

Autocorrelation in Weather

Pablo Lechon

1 Introduction

The goal in this practical is to write an `r` script that helps answer the question: *are temperatures of one year significantly correlated with the next year (successive years), across years in a given location?*

For this, we calculate the correlation between $n - 1$ pairs of years, where n is the total number of years. note that one cannot use the standard p-value calculated for a correlation coefficient, because measurements of climatic variables in successive time-points in a time series (successive seconds, minutes, hours, months, years, etc.) are *not independent*.

2 Data

We first load the data with the comand `load(filename, envir = globalenv())`, which consists of two columns: Year and Temperature. A quick plot to visualize these data can be seen in figure 1

It can be easily seen that temperatures between consecutive years are correlated. If we calculate the correlation between consecutive years by having two Year column next to each other, shifted one position, we obtain the value 0.32. This value is meaningless if we don't determine wether or not is significant. To do it, we calculate the p-value numerically

3 p-value

Calculating the p-value is the same as answering the question: How likely wouldit be that, assuming no correlation between consecutive years, we obtained a value of 0.32. To calculate it numerically we make many more comparisons between two shuffled columns of years. This assures that there will be no correlation between them. We are, in fact, generating a normal distributions of correlations with mean = 0. This distribution can be seen in figure 2

To calculate the p-value we have to obtain the ratio between the correlation coefficients coming from the normal distribution that are higher than the actual correlation coef-ficient. This fraction of the order of 10^{-4} , wich means that the probability of finding a correlation of 0.32 by chance, is that low. Therefore, we conclude that the correlation is very significant, yet weak.

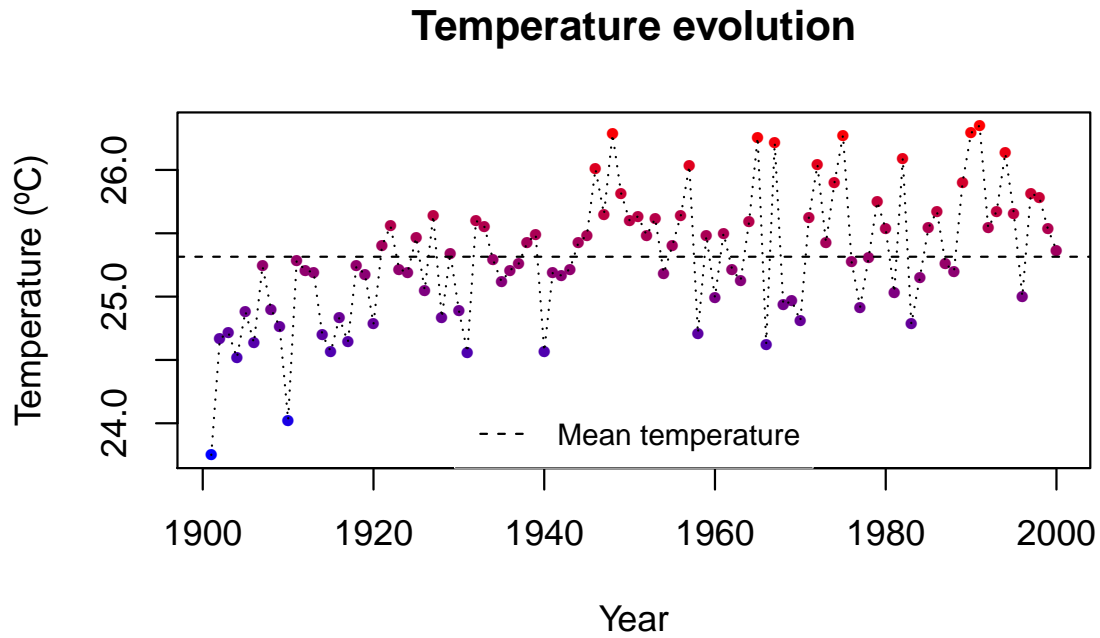


Figure 1: Temperature time series. Redder indicates warmer. A dashed line connects the points for better visualization. The mean temperature is represented with a horizontal dashed line

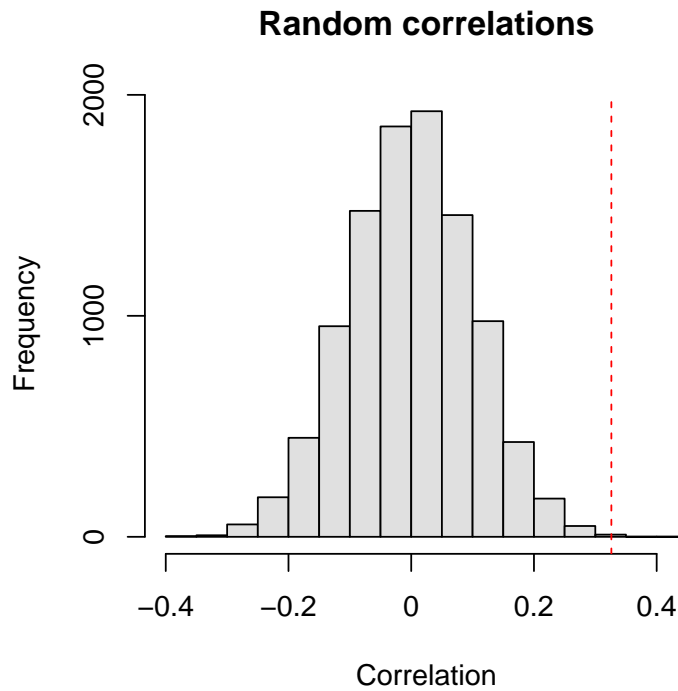


Figure 2: Normal distribution of correlation numerically generated. The vertical red line represents the correlation between consecutive years