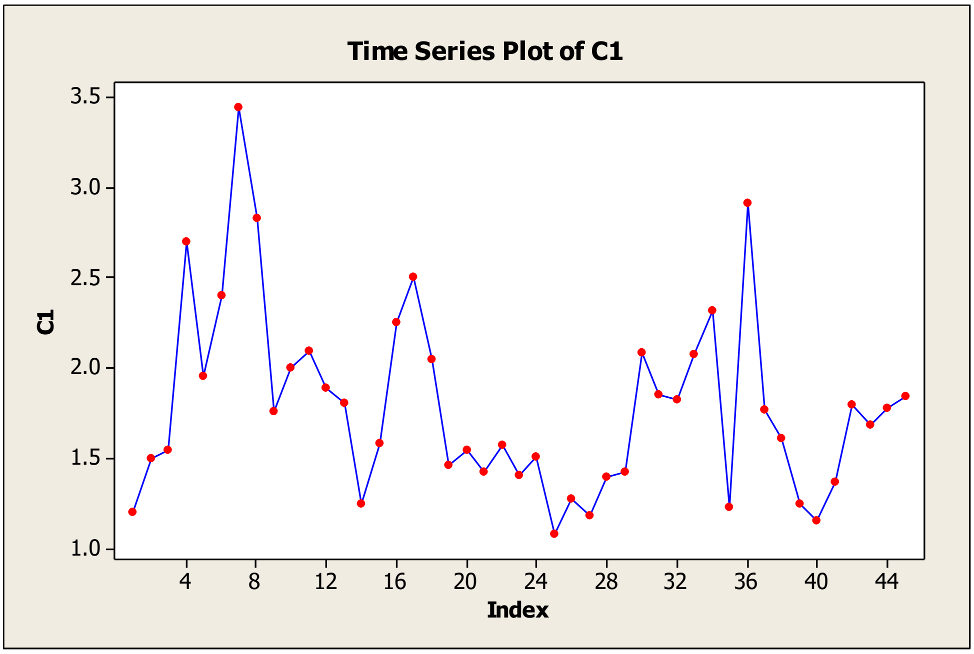
Tutorial set #5

Question 6:

For the attached two sets of data (data1) and (data2), do the following:

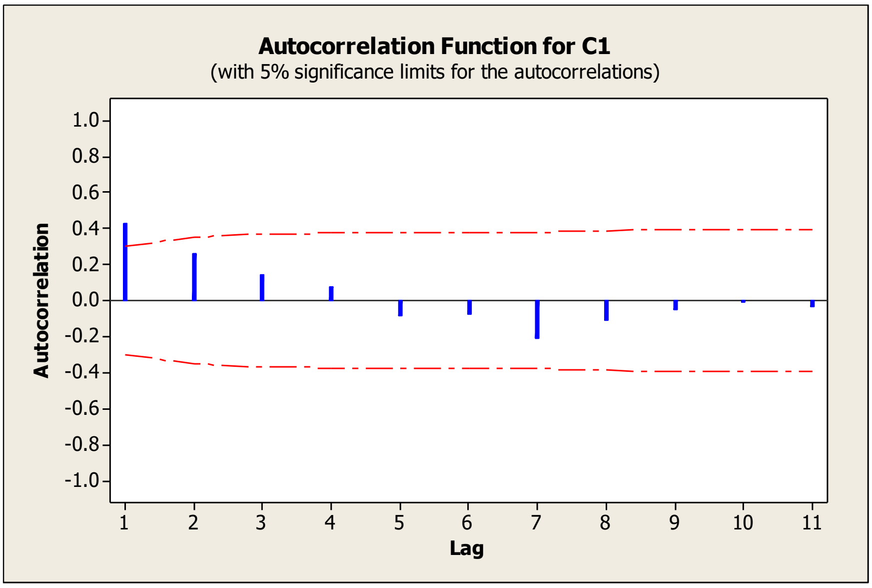
1- Plot the series, and check its stationarity in mean and variance.
2- plot the ACF and PACF , suggest a preliminary model for the data.
3- Fit the suggested models, and get acquainted with the MINITAB output.

1- data 1:

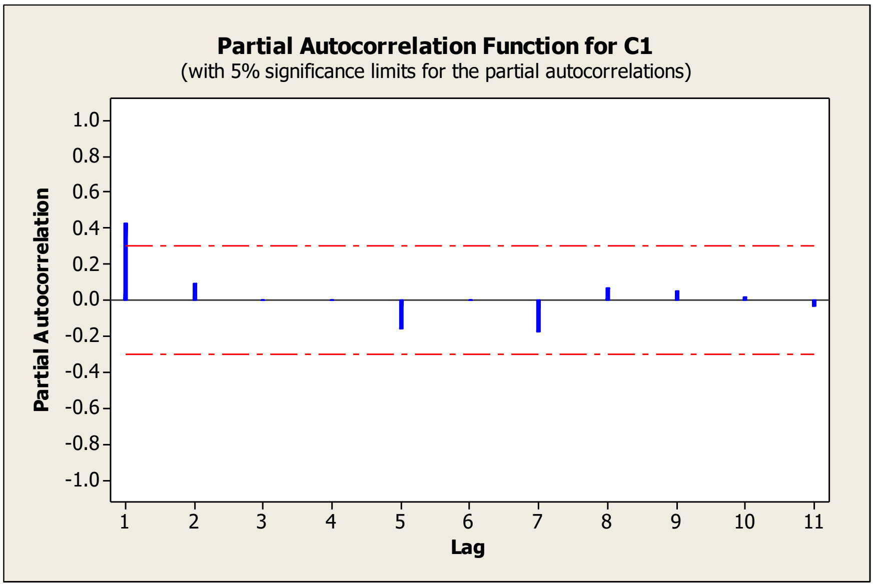


From the above figure, the data seems to be stationary in the mean. But, the variance seems to be different at the ends of the series than the middle, however, it is very serious problem, we will leave dealing with such situations later.

a- Now, we will examine the ACF and PACF for the data, to identify a tentative model for the data:



we notice from the correlogram that it takes a declining exponential form, the decline to zero is fast which is an indication of the series stationarity. Also, it is a format of the Autoregressive models AR(.). Next, we will examine the PACF to identify the order of the model:

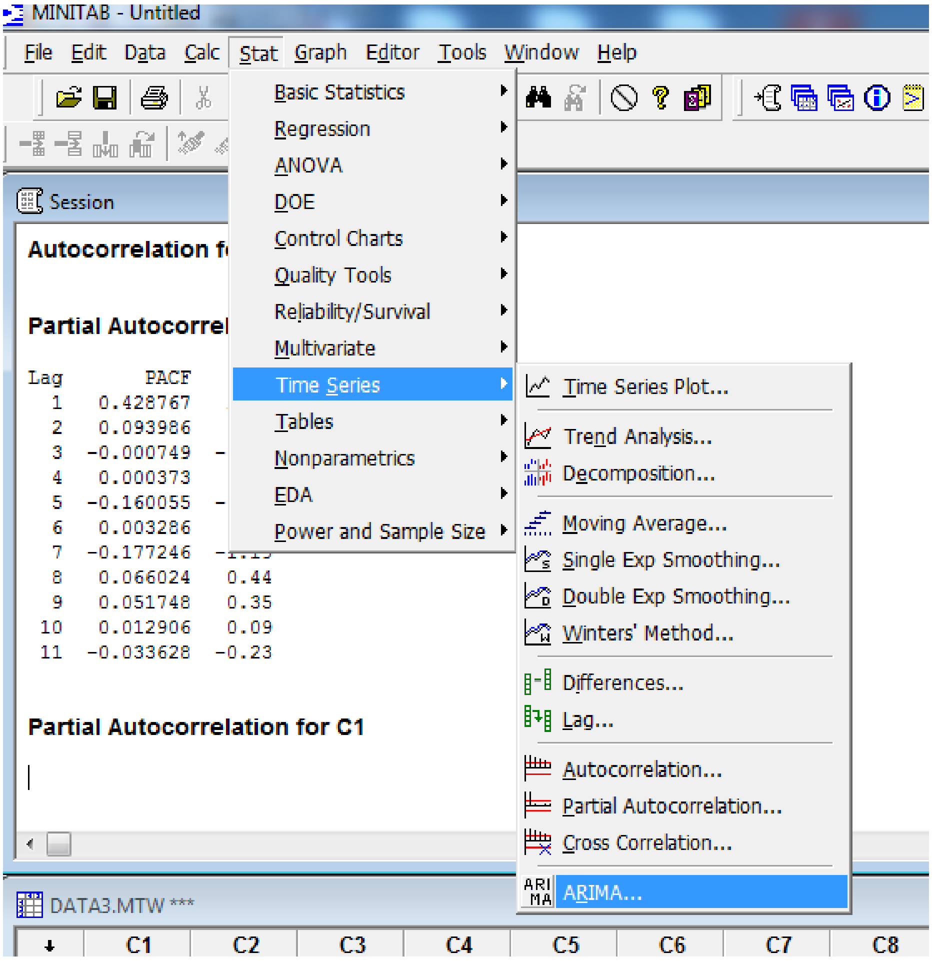


As we notice, the PACF has only one value that does equal to zero, whereas the rest of the values do not differ significantly from zero (they are within the 95% Confidence limits). Thus we can say tentatively that the appropriate model for the data set is AR(1).

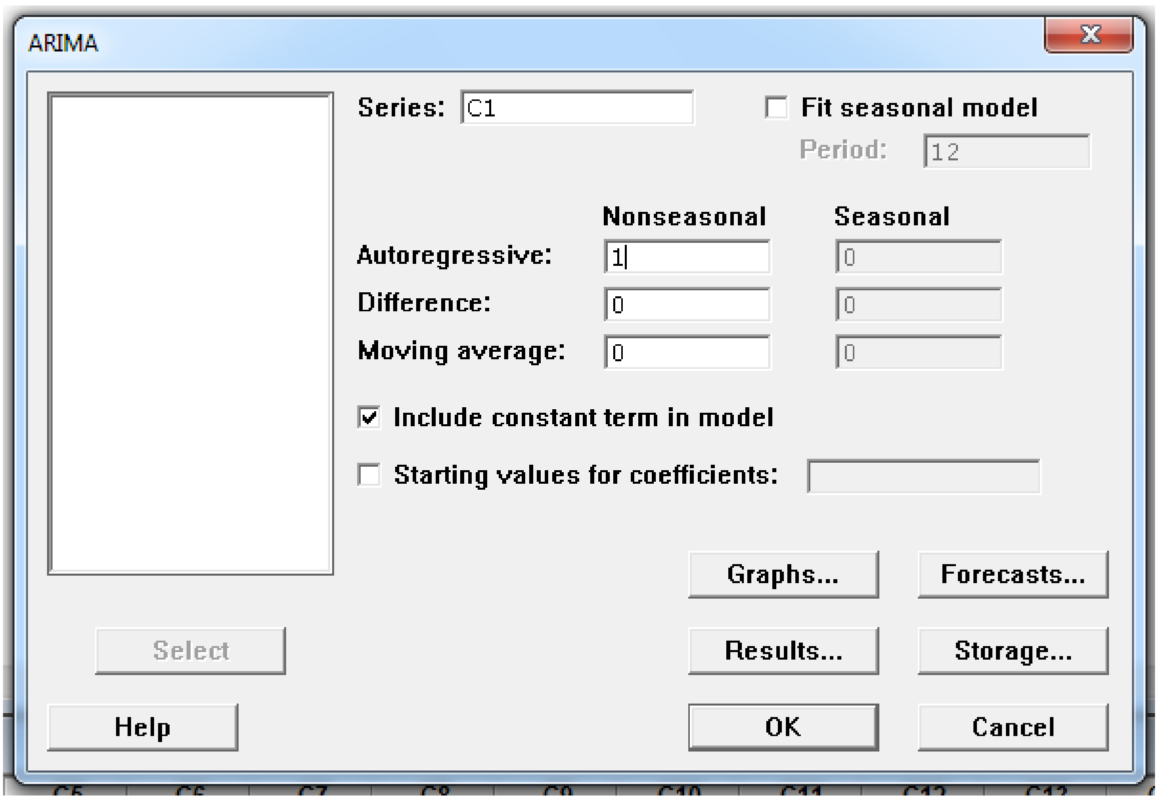
b- Fitting the model in MINITAB:

Wefit the model in MINITAB as follows:

Stat ---> Time Series --->ARIMA

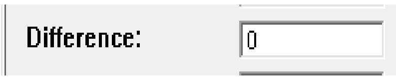


so the following window appears:



where the program asks for entering the column that contain the data set (here it is in C1).

Then we choose the model we wish to fit to the data , here, for example it is AR(1), thus we enter 1 here: 

, and if we want to take differences of any order, then we enter this in the dialogue (for our case we, do not require any differences, so we enter 0): 

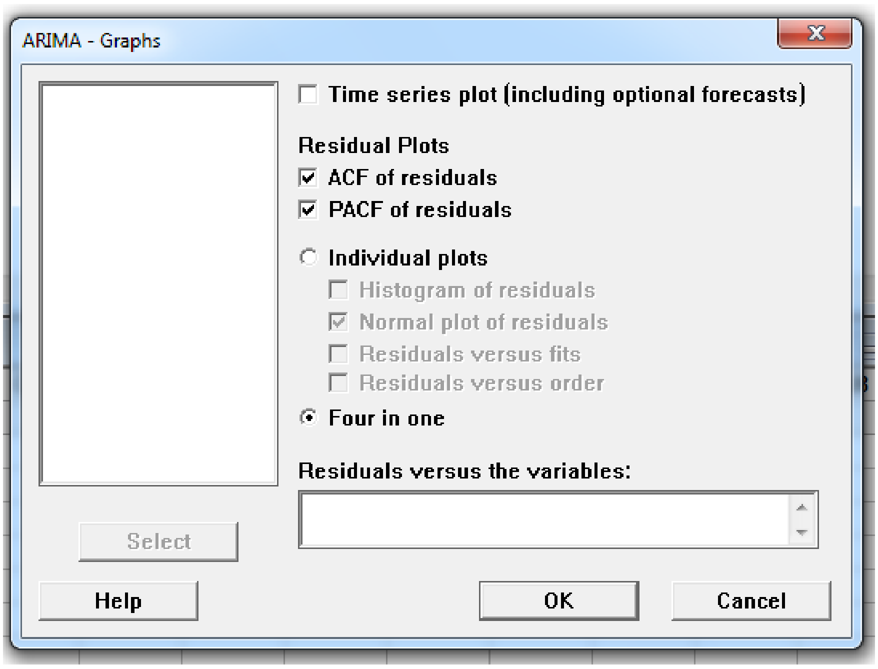
.

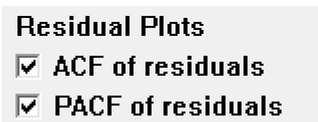
for the third dialogue 

this is for the moving average models, we will see this later.

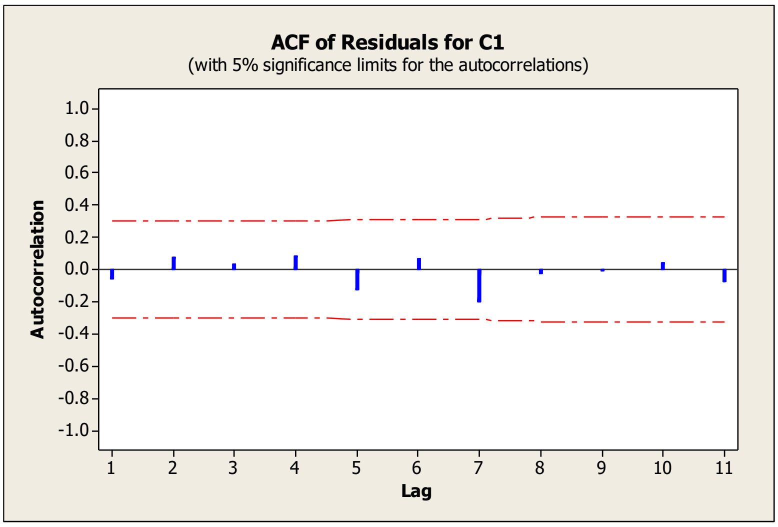
We must also get some diagnostic checks that help us in deciding on the quality of our fitted model. We choose the command 

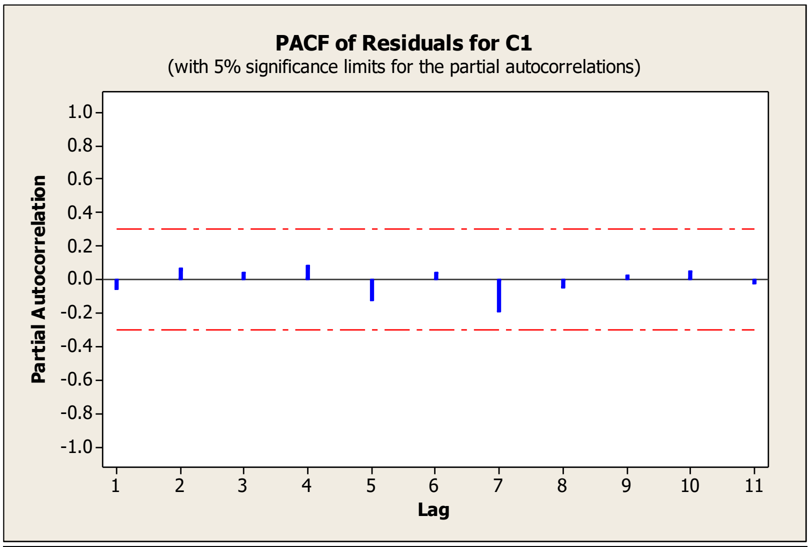
in the previous window, and we get the following output :



There are many options we can use, however, for the time being we will tick the options: 

which will plot for us the ACF and the PACF of the residuals of the model. So, if the model was successful in modelling the autocorrelation pattern of the time series, then we would expect the residuals will follow white noise process, i.e. it will be random variables that are independent, having mean zero, and constant variance. Thus, the ACF and PACF should not show any indication of autocorrelation pattern left in the time series data:





From the above two figures, we notice that there is no indication of any autocorrelation pattern left in the residuals of the model, as all the ACF and PACF coefficients lie within the 95% confidence limits. Thus we can say the model has succeeded in modelling the autocorrelation structure of the data.

But if we found the contrary, then we should go back and search for a better model for our data.

Notice: It is not enough just to check the ACF and PACH of the residuals, but there are actually several significance tests that must be performed such as the randomness test, testing that the residual mean equal to zero, and that it follows the normal distribution, we will come to see these tests later.

Now, let's have a look at the program output of the fit:

```
Session

Final Estimates of Parameters

Type          Coef   SE Coef      T      P
AR   1      0.4421    0.1365   3.24  0.002
Constant   0.99280   0.06999  14.19  0.000
Mean        1.7795    0.1254


Number of observations:  45
Residuals:     SS =  9.47811 (backforecasts excluded)
               MS =  0.22042  DF = 43


Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag             12      24      36   48
Chi-Square     4.9     8.9    30.9    *
DF              10      22      34    *
P-Value      0.899   0.994   0.620    *
```

From the output above, we see that the program provides us with the following:

1- The AR(1) coefficient estimate, since we have fitted the model $y_t = \delta + \phi_1 y_{t-1} + \varepsilon_t$, thus the estimated value for ϕ_1 is $\hat{\phi}_1 = 0.4421$, and the standard error of this estimate is 0.1365, and the P_value for testing the hypothesis $H_0: \phi_1 = 0$ equals 0.002, this means that we reject H_0 and deduce that the values of the series at time t depends on the values at time $t - 1$, hence it must remain in the model.

2- We also get an estimate for $\hat{\delta} = 0.99280$, which is the y-intercept , the program calls it the (constant) which equals (get back to the lecture notes): $\hat{\delta} = \hat{\mu}(1 - \hat{\phi}) = 1.7795(1 - 0.4421)$, together with P_value for testing the hypothesis $H_0: \delta = 0$ equals zero, this means that we reject H_0 and deduce that the constant should be kept in the model.

3- We also get an estimate for the mean of the series which is $\hat{\mu} = 1.7795$.

4- The estimated variance for the white noise process is $\widehat{\sigma_\varepsilon^2} = 0.22042$, with 43 degrees of freedom, where the number of observations of the series was 45, and we lost two degrees of freedom when we estimated ϕ_1 and δ. The estimated variance $\widehat{\sigma_\varepsilon^2}$ is useful in constructing C.I. for the forecasts (we will see this later), it also is useful choosing among several models which are suitable for modelling the time series data, in which we choose the one with the lowest $\widehat{\sigma_\varepsilon^2}$.

5- The program output also, provides us with a significance test that there is no correlation left between the residuals:

The hypothesis here is $H_0: \rho_1 = \rho_2 = \cdots = \rho_q = 0$ for any number of time lags q between the residuals of the model. The scientists Box and Cox designed a test for his hypothesis, and the program provide us with the results of this tests for several values of q: 12, 24, 36, 48. The test static for testing this hypothesis follows a chi-square distribution with q-k degrees of freedom, where k is the number of estimated parameters in the model. We, of course, want to accept the hypothesis H_0 , i.e. we hope that our model we proposed for the data was able to model all the autocorrelation structure in the data, and hence we do not expect to find any left autocorrelation structure in the model residuals for any time lag. From the output above we note that the P_values are 0.899, 0.994, 0.620 , and thus we accept H_0 and deduce that the model is adequate.

data set 2:

give it as a homework .