Is annual temperatures are correlated in successive years in Florida?

01_Awesome_Aardvarks

December 13, 2022

1 Introduction

Key West is an island located in Florida. Warmer temperatures in Key West can result in less dissolved oxygen in water which may associated with fish dying. Higher temperatures can increasing risk of coral bleaching as well(Group 2009).

In this study, we analyzed whether the temperature in one year is significantly correlated with the next year(successive year), across years in Key West.

2 Materials & Methods

2.1 Computing Tools

We use the R language because the R language is convenient in the processing of data frame, and the ability to analyze small dataset (here the data has 100 observations) is excellent. The operating system is Ubuntu 22.04.

2.2 Data

The data recorded the temperatures in Florida from 1900 to 2000. The scatter plot between temperature and year is shown in Figure 1. Here we show some of the data.

```
> str(ats)
  'data.frame':    100 obs. of  2 variables:
  $ Year: int   1901 1902 1903 1904 1905 1906 1907 1908 1909 1910 ...
  $ Temp: num   23.8 24.7 24.7 24.5 24.9 ...
```
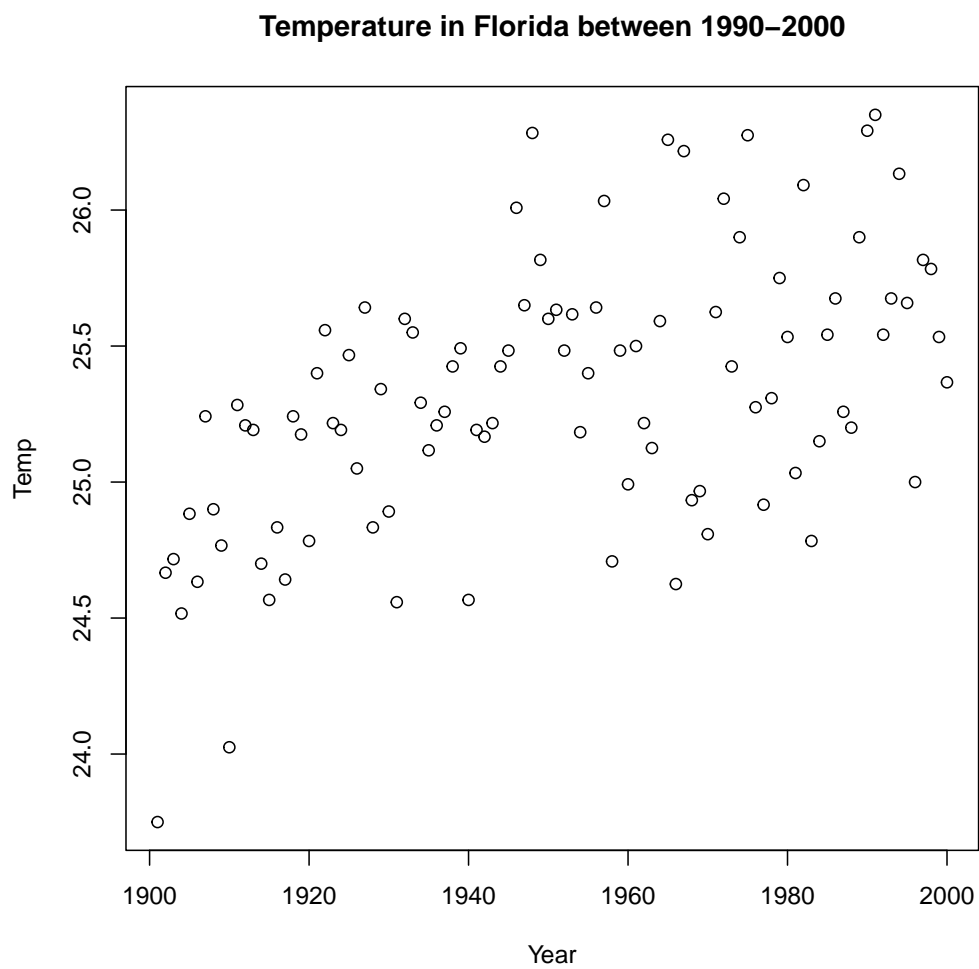
Temperature in Florida between 1990–2000

Figure 1: The scatter plot between temperatures and years.

2.3 Permutation Analysis

A permutation test (also called re-randomization test) is an exact statistical hypothesis test making use of the proof by contradiction. The properties of random permutations were proved by Hemerik & Goeman (2018) through the Conditional Monte Carlo test they proposed.

The test statistic is the correlation coefficient between temperature and successive years. The null hypothesis here is the temperature of one year isn't correlated with the next year across years in Florida. The alternative hypothesis here is temperature of one year is correlated with the next year across years in Florida. We did 10000 times shuffles of successive years.

2.4 Estimation of p-value

Instead of standard p-value calculation, we calculated the estimation of p-value in equation 1 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.

$$estimated\_p = \frac{\sum \left(cor\_of\_random \geq temp\_cor \right)}{length(cor\_of\_random)} \tag{1}$$

3 Results & Discussion

The estimated p-value after 10000 times shuffles is around 1×10^{-4} which is much less than 0.05. This p-value supports the alternative hypothesis and rejects the null hypothesis.

The distribution of the random correlation coefficients is shown in Figure 2. The red line in Figure 2 is the observed Pearson correlation coefficient r before shuffling the successive years. The observed r is larger compared to the distribution of possible r. In other words, when the empirical value is very far from the permutation results, we can be pretty sure that: when the variables are truly independent, a r as strong as the observed r is unlikely to appear in a situation of pure chance.
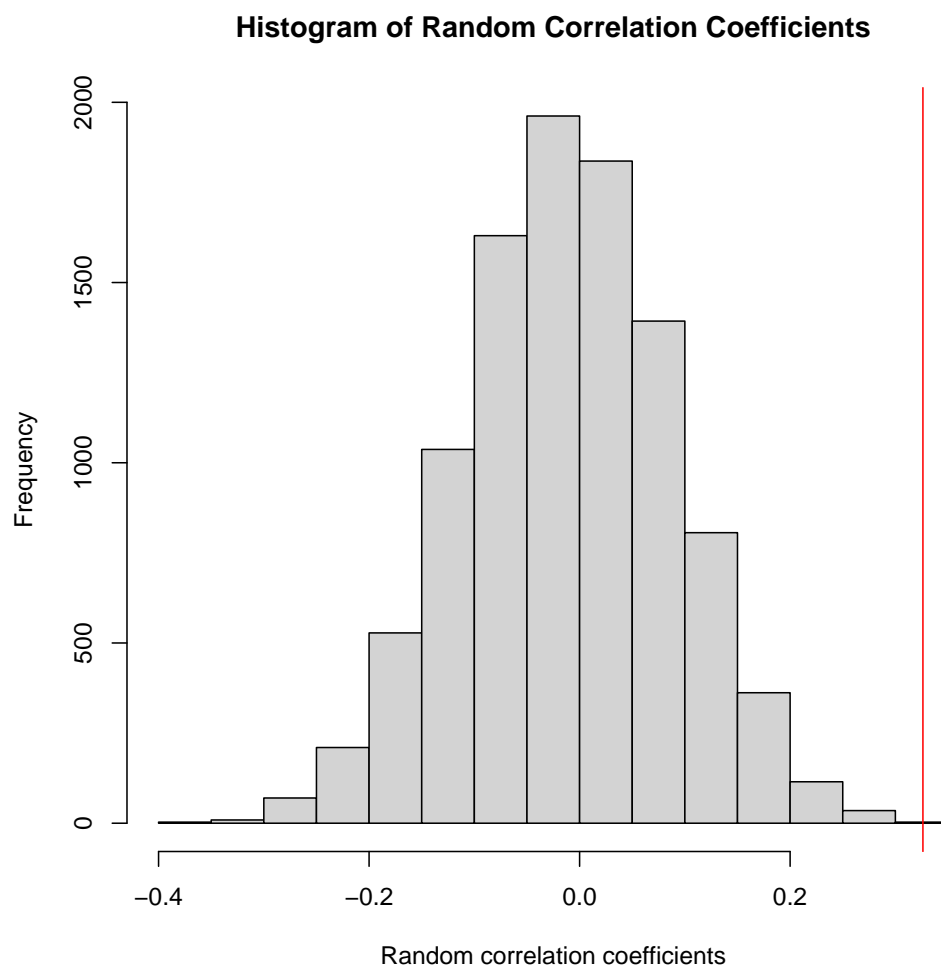
Figure 2: Correlation coefficients after permutation.

References

Group, I. C. C. (2009), *Temperature and Evapotranspiration*, pp. 13–14.

Hemerik, J. & Goeman, J. (2018), 'Exact testing with random permutations',
27(4), 811–825.
URL: *http://link.springer.com/10.1007/s11749-017-0571-1*