Analyzing Data and Selecting Software for Six Sigma

*(Published in Quality World August 2006)*

By Hemant Urdhwareshe

In six sigma approach, we convert a practical problem in to a statistical model using a process map. It takes a form of a mathematical equation:

$$Y=f(X1,X2,X3….. X_n)$$

where Y is the key process output (KPOV) and Xs are the key process inputs (KPIVs)

We then evaluate the process data to build (model) relationship of KPOV and KPIVs. This is easier said than done. It is in this context that Six Sigma implementation requires application of many advanced statistical tools. These tools involve complex calculations. Such calculations need not be done manually in twenty-first century. There are

good softwares that can make life of black and green belts much easier. Six Sigma is a data based approach for taking decisions to improve processes. Belts frequently need to analyze large amount of data in different ways so that it can provide strong clues to the solutions. With ISO 9000, QS-9000 and ISO/TS 16949, many companies do a good job of collecting data. This is an important step for making improvement. However, in absence of good understanding of tools to analyze data, it simply is an addition of cost for the company. It is extremely important for companies to train their key people in analyzing data and interpreting the results of analysis. Some of the tools that help analysis of data are:
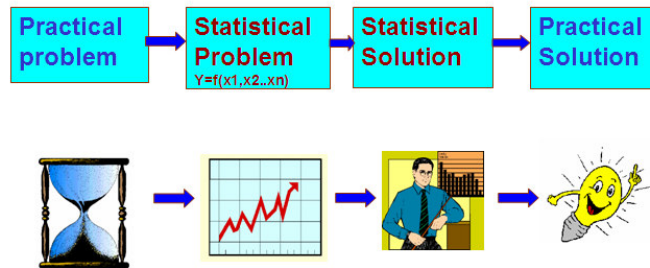
- Pareto Analysis for prioritizing improvement areas, projects
- Statistical Process Control (SPC) for assessing stability of processes and identifying presence of special or assignable causes
- Measurement Systems Analysis (MSA) to assess whether it is adequate for measuring process output
- Graphical Analysis of data using various tools such as:
  - Box Plots and Dot Plots
  - Histograms
  - Scatter Plots
  - Stem and Leaf Plots

- o Pie Charts
  - o Multi Vari Charts
- Correlation and Regression Analysis
- Multiple regression
- Hypothesis Tests such as t test, F test, Chi-Square Test
- Analysis of Variance (ANOVA)
- Planning and analyzing experiments for screening and characterization
- Response Surface Methods to optimize process
- Advanced control charts such as Cusum and Exponentially Weighted Moving Average (EWMA) Charts

This list is not all inclusive but these are the most commonly used tools in Six Sigma Projects. Often, process data is not normally distributed making the analysis more complex. Moreover, when data is analyzed by belts, it may not give the desired result in the first attempt. In case project requires analysis of field failures, it needs to be analyzed using different distributions such as Weibull and Lognormal. We now realize why softwares are necessary! In absence of softwares, belts will retire analyzing the data!

Some examples of data analysis are shown here to illustrate how simple graphical tools can be very useful in providing a clue to solve a problem or improve a process.

Figure 1 shows multi-vari chart of how strength of welding depends upon three factors Current, Force and time. The chart shows variation in strength with all combinations
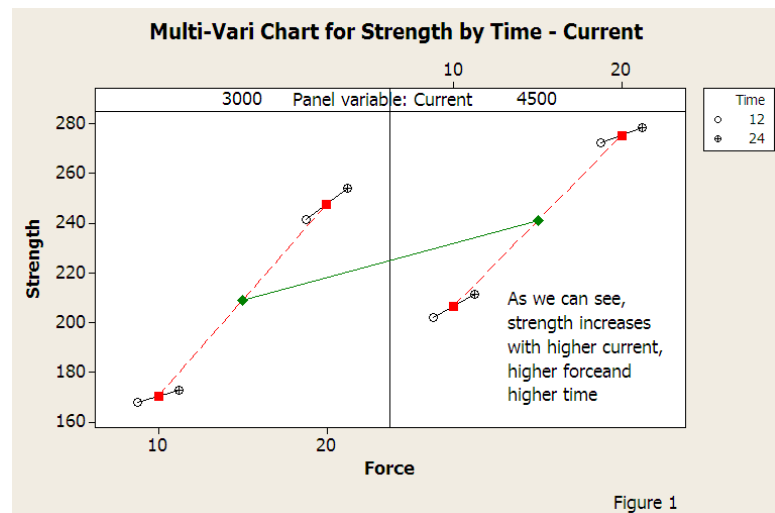

Figure 1

Figure 2 shows Scatter-plot of marks obtained by students and study time. We can clearly see a relationship between the two.
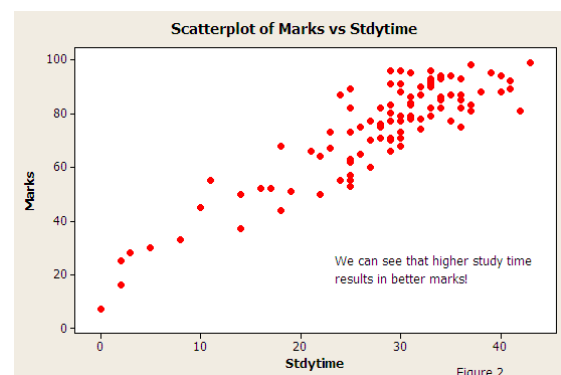

Figure 2

Figure 3 shows box plot of how paint gloss (shine of the paint) is affected by paint source, gun type and distance of painting.
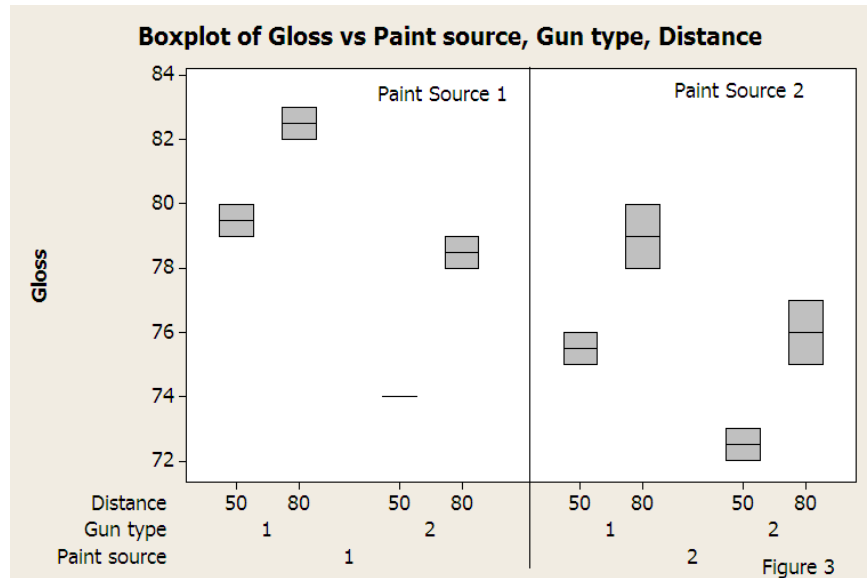
Boxplot of Gloss vs Paint source, Gun type, Distance

Figure 3

Figure 4 shows **matrix scatter plot** of units consumed in a factory and factors such as temperature, working days, production. We are not able to see strong relationships between individual factors. However, Multiple regression can show how they are related.
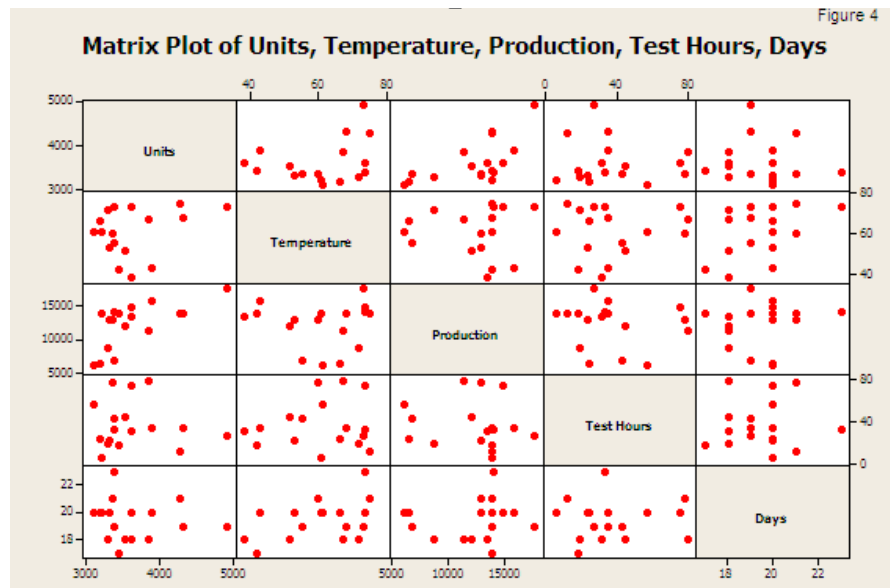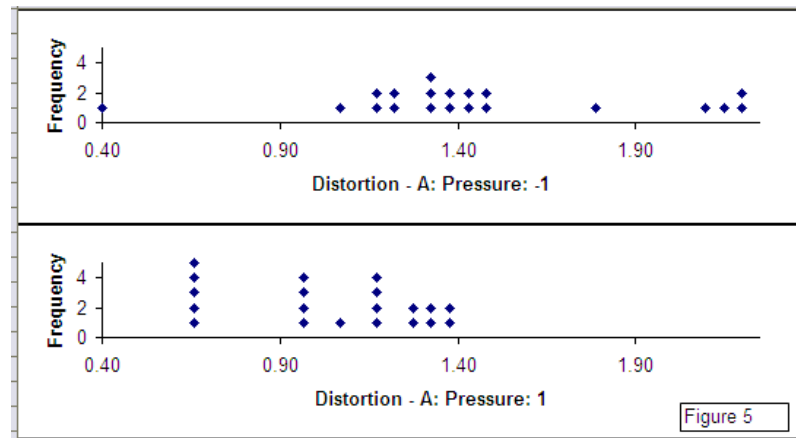
Figure 4

Matrix Plot of Units, Temperature, Production, Test Hours, Days

Figure 5 shows a **dot plot** generated in SigmaXL for distortion in plastic moulds as a function of injection pressure

Figure 5

There are many good types of software for analysis of such data. But choosing the right software is important while a company launches a Six Sigma Program. Frequently, choice of software depends upon recommendation of the consultant. Most commonly used softwares are:

- o Minitab
- o Statgraphics
- o Statistica
- o JMP
- o SigmaXL

I have not used all but used Minitab when I worked with Cummins India and later for Black Belt Training in various companies. I have also used SigmaXL for Six Sigma Green Belt Training. Both are good softwares. However, SigmaXL is Excel based user friendly software that can support green belts where analysis usually does not require very complex tools. Minitab on the other hand can support analysis needs of even a very highly complex nature. It is therefore an obvious choice over SigmaXL if company is launching a massive companywide Six Sigma Program with a pool of full time black belts and green belts. However, there are many companies, who launch Six Sigma program to assess its value. There are other companies who may not need many of the advanced tools such as response surface methods (RSM). This could be true for service companies. Why would a company prefer lower end software to the best in class? Frequently reason is cost. Cost of a Minitab single user license (about US$1195) may be about six times that of a SigmaXL license (about US$199). With the recent Version 5 of SigmaXL, its capabilities are enhanced and users need to carefully evaluate their choice. Table 1 provides a brief comparison of these two softwares.

Table 1

Comparison of Minitab and SigmaXL capabilities

	Minitab	SigmaXL
Graphical Tools	Excellent capabilities	Very good capabilities
Hypothesis Tests	Yes	Yes
Power and Sample Size	Yes	Yes
One Way and Two way Anova	Yes. Anova with more than 2 factors possible	Yes. Uses Excel Add-ins for 2 way. More than 2 way is not possible
Correlation and Linear Regression	Yes	Yes
Multiple Regression	Yes	Yes
Process Capability	Yes	Yes
Measurement Systems Analysis	Yes	Yes. Meets most of the needs.
Control Charts	Yes. Cusum and EWMA also possible	Yes. Cusum and EWMA not available. These are not very commonly used.
Design of Experiments	Yes. General Full Factorial is also available.	Yes. Two level Full and Fractional factorials
Anova: General Linear Model	Yes	No
Response Surface Methods	Yes	No
Reliability Analysis	Yes. Comprehensive analysis possible	Limited to Weibull. This was added in version 5.
Single User License Price (Please check current prices on their websites minitab.com and sigmaxl.com)	US$1195	US$199

As expected, we can see that capabilities of Minitab are better. However, its price is about 6 times that of SigmaXL. If a company desires to start Six Sigma program for green belts and does not expect complex cases of data analysis, they can choose SigmaXL rather than Minitab. They can later add Minitab for some of the belts who become experts in various tools and handle complex processes that need more tools and features.

Some of the limitations of SigmaXL are:

- Multi-way Anova is not possible. This may be required in some situations even in green belt projects
- General Linear Model is not available
- Not-normal data can be handled but there is limited capability. Minitab capability of selecting distributions from a dozen is excellent. SigmaXL has Box-Cox Transformation and Weibull distribution.
- SigmaXL has significantly improved its capabilities in design of experiments with 2 level-factorials in version 5. However, it cannot handle complex experiments with more than 3 levels.
- Response Surface Methods (RSM) are not available in SigmaXL. Minitab has excellent capabilities in RSM.

In spite of these limitations, SigmaXL is an excellent software for the value for green belt and general analysis needs in quality improvement. Considering its low price, number of installations can be more making it available to more employees than more expensive softwares.

There are others who tend to prefer Statgraphics, JMP or Statistica to Minitab. It seems that users get familiar with one of the softwares and tend to prefer the familiar software. Statistics will remain same as formulae do not change!

Note:

About Author:

Hemant Urdhwareshe is Director, Institute of Quality and Reliability, Pune. He is
- *Qimpro Foundation Silver Standard 2003 Award Winner*
- *First ASQ Certified Six Sigma Black Belt in India*
- *ASQ Certified Quality Engineer, Quality Manager and Reliability Engineer.*
- *The only person in India with 4 or more certifications from ASQ (as on Mar 2005)*
- *Senior Member ASQ*

Email: hemant@world-class-quality.com   *Phone: 020-25367952*

Visit: www.world-class-quality.com