ESTIMATING CONTRACT TIMES FOR TRANSPORTATION PROJECTS: CREATING A STATISTICAL MODEL TO ESTIMATE TIMES USING BID QUANTITIES

CSCE, MAY 2017
GUILLERMO NEVET
PAUL GOODRUM
UNIVERSITY OF COLORADO BOULDER
Presentation Agenda

- Problem statement.
- Current methods for time estimation
- Explain the development process for the statistical model.
- Path forward.
Problem Statement

Low accuracy

Lack of commitment
Approved FHWA Contract Time Determination Techniques

- Bar Charts Method
- Critical Path Method
- Estimated Cost Methods
First Steps

- Clean and Organize Data to get same units and grouping similar variables (e.g. piping)
- Convert Engineer’s Estimate to 2015 USD using the NHCCI index
  (https://www.fhwa.dot.gov/policyinformation/nhcci.cfm)
The data

<table>
<thead>
<tr>
<th>Bid Item</th>
<th>Freq.</th>
<th>Bid Item</th>
<th>Freq.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crushed aggregate course</td>
<td>177</td>
<td>Commercial Asphalt mix 3_4</td>
<td>20</td>
</tr>
<tr>
<td>Plant mix 3 4</td>
<td>96</td>
<td>Plant mix 9mm</td>
<td>19</td>
</tr>
<tr>
<td>Excavation unclassified</td>
<td>92</td>
<td>Commercial mix 3 8</td>
<td>6</td>
</tr>
<tr>
<td>Gen. Asphalt commercial mix</td>
<td>82</td>
<td>Plant mix 1 2</td>
<td>4</td>
</tr>
<tr>
<td>Special borrow neat line</td>
<td>67</td>
<td>Plant mix 3 8</td>
<td>4</td>
</tr>
<tr>
<td>Embankment in place</td>
<td>57</td>
<td>Commercial mix mm</td>
<td>3</td>
</tr>
<tr>
<td>Steel</td>
<td>41</td>
<td>Concrete class structure</td>
<td>2</td>
</tr>
<tr>
<td>Excavation borrow</td>
<td>32</td>
<td>Concrete class deck</td>
<td>1</td>
</tr>
<tr>
<td>Concrete Class DD Bridge</td>
<td>31</td>
<td>Concrete class SD repair</td>
<td>1</td>
</tr>
<tr>
<td>Concrete General</td>
<td>31</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Multivariate Linear Regression

- MLR is similar to univariate linear regression, but instead of having one independent variable, there can be several estimators.
- Our dependent or response variable is charge days and the independent variables are bid quantities, EE, locations, and project types.

```r
> reg.gen<-lm(chgdays ~ sewer + perfpipe + pvc + concpipe + classd + pavmark + muckexc + rockexc + concrete + concpav + strback10 + emban100 + asphalt10+ unclexc10 + strex10+ asphre10 + agg10 + lnu sd10, data = newrun3)
```
Unusual Observations.

---- Unusual Observations

--> Outliers (Excessive Residuals, Y direction)
plot(rstudent(run19.lmo),pch=19) # Plot the Studentized residuals in an Index plot
outlierTest(run19.lmo) # Actual statistical test for largest outlier
influenceIndexPlot(run19.lmo) # As before
marginalModelPlots(run19.lmo, id.n=4)

--> Leverage (Hat Values, X direction)
plot(hatvalues(run19.lmo)) # Plot of all Hat Values
leveragePlots(run19.lmo, id.n = 4) # Specific Leverage plots by predictor (and fitted values) like avPlots()
influenceIndexPlot(run19.lmo, id.n=4) # As before

--> Influence (Y and X direction simultaneously)
(crit.cd <- qf(.5, 3,39, lower.tail = FALSE)) # Critical F for assessing Cook's Distances
plot(cooks.distance(medical2.lmo), type="b") # Plot of Cook's Distances
abline(h=crit.cd, col = "red", lwd=6) # Cut off line (critical values) for Cook's D
influenceIndexPlot(fuel2.lmo, id.n=4) # Again, as before
influencePlot(fuel2.lmo) # A new one and useful too
Variable Transformations

Transform data to fit a linear model if distribution is not normal, for both, the response (dependent) variable and predictor (independent) variables.
Variable Transformation (cont.)

- More advanced method
  - Suggests the power transformation for a variable in presence of all other variables, which is better than older methods that transformed a variable without considering others.

```r
#----- Transformation ----#
#predictor#
boxTidwell(chgdays ~ sewer + perfpipe + pvc + concpipe + classd + pavmark + muckexc + rockex)

#----- response ----#
powerTransform(chgdays ~ sewer + perfpipe + pvc + concpipe + classd + pavmark + muckexc + rockex)
```
MLR model for time estimation

Select project type

Select Project Size*

Input bid quantities

Get Estimated Charge Days

e.g. Resurfacing

e.g. tons of asphalt paving

e.g. 50 charge days

*Based on a standardized Engineer’s Estimate using the NHCCI index

https://www.fhwa.dot.gov/policyinformation/nhcci.cfm
Validation

- Create the model using 80% of the data ($R^2$)
- Use the remaining 20% to validate (median percent error)
- Explained variability: 75%
- K-Fold cross validation

\[
\text{Percent Error} = \frac{|\text{Predicted Value} - \text{Observed Value}|}{\text{Observed Value}} \times 100
\]

\[
Y = 44.532 + 9.253E - 6 \times X_1 + .008 \times X_2 + .001X_3 + 5.421E - 5 \times X_5 + .002 \times X_5 + \varepsilon^1
\]

Goodness of fit $F=121.354$, significance = .000, Adjusted $R^2=.746$, Mean Percent Error= 44.59%, Median Percent Error= 29.54%.
Predicted vs. Actual Durations (log scale)
Path Forward

- Taylor et al. (2013) developed a similar tool for Kentucky Transportation Cabinet.
- Creating a multi-state time determination tool (CO, GA, MS, MT)
Path Forward (cont.)

- Develop a Machine Learning Estimation Model (analyzing Case-Based Reasoning and Neural Networks).
Questions?