Why Correlation Matters in Cost Estimating

Stephen A. Book
The Aerospace Corporation
P.O. Box 92957
Los Angeles, CA 90009-29597
(310) 336-8655
stephen.a.book@aero.org

32nd Annual DoD Cost Analysis Symposium
Williamburg, VA

2-5 February 1999

Contents

Space-System Cost Elements

0.0	TOTAL BRILLIANT EYES PROGRAM
1.0	SPACE BRILLIANT EYES SYSTEM
1.1	System-Level Costs
1.2	Space Vehicle (SV) Segment
1.2.1	SV Program Level
1.2.2	Space Vehicle Prime Mission Equipment
1.2.2.1	Space Software
1.2.2.2	Space Vehicle
1.2.2.2.1	Space Vehicle IA&T
1.2.2.2.2	Sensor Payload
1.2.2.2.3	Insertion Vehicle
1.2.2.2.4	Survivability
1.2.3	Prototype Lot
1.2.4	Spare Parts
1.2.5	Technology and Producibility
1.2.6	Aerospace Ground Equipment
1.2.7	Launch Support
1.3	Engineering Change Orders (ECOs)
1.4	Other Government Costs
1.5	Risk

2.0	GROUND BRILLIANT EYES SYSTEM
2.1	Ground System IA&T
2.2	Ground Segment
2.2.1	Ground Segment Program Level
2.2.2	Ground Software
2.2.3	Ground Control
2.3	Ground Military Construction
2.4	Ground Operations and Support
2.5	Ground ECO
2.6	Ground Other Government Costs
2.7	Ground Risk
3.0	LAUNCH BRILLIANT EYES SYSTEM
3.1	Launch System IA&T
3.2	Launch Segment
3.2.1	Launch Segment Program Level
3.2.2	Launch Facility
3.2.3	LV Production Facility
3.2.4	Launch Vehicle
3.2.5	Launch Vehicle Integration
3.2.6	Launch Operations
3.3	Launch ECO
3.4	Launch Other Government Costs
3.5	Launch Risk

Contents

- **A Cost Analysis is a Risk Analysis**
- **Measuring Uncertainty in Cost**
- **What is Correlation?**
- **Correlation Impacts Uncertainty**
- **Correlation Also Impacts the Estimate!**
- **Summary**

Typical Cost-Estimating “Roll-Up” Procedure

- List Cost Elements in Work-Breakdown Structure (WBS)
- Calculate “Single Best Estimate” of Cost for Each WBS Element
- Sum All Single Best Estimates
- Define Result to be “Single Best Estimate” of Total-System Cost

What Does “Best” Estimate Signify?

- “Most Likely” Cost? (“Mode”)
- 50th-Percentile Cost? (“Median”)
- Average Cost? (“Mean”)
- These Three are Almost Always Different

Cost-Element Probability Distributions

- Mean, Median, Mode are Statistical Characteristics of Probability Distributions
- Use of These Terms Implicitly Assumes that Costs Have Probability Distributions
- Indeed, Even Admission that “Best” Estimate is not the Only Possible Estimate Implicitly Assumes That Costs Have Probability Distributions

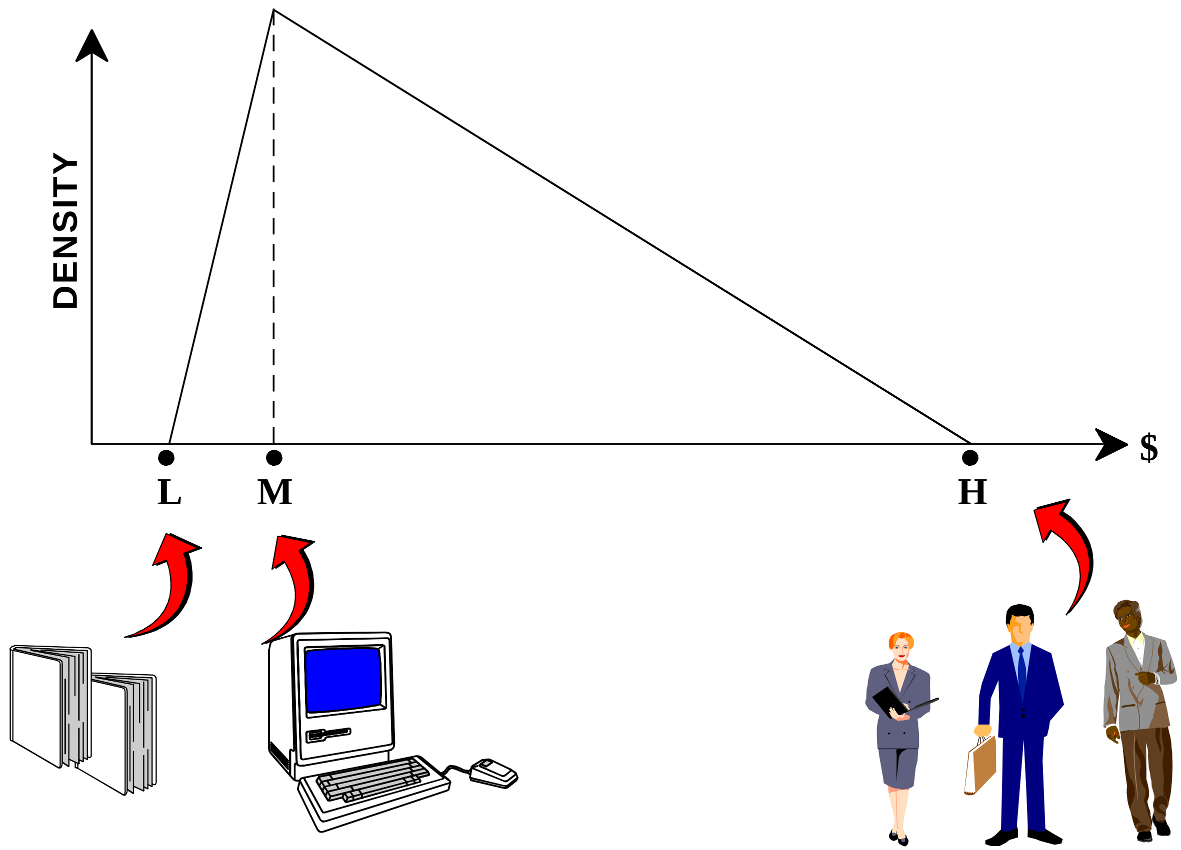
Cost-Risk Analysis

- **Definition of "Cost Risk"**
  - **Inadequacy of Forecasted Funding Requirements to Assure that Program Can Be Completed and Meet Its Stated Objectives**
- **Element Costs Are Uncertain Quantities (i.e., Random Variables) That Have "Probability" Distributions**
  - **Combine Element Cost Distributions to Generate Cumulative Distribution of Total System Cost**
  - **Read off 70th Percentile Cost, 90th Percentile Cost, etc., From Cumulative Distribution to Estimate Amount of Extra Dollars Needed to Cover Risk**
  - **Quantify Confidence in Ability to Complete Program Funded at the "Best" (or any other) Estimate of System Cost**

Contents

- A Cost Analysis is a Risk Analysis
- **Measuring Uncertainty in Cost**
- What is Correlation?
- Correlation Impacts Uncertainty
- Correlation Also Impacts the Estimate!
- Summary

Triangular Distribution of Element Cost

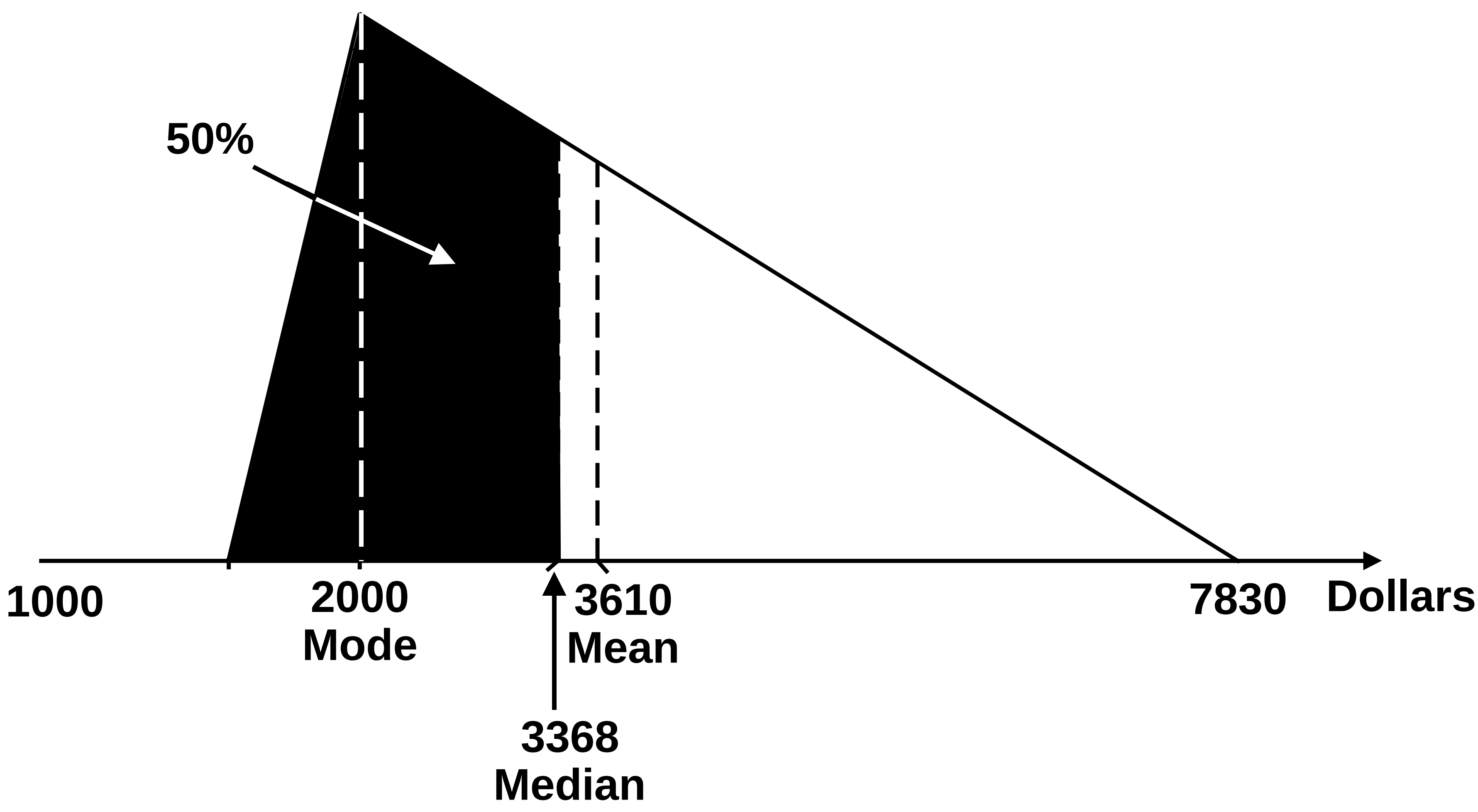


Optimistic Cost

Best-Estimate Cost (Mode)

Cost Implication of Technical Assessment

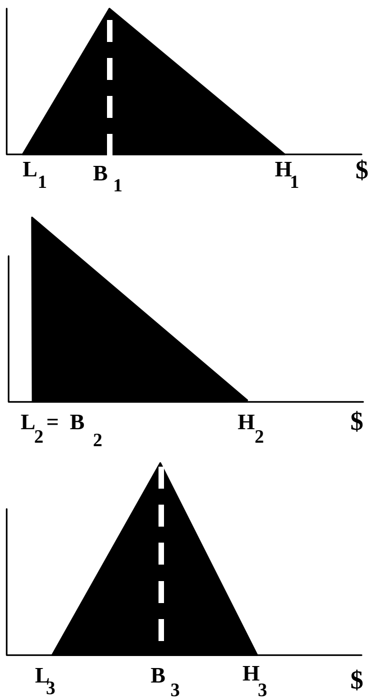
Triangular Distribution of WBS-Element Cost



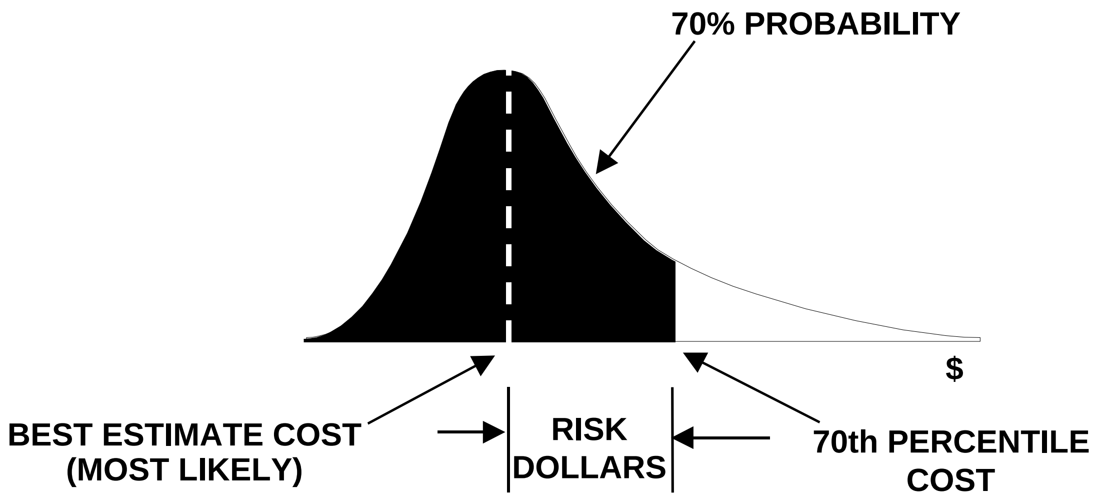
THE AEROSPACE CORPORATION

Cost-Risk Procedure

COST-ELEMENT TRIANGULAR DISTRIBUTIONS



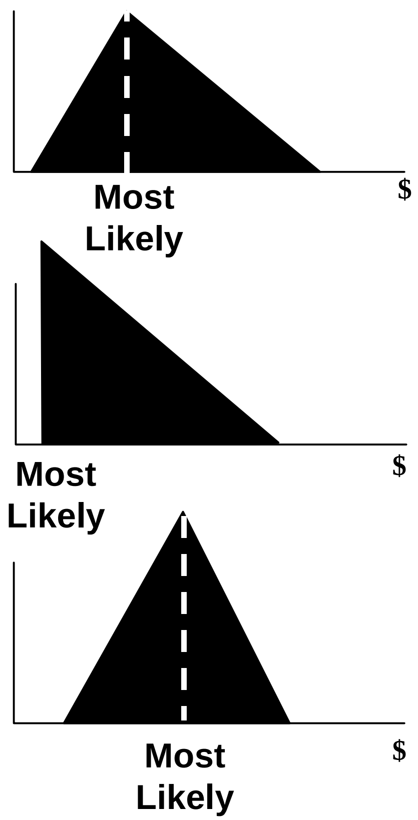
MERGE ELEMENT COST DISTRIBUTIONS INTO TOTAL-COST DISTRIBUTION



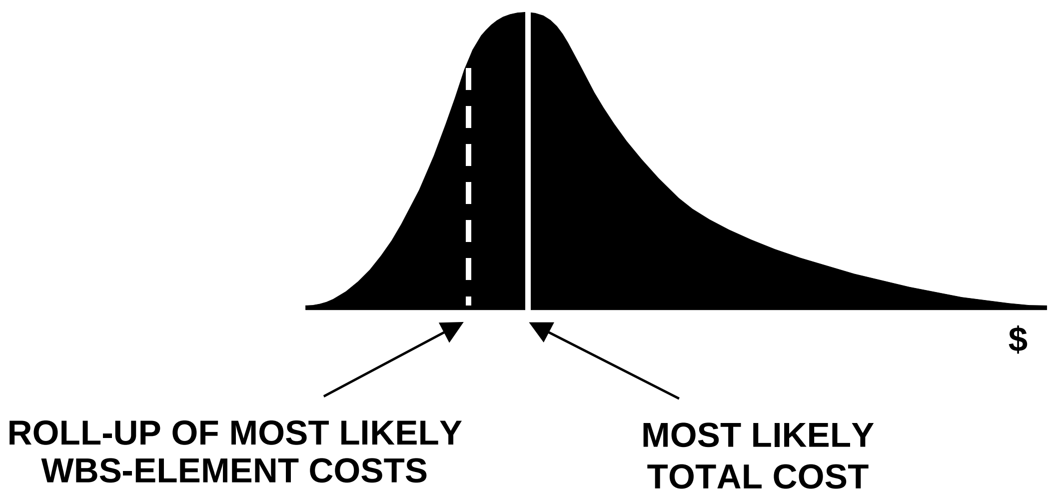
Note: Addition of risk dollars brings confidence that total appropriation (best estimate plus risk dollars) is sufficient to fund program.

When WBS Elements Are Few

WBS-ELEMENT TRIANGULAR COST DISTRIBUTIONS

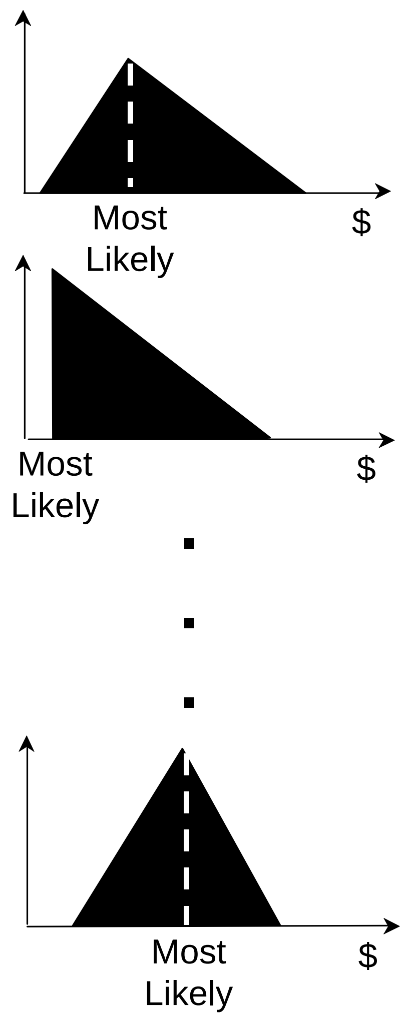


MERGE WBS-ELEMENT COST DISTRIBUTIONS INTO TOTAL-COST LOGNORMAL DISTRIBUTION

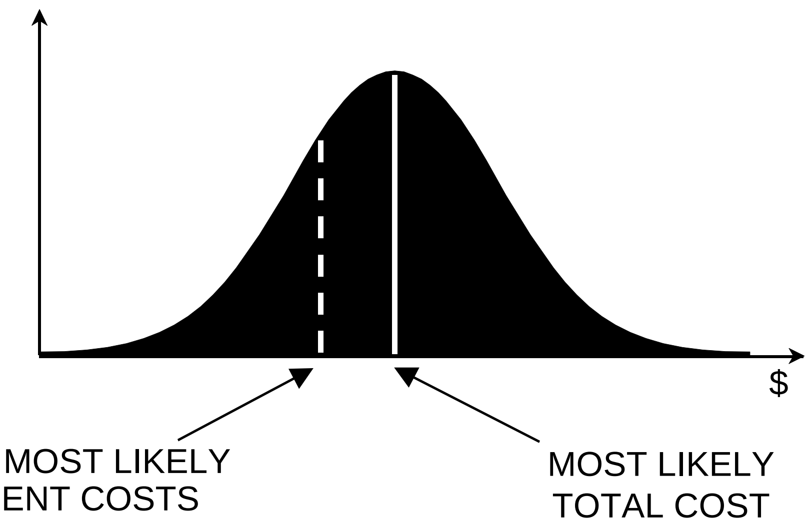


When WBS Elements Are Many

WBS-ELEMENT TRIANGULAR COST DISTRIBUTIONS



MERGE WBS-ELEMENT COST DISTRIBUTIONS INTO TOTAL-COST NORMAL DISTRIBUTION



Contents

- A Cost Analysis is a Risk Analysis
- Measuring Uncertainty in Cost
- **What is Correlation?**
- Correlation Impacts Uncertainty
- Correlation Also Impacts the Estimate!
- Summary

Pearson "Product-Moment" Correlation

- **Suppose X and Y are Two Random Variables**
  - $m_X = E(X),\ \ m_Y = E(Y)$ **are their Expected Values ("Means")**
  - **True Theorem:** $E(X+Y) = E(X) + E(Y) = m_X + m_Y$
  - **False Theorem:** $E(XY) = E(X)E(Y)$
  - $Cov(X,Y) = E(XY) - E(X)E(Y) =$ **"Covariance" of X and Y**

- $Var(X) = Cov(X,X) = E\left(X^2\right) - \left[E(X)\right]^2 =$ **"Variance" of X**

  $Var(Y) = Cov(Y,Y) = E\left(Y^2\right) - \left[E(Y)\right]^2 =$ **"Variance" of Y**

- $Corr(X,Y) = \dfrac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} =$ **"Correlation" of X and Y =** r_{XY}

  $Var(X) = s_X^2,\ \ Var(Y) = s_Y^2,$

  $Cov(X,Y) = r_{XY}\, s_X\, s_Y$

Variance of a Sum

- $Var(X+Y) = E\left([X+Y]^2\right) - [E(X+Y)]^2$

$$= E\left(X^2 + 2XY + Y^2\right) - [E(X) + E(Y)]^2$$

$$= E\left(X^2\right) + 2E(XY) + E(Y)^2 - [E(X)]^2 - 2E(X)E(Y) - [E(Y)]^2$$

$$= E\left(X^2\right) - [E(X)]^2 + E\left(Y^2\right) - [E(Y)]^2 + 2[E(XY) - E(X)E(Y)]$$

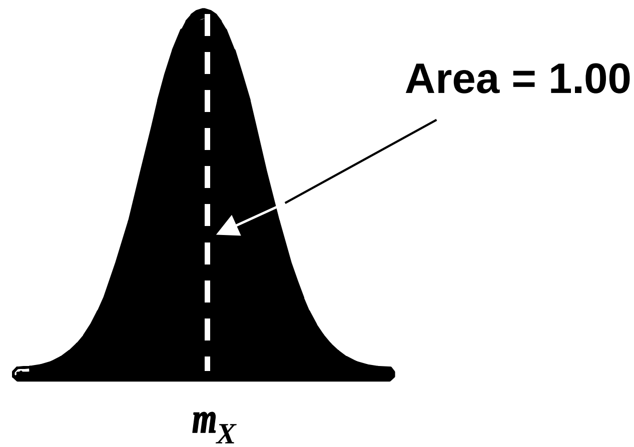
$$= Var(X) + Var(Y) + 2Cov(X,Y)$$

$$= Var(X) + Var(Y) + 2Corr(X,Y)\sqrt{Var(X)Var(Y)}$$

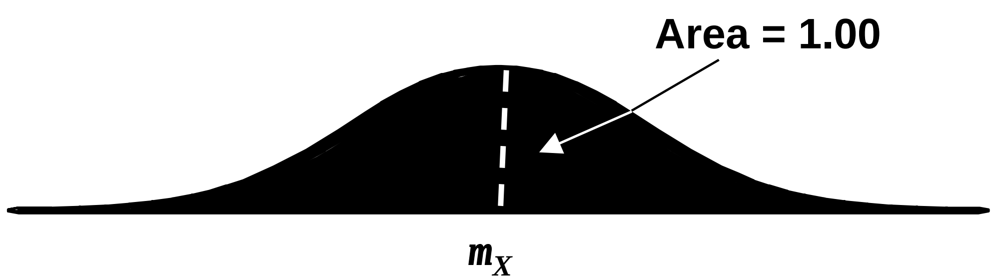
$$= s_X^2 + s_Y^2 + 2r_{XY}s_Xs_Y$$

Variance Measures Uncertainty

- s_X^2 Small



- s_X^2 Large



Contents

- A Cost Analysis is a Risk Analysis
- Measuring Uncertainty in Cost
- What is Correlation?
- **Correlation Impacts Uncertainty**
- Correlation Can Also Impact the Estimate!
- Summary

THE AEROSPACE CORPORATION

Correlation Affects the Variance

- $X_1,\ X_2,\ \ldots,\ X_n$ **are Costs of WBS Elements (Random Variables)**
- **Total Cost =** $\sum_{k=1}^{n} X_k = X_1 + X_2 + \ldots + X_n$
- **Mean of Total Cost =** $E\left(\sum_{k=1}^{n} X_k\right) = \sum_{k=1}^{n} E(X_k) = \sum_{k=1}^{n} m_k$
- **Variance of Total Cost =** $Var\left(\sum_{k=1}^{n} X_k\right)$

$$= \sum_{k=1}^{n} s_k^2 + 2\sum_{j=2}^{n} \sum_{i=1}^{j-1} r_{ij} s_i s_j$$

Does Correlation Matter?

- **If WBS-Element Costs are Uncorrelated (all $\rho_{ij} = 0$),**

  Variance of Total Cost = $\sum_{k=1}^{n} \sigma_k^2$

- **If WBS-Element Costs are Correlated,**

  Variance of Total Cost = $\sum_{k=1}^{n} \sigma_k^2 + 2\sum_{j=2}^{n} \sum_{i=1}^{j-1} \rho_{ij} \sigma_i \sigma_j$

  - **Positive Correlations Increase Dispersion**
  - **Negative Correlations Reduce Dispersion**

- **If (“Worst” Case) All Correlations $\rho_{ij} = 1$,**

  Variance of Total Cost = $\left(\sum_{k=1}^{n} \sigma_k \right)^2 >> \sum_{k=1}^{n} \sigma_k^2$

- **“Ignoring” Correlation Issue is Tantamount to Setting all $\rho_{ij} = 0$**

Yes, Correlation Matters

- **Suppose for Simplicity**
  - **There are *n* Cost Elements** $C_1,\ C_2,\ \ldots,\ C_n$
  - **Each** $Var(C_i) = s^2$
  - **Each** $Corr(C_i, C_j) = r < 1$
  - **Total Cost** $C = \sum_{k=1}^{n} C_i$
- $Var(C) = \sum_{k=1}^{n} Var(C_i) + 2r \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \sqrt{Var(C_i)\ Var(C_j)}$

$$= ns^2 + n(n-1)rs^2$$

$$= ns^2\left(1 + (n-1)r\right)$$

Correlation	0	ρ	1
$Var(C)$	ns^2	$ns^2\left(1+(n-1)r\right)$	n^2s^2

Magnitude of Correlation Impact

- **Percent Underestimation of Total-Cost Sigma $\left(\sqrt{Var(C)}\right)$ When Correlation Assumed to be 0 instead of ρ is 100% times ...**

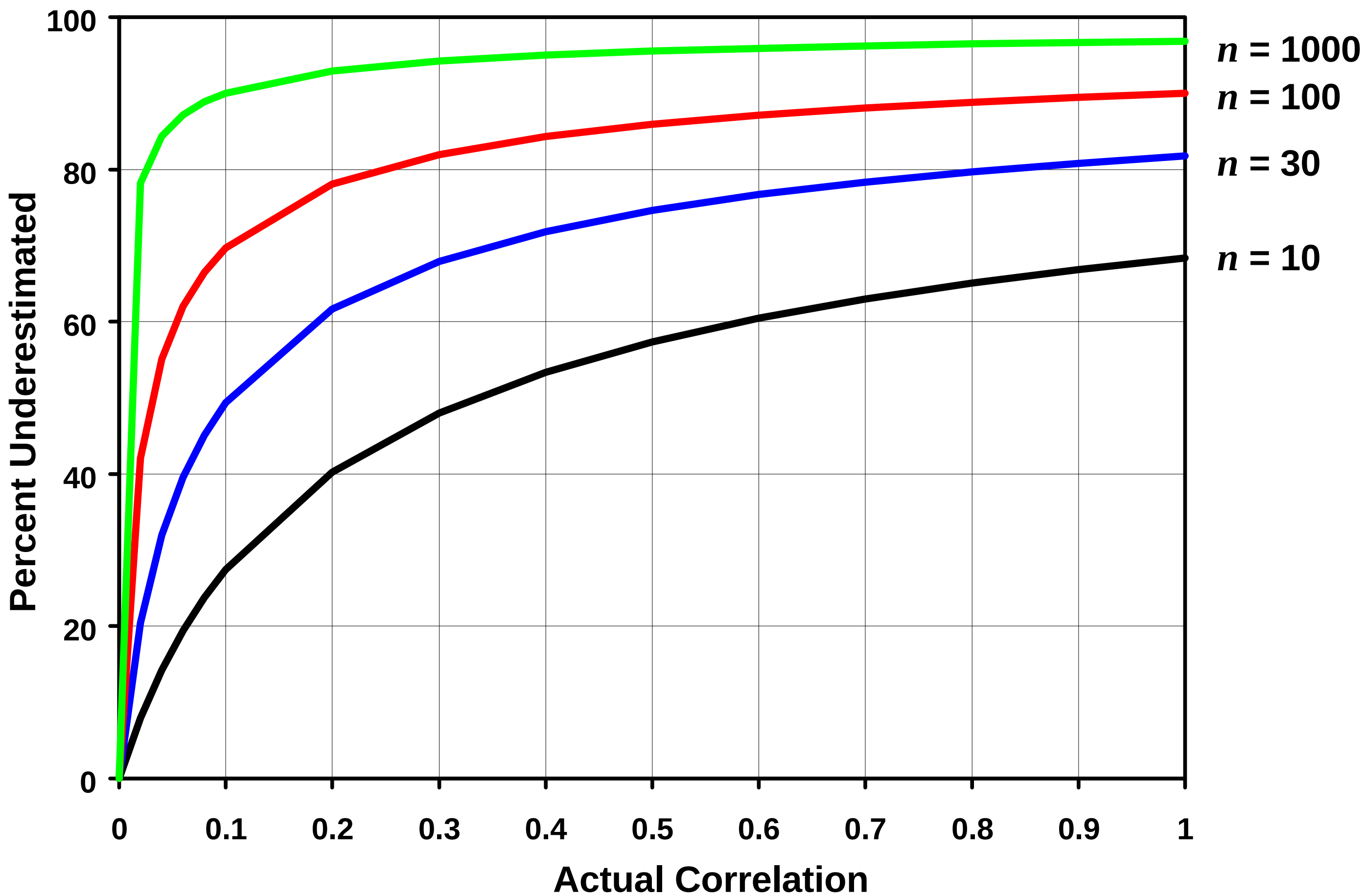
$$\frac{\sqrt{ns}\sqrt{1+(n-1)r}-\sqrt{ns}}{\sqrt{ns}\sqrt{1+(n-1)r}}=1-\sqrt{\frac{1}{1+(n-1)r}}$$

- **Percent Overestimation of Total-Cost Sigma $\left(\sqrt{Var(C)}\right)$ When Correlation Assumed to be 1 instead of ρ is 100% times ...**

$$\frac{ns-\sqrt{ns}\sqrt{1+(n-1)r}}{\sqrt{ns}\sqrt{1+(n-1)r}}=\sqrt{\frac{n}{1+(n-1)r}}-1$$

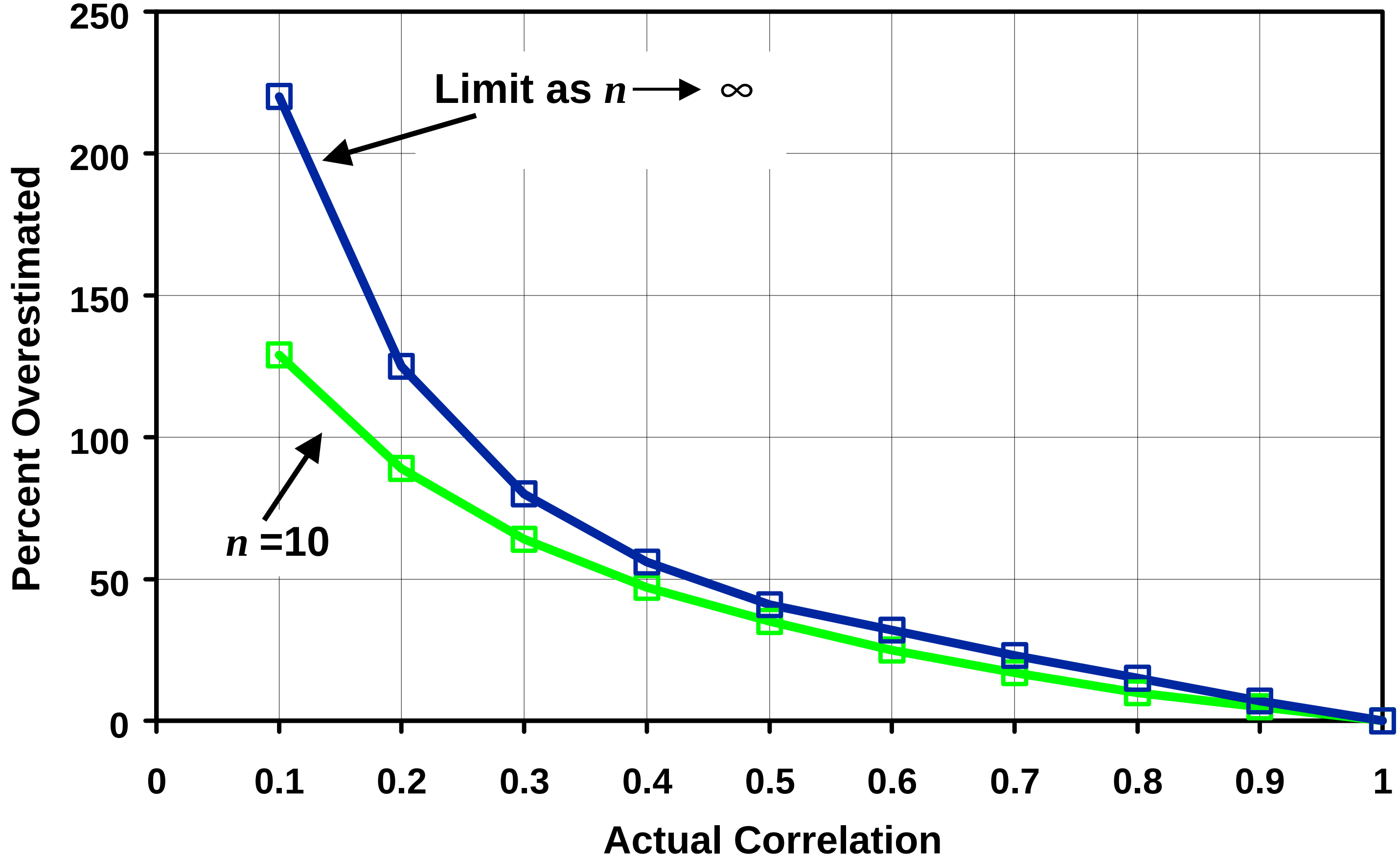
Maximum Possible Underestimation of Total-Cost Sigma

- **Percent Underestimated When Correlation Assumed to be 0 Instead of ρ**



Maximum Possible Overestimation of Total-Cost Sigma

- **Percent Overestimated When Correlation Assumed to be 1 Instead of ρ**



Impact of Nonzero Correlation Value

- **Percent Underestimation of Total-Cost Sigma** $\left(\sqrt{Var(C)}\right)$ **When Correlation Assumed to be** ρ_1 **instead of** $\rho_2 > \rho_1$ **is 100% times ...**

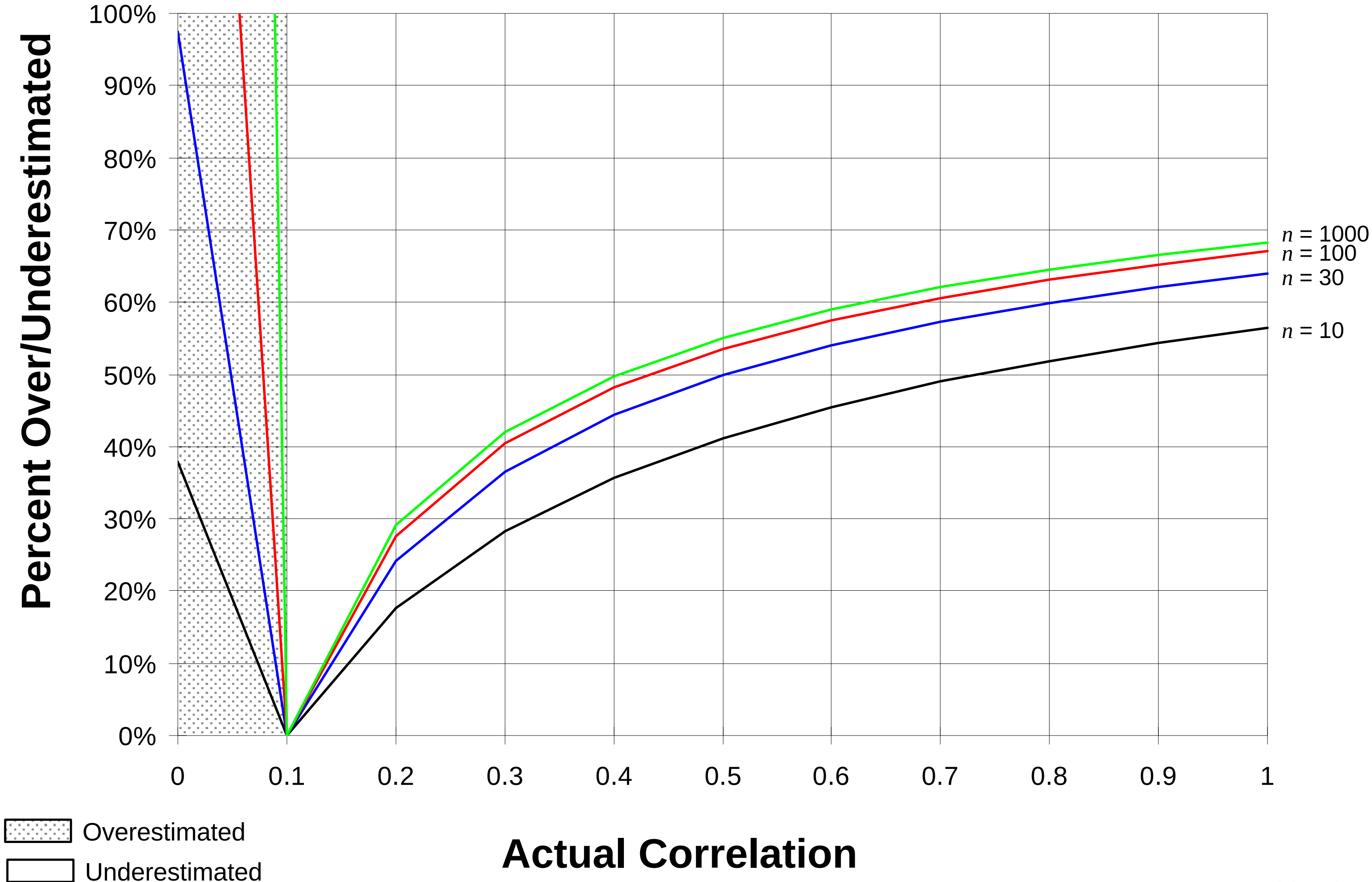
$$1-\sqrt{\frac{1+(n-1)\mathbf{r}_1}{1+(n-1)\mathbf{r}_2}}$$

- **Percent Overestimation of Total-Cost Sigma** $\left(\sqrt{Var(C)}\right)$ **When Correlation Assumed to be** ρ_2 **instead of** $\rho_1 < \rho_2$ **is 100% times ...**

$$\sqrt{\frac{1+(n-1)\mathbf{r}_2}{1+(n-1)\mathbf{r}_1}}-1$$

Maximum Possible Over- and Under-Estimation of Total-Cost Sigma

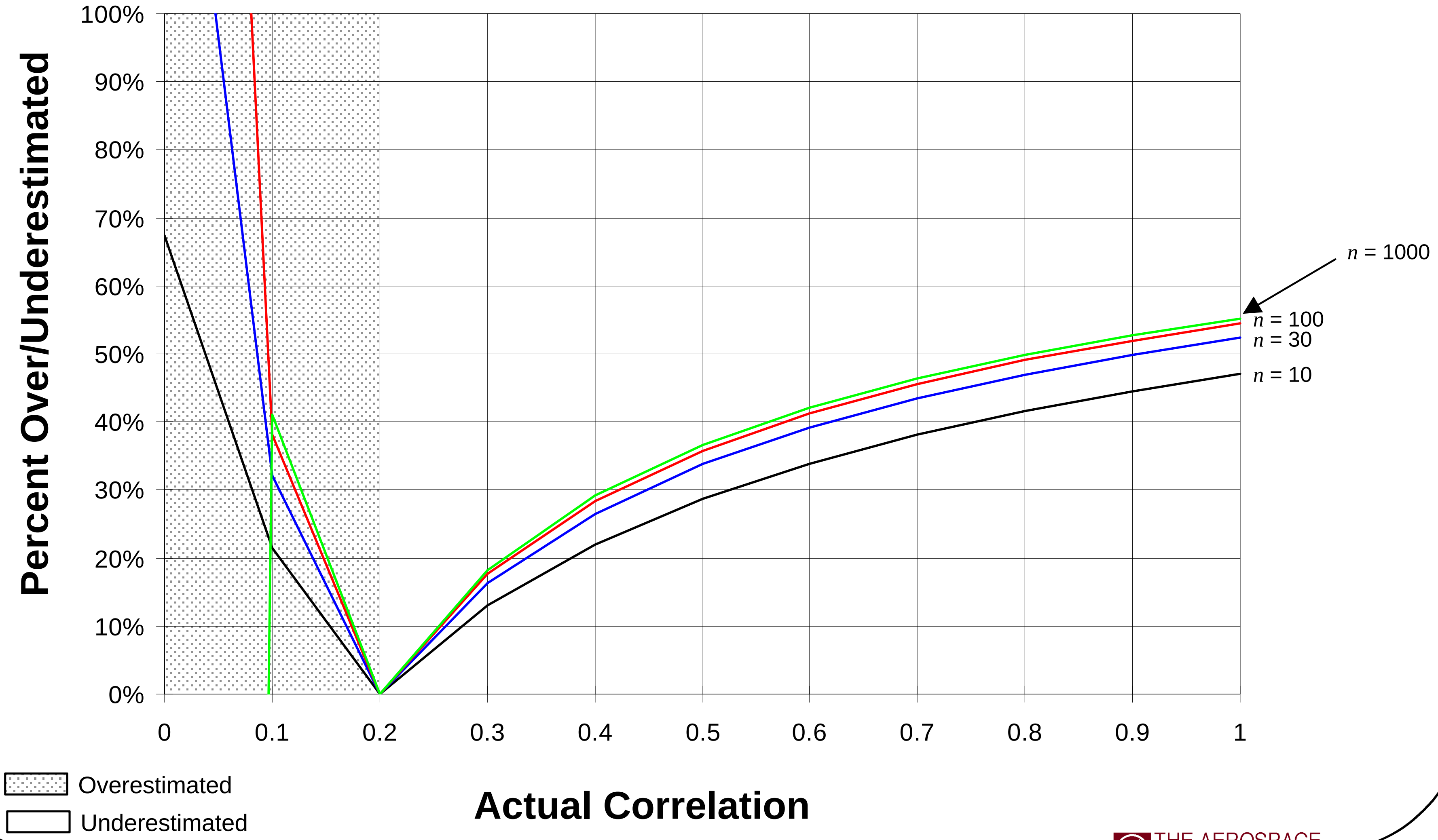
- **Percent Over/Underestimated When Correlation Assumed to be 0.1 Instead of ρ**



THE AEROSPACE CORPORATION

Maximum Possible Over- and Under-Estimation of Total-Cost Sigma

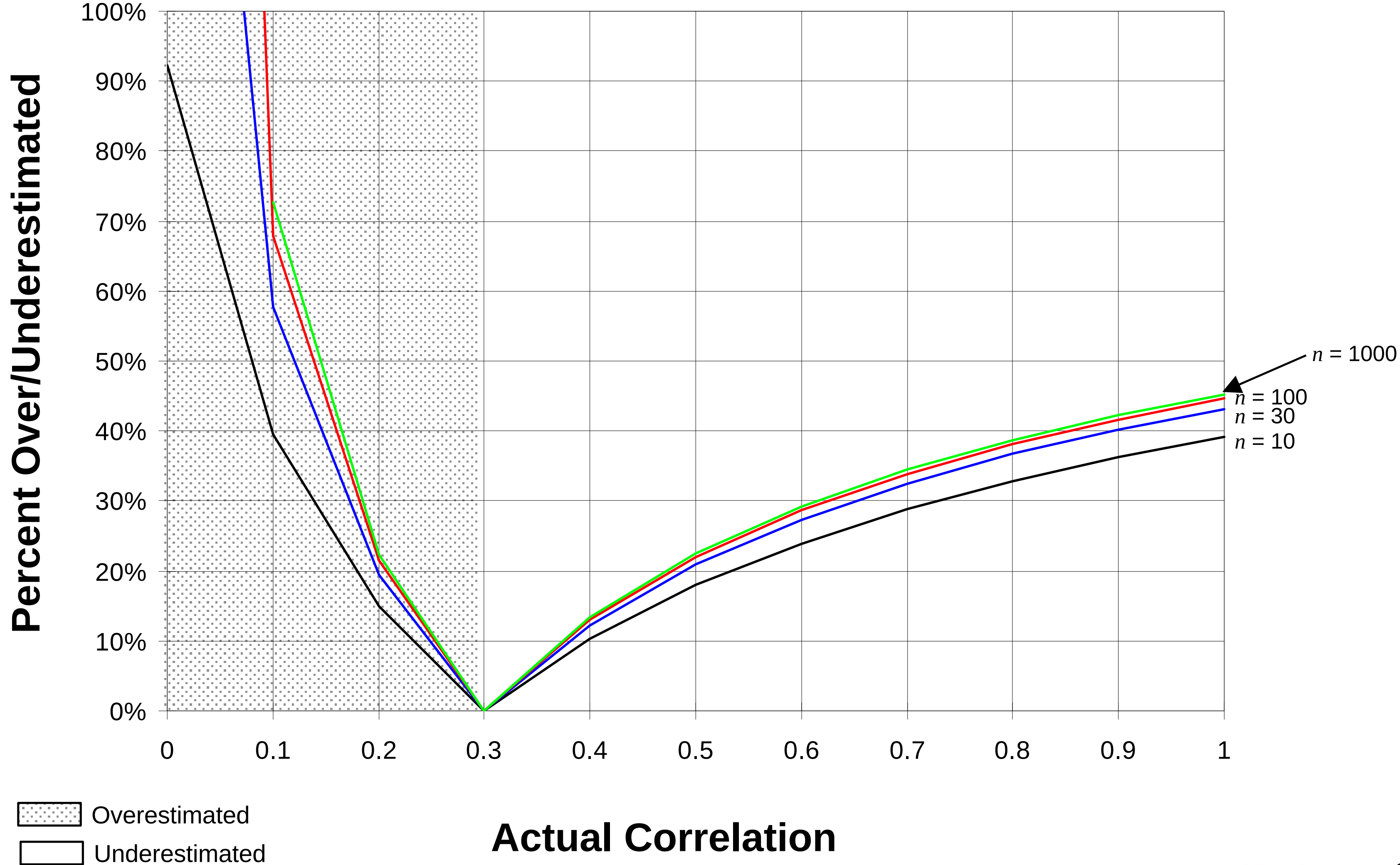
- **Percent Over/Underestimated When Correlation Assumed to be 0.2 Instead of ρ**



THE AEROSPACE CORPORATION

Maximum Possible Over- and Under-Estimation of Total-Cost Sigma

- **Percent Over/Underestimated When Correlation Assumed to be 0.3 Instead of ρ**



THE AEROSPACE CORPORATION

Selection of Correlation Values

- "Ignoring" Correlation is Equivalent to Assuming that Risks are Uncorrelated, e.g., All Correlations are Zero
- Reasonable Choice of Nonzero Values Brings You Closer to Truth
  - 0.2 is at "Knee" of Curve on Previous Charts, thereby Providing Most of the Benefits at Least Loss of Accuracy
- Square of Correlation Represents Percentage of Variation in one WBS Element's Cost Attributable to Influence of Another's

Correlation	% Influenced
0.00	0%
±0.10	1%
±0.32	10%
±0.50	25%
±0.71	50%

A Technical Problem

- **If WBS-Element Costs Are Correlated, Usual Monte Carlo Procedure (Summing Random Numbers Representing WBS-Element Costs) Does Not Produce Total Cost since Random Numbers are Independent**
- **Ideal Way to Simulate Total Cost Would be to Specify Inter-WBS-Element Correlations, Generate Correlated Random Numbers, and Sum Them**
- **Probability Theory Allows This to be Done *Exactly* <u>only</u> for Gaussian Cost Distributions, Not in General**
  - **Not Obvious How to Generate Correlated Random Numbers that Correctly Model WBS-Element Cost Distributions**
  - **Combining Process Must Provide <u>Approximately</u> Correct Values of Mean, Sigma, Percentiles of Total-Cost Distribution**
  - **Drs. M.S. Goldberg and P.M. Lurie of IDA (Institute for Defense Analyses) have been Working on this Problem for Several Years**

There is a Second Type of Correlation

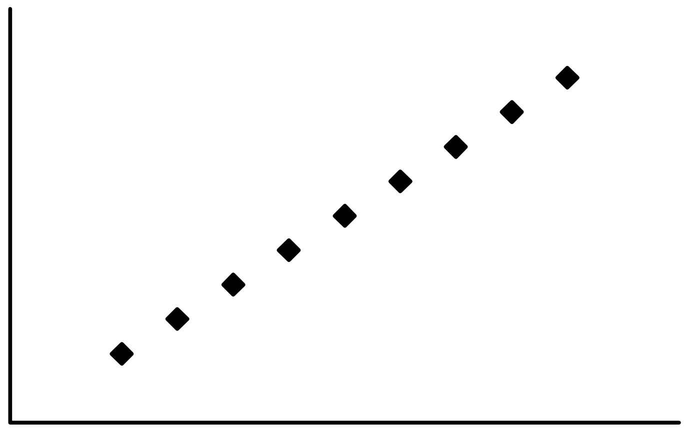
- **Pearson Product-Moment Linear Correlation** $\boldsymbol{r}(X,Y)$
  - $\boldsymbol{r}(X,Y)=\pm 1$ **if and only if X and Y are linearly related, i.e., the least-squares linear relationship between X and Y allows us to predict Y precisely, given X**
  - $\boldsymbol{r}^2(X,Y)$ **= proportion of variation in Y that can be explained on the basis of a least-squares linear relationship between X and Y**
  - $\boldsymbol{r}(X,Y)=0$ **if and only if the least-squares linear relationship between X and Y provides no ability to predict Y, given X**

- **Spearman Rank Correlation** $\boldsymbol{r}_s(X,Y)$
  - $\boldsymbol{r}_s(X,Y)=1$ **if and only if the largest value of X corresponds to the largest value of Y , the second largest, ... , etc.**
  - $\boldsymbol{r}_s(X,Y)=-1$ **if and only if the largest value of X corresponds to the smallest value of Y, etc.**
  - $\boldsymbol{r}_s(X,Y)=0$ **if and only if the rank of a particular X among all X values provides no ability to predict the rank of the corresponding Y among all Y values**

How Do The Two Types of Correlation Differ?

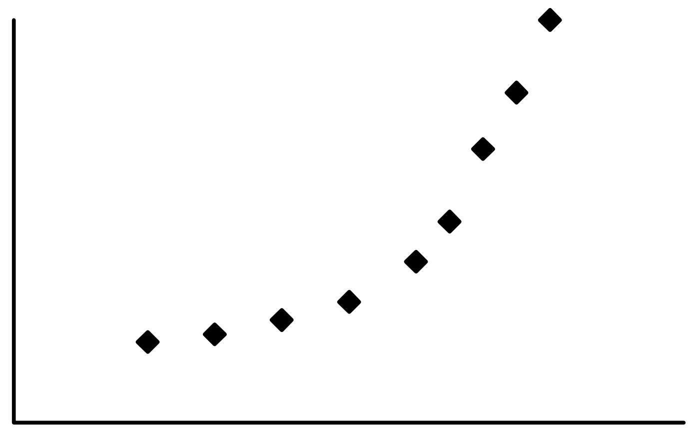
- <u>Pearson</u> Correlation
  - Measures Extent of <u>Linearity</u> of a Relationship Between Two Random Variables
  - Plays an Explicit, Well-defined Role in Establishing the Sigma Value (as well as the Range) of the Total-Cost Distribution
- <u>Spearman</u> Correlation
  - Measures Extent of <u>Monotonicity</u> of a Relationship Between Two Random Variables
  - Does Not Appear Explicitly in the Formula for the Total-Cost Sigma (Its impact on Sigma is Not Known)
- P.R. Garvey of The MITRE Corporation has Done Research on the Issue of Pearson vs. Spearman Correlation

Linear vs. Rank Correlation



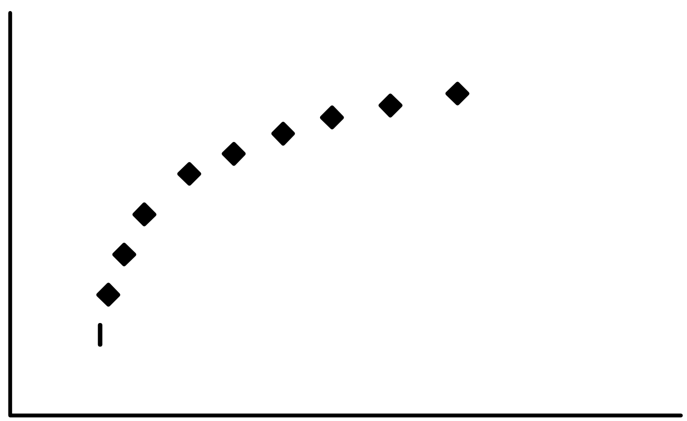
$r = 1.0$

$r_s = 1.0$



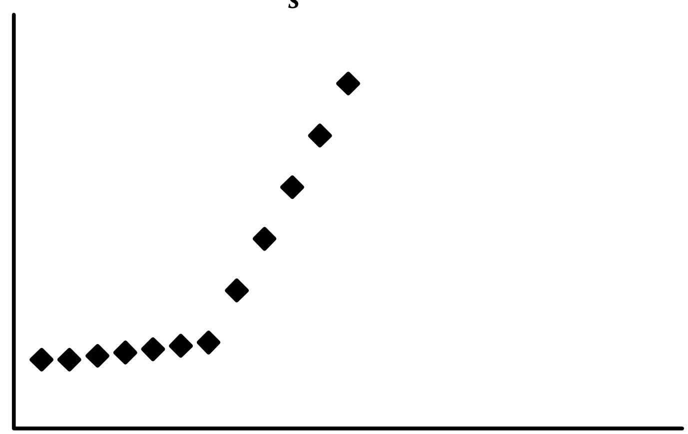
$r = 0.8$

$r_s = 1.0$



$r = 0.8$

$r_s = 1.0$



$r = 0.6$

$r_s = 1.0$

Spearman Rank Correlation Coefficient

- **Data Structure**

CASE	RANK OF X_i VALUE	RANK OF X_j VALUE	DIFFERENCE	SQUARED DIFFERENCE
#1	r_1	c_1	$d_1 = c_1 - r_1$	d_1^2
#2	r_2	c_2	$d_2 = c_2 - r_2$	d_2^2
#3	r_3	c_3	$d_3 = c_3 - r_3$	d_3^2
⋮	⋮	⋮	⋮ ⋮	⋮
#4	r_n	c_n	$d_n = c_n - r_n$	d_n^2
SUMS	$\frac{n(n+1)}{2}$	$\frac{n(n+1)}{2}$	$\sum d = \sum c - \sum r = 0$	$\sum d^2$

$$r\left(X_i, X_j\right) = 1 - \frac{6\sum d^2}{n\left(n^2 - 1\right)}$$

- **Statistics Theorem: Spearman Rank Correlation Coefficient Equals Pearson (Linear) Correlation Coefficient Calculated Between the Two Sets of Ranks**

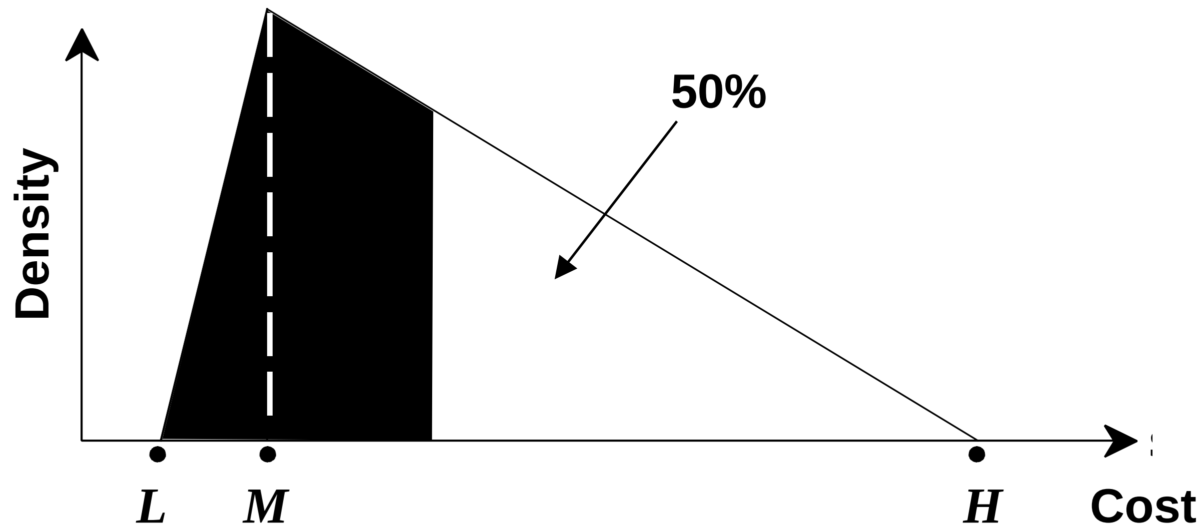
Crystal Ball, @Risk

- **Commercially Available Software Packages that are Add-ons to Additional Commercial Software Such As Windows, Excel, or Lotus on PC or Mac**
  - **Crystal Ball Marketed by Decisioneering, Inc., 2530 S. Parker Road, Suite 220, Aurora, CO 80014, (800) 289-2550**
  - **@Risk Marketed by Palisade Corporation, 31 Decker Road, Newfield, NY 14867, (800) 432-7475**
- **Inputs**
  - **Parameters Defining WBS-Element Distributions**
  - **Rank Correlations Among WBS-Element Cost Distributions**
- **Mathematics**
  - **Monte-Carlo and Stratified Random Sampling (Latin Hypercube)**
  - **Virtually All Probability Distributions That Have Names Can Be Used**
  - **Suggests Adjustments to Inconsistent Input Correlation Matrix**
- **Outputs**
  - **Percentiles of Program Cost**
  - **Cost Probability Density and Cumulative Distribution Graphics**

Contents

- A Cost Analysis is a Risk Analysis
- Measuring Uncertainty in Cost
- What is Correlation?
- Correlation Impacts Uncertainty
- **Correlation Also Impacts the Estimate!**
- Summary

Example: Triangular Cost Distribution



- **Probability Density Function**
- **Total Area = 1.00**
- **Three Parameters *L*, *M*, *H* Completely Specify Distribution**
- **Mean, Median, Sigma, All Percentiles can be Expressed in Terms of *L*, *M*, and *H***

Statistical Metrics of Triangular Distribution

- **Mode = M (most likely value of cost)**
- **Median =** $T_{.50} = L + \sqrt{0.50(M-L)(H-L)}$ if $M - L \geq 0.50(H-L)$

  $= H - \sqrt{0.50(H-L)(H-M)}$ if $M - L \leq 0.50(H-L)$

- T_p **= Dollar Value at Which** $P\{\text{Cost} \leq T_p\} = p$

  $T_p = H - \sqrt{(1-p)(H-L)(H-M)}$ if $p \geq \dfrac{M-L}{H-L}$

  $T_p = L + \sqrt{p(M-L)(H-L)}$ if $p \leq \dfrac{M-L}{H-L}$

- **Mean =** $\dfrac{L+M+H}{3}$

$$s = \sqrt{\frac{L^2 + M^2 + H^2 - LM - LH - MH}{18}}$$

THE AEROSPACE CORPORATION

Sum of Five Uncorrelated Triangular Distributions

- **Monte-Carlo Simulation Using @Risk™ Software**
- **Input Parameters: L = 100, M = 200, H = 500**
- **Comparison of Rolled-Up vs. Correct Total-Cost Statistics**

STATISTIC	EACH INPUT DISTRIBUTION	ROLLED-UP VALUE	CORRECT VALUE	OVER-ESTIMATE: ROLL-UP, MINUS CORRECT
MEAN	266.67	1333.33	1333.33	0.00
MODE	200.00	1000.00	1236.94	-236.94
MEDIAN	255.05	1275.25	1328.53	-53.28
STANDARD DEVIATION	84.98	190.03	190.03	0.00
50th PERCENTILE	255.05	1275.25	1328.53	-53.28
70th PERCENTILE	310.25	1551.30	1432.52	118.78
90th PERCENTILE	390.46	1952.30	1573.66	378.64

Sum of Five Correlated Triangular Distributions

- **Monte-Carlo Simulation Using @Risk™ Software**
- **Input Parameters: L = 100, M = 200, H = 500**
- **All Pairwise Correlations are 0.10**
- **Comparison of Rolled-Up vs. Correct Total-Cost Statistics**

STATISTIC	EACH INPUT DISTRIBUTION	ROLLED-UP VALUE	CORRECT VALUE	OVER-ESTIMATE: ROLL-UP, MINUS CORRECT
MEAN	266.67	1333.33	1333.33	0.00
MODE	200.00	1000.00	1192.82	-192.82
MEDIAN	255.05	1275.25	1319.20	-43.95
STANDARD DEVIATION	84.98	190.03	224.85	-34.82
50th PERCENTILE	255.05	1275.25	1319.20	-43.95
70th PERCENTILE	310.25	1551.30	1444.29	107.01
90th PERCENTILE	390.46	1952.30	1635.07	317.23

THE AEROSPACE CORPORATION

Sum of Five Correlated Triangular Distributions

- **Monte-Carlo Simulation Using @Risk™ Software**
- **Input Parameters: L = 100, M = 200, H = 500**
- **All Pairwise Correlations are 0.30**
- **Comparison of Rolled-Up vs. Correct Total-Cost Statistics**

STATISTIC	EACH INPUT DISTRIBUTION	ROLLED-UP VALUE	CORRECT VALUE	OVER-ESTIMATE: ROLL-UP, MINUS CORRECT
MEAN	266.67	1333.33	1333.33	0.00
MODE	200.00	1000.00	1187.18	-187.18
MEDIAN	255.05	1275.25	1305.28	-30.03
STANDARD DEVIATION	84.98	190.03	281.86	-91.83
50th PERCENTILE	255.05	1275.25	1305.28	-30.03
70th PERCENTILE	310.25	1551.30	1469.59	81.71
90th PERCENTILE	390.46	1952.30	1715.72	236.58

Sum of Five Correlated Triangular Distributions

- **Monte-Carlo Simulation Using @Risk™ Software**
- **Input Parameters: L = 100, M = 200, H = 500**
- **All Pairwise Correlations are 0.50**
- **Comparison of Rolled-Up vs. Correct Total-Cost Statistics**

STATISTIC	EACH INPUT DISTRIBUTION	ROLLED-UP VALUE	CORRECT VALUE	OVER-ESTIMATE: ROLL-UP, MINUS CORRECT
MEAN	266.67	1333.33	1333.33	0.00
MODE	200.00	1000.00	1181.76	-181.76
MEDIAN	255.05	1275.25	1297.63	-22.38
STANDARD DEVIATION	84.98	190.03	329.14	-139.11
50th PERCENTILE	255.05	1275.25	1297.63	-22.38
70th PERCENTILE	310.25	1551.30	1494.05	57.25
90th PERCENTILE	390.46	1952.30	1777.48	174.82

THE AEROSPACE CORPORATION

Sum of Five Correlated Triangular Distributions

- Monte-Carlo Simulation Using @Risk™ Software
- Input Parameters: L = 100, M = 200, H = 500
- All Pairwise Correlations are 0.80
- Comparison of Rolled-Up vs. Correct Total-Cost Statistics

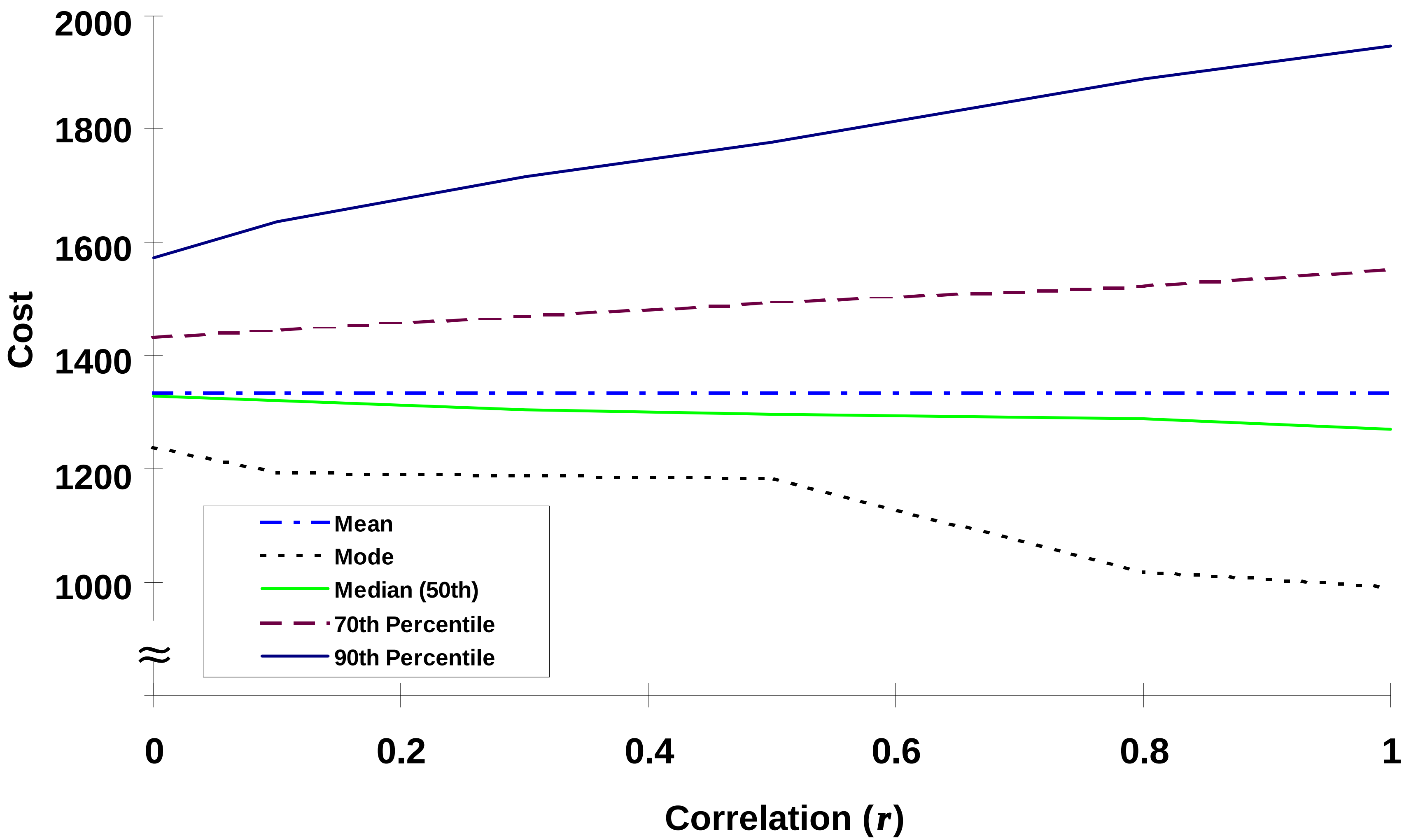
STATISTIC	EACH INPUT DISTRIBUTION	ROLLED-UP VALUE	CORRECT VALUE	OVER-ESTIMATE: ROLL-UP, MINUS CORRECT
MEAN	266.67	1333.33	1333.33	0.00
MODE	200.00	1000.00	1018.70	-18.70
MEDIAN	255.05	1275.25	1289.00	-13.75
STANDARD DEVIATION	84.98	190.03	389.44	-199.41
50th PERCENTILE	255.05	1275.25	1289.00	-13.75
70th PERCENTILE	310.25	1551.30	1522.95	28.35
90th PERCENTILE	390.46	1952.30	1887.31	64.99

Sum of Five Correlated Triangular Distributions

- **Monte-Carlo Simulation Using @Risk™ Software**
- **Input Parameters: L = 100, M = 200, H = 500**
- **All Pairwise Correlations are 1.00**
- **Comparison of Rolled-Up vs. Correct Total-Cost Statistics**

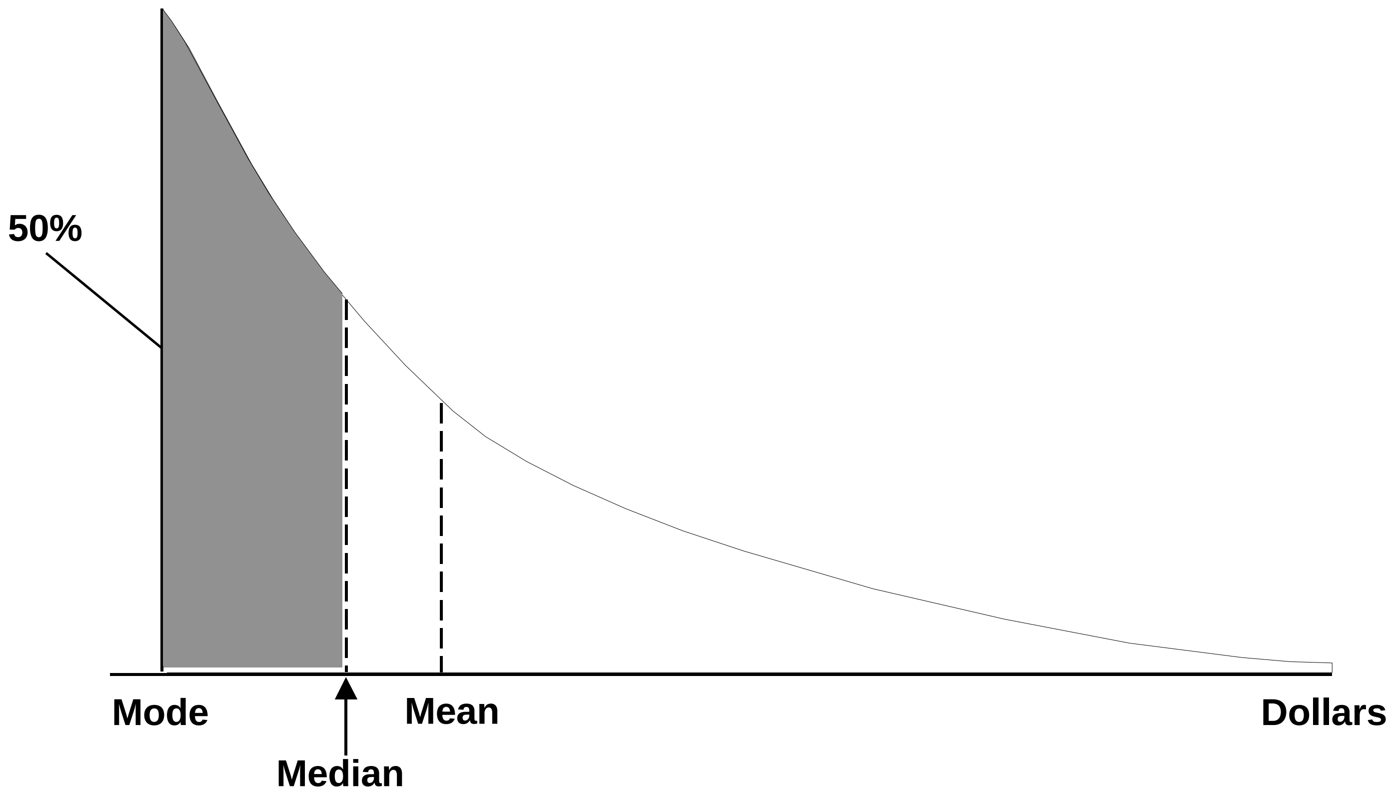
STATISTIC	EACH INPUT DISTRIBUTION	ROLLED-UP VALUE	CORRECT VALUE	OVER-ESTIMATE: ROLL-UP, MINUS CORRECT
MEAN	266.67	1333.33	1333.33	0.00
MODE	200.00	1000.00	990.69	9.31
MEDIAN	255.05	1275.25	1271.09	4.16
STANDARD DEVIATION	84.98	190.03	424.92	-234.89
50th PERCENTILE	255.05	1275.25	1271.09	4.16
70th PERCENTILE	310.25	1551.30	1550.07	1.23
90th PERCENTILE	390.46	1952.30	1946.27	6.03

Effects of Increasing Correlation on the Sum of Five Triangular Distributions

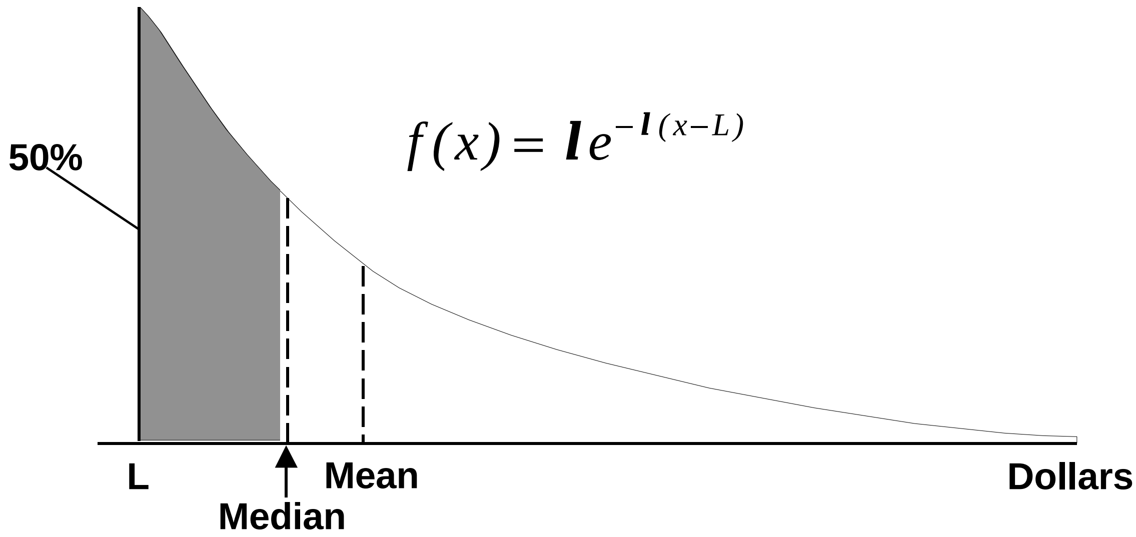


Example : Exponential Distribution

- **Five WBS Elements**
- **WBS-Element Cost Distributions Each Have Exponential Distribution**



Exponential Cost Distribution



$$f(x) = \lambda e^{-\lambda(x-L)}$$

- *f(x)* **is the Exponential Probability Density Function**
- **Total Area = 1.00**
- **Two Parameters L, λ Completely Specify Distribution**
- **Mean, Median, Sigma, All Percentiles can be Expressed in Terms of L and λ**

Statistical Metrics of Exponential Distribution

- **Mode = L**

- **Median = 50th Percentile =** $L + \dfrac{\ln 2}{l}$

- E_p **= Dollar Value at Which** $P\{\text{Cost} \leq E_p\}$ **is** p

$$E_p = L - \frac{\ln(1-p)}{l}$$

- **Mean =** $L + \dfrac{1}{l}$

- $Q = \dfrac{1}{l}$

Sum of Five Uncorrelated Exponential Distributions

- **Monte-Carlo Simulation Using @Risk™ Software**
- **Input Parameters: L = 100, M = 200, H = 500**
- **Comparison of Rolled-Up vs. Correct Total-Cost Statistics**

STATISTIC	EACH INPUT DISTRIBUTION	ROLLED-UP VALUE	CORRECT VALUE	OVER-ESTIMATE: ROLL-UP, MINUS CORRECT
MEAN	266.67	1333.33	1333.33	0.00
MODE	100.00	500.00	1153.04	-653.04
MEDIAN	215.52	1077.62	1287.83	-210.21
STANDARD DEVIATION	166.67	372.68	372.68	0.00
50th PERCENTILE	215.52	1077.62	1287.83	-210.21
70th PERCENTILE	300.66	1503.31	1486.16	17.15
90th PERCENTILE	483.76	2418.82	1809.35	609.47

Sum of Five Correlated Exponential Distributions

- **Monte-Carlo Simulation Using @Risk™ Software**
- **Input Parameters: L = 100, M = 200, H = 500**
- **All Pairwise Correlations are 0.10**
- **Comparison of Rolled-Up vs. Correct Total-Cost Statistics**

STATISTIC	EACH INPUT DISTRIBUTION	ROLLED-UP VALUE	CORRECT VALUE	OVER-ESTIMATE: ROLL-UP, MINUS CORRECT
MEAN	266.67	1333.33	1333.33	0.00
MODE	100.00	500.00	1027.80	-527.80
MEDIAN	215.52	1077.62	1249.68	-172.06
STANDARD DEVIATION	166.67	372.68	440.96	-68.28
50th PERCENTILE	215.52	1077.62	1249.68	-172.06
70th PERCENTILE	300.66	1503.31	1481.14	22.17
90th PERCENTILE	483.76	2418.82	1903.46	515.36

THE AEROSPACE CORPORATION

Sum of Five Correlated Exponential Distributions

- **Monte-Carlo Simulation Using @Risk™ Software**
- **Input Parameters: L = 100, M = 200, H = 500**
- **All Pairwise Correlations are 0.30**
- **Comparison of Rolled-Up vs. Correct Total-Cost Statistics**

STATISTIC	EACH INPUT DISTRIBUTION	ROLLED-UP VALUE	CORRECT VALUE	OVER-ESTIMATE: ROLL-UP, MINUS CORRECT
MEAN	266.67	1333.33	1333.33	0.00
MODE	100.00	500.00	981.76	-481.76
MEDIAN	215.52	1077.62	1205.95	-128.33
STANDARD DEVIATION	166.67	372.68	552.77	-180.09
50th PERCENTILE	215.52	1077.62	1205.95	-128.33
70th PERCENTILE	300.66	1503.31	1498.84	4.47
90th PERCENTILE	483.76	2418.82	2053.84	-364.98

Sum of Five Correlated Exponential Distributions

- **Monte-Carlo Simulation Using @Risk™ Software**
- **Input Parameters: L = 100, M = 200, H = 500**
- **All Pairwise Correlations are 0.50**
- **Comparison of Rolled-Up vs. Correct Total-Cost Statistics**

STATISTIC	EACH INPUT DISTRIBUTION	ROLLED-UP VALUE	CORRECT VALUE	OVER-ESTIMATE: ROLL-UP, MINUS CORRECT
MEAN	266.67	1333.33	1333.33	0.00
MODE	100.00	500.00	767.35	-267.35
MEDIAN	215.52	1077.62	1165.85	-88.23
STANDARD DEVIATION	166.67	372.68	645.50	-272.82
50th PERCENTILE	215.52	1077.62	1165.85	-88.23
70th PERCENTILE	300.66	1503.31	1497.09	6.22
90th PERCENTILE	483.76	2418.82	2155.59	263.23

Sum of Five Correlated Exponential Distributions

- **Monte-Carlo Simulation Using @Risk™ Software**
- **Input Parameters: L = 100, M = 200, H = 500**
- **All Pairwise Correlations are 0.80**
- **Comparison of Rolled-Up vs. Correct Total-Cost Statistics**

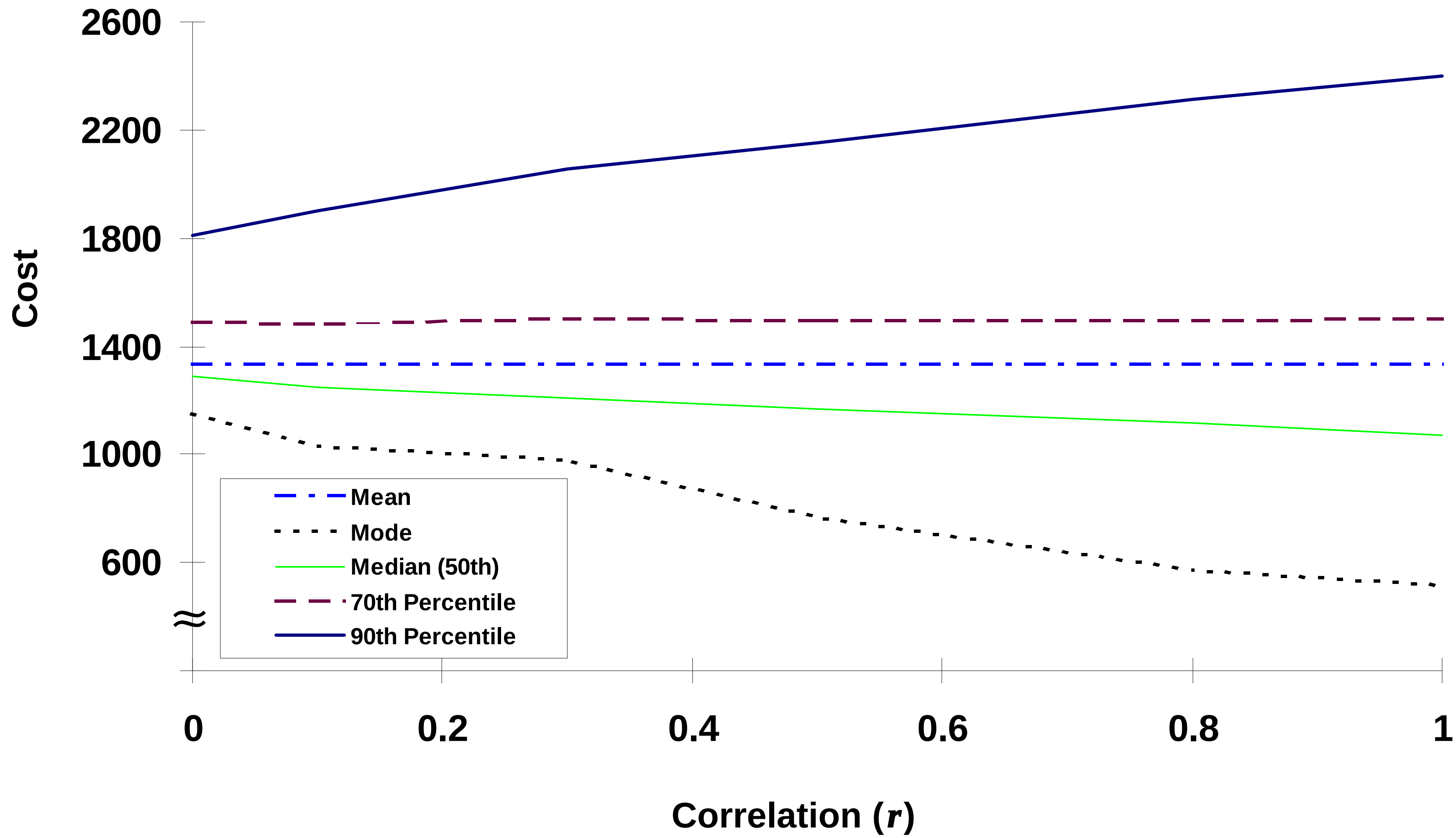
STATISTIC	EACH INPUT DISTRIBUTION	ROLLED-UP VALUE	CORRECT VALUE	OVER-ESTIMATE: ROLL-UP, MINUS CORRECT
MEAN	266.67	1333.33	1333.33	0.00
MODE	100.00	500.00	574.59	-74.59
MEDIAN	215.52	1077.62	1119.08	-41.46
STANDARD DEVIATION	166.67	372.68	763.76	-391.08
50th PERCENTILE	215.52	1077.62	1119.08	-41.46
70th PERCENTILE	300.66	1503.31	1492.81	10.50
90th PERCENTILE	483.76	2418.82	2311.91	106.91

Sum of Five Correlated Exponential Distributions

- **Monte-Carlo Simulation Using @Risk™ Software**
- **Input Parameters: L = 100, M = 200, H = 500**
- **All Pairwise Correlations are 1.00**
- **Comparison of Rolled-Up vs. Correct Total-Cost Statistics**

STATISTIC	EACH INPUT DISTRIBUTION	ROLLED-UP VALUE	CORRECT VALUE	OVER-ESTIMATE: ROLL-UP, MINUS CORRECT
MEAN	266.67	1333.33	1333.33	0.00
MODE	100.00	500.00	512.50	-12.50
MEDIAN	215.52	1077.62	1071.97	5.65
STANDARD DEVIATION	166.67	372.68	833.33	-460.65
50th PERCENTILE	215.52	1077.62	1071.97	5.65
70th PERCENTILE	300.66	1503.31	1501.15	2.16
90th PERCENTILE	483.76	2418.82	2400.81	18.01

Effects of Increasing Correlation on the Sum of Five Exponential Distributions



Contents

THE AEROSPACE CORPORATION

Summary

- Every Cost-Analysis Job Requires a Risk Analysis
- WBS-Element Risks(and therefore Costs) are Typically Correlated
- Correlations Impact the Probable Cost Range
- Correlations Also Impact Cost Estimates, namely the Mode, Median, and other Percentiles of the Cost Probability Distribution
- Your Estimate and Range will be Closer to the Truth if You Use Reasonable Nonzero Correlations Rather Than Zeroes