Homework 3

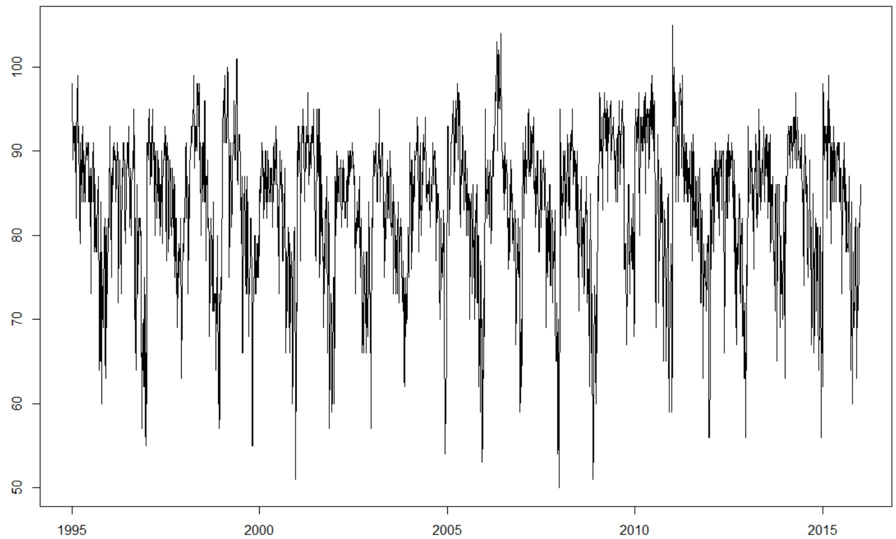
Question 7.1
Describe a situation or problem from your job, everyday life, current events, etc., for which exponential smoothing would be appropriate. What data would you need? Would you expect the value of α (the first smoothing parameter) to be closer to 0 or 1, and why?

Answer:
Since we are dealing with time series, the data that is being collected is sequential. For example, we are collecting data from a site that posts the current value of bitcoin. Exponential smoothing can very handy in these situations. Since the price of bitcoin is set by the market, and it can volatile at times, hence it's safe to assume that α should be closer to 1. And we'll trust are previous estimate $S_t - 1$

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).

Answer: First, lets load the data and plot it using ts().

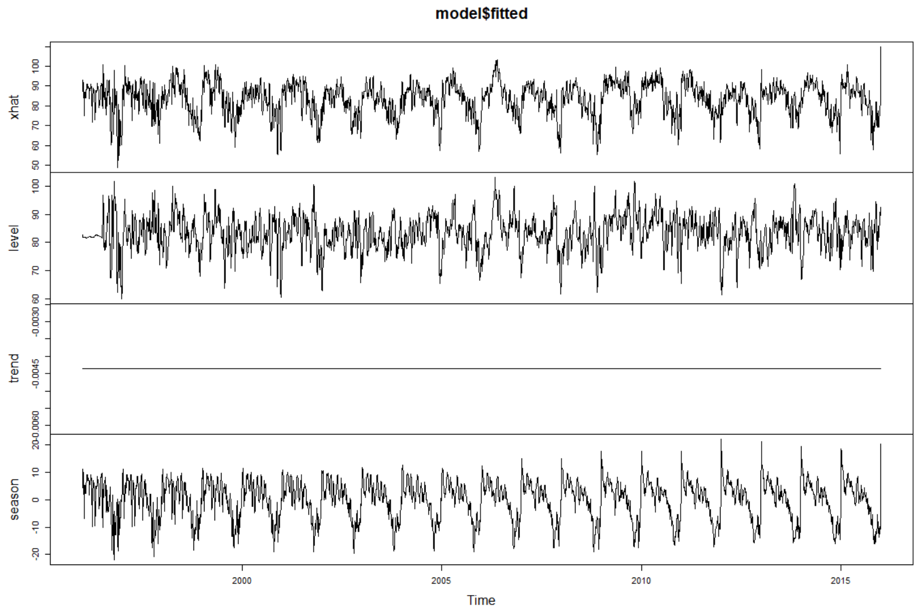
```
# load our data
data3<- read.table("6.2tempsSummer2018.txt",stringsAsFactors = FALSE,header =TRUE)
crimestat<- ts(as.vector(unlist(data3[,-1])),frequency=123,start=1995,end = 2016)
plot(crimestat)
```



Next we'll use HoltWinters(), and since we know there is some associate randomness to it, and that it's seasonal, we'll let R function decide our Alpha value.

```
model<- HoltWinters(crimestat)
plot(model)
model$SSE
model$coefficients
plot(model$fitted)
```

The alpha I got was 0.6566, and gamma= .599 with an SSE 70265




```
> model$coefficients
            a             b            s1            s2            s3            s4            s5            s6            s7            s8            s9 
 73.925574617  -0.004362918  18.216428214  13.470486156  11.944871460  12.292692713  12.312864902  11.109882975   9.866595063   9.156906982   6.566089029 
          s10           s11           s12           s13           s14           s15           s16           s17           s18           s19           s20 
  3.243814051   4.563486239   3.464816073   3.195197001  -0.170523259   3.636175181   7.071313106   5.482445141   7.485377121   7.097571731   7.554282184 
          s21           s22           s23           s24           s25           s26           s27           s28           s29           s30           s31 
  9.748063689   8.103755983   8.635998586   5.752932351   6.340960891   5.869994715   5.615399266   6.995146612   7.477143549   4.651120050   4.647499702 
          s32           s33           s34           s35           s36           s37           s38           s39           s40           s41           s42 
  7.243811294   5.063365887   6.104775453   4.916520069   1.844314629   2.959875732   1.623876018   3.762771950   2.389597099   0.326007765  -0.502223489 
          s43           s44           s45           s46           s47           s48           s49           s50           s51           s52           s53 
 -1.710768557  -1.349362373  -1.219018945   0.081094365   1.600594482   4.464145064   6.038054936   7.103049587   8.348798856   8.468244610   5.388415843 
          s54           s55           s56           s57           s58           s59           s60           s61           s62           s63           s64 
  2.574757295  -1.833812744   1.451313829   0.560641734   0.857405199   2.241979709  -0.805247246   3.621936991   4.400010084   2.712620292   6.233827276 
          s65           s66           s67           s68           s69           s70           s71           s72           s73           s74           s75 
  0.775077694   1.355136231   1.399034429   1.232451991   0.868575820   2.493797309   1.531877360   4.051513046   3.444494107  -0.393631197   0.883099240 
          s76           s77           s78           s79           s80           s81           s82           s83           s84           s85           s86 
  3.751238432   1.593999267   0.224925421  -1.694242716   0.026657372   0.447680665  -2.666408450  -1.419201438  -0.966299092  -4.266784221  -4.125074077 
          s87           s88           s89           s90           s91           s92           s93           s94           s95           s96           s97 
 -3.441611491  -3.674269872  -4.296998862  -4.965977827  -4.900079856  -7.108443258  -7.167700947  -5.978988087  -9.954083942  -7.824242105  -6.583006108 
          s98           s99          s100          s101          s102          s103          s104          s105          s106          s107          s108 
 -9.385317001  -7.854064600 -13.229722369 -15.105595590 -15.577156919 -14.300094064 -12.393635404 -13.889398338 -14.321091607 -13.725301315 -11.754919596 
         s109          s110          s111          s112          s113          s114          s115          s116          s117          s118          s119 
-15.076379257 -14.173365002  -9.256194503  -3.997967915  -5.676145627 -10.248050695 -10.254203353 -12.747618552 -10.461010084 -12.052170605  -9.028958544 
         s120          s121          s122          s123 
-11.395093619  -9.757278981  -8.162953051  15.369867491 
```

As we can see above, we can extract an important information from the function HoltWinter(), the coefficients up to s82 (September 20th) are all positive, and from s82 onward, all negative coefficients. This tells us the summer officially ends on s82, September 20th.

Question 8.1

Describe a situation or problem from your job, everyday life, current events, etc., for which a linear regression model would be appropriate. List some (up to 5) predictors that you might use.

Answer:

Let the variable of interest be the number of traffic light a driver encounters. Here are some possible predictors that may be considered. Distance the driver traveled Time spent driving Categorical response: '0' for majority time spent on the interstate and '1' for non-interstate road The current density population of the driver's city Type of car the driver drives.

Question 8.2

Answer: First we note that the lower the AIC value the better the model. So, initially I load the txt file and use the function glm(). And we got an AIC of 650.

```
data2<- read.table("5.1uscrimeSummer2018.txt",stringsAsFactors = FALSE,header = TRUE)
fit <-glm(data2$Crime~., data=data2)
summary(fit)
```

```
Coefficients:
(Intercept)           M          So          Ed         Po1         Po2          LF         M.F         Pop          NW          U1          U2
 -5.984e+03   8.783e+01  -3.803e+00   1.883e+02   1.928e+02  -1.094e+02  -6.638e+02   1.741e+01  -7.330e-01   4.204e+00  -5.827e+03   1.678e+02
     Wealth        Ineq        Prob        Time
  9.617e-02   7.067e+01  -4.855e+03  -3.479e+00

Degrees of Freedom: 46 Total (i.e. Null);  31 Residual
Null Deviance:      6881000
Residual Deviance: 1355000     AIC: 650
```

To get a better model with a lower AIC, I scaled the dataset. Using scale() function. Afterwards I got an AIC of 89.99.

```
Coefficients:
(Intercept)          V1          V2          V3          V4          V5          V6          V7          V8          V9         V10         V11
 -1.697e-16   2.854e-01  -4.710e-03   5.447e-01   1.482e+00  -7.911e-01  -6.936e-02   1.326e-01  -7.215e-02   1.118e-01  -2.716e-01   3.664e-01
        V12         V13         V14         V15
  2.399e-01   7.290e-01  -2.854e-01  -6.375e-02

Degrees of Freedom: 46 Total (i.e. Null);  31 Residual
Null Deviance:      46
Residual Deviance: 9.058      AIC: 89.99
```

Hence our model is:
$\hat{y}$=**-1.697448e-16 + 2.853992e-01(M)+ -4.710274e-03(So)+ 5.447226e-01(Ed) + 1.481515e+00 (Po1)+ -7.910746e-01(Po2)+ -6.936144e-02(LF) + 1.326225e-01(MF)+ -7.215404e-02(Pop)+ 1.117842e-01 (NW)+ -2.716280e-01 (U1)+ 3.664117e-01(U2)+ 2.399190e-01 (Wealth)+ 7.290099e-01(Ineq)+ -2.854309e-01(Prob)+ -6.374822e-02 (Time)**

Next let's insert the sample that we are trying to estimate into our model. Here is our Sample point:
M = 14.0 So = 0 Ed = 10.0 Po1 = 12.0 Po2 = 15.5 LF = 0.640 M.F = 94.0 Pop = 150 NW = 1.1 U1 = 0.120 U2 = 3.6 Wealth = 3200 Ineq = 20.1 Prob = 0.04 Time = 39.0

Inserting the data point into the model:
$\hat{y}$=**-1.697448e-16 + 2.853992e-01(14)+ -4.710274e-03(0)+ 5.447226e-01(10) + 1.481515e+00 (12)+ -7.910746e-01(15.5)+ -6.936144e-02(.640) + 1.326225e-01(94)+ -7.215404e-02(150)+ 1.117842e-01 (1.1)+ -2.716280e-01 (.12)+ 3.664117e-01(3.6)+ 2.399190e-01 (3200)+ 7.290099e-01(20.1)+ -2.854309e-01(0.04)+ -6.374822e-02 (39)**

we get the estimate value $\hat{y}$= **13.971318**.