## Methodology

The task at hand is to try to find a Life Cycle Cost Estimate (LCCE) for a piece of commercial off-the-shelf (COTS) military equipment. We were given a sample of ten products with data corresponding to Operating Capacity, Horse Power, Weight, and Unit Price. We took the approach of trying to create a prediction model using regression analysis given the data from the sample.

Since our goal is to create a LCCE, Unit Price was identified as the response variable and Operating Capacity, Horse Power, and Weight were identified as potential predictors. We first utilized MiniTab to receive feedback on the data. We keyed in on two analysis tools, the Coefficient of Determination (R^2) and the R^2 Adjusted. The R^2 value measure how well future outcomes are likely to be predicted by the model. An R^2 value of 0 represent no correlation and an R^2 model of 1 represents perfect correlation. The R^2 adjusted takes the analysis a step further by factoring in how well a predictor improves the model against what would be expected by chance.

MiniTab optimizes for the highest R^2 value based on Multiple Linear Regression. This yielded and equation using Operating Capacity, Horse Power, and Weight as predictors with an R^2=67.3% and R^2 Adjusted=50.9%. Since MiniTab does not optimize for R^2 adjusted we then proceeded to test the additional five combinations of predictors to try to find the highest R^2 Adjusted. This method would not be realistic in many large data sets with a high number of possible predictors, in those cases a heuristic would need to be utilized. The results of we found interesting but not completely surprising. We found that only using the predictors of Operating Capacity and Horse Power was able to obtain a model with an R^2=65.4% and an R^2 Adjusted=55.6%.

We continued further in our testing and modeled each individual predictor using a quadratic regression model and a cubic regression model. Our results netted only R^2 and R^2 Adjusted values that were smaller than we found using multiple linear regression models. The last thing we tried was to duplicate our results and for this we used Excel and R. Our results matched so we were confident with our results such that we proceeded to look further at the output data we received.

## Results

At this point we have narrowed our search for a LCCE down to two possible models. The requirement for the new equipment states an Operation Capacity of 5700, Horse Power of 83, and Weight of 1350. One risk associated with the requirements stated is that the Weight is slightly higher than any of the sample data. This means that the prediction is being extrapolated and must be considered as a potential technical issue. Below are the two models being considered with the requirements being inputted.

Model 1:

Unit Price = 39367 + 6.85 Operating Capacity - 215 Horse Power + 0.87 Weight

$72312 = 39367 + 6.85 \* (5700) - 215 \* (83) + 0.87 \* (13500)

R-Sq = 67.3% R-Sq(adj) = 50.9%

Model 2:

Unit Price = 35453 + 8.69 Operating Capacity - 169 Horse Power

$70959 = 35453 + 8.69 \* (5700) - 169 \* (83)

R-Sq = 65.4% R-Sq(adj) = 55.6%

The two models yielded similar results with Model 1 resulting in an estimated cost of $72312 vs. $70959 for Model 1. The $1353 difference represents a 1.9% difference. A difference between the two models is Model 1 does characterize that the sample with Operating Capacity of 6000, Horse Power of 83, Weight of 12456, and Unit Price of $68275 but Model 2 does not. Looking at the residual for this sample could give further insight. One aspect of both equations that we found interesting is that Horse Power has a negative result on the Unit Cost. Logically higher Horse Power would result in a higher price but as often found in regression models, prediction models do not always follow what would logically be expected.