SCHOOL OF ELECTRICAL ENGINEERING AND COMPUTING

BIG DATA & DATA ANALYTICS
LAB PROJECT 3
(DATASET DESCRIPTION & PREPARATION)

This **LAB** project is based on a meteorological dataset about forest fires in the northeast region of Portugal. The dataset is available from the UCI Machine Learning Repository (Lichman, 2013):

http://archive.ics.uci.edu/ml/datasets/Forest+Fires

PREPARATION

In preparation for Lab Project 3, load the forestfires.csv dataset. Use the dataset to practice the following topics:

- Create correlation tables and correlation matrices
- Reporting of results in APA style (including tests discussed in earlier weeks, e.g., ANOVA, *t*-test)
- Storing the results of ggplot and other plotting tools to automatically generate png and pdf files
- Creating rtf files and include text, tables, and figures
- Running R files from the command line with "RScript"
- Simple linear regressions: coefficient estimates, predicted values, residuals, standard errors, confidence intervals, *t*-values, residual standard error, R^2
- Analysing and interpreting the results of simple linear regressions
- Find out more about the function "describe" to explore a dataset and find out whether there are missing values.
- Create density plots with ggplot()
- Install and load the library "moments". Search online to find out more about the functions "skewness()" and "kurtosis()" of the library "moments" and the interpretation of their results. What is skewness? What is kurtosis? (e.g., http://www.r-bloggers.com/measures-of-skewness-and-kurtosis)

REFERENCES

Lichman, M. (2013). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.

DATASET

Forestfires	*Forest Fires Dataset*

Description

A meteorological dataset about forest fires in the northeast region of Portugal. Some of the data includes measures from the Fire Weather Index (FWI).

Usage

Forestfires

Format

A data frame with 517 observations on the following 13 variables.

`X`	x-axis spatial coordinate within the Montesinho park map: 1 to 9
`Y`	y-axis spatial coordinate within the Montesinho park map: 2 to 9
`month`	month of the year: 'jan' to 'dec'
`day`	day of the week: 'mon' to 'sun'
`FFMC`	FFMC (Fine Fuel Moisture Code) index from the FWI system: 18.7 to 96.20. This is a numerical rating of the moisture content of surface litter and other cured fine fuels. It shows the relative ease of ignition and flammability of the fine fuels.
`DMC`	DMC (Duff Moisture Code) index from the FWI system: 1.1 to 291.3. The DMC is a numerical rating of the average moisture content of loosely compacted organic layers of moderate depth.
`DC`	DC (Drought Code) index from the FWI system: 7.9 to 860.6. is is a numerical rating of the moisture content of deep, compact, organic layers. It is a useful indicator of seasonal drought and shows the likelihood of fire involving the deep duff layers and large logs.
`ISI`	ISI (Initial Spread Index) index from the FWI system: 0.0 to 56.10. This indicates the rate fire will spread in its early stages.
`temp`	temperature in Celsius degrees: 2.2 to 33.30
`RH`	relative humidity in %: 15.0 to 100
`wind`	wind speed in km/h: 0.40 to 9.40
`rain`	outside rain in mm/m2 : 0.0 to 6.4
`area`	the burned area of the forest (in ha): 0.00 to 1090.84

Source

Lichman, M. (2013). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.

Cortez, P. & Morais, A. (2007). A Data Mining Approach to Predict Forest Fires using Meteorological Data. In J. Neves, M. F. Santos and J. Machado Eds., *New Trends in Artificial Intelligence*, Proceedings of the 13th EPIA 2007 - Portuguese Conference on Artificial Intelligence, December, Guimaraes, Portugal, pp. 512-523, 2007. APPIA, ISBN-13 978-989-95618-0-9.

FWI (Fire Weather Index) descriptions are taken from: http://www.malagaweather.com/fwi-txt.htm