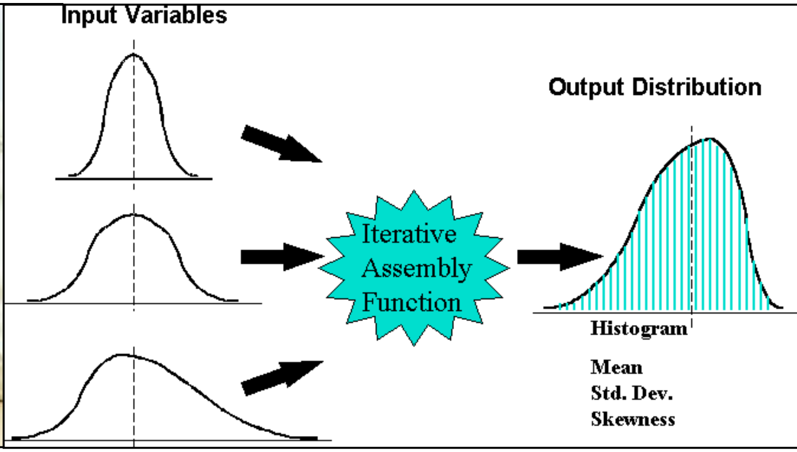
Use of Monte Carlo Simulation for a Peer Review Process Performance Model

Presenter: Emerald Russo, Systems Engineering
US Combat Systems, BAE Systems

Credit for photo references found at end of presentation.



Agenda

- Introduction to BAE Systems
- Performance Model Description
- Business and Project Goals
- Model Output & Stakeholder Description
- Model Input Description
- Data Collection
- Modeling/Analytical Techniques Used
- Results and Benefits
- Lessons Learned
- Summary
- References & Credits

Introduction to BAE Systems

- BAE Systems: US Combat Systems
- BAE Systems Land & Armaments, U.S. Combat Systems is a world-leading developer and producer of a full spectrum of gun systems, weapon launching systems and containers, as well as armored combat systems, such as the Bradley Combat System and next-generation systems for manned and unmanned ground vehicles.
- The division has several facilities located around the U.S. The relevant facilities to this presentation are: Santa Clara, CA; York, PA; Orlando, FL.
- In October 2008 USCS (legacy Ground Systems) received CMMI level 5 Certification.
- **Photo: Bradley Fighting Vehicle**

Performance Model Description (1 of 2)

- This model is a result of an annual review of peer review statistical data and pilot usage of Monte Carlo simulation tools for a trade study.
  - Annual review of peer review statistical data identified multiple independent contributor factors to the peer review process
    - Monte Carlo simulation is ideal for application to complex processes with multiple independent variables
  - This model was generated using a Monte Carlo extension (Crystal Ball or @RISK) to Microsoft excel in conjunction with predictive equations from regression studies of historical peer review data with Minitab.
    - By using historical peer review data as the basis for assumptions, the model has the ability to create thousands of realistic simulations of peer review results to help predict ranges of outcomes based on one or more inputs by the user.

Performance Model Description (2 of 2)

- Who developed process performance model?
  - Emerald Russo, member of the BAE USCS Measurement & Analysis group, main developer of the Peer Review Process Performance Model

- Who uses the model?
  - The purpose of this model is intended as a planning aid for anyone responsible for planning a peer review of a document.

- How often is the model used?
  - Historical data suggests that peer reviews are held in clusters for projects, usually held before major project milestones, more peer reviews are held earlier in project lifecycle.

Business and Project Goals

Organizational Business Goals	
Quality	Deliver products using controlled processes to meet customer expectations for product quality.
Capability	Continuously improve the organization's process capability to develop and integrate systems and products.
Cost	Develop products for USCS programs within defined budgets and at an affordable/competitive cost by using efficient tools and processes.
Time to Market	Consistently meet schedules and continuously improve the time to develop products.

Product or process performance objectives that the process performance model supports:

- At the org. level: Quality, Cost, and Capability business goals
- At the project level:
  - Plan for peer reviews with the intention to ensure highest quality of documents released without wasting resources such as time or money.
  - As a process performance model, it serves to drive continuous improvement of the modeled process.

Model Output & Stakeholder Description

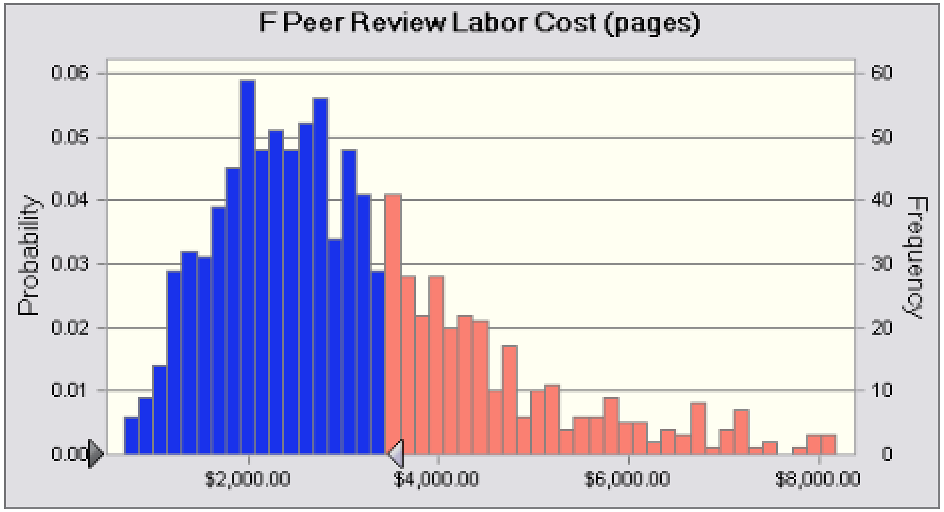
Output	Stakeholder Audience
Expected Defects	Planner / Author
Reviewers Needed	Planner
Estimated Peer Review Hours Expended	Project manager / Planner
Estimated Total Labor Cost	Project manager
Estimated Cost per Defect	Project manager / Organization

This model can guide a peer review moderator in planning decisions such as:

- How many people are needed to attend a peer review to assure that the number of defects escaped is less than or equal to ‘x’, y% of the time.
- It will also help in cost planning by helping projects budget for peer reviews based on average cost values of peer reviews by knowing how many documents and pages are required to be reviewed before release.

Example Forecast of Expected Peer Review Cost

- This is an example forecast (output) using Crystal Ball to predict the labor cost of a peer review for a functional document with an input of 20 pages.
- For functional documents, the mean amount of total PR hours is 34.1 hours per peer review.
- Using the example labor rate of $100, this would translate into roughly $3,410.
- The chart shows that there is a 67% chance that the input peer review will cost $3,410 or less.



Additional Statistics	
Statistics	**Forecast Values**
Trials	1000
Mean	$3,178.17
Median	$2,759.81
Standard Deviation	$1,782.30

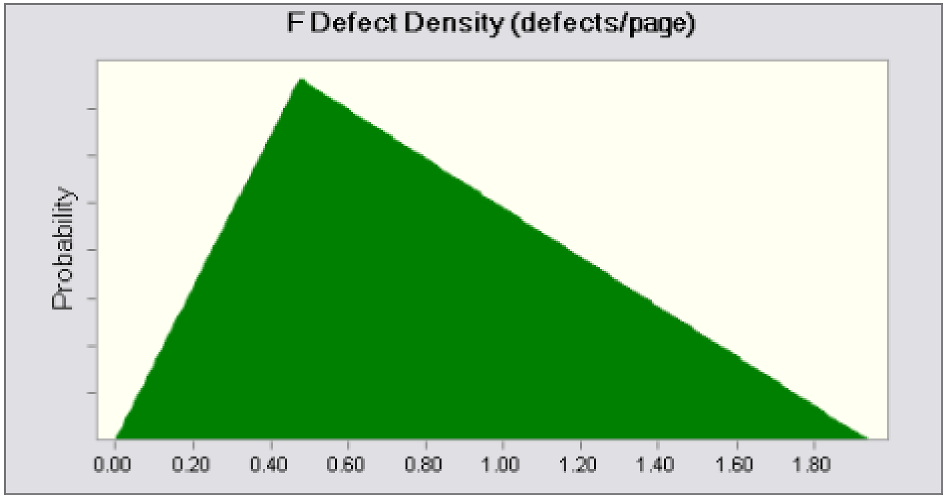
Summary:

- Certainty Level is 66.7%
- Certainty range is from -Infinity to $3,410.00
- Entire range is from $696.43 to $20,481.89
- After 1,000 trials, the std. error of the mean is $56.36

Model Input Description

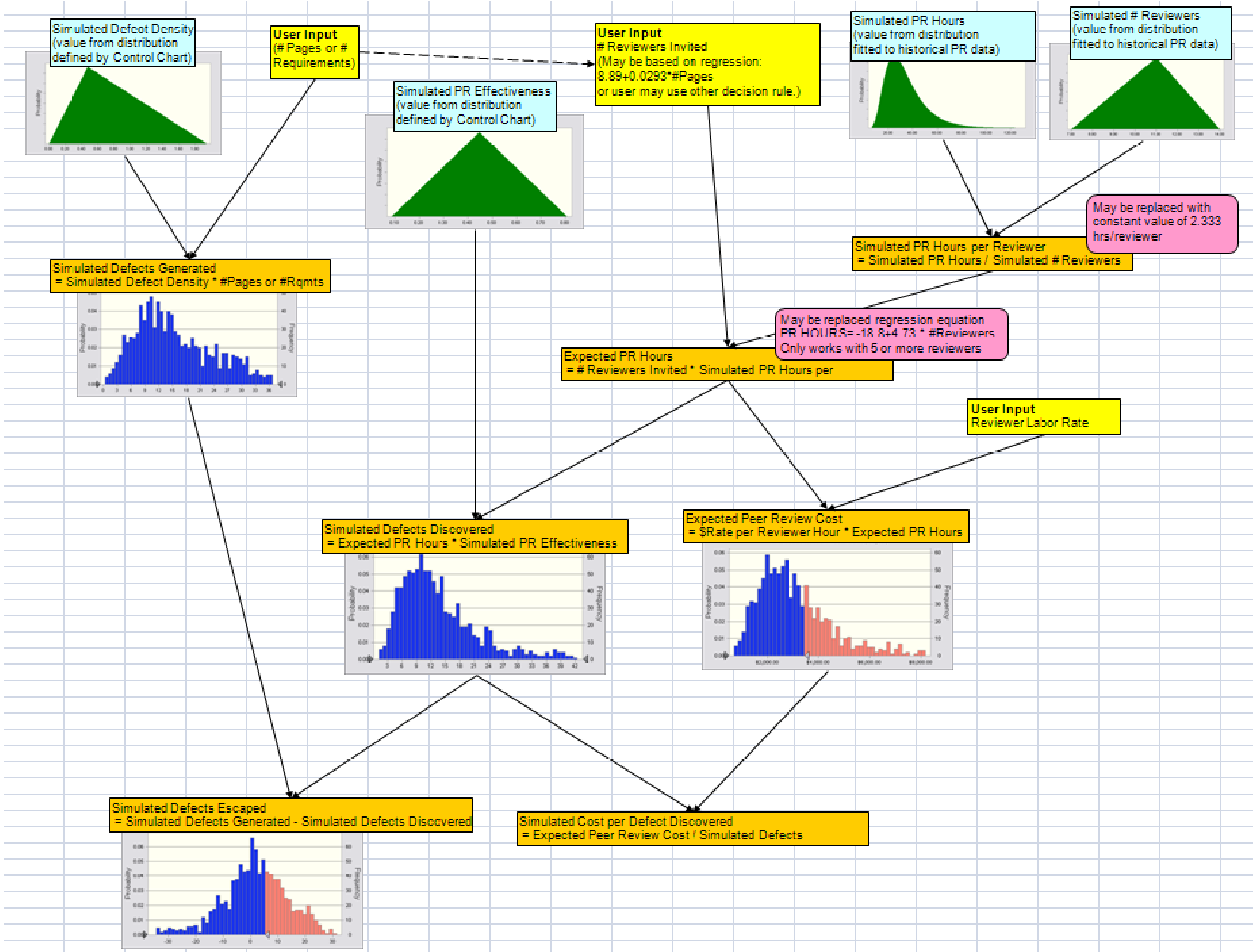
Input Name	Data Type	Controlled/ Uncontrolled	Amount of historical data	Significant/ correlation	Other Data
Defect Density	Simulated	Uncontrolled	All documented peer reviews, segmented by sizing information (ex. Pages, requirements, or SLOC)	YES/HIGH with document size	Value from distribution defined by control chart (historical data)
Document Size	constant	Controlled		YES/varying, see all others	User Input of # Pages or # Requirements
PR Effectiveness	Simulated	Uncontrolled	All documented peer reviews	YES/HIGH with PR Hours	Value from distribution defined by control chart
Reviewers	constant	Controlled	All documented peer reviews	Possible/Very Low with document size	User input, regression equation using document size as input, or value from distribution fitted to historical PR data
PR Hours	Simulated	Uncontrolled	All documented peer reviews	YES/High with Reviewers	Value from distribution fitted to historical PR data
Labor Rate	constant	Controlled	N/A	N/A	From project contract agreement or generally accepted labor rate per location

Example Assumption for Simulated Defect Density



- This is an example assumption (input) using Crystal Ball to define defect density (in defects per page) values for functional documents. The minimum value is the lower control limit from the baseline value of defect density control chart for functional documents, value = 0. The most likely value and maximum values are also from the baseline control chart for defect density of functional documents. By defining the defect density as a range of commonly observed values, the model can randomly choose potentially valid defect density values to simulate an individual peer reviews. These results are used for the forecast models.

Map of Inputs & Outputs



Data Collection

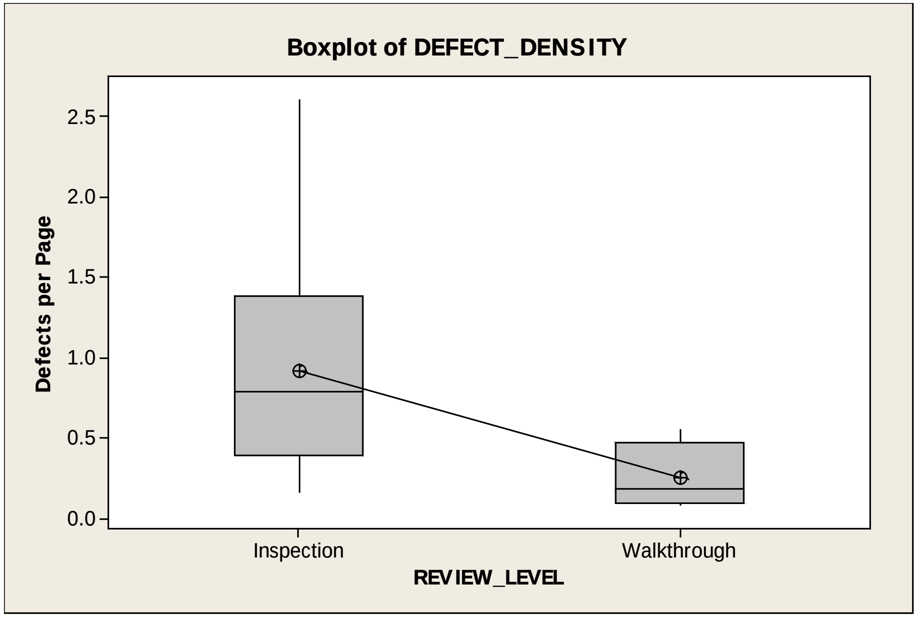
- Data was collected using database tool, Data Drill
- Data sampling not necessary, due to amount of data available at time of model creation
- Previous to this model, data was tracked in control charts and classified as “out of control” if falling outside of organizational baseline values, baseline values are re-evaluated on at least annual basis.
- Original pool of data used to create model ranged from peer reviews held in February 2006 to April 2008 for projects hosted at York PA BAE site.

Modeling/Analytical Techniques Used

- **Analytical method used:**
  - Data with medium to high linear correlation was estimated with regression equations.
    - Analysis of Variance used to decide if data should be grouped/separated for baseline process
    - Regression equations generated using regression analysis and general linear model functions of statistical program Minitab
  - Other data was estimated by fitting distributions to the historical data and using Monte Carlo simulation to simulate realistic data values.

Example of One-way ANOVA Studies for Significance

Plot Title	Significance
DEFECT_DENSITY vs REVIEW_LEVEL	Yes
PR_EFFECTIVENESS vs REVIEW_LEVEL	None
DEFECT_DENSITY vs PROJECT_PHASE	None
PR_EFFECTIVENESS vs PROJECT_PHASE	None



- This indicates that Inspectional peer reviews generally find more defects per page than walk-through peer reviews.

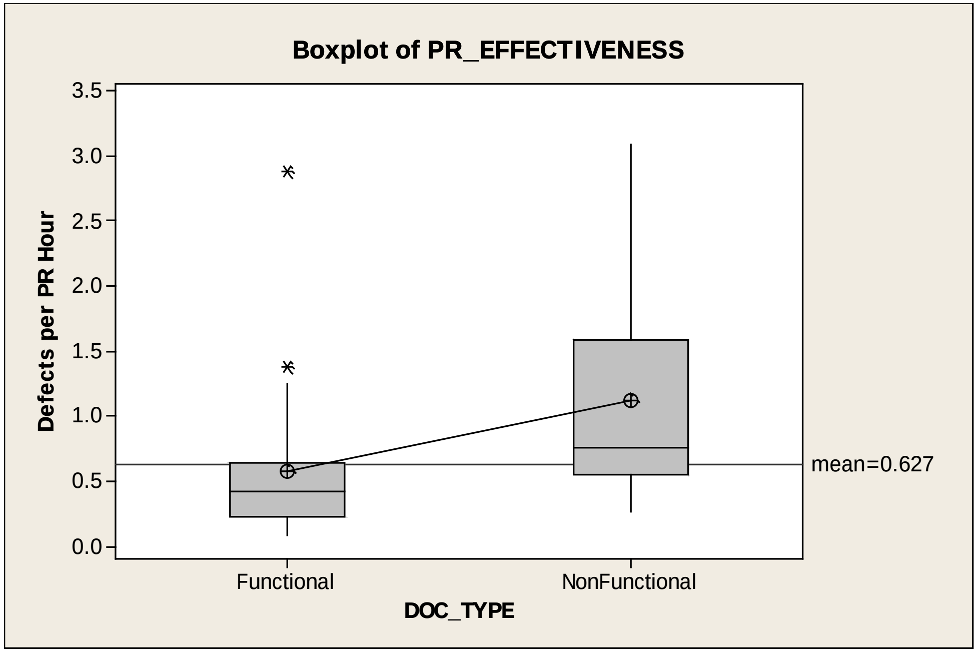
- NOTE: Studies based on baseline data, not segregated by document type.

Example Using Analysis of Variance: Defect Density, PR Effectiveness by Project & Document type

- Defect Density by Project & Document type
  - Analyzed with general linear model, found no significance
- PR Effectiveness by Project & Document type
  - Analyzed with general linear model, found project is not significant but document type is significant

Upon Further Analysis:

- One-way ANOVA of PR Effectiveness vs. Document type reveals high significance
- According to GSD Organizational policy, all following analyses will discriminate by document type



General Linear Model Results: Significance, Correlation

Plot Title	Significance	Correlation
PR_DEFECT_COUNT vs PR_HOURS	Yes	None
PR_DEFECT_COUNT vs PLANNING_TIME KICK_OFF_TIME INSPECTION_TIME REPORT_BACK_TIME	Yes Yes Yes None	Medium
PR_DEFECT_COUNT vs PLANNING_TIME INSPECTION_TIME KICK_OFF_TIME	Yes Yes Yes	Medium
WORK_OFF_TIME vs PR_DEFECT_COUNT	Yes	Medium-low
PR_DEFECT_COUNT vs NUMBER_OF_ REVIEWERS NUMBER_OF_PAGES NUMBER OF REQUIREMENTS	None May be None	Low
PR_DEFECT_COUNT vs NUMBER_OF_PAGES	Yes	Low-None

- NOTE: Studies based on baseline data, not segregated by document type.

Significance, Correlation, & Predictive Equations

EXAMPLE REGRESSION ANALYSIS

Regression Analysis: PR_DEFECT_COUNT versus NUMBER_OF_PAGES

The regression equation is
PR_DEFECT_COUNT = 8.80 + 0.223 NUMBER_OF_PAGES

30 cases used, 5 cases contain missing values

Predictor	Coef	SE Coef	T	P
Constant	8.798	3.002	2.93	0.007
NUMBER_OF_PAGES	0.22282	0.07462	2.99	**0.006** **Significant**

S = 11.5076 **R-Sq = 24.2%** **R-Sq(adj) = 21.4%** **Low-No correlation**

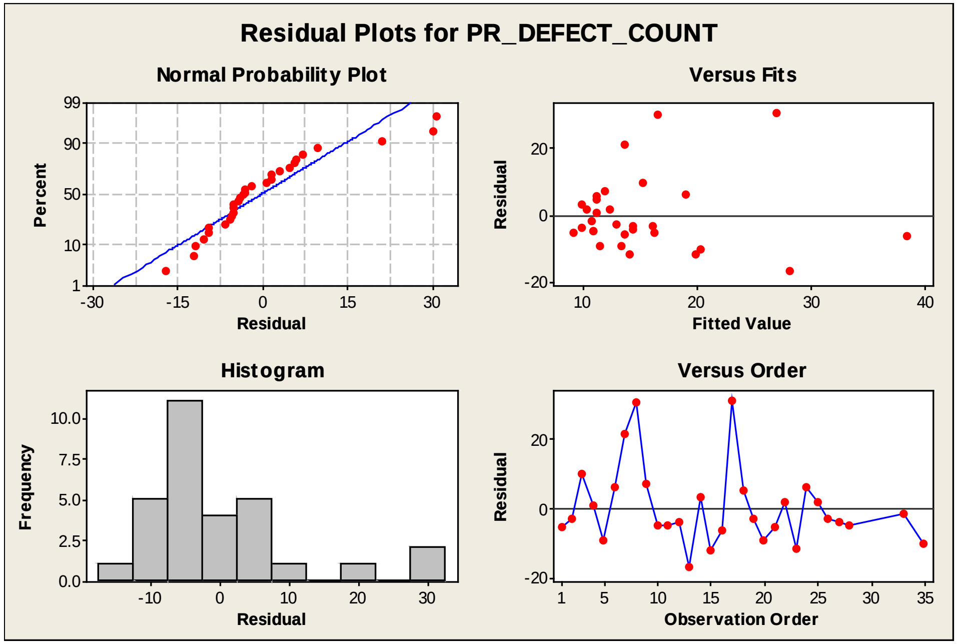
Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	1180.9	1180.9	8.92	0.006
Residual Error	28	3707.9	132.4		
Total	29	4888.8			

Unusual Observations

Obs	NUMBER_OF_PAGES	PR_DEFECT_COUNT	Fit	SE Fit	Residual	St Resid
8	35	47.00	16.60	2.15	30.40	2.69R
16	133	32.00	38.43	8.06	-6.43	-0.78 X
17	82	58.00	27.07	4.50	30.93	2.92R

R denotes an observation with a large standardized residual.
X denotes an observation whose X value gives it large leverage.



BAE SYSTEMS

Results and Benefits

- Results
  - Tool allows multiple stakeholders information about peer reviews before they occur, results in less out of control instances
- Benefits
  - Less out of control instances results in higher quality documentation.
- Overall Acceptance & Buy-in by Stakeholders of model
  - Planners were initial intended users and had the most chances for buy-in before model was created, other outputs were identified later as tools for other potential users
- Effort Invested
  - Baseline activity was performed separately from performance model development but that information was essential and would have to be done before a model could be created.
    - Including all baseline activity, approx 80-100 labor hours of work invested in development of this model.
- Effort to Maintain & Use Model
  - Model should be maintained in line with baseline values, estimated at 15 labor hours
  - Model usage is very quick, less than 10 minutes to run simulation model and create output report (assumes user has all input information before attempting to use model)

BAE SYSTEMS

Lessons Learned

- Challenges
  - Not all intended stakeholders have made use of model, currently planners are the only known users, may
  - Defining rules regarding when to use regression equations and when to fit distributions to historical data
  - Wide variety of projects in different lifecycle phases made baseline process difficult when it came to understanding what was in or out of control

- What Worked Well
  - Data collected from database was easily imported to Minitab statistical program, sped up baseline process
  - Monte Carlo Excel add-ins (@RISK & CrystalBall) both have user-friendly reporting features, allows for quick production of performance model results for analysis

Summary

- Baseline efforts resulted in following recommendations:
  - Allow 15 minutes to 1 hour for kick-off meeting (average value not available due to lack of data)
    - Allows reviewers to become familiar with document, affiliated project, and establishes ground rules.
  - Process of data recording in peer review forms needs to be better defined
    - Several cases of “out of control” data were due to inconsistencies in data collection.
  - When possible, plan for an Inspectional peer review
    - Analysis proves that Inspectional reviews find more defects per page than walkthroughs.
- Suggestion: Incorporate regular usage of process performance model in peer review process
  - This should help cut down on planning time and improve future peer review execution by allowing planners to understand range of expected outcomes and to react quickly to incoming out-of-control data

References/Credits

- Cover images from following sources:
  - Image Description: People at a table reviewing documents.
    - PowerPoint Clipart
  - Image Description: Picture of Input curves and output distribution.
    - http://adcats.et.byu.edu

- More information on BAE USCS Monte Carlo Tool Deployment study from Fred Oleson, speaker at 2009 NDIA CMMI Conference (on this topic)