STATISTICAL APPROACH FOR VALIDATING A MODIFIED LABORATORY METHOD

BY

Khalid Saleh AL-Ghamdi

A Thesis Presented to the
DEANSHIP OF GRADUATE STUDIES

KING FAHD UNIVERSITY OF PETROLEUM & MINERALS

DHAHRAN, SAUDI ARABIA

In Partial Fulfillment of the
Requirements for the Degree of

MASTER OF SCIENCE

In

APPLIED STATISTICS

March 2018

STATISTICAL APPROACH FOR VALIDATING A MODIFIED LABORATORY METHOD

Khalid Saleh AL-Ghamdi

APPLIED STATISTICS

March 2018

KING FAHD UNIVERSITY OF PETROLEUM & MINERALS

DHAHRAN- 31261, SAUDI ARABIA

DEANSHIP OF GRADUATE STUDIES

This thesis, written by **Khalid Saleh AL-Ghamdi** under the direction of his thesis advisor and approved by his thesis committee, has been presented and accepted by the Dean of Graduate Studies, in partial fulfillment of the requirements for the degree of **MASTER OF SCIENCE IN APPLIED STATISTICS**

Prof. Muhammad Riaz
(Advisor)

Dr. Hussain Al-Attas
Department Chairman

Dr. Saddam A. Abbasi
(Co-Advisor)

Dr. Salam A. Zummo
Dean of Graduate Studies

Dr. Nasir Abbas
(Member)

7/5/18
Date

Dr. Walid Sabah. Al-Sabah
(Member)

Dr. Mohammad Hafidz Omar
(Member)

To My Parents and Family

ACKNOWLEDGMENTS

I am a chemist by both academic training and career, so when I started studying applied statistics, I came across multiple difficulties. Firstly, I'm part time student and my workplace –Ras Tanura petroleum refinery QC/QA laboratory - is about one hour drive to my educational institute (KFUPM). Secondly, I was not familiar with the nomenclature used by statisticians. Thirdly, I had my parents and my family to take care of. Nonetheless, I was able to pass through this journey successfully. God of Allah created for me facilitators who helped me to continue the journey until the end. Starting with my academic and thesis advisor Prof. Muhammad Riaz for his unlimited support and guidance and Prof.Anwar Joarder who helped me a lot in overcoming the shock of studying statistics, which was the first course in the applied statistics program "Mathematical Statistics". Almost all professors who taught me during applied statistics program at KFUPM were very helpful. The appreciation also goes to my employer Saudi Aramco company represented by Ras Tanura Refinery engineering department for its support. In fact, the idea of this thesis work was solely developed due to the frequent technical challenges that I face at Ras Tanura Refinery QC/QA laboratory.

Special appreciation goes to my beloved family; wife (Lubna), my daughters (Ruaa, Rasha, Rahaf and Sarah) and my son (Abdul-Rahaman) for their cooperation in availing their time for my use in solving applied statistics homework's and prepare for exams.

Finally, I am very thankful to my deceased parents -*May Allah be merciful to them*- who kept investing on me since I was a newborn baby. Their help and support cannot be quantified by all mathematical and statistical tools that humankind has invented and will invent.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS/NOTATIONS

[m]	Bracketed Parenthesis indicates concentration of the measurand "m"
Σ	Capital Greek sigma means add up all the values.
Π	Capital Greek symbol pronounced "pi". Means multiplication of the values given in the data set. e.g. {X} =1, 4, 5, 3. The Π(X) = 1*4*5*3= 60.
θ	The Greek symbol pronounced "theta". Denoting the Characteristic parameters, i.e. the property under the study.
∧	Denotes an estimate of a parameter, e.g. for parameter θ, its estimate is $\hat{\theta}$.
$\bar{x}$	Sample Mean.
μ	The Greek letter pronounced "meo" and represents the population mean. Also called True Value.
n	Sample Size. The total number of observations.
N	Population Size.
σ	Population Standard Deviation.
s	Sample Standard Deviation.
σ^2	Population Variance.
μ	Population Mean/ True Value
y	Response Variable. Also called Dependent Variable.
x	Predictor Variable, independent variable. Also called Regressor.
χ^2	Pronounced "Chi Squared" was introduced by K.Pearson. Originally, it refers the sum of squares of the deviations between observed and expected values therefore, it is a suitable indicator of the goodness-of-fit.
Φ	Is the cumulative distribution function (cdf) of the standard normal distribution that yields the P-Value; $P = \Phi(\frac{\bar{y}-\bar{x}}{s*\sqrt{n}} * \sqrt{n})$ Where; s: Sample Standard Deviation. n: Sample Size. $\bar{x}$: Average of group (x) of sample size (n). $\bar{y}$: Average of group (y) of sample size (n).
H_o	The Null Hypothesis.
H_1	The Alternative Hypothesis.
r	Correlation Coefficient.
R^2	Coefficient of Determination.
R&r	In the context of precision of measurements; R is denoting Reproducibility Limits while r is denoting Repeatability Limits.

m	In the context of Design Of Experiment (DOE), m signifies either the number of independent variables Xs or the levels for the predictor variables X_s (number of factor levels)
P	The number of regressors, number of independent variables or the number of predictors.
S_R	Inter-laboratory Precision. Also called Inter-laboratory Reproducibility.
β_s	The number of regression coefficients including the intercept. So for predictor variable X with five levels, the total number of regression coefficients $\beta = 6$.
α	(Alpha) False positive
β	(Beta) False negative
ε	(epsilon)
F(A)	Fluoride Peak Area.
F(H)	Fluoride Peak Height.
Cl (A)	Chloride Peak Area.
Cl (H)	Chloride Peak Height.
Br (A)	Bromide Peak Area.
Br (H)	Bromide Peak Height.
STD	Standard
ppm	Part Per Million
PA	Peak Area
PH	Peak Height
AD	Anderson-Darling
RJ	Rayn-Joiner
KS	Kolmogorov-Smirnov
$CC\alpha$	Probability of detecting false negative $CC\alpha$
$CC\beta$	Minimum detectable net concentration. Probability of detecting false negative.

ABSTRACT

Full Name : [Khalid Saleh AL-Ghamdi

Thesis Title : [Statistical Approach for Validating A Laboratory Modified Method]

Major Field : [Applied Statistics]

Date of Degree : [May 2018]

The standardized testing procedures (SOPs) that are recognized globally are used as reference methods for testing. Developing and standardizing laboratory testing methods helps different laboratories around the world in producing data that are not significantly different from each other. Typically, any laboratory adopting any of these SOPs has to strictly comply with the requirements of adopted SOP. In real life, some of these SOPs are hardly to fulfill their requirements due to operational difficulties and cost constraints. So sometimes end-users of these SOPs tend to deviate from rules and conditions dictated by these SOPs in order to make them easier for implementation. Here comes the importance of validating the validity of the modified method and checking its performance against the original method.

Into this thesis, the original SOP is called UOP991 with a working range of (0.1-to-1) part per million (ppm) and the proposed modified method was given a name High Range "HR" with a working range ten times higher than the original method which will make the working range from (1-to-10 ppm). Three changes were done on the original method. These changes are; elongating the calibration frequencies for the measuring equipment to be monthly instead of upon use, change the method working range to be calibrated using high range

regressors/calibrants and calibrating the measurand Fluoride by using the regression method Linear Least-Squares instead of the Quadratic one, which is mandated as per the UOP991.

Statistical tools in regression analysis, detecting outliers, testing data for normality and homoscedasticity and multiple comparisons procedures played a pivotal role in validating the validity of the modified method "HR" against the original method UOP991. This thesis proved both statistically and experimentally that test results obtained from the HR method are not significantly different from the ones obtained from original method UOP991. Also, the operational life of the built calibration models were tested over a period of one year, which helps in proving the validity of the new calibration frequency to be on monthly basis instead of the upon use as dictated by UOP991. Moreover, test results generated from the Fluoride calibration models built by using the regression analysis Linear Least-Squares is as good as the ones produced by Fluoride models calibrated by quadratic function.

ملخص الرسالة

الاسم الكامل: خالد صالح الغامدي

عنوان الرسالة: طريقة إحصائية لشرعنّة طريقة تحليل مخبّري معتمّدة ولكن محوّرة.

التخصص: الإحصاء التطبيقي

تاريخ الدرجة العلمية: مارس 2018

طُرق التحليل المخبّري المعتمدة عالمياً هي طُرق خضعت لمعاييّر عالميا قبل إعتمادها وأستخدامها كمرجع موثوق لإجراء تحليل مخبري ما. هذه الطُرق المخبريّة المعتمدة عالمياً تُعرف بإسم Standard Operating Procedures (SOPs). والإلتزام بهذه الطُرق المخبّرية يُعتبر مُلزماً لجميع مستخدميها كي تُصبح نتآئج التحليل الناتجة عن جميع المختبرات المستخدمة لنفس طريقة التحليل مُتقاربة جداً. بعض هذه الطُرق لديها نُقاط ضعف تجعل إستخداما والإلتزام بها من قبل مختبرات التحليل امراً مُكلفاً إقتصادياً وعملياً. لدى هذه الرسالة اثبتت عملياً وإحصائياً أن طريق التحليل المخبري المرموز لها ب: UOP991 يُمكن جعلها اكثر سهولة واقل تكلفة عملياً وذلك بإجراء ثلاثة تغُيرات عليها. هذه التغيُرات الثلاثة هي: اولاً، إطالة عمر المُعايرة Calibration وجعلهُ شهرياً بدلاً من عمل معايرة لجهاز التحليل مع كل إستخدم لهُ. ثانياً ، تغيير نطاق التحليل من المدى المُعتمد وهو من 0.1 –الى- 1 جزء من الملّيون (ppm) إلى النطاق الجديد وهو من 1- الى-10 جزء من المليون. ثالثاً، إستبدال طريقة المُعايرة المذكورة في الطريقة الاصلية UOP991 للعنصر فلوريد من المُعايرة بإستخدام التحليل التراجعي التربيعي Quadratic الى التحليل التراجعي الخطيّ Linear Least-Squares . وإثبات بإستخدام المفاهيم الإحصائية ان نتآئج التحليل الصادرة عن طريقة التحليل المُحوّرة و المُسماة HR لإتختلف بشكل كبير عن نتآئج التحليل الصادرة بإستخدام الطريقة الاصليّة UOP991.

إسُتخدمت طُرق إحصائية مثل: التحليل التراجعي Regression Analysis و طُرق مُعالجة البيانات من حيث وجود القرآئات الشاذة Outliers و التأكيد على ان البيانات الرقميّة ذات توزيع طبيعي Normally Distributed و إثبات ثبات

التباين بين القرآئات Homoscedasticity وعمل مقارانات إحصائية بين القرآئات Multiple Comparisons Procedures في إثبات صلاحيّة طريقة التحليل المُحوّرة HR ومُقاربتها من حيث الأدآء للطريقة الاصلّية UOP991.

CHAPTER 1

INTRODUCTION

The scope of work, work significance work objectives and their enablers are discussed into this chapter. Also, challenges and strategy followed on this research work are explained too. Finally, summary tables for the developed calibration models are illustrated in tables. Scope of Work

1.1.Scope of Work

This thesis is studying the validity of using a proposed modified method called High Range (HR) as an alternative for the original low range standard operating procedure (SOP) method called UOP991. The scope of UOP991 is for determining the concentrations of chemical elements; Fluoride (F), Chloride (Cl) and Bromide (Br) in petroleum products at trace concentration levels (i.e. at concentration level less than one part per million $\leq$ 1-ppm) in liquid organics by a technique called Ion Chromatography (IC) hyphenated with Combustion system, so this instrumental analytical technique is abbreviated CIC [1].

1.2. Work Objectives

The objectives of this thesis are summarized into the following points:

1.2.1. Checking the capability of the proposed modified method "HR" in reproducing data that are not significantly different from the ones produced by the standard method "UOP991".

1.2.2. Extending the calibration frequencies of the chemical measuring equipment "CIC" to be monthly instead of upon use. This is explored by checking the stability of six HR calibration models that were built over one-year period and elven calibration models that were built per the standard procedure UOP991.

1.2.3. Developing an in-house quality control "QC" Chloride audit sample and checking its stability "*shelf-life*". This was achieved by sampling a Naphtha sample from the refinery plant and testing its chloride content by both the five HR calibration models and by the eleven UOP991 calibration models over a period of one year. This sample was named as "SQC (1.0743-ppm) Oct 14/2015. Note that this Chloride QC audit sample is named by its sampling date and has a measurand concentration level around one ppm.

1.2.4. Using the Linear Regression method for modelling the Fluoride measurand instead of the quadratic one which was recommended by the UOP991 Standard Operating Method (SOP).

The ultimate objective of the whole thesis work is developing an economically design of calibration practice which is replacing the need to follow the expensive and labor intensive procedure covered by the standard testing method UOP991. So the proposed modified method “HR” was validated by single laboratory, using calibration standards – which are in the context of statistical science are called regressors – which are ten times higher in concentrations than the low concentration ones that are covered by the standard method UOP991. So these high in concentration calibration standards/regressors are easier to prepare and more stable. Finally, extending the calibration frequencies required for this lab chemical measuring equipment “CIC” to be monthly rather than upon use.

1.3. Enablers

The enablers used for achieving the thesis objectives are:

1.3.1. Checking the limit of quantification (LOQ) for the modified method “HR” in reproducing test results that are not significantly different from the ones produced by the standard method “UOP991”; hence, capability of quantification of the HR method.

1.3.2. Building eleven UOP991 calibration models and five HR calibration models over one year period and checking their validity/stability by analyzing samples with known measurand concentration. This was done only for the Chloride measurand due to its criticality and availability of samples with known chloride concentration. Data for chloride generated by these calibration models were compared by looking for significance of differences among them using statistical tools.

1.3.3. Obtaining real life samples from the petroleum refinery process and testing its chloride concentration level by both UOP991 and HR calibration models over one year to assure the stability of this Naphtha based sample.

1.3.4. Developing both linear and quadratic calibration functions for the built Fluoride calibration models then testing the significance difference in the obtained Fluoride values by these two methods; UOP991 and the modified one

HR for the same calibration model. Also doing multiple comparisons looking for significance differences for the obtained test results, such simple but useful mathematical tool “Two Points Linear Interpolation” was used to confirm the linearity of the curve that was obtained for the chosen Fluoride calibration/ regressors concentration levels.

1.4. Work Significance

Determining the concentrations of halides -these chemical elements; Fluoride (F), Chloride (Cl) and Bromide (Br) are commonly called halides- is quite important at any petroleum refining industry. Since they play a major role in corrosion at the refinery processing plants. Need not to say that corrosion is one of the common permanent factors that negatively impact the economics of running any chemical plants. So monitoring the presence and concentration levels of these halides is very essential to control corrosion rate. All petroleum refineries laboratories conduct halides analysis in both organic and aqueous samples. Different laboratory analytical methods are involved for different halide types, different concentration levels and for different sample matrices. This thesis is trying to enhance the lab performance regarding testing halides in liquid organic samples by proposing a modified lab test method called (HR) as an alternative for the original one (UOP991). The main motivations behind this proposed method (HR UOP991) are;

1.4.1. Lowering the Operational Costs of the Lab Measuring Instrument. According to the UOP991, the measuring equipment (CIC) has to be calibrated each time the samples to be measured. This is done by preparing and running four multi-elements calibration standards which are ; 0.1-ppm, 0.2-ppm, 0.5-ppm and 1-ppm standards and before these standards are run by the CIC chemical measuring equipment, a blank has to be measured several times till the background readings for these elements (F,Cl & Br) are close to zero. Usually this will require running the blank sample about five times. Each single run by the CIC requires about half an hour. So one calibration sequence will consume at least four to five hours. This is regardless of the time needed to test each sample one time by the CIC. Spending five hours to calibrate the lab measuring equipment each time sample to be tested is too much time loss for petroleum refinery operational Quality Control (QC), Quality Assurance (QA) laboratory. So extending the calibration frequencies while maintaining the validity of the generated tests results is one of the main goals behind this thesis work.

1.4.2. Developing In-House Quality Control Audit Sample. It is a rule of thumb that the best QC sample is the one that has a sample matrix if not the same, then almost the same as the routinely measured samples. But in reality this is not possible most of the time. So lab analysts usually prepare their own synthetic QC audit samples. For this study, it was possible to obtain a process stream naphtha sample which has a sample matrix very similar to most of routinely

tested samples. This sample was collected from the refinery process plant on October 14,2015 and tested for the first time for its chloride concentration level which was found to be [Cl]=1.0743-ppm. Then since that date, this QC audit sample was tested for a period of 14 consecutive months by both the original UOP991 and the proposed HR method to check the stability and validity of the obtained test results for chloride content in this SQC sample.

1.5. Challenges

This fifteen months work experienced some operational difficulties, the major ones were;

1.5.1. **Interferences Due to High Background Contamination.** Several challenges were faced during the course of this thesis work but the two major ones were; detecting elements at trace level (i.e. less than 1-ppm) while their natural abundance (i.e. their concentrations levels that exist naturally on earth's crust) is extremely high. Recall that about 126-ppm of Earth's crust is chlorine (Cl), Fluorine (F) is about 600-to-700-ppm and Bromide (Br) is the least one which is about 2.5-ppm of the Earth's crust, so very high background affecting the performance of both the original UOP991 and the modified one HR UOP991. This effect was found very prominent in case of Fluorine. So how can I get chemical reagents to prepare the calibration standards –which are the regressors-, and blank samples in workplace, which is free from Fluorine, Chlorine and Bromide!

1.5.2. **Stability of Calibration Standards (Regressors), Blank Sample and Samples.** All the calibration standards prepared for calibrating the measuring instrument are Toluene based and this chemical "Toluene" is highly volatile at room temperature (i.e. 25^{o}C). So whenever these trace concentrations levels (0.1-ppm, 0.2-ppm,0.5-ppm and 1.0-ppm) were undergoing preparation at room temperature, some loss of the measurands concentration was taking place simultaneously due to the volatility of Toluene. Now here it comes one of the advantages of using the modified method (HR UOP991) since the prepared concentration levels were (1-ppm, 2-ppm, 5-ppm and 10-ppm) so the effect of measurands loss due to evaporation of the solvent Toluene is less. Note that this one way in which errors is occurring and cannot be controlled. The other one was when the whole calibration standards, blank and routine samples being loaded on the measuring equipment auto-sampler, each time the sample vial is punctured, the vial septum will remain opened partially for about half-hour till second injection take place, so evaporation of sample will occur and consequently the repeatability values will be adversely affected.

1.5.3. **Lack of Certified Reference Materials.** These are chemicals with known composition with respect to both the chemical types they contain and their concentration levels associated with their uncertainty values. Usually, these materials are used for either calibrating the lab measuring equipment or as quality control standard to monitor the measuring equipment performance. In real life laboratory analytical work, most of the time it is not possible to have a reference material that has matrix that is identically match to samples being

analyzed. So for this study, a process stream sample (PLT15 C-300 Bottom) collected on October 14, 2015 was used as quality audit sample for the parameter Chloride. For the other two elements; Fluoride and Bromide, no quality audit samples were used.

1.5.4. **Measuring Equipment:** The CIC is very delicate lab analyzer. It is a hyphenated technique of Ion Chromatograph (IC) connected to Combustion System. This piece of technology is relatively new and not widely used at operational laboratory due to its cost and being single application lab analyzer. So to crosscheck the performance of this equipment by other laboratory in the region that has a similar measuring equipment was not possible and whenever this lab measuring equipment was malfunction, the whole study was kept on hold.

1.5.5. **Method Validation:** One of the simplest and reliable ways to do bias study is to compare the obtained test result with results of other laboratories around the globe that are doing the same test using the same Standard Operating Procedure (SOP). This is called inter-laboratory study (also known as Proficiency Testing "PT"). Unfortunately, doing halides determination in liquid organics by CIC is not common technique due the cost of this highly sophisticated equipment and being a single application lab analyzer. Nonetheless, I was able to compare the PT sample (ASTM PT NP1602 Naphtha) for its reported Chloride test result by other SOPs; ASTM D5808 and ASTM D5194 vs. UOP991 and there was some sort of the agreement among their readings. See the experimental and findings

section of this thesis – chapter 3-. Chapter five for the method validation part and chapter six for findings and conclusion.

1.5.6. **Design of Experiment (DOE):** Sometimes, it is much easier to construct a work from scratch rather than fixing an already existing work. DOE is step number one in developing any new testing method. DOE phase can be metaphorically represented as the engineering design stage of any building construction project. However, to do DOE work on an already existing method and considering the limitations aforementioned into this challenges section makes a DOE part is quite difficult to do it correctly. Therefore, this thesis work is improving the quality of a process that is already in action.

1.6. Strategy Followed

Working outside the scope of SOP like UOP991 is considered as an invalid action, unless a validation work is done to prove the quality of the obtained test results from the modified SOP. So the strategy followed into this work are covered by the following points;

1.6.1 **Standardizing the Language:** On the field of Metrology – the science of measurement-, the language is well developed. Academia vocabulary was abandoned and replaced by internationally recognized terminologies such as the ones covered by the international standard "International Vocabulary of Metrology (VIM) ISO Guide99 and ASTM E456 [2-3]. So in analytical chemistry, the term Analyte is widely used to denote the specie under the study while in metrological standardized work it is replaced by the term Measurand. Therefore, Fluoride, Chloride and Bromide are referred to into this study as the measurands. Also, in statistical science in particular on the context of regression analysis, the term *regressor* variable is used to denote the predictor variable x that will predict the response value y. Its equivalent term in analytical chemistry is *calibrant*, but the ISO Guide 99 is recommending using the term *Calibrator*. Using such standardized, globally recognized language will make the understanding of this work much simpler by people from different part of this world.

1.6.2 **Glossary:** The work done on this thesis dictates the need to develop a glossary section that shows the definitions for all technical terminologies mentioned within this thesis. Whenever the definition is taken directly from a reference standardization body such as American Society for Testing & Materials (ASTM) or the International Standardization Organization (ISO), the abbreviation of the organization is mentioned.

1.6.3 **Symbols & Abbreviations:** Master list was developed to cover all symbols and abbreviations that are mentioned within this work.

1.6.4 **Literature Search:** Only standardized testing methods in particular the ASTM and ISO were considered in the literature search. Although some academic publications were considered, yet they were given less weight into this work. Reason for this approach is due to nature of this new global market which operates mainly via standardized, globally recognized Standard Operating Procedures (SOPs). Although this work was covering three elements; F,Cl and Br, yet more emphasis was given to the Chloride (Cl) since it's the most common corrosive element that is more encountered at petroleum refineries. Some of common SOPs used at petroleum refineries to monitor the chloride (Cl) content in liquid organic samples are [4-9].

1.6.5 **Method Validation:** The process of establishing the performance characteristics and assuring that the (developed method/modified) is fit for purpose is referred to as method validation. The typical method performance characteristics are; Selectivity, LOD & LOQ, Working Range, Accuracy, Precision, Calibration and Traceability, Linearity, Ruggedness. A number of protocols and guidelines were developed to address the systematic way of validating a laboratory testing procedures, some of which are [10-17]. This work is considered as validation by single laboratory.

Note that for any laboratory in order to be an ISO17025 certified/accredited lab [13], it has to prove that its testing methods are valid analytical laboratory method such as ISO and ASTM SOPs and if the original method is modified, then the lab should conduct a validation work for the modification done on the original SOP.

1.6.6 **Statistical Calculations:** Some of international standards related to statistical concepts mentioned by this thesis work are [18-22]. All calculations were carried out by using the statistical software package MINITAB17 [23].

1.7. Summary of Experimental Work

The main theme of this work is to check is there any significant difference between the original testing method UOP991 and the proposed modified High Range method (HR)? So to explore this, a plenty of time was needed to develop multiple regression (calibration) models both for the low range original method UOP991 and the high range one (HR) . Also note that these regression models have to be constructed over a reasonable intervals in order to assess their operational life which is in turn is important to set the calibration frequencies of the lab measuring equipment CIC. The table below summarizes these calibration/ regression models that were done.

Table 1.1: Summary Data for UOP991 Method Calibration Datasets.

Calibration Model No.	Date	No. of Elements	No. of Concentration Levels	No. of Replicates	Peak Area (PA)/ Peak Height (PH)
1	JAN26, 2016	F, Cl	except for Cl are 3	Single Observation	✓
2	FEB22, 2016	Cl	4	Single Observation	✓
3	April25,2016	F, Cl	4	Single Observation	✓
4	May12,2016	F, Cl	4	5 replicates	✓
5	June1,2016	F, Cl	4	3 replicates	✓
6	June9,2016	F, Cl	4	4 replicates	✓
7	August15,2016	Cl	5	2 replicates	✓
8	August30,2016	Cl	5	3 replicates	✓
9	Sep11,2016	F, Cl	4	3 replicates	✓
10	Oct19,2016	Cl	5	Single Observation	✓
11	Oct29,2016	F, Cl	4	3 replicates	✓

Table 1.2: Summary Data for (HR) Method Calibration Datasets.

Calibration Model No.	Date	No. of Elements	No. of Concentration Levels	No. of Replicates	Peak Area (PA)/ Peak Height (PH)
1	April26,2016	F, Cl	10	4 replicates	✓
2	May17,2016	F, Cl	4	Single Observation	✓
3	June10,2016	F, Cl	10	4 replicates	✓
4	August21,2016	Cl	10	3 replicates	✓
5	Sep10,2016	F, Cl	10	3 replicates	✓
6	Oct29,2016	F, Cl	7	3 replicates	✓

CHAPTER 2

STATISTICAL CONCEPTS

Statistics is an empirical science in origin. Based on experimental observations, statistical concepts were developed. This chapter is giving brief background about the statistical concepts used into this thesis work. More elaboration about some them and statistical terminologies can be seen in the glossary section and cited references. Detailed discussion about the science of *metrology* can be found in [25-28].

2.1 Metrology

2.2.1 Dynamic and Static Measurement: Based on the nature of the quantity being measured; measurements are classified into two categories; Dynamic and Static Measurements. So, since this work is dealing with highly volatile samples – recall that these samples are organic liquids based on Toluene and Naphtha which do evaporate continuously with time at room temperature- then the measurement could be either Dynamic Measurement (this is the case when the quantity of measurement varies as a function of time) or Static Measurement. Static measurement is defined as measurement of a quantity that is assumed to remain unchanged for the amount of time required to perform the measurement.

2.2.2 Direct and Indirect Measurement: Based on the method used by which measurement results are obtained: Direct measurement is performed by using measuring instrument that store a unit or scale for the quantity being measured such as using scaled ruler for measuring the length. On the other hand, Indirect measurement in which the value of a quantity is determined based on results for direct quantities functionally related to the quantity being determined. Typical example of indirect measurement is the use of calibration function that is developed by running series of standards (rgressors x_s), recording their responses y_s obtained from the measuring device then developing this mathematical model $Y = a + bX + \varepsilon$. This mathematical model (i.e. the calibration function) will be used to correlate the measuring instrument response value y of the measurand to the measurand actual concentration level x.

2.2 Calibration and Interpretation of Regression Function

Calibration from chemical analytical perspective, can be represented by a set of operations that relates quantities $x_{i's}$ in the sample domain with quantities $y_{i's}$ in the signal domain and mathematically expressed as $y = f(x)$. These calibration data (pairs of analytical quantities-x_i and measuring equipment response for each analytical quantity y_i) are used for developing the calibration function. Depending on the type of the measuring equipment, different types of calibration are used. For instance, *direct reference calibration* could be used in X-Ray based analysis .This is the case of absolute measurement –i.e. one calibration function is valid for all-. But most of the instrumental laboratory analysis is done via the *indirect reference calibration* method and the frequently used calibration model for the indirect reference calibration is the linear model, $y = a + bx + \varepsilon$. More details about the laboratory calibration methods and requirements can be found on these references [25-29] and a tutorial review papers addressing the topic of analytical calibration based on least-squares linear regression for instrumental techniques is cited in references [30-41]. Various types of regression method are used for calibrating laboratory measuring equipment, but the most widely used one is the least squares regression, which works by findings the best curve through the date that minimizes the residuals sum of squares. Linear model of calibration was developed by Lieberman, Miller, and Hamilton (1967) [34]. A concise statistical background about regression analysis for single numerical predictor variables can be found into this ASTM standard [42].

2.2.1 Linearity of the Calibration Function. It is one of the performance characteristics in method validation. The term linearity is quite confusing and it is recommended to use the term calibration *curve* function instead of calibration linear function, why? Because in statistical science when the regression functions is said to be linear, then this means linear with respect to the power of rgressors' coefficients being one. From chemical analytical perspective saying a linear regression line signifies that the relationship between the measuring equipment response and the measurand is linear –which is a requirement for valid analytical method- but still some measurands behave in nonlinear way and a calibration curve -could be quadratic- still can be used as the calibration function in valid instrumental analytical method. So mathematically speaking, a quadratic function could be linear with respect to its repressors' coefficients but once it is plotted, it is a curve, while linear line is also a linear function.

2.2.2 Importance of Linearity of the Calibration Function. Linearity is a measure of a method's ability to give a response that is directly proportional to the concentration of the measurand (i.e. material under the study); hence, the linearity region is the one used as the method working range. The linearity of the built calibration models was checked by two ways; one by the values of the correlation coefficient – as demonstrated in chapter three- and the second way by reproducing quality control audit values that are having concentration levels on both extremes of the calibration line – as demonstrated in chapter four-. Note

that the calibration function is developed based on the least square method. A standardized procedure for regression analysis is found in [42]. But in real life, almost all analytical chemical laboratories are using the inverse of this regression line for predicting the measurand concentration level based on the obtained measuring equipment response value and it turned to be that an inverse calibration predicts better than classical calibration which is based on least square method [43]. **Sensitivity.** Is the change in the analytical response (y-value/dependent variable) divided by the corresponding change in measurand concentration (x-value/regressor) so at a given value of the measurand z_o, sensitivity is defined as $Sensitivity = \left(\frac{dY}{dZ}\right)_{z_o}$.

If the calibration is a linear, the **sensitivity** is just calibration slope b at every value of measurand concentration. Two important method validation performance characteristics are directly derived from knowing the slope of the calibration line. These parameters are the Limit Of Detection (LOD) and Limit Of Quantification (LOQ). The second importance of having linearity in calibration function is the capability of identifying the **Limit of Detection (LOD).** It is the lowest concentration of measurand that can be detected and reliably distinguished from zero (or the noise level of the system), but not necessarily quantified. $Y_{LOD} = Y_{blank} + 3.3s_{blank}$ where; Y_{LOD} is the method limit of detection, Y_{blank} is the measuring equipment response to the blank sample and s_{blank} is the standard deviation of 10 readings for measured blank.

2.3 Modeling via Multiple Linear Regression

When the response variable y is related to more than one regressor ($x_1, x_2, \ldots x_k$) where k is the number of regressors), the regression equation will be;

$$Y = \beta_o + \beta_1 X_1 + \beta_2 X_2 + \cdots \beta_k X_k + \varepsilon. \quad (2\text{-}1)$$

Epsilon ε is the error term for the general regression model. The model is called linear in the parameters $\beta_o, \beta_1, \beta_2, \ldots, \beta_k$. β_o shows the intercept of the regression plane (since we're doing three dimensional regression) with the y-axis. The parameter β_1 indicates the expected change in the response (Y) per unit change in the regressor/calibration standard X_1 when X_2 *and* X_3 are held constant. In other words, β_1 is the method sensitivity (i.e. the change in the measuring equipment response "peak area" when there is change in the concentration level of the regressors/calibration standard). Similarly, β_2 represents the change in the response (Y) when the concentration levels of both X_1 and X_3 are held constant. Finally, β_3 represents the change in the response (Y) when the concentration levels of measurands X_1 and X_2 are held constant. Into this thesis, the multiple linear regressor modelling was explored for modelling three high range models; HR April28, HR June 10 and HR Oct29 and two standard models; UOP991 June9 and UOP991 Oct29. The regressors were the three measurands; Fluoride (F), Chloride (Cl) and Bromide (Br). The measuring equipment responses (i.e. the peak areas) were used as the response ($Y_{i's}$) and the regressors were the calibration standards of Fluoride, Chloride and Bromide. The multiple regression equation is;

$$Y = \beta_o + \beta_1 F + \beta_2 Cl + \beta_3 Br + \varepsilon$$

Since this thesis is dealing with regression models that have small sample sizes, using multiple linear regression will help in overcoming the issue of dealing with small sample sizes, improving normality assumption, reducing the error of the regression plane and stabilizing the estimation of the variance.

2.4 Note Regarding The Term Blank:

A blank theoretically consists of all chemicals in the sample except the measurand and run through the entire analytical procedure in the same way as the analyzed sample. Using Blanks has mainly four purposes;

1. Cleaning the analytical measuring system between samples consecutive runs i.e. flushing the system and this will eliminate cross contamination between samples analyzed.
2. Diluting the samples that contain the measurand with concentrations levels of falling outside the working range of the analytical method.
3. Blank Correction/Subtraction. The obtained measuring equipment response arising from measuring the blank sample will be subtracted from the observed response coming from the measurand. This will help in identifying the net measurand concentration in the sample. Ideally, the blank matrix is preferred to match the sample matrix; but in real life, this is almost impossible.

   In case of indirect reference measurements which hare based on empirical calibration function / regression function obtained experimentally and typically based on linear model shown in equation 2-1;

$$y = a + bx + \varepsilon \qquad (2\text{-}2)$$

The intercept a corresponds to the experimental *blank* –i.e. the observed measuring equipment signal coming from the blank- and the slope b represents the method sensitivity. When blank subtraction is conducted, the calibration function/regression function will look like $y = bx + \varepsilon$

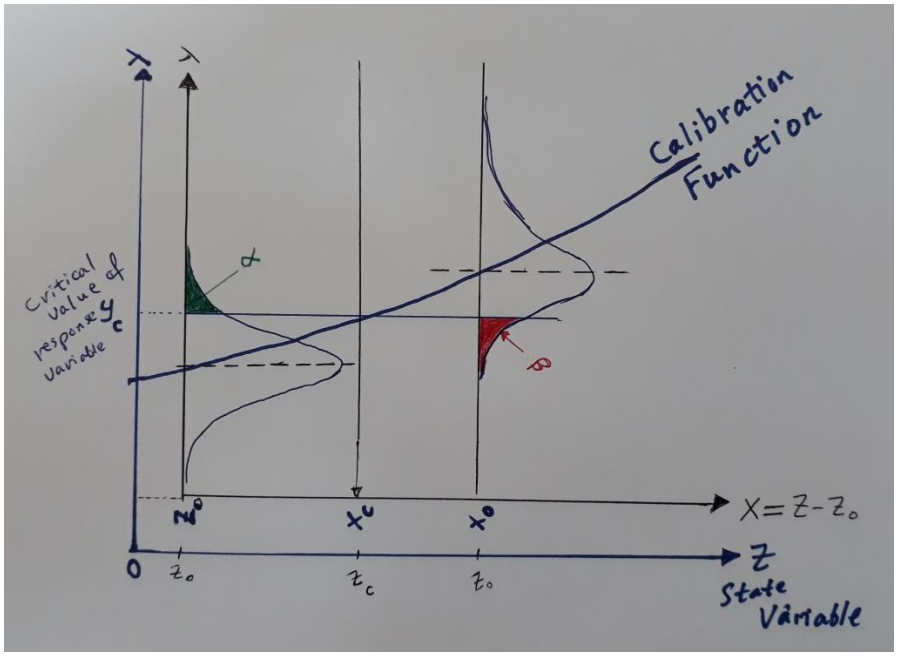


Figure 2.1. Illustration of Critical Values, Regression Curve and Probability for Normally Distributed Data

Limits that characterize the capability of detection of analytical testing procedures are mainly the lower, upper and thresholds. These limits help in distinguishing between real measurement values and zero values (when no measurand is present in the sample or zero value of measurand in case of measuring blank sample). When we present values, which are near to the limit of detection, these values are called critical, the term *critical value* standards for the guarantee of presence of measurand in the

sample. These critical values y_c (measuring equipment response variable) and x_c (independent variable which is the amount of measurand/regressor) shown in figure 2.1 are the fundamental measures to characterize the method of detection and quantification. If the response variable value y_i obtained from the measuring equipment for the measured sample is greater than critical value of the response variable ($y_i \geq y_c$), then this means the measurand has been detected. The corresponding measurand critical value x_c is usually estimated from the built calibration/regression function.

The critical value y_c represents the smallest measured concentration level that can be distinguished from the basic state (i.e. the blank response y_{BLK}) with a given level of significance $P = 1 - \alpha$. Simply, y_c is the measuring equipment observed response that is mainly due to the presence of measurand (i.e. the independent variable x_c) that exists in the sample. Definitions of symbols shown in in figure 2.1 are:

$\boldsymbol{y_c}$**:** Critical value of response variable which is the lowest signal which can significantly be distinguished from that of blank sample.

$\boldsymbol{y}$: Response value, the exceeding of which leads – at given error probability α- to the decision that the observed system is not in its basic state – by basic state it is meant measurand does not exist in the sample-. Note that the lower detection limit is the lowest measurand concentration which produces a signal (response y) that is greater than the critical response value $y \geq y_c$.

$\boldsymbol{x_c}$: Critical value of the net state variable which is directly corresponding to the y_c. The value of x_c is allocated to the critical response value obtained from the built calibration/regression function.

$\boldsymbol{x_o}$: the value of the measurand in the blank which is supposed to be zero.

$\boldsymbol{X}$: is the measurnad concentration.

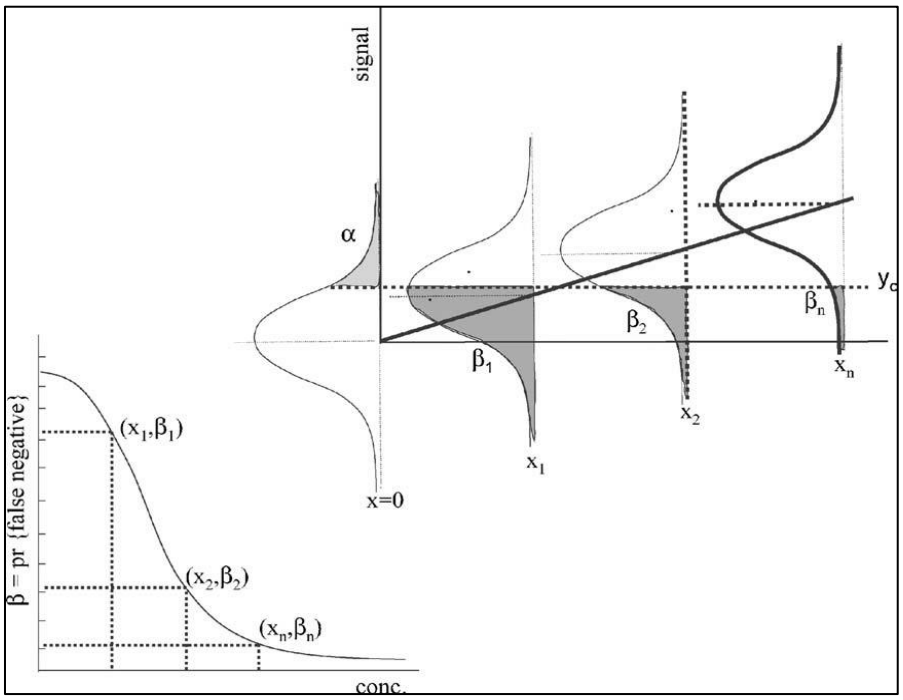


Figure 2.2. At given value of *α*, this is how the probability of false negative β decreases as the concentration (conc.) of the measurand increase. x_n is n number of calibration standards/regressors, x_o means the concentration level of the regressor is zero (i.e. Blank) and y_c is the lowest response signal which can be distinguished from the signal coming from the blank.

For a monotonic case and when the relationship between the response variable – measuring equipment output y - versus the concentration level (conc.) of the calibration standards /regressors; $x_o, x_1, x_2, \ldots x_n$, -at given probability of α (false

positive), the probability of erroneously not detecting the measurand – i.e. false negative β - degreases as illustrated in figure 2.2. Here we can appreciate why it is preferred to calibrate the measuring equipment according to the proposed High Range (HR) method. So working with HR method is preferred over UOP991 since it will minimize the error.

4. Determining the method limit of detection (LOD) and method limit of quantification (LOQ). Recall that the regression line is described by the function $y = a + bx + \varepsilon$. The intercept of the regression line is given the letter a.

   It is advisable to use LODs in measurand concentration units rather than equipment signal units. So from the calibration curve equation, LOD for particular measurand Z will be:

$$Z_{LOD} = \frac{Y_{LOD} - a}{b} \qquad (2\text{-}3)$$

   Wher; a is the intercept and b is the slope of the calibration curve. If there is no intercept, then LOD= $3.3\frac{s}{b}$. where b is calibration curve slope and s is the blank readings standard deviation. Limit of detection in chemical analysis is not an easy subject to tackle, a recently published monograph that is giving detailed discussion of this topic is the one written by Edward [44].

   Once the LOD is identified, the Limit of Quantification (LOQ). Which is the lowest concentration of measurand that can be determined quantitatively with

an acceptable precision and accuracy under the stated experimental conditions. The simplest way of determining the LOQ is

$$Y_{LOQ} = Y_{blank} + 10s_{blank} \quad (2\text{-}4)$$

The LOQ could also be determined from the calibration curve function as

$$Y_{LOQ} = 10\left(\frac{s}{b}\right) \quad (2\text{-}5)$$

Where: s is the blank standard deviation and b is the slope of the calibration curve.

The relationships between limit of detection for the blank (LOB), LOD and LOQ is illustrated in figure 2.1

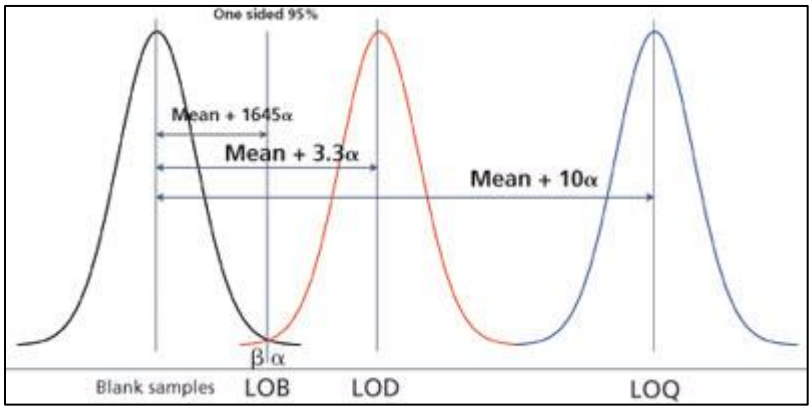


Figure 2.3: Blank samples, their mean value and how it is related to LOD and LOQ

2.5 Outliers.

By definition, an outlier is a data point that deviates markedly from others. When the number of available data points is very large and numbers of outliers are limited, then removing these outliers' observations is not a risky job. But in case of this study, the numbers of observations are limited and some observations which are potential to be outliers are not necessarily an outlier. Recall that this study is dealing with liquid organic volatile samples that do evaporate at room temperature. So sample stability was one of the challenges that the data gathering part of this work was consistently facing. Moreover, each obtained data point of this work was consuming a lot of time and laborious work, so considering suspicious observation as an outlier is not favored and removing a lot of suspicious outlier observations could end up with having biased data.

Two international standards were addressing the subject of detecting and dealing with outliers observations and they were used as a guide for the data treatment of this study, these ASTM and ISO standards are [45-46]. During this work, more than one test of outliers' detection was used for inspecting the same data set. This will help in overcoming the limitation of each test. Some of the very basic fundamental outliers tests such as Grubbs [47-48]. A paper written by Hereman [49] is summarizing statistical tests that are used for defining, identifying and handling outliers.

2.6 Normality Tests

Normality assumption is very fundamental requirement for the statistical assessment done on this research work. Typically, the larger the sample size, the better is the performance of the normality tests. But according to the ISO standard [50] sample size less than eight could be sufficient for normality tests. For this thesis work, normality test was checked by residual plots and this is illustrated on the residual plots shown in chapter four.

2.7 Multiple Comparisons Procedures

ANOVA does not tell us which means are not equal, it only tells us that some significant difference exists. On the context of One-Way ANOVA, multiple comparisons among means were conducted according to Tukey's, Fisher's and Dennett's procedures. The overall error rate was set to 0.05. These pairwise comparisons tests were done by using the statistical package Minitab17.

- **Tukey's Procedure:** This method was invented by Tukey (1953). This method tests the hypothesis for which the overall significance level is exactly "α" when the sample sizes are equal (i.e. balanced design). Tukey's procedure relies on the assumption that the samples means are independent random samples, each

containing an equal number of observations. For an unequal sample size, a modified Tukey method called "Tukey-Kramer" is used. Tukey's-Kramer method is considered as a conservative multiple comparison procedure, it fixes the overall family error rate "α", but will allow for rejecting more tests. This thesis is considering the Tukey's method for completed randomized, one-factor design.

- **Fisher's Least Significant Difference (LSD) Procedure:** This procedures was developed in 1935 and is comparing all pairs of means wile controlling the error rate α for each individual pairwise comparison but does not control the experimentwise (i.e. family) error rate.

- **Dunnett's Procedure:** This procedure was developed by Charles Dunnett (1964). It compares each treatment mean with one selected treatment called "Control".

2.8 Box-Cox Transformation

Predicting the outcome from a predictor variable is typically done at chemical laboratories by using linear modeling (i.e. $y = \beta_o + \beta_i X_i + \varepsilon$). But in some cases, the relationship between the response and predictor variable is not linear; such the case of fluoride which the original testing procedure UOP991 is recommending to use the quadratic function for building the regression model ($y = \beta_o + \beta_i^2 + \varepsilon$) . In 1964, Box and Cox proposed a transformation parameter λ, so the transformed peak responses can be expressed mathematically as; $\grave{y} = (\frac{y^{\lambda}-1}{\lambda})$. In many instances, data transformations methods are selected empirically. Therefore, this thesis work selected to conduct data transformation on the fluoride (F) peak responses obtained from the measuring equipment. The Box-Cox technique was used where the power of transformation λ was chosen to be 0.5, so the transformed peak area $\grave{y}$ is defined as $\grave{y}=y^{\lambda}$ where λ=0.5. The Box-Cox transformation requires the data to have a positive responses – which is the case for all measured peak areas responses- and sample size of $(10\sim20)$ are considered sufficient for estimating the power of transformation λ which is selected to be λ=0.5. The Box-Cox transformation is used for correcting these two conditions; the process data are not normally distributed and subgroup variance is unstable due to the variation in the data is proportional to the subgroup.

2.9 P-Value

The smallest significant level at which the H_o would be rejected is the p-value. It is used extensively in the decision rule of the hypothesis testing. But the P-Value concept should be used with cautious since it is biased towards sample size. As the sample size increases, the p-value will decrease, figure 2.2 illustrates this relationship, so P-Value is biased toward large sample size [51]. On the year 2016, the American Statistical Association issued a policy statement on p-values that worth reading [52]. Basic statistical calculations that are used into this thesis as part of data treatments are summarized in this ASTM standard [53].

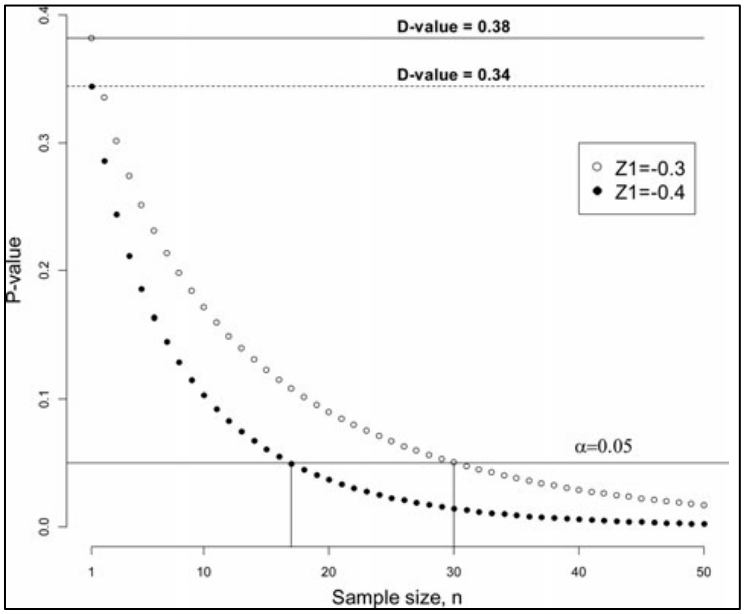


Figure 2.4: Graph showing how the p-value decreases as the sample size increases.

CHAPTER 3

EXPERIMENTAL WORK

The term chromatography is the general name for a wide range of physicochemical separation processes in which the components to be separated are distributed between a stationary phase (which is the separation column) and a mobile phase (which is the *eluent* : the term eluent is synonymous with mobile phase. Eluent is the solvent that flows through the separation column in the measuring system). The term ion is referring to the oxidation state of the material to be measured. IC technique is a combination of chromatography and ion exchange which was developed around 1975 and applied to the determination of ions. This chapter is devoted to the analytical setup; measuring equipment and calibration standards/ regressors used for building the calibration models.

3.1 Operating Principle of Combustion Ion Chromatograph (CIC)

It consists of four units as shown in figures 3.1 and 3.2. Unit numbered one is the auto-sampler, where all samples vials will be loaded. Unit numbered two is the combustion part, where the sample will be burned at temperature around 1050^{o}C. Unit numbered three is the solutions handling module which will trap all combustion products in aqueous liquids then forward them to module number four which is the Ion Chromatograph (IC) where separation will occur via the IC separation column followed by detection and quantification by the conductivity detector. Concise introduction about Ion Chromatograh (IC) is covered by this introductory level monograph [54].

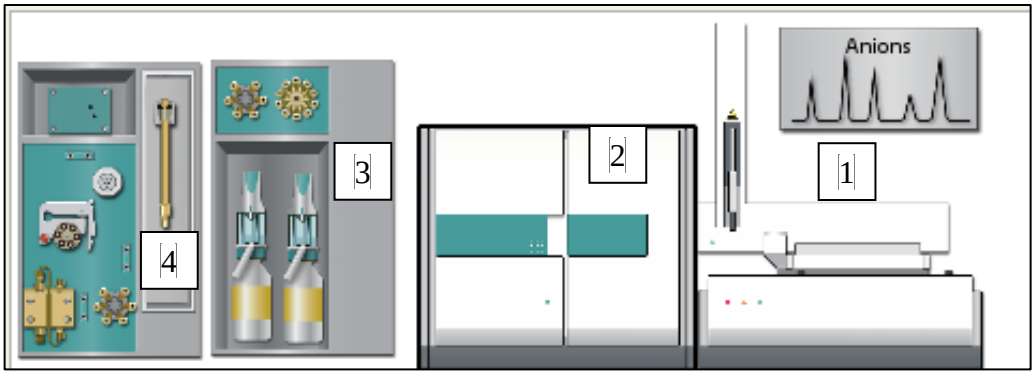


Figure 3.1. Main components of CIC

3.2 Instrumentation

The exact CIC measuring system used for generating data for this thesis work is shown in figure 3.2. This high-tech lab analytical measurement system is used at the refinery laboratory to monitor the concentration of Chloride in the refinery feed and produced Naphtha. Chloride is very corrosive chemical, so its presence and concentration determination/monitoring is a common test in petroleum refineries.

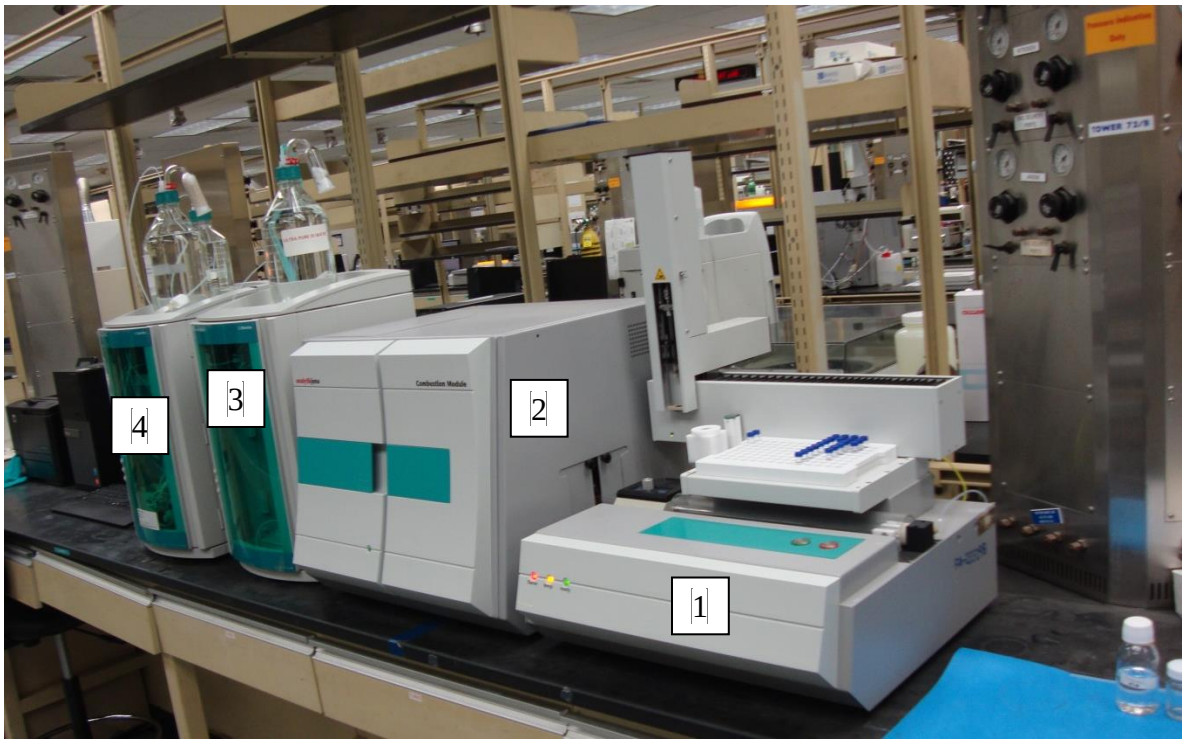


Figure 3.2. The CIC Measuring Equipment. This is Metrohm CIC. It consists of 930 Compact IC Flex, 920 Absorber Module and Analytikajena Combustion Module.

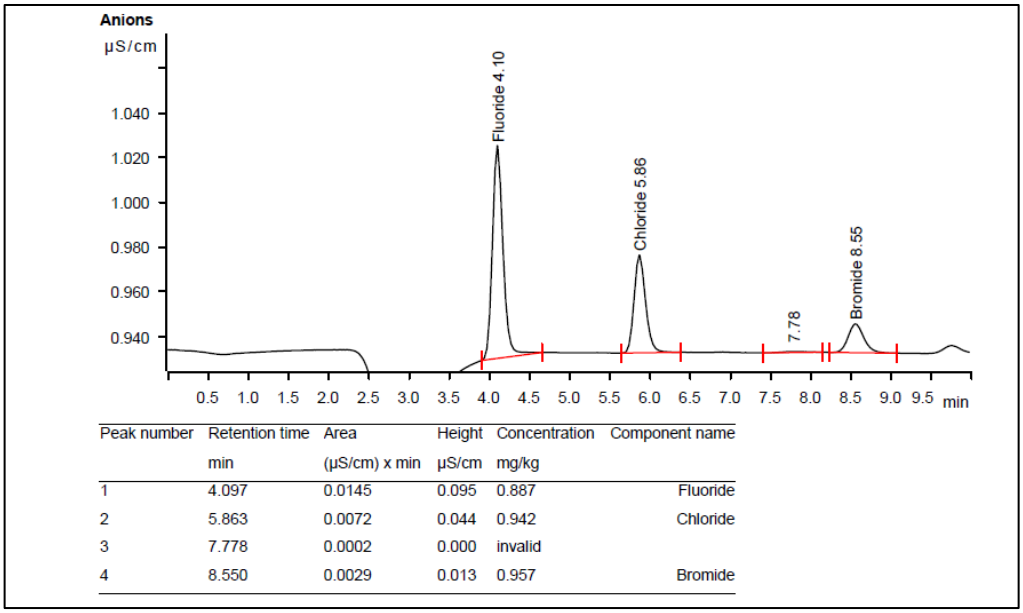
3.2.1 Mobile Phase

Mixture of sodium carbonate Na_2CO_3 (about 0.6814-g) and sodium bicarbonate $NaHCO_3$ (about 0.168-g) were dissolved into 2-Liter of deionized water. The mobile phase works a carrier for the sample once the sample is

injected into the IC , passes through the separation column then being detected by the detector. This mobile phase in the context of IC is called eluent.

3.2.2 Separation Column

This micro-column is packed with ion exchange resin where the injected sample will interact with this stationary phase (i.e. the ion exchange resin). Based on the affinity between the nature of the chemicals (F,Cl and Br) and the nature of this mobile phase, some measurands will show interest to stay longer inside the separation column, hence this is called retention time. Figure 3.3 shows the ion chromatogram for the three separated measurands (F,Cl and Br).



Peak number	Retention time min	Area (µS/cm) x min	Height µS/cm	Concentration mg/kg	Component name
1	4.097	0.0145	0.095	0.887	Fluoride
2	5.863	0.0072	0.044	0.942	Chloride
3	7.778	0.0002	0.000	invalid	
4	8.550	0.0029	0.013	0.957	Bromide

Figure 3.3. This Ion Chromatogram shows that Fluoride will be eluted first (about 4.097minutes), Chloride (about 5.863 minutes) and bromide will be eluted lastly about 8.550-minutes.

3.2.3 Detection

While separation is the central issue in chromatography, detection of the separated species is of comparable importance. Different types of detectors are used with IC. The universal and most frequently used IC detector is the conductivity detector (CD) where the unit of detector response (response is the magnitude of change of signal per unit change in concentration of measurand) is expressed in µS/cm.

3.3 Calibration Standards (Regressors) Preparation and Measuring

The calibration method for both the standard UOP991 (i.e. the low range original method) and the modified High Range method (HR) was done according to the calculation formulas given in UOP991 [1]. Below is an illustrative example done for the UOP991 method?

3.3.1 Calibration Standards/Reagents Calculated Values

Note that each time calibration to be conducted, the following reagents/calibration standards will be prepared. The illustrative calculations shown below were the ones done on April 25, 2016 for calibrating the standard UOP991 method.

3.3.2 Preparation of the Mobile Phase

0.6803-g of Na_2CO_3 (Panreac#141648.1210/Lot#122500), 0.1702-g of $NaHCO_3$ (Panreac#141638.1211/Lot#469932) both dissolved into two liters of type-I deionized water.

3.3.3 MSN Regenerant Reagent Preparation

The role of this chemical is just to regenerate the strength of the stationary phase of the separation column. It was prepared by dissolving about 5.5-mL of concentrated sulfuric acid H_2SO_4 (Sigma-Aldrich#30743/Lot#SZBB3070V) into one liter of type-I deionized water.

3.3.4 Preparation of the Standards Stock Solutions

3.4.4.1 Fluoride Standard Stock Solution [F]=1034.444-mg/kg:

Starting from 4-Fluorobenzoic Acid (Aldrich#418846/Lot#MKBJ6402V, purity 99%), 0.2015-g added to 13.0110-g of methanol (Sigma-Aldrich#32213/Lot#SZBA2950V) followed by Toluene (13.0141-g) Scharlau#TO00792500/Lot#0714Batch15074405.

3.4.4.2 Chloride Standard Stock Solution [Cl]=1071.33737-mg/kg:

0.1057-g was taken from 2,4,6-Trichlorophenol, purity 98% (Aldrich#T55301/Lot#MKBG239V) and dissolved in 52.0094-g of Toluene.

3.4.4.3 Bromide Standard Stock Solution [Br]=1067.93358 mg/kg:

0.1055-g of 4-Bromoacetanilide (Aldrich#16165-g/Lot#S79241-349, purity 98%) was dissolved in 36.0058-g of Toluene.

3.3.5 Preparation of the Calibration Mixture (CAL MIX) Standard

For the measurand Fluoride standard, 0.2061-g taken from Fluoride stock standard, 0.2051-g from the Chloride Stock Standard and 0.2074-g from the Bromide stock standard were diluted into 20.0662-g of Toluene. This will make the concentrations of Fluoride, Chloride and Bromide CAL MIX as follows: [F]=10.30704 mg/kg, [Cl]=10.62284 mg/kg and [Br]=10.70783 mg/kg.

3.4.5.1 Preparation of the Calibration Working Standards (Regressors). 0.2061-g taken from Fluoride stock standard, 0.2051-g from the Chloride Stock Standard and 0.2074-g from the Bromide stock standard were diluted into 20.0662-g of Toluene. This will make the concentrations of Fluoride, Chloride and Bromide CAL MIX as follows: [F]=10.30704 mg/kg, [Cl]=10.62284 mg/kg and [Br]=10.70783 mg/kg.

3.4.5.2 Preparation of the Calibration Working Standards (Regressors)

Starting from the CAL MIX standard, a series of calibration working standards were prepared. Toluene was the solvent used to dilute in order to prepare each of these calibration standard whose concentration values in (mg/kg) are shown in the table 3.1.

Table 3.1: Prepared Calibration Standards for Fluoride, Chloride and Bromide

Standard No. (Regressor No.)	**Fluoride (F)**	**Chloride (Cl)**	**Bromide (Br)**
1	0.10346	0.10663	0.10748
2	0.25804	0.26594	0.26807
3	0.49641	0.51162	0.51571
4	0.93796	0.96670	0.97443

3.4 Chromatograms of Measured Calibration Standards/Regressors

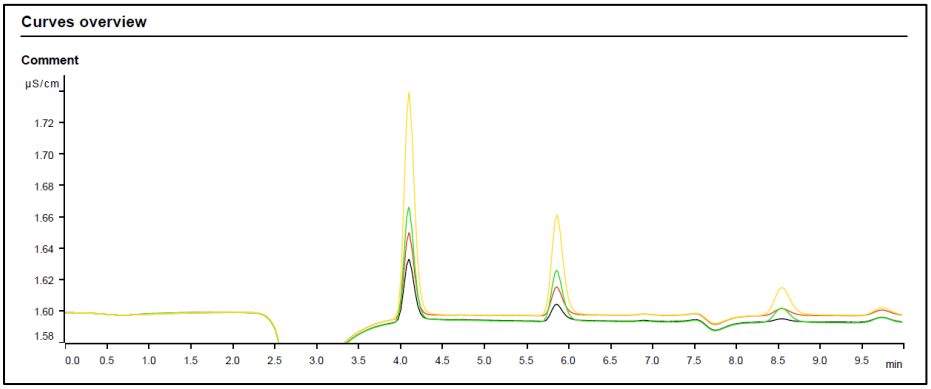


Figure 3.4. Calibration Working Standards Ion Chromatogram Overlay. This graph shows no shift in the retention time for each element (F,Cl & Br).

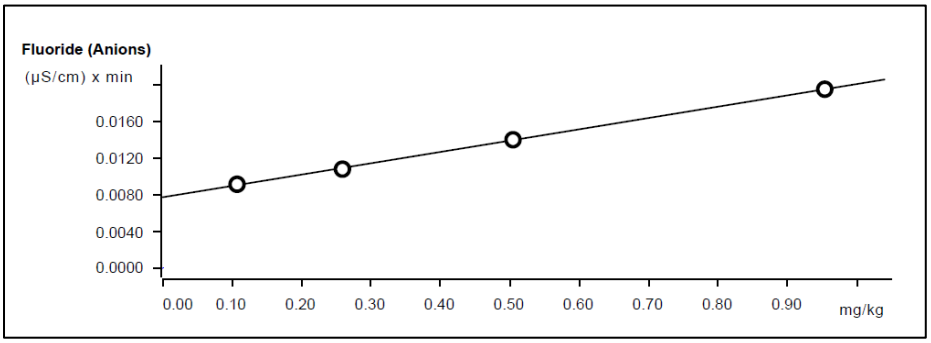


Figure 3.5. Calibration Curve for the Fluoride Calibration Standards.

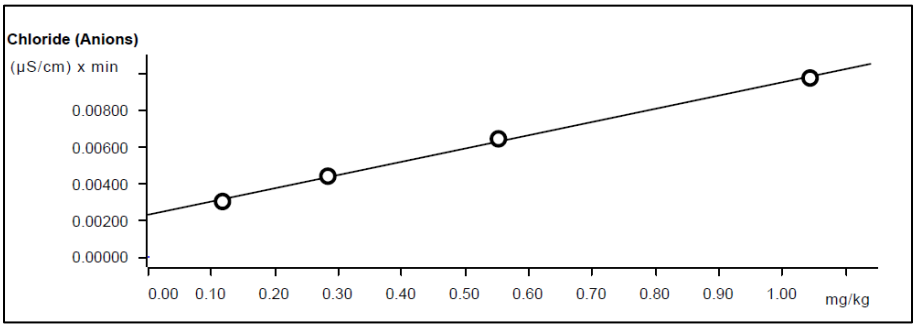


Figure 3.6. Calibration Curve for the Chloride Calibration Standards.

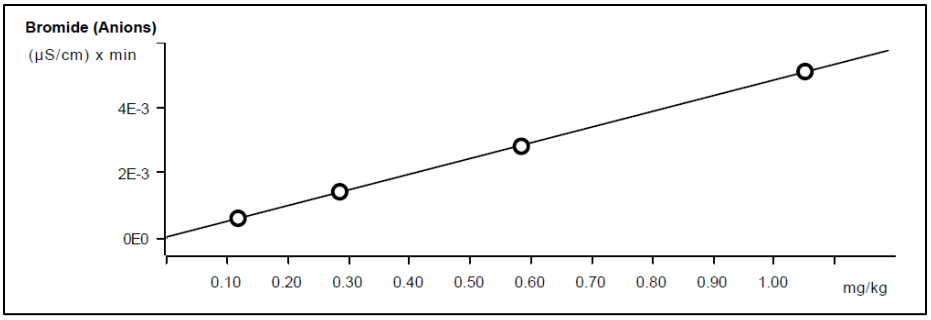


Figure 3.7. Calibration Curve for Bromide Calibration Standards.

3.5 Why the Calibration Curves in IC are Non-Linear?

Two factors contribute for having nonlinear calibration curve in Ion chromatography. The first is an instrumental based reason and has to do with the technique of IC with suppressor column while the second one has to do with the chemical nature of the chemical specie itself (the measurand) and this is called chemical kinetics of the measurand. For the IC part, If chemical suppression is used, slightly curved calibration curves are usually obtained. In such cases, the calibration values must be located as close as possible to the contents of the sample (this calibration technique is called *bracketing*) in order to obtain the most accurate results. Bracketing technique requires more work in identifying the region of curvature in the regression curve; consequently, more calibration standards/regressors are needed. While for the chemical kinetics part, For anion (F^-) of weak acid like Hydrofluoric Acid (HF), the IC suppressor will convert all fluoride ions (F^-) back to HF. When this acid is diluted with water, the dissociation of this weak acid HF to give F^- ions is very limited (recall that fluoride element has the greatest electron affinity among all other elements). In fact, as the concentration of the HF increases, the amount of released F^- decreases, which explains the nonlinearity at high concentration of HF.

This thesis work is proposing modeling the Fluoride responses by using linear regression rather than the quadratic one which is recommended by the original method UOP991 and the motivation behind this change in modelling are

mainly; the simplicity of estimating the method LOD and LOQ, better forecasting of measurand concentration level by just extrapolating the regression line and the slope of the linear line can easily be quantified and recall that the method sensitivity can easily estimated from the slope value. Moreover, the estimated regression line standard deviation can be used directly as one of the contributors of the overall test result global uncertainty.

3.6 Economical Design of Calibration Protocol

Designing an experiment is a quite big subject and this thesis work is not about redesigning the whole testing procedure; instead, it is about improving the performance of a designed one (i.e. UOP991) and approving the acceptance of the proposed designed one. So this thesis is proposing an *economically design* for calibrating measuring equipment CIC and assuring its performance.

3.7 Peak Area or Peak Height

The connection between the calibration working standards (i.e. the regressors) and the measuring equipment responses is called calibration and it is described mathematically by calibration functions. These calibration functions could be linear,

quadratic, polynomial…etc. In chromatographic measuring methods, the equipment responses can be quantified and represented by the measured Peak Area (PA) or Peak Height (PH). Typically, peak area is the preferred way of relating the measured calibration standard concentration level to the measuring equipment response; reasons behind preferring peak area over peak height are discussed into Leonid's paper [55]. This thesis work checked the measuring equipment responses with respect to both peak area (PA) and peak height (PH) and it turned to be results obtained by considering the peak area are more accurate. Therefore, the whole study was considering only the peak area.

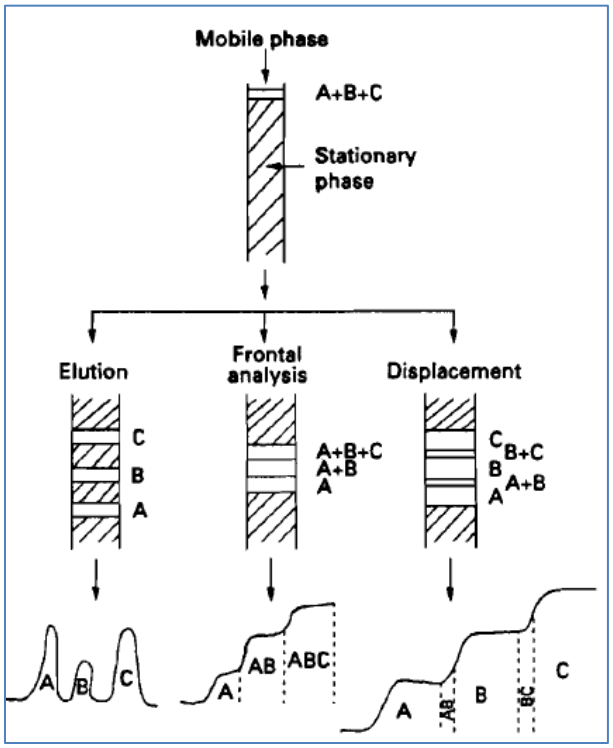


Figure 3.8. Segregation of the three Measurands (A=Fluoride ,B=Chloride and C=Chloride) that sample/regressor contains by Chromatographic method and peaks formation (i.e. responses).

For quantitative analysis, if it necessary to establish a relationship between the magnitude of the detector signal (i.e. the measuring equipment response) and the

measurand amount/regressor concentration level x. In chromatographic method, the response is quantified numerically as either the signal peak area (PA) or signal peak height (PH). Both PA and PH are proportional to the to the measurand/regressor concentration level, so any of them can be utilized for quantification after calibration/regression. The accuracy of the quantification is affected by the quality of the measured peak area or peak height. Peak distortions such as; peak skewness and kurtosis will affect the quality of measure responses in case peak height is used. Which is not the case if peak are is used to represent the measuring equipment responses. Figure 3.9 shows a Gaussian peak where the peak height is denoting the maximum response observed from the measurand "C_{max}".

On the other hand, the quality of the measured peak area can be adversely affected if there is a poor peaks resolutions (i.e. two peaks are partially overlapping) and this is a case where difficulties of accurately defining the peak boundaries is faced.

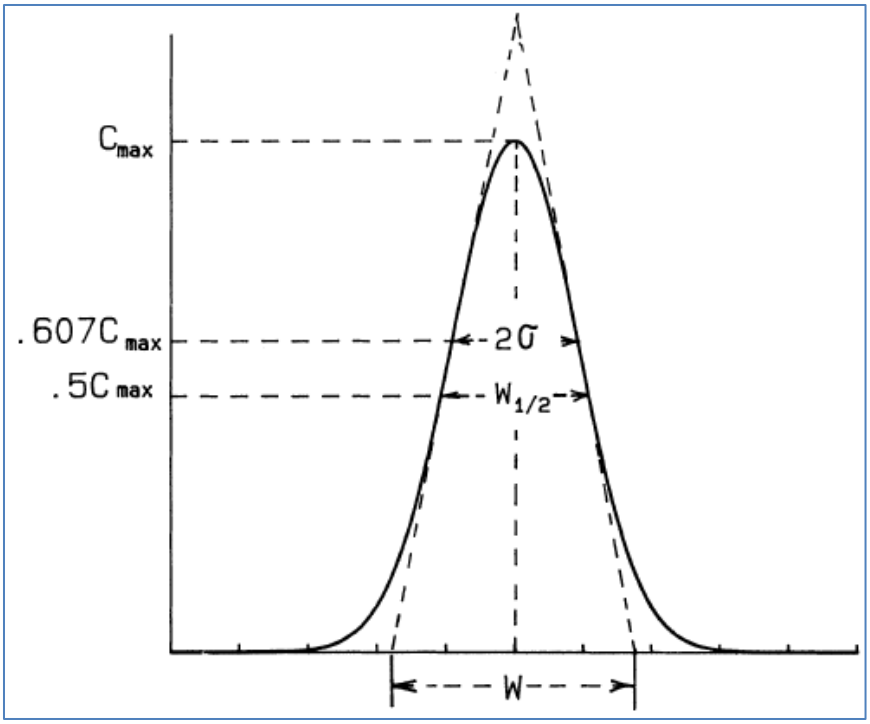


Figure 3.9. In case of peak height (PH), the measuring equipment response y is directly proportional to the PH which is given by the symbol C_{max} . While in case of peak area (PA), the measuring equipment software will integrate the entire area below the peak.

3.8 How Often Do We Need to Calibrate?

The original SOP UOP991 was stating clearly to calibrate with each sample, yet this thesis proved experimentally that for both the standard (low range) UOP991 and the modified (High Range) HR, calibrating the measuring equipment once per a month is sufficient.

3.9 How Many Points Needed for Calibration?

The more number of calibration data points, the better is the calibration, but the more costly will be. Two points calibration could be sufficient provided that the relationship between the measuring equipment response and the measuring samples/standards are linear. To check for this linearity, two points linear interpolation technique could be used. ISO standard [56] is recommending at least five calibration data points each with two replicates. This is valid only when there is linearity in the calibration curve.

3.10 Shall Blank Be Part of Calibration Curve Data Points?

There are several purposes behind using blank. The most commons ones are;

3.10.1 Blank Correction

To check for purity of the used chemical reagents that were involved in calibration working standards preparation (so ultimately, blank subtractions will be done for all measured calibration standards data points).

3.10.2 Memory Effect

Typically blank is run between two consecutive samples/calibration standards in order to assure no residuals (i.e. carryover) from the previous sample in the measuring equipment. Otherwise, the measuring equipment might show continuous increase in the response regardless of the actual concentration of the measurand in the sample/standard.

3.10.3 Limit of Detection

Recall that one of the classical approaches for determining the LOD of a testing method is by using the standard deviation of blank runs.

> Note that blank is not used as one of the calibration stanrdard (regressors). The regression parameters do not include β_o which is the y-axis intercept of the regression line.

3.11 Running Sequence of Calibration Standards/Regressors

Measuring the calibration standards starting from the calibrant with concentration level firs, and going in ascending order is highly recommended. Particularly, when there is a possibility of losing measurand due to its low stability or cross contamination.

3.12 Obtaining Matrix-Matched Calibration Standards/Regressors

In real life, this is possible in limited cases, most of the time the matrix of the prepared calibration standard is not the same as the matrix of the measured samples. Some of the commonly used methods to overcome the matrix-matched difficulty are; standard addition calibration method and internal standard calibration method. This thesis work was carried by using the external calibration method which is a matrix-free method.

CHAPTER 4

CAPABILITY OF DETECTION

In manufacturing, the term process capability is widely used. But in the field of laboratory test methods development and validation, the equivalent term *capability of detection* is the dominant one. This chapter is addressing standardized procedures for conducting calibration and quantifying the lower limit of detection. One of the ways for validating the validity of the modified method "HR" is by comparing the critical values of detection versus a given known value as detailed into this chapter. One of the six developed HR calibration model was used as example.

4.1 Introduction

In order to characterize a measurement process, a laboratory or measurement testing procedure, the minimum detectable value can be stated if appropriate data are available for each relevant level. Different measurement methods have different minimum detection limit. The minimum detectable value of the measuring method may be used as a factor to be considered in selecting the measurement process. The simple definition of detection limit which is often used for calculating the instrumental minimum detection limit is defined as the concentration of substance (*measurand*) which produces a signal (equipment *response*) three times greater than the standard deviation of noise for a blank.

This definition helps in comparing the performance of different instrumental methods. This approach for quantifying detection limits is focusing mainly on the probability for false positive i.e. the likelihood of obtaining a positive response for a measurand which is truly does not exist in the sample. The probability of detection is called true positive, when the instrumental response is purely due to the existence of measurand in the sample. One well known detection-limit definition that considers both the *true positive probability* (β) – also called false negative probability- and the *false positive probability* (α) is the calibration-based method developed by Hubaux and Vos.

The capability of detection based on the linear calibration is well-defined in the ISO standard 11843 and it is done based on the Neyman-Pearson test that considers both types of errors; false positive (α) and false negative (β). Consequently, the minimum detectable net concentration is commonly named *CCβ*. M.C. Ortiz illustrated clearly in his paper [58] –figure 1- how the probability of detecting false negative *CCα* decreases as the measurnad concentration increases (response vs. predictor for monotonic case).

Two influential factors contribute to the capability of detection, these factors are mainly; the chemistry involved in the testing procedures and the instrumental capability of the measuring equipment, together; testing procedure and measuring equipment make the overall Chemical Measuring Systems (CMS).

4.2 Chemical Measuring Systems (CMS)

CMS are used for estimating qualitatively and quantitatively the chemical substance. In analytical chemistry, the specie under the study is called *analyt* but its standardized equivalent term is *measurand* and denoted by the uppercase X. The measurand's numerical value is denoted by the lowercase x. Note that measuring system is mainly means the measuring equipment such as the lab analyzer, but the meaning of measuring system could also be extended to cover also the whole analytical setup which consisting of ; the lab measuring equipment (analyzer), the testing methodologies i.e. lab testing method "SOP" and the operators too, so the whole analytical process consists of; man, materials and machine. Note that term method used into this thesis can be considered synonymous with the term measuring equipment.

The measuring equipment used in generating these real life data are obtained from a laboratory measuring equipment called Combustion-Ion-Chromatograph "CIC" and the chemistry of the testing procedure is based on the laboratory standard operating procedure (SOP) called UOP991 [1], CIC and UOP991 together makeup the chemical measuring system used by this thesis work.

The CMS input quantities were the materials with known concentration levels of the measurands (Fluoride and Chloride). These materials of known concentration levels of measurands are called calibration standards (*calibrants*/*calibrators*) and in the context of regression analysis are called *regressors*. Five calibration standards

(X_1,X_2,…X_5) were used for calibrating the CIC. These calibration standards supposed to have measurands concentration levels spanning the range from 0.1-ppm to 1.0-ppm according to the original SOP UOP991. While for the modified method, i.e. the High Range (HR), the calibration standards were starting from about 1-ppm to 10-ppm which is ten times higher in range than the original SOP UOP991. Hence, recall Ortiz [58] not that for nontonic case, the probability of detecting false negative decreases as the measurand concentration increases.

The output quantities of the CMS are the measuring equipment *responses* denoted by Y_i. so five calibration standards would give five responses ($y_1, y_2, \ldots . y_5$). The CMS response Y is a function of the input quantity X. There are some criteria to be considered in selecting the calibration standards (regressors). Such criteria like; matrix-matched between the calibration standards and the real sample, this is in addition to the stability of the calibration standards(X_n). By stability we mean that the calibration standard supposed to be time independent i.e. $X \neq x(t)$. In real life, this condition is sometimes not possible, particularly for this thesis work where the measuring equipment calibration standards were prepared in chemical volatile solvent called Toluene. This chemical evaporates quickly with time at room temperature. So it is very expected that there will be variations between the calculated values of the prepared calibration standards (regressors) versus the measured ones. This in addition to the built-in variations in the measuring equipment response values $y_{i\prime s}$.

The output of any CMS is generally characterized by a probability density function (PDF) which is described by two population parameters; a location parameter also called a centrality parameter and given the symbol δ and the scale parameter which is the standard deviation represented by the symbol σ. Both of these population parameters δ and σ are directly involved in the calculation of the minimum detectable measurand X_d which is defined by ISO11843-2 [56] as:

$$X_d = \delta * \frac{\hat{\sigma}}{\hat{b}} * \sqrt{\frac{1}{k} + \frac{1}{I*J} + \frac{(\bar{X})^2}{S_{xx}}} \quad (4\text{-}1)$$

Where;

$\delta(v; \alpha: \beta)$: is the non-centrality parameter, its value is defined in such a way that a random variable following the non-central t-distribution with $v = I * J - 2$ degrees of freedom and the non-centrality parameter δ,$T(v, \delta)$ satisfies the equation: $P[T(v, \delta) \leq t_{1-\alpha}(v)] = \beta$ where $t_{1-\alpha}(v)$ is the (1-α) quantile of the t-distribution with v degrees of freedom.

$\hat{\sigma}$: is the estimate of residual standard deviation of the calibration curve (regression line). Practically, it is the standard deviation of all the obtained responses of the set of calibration standards.

$\hat{b}$: is the estimate of the functional relation curve (i.e. regression line) slope. The physical interpretation of the slope is that it shows the sensitivity of the established calibration curve. It is a common practice to design the calibration curve to be linear.

k: is the number of measurements on the actual test sample. For this thesis work was considered 1.

I: is the number of calibration standards (i.e regressors) including the blank standard. For this thesis work I=6.

J: is the number of replications of measurements on the calibration standards including the blank. Due to the high volatility of Toluene, no replicates work were done, so $J = 1$.

The functional relationship between the measuring system response R and the measurand X is $R(X)$ and estimated by $r(x)$. This functional relationship is required to be monotonic, which means that the response values do not have regions with either zero slope or reversals slope. Mathematically this functional relationship is described as:

$$r(x) = a + bx + \varepsilon \qquad (4\text{-}2)$$

Where;

a: is the intercept of the calibration curve and estimated by $\hat{a}$.

b: is the slope of the calibration curve and estimated by $\hat{b}$.

ε: is the error term which supposed to ~NID with $(0, \sigma^2)$.

Note that the intercept of the calibration curve (regression line) is directly contributing to the measuring equipment critical value of response variable y_c. The critical value of the response variable y_c is the response variable value exceeding of which will lead, for a given error probability α to the decision that the amount of measurand in the analyzed material is greater than the amount of measurnad in the blank. According to ISO11843-2 , it can be estimated by this formula:

$$y_c = \hat{a} + t_{0.95,v} * \hat{\sigma} * \sqrt{\frac{1}{k} + \frac{1}{I*J} + \frac{\bar{x}^2}{s_{xx}}} \quad (4\text{-}3)$$

Where

$t_{(0.95,v)}$ is the 95% quantile of the *t*-distribution with $v = I * J - 2$ degrees of freedom.

For computing the net state variable x_c which is corresponding to the critical response variable y_c , the formula is:

$$x_c = t_{0.95}(v)\frac{\hat{\sigma}}{\hat{b}}\sqrt{\frac{1}{K} + \frac{1}{I \cdot J} + \frac{\bar{x}^2}{s_{xx}}} \quad (4\text{-}4)$$

Where;

$t_{0.95}(v)$: is the 95% quintile of the t-distribution with $v = I \cdot J - 2$ degrees of freedom.

I: is the number of calibration standards.

J: is the number of replicates for each calibration standards which is considered to be one.

Detailed discussion about probability distribution types and their parameters estimation can be found in statistical textbooks; but I'd recommend the one written by Nick. Thomopoulos [57].

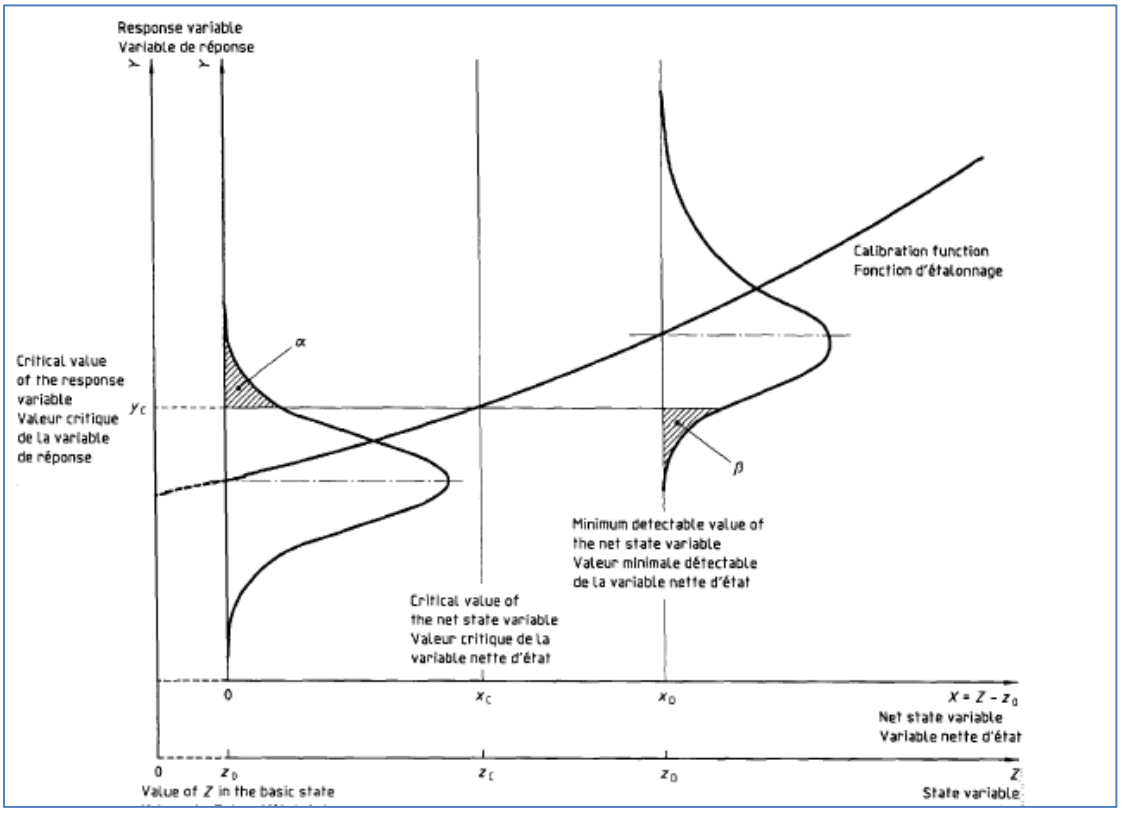


Figure 4.1. Relationship of Critical Response Variable (y_c), Critical Measurand Variable (x_c) and False Negative β And False Positive α.

4.3 ISO Standards Related to Detection Limits

International Standardization Organization (ISO) issued seven parts of ISO11843 dedicated to the subject of capability of detection [56, 59-64]. This thesis used ISO11843-1 (capability of detection; Terms and Definitions), ISO11843-2 (Methodology in the Linear Calibration Cases) and ISO11843-4 (Methodology for Comparing the Minimum Detectable Value with a Given Value) since they are relevant to this thesis work. The ISO11843-2 approach is describing the procedure for estimating the limit of detection based on the upper and lower prediction limits of the regression line as shown in figure 1 of ISO 11843-1.

4.4 Calibration

Calibration is an essential part of most analytical measurement systems. It requires a set of reference materials with known measurand concentration levels; these reference materials are called calibration standards (*calibrants*) and in the context of regression analysis are called *regressors*. These calibration standards ($x_{i's}$) will be measured/analyzerd by the measuring equipment and the corresponding equipment responses values will be obtained ($y_{i's}$). Graphically, the calibration function is a plot of the measuring equipment response variables ($y_{i's}$) and is usually represented by the ordinate (Y-axis) while the values of the used calibration standards ($x_{i's}$) are

plotted on the abscissa (X-axis). Calibration is a conceptual thing and cannot be determined experimentally. It is estimated through regression analysis [56]. Calibration function is describing the functional relationship between the expected value of the response variable (y) and the value of the state variable (x) . Note that it is assumed that the net state variable "also called measurand" is non-negative and the calibration function is monotonically increasing.

In graphical representations of a calibration function, the response variable is usually represented by the ordinate (y) and the net state variable (x) by the abscissa. Note that in instrumental chemical calibration, the built regression function is used in reverse manner i.e. the value of the response (y) is obtained experimentally, to predict its equivalent measurnad value(x). Therefore, the calibration is called an *inverse* calibration/ *inverse* regression.

4.5 Regression Analysis

The accuracy of test results obtained by most of instrumental chemical analysis is heavily depending on the quality of the built calibration model. This calibration model is a mathematical function that helps in estimating the value of a new unknown quantity "measurand" which is also called dependent variable by using the response –dependent variable- as an input data. For univariate model, the least squares method is used for constructing the calibration function. For studying the relationship between the response variable and the predicator variable, regression analysis is the statistical procedure used for this purpose. This thesis work is considering one single

case of regression/calibration; where a numerical response variable – call it measurand- and only one single numerical predictor variable – call it calibrant/regressor-. But this predicator variable was measured at different concentration levels. In instrumental chemical analysis, regression analysis is the most dominant statistical tool used for building the calibration model.

"The regression model can be useful for developing process knowledge through description of the variable relationship, in making predictions of future values, and in developing control methods for the process generating values of the variables". [65].

The validity of the calibration function depends on two conditions; the measurements from which the calibrating function was calculated are representative of the normal conditions under which the measurement system operates and the measurement system should be in a state of control.

When linearity is essential for the validity of the calibration function; which is the case most of the time, then bracketing technique should be considered. This bracketing technique is simply zooming in selecting regressors "calibrants" concentration levels that are covering the analytical range of interest. Although this bracketing technique is time consuming, yet it is widely used and preferred due to its greater accuracy than other approaches in calibration. When there is a departure in

the linearity of the calibration (curve/line), here it comes the importance of doing bracketing technique in which both the smallest and the largest (calibration standards/regressors or calibrants) concentration levels are selected in such a way that they cover the expected value of the measurand. By doing so, the predicted concentration level of the measurand becomes more accurate. With time, the established calibration function might need to be re-estimated, so the validity of the built calibration function can also be monitored by using control charts. Other practical tool for monitoring the stability of the built calibration function is via using a quality control audit sample which is the technique used in this thesis work and illustrated in chapter five.

4.6 Building of High Range Calibration Models

Six High Range (HR) calibration models were developed as an alternative to the standard low range UOP991 method. These methods are named based on the date at which the measuring equipment was calibrated in. These HR models are; HR April28, HR May17, HR June10, Aug21, HR Sep10, and HR Oct29. All calibration and validation works were done on the year 2016. The strategy followed on constructing these HR calibration models -figures 2 and 4- were based on running a series of calibration standards (independent variables $x_{i's}$), obtaining their respective measuring equipment responses (dependent variables $y_{i's}$). Responses obtained values -shown in table 1- were checked for potential presence of outliers by using multi-statistical tests such as; residuals analysis –figures 1 and 3-, Grubb's test and

Normality plot. The measuring responses with respect to peak areas were considered instead of peak heights. Below is a typical illustration about how each HR calibration model functions was constructed for two measurands; Fluoride (F) and Chloride (Cl), the same methodology was followed in constructing all the six HR calibration models shown in table number four. Tables 3 and 4 are showing the summary of the regression analysis such as ANOVA and other statistical values for both Fluoride and Chloride measurands respectively. Table 5 shows the summary of the built calibration models for these six HR calibration models considering the Fluoride and Chloride measurands.

Checking the performance and validity of these calibration functions is addressed in chapter five of this thesis.

Table 4.1: Input Raw Data for Measured Responses (Peak Areas) for Fluoride and Chloride Measurands.

Calibration Standard Level No.	Fluoride Calibration Standard (ppm)	Observed Peak Areas (uS/cm)	Chloride Calibration Standard (ppm)	Observed Peak Areas (uS/cm)
1	0.8536	0.01236	0.91059	0.00622
2	1.6108	0.02188	1.71839	0.0122
3	2.2058	0.0298	2.35318	0.01754
4	2.7496	0.03364	2.93327	0.02028
5	4.0989	0.0546	4.28657	0.03304
6	4.6751	0.0775	4.88917	0.04648
7	5.4221	0.07792	5.67029	0.045575
8	6.0262	0.0857	6.30206	0.05048
9	6.8784	0.10338	6.9212	0.05814
10	7.3554	0.109	7.40114	0.06084

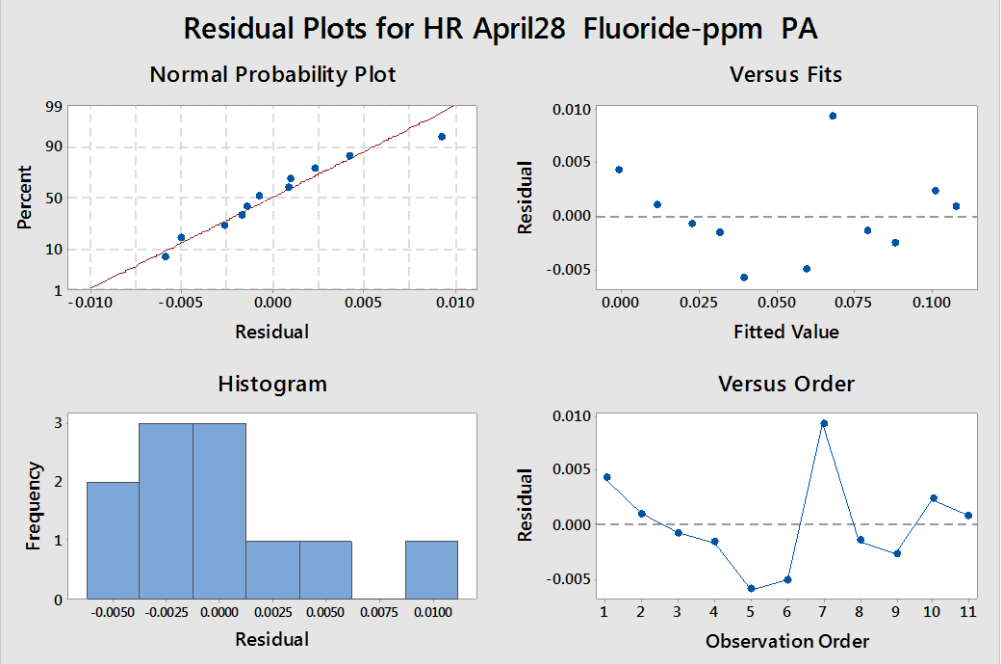


Figure 4.2: Residual Plots for HR April 28 Fluoride Model

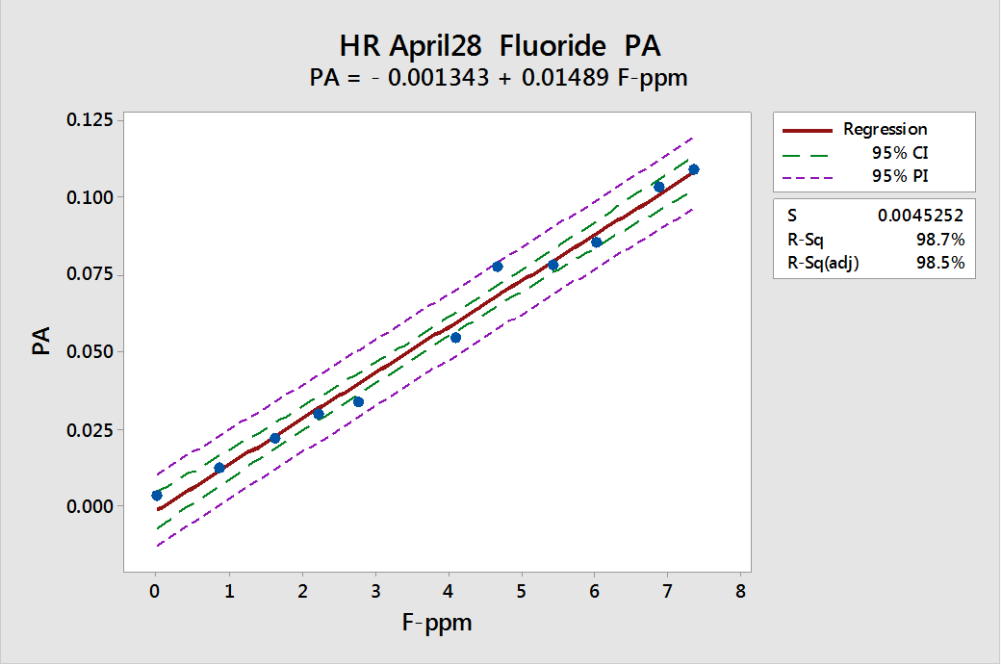


Figure 4.3: Calibration Curve for the HR April28 Fluoride Model.

Table 4.2: Fluoride Regression Analysis. PA versus F-ppm Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	0.013768	0.013768	672.34	0.000
F-ppm	1	0.013768	0.013768	672.34	0.000
Error	9	0.000184	0.000020		
Total	10	0.013952			

Model Summary	S	R-sq	R-sq(adj)	R-sq(pred)
	0.0045252	98.68%	98.53%	98.20%

Coefficients

Term	Coefficients	SE Coef	T-Value	P-Value	VIF
Constant (Intercept)	-0.00134	0.00258	-0.52	0.615	
F-ppm (Slope)	0.014888	0.000574	25.93	0.000	1.00

Regression Equation	PA = -0.00134 + 0.014888 F-ppm

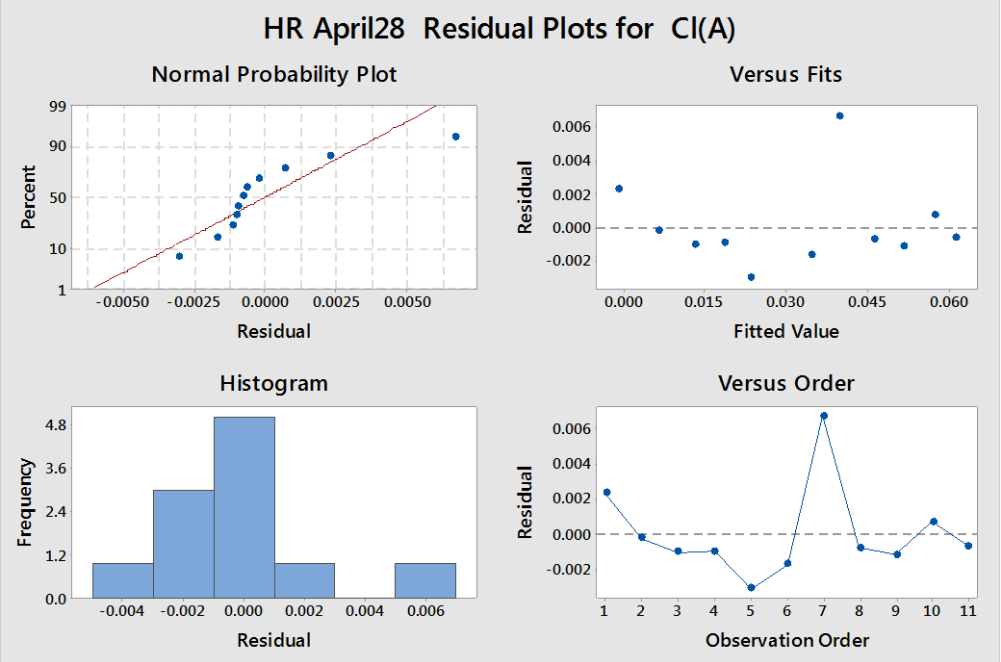


Figure 4.4: Residual Plots for HR April28 Chloride Model

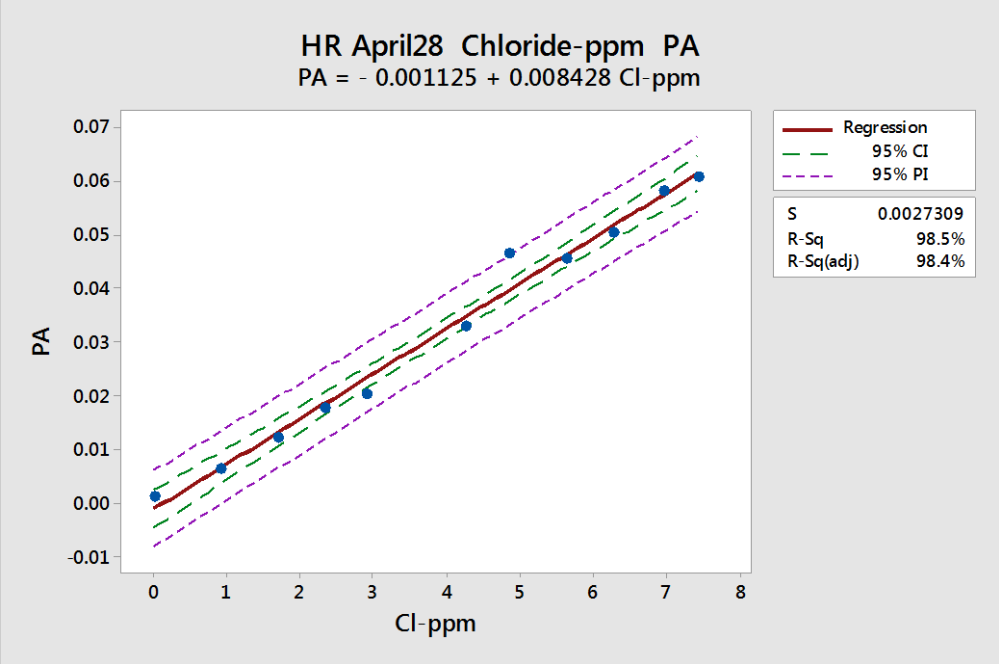


Figure 4.5: Calibration Curve for the HR April28 Chlroide Model.

Table 4.3: Chloride Regression Analysis: PA versus Cl-ppm Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	0.004538	0.004538	608.45	0.000
F-ppm	1	0.004538	0.004538	608.45	0.000
Error	9	0.000067	0.000007		
Total	10	0.004605			

Model Summary	S	R-sq	R-sq(adj)	R-sq(pred)
	0.0027309	98.54%	98.38%	98.04%

Coefficients

Term	Coefficients	SE Coef	T-Value	P-Value	VIF
Constant (Intercept)	-0.00113	0.00158	-0.71	0.493	
F-ppm (Slope)	0.008428	0.000342	24.67	0.000	1.00

Regression Equation	PA = -0.00113 + 0.008428 Cl-ppm

Table 4.4: HR Calibration Models Regression Functions

Model Name	Measurand	Calibration Functions	R^2 (adj)	Standard Deviation
HR April28,2016	Fluoride	PA = - 0.001981 + 0.01523 F-ppm	98.21%	0.0046497
	Chloride	PA = - 0.001645 + 0.008623 Cl -ppm	98.61%	0.0023968
	Bromide	PA = 0.000485 + 0.002756 Br -ppm	95.71%	0.0012930
HR May17,2016	Fluoride	PA = - 0.000372 + 0.01464 F-ppm	99.81%	0.0028279
	Chloride	PA = - 0.001495 + 0.008012 Cl-ppm	99.95%	0.0007461
HR June10,2016	Fluoride	PA = - 0.006466 + 0.01613 F-ppm	99.38%	0.0040475
	Chloride	PA = 0.000029 + 0.008534 Cl-ppm	99.63%	0.0016259
HR Aug21, 2016	Chloride	PA = 0.00131 + 0.007534 Cl-ppm	98.28%	0.0029560
HR Sep10, 2016	Fluoride	PA = -0.001021 + 0.01568 F-ppm	99.09%	0.0044520
	Chloride	PA = -0.000856 + 0.008415 Cl-ppm	99.04%	0.0026407
HR Oct29, 2016	Fluoride	PA = - 0.00280 + 0.014058 F-ppm	99.32%	0.0043982
	Chloride	PA = -0.000092 + 0.007653 Cl-ppm	99.88%	0.0009403
	Bromide	PA = 0.000767 + 0.002515 Br-ppm	99.50 %	0.000543

4.7 Validation Via Capability of Detection

Assessment of capability of detection is part of the validation process of any laboratory testing method. Capability of detection is referring to the minimum detectable limit that a valid laboratory measuring method can achieve. For the same measurement method, the quantified minimum detectable values could vary from one laboratory to another. Several factors contribute to this variability, but the most influential ones are; the performance of the used measuring system and the quality of the conducted calibration.

Two standardized approaches were used for evaluating the capability of the high range (HR) method in detecting measurands (F and Cl) concentration levels that are less than one part per million (<1-ppm). The first approach was according to the ISO11843-2 [56], in which the minimum critical value of the net measurand concentration X_c and the minimum detectable value X_d of measurand concentration were calculated based on the obtained critical response values of the chemical measurement system as per the formulas shown in section 4.2. Table 5 summarizes these calculated values for the six HR calibration models. Note that ISO11843-2 is addressing only cases where the relationship between the calibration standards and CMS responses values is linear. For chloride, it known as per the original method (UOP991) that the relationship is linear, while UOP991 was mentioning that for calibrating the CMS for the Fluoride measurand, a quadratic calibration function should be considered. This thesis proved that Fluoride data generated from the linear

regression was not significantly different from the ones generated from the quadratic one, see chapter five.

Table 4.5: Calculated critical values of; the response variable (y_c), net state variable (x_c) and minimum detectable values (x_d) , and the calculated values of the lowest concentration level of the prepared calibration standards (CAL STD#1) of each calibration set for the HR calibration models.

Model Name	**Measurand**	**(y_c)** µS/cm calculated	**(y_c)** µS/cm measured	**(x_c)** ppm	**(x_d)** ppm	**CAL STD#1** x_g -ppm
HR April28	Fluoride	0.0082050	0.0028826	0.64112	1.25042	0.8536
	Chloride	0.0046487	0.001178	0.68568	1.33728	0.90172
HR May17	Fluoride	0.007474	0.0148	0.53596	1.01497	1.002388
	Chloride	0.000575	0.0066	0.25838	0.489	0.961647
HR June10	Fluoride	0.00202	0.0130	0.52617	1.02623	1.0104
	Chloride	0.00343	0.00985	0.399	0.77917	0.9965
HR Aug21	Chloride	0.0075648	0.011033	0.83021	1.619217	1.12637
HR Sep10	Fluoride	0.002541	0.0114	0.00168	0.00327	0.9153
	Chloride	0.001053	0.0095	0.00070	0.00136	1.00154
HR OCT29	Fluoride	0.007277	0.0135	0.71682	0.99184	1.38420
	Chloride	0.002018	0.0072	0.27581	0.94145	0.532599

In ISO11843-2 [52] the estimation of the measurand minimum detectable values X_d was stipulating that calibration function should be linear and the standard deviation of residuals should be either constant or linearly related to the measurand concentration level. These two stipulations (i.e. having linear calibration curve or constant standard deviation of residuals) are "*often doubtful for measurand with concentration levels close to zero*" ISO11843-4 [61], such like the case of this thesis work where the measurands concentrations are less than one part per million (<1-ppm). Beside these stipulations, there are basic assumptions that all generated data should fulfill, such assumptions like; all CMS response values $Y_{i,s}$ should be normally distributed. Also, the chemical matrix of the used reference materials (i.e. the calibration standards) should be very similar to the test materials (i.e. the samples).

The assumptions aforementioned by ISO11843-2, are sometimes practically, not doable. So the performances of the built HR calibration models were evaluated based on the methodology given by ISO11843-4. This methodology provides two things;

1. A criterion for judging whether the minimum detectable value is less than a given level of the net measurand concentration level.
2. An experimental design for testing the conformity of this criterion.

The criterion for each measurnad of the HR calibration model was estimated by using this formula:

$$\frac{\bar{y}_g - \bar{y}_b}{\sqrt{s_b^2 + s_g^2}} \geq \frac{2z_{1-\alpha}}{\sqrt{J}} \quad (4\text{-}5)$$

And the formula for calculating the 100(1-γ)% lower Confidence Limit (CL) is:

$$CL = \frac{\bar{y}_g - \bar{y}_b}{\sqrt{s_b^2 + s_g^2}} - \frac{t_{1-\gamma} * \nu}{\sqrt{N}} \quad (4\text{-}6)$$

Where;

$\bar{y}_g$: Observed mean response of a measurand with the net state variable equals to the given value x_g. This given value which will be tested to determine whether it is greater than the minimum detectable value. So the calculated calibration standard lowest value is considered the given value x_g, in cased of the calibration model named HR April28, the given value $x_g = 1.12637$-ppm for the chloride measurand.

$\bar{y}_b$: Observed mean response of the basic state (i.e. the blank).

s_b^2: Estimate variance of responses for the blanks

s_g^2: Estimate variance of responses for a measurand with net state variable (i.e. net measurand concentration) equals to x_g

N: Number of replicates of measurements on each calibration standard in assessment of capability of detection. Since only one replicate was don, then N value considered to be =1.

v: Number of degrees of freedom.

$t_{1-\gamma}$: Quantile of the *t*-distribution with v degrees of freedom.

In table 6, the criteria numbers are shown for each measurands of the calibration model. These criteria numbers are calculated according to ISO11843-4

J: Number of replications of measurements on the calibration standard including the value of the net state variable (blank sample) in an application of the method.

$z_{1-\alpha}$: quantile of the standard normal distribution, where $(1-\alpha)$ is the confidence level.

For the sake of simplicity, J value was set to equal N value, so J=N=1. An approximation for the $100(1-\gamma)\%$ lower confidence limit (LCL) is given by equation 6. The decision rule is; if the LCL for $\frac{\bar{y}_g-\bar{y}_b}{\sqrt{s_b^2+s_g^2}}$ satisfies criterion shown in equation 5, then a minimum detectable value less than or equal to the given value x_g is confirmed. Recall that x_g is the value of calibration standard with the lowest concentration level. So from Table 6, for the calibration model HR April28, concerning the measurnad Chloride, it has LCL 13.69 which is less than the criterion 14.53, then it is confirmed that this model has detectable capability less than the given value $x_g = 0.91059 - ppm$. For the Fluoride measurand, LCL was found to be 13.63 while its respective criterion calculated to be 14.46, so the HR April28 calibration model is capable to detect Fluoride value less than the given value $x_g = 0.8536 - ppm$, similarly for the remaining of data shown in table 6 for the other HR calibration models. Based on the data displayed on table 6, all the HR calibration models are capable of detecting Fluoride and

Chloride concentrations levels less than the concentration of the lowest calibration Fluoride/Chloride concentration levels (x_g) values.

Table 4.6: Capability of detection for HR calibration models according to ISO11843-4:2003 (E)

Models		**Fluoride**	**Chloride**
HR April28	**Criterion**	**14.46**	**14.53**
	Lower Confidence Limit (LCL)	13.63	13.69
	Given Value x_g-ppm	0.8536	0.90172
HR May17	**Criterion**	**1.032**	**1.00**
	Lower Confidence Limit (LCL)	0.200	0.1727
	Given Value x_g-ppm	1.002388	0.961647
HR June10	**Criterion**	**NA**	**NA**
	Lower Confidence Limit (LCL)	NA	NA
	Given Value x_g-ppm	1.0104	0.9965
HR Aug21	**Criterion**	**NA**	**3.45**
	Lower Confidence Limit (LCL)	NA	1.25
	Given Value x_g-ppm	NA	1.6192171
HR Sep10	**Criterion**	**2.97**	**25.26**
	Lower Confidence Limit (LCL)	2.13	24.43
	Given Value x_g-ppm	0.9153	1.00154
HR OCT29	**Criterion**	**26.45**	**17.21**
	Lower Confidence Limit (LCL)	25.62	16.38
	Given Value x_g-ppm	1.38420	0.532599

CHAPTER 5

METHOD VALIDATION

Method validation is considered as one of the quality assurance measures exercised by laboratory prior accepting a modified method. The proposed HR method is a modification of the original UOP991 method, where the three changes which are ; changing the calibration working range, extending the calibration frequencies to be monthly and modelling the fluoride measurnad by the linear regression instead of the quadratic are evaluated.

1.1. Introduction

Traditionally, validation has been seen as a process of assessing the scientific validity of an alternative method. Note that by alternative method, it is meant both; method that is partially deviating from the original method –which is the scope of this thesis work- or a method that is completely different but used for analyzing the same sample and measurands but via utilizing totally different analytical testing procedure and technology. The method validation concept was introduced in the early 90'es. In 1994, the International Conference on Harmonization of Technical Requirements For Registration of Pharmaceuticals for Human Use (ICH) issued a harmonized guide for validating analytical procedure [10]. This is in addition to the Harmonized Guidelines For Single Laboratory Validation of Method of Analysis that was developed in 2002 by the International Union of Pure and Applied Chemistry (IUPAC) [11] and the ISO 11726 guidelines for validation of alternative method [12]. This thesis work is addressing validating a modified method with respect to the

acceptance of test results obtained from this modified method. But strictly speaking, validation should refer to the whole *analytical system* rather than the analytical *method*. By analytical system which comprising a defined method protocol, a defined concentration range for the measurand, and the type of materials used in the validation work. This validation work is done within-laboratory, also called Single-Laboratory Validation.

The use of non-standardized test methods or a method that is slightly derived from standardized method has to be validated. So in this thesis, the modified method is called the High Range and denoted by HR. This HR method was originally derived from a standardized method called UOP991 [1]. The only deviations of the HR method from UOP991 are identified into two points; UOP991 quantitative working range is from 0.1-ppm to about 1-ppm and it dictates using quadratic function to calibrate the measuring equipment for the quantification of Fluoride measurand. While the modified method HR is calibrated for covering quantitative working range from about 1-ppm to 10-ppm, which is ten times higher than the UOP991 and linear regression was used to calibrate for the Fluoride measurand instead of the quadratic.

5.2 Why Method Validation is Needed?

The main motivation behind deviating from the UOP991 and validating the modified method HR is to make the original method i.e. UOP991 both rapid and easier to perform so a modified method called "HR" was proposed as an alternative for the UOP991. Note that UOP991 requires to do calibration each time sample to be tested and its calibration procedure requires to calibrate the chemical measuring equipment with five calibration standards, each run three times, then select the best run file for each calibrant "regressor", so it is both labor and time intensive. This makes the calibration procedure consumes about seven hours to calibrate the measuring equipment. So what about the time left -among eight working hours- for measuring samples and quality audit samples? Moreover, it is much simpler to prepare the chemical calibration standards "regressors" in high range rather than in very minute scale, i.e. from 0.1-to-1-ppm. Once the derived method is validated, it will be able to do comparability of results obtained within the same laboratory or with different laboratory to assure that data generated by these two methods for the same tested samples are not significantly different.

From regulatory perspective and according to the international standard ISO17025 [13], "*the laboratory shall validate non-standard methods, laboratory-developed methods and standard methods used outside their intended scope or otherwise modified. The validation shall be as extensive as is necessary to meet the needs of the given application or field of application*". Notice the acceptance criteria here set by this ISO17025 that "the validation shall be as extensive as necessary to meet the needs of the application.

The work at which this thesis data were generated was conducted at ISO17025 certified and accredited laboratory, so full compliance with ISO17025 requirements is a must. Method validation in the field of laboratory analysis is very active subject and well standardized too – with the exception of ways of identifying method Limits Of Detection "LOD" and Limit Of Quantification "LOQ"-, some of these validation protocols references are [14-18].

Statistical concepts play a pivotal role in validating the performance of modified method. These basic essential statistical tools needed for validating analytical method are summarized nicely into this reference [19]. Precision is one of the fundamental elements in method development and validation and it is mainly referring to ways of quantifying the repeatability, reproducibility and site-precision. Review paper shown in reference [20] is summarizing the uses and misuses of the term precision in chemical analysis.

5.3 Strategy Followed in Method Validation

This thesis work was conducted at petroleum refinery quality control, quality assurance laboratory (QC/QA) that operates 24/7 and has only one chemical measuring equipment that operates according to the reference method UOP991 and it is used almost daily. This situation contributed in shaping the strategy used in validating the modified method HR to be as follows:

5.3.1 Lean Design of Validation Studies

Data shown into this thesis work are characterized by being small in size, almost no repeatability of measurements and validation was done within the same laboratory. Each calibration standard, supposed to be run in duplicate, preferably triplicates. But due to the high volatility of the chemical from which these standards were prepared in – this chemical is Toluene-, only single run was done for each tested calibration standard. Ultimate object is to come up with an economical design of calibration.

5.3.2 Single Laboratory Validation

In single laboratory validation, there is a particular danger that laboratory bias may be overlooked, so trueness – which is the quantitative measure of bias- was conducted. Trueness is the closeness of agreement between a test result and the accepted value of the property being measured, smaller bias values indicate greater trueness. In real life, it is quite difficult to find a reference material – these are chemicals with known/traceable measurand value- that is matching the property being measured with respect to both concentration level and sample matrix. So the

trueness (bias) of the data obtained experimentally was checked by different ways. Firstly, by in-house preparation of Quality Control "QC" Chloride audit samples with known measured values at different concentration levels and reproduced their values by using both the UOP991 built methods and the built methods of the modified method HR as shown in table 1. Secondly, by reproducing the low range UOP991 prepared calibration standards values by using the HR calibration models and vice versa as detailed in cross validation part, section 4.3.3. Thirdly, by comparing with alternative testing procedures, see clause 4.3.4 and finally by doing retrospective and prospective validation as shown in section 4.3.5. This integrated approach will synergize and enlarge the strength of the final finding which is that test results obtained from UOP991 is not really significant from the ones obtained from the modified method High Range method "HR".

Each of these chloride quality control "QC" audit samples that are shown in table 5.1 were prepared in-house with the exception of FLUKA which was a water based reference material and the ASTM PT Sample "PT NP1602" which was proficiency testing "PT" Naphtha that was tested on February 2016 by three laboratories around the globe. Data shown in tables (5.2-5.7) are selected from the best models among the 11 UOP calibration models and 6 of the HR calibration models.

Table 5.1: Chloride QC Audit Samples that were used to verify the trueness of the modified HR Method and validity of the built calibration models for both UOP991 and the HR method.

No.	Chloride QC Audit Sample Name	Chloride Concentration Level in-ppm	Remarks
1	July 28,2016	0.10698	Prepared in House
2	September 22,2016	0.11829	
3	July 28, 2016	0.49799	
4	October 14,2015	1.0743	
5	FLUKA Aqueous Chloride Standard	10.00	Certified Reference Materials.
6	ASTM PT NP1602	1.00	According to ASTM D5808.
		0.5-ppm, 1.4-ppm and 1.4-ppm	Three Readings according to ASTM D5194. These two readings obtained by three different laboratories for the same sample

Table 5.2: Reproduced Data for Chloride QC Audit Samples (0.10698-ppm Sep22) by UOP991 and HR Method.

Run Replicate No.	UOP991 April 25	HR June 10
1	0.113780173	0.199140961
2	0.113780173	0.199140961
3	0.125498008	0.21215671
4	0.090344504	0.173109462
5	0.102062339	0.186125212
6	0.113780173	0.199140961
7	0.113780173	0.199140961
8	0.125498008	0.21215671
9	0.137215843	0.225172459
10	0.125498008	0.21215671

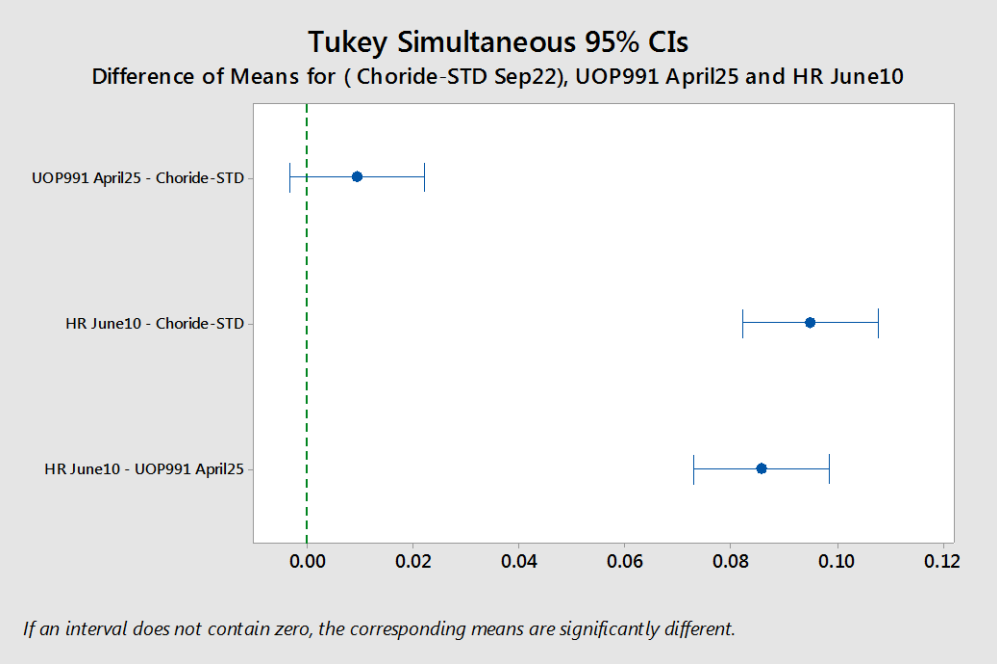


Figure 5.1: Turkey's multiple comparisons for the reproduced Chloride QC audit sample (0.10698-ppm Sep22)

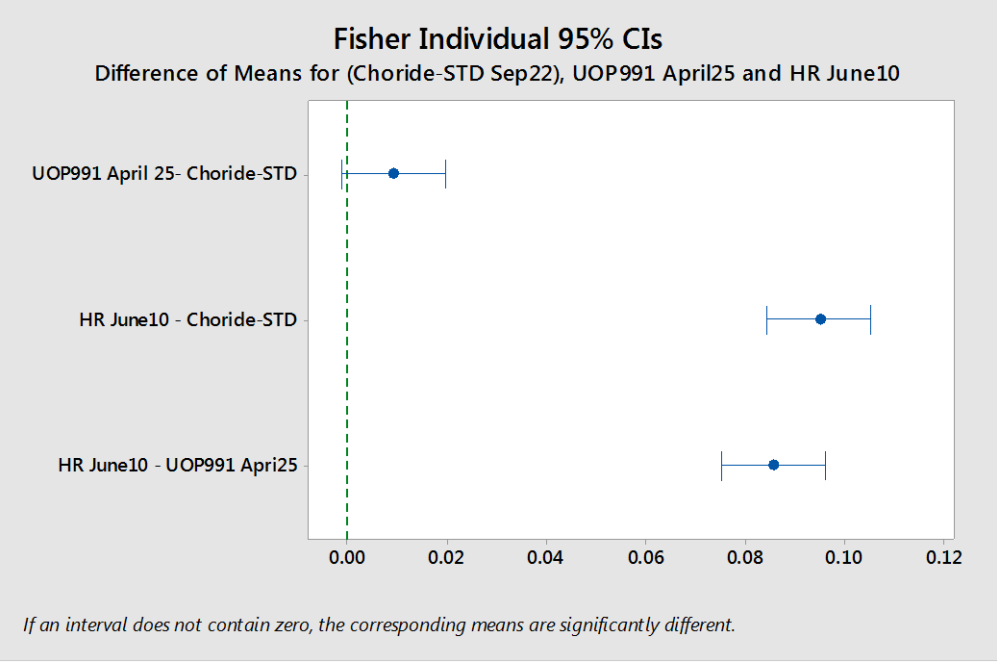


Figure 5.2: Fisher LSD for the reproduced Chloride QC audit sample (0.10698-ppm Sep22)

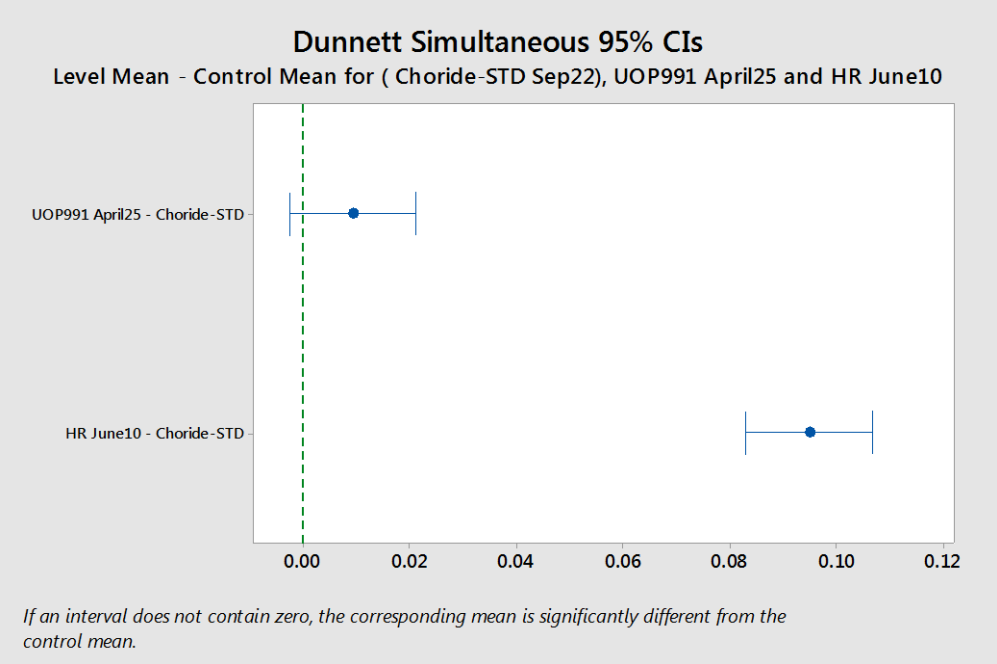


Figure 5.3: Dunnett multiple comparisons for the reproduced Chloride QC audit sample (0.10698-ppm Sep22)

Table 5.3: Reproduced Data for Chloride QC Audit Samples (0.11829-ppm Sep22) by UOP991 and HR Method.

Run Replicate No.	UOP991 October 19	UOP991 October 29	HR June 10
1	0.104812223	0.183963605	0.277829857
2	0.147651007	0.226613591	0.312983361
3	0.133371412	0.212396929	0.301265526
4	0.076253034	0.155530281	0.254394188
5	0.104812223	0.183963605	0.277829857
6	0.104812223	0.183963605	0.277829857

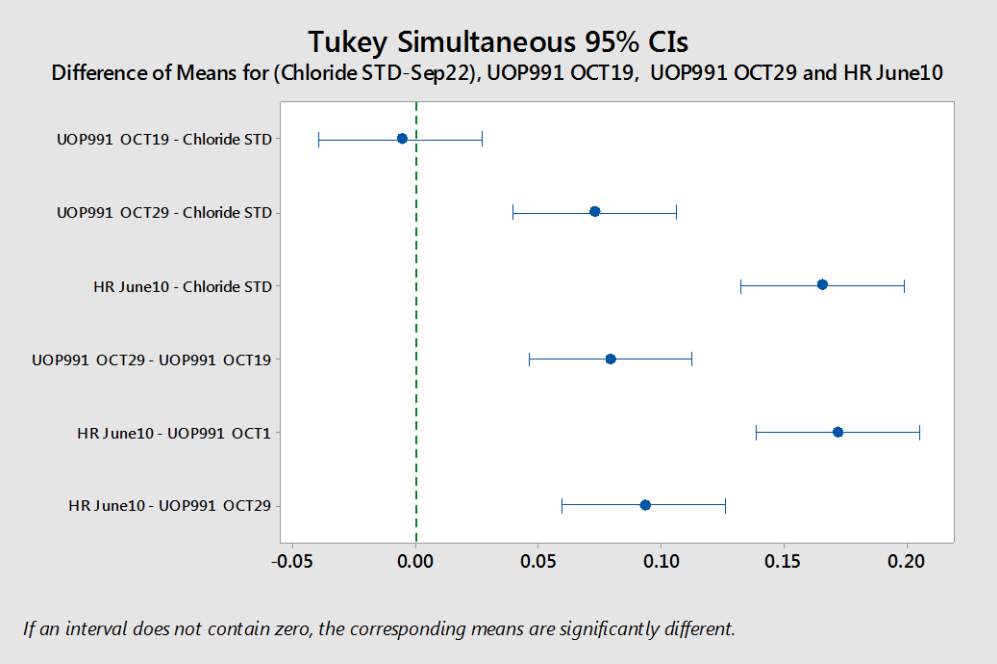


Figure 5.4: Turkey's multiple comparisons for the reproduced Chloride QC audit sample (0.11829-ppm Sep22)

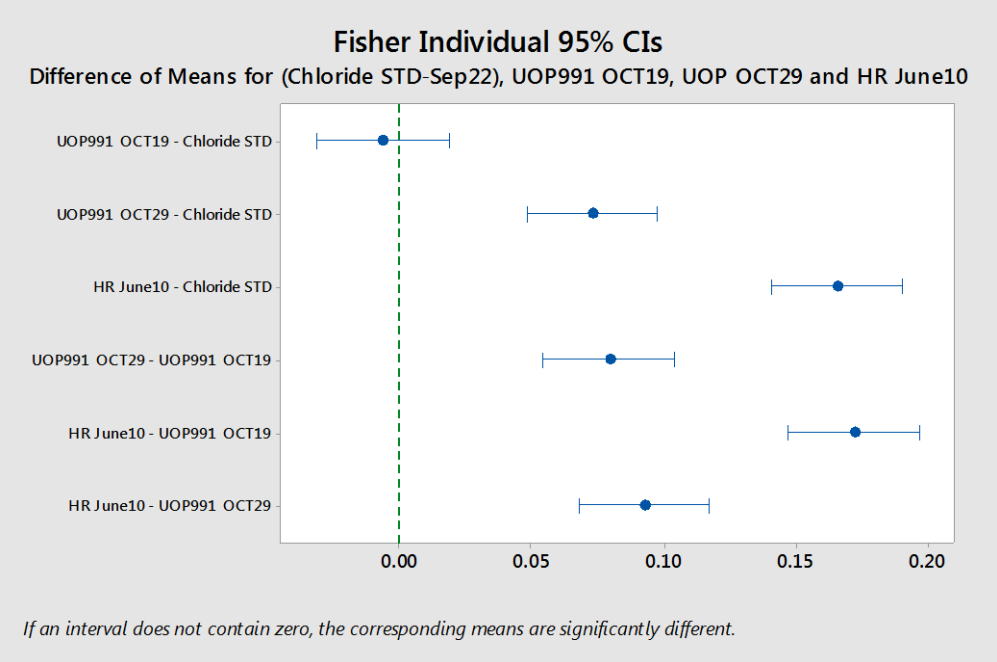


Figure 5.5: Fisher LSD for the reproduced Chloride QC audit sample (0.11829-ppm Sep22)

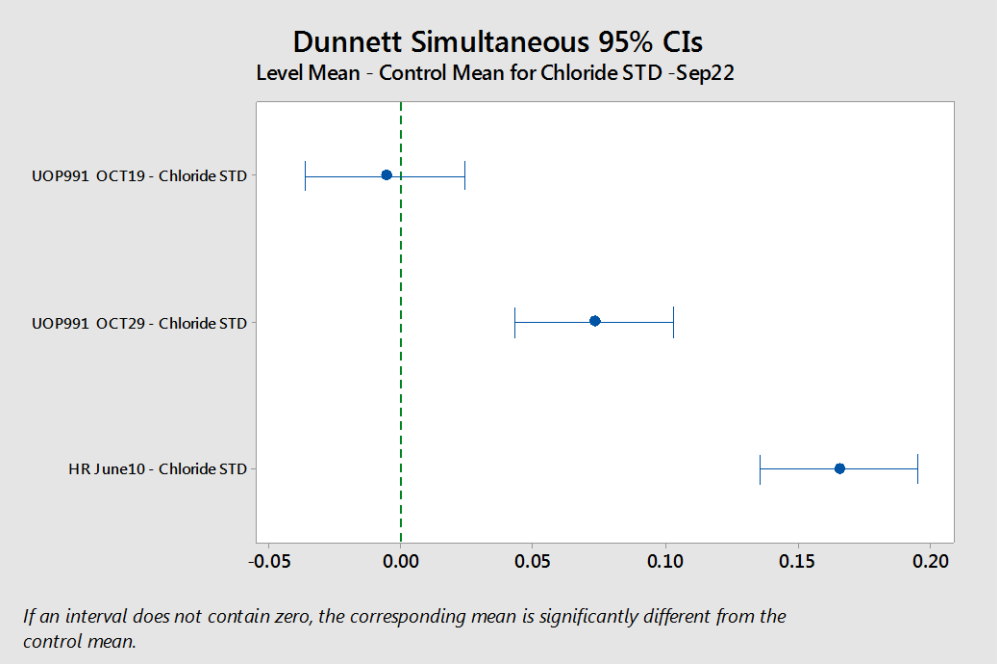


Figure 5.6: Dunnett multiple comparisons for the reproduced Chloride QC audit sample (0.11829-ppm Sep22)

Table 5.4: Reproduced Data for Chloride QC Audit Samples (0.49799-ppm July28) by UOP991 and HR Method.

Run Replicate No.	UOP991 October 29	HR June 10
1	0.43986352	0.488750879
2	0.454080182	0.500468713
3	0.454080182	0.500468713
4	0.496730168	0.535622217
5	0.454080182	0.500468713
6	0.411430196	0.46531521
7	0.43986352	0.488750879
8	0.425646858	0.477033044
9	0.454080182	0.500468713
10	0.411430196	0.46531521

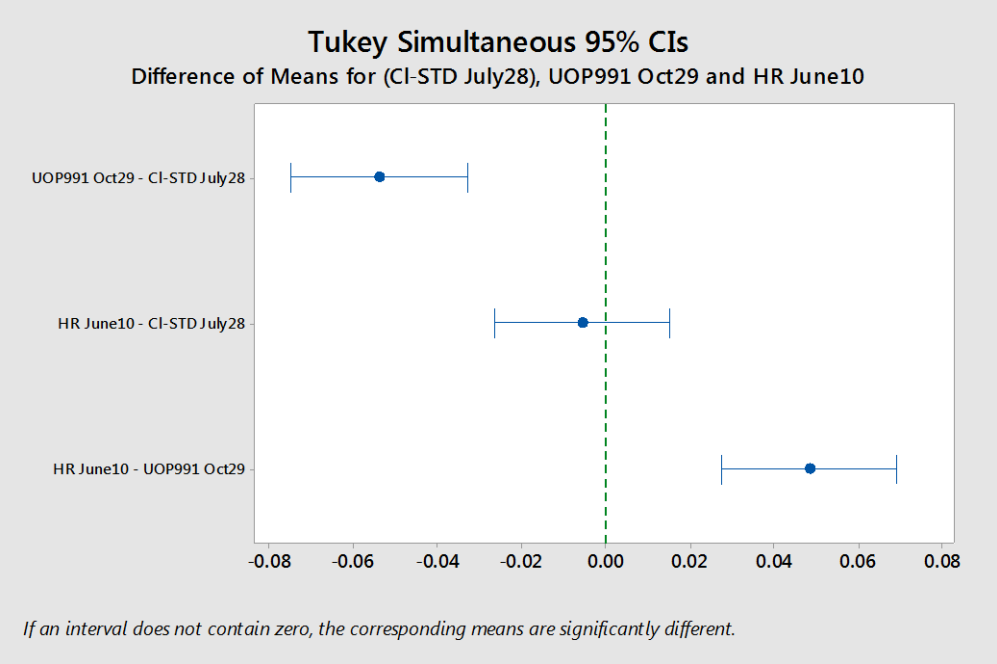


Figure 5.7: Turkey's multiple comparisons for the reproduced Chloride QC audit sample (0. 49799-ppm July28)

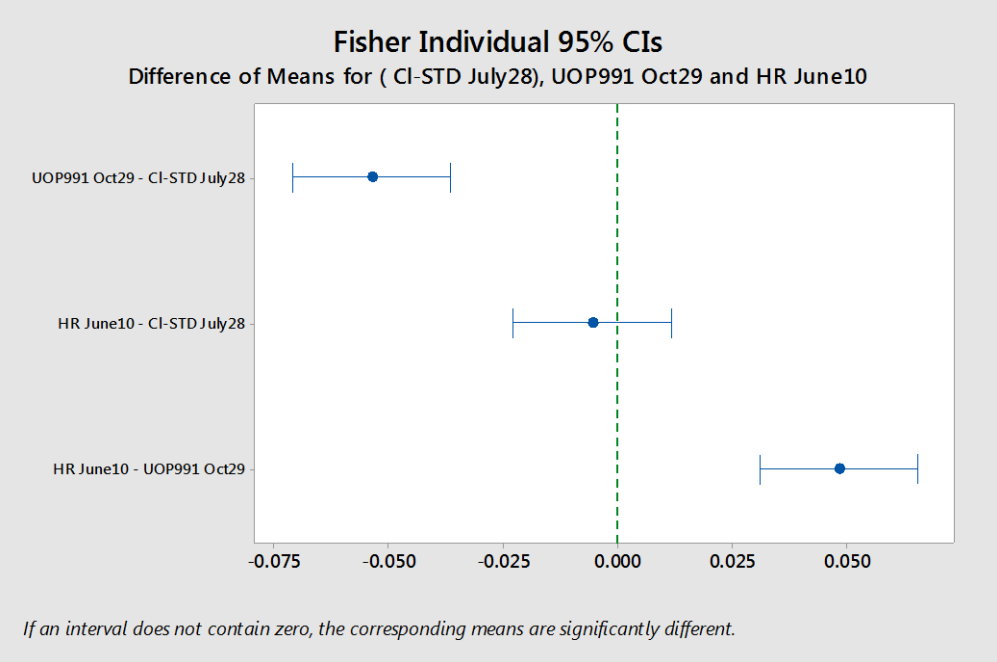


Figure 5.8: Fisher LSD for the reproduced Chloride QC audit sample (0. 49799-ppm July28)

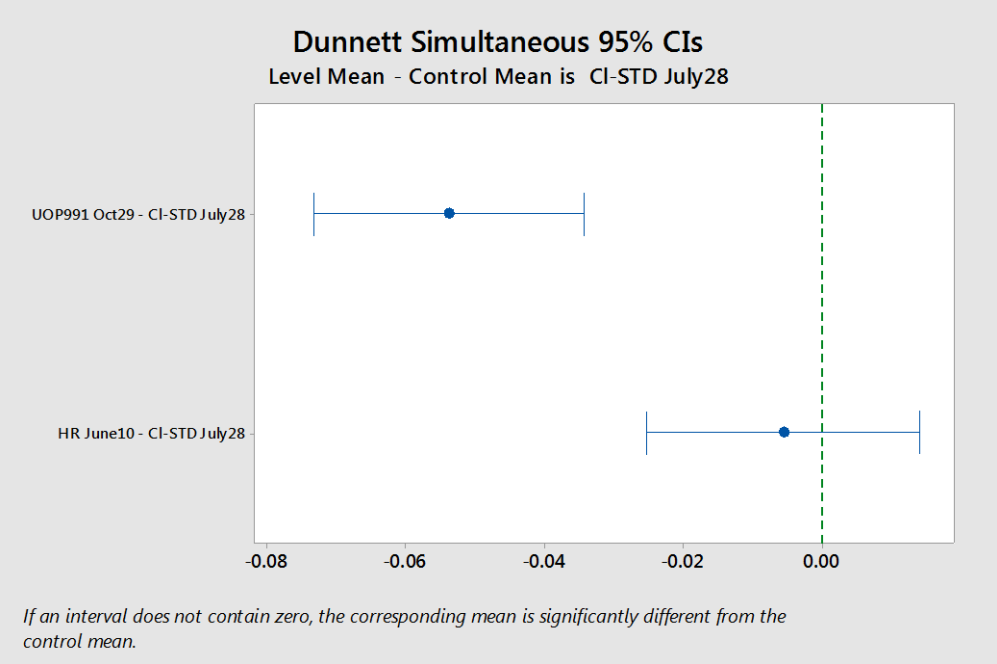


Figure 5.9: Dunnett multiple comparisons for the reproduced Chloride QC audit sample (0. 49799-ppm July28)

Table 5.5: Reproduced Data for Chloride QC Audit Samples (1.0743-ppm Oct14, 2015) by UOP991 and HR Method.

Run Replicate No.	UOP991 October 29	HR June 10
1	1.17912994	1.098078275
2	1.150696616	1.074642606
3	1.164913278	1.086360441
4	1.164913278	1.086360441
5	1.093829969	1.027771268
6	1.136479955	1.062924772
7	1.150696616	1.074642606
8	1.093829969	1.027771268
9	1.17912994	1.098078275
10	1.093829969	1.027771268

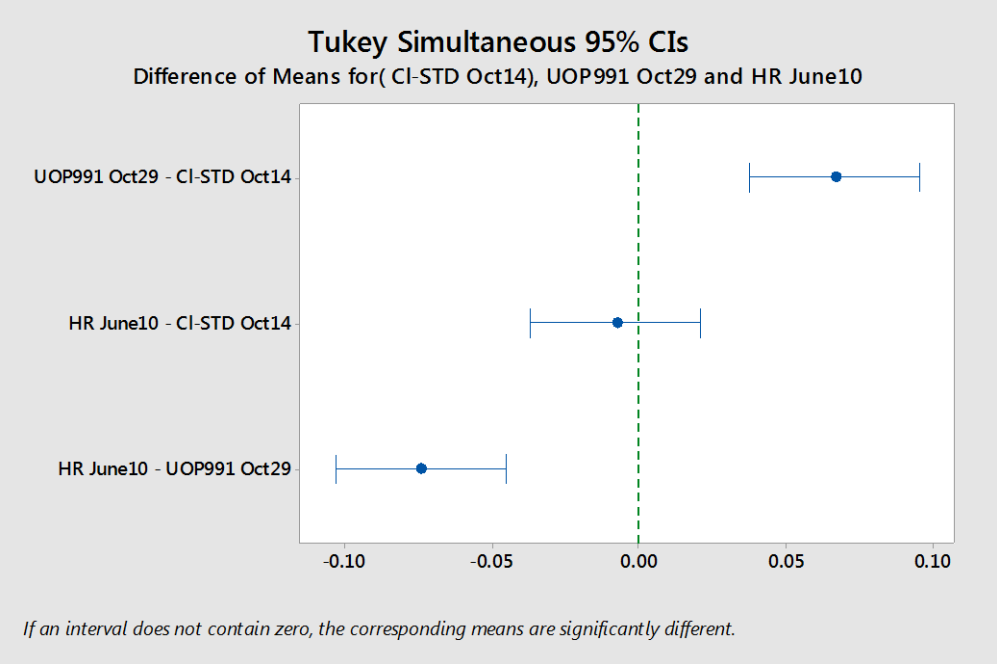


Figure 5.10: Turkey's multiple comparisons for the reproduced Chloride QC audit sample (1.0743-ppm Oct14, 2015)

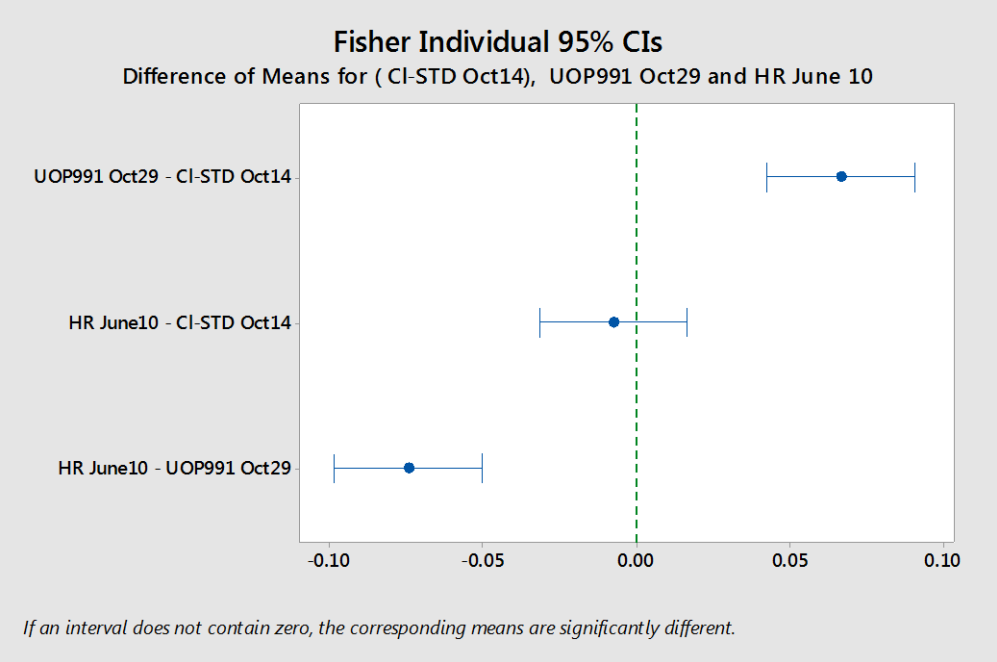


Figure 5.11: Fisher LSD for the reproduced Chloride QC audit sample (1.0743-ppm Oct14, 2015)

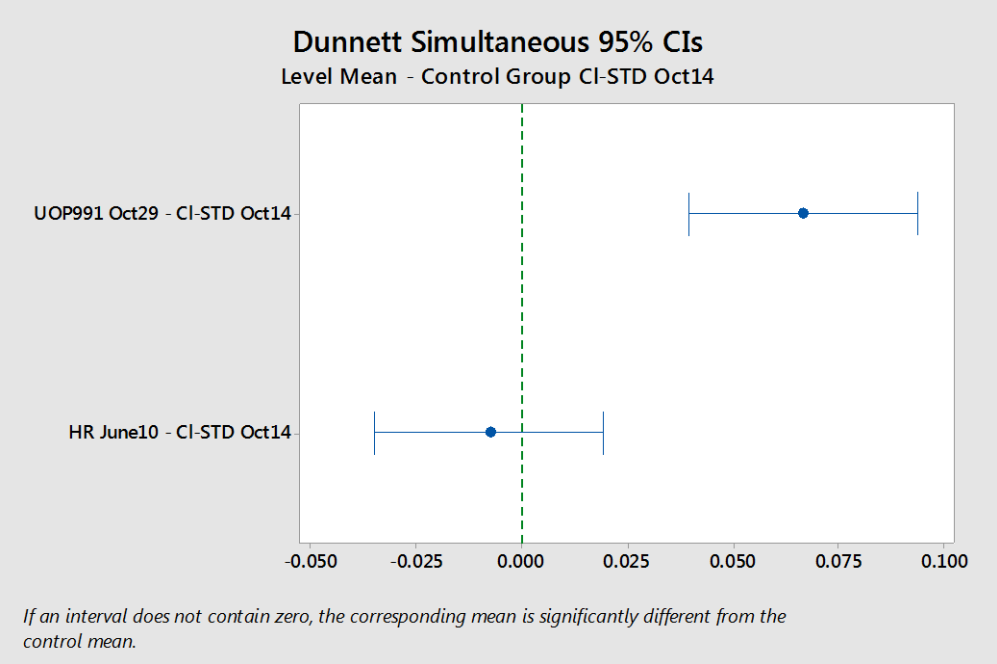


Figure 5.12: Dunnett multiple comparisons for the reproduced Chloride QC audit sample (1.0743-ppm Oct14, 2015)

Table 5.6: Reproduced Data for Chloride QC Audit Samples (10.00-ppm FLUKA) by UOP991 and HR Method.

Run Replicate No.	UOP991 October 29	HR June 10
1	14.05942565	11.71443637
2	13.94569235	11.6206937
3	13.68979244	11.40977267
4	14.42905886	12.01910007
5	14.42905886	12.01910007
6	14.20159227	11.83161472

*: Fluka is an aqueous standard that has Chloride concentration level =1.00. It was tested by is manufacturer by using IC only without the combustion part.

Table 5.7: Reproduced Data for Chloride QC Audit Samples (ASTM PT Naphtha Sample Dated February, 2016) by UOP991 Method.

Run Replicate No.	UOP991 January 26	UOP991 April 25	UOP991 May 12	UOP991 October29
1	1.092042589	0.941038657	0.989398357	0.795280068
2	1.148080695	0.993101653	1.042406573	0.852146716
3	1.106052115	0.954054406	1.002650411	0.80949673
4	1.134071168	0.980085904	1.029154519	0.837930054
5	1.134071168	0.980085904	1.029154519	0.837930054
6	1.162090221	1.006117402	1.055658627	0.866363378
7	1.106052115	0.954054406	1.002650411	0.80949673
8	1.120061642	0.967070155	1.015902465	0.823713392
9	1.106052115	0.954054406	1.002650411	0.80949673

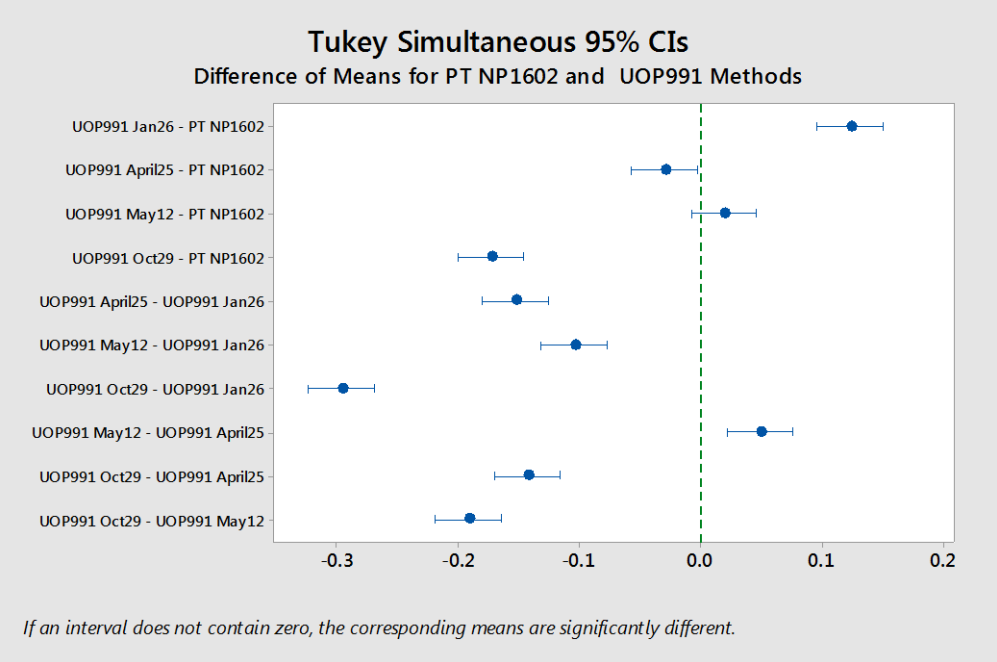


Figure 5.13: Turkey's multiple comparisons for the reproduced Chloride QC audit sample (ASTM PT NP1602)

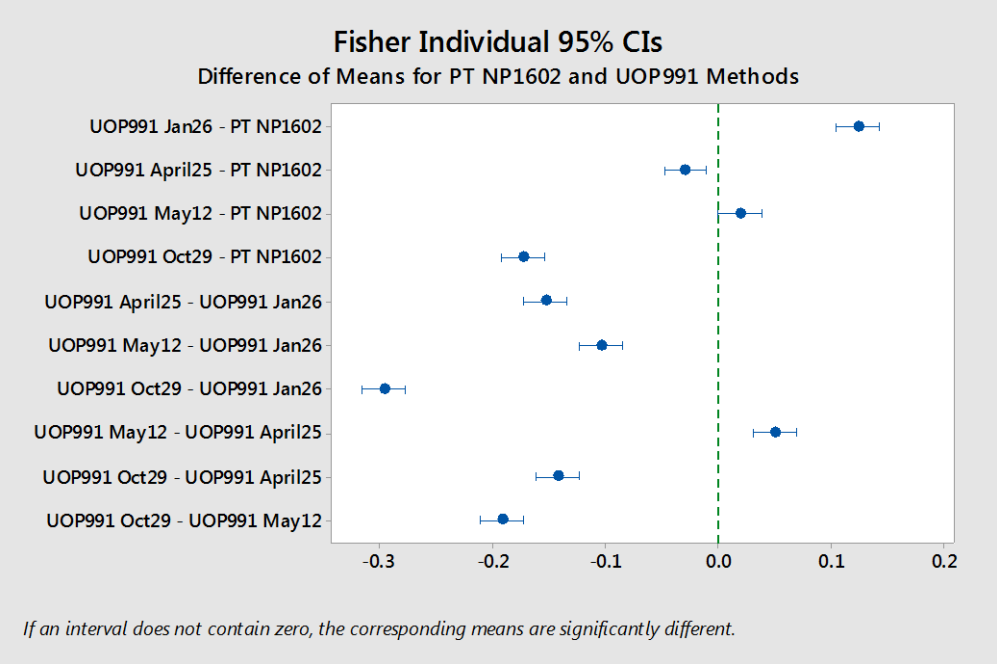


Figure 5.14: Fisher LSD for the reproduced Chloride QC audit sample (ASTM PT NP1602)

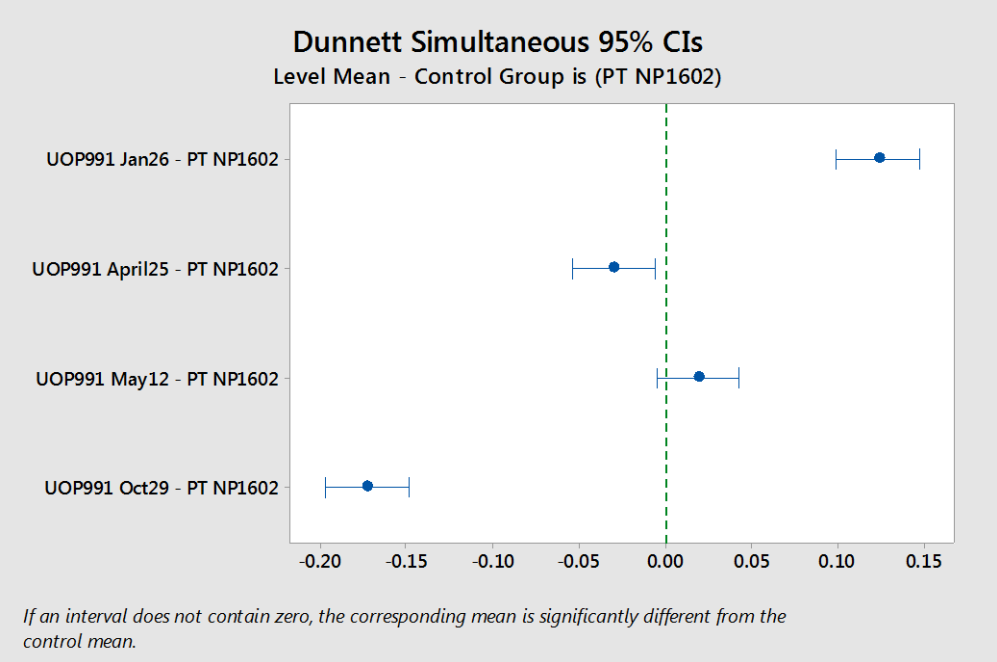


Figure 5.15: Dunnett multiple comparisons for the reproduced Chloride QC audit sample (ASTM PT NP1602)

5.3.3 Validation by Reproducing Low Range Chloride Calibration Standards by the High Range Method and Vice versa

In this section, the projection power of the high range HR calibration models in predicting the chloride measurand that falls in the low range standard UOP991 method was tested by reproducing the low range calibration standards values by using the HR calibration models. These low range chloride standards were prepared in-house and used to calibrate the measuring equipment according to UOP991 procedure. Total of 11 chloride low range calibration standards values were reproduced by using their measured peak areas – obtained from the lab measuring equipment which was calibrated according the UOP991- as an inputs into the built 5 high range calibration models. HR calibration models that were able to reproduce chloride concentration level that is not significantly different from the ones obtained by the low range UOP991 are shown in table 5.8.

Table 5.8: Summary table showing reproduced Low Range Chloride Calibration Standards Values by using the High Range Calibration Models.

Chloride Standards Set No.	Low Range Chloride-ppm Standards	Low Range Chloride Preparation Date	Reproduced Chloride Values by HR August21	
1	0.114	February 22, 2016	0.26283	
	0.2714		0.40116	
	0.5123		0.63632	
	0.9767		1.106653	
	Low Range Chloride-ppm	**Low Range Chloride Preparation Date**	**Reproduced Chloride Values by HR Sep10,2016**	**Reproduced Chloride Values by HR October29,2016**
2	0.10663	April 25, 2016	0.137374	0.133118
	0.26594		0.279976	0.288057
	0.51162		0.505764	0.533376
	0.9667		0.921687	0.985281
	Low Range Chloride-ppm	**Low Range Chloride Preparation Date**	**Reproduced Chloride Values by HR May17,2016**	**Reproduced Chloride Values by HR September 10,2016**
3	0.11014	May12, 2016	0.199076	0.1136066
	0.2687		0.3313779	0.239572
	0.51998		0.5922366	0.487982
	0.98013		1.0103594	0.886036
	Low Range Chloride-ppm	**Low Range Chloride Preparation Date**	**Reproduced Chloride Values by HR August21,2016**	
4	0.10868	June 1, 2016	0.276663	
	0.20816		0.304329	
	0.52363		0.553326	
	1.03164		0.875639	
	Low Range Chloride-ppm	**Low Range Chloride Preparation Date**	**Reproduced Chloride Values by HR August21,2016**	
5	0.10002	June 9, 2016	0.183289	
	0.21384		0.3354544	
	0.31048		0.449578	
	0.41258		0.536035	
	0.513386		0.615576	
	0.61279		0.702033	
	0.71203		0.819615	
	0.80379		0.864573	
	0.87777		0.933739	
	0.977681		0.951035	

Table 5.8 cont'd: Summary table showing reproduced Low Range Chloride Calibration Standards Values by using the High Range Calibration Models.

Chloride Standards Set No.	Low Range Chloride-ppm	Low Range Chloride Preparation Date	Reproduced Chloride Values by HR August21
6	0.103	August 30, 2016	0.0968322
	0.208		0.2074975
	0.383		0.3687923
	0.786		0.7746576
	0.928		0.89915161
	Low Range Chloride-ppm	**Low Range Chloride Preparation Date**	**Reproduced Chloride Values by HR August 21,2016**
7	0.1172367	September 11	0.091299
	0.283277		0.27113
	0.551989		0.53396
	1.04311		0.99598
	Low Range Chloride-ppm	**Low Range Chloride Preparation Date**	**Reproduced Chloride Values by HR August 21,2016**
8	0.103	October 19, 2016	0.055332
	0.3215		0.1798312
	0.5101		0.3873288
	0.7611		0.650159
	0.9845		0.885323
	Low Range Chloride-ppm	**Low Range Chloride Preparation Date**	**Reproduced Chloride Values by HR**
9	0.1049623	October 29,2016	0.3439638
	0.255472		0.4515171
	0.490675		0.688315
	0.941455		1.092834

5.3.4 Validation by Using Data from Alternative Method

Into this section, a proficiency testing (PT) sample called (ASTM PT1602) is obtained from the American Society for Testing and Materials (ASTM) and dated February 2016. This sample was tested by some different laboratories around the globe using testing procedures different from UOP991. Three laboratories tested this PT sample according to ASTM D5194 [7] and reported its chloride concentration level to be; 0.5-ppm and two other laboratories reported 1.4-ppm. Other lab tested the PT sample according the other testing procedure ASTM D5808 [5] and reported its chloride content to be 1.00-ppm. So the chloride test result of 1.00-ppm is considered to be the most probable one. This ASTM PT1602 sample was tested by five high range calibrated method and the reproduced chloride test results of this PT1602 ASTM sample were very close to 1.00-ppm . This shows clearly that although the 1.00-ppm chloride content that was obtained by using testing procedure [5] different from the one covered by the HR method, yet the HR modified UOP991 method was capable to reproduce chloride data at the upper concentration limit that is not far from the one produced by the ASTM D5808 [5]. Two more findings to gain from this exercise are: the operational life of the built HR calibration model is not less than one month and the shelf life of the PT ASTM 1602 sample can be ten months. See table 5.9

Table 5.9: Reproduced Data for Chloride QC Audit Samples (ASTM PT Naphtha Sample Dated February, 2016) by HR Methods.

Run Replicate No.	HR April 28	HR May17	HR June 10	HR September10	HR October29
1	1.011382114	1.022840739	0.781696742	0.89792038	0.959457715
2	1.057839721	1.072765851	0.828568081	0.945454545	1.011103938
3	1.022996516	1.035322017	0.793414577	0.909803922	0.97236927
4	1.046225319	1.060284573	0.816850246	0.933571004	0.998192382
5	1.046225319	1.060284573	0.816850246	0.933571004	0.998192382
6	1.069454123	1.085247129	0.840285915	0.957338087	1.024015494
7	1.022996516	1.035322017	0.793414577	0.909803922	0.97236927
8	1.034610918	1.047803295	0.805132412	0.921687463	0.985280826
9	1.022996516	1.035322017	0.793414577	0.909803922	0.97236927

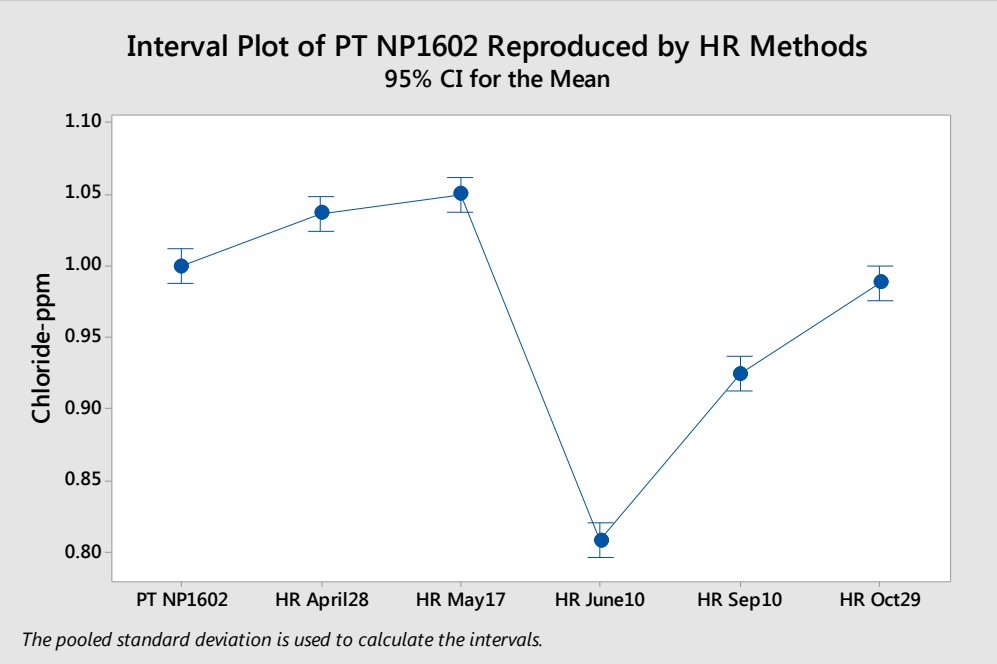


Figure 5.16: Interval Plot for the Chloride QC audit sample (ASTM PT NP1602) means reproduced by HR models.

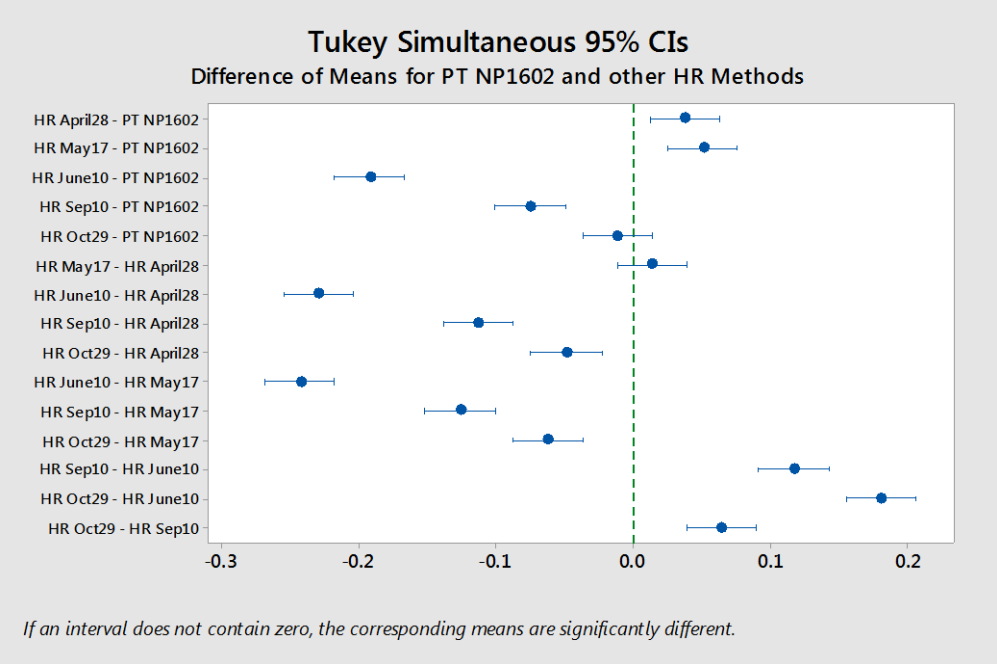


Figure 5.17: Tukey multiple comparisons for the reproduced Chloride QC audit sample (ASTM PT NP1602). Note how calibration models named HR Oct29 and HR May17 are equivalent in their performance.

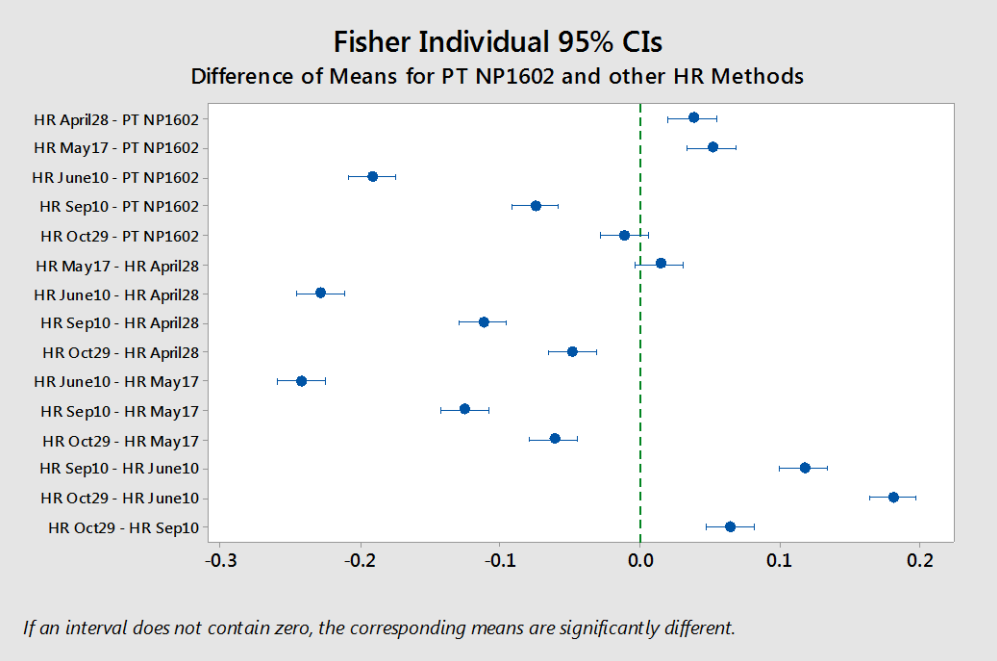


Figure 5.18: Fisher LSD for the reproduced Chloride QC audit sample (ASTM PT NP1602). Note how calibration models named HR Oct29 and HR May17 are equivalent in their performance.

5.3.5 Validation via Using Retrospective and Prospective Lab Data

Validation studies could be done by grouping data into two groups; one group dated older than the date of which the calibration model was built hence *retrospective*. The second group of data was gathered at dates after the date in which the calibrated model was built, hence *prospective*. By doing so, we will be able not only to check the calibration model performance with respect to reproducing these two types of data, but also checking the operational life of the built calibration models.

Table 5.10 shows chloride tests results for actual samples, the same samples were retested by five calibrated methods – from A to E-, where A,B and C are methods calibrated as per UOP991 and D/E are high range calibrated models.

If the acceptable criterion is that the difference between the obtained samples tests results and any of the data reproduced by these calibrated models – A-to- E- to be ±0.3-ppm, then all the calibrated models are capable to produce data that are not significant different than the original obtained sample test results for chloride that were originally obtained by method developed according to the UOP991. In fact, the HR calibrated models lettered D and E are showing better performance than the low range UOP991 calibrated models – A-to-C-.

Table 5.10: Low range real samples Chloride tests results reproduced by three UOP991 calibrated models and two HR calibrated models*.

Sample No.	Sample Testing Date	Sample Test Results Chloride-ppm	A	B	C	D	E
J24 D101	OCT31,2016	0.0616	0.2935	0.1991	0.2340	0.1138	0.1427
J24 C-201	OCT31,2016	0.0542	0.2935	0.1991	0.2340	0.3013	0.3518
J80-D106	OCT31,2016	0.2469	0.5177	0.4074	0.4461	0.3130	0.3648
QRT-1 with 5%SLOP	NOV01,2016	0.2911	0.5317	0.4204	0.4593	0.4536	0.5216
QRT-1 with 10%SLOP	NOV01,2016	0.4669	0.6998	0.5766	0.6183	0.1138	0.1427
J88 D-102	NOV01,2016	0.0433	0.2935	0.1991	0.2340	0.1255	0.1558
PLT488 D-102	NOV02,2016	0.0644	0.3075	0.2122	0.2473	0.1255	0.1558
PLT493 D-102	NOV02,2016	0.0646	0.3075	0.2122	0.2473	0.1021	0.1296
Toluene Blank	NOV03,2016	0.0356	0.2795	0.1861	0.2208	0.1138	0.1427
J24 C-201	NOV03,2016	0.0472	0.2935	0.1991	0.2340	0.1724	0.2080
J80-D106	NOV07,2016	0.0905	0.3635	0.2642	0.3003	0.0669	0.0904
PLT488 D-102	NOV07,2016	0.0016	0.2375	0.1471	0.1810	0.0669	0.0904
PLT493 D-102	NOV07,2016	0.0039	0.2375	0.1471	0.1810	0.0669	0.0904
J24 C-201	NOV07,2016	0.000	0.2375	0.1471	0.1810	0.0786	0.1035
TK-1144-A310	NOV11,2016	0.0073	0.2515	0.1601	0.1943	0.0786	0.1035
PLT488 D-102	NOV15,2016	0.0155	0.2515	0.1601	0.1943	0.0903	0.1166
PLT493 D-102	NOV15,2016	0.0196	0.2655	0.1731	0.2075	0.2427	0.2864
TK-1754 VGO	NOV17,2016	0.1842	0.4476	0.3423	0.3798	0.1138	0.1427

* **A** :UOP991 January 26, 2016; **B**: UOP991 April25,2016; **C**: UOP991 May12, 2016; **D**:HR June10,2016 and **E**: HR October 29,2016.

Tables 5.11 and 5.12 are showing reproduced data for two in-house prepared chloride QC audit samples; 0.6324-ppm and 0.5-ppm. Both of these QC audit samples were prepared during the year 2017 and their peak areas obtained by using the standard UOP991 were used to reproduce the chloride QC audit concentration levels by using both UOP991 and HR models that were built in the year 2016.

Data shown in tables 5.11 and 5.12 prove that these 2016 calibrated models both low range as per UOP991 and the modified ones HR models are still capable to perform well within the preset equivalence condition of (E= ±0.3-ppm) difference.

Table 5.11: Chloride (0.6324-ppm) QC audit tests results reproduced by UOP991 calibrated models.

Sample No.	Sample Testing Date	Measured Values for Cl=0.6324 ppm	UOP991 January 26	UOP991 April 25	UOP991 May 12
QC Chloride	NOV17,2016	0.524	0.7698	0.6417	0.6846
QC Chloride	NOV20,2016	0.5114	0.7418	0.6156	0.6581
QC Chloride	NOV21,2016	0.5218	0.7558	0.6287	0.6713
QC Chloride	NOV22,2016	0.4979	0.7278	0.6026	0.6448
QC Chloride	NOV23,2016	0.5161	0.7558	0.6287	0.6713

Table 5.12: Chloride (0.5-ppm) in-house prepared QC audit tests results obtained during the year 2017 by using UOP991 method and their measured values were reproduced by 2016 HR calibration models.

Chloride QC Audit Sample 0.5-ppm Run No.	HR Method June10,2016	HR Method October 29,2016
1	0.5591	0.6392
2	0.5473	0.6262
3	0.5239	0.6000
4	0.5708	0.6523
5	0.5239	0.6000
6	0.5005	0.5739
7	0.5005	0.5739
8	0.4653	0.5347
9	0.5005	0.5739
10	0.5005	0.5739
11	0.5122	0.5870
12	0.4888	0.5608
13	0.5005	0.5739
14	0.4770	0.5478
15	0.5005	0.5739

5.4 Quadratic vs. Linear Calibration Model for Fluoride Measurand

According to the original reference testing procedure UOP991[1], the measuring equipment has to be calibrated using the quadratic function – equation 5.1- for the fluoride measurand. Typically, quadratic calibration function is used when the measuring equipment response does not follow a linear model over the working range, i.e. there is a curvature in the relationship between the response and predictor.

$$Y = \beta_2 X^2 + \beta_1 X + \beta_o + \varepsilon \quad (5.1)$$

Where;
Y: is the measuring equipment response variable.
X: is the independent variable.
β_2 &β_1: are calibration function parameters.
β_o : is the intercept.
ε: the error term.

Quadratic calibration model is second order model often approximated in the shape of a ∪ or an inverted ∩. Solving quadratic function and determining the intercept points, is not as easy as the case of the linear one. Dealing with quadratic functions in calibrating lab measuring equipment are not favored by analytical chemists due to the following:

1. They may require more calibration standards, i.e. regressors to capture the region of curvature of the calibration curve. The more calibration standards needed, then the higher will be cost of calibration.
2. Defining the limit of detection and limit of quantification (LOD/LOQ) becomes more complicated.

3. The effect of outlier presence in the calibration data has stronger negative impact than in the case of linear calibration.
4. The correction for bias is more complicated than the linear model.
5. The uncertainty analysis is difficult.

Because of these five obstacles points associated with quadratic calibration, straight line –*linear regression*- is by far the most popular model, and it described mathematically as:

$$Y = \beta_1 X + \beta_o + \varepsilon \qquad (5.2)$$

To use the linear calibration instead of the quadratic one, it has to be proved that the analytical range of the calibration is linear in nature. This is approved by using a simple- but useful- mathematical tool called Two Points Linear Interpolation (TPLI) as illustrated in section 5.4.1.

5.4.1 Two Points Linear Interpolation (TPLI)

Fluoride (F) is one of the three measurands covered by the reference testing procedure UOP991 [1]. According to UOP991, the quadratic function is supposed to be used for building the calibration model of Fluoride and reason why UOP991 recommends to use quadratic formula could be attributed to the dissociation mechanism of HF acid which is formed during the analytical process. So the formation and dissociation mechanism; hence *chemical kinetics* dictates the type of calibration function to be used. But on the other hands, calibrating analytical method by using quadratic and

higher-order polynomials functions are sometimes not preferred by analytical scientists and reasons aforementioned.

So in order to overcome these difficulties associated with modeling the calibration model by using quadratic function, the linear one was used in this thesis work. Three ways were used to prove that using linear calibration function instead of the quadratic function gives values that are not significantly different. These approaches are; checking the linearity by two points linear interpolation (TPLI), by visual illustration of the quadratic calibration curve versus the linear one and by testing for equivalency by two one-sided test as per ASTM E2935[66].

Using TPLI for estimating the measuring equipment responses values of Fluoride and comparing them with the actual observed values from the measuring equipment. Table 5.13 shows the measuring equipment responses values versus the ones obtained by TPLI for some selected UOP991 models and similarly table 5.14 shows the modified HR methods. These tables show clearly that there is no significant difference between Fluoride response values obtained by quadratic function and the linear function for both the reference testing method UOP991 and the modified one HR method.

Table 5.13: Reproduced fluoride responses in peak area (µS/cm) for some selected UOP991 quadratic calibration models.

Fluoride Standards Set No.	**UOP991 April 25 Model**			
	Level No.	Concentration of Fluoride (ppm)	Measured Responses (µS/cm)	Estimated Responses by TPLI
1	1	0.10346	0.0052	
	2	0.25804	0.0067	0.006923
	3	0.49641	0.0107	0.009579
	4	0.93796	0.0145	
Fluoride Standards Set No.	**UOP991 May12 Model**			
	Level No.	Concentration of Fluoride (ppm)	Measured Responses (µS/cm)	Estimated Responses by TPLI
2	1	0.11095	0.0041	
	2	0.27024	0.005825	0.005929
	3	0.5238	0.00893	0.008839
	4	0.98732	0.01416	
Fluoride Standards Set No.	**UOP991 June 1 Model**			
	Level No.	Concentration of Fluoride (ppm)	Measured Responses (µS/cm)	Estimated Responses by TPLI
3	1	0.0979	0.00265	
	2	0.18751	0.0031	0.003334
	3	0.4717	0.00553	0.005505
	4	0.92931	0.009	
Fluoride Standards Set No.	**UOP991 June 9 Model**			
	Level No.	Concentration of Fluoride (ppm)	Measured Responses (µS/cm)	Estimated Responses by TPLI
4	1	0.10002	0.003375	
	2	0.21384	0.0044	0.004676
	3	0.31048	0.007375	0.005781
	4	0.41258	0.0077	0.006949
	5	0.5133855	0.0087	0.008101
	6	0.61279	0.01	0.009238
	7	0.71203	0.01085	0.010373
	8	0.80379	0.012125	0.011422
	9	0.87777	0.013025	0.012268
	10	0.97681	0.0134	
Fluoride Standards Set No.	**UOP991 Sep11 Model**			
	Level No.	Concentration of Fluoride (ppm)	Measured Responses (µS/cm)	Estimated Responses by TPLI
5	1	0.107142	0.009133	
	2	0.25889	0.0105	0.010896
	3	0.50446	0.01366	0.01375
	4	0.95329	0.018966	

Fluoride Standards Set No.	UOP991 Oct29 Model			
6	Level No.	Concentration of Fluoride (ppm)	Measured Responses (µS/cm)	Estimated Responses by TPLI
	1	0.1105806	0.003266	
	2	0.269147	0.004566	0.005107
	3	0.516939	0.00766	0.007985
	4	0.991847	0.0135	

Table 5.14: Reproduced fluoride responses in peak area (µS/cm) for some selected HR quadratic calibration models.

Fluoride Standards Set No.	HR April 28 Model			
1	Level No.	Concentration of Fluoride (ppm)	Measured Responses (µS/cm)	Estimated Responses by TPLI
	1	0.8536	0.01236	
	2	1.6108	0.02188	0.0236
	3	2.2058	0.0298	0.0325
	4	2.7496	0.03364	0.0405
	5	4.0989	0.0546	0.0606
	6	4.6751	0.0775	0.0692
	7	5.4221	0.07792	0.0803
	8	6.0262	0.0857	0.0892
	9	6.8784	0.10338	0.1019
	10	7.3554	0.109	
Fluoride Standards Set No.	**HR May17 Model**			
2	Level No.	Concentration of Fluoride (ppm)	Measured Responses (µS/cm)	Estimated Responses by TPLI
	1	1.002388	0.0148	
	2	2.1905541	0.0334	0.032282793
	3	5.1254579	0.0713	0.086385767
	4	10.98601	0.1617	
Fluoride Standards Set No.	**HR June 10 Model**			
3	Level No.	Concentration of Fluoride (ppm)	Measured Responses (µS/cm)	Estimated Responses by TPLI
	1	1.0104	0.013275	
	2	2.0032	0.027625	0.0297
	3	3.08	0.043575	0.0475
	4	3.7806	0.0547	0.0591
	5	5.0609	0.073775	0.0802
	6	5.9983	0.0829	0.0957
	7	7.2147	0.108475	0.1158
	8	8.2132	0.124225	0.1323
	9	9.1074	0.1394	0.1471
	10	10.58166	0.1715	

Table 5.14 cont'd: Reproduced Fluoride Responses in Peak Area (µS/cm) for some selected HR Quadratic Calibration Models.

Fluoride Standards Set No.	**HR Sep10 Model**			
	Level No.	Concentration of Fluoride (ppm)	Measured Responses (µS/cm)	Estimated Responses by TPLI
4	1	0.9153	0.015833	
	2	1.8176	0.0299	0.0302
	3	2.7746	0.044833	0.0455
	4	3.6638	0.05533	0.0597
	5	4.5951	0.066833	0.0746
	6	5.4576	0.081433	0.0884
	7	6.2834	0.0927	0.1016
	8	7.2777	0.11676	0.1175
	9	8.2792	0.12553	0.1335
	10	10.5224	0.16933	
Fluoride Standards Set No.	**HR Oct29 Model**			
	Level No.	Concentration of Fluoride (ppm)	Measured Responses (µS/cm)	Estimated Responses by TPLI
5	1	0.991847	0.0135	
	2	2.179411	0.026	0.030556
	3	3.325837	0.040533333	0.047021
	4	4.36172	0.055266667	0.061899
	5	6.3577	0.0817	0.090566
	6	8.6139417	0.118266667	0.12297
	7	10.8812098	0.1555333	

5.4.2 Modeling Via Multiple Linear Regression

Since the samples sizes used for calibrating each model were quite small and sometimes even no replicates were available, calibration via using the multiple linear regression was done on four of the modified high range (HR) models and on two of low range UOP991 calibration models. The dummy variables used for coding are as follows;

Measurand/Regressor	D1	D2
Fluoride (F) -Base Level	0	0
Chloride (Cl)	1	0
Bromide (Br)	0	1

Since we've three measurands, then the number of dummy variables will be (3-1) = 2. These dummy variables are the coded values of the measurands. The Fluoride is selected arbitrarily to be the base level.

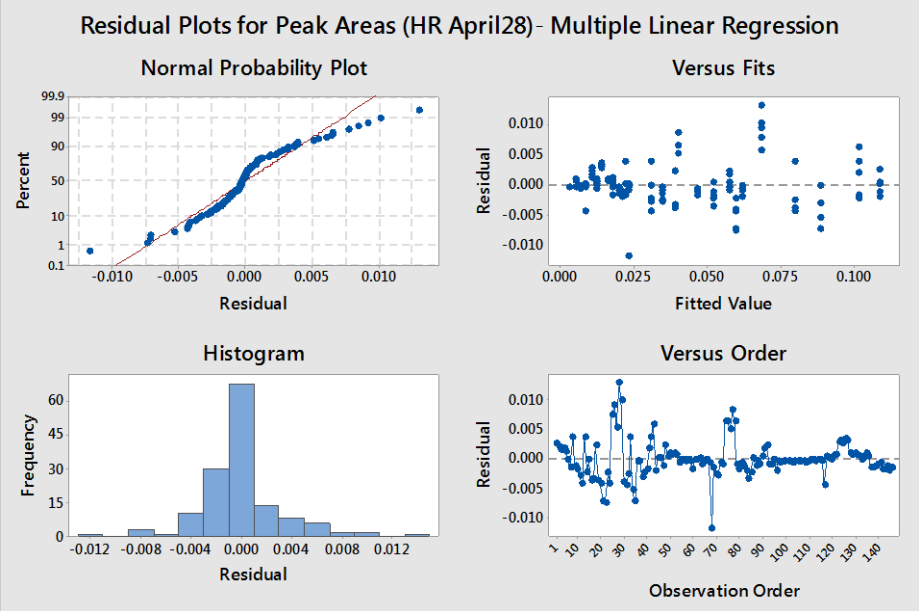


Figure 5.19. Residual Plots for HR April28 by using multiple linear regression.

Table 5.15: Multiple Linear Regression Output for HR April28 Calibration Model.

Model Summary: HR April28

S	R-sq	R-sq(adj)	R-sq(pred)
0.0032238	98.88%	98.85%	98.82%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-0.002517	0.000719	-3.50	0.001	
PPM	0.015167	0.000167	90.69	0.000	1.86
PPM*D1	-0.006494	0.000137	-47.35	0.000	1.51
PPM*D2	-0.012384	0.000269	-45.95	0.000	5.83

Regression Equations:

Peak Area (PA)= -0.002517+0.015167PPM-0.006494 PPM*D1 - 0.012384 PPM*D2

From table 5.15, the model will be;

Peak Area (PA) = -0.002517+0.015167PPM-0.006494 PPM*D1 - 0.012384 PPM*D2 where the response Y is the peak area, PPM is denoting slope for Fluoride, PPM*D1 is denoting the addition to the slope for Chloride and PPM*D2 is denoting the addition to the slope for Bromide, so the regression function for each measurand can be written as:

$$Y = -0.002517 + 0.01516\ F - ppm$$

$$Y = -0.002517 + 0.008673\ Cl - ppm$$

and $Y = -0.002517 + 0.002783\ Br - ppm.$

Peak area Y is expressed in uS/cm and measurand concentration in ppm.

β_o represents the mean response for the base level (Fluoride). The estimated betas are:

$$\hat{\beta}_o = -0.002517,\quad \hat{\beta}_1 = 0.015167\ ,\quad \hat{\beta}_2 = 0.008673\ ,\quad \hat{\beta}_3 = 0.002783$$

$\hat{\beta}_1$ represents the slope of the Fluoride regression line, $\hat{\beta}_2$ represents the difference in slope between the Chloride regression line and the Fluoride regression line and $\hat{\beta}_3$ represents the difference in slope between the Bromide regression line and the Fluoride regression line.

From the MINITAB output shown in table 5.15, the *p*-values are all (0.000) and are less than α=0.05, so there is a significant difference between the responses obtained from these three measurands (F, Cl and Br). Also note that the variance inflation factors (VIF) are less than 5 for Fluoride and Chloride and 5.83 for Bromide. The whole model

$R^2_{adj} = 98.85\%$ which is quite good while when bromide was modeled by OLS – see table 4.4, the obtained $R^2_{adj} = 95.71\%$.

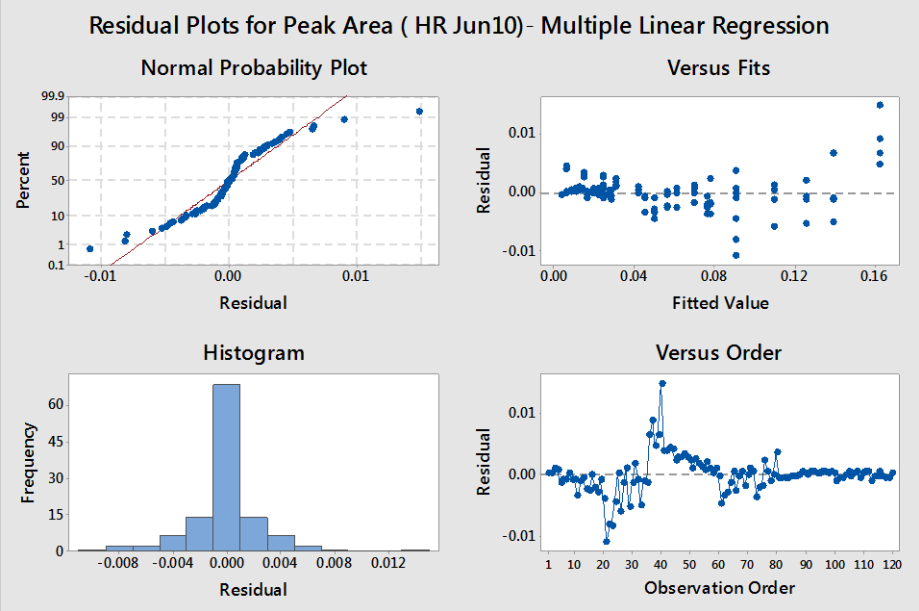


Figure 5.20. Residual Plots for HR June10 by using multiple linear regression

Table 5.16: Multiple Linear Regression Output for HR June10 Calibration Model.

Model Summary: HR June10

S	R-sq	R-sq(adj)	R-sq(pred)
0.0028036	99.59%	99.57%	99.54%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-0.006466	0.000937	-6.90	0.000	
PPM	0.016130	0.000147	109.57	0.000	3.16
PPM*D1	-0.007596	0.000210	-36.23	0.000	6.53
PPM*D2	-0.013795	0.000200	-68.82	0.000	7.19

Regression Equations:

D2	D1			
0	0	Peak Area	=	-0.006466 + 0.016130 PPM - 0.007596 PPM*D1 - 0.013795 PPM*D2
0	1	Peak Area	=	0.000029 + 0.016130 PPM - 0.007596 PPM*D1 - 0.013795 PPM*D2
1	0	Peak Area	=	0.000879 + 0.016130 PPM - 0.007596 PPM*D1 - 0.013795 PPM*D2

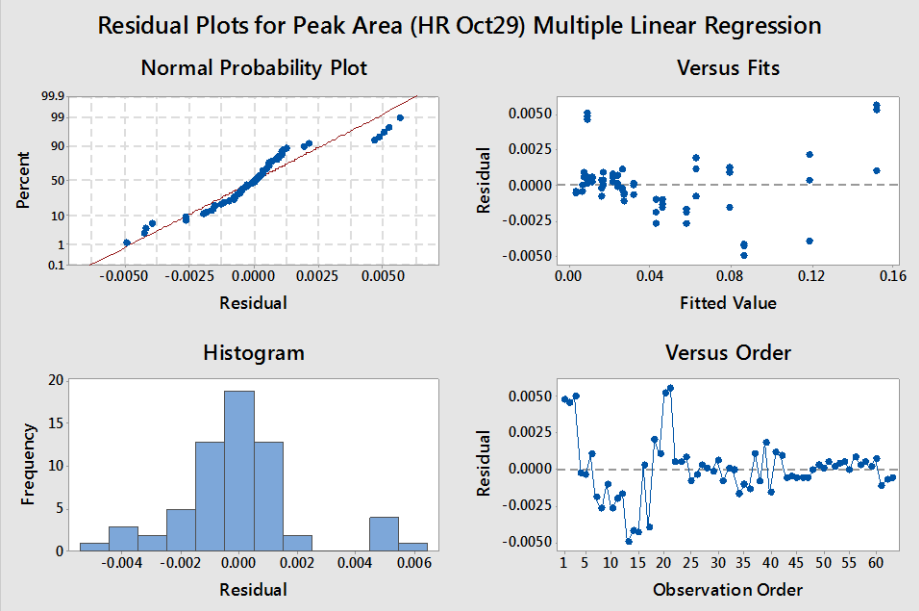


Figure 5.21. Residual Plots for HR Oct29 by using multiple linear regression

Table 5.17: Multiple Linear Regression Output for HR Oct29 Calibration Model.

Model Summary: HR Oct29

S	R-sq	R-sq(adj)	R-sq(pred)
0.0021518	99.72%	99.70%	99.65%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-0.005659	0.000881	-6.42	0.000	
PPM	0.014448	0.000142	101.59	0.000	2.82
PPM*D1	-0.006703	0.000207	-32.45	0.000	5.10
PPM*D2	-0.011934	0.000206	-58.02	0.000	5.14

Regression Equations:

D2	D1			
0	0	Peak Area	=	-0.005659 + 0.014448 PPM - 0.006703 PPM*D1 - 0.011934 PPM*D2
0	1	Peak Area	=	-0.000731 + 0.014448 PPM - 0.006703 PPM*D1 - 0.011934 PPM*D2
1	0	Peak Area	=	0.000767 + 0.014448 PPM - 0.006703 PPM*D1 - 0.011934 PPM*D2

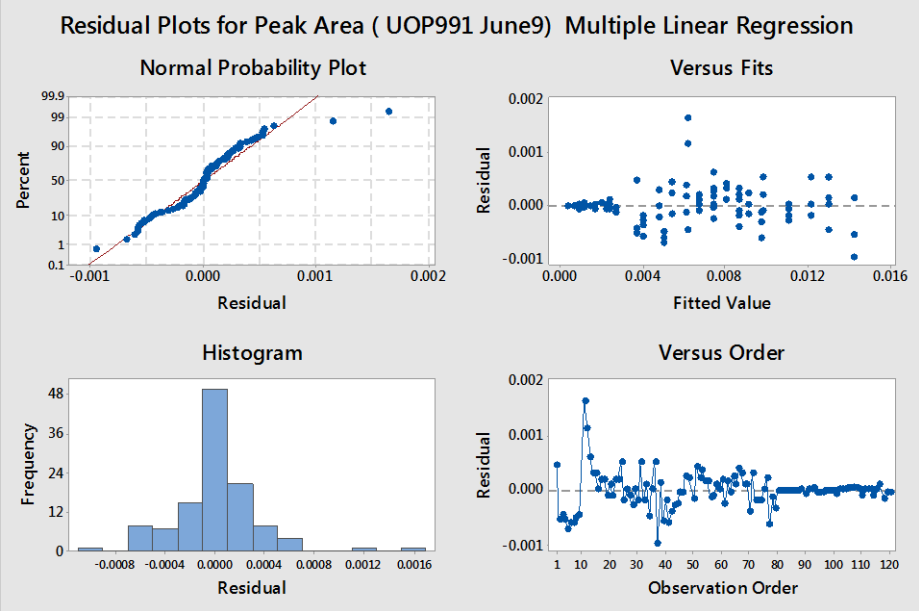


Figure 5.22. Residual Plots for UOP991 June9 by using multiple linear regression

Table 5.18: Multiple Linear Regression Output for UOP991 June9 Calibration Model

Model Summary:UOP991 June 9

S	R-sq	R-sq(adj)	R-sq(pred)
0.0003396	99.27%	99.24%	99.17%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	0.002425	0.000123	19.72	0.000	
PPM	0.012013	0.000200	59.92	0.000	3.25
PPM*D1	-0.005471	0.000282	-19.40	0.000	7.29
PPM*D2	-0.009572	0.000271	-35.34	0.000	7.97

Regression Equations:

D1	D2			
0	0	Peak Area	=	0.002425 + 0.012013 PPM - 0.005471 PPM*D1 - 0.009572 PPM*D2
0	1	Peak Area	=	0.000041 + 0.012013 PPM - 0.005471 PPM*D1 - 0.009572 PPM*D2
1	0	Peak Area	=	0.003324 + 0.012013 PPM - 0.005471 PPM*D1 - 0.009572 PPM*D2

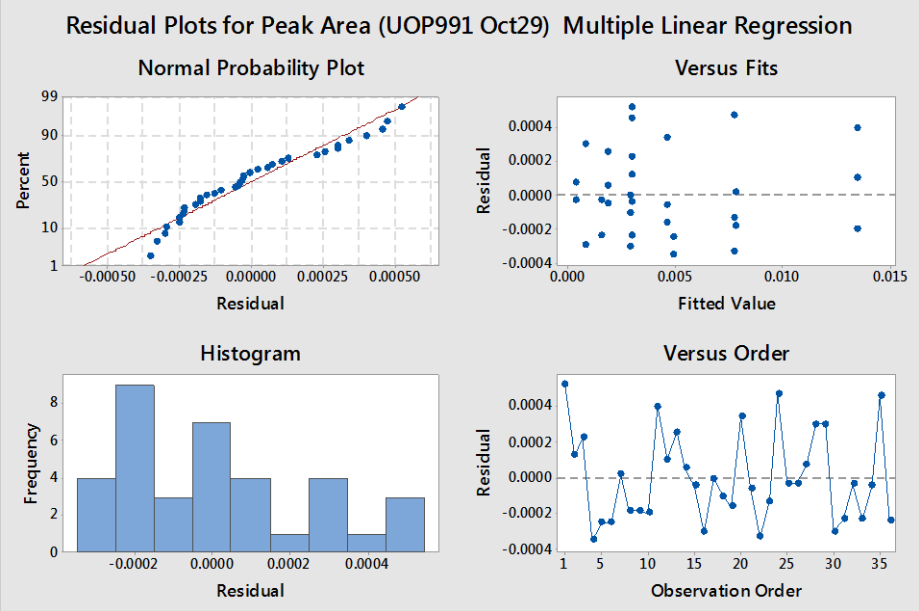


Figure 5.23. Residual Plots for UOP991 Oct29 by using multiple linear regression

Table 5.19: Multiple Linear Regression Output for UOP991 Oct29 Calibration Model

Model Summary:UOP991 Oct29

S	R-sq	R-sq(adj)	R-sq(pred)
0.0002702	99.53%	99.45%	99.26%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	0.001666	0.000135	12.32	0.000	
PPM	0.011827	0.000234	50.52	0.000	2.82
PPM*D1	-0.004793	0.000340	-14.10	0.000	4.45
PPM*D2	-0.008728	0.000339	-25.78	0.000	4.48

Regression Equations:

D1	D2			
0	0	Peak Area	=	0.001666 + 0.011827 PPM - 0.004793 PPM*D1 - 0.008728 PPM*D2
0	1	Peak Area	=	-0.000001 + 0.011827 PPM - 0.004793 PPM*D1 - 0.008728 PPM*D2
1	0	Peak Area	=	0.001106 + 0.011827 PPM - 0.004793 PPM*D1 - 0.008728 PPM*D2

From table 5.19, the regression models for the measurand Fluoride (F) , Chloride (Cl) and Bromide (Br) are:

$$Y = -0.005659 + 0.014448\ F - ppm$$

$$Y = -0.000731 + 0.007745\ Cl - ppm$$

$$Y = 0.000767 + 0.002514\ Br - ppm$$

Table 5.20 shows the performance of both the high range models calibrated according to the separate but simple linear regressions and the one calibrated according to the Multiple Linear Regression (MLR) in reproducing the measurand concentration values that were obtained from the chemical preparation of these calibration standards (CAL STDs). Although no significant improvement can be seen, yet the motivation behind doing MLR was to overcome the deficiency in having small sample sizes. Also, values obtained from the MLR double assure the findings obtained by using calibration models that were built based on simple linear regression.

Table 5.20: Comparisons for concentration levels of F, Cl and Br prepared in-house versus their estimated concentration levels values estimated from both the HR models developed by OLS and MLR.

UOP991 OCT29 CAL STDs	**Reproduced by HR OCT29 (OLS)**	**Reproduced by HR OCT29 (MLR)**
F-ppm Prepared	**F-ppm** Estimated	**F-ppm** Estimated
0.110581	0.707779623	0.70771041
0.269147	0.92192691	0.921857697
0.516939	1.326135105	1.326065891
0.991847	0.617801772	0.617732558
UOP991 OCT29 CAL STDs	**Reproduced by HR OCT29 (OLS)**	**Reproduced by HR OCT29 (MLR)**
Cl-ppm Prepared	**Cl-ppm** Estimated	**Cl-ppm** Estimated
0.10496237	0.343963848	0.343963848
0.25547255	0.451517108	0.451517108
0.490675	0.688315042	0.688315042
0.941455	1.092834087	1.092834087
UOP991 OCT29 CAL STDs	**Reproduced by HR OCT29 (OLS)**	**Reproduced by HR OCT29 (MLR)**
Br-ppm Prepared	**Br-ppm** Estimated	**Br-ppm** Estimated
0.1058488	-0.172564612	-0.172633254
0.25763	0.052882704	0.052903739
0.494819	0.238170974	0.238265712
0.94940616	0.887872763	0.888225935

5.4.3 Modeling Fluoride Response Via; Linear, Quadratic and Box-Cox Transformation

According to the original testing procedure UOP991, the calibration function of Fluoride (F) has to be quadratic and this is due to the non-linearity behavior of fluoride-measured responses. In this section, three regression functions for modeling fluoride were built; Linear (L), Quadratic (Q) and Transformed (T) one via Box-Cox Transformation. Three fluoride high range (HR) calibration

models and two fluoride UOP991 calibration models were considered as seen in tables 5.15 and 5.16. Cross validation of these five fluoride models was done by using the UOP991 prepared measured fluoride peak areas and reproduce them by using the modified HR fluoride calibration methods and vice versa. From figures 5.19-5.21, it is shown clearly that there is no significant difference between the measured fluoride peak areas obtained by using the standard low range method UOP991 versus the reproduced ones from the modified HR calibrated methods. Also fluoride peak areas obtained from the Linearly modeled calibration functions PA(L) are not significantly different from the ones obtained from the Quadratic calibration models PA(Q).

Table 5.17 shows that the three calibration models (i.e. linear, quadratic and transformed one) of the HR April28 are capable to predict Fluoride concentrations levels as low as ≈0.6-ppm. While table 5.18 shows that the prediction power of the original testing procedure UOP991 Oct29 is capable to predict Fluoride concentration as high as ≈2-pmm.

Table 5.21: UOP991 Fluoride Calibration Functions Models; linear, quadratic and Box-Cox transformation calibration functions. Fluoride (F) concentration in ppm and PA is the Peak Area (response variable) for Fluoride.

UOP991 June 9 Calibration Models			
Calibration Function Type	**Calibration Function**	$\mathbf{R^2_{adj}}$	**Standard Deviation**
Linear	PA=0.002683 + 0.011588 F	95.63%	0.0006986
Quadratic	PA = 0.001767 + 0.01620 F - 0.004266 F^2	96.33%	0.0006404
Box-Cox Transformation (λ=0.5)	$(PA)^{0.5}$=0.05763 + 0.06484 F	92.76%	0.0051080
UOP991 Oct29 Calibration Models			
Calibration Function Type	**Calibration Function**	$\mathbf{R^2_{adj}}$	**Standard Deviation**
Linear	PA = 0.001666 + 0.01183 F	99.49%	0.0002945
Quadratic	PA = 0.002084 + 0.009480 F + 0.002067 F^2	99.65%	0.0002428
Box-Cox Transformation (λ=0.5)	$(PA)^{0.5}$=0.050287 + 0.06741 F	99.37%	0.0018657

Table 5.22: HR Fluoride Calibration Functions Models; linear, quadratic and Box-Cox transformation calibration functions. Fluoride (F) concentration in ppm and PA is the Peak Area (response variable) for Fluoride.

HR April28 Calibration Models			
Calibration Function Type	**Calibration Function**	**R^2_{adj}**	**Standard Deviation**
Linear	PA = - 0.002471 + 0.01516 F	98.19%	0.0045234
Quadratic	PA = - 0.001303 + 0.01437 F + 0.000095 F^2	98.16%	0.0045584
Box-Cox Transformation (λ=0.5)	$(PA)^{0.5}$ =0.09621 + 0.033336 F	97.37%	0.0120412
HR June10 Calibration Models			
Calibration Function Type	**Calibration Function**	**R^2_{adj}**	**Standard Deviation**
Linear	PA = -0.00647 + 0.016130 F	99.19%	0.0044401
Quadratic	PA = 0.002540 + 0.01178 F + 0.000380 F^2	99.59%	0.0031680
Box-Cox Transformation (λ=0.5)	$(PA)^{0.5}$ =0.10827 + 0.029795 F	98.46%	0.0113762
HR Oct29 Calibration Models			
Calibration Function Type	**Calibration Function**	**R^2_{adj}**	**Standard Deviation**
Linear	PA = - 0.005659 + 0.01445 F	99.48%	0.0035458
Quadratic	PA = 0.002201 + 0.01053 F + 0.000330 F^2	99.89%	0.0015931
Box-Cox Transformation (λ=0.5)	$(PA)^{0.5}$ = 0.10272 + 0.027753 F	99.02%	0.0093564

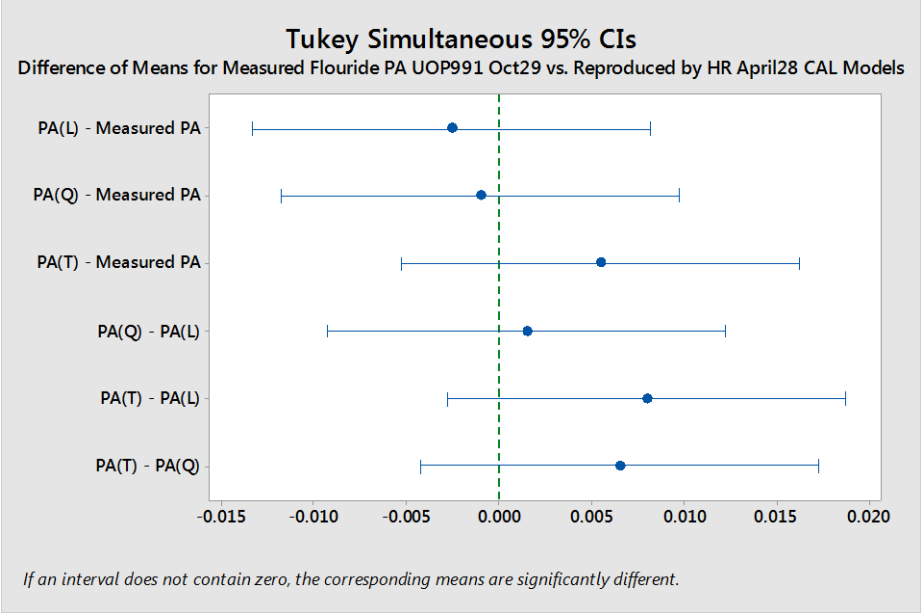


Figure 5.24: Tukey's multiple comparison for fluoride peaks areas obtained from differently calibrated models.

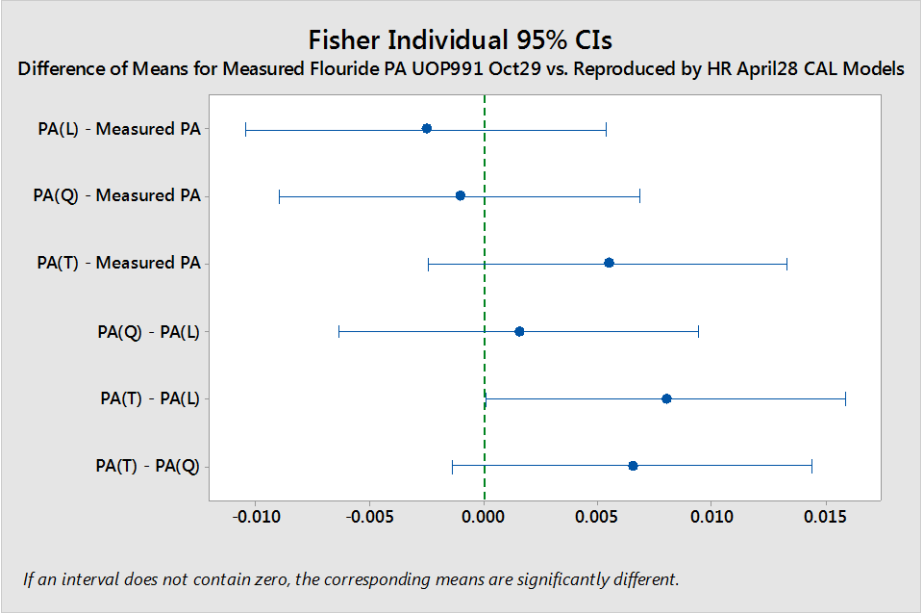


Figure 5.25: Fisher LSD for fluoride peaks areas reproduced from differently calibrated models.

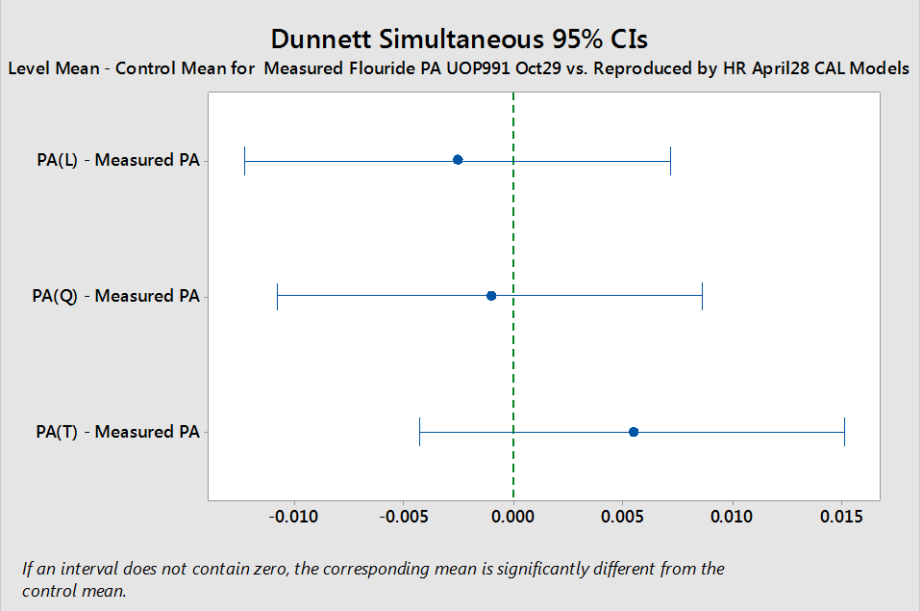


Figure 5.26: Dunnett multiple comparisons. The control group (Measured PA) represents the measured fluoride peak areas obtained from method calibrated as per UOP991 Oct29.

Table 5.23: Reproducing Fluoride UOP991 June9 Calibration Standards (CAL STDs) by Using HRApril28 Calibration Models; Linear, Quadratic and Transformed

UOP991 June 9 CALSTDS		**Linear**	**Quadratic**	**Transformed (λ=0.5)**
		PA= -0.002471+0.01516 F	**PA= -0.001303+ 0.014378 F+ 0.000095 F^2**	**$(PA)^{0.5}$ = 0.09621+ 0.033336 F**
F-ppm	**Measured PA** (uS/cm)	**Estimated PA** (uS/cm)	**Estimated PA** (uS/cm)	**Estimated PA** (uS/cm)
0.10002	0.003375	-0.000954697	0.000155766	0.009909061
0.21384	0.0044	0.000770814	0.00186611	0.01067886
0.31048	0.007375	0.002235877	0.003360336	0.011355068
0.41258	0.0077	0.003783713	0.004980925	0.01209203
0.51338	0.0087	0.005311924	0.006623245	0.012842379
0.61279	0.01	0.006818896	0.008283877	0.013604415
0.71203	0.01085	0.008323375	0.009982509	0.014387099
0.80379	0.012125	0.009714456	0.011589337	0.015130265
0.87777	0.013025	0.010835993	0.01291016	0.015743057
0.97681	0.0134	0.01233744	0.014713818	0.01658247

Table 5.24: Reproducing Fluoride HR April28 Calibration Standards (CAL STDs) by Using UOP991 Oct29 Calibration Models; Linear, Quadratic and Transformed

HR April28 CAL STDs		**Linear**	**Quadratic**	**Transformed (λ=0.5)**
		PA= 0.001666 + 0.01183F	**PA= 0.002084 + 00948F + 0.002067F^2**	**(PA)$^{0.5}$ = 0.050287 + 0.06741 F**
F-ppm	**Measured PA** (uS/cm)	**Estimated PA** (uS/cm)	**Estimated PA** (uS/cm)	**Estimated PA** (uS/cm)
1.010	0.013275	0.013619032	0.013772809	0.014018102
2.003	0.027625	0.025363856	0.029368815	0.034344508
3.08	0.043575	0.0381024	0.050890789	0.066517465
3.781	0.0547	0.046390498	0.067467587	0.093108739
5.061	0.073775	0.061536447	0.103002801	0.15322705
5.998	0.0829	0.072625889	0.133317723	0.206690622
7.2147	0.108475	0.087015901	0.178070625	0.287971679
8.2132	0.124225	0.098828156	0.21937804	0.364742089
9.1074	0.1394	0.109406542	0.259868919	0.441184003
10.581	0.1715	0.126847038	0.333843286	0.583079921

5.4.4 Visual Illustration of the Calibration Models Curvatures

Visual illustration of the curvature of the Fluoride calibration curve obtained by the standard reference testing method UOP991 (quadratic calibration) in comparison with the curvature obtained by the modified method HR in which the linear calibration is used. The suitability of using the linear regression function over the quadratic function for modeling the Fluoride responses was checked by considering the calibration functions obtained for Fluoride from both UOP991 Linear and Quadratic and similarly from the HR calibration models of Fluoride Linear and Quadratic functions as shown in table 5.15. These functions – only data set of October 29 was considered- are summarized in the table 4.4.2.1 and plotted in figurers numbers 4.4.2.1 and 4.4.2.2. The two graphs show clearly that of the range of 0.1-ppm to 1-ppm, the calibration curve is quite linear both at standard range covered by UOP991 and at high range.

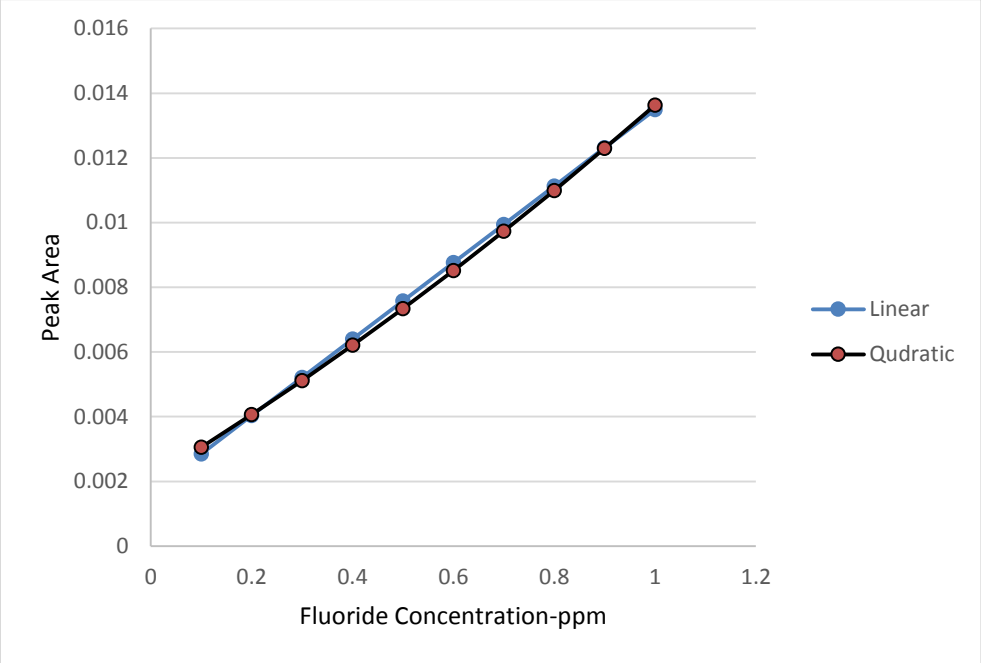


Figure 5.27: Overlay Plot of UOP991 Oct29, 2016 calibration functions (linear vs. quadratic) for fluoride

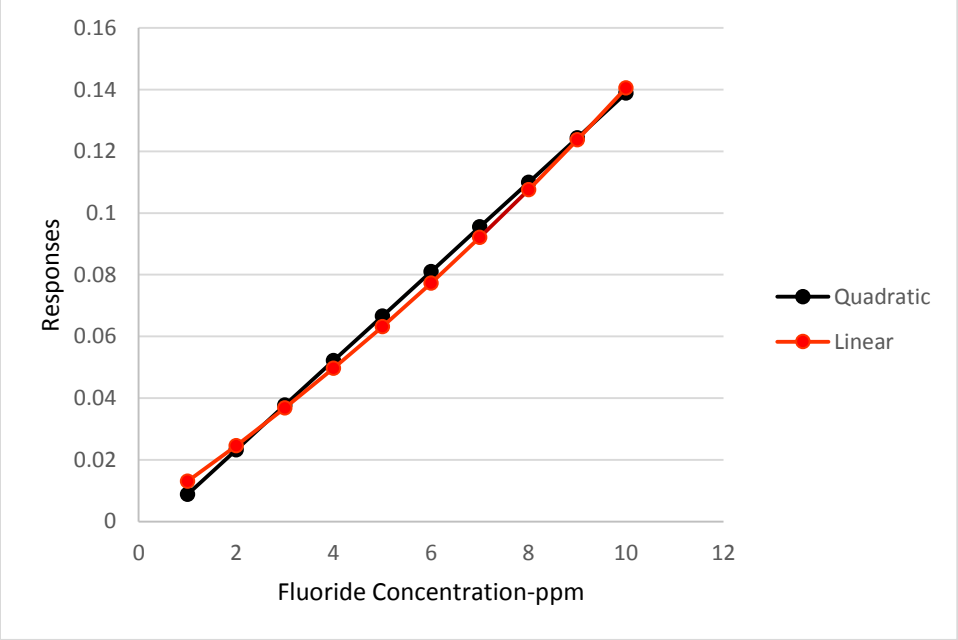


Figure 5.28: Overlay Plot of HR Oct29, 2016 calibration equations (linear vs. quadratic) for fluoride

5.4.5 Testing for Equivalence by Two One-Sided "TOST" (ASTM E2935)

Using the statistical technique TOST requires having replicates of data for each sample. Data used for calibration most of them were not having replicates, while the Chloride QC Audit Samples were reproduced by both the calibration models of the UOP991 and the calibration models of the modified testing procedure referred to High Range (HR) method. The value of the Equivalence (E) between the two methods (i.e. UOP991 vs. HR method) was chosen to be 0.3-ppm, which means that the maximum acceptable difference between the generated test results by these two methods for the same sample should be within ±0.3-ppm.

ASTM E2935 definition for E is "equivalence, condition that two population parameters differ by no more than predetermined limits". There are no general guidelines for setting the value of equivalence "E" and its value varies from one case to another. But the ISO standard [15] does mention a statement about the Acceptability Limit (AL) which is:

"The data generated in some parts of the validation study are evaluated using the so-called Acceptability Limits (AL) and no statistical evaluation of the data is conducted. These AL are based on experts' opinion and data generated in existing validation studies" ISO 16140-2:2016(E)

5.4.4.1 Brief About ASTM E2935

One of the applications of this SOP [66] is for checking the equivalence between the means of two sets of data; one set of data from a modified testing process - which is in this thesis work is the HR method- and the other set of data is from the original testing process (i.e. like UOP991). Two principle types of equivalence are covered by this ASTM E2935; means equivalence and non-inferiority.

Three types of Two One-Sided Tests (TOSTs) covered by ASTM E2935 and they are;

1. TOST procedure for statistical analysis of means equivalence - two independent samples design- which is the case of this thesis work (i.e. HR method vs. UOP991 method). This approach was conducted by using the Chloride Quality Control Audit Samples at three different concentration levels as detailed in section 5.3.4.2.
2. TOST procedure for statistical analysis of means equivalence - paired samples design. This is not applicable to our work.
3. TOST procedure for statistical analysis of *bias*. This can be used only whenever an Accepted Reference Value (ARV) is available. For this thesis work, no organic based ARV chemical was available, so this approach was not tried by this thesis.

ASTM E2935 is recommending using the TOST as an alternative for using the conventional t-Test. Section X1.3 of ASTM E2935 is titled "Criticism of the Use of the Conventional t Test for Equivalence Testing" is explaining the deficiency of the t-Test with respect to both consumer's risk and the impact of the populations means variances in masking the critical differences between the two samples means.

5.4.4.2 Case Study of Using ASTM E2935 TOST for Two-Independent Samples Design: Chloride Audit Samples

Tables numbers 5.16 -5.17 show that the obtained values for LCL and UCL are less than the chosen E= ±0.3 which proves that data generated by these two methods; original method UOP991 and the modified one HR method are equivalent.

Table 5.25. Two one-sided test for chloride QC audit sample (Cl=0.11829-ppm).

Chloride QC Audit Samples Concentration Level-ppm	**Measured Chloride Peak Areas** (ppm) **by UOP991 Oct29**	**Measured Chloride Peak Areas** (ppm) **by HR Oct29**
Cl=0.11829-ppm	0.183963605	0.404260813
	0.226613591	0.442995481
	0.212396929	0.430083925
	0.155530281	0.378437702
	0.183963605	0.404260813
	0.183963605	0.404260813
	UCL=0.244657995	
	LCL=0.194631316	

Table 5.26. Two one-sided test for chloride QC audit samples (0.5327 and 1.0743 ppm).

Chloride QC Audit Samples	**Measured Chloride** (ppm) **by UOP991 Oct29**	**Measured Chloride** (ppm) **by HR Oct29**	**Chloride QC Audit Samples**	**Measured Chloride** (ppm) **by UOP991 Oct29**	**Measured Chloride** (ppm) **by HR Oct29**
Chloride = 0.5327-ppm	0.837930054	0.998192382	Chloride=1.0743-ppm	0.709980097	0.88198838
	0.709980097	0.88198838		1.392379869	1.50174306
	0.781063406	0.946546159		1.278646574	1.398450613
	0.738413421	0.907811491		1.435029855	1.540477728
	0.653113449	0.830342156		1.363946545	1.475919948
	0.638896787	0.8174306		1.392379869	1.50174306
	0.624680125	0.804519045		1.264429912	1.385539057
	0.297696901	0.50755326		1.307079898	1.424273725
	0.553596815	0.739961265		1.335513221	1.450096837
	0.525163492	0.714138154		1.307079898	1.424273725
	0.596246801	0.778695933		1.32129656	1.437185281
	0.610463463	0.791607489		1.307079898	1.424273725
	0.596246801	0.778695933		1.32129656	1.437185281
	0.567813477	0.752872821		1.335513221	1.450096837
	0.525163492	0.714138154		1.17912994	1.308069722
	0.51094683	0.701226598		1.150696616	1.282246611
	0.482513506	0.675403486		1.164913278	1.295158167
	0.539380154	0.727049709		1.164913278	1.295158167
	0.482513506	0.675403486		1.093829969	1.230600387
	0.454080182	0.649580374		1.136479955	1.269335055
	0.454080182	0.649580374		1.150696616	1.282246611
	0.454080182	0.649580374		1.093829969	1.230600387
	0.411430196	0.610845707		1.17912994	1.308069722
	0.454080182	0.649580374		1.093829969	1.230600387
	0.454080182	0.649580374		1.235996588	1.359715946
	0.468296844	0.66249193		1.207563264	1.333892834
	0.43986352	0.636668819		0.894796702	1.049838606
	0.454080182	0.649580374		0.909013364	1.062750161
	0.425646858	0.623757263		0.823713392	0.985280826
	0.454080182	0.649580374		0.823713392	0.985280826
	1.17912994	1.308069722		0.80949673	0.97236927
	UCL =0.252583489			**UCL= 0.208881706**	
	LCL =0.118882405			**LCL=0.051187069**	

CHAPTER 6

FINDINGS, OBSERVATIONS AND RECOMMENDATIONS

It took about one year to build experimentally 11 regression models according to the original testing procedure UOP991 and six "HR" model for the proposed method. This chapter is summarizing the findings, observations and recommendations.

6.1. Findings

The findings of this these work can be summarized into the following points:

6.1.1. Equivalency Between the Two Methods

This thesis work utilized very fundamental statistical concepts to prove that the modified method "HR" is capable to produce test results that are not significantly different from the ones produced by the origin reference method UOP991. The term equivalence is used indicate a formal statistical demonstration of similarity between the two methods under considerations. But note that the statistical differences obtained from the two methods; UOP991 vs. HR method does not necessarily dictate that these differences are tangible. There is a difference between data being *statistically* different vs. *practically* different.

6.1.2. Extending the Calibration Intervals

The calibration interval for the measuring equipment covered by the original method can be easily extended to be monthly calibration instead of upon use. This finding has huge impact in reducing the operational cost of this highly delicate lab equipment and having an economically design of calibration practice.

6.1.3. Accuracy Improvement

Since the modified HR method proved to be fit for purpose, there is no need to dilute samples with measurand concentration falls in the range of 1-to-10 ppm for the Fluoride and chloride measurand. Avoiding the dilution step as part of sample preparation will reduce the errors associate with sample dilution. Also, it is much easier and more accurate for the operator to prepare calibration standards in higher concentration levels, i.e. using the HR method, than in lower concentration level as mentioned by the original method UOP991.

6.1.4. Developing Stable QC Audit Sample

Based on the one year data collected for the frequently tested in-house chloride QC audit sample, it is proven that chloride at concentration level around 1-ppm in naphtha sample matrix can stay stable for a period not less than one year.

6.1.5. Linear Regression vs. Quadratic Regression

UOP991 was recommending to use the quadratic function for calibrating the chemical measuring equipment for fluoride (F) determination but this recommendation was challenged experimentally and the statistical comparisons work done by this thesis proved that the measurand Fluoride calibration curve can be modeled via the linear least square method rather than the quadratic one.

6.1.6. Modeling Bromide Measurand

The measurnad Bromide was not studies extensively by this validation work. Reasons being are both; no interest into this particular chemical element and the difficulty in preparing bromides calibration standards using the chemical recommended by the original method UOP991 which is *4-Bromoacetanilide*. This chemical does not dissolve easily and get recrystallize readily. So preparing a reliable regressor using such chemical represents is not recommended. Nonetheless, some UOP991 and HR calibration models were built and the best one of them prove that the modified method (HR) is capable in reproducing bromide concentration level as low as 0.3-ppm, see table 6.1 .

Table 6.1. Reproducing UOP991 Bromide Calibration Standards (in ppm) Prepared on Jun9 by HR April28 and May17 (OLS) Calibration Models.

Prepared Bromide Calibration Standards (CAL STDs) UOP991 June9		**Reproduced Bromide-ppm by HR CAL Models**		
Level No.	**Br-ppm**	**HR April28**	**HR Ma17**	
L1	0.10906	-0.062378	-0.03394	
L2	0.23317	0.099644	0.16336	
L3	0.33854	0.180655	0.26201	
L4	0.44987	0.261666	0.36066	
L5	0.55978	0.342677	0.45931	
L6	0.66817	0.443940	0.58262	
L7	0.77638	0.585710	0.75525	
L8	0.87644	0.646468	0.82924	
L9	0.95711	0.727479	0.92789	
L10	1.06509	0.801199	1.01766	

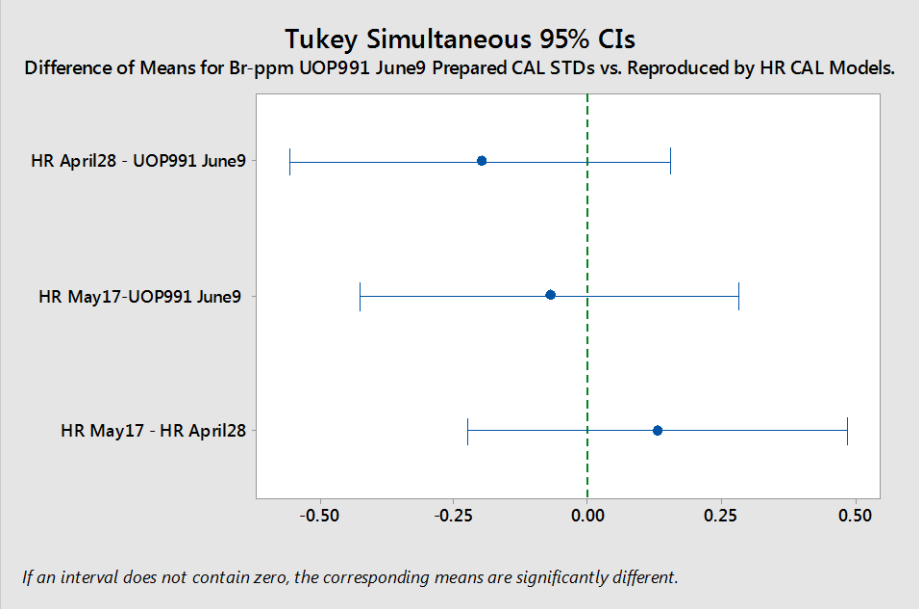


Figure 6.1. Tukey's multiple comparisons for Bromide standards reproduced by HR methods

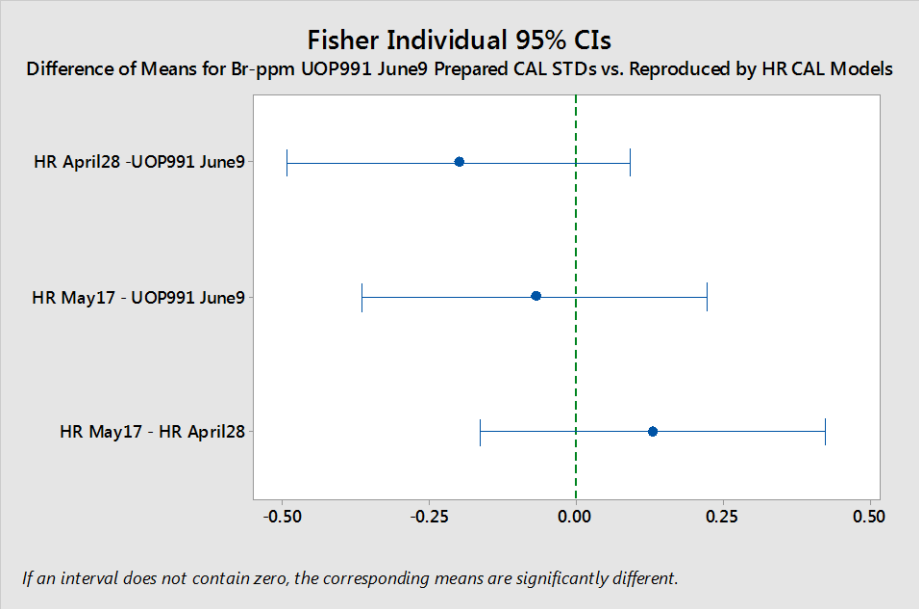


Figure 6.2. Fisher LSD for UOP991 Bromide standards reproduced by HR methods

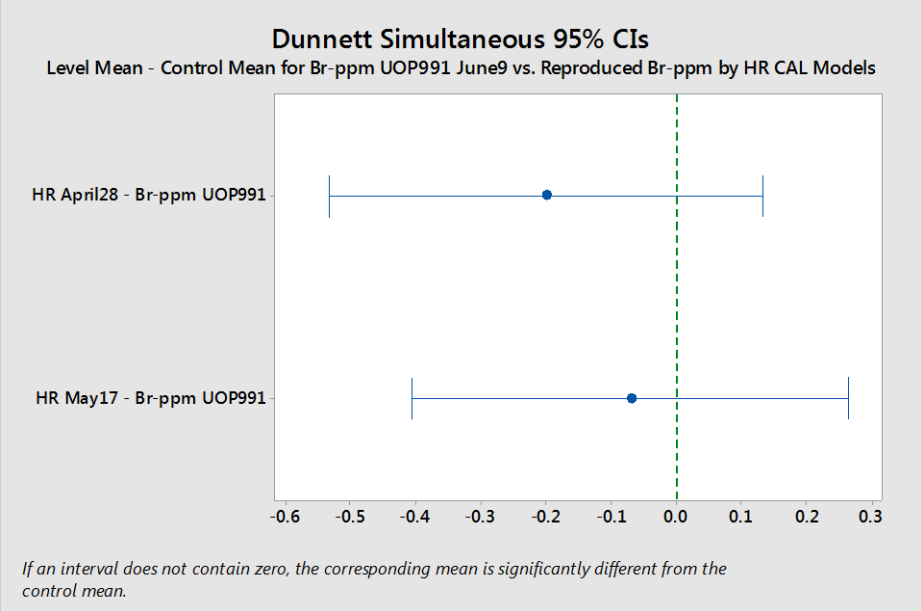


Figure 6.3. Dunnett means comparisons for UOP991 Bromide standards reproduced by HR methods

6.2. Acceptance Criteria

Method should meet performance criteria that are linked to the process specifications and requirements so when statistical inferences show a statistical significance difference between the two methods – original method UOP991 vs. HR- while the practical impact is not significant, then the performance of the method will depend on what is acceptable practically. So the acceptance criteria of any testing procedure have to do with the practical impact rather than based on the statistical one. There is no exact approach for setting the criteria of equivalency acceptance, but there is a consensus that the control requirement of a process are considered as the first criterion to be taken in consideration. A number of approaches are proposed in the literature for setting acceptance criterion for equivalence, depending on the purpose of method and usage. Some of these approaches are discussed by Marian [66].

Two one-side test (TOST) [65] is frequently used when assessing mean equivalence. This approach is based on calculating the confidence interval for the mean difference between methods, and this calculated confidence interval has to lie within $E = \pm\theta$ for mean equivalence.

Setting equivalence criteria for method equivalency could be a challenging task. The use of the two one-sided tests (TOST) is usually applied when assessing mean equivalence. This task could be made simple if the performance criteria are linked to the process requirement. By process, we mean the end-user of the produced lab data of this modified "HR" method. so for this thesis work, the equivalence criteria E was set to be: $E = \pm 0.3 - ppm$.

The international standard ASTM E2935 [65] titled "Conducting Equivalence Testing in Laboratory Applications" is used by this thesis work to check if the difference between the two methods UOP991 and HR method is within the set value of $E = \pm 0.3 - ppm$.

6.3. General Observations and Recommendations

This section is addressing general remarks about this research work, these remarks worth consideration into interpreting some data that were obtained. These observations are as follows:

1. **Life Cycle of Method Validation:** All method validation work and findings can be represented by a picture. This picture reflects the situation of particular moment of time, consequently, the validity of this picture will deteriorate with time and here it comes the importance of conducting a frequent validation during the working lifetime of a method. Recall that validation work is both time and labor intensive work, so considering economical design of method validation is necessary. Using internal quality control tool such as control charts could help in monitoring the performance of the modified method and consequently identified the operational life of the modified method and setting the frequencies for revalidation work.

2. **Sample Size**: For practical and economical reasons, validation studies can only be empirically tested by small samples sizes, so this thesis work was depending on both using small sample sizes and no replicates and this is because of; operational cost, instability the calibration standards and other operational difficulties.

3. **Normality Tests:** Normality tests such as Kolmogrov-Smirnove requires large number of data points ≥30 readings. But in real life and having this validation study being conducted at petroleum refinery QA/QC 24/7 working laboratory, it is quite difficult in generating such a large number of data points.

4. **Inverse Calibration:** The vast majority of prediction work done at any chemical analytical laboratory is based on an inverse calibration and not on the direct regression analysis. By *inverse* we mean the regression model is built and based on the measuring equipment response value –which is the dependent variable Y- , the value of the independent variable X–which is the measurnad-, is predicted. Until the end the year 2017, there is no and ISO or ASTM standard that addresses the subject of inverse calibration/regression and this is an area of standardization work that worth consideration.

5. **Process Capability:** It is a measure of inherent process variability which represents the variation that remains after all known removable assignable causes have been eliminated. The concept of process capability is mainly used in manufacturing industry, it requires the process to be stable and very large data points are needed in order to estimate the process capability indices correctly. Consequently, the concept of process capability is not

doable in case of this thesis work. But if the analytical testing process is stable, then capability analysis can be done as part of method validation, such example is the one done by A.Bouabidi et al [67].

"The single important difference between performance and capability is that for performance, there is no requirement for the process to be in statistical control or for the process to be controlled using a control chart" ISO22514-4 [68]

"Ideal process implies that the long-term standard deviation is equal to the short-term standard deviation" ISO22514-4 [68]

"A process is in control with respect to its measures of location and spread" ASTM E2281 [69]

6. **Negative Values**: Obtaining a negative concentration levels values for the measurand. This can be explained by two ways. The first one is by recalling the formula (Regression Line) used to calculate the measurand concentration (X) from the Inverse Calibration formula. $X = \frac{(Y-\beta_o)}{\beta_1}$, so when the intercept of the calibration line (β_o) is greater than the response of the measuring equipment (i.e. Y), then the calculated measurand value will be negative. Having an outlier among the response mean values $(Y_{i's})$ could contribute to push upward the regression line, consequently the higher will be the intercept value (β_o) . Moreover, having non-linear response of the measurand could also be a reason for getting negative values for the measurand concentration level , recall that UOP991 was using a quadratic function for developing the regression curve for the fluoride, while this

thesis work was using simple linear regression model for the three measurands of interest (fluoride, chloride and bormide).

Another interpretation for having a negative concentration levels values could be attributed to the amount of measurand that exists as a background. By background we mean either the *natural abundance* of the measurand of interest or the amount of measurand that already exist as part of some or all of the chemical reagents used in conducting the experiment (*chemical purity*).

"*Negative values of the response variable shall not be discarded or altered if these arise. For example negative values shall not be replaced by zero"* ISO11843-3.

6.4. Capable But Unstable

This research work proved experimentally and statistically, that data generated by the original low range testing procedure UOP991 for the measurands Fluoride and Chloride are not both statistically and practically different from the ones generated by the modified high range method "HR", but if the concentration level either fluoride or chloride is less 0.3-ppm, then it is preferable to adhere to the original method UOP991, i.e.; *bracketing* technique in calibration must be used.

The sample preparation step which requires to dilute the high concentration sample (concentration range from 1-to-10 ppm) in order to fall into the low range analytical method UOP991 (0.1-to1 ppm) is not required. Note that the process of sample preparation such as in dilution step, will add random variation to the measured test results. Furthermore, this thesis work confirmed that calibrating the chemical measuring equipment (CME) once per a month is considered acceptable so there is no need to calibrate the CME each time samples to be analyzed. This finding of elongating the calibration frequencies will reduce significantly the operational cost of this CME. So the HR method is capable to determine the concentration levels of Fluoride and Chloride at trace concentration level, i.e. $\leq$ 1-ppm but the calibrated model (regression function) has to be built on monthly basis.

6.5. Conclusion

Finally, this thesis proved statistically and experimentally that the high range method is capable of reproducing data for Fluoride and Chloride that are not significantly different from the low range standard lab testing method UOP991. But if the concentration of the measurands is ≤0.3-ppm, then it is preferably to use regression functions that are built as per the original method UOP991. Bromide element was excluded from this work due to the difficulty in preparing the bromide standard by using the starting raw material chemical 4-Bromoacetanilide which does not dissolve readily in the solvent Toluene even by using ultrasonic bath and easily recrystallizes with time; phase separation occurs within short period of time. This thesis was mainly about studying the linearity of the calibration curves developed by different datasets over one year. The time domain (i.e. for how long the validity of calibration method will last) and the time frequency (i.e. how often the measuring equipment needs to be calibrated) are covered by these one year of calibration data sets (11 calibration models as per UOP991 and 6 calibration models for the HR method). In regression analysis, cross-validation is an effective and popular approach for testing the performance of the regression model. The principle of cross-validation is to leave out part of the data, build the model, and then predict the left-out samples. So for method validation by single laboratory which is the case of this thesis; both retrospective and prospective data were used for validating the performance of the built calibration models.

Till this year (2018), there is no an international recognized standard such as ASTM or ISO that addresses *inverse* calibration – if possible, call it an inverse regression-. ASTM E3080 [42] is addressing normal regression analysis, so this is a potential area for future standards development since the vast majority of any instrumental chemical analysis is based on an inverse calibration. For setting the proper calibration "regression" intervals, there exist a guideline developed by ILAC [70].

Glossary

1. **Abundance %:** Mass fraction % of the element in the earth's lithosphere (upper 16 km) plus hydrosphere (oceans) plus atmosphere.

2. **Accepted Reference Value (ARV):** a value that serves as an agreed-upon reference for comparison, and which is derived as: (1) a theoretical or established value, based on scientific principles, (2) an assigned or certified value, based on experimental work of some national or international organization, or (3) a consensus or certified value, based on collaborative experimental work under the auspices of scientific or engineering group. ASTM D3764-15

3. **Accuracy:** It is the closeness between the obtained test result value and the true-or-the reference value. The accuracy term include both; the trueness and precision.

4. **Actual State**: Which is the test sample. ISO11843-1

5. **Alpha (α) and Beta (β): 1.** Alpha α is the probability of erroneously detecting that a system is not in the basic state, i.e. blank contains some measurand. β is the probability of erroneously not detecting that the a system is not in the basic state when the value of the state variable, i.e. the concentration/amount of measurand is equal to the minimum detectable value x_d. **2.** The type I-error is the error that consists in rejecting the null hypothesis (H_o) when H_o is true. It represents the false positive and denoted by α. While **Beta** (β) is the type II-error. It is the error that consists in rejecting the null hypothesis, i.e. accepting H_o, when H_o is false. The probability of type-II error is represented by β. The ***power*** of the statistical comparison is defined by (1- β).

6. **Analyte:** Specific compound to be measured quantitatively in a mixture of compounds. ASTM D4175-16C. Note that the word analyte is used mainly in analytical chemistry to refer to the chemical specie under the study, but the ISO standard (The VIM, ISO Guide:99) is recommending to use the term measurand instead of analyte. Analyte a chemical substance for which quantitative content information, such as its concentration or quantity is sought.

7. **Analytical Measurement System:** A collection of one or more components or subsystems, such as samplers, test equipment, instrumentation, display devices, data handlers, printouts or output transmitters, that is used to determine a quantitative value of a specific property for an unknown sample in accordance with a test method. ASTM D6299-17. Basic properties of measurement system are repeatability, reproducibility, linearity, bias, stability, consistency and resolution. ASTM E2782-17.

8. **Analytical Procedure:** It is the way the analysis is performed. Typically, any analytical procedure covers the following: the sample, the reference standard and chemical reagents preparation, use of the apparatus, the apparatus calibration method, the calculations and the way of reporting the sample test result.

9. **Analyzer Output:** a signal (pneumatic, electrical, or digital), proportional to the property being measured that is suitable for readout or control instrumentation external to the analyzer system. ASTM D3764-15

10. **ANDERSON-DARLING (AD) NORMALITY TEST:** It is a test for randomness (normality) in the least squares residuals. If the P-Value of the Anderson-Darling test is less than the chosen alpha "α" level (typically α =0.05 or 0.10), then the data are not normally distributed. H_o: data follows normal distribution vs. the alternative hypothesis H_1: data does not follow normal distribution. Anderson-Darling test is used to check if a sample of data came from a population with specific distribution. This specific distribution could be; normal, log-normal, exponential distribution, etc. The Anerson-Darling normality test was developed in 1954.

11. **ANOVA :** ANOVA is the acronym for analysis of variance. It is important to note that ANOVA technique is not about analyzing the population variance. In fact, we are analyzing the treatments means (μ_{is}) by identifying sources of variability of the data. In the simplest form of ANOVA, ANOVA can be considered as an extension of the test of hypothesis for equality of two means. In ANOVA, we are trying to know, if the discrepancies among the treatments are due to chance fluctuations or are these discrepancies due to inherent differences among the population.

12. **Arithmetic Mean ($\overline{x}$) vs. Median:** Arithmetic mean is a quantitative measure of central tendency. It is the sum of all data values ($\sum_i^n x$) divided by the number of data values (n). Arithmetic mean is not the only measure of central tendency, and in fact it has some rather unfortunate properties. The most serious failing of the arithmetic mean is that it is highly sensitive to outliers. Any single extremely large or extremely

small value in the data set will have a big effect on the value of the arithmetic mean. On the other hand, Median is a measure of central tendency that is not affected by the presence of outliers. Median is the middle value in the data set.

13. **Assignable Cause:** A factor that contributes to variation and that is feasible to detect and identify. By feasibility it is meant that this factor has economical or otherwise contribution to variation in the process or product output. ASTM D6299-17 and ASTM E2587-10.

14. **Basic State:** Chemical composition of the blank material

15. **Bias:** A systematic error that contributes to the difference between a populations mean of the measurements or test results and an accepted reference or true value. ASTM E177, E456.

16. **Bias and Unbiased Estimators:** A point estimated $\hat{\theta}$ is said to be an unbiased estimator of θ if the $E(\hat{\theta}) = \theta$ for every possible value of θ. If $\hat{\theta}$ is biased, then $E(\hat{\theta}) \neq \theta$ and $\{E(\hat{\theta}) - \theta\}$ is called bias of $\hat{\theta}$. Note that $MSE(\theta) = V(\theta) + Bias\ of\ \hat{\theta}$. If there is no bias exists, then $MSE(\theta) = V(\theta)$.

17. **Blank:** Solution which is similar in composition and contents to the sample solution but does not contain the analyte (measurand) being measured. ASTM D4175-16C.

18. **Bracketing Method:** When the measurement instrument responses to the measured calibrants (reference materials) are nonlinear, then the bracketing method is used to overcome the issue of nonlinearity in the measurement. This method consists of surrounding as tightly as possible (bracketing) each unknown quantity by two reference materials (i.e. calibrants/regressors) and extracting a transformed value for the unknown quantity from measurements of both the unknown quantity and the values of the two reference materials. ISO11095:1996(E)

19. **Calibrant or Calibrator:** Measurement standard used in the measuring instrument calibration process. In statistical sense, calibrant is called regressor/independent variable.

20. **Calibration Function:** Is the output of a calibration procedure. The calibration function is used to make transformations of future measurement results. The term "transformation" refers to either; a correction of the future measurements if both the accepted values of the reference materials (RMs) and the observed values have the same units or a translation from the units of the observed measurement to the units of the RMs.

21. **Calibration Model Validation:** The process of testing a calibration model with validation samples to determine bias between the estimates from the model and the reference method, and to test the agreement between estimates made with the model and the reference method. ASTM E1655-12

22. **Calibration Samples**: Also called Calibration Standards. Are set of samples with known concentration values for the intended property to be measured. These calibration samples are used to calibrate the measuring instrument by developing the calibration function (also called regression function).

23. **Calibration: 1.** Is the process used to create a mathematical model relating two types of measured data. So in the context of instrumental chemical analysis, these two types of data could be represented by the measuring instrument response value for the measured calibration standard (also called Calibrant/independent variable or Regressor). **2.** Process of establishing a relationship between a measurement device and a known standard value(s). ASTM E2782-17

24. **Capability Indices**: Are the ratios of the process spread and the specifications spread. They are unitless values so that they can be used in comparison of different processes capability. Many practitioners consider 1.33 to be a minimum acceptable value for capability indices, and most PR actioners believe a value less than 1 is not acceptable.

25. **Characteristic:** A property of items in a sample or population which, when measured, counted, or otherwise observed, helps to distinguish among the items. ASTM E3080-16

26. **Chemical Kinetics**: The study of reaction rates. The reaction rates depend on the concentration of reactants (and products) and trhe rate constant that are characteristic of the reaction. The rate of chemical reaction ($A + B \rightleftarrows C + D$) is the rate of change of the concentrations of the reactants (A & B) or products (C&D). It is expressed mathematically as:

$$reaction\ rate\ (\upsilon) = \frac{-d[reactants]}{dt} = \frac{d[products]}{dt}$$

υ: Reaction rate.

d[ ] : the change in the concentration of the reactants/products.

dt: the change in time.

The linear dissociated of certain chemical specie could be linear or nonlinear.

27. **Chromatogram:** In column chromatography – which is an instrumental technique for separating sample into its constituents and quantify their amounts -, the recording of the detector signal as a function of elution time or elution volume is termed a *chromatogram*.

28. **Collaborative Studies:** Like proficiency testing collaborative study also called inter-laboratory method performance study or collaborative trial.

29. **Combustion Ion Chromatography (CIC):** an analytical system consisting of oxidative pyrohydrolytic combustion followed by ion chromatographic detection. ASTM D7994.

30. **Common (Chance, Random) Cause:** For quality assurance programs, one of generally numerous factors, that contributes to variation, and that is not feasible to detect and identify. ASTM D6299-17.

31. **Confidence Interval (C.I):** A confidence interval shows the likely range in which the mean would fall if the sampling exercise were to be repeated. So confidence interval is directly proportion to standard error of the mean($SE_{\bar{x}}$). The higher the confidence, the wider the confidence interval (C.I). There are different levels of confidence intervals. Confidence intervals. Confidence intervals are always two-tailed because the parameter may be larger or smaller than our estimate of it. For small sample size($n < 30$), the student's t-distribution is used. So the calculation for the confidence interval of a mean will be:

$$Confidence\ Interval = (t - value)X(Standard\ Error)$$

$$C.I_{95\%} = \left(t_{\alpha=0.025,d.f}\right)X\left(\sqrt{\frac{s^2}{n}}\right)$$

A confidence interval represented by $(1-\alpha)$ is always calculated by selecting a confidence level – also called significant level "α". This significant level is a measure of the degree of reliability of the confidence level. In case of using standardized normal distribution, the formula of confidence level will be:

$$C.I = \bar{y} \pm Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$$

Where σ is the population standard deviation and $Z = \frac{\bar{y}-\mu}{\frac{\sigma}{\sqrt{n}}}$.

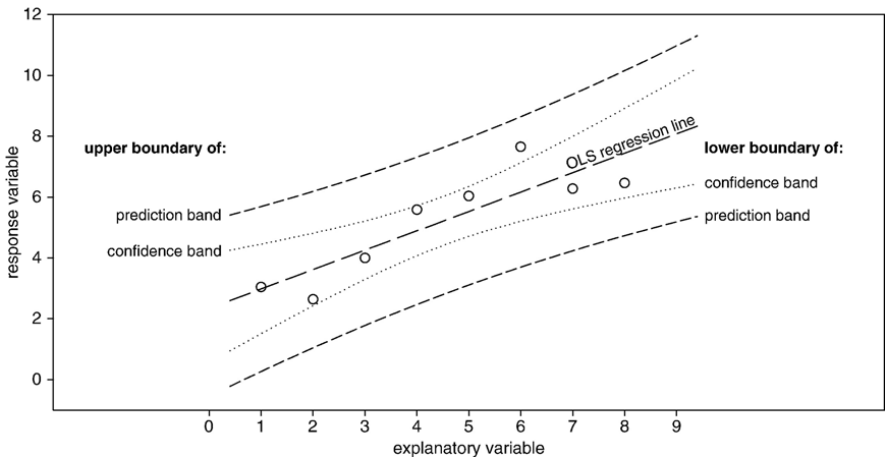
$\bar{y}$: Sample mean , μ : Population mean and n: Sample Size.

Confidence Interval (C.I) is an estimate for the location of the characteristic parameter(θ). It has a Lower Limit (L) and Upper Limit (U), so the interval is [L,U].

These interval statistics L and U are estimated with confidence level –significant level $(1-\alpha)$ where $\Pr(L \leq \theta \leq U) \geq (1-\alpha)$.
Confidence level –i.e. $(1-\alpha)$- is expressed in percentage and it is the probability associated with the estimated confidence interval. Typically, the confidence level is taken to either 95% - which is the most of the time- or at 99%.

Note: Correct Interpretation of C.I:
The correct interpretation for the C.I is that; if we found the population mean μ is falling in the interval $(0.5 \leq \mu \leq 1.5)$ at 95% confidence level (significant level), this means that 95% this calculated C.I will contain the true value of the population mean. It does not meant that 95% the true population mean falls within this calculated C.I. Since the stated probability level (i.e. 95%) refers to the properties of the C.I and not to the parameter (i.e. the population meanμ) itself.

32. **Confidence Interval (CI) vs. Prediction Interval (PI):** In the context of regression line, the range of the PI is wider than the range of the CI and this because the a PI applies to an individual value whereas the CI applies to a mean response of all observations.



33. **Confidence Level (CL):** Is a measure of the degree of reliability. It is called significant level. The higher the confidence level, the more strongly we believe that the value of the estimated parameter lies within the calculated confidence interval. 90%, 95% and 99% are the commonly used confidence level. But the popular choice is the 95%. Confidence level –i.e. $(1-\alpha)$- is expressed in percentage and it is the probability associated with the estimated confidence interval. Typically, the confidence level is taken to either 95% - which is the most of the time- or at 99%.

34. **Continuous Sample Space:** If the sample space contains an interval (either finite or infinite) of real number.

35. **Critical Value:** It is the value of the sample criterion which be exceeded by chance with some specified (small) probability on the assumption that all the observations did indeed constitute a random sample from a common system of causes, a single parent population, distribution or universe. The specified small probability is called "significant level" or "percentage point" and can be thought of as the risk of erroneously rejecting a good observation. *ASTM E178-16a*

36. **Critical Value of the Net State Variable X_c:** Value of the net state variable "X", the exceeding of which leads, for a given error probability, α, to the decision that the observed system is not in its basic state. ISO11843.

37. **Decision Rule:** A rule that describes how measurement uncertainty will be accounted for when stating conformity with a specified requirement. ISO17025

38. **Degrees of Freedom:** The number of independent data points minus the number of parameters that have to be estimated before calculating the variance. ASTM E3080-16

39. **Density:** The mass of liquid per unit volume at 15ºC and its saturation pressure with the standard unit of measurement being kilograms per cubic meter, and it is customarily used in g/cm^3. ASTM D4175-16C.

40. **Dependent Variable vs. Independent Variable:** In regression model($y = \beta_o + \beta_1 X + \varepsilon$), X is called the independent variable (or predictor) and y is called the dependent variable (or response). For the sake of preventing confusion with the concept of statistical independence, X will be referred to as the predictor or regressor. This model $(y = \beta_o + \beta_1 X + \varepsilon)$ is called linear regression model. Linear with respect to the coefficient (β_1) of the regressor and simple because this model has only one predictor (X). In case the regression model contains multiple predictors (e.g. $y = \beta_o + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \varepsilon$), then the model is called multiple linear regression since it has more than one predictor / regressor (X) variables.

41. **Detection Limit (DL):** Also called limit of detection (LOD). It is the smallest amount or concentration of analyte (measurand) in the sample that can be reliably distinguished from zero.The intercept of the calibration function is used for determining the detection limit.

42. **Determination:** The process of carrying out series of operations specified in the test method whereby a single value is obtained. ASTM D4175-16C.

43. **Discrete Sample Space:** If the sample space consists of a finite or countable infinite set of outcomes.

44. **Distillate:** In the petroleum industry, an overhead or side stream liquid from a distillation process. Distillates can be produced directly from crude oil (called straight-run distillate). ASTM D4175-16C.

45. **Dixon Test (Q-Test):** This test is used for detecting outliers in a small sample size. This Q-test is vulnerable to the effects of outlier masking (i.e. existence of two outliers which one of them will hide the detection of the other one).
    Dixon test is based on the difference between the observations and can be used in situations where quick check is needed or when calculating the sample standard deviation is avoided. The Dixon test also called the Q-Test or Q-Statistic.

    H_o: All measurements come from the same population.
    H_1: At least one measurement is not coming from the same population.
    Mathematically it is expressed as follows:
    $Q = \frac{|Suspected\ Observation - Nearest\ Value|}{(Largest\ Obsrevation - Smalles\ Value)}$
    The Q-test is less powerful in detecting an outlier than the criterion of the Grubbs test and also it suffers from outlier masking.
    The decision rule in the Q-test is that if the calculated Q-test value exceeds the critical tabulated Q- value at the same P-value, the suspected observation is rejected.

46. **Duplicate vs. Replicates Readings:** Duplicate analysis, when paired determinations on the same sample almost at the same time. So in duplicated readings, one run of experiment will generated a pair of readings. While in case of replicate analysis, each run of readings, the whole analysis (i.e. the whole experiment) has to be repeated. So for testing the lack of fit (LOF), the obtained observations have to be true replicates, not just duplicate readings.

47. **Durbin-Watson Test:** This used for assessing the independence of the error term (ε). On other words, to test for the presence of autocorrelation in residuals. Autocorrelation means that adjacent observations are correlated. Having an autocorrelation will negatively impact the estimation of the standard errors of the regression coefficient;
    H_o: No residual correlation.
    H_1: Positive residuals correlation.
    Note that the observations obtained from the experiment are of the type time series and they need to be checked for potential presence of autocorrelation (i.e. interdependency) among these observations prior using them for constructing the regression model.

48. **Error Assumptions:** In statistical analysis, typically the error (ε) is assumed to be normally (N), independently (I) distributed (D) with mean equals zero and variance equals$(\sigma)^2$. This is abbreviated as; $(\varepsilon \sim \text{NID}\ (0, \sigma^2))$:
    a. Assessing the Normality of the error term by (Anderson-Darling, Shapiro-Wilk, Kolmogorv-Smirnov).
    b. Assessing the Homoscedasticity (i.e. errors having equal variance) by plotting the studentized residuals versus the predicted $\hat{y}$ or versus each of the predictors. A wedge-shaped pattern indicates homoscedasticity.
    c. Assessing the independence of the error. Numerically, it can evaluated by using the Durbin-Watson test.
    d. Assessing the linearity assumption. This could be done numerically via the lack of fit (LOF) test, which is obtained from Analysis Of Variance (ANOVA).

    Note that testing the assumption of independence (I) and of normality (N) of the error term $\varepsilon \sim NID$ is crucial for the validity of the Lack Of Fit (LOF) of the calibration function.

49. **Error:** **1.** The difference between a random variable and its estimate mean. In statistical sense, any deviation of an observed value from the true but generally unknown value when expressed as a fraction or percentage of the value measured, it is called a relative error. ASTM E1547. **2.** In statistical usage does not connote a mistake or blunder but usually refers to the chance deviation between an observation and its expected or average value. Statistical Manual of the Association of Analytical Chemists. www.aoac.org

50. Estimation: The process of providing a numerical value for a population parameter on the basis of information collected from sample. If a single figure is calculated for the unknown parameter, the process is called point estimation. If an interval is calculated which is likely to contain the parameter, then the procedure is called interval estimation. Some of the common estimation methods are; method of moment, method of maximum likelihood, least-squares method and Bayesian method. Estimate is the statistic /function used for estimation.

51. Estimator: A well-defined function that is dependent on the observations in a sample. The resulting value for a given sample may be an estimate of a distribution parameter (a point estimator) associated with the underlying population. The arithmetic average of a sample is for example an estimator of the distribution mean. ASTM D4175-16C.

52. Event: An event is a subset of the sample space of a random experiment.

53. Experiment: A test or series of tests in which purposeful changes are made to the input variables of a process or a system so that we may observe and identify the reasons for changes that may be observed in the output response.

54. Factors: Are independent variables whose effect on the response variable is a main objective of the study.

55. Family Error Rate (α^*):

$$\alpha^* = 1\text{-}(1\text{-}\alpha)^k$$

k: is the total number of comparisons.

α : is the significance level of two methods (for single comparison)

α^*: is the Family Error Rate. Also called overall type (I) error.

For comparing two methods, k will be $k = \binom{2}{2} = \frac{2i}{2(2-2)i} = 1$, where two methods were selected from total of two methods under comparison.

If three methods were compared with respect to two variables, then

$$k = \binom{3}{2} = \frac{3i}{2(3-2)i} = 3$$

And for four methods compared with respect to two variables, then

$$k = \binom{4}{2} = \frac{4i}{2(4-2)i} = 6$$

At the beginning, the α=0.05, but when the number of method increased, the new level of significance (called α^*; family error rate) will be increased. So for k=6, the new level of significant (α^*) will be ; $\alpha^* = [1 - (1 - 0.05)^6] = 0.18$

Notice how the level of significance increased from $\alpha = 0.05$ to the new level $\alpha^* = 0.18$. This is called inflation of level of significance. So the probability of type (I) error increased from originally $\alpha = 0.05$ to $\alpha^* = 0.18$.

56. **Forecasting VS. Prediction:** Forecasting is about the future (tomorrow's temperature) while forecasting is about finding out the unobserved present. If you want to determine how much your house will sell for, you could make a prediction based on the prices of houses in your neighborhood. The term "forecasting" is used when it is a time series and we are predicting the series into the future. Hence "Weather Forecasts". A time series forecast can be made for any data item collected on a regular basis. Also note that "prediction" is the act of predicting in a cross-sectional setting, where the data are a snapshot in time (e.g. sampling one time from databases). In this case, we use information on a sample of records to predict the value of other records which could be a value that will be observed in the future. Prediction usually is done by regression where modeling is done for current data, using regression and the relationship obtained from this regression curve will be used to predict the unobserved data current finite population. Typical example of applying predication is in instrumental chemical analysis where set of calibration standards are prepared (these are regressors) then each calibrant (regressor) is measured by the measuring instrument to obtain its respective response. The collected instrumental responses for their respective regressors will be plotted and this constructed calibration model will be used for predicting the identity of the unknown sample measured response from these currently available finite population of calibrations chemical standards "calibrants" / regressors. On other words, we predict the measurand concentration level from the working range of the calibration curve, while in case the measurand concentration level is outside the calibration range, then the calibration function is used to forecast measurand concentration level.

57. **Good-Of-Fit Tests**: These tests are used for testing if a given set of data follows a particular probability distribution. "Goodness-of-fit" is given this name because this test means how well the observed distribution of data fits with the distribution that is expected if the variables are independent. So if a set of data is assumed to follow a normal distribution, then the data must be a good fit to this distribution with high degree of confidence. The hypothesis will be:
    H_o: The data following specific distribution.
    H_1: The data does not follow specific distribution.
    Some of the commonly used goodness-of-fit tests are; Anderson-Darling, Kolmogorov-Smirnov, Shapiro-Wilk, Ryan-Joiner and Normality Plot.

58. **Grubbs Test:** This test is used for detecting an outlier observation. In which, the observations of the measurements are arranged in ascending order, forming a montonic series; $\{X_i\}$, i=1,2,…, X_n, where X_1 is the minimum value (X_{min}) and (X_n) is the maximum value (X_{max}). So if the smallest observation is an outlier, the Grubbs statistics has the form:

$$G = \frac{|\bar{X} - X_1|}{s}$$

And for testing the highest observation for its potential of being an outlier reading, the Grubbs formula will be:

$$G = \frac{|X_n - \bar{X}|}{s}$$

Where $\bar{X}$ is the sample mean of the observations including the suspecting reading, n is the number of observations and sample standard deviation is (s).

These two equations represent one-sided Grubbs test statistics. For the two-sided, the Grubbs test statistics will be:

$$G = \max\{\frac{|\bar{X} - X_1|}{s}, \frac{|X_n - \overline{X}|}{s}\}$$

So the maximum value obtained from this formula is considered an outlier. The German standard (DIN EN53804-1) recommends at least 30 observations for reliable performance of this statistical test. Grubbs test is a valid procedure for evaluating a single outlier in a sample which is normally distributed. Also, if the sample has more than one outlier, then the test suffers from the effect of outlier masking (i.e. one outlier observation will not be detected due to the presence of another outlier in the same set of sample). The decision rule is; if the calculated Grubbs value is greater than the tabulated Grubbs value, then reject the H_o (i.e. the suspected value is not an outlier. Otherwise the suspected value is an outlier.

59. **Halogen (X):** A generic term which includes elements; fluorine (F), Chlorine (Cl) and Bromine (Br).

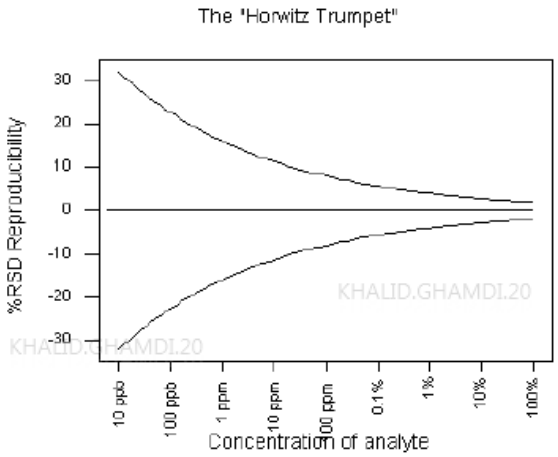
60. **Homogeneity:** The condition of the population under which all items of the population are identical with respect to the characteristic(s) of interest. ASTM D4175.

61. **Homogenous:** The condition of a material in which the relevant properties (composition, structure, density and so forth) are not a function of position for sample size used, so that a small sample taken from any location in an original body is representative of the whole. ASTM C1145.

62. **Homoscedasticity vs. Heteroscedasticity**: One of the assumptions necessary for the validity of the regression inferences is that the error ε should have a constant variance σ^2 for all levels of the independent variable(s). Variances that satisfy this property are called homoscedastic. In contrast, unequal variance(s) for different levels of the independent variable(s) are called heteroscedastic. One of the commonly used tests for testing the presence of heteroscedasticity in data are the residual plots.

63. **Horwitz Function (σ_H):** This function is used for estimating the inter-laboratory precision (i.e. reproducibility "S_R") and it is suitable for use when the measurand concentration level is greater than 120-ppb (part per billion). Note that the right way of estimating the inter-laboratories precision (S_R) is to have a minimum of eight laboratories examining at least five materials to obtain a reasonable estimated of (S_R). Acceptable values for inter-laboratories reproducibility (S_R) are between (0.5 and 2) times the calculated values.

    Horwitz noticed that this trend in (%RSDR: Relative Standard Deviation of Reproducibility) regardless of the nature of measurand and test material (i.e. matrix effect) or the physical principle underlying the measurement method (i.e. the testing procedure). So the test method reproducibility $\sigma_H = 0.02C^{0.8495}$ where c is the measured concentration and can be estimated from the linear form which is:



64. **Identical and Independently Distributed (IID):** If each random variable has the same probability distribution as the others and all are mutually independent, then the random variables are called (IID). Notice that an independent and identically distributed (IID) sequence does not imply that the probabilities for all elements of the sample set must be the same.

65. **Identification:** Is the characterization of the substance being analyzed.

66. **Independent and Identical Distribution Random Variable (IID):** In probability theory and statistics, identical means each random variable has the same probability distribution as the other random variables and all are mutually independent.

67. **Indirect Measurement:** a correlated quantitative measurement result obtained using a measurement principle that produces values that do not express the desired characteristic property but which can be modified empirically, using mathematical modeling techniques, to estimate the necessary defining units of the property of interest. ASTM D3764-15

68. **Inference:** Drawing a conclusion based on the statistical data.

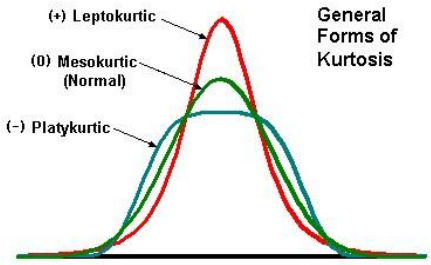
69. **Intermediate Precision:** It is the within-laboratory reproducibility, also called site precision. It is a measure of repeatability of test results within the same laboratory by varying the; analysts, used measuring instrument and different days.

70. **ISO:** The International Organization for Standardization is a worldwide federation of national standards bodies (ISO member bodies). www.iso.org

71. **IUPAC**: International Union of Pure & Applied Chemistry. www.iupac.org

72. **KOLMOGOROV-SMIRNOVE (KS) NORMALITY TEST:** This nonparametric normality test compares the empirical cumulative distribution function of the sample data with the distribution if the date were normal. If the observed vertical difference is sufficiently large, the test will reject the null hypothesis (H_o) of population normality. If the P-Value of this test is less than the chosen alpha level, then the null hypothesis will be rejected and the conclusion is that the population is not normally distributed. H_o: data follows normal distribution vs. the alternative hypothesis H_1: data does not follow normal distribution.

73. **Kurtosis:** It refers to how sharply peaked a data distribution is. Kurtosis value close to zero indicates normally peaked data. Negative values indicate that the distribution is flatter than normal. If the kurtosis found to be positive value, then the data distribution has a peak sharper than the normal peak.



74. **Laboratory:** Body that performs one or more of the following activities; Calibration, Testing and Sampling associated with subsequent calibration or testing. ISO17025

75. **Lack Of Fit (LOF):** This statistical concept is used in the regression design of experiment (DOE). Its purpose is to check if a specific type of regression function adequately fit the data. The LOF test assumes that the observations of the response variable (y) for a given predictor (x) are; independent, normally distributed, constant variance of the distributions of (y) and the only the First-Order (or Straight-Line) is doubt. In order for the LOF to be calculated, it requires that we have replicate observations on the response for at least one level of the predictor. Note that the observations should be replicates and not duplicates. The reason it is called Lack-Of-Fit is; when the developed regression model is passing far from the measured points (hence; Lack Of Fit). The LOF is obtained as the difference between the pure error (PE) and the residuals of the response (y-residuals; which is the distance from the model to the measured y). This distances is called pure error (PE) and mathematically expressed as $\sum(y_{ij} - \bar{y}_j)^2$ where y_{ij} represents the observation of the response (i) for the predictor with level number (j).

$$\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \bar{y}_j)^2$$

a: the total number of the levels of the predictor (i.e. number of the factor levels).
n: the total number of all observations (responses).

In the LOF, the hypothesis is :

H_o: There is a linear association between the response (y) and the predictor (x) (i.e. $\beta_1 = \beta_2 = \beta_3 = \cdots, \beta_j = 0$).

H_1: There is no linear association between the response (y) and the predictor (x) (i. at least one of the $\beta_j \neq 0$.

Decision Rule:

$$\text{Reject } H_o \text{ if } F_o(for\ LOF) > F_{\alpha,m-p,n-m}$$

Where;

α: Chosen level of significant.

m: Number of different levels of the predictor variable (X). Number of replicates.

n: Number of observations in the sample.

p: is the number of regression coefficients (β_{is}) including the intercept (β_o).

Note that $(m-p)$ is the numerator degrees of freedom represented by v_1 and $(n-m)$ is the denominator degrees of freedom represented by (v_2). The whole term $(F_{\alpha,m-p,n-m})$ represents the tabulated value from the percentage points of the F-Distribution. The test statistic for LOF is:

$$F_o = \frac{MS_{LOF}}{MS_{PE}} = \frac{(\frac{SS_{LOF}}{m-p})}{(\frac{SS_{PE}}{n-m})}$$

76. Latent Variable (X) vs. Observed Variable (Y): Latent variable is also called the independent variable, explanatory variable, regressor variable or predictor, while the observed variable is the response value for each latent variable. In the context of instrumental chemical analysis, latent variable can be represented as the measured sample while the observed variable is the measuring instrument response to this measured sample. The Expectation E[X]: Mathematically is expressed as;

$$\bar{X} = \left(\frac{\sum_{i=1}^{n} x_i}{n}\right)_t$$

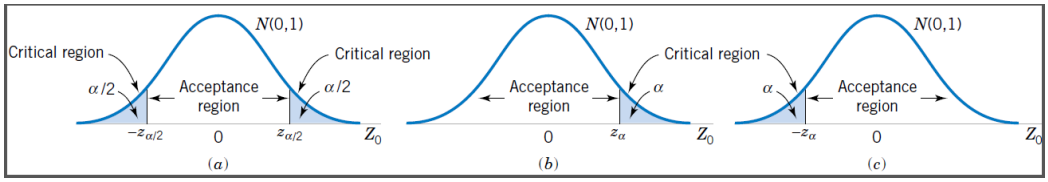
Where;

t: is the observational time interval.

x_i: are the estimates of the random quantity, whose variation in time forms a random

77. Levels: Are the values of the factors. Values of predictors. The different values of an independent variable used in regression are called levels.

78. **Level of Significance (α):** This Greek letter is used to denote the probability of denoting type (I) error (Consumer's Risk). On the context of testing hypothesis, the sum of the two sided shaded area (figure *a*) represents the rejection area. For one-sided upper limit, the rejection part is the upper shaded area (figure *b: upper tail*) and for one-sided lower limit, this is represented by the lower shaded area (figure *c: lower tail*) as shown in figure 75.1. The value of α is usually set in advance, with commonly chosen values being α=0.10 (90% confidence interval), 0.05 (95% confidence interval), and 0.01(99% confidence interval).



Standardized Normal Distriubtion Curve with Critical Regions

79. **Levene's Test (modified):** The test is used to test for equal variances in all treatments when the observations responses(Y_s) are not normally distributed. This modified Leven's test used the absolute deviation of the observations y_{ij} in each treatment from the treatment median (let it be $\tilde{y}_i$). These deviations are denoted by :

$$d_{ij} = |y_{ij} - \tilde{y}_{ij}| \begin{cases} i = 1,2,\dots a. \\ j = 1,2,\dots n. \end{cases}$$

Where;

a : is the number of treatments.

i : is the treatment number.

j : is the observation number.

n : is the total number of observations.

The modified Leven test then evaluates whether or not the means of theses deviations are equal for all treatments. If the mean deviations are equal, the variances of the observations in all treatments will be the same. The test statistic for Levene's test is simply the used ANOVA F statistic for testing equality of means applied to the absolute deviations.

80. **Linear Model:** This term refers to the fact that the model ($Y_i = \beta_o + \beta_1 X_{i\,1} + \beta_2 X_{i\,2} + \cdots + \beta_{p-1} X_{i\,,p-1} + \varepsilon_i$) is linear in the parameters (β_s); - i.e. these beta parameters each of them has an exponent of one- and it does not refer to the shape of the response surface of this regression model described by this function Y_i .

81. **Linearity:** The linearity of an analytical procedure is its ability (within a given range) to obtain test results which are directly proportional to the concentration (amount) of measurand (i.e. analyte) in the sample. Validation of Analytical Procedures: ICH Harmonized Tripartite Guideline Q2(R1) www.ich.org

82. **Lower Limit of Quantification (LLQ):** Method detection limit has to do with the intercept of the calibration function$y = ax + b + \varepsilon$.

83. **Masking:** In the context of detection of an outlier observations, masking means the presence of more than one outlier, making each outlier difficult to detect.
a: Is the slope of the calibration line.
b: Is the intercept of the calibration line.
ε: Is the error term which supposed to be $NID(0, \sigma^2)$.

84. **Mathematical Model/Method of Least Squares / Regression Line/ Best Fit Line:** In experimental work, a mathematical relationship $y = f(x)$ is developed between two variables (x) which is the independent variable and (y) is the dependent variable (called the response). So experimentally, a pairs of data points are obtained $(x_1, y_1), (x_2, y_2), (x_3, y_3), \ldots . (x_n, y_n)$. the curve $y = f(x)$ represents the mathematical model of these pairs of data.
The interest is to fit a straight line $y = mx + b$ to data. Usually, the data will not lie on a line (this could be due to experimental error or variations in experimental conditions), so the challenge is to find a line that fits the data "best" according to some criteria. One criterion for selecting the line of best fits is to choose (m) and (b) to minimize the function $g(m, b) = \sum_{i=1}^{n} (mx_i + b - y_i)^2$. This is called the method of least squares, and the resulting line is called regression line or the least squares line of best fit.

85. **Matrix:** All Components of the sample. ISO 16140-1:2016.

86. **Measurand**. **1.** Quantity subject to measurement. In analytical chemistry, the widely used term is Analyte (Analyte: is usually the concentration of a substance with a statement of its uncertainty and the identity of the subjstance) while the ISO standard (ISO/IEC Guide:99:2007) is recommending the term Measurand. **2.** The measurable quantity subject to measurement. ASTM D4175-16C. 2.

87. **Measurement Bias:** Estimate of a systematic measurement error. ISO/IEC 99:2007.

88. **Measurement Error:** Measured quantity value minus a reference quantity value. ISO/IEC Guide 99:2007.

89. **Measurement Procedure:** Detailed description of a measurement according to one or more measurement principles and to a given measurement method, based on a measurement model and including any calculation to obtain a measurement result. A measurement procedure is sometimes called a Standard Operating Procedure (SOP). ISO/IEC: Guide 99:2007.

90. **Measurement Result**: (Result of Measurement). Set of quantity values being attributed to a measurand together with any other available relevant information. ISO/IEC Guide 99:2007.

91. **Measurement Standards:** Realization of the definition of a given quantity, with stated quantity value and associated measurement uncertainty, used as reference. Example 1-Kg mass measurement standard with an associated standard measurement uncertainty of 3-µg. ISO/IEC Guide 99:2007.

92. **Measurement System: 1.** Represents not only a measuring instrument but also the set of procedures, operators and environment conditions associated with that instrument. ISO 11095:1996 (E) **2.** The collection of hardware, software, procedures and methods, human effort, environmental conditions, associated devices, and the objects that re measured for the purpose of producing a measurement. ASTM E2782-17

93. **Measurement Uncertainty:** A non-negative parameter characterizing the dispersion of the quantity being attributed to a measurand. The parameter could be a standard deviation or specific multiple of it or the half-width of the confidence interval, having a stated coverage probability. ISO/IEC Guide 99:2007.

94. **Measurement:** Process of experimentally obtaining one or more quantity values that can reasonably be attributed to a quantity. ISO/IEC Guide 99:2007.

95. **Memory Effect:** In the context of statistical quality control, memory effect concept is used for quality control charts that are capable of relating each observation to its previous observation. The purpose of this is to predict any assignable cause shift in the process mean. While, memory effect is interpreted in the context of instrumental chemical analysis as the residual (carryover) from previous sample that was injected in the measuring instrument. So if two consecutive samples injections are done on the measuring instrument, the second injection might show sample results higher the actual measurand concentration. Memory effect is considered as an advantage of certain types of control charts like EWMA and CUSUM but it is a nuisance factor in instrumental analysis. Normally, running, the sample multiple times then selecting the best reading or running blanks between two consecutive samples is the usual remedies to reduce the impact of the instrumental memory effect.

96. **Method:** Synonym with the term measurement procedure.

97. **Method Detection Limit (MDL) vs. Instrument Detection Limit (IDL):** In method validation, it is important to consider the MDL and not the IDL. Reason being that the IDL is referring to the measuring instrument response that is based on the analysis of sample or reagent blank that are presented directly to the instrument (i.e. no sample preparation part is considered). While the MDL is the measuring instrument response that is generated due the whole measurement procedure (i.e. sample went through sample treatment prior being presented to the measuring instrument).

98. **Method of Analysis:** Is the detailed set of directions, from the preparation of the test sample to the reporting of the results that must be followed exactly for the results to be accepted for the stated purpose. AOAC Official Methods of Analysis (2012) Appendix K.

99. **Method Validation:** "The word validation originates from the Latin validus meaning strong, and suggests that something has been proved to be true, useful and acceptable standard". Journal of Chromatography A, 1232(2012) 101-109. Elements of method validation are; Specificity, Accuracy, Linearity, Intermediate Precision, Repeatability, Reproducibility, Limit of Detection, Limit of Quantification and Range.The objective of validation of an analytical procedure is to demonstrate that the procedure is suitable for its intended purpose. So validation work is proved via documented evidence that the alternative test method is equivalent to the standard method.

100. **Minimum Detectable Value X_d of the Net State Variable:** Value of the net state variable in the actual state that will lead, with probability $(1-\beta)$, to the conclusion that the system is not in the basic state.

101. **Model:** A formalized mathematical expression of the process assumed to have generated the observed data.

102. **Model Validation:** "The process of determining the correctness of the assumptions and governing equations implemented in a model when applied to the entire class of problems addressed by the model". ASTM E176.

103. **Model Error (ε_i) Term vs. Residual(e_i):** The ith residual is the difference between the observed value (Y_i) and the fitted value (also called predicted value) $(\hat{Y}_i)$ and both are known values. So;

$$e_i = Y_i - \hat{Y}_i$$

104. **Multiple Comparisons Tests:** When the null hypothesis (H_o);

$$\tau_1 = \tau_1 = \tau_2 = \cdots = \tau_a = 0.$$

These τ'hous means there is no significant difference among the treatments. Analysis of Variance (ANOVA) doesn't identify which means are different. Method of investigating this issue are called multiple comparisons methods. The most commonly used ones are; Fisher's Least Square Difference F(LSD), Tukey's Test, Dunnet's Test and Scheffe's Test .

Tukey's test for comparison is the best for all-possible pairwise comparisons when sample sizes are unequal or confidence intervals are needed. Also it is very good even with equal samples sizes without confidence interval. Tukey's approach is specifically for comparing group of means in an ANOVA setting. **Dunnett's test** on the other hand, is comparing one sample (called control sample) to each of the others, but not comparing the others to each other.

Fisher's LSD is not a multiple comparison method, but instead contrasts the individual confidence intervals for the pairwise differences between means using an individual error rate. Fisher's LSD method inflates the family error rate.

105. **Multiple Linear Regression:** Is a regression model that involves more than one regressor (dependent variable/predictor/calibrant). So for model with two predictor variables the model could be represented as; $Y_i = \beta_o + \beta_1 X_1 + \beta_2 X_2 + \varepsilon_i$. This is a first order model with two regressors ($X_1 \; and \; X_2$).
The term linear is a linear function of the unknown population parameters (i.e. $\beta_1 and \; \beta_2$).
The parameter (β_1) indicates the expected change in response (Y_i) per unit in (X_1) when (X_2) is held constant. Similarly, (β_2) indicates the expected change in response (Y_i) per unit change in (X_2) when (X_1) is held constant.
When the effect of (X_1) on the mean response does not depend on the level of (X_2) and correspondingly the effect of (X_2) does not depend on the level of (X_1) , the two predictors (i.e. the two regressors/calibrants $X_1 \; and \; X_2$) are said to have an additive effect (i.e. no interaction exists between them. On other words, no multicollinearity exist).

NOTES:

1. *In general, any regression model that is linear in the parameters (i.e. the coefficients $\beta_i s$) is a linear regression model regardless of the shape of the surface that it generates.*
2. *Based on the Parsimony principle in building regression models, the number of (predicotrs/regressors/calibrants/independent variables) p should be less thant the number of observations $(n) \Rightarrow p < n$.*
3. *In regression analysis, the term linear is used because the regression function is linear since its equation determines a straight line n the XY plane. Another interpretation for the using the term linear is because the parameters β_o and β_1 are linear in the parameters.*
4. The vertical distances are called the residuals of the data points, so the purpose of the least square method function $g(m, b)$ is to minimize the sum of the squares of these residuals.

$$\text{Slope} \quad m = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)}$$

$$\text{Intercept} \quad b = \frac{1}{n}(\sum_{i=1}^{n} y_i - m\sum_{i=1}^{n} x_i)$$

Ref; CALCULUS. Anton,Bievens and Davis. 10th edition. John Wiley & Sons INC. 2012 USA.

106. **Naphtha:** A general term applied to refined petroleum product not less than 10% of which distill below 175 ^{o}C and less than 95% of which distills below 240 ^{o}C when subject to distillation according to ASTM D86. Naphtha is used in making various chemicals and as solvent.

107. **Natural Abundance:** Abundance (mass fraction %) of the element in the earth's Lithosphere (i.e. upper 16-km of the solid land) plus Hydrosphere (oceans) plus Atmosphere (Atmosphere is the envelope of gases surrounding the Earth).
Fluorine (F) with the mass number 19 (^{19}F) has 100% natural abundance. Only one isotope exists for Fluorine which is (^{19}F).
Chlorine (Cl) with mass number 35, (^{35}Cl) has 75.8% natural abundance. Chlorine has two isotopes, one with mass number 35 and the second 37 (^{35}Cl and ^{37}Cl).
Bromine (Br) has two isotopes: (^{79}Br) with mass number 79 and (^{81}Br). The former represents 50.7% and the latter represents 49.3% natural abundance.

About 126-ppm of Earth's crust is chlorine (Cl), (600-to-700 ppm) of Earth's crust is Fluorine (F) and Bromine (Br) represents about 2.5-ppm of Earth's crust.

108. **Net State Variable (X):** Which is the test samples. It is the difference between the state variable (Z) and its value in the basic state (Z_o), so $= Z - Z_o$. The value in the basic state means the measurand concentration in the blank.

109. **Noise vs. Drift**: Noise is the amplitude expressed in amperes (A) or Hertz (Hz) of the baseline envelope which includes all random variations of the detector signal of the frequency on the order of 1 cycle/min or greater. This is the observed noise. On the other hand, drift is the average slope of the noise envelope expressed in amperes per hour or Hertz per hour as measured over a period of $^{1}/_{2}$ hour. So drift is a noise per unit time. ASTM E697-11

110. **Normal Distribution vs. Student's t-Distribution:** Both normal and Student's distributions assume there is no skewness in the data distribution curve. Student's distribution is used when the sample size is $n < 30$. Student's t-distribution is the pseudonym of William S. Gosset.

111. **Normal Probability Distribution:** A random variable X is said to have a normal probability distribution with parameters μ and σ^2, if it has a probability density function given by:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{\frac{-(x-\mu)}{2\sigma^2}} \;,\; -\infty < x < +\infty \,, -\infty < \mu < +\infty \;\; and \;\; \sigma > 0$$

If $\mu = 0$ and $\sigma = 1$, then the normal probability distribution is called Standard Normal Distribution.

112. **One-Point Calibration:** This fast calibration method requires using only a blank (calibrating standard where the concentration of the measurand is almost zero) and one reference material for calibration. This calibration method can be used only when the calibration function is linear, so it is used to check the linearity of the calibration method. Due to the weakness of this calibration method, the one –point calibration method is not widely used.

113. **Ordinary Least Square Regressions (OLSR):** Where all data points are given equal weight. Weight is the coefficient assigned to observations in order to manipulate their relative influence in subsequent calculations. For example, in weighted least squares, noisy observations are weighted downwards, while precise data are weighted upwards. The disadvantages of (OLSR) are:
    - Special effects like the matrix influence the accuracy of the predicted value. On other words, the (OLSR) model is affected by the influence of several operating conditions on measurement response.
    - In trace analysis (i.e. when the concentration of the measurand is below 1-ppm) it is a basic requirement that the analytical method used to have a variance as low as possible to assist threshold level decisions.
    - If the variance homogeneity is violated, then using the Weight Least Squares Regression (WLSR) is recommended over the (OLSR).

Coefficient of Variation: Also called Relative Standard Deviation (RSD. Typically RSD value needs to be =1 and is defined mathematically as $= \frac{\sigma}{\bar{X}}$. and estimated by sample RSD which is $RSD = \frac{\sqrt{\frac{\sum_i^n (x_i - \bar{X})^2}{n-1}}}{\bar{X}}$

Where: x_i is the observation ith, n is the total number of observations (i.e. sample size) and $\bar{X}$ is the average of all observations.

114. **Outlier:** A result far enough in magnitude from other results to be considered not a part of the set. Outliers are typically detected by box plots, stem-and-leaf plots, scatter plots and residual plots. Once an outlier observation is determined, it will be tested to check if tis either a leverage point or an influential point. Influential point has a negative impact on the regression model, so it has to be removed. Note that "one should be distinguish between data to be used to estimate a central value from date to be used to assess variability. When the purpose is to estimate a standard deviation, it might be seriously underestimated by dropping too many outlying observations". ASTM E178-16a.

115. **Parameters vs. Variables:** Variable is any characteristic whose value may change from one object to another in the population. So the heights and weights are called variables. The averages and standard deviation of these heights and weights are called parameters.

116. **Parametric:** A term referring to a statistical technique that assume the nature of the underlying frequency distribution is known. ASTM D4175-16C.

117. **Parsimony Principle:** In situations where two competing models are found to have essentially the same predictive power, the model with the lower number of regression coefficients (i.e.β_{is}) is selected. β_{is} denotes the number of coefficients – including the intercept (β_o).

118. **Percentile:** The set values that divide the sample into 100 equal parts.

119. **Population: 1.** A population consists of the totality of the observations with which we are concerned. A sample is a subset of observations selected from a population. **2.** The totality of items or units of material under consideration. ASTM E3080

120. **Precision**: **1.** The degree of agreement between two or more test results on the same property obtained using the same test method on identical test material. ASTM D4175-16C. The total precision (σ_{Tot}) is a combination of precession under repeatability conditions (r) σ_r and under reproducibility conditions (R), so $\sigma_{Tot} = \sigma_r + \sigma_R$. **2.** The precision of analytical procedure expresses the closeness of agreement (degree of scatter) between a series of measurements obtained from multiple sampling of the same homogenous sample under the prescribed conditions. Precision may be considered at three levels: repeatability, intermediate precision and reproducibility. Precision should be investigated using homogenous authentic samples. However, if it is not possible to obtain a homogenous sample it may be investigated using artificial prepared samples or a sample solution. The precision of analytical procedure is usually expressed as the variance, standard deviation or coefficient of variation of a series of measurements. Ref: ICH Harmonized Tripartte Guideline Q2(R1)-2005.

121. **Probability (P) and Expectation (E):** The probability of the event P(X) which has an outcome value (X) and Expectation (E) is expressed as P(X)=E. So if the probability P of getting $100 (value is $100) is 0.01, then the expectation is $1.

122. **Probability:** A numerical measure between 0 and 1 assigned to events in a sample space. Higher numbers indicate the event is more likely to occur.

123. **Process:** Historically, the term process has been used to suggest the observation of a system over time.

124. **Process Capability Indices:** These indices compare the variability of a process quality measure against product specifications or tolerances and assume the process in statistical control. Capability indices are the ratios of the process spread and the specifications spread. They are unitless values so that they can be used in comparison of different processes capability. Many practitioners consider 1.33 to be a minimum acceptable value for capability value for capability indices; and most practitioners believe a value less than 1 is not acceptable. Minitab 18.

125. **Proficiency Testing: 1.** Evaluation of participant performance against pre-established criteria by means of inter-laboratory comparisons. ISO17025. **2.** Proficiency testing Involves the use of inter-laboratory "comparisons to determine the performance of participants (which may be laboratories, inspection bodies, or individuals) for specific test or measurement. Purposes of proficiency testing can be found in ISO17043:2010.

126. **Protocols, Standards and Guidelines:** Protocol is a formal set of conventions governing a communication process. ASTM F1457. Standard is a concept that has been established by authority, custom, or agreement to serve as a model or rule in the measurement of quality or establishment of a practice or procedure. ASTM E7. While guideline is defined by ASTM E631 as a written statement or outline of policy, practice or conduct.

127. **Pure Error:** In the context of the test for lack of fit (LOF), the full model error sum of squares is called the pure error sum of squares and is denoted by (SSPE). Note that SSPE is made up of the sums of squared deviations at each level of the predictor variable (X). So at the level of $X = X_j$, this sum of squared deviations will be:
$\sum_i (Y_{ij} - \bar{Y}_j)^2$ also it could be written as; $\sum_{i=1}^{a} \sum_{j=1}^{n} (y_{ij} - \bar{y}_j)^2$
Where;

Y_{ij}: observation (response) number i of the predictor number$j$.

n: is the total number of all observations (responses).

a: is the total number of levels of the predictor (number factors levels).

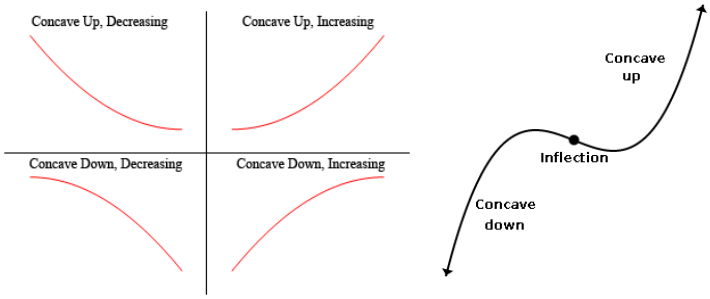
i: is the number of replicates.

j: is the number of predictors.

128. **P-Value:** The smallest significant level at which the null hypothesis (H_o) will be rejected. Value of P-value greater than α-value (probability of type "I" error) indicates that the null hypothesis is unlikely to be true. The smaller is the P-value, the greater is the probability of rejecting the null hypothesis (H_o). P-value is usually used as an index showing the statistical significant of comparison between two groups. P-value is a common index for the strength of evidence.
The value of the P-value can be made small enough with increasing the sample size (n). So the null hypothesis will always be rejected regardless of the type "I" error α. So the widely used rule of thumb that reject the null hypothesis (H_o) if the ($P < \alpha$) is not necessary correct. So the P-value is biased towards the sample size. The greater will be the sample size, the smaller will be the P-value. The ability of getting the p-values arbitrarily small with increased sample size (n) contributes to false positive.
The formula for calculating the P-value for the comparison between two means is;

$$P = \Phi\left(\frac{\bar{Y} - \bar{X}}{s\sqrt{2}} * \sqrt{n}\right)$$

129. **Quadratic Model:** when the relationship between the response and the predictor variable (regressor/calibration standard) is not linear and described by curvature. Mathematically is written as $y = \beta_o + \beta_1 x + \beta_2 x^2 + \varepsilon$, where;
β_o: is the y-intercept of the curve.
β_1: is the shift parameter.
β_2: is the rate of the curvature.
If the value of β_2 is positive, then the curvature will be concave upward and if it is negative, then the curvature will concave downward



130. **Qualitative Method:** Method of analysis whose response is either the presence of absence of the analyte (i.e. the measurand) detected either directly or indirectly in a specified test portion.

131. **Quality Assurance (QA) vs. Quality Control (QC):** Quality assurance addresses the activities the laboratory undertakes to provide confidence that quality requirements will be fulfilled, whereas quality control describes the individual measures which are used to actually fulfill the requirements. ISO9000.

132. **Quality Control Sample (QC Sample): 1.** For use in quality assurance program to determine and monitor the precision and stability of a measurement system; a stable and homogenous material having physical or chemical properties, or both, similar to those of typical samples tested by the analytical measurement system. The material is properly stored to ensure sample integrity, and is available in sufficient quantity for repeated long-term testing. ASTM D4175-16C. **2.** For use in assurance programs to determine and monitor the precision and stability of a measurement system, a stable and homogeneous material having physical or chemical properties or both similar to those of typical samples tested by the analytical measurement system. The material is properly stored to ensure sample integrity and is available in sufficient quantity for repeated, long term testing. ASTM D6299

133. **Quantiles:** Divisions of a probability or frequency distribution into equals, ordered subgroups for example; quartiles of percentiles.

134. **Quantitative Method:** Method of analysis whose response is the amount (count or mass) of the analyte measured either directly (e.g. enumeration in a mass or volume) or indirectly (e.g. color absorbance, impedance, etc.).

135. **Quantity:** Property of a phenomenon, body or substance where the property has a magnitude that can be expressed as number and a reference. ISO/IEC 99:2007 (VIM).

136. **Quasi Outlier:** This term refers to the statistic pertaining to the selected compatibility criterion exceeds a critical value that corresponds to the 95% confidence interval, but does not exceed a critical value that corresponds to the 99% confidence interval.

137. **Random Error:** The chance variation encountered in all experimental work despite the closest possible control of variables. It is characterized by the random occurrence of both positive and negative deviations from the mean value for the method, the algebraic average of which will approach zero in a long series of measurements. ASTM E1547.

138. **Random Experiment:** Any experiment that can result in different outcomes (values), even though it is repeated in eh same manner every time.

139. **Reagent Blank vs. Sample Blank:** Reagent blank is referring the reagent used during the analytical process (including solvents used for sample preparation) and this blank is analyzed in order to determine whether the reagent contribute to the measurement signal, while the sample blank is a real life sample with no analyte (measurand) present. The purpose of using blanks in analytical measurements are mainly; for obtaining more accurate readings about the concentration of certain measurand by subtracting the instrumental signal that is coming from the blank from the instrumental signal that is coming from the analyte (measurand) that exists in the sample. Or, the blank could be used for flushing the analytical measurement system (i.e. the instrument) consequently, reduces any memory effect (i.e. samples residue that the analytical measurement system still contaminated with.

140. **Reagent Water:** Water that is used specifically as a component of an analytical process meets or exceeds the specifications for these waters. (ASTMD1193).Reagent water is classified into four types; type I, II, III and IV. Historically, this classification was based on the process which was used to purify the water. But currently, these classifications signify the grade of the reagent water. Type (I) water is water with electrical resistivity greater than or equal $18M\Omega$.

141. **Reference Limits of the Product Characteristics:** $X_{0.135\%}$ and $X_{99.865\%}$ are the quantiles of the distribution of the product characteristic – characteristic is the distinguishing feature-. If the distribution of the product characteristic is normal with mean μ and standard deviation σ, the limits are $\mu \pm 3\sigma$ if traditional 0.135% and 99.865% quantiles are used. In normal distribution with mean μ and standard deviation σ , the reference interval corresponding to the traditional 0.135% and 99.865% quantiles has limits $\mu \pm 3\sigma$, and has length 6σ. ISO 22514-1:2014 (E)

142. **Reference Material:** A substance or an artifact for which one or more properties are established sufficiently well to be used to validated a measurement system. Reference materials could be developed in-house by a user, provided by external body other than the end-user such as the ready-made reference materials or a reference material could be a certified reference material – this is high grade/quality- with certified values by an organization recognized as competent to do so. ISO 11095:1996(E)

143. **Reference Method: 1.** Preexisting recognized analytical method against which the candidate method will be compared. Appendix J: AOAC International Methods. 2012 AOAC International. **2.** Internationally recognized and widely accepted method. ISO16140-1:2016.

144. **Relative Uncertainty:** Uncertainty expressed as relative standard deviation.

145. **Repeatability:** The precision of a method expressed as the agreement attainable between independent determinations performed at essentially the same time (duplicates) by one analyst using the same apparatus and techniques. ASTM E344. Good repeatability indicates that the random errors are small. So random errors can be discovered by repeatability.

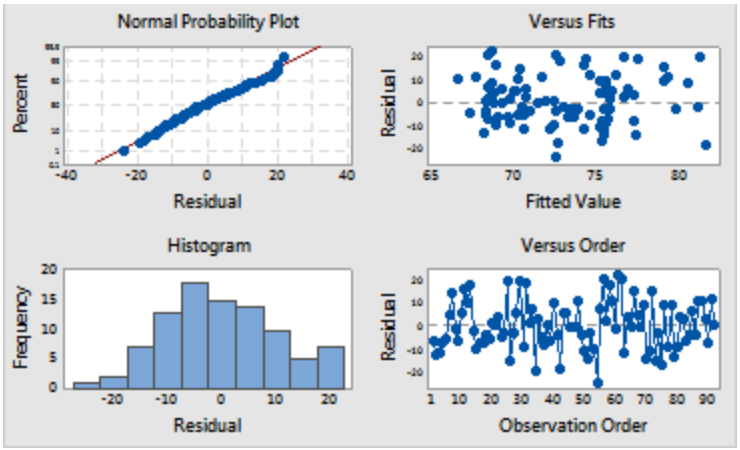
146. **Replicate: 1.** Are multiple observations with identical values. **2.** In experimenting or testing, one of two or more runs with the same specified experimental or test conditions and with each experimental or test condition being established independently of all previous runs. ASTM D123.

147. **Reproducibility:** The closeness of agreement between the results of successive measurements for the same test specimen, or of test specimens taken at random from a homogeneous supply, but changing conditions such as operator, measuring instrument laboratory, or time. ASTM E284. It is evaluated by means of an inter-laboratory trial. Systematic errors can be discovered experimentally either by comparing a given result with a measurement of the same quantity performed by a different method or by using more accurate measuring instrument. Good reproducibility indicates that both random and systematic errors are small.

148. **Residual (e_i):** The difference between the actual response (y_i) and the predicted response from the regression line $\hat{y}_i$ so $e_i = y_i - \hat{y}_i$. The residual describes the error in the fit of the model to the ith observation y_i. The sum of squares of residuals (SSRes) is used to provide information about the adequacy of the fitted model.

149. **Residual Plot:** Use to examine the goodness of model fit in regression and ANOVA. Examining residual plots helps in determining if the ordinary least squares assumptions are being met. If these assumptions are satisfied, then ordinary least squares regression will produce unbiased coefficient estimates with the minimum variance. These residuals plots are:

    Histogram of the Residual Plot: An exploratory tool to show general characteristics of the residuals including typical values, spread, and shape. A long tail on one side may indicate a skewed distribution. If one or two bars are far from the others, those points may be outliers.



***Normal Probability Plot of residuals*.** The points in this plot should generally form a straight line if the residuals are normally distributed. If the points on the plot depart from a straight line, the normality assumption may be invalid.

***Residuals Versus Fitted Values*.** This plot should show a random pattern of residuals on both sides of 0. If a point lies far from the majority of points, it may be an outlier. There should not be any recognizable patterns in the residual plot. For instance, if the spread of residual values tend to increase as the fitted values increase, then this may violate the constant variance assumption.

***Residuals Versus Order of Data*.** This is a plot of all residuals in the order that the data was collected and can be used to find non-random error, especially of time-related effects. This plot helps you to check the assumption that the residuals are uncorrelated with each other (i.e. independence of the error term).

150. **Response & Explanatory Variable:** The response variable usually given the symbol (Y) and plotted in the ordinated while the explanatory variable (also called independent variable) is given the symbol (X) and plotted in the abscissa of the calibration graph.

151. **Result:** The value obtained by following the complete set of instructions of a test method. It may be obtained from a single determination or several determinations, depending on the instruction of the test method. STM D4175-16C.

152. **Risk Management:** In the context of evaluating the equivalence between two laboratory testing procedures, risk management is referred to the amount of laboratory data needed to control the risk of making wrong decision in accepting or rejecting the equivalence between two testing methods. These risks are the type (I) error (Producer's Risk "α") and type (II) error (the consumer's risk "β").

153. **Robust Estimation**: Estimation method that is insensitive to small departures from assumptions about the underlying probability model of the data.

154. **RSD vs. Uncertainty:** Relative Standard Deviation (RSD) reflects the precision of the obtained observations, while the uncertainty concept reflects both the precision and accuracy of the observations.

155. **Ruggedness vs. Sensitivity:** Ruggedness is a measure of the analytical method capacity to remain unaffected by external interferences while sensitivity is a measure the of the analytical method capacity to remain unaffected from internal interferences.

156. **Ruggedness:** it is the ability of the analytical method to resist having changes in the results produced when minor deviations are made from the experimental conditions described in the procedure. It is defined by ASTM E-456 as "the insensitivity of a test method to departures from specified test or environmental conditions". The ruggedness of a method is tested by deliberately introducing small changes to the procedure and examining the effect on the obtained results. So ruggedness is a measure of reproducibility of test results under the variation in conditions such as; different laboratories, different analysis, different measuring equipment. Ruggedness is required as part of the method development.

157. **Ryan-Joiner (RJ) Normality Test:** This test is similar to Shapiro-Wilk. The Ryan-Joiner test calculates the correlation between the sample data and the normal scores of these sample data. If the correlation coefficient (RJ) is close to one, then the population is likely to be normal. The Ryan-Joiner statistic (RJ) assesses the strength of this correlation, if it falls below the chosen critical value (for example $\alpha = 0.05$), then the null hypothesis is rejected and we conclude that the population is not normally distributed. This Normality Test was developed by Thomas A. Ryan and Brian Joiner in 1976.

158. **Sample: 1.** In statistical science, a group of observations or test results, taken from a larger collection of observations or test results, which serves to provide information that may be used as a basis for making a decision concerning the larger collection. ASTM E2586. **2.** In chemical/physical sciences sense, sample is a part of physical material to be studied, while in statistical term, sample is defined as set of values extracted randomly from an overall population of values. *J.M.Andrade et al. Analytica Chemica Acta 838(2014) 1-12.*

159. **Sample Size:** Number of observed values in the sample. ASTM E2586

160. **Sample Space**: Is the set of all possible outcomes of a random experiment. The sample space could be discrete or continuous.

161. **Sample Standard Error ($SE_{\bar{x}}$) & Population Standard Deviation (SD "σ"):** Is a measure of the unreliability of the data. Standard error of the sample mean is defined as:

$$SE_{\bar{x}} = \sqrt{\frac{s^2}{n}}$$

Where s^2 is the sample variance and n is the sample size. The standard error is also called the error of the mean. On the other hand, the population standard deviation (SD) is defined as;

$$SD = \frac{\sum_{i=1}^{n}(x_i - \bar{x})}{\sqrt{n-1}}$$

Standard deviation of a population is given by a symbol (σ) and defined mathematically as;

$$\sigma = \sqrt{(X-\mu)^2}$$

μ: Population Mean and is defined as $\mu = E[X]$

X: An observation (Random Variable) in the population.

n: is the sample mean. Population variance (σ^2) is the expected squared difference between the observation (X) and the population mean (μ). If the population mean (μ), is known, then the estimator of the population variance (σ^2) is; $\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{n}$.

If the population mean (μ) is unknown, then the estimator of the population variance (σ^2) is S^2;

$$S^2 = \frac{\sum_{i=1}^{n}(X_i-\bar{x})^2}{n-1}.$$

162. Sampling Distribution: The probability distribution of a statistic is called sampling distribution. Typical example is the normal distribution of a random variable X which is described by the statistics:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{\frac{(x-u)^2}{2\sigma^2}} \quad +\infty < \mu < -\infty$$

Where;

μ: is the mean of the distribution.

σ >0: is the distribution variance.

The sampling distribution describes how the statistic varies in value across all samples that might be selected. Note that in this context, the word statistic mathematically means function.

163. Selectivity: Is the property of a measurement system used with a specified measurement procedure, whereby it provides measured quantity values for one or more measurands such that the values of each measurand are independent of other measurands or other quantities in the phenomenon, body or substance being investigated.

164. **Shapiro-Wilk Test (W):** This test is used for checking the departure from normality. It is applicable when the sample size (n) is $(8 \leq n \leq 50)$ only. This test is based on the regression of the order statistics upon their expected values. It is an analysis of variance type test for a complete sample. The test statistic is the ratio of the square of a linear combination of the sample order statistic to the usual estimate of variance. This test was developed in 1965 by Shapiro and Wilk and it detects departures from normality due to either skewness or kurtosis, or both. The value of W lies between zero and one. Small values of W lead to the rejection of normality whereas a value of one indicates normality of the data.

165. **Significance:** In the context of statistical science, significance means that the observation (obtained/occurred) not by chance.

166. **Simple Linear Regression:** When the regression model has only one explanatory variable (i.e. one independent variable "X"). Mathematically this model is written as:

$$y = \beta_o + \beta X + \varepsilon_i$$

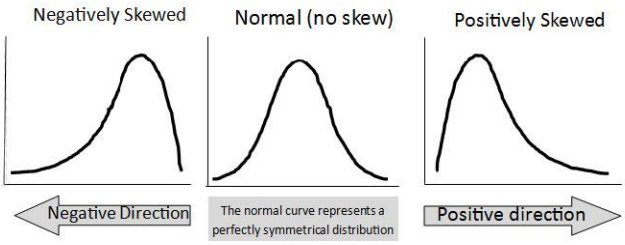
Where;

β_o: is the intercept.

β: is the slope which signifies the correlation between (y) and (x).

ε_i: is the error term. The errors are assumed to be independent from each other and normally distributed with mean equals zero and variance equals σ^2 ; $\varepsilon_i \sim NID(0, \sigma^2)$.

167. **Skewness**: It refers to the lack of symmetry. A distribution is skewed if one tail extends farther than the other. A skewness value close to zero indicates symmetric data, if the skewness value is negative, then the data left skew and if the skewedness value is positive, then this indicates that the data are skewed to the right



168. **Slope & Sensitivity:** In the context of calibration curve, the sensitivity of the developed method is obtained from the slope b of the calibration curve equation. Sensitivity value can be obtained from the slope of the calibration (regression) function. In the linear case of calibration curve (i.e. $y = bX + a + \varepsilon$, the slope is defined as:

$$Slope\ (b) = \frac{\Delta y}{\Delta x} = \frac{change\ in\ the\ response\ variable\ (y)}{change\ in\ the\ predictor-regressor,calibrant-variable\ (x)}.$$

Otherwise, the slope (b) is defined as $\frac{\partial y}{\partial x}$. The higher the slope value, the more sensitive the measuring instrument is. The intercept is given the symbol (a) and the error term is (ε).

169. **Specification**: Document stating requirements, where requirement is a need or expectation that is stated, generally implied, or obligatory. ISO 22514-1:2014 (E)

170. **Specification Limit**: Limiting value stated for a characteristic, where characteristic is the distinguishing feature. ISO 22514-1:2014 (E)

171. **Specificity:** The ability to distinguish the measurand from other substances that coexist in the sample. UPAC (International Union of Pure & Applied Chemistry) is recommending using the term selectivity instead of specificity.

172. **Specification Limits:** Are boundary points that define the acceptable values for an output variable (i.e. for a quality characteristic) of a particular product. These boundary points could be determined by product designers. Specification limits may be two-sided, with upper and lower limits, or one-sided, with either an upper or a lower limit.

173. **Standardizing a Variable:** This involves two actions; centering and scaling. Centering the variable, involves taking the difference between each observation and the man of all observations of the variables. While Scaling involves expressing the centered observations in units of the standard deviation of the observations for the variable. Thus, the usual standardization of the response variable (y) and the predictor variable (x) is as follows:

$$\frac{y_i - \bar{y}}{S_y} \text{ and } \frac{x_i - \bar{x}}{S_x}$$

$\boldsymbol{S_y}$**:** Standard deviation observations of the response variable.

$\boldsymbol{S_x}$**:** Standard deviation of the predictor variable.

174. **Stable Process:** Constant mean process subject only to random causes. This random variation does not necessarily be large or small, within or outside of specification, but rather that the variation is predictable using statistical technique. ISO 22514-1:2014 (E)

175. **Statistic: 1.** Any function of the observations in a sample that does not contain unknown parameters. E.g. the sample mean$\bar{X} = \frac{\sum_{i=1}^{n} x_i}{n}$. **2.** is any quantity whose value can be calculated from sample data. Statistic could also mean the mathematical function used for estimation.

176. **Statistical Significance vs. Practical Significance:** The entire null hypothesis (H_o) was rejected at the selected significance level , this is the statistical significance. Recall that the null hypothesis (H_o) is the hypothesis that we need to refute. Data that shows statistical significance not necessarily implies a practical significance such example like when the sample size gets very large (i.e. n is > >>), then calculated P-value will be less than the level of significance α; consequently the H_o is rejected at the specified level of significance (since P<α). But this statistical significance not necessarily indicates any real life practical significance.

177. **Statistics:** The mathematics of collection, organization and interpretation of numerical data, especially the analysis of population characteristics by inference from sampling.
Student's t-distribution is one from the family of normal distributions for finite samples. Samples obtained from analytical experiments are small in size, so they are studied statistically by the t-distribution rather than the normal distribution which requires very large sample size.

178. **Student's t-test:** (William Sealey Gosset, 1876-1937) is any statistical hypothesis test in which the test statistic follows a *student's t-distribution* when the null hypothesis is supported.

179. **Systematic Measurement Error**: Component of measurement error that is replicate measurements remains constant or varies in a predictable manner.

180. **Target Value:** In the context of process capability, the target value is the ideal value of a process based on the specifications or customer requirements. It is the preferred or reference value of a characteristic stated in a specification. Characteristic is the distinguishing feature (of an item) ISO 22514-1:2014 (E)

181. **Test Result:** The value of a characteristic obtained by carrying out a specified test method. ASTM E2282.

182. **Test Result:** Outcome of an analytical procedure or method.

183. **Test Statistics:** Is used to assess a particular hypothesis in relation to some population. The essential requirement of such a statistic is a known distribution when the null hypothesis (H_o) is true.

184. **Toluene:** Methyl benzene, C_7H_8 with molecular weight 92.13 g/mole. Clear, colorless and highly flammable liquid; odor like benzene with freeze point -94.99 ^oC and boiling point 110.6^oC.

185. **Traceability:** A property of the result of a measurement or the value of a standard whereby it can be related to stated references, usually national or international standards, through an unbroken chain of comparisons all having stated uncertainties. ASTM D4175-16C.

186. **Treatments:** A factor can have different levels referred to as the treatment or factor level. While in the context of analytical chemistry, treatment is referring the all activities that the sample was exposed to prior testing it.

187. **True Quantity Value:** Not known to mankind. All determined values of measurands are an estimates with different degrees of uncertainties.

188. **True Value:** In statistics, the value towards which the average of single results obtained by N laboratories tends, when N becomes very large. ASTM D4175. Also defined by ASTM D E170 as a value of measurand that would be obtained by a perfect measurement. Others defined the value of reference material as true value.

189. **t-Statistic:** It was developed by Gossett and published under the pseudonym "student". Mathematically expressed as:

$$t_{n-1} = \frac{\bar{x}-\mu}{\frac{s_x}{\sqrt{n}}} \text{ with } \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \text{ and } s_x^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}$$

Where the $x_i(i = 1,2,..,n)$ are normally distributed "Gaussian", independently, and identically distributed (NIDD) with mean μ and variance σ^2 :

$$x_i \in NIID\ (\mu, \sigma^2)$$

190. **Type (I) Error:** Denoted "**α**" which is finding something which is not there. Rejecting null hypothesis (H_o) when it is true (consumer's risk).

191. **Type (II) Errors:** Denoted "**β**" which is not finding something which is there. When a false hypothesis is accepted (producer's risk).

192. **Validation Range:** The part of the concentration range of an analytical method that has been subjected to validation.

193. **Validation Samples:** A set of samples with known concentration levels used in validating the calibration model.

194. **Validation: 1.** Where the specified requirements are adequate for an intendent use. ISO/IEC Guide 99:2007. Validation is the process of demonstrating or confirming the performance characteristics of a method of analysis. AOAC Official Methods of Analysis (2012) Appendix K. **2.** Establishment of the performance characteristics of a method and provision of objective evidence that the performance requirements for a specified intended used are fulfilled. ISO16140-1:2016.

195. **Variable Data vs. Attribute Data:** variable data is a characteristics measured on a continuous numerical scale, while the attributes data are characteristics measured as percentage, fractions, or counts of occurrences in a defined interval of time or space. ASTM E2587-10.

196. **Variance**: Is the square of the standard deviation of population (σ) or square of the sample standard deviation (s). The variance is denoted (σ^2) or (s^2). The population variance σ^2 is defined as $\sigma^2 = \frac{\sum_{i=1}^{N}(x_i-\mu)}{N}$ and sample variance $s^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})}{n-1}$ where N is population size, n is the sample size, μ is the population mean, $\bar{x}$ is the sample mean and x_i is the ith observation.

197. **Variation:** is the difference between values of a characteristic. Where characteristic is the distinguishing feature. ISO 22514-1:2014 (E)

198. **Verification**: Demonstration that a validated method functions in the user's hands according to the method's specifications determined in the validation study and is fit for its purpose. IOS16140-1:2016.

199. **Water:** A clear, colorless, odorless and tasteless liquid. Chemically one molecule of water consists of two atoms of hydrogen (H) and one atom of oxygen (O) and given molecular formula H_2O. Freezing point is 0^oC and its boiling point 100^oC. The water specific gravity at 4^oC is 1.000.

200. **Working Range:** Is the difference between the highest value of the measurand (this is the independent variable) and the lowest value of the measruand in the sample. The working range of an analytical method is the concentration range over which results are obtained that are fit for a specific purpose.

201. **Z-Score:** Standardized and dimensional measure of the difference between an individual result in a data set and the arithmetic mean of the date set (by dividing the actual difference from the mean by the standard deviation for the data set). Mathematically it is written as (ASTM D4175-16C):

$$Z = \frac{|x_i - \bar{x}|}{\sigma}$$

Where;

x_i: Observation i

$\bar{x}$: The average of observations in the sample set.

σ : Standard deviation of all observations in the sample.

References

1. UOP991-17. Trace chloride, fluoride and bromide in liquid organics by combustion ion chromatography, ASTM International, West Conshohocken, PA, 2017, www.astm.org.

2. ISO/IEC Guide 99:2007. International Vocabulary of Metrology – Basic and general concepts and associated terms (VIM). www.iso.org .

3. ASTM E456-13, Standard Terminology Relating to Quality and Statistics, ASTM International, West Conshohocken, PA, 2013, www.astm.org.

4. ASTM E59-14, Standard Test Method for Total Fluorine, Chlorine and Sulfur in Aromatic Hydrocarbons and Their Mixtures by Oxidative Pyrohydrolytic Combustion Followed by Ion Chromatography Detection (Combustion Ion Chromatography-CIC), ASTM International, West Conshohocken, PA, 2014, www.astm.org.

5. ASTM D5808-14, Standard Test Method for Determination of Chloride in Aromatic Hydrocarbons and Related Chemicals by Microcoulometr, ASTM International, West Conshohocken, PA, 2014, www.astm.org.

6. ASTM D4929-16, Standard Test Method for Determination of Organic Chloride Content in Crude Oil, ASTM International, West Conshohocken, PA, 2016, www.astm.org.

7. ASTM D5194-13, Standard Test Method for Trace Chloride in Liquid Aromatic Hydrocarbons, ASTM International, West Conshohocken, PA, 2013, www.astm.org.

8. ASTM D7959-15, Standard Test Method for Chloride Content Determination, ASTM International, West Conshohocken, PA, 2015, www.astm.org.

9. Chloride in Petroleum Distillates by Microcoulometry. UOP779-08. www.astm.org.

10. International Conference on Harmonization Of Technical Requirements For Registeration Of Pharmaceuticals For Human Use. Validation of Analytical Procedures. Q2(R1) October 27th,1994. www.ich.org.

11. Michael Thompson, Stephen L. Ellison and Roger Wood (2002). Harmonized Guidelines For Single Laboratory Validation Of Methods Of Analysis.Journal of Pure and Applied Chemistry, 74 (5), 835-855. www.iupac.org.

12. Solid Mineral Fuels -Guidelines for the Validation of Alternative Methods. ISO11726:2017. www.iso.org.

13. General Requirements For The Competence of Testing and Calibration Laboratories. ISO17025:2017. www.iso.org.

14. The Fitness for Purpose of Analytical Methods. A Laboratoy Guide to Method Validation and Realted Topics (2nd ed.2014). www.eurachem.org.

15. Microbiology of the food chaing-Method Validation- Part 2: Protocol for the Validation of Alternative (Proprietary) Methods against a Reference Method. ISO16140-2. International Standardization Organization. www.iso.org.

16. Guideline for Analytical Method Validation. Compressed Gas Association, CGA M-6-2013. www.cganet.com.

17. ASTM E2857-16, Standard Guide for Validating Analytical Methods. ASTM International, West Conshohocken, PA, 2016, www.astm.org.

18. Vicki Barwick, Stephen L. R. Ellison, et al. (2017). Method Validation in Analytical Sciences: Discussions on Current Practice and Future Challenges. Journal of Accreditation and Quality Assurance, 22, 253-263

19. Soumia Belouafa, et al. (2017). Statistical Tools and Approaches to Validate Analytical Methods: Methodology and Practical Examples. International Journal of Metrology and Quality Engineering, 8(9), 1-10.

20. Michael Thomson (2012). Precision in Chemical Analysis: A Critical Survey of Uses and Abuses. Journal of Analytical Methods, 4 ,1598-1611.

21. Klaus Danzer and Lloyd A. Currie (1998). Guideliens for Calibration in Analytical Chemistry. Journal of Pure and Applied Chemistry, 70, 993-1014.

22. Statistical Manual of Association Of Analytica Chemists. W.J. Youden and E.H. Steiner. ISBN: 0-935584-15-3 . www.aoac.org.

23. MINITAB 17. www.minitab.com.

24. A.E. Fridman.The Quality of Measurement , ISBN 978-1-4614-1477-3. Springer (2012)

25. Giovanni Battista Rossi. Measurement and Probabilty. Aprobabilistic Theory of Measurement with Applications. ISBN 978-94-017-824-3. Springer (2014).

26. Semyon G. Rabinovich. Evaluating Measurement Accuracy .A Practical Approach. ISBN 978-1-4419-1455-2. Springer (2010).

27. Franco Pavese, Alistair B. Forbes (Eds.). Data Modeling for Metrology and Testing in Measurement Science. ISBN 13: 978-0-8176-4592-2. Birkhauser (2009). www.birkhauser.com.

28. G.Meinrath , P.Schneider. Quality Assurance for Chemistry and Environmental Science. Metrology from pH Measurement to Nuclear Waste Disposal. ISBN 978-3-540=71271-8. Springer (2007).

29. K.Danzer.Analytical Chemistry. Theoretical and Metrological Fundamentals. ISBN 978-3-540-35988-3. Springer (2007).

30. Francisco Raposo. Evaluation of analytical calibration based on least-squares linear regression for instrumental techniques: A tutorial review. Trends in Analytical Chemistry 77 (2016) 167-185.

31. Guidelines for Calibration in Analytical Chemistry. International Union of Pure and Applied Chemistry. www.iupac.org.

32. Calibration and Evaluation of Analytical Methods and Estimation of Preformance Characteristics. ISO8466:1990. www.iso.org.

33. International Standard on Linear Calibration Using Rreference Materials. ISO 11095:1996 www.iso.org.

34. R.J.Carroll, C.H.Spiegelman and J.Sacks (1988). A Quick and Easy Multiple-Use Calibration-Curve Procedure. Technometrics,30:2,137-141.

35. Luis C. Rodrigues, Jesus L. Sanchez et al (2001). Calibration in Chemical Measurement Processes, part I. Trends in Analytical Chemistry, 20:4,195-206.

36. Luise C. Rodrigues, et al (2001). Calibration in Chemical Measurement Processes, part II. Trends in Analytical Chemistry, 20:11,620-636.

37. ASTM D7578-15, Standard Practice for Calibration Requirements for Elemental Analysis of Petroleum Products and Lubiricants, ASTM International, West Conshohocken, PA, 2015, www.astm.org.

38. Guidelines for the Determination of Calibration Intervals of Measuring Instruments. International Laboratory Accreditation Cooperation (ILAC). www.ilac.org.

39. J.M. Jurado et al (2017). Some Practical Considerations for Linearity Assessment of Calibration Curves as Function of Concentration Levels According to the Fitness-for-Prupose. Talanta, 172, 221-229.

40. J.M.Andrade, M.G. Estevez-Perez (2014). Statistical Comparison of the Slopes of Two Regression Lines: A Tutorial. Analytica Chimica, 838, 1-12.

41. Irma L. and Fraco Magno (2007). A Statistical Overview on Univariate Calibration, Inverse Regression, and Detection Limits. Mass Spectrometry Reviews,26, 1-18.

42. ASTM E3080-17, Standard Practice for Regression Analysis, ASTM International, West Conshohocken, PA, 2016, www.astm.org.

43. V. Centner D. L. Massart. S. de Jong (1998). Inverse Calibration Predicts Better than Classical Calibration. Fresenius Journal of Analytical Chemistry, 361, 2-9.

44. E.Voigtman. Limits of Detection in Chemical Analysis. ISBN 978-1-119-18897-1. Wiley (2017).

45. ASTM E178-16Standard Practice for Dealing With Outlying Observations, ASTM International, West Conshohocken, PA, 2016, www.astm.org.

46. Statistical Interpretation of Data. Part 4: Detection and Treatments of Outliers. ISO 16269. www.iso.org.

47. Frank E. Grubbs (1969). Procedures for Detecting Outlying Observations in Samples. Technometrics,11:1,1-211.

48. Accuracy of Measurement Methods and Results. ISO5725-2. www.iso.org

49. Herman A, Ryan Gottfredson and Harry Joo (2013). Best-Practice Recommendations for Defining, Identifying and Handling Outliers. Orgnaizational Research Methods, 16:2, 270-301.

50. Statistical Interpretation of Data:Tests for Departure From the Normality Distribution. ISO5479. www.iso.org.

51. Eugene Demidenko (2016): The P-Value You Can't Buy.The American Statistician Journal. 70:1,33-38. www.amstat.org.

52. Ronald L Wasserstein and Nicole A. Lazer (2016): The ASA's statement on P-Values: context, process, and prupose. The American Statisticain Journal. 70:2, 129-133 www.amstat.org.

53. ASTM E2586-16, Standard Practice for Calculating and Using Basic Statistics, ASTM International, West Conshohocken, PA, 2016, www.astm.org.

54. Kai Henning Viehweger (Editor) , et al (2007). Practical Ion Chromatography. An Introduction. www.metrohm.com.

55. Leonid D. Asnin (2016). Peak Measuremetn and Calibration in Chromatographic Analysis. Trends in Analytical Chemistry, 81,51-62.

56. Capability of Detection. Methodology in the Linear and Non-Linear Calibration Cases. ISO11843-2:2000(E). www.iso.org.

57. Nick T. Thomopoulos. Statistical Distributions; Applications & Parameters Estimates. ISBN: 978-3-319-65112-5. Springer (2017).

58. M.C.Ortiz et al (2006). Capability of Detection and Three-Way Data. Analytica Chimica Acta,559, 124-136

59. Capability of Detection. Terms and Definitions:1997. ISO11843-1 (E/F). www.iso.org.

60. Capability of Detection. Methodology for Determinatio of the Critical Values for the Response Variable When No Calibration Data are Used . ISO11843-3:2003 (E). www.iso.org.

61. Capability of Detection. Methodology for Comparing the Minimum Detectable Value with a Given Value . ISO11843-4:2003 (E). www.iso.org

62. Capability of Detection. Methodology in the Linear and Non-Linear Calibration Cases. ISO11843-5:2008 (E). www.iso.org

63. Capability of Detection. Methodology for the Determination of the Critical Value and the Minimum Detectable Value in Poisson Distributed Measurements by Normal Approximations. ISO11843-6:2013 (E). www.iso.org

64. Capability of Detection. Methodology Based on Stochastic Properties of Instrumental Noise. ISO11843-7:2012 (E). www.iso.org

65. ASTM E2935-16, Standard Test Method for Conducting Equivalence Testing in Laboratory Applications, ASTM International, West Conshohocken, PA, 2016, www.astm.org.

66. Marion J. Chatfield and Phil J. Borman. (2009). Acceptance Criteria for Method Equivalence Assessments. Analytical Chemistry, 81,9841-9848.

67. A. Bouabidi et al (2012). Usefulness of Capability Indices in the Framework of Analytical Method Validation. Analytica Chimica Acta, 714, 47-56.

68. Statistical Methods in Process Management- Capability and Performance. ISO22514-4:2016. www.iso.org.

69. ASTM E2281-15 Standard Practice for Process Capability and Performance Measurement, ASTM International, West Conshohocken, PA, 2016, www.astm.org.

70. Guidelines for the Determination of Calibration Intervals of Measuring Instruments. International Laboratory Accrediation Cooperation (ILAC). Guidance Series (ILAC-G24). www.ilac.org.

Vitae

Name	:Khalid Saleh AL-Ghamdi
Nationality	:Saudi
Date of Birth	:Born in Dammam, Saudi Arabia 1975.
Email	:ghamks0n@gmail.com
Address	:Dhahran 31311, Street B12. AL-Jami'a Area 7474, Saudi Arabia
Academic Background	:BSc. Chemistry, 1999.
	MSc. Physical Chemistry 2003
	MSc. Applied Statistics 2018
	Chartered Chemist, Royal Society of Chemistry, UK.
	Member of Quality and Statistics Committee of ASTM, USA
	Unit Supervisor at Ras Tanura Refinery QC/QA Laboratory