Data Analysis in Support of Defensible Estimates

Melissa Winter
ICEAA SoCal 12/17/14
Agenda

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

- 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
Data Collection Sources

- 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
Collecting and Normalizing Data

- 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®
Analyzing Data in 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!
TrueFindings®

- 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
Distribution Finder

Boxplot, Histogram, CDF

- Boxplot showing the distribution of 'Calibrated Manufacturing Complexity'
- Histogram with data points
- CDF (Cumulative Distribution Function)
- Statistical data:
  - Min: 2.224
  - Max: 4.211
  - 25%: 3.177
  - 75%: 3.515
  - Mean: 3.274
  - Median: 3.222
  - Mode: 3.515
  - Standard Deviation: 0.591382134
  - Coefficient: 0.180552645

© 2014 PRICE Systems, LLC All Rights Reserved  |  Decades of Cost Management Excellence
Dependency Finder

Correlations

| Independent Variable | Dependent Variable          | |R Value|
|----------------------|----------------------------|-------|
| Horsepower           | Calibrated Manufacturing Complexity | 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 Manufact...     | 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 Manufact...     | 0.025 |
| Calibrated Manufact...| Number of Drive Whe...     | 0.025 |
Curve Finder

Regression Analysis

![Graph showing regression analysis for Horsepower vs. Calibrated Manufacturing Complexity with trendlines for Logarithmic, Polynomial, Linear, Power, and Exponential models.](image)

**Curve Equation**

<table>
<thead>
<tr>
<th>Trendline</th>
<th>Solved [Calibrated Manufacturing Complexity]</th>
<th>Equation</th>
<th>R^2</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logarithmic</td>
<td>3.74</td>
<td>0.939089752108198*ln([Horsepower]) - 2.2673384159969</td>
<td>0.518</td>
<td>0.409</td>
</tr>
<tr>
<td>Polynomial</td>
<td>3.722</td>
<td>-8.465227240073281<em>0.07</em>([Horsepower]^2) + 0.0298090862217707*[Horsepower] + 2.23825212156051</td>
<td>0.516</td>
<td>0.428</td>
</tr>
<tr>
<td>Linear</td>
<td>3.696</td>
<td>0.00216225317217645*[Horsepower] + 2.39902544377445</td>
<td>0.515</td>
<td>0.41</td>
</tr>
<tr>
<td>Power</td>
<td>3.724</td>
<td>0.580022116849255*[Horsepower]^0.290665333777005</td>
<td>0.45</td>
<td>0.437</td>
</tr>
<tr>
<td>Exponential</td>
<td>3.668</td>
<td>2.465126723212725<em>e^(0.000682357434055067</em>[Horsepower])</td>
<td>0.430</td>
<td>0.441</td>
</tr>
</tbody>
</table>

Define [Horsepower] = 600

Observations: 14
Multi-Curve Finder

Multi-Variate Linear Regression Analysis

Calibrated Manufacturing Complexity = 0.002089450027361841 * Horsepower - 0.0005386895604978165 * Range - 0.0

<table>
<thead>
<tr>
<th>Independent Variables</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Horsepower</td>
<td>600</td>
</tr>
<tr>
<td>Range</td>
<td>300</td>
</tr>
<tr>
<td>Road Speed</td>
<td>55</td>
</tr>
</tbody>
</table>

Regression Table:

<table>
<thead>
<tr>
<th>Statistics</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple R</td>
<td>0.7249501148786306</td>
</tr>
<tr>
<td>R Square</td>
<td>0.5255613683002086</td>
</tr>
<tr>
<td>Adjusted R Square</td>
<td>0.38322977905079078</td>
</tr>
<tr>
<td>Standard Error</td>
<td>0.444483589916397711</td>
</tr>
<tr>
<td>Observations</td>
<td>14</td>
</tr>
</tbody>
</table>

Anova Table:

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>Degrees of Freedom</th>
<th>Mean Square</th>
<th>P-Value</th>
<th>F-Statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>2.188535</td>
<td>3</td>
<td>0.729531</td>
<td>0.050525</td>
<td>3.692513</td>
</tr>
<tr>
<td>Error</td>
<td>1.975615</td>
<td>10</td>
<td>0.19755</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>4.164187</td>
<td>13</td>
<td>0.320322</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Coefficients Table:

<table>
<thead>
<tr>
<th>Name</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Test</th>
<th>P-Value</th>
<th>Upper Limit</th>
<th>Lower Limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Horsepower</td>
<td>0.0020</td>
<td>0.0007</td>
<td>2.8061</td>
<td>0.0185</td>
<td>0.0037</td>
<td>0.0003</td>
</tr>
<tr>
<td>Range</td>
<td>-0.0000</td>
<td>0.0015</td>
<td>-0.3410</td>
<td>0.7400</td>
<td>-0.0029</td>
<td>-0.0042</td>
</tr>
<tr>
<td>Road Speed</td>
<td>-0.0002</td>
<td>0.0181</td>
<td>-0.1210</td>
<td>0.9054</td>
<td>-0.0181</td>
<td>-0.0121</td>
</tr>
<tr>
<td>Intercept</td>
<td>2.7020</td>
<td>0.8422</td>
<td>3.2082</td>
<td>0.0093</td>
<td>4.5785</td>
<td>0.8254</td>
</tr>
</tbody>
</table>
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
Summary

- 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