

32nd Annual Department of Defense Cost Analysis Symposium

Do Not Use RANK CORRELATION in Cost Risk Analysis

Paul R. Garvey

Chief Scientist

The Economic and Decision Analysis Center

The MITRE Corporation

Main Points

- Correlation exists between work breakdown structure (WBS) cost element costs and between cost and schedule*
- Correlation is a necessary consideration in cost risk analyses; however, subtleties associated with correlation must be well understood to avoid an improperly specified risk model
- Two measures of correlation are commonly used in cost risk analyses; they are Pearson's *product-moment correlation* coefficient and Spearman's *rank correlation* coefficient
- From a work breakdown structure perspective, Pearson's product-moment correlation coefficient is the only appropriate measure of correlation for cost risk analyses

* http://www.acq.osd.mil/te/programs/se/risk_management/papers_speeches_briefs/index.htm

Capturing Correlation

A Necessary Consideration

WBS Cost Element (CE)

Prime Mission Equipment (PME) Segment

CE1 Hardware (HW)

$$HW_{Cost} = \sum_i HW_{i_{Cost}}$$

CE2 Software (SW)

$$SW_{Cost} = 8.2 I_{SW} (Size) e$$

CE3 Integration & Assembly (I&A)

$$I \& A_{Cost} = Sched_{I\&A} (I_{I\&A})(Staff_{I\&A})$$

← Labor rate for SW Development Effort
 ← Staff-level for I&A
 ← Schedule for I&A

Associated System Costs

CE4 Systems Engineering (SE)

$$SE_{Cost} = PrgmSched (I_{SE})(Staff_{SE})$$

CE5 Program Management (PM)

$$PM_{Cost} = PrgmSched (I_{PM})(Staff_{PM})$$

CE6 System Test & Evaluation (STE)

$$STE_{Cost} = PrgmSched (I_{STE})(Staff_{STE})$$

•
•
•

CEm Training

$$Training_{Cost} = f_{Training} (HW_{Cost} + SW_{Cost})$$

← Overall Program Schedule
 ← Cost factor for Training (e.g., 0.02)

•
•
•

Capturing Correlation

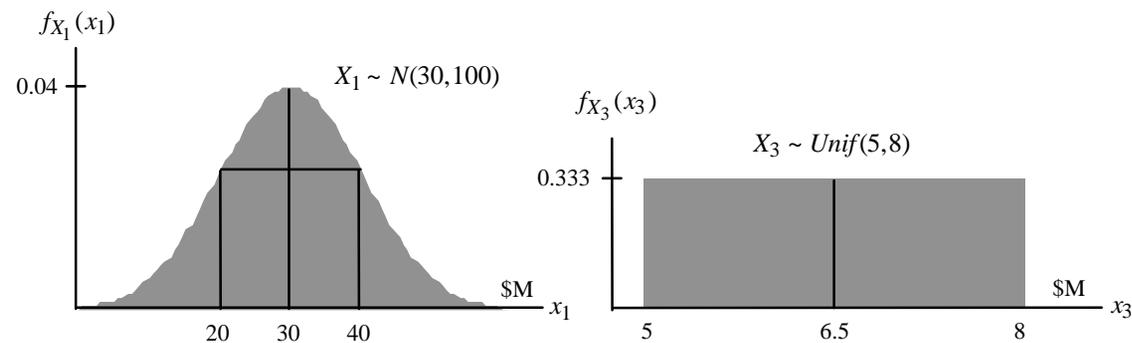
A Necessary Consideration

- In modeling cost risk, correlation is a necessary consideration...why?
- Correlation can exist in a WBS between cost element costs, as well as between the cost of a cost element and the variables that define it (e.g., weight, schedule)
- For example, in the WBS associated with the preceding chart
 - A perfect linear correlation exists between the cost of the cost element *Training* and the cost given by the sum of the cost element costs for *Hardware (HW)* and *Software (SW)*
 - A cost-schedule correlation exists between the first three associated system costs and the program's schedule

Capturing Correlation

A Necessary Consideration

Example 5-9 Suppose the total cost of a system is given by $Cost = X_1 + X_2 + X_3$. Let X_1 denote the cost of the system's prime mission product — PMP. Let X_2 denote the cost of the system's systems engineering, program management, and system test. Suppose X_1 and X_2 are *dependent* random variables and $X_2 = \frac{1}{2}X_1$. Let X_3 denote the cost of the system's data, spare parts, and support equipment. Suppose X_1 and X_3 are *independent* random variables with distribution functions given below.



Capturing Correlation

A Necessary Consideration

The intervals below contains approximately 95 percent of the probability under a normal PDF. For PDF A (refer to the figure) this interval is

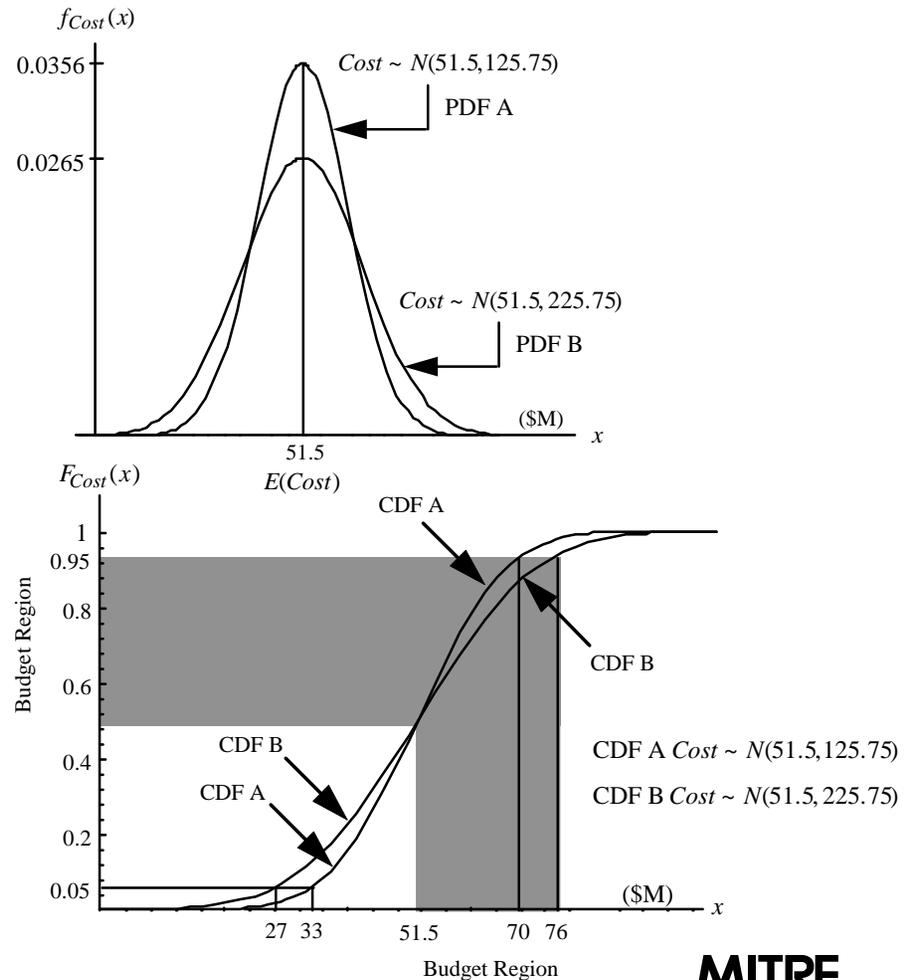
[29.08 - 73.92]; range = 44.84 (\$M)

For PDF B, (refer to the figure) which accounts for correlation, this interval is

[21.46 - 81.54]; range = 60.08 (\$M)

This is a 34 percent increase (\$15.24M) in the “95-percent” dollar range when compared to the same interval for PDF A.

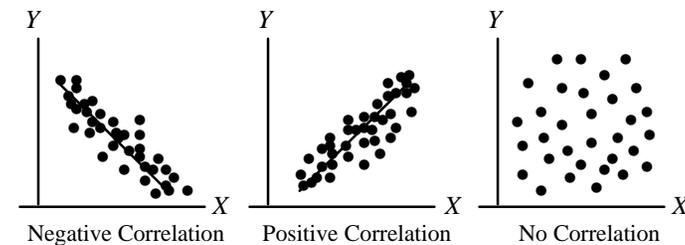
Ignoring correlation in cost uncertainty analysis is a common “analyst-omission”. The omission can be serious. Illustrated in the figure, failing to capture positive correlation in the analysis leads decision-makers to falsely conclude the variability (uncertainty) in a system’s cost is less than it “truly” is.



Correlation

Is Easily Misinterpreted!

- Although statistical theory provides a number of ways to measure correlation, two common measures are Pearson's product-moment correlation and Spearman's rank correlation
- Pearson's product-moment correlation measures the linearity between two random variables; if two random variables are *perfectly linear* (e.g., $Y = aX + b$) with positive slope, then the Pearson product-moment correlation between X and Y , denoted by $\rho_{X,Y}$, is +1
- Spearman's rank correlation measures the monotonicity between two random variables; if two random variables are perfectly *monotonically increasing* (e.g., $Y = X^2$), then the Spearman rank correlation, denoted by $r_{X,Y}$, is +1



Correlation

Sums of Random Variables

- Sums of random variables involve *only* Pearson's product-moment correlation *not Spearman's rank correlation*; why?
- You can see this from theory...consider the simple case of *Cost* equal to the sum of two WBS cost element costs, say X_1 and X_2 ; then

$$Cost = X_1 + X_2$$

$$E(Cost) = E(X_1 + X_2) = E(X_1) + E(X_2)$$

$$\begin{aligned} Var(Cost) &= E(Cost - E(Cost))^2 = E(Cost^2) - (E(Cost))^2 \\ &= E(X_1^2 + 2X_1X_2 + X_2^2) - [E(X_1) + E(X_2)]^2 \\ &= E(X_1^2) + 2E(X_1X_2) + E(X_2^2) - [(E(X_1))^2 + 2E(X_1)E(X_2) + (E(X_2))^2] \\ &= E(X_1^2) - (E(X_1))^2 + E(X_2^2) - (E(X_2))^2 + 2(E(X_1X_2) - E(X_1)E(X_2)) \\ &= Var(X_1) + Var(X_2) + 2Cov(X_1X_2) \\ &= Var(X_1) + Var(X_2) + 2\mathbf{r}_{X_1, X_2} \mathbf{s}_{X_1} \mathbf{s}_{X_2} \end{aligned}$$

↑ This is Pearson's product-moment correlation coefficient
not Spearman's rank correlation coefficient!

Correlation

Pearson's and Spearman's Correlation Coefficients Can be Very Different!

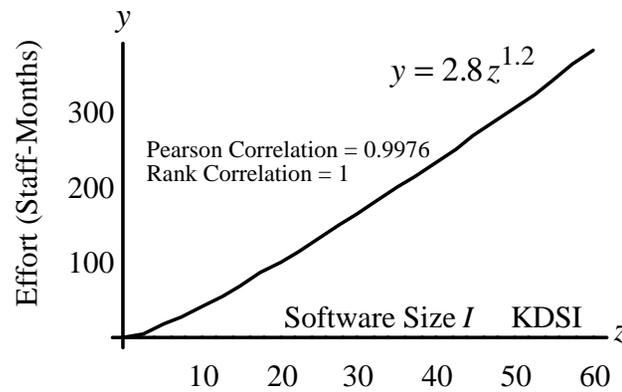


Figure 5-9A

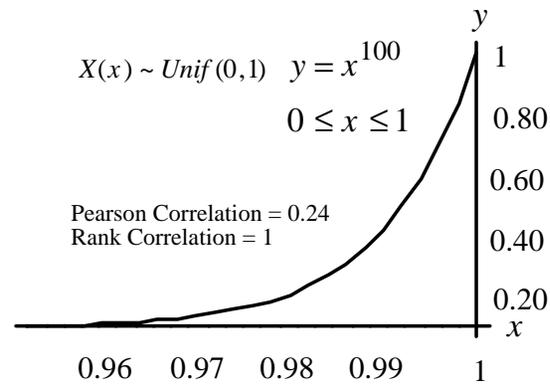


Figure 5-9B

Correlation

Avoid *Explicitly* Dealing with Either Correlation Coefficient

- Avoid “dealing with” either measure explicitly!!
 - Some reasons...
 - The variance of a sum of WBS cost elements is a function of Pearson’s product-moment correlation measure...not rank correlation! So, what does this sum mean if the cost elements were given a specified rank correlation?...the answer is not clear!
 - A less experienced analyst may over-specify the correlation “structure” in a WBS cost-risk model...Pearson correlations may “naturally” exist in the WBS by virtue of the numerous functional relationships typically defined between cost equations and cost elements...analysts may inadvertently specify rank correlations on top of existing Pearson correlations...a mix of rank and Pearson correlations can result!
 - Explicitly specifying Pearson correlations between pairs of cost elements may result in infeasible correlations...Certain random variables have correlation bounds...

$$\begin{array}{l}
 X_1 \sim \text{Normal}(\mathbf{m}_1, \mathbf{s}_1^2) \\
 X_2 \sim \text{LogNormal}(\mathbf{m}_2, \mathbf{s}_2^2)
 \end{array}
 \quad
 -1 < -\frac{\mathbf{s}_2}{\sqrt{e^{\mathbf{s}_2^2} - 1}} < \mathbf{r}_{X_1, X_2} < \frac{\mathbf{s}_2}{\sqrt{e^{\mathbf{s}_2^2} - 1}} < 1$$

Correlation

Summary and Recommendation

- Recommendation
 - *Functionally* relate your cost equations in the WBS model that addresses the “relatedness” of the variables...then let the mathematics (or simulation tool) run its course!...your time as an analyst is best spent doing this...it is also easier to brief functional relationships than to brief absolute correlation measures
 - Crystal Ball and @Risk use *rank correlation*. Rank correlation is easier to simulate than Pearson correlation; however, as we’ve seen rank correlation is not appropriate for cost risk analyses
 - If Pearson correlation coefficients need to be explicitly specified in a cost risk analysis, then the Lurie-Goldberg (IDA) algorithm is recommended...“An Approximate Method for Sampling Correlated Random Variables from Partially-Specified Distributions,” *Management Science*, Vol 44, No. 2, February 1998, pp 203-218.