State of California—Health and Human Services Agency
California Department of Public Health

KAREN L. SMITH, MD, MPH
Director & State Health Officer

Gavin Newsom
Governor

May 10, 2019

TO: Participants in the February 2019 Voluntary Proficiency Test in Forensic Alcohol Analysis

SUBJECT: Assigned Values and Expected Ranges of Results for the February 2019 Proficiency Test in Forensic Alcohol Analysis

Enclosed is a summary of the descriptive statistics for the Department's February 2019 voluntary proficiency test in forensic alcohol analysis. The Department prepared four blood-alcohol pools (01239A, 01239B, 01289A, and 01289B) for this proficiency test. Included in the summary are the target formulation values for the pools, the test pools' true values as determined by the Department's analyses, the peer-group or consensus values and the standard deviations, general descriptive statistics, and graphical summaries of the distribution of participant results. A total of 27 laboratories elected to voluntarily participate in this proficiency test, with 23 laboratories submitting test results.

With the 2017 revisions to the Title 17 regulations, the Department is no longer authorized to evaluate participants' performances on proficiency tests. Instead, staff of each individual laboratory must evaluate the laboratory's results to determine whether they are consistent with expected test results [17 CCR §1220.1 (b)]. The comments below describing the procedures historically used by the Department when evaluating test results are advisory in nature and intended to assist the laboratory staff in evaluating their own results.

Historically, the Department has determined the acceptable limits of performance based on reported results that are within the range representing ±5% of the 99% confidence interval of the peer group mean, where the range has been truncated to two significant figures (Table 1). This range was described as the "Tier #2 interval." The Department also calculated a narrower "Tier #1 interval," which represents the range of reported results that are within ±5% of the 95% confidence interval of the peer group mean where the range is based on the results reported to three significant figures (Table 1). Tier #1 was expected to include those laboratories demonstrating a high degree of accuracy. The second, wider tier was intended to include those laboratories not as close to the central tendency as required by the first tier, but still accurate and therefore adequately competent.

One of the recent revisions to the Title 17 regulations was to permit the expression of results to either two or three decimal places. When reporting results to the second decimal place, the digit in the third decimal place must be deleted [17 CCR §1220.4 (b)]. The regulations are silent with respect to the procedures for determining the third decimal place.

Many participants [10 out of 23] reported results to three decimal places. Under these circumstances, the wider second tier based on two decimal place results, which again historically was used by the Department to evaluate the laboratories' results, is no longer appropriate.

The IUPAC[1] International Harmonized Protocol for the Proficiency Testing of Analytical Chemistry Laboratories (Harmonized Protocol) recommends the use of z-scores for evaluating proficiency test data. However, the Harmonized Protocol notes that that the interpretation of the z-scores is based on the normal distribution of reported results, in which case the z-scores can be expected to follow the standard normal distribution. As indicated in Table 2, none of the results in this proficiency test was found to be normally distributed. Accordingly, the use of z-scores may not be completely appropriate, but they still may be useful to identify outlier and/or warning level results. The expression for calculating a z-score is included in Table 2. Generally a score between -2 and +2 ($|z| \leq 2$) is considered satisfactory or acceptable. A score outside the range -3 to +3, inclusive ($|z| \geq 3$) is considered unsatisfactory or unacceptable and the laboratory must take corrective actions. Z-scores between -3 and -2 or +2 and +3 ($2 < |z| < 3$) are considered questionable and these two ranges should be used as warning limits. Scores within the warning limit ranges in two or more consecutive test events could be considered unacceptable.

The proficiency test results expressed as *z*-scores for the participants whose results were used to determine the peer group mean and statistics in the February 2019 test are summarized in Figure 7[2]. Participants are identified by codes. An enclosure with the current correspondence provides codes for the results submitted by your laboratory.

Another approach for evaluating proficiency test data, which is non-parametric and does not require the data to be converted to a standard normal form, divides the test data at regular intervals or quantiles[3]. The quartile is a type of quantile: the first quartile (Q_1) is defined as the middle number between the lowest number and the median of the data set. The second quartile (Q_2) is the median of the data set. The third quartile (Q_3) is the middle number between the median and the highest number of the data set[4]. The interquartile range (IQR), a measure of the dispersion of the data, is the difference between the upper and lower quartiles ($IQR = Q_3 - Q_1$) and represents 50% of the data range. Boundaries (called fences) are set at $Q_1 - 1.5$ IQR (lower fence) and $Q_3 + 1.5$ IQR (upper fence) to identify potential outliers in the tails of the distribution.

In Figure 5, the quartile data from pools 01239A and 01239B are presented as box and whisker or Tukey plots with the quartiles and fences shown. The IQR is represented as the box. The median of the data is shown by a black line and the mean of the data is shown by a red line inside the box. Lines ("whiskers") are drawn at 10% (lower) and 90% (upper) of the data range. Figure 6, presents the same data for pools 01289A and 01289B. These figures can be used by the participants to evaluate their data.

[1] International Union of Pure and Applied Chemistry (IUPAC)
[2] When calculating z-scores, the Department used the round even mean of the three decimal place duplicate results reported by the participants since this represents the best estimate of the sample concentration.
[3] See Statistics and Chemometrics for Analytical Chemistry Sixth Edition, Miller and Miller (p. 158)
[4] There are two ways to calculate quartiles (generally, percentiles): Standard and Cleveland methods. By default SigmaPlot software uses the Standard method.

A copy of this report is available on Food and Drug Laboratory webpage:

Sincerely,

Clay Larson, Chief
Abused Substances Analysis Section
Food and Drug Laboratory Branch

For questions or additional information, contact the Food and Drug Laboratory Branch:

Phone – (510) 412-6220

Web - https://www.cdph.ca.gov/Programs/CEH/DFDCS/Pages/FALP.aspx

Email - fdlb.info@cdph.ca.gov

Statistical Data for February 2019 Proficiency Test in Forensic Alcohol Analysis

Table 1 CDPH Tier #1 and Tier #2 Acceptable Ranges (grams%)

Pool #	Pool Date Code	Peer Group Mean	Tier #1	Tier #2
#1A	01239A	0.059	0.054 – 0.064	0.05 – 0.06
#1B	01239B	0.127	0.119 – 0.135	0.11 – 0.13
#2A	01289A	0.161	0.150 – 0.172	0.14 – 0.17
#2B	01289B	0.255	0.239 – 0.271	0.23 – 0.27

Table 2 Summary of Test Pool Data

Parameter		Pool1A (01239A)	Pool 1B (01239B)	Pool 2A (01289A)	Pool 2B (01289B)
Pre-distribution Data	Target Value	0.060	0.130	0.160	0.260
	True Value[5]	0.058	0.124	0.157	0.250
	Standard Deviation	0.0003	0.0005	0.001	0.001
Descriptive statistics	Mean	0.059	0.127	0.161	0.255
	Adjusted Mean[6]	0.059	0.126	0.160	0.254
	Standard Error[7]	0.0003	0.0004	0.0007	0.0012
	Median	0.059	0.126	0.160	0.254
	Standard Deviation	0.0016	0.0024	0.0045	0.0071
	Minimum	0.054	0.122	0.152	0.228
	Maximum	0.064	0.136	0.178	0.274
	Count	38[8]	38[8]	38[8]	38[8]
Non-parametric statistics (SigmaPlot)	Q1 (25%)	0.058	0.125	0.158	0.253
	Q3 (75%)	0.060	0.128	0.162	0.256
	IQR	0.002	0.003	0.004	0.003
	Lower Fence	0.0550	0.1205	0.1520	0.2485
	Upper Fence	0.0630	0.1325	0.1680	0.2605
Histogram		Figure 1	Figure 2	Figure 3	Figure 4
	Normal distribution?[9]	No ($p<0.001$)	No ($p=0.00$)	No ($p<0.001$)	No ($p<0.001$)
	Box Plot/Kernel Density Plot (python)	Figure 5	Figure 5	Figure 6	Figure 6
Robust mean, X*[10]		0.0589	0.1263	0.1601	0.2544
Robust standard deviation, σ_{rob}		0.0012	0.0015	0.0300	0.0024
Fitness-for-purpose standard deviation, σ_p[11]		0.0018	0.0034	0.0042	0.0062
Consensus value (X_a) - determined as Mode ($\mu_{1/2}$) of Gaussian Kernel distribution		0.0586	0.1262	0.1599	0.2543
Uncertainty of the consensus value, X_a , S.E.[12]		0.0002	0.00025	0.0049	0.00040
$X_a \pm$ S.E.		0.0586 ± 0.0002	0.1262 ± 0.0002	0.1600 ± 0.0049	0.2543 ± 0.0004
z-score		$z = \frac{X - X_a}{\sigma_p}$	$z = \frac{X - X_a}{\sigma_p}$		$z = \frac{X - X_a}{\sigma_p}$

[5] Based on CDPH's Headspace Gas Chromatographic Method
[6] Mean determined from participant data after the removal of outlier(s) using MAD method (https://www.itl.nist.gov/div898/handbook/eda/section3/eda35h.htm)
[7] Standard Error of the Mean
[8] A total of 27 laboratories participated and analyzed a total of 46 sample sets.
[9] Shapiro-Wilk test used at 0.05 significance level.
[10] Robust mean of the results reported by the participants was calculated using Algorithm A in Annex C of ISO 13528:2005.
[11] The Department has determined a value for σ_p as 2.5% of robust mean for roughly symmetrical distributions based on the uncertainties associated with the reported results on recent tests together with the 5% accuracy and precision standard of performance requirements set forth in the regulations. In case of skewed, non-normal distributions, the revised, derived Horwitz equation (σ_p') is used : $\boldsymbol{\sigma_p' = 0.02*\mu_{1/2}^{\ 0.8495}}$
[12] calculated per recommendation of ISO 13528: $U_{xa} = \sigma_{rob}$/SQRT(N)

Figure 1

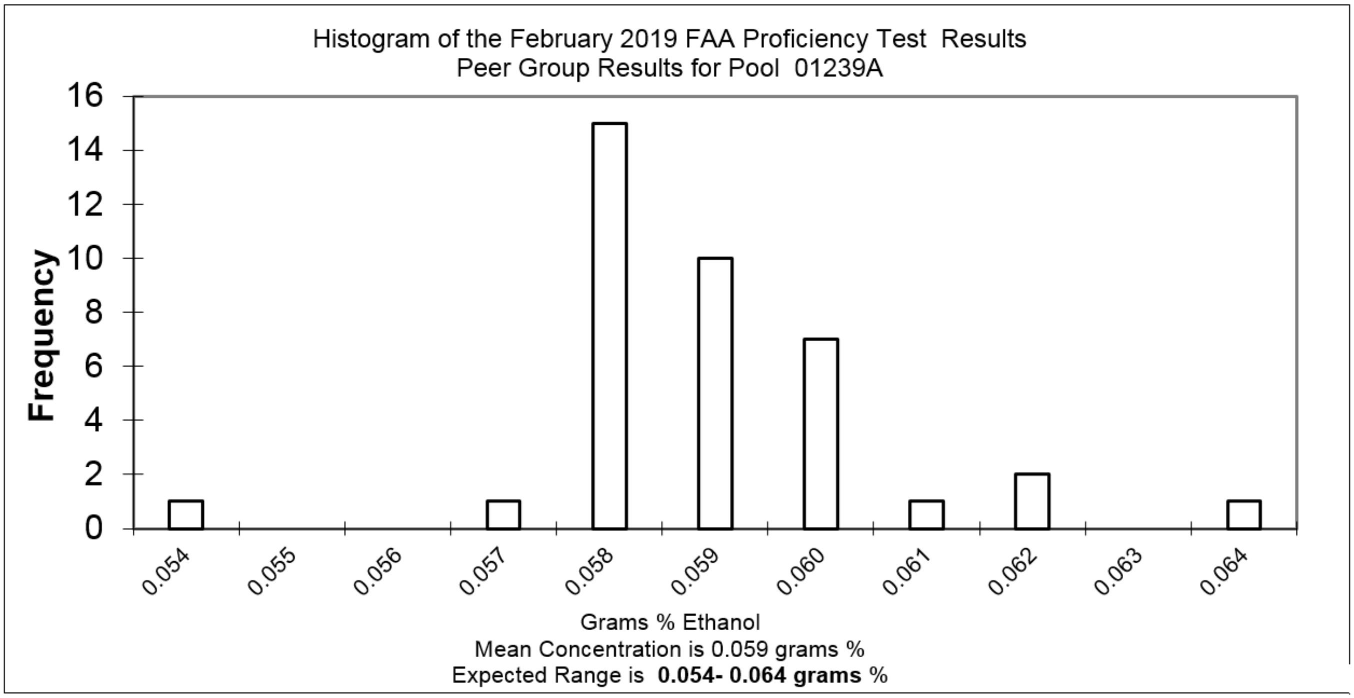


Figure 2

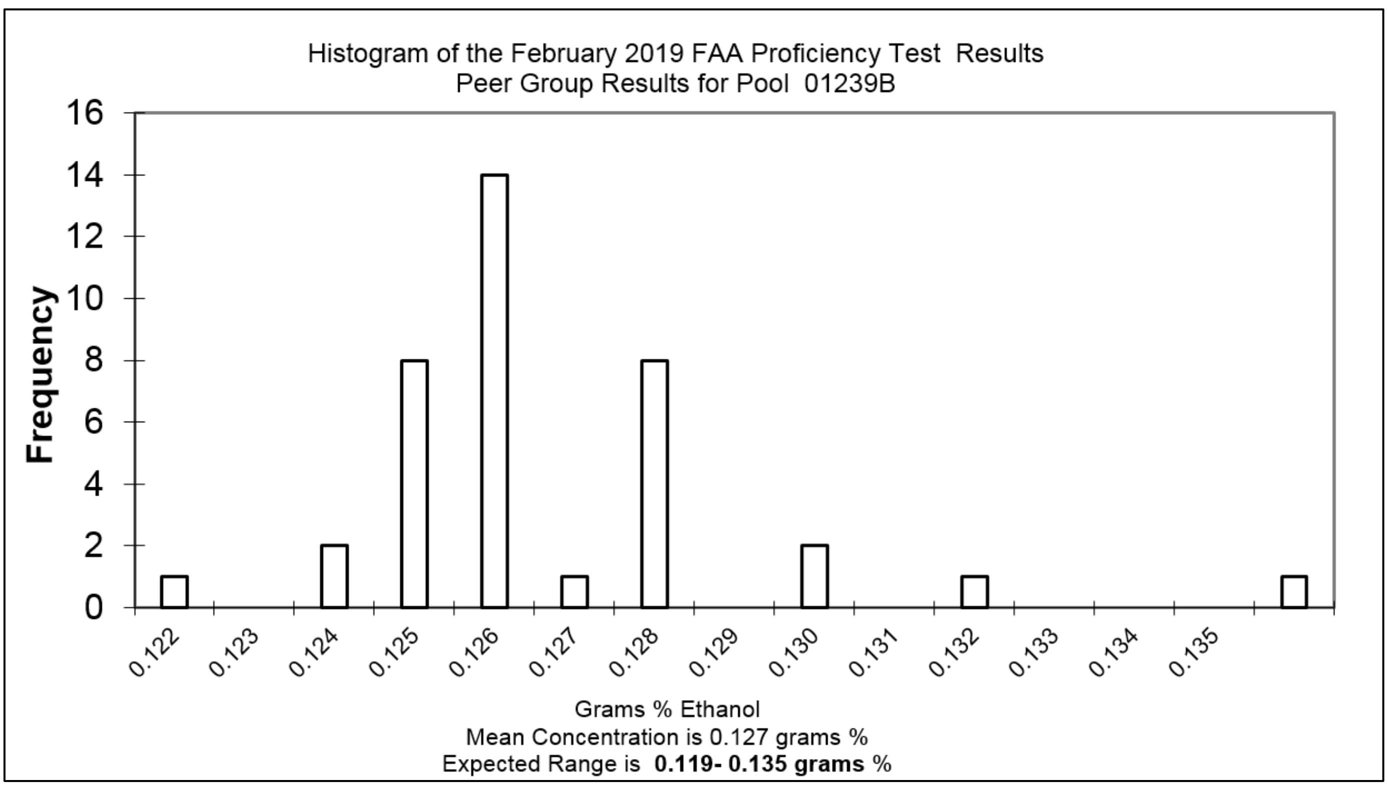


Figure 3

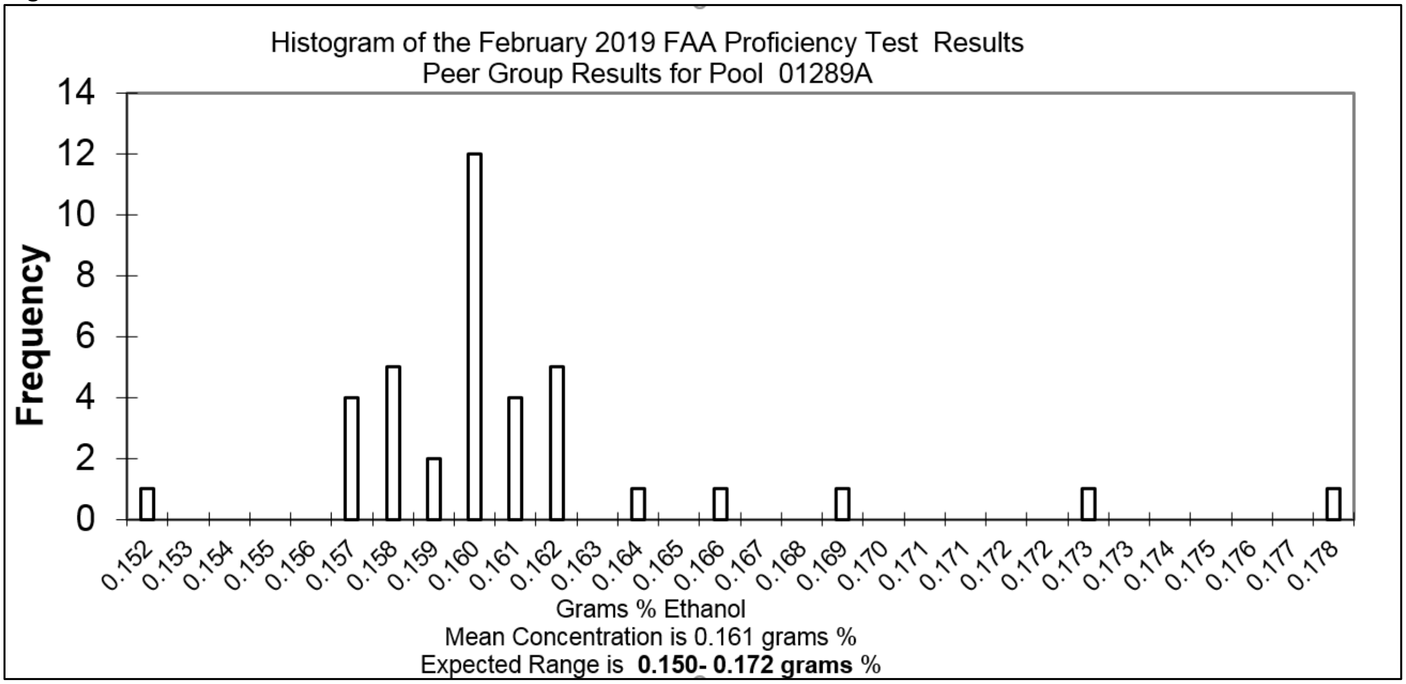


Figure 4

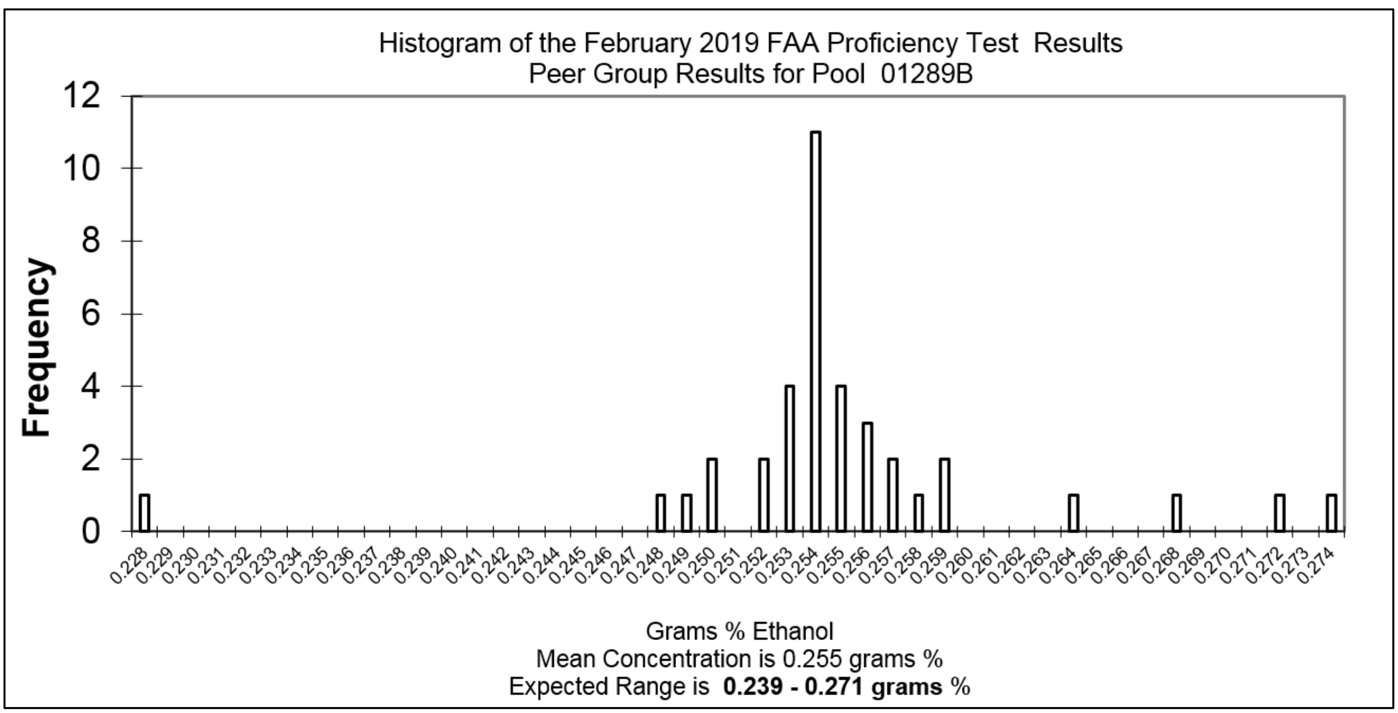


Figure 5 – Python/SigmaPlot Analysis of pools 01239A & 01239B (pools 1 & 2)

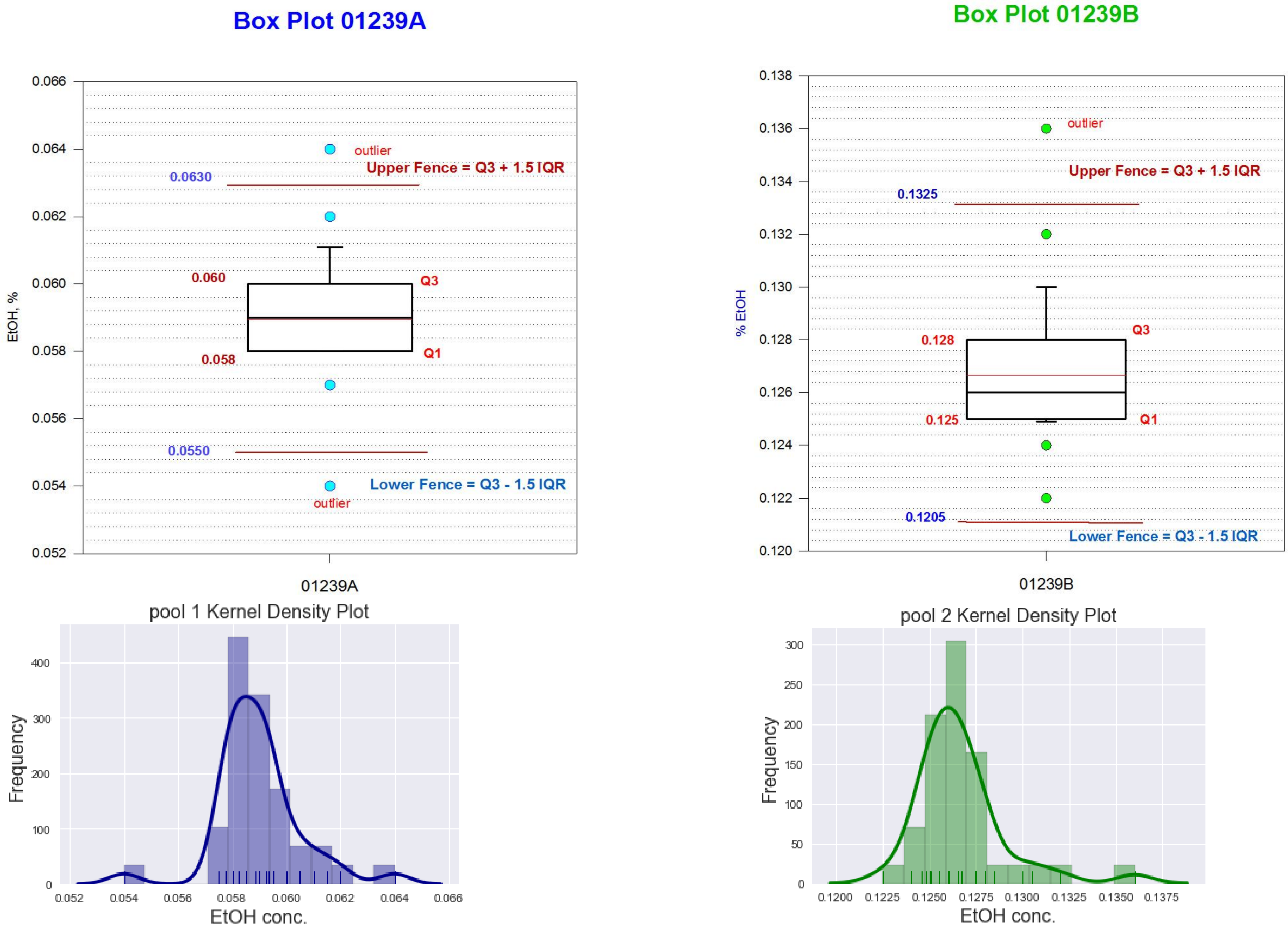


Figure 6 – Python/SigmaPlot analysis of pools 01289A & 01289B (pools 3 & 4)

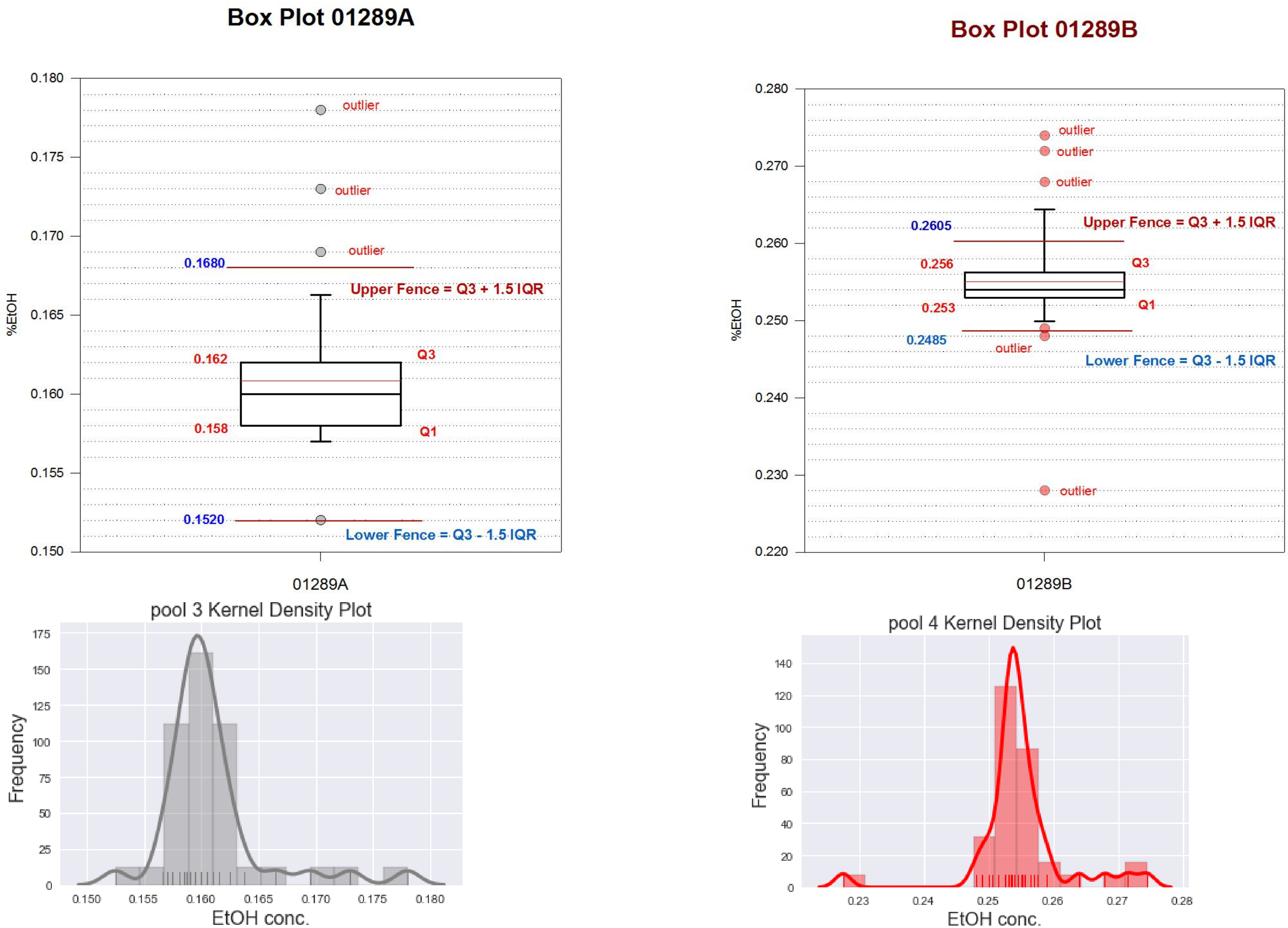
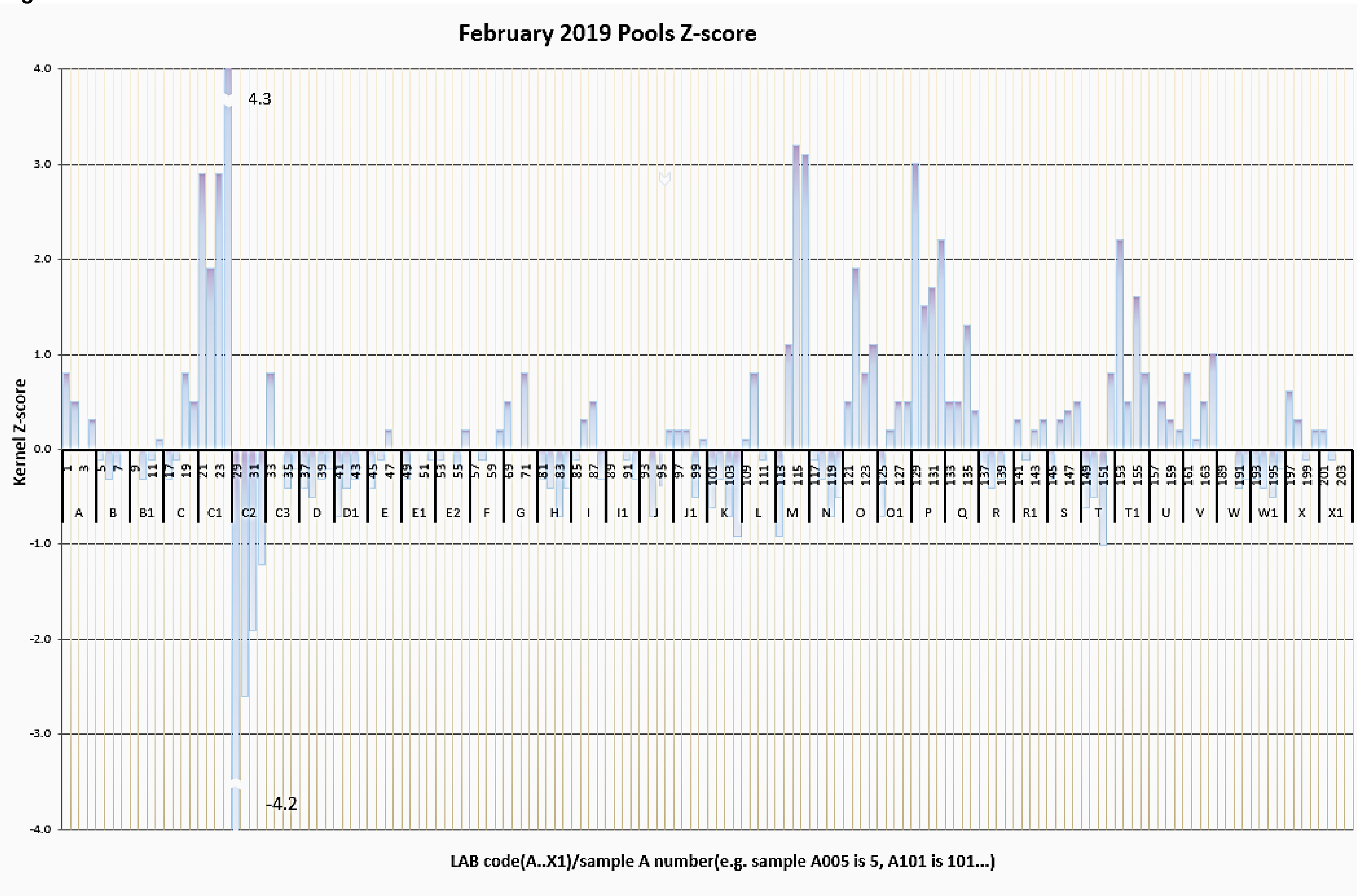


Figure 7


February 2019 Pools Z-score
Kernel Z-score
4.0
3.0
2.0
1.0
0.0
-1.0
-2.0
-3.0
-4.0
4.3
-4.2
A
B
B1
C
C1
C2
C3
D
D1
E
E1
E2
F
G
H
I
I1
J
J1
K
L
M
N
O
O1
P
Q
R
R1
S
T
T1
U
V
W
W1
X
X1
LAB code(A..X1)/sample A number(e.g. sample A005 is 5, A101 is 101...)