Big Data Meets Earned Value Management

We have lots data. How can we use it to make predictive and prescriptive forecasts of future performance to increase Probability of Program Success?

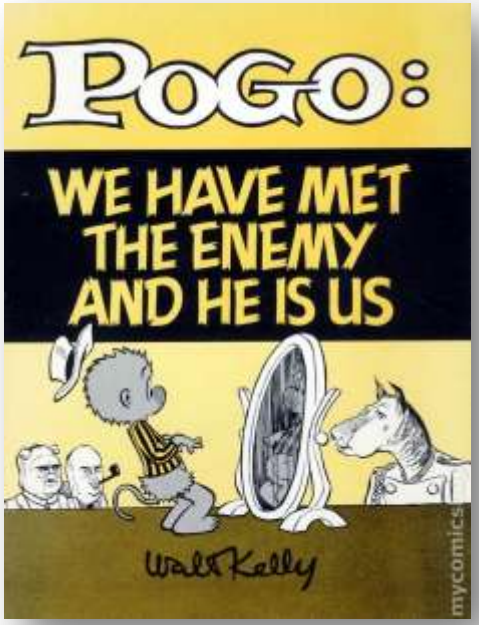
Glen B. Alleman
Thomas J. Coonce

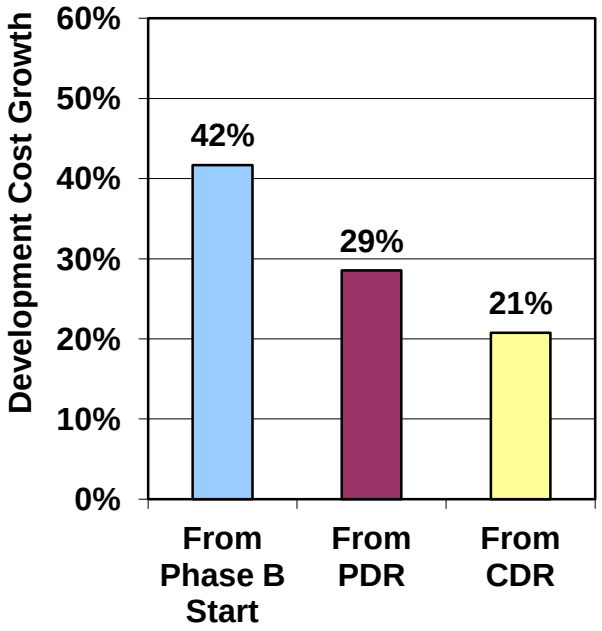
The *Killer Question* For Every Manager Of A Complex, High Risk Program Is …

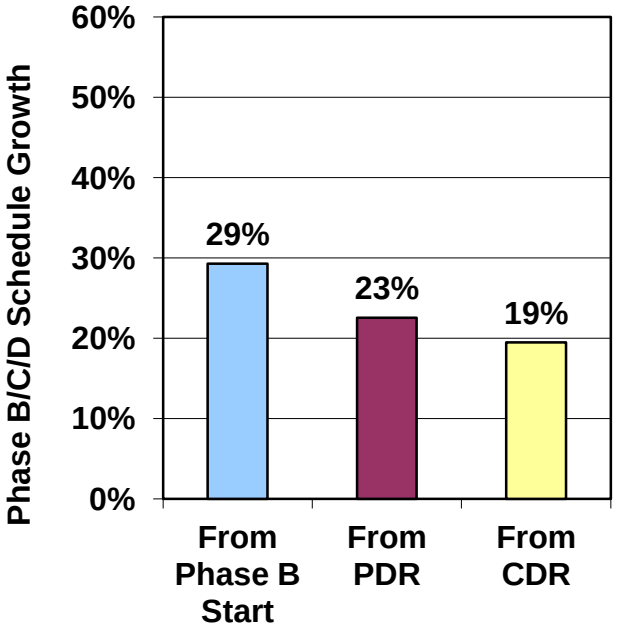
… How Can I See An Unanticipated Estimate At Completion (EAC) Coming Before It's Too Late?

"What's in Your Estimate at Completion?", Pat Barker and Roberta Tomasini, *Defense AT&L*, March-April, 2014

Here's *WHY* We Need Better Ways To Forecast Estimate At Complete …







… the root cause starts on day one, with a less than credible PMB.

Three Types Of Data Are Available In The Big Data Repositories

- ***Descriptive*** – looking in the past we can learn what happened, but it's too late to take corrective action.
- ***Predictive*** – using past performance we can answer the question what will happen if we do nothing but do the same as we've done in the past.
- ***Prescriptive*** – past performance data used to make predictions and suggest decision options to take advantage of the predictions

Prescriptive analytics not only *anticipates* what will happen and when it will happen, but ***why*** it will happen.

Descriptive Analytics

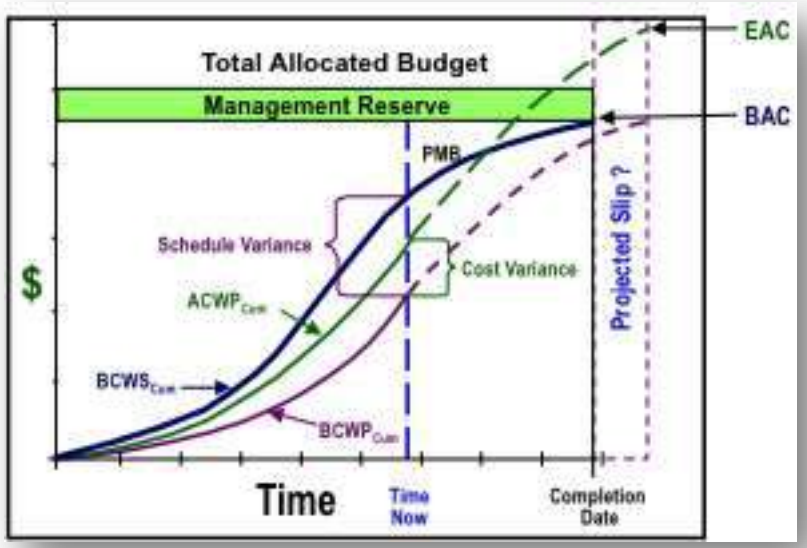
The EVM repositories provide the raw material for Descriptive Analytics through the IPMR (DI-MGMT-81861) submittals

- Descriptive Analytics – condensing big data into smaller, useful nuggets of information.
- Most raw Earned Value data is not suitable for human consumption since it is reported by WBS without the connectivity to the product or programmatic topology
- Descriptive data summarizes what happened in the past, many times 45 days in the past.
- Correlations between WBS elements not defined nor correlations between risk, technical performance or Systems Engineering attributes – MOE, MOP, KPP†

† The Defense Acquisition Guide defines how to apply Measures of Effectiveness, Measures of Performance, Technical Performance Measures, and Key Performance Parameters to assess program performance

DAU Gold Card's EAC Formula Uses *Predictive* Analytics, But …

- Past variances are wiped out with "Cumulative to Date" data
- No adjustment for risk
- Not statistically corrected for past performance



ESTIMATE @ COMPLETION = ACTUALS TO DATE + [(REMAINING WORK) / (PERFORMANCE FACTOR)]

EAC_{CPI} = $ACWP_{CUM}$ + [(BAC – $BCWP_{CUM}$) / CPI_{CUM}]

$EAC_{Composite}$ = $ACWP_{CUM}$ + [(BAC – $BCWP_{CUM}$) / (CPI_{CUM} * SPI_{CUM})]

TO COMPLETE PERFORMANCE INDEX (TCPI) § #

$TCPI_{Target}$ = Work Remaining / Cost Remaining = (BAC – $BCWP_{CUM}$) / (**Target** – $ACWP_{CUM}$)

§ To Determine the TCPI for BAC, LRE, or EAC Substitute TARGET with BAC, LRE, or EAC

To Determine the Contract Level TCPI for EAC, You May Replace BAC with TAB

Prescriptive Analytics

- Is a type of Predictive Analytics
- Used when we need to ***prescribe*** an action so leadership can take the data and act.
- Predictive analytics doesn't predict one future outcome – but Multiple outcomes based on the decision makers actions.
- Prescriptive analytics requires a predictive model with two additional components:
  - Actionable data.
  - Feedback system that tracks the outcome produced by the action taken..

Prescriptive Analytics Is The Foundation For Corrective Actions

- Prescriptive Analytics is about making decisions based on data.
- Prescriptive analytics requires a predictive model with two components:
  - Actionable data
  - Feedback from those actions
- Prescriptive models predict the possible consequences based on different choices of action.

Milestones are *rocks* on the side of the road. The Roman Milestone was a measure back to Rome. You only know that distance after you pass the milestone.

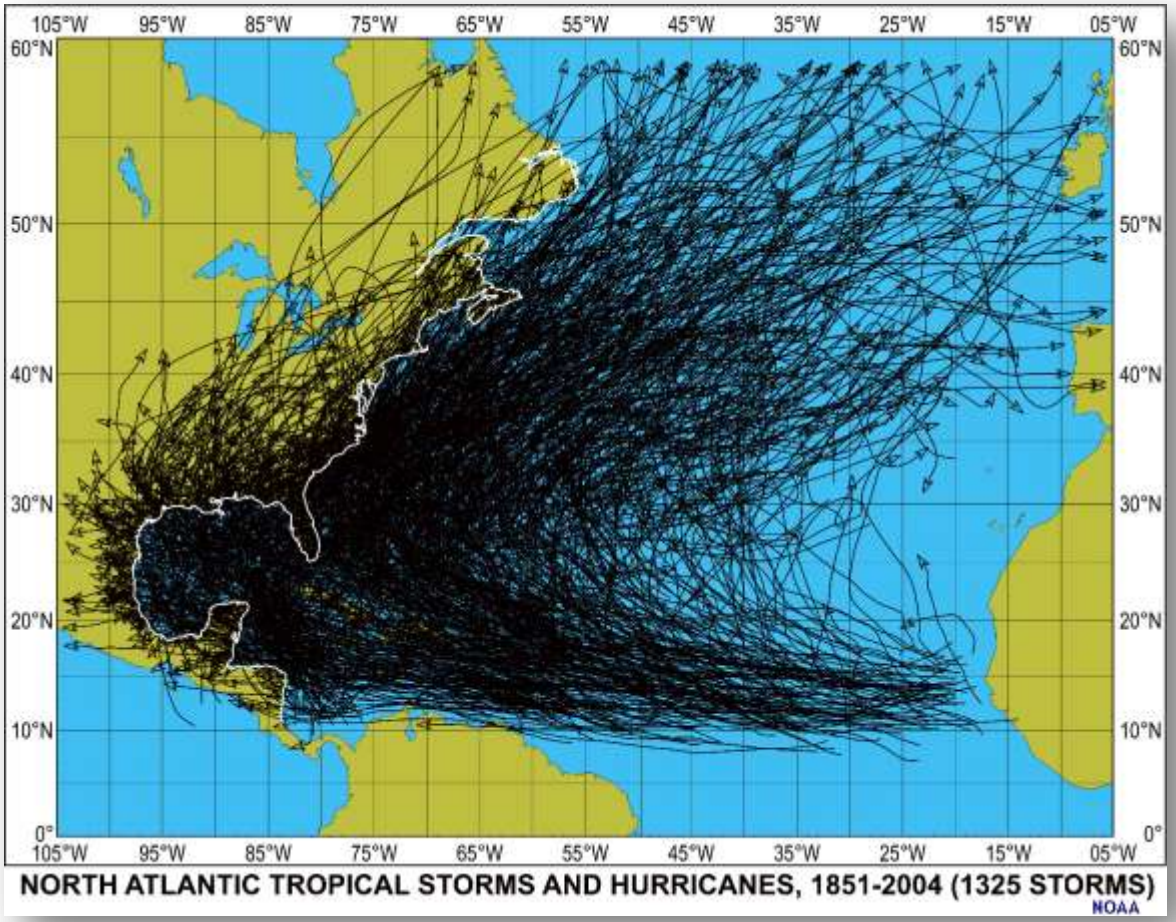
There Is Untapped Value In An Earned Value Data Repository

To extract this value we need to overcome some limitations in today's repositories

- Most data is of little value at the detail level since it is uncorrelated in the reporting process
  - Making correlations between cause and effect is difficult for humans, but statistical process algorithms can do this for us
- With correlated data in hand, we can start generating descriptive analytics
  - But drivers of variance are not visible in the repository
  - Variances from past can be calculated, but not used in future forecasts
- There is no built-in mechanism to see patterns in the data
  - Standard tools produce linear, non-statistical, non-risk adjusted forecasts

All Programmatic Forecasting Is Probabilistic, Driven By Underlying Statistical Processes

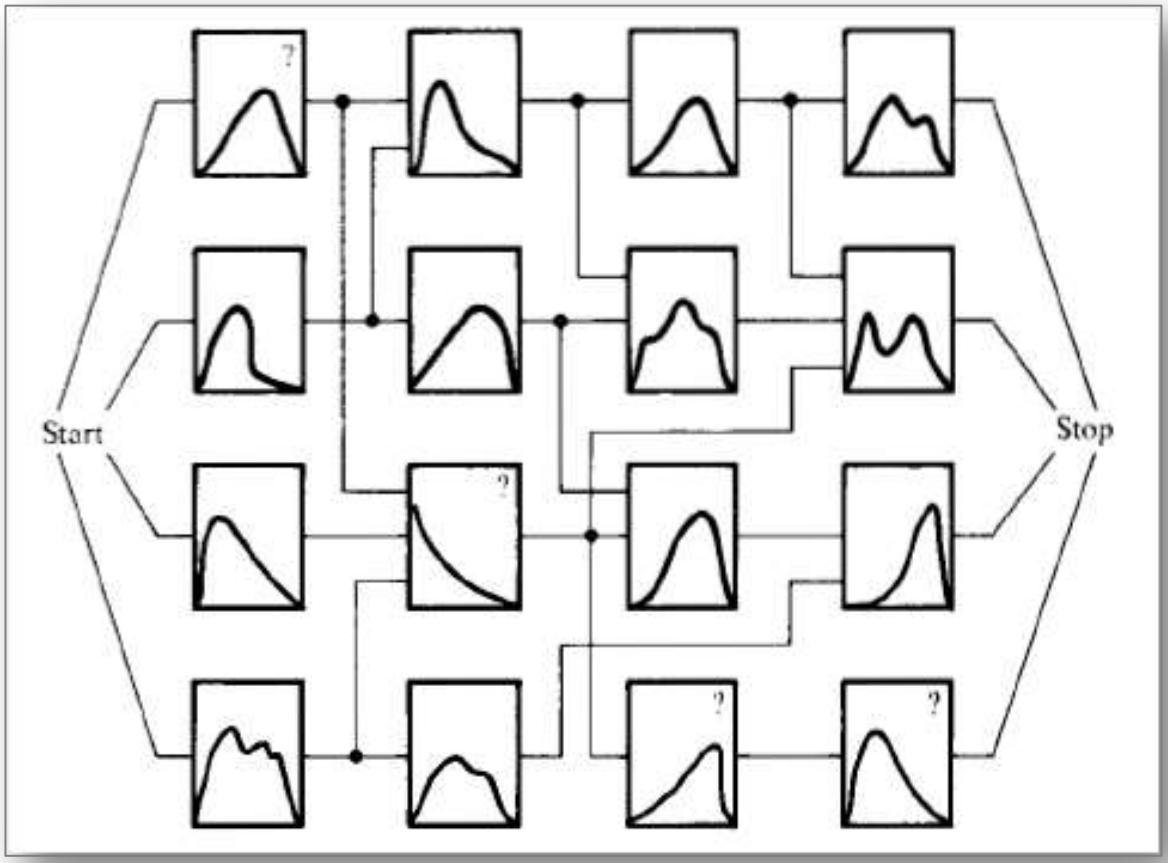
If we make forecasts about program performance that are not statistically and risk adjusted – we're gonna get wet.



Schedule, Related Cost And Technical Elements Are Probabilistic

The IMS doesn't help us much either, since the correlative drivers are themselves non-linear stochastic processes

A Stochastic process is a collection of random variables used to represent the evolution of some random value or system, over time.



The Ability To Forecast Future Performance Starts With A Tool That Provides…

- Forecasting of future performance, using time series of the past using Autoregressive Integrated Moving Average (ARIMA) algorithm
- Confidence intervals of these forecasts for past performance
- Correlation between the time series elements (CPI, SPI, WBS element)
- Deeper correlations between these Earned Value elements as risk retirement, increase effectiveness and performance and any other recorded measure of the program.

http://cran.us.r-project.org/

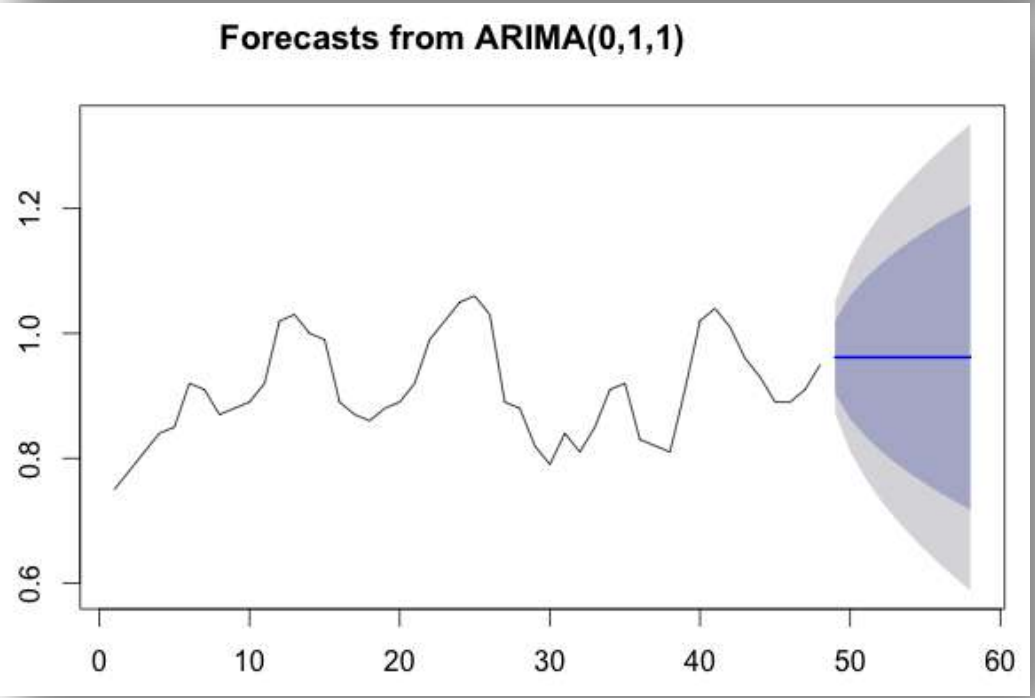
The combination of some data and an aching desire for an answer does not ensure that a reasonable answer can be extracted from a given body of data. - John Tukey

A Quick Look At Where We're Going Starting with Forecasting CPI/SPI

If we want to credibly forecast the future with the past, we'll need better tools. We've got the data, just need to use it

- We have a time series of CPI, SPI, in the repository
- What's possible behaviors in the future can we discover from the past behavior?
- The R code on the top answers that in 4 lines.

```
> CPITS=ts(CPI)
> CPITSARIMA=arima(CPITS, order=c(0,1,1))
> CPITSFORECAST=forecast(CPITSARIMA)
> plot(CPITSFORECAST)
```



WE'VE GOT A DARK SECRET...

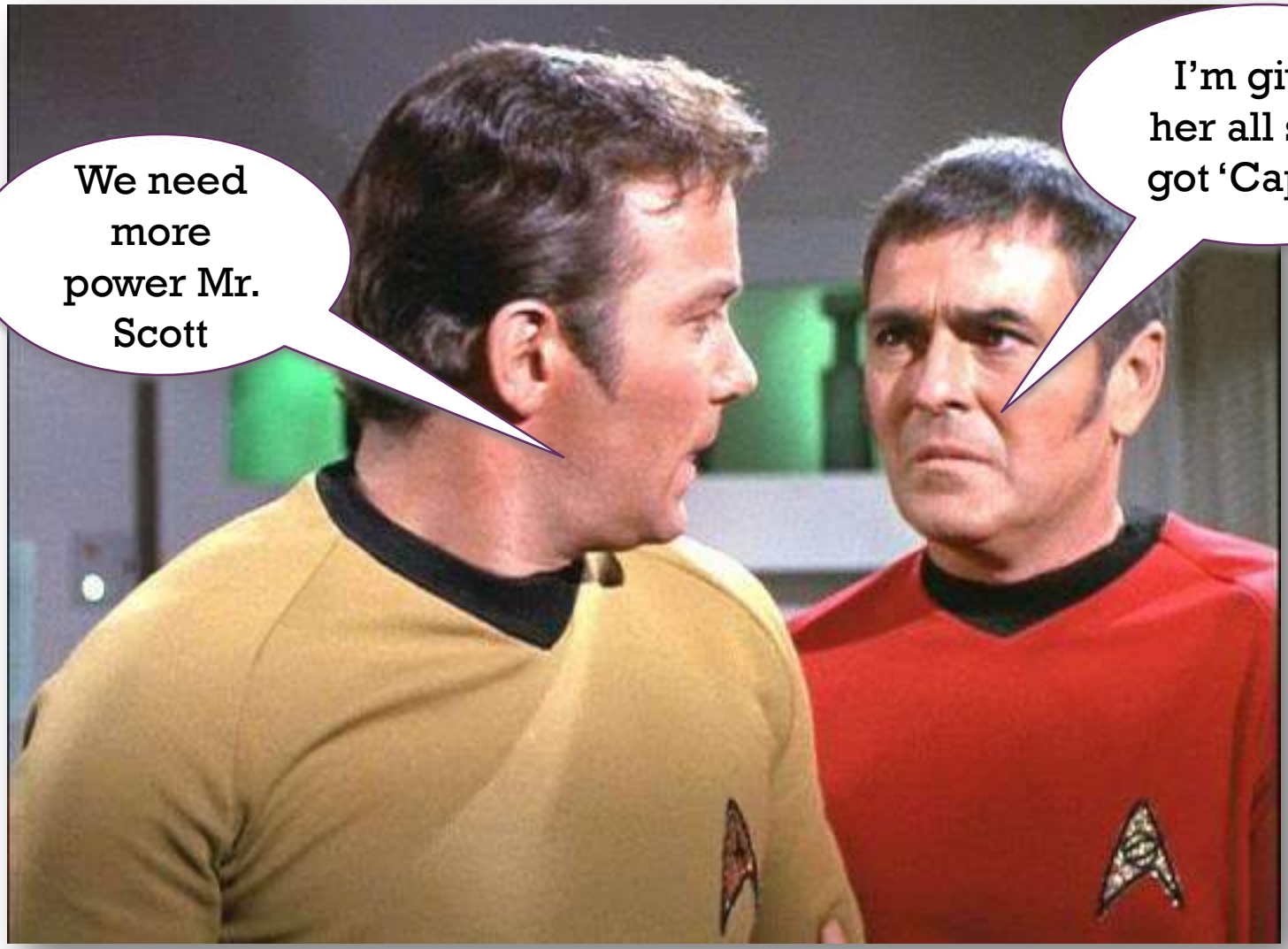
- The Units of Measures for Earned Value Management are Dollars
- Cumulative indices wipe out all the variances
- Forecasts of future performance are not statistically adjusted
- There is no correlative information drivers of variances
- None of these forecasts use the risk register to adjust their value

Since ARIMA Is A Well Traveled Path, We Need More and Better Tools

To provide better forecasts of EAC, we need more data. CPI/SPI needs to be augmented with technical program data

- The Earned Value Management Performance measures need to be connected to:
  - Risk retirement and buy down status Technical Performance Measure compliance
  - Measures of Effectiveness and Measures of Performance
- Work Breakdown Structure correlations for each work activity
  - Correlations between performance and work performed is available in the repository
  - We're missing the tool to reveal these correlations, drivers, and corrective actions to keep the program GREEN

We Need More Power To See Into The Future And Take Corrective Actions



Principal Component Analysis (PCA) Gets *More Power* from our data

Principal component analysis (PCA) is a statistical procedure that uses orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components.

We want to convert a larger set of program performance variables – SPI, CPI, Risk Retirements, TPM, MOE, MOP, KPP, Staffing, and others, into a small set of *drivers of variance*.

PCA can provide visibility to *what are the connections between EAC growth and the source of that growth* in a statistically sound manner, not currently available with IPMR reporting using CPI/SPI

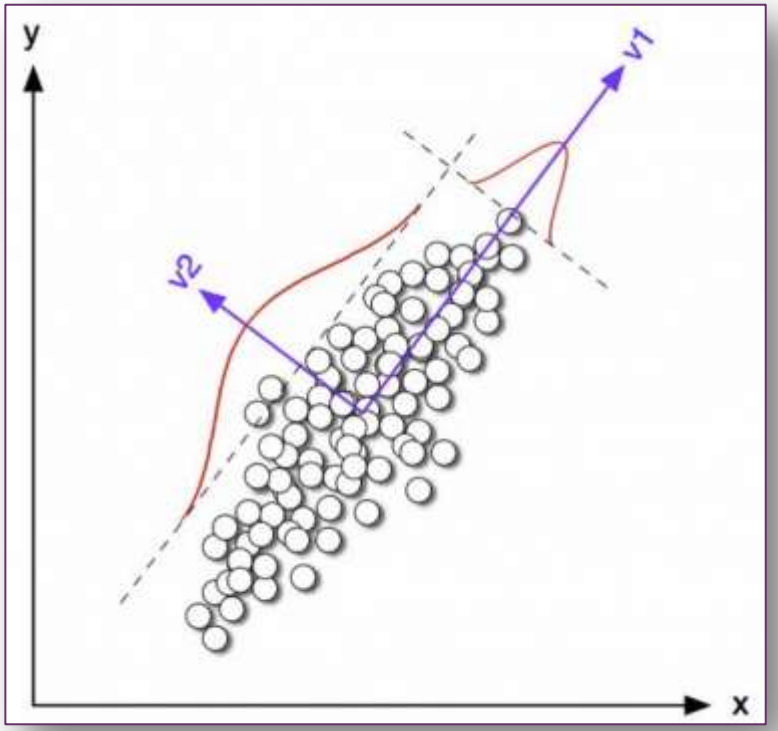
What Can PCA Tell Us?

With "all" the data in a single place – which it is not – we need a way to reduce the dimensionality to provide analysis

- If data lies in high dimensional space (more than just CPI/SPI), then large amount of data is required to learn distributions or decision rules.
- For each WBS element 9 dimensions (CPI, SPI, WBS, TPM, MOE, MOP, KPP, Risk, Staffing Profiles).
- Each dimension has 36 levels (36 months of data).
- We could produce a 9 dimension scatter plot for the 36 months of data and it'd look like a *big blob*.
- We need to know *what are the drivers in this Blob of data?*

From 2 Dimensions (SPI/CPI) to 8 Dimensions and Back Again

- Two components, for example – SPI and CPI
- Discover the correlation between these two data samples
- Locate in the individual samples the time the drivers started impacting the program
- Extend this to 8 dimensions
- Similar to Joint Confidence Level, but with actual data



$$PC_i = a_1X_1 + a_2X_2 + a_3X_3 + \ldots + a_8K_8$$

Program Performance Dimensions

PCA data can be simple 2 dimensional – CPI/SPI or more complex and represent other "attributes" driving EAC

Variable	Information that mat drive Unanticipated EAC
CPI/SPI	CPI for program, time phased by reporting period
TPM	Technical Performance Measures, with control bands as program moves left to right. These can be any measure technical compliance ▪ Weight ▪ Throughput ▪ Information Assurance validation ▪ Any of the JROC KPPs
Risk	Risk ▪ Risk retirement buy down plan ▪ Risk handling and planned reduction
Margin	Cost and schedule margin burn down to plan

Call to Action for increased visibility to Unanticipated EAC Growth using BIG Data

0	Normalize data in the Central Repository in preparation for analysis	EV, WBS, Timeline units normalized with No missing items
1	Apply ARIMA to normalized data to forecast CPI, SPI, and Calculated EAC	Product EAC now based on statistically sound forecast
2	Adjust ARIMA parameters using past performance compliance	Tune ARIMA for program phase sensitivity
3	Integrate external data with EV repository data to build correlations the EAC forecasts	Add risk register and SEMP measures to ARIMA
5	Apply Principal Component Analysis (PCA) to identify correlated *drivers* of EAC growth	Research the Use multivariate forecasting for EAC