Chapter 1: Introduction

Question 1:

1- Define a time series? mention some examples of time series data.
   A time series is a collection of observations of some phenomenon (ظاهرة) collected sequentially over a period of time.

Some examples are:

1. Monthly rain amounts.
2. Daily car accidents.
3. Weekly sales.

2- Mention four goals of time series analysis.

a- Description
Which indicate describing and portraying the available information that shows how the studied phenomenon evolve over time. That is, describe the main features of the time series, which will help in determining the best model that can be appropriate to achieve the other goals of the analysis, and get to know the upward and downward movements in the time series and to identify the major components such as **trend**, **stationary** and **seasonal changes**.

b- Interpretation
Interpretation means explaining the changes occurring in the phenomenon using other time series that are related to it, or by using environmental factors affecting the phenomenon.

for example:

1. How the sea level is affected by temperature
2. How sales are affected by advertising.

c- Control

In production lines (in the factories), one may get time series that designate the product quality in the manufacturing process, and the goal here might be to control product quality so that it does not go below a specified level.

d- Forecasting

*Forecasting is considered one of the most important goals of time series analysis. As one might want to know or expect the future values of a time series. Analysis of time series usually **starts by identifying an appropriate model** that explains the evolution pattern of the series, and **then uses the model to extrapolate this pattern into the future**.*

3- *Explain briefly the components of a time series.*

a- trend component

*If there **exist a long-term increase (or decrease)** in the level of the series, then we say there exist a trend component in the series.*
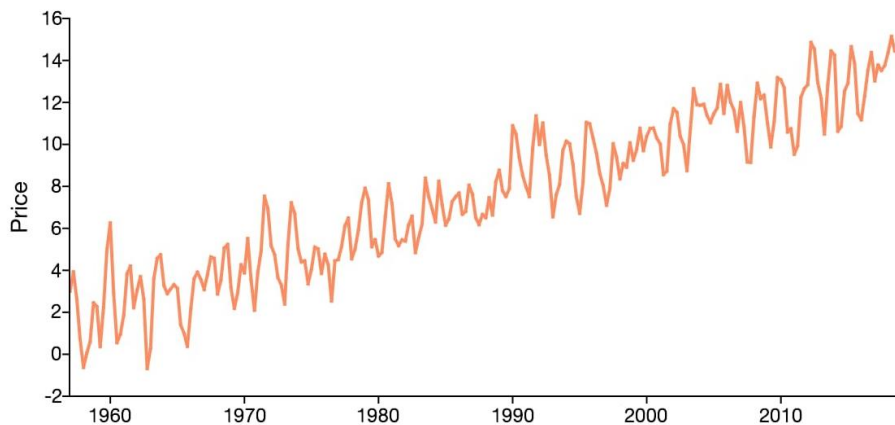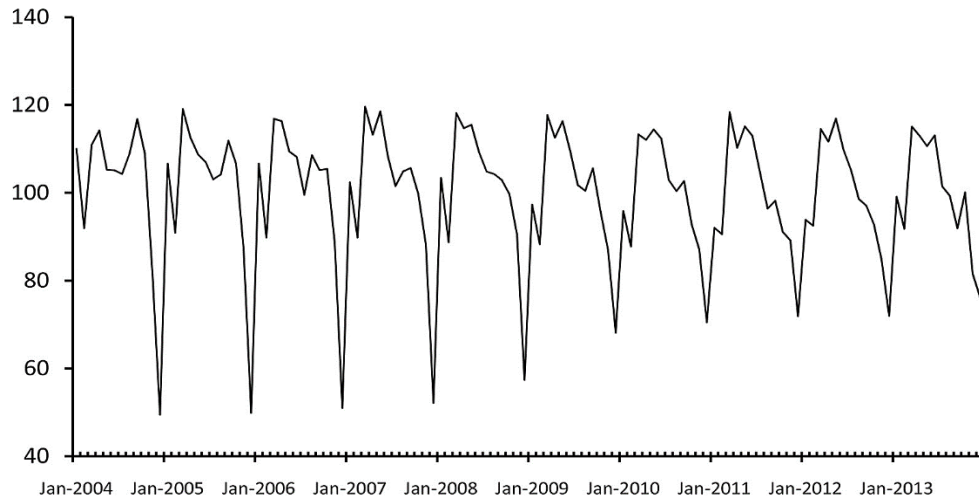

for example:

1. *Time series of the number of births.*
2. *Time series of the number of pilgrims.*
3. *Time series of the prices of goods annually*

b- seasonal component

*A seasonal pattern exists when a series is influenced by seasonal factors (**e.g., the quarter of the year, the month, or day of the week**). Seasonality is always of a fixed and known period.*

*For example: the electric power consumption reaches its peak in summer and fall in winter. Seasonal changes occur at periods **less than a year**, such as hour, day, week, month, quarter, etc.*

c- cyclical component

*Similar to seasonal variation, but they appear **in long periods of time (more than one year),** and to discover the cyclical variation one need a very long annual series.*

for example:

1. *climate changes need data of fifty years or more to discover its cycle*
2. *economic cycles need a long period of time, for example five or ten Years, to appear.*

d- Random component

*After getting rid of seasonal, trend, or cyclical components from the data, we are left with **a residual series**, which represent the irregular changes. These changes differ from the other components, as they can't be predicted, and they **do not occur according to any law or system.***

4- *Explain the difference between time series models and causal models.*

a- Causal models (Regression model):

*This approach is based on identifying the variable(s) that may have a causal relationship with the variable under study that we want to predict, this variable is called the **dependent variable**, then determine the appropriate statistical model or appropriate functional relationship which explains how the dependent variable is associated to **the independent or explanatory variables**. Using this model, we can predict the dependent variable under study.*

The main disadvantages of this approach are:

1. *Difficulty of identifying all the explanatory variables that are related to the dependent variable.*
2. *Requires the availability of detailed historical information about all the explanatory variables, and the ability of knowing these variables or predicting them.*

b- time series models:

This approach relies on analyzing historical data of the variable under study in order to determine the pattern it follows. Assuming that this pattern will continue in the future, we use it to predict future values of the variable. The approach assumes that there exists a stochastic process able to produce the time series under study, and by imposing the stationarity condition, we then will be able to use the mathematical model that describes the correlation structure of the data to obtain reliable forecasts.

5- *When measuring forecast accuracy (or error size), explain why don't we just use the sum of the errors.*

*Because by definition, the errors can be positive and negative, and using the sum will not provide us with an accurate measure of their size $\sum \varepsilon_i = 0$, thus we resort to **squaring** or taking the **absolute value** of the errors to get a good estimate of their size.*

6- *Is the best forecasting technique always the most accurate one? explain.*

The best forecasting method is not necessarily the method that achieves the highest accuracy or the smallest forecasting errors, but one method may be used because of type of the required forecast (point or confidence interval forecast), another because of only small number of observations are available, a third because it has a low cost, and a fourth because its theoretical assumptions comply with the data set in hand.

7- *Give examples from the real life of a time series that have:*

a- *seasonal component of length a year.*
   1. *Yearly amount of rain for a period of 50 years.*
   2. *Yearly average riyal exchange rate against dollar for a period of 50 years.*

b- *seasonal component of length a month.*
   Number of monthly car accidents for a period of 50 years.

c- *seasonal component of length a week.*
   Number of patients visiting a clinic weekly for a period of a year.

Question 2*:

The following table shows yearly accidents in a certain city, and estimated values of the yearly accidents calculated from some fitted model:

Year	2018	2019	2020	2021
accidents	190	204	289	349
Estimated accidents	197	199	299	344

a- *calculate the estimated errors.*
b- *calculate mean squared errors MSE, the mean absolute deviances MAD, and mean absolute percentage errors MAPE.*

Year	2018	2019	2020	2021	Total
accidents y_t	190	204	289	349	
estimated sales $\hat{y}_t$	197	199	299	344	
estimated errors $\varepsilon_t = y_t - \hat{y}_t$	-7	5	-10	5	
absolute errors $\|\varepsilon_t\|$	7	5	10	5	27
Absolute relative errors $\left\|\frac{y_t - \hat{y}_t}{y_t}\right\|$	0.036842	0.0279	0.0189	0.04	0.110281
Squared errors $(\varepsilon_t)^2$	49	25	100	25	199

$$MAD = \frac{1}{k}\sum_{i=1}^{k}|\varepsilon_t| = \frac{27}{4} = 6.75$$

$$MAPE = \frac{100}{k}\sum_{i=1}^{k}\left|\frac{y_{t-}\hat{y}_t}{y_t}\right| = \frac{100}{4}(0.110281) = 2.757016$$

$$MSD = \frac{1}{k}\sum_{i=1}^{k}(y_{t-}\hat{y}_t)^2 = \frac{199}{4} = 49.75$$

By R:

```
> y=c(190,204,289,349)
> yhat=c(197,199,299,344)

> e=yhat-y
> sum(e)
[1] -7
>
> k=4
> MSD=sum(e^2)/k
> MSD
[1] 49.75

> MAD=sum(abs(e))/k
> MAD
[1] 6.75

> MAPE=(100/k)*sum(abs(e/y))
> MAPE
[1] 2.757016
```

$$\sum \varepsilon_t$$

$$MSD = \frac{1}{k}\sum_{i=1}^{k}(y_t - \hat{y}_t)^2$$

$$MAD = \frac{1}{k}\sum_{i=1}^{k}|\varepsilon_t|$$

$$MAPE = \frac{100}{k}\sum_{i=1}^{k}\left|\frac{y_t - \hat{y}_t}{y_t}\right|$$

Question 3:

The following table shows the loans financed by a bank (in millions of dollars) in the period 1995 to 2001:

Year	1995	1996	1997	1998	1999	2000	2001
Sales	12	13	11	13	12	14	11

a- *Use the method of simple moving averages method to find all the possible forecasts using k=3, and k=4. Also find the* <u>mean absolute deviance for k=4</u>

Year	1995	1996	1997	1998	1999	2000	2001	2002
Sales	12	13	11	13	12	14	11	
K=3	-	-	-	12	12.33	12	13	12.33
K=4	-	-	-	-	12.25	12.25	12.5	12.5

simple moving average of order k=3:

$$ma_{1998}(3) = \frac{y_{1995}+y_{1996}+y_{1997}}{3} = \frac{12+13+11}{3} = 12$$

$$ma_{1999}(3) = \frac{y_{1996}+y_{1997}+y_{1998}}{3} = \frac{13+11+13}{3} = 12.33$$

$$ma_{2000}(3) = \frac{y_{1997}+y_{1998}+y_{1999}}{3} = \frac{11+13+12}{3} = 12$$

$$ma_{2001}(3) = \frac{y_{1998}+y_{1999}+y_{2000}}{3} = \frac{13+12+14}{3} = 13$$

simple moving average of order k=4:

$$ma_{1999}(4) = \frac{y_{1995}+y_{1996}+y_{1997}+y_{1998}}{4} = \frac{12+13+11+13}{4} = 12.25$$

$$ma_{2000}(4) = \frac{y_{1996}+y_{1997}+y_{1998}+y_{1999}}{4} = \frac{13+11+13+12}{4} = 12.25$$

$$ma_{2001}(4) = \frac{y_{1997}+y_{1998}+y_{1999}+y_{2000}}{4} = \frac{11+13+12+14}{4} = 12.5$$

The MAD for k=4

Year	y_t	$\hat{y}_t$	$\|\varepsilon_t\| = \|y_t - \hat{y}_t\|$
1999	12	12.25	0.25
2000	14	12.25	1.75
2001	11	12.5	1.5
	$\sum_{i=1}^{k}\|\varepsilon_t\|$		3.5

$$MAD = \frac{1}{k}\sum_{i=1}^{k}|\varepsilon_t| = \frac{3.5}{3} = 1.167$$

b- *Forecast the amount of loans that the bank will finance in the year 2002 using the simple moving average method.*

$$ma_{2002}(3) = \frac{y_{1999} + y_{2000} + y_{2001}}{3} = \frac{12+14+11}{3} = 12.33$$

$$ma_{2002}(4) = \frac{y_{1998} + y_{1999} + y_{2000} + y_{2001}}{4} = \frac{13+12+14+11}{4} = 12.5$$

By R:

```
> sales <- read.table("C:/Desktop/sales.txt",header = TRUE)
> install.packages("TTR")
> library(TTR)
> SMA(sales$Sales,n=3)
[1]       NA       NA 12.00000 12.33333 12.00000
[6] 13.00000 12.33333
#OR
> install.packages("forecast")
> library(forecast)
> ma(sales$Sales,order=3,centre=FALSE)
Time Series:
Start = 1
End = 7
Frequency = 1
[1]       NA 12.00000 12.33333 12.00000 13.00000
[6] 12.33333       NA
```

c- *Estimate the initial value $\hat{y}_0(1)$ using the mean of the series, then use the single exponential smoothing method to find all the forecasts, use $\alpha = 0.75$, and then $\alpha = 0.95$. Which one gives better forecasts? explain.*

$$\hat{y}_0(1) = \bar{y} = \frac{12+13+11+13+12+14+11}{7} = 12.2857$$

now, simple exponential smoothing (SES) using $\alpha = 0.75$:

$$\hat{y}_t(1995) = 0.75y_{1995} + (1 - .75)\hat{y}_0(1)$$
$$= (0.75)(12) + (1 - .75)(12.2857) = 12.0714$$

$$\hat{y}_t(1996) = 0.75y_{1996} + (1 - .75)S_{1995}$$
$$= (0.75)(13) + (1 - .75)(12.0714) = 12.7679$$

$$\hat{y}_t(1997) = 0.75y_{1997} + (1 - .75)S_{1996}$$
$$= (0.75)(11) + (1 - .75)(12.7679) = 11.442$$

$$\hat{y}_t(1998) = 0.75y_{1998} + (1 - .75)S_{1997}$$
$$= (0.75)(13) + (1 - .75)(11.442) = 12.6105$$

$$\hat{y}_t(1999) = 0.75y_{1999} + (1 - .75)S_{1998}$$
$$= (0.75)(12) + (1 - .75)(12.6105) = 12.1526$$

$$\hat{y}_t(2000) = 0.75y_{2000} + (1 - .75)S_{1999}$$
$$= (0.75)(14) + (1 - .75)(12.1526) = 13.5382$$

$$\hat{y}_t(2001) = 0.75y_{2001} + (1 - .75)S_{2000}$$
$$= (0.75)(11) + (1 - .75)(13.5382) = 11.6346$$

In the same manner, for SES using $\alpha = 0.95$, we get:

$$\hat{y}_t(1995) = 0.95y_{1995} + (1 - .95)\hat{y}_0(1)$$

$$= (0.95)(12) + (1 - .95)(12.2857) = 12.0143$$

$$\hat{y}_t(1996) = 0.95y_{1996} + (1 - .95)S_{1995}$$

$$= (0.95)(13) + (1 - .95)(12.0143) = 12.9507$$

$$\hat{y}_t(1997) = 0.95y_{1997} + (1 - .95)S_{1996}$$

$$= (0.95)(11) + (1 - .95)(12.9507) = 11.0975$$

$$\hat{y}_t(1998) = 0.95y_{1998} + (1 - .95)S_{1997}$$

$$= (0.95)(13) + (1 - .95)(11.0975) = 12.9049$$

$$\hat{y}_t(1999) = 0.95y_{1999} + (1 - .95)S_{1998}$$

$$= (0.95)(12) + (1 - .95)(12.9049) = 12.0452$$

$$\hat{y}_t(2000) = 0.95y_{2000} + (1 - .95)S_{1999}$$

$$= (0.95)(14) + (1 - .95)(12.0452) = 13.9023$$

$$\hat{y}_t(2001) = 0.95y_{2001} + (1 - .95)S_{2000}$$

$$= (0.95)(11) + (1 - .75)(13.9023) = 11.1451$$

y_t	$\alpha = 0.75$ $\hat{y}_t$	$(y_t - \hat{y}_t)^2$	y_t	$\alpha = 0.95$ $\hat{y}_t$	$(y_t - \hat{y}_t)^2$
12	12.0714	0.0051	12	12.0143	0.0002
13	12.7679	0.0539	13	12.9507	0.0024
11	11.442	0.1954	11	11.0975	0.0095
13	12.6105	0.1517	13	12.9049	0.0090
12	12.1526	0.0233	12	12.0452	0.0020
14	13.5382	0.2133	14	13.9023	0.0095
11	11.6346	0.4027	11	11.1451	0.0211
$\sum(y_t - \hat{y}_t)^2$		**1.0453**	$\sum(y_t - \hat{y}_t)^2$		**0.0538**

$$MSD_{\alpha=0.75} = \frac{1}{k}\sum_{i=1}^{k}(y_{t-}\hat{y}_t)^2 = \frac{1.0453}{7} = 0.15$$

$$MSD_{\alpha=0.95} = \frac{1}{k}\sum_{i=1}^{k}(y_{t-}\hat{y}_t)^2 = \frac{0.0538}{7} = 0.008$$

So we see that the SES with smoothing parameter $\boldsymbol{\alpha = 0.95}$ gives forecasts the smallest MSD, thus, we use it for forecasting the value of the loans the bank will finance for the year 2002.

By R:

```
> sales <- read.table("C:/Desktop/sales.txt",header = TRUE)
> y <- sales$Sales
> fitted75 <- array(dim = 7)
> fitted75[1] <- y[1]*0.75 + mean(y)*0.25
> for(i in 2:7){
    fitted75[i] <- y[i]*0.75+fitted75[i-1]*0.25
}
> fitted75
[1] 12.07143 12.76786 11.44196 12.61049 12.15262
[6] 13.53816 11.63454
> e75 <- y-fitted75
> MSD75 <- sum(e75^2)/7
> MSD75
```

```
[1] 0.1493251
> fitted95 <- array(dim = 7)
> fitted95[1] <- y[1]*0.95 + mean(y)*0.05
> for(i in 2:7){
    fitted95[i] <- y[i]*0.95+fitted95[i-1]*0.05
}
> fitted95
[1] 12.01429 12.95071 11.09754 12.90488 12.04524
[6] 13.90226 11.14511
> e95 <- y-fitted95
> MSD95 <- sum(e95^2)/7
> MSD95
[1] 0.007693186
```