7/08 Quarterly Report
UST 2004 Tank Baseline Regression Analysis
University of Rhode Island Department of Computer Science and Statistics
Professor R. Choudary Hanumara rch@cs.uri.edu

From the 97 facilities sampled, data on 301 tanks were collected in 4 categories for a number of variables. Not all variables were recorded/or applied to each of the tanks so there are unequal observations. The Table below gives a summary profile of the tanks:

Descriptive Statistics

Variable	Frequency	Average	Percentage
Age of the tank	289	13.7 yrs	
Capacity of the tank	297	8575 gls	
Product stored			
Gasoline	243		80.7
Diesel	33		11.0
Heating oil	7		2.3
Fuel oil	3		1.0
Kerosene +Waste oil	11		3.7
Other	4		1.3
Type of tank			
Steel	28		9.3
Fiberglass	172		57.1
Steel with jacket	91		30.2
Other	10		3.4
Repair status			
No	170		56.5
Yes	11		3.7
Unknown	120		39.8
Col. W			
SW	121		40.3
DW	179		59.7
Tank manifold			
Yes	74		26.0
No	211		74.0
Tank compartment			
Yes	23		7.9
No	269		92.1
Piping			
SW	59		19.6
DW	221		73.4
Unknown	21		7.0
Safe/suction pump			
Yes	27		10.0
No	243		90.0

Interior liner		
Yes	20	6.6
Unknown	281	93.4
Impressed current		
Yes	18	6.0
Unknown	283	94.0
Sacrificial Anodes		
Yes	20	6.6
Unknown	281	93.4

The compliance status of each tank was assessed on a number of variables (checks) under categories Tank Corrosion Protection, Tank Leak Detection, Piping Leak Detection, Spill Prevention and Vapor Recovery. The average number of checks (Y1) and noncompliances (Y2) were 41.5 and 5.2 , respectively.

The University of Rhode Island Department of Computer Science and Statistics investigated four regression models to analyze the count data on non- compliance and the profile variable data on tanks. The independent variables were tank age (X1) and tank capacity (X2), and the qualitative variables as listed in the table above. These qualitative variables were introduced into the model as dummy variables and thus the number of independent variables used in the modeling were 20. The dependent variables in the first two regression models were Y3=Y2/Y1 and Y4=Arcsin(Y3**.5), applying standard linear model techniques. Since the non-compliance data Y2 is count by nature the next two modeling techniques used were Poisson regression and Negative binomial regression. In the Poisson regression model, there was evidence of over dispersion. Hence, Negative binomial regression was selected to be the appropriate choice for modeling. A fit of Negative binomial regression indicated that type of piping (single/double walled), safe or U.S. suction pipe, pressure submersible pump system, and impressed current cathodic protection (tank corrosion protection) are significant variables explaining the non-compliance status of the tanks. We note the same variables were also selected in the linear and Poisson regression modeling.