Question 7.1

Describe a situation or problem from your job, everyday life, current events, etc., for which exponential smoothing would be appropriate. What data would you need? Would you expect the value of α (the first smoothing parameter) to be closer to 0 or 1, and why?

Answer:

Exponential smoothing is a forecasting method for time series data, especially for the ones following a systematic trend or seasonality. One good example of the usages of exponential smoothing will be the prediction of the number of customers in a restaurant, based on the observations of the previous records.

In this example, we could think that all of the randomness, seasonality and trend can be relevant and important for the forecast. Here is a possibility that each factor influences the number of customers.

- **Randomness:** if the menu was on air, or suddenly became viral on the social media, there can be more demand than usual. However, as it is not a usual case, we can consider α closer to 1, giving more trust to the history of the business.
- **Seasonality:** when it is weekend, there can be more customers than the usual weekday average. And as it will be a repeating pattern, seasonality should be considered into when using exponential smoothing method.
- **Trend:** trend refers to the slope in the data. For example, we could think of a successful promotion strategy making a result. As the place becomes popular, the business will constantly grow. We can decide how quickly the model will react to the change in the trend by increasing the beta value.

Question 7.2

Using the 20 years of daily high temperature data for Atlanta (July through October) from Question 6.2 (file `temps.txt`), **build and use an exponential smoothing model** to help make a judgment of **whether the unofficial end of summer has gotten later over the 20 years**. (Part of the point of this assignment is for you to think about how you might use exponential smoothing to answer this question. Feel free to combine it with other models if you'd like to. There's certainly more than one reasonable approach.)
Note: in R, you can use either `HoltWinters` (simpler to use) or the `smooth` package's `es` function (harder to use, but more general). If you use `es`, the Holt-Winters model uses `model="AAM"` in the function call (the first and second constants are used "A"dditively, and the third (seasonality) is used "M"ultiplicatively; the documentation doesn't make that clear).

<u>Answer:</u>

In this question, I first tried to fit the temperature data from 1996 to 2015 with the Holt-Winters method to see the development of the highest temperature of the days for the 20 years. Then, in order to answer verify if the unofficial summer has gotten later, I compared the prediction from 2006 to 2016 by Holt-Winters model with the fitted values of the same period. I used CUSUM to see if there was any difference in the end of unofficial summer in the predictions and the fitted value.

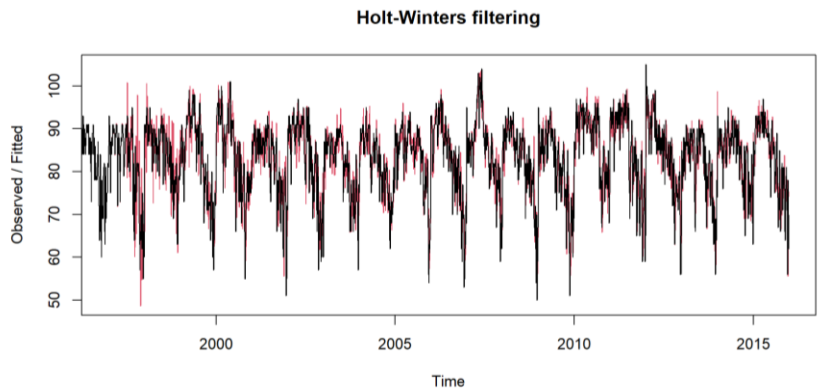
1. Generate a model with Holt-Winters method with the data from 1996 - 2015

First of all, I wanted to see the smoothed development of the highest temperature from 1996 to 2015 using Holt-Winters method. The first task to be done before modelling was transforming the data into a time series data. As I wanted the model to find the best α, β, γ value by minimizing the residual sum of the squares, I set the variables to NULL. Also, as the seasonal variations are almost constant throughout the series, I used "additive" method for building the model.

```
## make the data into time series
temps_vec <- unlist(temps_1)
temps_ts <- ts(temps_vec, start = 1996, frequency = 123)
## run the Holt Winter method
model <- HoltWinters(temps_ts, alpha = NULL, beta = NULL,
                     gamma = NULL, seasonal = "additive")
plot(model)
```

```
Smoothing parameters:
 alpha: 0.6610618
 beta : 0
 gamma: 0.6248076
```

The result of the exponential smoothing is as in the graph below. From the visualization, I could not observe any explicit trend of the temperature going up in general.



2. Verifying if there is any sign of unofficial summer getting later using CUSUM

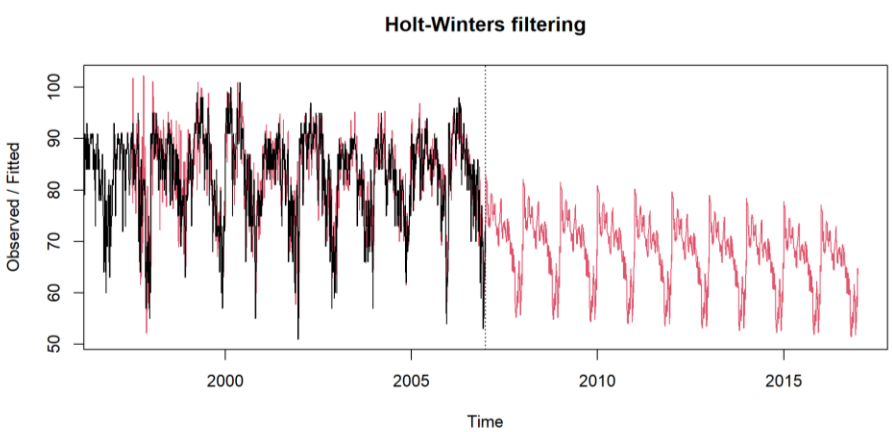
In order to answer this question, I thought of making a prediction for 2006 - 2015 using Holt-Winters model built from 1996 – 2005 data, as the forecasted values from Holt-Winters would reflect the trend and seasonality of the period. By comparing the actual development of the temperature and the forecasted value, we could have an idea on how much the actual temperatures deviated from the historical development.

First, I built the Holt-Winters model using 10-year data from 1996 to 2005. And based on the model, the temperatures for the next 10 years (2006 - 2015) were predicted.

```
### run Holt Winter with the data from 1996 to 2005

temps_2 <- temps[,2:12]
head(temps_2)
temps_vec_2 <- unlist(temps_2)
temps_ts_base <- ts(temps_vec_2, start = 1996, frequency = 123)

model_2 <- HoltWinters(temps_ts_base, alpha = NULL, beta = NULL,
                    gamma = NULL, seasonal = "additive")

# prediction for the next 10 years
temp_pred <- predict(model_2, 1230)
plot(temp_pred)
```

```
Smoothing parameters:
 alpha: 0.6386924
 beta : 0
 gamma: 0.7263512
```

The red line from the year 2006 in the graph below is the predicted temperature for the next 10 years. We could observe that the prediction in general is lower than the fitted value from 1996 to 2005. Using the same parameters for α, β, γ, the actual data from 2006 to 2015 were fitted as well.

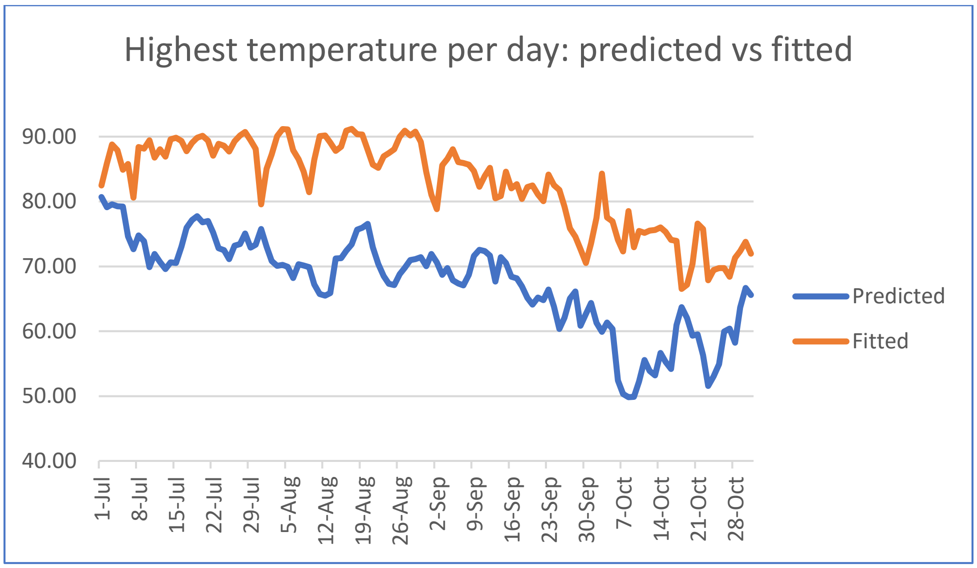


As for the next step, I tried to find the average of the highest temperature of each day, for both 2006 – 2015 fitted and predicted values. Then, I wrote the saved vector into csv file to continue with the CUSUM analysis on Excel if there was any delay in the unofficial end of the summer.

```
## find the mean highest temperature of each day for 10 years

temp_pred_mean <- rep(0, 123)
temp_day <- rep(0,10)

for (i in 1:123){
  for (j in 1:10){
    k = i + 123 * (j - 1)
    temp_day[j] <- temp_pred[k,]
  }
  temp_pred_mean[i] <- mean(temp_day)
}

temp_pred_mean
```

I attached my CUSUM analysis as a separate Excel file. I did CUSUM separately on the predicted and fitted value, using a mu value with the average of the July temperatures each. C and T value were set to 5 and 25 for both of the data sets. The unofficial end of summer of the predicted data was at 23rd September, while that of the fitted data of the actual temperatures was at 28th September. So, the summer was longer actually from 2006 to 2015 by 5 days than the prediction based on 1996 – 2005. However, it is hard to say that there was a huge delay.

The graph below is the comparison of the temperatures from the predicted and fitted data. We can see that the fitted data, which means the actual temperature, was higher than the prediction. So, we could conclude that meanwhile the unofficial end of the summer was not delayed, we could still observe that the temperature was higher than how it is expected to be, based on 1996 – 2005 observations.



To end the analysis, there is one thing that can be improved from this analysis. The line graph on the left is the prediction of the 10 years from 1996 – 2005 data, and the one on the right side is the prediction of the 10 years from 1996 – 2015 data. As you can see, when we included the year 2006 – 2015, the temperatures are in general higher. However, when we just fitted the 20-year data in the beginning of the analysis, we did not observe an obvious trend of temperature going up. This is the point where we need further investigations or improvement to the model.