



# Data Analysis in Support of Defensible Estimates

---

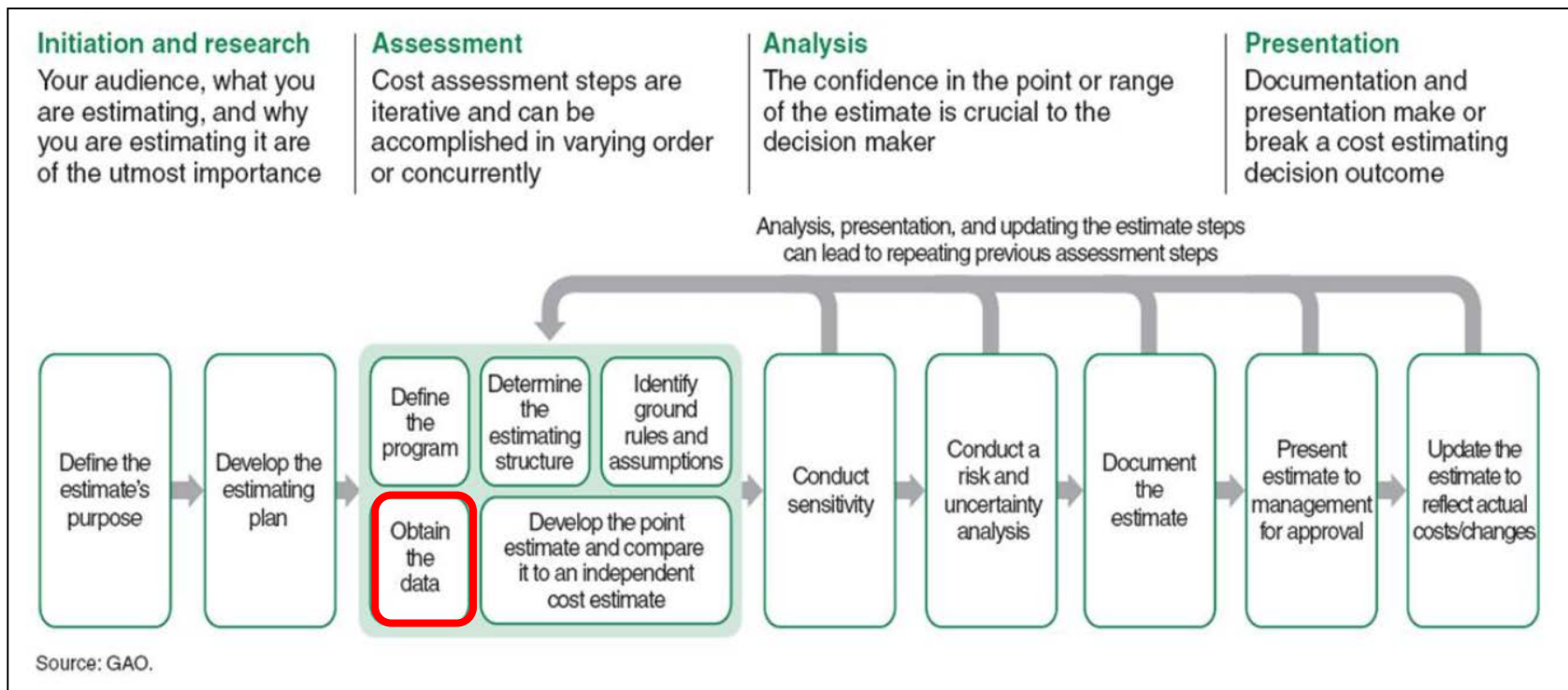
Melissa Winter

ICEAA SoCal 12/17/14

- Introduction to GAO 12-Step Process
- Creating Data Collection Plan
- Data Collection Sources
- Interviewing Sources
- Collecting and Normalizing Data
- Analyzing Data in TrueFindings<sup>®</sup>
- Applying Data Analysis to TruePlanning<sup>®</sup> Estimate
- Validating / Defending TruePlanning<sup>®</sup> Inputs

# Introduction to GAO 12-Step Process

- GAO created the 12-Step process in 1972; still valid today
- Provides repeatable process for developing cost estimates
- This module focuses on Step 6: “Obtain the Data”



- Document the plan...write it down!
- Data collection is lengthy process; continues throughout estimate
- Four main types of data:
  - Cost
  - Schedule
  - Program
  - Technical
- What types of data do you need?
- What types of data are available?
- Develop plan to bridge the gap between these two questions

- Some sources require advance notice for access / clearance
- Sources include:
  - Program Management Plan
  - Cost Analysis Requirements Document (CARD)
  - Integrated Master Schedule
  - Specifications
  - Drawings
  - Size, Weight & Power (SWAP)
  - Labor Rates and Inflation Tables
  - Earned Value data
  - Publicly available sources
  - Paid access sources (i.e. Haystack)

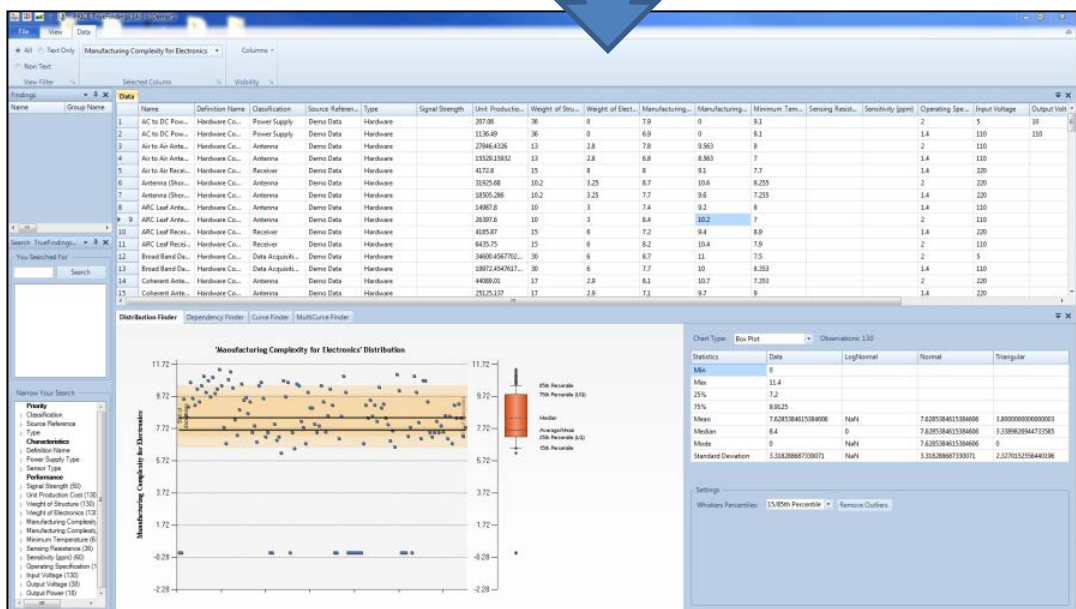
# Interviewing Sources

- Part of the “art” of cost estimating
- Avoid “yes or no” questions or seeking point estimate values
- Bad example:
  - Question: Are we buying 100 widgets this year?
  - Answer: Yes!
- Better example:
  - Question: How many widgets are we buying this year?
  - Answer: We want to buy 100, but we may buy between 80 and 120, based on price.
- Second example provides idea of uncertainty around quantity
- As always, documentation is critical

- Collection is just one step
- Data points are often not in the format we need
- Normalization aligns data points in same format for comparison
  - Cost units: may require inflation or currency adjustments
  - Size units: metric units or imperial units?
  - Groupings: mission types, commodities, recurring vs. nonrecurring costs
  - Technology maturity: solid state electronics or vacuum tubes?
- Document any ground rules or assumptions used for normalization
  - Exchange rates
  - Inflation indices
  - Technology or grouping definitions
- Normalization may occur before, during, or after importing data into TrueFindings®

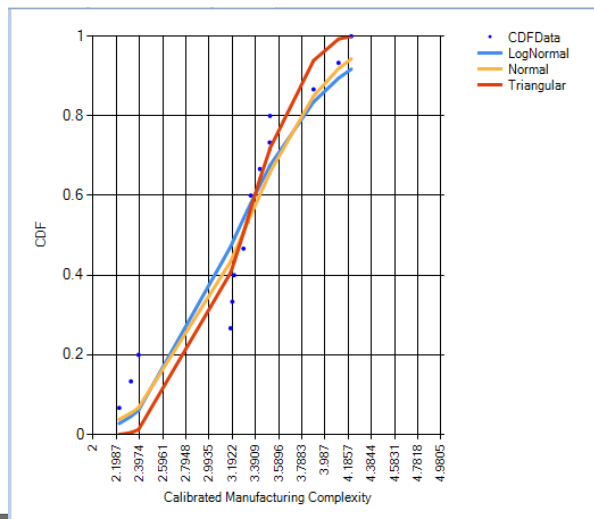
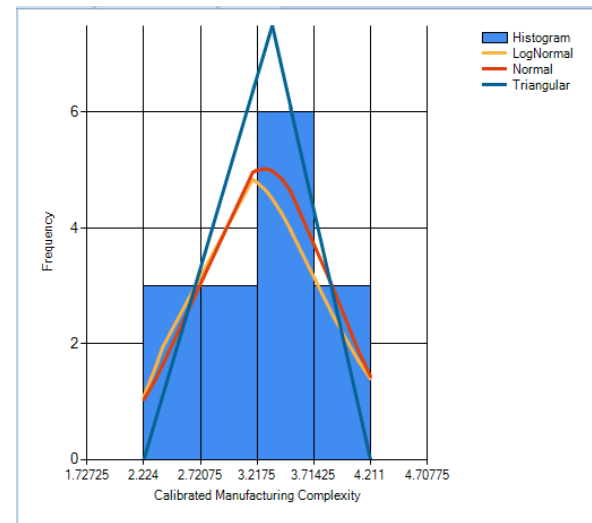
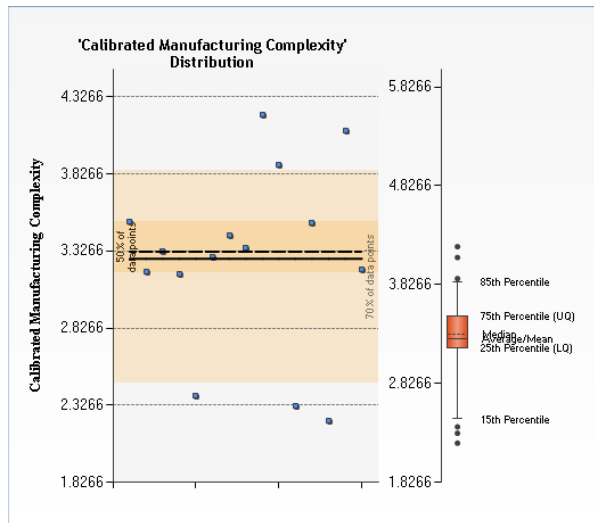
- Analogy: Find similar data point from historical program that directly applies to new program
- Data distribution: Utilize mean or other statistical information from data set to apply to new program
- Correlation: Is there a relationship between variables?
- Bivariate regression: Determine how a change in one independent variable impacts cost or labor requirements
- Multivariate regression: Determine how changes in multiple independent variables impact cost or labor requirements
- Analysis may be conducted on inputs, results or metrics!

- Data analysis tool available with TruePlanning® framework
- Create user defined database in Excel
- Link analysis to estimate



- Research potential input values
- Determine risk and sensitivity analysis ranges
- Compare model results to historical costs

## Boxplot, Histogram, CDF

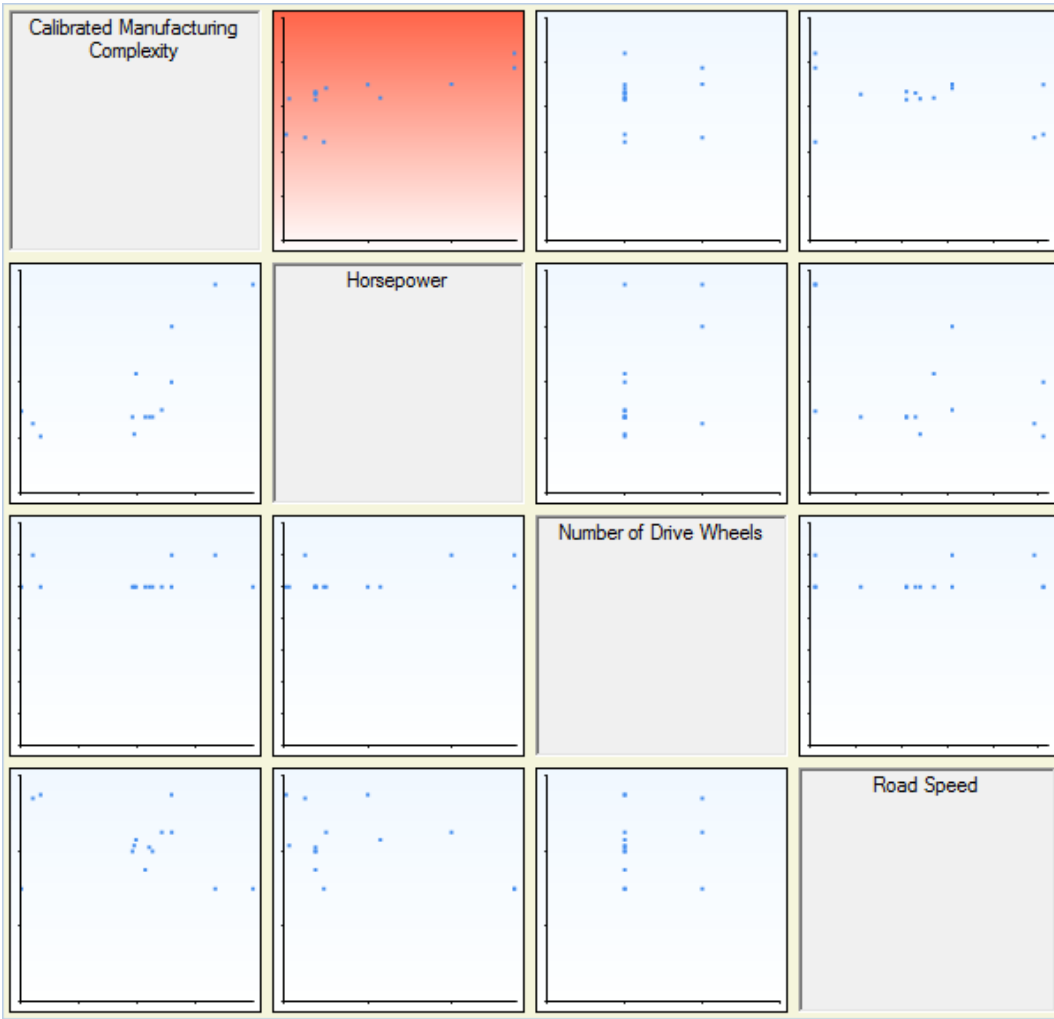


Statistics >>

Chart Type:  Observations: 15

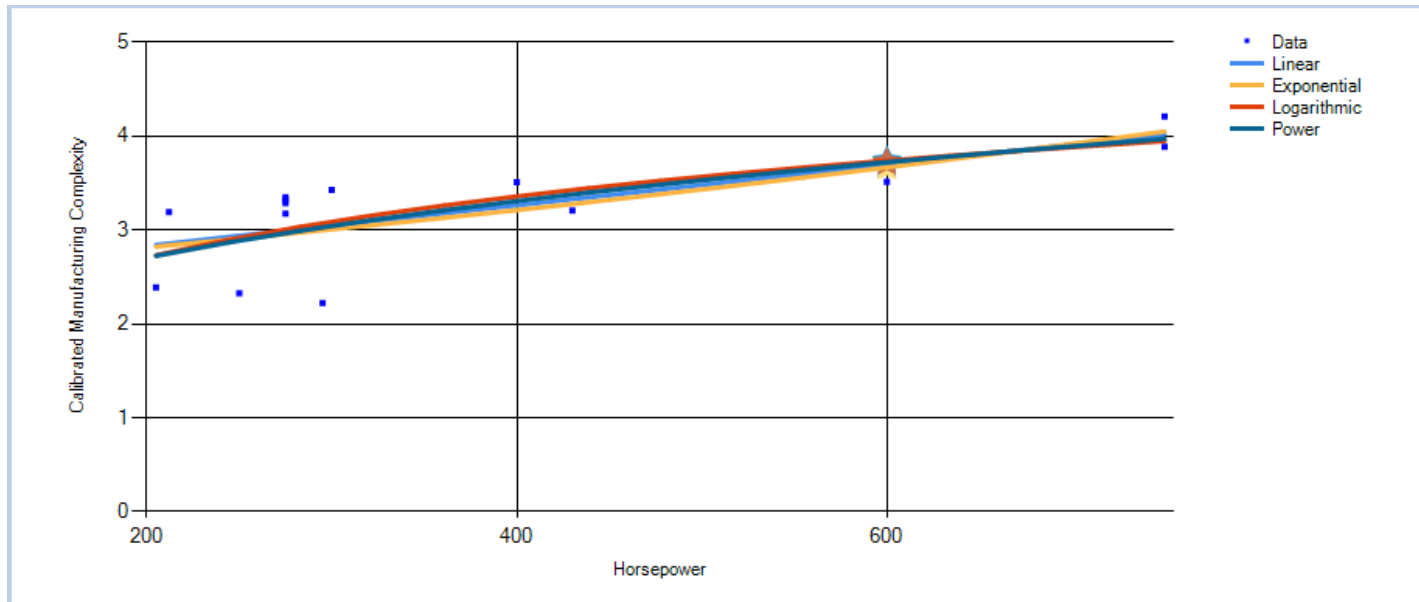
Statistics	Data	LogNormal	Normal	Triangular
Min	2.224			
Max	4.211			
25%	3.177			
75%	3.515			
Mean	3.2754	3.282079495...	3.2754	3.261666666...
Median	3.322	3.221579221...	3.2754	3.281677171...
Mode	3.515	3.103903815...	3.2754	3.35
Standard De...	0.591382134...	0.639054389...	0.591382134...	0.405594676...
Coefficient ...	0.180552645...	0.194710210...	0.180552645...	0.124351970...

## Correlations



Statistics		
Independent Variable	Dependent Variable	R Value
Horsepower	Calibrated Manufactur...	0.717
Calibrated Manufact...	Horsepower	0.717
Road Speed	Horsepower	0.452
Horsepower	Road Speed	0.452
Number of Drive Wh...	Horsepower	0.448
Horsepower	Number of Drive Whe...	0.448
Road Speed	Calibrated Manufactur...	0.389
Calibrated Manufact...	Road Speed	0.389
Road Speed	Number of Drive Whe...	0.078
Number of Drive Wh...	Road Speed	0.078
Number of Drive Wh...	Calibrated Manufactur...	0.025
Calibrated Manufact...	Number of Drive Whe...	0.025

## Regression Analysis



Curve Equation

Define [Horsepower] =  Observations: 14

	Trendline	Solved [Calibrated Manufacturing Complexity]	Equation	R <sup>2</sup>	Standard Error
<input checked="" type="checkbox"/>	Logarithmic	3.74	$0.939089752108198 \cdot \ln([\text{Horsepower}]) - 2.2673384159969$	0.518	0.409
<input type="checkbox"/>	Polynomial	3.722	$-8.46523734007328E-07 \cdot ([\text{Horsepower}]^2) + 0.00298090862217707 \cdot [\text{Horsepower}] + 2.23825312156051$	0.516	0.428
<input checked="" type="checkbox"/>	Linear	3.696	$0.00216225317217645 \cdot [\text{Horsepower}] + 2.39902544377445$	0.515	0.41
<input checked="" type="checkbox"/>	Power	3.724	$0.580022116849256 \cdot ([\text{Horsepower}]^{0.29066533377005})$	0.45	0.437
<input checked="" type="checkbox"/>	Exponential	3.668	$2.46512672312725 \cdot e^{(0.000662357434055087 \cdot [\text{Horsepower}])}$	0.439	0.441

## Multi-Variate Linear Regression Analysis

Distribution Finder
Dependency Finder
Curve Finder
MultiCurve Finder
Findings

Regression

Dependent Variable: Calibrated Manufacturing Comple

Selected Independent Variables:

- Engine Weight
- Horsepower
- Normalized Unit Cost
- Number of Drive Wheels
- Quantity
- Range
- Road Speed
- Total Production Cost
- Volume

MultiCurve Equation

Linear Model Calibrated Manufacturing Complexity = 0.00209450027361841 \* Horsepower - 0.000536695504978165 \* Range - 0.0

Dependent Variable: Calibrated ... 3.676

Independent Variables:

Horsepower	600
Range	300
Road Speed	55

Statistics

**Regression Table**

Statistics	Value
Multiple R	0.7249561148788306
R Square	0.52556136850020829
Adjusted R Square	0.38322977905027078
Standard Error	0.44448298991689711
Observations	14

**Anova Table**

Source	Sum of S...	Degrees...	Mean Sq...	P-Value	F-Statistic
Regressi...	2.188535...	3	0.729511...	0.050525...	3.692513...
Error	1.975651...	10	0.197565...		
Total	4.164187...	13	0.320322...		

**Coefficients Table**

Name	Coeffici...	Std. Err...	t Test	P-value	Upper...	Lower...
Horse...	0.0020...	0.0007...	2.8061...	0.0185...	0.0037...	0.0004...
Range	-0.000...	0.0015...	-0.341...	0.7400...	0.0029...	-0.004...
Road S...	-0.002...	0.0181...	-0.121...	0.9054...	0.0382...	-0.042...
Interce...	2.7020...	0.8422...	3.2082...	0.0093...	4.5785...	0.8254...

Calibrated Manufacturing Complexity

# Applying Data Analysis to TruePlanning®

- Link analysis directly to TruePlanning® inputs or metrics
- Improves validity / defensibility of estimate
- Analysis based on your programmatic data
- Underlying analysis can be viewed by clicking on database icon

The screenshot displays the PRICE TruePlanning 14.0 software interface. On the left, a 'Product Breakdown Structure' tree shows a hierarchy starting with 'Universal Driver Training System Hardware', followed by 'Hardware System Cost Object', 'Hardware Assembly Cost Object', and 'Crew Cab Hardware'. Below these are sub-items like 'Visual System Hardware', 'Motion System Hardware', 'Instructor Station Hardware', 'Instructor Station Server', 'Instructor Station Workstation', 'Instructor Station Network Device (Router)', and 'Instructor Station Peripheral Device (Printer)'. The main window shows the 'Input Sheet: Crew Cab Hardware' with a 'Simple Estimate' view. It includes summary statistics: Cost (\$52,789,781), Project Cost (\$83,774,686), Labor Requirement (620,665.64 hours), and Project Labor Requirement (1,052,800.59 hours). A table below lists various parameters with their values and units. Row 11, 'Manufacturing Complexity for ...', is circled in red, showing a value of 7.869.

	Value	Units	Spread	Notes	Analyzer
1 Start Date	7/1/...				
2 Quantity Per Next Higher ...	1.00				
3 Additional Units					
4 Number of Additional Prod ...	0.00				
5 Number of Additional Prot ...	0.00				
6 Technical Description					
7 Equipment Type	None				
8 Operating Specification	1.20				
9 Weight of Structure	1,000.000	lbs			
10 Weight of Electronics	200.000	lbs			
11 Manufacturing Complexity for ...	7.869				
12 Percent of New Structure	100%	%			
13 Manufacturing Complexity for ...	8.300				
14 Percent of New Electronics	100%	%			
15 Engineering Complexity	1.000				
16 Manufacturing Process Index	0.000				
17 External Integration Complexi...	3.00				
18 External Integration Complexi...	3.00				

- Collect relevant data: cost, programmatic and technical
- Normalize data to comparable terms
- Look for relationships and correlations in data
- Use statistically significant findings to drive inputs of forward estimates