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

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

EDMUND G. BROWN JR.
Governor

July 23, 2018

TO: Participants in the May 2018 Voluntary Proficiency Test in Forensic Alcohol Analysis

SUBJECT: Assigned Values and Expected Ranges of Results for the May 2018 Proficiency Test in Forensic Alcohol Analysis

Enclosed is a summary of the descriptive statistics for the May 2018 proficiency test in forensic alcohol analysis. The Department prepared four test blood-alcohol pools (04248A, 04248B, 04308A, and 04308B) 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. As described in the Addendum, "Alcohol Concentration Stability of Test Samples Used in the May 2018 Proficiency Test," there were losses in alcohol concentration in some samples from Pools 04308A and 04308B.

With the recent 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 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

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

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 [19 out of 27] 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 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 were 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 May 2018 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 pools 04248A and 04248B are presented as box and whisker or Tukey plots with the quartiles and fences shown. 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. Figure 6, presents the same data for pools 04308A and 04308B. These figures can be used by the participants to evaluate their data.

[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)

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

Addendum: Alcohol Concentration Stability of Test Samples Used in the May 2018 Proficiency Test

The test samples used in the Department's proficiency tests in forensic alcohol analysis were prepared from human whole blood components (blood cells and fresh frozen plasma) combined to produce a normal hematocrit. For the May 2018 proficiency test, the Department prepared four blood-alcohol test pools (04248A, 04248B, 04308A, and 04308B[4]). The blood samples contain a preservative sodium fluoride (2 mg/mL).

The Department's pool qualification procedures include the analyses of selected samples to determine the homogeneity and true value of the samples. Selected samples are also tested for any microbiological contaminants that may have been present in the blood components or introduced during the preparation of the pools and some samples are subjected to an aging study (10 days storage at room temperature) to evaluate the stability of the alcohol concentrations in the pools. After the samples are shipped to the participants, the Department monitors the stability of the concentrations of the blood alcohol test pools throughout the participants' testing period. The goal here is to ensure that all of the participants are analyzing the same samples.

FDLB has conducted proficiency tests in forensic alcohol analyses for more than 30 years and has almost never encountered any problems with the stability of alcohol concentrations in the test pools during the testing period and for an extended period thereafter. However, with the test samples included in the May 2018 proficiency test, the monitoring analyses suggested that there may be problems with the stability of two of the four test pools. The monitoring phase analyses included tests of 156 samples, which were analyzed by two different methods. The tests revealed that a total of four samples from two pools appeared to lose alcohol during the end of the monitoring period yielding results that were outliers as determined by statistical tests. The absolute variations here were relatively small: - 0.006 % for Pool 04308A and - 0.009 % for Pool 04308B, but statistically significant. These two pools were shown to be contaminated with a bacteria, *Acinetobacter lwoffii,* however, subsequent testing of the contaminated samples did not show that there was a loss of alcohol. There did appear to be some physical changes in the affected blood samples. These samples appeared darker in color and were more hemolyzed than non-affected samples. The correlation was not perfect, however, as some darker, hemolyzed samples did not suffer from loses in alcohol concentration. The Department is continuing to evaluate the situation and is conducting more microbiological tests and additional monitoring analyses.

Since the alcohol concentration stability problems appeared to affect only a small number of samples during the testing period, the Department concluded that there is still value in publishing the results obtained for the two affected test pools. A laboratory should consider the alcohol concentration stability when evaluating its results. For example, in evaluating Z-scores for samples from Pools 04308A and 04308B, it is possible that a negative Z-score could be partially or completely explained by a loss of alcohol in the sample. The results shown in the summary are footnoted to explain the alcohol stability issues and the samples affected.

[4] Blood-alcohol Pools 04248A, 04248B, 04308A, and 04308B are designated as Pools 1A, 1B, 2A, and 2B, respectively, in the report and of the laboratories' annotated report forms.

Statistical Data for May 2018 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	04248A	0.048	0.042 – 0.055	0.04 – 0.05
#1B	04248B	0.107	0.100 – 0.114	0.09 – 0.11
#2A	04308A	0.139	0.130 – 0.148	0.12 – 0.14
#2B	04308B	0.237	0.223 – 0.251	0.22 – 0.25

Table 2 Summary of Test Pool Data

Parameter		Pool1A (04248A)	Pool 1B (04248B)	Pool 2A (04308A)	Pool 2A (04308B)
Pre-distribution Data	Target Value	0.050	0.110	0.140	0.240
	True Value[5]	0.048	0.106	0.139	0.235
	Standard Deviation	0.0003	0.0002	0.0007	0.0013
Descriptive statistics	Mean	0.0480	0.107	0.139	0.237
	Adjusted Mean[6]	0.0480	0.107	0.139	0.237
	Standard Error[7]	0.0001	0.0003	0.0006	0.0006
	Median	0.0484	0.107	0.140	0.237
	Standard Deviation	0.0008	0.0019	0.0043	0.0039
	Minimum	0.046	0.100	0.120	0.227
	Maximum	0.050	0.111	0.148	0.246
	Count	46[8]	46[8]	46[8]	46[8]
Descriptive statistics (box plot)	Q1 (25%)	0.048	0.106	0.138	0.236
	Q3 (75%)	0.049	0.108	0.141	0.239
	IQR	0.001	0.002	0.003	0.003
	Lower Fence	0.0465	0.103	0.133	0.232
	Upper Fence	0.0505	0.111	0.146	0.244
Histogram		Figure 1	Figure 2	Figure 3	Figure 4
Normal distribution?[9]		No (p<0.001)	No (p<0.001)	No (p<0.001)	No (p<0.016)
Box Plot (SigmaPlot™)		Figure 5	Figure 5	Figure 6	Figure 6
Robust mean, X*[10]		0.0481	0.1068	0.1397	0.2371
Robust standard deviation, σ_{rob}		0.0001	0.0016	0.0023	0.0019
Fitness-for-purpose standard deviation, σ_p[11]		0.0015	0.0030	0.0038	0.0059
Consensus value (X_a) determined as Mode ($\mu_{1/2}$) of Gaussian Kernel distribution		0.0482	0.1068	0.1399	0.2373
Uncertainty of the consensus value, X_a , S.E.[12]		0.0001	0.00025	0.00032	0.00051
X_a ± S.E.		0.0482 ± 0.0001	0.1068 ± 0.0002	0.1399 ± 0.0003	0.2373 ± 0.0005
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)
[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 **: $\sigma_p' = 0.02*\mu_{1/2}^{0.8495}$**
[12] Determined as the Standard Error of Mode using bootstrap simulation technique with bandwidth of $0.75*\sigma_p$

Figure 1

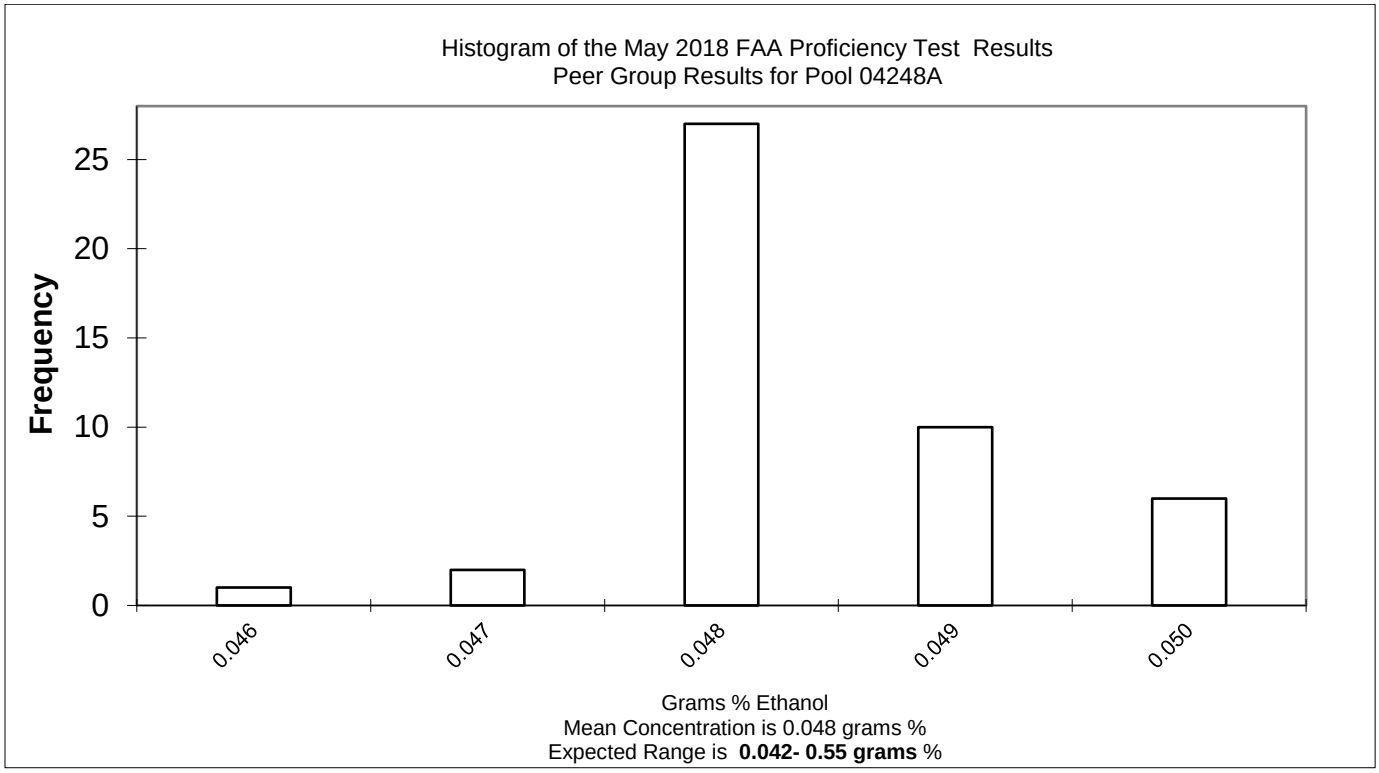


Figure 2

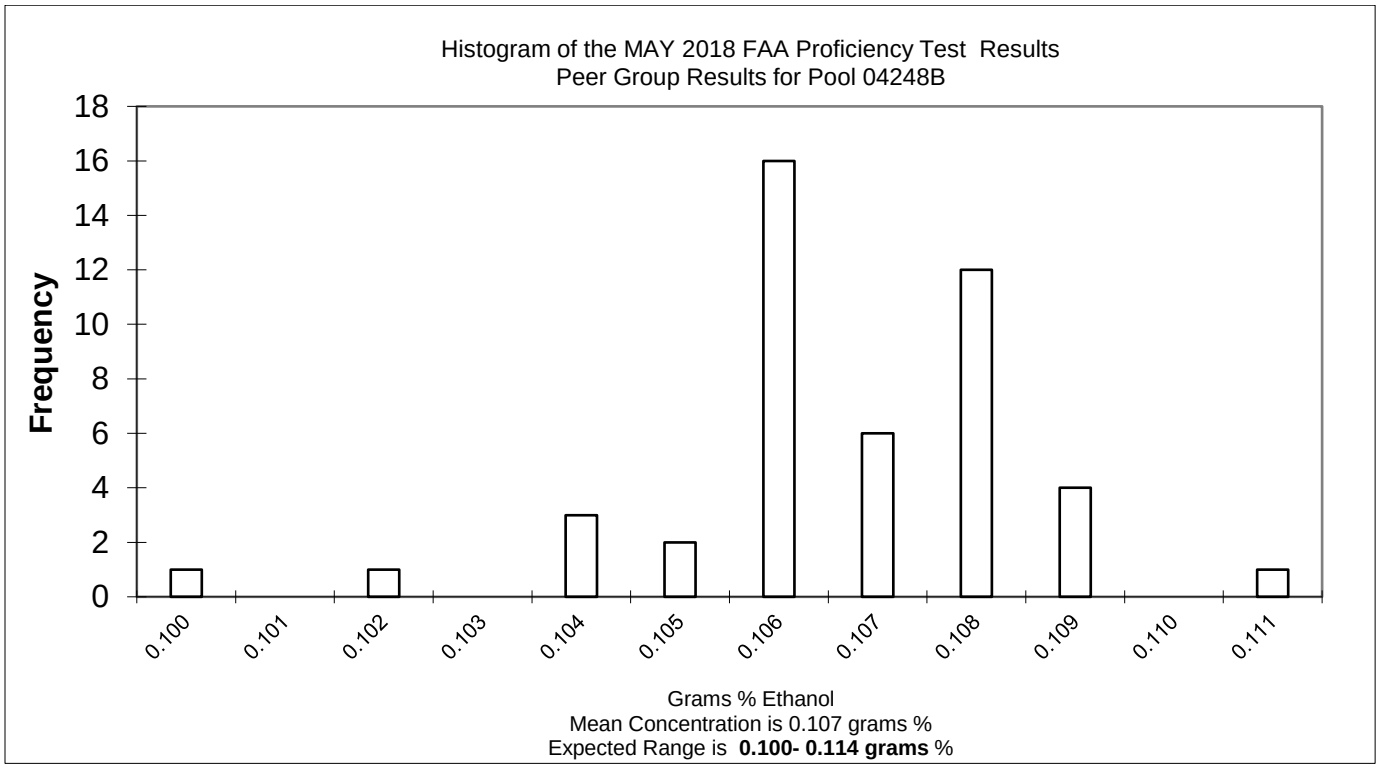


Figure 3

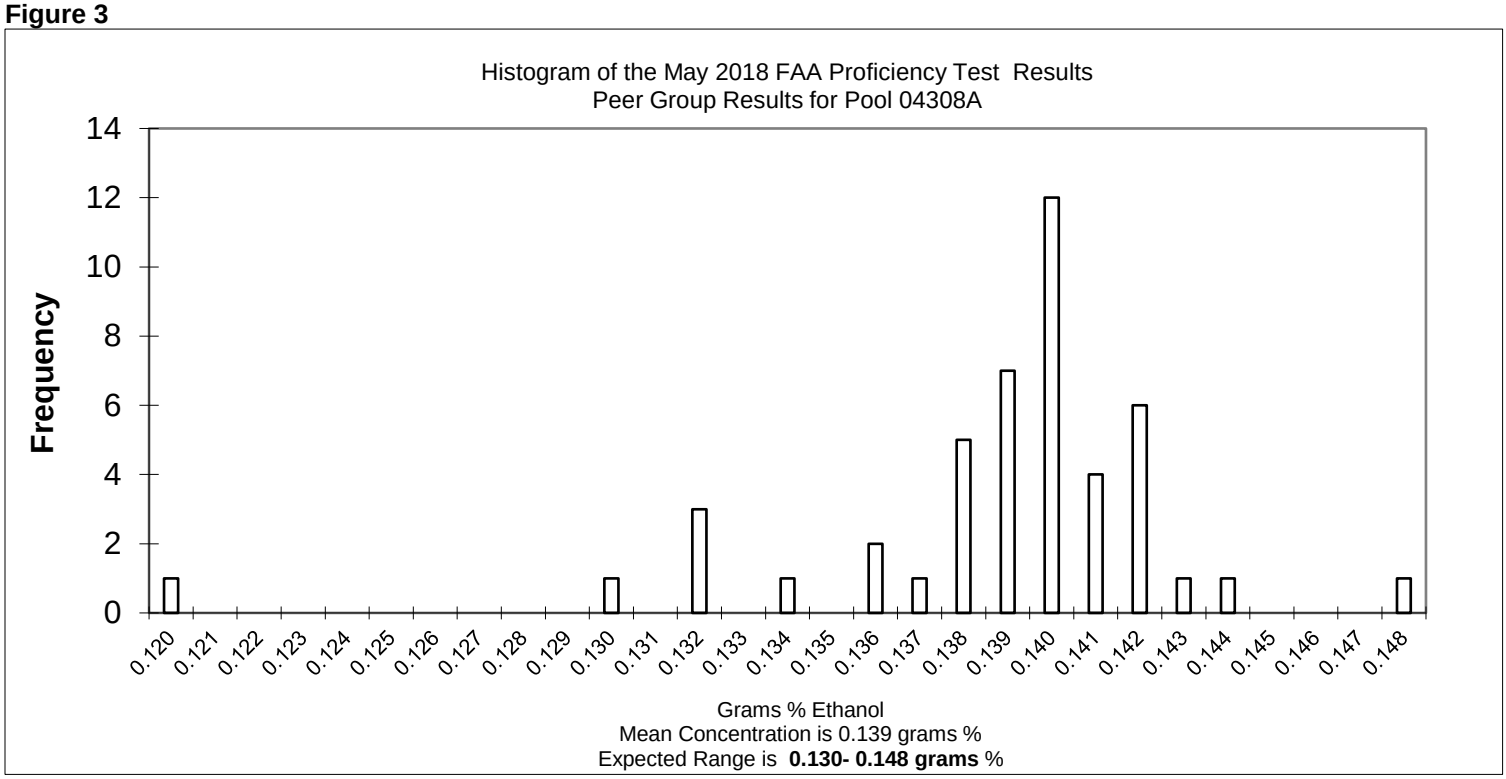


Figure 4

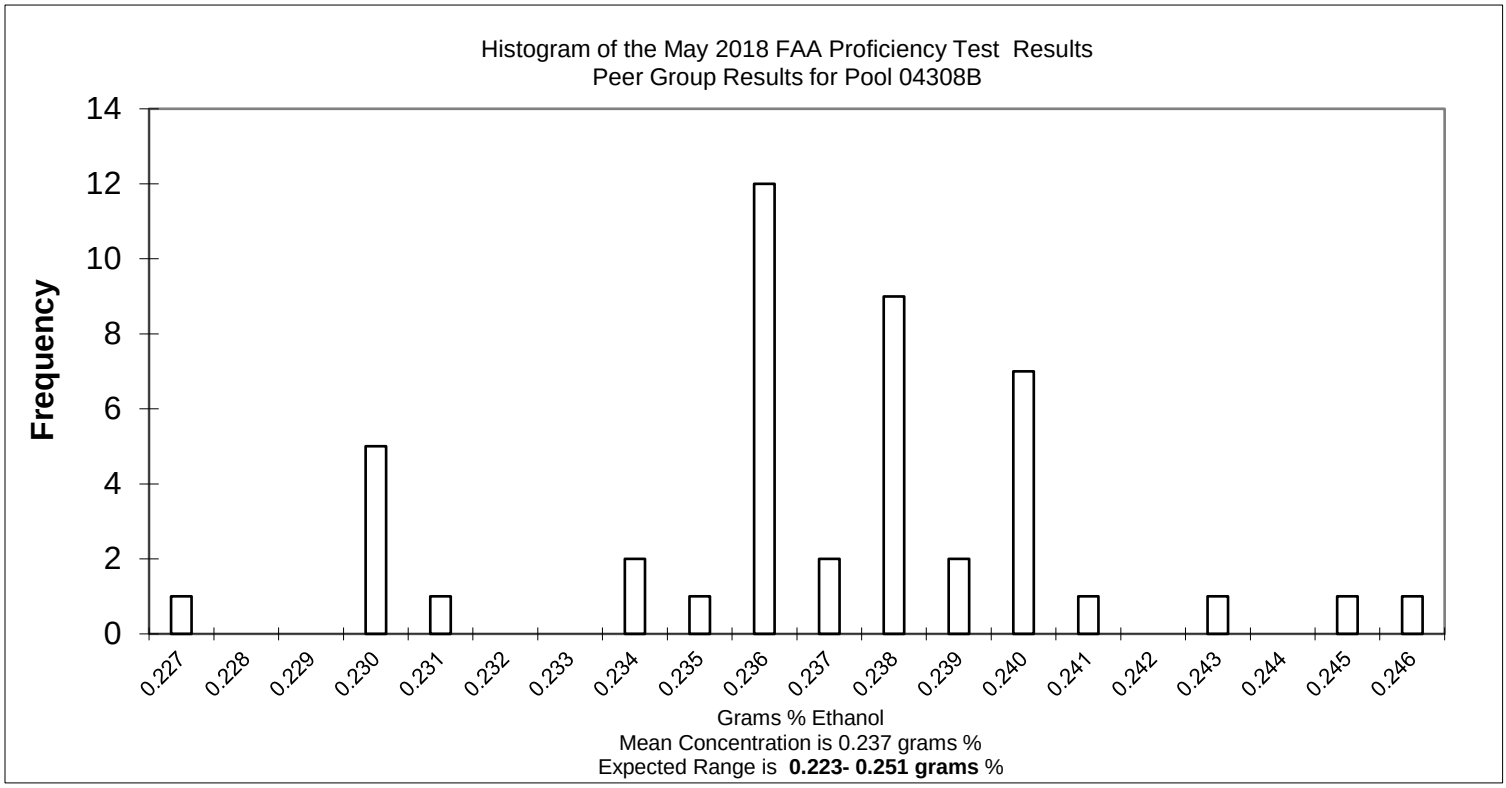
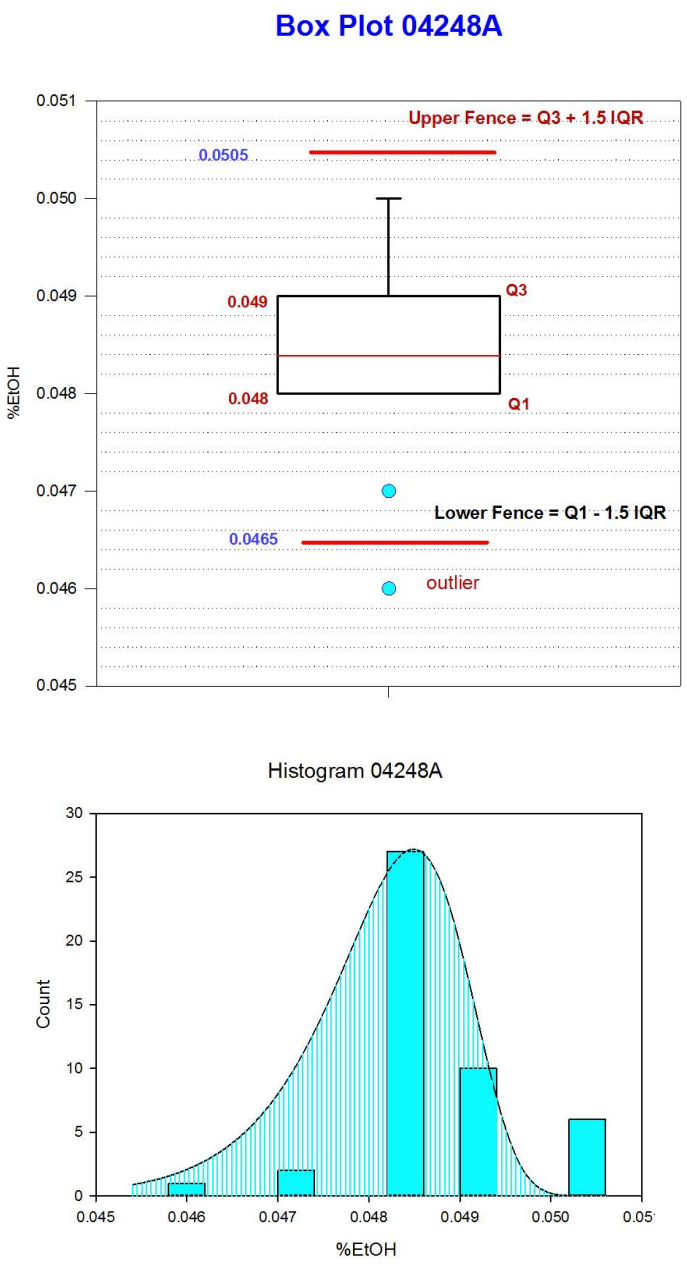


Figure 5 – SigmaPlot™ analysis of pools 04248A & 04248B



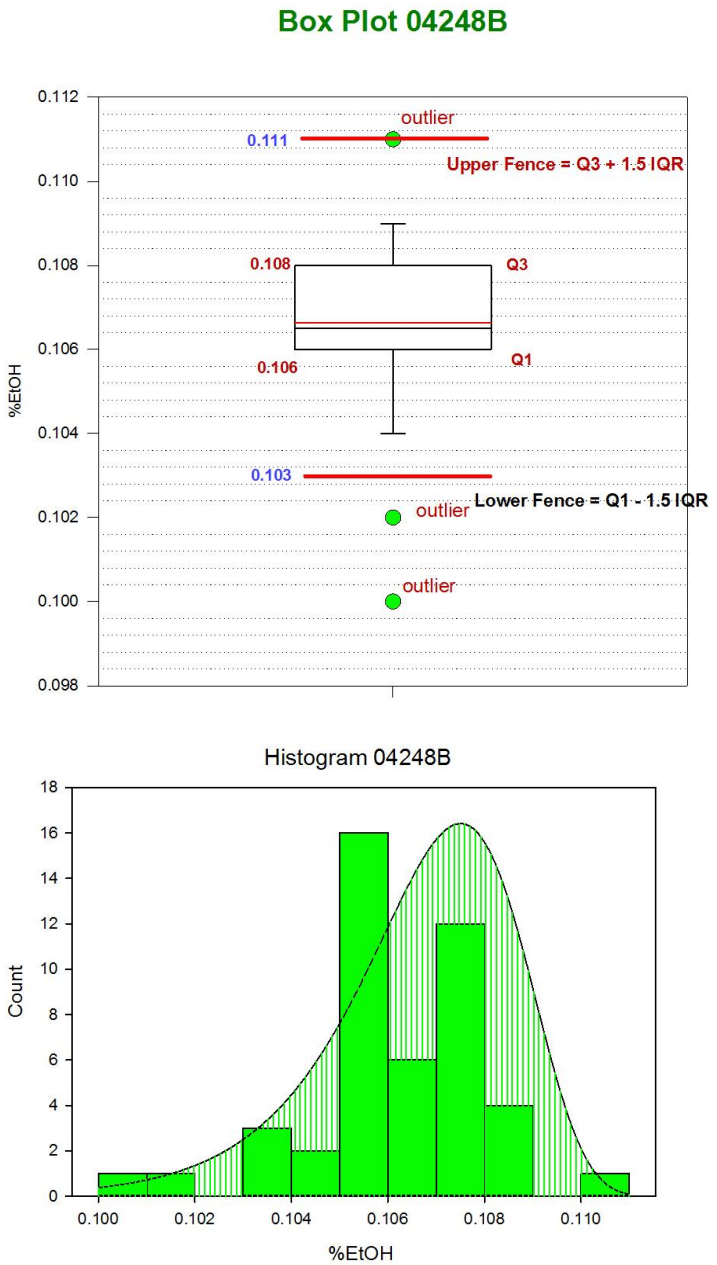
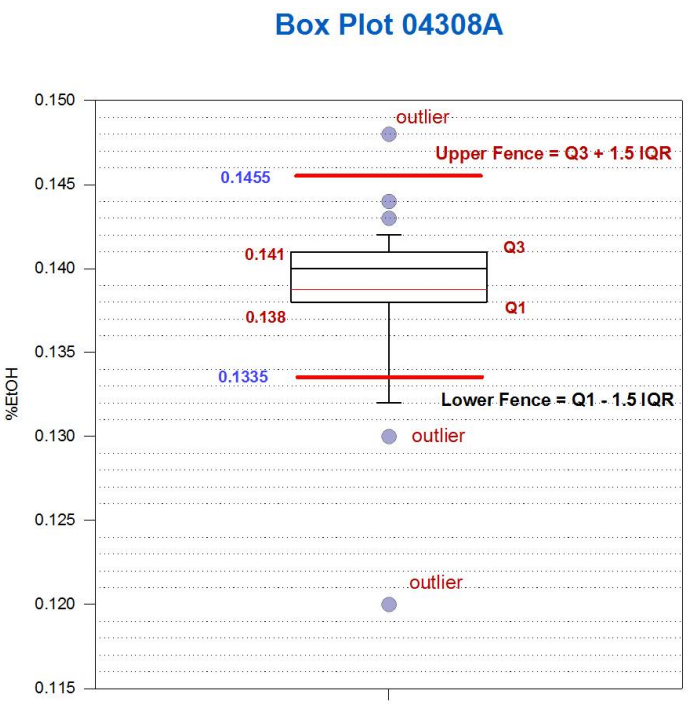
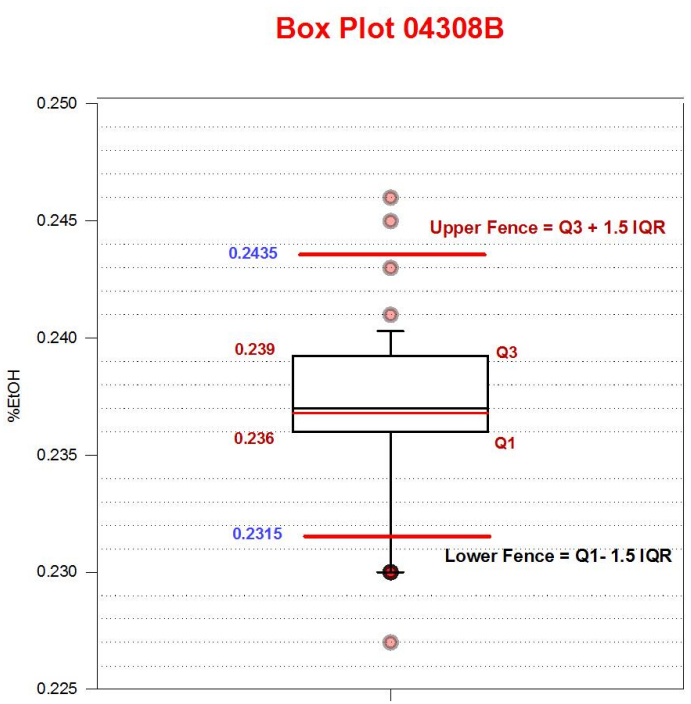
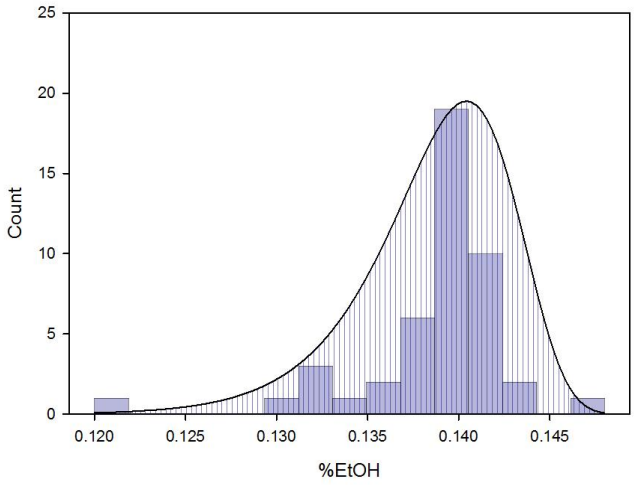


Figure 6 - SigmaPlot™ analysis of pools 04308A & 04308B







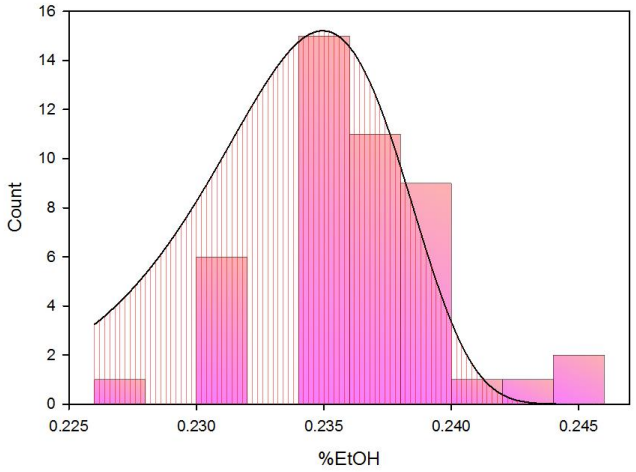
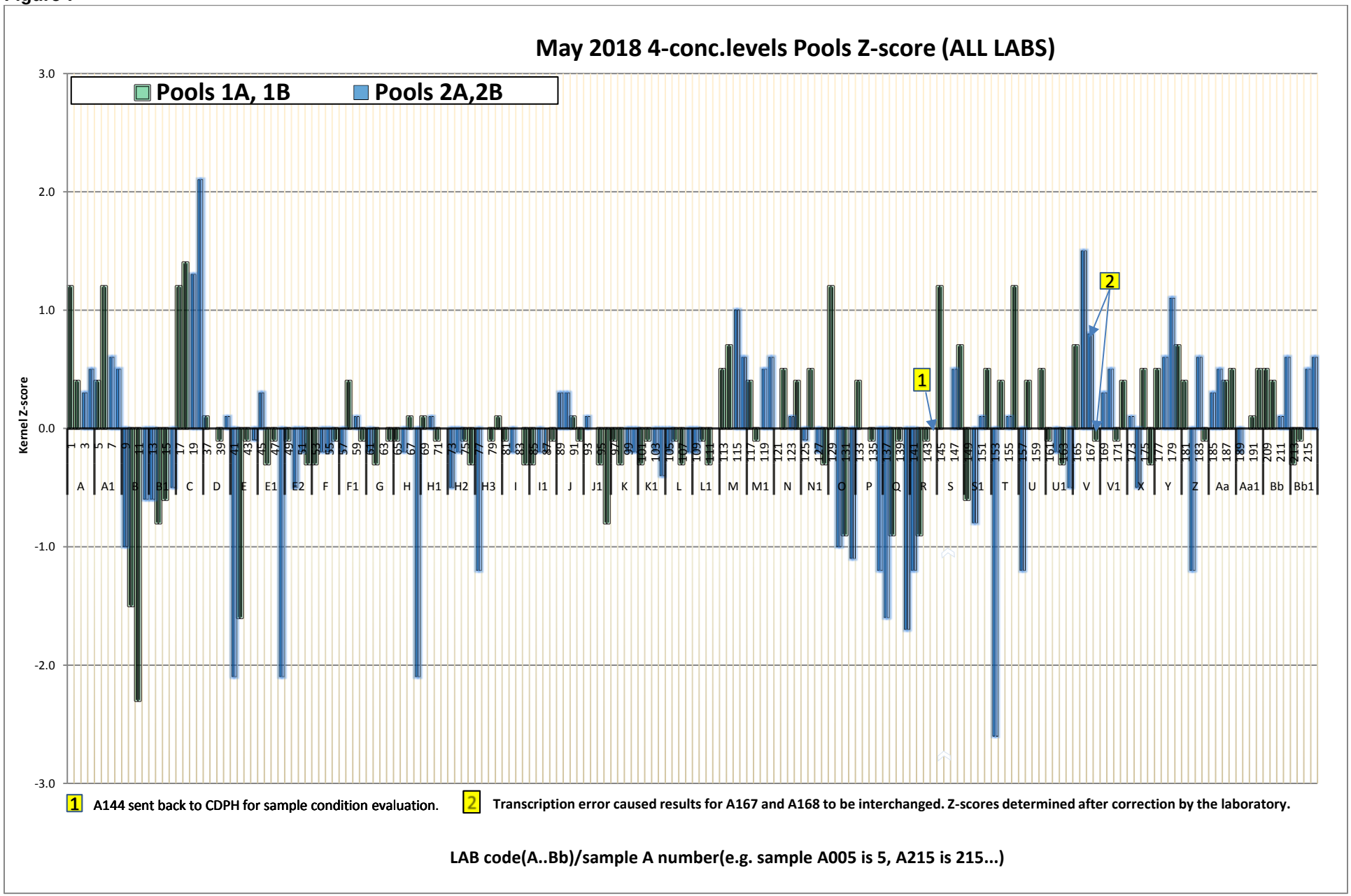


Figure 7


May 2018 4-conc.levels Pools Z-score (ALL LABS)
Pools 1A, 1B
Pools 2A,2B
Kernel Z-score
3.0
2.0
1.0
0.0
-1.0
-2.0
-3.0
A
A1
B
B1
C
D
E
E1
E2
F
F1
G
H
H1
H2
H3
I
I1
J
J1
K
K1
L
L1
M
M1
N
N1
O
P
Q
R
S
S1
T
U
U1
V
V1
X
Y
Z
Aa
Aa1
Bb
Bb1
1
2
1 A144 sent back to CDPH for sample condition evaluation.
2 Transcription error caused results for A167 and A168 to be interchanged. Z-scores determined after correction by the laboratory.
LAB code(A..Bb)/sample A number(e.g. sample A005 is 5, A215 is 215...)