

in the process of editing tabulations for publication. These errors are generally corrected at the levels of aggregation at which data are published, but often are not carried back to the basic records for the individual respondent unit, in machine sensible form. Carrying back such corrections to the basic files is prerequisite to the kind of analysis of interrelationships which is here proposed. Some method of insuring that this is done, and that the basic records meet appropriate standards as archives, is essential. Such tasks generally receive a low priority in the statistical agency whose primary task is production and publication, rather than analysis, of data.

As a technical device, the use of master samples can achieve a high degree of unification of information about the individual reporting unit. Moreover, proper experimental designs provide for the straightforward estimation of sampling variances; differences among subpopulations can be measured with specified precision if this requirement is stated in advance; variances due to differences among samples can often be eliminated; costly matching of units at the later stages can be avoided. Inconsistencies arising from many kinds of differences between surveys can also be avoided.

The use of the current population survey for special questions (veteran status, duration of unemployment, preferred number of hours, work, etc.) permits a number of useful comparisons with standard information about labor force status. Similarly, it may often be practical to use master samples to obtain information about subjects vital to some of the newer Federal welfare programs. Hence, an integrated system of master samples of households could be used to collect information about income, education, health, crime, employment, social services, housing, demography, voting registration, and the effects of opening or closing industrial plants. Not only could information be compiled about each of these subjects, but analyses could be performed which inter-related several of the subjects: education-income-crime rates, health-housing-education, etc., without loss of information or the introduction of uncertainty arising from variances between samples or from different survey practices. Indeed, the judicious use of master samples can lead economically to conformity with the general principles stated above.

The fourth principle is ability to produce any information computable from the basic records. The principle acquires new power when combined with the other three principles, because a great deal more becomes computable. The concept of "inherently computable" is taken literally, and includes kinds of statistical operations not now widely used. The paucity of current use derives from unfavorable economics, unsuitable organization of data, insufficient available detail, failure to use known techniques, obstacles growing out of confidentiality restrictions on data intended for input to the analysis, and current habits and practices deriving from all of the other obstacles. In short, the current ways of doing business fall far short of the potentiality of advanced statistical techniques applied to a well-organized body of Federal data. The present report suggests the means for mitigation or elimination of the shortcomings of the statistical system built before modern computers became available.

In these terms, the notion of "inherently computable" takes on new meaning. Obviously included are the routine computation of averages, cross-tabulations, correlations, curve fittings, time series analysis, seasonal adjustments, distribution statistics, and the application of other techniques of mathematical statistics. But it would also now be possible to test the reconciliation of one series of data against others. Test adjustments of all sorts, even very detailed and burdensome adjustments involving manipulation of very large matrices, could be countenanced. Errors could be studied, including those for whose estimation there is little theoretical foundation—the myriad kinds of inconsistencies of definition, practice, error rate, personnel, etc.—when data from two or more sources are used. Propagation of errors through the system, especially in the major synthetic series (national income and product accounts, Federal Reserve Board production indexes, price indexes) could also be studied and estimated. One set of objectives would be error detection and measurement in the primary collections. Another would be consistency testing and anomaly detection in two or more collections from different agencies, geographic regions, time periods, etc. Computations of this sort could also be used to assist in the setting of quality standards for Federal data and for validation or certification of particular bodies of data.

There is already a praiseworthy trend toward the use of computers in the production of standard series of data by several agencies. The notion of "inherently computable" includes the generation of new series for special purposes