Statistical Methods Applied to Project Management

Walt Lipke

PMI - Oklahoma City Chapter

+1 405 364 1594

waltlipke@cox.net

www.earnedschedule.com

Abstract

- An objective of project management is to have the capability to reliably predict cost and schedule outcomes
- The application of statistical methods to the cost and schedule indicators from EVM and ES is a well-founded means for providing the project management objective

Overview

- Forecasting with EVM & ES
- Discussion of Statistical Method
- Application to Real Data
- Analysis & Results
- Summary
- Final Remarks

Forecasting with EVM & ES

- IEAC = BAC / CPI
  - *IEAC = Independent Estimate at Completion*
  - *BAC = Budget at Completion*
  - *CPI = Cost Performance Index*
    = EV / AC
- IEAC(t) = PD / SPI(t)
  - *IEAC(t) = IEAC(time)*
  - *PD = Planned Duration*
  - *SPI(t) = Schedule Performance Index (time)*
    = ES / AT

Forecasting Background

- IEAC & CPI studies by Dr. Christensen et al (1990 – 2004)
  - IEAC = BAC / CPI is Low Bound
  - |CPI(final) – CPI(20%)| ≤ 0.10
  - US DOD Acquisition Data
- IEAC(t) & SPI(t) studies by K. Henderson, Dr. Vanhoucke & S. Vandevoorde (2003 – ….)
  - Henderson & Vandevoorde validated ES concept with real data
  - Using simulation Vanhoucke & Vandevoorde showed ES to be a better schedule predictor than other EVM-based methods

Forecasting Dilemma ….

- Without broad-based data from a variety of EVM & ES applications empirical study is incomplete
- Simulations may not be representative

- Statistical methods are long standing calculation techniques for inferring outcomes

Statistical Method

- Confidence Limits: the range of possible values which encompass the true value of the mean, at a specified level of confidence
- Mathematically

$$CL = Mean \pm Z * \sigma/\sqrt{n}$$

Mean = estimate of average from the sample

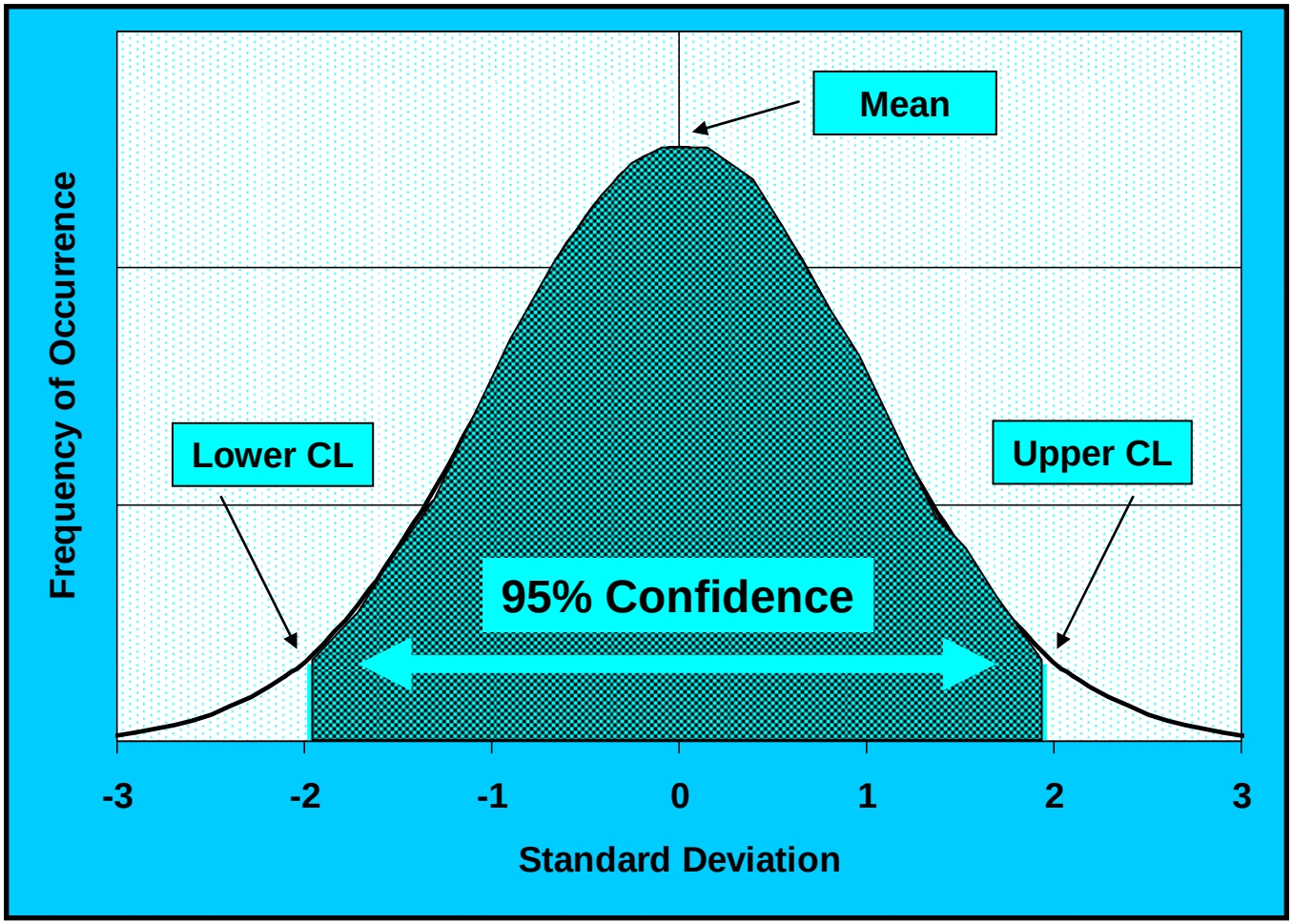
Z = value related to prescribed area within the Normal distribution

[generally 90% or 95% level of confidence]

σ = estimate of the Standard Deviation

n = number of observations in the sample

Confidence Limits



Complexity Elements

- Normality of Data
  - CPI & SPI(t) distributions appear lognormal
  - Mean is logarithm of cumulative value of index
  - $\sigma = \sqrt{(\Sigma(\text{ln period index(i)} - \text{ln cum index})^2 / (n - 1))}$
- Finite Population
  - $AF_C = \sqrt{((BAC - EV) / (BAC - (EV/n)))}$
  - $AF_S = \sqrt{((PD - ES) / (PD - (ES/n)))}$
- Fewer than 30 Observations
  - Use Student-t Distribution

Use of Confidence Limits

- Intent is to show that Confidence Limits are reliable forecasts of bounds for cost and schedule outcomes
- $CL_{(\pm)}$ = ln index(cum) ± Z ∗ ($\sigma/\sqrt{n}$) ∗ AF
- Forecast at Completion
  - $IEAC_{(\text{low or high})} = BAC / EXP(CL_{(\pm)})$
  - $IEAC(t)_{(\text{low or high})} = PD / EXP(CL(t)_{(\pm)})$

Study Method

- IEACs are iteratively computed for each newly added observation
- Upper and Lower Confidence Limits are tested using the statistical hypothesis test, *Sign Test*, at 0.05 significance
  - Final Cost < $IEAC_H$
  - Final Cost > $IEAC_L$
  - Final Duration < $IEAC(t)_H$
  - Final Duration > $IEAC(t)_L$

Study Method

- Desired test result is the alternative hypothesis, Ha (shown on previous chart)
- Test results are tabulated as Ha when value of test statistic is in the critical region (0.05) – and Ho when it is not
- From the Ha results for the projects, the probability of obtaining reliable results is computed

Study Method

- Testing is conducted for various confidence levels and data sets
  - Confidence Levels: 90%, 95%, 98%
  - Data Set: 10-100%, 30-100%, 60-100%
- Expectation: as Conf Level & Data% increase, reliability of forecast increases
- By combining confidence levels and data sets a generally reliable project cost and duration forecasting is sought

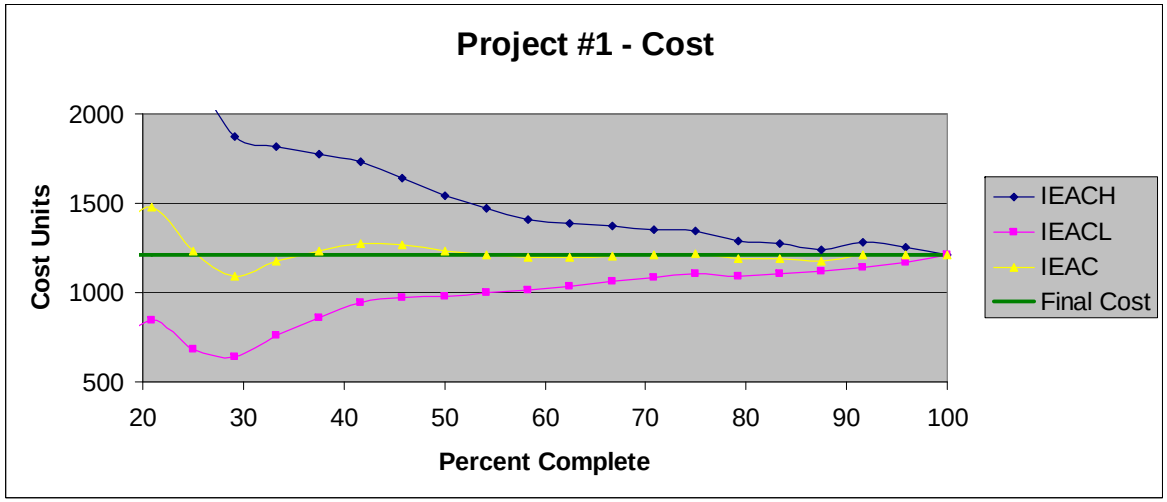
Real Data - Characterized

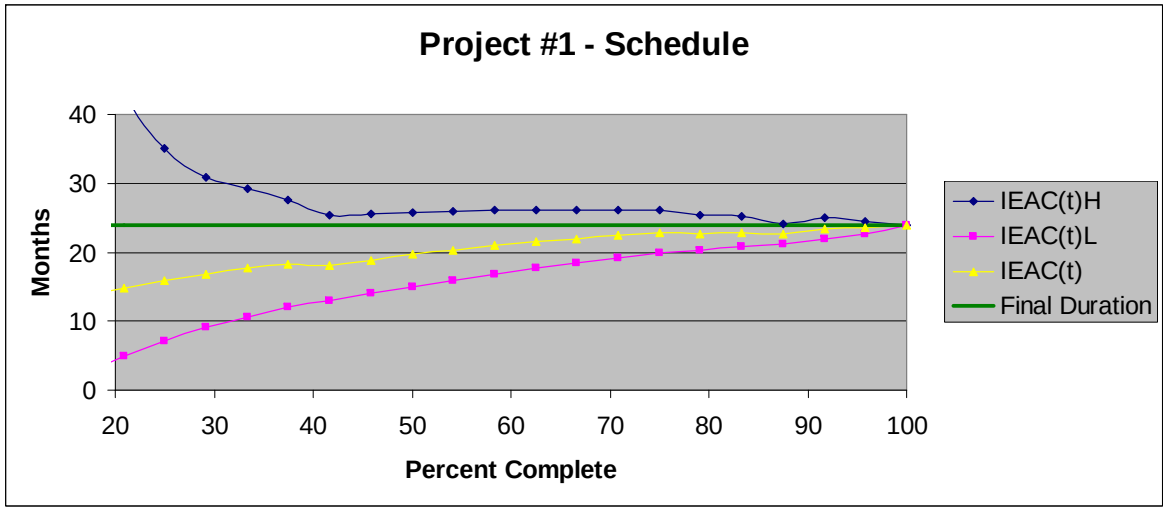
- Twelve projects – low risk, high technology products
  - 497 months of EVM data
  - No re-plans
  - Data from single MIS under one manager
  - Cost range: $291K - $6.08M
  - Duration range: 17 – 50 months
  - CPIcum range: 0.481 – 1.051
  - SPI(t)cum range: 0.739 – 1.000
  - With one exception, SPI(t) > CPI

Real Data - Observations

- Cost & schedule standard deviations are comparable
  - Variation greater than seen previously
- Change in index values greater than expected
  - Four projects had changes greater than 0.10 between 80 and 100 percent complete
  - Seven had changes greater than 0.05
  - *Not supportive of Christensen CPI stability*

Forecast Result (90% Confidence)





Project #1 Observations

- Difference between upper & lower CLs becoming smaller as percent complete increases
- CPI is very stable between 50 and 100%
- SPI(t) consistently worsens
  - $IEAC(t)_H$ beginning at 30% complete proved to be very close to the eventual final duration

Test Result – One Scenario

Hypothesis Test Results @ 98% Confidence ≥10% Complete													
	***** Project Number *****												
Bounds	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**Probability**
Cost High	Ha	Ha	Ho	Ha	Ho	Ha	Ho	Ha	Ha	Ho	Ha	Ha	0.927
	0.000	0.000	0.500	0.044	0.500	0.000	0.844	0.000	0.000	0.116	0.000	0.000	
Cost Low	Ha	Ho	Ha	Ha	Ha	Ha	Ha	Ha	Ha	Ha	Ha	Ha	1.000
	0.000	0.804	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
Schedule High	Ha	Ha	Ha	Ha	Ha	Ha	Ha	Ha	Ha	Ha	Ho	Ha	1.000
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.000	0.000	0.132	0.000	
Schedule Low	Ha	Ha	Ho	Ha	Ha	Ha	Ha	Ha	Ha	Ha	Ha	Ho	0.997
	0.000	0.000	0.791	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	
									Composite Probability				1.000

Test Result Observations

- Ha & Ho are the alternate and null hypothesis test results for each project
- The numbers beneath Ha & Ho are the computed values for the test statistic
- The probability results for both cost & schedule forecasts indicate high reliability throughout the entire period

Compiled Test Results

Prediction Probability						
Bounds		*** 90% Confidence ***			95% Confidence	98% Confidence
		≥ 10% Complete	≥ 30% Complete	≥ 60% Complete	≥ 10% Complete	≥ 10% Complete
Cost	High	0.613	0.613	0.927	0.613	0.927
	Low	1.000	1.000	0.981	1.000	1.000
Schedule	High	1.000	1.000	1.000	1.000	1.000
	Low	0.997	0.981	0.997	0.997	0.997

Compiled Test Analysis

- In general, expectation realized
  - As confidence level increases, probability of obtaining Ha increases
  - As data is restricted nearer to the project completion, the probability of obtaining Ha increases
- Safest forecast regardless of data set is 98% confidence level
- Trade-off: the larger the confidence, the greater the likelihood of overstating the upper and lower limits

Compiled Test Analysis

- Reliable forecasts are seen for the 90% confidence at 60% complete scenario
  - Compares favorably to previous work, where it was determined that 60% complete is the generalized stability point for the CPI
  - Adds credence to assertion that as index becomes more stable a lower confidence level can be applied with the expectation of obtaining reliable forecasts

Compiled Test Analysis

- Recall comparison of final values of cost and schedule indexes: SPI(t) > CPI
  - Achieving schedule likely has priority
  - Focus on schedule possibly caused costs to be skewed high
- The tendency toward high cost could explain generally lower probability values for $IEAC_H$
- Application of 90% confidence level at 10% complete conjectured to be generally reliable

Summary

- Statistical forecasting of high and low outcomes tested for reliability
  - Confidence Levels: 90%, 95%, 98%
  - Data Sets: 10%, 30%, 60%
- Generally, greater reliability the higher the confidence level and the larger the percent complete
  - Schedule forecast more reliable than for cost
- Due to unique characteristics of data tested, 90% confidence postulated to be appropriate for most circumstances

Final Remarks

- The method put forth is generally applicable and encouraged – independent of size or type of project
- The statistical method has the potential to greatly enhance management information for the purpose of project control
- Tool for trialing available at the calculators page of the Earned Schedule website *(Statistical Prediction Calculator)*

References

- "Statistical Methods Applied to EVM: The Next Frontier," *CrossTalk*, June 2006: 20-23
- "Prediction of Project Outcome: The Application of Statistical Methods to Earned Value Management and Earned Schedule Performance Indexes," *International Journal of Project Management*, February 2009 (pending)
- Earned Schedule Website: www.earnedschedule.com