

Analysis of Key West Annual Mean Temperatures

Groupwork Practical: Autocorrelation in Florida Weather

Group Members:

Zhang, Tianye (tianye.zhang24@imperial.ac.uk)

Li, Yibin (yibin.li24@imperial.ac.uk)

Pearce, Saskia (saskia.pearce21@imperial.ac.uk)

Li, Kaiwen (kaiwen.li21@imperial.ac.uk)

Objective and Hypotheses

The objective is to evaluate whether there is a significant correlation between the mean annual temperatures (**Temp**) of successive years in Key West using a permutation-based significance test.

- H_0 : There is no correlation between temperatures of successive years (**Temp**).
- H_1 : There is a significant correlation between temperatures of successive years (**Temp**).

Data Overview

The dataset contains 100 observations of annual mean temperatures (**Temp**) recorded from 1901 to 2000 in Key West, Florida. The mean temperature is 25.31°C with a standard deviation of 0.495°C.

Methods

To assess the correlation between successive years:

1. The observed Pearson correlation coefficient (r_{observed}) between temperatures of successive years was calculated as $r_{\text{observed}} = 0.326$. The relationship between temperatures of successive years is illustrated in Figure 1.

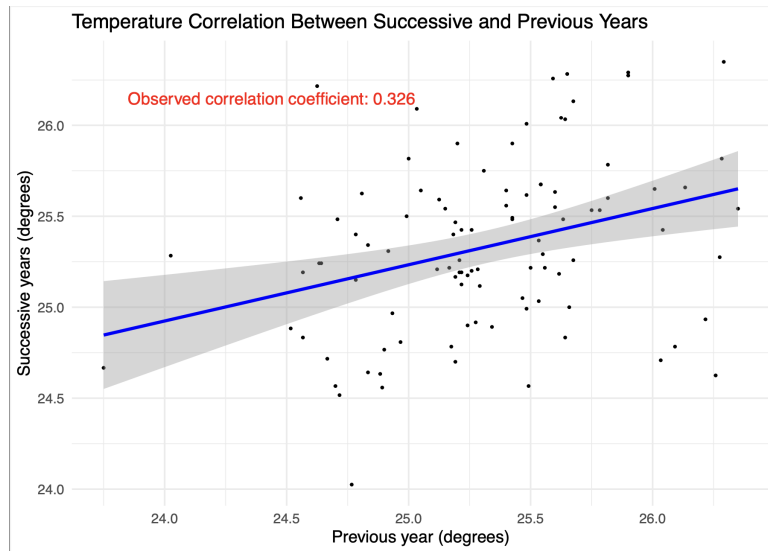


Figure 1: Scatterplot showing the relationship between successive years' temperatures, with a fitted regression line and observed correlation coefficient.

A permutation test with $n_{\text{sim}} = 10,000$ iterations was performed by shuffling **Temp** and recalculating the correlation for each randomly permuted dataset. To ensure reproducibility, the random seed was set using `set.seed(123)`.

The p-value was computed as:

$$p_{\text{sim}} = \frac{1}{n_{\text{sim}}} \sum_{i=1}^{n_{\text{sim}}} \mathbf{1}\{|r_{\text{random},i}| \geq |r_{\text{observed}}|\},$$

where $r_{\text{random},i}$ is the i -th random correlation, and $\mathbf{1}$ is the indicator function.

As the number of permutations (n_{sim}) increases, the empirical p-value (p_{sim}) converges to the true p-value (p) due to the **law of large numbers**. In the code, permutations were tested with $n_{\text{sim}} = 10, 100, 1,000$, and $10,000$, yielding progressively more stable results. For smaller n_{sim} , the randomness of permutation leads to greater variability in p_{sim} , while larger n_{sim} reduces this variability.

By the **central limit theorem**, the empirical p-value (p_{sim}) for a finite number of permutations approximately follows a normal distribution:

$$p_{\text{sim}} \sim N\left(p, \frac{p(1-p)}{n_{\text{sim}}}\right).$$

This implies that as n_{sim} increases, the variance of p_{sim} ($\text{Var}(p_{\text{sim}})$) decreases:

$$\text{Var}(p_{\text{sim}}) = \frac{p(1-p)}{n_{\text{sim}}}.$$

Thus, the precision of the p-value estimate improves with more permutations, eventually converging to the true p-value (p). However, diminishing returns are observed beyond a certain point, as further increasing n_{sim} provides only marginal gains in accuracy.

Results

The observed Pearson correlation coefficient between temperatures of successive years was $r_{\text{observed}} = 0.326$. A permutation test with $n_{\text{sim}} = 10,000$ yielded a p-value of $p = 0.001$, indicating that the observed correlation is highly significant.

Figure 2 shows the null distribution of random correlation coefficients. The observed $r_{\text{observed}} = 0.326$ (red dashed line) lies far in the right tail, demonstrating strong evidence against the null hypothesis (H_0) of no correlation.

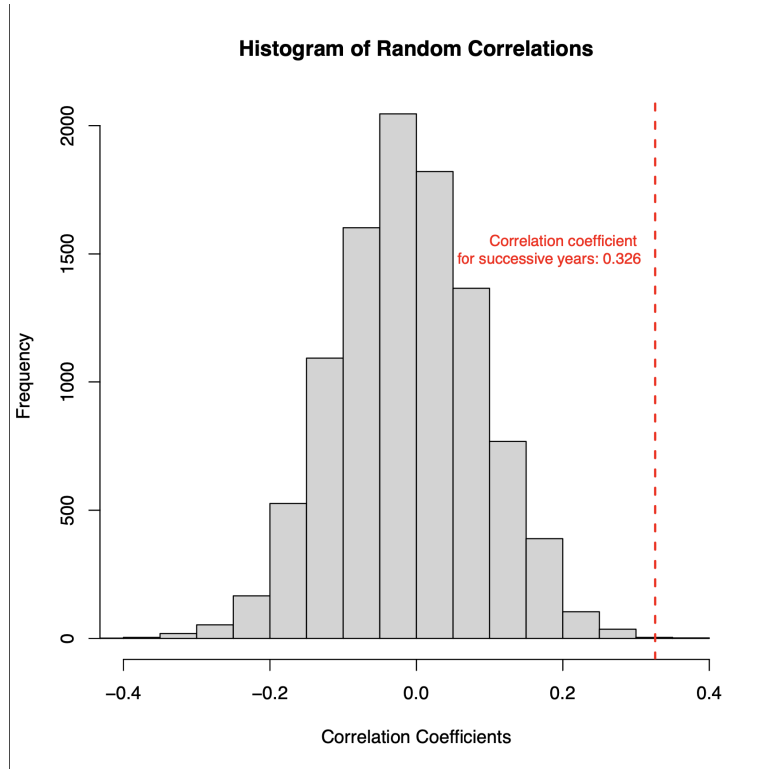


Figure 2: Histogram of random correlation coefficients from 10,000 permutations. The red dashed line represents the observed correlation $r_{\text{observed}} = 0.326$.

These results provide compelling evidence of a significant temporal dependence in Key West annual temperatures, suggesting that temperature in one year influences that of the following year.

Conclusion

These results provide compelling evidence of a significant temporal dependence in Key West annual temperatures. The observed correlation coefficient ($r_{\text{observed}} = 0.326$) suggests that temperatures in one year are positively correlated with those of the following year. This indicates that temperature dynamics in Key West exhibit a degree of continuity over time, which may reflect underlying climatic patterns or local environmental factors influencing year-to-year variations.

The statistically significant p-value ($p = 0.001$) derived from the permutation test strongly rejects the null hypothesis (H_0) of no correlation, emphasizing that the observed relationship is unlikely to have occurred by random chance. The findings underscore the importance of examining temporal dependencies in climate data to better understand trends and potential long-term impacts.