Are temperatures of one year significantly correlated with the next year (successive years), across years in a given location?

Lucy Goodyear

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.

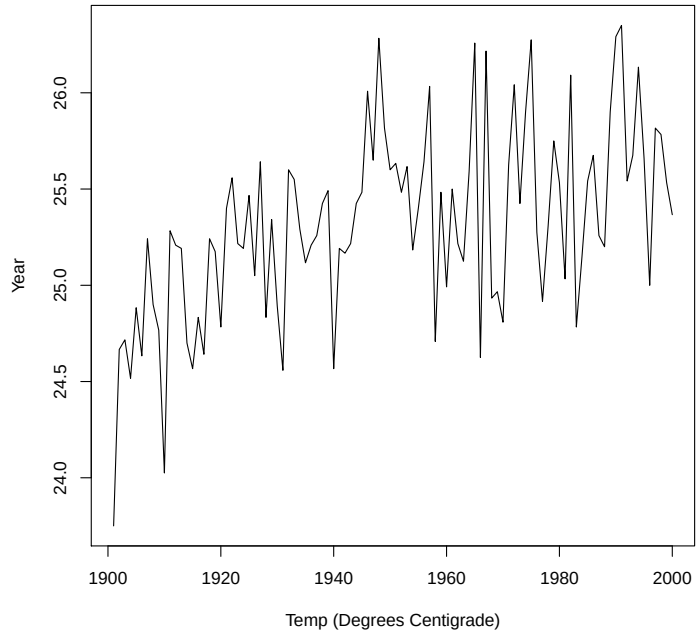

Figure 1: Key West Annual Mean Temperature

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 temperatures in order to compare significance.

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. As such, the calculation in the last line of the code is used to provide an estimated p-value.

2 Code

Below is the code used to calulcate the autocorrelation and an estimate of its significance.

```
> # loads and plots a dataset of temperatures by year in Key West. It creates two
> # lists offset from each other by one year, which allows the correltaion to be
> # calculated for successive years. It then calculates the correlation coefficient
> # for 10,000 random permutations of the temperature in order to compare significance.
>
> # Author: Lucy Goodyear (lucy.goodyear19@imperial.ac.uk)
> # Version: 0.0.1
>
> # clear workspace
> rm(list=ls())
> # dev.off()
>
> # imports
> library(dplyr)
> # load weather data
> load("../Data/KeyWestAnnualMeanTemperature.Rdata")
> # plot tmperatures over time on line graph
> pdf("../Results/KWAMT.pdf")
> plot(ats$Year,
+      ats$Temp,
+      xlab = "Temp (Degrees Centigrade)",
+      ylab = "Year",
+      type = "l")
> dev.off()

null device
          1

> # create two lists of tmperatures, one with the first row deleted and then
> # realigned with the second list so temps between suuccesive 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)
> cat("Correlation between successive years is", CorCoeff)

Correlation between successive years is 0.3261697

> # create matrix of 10000 random permututions 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
+ }
> # generate histogram comparing p-values
> pdf("../Results/KWAMT_corr.pdf")
> hist(RdmCors,
+      xlim = c(-0.4, 0.4),
+      xlab = "Correlation Coefficients",
+      main = NULL)
> abline(v = CorCoeff, col = "red", lwd = 1, lty = 2)
> text(CorCoeff, 1950, "correlation coefficient for\n successive years")
> dev.off()

null device
          1

> # calculate out estimated p-value (the fraction of correlation coefficients more
> # significant than that of successive years (CorCoeff))
> p_estimate <- (sum(RdmCors > CorCoeff) + sum (RdmCors < -CorCoeff)) / 10000
> # print p_estimate to screen
> cat("P-value estimate is", p_estimate)

P-value estimate is 0.001
```

3 Results

Running the code above gives us an autocorrelation between successive years of 0.3261697, which indictates moderate positive correlation.

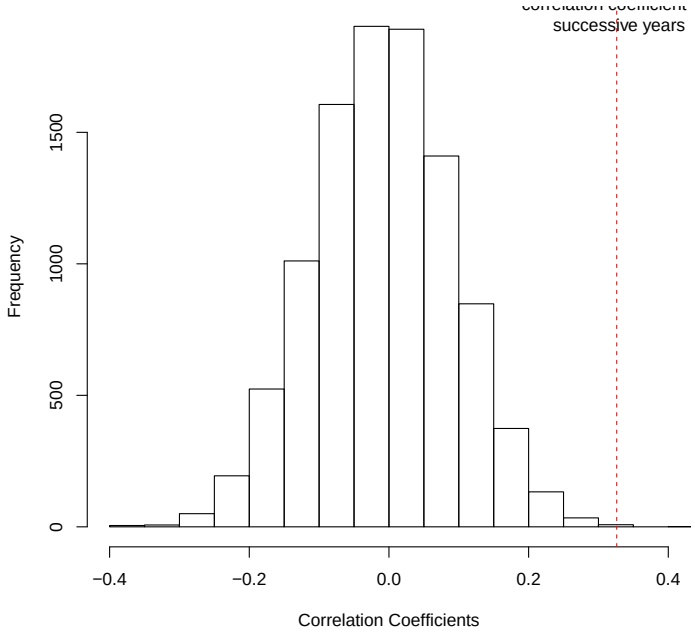

Figure 2: Hisotgram showing frequency of correlation results over the 10,000 perumtations

Using the result from the final line of code, we can see how signifcant this value is. Since the numbers are generated at random, the number of permutations that show higher correlation than successive years fluctuates between 4 and 14 each time you run the code, giving a range of significance values of 0.0004 to 0.0014, both of which are far below the required 0.05 and indicate a strong significance in the correlation between temperatures of one year and the next. This can be seen clearly from Figure 2, as well as the distribution of the correlation coefficients over the 10,000 permutations, which roughly centres around 0 (no correlation).