

# Code Book

Adler Fonseca de Castro

6/25/2020

## Study Design

The purpose of this coursework is to demonstrate the ability to collect, work with, and clean a data set. The goal is to prepare tidy data that can be used for later analysis.

The data used here represent measurements collected from accelerometers from the Samsung Galaxy S smartphone. A full description is available [here](#). For this work, it was downloaded from this link and extracted to the UCI HAR Dataset folder.

To replicate the processing steps, run the script `run_analysis.R` in the same directory as the mentioned folder. The output is the tidy data set saved as `tidy_dataset.txt`

## Code Book

Regarding the raw data, there is a detailed explanation in the README file inside the data folder. For the purposes of this coursework, it is enough to quote:

“The experiments have been carried out with a group of 30 volunteers within an age bracket of 19-48 years. Each person performed six activities (WALKING, WALKING\_UPSTAIRS, WALKING\_DOWNSTAIRS, SITTING, STANDING, LAYING) wearing a smartphone (Samsung Galaxy S II) on the waist. Using its embedded accelerometer and gyroscope, we captured 3-axial linear acceleration and 3-axial angular velocity at a constant rate of 50Hz. The experiments have been video-recorded to label the data manually. The obtained data set has been randomly partitioned into two sets, where 70% of the volunteers was selected for generating the training data and 30% the test data.”

For each record it is provided:

- A 561-feature vector with the resulting variables. The label of each of those features is listed in `features.txt`. Each feature vector is a row on the text file.
- Its activity label, listed in `train/y_train.txt` and `test/y_test.txt`.
- An identifier of the subject who carried out the experiment, listed in `train/subject_train.txt` and `test/subject_test.txt`

The measurements are recorded in the `train/X_train.txt` and `test/x_test.txt` files.

The processing steps consisted in:

1. Reading the data from the mentioned files, using the information on `features.txt` to appropriately label each variable in the measurements.
2. Merging the training and test sets to create one data set.
3.
  1. Extracting only the measurements on the mean and standard deviation for each measurement. Those are labeled with “`mean()`” and “`std()`” respectively.
  2. Tidying the labels by removing the characters “`()`” and replacing “`-`” with “`.`”.
4. Assigning descriptive names for the activities in the data set.

5. From the data set in step 4, creating a second, independent tidy data set with the average of each variable for each activity and each subject.
6. Outputting this data set with `write.table()`.