Improve Forecasts: Use Defect Signals

Paul Below
paul.below@qsm.com

© Quantitative Software Management, Inc.

Introduction

- Large development and integration project testing phases can extend over many months
- It is important to predict when the defect detection rate will drop to acceptable levels
- SLIM Control can be used to track and forecast defect detection rates
- The peak of the defect Rayleigh curve is an important point in the testing phase
- Control charts are another tool
Cumulative Defects Found by Month, Project XYZ

Project has found fewer defects than estimated. What will happen next?

Defects Found by Month, Project XYZ

Has the peak of the curve really been reached? Or was the last month misleading? If only we had another signal!

Date we need to forecast
Comment on Defect Prediction

- It is possible to create very useful estimates of defects based on only a few key metrics. For example, I created a regression analysis to predict defects based on over 2000 recently completed software projects from the QSM database. This resulted in an adjusted R square of 0.537 using only the input variables the log of peak staff, the log of ESLOC, and the log of production rate (ESLOC per calendar month).
- Standardized residuals show a normal distribution with mean close to zero.
- Large projects have multiple testing phases. Such models, with multiple control charts, can be used throughout. For example, with one Fortune 500 client, I found that merely using the number of prerelease defects was an excellent predictor of their go live release defects (R Square of over 0.7).

Control Charts

- Every process displays variation. Some processes display controlled variation and others display uncontrolled variation.
- In the 1920’s, Control Charts were invented by Walter A. Shewhart. In most of the rest of the 20th century the concepts were popularized by people such as W. Edwards Deming.
- Control Charts are a major part of Statistical Process Control, and are integrated into Six Sigma and CMMI.

"The engineer desires to reduce the variability in quality to an economic minimum. In other words, he wants (a) a rational method of prediction that is subject to minimum error, and (b) a means of minimizing variability in the quality of a given produce at a given cost of production."  
Walter A. Shewhart
Control Charts and Signals

- Purpose here is **not** to determine if the testing is in control or to improve product quality (both have been done, but are outside scope of this presentation, see sources at end)
- Purpose here **is** to determine if there has been a shift in quality of the underlying product
- I use Individuals and Moving Range charts ($XmR$), suitable for many situations including the periodic collection of items found in a given time
- It is important to calculate the control limits properly (see resources at end, or use a good stats tool)

Rules, Rules...

- There are a number of rules that are used to detect signals. The number of rules used and the definitions of the rules vary slightly from one source to another. However, the traditional use of control charts is best met by keeping the number of rules to a minimum, thereby reducing the chance of obtaining a false signal. But in our case, we don’t want to miss a signal.
- In IBM SPSS 22, for example, there are 11 possible rules that can be turned on or off:
  - One point above +3 Sigma, or one point below -3 Sigma
  - 2 out of last 3 above +2 Sigma, or 2 of 3 below -2 Sigma
  - 4 out of last 5 above +1 Sigma, or 4 out of 5 below -1 Sigma
  - 8 points above center line, or 8 below center line
  - 6 in a row trending up, or 6 trending down
  - 14 in a row alternating up and down
Why the Rules?

- All uses of control charts walk this decision line. Shewhart originally used 3 sigma limits because he wanted to minimize false signals, which would incur the unnecessary cost of researching a problem that didn’t exist. In other words, when he saw a signal he wanted to be almost completely certain it was real.

<table>
<thead>
<tr>
<th>Possible States</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Possible Decision</td>
<td>Problem</td>
</tr>
<tr>
<td>Investigate</td>
<td>Correctly Accepted</td>
</tr>
<tr>
<td>Ignore</td>
<td>Type I error</td>
</tr>
</tbody>
</table>

Another Example

- In this example, the balance between testing and fixing has not remained constant.
- For defects detected, five of the points show up as red, meaning they violated one of the rules.
- These violations are not unusual. As mentioned previously, defect metrics commonly follow a Rayleigh distribution.
- What about backlog burndown?
Balance between fixed and found has shifted. Something has changed.

Control charts can be used to determine whether apparent changes in defect rates are significant. Especially, has the peak of the defect detection Rayleigh curve been reached?

One use for this knowledge is to create and improve forecasts of Program completion, or software quality at key Program milestones.

“...since we can make operationally verifiable predictions only in terms of future observations, it follows that with the acquisition of new data, not only may the magnitudes involved in any prediction change, but also our grounds for belief in it.”
Walter A. Shewhart
Example from SLIM Training

Defect Closure & Backlog - Web Gateway

SLIM Control Example: But wait, there’s more!

From the SW-Wed Gateway example packaged with SLIM Control 8.1

Defect Closure & Backlog - Web Gateway

Cummulative Found, Fixed & Open Defects

Defect closure is not keeping up with discovery rates.

Current Plan Actuals Green Control Bound Yellow Control Bound Red Control Bound
Let's Look at a Defect Forecast

SW-Wed Gateway example packaged with SLIM Control 8.1

Arbitrary: Weighted every metric evenly that had Excellent Goodness of Fit

SLIM Control Example

Forecast predicts we are still a couple of months away from peak of defect discovery Rayleigh curve.
SLIM Control Example
XmR for Defects Discovered

No rule “violations”. Therefore, no significant change (month to month variation in defects discovered can be explained by random variation).

SLIM Control Example
XmR for Defects Fixed

No significant change. Note though that the moving range chart picked up on month 2 as almost 3 sigma, when they ramped up fixing.
SLIM Control Example
XmR for Defects Fixed / Defects Discovered

No significant change. Control charts agree with SLIM Control Forecast, in that we have no evidence that curve peak has been reached.

Resources


Paul Below
paul.below@qsm.com
http://www.qsm.com
Extra Material

XmR Control Limits

- Individuals Chart (X)
  - Center line is the mean of X, mX, the average of all the individual values
  - Upper Control Limit is mX + (2.66 * mR)
  - Lower Control Limit is mX – (2.66 * mR)

- Moving Range Chart (mR)
  - Center line is the mean of ranges, mR
  - Upper Control Limit is 3.268 * mR

Mean Range, mR, is the average of a moving range of 2. For each data point, except the first, the moving range of 2 is the difference between that data point and the preceding one. Those ranges are averaged to get the mR. For in depth explanation, see Chapter 3 of Wheeler (on resources slide).