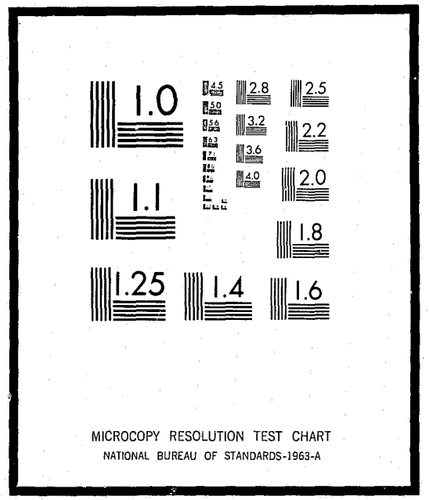
If you have issues viewing or accessing this file contact us at NCJRS.gov.

NCJRS

This microfiche was produced from documents received for inclusion in the NCJRS data base. Since NCJRS cannot exercise control over the physical condition of the documents submitted, the individual frame quality will vary. The resolution chart on this frame may be used to evaluate the document quality.



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

Microfilming procedures used to create this fiche comply with the standards set forth in 41CFR 101-11.504

Points of view or opinions stated in this document are those of the author(s) and do not represent the official position or policies of the U.S. Department of Justice.

U.S. DEPARTMENT OF JUSTICE
LAW ENFORCEMENT ASSISTANCE ADMINISTRATION
NATIONAL CRIMINAL JUSTICE REFERENCE SERVICE
WASHINGTON, D.C. 20531

Date filmed 6/27/75

ST. LOUIS HIGH IMPACT ANTI-CRIME PROGRAM

CONTROL OF REGRESSION ARTIFACT ERROR IN EVALUATING THE EFFECTIVENESS OF CRIME REDUCTION PROGRAMS

BY

Denise Corcoran, M.S.

Nelson B. Heller, Ph.d.

June, 1974

High Impact Evaluation Unit

Missouri Law Enforcement Assistance Council
Region 5
812 Olive Street - Room 1032
St. Louis, Missouri 63101

Floyd D. Richards, Executive Director

14172

CONTROL OF REGRESSION ARTIFACT ERROR IN EVALUATING THE EFFECTIVENESS OF CRIME REDUCTION PROGRAMS

by

Denise Corcoran, M.S., and Nelson B. Heller, Ph.d.

ABSTRACT

A common method of evaluating the results of social programs designed to alleviate specific problems involves a "before and after" comparison of the performance of the treatment group (i.e., clients, or geographic areas being served). When the selection of the treatment group depends on prior performance (eg., high crime, low I.Q., etc.) rather than on a random scheme, this type of evaluation may produce erroneously inflated results in favor of the project's impact, by overestimating the "before" levels. This study presents analytical techniques for estimating the magnitude of this bias, called regression artifact. These techniques were used to analyze the results of the St. Louis High Impact Anti-Crime Program's Foot Patrol Project, which was implemented in 1972 in the highest crime areas in the city.

TABLE OF CONTENTS

TABLE OF CONTENTS
(continued)

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES
(continued)

PREFACE

Considerable attention has recently been focused on the need to evaluate government programs, particularly those aimed at correcting social problems. In the field of law enforcement and criminal justice, Congress has specifically directed the Law Enforcement Assistance Administration, through the Safe Streets Act of 1973, to examine its own and local projects to find out "what works and what doesn't work". Although many criminal justice practitioners and researchers are eager to respond, the art of evaluating crime control programs is still very much in its infancy, and pitfalls abound, likely to mislead even the most well-intentioned evaluators. This study is an examination of one such pitfall, called the "regression artifact", which presents itself in the very common situation in which project treatment resources are administered only to those clients or geographic areas most in need of service. For example, intensive police patrols are usually deployed to high crime areas where the need for crime reduction appears most acute.

Scientists who can perform experiments in laboratories are careful to establish controlled conditions which permit rigorous interpretation of the results. Social and political scientists, however, can rarely control the experimental programs with whose evaluations they are charged. Crime control programs are implanted in political and social enviroments far from the tranquility of the research laboratory. While the laboratory scientist can withhold treatment from a "control" group of the population under study, moral and political considerations often make the establishment of similar control groups for criminal justice programs impractical or impossible. Is it fair to withhold police patrols from some high crime areas while intensifying them

in others? Should some seriously delinquent juveniles be left untreated when others are receiving the benefits of new programs?

When treatment programs serve only those persons or areas most in need of service, it is common practice to use a comparison of the treated group's performance before and after treatment as the basic evaluative index. If crime and recidivism rates, and the other measures of effectiveness employed in evaluating criminal justice programs, were not frequently erratic and subject to apparently random fluctuations, then this sort of straightforward before-after comparisons could be quite reliable. However, the random nature of these variables introduces a form of estimation error which is rather subtle and has been overlooked in an alarmingly high proportion of evaluative studies, although its magnitude can be substantial. This, of course, is the regression artifact, the subject of this study.

Without going into a full discussion of the mechanism responsible for this type of error, since it is covered thoroughly on the following pages, the nature of the problem can be illustrated rather dramatically by considering a simple coin tossing experiment.

Imagine a room in which 20 individuals are each given a perfectly fair coin to toss, that is, a coin which has been tested to verify that a head is as likely to come up as a tail on any toss. Also, imagine that a "coin fairness" program has been instituted by the U.S. Treasury Department to identify and correct any coins which are <u>not</u> fair, and that a program staff person has been assigned to deal with "problem" coins in the aforementioned room. Further, because of the usual time pressures and paucity of available data, the staff person is <u>not</u> aware that all the coins in question are perfectly fair. Consequently, his first activity is to "test" all coins in the room by having each individual toss his coin ten times

and report the number of heads observed. Since five heads would be expected on the average from fair coins, the coins in the room are then ranked according to the absolute difference between the number of heads observed and five. Clearly, those with the greatest differences are most in need of "fairness correction." Unfortunately, sufficient funds are available to treat only six of the twenty coins, so the highest ranked coins are singled out for treatment. Being a careful evaluator, the staff person carefully notes the performance observed for each of these six coins before treatment, and computes the average difference as an index of the "coin unfairness" observed prior to treatment. Next, the six coins are carefully "treated" and then retested by having them each tossed an additional ten times. Encouragingly, the average number of heads observed is found to be much closer to five per coin, and a computation of the before-to-after improvement ratio proudly indicates the effectiveness of the treatment program.

Is this a completely ridiculous analog to real-world crime control program evaluation? After all, all coins were in fact perfectly fair, and those observed originally as having been "unfair" in the pretest were actually behaving consistently with their fairness-- when the probability of tossing a head is 0.5, the average number of heads in ten tosses will be five, but not every set of ten tosses will yield exactly five heads (in fact, the chances are only about 25 per cent that exactly five will be observed). Therefore, it is entirely likely that for at least six of the twenty coins the number of heads observed will be significantly different from five per coin. Of course, when these six coins are retossed, assuming the "fairness treatment" has not in fact made them less fair, the expected number of heads will still be five per coin and the chances that all six will again behave as divergently as they did on the pretest are quite small. In other words, the treatment program will appear to have "corrected" the fairness of the coins when in fact they were perfectly fair from the start.

This simple example points out that the process used to select members of a population for treatment may cause a program to look good when program services actually have no effect whatsoever on the treated population. Of course, in the world of crime reduction programs, the populations treated will not be composed of members whose needs for treatment are equally great. Some members will always be seriously in need of services while others will not. Because of random fluctuations in the performance measures, however, is is never possible to discriminate with certainty between members truely in need of service and those whose performance exhibited exceptional need on the pretest due to random variation. Consequently, following treatment, some members will revert to their normal levels of performance (a change which would have occurred without any treatment at all), while other members may make meaningful improvements as a result of treatment. The net effect is that before-after evaluation methods, unless carefully controlled for regression artifact, may erroneously indicate inflated success levels, even for programs having no effect whatsoever.

In the study presented here a fuller description of the regression artifact phenomenon and a method for estimating the magnitude of artifact related error are given. The study was conducted as a component of the Impact Evaluation Program established by the Missouri Law Enforcement Assistance Council - Region 5 to evaluate projects funded by the St. Louis High Impact Anti-Crime Program. Co-author Denise Corcoran analyzed the results of the study for her Master of Science thesis, submitted to the Department of Computer Science at Washington University. Co-author Nelson Heller served the dual roles of Director of Program Evaluation for the St. Louis Impact Program, and thesis adviser (he is an Affiliate Professor of the Department of Computer Science at Washington University). The following

report is taken almost entirely from Ms. Corcoran's thesis.

Nelson B. Heller
St. Louis, June 1974

CONTROL OF REGRESSION ARTIFACT
ERROR IN EVALUATING THE EFFECTIVENESS
OF CRIME REDUCTION PROGRAMS

1. FORMULATION OF THE PROBLEM

1.1 SCOPE OF THE PROBLEM

As one component of St. Louis' High Impact Anti-Crime Program, the Foot Patrol Project was designed and implemented to determine the crime reduction effect of intensive police foot patrols in high crime areas. The project's primary objective was to decrease the number of robberies and burglaries in the areas patrolled by providing a concentrated police presence in the form of foot patrolmen. The process for selecting the experimental group of blocks for the project involved identifying the six reporting areas (called "Pauly" blocks in St. Louis) within the city that ranked highest in crime for the specific target crimes considered. The ranking procedure was based on the 10-month crime totals for the period

*The numbers in parentheses in the text indicate references in the Bibliography.

from January to October of 1971. The project was implemented in July 1972, and was evaluated for its effectiveness for the period ending December 1972. Figure 1 displays graphically the above details. The statistics from the experimental period indicated a reduction of the target crimes in these six Pauly blocks. This leads to an important question: How effective was the project in meeting its objectives?

This is just one example of the following general type of analysis used in social experimentation: collect observations which measure a characteristic of the population considered, choose an extreme subgroup for treatment, collect data on this subgroup after the treatment. To determine the effectiveness of the treatment, a common approach is to measure the net change of the subgroup before and after the treatment. The problem with this type of "before-and-after" analysis is that, in most cases, some of the change in the observations may be due to sources other then those controlled by the experiment.

Thus, one purpose of this study was to examine the erratic behavior in crime rates before the Foot Patrol Project went into existence. It is only in this perspective that a true measure of the effectiveness of the project can be determined. Although the major part of this study revolves around this specific application, the basic analytical techniques can be used for other similar experimental situations.

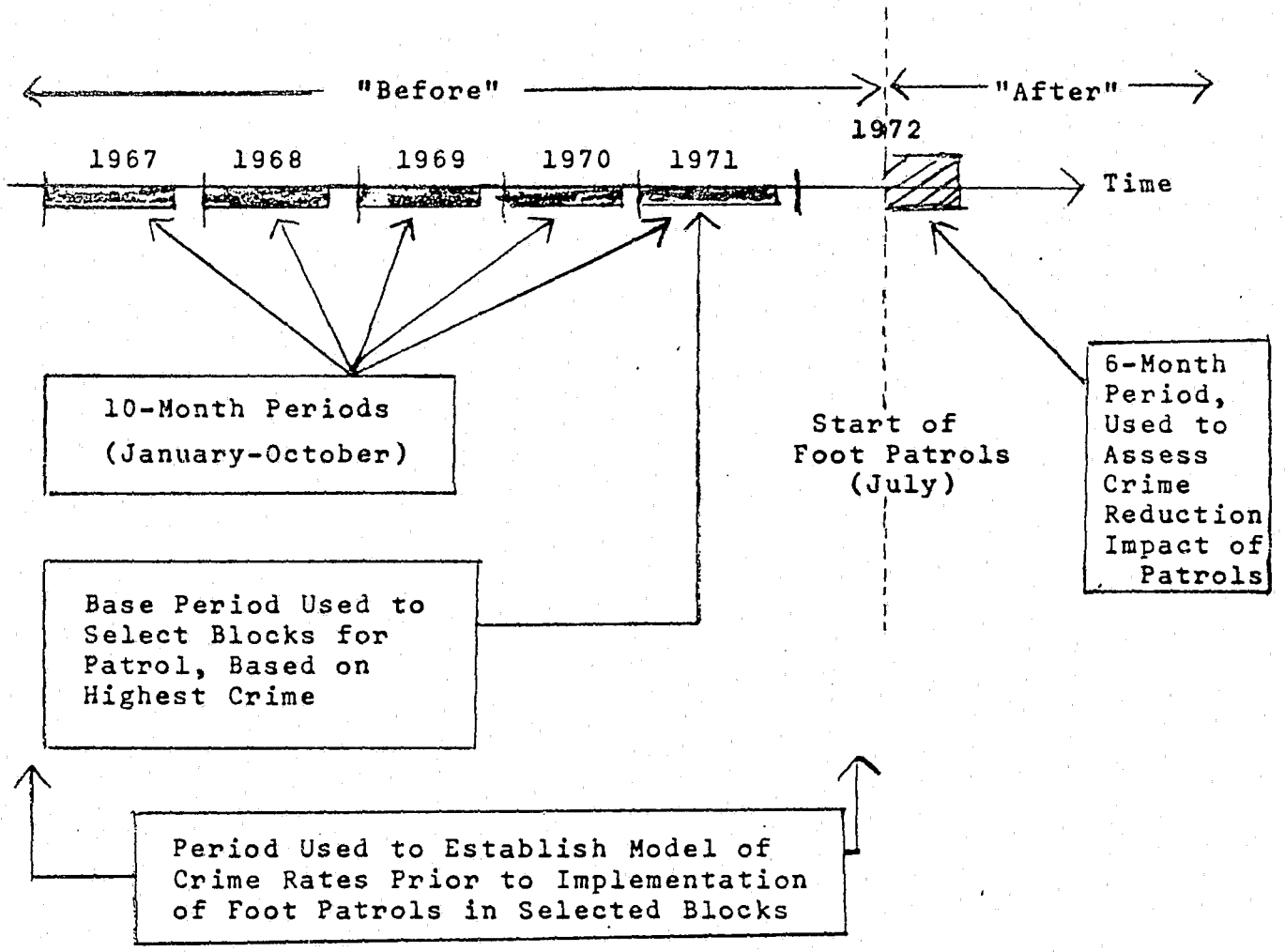


FIGURE 1. Time Line for Foot Patrol Project.

The core of this research effort was mainly concerned with two aspects of the project:

1) the design of the experiment
2) the method of evaluating the program after its implementation.

In discussing the design of an experiment in general terms, a bias may result from the process of selecting a particular extreme group to which a remedial treatment is administered. Relating this to the Foot Patrol Project, the underlying assumption in the design of this program was that the six Pauly blocks actually chosen were, in fact, the six highest crime areas in the city, for the period in which they were selected. However, the erratic behavior of crime within the population of Pauly blocks may actually have produced a selection bias. But the problem does not end here.

In the actual evaluation of the project, it was assumed that any significant reduction in the overall crime total for the six Pauly blocks, from that for the base period in which they were chosen, would be sufficient evidence of the successful impact of the project. However, since the group of Pauly blocks may have been chosen at an exceptionally high point in the histories of their own crime rates, caused by unusually large "random" crime increases, the absence of another random increase

may well reduce the magnitude of crime in the subsequent "after" time period to a more normal level, regardless of the existence of the Foot Patrol Project. Thus, an abnormal period of crime in the six foot-patrolled Pauly blocks may have coincided with the period in which they ranked the highest among all the Pauly blocks.

Thus, this study will attempt to evaluate the underlying random process which affects both the whole population of Pauly blocks and each individual Pauly block within the population simultaneously.

The remaining sections in this chapter include the following topics: a general discussion of the problem of "regression artifact", as encountered in evaluative research; previous research concerning this problem; and a general formulation of the problem as it relates to the evaluation of social programs. Chapter 2 presents a theoretical approach to this artifact problem using order statistics, for those cases in which certain simplifying assumptions of the model are valid. Chapters 3 and 4 discuss the design of a simulation model and the general logic of the computer program written to carry out the simulation. Chapter 5 describes the procedures used to validate the computer model in order to justify its use in situations too complex to be solved using order

statistics techniques. Chapter 6 presents the results of the simulation experiment, which replicates the selection process for the six highest Pauly blocks, based on generated crime rates, for a given number of runs. Finally, a summary of this paper with conclusive remarks about the results of the simulation experiment are presented in Chapter 7.

1.2 REGRESSION ARTIFACTS - WHAT ARE THEY?

One of the most pervasive phenomena in the study of change has been coined "regression artifact."* It is the natural tendency for those subjects selected as most deviant on an initial measure to average nearer to the mean on a second measurement. Since this so-called regression is a situation to be found in real life, it is important to understand why it occurs, so as to avoid misinterpreting any causal inferences.

All measures contain some component of "error" - to a greater degree for measures of behavioral characteristics than for those of physical properties. Thus, it is possible that those initially high on a measure are there partly because chance errors favored them on the day they were examined. Similarly, those low on an initial measure fell down because chance errors worked to

*Also called regression fallacy or regression effect.

their disadvantage on this testing. Since it is atypical that chance hits in the same manner on two successive occasions, it is a stochastic expectation that both those originally high and those originally low would regress toward the mean.

For experimental work, the regression phenomenon becomes important if subjects are selected bacause of their extreme scores on some variable, which is to be measured after a certain treatment has been administered to the group. Whenever a group is chosen for treatment because they were high on an initial measure, the effect of the treatment may, in part, be counteracted by the regression effect. Moreover, when a group is chosen because they were low on an initial measure, at least some of their gain on a subsequent measure may be attributed to the rival hypothesis of regression effects. A familiar example of this situation deals with I.Q. scores. Frequently, those children who score the lowest on an I.Q. test are selected for remedial training. After some time they are again tested, resulting in an improvement in their I.Q. scores. However, since the scores would be expected to rise anyway due to regression, the contribution of the training program is unknown. The regression effects are

created partly by chance factors in the testing situation, by measurement error, and by other temporary influences, all of which had worked to the disadvantage of these subjects on an initial test.

However, it is necessary to point out that a regression artifact is not always encountered with extreme scores. For the situation in which a group is originally chosen by a random process, but yet turn out to have an extreme mean, regression effects may be prevented. But for a group selected because of an extreme variable, we can expect the mean of this group to regress toward the mean of the population from which it was selected.

1.3 PREVIOUS RESEARCH

The phrase "regression artifact" appears to have taken its name from F. Galton's observation that "the progeny of all exceptional individuals tends to regress towards mediocrity."[1] That is, Galton thought the heights of people were becoming more uniform because the sons of the tallest fathers and of the shortest fathers were closer to the average than the fathers had been. Later on, Galton realized the fallacy in his own findings, due to the fact that he was simply looking at selected members of the population (sons of the tallest and shortest fathers). In his paper,

"Regression toward Mediocrity in Hereditary Stature", he recognized that not only did exceptional parents have offspring more mediocre than themselves, but also exceptional offspring came from parentage more mediocre than they.[2] Thus, in an era in which the average stature does not change, the heights of sons of tall fathers average shorter than their fathers, but the heights of fathers of tall sons average shorter than their sons.

Although the regression phenomenon has been known for more than a half century, such results were not used until some time later in interpreting scores obtained in mental and educational tests. Robert Thorndike, involved in educational experimentation, pointed out the importance of the reliability of a test when one is using the obtained score on a test as an estimate of the student's "true" score. He considered a highly reliable test one which yields a high correlation between obtained scores with the corresponding true scores. The formula that Thorndike used to estimate the true score was

$$X_{true} = r\, X_{obtained},$$

where the lefthand side denotes the true score as a deviation from the mean of the group being measured, and

the right side equals the reliability coefficient, assumed to be a certain value, times the obtained score expressed as a deviation from the mean of the group. Thorndike emphasized that it is the true scores from an initial test, not the obtained scores, which must be used as the basis to match those groups to be compared on subsequent tests. However, throughout the discussion, Thorndike does not attempt to define what he means by a student's "true" score. Moreover, he fails to mention how to compute the reliability coefficient that he uses in his formula.

Further development was made by Frederic M. Lord in regard to educational testing. In "Measurement of Growth", he derives a regression formula to estimate the student's true gain between successive tests, as based on his observed initial and final scores.(3) However, such a method is based on the assumption that initial and final tests are of "identical or equated forms". In a follow-up paper, he discusses some of the basic controversial measurement problems.(4)

The most commom problem deals with a frequent attempt made by educators to compare the magnitude of a gain in the score of a good student to that of a poor student on a retest. Since it is highly probable that one would observe smaller gains for higher initial

scores, such an occurence might lead to the erroneous conclusion that the good students are actually learning less than the poor students. As Lord points out, the flaw in this type of evaluation is that a comparison of the gain of two individuals cannot be made unless they start at exactly the same point on the score scale. For to compare gains of people at different parts of the score scale is to imply that the magnitude of a gain from different points on the score scale may be treated in terms of "equal" units. This would be a subject for serious debate.

Since very little of the experimental and evaluative research work in the social psychological sciences are subject to controlled observation, regression effects must not be ignored as a possible explanation for many test results. However, a survey of the literature indicates that at the present time there has been no rigorous attempt to measure the magnitude of the artifact, but merely a recognition that the problem exists.

1.4 GENERAL APPROACH TO THE PROBLEM OF REGRESSION ARTIFACT IN EVALUATIVE RESEARCH

The artifact problem encountered in the evaluation of experimental social programs may lead to erroneous conclusions about the outcome of a program if the clients or

areas served by it are chosen on the basis of their poor performance or exceptional need during the period preceding implementation, as opposed to some kind of randomized selection procedure. The most common measure of effectiveness of such programs involves a "before and after" comparison of the behavior of the experimental, or treated group - that is,

$$\text{Effectiveness} = \frac{\text{Performance measured before experiment}}{\text{Performance measured after experiment}}$$

Any change in performance is attributed to the program. The extent to which this type of measure is inflated by the artifact depends on the random or irregular behavior of the performance measure for the experimental group and the population from which it is selected. Therefore, a more accurate estimate of the effectiveness is given by the overall change in effectiveness minus an estimate of the artifact.

To measure the magnitude of the artifact requires knowledge of the following: the size of the population (i.e., clients or areas), the size of the experimental or treated group, and a measure of both the regular behavior and the irregular, or random, variability of the performance measure for the population. Since most experiments are performed in a time setting, an estimate of the expected performance and random error component can be computed using a time series model. Each estimate is used in determining

the artifact: the expected performance represents a measure of normal behavior and the error component represents a measure of erratic behavior. Although there is no set way to compute these estimates for a given model, the least-squares regression model technique is suggested because it gives measures for the testing of significance of its estimated parameters, as well as minimizing the sum of squares of the error terms.

Following the format of the "before and after" effectiveness measure, the measure of the artifact suggested is the ratio of a "before" measure, based on actual behavior during the selection period, to an estimated "after" measure, based on behavior expected from the areas treated if they had been selected during a period of more normal performance for themselves. In other words, the "before" measure reflects the extremes of performance during the selection period, and the "after" measure represents the more normal behavior of the experimental group, or the level of behavior expected when no random or irregular fluctuations are present. That is,

$$\text{Artifact} = \frac{\text{"Before" Measure}}{\text{"After" Measure}} = \frac{\text{Biased Behavior}}{\text{Expected Behavior}} .$$

Because of the random nature of the extremes in performance

of the clients, or areas, the artifact measure, itself, is best treated as a random variable. Therefore, the information which is necessary for the evaluation of a social experiment must include knowledge of random variations in the behavior of the performance measure for the population under study, and how the process of selecting for treatment only those units exhibiting poorest performance during a specified period leads to "before" performance estimates which are misleadingly extreme.

A computer simulation was used in this study to provide estimates of all the factors described above. Included as input to the simulation model is a time series performance model which is based on observed performance for time periods prior to implementation of the program. The time series model is used to estimate:

1) the distribution of irregular or random performance variations for each client or area in the population under study,
2) an estimate of the expected or normal performance of each client or area for the selection period, and,
3) an estimate of the expected behavior for each client or area during the subsequent period when the program is underway.

The procedure for estimating the magnitude of the artifact, based on data for the period prior to implementation

of a program, is illustrated for the case of the St. Louis Foot Patrol in the following chapters. The probability distribution for the artifact measure, and its expected value, are derived using the time series model, and a random number generator for the irregular fluctuations. At each iteration of the simulator the performance of every client or area is determined. The set of those exhibiting poorest performance is identified, and their performance average is computed. This average is then divided by the average performance for these same clients or areas when the random fluctuation has been set to zero. The iterative procedure is repeated a specified number of times (eg., 400-500 times for the Foot Patrol analysis). Probability statements can then be made, on the basis of the estimated distribution, about the magnitude of the artifact for the actual period used for the selection of the experimental group.

If the experimental period is, in fact, the same duration as the selection period but in a subsequent time interval, then the simulation may also be used to evaluate the extent of the change in the behavior of the experimental group during the program, which can be attributed to the artifact. This involves a two-stage process: first, simulating the behavior of the population

prior to the project to determine the members of the experimental group, and second, simulating the behavior of the selected group for the subsequent period. Thus, the "before and after" measure of the artifact for the experimental group is

$$\text{Artifact} = \frac{\text{"Before" Simulated Behavior}}{\text{"After" Simulated Behavior}} .$$

By repeating this process a number of times, the distribution of this random variable can be estimated. It is then used to determine the probability that the actual change in the behavior of the experimental group during the program period is due to an artifact alone. The actual change during the experimental period minus the mean artifact measure from the simulated distribution may be used as an estimate of the true change in the experimental group attributed to the program. The validity of the output results will, of course, depend on the extent to which the simulation model correctly describes the behavior of the population.

1.5 PREDICTIVE MODELS FOR CRIME TOTALS IN EACH PAULY BLOCK

The main objective of this research effort is to test hypotheses about the behavior of crime within the

system of Pauly blocks. However, it is necessary to define the structure of this system before any analysis work can be performed.

The two important features of this system are as follows:

1) Components

The components of the system are all the Pauly blocks in the City. The performance of the system as a whole and of each component in the system is measured in terms of crime rates.

2) Variables

As the performance measure, crime rate is the key variable of this system. Formulated as a time-series model, the crime rate can be explained in terms of three variables: trend, seasonality and a random "error." The first two variables relate to a systematic change in crime rate behavior; the latter, an erratic change. It is this "error" variable which leads to, and is used as a basis for measuring, the regression artifact in crime rates.

Assuming that crime rates can be modeled as a linear combination of the variables described above, the following expression may be used to relate crime for

each Pauly block and time period to trend, seasonality, and random fluctuation,

$$C(i,t) = A(i) + T(i,t) + S(i,t) + e(i,t), \qquad (1\text{-}1)$$

where $C(i,t)$ denotes the crime rate for Pauly block i, during time period t,
$A(i)$ is a constant factor giving the average crime rate for Pauly block i,
$T(i,t)$ is the trend factor for Pauly block i and time period t,
$S(i,t)$ is the seasonality factor for Pauly block i and time period t,
$e(i,t)$ is the random fluctuation component in the crime rate for Pauly block i, and time period t.

The underlying, or expected, crime rate for a given Pauly block and time period is obtained by estimating the parameters $A(i)$, $T(i,t)$ and $S(i,t)$ from equation (1-1). That is,

$$\hat{C}(i,t) = \hat{A}(i) + \hat{T}(i,t) + \hat{S}(i,t) \qquad (1\text{-}2)$$

The estimate $\hat{C}(i,t)$ represents the expected norm of crime behavior for each Pauly block and time interval. Any deviation from this norm can be characterized as a

random fluctuation which depends on other variables that affect crime rates, but which are not explicitly included in the model. An estimate of this random variable $e(i,t)$ can be obtained from the equation

$$\hat{e}(i,t) = C(i,t) - \hat{C}(i,t).$$

The estimates of $e(i,t)$ may be used to compute a measure of the regression artifact, as described in Section 6.

1.6 USE OF A SIMULATION

A computer simulation model was used to study the behavior of the "error" component in crime rates over time for single Pauly blocks and for groups of Pauly blocks. One of the main reasons for selecting a simulation rather than an algebraic model is that fairly complex processes may be modeled more readily in simulation. In this case, the size of the population of Pauly blocks, compounded by differences in their crime behavior, make the development of a realistic, analytic model of the selection process for the Foot Patrol Project very difficult.

The simulation is programmed to generate random numbers having the probability distribution of the error factors for each Pauly block being modeled. The

random number generator makes use of an estimated distribution function for the error terms. The function was computed by fitting a form of the basic model, discussed in the previous section, to the actual time-series of crime data for each block. Both the generated error term and an estimate of the normal crime rate during the actual selection period for the project for each Pauly block (also computed from the basic model) were used to replicate the process of selecting the six highest crime Pauly blocks from a population of Pauly blocks affected by erratic crime behavior. Given this set of the highest crime blocks for each run, a measure of the selection bias is then recorded. Thus, the simulation model is run a sufficient number of times to produce a measure of the expected performance of the system of Pauly blocks in terms of its generated crime rate and its expected crime rate for the selection period. However, reliability of the simulation output depends on how accurately the model describes the real system. If validation of the simulated model can be established, it then becomes an effective tool for evaluating events that have occurred in the real system.

2. ORDER STATISTICS MODEL

2.1 DEFINITION AND ASSUMPTIONS OF ORDER STATISTICS

Methods for analyzing order statistics have become an extremely useful tool in statistical inference because some of their properties are not dependent upon the distribution from which the random sample is obtained. Thus, assuming that all the observations of a random sample $X_1, X_2, \ldots, X_n$ have the same density function $f(x)$ and are independently distributed, order statistics can create order from this mass of data by putting the observations in numerical sequence. The result is a permutation of the original observations X_i, denoted by $[Y_{(1)}, \ldots, Y_{(n)}]$, such that $Y_{(1)} < \ldots Y_{(n)}$.

This vector of ordered observations is referred to as the order statistic. Then Y_i, $i = 1,2, \ldots, n$, is called the $i^{\underline{th}}$ order statistic of the random sample $X_1, X_2, \ldots, X_n$. In principle, it is possible to derive the distribution of the individual components of the order statistic or the joint distribution of several of them from the distribution of the complete order statistic. However, beyond the joint distribution of two order statistics, the task becomes quite burdensome for manual computation. For this reason, the usefulness of these results has been limited for any kind of analytical work.

Generally speaking, various other quantities based on order can be thought of as order statistics. For example, the average of the i^{th} and j^{th} order statistic, $\frac{Y(i) + Y(j)}{2}$, is included in this branch of statistics. It is this area of order statistics that is of great consequence in analyzing regression artifacts.

2.2 ANALYTICAL MODEL BASED ON ORDER STATISTICS

In this section, general results for the distribution of the average of the two highest order statistics have been derived - that is, for the distribution of the statistic, $\left[\frac{Y(n-1) + Y(n)}{2}\right]$, for a given sample of size n.

Let $X_1, X_2, \ldots, X_n$ denote a random sample from a distribution of the continuous type, having a probability density f(x)*. Let Y(r) and Y(s) denote the r^{th} and s^{th} order statistic, such that Y(r) < Y(s), Then the joint dsitribution of Y(r) and Y(s), denoted by f(Y(r), Y(s)), for a given sample of size n can be expressed as

$$f(Y(r), Y(s)) = \frac{n!}{(r-1)!(s-r-1)!(n-s)!} [F(Y(r))]^{r-1} * [F(Y(s)) - F(Y(r))]^{s-r-1} [1-F(Y(s))]^{n-s} f(Y(r)) f(Y(s)),^{(5)} \quad (2\text{-}1)$$

*The discrete case has been omitted due to the complexity of the expressions for manual computation.

where F denotes the cumulative distribution function. Thus, it follows that the joint distribution of the two highest order statistics, namely Y(n-1) and Y(n), is

$$f(Y(n-1), Y(n)) = n(n-1) [F(n-1)]^{n-2} f(Y(n-1)) * f(Y(n)), \quad (2\text{-}2)$$

This result is necessary to obtain the distribution of the average of $Y_{(n-1)}$ and $Y_{(n)}$.

Now let M define a new random variable, such that

$$M = \left[\frac{Y_{(n-1)} + Y_{(n)}}{2}\right].$$

The cumulative distribution function of the random variable M, denoted by $F_M(z)$ - that is, the probability that M is less than or equal to some arbitrary number z - can be expressed as

$$F_M(z) = P(M \leq z) = P\left[\frac{Y_{(n-1)} + Y_{(n)}}{2}\right] \leq z = P\left[Y_{(n-1)} + Y_{(n)} \leq 2z\right].$$

But $P[Y_{(n-1)} + Y_{(n)} \leq 2z]$ is the volume of f(Y(n-1), Y(n)) in the region Y(n-1) + Y(n) $\leq$ 2z, with an additional constraint that Y(n-1) $\leq$ Y(n), and f(Y(n-1), Y(n)) = 0 outside the defined boundaries.

Thus,

$$F_M(z) = P[Y(n-1) + Y(n) \leq 2z]$$

$$= \iint_R f(Y(n-1), Y(n)) \, d(Y(n-1)) \, d(Y(n)) \qquad (2\text{-}3)$$

where R is the region in which $f(Y(n-1), Y(n))$ is defined (see Figure 2). Using the "change of variable" technique,[6] let

$$s = Y(n-1)$$

$$t = Y(n-1) + Y(n),$$

which results in the transformation

$$Y(n-1) = s$$

$$Y(n) = t-s$$

and

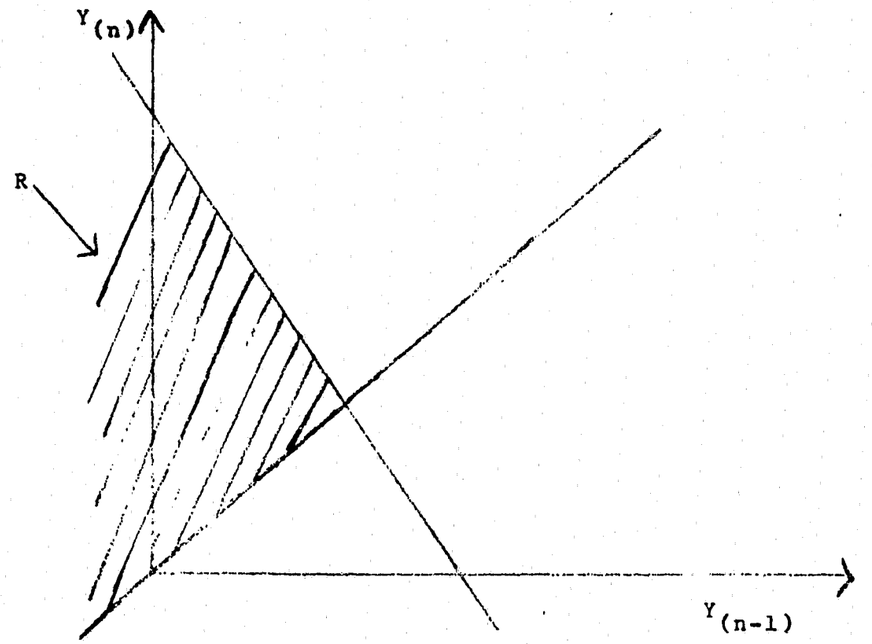
$$d(Y_{(n-1)}) = ds$$

$$d(Y_{(n)}) = dt.$$

This gives the identity

$$f(Y_{(n-1)}, Y_{(n)}) \, d(Y_{(n-1)}) \, d(Y_{(n)}) = f(s, t-s)dsdt.$$

The region R' on which $f(s,t-s)$ is defined can be obtained by using a similar transformation on the boundaries of R.



$$R: \begin{cases} Y_{(n-1)} \leq Y_{(n)} \\ Y_{(n-1)} + Y_{(n)} \leq 2z \end{cases}$$

FIGURE 2. Region on which the joint density function of Y(n-1) and Y(n) is defined.

The boundary $Y_{(n-1)} \leq Y_{(n)}$ yields the new boundary $s \leq \frac{t}{2}$; the boundary $Y_{(n-1)} + Y_{(n)} \leq 2z$ gives the new boundary $t \leq 2z$ (see Figure 3).

Using the change of variable technique, the cumulative distribution function of M can now be expressed as

$$F_M(z) = \int_{-\infty}^{2z} \int_{-\infty}^{\frac{t}{2}} f(s,t-s)dsdt. \qquad (2.4)$$

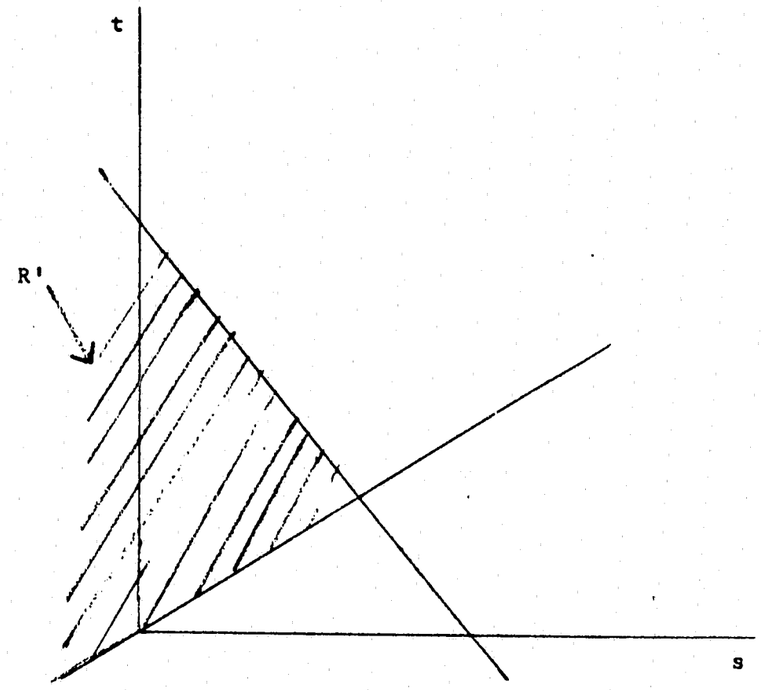
The probability density function, $f_M(z)$, is computed by taking the derivative of $F_M(z)$ with respect to z. That is,

$$f_M(z) = \frac{d(F_M(z))}{dz}$$

$$= \frac{d}{dz}\left(\int_{-\infty}^{2z} \int_{-\infty}^{\frac{t}{2}} f(s,t-s)dsdt\right).$$

The derivative of the right-hand expression with respect to z, which appears only in the upper limit of the outer integral, is the inner integral evaluated at t = 2z times the derivative of 2z with respect to z:

$$f_M(z) = \int_{-\infty}^{2} \int_{-\infty}^{z} f(s,t-s)ds.$$



$$R' : \begin{cases} s \leq t/2 \\ t \leq 2z \end{cases}$$

FIGURE 3. Transformed region on which the joint density function of Y(n-1) and Y(n) is defined.

Using the expression for the joint distribution of the two highest order statistics from Equation (2-2),

$$f_M(z) = 2n(n-1) \int_{-\infty}^{z} [F(2z-s)-F(s)]^{n-2} \cdot f(s)\, f(2z-s)\, ds. \qquad (2\text{-}5)$$

To apply this result to a specific probability distribution, one need only substitute the appropriate cumulative districution function and density function in the latter equation.

2.3 ANALYTICAL MODEL VS. SIMULATION MODEL

The validity of the order statistics model is contingent on the strength of its underlying assumptions, as applied to the "real world" situation. Thus, to use this model, it is necessary to know the specific distribution of the random sample and also justify that each and every observation is identically distributed according to this density function.

With reference to crime rates for the various sections of the city of St. Louis, the necessary condition of identical distributions is highly improbable, since the magnitude of crime is greatly influenced by geographical location. Such a discrepancy is also complicated by the lack of knowledge of the particular distribution.

Furthermore, manual computation limits the results to a small number of specific continuous distribution functions, to a very small sample size and to only the average of the two extreme observations in the sample. This is due mainly to the impossible task of manually evaluating cumbersome integrals and/or summations that are encountered throughout the computation. However, this approach can become a valuable tool if such limitations can be overcome through the use of computers.

On the other hand, the main advantage of a simulation model is that it can reproduce system behavior, given <u>any</u> distribution function. Furthermore, if the specific distribution which governs the observations is unknown, the empirical distribution may be used as a substitute in the model. This also relaxes the need for every area in the city to have an identical distribution of crime rates. The flexibility of the sample size and the number of replications of the simulated experiment can allow for a greater degree of reliability in the output analysis. For these reasons, the simulation model was selected as the means for evaluating the experimental results, as they exist in the real situation. However, there is always the question of how much confidence we can place in the simulation model in using it to represent

the true system. It is for this purpose that the analytical results become very important. Consequently, deviations of the simulated results from the analytical results have been used to determine the validity of the simulation model in a subsequent chapter.

2.4 USE OF ORDER STATISTICS FOR VALIDATION OF THE SIMULATION MODEL

The results derived in Section 2.2 have been applied to three distributions for the purpose of validating the simulation - namely, the uniform, exponential and triangular distributions. A sample of the procedure for each distribution will now follow: first, the joint distribution of the two highest order statistics will be given for a certain sample size, and second, the distribution of the average of the two highest order statistics will be computed.

Uniform Distribution

Let $X_1, X_2, \ldots, X_n$ be uniform on [0,1] and $Y_1, Y_2, \ldots, Y_n$ be the corresponding order statistics. Then the joint distribution of Y(n-1) and Y(n) is

$$f(Y(n-1), Y(n)) = n(n-1)[Y(n-1)]^{n-2},$$

as derived from Equation (2-2). For a sample of size three

$$f(Y(2), Y(3)) = 6Y(2) .$$

Since the joint density function of Y(2) and Y(3) is defined only on the [0,1] interval, the distribution of $f_M(z)$, where $z = \frac{Y(2) + Y(3)}{2}$, is also defined only on the [0,1] interval. For $0 \leq z \leq 1$, different calculations are required in the right and left halves of the interval [0,1] (see Figure 4).

Having used the change of variable technique to obtain $f_M(z)$ in 2.2, the procedure for determining the corresponding constraints for the uniform distribution have been outlined in Table 1. Thus, for $0 \leq z \leq 1/2$,

$$f_M(z) = 2\int_0^z f(s, 2z-s)ds$$

$$= 6z^2$$

and for $1/2 \leq z \leq 1$,

$$f_M(z) = 2\int_{2z-1}^z f(s, 2z-s)ds$$

$$= 24z - 18z - 6 .$$

Figure 5 shows the population distribution from which the random variables $X_1, X_2, \ldots, X_n$ were derived and the

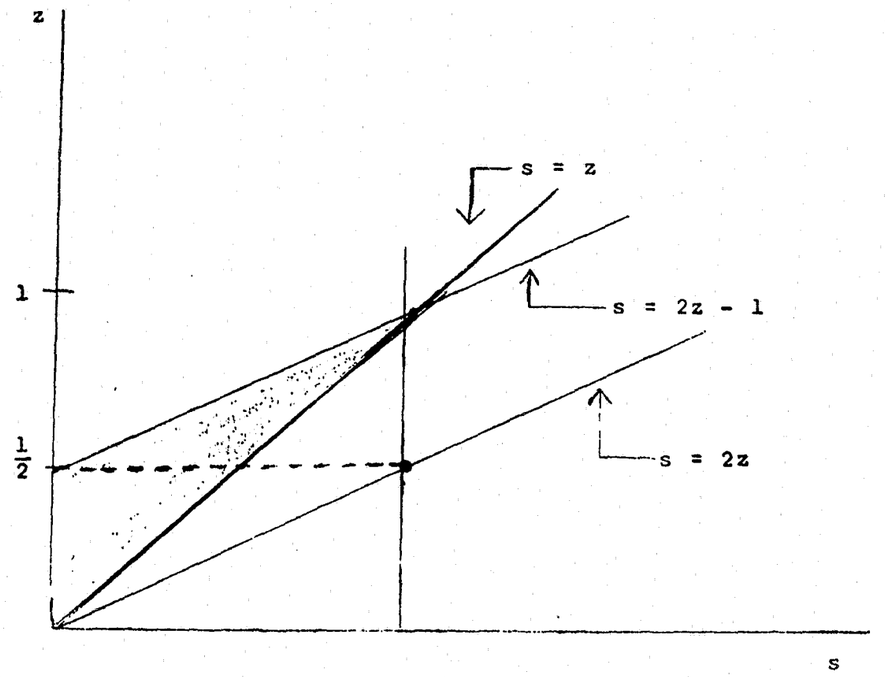
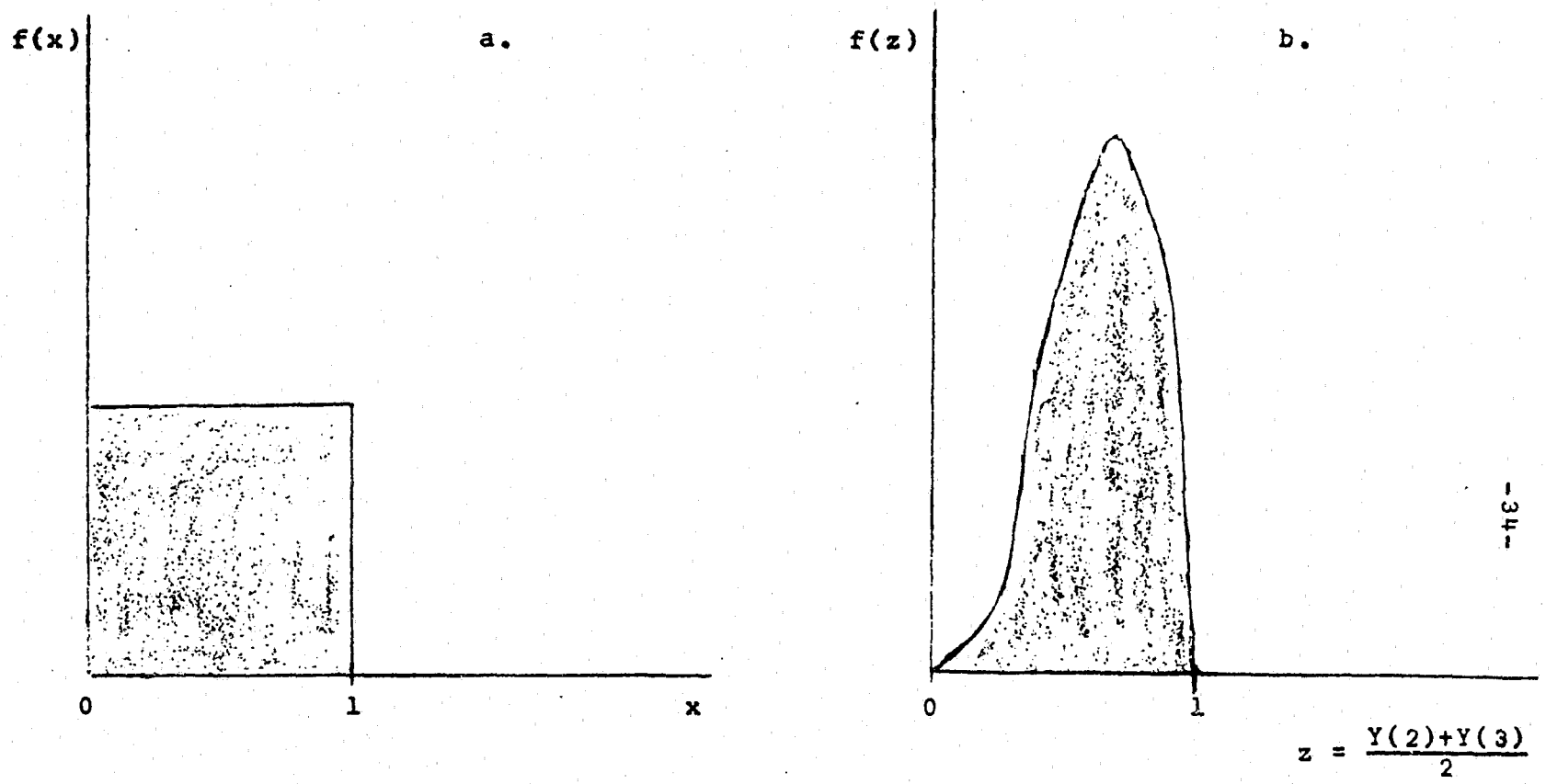


FIGURE 4. Graphical Representation of $f_M(z)$ in terms of transformed variables for the uniform distribution.

Table 1. Table of Transformed Variables (Corresponding to Figure 4)

Constraints for Original Variables	Equivalent Constraints for Transformed Variables	
$Y_{(1)} \geq 0$	$s \geq 0$	
$Y_{(2)} \geq 0$	$2z-s \geq 0$	$(s \leq 2z)$
$Y_{(1)} \leq Y_{(2)}$	$s \leq 2z-s$	$(s \leq z)$
$Y_{(2)} < 1$	$2z-s \leq 1$	$(s \geq 2z - 1)$



Population Distribution of x:
Uniform Distribution

Distribution of z,
Given the Uniform Distribution

FIGURE 5. Population and Transformed Distribution: Uniform

corresponding distribution of the average of the two highest observations as related to the order statistics.

Exponential Distribution

Let X be exponentially distributed, where $f(X) = \lambda e^{-\lambda X}$ for $0 < \infty$, so that the joint distribution of the order statistics Y(n-1) and Y(n) is

$$f(Y(n-1),\ Y(n)\ ) = n(n-1)[1 - e^{-\lambda Y(n-1)}]^{n-2}$$

$$[\lambda e^{-\lambda Y(n-1)}]\ [\lambda e^{-\lambda Y(n)}].$$

For n = 3,

$$f(Y(2),\ Y(3)\ ) = f(Y(2),Y(3)) = 6[1-e^{-\lambda Y(2)}][\lambda^2 e^{-\lambda(Y(2)+Y(3))}]$$

In terms of the transformed variables, where it is required that $s \leq 2z - s$ (Or $s \leq z$),

$$f_M(z) = 2 \int_{-\infty}^{z} f(s,\ 2z - s)ds$$

$$= 4 \int_{0}^{z} \lambda^2 e^{-2\lambda z} ds$$

$$= 12z\lambda^2 e^{-2\lambda z} + 12\lambda e^{-3\lambda z} - 12\lambda e^{-2\lambda z}$$

Figure 6 shows the probability density function of X_1, X_2, ..., X_n and the probability distribution obtained for the average of the two highest order statistics, Y(n-1) and Y(n).

Triangular Distribution

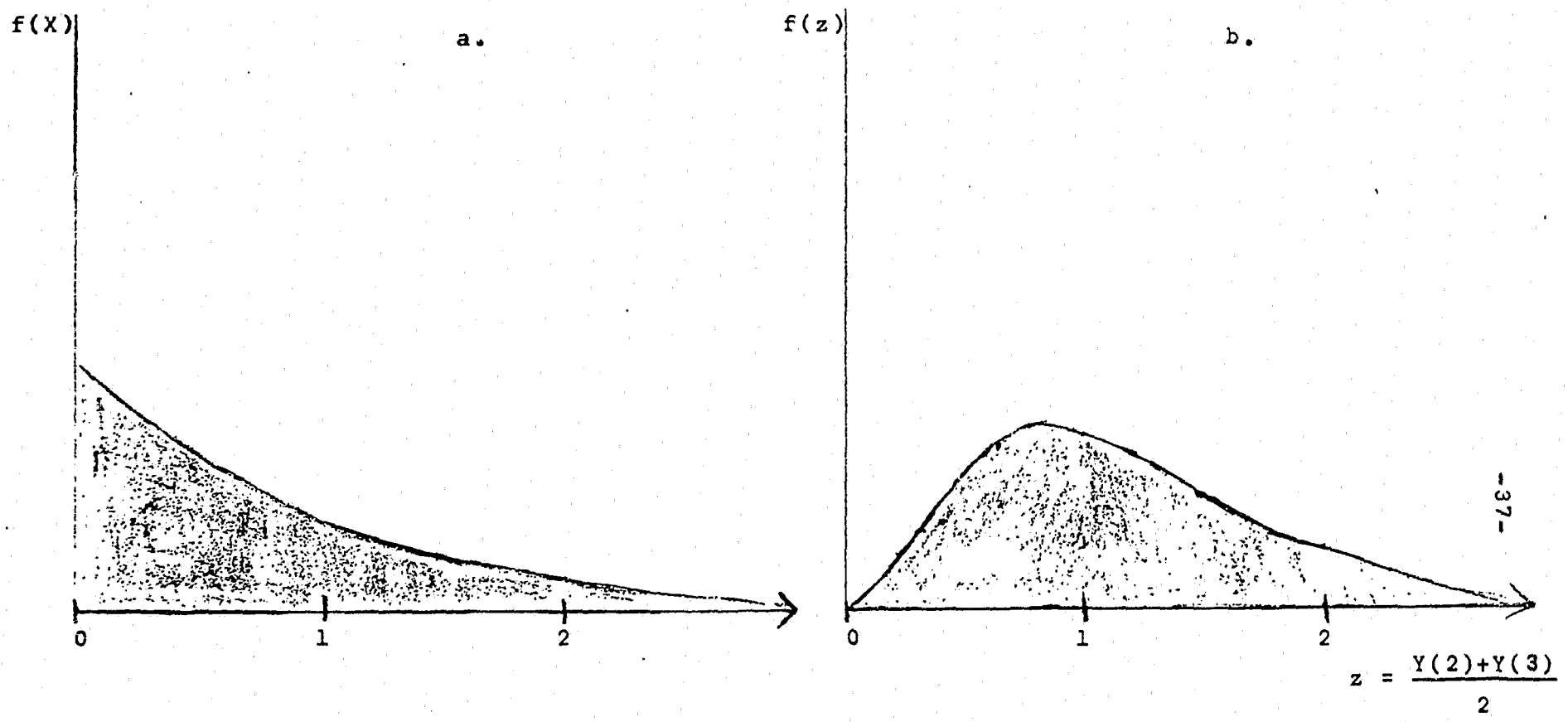
The computational procedure for the average of the two highest order statistics, $\frac{Y(n-1) + Y(n)}{2}$, for the triangular distribution is similar to that for the previous distributions. Thus, the results for this distribution have been summarized.

Let X have the density function

$$\begin{aligned} f(X) &= X && 0 \leq X \leq 1 \\ &= 2 - X && 1 \leq X \leq 2 \\ &= 0 && \text{Elsewhere} \end{aligned} .$$

Then the probability distribution of $\frac{Y(n-1 + Y(n)}{2}$, denoted as $f_M(z)$, for a sample size of two has been evaluated as follows:

$$\begin{aligned} f_M(z) &= 8z^3 && 0 \leq z \leq 1/2 \\ &= -8z^3+16z^2-8z+4/3 && 1/2 \leq z \leq 1 \\ &= 8z^3-32z^2+40z-44/3 && 1 \leq z \leq 3/2 \\ &= -8/3z^3+16z^2-32z+64/3 && 3/2 \leq z \leq 2 \end{aligned} .$$



Population Distribution of X:
Exponential Distribution ($\lambda = 1$)

Distribution of z,
Given the Exponential Distribution

FIGURE 6. Population and Transformed Distribution: Exponential

For this particular sample size, these expressions are equivalent to the average of any two random variables, regardless of order, from the triangular distribution. The derivation of $f_M(z)$ for a sample size greater than two was too cumbersome for manual computation.

The results derived for each distribution have been summarized in Chapter 5, and used for tests of significance to determine the validity of the simulation model.

3. DESIGN OF SIMULATION EXPERIMENT

The first section of this chapter outlines the data specifications, as the crime type, the time period, etc.; the last two sections deal with the statistical design of the simulation model, including the estimation of the parameters of the model and the design of the sample of Pauly blocks used to represent the complete population in the model.

3.1 DATA BASE

The Foot Patrol Project will be used to illustrate the types of input data required by the simulation model. The specifications of the data are as follows:

Form of Data: Crime rates for specific blocks and time periods.

Type of Crime: Suppressible robbery and burglary.

Time Periods: January - October for five years (1967-1971).

The term "crime rate" here refers to the number of crimes reported for a given time period. The St. Louis Metropolitan Police Department classifies as "suppressible" all crimes which could conceivably have been prevented by an officer on routine patrol had he been near enough to view the incident. In general, suppressible crimes are those which take place outdoors. In regard to the time period, crime data at the Pauly block level was not

available for years prior to 1967. Since the Foot Patrol Project had been implemented in 1972, the time range of interest in assessing crime patterns prior to implementation was limited to the five years from 1967 to 1971. The historical crime data, as outlined above, was obtained for each Pauly block represented in the simulation. All data was obtained from crime tapes at the St. Louis Police Department.

3.2 STATISTICAL DESIGN

The simulation model was used to test the hypothesis that the artifact situation, as introduced in Section 1.2, existed in the evaluation of the Foot Patrol Project. It was also used to compute the probability distribution of the inflation in crime - that is, the ratio of the generated crime rate to the expected, or normal, crime rate - for the six highest Pauly blocks (before the Foot Patrol Project). The inflation in crime can be attributed to the selection process and the random fluctuation in crime in each Pauly block; the above ratio used to measure this inflation will be referred to as the artifact ratio. The simulated distribution of the artifact ratio can be used to make probability statements about the behavior of crime rates in the system of Pauly blocks. Moreover, the mean of the distribution is an indicator of the expected bias - that is, the amount of crime reduction which can be

anticipated in a subsequent period even without the presence of the project.

In order to produce such output, estimation of the parameters of the model, formulated in Equation (1-1), was necessary to satisfy the input specifications of the computer simulation. To do this, the basic model had to be modified due to the small number of observations for each Pauly block. Thus, a time-series model, which incorporated a uniform trend factor for each Pauly block, was used. Since the use of a single model might result in a poor fit, if, in fact, behavior among the Pauly blocks differed significantly, a block factor was included to explain some of this variation.

Let $C(i,y)$ ($y = 1,2, \ldots, 5$ for the time range 1967-1971) be the 10-month total of crime in Pauly block i during year y. Then the time-series model for $C(i,y)$ can be expressed in the form

$$C(i,y) = a + \sum_{w=1}^{5} t(w)X(w,y) + \sum_{p=1}^{z} b(p)\,g(p,i) + e(i,y) \qquad (3\text{-}1)$$

where a is a constant, $t(w)$ is a correction, or trend, factor for each year, $X(w,y)$ is a "0-1" variable such that

$$X(w,y) = \begin{cases} 1 & \text{if } w = y \\ 0 & \text{otherwise} \end{cases},$$

b(p) is a correction factor for each block, z is the total number of blocks in the sample, g(p,i) is a "0-1" variable such that

$$g(p,i) = \begin{cases} 1 & \text{if } p = i \\ 0 & \text{otherwise} \end{cases}, \text{ and}$$

e(i,y) is the residual error term for block i and year y. The seasonal factor was omitted from the model since the same ten months of each year were used for each of the five years under study.

The input data for the time-series model was based on crime data for the sample of Pauly blocks used to represent the population. The design of the sample is discussed in the next section. Estimates of the parameters in Equation (3-1) were obtained using the BMD "Multiple Regression" Library Program. A sample of the input format for the program is shown in Table 2 . Using the estimates obtained for the parameters, the expected total crime for each block and year can be computed as

$$\hat{C}(i,y) = \hat{a} + \sum_{w=1}^{5} \hat{t}(w)X(w,y) + \sum_{p=1}^{z} \hat{b}(p)g(p,i) \qquad (3\text{-}2)$$

The results of the regression analysis indicated that the estimates for each year correction factor are significant at the 95% level; the estimates for the Pauly block factors

TABLE 2. Sample Input for Time-Series Model
(Number of blocks = 3, Number of years = 2)

		Dependent Variable Total Crime	Independent Variables				
			Year 1	Year 2	Block 1	Block 2	Block 3
YEAR 1	Block 1	C_{11}	1	0	1	0	0
	Block 2	C_{21}	1	0	0	1	0
	Block 3	C_{31}	1	0	0	0	1
YEAR 2	Block 1	C_{12}	0	1	1	0	0
	Block 2	C_{22}	0	1	0	1	0
	Block 3	C_{32}	0	1	0	0	1

showed very little statistical significance. The last result suggests that there was no a significant difference in the year to year crime behavior among Pauly blocks. But this is not conclusive since these estimates are based on a small number of observations. Moreover, the t-statistic used for testing significance is based on the assumption that the distribution of e(i,y) is normal $(0,\sigma^2)$, which may not be valid in this case.

The estimates obtained from Equation (3-2) were utilized in the simulation model in two ways:

1) to compute $\hat{C}(i,5)$.

Since year 5, corresponding to 1971, was the year in which high crime caused a Pauly block to be selected for the Foot Patrol Project, $\hat{C}(i,5)$ is an estimate of the normal activity expected for Pauly block i during the selection period. The estimate for each block remains constant throughout the simulation experiment, only the random component varies from year to year.

2) to compute $\hat{e}(i,y)$ for every year, and every Pauly block.

An estimate of $\hat{e}(i,y)$ can be obtained from the formula

$$\hat{e}(i,y) = C(i,y) - \hat{C}(i,y).$$

The frequency distribution of $\hat{e}(i,y)$ for Pauly blocks is then computed; the distribution serves as the basic input for generating the variables e(i,5) - that is, the random

crime fluctuation for each block during the selection period. This, in turn, is used to generate an estimate of the crime rate, denoted as C'(i,5), which determines the crime performance of the system of Pauly blocks during the selection year.

3.3 SAMPLE DESIGN

The motivation for using a sample of Pauly blocks rather than the whole population of blocks was to reduce the cost of running the computer simulation. The main expense arises with the generation of a random number for each Pauly block and year under study, for a large number of runs. The total number of Pauly blocks in the city is approximately 500.

There is another reason for limiting the number of blocks used in the simulation experiment. Since many of the Pauly blocks have such low mean crime rates, the probability that they would rank among the top six Pauly blocks for any simulation run would be close to zero. Thus, these blocks can be safely omitted from the model due to the fact that the artifact ratio depends only on the generated crime rate for the six highest crime blocks, in this particular application of the problem. The use of a sample of only the topmost elements to represent an entire population is also valid for the general situation.

Those elements in the population with a relatively low estimated average would be chosen for an experiment only in rare instances in which they might experience very abnormal behavior. Under such circumstances, there is little harm in excluding these cases from participation in the simulation experiment.

In view of the above discussion, the sample of blocks finally chosen for the experiment were the 40 blocks which ranked the highest in target crimes in the indicated 10-month period of 1971. The minimal size of the sample had to be 37 in order to insure that those six Pauly blocks actually chosen for the project would be included in the sample range for each of the five years considered. A check was made for those blocks whose crime ranked between forty-first and fiftieth highest in 1971. The highest rank among those ten blocks for any of the five years was twenty-fifth. Thus, these blocks, as evidenced by their history of crime rates, would almost certainly not be among the six highest in crime for any simulation of 1971 crime.

For those 40 Pauly blocks chosen for the sample, various statistics were estimated that characterized the behavior of the system. Figure 7 shows the overall average crime rate for the set of sample blocks for the 10-month periods for each of the five years under study. The plot of this statistic shows a general upward trend for the five years.

FIGURE 7. Mean Crime of Sample for Each of Five Years.

However, the last three years suggest that average crime for the sample set had reached a fairly constant level. Figure 8 shows a plot of each Pauly block in the sample, giving its mean, highest, and lowest crime rates for the five year period in descending order. Those blocks depicted with a dotted line represent the Pauly blocks actually chosen for patrol in the project. This graph, based on the history of crime rates, supports the hypothesis that an artifact situation exists within the system. In considering each block separately, many of them show great variability about their own mean crime rates. In considering the whole system of blocks, it can be observed that those six blocks chosen for the project were not very different from the other blocks in the sample, as indicated by the great amount of overlap in the crime rate ranges of the blocks. In other words, these six blocks might well have not been chosen for patrol if the time period on which the selection was to be made was changed. Figure 9 plots the estimated crime for each Pauly block for 1971 in descending order, and gives a range of possible variation estimated by the standard error of the regression model of Equation (3-1). In this case, the estimated crime for each block is the average number of crimes for 1971. The graph suggests that

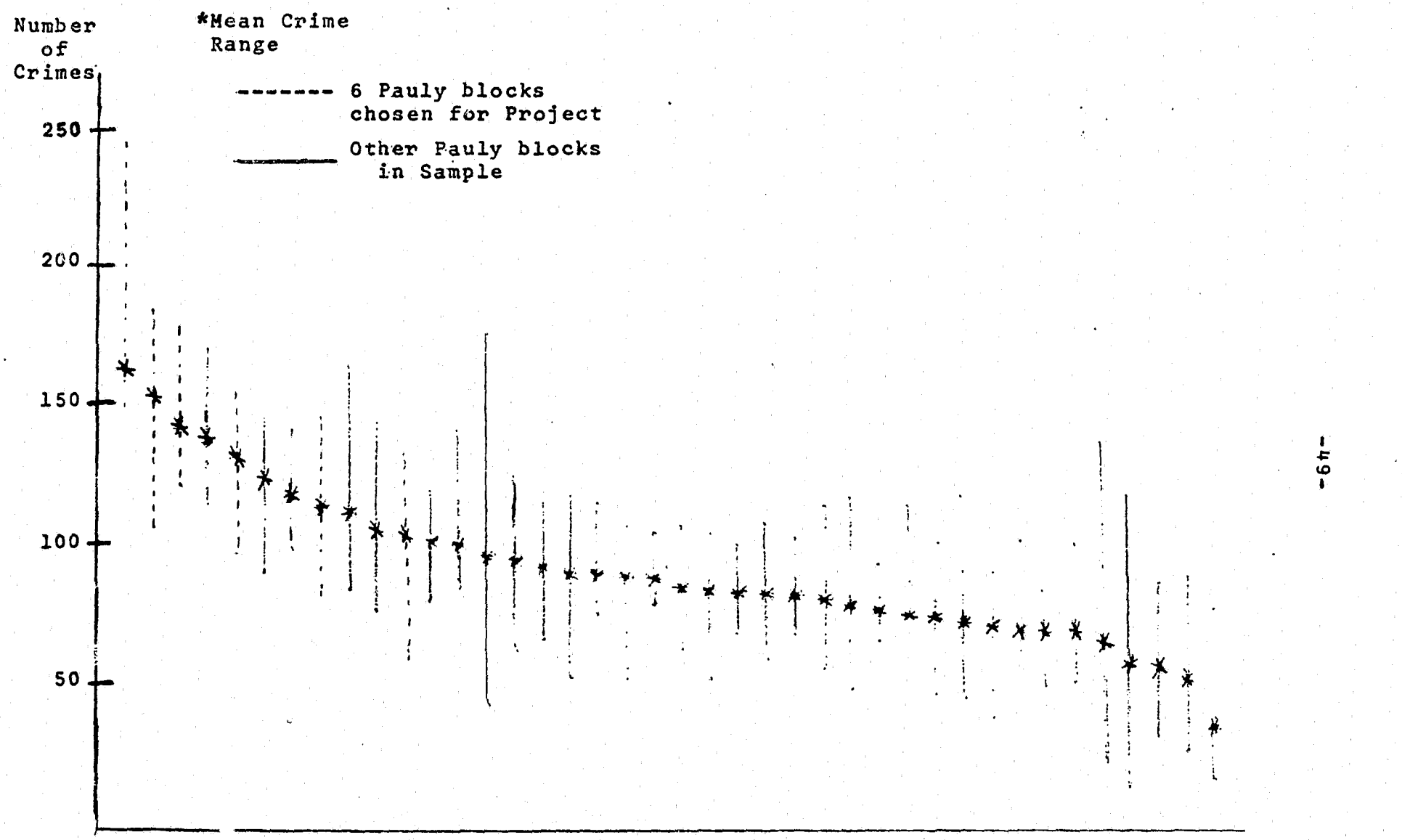


FIGURE 8. Plot of Mean and Range of Crime Rates for Sample Blocks Based on 10-Month Periods, 1967-1971.

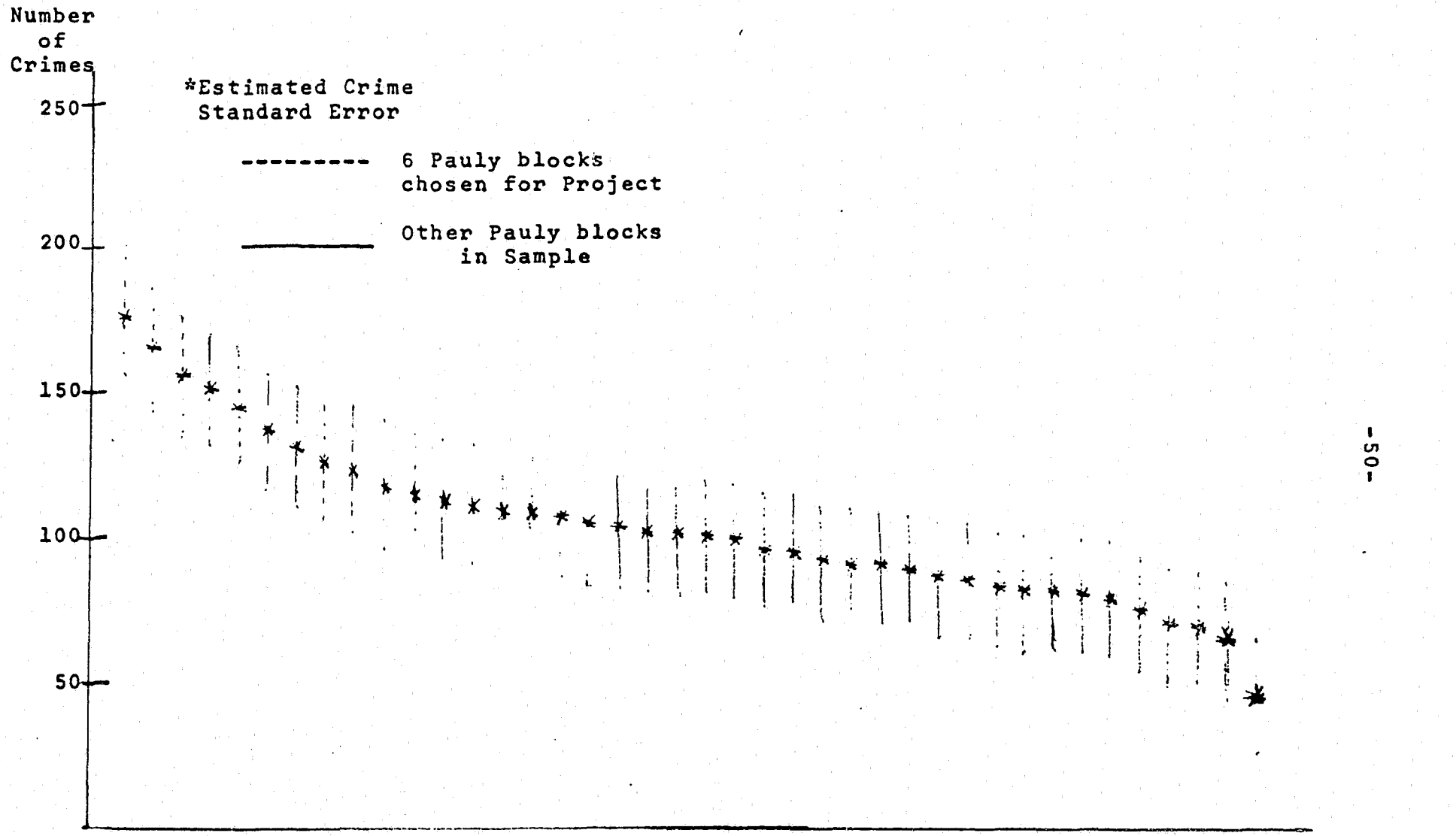


FIGURE 9. Plot of Estimated Crime for 1971 and Standard Error Based on Time-Series Model.

if the selection of the six highest blocks could actually be repeated, a different set could be expected each time due to the variability in the system. These results strongly suggest that an artifact situation exists in the system. The remaining chapters focus on simulating the behavior of crime during the base period 1971 to determine the magnitude of the artifact that can be expected using this particular year as the selection period.

4. FORMULATION OF COMPUTER PROGRAM

This chapter gives the details of the structure of the computer simulation, beginning with the initial conditions of the program and ending with the printout of the final results. The overall logic of the program is discussed, followed by a description of the function of each sub-program. Flow charts are included for easy reference.

4.1 INITIAL CONDITIONS AND INPUT DATA

In this program the following input data is required for each element of the sample considered for the simulation experiment: an estimated mean for the period used in selecting a treatment group and an estimate of the random error for each of those periods used for the time-series model. For the specific application of the Foot Patrol Project, the input data for each block in the system consisted of an underlying average of crime for 1971 and a time-series of the error terms for the five years of crime statistics available for the period prior to implementation of patrols. To duplicate the actual conditions of the Foot Patrol Project, during each iteration of the simulator, after an estimate of the 1971 crime rate was computed by the simulator for each of the 40 blocks modeled, the six highest crime blocks were identified and

their estimated crime rates averaged. The number of simulation runs (iterations) was set to 400.

4.2 GENERAL LOGIC OF THE COMPUTER SIMULATION

The simulation model was embodied in a computer program in FORTRAN. In general, the chronology of the program proceeds in three stages:

1) Preliminary computation
2) Simulation experiment
3) Analysis

The logical structure of the program, as defined by these stages, is shown in Figures 10, 11 and 12. A discussion of the flow of the program within each stage follows.

Stage 1:

Since the random number generator in this program utilizes the cumulative distribution of the random variable being simulated, the time series data for each block has to be converted into such a distribution. (The details of the random number generator are given in the next section). This involves, first constructing a frequency histogram for the error terms. The cumulative distribution can be calculated directly from the results of the histogram. That is,

$$F_i = \frac{\sum_{j=1}^{i} f_j}{n} \quad , \qquad 1 \leq i \leq t$$

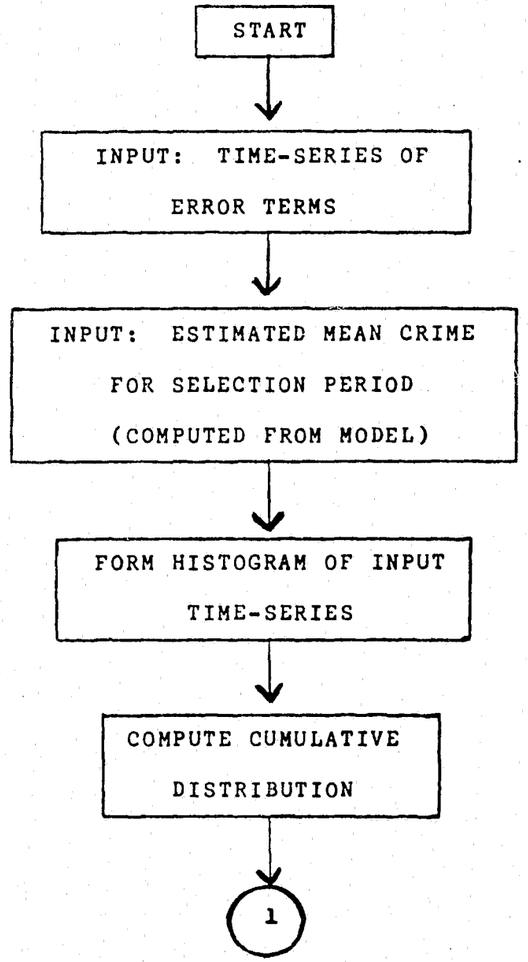


FIGURE 10. Flow chart of Preliminary Computation.

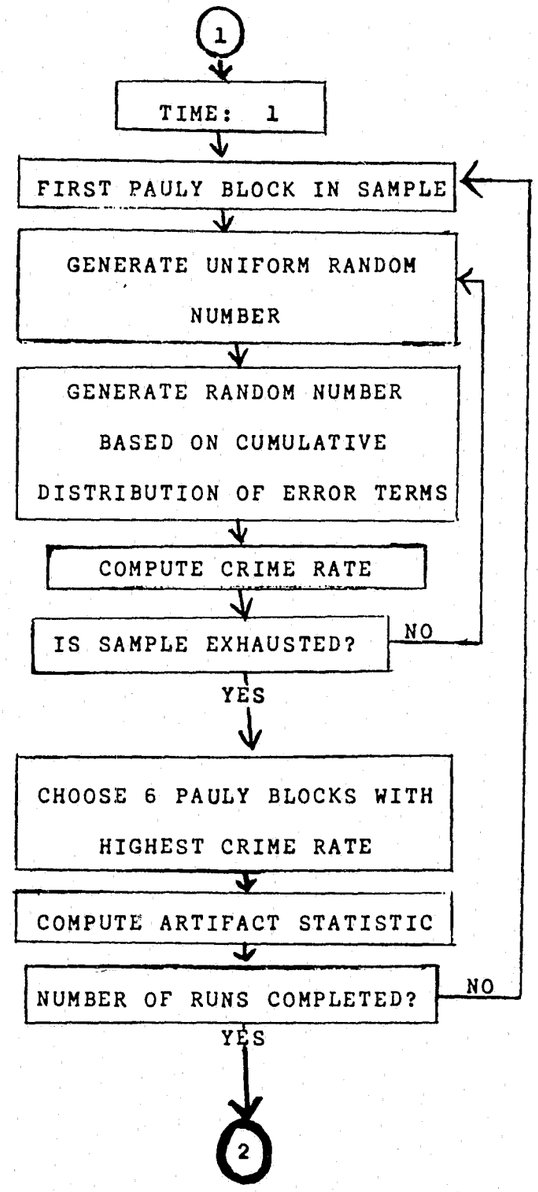


FIGURE 11. Flow chart of Simulation Experiment.

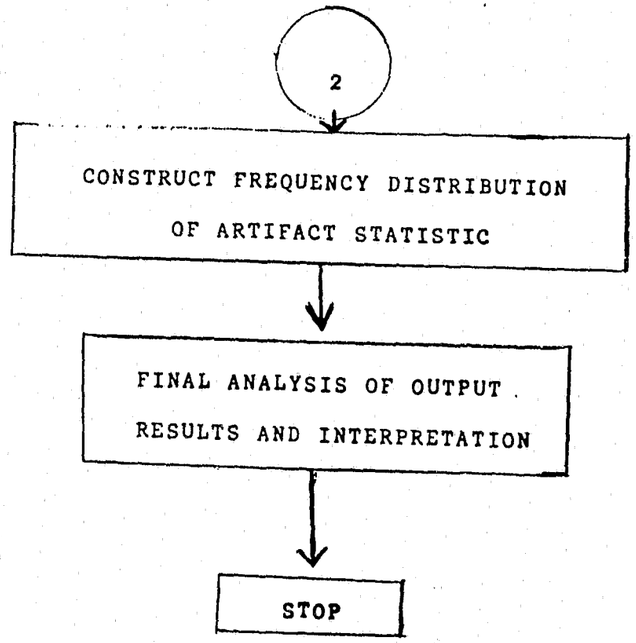


FIGURE 12. Flow chart of Analysis Stage.

where F_i is the cumulative probability for the i^{th} interval,

f_j is the number of observations in the j^{th} interval ($j \leq i$),

n is the total number of observations,

t is the number of intervals.

The cumulative probability distribution has to be stored as an array in the program to be used as the input for each iteration of the simulation experiment.

Stage 2:

Having organized the input data in the necessary form, the simulation experiment can now proceed. The experiment can be described in terms of a two-stage process. The first stage includes those events occurring within the system in a single unit of time - in this case, the selection period for the Foot Patrol Project. In more specific terms, the simulation generates crime rates for each block in the sample for the initial 10-month period. Using the simulated crime rates for the 40 sample blocks, the six blocks whose rates are highest are "selected" for foot patrol operations. The artifact ratio is then estimated. The second stage involves a replication of the simulation experiment: the estimation of crime rates for the base year, the selection of the six highest blocks and the estimation of the artifact ratio. This

process is repeated for the number of runs specified.

Stage 3:

After the completion of the experiment, a frequency distribution is constructed for the artifact statistic, and the mean and standard deviation are computed. The distribution is then used to make probabilistic statements about the magnitude of the regression artifact.

The next section explains the subroutines used for storage of pertinent information, the process of generating random variates, and the recording of statistics concerning the performance of the system.

4.3 DATA GENERATION

4.3.1 Random Number Generator

The subroutine RANDN is used to generate random numbers based on the inverse transformation method.[7] The advantage of this method is that its flexibility lends itself to be applied to any probability distribution, both theoretical and empirical, discrete and continuous.

Since uniformly distributed random variates play a major role in the generation of random variates drawn from other probability distributions, the basic prerequisite for this method is that a sequence of independent random variates, each with a uniform distribution on the interval [0,1], can be generated. The particular source which

was used to generate the uniform random numbers was a subprogram in the Scientific Subroutine Package.

The rationale for such a technique is as follows. Let f(x) represent the density function of the particular statistical population for which it is desired to generate random variates, X_s. Let F(x) denote the corresponding cumulative distribution function, that is, the probability that a random variable X takes on the value of x or less. For the continuous case, this can be computed by

$$F(x) = \text{Prob}(X \leq x) = \int_{-\infty}^{x} f(t)dt \quad ,$$

and for the discrete case,

$$F(x) = \text{Prob}\,(X \leq x) = \sum_{\substack{s \\ \text{such that} \\ X_s \leq x}} P_s \quad ,$$

where p_s denotes the probability of the random variable taking on the value X_s. Let u denote a uniform random variate such that the probability density function is

$$r(u) = \begin{cases} 1 & 0 \leq u \leq 1 \\ 0 & \text{Elsewhere} \end{cases}$$

and the cumulative distribution is

$$R(u) = \begin{cases} 0 & u < 0 \\ u & 0 \leq u \leq 1 \\ 1 & u > 1. \end{cases}$$

The cumulative distribution of x, F(x), also defined over the range 0 to 1, can be used as the intermediary in the transformation process of generating a uniform random variate u to the generation of the random variate x, of the desired probability distribution. Thus, the procedure involves generating uniformly distributed numbers and setting F(x) = u. At this point, one can approach the problem in one of two ways, depending on the form of F(x).

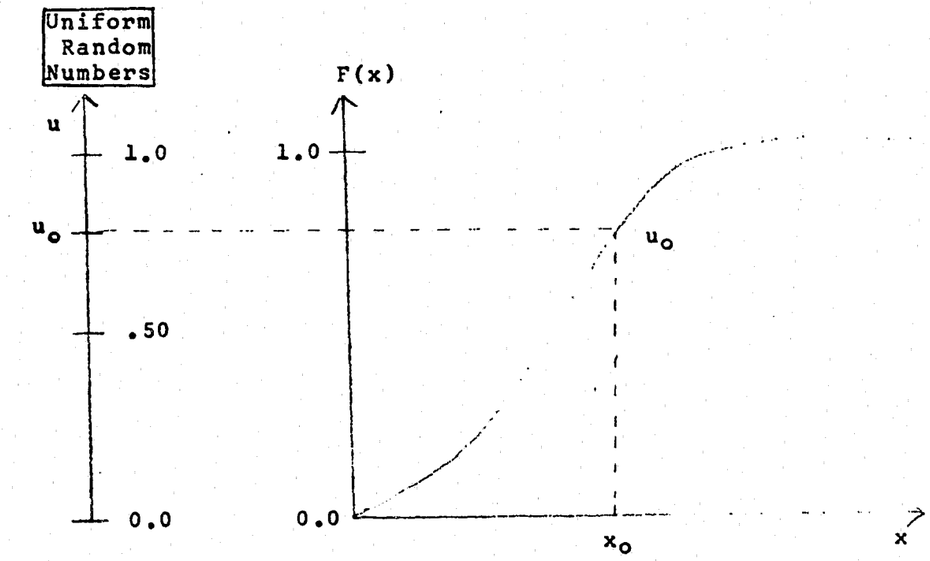
Case I:

The following procedure can be used only if x is uniquely determined by u = F(x), that is, if there exists an inverse function of x, $F^{-1}(u)$ (see Figure 13). Then it follows that for any particular value of u, say u_o, it is possible to find the value of x, namely x_o, which corresponds to it, through the inverse function of F, if it is known. That is,

$$x_o = F^{-1}(u_o),$$

where $F^{-1}(u_o)$ is the inverse transformation of F, taking u_o from the unit interval to the domain of x.

This procedure was used for some of the distributions that were involved in validating the simulation model, as summarized in Table 3 .



Cumulative distribution function of x

FIGURE 13. Generation of Random Numbers from Specific Continuous Distribution.

TABLE 3. Generation of Random Numbers by Inverse Transformation Method.

Distribution for Generation of Random Numbers	Corresponding Density Function	Transformation Function (where u is uniform random variate)
Exponential	$f(x) = \lambda e^{-\lambda x}$, $\lambda > 0$, $x \geq 0$	$x = (\frac{1}{2}) \log u$
Triangular	$f(x) = \begin{cases} 4 & 0 \leq x \leq 1 \\ 2-x & 1 \leq x \leq 2 \end{cases}$	$x = \begin{cases} \sqrt{2u} & 0 \leq u \leq 1/2 \\ 1+\sqrt{2u-1} & 1/2 \leq u \leq 1 \end{cases}$

However, for many probability functions it is extremely difficult or impossible to express x in terms of an inverse transformation, $F^{-1}(u)$. Thus, a more general version of this approach must be taken, which is particularly applicable to 1) empirically estimated probability distributions, 2) discrete distributions, and 3) some continuous functions with no simple inverse transformation function and which can be approximated by a discrete distribution.

Case II:

In this case, it is necessary to compute numerically the cumulative distribution for a given interval. This may be expressed as

$$F_j = \frac{1}{n} \sum_{i=1}^{j} f_i \qquad 1 \leq j \leq n \ ,$$

where F_j is the cumulative distribution up to and including the $j^{\underline{th}}$ interval, f_i is the number of observations falling in the $i^{\underline{th}}$ interval and n is the total number of observations. A uniform random variate u_o is then generated, and a searching process is performed to find the intervals for which the relation

$$F_{j-1} < u_o \leq F_j$$

holds. Having determined the proper interval, the method used to find x_o is arbitrary. Some of the options include using the lower interval limit or the midpoint of the $j^{\underline{th}}$

interval. For the simulation experiment in this study, x_o was determined by interpolating within the j^{th} interval, as bounded by x_j and x_{j+1} (see Figure 16). Since the random variable of interest assumes only integer values, then the largest integer in x_o was used as the variate. Moreover, the estimated cumulative distribution of error terms was used as the input distribution for generating the random variates. Then, at the end of this subroutine, the sum of the generated error term and the average crime rate (estimated from the time-series model) for the selection period for each Pauly block is computed; this sum represents the simulated crime behavior of each Pauly block in the sample during the selection period.

4.3.2 Description of Record-Keeping

The following subroutines have been developed to perform the record-keeping process for all three stages of the computer simulation, as discussed in the previous section. The variable names used in these subroutines are defined in Appendix 9.

HIST SUBROUTINE

The function of this subroutine is to compute the frequency distribution for an array of numbers. The subroutine first determines the limits for each of the intervals in the distribution, then takes each number of the input array and performs a search for the appropriate

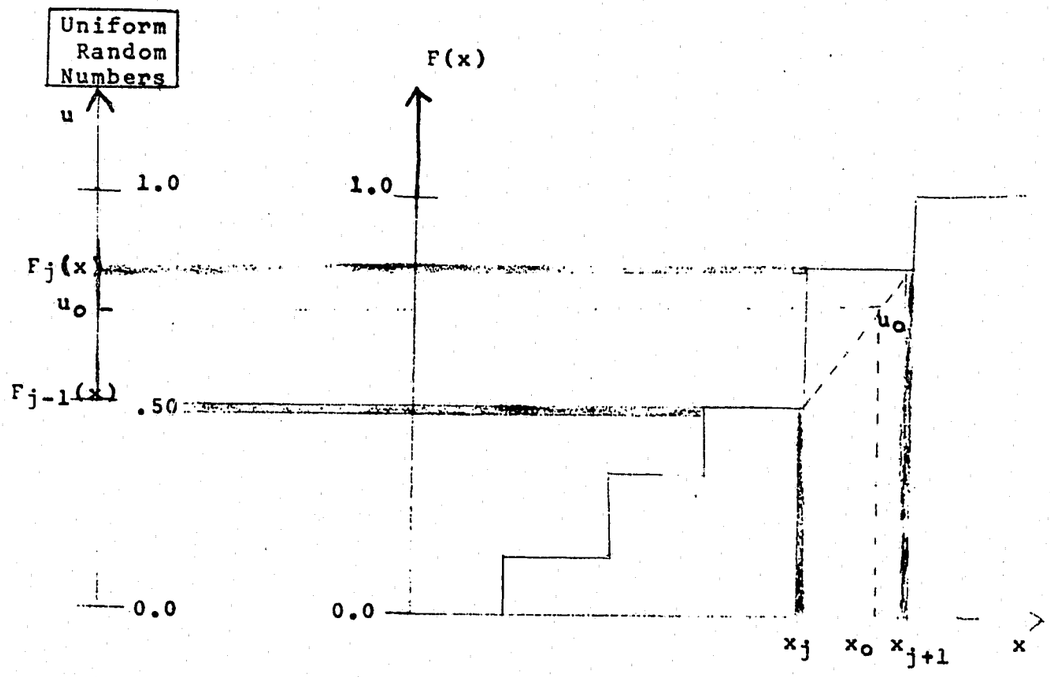


FIGURE 16. Generation of Random Numbers from Specific Discrete Distribution.

interval in which it falls. After each number has been placed in the correct interval, the frequency count for this interval is updated.

This subroutine is called from the main program in two instances: to form the frequency distribution of the time-series of error terms at the start of the program and the frequency distribution of the artifact statistic computed in the simulation experiment. For the latter, the distribution is stored in an array and printed after the experiment is completed.

CMPROB SUBROUTINE

After the frequency distribution of the error terms is determined for each Pauly block, the program proceeds to this subroutine to compute the cumulative distribution. Stored as an array, this distribution remains unchanged throughout the program.

MAXM SUBROUTINE

Following the subprogram RANDN, this subroutine ranks the Pauly blocks in descending order by estimated crime rate. To do this, the method involves the comparison of two numbers, as they appeared in the original array, and interchanging the order of the two numbers if, in fact, the latter number was the larger. The logical check continues until the six highest crime rates have been placed in the proper order sequence. This subset is stored in an array for each run, along with the

corresponding expected crime rate for each of the six highest Pauly blocks.

AVERG SUBROUTINE

This subroutine computes the mean, variance and standard deviation for a series of numbers. The subroutine is called a number of times from the main program. It is needed to compute the above statistics for the generated crime rates of the six highest blocks and similar statistics for the expected crime rates of these blocks. These results are used to compute the artifact ratio for each run, which is stored in an array. It is also used to compute the mean and standard deviation of the distribution of the artifact ratio at the end of the program.

4.3.3 Generation of Statistics and Final Output

The artifact ratio is computed at the end of each run of the experiment, based on those six Pauly blocks which have the highest generated crime rates.

Let B(n) denote the set of numbers identifying the six highest crime blocks for the $n^{\underline{th}}$ run. Then for the $i^{\underline{th}}$ block in this set C'(i,n) represents the generated crime rate for this run and $\hat{C}(i,5)$ represents the estimated norm of crime behavior for the 1971 selection period, as determined by basic model. The artifact ratio for the

$n^{\underline{th}}$ run is computed as

$$A(n) = \frac{\sum_{i \in B(n)} C'(i,n)}{\sum_{i \in B(n)} \hat{C}(i,5)} .$$

From the identity

$$C'(i,n) = \hat{C}(i,5) + e'(i,n),$$

where e'(i,n) is the generated error term for block i,

an alternative expression of the artifact ratio is

$$A(n) = 1 + \frac{\sum_{i \in B(n)} e'(i,n)}{\sum_{i \in B(n)} \hat{C}(i,5)} .$$

A printout of the final results displays the frequency distribution of this statistic for 400 runs of the experiment, and the mean and standard deviation of this distribution. A sample report is shown in Table 4.

CONTINUED

1 OF 2

TABLE 4. Sample Output Report of Program.

Interval	Lower Interval Limit of Artifact Statistic	Observed Frequency
1	0.80	0
2	0.85	0
3	0.90	0
4	0.95	0
5	1.00	0
6	1.05	0
7	1.10	2
8	1.15	8
9	1.20	13
10	1.25	40
11	1.30	66
12	1.35	106
13	1.40	74
14	1.45	59
15	1.50	23
16	1.55	7
17	1.60	2
18	1.65	0
19	1.70	0
20	1.75	0

Number of Runs = 400 Mean = 1.3845

Standard Deviation = .08578

5. VALIDATION OF SIMULATION MODEL

The distribution of the average of a set of extreme observations is a valuable tool for determining the inflation that can be expected in using the mean of these observations as an estimate of the population mean. For this reason, the crux of the simulation model is to establish a basis of confidence in the reliability of the distribution of the average, which it generates, as being representative of the population distribution. This chapter includes two approaches used in validating the simulation model: the order statistics results, derived in Chapter 2, and a coin-tossing experiment.

5.1 VALIDATION OF MODEL FROM ORDER STATISTICS RESULTS

For the case of continuous probability distributions, the order statistics results were used in validating the simulation model in order to achieve confidence in its output. Table 5 summarizes these results, for each of three density functions and a given sample size. These results represent the theoretical probability distributions of the average of extreme observations; the simulation results represent sample probability distributions.

To evaluate the reliability of the simulation distribution, the Kolmogorov-Smirnov statistic,[8] based on the cumulative distribution function, has been used to

TABLE 5. Distribution of the average of the two highest observations for a given sample size.

Density Function	Range of z $z = \frac{y_{(n-1)}+y_{(n)}}{2}$	SAMPLE SIZE		
		n = 2	n = 3	n = 4
Uniform [0,1]	$0 \le z \le 1/2$	$4z$	$6z^2$	$8z^3$
	$1/2 \le z \le 1$	$4-4z$	$24z-18z^2-6$	$-56z^3+96z^2-48z+8$
Exponential	$0<z<\infty$	$4z\lambda^2 e^{-2\lambda z}$	$12z\lambda^2 e^{-2\lambda z}+12\lambda e^{-3\lambda z}$ $-12\lambda e^{-2\lambda z}$	$24\lambda^2 z e^{-2\lambda z}$ $+48\lambda e^{-32z}$ $-36\lambda e^{-2\lambda z}-12\lambda e^{-4\lambda z}$
Triangular	$0 \le z \le 1/2$	$8/3z^3$		
	$1/2 \le z \le 1$	$-8z^3+16z^2-8z+4/3$	———	———
	$1 \le z \le 3/2$	$8z^3-32z^2+40z-44/3$		
	$3/2 \le z \le 2$	$-8/3z^3+16z^2-32z+64/3$		

compare the simulation results with the theoretical results derived from the order statistics. Thus, if $F_n(X)$ denotes the sample distribution function, generated by the simulation, and F(X) denotes the theoretical (or population) distribution function, obtained from the order statistics model, the hypothesis to be tested assumes that F(X) is the appropriate population distribution from which $F_n(X)$ has been obtained.

It is the absolute difference $|F_n(X) - F(X)|$ that is used in the Kolmogorov-Smirnov test. The rationale of the test is that if the sampling distribution, $F_n(X)$, differs from the expected distribution by too much, this may serve as grounds to reject the hypothesis that F(X) is the correct assumed distribution from which $F_n(X)$ has been derived. The statistic then becomes

$$Dn(X) = \max_{\text{all } X} |F(X) - Fn(X)|.$$

A test of validity was performed for each density function used in the order statistics model: the uniform, exponential and triangular distributions. However, due to the limitations of these results, the simulation model was set up to form the distribution of the average of only the two highest observations from a certain sample

size, n. The distribution of the average was based on 75 observations - that is, the number of simulation runs - for each density function.

Figures 15, 16 and 17 display the cumulative distributions from both the theoretical and simulation models for each density function; Tables 6 , 7 and 8 , following the graphs, indicate the absolute deviation between the two cumulative distributions for a given interval. The additional columns pertain to the Kolmogorov-Smirnov test.

In summarizing the test results, for each density function the null hypothesis was accepted at the 95% confidence level. That is to say, F(X) is the population distribution function for the sample distribution, Fn(X).

In Table 9 , the mean of the distribution of $\frac{Y(n-1) + Y(n)}{2}$ - that is, the average of the two highest observations for a sample size n, is given for both the order statistics model and the simulation model. It can be be observed that the sample mean (derived from the simulation results) is very close to the theoretical mean.

This provides validation of the logical structure of the simulator for these simple distributions, necessary in order to achieve confidence in simulation results for more complicated density functions and averages of the set of the n highest observations when n is greater than two.

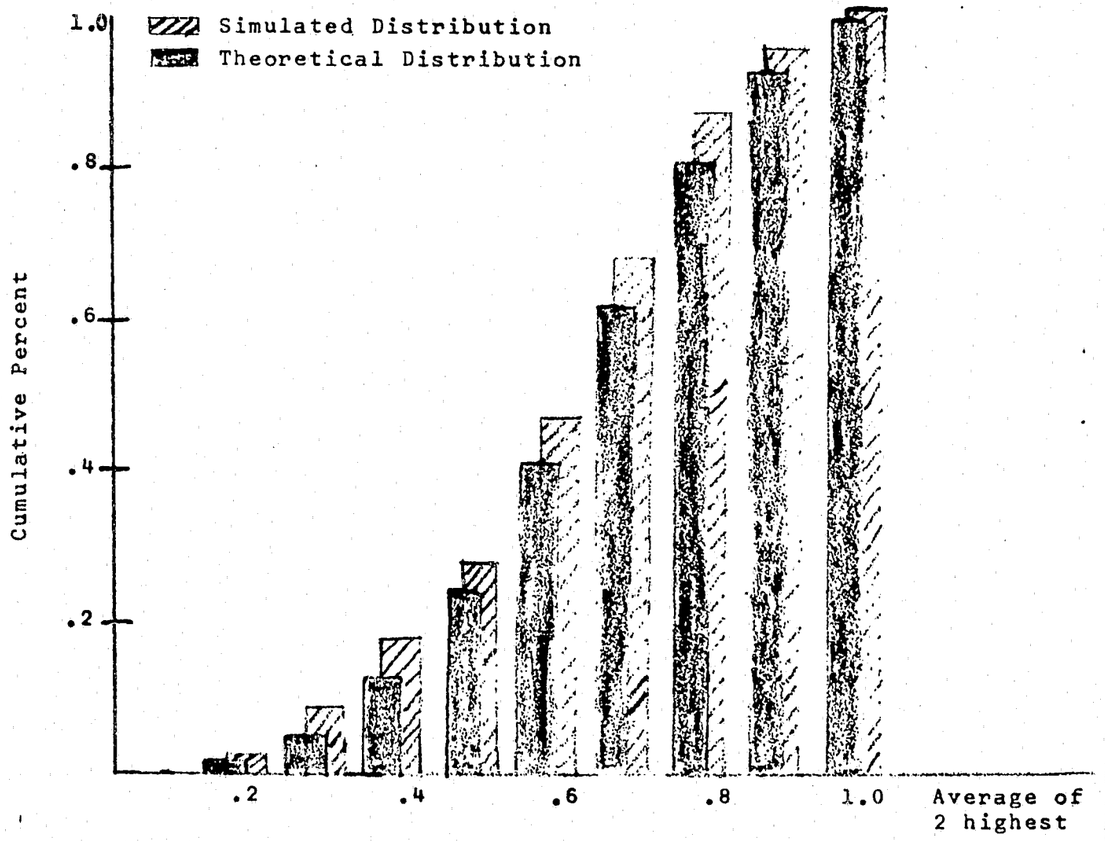


FIGURE 15. Uniform Cumulative Distributions - Average of Top 2 out of 3 Random Numbers.

TABLE 6. Comparison of Simulated and Theoretical Uniform Distributions.

Range of Average	Absolute Deviation of Simulated from Theoretical Distribution	Maximum Deviation	Test Statistic (95% Confidence)
.0-.1	.0113		
.1-.2	.0106		
.2-.3	.0260		
.3-.4	.0453		
.4-.5	.0513		
.5-.6	.0640		
.6-.7	.0660		
.7-.8	.0667	.0667	.1550
.8-.9	.0220		
.9-1.0	.0080		

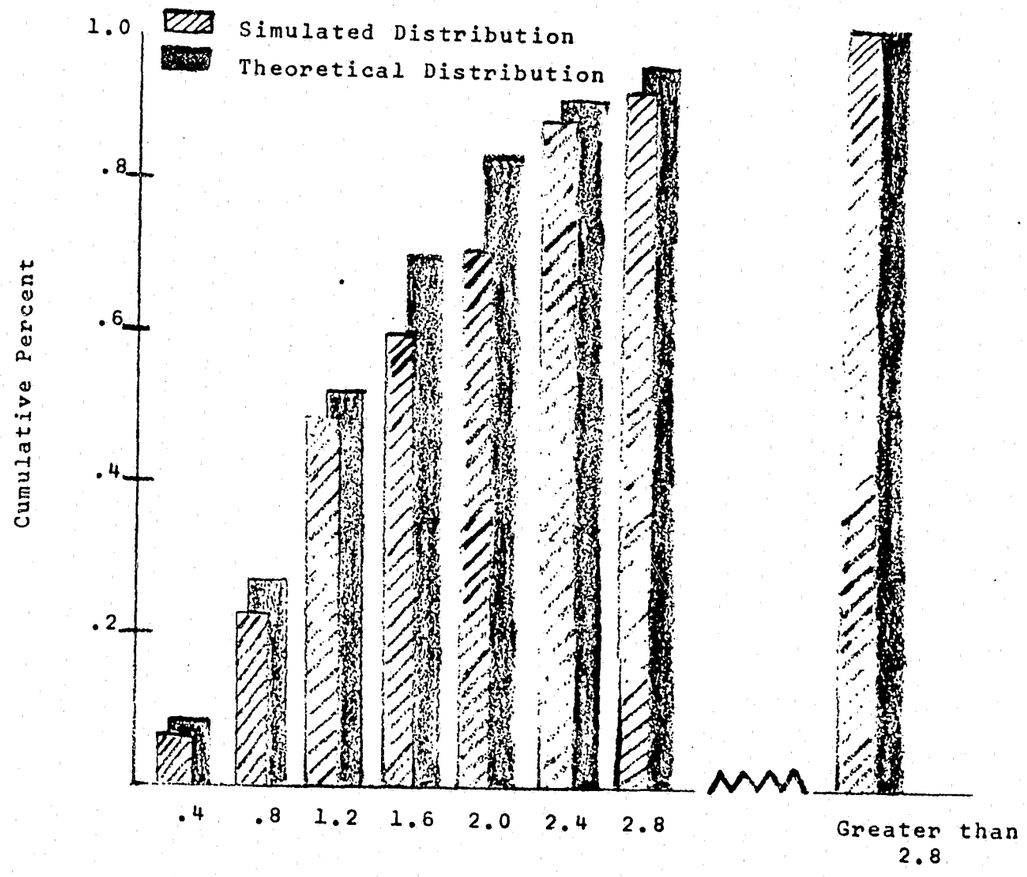


FIGURE 16. Exponential Cumulative Distributions ($\lambda = 1$) - Average of Top 2 out of 3 Random Numbers.

TABLE 7. Comparison of Simulated and Theoretical Exponential Distributions.

Range of Average	Absolute Deviation of Simulated from Theoretical Distribution	Maximum Deviation	Test Statistic (95% Confidence)
.0-.4	.0152		
.4-.8	.0470		
.8-1.2	.0297		
1.2-1.6	.1114		
1.6-2.0	.1319	.1319	.1550
2.0-2.4	.0232		
2.4-2.8	.0281		
Greater than 2.8	.0000		

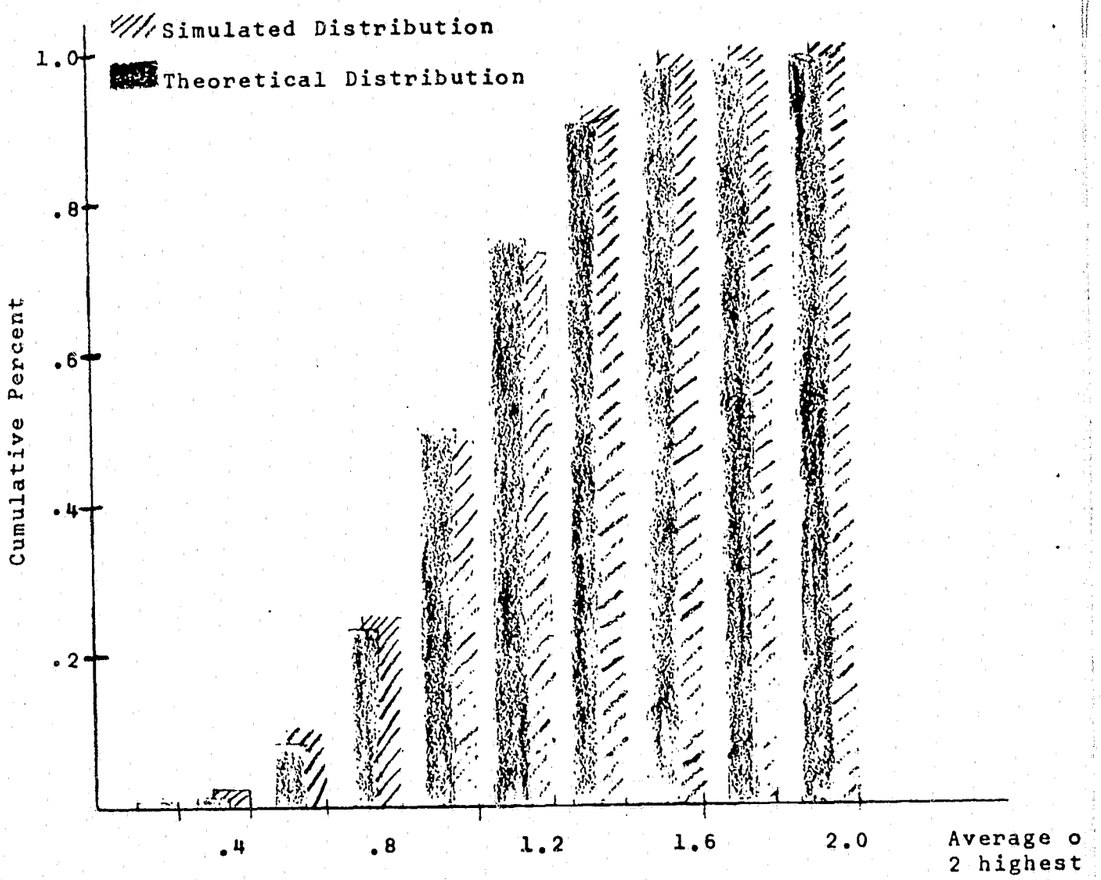


FIGURE 17. Triangular Cumulative Distributions - Average of Top 2 out of 2 Random Numbers.

TABLE 8. Comparison of Simulated and Theoretical Triangular Distributions.

Range of Average	Absolute Deviation of Simulated from Theoretical Distribution	Maximum Deviation	Test Statistic (95% Confidence)
.0-.2	.0010		
.2-.4	.0096		
.4-.6	.0205	.0205	.1550
.6-.8	.0018		
.8-1.0	.0198		
1.0-1.2	.0147		
1.2-1.4	.0204		
1.4-1.6	.0057		
1.6-1.8	.0065		
1.8-2.0	.0050		

Table 9. Mean of Distribution of $\frac{Y(n-1) + Y(n)}{2}$

Distribution	Simulation Model	Order Statistics Model
Uniform [0,1] (top 2 out of 3)	.588	.625
Exponential $\lambda = 1$ (top 2 out of 3)	1.482	1.333
Triangular (top 2 out of 3)	1.008	.968

5.2 COIN-TOSSING EXPERIMENT

Since the order statistics model, as derived in Chapter 2, is not applicable to discrete random variables, a coin-tossing experiment was used as a means of validating the simulation results for the binomial distribution. The experiment proceeded in the following manner:

1) Toss five identical coins simultaneously.
2) Record the number of heads obtained (the maximum being five heads).
3) Repeat steps 1-2 for a total of five times.
4) Choose three out of the five tosses which have the highest number of heads.
5) Average the number of heads from the top three tosses obtained in Step 4.
6) Repeat steps 1-5 for 75 runs.
7) Construct a histogram for the distribution of the average (of the highest three tosses for each trial).

A similar procedure was followed for the simulation experiment. Five random numbers were generated from the binomial distribution,

$$P(X = x) = \binom{n}{x} p^5 q^{n-x} \quad ,$$

where x is the number of heads obtained in one trial. In order to reproduce the initial conditions of the coin-tossing experiment, the sample size, n, was set equal to

the number of coins tossed and p(the probability of getting a head on a single toss) was set to .5, assuming the coins to be unbiased. Thus, the binomial probability function for the simulation experiment was

$$P(X = x) = \binom{5}{x}\left(\frac{1}{32}\right).$$

The three highest tosses of heads out of five were then averaged, and this was also repeated for 75 trials. The frequency distribution of this average was then tabulated.

In comparing the distributions of the average obtained from each experiment, it is necessary to test the hypothesis that these two distributions were obtained from the same population distribution. For this case, the two-sample Kolmogorov-Smirnov test must be used.

Although random sampling fluctuations can introduce a difference in the two sample distributions even if the samples are from the same population distribution, it is a large discrepancy between the sample distribution functions that serves as a reasonable basis to reject the null hypothesis of the test.

A comparison of the cumulative distributions is shown in Figure 18, and the deviation between the two distributions for each interval is given in Table 10 which

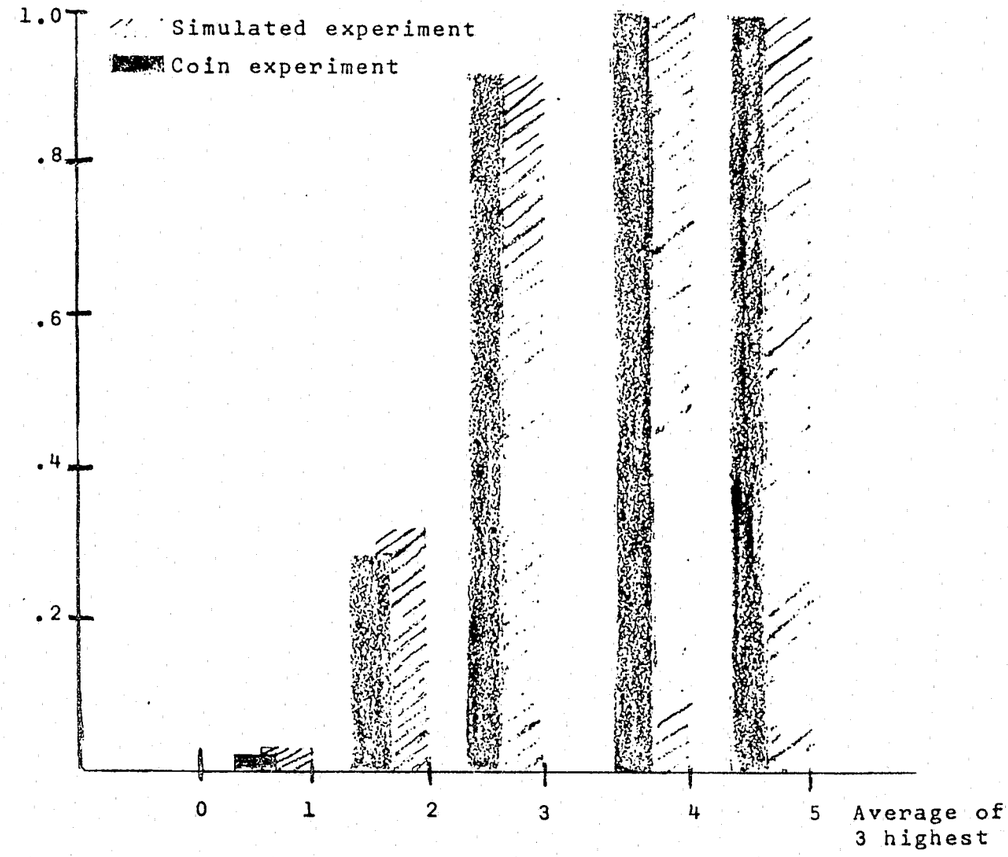


FIGURE 18. Sample Cumulative Distributions: Binomial (p=.5), Average of top 3 out of 5 Random Numbers.

TABLE 10. Comparison of 2 Sample Binomial Distributions.

Range of Average	Absolute Deviation of Simulated from Coin Experiment Distribution	Maximum Deviation	Test Statistic (95% Confidence)
0.-1.	.0000		
1.-2.	.0145		
2.-3.	.0400	.0400	.1550
3.-4.	.0000		
4.-5.	.0000		

follows. Since the maximum deviation between the two cumulative distributions is less than the critical value at the 95% confidence level, the hypothesis that the two sample distributions were derived from the same population distribution is accepted.

In this case, the population of the two samples is assumed to be binomial. Theoretically speaking, the coin can be categorized as a binomial experiment. To verify this, a Chi-Square goodness of fit test was used to determine how good the sampling distribution from the coin-tossing experiment approximates the population distribution. The Chi-Square statistic for the test was 5.7; for five degrees of freedom at the 95% confidence level, the critical value is 11.1. Thus, the hypothesis that the coin experiment has a binomial distribution can be accepted.

From the results of the foregoing tests (Chi-Square and Kolmogorov-Smirnov), one can conclude that the distribution of $\frac{Y(n-1) + Y(n)}{2}$, as generated by the simulation model, is also representative of the true (or theoretical) distribution.

Finally, the means for each of the sample distributions of $\frac{Y(n-1) + Y(n)}{2}$ were as follows:

Simulation experiment: 3.09

Coin-tossing experiment: 3.13

These results validate the use of the simulation model for processes involving discrete random variables. This is a particularly useful capability, since in many real-life experiments the related probability density functions are unknown and empirical estimates in the form of histograms must be used, thereby representing a continuous phenomenon with a discrete approximation.

6. SIMULATION EXPERIMENT AND RESULTS

The simulation experiment was run a number of times, computing the probability distribution of the artifact ratio and its expected value. These results were the measures used to reflect the crime performance of the system of Pauly blocks. This was done for three different experiments, in which the input data had been modified. A description of these experiments and their results are presented in this chapter. Since conclusions based on results of the simulation runs depend on the validity of the assumptions made in this study, some of these have been tested at the end of the chapter, along with suggestions for validating them.

6.1 SIMULATION EXPERIMENT

Crime rates during the 10-month selection period January to October 1971 for the Foot Patrol Project were simulated for a set of 40 Pauly blocks. The purpose of the simulation was to study the variability in the crime rates in each block and how this would effect this selection of high crime blocks for foot patrol. The simulator provided estimates of the results which could be expected if this selection process were replicated many times. The magnitude of the regression artifact for the set of highest blocks for the selection period depends on the deviation of the simulated crime rates for the set from its estimated normal crime behavior. In

repeating the selection process, the magnitude of the regression artifact itself would be expected to change and, therefore, must be considered a random variable, for which a probability distribution may be estimated on the basis of 400 iterations of the simulation experiment. Thus, the crime performance of the system of Pauly blocks can be evaluated in terms of the magnitude of the estimated artifact for each run of the simulator, and the probability of getting this bias.

Three replications of the experiment, using an estimate of the expected crime rates for each Pauly block for the selection period and the distribution of error terms as the input data, were made. Each experiment used a different random number seed for the program's random number generator. The average of these results in terms of the probability distribution of the artifact rates was used as the final distribution for any analysis.

Other experiments were made with the input data modified. The basic time series model used to estimate the error terms for each block for the input data, was assumed to be a linear, additive model. However, this model will actually overestimate the error terms, and therefore, overestimate the artifact ratio, if there is another model that is a "better fit" for the crime data." "Better fit", in this case, refers to a smaller sum of squares of unexplained variation about the regresson line

or curve. To allow for such a possibility, a sensitivity analysis of the output results of the artifact ratio was performed by reducing the series of error terms by a constant fraction. Two additional experiments of this nature were made, with the error terms reduced to 1/2 and 3/4 of their original size. Three runs were also made for each of these experiments, starting with a different seed each time. The results of each experiment are presented in the next section; moreover, a graphic representation of each experiment displays the findings of the sensitivity analysis.

6.2 OUTPUT RESULTS OF SIMULATION EXPERIMENT

Before presenting the results for each simulation experiment, it is necessary to point out that the particular results obtained are dependent on three major factors:

1) the validity of the assumptions on which the time series model was based.
2) the size of the treatment group for the experiment.
3) the size of the population from which the treatment group is selected.

Thus, caution must be taken in using these results to make any generalizations about the nature of the regression artifact in other situations. Each experiment must be treated separately within its own setting. However, it is the approach to evaluate this artifact which may be applied to other cases.

The results for each experiment are presented in the form of a cumulative probability distribution in Table 11. In using the cumulative distribution, probabilistic statements can be made about the occurrence of a selection bias, measured by the artifact ratio, of any given size. That is, let A_o denote the "before" artifact ratio, based on actual crime rates and estimated mean crime rates for those six Pauly blocks chosen for the project in 1971. The chances of getting an artifact at least as large as A_o can then be determined. This also gives an indication of the crime performance of the system of Pauly blocks, if it could be observed for a large number of "selection" periods.

For the Foot Patrol Project, the base period, used to select blocks for patrol, was January to October 1971. If B denotes the indices of set of the six highest crime Pauly blocks, chosen for patrol in the project, then the artifact measure, in this case, is

$$A_o = \frac{\sum_{i \varepsilon B} C(i,5)}{\sum_{i \varepsilon B} \hat{C}(i,5)} ,$$

where $C(i,5)$ represents the actual crime rate and $\hat{C}(i,5)$ represents the estimated mean crime rate for the selection period for each of the six Pauly blocks chosen for the

TABLE 11. Output Results of Simulation Experiments.

Range of Artifact Ratio	FREQUENCY DISTRIBUTION			CUMULATIVE DISTRIBUTION		
	Experiment 1	Experiment 2	Experiment 3	Experiment 1	Experiment 2	Experiment 3
1.00-1.05	0	0	1	0.000	0.000	0.003
1.05-1.10	0	0	43	0.000	0.000	0.025
1.10-1.15	2	3	201	0.005	0.008	0.350
1.15-1.20	5	30	137	0.018	0.080	0.860
1.20-1.25	17	105	18	0.060	0.350	1.000
1.25-1.30	40	160	0	0.160	0.750	--
1.30-1.35	67	87	0	0.330	0.960	--
1.35-1.40	95	14	0	0.580	0.990	--
1.40-1.45	78	1	0	0.740	1.000	--
1.45-1.50	55	0	0	0.900	--	--
1.50-1.55	30	0	0	0.970	--	--
1.55-1.60	8	0	0	0.990	--	--
1.60-1.65	3	0	0	1.000	--	--
1.65-1.70	0	0	0	--	--	--
Mean Standard Deviation	1.388 0.0842	1.258 0.0478	1.141 0.0336			

Experiment 1: e(i,j) - original input data

Experiment 2: 3/4 e(i,j)

Experiment 3: 1/2 e(i,j)

project. This ratio was computed to be 1.27. The likelihood of this selection bias is given in Table 12. A graphical representation of this table is shown in Figure 19, in which the cumulative distribution is approximated as a continuous function. The height of the shaded section to the left of A_o denotes the probability of getting a bias less than 1.27; the height of the shaded section to the right denotes the probability of getting a bias of 1.27 or greater. As can be observed from these graphs, the chances of getting a selection bias greater than 1.27 quickly diminishes as the variability of random fluctuations in crime within each block is reduced in magnitude.

With the unscaled error terms estimated from the original data in experiment 1, the distribution function indicates that 78 out of 100 times, or selection periods, the artifact ratio will be at least 1.27. This indicates that for the six blocks chosen for patrol in the Foot Patrol Project, it is highly probable that the reduction in the target crime in the subsequent time period, comparable to the selection period in terms of duration and season, will be at least 27%, due to the selection process.

TABLE 12. Probability Statements about the Artifact Ratio for Each Experiment.

	Prob($A \leq 1.27$)	Prob($A \geq 1.27$)
Experiment 1	.22	.78
Experiment 2	.83	.17
Experiment 3	1.00	0.00

Experiment 1: e(i,j) - original data

Experiment 2: 3/4 e(i,j)

Experiment 3: 1/2 e(i,j)

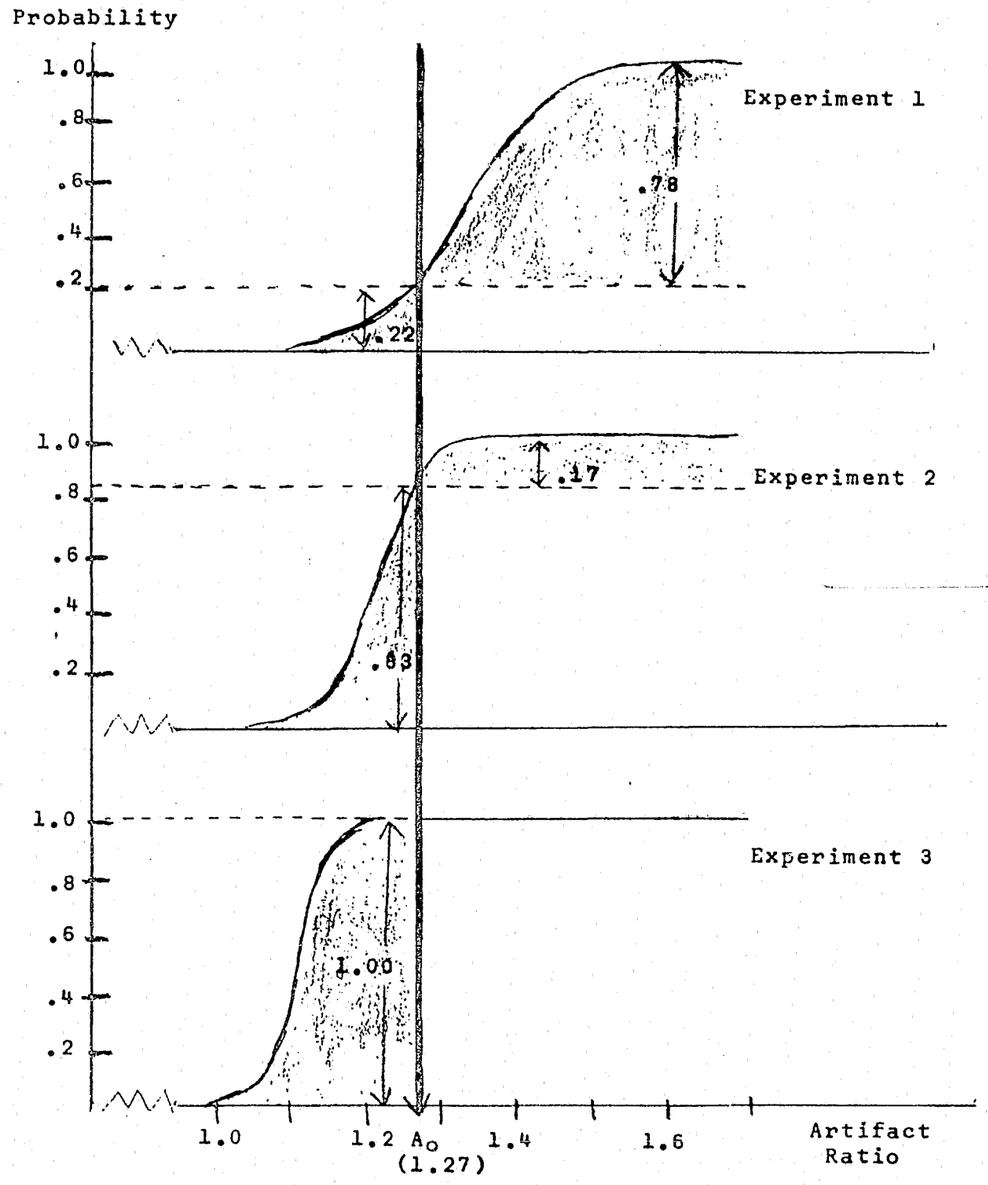


FIGURE 19. Cumulative Distribution for Each Experiment Showing the Probabilities Relating to a Change as Large as that Observed in the Foot Patrol Project.

The average artifact for each experiment has been shown in Table 11. The 1.39 average for experiment 1 indicates that an inflation of 39% above the normal crime totals can be expected in the selection period for those six blocks chosen, considering the manner in which they were selected.

The other two experiments suggest that the expected bias is a function of the magnitude of the random crime fluctuations in each block. Figure 20 shows a graphical representation of these results.

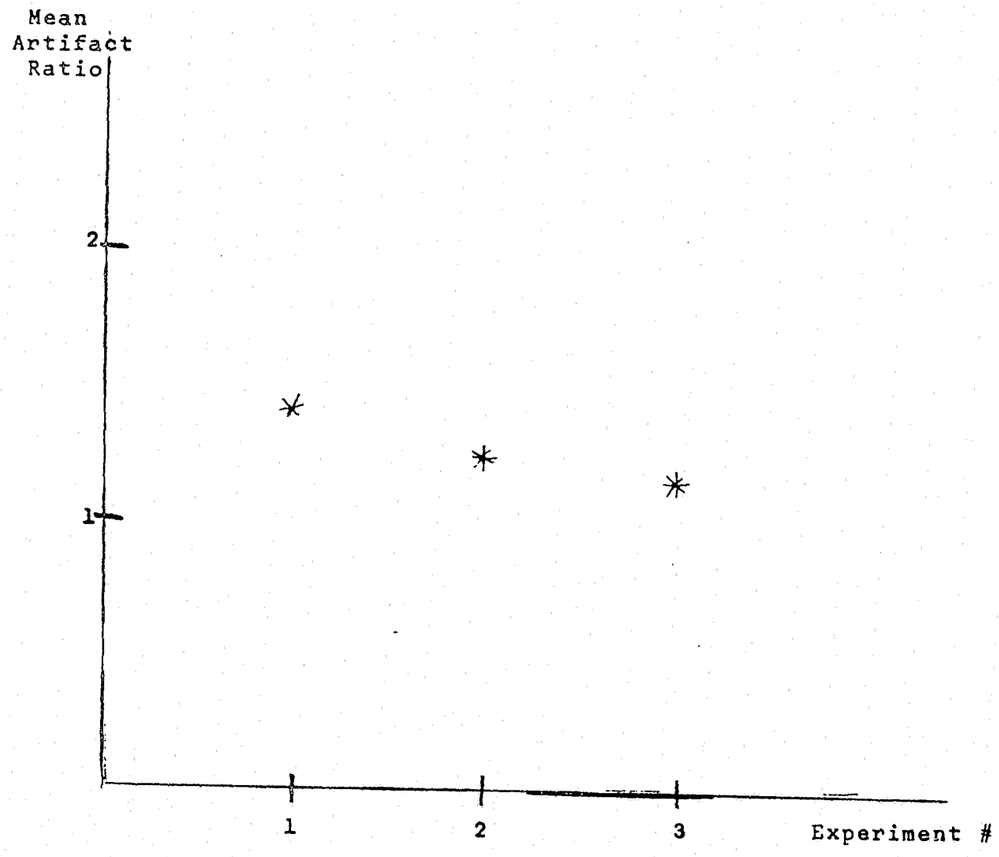
Finally, as presented in Table 11, the artifact ratio for each of the three experiments was never less than one. This indicates that those six Pauly blocks chosen for treatment were always above their normal crime totals. Since the crime behavior of the blocks in the sample is not significantly different, the tendency to select those blocks with a positive error - that is, above their normal crime rates - can be expected.

6.3 BASIC ASSUMPTIONS

This section presents some of the basic assumptions on which the study is based, and suggests means of validating them.

Assumption:

The model used to estimate the error terms, e(i,j), was assumed to be a linear, additive function of trend factors for each year and block factors for each Pauly



Experiment 1: e(i,j)
Experiment 2: 3/4 e(i,j)
Experiment 3: 1/2 e(i,j)

FIGURE 20. Plot of the Results of Each Experiment in Terms of the Mean Artifact Ratio.

block in the sample. To verify this assumption, it is necessary to measure how well any part of the model contributes to its ability to explain the observations. Tests of significance may be used for this kind of analysis, testing whether the parameter estimates of the model are significantly different from zero. If they are not, then the related factor is not a useful part of the model.

Assumption:

The use of uniform trend factors for all Pauly blocks assumes that the time trends in crime are uniform for all the blocks. The small number of observations for each Pauly block (i.e., five years of data) does not permit verification of the assumption at any reasonable level of significance. However, to determine significant differences, the "F to remove" statistic for the trend terms in the regression model indicates how well they contribute to explaining the variance. If trend varies widely from block to block, this statistic will be small and indicate low significance for trend factor estimates.

To minimize the variance of error in this model, the use of uniform trend factors for all Pauly blocks can be slightly modified to differentiate between those blocks with an upward trend and those with a downward trend in crime. This would involve plotting the crime for each of

the 40 blocks individually for the five years and then through a scanning process of these plots, separate the blocks into two groups, according to an increasing or decreasing trend in crime. The use of two time series models - each with a uniform trend factor for its group - may reduce some of the unexplained variation in the single model used for all Pauly blocks. Another possibility is to just screen out those blocks with decreasing crime rates from the sample of the 40 highest blocks, since their expected, or normal, crime behavior was at an all time low during the selection period. Under such circumstances it would be difficult for these blocks to compete for the experimental treatment with other blocks whose normal crime rates have, on the other hand, reached a peak during the selection period

Assumption:

When using data on the number of reported crimes, as opposed to the actual number of crimes for each Pauly block, two assumptions are being made. First, it is assumed that the reporting rate for each block does not change over time. Second, it is assumed that reporting rates do not differ among the blocks. Thus, changes in reporting rates as a function of time and geography could be a contributing factor in the erratic behavior of the

system of Pauly blocks. To test these assumptions, data on crime victimizations and crime reporting rates in each block for the years of interest are required. The high cost of obtaining this type of information has prevented its collection by criminal justice agencies.

7. CONCLUSIONS

Regression artifact is an unavoidable problem in the evaluation of experimental programs which are designed to select a sample of clients or areas from the population for treatment, based on performance during a given period prior to the experiment. Ignoring the artifact situation in a comparison of the "before and after" performance of the treatment group may inflate the effectiveness measure used to evaluate the program. In view of this problem, the objective of this project was to develop analytical techniques for estimating the magnitude of the artifact, and consequently, for determining a more accurate measure of the effectiveness of such programs. For simple cases, the use of order statistics provides a theoretical approach to the artifact problem. For more complex situations, a computer simulation was developed.

A computer program was constructed to simulate the process of selecting for treatment those elements in the population which exhibit the poorest performance during a specified period. To satisfy the specifications of the simulator, estimates of the expected performance during the selection period and the distribution of random performance variations for each element in the population are necessary. A time series model may be used to estimate

the above factors. For each run of the experiment, a measure of the artifact, based on the performance of the group selected for treatment, is computed. The output results of the experiment provide a probability distribution of the artifact measure, and its expected value, using the time series model, and a random number generator for the irregular fluctuations. In determining the reliability of these results, the order statistics techniques were used for validating the computer model for simple cases involving continuous probability distribution functions. A coin-tossing experiment was used for validation of the simulation model for discrete probability distribution functions.

Use of the analytic techniques to study an actual evaluation situation, the St. Louis Police Foot Patrol Project, illustrates by way of a specific example that the regression artifact is almost certainly responsible for some of the apparent crime reductions attributed to the project. Based on the simulation results for this project, it can be concluded that the mere comparison of a "before" measure to an "after" measure may not always be a true indication of the effectiveness of a program.

8. APPENDIX

APPENDIX 8.1

Names of Variables Used in Computer Program

Input Variables

XINT (1)	lower bound on first class interval of histogram
NINT	number of class intervals
WIDTH	width of each class interval
LINT	number of class limits (i.e., NINT + 1)
NRUNS	number of runs of each simulation experiment
M	number of highest Pauly blocks in experimental set
LB	total number of Pauly blocks in sample
LT	number of unit time intervals
CRIME (I, J)	number of crimes for i^{th} Pauly block, j^{th} time interval
P(I)	estimated mean crime for 1971 for Pauly block i

Variables Determined by Program

XINT(K)	lower class limit for k^{th} interval
F(K)	frequency within k^{th} interval of histogram
CPROB(I, K)	cumulative probability up to k^{th} interval for i^{th} Pauly block
X	uniform random number generated

APPENDIX 8.1
(continued)

RANDN(I)	random error term generated by use of cumulative distribution for i^{th} Pauly block; also represents generated crime rate for i^{th} Pauly block
SMALL	smallest of two random numbers being compared
SUMD2	sum of squares for series of numbers
AMEAN	average of series of numbers
VAR	variance of series of numbers
STDV	standard deviation of series of numbers
STAT(N)	artifact ratio for n^{th} run

9. BIBLIOGRAPHY

1. F. Galton, "Typical Laws of Heredity", Proceedings of the Royal Institute of Great Britain, 1879.

2. F. Galton, "Regression Toward Mediocrity in Hereditary Stature", Journal of the Anthropological Institute of Great Britain, 1886.

3. P. J. Rulon, "Problems of Regression", Harvard Educational Review, 1941.

4. Frederick M. Lord, "Measurement of Growth", Educational and Psychological Measurement, 16, 1956.

5. Frederick M. Lord, "Further Problems in the Measurement of Growth", Educational and Psychological Measurement, 18, 1958.

6. H. A. David, Order Statistics, John Wiley & Sons, Inc., New York, 1970.

7. Herbert Maisel, Simulation of Discrete Stochastic Structure, Science Researc Association, Chicago, 1972.

8. H. D. Brunk, Mathematical Statistics, Blaisdell Publishing Company, Waltham, Massachusetts, 1965.

9. Donald Campbell, Experimental and Quasi-Experimental Designs for Research, Rand McNally and Company, Chicago, 1963.

10. Donald Campbell and Albert Erlebacher, "How Regression Artifacts in Quasi-Experimental Evaluations Can Mistakenly Make Compensatory Education Look Harmful," in Compensatory Education: A National Debate (J. Hellmuth, ed.), Brunner/Mazel, New York, 1970.

11. Victor Cicirelli, "The Relevance of the Regression Artifact Problem to the Westinghouse - Ohio Evaluation of Head Start: A Reply to Campbell and Erlebacher", in Compensatory Education: A National Debate (J. Hellmuth, ed Brunner/Mazel, New York, 1970.

12. John Evans and Jeffry Schiller, "How Preoccupation With Possible Regressior Artifacts Can Lead to a Faulty Strategy For the Evaluation of Social Actic Programs: A Reply to Campbell and Erlebacher", in Compensatory Education: A National Debate (J. Hellmuth, ed.), Brunner/Mazel, New York, 1970.

13. Donald Compbell and Albert Erlebacher, "Reply to the Replies", in Compensa Education: A National Debate (J. Hellmuth, ed.), Brunner/Mazel, New York,

14. Denise Corcoran, Problem of Regression Artifact in the Evaluation of Exper mental Programs, Masters thesis, Department of Applied Mathematics and Com Science, Washington University, June 1974.