Forestal and Environmental modelling and data science

Temperature Linear Regression

Spring 2023

Steffen Manfred Noe
Emílio Graciliano Ferreira Mercuri (post-doctoral researcher)
Estonian University of Life Sciences (EMÜ)

Outline

- Understand what is a Linear Regression
- Work with Temperature data from different cities
- Calculate the linear regression and correlation using data from Kaggle
- Explore the Google Colab facility

How temperature have changed in your city during the last 10-20 years?

- It went warmer or cooler?
- If yes, how many degrees celsius?
- Is Global Warming myth or reality?

Linear Regression (LR)

- LR is a linear approach for modelling the relationship between a scalar response and one or more explanatory variables (also known as dependent and independent variables).
  - For 1 explanatory variable is called **simple linear regression**;

$$y = a_0 + a_1 x$$

independent variable
(explanatory variable)

dependent variable
(scalar response)

a_0, a_1 -> constants determined by a minimization process

  - For +1 explanatory variable the process is called **multiple linear regression**

Linear Regression applied to Temperature time series

- Applying to our data set

$$T = a_0 + a_1 t$$

time (years) = independent variable (explanatory variable)

Temperature = dependent variable (scalar response)

a_0, a_1 -> constants determined by a minimization process

Minimization process

When we talk about linear regression we mean "fitting a straight line to the data." Thus,

$$f(x) = a_0 + a_1 x$$

In this case, the function that we'll minimize is:

$$S(a_0, a_1) = \sum_{i=0}^{n} [y_i - f(x_i)]^2 = \sum_{i=0}^{n} (y_i - a_0 - a_1 x_i)^2$$

Equations (2) become:

$$\frac{\partial S}{\partial a_0} = \sum_{i=0}^{n} -2(y_i - a_0 - a_1 x_i) = 2\left[a_0(n+1) + a_1 \sum_{i=0}^{n} x_i - \sum_{i=0}^{n} y_i\right] = 0$$

$$\frac{\partial S}{\partial a_1} = \sum_{i=0}^{n} -2(y_i - a_0 - a_1 x_i)x_i = 2\left[a_0 \sum_{i=0}^{n} x_i + a_1 \sum_{i=0}^{n} x_i^2 - \sum_{i=0}^{n} x_i y_i\right] = 0$$

... Complete deduction is on the Notebook

$$a_1 = \frac{\sum_{i=0}^{n} y_i(x_i - \bar{x})}{\sum_{i=0}^{n} x_i(x_i - \bar{x})} \quad , \quad a_0 = \bar{y} - a_1 \bar{x}$$

Let's code :)

- Open your Google Drive
- Create / find your folder, organize it
- Start a new Google Colab session

A copy of the notebook is on https://github.com/stenoe/FEDS