TECHNIQUES FOR COST/BENEFIT ANALYSIS

The ultimate goal of all safety projects is to reduce accidents in the most cost-effective manner. The use of the incremental cost/benefit ratio and of dynamic and integral programming is appropriate for selecting safety improvements that will optimize safety benefits for every dollar spent.

Dynamic programming is an optimization technique that transfers a multistage decision problem in a series of one-stage decision problems. It is used to allocate money to obtain the maximum possible benefits under a fixed-budget. The three possible levels of dynamic programming are:

1. Single-stage (used to evaluate a single project with several alternatives),
2. Multistage (involves selection among several projects with several alternatives each), and
3. Multistage with a time factor (used where several alternatives are considered and various time periods are involved).

Basic input into the dynamic programming model consists of (a) initial costs and maintenance costs for each project alternative, (b) accident benefits for each project alternative, and (c) budget available for improvements. Dynamic programming can take advantage of greater benefits by choosing a project that will yield greater benefits, even though it shows a lower cost/benefit ratio than another alternative. Therefore, expenditure for a group of projects chosen by dynamic programming can yield greater dollar benefits than expenditure for the same projects chosen through the cost/benefit technique. In some instances, the same projects will be selected. The optimal selection of projects is sometimes fairly obvious by using manual techniques; however, where many projects are involved, computer analysis must be used.

USING DATA ANALYSIS AND REPORTING TECHNIQUES (DART) FOR PROBLEM IDENTIFICATION

John W. Larmer, GENASYS Corporation

The Data Analysis and Reporting Techniques (DART) system was developed by GENASYS Corporation under a series of NHTSA contracts. It is a computer software statistical system that was specifically designed to assist in the problem identification process for the acquisition, selection, and analysis of state accident data.

The objective of the DART system is to overcome the lack of integrated data that resulted from poor communication between state agencies. Historically, a state agency designed a traffic records system to meet its particular needs. Other agencies requiring the same information were rarely consulted. As a result, several components of the overall traffic records system were developed, and each was organized and operated differently. This presented a basic problem in the development of an integrated system.

DART was thus developed in an effort to deal with these partially integrated traffic records systems. While there was little doubt that a tool for analysis was needed immediately within the highway safety management program, it was unlikely that an integrated traffic records system would be a reality within the near future. DART was designed to address this problem; consequently, it allows for upgrading of the traffic records system as integration capability is achieved.

An initial decision was made to create a subsystem that would require information from all the other data subsystems. Thus, the safety analysis and reporting system required group analysis of accidents so that the problem identification process could be implemented. This emphasis on problem identification has placed the burden on states to develop better data, better records systems, and better data analysis capabilities. However, dependence on accident records systems has revealed the inadequacy of many state systems to support a coordinated problem identification process. The analytical techniques used by many states are merely elementary analyses—an indication of a potential problem.

The use of DART in the problem identification process is one way to improve a state's analytical capability. Problem identification is simply an iterative process of comparative analyses that narrows the potential problem populations until the true problem is identified. The effectiveness of countermeasures can be evaluated by performing the same analysis after a period of time has passed.

Analysis through use of the DART system must be performed on a static data file. At least one full year—ideally multiple years—of data is used to develop trends and to evaluate the impact of ongoing programs.

The first step is to produce a full file univariate that will show areas of high accident involvement. A full file univariate is a one-way frequency distribution of each data element present on the data file. The file also reduces each subset of all data elements to a relative percentage or weight in the overall population under analysis. High involvement does not necessarily mean that a problem has been identified; this step in the process is designed to lead the researcher on into comparative analysis.

The next step in our problem identification process is to isolate the potential problem groups highlighted on the univariate and to look at attributes that may contribute to their accidents. The researcher may decide to break the groups down further—for example, into males and females within specific counties. The selection process allows a comparison of county versus county, male versus female, etc., by any attribute selected and in any combination. The DART output of such a comparison of accidents would be a table showing these figures.

The frequency of accidents by each age group can then be compared, and any discrepancies can be determined. Each age group is reduced to a common denominator, that is, to its relative weight percentage within the population. By using this common denominator as the basis for comparison, it can be surmised that potential problems rest within different age groups by sex. The DART report would be (a) a table showing both accident frequencies for males and females in each age group and the percentages representing those frequencies, and (b) a graph comparing the percentages. The comparison highlights overrepresentation in various age groups.

The process has thus far succeeded in illustrating vast differences between males and females by age group. In order to establish what can be expected, these facts must now be compared with normalizing data. Normalizing data are external factors used to determine the appropriate weight to be given to the accident figures for each sex and age group. If further analysis is to be based on exposure, one criterion that may establish exposure to accidents would be the number of licensed male and female drivers. The computer would be asked to supply a report on the number of licensed drivers by sex and age.

In the next step, the number of licensed drivers by sex and age is reduced to a weight (percentage) within the population 16 to 30 years of age. The percentage of accidents involving males can then be compared against that of male licensed drivers, and the percentage of accidents involving females against that of female licensed drivers. The DART graphs showing these comparisons would indicate the extent of overrepresentation in particular age groups with respect to exposure.

The main thrust of this presentation was to illustrate that DART is a tool that can be used efficiently with accident and related exposure data. The use of such a tool can upgrade the skills of the analyst and uncover areas of data collection that need improvement. The challenge is to broaden the awareness, acceptance, and use of tools such as DART for highway safety management.

USE OF RAPID FOR PROBLEM IDENTIFICATION AND EVALUATION

David B. Brown, Auburn University, Alabama

The RAPID (Records Analysis for Problem Identification and Definition) system is a user-oriented computer system specifically designed to process state accident data. As opposed to a standardized report generator, RAPID enables the user to specify informational needs. By responding to simple queries, the user is guided through the process, interacting directly with the computer. RAPID has two specialized software modules: (a) ACT, which automatically generates priorities among cities within population subgroupings for any user-defined subset of accidents, and (b) AIM, which finds high-accident concentration areas on the roadway for any user-specified accident type. RAPID also draws on the resources of the Statistical Package in the Social Sciences (SPSS), automatically furnishing the appropriate SPSS labels, codes, and ranges, as well as all control statements and format specifications.

No computer knowledge is required to use RAPID. Although RAPID uses a standard statistical package to produce output, the user does not need to understand how to assemble statistical control statements. This is handled entirely by the RAPID system.

RAPID has been installed for Alabama, South Carolina, Kentucky, Tennessee, and Delaware. It is a portable package that can be applied to any state's accident data as well as to other types of data where statistical processing is required. It is available on commercial time-sharing systems if state resources do not permit in-house installation.

RAPID provides the user with the following capabilities:

- Create a subset, including variables, of the master data base according to any logical specification. Subsets requested could include all pedestrian accidents, alcohol-related accidents in a given county or city, or motorcycle accidents between milepost 235 and 240 on Interstate 85.
- Obtain labeled univariant frequency distributions for the variables chosen to be included in subsets. The production of total statewide univariant distribution for all variables falls within this capability.
- Obtain labeled histograms of frequency distributions.
- Obtain fully labeled bivariant (crosstab) analyses for any of the subset variables.
- Perform up to eight levels of multivariant analyses for any subset produced. For example, in the three-level analysis, a crosstab of accident time of day by day of the week could be produced for all severity levels.
- Obtain a correlation table for all combinations of subject variables.
- Find high accident locations according to user-specified criteria. (Locations are specified by road codes and mileposts.) The interactive nature of this task enables the user to try any number of alternative criteria in order to obtain the number and type of high accident locations the user needs to work with.
- Obtain univariant distributions of any or all variables for the locations found to be high accident locations. (The same capability also exists for any other location specified by the user.) A separate report is produced for each high accident location and for each accident. This condenses the information for ease of review before location investigations.
- Obtain any of the reports specified by the above capabilities for any or all of the high accident locations.
- Create a logically restricted subset from any previously created subset. Because the user can obtain many different logical restrictions (for example, composite geographical areas) from a subset without rereading the master data base, which is usually stored on tape, computer time can be greatly reduced.
- Integrate demographic information and thereby establish priorities among political subdivisions for various accident types. For example, RAPID produces priority lists for cities by population grouping according to the number of motorcycle accidents divided by any one of several demographic indexes, such as population, miles driven, or number of registered motorcycles.
- Obtain further statistical analyses (RAPSTAT), including analysis of variance, breakdown analysis, regression analysis, scatter diagrams, and a variety of student's t-test options.
- Obtain accident report numbers for any subset of accident records so that hard copy for particular types of accidents can be retrieved.

The RAPID system can be best explained by tracing the data from the origin to the final output report.

When a pedestrian accident occurs, an officer in the field records the accident on a standardized form, which is sent to a central point for data entry. Along with thousands of other records, it becomes part of the state's accident records data base, which is generally stored on tape.

The accident records data base is generally not constructed with problem identification in mind. In fact, it contains virtually all of the "codable" elements from the accident records. Many of these are not required for problem identification work, and they are generally not in a form compatible with problem identification. For example, the pedestrian's actual age is probably coded on the tape, whereas age intervals (e.g., 0-4, 5-7, 8-9, 10-15, 16-21, etc.) would be much more useful for problem identification and cross tabulation. In addition, certain calculations and other data manipulation might be required to satisfy the requirements of problem identification. For these reasons, the data base must be cleaned up before it can be used for problem identification. This may be done once a year for the data base compiled from the previous year. The program that reformats and puts the data elements into their proper intervals is known as the BASE program.

The BASE program is then run to create a new tape, the RAPID master data base, which is totally compatible with RAPID formats and objectives. RAPID can work on any properly formatted data base. The arrangement, number, and type of variables are totally flexible and may be specified by the user during the development of the BASE program. Since the new, properly formatted tape is too large to generate statistical reports efficiently, a subset of the RAPID master data base can be created on highway speed direct access storage whenever processing is required. This subset can be either retained for repeated use (cataloged) or used immediately and deleted.