

Forecasting Vehicle Quality: Developing a dynamic data-driven method to identify homogeneous sub-populations of vehicles

Sotirios Loustas*

March 2017

Abstract

This thesis aims to develop a dynamic data-driven method to identify homogeneous sub-populations/clusters of vehicles to be used in Scania's Basic Warranty Forecast but also for e.g. EPC and contract calculations. The main goals are:

1. To define “good enough” homogeneous sub-populations/clusters as well as minimum population size.
2. To identify factors governing lambda and failure rate of homogeneous sub-populations/clusters of complete vehicles
3. To identify suitable method(s) to find homogenous sub-populations/clusters of complete vehicles
4. To identify homogeneous sub-populations/clusters of complete vehicles.

To achieve those targets, data from different sources are combined in a way to maximize inclusion of recent events and a variety of clustering algorithms are applied (k-means, ward's algorithm, EM algorithm). The results are then validated using cluster validation metrics and also, empirically.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden.
E-mail: sotloustas@gmail.com. Supervisor: Taras Bodnar.