HUDAP

Hebrew University Data Analysis Package

Computation Authority, the Hebrew University of Jerusalem

Hudap Package is a data analysis software which contains mainly data analysis techniques based on methods introduced by Professor Louis Guttman. These methods are essentially *intrinsic*, namely, the data is treated in terms of internal inequalities without need for an explicit model.
The main modules in the package are:

- **MONCO**: Weak Monotonicity Coefficients (correlations)
  This coefficient, designated as μ_2, expresses the extent to which replies to one question increase in a particular direction as the replies to the other question increase, without assuming that the increase is exactly according to a straight line. This coefficient varies between -1 and $+1$. $\mu_2=+1$ implies a fully monotone relationship with positive or rising trend; and $\mu_2=-1$ implies a fully monotone relationship which is of negative or descending trend. *Weak* monotonicity is meant here. Ties in one variable may be untied in the other without penalty.
  μ_2 can equal $+1$ or -1 even though the marginal distributions of the two variables differ from each other. Therefore a weak monotonicity coefficient is especially appropriate in condition where marginal distributions differ from item to item, as is the case in most research studies.

- **WSSA1**: Smallest Space Analysis (better called Similarity structural Analysis)
  This is a practical technique for aiding comprehension of the structure of interrelationships among variables. The analysis may be regarded as being essentially geometric. SSA treats each variable as a point in a Euclidean space in such a way that the higher the correlation between two variables, the closer they are in the space. The space of smallest dimensionality is used that makes possible such an inverse relationship between the observed correlations and the geometric distances. The empirical data to be anlysed are not limited to coefficients of similarity. They can be also dissimilarity coefficients (distances). In such a case, the monotonicity condition becomes: the smaller dissimilarity coefficient between variables, the closer their points are in the space.
  Subgroups of the population can be superimposed into the fixed SSA space diagram of the content variables via external variable feature.

  WSSA1 with Facets:
  There is an option to superimpose facet elements for each variable, and produce facet diagrams. This option facilitates viewing regional correspondence between the empirical distribution of the variables and their faceted definition. In the graphic representation, each element in each facet diagram is labeled by a different color.
  Moreover the user can request, interactively, regionalization of the variables for Axial, Modular or Polar models. Various options are offered to the user. In Axial model, the lines can be parallel or not. In Modular model, the curves can be circles or ellipses, concentric or not. The user can choose an initial or fixed *center* (for Modular or Polar models), *direction* (for Axial model).

- **POSAC**: Partial Order Scalogram Analysis with base Coordinates
  While SSA deals with space of variables, POSAC deals with space of subjects.
  A partial order analysis begins with some *n* different criteria (items) for the stratification of

the population, each item being ordered in a sense common to all items.
Each member of the population has an observed profile composed of *n* structs, one struct from each item. The overall partial-order is an automatic consequence of the simple orders on each item separately.
By definition, a profile is higher than another, if and only if it is higher on at least one item and not lower on any other item. Similar definition holds for “lower” relation. A profile is equal to another if both are equal on all items. When the relation between two profiles is “higher”, “lower” or “equal”, these profiles are said to be *comparable*. Two profiles are *incomparable* if and only if one profile is the higher on at least one item while the other profile is also the higher on at least one item.
The complete `POSAC` output provides a two-dimensional representation of the partial order of the set of profiles.
`POSAC` calculates a mathematically optimal pair of base axes (x and y) for the empirical partial order. These base coordinates usually have a substantive meaning for the partial order, though there need not always be items that correspond to the directions of these base coordinates.
Each profile appears as a point in the two-dimensional space. Any two profiles that are comparable will have their two points on a line with a positive slope, namely the joint direction ($x+y$). Two incomparable profiles have their points on a line with a negative slope or the lateral direction ($x-y$). All four directions in the two-dimensional space (x, y, *joint* and *lateral*) have a role in interpreting the results.
Detailed analysis of the systematic differences among the items is made in terms of *n* `POSAC` diagrams, one for each item.
Furthermore, for specified categories of a given external variable (an item not in the original set of the *n* items) representing an external criterion or trait, trait-diagrams are presented depicting the proportion of subjects possessing that trait among all those sharing the same profile of *n* items. The trait may be specified also as a combination (intersection) of response categories from different external variables.
As in WSSA1, the user can request, interactively, regionalization of the profiles for each item diagram. Various options are available. The lines can be parallel or not. The user can choose an initial or fixed line *direction*.

MPOSAC

`POSAC` was technically limited to two dimensions, until recently when a new algorithm allowed for processing `POSAC` in any dimensionality. The related computer module is termed `MPOSAC`, namely Multidimensional `POSAC`.

- **DISCO**: **Discrimination Coefficient** (Effective ANOVA).
When two or more populations have distributions on the same numerical variable x, it is of interest to know to what extent these distributions overlap. One motivation for this interest is the problem of discriminant analysis.
Guttman presented two discrimination coefficients for this purpose without making any assumptions as to how their ditributions may differ in other respects. Both coefficients express the loss due to overlap as a direct function of the variance between the arithmetic means of the distributions. One coefficient is called *disco* for "discrimination coefficient". The other is called *odisco*; it is more relaxed than *disco* in a certain sense of overlap. The "*o*" at the beginning of *odisco* is meant to indicate that some overlap is allowed. Each is distribution-free, avoiding unrealistic assumptions of normality of population distributions and equality of variances within the populations.
Both coefficients vary between 0 and 1. They equal 0 if there are no differences among the means. Each equals 1 if perfect discrimination holds in its sense. Disco equals 1 when there is no overlap between the distributions. Odisco equals 1 when there is overlap only

between the means of the respective distributions. Such efficacy coefficients are always appropriate since consistent estimation is ensured.

- **INTRACLASS** Correlation
  In some cases involving pairs of measurements there is no way to distinguish between the pairs. For example, given IQ measurements of a pair of identical twins, how can it be decided which measurement is *x* and which is *y*? In such cases a different type of correlation coefficient, called the intraclass correlation coefficient which treats the pairs of measurements symmetrically, is needed to asses the relation between them.
  Moreover, this coefficient is not limited to pairs of variables. It can be defined on more than two variables. In this context, a case or observation is called a *class*. Each class, has a number of members represented by *scores*. The classes can even have unequal number of members.

Hudap contains also some descriptive statistics as frequencies and crosstabulations. It allows the recoding of defined variables as well as computation of new variables. Looping is a nice feature of the package.

A Windows version of Hudap is available. In this version the user can choose one of two modes:
- **Edit Mode** : the user writes the job in Hudap language in the Editor area and then submits it to the system.
- **Assist Mode**: knowledge of Hudap language is not required. Instead various menus and windows drive the user to process the main Hudap analyses.

This version includes also two and three-dimensional graphic representation of space configurations with interactive rotations.

A manual (as a PDF file from Help menu) accompanies Hudap. This manual contains explanations on Hudap language and a detailed documentation on each module with examples.
The mathematics of the methods as well as Windows interface examples are given as appendices at the end of the manual.

The price of Hudap Package is $350.

To purchase Hudap, a purchase order must be sent to Reuven Amar by Mail or by Fax.

For assistance and consultation on the practice as well as for further information, please contact Reuven Amar

Address:
Computation Authority
The Hebrew University of Jerusalem
Mount Scopus, Jerusalem, Israel

Electronic mail: reuben@mscc.huji.ac.il

Phone: 972-2-5881088
Fax: 972-2-5828826