ILI prediction using Holt-winter method:

Data Description:

Dimensions of data: 1458 obs and 7 variables

Data Field :

Pincode: AKAMRUP, DIBRUGARH, TAWANG, PAPUM, IMPHAL, KHASI, KHASI, KOHIMA, AIZAWA
Here we have 8 distincts pincodes

YEAR
WEEK
ILITOTAL : ILI
wcr : Weekly ratio of ILI
Date
Nform : Confirmed cases

Here ILI is our Response variable. And we are predicting ILI for next 8 weeks according to Pincode.

Tool used : R

Techniques used : Holt- winter forecasting model

Holt- winter:

Time Series techniques work on numerical data collected over a considerable period of time. It is further used to generate future values of the series (termed as forecast).

Holt's method is used to capture seasonality. The Holt-Winters seasonal method comprises the forecast equation and three smoothing equations — one for the level ℓtℓt, one for trend btbt, and one for the seasonal component denoted by stst, with smoothing parameters α, β∗ and γ.

There are two methods: The additive method is preferred when the seasonal variations are roughly constant through the series, while the multiplicative method is preferred when the seasonal variations are changing proportional to the level of the series.

Seasonal and Non-Seasonal Patterns

Seasonality is a component of a Time Series, which defines the repetitive movement around the trend line in a specific period of time. It is measured for time intervals of days, weeks, months, or quarters.
Non-Seasonal Time Series doesn’t follow any trend.

To estimate the trend component on seasonal or non-seasonal Time Series, Holt-Winters’ smoothing methods are used.

Holt-Winters

Holt-Winter is used for exponential smoothing to make short-term forecasts by using “additive” or “multiplicative” models with increasing or decreasing trend and seasonality. Smoothing is measured by beta and gamma parameters in Holt’s model.

- If the beta parameter is set to FALSE, the function performs exponential smoothing
- The gamma parameter is used for the seasonal component. If the gamma parameter is set to FALSE, a non-seasonal model is fitted

When the gamma and beta values are set between 0 and 1, the values close to 0 (zero) specifies that weight is placed on the most recent observations while constructing the forecast of future values.

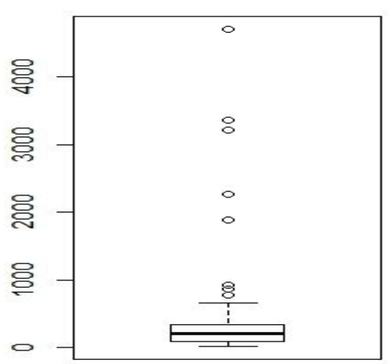
Here we’ll build holt winter forecasting model on whole data set.

```
#import data
data2=read.csv("C:/Users/shwetag/Downloads/ILI_Zip.csv",header=T)

head(data2,4)
  Pincode    YEAR   WEEK   ILITOTAL      wcr        Date
1  781014  1997   40       44             1         1997-09-29
2  786001  1997   40       28             1         1997-09-29
3  791111   1997   40       79             1         1997-09-29
4  790104  1997   40       37             1         1997-09-29


 # Rename all levels of pincode
 data2$Pincode[data2$Pincode=="781014"] <- "AKAMRUP"
 data2$Pincode[data2$Pincode=="786001"] <- "DIBRUGARH"
 data2$Pincode[data2$Pincode=="791111"] <- "TAWANG"
 data2$Pincode[data2$Pincode=="790104"] <- "PAPUM"
 data2$Pincode[data2$Pincode=="795001"] <- "IMPHAL"
 data2$Pincode[data2$Pincode=="793108"] <- "WKHASI"
 data2$Pincode[data2$Pincode=="793005"] <- "EKHASI"
 data2$Pincode[data2$Pincode=="797001"] <- "KOHIMA"
 data2$Pincode[data2$Pincode=="796001"] <- "AIZAWL"


 #create dummy nform variable
 median(data2$ILITOTAL)
 [1] 469.5
 nform =sample(rpois(1,469),1458,replace=T)
 length(nform)
 [1] 1458
 #create dataframe with nform
 data_new=data.frame(data2,nform)


> dim(data_new)
[1] 1458    7

> str(data_new)
'data.frame':      1458 obs. of  7 variables:
 $ Pincode : chr  "AKAMRUP" "DIBRUGARH" "TAWANG" "PAPUM" ...
 $ YEAR    : int  1997 1997 1997 1997 1997 1997 1997 1997 1997 1997 ...
 $ WEEK    : int  40 40 40 40 40 40 40 40 40 41 ...
 $ ILITOTAL: int  44 28 79 37 50 3 5 244 80 57 ...
 $ wcr     : num  1 1 1 1 1 ...
 $ Date    : Factor w/ 162 levels "1997-09-29","1997-10-06",..: 1 1 1 1 1 1 1 1 1 2 ...
 $ nform   : int  467 234 265 463 139 28 34 157 38 367 ...
```

```
#required data
data_new=data_new[,c(1,4,5,6)]

#####.........Time series analysis

#..........................holt winter method
library(stats)
library(forecast)
library(zoo)
library(data.table)
library(lubridate)
```

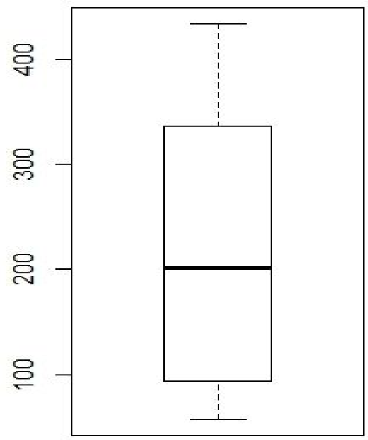
Forecast ILI for next 8 weeks corresponding to AKAMRUP city:

```
#data corresponding to the area AKAMRUP
data3=data2[data2$Pincode=="AKAMRUP",c('Date', 'ILITOTAL')]
boxplot(data3$ILITOTAL)
```

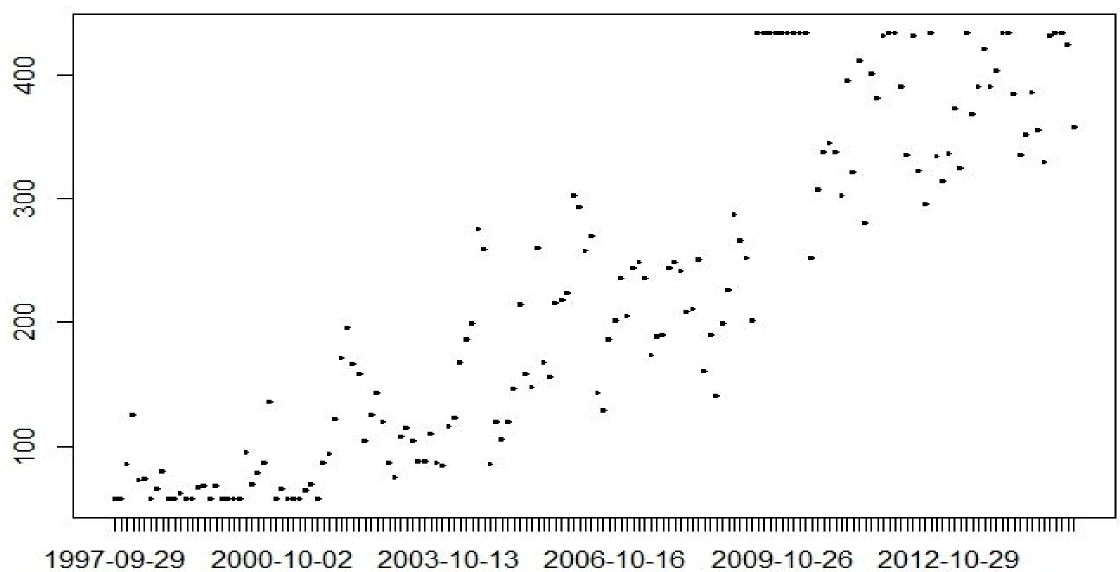


Replacing extreme values with percentiles

```
#replace outliers with percentile

pcap <- function(x){
  for (i in which(sapply(x, is.numeric))) {
    quantiles <- quantile( x[,i], c(.1, .9 ), na.rm =TRUE)
    x[,i] = ifelse(x[,i] < quantiles[1] , quantiles[1], x[,i])
    x[,i] = ifelse(x[,i] > quantiles[2] , quantiles[2], x[,i])}
  x}
data4 = pcap(data3)
boxplot(data4$ILITOTAL)
```



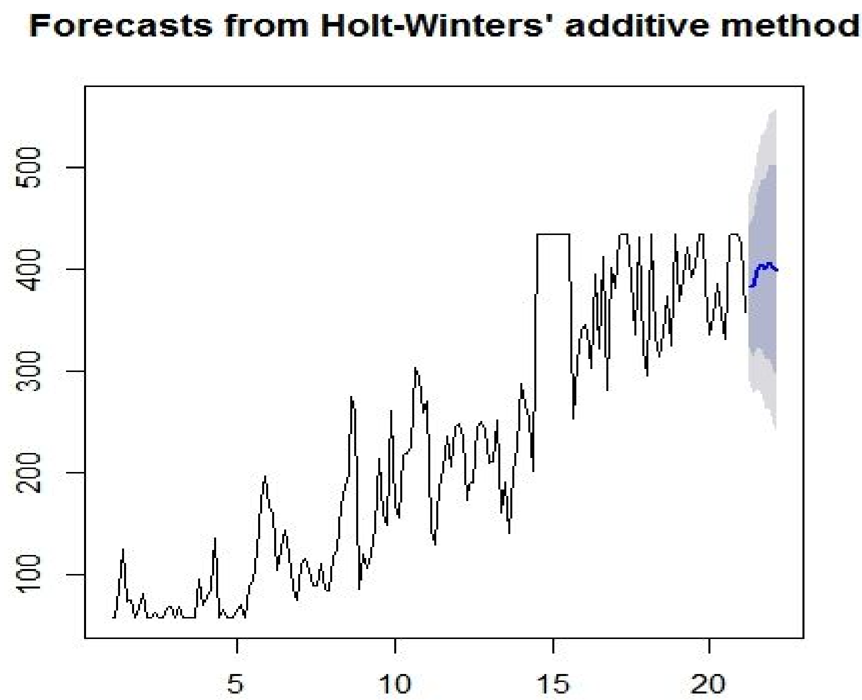
plot(data4$Date , data4$ILITOTAL)



ts = ts(data4$ILITOTAL, freq=8)
model <- hw(ts, h=8)
summary(model)

```
Forecast method: Holt-Winters' additive method

Model Information:
Holt-Winters' additive method

Call:
 hw(y = ts, h = 8)

  Smoothing parameters:
    alpha = 0.5383
    beta  = 1e-04
    gamma = 1e-04

  Initial states:
    l = 82.8436
    b = 1.3645
    s=7.2772 2.0356 8.6147 3.7893 -10.4581 -9.0299
           -3.2584 1.0295

  sigma:  47.1133

     AIC     AICc      BIC
2098.418 2100.878 2138.557

Error measures:
                    ME     RMSE      MAE       MPE     MAPE      MASE       ACF1
Training set 0.9795461 47.11329 33.77645 -4.289569 19.26869 0.5434212 0.04364736

Forecasts:
       Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
21.250       382.5961 322.2180 442.9742 290.2558 474.9365
21.375       382.5434 313.9704 451.1165 277.6700 487.4169
21.500       398.1631 322.2721 474.0542 282.0977 514.2285
21.625       404.3795 321.8138 486.9452 278.1062 530.6528
21.750       399.1910 310.4488 487.9331 263.4716 534.9104
21.875       405.8017 311.2835 500.3199 261.2486 550.3549
22.000       400.9320 300.9688 500.8953 248.0514 553.8127
22.125       398.0341 292.9053 503.1628 237.2535 558.8147
```

plot(model)



Forecast for next 8 weeks :

m1=m[,2]
M[,2]

[1] 322.2180   313.9704   322.2721   321.8138   310.4488   311.2835   300.9688   292.9053

Do the same analysis for remaining Pincodes …..