HOMEWORK ASSIGNMENT 4

Zachary Lazerick

16 February 2023

PROBLEM 1: Describing Association (Part 1)

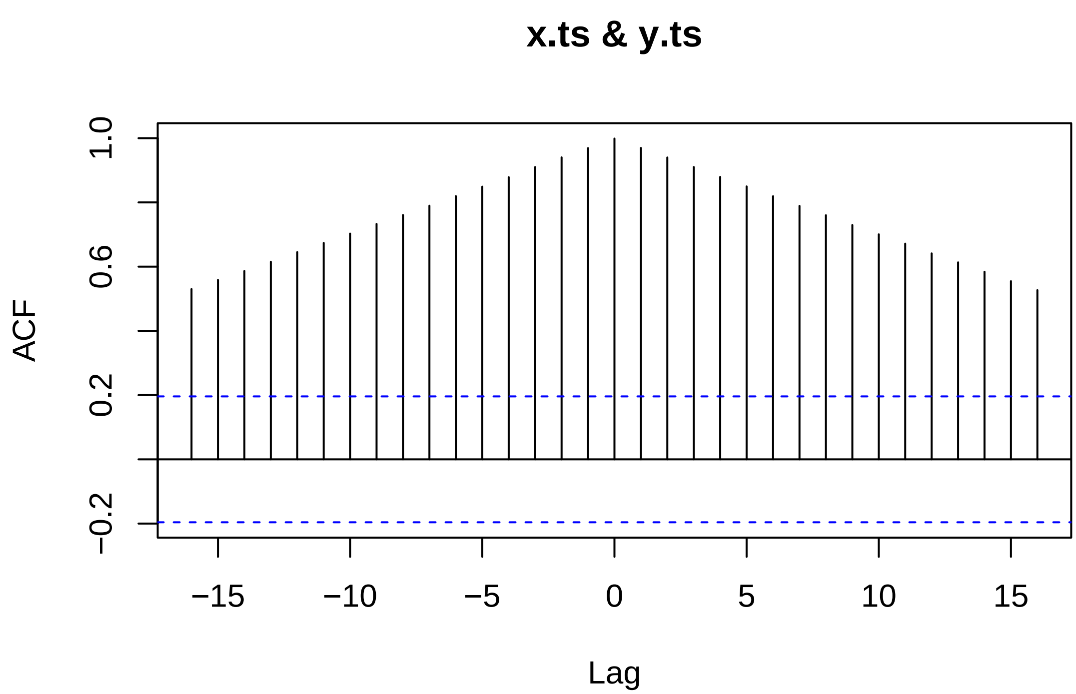
(a) Define the following quantities in R for k = 1:

```
> w <- 1:100
> x <- w + k * rnorm(100)
> y <- w + k * rnorm(100)
```

```r
k = 1; w <- 1:100
x <- w + k * rnorm(100); y <- w + k * rnorm(100)
x.ts <- ts(x); y.ts <- ts(y)
```

(b) Produce a ccf plot of the two time series x and y. What do you notice?

```r
xy.ccf <- ccf(x.ts, y.ts)
```

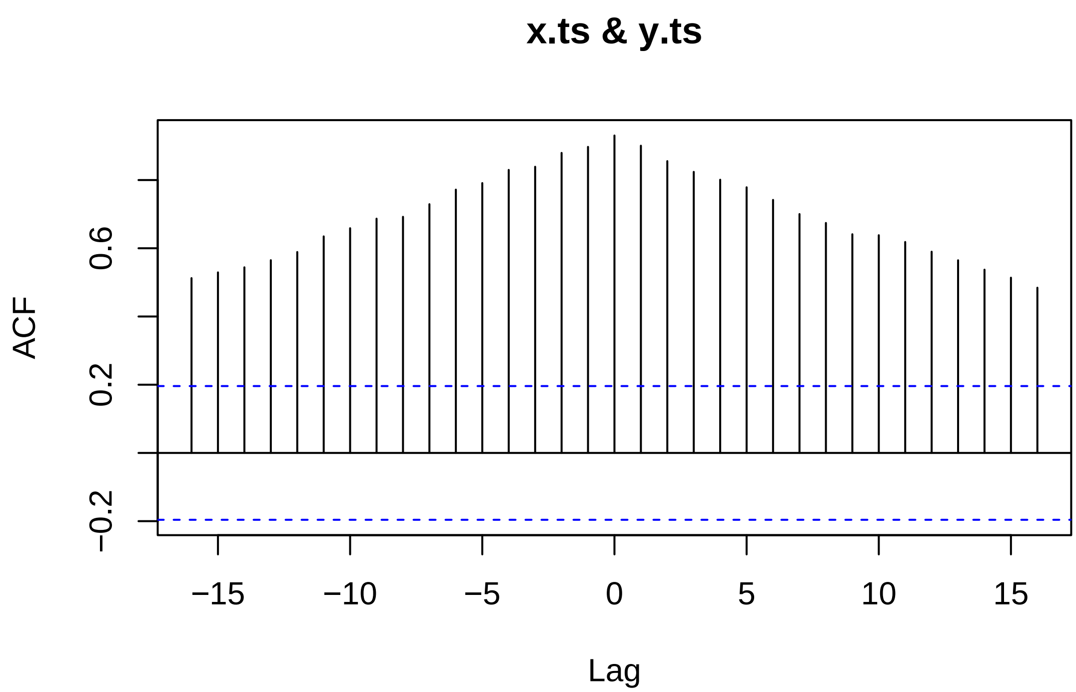


For positive values of lag, as lag increases, the acf at said lag decreases. As for negative values of lag, as lag decreases, the acf at said lag decreases. The rate of change in acf appears to be the same regardless of lag direction, suggesting that neither variable is leading or lagging behind the other.

(c) Repeat the above for `k = 10`:

```
k = 10; w <- 1:100
x <- w + k * rnorm(100); y <- w + k * rnorm(100)
x.ts <- ts(x); y.ts <- ts(y)

xy.ccf <- ccf(x.ts, y.ts)
```

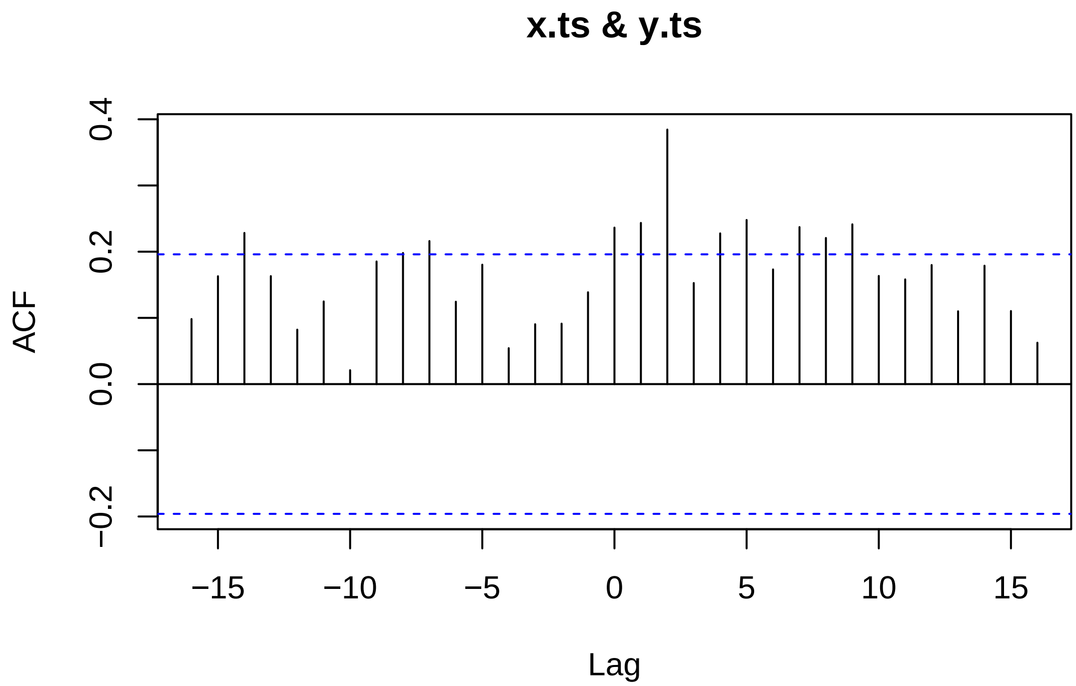


This has similar behavior to the ccf for $k = 1$. For positive values of lag, acf decreases, and for negative values of lag, acf decreases also. However, unlike the previous ccf plot, the decreases are not uniform and symmetric. Despite this, the correlation is significant at all levels of lag.

(d) Repeat the above for `k = 50`:

```
k = 50; w <- 1:100
x <- w + k * rnorm(100); y <- w + k * rnorm(100)
x.ts <- ts(x); y.ts <- ts(y)

xy.ccf <- ccf(x.ts, y.ts)
```



This plot is not similar to the previous two. This is because the correlation between the variables is significant at several different levels of lag. Also, the values of correlation are the smallest among the two graphs. This suggests that the effect of rnorm() on the creation of the variables x and y has a more pronounced effect the greater the scalar k is.

PROBLEM 2: Describing Association (Part 2)

(a) Define the following quantities in R:

```
> Time <- 1:370
> x <- sin(2 * pi * Time / 37)
> y <- sin(2 * pi * (Time + 4) / 37)
```

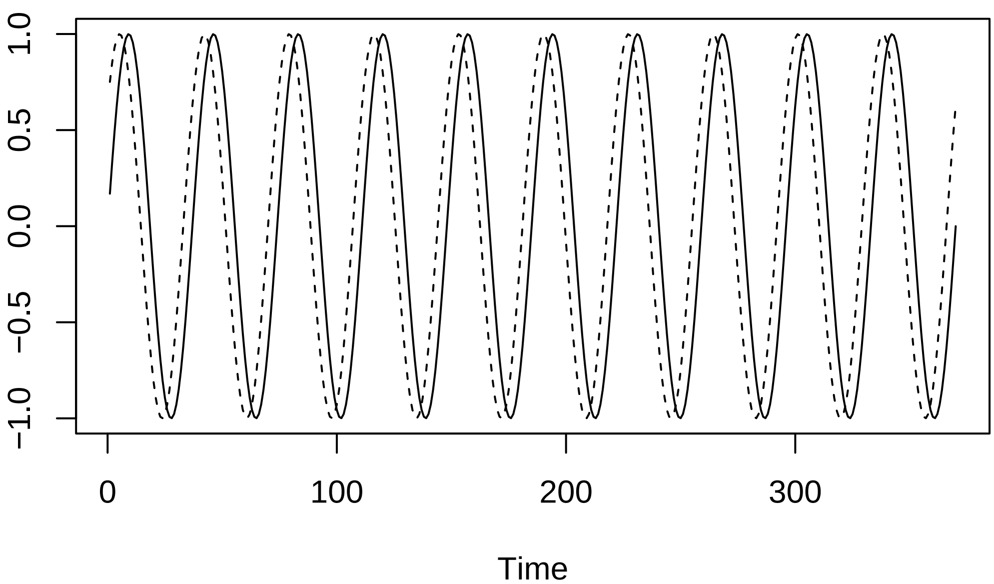
```
Time <- 1:370
x <- sin(2 * pi * Time / 37); y <- sin(2 * pi * (Time + 4) / 37)
```

(b) Based on the equations alone, what is the period of each quantity? What is the lag between them?

Based on the equations alone, the period of each quantity is 2π. The lag between them is 4.

(c) Convert each variable to a time series using the `ts()` function. Then plot them on the same plot using `ts.plot()` and `cbind()`. Is your answer to part (b) confirmed?

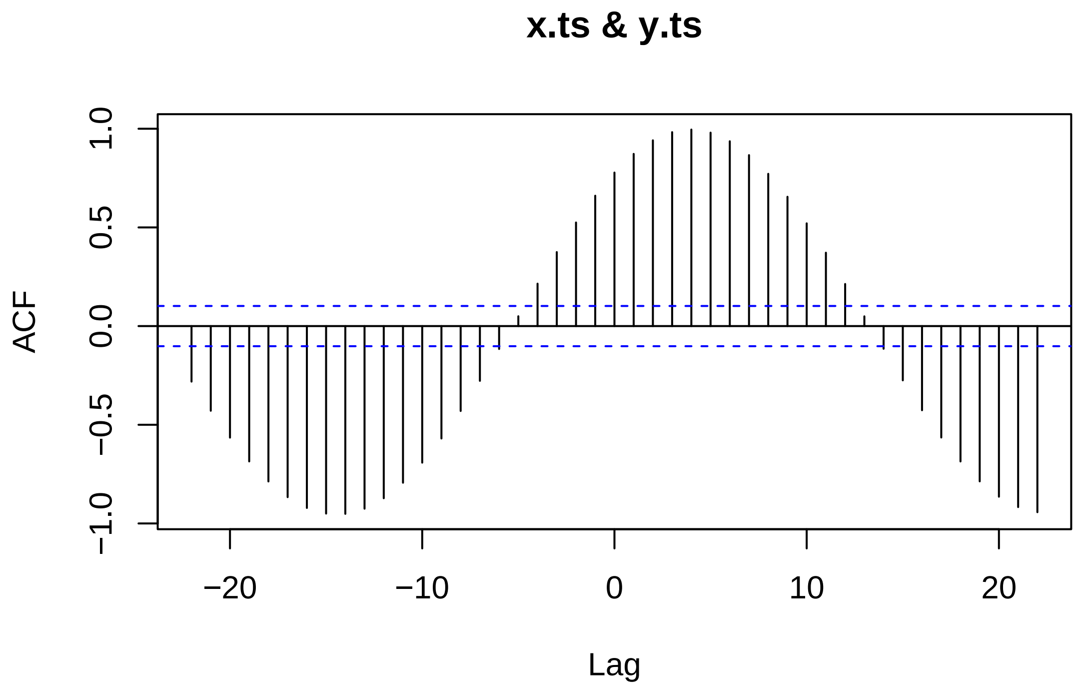
```
x.ts <- ts(x); y.ts <- ts(y)
ts.plot(cbind(x.ts, y.ts), lty = 1:2)
```



Yes, my answer is confirmed. Both series have the same shape and structure. Y leads X by 4 time intervals.

(d) Produce a `ccf` plot of the two time series `x` and `y`. What do you notice?

```
xy.ccf <- ccf(x.ts, y.ts)
```

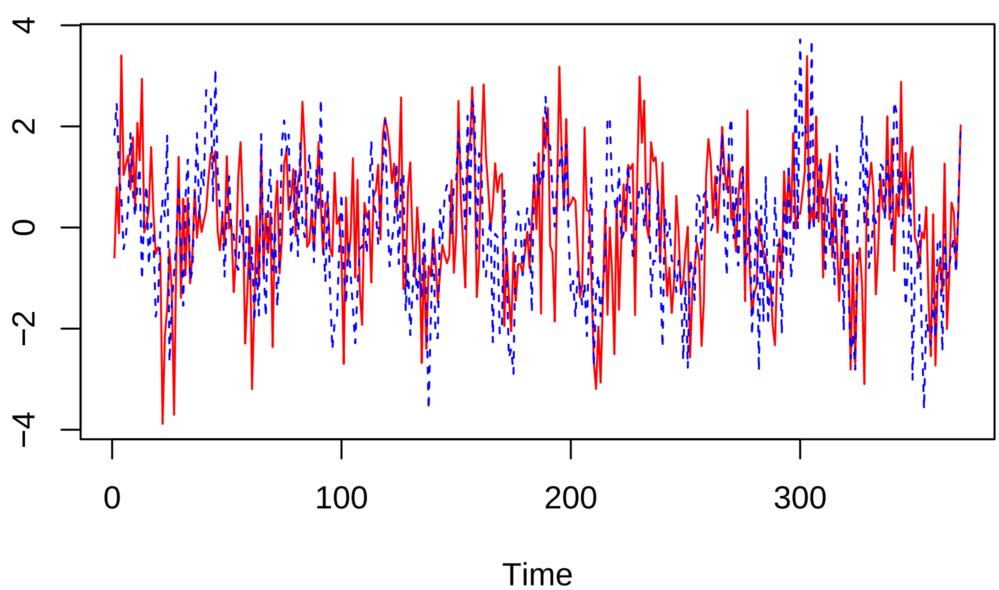


I notice a sin-like wave in the ccf plot. This suggests that x and y are going in-and-out of phase. This makes sense because of how the variables were defined. The changes in correlation between lags appear to be largely uniform.

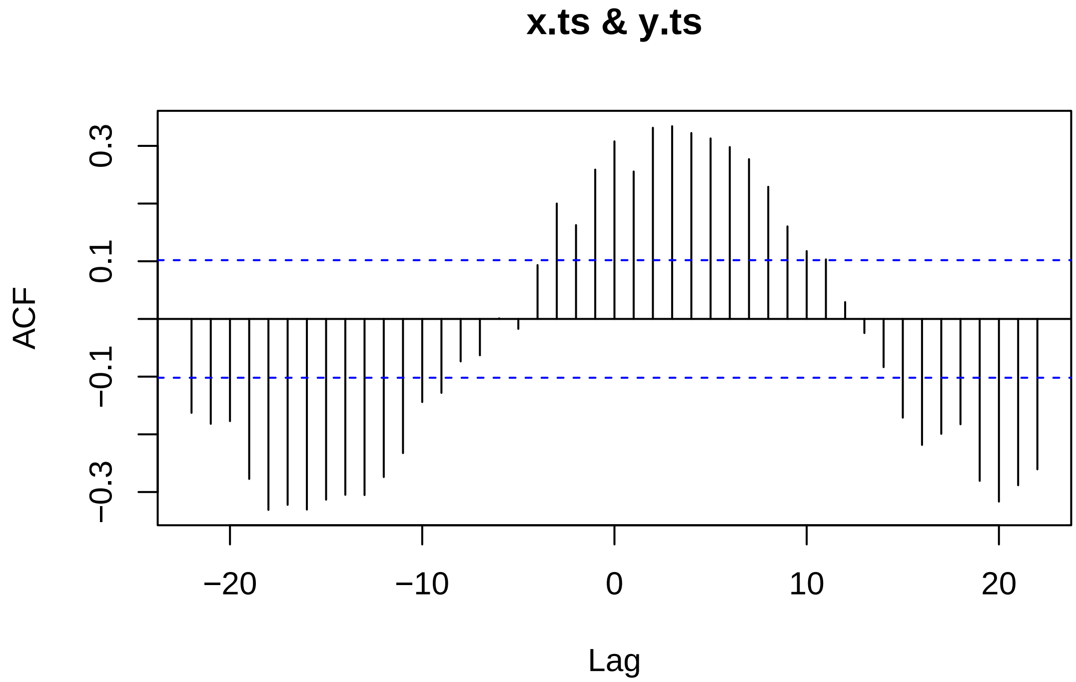
(e) Investigate the effect of adding independent random variation to both `x` and `y` using `rnorm()`. Describe the results.

```
x <- x + rnorm(370); y <- y + rnorm(370)
x.ts <- ts(x); y.ts <- ts(y)

ts.plot(cbind(x.ts, y.ts), lty = 1:2, col = c('red', 'blue'))
```



```r
xy.ccf <- ccf(x.ts, y.ts)
```



With the random variation term added, the overall shape of the ccf plot remains the same. It is still shaped like a sin-wave. However, the changes in correlation as lag increments by 1 in either direction, are not consistent. This is the work of the random error term at play.