FREQUENTLY ASKED QUESTIONS

4

(courtesy of previous students)

Assignment 4 - CSE 6329

DMAIC Report with root cause analysis and data analysis

**1) In section 4 - Analyze High Defect Level Problem - what exactly needs to be done? Is it like A2? We've to generate graphs for some cases (graph by team size, development process, etc.) and add those graphs to the document?**

A) In phase 1 (task 4.1) of this section your goal is to identify potential causes by examining three factors that might be correlated to the high defect rate: project manager, programming language and development method. In part 1 of the course we discussed many different kinds of analysis methods (see the suggestions in the SOW) and also several useful graphs such as box plots, bar charts, control charts, scatter diagrams, pareto charts and so forth. We also discussed a lot of other data analysis techniques such as looking at the data, sorting the data, distributions, central tendency measures, measures of dispersion, and confidence levels. You may find some of these helpful in doing analysis or in showing the results of your analysis. The key is deciding what sort of analysis to use, applying that analysis, and recording what you learn in the tables and perhaps illustrating using an appropriate graph. Note that the results of many analysis techniques are best shown by using graphs of some kind. For example, X-Y charts are one way to show a correlation. And a causal model is often used after other forms of analysis to show all the causes of some problem in one diagram. Note that the SOW gives you some suggestions about the analysis methods to use in each phase.

In phase 2 (task 4.2), your goal is to apply a somewhat more sophisticated analysis to see whether there are any correlations between size or release date and defect levels. In class, we discussed several potentially suitable techniques such as analysis of variations, correlation analysis (Pearson and Spearman coefficients), regression analysis, and Bayesian statistics. You should select techniques that make sense and see what they tell you. Note that many of the statistical formulas mentioned in class are available in Excel, so it would be wise to put your data into Excel, refine the data appropriately, and use Excel formulas to compute things like correlation coefficients.

The file "**IPC Defect Data**" (SH11), which was provided to you in one of the supplementary handouts, contains defect data for all of the company's products. Your job is to analyze this data to determine what factor or factors, if any, appear to be causing the high defect problem. In phase 1 you look for possible causes due to the project manager, development method or programming language. In phase 2 you further evaluate correlations related to release date and size. In the first half of the course, you were presented with many methods of data analysis. Your job is to show that you know how to use what you learned when presented with real data.

I suggest that you go through the methods you learned in the first half of the course (as a team if you have a teammate), determine which of those methods might be suitable for analyzing the defect data in each phase, apply those methods, and see if you learn anything. In your DMAIC report (tables 4.1 and 4.2), you should indicate the following:

-- which methods you tried (and the results of applying each method, even if the results were that you didn't learn anything or you learned that something is NOT a strongly correlated to the defect problem);

-- which method or methods provided information that suggests strong correlation and thus possible causes of the high defect problem

Then in Table 4.3.a you should:

-- summarize the likely cause(s) of the high defect rate problem

And in task 4.3.b you should provide:

-- a cause map showing cause(s) of the high defect rate problem;

Finally, in task 4.4 you should provide:

-- any further data collection and/or analysis you think might be useful.

Note that when you apply an analysis method, you may have to apply it several times to different parts of the data before you find something that shows a significant relationship. For example, suppose you examine programming language and decide that it has no apparent effect on defect rate. You would report that as an analysis you tried that showed no relationship. But you might examine other factors and determine that one or more of the others had a significant (or perhaps moderate or small) effect on the defect rate.

Note: A significant part of your credit for A4 is how well you have used the analysis methods from the first half of the course to analyze the data. In other words, did you select appropriate analysis methods and did you apply them correctly. You do not have to draw any graphs, although you may do so as part of your analysis process and may choose to show your graph if you think it will help explain your findings.

**2) For section 5, Improvements and Recommendations, we are describing root causes and providing some solution to the root causes. Didn’t we already describe the root causes in sections 3 and 4?**

A) The company now knows the potential root causes for the two problems, but does not necessarily know which ones are most important to fix first. Section 5 is intended as a summary of your findings and an overall recommendation about the order in which problems should be addressed. In section 5, you are combining the findings of sections 3 and 4 and putting them in priority order. In other words, you are supposed to consider what you found in sections 3 and 4 and decide, of all the root causes you have identified, which one is most important for the company to fix, which is second most important, etc. For each of them, you are to describe the root cause and explain why it is important (in table 5.1). And you list them in priority order. Why? Because the company may not choose to do all of them and part of your recommendation is telling them which one(s) would be most beneficial, in your opinion.

**3) In section 5.2 under recommendations, what do we have to do? Do we suggest a solution to each root cause? Do we have to give some flowgraphs for those? In the template it says to just write one paragraph.**

A) You should indicate what you think the company should do for each of the root causes listed in section 5.1. One paragraph should be sufficient for each root cause, although you might decide to show a diagram or graph if you think it will help explain your recommendation. For example, should they do better training? Or should they change some parts of their process? Or should they change some of their policies?