

Manuel Escabias¹, Ana M. Aguilera¹, M. Carmen Aguilera-Morillo², and Mariano J. Valderrama¹

¹Department of Statistics and O.R. University of Granada, Granada,

²Department of Statistics University Carlos III, Madrid



August 10, 2018

Datasets

<https://consigna.ugr.es/f/MHI4GocFBHXB9c2C/Data.zip>

Monthly weather

Classical dataset (Ramsay and Silverman, 2005) available in R's `fda` package.

Discretely observed curves are mean monthly temperatures in 35 canadian weather stations. Annual cummulated rainfall is also registered.

- ▶ `MonthlyTemp.csv`. File with discretely observed curves. 35 rows (weather stations) and 12 columns (months). Discrete observations x_{ij} are mean monthly temperature in weather station i in month j .
- ▶ `CumPrec.csv`. File with annual cummulated rainfall. 35 rows and 1 column. y_i is the annual cummulated precipitation in weather station i .
- ▶ `CumPrec01.csv`. File with the binary variable that indicates if in a weather station cummulated annual rainfall is bigger than the mean of the annual cummulated rainfalls. 35 rows and 1 column.

DailyTemp

Classical dataset (Ramsay and Silverman, 2005) available in R's `fda` package.

Discretely observed curves are mean daily temperatures in 35 canadian weather stations. Annual cummulated rainfall is also registered.

- ▶ `DailyTemp.csv`. File with discretely observed curves. 35 rows (weather stations) and 365 columns (days). Discrete observations x_{ij} are mean daily temperature in weather station i in day j .
- ▶ `CumPrec.csv`. File with annual cummulated rainfall. 35 rows and 1 column. y_i is the annual cummulated precipitation in weather station i .
- ▶ `CumPrec01.csv`. File with the binary variable that indicates if in a weather station cummulated annual rainfall is bigger than the mean of the annual cummulated rainfalls. 35 rows and 1 column.

Growth

Classical dataset (Ramsay and Silverman, 2005) available in R's `fda` package.

Heights of 39 boys and 54 girls from age 1 to 18 and the ages at which they were collected. The ages are not equally spaced.

- ▶ `GrowthHeights.csv`. 93 rows and 31 columns. Heights of 54 girls (first 54 rows) and 39 boys (last 39 rows) at 31 ages between 1 and 18.
- ▶ `GrowthAges.csv`. 1 row and 31 columns with the ages where heights are observed.
- ▶ `Gender.csv`. 1 column and 93 rows with 1's and 0's. 1 indicating girl and 0 boy.

Medfly

Popular dataset for functional data analysis. The medfly data consist of records of the number of eggs laid by 50 fruit flies on each of 26 days, along with each individual's total lifespan.

- ▶ MedflyEgg.csv. 50 rows and 26 columns with egg counts of 50 fruit flies during the 26 days of lifespan
- ▶ MedflyLife.csv. 50 rows and 1 column with the life time of 50 fruit flies.

Tecator

215 samples of finely chopped pure meat with different fat, water and protein content. For each sample of meat a 100 channel spectrum of absorbances was recorded. 215 curves (rows) were observed at 100 wavelength (columns) equispaced between 850 and 1050.

- ▶ Spectro_xt_Eq.csv: 215 rows (curves) and 100 columns of spectro.
- ▶ Spectro_t_Vec.csv: 1 row and 100 columns with the wavelengths. 100 equispaced values between 850 and 1050.
- ▶ Spectro_xt_Uneq.csv: 215 rows with different number of columns for each row. Curves of spectro.
- ▶ Spectro_t_Mat_Uneq.csv: 215 rows with different number of columns for each row. Same format than Spectro_xt_Uneq.csv. Wavelengths where curves of spectro are observed.

Tecator

215 samples of finely chopped pure meat with different fat, water and protein content. For each sample of meat a 100 channel spectrum of absorbances was recorded. 215 curves (rows) were observed at 100 wavelength (columns) equispaced between 850 and 1050.

- ▶ Y_fat.csv: 215 rows and 1 column. Response variable with the percentage of fat of each piece of meat.
- ▶ Y_Protein.csv: 215 rows and 1 column. Response variable with the percentage of protein of each piece of meat.
- ▶ Y_water.csv: 215 rows and 1 column. Response variable with the percentage of water of each piece of meat.
- ▶ Y_water_bin.csv: 215 rows and 1 column. Binary response variable that shows if a piece of meat has more than 50% of water.

aemet

aemet data. Weather data of 73 weather stations in Spain. Available in `fda.usc` package. Similar to `DailyTemp` for spanish weather stations.

- ▶ `aemet_temp.csv`. 73 rows (weather stations) and 365 columns (days) of temperature. For each one of 365 days, the registered temperature is the mean temperature for the same day over the period 1980-2009.
- ▶ `aemet_logprec.csv`. 73 rows (weather stations) and 365 columns (days) of the logarithm of precipitation.
- ▶ `aemet_wind.csv`. 73 rows (weather stations) and 365 columns (days) of wind speed.
- ▶ `aemet_north.csv`. 73 rows (weather stations) and one column of 0's and 1's. A weather station takes value 1 if its latitude is bigger than 40 (approximately the center of Spain).