Dataset Documentation:

Multivariate Data Analysis, Seventh edition

"The world's leading authority on applied multivariate data analysis based on number of citations, as reported by Google.Scholar."

A number of datasets are available to enable students and faculty to perform the multivariate analyses described in the textbook. While some techniques require specialized datasets (e.g., multidimensional scaling, conjoint analysis and structural equation modeling), many of the techniques are performed using conventional survey data.

HBAT

HBAT is a common dataset developed for use with many of the techniques to allow students to see the interrelationships among techniques as well as the techniques themselves. The HBAT dataset has several forms utilized throughout the text:

- **HBAT** – the primary database described in the text which has multiple metric and nonmetric variables allowing for use in most of the multivariate techniques.

- **HBAT_200** – an expanded dataset, comparable to HBAT except for 200 rather than 100 respondents, that allows for multiple independent variables in MANOVA while still maintaining adequate sample size in the individual cells.

- **HBAT_MISSING** – a reduced dataset with 70 respondents and missing data among the variables. It is utilized in illustrating the techniques for diagnosis and remedy of missing data described in Chapter 2.

- **HBAT_SPLITS** – contains two variables that split the HBAT dataset into 50/50 and 60/40 subsamples. This dataset can be merged with the original HBAT dataset if desired.

Other HBAT Datasets

In addition to these primary datasets, there are several other datasets used with specific techniques, including conjoint analysis, multidimensional scaling and structural equation modeling. These datasets include:

- HBAT_CPLAN and HBAT_CONJOINT – the datasets to perform the "full profile" conjoint analysis available in SPSS. HBAT_CPLAN details the stimulus profile descriptions and HBAT_CONJOINT contains the actual responses to the stimulus profiles.

- HBAT_MDS, HBAT_CORRESP and HBAT_CORRESP_INDIV – the datasets used for the multidimensional scaling and correspondence analyses in the text.

- HBAT_SEM – the original data responses from 400 individuals which are the basis for the structural equation analyses of Chapters 10, 11 and 12. This dataset can be used to derive the covariance matrices used as input to structural equation programs such as LISREL, EQS or AMOS.

- HBAT.COV, HBATF.COV and HBATM.COV – These three covariance matrices represent the overall sample, female respondents and male respondents, respectively. They are used in the analyses described in Chapters 10, 11 and 12.

- HBAT_SEM_NOMISSING – the original dataset of 400 responses has two individuals with missing data. To facilitate SEM analysis without having to address issues of missing data, this dataset replaces the missing values so that the resulting sample is 400 complete responses. It can be used to perform any SEM analysis, although it should be noted that very small differences from the results obtained with HBAT_SEM and reported in the text may occur due to elimination of these two cases.

Other Datasets

Finally, two additional datasets are provided to allow students access to data other than the HBAT data files described in the textbook:

- HATCO – this dataset has been utilized in past versions of the textbook and provides a simplified set of variables amenable to all of the basic multivariate techniques.

- SALES – this dataset concerns sales training and is comprised of 80 respondents, representing a portion of data that was collected by academic researchers.

Dataset Formats

Given the widespread interchangeability of data formats among statistical programs, all of the datasets are provided in two formats. First is the .SAV format used in SPSS, which allows for documentation of variable descriptions, etc. in a standard format. Also, all of the datasets are contained in an EXCEL workbook, amenable to input to any statistical program.