Analysis of Key West Annual Mean Temperatures

Objective and Hypotheses

The objective is to evaluate whether there is a significant correlation between the years (`Year`) and the annual mean temperatures (`Temp`) of Key West, using permutation tests.

- H_0: There is no correlation between `Year` and `Temp`.
- H_1: There is a significant correlation between `Year` and `Temp`.

Data Overview

The dataset contains 100 observations with two variables: **Year**, representing the year of observation, and **Temp**, representing the annual mean temperature in degrees Celsius. The statistical analysis indicates that the average annual temperature is 25.31°C, with a standard deviation of 0.495°C.

Methods

The Pearson correlation coefficient between `Year` and `Temp` was calculated as $r_{\text{observed}} = 0.68$. A permutation test with $n_{\text{sim}} = 10,000$ iterations was conducted by shuffling `Temp` while keeping `Year` fixed, generating a null distribution of correlation coefficients. The p-value was computed as:

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

where $r_{\text{random},i}$ is the i-th random correlation, and $\mathbf{1}$ is the indicator function that equals 1 if the condition is true and 0 otherwise.

Although $n_{\text{sim}} = 10,000$ was chosen for this analysis to balance computational cost and precision, it is worth noting that as $n_{\text{sim}} \to \infty$, the empirical p-value p_{sim} converges to the true p-value p according to the law of large numbers. By testing various values of n_{sim} (e.g., 10, 100, 1000, 10,000, and beyond), it became evident that due to the large observed value of $r_{\text{observed}} = 0.68$, the empirical p-value p_{sim} was consistently 0. However, it remains important to emphasize that as n_{sim} increases, the empirical p-value p_{sim} will still theoretically converge to a fixed value as $n_{\text{sim}} \to \infty$. Furthermore, by the central limit theorem, p_{sim} approximately follows a normal distribution:

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

where p is the true p-value. As n_{sim} increases, the variance $\frac{p(1-p)}{n_{\text{sim}}}$ decreases, resulting in a more precise estimate of the p-value. Thus, larger n_{sim} values lead to greater accuracy, though diminishing returns may be observed beyond a certain point.

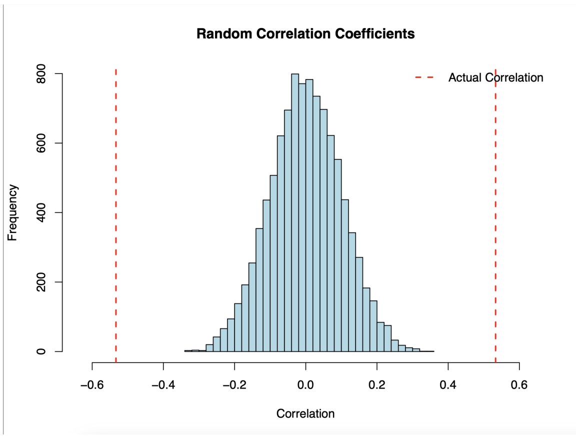

Figure 1: Histogram of random correlation coefficients. Red dashed lines indicate the observed correlation.

Results

The observed Pearson correlation coefficient between `Year` and `Temp` was $r_{\text{observed}} = 0.68$. A permutation test with $n_{\text{sim}} = 10,000$ iterations yielded a p-value of $p = 0$, indicating that no random correlation coefficients in the null distribution were as extreme as the observed value.

Figure 1 shows the null distribution of 10,000 random correlation coefficients, centered around 0. The red dashed lines represent the observed correlation $r_{\text{observed}} = 0.68$, which lies far in the tail, beyond any randomly generated values. This result strongly rejects the null hypothesis (H_0), providing robust evidence of a significant positive relationship between `Year` and `Temp`. The observed correlation suggests a potential warming trend in Key West over time.