A Study on Statistical Process Control (SPC) in Software Reliability

Dr. R. Satya Prasad, V. Goutham

1Dept. of Computer Science & Engg., Acharya Nagrjuna University, Nagarjuna Nagar, Guntur, Andhra Pradesh, India.
2Dept. of Computer Science & Engg., Vatsalya Group of institutions, Hyderabad, Andhra Pradesh, India.

Abstract—Statistical process control (SPC) is a statistical based tool to study how best one can describe and analyze the data and then draw conclusion or inferences based on available data. Critical business applications require reliable software and developing reliable software is a difficult problem that facing the software industry. To maintain consistency and capability of the software is determined by using statistical and process control methods. Therefore, Statistical Process Control (SPC) methods are simple and easy to construct and interpret. This paper reviews the use of SPC to measure and analyze the software reliability. This paper concentrates on statistical techniques and the usage of SPC in Software Reliability.

Index Terms—Error, Failure, Fault, Statistical Process Control, software reliability, Statistical Control

I. INTRODUCTION

The software reliability is defined as the probability of failure free software operations for a specified period time in a specified environment. We need a number of definition to introduce the concept of software reliability[15].

Error—These are human actions that result in developing software containing a fault. Example of such fault are omission or misinterpretation of the user’s requirement a coding error etc.

Fault—These are manifestation of an error in the software. If encountered a fault may cause a failure of the software.

Failure—It is the inability of the software to perform its mission for function with in specified time. Failures are observed during testing and operations.

Software Reliability—It is the probability that software will not cause the failure of a product for a specified time under under specified conditions. It is concerned with the time between failure or its reciprocal, the failure rate. Statistical process control is a statistical based approach which can determine whether a process is stable or not by discriminating between the presence of common cause variation and assignable cause variation[3][15].

The predicted number of defects in a software reliability is statistically estimated from test time and defect data. There are many different software reliability growth models in the literature and many types of data and different statistical techniques that can be used in combination with these models. Therefore fitting a mathematical function to data involves estimating the function’s parameter from the data. The two popular estimation technique are maximum likelihood and least squares. We found that least squares provides the best point estimation and the maximum likelihood is better for deriving confidence interval. Most software reliability growth models fall in to two classes Concave and S-shaped [14].

A Concave models are so called because they bend downward. S shaped models on the other hand are first convex and then concave. This reflects that early testing is not as efficient as later testing, hence there is an ramp-up period during which defect detection rate increases. A software process control is defined as an organized collection of methods used in the development of software product. Software process control can be defined as software that automatically monitors and controls a process to assure the output of the process conforms to manufacturing specifications[2][9]. SPC chart methods used for displaying data, it is the basis on which software organization make individual sound and successful project commitments. Yet software organization find this important activity as challenging. Even today it is clear that software industry in general does not estimate project throughly and does not use estimation appropriately. With an effective estimations, the whole organization’s attention can be focused on the
opportunities and risk in a project. However, there are significant challenges to applying SPC to software reliability like incorrect statistical technique, confusing threshold and control limits, measurement of behaviour change[2].

II. Statistical Interface and Estimation - The problem of point estimation is that estimation of the parameters of a population. It is assumed that the type of population distribution involved is known but the distribution parameter are unknown and they must be estimated using collected failure data[11]. The two most common techniques for estimating a mathematical function parameters from test data are maximum likelihood and least squares. The maximum likelihood technique estimates parameter by solving a set of simultaneous equations that maximise the likelihood that the observed data came from a function with those parameter values. It is generally considered to be the best statistical estimation for large sample size. The least square approach is to fit the curve described by the function to the data and estimate the parameters from the best fit to the curve[12]. This technique is generally considered to be best for small to medium sample size. Both statistical technique can be used to derive confidence interval[1][6].

Statistical Process Control - SPC is an analytical decision making tool which allows you to see when a process is working correctly and when it is not. Variation is present in any process, deciding when the variation is natural and when it needs correction is the key to quality control.

What are process control charts?
Control means predictability. If the variation in the behavior of a process is predictable in statistical terms, that process is said to be in control. This means that, we can expect (within certain limits) what the outcome will be the next time we perform the same process. In this way, we can prepare more accurate project plans, do better cost estimations and schedule activities in more reasonable basis. The aim of Statistical Process Control is: firstly to detect assignable causes of variation in the processes and provide process control; secondly to enable monitoring of the improvement in processes (that are already under statistical control) by demonstrating the chance causes; and Shewhart Control Charts are good means to achieve Statistical Process Control[5],[12]. Control charts show the variation in a measurement during the time period that the process is observed. Control charts monitor processes to show how the process is performing and how the process and capabilities are affected by changes to the process. This information is then used to make quality improvements.

Control charts are also used to determine the capability of the process. They can help identify special or assignable causes for factors that impede peak performance.

If a process is in control, the outcomes of the process can be accurately predicted. In an out of control process, there is no way of predicting whether the results will meet the target. An out of control process is like driving a bus in which the brakes may or may not work and you have no way of knowing. If a process is out of control, the next step is to look for the assignable causes for the process output, to look for the out-of-controlness. If this out-of-controlness is considered negative, such as multiple defects per part, the reasons for it are investigated and attempts are made to eliminate it. The process is continuously analyzed to see if the changes work to get the process back in control. On the other hand, sometimes the out-of-control outcomes are positive, such as no defects per part. Then the assignable cause is sought and attempts are made to implement it at all times. If successful, the averages are lowered and a new phase of the process is begun. A new set of capabilities and control limits is then calculated for this phase[10].

Control charts have four key features: 1) Data points are either averages of subgroup measurements or individual measurements plotted on the x/y axis and joined by a line. Time is always on the x-axis. 2) The Average or Center Line is the average or mean of the data points and is drawn across the middle section of the graph, usually as a heavy or solid line. 3) The Upper Control Limit (UCL) is drawn above the centerline and often annotated as "UCL". This is often called the "+ 3 sigma" line. 4) The Lower Control Limit (LCL) is drawn below the centerline and often annotated as "LCL". This is called the "- 3 sigma" line. The x and y axes should be labeled and a title specified for the chart[12].

III. Implement of SPC techniques - SPC has been widely used in manufacturing industries in order to control variability and improve processes[9]. The basic tools used for statistical control are [15][8]: Scatter Diagram, Run chart, Histogram, Bar charts, Pareto Chart, Shewhart Control Charts

Scatter Diagram - A scatter diagram is a tool for analyzing relationships between two variables. One variable is plotted on the Review Hours (x-axis) and the other is plotted on the Number of defects(y-axis). The pattern of their intersecting points can graphically show relationship patterns. Plot the data on the chart, using concentric circles to indicate repeated data points. Most often a scatter diagram is used to prove or disprove cause-effect relationships. A sample Scatter Diagram can be seen in Figure 2.1
Run Chart-Run Charts are specialized, time-sequenced form of scatter diagrams that can be used to examine data quickly and informally for trends or other patterns that occur over time. They look like control charts, but without the control limits and center line. It displays data in the time sequence in which they occurred. In Run Charts, there should be at least 20-25 points and y-axis should be 1 ½ times the range expected [13]. A sample Run Chart is given in Figure 2.2.

Run is nothing but one or more consecutive data points on the same side of the median. Excludes data points on the median.

Pareto Chart-Pareto chart is another form of bar chart. However, the occurrences are ordered with respect to their frequencies. X-axis contain Project phases with Y-axis as Rework Percentage. This (see Figure 2.5) Pareto charts are good means to visualize the ranking of an Rework Percentage with different Project Phases.

Shewhart Control Charts - Shewhart proposed that it is possible to define limits within which the results of routine efforts must lie to be economical. In order to detect assignable causes, Shewhart utilized statistics and control charts [10]. First of all, a sample of data is collected for the subject measure (i.e., number of defects in a piece of code) then, its mean and variance are calculated. The lower and upper control limits are derived and data is analyzed using the statistical evidence on hand. By analyzing the data values with respect to upper and lower control limits together with their location in the zones, assignable causes are detected.

Statistical analysis may be performed by implementing X and R charts for sample data. The limits for an Individuals Chart are calculated as:

\[
\text{Center Line } = \bar{X}
\]

\[
\text{UCL} = \bar{X} + 3 \frac{MR}{d_2}
\]

\[
\text{LCL} = \bar{X} - 3 \frac{MR}{d_2}
\]

where \(X\) is the average of individual values, \(MR\) is the average of moving ranges and \(d_2\) is a constant which depends on the number of individual values that moving range is calculated [12].

The fraction of defectives in a sample can be explained with a binomial distribution and p-chart is drawn to investigate statistical control. The other major charts that may be utilized for the analysis of attribute data are the following:
**np charts:** When we want to work on the number of defectives instead of the fraction.

**c-charts:** If we want to emphasize the number of defects (of a certain type).

**u-charts:** If we want to emphasize the number of defect per unit. The limits for u-chart are calculated as:

\[
\bar{u} = \frac{\sum d_i}{n},
\]

\[
UCL/LCL = \bar{u} \pm 3 \sqrt{\frac{\bar{u}}{n}},
\]

where, \( di \) : number of defects for measurement \( i \), \( ni \) : sample size for measure \( i \). Shewhart charts are good indicators for detecting out-of-control situations in a process. For ease of understanding, the charts are divided into zones (A, B and C), which represent values of “Mean ± 1/2/3 Standard Deviation” as seen in Figure 3.

![Sample Control Chart](image)

**REFERENCES**


**Author Profile:**

**First Author:** Dr. R. Satya Prasad received Ph.D. degree in Computer Science in the faculty of Engineering in 2007 from Acharya Nagarjuna University, Andhra Pradesh. He received gold medal from Acharya Nagarjuna University for his outstanding performance in Masters Degree. He is currently working as Associate Professor and H.O.D. in the Department of Computer Science & Engineering, Acharya Nagarjuna University. His current research is focused on Software Engineering. He has published several papers in National & International Journals.

**Second Author:** V. Goutham is an Associate Professor in the Department of Computer Science and Engineering at Vatsalya Group of institutions affiliated to J.N.T.U Hyderabad. He received M.Tech from Andhra University and B.Tech from J.N.T.U Hyderabad. He Worked for various MNC Companies in Software Testing and Quality as Senior Test Engineer. His research interests are Software Reliability Engineering, software testing, software Metrics, and Data mining.