A Statistical Evaluation of TriX Software's Performance

In order to increase the efficiency of the loose freight process, recently we had a trial with TriX Software's solution to the problem. We would like to statistically evaluate the efficiency enhancement based on the test results before fully convert to the new software.

Currently, 2% of trucks are more than 37 feet filled and only 7% of trucks are between 35 and 37 feet filled. In the evaluation process, we implemented both Bootstrapping and Monte-Carlo simulation methods to evaluate the statistical significance of the new software. We only focus on trucks that are filled between 35 and 37 feet, and those filled above 37 feet.

- Bootstrapping

Based on the current distribution of fill percentage of trucks, we first simulated a sample of 1,000 trucks' fill levels under the assumption that they are independent to each other. Then we performed Bootstrapping and computed the estimated value, the confidence interval (lower bound & upper bound) and the standard error. Results are summarized as follows:

Confidence Interval	Truck Fill Level	Number of bootstraps	Est. Value	CI Lower Bound	CI Higher Bound	Std. Error
90%	30-35	100	0.198	0.175	0.220	0.001403
90%	30-35	1,000	0.198	0.178	0.218	0.000392
90%	30-35	2,500	0.174	0.154	0.194	0.000238
90%	30-35	5,000	0.201	0.181	0.222	0.000178
90%	<10	100	0.085	0.072	0.100	0.000830
90%	<10	1,000	0.096	0.081	0.112	0.000303
90%	<10	2,500	0.091	0.076	0.106	0.000179
90%	<10	5,000	0.084	0.070	0.098	0.000123

From the table we can see that as the number of bootstraps increases, the range of the confidence interval doesn't change much, but the sizes of the standard errors decrease, indicating that we are getting more confident about where the true percentage of the specified fill level lies. Such behavior is in line with what we have expected.

Similarly, we simulated the case if 4,989 trucks went out, which is the same case as test day. Results are presented as follows:

Confidence Interval	Truck Fill Level	Number of bootstraps	Est. Value	CI Lower Bound	CI Higher Bound	Std. Error
95%	35-37	200	0.071	0.064	0.078	0.000266
95%	37+	200	0.021	0.017	0.025	0.000145
90%	35-37	200	0.071	0.066	0.077	0.000250
90%	37+	200	0.021	0.018	0.025	0.000143

The percentages of the '35-37' group and the '37+' group are 0.074(=368/4989) and 0.022(=108/4989) respectively in the test day. Since both of them are within the corresponding 90% confidence interval computed based on the current distribution of fill percentage, we can conclude that TriX's software did NOT increase the efficiency of the loose freight process of each group separately.

- Monte-Carlo simulation

Under this approach, we made a hypothesis that there is no difference between TriX's algorithm and our current solution and tested if the Monte-Carlo simulation result was in accordance with the hypothesis.

Specifically, we created 1,000 simulated lists of 4,989 trucks and computed the probability that the percentages of desired groups are no less than those of the test day. The simulation results are presented as follows:

Truck Fill Level	Number of simulations	Est. Probability
35-37	1,000	0.167
37+	1,000	0.239
35-37 & 37+ jointly	1,000	0.034

From the results we can see, for group '35-37', about 16.7% simulated days using our current plan is better than TriX's performance. Similarly for group '37+', 23.9% of the simulated days has more trucks than TriX's test day. The results indicate that TriX's algorithm is no better than the current one on each group separately.

However, when we look at two groups jointly, the percentage is a lot lower, since we are counting the days that both groups are better performed than TriX's result. Only 3.4% of the simulated days are better than TriX's test day, which means when we consider two groups jointly, Trix's algorithm is better than the current one at 90% confidence interval.

- Conclusion

The above methods we adopted involved stochastic simulation and sampling to derive useful statistics from limited test data. Bootstrapping technique is especially helpful as it can provide satisfactory estimates even when the sample size is small. But it can be biased when the sample is severely skewed. Monte-Carlo method can avoid such situation by sampling data from the entire distribution, while its usage is limited when the underlying distribution is unavailable.

Both approaches prove that TriX's algorithm doesn't increase the percentage of either '35-37' group or '37+' group separately. However, in the context that our goal is to increase the percentages of both '35-37' and '37+' groups, we can conclude from the Monte-Carlo simulation results that TriX's software does make a difference at 90% confidence level. Therefore, we would recommend converting to TriX's software.