Data Analysis & Statistical Methods for Command File Errors

Leila Meshkat, Bruce Waggoner,
Larry Bryant
Jet Propulsion Laboratory
California Institute of Technology
12 June 2014

Outline

- Background
  - Definitions
  - Operations Best Practices and the Data Set
- Data Analysis
  - Correlation Analysis
  - Regression Analysis
  - Goodness of Fit Test
  - Principal Component Analysis
  - Sigma Tool
- Summary & Conclusions

Background

- Command File Errors (CFEs) are one of the main contributors to operational problems on space missions. Managing CFE's entails making good decisions about activities that are aimed at keeping them within acceptable bounds and preventing the occurrence of critical CFE's within reason.

Background: Definitions

- **SCMF** (Spacecraft Message File) - the binary file sent to the spacecraft - common to most missions
  - SCMFs can contain one to thousands of commands, usually multiple commands form sequences
- **Commands** are the discreet instructions issued to the spacecraft by the ground; these can be real-time or sequenced
  - Some commands are simple and some can be very complex due to parameterization
- **Sequences** are sequences of commands
  - Some sequences are simple and some can be complex
- **Blocks** are groups of commands that can be reused
  - Blocks can behave differently if they are parameterized
  - Blocks can be simple or complex
- **Activity Level** - A subjective estimate of the workload on the flight team. This factor was determined by averaging the number of concurrent sequence development activities from the development schedule.
- **Novelty** - A subjective estimate of the 'newness' of flight activities being executed on the spacecraft. For our purposes, novelty should is usually defined to be a minimum of one month in duration, but usually more. For example, the first three months after launch or a mission phase transition would almost universally be a period of novel operations.

Background: Operations Best Practices

- If the flight team radiates a larger number of SCMFs, they will make more errors
  - This has always been an argument for sequencing versus real time commanding
- If the flight team issues a larger number of commands, they will make more errors
  - We are usually sensitive to not sending extraneous command to the spacecraft
- If the flight team is experiencing a high level of development activity, they will make more errors
  - This is just human factors common sense
- If the flight team uses blocks they will make fewer errors
  - Reusing blocks will reduce errors since you are using a proven product
- If the flight team is doing something novel on the spacecraft, they will make more errors
  - This is also human factors common sense

Data Analysis

- Data Sources:
  - Information about the SCMF's sent to the spacecraft, in terms of the number of blocks, commands and files in each month of the mission and the number of CFE's observed.
  - Information about the level of activity (in terms of the products produced) by the flight team during each month.
  - Information about the novelty levels associated with each month of the mission.
  - General information in the databases in terms of the various errors that have occurred during the lifecycle of the flight projects and details about their causes and mitigations.
  - Our main goal in data analysis was to validate the results of the sigma tool. This is the first of the tools to be adopted by flight teams as it is the simplest to use. The sections below describe the different types of analysis conducted for this purpose.

Correlation Analysis

- At the onset of the study, we decided to look into the correlations between the following variables in each month of the mission:
  - Command File Errors
  - Number of files
  - Number of commands
  - Number of blocks
  - Activity levels during the month (in terms of products produced by the flight teams.
- Experiments Conducted:
  - Phasing
  - Time shifting for activity levels
- Result: Add Novelty Factor

Regression Analysis

- The goal of this analysis was to determine how much of the variability in the Command File Errors can be explained with a nonlinear function of the variables in question.
- Of course the caveats of this study are that
  - (1) the CFE's are not continuous variables and therefore they can't be predicted as a continuous function of the variables and
  - (2) we know that the behavior of the system is probabilistic rather than quadratic, so a quadratic equation does not completely capture the variability of the CFE's.
- For the flight projects analyzed, the R-squared value was approximately 50% based on all the variables.
- When we conducted a regression analysis only on the Novelty factors and the number of SCMF's, the R-squared value was reduced to about 40%.
- Given the caveats listed above, we conclude that the variables in question, especially the novelty levels and the number of files are very significant factors.

Chi-Squared Goodness of Fit Test

- We used a chi-squared goodness of fit test to test the hypothesis that the Binomial distribution with the parameters we had used was in fact the correct distribution for the distribution of the command file errors.
- For this purpose, we binned the files into those with errors and those without errors.
- We then calculated the expected number of files with errors based on our distribution, and also assessed the observed number of errors in each case.
- Then, using the equation $X^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$, where Oi is the observed frequency for bin i, and Ei is the expected frequency for that bin based on the hypothesis that our distribution is correct, we obtain the value to be compared with a chi-squared distribution with one degree of freedom to assess the goodness of fit.

Principal Component Analysis

- Since the variables SCMF, Commands and Blocks are correlated, we conducted a Principal Component and Factor Analysis to determine an equation for “Adjusted SCMF”.
  - The idea is that all SCMF’s are not equal – the ones with more blocks and more commands are slightly more error prone.
  - Equation obtained from Principal Component analysis. This equation can be used for the parameter “n” in the Binomial Distribution.
    - the “p” parameter will depend on the novelty level of the mission.

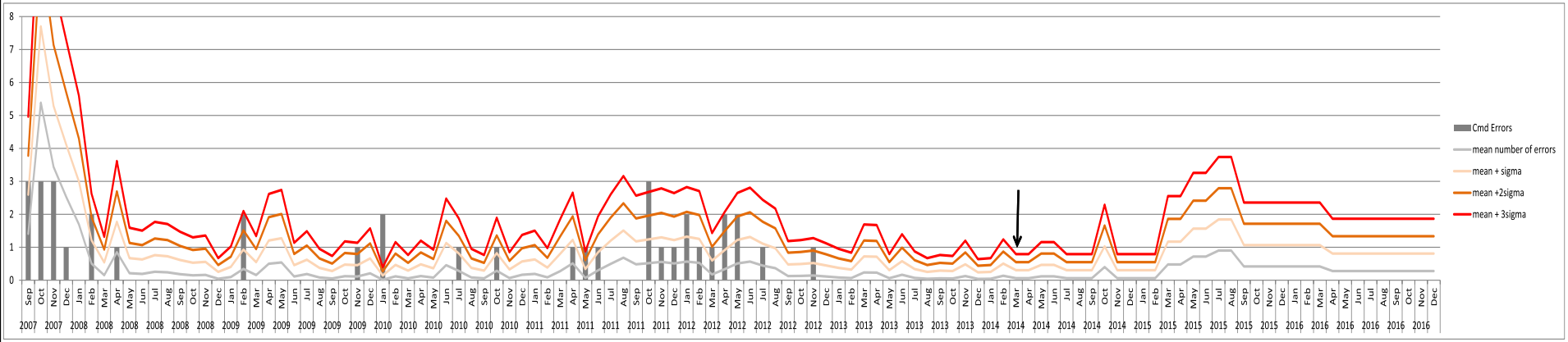
Sigma Tool

- Use Excel to enable cheap, multi-platform deployment
- Limit input parameters
  - Project duration built into spreadsheet
  - Number of Files sent per month
  - High or low novelty to be applied to any given month
    - High novelty raises the risk of an error by 50% for any given ‘trial’ for that month
- Use a binomial distribution to model each SCMF radiated as a trial which might result in a commanding error
- Calculate the mean expected number of errors in a given month
- Calculate the one, two and three sigma levels associated with the mean expected number of errors
- Capture the data in a spreadsheet and in a graphical output

Sample Project Results!

- One look at the plot gives the answers to our example questions and a lot more….
  - The post-launch CFEs were not unexpected, in fact the team did great!
  - We should expect a CFE every two months during novel operations
  - Our current drought of CFEs may not ***just*** be because we are good, but because we are commanding so little and so repetitively
  - There were times during Novel operations when we were ***appropriately*** concerned

0.004/0.006 Error Rate



Summary

- We discussed the distribution of command file errors as a function of several key variables and used a variety of empirical analysis techniques to refine and validate these distributions.
- The distribution used for this purpose is a Binomial Distribution. The parameters for this distribution are obtained by considering the number of trials to be a function of the number of files and the number commands and blocks in each file.
- The error rate is dependent on the level of novelty of the operations and is derived from similar data from past missions by the Mission Assurance Operations Manager.
- Building this model has resulted in a tool, which we call the “Sigma Tool”. This tool is used by management to assess the range of the CFE’s and determine what to expect based on mission plans.

Conclusions

- Here is our assessment of the current best practices:
  - If the flight team radiates a larger number of SCMFs, they will make more errors
    - True! This seems to be the primary driver for CFEs
  - If the flight team issues a larger number of commands, they will make more errors
    - True, but not as important as SCMFs radiated or novelty
  - If the flight team is experiencing a high level of development activity, they will make more errors
    - False, it appears staffing takes this risk down, in other words we manage this one.
  - If the flight team uses blocks they will make fewer errors
    - Not true. We suspect that blocks are a double edged sword as they are sometimes used ‘inappropriately’ (screwdrivers don’t make good hammers).
  - If the flight team is doing something novel on the spacecraft, they will make more errors
    - True! The second biggest driver.

Questions