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 28, 2020

TO: Participants in the April 2020 Voluntary Proficiency Test in Forensic Alcohol Analysis

SUBJECT: Assigned Values and Expected Ranges of Results for the April 2020 Proficiency Test in Forensic Alcohol Analysis

Enclosed is a summary of the descriptive statistics for the Department's April 2020 voluntary proficiency test in forensic alcohol analysis. The Department prepared four blood-alcohol pools (03020A, 03020B, 03090A, and 03090B) 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 24 laboratories elected to voluntarily participate in this proficiency test, with 22 laboratories submitting test results within the time limit.

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.

The majority of participants [16 out of 22] 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.

Z-score calculation formula:

$$z = \frac{\mathbf{X} - \mathbf{X}_a}{\sigma_p}$$

where: $\mathbf{X}$ is a participant result, $\mathbf{X}_a$ is a consensus value and σ_p is the fitness-for-purpose standard deviation, see also Table 2,

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 April 2020 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

[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. The Department used the Standard method.

(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 03020A and 03020B 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 03090A and 03090B. These figures can be used by the participants to evaluate their data.

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 April 2020 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	03020A	0.070	0.064 – 0.077	0.06 – 0.07
#1B	03020B	0.152	0.142 – 0.162	0.13 – 0.16
#2A	03090A	0.118	0.110 – 0.126	0.10 – 0.12
#2B	03090B	0.213	0.200 – 0.226	0.19 – 0.22

Table 2 Summary of Test Pool Data

Parameter		Pool1A (03020A)	Pool 1B (03020B)	Pool 2A (03090A)	Pool 2B (03090B)
Pre-distribution Data	Target Value	0.07	0.15	0.12	0.22
	True Value[5]	0.071	0.150	0.117	0.210
	Standard Deviation	0.0004	0.0008	0.0010	0.0010
Descriptive statistics	Mean	0.070	0.152	0.117	0.213
	Adjusted Mean[6]	0.070	0.152	0.118	0.213
	Standard Error[7]	0.0002	0.0003	0.0004	0.0007
	Median	0.070	0.152	0.118	0.212
	Standard Deviation	0.0014	0.0020	0.0021	0.0043
	Minimum	0.067	0.147	0.110	0.200
	Maximum	0.074	0.157	0.120	0.228
	Count	35[8]	35[8]	35[8]	35[8]
Non-parametric statistics (SigmaPlot)	Q1 (25%)	0.070	0.152	0.116	0.211
	Q3 (75%)	0.071	0.153	0.118	0.214
	IQR	0.001	0.001	0.002	0.003
	Lower Fence	0.0685	0.1505	0.1130	0.2095
	Upper Fence	0.0725	0.1545	0.1210	0.2155
Histogram		Figure 1	Figure 2	Figure 3	Figure 4
Normal distribution?[9]		No (p=0.001)	No (p=0.022)	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]		0.0700	0.1521	0.1178	0.2124
Robust standard deviation, σ_{rob}		0.0012	0.0016	0.0016	0.0028
Fitness-for-purpose standard deviation, σ_p[11]		0.0021	0.0040	0.0033	0.0054
Consensus value (X_a) - determined as Mode ($\mu_{1/2}$) of Gaussian Kernel distribution		0.0702	0.1522	0.1178	0.2122
Uncertainty of the consensus value, X_a , S.E.[12]		0.00037	0.00064	0.00550	0.00310
$X_a \pm$ S.E.		0.0702 ± 0.0004	0.1522 ± 0.0006	0.1178 ± 0.0055	0.2122 ± 0.0031
z-score		$z = \frac{X - X_a}{\sigma_p}$	$z = \frac{X - X_a}{\sigma_p}$	$z = \frac{X - X_a}{\sigma_p}$	$z = \frac{X - X_a}{\sigma_p}$

[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 22 laboratories participated and analyzed a total of 35 sample sets.

[9] Shapiro-Wilk test used at 0.05 significance level (i.e., P-value greater than 0.05).

[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 **: $\sigma_p' = 0.02*\mu_{1/2}^{0.8495}$**

[12] Derived from bootstrep procedure with bandwidth (h) equal 0.75*σ_p

Figure 1

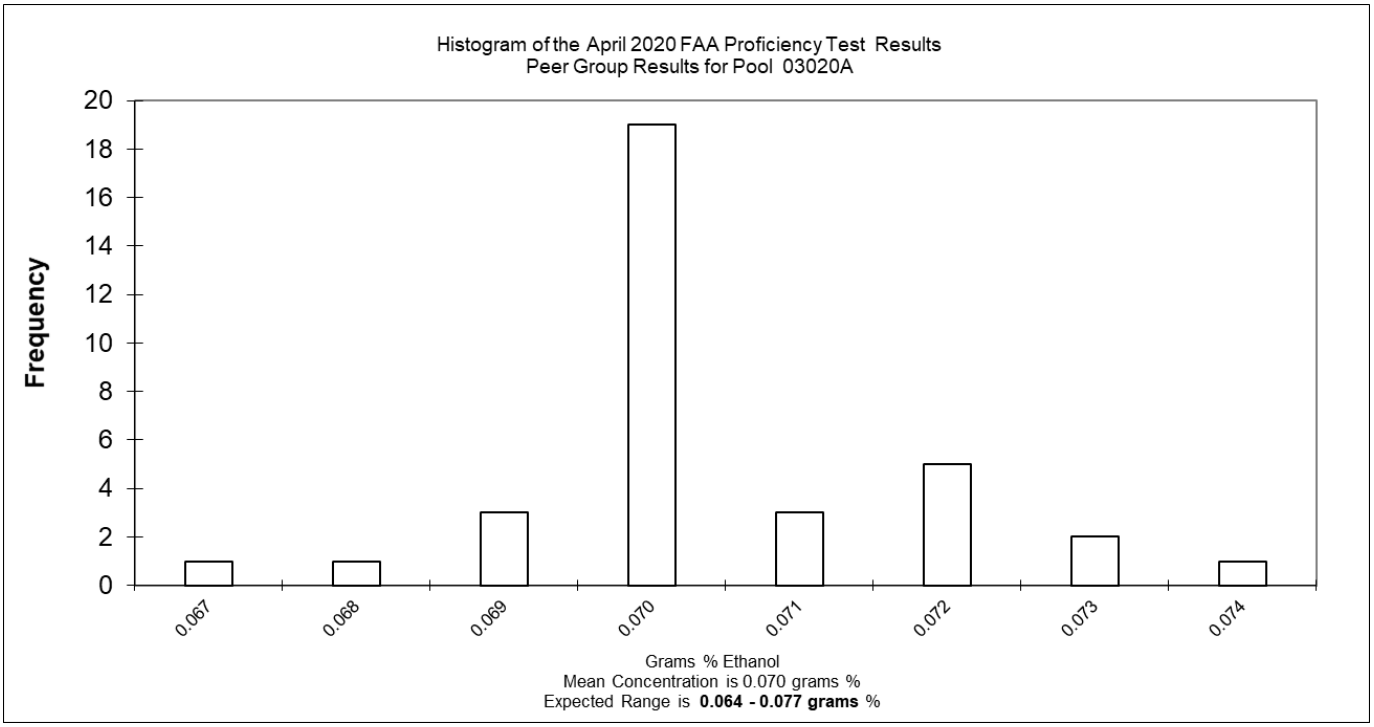


Figure 2

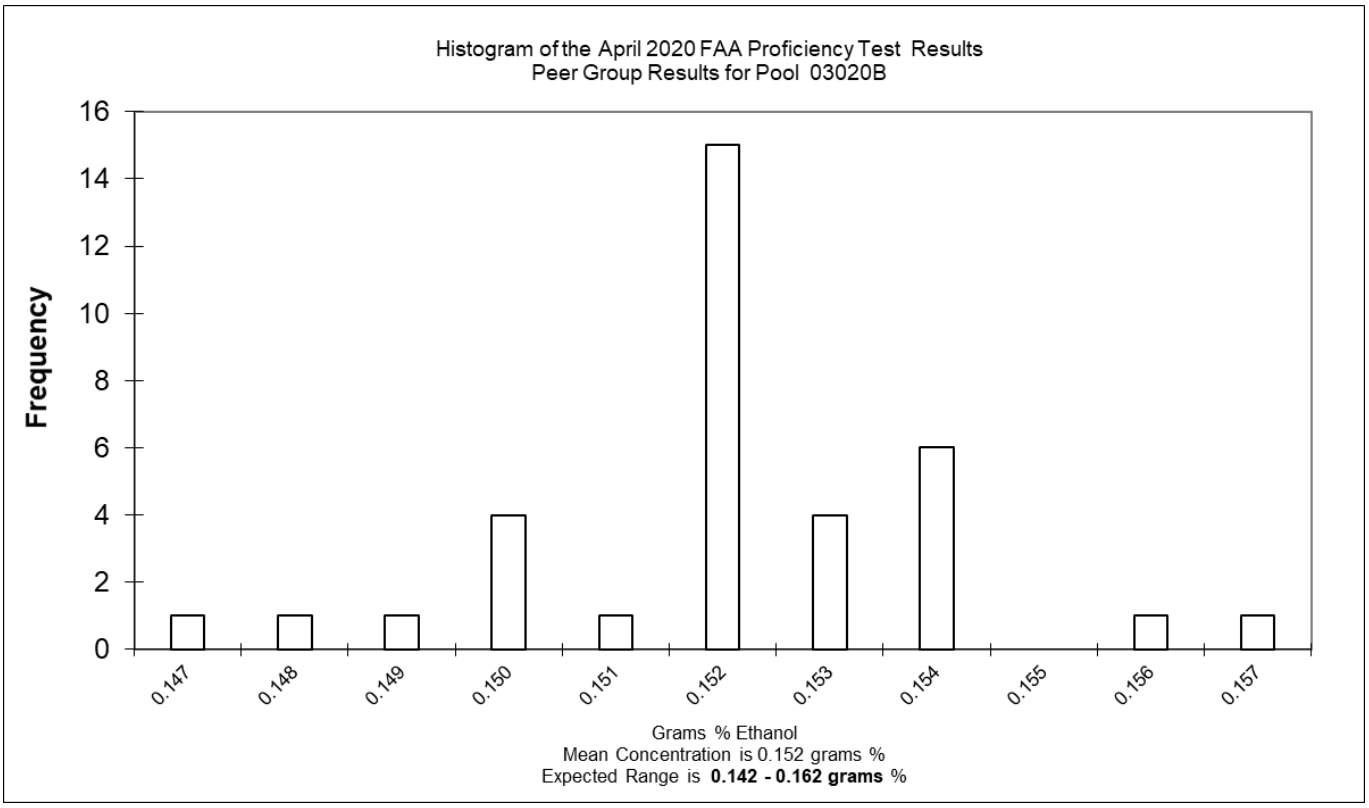


Figure 3

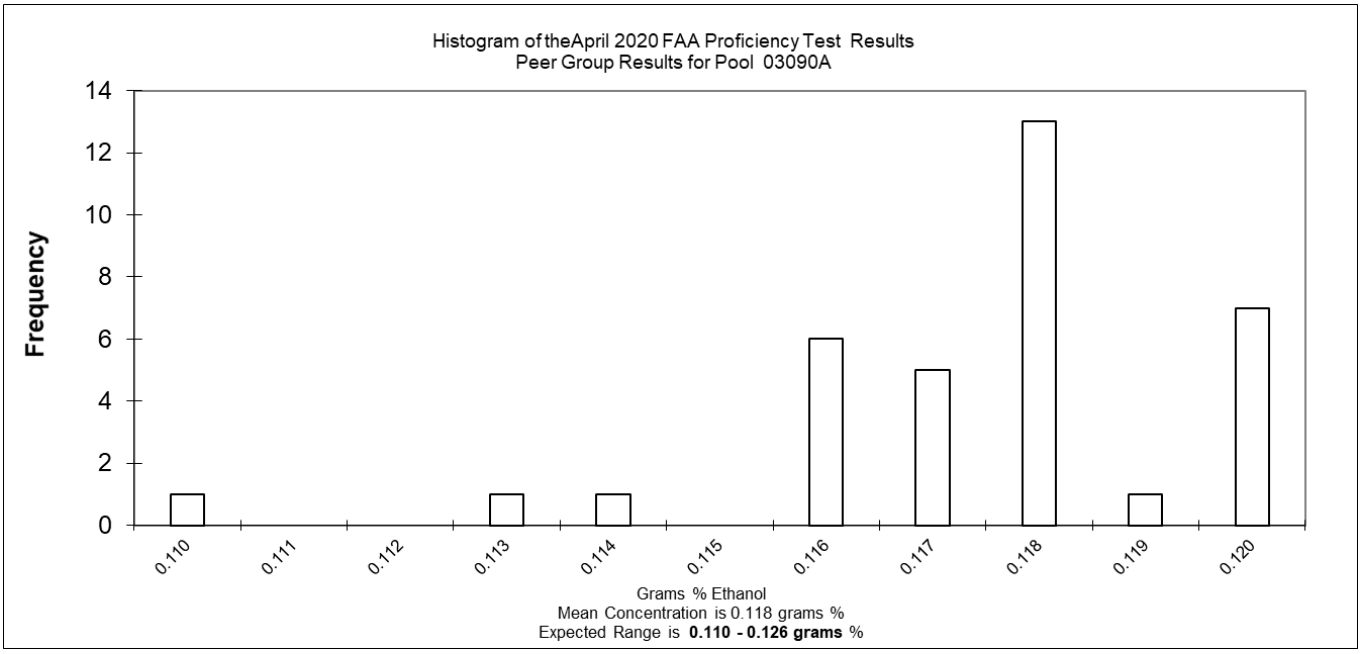

Histogram of theApril 2020 FAA Proficiency Test Results
Peer Group Results for Pool 03090A
Frequency
14
12
10
8
6
4
2
0
0.110
0.111
0.112
0.113
0.114
0.115
0.116
0.117
0.118
0.119
0.120
Grams % Ethanol
Mean Concentration is 0.118 grams %
Expected Range is **0.110 - 0.126 grams** %

Figure 4

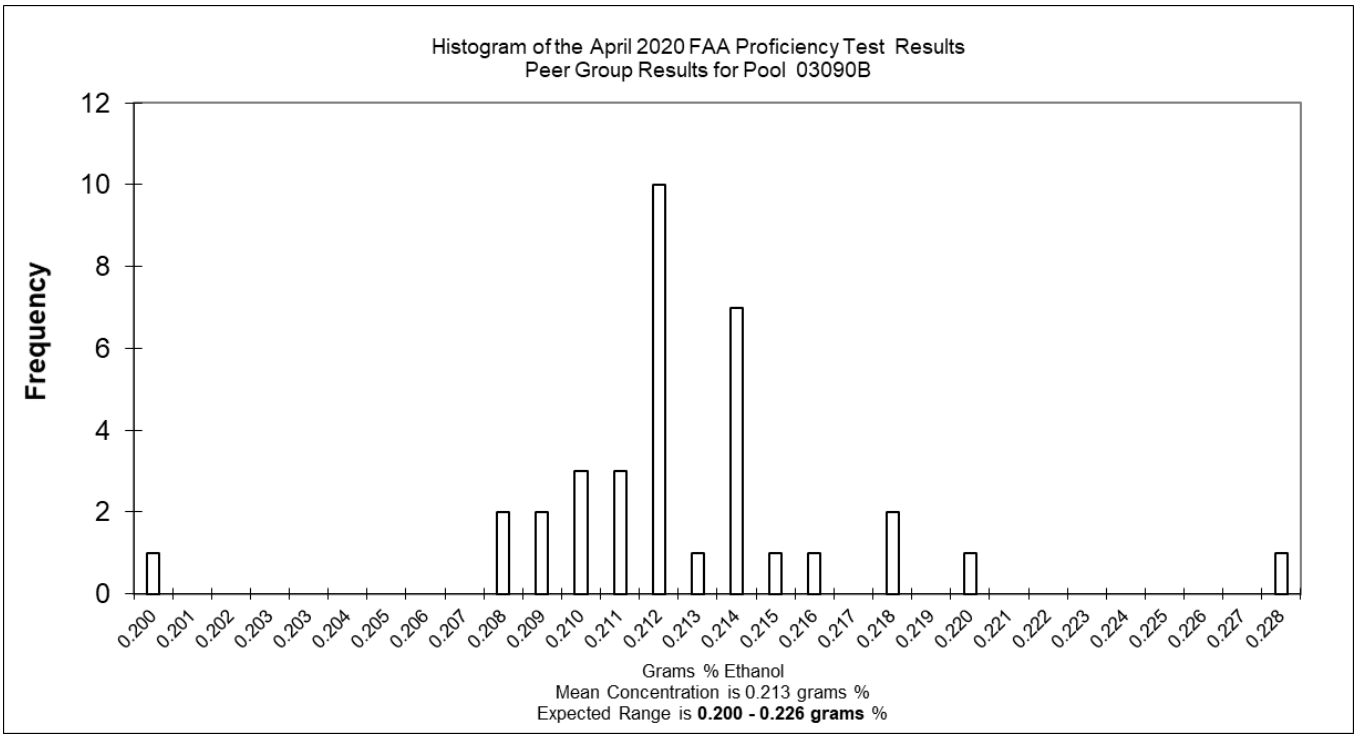
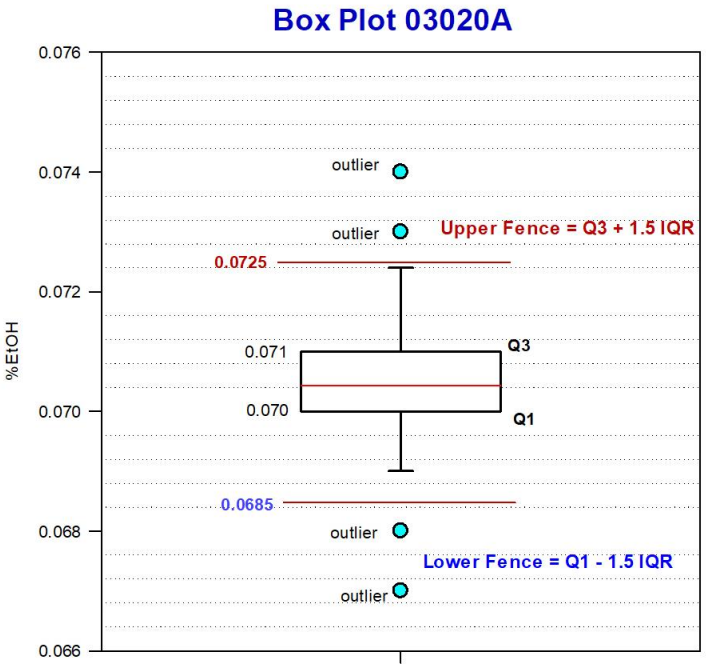
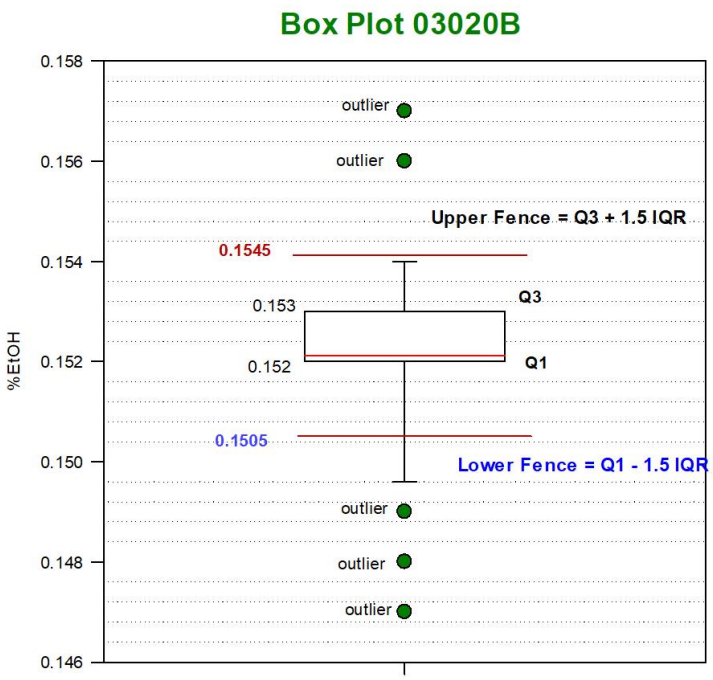
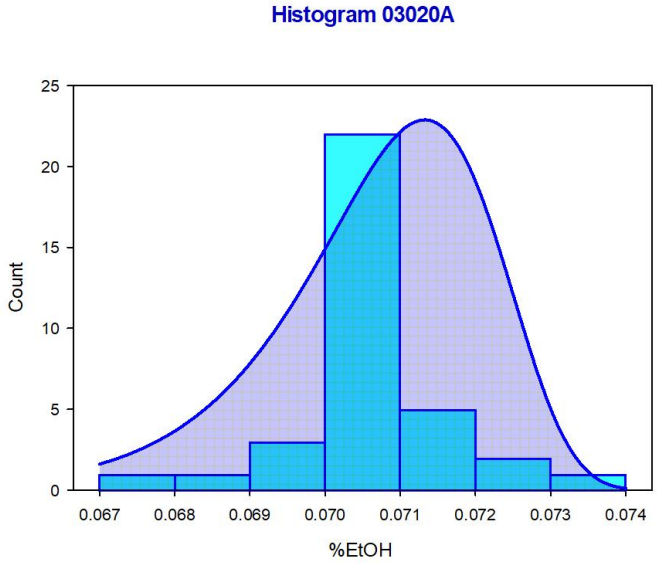

Histogram of the April 2020 FAA Proficiency Test Results
Peer Group Results for Pool 03090B
Frequency
12
10
8
6
4
2
0
0.200
0.201
0.202
0.203
0.203
0.204
0.205
0.206
0.207
0.208
0.209
0.210
0.211
0.212
0.213
0.214
0.215
0.216
0.217
0.218
0.219
0.220
0.221
0.222
0.223
0.224
0.225
0.226
0.227
0.228
Grams % Ethanol
Mean Concentration is 0.213 grams %
Expected Range is **0.200 - 0.226 grams** %

Figure 5 – SigmaPlot Analysis of Pools 03020A & 03020B (Pools 1A & 1B)







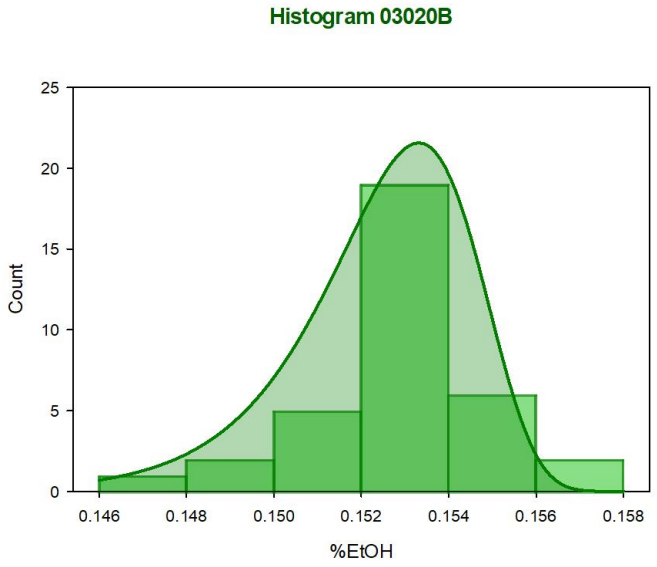
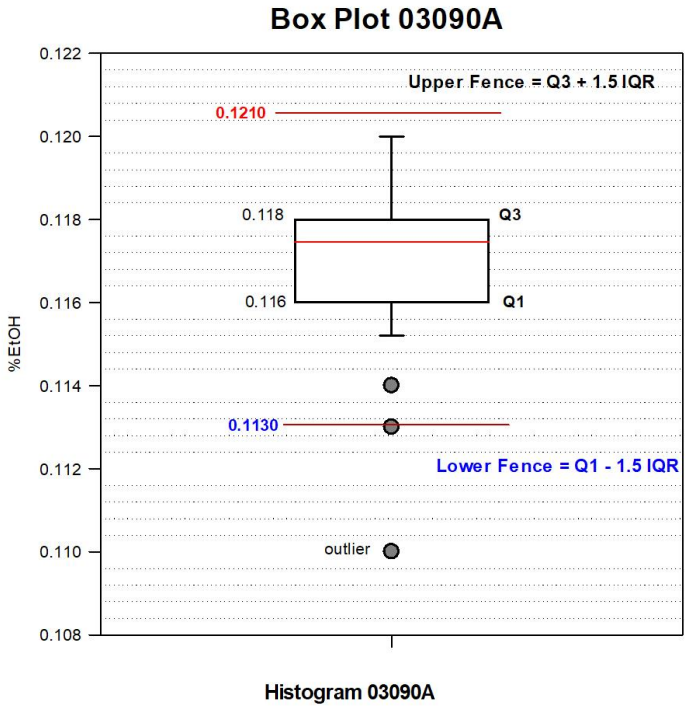
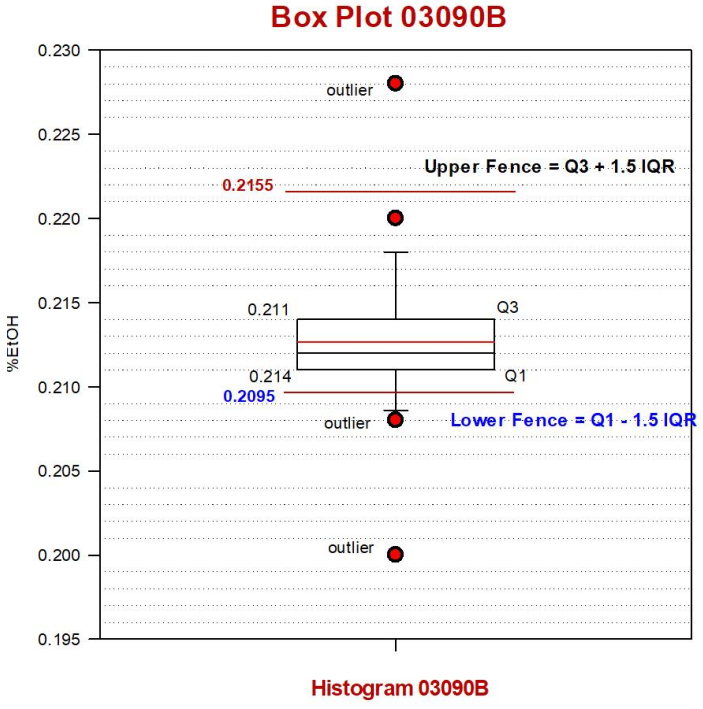
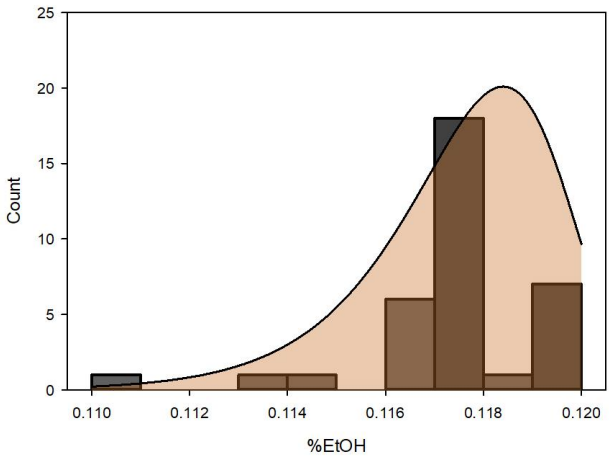


Figure 6 – SigmaPlot analysis of Pools 03090A & 03090B (Pools 2A & 2B)







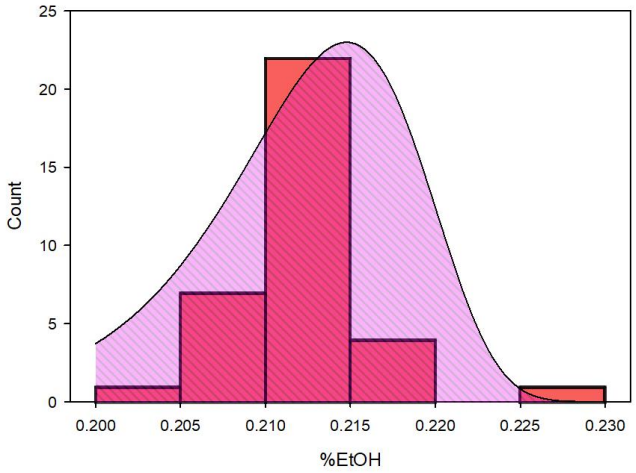
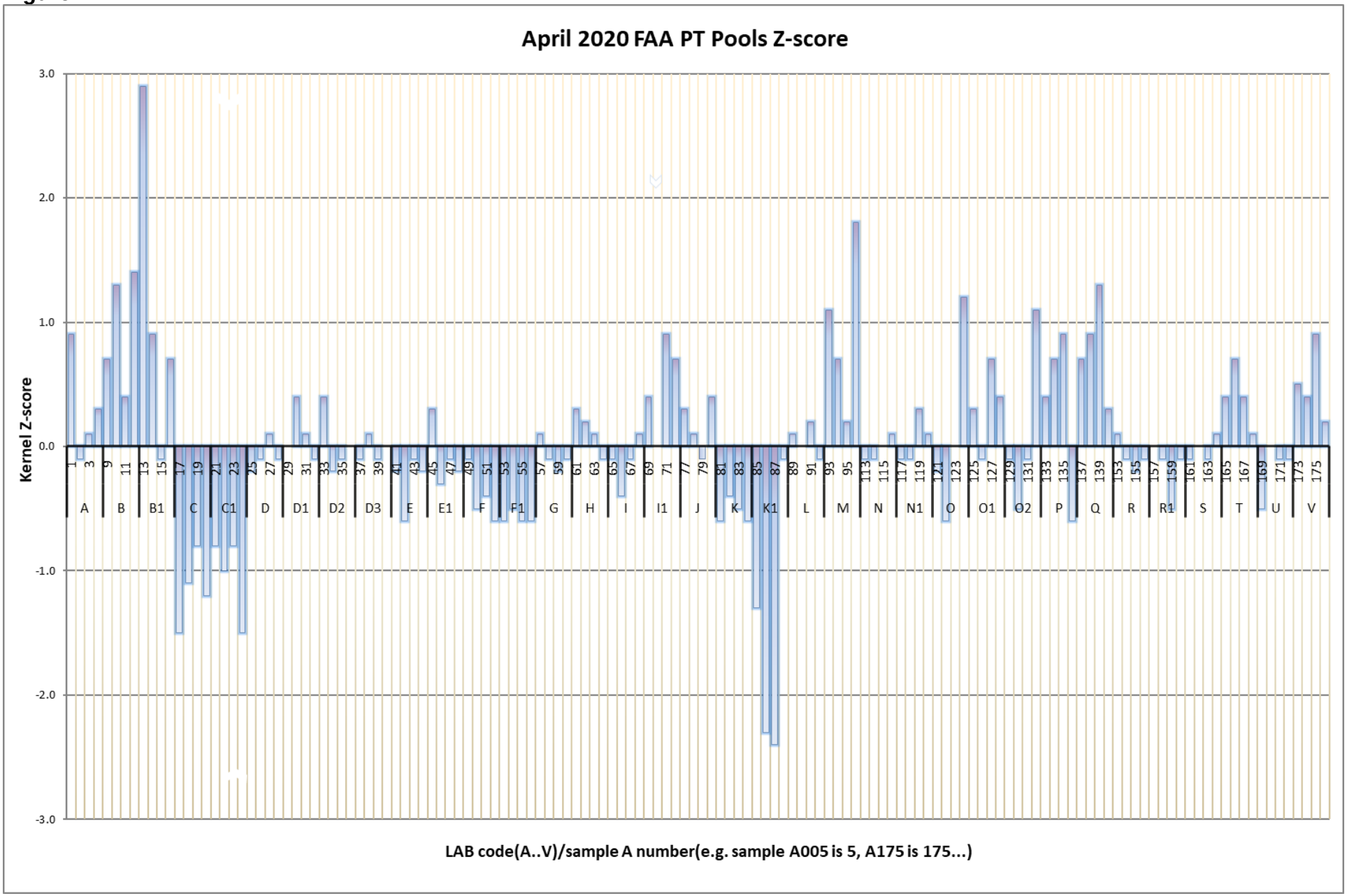


Figure 7


April 2020 FAA PT Pools Z-score
Kernel Z-score
3.0
2.0
1.0
0.0
-1.0
-2.0
-3.0
A
B
B1
C
C1
D
D1
D2
D3
E
E1
F
F1
G
H
I
I1
J
K
K1
L
M
N
N1
O
O1
O2
P
Q
R
R1
S
T
U
V
LAB code(A..V)/sample A number(e.g. sample A005 is 5, A175 is 175...)