Random Intensity Datasets
Brian T. Luke (lukeb@ncifcrf.gov)

These 30 datasets contain 300 features and between 30 and 300 Cases and Controls. The datasets are constructed with random peak intensities so that they contain no biological information. Structure of the Datasets contains a general description of datasets that can be used by programs within the BioMarker Development Kit (BMDK). Since the Cases and Controls are stored in different files, the class indices are not included in the data. Each feature has a single label, but they are simply "F-00001" through "F-00300". Each dataset has an associated document that describes the results of an analysis using the BioMarker Development Kit (BMDK), and classifiers based on a decision tree (DT) and a medoid classification algorithm (MCA). To reduce the amount of repeated information in these tables of results, Description of the Tables gives details about each table.

Since all 30 datasets have 300 peak intensities, the first step was to set a maximum intensity for each peak. The maximum intensity for each peak was set to a random number between 0.0 and 200.0. The smallest peak (peak 64) had a maximum intensity of 1.055, while the largest peak (peak 131) had a maximum intensity of 197.9. A different seed to the random number generator was used for each dataset so that the first sample, for example, had a different intensity for each peak in each dataset. Since the average maximum intensity was approximately 100.0, the average intensity across all peaks for a sample was approximately 50.0. To ensure that no sample varied significantly from this average, each samples spectrum was scaled so that the sum of all peak intensities was exactly 15000.0.

Each dataset was constructed to contain the same number of Cases and Controls (30, 42, 60, 90, 150, and 300 Cases and Controls). For each number of Cases and Controls, a total of five random datasets were constructed, producing 30 unique datasets. For each spectrum in each dataset, peak 64 should have one of the lowest intensities, but it is possible to have a lower intensity in another peak, even peak 131, since the intensity is set to a random value between zero and the maximum allowed.

Analysis	#Cases #Controls	#Features	Case Dataset	Control Dataset	Analysis
Random_Intensity_30_1a	30	300	case_30_1a.txt	control_30_1a.txt	Tables
Random_Intensity_30_2a	30	300	case_30_2a.txt	control_30_2a.txt	Tables
Random_Intensity_30_3a	30	300	case_30_3a.txt	control_30_3a.txt	Tables
Random_Intensity_30_4a	30	300	case_30_4a.txt	control_30_4a.txt	Tables
Random_Intensity_30_5a	30	300	case_30_5a.txt	control_30_5a.txt	Tables
Random_Intensity_42_1a	42	300	case_42_1a.txt	control_42_1a.txt	Tables
Random_Intensity_42_2a	42	300	case_42_2a.txt	control_42_2a.txt	Tables
Random_Intensity_42_3a	42	300	case_42_3a.txt	control_42_3a.txt	Tables
Random_Intensity_42_4a	42	300	case_42_4a.txt	control_42_4a.txt	Tables
Random_Intensity_42_5a	42	300	case_42_5a.txt	control_42_5a.txt	Tables
Random_Intensity_60_1a	60	300	case_60_1a.txt	control_60_1a.txt	Tables
Random_Intensity_60_2a	60	300	case_60_2a.txt	control_60_2a.txt	Tables
Random_Intensity_60_3a	60	300	case_60_3a.txt	control_60_3a.txt	Tables
Random_Intensity_60_4a	60	300	case_60_4a.txt	control_60_4a.txt	Tables
Random_Intensity_60_5a	60	300	case_60_5a.txt	control_60_5a.txt	Tables
Random_Intensity_90_1a	90	300	case_90_1a.txt	control_90_1a.txt	Tables
Random_Intensity_90_2a	90	300	case_90_2a.txt	control_90_2a.txt	Tables
Random_Intensity_90_3a	90	300	case_90_3a.txt	control_90_3a.txt	Tables
Random_Intensity_90_4a	90	300	case_90_4a.txt	control_90_4a.txt	Tables
Random_Intensity_90_5a	90	300	case_90_5a.txt	control_90_5a.txt	Tables
Random_Intensity_150_1a	150	300	case_150_1a.txt	control_150_1a.txt	Tables
Random_Intensity_150_2a	150	300	case_150_2a.txt	control_150_2a.txt	Tables
Random_Intensity_150_3a	150	300	case_150_3a.txt	control_150_3a.txt	Tables
Random_Intensity_150_4a	150	300	case_150_4a.txt	control_150_4a.txt	Tables
Random_Intensity_150_5a	150	300	case_150_5a.txt	control_150_5a.txt	Tables
Random_Intensity_300_1a	300	300	case_300_1a.txt	control_300_1a.txt	Tables
Random_Intensity_300_2a	300	300	case_300_2a.txt	control_300_2a.txt	Tables
Random_Intensity_300_3a	300	300	case_300_3a.txt	control_300_3a.txt	Tables
Random_Intensity_300_4a	300	300	case_300_4a.txt	control_300_4a.txt	Tables
Random_Intensity_300_5a	300	300	case_300_5a.txt	control_300_5a.txt	Tables

(Last updated 5/3/07)