Name: Your name here
Due: 2024/09/30

Homework 3

Be sure to submit **both** the .pdf and .qmd file to Canvas by Monday, September 30th at 11:59 pm.

0. [1 pt] With whom did you work on this assignment?

1. [11 pt] We will focus on a data set describing weekly avocado sales volume and price in the United States between 2015 and 2018 for this question.

   a) [2 pt] Read the data in (naming it `avocado`), filter to sales of conventional avocados in Las Vegas, and create a new column, called `volume1000`, that represents the total volume of sales in 1000s.

   b) [3 pt] Create a `ts` object with the `volume1000` vector (called `avo_ts`), create an additive decomposition of that time series (called `avo_decomp`), and plot that decomposition.

   > **Tip**
   >
   > Be sure to pay attention to how the data set is arranged with respect to date.

   c) [1 pt] Describe the time series decomposition in terms of its trend and seasonal components.

   d) [1 pt] You should notice that the trend and random components are significantly (in the english way, not the statistics way) shorter than the other two series. Why is this?

   e) [1 pt] You should see a rather large spike in avocado sales during the beginning of each year in the decomposition. Hypothesize an explanation for this spike.

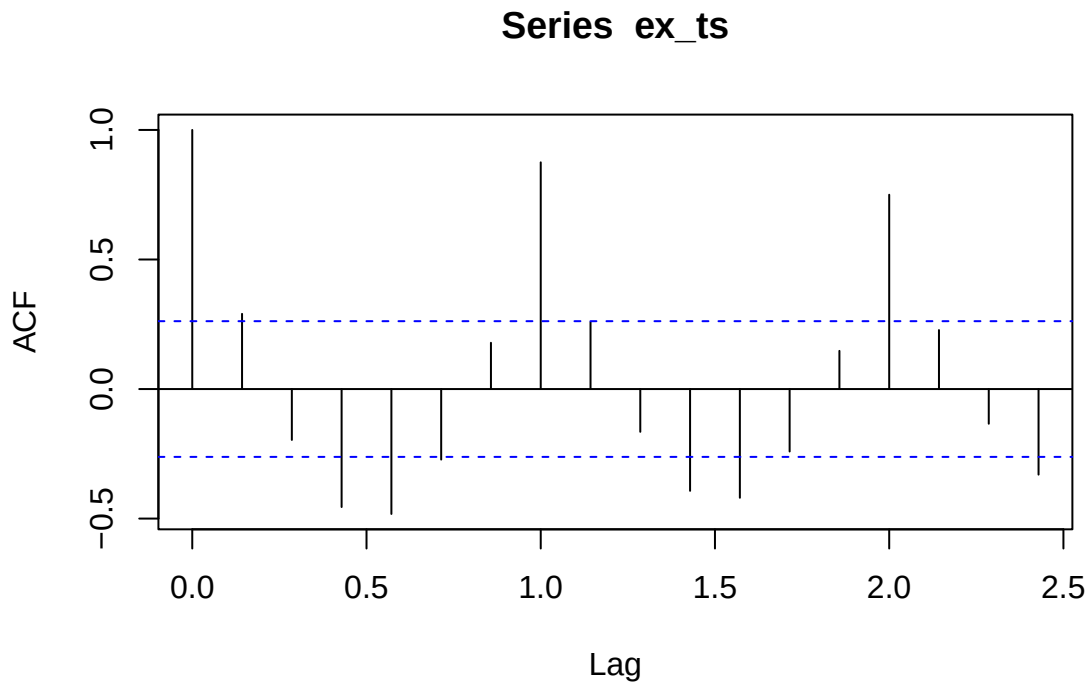
   f) [1 pt] The code below attempts to create a data frame of the random component over time, but R kicks back an error. Investigate the source of this error, and describe what is causing the problem.

```
res_tbl <- tibble(
  dt = avocado$Date,
  res = avo_decomp$random
)
```

g) [2 pt] We are going to ignore the problem for now (:]). Provide a correlogram of the residual error series and comment on whether there appears to be leftover serial correlation.

2. [1 pt] The code below creates a time series out of the same seven values on repeat, and generates a correlogram for that series. Describe what is happening in the correlogram at time lags that are a multiple of seven. Why should that make sense?

```
ex_ts <- ts(
  rep(-3:3, 8),
  start = c(1,1),
  end = c(8, 7),
  freq = 7
)
acf(ex_ts)
```



3. [6 pt] The rest of the questions on this assignment are fairly math-heavy, but in order to understand time series analysis, it is essential to understand the concepts of stationarity, covariance, and correlation. If the following questions feel hard, *that is okay* - it does not mean you cannot learn time series. I encourage you to ask for help!

Note: When I use the word *prove* below, I do not mean in a rigorous mathematical sense (though, by all means give it a shot!). I only ask that you provide logic and sound mathematical reasoning (meaning I should see an equation or two in each response).

a) [3 pt] *Prove* that a model cannot be second-order stationary if the model is not stationary in the mean.

> Hint
>
> What do we know about the autocovariance function, $\gamma(k)$ for a second-order stationary process?

b) [3 pt] Recall that the population autocorrelation function is given by $\rho_k = \frac{\gamma_k}{\sigma^2}$. *Prove* that ρ_0 must always equal 1.

4. [4 pt] (This question is admittedly a nasty piece of work, but valuable to understand). Recall that the sample autocorrelation function, r_k, is defined as

$$r_k = \frac{c_k}{c_0}$$

but the population autocorrelation function, ρ_k is defined as

$$\rho_k = \frac{\gamma_k}{\sigma^2}.$$

Why do we define $r_k = \frac{c_k}{c_0}$, rather than $r_k = \frac{c_k}{s^2}$, which is arguably more natural? Support your answer with something proof-ish.