

# Autocorrelation in Weather: Results

Lucy Goodyear

21st October 2019

## 1 Introduction

Our null hypothesis states that temperatures of one year are not significantly correlated with the next year (successive years), across years in a given location. The alternative hypothesis states that they are correlated. Our sample data is from Key West, FLorida, as shown in figure 1.

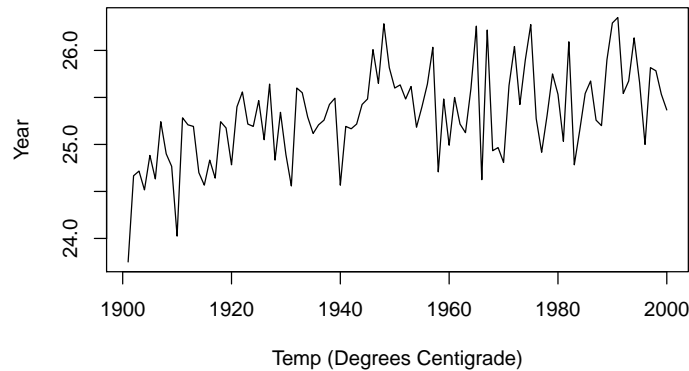


Figure 1: Key West Annual Mean Temperature

The code below calculates the correlation coefficient for temperature between successive years and then calculates the correlation coefficient for 10,000 random permutations of the temperature in order to compare significance.

Normally, for the correlation to be significant, we would expect a significance level of 0.05 or less. However, since climate data measured in a time series is not independent, we cannot use a standard p-value calculation. This means that our significance could be slightly greater than 0.05 and we wouldn't necessarily reject the null hypothesis.

## 2 Code

Below is the code used to calculate the autocorrelation and an estimation of its significance.

```
> # imports
> library(dplyr)
> # load weather data
> load("../Data/KeyWestAnnualMeanTemperature.Rdata")
> # plot temperatures over time on line graph
> pdf("KWAMT.pdf", 6, 4)
> plot(ats$Year,
+      ats$Temp,
+      xlab = "Temp (Degrees Centigrade)",
+      ylab = "Year",
+      type = "l")
> # create two lists of temperatures, one with the first row deleted and then
> # realigned with the second list so temps between successive years can be
> # easily compared
> Temp_t0 <- c(ats[2:100,2])
> Temp_t1 <- c(ats[1:99,2])
> # calculate the correlation between successive years
> CorCoeff <- cor(Temp_t0, Temp_t1)
> print(CorCoeff)

[1] 0.3261697

> # create matrix of 10000 random permutations of temperature column
> Temps1 <- replicate(10000, sample(ats$Temp,replace = F))
> # for each permutation, realign as before and then calculate correlation
> RdmCors <- vector("numeric", 10000)
> for (i in 1 : 10000) {
+   RdmTemps <- cor( Temps1[2:100,i], Temps1[1:99,i])
+   RdmCors[i] <- RdmTemps
+ }
> # calculate the fraction of correlation coefficients more significant than
> # that of successive years (CorCoeff)
> sum(RdmCors > CorCoeff) + sum (RdmCors < -CorCoeff)

[1] 11
```

## 3 Results

Running the code above gives us an autocorrelation between successive years of 0.3261697, which indicates a slight correlation.

Using the result from the final line of code, we can see how significant this value is. Since the numbers are generated at random, this value fluctuates

between 4 and 14, giving a very rough average of 9, which is a significance value of 0.09.

As stated in the introduction, this value would usually mean a failure to reject the null hypothesis. However, given the dependent nature of the variables, more statistical tests would be recommended before informed conclusions can be drawn.