User’s Guide

Premium SQC for Excel™

Version 3.02f

Disclaimer

Microsoft® and Excel are registered trademarks of Microsoft Cooperation.

SQC for Excel™ and **Premium SQC for Excel™** are provided "as is" without warranty. To the best of our knowledge, the statistical calculations provided in the package are accurate. However, this validity is not guaranteed. The user is responsible for decisions made based on the results generated by the package.

The spreadsheets generated by the **SQC for Excel™** or **Premium SQC for Excel™** program can only be used when a license for the **SQC for Excel™** or **Premium SQC for Excel™** programs are purchased. Please do **not** distribute the stand-alone SQC worksheets generated by the **SQC for Excel™** or **Premium SQC for Excel™** without written permission of BaRaN Systems Ltd.

BaRaN Systems (BSL)
458 Plantation Dr. PMB 100
Lake Jackson, TX 77566

Phone: (403) 449-6554
Web www.baran-systems.com

Document No. Sfx_Premium_101.doc

Contents

Introduction

Before You Begin

Complete and mail the registration card.

Make a back-up of your original disk. Do not use your original disk as your working copy. If you happen to overwrite it, you will have lost your **Premium SQC for Excel™** program. Keep your original disk locked. Either copy it to your hard disk or make a working copy on another floppy.

The spreadsheets generated by the **Premium SQC for Excel™** or **SQC for Excel™** programs can only be used when a license for the **Premium SQC for Excel™** or **SQC for Excel™** program are purchased. Please do **not** distribute the stand-alone SQC worksheets generated by the **SQC for Excel™** or **Premium SQC for Excel™** without written permission of BaRaN Systems Ltd.

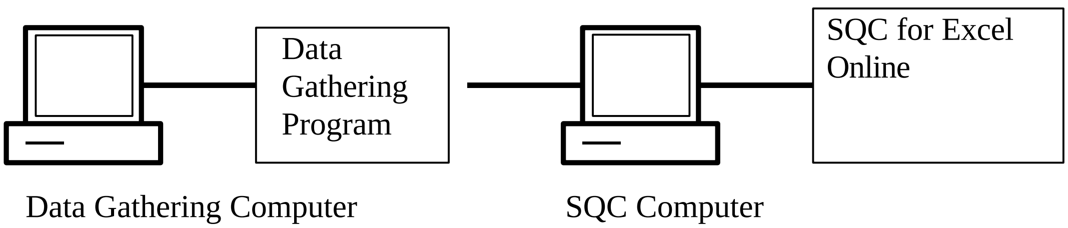
Microsoft® Excel Version

The **SQC for Excel™** and **Premium SQC for Excel™** are developed for **Microsoft® Excel 97** or version 8 or **Microsoft® Excel 2000** or version 9.

Online SQC for Excel™

Introduction

The **Online SQC for Excel**™ allows the users to connect their Excel SQC charts to the auxiliary online measurement. The Access to your DCS (Digital Computer Systems) or any other computer should be functional before installing this program.



How the Program Works

The Online program utilizes the following steps:

1- Input Data Processing

The program runs on an internal timer set by the user. At a fixed sampling time, Ts. The program runs to update the history buffer and all appropriate calculations and reports.

The program assumes that at each sample time, Ts, the data fields are updated by an auxiliary program.

The data fields can be set to collect data directly with programs such as @aGlance. In this case the user connects the data fields to the auxiliary program manually.

The Data can be collected automatically with an External Data Gathering Program (see the last section of this chapter).

2- History Buffer

When the online option is activated, at each sample time of Ts, the program will copy the data fields from the ONLINE worksheet to the history buffer in the DATA worksheet.

ONLINE worksheet	DATA Worksheet
Ts = 1.0 Min. 10:45 13 12 11 12	10:30 10 19 18 17 10:31 11 11 12 14 10:32 12 12 14 12 10:33 13 12 11 11 … 10:44 14 13 12 11 10:45 13 12 11 12

For example at 10:45 a new sample arrives and the program copies it to the history buffer.

ONLINE worksheet	DATA Worksheet
Ts = 1.0 Min. 10:46 14 14 10 12	10:30 10 19 18 17 10:31 11 11 12 14 10:32 12 12 14 12 10:33 13 12 11 11 … 10:44 14 13 12 11 10:45 13 12 11 12 10:46 14 14 10 12

At 10:46 the program executes again and copies the 10:46 data to the history buffer.

The maximum number of records in the history buffer, Nh, is set to 1000 records. Then the user can change this setting by choosing Setup from SQC menu. The new data is copied from the ONLINE worksheet and is placed at the end of the history buffer until the buffer size reaches the Nh. If the number of records in the history buffer exceed the Nh, the program automatically scrolls the buffer and deletes the first record (or the oldest record).

Record 1	Record 2	Record 3	Record 4
Record 2	Record 3	Record 4	Record 5
Record 3	Record 4	Record 5	Record 6
…	…	…	…
Record 999	Record 1000	Record 1001	Record 1002
Record 1000	Record 1001	Record 1002	Record 1003

3- Calculation Buffer

The SQC for Excel™ has an internal calculation buffer, Nc. This is the number of RECORDS or number of points in the SQC chart. You set this number by selecting the initial SCQ data. For example the “Quick Tour Example had 15 sub-groups or 15 records).

The **Online SQC for Excel**™ program will connect the CALCULATION worksheet data buffer to the newest data in the history buffer of the DATA worksheet.

Ts = 1.0 Min.

10:45 13 12 11 12

ONLINE worksheet

10:30 10 19 18 17
10:31 11 11 12 14
10:32 12 12 14 12
10:33 13 12 11 11
…
10:44 14 13 12 11
10:45 13 12 11 12

DATA Worksheet

01 10:31 11 11 12 14
02 10:32 12 12 14 12
03 10:33 13 12 11 11
…
14 10:44 14 13 12 11
15 10:45 13 12 11 12

CALCULATION Worksheet

Ts = 1.0 Min.

10:46 14 14 10 12

ONLINE worksheet

10:30 10 19 18 17
10:31 11 11 12 14
10:32 12 12 14 12
10:33 13 12 11 11
…
10:44 14 13 12 11
10:45 13 12 11 12
10:46 14 14 10 12

DATA Worksheet

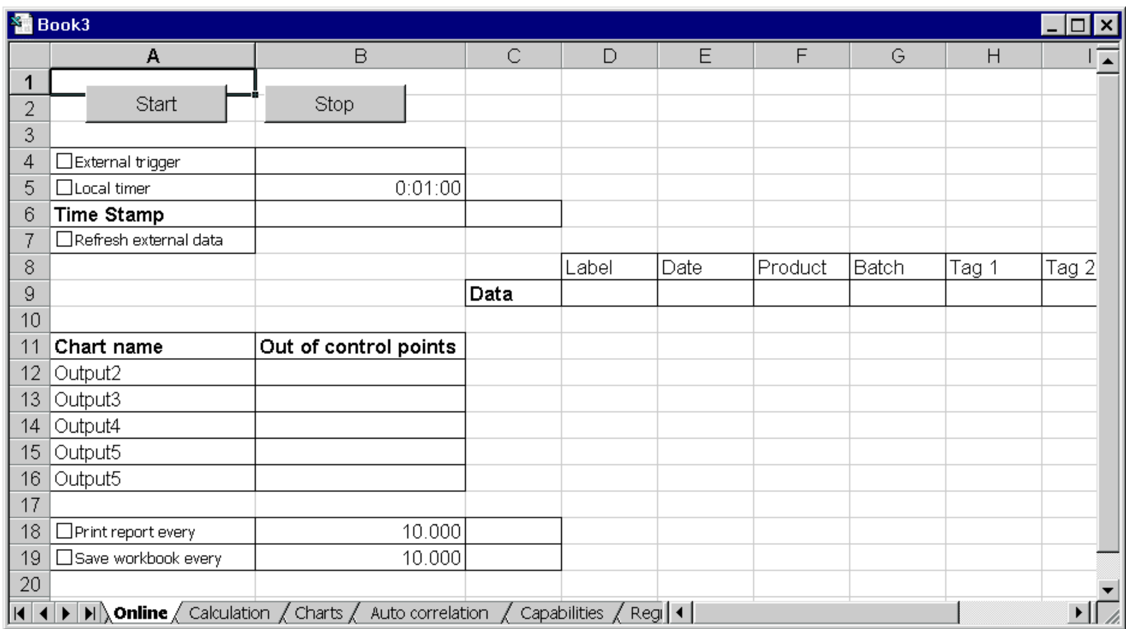
01 10:32 12 12 14 12
02 10:33 13 12 11 11
03 10:34 12 11 14 15
…
14 10:45 13 12 11 12
15 10:46 14 14 10 12

CALCULATION Worksheet

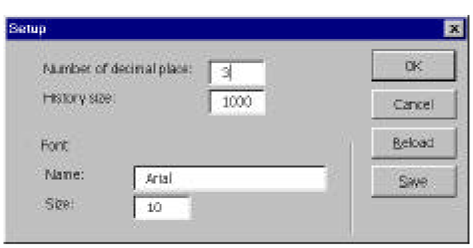
How to Setup Online Parameters

The Online Parameters are:

Variable	Description	Value
Start and Stop Button	These buttons start and stop the **Premium SQC for Excel™** program.	
External Trigger	The content of the cell B3 is changed externally.	
Internal Trigger	The program runs at sample Time, Ts, entered in the cell B4.	
Sample Time (Ts)	The time program will run if the Internal trigger is selected	
Data	The data to be used by Premium SQC for Excel™	
Output	The data to be sent to the External source	
Reports and Saving	This option allows the user to print a report and save the data to a disk.	



From SQC menu, choose Setup to change the history buffer size.



Quick Tour

Purpose of the Quick Tour

The purpose of the quick tour is to provide the user with a quick overview of the functionality of **SQC for Excel** by stepping through an example problem. By so doing, the user will become familiar with the basic steps involved in doing an analysis and will readily acquire a feel for the tool.

The example involves analyzing data from a manufacturing process. Five observations a day have been made over a fifteen day period and tabulated in the following table.

Day	No 1	No 2	No 3	No 4	No 5
1	13.6	12.6	13.2	13.1	12.1
2	13.5	12.8	13.0	12.8	12.4
3	13.0	13.0	12.1	12.2	13.3
4	13.9	12.1	12.7	13.4	13.0
5	14.4	12.4	12.6	12.4	12.5
6	13.3	12.8	13.0	13.0	13.1
7	13.4	13.3	12.0	13.0	13.1
8	14.2	12.7	12.2	12.9	12.5
9	14.0	13.2	12.4	16.0	13.0
10	14.1	13.7	13.4	12.2	12.5
11	13.3	12.7	12.6	13.0	12.7
12	13.9	12.4	12.7	12.4	12.8
13	13.2	12.3	12.6	13.1	12.7
14	13.2	12.8	12.8	12.3	12.6
15	13.3	12.4	13.0	12.3	12.2

The task is to develop online SQC charts for this data. It assumes that the external program will generate 5 data points every minute.

Files Required

There is one file required to run this example:

1. Process_Data worksheet in the **Cases.xls file**

Flow of Calculations

There are several steps in transforming a set of raw data points from an Excel worksheet into a set of charts or reports. These steps include:

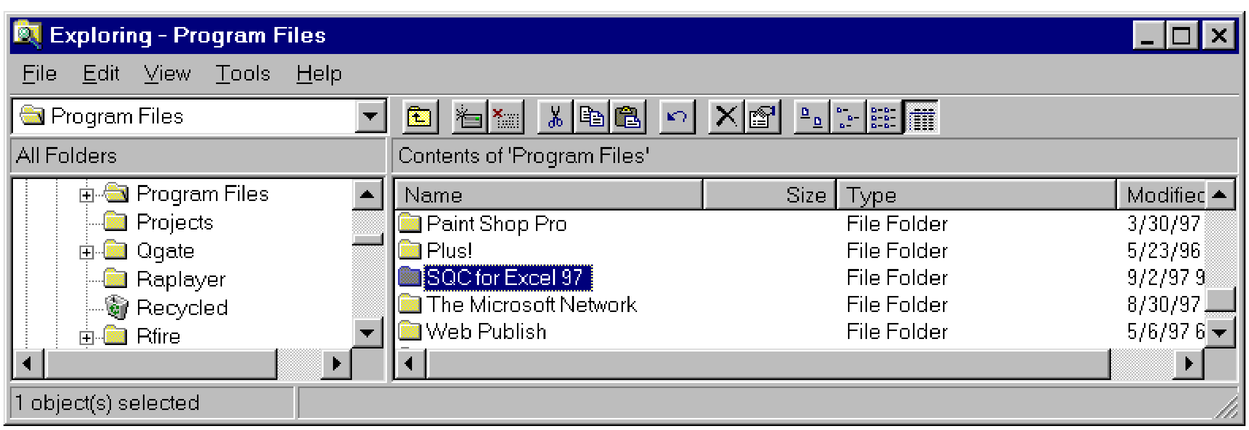
a- Opening the initial data file
b- Installing the **SQC for Excel™ program** (see installation chapter)
c- Loading the data to a working SQC spreadsheet.
d- Analyzing the data.
e- Adding **Premium SQC for Excel™** parameters
f- Setting the external data generator
g- Starting the **Premium SQC for Excel™**
h- Checking the results

Once the data has been selected and the load data option has been selected, flow of calculations will automatically proceed to the plotting phase. The report option needs to be specifically selected to generate printed output. Once a set of calculations has been completed, the user is able to restart the sequence at any point.

a- Opening the initial data file

The **SQC for Excel™** can be loaded into the Microsoft® Excel using the following procedure:

1. Start Microsoft® Excel
2. From **File** menu, choose **Open** and select **Cases.xls** file in the **Program Files➔ SQC for Excel 97** folder.

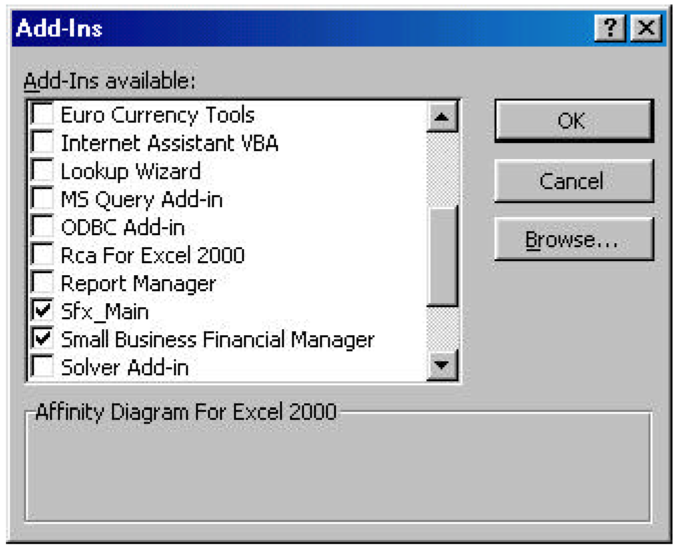


3. Click **OK**

b- Installing the SQC for Excel™

1. Check to see if you have **SQC** on the Excel menu bar, if not please do the following:

    a. From **Tools** menu, choose **Add-in…**



    b. Make sure that the **SFX_Main** add-in is checked
    c. Click **OK**

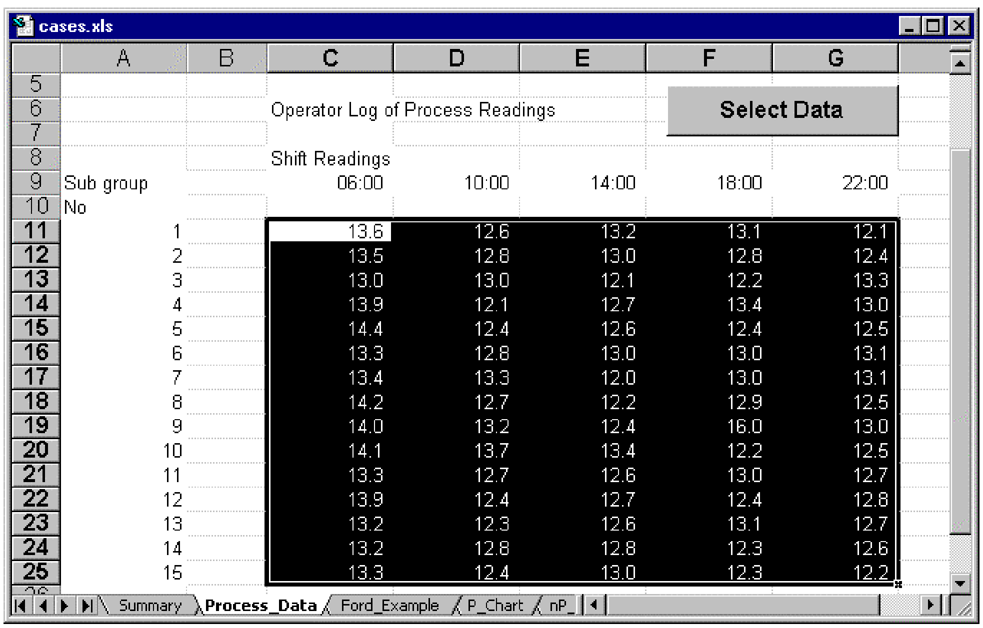
1. **Note**: If SFX**_Main SQC for Excel™** is not shown in the add-in list, please follow the directions given in the installation chapter or add it in by clicking on the **Browse…** button and look at Program **Files➔ SQC for Excel** folder**.**

c- Loading the data to a working SQC spreadsheet

Only data that has been selected on the raw data worksheet will be analyzed. To select data on the raw data worksheet, use the mouse or keypad to highlight the area of cells, which represents the data of interest. In the following example, the whole data set representing fifteen days of observations is selected.

Select data (C11 to G25)

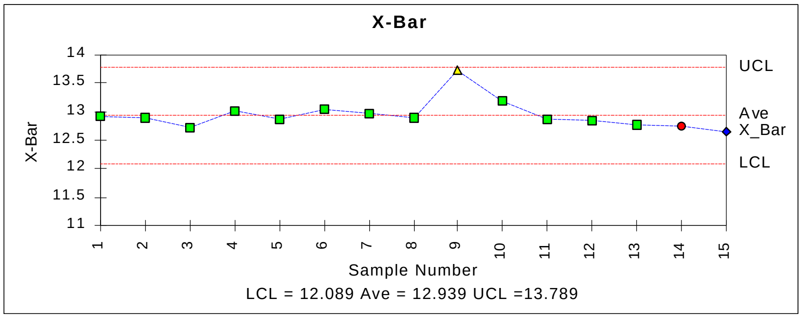
1. From **Window** menu, choose file **Cases.xls**
2. Select the **Process_Data** worksheet
3. Select the first data point (i.e. click on cell **C11**)
4. Hold the mouse button down and drag to the last data point (i.e. cell **G25**) or click on the **Select Data** Button.

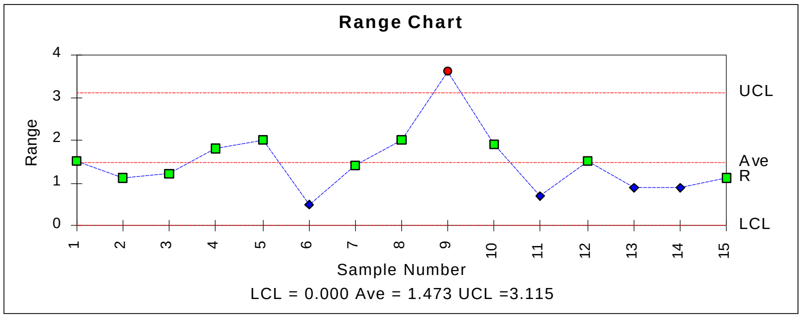


Sub group No	06:00	10:00	14:00	18:00	22:00
1	13.6	12.6	13.2	13.1	12.1
2	13.5	12.8	13.0	12.8	12.4
3	13.0	13.0	12.1	12.2	13.3
4	13.9	12.1	12.7	13.4	13.0
5	14.4	12.4	12.6	12.4	12.5
6	13.3	12.8	13.0	13.0	13.1
7	13.4	13.3	12.0	13.0	13.1
8	14.2	12.7	12.2	12.9	12.5
9	14.0	13.2	12.4	16.0	13.0
10	14.1	13.7	13.4	12.2	12.5
11	13.3	12.7	12.6	13.0	12.7
12	13.9	12.4	12.7	12.4	12.8
13	13.2	12.3	12.6	13.1	12.7
14	13.2	12.8	12.8	12.3	12.6
15	13.3	12.4	13.0	12.3	12.2

e- Analyzing the data.

Follow the instruction in the Quick Tour Chapter in the **SQC for Excel™** User manual.

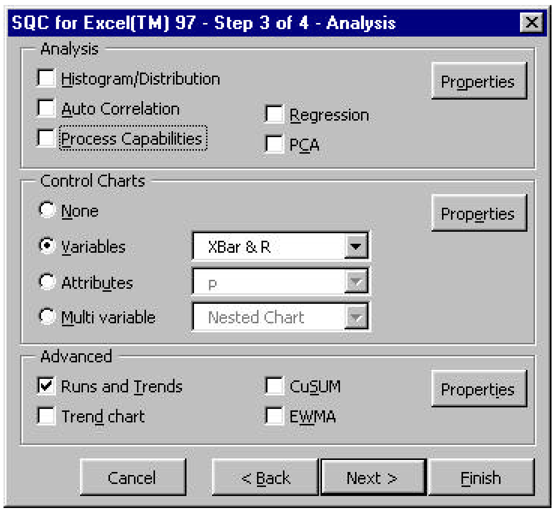




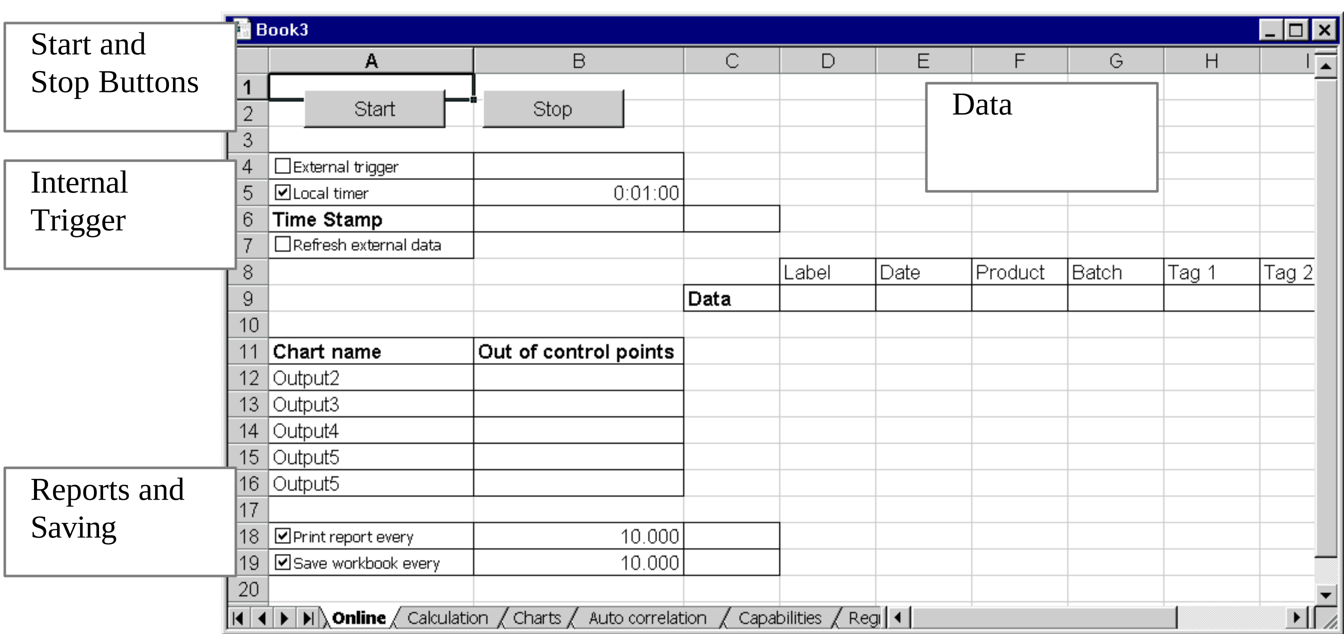
f- Adding Online SQC for Excel™ parameters

The **Premium SQC for Excel™** updates all worksheets selected in the Analysis dialog. To ensure that the **Premium SQC for Excel™** program does all calculations on time, please follow these steps:

1- From the **SQC** menu, choose the **Analysis** option.
2- Change the parameters to get the following setup.



3- Click on the NEXT button.
4- Select the worksheet TAB labeled "Online".



5- Set the following Parameters

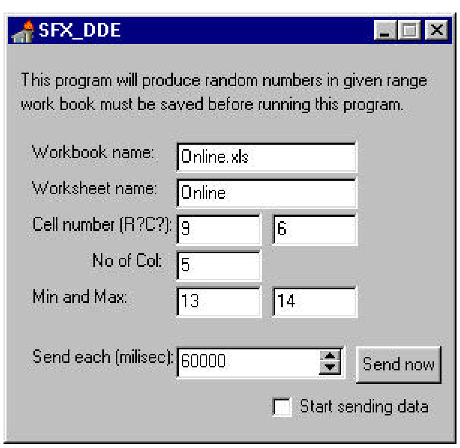
Variable	Description	Value
Start and Stop Button	These buttons start and stop the **Premium SQC for Excel™** program.	Do not use now
Trigger	External: The content of the cell B3 is changed externally. Internal: The program works at frequency entered in cell B4.	Select Internal trigger and enter 00:01:00 for 1 minute in cell B4.
Data	The data to be used by Premium SQC for Excel™	Enter =Now() in cell B8
Output	The data to be sent to the External source	Enter the UCL and LCL from the calculation worksheet. B13 → =Calculation!C12 B14 → =Calculation!C11
Reports and Saving	This option allows the user to print a report and save the data to a disk.	Select both options

6- Save the Spreadsheet as ONLINE.XLS

f- Setting the external data generator

Do the following steps to start the SQC_DDE.

1- Double click on SFX_DDE.EXE file in the SQC folder



Spreadsheet name

Worksheet Name

Cell reference

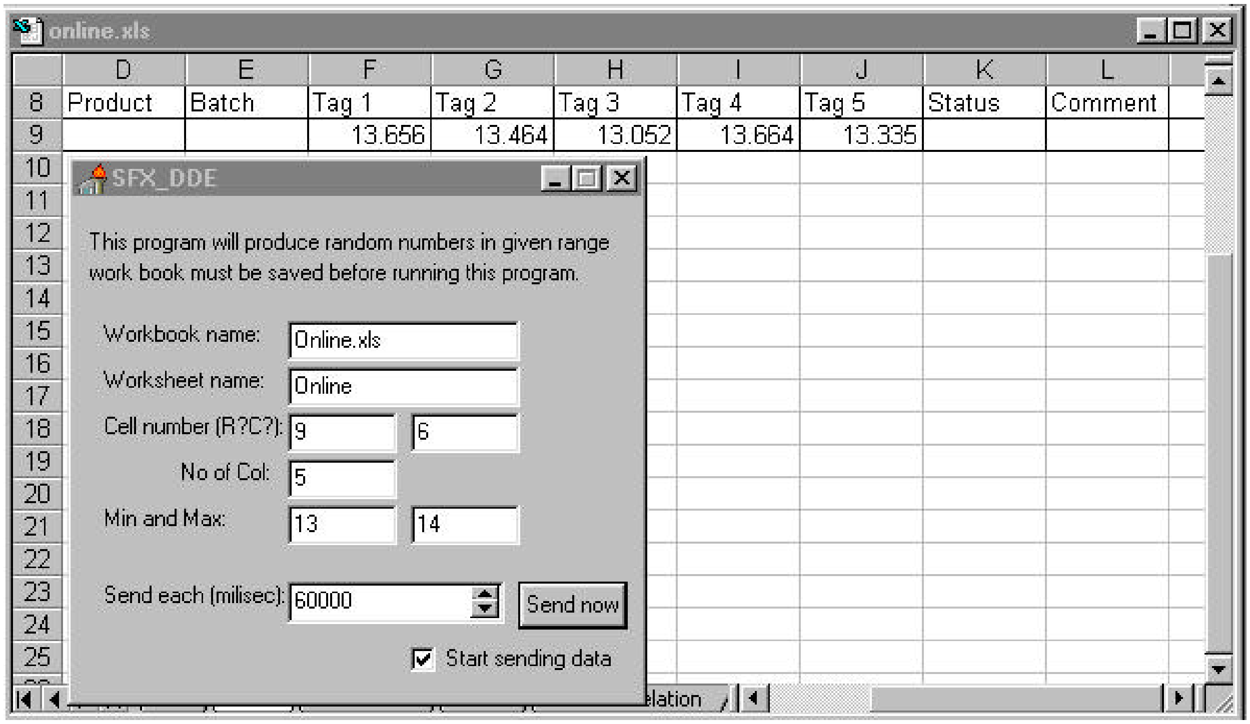
Number of columns

Min and Max

Frequency

2- Set the parameters as shown above

3- Check mark the Start Sending data

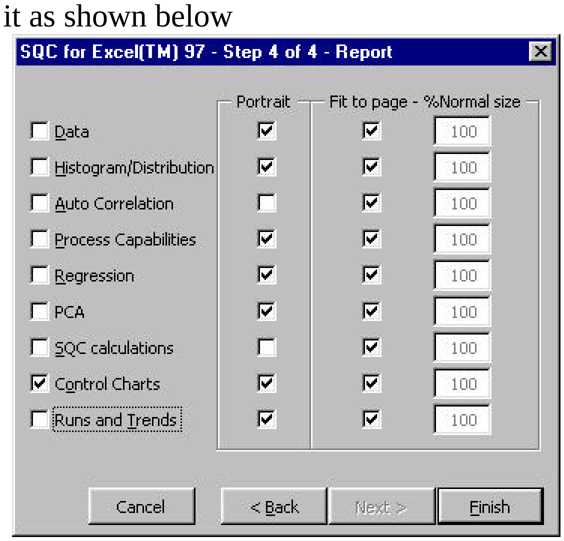


4- Check that the values in Data row in the ONLINE.XLS file is changing.

g- Starting the Online SQC for Excel™

Once the external data is being forwarded to the **Online.XLS**, you can start the **Premium SQC for Excel™** program with the following steps:

1- From SQC menu, choose Report and set it as shown below

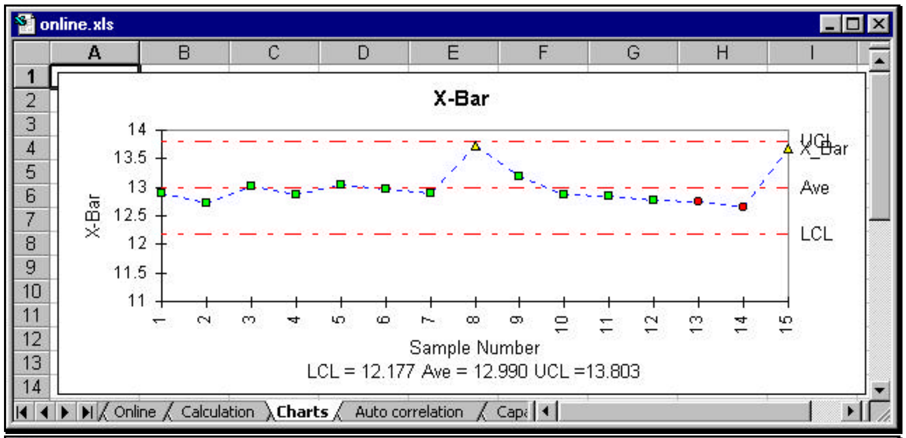


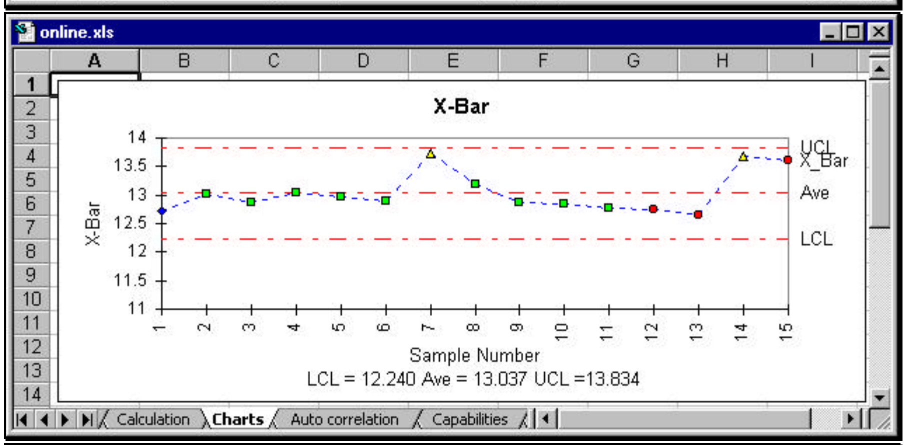
2- In the **Online** worksheet click on the **Start** button.

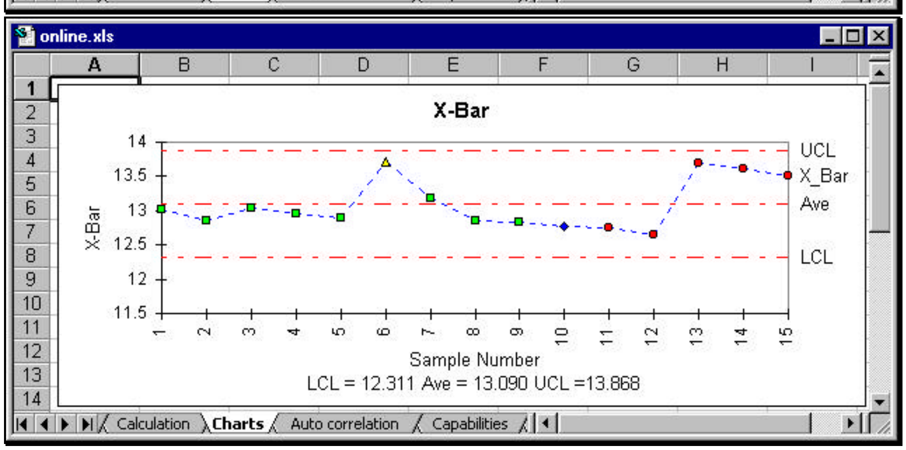
Note: You need to click the STOP button to stop the online program. In case of emergency press the ESCAPE key and then click on the END button. The ESCAPE key will make the program terminate incorrectly and you will need to start over again.

h- Checking the results

The SQC for Excel™ program automatically updates all relevant calculations and charts in the Online spreadsheet.



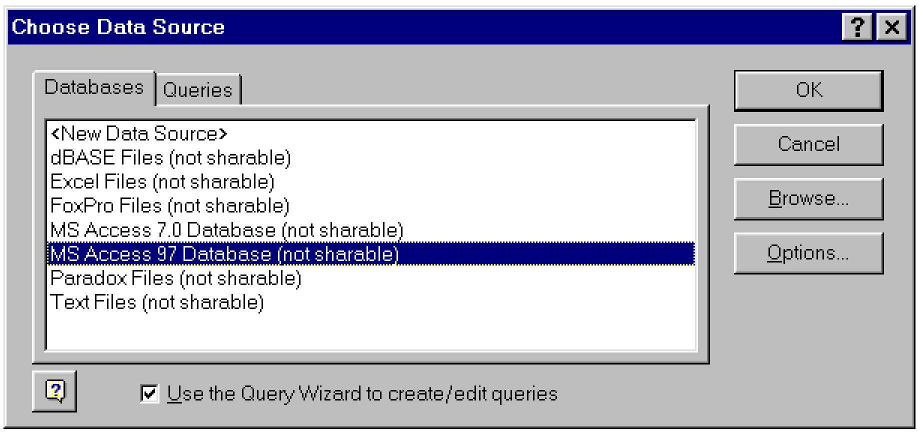




Online SQC Using External Database

Introduction

The Online functionality of SQC for Excel™ can be used with a variety of external databases. The supported databases are listed below:-



Once you have your data updated in one of the above databases, you can connect it to the SQC for Excel™ for online observation. The following example provides the information and step by step instructions on how to setup the connections between SQC for Excel™ and Microsoft® Access 97. You can use this example to set connections with other databases.

How to set the Access Database

The SQC for Excel™ data input should be EXACTLY the way the data is arranged in the DATA worksheet for your SQC charts. The data worksheet for the Online Quick tour is constructed as:

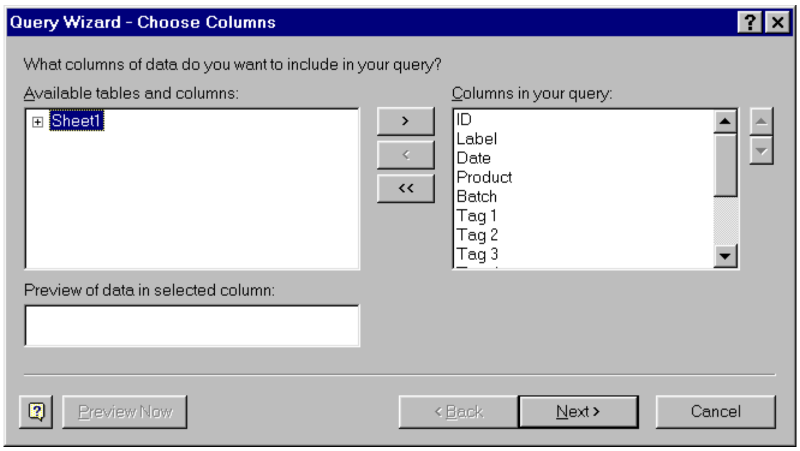
Label	Date	Product	Batch	Tag 1	Tag 2	Tag 3	Tag 4	Tag 5	Status	Comment
				13.6	12.6	13.2	13.1	12.1		
				13.5	12.8	13.0	12.8	12.4		
				13.0	13.0	12.1	12.2	13.3		
				13.9	12.1	12.7	13.4	13.0		

Therefore you need to create an Access database with exactly the same fields and connect your Access database to the data-gathering program.

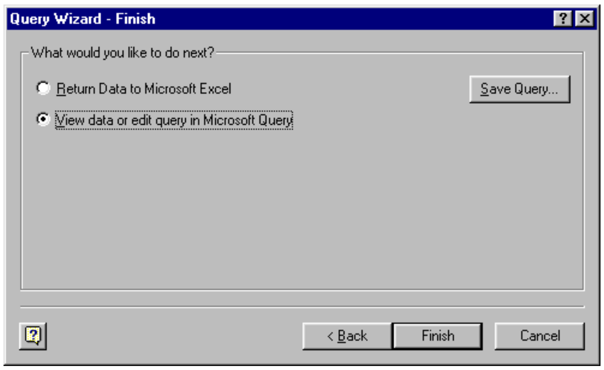
How to set the Online Worksheet

Once the database is setup and connected to the data-gathering program (i.e. collecting data at every minute), you need to setup the ONLINE worksheet. Please follow these steps:-

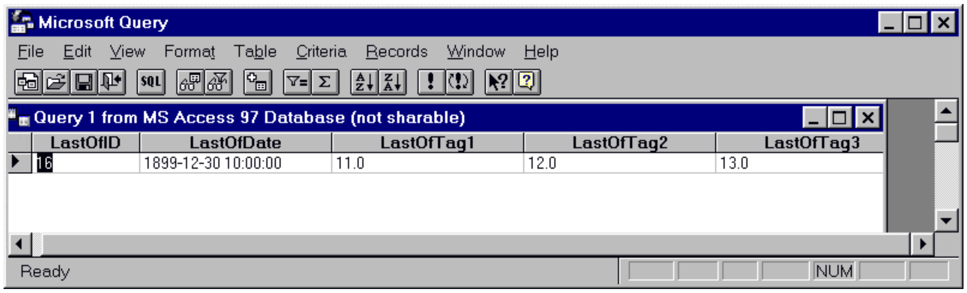
Step 1: Select the ONLINE worksheet.
Step 2: From **Data** menu, choose **Get External Data ➔ Create New Query…**option.
Step 3: Choose "MS Access 97 Database" and click OK.
Step 4: Click on your data (Sheet1) and click on the ">" arrow to copy all fields across.



Step 5: Click on NEXT button until the following dialog.



Step 7: Select the "View Data… " and click on the **Finish** button.

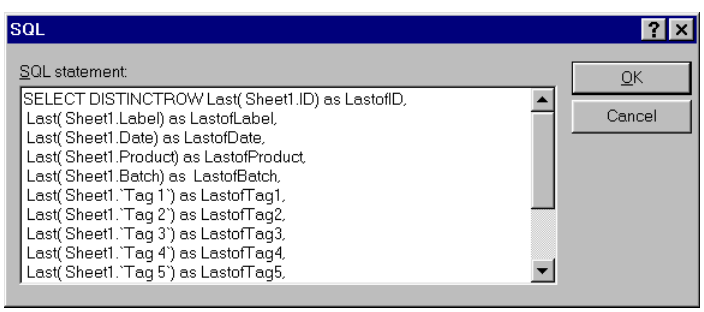


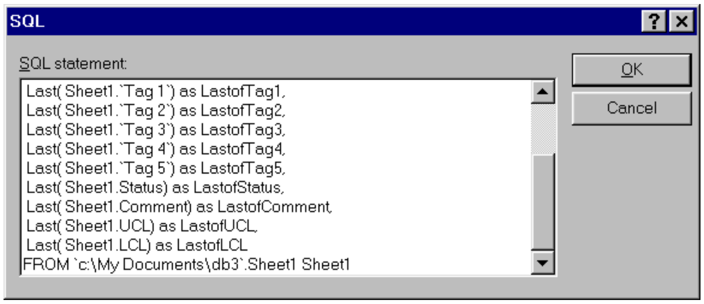
Step 8: Write the following SQL commands in a word processor or Excel and copy them to the local buffer.

```
SELECT DISTINCTROW
Last( Sheet1.ID) as LastofID,
Last( Sheet1.Label) as LastofLabel,
Last( Sheet1.Date) as LastofDate,
Last( Sheet1.Product) as LastofProduct,
Last( Sheet1.Batch) as  LastofBatch,
Last( Sheet1.`Tag 1`) as LastofTag1,
Last( Sheet1.`Tag 2`) as LastofTag2,
Last( Sheet1.`Tag 3`) as LastofTag3,
Last( Sheet1.`Tag 4`) as LastofTag4,
Last( Sheet1.`Tag 5`) as LastofTag5,
Last( Sheet1.Status) as LastofStatus,
Last( Sheet1.Comment) as LastofComment,
Last( Sheet1.UCL) as LastofUCL,
Last( Sheet1.LCL) as LastofLCL
FROM `c:\My Documents\db3`.Sheet1 Sheet1
```

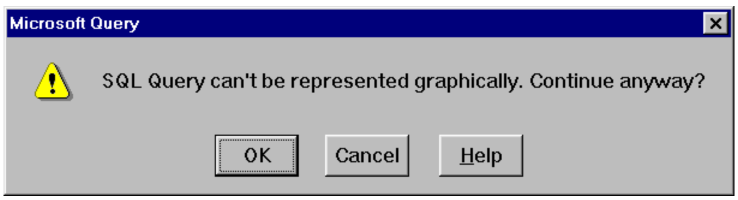
Note that all of the fields are shown here. The Tag 1 is placed in a single apostrophe to preserve the space. The address of the database should be the address of your database.

Step 9: Click on the SQL button on the QUERY tool bar or choose SQL from View menu. Paste the above commands instead of the one that exists there. The above SQL will update the LAST record of the database for SQC for Excel™ access. Click the OK button.





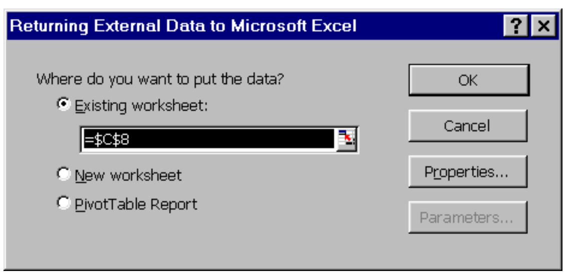
Step 10: Click on the following dialog.



Note: After this dialog you should not have any errors. However if you get one check your SQL commands.

Step 11: From File Menu, choose Close and Return Data to Excel.

Step 12: Enter Cell C8 as the data entry

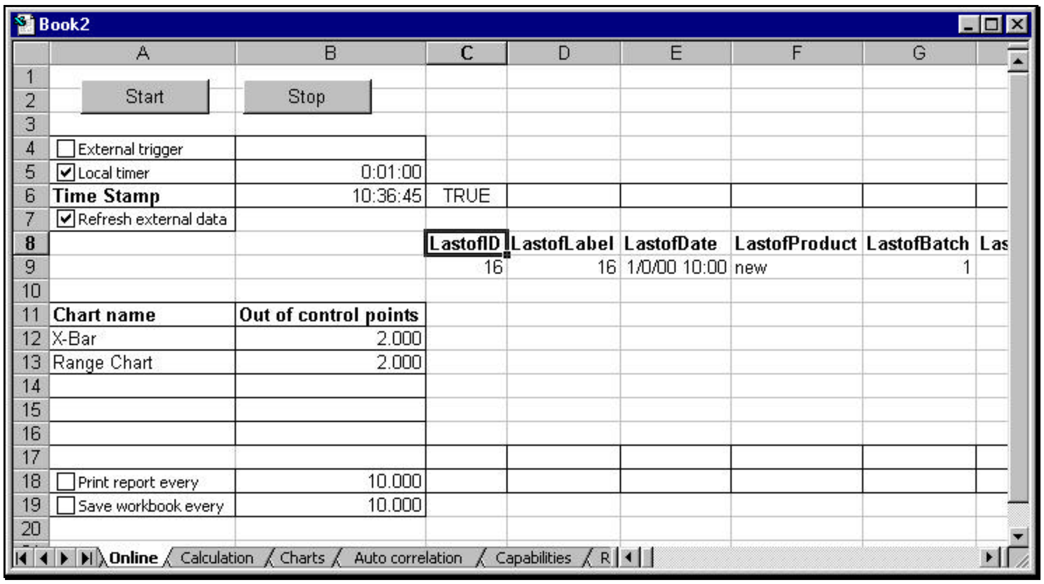


How to Start the Online Program

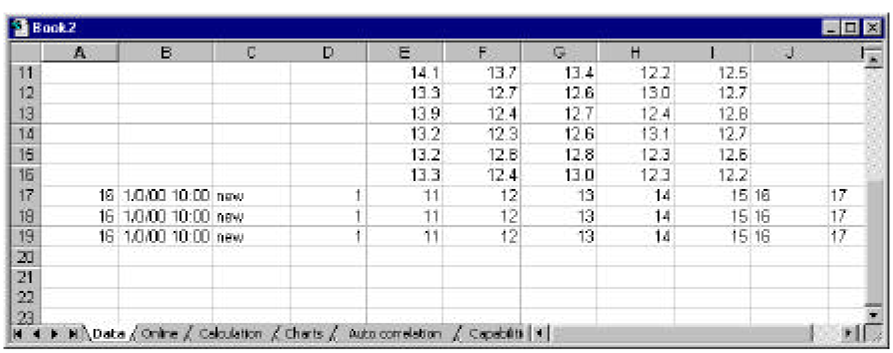
Once the external database is connected to ONLINE worksheet, please do the following steps:

Step 1: Make sure you clicked the

a- Local time check box
b- Refresh External data
c- Entered the sample time of 0:0:1



Step 2: Click the Start button. Note that the SQC menu on the menu bar is now dimmed. You need to click the Stop button to stop the Online SQC for Excel™ program. After a minute you will see the program writes the time stamp in cell B5 and copies the data to all appropriate spreadsheets. For example the CALCULATION worksheet after three minutes is shown.



Nested Charts

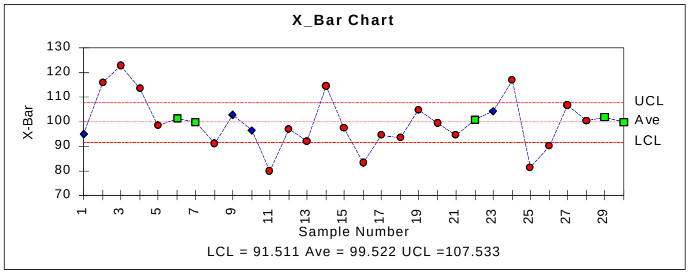
Introduction

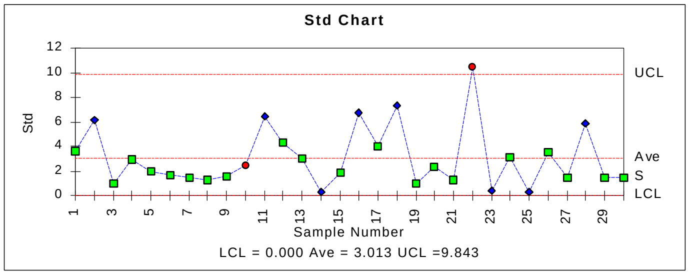
This chapter describes how to construct nested SQC charts. This allows the user to examine multiple sources of variations.

Suppose a process generates 15 batches of a product For each batch 2 samples are taken and the thickness of each sample is measured 2 times. Using the standard SQC charting techniques described in the previous sections, we can analyze the process using X-Bar and S Charts:

	A	B	All				
First SubGroup	1.000	1.000	1.000	Limits are calculated using zone			AB
Last SubGroup	30.000	30.000	30.000				
Std-							
Average	3.013	3.013	3.013				
LCL	0.000	0.000	0.000	Std_Ave * [0.000 B3]			
UCL	9.843	9.843	9.843	Std_Ave * [3.267 B4]			
X bar-							
Average	99.522	99.522	99.522				
LCL	91.511	91.511	91.511	X_BaStd_Ave - Std_Ave * [2.659 A3]			
UCL	107.533	107.533	107.533	X_BaStd_Ave + Std_Ave * [2.659 A3]			
X Average	99.522	99.522	99.522				
Sigma	3.776	3.776	3.776	Std_Ave / [0.798 C4]			
LCL	91.970	91.970	91.970	X_Ave -2 * Sigma			
UCL	107.073	107.073	107.073	X_Ave +2 * Sigma			

Analyzing the above data yields the following X Bar and S charts.





Theory

Considering the above charts, the X-Bar chart is severely out-of-control. In interpreting these charts, one should realize that the control limits in the X-Bar chart were calculated on the basis of the standard deviations within the subgroup on the basis of the within-sample variability. The subsequent subgroups, however, are not only subject to within-sample variability, but also to between sample to sample and batch to batch variability.

What are the sources of variation in a product response? There are three major sources of variation in a process.

1. Measurement variability
2. Sampling variability (within batch)
3. Batch to batch variability

Each of the three sources of variability adds variation or error to the results. The sample pulled from the batch and tested yields to:

$$E = X - X_{bar} \text{ (Average)}$$

Where

E	is the error,
X	is the measurement and
X_{bar} (Average)	is the population average.

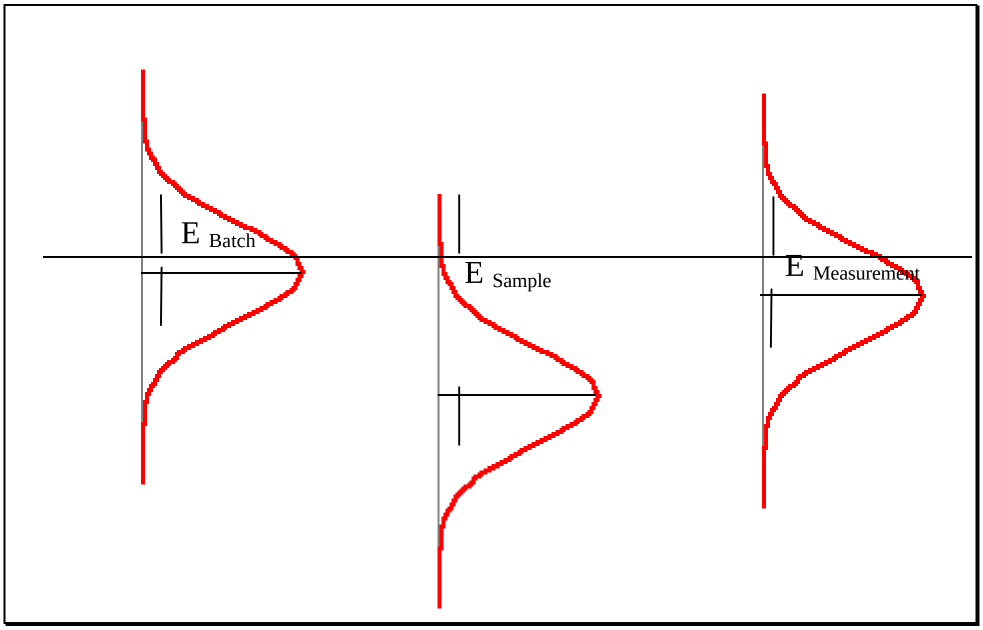
The overall error has three components corresponding to three sources of variations:-

$$E = E_{Batch} + E_{Sample} + E_{Measurement}$$

Where

E_{Batch}	is the error caused by batch to batch variability,
E_{Sample}	is the error caused by batch to sampling variability and
$E_{Measurement}$	is the error caused by batch to measurement variability.

By definition, these errors have zero averages and can be assumed to be represented as normal distributions with fixed variances.



The standard deviation, σ , calculated from the data represents an estimate of the combination of these error components.

$$\sigma^2 = \sigma_{Batch}^2 + \sigma_{Sample}^2 + \sigma_{Measurement}^2$$

where

σ^2 is the total estimated variable

σ_{Batch}^2 is the batch-to-batch estimated variance

σ_{Sample}^2 is the sample estimated variance

$\sigma_{Measurement}^2$ is the measurement variance

Calculation

The following variances are calculated:-

Measurement Variance

The estimated variance of measurement, $\sigma_{Measurement}^2$, is calculated as the average of all measurement variances.

$(\sigma_{Measurement}^2)_J$ = Variance ($X_{I,J}$ I=1,2) for all J=1,2,…,15

$(\sigma_{Measurement}^2)$ = Average ($(\sigma_{Measurement}^2)_J$, J=1,2,…15)

Sample Variance

If Variance of the samples equals the average of the variances of the sample averages, then

$Vs = (\sigma_{Sample}^2) + (\sigma_{Measurement}^2) / k$

where

k is the number of measurements per sample

$\sigma_{Measurement}^2$ is the estimated variance of measurement

σ_{Sample}^2 is the estimated variance of sample

Batch Variance

The Batch to Batch variance is determined from the batch averages. The following procedure is used to calculate batch to batch variance σ_{Batch}^2

$Vp = (\sigma_{Batch}^2) + (\sigma_{Sample}^2)/n + (\sigma_{Measurement}^2) / (k * n)$

Vp is the variance of the batch averages
k is the number of measurements per sample and
n is the number of samples taken per batch.

Nested Control Chart Example

Based on the statistical definition provided before, we can construct control charts for batch process variability.

Variable	Description	Value
k	Number of Measurements per sample (Sub-group size)	2
n	Number of Samples per Batch	2
Nc	Total Number of Batches (Calculation buffer size)	15

The SQC for Excel™ automatically generates the following calculations:-

Step 1: Develop S charts based on all measurements.
Subgroup size = Ns = 2 ➔ k
Number of subgroups = I*J = 2 * 15= 30

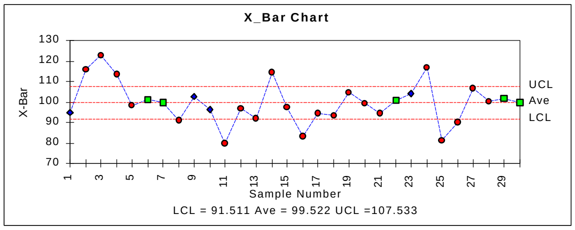
nested_1.xls

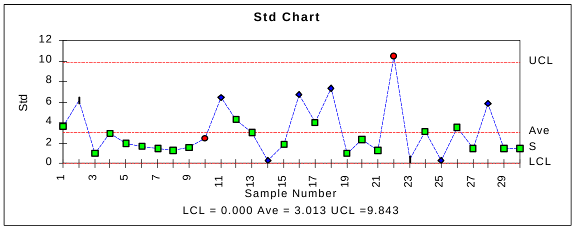
	A	B	C	D
1				
2				
3	Product	Sample	Test1	Test2
4	1	1	92.32	97.38
5	1	2	120.12	111.36
6	2	1	123.21	121.86
7	2	2	111.58	115.66
8	3	1	99.43	96.73
9	3	2	102.26	99.9
10	4	1	100.73	98.67
11	4	2	89.9	91.71
12	5	1	101.43	103.61
28	13	1	81.22	81.57
29	13	2	92.68	87.72
30	14	1	105.31	107.38
31	14	2	95.86	104.17
32	15	1	100.81	102.88
33	15	2	101.01	98.91
34				

Sheet1 Sheet2 Sh

	A	B	All		
First SubGroup	1.000	1.000	1.000	Limits are calculated using zone	AB
Last SubGroup	30.000	30.000	30.000		
Std-					
Average	3.013	3.013	3.013		
LCL	0.000	0.000	0.000	Std_Ave * [0.000 B3]	
UCL	9.843	9.843	9.843	Std_Ave * [3.267 B4]	
X bar-					
Average	99.522	99.522	99.522		
LCL	91.511	91.511	91.511	X_BaStd_Ave - Std_Ave * [2.659 A3]	
UCL	107.533	107.533	107.533	X_BaStd_Ave + Std_Ave * [2.659 A3]	
X Average	99.522	99.522	99.522		
Sigma	3.776	3.776	3.776	Std_Ave / [0.798 C4]	
LCL	91.970	91.970	91.970	X_Ave -2 * Sigma	
UCL	107.073	107.073	107.073	X_Ave +2 * Sigma	

Step 2: Construct nested S charts based on the averages of each batch. This is similar to having the X-Bar and S charts constructed for each batch average.





Step 3: Calculate the Sample Average and Variance for each sample (2 measurements per sample)

No.	X[1]	X[2]	X-bar	Var
1	92.320	97.380	94.850	12.802
2	120.120	111.360	115.740	38.369
3	123.210	121.860	122.535	0.911
4	111.580	115.660	113.620	8.323
5	99.430	96.730	98.080	3.645
6	102.260	99.900	101.080	2.785
7	100.730	98.670	99.700	2.122
8	89.900	91.710	90.805	1.638
9	101.430	103.610	102.520	2.376
10	94.470	97.960	96.215	6.090

Step 4: Construct a new table based on Samples per Batch

Label	S[1]	S[2]
1.000	94.850	115.740
2.000	122.535	113.620
3.000	98.080	101.080
4.000	99.700	90.805
5.000	102.520	96.215
6.000	79.960	96.690
7.000	92.185	114.210
8.000	97.285	83.120
9.000	94.210	93.560
10.000	104.620	99.030
11.000	94.200	100.690
12.000	103.935	117.050
13.000	81.395	90.200
14.000	106.345	100.015
15.000	101.845	99.960

Note that the S(1) and S(2) for the first record are the averages of the first and second records from the previous table. The table size is ½size of the original table.

Nested Charts									
	A	B	All						
First SubGroup	1.000	1.000	1.000	Limits are calculated using zone			AB		
Last SubGroup	15.000	15.000	15.000	K=	2	J=	2		
Var(m)	14.890	14.890	14.890	13%					
Var(s)	58.432	58.432	58.432	53%	Var(s) = Vs -Var(m)/K				
Var(b)	37.563	37.563	37.563	34%	Var(b) = Vp - Var(s)/K + Var(s)/J*K				
Var(t)	110.885	110.885	110.885		Var(t) = Var(m) + Var(s) + Var(b)				
Vs	65.878	65.878	65.878						
Vp	70.501	70.501	70.501		Vp = Var(b) + Var(s)/k + Var(s)/j*k				
SQRT(Vs)	8.117	8.117	8.117	Sigma S = SQRT(Vs)					
LCL	0.000	0.000	0.000	Sigma S * [0.000 B3]					
UCL	26.517	26.517	26.517	Sigma S * [3.267 B4]					
SQRT(Vp)	8.397	8.397	8.397	Sigma(X) = SQRT(Vp)					
Average	99.522	99.522	99.522						
LCL	77.195	77.195	77.195	X_Bar_Ave - Sigma(X) * [2.659 A3]					
UCL	121.848	121.848	121.848	X_Bar_Ave + Sigma(X) * [2.659 A3]					

How to Calculate

1- Calculate the **Var(m)** = Average of all Variances in the Table 1.

2- Calculate the Average and Variance for table

Label	S[1]	S[2]	Var	Average
1.000	94.850	115.740	218.196	105.295
2.000	122.535	113.620	39.739	118.078
3.000	98.080	101.080	4.500	99.580
4.000	99.700	90.805	39.561	95.253
5.000	102.520	96.215	19.877	99.368
6.000	79.960	96.690	139.946	88.325
7.000	92.185	114.210	242.550	103.198
8.000	97.285	83.120	100.324	90.203
9.000	94.210	93.560	0.211	93.885
10.000	104.620	99.030	15.624	101.825
11.000	94.200	100.690	21.060	97.445
12.000	103.935	117.050	86.002	110.493
13.000	81.395	90.200	38.764	85.798
14.000	106.345	100.015	20.034	103.180
15.000	101.845	99.960	1.777	100.903

3- Calculate the **Vs** or the Average of all Variances in the Table 2.

4- Calculate the **Vp** or the Variance of all Averages in the Table 2.

5- Calculate the **Var(s) = Vs -Var(m)**/K where **k** is 2

6- Calculate the **Var(b) = Vp - Var(s)/K + Var(s)/J*K** Where K and J are 2 and 2 respectively.

7- Calculate **Var(t) = Var(m) + Var(s) + Var(b**) and the percentage of each Var out of total variance **Var(t)**

8- Calculate the Sigma for control charts as **Sigma= SQRT(Vs)**

9- Calculate the S Chart Lower Control Limit **LCL = Sigma* [0.000 B3]**

10- Calculate the S Chart Upper Control Limit **LCL = Sigma* [3.267 B4]**

11- Calculate the overall average as Average of all Averages in the Table 2.

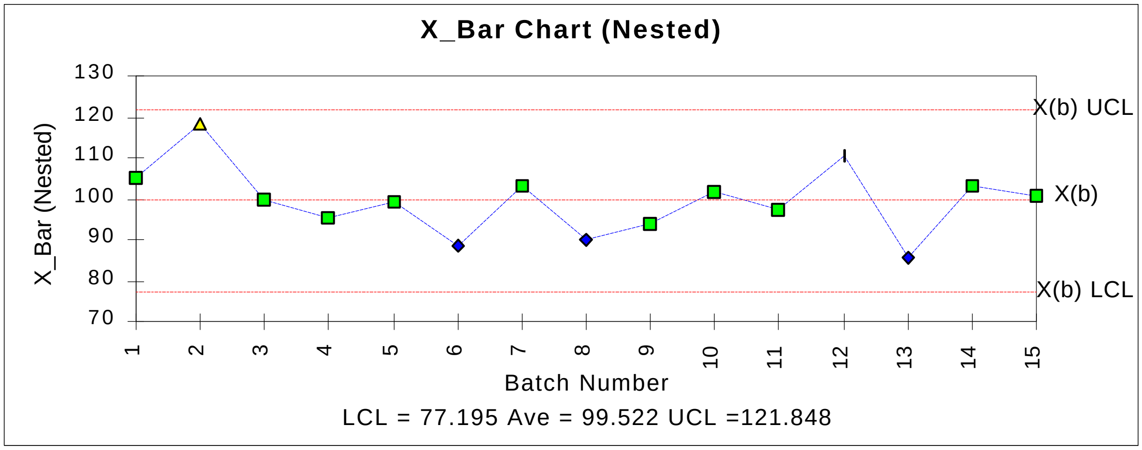
12- Calculate the Nested Chart Upper Control Limit

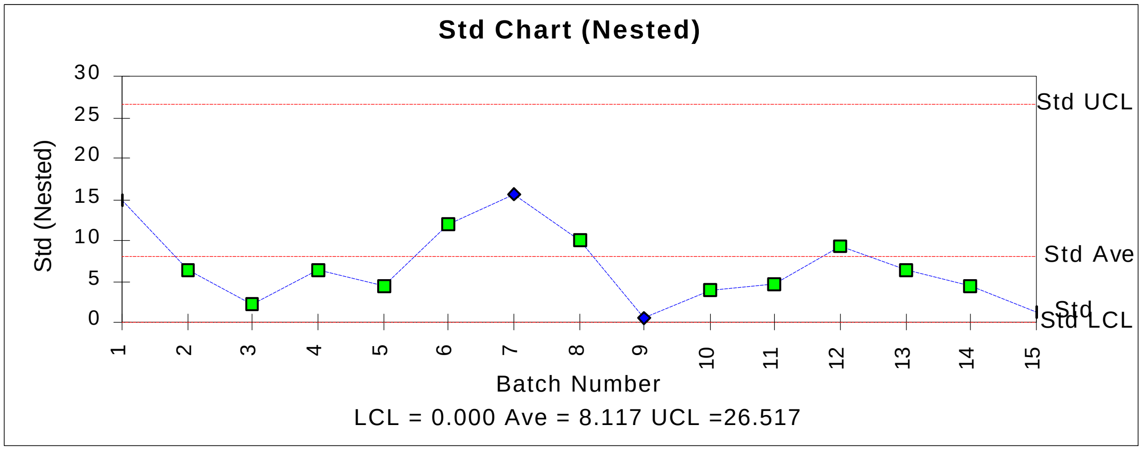
UCL = X_Bar_Ave - Sigma(X) * [2.659 A3]

13- Calculate the Nested Chart Lower Control Limit

LCL = X_Bar_Ave + Sigma(X) * [2.659 A3]

14- Plot the charts.





How to Analyze Nested Control Charts

1. Establish control limits for the nested control charts.

2. Use the first S chart to monitor within the sample variability. Out of control situations have to be interpreted as uniformity problems. Note do not analyze X-Bar charts.

3. Use the nested S chart to monitor the posted standard deviation **Var(b**). When out of control, first check the S chart. If this shows no out of control, then there is a between sample problem.

4. Finally, use individual Nested Control charts to monitor the **Var(b)** as well as the trend of the run averages. In the case of out of control, check within the sample and between the sample charts for out of control. IF the latter is OK, then a batch problem exists.

The material in this chapter is based on the following references:

- Rene Penning De Vnes, "Control Charts in IC Manufacturing"

- W.H McNeese and R.A. Klein, "Statistical Methods for the Process Industries", ASQC Quality Press, 1991

How to Setup Nested Charts.

The nested charts are calculated for you automatically with the Premium version of **SQC for Excel** ™.

Please do the following steps:-

1- Prepare Original Data

The data should be prepared as Batch, Sample and Measurement. For example in the previous section, the data was organized as:

- 2 Measurements per Sample
- 2 Samples per Batch
- 15 Batches

However it is possible to have access to data with more samples or sub-groups per Batch. The following example provides a example of how to use MS Excel to collect the data.

Assume that you have collected
- 10 measurements per Sample
- 10 samples per Batch
- 5 Batches

However you are interested in analyzing only 5 Sample per batch instead of 10. The following steps will provide you with a detailed procedure to do this:

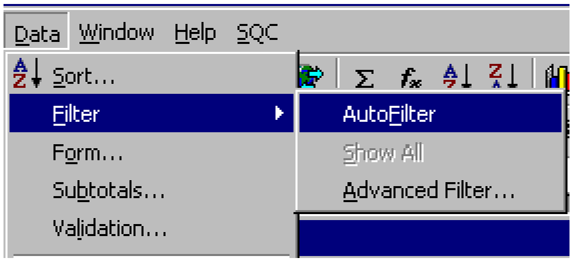
Step 1: Open the spreadsheet with data.

data.xls

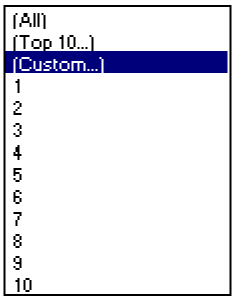
	L	M	N	O	P	Q	R	S	T	U	V	W
1	Batch	Sample Num.	Tag1	Tag2	Tag3	Tag4	Tag5	Tag6	Tag7	Tag8	Tag9	Tag10
2	1	1	7065	7135	7055	7065	7120	7030	7150	7045	7155	7065
3	1	2	7035	7015	7045	6990	6905	6885	6890	6910	6940	6825
4	1	3	6835	6875	6990	6905	6895	6985	6915	6910	6925	6890
5	1	4	7000	6970	6975	6865	6940	7045	7035	7145	7030	7045
6	1	5	7035	7035	6930	6990	7015	6940	6955	6885	6975	7065
7	1	6	6995	6835	6960	6985	6885	7015	7080	7135	7135	7145
8	1	7	7080	7110	6965	7125	7020	7050	7020	6990	7050	6985
9	1	8	7055	7050	7165	7070	7100	6920	7060	7170	7175	7150
10	1	9	7005	7080	7005	7015	7005	6895	6895	6980	6955	7010
11	1	10	7030	7160	6970	6995	7250	7035	7000	6980	6945	6940
12	2	1	6985	6970	6985	6970	7055	7055	7110	6900	6975	7070
13	2	2	6960	6940	6950	6930	6930	6970	6935	6955	6925	6840
14	2	3	7045	7190	7025	7010	7085	7195	7125	7090	7080	7005
50	5	9	7030	7115	7055	6980	6920	6965	7030	6985	7000	7025
51	5	10	7025	7130	7045	7125	7025	7015	7055	7075	7095	7115

A

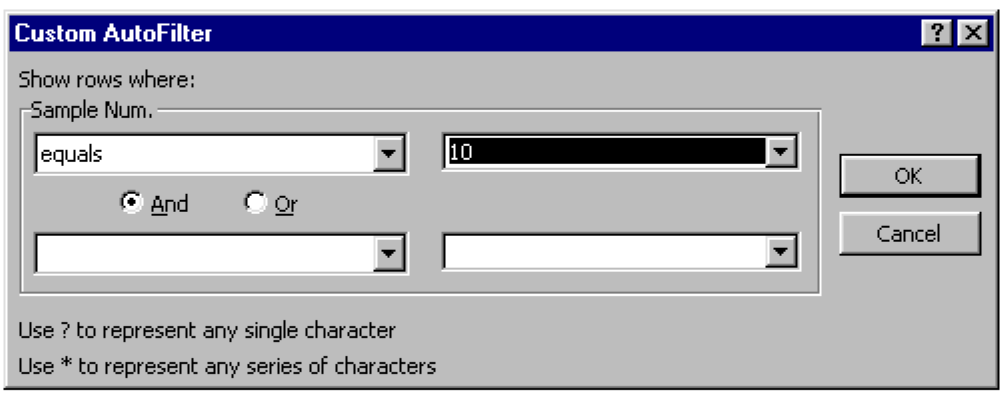
Step 2: From the **Data** menu, choose **Filter** and then **AutoFilter**

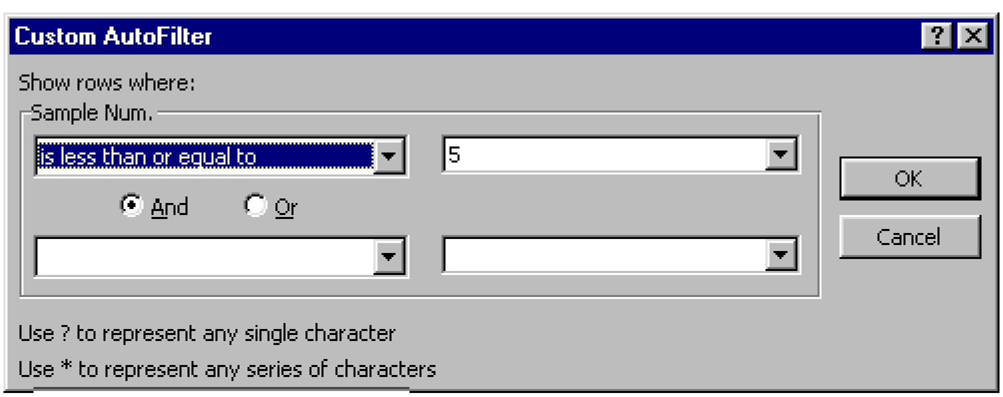


Step 3: In the cell labeled Sample Number, click on the down arrow and then select **Custom** option.



Step 4: Click on the selection arrow and choose the "Less than and Equal to" option.





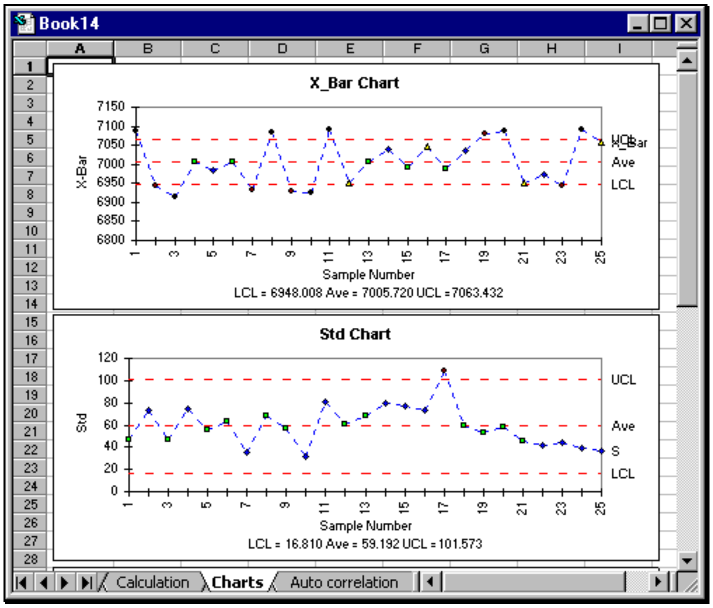
Step 5: all data with the Sample Number of less than or equal 5 are selected.

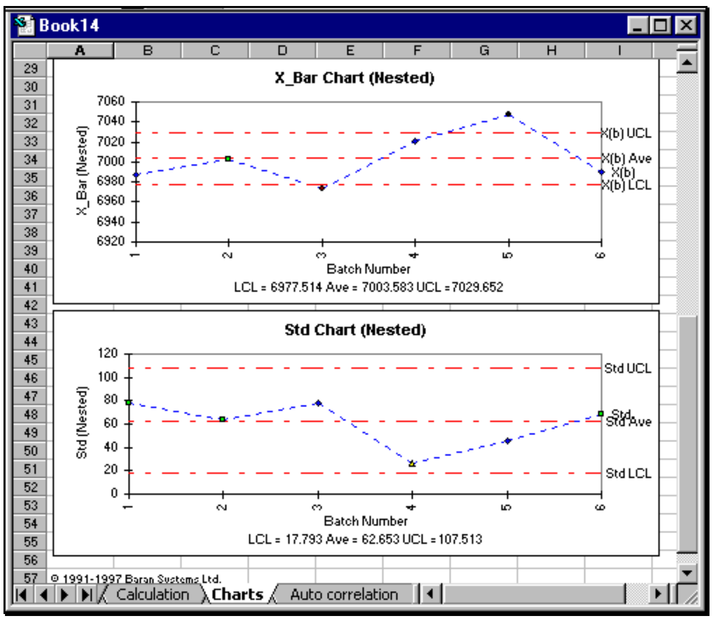
	L	M	N	O	P	Q	R	S	T	U	V	V
1	Batc	Sample Num.	Tag1	Tag2	Tag3	Tag4	Tag5	Tag6	Tag7	Tag8	Tag9	Tac
2	1	1	7065	7135	7055	7065	7120	7030	7150	7045	7155	7
3	1	2	7035	7015	7045	6990	6905	6885	6890	6910	6940	6
4	1	3	6835	6875	6990	6905	6895	6985	6915	6910	6925	6
5	1	4	7000	6970	6975	6865	6940	7045	7035	7145	7030	7
6	1	5	7035	7035	6930	6990	7015	6940	6955	6885	6975	7
12	2	1	6985	6970	6985	6970	7055	7055	7110	6900	6975	7
13	2	2	6960	6940	6950	6930	6930	6970	6935	6955	6925	6
14	2	3	7045	7190	7025	7010	7085	7195	7125	7090	7080	7
15	2	4	6805	7000	6890	6955	6905	6965	6980	6965	6915	6
16	2	5	6955	6885	6935	6975	6930	6895	6915	6940	6930	6
22	3	1	7145	6990	7015	7110	6970	7030	7165	7135	7135	7
23	3	2	6980	6900	6890	6825	6935	7005	6980	7010	6990	6
24	3	3	7145	6985	6985	6915	6955	7010	7090	7045	6965	6
42	5	1	6960	6880	6880	6940	6965	6940	6960	7015	6960	7
43	5	2	6965	7000	6980	7005	6920	6975	7005	7010	6885	7
44	5	3	6885	6975	6975	6940	6915	6965	6970	6905	7020	6
45	5	4	7015	7125	7135	7095	7040	7125	7105	7060	7095	7
46	5	5	7030	7045	7085	7080	7045	7080	7035	6990	7060	7

Step 6: Highlight the data and choose **New** from **SQC** Menu.

	O	P	Q	R	S	T	U	V	W	X	Y
1	Tag2	Tag3	Tag4	Tag5	Tag6	Tag7	Tag8	Tag9	Tag1		
2	7135	7055	7065	7120	7030	7150	7045	7155	7065		
3	7015	7045	6990	6905	6885	6890	6910	6940	6825		
4	6875	6990	6905	6895	6985	6915	6910	6925	6890		
5	6970	6975	6865	6940	7045	7035	7145	7030	7045		
6	7035	6930	6990	7015	6940	6955	6885	6975	7065		
12	6970	6985	6970	7055	7055	7110	6900	6975	7070		
13	6940	6950	6930	6930	6970	6935	6955	6925	6840		
14	7190	7025	7010	7085	7195	7125	7090	7080	7005		
15	7000	6890	6955	6905	6965	6980	6965	6915	6925		
16	6885	6935	6975	6930	6895	6915	6940	6930	6875		
22	6990	7015	7110	6970	7030	7165	7135	7135	7200		
23	6900	6890	6825	6935	7005	6980	7010	6990	6985		
24	6985	6985	6915	6955	7010	7090	7045	6965	6970		
42	6880	6880	6940	6965	6940	6960	7015	6960	7015		
43	7000	6980	7005	6920	6975	7005	7010	6885	7005		
44	6975	6975	6940	6915	6965	6970	6905	7020	6890		
45	7125	7135	7095	7040	7125	7105	7060	7095	7110		
46	7045	7085	7080	7045	7080	7035	6990	7060	7120		

Step 7: Check the results. Make sure you select NESTED CHART


Book14
X_Bar Chart
X-Bar
Sample Number
LCL = 6948.008 Ave = 7005.720 UCL = 7063.432
Std Chart
Std
Sample Number
LCL = 16.810 Ave = 59.192 UCL = 101.573
Calculation
Charts
Auto correlation


Book14
X_Bar Chart (Nested)
X_Bar (Nested)
Batch Number
LCL = 6977.514 Ave = 7003.583 UCL = 7029.652
Std Chart (Nested)
Std (Nested)
Batch Number
LCL = 17.793 Ave = 62.653 UCL = 107.513
© 1991-1997 Baran Systems Ltd.
Calculation
Charts
Auto correlation

PCA - Partial Component Analysis

Introduction

Statistical process control, SPC, is being implemented across the industry without consideration to the multivariable nature of the measurements. The objective of this chapter is to provide basic statistical theory and application of the Principal Component Analysis (PCA) technique to evaluate and understand process unit data. The theory is presented from the user's point of view and more detail of mathematical derivations can be found in the references. The technology is mainly being introduced to the chemical industry by John MacGregor of McMaster University.

The statistical process control, SPC, and statistical quality control, SQC, foundations are based upon the detection of a statistically significant change in a given process. When a process can be described by a single measurement, such as making an object with a desired length, implementation of these techniques becomes an easy task.

However in the majority of the industries, a given process such as a reactor can not be described completely by a single measurement. Implementation of SQC and SPC technology will yield hundreds of control charts which will identify some causes but not the status of the process.

The multivariable nature of processes makes it difficult to provide the unit operator with the information with respect to the health of the process based on SQC and SPC technology due to the number of charts generated.

The PCA technique presented in this chapter will reduce a large multivariable system (i.e. 10 tags with 200 measurements) to a manageable system (i.e. 2 tags with 200 measurement) while preserving most of the information in the original system.

Theory

Principal Component Analysis, PCA, is a mathematical technique, which examines a multivariable data set and determines the underlying associative patterns based on variance. Process computers routinely collect thousands and thousands of data from multitude of plant sensors every second. This data overload will cause the operators to ignore the majority of the received information and concentrated on few measurements. However statistical method such as PCA are capable of compressing the information down to a manageable size by retaining most of the information.

Data processing and scaling

The multivariable data gathered either online or from historical data bases to represent a process usually have different mean and standard deviations. This will yield some inconsistency in statistical evaluation of the data.

The implementation of PCA is achieved by

1. Data processing and scaling
2. Calculation of Similarity matrix
3. Calculation of Eigen vector and Eigen matrix
4. Sorting of Eigen vector
5. Loading Factors
6. Score Factors

which are described in the following sections.

Example

Assume a simple database of 5 variables (m=5) and 12 observations (n=12). The objective is to condense the data to a single variable (k=3) with same number of observations (n=12). The variables are:

1.Population
2.School
3.Employment
4.Professional Services
5.House

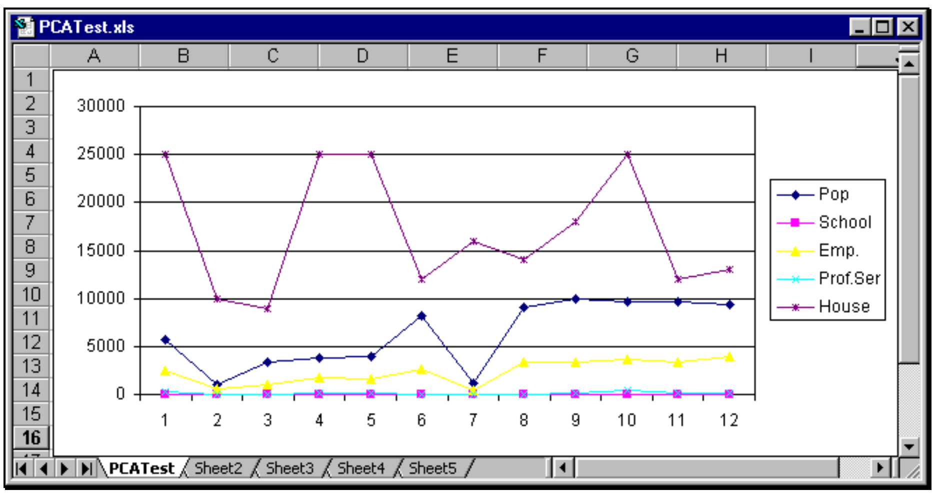
with the following observations:

Pop	School	Emp.	Prof.Ser	House
5700	12.8	2500	270	25000
1000	10.9	600	10	10000
3400	8.8	1000	10	9000
3800	13.6	1700	140	25000
4000	12.8	1600	140	25000
8200	8.3	2600	60	12000
1200	11.4	400	10	16000
9100	11.5	3300	60	14000
9900	12.5	3400	180	18000
9600	13.7	3600	390	25000
9600	9.6	3300	80	12000
9400	11.4	4000	100	13000

Data processing and scaling

The multivariable data gathered either online or from historical databases represent a process usually having different means and standard deviations. This will yield some inconsistency in statistical evaluation of the data.

Assume a multivariable data set consists of **m** data points (i.e. 5 classes of measurement) and **n** measurements (i.e. 12 in the above example).

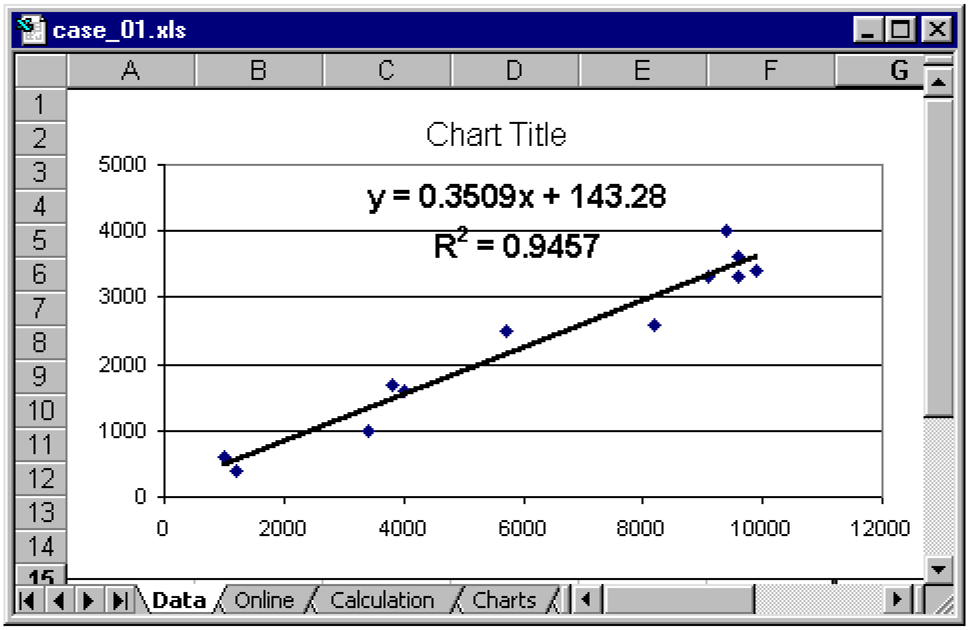


However, if one calculates the (Data - Mean)/(Standard Deviation) as the normalized measurement, then each class of data has the same effect on the overall analysis of the data.

Similarity matrix calculation.

The degree of relationship existing between three or more variables is called multiple correlation. The fundamental principles involved in the problems of multiple correlation are analogous to those of single input, single output or simple correlation.

In case of a simple correlation, the relationship between two variables, x and y is investigated. The variables x and y as shown below, seem to lie near a line, hence the correlation is called linear.



The linear relationship between variables x and y can be described as

Y = SLOPE * X + INTERCEPT

after estimation of SLOPE and INTERCEPT using least square regression, the goodness of the regression is defined as

R = SQRT (explained variation / Total variation)

$$R = \text{SQRT}\left(\Sigma(Y_{est} - Y_{ave})^2 / \Sigma(Y - Y_{ave})^2 \right)$$

The quantity, R , called the coefficient of correlation, varies between +1 and -1.

In a multivariable case the coefficients of correlation for the m variables needs to be calculated. This is called the similarity matrix or correlation matrix R.

Defining

$$X_i = SLOPE * X_j + INTERCEPT$$

$$R = SQRT\text{(explained variation / Total variation)}$$

$$R = SQRT(\Sigma(X_{i\ est} - X_{i\ ave})^2 / \Sigma(X_i - X_{i\ ave})^2)$$

and

$$R = \begin{matrix} R_{1,1}, & R_{1,2}, & R_{1,3}, & \dots, & R_{1,m} \\ R_{2,1}, & R_{2,2}, & R_{2,3}, & \dots, & R_{2,m} \\ R_{3,1}, & R_{3,2}, & R_{3,3}, & \dots, & R_{3,m} \\ \dots, & \dots, & \dots, & \dots, & \dots \\ R_{m,1}, & R_{m,2}, & R_{m,3}, & \dots, & R_{m,m} \end{matrix}$$

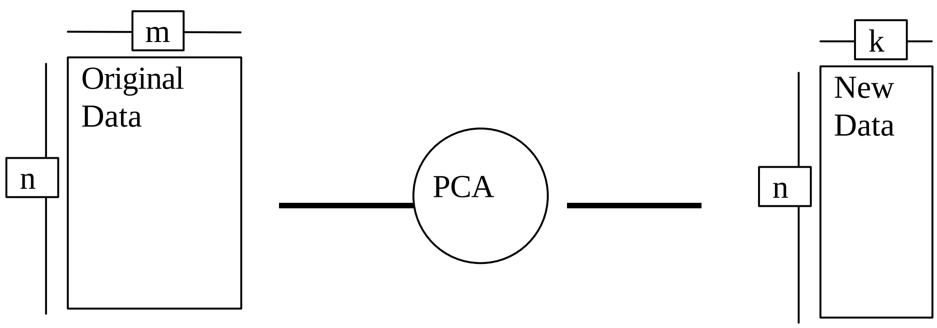
and in the case of the multivariable system (m=5) R can be represented as:

	Pop	School	Emp.	Prof.Ser	House
Pop	1.000	0.010	0.972	0.439	0.022
School	0.010	1.000	0.154	0.691	0.863
Emp.	0.972	0.154	1.000	0.515	0.122
Prof.Ser	0.439	0.691	0.515	1.000	0.778
House	0.022	0.863	0.122	0.778	1.000

Calculation of Eigen vector and Eigen matrix

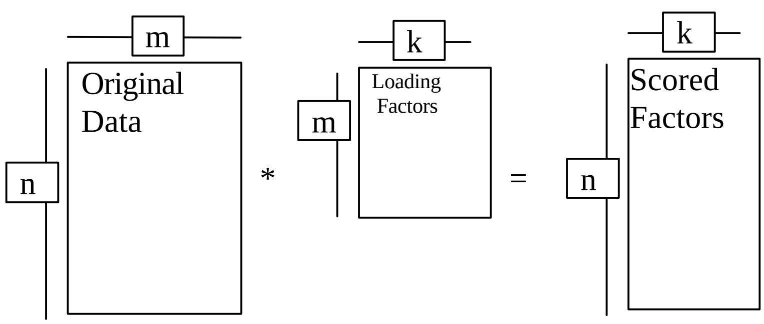
The correlation matrix represents the correlation of the variables with respect to each other. The PCA technique is based on using the correlation matrix to calculate a new set of linear coefficients, called **Loading Factors**.

These coefficients will be used in condensing the original data set of **m** variables by **n** observations to a manageable **k** variables by **n** observations.



Where **k** is usually much less than **m**.

Hence PCA calculates a new set of **k** variables, called **Scored Factors**, based on the original data.



There are many different algorithms to calculate the **Factor Loading** matrix. The algorithm used here to calculate the **Factor Loading** matrix uses the Eigen vector, **e** and Eigen value matrix, **E**, of the correlation matrix, **R** calculated by the Jaccobian decomposition method.

In other words by calculating the Eigen vector of the correlation matrix one will evaluate the importance or the weighting of each variable.

For example assume an Eigen vector calculated and sorted for a data set with 5 measurements is:

	Value 1	Value 2	Value 3	Value 4	Value 5
Eigen Value	2.873	1.797	0.215	0.100	0.015
Difference	1.077	1.582	0.115	0.085	

The amount of information on each Eigen vector element can be evaluated by calculation of the Percentage of Information or Proportion,

Proportion = $\mathbf{e}_i \; / \; \Sigma{:}\mathbf{e}_i$ for i = l,...,5

which is

	Value 1	Value 2	Value 3	Value 4	Value 5
EigenValue	2.873	1.797	0.215	0.100	0.015
Difference	1.077	1.582	0.115	0.085	
Proportion	0.575	0.359	0.043	0.020	0.003
Cumulative	0.575	0.934	0.977	0.997	1.000

This means the first measurement can represent about 57.5% of the information and the combination of the first and second measurement will represent 93.4% (or 57.5+35.9%) of the data.

In the above example the order of significance, relative weighting of the Eigen Vector, and the data order is the same (the first Eigen value is the largest one and so on).

However one needs to sort the Eigen vector in descending order of significance. Assume a data based with 5 measurements and the following Eigen vector:

	1	2	3	4	5
Value	1.422	1.1	0.678	0.3	1.5
Proportion	28.4%	22.0%	13.6%	6.0%	30.0%

Sorting of Eigen vector

The order of the Eigen Vector is 5,1,2,3 and 4, therefore the 5th element is the most significant one.

By sorting the Eigen vector in a descending order

	5	1	2	3	4
Value	1.5	1.422	1.1	0.678	0.3
Proportion	30.0%	28.4%	22.0%	13.6%	6.0%
	30.0%	58.4%	80.4%	94.0%	100.0%

Hence by using only 5, 1 and 2 one can utilize 80.44% of information within the original data set and the system size has been reduced form 5 to 3.

Loading Factors

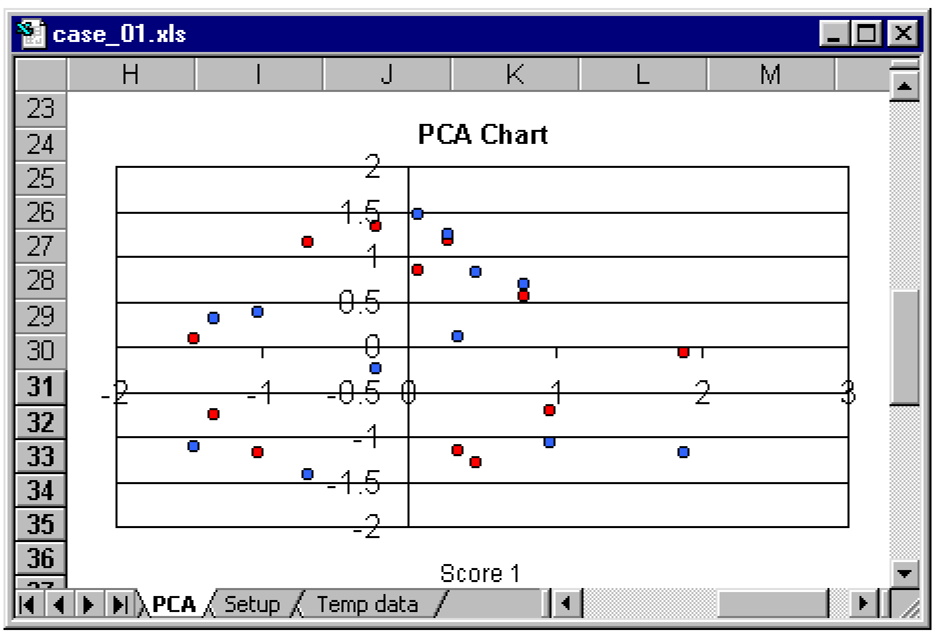
The program calculates the following Loading Factor Matrix

	Factor 1	Factor 2	Factor 3
Pop	0.581	0.806	0.028
School	0.767	-0.545	0.319
Emp.	0.672	0.726	0.115
Prof.Ser	0.932	-0.104	-0.308
House	0.791	-0.558	-0.065

Score Factors

The program calculates the following Score Factors

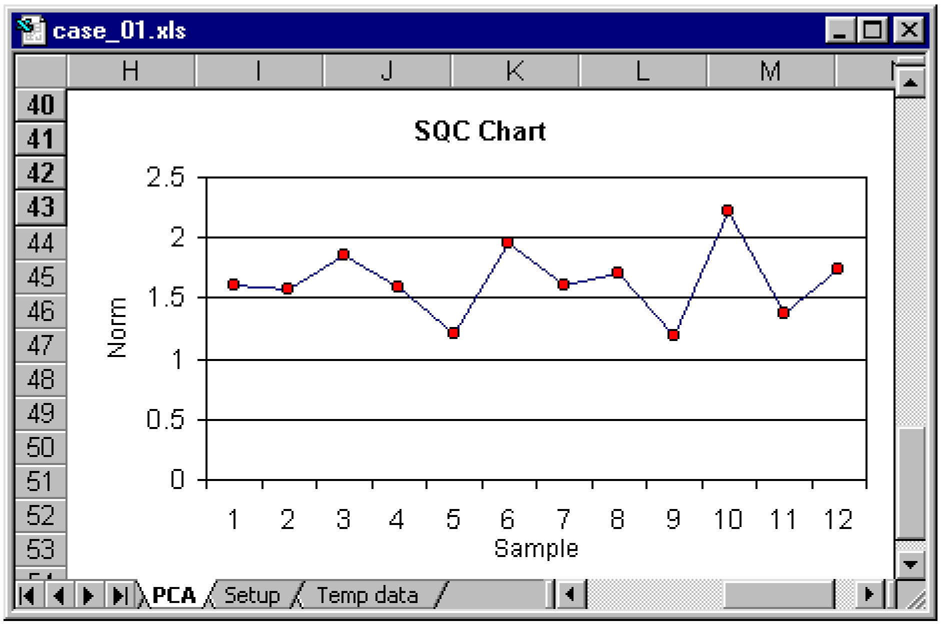
Score 1	Score 2	Score 3
0.970	-0.713	-1.057
-1.331	-0.759	0.320
-1.472	0.090	-1.118
0.460	-1.291	0.814
0.333	-1.162	0.113
-0.692	1.150	-1.430
-1.023	-1.175	0.373
0.057	0.855	1.472
0.785	0.566	0.692
1.880	-0.059	-1.184
-0.227	1.330	-0.244
0.262	1.167	1.251



This chart represents the Score2 and Score3 plotted against Score1.
Please observe that the data is normalized with the mean centering and standard deviation scaling. Hence any point outside +- 3 sigma is out of control.

M-SQC

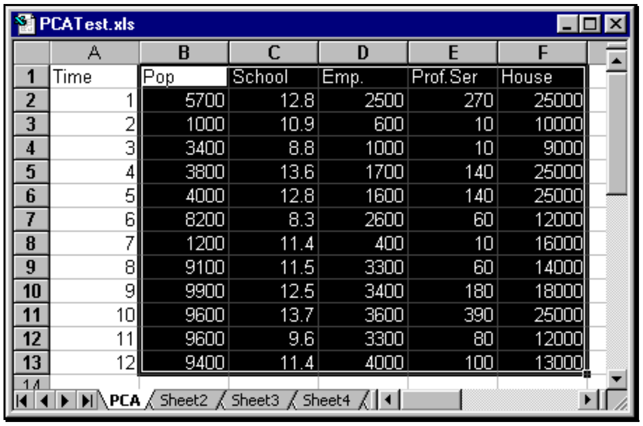
Based on the Scored data, the Norm of the data (i.e the dissonance from the center) is plotted as individual X-chart.

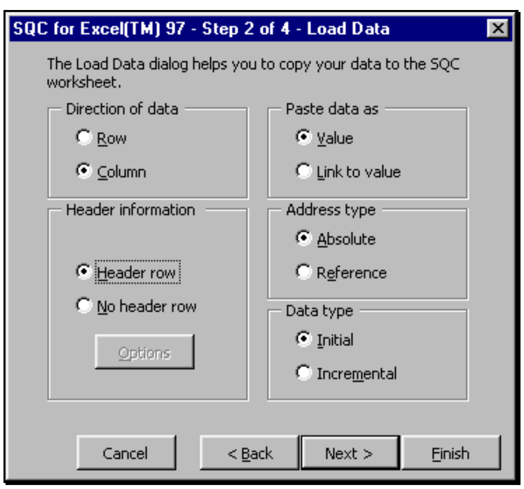


x = SQRT (Σ (S_i)^2 for i = 1,2 and 3)

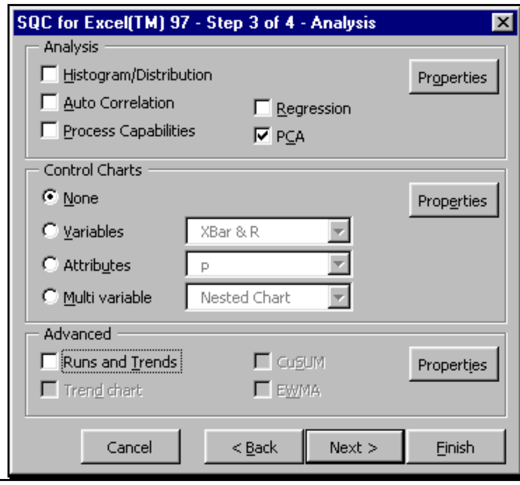
How to Use PCA Program

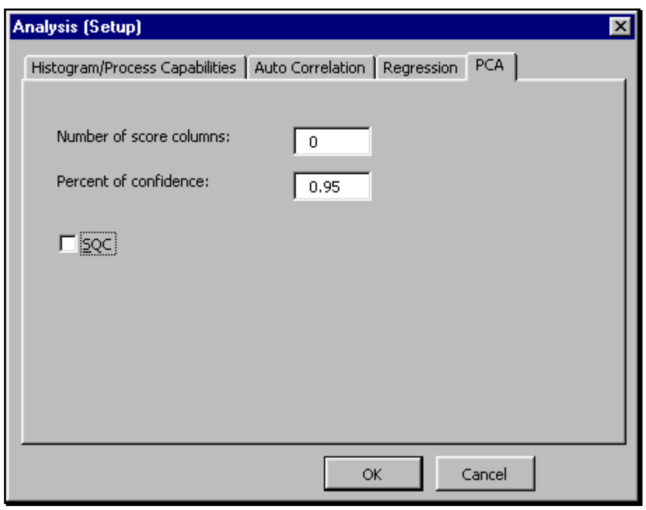
Do the following steps to develop the PCA charts.

<table>
<tr><th>Step</th><th>Dialog</th></tr>
<tr><td>Select Data</td><td>


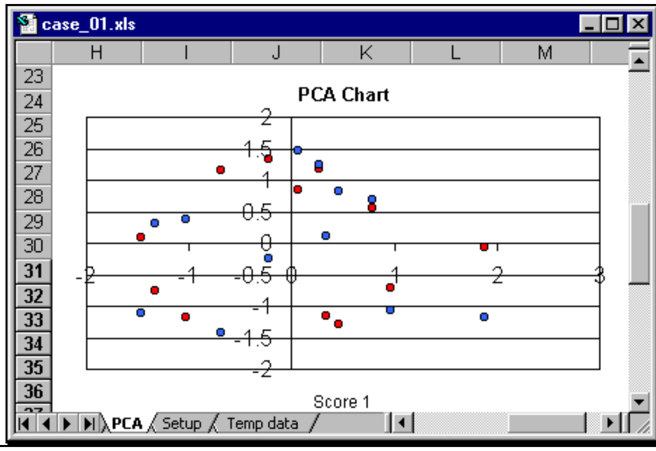
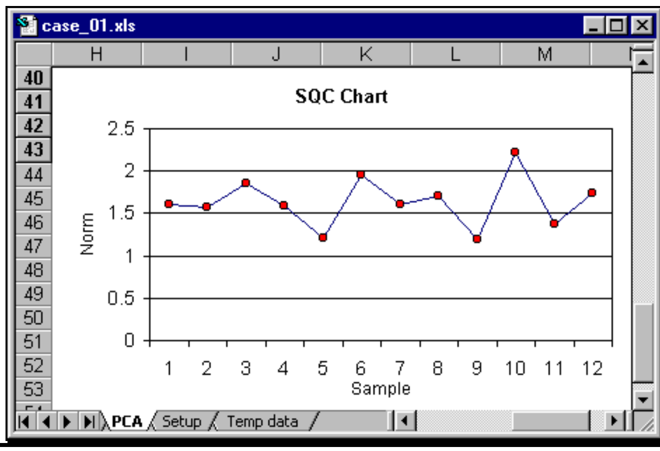
Choose PCA worksheet and select data from B1 to F13 (See Quick Tour chapter for more detail).</td></tr>
<tr><td>Start SQC</td><td>From SQC menu, choose New option</td></tr>
<tr><td>Select Type of Data</td><td>Make sure you click on Header row if you have selected the header in your data (i.e. Row 1)


</td></tr>
</table>

<table>
<tr>
<td>Select PCA</td>
<td>Make sure you ONLY select PCA to speed up the operation


</td>
</tr>
<tr>
<td>PCA Parameters</td>
<td>You can select 3 parameters for PCA
1- Number of Score Factors (i.e for the above example)
2- % Accumulation
3- SQC Charts


</td>
</tr>
</table>

Results	Check the PCA chart 
	

References

The following is a list of references that provide more information on concepts and terminology used in this paper.

1- A. Hoskuldsson. PLS regression methods. Journal of Chemometrics, 2:211-22S, l988.

2- Arthur Jutan, J. F. MacGregor, and J. D. Wright. Multivariate computer control of a butane hydrogenolysis reactor. part ii: Data collection, parameter estimation, and stochastic disturbance identification. AIChE Journal, 23(5):742-750, September 1977.

3- J. C. Liao. Fermentation data analysis and state estimation in the presence of incomplete mass balance. Biotechnology and Bioengineering, 33:613-622, l989.

4- John F. MacGregor, Thomas E. Marlin, James Kresta, and Bert Skagerberg. Multivariate statistical methods in process analysis and control. In CPC-IV, 1991.

5- Rolf Manne. Analysis of two partial-least-squares algorithms for multivariate calibration. Chemometrics and Intelligent Laboratory Systems, 2:187-197, l987.

6- G. Musumarra, G. Scarlata, G. Romano, and S. Clementi. Identification of drugs by principal component analysis of rf data obtained by TLC in different effluent systems. Journal of Analytical Toxicology, 7:286-292, l983.

7- J. J. Roffel, J. F. MacGregor, and T. W. Hoffman. The design and implementation of a multivariable controller for a continuous polybutadiene polymerization train. In Proceedings of the 2nd IFAC Symposium on Dynamics and Control of Chemical Reactors, Distillation Columns, and Batch Processes. Pergamon Press, l989.

I

N

O

P

Q

R