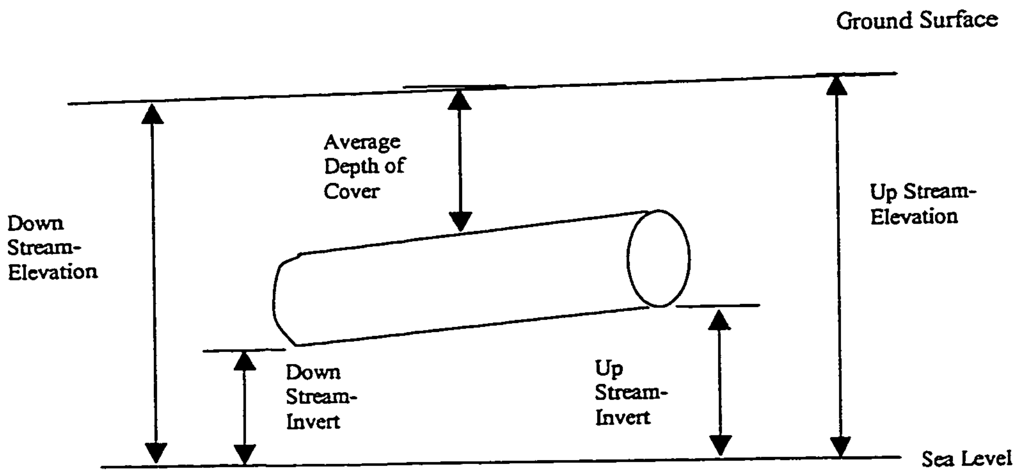

Figure 3-3 Pipe Average Cover Depth Calculation

$$D = \frac{(UPS\_ELE - UPS\_INERT) + (DWS\_ELE - DWS\_IINVERT)}{2} - \frac{Diameter}{1000} \qquad \text{Eq 3-1}$$

Where D is the average cover depth of pipe in meters, UPS_ELE, DWS_ELE is the up- and down- stream elevations (above sea level), respectively. UPS_INVERT, DWS_INVERT represent the up- and down- stream invert elevations (above the sea level), respectively. Diameter is the pipe diameter measured in millimeters.

3.4 SAMPLE SIZE EVALUATION

The manhole (from) and manhole (to) identify sewer pipe. The length of pipe varies for different type of pipe. The longer of the pipe is, the higher deficient probability may occur. When the sewer system was designed, the length of pipe may be limited with in a reasonable range. However in this research the effect of pipe length is neglected. The sample size is the pipe route (from manhole to manhole) number inspected from the corresponding neighborhood.

Table 3-1 presents the results of a sample size calculation performed to determine the confidence level and margin of error of the selected data records. The population size for each neighborhood is the total valid records for this neighborhood. The valid records in sample size are actual inspected valid records. The sample records were randomly chosen for inspection in each neighborhood. Additionally, it is imperative that the sample size for each neighborhood be evaluated in order to build a satisfactory model.

Table 3-1 Sample Size Evaluation

	Neighborhood	Total Valid Records In this Cadastral Area (Population Size)	Actual Inspected CCTV Valid Records (Sample Size)	Theoretical Sample Size Required	
				10% Error	5%Error
934+36	RiverDale, Boyle & McCauley	1455	448	94	314
928+36	Ritchie	1605	229	94	320
922+40	Millwoods	1955	25	95	332
925+40	Millwoods	2250	70	96	340
922+36	Millwoods	915	12	90	278
Total		*8180*	*784*	*469*	*1584*

Note: Valid records mean the records excluding pipes with
Pipe Size =< 150 or >=1200mm
Built year equals to 9999
Depth of pipe <=0 or >=300m

The last two columns in Table 3-1 are the theoretical sample size computed by the equation 3-5 and 3-6 (Gower and Kelly 1993).

$n_1 = z^2pq / (d^2) = 100$ (± 10% margin of error) Eq 3-2

$n_1 = z^2pq / (d^2) = 400$ (± 5% margin of error) Eq 3-3

where;

n_1 = initial sample size

$z = 1.96$ (rounded to 2) for a 95% confidence interval

P = an estimation of the proportion of the population that has the characteristics of interest. Since the variability of population isn't known in advance, we assume the proportion of population with deficient pipe is 50%, take $p = 0.5$.

$q = 1 - p = q = 0.5$

d = the specified margin of error, in this case ±10% and ±5% respectively.

Here, the margin of error is used to indicate the confidence interval for estimation. We assumed 50% of the population with the deficient pipes might be reported as having a margin of error of ± 5% with a 95% confidential level. This means that true percentage would be expected to lie within the range of 45% (= 50 % -5%) to 55% (= 50% + 5%) with a 95% confidence level.

The following equation is used to calculate the modified sample size n_2;

$n_2 = n_1 * N / (N + n_1)$ Eq 3-4

Using the values of n_1 calculated from Eq 3-2 and Eq 3-3 gives:

$n_2 = 100 * N / (N + 100)$ (for ±10% margin of error) Eq 3-5

$n_2 = 400 * N / (N + 400)$ (for ±5% margin of error) Eq 3-6

Where N is the population size for a given neighborhood. For example, the neighborhood "Ritchie" has a population size of 1605, or N = 1605, Therefore;

$n_2 = 100 * 1605 / (100 + 1605) = 94$ (for ± 10 % margin of error)

$n_2 = 400 * 1605 / (400 + 1605) = 320$ (for ± 5 % margin of error)

Totally, the actual sample size is 784 lying within a range from 469 (with 10% margin of error) and 1584 (with 5% margin of error). Therefore, we can accept the statistics obtained from this actual sample size with 95% confidence level and 5%-10% margin of error.

3.5 DATA CATEGORIES

Each attribute value is broken down into several classes in order to calculate the deficiency frequency for each class. These deficiency frequencies are utilized later to determine the categories and the interval for each attribute. The breakdown detail for each attribute is illustrated in Figure 3-4. For each class, the *Frequency* field is the total count; the *Deficiency* is the deficiency frequency and the *Percentage* (Eq 3-7) is the result of deficiency frequency divided by total count.

$$Percentage = \frac{DeficientFrequency}{TotalFrequency} \qquad \text{Eq 3-7}$$

$$NormalizedWeight = \frac{Percentage}{\sum Percentage} \qquad \text{Eq 3-8}$$

Ranking and *Normalized Weight* (Eq 3-8) are consistent with the Percentage. Ranking is based on the magnitude of percentage. The *Normalized Weight* visually provides the proportion of the deficiency percentage, so is used to categorize the pipe attributes.

The categories for each parameter are based on Normalized Weight shown in Table 3-2. For pipe material and waste type, they are naturally categorized into different categories. For instance, there are three kinds of waste type, therefore waste type are classified into three categories. For other attributes (pipe depth, pipe age and pipe diameter), the procedure is to compare the *Normalized Weight* for each classes, then put all classes with as similar *Normalized Weight* as possible into one category, while at the same time considering the consecutive nature of numbering schemes. For example, for pipe age the classes are as follows:

Pipe Age	Normalized Weight	Category
0-9	0.050	1
10-19	0.071	1
20-29	0.082	1

The Normalized Weights are close for the above pipe age, therefore, they may be categorized into 1, which means pipe ages 0 to 29 fall into one category in statistical analysis. Similar rules are applied for categorizing other attributes.

However, in some cases, the consecutiveness of the attribute values must be considered when categorizing data groups. For example, pipe age is grouped from 30 to 59 as one category although the Normalized Weights are somewhat different among the intervals 30-39, 40-49, 50-59. The pipe deficiency percentage and the categories for all attributes in this research are presented in the Table 3-2.

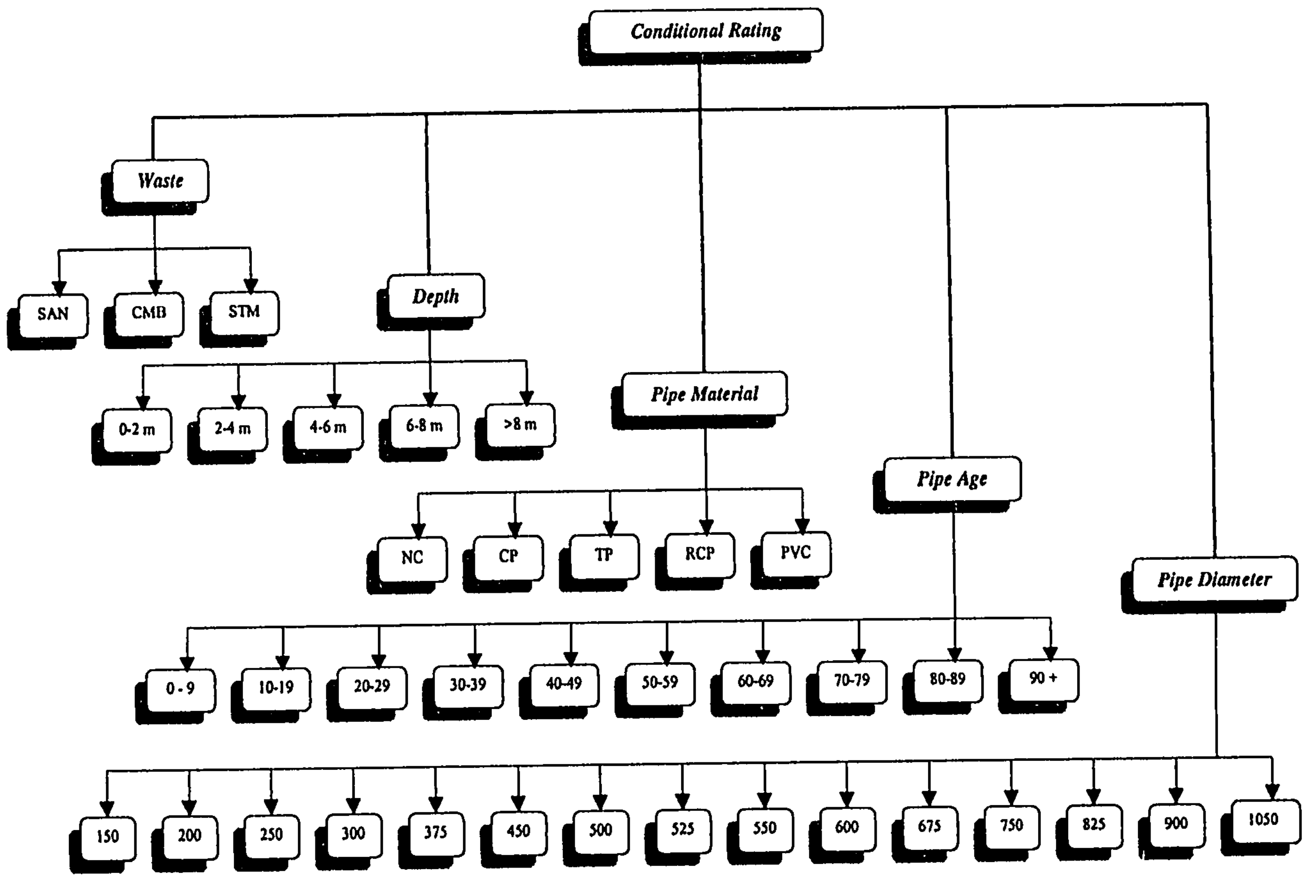

Figure 3-4 Breakdown of Pipe Attributes

Table 3-2 Pipe Categories Based on Deficiency Percentage

WasteType	*Frequency*	*Deficiency*	*Percentage*	*Ranking*	*Normalized Weight*	*CATEGORY*
SAN	100	55	55.00%	1	0.435	1
STM	147	34	23.13%	3	0.183	2
CMB	537	260	48.42%	2	0.383	3
	784	*349*	126.55%		1.000	

PipeMaterials	*Frequency*	*Deficiency*	*Percentage*	*Ranking*	*Normalized Weight*	*CATEGORY*
CP	158	51	32.28%	4	0.171	1
TP	474	261	55.06%	2	0.292	2
RCP	104	11	10.58%	5	0.056	3
NC	42	24	57.14%	1	0.303	4
PVC	6	2	33.33%	3	0.177	5
	784	*349*	188.39%		1.000	

Diameter(mm)	*Frequency*	*Deficiency*	*Percentage*	*Ranking*	*Normalized Weight*	*CATEGORY*
150	2	1	50.00%	2	0.127	1
200	156	94	60.26%	1	0.153	1
250	195	114	58.46%	1	0.149	1
300	149	70	46.98%	3	0.119	1
375	86	39	45.35%	3	0.115	1
450	59	13	22.03%	5	0.056	2
500	5	1	20.00%	5	0.051	2
525	13	4	30.77%	4	0.078	2
550	3	0	0.00%	7	0.000	3
600	26	4	15.38%	6	0.039	3
675	30	3	10.00%	6	0.025	3
750	24	3	12.50%	6	0.032	3
825	10	1	10.00%	6	0.025	3
900	17	2	11.76%	6	0.030	3
1050	9	0	0.00%	7	0.000	3
	784	*349*	393.50%		1.000	

PipeAge	*Frequency*	*Deficiency*	*Percentage*	*Ranking*	*Normalized Weight*	*CATEGORY*
0-9	15	3	20.00%	4	0.050	1
10-19	84	24	28.57%	4	0.071	1
20-29	144	47	32.64%	3	0.082	1
30-39	14	6	42.86%	2	0.107	2
40-49	139	48	34.53%	3	0.086	2
50-59	33	8	24.24%	4	0.061	2
60-69	7	4	57.14%	1	0.143	3
70-79	6	2	33.33%	3	0.083	3
80-89	339	205	60.47%	1	0.151	3
90+	3	2	66.67%	1	0.166	3
	784	*349*	400.46%		1.000	

Depth(m)	*Frequency*	*Deficiency*	*Percentage*	*Ranking*	*Normalized Weight*	*CATEGORY*
0-2	33	20	60.61%	1	0.273	1
2-4	477	217	45.49%	3	0.205	1
4-6	199	80	40.20%	3	0.181	1
6-8	52	26	50.00%	2	0.225	2
8+	23	6	26.09%	4	0.117	2
	784	*349*	222.39%		1.000	

3.6 REAL POSSIBLE DATA COMBINATIONS

From the above data analysis, the theoretical possible data combinations for pipe attributes are 3 (Waste Type) × 5 (Material) × 3 (Diameter) × 3 (Age) × 2 (Depth) = 270. Although this number of combinations is theoretically possible, in reality, some combinations do not exist at all. For example, PVC pipes with ages between 60-90+ for sanitary waste types are unexisting since PVC material was not even use in the period of 1908 – 1938. All the pipe combinations that do not exist were eliminated. There are two ways to eliminate any impossible pipe combinations. One is to delete the records from the raw data so that all real possible data are used in the statistical analysis. The other way is to define a Cell Structure Variable with Structural Zero in the statistical analysis, which was used for this research (SPSS, 1997).

If a count for one combination of pipe attributes is zero, there are two possible interpretation for it. If this combination does not exist at all, the zero is necessary zero, called *structure zero*. Then the value of Cell Structure Variable is specified into zero; If this combination does exist, however the count in this sampling data is zero due to chance variation, called *sample zero*. Then the value of Cell Structure Variable is specified into 1. Table 3-3 illustrates the explanation. Also see Appendix A which shows all sample data with the column Cell Structure Variable to indicate zero count feature.

Table 3-3 Cell Structure Variable Definition

Count of Deficiency Pipe for One Pipe Combination Attributes	*Interpretation of Zero*	*Cell Structure Variable*
0	Necessary Zero	0
0	By Chance Zero	1

CHAPTER 4 STATISTICAL MEASURES

4.1 INTRODUCTION

Statistics refers to the analysis and interpretation of data with a view toward objective evaluation of the reliability of the conclusions based on the data (Zar, 1996). Statistical methods are normally applied to measurements in a sample selected from a population of interest. Statistics describe samples, while parameters describe populations. The two main types of statistical analysis are descriptive methods for summarizing the sample and population and inferential methods for making predictions about population parameters using sample statistics (Agresti, 1997). For sewer systems, inferential methods are utilized to predict the population parameters such as the pipe deficiency probability by using the sample statistics.

Once the initialing analysis for the historical data is completed, the next step is to investigate statistical methods to predict the deficiency probability for different sewer systems, so that more objective evidence can be offered for the decision-maker to determine which kind of sewer system should be inspected.

Statistical methods are so powerful that there are a variety of applications in civil engineering. In order to limit the statistical application scope in infrastructure rehabilitation, this chapter only presents the basic statistical concepts and principles applied to the statistical methods detailed in the Chapters 5 and 6.

The variables used in infrastructure analysis are defined in Section 2 of this chapter. This secticn is very important in that it specifies the correct variable types for sewer pipe attributes so that an appropriate model may be selected. The use of contingency tables and the chi-square test is the basic measure for variable independence analysis and goodness-of-fit analysis, presented in Section 3. The last section details the sample distribution to determine which distribution correctly reflects the sample data for the sewer system.

4.2 VARIABLE DEFINITIONS

Variables are classified as *continuous* or *discrete*, according to the number of values they can attain. The continuous vs. discrete difference is, in practice, a distinction between variables that can assume many values versus variables that take on relatively few values. Pipe age, pipe diameter and average depth of cover are considered continuous variables since they each contain numerous values. Pipe material and waste type only contain distinct variables and are thus defined as being discrete.

Nominal variables are *qualitative* – each distinct level differs in quality, not in quantity. For example, pipe status is defined as either being in a deficient or non-deficient state. Interval variables are *quantitative* – each distinct level has numerous quantities of interest. For example, pipe diameter takes on values of 230mm, 250mm, 300mm, etc.

Most statistical analysis distinguishes between *response* (or "dependent") variables and *explanatory* (or "independent") variables. For example, in determining the relationship of income with gender, income represents the response variable and gender the explanatory variable. However, income may depend on gender, not gender on income.

Categorical variables are those for which the measurement scale consists of a set of categories. For sewer systems, waste type is measured as "sanitary", "storm" and "combined".

For categories which levels do not have a natural ordering are called *nominal categories*. These include pipe status with categories of "pipe deficiency" and "pipe non-deficiency". Conversely, many categorical variables do have ordered levels. Such variables are called *ordinal*. An example of an ordinal variable is size of automobile (i.e. subcompact, compact, mid-size, and large). An *interval* variable is one that does have numerical distances between any two levels of the scale. An example might be blood pressure level. For sewer systems, pipe age and pipe diameter are considered to be interval variables. The minimum distance between two pipe ages is taken as 1 year and the minimum interval between pipe diameters as 1 mm.

4.3 CONTINGENCY TABLE AND CHI-SQUARE

The objective of this research is to determine the factors that contribute to pipe deficiency and their magnitudes. As discussed in Chapter 3, five factors contributing to pipe deficiency were chosen for analysis. The association of each factor to pipe deficiency was performed to evaluate the extent of contribution to deterioration. There is said to be an *association* between two variables if the distribution of the response variable changes as the value of the explanatory variable changes.

Besides determining whether two variables are associated, one should determine whether the association is strong enough to have practical importance. For categorical variables, the common way to exam the association between two variables is to analyze the contingency table. When the data are cross-classified (the purpose of which is usually to uncover the interrelationships between the variables) the various classifications along with their constituent frequency counts are referred as a *contingency table.*

Data for the analysis of categorical variables are displayed in contingency tables. This type of table displays the number of subjects observed at all the combinations of possible outcomes for the variables. Table 4-1 contains the results of the observed frequencies for each combination of the relationship between pipe deficiency and waste type.

Table 4-1 Contingency Table Pipe Status vs. Waste Type (Observed Frequencies)

Waste Type	*Deficiency*	*Non-dificiency*	*Total*
SAN	55	45	***100***
STM	34	113	***147***
CMB	260	277	***537***
Total	***349***	***435***	***784***

4.3.1 Independence and Dependence

Whether an association exists between waste type and pipe status is a matter of whether the three waste types differ in their conditional distribution on pipe deficiency. Two categorical variables are statistically *independence* if the population conditional distributions on one of them are identical at each category of the other. Conversely, the variables are statistically *dependent* if the conditional distributions are not identical. In other words, two variables are statistically independent if the percentage of the population in any particular category of one variable is the same across all categories of the other variable. The percentage proportions for each category are illustrated in Table

4-2. The following example illustrates how to calculate the percentage using the case of STM and Deficiency.

$$Percentage = \frac{34}{34+113} = 23\%$$

Using this approach yields conditional distributions for SAN, STM and CMB as (55.0, 45.0), (23.0, 77.0), (48.0, 52.0), respectively.

Table 4-2 Waste Type Percentage for Pipe Status (Observed Percentages)

Waste Type	*Deficiency*	*Non-dificiency*	*Total %*	*N*
SAN	55%	45%	100%	**100**
STM	23%	77%	100%	**147**
CMB	48%	52%	100%	**537**

Since the Observed deficiency Percentages for SAN (55%) and STM(23%) and CMB (48%) are not identical, the sample distribution is not identical. However, since observed percentages are sample conditional distributions, even if they are independent, we would not expect the sample conditional distribution to be identical. Because of the sampling variability, each sample percentage typically differs from the true population percentage.

A statistical test is employed to test the sample distribution instead of the population distribution. The null hypothesis for the test determines if the two categorical variables are statistically independent. The hypotheses are:

H_0: The variables are statistically independent

Ha: The variables are statistically dependent.

Employing a chi-square test suggests whether to accept the null hypothesis or the alternate hypothesis.

4.3.2 Chi-Square Test Statistic

The chi-square test compares the observed frequencies in the cells of the contingency table with values expected from the null hypothesis if variables are independent. The expected frequency, f_e, for a cell equals the product of the row and column totals for that cell divided by the total sample size as shown in Equation 4-1.

$$f_e = \frac{TotalRow \times TotalColumn}{SampleSize} \qquad \text{Eq 4-1}$$

For example; for a cell with CMB waste type and deficiency pipe status, the total row is 537, total column is 349, and the sample size is 784. Therefore, the expected frequency may be calculated as:

$$f_e = \frac{537 \times 349}{784} = 239$$

The expected frequencies calculated for other cells are shown Table 4-3. Chi-square statistics test how closely the expected frequencies compare to the observed frequencies. The general form of the chi-square statistics is presented in Equation 4-2.

Table 4-3 Contingency Table Pipe Status vs. Waste Type (With Expected Frequencies in Parenthesis)

Waste Type	*Deficiency*	*Non-dificiency*	*Total*
SAN	55 (45)	45 (55)	***100***
STM	34 (65)	113 (82)	***147***
CMB	260 (239)	277 (298)	***537***
Total	***349***	***435***	***784***

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e} \qquad \text{Eq 4-2}$$

The summation is taken over all cells in the contingency table. For each cell, the difference between the observed and expected frequencies is squared and then divided by the expected frequency. If H_o is true, f_o and f_e tend to be close for each cell, and χ^2 is relatively small. If H_o is false, at least some f_o and f_e values tend not to be close leading to large $(f_o - f_e)^2$ values and a large test statistic. The larger the χ^2 value, the greater the evidence against the null hypothesis of independence.

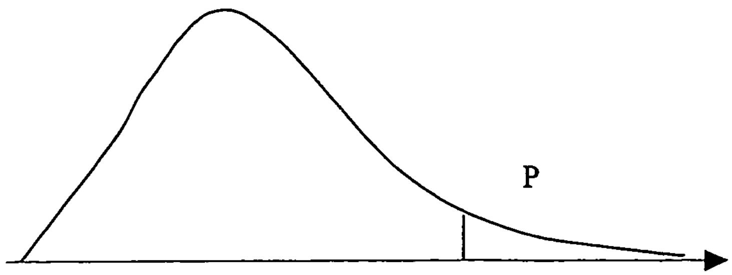

Figure 4-1 The P-Value for the Chi-square test of independence is the right-hand tail probability above the observed value of the test statistic.

The main properties of χ^2 test follow the following rules (Agresti, 1997):

- It is concentrated on the positive portion of the real line. The χ^2 test cannot be negative (i.e. the minimum possible value = 0). When $f_o = f_e$ in each cell in Eq 4-2; the variables are completely independent in the sample.
- The precise shape of the distribution depends on the degrees of freedom. For testing independence in a table with r-rows and c-columns, the degree of freedom is:

$df = (r-1)(c-1)$

(i.e. for a 3×2 table: $df = (3-1) \times (2-1) = 2$)

- The larger the χ^2 value, the stronger the evidence against the null hypothesis of independence. Values of χ^2 greater than the observed value are ones providing even greater evidence against the null hypothesis than the observed data.
- The P-value equals the right tail probability under the χ^2 curve above the observed χ^2 value (see Figure 4-1).

The chi-square test is utilized to verify the relationship between two variables. For a given significance level, if the observed χ^2 Value > χ^2 Value in Chi-square Table, the null hypothesis is rejected. Therefore the two variables are dependent. To investigate the relationship between the pipe status and waste type, the observed chi-square is calculated from Table 4-3: the result is 34.9812. Based on 95% significance level, with the df=2, check the statistic χ^2 table in any of statistical textbook to get the χ^2 Value = 5.99. Compare Chi-square 34.9812 > 5.99, we reject the null hypothesis. Therefore the pipe status and waste type is dependent.

4.4 THE SAMPLE DISTRIBUTION

The *sample distribution* of data is the distribution we actually observe, graphically displayed as a histogram of the data, or numerically described by statistics. The larger the sample size, n, the closer the sample distribution resembles the population parameters. The *population distribution* is the distribution from which the sample is selected. The distribution is usually unknown and inferences may be made about its characteristics.

The *sampling distribution* of a statistic is the probability distribution of a sample statistic. A sampling distribution describes the variability that occurs in the value of the statistic of a certain size. This distribution determines the probability of the statistic falling within a certain distance of the population parameter (Agresti, 1997).

The data in a multi-way cross-tabulation can originate in many different ways. Subsequently, there are three ways to gather sample data, namely, the total sample size is fixed, not fixed, and both sample size and margin sizes are fixed. For instance, you may wish to determine whether there is a relationship between construction productivity and weather. There are three ways to gather sample data. One may choose to select the historical records of 1000 days in productivity and weather. In this case, the sample size 1000 days is fixed. Alternatively, one may decide to examine records of two projects in productivity and weather. In this case, the sample size is not fixed since the actual project duration (days) is changeable. Another way one might obtain data is to select 500 days productivity during sunny days and 500 productivity records during raining days. In this case, the sample size is fixed not only for total number of selected records but also for the margin size (each group). Regardless of how you conducted the study, the results can be displayed in a cross-tabulation of productivity and weather. How you conduct the study determines, in part, which sampling model is appropriate for the data.

The two distributions most often used to describe the distribution of counts in a cross-tabulation are the *Poisson* distribution and *multinomial* distribution. The Poisson distribution is useful for modeling rare events such as suicides. A Poisson sampling model for a cross-tabulation table arises when the total sample size is not fixed and the number of cases in each cell of the table is independent of others and has a Poisson distribution.

The multinomial distribution is a generalization of the binomial distribution to more than two events. Under a multinomial sampling model, each cell of the cross-tabulation table has a probability that indicates how likely an observation is to fall into it. The sum of the probability across all cells is 1. The total sample size in a multinomial model is fixed; thus, the cell counts are not independent because they must sum to the total. In this case, prior knowledge of the number of observations included in the sample is available. If the row or column totals (margins) are fix, such as when selecting 500 sunny days and 500 raining days, the distribution for the entire table is called the *Product Multinomial* distribution.

Multinomial sampling is more commonly encountered in the analysis of categorical data. Fortunately, most of the results on modeling under these three distributions turn out to be the similar. No matter of what sampling distribution you choose from those three, your analysis results make no significant difference.

For sewer systems, the multinomial sampling model is chosen due to the sample size is fixed and the sum of the probability for all cells is 1. In addition, the SPSS software supports multinomial sampling.

More generally, consider observing a random variable, Y, which can have one of k possible outcomes: b_1, b_2 ..., b_k with probabilities π_1, π_2..., π_k. The outcome cells are assumed mutually exclusive and $\sum_{i=1}^{k} \pi_i = 1$. For example, Y could represent the answer to the question "Do you wear a seatbelt?" with k=3 possible responses; b_1= "Always", b_2

= "Sometimes", and b_3 = " Never". If the n people are interviewed on this case, the random vector is used to summarize the responses. The random vector is $X = (X_1, X_2 \ldots, X_k)$ where X_i = Number of times (out of n) that $Y = b_i$, $i = 1, \ldots k$. When the observed Y's are independent and identical distributed, then X has a multinomial distribution with parameters n, $\pi = (\pi_1, \pi_2, \ldots, \pi_k)$. The probability of any particular outcome $\mathbf{x} = (x_1, x_2, \ldots, x_k)$, is then

$$\Pr(X = x) = n! \prod_{i=1}^{k} \frac{\pi_i^{x_i}}{x_i!} \qquad \text{Eq 4-3}$$

where $0 \le x \le n$, $0 \le \pi^i \le 1$, i= 1,..., k, and $\sum_{i=1}^{k} x_{i=n}$, $\sum_{i=1}^{k} \pi_{i=1}$.

When k = 2, the multinomial distribution reduces to the binomial distribution described later in Chapter 6.

CHAPTER 5 LOGLINEAR MODEL

5.1 INTRODUCTION

When examining the relationship between two qualitative variables, the traditional approach is to construct a contingency table and compute the appropriate Chi-square statistic to test the hypothesis of independence. However, what happens when you have more than two categorical variables? One approach would be to construct a series of two-way tables and compute corresponding Chi-square statistics for each one. However, this strategy is fatally flawed in the sense that the results will be confounded by interaction between variables. The ideal solution would allow one to examine the relationships among all of the variables simultaneously, including interactions among specified groups of variables. One such approach is the use of loglinear models (SPSS 1997). *The loglinear model* is appropriate for contingency tables in which each classification is a responsible variable meaning there is no distinction between all variables. *The Logit model* takes one variable as a response variable and the others as explanatory variables.

This chapter first compares the loglinear model vs. the Logit model to decide which suits the analysis presented in this research. The second section provides a comparison of the role of Pearson statistics and Likelihood-ratio statistics in model selection. Subsequently, the model selection process is presented in detail providing solutions to questions including how to determine the significance of each main effect and which interactions should be included in the model to increase its accuracy. Finally, the model diagnoses techniques to validate the model from different perspectives are presented. Goodness-of-fit statistics are used to provide broad summaries of the how the model fits the data, whereas the residual diagnosis provides a microscope insight of the model fitting. Outcome analysis is utilized to denote the results obtained from the model.

5.2 LOGLINEAR MODEL AND LOGIT MODEL

5.2.1 Loglinear Model

Loglinear models are used to study the association patterns among categorical variables and predict the frequency for each possible combination. All variables are identical to be explanatory variables, none of them is taken to be responsible variable. The expected frequency of each combination of the variables is the real response variable. Loglinear analysis resembles a correlation analysis more than a regression analysis, however, it focuses on studying associations between pairs of variables rather than modeling the response of one in terms of the others.

Berenson et. al 1983 define loglinear models as a class of mathematical models designed to assist in uncovering associations that exist among categorical variables. In statistical analysis, a loglinear model is a model that is an expression of how observed data (i.e. observed frequency in each cell) are affected by variables and their combinations. Loglinear refers to a procedure whereby a complicated relationship may be transformed into a linear relationship by the use of logarithms (Zar, 1996).

A general loglinear analysis analyzes the frequency counts of observations falling into each cross-classification category. Each cross-classification constitutes a cell, and each categorical variable is termed a factor. Thus, the dependent variable is the number of cases (frequency) in a cell of the cross-tabulation, and the explanatory variables are factors. The mathematical model for the expected count in a cell is given by:

$$m_i = e^{x_i\beta} \quad \text{Eq 5-1}$$

where; i = 1..., r

Here, the m_i term is the expected cell count for the *i*th cell, x_i is the *i*th row of the design matrix, β is the vector of parameter, and r is the number of cells (SPSS, 1997). Taking logarithm *of* to both sides yields equation 5-2:

$$Ln(m_i) = Ln(z_i) + x_i\beta \quad \text{Eq 5-2}$$

where; i= 1...r

In other words, loglinear analysis allows the model to consider several variables at once and multiple categories in each variable. Additionally, the loglinear model formula expresses the logs of cell expected frequencies in terms of dummy variables for the categorical variables.

5.2.2 Logit Model

Loglinear models are used to study the association patterns among categorical variables and predict the frequency for each possible combination. No distinction is made between the variables that compose the cross-classification. In many cases, interest centers upon the relationship of a set of design or explanatory variables on at least one response variable. As the name suggests, the behaviour of response variables are thought to be explained by the explanatory variables. *Logit Models,* a special class of loglinear models, are used to model the relationship between one or more dependent categorical variables and a set of independent categorical variables (as well as covariates). A special case of the logit model is the multinomial model. This model is appropriate when it is natural to regard one or more categorical variable(s) as the response variable(s) and others as the explanatory variables. At each setting or combination of the explanatory variable, the subtotal sample size is fixed and the cell counts of the response variables follow a multinomial distribution (Bererson et. al, 1983). In a sewer system, the response variable is Pipe Status and the explanatory variables are Pipe Material, Pipe Diameter, Pipe Age, Average Depth of Cover, and Waste Type. For the purpose of this research sample set, deficiency are comprised of all records rated 1 or 2 using the initial condition rating system (out of 3) and all records rated 4 or 5 using the current structural condition rating system (out of 5). All other ratings are considered Non-Deficient. Recall we category the parameter into different group so that we can minimize the error occurring in data collection and all parameters are ready for the logit analysis. Each variable and the their nominal values for all categories are displayed in Table 5-1. For instance, the explanatory variable Pipe Age is comprised of three categories denoted by nominal values of 1 for Ages 1-29 years; 2 for 30-59 and 3 for 60-90+.

Table 5-1 Variable Categories

Response Variable:				
Pipe Status	1:Deficiency			
	0:Non-deficiency			
Explanatory Variables:				
Waste Type	Pipe Materials	Pipe Diameter (mm)	Pipe Age (years)	Depth(m)
1: SAN	1: CP	1: 150-375	1: 0-29	1: 0-6
2: STM	2: TP	2: 450-525	2: 30-59	2: 6-8+
3: CMB	3: RCP	3: 550-1050	3: 60-90+	
	4: NC			
	5: PVC			

As previously mentioned, a special case of the logit model is the multinomial model. Sometimes they are referred to as *Multinomial Logit Models,* since for each combination of values of the independent variable there is a multinomial distribution of the dependent variable and the cell counts across combinations are independent (SPSS, 1997).

If you consider one of the variables as the dependent variable, instead of modeling the counts of cases for each cell, one can model the ratio of the counts of the dependent variable for each of the combinations of values of the independent variables. The multinomial logit model (only including the main effect for the time being) is as follows:

$$Ln\left(\frac{M_{ijkln1}}{M_{ijkln0}}\right) = \lambda + \alpha_i + \beta_j + \gamma_k + \mu_l + \upsilon_n \qquad \text{Eq 5-3}$$

where;

M_{ijkln1} is the deficiency frequency in cell (i, j, k, l, n)

M_{ijkln0} is the non-deficiency frequency in cell (i, j, k, l, n)

λ is the baseline term, constant

α_i is the term due to the Waste Type, i= 1, 2, 3

β_j is the term due to the Pipe Material, j=1, 2, 3, 4, 5

γ_k is the term due to the Pipe Diameter, k = 1, 2, 3

μ_l is the term due to the Pipe Age, l = 1, 2, 3

ν_n is the term due to the Depth of Cover, n = 1, 2

Odds are the ratio of the probability that an event will occur to the probability that the event will not occur. That is:

Odds = Probability of Success / Probability of Failure

In this case, odds = F_{ijkln1} / F_{ijkln0} for cell (i, j, k, l, n).

5.3 MODEL SELECTION

5.3.1 The Likelihood-Ratio Chi-Square

Prior to proceeding with model selection, a special case of the Chi-square test, Likelihood-Ratio Chi-square is performed. Recalling, the Pearson statistics, shown in equation 5-4, summarizes the differences between the expected frequency f_e and the observed frequency f_o. Large differences yields large values of the statistics and more evidence that the model is inadequate (Agresti, 1997).

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e} \qquad \text{Eq 5-4}$$

The Likelihood-Ratio chi-square is computed by the following equation (Kennedy, 1992):

$$G^2 = 2\sum f_o Ln(f_o / f_e) \qquad \text{Eq 5-5}$$

where f_o is the observed frequency for cell i and f_e is the expected frequency for cell i. It equals the difference between the values for the model being tested (*Custom model*) and the most complex model (*Saturated model*).

A *Saturated* model is a perfect fit using the maximum possible number of parameters leaving zero degrees of freedom and zero residual between the observed data and the predicted data. A *Custom* model (unsaturated model) is a model from which some terms are eliminated based on the statistical non-significant. The likelihood-ratio statistic and its *P-value* of a custom model describe whether the selected model is statistically different from a *Saturated* model (SPSS, 1997).

When there is a perfect agreement between the custom model and the saturated model, (i.e. all cells with $f_o = f_e$,), therefore, $G^2=0$. As the discrepancy increases, the value of G^2 will increase. Since the large values of X^2 and G^2 indicate a poor fit, the P-value for testing a model is the right-hand tail probability above the observed value.

In loglinear analysis, we use the Likelihood-Ratio Chi-square test instead of the traditional Pearson Chi-square because it is additive for nested models, whereas the Pearson statistic, in general, is not.

Likelihood-Ratio Chi-square statistic has the desirable property that it is additive, meaning the sum of the chi-square values for the individual effects in the model equals the chi-square for the total model. Therefore, if one considers the difference between two Likelihood Chi-square statistics for related models, the result would be another Likelihood-Ratio Chi-square statistic. This property enables one to make two important inferences: 1) nested models can be compared; and 2) individual effects may be assessed.

5.3.2 Main Effect Tests

A model is said to be nested within another model if the effects in the nested model are a subset of the effects in the more complex model (Kennedy, 1992).

For example:

$$Ln(F) = \theta + \lambda_x + \lambda_y \qquad \text{Eq 5-6}$$

is nested within the model:

$$Ln(F) = \theta + \lambda_x + \lambda_y + \lambda_z \qquad \text{Eq 5-7}$$

Notice that all of the effects in equation 5-6 are also included in equation 5-7. By computing Likelihood-Ratio Chi-Square statistics for each of these models and then taking the difference, another Likelihood-Ratio Chi-Square statistic that tests the relative advantage of the more complex model (Eq 5-7) over the simple model (Eq 5-6) in predicting the odds ratio is obtained. In such cases, the simpler model that still explains the observed data is chosen.

This research tests the effects of each parameter individually (i.e. pipe material, pipe diameter, pipe age, depth of cover and waste type) to the logit model which is further explained in the following section. For demonstration purpose, take the effect of Pipe Age and to Pipe Deficiency as an example. The basic model includes all explanatory variables (i.e. only the main effects, not including any interactions). The nested model is the model excluding the effect of Pipe Age from the basic model. The basic logit model is presented in equation 5-8:

$$Ln\ (F) = Constant + Material + Diameter + Age + Depth + Waste \qquad \text{Eq 5-8}$$

Where F is the ratio of Deficiency Frequency to Non-deficiency Frequency for a given Combination. The nested logit model is easy to design by eliminating the attribute to be evaluated from the basic model. For example, to evaluate the effect of Pipe Age to the Pipe Deficiency, simply remove the Pipe Age term from the basic logit model. The nested logit model is presented in equation 5-9:

$$Ln\ (F) = Constant + Material + Diameter + Depth + Waste \qquad \text{Eq 5-9}$$

The SPSS analysis results of running the logit model for the basic and nested model separately to determine G^2 and df are presented in Tables 5-2 and 5-3, respectively.

Table 5-2 SPSS Printout for Basic Model (including all five explanatory variables)

Goodness-of-fit Statistics

	Chi-Square	DF	Sig.
Likelihood Ratio	83.8327	258	1.0000
Pearson	79.1221	258	1.0000

Table 5-3 SPSS Printout for Nested Model Excluding Pipe Age Effect

Goodness-of-fit Statistics

	Chi-Square	DF	Sig.
Likelihood Ratio	115.8257	260	1.0000
Pearson	120.8032	260	1.0000

Comparing the Likelihood-ratio Chi-square yields the following results:

Difference G^2 = Basic Model G^2 – Nested Model G^2

= 115.8257 – 83.8327 = 31.993

Difference *df* = Basic Model *df* – Nested Model *df*

= 260-285 = 2

Based on a 95% significance level,

P-Value = 1 – CDF.CHISQ (Diff G^2, Diff *df*) = 0.00

This formula is used to calculate the P-Value based on the difference of Likelihood ratio and difference of degree of freedom. P-Value = 0.00 means the effect of Pipe Age to the basic model is significant.

The other nested models and the SPSS printouts are presented in Table 5-4 to 5-7.

Table 5-4 SPSS Printout for Nested Model Excluding Material Effect
Ln (F) = Constant + Diameter + Age + Depth + Waste

Goodness-of-fit Statistics

	Chi-Square	DF	Sig.
Likelihood Ratio	87.9608	262	1.0000
Pearson	82.3800	262	1.0000

Table 5-5 SPSS Printout for Nested Model Excluding Diameter Effect
Ln (F) = Constant + Material + Age + Depth + Waste

Goodness-of-fit Statistics

	Chi-Square	DF	Sig.
Likelihood Ratio	113.1259	260	1.0000
Pearson	98.4224	260	1.0000

Table 5-6 SPSS Printout for Nested Model Excluding Depth Effect
Ln (F) = Constant + Material + Diameter + Age + Waste

Goodness-of-fit Statistics

	Chi-Square	DF	Sig.
Likelihood Ratio	84.7783	259	1.0000
Pearson	82.7879	259	1.0000

Table 5-7 SPSS Printout for Nested Model Excluding Waste Effect

Ln (F) = Constant + Material + Diameter + Age + Depth

Goodness-of-fit Statistics

	Chi-Square	DF	Sig.
Likelihood Ratio	91.9839	260	1.0000
Pearson	86.0376	260	1.0000

The overall test results are summarized in Table 5-8. The results indicate that Pipe Diameter and Pipe Age have the greatest contribution among the five parameters to Pipe Deficiency, and that Pipe Material and Average Depth of Cover have the least influence. The influence of Waste Type was found to be significant. It should be noticed that all five parameters were included in the model because the effects of Pipe Material and Depth of Cover do exist in predicting pipe failure probability even though they have less influence than the other three parameters.

Table 5-8 Results of Main Effect Testing

To test this effect	Compare this model	To this model (Basic Model)
Material	Ln F=Constant+Diameter+Age+Depth+Waste	Ln F=Constant+Material+Diameter+Age+Depth+Waste
Diameter	Ln F=Constant+Material+Age+Depth+Waste	Ln F=Constant+Material+Diameter+Age+Depth+Waste
Age	Ln F=Constant+Material+Diameter+Depth+Waste	Ln F=Constant+Material+Diameter+Age+Depth+Waste
Depth	Ln F=Constant+Material+Diameter+Age+Waste	Ln F=Constant+Material+Diameter+Age+Depth+Waste
Waste	Ln F=Constant+Material+Diameter+Age+Depth	Ln F=Constant+Material+Diameter+Age+Depth+Waste

	Likelihood Ratio Chi-Square	DF	LRC (Differ)	DF(Differ)	P-Value	Conclusion
(1)	(2)	(3)	(4)	(5)	(6)	(7)
Basic Model	83.8327	258				
Excluding Material Model	87.9608	262	4.1281	4	0.39	Non-Significant
Excluding Diameter Model	113.1295	260	29.2968	2	0	Significant
Excluding Age Model	115.8257	260	31.993	2	0	Significant
Excluding Depth Model	84.7783	259	0.9456	1	0.33	Non-Significant
Excluding Waste Model	91.9839	260	8.1512	2	0.02	Significant

Column (4): LRC(Differ) = LRC of Nested model--LRC of basic model
Column (5): DF(Differ) = DF of Nested model--DF of basic model
Column (6): P-Value=1-CDF.CHISQ(LR,DF), based on 95% significance of level
P<=0.05 The effect is significant, otherwise P>0.05 nonsignificant.

5.3.3 *K*-Way Test to Determine Interactions

One question that often arises is "how to reflect the effects of interactions among the parameters into the logit model?" Theoretically, the number of interactions are C_5^5 + C_5^4 + C_5^3 + C_5^2 = 26. How does one determine which interaction term should be included in the model? *K-way effects tests* and *partial association tests* may be used to determine solutions for these questions.

For example, assume three parameters, *X, Y, Z.* The 3-way interaction is *X*Y*Z* term and the 2-way interactions are *X*Y, Y*Z, Z*X.* The *K*-Way Effects Test tests the hypothesis that the *K*-Way effects are zero and helps in assessing the level of complexity required in the model (Kennedy, 1992). If the P-value ≤ 0.05, it means that the effect is significantly different than 0, we reject the hypothesis. Therefore, these interactions should be included in the model. Table 5-9 displays the results of the *K*-Way Effects Test from the SPSS analysis.

Table 5-9 *K*-Way Effects Test Results

Tests that K-way effects are zero				
K	**DF**	**L.R.Chisq**	**P-Value**	**Conclusion**
5	152	0.636	1.000	Non-significant
4	332	28.411	1.000	Non-significant
3	470	230.623	1.000	Non-significant
2	527	1965.784	0.000	Significant
1	539	4429.469	0.000	Significant

Table 5-9 indicates that the 5-way, 4-way and 3-way effects are non significant, therefore 2-way effects and 1-way (i.e. main effects) are included in the model. However, even for a 2-way interaction, there are C_5^2 = 10 terms. As a result, we have to make use of a Partial Association Test to decide which 2-way interaction should be present in the model.

The Partial Association Test results from SPSS are tabulated in Table 5-10. Partial associations allow us to test the significance of each individual effect in the model (SPSS, 1997). In general, terms with partial associations that do not differ from 0 (i.e. probability ≥0.05, which have a non-significant partial chi-square) can be omitted from the model without sacrificing predictive accuracy.

Table 5-10 Tests of Partial Association

Effect Name	DF	Partial Chisq	Prob	Conclusion
AGE*DEPTH	2	9.532	.0085	Significant
AGE*DIAMETER	4	2.869	.5800	Non-significant
DEPTH*DIAMETER	2	11.023	.0040	Significant
AGE*MATERIAL	8	107.776	.0000	Significant
DEPTH*MATERIAL	4	7.393	.1165	Non-significant
DIAMETER*MATERIAL	8	179.614	.0000	Significant
AGE*WASTE	4	330.301	.0000	Significant
DEPTH*WASTE	2	41.869	.0000	Significant
DIAMETER*WASTE	4	35.872	.0000	Significant
MATERIAL*WASTE	8	170.324	.0000	Significant
AGE	2	55.639	.0000	Significant
DEPTH	1	592.231	.0000	Significant
DIAMETER	2	578.239	.0000	Significant
MATERIAL	4	815.908	.0000	Significant
WASTE	2	412.212	.0000	Significant

Table 5-10 suggests that *Age*Diameter* and *Depth*Material* are Non-significant and can be excluded from the model. At this point, we know which effects and which interactions should be included in this model. Finally, the multinomial logit model presented in equation 5-3 including the main effects and the interactions may be written as follows:

$$Ln(F) = \lambda + \alpha_i + \beta_j + \gamma_k + \mu_l + \upsilon_n + \mu_l \times \upsilon_n + \mu_l \times \beta_j + \mu_l \times \alpha_i + \upsilon_n \times \alpha_i + \beta_j \times \alpha_i + \upsilon_n \times \gamma_k + \beta_j \times \gamma_k + \alpha_i \times \gamma_k \quad \text{Eq 5-10}$$

where F is the ratio of deficiency frequency over the non-deficiency frequency for a given cell. All other letter's meanings are as same as those in Eq 5-3. The product terms denote the interaction between those two parameters. For practical purposes, this ratio is transferred into a *failure probability* for each cell as shown in equation 5-11 below:

Deficiency Probability = Deficiency Frequency / Total Frequency Eq. 5-11

The deficiency probabilities for all combinations obtained from equation 5-10 are illustrated in Table 5-13 in descending sequence. Tables in Appendix 3 show the sorted deficiency probabilities by pipe age, average depth of cover pipe diameter, material and waste type.

5.4 LOGIT MODEL DIAGNOSIS

5.4.1 Goodness-of-Fit Statistics

The first step in model diagnosis is the examination of goodness-of-fit. Goodness-of-fit tests use the properties of a hypothesized distribution to assess whether data are generated from that distribution (Read, 1988). The fit of the model is usually assessed by comparing the frequencies expected in each cell, given by $n\pi$, against the observed frequencies. If there were substantial discrepancy between the observed frequencies and the expected frequencies from the null model, then it would be wise to reject the null model.

Two goodness-of-fit statistics are reported in the Logit procedure: 1) the Pearson Chi-square statistic; and 2) the Likelihood-Ratio Chi-square statistic. As previously mentioned, the Likelihood-Ratio Chi-square statistic has a definite advantage because it is additive for nested models, whereas the Pearson Chi-square statistic in general is not. Therefore, the Likelihood-Ratio Chi-Square is the preferred test. The likelihood-ratio statistic compares the extent that the selected model fits the data to the fit of a corresponding Saturated Model.

The Goodness-of-Fit test results for the custom model are presented in Table 5-11. The Likelihood-Ratio Chi-square with *Sig.* = 1.000 indicates that the custom model fits the data with almost perfect fitting to the observed data through comparison of the custom model with the corresponding saturated model.

Table 5-11 Goodness-of-Fit Test

Goodness-of-fit Statistics

	Chi-Square	DF	Sig.
Likelihood Ratio	19.2195	220	1.0000
Pearson	24.7592	220	1.0000

If the *P-value* in the Goodness-of-Fit test ≤ 0.05 (labeled *Sig.* in the output table), this indicates that the custom model cannot adequately describe the data as well as the saturated model and should therefore include more parameters.

5.4.2 Residuals

Another step in model diagnosis is the examination of residuals. Goodness-of-Fit statistics provide only broad summaries of how models fit data. Chi-square test statistics and P-value analysis summarize the strength of evidence against the null hypothesis. If χ^2 is too large for testing independence, then somewhere in the data set there is a departure from that predicted. The test statistic does not indicate, however, whether all cells deviate greatly from independence or only one or two of the data sets.

After selecting a model, further insight is obtained by conducting a microscopic mode of analysis. A cell-by-cell comparison of observed and expected frequencies reveals the nature of the variation. The difference $(f_o - f_e)$ between an observed and expected frequency is called a residual. This step assists in evaluating the fit for each observation, identifying possible outliers, and often provides hints to improving the model.

How do we know whether a residual is large enough to indicate a significant departure from independence? The following section discussed two kinds of residuals: 1) *Standard Residual; and* 2) *Adjusted Residual.* Each residual has its own criteria to judge whether to accept the residual.

1. The Standardized residuals

The standard residuals take into account the size of the fitted data. Standardized residuals of a multinomial logit model may be calculated by the following equation:

$$Std.R = \frac{f_o - f_e}{\sqrt{f_e(1 - \frac{f_e}{N})}} \qquad \text{Eq 5-12}$$

where f_o is the observed frequency, f_e is the expected frequency, and N is the sample size.

If the model holds true, the standardized residuals are asymptotically normal with the mean equal to 0 and the variance less than 1. Figure 5-1 illustrates a Standardized Residual histogram generated from the designed model. The variance equals 0.28 * 0.28 = 0.0784 < 1 and the mean = 0.00, which indicates that the selected model fits the observed data at an acceptable level.

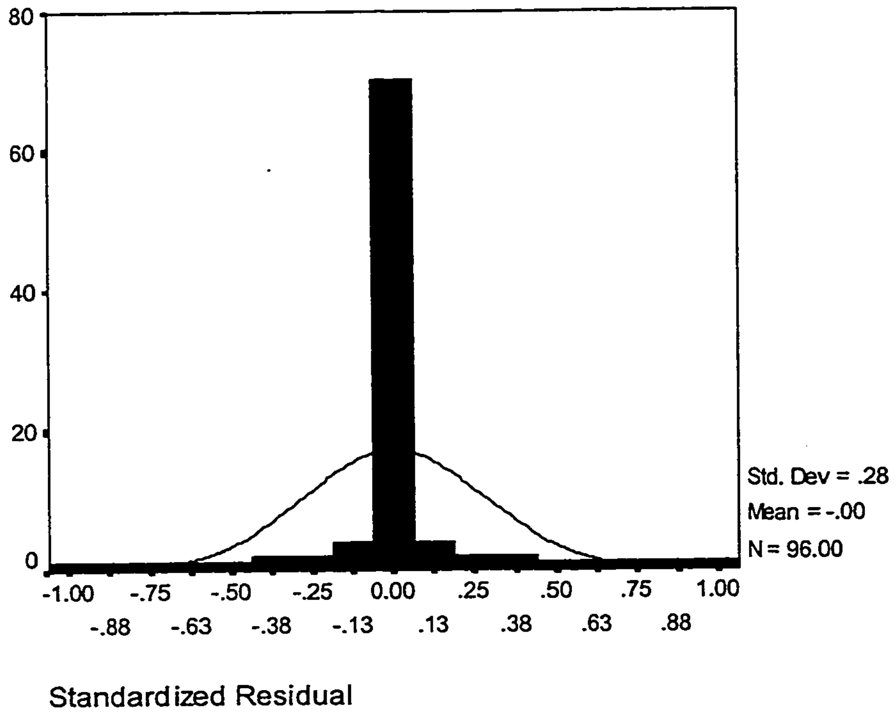

Figure 5-1. Standardized Residual Histogram

Generally speaking, standardized residuals with an absolute value in excess of 2.0 would be cause for concern (Berenson et. al, 1983). This, however, is not the case here because the maximum absolute value is 1.0 for the designed model.

2. *The Adjusted Residuals (Q-Q Plot)*

The Adjusted Residual is the standardized residual divided by its estimated standard error. If the designed model holds true, the Adjusted Residual behaves like a normal distribution with mean = 0 and standard deviation = 1. For the model to hold true, the adjusted residuals must behave like standard normal variables. A large adjusted residual (i.e. exceeding 3 in absolute value) provides evidence of lack of fit in that cell (Agresti, 1997).

The plot Q-Q is the quantiles plot of a variable's distribution against the quantiles of test distribution. Q-Q plots are generally used to determine whether the distribution of a variable matches a given distribution. If the selected variable matches the test distribution, the points cluster around a straight line. Figure 5-2 is the output of a Q-Q plot from SPSS, which indicates that the Adjusted Residual from the designed model fits a normal distribution. Therefore, the designed model is correct.

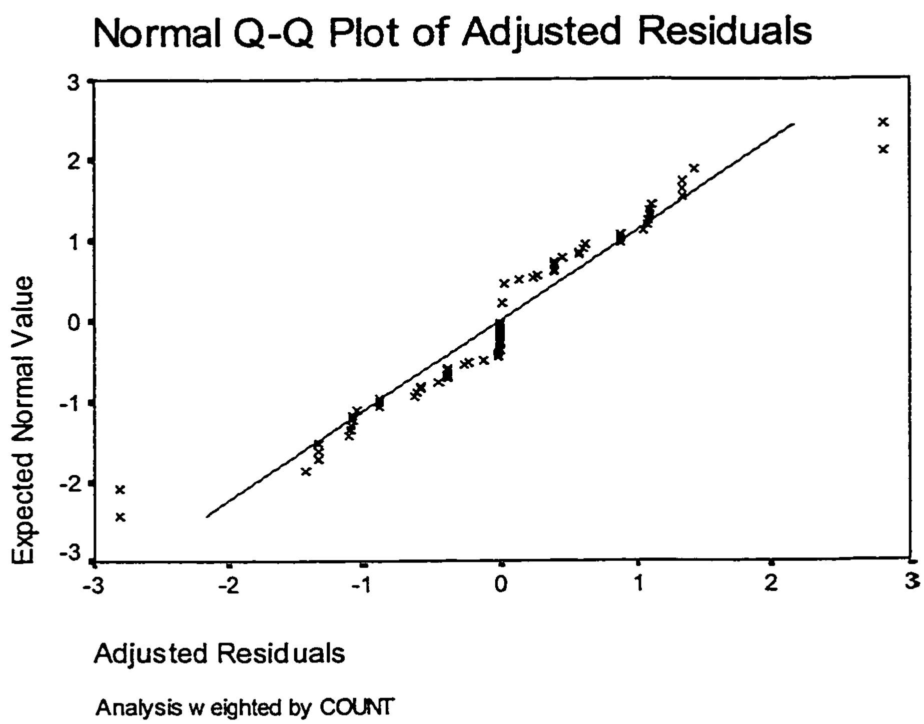

Figure 5-2 Q-Q Plot of Adjusted Residuals

5.4.3 Entropy Test

The third step in model diagnosis involves using an Entropy Test. Entropy is used to test the independence between the response variable and the explanatory variables and is defined by equation 5-13 (Haberman, 1982):

$$H(p) = -\sum_{j=1}^{s} p_j \log p_j \qquad \text{Eq 5-13}$$

where j is the integral, 1 ≤j ≤s, with probability p. The results of Entropy testing for the designed model from SPSS are presented in Table 5-12.

Table 5-12 Entropy Testing Results

Sourse of Dispersion	Entropy	Concentration	DF
Due to Model	111.967	91.8062	49
Due to Residual	426.734	295.4772	734
Total	538.7013	387.2834	783
*Chisq Test: P - Value = 1 - CDF.CHISQ(2*111.976) = 0.00*			

The dependence between Pipe Status and the chosen five parameters are significant if the *P-value* is close to zero.

Model diagnoses from the goodness-of-fit test and the residual examination indicates that the multinomial logit model we designed fits the observed data quite well. Therefore, we can accept this model as the final statistical model to provide a viable indicator for determining possible candidate sewers for CCTV inspection.

5.5 OUTCOME ANALYSIS

The outcome of Loglinear model analysis is illustrated in Table 5-9. This table shows the likely chance of pipe deficiency for 26 categories currently used in sewer system. Table 5-9 also tells the rehabilitation priority ranking. High rehabilitation priority will be placed on the pipes with high deficiency probability. If the pipe attributes are known, such as pipe age: 70 years; average depth of cover: 3 meter; pipe diameter: 200 mm; pipe material: CP and the waste type: CMB, the deficiency probability can be read from Table 5-9 is 82%, which is one of the pipes with the highest likely chance to failure. Recommended action is to make to inspect this kind of pipe to see what repairs need to do before its performance gets worse.

Table 5-13 Pipe Deficiency Probability Ranking

Pipe Age	Cover Depth(m)	Diameter (mm)	Material	Waste Type	Observed Def Frequency	Obs NonDef Frequency	Observed Frequency	Expected Def Frequency	Observed Def Probability	Expected Def Probability
60-90+	0-6	150-375	CP	CMB	14	3	17	14	82	82
0-29	6-8+	150-375	TP	SAN	26	12	38	26	68	68
60-90+	0-6	150-375	NC	CMB	24	13	37	24	65	65
60-90+	0-6	150-375	TP	CMB	159	89	248	158	64	64
0-29	0-6	150-375	TP	CMB	4	3	7	4	57	62
0-29	0-6	150-375	TP	SAN	22	15	37	22	59	59
30-59	0-6	150-375	CP	STM	10	7	17	10	59	59
30-59	0-6	150-375	TP	STM	6	5	11	6	55	55
0-29	6-8+	150-375	TP	CMB	1	1	2	1	50	50
30-59	0-6	450-525	RCP	CMB	1	1	2	1	50	50
60-90+	0-6	150-375	PVC	CMB	1	1	2	1	50	50
30-59	0-6	150-375	TP	CMB	25	32	57	25	44	45
60-90+	0-6	550-1050	RCP	CMB	4	5	9	4	44	44
60-90+	0-6	450-525	TP	CMB	7	15	22	7	32	34
0-29	6-8+	150-375	RCP	SAN	1	2	3	1	33	33
30-59	0-6	450-525	CP	CMB	1	2	3	1	33	33
0-29	0-6	150-375	CP	STM	11	24	35	11	31	31
30-59	0-6	150-375	CP	CMB	10	24	34	10	29	29
30-59	0-6	150-375	TP	SAN	4	10	14	4	29	29
60-90+	0-6	550-1050	TP	CMB	1	5	6	1	17	24
0-29	0-6	550-1050	TP	CMB	1	2	3	1	33	22
0-29	0-6	450-525	RCP	STM	2	8	10	2	20	20
30-59	0-6	450-525	TP	CMB	3	11	14	3	21	19
0-29	0-6	450-525	CP	STM	2	10	12	2	17	17
30-59	0-6	550-1050	TP	CMB	1	6	7	1	14	13
0-29	0-6	550-1050	RCP	STM	2	36	38	2	5	5

The deficiency probability from 82% to 5% covers 26 types of sewer pipes. The expected deficiency probability vs. observed deficiency probability is graphically presented in Figure 5-3. The graph indicates an fairly smooth fit of the designed model with the sample data.

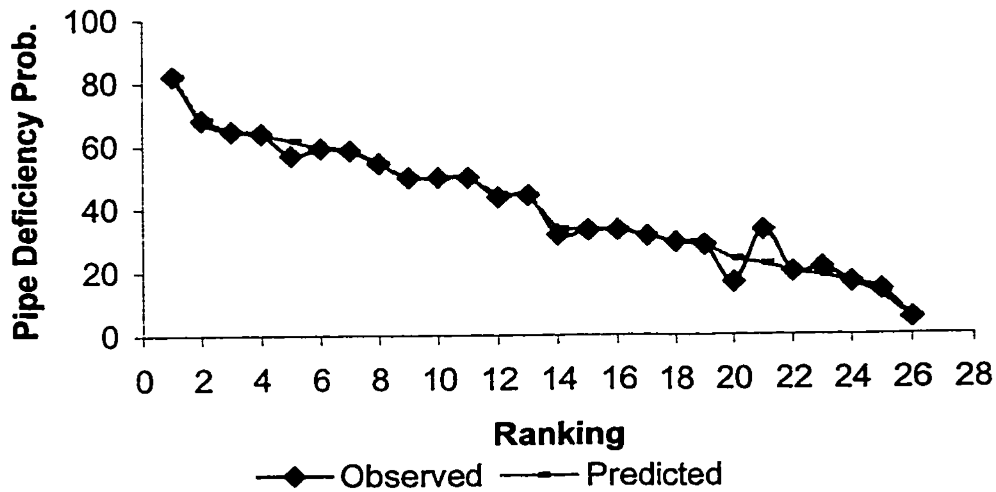

Figure 5-3 Expected and Observed Deficiency Probability vs. Ranking

The pipe deficiency probability ranking is an essential factor in sewer rehabilitation scheduling as it provides the decision-maker with a priority. A higher priority is placed on the higher deficiency probability of the sewer system. At the same time the decision-maker must take into account the different cost for different pipe deficiency, and the importance location factors etc. CCTV inspection can be scheduled to optimize inspection so that necessary repairs are performed in a proactive manner prior to collapse.

There are several factors involved in determining the threshold of the deficiency probability. If the pipe deficiency probability exceeds the threshold, inspection to that pipe must be proceeded to prevent the severe deficiency. If the pipe deficiency probability is lower than the threshold probability, low priority is placed in that pipe inspection. Budget availability, historical record and experience will play important role in determining the threshold value.

The other important application of deficiency probability of pipes is to determine budget allocation. The pipe deficient probability will be taken into account during the budget allocation. Neighborhood importance factor, pipe type constitution, and total length of pipe type may also affect budget allocation.

CHAPTER 6 LOGISTIC MODEL

6.1 INTRODUCTION

The deficiency probability for each combination of sewer pipe type was obtained from the Logit model analysis previously discussed in Chapter 5. This ranking provided valuable information on the type of sewer systems most likely to be in a deficient state and subsequently, assists the decision-maker in planning and scheduling of inspection and rehabilitation.

Recalling, from the Logit model analysis, all pipe attributes are considered categorical variables with each attribute further classified into several categories (i.e. pipe age contains three categories 0-29, 30-59 and 60-90+). Now that this approach has been proven mathematically valid, why not examine other statistical models to assess their validity in modeling the data? One such model, the Logistic Model, is described in this chapter. The explanatory variables in the Logistic Model may be quantitative or qualitative, which makes it possible to utilize the data previously used in the development of the Loglinear models. Additionally, the response variable, Pipe Status, is binary. The main difference between the logistic model and the linear regression model is that the outcome variable in logistic regression is binary.

The first section in this chapter defines the variable types used in the Logistic Model including Binary variables, Covariates, Dummy variables, etc. The Logistic Model is explained in detail in the second section. The third section outlines the model selection through initiation of the SPSS software. The last section discusses the result analysis and conclusions obtained from the Logistic Model.

6.2 VARIABLE DEFINITIONS

6.2.1 Binary Response Variable

A nominal scale variable having two categories is said to be *dichotomous*. For example, when considering whether a hospital patient has an illness, the categories would be "yes" or "no". The term binary response variable refers to any variable having only two possible outcomes. In sewer systems, the sewer status is the binary response variable and may be classified as either Deficiency or Non-deficiency.

6.2.2 Covariate

When a regression model simultaneously handles both quantitative and qualitative explanatory variables, the model combines elements of standard regression analysis, for which the predictors are quantitative, and analysis of variance, for which the predictors are qualitative.

Regression model can compare the mean of the response variable for several groups, treated as categories of a qualitative explanatory variable. However, in many applications, it is natural to do this while simultaneously controlling another quantitative variable. For example, when comparing the mean income for men and women, we might control differing levels of job experience between men and women (Agresti, 1997). The quantitative control variable is called a covariate. Typically, a covariate is correlated with the response variable and is also associated with the qualitative predictor, in the sense that some groups tend to have higher values than others on the covariate.

In the logistic regression model, the response variable is binary variable; all explanatory variables are considered as covariate. For sewer system all explanatory variables consist of two types: quantitative and qualitative variables. The quantitative variables such as pipe age, pipe diameter, and average depths of cover are covariate, while the qualitative variables such as pipe material and waste type are considered categorical covariates.

6.2.3 Dummy Variables

Dummy variables are artificial explanatory variables in a regression model, which can represent the categories of the qualitative variable. Each variable takes on only two values, 0 and 1, and indicates whether an observation falls in a particular group.

There are three categories: Sanitary; Storm; and Combined when considered waste type. The first artificial variable, denoted by Z_1, equals 1 for all observations from the sanitary group and equals 0 otherwise. The second, denoted by Z_2, equals 1 for observations from the Storm group and equals 0 otherwise. That is $Z_1=1$ and $Z_2=0$ for observations from the Sanitary group, and $Z_1= 0$ and $Z_2 =1$ for observations from the Storm group. Therefore, if $Z_1=Z_2=0$, the observation is from the third group, Combined.

The dummy variables identify the group that an observation represents. For example, the combination ($Z_1=0$, $Z_2=0$) occurs for all subjects in the third group. The dummy variables denote classification, not magnitude, of an observation on the qualitative predictor.

Dummy variable coding works because it assumes no distance between groups. Recall, that we assigned the nominal value for each sewer type as presented in Table 6-1. If we set $Z =1$ for the Sanitary group, $Z=2$ for the Storm group, and $Z=3$ for the Combined group, an ordering, as well as equal distance between groups, is assumed when substituting into the logistic model. The qualitative variable is also treated as if it were quantitative, which is improper. It takes only one term in a regression model to represent the linear effect of a qualitative explanatory variable, whereas it requires 2 terms to represent the 3 categories of a qualitative variable. The logistic model takes the ordinary multiple regression form, however, dummy variables are considered for each qualitative predictor.

Table 6-1 The Two Dummy Variables for Three Groups

Group	Z1	Z2	Nominal Value
Sanitary	1	0	1
Storm	0	1	2
Combined	0	0	3

6.2.4 Interval Variables

An interval variable is one that does have numerical distances between any two levels of thee scale. These are called interval scales. Pipe age, pipe diameter, and average depth of cover are considered to be interval variables within the interval scale. The interval between 15 years and 20 years of pipe age is the same as the interval between 20 years to 25 years. This observation is also similar for the difference between 200 mm and 250mm of pipe diameter and between 250 mm and 300 mm.

6.3 LOGISTIC REGRESSION MODEL

6.3.1 Odds

The odds for a binary response variable equal to the number of successes divided by the number of failures. Suppose we have two outcomes for the response variables:

Success and Failure. π denotes the probability of success and $(1-\pi)$ denotes the failure probability. The odds of success may be defined as:

$$\text{Odds} = \text{Probability of success / Probability of failure} = \pi / (1-\pi) \qquad \text{Eq 6-1}$$

For instance, if the probability of success equals 0.75, then the probability of failure equals 0.25, and the odds of success is then calculated as being 0.75/0.25 = 3, meaning that a success is three times as likely to occur as a failure. We expect about three successes for every failure. For sewer systems, the pipe deficiency is the variable interest. The odds of pipe deficiency is represented by:

$$\text{Pipe Deficiency Odds} = \text{Prob. of Pipe Deficiency / Prob. of Non-Deficiency} \qquad \text{Eq 6-2}$$

Subsequently, the pipe deficiency probability is the function of the Odds:

$$\text{Deficiency Probability} = \text{Odds} / (\text{Odds} + 1) \qquad \text{Eq 6-3}$$

6.3.2 Definition of the Logistic Model

Many categorical response variables have only two categories, which refer to binary response variables. The observation for each subject might be classified as a "success" or a "failure" and could be represented by either 1 or 0. For sewer systems the pipe deficiency category is of interest. Therefore, the binary response is the Pipe Status with Deficiency (Y=1), and Non-Deficiency (Y=0).

The mean of 0 and 1 scores, which is the sum divided by the total sample size, equals the proportion of interest category. Therefore, the mean of Y is the probability of Y=1.

For binary response variables, the model describes how the proportion of interest category depends on the explanatory variables. For example, let $\pi = E(Y)$ denote the true proportion of interest category. Now π also represents the probability that a randomly selected subject has an interest category response and this probability varies according to the values of the explanatory.

Consider the following example, where there is only one explanatory variable X, the logistic regression model may be written as:

$$Log(\frac{\pi}{1-\pi}) = \alpha + \beta X \qquad \text{Eq 6-4}$$

The ratio $\pi/(1-\pi)$ is equal to the odds. When $\pi = 0.65$, the odds equals 0.65/0.35 =1.86, meaning that a success is 1.86 times as likely as a failure. The equation uses the LOG of the odds, log $[\pi/(1-\pi)]$, called the *logistic transformation*, or *logit* for short. Equation 6-4 can be rewritten as following:

$$Logit(\pi) = \alpha + \beta X \qquad \text{Eq 6-5}$$

As π increases from 0 to 1, the odds increase from 0 to ∞ and the Logit increases from $-\infty$ to ∞. The probability π = 1/2 has odds equal to 1 (= (1/2)/(1-(1/2)), and the Logit equals to 0 (the π values above ½ have positive Logits, below ½ have negative Logits).

The logistic response curves, shown in Figure 6-1, have an S-shape. Using the curves, the probability of a success falls between 0 and 1 for all possible X-values.

The parameter β in this model refers to whether the curve increases or decreases as X increases. For $\beta>0$, π increases as X increases. For $\beta<0$, π decreases as X increases; the probability of success tends toward 0 for large values of X, as in curve (2).

If $\beta = 0$, π does not change as X changes. Therefore, the curve flattens to a horizontal straight line (Agresti, 1997).

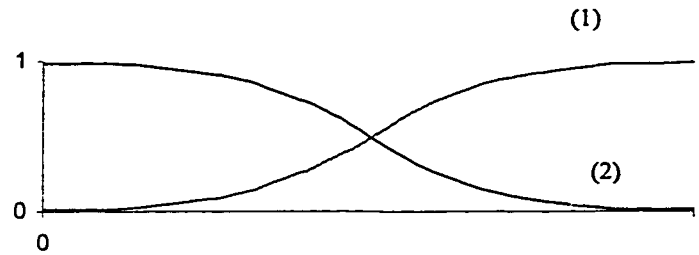

Figure 6-1 Logistic Regression Model

6.3.3 Distribution Assumption

For binary data, it makes more sense to assume a binomial distribution rather than a normal distribution for the response.

For observations on a categorical variable with two categories, the binomial distribution applies when the following three conditions hold true (Agresti, 1997):

- For a fixed number of observation *n*, each falls into one of two categories.
- The probability of falling in each category, π for the first category and $(1-\pi)$ for the second category, is the same for every observation.
- The outcomes of successive observations are independent; that is, the category that occurs for one observation does not depend on the outcomes of other observations.

The *Bernoulli* distribution for binary random variables specifies probabilities $P(Y = 1) = \pi$, and $P(Y=0) = 1-\pi$ for the two outcomes, for which $\pi = E(Y)$. When Y_i has a Bernoulli distribution with parameter π_i, the probability density function is:

$$f(y_i,\pi_i) = (1-\pi_i)\exp\left[y_i \log\left(\frac{\pi_i}{1-\pi_i}\right)\right] \qquad \text{Eq 6-6}$$

Where $y_i = 0$ or 1.

6.3.4 Logistic Model for Sewer System

The same five explanatory variables used in Chapter 5 were initially selected for model development. Of these variables, pipe age, pipe diameter, and average depths of cover are defined as interval co-variables in the Logistic model to take the advantage of data knowing. For instance, since the accurate pipe age is known, say 53 years, and we do not need to classify it into the period of 39-59 years in the logistic model analysis. Pipe material and waste type are qualitative values, and therefore, are defined as categorical co-variables in the logistic model. Categorical variables in any logistic model works similar to dummy variables. The response variable of the logistical model is given in binary format (i.e. response variable = Pipe Status, with deficiency assigned the value 1, and non-deficiency value 0).

For binary response variables, the logistic model describes how the proportion of deficiency pipe depends on the chosen five explanatory variables. For example, let π denote the probability that a randomly selected subject has a deficient response.

$$\log\left(\frac{\pi}{1-\pi}\right) = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 \qquad \text{Eq 6-7}$$

where α is the interception, X_1 is Pipe Age, X_2 is Pipe Diameter and X_3 is Average Depth of Cover. X_4 is the effect of Waste type and X_5 is the effect of Pipe Material. β_1-β_5 is the regression coefficient for individual variables and π / $(1-\pi)$ is the Odds.

6.4 LOGISTICAL MODEL SELECTION

6.4.1 Linear Regression Variable Selection Method (Backward Stepwise)

Linear regression variable selection method allows the users specify how independent variables are entered into the analysis. Backward Elimination procedure begins with the complete regression model, one that includes all possible independents variables, and attempts to eliminate them from the model one at a time. For the logistic model, the Backward Stepwise procedure is chosen to perform the model selection. Backward Stepwise elimination enters all of the variables in the block in a single step and then removes them one at a time based on removal criteria.

For categorical covariates, the reference category is required to be determined, so that all other categories can be compared to this reference one. Such as for the pipe waste type, there are three categories: Sanitary, Storm and Combined. Specify the last category: Combined type as the reference category, the effect of Sanitary and Storm pipe types to the pipe status will be compared to that of pipes with Combined waste type. SPSS offers an easy way to specify the reference category. Either the *First* or the *Last* category in the variable can be chosen to be the reference category by selecting the "*First*" or "*Last*" option in SPSS.

6.4.2 Main Effects In the Model

Five explanatory variables and one response variable initially are included in the Logistic Model. Pipe Age; Pipe Diameter; Pipe Material; Average Cover Depth and Sewer Type. Response variable is Pipe Status.

Use the unweighted raw data to run the SPSS software by choosing the Backward Likelihood Ratio Elimination Stepwise mode. For dummy variables, choose the last category as reference category.

Model selection consists of two steps: To exam the significance for each parameter by performing Wald test and to determine which parameters should be included in logistical regression model by likelihood-ratio test.

Step 1

To examine the significance for each parameter by performing Wald test. Wald statistic is the square of z test. The hypothesis H_0: β_i =0 states that X_i has no effect on the probability π that Y =1. It has a chi-square distribution with df = 1. It has the same P-value as the Z statistic for two-side alternative H_a: $\beta \neq 0$.

Step 2

To determine which parameters should be included in logistical regression model by likelihood-ratio test. Likelihood-ratio test compares two models, by testing the extra parameters in the complete model equal zero. The test refers a key ingredient of maximum likelihood inference, the likelihood function, denoted by L. The formula for likelihood-ratio test statistic is

$$-2\log\left(\frac{L_0}{L_1}\right) = (-2\log L_0) - (-2\log L_1) \qquad \text{Eq 6-8}$$

It compares the maximized values of (-2logL) when the null hypothesis is true and when it need not be true (Agresti, 1997). Figure 6-2 displays the prediction of logistic regression model. The result shows Average Cover Depth and the Pipe Material are non-significant to the Pipe Deficiency so that they are excluded in the final model. This conclusion is same as the previous one we draw from Logit Model analysis (See Chapter

5). Here we use the parsimonious rule to determine the terms in regression equation. The logistic regression model (only main effects) is:

$$Log(\frac{\pi}{1-\pi}) = -0.0839 + 0.0259X_1 - 0.0039X_2 + 1.2191Z_1 + 0.3085Z_2 \qquad \text{Eq 6-9}$$

X_1 is the pipe age (year); X_2 is the pipe diameter (mm); Z_1, Z_2 are dummy variables for waste type. β_1-β_4 are read from the SPSS printout (Figure 6-2).

- For Sanitary waste type $\beta_3 = 1.2191$, Z_1=1; Z_2=0
- For Storm waste type $\beta_4 = 0.3085$, Z_1=0; Z_2=1
- For Combined waste type Z_1=0; Z_2=0

The equation expressing the logistic regression model directly in terms of π (the deficiency probability):

$$\pi = \frac{e^{(-0.0839+0.0259X_1-0.0039X_2+1.2191Z_1+0.3085Z_2)}}{1+e^{(-0.0839+0.0259X_1-0.0039X_2+1.2191Z_1+0.3085Z_2)}} \qquad \text{Eq 6-10}$$

Figure 6-2 SPSS Printout of Logistic Regression Model

------------------------ Variables in the Equation ------------------------

Variable	B	S.E.	Wald	df	Sig	R	Exp(B)
AGE	.0259	.0042	37.3441	1	.0000	.1811	1.0262
DIAMETER	-.0039	.0006	38.3782	1	.0000	-.1838	.9961
WASTETYP			14.9710	2	.0006	.1009	
WASTETYP(1)	1.2191	.3221	14.3221	1	.0002	.1069	3.3840
WASTETYP(2)	.3085	.2865	1.1592	1	.2816	.0000	1.3613
Constant	-.0839	.3001	.0782	1	.7798		

----------------- Model if Term Removed -----------------

Term Removed	Log Likelihood	-2 Log LR	df	Significance of Log LR
AGE	-482.173	41.112	1	.0000
DIAMETER	-485.820	48.405	1	.0000
WASTETYP	-469.342	15.449	2	.0004

--------------- Variables not in the Equation ----------------

Residual Chi Square 2.327 with 5 df Sig = .8022

Variable	Score	df	Sig	R
DEPTH	.5443	1	.4607	.0000
MATERIAL	1.5413	4	.8193	.0000
MATERIAL(1)	.0175	1	.8949	.0000
MATERIAL(2)	.6314	1	.4269	.0000
MATERIAL(3)	.4368	1	.5087	.0000
MATERIAL(4)	.0426	1	.8365	.0000

No more variables can be deleted or added.

6.4.3 Interaction Determination

Cross-product terms allow interaction among explanatory variables in their effects on the response. For current logistic model, there are three explanatory variables: Pipe Age, Pipe Diameter and Sewer Type. The possible cross-product terms are Pipe Age × Sewer Type, Pipe Age × Pipe Diameter and Pipe Diameter × Sewer Type. Based on the logit model analysis presented in Chapter 5 (Partial correlation) we conclude that no interaction exists between Pipe Age and Sewer Type. Therefore, we just exam the effect of the other two cross products to the Pipe Status. To compare the effect of these two cross products we need to compare the following two logistic models.

$$\log(\pi) = \alpha + \beta_1 Age + \beta_2 Diameter + \beta_3 Waste \qquad \text{Eq 6-11}$$

$$\log(\pi) = \alpha + \beta_1 Age + \beta_2 Diameter + \beta_3 Waste + \gamma_1 Age \times Sewer + \gamma_2 Diameter \times Waste \qquad \text{Eq 6-12}$$

The first model only contains the main effects (Simple model). The second model contains the main effect along with the cross products (Full model). The difference of Likelihood-ratio (-2LogL) for the two models will be calculated, which is an approximate chi-square statistic with *df* given by the extra parameters in the full model. Run SPSS for these two models to obtain the Logistic regression analysis printouts:

For simple model:

-2Log Likelihood = 923.235

For full model:

-2Log Likelihood = 915.429

The different of *–2Log L = 923.235-915.429 = 7.806*, the *df* for the cross-product is 4, check the chi-square table to get the P value = 0.1 > 0.05, so we accept the hypothesis, that is the effects of cross-products are non-significant.

6.5 RESULT ANALYSIS

An interpretation of the logistic regression coefficient β is as an effect on the odds. Specifically, applying antilogs to both sides of the logistic regression equation, take the simple example:

$$Log(\frac{\pi}{1-\pi}) = \alpha + \beta X \qquad \text{Eq 6-13}$$

$$\frac{\pi}{1-\pi} = e^{(\alpha+\beta X)} = e^{\alpha}(e^{\beta})^{X} \qquad \text{Eq 6-14}$$

The right side of the equation has a constant multiplied by another constant raised to the X power. This exponential relationship implies that every unit increase in X has a multiplicative effect of the e^{β} on the odds.

The effects of three pipe attributes to pipe deficiency obtained from the logistic model equation 6-9 are presented in Table 6-2.

Table 6-2 Logistic Regression Model Outcomes

Attributes	Coefficient β	EXP(β)	Meaning
Age (year)	0.02590	1.02624	Deficiency odds increase by 2.62% when age increases one year.
Diameter(mm)	-0.00390	0.99611	Deficiency odds decrease by 0.39% when the diameter increases 1 mm.
Combined		1.00000	The deficiency odds for Combined type are assumed 1.
Sanitary	1.21910	3.38414	The deficiency odds for Sanitary type is as 3.38414 times as that of Combined .
Storm	0.03085	1.36138	The deficiency odds for Storm type is as 1.36138 times as that of Combined .

6.5.1 Age Effect to the Pipe deficiency

The logistic regression coefficient β for Age is 0.0259. *EXP(β) = EXP(0.0259) = 1.0262*. When the age increase by one year, the estimated odds of the pipe failure is multiplied by 1.0262; that is it increases by 2.62%. When the age increases 10 years, the estimated odds of pipe failure increases $(1.0262)^{10} = 1.295$ times. For example, the odds of pipe failure at age $X_1 = 50$ years is 1.295 times of that odds for pipe age $X_1 = 40$. That is the chance of pipe deficiency likeliness over the pipe non-deficiency likeness will increase 29.5% comparing to the pipes with age decreasing by 10 years.

6.5.2 Diameter effect to the Pipe deficiency

The logistic regression coefficient β for Diameter is 0.0039, *EXP(β) = EXP(-0.0039) = 0.9961*. Notice the β is negative, therefore when the pipe diameter increases by 1 centimeter, the estimated odd of pipe failure is multiplied by 0.996; that is a decreases of 0.39%. When the diameter increases 100 centimeter, the odd of pipe failure is $(0.9961)^{100} = 0.677$ times the original one. That is the chance of pipe deficiency likeliness over the pipe non-deficiency likeness will increase 32.3% comparing to the pipes with diameter decreasing by 100 mm.

6.5.3 Waste Type Effect to the Pipe Deficiency

If the effect of combined waste type to the odds of pipe failure is 1, the effect of sanitary waste type to the odds of pipe failure is 3.384; the storm effect is 1.3613. Sanitary pipe has the biggest effect to the pipe deficiency; the effect of storm pipe is the next to sanitary; combined is the last one.

6.5.4 Confidential Interval for Each Coefficient

For each coefficient in the logistic model, since the standard error is known, the odds interval can be calculated at the 95% confidence level. The confidence interval for the log odds ratio is:

$$\text{Coefficient Value} \pm 1.96 * (\text{Standard Error})$$

For example, for pipe age, the coefficient value in logistic model is 0.0259, the log odds confidential interval (shown in Table 6-2) is 0.259 ± 1.96 * 0.0042 = (0.034, 0.018). Applying the antilog to each end point, the odds confidence intervals are calculated and presented in last column of Table 6-3.

Table 6-3 Logistic Regression Model Confidential Interval for Each Parameters

Parameter	Coefficient	Standard Error	Log Odds Confid. Interval	Odds Confid.Interval
Age	0.0259	0.0042	(0.034, 0.018)	(1.018, 1.035)
Diameter	-0.0039	0.0006	(-0.003, -0.005)	(0.995,0.997)
Sanitary	1.2191	0.3221	(1.850, 0.588)	(1.800, 6.362)
Storm	0.03085	0.2865	(0.592, -0.531)	(0.588, 1.808)
Constant	-0.839	0.3001	(-0.251,-1.427)	(0.240, 0.778)

CHAPTER 7 TASES PROGRAM

7.1 INTRODUCTION

The concept of a windows-based computer program named Trends Analysis Sewer Evaluation System (TASES) was developed to serve as a tool for determining sewer rehabilitation plan/strategy for the proactive implementation and scheduling of the Local Sewer Rehabilitation Program. TASES is intended to assist City of Edmonton personnel in determining the projected level of deficiency of a given sewer based on "objective" criteria developed from historical records. The initial and introductory screens are presented in Figure 7-1 and 7-2, respectively.

The conceptual framework for TASES is written in Visual Basic 4.0. The user is instructed to input the Pipe ID number and information on the five attributes used in the Trends Analysis: 1) waste type; 2) pipe material; 3) pipe age (year of construction); 4) average depth of cover; and 5) pipe diameter. The deficiency weights developed from the Trends Analysis, combined with the interaction between these attributes, are used to calculate the expected level of deficiency. TASES utilizes the outputs to inform the user as to whether the pipe is in the Urgent, Moderate, or Normal Needs category. All inputted information is saved in a database and can print a hard copy file of candidate sewers for inspection based on ranking. A graphical taskbar allows the user to visualize the predicted level of deficiency. Figure 7-3 illustrates the main program screen of TASES (Ariaratnam, et al. 1998).

One feature of TASES is an information box that solicits identification of the person performing the analysis and the date. This feature is intended to create a historical record for accountability. All information is saved in a database and may be printed by overall records or Needs category (i.e. Urgent, Moderate, or Normal).

All pipe deficiency probabilities used in TASES program are the analysis result of Logit Model. The deficiency probabilities felling within 66.7% to 100% range are in the Urgent level; From 33.3% to 66.7% are in the Moderate level and from 0 % to 33.3% are in Normal level. The thresholds for Urgent, Moderate and Normal are taken based on the experience and budgetary. Those thresholds need updated periodically to reflect the latest situation. If the pipe deficiency probability lies in the Urgent level, which means CCTV inspection required for this kind of pipes to prevent sever deterioration occurring; the Moderate level means CCTV inspection needed. An inspection schedule needs to be set. This schedule also takes into account the budget constrains. No special actions need for Normal level pipes since they are in lower deficiency probability.

TASES does not require an interface with any mainframe computer to operate and can be utilized by field personnel through a PC-based system. This provides the capability to determine sewer sections that are likely candidate for immediate inspection while in the field. It is recommended that future enhancement be made to interface TASES to the DRAINS database thus requiring the user to input Pipe ID as the only parameter.

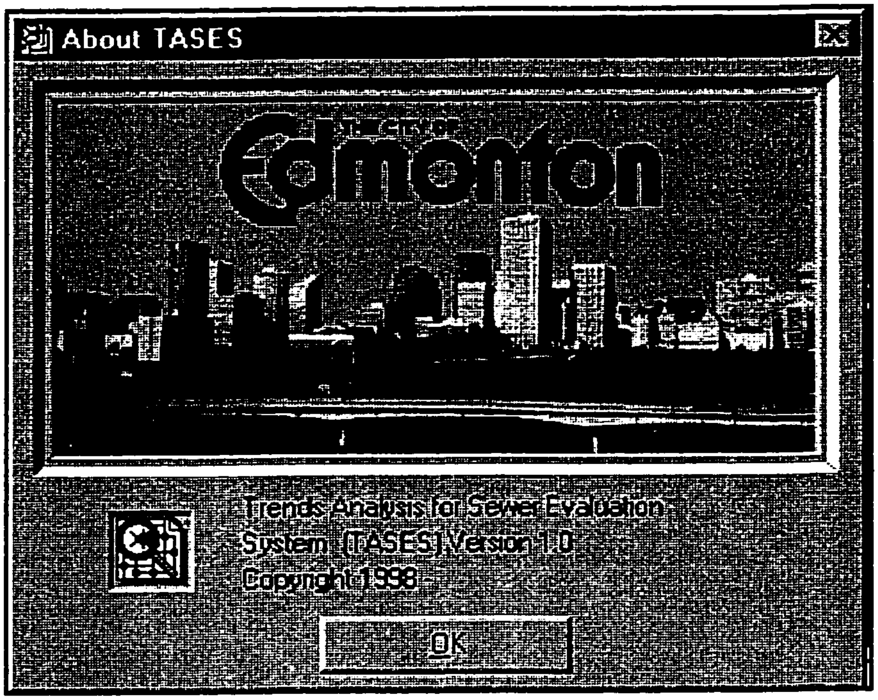

Figure 7-1 TASES Program Initial Screen

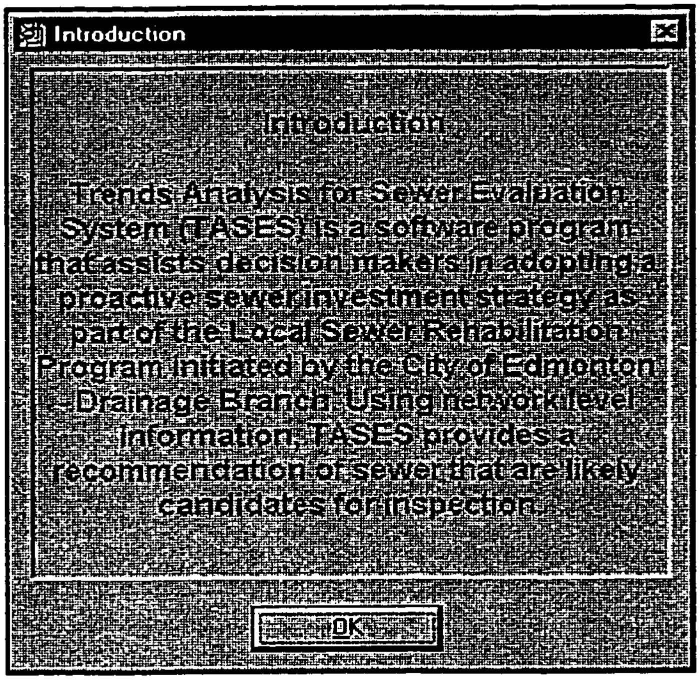

Figure 7-2 TASES Introductory Screen

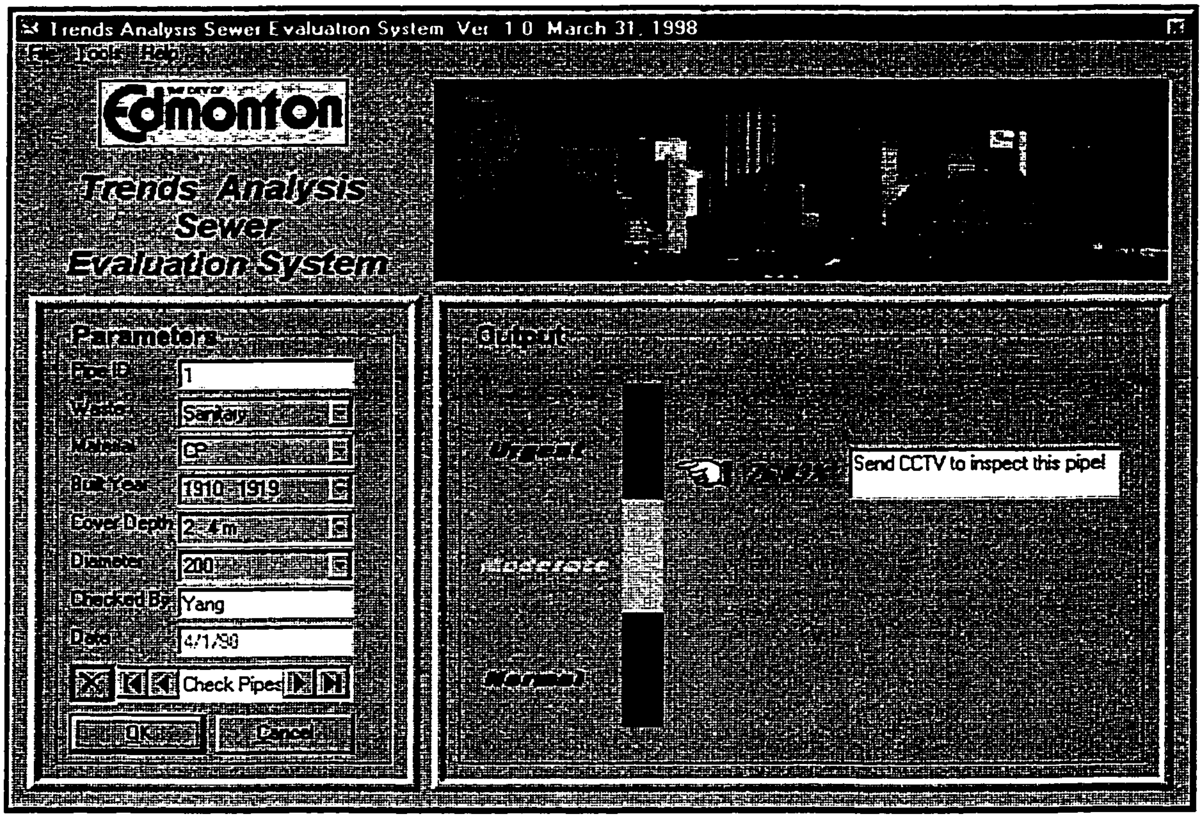

Figure 7-3 TASES Program Main Screen

7.2 FLOW CHART FOR TASES

TASES program consists of two components: pipe attributes inputs and analysis outcome display. Information on sewer pipe ID and attributes (i.e. Pipe Unique ID number; Waste Type; Material; Built Years; Average Cover Depth and Pipe Diameter) are required to initiate the program. Once all pipe attributes have been selected, the pipe combination is set. The deficiency probability for this pipe combination is matched from the Logit ranking model, that is, the accurate deficiency probability is obtained. Based on the threshold for three levels: Urgent (66.7% - 100%); Moderate (33.3% - 66.7%) and Normal (0% - 33.3%) the obtained deficiency probability will fall into one of these levels. At the same time, recommendation for the rehabilitation will be shown in the user interface. The pipe attributes and the deficiency probability, along with the recommended rehabilitation measures are stored into a database so that these records may be retrieved at later time. The flowchart of TASES is presented in Figure 7-4.

7.3 CONCLUSIONS

TASES program plays an essential role in the implementation of the outcome of this research project. The Window's interface screen makes the program easy to use. Additionally, TASES may be used either in the office or on-site.

TASES program still needs to be sophisticated enough to update itself in order to reflect the most current conditions in the sewer system. When the latest sample data are available and inputted into TASES, the deficiency probabilities must be generated and updated automatically by TASES program. In some cases, if the sample data are extracted from a specific area, such as, an industrial area, and are inputted into TASES, The program must be able to generate the current deficiency probabilities to reflect any unique feature for that area.

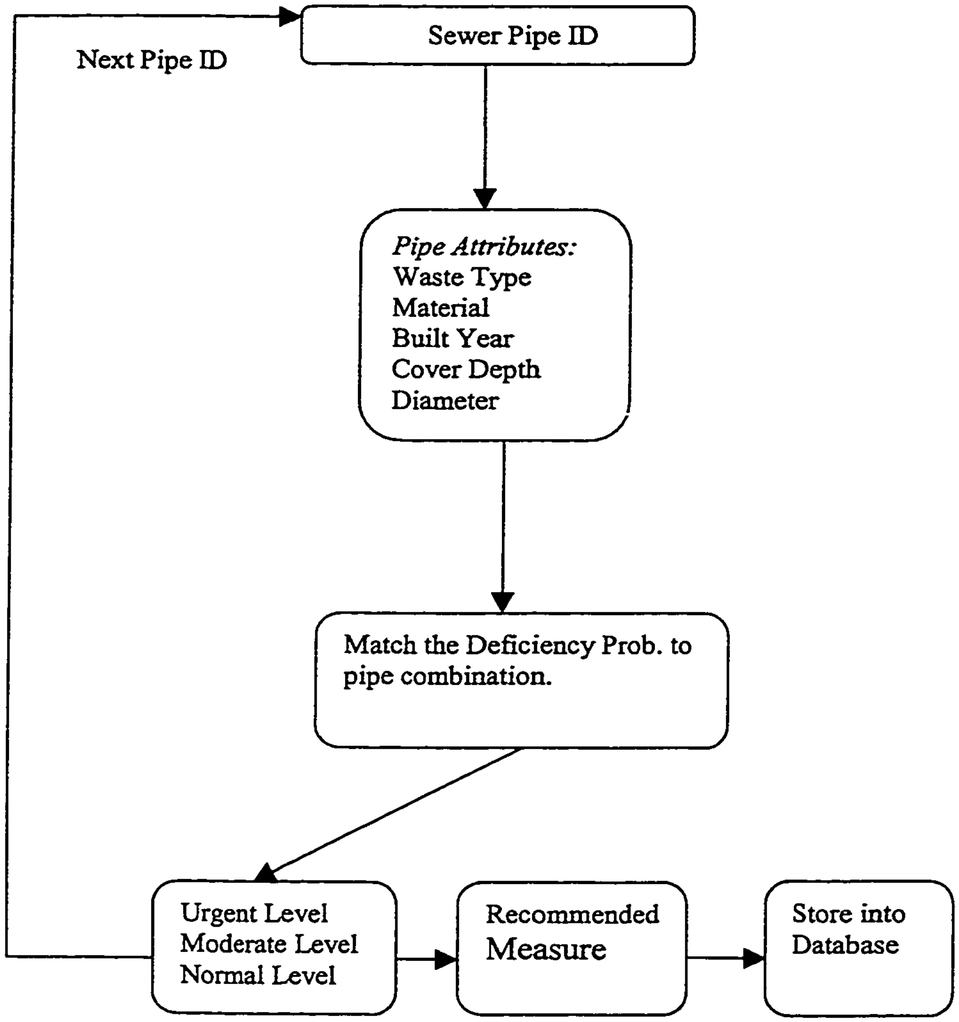

Figure 7- 4 TASES Program Flow Chart

CHAPTER 8 CONCLUSIONS

8.1 CONCLUSIONS

Due to their low visibility, rehabilitation of underground sewer system is often neglected until a catastrophic failure occurs. This, more often than not, results in costly and difficult rehabilitation due to the urgent nature of ensuring that the sewer system is operational. A majority of sewer repair projects are executed on a "reactionary" basis, rather than adopting a "proactive" approach. There are two main reasons for this: the first is the availability of adequate information regarding the condition of the sewer system; and the second is the ineffectiveness of predicting sewer deficiency prior to failure so that inspection and repairing could be performed prior to failure of the system.

The main contribution of this thesis is the development of two statistical models to assess sewer infrastructure inspection requirements. These pipe deficiency probability ranking and prediction models are developed to improve the objectivity of the pro-activity.

Recalling, the first objective of this research is to examine which pipe attribute among the five significantly contributes to the pipe deficiency. The Likelihood-Ratio Chi-square test in the Logit model analysis (Chapter 5) revealed that pipe age, pipe diameter, and waste type significantly contributed to deterioration, while the pipe material and the average depth of cover have little effect. The same conclusion is drawn from the Logistic model analysis (Chapter 6); however, the pipe material and average depth of cover are included in the Logit model analysis although they are not significant to the pipe deficiency. The reason for including those two parameters in the model is to improve the accuracy of the Logit model; however, pipe material and average depth of cover are excluded in the logistic model based on the parsimonious rule set by the SPSS program.

Main effects test and the *K*-way test, along with the partial association test, were employed to determine the various used in the Logit model as suggested by equation 5-10. A detailed Logit model diagnosis (i.e. goodness-of-fit test; residual analysis and entropy test), proved that the designed model fits the observed data quite well. The deficiency probability ranking obtained from the Logit model is illustrated in Table 5-13, and is required achieve the second objective of this research. The deficiency probabilities for 26 types of pipe currently used in the sewer system of City of Edmonton were developed. For a given sewer pipe type (i.e. based on the five attributes) the likely chance of failure can be determined from Table 5-13. The inspection requirement priorities are placed on the pipes with high deficiency probabilities. The deficiency probabilities for all existing pipe types are invaluable information for the decision-maker to initiate the proactive sewer rehabilitation strategy. TASES program presented in Chapter 7 is developed to facilitate the implementation of proactive actions using the result of Logit model.

In the logistic regression model analysis, five attributes are assumed to contribute to pipe deficiency and considered as explanatory variables. The response variable, pipe status, is a binary variable. Pipe age, pipe diameter, and average depth are not classified into categories in the logistic model; however, pipe material and waste type are considered as categorical covariates. The Wald test and likelihood-ratio test excluded the pipe material and average depth of cover from the initial logistic model. The final logistic model is illustrated in equation 6-9. The outcome of the logistic model is explained in detail in section 6.5 "Result Analysis" and an overview of outcomes are presented in the Table 6-2.

The deficiency probability for a given pipe type (i.e. age, diameter, and waste type) is calculated by equation 6-10. The logistic model serves as a prediction model to predict the likelihood of a sewer pipe entering a deficient state through aging. The effects of pipe age, pipe diameter, and waste types to pipe deterioration are quantitative by logistic model. This information should also assist the decision-maker in short-term and long-term planning for proactive strategy, which reaches the third objective of this research.

An early schedule may be developed to inspect the sewer system according to the deficiency probability magnitude. This action should minimize any adverse effect to the general public, such as the traffic congestion.

The subjectivity of annual budgetary allocation could be decreased by considering the sewer pipe deficiency probability along with other factors including pipe constitution and the importance of neighborhood to the whole city.

The methodology depicted in this thesis can also be applied to other infrastructure such as highways, railroads, bridges, etc. provided that sufficient historical data records are available.

8.2 RECOMMENDATIONS FOR FUTURE WORK

8.2.1 More Objective Pipe Status Evaluation

Currently, sewer system condition assessment and evaluation involves subjective judgement and experience. Specifically, how to define pipe failure or not failure to a great extent depends on experience. Although the structural rating system, such as the one employed by the City of Edmonton, attempts to subjectively reflect the pipe performance condition, the criteria to determine pipe deficiency or non-deficiency is still based largely on experience. It is recommended that the objectivity of pipe status evaluation be improved by using state-of-the-art inspection techniques (presented in Chapter 2) to inspect and record the pipe conditions as accurately as possible. In addition, a sophisticated pipe condition assessment system should be established to reflect pipe performance.

Structural failure and hydraulic concerns are two main factors contributing to pipe deterioration. Only structural rating is used in the analysis because historical hydraulic records are not available. It is recommended that hydraulic effects on pipe deterioration be considered. Subsequently, historical pipe hydraulic information should be recorded and stored for future analysis.

8.2.2 Logistic Model Validation

A common method to validate a regression model is to compare the difference between the observed value and the expected value of interest parameter. Since observed deficiency odds ratios for all possible combinations of pipe age, diameter, and waste type are not available, the logistic model cannot be validated by this method. Additionally, the logistic model cannot be diagnosed using any of the methods employed for the Logit model, such as goodness-of-fit test.

Although the sample data is analyzed by utilizing two statistical models, the results from the two models cannot be compared with each other because the variables are dealt in different ways in those models.

One possible approach to validating the logistic model is through "sensitivity analysis". This entails separating the data into two groups (20% and 80%) and initially using only 80% of the data to perform the logistic model analysis. Once the logistic model has been established, the remaining 20% of sample data may be used to predict the deficient probability. Comparing the observed deficient probability and the observed probability for the 20% sample data to validate the logistic model. The sample data in this research is not enough to utilize this method. Once additional data is collected, validation of the logistic model can be performed.

8.2.3 TASES Program Improvements

TASES program currently requires manual data entry of the Pipe ID and the pipe attributes. It is recommended that future enhancement be made to link the TASES program to the existing DRAINS database thus requiring the user to input Pipe ID as the only parameter. TASES would then be able to navigate all pipe records and retrieve the pipe deficiency probability automatically.

The other feature recommended to improve in the TASES program is self-updating capability. When TASES is implemented in the sewer rehabilitation, it must be able to update itself to reflect the most current change in the sewer system. To realize this function, the SPSS software package should be integrated in TASES. TASES will then be able to update the ranking model and the prediction models as long as new sample data are inputted. The decision-maker should continue collecting the latest sample data and entering these into TASES, therefore, generating both ranking and prediction results automatically.

CHAPTER 9 REFERENCES

Agresti, Alan, *Categorical Data Analysis*, John Wiley & Sons, New York, 1990.

Agresti, Alan and Finlay, Barbara, *Statistical Methods for the Social Science*, Third Edition, Prentice-Hall , NJ, 1997

Ariaratnam, S.T., A. El-Assaly, and Y. Yang, "Local Sewer Rehabilitation Strategy – Stage I: Trends Analysis and Methodology Report", Report Submitted to the City of Edmonton, University of Alberta, Department of Civil and Environmental Engineering, Edmonton, AB, 1998.

Bechhofer R.E, Santerner T. J., and Goldsman D. M. *Design and Analysis of Experiments for Statistical Selection Screening, and Multiple Comparisons*, John Wiley & Sons, Inc. New York, 1995.

Bennett, R.J., *Spatial Time Series,* Pion Limited, U.K., 1979.

Berenson, Mark L., David M. Levine, and Matthew Goldstein, *Intermediate Statistical Methods and Applications,* Prentice-Hall Inc., New Jersey, 1983.

Bowerman, B., and R. O'Connell, *Forecasting and Time Series: An Applied Approach,* Duxbury Press, CA, 1993.

CH2M Hill Engineering Ltd., "Inner City Prioritization Study: Drainage System Assessment," *Report Prepared for the City of Edmonton, Drainage Branch,* Edmonton, AB, March 1993.

Chatfield, C., *The Analysis of Time Series: An Introduction,* Chapman and Hall, N.Y., 1984.

Chouinard, L.E., G.R. Andersen, and V.H. Torrey III, "Ranking Model Used For Condition Assessment of Civil Infrastructures System," *ASCE J. of Infrastructure Systems,* Vol.2, No. 1, 1996, pp. 23-29.

City of Edmonton, "Sewer Physical Condition Classification Manual," *Internal Report,* Edmonton, AB, March 1996.

City of Edmonton, "Standard Sewer Condition Rating System Report," *Internal Report,* Edmonton, AB, March 1996.

City of Edmonton, "Standard Sewer Condition Rating System Report: Technical Memorandum," *Internal Report*, Edmonton, AB, March 1996.

Existing Sewer Evaluation & Rehabilitation, WEF Manual of Practice FD – 6, ASCE Manual and Report on Engineering Practice No. 62, Water Environment Federation (Alexandria, VA) and American Society of Civil Engineers (New York, NY), 1994.

Gower, Allen, and Karen Kelly, *How Big Should The Sample Be*? Social Survey Methods Division, Statistics Canada, 1993, P1-P14

Greenwood, P. E. and Niskulin M. S., *A Guide to Chi-Square Testing*, John Wiley & Sons, Inc. New York, 1996.

Harberman, Shelby J. Analysis of Dispersion of Multinomial Responses, *J. of American Statistical Association*, Vol. 77, No. 379, Sept. 1982. P568-P580

Lindsey, J. K. *Modeling Frequency and Count Data*, Clarendon Press, Oxford, 1995.

Kennedy, John J. *Analyzing Qualitative Data: Log-Linear Analysis for Behavioral Research*, Second Edition, Praeger Publishers, New York, 1992.

Khazanie, Ramakant, *Elementary Statistics in a World of Applications*, Third Edition, Scott, Foresman and Company, Glenview, Illinois, 1990.

Ott, L., *Understanding Statistics,* PWS Publishers, Boston, MA, 1985.

Read, Timothy R.C. and Cressie Noel A.C. *Goodness-of-Fit Statistics for Discrete Multivariate Data,* Spinger-Verlag New York Inc. 1988.

SPSS Advanced Statistics 7.5, SPSS Inc., Chicago, Illinois, 1997.

SPSS Base 8.0 Application Guide, SPSS Inc., Chicago, Illinois, 1998.

Wirahadikusumah, Reini, Abraham, Dulcy M., Iseley, Tom and Prasanth, Ravi K. *Asseement Technologies for Sewer Rehabilitation*, Automation in Construction, 1998, P259-P270

Zar, J.H., *Biostatistical Analysis,* Second Edition, Prentice-Hall Inc., New Jersey, 1996.

Zar, J.H., *Biostatistical Analysis,* Third Edition, Prentice-Hall Inc., New Jersey, 1984.

Najafi, M, and Iseley T, *Evaluation of Advancements in Microtunneling*, Constuction Congress V: Management Engineered Constrution in Expanding Global Markets, Minneapolis, Minnesota, Oct. 4-8, 1997

APPENDIX A

POSSIBLE PIPE COMBINATIONS

No.	Age	Depth(m)	Diameter	Material	Waste	Status	Count	StrVariable
1	0-29	0-6	150-375	CP	SAN	1	0	0
2	0-29	0-6	150-375	CP	STM	1	11	1
3	0-29	0-6	150-375	CP	CMB	1	0	0
4	0-29	0-6	150-375	TP	SAN	1	22	1
5	0-29	0-6	150-375	TP	STM	1	0	0
6	0-29	0-6	150-375	TP	CMB	1	4	1
7	0-29	0-6	150-375	RCP	SAN	1	0	0
8	0-29	0-6	150-375	RCP	STM	1	0	0
9	0-29	0-6	150-375	RCP	CMB	1	0	0
10	0-29	0-6	150-375	NC	SAN	1	0	0
11	0-29	0-6	150-375	NC	STM	1	0	0
12	0-29	0-6	150-375	NC	CMB	1	0	0
13	0-29	0-6	150-375	PVC	SAN	1	0	0
14	0-29	0-6	150-375	PVC	STM	1	0	0
15	0-29	0-6	150-375	PVC	CMB	1	0	0
16	0-29	0-6	450-525	CP	SAN	1	0	0
17	0-29	0-6	450-525	CP	STM	1	2	1
18	0-29	0-6	450-525	CP	CMB	1	0	0
19	0-29	0-6	450-525	TP	SAN	1	0	0
20	0-29	0-6	450-525	TP	STM	1	0	0
21	0-29	0-6	450-525	TP	CMB	1	0	0
22	0-29	0-6	450-525	RCP	SAN	1	0	0
23	0-29	0-6	450-525	RCP	STM	1	2	1
24	0-29	0-6	450-525	RCP	CMB	1	0	0
25	0-29	0-6	450-525	NC	SAN	1	0	0
26	0-29	0-6	450-525	NC	STM	1	0	0
27	0-29	0-6	450-525	NC	CMB	1	0	0
28	0-29	0-6	450-525	PVC	SAN	1	0	0
29	0-29	0-6	450-525	PVC	STM	1	0	0
30	0-29	0-6	450-525	PVC	CMB	1	0	0
31	0-29	0-6	550-1050	CP	SAN	1	0	0
32	0-29	0-6	550-1050	CP	STM	1	0	0
33	0-29	0-6	550-1050	CP	CMB	1	0	0
34	0-29	0-6	550-1050	TP	SAN	1	1	1
35	0-29	0-6	550-1050	TP	STM	1	0	0
36	0-29	0-6	550-1050	TP	CMB	1	1	1
37	0-29	0-6	550-1050	RCP	SAN	1	0	0
38	0-29	0-6	550-1050	RCP	STM	1	2	1
39	0-29	0-6	550-1050	RCP	CMB	1	0	0
40	0-29	0-6	550-1050	NC	SAN	1	0	0
41	0-29	0-6	550-1050	NC	STM	1	0	0
42	0-29	0-6	550-1050	NC	CMB	1	0	0
43	0-29	0-6	550-1050	PVC	SAN	1	0	0
44	0-29	0-6	550-1050	PVC	STM	1	0	0
45	0-29	0-6	550-1050	PVC	CMB	1	0	0
46	0-29	6-8+	150-375	CP	SAN	1	0	0
47	0-29	6-8+	150-375	CP	STM	1	0	0
48	0-29	6-8+	150-375	CP	CMB	1	0	0
49	0-29	6-8+	150-375	TP	SAN	1	26	1
50	0-29	6-8+	150-375	TP	STM	1	0	0

No.	Age	Depth(m)	Diameter	Material	Waste	Status	Count	StrVariable
51	0-29	6-8+	150-375	TP	CMB	1	1	1
52	0-29	6-8+	150-375	RCP	SAN	1	1	1
53	0-29	6-8+	150-375	RCP	STM	1	0	0
54	0-29	6-8+	150-375	RCP	CMB	1	0	0
55	0-29	6-8+	150-375	NC	SAN	1	0	0
56	0-29	6-8+	150-375	NC	STM	1	0	0
57	0-29	6-8+	150-375	NC	CMB	1	0	0
58	0-29	6-8+	150-375	PVC	SAN	1	0	0
59	0-29	6-8+	150-375	PVC	STM	1	0	0
60	0-29	6-8+	150-375	PVC	CMB	1	0	0
61	0-29	6-8+	450-525	CP	SAN	1	0	0
62	0-29	6-8+	450-525	CP	STM	1	0	0
63	0-29	6-8+	450-525	CP	CMB	1	0	0
64	0-29	6-8+	450-525	TP	SAN	1	0	0
65	0-29	6-8+	450-525	TP	STM	1	0	0
66	0-29	6-8+	450-525	TP	CMB	1	0	0
67	0-29	6-8+	450-525	RCP	SAN	1	0	0
68	0-29	6-8+	450-525	RCP	STM	1	0	0
69	0-29	6-8+	450-525	RCP	CMB	1	0	0
70	0-29	6-8+	450-525	NC	SAN	1	0	0
71	0-29	6-8+	450-525	NC	STM	1	0	0
72	0-29	6-8+	450-525	NC	CMB	1	0	0
73	0-29	6-8+	450-525	PVC	SAN	1	1	1
74	0-29	6-8+	450-525	PVC	STM	1	0	0
75	0-29	6-8+	450-525	PVC	CMB	1	0	0
76	0-29	6-8+	550-1050	CP	SAN	1	0	0
77	0-29	6-8+	550-1050	CP	STM	1	0	0
78	0-29	6-8+	550-1050	CP	CMB	1	0	0
79	0-29	6-8+	550-1050	TP	SAN	1	0	0
80	0-29	6-8+	550-1050	TP	STM	1	0	0
81	0-29	6-8+	550-1050	TP	CMB	1	0	0
82	0-29	6-8+	550-1050	RCP	SAN	1	0	0
83	0-29	6-8+	550-1050	RCP	STM	1	0	0
84	0-29	6-8+	550-1050	RCP	CMB	1	0	0
85	0-29	6-8+	550-1050	NC	SAN	1	0	0
86	0-29	6-8+	550-1050	NC	STM	1	0	0
87	0-29	6-8+	550-1050	NC	CMB	1	0	0
88	0-29	6-8+	550-1050	PVC	SAN	1	0	0
89	0-29	6-8+	550-1050	PVC	STM	1	0	0
90	0-29	6-8+	550-1050	PVC	CMB	1	0	0
91	30-59	0-6	150-375	CP	SAN	1	0	0
92	30-59	0-6	150-375	CP	STM	1	10	1
93	30-59	0-6	150-375	CP	CMB	1	10	1
94	30-59	0-6	150-375	TP	SAN	1	4	1
95	30-59	0-6	150-375	TP	STM	1	6	1
96	30-59	0-6	150-375	TP	CMB	1	25	1
97	30-59	0-6	150-375	RCP	SAN	1	0	0
98	30-59	0-6	150-375	RCP	STM	1	0	0
99	30-59	0-6	150-375	RCP	CMB	1	0	0
100	30-59	0-6	150-375	NC	SAN	1	0	0
101	30-59	0-6	150-375	NC	STM	1	0	0
102	30-59	0-6	150-375	NC	CMB	1	0	0
103	30-59	0-6	150-375	PVC	SAN	1	0	0
104	30-59	0-6	150-375	PVC	STM	1	0	0
105	30-59	0-6	150-375	PVC	CMB	1	0	0

No.	Age	Depth(m)	Diameter	Material	Waste	Status	Count	StrVariable
106	30-59	0-6	450-525	CP	SAN	1	0	0
107	30-59	0-6	450-525	CP	STM	1	0	0
108	30-59	0-6	450-525	CP	CMB	1	1	1
109	30-59	0-6	450-525	TP	SAN	1	0	0
110	30-59	0-6	450-525	TP	STM	1	0	0
111	30-59	0-6	450-525	TP	CMB	1	3	1
112	30-59	0-6	450-525	RCP	SAN	1	0	0
113	30-59	0-6	450-525	RCP	STM	1	0	0
114	30-59	0-6	450-525	RCP	CMB	1	1	1
115	30-59	0-6	450-525	NC	SAN	1	0	0
116	30-59	0-6	450-525	NC	STM	1	0	0
117	30-59	0-6	450-525	NC	CMB	1	0	0
118	30-59	0-6	450-525	PVC	SAN	1	0	0
119	30-59	0-6	450-525	PVC	STM	1	0	0
120	30-59	0-6	450-525	PVC	CMB	1	0	0
121	30-59	0-6	550-1050	CP	SAN	1	0	0
122	30-59	0-6	550-1050	CP	STM	1	0	0
123	30-59	0-6	550-1050	CP	CMB	1	0	0
124	30-59	0-6	550-1050	TP	SAN	1	0	0
125	30-59	0-6	550-1050	TP	STM	1	0	0
126	30-59	0-6	550-1050	TP	CMB	1	1	1
127	30-59	0-6	550-1050	RCP	SAN	1	0	0
128	30-59	0-6	550-1050	RCP	STM	1	0	0
129	30-59	0-6	550-1050	RCP	CMB	1	0	0
130	30-59	0-6	550-1050	NC	SAN	1	0	0
131	30-59	0-6	550-1050	NC	STM	1	0	0
132	30-59	0-6	550-1050	NC	CMB	1	0	0
133	30-59	0-6	550-1050	PVC	SAN	1	0	0
134	30-59	0-6	550-1050	PVC	STM	1	0	0
135	30-59	0-6	550-1050	PVC	CMB	1	0	0
136	30-59	6-8+	150-375	CP	SAN	1	0	0
137	30-59	6-8+	150-375	CP	STM	1	0	0
138	30-59	6-8+	150-375	CP	CMB	1	0	0
139	30-59	6-8+	150-375	TP	SAN	1	0	0
140	30-59	6-8+	150-375	TP	STM	1	0	0
141	30-59	6-8+	150-375	TP	CMB	1	0	0
142	30-59	6-8+	150-375	RCP	SAN	1	0	0
143	30-59	6-8+	150-375	RCP	STM	1	0	0
144	30-59	6-8+	150-375	RCP	CMB	1	0	0
145	30-59	6-8+	150-375	NC	SAN	1	0	0
146	30-59	6-8+	150-375	NC	STM	1	0	0
147	30-59	6-8+	150-375	NC	CMB	1	0	0
148	30-59	6-8+	150-375	PVC	SAN	1	0	0
149	30-59	6-8+	150-375	PVC	STM	1	0	0
150	30-59	6-8+	150-375	PVC	CMB	1	0	0
151	30-59	6-8+	450-525	CP	SAN	1	0	0
152	30-59	6-8+	450-525	CP	STM	1	0	0
153	30-59	6-8+	450-525	CP	CMB	1	0	0
154	30-59	6-8+	450-525	TP	SAN	1	0	0
155	30-59	6-8+	450-525	TP	STM	1	0	0
156	30-59	6-8+	450-525	TP	CMB	1	0	0
157	30-59	6-8+	450-525	RCP	SAN	1	0	0
158	30-59	6-8+	450-525	RCP	STM	1	0	0
159	30-59	6-8+	450-525	RCP	CMB	1	0	0
160	30-59	6-8+	450-525	NC	SAN	1	0	0

No.	Age	Depth(m)	Diameter	Material	Waste	Status	Count	StrVariable
161	30-59	6-8+	450-525	NC	STM	1	0	0
162	30-59	6-8+	450-525	NC	CMB	1	0	0
163	30-59	6-8+	450-525	PVC	SAN	1	0	0
164	30-59	6-8+	450-525	PVC	STM	1	0	0
165	30-59	6-8+	450-525	PVC	CMB	1	0	0
166	30-59	6-8+	550-1050	CP	SAN	1	0	0
167	30-59	6-8+	550-1050	CP	STM	1	0	0
168	30-59	6-8+	550-1050	CP	CMB	1	1	1
169	30-59	6-8+	550-1050	TP	SAN	1	0	0
170	30-59	6-8+	550-1050	TP	STM	1	0	0
171	30-59	6-8+	550-1050	TP	CMB	1	0	0
172	30-59	6-8+	550-1050	RCP	SAN	1	0	0
173	30-59	6-8+	550-1050	RCP	STM	1	0	0
174	30-59	6-8+	550-1050	RCP	CMB	1	0	0
175	30-59	6-8+	550-1050	NC	SAN	1	0	0
176	30-59	6-8+	550-1050	NC	STM	1	0	0
177	30-59	6-8+	550-1050	NC	CMB	1	0	0
178	30-59	6-8+	550-1050	PVC	SAN	1	0	0
179	30-59	6-8+	550-1050	PVC	STM	1	0	0
180	30-59	6-8+	550-1050	PVC	CMB	1	0	0
181	60-90+	0-6	150-375	CP	SAN	1	0	0
182	60-90+	0-6	150-375	CP	STM	1	0	0
183	60-90+	0-6	150-375	CP	CMB	1	14	1
184	60-90+	0-6	150-375	TP	SAN	1	0	0
185	60-90+	0-6	150-375	TP	STM	1	0	0
186	60-90+	0-6	150-375	TP	CMB	1	159	1
187	60-90+	0-6	150-375	RCP	SAN	1	0	0
188	60-90+	0-6	150-375	RCP	STM	1	0	0
189	60-90+	0-6	150-375	RCP	CMB	1	0	0
190	60-90+	0-6	150-375	NC	SAN	1	0	0
191	60-90+	0-6	150-375	NC	STM	1	0	0
192	60-90+	0-6	150-375	NC	CMB	1	24	1
193	60-90+	0-6	150-375	PVC	SAN	1	0	0
194	60-90+	0-6	150-375	PVC	STM	1	0	0
195	60-90+	0-6	150-375	PVC	CMB	1	1	1
196	60-90+	0-6	450-525	CP	SAN	1	0	0
197	60-90+	0-6	450-525	CP	STM	1	0	0
198	60-90+	0-6	450-525	CP	CMB	1	1	1
199	60-90+	0-6	450-525	TP	SAN	1	0	0
200	60-90+	0-6	450-525	TP	STM	1	0	0
201	60-90+	0-6	450-525	TP	CMB	1	7	1
202	60-90+	0-6	450-525	RCP	SAN	1	0	0
203	60-90+	0-6	450-525	RCP	STM	1	0	0
204	60-90+	0-6	450-525	RCP	CMB	1	0	0
205	60-90+	0-6	450-525	NC	SAN	1	0	0
206	60-90+	0-6	450-525	NC	STM	1	0	0
207	60-90+	0-6	450-525	NC	CMB	1	0	0
208	60-90+	0-6	450-525	PVC	SAN	1	0	0
209	60-90+	0-6	450-525	PVC	STM	1	0	0
210	60-90+	0-6	450-525	PVC	CMB	1	0	0
211	60-90+	0-6	550-1050	CP	SAN	1	0	0
212	60-90+	0-6	550-1050	CP	STM	1	0	0
213	60-90+	0-6	550-1050	CP	CMB	1	0	0
214	60-90+	0-6	550-1050	TP	SAN	1	0	0
215	60-90+	0-6	550-1050	TP	STM	1	0	0

No.	Age	Depth(m)	Diameter	Material	Waste	Status	Count	StrVariable
216	60-90+	0-6	550-1050	TP	CMB	1	1	1
217	60-90+	0-6	550-1050	RCP	SAN	1	0	0
218	60-90+	0-6	550-1050	RCP	STM	1	0	0
219	60-90+	0-6	550-1050	RCP	CMB	1	4	1
220	60-90+	0-6	550-1050	NC	SAN	1	0	0
221	60-90+	0-6	550-1050	NC	STM	1	0	0
222	60-90+	0-6	550-1050	NC	CMB	1	0	0
223	60-90+	0-6	550-1050	PVC	SAN	1	0	0
224	60-90+	0-6	550-1050	PVC	STM	1	0	0
225	60-90+	0-6	550-1050	PVC	CMB	1	0	0
226	60-90+	6-8+	150-375	CP	SAN	1	0	0
227	60-90+	6-8+	150-375	CP	STM	1	0	0
228	60-90+	6-8+	150-375	CP	CMB	1	0	0
229	60-90+	6-8+	150-375	TP	SAN	1	0	0
230	60-90+	6-8+	150-375	TP	STM	1	0	0
231	60-90+	6-8+	150-375	TP	CMB	1	0	0
232	60-90+	6-8+	150-375	RCP	SAN	1	0	0
233	60-90+	6-8+	150-375	RCP	STM	1	0	0
234	60-90+	6-8+	150-375	RCP	CMB	1	0	0
235	60-90+	6-8+	150-375	NC	SAN	1	0	0
236	60-90+	6-8+	150-375	NC	STM	1	0	0
237	60-90+	6-8+	150-375	NC	CMB	1	0	0
238	60-90+	6-8+	150-375	PVC	SAN	1	0	0
239	60-90+	6-8+	150-375	PVC	STM	1	0	0
240	60-90+	6-8+	150-375	PVC	CMB	1	0	0
241	60-90+	6-8+	450-525	CP	SAN	1	0	0
242	60-90+	6-8+	450-525	CP	STM	1	0	0
243	60-90+	6-8+	450-525	CP	CMB	1	0	0
244	60-90+	6-8+	450-525	TP	SAN	1	0	0
245	60-90+	6-8+	450-525	TP	STM	1	0	0
246	60-90+	6-8+	450-525	TP	CMB	1	0	0
247	60-90+	6-8+	450-525	RCP	SAN	1	0	0
248	60-90+	6-8+	450-525	RCP	STM	1	0	0
249	60-90+	6-8+	450-525	RCP	CMB	1	0	0
250	60-90+	6-8+	450-525	NC	SAN	1	0	0
251	60-90+	6-8+	450-525	NC	STM	1	0	0
252	60-90+	6-8+	450-525	NC	CMB	1	0	0
253	60-90+	6-8+	450-525	PVC	SAN	1	0	0
254	60-90+	6-8+	450-525	PVC	STM	1	0	0
255	60-90+	6-8+	450-525	PVC	CMB	1	0	0
256	60-90+	6-8+	550-1050	CP	SAN	1	0	0
257	60-90+	6-8+	550-1050	CP	STM	1	1	1
258	60-90+	6-8+	550-1050	CP	CMB	1	0	0
259	60-90+	6-8+	550-1050	TP	SAN	1	0	0
260	60-90+	6-8+	550-1050	TP	STM	1	0	0
261	60-90+	6-8+	550-1050	TP	CMB	1	0	0
262	60-90+	6-8+	550-1050	RCP	SAN	1	0	0
263	60-90+	6-8+	550-1050	RCP	STM	1	0	0
264	60-90+	6-8+	550-1050	RCP	CMB	1	1	1
265	60-90+	6-8+	550-1050	NC	SAN	1	0	0
266	60-90+	6-8+	550-1050	NC	STM	1	0	0
267	60-90+	6-8+	550-1050	NC	CMB	1	0	0
268	60-90+	6-8+	550-1050	PVC	SAN	1	0	0
269	60-90+	6-8+	550-1050	PVC	STM	1	0	0
270	60-90+	6-8+	550-1050	PVC	CMB	1	0	0

No.	Age	Depth(m)	Diameter	Material	Waste	Status	Count	StrVariable
271	0-29	0-6	150-375	CP	SAN	2	0	0
272	0-29	0-6	150-375	CP	STM	2	24	1
273	0-29	0-6	150-375	CP	CMB	2	14	1
274	0-29	0-6	150-375	TP	SAN	2	15	1
275	0-29	0-6	150-375	TP	STM	2	0	0
276	0-29	0-6	150-375	TP	CMB	2	3	1
277	0-29	0-6	150-375	RCP	SAN	2	0	0
278	0-29	0-6	150-375	RCP	STM	2	4	1
279	0-29	0-6	150-375	RCP	CMB	2	3	1
280	0-29	0-6	150-375	NC	SAN	2	0	0
281	0-29	0-6	150-375	NC	STM	2	0	0
282	0-29	0-6	150-375	NC	CMB	2	0	0
283	0-29	0-6	150-375	PVC	SAN	2	0	0
284	0-29	0-6	150-375	PVC	STM	2	0	0
285	0-29	0-6	150-375	PVC	CMB	2	0	0
286	0-29	0-6	450-525	CP	SAN	2	0	0
287	0-29	0-6	450-525	CP	STM	2	10	1
288	0-29	0-6	450-525	CP	CMB	2	0	0
289	0-29	0-6	450-525	TP	SAN	2	0	0
290	0-29	0-6	450-525	TP	STM	2	0	0
291	0-29	0-6	450-525	TP	CMB	2	3	1
292	0-29	0-6	450-525	RCP	SAN	2	0	0
293	0-29	0-6	450-525	RCP	STM	2	8	1
294	0-29	0-6	450-525	RCP	CMB	2	0	0
295	0-29	0-6	450-525	NC	SAN	2	0	0
296	0-29	0-6	450-525	NC	STM	2	0	0
297	0-29	0-6	450-525	NC	CMB	2	0	0
298	0-29	0-6	450-525	PVC	SAN	2	0	0
299	0-29	0-6	450-525	PVC	STM	2	0	0
300	0-29	0-6	450-525	PVC	CMB	2	0	0
301	0-29	0-6	550-1050	CP	SAN	2	0	0
302	0-29	0-6	550-1050	CP	STM	2	2	1
303	0-29	0-6	550-1050	CP	CMB	2	0	0
304	0-29	0-6	550-1050	TP	SAN	2	0	0
305	0-29	0-6	550-1050	TP	STM	2	0	0
306	0-29	0-6	550-1050	TP	CMB	2	2	1
307	0-29	0-6	550-1050	RCP	SAN	2	1	1
308	0-29	0-6	550-1050	RCP	STM	2	36	1
309	0-29	0-6	550-1050	RCP	CMB	2	10	1
310	0-29	0-6	550-1050	NC	SAN	2	0	0
311	0-29	0-6	550-1050	NC	STM	2	0	0
312	0-29	0-6	550-1050	NC	CMB	2	0	0
313	0-29	0-6	550-1050	PVC	SAN	2	0	0
314	0-29	0-6	550-1050	PVC	STM	2	0	0
315	0-29	0-6	550-1050	PVC	CMB	2	0	0
316	0-29	6-8+	150-375	CP	SAN	2	0	0
317	0-29	6-8+	150-375	CP	STM	2	0	0
318	0-29	6-8+	150-375	CP	CMB	2	1	1
319	0-29	6-8+	150-375	TP	SAN	2	12	1
320	0-29	6-8+	150-375	TP	STM	2	0	0
321	0-29	6-8+	150-375	TP	CMB	2	1	1
322	0-29	6-8+	150-375	RCP	SAN	2	2	1
323	0-29	6-8+	150-375	RCP	STM	2	0	0
324	0-29	6-8+	150-375	RCP	CMB	2	0	0
325	0-29	6-8+	150-375	NC	SAN	2	0	0

No.	Age	Depth(m)	Diameter	Material	Waste	Status	Count	StrVariable
326	0-29	6-8+	150-375	NC	STM	2	0	0
327	0-29	6-8+	150-375	NC	CMB	2	0	0
328	0-29	6-8+	150-375	PVC	SAN	2	3	1
329	0-29	6-8+	150-375	PVC	STM	2	0	0
330	0-29	6-8+	150-375	PVC	CMB	2	0	0
331	0-29	6-8+	450-525	CP	SAN	2	0	0
332	0-29	6-8+	450-525	CP	STM	2	0	0
333	0-29	6-8+	450-525	CP	CMB	2	0	0
334	0-29	6-8+	450-525	TP	SAN	2	0	0
335	0-29	6-8+	450-525	TP	STM	2	0	0
336	0-29	6-8+	450-525	TP	CMB	2	0	0
337	0-29	6-8+	450-525	RCP	SAN	2	1	1
338	0-29	6-8+	450-525	RCP	STM	2	0	0
339	0-29	6-8+	450-525	RCP	CMB	2	0	0
340	0-29	6-8+	450-525	NC	SAN	2	0	0
341	0-29	6-8+	450-525	NC	STM	2	0	0
342	0-29	6-8+	450-525	NC	CMB	2	0	0
343	0-29	6-8+	450-525	PVC	SAN	2	0	0
344	0-29	6-8+	450-525	PVC	STM	2	0	0
345	0-29	6-8+	450-525	PVC	CMB	2	0	0
346	0-29	6-8+	550-1050	CP	SAN	2	0	0
347	0-29	6-8+	550-1050	CP	STM	2	0	0
348	0-29	6-8+	550-1050	CP	CMB	2	1	1
349	0-29	6-8+	550-1050	TP	SAN	2	0	0
350	0-29	6-8+	550-1050	TP	STM	2	0	0
351	0-29	6-8+	550-1050	TP	CMB	2	0	0
352	0-29	6-8+	550-1050	RCP	SAN	2	1	1
353	0-29	6-8+	550-1050	RCP	STM	2	11	1
354	0-29	6-8+	550-1050	RCP	CMB	2	1	1
355	0-29	6-8+	550-1050	NC	SAN	2	0	0
356	0-29	6-8+	550-1050	NC	STM	2	0	0
357	0-29	6-8+	550-1050	NC	CMB	2	0	0
358	0-29	6-8+	550-1050	PVC	SAN	2	0	0
359	0-29	6-8+	550-1050	PVC	STM	2	0	0
360	0-29	6-8+	550-1050	PVC	CMB	2	0	0
361	30-59	0-6	150-375	CP	SAN	2	0	0
362	30-59	0-6	150-375	CP	STM	2	7	1
363	30-59	0-6	150-375	CP	CMB	2	24	1
364	30-59	0-6	150-375	TP	SAN	2	10	1
365	30-59	0-6	150-375	TP	STM	2	5	1
366	30-59	0-6	150-375	TP	CMB	2	32	1
367	30-59	0-6	150-375	RCP	SAN	2	0	0
368	30-59	0-6	150-375	RCP	STM	2	0	0
369	30-59	0-6	150-375	RCP	CMB	2	0	0
370	30-59	0-6	150-375	NC	SAN	2	0	0
371	30-59	0-6	150-375	NC	STM	2	0	0
372	30-59	0-6	150-375	NC	CMB	2	1	1
373	30-59	0-6	150-375	PVC	SAN	2	0	0
374	30-59	0-6	150-375	PVC	STM	2	0	0
375	30-59	0-6	150-375	PVC	CMB	2	0	0
376	30-59	0-6	450-525	CP	SAN	2	0	0
377	30-59	0-6	450-525	CP	STM	2	4	1
378	30-59	0-6	450-525	CP	CMB	2	2	1
379	30-59	0-6	450-525	TP	SAN	2	0	0
380	30-59	0-6	450-525	TP	STM	2	0	0

No.	Age	Depth(m)	Diameter	Material	Waste	Status	Count	StrVariable
381	30-59	0-6	450-525	TP	CMB	2	11	1
382	30-59	0-6	450-525	RCP	SAN	2	0	0
383	30-59	0-6	450-525	RCP	STM	2	1	1
384	30-59	0-6	450-525	RCP	CMB	2	1	1
385	30-59	0-6	450-525	NC	SAN	2	0	0
386	30-59	0-6	450-525	NC	STM	2	0	0
387	30-59	0-6	450-525	NC	CMB	2	0	0
388	30-59	0-6	450-525	PVC	SAN	2	0	0
389	30-59	0-6	450-525	PVC	STM	2	0	0
390	30-59	0-6	450-525	PVC	CMB	2	0	0
391	30-59	0-6	550-1050	CP	SAN	2	0	0
392	30-59	0-6	550-1050	CP	STM	2	0	0
393	30-59	0-6	550-1050	CP	CMB	2	6	1
394	30-59	0-6	550-1050	TP	SAN	2	0	0
395	30-59	0-6	550-1050	TP	STM	2	0	0
396	30-59	0-6	550-1050	TP	CMB	2	6	1
397	30-59	0-6	550-1050	RCP	SAN	2	0	0
398	30-59	0-6	550-1050	RCP	STM	2	0	0
399	30-59	0-6	550-1050	RCP	CMB	2	7	1
400	30-59	0-6	550-1050	NC	SAN	2	0	0
401	30-59	0-6	550-1050	NC	STM	2	0	0
402	30-59	0-6	550-1050	NC	CMB	2	1	1
403	30-59	0-6	550-1050	PVC	SAN	2	0	0
404	30-59	0-6	550-1050	PVC	STM	2	0	0
405	30-59	0-6	550-1050	PVC	CMB	2	0	0
406	30-59	6-8+	150-375	CP	SAN	2	0	0
407	30-59	6-8+	150-375	CP	STM	2	0	0
408	30-59	6-8+	150-375	CP	CMB	2	3	1
409	30-59	6-8+	150-375	TP	SAN	2	0	0
410	30-59	6-8+	150-375	TP	STM	2	0	0
411	30-59	6-8+	150-375	TP	CMB	2	0	0
412	30-59	6-8+	150-375	RCP	SAN	2	0	0
413	30-59	6-8+	150-375	RCP	STM	2	0	0
414	30-59	6-8+	150-375	RCP	CMB	2	0	0
415	30-59	6-8+	150-375	NC	SAN	2	0	0
416	30-59	6-8+	150-375	NC	STM	2	0	0
417	30-59	6-8+	150-375	NC	CMB	2	0	0
418	30-59	6-8+	150-375	PVC	SAN	2	0	0
419	30-59	6-8+	150-375	PVC	STM	2	0	0
420	30-59	6-8+	150-375	PVC	CMB	2	0	0
421	30-59	6-8+	450-525	CP	SAN	2	0	0
422	30-59	6-8+	450-525	CP	STM	2	0	0
423	30-59	6-8+	450-525	CP	CMB	2	0	0
424	30-59	6-8+	450-525	TP	SAN	2	0	0
425	30-59	6-8+	450-525	TP	STM	2	0	0
426	30-59	6-8+	450-525	TP	CMB	2	1	1
427	30-59	6-8+	450-525	RCP	SAN	2	0	0
428	30-59	6-8+	450-525	RCP	STM	2	0	0
429	30-59	6-8+	450-525	RCP	CMB	2	0	0
430	30-59	6-8+	450-525	NC	SAN	2	0	0
431	30-59	6-8+	450-525	NC	STM	2	0	0
432	30-59	6-8+	450-525	NC	CMB	2	0	0
433	30-59	6-8+	450-525	PVC	SAN	2	0	0
434	30-59	6-8+	450-525	PVC	STM	2	0	0
435	30-59	6-8+	450-525	PVC	CMB	2	0	0

No.	Age	Depth(m)	Diameter	Material	Waste	Status	Count	StrVariable
436	30-59	6-8+	550-1050	CP	SAN	2	0	0
437	30-59	6-8+	550-1050	CP	STM	2	0	0
438	30-59	6-8+	550-1050	CP	CMB	2	0	0
439	30-59	6-8+	550-1050	TP	SAN	2	0	0
440	30-59	6-8+	550-1050	TP	STM	2	0	0
441	30-59	6-8+	550-1050	TP	CMB	2	1	1
442	30-59	6-8+	550-1050	RCP	SAN	2	0	0
443	30-59	6-8+	550-1050	RCP	STM	2	1	1
444	30-59	6-8+	550-1050	RCP	CMB	2	0	0
445	30-59	6-8+	550-1050	NC	SAN	2	0	0
446	30-59	6-8+	550-1050	NC	STM	2	0	0
447	30-59	6-8+	550-1050	NC	CMB	2	0	0
448	30-59	6-8+	550-1050	PVC	SAN	2	0	0
449	30-59	6-8+	550-1050	PVC	STM	2	0	0
450	30-59	6-8+	550-1050	PVC	CMB	2	0	0
451	60-90+	0-6	150-375	CP	SAN	2	0	0
452	60-90+	0-6	150-375	CP	STM	2	0	0
453	60-90+	0-6	150-375	CP	CMB	2	3	1
454	60-90+	0-6	150-375	TP	SAN	2	0	0
455	60-90+	0-6	150-375	TP	STM	2	0	0
456	60-90+	0-6	150-375	TP	CMB	2	89	1
457	60-90+	0-6	150-375	RCP	SAN	2	0	0
458	60-90+	0-6	150-375	RCP	STM	2	0	0
459	60-90+	0-6	150-375	RCP	CMB	2	0	0
460	60-90+	0-6	150-375	NC	SAN	2	0	0
461	60-90+	0-6	150-375	NC	STM	2	0	0
462	60-90+	0-6	150-375	NC	CMB	2	13	1
463	60-90+	0-6	150-375	PVC	SAN	2	0	0
464	60-90+	0-6	150-375	PVC	STM	2	0	0
465	60-90+	0-6	150-375	PVC	CMB	2	1	1
466	60-90+	0-6	450-525	CP	SAN	2	0	0
467	60-90+	0-6	450-525	CP	STM	2	0	0
468	60-90+	0-6	450-525	CP	CMB	2	0	0
469	60-90+	0-6	450-525	TP	SAN	2	0	0
470	60-90+	0-6	450-525	TP	STM	2	0	0
471	60-90+	0-6	450-525	TP	CMB	2	15	1
472	60-90+	0-6	450-525	RCP	SAN	2	0	0
473	60-90+	0-6	450-525	RCP	STM	2	0	0
474	60-90+	0-6	450-525	RCP	CMB	2	0	0
475	60-90+	0-6	450-525	NC	SAN	2	0	0
476	60-90+	0-6	450-525	NC	STM	2	0	0
477	60-90+	0-6	450-525	NC	CMB	2	0	0
478	60-90+	0-6	450-525	PVC	SAN	2	0	0
479	60-90+	0-6	450-525	PVC	STM	2	0	0
480	60-90+	0-6	450-525	PVC	CMB	2	0	0
481	60-90+	0-6	550-1050	CP	SAN	2	0	0
482	60-90+	0-6	550-1050	CP	STM	2	0	0
483	60-90+	0-6	550-1050	CP	CMB	2	5	1
484	60-90+	0-6	550-1050	TP	SAN	2	0	0
485	60-90+	0-6	550-1050	TP	STM	2	0	0
486	60-90+	0-6	550-1050	TP	CMB	2	5	1
487	60-90+	0-6	550-1050	RCP	SAN	2	0	0
488	60-90+	0-6	550-1050	RCP	STM	2	0	0
489	60-90+	0-6	550-1050	RCP	CMB	2	5	1
490	60-90+	0-6	550-1050	NC	SAN	2	0	0

No.	Age	Depth(m)	Diameter	Material	Waste	Status	Count	StrVariable
491	60-90+	0-6	550-1050	NC	STM	2	0	0
492	60-90+	0-6	550-1050	NC	CMB	2	3	1
493	60-90+	0-6	550-1050	PVC	SAN	2	0	0
494	60-90+	0-6	550-1050	PVC	STM	2	0	0
495	60-90+	0-6	550-1050	PVC	CMB	2	0	0
496	60-90+	6-8+	150-375	CP	SAN	2	0	0
497	60-90+	6-8+	150-375	CP	STM	2	0	0
498	60-90+	6-8+	150-375	CP	CMB	2	0	0
499	60-90+	6-8+	150-375	TP	SAN	2	0	0
500	60-90+	6-8+	150-375	TP	STM	2	0	0
501	60-90+	6-8+	150-375	TP	CMB	2	0	0
502	60-90+	6-8+	150-375	RCP	SAN	2	0	0
503	60-90+	6-8+	150-375	RCP	STM	2	0	0
504	60-90+	6-8+	150-375	RCP	CMB	2	0	0
505	60-90+	6-8+	150-375	NC	SAN	2	0	0
506	60-90+	6-8+	150-375	NC	STM	2	0	0
507	60-90+	6-8+	150-375	NC	CMB	2	0	0
508	60-90+	6-8+	150-375	PVC	SAN	2	0	0
509	60-90+	6-8+	150-375	PVC	STM	2	0	0
510	60-90+	6-8+	150-375	PVC	CMB	2	0	0
511	60-90+	6-8+	450-525	CP	SAN	2	0	0
512	60-90+	6-8+	450-525	CP	STM	2	0	0
513	60-90+	6-8+	450-525	CP	CMB	2	0	0
514	60-90+	6-8+	450-525	TP	SAN	2	0	0
515	60-90+	6-8+	450-525	TP	STM	2	0	0
516	60-90+	6-8+	450-525	TP	CMB	2	2	1
517	60-90+	6-8+	450-525	RCP	SAN	2	0	0
518	60-90+	6-8+	450-525	RCP	STM	2	0	0
519	60-90+	6-8+	450-525	RCP	CMB	2	0	0
520	60-90+	6-8+	450-525	NC	SAN	2	0	0
521	60-90+	6-8+	450-525	NC	STM	2	0	0
522	60-90+	6-8+	450-525	NC	CMB	2	0	0
523	60-90+	6-8+	450-525	PVC	SAN	2	0	0
524	60-90+	6-8+	450-525	PVC	STM	2	0	0
525	60-90+	6-8+	450-525	PVC	CMB	2	0	0
526	60-90+	6-8+	550-1050	CP	SAN	2	0	0
527	60-90+	6-8+	550-1050	CP	STM	2	0	0
528	60-90+	6-8+	550-1050	CP	CMB	2	1	1
529	60-90+	6-8+	550-1050	TP	SAN	2	0	0
530	60-90+	6-8+	550-1050	TP	STM	2	0	0
531	60-90+	6-8+	550-1050	TP	CMB	2	0	0
532	60-90+	6-8+	550-1050	RCP	SAN	2	0	0
533	60-90+	6-8+	550-1050	RCP	STM	2	0	0
534	60-90+	6-8+	550-1050	RCP	CMB	2	0	0
535	60-90+	6-8+	550-1050	NC	SAN	2	0	0
536	60-90+	6-8+	550-1050	NC	STM	2	0	0
537	60-90+	6-8+	550-1050	NC	CMB	2	0	0
538	60-90+	6-8+	550-1050	PVC	SAN	2	0	0
539	60-90+	6-8+	550-1050	PVC	STM	2	0	0
540	60-90+	6-8+	550-1050	PVC	CMB	2	0	0

APPENDIX B

PIPE MATERIAL CODES

Code	Description
ABS	Acrylonitrile Butadiene Styrene
ACP	Asbestos Cement Pipe
BRK	Brick
CBL	Concrete Block Pipe
CIP	Cast Iron Pipe
CMP	Corrugated Metal Pipe
CON	Poured-In-Place Concrete
CP	Non-Reinforced Concrete Pipe
CPP	Cured-In-Place Pipe
DB	Double Barrel Pipe
DIP	Ductile Iron Pipe
EYE	Eye Pipe
FRP	Fiberglas Reinforced Pipe
NC	Non-Corrode Pipe
ORG	Orangeberg
OVL	Oval Pipe
PEP	Polyethylene Pipe
PLP	Plastic Lined Pipe
PMP	Perferated Metal Pipe
PVC	Polyvinylchloride Pipe
RCP	Reinforced Concrete Pipe
RPM	Reinforced Plastic Mortar Pipe
STP	Steel Pipe
TP	Clay Tile Pipe
VCP	Vitrified Clay Pipe
VSG	Vitrified Segmented Duct
WT	Weeping Tile

APPENDIX C

TABLE C-1 *PIPE DEFICIENCY SORTED BY PIPE AGE*

Pipe Age	Cover Depth(m)	Diameter (mm)	Material	Waste Type	Observed Def Freq	Obs NonDef Freq	Observed Freq	Expected Def Prob	Observed Def Prob	Expected Def Prob
0-29	6-8+	150-375	TP	SAN	26	12	38	26	68	68
0-29	0-6	150-375	TP	CMB	4	3	7	4	57	62
0-29	0-6	150-375	TP	SAN	22	15	37	22	59	59
0-29	6-8+	150-375	TP	CMB	1	1	2	1	50	50
0-29	6-8+	150-375	RCP	SAN	1	2	3	1	33	33
0-29	0-6	150-375	CP	STM	11	24	35	11	31	31
0-29	0-6	550-1050	TP	CMB	1	2	3	1	33	22
0-29	0-6	450-525	RCP	STM	2	8	10	2	20	20
0-29	0-6	450-525	CP	STM	2	10	12	2	17	17
0-29	0-6	550-1050	RCP	STM	2	36	38	2	5	5
30-59	0-6	150-375	CP	STM	10	7	17	10	59	59
30-59	0-6	150-375	TP	STM	6	5	11	6	55	55
30-59	0-6	450-525	RCP	CMB	1	1	2	1	50	50
30-59	0-6	150-375	TP	CMB	25	32	57	25	44	45
30-59	0-6	450-525	CP	CMB	1	2	3	1	33	33
30-59	0-6	150-375	CP	CMB	10	24	34	10	29	29
30-59	0-6	150-375	TP	SAN	4	10	14	4	29	29
30-59	0-6	450-525	TP	CMB	3	11	14	3	21	19
30-59	0-6	550-1050	TP	CMB	1	6	7	1	14	13
60-90+	0-6	150-375	CP	CMB	14	3	17	14	82	82
60-90+	0-6	150-375	NC	CMB	24	13	37	24	65	65
60-90+	0-6	150-375	TP	CMB	159	89	248	158	64	64
60-90+	0-6	150-375	PVC	CMB	1	1	2	1	50	50
60-90+	0-6	550-1050	RCP	CMB	4	5	9	4	44	44
60-90+	0-6	450-525	TP	CMB	7	15	22	7	32	34
60-90+	0-6	550-1050	TP	CMB	1	5	6	1	17	24

APPENDIX C

TABLE C-2 *PIPE DEFICIENCY SORTED BY COVER DEPTH*

Pipe Age	Cover Depth(m)	Diameter (mm)	Material	Waste Type	Observed Def Freq	Obs NonDef Freq	Observed Freq	Expected Def Prob	Observed Def Prob	Expected Def Prob
0-29	0-6	150-375	TP	CMB	4	3	7	4	57	62
0-29	0-6	150-375	TP	SAN	22	15	37	22	59	59
0-29	0-6	150-375	CP	STM	11	24	35	11	31	31
0-29	0-6	550-1050	TP	CMB	1	2	3	1	33	22
0-29	0-6	450-525	RCP	STM	2	8	10	2	20	20
0-29	0-6	450-525	CP	STM	2	10	12	2	17	17
0-29	0-6	550-1050	RCP	STM	2	36	38	2	5	5
30-59	0-6	150-375	CP	STM	10	7	17	10	59	59
30-59	0-6	150-375	TP	STM	6	5	11	6	55	55
30-59	0-6	450-525	RCP	CMB	1	1	2	1	50	50
30-59	0-6	150-375	TP	CMB	25	32	57	25	44	45
30-59	0-6	450-525	CP	CMB	1	2	3	1	33	33
30-59	0-6	150-375	CP	CMB	10	24	34	10	29	29
30-59	0-6	150-375	TP	SAN	4	10	14	4	29	29
30-59	0-6	450-525	TP	CMB	3	11	14	3	21	19
30-59	0-6	550-1050	TP	CMB	1	6	7	1	14	13
60-90+	0-6	150-375	CP	CMB	14	3	17	14	82	82
60-90+	0-6	150-375	NC	CMB	24	13	37	24	65	65
60-90+	0-6	150-375	TP	CMB	159	89	248	158	64	64
60-90+	0-6	150-375	PVC	CMB	1	1	2	1	50	50
60-90+	0-6	550-1050	RCP	CMB	4	5	9	4	44	44
60-90+	0-6	450-525	TP	CMB	7	15	22	7	32	34
60-90+	0-6	550-1050	TP	CMB	1	5	6	1	17	24
0-29	6-8+	150-375	TP	SAN	26	12	38	26	68	68
0-29	6-8+	150-375	TP	CMB	1	1	2	1	50	50
0-29	6-8+	150-375	RCP	SAN	1	2	3	1	33	33

APPENDIX C

TABLE C-3 *PIPE DEFICIENCY SORTED BY PIPE DIAMETER*

Pipe Age	Cover Depth(m)	Diameter (mm)	Material	Waste Type	Observed Def Freq	Obs NonDef Freq	Observed Freq	Expected Def Prob	Observed Def Prob	Expected Def Prob
0-29	0-6	150-375	TP	CMB	4	3	7	4	57	62
0-29	0-6	150-375	TP	SAN	22	15	37	22	59	59
0-29	0-6	150-375	CP	STM	11	24	35	11	31	31
30-59	0-6	150-375	CP	STM	10	7	17	10	59	59
30-59	0-6	150-375	TP	STM	6	5	11	6	55	55
30-59	0-6	150-375	TP	CMB	25	32	57	25	44	45
30-59	0-6	150-375	CP	CMB	10	24	34	10	29	29
30-59	0-6	150-375	TP	SAN	4	10	14	4	29	29
60-90+	0-6	150-375	CP	CMB	14	3	17	14	82	82
60-90+	0-6	150-375	NC	CMB	24	13	37	24	65	65
60-90+	0-6	150-375	TP	CMB	159	89	248	158	64	64
60-90+	0-6	150-375	PVC	CMB	1	1	2	1	50	50
0-29	6-8+	150-375	TP	SAN	26	12	38	26	68	68
0-29	6-8+	150-375	TP	CMB	1	1	2	1	50	50
0-29	6-8+	150-375	RCP	SAN	1	2	3	1	33	33
0-29	0-6	450-525	RCP	STM	2	8	10	2	20	20
0-29	0-6	450-525	CP	STM	2	10	12	2	17	17
30-59	0-6	450-525	RCP	CMB	1	1	2	1	50	50
30-59	0-6	450-525	CP	CMB	1	2	3	1	33	33
30-59	0-6	450-525	TP	CMB	3	11	14	3	21	19
60-90+	0-6	450-525	TP	CMB	7	15	22	7	32	34
0-29	0-6	550-1050	TP	CMB	1	2	3	1	33	22
0-29	0-6	550-1050	RCP	STM	2	36	38	2	5	5
30-59	0-6	550-1050	TP	CMB	1	6	7	1	14	13
60-90+	0-6	550-1050	RCP	CMB	4	5	9	4	44	44
60-90+	0-6	550-1050	TP	CMB	1	5	6	1	17	24

APPENDIX C

TABLE C-4 *PIPE DEFICIENCY SORTED BY PIPE MATERIAL*

Pipe Age	Cover Depth(m)	Diameter (mm)	Material	Waste Type	Observed Def Freq	Obs NonDef Freq	Observed Freq	Expected Def Prob	Observed Def Prob	Expected Def Prob
0-29	0-6	150-375	CP	STM	11	24	35	11	31	31
30-59	0-6	150-375	CP	STM	10	7	17	10	59	59
30-59	0-6	150-375	CP	CMB	10	24	34	10	29	29
60-90+	0-6	150-375	CP	CMB	14	3	17	14	82	82
0-29	0-6	450-525	CP	STM	2	10	12	2	17	17
30-59	0-6	450-525	CP	CMB	1	2	3	1	33	33
60-90+	0-6	150-375	NC	CMB	24	13	37	24	65	65
60-90+	0-6	150-375	PVC	CMB	1	1	2	1	50	50
0-29	6-8+	150-375	RCP	SAN	1	2	3	1	33	33
0-29	0-6	450-525	RCP	STM	2	8	10	2	20	20
30-59	0-6	450-525	RCP	CMB	1	1	2	1	50	50
0-29	0-6	550-1050	RCP	STM	2	36	38	2	5	5
60-90+	0-6	550-1050	RCP	CMB	4	5	9	4	44	44
0-29	0-6	150-375	TP	CMB	4	3	7	4	57	62
0-29	0-6	150-375	TP	SAN	22	15	37	22	59	59
30-59	0-6	150-375	TP	STM	6	5	11	6	55	55
30-59	0-6	150-375	TP	CMB	25	32	57	25	44	45
30-59	0-6	150-375	TP	SAN	4	10	14	4	29	29
60-90+	0-6	150-375	TP	CMB	159	89	248	158	64	64
0-29	6-8+	150-375	TP	SAN	26	12	38	26	68	68
0-29	6-8+	150-375	TP	CMB	1	1	2	1	50	50
30-59	0-6	450-525	TP	CMB	3	11	14	3	21	19
60-90+	0-6	450-525	TP	CMB	7	15	22	7	32	34
0-29	0-6	550-1050	TP	CMB	1	2	3	1	33	22
30-59	0-6	550-1050	TP	CMB	1	6	7	1	14	13
60-90+	0-6	550-1050	TP	CMB	1	5	6	1	17	24

APPENDIX C

TABLE C-5 *PIPE DEFICIENCY SORTED BY WASTE TYPE*

Pipe Age	Cover Depth(m)	Diameter (mm)	Material	Waste Type	Observed Def Freq	Obs NonDef Freq	Observed Freq	Expected Def Prob	Observed Def Prob	Expected Def Prob
30-59	0-6	150-375	CP	CMB	10	24	34	10	29	29
60-90+	0-6	150-375	CP	CMB	14	3	17	14	82	82
30-59	0-6	450-525	CP	CMB	1	2	3	1	33	33
60-90+	0-6	150-375	NC	CMB	24	13	37	24	65	65
60-90+	0-6	150-375	PVC	CMB	1	1	2	1	50	50
30-59	0-6	450-525	RCP	CMB	1	1	2	1	50	50
60-90+	0-6	550-1050	RCP	CMB	4	5	9	4	44	44
0-29	0-6	150-375	TP	CMB	4	3	7	4	57	62
30-59	0-6	150-375	TP	CMB	25	32	57	25	44	45
60-90+	0-6	150-375	TP	CMB	159	89	248	158	64	64
0-29	6-8+	150-375	TP	CMB	1	1	2	1	50	50
30-59	0-6	450-525	TP	CMB	3	11	14	3	21	19
60-90+	0-6	450-525	TP	CMB	7	15	22	7	32	34
0-29	0-6	550-1050	TP	CMB	1	2	3	1	33	22
30-59	0-6	550-1050	TP	CMB	1	6	7	1	14	13
60-90+	0-6	550-1050	TP	CMB	1	5	6	1	17	24
0-29	6-8+	150-375	RCP	SAN	1	2	3	1	33	33
0-29	0-6	150-375	TP	SAN	22	15	37	22	59	59
30-59	0-6	150-375	TP	SAN	4	10	14	4	29	29
0-29	6-8+	150-375	TP	SAN	26	12	38	26	68	68
0-29	0-6	150-375	CP	STM	11	24	35	11	31	31
30-59	0-6	150-375	CP	STM	10	7	17	10	59	59
0-29	0-6	450-525	CP	STM	2	10	12	2	17	17
0-29	0-6	450-525	RCP	STM	2	8	10	2	20	20
0-29	0-6	550-1050	RCP	STM	2	36	38	2	5	5
30-59	0-6	150-375	TP	STM	6	5	11	6	55	55

APPENDIX 4

VISUAL BASIC CODES FOR TASES PROGRAM

```
Private Sub cmdOK_Click()
   Unload Me
   If Not Main.Visible Then
      Main.Show 1
   End If
End Sub


Private Sub Form_Load()

   Timer1.Interval = 1000   ' Set interval.

End Sub

Private Sub Timer1_Timer()
   If Not Main.Visible Then
      Unload Me
      Main.Show 1
   End If
End Sub
```

```
Private Sub Command1_Click()
   Main.txtDate = calDate
   Unload Me
End Sub

Private Sub Form_Load()
   calDate = Date
End Sub
```

```
Dim DB As Database, RS As Recordset

Private Sub cmdCancel_Click()
   Unload Me
End Sub

Private Sub cmdOK_Click()
Dim NoRecord As Boolean
Dim temRank As Double
Dim msg As String

   NoRecord = False
   temRank = 0: msg = ""

   'Waste Type
   Select Case Cmbwaste.Text 'Cmbwaste.ListIndex
      Case "Sanitary"
```

```
        temRank = temRank + 0.574
    Case "Storm"
        temRank = temRank + 0.081
    Case "Combined"
        temRank = temRank + 0.346
    Case Else
        msg = "Waste Type, "
        NoRecord = True
End Select

'Material
Select Case cmbmaterial.Text
    Case "CP"
        temRank = temRank + 0.201
    Case "TP"
        temRank = temRank + 0.717
    Case "RCP"
        temRank = temRank + 0.027
    Case "NC"
        temRank = temRank + 0.055
    'Case "PVC", "CMP", "EYE", "PE"
    '    NoRecord = True
    Case Else
        msg = msg & "Material, "
        NoRecord = True
End Select

'Build Year
Select Case cmbyear.Text
    Case "1910 - 1919"
        temRank = temRank + 0.469
    Case "1920 - 1929"
        temRank = temRank + 0.098
    Case "1930 - 1939"
        temRank = temRank + 0.167
    Case "1940 - 1949"
        temRank = temRank + 0.045
    Case "1950 - 1959"
        temRank = temRank + 0.043
    Case "1960 - 1969"
        temRank = temRank + 0.05
    Case "1970 - 1979"
        temRank = temRank + 0.12
    Case "1980 - 1989"
        temRank = temRank + 0
    Case "> 1990"
        temRank = temRank + 0.008
    Case Else
        msg = msg & "Build Year, "
        NoRecord = True
End Select

'Depth
Select Case cmbdepth.Text
    Case "0 - 2 m"
        temRank = temRank + 0.093
```

```
    Case "2 - 4 m"
      temRank = temRank + 0.296
    Case "4 - 6 m"
      temRank = temRank + 0.14
    Case "6 - 8 m"
      temRank = temRank + 0.239
    Case "8 - 10 m"
      temRank = temRank + 0.232
    Case "> 10 m"
      temRank = temRank + 0.232
    Case Else
      msg = msg & "Depth, "
      NoRecord = True
End Select

'Diameter
Select Case cmbdiameter.Text
    Case "200"
      temRank = temRank + 0.348
    Case "250"
      temRank = temRank + 0.162
    Case "300"
      temRank = temRank + 0.161
    Case "375"
      temRank = temRank + 0.084
    Case "450"
      temRank = temRank + 0.154
    Case "525"
      temRank = temRank + 0.048
    Case "600"
      temRank = temRank + 0.042
    Case Else
      msg = msg & "Diameter."
      NoRecord = True
End Select

'Normalize maximum = 2.404, minimum = 0.251, differnce = 2.153
temRank = (temRank - 0.251) / 2.153
txtRank = temRank

If NoRecord Then
   txtComment.Text = "No deficency record for this " & msg
   temRank = 0.5
   txtRank = -1
ElseIf temRank > 0.9 Then
   txtComment.Text = "Danger!!!"
ElseIf temRank > 0.66 Then
   txtComment.Text = "Do somthing!!"
ElseIf temRank > 0.5 Then
   txtComment.Text = "Pay attention to!"
ElseIf temRank > 0.33 Then
   txtComment.Text = "OK"
Else
   txtComment.Text = "Relax"
End If
```

```
        'Move arrow around, top = 430, bottom = 3630, Y1 = 600, Y2 = 3840
        sspNote.Top = 3630 - CInt((3630 - 430) * temRank)
        linNote.Y1 = 3840 - CInt((3840 - 600) * temRank)
        linNote.Y2 = linNote.Y1
        lblNote.Caption = Format(temRank, "##.00%")

    End Sub

    Private Sub OpenFile()
        'datHistogram.DatabaseName = App.Path & "\CityofED.MDB"
        'datHistogram.Recordset.MoveLast
    On Error Resume Next
        Set DB = Workspaces(0).OpenDatabase(FileName)
        Set RS = DB.OpenRecordset("Histogram")
        Set datHistogram.Recordset = RS
        datHistogram.Recordset.MoveLast
        datHistogram.Recordset.AddNew
    End Sub


    Private Sub mnuFileExit_Click()
        Unload Me
    End Sub

    Private Sub mnuFileNew_Click()
    On Error Resume Next
        'RS.Close: DB.Close
        'FrmParameter.Enabled = False
        diaFile.DialogTitle = "New"
        ' Set filters.
        diaFile.Filter = "All Files (*.*)|*.*|Database Files (*.mdb)|*.mdb"
        ' Specify default filter.
        diaFile.FilterIndex = 2
        ' Display the File Open dialog box.
        diaFile.ShowOpen
        FileName = diaFile.FileName
        FileCopy App.Path & "\CityofED.CFG", FileName
        OpenFile
        FrmParameter.Enabled = True
    End Sub

Private Sub mnuFileOpen_Click()
On Error Resume Next
    'RS.Close: DB.Close
    FrmParameter.Enabled = False
    diaFile.DialogTitle = "Open"
    ' Set filters.
    diaFile.Filter = "All Files (*.*)|*.*|Database Files (*.mdb)|*.mdb"
    ' Specify default filter.
    diaFile.FilterIndex = 2
    ' Display the File Open dialog box.
    diaFile.ShowOpen
    FileName = diaFile.FileName
    If Dir(FileName) Then
        OpenFile
        FrmParameter.Enabled = True
```

```
   End If
End Sub

Private Sub mnuFileSaveAs_Click()
Dim OldFile As String
Dim response As Integer
On Error GoTo CloseError
   OldFile = FileName
   diaFile.DialogTitle = "Save As"
   ' Set filters.
   diaFile.Filter = "All Files (*.*)|*.*|Database Files (*.mdb)|*.mdb"
   ' Specify default filter.
   diaFile.FilterIndex = 2
   ' Display the Save As dialog box.
   diaFile.ShowSave
   FileName = diaFile.FileName

   If Dir(FileName) <> "" Then       ' File already exists, so ask if the user wants to overwrite the file.
      response = MsgBox("Overwrite existing file?", vbYesNo + vbQuestion + vbDefaultButton2)
      If response = vbNo Then Exit Sub
   End If
   FileCopy OldFile, FileName 'does not work due to the oldfile is being used
   Exit Sub
CloseError:
   MsgBox Err.Description
End Sub


Private Sub mnuHelpAbout_Click()
   About.Show 1
End Sub


Private Sub mnuHelpIntroduction_Click()
   Introduction.Show 1
End Sub


Private Sub mnuToolsCalander_Click()
   Calander.Show 1
End Sub


Private Sub Picture3_Click()
On Error Resume Next
   datHistogram.Recordset.Delete
   datHistogram.Refresh
End Sub
```

```
Public FileName As String
```