國立雲林科技大學
97 學年度博士班招生考試試題

所別：工業工程與管理研究所
科目：統計學

1. (10%) If U is a uniform random variable in the interval between 0 and 1, what is the distribution of $1-U$? Prove your results.

2. (10%) A particle randomly moves in a horizontal line. In each move, it has probability 1/3 to move right and probability 2/3 to move left. The particle is initially at coordinate (0,0). What is the probability that the particle will move to (2,0) before it moves to (-3, 0)?

3. (10%) Suppose that X and Y have the following joint p.d.f.: $f(x,y)=c(x+y)$ for $0<x<y<1$. Determine the constant c and the conditional p.d.f. of Y given that $X=x$.

4. (10%) Random variable X has the following probability function: $f(x)=q^{x-1}p, x=1,2,3,\ldots,$ where $p+q=1$.

(a) Determine the moment generating function of X.

(b) If X_1, X_2, X_3 are 3 independent random variable with the p.d.f $f(x)$, what is the distribution of $X_1+X_2+X_3$?

5. (60%) Read the following case and explain the need for care in selecting rational subgroups and suggest procedures to evaluate different variability: within-wafer, between-wafer, and between-lot variation.

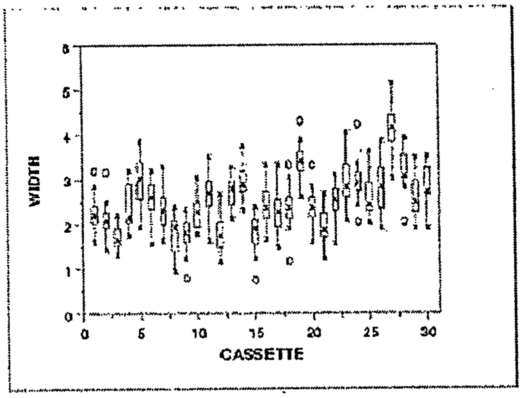
Case Study for SPC in Batch Processing Environment

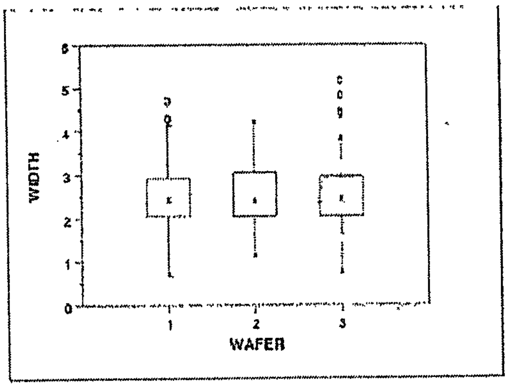
One of the assumptions in using classical Shewhart SPC charts is that the only source of variation is from part to part (or within subgroup variation). This is the case for most continuous processing situations. However, many of today's processing situations have different sources of variation. The semiconductor industry is one of the areas where the processing creates multiple sources of variation.

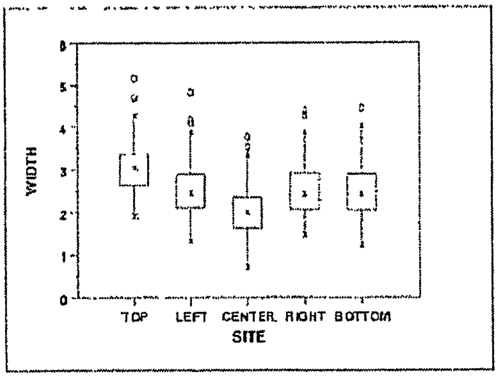
In semiconductor processing, the basic experimental unit is a silicon wafer. Operations are performed on the wafer, but individual wafers can be grouped multiple ways. In the diffusion area, up to 150 wafers are processed in one time in a diffusion tube. In the etch area, single wafers are processed individually. In the lithography area, the light exposure is done on sub-areas of the wafer. There are many times during the production of a computer chip where the experimental unit varies and thus there are different sources of variation in this batch processing environment.

tHE following is a case study of a lithography process. Five sites are measured on each wafer, three wafers are measured in a cassette (typically a grouping of 24 - 25 wafers) and thirty cassettes of wafers are used in the study. The width of a line is the measurement under study. There are two line width variables. The first is the original data and the second has been cleaned up somewhat. This case study uses the raw data. The entire data table is 450 rows long with six columns.

Cassette	Wafer	Site	Raw Line Width	Sequence	Cleaned Line Width
1	1	Top	3.199275	1	3.197275
1	1	Lef	2.253081	2	2.249081
1	1	Cen	2.074308	3	2.068308
1	1	Rgt	2.418206	4	2.410206
1	1	Bot	2.393732	5	2.383732
1	2	Top	2.654947	6	2.642947
1	2	Lef	2.003234	7	1.989234
1	2	Cen	1.861268	8	1.845268
1	2	Rgt	2.136102	9	2.118102
1	2	Bot	1.976495	10	1.956495
1	3	Top	2.887053	11	2.865053
1	3	Lef	2.061239	12	2.037239
1	3	Cen	1.625191	13	1.599191
1	3	Rgt	2.304313	14	2.276313
1	3	Bot	2.233187	15	2.203187





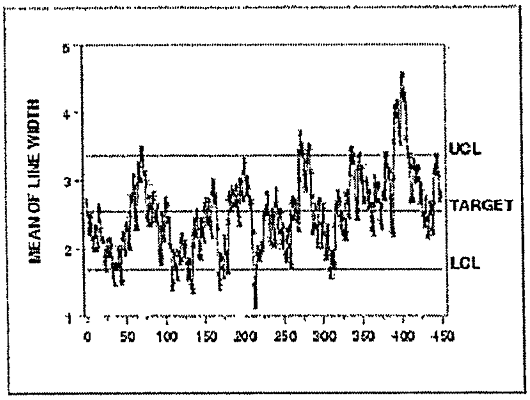


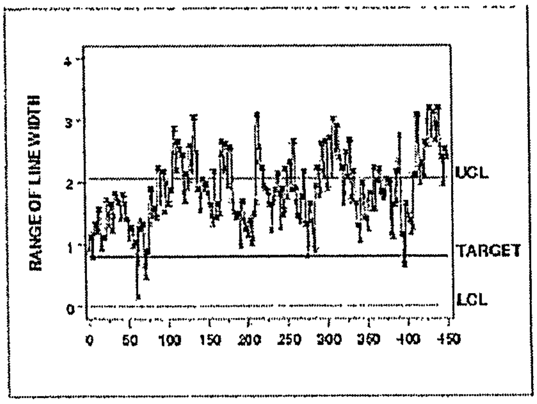
The above graphs show that there are differences between the lots and the sites.

There are various ways we can create subgroups of this dataset: each lot could be a subgroup, each wafer could be a subgroup, or each site measured could be a subgroup (with only one data value in each subgroup).

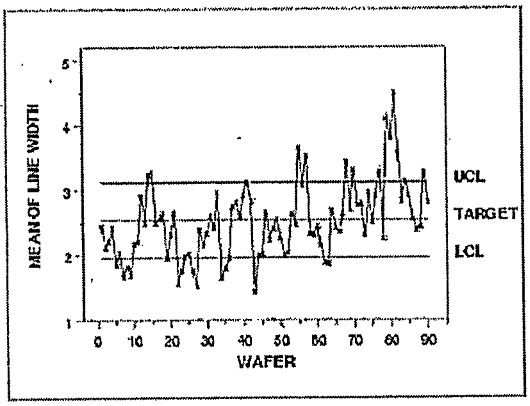
Recall that for a classical Shewhart Means chart, the average within subgroup standard deviation is used to calculate the control limits for the Means chart. However, on the means chart you are monitoring the subgroup mean-to-mean variation. There is no problem if you are in a continuous processing situation - this becomes an issue if you are operating in a batch processing environment.

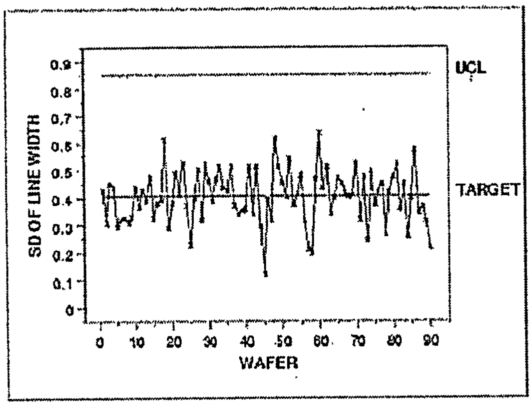
The first pair of control charts use the site as the subgroup. However, since site has a subgroup size of one we use the control charts for individual measurements. A moving average and a moving range chart are shown.



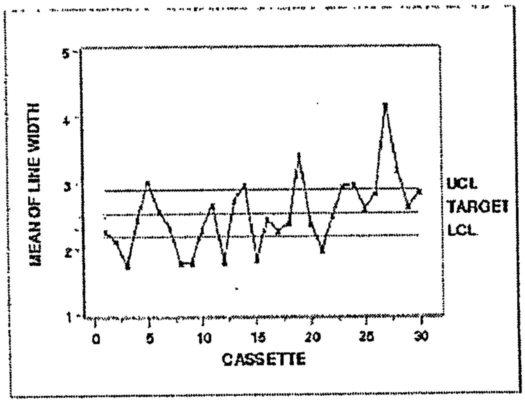


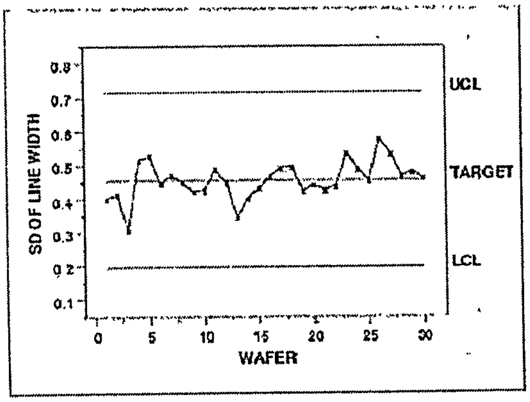
The next pair of control charts use the wafer as the subgroup. In this case, that results in a subgroup size of 5. A mean and a standard deviation control chart are shown.





The next pair of control charts use the cassette as the subgroup. In this case, that results in a subgroup size of 15. A mean and a standard deviation control chart are shown.





The largest source of variation in this data is the lot-to-lot variation. So, using classical Shewhart methods, if we specify our subgroup to be anything other than lot, we will be ignoring the known lot-to-lot variation and could get out-of-control points that already have a known, assignable cause - the data comes from different lots. However, in the lithography processing area the measurements of most interest are the site level measurements, not the lot means. How can we get around this seeming contradiction?

One solution is to chart the important sources of variation separately. We would then be able to monitor the variation of our process and truly understand where the variation is coming from and if it changes. For this dataset, this approach would require having two sets of control charts, one for the individual site measurements and the other for the lot means. This would double the number of charts necessary for this process (we would have 4 charts for line width instead of 2).

Another solution would be to have one chart on the largest source of variation. This would mean we would have one set of charts that monitor the lot-to-lot variation. From a manufacturing standpoint, this would be unacceptable.

We could create a non-standard chart that would plot all the individual data values and group them together in a boxplot type format by lot. The control limits could be generated to monitor the individual data values while the lot-to-lot variation would be monitored by the patterns of the groupings. This would take special programming and management intervention to implement non-standard charts in most floor shop control systems.