TN 4: ESTIMATING PARK ATTRACTIVENESS, POPULATION CENTER EMISSIVENESS, AND THE EFFECT OF DISTANCE (LOCATION) IN OUTDOOR RECREATION TRAVEL

BY F. J. CESARIO

ABSTRACT

This report contains a discussion of the results of the application of a two-stage methodology for estimating certain parameters of outdoor recreation travel. Stage I analysis results in extracting systematic travel (distance), population center (emissiveness) and park effects (attractiveness) parameters, from outdoor recreation trip-making data by the use of an analysis of covariance technique. A "reaction to travel parameter" is also determined. Stage II analysis is devoted to accounting for differences in estimated population center (emissiveness) and park (attractiveness) effects by use of multivariate analyses. The goal is to identify those population center characteristics (population, magnitudes, income, age, etc.) that appear to best account for observed variations in aggregate outdoor-recreation trip-making behaviour and to identify park characteristics, etc. that account for attractivity.

Stage I analysis of Ontario data obtained from camping permits showed that statistically significant population center, park and location effects on camper trip-making could be extracted from the 1968 Ontario data on all use of Ontario campgrounds.

Stage II analysis had some expected consequences and some unexpected consequences. Emissivenesses (city effects) were found to be explained by city size with income not being important. It is well known that the volume of visitor flow from an origin to a destination is proportional to the population of the origin, but income is usually also considered to influence the volume of flows. For attractiveness, it was found that such expected factors as size of park and numbers of campsites had a positive influence on park attractiveness. But, the complexity of the relation discovered has not been recognized in the past.

The ways in which the results can be put to practical use by planners and researchers are emphasized.

INTRODUCTION

In the planning and design of park systems, outdoor recreation agencies are continually faced with complex questions. How many parks are needed? Where should new parks be located? What existing parks should be expanded and in what ways? What mix of what activities should be provided on what sites? How many camping spaces should be provided? The list of such planning questions is virtually endless. Since recreation budgets are not unlimited, hard choices must be made in deciding exactly what levels and mixes of recreation opportunities to provide and where to provide them.

It can be realistically argued that rational planning would involve a systematic exploration of alternative plans and policies in a formal or systematic way: That is, an "informed" choice would take into account all (or as many as possible) of the consequences of alternative solutions to a particular planning problem before deciding upon a specific course of action. In many cases, the park use implications of alternative plans and policies are of paramount importance. For example, a recreation agency might wish to choose that alternative which promised to result in the greatest increase in regional recreation visits (distributed across properties). But then it is imperative that the planning agencies have on hand considerable relevant information with which to assess the visitor implications of each financially feasible plan of recreational development that may be considered (to which a total amount of resources would be distributed – consideration subject to constraints on total resources).

In an attempt to develop this relevant planning information, the last ten years or so have seen many outdoor recreation researchers working on the construction of "models" which purport to "explain" or "predict" various aspects of outdoor recreation trip-making behavior (see #References 6, 17, 18, 27, 35.) Models are used to relate various measures of park use to the factors that give rise to them. They have, in fact, provided many insights as to what factors are and are not important.

Yet, these modeling efforts have been plagued by some common difficulties. The modeler is inevitably faced with the complex problems of, first, selecting the appropriate variables to include in his model and, second, postulating how these variables might combine to give rise to a particular use pattern. In short, there are major unresolved problems concerning both the substance and the form of the relationships. In many cases the variables are arbitrarily chosen, perhaps on the basis of data availability, and a linear model is constructed (e.g., because that class of model is easily soluble). But using linear models is a rather restrictive approach particularly considering that the linear model may be a poor approximation to reality. Given that the there are problems with many, possibly most linear models, it is possible to do better. Such improvement is a goal of this study.

This report contains the results of a pilot study carried out, using a new and untested methodology developed by the author, to identify those factors which by themselves and in combination appear to be the most important influences shaping park use levels and related patterns in outdoor recreation systems. As will be seen, this new methodology represents a substantial departure from previous methodologies and, with appropriate qualifications, offers considerable promise of (1) allowing better prediction of visitor flows than has been possible, and (2) affording insights into the particular factors that are important in determining these flows.

Utilizing findings obtained in this study and possible future applications of the research methodology, outdoor recreation planners and researchers will hopefully be able to make improved decisions in their respective areas of responsibility.

FRAMEWORK FOR ANALYSIS

The purpose of this investigation is to test, on a pilot study basis, a research methodology to identify factors that appear to be important determinants of park use and to uncover the particular (aggregate) characteristics of specific parks and of specific population centers which give rise to observed variations in regional outdoor recreation trip-making behavior. In addition to revealing these population center and park effects, the effect of distance (i.e., location) on participation is considered.

For analysis purposes, assume that the region of interest (e.g., part of a province, an entire province, several provinces) is subdivided into N mutually exclusive population centers (e.g., counties, townships, census tracts). Assume that there are M outdoor recreation parks located in, and possibly out of, the region. The distance between population center i and park j, as measured, say by the number of road miles over the route of minimum distance, is denoted by $d(i,j)$. The array of $d(i,j)$) is an NxM matrix of trip distances. Let $v(i,j)$ be the number of visits of a particular type observed to emanate from population center i and terminate at park j during some time period of finite duration. Then the array of $v(i,j)$) is an NxM matrix of trip frequencies.

Let the total number of recreation trips emanating from population center i to the parks being considered be O(i). Let the total number of trips terminating at park j be denoted by V(*,j); and let the total number of trips taken in the region be denoted by W. Then:

(1) $O(i) = V(i,*) = \Sigma_k v(i,k)$ for k=1 to M

(2) $V(*, j) = \Sigma_k v(k,j)$ for k=1 to N

(3) $W = \Sigma_i O(i) = \Sigma_j V(j) = \Sigma_i \Sigma_j v(i,j)$ for i= 1 to N and j=1 to M.

Quantities of interest are displayed in Table 1. This analysis assumes that trip-making data of the above type are available for at least one time period (e.g., a summer season). Further, it is assumed that it is possible to measure certain overall socio-economic characteristics of the population centers as well as site specific characteristics of the parks under consideration. In this context, we then ask what it is possible to learn from these data.

TABLE 1: TRIP DISTRIBUTION TABLEAU

	Parks						
Origins	1	2	●●●			M	$\Sigma_j v_{ij}$
1	v_{11}	v_{12}	●●●			v_{1M}	O_1
2	v21	v_{22}	●●●			v_{2M}	O_2
●	●	●	●			●	●
●	●	●		●		●	●
●	●	●			●	●	●
N	v_{N1}	v_{N2}	●●●			v_{NM}	O_N
$\Sigma_i v_{ij}$	v_{*1}	v_{*2}				v_{*M}	W

METHODOLOGY

The methodology employed in this study is a two-stage estimation procedure. The first stage of the analysis extracts systematic population center and park effects from the data structured as in Table 1. The second stage is to identify the population center and park characteristics that appear to account for differences in these effects. Before discussing each stage of the analysis in detail, it is useful to present a brief overview of the underlying modeling postulates.

THEORETICAL BASIS

Basically, it is postulated that the number of recreation trips made from population center i to park j is a function of (1) certain population center characteristics such as population, size, income distribution and measures of other socio-economic factors, (2) certain park characteristics such as acreage, miles of shoreline and other natural or man-made features, and (3) spatial separation. Further, it is assumed that population center characteristics relate to each other in *a* certain way. Similar assumptions can be made with respect to park characteristics and spatial separation. Thus the number of visits from population center *i* to park j may be expressed as:

(4) V(i,j) = f(g1(population center characteristics), g2 (park characteristics), h(d(i,j))).

It is assumed that g1 and g2 are independent of h(d(i,j)) and that their particular functional forms are unknown. The functional form of h(d(i,j)) is presumed to be specified except for values of its unknown parameters. Viewed in this way the estimation problem is one of finding appropriate functions g1 and g2 as well *as any* unknown parameters of the particular distance function h(d(i,j)) employed.

In a lengthy development, Cesario (see #Reference 8) made the above concepts operational by formulating the following model structure:

(5) $V(i,j) = kE(i)A(j)h(D(i,j))\exp(\varepsilon(i,j))$

It is to be representative of the data, where k is *a* constant of proportionality; E(i) is an unobservable population center factor called "emissiveness"; A(j) is an unobservable park factor called "attractiveness"; h(d(i,j)) is a particular distance function; and $\varepsilon(i,j)$ is an error term. The quantities v(i,j) and d(i,j) are observable whereas E(i), A(j) and parameters of h(d(i,j)) are unobservable and must be estimated.

Cesario (see Reference 7, 8) indicates vectors [E(i)] and [A(j)] are each specified up to multiplicative constants (i.e. only ratios E(i)/E(k) and A(j)/A(l) have meaning). Thus, emissiveness of a population center, E(i), reflects its relative propensity (relative to other population centers) to emit trips under identical circumstances - (that is if all centers were confronted by the same availability of "supply" of recreation opportunities). Emissiveness thus serves as a relative measure of population center "participation" that is unencumbered by the existence of a differential supply. It is seen to be a metric that represents the combined effects of a multitude of population center characteristics on recreation trip-making.

It is possible, of course, to think of many reasons why one origin might be more emissive than another. One obvious reason is due to the possible differences in population sizes among population centers. Another reason might be differences in socio-economic composition of the population. Each of these factors needs to be investigated to discover the role each plays in giving rise to differences in E(i) values (this relates to estimating the form of g1).

In contradistinction to emissiveness, attractiveness of park j, A(j), reflects its relative ability (compared to other parks) to attract trips under identical circumstances (for example, as if all parks were equally accessible from each population center). Attractiveness thus serves *as a* measure of park "popularity". It is a metric that reflects the combined effects of a multitude of park characteristics on recreation trip-making. There are many potential reasons why one park might be more appealing than another: one might have excellent swimming facilities whereas another does not; another might have an excellent network of hiking traits, whereas another has none. Again there is a need to account for differential A(j) values in terms of these factors (to estimate the nature of g2).

Finally, the distance function h(d(i,j)) is a "correction" term to account for violations of the above ceteris paribus conditions. Various functional forms can be used. Common examples-are $h(d(i,j)) = d(i,j)^{\beta}$, $h(d(i,j)) = e^{\beta d(i,j)}$ and $h(d(i,j)) = d(i,j)^{\beta}\exp(\alpha d(i,j))$ where *a* and β are parameters to be estimated. Each of these functions has different behavioral implications (see Reference 14).

From the above, we see that the process of extracting relevant factors affecting park use may *be* thought of in two stages. Stage I involves the estimation of vectors [$\hat{E}(i)$] and [$\hat{A}(j)$] of Equation 5. Stage II then involves determining the combinations of population center characteristics and park characteristics that appear to account for whatever differences were found to exist in the $\hat{E}(i)$'s and $\hat{A}(j)$'s. This is the essence of the methodology being tested.

Stage I: Estimation of Vectors [$\hat{E}(i)$] and [$\hat{A}(j)$]; of attractiveness and emissiveness values

A logarithmic transformation of Equation 5 for a particular distance function (h) results in what is essentially an analysis of covariance model from which estimates [Ê(i)] and [Â(j)]can be ultimately obtained. Letting h(d(i,j)) = $d(i,j)^{\beta}$ for convenience, we form:

(6) $V(i,j) = kE(i)A(j)d(i,j)^{\beta}\exp(\varepsilon(i,j))$

Taking Logarithms of both sides of Equation 6 yields:

$\ln(v(i,j)) = \ln(k)+ \ln(E(i))+ \ln(A(j))+ \beta\ln(d(i,j))+ \varepsilon(i,j).$

Letting y(i,j) = ln(v(i,j)); m=ln(k); e(i)=ln(E(i)); a(j)=ln(A(j)); x(i,j)=ln(d(I,j))

produces the linear model:

(8) $y(i,j) = m + e(i) + a(j) + \beta x(i,j) + \varepsilon(i,j).$

In Equation 8, y(i,j) and x(i,j) are given and e(i), a(j), β are unknown. Estimation by the usual least-squares method yields vectors [ê(i)] and [â(j)] which, when transformed by taking antilogarithms, yield (biased) estimates of [Ê(i)] and [Â(j)]. That is, $e^{\hat{e}(i)} \approx E(j) \approx \hat{E}(j)$. The estimated parameters have the interpretation provided in the previous section.

If one is willing to make some rather strong assumptions on the distribution of ε(i,j), a test for the statistical significance of β can be made, as well as tests of hypotheses about [ê(i)] and [â(j)] and [Ê(i)] and [Â(j)]. These are tests are modifications of usual analysis of variance methods. (See #References 10, 29.) In addition, various contrasts (comparisons and grouping based on similar parameters) of population center and park parameters can be tested. This is to sort out homogeneous groupings. See TN 19 for an expanded discussion of these points.

As implied by previous discussion, the Stage I analysis provides information on the relative "importance" of population centers, parks, and location in accounting for the observed total variation in population center to park trips (i.e., visitor flows). For instance, if all Ê(i)'s are equal, differences between population centers do not contribute to variation. Likewise, if all Â(j)'s are equal, differences between parks do not contribute to the total variation in use between parks, at least in a simple way (it is possible that the effects of some factors may cancel out). If, on the other hand, some Ê(i)'s and/or Â(j)'s are different, then some variation in v(i,j) is due to these factors and further investigation into the causes of these differences would be in order. Finally, the magnitude of the distance parameter provides information for making inferences about location affecting visiting. For example, if is negative and large, it indicates that the elasticity of recreation trips with distance is large or, more simply, that visitor flows fall off rapidly with small increments in distance. On the other hand, if is identically 0, then (everything else being equal) the same visitor flows would be achieved for a park no matter where it was located. If is positive, then the plausibility of the results or the data would be in question (because increasing travel with destinations of equal attractiveness being further away is not to be expected – nor is no effect of distance, i.e. β=0).

STAGE II: ACCOUNTING FOR DIFFERENCES IN Ê(i) AND Â(j)

Assume that estimated parameter sets {Ê(i)|i= 1 to N} and {Â(j)|j=1 to M} are available from Stage I. Also assume that not all Ê(i)'s are the same and not all Â (j) are the same. The Stage II problem is to account for differences among Ê (i)'s and Â (j)'s. These quantities are taken to be the dependent variables in multivariate analyses which have as their purpose the selection of a set of independent variables that appear to account for variation in these dependent variables. In the Â(i) analysis, independent variables consist of various measurable characteristics of population centers which might make one center more emissive than another. In the Â(j) analysis, independent variables consist of various measurable characteristics of parks which might make one more attractive than another. For applications where few independent variables are involved, standard linear or nonlinear regression approaches might be used; in more complicated problems where many variables are expected to interact in complex ways, the Automatic Interaction Detector Technique (AID, Sonquist and Morgan 1964) is a useful approach.

While the use of standard regression analysis is commonplace, a few words of explanation about the potential usefulness of the AID technique might be in order. The technique developed by Sonquist and Morgan (1964) is a multivariate method of analysis which has as its purpose the "classification" of data into homogeneous groups. This empirical method is free of many of the traditional strong, sometimes false or poor approximation, assumptions implicitly employed in the analysis of data by techniques such as regression analysis. Given a set of data (measured on nominal, ordinal or interval scales) on some "independent' variables, the AID technique determines what variables combine to produce the greatest discrimination in group means of an interval scale dependent variable.

To carry out separation, the total population of observations is divided into mutually exclusive "terminal" subgroups through a sequence of binary splits . At each stage of the sequence, the dichotomization occurs so as to provide the largest reduction in the unexplained sum of squares of the dependent variable. The split into two groups is chosen that accounts for more of the total remaining sum of squares of the dependent variable than splitting of any other “next level” split (further splitting of current combinations). For example if there has been a split on gender and for males a split between under 40 and above, one has 3 “current” groups and the possibility of further splitting them on age or any other independent variable. In this way, the mean values of the dependent variable will be as different as possible between groups, but as equal as possible within groups(largest differences given the splitting sequence – differences depend on the sequence so if one forces an initial split on age, one may produce a quite different pattern of differences than occurs when gender is the first split).

The output of the analysis is a so-called "tree diagram" which shows the subgroups formed at each iteration (splitting), along with associated statistics. (A very readable example of the use of AID in a marketing context is given by Assael, see #Reference 2. Cesario, see Reference 9, discusses a hypothetical use of AID in the present context.)

The relevance of this AID technique to the Stage II problem is obvious. One has emissiveness and attractiveness factors estimated in Stage 1 for large numbers of population centers and parks, respectively, together with data on particular population center and park characteristics hypothesized to give rise to emissiveness and attractiveness differentials. AID is used to "classify" population centers and parks into homogeneous groupings. Beyond the informative aspects of these AID results - finding out which variables are important and which are not - it becomes possible for planning purposes to make informed estimates of the attractiveness of any new or altered parks for which appropriate measurements are available, and to assess the associated park use implications. This systematic approach may be contrasted with the procedure of defining attractiveness in an ad hoc way. (See TN 1 and other references cited there.) It follows that results of AID analysis will provide researchers with a more objective basis for the construction of improved visitation models than has occurred in other research of which the author is aware.

DATA BASE

To implement the proposed two-stage methodology described in the previous section, it was necessary to construct (v(i,j)) and (d(i,j)) matrices for one or more visitor types in a particular Canadian region, and for the Stage II analysis to collect information on the characteristics of population centers and parks in the region.

THE STUDY REGION

For statistical reasons it was desirable to select a study region for which a reasonably large data tableau of the type shown in Table 1 could be constructed. To minimize bias it was desirable to also work with a matrix containing as many positive (hopefully large) entries as possible. Recall that the parameters $\hat{E}(i)$ and $\hat{A}(j)$, which are antilogarithms of $\hat{e}(i)$ and $\hat{a}(j)$ respectively, are biased estimates of $\hat{E}(i)$ and $\hat{A}(j)$. This bias is attributable to the fact that in the "log-linear" model given by Equation 8 undue weight is given to small visitor flows. GLS, generalizes least squares allows for weighting down small flows but there are other non-linearity issues influencing bias (see TN 19 re GLS and SAS Proc NLIN references re non-linear estimation as an alternative to transformation, e.g. SAS/STAT User's Guide, Vol. 2 ch. 29, 1990, Cary, INC, SAS Institute).

Existing data sources were examined with bias and other considerations in mind. After examining the potential for each region in Canada to provide an appropriate data base for estimation purposes, it was determined that the region of Southern Ontario bordered roughly on the North by the cities of North Bay and Sudbury, on the East by Ottawa, on the South by Lakes Ontario and Erie, and on the West by Lake Huron and Georgian Bay, would be suitable. This region is shown in Figure 1. The region of Southern Ontario contains recreation parks operating under several different jurisdictions, - e.g., Parks Canada, Ontario Provincial Government and various Conservation Authorities. Unfortunately, due to paucity of data it was not possible to include all of the region's parks in the study. Concentration was focused on the analysis of data for as many provincial parks as possible.

FIGURE 1

STUDY REGION

VISITOR TYPES

Consistent with the Canadian Outdoor Recreation Demand Study classification, visits to recreation areas can be broken down into the following types:

A. Day-use, main destination
B. Day-use, stopover
C. Overnight-use, main destination
D. Overnight-use, stopover

Visit type A (day-use, main destination) is occurs when a party originates a recreation trip at home, visits only one park and returns home on the same day. Visit type B (day-use, stopover) leaves home and visits two or more sites on the same day before returning home. Visitor type C (overnight-use, main destination) embarks on *a* camping trip from home and travels to one and only one camping site before returning home after completing the stay. Visitor type D (overnight-use, stopover) is typified by vacation campers. This party leaves the home and whether or not an ultimate single destination is the goal, stops for short periods of time at many different campgrounds on a single trip before returning home. Each of the above visitor types is motivated by a different set of factors (each has different amounts of leisure time available for the trip) and conceivably can be considered separately for modeling purposes. It is important to note that the methodology being applied in this study is most relevant for the *analysis* of the trip-making patterns of visitor types A and C. Attention *was* focused, then, on securing data with which to construct (v(i,j)) matrices for one or both of these visitor types.

One should note that the text above has been altered to refer to visits. This is to avoid the logical problem associated with classifying a visitor when, in fact, one can only classify a visit, given that a person making a day visit today can be making a Type C visit the next weekend.

In connection with Visit Type A, data for day-use visit types at thirty-five parks were available from CORD Study surveys conducted in 1969; however, there were problems in defining relevant sampling rates. In addition, 1971 data (10 percent sampling rate) were available from the Ontario Department of Lands and Forests (ODLF). Due to very serious problems in establishing the populations to which the observations in both surveys belonged (i.e., total visitor flow by type at a park on any given sampling day) these data could not be used with any confidence the flows recorded resembled actual day-use trip patterns.

With reference to Visit Type B, camper data for main destination and stopover types were also available from CORD Study tabulations. Further, the ODLF provided detailed tabulations based on a 100 percent canvass of camping permits issued during 1968. Unfortunately, ODLF data were not broken down into main destination and stopover categories. The sample did, however, provide an accurate estimate of the total population of campers (the CORD Study data on campers and day-use visitors suffer from problems in defining sampling rates).

In collaboration with the Outdoor Recreation Research Section, Parks Canada, it was decided that Visit Type C (main destination campers) would be the most appropriate group to consider in the pilot study. A mixture of CORD Study and ODLF data were then used to construct a relevant matrix, (v(i,j)). The ODLF data were used to define an initial (v(i,j)). But since the ODLF camper data did not distinguish between main destination and stopover use, a correction procedure, based on an analysis of CORD Study main destination/stopover data, was developed to estimate of the number of campers in main destination and stopover categories. *The estimation procedure was described in an Appendix to this TN in its original version but is omitted here since it is not relevant 30 years or more later.*

Compilation of the Data Set

The population centers chosen for use in this study were 46 cities in the region shown in Figure 1 having a 1970 population greater than 9,000, according to Statistics Canada estimates (i.e., N = 46). in addition to population sizes, average income data were the only socio-economic information available for use in the Stage II analysis. (Actually, data on average income were only available for 30 out of the total of 46 population centers.) Table 2 presents a list of the 46 Southern Ontario population centers used in the study with population data and average income estimates that were available.

For the 46 population centers identified in Table 2, it was possible to compile main-destination camper estimates for 50 provincial parks in the Southern Ontario region (i.e., N = 50). In anticipation of the Stage II analysis, data on park characteristics were collected from various published reports of ODLF and others. Information on 15 characteristics was available for each of the 50 parks included in the study. Table 3 presents the park characteristics information collected and gives the coding system used in the AID analysis. Table 4 lists the 50 parks included in this study along with (coded) associated characteristics.

EMPIRICAL RESULTS

The results of applying the previously described two-stage methodology in the analysis of the Southern Ontario data are described in this section. Results for each stage of the analysis are given separately.

Stage I Results

Parameters of Equation 8, the transformed version of Equation 6, were estimated both for "unadjusted" and "adjusted" v(i,j) data. The distance function $h(d(i,j))=d(i,j)^{\beta}$ was used in both cases. The estimates of k and β are given in Table 5. Recall that unadjusted data include both main destination and stopover campers, while adjusted data *are* estimates of main destination campers only.

As suspected, there is considerable correspondence between the two sets of results. The magnitudes of the estimated values of k are of little importance since k acts merely as a scaling parameter in the model.

As mentioned before, the parameters have certain behavioral implications; hence, the sign and magnitude of β's estimated value is of considerable importance. In this analysis, the estimates of β using unadjusted and adjusted data were both significantly different from 0 (beyond the 0.0001 level). β being negative in both cases is in conformance with the usual hypothesis that visitor flow is a decreasing function of distance; that is, all other things being equal, the farther a park is located away from a population center the fewer will be its visitors. The numerical estimates of β provide a clue to the particular rate at which visitor flow declines with distance.

The estimated emissiveness, $\hat{E}(i)$, and attractiveness, $\hat{A}(j)$, parameters are given in Tables 6 and 7 respectively. Again, results are given for analyses which used both unadjusted and adjusted v(i,j) observations. Since the estimates obtained by using the analysis of convariance procedure to estimate Equation 8 are in natural logarithms, transformed results are also provided. Correspondence between the adjusted and unadjusted results is again apparent. Nevertheless, due to the crudeness of the main destination/stopover adjustment process there are bound to be some irregularities in these tables, especially Table 7. For instance, Chutes gets a relatively high attractiveness value for main destination use even though from an independent analysis it appears that this park, being adjacent to the Trans-Canada highway, serves primarily stopover users. The number of these possible irregularities is hopefully quite small and one would suspect that the AID results would not be very much distorted.

TABLE 2: POPULATION CENTER DATA

Code No.	Population Center (City)	Population 1969(a)	Ave. Income 1969(b)
1	Brantford	62,594	5,888
2	Ottawa	297,701	6,992
3	St. Thomas	24,520	5,900
4	Leamington	10,082	-
5	Windsor	198,997	6,969
6	.Kingston	59,029	6,138
7	Owen Sound	18,189	5,482
8	Burlington	78,590	-
9	Georgetown	14,964	-
10	Oakville	58,007	7,936
11	Belleville	34,190	6,146
12	Trenton	14,251	-
13	Chatham	34,158	6,306
14	Wallaceburg	10,637	-
15	Sarnia	56,407	7,264
16	Smith's Falls	9,701	-
17	Brockville	19,565	6,016
18	London	211,699	6,344
19	Simcoe	10,447	-
20	Cobourg-Port Hope	20,163	-
21	Ajax	11,273	-
22	Oshawa	86,185	6,813
23	Whitby	22,103	6,390
24	Woodstock	25,314	5,878
25	Brampton	39,232	6,902
26	Peterborough	57,337	6,277
27	Hawkesbury	9,240	-
28	Pembroke	16,431	5,737
29	Barrie	26,212	6,006
30	Midland	10,646	-
31	Orillia	20,542	5,040
32	Cornwall	46,576	5,455
33	Lindsay	12,483	-
34	Galt	36,734	5,756
35	Kitchener-Waterloo	138,345	6,165
36	Preston	15,385	-
37	Niagara Falls	63,054	6,048
38	Welland	42,622	6,263
39	Guelph	56,603	6,087
40	Hamilton-Dundas	321,277	6,506
41	Aurora	12,338	-
42	Toronto(c)	1,253,000	6,741
43	Newmarket	15,267	-
44	Richmond Hill	27,339	-
45	North Bay	38,966	5,990
46	Sudbury-Copper Cliff	92,131	6,371

(a) **Source: Estimated from 1966 and i971 data provided in Ontario Population Statistics, 1971, prepares by Planning Research Section, Municipal Planning and Development Branch, Urban and Regional Development Division.**

(b) **Source: Inside - Taxation, 1972, *a* publication of the Department of National Revenue and Taxation.**

(c) **Includes Agincourt, Etobicoke, Don Mills. New Toronto, Scarborough, West Hill,**

(d) **Willowdale, and Toronto**

TABLE 3: PARK CHARACTERISTIC CODES

Variable	Descripion	Codes
X1	Number of acres	1 = 0-183 acres 2 = 184-849 acres 3 = 850 acres or more
X2	Number of camping units	1 = 0-100 units 2 = 101-275 units 3 = 276 units and over
X3	Length of swimming beach	1 = 0-800 feet 2 = 801-1900 feet 3 = 1901 feet or more
X4	Park class	1 = Recreational 2 = Natural Environment 3 = Primitive
X5	Acres of picnic area	1 = 0-2.5 acres 2 = 2.6-35 acres 3 = 36 acres or more
X6	Miles of hiking trails	1 = 0 miles 2 = 1 mile or more
X7	Availability of modern comfort stations	1 = Yes 2 = No
X8	Availability of museums/ exhibition centers	1 = Yes 2 = No
X9	Availability of boat-launch ramps	1 = Yes 2 = No
X10	Availability of boats for hire	1 = Yes 2 = No
X11	Availability of trailer sanitary stations	1 = Yes 2 = No
X12	Availability of showers	1 = Yes 2 = No

Using the appropriate distributional assumptions, it was determined that the proportions of the total variance in the data accounted for by population centers in the aggregate and by parks in the aggregate were significantly different from 0. Tests of specific contrasts of emissiveness and attractiveness parameters were not conducted.

The results are interpreted as follows. First, distance (location) is an important determinant of park use at a particular site and should be explicitly considered in decisions concerning the location of parks. It is, of course, no great surprise to learn that park use decreases as distance from population centers increases; any other result would be looked upon with a considerable amount of skepticism. The importance of this result in this study is that by use of the function h(d(i,j)) with an estimated β, the v(i,j) data are essentially "corrected" to eliminate the distance effect. To put it another way, the location effect due to distance (as indicated by the estimated value of β) is essentially removed from the v(i,j) data so that the estimated E(i)'s and A(j)'s represent the quantities that would be found to prevail if distance between every population and every park were equal. This interpretation of E(i) and A(j) is completely consistent with the interpretation provided in the operational definitions of these concepts provided earlier.

TABLE 4: CODED CHARACTERISTICS OF 50 ONTARIO PARKS

Park	Coded Park Characteristics											
Park	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12
Rock Point	2	1	2	1	2	1	2	3	1	2	1	2
Mara	1	1	1	1	2	1	1	3	2	2	1	2
Sibbald Point	2	3	3	1	3	2	1	1	1	2	1	2
Earl Rowe	3	3	3	1	3	1	1	3	1	2	1	2
Craigleith	1	2	3	1	2	1	1	2	2	2	2	2
Inverhuron	2	3	3	2	2	1	2	3	1	2	2	2
Point Farms	2	2	2	1	2	1	1	3	1	2	2	2
Sauble Falls	1	2	-1	1	2	1	1	3	2	2	1	2
Bass Lake	1	2	1	1	2	1	1	3	1	2	1	1
Selkirk	1	2	2	1	2	1	2	3	1	2	1	2
Devil's Glen	1	1	1	1	2	1	2	3	2	2	2	2
Turkey Point	2	3	2	2	2	1	1	3	1	2	1	2
Wheatley	2	2	3	1	2	1	1	3	1	2	1	2
Rondeau	3	3	3	2	3	2	2	1	1	2	1	2
Holiday Beach	2	1	2	1	3	1	1	3	1	2	1	2
Ipperwash	1	2	2	1	2	1	1	3	1	2	1	2
Long Point	2	3	2	1	2	1	1	3	1	2	1	2
Pinery	3	3	3	2	2	2	2	3	2	1	2	2
Antoine	1	1	1	1	2	1	2	3	1	2	1	2
Samuel de Champlain	3	2	2	2	2	2	2	2	1	2	1	2
Marten River	3	2	2	1	3	1	1	3	1	2	1	2
Mississagi	3	1	2	2	2	1	2	3	1	2	1	2
Chutes	2	1	1	1	2	2	2	3	2	1	1	1
Killarney	3	2	2	2	1	2	2	3	1	1	1	2
Windy Lake	2	1	3	1	3	1	2	3	1	2	1	2
Fitzroy	2	2	1	1	2	1	1	3	1	2	1	2
Rideau River	1	2	2	1	2	1	1	3	1	2	1	2
Silver Lake	1	2	1	1	1	1	1	3	1	2	1	2
South Nation	1	1	1	1	2	1	2	3	1	2	2	2
Six Mile Lake	1	2	1	1	2	1	2	3	1	2	1	2
Balsam Lake	3	3	2	1	2	2	1	3	1	2	1	2
Darlington	2	3	2	1	3	1	1	1	1	2	1	2
Emily	1	2	2	1	2	1	1	3	1	2	1	2
Presqu'ile	3	3	3	2	3	2	1	1	2	2	1	2
Serpent Mound	1	2	1	2	2	1	1	2	1	2	1	2
Arrowhead	3	2	2	1	1	1	1	3	1	2	1	2
Grundy Lake	3	3	3	2	2	2	1	3	1	2	1	2
Killbear Pt	3	3	3	2	2	2	2	3	1	2	1	2
Mikisew	1	2	2	1	2	1	2	3	1	2	1	2
Restoule	3	2	3	2	1	1	2	3	1	2	1	2
Sturgeon Bay	1	1	1	1	1	1	2	3	1	2	1	2
Algonquin	3	3	3	1	2	2	2	3	1	2	1	2
Carson Lake	1	1	1	1	1	1	2	3	1	2	1	2
Bonncherre	2	1	2	1	1	1	1	3	1	2	1	2
Black Lake	1	2	1	2	2	1	1	3	1	2	1	2
Bon Echo	3	3	3	1	2	2	2	3	1	2	1	2
Lake St. Pete	1	1	2	2	2	2	1	3	1	2	1	2
Outlet Beach	2	3	3	1	3	1	1	3	1	2	i	2
Oastler Lake	1	2	1	1	1	1	2	3	1	2	1	2
Driftwood	2	1	3	1	1	1	2	3	1	2	i	2

TABLE 5: STAGE I ESTIMATES OF K AND β

	Parameters	
	k	β
Unadjusted data	exp(9.878)	-1.291
Adjusted data	exp(10.770)	-1.579

One conclusion from these results is that the nature of the population centers having access to a park should *be* explicitly considered in planning decisions. Take any two population centers, for example London and Hamilton-Dundas to the east of London. The Stage I analysis showed (Table 6) that these two population centers are approximately equal in emissiveness. This means that if *a* park were equally distant from each of these population centers, ceteris paribus, it would be expected to receive the same number of visits from London that it would receive from Hamilton-Dundas. On the other hand, consider the comparison between London and Sarnia, a population center to the west of London. It is seen from Table 6 that London is about three times as emissive as Sarnia. A park equidistant from these two population centers would therefore receive a greater share of its visits from London. If virtually the same park were being contemplated for location in either of the above sites, its use by the three population centers would clearly be maximized at the former site, i.e., at *a* position equidistant from London and Hamilton-Dundas. (Of course, the real-world analysis is not so simple since many other population centers are involved and the ceteris paribus clause is in general violated; the distance or location effect must be incorporated explicitly into the analysis.)

It has been shown that the nature of parks significantly influences use patterns. Take any two provincial parks, say Rondeau and Black Lake. The Stage I analysis showed (Table 7) that Rondeau is about twice as attractive as Black Lake. This means that, ceteris paribus, Rondeau would draw twice as many visitors from a city as would Black Lake, indicating that Rondeau is twice as "popular" based upon the analysis of these data. Other similar comparisons could be made using the results in Table 7.

STAGE II RESULTS

The questions raised by Stage I revolve around what makes one population center more or less emissive than others and what makes one park more of less attractive than others. To answer these questions multivariate analyses were performed in Stage II to relate the emissiveness and attractiveness indices of Tables 6 and 7 to particular characteristics of population centers and parks, respectively.

TABLE 6: TWO SETS OF CAMPING EMISSIVENESS INDICES FOR 46 ONTARIO POPULATION CENTERS[(a)]

Rank	Study Code #	Population Center	Adjusted $\ln(\hat{E})$	Data $\hat{E}$	Unadjusted $\ln(\hat{E})$	Data $\hat{E}$
1	42	Toronto	3.697	40.326	3.962	52.562
2	40	Hamilton-Dundas	1.943	6.980	2.027	7.591
3	18	London	1.876	6.527	2.097	8.142
4	2	Ottawa	1.796	6.025	2.142	8.516
5	5	Windsor	1.738	5.686	2.126	8.381
6	35	Kitchener-Waterloo	1.534	4.637	1.659	5.254
7	8	Burlington	1.215	3.370	1.309	3.702
8	1	Brantford	1.057	2.878	1.266	3.547
9	22	Oshawa	0.811	2.250	0.864	2.373
10	15	Sarnia	0.778	2.177	1.019	2.770
11	10	Oakville	0.684	1.982	0.722	2.059
i2	39	Guelph	0.662	1.939	0.717	2.048
13	37	Niagara Fall	0.528	1.696	0.678	1.970
14	24	Woodstock	0.499	1.647	0.588	1.800
15	25	Brampton	0.448	1 565	0.450	1.568
16	6	Kingston	0.438	1.550	0.545	1.725
17	34	Galt	0.367	1.443	0.396	1.486
18	46	Sudbury-Copper Cliff	0.319	1.376	0.435	1.545
19	26	Peterborough	0.312	1.366	0.344	1.411
20	38	Welland	0.209	1.232	0.294	1.342
21	13	Chatham	0.205	1.228	0.322	1.380
22	3	St. Thomas	-0.003	0.997	0.058	1.060
23	45	North Bay	-0.069	0.933	-0.029	0.971
24	29	Barrie	-0.278	0.757	-0.313	0.731
25	11	Belleville	-0.296	0.744	-0.331	0.718
26	9	Georgetown	-0.411	0.663	-0.449	0.638
27	41	Aurora	-0.568	0.567	-0.625	0.535
29	17	Brockville	-0.574	0.563	-0.681	0.506
29	36	Preston	-0.630	0.533	-0.678	0.508
30	23	Whitby	-0.654	0.520	-0.762	0.467
31	44	RichmondHill	-0.686	0.504	-0.798	0.450
32	21	Ajax	-0.747	0.474	-0.837	0.433
33	20	Cobourg-Port Hope	-0.755	0.470	-0.863	0.422
34	14	Wallaceburg	-0.802	0.448	-0.939	0.391
35	7	Owen Sound	-0.815	0.443	-0.853	0.426
36	4	Leamington	-0.851	0.427	-0.952	0.386
37	43	New Market	-0.876	0.416	-0.980	0.375
38	32	Cornwall	-0.886	0.412	-1.029	0.357
39	19	Simcoe	-0.999	0.368	-1.116	0.328
40	12	Trenton	-1.012	0.363	-1.162	0.313
41	31	Orillia	-1.168	0.311	-1.272	0.280
42	33	Lindsay	-1.317	0.268	-1.488	0.226
43	28	Pembroke	-1.408	0.245	-1.602	0.201
44	27	Hawkesbury	-1.563	0.210	-1.957	0.141
45	16	Smith Falls	-1.713	0.180	-2.067	0.127
46	30	Midland	-2.03	0.131	-2.235	0.107

(a) In this table the population centers are ranked according to emissiveness values using adjusted data.

EMISSIVENESS ANALYSIS

Due to the limited amount of data available on population center characteristics, it was decided that linear regression analysis was the appropriate method of relating camping emissiveness indexes to population center factors. Two regression runs were made. The first run tested the relationship

(9) $Y(i) = b_0 + b_1 \ln(P(i)) + \varepsilon(i).$

WHERE

$Y(i) = \ln(\hat{E}(i))$

$P(i)$ = population of center b_0, b_1 = parameters

using all forty¯six population centers as observations.

The second run tested the relationship:

(10) $Y(i) = b_0 + b_1 \ln(P(i)) + b_2 \ln(I(i)) + \varepsilon(i)$

WHERE I(i) represents average income of population center *i* and b+ is an additional parameter. In both of the runs, emissiveness indexes obtained from Stage I using adjusted v(i,j) data were the dependent variable. The BMD stepwise regression program (a standard package being used in the 70s – see Dixon, W.J. (Ed). BMD Biomedical Computer Programs. U. of California Press, Berkeley, 1973) was employed. Table 8 contains the results of the two runs.

It is seen that population is an overwhelmingly important and statistically significant explanatory variable. Results of run 1 (with 46 observations)) and the first step of run 2 (with 30 observations are similar and indicate that in this outdoor recreation system, population accounts for about 76 percent of the total variance in emissiveness. Furthermore, it is clear that emissiveness appears to be directly proportional to population. From step 2 of run 2 it is seen that the income variable is unimportant.

A plot of $\ln(\hat{E}(i))$ versus $\ln(\hat{E}(i))$ using all 46 observations shows the functional relationship obtained from run 1 is not as good as one might hope. Extreme values are fit well, but there is still much dispersion about the predicted straight line. Examination of the residuals uncovered no unusual departures from the assumed distribution, and no sub-regional pattern of residuals were apparent.

That population has an important influence in recreation trip-making is not surprising; this result merely confirms what many have intuitively sensed over the years. It may be possible, by disaggregating population in different ways, or by including additional variables, or by using different functional forms, to account for additional variance in emissiveness. But it can be effectively argued that, based on the above results, these refinements could easily be judged unnecessary, especially in view of the crude nature of flow and other estimates used in this analysis.

TABLE 7: TWO SETS OF CAMPING ATTRACTIVENESS INDICES OR 50 ONTARIO PROVINCIAL PARKS[a]

Rank	Code #	Provincial Park	Adjusted Data ln(Â)	Â	Unadjusted Data Ln(Â)	Â
1	42	Algonquin	2.952	19.144	3.115	22.533
2	38	Killbear Poi	1.805	6.080	1.914	6.780
3	37	Grundy Lake	1.729	5.635	1.924	6.848
4	18	Pinery	1.356	3.881	1.349	3.854
5	49	Oastler Lake	1.186	3.274	1.248	3.483
6	48	Outlet Beach-	0.867	2.380	0.861	2.366
7	46	Bon Echo	0.771	2.162	0.833	2.300
8	23	Chutes	0.747	2.111	1.020	2.773
9	21	Marten River	0.730	2.075	0.990	2.691
10	39	Mikisew	0.665	1.944	0.747	2.111
11	9	Bass Lake	0.635	1.887	0.673	1.960
12	34	Presqu'ile	0.582	1.790	0.638	1.893
13	40	Restoule	0.451	1.570	0.465	1.592
14	14	Rondeau	0.433	1.542	0.446	1.562
15	36	Arrowhead	0.425	1.530	0.528	1.696
16	8	Sauble Falls	0.367	1.443	0.316	1.372
17	41	Sturgeon Bay	0.305	1.357	0.369	1.446
18	35	Serpent Moun	0.290	1.336	0.323	1.381
19	30	Six Mile Lak	0.274	1.315	0.282	1.326
20	6	Inverhuron	0.232	1.261	0.128	1.137
21	16	Ipperwash	0.199	1.220	0.214	1.239
22	27	Rideau River	0.132	1.141	0.332	1.394
23	20	Samuel de Ch	0.131	1.140	0.287	1.332
24	28	Silver Lake	0.080	1.083	0.143	1.154
25	5	Craigleith	0.080	1.083	0.102	1.107
26	31	Balsam Lake	0.049	1.050	0.000	1.000
27	33	Emily	-0.044	0.957	-0.052	0.949
28	2	Mara	-0.102	0.903	-0.099	0.906
29	26	Fitzroy	-0.128	0.880	-0.018	0.982
30	3	Sibbald Poin	-0.157	0.855	-0.155	0.853
31	50	Driftwood	-0.166	0.847	-0.039	0.962
32	17	Long Point	-0.190	0.827	-0.330	0.719
33	45	Black Lake	-0.256	0.774	-0.233	0.792
34	32	Darlington	-0.363	0.696	-0.333	0.717
35	22	Mississagi	-0.373	0.689	-0.341	0.711
36	24	Killarney	-0.376	0.687	-0.345	0.708
37	4	Earl Rowe	-0.438	0.645	-0.303	0.739
38	7	Point Farms	-0.489	0.613	-0.579	0.560
39	12	Turkey Point	-0.653	0.520	-0.790	0.454
40	47	Lake St. Pet	-0.838	0.433	-0.996	0.369
41	29	South Nation	-0.895	0.409	-0.915	0.401
42	43	Carson Lake	-0.918	0.399	-0.937	0.392
43	44	Bonnecherre	-1.158	0.314	-1.301	0.272
44	25	Windy Lake	-1.183	0.306	-1.466	0.231
45	11	Devil's Glen	-1.215	0.297	-1.326	0.266
45	13	Wheatley	-1.215	0.297	-1.447	0.235
47	15	Holiday Beac	-1.219	0.296	-1.472	0.229
48	19	Antoine	-1.349	0.259	-1.582	0.206
49	1	Rock Point	-1.551	0.212	-1.668	0.189
50	10	Selkirk	-2.203	0.110	-2.519	0.081

[a] In this table the parks are ranked according to attractiveness values using adjusted data.

TABLE 8: EMISSIVENESS REGRESSION RESULTS

Run	Step	Parameters	Std. Error	R^2	S. E. of Estimate
1	1	b0 = - 9.742	-	0.7677	0.5588
		b1 = 0.926	0.07676		
2	1	b0 = -10.689	-	0.7563	0.5794
		b1 = 1.007	0.10799		
	2	b0 = - 8.808	-		
		b1 = 1.017	0.12749	0.7566	0.5898
		b2 = - 0.229	1.38526		

Attractiveness Analysis

Data on park characteristics were somewhat more plentiful than was the case with data on population centers. Therefore, it was reasonable to employ AID to find those park characteristics that were most closely associated with differences in attractiveness. The particular park characteristics employed in the *analysis* were given in Tables 3 and 4.

For exploratory purposes, results were obtained for both adjusted and unadjusted Stage I attractiveness results. It turned out, however, that the Stage II results were nearly identical. Thus, only results from using the adjusted data are reported. The AID tree diagram for the AID analysis using adjusted data is given as Figure 2. (It is important to note that in this run the stopping rule used was such that as soon as a split resulted in a subgroup with fewer than twelve observations, no further splits were allowed.)

Figure 2 is interpreted as follows. Group 1 (G1) is the total population of 50 observations (i.e., n1 = 50). The mean value of the dependent variable, the logarithms of the Stage I attractiveness estimates, is ml = 0.035. After examining all possible binary splits for each possible "independent" variable in Table 3, the maximum reduction in the unexplained sum of squares is obtained by splitting Group 1 on the basis of "total park acreage". All those parks with 850 acres or less are in Group 2 (n2 = 35); all those parks with greater than 850 total acres are in Group 3 (n3 = 15). The mean attractiveness value for Group 2 is m2 = —0.241; the mean attractiveness value for Group 3 is m3 = 0.681. The horizontal axis of the tree diagram is scaled accordingly.

After Step 1, additional binary splits are formed by applying exactly the same procedure to Groups 2 and 3 that was applied to the original set of data, Group 1. Then the procedure is again applied to the subgroups formed at the end of Step 2; and so on down the line. The vertical axis of the tree diagram thus shows the groups that were formed at each step of the analysis. For example, Groups 4 and 5 were obtained by splitting Group 2 on the basis of "number of campsites". On the other hand, Groups 6 and 7 were formed by splitting Group 3 on the basis of "length of swimming beach". Thus while, in a way, total acreage is the most influential variable in accounting for differences in park attractiveness, the second most influential variable is "number of campsites" for parks with a total acreage of 850 acres or less and "length of swimming beach" for parks with *a* total acreage of greater than 850 acres. These results could be interpreted in a number of different ways. Most likely, a "severe/important" interaction has been detected; that is, the relative importance of an independent variable depends on the levels of its predecessors.

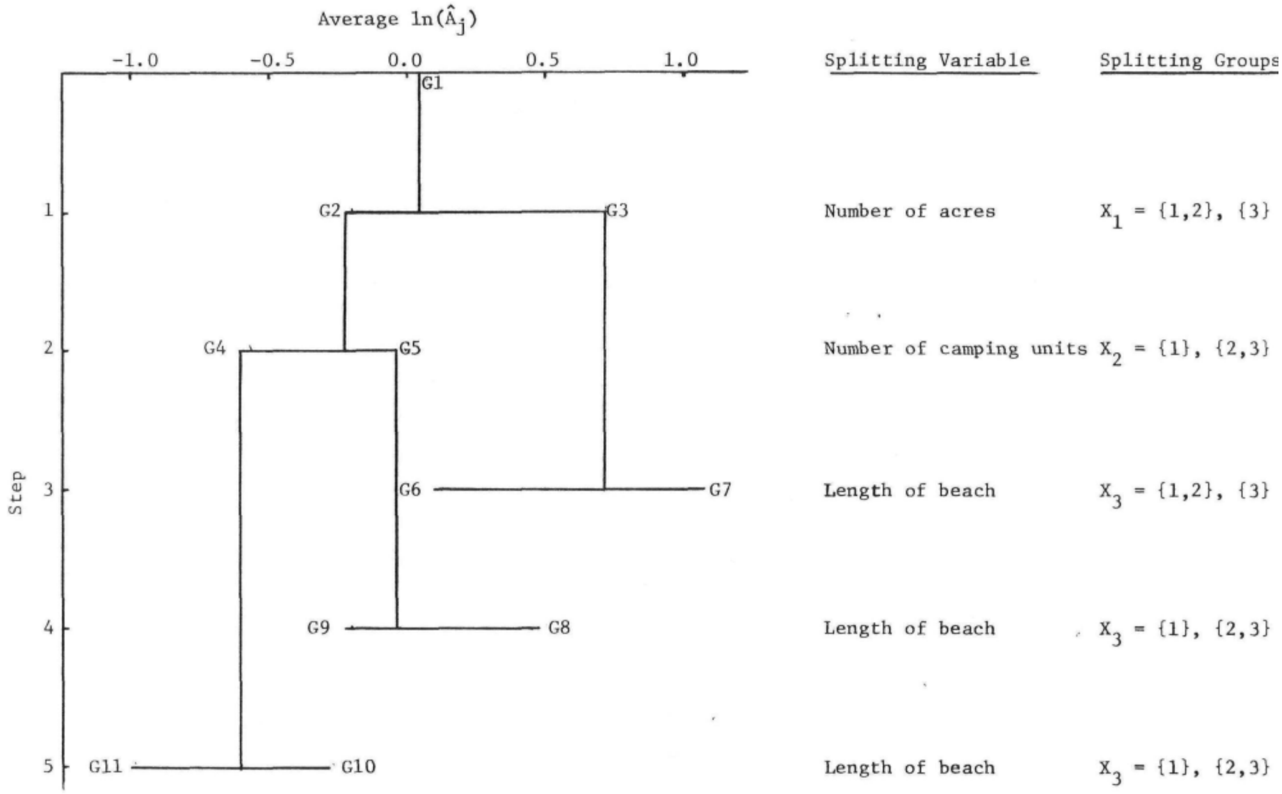

FIGURE 2. AID CLASSIFICATION TREE FOR 50 ONTARIO PROVINCIAL PARKS

A potentially important interaction is likely occurring whenever both the upper and lower branches of an AID tree split on different variables; "mild" interaction occurs when upper and lower branches split on the same variable. For example, continuing down the AID tree of Figure 2, it is seen that at high coded values of X2, the next-best variable is X3 (length of swimming beach). Similarly, for lower codes of X2 the next best variable is X3. The difference between the average value of attractiveness in group 9 (i.e., -0.224) and the average value of attractiveness in group 8 (i.e., 0.306) can be construed *as a* measure of the "main effect" of X3 at high levels of X2. The corresponding difference for groups 11 and 10 gives an estimate of the main effect of X3 at low levels of X2.

The difference *between two differences* then provides a statistical measure of the interaction between X2 and X3. If this difference of differences is negligible, then there is no interaction. If, on the other hand, the difference between average attractiveness of groups 8 and 9 *is* markedly different from the difference between the average attractiveness values for groups 10 and 11, then the interaction is “substantive”. Under suitable assumptions, the statistical significance of the magnitude of this interaction can be assessed. In the AID tree of Figure 2, the difference of the differences is equal to 0.255, and therefore there is some evidence of a mild interaction.

Figure 2 shows that *a* total of six terminal groups were formed: these are groups numbered 6, 7, 8, 9, 10 and 11. Each of the original observations belongs to one (and only one) of these terminal groups. Table 9 gives some relevant information *for each* group. The groups are arranged according to group means. Each terminal group is defined by a particular combination of codes for independent variables. For example, group number 9 of Figure 2 is seen to be defined by X1 = (1,2), X2 = (2,3), and X3 = (2,3). No other independent variables are required to define terminal group 9. Enumeration of all terminal groups in the above fashion shows that even in this simple structure, the terminal groups are defined by widely different combinations of characteristics. Table 10 presents the parks contained in each of these groups. Tables 9 and 10 illustrate a common and interesting result: two terminal groups (9 and 10) have mean attractiveness values that are approximately equal. Yet each group is defined by a different set of characteristics. Splits were obtained on only three of the thirteen independent variables. Due to the relatively small sample size, the particular stopping rule used in this analysis meant further splits, which could well be spurious, would not be obtained. It is possible to conclude, however, that the three variables for which splits were obtained are the most important discriminatory variables for this set of parks; it is conceivable that with a larger number of parks, other less important variables might be found. Of course, if the parks differed (e.g., Quebec parks were included, different variables could reflect a difference in the relation being studied).

TABLE 9: TERMINAL GROUPS FORMED BY AID CLASSIFICATION TECHNIQUE

Terminal Group #	# of Observations	Mean Attractiveness		Defining Characteristics
		Log	Transformed*	
7	9	1.071	2.918	X1=(3),X3=(3)
8	8	0.306	1.358	X1=(1,2),X2=(2,3),X3=(1)
	6	0.097	1.102	X1=(3),X3=(1,2)
9	14	-0.224	0.799	X1=(1,2),X2=(2,3),X3=(2,3)
10	7	-0.233	0.792	X1=(1,2),X2=(1),X3=(1)
11	6	-1.013	0.363	X1=(1,2),X2=(1),X3=(2,3)

*** Transformed refers to e^x x=log(mean attractiveness).**

SOME PRACTICAL APPLICATIONS

It was mentioned previously that results of successful applications of the proposed two-stage procedure for identifying important factors accounting for variations in aggregate outdoor recreation trip-making behavior would be useful to both planners and researchers.

An Application for Planners

An important park planning function is optimizing park design by establishing appropriate facilities, subject to physical, financial and political constraints. In the design of new parks, or changing existing parks, the planner is faced with choices. For any particular park she/he is, of course, interested in ascertaining the visitor implications of each of several alternative design possibilities: she/he would like to know for each design possibility how *a* park would compare with alternatives in terms of attractiveness. Results such as those provided by this analysis are ideally suited for providing appropriate information.

The AID tree shown in Figure 2 may be used to classify new observations, i.e., new or improved parks, using only values of the park's relevant independent variables. Suppose, for example, that the coded values of X1, X2 and X3 are *given* by i, 3 and *1,* respectively for *a* new park. The park might have, for example, 700 acres, 200 campsites and *a* swimming beach that was fifty feet in length. The AID tree yields an estimate of the park's attractiveness simply by determining to which terminal group the new observation belongs. Beginning at the top of the AID tree, it is seen that the left branch is taken because X1 = 1 for the new park. At the next split the right branch is taken because X2 = 2. The next branch yields terminal group 9 because X3 = 2. The predicted value of attractiveness for this hypothetical park is graphically seen to be approximately -0.224 in logarithmic terms. Thus an informed estimate of attractiveness is obtained.

TABLE 10: HOMOGENEOUS GROUPINGS OF 50 ONTARIO PROVINCIAL PARKS AS OBTAINED BY AID ANALYSIS

Terminal Group	Mean Attractiveness Log	Transformed	Park Group	
7	1.071	2.918	Algonquin	Presqu'ile
			Killbear Pt.	Restoule
			Grundy Lake	Rondeau
			Pinery	Earl Rowe
			Bon Echo	
8	0.306	1.358	Oastler Lake	Six Mile Lake
			Bass Lake	Silver Lake
			Sauble Falls	Fitzroy
			Serpent Md	Black Lake
6	0.097	1.102	Marten River	Balsam Lake
			Arrowhead	Mississagi
			Samuel de C	Killarney
9	-0.224	0.799	Outlet Beach	Sibbald Pt.
			Mikisew	Long Point
			Invenhuron	Darlington
			Ipperwash	Point Farms
			Rideau River	Turkey Pt.
			Emily	Wheatley
			Craigleith	Selkeirk
10	-0.233	0.792	Chutes	Carson Lake
			Sturgeon Bay	Devils Glen
			Mara	Antoine
			South Nation	
11	-1.013	0.363	Driftwood	Windy Lake
			Lake St. Peter	Holiday Beach
			Bonnecherre	Rock Point

If no information were available on a park's characteristics, compared to the characteristics of other parks that give rise to differential attractiveness values, the best estimate that could be made would be to assume the average value of the total sample, which from Figure 2 can be seen to be equal to 0.035 (in logarithmic terms). This is, in a sense, is an "unconditional" predicted value. When the unique characteristics of the new observation are used in the assessment, the "conditional" prediction thereby obtained represents a substantial refinement over an unconditional prediction. In this particular case it is seen that a park with the characteristics cited above is less attractive than the average park. Translated into absolute numerical scales, the implied difference in attractiveness magnitudes is about 0.8:l. Interpreted, this implies that (all other things being equal) the new park would draw about eighty percent of the visitors that the *"average park"* would draw. Planning can be performed more effectively by making use of such information.

An Application for Model Building

Several approaches for relating park use to the factors which "explain" it were alluded to in the Introduction. It was noted that researchers have attempted to construct models which use as dependent variables various measures of park use and as independent variables various suspected influencing factors. Independent variables that are specific to population centers and to parks are, of course, generally considered to be very relevant. But, due to the virtually infinite number of *possible* combinations of variables that can be hypothesized to characterize population center and park effects on outdoor recreation park use, the selection of particular variables to include in models is a difficult undertaking. As a result, subjective ad hoc selection procedures are often adopted. The result provided by this study can be viewed as providing a somewhat objective basis for selecting variables.

Viewed in this general sort of way, the overall results of the study can be taken as being supportive of previous modeling efforts that have used population and park size-related measures as independent variables. It was found, for instance, from regressions on the $\hat{E}(i)$'s that population size by itself is a very good indicator of population center emissiveness (i.e., visit generation) and from AID analysis on the $\hat{A}(j)$'s that acreage and other size-related characteristics of parks are important indicators of attractiveness (i.e., popularity). These measures may now be used with some justification whereas before their use was based strictly on what might be termed "informed judgment".

Concentrating on the AID analysis, attractiveness results, in particular, can serve to lead future research and model-building investigations into potentially fruitful directions. For example, regarding the AID classification tree of Figure 2, it was pointed out that both mild and important/severe interaction existed in the system under study, and that no unique set of variables would be expected to suffice for the modeling of visitor flows to all parks. On the basis of these findings the researcher might proceed to develop separate regression models for observations in each terminal box and thereby obtain coefficients with which to judge the relative partial effects of the associated independent variables on each subgroup.

GENERAL CONCLUSIONS

The objective of the analysis was to test a proposed methodology for drawing certain inferences about important factors influencing park usage. The methodology proved to be *a* promising way of exploring emissiveness, attractiveness and location.

As far as specific numerical findings are concerned, two results became clear. Evidence was provided that population size is by itself a useful indicator of participation differences in the region under study. There is thus some justification for using a population measure as a basis for park use models.

On the other side of the coin, park size was seen to be very important factor in attracting trips to *a* particular recreation area. It was shown that large parks attract large numbers of people relative to small parks, all other things being equal. Further, the number of camping sites appears to be a very important variable for small parks, whereas length of swimming beach is only important in large parks. It can be concluded that size-related measures are useful as indicators of recreation site attractiveness in modeling efforts.

Advantages and Limitations

The new two-stage methodology, while having many advantages over existing methodologies that attempt to perform *similar* analyses, is not without its limitations. To properly assess its ultimate usefulness, it would be necessary to conduct other tests. Nevertheless it is possible to systematically list some of the advantages and disadvantages.

The proposed procedure is, in a sense, economical. First, data requirements for the first stage are quite meager - only matrices (v(i,j)) and (d(i.j)) are required. And if significant differences do not appear in the estimates (Ê(i)) and (Â(j)) after application of Stage I, the second stage of the analysis need not be undertaken, consequently effecting a savings in data collection costs since population center and/or park characteristics data would not be required.

It is clear that classification is regarded as a central requirement of the approach tested in this report. In this respect, the AID technique used in Stage II has many advantages over other methods of analysis in that it determines the joint effects of many variables by analyzing variance between grouped data by making the very minimal number of statistical and algebraic assumptions. Other methods such as regression analysis require very strong assumptions about data that are possibly poorly approximated in many analyses of non experimental data. Other advantages of AID, such as the detection of "mild" and "important/severe" interactions, have already been mentioned.

By far the most useful feature of the methodology under test here is the systematic extraction, from observed data on recreation choices, of factors influencing park use. That is, the underlying assumption is that these factors are revealed by people who make observable choices in their recreational pursuits. This "deductive" sort of reasoning stands in contradistinction to the "inductive" reasoning that has characterized many previous modeling studies.

There are, of course, many real limitations of both stages of the analysis. In the first stage, transformation of the model (Equation 6) into logarithmic form produces biased estimates of its parameters. Estimation in Stage I by Equation 8 involves making adjustment when there are many zeros in an observed spatial interaction matrix since the logarithm of 0 does not exist (e.g., adding a small amount to zero flows). Finally, due to heteroscedasticity problems (variance of large flows differing from the variance in small flows – GLS is used to address this) hypothesis-testing procedures can yield deceptive results (when estimation is based on assuming all flows have the same variance, homoscedasticity). The potential user of this methodology should take these considerations into account.

With regard to the second stage of the analysis, the AID classification technique produces results that are not easily interpreted in the analysis of trip-making systems of small dimension. And, since the method produces a sequence of dichotomous splits, only the best "local" split is obtained at each step of the sequence. Thus a series of locally optimal decompositions is obtained where each split is contingent upon the previous one (as mentioned above). It is possible that a variable "far down the line" could be almost as discriminating as one involved in an early split. Had chance resulted in the program splitting observations differently in the beginning, a different AID tree would have resulted (one addresses this problem by running the analysis several times with judicious elimination of variables). The codification of independent variables for AID analysis must also be done judgmentally. Overly coarse coding loses information.

Recommendations for Further Research

All in all, the methodology described in this report appears promising. Only by applying the method to other situations and comparing results with alternative methods can its usefulness be ascertained. The following recommendations for further study can be given:

1. This pilot study concentrated attention on only one region in Canada. Much can be learned

about the generality of the results by applying the two-stage methodology to analyze corresponding outdoor recreation travel patterns in other regions.

2. This pilot study concentrated attention on one visitor type - main destination campers. It would be interesting to compare these results with outdoor recreation travel parameters estimated for the other relevant visitor- type (main destination day-users) in the same region.
3. This pilot study used data at its highest level of aggregation - the total visitors traveling from i to j during some long time period. It is possible to explore, by use of the proposed methodology, travel patterns of various population subgroups. For example, it is theoretically possible by use of CORD Study data to use various characteristics, such as *age* and income to define groups for which models would be developed. Parties travel and do not have an age so it is easier to see using travel party groupings. It *is* also possible to pinpoint dates of travel so weekend of long weekend travel, long-medium-short stay and other classifications can result in models with different parameters. Understanding of outdoor recreation travel patterns might be enhanced on the basis of some of the above identifiable characteristics.
4. The Stage I analysis produces biased estimates of parameters. Techniques that are not fraught with the many problems of using an analysis of covariance on a transformed version of the postulated model are needed (see Cesario 1974). In this regard, change in computing power and estimation methods means that non-linear regression is a powerful tool in the new millennium (e.g. see SAS Proc NLIN reference earlier).
5. Stage II results using the AID technique as a statistical tool would call for large numbers of observations. Its usefulness in the outdoor recreation context is therefore jeopardized because the number of parks associated with reasonably large flows from a set of cities is limited. Nevertheless it is desirable to increase the data base for making estimates by e.g. getting flow data and park relations to cities for e.g., all of Canada. If there is a common structure across Canada, it could be detected and thus knowledge increased.
6. The two-stage methodology presented in this report was used to uncover the structure of some existing data and, second, interpret this structure so that we may come to an understanding of how it might have come about. Other methodologies, some of which were mentioned in the Introduction, purport to answer some of the same questions. (To this author's knowledge no other procedure addresses all of the same questions). A useful exercise would involve a comparison of corresponding results obtained from this study with those obtained from applying other methodologies to the same or similar data.