DOCUMENT RESUME

ED 061 980 LI 003 650

AUTHOR Lazorick, Gerald J.
TITLE Demand Models for Books in Library Circulation Systems. Final Report.
SPONS AGENCY National Science Foundation, Washington, D.C. Office of Science Information Services.
PUB DATE Jul 70
NOTE 233p.; (48 References)

EDRS PRICE MF-$0.65 HC-$9.87
DESCRIPTORS *Books; *Library Automation; *Library Circulation; Library Materials; *Mathematical Models; *Operations Research

ABSTRACT

This research is a study of demands for books in library circulation systems. Demand data for random samples of books were collected and fitted to various standard distributions. The numbers of demands for collections of books are shown to be Negative Binomially distributed. As is shown, this implies that the numbers of demands for individual books in the collection are Poisson distrubuted and that the demand rate varies from book to book according to a Gamma distribution. Using these facts and assuming Exponentially distrubuted loan intervals, a model is developed which will predict the availability and unavailability of a book in a library. The practicality of using the model is demonstrated. (Author)

FINAL REPORT

DEMAND MODELS FOR BOOKS IN LIBRARY CIRCULATION SYSTEMS

by

Gerald J. Lazorick
Associate Professor
School of Information & Library Studies
Faculty of Engineering and Applied Sciences
State University of New York at Buffalo

This research was supported by Grant GN-706
from the Office of Science Information Services
of the National Science Foundation.

July, 1970

TABLE OF CONTENTS

2

Page No.

Page No.

ABSTRACT

This research is a study of demands for books in library circulation systems. Demand data for random samples of books were collected and fit to various standard distributions. The numbers of demands for collections of books are shown to be Negative Binomially distributed. As is shown, this implies that the numbers of demands for individual books in the collection are Poisson distributed and that the demand rate varies from book to book according to a Gamma distribution. Using these facts and assuming Exponentially distributed loan intervals, a model is developed which will predict the availability and unavailability of a book in a library. The practicality of using the model is demonstrated.

ACKNOWLEDGEMENTS

The author expresses his thanks to Mr. Bernard I. Bayer for his helpful editing comments, and the staff of the Technical Information Dissemination Bureau for helping to prepare this document.

ACKNOWLEDGEMENTS

The author expresses his thanks to Professors Warren Thomas (Major Professor), Richard Schmidt, Wayland P. Smith and Vincent E. Giuliano for their suggestions, Mr. Bernard I. Bayer for his helpful editing comments, and the staff of the Technical Information Dissemination Bureau for helping to prepare this document. Thanks is also expressed to the author's wife, Sylvia, for the many ways in which she helped during the course of the research.

Partial support for this research was received from the National Science Foundation (Grant Number GN-706).

LIST OF TABLES

Table		Page No.
6.2	CHI SQUARED GOODNESS OF FIT TABLE FOR A NEGATIVE BINOMIAL FIT TO THE FREQUENCIES OF KNOWN DEMAND FOR A SAMPLE OF BOOKS FROM THE SUNY-BUFFALO LIBRARY COLLECTION DURING THE PERIOD SEPTEMBER 1, 1966 - AUGUST 31, 1967	94
6.3	VARIANCE TEST TABLE FOR KNOWN DEMAND SAMPLE OF BOOKS DEMANDED FROM SEPTEMBER 1, 1966 - AUGUST 31, 1967	96
6.4	FREQUENCY DISTRIBUTION OF DEMANDS FOR P CLASS BOOKS DURING THE PERIOD MARCH 3, 1968 - MAY 11, 1968	98
6.5	CHI SQUARED GOODNESS OF FIT TABLE FOR A NEGATIVE BINOMIAL FIT TO THE FREQUENCIES OF DEMAND FOR THE SAMPLE OF P CLASS BOOKS DEMANDED DURING MARCH 3, 1968 - MAY 11, 1968	99
6.6	VARIANCE TEST TABLE FOR DEMANDS FOR P CLASS SAMPLE OF BOOKS DEMANDED FROM MARCH 3, 1968 - MAY 11, 1968	101
7.1	DAILY DEMAND FREQUENCIES OF TEN RANDOM SAMPLES OF ONE HUNDRED P CLASS BOOKS DEMANDED DURING MARCH 3, 1968 - MAY 11, 1968	106
7.2	SUMMARY OF THE CHI SQUARED GOODNESS OF FIT TEST FOR A POISSON FIT TO THE FREQUENCIES OF DAILY DEMAND FOR SAMPLES OF BOOKS DEMANDED DURING THE PERIOD MARCH 3, 1968 - MAY 11, 1968	108
7.3	SUMMARY OF THE RESULTS OF THE VARIANCE TEST RUN ON SAMPLES OF DAILY DEMAND FREQUENCIES OF P CLASS BOOKS DEMANDED DURING THE PERIOD MARCH 3, 1968 - MAY 11, 1968	110

Table		Page No.

Table		Page No.

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Background

Ever since libraries have been in existence, methods were used by librarians to control the availability of library materials so as to satisfy the demands of patrons. Traditionally, librarians have controlled the availability of books by setting the loan intervals or by ordering multiple copies of books in high demand. Circulation department librarians when exercising control in this manner have given attention, although probably not overtly, to the two most important objectives of a circulation department in a library. These are: (1) to have a book available when it is demanded and (2) to allow a patron to charge out a book for as long as he needs it. In establishing loan intervals, the librarian must and does give attention to these objectives. Little has been done to analytically determine these loan intervals. In all cases in all libraries they have been subjectively, although thoughtfully established. Analytic methods have not been used probably because the kinds of data which are needed for a thorough analytic evaluation are not easily available and because collection of data of the type needed is time consuming and expensive.

The advocates of the use of computers in libraries are quick to point out that automated circulation systems will not only reduce patron time, but will also provide data concerning book usage, which will be useful in the planning functions of the library. Trueswell (1964) points out:

> "Mechanization of library procedure cannot directly improve the probability of not finding the book wanted, but instead merely makes the statistic more obvious. However, one of the benefits of a mechanized punched card control system is the fact that complete analyses can be made about the use of books and their frequency of circulation. From these analyses it would be possible to develop new operating procedures relative to charge periods, duplicate copies and non-circulation copies. These new procedures would reduce the probability of the user being unable to find the book he wanted."

In undertaking this research, this investigator contended that librarians and the computer specialists did not really know how they would use the data which were soon to inundate them. What does it mean when book X circulates 1, or 5, or 25 times a year? How can the librarian use this knowledge? Morse (1968, pp. 141-142) apparently shares this investigator's concern when he states:

> "It is the author's belief based on discouraging experience, that neither the computer experts nor the librarian (for different reasons) really knows what data would be useful for the librarian to have collected, analyzed, and displayed, so he can make decisions with some knowledge of what the decision implies. What is needed before the computer designs are frozen is for models, of the sort developed in this book, to be played with, to see which of them could be useful and to see what data are needed and in what form, in order that both models and computers can be used most effectively by the librarian.
>
> If it is so designed, the computer not only can gather the needed data but also can efficiently help to carry out the operating policy, once decided on .... Again what is needed is some experimenting with models of operation of the sort discussed in this book, so as to insure that the computer will utilize, as completely as possible, the data it gathers and stores and also to prevent the system from being overdesigned to do things that turn out to be inutile or vacuous."

One of the most important variables that must be predicted before analytically derived decision rules concerning library circulation problems could be formulated is the demand for a book. A knowledge of how many times a book is going to be demanded is most important in the establishing of decision rules for library circulation systems. Predicting book circulation or book use is not sufficient for determining circulation department policy. One must be able to predict or estimate demand. It is not unlike the sales problem in which future sales are predicted on the basis of past sales and not on demand, possibly resulting in a significant number of dissatisfied or lost customers.

1.2 Definition of a Demand for a Book

For the purposes of this research, a demand for a specific book is an expressed or unexpressed need for a book once it has been identified as being available in the library.

An expressed need is one in which the library system knows that the need exists either because the patron found the book he wanted and charged it out or did not find the book he wanted and asked that it be saved or located.

An <u>unexpressed need</u> is one in which the library system does not know that the need exists either because the patron found the book he wanted, used the book in the library but did not charge it out, or one in which the patron did not find the book in the library but did not indicate that the book should be saved or found.

A book found by browsing becomes a demand once a book is found which the patron will use. An unsatisfied need of a browser then does not satisfy the definition of demand made above.

There are also demands for books which are not in the collection of the library, such as those books for which no cards are found in the card catalog. This type demand doesnot meet the definition of demand given above.

1.3 Distinction between Book Demand, Book Use and Book Circulation

When a patron has identified that a book he needs is in the library and attempts to get the book, a demand has occurred. If he finds the book and uses it in the library or if he charges out the book for home use, we say that a book use has occurred. The latter type use is called a circulation. The literature is confusing as to what actually has been studied in particular investigations. This research reported here is a study of book demand which includes book use and circulations as well as unsatisfied demand.

1.4 Definitions of Terms

1. A loan interval is the interval of time for which a library patron may borrow a book from a library.

2. A closed stack library is one in which access to the collection is not available to the patron (no browsing).

3. An open stack library is one in which access to the collection is available to the patron (browsing allowed).

4. The demand for a library book is an expressed (known to the library system) or unexpressed

(unknown to the library system) need for a book which has been identified as being in the library's collection.

5. A satisfied demand for a book occurs when the patron finds the book which he seeks.

6. An unsatisfied demand occurs when a patron does not find the book which he seeks.

7. A book use occurs when a demand is satisfied and the patron uses the book in the library or charges it out.

8. A book circulation occurs when a demand is satisfied and the patron charges out the book.

9. A known demand is one which becomes known to the library system either because the patron charges out the book, does not find the book but asks the circulation department staff to find it, or finds that the book is already charged out and asks that the book be saved when it is returned.

10. An unknown demand occurs when the patron does not find the book he seeks, does not ask the circulation staff to find it, and does not ask that it be saved, or else the patron finds the book he seeks and uses it in the library.

1.5 Objectives of the Research

This investigation is a study of book demand and was conducted in order to:

1. Provide evidence that one cannot reject the hypothesis that both the known and unknown demand for books is Poisson distributed.

2. Develop a practical model to predict total and unknown demand based on known demand data readily available in library circulation systems.

1.6 Approach to the Problem

In order to fulfill Objective 1, book demand data were collected to verify the hypothesis that the number of demands for individual books is Poisson distributed cannot be rejected. This was done by randomly selecting books from the entire SUNY/Buffalo Library collection and collecting data concerning the number of times each book was charged out of the library over a period of time.

In addition, for all books in the P Class (Language and Literature) collection all known and unknown demands were recorded from February 13, 1968 through May 11, 1968. Random samples of P Class books were then selected for analysis.

With both sets of demand data, various standard distributions were fit to the number of demands for samples of books. Ultimately, the Negative Binomial distribution was fit to the data. This indicated that the demands for individual books are Poisson distributed, but that the mean demand rate varies from book to book, according to a Gamma distribution. The Poisson hypothesis was further verified by the data on the number of demands per day of random samples of books, and showed that the number of demands per day was Poisson distributed.

Further evidence is given to permit non-rejection of the

Poisson hypothesis by the analysis of 18 samples of data presented by Fussler and Simon (1961) concerning book use in various subject areas at the libraries at Yale University, Northwestern University, the University of California at Berkeley, and the University of Chicago.

In order to fulfill Objective 2, the knowledge that demands for individual books is Poisson distributed was used to develop a model with which the number of demands for a book could be predicted, from information about known demand, specifically, the number of times the book circulates over a period of time. At this point, it was demonstrated that the known demand information could provide insights as to the future availability, or unavilability, of the book in the library.

1.7 Conventions Employed in the Report

The decimal numbering system was used to number Chapters, Sections, and Subsections. Appendices are distinguished from the body of the report by using capital letters. When appropriate, equations are numbered consecutively throughout each chapter. The equation numbers are enclosed within brackets. Tables are numbered consecutively throughout each chapter and appendix. The decimal numbering system is also employed here. References are always made to the smallest subsection. Thus, a reference to a section implies a reference to the appropriate subsection.

CHAPTER 2

REVIEW OF RELATED RESEARCH

Studies of the use of books, journals and facilities in libraries are abundant within the literature. DeWeese (1967), for example, published a bibliography of 547 use studies found in the literature. In his thesis, Jain (1967) indicates one of his purposes is to catalog the important history of statistical analysis and mathematical modeling of library usage. He listed 84 references and provided detailed appendices so that "in the future, research workers would only have to consult his thesis for early (up to 1965) references and their summaries."

Although there are many reported studies of book usage models, there are no reported studies of book demand. There is, however, one mathematical prediction model developed by Morse in a recently published highly significant book which attempts to apply the analytic method of operations research-systems analysis to the operating problems in the Library (Morse, 1968). The model will be explained later in this section.

It is the author's purpose here to review the writings believed to be relevant to the development of models to predict demand even though most of the following studies relate to the prediction of book usage and book circulations. This review does not cover journal usage or journal circulations.

2.1 The First Book Use Study

The earliest published use study was conducted by Ranck (1911) in the Grand Rapids Public Library. His study, as do most other use studies, provides breakdowns of the used books in a library by various characteristics such as class, grade average of students, etc.

2.2 The Fussler and Simon Data at the University of Chicago Library

The first truly large and thorough study of book usage was made by Fussler and Simon (1961) at the University of Chicago. The fundamental question to which they directed their attention was:

> "Will any kind of statistical procedure predict with reasonable accuracy, the frequencies with which groups of books with defined characteristics are likely to be used in a research library?"

Their study was based on certain assumptions, some of which were later examined empirically. These assumptions were: (1) The recorded circulation use of books is a reasonably reliable index of all use, including browsing use, within the library. (2) There are some patterns in the use of books common to major research libraries. (3) Within homogeneous subject areas and types of books (i.e., monographs, serials), use is a suitable initial criterion for the segregation of materials into different levels of accessibility. (4) Economic factors make it highly desirable to segregate books, on the basis of their value and use, into two or more levels of accessibility. Fussler and Simon used use data on a sample of 600 books (400 selected using a systematic sampling plan and 200 using a stratified random sampling plan). With these data, they studied 22 functional relationships using "use" as a dependent variable and various independent variables. The 22 functions fell into three broad categories:

(1) Functions for libraries with no record of prior use.

(2) Functions that require five-year past use records.

(3) Functions employing long records of past use. Fussler and Simon conclude that:

> "By far the best predictor of the future use of a title is its past use. Because of the low probability of use in any one year for titles in the marginal-value range of titles in a library the size of the University of Chicago Library, a fifteen or twenty year long observation period produces considerably better results than will an observation period of five years.
>
> Some research libraries have no records of past use. If these libraries wish to commence storage immediately our results also suggest the best possible functions and suggest to them the extent of the errors that will arise. Our results also suggest the wisdom of postponing storage for perhaps five years and collecting records of the use during that time. If a library does wish to initiate a storage plan without waiting to collect such records, it might well consider setting up a system by which high-use books that were sent to storage could easily be restored to the working collection."

Fussler and Simon also point out that their assumption of a stochastic model does not suggest that the Poisson distribution will approximate the distribution of books within a library (or within a given subject area) by the frequency of their use during some period of time. They say that their observed distributions do not resemble the Poisson closely, having a much higher variance. There are too many observations at "zero" and at multiple use points. Chapter 8 of the research reported here is devoted to an analysis of some of the Fussler and Simon data.

2.3 Deterministic Book Use Models

Most of the reported mathematical models of book use have been deterministic. The first mathematical model was reported by Shaffer and Ernst (1954). Their model was developed from a non-random sample of 237 books which were borrowed from the M.I.T. Science Library during the period May 5 - 10, 1954. After attempting to fit various models, they found that the average annual circulation I(t) of a t year old book was:

$$I(t) = 10\, e^{-0.9t} + 1.1\, e^{-0.016t}$$

Rothkopf (1962) using a non random sample of 97 books which were at least three years old attempted to fit various models to the data and considered that a good fit was obtained in which $\overline{C}(t)$, the average annual circulation of the book during the first t years in the library, was related to $\bar{C}(3)$, the average circulation of a book during the first three years in the library by the model:

$$\frac{\bar{C}(t)}{\bar{C}(3)} = \sqrt{\frac{t-.3}{2.7}}$$

Dawson, Aldrin and Gould (1962) using a sample of 202 books developed a mc lel based on two assumptions.

(1) Within a homogeneous class of books, there is a maximum value of initial circulation rate. The most popular books in that class will tend to achieve that rate, r.

(2) A particular book will eventually decrease in popularity and the circulation rate will fall off inversely with time. They fit various models and found the best fit which would allow the determination of C(t), the cumulative circulation of a book during the first t years in the library, to be:

$$C(t) = 5.12 \log \left(1 + \frac{t}{0.77}\right)$$

Leimkuhler (1966) proposed a circulation model with a constant rate of obsolescence in which the average manual circulation rate for books, C(t), whose age is t, can be determined using C_0, the average annual circulation rate in the first year of acquisition ($C_0 \geq 0$) and B a constant rate of obsolescence ($0 \leq B \leq 1$) using the following model:

$$C_t = C_0 (1 - B)^t$$

2.4 <u>Probabilistic Models of Book Use</u>

Dawson (1962) developed a Markov model for predicting the future circulation of a book on the basis of its immediate past circulation. His model was developed from a sample of 305 books. His states were 0 circulations per year, 1 or 2 circulations per year and 3 or more circulations per year. He tried several functions and found the following to be close approximations to his observed transition probabilities:

$$p_{00}(n) = \frac{9}{10} - \left(\frac{3}{8}\right)\left(\frac{8}{10}\right)^n$$

$$p_{01}(n) = \frac{1}{10} + \left(\frac{2}{8}\right)\left(\frac{8}{10}\right)^n$$

$$p_{02}(n) = \left(\frac{1}{8}\right)\left(\frac{8}{10}\right)^n$$

$$p_{10}(n) = \frac{4}{10}$$

$$p_{11}(n) = \frac{4}{10}$$

$$p_{12}(n) = \frac{2}{10}$$

$$p_{20}(n) = \frac{6}{10} - \left(\frac{5}{8}\right)\left(\frac{8}{10}\right)^n$$

$$p_{21}(n) = \frac{4}{10}$$

$$p_{22}(n) = \left(\frac{5}{8}\right)\left(\frac{8}{10}\right)^n$$

Morse (1963, 1965, 1968) has suggested a Markov model which has been fit to circulation data of certain classes of books at the Massachusetts Institute of Technology. For all books having a mean circulation m in a given year, the mean circulation $\bar{n}$ for the subsequent year given by his model is:

$$\bar{n} = a + bm$$

where a is the "residual circulation" after time has removed the initial popularity and b = rate of diminution of circulation with time ($0 < b < 1$). Morse says that the actual circulations in the subsequent year are clustered about the expected value in accord with the Poisson distribution. Morse suggests that a Markov model is appropriate in which the transition probability (P_{mn}) can be found by:

$$P_{mn} = \frac{\bar{n}^{n}}{n!}(e)^{-\bar{n}}$$

$$\bar{n} = (a + bm)^{n}$$

2.5 The Jain Study at Purdue University

Since Morse (1963, 1965) suggested that the uses of a book followed a Poisson distribution, Jain (1967) examined the frequencies of uses of homogeneous groups of books from the Chemistry, Physics and Pharmacy Libraries at Purdue University and found that the "zero use" class does not follow the same probability law as the remaining classes. Jain, because of his findings with the "zero use" class of books went on to develop a model by splitting the probability that the book will be used into two components: (1) the probability of a book not being used and (2) the probability of uses if the book is used. He called the latter the "Pn Model." He proposed a truncated Poisson distribution as a possible fit to observed frequency distributions of book uses. He let X_n = the number of uses of a given book during the nth year after its acquisition in the library and then:

$$P(X_n = j) = \begin{cases} p_n & \text{if } j = 0 \\ (1 - p_n)\dfrac{e^{-\lambda}\lambda_n^{\,j}}{(1 - e^{-\lambda_n})\,j!} & \text{if } j = 1, 2, \ldots \end{cases}$$

where $0 \leq p \leq 1$ and $\lambda_n > 0$ and where, in general, $\lambda_n \neq -\log p$. He then studied the probability that $X_n = 0$, p_n, and developed his "pn function."

$$p_n(\beta, \gamma) \begin{cases} \dfrac{n - 1 + \beta}{n - 1 + \beta + \gamma} & \text{if } n = 1, 2, \ldots \\ 0 & \text{if } n = 0 \end{cases}$$

Where β, γ are parameters which depend on the class of books of which the book under consideration is a member and n is the year after acquisition.

Jain (1966) previously had developed a concept which he called "relative use" which he used in testing his Pn model and his pn function after he estimated the required parameters. Jain developed his method of relative use because, he felt, that the two methods previously used to study samples of book usage data were inadequate, namely the collection method in which random samples (or samples) of the entire collection were identified through the shelf list and then studied, and the check out sample method in which books checked out for a period of time were studied. The first method is criticized because it is difficult to collect data and because there are problems of missing data and of lack of control on the methods of recording usage histories in the past. The latter method is criticized primarily because one cannot draw inferences about the entire collection. Because of these deficiencies, Jain developed his method of relative use in which three independent samples of monograph titles are obtained from the total collection (S), the home used material (H) and the in-library used material (I). These samples are divided into a certain number of groups on the basis of the following characteristics of the title: language, country of publication, year of publication and year of accession. Using S_i as the number of titles in sample S which belong to the ith group, H_i as the number of titles in sample H which belong to the ith group, and I_i as the number of titles in sample I which belong to the ith group, he defined relative use R as follows:

$$R_i = \text{relative use of the ith group} = \frac{H_i + I_i}{S_i} (100)$$

As a further illustration of how the relative use method could be used to estimate the percentage of monographs used in the ith group, Jain presented the following using as an example only the checkouts for home use:

Let:

m = number of groups of monographs.

S_i = number of monographs in the ith group in the shelf list sample, $i = 1, 2, \ldots, m$.

H_i = number of monographs in the ith group in the 1966 Spring check-out sample, $i = 1, 2, \ldots, m$.

h_i = number of monographs which are included the ith group of 1966 Spring check out sample and the ith group of the shelf list sample, $i = 1, 2, \ldots, m$.

then, relative use for the ith group of monographs is:

$$\frac{H_i}{S_i}(100)$$

while the collection method estimate of the percentage of monographs used in the ith group is:

$$\frac{h_i}{S_i}(100)$$

If f = fraction of the total collection included in the shelf list sample S, then the estimate of the percentage of monographs used in the ith group based on relative use is:

$$f\left(\frac{H_i}{S_i}\right)100$$

Jain points out that it is not clear from the above that the relative use method of getting the estimate is superior to the collection method, but he states that if the fraction of the total collection is small or if the period for which usage data is recorded is short then it seems that the ratio:

$$\frac{h_i}{\text{expected value of } h_i}$$

will fluctuate widely around 1 for a given value of f_i. In this case, the relative use estimates are expected to do much better than the collection method estimate.

Jain compared the collection method estimate with the relative use estimate and the actual number of books used in selected subsets of a sample of books. Relative use gave values closer to the true value of use. Jain tested his Pn model and pn function by comparing estimates from the model with estimates of the relative use.

After estimating the parameters in his pn function, he fit the function to his relative use estimates for five subgroups of titles from his shelf list sample. The subgroups ranged in size from 22 - 67. He used the method of least squares, and the fit was poor (R^2 from .12 to .66). He also fit the pn function to four subgroups from the check out sample and found R^2 values of .89 - 94. He then fit the function to three subgroups of the in-library use sample and had poorer results than the home use sample but better results than the shelf list sample. Using the same subgroups he fit his Pn model, a generalized exponential model, a generalized logarithmic model, and the Morse model. He concluded that each of the models does equally well for predicting Pn. For certain subgroups, he showed that all models were equally poor.

Jain concludes that the pn function (no use) is quite efficient in describing the probability of no use of a book for various ages and that the Pn Model does as well as the three other book usage models when fitted to the data from Purdue Libraries. He further concludes that age is a significant variable in studying the usage of monographs. He points out that while usage rates of individual monographs have considerable variation even over a short period of time, the usage rates of various age groups do not show any significant difference over time.

2.6 The Morse Book Demand Model

Morse (1968, pp. 124-132) develops a model to estimate unsatisfied demand for a book. Earlier calling the number of uses per period of time the service rate of the book, μ, he shows that the expected circulation, R_1, is:

$$R_1 = \frac{\mu\lambda}{\mu + \lambda}$$

where λ is the demand rate. The expected unsatisfied demand is:

$$U_1 = \lambda - R_1 = \frac{\lambda^2}{\lambda + \mu}$$

Morse using the conditional probability relationship:

$$P(m|\lambda)\,P(\lambda) = P(\lambda|m)\,P_m$$

$$P(\lambda) = P(\lambda, m)\,P(m, \lambda)$$

where $P(\lambda,m) = P(m,\lambda)$ is the joint probability that a book has a demand λ and also circulates m times a year. Morse uses this relationship and goes from a formula giving expected circulation in terms of demand to one giving estimated demand in terms of known (or predicted) circulation. He develops the model because:

"Some of the things we would like to know, for each homogeneous class of books, are: the expected value of the demand for those books of a class that circulate m times a year; the mean value $D = \bar{\lambda}$ of the per book demand, averaged over all books of the class; and thus, by subtraction, the mean unsatisfied demand $\bar{U} = D - \bar{R}$ per book for all the books in the class. We would like to know how the unsatisfied demand would change in value if some of the books were duplicated of if the loan period $1/\mu$ were changed in value."

Morse points out that the calculations aren't easy because λ is not a definite number, the exact number of would-be borrowers who come to the library for a given book during the year. It is the expected demand. He points out that since λ is a continuous variable, $P(\lambda|m)$ and $P(\lambda)$ are probability densities. Secondly, he points out, the conditional probability $P(m|\lambda)$ is a quite complicated function and that it is not a bad approximation to use the Poisson formula:

$$P(m|\lambda) \simeq P_m\left(\frac{\mu\lambda}{\mu+\lambda}\right) = \frac{1}{m!}\left(\frac{\mu\lambda}{\mu+\lambda}\right)^m \exp\left(\frac{-\mu\lambda}{\mu+\lambda}\right)$$

Where:

$\dfrac{\mu\lambda}{\mu+\lambda}$ is the expected circulation.

He points out that this approximation is better than some made later in his development.

Morse points out that there is no way of measuring the distribution of demand $P(\lambda)$ for a class of books. He states that it is possible to determine $P(m)$ however, and using data collected in three different samples of 876, 204 and 97 titles in the collection of the Library of Science at MIT, found that the distribution is roughly geometric, though a semi-log plot showed a slight downward curvature, rather than a straight line. He points out that $P(m)$ is proportional to γ^m but that $P(0)$ is not $1 - \gamma$. The straight line does not extend back to $P(0)$.

Morse states that since the knowledge of $P(m|\lambda)$ and $P(m)$ alone does not enable one to find $P(\lambda)$ and $P(\lambda|m)$, $P(\lambda)$ has to be guessed and then the guess verified indirectly. Morse assumes that $P(\lambda)$ is exponential since it is an extension of the geometric distribution to the continuous variable λ. He tried:

$$P(\lambda) = \left(\frac{1}{D}\right) e^{-\frac{\lambda}{D}} \qquad \int_0^{\infty} P(\lambda)\,d\lambda = 1$$

as the probability that the demand for a randomly chosen book of the class is between λ and $\lambda + 1$. Then:

$$D = \int_0^\infty \lambda P(\lambda)\, d\lambda$$

and since:

$$P(m) = \int_0^\infty P(m|\lambda)\, P(\lambda)\, d\lambda$$

substituting:

$$P(m) = \frac{1}{m!D} \int_0^\infty \left(\frac{\mu\lambda}{\mu+\lambda}\right)^m \exp\left(-\frac{\lambda}{D} - \frac{\mu\lambda}{\mu+\lambda}\right) d\lambda$$

Morse states that numerically evaluating the integral for various values of m, λ and D produced a distribution P(m) which is nearly geometric when plotted against m on semi-log paper. Thus, he approximated P(m) with a geometric distribution designed to produce the same mean circulation as does the formula shown above and data collected by Pourny (1962). Morse found a modified geometric distribution to fit:

$$P(m) = \frac{1}{D+1}\left(\frac{Q}{Q+1}\right)^m \qquad \text{for } m > 0$$

$$P(0) = \frac{Q+1-R}{Q+1}$$

And
$$P(\geq m) = P(m) + P(m+1) + \ldots$$

$$\simeq \frac{Q+1}{D+1}\left(\frac{Q}{Q+1}\right)^m$$

where Q is set so that

$$\overline{R} = P(1) + 2P(2) + 3P(3) + \ldots = \frac{Q(Q+1)}{D+1}$$

or

$$Q = \left\{ \overline{R}\,(D+1) + \frac{1}{4} \right\}^{\frac{1}{2}} - \frac{1}{2}$$

Constant Q is the parameter of the modified geometric distribution.

Morse points out that the value of the mean circulation R can be calculated from the circulation distribution

$$P(m) = \frac{1}{m!\,D} \int_0^{\infty} \left(\frac{\mu\lambda}{\mu+\lambda} \right)^m \exp\left(-\frac{\lambda}{D} - \frac{\mu\lambda}{\mu+\lambda} \right) d\lambda$$

but that the individual P(m) cannot. He does this by using the series expansion $e^x = 1 + x + (x^2/2!) + (x^3/3!) + \ldots$

$$\overline{R} = \frac{1}{D} \int_0^{\infty} \left[\left(\frac{\mu\lambda}{\mu+\lambda} \right) + \frac{2}{2!} \left(\frac{\mu\lambda}{\mu+\lambda} \right)^2 + \frac{3}{3!} \left(\frac{\mu\lambda}{\mu+\lambda} \right)^3 + \ldots \right] \exp\left(-\frac{\lambda}{D} - \frac{\mu\lambda}{\mu+\lambda} \right) d\lambda$$

$$= \frac{\mu}{D} \int_0^{\infty} \left(1 - \frac{\mu}{\mu+\lambda} \right) \exp\left(-\frac{\lambda}{D} \right) d\lambda$$

$$= \mu \left[1 - \frac{\mu}{D} \exp\left(\frac{\mu}{D} \right) E_1 \left(\frac{\mu}{D} \right) \right]$$

where $E_1(x)$ is the exponential integral of x.

He gives the approximate formulas for the above as

$$\frac{D-\overline{R}}{\mu} \simeq 1.11\left(\frac{D}{\mu}\right)^{1.87} \simeq 2.5\left(\frac{R}{\mu}\right)^{2.06}$$

Thus given $\overline{R}$ and μ, D can be calculated and once D is known, Q can be calculated. Once Q is known, P(m) can be calculated.

Morse points out that $D - \overline{R}$ is the mean value of the unsatisfied demand per book of the class, since D is the estimated value of the mean demand.

Morse states that his choice of a simple model and the approximations he has made may result in values of D being off by 10 or even 30 percent but that this does not matter for the needs because one is generally interested in whether D is 2 or 200 and in whether some action increases or decreases $D - \overline{R}$ by factors of 2 or more. The formulas, he states are adequate for such purposes.

CHAPTER 3

THE NEGATIVE BINOMIAL DISTRIBUTION

The negative binomial distribution has been used in many diverse experimental situations. One of the oldest applications was made by Greenwood and Yule (1920). They found that the number of industrial accidents among workers during relatively short periods of time may follow a negative binomial distribution. Apparently, the distribution over periods of equal lengths of time and for the ith worker would be Poisson, but accident proneness from worker to worker varies in such a way that the numbers of accidents for a group of workers may follow a negative binomial distribution. It has also been found that counts of insects on experimental plots often conform to a negative binomial distribution, presumably because their gregariousness or hatching habits introduce "contagion" in the sense that the location of one insect in a plot increases the probability that there are others in the plot.

Kendall (1949) reports that when the birth and death rates per individual within a living population are constant, and there is a constant rate of immigration, population size may conform to a negative binomial population.

The numbers of pinholes in 50-foot samples of enameled wire have been found by Wise (1946) to conform to a negative binomial distribution. Presumably, the rare event of a pinhole in such a wire has a changing probability from batch to batch. Feller (1957, p. 270) gives an example from records of damage caused by lightning as observed during n time intervals of length t. In this situation, it seems that lightning strikes rarely enough to make a Poisson distribution applicable but the damage caused varies from strike to strike in such a way that the total result is well fitted by a negative binomial distribution.

3.1 Derivation of the Negative Binomial Distribution

The negative binomial distribution can be derived in several ways. One way produces the negative binomial distribution in a waiting time experiment under a binomial sampling situation. The number of trials, n, are usually fixed and the number of events, x, are observed. In contrast, in a negative binomial sampling situation, the number of occurrences of the event, r, are fixed and the number of failures, x, $(x = n - r)$, is the random variable. The total number of trials then is $x + r$. The trials always end with the rth occurrence of a success. For example, when considering the probability that 3 occurrences of a tail (failure) will occur before the occurrence of a head (success) in a coin tossing experiment, the probability is:

$$P\{TTT\} = (\tfrac{1}{2})^3$$

The probability of a success on the fourth toss is:

$$P\{H \text{ on 4th Toss}\} = \tfrac{1}{2}$$

Then:

$$P\{TTTH\} = (\tfrac{1}{2})^3 (\tfrac{1}{2})$$

If $r = 2$ and $x = 3$, then by the binomial probability function,

$$P\{3T, 1H\} = \binom{4}{1} (\tfrac{1}{2})^3 (\tfrac{1}{2})$$

and the probability that the 5th toss would result in a head is 1/2, therefore:

$$P\{3T, 2H\} = \binom{4}{1} (\tfrac{1}{2})^3 (\tfrac{1}{2})^2$$

In General, if

p = probability of a success
q = probability of a failure
r = the number of successes

The probability of r - 1 successes in n - 1 trials is:

$$\binom{n-1}{r-1} p^{r-1} q^{n-r}$$

and the probability of the rth success on the nth trial is:

$$= \binom{n-1}{r-1} p^{r} q^{n-r} \qquad n \geq r$$

$$= \qquad 0 \qquad n < r \qquad [3.1]$$

Another development of the negative binomial distribution follows. Assume that death occurs after the rth attack of a poison. The probability that an individual survives n exposures is the probability that he will be attacked 0, 1, ..., r-1 times during n exposures is:

$$\sum_{j=0}^{r-1} \binom{n}{j} q^{n-j} p^{j}$$

Consider n-1 exposures, death occurs after r attacks the probability of surviving n - 1 exposures is:

$$\sum_{j=0}^{r-1} \binom{n-1}{j} q^{n-1-j} p^{j}$$

The probability of dying during the nth exposure is the probability of surviving n - 1 exposures minus the probability of surviving n exposures:

$$\sum_{j=0}^{r-1} \binom{n-1}{j} q^{n-1-j} p^{j} - \sum_{j=0}^{r-1} \binom{n}{j} q^{n-j} p^{j} \qquad [3.2]$$

And since:

$$\binom{n}{j} = \binom{n-1}{j} + \binom{n-1}{j-1}$$

Equation [3.2] above is equal to:

$$\sum_{j=0}^{r-1} \left\{ \binom{n-1}{j} q^{n-1-j} p^{j} - \left[\binom{n-1}{j} q^{n-j} p^{j} + \binom{n-1}{j-1} q^{n-j} p^{j} \right] \right\}$$

Combining,

$$\sum_{j=0}^{r-1} \left\{ \binom{n-1}{j} \left[q^{n-1-j} p^{j} - q^{n-j} p^{j} \right] - \binom{n-1}{j-1} q^{n-j} p^{j} \right\}$$

Factoring out p^j:

$$\sum_{j=0}^{r-1} \left\{ \binom{n-1}{j} p^{j} \left[q^{n-1-j} - q^{n-j} \right] - \binom{n-1}{j-1} q^{n-j} p^{j} \right\}$$

And,

$$\sum_{j=0}^{r-1} \left\{ \binom{n-1}{j} p^{j} \left[q^{n-j} q^{-1} - q^{n-j} \right] - \binom{n-1}{j-1} q^{n-j} p^{j} \right\}$$

$$= \sum_{j=0}^{r-1} \left\{ \binom{n-1}{j} p^{j} q^{n-j} \left[q^{-1} - 1 \right] - \binom{n-1}{j-1} q^{n-j} p^{j} \right\}$$

Since p = 1 - q and 1/q - 1 = (1 - q)/q = p/q, [3.2] becomes

$$\sum_{j=0}^{r-1} \binom{n-1}{j} p^{j+1} q^{n-j-1} - \binom{n-1}{j-1} q^{n-j} p^{j}$$

Putting successively j = 0, 1, ... , r-1 in the equation, the positive term of each value of j is cancelled by the negative term at the next value of j. Only the positive terms at j = r - 1 remain and the probability of death during the nth exposure reduces to:

$$\binom{n-1}{r-1} p^{r} q^{n-r} \qquad \text{for } n \geq r$$

$$0 \qquad \text{for } n < r \qquad [3.3]$$

This form of the negative binomial distribution is called the Pascal or Polya distribution.

A more specialized form of the negative binomial (which generalizes the Poisson) can be developed by considering a random variable, x, which follows the Poisson distribution but with its parameter regarded as a random variable independently following a particular continuous probability distribution. This approach was developed by Greenwood and Yule (1920). They said that while for one person the probability of 0, 1, 2, ..., accidents may follow a Poisson distribution with parameter λ:

$$f(x) = \frac{e^{-\lambda}\lambda^{x}}{x!} \qquad x = 0, 1, 2, \ldots \tag{3.4}$$

The mean number, λ, of the accidents may vary from person to person. The particular probability distribution proposed by Greenwood and Yule was the Gamma (or Pearson Type III), whose continuous density function is:

$$g(\lambda) = \frac{C^{r}}{\Gamma(r)} e^{-c\lambda} \lambda^{r-1} \tag{3.5}$$

Then the joint density function of x and λ is the product of [3.4] and [3.5]. Thus:

$$h(x,\lambda) = \frac{C^{r}}{(r-1)!} e^{-c\lambda} \lambda^{r-1} \frac{e^{-\lambda}\lambda^{x}}{x!}$$

Since:

$$\Gamma(r) = (r-1)! = \int_{0}^{\infty} Z^{r-1} e^{-Z} \, dZ$$

Therefore:

$$h(x,\lambda) = \frac{C^{r}}{x!(r-1)!} e^{-\lambda(C+1)} \lambda^{x+r-1}$$

The density function of x is obtained by integrating on λ over its range, zero to infinity:

$$g(x) = \frac{C^r}{x!(r-1)!} \int_0^\infty e^{-\lambda(C+1)} \lambda^{x+r-1} d\lambda$$

$$= \frac{C^r}{x!(r-1)!} \int_0^\infty \lambda^{x+r-1} e^{-\lambda(C+1)} d\lambda$$

Let:

$$Z = \lambda(C+1)$$

$$dZ = (C+1)\,d\lambda$$

$$d\lambda = \frac{dZ}{C+1}$$

$$\lambda = \frac{Z}{C+1}$$

Then:

$$g(x) = \frac{C^r}{x!(r-1)!} \int_0^\infty \left(\frac{Z}{C+1}\right)^{x+r-1} e^{-Z} \frac{dZ}{C+1}$$

$$= \frac{C^r}{x!(r-1)!(C+1)^{x+r}} \int_0^\infty Z^{x+r-1} e^{-Z} dZ$$

$$= \frac{(x+r-1)!}{x!(r-1)!} \cdot \frac{C^r}{(C+1)^{x+r}}$$

Since:

$$\Gamma(x+r) = (x+r-1)! = \int_0^\infty Z^{x+r-1} e^{-Z}\, dZ$$

Rearranging:

$$g(x) = \frac{(x+r-1)!}{x!(r-1)!} \left(\frac{C}{C+1}\right)^r \left(\frac{1}{C+1}\right)^x \qquad [3.6]$$

This is another form of the negative binomial distribution. Letting x = n - r and p = [C/(C + 1)], q is 1/(C + 1), the probability of x is:

$$\binom{n-1}{r-1} p^r q^{n-r} \qquad [3.7]$$

[3.7] is identical to [3.1] and [3.3], though now r is not integer.

Frequently, [3.6] is developed with the following parameters:

$$\frac{P'}{q'} = \frac{1}{C+1}$$

$$q' = 1 + P'$$

$$q' = \frac{C+1}{C}$$

p' and q' are not probabilities. Using these parameters, the negative binomial becomes:

$$\binom{n-1}{r-1} q'^{-r} \left(\frac{p'}{q'}\right)^{x} \qquad [3.8]$$

Where q' = 1 + p' and r is positive. T negative binomial gets its name because successive terms in the expansion of:

$$p^{r}(1-q)^{-r}$$

are the probabilities [3.7] and successive terms in the expansion of

$$(q'-p')^{-r}$$

are the probabilities [3.8].

3.2 Parameters of the Negative Binomial Distribution

The parameters of the negative binomial [3.7] are p and r whereas the parameters of [3.8] are p' and r. The Gamma distribution used in the development of the negative binomial distribution has parameters C (scale) and r (shape The Poisson distribution has the parameter, λ .

3.3 Mean and Variance of the Negative Binomial Distribution

The probability generating function for the negative binomial distribution [3.7] is:

$$P(S) = p^{r}(1-qS)^{-r} \qquad [3.9]$$

Then,

$$E(x) = P'(1)$$

$$= \left. \frac{-(p^{r})\left[r(1-qS)^{r-1}(-q)\right]}{(1-qS)^{r}(1-qS)^{r}} \right|_{S=1}$$

$$E(x) = \frac{rq}{p} \qquad [3.10]$$

$$Var(x) = P'(1) + P''(1) - \left[P'(1)\right]^{2}$$

And, since:

$$P''(1) = \left. \frac{-(rp^{r}q)\left[(r+1)(1-qS)^{r}(-q)\right]}{(1-qS)^{r+1}(1-qS)^{r+1}} \right|_{S=1}$$

$$= \frac{rp^{r}q\,(r+1)\,q}{(1-q)^{r+2}}$$

$$= \frac{q^{2}(r^{2}+r)}{p^{2}}$$

And,

$$\left[P'(1) \right]^2 = \frac{r^2 q^2}{p^2}$$

Therefore:

$$Var(x) = \frac{rq}{p} + \frac{r^2 q^2}{p^2} + \frac{rq^2}{p^2} - \frac{r^2 q^2}{p^2}$$

$$Var(x) = \frac{rq}{p^2} \qquad [3.11]$$

For the alternate form of the negative binomial [3.9], the mean and variance are determined by using the generating function:

$$P(s) = (q' - p's)^{-r}$$

Then,

$$E(x) = P'(1)$$

$$P'(s) = -r(q - p's)^{-r-1}(-p')$$

$$P'(1) = rp'$$

$$E(x) = rp' \qquad [3.12]$$

Since $q' = 1 + p'$

And, $$P''(s) = -(r+1)p'r(q'-p's)^{-r-2}(-p')$$

$$P''(1) = (r+1)r(p')^2$$

Therefore,

$$Var(x) = rp' + (r+1)(p')^2 r - r^2(p')^2$$

$$= rp'(1+p')$$

$$Var(x) = rp'q' \qquad [3.13]$$

since $q' = 1 + p'$

3.4 Estimation of the Parameters of the Negative Binomial Distribution

Three methods have been used to estimate parameters of the negative binomial distribution: The Method of Moments, The Maximum Likelihood Method, and the Proportion of Zeros Method.

3.4.1 The Method of Moments

The parameters have been shown to be p and r and the mean and variance were shown to be rq/p and rq/p^2. If the first two moments are estimated from the sample moments, then the ratio of the mean to the variance provides an estimate of p. That is, if the mean of the sample is m and the variance is S^2, then m/S^2 estimates p. Since, m = rq/p and noting that q = 1 - p, an estimate of r is given by r = mp/(1 - p). The efficiency of estimating p and r by this method has been derived by Fisher (1941) and it's reciprocal is given by:

$$\frac{1}{E} = 1 + 2\left\{\frac{1}{3} q \frac{2}{(r+2)} + \frac{1}{4} q^2 \frac{2\cdot 3}{(r+2)(r+3)} + \frac{1}{5} q^3 \frac{2\cdot 3\cdot 4}{(r+2)(r+3)(r+4)} + \ldots\right\}$$

3.4.2 The Method of Maximum Likelihood

The maximum likelihood method of fitting the negative binomial distribution has been discussed by Haldane (1941), Anscombe (1950) and Sichel (1951). Let f_n be the observed frequency of n events and Z the highest value of n observed. The total number of observations, N, is given by:

$$N = \sum_{n=0}^{z} f_n$$

The mean of the sample, m, is given by:

$$\frac{1}{N}\sum_{\eta=0}^{Z} \eta r_\eta$$

The likelihood is given by:

$$\prod_{\eta=0}^{Z} \left[P(\eta) \right]^{f_\eta}$$

And the logarithm of the likelihood, L, is given by:

$$L = \sum_{\eta=0}^{Z} f_\eta \log P(\eta)$$

$$= f_0\, r \log p + \sum_{\eta=1}^{Z} r_\eta \left[\sum_{x=0}^{n-1} \log (r+x) - \log \eta! + r \log p + \eta \log (1-p) \right\}$$

Differentiating with respect to p:

$$\frac{\partial L}{\partial p} = \sum_{\eta=0}^{Z} f_\eta \left\{ \frac{r}{p} - \frac{\eta}{(1-p)} \right\}$$

$$= \frac{Nr}{p} - \frac{\sum_{\eta=0}^{Z} \eta f_\eta}{(1-p)}$$

Equating to zero and substituting m for:

$$\frac{1}{N}\sum_{\eta=0}^{Z} \eta f_\eta$$

and q for 1 - p, gives:

$$m = \frac{rq}{p} \qquad [3.14]$$

The mean of the sample, m, is therefore the maximum likelihood estimator of the distribution mean.

Differentiating L with respect to r, then,

$$\frac{\partial L}{\partial r} = f_0 \log p + \sum_{\eta=1}^{z} f_\eta \left\{ \sum_{x=0}^{n-1} \frac{1}{(x+r)} + \log p \right\}$$

Equating to zero gives:

$$\sum_{\eta=1}^{z} f_\eta \sum_{s=0}^{n-1} \frac{1}{(x+r)} = N \log \frac{1}{p}$$

Eliminating p from the equation by solving [3.14] for p yields:

$$N \log \left(1 + \frac{m}{r}\right) - \frac{f_1 + f_2 + f_3 + \ldots + f_n}{r}$$

$$- \frac{f_2 + f_3 + \ldots + f_n}{(r+1)} \ldots - \frac{f_n}{(r+x-1)} = 0$$

The solution of this equation for other than r = infinity provides an estimate of r. An iterative process must be used by guessing a value for r and iteratively adjusting

r. The estimate obtained by the method of moments is a good initial guess. The use of computers has greatly improved this iterative process. This method of maximum likelihood was used to estimate the parameters of the negative binomial distribution in this research.

3.4.3 The Proportion of Zeros Method

A third method of estimating r is from the proportion of zeros. If f_o is the observed number of zeros, then an estimate of r (Anscombe, 1950) is:

$$\frac{f_o}{N} = \left(1 + \frac{m}{r}\right)^{-r}$$

Taking logarithms of both sides and putting:

$$a = -\frac{\log f_o}{N}$$

yields

$$\frac{a}{r} = \log\left(1 + \frac{m}{r}\right)$$

This equation can also be solved in an iterative manner.

CHAPTER 4

STATISTICAL TECHNIQUES USED IN THE ANALYSIS OF DATA

Since extensive use has been made of certain statistical techniques, this chapter will explain the rationale and subtleties of applying the techniques to the data used in this research.

4.1 Estimating Parameters of the Negative Binomial Distribution

In order to calculate the theoretical frequencies of occurrence of events in the negative binomial distribution, it is necessary to estimate the parameters r and p of the distribution:

$$f(x) = \binom{x-1}{r-1} p^{r} q^{x-r}$$

Using the method of maximum likelihood, it has been shown in Section 3.4 that an estimate of the mean (m) of the distribution can be determined by:

$$m = \frac{rq}{p}$$

In Section 3.3, it was shown that the mean of the negative binomial distribution was given by:

$$E(x) = \frac{rq}{p}$$

and the variance by:

$$Var(x) = \frac{rq}{p^2}$$

By taking the ratio of the mean to the variance of the sample data, one can obtain an estimate for the parameter p. Thus:

$$p = \frac{\text{Mean}}{\text{Variance}} = \frac{\frac{rq}{p}}{\frac{rq}{p^2}}$$

In order to obtain the estimate of the parameter r by the method of maximum likelihood an iterative procedure must be used. In Section 3.4, it was shown that the parameter r could be determined from the equation:

$$N \operatorname{Log}\left(1+\frac{m}{r}\right) - \frac{f_1+f_2+f_3+\ldots+f_n}{r} - \frac{f_2+f_3+\ldots+f_n}{(r+1)} - \ldots - \frac{f_n}{(r+x+1)} = 0 \qquad [4.1]$$

Where:

r = the parameter
N = the sample size
f_i = the frequency of occurrence of i
m = the sample mean

The procedure, then, is to guess at an initial value of r and solve the above equation, then successively alter r until the value of the left hand side of the above equation becomes zero. It is sometimes judicious to use the value of r estimated using the method of moments (Section 3.4). An illustration of the procedure of estimating p and r is given below for the following book use data.

No. of Uses	Frequency
0	85
1	21
2	12
3	3
4	4
5	7

The mean of the above data is .79545 and the variance is 1.91967. By taking the ratio of the mean to the variance, it is found that the parameter p = .41437.

Using r = .400, the value of the left hand side of the equation is 2.21559. The estimate of r is then increased by .001 until the value of the left hand side of equation 4.1 becomes negative. Table 4.1 shows the tabular values of the numeric value of the left hand side of the equation.

Since the numeric value of the left hand side of the equation at $r = .427$ is $.00841$ and the numeric value is $-.06435$ for $r = .428$, we could use linear interpolation to determine the r for which the le hand side of the equation is zero. This method was used in all analyses in which the parameters of the negative binomial distribution had to be estimated.

4.2 Using the Chi Squared Test of Goodness of Fit

The Chi Squared test is the oldest goodness of fit test. A random sample of size n is obtained from the population and f_i denotes the frequency of x_i in the sample. If the population is discrete and has the probability function, $p\{x_i\} = P\{X = x_i\}$, the expected value, e_i equals $np\{x_i\}$; f_i will not usually be equal to $np\{x_i\}$ and the differences $f_i - np\{x_i\}$ are related to the extent of departure of the actual population probabilities from $p\{x_i\}$

TABLE 4.1

TABULATION OF VALUES USED TO ILLUSTRATE THE CALCULATION OF r

r	numerical value
.400	2.21559
.401	2.12479
.403	2.03475
.	.
.	.
.	.
.425	0.15573
.426	0.08177
.427	0.00841
.428	-.06435
.429	-.13653

The differences are squared, weighted by the reciprocal of the theoretical frequencies, $np\{x_i\}$, and summed yielding:

$$\chi_s^2 = \sum_{i=1}^{k} \frac{(f_i - np[x_i])^2}{np[x_i]}$$

This statistic has approximately the Chi Squared distribution with k-j-1 degrees of freedom. The value of j being the number of parameters estimated from the sample data.

For a test based on this statistic, we need to be able to calculate the probability of obtaining a value of the statistic as large as, or larger than, the observed value.

To illustrate the use of this test, it is hypothesized that the data shown in Section 4.1 is negative binomially distributed. The tabulation of the components of the negative binomial distribution with parameters p = .41437 and r = .427 are shown in Table 4.2.

The Chi Squared statistic is 10.9543.

TABLE 4.2

CHI SQUARED TABLE USED FOR ILLUSTRATING
CHI SQUARED TEST CALCULATIONS

x_i	$f_i\{x_i\}$	$p\{x_i\}$	$np\{x_i\}$	$(f_i\{x_i\} - np\{x_i\})$	$\frac{(f_i\{x_i\} - np\{x_i\})^2}{np\{x_i\}}$
0	85	.68647	90.614	-5.6140	.3478
1	21	.17166	22.659	-1.6591	.1215
2	12	.07173	9.468	2.5316	.6769
3	3	.03398	4.485	-1.4853	.4919
4	4	.01705	2.251	1.7494	1.3598
5	7	.01911	2.522	4.4788	7.9564
	132	1.00000	132.000	.0000	10.9543

The probability of obtaining a value as large or larger than the observed statistic is .013 (from a Chi Squared table with 3 degrees of freedom). We conclude then that the data are not well fitted to a negative binomial distribution.

In order to prove the quantity Chi Squared $= \sum_{i=1}^{k} (f_i - np\{x_i\})^2 / np\{x_i\}$ is distributed as Chi Squared when the null hypothesis is true, it is necessary that the $np\{x_i\}$'s be large. That is, the proof is strictly valid when the $np\{x_i\}$'s tend to infinity in the limit. Because of this many authors recommend that the $np\{x_i\}$'s be greater than or equal to 5. Some authors recommend that the $np\{x_i\}$'s be greater than 10. The point is controversial, however, and Cochran (1954) using m_i as the expected frequency states that:

> "It is my opinion that these recommendations are too conservative and may on occasion result in a substantial loss of power in the test, I give this as an opinion, because not enough research has been done to make the situation quite clear. However the exact distribution of the $\Sigma(f_i - m_i)^2/m_i$ when the expectations are small, has been worked out in a number of particular cases by Sukhatme (12), Neyman and Pearson (3) and Cochran (4), (5). These

results indicate that the χ^2 tables give an adequate approximation to the exact distribution even when some m_i are lower than 5."

The loss of power from the greater or equal to five rule occurs because the rule requires grouping of frequencies in cells at the tails of the distribution where frequently the differences between the alternative hypothesis and null hypothesis stand out most clearly, so that grouping would cover up the most serious differences between the two distributions. Cochran (1954) uses the following illustration. Suppose that we have a sample of N = 100. The null hypothesis is that the data follow a Poisson distribution with known mean equal to 1, when in fact the data follows a negative binomial distribution:

$$(q - p)^{-n}$$

where n = 2, q = 1.5, and p = 0.5. What is the chance of rejecting the null hypothesis at the 5% level of significance? Cochran presents the data in Table 4.3.

If an m_i as low as 1.00 is allowed, 5 cells are used in the Chi Squared test with 4 degrees of freedom since

TABLE 4.3

TABLE USED TO ILLUSTRATE POWER OF CHI SQUARED GOODNESS OF FIT TESTS

	Expected Frequencies		$\frac{(m_i - m_i')^2}{m_i}$		
	Poisson	Neg.Bin	Ungrouped	Grouped	
i	m_i	m_i'		$m_i \geq 5$	$m_i \geq 10$
0	36.79	44.44	1.590	1.590	1.590
1	36.79	29.63	1.393	1.393	1.393
2	18.39	14.82	0.693	0.693 }	0.009
3	6.13	6.58	0.033 }	1.181 }	
4+	1.90	4.53	3.640 }		
			7.349	4.857	2.992

the mean is known. To make all m_i greater than 5, the last two classes are pooled and there are 3 degrees of freedom. To make all the m_i's greater than 10, the last 2 classes are pooled and there are two degrees of freedom. Cochran calculates the values of the power function:

$$\lambda = \sum \frac{(m_i' - m_i)^2}{m_i}$$

which is a non-central Chi Squared distribution with parameter, λ, of non-centrality. The larger the value of λ, the higher the power. The contributions to λ is shown in the fourth column for ungrouped data; 5th column for grouping the frequencies of the last two cells; and, the last column for grouping the frequencies of the last three cells. The probabilities of rejecting the null hypothesis from non-central Chi Squared distribution tables are .56 for the ungrouped data, .43 for an m_i greater than or equal to 5, and .32 for an m_i greater than or equal to 10. Cochran points out that the power function slightly favors the ungrouped data because it is an asymptotic result. Cochran then calculates the sample N that would be needed in the grouped cases in order to have the same power as in the ungrouped case. He found that an N of 136 would be required when the frequencies of the last two cells are grouped, and an N of 191 when the last three are grouped. These are to be compared against an N = 100 for the ungrouped data.

Cochran gives the following rule for goodness of fit tests for unimodal distributions:

> "Here the expectations will be small only at one or both tails. Group so that the minimum expectation at each tail is at least 1."

For all uses of the Chi Squared goodness of fit test in this research, grouping was kept at a minimum because of the few number of cells and because of the loss of two degrees of freedom for estimating the parameters of the Negative Binomial distribution. Cochran's rule above was followed in all cases.

4.3 <u>The Variance Test</u>

One of the ways that data distributed according to a negative binomial distribution is distinguished from the Poisson is by the magnitude of the variance. The Poisson distribution has a variance equal to the mean, whereas the negative binomial distribution always has a variance greater than the mean. These are contrasted with the binomial distribution which has a variance which is always less than the mean. Rejecting the null hypothesis that the variance is equal to the mean will permit one to conclude at a specified significance level that there is no reason to believe that the data is Poisson distributed. Cochran (1954) states that a comparison between the observed variance of the observations x_i and the variance predicted from Poisson theory will frequently be more sensitive than the goodness of fit test. The variance test is made by calculating

$$\chi_v^2 = \sum_{i=1}^{N} \frac{(x_i - \bar{x})^2}{\bar{x}}$$

The quantity Chi Squared is referred to the Chi Squared tables with N-1 degrees of freedom.

To illustrate the use of the variance test, the data in Section 4.1 will be used. The mean of the sample is .79545 and the variance is 1.91967. An alternative for

Chi Squared, if the calculation is made from the frequency distribution of x_j, is:

$$\chi_v^2 = \sum \frac{f_j (j - \bar{x})^2}{\bar{x}}$$

Where

j = the cell
$\bar{x}$ = the mean of the sample
f_j = the frequency of occurrence of j

The calculations for the data of Section 4.3 are tabulated in Table 4.4. The calculated Chi Squared is 316.14452. The value of Chi Squared can be determined using the normal approximation to the Chi Squared distribution:

$$\chi_p^2 = \frac{1}{2} \left(Z_p + \sqrt{2k-1} \right)^2$$

Where:

p = 1 - significance level
Z_p = the standard normal value at probability p
k = sample size

TABLE 4.4

VARIANCE TEST TABLE

j	f_j	$(j-\bar{x})$	$(j-\bar{x})$	$f_j\,(j-\bar{x})$	$f_j\,(j-\bar{x})^2/\bar{x}$
0	85	-.79545	.63274	53.78290	67.61317
1	21	.20455	.04184	.87864	1.10458
2	12	1.20455	1.45094	17.41128	21.88859
3	3	2.20455	4.86004	14.58012	18.32939
4	4	3.20455	10.26914	41.07656	51.64939
5	7	4.20455	17.67824	123.74768	155.56940

Chi Squared = 316.14452

At a significance level of 5%, Z is 1.65, k is 132, and the value of Chi Squared is calculated to be 159.31. Thus, we would conclude that the hypothesis should not be accepted, that the variance is greater than the mean and thereby assume that the data is negative binomially distributed.

CHAPTER 5

THE COLLECTION OF BOOK DEMAND AND USE DATA

Several types of demand occur for books in a library. This investigator identifies two major kinds of demand: known and unknown. A known demand is a demand by a patron which the library system knows has been made. For example, if a person charges out a book, the library system knows a demand has been made. The particular book has been both demanded and used. An unknown demand occurs when the library system does not know that a demand has been made. For example, a patron using the card catalog has identified that the book he wants is in the collection, looks for the book and doesn't find it. He does not check to see if the book is charged out nor does he ask that the book be found or held for him. Previous investigators have studied only known demand.

5.1 Types of Demand

Libraries generally are of two types: closed stack or open stack. In a closed stack library (one in which a patron has no access to the library stacks), a patron must identify his demand to a library employee in some manner. Generally, he does this by filling out some sort of form, for example, a book call slip. In this type library, all demands are known. The demands could be satisfied or unsatisfied. The satisfied demand can result in a home use in which the patron physically removes the book from the library for a period of time or an in-house use in which the patron uses the book in the library. This type of library can also experience unsatisfied book demands. Thus, the library has a record of both known and unknown satisfied or unsatisfied demands.

An open stack library, on the other hand, permits the patrons access to the book stacks of the library. This type library can also experience both known and unknown demands. Both types of demands can be satisfied or unsatisfied. However, the open stack library generally only has a record of satisfied home use demands. The library system may never become aware of unsatisfied demands or of satisfied in-house demands.

Studies of book use previously conducted always consider known home use demands, sometimes consider known in-house use demands but never consider unsatisfied demands.

5.2 Methods Used For Collecting Demand Data

Previous use studies generally involve random samples or samples of the books from selected subsets of the entire collection. Two methods for collecting data will be described using the terminology used by Jain (1967).

A "check out sample" is one in which the past use (generally satisfied home use demands) of books "checked out" during a specified period of time are recorded. Dawson, Aldrin and Gould (1961) used this type of sample.

The second method is called a "collection sample." This method involves choosing a sample of the collection of books in the library or of the collection of the selected subsets of book in the library and collecting information about past or present use of the books. Fussler and Simon (1961) used this type of approach.

All previous investigators, with the exception of Jain (1967) who used his "relative use" measure previously described in Section 2.5, used either the "collection method" or the "check out" method for the collection of their use data. This investigator collected random samples of book demand using the "collection method."

5.3 Problems and Difficulties in Collecting Demand Data

Certain problems arise with the collection of all types of book demand data caused by the type of library, the type of records which are kept and by the method of collecting data. With regard to the method of collecting data, Jain (1967) examined the advantages and disadvantages of the "collection method" and "check out" method and showed the following advantages and disadvantages of each:

Characteristics	Collection Sample	Check-Out Sample
Can one draw inferences about the whole collection?	yes	no

Characteristics	Collection Sample	Check-Out Sample
Can information be obtained on the rate of usage of the same group of books over a long period of time?	yes	no
Is it relatively easy to design a sampling scheme and collect data?	no	yes
Are the problems of missing data and of lack of control on the methods of recording usage histories in the past avoided?	no	yes

The collection method is the only way that inferences can be made about the entire collection or subsets of the collection.

There are problems caused by the type of library and the method in which book usage is recorded. Theoretically in a closed stack library, all demand is known. If the library keeps the records of all demands, demand studies would be no problem. This is not the case. Records generally are not kept. Since the data collected for this study were collected in an open stack library, the fourth characteristic shown above will now be considered. Book use records are generally available only for home-use demands. A book card or some record is generally affixed to the inside back cover of the book and a patron on checking out a book will sign his name or a date charged or date due is written or stamped into the book. Theoretically, this record of home use should provide for a good record of past home use of the book. However, the amount of space provided for names or dates is finite. The cards are generally thrown out when they are "filled up" with signatures or the "date due" slips are removed when they are "filled up" with dates.

When looking at these usage records for long periods of usage, it is sometimes difficult to determine whether the first date stamped is the first home use of the book or is the first home use on a new date due slip. That is, a date due slip had already been filled up with dates, removed and discarded. Other problems sometimes arise such as attempting to determine the date at which a book was added to the collection. For example, a book with a date due slip inside its back cover with an imprint date of

1910 and with no record of home use on the date due slip may be: (1) a book which was added to the collection in 1910 with no home use since that time; (2) A book added in 1910 with a thousand uses until 1968 at which time a new date due slip was placed in the book; (3) A book which was added to the collection in 1968 at which time a date due slip was placed into the book.

5.4 Identification of Books to be Included in Random Samples

The books identified for the analyses performed during this investigation were identified using a random sampling scheme. A random sample of the entire State University of New York at Buffalo library collection was collected from the library's shelf list in order to study the number of home uses of books during a period of time, and a random sample of Library of Congress P class (Language and Literature) books were identified to be studied to determine the number of demands for the books.

The population sampled was the library's entire book collection. This was performed by randomly identifying the books by using the library's shelf list. The shelf list consists of approximately 400 drawers (each approximately 16" in length) of 3 x 5 cards - each containing catalog card information about one title. The shelf list is arranged in call number order so that, theoretically, the cards are arranged in the same manner as the books are on the shelves.

In order to collect a random sample from the shelf list, it was decided to randomly identify an eighth of an inch of the total thickness of cards comprising the shelf list, pull out a card from this eighth of an inch and record the appropriate identifying information.

To select the random samples, the drawers were numbered, the depths of the drawers measured and a computer program written to generate as many random numbers as cards to be included in the sample. The program then converted the random number into a drawer number and a depth (in inches) into the drawer at which a card was to be pulled. A portion of the output of the program is shown in Table 5.1. The first line in the Table identifies a card in the 50th drawer at 4/8 of an inch into the drawer. The

TABLE 5.1

A PORTION OF THE RANDOMLY GENERATED SHELF LIST DEPTHS FOR RANDOM TITLE IDENTIFICATION

Drawer Number	Inches
50	0 4/8
51	1 4/8
51	9 3/8
53	5 3/8
53	7 7/8
53	14 7/8
54	11 1/8
54	12 7/8
55	1 5/8
55	2 3/8
55	5 7/8
55	12 1/8
55	14 1/8

second line identifies a book in drawer 51 at 1 4/8 inches into the drawer.

Two hundred random shelf list depths were generated at a time. For example, if 1000 random depths were desired, five sets of 200 random depths were generated. Each 200 locations were randomly generated over the entire shelf list depth.

This was done because shelf list drawers are, necessarily, not full and prior to selecting the desired cards, it was necessary to compress the cards toward the front of the drawer. In all drawers, then, the back portion of the drawer would contain no cards and the randomly generated location may be empty. Two hundred (200) random locations might only result in 78 cards or 114 cards, etc.

Upon identification of a card, the call number was recorded as well as some other identifying information, such as number of copies, location (branch library, etc.) and a short title. This information was then used to retrieve the book for the book demand data to be used in the analyses.

5.5 Samples of Book Demand and Uses

This investigator collected demand for known home use of books for a sample of 702 books from the library's entire collection for the period 9/1/66 - 8/31/67. In addition, all demands (known and unknown) were recorded for the P class (Language and Literature) books using a questionnaire method for the period March 3, 1968 through May 11, 1968. During this period, 12,898 demands were recorded for 10,422 different books. The P class shelf list was then sampled and 732 books randomly selected for analysis.

5.5.1 The Collection of Home-Use Book Demand Data For The SUNY Buffalo Collection

Ten sets of 200 random depths were generated in order to identify call numbers of titles from the entire SUNY Buffalo shelf list. The procedure described in 5.4 was followed. Because of empty space in shelf list drawers, 1167 titles were identified. A search was made in the stacks and: (1) if the book was found, the number of charges (home use demand) as recorded in the back of the book was copied; (2) if the book was not on the shelf, the book was recalled if it were charged out to another patron and the number of charges as recorded in the back of the book was copied upon its return; (3) if the book was not on the shelf and not charged out, a search for the book was made one week later and the number of charges recorded if it were found. If it was not found upon a second trial, it was assumed to be a missing book.

A tabulation of the number of books identified by Library are shown in Table 5.2. For example, 48 titles were searched for the Health Sciences Library, 44 were found, 3 were either not found or not returned by the borrower upon recall, 1 did not circulate.

The 1020 books which were found were checked to see if they were charged out for home use during the period 9/1/66 through 8/31/67. Because of the missing data or uncertain data problem described in Section 5.3, 318 books were eliminated from consideration. The investigator was sure that a book was on the shelf available to patrons if: (1) a charge date was recorded on the date due slip earlier than 9/1/66; (2) the book contained a University of Buffalo

TABLE 5.2

NUMBER OF BOOKS RANDOMLY SELECTED FOR STUDY IN EACH OF THE SUNY-BUFFALO LIBRARIES

Library	Number Found	Number Not Found	Number Not In Circulation	Total
Lockwood	791	71	25	887
Art and Music	37	11	1	49
Health Sciences	44	3	1	48
Ridge Lea	22	3	1	26
Harriman	-	-	9	9
Science & Engineering	101	14	3	118
Chemistry and Physics	25	5	-	30
Total	1020	107	40	1167

stamp (this means it was cataloged prior to 1963); (3) an identifying book plate was in the book (book plates were not put in books later than 9/1/66). The investigator was sure that the book was not on the shelf during 9/1/66 through 8/31/67 if the imprint date was later than 1967 (for example, 1968), and was not sure if the book was on the shelf for the specified period if none of the above described conditions were appropriate.

A tabulation of the number of home use demands for the period by sample is shown in Table 5.3. For example, sample 6 showed that 52 books were not charged out during the period 9/1/66 through 8/31/67, 14 were charged out once, 7 were charged out twice and one book was charged out six times.

5.6 Collection of Book Demand Data for Language and Literature (P Class) Books

The Language and Literature collection at the SUNY/Buffalo Libraries comprises the largest class of books within the library system (perhaps one quarter) of the 450,000 title collection. In addition, the entire class

TABLE 5.3

HOME USE DEMANDS OF TEN RANDOM SAMPLES
OF BOOKS DURING THE PERIOD 9/1/66 - 8/31/67

Sample Number	Number of Home Use Demands								Total
	0	1	2	3	4	5	6	7	
1	41	9	2	0	0	0	0	0	52
2	40	8	2	1	0	0	0	0	51
3	54	12	6	1	2	0	0	0	75
4	61	11	5	2	0	0	0	1	80
5	59	12	4	0	2	0	0	0	77
6	52	14	7	0	0	0	1	0	74
7	55	14	5	0	0	0	2	0	76
8	53	6	3	2	0	0	0	0	64
9	54	13	3	2	0	1	0	0	73
10	61	13	5	1	0	0	0	0	80
Totals	530	112	42	9	4	1	3	1	702

(with the exception of fiction books) are shelved on one level of Lockwood Library. The level has one stack entrance. There is an emergency exit which is not to be used by patrons. In addition, there are essentially no reader spaces on the level so that patrons wishing to use P Class books must physically remove the books from the stack level. Because of these reasons, the P Class books were selected for intensive demand study for this investigation. During the time of the data collection (Spring 1968), Lockwood Library was open 96 hours per week. (From 7:30 a.m. to 11:00 p.m. on Mondays through Fridays; from 7:30 a.m. to 5:00 p.m. on Saturdays; and 1:00 p.m. to 11:00 p.m. on Sundays). It was decided that the best way to get demand data (both known and unknown) was to ask patrons what they were looking for. In order to accomplish this, data collectors were stationed at the stack entrance. As each patron entered the stack entrance, they were given a questionnaire (See Figure 5.1 and Appendix A). The questionnaire was designed to find out what the patron looked for, what he found, and whether he would use what he found in the library or charge it out for home use. Data collection was begun on February 13, 1968 and continued through May 11, 1968. Data was collected on each day except

FIGURE 5.1

QUESTIONNAIRE

Date________________ Time________________

THE RESULTS OF THIS QUESTIONNAIRE WILL HELP THE LIBRARY IMPROVE ITS SERVICE. YOUR COOPERATION WILL BE APPRECIATED.

For each book which you looked for, found or used, list the complete call number on the left and circle your answer to each of the three questions on the right. Please circle the answer to the question even if you did not find the book you wanted.

Call Number	Question	Answer
________ ________	1) Did you have the call number when you came into the stacks?	YES NO
________ ________	2) If you had the call number, did you find the book you looked for?	YES NO
________ ________	3) If you found (or would have found) the book, will you (would you)?	Use it in the Library Charge it Out Not sure if I will(would) charge it out or not

Call Number	Question	Answer
________ ________	1) Did you have the call number when you came into the stacks?	YES NO
________ ________	2) If you had the call number, did you find the book you looked for.	YES NO
________ ________	3) If you found (or would have found) the book, will you (would you)?	Use it in the Library Charge it Out Not sure if I will(would) charge it out or not

Call Number	Question	Answer
________ ________	1) Did you have the call number when you came into the stacks?	YES NO
________ ________	2) If you had the call number, did you find the book you looked for?	YES NO
________ ________	3) If you found (or would have found) the book, will you (would you)?	Use it in the Library Charge it Out Not sure if I will(would) charge it out or not

Call Number	Question	Answer
________ ________	1) Did you have the call number, when you came into the stacks?	YES NO
________ ________	2) If you had the call number, did your find the book you looked for?	YES NO
________ ________	3) If you found (or would have found) the book, will you (would you)?	Use it in the Library Charge it Out Not sure if I will(would) charge it out or not

March 26, 1968 (Easter Sunday) when the library was closed. The questionnaire was checked for a period of 18 days. At that time, it was decided to ask the patrons for their student number if they were a student or their name if they were a faculty member. It was discovered at that time that books which were not recorded on the questionnaires were being charged out. In addition, call numbers were being imporperly recorded and coded by the data collectors. Detailed instructions were prepared and given to the data collectors (Appendix B). Each demand was recorded, coded and keypunched. The punched card format is shown in Appendix C. All data collected beyond the eighteen-day period were used for the analyses in this research.

In addition to the demand data collected on the questionnaires at the stack level, the charge cards for P Class books were checked daily to see which P Class books were charged out. The names or numbers of the person charging the book were compared with the questionnaires to determine if any individuals were not completely filling out a questionnaire. The data collection procedure was completed on May 11, 1969 and produced 51 days worth of demand information. Seven thousand eight hundred ninety-three (7,893) questionnaires were distributed during the 51-day period.

P Class stack level is shown in Table 5.4. The table shows: the day of the study; the number of questionnaires distributed that day (this number includes a questionnaire for each patron who refused to fill out a questionnaire); the number of questionnaires not returned; the number that were useable (some questionnaires did not include P class books or contained unintelligible information and were thus not useable); the number of questionnaires which contained no P's; the number of questionnaires which were refused; the number of questionnaires for patrons who were browsing but who found nothing; the total number of demands for the day; the total number of recorded demands for the day; the total number of unrecorded demands (these were for books charged out but which were not recorded on questionnaires and are traceable either to those refusing to fill out a questionnaire or unintelligibly filling out a questionnaire); and the number of unrecorded traced demands (these are for books charged out which were traceable to particular questionnaires but which were not recorded on the questionnaires). The number of unrecorded demands appears to be high, ranging from 12% to 53% of the total demands (averaging 33%). The total demands then are somewhat deflated and the number of demands would be

TABLE 5.4

TABULATION OF DATA FROM MARCH 3,1968 THRU MAY 11,1968

DAY	NUMB. OF QUEST.	QUEST. NOT RETURNED		USABLE QUEST.		NO P's QUEST.		UNUSABLE QUEST.		REFUSED		BROWSERS		TOTAL DEMANDS	RECORDED DEMAND		UNRECORDED DEMAND		TRACED UNREC. DEMAND	
	NO.	NO.	PC	NO.	PC	NO.	PC	NO.	PC	NO.	PC	NO.	PC	NO.	NO.	PC	NO.	PC	NO.	PC
19	168	7	4.2	74	44.0	50	29.8	11	6.5	10	6.0	16	9.5	289	196	66.4	86	30.7	8	2.9
20	169	24	14.2	67	39.6	40	23.7	9	5.3	21	12.4	8	4.7	286	174	60.8	92	32.2	20	7.0
21	112	14	12.5	50	44.6	29	25.9	8	7.1	7	6.3	5	4.5	158	98	62.0	59	37.3	1	.6
22	107	8	7.5	56	52.3	21	19.6	3	2.8	13	12.1	6	5.6	227	131	57.7	96	42.3	0	0.0
23	171	30	17.5	79	46.2	29	17.0	13	7.6	13	7.6	7	4.1	283	150	53.0	128	45.2	5	1.8
24	175	16	9.1	84	48.0	54	30.9	4	2.3	7	4.0	10	5.7	258	174	67.4	81	31.4	3	1.2
25	147	12	8.2	70	47.6	45	30.6	6	4.1	10	6.8	4	2.7	223	138	61.9	82	36.8	3	1.3
26	130	11	8.5	63	48.5	32	24.6	8	6.2	7	5.4	9	6.9	247	129	52.2	114	46.2	4	1.6
27	160	21	13.1	85	53.1	27	16.9	5	3.1	11	6.9	11	6.9	250	191	76.4	59	23.6	0	0.0
28	62	12	19.4	29	46.8	17	27.4	1	1.6	3	4.8	0	0.0	123	76	61.8	47	38.2	0	0.0
29	43	4	9.3	22	45.9	17	35.4	3	6.3	1	2.1	1	2.1	116	54	46.6	62	53.4	0	0.0
30	103	11	10.7	35	34.0	40	38.8	7	6.8	7	6.8	3	2.9	157	86	54.8	71	45.2	0	0.0
31	151	13	8.6	62	41.1	50	33.1	9	6.0	14	9.3	3	2.0	232	158	68.1	59	25.4	15	6.5
32	120	6	5.0	54	45.0	38	31.7	6	5.0	16	13.3	0	0.0	158	129	81.6	19	12.0	10	5.3
33	97	5	5.2	45	46.4	32	3	3	3.1	10	10.3	2	2.1	161	111	68.9	23	14.3	27	16.8
34	95	6	6.3	38	40.0	33	34.7	7	7.4	9	9.5	2	2.1	143	94	65.7	23	16.1	26	18.2
35	91	1	1.1	55	60.4	24	26.4	1	1.1	10	11.0	0	0.0	126	108	85.7	17	13.5	1	.8
36	85	2	2.4	50	58.8	17	20.0	4	4.7	8	9.4	4	4.7	150	98	65.3	32	21.3	20	13.3
37	177	11	6.2	101	57.1	33	18.6	12	6.8	8	4.5	12	6.8	268	175	65.3	81	30.2	12	4.5
38	127	15	11.8	63	49.6	32	25.2	10	7.9	4	3.1	3	2.4	162	115	71.0	46	28.4	1	.6
39	186	13	7.0	96	51.6	55	29.6	3	1.6	9	4.8	10	5.4	322	200	62.1	88	27.3	34	10.6
40	152	14	9.2	64	42.1	45	29.6	11	7.2	8	5.3	10	6.6	289	142	49.1	119	41.2	28	9.7
41	171	26	15.2	58	33.9	60	35.1	9	5.3	13	7.6	4	2.3	265	135	50.9	111	41.9	19	7.2
42	90	4	4.4	50	55.6	20	22.2	4	4.4	10	11.1	2	2.2	175	112	64.0	53	30.3	10	5.7
44	251	23	9.2	129	51.4	54	21.5	22	8.8	13	5.2	10	4.0	421	258	61.3	118	28.0	45	10.7
45	219	35	16.0	95	43.4	57	26.0	16	7.3	14	6.4	2	.9	370	238	62.8	127	33.5	14	3.7

TABLE 5.4 (CONTINUED)

TABULATION OF DATA FROM MARCH 3,1968 THRU MAY 11,1968...CONTD.

DAY	NUMB. OF QUEST.	QUEST. NOT RETURNED		USABLE QUEST.		NO P#S QUEST.		UNUSABLE QUEST.		REFUSED		SPONSERS		TOTAL DEMANDS	RECORDED DEMAND		UNRECORDED DEMAND		TRACED UNREC. DEMAND	
	NO.	NO.	PC	NO.	PC	NO.	PC	NO.	PC	NO.	PC	NO.	PC	NO.	NO.	PC	NO.	PC	NO.	PC
46	243	17	7.0	121	49.8	54	22.2	18	7.4	16	6.6	17	7.0	416	238	57.2	155	37.3	23	5.5
47	166	12	7.2	75	45.2	49	29.5	13	7.8	12	7.2	5	3.0	286	169	59.1	117	40.9	0	0.0
48	222	21	9.5	101	45.5	57	25.7	11	5.0	22	9.9	10	4.5	311	208	66.9	103	33.1	0	0.0
49	101	11	10.9	55	54.5	27	26.7	3	3.0	5	5.0	0	0.0	187	143	76.5	25	13.4	19	10.2
50	108	8	7.4	53	49.1	30	27.8	6	5.6	9	8.3	2	1.9	162	115	71.0	47	29.0	0	0.0
51	193	12	6.0	93	46.7	58	29.1	8	4.0	21	10.6	7	3.5	396	204	51.5	171	43.2	21	5.3
52	227	32	14.1	110	48.5	48	21.1	13	5.7	16	7.0	8	3.5	431	250	58.0	122	28.3	59	13.7
53	222	22	9.9	84	37.8	69	31.1	14	6.3	25	11.3	8	3.6	302	168	55.6	127	42.1	7	2.3
54	176	14	8.0	80	45.5	46	26.1	14	8.0	15	8.5	7	4.0	290	176	60.7	104	35.9	10	3.4
55	172	12	7.0	79	45.9	46	26.7	9	5.2	25	14.5	1	.6	236	135	57.2	98	41.5	3	1.3
56	125	18	14.4	62	49.6	36	28.8	4	3.2	5	4.0	0	0.0	258	142	55.0	96	37.2	20	7.8
57	128	14	10.9	67	52.3	28	21.9	2	1.6	7	5.5	10	7.8	197	125	63.5	57	34.0	5	2.5
58	184	11	6.0	85	46.2	55	29.9	8	4.3	15	8.2	10	5.4	334	197	59.0	119	35.6	18	5.4
59	236	30	12.7	97	41.1	62	26.3	21	8.9	18	7.6	8	3.4	336	191	56.8	114	33.9	31	9.2
60	224	18	8.0	104	46.4	57	25.4	13	5.8	27	12.1	5	2.2	350	212	60.6	110	31.4	28	8.0
61	175	21	12.0	71	40.6	39	22.3	14	8.0	21	12.0	9	5.1	295	141	47.8	122	41.4	32	10.8
62	149	9	6.0	62	41.6	38	25.5	19	12.8	18	12.1	3	2.0	198	142	71.7	27	13.6	29	14.6
63	100	21	21.0	36	36.0	29	29.0	11	11.0	2	2.0	1	1.0	120	72	60.0	37	30.8	11	9.2
64	119	14	11.8	57	47.9	35	29.4	3	2.5	6	5.0	4	3.4	176	118	67.0	43	24.4	15	8.5
65	219	34	15.5	78	35.6	63	28.8	8	3.7	23	10.5	13	5.9	341	206	60.4	126	37.0	9	2.5
66	208	26	12.5	105	50.5	41	19.7	11	5.3	20	9.6	4	1.9	431	192	44.5	197	45.7	42	9.7
67	172	21	12.2	81	47.1	38	22.1	5	2.9	22	12.8	5	2.9	304	176	57.9	107	35.2	21	6.9
68	157	28	17.8	66	42.0	25	15.9	16	10.2	17	10.8	5	3.2	255	127	49.8	100	39.2	28	11.0
69	154	12	7.8	76	49.4	29	18.8	17	11.0	13	8.4	7	4.5	238	135	56.7	85	35.7	18	7.6
70	91	23	25.6	41	45.5	19	21.1	7	7.8	0	0.0	0	0.0	189	117	61.9	66	34.9	6	3.2

increased if it were possible to get patrons to cooperate more with questionnaire surveys. On the other hand, the probability of a person having an unsatisfied or satisfied in-house demand given that he had a home use demand was calculated to be approximately .2, which is not high enough to affect the curve fitting analysis which will be performed in Section 6.2 especially since there is no reason to believe that unrecorded demands occurred in any systematic manner. It is reasonable to assume that the unrecorded demands occurred in a random manner. This means that although the mean number of demands and actual frequencies of demands may be understated, the analysis performed should be reasonably valid. A summary of the totals for the 51-day period are shown in Table 5.5.

5.6.1 <u>Records of P Class Demand Data</u>

After keypunching and verification, the demand data was entered onto magnetic tape and maintained in day-time and call number order. A page listing from this file is shown in Figure 5.2. The punched card format and master file tape record format are shown in Appendix C.

TABLE 5.5

QUESTIONNAIRE SUMMARY OF P CLASS DATA

Questionnaire Status	Number	Percent
Not Returned	805	10.27
Useable	3,613	46.08
Without P's	2,029	25.88
Not Useable	463	5.91
Refused	636	8.11
Browsing	293	3.75
Total	7,893	100.00

Demand Status	Number	Percent
Recorded	7,759	60.16
Unrecorded	4,378	33.94
Traced	761	5.90
Total	12,898	100.00

FIGURE 5.2

SAMPLE PAGE FROM MASTER FILE LISTING

PA	4440PA6756	S17	V1	1912	121427	491411	21211	044	866
PA	4441PA6756	S17	V2	1912	121427	491411	21211	044	867
PA	4442PA676	G154				381337	12321	075	868
PA	4443PA6766	07			121427	201646	0		869
PA	4444PA6787	A2		1924		270932	21211	010	870
PA	4445PA6787	A2F+		1966	142923	69	9		871
PA	4446PA6787	A5		1963	121427	201646	11211	130	872
PA	4447PA6791	V4HA			121427	601658	21211	170	873
PA	4448PA6791	V5575		1965	121427	201646	11211	130	874
PA	4449PA6792	A3GA			121427	591246	11211	041	875
PA	4450PA6798	A2			142814	351224	12221	029	876
PA	4451PA6801	A107				551157	21211	038	877
PA	4452PA6801	C6			147174	40	9		878
PA	4453PA6802	A1F14	V1		140574	231404	11211	091	879
PA	4454PA6803	B24P4			147512	571520	12221	045	880
PA	4455PA6804	A2		1944		20	7		881
PA	4456PA6804	A2		1944	114886	20	8		881
PA	4457PA6804	B7B55				501402	12121	026	882
PA	4458PA6806	A3W4			140920	25	9		883
PA	4459PA6806	A3W4			140910	451236	11211	052	883
PA	4460PA6806	A3W4			134921	521854	11211	177	883
PA	4461PA6807	A1M3			140910	451236	0		884
PA	4462PA6807	A5			140910	451236	11211	052	885
PA	4463PA6807	A5CA5			142960	69	9		886
PA	4464PA6807	A5D34				200848	11221	003	887
PA	4465PA6807	A5D34		1952	125840	390813	11211	001	888
PA	4466PA6807	A5D7		1945	140910	441415	11221	143	889
PA	4467PA6807	A5D7		1965	140910	461800	11211	206	889
PA	4468PA6807	A5D75			140910	451236	0		890
PA	4469PA6807	A5HP			067928	441211	11211	070	891
PA	4470PA6807	A5HB			140910	441415	11221	143	891
PA	4471PA6807	A5D77			140910	451236	21221	052	892
PA	4472PA6807	B7LA				28	9		893
PA	4473PA6821	F6			126442	61	9		894
PA	4474PA6823	S5		1966		651612	11211	148	895
PA	4475PA6825	08			134921	521854	11211	177	896
PA	4476PA6825	08				551309	12221	067	896
PA	4477PA6825	F55			097633	53	9		897
PA	4478PA6825	H5			135080	341415	0		898
PA	4479PA6825	K57			097633	541822	21211	112	899
PA	4480PA6825	M3			123390	48	9		900
PA	4481PA6825	M3				501402	12121	026	900
PA	4482PA6825	M3		1963	135080	341415	0		901
PA	4483PA6825	M3		1963	126442	69	9		901
PA	4484PA6825	K3				501402	12121	026	902
PA	4485PA6825	08				381337	12321	025	903
PA	4486PA6825	08				501402	12121	026	903
PA	4487PA6825	08				551309	12221	067	903
PA	4488PA6825	P617			126442	58	9		904
PA	4489PA6825	P72			140920	25	9		905
PA	4490PA6825	P72			123390	48	9		905
PA	4491PA6825	Q27			152675	311555	21211	076	906
PA	4492PA6825	R27				501402	12121	026	906
PA	4493PA6825	R27		1966		651612	12221	148	907
PA	4494PA6825	S37			097633	541822	21211	112	908
PA	4495PA6825	S4			135080	341415	0		909
PA	4496PA6825	S4			097633	541822	21211	112	909
PA	4497PA6825	S4			147512	571520	21211	045	909
PA	4498PA6825	S4			126442	69	9		909
PA	4499PA6826	C65				501402	12121	026	910
PA	4500PA6826	T7			126442	58	9		911

5.6.2 Identifying a P Class Random Sample for Analysis

The data collected represents demands for books demanded. Ten thousand four-hundred twenty (10,420) different books were demanded during the 51-day period. The collection comprises over 125,000 titles. In order to analyze the demands for books, a random sample of titles was selected from the P Class shelf list using the shelf list sampling procedure described in Section 5.4. The random shelf list locations were generated in samples of 200 as described in Section 5.4. Five random samples were collected. After recording the call numbers from the shelf list, the call numbers were checked against the call number ordered master file to see if the books were demanded during the period of the study. A summary of the number of demands for each sample is shown in Table 5.6.

TABLE 5.6

SUMMARY OF RANDOM SAMPLE OF P CLASS BOOK DEMANDS

Sample Number	Number of Demands 0	1	2	3	4	Total
1	131	10	3	0	0	144
2	137	16	0	0	0	153
3	122	13	4	2	0	141
4	125	16	2	0	0	143
5	135	10	4	0	2	151
Totals	650	65	13	2	2	732

CHAPTER 6

ANALYSIS OF FREQUENCIES OF KNOWN AND UNKNOWN DEMAND FOR BOOKS

The analyses performed in this chapter use the collected random sample of known demand (described in Section 5.5.1) and the random sample of total demand for P class books (described in Section 5.5.2) of the collection of the State University of New York at Buffalo Libraries.

6.1 The Known Demand Sample

The known demand sample consists of 702 books randomly identified in the SUNY/Buffalo shelf list. The frequency distribution of the number of recorded charges are shown in Table 6.1. The mean number of known demands is .38319. The variance is .71886. We hypothesize that the number of known demands for books in the sample can be described by the negative binomial distribution. Using the method of maximum likelihood, the parameter, r, of the distribution was found to be .47066. P was estimated using the ratio of the mean to the variance and was found to be .53305.

TABLE 6.1

FREQUENCY DISTRIBUTION OF KNOWN DEMANDS FOR
A SAMPLE OF BOOKS FROM THE SUNY-BUFFALO COLLECTION DURING
THE PERIOD SEPTEMBER 1, 1966 - August 31, 1967

No. of Known Demands		Frequency
0		530
1		112
2		42
3		9
4		4
5		1
6		3
7		1
	Total =	702

6.1.1 The Chi Squared Goodness of Fit Test

The negative binomial distribution with parameters $p = .53305$ and $r = .47066$ was fit to the known demand data sample. The Chi Squared table is shown in Table 6.2. The Chi Squared column shows the individual components of Chi Squared. The calculated Chi Squared is 6.7451. The P value (the probability of getting a value of Chi Squared as high or higher than the observed) determined from a Chi Squared table was found to be .24. Since the 7 demand cell had a theoretical frequency of less than 1, the frequencies of the 6 and 7 cells were combined into cell 6. The theoretical cell frequencies are less than five in two of the cells. Grouping the frequencies of the last two cells would have resulted in a reduced calculated value of Chi Squared of 3.6526. At 2 degrees of freedom, the P value would be .17. Both are reasonably high values of P. A complete Chi Squared Table is shown in Appendix D. The results of the test indicate that the known demands for the sample of books are negative binomially distributed.

6.1.2 The Variance Test

The variance test was run in order to test the hypothesis that the variance was equal to the mean. The

TABLE 6.2

CHI SQUARED GOODNESS OF FIT TABLE FOR A NEGATIVE BINOMIAL FIT TO THE FREQUENCIES OF KNOWN DEMAND FOR A SAMPLE OF BOOKS FROM THE SUNY-BUFFALO LIBRARY COLLECTION DURING THE PERIOD SEPTEMBER 1, 1966 - August 31, 1967

Value	Observed Frequency	Probability	Theoretical Frequency	Chi Squared
0	530	.74370	522.077	.1203
1	112	.16345	114.742	.0655
2	42	.05612	39.396	.1721
3	9	.02158	15.149	2.4959
4	4	.00874	6.135	.7433
5	1	.00365	2.562	.9526
6	4	.00276	1.936	2.1954
			Total Chi Squared =	6.7451

alternative hypothesis was that the variance was greater than the mean, thus,

H_o: Variance equals the mean

A_o: Variance is greater than the mean

The table of computations for the variance test is shown in Table 6.3. A complete Chi Squared calculation table is shown in Appendix E. The calculated value of Chi Squared is 1322.638, whereas the theoretical Chi Squared calculated using the normal approximation to Chi Squared at a 5% significance level and 701 degrees of freedom is 764. We would thus not accept the hypothesis that the variance is equal to the mean and would conclude that the variance is greater than the mean.

6.2 Sample of P Class Book Demands

A sample of P Class books randomly identified from the P class shelf list was matched with the magnetic tape master file in order to determine the number of times that the 732 books were demanded during the period March 3, 1968 through May 11, 1968. The distribution of the number

TABLE 6.3

VARIANCE TEST TABLE FOR KNOWN DEMAND
SAMPLE OF BOOKS DEMANDED FROM
SEPTEMBER 1, 1966 - AUGUST 31, 1967

Value	Frequency	Chi Squared
0	530	201.930
1	112	112.635
2	42	288.947
3	9	162.027
4	4	137.503
5	1	55.998
6	3	248.608
7	1	114.990

Total Chi Squared =1322.638

of demands is shown in Table 6.4. The mean of the distribution is .14344 and the variance is .20785. We again hypothesize that the demand is negative binomially distributed and, using the method of maximum likelihood, estimate the parameter r to be .41276. The parameter p was estimated, using the ratio of the mean to the variance, to be .69011.

6.2.1 The Chi Squared Goodness of Fit Test

The negative binomial distribution with parameters r = .41276 and p = .69011 was fit to the P class demand data. The Chi Squared table is shown in Table 6.5. No grouping was done in the tail of the distribution. The calculated value of Chi Squared was 6.2817. The P value was determined from a Chi Squared table at 2 degrees of freedom to be .048. Grouping the frequency for cell 4 with cell 3 would have given a P value of .021. The data is not exceptionally well fitted by a negative binomial distribution.

TABLE 6.4

FREQUENCY DISTRIBUTION OF DEMANDS
FOR P CLASS BOOKS DURING THE
PERIOD MARCH 3, 1968 - MAY 11, 1968

No. of Demands		Frequency
0		650
1		65
2		13
3		2
4		2
	Total =	732

TABLE 6.5

CHI SQUARED GOODNESS OF FIT TABLE FOR A NEGATIVE BINOMIAL FIT TO THE FREQUENCIES OF DEMAND FOR THE SAMPLE OF P CLASS BOOKS DEMANDED DURING MARCH 3, 1968 - MAY 11, 1968

Value	Observed Frequency	Probability	Theoretical Frequency	Chi Squared
0	650	.85805	628.093	.7641
1	65	.10975	80.337	2.9280
2	13	.02402	17.583	1.1944
3	2	.00599	4.385	1.2969
4	2	.00219	1.603	.0983
			Total Chi Squared =	6.2817

6.2.2 The Variance Test

The variance test was again run in order to test the hypothesis that the variance is equal to the mean. The results of the variance test are shown in Table 6.6. The calculated Chi Squared is 1062.508. At a 5% level of significance and 731 degrees of freedom, the theoretical Chi Squared was calculated to be 795. The normal approximation to Chi Squared was used to calculate the theoretical Chi Squared. Thus, we would again not accept the hypothesis that the variance is equal to the mean and conclude that the variance is greater than the mean.

6.3 Analysis of the Distributions of Known and Unknown Demand

On the basis of the variance test alone, we would conclude that the variance is greater than the mean which provides suspicions that the data is negative binomially distributed. The results of the Chi Squared goodness of fit test provides reasonably strong evidence that the known demands (Section 6.1) are negative binomially distributed, but weak evidence that the total demand (known

TABLE 6.6

VARIANCE TEST TABLE FOR DEMANDS FOR
P CLASS SAMPLE OF BOOKS DEMANDED FROM
MARCH 3, 1968 - MAY 11, 1968

Value	Frequency	Chi Squared
0	650	92.950
1	65	333.840
2	13	313.495
3	2	114.160
4	2	208.052
	Total Chi Squared =	1062.508

and unknown) for P class books (Section 6.2) are negative binomially distributed. We are hampered in the latter goodness of fit test with few cells and the fact that we lose 3 degrees of freedom because of estimating two parameters of the distribution. This is typical of book use studies since exceptionally large samples are required before a sufficient number of values are found in the tail of the distribution to reach sound statistical conclusions. The fact that we accept the alternative hypothesis that the variance is greater than the mean and have evidence from the Chi Squared tests, the demands for both groups of books are negative binomially distributed.

6.4 Implications of Negative Binomially Distributed Demands

In deriving the negative binomial distribution in Section 3.1, it was shown in equation [3.4] that if the probability of 0, 1, 2, ... events follow a Poisson distribution with parameter λ:

$$f(x) = \frac{e^{-\lambda} \lambda^{x}}{x!} \qquad x = 0, 1, 2, \ldots$$

and the mean number, λ, of events vary in such a way that they can be described by the Gamma probability distribution with density function [3.5]:

$$g(\lambda) = \frac{C^r}{\Gamma(r)} e^{-C\lambda} \lambda^{r-1}$$

Then, the joint density of x and λ is the product of the above equations, and leads to a negative binomial distribution of the variable x. Thus, concluding here that the demands for groups of books is negative binomially (as was done in the samples analyzed in this section), permits us to at least assume that the number of demands for individual books varies according to a Poisson distribution and that the demand rates vary from book to book according to a Gamma distribution.

CHAPTER 7

DAILY DEMAND FOR P CLASS BOOKS

Using the master file of demand data for P class books, ten random samples of 100 demanded books were collected. The daily demand for each of the samples was tabulated and analyzed to determine if the number of demands per day for each of the samples was Poisson distributed. The Chi Squared goodness of fit and the variance tests were used to determine if the number of demands per day, X, was distributed according to

$$p(X) = \frac{e^{-\mu}\mu^{x}}{X!}$$

Where:

μ = the mean daily demand for the sample

X = the number of demands

7.1 Randomly Selecting Books to be Included in the Samples

The P class books demanded during the 51-day period from March 3, 1968 through May 11, 1968 were numbered from 1 to 14,222. Ten sets of 100 random numbers were generated within this range identifying those books which would be included in each of the samples. The number of daily demands was tabulated for each of the books in each of the samples. The demand frequencies for each of the samples are shown in Table 7.1 together with the mean demand and variance of each of the samples of 100 books.

In general, the demand distribution did not exhibit very long tails except for sample 3 in which there were 13 demands on one day. The means and variances were calculated using the actual data and not the data grouped into the greater than or equal to seven category.

TABLE 7.1

DAILY DEMAND FREQUENCIES OF TEN RANDOM
SAMPLES OF ONE HUNDRED P CLASS BOOKS DEMANDED
DURING MARCH 3, 1968 - MAY 11, 1968

Sample Number	Number of Daily Demands 0	1	2	3	4	5	6	7	Sample Size	Mean	Variance
1	2	9	16	13	6	2	1	2	51	2.63	2.36
2	5	14	10	9	6	3	2	2	51	1.51	5.01
3	4	10	13	6	7	7	2	2*	51	2.27	3.14
4	7	14	10	8	6	4	2	0	51	2.43	2.78
5	8	12	8	11	7	1	1	3	51	2.33	3.63
6	8	10	11	10	5	6	0	1	51	2.31	3.84
7	7	10	13	11	5	3	1	1	51	2.31	4.25
8	9	7	13	11	2	7	0	2	51	2.43	4.25
9	6	11	12	9	7	3	3	0	51	2.41	2.77
10	7	9	14	10	6	1	3	1	51	2.37	2.92

*One book had 13 demands

7.2 The Chi Squared Goodness of Fit Test

A Chi Squared goodness of fit test was run on each of the samples. Because of the low theoretical cell frequencies in the tails of the distribution (less than one), it was necessary to group the data in all but the 9th sample. These data were grouped into the "greater than or equal to six" cell except for sample 3 in which the data was grouped into the "greater than or equal to seven" cell. Table 7.2 shows the sample number, calculated value of Chi Squared (Chi Squared Observed) and P values for each of the calculated Chi Squared. The P value is the probability of getting a value of Chi Squared as high or higher than the observed if the hypothesis is true. A tabulation of the Chi Squared goodness of fit tests for all samples can be found in Appendix E.

7.3 The Variance Test

The variance test was run on each of the samples in order to determine whether there is sufficient evidence to reject the hypothesis that the variance is equal to the mean. The alternative hypothesis is that the

TABLE 7.2

SUMMARY OF THE CHI SQUARED GOODNESS OF FIT TEST
FOR A POISSON FIT TO THE FREQUENCIES OF DAILY DEMAND
FOR SAMPLES OF BOOKS DEMANDED DURING THE
PERIOD MARCH 3, 1968 - MAY 11, 1968

Sample Number	Observed Chi Squared	Degrees of Freedom	P Value
1	3.16	5	.68
2	4.16	5	.53
3	5.12	6	.41
4	3.29	5	.66
5	8.74	5	.13
6	6.48	5	.27
7	1.36	5	.93
8	13.71	5	.02
9	1.59	5	.90
10	5.85	5	.33

variance is greater than the mean. Concluding that the variance is equal to the mean is important since we are, in effect, testing to see if the daily demand is Poisson distributed. The values in the tails of the distributions were not grouped for this test. A summary of the sample number, means, variances, calculated Chi Squared (Chi Squared Observed), degrees of freedom and P values are summarized in Table 7.3. The results of the complete tests are shown in Appendix I. The P value is, as before, the probability of getting a value of Chi Squared as high or higher than observed.

7.4 Analysis of Daily Demand

Nine of the ten Chi Squared goodness of fit tests would tend to confirm the hypothesis of Poisson distributed daily demand since they showed reasonably high P values. Sample eight has a low P value. The variance test, on the other hand, results in the acceptance of the variance equals mean hypothesis at the 5% level of significance in only six of the ten samples. Samples 2, 3, 5 and 8 would tend to be not accepted as having their variances equal to the mean. A closer look at the results of both the Chi Squared goodness

TABLE 7.3

SUMMARY OF THE RESULTS OF THE VARIANCE TEST RUN ON SAMPLES OF DAILY DEMAND FREQUENCIES OF P CLASS BOOKS DEMANDED DURING THE PERIOD MARCH 3, 1968 - MAY 11, 1968

Sample Number	Mean	Variance	Chi Squared Observed	Degrees of Freedom	P Value
1	2.63	2.36	44.88	50	.677
2	2.51	5.01	74.41	50	.015
3	2.72	3.14	88.20	50	.005
4	2.24	2.78	62.26	50	.119
5	2.43	4.17	85.76	50	.005
6	2.33	3.63	62.29	50	.118
7	2.31	3.84	61.80	50	.128
8	2.43	4.25	72.60	50	.021
9	2.41	2.77	57.37	50	.124
10	2.37	2.92	61.50	50	.134

of fit test and the results of the variance test shown in Appendices II and I reveals some interesting facts. Sample 2, which had a good P value on the goodness of fit test and has a low P value on the variance test had frequencies of 7 and 9 pulled into the greater than or equal to six cell. Looking at the variance test table shows that the bulk of the Chi Squared is caused by the value of 9 in the table. Eliminating this value would lower the total Chi Squared which would give a P value closer to .12 which would be accepted at the 5% level. Sample 3 has similar characteristics. However, here the distribution would have a better fit if the one value of 13 were eliminated. Samples 5 and 8 exhibit poor values of Chi Squared for most of the cells in the goodness of fit and variance test analyses. A check on confidence limits for daily mean demand was made and mean demand was out of control on days 51, 50 and 66 and almost out of control on day 52. This check was made using the first five samples (500 books) and calculating a mean demand per day of .254. The standard deviation of the daily mean demands was calculated to be .1058. Using 95% confidence limits for the daily mean demands with the normal distribution, the limits were .0467 and .4613. No days were outside the lower confidence limit. Day 51 showed a mean demand of .520, day 52 showed .440, day 60 showed .480; and day 66 showed .520. Interestingly, the problems with the variance test in samples 2 and 3 occurred with frequencies for both of those days. We could conclude that something peculiar occurred on those days to cause a shift in the mean demand for books. One could possibly conclude that the mean demand varies on a daily basis. If this is so, the daily demand data would fit a negative binomial distribution if the demand rates for the group of books varied from day to day according to a gamma distribution. A further point could be made here concerning a possible shift in the mean demand late in the semester. All four of the high mean demand days occurred in the last 20 days of the 51-day collection procedure. Although analysis of the demand data summary appeared to indicate a slight shift in the number of demands late in the semester, no attempt was made in this study to determine where the shift occurred. Also, there appeared to be no significant day to day variation in mean demand although the point was not formally pursued. The data collected in the ten random samples, does appear to support, although not exceptionally strongly, the Poisson distributed demand hypothesis. This evidence, again, does not support the hypothesis that the demand for individual books is Poisson distributed. It does support the hypothesis that daily demand is Poisson distributed. Not rejecting this hypothesis, however, does provide evidence that the demand for individual books is Poisson distributed. This point will be pursued in Section 7.5.

7.5 Significance of the Distribution of Daily Demands

It is assumed here that the demand $(x_1, \ldots, x_n)$ for individual books are independently distributed with parameters $(\mu_1, \ldots, \mu_n)$. Then,

$$x = \sum_{i=1}^{n} x_i$$

has the Poisson distribution with parameter,

$$\sum_{i=1}^{n} \mu_i$$

Consider the case for n = 2. The variables x_1, x_2 takes the values 0, 1, 2, ... We find the probability that their sum $(x_1 + x_2)$ is equal to r. The probability that x_1 has the value s and x_2 the value (r-s) is:

$$P\{x_1 = s\} P\{x_2 = r-s\} = e^{-\mu_1} \frac{\mu_1^{s}}{s!} e^{-\mu_2} \frac{\mu_2^{(r-s)}}{(r-s)!}$$

The probability that x = r is given by summing the above equation over all values of s (0 to r):

$$\sum_{s=0}^{r} e^{-\mu_1} e^{-\mu_2} \frac{\mu_1^{s} \mu_2^{(r-s)}}{s!(r-s)!} = e^{-(\mu_1+\mu_2)} \sum_{s=0}^{r} \frac{\mu_1^{s} \mu_2^{(r-s)}}{s!(r-s)!} \qquad [7.1]$$

the binomial expansion of $(\mu_1 + \mu_2)^r$ is:

$$\sum_{s=0}^{r} \binom{r}{s} \mu_1^{s} \mu_2^{(r-s)} = r! \sum_{s=0}^{r} \frac{\mu_1^{s} \mu_2^{(r-s)}}{s!(r-s)!}$$

the summation in [7.1] is

$$\frac{(\mu_1 + \mu_2)^r}{r!}$$

and,

$$P\{x=r\} = e^{-(\mu_1+\mu_2)} \frac{(\mu_1+\mu_2)^2}{r}$$

which is a Poisson frequency distribution function with parameter $(\mu_1 + \mu_2)$. This can be extended to n values.

Thus, when we test the daily demand for a sample of books (not necessarily random), we are, in addition to showing that the demand for the whole sample is Poisson distributed, providing evidence that the demands for the individual books comprising the sample are Poisson distributed. It could happen that some arbitrary distributed demands for books could combine in such a way that the demand for the whole sample is Poisson, but it is unlikely.

CHAPTER 8

BOOK USE IN SEVERAL INSTITUTIONS

As part of their study, "Patterns In The Use Of Books In Large Research Libraries," Fussler and Simon collected data and made comparisons of book-use in several institutions. They used their collected data to attempt to find out if their findings could be generalized to other libraries.

8.1 The Fussler and Simon Data

The Fussler and Simon data collection procedure involved deriving lists of books that were held at the University of Chicago library and other university libraries. They compared sample lists in biology, Teutonic languages and literature, and philosophy against the holdings of the Yale University library; the lists of physics, Teutonic languages and literature, and economics against the holdings of the Northwestern University library; and the lists of economics, Teutonic languages and literature, and biology with the lists of the University of California at Berkeley library. The original University of Chicago lists were developed by a random systematic sampling scheme for selecting monographs for the use studies in the University of Chicago library. For each comparison sample, the titles on the original list that were held by both institutions formed the new sample. Each comparison-sample formed a group of similar items that were available to patrons in different institutions. Fussler and Simon selected the particular libraries so as to include one much larger library, one smaller library, and one about the same size as the University of Chicago library. They then attempted to answer the question, "Are titles that are used little or much in one institution used little or much in the other?" They examined each use-group of titles (identified by use in 1949-1953) in each comparison-sample to see how the groups behaved in 1954-1958 in the two institutions in each pair. For example, the data presented for the Chicago-Yale comparison for Philosophy are shown in Table 8.1. Of the titles that had no use at Yale in 1949-1953, 58 were not used at Chicago in 1954-1958; 70 were not used at Yale during the same period.

TABLE 8.1

COMPARISONS OF THE USE OF THE SAME
PHILOSOPHY TITLES AT CHICAGO AND YALE

Use at Yale in 1949-53	Number in Group	Use at Chicago 1954-1958 Number of Cases Used 0	1	2	3	4	5+	Total Use	Total Use (0-5)	Use at Yale 1954-1958 Number of Cases Used 0	1	2	3	4	5+	Total Use	Total Use (0-5)
0	40	58	11	6	0	1	3	57	42	70	7	1	1	0	0	12	12
1	21	13	3	2	0	0	0	7	7	10	4	3	0	1	0	14	14
2	14	7	3	1	1	0	2	18	18	6	4	1	1	2	0	17	17
3	6	1	0	0	1	0	0	3	3	0	1	0	0	0	1	6	6
4+	8	6	4	3	1	3	2	65	35	4	2	2	1	3	7	105	56
TOTALS	89	85	21	12	3	4	7	150	105	90	18	7	3	6	8	154	105

8.2 Fitting the Negative Binomial Distribution

Although the Fussler and Simon data are used in this research to perform an analysis to determine if book use for samples of books at other institutions is negative binomially distributed, Fussler and Simon collected and used the data for a much different purpose. For the research reported here, the frequency of book use during the period 1954-1958 in the specified library and subject area were extracted from the Fussler and Simon data. The data are presented in Table 8.2.

The mean use for each sample was calculated by using the total number of uses for the books in the sample and the total number of books in the sample. This results in a different mean for each sample than that which would be calculated using the frequency data, since the number of uses of frequently used books are necessarily grouped into the 5+ use category. Using the mean calculated in this latter manner would, in general, result in an inaccurate fit of the negative binomial distribution to the data. The parameter, r, was estimated for each of the samples using the method of maximum likelihood. The estimated

TABLE 8.2

NUMBER OF USES OF BOOKS IN SAMPLES IN SPECIFIED SUBJECT AREAS IN THE CHICAGO, YALE, NORTHWESTERN AND BERKELEY LIBRARIES DURING THE PERIOD 1954 - 1958

Subject Area	Library	Sample Selected To Be Compared With Books Used 1949-1953 at	Number of Uses 0	1	2	3	4	5	Sample Size	Mean Use
Philosophy	Chicago	Yale	85	21	12	3	4	7	132	1.13636
	Yale		90	18	7	3	6	8	132	1.16667
Teutonic Languages and Literature	Chicago	Yale	109	19	7	5	2	7	149	1.02013
	Yale		112	22	6	3	1	5	149	.75168
	Chicago	Northwestern	39	5	3	1	4	9	61	2.00000
	Northwestern		46	7	2	3	1	2	61	.77049
	Chicago	California	125	23	2	8	1	10	169	1.00592
	California		125	21	7	7	3	6	169	.77515
Biology	Chicago	Yale	81	10	4	7	2	11	115	1.99130
	Yale		89	7	3	3	2	11	115	1.83478
	Chicago	California	72	17	7	5	0	5	106	.98113
	California		59	18	10	5	6	8	106	1.57547
Physics	Chicago	Northwestern	28	14	9	4	6	20	81	4.87654
	Northwestern		44	13	4	3	3	14	81	1.81481
Economics	Chicago	Northwestern	55	16	7	1	0	10	89	2.20225
	Northwestern		52	20	6	2	2	4	89	.88764
	Chicago	California	69	16	5	2	0	2	94	.52128
	California		57	17	8	3	3	6	94	1.03191

value of p, was therefore determined by using the relationship, $p = r/(m+r)$. After estimating the parameters from the sample data, the negative binomial distribution was fit to the data using Pearson's Chi Squared Goodness of Fit test. A summary of the sample size (N), sample mean (m), frequency of use, values of the parameters r and p, calculated values of Chi Squared, degrees of freedom (df) and P values are displayed in Table 8.3. The P value is the probability of getting a value of the Chi Squared statistic as high or higher than the observed Chi Squared. The original data presented by Fussler and Simon and the results of the Goodness of Fit test on each sample are shown in Appendix J. An analysis is given in Section 8.4.

8.3 Variance Test

A variance test was run on each of the samples using the following hypotheses:

Hypothesis: $\sigma^2 = \mu$

Alternative: $\sigma^2 > \mu$

For each of the samples a Chi Squared statistic was calculated using:

$$\chi^2_{observed} = \sum_{i=1}^{n} \left[\frac{f_i \, (i-m)^2}{m} \right]$$

TABLE 8.3

VALUES OF N, m, r, p, df, χ^2 AND P FOR THE NEGATIVE BINOMIAL DISTRIBUTIONS WHICH WERE FIT TO THE DATA FOR SPECIFIED SUBJECT AREAS AND INSTITUTIONS

Subject Area	Library	Sample Size (N)	Mean (m)	Values (r)	(p)	χ^2	df	P
Philosophy	Chicago	132	1.13636	.23385	.17067	4.39	3	.225
	Yale	132	1.16667	.18322	.13573	4.59	3	.206
Teutonic Languages and Literature	Chicago	149	1.02013	.14014	.12078	3.70	3	.296
	Yale	149	.75168	.14971	.16609	6.46	3	.082
	Chicago	61	2.00000	.17298	.07960	4.47	3	.292
	Northwestern	61	.77049	.15156	.16437	1.86	3	.607
	Chicago	169	1.00592	.13393	.11750	12.44	3	.002
	California	169	.77515	.17028	.18011	2.42	3	.492
Biology	Chicago	115	1.99130	.11445	.05435	5.55	3	.144
	Yale	115	1.83478	.07719	.04037	.37	3	.946
	Chicago	106	.98113	.22101	.18385	4.54	3	.210
	Northwestern	106	1.57547	.29627	.15829	3.54	3	.310
Physics	Chicago	81	4.87654	.30428	.05873	6.94	3	.079
	Northwestern	81	1.81481	.30362	.14332	2.65	3	.457
Economics	Chicago	89	2.20225	.61045	.21703	24.38	3	.001
	Northwestern	89	.88764	2.12229	.70509	18.80	3	.001
	Chicago	94	.52128	1.48549	.74024	4.58	1	.035
	California	94	1.03191	3.17793	.75556	37.45	3	.001

Where:

i = the number of uses
f_i = the frequency of occurence of the ith use
m = the mean of the sample

A theoretical Chi Squared was calculated using the normal approximation for Chi Squared:

$$\chi^2_{theoretical} = \frac{1}{2} (Z_p + \sqrt{2k - 1})^2$$

Where:

Z_p = value of a standard normal deviate at a significance level of 1-p
k = degrees of freedom = N-1

The normal approximation was used since there are greater than 60 degrees of freedom. For all samples the value of Chi Squared was calculated at a significance level of 5%. A summary of the results of the test is displayed in Table 8.4. It should be pointed out that the variance test tables are shown in Appendix K. The Chi Squared observed will be understated in all samples because Fussler

TABLE 8.4

RESULTS OF VARIANCE TEST ON SAMPLES
OF BOOK USES IN SPECIFIED INSTITUTIONS

Subject Area	Library	Sample Size	Calculated Variance	Mean	Calculated χ^2	df	Theoretical χ^2	Action
Philosophy	Chicago	132	1.91967	1.13636	234.80	131	159.31	DNA
	Yale	132	2.17922	1.16667	223.13	131	159.31	DNA
Teutonic Language and Literature	Chicago	149	1.64475	1.02013	263.10	148	133.66	DNA
	Yale	149	1.21087	.75168	252.69	148	133.66	DNA
	Chicago	61	3.71311	2.00000	129.50	60	78.75	DNA
	Northwestern	61	1.48415	.77049	119.17	60	78.75	DNA
	Chicago	169	1.80811	1.00592	326.83	168	199.00	DNA
	California	169	1.50697	.77515	334.93	168	199.00	DNA
Biology	Chicago	115	2.68009	1.99130	223.86	114	136.95	DNA
	Yale	115	2.54539	1.834478	233.93	114	136.95	DNA
	Chicago	106	1.59470	.98113	181.14	105	104.40	DNA
	Northwestern	106	2.57008	1.57547	186.26	105	104.40	DNA
Physics	Chicago	81	4.16944	4.87654	198.85	80	101.53	DNA
	Northwestern	81	3.73920	1.81481	173.16	80	101.53	DNA
Economics	Chicago	89	2.56359	2.20225	167.59	88	111.75	DNA
	Northwestern	89	1.64837	.88764	165.56	88	111.75	DNA
	Chicago	94	.91650	.52128	125.65	93	116.28	DNA
	California	94	2.13910	1.03191	194.53	93	116.28	DNA

Note: DNA means Do Not Accept
DNR means Do Not Reject

and Simon grouped frequencies with greater than 5 uses into their 5+ use category.

8.4 Analysis

Although not specifically for their comparisons of book-use in several instituions, but elsewhere in their study, Fussler and Simon make the point that the distribution of books by the frequency of their use is not Poisson distributed. Fussler and Simon devoted their Appendix H to this premise and state:

> "The assumption of a stochastic model does not suggest that the Poisson distribution will approximate the distribution of books within a library (or within a given subject area) by the frequency of their use during some period of time. If a library contained 10 books, each with an expectation of 1,000 uses per year and 100 books with an expectation of .01 uses per year, the expected distribution of a sample drawn from such a library, instead of being Poisson will show a scatter of books around 1,000 uses, and a scatter of books at zero use and slightly above, with practically nothing between the scatters. We make this point because the fast-falling, convex-to-the-origin curve stemming from a binomial process may immediately suggest the Poisson to many readers.

> And, in fact, the observed distributions do not resemble the Poisson closely, having a much higher variance. There are too many observations at "zero" and at multiple use points....."

Fussler and Simon then demonstrate this point by showing the distribution of the number of uses for monographs in economics and Teutonic languages and literature samples with the points of a Poisson distribution superimposed on the actual distribution.

Indeed Fussler and Simon are correct in their conclusions that the frequencies of number of uses of books in collections are not Poisson distributed. Their example using the 10 high use books and 100 low use books is, however, somewhat ambiguous. Their point is valid. Such use would not lead to a Poisson distribution just as normally distributed frequency of use or some other arbitrarily distributed frequency of use would probably not fit. It is the intention here to show that the Fussler and Simon data do fit the negative binomial distribution reasonably well.

In all cases, the hypothesis that the variance is equal to the mean was not accepted at the 5% level of significance. This is an important point since negative binomially distributed random variables necessarily have a variance higher than the mean, whereas the variance would be equal to the mean for the Poisson distribution and less than the mean for the binomial distribution. The variance test is frequently used when there are too few values to use a Chi Squared Goodness of Fit test. Cochran (1954) makes the point that he has frequently found the variance test to be significant when the Goodness of Fit test was not. He indicates that there is a striking difference in the power of the variance test over the Goodness of Fit test. Concluding that the variance is not equal to the mean here does not in itself lead one to the conclusion that the data are negative binomially distributed. It does lead one to conclude that the data are not Poisson distributed and provides evidence that the sample data are negative binomially distributed since the alternative hypothesis is that the variance is greater than the mean.

The results of the eighteen goodness of fit tests indicate that the sample data fits the negative binomial distribution in 13 of the 18 samples. The samples which do not fit well are the four samples in the economics subject area at the Chicago, Nortwestern and University of California at Berkeley libraries, and the Teutonic language sample collected at the University of Chicago library in order to compare use at that library with the use at the University of California at Berkeley library. One could probably conclude by looking at the results of these Goodness of Fit tests that perhaps there is something peculiar about the usage of economics books which causes the poor fit of economics book usage data with a negative binomial distribution and thus would lead one to conclude that number of uses of individual economics books in libraries are not Poisson distributed.

A closer look at the economics samples reveals that the ratio of the number of books with zero use to the number of books with greater than five uses is quite high. As a matter of fact, the ratios for the economics samples at Chicago when compared with California, and the Northwestern samples have the highest ratios of all the samples. The economics sample at Chicago (when compared with Northwestern) does not. However, of the 89 books in the sample, 71 were not used or used once, eight books had 17 uses and 10 books had 173 uses. This is indeed a most peculiar distribution. The fit of all the distributions would have been more accurate if it were possible to analyze the data in the tail of the distribution. One could only conclude that there is something which causes the uses of these samples of economics books to be not negative binomially distributed. The other samples do, however, support the hypothesis of negative binomially distributed uses of samples of books both with the variance test and the Chi Squared Goodness of Fit test. Thus, the hypothesis of Poisson distributed uses of individual books tends to be supported by the Fussler and Simon data.

CHAPTER 9

THE PREDICTION OF DEMAND FOR BOOKS

The models developed in this chapter with the exception of that in Section 9.7, are models for a book which has only one copy available. Models for multiple copy books have not been considered in this investigation.

9.1 A Two Stage Book Use Model

Consider a book which has a demand rate equal to λ per year and which is permitted to be charged out of the library for a period of time $1/\mu$ where μ is the permissible loan rate per year for the book. $1/\mu$ is called the loan interval. Suppose that we say that a book is in state 0 when it is on the shelf and available for use, and is in state 1 when the book is charged out. Assume that only the two states are permissible, specifically, the book is in state 0 if it is not charged out and state 1 if it is charged out. Given that the book is in state 0 at time t there is a probability $\lambda\Delta t+0(\Delta t)$ of a transition to state 1 during the interval $(t,t+\Delta t)$, independently of all occurrences before t. If the book is in state 1 at time t, there is a probability $\mu\Delta t+0(\Delta t)$ that there will be a transition to state 0 during the interval $(t, t + \Delta t)$, independently of all occurrences before t. We assume that the probability of more than one transition during the interval $(t, t + \Delta t)$ is negligible and designate these probabilities above as $0\Delta t$. An alternate specification of this process is in terms of two sequences of mutually independent random variables $(T_1, T_2, \ldots)$ and $(T_1', T_2', \ldots)$ exponentially distributed with parameters equal respectively to λ and μ. If the process starts in state 0 there is a transition to state 1 at time T_1, a transition to state 0 after a further time, T_1'. This two stage book use process is shown in Figure 9.1. Let $p_1(t)$ and $p_0(t)$ be the probability distributions at time t and let $p_1(0)$ and $p_0(0) = 1 - p_1(0)$ be specified.

Using the above probabilities, it follows that:

$$p_0(t+\Delta t) = p_1(t)[\mu\Delta t + 0(\Delta t)] + p_0(t)[1 - (\lambda\Delta t + o(\Delta t))]$$

$$p_1(t+\Delta t) = p_0(t)[\lambda\Delta t + 0(\Delta t)] + p_1(t)[1 - (\mu\Delta t + 0(\Delta t))]$$

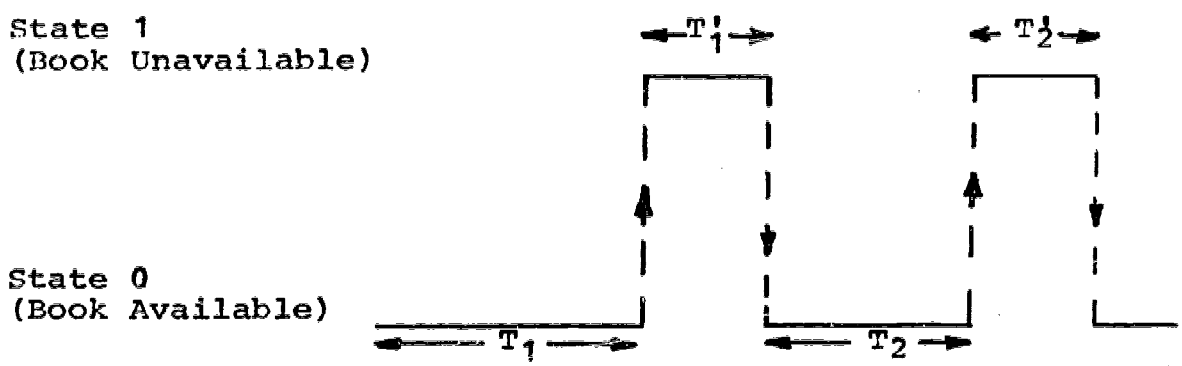
Then letting $0(\Delta t) \to 0$

$$p_0(t+\Delta t) = p_1(t)\mu\Delta t + p_0(t) - p_0(t)\lambda\Delta t$$

$$p_1(t+\Delta t) = p_0(t)\lambda\Delta t + p_1(t) - p_1(t)\mu\Delta t$$

FIGURE 9.1

A TWO STAGE BOOK USE PROCESS



And

$$p_0(t+\Delta t) - p_0(t) = -p_0(t)\lambda\Delta t + p_1(t)\mu\Delta t$$

$$p_1(t+\Delta t) - p_1(t) = p_0(t)\lambda\Delta t - p_1(t)\mu\Delta t$$

Dividing by Δt

$$\frac{p_0(t+\Delta t) - p_0(t)}{\Delta t} = -\lambda p_0(t) + \mu p_1(t)$$

$$\frac{p_1(t+\Delta t) - p_1(t)}{\Delta t} = \lambda p_0(t) - \mu p_1(t)$$

Letting $\Delta t \rightarrow 0$

$$\frac{dp_0(t)}{dt} = -\lambda p_0(t) + \mu p_1(t) \qquad [9.1]$$

$$\frac{dp_1(t)}{dt} = \lambda p_0(t) - \mu p_1(t) \qquad [9.2]$$

We know

$$p_1(t) = 1 - p_0(t)$$

Therefore for [9.1]

$$\frac{dp_0(t)}{dt} = -\lambda p_0(t) + \mu[1 - p_0(t)]$$

$$p_0'(t) = -\lambda p_0(t) + \mu - \mu p_0(t)$$

Rearranging

$$P_o'(t) = -(\lambda+\mu)P_o(t)+\mu$$

Taking La Place transforms

$$\mathcal{L}\{P_o(t)\} = u(s)$$

$$\mathcal{L}\{P_o'(t)\} = s\,u(s)-P_o(0)$$

$$\mathcal{L}\{\mu\} = \frac{1}{s}$$

Then

$$s\,u(s) - P_o(0) + (\lambda+\mu)u(s) - \frac{\mu}{s} = 0$$

$$u(s)[s+(\lambda+\mu)] = \frac{\mu}{s} + P_o(0)$$

$$u(s) = \frac{\mu}{s[s+(\lambda+\mu)]} + \frac{P_o(0)}{s+(\lambda+\mu)} \qquad [9.3]$$

By partial fractions

$$\frac{\mu}{s[s+(\lambda+\mu)]} = \frac{A_1}{s} + \frac{A_2}{s+(\lambda+\mu)}$$

$$\mu = A_1[s+(\lambda+\mu)]+A_2 s$$

When s = 0

$$\mu = A_1(\lambda+\mu)$$

$$A_1 = \frac{\mu}{\lambda+\mu}$$

When s = $-(\lambda+\mu)$

$$\mu = -(\lambda+\mu)A_2$$

$$A_2 = -\frac{\mu}{\lambda+\mu}$$

Then substituting into [9.3]

$$u(s) = \frac{\mu}{(\lambda+\mu)s} - \frac{\mu}{(\lambda+\mu)}\frac{1}{[s+(\lambda+\mu)]} + \frac{p_o(0)}{[s+(\lambda+\mu)]}$$

Taking Inverse La Place transforms

$$p_o(t) = \frac{\mu}{\lambda+\mu} + p_o(0) - \frac{\mu}{\lambda+\mu}\, e^{-(\lambda+\mu)t} \qquad [9.4]$$

We know $p_o(t) = 1 - p_1(t)$

Therefore for [9.2]

$$\frac{dp_1(t)}{dt} = \lambda[1 - p_1(t)] - \mu p_1(t)$$

$$p_1'(t) = \lambda - \lambda p_1(t) - \mu p_1(t)$$

$$p_1'(t) = \lambda - (\lambda + \mu) p_1(t)$$

Rearranging

$$p_1'(t) + (\lambda + \mu) p_1(t) = \lambda$$

As before

$$p_1(t) = \frac{\lambda}{\lambda + \mu} + \left\{ p_1(0) - \frac{\lambda}{\lambda + \mu} \right\} e^{-(\lambda + \mu)t} \qquad [9.5]$$

For $p_0(t)$ and $p_1(t)$ when $t \to \infty$ we have the equilibrium distribution:

$$p_0 = \frac{\mu}{\lambda + \mu} \qquad [9.6]$$

$$p_1 = \frac{\lambda}{\lambda + \mu} \qquad [9.7]$$

From the above development, we could determine the probability of the availability of the book, p_0, and the probability that the book is unavailable, p_1 (the probability that the book is charged out).

9.2 Comments Concerning the Two Stage Book Use Model

The model for the availability and unavailability of a book was developed on the assumption that the number of demands are Poisson distributed with mean λ and that the loan intervals are exponentially distributed with mean $1/\mu$. The model further explicitly considered only known demands (book charges) as permissible uses and a book which was charged out was in state 1 and unavailable whereas a book which was not charged out was available. This latter point, however, does not consider unknown demands which are satisfied but for which the book is used in the library. Considered in this way, the p_0 and p_1 developed in Section 9.1 could be considered as:

p_0 = the probability that the book is demanded and not available or is demanded, is available and used in the library (i.e., the probability the book is not charged out).

p_1 = the probability that the book is charged out.

Thus, we are sure of the probability that the book is charged out, p_1, and we are sure of the probability that the book is not charged out, p_o. We are unsure, however, of the components of p_o, namely, the probability that the book is used in the library or is unavailable. The latter probability, the probability of a book being used in the library is, however, very small when one considers the demand rates for books in libraries (such as they are), since books used within libraries are generally off the shelf for at most a day whereas books charged out of the library are off the shelf for 3 weeks, a month, a semester or for a longer period of time. The P class demand data (Section 5.6) showed no day in which there were two demands for the same book. Thus, unavailability of books in a library is not caused by unknown satisfied demands but by known satisfied demands and the p_o could be considered as described with little error.

9.3 Use of the Two Stage Book Use Model for Loan Interval Establishment for a Book with a Known Demand Rate

It is assumed that the demand rate for a book, λ, is known and that the mean loan interval is the loan interval, $1/\mu$, established by the library. The number of demands for the book is Poisson distributed with mean, λ, above. The particular library can establish a goal as to the percent of the time that the library wishes to have the book available for a patron. Using [9.6] and [9.7], it is possible to establish a loan interval for the book. Thus, if the probability of availability is:

$$P_0 = \frac{\mu}{\lambda + \mu}$$

Then

$$\mu = \frac{P_0 \lambda}{P_1}$$

Since $p_1 = 1 - p_0$

For example, if the library had as its goal an availability of 75% and the book had a demand rate of 3 demands per year:

$$\mu = \left(\frac{.75}{.25}\right)(3)$$

$$\mu = 9$$

Thus, the loan interval for the book should be 1/9 year or approximately six weeks. This model assumes the λ is known.

9.4 Use of the Two Stage Book Use Model for Loan Interval Establishment for a Book with a Demand Rate Which is Unknown

Using equations [9.6] and [9.7], it is possible to determine the mean rate of book charges and the mean rate of unsatisfied and unknown demands. Since a patron could charge out a book only if it is available, p_o, and the demand rate is λ, the mean number of persons charging out a book is $p_o\lambda$ or

$$\lambda\left(\frac{\mu}{\lambda+\mu}\right) = \frac{\lambda\mu}{\lambda+\mu} \qquad [9.8]$$

If the number of charges, N, for a period of time and μ for the period of time were known, it is possible to estimate λ by

$$N = \frac{\lambda\mu}{\lambda+\mu}$$

$$\lambda = \frac{\mu N}{\mu - N} \qquad [9.9]$$

For example, if a library had a loan interval, $1/\mu$, equal to 1 month (μ would be 12) which was charged out five times during the past year, we could estimate λ as

$$\lambda = \frac{(12)(5)}{12-5}$$

$$\lambda = 8.59$$

the availability of the book could be determined by substituting [9.9] in [9.6]:

$$P_0 = \frac{\mu^2 - \mu N}{\mu^2}$$

$$P_0 = \frac{(12)^2 - (12)(5)}{(12)^2}$$

$$P_0 = .582$$

In this way, the book is estimated to be unavailable slightly more than 58% of the time.

The problem in estimating the demand rate and unavailability from this approach lies in the fact that the charge rate must be estimated from the number of charges of the book for a period of time. Two problems exist with this approach: (1) as Jain (1967), Morse (1965), and others have shown that book use diminishes as time increases.

Since this is so, it is necessary to select a reasonably short period of recent book use history for determining N. As a result, (2) because of the small sample size, the confidence interval estimate for the mean is rather wide. Brownlee (1960 p. 140) shows that it is possible to establish confidence intervals for the population parameter for a Poisson distribution from an observation as follows:

$$\mu_U = \frac{1}{2}\chi^2_{1-\alpha/2}\left[2(x_o+1)\right] \qquad [9.10]$$

$$\mu_L = \frac{1}{2}\chi^2_{\alpha/2}\left[2x_o\right] \qquad [9.11]$$

Where μ_U is the upper confidence interval for the population parameter, μ_L is the lower confidence interval for the population parameter, $1-\alpha/2$ is the cumulative probability at the upper limit, $\alpha/2$ is the cumulative probability at the lower limit, x_o is the observation and the term in brackets is the degrees of freedom. χ^2 is the value of Chi Squared from a Chi Squared table at the specified level of confidence and degrees of freedom. For example, in the problem given above, a library observing five charges in the past year would be 95% confident that the mean number of charges for the book is between 1.62 and 11.65 as calculated by:

$$\bar{N}_U = \frac{1}{2} \chi^2_{.975} [2(5+1)]$$

$$= \frac{1}{2} (23.3)$$

$$= 11.65$$

$$\bar{N}_L = \frac{1}{2} \chi^2_{.025} [(2)(5)]$$

$$= \frac{1}{2} (3.25)$$

$$= 1.62$$

It is thus possible from a single observation of the two stage book use process to establish confidence intervals for the mean number of uses of the book. A library could, if it wished, determine the percent unavailability of a book based on the methods developed in this section.

9.5 <u>The Concept of Critical λ</u>

There is a certain demand rate at which it will be impossible for the library to meet its availability goal. This investigator calls this demand rate the "critical λ". Using equations [9.6] and [9.7], it is possible to determine the critical λ for a book based on the library's availability goal. Thus:

$$p_0 = \frac{\mu}{\lambda + \mu}$$

$$\lambda = \frac{\mu - p_0 \mu}{p_0}$$

$$\lambda = \frac{\mu p_1}{p_0}$$

Since $p_1 = 1 - p_0$

Thus, a library with an availability goal of 75% and a μ of 12 per year would have a critical demand rate, λ_c, of

$$\lambda_c = \frac{(12)(.25)}{(.75)}$$

$$\lambda_c = \frac{3}{.75} = 4$$

Any book in the collection which would have a demand rate higher than 4 would be unavailable more than 25% of the time. Further, this book would have a critical charge rate N_c (known demand rate) of $p_o \lambda_c$ and using [9.8]:

$$N_c = \frac{(4)(12)}{12+4} = 3$$

If a book in this library would have a charge rate higher than 3 per year, the library would not meet its availability goal.

9.6 Using the Critical λ Concept

Based on the critical demand rate, λ_c, and the critical charge rate, N_c, the library could annually statistically determine which books are not available more than $100p_o$% of the time in the following manner. The library wants to test a null hypothesis $N = N_c$ and in so doing is testing the hypothesis that $\lambda = \lambda_c$. They do this by observing the number of charges, x_o, which the book had in the past year and use the following test statistic:

$$P_o\{x \leq x_o\} = 1 - P\{\chi^2[2(x_o+1)] < 2N_c\} \qquad [9.12]$$

Where x_o is the observed number of charges, χ^2 is the value of Chi Squared from a Chi Squared table, the term in brackets is the number of degrees of freedom, and N is the critical demand rate. To illustrate the use of the test, suppose a library had a μ = 12, p_o = .75, λ_c = 4, N_c = 3 and observed x_o = 7. We calculate the probability of getting x = 7 by using [9.12] and thus:

$$P\{x \leq 7 \mid N=3\} = 1 - P\{[\chi^2[16]\} < 6\}$$

$$P\{x \leq 7 \mid N=3\} = 1 - .014$$

$$P\{x \leq 7 \mid N=3\} = .986$$

Thus the probability of observing 7 charges or less if the book had a charge rate of 3 is .986.

The library could further determine a critical number of charges. That is, that number of charges at which they would conclude that their availability goal is not being met. Suppose the library wants to determine the number of observed charges x_c at which they are willing to conclude that the charge rate is equal to the critical charge rate. Using [9.12],

$$P\{x \leq x_c\} = 1 - P\{\chi^2[2x_c+1] < 2\mu\} \qquad [9.13]$$

We calculate [9.13], for a x_c of 0, 1, 2, ... until the probability is at a level at which the library is willing to risk rejecting the hypothesis that $N = N_c$ when it is true. To illustrate the use of [9.13], assume that a library has a $\lambda_c = 4$, $p_o = .75$, $N_c = 3$ and is willing to take a 5% risk of rejecting the hypothesis that $N_c = 3$ when it is true. We calculate [9.13] for $x_o = 0, 1, 2, 3, 4,$ .... and find:

$$P\{x \leq 0\} = .05$$
$$P\{x \leq 1\} = .20$$
$$P\{x \leq 2\} = .43$$
$$P\{x \leq 3\} = .65$$
$$P\{x \leq 4\} = .82$$
$$P\{x \leq 5\} = .92$$
$$P\{x \leq 6\} = .97$$
$$P\{x \leq 7\} = .99$$

A library then which is trying to maintain the goals stated in this illustration and would not want to reject their hypothesis of an $N_c = 3$ when it is true of .05 (1 - .95) would have to establish a critical number of charges, x_c, of 5. A library using this critical number of charges concept could, upon observing 7 charges of a book, conclude that the demand rate for the book is at least at the critical level λ_c and could adjust their charge rate. They could, for example, cut their loan interval in half. If μ were thus doubled in this manner, the new critical demand rate would be (using Section 9.5):

$$\lambda_c = \frac{(24)(.25)}{(.75)}$$

$$\lambda_c = 8$$

The charge rates for the specific books could be adjusted on an annual (or periodic) basis in this manner.

9.7 Establishing Confidence Intervals for Demand Rates for Books in a Collection

One of the questions which could be answered from this investigation is, "what is the probability of a specific book in a collection having a demand rate exceeding some specific value?" Concluding, as has been done in Chapters 6, 7 and 8, that the number of demands for a group of books is negative binomially distributed permits the conclusion that not only are the number of demands for specific books

Poisson distributed but that the demand rates for these books vary according to a Gamma distribution. Using the Gamma distribution and collected known demand data (such as that of Section 6.1) or of total demand data (such as that of Section 6.2) it is possible to predict the probability of getting demand rates or charge rates of specified values. Prior to doing this, however, it will be worthwhile to look at the Gamma distribution and its peculiarities. From [3.5], λ is distributed as

$$g(\lambda) = \frac{C^r}{\Gamma(r)} e^{-C\lambda} \lambda^{r-1}$$

where the parameters are C and r. r is the shape parameter and C is the scale parameter. Changing C changes the scale of the two axes. When r = 1 the distribution becomes the exponential distribution. The cumulative distribution is

$$F(\lambda) = \int_0^{\lambda} \frac{C^r}{\Gamma(r)} e^{-C\lambda} \lambda^{r-1} d\lambda \qquad [9.14]$$

when $\lambda > 0$ and is 0 when $\lambda \leq 0$. The integral must be evaluated by numerical methods unless r is integer in which case $F(\lambda)$ could be found by:

$$F(\lambda) = 1 - e^{-c\lambda} \sum_{i=0}^{r-1} \frac{(c\lambda)^i}{i!}$$

The r's for the collected demand data are not integer. In order to calculate probabilities from this function, it is necessary to numerically calculate probabilities from [9.14] or use the tables of the Incomplete Gamma Function which were developed by Pearson (1965). The tables present tabular values of an I(v,p) function where:

$$I(v,p) = \frac{\int_0^v e^{-v} v^p \, dv}{\int_0^\infty e^{-v} v^p \, dv} \qquad [9.15]$$

where the I(v,p) function always falls between 0 and 1. Pearson makes a further transformation because of the wide range of the value of the argument v. He uses

$$u = \frac{v}{\sqrt{p+1}}$$

Pearson's tables thus really present

$$I(u,p) = \frac{\int^{u\sqrt{p+1}} e^{-v} v^{p} dv}{\int_{0}^{\infty} e^{-v} v^{p} dv}$$

I(u,p) provides the probability of a deviation exceeding a certain size.

In order to use Pearson's tables, it is necessary to transform [9.14] in the following way. We substitute p + 1 for r and get:

$$F(\lambda) = \int_{0}^{\lambda} \frac{c^{p+1}}{\Gamma(p+1)} e^{-c\lambda} \lambda^{p} d\lambda \qquad [9.16]$$

This transformation prevents a negative factorial in the gamma function. We then let $u = \lambda / \sqrt{p+1}$. We then get:

$$I(u,p) = \frac{\int_{0}^{u\sqrt{p+1}} \frac{c^{p+1} e^{-c\lambda} \lambda^{p}}{\Gamma(p+1)} d\lambda}{\int_{0}^{\infty} \frac{c^{p+1} e^{-c\lambda} \lambda^{p}}{\Gamma(p+1)} d\lambda}$$

which upon cancelling common terms becomes

$$I(u,p) = \frac{\int_{0}^{u\sqrt{p+1}} e^{-\lambda} \lambda^{p} d\lambda}{\Gamma(p+1)}$$

which is identical to [9.15] and we are prepared to use Pearson's tables. We illustrate their use for establishing a confidence interval for the P class demand data (Section 6.2). We concluded that the number of demands for the sample of books in that class was negative binomially distributed with mean = .14344 and variance = .20785. The parameters r and p were calculated to be .41276 and .69011. We want to determine a one sided 95% confidence interval for the variable λ knowing that r = .41276. We consult Pearson's tables to find the value of u for which p = 1.41276 and I(u,p) = .05. We find that:

$$I(.3,1.4) = .0386136$$
$$I(.4,1.4) = .0693255$$
$$I(.3,1.5) = .0334023$$
$$I(.4,1.5) = .0614995$$

Using 2 way interpolation, we calculate u, for .05 to be .33967. We transform u to λ using the transformation $\lambda = u\sqrt{p+1}$ and λ is .5278. Since the λ is the demand rate for only a 51 day period, we extend the rate to a yearly basis, and say that we are 95% confident that books in the P class have demand rates which do not exceed 3.78 per year. The methods of the previous sections could then be used to determine the consequences of such a demand rate.

9.8 Comments and Constraints Concerning Use of the Demand Models

The models and use of the models in this Chapter assume that the number of demands for a book are Poisson distributed, the number of demands for a group of books is negative binomially distributed and that the demand rates vary from book to book according to a Gamma distribution. Sections 9.1 through 9.6 deal with a book in which it is assumed that a single copy of the book is in the collection. The models and their use do not apply to multiple copy books. The establishment of a confidence interval for the demand rate in Section 9.7 does not assume that there is only one copy of the book available. Total demand is independent of the number of copies of the book which are available. Known demand (home use) is not. The models developed in 9.1 through 9.6 further assume that loan intervals are exponentially distributed. This investigator was unable to verify that the loan intervals were exponentially distributed using SUNY/Buffalo data because of the peculiar nature of the loan policy at that institution. The policy is that any book may be charged out for a semester loan subject to recall if it is requested. The minimum loan interval is three weeks. This investigator was unable to fit a known distribution to the SUNY/Buffalo loan interval data although, if anything, the data was possibly Gamma distributed although there was insufficient evidence to reach this conclusion. The SUNY/Buffalo system assumes that a person needing a book will request it if it is unavailable. This is simply not the case. Virtually all of the respondents in the P class demand study did not ask that the book be saved or found when they did not find the book. The SUNY/Buffalo loan interval data showed a mean of about one half of a semester. This would be expected in an institution which has a semester loan policy since we would expect half of the book charges to occur before and half after the midsemester point. Nevertheless, the developed models would be applicable to a library which had a fixed loan interval (2 weeks, 5 weeks, a month, etc.). A study performed at Purdue University* showed that the loan intervals for books at that library were, in fact, exponentially distributed with a mean equal to about the specified loan interval. We would conclude then that the models would be applicable in a typical library environment if the number of demands for a book are Poisson distributed and the loan intervals are exponentially distributed.

*Informal communication from Edward O'Neill presently at the State University of New York at Buffalo.

CHAPTER 10

SUMMARY OF RESULTS, OBSERVATIONS AND RECOMMENDATIONS FOR FURTHER WORK

10.1 Summary of Results and Observations

In proposing this research, this investigator had two objectives: (1) to provide evidence that one cannot reject the hpothesis that the number of demands for books is Poisson distributed and (2) to develop a practical model with which it would be possible to predict demand based on known demand data generated in library circulation systems.

An attempt was made to fulfill objective 1 by collecting data about the number of demands for books and determining how the demands for books were distributed. Since there is so little data available about the demand for any one book to statistically conclue the type of distribution, demands for groups of books were used. The investigation disclosed that the number of demands for groups of books appears to be negative binomially distributed. Reaching this conclusion establishes an experimental basis for concluding that the number of demands for individual books is Poisson distributed and that the demand rate varies from book to book according to a Gamma distribution.

Evidence of this was provided by using the number of demands for a large sample of P class books in the collection of the SUNY-Buffalo libraries and by using a large random sample of books from the entire SUNY-Buffalo library collection. Both these samples appeared to have Negative Binomially distributed demands, although the P class demand sample provided weaker evidence with the Chi Squared goodness of fit test. A further analysis of the number of demands per day for samples of P class demanded books showed that the demand per day for these samples was distributed according to Poisson distributions. As has been shown, the results of the analyses on all three sets of data permits the inference that the number of demands for individual books is, indeed, Poisson distributed. Using the sample from the entire collection permits the conclusion that notonly do P class books exhibit Poisson distributed demand properties but that there is evidence that the demand for all books is Poisson distributed. To permit generalization of these results to other libraries, 18 samples of book use data in several subject areas at four other libraries showed that, for the most part, with the exception of economics books, the number of uses for groups of books at other libraries was Negative Binomially distributed permitting some generalization of the results to other subject areas and libraries.

Although this investigator wanted to develop a practical model for predicting demand for individual books (Objective 2), the results here provide a good mechanism for predicting demand for collections of books. By accepting the Negative Binomial distribution as an adequate representation of the distribution of demands for a collection of books, any library can, with the collection of demand information about a small sample of books, draw inferences about the demand for the entire collection. This was demonstrated in Chapter 9 using the Gamma Distribution. The investigation also showed that it is possible to predict the demand rate for a book by using an observation of the number of known demands for a period of time although the confidence interval estimate for the demand rates are somewhat broad. A more practical approach may depend on a library availability goal and the predetermination of a critical demand rate. Using the critical demand rate permits testing a hypothesis that the demand rate is at a critical level and provides an action point for permitting library managers to alter their loan interval or add a copy of a book. The prediction of a demand rate as developed in this research is based on the assumption that only one copy of a book is available.

10.2 Recommendations for Further Work

This research has merely presented a beginning for demand model development. The results presented here are based on Poisson distributed demand and exponentially distributed loan intervals. Further work should be performed in order to determine the nature of the distribution of loan intervals so that models could be developed for predicting demand for other distributions of loan intervals. Investigations should also be performed for predicting demand based on the availability of multiple copies of a particular book. The effects of deviations, in practice, from Poisson distributed demand and exponentially distributed service times should also be examined. Such an investigation may show that it is not necessary to collect extensive loan interval or demand data because, perhaps, an assumption of a Poisson distribution for demand and an exponential distribution for loan intervals may be more than adequate for a library's purpose in attempting to determine book availability or unavailability. Further impetus is given to this recommendation by the fact that collecting data concerning demand for books in libraries is tedious, time-consuming and expensive.

REFERENCES

1. Cochran, E. G. (1954). "Some Methods for Strengthening the Common Tests." Biometrics. Vol. 10, p. 417.

2. Dawson, C. S. (1962). A Markov Model for Library Circulation. Unpublished M. S. Thesis. Massachusetts Institute of Technology.

3. Dawson, C. S. Aldrin, E. E. and Gould, E. P. (1962). Increasing Effectiveness of the M.I.T. Science Library by the Use of Circulation Statistics. Unpublished Study. Massachusetts Institute of Technology.

4. Deweese, L. C. (1967). A Bibliography of Use Studies. Reprinted in: Jain, A. K. (1967). A Statistical Study of Book Use. Unpublished Ph.D. Dissertation. Purdue University.

5. Feller, W. F. (1957). An Introduction to Probability Theory and Its Applications. New York, John Wiley and Sons.

6. Fussler, H. H. and Simon, J. L. (1961). Patterns in the Use of Books in Large Research Libraries. The University of Chicago Library.

7. Greenwood, M. and Yule, G. U. (1920). "An Inquiry into the Nature of Frequency Distributions Representative of Multiple Happenings with Particular Reference to Multiple Attacks of Disease or of Repeated Accidents. <u>Journal of the Royal Statistical Society</u>. Vol. 83, p. 255.

8. Haldane, J. B. S. (1949). "The Fitting of Binomial Distributions." <u>Annals of Eugenics</u>. Vol. 11, p. 179.

9. Jain, A. K. (1966). "Sampling and Short-Period Usage in the Purdue Library." <u>College and Research Libraries</u>. Vol. 27, p. 211.

10. Jain, A. K. (1967). <u>A Statistical Study of Book Use</u>. Unpublished Ph.D. Dissertation. Purdue University.

11. Kendall, D. G. (1949). "Stochastic Processes and Population Growth." <u>Journal of the Royal Statistical Society</u>, B. Vol. 11, p. 230.

12. Leimkuhler, F. F. (1966). "Systems Analysis in University Libraries." <u>College and Research Libraries</u>. Vol. 27, p. 13.

13. Lindgren, B. W. and McElrath, G. W. (1959). <u>Introduction to Probability and Statistics</u>. New York, The Macmillan Company.

14. Meier, F. A. (1969). "Non-Normal Statistical Distributions and Their Use in Industrial Engineering." Technical Papers of the Twentieth Institute Conference and Convention of the American Institute of Industrial Engineers. Shamrock Hilton Hotel. Houston, Texas.

15. Mood, A. M. and Graybill, F. A. (1963). Introduction to the Theory of Statistics. New York. The Macmillan Company.

16. Morse, P. M. (1963). "Probabilistic Models for Library Operations." Unpublished Paper. Massachusetts Institute of Technology.

17. Morse, P. M. (1965). "On the Prediction of Library Use." Appendix N. Intrex. The M.I.T. Press.

18. Morse, P. M. (1968). Library Effectiveness. The M.I.T. Press.

19. Pearson, K. (1965). Tables of the Incomplete Gamma Function. Cambridge. Cambridge University Press.

20. Pourny, J. L. (1962). Missing Books - Non Available Books. Unpublished Study. Massachusetts Institute of Technology.

21. Ranck, S. H. (1911). "The Problem of the Unused Book." Library Journal. Vol. 36, p. 428.

22. Rothkopf, M. (1962). Two Phases of an Operations Research Study of M.I.T.'s Science Library. Unpublished Study. M.I.T.

23. Shaffer, B. and Ernst, M. (1954). A Survey of Circulation Characteristics of Some General Library Books. Unpublished Study. Massachusetts Institute of Technology.

24. Sichel, H. S. (1951). "The Estimation of the Parameters of a Negative Binomial Distribution with Special Reference to Psychological Data." Psychometrika. Vol. 16, p. 107.

25. Trueswell (1964). "Two Characteristics of Circulation Control and Their Effects on the Implementation of Mechanized Circulation Control Systems." College and Research Libraries. Vol. 25, p. 285.

RELATED REFERENCES

1. Anscombe, F. J. (1948). "The Transformation of Poisson, Binomial and Negative Binomial Data." Biometrika. Vol. 35, p. 246.

2. Anscombe, F. J. (1949). "The Statistical Analysis of Insect Counts Based on the Negative Binomial Distribution." Biometrics. Vol. 5, p. 164.

3. Anscombe, F. J. (1950). "Sampling Theory of the Negative Binomial and Logarithmic Series Distributions." Biometrika. Vol. 37, p. 358.

4. Bartko, J. J. (1961). "The Negative Binomial Distribution." Virginia Journal of Science. p. 18, January, 1961.

5. Bliss, C. I. (1953). "Fitting the Negative Binomial Distribution to Biological Data." Biometrics. Vol. 9, p. 177.

6. Brownlee, K. A. (1961). Statistical Theory and Methodology in Science and Engineering. New York. John Wiley and Sons.

7. Cohen, A. C. (1960). "Estimating the Parameter in a Conditional Poisson Distribution." Biometrics. Vol. 16, p. 203.

8. Cohen, A. C. (1960). "An Extension of a Truncated Poisson Distribution." Biometrics. Vol. 16, p. 446.

9. Cornell, R. G. (1965). "Spearman Estimation For a Simple Exponential Model." Biometrics. Vol. 24, p. 859.

10. Cornell, R. G. and Speckman, J. A. (1967). "Estimation For a Simple Exponential Model." Biometrics. Vol. 26, p. 717.

11. David, F. N. and Johnson, N. L. (1952). "The Truncated Poisson." Biometrics. Vol. 8, p. 275.

12. Doss, S.A.D.C. (1963). "On the Efficiency of Best Asympototically Normal Estimates of the Poisson Parameter Based on Singly and Doubly Truncated or Censored Samples." Biometrics. Vol. 19, p. 588.

13. Finney, D. J. and Vorley, G. C. (1955). "An Example of the Truncated Poisson Distribution." Biometrics. Vol. 11, p. 387.

14. Fisher, R. A. (1941). "The Negative Binomial Distribution." Annals of Eugenics Vol. 11, p. 179.

15. Fisher, R. A. (1964). "The Significance of Deviations from Expectations in a Poisson Series." Biometrics. Vol. 20, p. 265.

16. Hinz, P. and Garland, J. (1967). "Simplified Techniques for Estimating Parameters of Some Generalized Poisson Distributions." Biometrika. Vol. 54, p. 555.

17. Katti, S. K. and Gurland, J. (1961). "The Poisson Pascal Distribution." Biometrics. Vol. 17, p. 527.

18. Lancaster, H. O. (1950). "The Exact Partition of and Its Application to the Problem of the Pooling of Small Expectations. Biometrika. Vol. 37, p. 367.

19. Lewis, P. A. W. (1965). "Some Results on Tests for Poisson Process." Biometrika. Vol. 52, p. 67.

20. Moore, P. G. (1952). "The Estimation of the Poisson Parameter from a Truncated Distribution." Biometrika. Vol. 39, p. 247.

21. Newbold, E. M. (1927). "Practical Applications of the Statistics of Repeated Events, Particularly to Industrial Accidents." Journal of the Royal Statistical Society. Vol. 90, p. 487.

22. Plackett, R. L. (1953). "The Truncated Poisson Distribution." Biometrics. Vol. 9, p. 485.

23. Rao, C. R. and Chakravarti. (1956). "Some Small Sample Tests of a Significance for a Poisson Distribution." Biometrics. Vol. 12, p. 264.

24. Sprott, D. A. (1958). "The Method of Maximum Likelihood Applied to the Poisson Binomial Distribution. Biometrics. Vol. 14, p. 97.

25. Student. (1925). "An Explanation of Deviations from Poisson's Law in Practice. Biometrika. Vol. 12, p. 211.

21. Taylor, C. J. (1961). "The Application of the Negative Binomial Distribution to Stock Control Problems." Operational Research Quarterly. Vol. 12, p. 81.

22. Watson, G. S. (1959). "Some Recent Results in Chi-Square Goodness-of-Fit Tests." Biometrics. Vol. 15, p. 440.

23. Whitaker, L. (1914). "On the Poisson Law of Small Numbers." Biometrika. Vol. 10, p. 36.

APPENDIX A

QUESTIONNAIRE FOR P CLASS BOOK DEMANDS

Date ____________________ Time ____________________

THE RESULTS OF THIS QUESTIONNAIRE WILL HELP THE LIBRARY IMPROVE ITS SERVICE. YOUR COOPERATION WILL BE APPRECIATED.

For each book which you looked for, found or used, list the complete call number on the left and circle your answer to each of the three questions on the right. Please circle the answer to the question even if you did not find the book you wanted.

Call Number	Question	Answer
______________	1) Did you have the call number when you came into the stacks?	YES NO
______________	2) If you had the call number, did you find the book you looked for?	YES NO
______________	3) If you found (or would have found) the book, will you (would you)?	Use it in the Library Charge it Out Not sure if I will(would) charge it out or not

Call Number	Question	Answer
______________	1) Did you have the call number when you came into the stacks?	YES NO
______________	2) If you had the call number, did you find the book you looked for.	YES NO
______________	3) If you found (or would have found) the book, will you (would you)?	Use it in the Library Charge it Out Not sure if I will(would) charge it out or not

Call Number	Question	Answer
______________	1) Did you have the call number when you came into the stacks?	YES NO
______________	2) If you had the call number, did you find the book you looked for?	YES NO
______________	3) If you found (or would have found) the book, will you (would you)?	Use it in the Library Charge it Out Not sure if I will(would) charge it out or not

Call Number	Question	Answer
______________	1) Did you have the call number, when you came into the stacks?	YES NO
______________	2) If you had the call number, did your find the book you looked for?	YES NO
______________	3) If you found (or would have found) the book, will you (would you)?	Use it in the Library Charge it Out Not sure if I will(would) charge it out or not

APPENDIX B

INSTRUCTIONS TO DATA COLLECTORS

1) Assign a patron number to a sheet, record the date and time and hand the sheet to the patron. Tell him that if he needs additional sheets that they are available.

   Patrons are to be assigned numbers sequentially beginning in the morning of each day with "001."

   NOTE: A patron is to be assigned a number even if he refuses a sheet. The word "Refused" is to be written in the top right hand corner of the sheet of a patron who refuses to take one or who turns it in uncompleted.

2) When the sheet is turned in, the data collector is to scan the sheet and
   a) If the sheet contains no call numbers beginning with "P", he is to write the phrase, "Not Applicable" in the top right hand corner of the sheet.
   b) If the sheet contains call numbers beginning with "P" but call numbers are not complete; the handwriting is illegible, or it is clear that the patron did not answer the questions properly, the phrase "Can't Use" is to be written on the sheet.

3) After the patron has filled out and turned in the data sheet, the data collector is to enter the appropriate data on an 80 column data sheet. The following codes are to be used.

Columns	
1-10	Classification Letters (e.g: PB, PA, PQ, P)
11-20	Classification (e.g: P23, PB15, PQ2673, PA1)
21-30	Second Line of Call Number (e.g: 04Z73, T86, G5, G8G83)
31-40	Third Line of Call Number if there is one
41-50	Fourth Line of Call Number
51-60	Leave Blank
61&62	Day of the Study (e.g: 15, 19, 63) Note: This is not the day of the month. The first day of the study was day 01. The last day is day 70.
63,64,65,66	The time of the day on the 24 hour clock. (e.g: 1213, 1516, 2202) Note: The 24 hour clock begins at midnight, one minute after midnight would be 0001, 1:00A.M. would be 0100, 3:15 in the afternoon would be 1515, 8:00 P.M. would be 1600.
71	Enter a "1" if the answer to question 1 was "YES." Enter "2" if it was "NO."

Columns	
72	Enter "1" if the answer to question 2 was "YES." Enter a "2" if the answer to question 2 was "NO."
73	Enter "1" if the answer to question 3 was "Use it in the Library." Enter a "2" if the answer was "Charge it Out." Enter "3" if the answer was "Not sure if I will(would) charge it out or not."
74	Leave Blank
75	Enter "1" if the patron's data sheet was useable. Enter "2" if the patron's data sheet contained no "P" classifications. Enter "3" if the data sheet was unuseable but contained "P" classifications. Enter "4" if the person refused to cooperate. Enter "5" if the person browsed but used no books.
78,79,80	The patron's number.

4) The data collector is to initial each data sheet after he has transferred the information to the 80 column summary sheet.

5) It is important to Print Clearly on the data sheet and to distinguish between the letter 2 and the numeral 2. 2's written as Z look like Z's. Therefore, always cross hatch the Z e.g. Ƶ. Also the letter I sometimes looks like a 1. Therefore, always write the I as I. The letter O is written as Φ.

Examples

Attached is a Xerox Copy of some entries on a sample 80 column summary sheet.

Examples:

1st Line

Patron 001 listed the book PR 3142
G6

Columns

61–62 It was the 15th day of the study
63–66 It was 7:20 A.M.
71–73 His answers to the three questions were: "YES", "YES", and "Charge it Out"
75 His data sheet was useable

2nd – 7th Line

Patron 002 listed six books with the call numbers shown.

Columns

61–62 It was the 15th day of the study
63–66 It was 8:22 A.M.
71–73 His answers were: "YES", "NO", "Use in Library"
75 His data sheets were useable.

8th Line

Patron 005 refused to cooperate

Column

75 Coding "4" indicates he refused to cooperate

9th Line

Patron 003 turned in a data sheet which contained no "P" classification. He arrived on the 15th day at 11:20 A.M.

10th Line

Patron 004 turned in a useable data sheet with one "P" item on it and one "J" item on it. No entry is made for the "J" item. The call number of the "P" item was:

PQ56
A123Z56
Vol 2
1965

He arrived on the 15th day at 1:20 P.M. He didn't answer the Ist and 2nd questions, but answered the 3rd question, "Not Sure if I will(would) charge it out or not." His data sheet was useable.

11th Line

Patron 006 turned in a data sheet which contained the following call number:

PE42

This is not a complete call number and is unuseable. He arrived at 3:10 P.M. on the 15th day.

6) The last data collector working each day will take the 80 column data sheets and the questionnaires for that day and put them all in a brown envelope. He will mark the number of the day on the envelope. The circulation data collector will pick this up the following morning.

7) If in doubt about the correct day number

Monday	March	4 = day 2
"	"	11 = day 9
"	"	18 = day 16
"	"	25 = day 23
"	April	1 = day 30
"	"	8 = day 37
"	"	15 = day 44
"	"	22 = day 51
"	"	29 = day 58
"	May	6 = day 65
Saturday	May	11 = day 70 (last day)

Circulation Data Collector

1) Pick up the 80 column summary sheet downstairs.

2) From the circulation desk you will receive:

   a) A stack of "P" classification charge cards
   b) A list of "P" classification "saves" and "traces"
   c) A list of "P" classification saved books which were charged out.

3) You are to compare the charge cards with the 80 column list. If the item was charged out, you are to enter a "1" in column 74. If it was not charged out you are to enter a "2" in column 74.

4) You will probably have some cards for books which were not listed on the 80 column summary sheet. If you do, fill out a small data sheet for each book. Do not enter any information on the 80 column summary sheet.

5) You should then have:

   a) 80 column summary sheets
   b) A list of "P" classification saves and traces
   c) A list of "P" classification saved books which were charged out
   d) A small stock of small data sheets from (4) above
   e) The questionnaires corresponding to the 80 column data sheets.

6) Put all the information in a brown envelope and deliver to Ralph Hall.

SAMPLE

IBM

FORTRAN Coding Form

FORTRAN STATEMENT

PR	PR3142	G6			150720	112		1	001
PR	PR5853	P4			150822	221		1	002
PR	PR5851	Z3			150822	121		1	002
PR	PR5851	K5			150822	121		1	002
PR	PR5851	A12			150822	121		1	002
PR	PR5819	N1L1			150822	121		1	002
PR	PR56	S7			150822	121		1	002
Refused					151402			4	005
NO P's					151120			2	003
PQ	PQ56	A123Z56	VOL2	1965	151320		3	1	004
Unreachable					151510			3	006

APPENDIX C

PUNCHED CARD FORMAT

Columns	
1 - 4	Classification letter
11 - 20	Classification
21 - 30	2nd line of call number
31 - 40	3rd line of call number
41 - 50	4th line of call number
51 - 60	Last name or student number of patron
61 - 62	Day of the study
63 - 66	Time of the day
71	1 if patron had call number 2 if patron did not have call number
72	1 if desired book was found 2 if desired book was not found
73	1 if patron said he would use it in the library 2 if patron said they would charge it out 3 if patron did not know
74	1 if book was charged out 2 if book was not charged out
75	0 if book was charged out and did not appear on the questionnaire 1 if the questionnaire was useable 2 if no P's on questionnaire 3 if the questionnaire was not useable 4 if questionnaire was refused 5 if patron was a browser and found nothing 6 if request that a book be saved 7 if request that a book be traced 8 if a book from the save shelf was charged out 9 if a book was charged out but not recorded on the questionnaire

APPENDIX D

TABLE USED TO CALCULATE CHI SQUARED FOR FITTING THE NEGATIVE BINOMIAL DISTRIBUTION TO THE SHELF LIST SAMPLE FROM THE SUNY-BUFFALO COLLECTION

The table in this appendix shows the cell value, cell frequency, theoretical probability, theoretical cell frequency, difference between the observed cell frequency and theoretical frequency, the square of the difference, and the component of Chi Squared for each cell for the shelf list sample data in Chapter 6.

TABLE D.1

CHI SQUARED TEST TABLE FOR KNOWN DEMAND SAMPLE
COLLECTED FOR THE SUNY - BUFFALO COLLECTION
DURING THE PERIOD SEPTEMBER 1, 1966 - AUGUST 31, 1967

VALUE	OBSERVED FREQUENCY	PROBABILITY	THEORETICAL FREQUENCY	DIFFERENCE	CHI SQUARED
0	530	.74370	522.077	7.9226	.1202
1	112	.16345	114.742	-2.7419	.0655
2	42	.05612	39.396	2.6038	.1721
3	9	.02158	15.149	-6.1492	2.4960
4	4	.00874	6.135	-2.1355	.7433
5	1	.00365	2.562	-1.5623	.9526
6	4	.00276	1.938	2.0625	2.1955
	702	1.00000	702.000	.0000	6.7452

APPENDIX E

TABLE USED TO CALCULATE CHI SQUARED FOR THE VARIANCE TEST ON THE SHELF LIST SAMPLE FROM THE SUNY-BUFFALO COLLECTION

The table in this appendix shows the cell value, cell frequency, the square of the difference of the cell value and mean, the frequency of occurrence of the cell value multiplied by the difference, and the component of Chi Squared for the cell for the shelf list sample in Chapter 6.

TABLE E.1

VARIANCE TEST TABLE FOR KNOWN DEMAND SAMPLE
COLLECTED FOR THE SUNY - BUFFALO COLLECTION
DURING THE PERIOD SEPTEMBER 1, 1966 - AUGUST 31, 1967

i	f_i	$(i - m)^2$	$f_i(i - m)^2$	$f_i(i - m)^2/m$
0	530	.145	76.935	201.930
1	112	.383	42.914	112.635
2	42	2.621	110.089	288.947
3	9	6.859	61.732	162.027
4	4	13.097	52.389	137.503
5	1	21.335	21.335	55.998
6	3	31.573	94.719	248.608
7	1	43.811	43.811	114.990
				1322.638

APPENDIX F

TABLE USED TO CALCULATE CHI SQUARED FOR FITTING THE NEGATIVE BINOMIAL DISTRIBUTION TO THE SAMPLE OF P CLASS BOOKS

The following table shows the cell value, cell frequency, theoretical probabilities, theoretical cell frequencies, the difference between the observed and theoretical cell frequencies, and the component of Chi Squared for each cell for the P class demand data in Chapter 7.

TABLE F.1

CHI SQUARED TEST TABLE FOR P CLASS SAMPLE COLLECTED FOR THE SUNY - BUFFALO P CLASS COLLECTION DURING THE PERIOD MARCH 3, 1968 - MAY 11, 1968

VALUE	OBSERVED FREQUENCY	PROBABILITY	THEORETICAL FREQUENCY	DIFFERENCE	CHI SQUARED
0	650	.85805	628.093	21.9074	.7641
1	65	.10975	80.337	-15.3370	2.9280
2	13	.02402	17.583	-4.5826	1.1944
3	2	.00599	4.385	-2.3847	1.2969
4	2	.00219	1.603	.3969	.0983
	732	1.00000	732.000	.0000	6.2817

APPENDIX G

TABLE USED TO CALCULATE CHI SQUARED FOR THE VARIANCE TEST ON THE RANDOM SAMPLE OF P CL^^S BOOKS

The following table shows the cell value, cell frequency, the square of the difference between the cell value and the mean, the frequency of the cell value multiplie' by the difference, and the component of Chi Squared for the cell for the data in Chapter 6.

TABLE G.1

VARIANCE TEST TABLE FOR P CLASS DEMAND SAMPLE COLLECTED FOR THE SUNY - BUFFALO P CLASS COLLECTION DURING THE PERIOD MARCH 3, 1968 - MAY 11, 1968

i	f_i	$(i - m)^2$	$f_i(i - m)^2$	$f_i(i - m)^2/m$
0	650	.020	13.292	92.950
1	65	.734	47.739	333.840
2	13	3.448	44.830	313.495
3	2	8.162	16.325	114.160
4	2	14.876	29.753	208.062
				1062.508

APPENDIX H

TABLES USED TO CALCULATE CHI SQUARED FOR FITTING THE POISSON DISTRIBUTION TO THE RANDOM SAMPLES OF P CLASS DEMANDED BOOKS

The following tables show the cell value, cell frequency, theoretical probabilities, theoretical cell frequencies, the difference between the observed and theoretical cell frequencies, the square of the difference, and the component of Chi Squared for each cell for the data in Chapter 7.

TABLE H.1

CHI SQUARED GOODNESS OF FIT TABLE FOR FITTING THE POISSON DISTRIBUTION TO SAMPLE 1 OF P CLASS DEMANDS FOR THE PERIOD MARCH 3, 1968 - MAY 11, 1968

VALUE	OBSERVED FREQUENCY	PROBABILITY	THEORETICAL FREQUENCY	DIFFERENCE	CHI SQUARED
0	2	.07226	3.685	-1.6853	.7707
1	9	.18986	9.683	-.6829	.0482
2	16	.24943	12.721	3.2791	.8452
3	13	.21845	11.141	1.8591	.3102
4	6	.14349	7.318	-1.3180	.2374
5	2	.07540	3.845	-1.8454	.8856
6	3	.05111	2.607	.3934	.0594
	51	1.00000	51.000	.0000	3.1566

TABLE H.2

CHI SQUARED GOODNESS OF FIT TABLE FOR FITTING THE POISSON DISTRIBUTION TO SAMPLE 2 OF P CLASS DEMANDS FOR THE PERIOD MARCH 3, 1968 - MAY 11, 1968

VALUE	OBSERVED FREQUENCY	PROBABILITY	THEORETICAL FREQUENCY	DIFFERENCE	CHI SQUARED
0	5	.08128	4.145	.8547	.1762
1	14	.20400	10.404	3.5960	1.2429
2	10	.25600	13.056	-3.0560	.7153
3	9	.21417	10.923	-1.9227	.3384
4	6	.13438	6.853	-.8534	.1063
5	3	.06745	3.440	-.4399	.0563
6	4	.04272	2.179	1.8213	1.5225
	51	1.00000	51.000	.0000	4.1579

TABLE H.3

CHI SQUARED GOODNESS OF FIT TABLE FOR FITTING THE POISSON DISTRIBUTION TO SAMPLE 3 OF P CLASS DEMANDS FOR THE PERIOD MARCH 3, 1968 - MAY 11, 1968

VALUE	OBSERVED FREQUENCY	PROBABILITY	THEORETICAL FREQUENCY	DIFFERENCE	CHI SQUARED
0	4	.05384	2.746	1.2542	.5728
1	10	.15732	8.023	1.9767	.4870
2	13	.22981	11.720	1.2797	.1397
3	6	.22380	11.414	-5.4138	2.5679
4	7	.16346	8.336	-1.3365	.2143
5	7	.09551	4.871	2.1290	.9305
6	2	.04650	2.371	-.3715	.0582
7	2	.02976	1.518	.4822	.1532
	51	1.00000	51.000	.0000	5.1236

TABLE H.4

CHI SQUARED GOODNESS OF FIT TABLE FOR FITTING THE POISSON DISTRIBUTION TO SAMPLE 4 OF P CLASS DEMANDS FOR THE PERIOD MARCH 3, 1968 - MAY 11, 1968

VALUE	OBSERVED FREQUENCY	PROBABILITY	THEORETICAL FREQUENCY	DIFFERENCE	CHI SQUARED
0	7	.10696	5.455	1.5450	.4376
1	14	.23908	12.193	1.8069	.2678
2	10	.26721	13.628	-3.6277	.9657
3	8	.19910	10.154	-2.1541	.4570
4	6	.11126	5.674	.3257	.0187
5	4	.04974	2.537	1.4633	.8440
6	2	.02665	1.359	.6409	.3022
	51	1.00000	51.000	.0000	3.2930

TABLE H.5

CHI SQUARED GOODNESS OF FIT TABLE FOR FITTING
THE POISSON DISTRIBUTION TO SAMPLE 5
OF P CLASS DEMANDS FOR THE PERIOD
MARCH 3, 1968 - MAY 11, 1963

VALUE	OBSERVED FREQUENCY	PROBABILITY	THEORETICAL FREQUENCY	DIFFERENCE	CHI SQUARED
0	8	.08791	4.483	3.5166	2.7583
1	12	.21375	10.901	1.0987	.1107
2	8	.25986	13.253	-5.2529	2.0820
3	11	.21060	10.741	.2594	.0063
4	7	.12801	6.529	.4715	.0341
5	1	.06225	3.175	-2.1747	1.4897
6	4	.03762	1.919	2.0814	2.2579
	51	1.00000	51.000	.0000	8.7390

TABLE H.6

CHI SQUARED GOODNESS OF FIT TABLE FOR FITTING
THE POISSON DISTRIBUTION TO SAMPLE 6
OF P CLASS DEMANDS FOR THE PERIOD
MARCH 3, 1968 - MAY 11, 1968

VALUE	OBSERVED FREQUENCY	PROBABILITY	THEORETICAL FREQUENCY	DIFFERENCE	CHI SQUARED
0	8	.09697	4.945	3.0545	1.8866
1	10	.22626	11.539	-1.5393	.2053
2	11	.26397	13.462	-2.4625	.4504
3	10	.20531	10.471	-.4708	.0212
4	5	.11976	6.108	-1.1078	.2009
5	6	.05589	2.850	3.1496	3.4802
6	1	.03184	1.624	-.6238	.2397
	51	1.00000	51.000	.0000	6.4843

TABLE H.7

CHI SQUARED GOODNESS OF FIT TABLE FOR FITTING
THE POISSON DISTRIBUTION TO SAMPLE 7
OF P CLASS DEMANDS FOR THE PERIOD
MARCH 3, 1968 - MAY 11, 1968

VALUE	OBSERVED FREQUENCY	PROBABILITY	THEORETICAL FREQUENCY	DIFFERENCE	CHI SQUARED
0	7	.09889	5.043	1.9566	.7591
1	10	.22880	11.669	-1.6688	.2387
2	13	.26470	13.500	-.4997	.0185
3	11	.20414	10.411	.5889	.0333
4	5	.11808	6.022	-1.0221	.1735
5	3	.05464	2.787	.2134	.0163
6	2	.03075	1.568	.4317	.1189
	51	1.00000	51.000	.0000	1.3582

TABLE H.8

CHI SQUARED GOODNESS OF FIT TABLE FOR FITTING
THE POISSON DISTRIBUTION TO SAMPLE 8
OF P CLASS DEMANDS FOR THE PERIOD
MARCH 3, 1968 - MAY 11, 1968

VALUE	OBSERVED FREQUENCY	PROBABILITY	THEORETICAL FREQUENCY	DIFFERENCE	CHI SQUARED
0	9	.08791	4.483	4.5166	4.5500
1	7	.21375	10.901	-3.9013	1.3961
2	13	.25986	13.253	-.2529	.0048
3	11	.21060	10.741	.2594	.0063
4	2	.12801	6.529	-4.5285	3.1412
5	7	.06225	3.175	3.8253	4.6090
6	2	.03762	1.919	.0814	.0035
	51	1.00000	51.000	.0000	13.7109

TABLE H.9

CHI SQUARED GOODNESS OF FIT TABLE FOR FITTING THE POISSON DISTRIBUTION TO SAMPLE 9 OF P CLASS DEMANDS FOR THE PERIOD MARCH 3, 1968 - MAY 11, 1968

VALUE	OBSERVED FREQUENCY	PROBABILITY	THEORETICAL FREQUENCY	DIFFERENCE	CHI SQUARED
0	6	.08965	4.572	1.4279	.4459
1	11	.21623	11.028	-.0277	.0001
2	12	.26075	13.298	-1.2982	.1267
3	9	.20962	10.691	-1.6906	.2674
4	7	.12638	6.445	.5546	.0477
5	3	.06096	3.109	-.1090	.0038
6	3	.03641	1.857	1.1431	.7037
	51	1.00000	51.000	.0000	1.5953

TABLE H.10

CHI SQUARED GOODNESS OF FIT TABLE FOR FITTING THE POISSON DISTRIBUTION TO SAMPLE 10 OF P CLASS DEMANDS FOR THE PERIOD MARCH 3, 1968 - MAY 11, 1968

VALUE	OBSERVED FREQUENCY	PROBABILITY	THEORETICAL FREQUENCY	DIFFERENCE	CHI SQUARED
0	7	.09324	4.755	2.2448	1.0597
1	9	.22122	11.282	-2.2822	.4617
2	14	.26243	13.384	.6161	.0284
3	10	.20754	10.585	-.5845	.0323
4	6	.12310	6.278	-.2781	.0123
5	1	.05841	2.979	-1.9789	1.3146
6	4	.03406	1.737	2.2629	2.9480
	51	1.00000	51.000	.0000	5.8569

APPENDIX I

TABLES USED TO CALCULATE CHI SQUARED FOR THE VARIANCE TEST ON THE RANDOM SAMPLES OF P CLASS DEMANDED BOOKS

The following tables show the cell value, cell frequency, the square of the difference between the cell value and the mean, the frequency of the cell value multiplied by the difference, and the component of Chi Squared for the cell for the data in Chapter 7.

TABLE I.1

VARIANCE TEST TABLE FOR SAMPLE 1
OF THE P CLASS DEMANDS DURING
THE PERIOD MARCH 3, 1968 - MAY 11, 1968

i	f_i	$(i - m)^2$	$f_i(i - m)^2$	$f_i(i - m)^2/m$
0	2	6.903	13.807	5.255
1	9	2.649	23.837	9.072
2	16	.394	6.299	2.397
3	13	.139	1.804	.687
4	6	1.884	11.303	4.302
5	2	5.629	11.258	4.285
6	2	11.374	22.748	8.658
7	1	19.119	19.119	7.277
				41.933

TABLE I.2

VARIANCE TEST TABLE FOR SAMPLE 2
OF THE P CLASS DEMANDS DURING
THE PERIOD MARCH 3, 1968 - MAY 11, 1968

i	f_i	$(i - m)^2$	$f_i(i - m)^2$	$f_i(i - m)^2/m$
0	5	6.299	31.495	12.549
1	14	2.279	31.913	12.715
2	10	.260	2.599	1.036
3	9	.240	2.163	.862
4	6	2.221	13.324	5.309
5	3	6.201	18.603	7.412
6	2	12.181	24.363	9.707
7	1	20.162	20.162	8.033
8	0	30.142	0.000	0.000
9	1	42.123	42.123	16.783
				74.406

TABLE I.3

VARIANCE TEST TABLE FOR SAMPLE 3
OF THE P CLASS DEMANDS DURING
THE PERIOD MARCH 3, 1968 - MAY 11, 1968

i	f_i	$(i - m)^2$	$f_i(i - m)^2$	$f_i(i - m)^2/m$
0	4	8.536	34.142	11.686
1	10	3.692	36.924	12.638
2	13	.849	11.041	3.779
3	6	.006	.037	.013
4	7	1.163	8.141	2.787
5	7	4.320	30.239	10.350
6	2	9.477	18.954	6.487
7	1	16.634	16.634	5.693
8	0	25.791	0.000	0.000
9	0	36.947	0.000	0.000
10	0	50.104	0.000	0.000
11	0	65.261	0.000	0.000
12	0	82.418	0.000	0.000
13	1	101.575	101.575	34.767
				88.202

TABLE I.4

VARIANCE TEST TABLE FOR SAMPLE 4
OF THE P CLASS DEMANDS DURING
THE PERIOD MARCH 3, 1968 - MAY 11, 1968

i	f_i	$(i - m)$	$f_i(i - m)^2$	$f_i(i - m)^2/m$
0	7	4.997	34.976	15.647
1	14	1.526	21.363	9.557
2	10	.055	.554	.248
3	8	.585	4.678	2.093
4	6	3.114	18.685	8.359
5	4	7.644	30.574	13.678
6	2	14.173	28.346	12.681
				62.263

TABLE I.5

VARIANCE TEST TABLE FOR SAMPLE 5
OF THE P CLASS DEMANDS DURING
THE PERIOD MARCH 3, 1968 - MAY 11, 1968

i	f_i	$(i - m)^2$	$f_i(i - m)^2$	$f_i(i - m)^2/m$
0	8	5.912	47.292	19.451
1	12	2.049	24.586	10.112
2	8	.186	1.489	.612
3	11	.323	3.557	1.463
4	7	2.461	17.224	7.084
5	1	6.598	6.598	2.714
6	1	12.735	12.735	5.238
7	1	20.872	20.872	8.585
8	1	31.010	31.010	12.754
9	1	43.147	43.147	17.746
				85.758

TABLE I.6

VARIANCE TEST TABLE FOR SAMPLE 6
OF THE P CLASS DEMANDS DURING
THE PERIOD MARCH 3, 1968 - MAY 11, 1968

i	f_i	$(i - m)^2$	$f_i(i - m)^2$	$f_i(i - m)^2/m$
0	8	5.444	43.555	18.667
1	10	1.778	17.778	7.619
2	11	.111	1.222	.524
3	10	.444	4.444	1.905
4	5	2.778	13.889	5.952
5	6	7.111	42.667	18.286
6	0	13.444	0.000	0.000
7	1	21.778	21.778	9.333
				62.286

TABLE I.7

VARIANCE TEST TABLE FOR SAMPLE 7
OF THE P CLASS DEMANDS DURING
THE PERIOD MARCH 3, 1968 - MAY 11, 1968

i	f_i	$(i - m)^2$	$f_i(i - m)^2$	$f_i(i - m)^2/m$
0	7	5.353	37.473	16.196
1	10	1.726	17.259	7.459
2	13	.098	1.280	.553
3	11	.471	5.181	2.239
4	5	2.844	14.218	6.145
5	3	7.216	21.648	9.356
6	1	13.589	13.589	5.873
7	0	21.961	0.000	0.000
8	1	32.334	32.334	13.975
				61.796

TABLE I.8

VARIANCE TEST TABLE FOR SAMPLE 8
OF THE P CLASS DEMANDS DURING
THE PERIOD MARCH 3, 1968 - MAY 11, 1968

i	f_i	$(i - m)^2$	$f_i(i - m)^2$	$f_i(i - m)^2/m$
0	9	5.912	53.204	21.882
1	7	2.049	14.342	5.899
2	13	.186	2.419	.995
3	11	.323	3.557	1.463
4	2	2.461	4.921	2.024
5	7	6.598	46.185	18.995
6	0	12.735	0.000	0.000
7	1	20.872	20.872	8.585
8	0	31.010	0.000	0.000
				59.843

TABLE I.9

VARIANCE TEST TABLE FOR SAMPLE 9
OF THE P CLASS DEMANDS DURING
THE PERIOD MARCH 3, 1968 - MAY 11, 1968

i	f_i	$(i - m)^2$	$f_i(i - m)^2$	$f_i(i - m)^2/m$
0	6	5.817	34.900	14.471
1	11	1.993	21.924	9.090
2	12	.170	2.035	.844
3	9	.346	3.114	1.291
4	7	2.523	17.658	7.321
5	3	6.699	20.097	8.333
6	3	12.875	38.626	16.016
				57.366

TABLE I.10

VARIANCE TEST TABLE FOR SAMPLE 10
OF THE P CLASS DEMANDS DURING
THE PERIOD MARCH 3, 1968 - MAY 11, 1968

i	f_i	$(i - m)^2$	$f_i(i - m)^2$	$f_i(i - m)^2/m$
0	7	5.629	39.403	16.608
1	9	1.884	16.955	7.146
2	14	.139	1.943	.819
3	10	.394	3.937	1.659
4	6	2.649	15.892	6.698
5	1	6.903	6.903	2.910
6	3	13.158	39.475	16.638
7	1	21.413	21.413	9.025
				61.504

APPENDIX J

TABLES USED TO CALCULATE CHI SQUARED FOR FITTING THE NEGATIVE BINOMIAL DISTRIBUTION TO THE FUSSLER AND SIMON BOOK USE DATA

The following tables show the cell value, cell frequency, theoretical probabilities, theoretical cell frequencies, the difference between the observed and theoretical cell frequencies, the differences squared, and the component of Chi Squared for each cell for the data in Chapter 8.

TABLE J.1

CHI SQUARED TABLE FOR FITTING
THE NEGATIVE BINOMIAL DISTRIBUTION
TO SAMPLES OF BOOK USE
FOR THE PHILOSOPHY SAMPLE AT CHICAGO
WHEN COMPARED WITH YALE

VALUE	OBSERVED FREQUENCY	PROBABILITY	THEORETICAL FREQUENCY	DIFFERENCE	CHI SQUARED
0	85	.66136	87.300	-2.2995	.0606
1	21	.12826	16.930	4.0697	.9783
2	12	.06562	8.662	3.3382	1.2865
3	3	.04053	5.350	-2.3500	1.0322
4	4	.02717	3.586	.4136	.0477
5+	7	.07706	10.172	-3.1719	.9891
	132	1.00000	132.000	.0000	4.3943

TABLE J.2

CHI SQUARED TABLE FOR FITTING
THE NEGATIVE BINOMIAL DISTRIBUTION
TO SAMPLES OF BOOK USE
FOR THE PHILOSOPHY SAMPLE AT YALE

VALUE	OBSERVED FREQUENCY	PROBABILITY	THEORETICAL FREQUENCY	DIFFERENCE	CHI SQUARED
0	90	.69357	91.5512	-1.5512	.0263
1	18	.10983	14.4975	3.5025	.8462
2	7	.05616	7.4131	-.4131	.0230
3	3	.03532	4.6622	-1.6622	.5926
4	6	.02429	3.2063	2.7937	2.4342
5+	8	.08083	10.6695	-2.6695	.6679
	132	1.00000	132.000	.0000	4.5902

TABLE J.3

CHI SQUARED TABLE FOR FITTING
THE NEGATIVE BINOMIAL DISTRIBUTION
TO SAMPLES OF BOOK USE
FOR THE TEUTONIC LANGUAGES SAMPLE AT CHICAGO
WHEN COMPARED WITH YALE

VALUE	OBSERVED FREQUENCY	PROBABILITY	THEORETICAL FREQUENCY	DIFFERENCE	CHI SQUARED
0	109	.74362	110.799	-1.7994	.0292
1	19	.09162	13.651	5.3486	2.0956
2	7	.04592	6.842	.1579	.0036
3	5	.02880	4.291	.7088	.1171
4	2	.01988	2.962	-.9621	.3125
5+	7	.07016	10.454	-3.4538	1.1411
	149	1.00000	149.000	.0000	3.6992

TABLE J.4

CHI SQUARED TABLE FOR FITTING
THE NEGATIVE BINOMIAL DISTRIBUTION
TO SAMPLES OF BOOK USE
FOR THE TEUTONIC LANGUAGES SAMPLE AT YALE

VALUE	OBSERVED FREQUENCY	PROBABILITY	THEORETICAL FREQUENCY	DIFFERENCE	CHI SQUARED
0	112	.76432	113.884	-1.8837	.0312
1	22	.09542	14.218	7.7824	4.2599
2	6	.04574	6.815	-.8153	.0975
3	3	.02733	4.072	-1.0722	.2823
4	1	.01795	2.675	-1.6745	1.0484
5+	5	.04924	7.337	-2.3368	.7443
	149	1.00000	149.000	.0000	6.4636

TABLE J.5

CHI SQUARED TABLE FOR FITTING
THE NEGATIVE BINOMIAL DISTRIBUTION
TO SAMPLES OF BOOK USE
FOR THE TEUTONIC LANGUAGES SAMPLE AT CHICAGO
WHEN COMPARED WITH NORTHWESTERN

VALUE	OBSERVED FREQUENCY	PROBABILITY	THEORETICAL FREQUENCY	DIFFERENCE	CHI SQUARED
0	39	.64548	39.374	-.3743	.0036
1	5	.10277	6.269	-1.2690	.2569
2	3	.05547	3.384	-.3837	.0435
3	1	.03698	2.256	-1.2558	.6991
4	4	.02700	1.647	2.3530	3.3616
5+	9	.13230	8.070	.9297	.1071
	61	1.00000	61.000	.0000	4.4717

TABLE J.6

CHI SQUARED TABLE FOR FITTING
THE NEGATIVE BINOMIAL DISTRIBUTION
TO SAMPLES OF BOOK USE
FOR THE TEUTONIC LANGUAGES SAMPLE AT NORTHWESTERN

VALUE	OBSERVED FREQUENCY	PROBABILITY	THEORETICAL FREQUENCY	DIFFERENCE	CHI SQUARED
0	46	.76059	46.396	-.3960	.0034
1	7	.09633	5.876	1.1239	.2150
2	2	.04635	2.827	-.8273	.2421
3	3	.02778	1.695	1.3054	1.0056
4	1	.01829	1.116	-.1157	.0120
5+	2	.05066	3.090	-1.0903	.3846
	61	1.00000	61.000	.0000	1.8627

TABLE J.7

CHI SQUARED TABLE FOR FITTING
THE NEGATIVE BINOMIAL DISTRIBUTION
TO SAMPLES OF BOOK USE
FOR THE TEUTONIC LANGUAGES SAMPLE AT CHICAGO
WHEN COMPARED WITH CALIFORNIA

VALUE	OBSERVED FREQUENCY	PROBABILITY	THEORETICAL FREQUENCY	DIFFERENCE	CHI SQUARED
0	125	.75067	126.863	-1.8632	.0274
1	23	.08872	14.994	8.0063	4.2752
2	2	.04439	7.502	-5.5019	4.0351
3	8	.02787	4.710	3.2900	2.2981
4	1	.01927	3.257	-2.2566	1.5637
5+	10	.06908	11.675	-1.6745	.2402
	169	1.00000	169.000	.0000	12.4396

TABLE J.8

CHI SQUARED TABLE FOR FITTING
THE NEGATIVE BINOMIAL DISTRIBUTION
TO SAMPLES OF BOOK USE
FOR THE TEUTONIC LANGUAGES SAMPLE AT CALIFORNIA

VALUE	OBSERVED FREQUENCY	PROBABILITY	THEORETICAL FREQUENCY	DIFFERENCE	CHI SQUARED
0	125	.74685	126.218	-1.2176	.0117
1	21	.10427	17.622	3.3784	.6477
2	7	.05002	8.453	-1.4534	.2499
3	7	.02967	5.014	1.9858	.7864
4	3	.01928	3.258	-.2583	.0205
5+	6	.04991	8.435	-2.4348	.7028
	169	1.00000	169.000	.0000	2.4190

TABLE J.9

CHI SQUARED TABLE FOR FITTING
THE NEGATIVE BINOMIAL DISTRIBUTION
TO SAMPLES OF BOOK USE
FOR THE BIOLOGY SAMPLE AT CHICAGO
WHEN COMPARED WITH YALE

VALUE	OBSERVED FREQUENCY	PROBABILITY	THEORETICAL FREQUENCY	DIFFERENCE	CHI SQUARED
0	81	.71655	82.403	-1.4032	.0239
1	10	.07755	8.918	1.0817	.1312
2	4	.04086	4.699	-.6989	.1040
3	7	.02724	3.133	3.8674	4.7746
4	2	.02005	2.306	-.3057	.0405
5+	11	.11775	13.541	-2.5412	.4769
	115	1.00000	115.000	.0000	5.5511

TABLE J.10

CHI SQUARED TABLE FOR FITTING
THE NEGATIVE BINOMIAL DISTRIBUTION
TO SAMPLES OF BOOK USE
FOR THE BIOLOGY SAMPLE AT YALE

VALUE	OBSERVED FREQUENCY	PROBABILITY	THEORETICAL FREQUENCY	DIFFERENCE	CHI SQUARED
0	89	.78055	89.763	-.7632	.0065
1	7	.05782	6.649	.3507	.0185
2	3	.02988	3.436	-.4362	.0554
3	3	.01986	2.284	.7161	.2245
4	2	.01466	1.686	.3141	.0585
5+	11	.09723	11.181	-.1814	.0029
	115	1.00000	115.000	.0000	.3664

TABLE J.11

CHI SQUARED TABLE FOR FITTING
THE NEGATIVE BINOMIAL DISTRIBUTION
TO SAMPLES OF BOOK USE
FOR THE BIOLOGY SAMPLE AT CHICAGO
WHEN COMPARED WITH CALIFORNIA

VALUE	OBSERVED FREQUENCY	PROBABILITY	THEORETICAL FREQUENCY	DIFFERENCE	CHI SQUARED
0	72	.68776	72.903	-.9026	.0112
1	17	.12406	13.150	3.8496	1.1269
2	7	.06181	6.552	.4481	.0307
3	5	.03735	3.959	1.0409	.2737
4	0	.02455	2.602	-2.6023	2.6023
5+	5	.06447	6.834	-1.8338	.4921
	106	1.00000	106.000	.0000	4.5368

TABLE J.12

CHI SQUARED TABLE FOR FITTING
THE NEGATIVE BINOMIAL DISTRIBUTION
TO SAMPLES OF BOOK USE
FOR THE BIOLOGY SAMPLE AT CALIFORNIA

VALUE	OBSERVED FREQUENCY	PROBABILITY	THEORETICAL FREQUENCY	DIFFERENCE	CHI SQUARED
0	59	.57919	61.394	-2.3941	.0934
1	18	.14443	15.310	2.6904	.4728
2	10	.07880	8.353	1.6472	.3248
3	5	.05077	5.382	-.3816	.0271
4	6	.03521	3.732	2.2677	1.3779
5+	8	.11160	11.830	-3.8296	1.2398
	106	1.00000	106.000	.0000	3.5357

TABLE J.13

CHI SQUARED TABLE FOR FITTING
THE NEGATIVE BINOMIAL DISTRIBUTION
TO SAMPLES OF BOOK USE
FOR THE PHYSICS SAMPLE AT CHICAGO
WHEN COMPARED WITH NORTHWESTERN

VALUE	OBSERVED FREQUENCY	PROBABILITY	THEORETICAL FREQUENCY	DIFFERENCE	CHI SQUARED
0	28	.42207	34.188	-6.1877	1.1199
1	14	.12089	9.792	4.2079	1.8082
2	9	.07420	6.010	2.9898	1.4873
3	4	.05365	4.346	-.3456	.0275
4	6	.04171	3.379	2.6215	2.0341
5+	20	.28748	23.286	-3.2859	.4637
	81	1.00000	81.000	.0000	6.9407

TABLE J.14

CHI SQUARED TABLE FOR FITTING
THE NEGATIVE BINOMIAL DISTRIBUTION
TO SAMPLES OF BOOK USE
FOR THE PHYSICS SAMPLE AT NORTHWESTERN

VALUE	OBSERVED FREQUENCY	PROBABILITY	THEORETICAL FREQUENCY	DIFFERENCE	CHI SQUARED
0	44	.55442	44.908	-.9080	.0184
1	13	.14421	11.681	1.3190	.1489
2	4	.08052	6.522	-2.5221	.9753
3	3	.05297	4.291	-1.2906	.3882
4	3	.03748	3.036	-.0359	.0004
5+	14	.13040	10.562	3.4376	1.1188
	81	1.00000	81.000	.0000	2.6500

TABLE J.15

CHI SQUARED TABLE FOR FITTING THE NEGATIVE BINOMIAL DISTRIBUTION TO SAMPLES OF BOOK USE FOR THE ECONOMICS SAMPLE AT CHICAGO WHEN COMPARED WITH NORTHWESTERN

VALUE	OBSERVED FREQUENCY	PROBABILITY	THEORETICAL FREQUENCY	DIFFERENCE	CHI SQUARED
0	55	.39353	35.024	19.9758	11.3931
1	16	.18809	16.740	-.7400	.0327
2	7	.11859	10.555	-3.5545	1.1971
3	1	.08079	7.190	-6.1903	5.3294
4	0	.05710	5.082	-5.0819	5.0819
5+	10	.16190	14.409	-4.4091	1.3492
	89	1.00000	89.000	.0000	24.3833

TABLE J.16

CHI SQUARED TABLE FOR FITTING THE NEGATIVE BINOMIAL DISTRIBUTION TO SAMPLES OF BOOK USE FOR THE ECONOMICS SAMPLE AT NORTHWESTERN

VALUE	OBSERVED FREQUENCY	PROBABILITY	THEORETICAL FREQUENCY	DIFFERENCE	CHI SQUARED
0	55	.47636	42.396	12.6040	3.7470
1	20	.29814	26.534	-6.5345	1.6092
2	6	.13726	12.216	-6.2161	3.1631
3	2	.05562	4.950	-2.9502	1.7582
4	2	.02101	1.870	.1301	.0091
5+	4	.01161	1.033	2.9667	8.5178
	89	1.00000	89.000	.0000	18.8044

TABLE J.17

CHI SQUARED TABLE FOR FITTING
THE NEGATIVE BINOMIAL DISTRIBUTION
TO SAMPLES OF BOOK USE
FOR THE ECONOMICS SAMPLE AT CHICAGO
WHEN COMPARED WITH CALIFORNIA

VALUE	OBSERVED FREQUENCY	PROBABILITY	THEORETICAL FREQUENCY	DIFFERENCE	CHI SQUARED
0	69	.63967	60.129	8.8710	1.3088
1	16	.24683	23.202	-7.2020	2.2355
2	5	.07968	7.490	-2.4899	.8277
3	2	.02405	2.261	-.2607	.0301
4	0	.00700	.658	-.6580	.6580
5+	2	.00277	.260	1.7396	11.6225
	94	1.00000	94.000	.0000	16.6827

TABLE J.18

CHI SQUARED TABLE FOR FITTING
THE NEGATIVE BINOMIAL DISTRIBUTION
TO SAMPLES OF BOOK USE
FOR THE ECONOMICS SAMPLE AT CALIFORNIA

VALUE	OBSERVED FREQUENCY	PROBABILITY	THEORETICAL FREQUENCY	DIFFERENCE	CHI SQUARED
0	57	.41034	38.572	18.4280	8.8041
1	17	.31876	29.963	-12.9634	5.6085
2	8	.16277	15.300	-7.3004	3.4833
3	3	.06867	6.455	-3.4550	1.8493
4	3	.02593	2.437	.5626	.1298
5+	6	.01353	1.272	4.7282	17.5455
	94	1.00000	94.000	.0000	37.5455

APPENDIX K

TABLES USED TO CALCULATE CHI SQUARED FOR THE VARIANCE TEST FOR THE FUSSLER AND SIMON BOOK USE DATA

The following tables show the cell value, cell frequency, the square of the difference between the cell value and the mean, the frequency of the cell value multiplied by the difference, and the component of Chi Squared for the cell for the data in Chapter 8.

TABLE K.1

VARIANCE TEST TABLE FOR BOOK USE OF THE PHILOSOPHY SAMPLE AT CHICAGO WHEN COMPARED WITH YALE

i	f_i	$(i - m)^2$	$f_i (i - m)^2$	$f_i (i - m)^2/m$
0	85	1.291	109.762	96.591
1	21	.019	.390	.344
2	12	.746	8.950	7.876
3	3	3.473	10.419	9.169
4	4	8.200	32.802	28.866
5	7	14.928	104.494	91.955
				234.800

TABLE K.2

VARIANCE TEST TABLE FOR BOOK USE OF THE PHILOSOPHY SAMPLE AT YALE

i	f_i	$(i - m)^2$	$f_i (i - m)^2$	$f_i (i - m)^2/m$
0	90	1.361	122.496	[illegible]4.988
1	18	.028	.504	.432
2	7	.694	4.858	4.164
3	3	3.361	10.083	8.642
4	6	8.028	48.168	4.129
5	8	14.694	117.552	100.758
				223.113

TABLE K.3

VARIANCE TEST TABLE FOR BOOK USE OF THE TEUTONIC LANGUAGES SAMPLE AT CHICAGO WHEN COMPARED WITH YALE

i	f_i	$(i - m)^2$	$f_i(i - m)^2$	$f_i(i - m)^2/m$
0	109	1.041	113.433	111.194
1	19	.000	.008	.008
2	7	.960	6.721	6.588
3	5	3.920	19.599	19.213
4	2	8.880	17.759	17.409
5	7	15.839	110.876	108.688
				263.099

TABLE K.4

VARIANCE TEST TABLE FOR BOOK USE OF THE TEUTONIC LANGUAGES SAMPLE AT YALE

i	f_i	$(i - m)^2$	$f_i(i - m)^2$	$f_i(i - m)^2/m$
0	112	.565	63.283	84.188
1	22	.062	1.357	1.805
2	6	1.558	9.350	12.439
3	3	5.055	15.165	20.175
4	1	10.552	10.552	14.037
5	5	18.048	90.241	120.053
				252.696

TABLE K.5

VARIANCE TEST TABLE FOR BOOK USE OF THE TEUTONIC LANGUAGES SAMPLE AT CHICAGO WHEN COMPARED WITH NORTHWESTERN

i	f_i	$(i - m)^2$	$f_i (i - m)^2$	$f_i (i - m)^2/m$
0	39	4.000	156.000	78.000
1	5	1.000	5.000	2.500
2	3	0.000	0.000	0.000
3	1	1.000	1.000	.500
4	4	4.000	16.000	8.000
5	9	9.000	81.000	40.500
				129.500

TABLE K.6

VARIANCE TEST TABLE FOR BOOK USE OF THE TEUTONIC LANGUAGES SAMPLE AT NORTHWESTERN

i	f_i	$(i - m)^2$	$f_i (i - m)^2$	$f_i (i - m)^2/m$
0	46	.594	27.308	35.443
1	7	.053	.369	.479
2	2	1.512	3.023	3.924
3	3	4.971	14.912	19.354
4	1	10.430	10.430	13.536
5	2	17.889	35.778	46.435
				119.170

TABLE K.7

VARIANCE TEST TABLE FOR BOOK USE OF THE TEUTONIC LANGUAGES SAMPLE AT CHICAGO WHEN COMPARED WITH CALIFORNIA

i	f_i	$(i - m)^2$	$f_i(i - m)^2$	$f_i(i - m)^2/m$
0	125	1.012	126.484	125.740
1	23	.000	.001	.001
2	2	.988	1.976	1.965
3	8	3.976	31.811	31.624
4	1	8.965	8.965	8.912
5	10	15.953	159.527	158.588
				326.829

TABLE K.8

VARIANCE TEST TABLE FOR BOOK USE OF THE TEUTONIC LANGUAGES SAMPLE AT CALIFORNIA

i	f_i	$(i - m)^2$	$f_i(i - m)^2$	$f_i(i - m)^2/m$
0	125	.601	75.107	96.894
1	21	.051	1.062	1.370
2	7	1.500	10.502	13.548
3	7	4.950	34.650	44.701
4	3	10.400	31.199	40.249
5	6	17.849	107.096	138.162
				334.923

TABLE K.9

VARIANCE TEST TABLE FOR BOOK USE OF THE BIOLOGY SAMPLE AT CHICAGO WHEN COMPARED WITH YALE

i	f_i	$(i - m)^2$	$f_i(i - m)^2$	$f_i(i - m)^2/m$
0	81	3.965	321.187	161.295
1	10	.983	9.827	4.935
2	4	.000	.000	.000
3	7	1.017	7.122	3.577
4	2	4.035	8.070	4.053
5	11	9.052	99.575	50.005
				223.865

TABLE K.10

VARIANCE TEST TABLE FOR BOOK USE OF THE BIOLOGY SAMPLE AT YALE

i	f_i	$(i - m)^2$	$f_i(i - m)^2$	$f_i(i - m)^2/m$
0	89	3.366	299.611	163.295
1	7	.697	4.878	2.659
2	3	.027	.082	.045
3	3	1.358	4.073	2.220
4	2	4.688	9.376	5.110
5	11	10.019	110.205	60.064
				233.393

TABLE K.11

VARIANCE TEST TABLE FOR BOOK USE OF
THE BIOLOGY SAMPLE AT CHICAGO
WHEN COMPARED WITH CALIFORNIA

i	f_i	$(i - m)^2$	$f_i(i - m)^2$	$f_i(i - m)^2/m$
0	72	.963	69.308	70.641
1	17	.000	.006	.006
2	7	1.038	7.267	7.406
3	5	4.076	20.379	20.771
4	0	9.114	0.000	0.000
5	5	16.151	80.757	82.310
				181.135

TABLE K.12

VARIANCE TEST TABLE FOR BOOK USE
OF THE BIOLOGY SAMPLE AT CALIFORNIA

i	f_i	$(i - m)^2$	$f_i(i - m)^2$	$f_i(i - m)^2/m$
0	59	2.482	146.444	92.953
1	18	.331	5.961	3.784
2	10	.180	1.802	1.144
3	5	2.029	10.146	6.440
4	6	5.878	35.270	22.387
5	8	11.727	93.819	59.550
				186.258

TABLE K.13

VARIANCE TEST TABLE FOR BOOK USE OF THE PHYSICS SAMPLE AT CHICAGO WHEN COMPARED WITH NORTHWESTERN

i	f_i	$(i - m)^2$	$f_i(i - m)^2$	$f_i(i - m)^2/m$
0	28	23.781	665.858	136.543
1	14	15.028	210.386	43.142
2	9	8.274	74.470	15.271
3	4	3.521	14.086	2.888
4	6	.768	4.610	.945
5	20	.015	.305	.063
				198.853

TABLE K.14

VARIANCE TEST TABLE FOR BOOK USE OF THE PHYSICS SAMPLE AT NORTHWESTERN

i	f_i	$(i - m)^2$	$f_i(i - m)^2$	$f_i(i - m)^2/m$
0	44	3.294	144.916	79.852
1	13	.664	8.631	4.756
2	4	.034	.137	.076
3	3	1.405	4.214	2.322
4	3	4.775	14.325	7.893
5	14	10.145	142.036	78.265
				173.164

TABLE K.15

VARIANCE TEST TABLE FOR BOOK USE OF THE ECONOMICS SAMPLE AT CHICAGO WHEN COMPARED WITH NORTHWESTERN

i	f_i	$(i - m)^2$	$f_i(i - m)^2$	$f_i(i - m)^2/m$
0	55	4.850	266.745	121.124
1	16	1.445	23.126	10.501
2	7	.041	.286	.130
3	1	.636	.636	.289
4	0	3.232	0.000	0.000
5	10	7.827	78.274	35.543
				167.587

TABLE K.16

VARIANCE TEST TABLE FOR BOOK USE OF THE ECONOMICS SAMPLE AT NORTHWESTERN

i	f_i	$(i - m)^2$	$f_i(i - m)^2$	$f_i(i - m)^2/m$
0	55	.788	43.335	48.820
1	20	.013	.252	.284
2	6	1.237	7.424	8.364
3	2	4.462	8.924	10.054
4	2	9.687	19.374	21.826
5	4	16.912	67.646	76.209
				165.557

TABLE K.17

VARIANCE TEST TABLE FOR BOOK USE OF THE ECONOMICS SAMPLE AT CHICAGO WHEN COMPARED WITH CALIFORNIA

i	f_i	$(i - m)^2$	$f_i (i - m)^2$	$f_i (i - m)^2/m$
0	69	.272	18.750	35.968
1	16	.229	3.667	7.034
2	5	2.187	10.933	20.973
3	2	6.144	12.288	23.573
4	0	12.101	0.000	0.000
5	2	20.059	40.118	76.960
				164.509

TABLE K.18

VARIANCE TEST TABLE FOR BOOK USE OF THE ECONOMICS SAMPLE AT CALIFORNIA

i	f_i	$(i - m)^2$	$f_i (i - m)^2$	$f_i (i - m)^2/m$
0	57	1.065	60.696	58.819
1	17	.001	.017	.017
2	8	.937	7.498	7.266
3	3	3.873	11.620	11.261
4	3	8.810	26.429	25.611
5	6	15.746	94.474	91.553
				194.527