Identification of files.

There are tree formats:

1. G_1_10_444_8_16_15_70_600000_1000000_muestra_tot_meA.csv
2. G_1_10_444_8_16_15_70_600000_1000000_muestra_70_tot_meA.csv
3. G_1_10_444_8_16_15_70_600000_1000000_galaxias_intA.csv

Format 1. The training sample contains 60% of the observations and the test sample the remaining 40%. They contain coordinates (x, y, z) of cells where the unit is the cell width, in addition, it indicates if the cell is occupied (column E) and if so for how many galaxies (column n).

G_1_10_444_8_16_15_70_600000_1000000_muestra_tot_meA.csv

G_1

G_0 luminosity distances

G_1 comoving distances

_10

Diameter of the universe in Gly

444

Number of cells per-axis X, Y, Z

Delimitation of the spherical crown:

8_16_

between 8 and 16 h of right ascension (equatorial coordinates J2000).

15_70

between 15 to 70 degrees of declination (equatorial coordinates J2000).

600000_1000000

Between 600 and 1000 million light-years from the earth.

muestra_tot_meA

muestra_tot_meA = The training sample (HyperLeda).

muestra_tot_mpA = The test sample (HyperLeda).

muestra_tot_mzA = Set of the two previous samples.

muestra_tot_az_meA = The training sample (ramdom)

muestra_tot_az_mpA = The test sample (ramdom)

muestra_tot_az_mzA = Set of the two previous samples.

muestra_tot_azaj_meA = The training sample (random samples using the density)

muestra_tot_azaj_mpA = The test sample (random samples using the density)

muestra_tot_azaj_mzA = Set of the two previous samples.

Format 2. It is the same as the previous one except that it adds the percentage of observations in the training sample:

G_1_10_444_8_16_15_70_600000_1000000_muestra_70_tot_meA.csv

_70_

The training sample contains 70% of the observations..

Formato 3. Similar a los anteriores.

G_1_10_444_8_16_15_70_600000_1000000_galaxias_intA.csv

It contains as many observations as there are galaxies in the sector. For each observation (galaxy) it registers the coordinates of the cell in which it is included, all the observations contain a 1 in the occupation column (column E) and a 0 or a 1 which corresponds to the number of galaxies of the formats 1 and 2 (column n), 1 means that the galaxy belongs to the training sample and 1 to the test sample. It can also contain other data (objname; pgc; al2000; de2000; r).