Time Series Analysis

111-1 Homework #02

> WANG, YIFAN (R10A21126@ntu.edu.tw)
>
> 2022-09-18

> * I use R in Jupyter Notebook

1.

Take what you have simulated for the Q1 of HW#01 and add a proper disturbance to it, for example, $\epsilon_t \sim N(0, 0.1^2)$, to make it more fluctuating and, therefore, unpredictable. Now, let's construct a mathematical model to predict it.

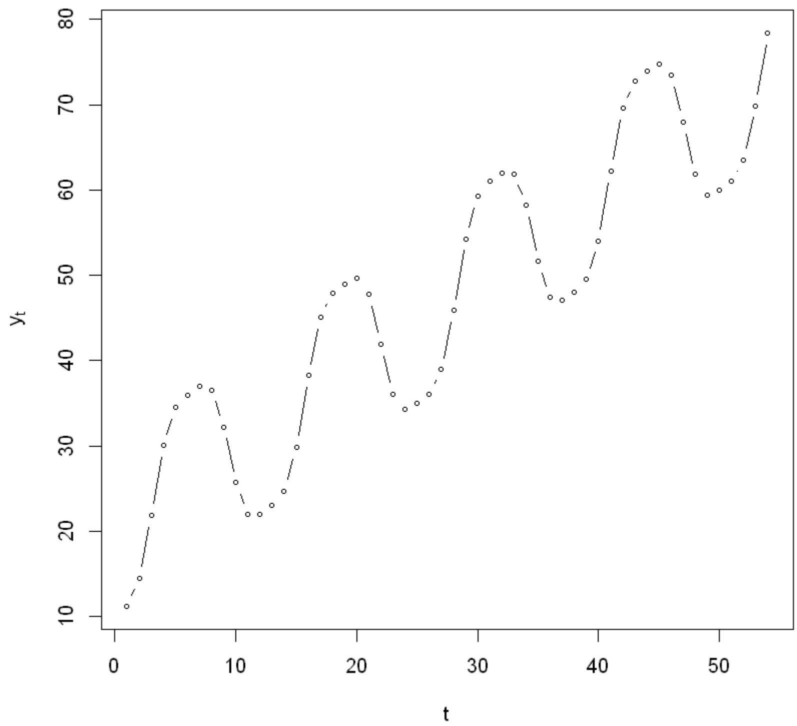
a. Extend the time series to y_{54} using the same formula/procedures in HW#01.
b. Yefine a proper disturbance for your time series.
c. Identify the number of periods in a season and then deseasonalize the series using $y_1, y_2, \ldots, y_{48}$.
d. Calculate the seasonality factors (depending on how many periods in a season) using $y_1, y_2, \ldots, y_{48}$.
e. Finalize the model and validate the model performance via MSE and MAPE using $y_1, y_2, \ldots, y_{48}$.
f. Use the static model to calculate $\hat{y}_{49}, \hat{y}_{50}, \ldots, \hat{y}_{54}$. Compare with the ground truth of $y_{49}, y_{50}, \ldots, y_{54}$ and calculate the prediction MSE and MAPE accordingly.
g. Modify the disturbance in (a) to change series (can be more or less fluctuating). Re-run the procedures (b)-(f). What can you conclude when comparing to the results in (e) & (f) with the previous disturbance.

In [2]:
```
library(latex2exp)
```

1.a Without disturbance

$$Y_t = 20 + t + 10\sin\left[\pi\left(\frac{1}{2}\cos\left(\frac{t}{2}\right) + 1\right)\right] \quad \text{for } t = 1, 2, \ldots, 54$$

In [3]:
```
y_t_a = array(54)
set.seed(23)
for (t in 1:54){
    y_t_a[t] = 20+t + 10*sinpi(cos(t/2)/2+1)
    }
plot(y_t_a,xlab = "t",ylab = TeX(r'($y_t$)'),type = 'b', cex = 0.5)
```



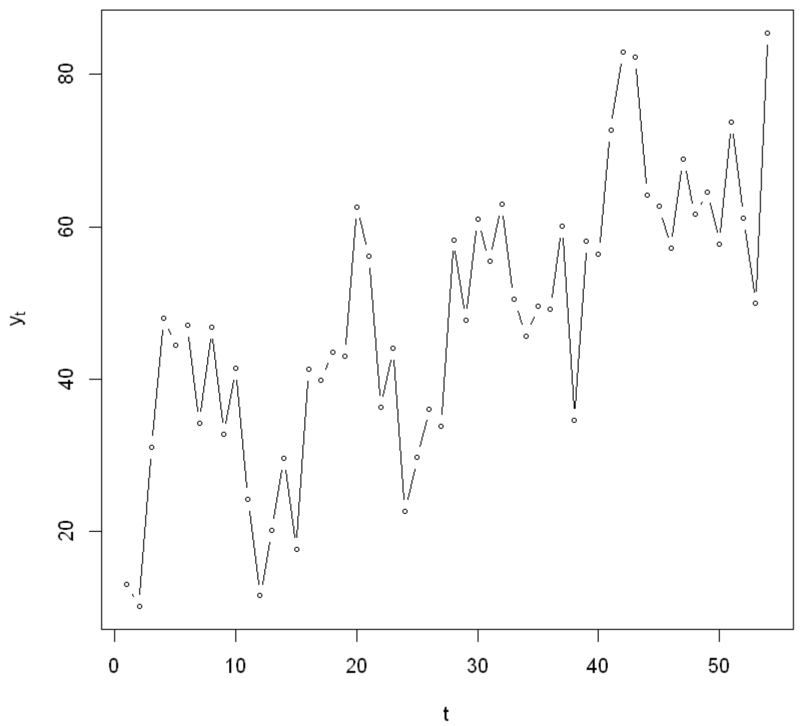
1.b With disturbance

$$Y_t = 20 + t + 10\sin\left[\pi\left(\frac{1}{2}\cos\left(\frac{t}{2}\right)+1\right)\right] + \epsilon_t \quad \text{for } t = 1, 2, \ldots, 54,$$

where ϵ_t is generated from a normal distribution $N(\mu = 0, \sigma = 10)$.

In [4]:

```
y_t = array(54)
set.seed(23)
for (t in 1:54){
    y_t[t] = 20+t + 10*sinpi(cos(t/2)/2+1) + rnorm(1,mean = 0,sd = 10)
    }
plot(y_t,xlab = "t",ylab = TeX(r'($y_t$)'),type = 'b', cex = 0.5)
```



In [5]:
```
y_t
```

13.1164370711084 · 10.1489287556822 · 31.0238185883322 · 48.015015077164 · 44.4821520244021 · 47.0736693472895 · 34.1693647387173 · 46.7482213177055 · 32.7053698699927 · 41.4480231407584 · 24.2118209377626 · 11.5542116966742 · 20.1198751535529 · 29.5534240937656 · 17.6563868461846 · 41.3470325915179 · 39.9069220778366 · 43.479584541564 · 43.0067732250638 · 62.6279726098902 · 56.148110823505 · 36.2703308615064 · 44.0009733335115 · 22.6399975605085 · 29.691859741904 · 36.0949066273609 · 33.8313604626106 · 58.2972884930595 · 47.6849520582198 · 60.9622012499553 · 55.4622206502166 · 62.9597055075933 · 50.4604374962847 · 45.6896547025791 · 49.599782244763 · 49.1650516432742 · 60.1215586604571 · 34.5722199380923 · 58.1207848385037 · 56.3486670112679 · 72.6691015649333 · 82.8718289095743 · 82.2876868639141 · 64.1437449394835 · 62.6439543182224 · 57.1341279733672 · 68.8940367244605 · 61.5980189646744 · 64.5935846403642 · 57.7054183718682 · 73.669970002591 · 61.114626015975 · 49.8998245097115 · 85.4402968242107

1.c

p = 12 (# of periods in a season)
There are 12 periods in a season.
Suppose we collect data every month, we have a 12-month season.

De-seasonalize the series (using $y_1, y_2, \ldots, y_{48}$)

$$\bar{Y}_t = \begin{cases} \frac{\left[Y_{t-(\frac{p}{2})} + Y_{t+(\frac{p}{2})} + \sum_{i=t+1-(\frac{p}{2})}^{t-1+(\frac{p}{2})} 2Y_i\right]}{2p} & p \text{ is even} \\ \frac{\sum_{i=t-(\frac{p-1}{2})}^{t+(\frac{p-1}{2})} Y_i}{p} & p \text{ is odd} \end{cases}$$

In this case ($p = 12$), for $7 \le t \le 42$, we have

$$\bar{Y}_t = \frac{\left[Y_{t-6} + Y_{t+6} + \sum_{i=t+1-6}^{t-1+6} 2Y_i\right]}{2 * 12}$$

For $7 \le t \le 42$, we run regression with t and $\bar{Y}_t$ and get the coefficients
L = 22.03068693
T = 0.915279317

Then we have deseasonalized $\hat{\bar{Y}}_t$ with $\hat{\bar{Y}}_t = L + tT$ for $1 \le t \le 48$

1.d

For $1 \le t \le 48$

$$S_t = \frac{\bar{Y}_t}{\hat{\bar{Y}}_t}$$

Then we get the average of the seasonality effect for each period in a season as $\bar{S}_t$. For example, $\bar{S}_1 = \frac{1}{4}(S_1 + S_{13} + S_{25} + S_{37})$. The values of $\bar{S}_t$ are the same for each season. They are also the same for the testing data.

1.e

For $1 \le t \le 48$

$$\hat{Y}_t = \hat{\bar{Y}}_t * \bar{S}_t$$

$$E_t = Y_t - \hat{Y}_t$$

We now caculate MSE and MAPE for $y_1, y_2, \ldots, y_{48}$

MSE = 82.18
MAPE = 20.30%

1.f

For $49 \le t \le 54$

$$\hat{\bar{Y}}_t = L + tT$$

$$\hat{Y}_t = \hat{\bar{Y}}_t * \bar{S}_t$$

$$E_t = Y_t - \hat{Y}_t$$

We now caculate MSE and MAPE for $y_{49}, y_{50}, \ldots, y_{54}$

MSE = 495.07

MAPE = 33.53%

Conclusion

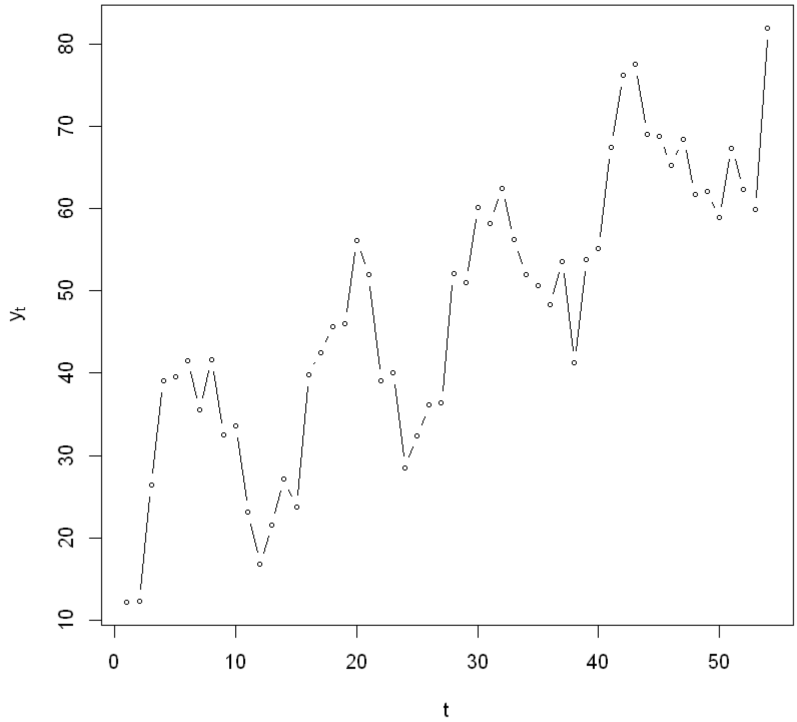
	MSE	MAPE
Training	82.18	20.30%
Testing	495.07	33.53%

1.g Modify the disturbance

$$Y_t = 20 + t + 10\sin\left[\pi\left(\frac{1}{2}\cos\left(\frac{t}{2}\right)+1\right)\right] + \epsilon_t \quad \text{for } t = 1, 2, \ldots, 54,$$

where ϵ_t is generated from a normal distribution $N(\mu = 0, \sigma = 5)$.

In [6]:
```
y_t2 = array(54)
set.seed(23)
for (t in 1:54){
    y_t2[t] = 20 + t + 10*sinpi(cos(t/2)/2+1) + rnorm(1,mean = 0,sd = 5)
    }
plot(y_t2,xlab = "t",ylab = TeX(r'($y_t$)'),type = 'b', cex = 0.5)
```




```
In [7]: y_t2
```

12.1503754016177 · 12.3223392967157 · 26.4574831053855 · 39.0480746168464 · 39.4991264902344 · 41.536216910049 · 35.55979615297 · 41.65219387958 · 32.478183976207 · 33.5691251776304 · 23.1203786833707 · 16.7868884069022 · 21.5633183942991 · 27.145672667459 · 23.738268981294 · 39.8063481100693 · 42.5078136527068 · 45.6911535452512 · 46.0033372849435 · 56.1550834658869 · 51.9711545865519 · 39.1004063600147 · 40.0588765833379 · 28.4696441880689 · 32.3459597731993 · 36.1002003110737 · 36.3941722883111 · 52.0829509304767 · 50.9878663152319 · 60.1290801759325 · 58.2282468509271 · 62.4687984336425 · 56.1937667959783 · 51.9392909852197 · 50.6103406575874 · 48.2775500190621 · 53.5833220562836 · 41.2868969608349 · 53.8153696405998 · 55.1842919956638 · 67.4578160974281 · 76.2263376387748 · 77.5578970966603 · 69.0718724602757 · 68.7232884839386 · 65.2801576522158 · 68.432455531897 · 61.7086060801294 · 62.0261950784773 · 58.8531865619676 · 67.3623904855087 · 62.306804211463 · 59.8888875491921 · 81.9349170338442

Repeat the steps above

	MSE	MAPE
Training	46.33	13.33%
Testing	270.95	23.63%

In general, the model performs better for data with smaller fluctuation.

2.

Construct the Holt-Winter's model (Triple Exponential Smoothing model) for the same data in 1-(a) using $y_1, y_2, \ldots, y_{48}$. Try to find the smoothing factors with better performance. Explain what you observe comparing with the results in 1-(e) & 1-(f)

We use the L_t first model in EXCEL
L_0, T_0 are from the regression model

SMOOTHING FACTORS WITH BETTER PERFORMANCE

Smoothing constant	Value
α	0.09
β	0.79
γ	0.84

	MSE	MAPE
Training	346.69	28.92%
Testing	221.47	20.46%

The validation data has greater fluctuation (worse MSE and MAPE) while the model performs better for testing data.