SOME APPLICATIONS OF STATISTICAL METHODS IN TRAFFIC ENGINEERING

by

JOHN CARL NELSON

B.S., Kansas State University, 1980

A MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Statistics

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1981

Approved by:

Doris Grosh

Major Professor

TABLE OF CONTENTS

ACKNOWLEDGMENT

I would like to thank Dr. Doris L. Grosh for her help during my years at Kansas State University, especially in preparation of this paper. Many thanks also to the other members of my advisory committee, Dr. G. A. Milliken and Dr. R. F. Nassar.

I. INTRODUCTION

The need for control of vehicular traffic in cities and on highways is not a new one. Ancient Rome had a severe problem with congestion which resulted in a ban on all nonpublic vehicles during daylight hours. London experienced traffic jams as early as the nineteenth century.(1) Expansion of highway systems in the United States has sometimes come too quickly for good control, resulting in poor safety.

The discipline which deals with these problems is known as traffic engineering. For centuries the only engineers involved in traffic were those involved in designing roads. Since the turn of the century, traffic engineering has become more than structural design. One definition of traffic engineering is

> It is the art and science of estimating traffic demand and highway capacity and measuring and determining the relationships between traffic variables--and the application of this knowledge to planning, designing, operating, and administering highway facilities in order to achieve the safe and efficient movement of persons and goods.(2,p.5)

The solutions to traffic problems require interactions between several disciplines in addition to engineering. Some of these are mathematics, physics, psychology, and statistics. This paper will deal with some of the contributions which are available from the field of statistics.

The early role of statistics in traffic engineering was mainly in the collection and summarization of data, along with some curve fitting and correlation study. Much of this early work was done at the Road Research Lab in England and by Greenshields and Norman in the United States. Later came hypothesis testing, modeling, and simulation.

Several variables associated with traffic were found to have well-known statistical distributions, making statistical methodologies more commonplace and necessary.(3) Some of these distributions are (under certain conditions),

Poisson - number of vehicles passing a point in the roadway in a given interval of time.

Exponential (or shifted exponential) - interarrival times of random vehicle arrivals.

Gamma (Erlang) - interarrival times of non-random vehicle arrivals.
- desired velocity, speed traveled when no other cars are present.

Normal - desired velocity (in some cases).
- interarrival times of vehicles within a moving queue.(4)

This paper will have two major sections. The first will take advantage of the Poisson/Exponential behavior of the random arrival of vehicles, and will borrow some results from queueing theory. The second section will show the results of three multiple regression analyses done on the safety effects of the 55-MPH speed limit in the United States.

II. QUEUEING THEORY

2.1 Stationary Queues

There are three basic queueing models which will be discussed: The Pure Birth Process, The Pure Death Process, and The Combined Birth and Death Process.

2.1.1 The Pure Birth Process

In many instances an analyst may be interested in the number of vehicles which have passed a certain point in the road within a given interval of time. This is analogous to the situation of vehicles arriving at a parking lot, but not leaving until after all vehicles have arrived. An example of this model is the parking lot at a factory. Workers arrive at the lot in the morning and arrivals accumulate until all workers have arrived, or until the lot is full. It is assumed that no workers leave (i.e. no departures) until the end of the working day. This is referred to as a Pure Birth Process in that the total number of vehicles in the lot increases or remains the same and never decreases (within a given interval of time). Other examples of this situation include parking lots for sporting events, concerts, and any other instance where there will temporarily be a great influx of traffic and little or no outflow.

In controlling this wave of arrivals, engineers must plan lane configurations and signal timing to efficiently accomodate all vehicles. It may be of interest to know the rate at which a given parking lot fills up, or equivalently, the probability of n cars in the lot at time t (n can be chosen to be the capacity of the lot, i.e. the probability of a full lot at time t). Plans can then be made to re-route traffic before congestion begins.

In many cases, the arrival of a vehicle at a parking lot can be considered a random event, not influenced by the activity of other vehicles. In this case, the number of vehicles in a lot at a given time will be a Poisson random variable, and the Pure Birth Process will be a Poisson Process. Several assumptions must be made in order to analyze the situation. These assumptions are the Poisson Postulates listed below.

<u>Notation</u>. Let

$N(t)$ = number of arrivals by time t, the "state" of the system.

$P_n(t) = Pr[N(t) = n]$ = probability of n arrivals by time t.

Note that $\{N(t + h) - N(t)\}$ = number of arrivals in a time interval of length h.

The Poisson Postulates

1) $N(0) = 0$. The counting of arrivals begins at time 0.

2) For any t, s, t', and s' with $t' > t + s$,

   $$\text{cov}\left\{\left[N(t + s) - N(t)\right], \left[N(t' + s') - N(t')\right]\right\} = 0.$$

   The process has independent increments, that is, the activity during one time interval does not affect the activity in another time interval.

3) For any $t > 0$, $0 < \left\{Pr\left[N(t) > 0\right]\right\} < 1$. In any time interval of any size, there is a positive probability that at least one arrival will happen and there is a positive probability that no arrivals will happen.

4) For any $t > 0$, $\lim_{h \to 0} \dfrac{\Pr[N(t + h) - N(t) \geq 2]}{\Pr[N(t + h) - N(t) = 1]} = 0.$

Two arrivals cannot occur simultaneously.

5) For any t, s, and t',

$\Pr\left[N(t + s) - N(t) = n\right] = \Pr\left[N(t' + s) - N(t') = n\right].$

The process has stationary increments, that is, the probability of n arrivals is constant for any two non-overlapping time intervals of the same length.

Note that if λ is the rate of arrivals, then

$$\Pr\left[N(t + h) - N(t) = 1\right] = \text{probability of one arrival in a time interval of length } h.$$
$$= \lambda h + o(h)$$

where $o(h)$ is a differential of higher order and can be neglected.(5)

Notation. Let

E_n denote the event that there are n arrivals by time t.

E_{n-1} denote the event that there are (n-1) arrivals by time t.

A_0 denote the event that there are no arrivals in a time interval of length h.

A_1 denote the event that there is one arrival in a time interval of length h.

Note that $\Pr(E_n) = P_n(t)$

$\Pr(E_{n-1}) = P_{n-1}(t)$

$\Pr(A_0) = \Pr\left[N(t + h) - N(t) = 0\right] = 1 - \lambda h$

$\Pr(A_1) = \Pr\left[N(t + h) - N(t) = 1\right] = \lambda h$

We wish to find an expression for $P_n(t)$, the probability of n arrivals after time t. In order to do this, it is necessary to solve two difference differential equations.

Note that from Postulate 2, E_i and A_j are independent for any i and j, so it follows by the definition of independence (6),

$$Pr(E_i \cap A_j) = \Big(Pr(E_i)\Big)\Big(Pr(A_j)\Big).$$

The difference-differential equations are obtained as follows:

$P_n(t + h)$ = probability of n arrivals by time (t+h).

= (probability of n arrivals by time t followed by no arrivals in interval h) + (probability of (n-1) arrivals by time t followed by one arrival in interval h.)

$$= Pr(E_n \cap A_0) + Pr(E_{n-1} \cap A_1).$$
$$= \{Pr(E_n)\}\{Pr(A_0)\} + \{Pr(E_{n-1})\}\{Pr(A_1)\}.$$
$$= \{P_n(t)\}\{1 - \lambda h\} + \{P_{n-1}(t)\}\{\lambda h\}$$
$$= P_n(t) - P_n(t)\{\lambda h\} + P_{n-1}(t)\{\lambda h\}. \qquad (2.1.1)$$

After transposing the first term on the right side and dividing through by h,

$$\frac{P_n(t + h) - P_n(t)}{h} = -\lambda\Big(P_n(t)\Big) + \lambda\Big(P_{n-1}(t)\Big)$$

and after taking the limit as $h \rightarrow 0$, the result is

$$\frac{d\,P_n(t)}{dt} = -\lambda\Big(P_n(t)\Big) + \lambda\Big(P_{n-1}(t)\Big). \qquad (2.1.2)$$

Now for the boundary condition we have

$P_0(t + h)$ = probability of no arrivals by time t followed by no arrivals in interval h.

$$= Pr(E_0 \cap A_0)$$
$$= \Big\{Pr(E_0)\Big\}\Big\{Pr(A_0)\Big\}$$
$$= \Big\{P_0(t)\Big\}\Big\{1 - \lambda h\Big\}$$
$$= P_0(t) - P_0(t)\{\lambda h\}. \qquad (2.1.3)$$

After transposing a term, dividing by h, and taking the limit as

$h \to 0$,

$$\frac{d\,P_0(t)}{dt} = -\lambda\Big(P_0(t)\Big). \tag{2.1.4}$$

In order to solve (2.1.2) and (2.1.4), one can invoke the Laplace Transform (7), resulting in the solution (see Appendix II):

$$P_0(t) = e^{-\lambda t} \tag{2.1.5}$$

$$P_n(t) = \frac{e^{-\lambda t}\ (\lambda t)^n}{n!} \tag{2.1.6}$$

which is the form of a Poisson random variable with parameter (λt). Hence this Pure Birth Process is a Poisson Process if the assumptions (which are the Poisson Postulates) are met. The probability of n arrivals by time t (for n = 0, 1, 2, $\cdots$) can then be evaluated by knowing only λ, the rate of arrivals, and using the above expression for $P_n(t)$.

2.1.2 Pure Death Process

A process complementary to the Pure Birth Process is the Pure Death Process. In the Pure Death Process there is a finite population whose members depart randomly (no two departures are simultaneous) until all are gone. The situation where this model is applicable is the converse of that for the Pure Birth Process. We are now concerned with a great outflow of traffic and no influx. This is the situation when the factory parking lot empties out at the end of the day, assuming there are no arrivals for a second shift. It is now of interest to obtain an expression to assess the probability of n vehicles remaining in the lot after time t.

The initial assumptions are changed somewhat:

1) $N(0) = N_0$.

2) Same as for Pure Birth Process.

3) For any $t > 0$, $0 < \left\{ Pr[N(t) < N_0] \right\} < 1$.

4) For any $t > 0$, $\lim_{h \to 0} \frac{Pr[N(t) - N(t + h) \geq 2]}{Pr[N(t) - N(t + h) = 1]} = 0$.

5) Same as for the Pure Birth Process.

Note that as λ was the rate of arrivals in the Pure Birth Process, μ is the rate of departures in the Pure Death Process. The probability of one departure in a time interval of length h (given that $n \geq 1$) is μh.

Notation. Let

E_n denote the event that there are n departures by time t.

E_{n+1} denote the event that there are (n+1) departures by time t.

D_0 denote the event that there are no departures in a time interval of length h.

D_1 denote the event that there is one departure in a time interval of length h.

Note that in the Pure Death Process, $P_n(t)$ denotes the probability of n remaining after time t. For both the Pure Birth and Pure Death Processes (and later in the Combined Birth and Death Process) $P_n(t)$ can be thought of as the probability of n in the system at time t.

Also, $Pr(E_n) = P_n(t)$

$Pr(E_{n+1}) = P_{n+1}(t)$

$Pr(D_0) = 1 - \mu h$

$Pr(D_1) = \mu h$.

Now $P_n(t + h)$ = probability of n departures by time (t+h).

= (probability of n departures by time t followed by no departures in interval h) + (probability of (n+1) departures by time t followed by one departure in interval h)

$$= \Pr(E_n \cap D_0) + \Pr(E_{n+1} \cap D_1)$$
$$= \left[\Pr(E_n)\right]\left[\Pr(D_0)\right] + \left\{\Pr(E_{n+1})\right\}\left[\Pr(D_1)\right]$$
$$= \left\{P_n(t)\right\}\left\{1 - \mu h\right\} + \left\{P_{n+1}(t)\right\}\left\{\mu h\right\}$$
$$= P_n(t) - P_n(t)\left\{\mu h\right\} + P_{n+1}(t)\left\{\mu h\right\}. \qquad (2.1.7)$$

After transposing the first term on the right, dividing by h, and taking the limit as $h \to 0$ (as done for the Pure Birth Process)

$$\frac{d\,P_n(t)}{dt} = -\mu\left(P_n(t)\right) + \mu\left(P_{n+1}(t)\right). \qquad (2.1.8)$$

For the boundary condition at $n = N_0$ (capacity), we have

$P_{N_0}(t + h)$ = probability of no departures by time t followed by no departures in interval h

$$= \Pr(E_{N_0} \cap D_0)$$
$$= \left\{\Pr(E_{N_0})\right\}\left[\Pr(D_0)\right]$$
$$= \left\{P_{N_0}(t)\right\}\left\{1 - \mu h\right\}$$
$$= P_{N_0}(t) - P_{N_0}(t)\left\{\mu h\right\}. \qquad (2.1.9)$$

After transposing and taking the limit as $h \to 0$,

$$\frac{d\,P_{N_0}(t)}{dt} = -\mu\left(P_{N_0}(t)\right). \qquad (2.1.10)$$

For the boundary condition at $n = 0$,

$P_0(t + h)$ = (probability of N_0 departures at time t followed by no departures in interval h) + (probability of (N_0-1) departures by time t followed by 1 departure in interval h)

$$= \Pr(E_0 \cap D_0) + \Pr(E_1 \cap D_1)$$
$$= \left[\Pr(E_0)\right]\left[\Pr(D_0)\right] + \left\{\Pr(E_1)\right\}\left\{\Pr(D_1)\right\}$$
$$= P_0(t)\left\{1\right\} + P_1(t)\left\{\mu h\right\}$$
$$= P_0(t) + P_1(t)\left\{\mu h\right\}. \qquad (2.1.11)$$

After transposing and taking the limit as $h \to 0$,

$$\frac{d\,P_0(t)}{dt} = \mu\left(P_1(t)\right). \tag{2.1.12}$$

Solutions for (2.1.8), (2.1.10), and (2.1.12) are obtained by using the Laplace Transform (see Appendix II):

$$P_n(t) = \frac{e^{-\mu t}\,(\mu t)^{N_0-n}}{(N_0-n)!} \qquad \text{for } n = 1, 2, 3, \cdots, N_0 \tag{2.1.13}$$

$$P_0(t) = 1 - \sum_{n=1}^{N_0} \frac{e^{-\mu t}\,(\mu t)^{N_0-n}}{(N_0-n)!} \tag{2.1.14}$$

Hence the Pure Death Process is a truncated Poisson Process.(5) The probability of n departures by time t (for n = 1, 2, 3, $\cdots$, N_0) can be evaluated by knowing μ, the rate of departures, and using the above expression for $P_n(t)$. The probability of an empty system is given by $P_0(t)$.

Note that if, in the Pure Birth Process,

$$P_n(t) = \frac{e^{-\lambda t}\,(\lambda t)^n}{n!}$$

n is substituted by (N_0-k), one obtains

$$P_k(t) = \frac{e^{-\lambda t}\,(\lambda t)^{N_0-k}}{(N_0-k)!}$$

which is the form of the Pure Death Process for $k \neq N_0$ with $\lambda = \mu$. Intuitively this should be true: an observer in the parking lot watches a Pure Death Process as vehicles depart at a rate μ. At the same time, another observer, by the roadway, watches a Pure Birth Process as the vehicles enter the flow of traffic at a rate λ. Both observers witness the same process, therefore $\lambda = \mu$.

2.1.3 Combined Birth and Death Process

The uses of the Pure Birth and Pure Death Processes are somewhat limited. Another model which finds more opportunities for application is the Combined Birth and Death Process. This is the case when a population experiences both arrivals and departures. An example of this process is a line of vehicles waiting to be served at an expressway toll booth. Vehicles arrive at the end of the line and depart once they have paid their toll. Other examples are entrance ramps to expressways and parking lots which are busy throughout the day. When vehicles arriving cannot immediately depart, they form a queue. The total population consists of vehicles being served, along with vehicles in queue. The same assumptions are made as were made for the Pure Birth and Pure Death Processes.

2.1.3a The Single Server Model

Note that from the assumptions, E_i, A_j, and D_k are jointly independent for all i, j, k. Hence

$$Pr(E_i \cap A_j \cap D_k) = \left\{Pr(E_i)\right\}\left\{Pr\ (A_j)\right\}\left\{Pr(D_k)\right\}.$$

Using the same notation as before and assuming there is one server:

$P_n(t + h)$ = probability of n in the system at time (t+h)

= [probability of n in the system at time t followed by no arrivals and no departures in interval h] +

[probability of (n-1) in the system at time t followed by one arrival and no departures in interval h] +

[probability of (n+1) in the system at time t followed by no arrivals and one departure in interval h] +

$$+ \Big[\text{probability of n in the system at time t followed by one arrival and one departure in interval h}\Big]$$

$$= Pr(E_n \cap A_0 \cap D_0) + Pr(E_{n-1} \cap A_1 \cap D_0) + Pr(E_{n+1} \cap A_0 \cap D_1) + Pr(E_n \cap A_1 \cap D_1)$$

$$= \{Pr(E_n)\}\{1 - \lambda h\}\{1 - \mu h\} + \{Pr(E_{n-1})\}\{\lambda h\}\{1 - \mu h\} + \{Pr(E_{n+1})\}\{1 - \lambda h\}\{\mu h\} + \{Pr(E_n)\}\{\lambda h\}\{\mu h\}$$

$$= P_n(t) - (\lambda + \mu)h\{P_n(t)\} + \lambda h\{P_{n-1}(t)\} + \mu h\{P_{n+1}(t)\} + h^2\lambda\mu\{P_n(t) - P_{n-1}(t) - P_{n+1}(t) + P_n(t)\}. \qquad (2.1.15)$$

After transposing terms and taking the limit as $h \to 0$,

$$\frac{d\,P_n(t)}{dt} = -(\lambda + \mu)\Big(P_n(t)\Big) + \lambda\Big(P_{n-1}(t)\Big) + \mu\Big(P_{n+1}(t)\Big). \qquad (2.1.16)$$

For the boundary condition when $n = 0$,

$$P_0(t + h) = \Big[\text{probability of none in the system at time t followed by no arrivals and no departures}\Big] + \Big[\text{probability of one in the system at time t followed by no arrivals and one departure}\Big] + \Big[\text{probability of none in the system at time t followed by one arrival and one departure}\Big]$$

$$= Pr(E_0 \cap A_0 \cap D_0) + Pr(E_1 \cap A_0 \cap D_1) + Pr(E_0 \cap A_1 \cap D_1)$$

$$= \{P_0(t)\}\{1 - \lambda h\}\{1\} + \{P_1(t)\}\{1 - \lambda h\}\{\mu h\} + \{P_0(t)\}\{\lambda h\}\{\mu h\}$$

$$= P_0(t) - \lambda h\{P_0(t)\} + \mu h\{P_1(t)\} - \mu\lambda h^2\{P_1(t)\} + \mu\lambda h^2\{P_0(t)\}. \qquad (2.1.17)$$

After transposing terms and taking the limit as $h \to 0$,

$$\frac{d\,P_0(t)}{dt} = -\lambda\Big(P_0(t)\Big) + \mu\Big(P_1(t)\Big). \qquad (2.1.18)$$

A solution for (2.1.16) assuming $N_0 = \infty$ is given by Saaty (8):

$$P_n(t) = e^{-[\lambda+\mu]t}\left[\left(\frac{\lambda}{\mu}\right)^{\frac{n}{2}} I_n[z] + \left(\frac{\lambda}{\mu}\right)^{\frac{n-1}{2}} I_{n+1}[z] + \left(1-\frac{\lambda}{\mu}\right)\left(\frac{\lambda}{\mu}\right)^{n} + \sum_{k=n+2}^{\infty}\left(\frac{\mu}{\lambda}\right)^{\frac{k}{2}} I_k[z]\right] \tag{2.1.19}$$

where $I_r[z] = \sum_{k=0}^{\infty} \frac{\left\{\frac{z}{2}\right\}^{r+2k}}{k!\ \Gamma(r+k+1)}$

is the modified Bessel function of the first kind with argument

$z = 2\sqrt{\lambda\mu}\,t.$

In many situations of the Combined Birth and Death Process, the process will reach a steady-state after a certain length of time. At this point the process acts much like the Wiener process which describes Brownian motion.(9,10) When the process reaches steady-state, the probabilities associated with each state of the system no longer depend upon time. Hence their derivative with respect to time vanishes. The resulting equations are (with the argument t deleted)

$$(\lambda+\mu)P_n = \lambda P_{n-1} + \mu P_{n+1} \tag{2.1.20}$$

$$\lambda P_0 = \mu P_1 . \tag{2.1.21}$$

Notation.

There are several quantities which will be of interest:

L = expected number in the system.

L_q = expected number in the queue.

W = expected waiting time in the system.

W_q = expected waiting time in the queue.

$\frac{\lambda}{\mu} = \rho$ = utilization rate.

The first case to be considered is one where there is one server and the system has infinite capacity ($N_0 = \infty$).

From (2.1.20) it can be shown that

$$P_{n+1} = \left(\frac{\lambda+\mu}{\mu}\right)P_n - \left(\frac{\lambda}{\mu}\right)P_{n-1} \tag{2.1.22}$$

and from (2.1.21)

$$P_1 = \left(\frac{\lambda}{\mu}\right)P_0 = \rho P_0 .$$

Letting n = 1 in the equation above

$$\begin{aligned} P_2 &= (\rho + 1)P_1 - \rho P_0 \\ &= (\rho + 1)\rho P_0 - \rho P_0 \\ &= \rho^2 P_0 + \rho P_0 - \rho P_0 = \rho^2 P_0 . \end{aligned}$$

Letting n = 2

$$\begin{aligned} P_3 &= (\rho + 1)P_2 - \rho P_1 \\ &= (\rho + 1)\rho^2 P_0 - \rho^2 P_0 \\ &= \rho^3 P_0 + \rho^2 P_0 - \rho^2 P_0 = \rho^3 P_0 \end{aligned}$$

and by mathematical induction

$$P_n = \rho^n P_0 . \tag{2.1.23}$$

Now $1 = P_0 + P_1 + P_2 + \cdots$

$$= P\,(1 + \rho + \rho^2 + \cdots)$$

$$= P_0\left(\frac{1}{1-\rho}\right) \qquad \text{(see Appendix I)}$$

which implies $P_0 = 1 - \rho$. Hence

$$P_n = \rho^n (1 - \rho). \tag{2.1.24}$$

Since N, the number in the system is a random variable, we can find its expected value by

$$\begin{aligned} L = E(N) &= \sum_{n=0}^{\infty} n(P_n) \\ &= \sum_{n=0}^{\infty} n\rho^n(1 - \rho) \\ &= \rho(1 - \rho)\sum_{n=0}^{\infty} n\rho^{n-1} \\ &= \rho(1 - \rho)(1 - \rho)^{-2} \qquad \text{(see Appendix I)} \end{aligned}$$

$$= \rho(1-\rho)^{-1} = \frac{\lambda}{\mu-\lambda}. \qquad (2.1.25)$$

With N being the number in the system, (N-1) will be the number in the queue, assuming there is one server, and its expected value is given by

$$L_q = E(N-1) = \sum_{n=1}^{\infty}(n-1)P_n \qquad N \geq 1$$
$$= \sum_{n=1}^{\infty}(n-1)\rho^n(1-\rho)$$
$$= \rho^2(1-\rho)\sum_{n=1}^{\infty}(n-1)\rho^{n-2}$$
$$= \rho^2(1-\rho)(1-\rho)^{-2} \qquad \text{(see Appendix I)}$$
$$= \rho^2(1-\rho)^{-1} = \frac{\lambda^2}{\mu(\mu-\lambda)} \qquad (2.1.26)$$

It has been proven (11,12,13) that the following relationships hold concerning L, L_q, W, and W_q;

$$L = \lambda W \qquad (2.1.27)$$

$$L_q = \lambda W_q. \qquad (2.1.28)$$

Hence $W = \frac{L}{\lambda} = \frac{\lambda}{\lambda(\mu-\lambda)} = \frac{1}{\mu-\lambda}$ (2.1.29)

and $W_q = \frac{L_q}{\lambda} = \frac{\lambda^2}{\lambda\mu(\mu-\lambda)} = \frac{\lambda}{\mu(\mu-\lambda)}$ (2.1.30)

Two other results which are often useful are

$$L - L_q = \frac{\rho}{1-\rho} - \frac{\rho^2}{1-\rho} = \frac{\rho(1-\rho)}{(1-\rho)} = \rho \qquad (2.1.31)$$

and $W - W_q = \frac{L}{\lambda} - \frac{L_q}{\lambda} = \frac{L - L_q}{\lambda} = \frac{\rho}{\lambda} = \frac{1}{\mu}.$ (2.1.32)

Hence values of L, L_q, W, and W_q can be easily found in the single server system with infinite queue-length capacity.

2.1.3b The Multiple Server Model

In the case where there is more than one server (i.e. more than one toll booth or more than one lane in an entrance ramp), the expressions for these quantities change somewhat. The departure rate μ is now dependent upon the state of the system. If there are two in the system and two servers are available, the departure rate is μ per server and 2μ for the system. If there are 3 available servers the departure rate for the system is 3μ when n = 3. In general, if s is the number of servers, then

$$\begin{aligned} \mu_n &= n\mu \qquad \text{for } n = 1, 2, \cdots, s \\ &= s\mu \qquad \text{for } n \geq s \end{aligned}$$

Under these conditions (2.1.20) and (2.1.21) become

$$(\lambda + \mu_n)P_n = \lambda P_{n-1} + \mu_{n+1} P_{n+1} \tag{2.1.33}$$

$$\lambda P_0 = \mu_1 P_1 \tag{2.1.34}$$

resulting in

$$P_{n+1} = \left(\frac{\lambda + \mu_n}{\mu_{n+1}}\right)P_n - \left(\frac{\lambda}{\mu_{n+1}}\right)P_{n-1} . \tag{2.1.35}$$

Hence

$$P_1 = \left(\frac{\lambda}{\mu_1}\right)P_0 = \left(\frac{\lambda}{\mu}\right)P_0 = \rho P_0$$

$$\begin{aligned} P_2 &= \left(\frac{\lambda + \mu_1}{\mu_2}\right)P_1 - \left(\frac{\lambda}{\mu_2}\right)P_0 \\ &= \left(\frac{\lambda + \mu}{2\mu}\right)\rho P_0 - \left(\frac{1}{2}\right)\rho P_0 \\ &= \left(\frac{1}{2}\right)(\rho + 1)\rho P_0 - \left(\frac{1}{2}\right)\rho P_0 \\ &= \left(\frac{1}{2}\right)\left(\rho^2 P_0 + \rho P_0 - \rho P_0\right) = \left(\frac{1}{2}\right)\rho^2 P_0 \end{aligned}$$

$$\begin{aligned} P_3 &= \left(\frac{\lambda + 2\mu}{3\mu}\right)P_2 - \left(\frac{\lambda}{3\mu}\right)P_1 \\ &= \left(\frac{1}{3}\right)(\rho + 2)\left(\frac{1}{2}\right)\rho^2 P_0 - \left(\frac{1}{3}\right)\rho^2 P_0 \\ &= \left(\frac{1}{3}\right)\left(\frac{1}{2}\right)\rho^3 P_0 + \left(\frac{1}{3}\right)\rho^2 P_0 - \left(\frac{1}{3}\right)\rho^2 P_0 = \left(\frac{1}{3!}\right)\rho^3 P_0 \end{aligned}$$

and by mathematical induction

$$P_n = \left(\frac{1}{n!}\right)\rho^n P_0 \qquad \text{for } n \leqq s \, . \qquad (2.1.36)$$

When $n \geqq s + 1$

$$\begin{aligned} P_{s+1} &= \left(\frac{\lambda + s\mu}{s\mu}\right)P_s - \left(\frac{\lambda}{s\mu}\right)P_{s-1} \\ &= \left(\frac{1}{s}\right)\rho P_s - \left(\frac{1}{s}\right)\rho P_{s-1} \\ &= \left(\frac{1}{s}\right)\left(\frac{1}{s!}\right)\rho^{s+1}P_0 \\ P_{s+2} &= \left(\frac{1}{s}\right)\left(\frac{1}{s!}\right)\rho^{s+2}P_0 \end{aligned}$$

and by mathematical induction

$$P_n = \left(\frac{1}{s^k}\right)\left(\frac{1}{s!}\right)\rho^{s+k}P_0 \qquad \text{for } n = s + k, \ k \geqq 0. \qquad (2.1.37)$$

The expression for P_0 is derived in a way similar to the way in which it was found for the single server case:

$$\begin{aligned} 1 &= P_0 + P_1 + P_2 + \cdots + P_s + P_{s+1} + \cdots + P_{s+k} + \cdots \\ &= P_0\left(1 + \rho + \frac{\rho^2}{2!} + \frac{\rho^3}{3!} + \cdots + \frac{\rho^s}{s!} + \frac{\rho^{s+1}}{(s)(s!)} + \cdots \right. \\ &\qquad \left. + \frac{\rho^{s+k}}{(s^k)(s!)} + \cdots\right) \\ &= P_0\left[\left(\sum_{n=0}^{s-1}\frac{\rho^n}{n!}\right) + \frac{\rho^s}{s!}\left(1 + \frac{\rho}{s} + \left[\frac{\rho}{s}\right]^2 + \left[\frac{\rho}{s}\right]^3 + \cdots\right)\right] \\ &= P_0\left[\left(\sum_{n=0}^{s-1}\frac{\rho^n}{n!}\right) + \frac{\rho^s}{s!}\left(\frac{1}{1 - \left[\frac{\rho}{s}\right]}\right)\right] \qquad \text{(see Appendix I)} \end{aligned}$$

Hence

$$P_0 = \left[\left(\sum_{n=0}^{s-1}\frac{\rho^n}{n!}\right) + \frac{\rho^s}{s!}\left(\frac{1}{1 - \left[\frac{\rho}{s}\right]}\right)\right]^{-1} \qquad (2.1.38)$$

The expressions for L and L_q are derived in the same way. L_q shall be derived first.

$$L_q = \sum_{n=s}^{\infty}(n - s)P_n \, .$$

After a change of variable, letting $k = n - s$,

$$L_q = \sum_{k=0}^{\infty} (k)(P_{s+k})$$
$$= \sum_{k=0}^{\infty} (k)\left(\frac{1}{s^k}\right)\left(\frac{1}{s!}\right)(\rho^{s+k})(P_0)$$
$$= \left(\frac{1}{s}\right)\left(\frac{1}{s!}\right)(\rho^{s+1})(P_0) \sum_{k=0}^{\infty} (k)\left(\frac{\rho^{k-1}}{s^{k-1}}\right)$$
$$= \frac{(P_0)(\rho^{s+1})}{(s)(s!)} \left(\frac{1}{1 - \left[\frac{\rho}{s}\right]}\right)^2$$
$$= \frac{(s)(P_0)(\rho^{s+1})}{(s!)(s - \rho)^2}. \tag{2.1.39}$$

The relationships between L, L_q, W, and W_q remain the same, resulting in

$$W_q = \frac{L_q}{\lambda} \tag{2.1.40}$$

$$W = W_q + \frac{1}{\mu} \tag{2.1.41}$$

$$L = \lambda W = \lambda\left(W_q + \frac{1}{\mu}\right) = L_q + \rho . \tag{2.1.42}$$

2.1.3c The Self-Service Model

The third case that will be considered is the situation when each member of the system has its own server. This is the case at a parking lot: each parking stall is a server and unless the lot is full there will be no queue. If we assume that the capacity of the lot is infinite (it will be shown later that this is not an outrageous assumption), the derivations for L and W are straight-forward. Since there is no queue, $L_q = W_q = 0$. The departure rate of the system is state dependent and takes on the form of the multiple server case with $n = \infty$:

$$\mu_n = n\mu \qquad \text{for all } n \geq 0.$$

From the multiple server model

$$P_n = \frac{\rho^n}{n!} P_0. \tag{2.1.43}$$

To find an expression for P_0, set $\sum_{n=0}^{\infty} P_n = 1$.

$$1 = P_0 + P_1 + P_2 + \cdots$$

$$= P_0 \left(1 + \rho + \frac{\rho^2}{2!} + \frac{\rho^3}{3!} + \cdots\right)$$

$$= P_0 (e^{\rho})$$ (see Appendix I)

Hence

$$P_0 = e^{-\rho} \tag{2.1.44}$$

$$P_n = \frac{\rho}{n!} e^{-\rho} \tag{2.1.45}$$

$$L = L_q + \rho = 0 + \rho = \rho \tag{2.1.46}$$

and

$$W = W_q + \frac{1}{\mu} = 0 + \frac{1}{\mu} = \frac{1}{\mu}. \tag{2.1.47}$$

If the capacity of the system is not infinite (which is almost always the case in applied problems),

$$P_0 = \sum_{n=0}^{N_0} \frac{\rho}{n!}$$ where N_0 is the capacity of the system.

For $\rho \leq 3$ and $N \geq 5$, conditions which cover almost all applicable situations, P_0 does not differ significantly from $e^{-\rho}$. Thus all of the above formulae for the infinite capacity system can be used without appreciable loss of precision.(14)

It should be noted that in the cases where arrival rates and departure rates are both state dependent, or even time dependent, values for L, L_q, W, and W_q can be found. Probabilities for the Pure Birth Process and Pure Death Process can also be found. The solution includes numerical solution of numerous iterative equations and probably does not find much applicability in traffic engineering.

Example of the uses of the queueing formulae.

Below is a list of simulated arrival times and departure times (in seconds) for ten trucks at a weigh station. This is a single-server Combined Birth and Death Process system.

Arrival	Interarrival Time	Departure	Interdeparture Time
24.7	24.7	1.1	1.1
27.2	2.5	2.3	1.2
28.2	1.0	16.7	14.4
44.3	16.1	35.9	19.2
50.9	6.6	41.0	5.1
56.1	5.2	42.6	1.6
66.6	10.5	50.5	7.9
70.9	4.3	53.1	2.6
72.3	1.4	56.8	3.7
80.7	8.4	66.9	10.1

Interarrival/interdeparture times are found by taking successive differences of the arrival/ departure times.

Sample statistics for the interarrival times are

$\bar{x}_A$ = mean interarrival time = 8.07 seconds/truck, and

s^2_A = variance of interarrival times = 55.24 (seconds/truck)^2.

The rate of arrivals, λ, is equal to the reciprocal of the mean interarrival time.

$$\lambda = \frac{1}{\bar{x}_A} = \frac{1}{8.07} = .1239 \text{ trucks/second} = 7.43 \text{ trucks/minute}$$

Sample statistics for the interdeparture times are

$\bar{x}_D$ = mean interdeparture time = 6.69 seconds/truck, and

s^2_D = variance of interdeparture times = 38.28 (seconds/truck)^2.

The rate of departures, μ, is equal to the reciprocal of the mean interdeparture time.

$$\mu = \frac{1}{\bar{x}_D} = \frac{1}{6.69} = .1494 \text{ trucks/second} = 8.96 \text{ trucks/minute}$$

Using equation (2.1.25), the expected number in the system is

$$L = \frac{\lambda}{\mu-\lambda} = \frac{7.43}{8.96 - 7.43} = 4.86 \text{ trucks}$$

Using equation (2.1.26), the expected number in the queue is

$$L_q = \frac{\lambda^2}{\mu(\mu-\lambda)} = \frac{(7.43)^2}{8.96(8.96 - 7.43)} = 4.03 \text{ trucks}$$

Using equation (2.1.29), the expected waiting time in the system is

$$W = \frac{1}{\mu-\lambda} = \frac{1}{8.96 - 7.43} = .65 \text{ minutes/truck} = 39.22 \text{ seconds/truck}$$

Using equation (2.1.30), the expected waiting time in the queue is

$$W_q = \frac{\lambda}{\mu(\mu-\lambda)} = \frac{7.43}{8.96(8.96 - 7.43)} = .54 \text{ minutes/truck} = 32.52 \text{ seconds/truck}$$

It should be noted that these data are simulated exponential distributions. The process above is assumed to be governed by the Poisson Postluates. If this assumption is not met, these formulae do not produce accurate results. In addition, when analyzing real data, a sample of size 10 would not, in most cases, be sufficiently large to produce precise and accurate results.

2.2 Computer Simulation

In many statistical analyses, especially those used in traffic engineering, large amounts of data are usually needed to insure results with a high level of precision. Computer simulation can help alleviate this problem. Based on assumptions concerning a random variable, one can have a computer generate "data" rather than having to collect it.

One such simulation package which can be used by traffic analysts has been developed by Thomas W. Rioux and Clyde E. Lee. Named "Texas", the package consists of these processors;

1) GEOPRO - geometry processor which analyzes the geometric configuration of a hypothetical or actual intersection. It points out areas of potential danger or congestion due to poor planning.

2) DVPRO - driver-vehicle processor which simulates arrivals of vehicles at an intersection. Inter arrival times can be chosen to be any one of the following random variables:

   a) uniform
   b) lognormal
   c) exponential
   d) shifted exponential
   e) gamma
   f) Erlang
   g) constant.

   Speed of each vehicle is simulated as a normal random variable where the user determines the mean and the 85th percentile. Also available is a processor named DISPRO which will compare a simulated data set to an actual data set by use of the Chi-square and/or Kolmogorov-Smirnoff tests for goodness-of-fit.

3) SIMPRO - simulation processor which passes each simulated driver-vehicle unit created by DVPRO through the intersection. With results obtained from GEOPRO, pertinent data are gathered and desired statistics are calculated.

Simulation, as done by packages like the Texas model, can be used in analyzing both present facilities and proposed plans. It is beneficial in evaluating roadway geometry, signal timing, and other traffic engineering problems.(15)

With any computer simulation, it is important for the user to realize that generated data is only as good as the validity of the model. Many times it may be necessary to make unrealistic assumptions when using simulation. This is often the case when using the queueing models previously mentioned. Simulation cannot replace real data, but it can complement it in many situations.

III. REGRESSION ANALYSIS

3.1 Linear Regression

The statistical method of linear regression deals with constructing a linear function of a set of variables which will produce an estimate of the mean of a random variable.

Notation. Let

Y = any observable random variable (its mean to be estimated)

$x_1, x_2, \cdots, x_p$ be p observable variables (used to estimate $\mu_{Y|\underline{x}}$)

ϵ = random error associated with the random variable Y.

Y is termed the response variable and the x_i's ($i = 1, 2, \cdots, p$) are termed predictor variables. $\mu_{Y|\underline{x}} = E(Y \mid x_1, x_2, \cdots, x_p)$

The general linear regression model is

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \epsilon \qquad (3.1.1)$$

where the β_i's ($i = 0, 1, 2, \cdots, p$) are the parameters of the model. β_0 is the intercept constant and β_i ($i = 1, 2, \cdots, p$) is the value of the change in the variable Y due to a unit change in the variable x_i ($i = 1, 2, \cdots, p$).

Note that regression analysis deals with the distribution of the response variable Y conditional upon known values of the predictor variables $x_1, x_2, \cdots, x_p$. One assumption made when regression analysis is used is that the random variables ϵ are independent and identically distributed normal random variables with mean equal to 0 and variance equal to σ^2 for all values of Y.

The simple linear regression model is a special case when

$p = 1$, resulting in

$$Y = \beta_0 + \beta x + \epsilon \tag{3.1.2}$$

A set of linear regression models takes the form

$$Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_p x_{pi} + \epsilon_i \tag{3.1.3}$$

which can be easily put into matrix form, resulting in

$$\underline{Y} = X\underline{\beta} + \underline{\epsilon} \tag{3.1.4}$$

where

$$\underline{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \cdot \\ \cdot \\ \cdot \\ Y_n \end{bmatrix}, \qquad X = \begin{bmatrix} 1 & x_{11} & x_{21} & \cdots & x_{p1} \\ 1 & x_{12} & x_{22} & \cdots & x_{p2} \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ 1 & x_{1n} & x_{2n} & \cdots & x_{pn} \end{bmatrix},$$

$$= \begin{bmatrix} \beta_1 \\ \beta_2 \\ \cdot \\ \cdot \\ \cdot \\ \beta_p \end{bmatrix}, \qquad = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \cdot \\ \cdot \\ \cdot \\ \epsilon_n \end{bmatrix}.$$

The values of the β_i's $(i = 0, 1, 2, \cdots, p)$ are usually not known and must be estimated.

<u>Notation</u>. Let

b_i $(i = 0, 1, 2, \cdots, p)$ = set of estimates for the set of parameters β_i $(i = 0, 1, 2, \cdots, p)$.

Note that the b_i's $(i = 0, 1, 2, \cdots, p)$ are estimates based on a set of values of the response variable. The estimates are, therefore, themselves random variables, and the variance of $\hat{\mu}_{Y|\underline{x}}$ is a function of their variances.

The resulting estimated regression line is

$$\hat{\mu}_{Y|\underline{x}_k} = b_0 + b_1 x_{1k} + b_2 x_{2k} + \cdots + b_p x_{pk}, \tag{3.1.5}$$

the matrix form of which is

$$\hat{\mu}_{Y|\underline{x}_k} = \underline{x}_k' \, \underline{b} \tag{3.1.6}$$

where $\underline{x}_k$ is a vector of observed values of the predictor variables, $\underline{b}$ is a vector of estimates of the parameters, and $\hat{y}_k$ is the resulting estimate of the mean response. (15) Two methods of parameter estimation will be taken up later.

Note that $\hat{\mu}_{Y|\underline{x}} = \hat{E}(Y \mid x_1, x_2, \cdots, x_p)$ and is a function of the random variables b_i $(i = 0, 1, 2, \cdots, p)$ and hence is itself a random variable.

In order to assess the predictive value of the model, the method of Analysis of Variance can be used. The total variability of the response value can be decomposed into two components: variability due to the regression model and variability due to random error.

Notation. Let

$$\text{SSTo} = \text{Sum of Squares Total} = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

$$= \text{measure of the total variability of Y.}$$

$$\text{SSR} = \text{Sum of Squares Regression} = \sum_{i=1}^{n} (\hat{\mu}_i - \bar{y})^2$$

$$= \text{measure of the variability of } \hat{\mu} \text{ about } \bar{y}.$$

$$\text{SSE} = \text{Sum of Squares Error} = \sum_{i=1}^{n} (y_i - \hat{\mu}_i)^2$$

$$= \text{measure of the variability of Y about } \hat{\mu}.$$

It can be proved (16) that SSTo = SSR + SSE, and an unbiased estimator of $\sigma^2(\epsilon_i)$ is $\dfrac{\text{SSE}}{n - p - 1}$.

Recall that $\sigma^2(\epsilon_i)$ is assumed to be a constant for all $i = 1, 2, \cdots, n$.

Definition. The coefficient of determination, R^2, is defined as (17)

$$R^2 = \frac{\text{SSR}}{\text{SSTo}} = 1 - \frac{\text{SSE}}{\text{SSTo}}. \tag{3.1.7}$$

R^2 is a measure of how well the regression model predicts the random variable Y. It is the "percent reduction" in the variability of the response variable when all predictor variables are considered. The matrix form of the coefficient of determination is

$$R^2 = \frac{\underline{b}'X'\underline{y} - n\,\bar{y}^2}{\underline{y}'\underline{y} - n\,\bar{y}^2} \tag{3.1.8}$$

where X, $\underline{y}$, and $\bar{y}$ are composed of observed values. Various methods are available for attempting to estimate the percent reduction in Y due to each x_i $(i = 1, 2, \cdots, p)$.(17)

3.1.1 Least Squares Estimation

The most widely used method of estimating the parameters is the Method of Least Squares. In this method, the values of the estimates of the regression function (3.1.5) are chosen in such a way as to minimize the sum of squares of the differences between each response observation and its estimate. The values of the b_i's $(i = 0, 1, 2, \cdots, p)$ are found such that

$$Q = \sum_{i=1}^{n} (y_i - b_0 - b_1x_{1i} - b_2x_{2i} - \cdots - b_px_{pi})^2 \tag{3.1.9}$$

is a minimum. To do this, the partial derivatives of Q with respect to each b_i $(i = 0, 1, 2, \cdots, p)$ are found and set equal to 0. Then the system of (p+1) equations

$$\frac{\partial Q}{\partial b_i} = 0 \tag{3.1.10}$$

is solved simultaneously, resulting in the set of estimates b_i $(i = 0, 1, 2, \cdots, p)$.

In matrix form, the vector of estimated parameters $\underline{b}$ has the form

$$\underline{b} = (X'X)^{-1}\ (X'\underline{y}) \tag{3.1.11}$$

where X is a matrix of the values of the predictor variables, that is, $X = [x_{ij}]$. Hence another form for the regression function (3.1.6) is

$$\hat{\mu}_{Y|\underline{x}_k} = \underline{x}'_k \ (X'X)^{-1} \ (X'\underline{y}). \qquad (3.1.12)$$

Note that $\sigma^2(X'X)^{-1}$ is the variance-covariance matrix of the set of random variables b_i ($i = o, 1, 2, \cdots, p$), and the resulting variance of the estimate is

$$\text{var}(\hat{\mu}_{Y|\underline{x}_k}) = \left[\underline{x}'_k \ (X'X)^{-1} \ \underline{x}_k\right]\hat{\sigma}^2 \qquad \text{where } \hat{\sigma}^2 \text{ is } \frac{SSE}{n - p - 1}.(19)$$

3.1.2 Stagewise Regression

Stage or Stagewise Regression is a modification of the Method of Least Squares. Although easier to compute, the resulting parameter estimates are not true least squares estimates.

The procedure involves constructing a simple linear model (3.1.2) using only the predictor variable that is most highly correlated with Y. The residuals ($\hat{\mu}_i - y_i$) of this model are calculated and are used to construct another simple linear model with the remaining predictor variable that is most highly correlated with the residuals. The two simple linear models are added and the residuals are calculated. The process then continues until all predictor variables are included.

Stagewise regression has limited applicability and usually has poorer predictive qualities than least squares regression. A comparison of these two methods follows.(16)

3.1.3 Numerical Comparison of Least Squares and Stagewise Estimation.

A linear regression model is to be built from a data matrix X and a response vector $\underline{y}$. The numerical values for this example are as follows:(16,20,21)

$$X = \begin{array}{c} \begin{array}{ccc} & x_1 & x_4 \end{array} \\ \begin{bmatrix} 1 & 7 & 60 \\ 1 & 1 & 52 \\ 1 & 11 & 20 \\ 1 & 11 & 47 \\ 1 & 7 & 33 \\ 1 & 11 & 22 \\ 1 & 3 & 6 \\ 1 & 1 & 44 \\ 1 & 2 & 22 \\ 1 & 21 & 26 \\ 1 & 1 & 34 \\ 1 & 11 & 12 \\ 1 & 10 & 12 \end{bmatrix} \end{array} \qquad \underline{y} = \begin{bmatrix} 78.5 \\ 74.3 \\ 104.3 \\ 87.6 \\ 95.9 \\ 109.2 \\ 102.7 \\ 72.5 \\ 93.1 \\ 115.9 \\ 83.8 \\ 113.3 \\ 109.4 \end{bmatrix}$$

x_2 and x_3 did not have significant predictive qualities and were omitted after an initial analysis.

From (3.1.11)

$$\underline{b} = (X'X)^{-1}(X'\underline{y}) = \begin{bmatrix} .57627 & -.07412 & -.01064 \\ -.02412 & .00251 & .00018 \\ -.01064 & .00018 & .00031 \end{bmatrix} \begin{bmatrix} 1240.5 \\ 10032 \\ 34733.3 \end{bmatrix}$$

$$= \begin{bmatrix} 103.10 \\ 1.44 \\ -.61 \end{bmatrix}$$

Hence the least squares regression equation for these data is

$$\hat{\mu}_{LS} = 103.10 + 1.44x_1 - .61x_4 \qquad (3.1.13)$$

To construct the stagewise regression equation, a simple linear regression model is formed between Y and x_4 (x_4 is the predictor variable most correlated with Y since $r_{yx_4} = -.821$ and $r_{yx_1} = .731$).

Thus

$$\underline{b}_1 = (X_4' X_4)^{-1} (X_4' \underline{y}) = \begin{bmatrix} .3446209 & -.0089233 \\ -.0089233 & .0082974 \end{bmatrix} \begin{bmatrix} 1240.5 \\ 34733.3 \end{bmatrix}$$

$$= \begin{bmatrix} 117.56657 \\ -.7396702 \end{bmatrix}$$

$$\text{where } X_4 = \begin{array}{c} \quad x_4 \\ \begin{bmatrix} 1 & 60 \\ 1 & 52 \\ 1 & 20 \\ 1 & 47 \\ 1 & 33 \\ 1 & 22 \\ 1 & 6 \\ 1 & 44 \\ 1 & 22 \\ 1 & 26 \\ 1 & 34 \\ 1 & 12 \\ 1 & 12 \end{bmatrix} \end{array} \qquad \text{and } \underline{y} = \begin{array}{c} Y \\ \begin{bmatrix} 78.5 \\ 74.3 \\ 104.3 \\ 87.6 \\ 95.9 \\ 109.2 \\ 102.7 \\ 72.5 \\ 93.1 \\ 115.9 \\ 83.8 \\ 113.3 \\ 109.4 \end{bmatrix} \end{array}$$

and the initial regression equation is

$$\hat{Y}_1 = 117.57 - .74x_4. \tag{3.1.14}$$

The residuals ($Z = \hat{Y}_1 - Y$) are computed for this model and another simple linear regression model is constructed using ($\hat{Y}_1 - Y$) as the response variable and x_1 as the predictor variable. If there were more than one predictor variable available at this point, the variable most highly correlated with the residuals would be chosen.

The vector of estimated parameters for this model is

$$\underline{b}_2 = (X_1' X_1)^{-1} (X_1' \underline{z}) = \begin{bmatrix} .2110041 & -.0179696 \\ -.0179696 & .0024083 \end{bmatrix} \begin{bmatrix} .01 \\ 561.88 \end{bmatrix}$$

$$= \begin{bmatrix} -10.094649 \\ 1.352995 \end{bmatrix}$$

$$\text{where } X_1 = \begin{array}{c} \quad x_1 \\ \begin{bmatrix} 1 & 7 \\ 1 & 1 \\ 1 & 11 \\ 1 & 11 \\ 1 & 7 \\ 1 & 11 \\ 1 & 3 \\ 1 & 1 \\ 1 & 2 \\ 1 & 21 \\ 1 & 1 \\ 1 & 11 \\ 1 & 10 \end{bmatrix} \end{array} \qquad \text{and } \underline{z} = \begin{array}{c} \hat{Y}_1 - Y \\ \begin{bmatrix} 5.22 \\ -4.88 \\ 1.50 \\ 4.73 \\ 2.69 \\ 7.87 \\ -10.44 \\ -12.59 \\ -8.22 \\ 17.52 \\ -8.67 \\ 4.59 \\ .69 \end{bmatrix} \end{array}$$

and the resulting regression equation is

$$\hat{Z} = -10.10 + 1.35x_1 \tag{3.1.15}$$

Since all significant predictor variables have been used, the final stagewise regression equation is found by adding all simple linear regression equations:

$$\hat{\mu}_{SW} = \hat{Y}_1 + \hat{Z} = 117.57 - .74x_4 + (-10.10 + 1.35x_1)$$

$$= 107.47 - .74x_4 + 1.35x_1 \tag{3.1.16}$$

Note that if there are p predictor variables

$$\hat{\mu}_{SW} = \hat{Y}_1 + \hat{Z}_1 + \hat{Z}_2 + \cdots + \hat{Z}_{p-1}$$

where $\hat{Z}_i$ is the estimated regression line using

$$(\hat{Y}_1 + \hat{Z}_1 + \hat{Z}_2 + \cdots + \hat{Z}_{i-1} - Y)$$

as a response variable and using the predictor variable most highly correlated with this response variable.

To summarize, the least squares solution regression equation is (3.1.13):

$$\hat{\mu}_{LS} = 103.10 - .61x_4 + 1.44x_1$$

and the stagewise regression equation for the same data is (3.1.16):

$$\hat{\mu}_{SW} = 107.47 - .74x_4 + 1.35x_1 .$$

The resulting difference in the two estimates is dependent upon the values of x_1 and x_4.

$$(\hat{\mu}_{LS} - \hat{\mu}_{SW}) = -4.37 + .13x_4 + .09x_1$$

Deviations of the stagewise estimate from the least square estimate, in this example, are of the order

$4.24\% \left\{\dfrac{4.37}{103.10}\right\}$ with regard to the contribution of the intercept,

$-21.3\% \left\{\dfrac{-.13}{.61}\right\}$ with regard to the contribution of x_4,

$$-6.75\% \quad \left\{\frac{-.09}{1.44}\right\}$$ with regard to the contribution of x_1.

3.2 The Penn State Study

Many opportunities arise within the discipline of traffic engineering in which a researcher can apply multiple regression analyses. Three such studies, which will be reported here, have attempted to evaluate the effects of impositions of the 55-MPH speed limit in the United States. Although causation cannot be inferred from a regression or correlation study, the research which will be presented here hopes to show some relationship between the 55-MPH speed limit and accident rates.

The first study to be reviewed is one by R. F. Heckard, J. A. Pachuta, and F. A. Haight at the Pennsylvania State University,(22) which was reported and further analyzed in an article by W. J. Kemper and S. R. Byington.(23) The study uses a time series regression model(17) with one predictor variable. Accident rates, injury rates, and fatality rates are regressed against time. According to the estimated regression line constructed by using the data from 1968-1973, an estimated fatality rate for 1974 is 3.94.* The actual fatality rate for 1974 was 3.57, 0.37 less than the prediction. At this point the study concludes that this result "proves" that the reduction in fatalities was caused by the enactment of the new speed limit.(23)

Apart from the fact that statistical relationships do not prove that anything causes anything, the analysis has one obvious shortcoming: the author assumes that the 1974 fatality rate value of 3.57 is significantly different from the predicted value of 3.94. Upon construction of a confidence interval for the

*3.94 fatalities/ 100 million vehicle-miles predicted value, it can be shown that the difference is significant with $\alpha = .01$, but no tests of significance or confidence intervals are reported for the other results.

The additional analyses done by Kemper et al. deal with some regression and correlation studies done on speed data and fatal accident data taken over the years 1970-1975. Four regression models were constructed:

1) Fatality rate regressed on mean travel speed (n = 6)
2) Fatal accident rate regressed on mean travel speed (n = 6)
3) One year changes in fatality rate regressed on one year changes in mean travel speed (n = 5)
4) All possible time changes in fatality rate regressed on all possible time changes in mean travel speed. (n = 15)

The procedure of taking successive differences of the independent variable data (procedure used in 3) is a common one used to remove the effect of auto-correlated error terms.(17) However, no test of auto-correlation was reported.

The results of 1, 2, and 3 were not reported because relatively small sample size yielded a variance in the estimated value so large that confidence intervals were of little use. The sample size in model 4 produced smaller confidence intervals, but the data seems to be somewhat "manufactured". The procedure of taking all possible differences within a data set is rather non-standard, resulting in a smaller mean square error (due to increase in sample size), but also showing a reduction in R^2. The results show a significant drop in the fatality rate, but the accuracy of the model is questionable.

In summary, the Penn State Study and the additional analyses of it done by Kemper et al. have several shortcomings. In light of the data and procedures that were reported,(23) the conclusion is questionable:

> The 55-MPH (89 km/h) speed limit - aided by driver education, increased enforcement, and drivers' desire to save fuel - probably caused most of the observed speed reductions during 1974 which saved a minimun of 4700 lives and 81000 injuries.(23, p.67)

3.3 The Michigan Study

Another study was performed in 1974 by N. Enüstün and A. H. Yang(24) for the Michigan Department of State Hgihways and Transportation. Monthly data was collected from January 1972 to June 1974 and four multiple regression models were constructed. The four dependent variables were:

a) Total freeway accidents
b) Total conventional road accidents
c) Fatal freeway accidents
d) Fatal conventional road accidents.

Each of these variables were regressed against seven predictor variables.

1) Total vehicle-miles
2) Average speed
3) Percentage of employment in population
4) Liquor sales
5) Sales tax
6) Motor fuel use
7) Car sales

The results of the regressions were somewhat mixed. All variables were kept in the model regardless of the significance of each estimated regression coefficient. The best resulting model was one used to predict fatal accidents on conventional roads. It had a coefficient of determination (R^2) of .773. The precision of this model was increased by .381 when average speed was included as a predictor in the presence of all other predictors (measured by the square of the partial correlation coefficient, r). The least effective model was the one for fatal accidents on freeway, having an R^2 of only .101. With the exception of the model for total accidents on conventional roads, the inclusion of average speed as a predictor resulted in a greater gain in precision when included than any other of the predictor variables. In the one exception, liquor sales had a higher r^2 and average speed was second.

The overall result of the Michigan Study was that average speed was a reasonably good predictor of accident rates. This result seem well-founded. Enüstün and Yang are quick to point out, however, that the association of two random variables does not imply causation. Another important point which is brought out in the article is that although average speed may be a good predictor, a change in average speed is not necessarily related to a change in speed limit. There are other factors involved and the effect on accident rates due to the 55-MPH speed limit cannot be determined with much precision.

3.4 The Maryland Study

The final regression study which will be examined is one done by H. S. Dawson, a research consultant.(25) Data were taken over numerous sections of a roadway in Maryland from 1970-1976. The only dependent variables used were total fatal accidents and fatal accident rate. Fifteen predictor variables were introduced into the initial multiple regression model:

1) The "time effect"
2) Speed limit
3) Average speed
4) Standard deviation in speed
5) Percent of vehicles traveling 16 km/h (10 mph) or less
6) Percent of vehicles traveling at the speed limit or less

7) Total traffic volume
8) Percent of volume traveling on Interstate roadways
9) Percent of volume traveling on weekends
10) Percent of volume traveling at night
11) Total number of speeding citations written
12) Percent of speeding citations written on weekends
13) Percent of speeding citations written at night
14) Daylight savings time or not
15) Energy crisis or not.

The "time effect" refers to the annual decline in fatal accident rates which has been occurring for many years due to higher technologies in the safety of vehicles and roadways.

Stage regression was used in construction of the predictor equation. Stage regression involves regressing the residuals of a given model against another predictor variable in a simple linear model and then combining all resulting simple linear models. This procedure does not provide the least square solutions for the regression coefficients, and its applicability is somewhat limited.(16)

Only four of the fifteen predictive variables were found to have significant contribution at the .05 level:

a) The "time effect"
b) Speed limit
c) Total traffic volume
d) Total number of speeding citations.

These four variables were fitted into the model, which produced an R^2 of .6756 for the fatal accident analysis and .7543 for the fatal accident rate analysis.

The data were tested for six properties of ill-conditioning:

1) Normality
2) Randomness
3) Lack of constant variance
4) Auto correlation
5) Multicollinearity
6) Spurious correlation.

No significance was detected except in the area of multicollinearity. This problem arises when some or all of the other predictor variables are highly correlated among themselves. The aptness of the model is not affected in the presence of multicollinearity, but the resulting variance of the predicted value is often very large, making precise inferences impossible.(26) Another problem which arises is the difficulty encountered when trying to assess the percent of explained variance due to a single predictor variable.

In trying to overcome the second problem, Dawson used two methods. First, by calculating the square of the partial correlation for each of the variables, he was able to get a rough hierarchy for the four variables. This is a commonly advised procedure. The second procedure used was on involving "commonality analysis". Here the explained variance is divided into (2^n-1) components, one for each of the n variables and (2^n-n-1) "common components".(27) A generalization of the process would be to say that the percent of variance explained by x_i is somewhere between R^2 for the simple linear model of y or x_i and r^2 (square of partial correlation coefficient of x_i) for the complete model. The width of the interval will be a function of the degree of multicollinearity of the x_i's. ($R^2 = r^2$ when all x_i's are jointly independent.)

Using this technique, Dawson states that the "time effect" accounted for 33 to 35 percent of explained variance. The lowering of the speed limit accounted for about 32 percent. The increase in law enforcement accounted for 14 to 17 percent, and the changes in traffic volume accounted for 10 to 15 percent of the explained variance.

This study, much more in depth than the two preceding, concludes that there is a relationship to the decline in fatality rate since the imposition of the 55-MPH speed limit. It should be noted, however, that the analysis was done on data from Maryland only and the results of the study cannot be necessarily assumed for the entire country.

With the availability of computers with statistical software, stagewise regression is an inappropriate technique. Several other methods, including forward selection and backward elimination procedures are available. These and other methods produce a model whose parameter estimates are unbiased least squares estimates.(16)

IV. CONCLUSION

This paper has tried to show a few instances where the field of applied statistics can contribute to solving some of the problems encountered in traffic engineering. The areas of queueing theory and regression analysis have much to offer the engineers, much more than is presented here. It is hoped that traffic engineering, along with all other scientific disciplines, may take advantage of these and the many other areas of statistics.

APPENDIX I

Some Mathematical Identities

$$\sum_{k=0}^{\infty} z^k = 1 + z + z^2 + z^3 + \cdots$$

$$= \frac{(1 - z)(1 + z + z^2 + z^3 + \cdots)}{1 - z}$$

$$= \frac{1 - z + z - z^2 + z^2 - z^3 + z^3 + \cdots}{1 - z}$$

$$= (1 - z)^{-1} \qquad |z| < 1$$

$$\sum_{k=0}^{\infty} kz^{k-1} = \sum_{k=0}^{\infty} \frac{d}{dz} z^k$$

$$= \frac{d}{dz}\left[\sum_{k=0}^{\infty} z^k\right]$$

$$= \frac{d}{dz}\left[(1 - z)^{-1}\right]$$

$$= (1 - z)^{-2} \qquad |z| < 1$$

$$\sum_{k=0}^{\infty} \frac{z}{k!} = 1 + z + \frac{z^2}{2!} + \frac{z^3}{3!} + \cdots$$

$$= e^z$$

APPENDIX II

Laplace Transforms

If, for a function $f(t)$ defined for $t > 0$, there exists a real number p, such that for p and all values greater than p, the integral $F(p)$ converges, then $F(p)$ is the Laplace Transform of $f(t)$, where

$$F(p) = \int_0^\infty e^{-pt} \; f(t)\, dt.$$

Notation.

$$F(p) = T\big[f(t)\big].$$

Solution to the Difference-Differential Equations of the Poisson Process

A. Pure Birth Process

$$\frac{d\, P_0(t)}{dt} = -\lambda\Big[P_0(t)\Big] \qquad (2.1.4)$$

$$\frac{d\, P_n(t)}{dt} = -\lambda\Big[P_n(t)\Big] + \lambda\Big[P_{n-1}\,(t)\Big] \qquad (2.1.2)$$

Let $Y_n = T\Big[P_n(t)\Big]$.

To solve (2.1.4) take the Laplace Transform of both sides.

$$\int_0^\infty e^{-pt} \left[\frac{d\, P_0(t)}{dt}\right] dt = \int_0^\infty e^{-pt} \Big[-\lambda\Big]\Big[P_0(t)\Big] dt$$

After factoring λ out of the integral on the right side, and integrating by parts on the left with

$$u = e^{-pt} \qquad\qquad du = -pe^{-pt}$$

$$dv = d\; P_0(t) \qquad\qquad v = P_0(t)$$

one obtains

$$\left[e^{-pt}\right]\left[P_0(t)\right]\Big|_0^{\infty} + p\int_0^{\infty} e^{-pt}\left[P_0(t)\right]dt = -\lambda\int_0^{\infty} e^{-pt}\left[P_0(t)\right]dt$$

which results in

$$pY_0 - P_0(0) = -(\lambda)(Y_0).$$

Solving for Y_0

$$Y_0 = \frac{1}{p+\lambda}$$

and taking the inverse transform of both sides

$$P_0(t) = e^{-\lambda t} .$$

Similarly, to solve (2.1.2), after taking the transform of both sides, one obtains

$$pY_n - P_n(0) = -(\lambda)(Y_n) + (\lambda)(Y_{n-1}).$$

Solving for Y_n, using $Y_0 = \dfrac{1}{p+\lambda}$ and mathematical induction,

$$Y_n = \frac{P_n(0) + (Y_{n-1})}{p+\lambda} = \frac{\lambda^n}{(p+\lambda)^{n+1}}$$

and taking the inverse transform of both sides

$$P_n(t) = \frac{\lambda^n t^n e^{-\lambda t}}{n!} = \frac{e^{-\lambda t}(\lambda t)^n}{n!}$$

B. Pure Death Process

$$\frac{d\,P_n(t)}{dt} = -\mu\left[P_n(t)\right] + \mu\left[P_{n+1}(t)\right] \tag{2.1.8}$$

$$\frac{d\,P_{N_0}(t)}{dt} = -\mu\left[P_{N_0}(t)\right] \tag{2.1.10}$$

$$\frac{d\,P_0(t)}{dt} = \mu\left[P_1(t)\right] \tag{2.1.12}$$

To solve (2.1.10) take the Laplace transform of both sides.

$$pY_{N_0} - P_{N_0}(0) = -(\mu)(Y_{N_0})$$

Solving for Y_{N_0}

$$Y_{N_0} = \frac{P_{N_0}(0)}{p+\mu} = \frac{1}{p+\mu}\ .$$

Taking the inverse transform of both sides

$$P_{N_0}(t) = e^{-\mu t}$$

To solve (2.1.8) take the transform of both sides.

$$pY_n - P_n(0) = -(\mu)(Y_n) + (\mu)(Y_{n+1})$$

Solving for Y

$$Y_n = \frac{\mu(Y_{n+1}) + P_n(0)}{p+\mu} = \frac{\mu(Y_{n+1})}{p+\mu}\ .$$

Let $n = (N_0-1)$. Then

$$Y_{N_0-1} = \frac{\mu}{(p+\mu)^2}\ .$$

Let $n = (N-2)$. Then

$$Y_{N_0-2} = \frac{\mu^2}{(p+\mu)^3}\ .$$

By mathematical induction, when $n = (N_0-k)$

$$Y_{N_0-k} = \frac{\mu^k}{(p+\mu)^{k+1}} = \frac{\mu^{N_0-n}}{(p+\mu)^{N_0-n+1}}.$$

Taking the inverse transform of both sides

$$P_{N_0-k}(t) = \frac{(\mu t)^k e^{-\mu t}}{k!} = \frac{(\mu t)^{N_0-n} e^{-\mu t}}{(N_0-n)!} = P_n(t).$$

To solve (2.1.12) take the transform of both sides.

$$pY_0 - P_0(0) = \mu Y_1$$

Solving for Y_0

$$Y_0 = \frac{\mu Y_1}{p}$$

$$= \frac{\mu(\mu^{N_0-1})}{p(p+\mu)^{N_0}} = \frac{\mu^{N_0}}{p(p+\mu)^{N_0}}$$

$$= \frac{1}{p} - \left[\frac{1}{(p+\mu)} + \frac{\mu}{(p+\mu)^2} + \frac{\mu^2}{(p+\mu)^3} + \cdots + \frac{\mu^{N_0-1}}{(p+\mu)^{N_0}}\right]$$

(by partial fractions)

$$= \frac{1}{p} - \left[Y_{N_0} + Y_{N_0-1} + Y_{N_0-2} + \cdots + Y_1 \right].$$

Taking the inverse transform of both sides

$$P_0(t) = 1 - P_{N_0}(t) + P_{N_0-1}(t) + P_{N_0-2}(t) + \cdots + P_1(t)$$

$$= 1 - \sum_{n=1}^{N_0} P_n(t)$$

$$= 1 - \sum_{n=1}^{N_0} \frac{e^{-\mu t} \ (\mu t)^{N_0-n}}{(N_0 - n)!} .$$

REFERENCES

(1) C. G. Kling, Urban Transportation: Problems and Prospects, (New York: Vantage Press, 1976).

(2) D. R. Drew, Traffic Flow Theory and Control, (New York: McGraw Hill, 1968).

(3) J. Riordan, Stochastic Service Systems, (New York: John Wiley & Sons, Inc., 1962).

(4) W. A. Jessiman, A Simulation Model of Two-Lane Two-Way Traffic Flow, Research Report R65-14, Massachusetts Institute of Technology, School of Engineering, Cambridge, Massachusetts, 1965.

(5) E. Parzen, Stochastic Processes, (San Francisco: Holden-Day, 1962).

(6) A. M. Mood, F. A. Graybill, and D. C. Boes, Introduction to the Theory of Statistics, (New York: McGraw Hill, 1974).

(7) L. Kells, Elementary Differential Equations, (New York: McGraw Gill, 1960).

(8) T. L. Saaty, Elements of Queuing Theory with Applications, (New York: McGraw Hill, 1961).

(9) N. Wiener, "Differential Space," Journal of Mathematical Physics of the Massachusetts Institute of Technology, Vol. 2, 1923, pp. 131-174.

(10) A. Einstein, Investigations on the Theory of the Brownian Movement, (New York: Dover, 1956).

(11) J. D. C. Little, "A Proof for the Queuing Formula: $L = \lambda W$," Operations Research, Vol. 9, No. 3, 1961, pp. 383-387.

(12) S. Stidham, "A Last Word on $L = \lambda W$," Operations Research, Vol. 22, No. 2, 1974, pp. 417-421.

(13) F. T. Barwell, Automation and Control in Transport, (Oxford, England: Permagon Press, 1973).

(14) F. S. Hillier and G. J. Lieberman, Introduction to Operations Research, (San Francisco: Holden-Day, 1980).

(15) T. W. Rioux and C. E. Lee, "TEXAS - A Microscopic Traffic Simulation Package for Isolated Intersections," Paper presented at the 56th Annual Meeting of the Transportation Research Board, January 1977, Center for Highway Research - The University of Texas at Austin.

(16) N. R. Draper and S. Smith, Applied Regression Analysis, (New York: John Wiley & Sons, Inc., 1966).

(17) J. Neter, W. Wasserman, and G. A. Whitmore, Applied Statistics, (Boston: Allyn and Bacon, 1978).

(18) G. W. Snedecor and W. G. Cochran, Statistical Methods, (Ames, Iowa: The Iowa State University Press, 1967, 6th Edition).

(19) F. A. Graybill, Theory and Application of the Linear Model, (North Scituate, Massachusetts: Duxbury Press, 1976).

(20) A. Hald, Statistical Theory with Engineering Applications, (New York: John Wiley & Sons, Inc., 1952).

(21) H. Woods, H. H. Steinour, and H. R. Starke, "Effect of Composition of Portland Cement on Heat Evolved during Hardening," Industrial and Engineering Chemistry, No. 24, 1932, Table I, pp. 1207-1214.

(22) R. F. Heckard, J. A. Pachuta, and F. A. Haight, "Safety Aspects of the National 55-MPH Speed Limit," The Pennsylvania State University, November 1976.

(23) W. J. Kemper and S. R. Byington, "Safety Aspects of the National 55-MPH Speed Limit," Public Roads, Vol. 41, No. 2, Septebmer 1977, pp. 58-67.

(24) N. Enustun and A. H. Yang, "The 55-MPH Limit: Effect on Accidents," Traffic Engineering, Vol. 45, No. 8, August 1975, pp. 22-75.

(25) H. S. Dawson, "Analysis of Fatal Highway Accident Trends of Maryland Highways, 1970-1976," Public Roads, Vol. 43, No. 2, September 1979, pp. 62-68.

(26) D. A. Belsley, E. Kuh, and R. E. Welsch, Regression Diagnostics, (New York: John Wiley & Sons, Inc., 1980).

(27) C. E. Wisler, "Partitioning the Explained Variation in a Regression Analysis," Technical Paper No. 3, Office of Program Planning and Evaluation, Office of Health, Education, and Welfare, Washington, D. C., July 23, 1969.

BIBLIOGRAPHY

Barwell, F. T., Automation and Control in Transport. Oxford, England: Permagon Press, 1973.

Belsley, D. A., Kuh, E., and Welsch, R. E., Regression Diagnostics. New York: John Wiley & Sons, 1980.

Dawson, H. S., "Analysis of Fatal Highway Accident Trends of Maryland Highways, 1970-1976." Public Roads, Vol. 43, No. 2, September 1979.

Draper, N. R. and Smith, S., Applied Regression Analysis. New York: John Wiley & Sons, Inc., 1966.

Drew, D. R., Traffic Flow Theory and Control. New York: McGraw Hill, 1968.

Einstein, A., Investigations on the Theory of the Brownian Movement. New York: Dover, 1956.

Enustun, N. and Yang, A. H., "The 55-MPH Limit: Effect on Accidents." Traffic Engineering, Vol. 45, No. 8, August 1975.

Graybill, F. A., Theory and Application of the Linear Model. North Scituate, Massachusetts: Duxbury Press, 1976.

Hald, A., Statistical Theory with Engineering Applications. New York: John Wiley & Sons, Inc., 1952.

Heckard, R. F., Pachuta, J. A., and Haight, F. A., "Safety Aspects of the National 55-MPH Speed Limit." The Pennsylvania State University, November 1976.

Hillier, F. S. and Lieberman, G. J., Introduction to Operations Research. San Francisco: Holden-Day, 1980.

Jessiman, W. A., A Simulation Model of Two-Lane Two-Way Traffic Flow. Research Report R65-14, Massachusetts Institute of Technology, School of Engineering, Cambridge, Massachusetts, 1965.

Kells, L., Elementary Differential Equations. New York: McGraw Hill, 1960.

Kemper, W. J. and Byington, S. R., "Safety Aspects of the National 55-MPH Speed Limit." Public Roads, Vol. 41, No. 2, September 1977.

Kling, D. G., Urban Transportation: Problems and Prospects. New York: Vantage Press, 1976.

Little, J. D. C., "A Proof for the Queuing Formula: $L = \lambda W$." Operations Research, Vol. 9, No. 3, 1961.

Mood, A. M., Graybill, F. A., and Boes, D. C., Introduction to the Theory of Statistics. New York: McGraw Hill, 1974.

Neter, J., Wasserman W., and Whitmore, G. A., Applied Statistics. Boston: Allyn and Bacon, 1978.

Parzen, E., Stochastic Processes. San Francisco: Holden-Day, 1962.

Riordan, J., Stochastic Service Systems. New York: John Wiley & Sons, Inc., 1962.

Rioux, T. W. and Lee, C. E., "TEXAS - A Microscopic Traffic Simulation Package for Isolated Intersections." Paper presented at the 56th Annual Meeting of the Transportation Research Board, January 1977, Center for Highway Research - The University of Texas at Austin.

Saaty, T. L., Elements of Queuing Theory with Applications. New York: McGraw Hill, 1961.

Snedecor, G. W., and Cochran, W. G., Statistical Methods. Ames, Iowa: The Iowa State University Press, 1967, 6th Edition

Stidham, S., "A Last Word on $L = \lambda W$." Operations Research, Vol. 22, No. 2, 1974.

Wiener, N., "Differential Space." Journal of Mathematical Physics of the Massachusetts Institute of Technology, Vol. 2, 1923.

Wisler, C. E., "Partitioning the Explained Variation in a Regression Analysis." Technical Paper No. 3, Office of Program Planning and Evaluation, Office of Health, Education, and Welfard, Washington, D. C., July 23, 1969.

Woods, H., Steinour, H. H., and Starke, H. R., "Effect of Composition of Portland Cement on Heat Evolved during Hardening." Industrial and Engineering Chemistry, No. 24, 1932.

SOME APPLICATIONS OF STATISTICAL METHODS IN TRAFFIC ENGINEERING

by

JOHN CARL NELSON

B.S., Kansas State University, 1980

AN ABSTRACT OF A MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Statistics

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1981

This report describes some basic statistical concepts which can be used in the problem-solving techniques of traffic engineering. The introduction includes a short history of the role of statistics in traffic engineering. Results from two major areas of statistics, queueing theory and regression, are included. The section on queueing theory includes some results concerning vehicles waiting at a toll booth or some other service situation. Several cases are considered. The section on regression analysis highlights the procedures and results of three studies done on the safety effects of the 55-MPH speed limit. A short statement on the role of statistics in traffic engineering concludes the paper.