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

TOMÁS J. ARAGÓN, M.D., Dr.P.H.
Director and State Public Health Officer

GAVIN NEWSOM
Governor

May 3, 2023

TO: Participants in the February 2023 Voluntary Proficiency Test in Forensic Alcohol Analysis

SUBJECT: Assigned Values and Expected Ranges of Results for the February 2023 Proficiency Test in Forensic Alcohol Analysis

Enclosed is a summary of the descriptive statistics for the February 2023 proficiency test in forensic alcohol analysis. The Department prepared four blood-alcohol test pools (01233A, 01233B, 01303A, and 01303B) 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, and graphical summaries of the distribution of participant results.

With the revisions[1] 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 results are advisory in nature and intended to assist the laboratories 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 the first tier, but still accurate and therefore adequately competent.

One of the 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. The majority of the participants [17 out of 21] 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.

[1] Revised Title 17 regulations filed with the Secretary of State on 1/26/17, with an effective date of 4/1/2017.

The IUPAC 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, the results from all four pools were found to be not 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 for the February 2023 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. The interquartile range (IQR), a measure of the dispersion of the data, is the difference between the upper and lower quartiles ($IQR = Q_3 - Q_1$). 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 data from all pools are presented as box and whisker plots.

- The box extends from the first quartile (Q1) to the third quartile (Q3) of the data.
- The median of the data is shown by a black line spanning the full width of the box and the mean of the data is shown by a red line inside the box.
- The fences are not explicitly shown in Figure 5, instead, the lower whisker represents the lowest data point(s) above the lower fence and the upper whisker represents the highest data point(s) below the upper fence. This figure can be used by the participants to evaluate their data.

The histograms and overlaying Kernel density plots for all 4 pools are depicted by Figure 6; the ticks on the lover part of the graph indicate actual data points.

[2] When calculating z-scores, the Department used (1) the round even mean of the three decimal place duplicate raw results (2) the round even mean of the four decimal place quadruplicate raw results reported by the participants; our simulation studies had shown that this represents the best/unbiased estimate of the sample concentration.

[3] See Statistics and Chemometrics for Analytical Chemistry Sixth Edition, Miller and Miller (p. 158)

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 2023 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	01233A	**0.069**	**0.063 – 0.076**	**0.06 – 0.08**
#1B	01233B	**0.126**	**0.118 – 0.134**	**0.11 – 0.13**
#2A	01303A	**0.162**	**0.151 – 0.173**	**0.14 – 0.17**
#2B	01303B	**0.269**	**0.253 – 0.285**	**0.24 – 0.28**

Table 2 Summary of Test Pool Data

Parameter		Pool1A (01233A)	Pool 1B (01233B	Pool 2A (01303A)	Pool 2B (01303B
Pre-distribution Data	Target Value	0.070	0.130	0.160	0.270
	True Value[4]	0.068	0.124	0.159	0.264
	Standard Deviation	0.0005	0.0008	0.0009	0.0011
Descriptive statistics	Mean	0.069	0.126	0.161	0.269
	Adjusted Mean[5]	0.069	0.126	0.162	0.270
	Standard Error[6]	0.0003	0.0005	0.0007	0.0012
	Median	0.069	0.126	0.161	0.270
	Standard Deviation	0.0022	0.0034	0.0047	0.0074
	Minimum	0.062	0.113	0.144	0.244
	Maximum	0.074	0.132	0.170	0.286
	Count	40[7]	40[7]	40[7]	40[7]
Descriptive statistics (box plot)	Q1 (25%)	0.068	0.124	0.159	0.266
	Q3 (75%)	0.070	0.128	0.164	0.272
	IQR	0.002	0.004	0.005	0.006
	Lower Fence	0.065	0.118	0.152	0.257
	Upper Fence	0.073	0.134	0.172	0.281
Histogram		Figure 1	Figure 2	Figure 3	Figure 4
Normal distribution?[8]		Failed (P=0.002)	Failed (P=0.007)	Failed (P<0.001)	Failed (P=0.020)
Box Plot / Kernel Density plots (Python)		Figure 5,6	Figure 5,6	Figure 5,6	Figure 5,6
Robust mean, X*[9]		0.069	0.126	0.161	0.269
Robust standard deviation, σ_{rob}		0.0014	0.0029	0.0033	0.0057
Fitness-for-purpose standard deviation, σ_p[10]		0.0017	0.0034	0.0042	0.0066
Mode ($\mu_{1/2}$) of Gaussian Kernel distribution		0.070	0.126	0.161	0.269
Consensus value (X_a)[11]		0.069	0.126	0.161	0.269
Uncertainty of the consensus value, X_a, S.E.[12]		0.00051	0.00166	0.00095	0.00129
$X_a \pm$ S.E.		0.0695 ± 0.0005	0.1260 ± 0.0017	0.1615 ± 0.0010	0.2695 ± 0.0013
z-score		$z = \frac{\mathbf{X} - \boldsymbol{X}^*}{\boldsymbol{\sigma_p}}$	$z = \frac{\mathbf{X} - \mathbf{X}_a}{\boldsymbol{\sigma'_p}}$	$z = \frac{\mathbf{X} - \mathbf{X}_a}{\boldsymbol{\sigma'_p}}$	$z = \frac{\mathbf{X} - \mathbf{X}_a}{\boldsymbol{\sigma'_p}}$

[4] Based on CDPH's Headspace Gas Chromatographic Method
[5] Mean determined from participant data after the removal of outlier(s) utilizing "Median Absolute Deviation" or MAD method
[6] Standard Error of the Mean
[7] A total of 21 laboratories participated and analyzed a total of 40 sample sets.
[8] Shapiro-Wilk test used at 0.05 significance level.
[9] Robust mean of the results reported by the participants was calculated using Algorithm A in Annex C of ISO 13528:2005.
[10] 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 **: $\sigma'_p = 0.02 * \mu_{1/2}^{0.8495}$**
[11] $\mu_{1/2}$ or X* was used for skewed / not skewed data distribution respectively
[12] Determined as the Standard Error of Mode using bootstrap simulation technique with bandwidth of $0.75 * \sigma_p$

Figure 1

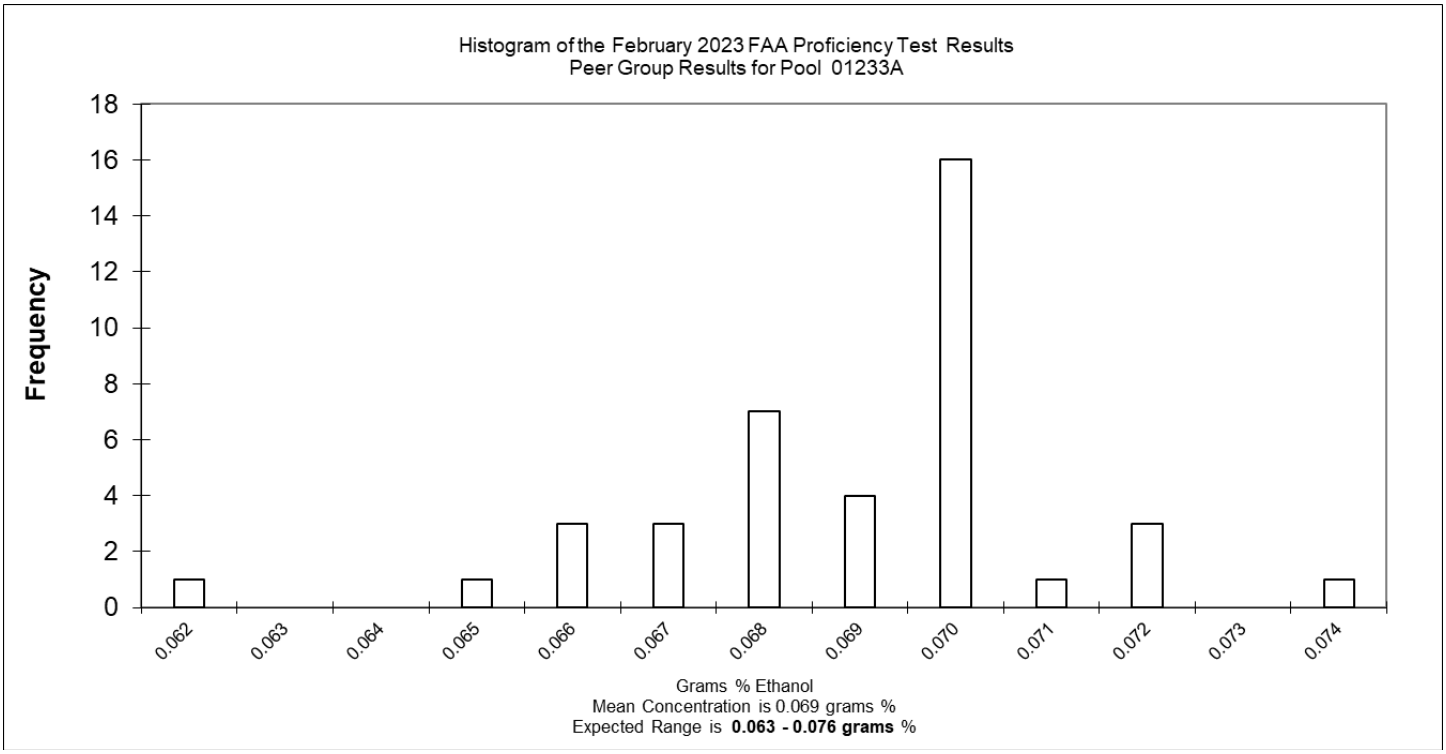


Figure 2

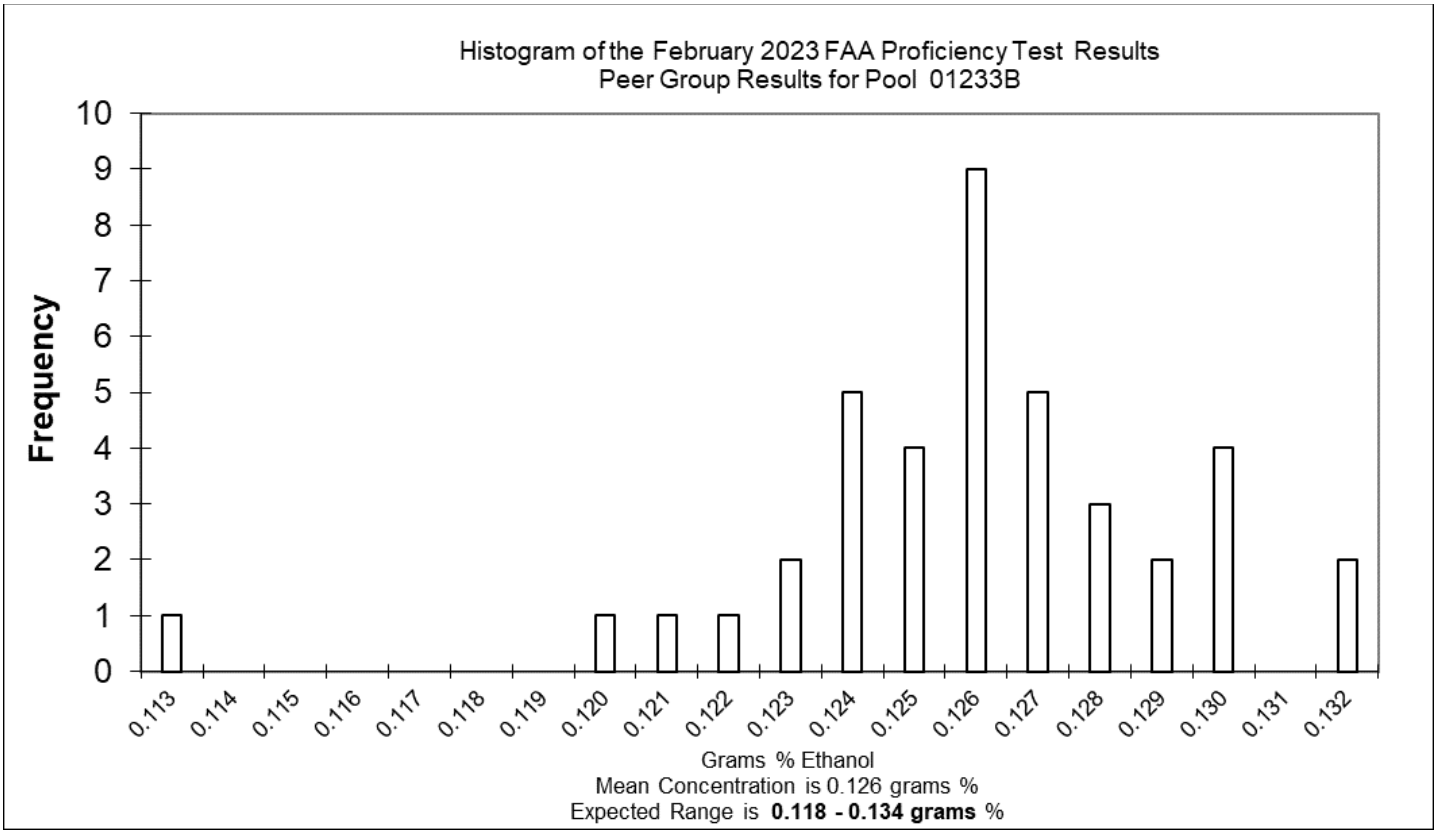


Figure 3

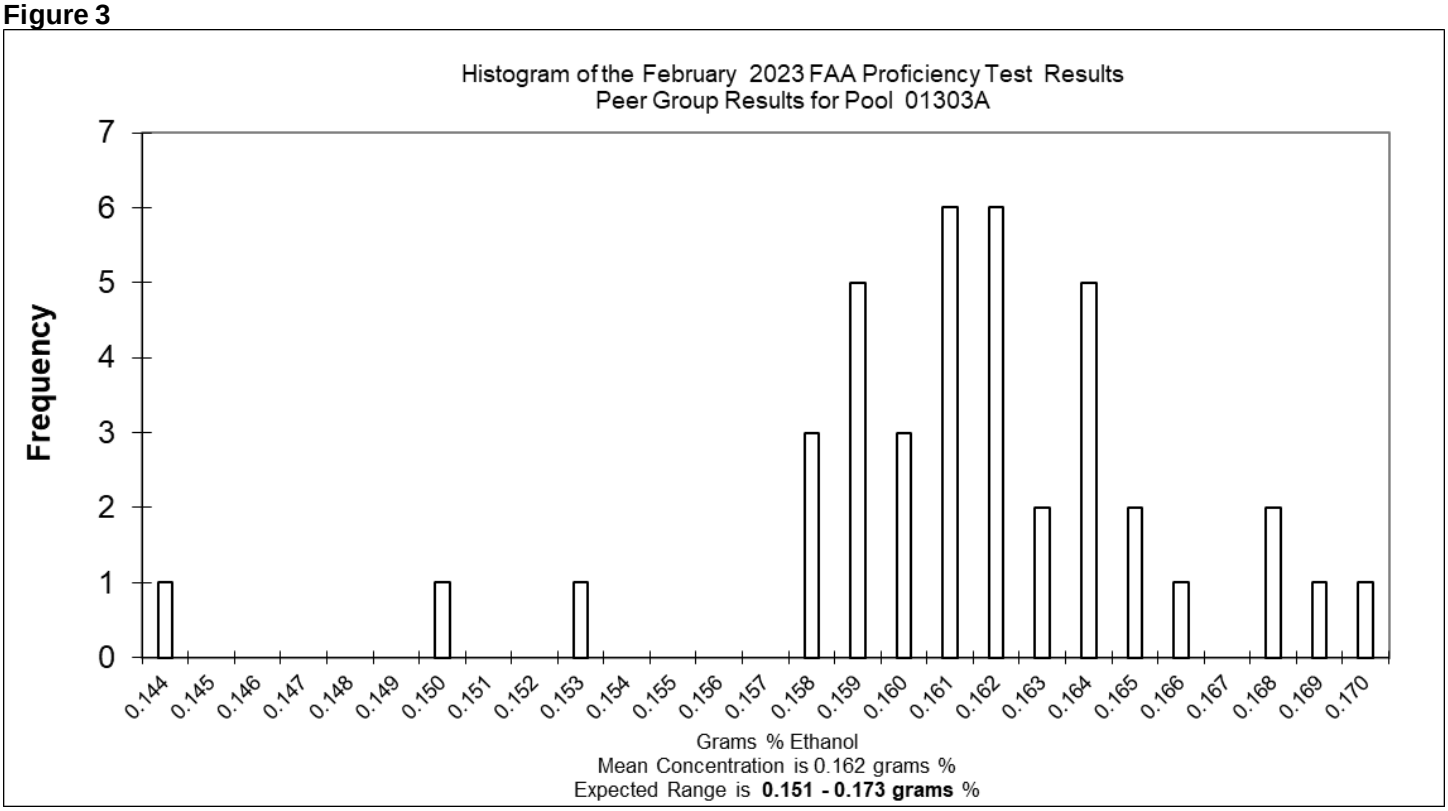


Figure 4

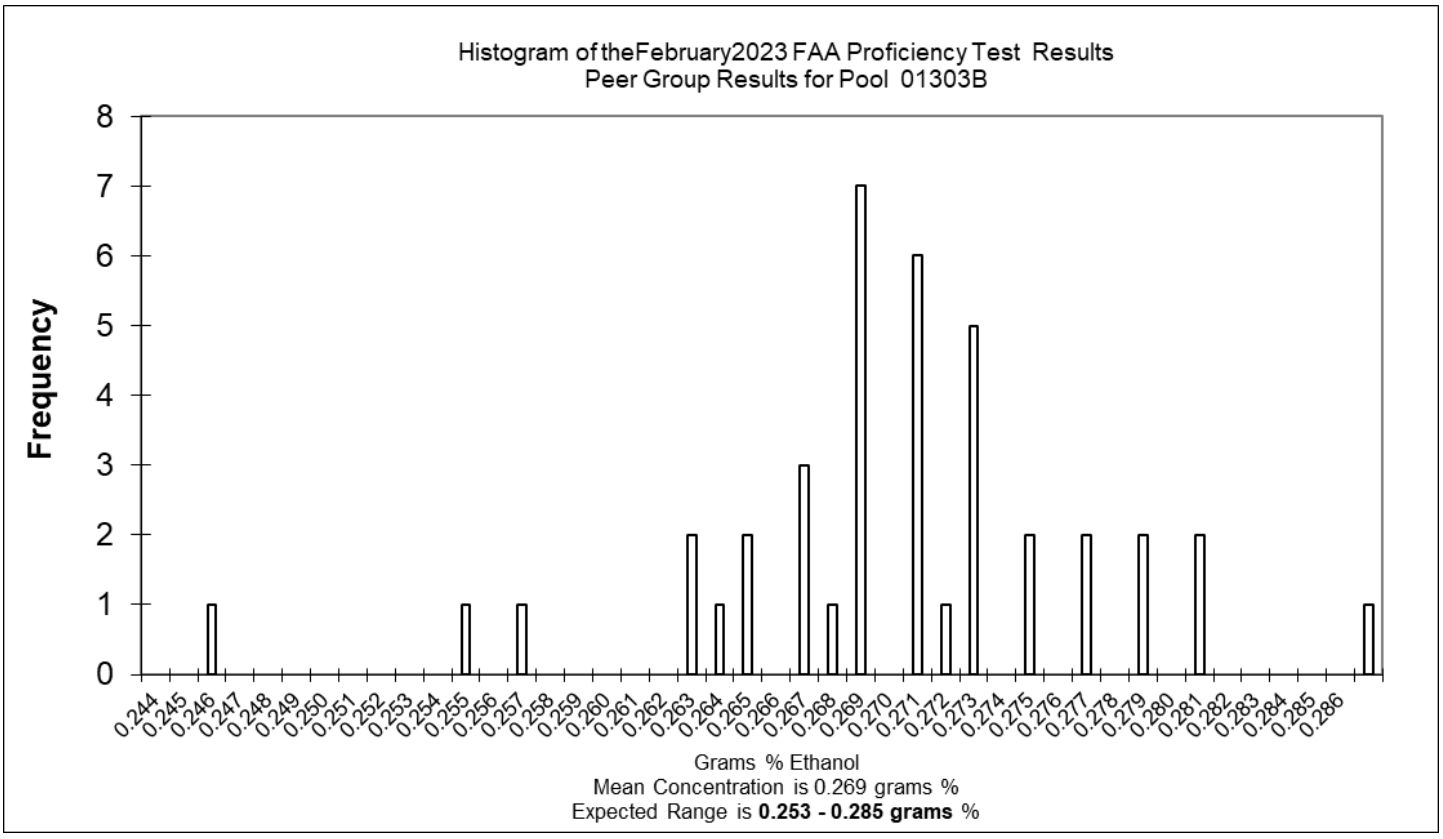


Figure 5 – Box plots (Python): analysis of pools 01233A,B & 01233A,B

- Box plot graphs are generated by Python program;
- Wiskers of these box plots represent the lowest (the highest) actual datapoint for a given pool that is still within IQR
- IQR ranges: pool 01233A (0.065 – 0.073), pool 01233B (0.118 – 0.134), pool 01303A (0.152 – 0.172) and pool 01303B (0.257 – 0.281)

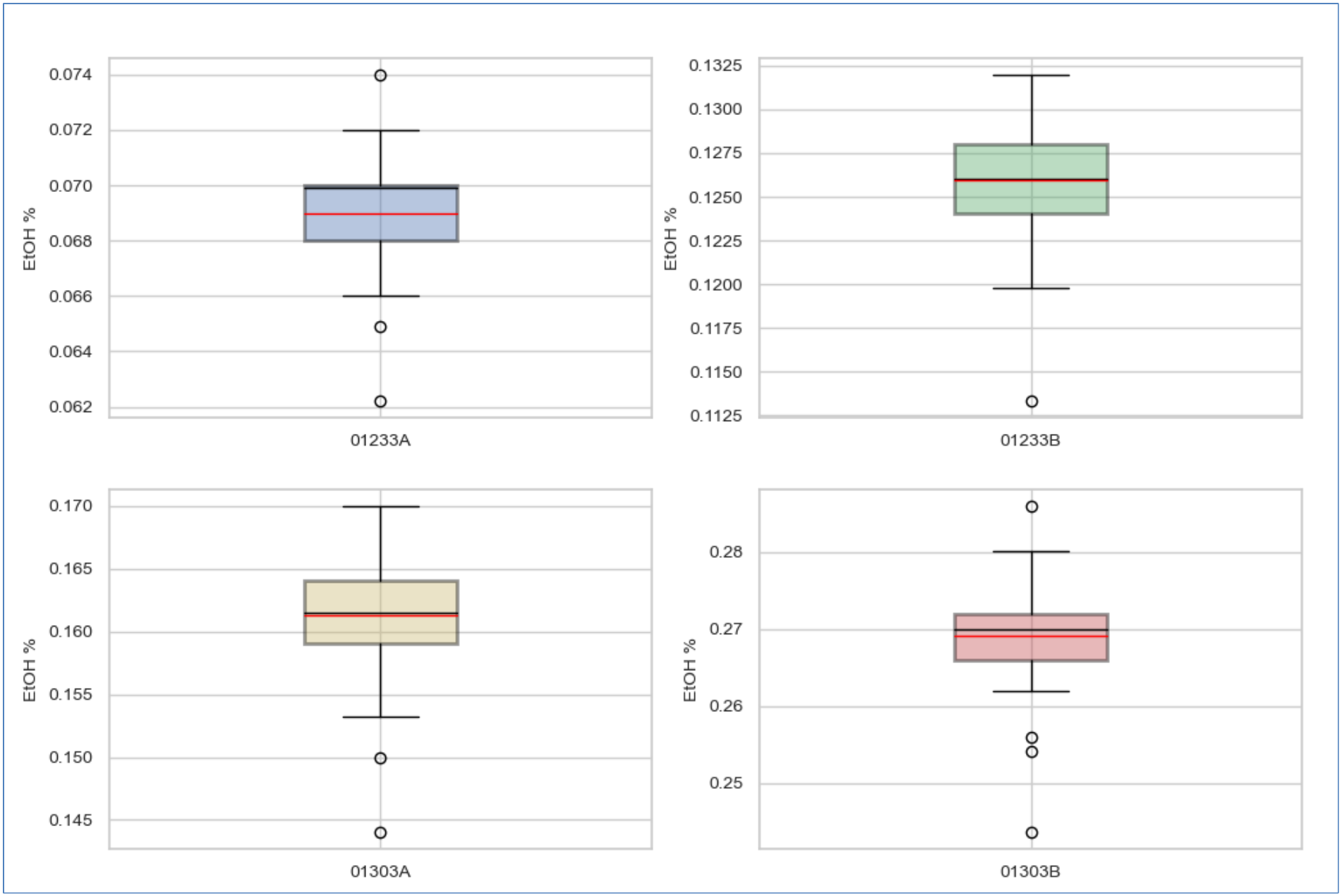


Figure 6 – Kernel density plots (Python) : analysis of pools 01303A,B & 01303A, B

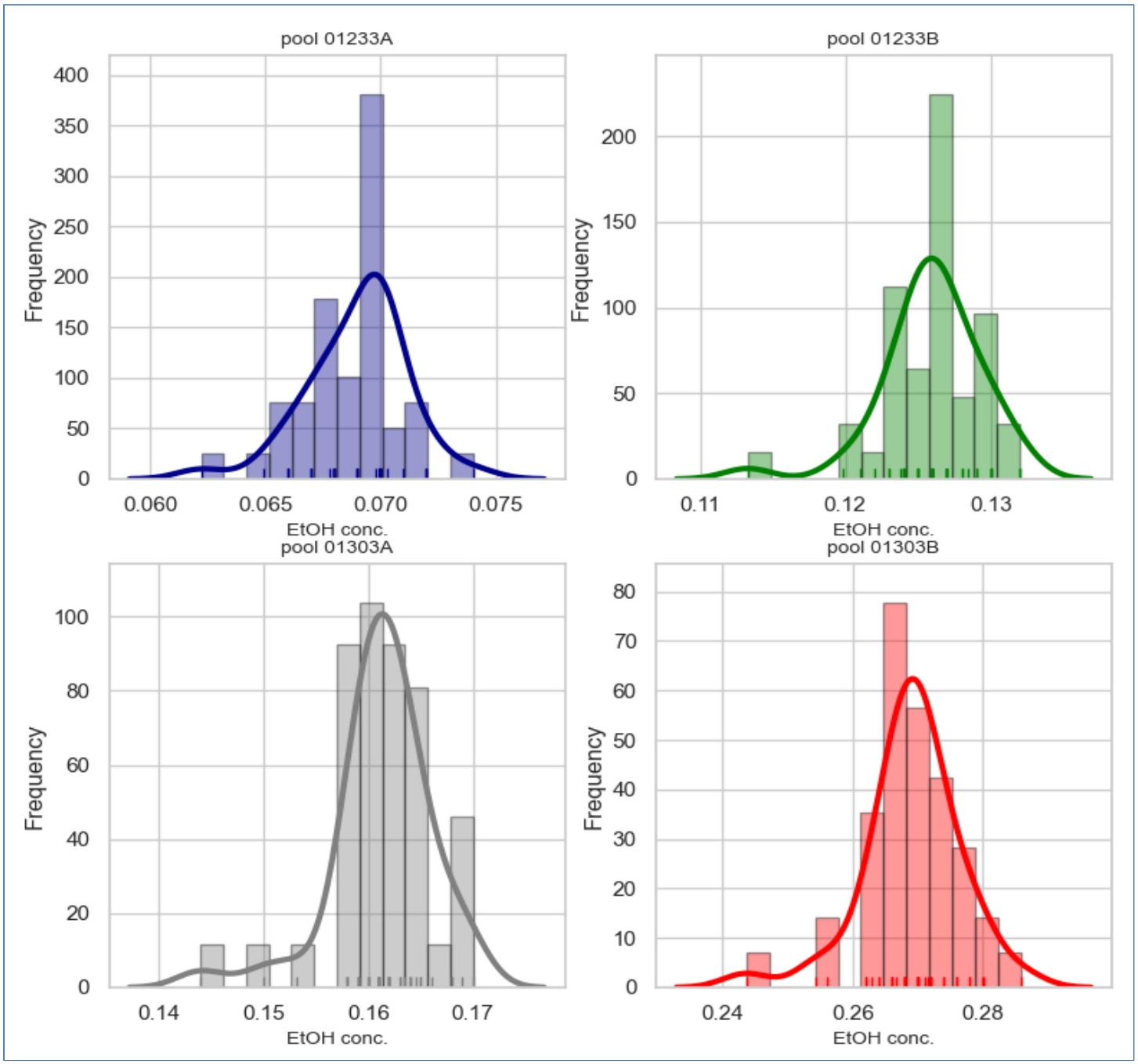
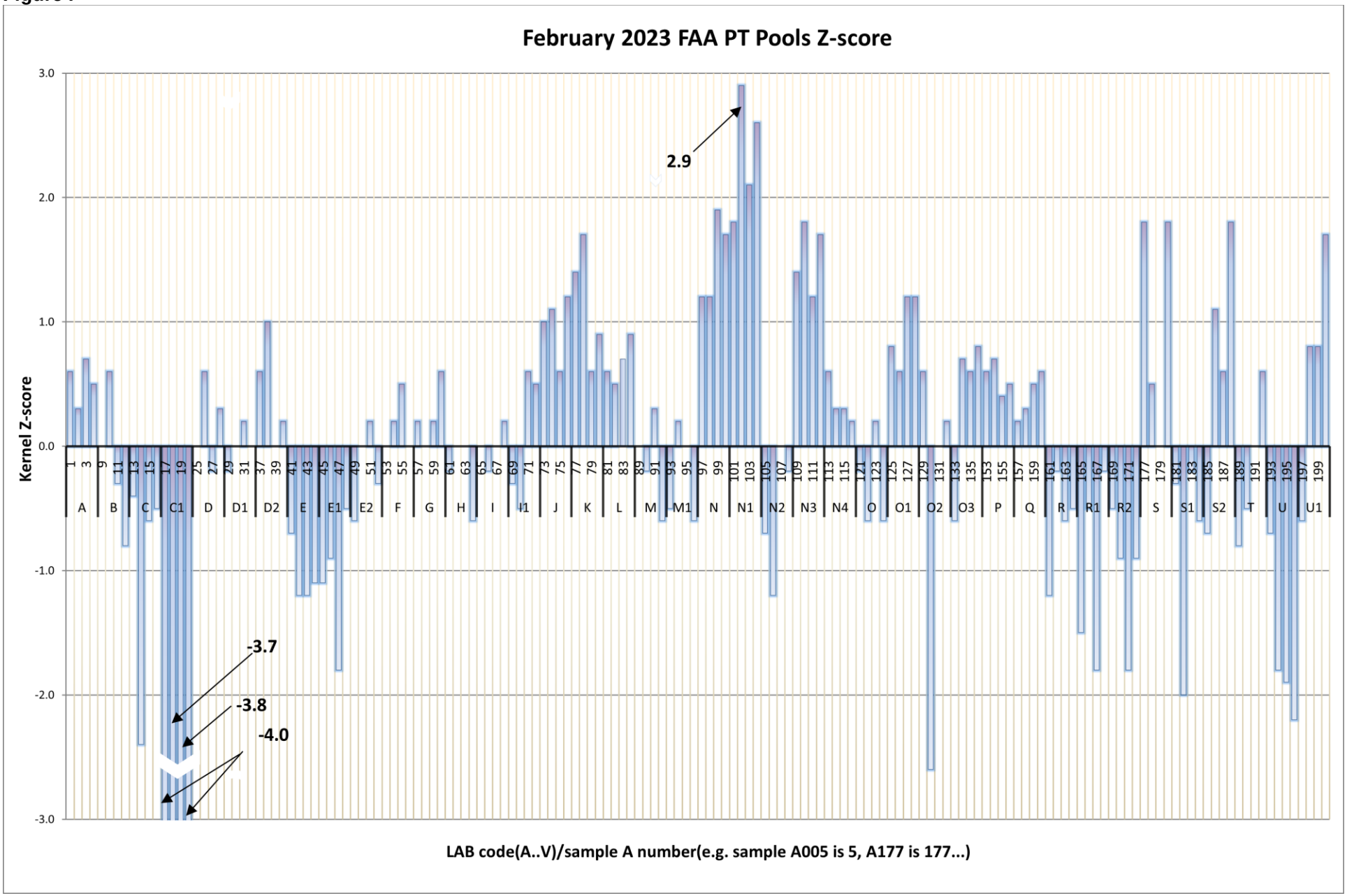


Figure 7


February 2023 FAA PT Pools Z-score
Kernel Z-score
3.0
2.0
1.0
0.0
-1.0
-2.0
-3.0
2.9
-3.7
-3.8
-4.0
A
B
C
C1
D
D1
D2
E
E1
E2
F
G
H
I
I1
J
K
L
M
M1
N
N1
N2
N3
N4
O
O1
O2
O3
P
Q
R
R1
R2
S
S1
S2
T
U
U1
LAB code(A..V)/sample A number(e.g. sample A005 is 5, A177 is 177...)