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

January 3, 2017

TO: Participants in the November 2016 Proficiency Test in Forensic Alcohol Analysis

SUBJECT: Assigned Values and Acceptable Ranges of Results for the November 2016 Proficiency Test in Forensic Alcohol Analysis

Attached is a summary of the descriptive statistics for the November 2016 proficiency test in forensic alcohol analysis. The Department prepared two test pools (10246 and 10316) 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.

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 is described as the "Tier #2 interval." The Department also calculates a "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. Tier #1 is expected to include those laboratories demonstrating a high degree of accuracy. The second, wider tier would include those laboratories not as close to the central tendency as the first tier, but still accurate and therefore adequately competent. Again, historically, the Department has used the wider second tier to evaluate the laboratories' results.

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 for pool 10246 in this proficiency test were not 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 November 2016 test are summarized in Figure 4. Participants are identified by codes. An attachment to this letter provides codes for participants from your laboratory. The figure is provided for educational purposes only and was not used to formally evaluate a laboratory's performance.

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[1]. The quartile is a type of quantile: the first quartile (Q_1) is defined as the middle value between the lowest value 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 value between the median and the highest value 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 3, the data from the two pools 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. 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
http://www.cdph.ca.gov/programs/DFDRS/Pages/FDLB-ForensicAlcoholProgram.aspx

Sincerely,

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

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

Statistical Data for November 2016 Proficiency Test in Forensic Alcohol Analysis

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

Pool	Peer Group Mean	Tier #1	Tier #2
#1	0.127	0.119 – 0.135	0.11 – 0.13
#2	0.206	0.193 – 0.219	0.19 – 0.22

Table 2 Summary of Test Pool Data

Parameter	POOL 1 (10246)	POOL 2 (10316
Pre-distribution Data	Target Value 0.13% True Value[2] 0.125 Standard Deviation[2] 0.009	Target Value 0.21 True Value[2] 0.205 Standard Deviation[2] 0.0011
Descriptive statistics	Mean 0.127[3] Adjusted Mean[4] 0.127 Standard Error[5] 0.0003 Median 0.126 Standard Deviation 0.0023 Minimum 0.122 Maximum 0.134 Count 44	Mean[3] 0.206 Adjusted Mean[4] 0.206 Standard Error[5] 0.0005 Median 0.206 Standard Deviation 0.0033 Minimum 0.196 Maximum 0.213 Count 44
Descriptive statistics (box plot)	Q1 (25%) 0.126 Q3 (75%) 0.128 IQR 0.002 Lower Fence 0.123 Upper Fence 0.131	Q1 (25%) 0.204 Q3 (75%) 0.208 IQR 0.004 Lower Fence 0.198 Upper Fence 0.214
Histogram	Figure 1	Figure 2
Normal distribution?[6]	No (p=0.017)	Yes (p=0.260)
Box Plot (SigmaPlot)	Figure 3	Figure 3
Robust mean, X*[7]	0.126	0.206
Robust standard deviation, σ_{rob}	0.0010	0.0030
Fitness-for-purpose standard deviation, σ_p[8]	0.0035	0.0051
Consensus value (X_a) determined as Mode ($\mu_{1/2}$) of Gaussian Kernel distribution	0.1265	0.2060
Uncertainty of the consensus value, X_a , S.E.[9]	0.0002	0.0005
$X_a \pm$ S.E.	0.1265 ± 0.0002	0.2060 ± 0.0005
z-score	$z = \frac{X - X_a}{\sigma_p}$	$z = \frac{X - X_a}{\sigma_p}$

[2] Based on CDPH's Headspace Gas Chromatographic Method

[3] Participant data were rounded to 3-decimal place values.The 3-decimal place values are consistent with published estimates of the uncertainty of forensic alcohol methods).

[4] Mean determined from participant data after the removal of outlier(s)

[5] Standard Error of the Mean

[6] Shapiro-Wilk test used at 0.05 significance level.

[7] Robust mean of the results reported by the participants was calculated using Algorithm A in Annex C of ISO 13528:2005.

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

[9] Determined as the Standard Error of Mode using bootstrap simulation technique with bandwidth of $0.75*\sigma_p$

Figure 1

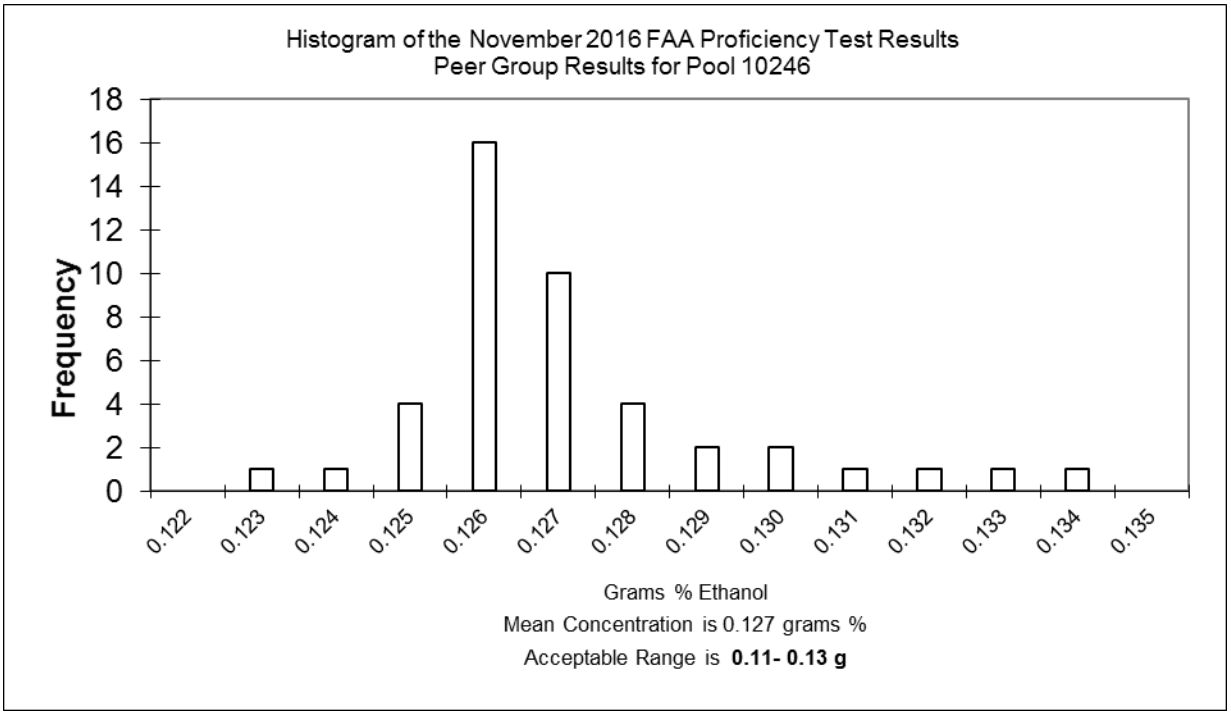


Figure 2

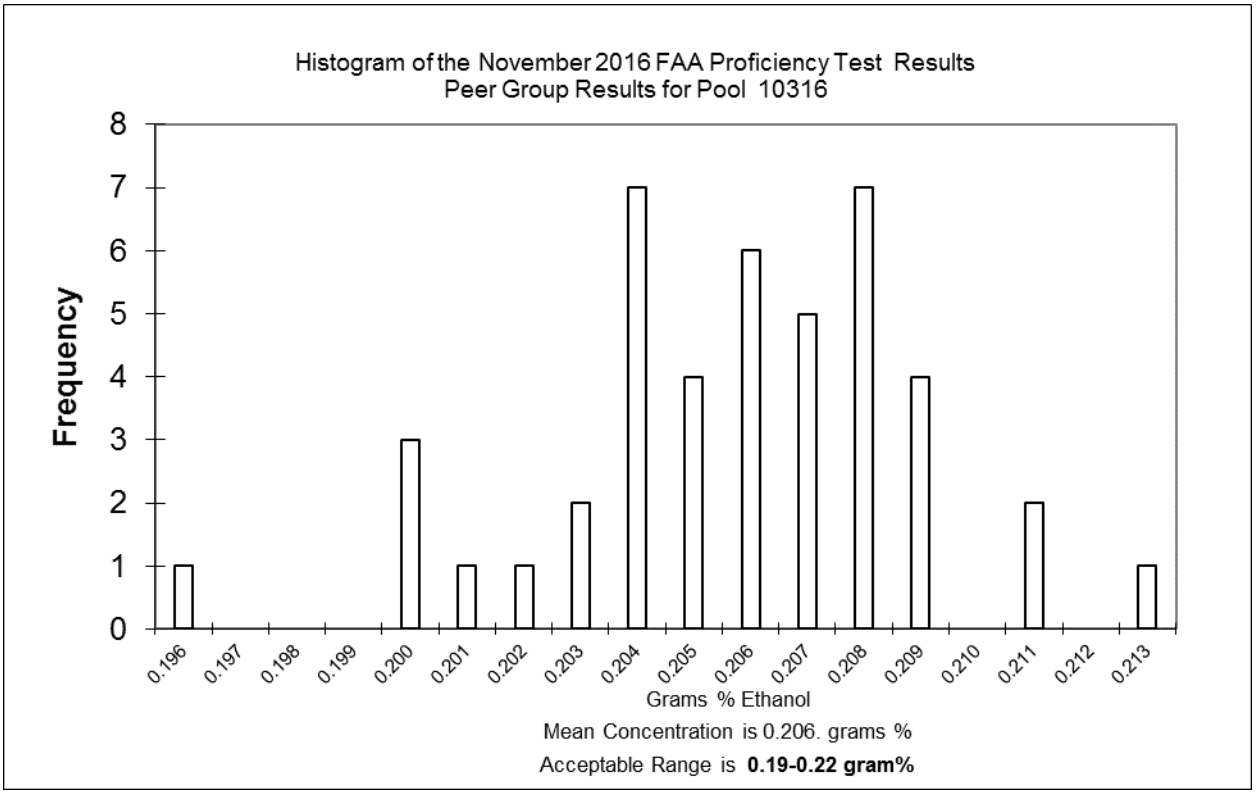


Figure 3 SigmaPlot analysis of pools 10246 & 10316

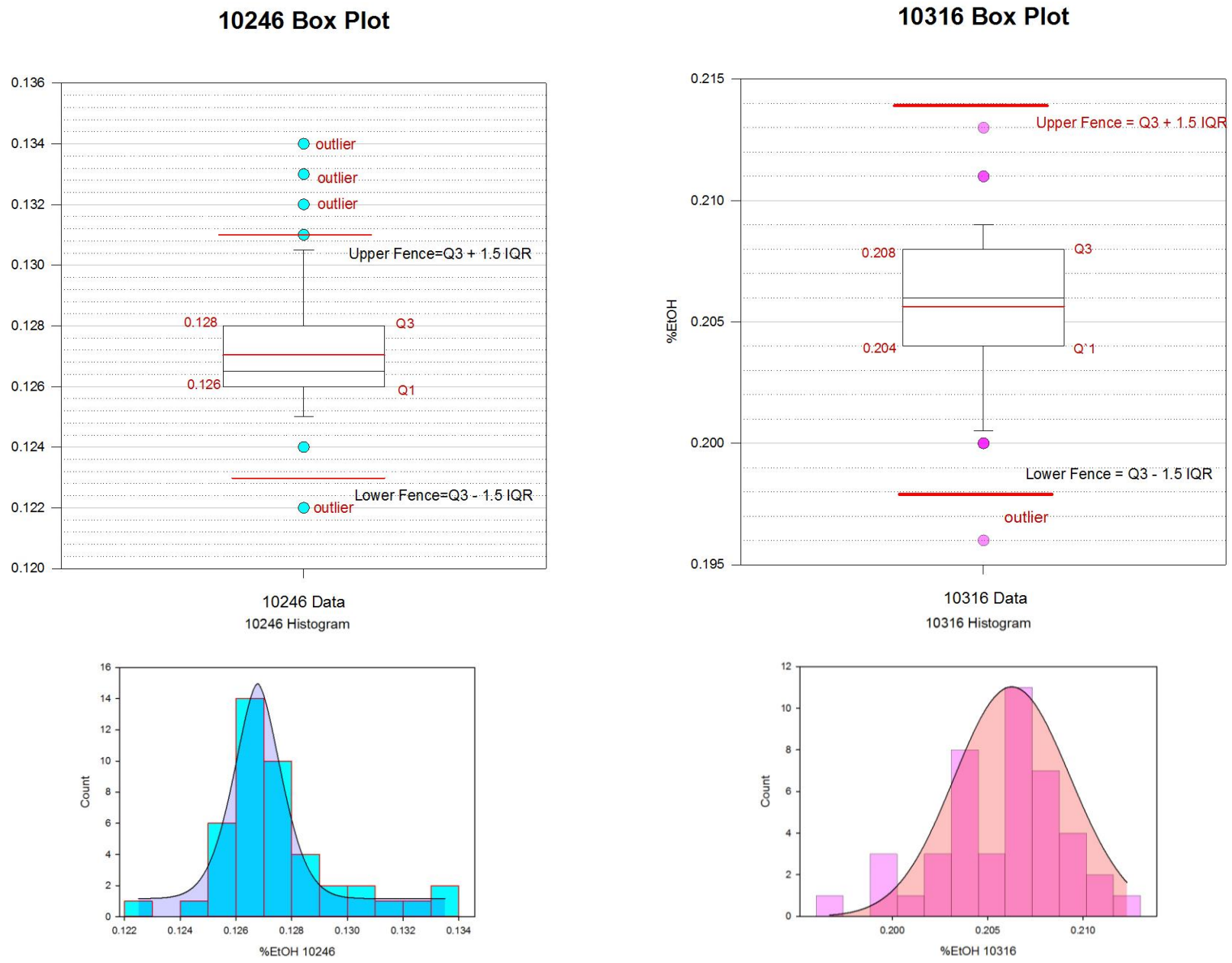
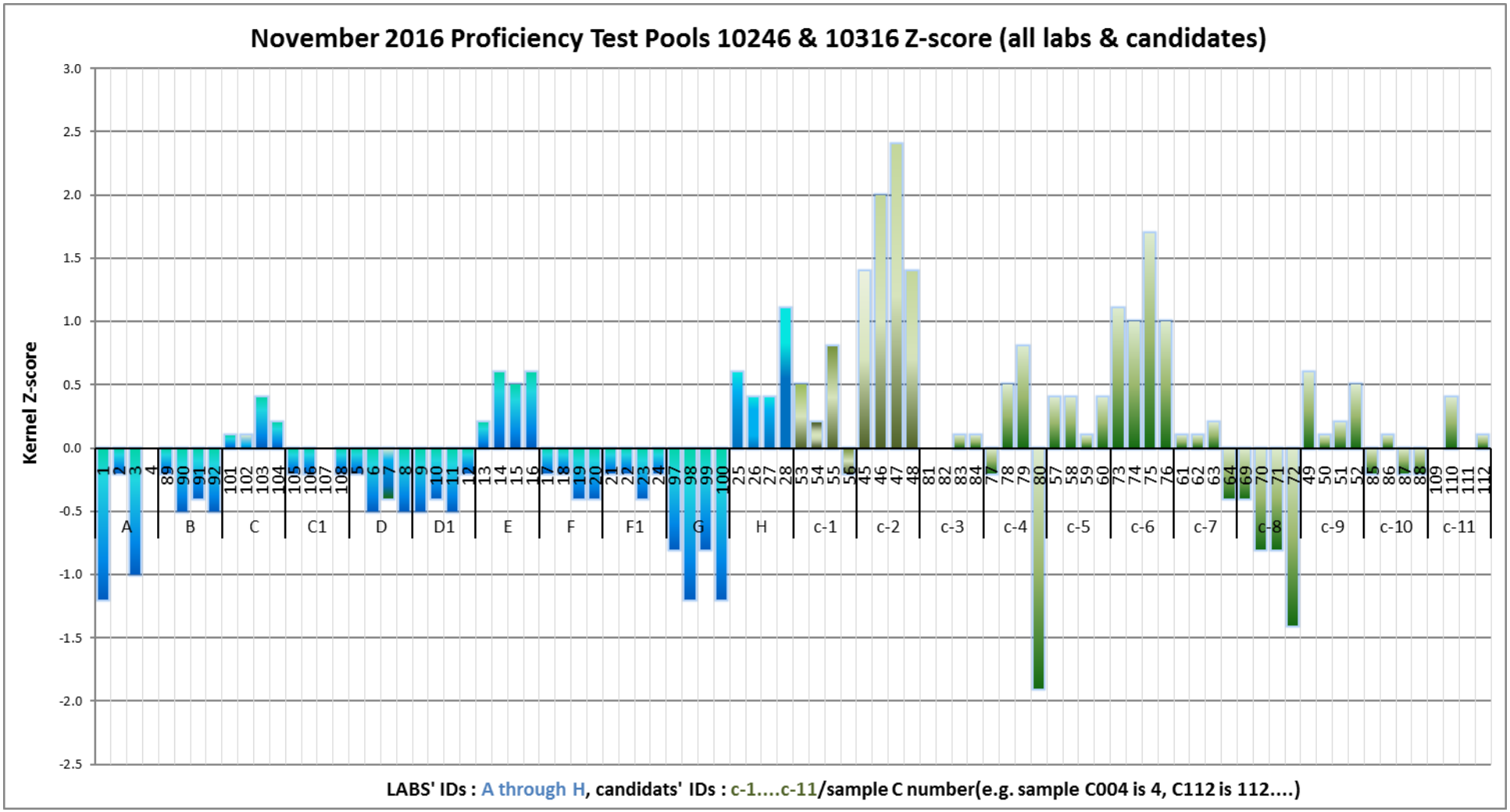


Figure 4


November 2016 Proficiency Test Pools 10246 & 10316 Z-score (all labs & candidates)
Kernel Z-score
3.0
2.5
2.0
1.5
1.0
0.5
0.0
-0.5
-1.0
-1.5
-2.0
-2.5
A
B
C
C1
D
D1
E
F
F1
G
H
c-1
c-2
c-3
c-4
c-5
c-6
c-7
c-8
c-9
c-10
c-11
LABS' IDs : A through H, candidats' IDs : c-1....c-11/sample C number(e.g. sample C004 is 4, C112 is 112....)