SECTION 4 TME SERIES ANAYSIS AND CONTROL CHARTS

Structure **Page Nos.**

4.0 INTRODUCTION

In the previous section, you were introduced to regression analysis, where it was desired to fit a linear (in case of linear regression) or a curve between the dependent and the independent variables. A typical decision making through use of regression analysis may involve plan for future growth of infrastructure of a city based on the growth pattern in population. Thus, prediction primarily focuses on estimating a future value based on the presented study data.

This section introduces the concept of Time series, which you learnt in BCS040/Block 3, and also, how they can be used for forecasting. The Unit begins with an introduction to time series and then shows how you can use a worksheet to forecast using time series data.

Further, the unit introduces the control charts that are used in determining the statistical quality control. You must go through the following two units of BCS040/Block 3 before performing various tasks/steps/activities listed in this section.

Unit 10: Forecasting and Time Series Analysis
Unit 11: Statistical Quality Control

4.1 OBJECTIVES

After performing the activities of this section, you should be able to:

- define time series data;
- use a spreadsheet to create time series data;
- use spreadsheet to fit the linear trends using the method of least squares;
- use spreadsheet for forecasting models.

4.2 TIME SERIES AND FORCASTING

In many situations, data may be collected over a period of time. This time period may be in terms of days, for example, for shopping analysis you may be interested in recording sales on Mondays, Tuesdays, etc.; a week when you want to plan weekly, a month, a year and so on. Such periodic data is called times series data. Block 3/Unit10/ Page44/ Table 3 which is reproduced in the Figure 1 shows the time series data of crop yields. This data is recorded for a time period of a year over 41 years. This data was collected by Government of AP for studying the changes in the cropping pattern. Such data may be used for first predicting a trend. In this example, it is expected to predict the future economic needs of the Agricultural sector.

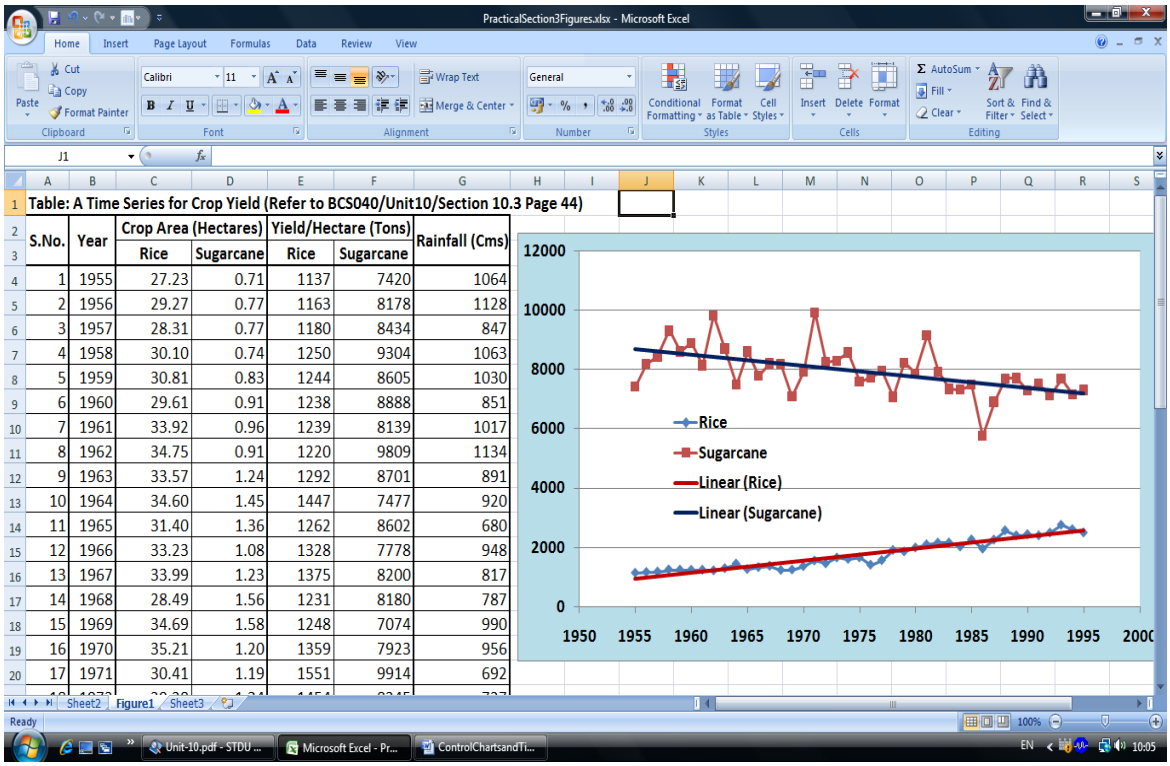
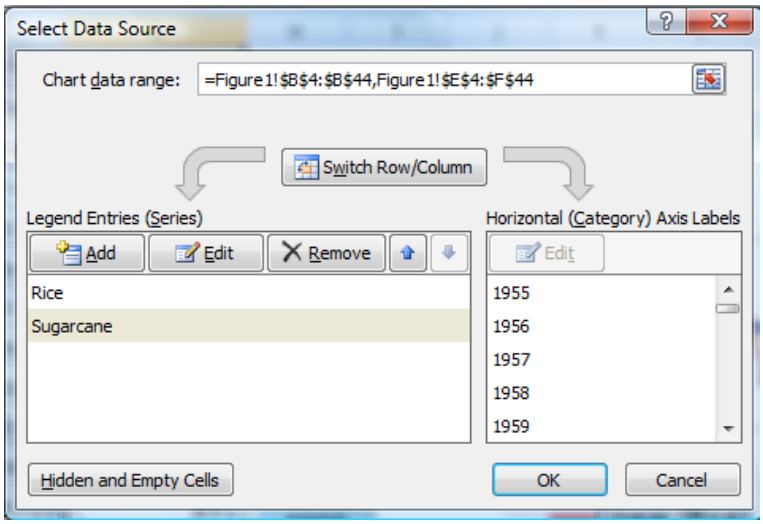
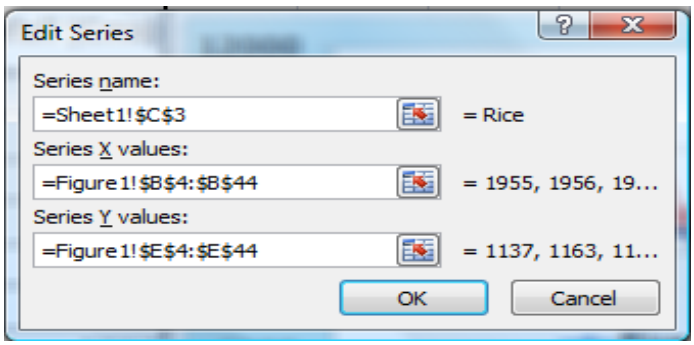


Figure 1: Long Term Trends for Crop yield

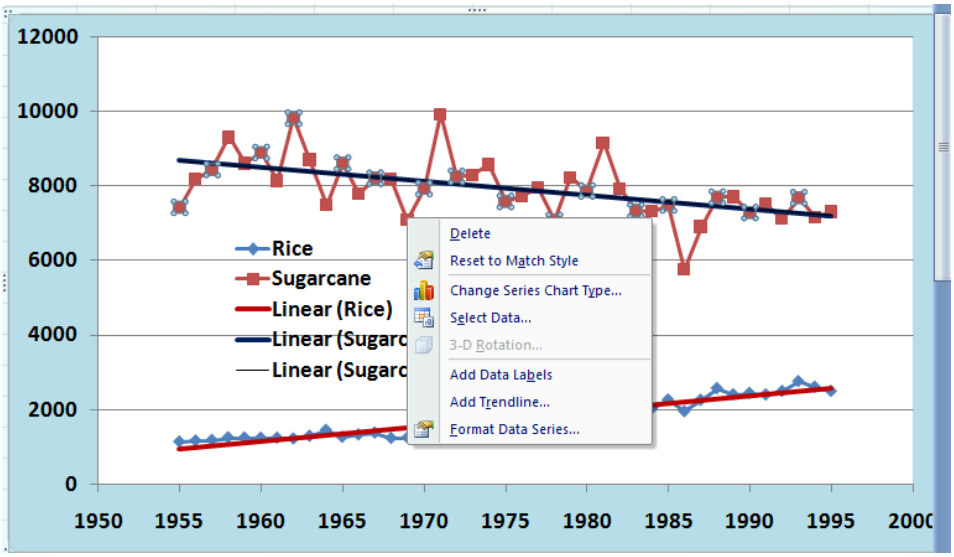
The figure shows the crop yield data and the long term trend for two crops Rice and Sugarcane over a period of 41 years. The Figure also shows the trend line of the two crop yields. To draw the graph, we have selected *a linear Scatter plot*. In the *Select Data…* option you can name the Series viz. "Rice" and "Sugarcane. For checking the details of series in a spreadsheet package, right click on the chart, and select *Select Data…* option. You will see the following window:



You may edit any series, for example, if you select *Rice* Data Series and click on *Edit* Button, you will see the following Range for the series:



Thus, you can modify the series as per your requirements. Now select any series on the chart and press right mouse button, the following Menu will be displayed:



Using this Menu, you can change series or chart type, select data, add data labels and format data series. You may explore these options on your own. One option which requires a mention due to its usefulness here is *Add Trendline…* On selecting this option, you will get the following Menu:

Please note that for Figure 1, you can select Trendline option as Linear. You can also change the Line Color and Line Style using options shown after arrow.

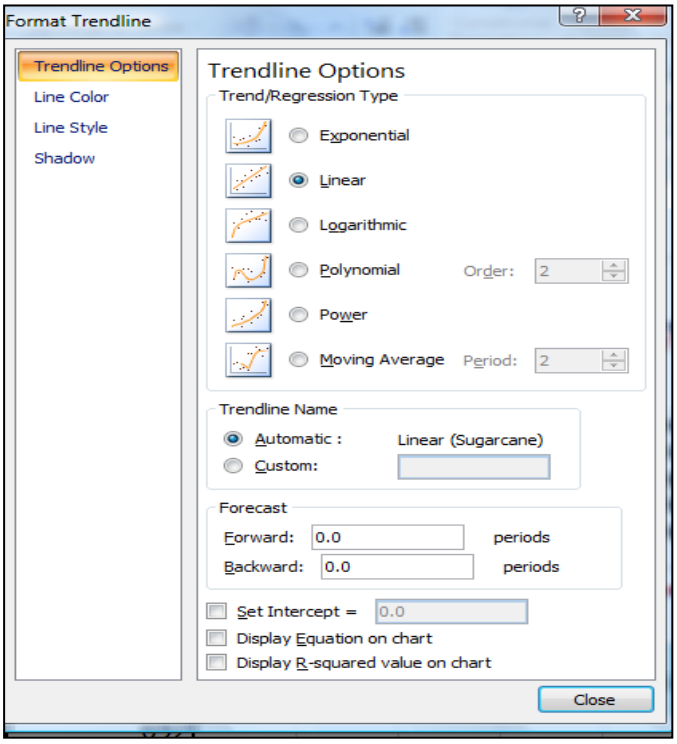
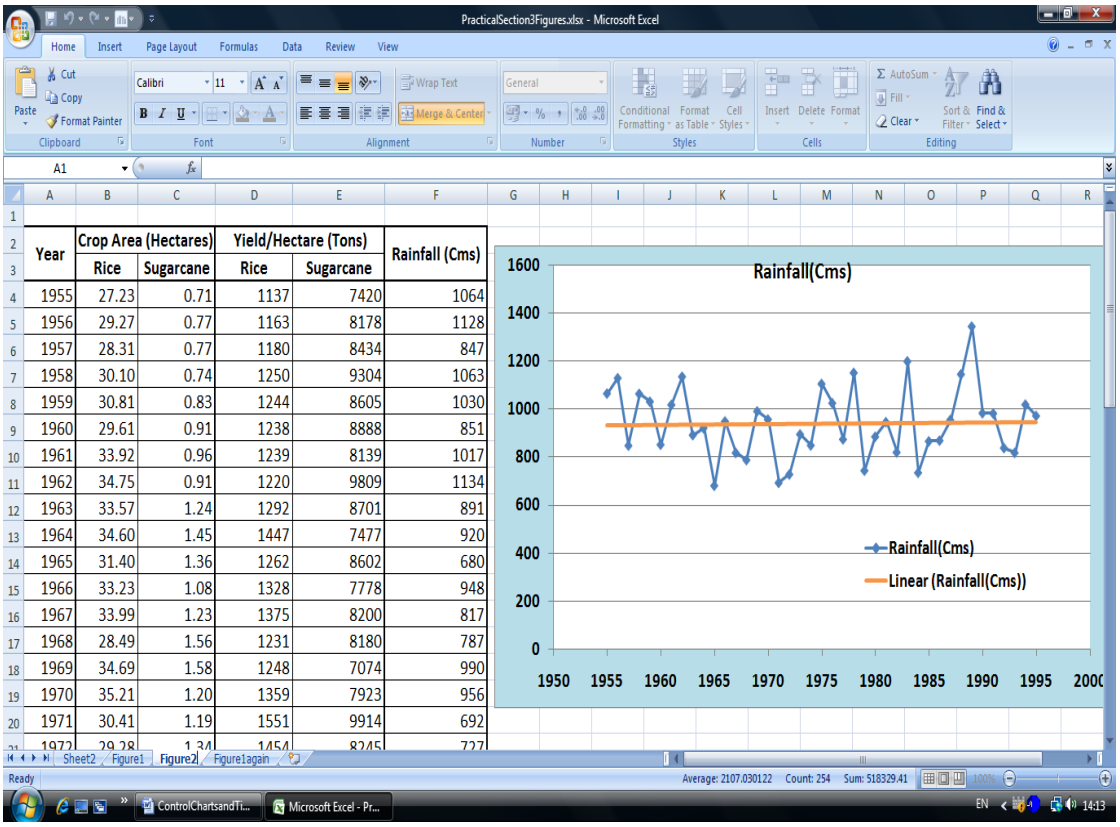


Figure 1(a): Trendline Options

Now observe the data series in Figure 1. There are fluctuations in the time series data. You can attribute these fluctuations on various factors that may include availability and quality of fertilizer, weather conditions etc. You may refer to page 44/Unit 10 for a complete discussion. We briefly present below the following:

Long term trends: From the Figure 1, the long term trend that can be drawn is that average yield of Rice crop per hectare has increased. On the other hand, the yield of Sugarcane per hectare has steadily dropped. These trends can be further analysed to ascertain the reasons of such the declining trends.

Another long term trend about rainfall is shown in Figure 2. You can see that average rainfall over the period of 41 years is almost unchanged.



Year	Crop Area (Hectares)		Yield/Hectare (Tons)		Rainfall (Cms)
	Rice	Sugarcane	Rice	Sugarcane	
1955	27.23	0.71	1137	7420	1064
1956	29.27	0.77	1163	8178	1128
1957	28.31	0.77	1180	8434	847
1958	30.10	0.74	1250	9304	1063
1959	30.81	0.83	1244	8605	1030
1960	29.61	0.91	1238	8888	851
1961	33.92	0.96	1239	8139	1017
1962	34.75	0.91	1220	9809	1134
1963	33.57	1.24	1292	8701	891
1964	34.60	1.45	1447	7477	920
1965	31.40	1.36	1262	8602	680
1966	33.23	1.08	1328	7778	948
1967	33.99	1.23	1375	8200	817
1968	28.49	1.56	1231	8180	787
1969	34.69	1.58	1248	7074	990
1970	35.21	1.20	1359	7923	956
1971	30.41	1.19	1551	9914	692

Figure 2: Trendline for Average Rainfall

Seasonal Variations: As a task, to start with, create the data for the seasonal variations given in the BCS040: Block 3/Unit 10/Page 46 and plot the time series in two possible ways. (i) Plot the figure akin to Fig.5/Unit 10 for all the six years, instead of three shown, with only four quarters along the x-axis. (ii) Plot the time series with years along the x-axis, showing quarters within each of the years. Note the variations that you observe in the curve drawn by the spreadsheet. Discuss with other students or counselors the "relevance" of seasonal variation in data besides commonly encountered examples in real life.

Cyclic variations: Using the data on Sugarcane Area, the cyclic variations are plotted in Figure 3 using spreadsheet package. You can observe the cyclic pattern in the plot. Did you understand the distinction between "a season" and "a cycle", which is discussed in section 10.3.3/page 48/Unit 10? Discuss the key points from there with other students or counselors.

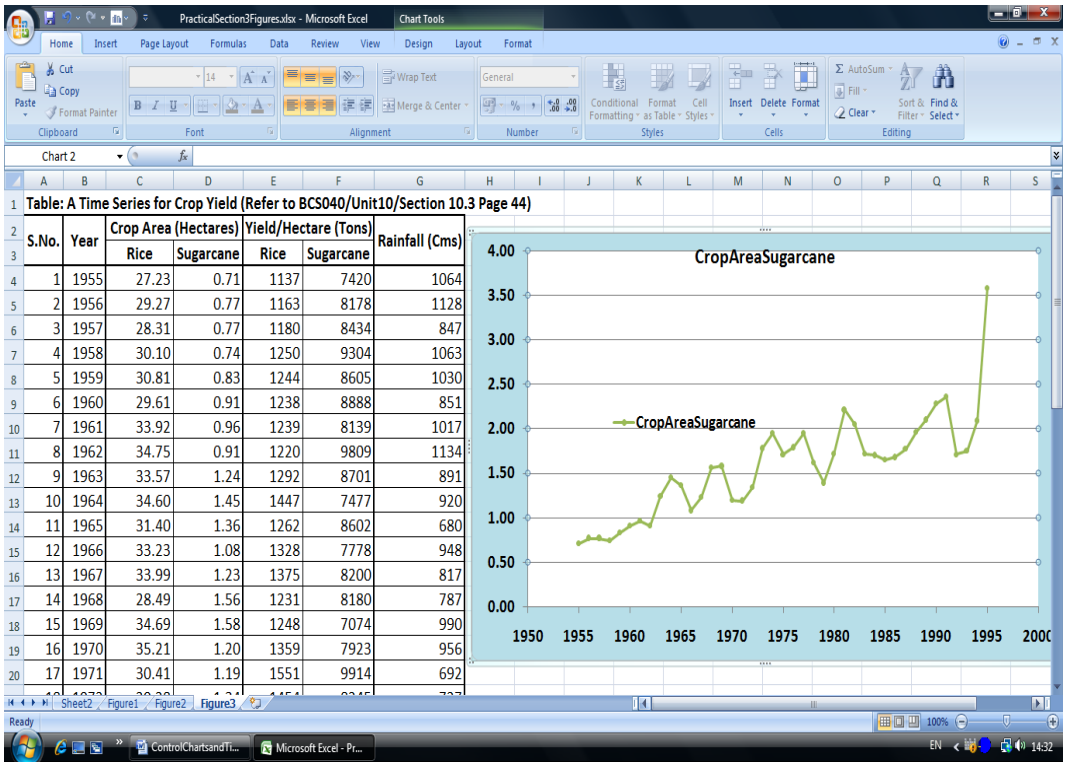


Figure 3: Cyclic variations

Forecasting Models: Please go through various forecasting models from Section 10.4/page 49/MCS040/Block3/Unit 10, viz., *additive* and *multiplicative* models. You may observe that Trendline actually is a forecasting curve line. In Figure 1 and Figure 2, you have used a linear fit. You can revise from Unit 9 or MCS040/Block3/Unit 10.5.1 the method of least squares method. Further, you can select the method of moving averages (MCS040/Block3/Unit 10.5.2) by simply selecting the Trendline Options as shown in Figure 1(a) or you can fit the Polynomial of second degree using least square method, by simple selecting Polynomial Options and keeping the order as 2 (see Figure 4).

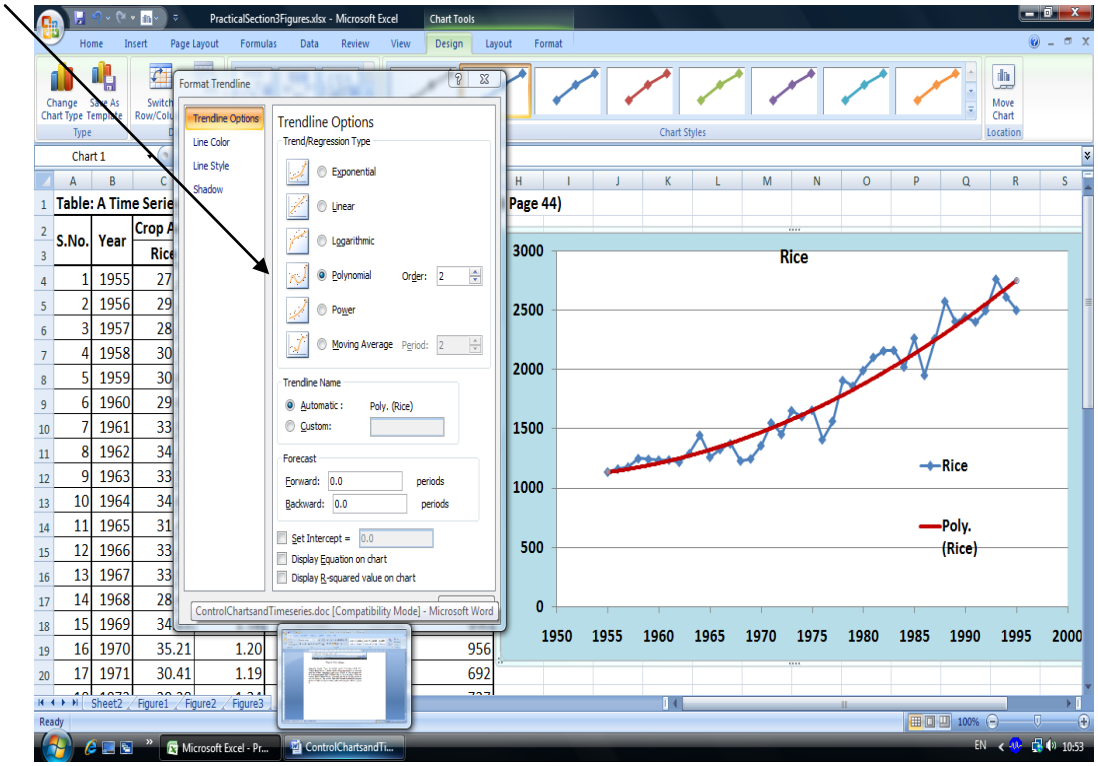
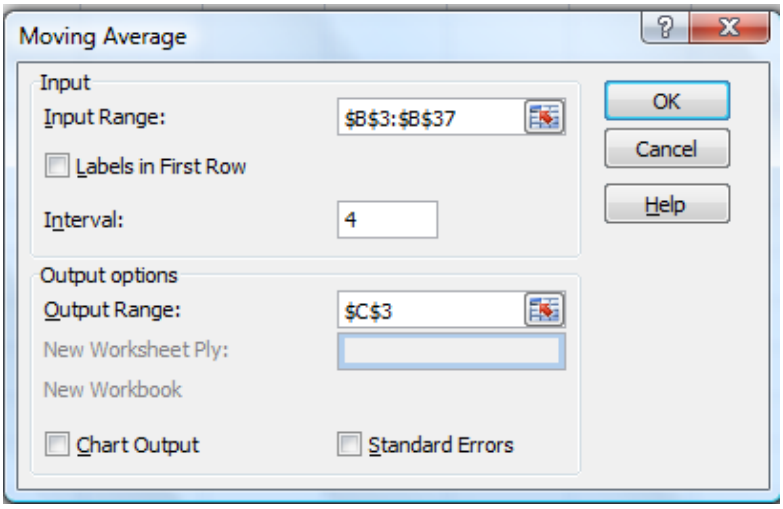


Figure 4: The Polynomial Trendline of Order 2

You can also select the Moving Average Trendline option from the same window. Another alternative approach for calculating moving averages is that you can use *Data → Data Analysis → Moving Average* option. On selection of this option, you will be presented the following Moving Average Window :



Select the input range, interval and output range. If you want a chart of moving averages then you can select Chart Output option box. Figure 5 shows moving averages with various intervals using repeated application of this command.

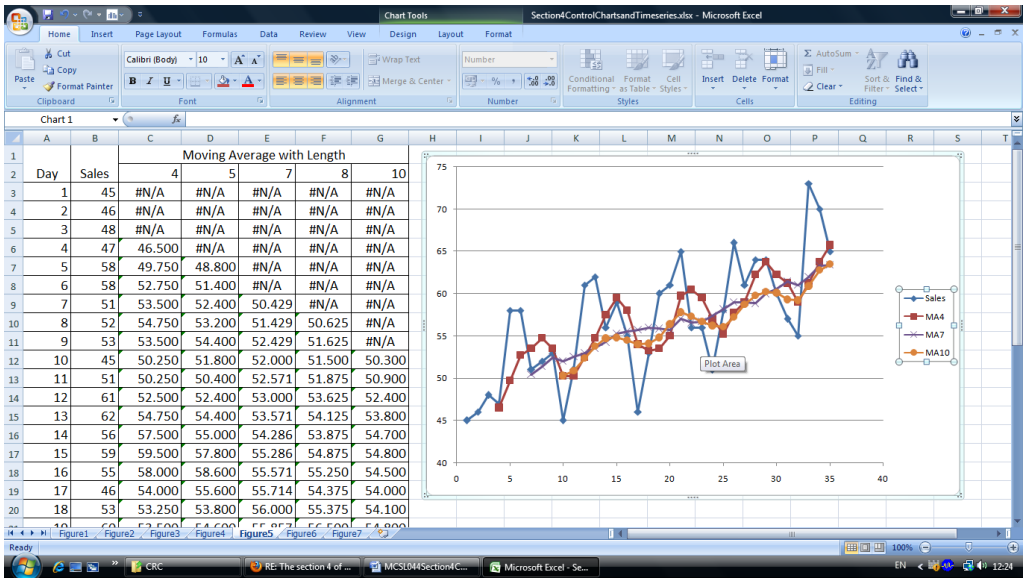


Figure 5: Finding Moving Averages

Thus, you can perform the long term trend analysis using spreadsheet packages.

☞ Check Your Progress 1

1) Develop all the spreadsheets as shown in Figures 1 to Figure 5.

.....

.....

.....

.....

.....

☞ **Lab Sessions 7**

1) Workout all the Examples given in the Unit 10 using spreadsheet package.
2) Workout the Exercises of BCS040/Blockl2/Unit 10 using spreadsheet wherever required, or generate sample data wherever possible using spreadsheet functions of random number generation. Alternatively, look for other examples from references with sample data.

4.3 CONTROL CHARTS

This section discusses about the use of a spreadsheet package in constructing a control chart. You should go through BCS-040 Block 3/ Unit 11 before going through this section. Let us first revise some of the terms used in that Block.

Quality and Quality characteristics: The term Quality may broadly be defined as conformance of a product to certain standards. Quality depends on certain characteristics related to the product. When you design a product, certain specifications or levels of tolerances are established on all important quality characteristics of the product. These characteristics are called the quality. For example, in case of a ball pen, the specifications on refill length may be that it should lie between 9.90 crns and 10.10 crns.

Controlling a Process: The control operations of a process has to be such that the quality characteristics of output product are maintained at the desired levels.

Statistical Process Control: It is a methodology that collects and analyses data on quality characteristics periodically from the process, to take appropriate actions in respect of quality monitoring of a process. Control charts are a most widely used technique for Statistical Process Control.

Variations and Stable/unstable process: A certain amount of variation in quality characteristics (for example, length of refill) is unavoidable despite stable conditions (for example, same machine settings, almost identical material quality and experience of operators). The small changes in quality characteristics may be attributed to small conditional changes. These small causes are inevitable and are called chance causes. The variation caused by such changes is called *chance cause* variation. However, some major chances, such as change in the machine settings, major drop in the quality of raw material etc., disturb the stable process. Such variations must be identified and corrected. Since, you can attribute change to one or more particular reasons, they are called assignable causes. Such causes may make a process unstable.

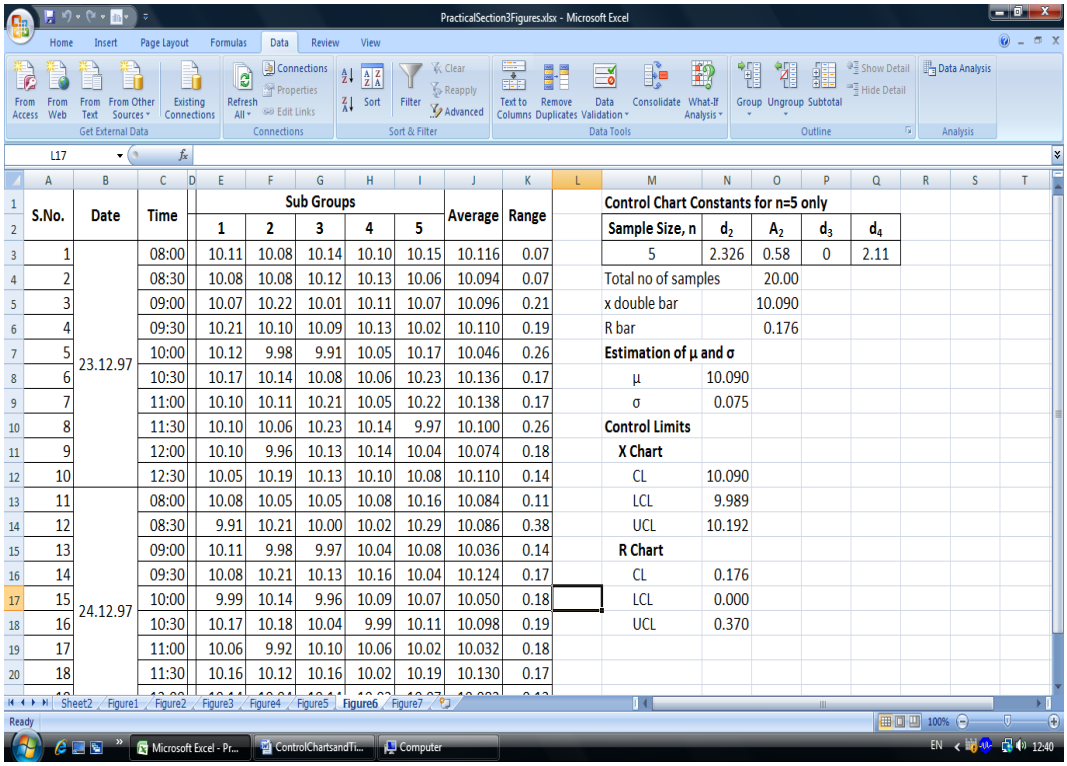
Control Chart Construction steps:

- Collect sample periodically
- Compute quality characteristics for the collected sample
- Plot sample number (x-axis) vs quality characteristic (y-axis)
- Check if the value of quality characteristics is between the *upper* and *lower control limits*.
- In general, the quality characteristics follow normal distribution. Therefore, the upper and lower limits may be determined by using "*Three Sigma Limits*" as: lower limit $\mu - 3\sigma$ and upper limit $\mu + 3\sigma$.
- If quality characteristics are a variable, then control in terms of both the mean and the variability is sought.

Control chart can tell us when process has become unstable, but not that why did it happen?

Problem 1: (Reference MCS040/Block 3/Unit 11/Page 83), Data for the length of refills and control chart constants (only for sample size ($n = 5$) are shown in Figure 6. This data is drawn from the Table 1 and Table 2 of the said Unit. You may now estimate μ and σ from the refill length data and calculate the control limits for $\bar{X}$ and R charts.

The said Unit shows the calculations that are performed with the help of spreadsheet and shown in Figure 6.



S.No.	Date	Time	Sub Groups 1	2	3	4	5	Average	Range
1	23.12.97	08:00	10.11	10.08	10.14	10.10	10.15	10.116	0.07
2		08:30	10.08	10.08	10.12	10.13	10.06	10.094	0.07
3		09:00	10.07	10.22	10.01	10.11	10.07	10.096	0.21
4		09:30	10.21	10.10	10.09	10.13	10.02	10.110	0.19
5		10:00	10.12	9.98	9.91	10.05	10.17	10.046	0.26
6		10:30	10.17	10.14	10.08	10.06	10.23	10.136	0.17
7		11:00	10.10	10.11	10.21	10.05	10.22	10.138	0.17
8		11:30	10.10	10.06	10.23	10.14	9.97	10.100	0.26
9		12:00	10.10	9.96	10.13	10.14	10.04	10.074	0.18
10		12:30	10.05	10.19	10.13	10.10	10.08	10.110	0.14
11	24.12.97	08:00	10.08	10.05	10.05	10.08	10.16	10.084	0.11
12		08:30	9.91	10.21	10.00	10.02	10.29	10.086	0.38
13		09:00	10.11	9.98	9.97	10.04	10.08	10.036	0.14
14		09:30	10.08	10.21	10.13	10.16	10.04	10.124	0.17
15		10:00	9.99	10.14	9.96	10.09	10.07	10.050	0.18
16		10:30	10.17	10.18	10.04	9.99	10.11	10.098	0.19
17		11:00	10.06	9.92	10.10	10.06	10.02	10.032	0.18
18		11:30	10.16	10.12	10.16	10.02	10.19	10.130	0.17

Control Chart Constants for n=5 only				
Sample Size, n	d_2	A_2	d_3	d_4
5	2.326	0.58	0	2.11

Total no of samples		20.00
x double bar		10.090
R bar		0.176
Estimation of μ and σ		
μ	10.090	
σ	0.075	
Control Limits		
X Chart		
CL	10.090	
LCL	9.989	
UCL	10.192	
R Chart		
CL	0.176	
LCL	0.000	
UCL	0.370	

Figure 6: Calculation of μ and σ for refill length

For creating Figure 6, you can use your own formulas.

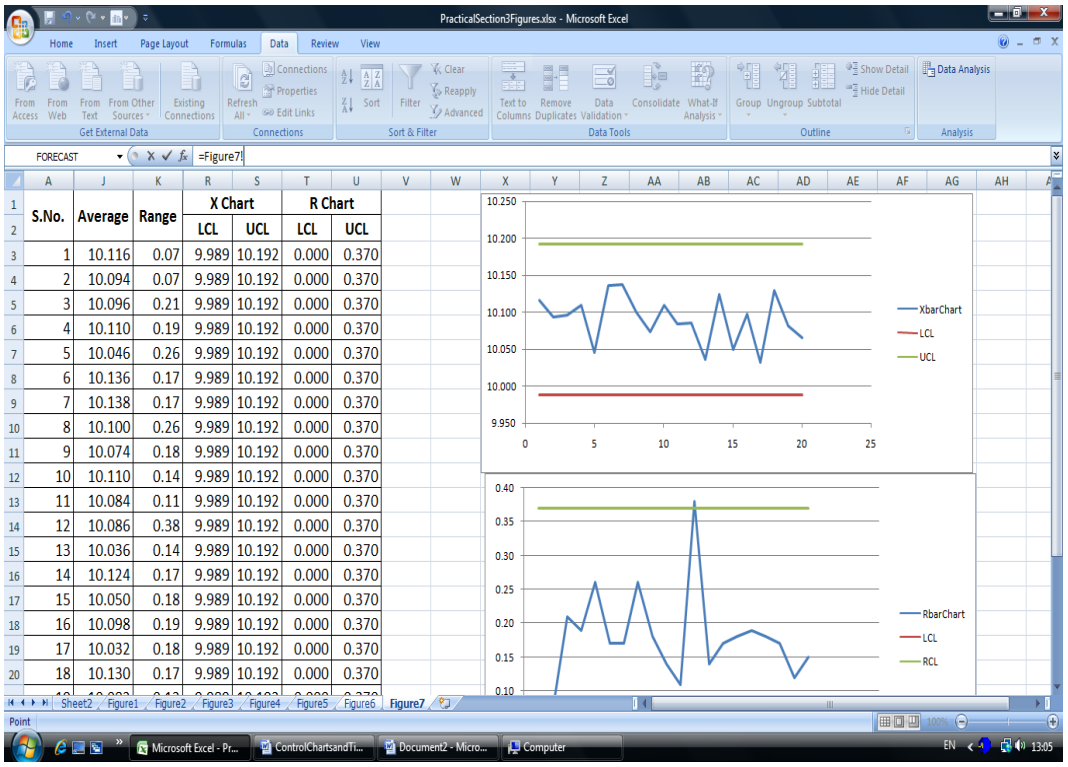
Steps

- You first create the sample data.
- The mean of the sample subgroup at a particular time is found in cell J3 using the formula $\bar{x}_i = AVERAGE(E3{:}I3)$ and similarly
- the Range for this subgroup is calculated at cell K3 using the formula $R_i = MAX(E3{:}I3) - MIN(E3{:}I3)$. For $i = 1,2,3,4,5$.
- On the basis of these mean and range values $\bar{\bar{x}}$ and $\bar{R}$ is calculated in the cell O5 using the formula *=SUM(J3:J22)/O4* and cell O6 using the formula *=SUM(K3:K22)/O4* respectively.
- On the basis of these values and control chart constants CL, LCL and UCL values are calculated for the $\bar{X}$ and R chart. Write the formula for these calculations and compare the values with the values calculated in the Unit.

Further,

- on the basis of these calculated values, you can create data as shown in Figure 7.
- This data is used to draw the control chart.
- You may use Scatter Plot to draw the charts as shown in Figure 7.
- You may keep the x-axis same for all the Ranges that are being plotted as Y axis.

Please note that in Figure 7, that the 12th point on the R-chart crosses the UCL line, which is indicative of "process out of control" situation with respect to variability (as this problem has been detected in R-chart). Consequently, we conclude that during this time period the process may not be stable.



S.No.	Average	Range	X Chart		R Chart	
			LCL	UCL	LCL	UCL
1	10.116	0.07	9.989	10.192	0.000	0.370
2	10.094	0.07	9.989	10.192	0.000	0.370
3	10.096	0.21	9.989	10.192	0.000	0.370
4	10.110	0.19	9.989	10.192	0.000	0.370
5	10.046	0.26	9.989	10.192	0.000	0.370
6	10.136	0.17	9.989	10.192	0.000	0.370
7	10.138	0.17	9.989	10.192	0.000	0.370
8	10.100	0.26	9.989	10.192	0.000	0.370
9	10.074	0.18	9.989	10.192	0.000	0.370
10	10.110	0.14	9.989	10.192	0.000	0.370
11	10.084	0.11	9.989	10.192	0.000	0.370
12	10.086	0.38	9.989	10.192	0.000	0.370
13	10.036	0.14	9.989	10.192	0.000	0.370
14	10.124	0.17	9.989	10.192	0.000	0.370
15	10.050	0.18	9.989	10.192	0.000	0.370
16	10.098	0.19	9.989	10.192	0.000	0.370
17	10.032	0.18	9.989	10.192	0.000	0.370
18	10.130	0.17	9.989	10.192	0.000	0.370

Figure 7: Control Charts (x-chart and R-chart) for the refill length data

Now you should try the following lab session:

☞ Check Your Progress 2

1) Develop the spreadsheets as shown in Figures 6 and Figure 7.

.....

.....

.....

.....

.....

☞ **Lab Sessions 8**

1) Perform all the examples given in the Unit 11 using spreadsheet package including example given in section 11.3.5 on control charts for attributes.

2) Perform all Exercises of BCS040/Block3/Unit 11 that you can perform using spreadsheet.

3) Write simple C functions to calculate various values required for drawing control charts.

4.4 SUMMARY

This section was an attempt to provide you details on some of the elementary methods of Unit 10 based on which you can perform time series analysis and forecasting using the spreadsheet. Time series analysis helped in performing analysis of data in terms of the components and the models discussed in the said Unit 10. The example demonstrated the use of spreadsheet package functions in evaluating trend.

This section also covered computational aspects relating to control charts in statistics using spreadsheet package. The purpose of these charts was to determine the process under statistical control.

4.5 ANSWERS TO CHECK YOUR PROGRESS

Check Your Progress 1:

Please use any spreadsheet package and enter all the data as shown in Figure 1 to Figure 5. You may download the file containing all these figures from the website www.ignou.ac.in from BCA page under MCSL044. Find out what formulas have been entered. Also find the errors in the entry of formula.

Check Your Progress 2:

Please use any spreadsheet package and enter all the data as shown in Figure 6 and Figure 7. You may download the file containing all these figures from the website www.ignou.ac.in from BCA page under MCSL044. Find out what formulas have been entered. Also find the errors in the entry of formula.

4.6 FURTHER READINGS

1. BCS-040: Statistical Techniques, IGNOU Material.

Web link:

- www.wikipedia.org