



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

CSCE, MAY 2017

GUILLERMO NEVETT

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

The screenshot shows the RStudio interface with a data table. The table has columns for 'cont\_id', 'sewer', 'perfpip', 'Bid Item', 'Freq.', 'Bid Item', and 'Freq.'. The data is sorted by frequency in descending order. The 'Bid Item' column contains various construction materials and services, and the 'Freq.' column shows the number of occurrences for each item. The table is displayed in a light yellow background with alternating rows.

cont_id	sewer	perfpip	Bid Item	Freq.	Bid Item	Freq.
1401	C17846	5.8664680	Crushed aggregate course	177	Commercial Asphalt mix 3_4	20
1402	C14606R	3.0445225				
1403	C18085	NA				
1404	C15201S	3.8501475				
1405	C16025	NA				
1406	C17766	NA	Plant mix 3 4	96	Plant mix 9mm	19
1407	C19655	4.9487600				
1408	C17890	NA	Excavation unclassified	92	Commercial mix 3 8	6
1409	C18155-COMBO	NA				
1410	C17622	NA				
1411	C19751	5.3890719	Gen. Asphalt commercial mix	82	Plant mix 1 2	4
1412	C17110	7.0892429				
1413	C17772	NA				
1414	C13898	8.1143246				
1415	C17988	NA	Special borrow neat line	67	Plant mix 3 8	4
1416	C16262	6.3189683				
1417	C13808	7.4151750				
1418	C14986	5.8636312				
1419	C20307	7.0527210				
1420	C14464	NA	Embankment in place	57	Commercial mix mm	3
1421	C19478A	NA				
1422	C19128	NA				
1423	C18070	NA	Steel	41	Concrete class structure	2
1424	C17024	5.6021190				
1425	C17800S	NA	Excavation borrow	32	Concrete class deck	1
1426	C17635	4.2626				
1427	C18095-ALT	NA				
1428	C13970T-ALT A	NA				
1429	C18151	NA	Concrete Class DD	31	Concrete class SD	1
1430	C17718	NA				
1431	C18908-COMBO	NA	Bridge	31	Concrete General	
1432	C13003	NA				
1433	C17735	NA				
1434	C18367	NA				
1435	C17169	7.5368972				
1436	C17757	5.8036280				
1437	C17136R-ALT	4.7706847				
1438	C13166	NA				
1439	C15790A	NA				
1440	C16222	NA				
1441	C15504	NA				
1442	C20306	NA				
1443	C13579-ALT	6.0544395				

# 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.

```
> reg.gen<-lm(chgdays ~ sewer + perpipe + pvc + concpipe + classd + pavmark + muckexc + rockexc + concrete + concpav + strback10 + emban100 + asphalt10+ unclexc10 + strexc10+ asphre10 + agg10 + lnu sd10, data = newrun3)
```

# Unusual Observations.

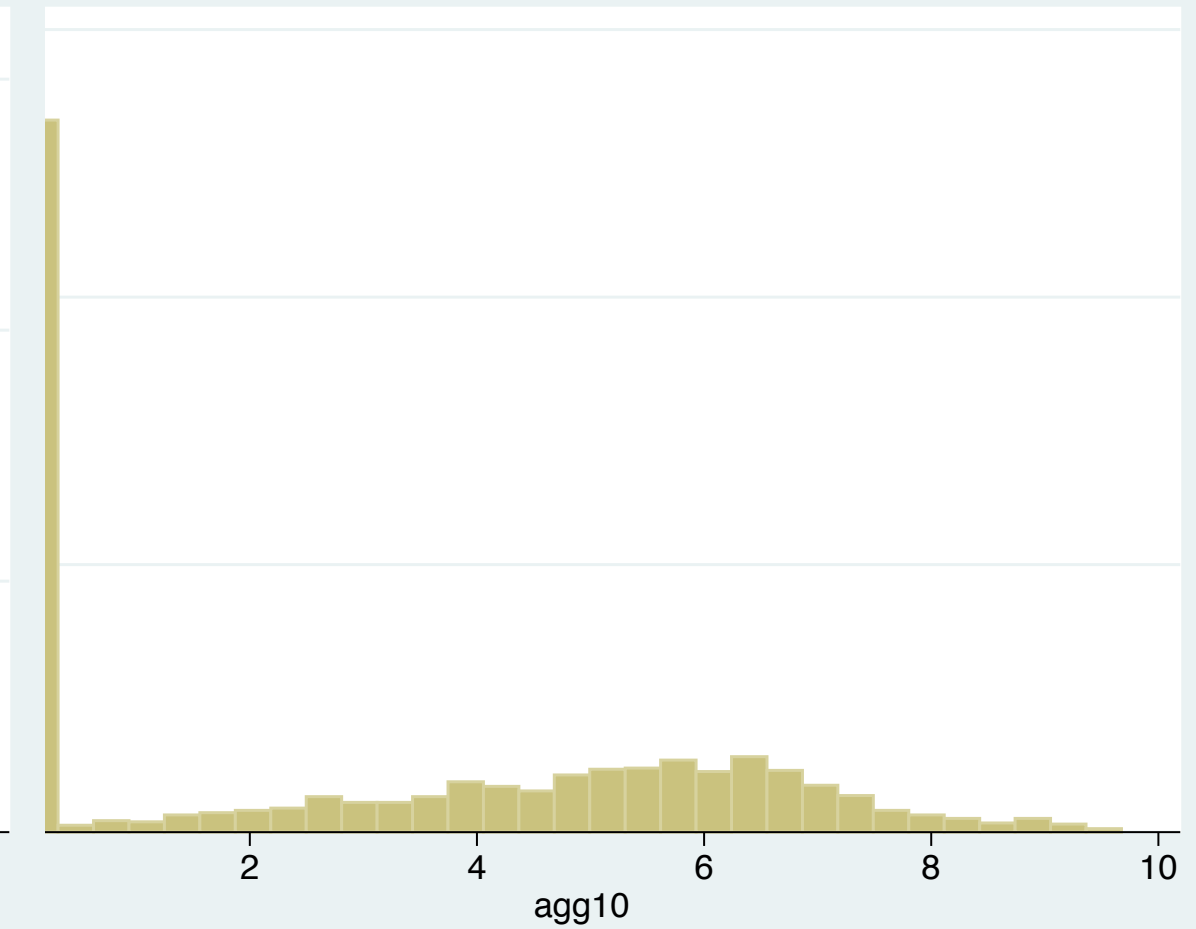
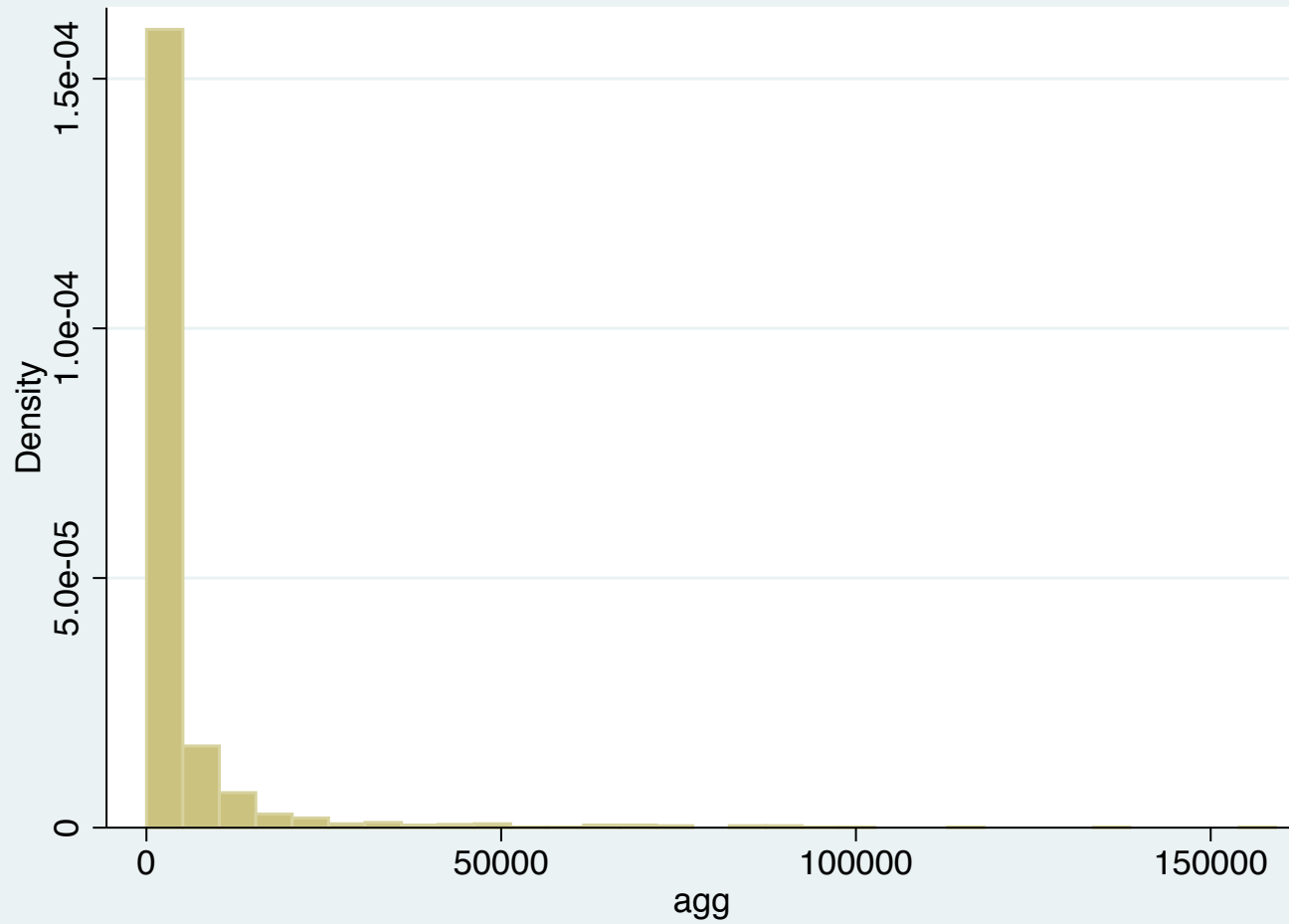
```
##### Unusual Observations #####
#--> Outliers (Excessive Residuals, Y direction)
plot(rstudent(run19.lmo),pch=19)      # Plot the Studentized residuals in an Index plot
outlierTest(run19.lmo)               # Actual statistical test for largest outlier
influenceIndexPlot(run19.lmo)        # As before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

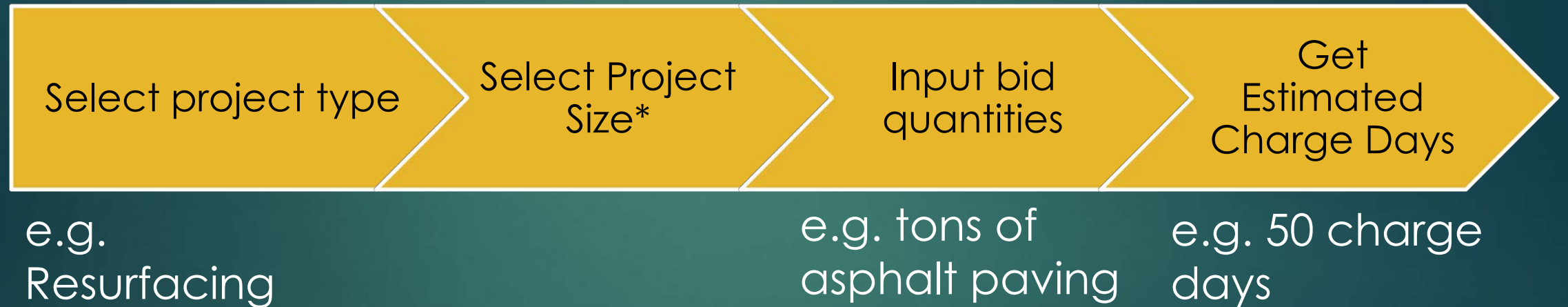


# Variable Transformation (cont.)

- ▶ 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.

```
##### Transformations #####  
#predictor#  
boxTidwell(chgdays ~ sewer + perfpipes + pvc + concpipes + classd + pavmark + muckexc + rockex  
  
#response#  
powerTransform(chgdays ~ sewer + perfpipes + pvc + concpipes + classd + pavmark + muckexc + ro
```

# MLR model for time estimation



\*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)

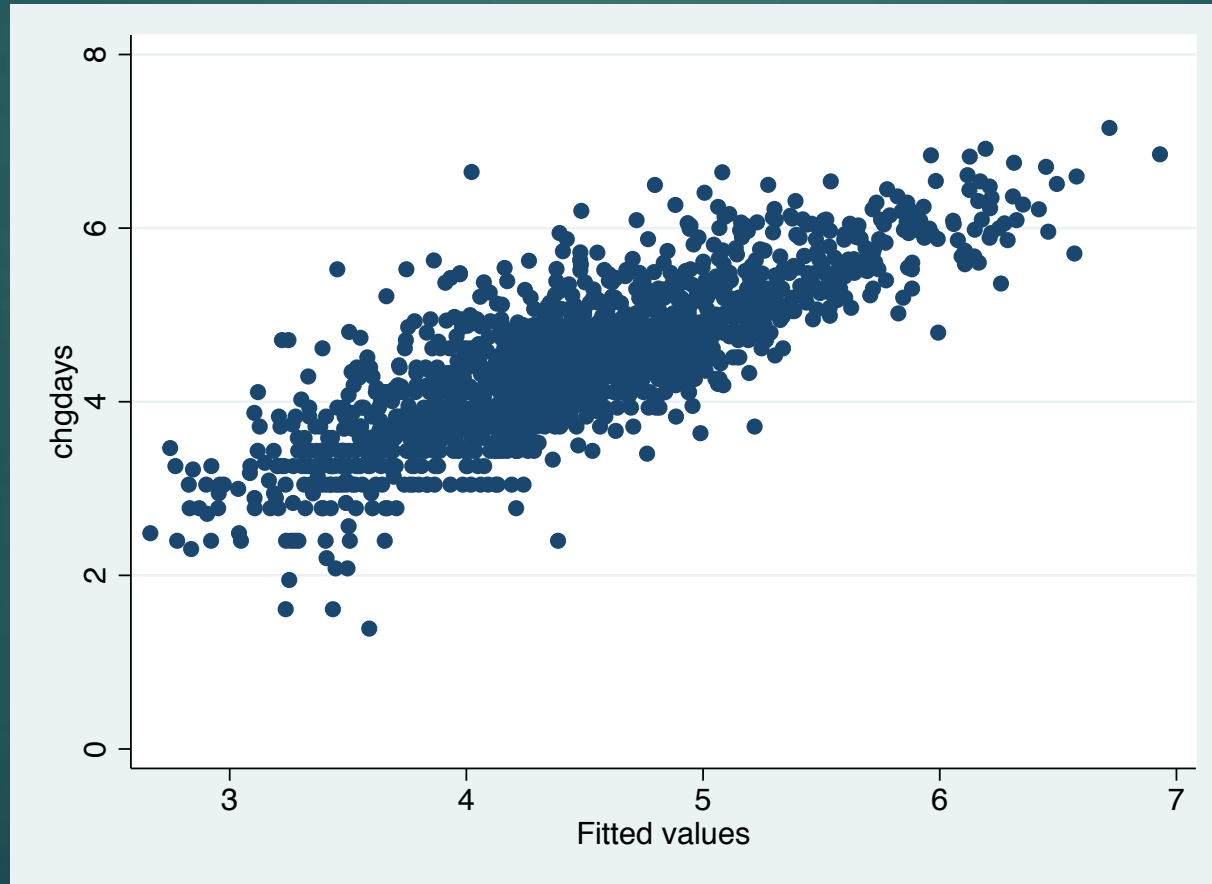
▶ Explained variability = 75%

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

▶ K-Fold

$Y = 44.532 + 9.253E - 6 * X_1 + .008 * X_2 + .001X_3 + 5.421E - 5 * X_5 + .002 * X_5 + \epsilon^1$   
*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?