EPAct/V2/E-89: Assessing the Effect of Five Gasoline Properties on Exhaust Emissions from Light-Duty Vehicles Certified to Tier 2 Standards

Final Report on Program Design and Data Collection

Appendix D
Identification of Extreme Values for Round Robin Laboratory Tests

Assessment and Standards Division
Office of Transportation and Air Quality
U.S. Environmental Protection Agency

National Renewable Energy Laboratory
U.S. Department of Energy

Coordinating Research Council

NOTICE

This technical report does not necessarily represent final EPA decisions or positions. It is intended to present technical analysis of issues using data that are currently available. The purpose in the release of such reports is to facilitate the exchange of technical information and to inform the public of technical developments.

EPA United States Environmental Protection Agency

EPA-420-R-13-004
April 2013

Identifying Extreme Values for Round Robin Laboratory Tests

The round robin laboratory tests presented independent measurements of 16 EPAct fuel properties for 27 fuels by up to 14 participating laboratories. Four additional fuels (identified as 17, 18, 19 and 29) were analyzed separately but used the below methodology. In all cases outliers were determined by application of ASTM E 178-08.

For each property, charts were prepared for the laboratory measurements for each EPAct fuel. These charts were used to classify the possible outliers, measurements clearly different than the others, as:

- A single measurement (N = 3 or more),
- Two measurements, one on either side of the mean of the measurements (N > 4), or
- Two measurements on the same side of the mean of the measurements (N > 4).

a. For testing a single fuel property measurement, the mean (M) and standard deviation (SD) are computed from all measurements on the property for a fuel. If the measurement being tested (x_i) is the largest, then T is computed from (x_i – M)/SD. If the measurement is the smallest, then T is computed from (M - x_i)/SD.

Table 1 from the standard practice, ASTM E 178-08, for a two-tailed test provides the critical value of T at the alpha=.05 significance level. Extreme values for the 64 tests greater than the critical value for N, the sample size, are removed as outliers.

b. For testing two measurements on either side of the mean, the range of values (w), w = (largest measurement – smallest measurement), is computed. The term, w/SD, is next calculated with the critical values supplied by the 5 Percent Significance Level and the number of observation (N) of Table 3. A total of 15 values were greater than the critical value.

For significant w/SD results, the T-values for the largest and smallest values are compared. If they are about the same, then both the largest and smallest values are removed as outliers. In the one case out of the 15 greater than the Table 3 critical value, the T-value of one measurement is significant (Table 1, alpha=0.05), but the other measurement of the pair is not significant (p > 0.05), only the measurement producing the significant T value is removed.

c. For testing two measurements on the same side of the mean, a variance (V) is computed for all the measurements and a variance ($V_{1,2}$ or $V_{n-1,n}$) is determined for the points excluding the two potential outliers (the two smallest or largest values). The ratios, $V_{1,2}/V$ and $V_{n-1,n}/V$, are determined and compared with the appropriate critical value for the Lower 5% Significance Level and for N observations. If the ratio is <u>less</u> than the critical values for one of the

comparisons, the two outliers are removed, unless one potential outlier of the two was identified as an outlier in a., above.

Seven tests were identified as having two outliers on the same side of the mean. Four of these cases were included in case a., i.e., having a single outlier. The decision to remove only the one point (under a.) for the four cases was evaluated and confirmed by reviewing the charts for the cases. For the other three cases, both outliers were removed.

The above two-outlier tests were not applied if N = 3 or 4. For these cases, it was considered unreasonable that 50% or 66.7% of the data would be removed as outliers.