Numeracy for Cost Analysts

Doing the Right Math, Getting the Math Right

Neal D. Hulkower, Ph.D.

Vice President, Operations

MCR Federal, LLC

nhulkower@mcri.com

(703) 584-7001

Acknowledgments

The author is grateful to the following for their review and suggestions:

Tim Anderson

Steve Book

Ray Covert

Nathan Menton

John Neatrour

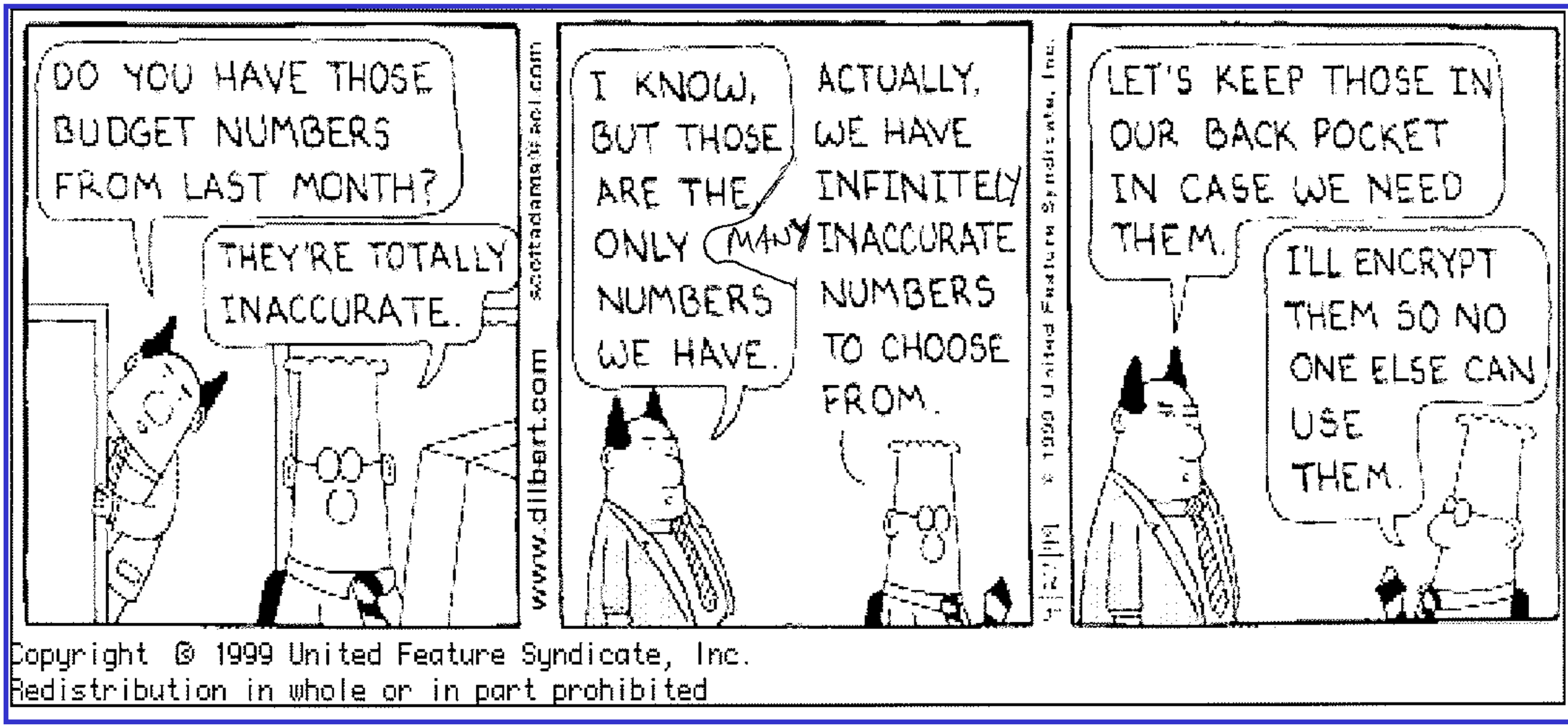
Outline

- Motivation
- Definitions
- Why Numeracy is Critical to Cost Analysis
- What Right Looks Like
- Good Math, Bad Math
- Limitations of Math
- Summary
- The Last Word
- Acronyms

Motivation

- **No one does costing right and it may never be possible to do so**
- **For now we must settle for doing it less wrong**
- **Proper mathematical and statistical techniques are necessary (but admittedly not sufficient) for doing costing less incorrectly**
  - **The use of incorrect techniques has been a contributor to chronic underestimation of cost and undermines confidence in the analysis**
  - **Math is one of the very few ingredients in an estimate for which there are objective criteria for goodness so when it is used, it needs to be done right**

Obligatory Dilbert Cartoon



"Numeracy"

Source: Wikipedia

"*Numeracy* is a contraction (or *portmanteau word*) of 'numerical literacy', and refers to an ability to reason with numbers and other mathematical concepts. The word was coined in 1959 by the UK Committee on Education, presided over by Sir Geoffrey Crowther...

"In the United States, numeracy it is also known as *Quantitative Literacy*, and is familiar to math educators and intellectuals. There is also substantial overlap between conceptions of numeracy and conceptions of *statistical literacy*."

“Numeracy”

Source: Merriam-Webster's Online Dictionary, 10th Edition

Main Entry: ***nu·mer·acy***

Pronunciation: \'n(y)üm-rə-sē, 'n(y)ü-mə-

Function: ***noun*** **Etymology: Latin** ***numerus*** **number + English** ***-acy*** **(as in** ***literacy*****)**

Date: 1959 :

the capacity for quantitative thought and expression

— ***nu·mer·ate*** **\'nüm-rət, 'nü-mə-; 'nyüm-, 'nyü-mə-** ***adjective***

MCR "Creating Customer-Focused Success"

"Innumeracy"

Source: Wikipedia

"*Innumeracy* is a *portmanteau* of 'numerical illiteracy'; it refers to a lack of ability to reason with numbers. The term *innumeracy* was coined by cognitive scientist Douglas Hofstadter and popularized by mathematician John Allen Paulos in his 1989 book, *Innumeracy: Mathematical Illiteracy and its Consequences*. Possible causes of innumeracy are poor teaching methods and standards and lack of value placed on mathematical skills. Even prominent and successful people will attest, sometimes proudly, to low mathematical competence, in sharp contrast to the stigma associated with illiteracy."

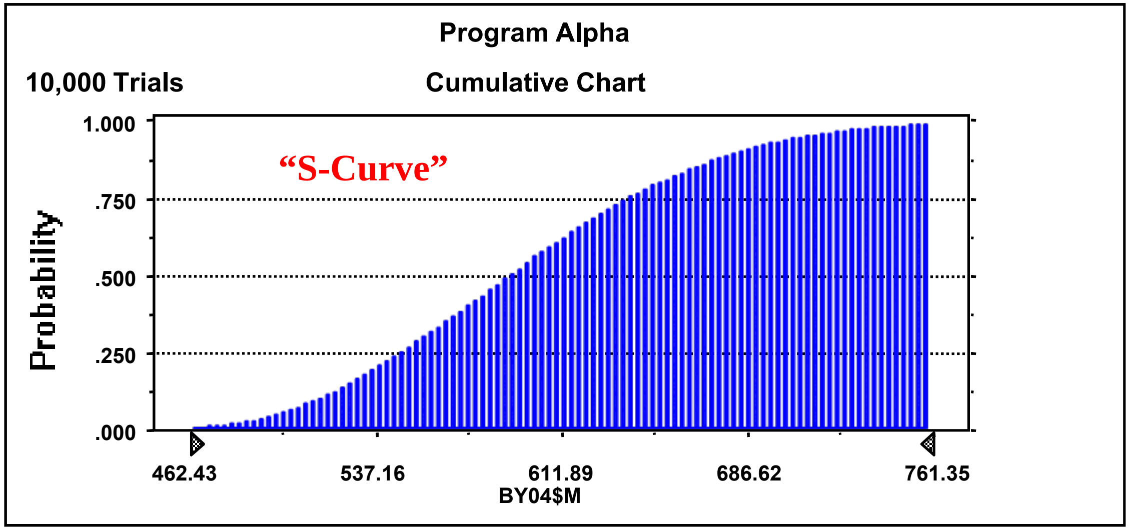
"The problem with remaining innumerate is that one can be fooled not only by numerates but also by other innumerates.".....NDH

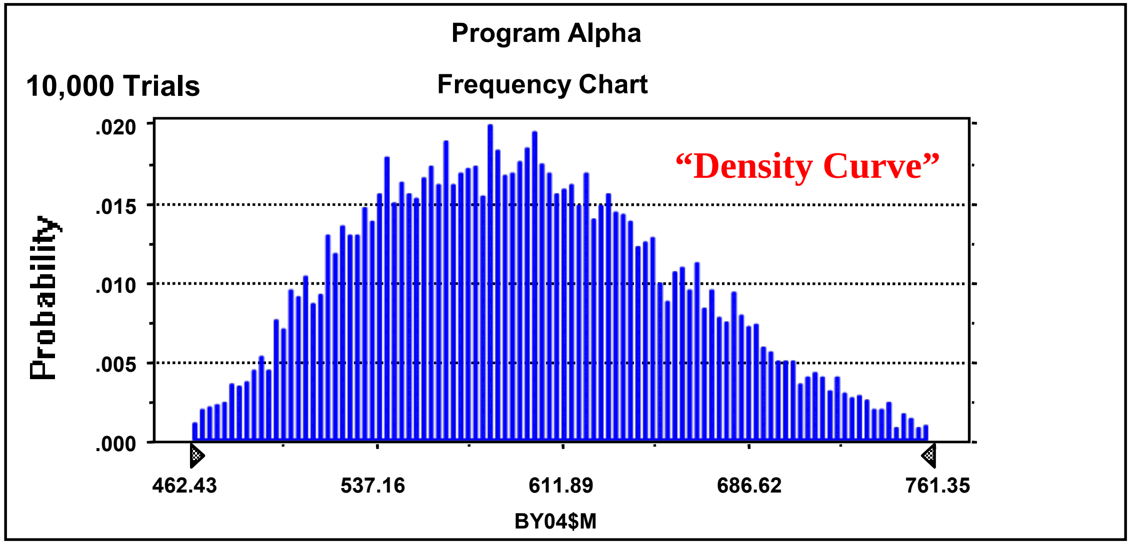
Why Numeracy is Critical to Cost Analysis

A Cost Estimate Is a Probability Distribution, Not a Single Number

Percentile	Value
10%	516.81
20%	538.98
30%	557.85
40%	575.48
50%	592.72
60%	609.70
70%	629.19
80%	650.97
90%	683.01

Statistics	Value
Trials	10,000
Mean	596.40
Median	592.72
Mode	---
Standard Deviation	63.18
Range Minimum	450.19
Range Maximum	796.68





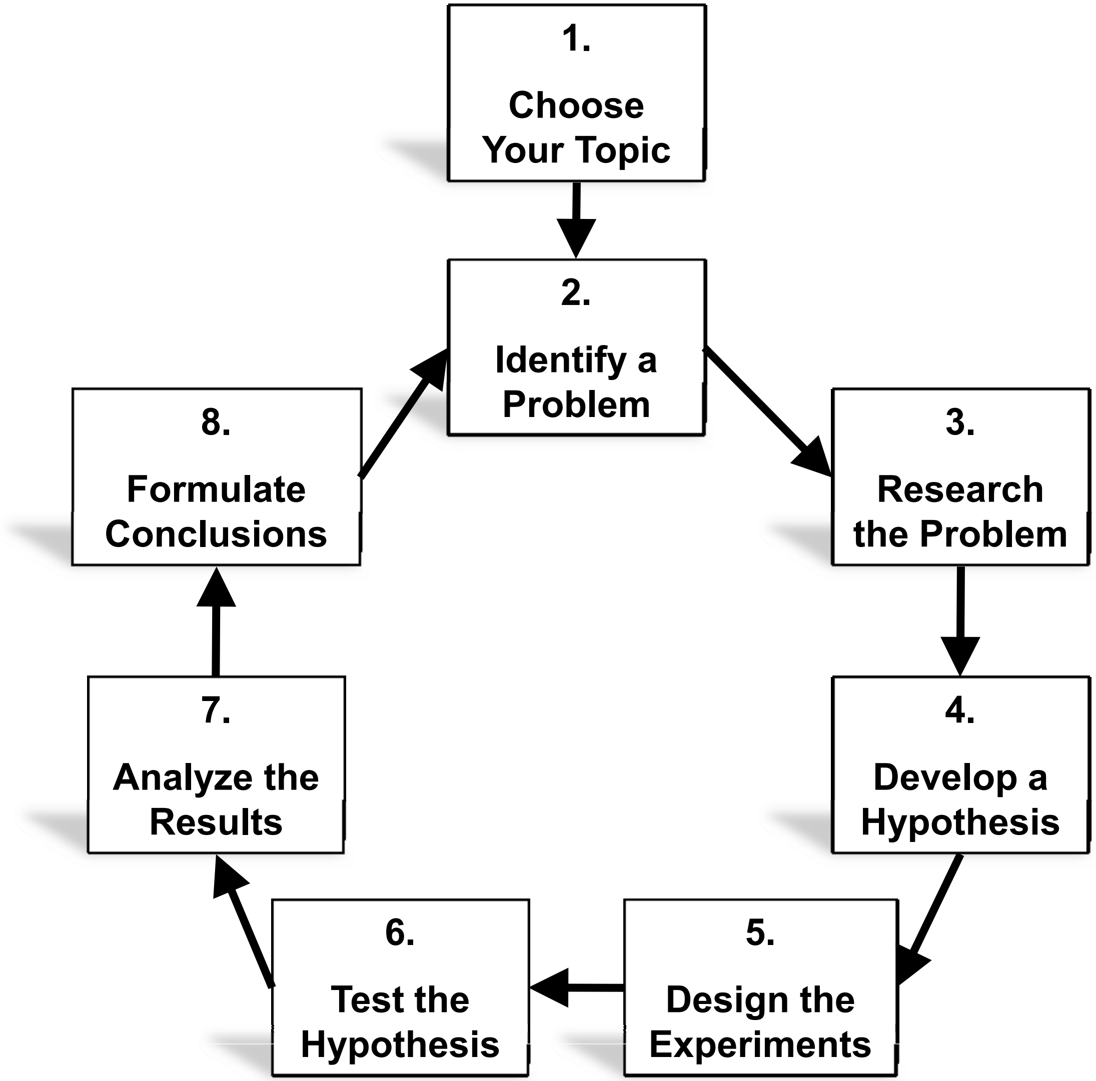
Some of the Math That Gets You to an Estimate

- **Curve fitting methods to establish Cost Estimating Relationships (CERs)**
  - Polynomial Fits
  - Regression techniques
    - Ordinary Least Squares (OLS) and Log-OLS
    - Iteratively Reweighted Least Squares (IRLS)
    - Minimum Percent Error – Zero Percent Bias (MPE-ZPB)
- **Statistical summing and other issues to get to the cost distributions**
  - Monte Carlo and Latin Hypercube
  - Method of Moments
  - Correlation coefficients
- **Spreading algorithms to get to annual funding**

Math is the language of analysis and of cost analysis in particular!

What Right Looks Like

If We Could Only Emulate the Sciences....



...to Develop the Best Math Model for Costing, it Would....

- Fit empirical data
- Be shaved by Occam's Razor
- Be predictive across the broadest set of inputs
- Satisfy an appropriate theory and/or philosophical consideration

...But Cost Analysis is Not a Science...

- **In 1605, Johannes Kepler completed his three laws of planetary motion**
  - **Derived from examination of data collected by Tycho Brahe**
  - **Phenomenologically, not theoretically based**
- **In 1687, Isaac Newton published his three laws of motion from which Kepler's laws can be derived putting them on a causal basis**
- **Cost Analysis is in the phenomenological stage of development <u>without a causal basis</u>**
  - **CER development not tied to any theory**
  - **Methods are either statistical or heuristic**

...So to the Greatest Extent Possible We Must

- **Define terms clearly**
- **Apply theorems properly**
  - **Be sure to satisfy hypotheses before invoking the conclusions!**
- **Use appropriate mathematical formulations**
  - **The world is rarely linear**
  - **Let economics, engineering, and common sense be guides**
- **Do calculations correctly**
  - **Statistical summing is not done arithmetically**
  - **No, you may never average percents, ratios, or averages to create CERs, factors or models**

“Creating Customer-Focused Success”

A Bad Model?

Bid the Price to Win

Subject: Management, marketing

The Rule

The bid price of an electro-optical system usually can be determined by the following relationship:

$$\frac{P_{c1} + P_{c2} + P_{c3} \cdots [\Delta P]}{S_b} \Rightarrow CH_4 + B_c \times (D_m)^{1/2} = \text{Price}$$

where P_c = your competitors' price
P = your past price
S_b = male bovine waste
CH_4 = methane gas
B_c = crystal ball
D_m = management directive

Basis for the Rule

The rule is based on empirical observations, a cynical view of management theory, and the second law of thermodynamics.

Cautions and Useful Range of the Rule

It works for equally well for systems, component, and study contracts.

Be wary of your perceptions about your competitor's price; they are working as diligently as you to reduce costs.

This applies to price; the actual cost requires quite a different equation.

Usefulness of the Rule

This rule is useful for understanding some of the considerations that may go into your price or the price of your competitors.

Notes and Explanation

This is not all tongue-in-check; electro-optics is sufficiently high-tech that the cost is a strong function of "extras." If desired, most systems and components can have costly qualifications, titivations, testing, and features. These may or may not be required for basic operation. Hence, a large part of the price often is based on the (perceived) actions of the competitor, capricious decisions from management, and a tad of bovine waste. Observe that the management directive is only to the 1/2 power. This reflects the fact that, in most technical companies, a management decision can be changed when faced with overwhelming technical evidence.

Source

Rule provided by Grant Milbouer, 1995.

Good Math, Bad Math

Telling Good from Bad

	Math Done Wrong	Math Done Right
Appropriate Math Model	Bad Math	Good Math
Inappropriate Math Model	Very Bad Math*	Bad Math

* AKA Junk Mathematics

Models Misused

- **Ignoring Phase of the Program**
  - **Example: COCOMO II has two versions, Pre Design and Post Architecture**
- **Not fully utilizing all inputs when available**
- **Failing to use a range for inputs in favor of a single value**
- **Setting ranges for estimating error without regard to statistics that are associated with the development of the CERs**

Very Bad Math: A Contrived Example

Number of Contractor O&S Staff	Total Number of O&S Staff	Contractors as a Percent of Total
9	10	90%
200	500	40%
40	50	80%

So on average, contractors represent

$$\frac{90\% + 40\% + 80\%}{3} = 70\%$$ **of the total O&S Staff**

First, the Average Must Be Weighted!

"Creating Customer-Focused Success"

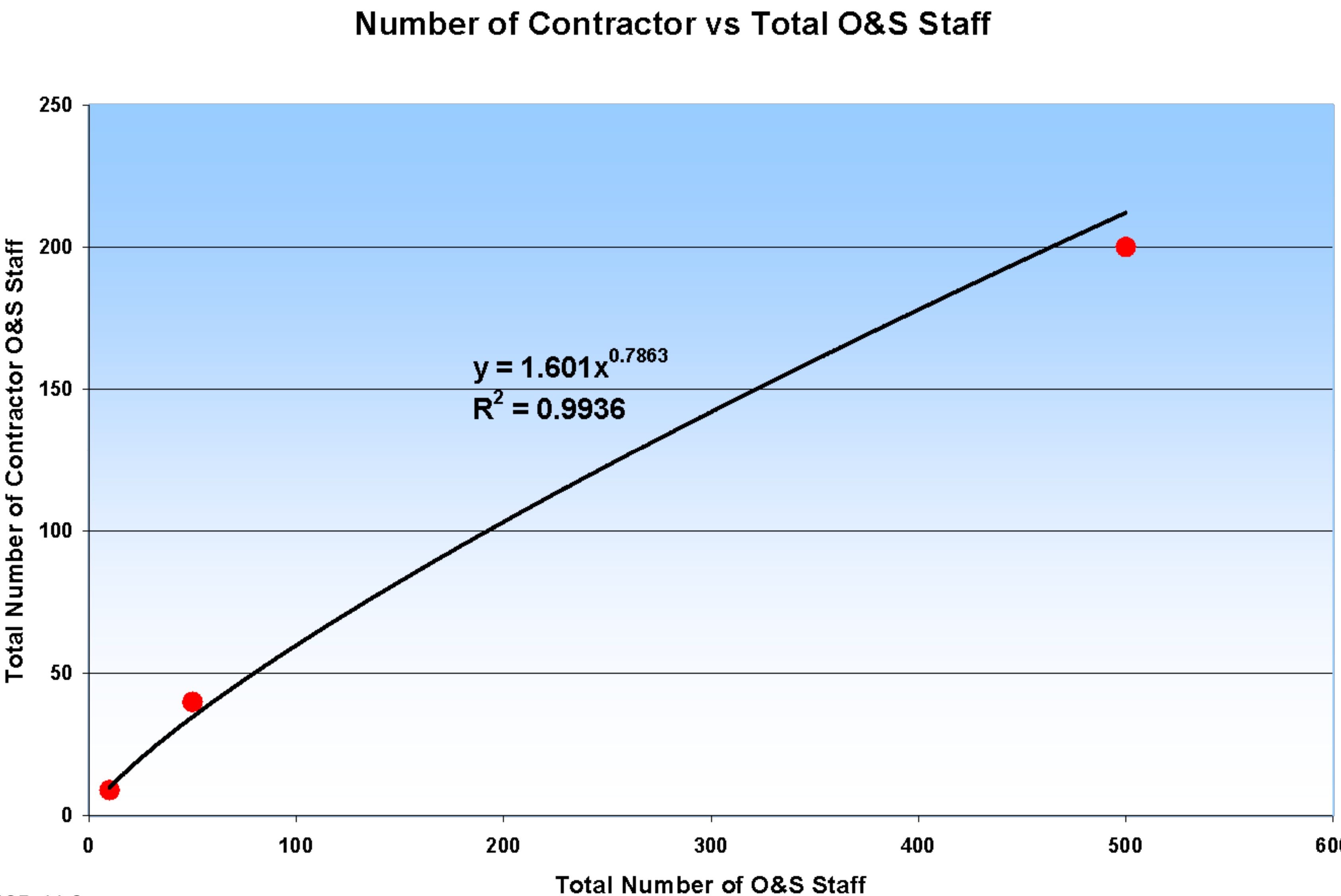
In general: $$\frac{1}{n}\sum_{i=1}^{n}\frac{a_i}{b_i} \neq \frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} b_i} = \frac{\sum_{i=1}^{n} b_i\left(\frac{a_i}{b_i}\right)}{\sum_{i=1}^{n} b_i}$$

In our example, the <u>weighted average</u> is:

$$\frac{\textbf{Total Number of Contractors}}{\textbf{Total Number of O\&S Staff}} = \frac{9+200+40}{10+500+50} = \frac{249}{560} = 44\%$$

"Creating Customer-Focused Success"

Second, a Linear Relationship is Not Appropriate



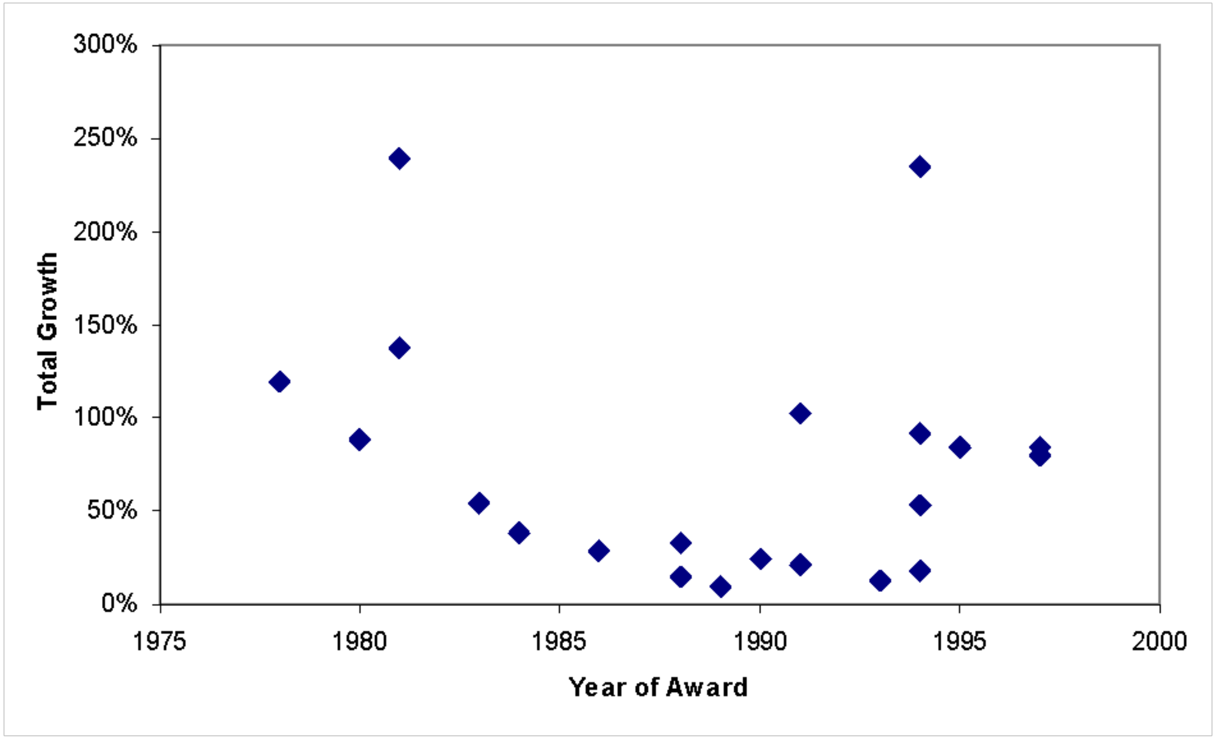
"Creating Customer-Focused Success"

Comparisons

Total Number of O&S Staff	Actual Number of Contractor O&S Staff	Estimated Number of Contractor O&S Staff Using Percent	Delta (Actual - Estimated Using Percent)	Estimated Number of Contractor O&S Staff Using Formula	Delta (Actual - Estimated Using Formula)	Delta (Estimated By Percent - Estimated Using Formula)
10	9	4	5	10	-1	-5
50	40	22	18	35	5	-12
500	200	222	-22	212	-12	10

Very Bad Math: Real Example #1

- **21 total programs**
  - **Mix of government programs**
  - **100% complete (two exceptions)**
- **Final contract values from $40M to $4,000M**
- **Average contract growth 78%**



Real Example #1: Percents Averaged

Program	Total Growth	Incidental ECPs	Technical ECPs Explicit	Technical ECPs Implicit	Cost Variance	Schedule Delay
1	103%	0%	8%	9%	57%	29%
2	119%	7%	5%	7%	100%	0%
3	89%	13%	12%	16%	25%	23%
4	54%	0%	10%	19%	26%	0%
5	39%	4%	1%	8%	26%	0%
6	10%	1%	6%	3%	FP	0%
7	21%	0%	1%	17%	3%	1%
8	15%	0%	12%	2%	0%	0%
9	24%	1%	23%	0%	-1%	0%
10	92%	0%	32%	30%	26%	4%
11	53%	0%	20%	30%	1%	3%
12	13%	0%	11%	-1%	3%	0%
13	80%	0%	5%	7%	53%	14%
14	29%	2%	17%	1%	5%	4%
15	18%	0%	8%	1%	6%	3%
16	235%	0%	16%	11%	206%	2%
17	240%	0%	25%	4%	103%	107%
18	138%	4%	17%	18%	85%	15%
19	33%	8%	13%	7%	0%	4%
20	84%	0%	0%	2%	82%	0%
21	85%	1%	3%	11%	45%	25%
Average	78%	2%	12%	10%	43%	11%

These five averages are computed without weighting

This is the sum of the five averages to the right!

Very Bad Math: Real Example #2

ProgramName	Subsystem	Non-Recurring	First Unit (T1)	NR/T1
Program 1	Subsystem X	13.6	17.9	0.76
Program 2	Subsystem X	0.8	0.6	1.26
Program 3	Subsystem X	6.5	6.2	1.05
Program 4	Subsystem X	3.7	2.1	1.78
Sum		24.6	26.8	

Method 1: Weighted Average = 24.6 / 26.8 = 0.92

Method 2: Simple Average = (0.76 + 1.26 + 1.05 + 1.78 / 4) = 1.21

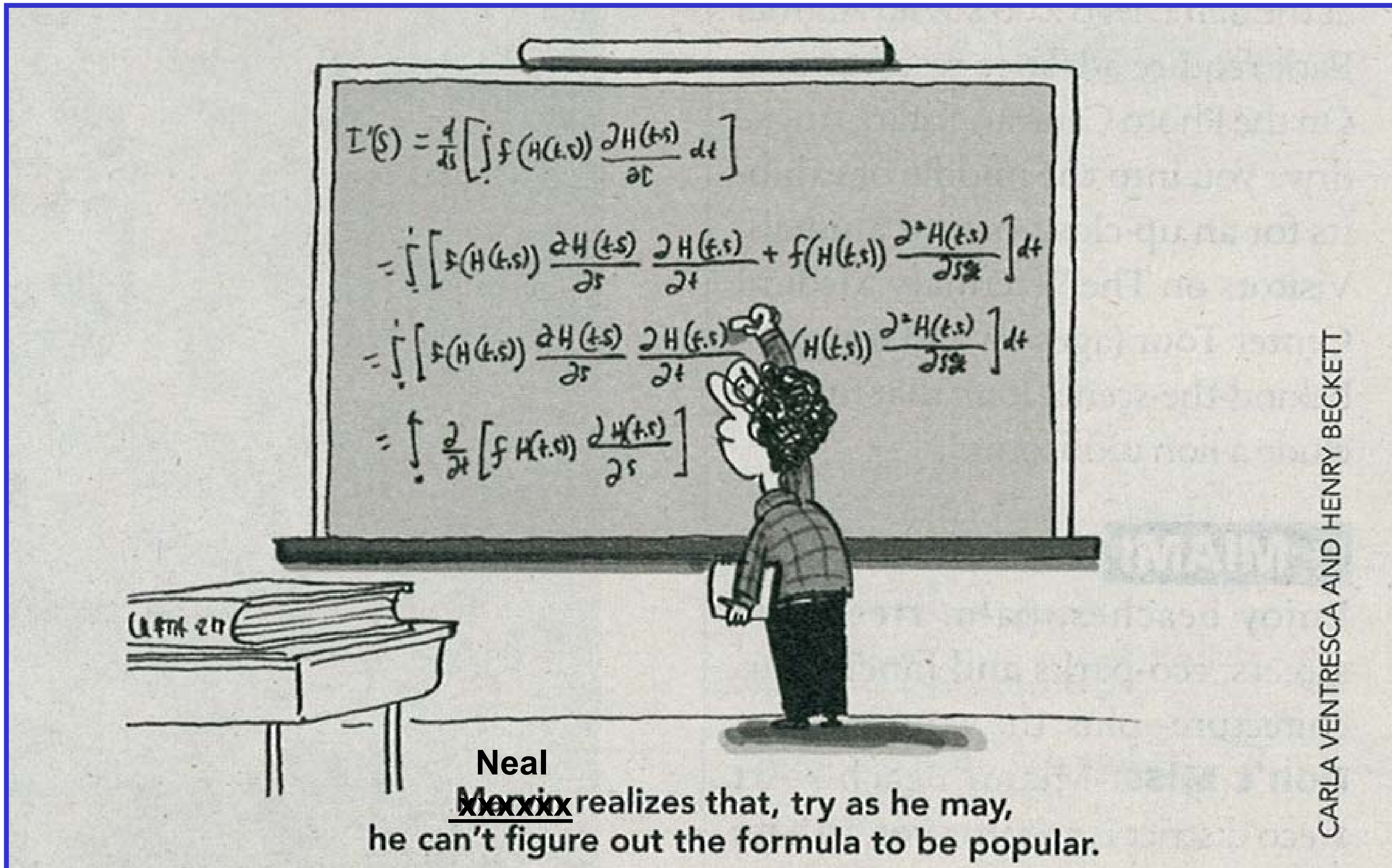
- **Use Weighted Average method to more heavily weight the larger systems**
- **Use the Simple Average method to weight each data point equally**
- **If Method 1 is used the NR Bus estimate would be reduced by $78M (BY03)**
- **Method 2 substantiated and recommended by our Statistician and Scientist**

What's Hot and What's Not Among the Numerate

Out	In	Comments
"Point Estimate"	Estimate with associated confidence level	"Point Estimate" is an undefined term. A proper cost estimate must reflect its probabilistic nature.
Arithmetic Summing of Estimates	Statistical Summing of Estimates	Only means of distributions can be arithmetically summed and nothing else!
OLS and Log-OLS	MPE-ZPB	The world is rarely linear and the assumptions for OLS are even more rarely satisfied. Multiplicative error makes more sense for cost estimates.
IRLS = Minimum Unbiased Percentage Error (MUPE)	IRLS ≠ MUPE	IRLS is biased in small samples (Goldberg and Sperling)
Learning Curves	Quantity as an Independent Variable (QAIV)	Learning slope is a significant cost driver but its selection rarely has a solid justification. Let the data dictate adjustment.
Black Box Cost Models	Transparent Cost Models	Should be obvious why.
Cost as an Independent Variable (CAIV)	Design to Cost (DTC)	Cost is a dependent random variable, not independent. The older terminology was better.
Innumeracy	Numeracy	'Nuff said

The Limitations of Math

Well for One Thing, Math Won't Make You Any Friends Among the Innumerate



Neal
xxxxxx realizes that, try as he may, he can't figure out the formula to be popular.

"Creating Customer-Focused Success"

Right Math Can Be Done Right But Still Yield A Counterintuitive Result

Average Unit Cost (FY07$K) of missile =

$$a+b(LQ^{c})(1+PLQ)^{d}(MV^{e})(f^{FirstLot})$$

Where:

LQ is lot size, PLQ is previous lot quantity

MV is maximum velocity (m/s), FirstLot = 0 if Yes and 1 if No

and

Standard Percent Error = 52.7%
Percent Bias = 0.0%
R^2 = 52.0%

a = -0.62

b = 0.00003625 ← Very small factor; gives more weight to other fit parameters

c = -1.199 ← Large negative value causes total cost to decrease when lot size is increased

d = -0.002

e = 2.747 ← Very large value causes cost to be nearly cubed for each additional m/s of maximum velocity

f = 0.612 ← Value less than one causes the first lot cost to be less than subsequent lots (opposite of learning curve theory)

Good Math Doesn't Guarantee that

- Programs won't overrun
  - Risk impacts may not be adequately considered
  - Budgets may not include a reasonable reserve
- Analyses will be accepted
  - Clients glaze over when math is shown
  - Clients often are inadequately trained in math and statistics
  - More attractive results (i.e., less scary) from less numerate costers prevail

“Creating Customer-Focused Success”

The Inadequacy of Models

- Costs of programs are determined by more than just labor and material
- Drivers include management, political factors, and budgetary environment which don’t readily lend themselves to quantification
- We still lack an overarching theory that would yield better cost models
  - No guarantee that one exists
  - Investigation into similar disciplines such as economics and finance is yielding some ideas that may point to a causal foundation for cost analysis

Other Challenges

- **Small data sets**
- **Assessing estimating errors**
- **Selection of figures of merit for models**
- **Determining most appropriate distributions for modeling impact of risk**
- **Creating joint cost and schedule probability distributions**

Summary

- In life and particularly in cost analysis, innumeracy isn't cool
- As cost analysts introduce more mathematical models and methods, care must be taken to follow the rules
  - Define terms
  - Satisfy hypotheses of theorems before applying them
  - Adopt the most appropriate model formulations
  - Do the calculations correctly
- Failure to do the right math and to get the math right will continue to introduce unnecessary error into our estimates and erode confidence in our numerate clients who know what good is

A Response to the Insistently Innumerate

"You are entitled to your own opinion but not to your own mathematics"...NDH

“Creating Customer-Focused Success”

Acronyms

BY	Base Year
CAIV	Cost as an Independent Variable
CER	Cost Estimating Relationship
COCOMO	Constructive Cost Model
DTC	Design to Cost
ECP	Engineering Change Proposal
ICE	Independent Cost Estimate
IRLS	Iteratively Reweighted Least Squares
MPE	Minimum Percent Error
MUPE	Minimum Unbiased Percentage Error
NDH	Neal David Hulkower
NR	Nonrecurring
NRO	National Reconnaissance Office
O&S	Operations and Support
OLS	Ordinary Least Squares
Ph.D.	Piled Higher and Deeper
POE	Program Office Estimate
QAIV	Quantity as an Independent Variable
T1	Theoretical First Unit Cost
WBS	Work Breakdown Structure
ZPB	Zero Percent Bias