التقييم الإحصائي للنتائج الكيميائية

١.٣ مقدمة :

إن نتائج التحليل الكيميائي عامة عديمة الفائدة ما لم تكن مقيمة إحصائيا. و عند قياس أي خاصة فيزيائية لا بد أن يكون هناك خطأ محتمل في قياسها و يمكن تقليص هدا الخطأ إلى حد مقبول و لكن لا يمكن تلافيه تماما.

٢.٣ تعريف بعض المصطلحات الإحصائية :

١.٢.٣ الدقة و المصداقية Precision and Accuracy:

٣ ـ ٢ ـ ١ ـ ١ الدقة Precision:

هي قياس للتوافق بين نتائج لقياسات متكررة، و كلما كان هذا الفارق صغيراً كانت الدقة جيدة.

٣ ـ ٢ ـ ١ ـ ٢ المصداقية Accuracy:

هي قياس لمدى قرب قياسات متكررة من القيمة الحقيقية في العينة.

و هناك نوعان من الأخطاء يؤثران على الدقة و المصداقية: الخطأ المنتظم و الخطأ العشوائي.

أ ـ الخطأ المنتظم Determinate error:

أسباب هذا الخطأ هي:

١. عجز في الطريقة المتبعة.
٢. خلل في الجهاز المستخدم للتحليل.
٣. المحلل الكيميائي.

و يؤثر هذا الخطأ على النتيجة في اتجاه واحد أي بعبارة أخرى تكون النتيجة إما أكبر أو أصغر من القيمة الحقيقية.

مثال ١ : في حالة عدم ترسيب المادة ترسيبا كاملا تكون النتيجة دائما وزنا أقل للراسب فيكون الخطأ سلبيا.

مثال ٢: إذا كانت المادة المراد قياسها تحتوي على شوائب تتفاعل مع الكاشف فإن كمية الكاشف المستخدم ستكون أكبر من الكمية المطلوبة فيكون الخطأ إيجابيا.

ملاحظة:

إذا كان الخطأ المنتظم ثابتا: فإنه سيؤثر في المصداقية و لن يؤثر على الدقة.

إذا كان الخطأ المنتظم غير ثابت: فإنه سيؤثر في المصداقية و الدقة معا.

ب ـ الخطأ العشوائي Random error:

مصدر هذا النوع من الخطأ مجهول و لا يمكن التحكم به و لحسن الحظ فإن له قيماً صغيرة تمتاز بالعشوائية و لا يمكن تقديرها بإتباع طرق الأخطاء.

ملاحظة:

يصعب تحديد الخطأ العشوائي و لكنه يحدث تغيراً في القيمة الحقيقية للمادة المراد قياسها سلبا و إيجابا بمقدار واحد أي نسبة الزيادة أو النقصان تكون متساوية حول القيمة الحقيقية.

و تعطى القياسات الكيميائية على شكل الأعداد التالية:

متوسط القيمة المقاسة أو متوسط القيمة.

١. متوسط الانحراف.
٢. الانحراف المعياري.
٣. الانحراف المعياري النسبي.

٣ ـ ٢ ـ ٢ المتوسط Mean و القيمة الوسطية Median:

٣ ـ ٢ ـ ٢ ـ ١ المتوسط The mean:

و هو مجموع القياسات مقسوما على عدد القياسات n:

$$\overline{X} = \frac{x_1 + x_2 + x_3 + \ldots + x}{n}$$

حيث:

$\overline{X}$: المتوسط Mean.

x_1, x_2, x_3...: القياسات الفردية.

n: عدد القياسات.

و يمكن كتابة كيفية حساب المتوسط بطريقة أخرى:

$$\overline{X} = \frac{\sum_{i=1}^{i=n} x_i}{n}$$

حيث $\sum_{i=1}^{i=n} x_i$ (سيجما Sigma) تعني مجموع القياسات الفردية من ١ حتى n.

مثال: أجريت معايرة لقياس نسبة الكلور فوجدت النتائج التالية: 6.87, 6.84 , 6.88, 6.88 , 6.85 , 6.83 احسب المتوسط.

الحل:

$$\overline{X} = \frac{6.87 + 6.84 + 6.88 + 6.88 + 6.85 + 6.83}{6} = 6.86\ \%$$

٣ - ٢ - ٢ - ٢ القيمة الوسطية The Median:

في بعض الحالات تستخدم القيمة الوسطية Median بدلا من المتوسط و تعرف بالقيمة الوسطية لأنها النتيجة التي تتوسط النتائج بمعنى أن نصف النتائج يكون أكبر منها و النصف الآخر أصغر منها من حيث القيمة، هذا في حالة كون عدد النتائج فرديا و في حالة كونه زوجيا يؤخذ المتوسط على أنه متوسط النتيجتين المتوسطتين.

مثال ١: احسب القيمة الوسطية فيما يلي: 2.81 , 3.56, 2.10 , 3.50 , 2.79.

الحل:

- أولا: نرتب النتائج ترتيبا تصاعديا: 2.10 , 2.79 , 2.81 , 3.50 , 3.56
- ثانيا: القيمة الوسطية تساوي النتيجة الموجود في الوسط بعد الترتيب و هي 2.81.

مثال ٢: احسب القيمة الوسطية فيما يلي: 5.99 , 6.01 , 5.90 , 5.51.

الحل:

أولا: نرتب النتائج ترتيبا تصاعديا: 5.51 , 5.90 , 5.99 , 6.01

ثانيا: نحسب القيمة الوسطية: التي تساوي متوسط النتيجتين المتوسطتين:

$$\text{median} = \frac{5.90 + 5.99}{2} = 5.95$$

٣.٢.٣ الانحراف المعياري Standard deviation **والانحراف المعياري النسبي** Relative standard deviation**:**

٣ـ٢ـ٣ـ١ الانحراف المعياري:

يحسب الانحراف المعياري sd كما يلي:

$$\text{sd} = \sqrt{\frac{\sum_{i=1}^{i=n}(x_i - \overline{X})^2}{n-1}}$$

مثال: احسب الانحراف المعياري فيما يلي: 6.87, 6.84 , 6.88, 6.88 , 6.85 , 6.83.

الحل:

أولا: نحسب $\sum_{i=1}^{i=n}\left(x_i - \overline{X}\right)^2$:

$$\left(x_1 - \overline{X}\right)^2 = (6.87 - 6.86)^2 = (0.01)^2 = 1\times 10^{-4}$$

$$\left(x_2 - \overline{X}\right)^2 = (6.84 - 6.86)^2 = (0.02)^2 = 4\times 10^{-4}$$

$$\left(x_3 - \overline{X}\right)^2 = (6.88 - 6.86)^2 = (0.02)^2 = 4\times 10^{-4}$$

$$\left(x_4 - \overline{X}\right)^2 = (6.88 - 6.86)^2 = (0.02)^2 = 4\times 10^{-4}$$

$$\left(x_5 - \overline{X}\right)^2 = (6.85 - 6.86)^2 = (0.01)^2 = 4\times 10^{-4}$$

$$\left(x_6 - \overline{X}\right)^2 = (6.83 - 6.86)^2 = (0.03)^2 = 9\times 10^{-4}$$

$$\sum_{i=1}^{i=n}\left(x_i - \overline{X}\right)^2 = 23\times 10^{-4}$$

ثانيا: نحسب الانحراف المعياري sd:

$$\text{sd} = \sqrt{\frac{23\times 10^{-4}}{6-5}} = \sqrt{\frac{0.0023}{5}} = 0.02$$

٣ ـ ٢ ـ ٣ ـ ٢ الانحراف المعياري النسبي Relative Standard Deviation:

يعبر الانحراف المعياري النسبي عن دقة نتائج التحليل و غالبا ما يستخدم بدلا من الانحراف المعياري. يحسب الانحراف المعياري النسبي rds كما يلي:

$$rsd=\frac{sd}{\overline{\overline{X}}}\times 100\%$$

مثال:احسب الانحراف المعياري النسبي فيما يلي: 6.87, 6.84 , 6.88, 6.88 , 6.85 , 6.83.

الحل: sd يساوي 0.02 و المتوسط يساوي 6.86 كما رأينا في المثال السابق.

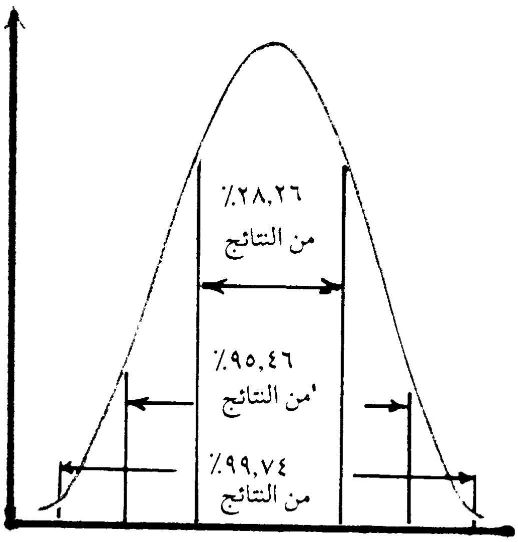
$$rsd=\frac{0.02}{6.86}\times 100$$

$$rsd=0.29\%$$

٤.٢.٣ منحنى التوزيع الطبيعي للنتائج Normal distribution:

نفترض أننا حللنا عنصراً ما في عينة ما عدة مرات (مثلا ١٠٠ مرة) باستخدام طريقة معينة. فإذا رسمنا العلاقة بين التردد (التواتر، التكرر) أي عدد القياسات التي تقع بنفس المقدار بالمقابل مع القياسات حصلنا على منحنى يسمى بمنحنى الخطأ الطبيعي Normal distribution أو منحنى جوسان Gaussian Curve أو المنحنى الجرسي Bell-shaped distribution (الشكل ١).

يبين هذا المنحنى أن الانحراف الناتج عن الخطأ العشوائي (نفترض أن ليس هناك خطأ منتظم) يتوزع بشكل متماثل حول القيمة الحقيقية (الانحراف = صفر) و ذلك لأن المتوسط يؤخذ في هذه الحالة على أنه القيمة الحقيقية و أن احتمال حدوث الانحراف الإيجابي يساوي احتمال حدوث الانحراف السلبي.



الشكل (٣ ـ ١): شكل المنحنى الطبيعي

لكل قياس فردي يمكن حسب الخطأ بالمعادلة التالية:

الخطأ = القيمة المرصودة (المقاسة) - القيمة الحقيقية

يلاحظ من الشكل (١) أن شكل المنحنى متناظر أي أنه يوجد مقابل كل خطأ موجب خطأ سالب له نفس القيمة المطلقة. كما توجد نسبة عالية لتكرار القياسات ذات الخطأ البسيط حيث أن أكثر من ٦٨٪ من القياسات تقع في المجال $\overline{X}$+sd و نسبة بسيطة لتكرار القياسات ذات الخطأ الكبير حيث ٩٩,٧٤٪ من القياسات تقع في المجال $\overline{X}$+3sd. و من أهم استعمالات منحنى التوزيع الطبيعي أنه يمكننا من معرفة جودة المتوسط للنتائج.

مثال:

إذا كان المتوسط يساوي ١٠,٠٠٪ و الانحراف المعياري يساوي ٠,٦٥٪ فإن ثلثي النتائج ينبغي أن تقع في المجال ١٠,٠٠+ ٠,٦٥٪ إذا كان توزيع النتائج طبيعيا.

٣.٢. طرق المقارنة بين نتائج التحليل الكيميائي:

هناك العديد من الاختبارات الإحصائية للمقارنة بين نتائج التحليل الكيميائي و منها :اختبار F و اختبار t.

٣ ـ ٣ ـ ١ اختبار F   F-test:

يستخدم اختبار F لمعرفة هل هناك فرق في الدقة بين طريقتين للتحليل أو محللين و تستخدم المعادلة التالية لهذا الغرض:

$$F = \frac{S_1^{\,2}}{S_2^{\,2}}$$

حيث s يساوي الانحراف المعياري. و بما أن F≥1 (انظر إلى الجدول ١) فيجب أن يكون الانحراف المعياري الذي يوضع في البسط أكبر من ذلك الذي يوضع في المقام.

بعد حساب قيمة F نرجع إلى جدول قيم F (الجدول ١) عند مستوى الثقة المطلوب حيث أن V_1 تمثل درجة حرية البسط و V_2 درجة حرية المقام علما بأن درجة الحرية = عدد العينات – ١ (1 (n –1).

الجدول (٣ ـ ١): قيم F عند حدود الثقة ٩٥٪

V1/ V2	٢	٣	٤	٥	٦	٧	٨	٩	١٠
٢	٩.٢٠	١٩.٢٠	١٩.٢٠	١٩.٣٠	١٩.٣٠	١٩.٤٠	١٩.٤٠	١٩.٤٠	١٩.٤٠
٣	٩.٥٥	٩.٢٨	٩.١٢	٩.٠١	٨.٩٤	٨.٨٩	٨.٨٥	٨.٨١	٨.٧٩
٤	٦.٩٤	٦.٥٩	٦.٣٩	٦.٢٦	٦.١٦	٦.٠٩	٦.٠٤	٦.٠٠	٥.٩٦
٥	٥.٧٩	٥.٤١	٥.١٩	٥.٠٥	٤.٩٥	٤.٨٨	٤.٨٢	٤.٧٧	٤.٧٤
٦	٥.١٤	٤.٧٦	٤.٥٣	٤.٣٩	٤.٢٨	٤.٢١	٤.١٥	٤.١٠	٤.٠٦
٧	٤.٧٤	٤.٣٥	٤.١٢	٣.٩٧	٣.٨٧	٣.٧٩	٣.٧٣	٣.٦٨	٣.٦٤
٨	٤.٤٦	٤.٠٧	٣.٨٤	٣.٦٩	٣.٥٨	٣.٥٠	٣.٤٤	٣.٣٩	٣.٣٥
٩	٤.٢٦	٣.٨٦	٣.٦٣	٣.٤٨	٣.٣٧	٣.٢٩	٣.٢٣	٣.١٨	٣.١٥
١٠	٤.١٠	٣.٧١	٣.٤٨	٣.٣٣	٣.٢٢	٣.١٤	٣.٠٧	٣.٠٢	٢.٩٨

مثال: قدر تركيز الرصاص في عينة بطريقة قياسية معروفة و طريقة قياسية جديدة و حصلنا على النتائج التالية:

الطريقة القياسية الجديدة ppm	الطريقة القياسية المعروفة ppm
١٢٧	١٢٩
١٢٥	١٣١
١٢٦	١٣٠
١٢٩	١٢٧
١٣١	١٢٥
١٣٠	١٢٨
١٢٣	-

هل دقة الطريقة الجديدة تختلف بشكل واضح عن الطريقة القياسية المعروفة (ثقة ٩٥٪)؟

الحل:

$$\overline{X_1} = 127$$

$$\overline{X_2} = 128$$

$$S_1^{\,2} = \frac{\sum_{i=1}^{i=n}\left(x_{i1} - \overline{X_1}\right)^2}{7-1} = \frac{50}{6} = 8.3$$

$$S_2^{\,2} = \frac{\sum_{i=1}^{i=n}\left(x_{i2} - \overline{X_2}\right)^2}{6-1} = \frac{24}{5} = 4.8$$

$$F = \frac{8.3}{4.8} = 1.73$$

قيمة F في الجدول تساوي ٤,٩٥ (٩٥٪ ثقة ، V1=7 - 1 و V2 =6 - 1). و بما أن قيمة F المحسوبة أصغر من قيمة F الجدولية، فيمكننا الحكم بثقة ٩٥٪ بأن الطريقة الجديدة لا تختلف من حيث الدقة بشكل واضح عن الطريقة القياسية المعروفة و أن الانحراف المعياري في كل الطريقتين ناتج عن أخطاء عشوائية.

٣ ـ ٣ ـ ٢ اختبار t:

يستخدم اختبار t للمقارنة بين مصداقية طريقتين للتحليل أو نتائج محللين و في هذه الحالة تستخدم المعادلات التالية:

أ ـ إذا كانت القيمة الحقيقية معروفة تستعمل المعادلة التالية:

$$\pm t = \left(\overline{X} - \mu\right)\frac{\sqrt{n}}{s}$$

حيث:

$\overline{X}$: المتوسط.

μ: القيمة الحقيقية.

s: الانحراف المعياري.

ب ـ إذا كانت μ غير معروفة و $\overline{X_1}$ و $\overline{X_2}$ معروفين، تحسب t كما يلي:

$$\pm t = \frac{\overline{X_1} - \overline{X_2}}{S}\sqrt{\frac{n1 \times n2}{n1 + n2}}$$

حيث:

$\overline{X}$: المتوسط.

n: عدد القياسات.

s: الأنحراف المعياري المشترك.

و يحسب s كما يلي:

$$s = \sqrt{\frac{\sum_{i=1}^{i=n}(x_{i1} - \overline{X}_1)^2 + \sum_{i=1}^{i=n}(x_{i2} - \overline{X}_2)^2}{n_1 + n_2 - 2}}$$

مثال ١:

قمت بتحليل النحاس في مادة قياسية تركيزها معلوم ppm ١١,٧ ، و كررت التحليل خمس مرات وكان متوسطها يساوي ppm ١٠,٨ و الانحراف المعياري يساوي s = ± 0.7 ppm. ما هي مصداقية نتائجك بثقة ٩٥٪؟ هل الخطأ عشوائي أم منتظم؟

الحل:

$$\pm t = (10.8 - 11.7)\frac{\sqrt{5}}{0.7}$$
$$\pm t = 2.9$$

من الجدول (٢) نجد بأن t الجدولية تساوي ٢,٧٧٦ (٩٥٪ ، $n - 1 = 4$) فنحكم بثقة ٩٥٪ أن طريقتك لم تعط النتيجة الصحيحة أي أنها غير صادقة بسبب وجود خطأ منتظم أثناء التحليل.

مثال ٢ : قمت بتقدير الكالسيوم في عينة من التربة عدة مرات بطريقة جديدة و طريقة قياسية معروفة وحصلت على النتائج التالية:

الطريقة الجديدة (٪)	الطريقة القياسية المعروفة (٪)
٢٠,١٠	١٨,٨٩
٢٠,٥٠	١٩,٢٠
١٨,٦٥	١٩,٠٠
١٩,٢٥	١٩,٧٠
١٩,٤٠	١٩,٤٠
١٩,٩٩	-

هل هناك فرق إحصائي واضح بين دقة و مصداقية الطريقتين عند مستوى الثقة ٩٥٪؟

الحل:

أولا : الطريقة الجديدة:

$$\sum_{i=1}^{i=n}\left(x_{i1} - \overline{X}_1\right)^2 = 2.262$$
$$s_1^{\ 2} = \frac{2.262}{5} = 0.452$$

ثانيا: الطريقة القياسية المعروفة:

$$\sum_{i=1}^{i=n}\left(x_{i2}-\overline{X_2}\right)^2=0.420$$
$$s_2{}^2=\frac{0.420}{4}=0.105$$

نحسب F:

$$F=\frac{0.425}{0.105}=4.30$$

قيمة F المحسوبة (٤,٣٠) أقل من قيمة F الجدولية (٦,٢٦)، لذا فإن الطريقتين متشابهتان من حيث الدقة أي أن الانحراف المعياري لكل منهما متقرب.

لاختبار مصداقية الطريقتين نطبق اختبار t.

أولا: نحسب S.

$$S=\sqrt{\frac{2.262+0.420}{6+5-2}}$$
$$S=0.546$$

و من ثم نحسب t:

$$\pm t=\frac{19.65-19.24}{0.546}\sqrt{\frac{6\times 5}{6+5}}$$
$$\pm t=1.23$$

نجد أن قيمة t الجدولية (ثقة = ٩٥٪ و درجة الحرية $n_1 + n_2 - 2 = 9$) تساوي ٢,٢٦ و بما أن t المحسوبة أصغر من t الجدولية فيمكن أن نحكم أن الطريقتين لا تختلفان من حيث المصداقية أي أن متوسطهما متقارب و الفرق بينهما ناتج عن خطأ عشوائي.

الجدول (٣ ـ ٢): قيم t عند حدود ثقة مختلفة

درجات الحرية (n-1)	حدود الثقة		
	٩٠٪	٩٥٪	٩٩٪
١	٦.٣١٤	١٢.٧٠٦	٦٣.٦٥٧
٢	٢.٩٢٠	٤.٣٠٣	٩.٩٢٥
٣	٢.٣٥٣	٣.١٨٢	٥.٨٤١
٤	٢.١٣٢	٢.٧٧٦	٤.٦٠٤
٥	٢.٠١٥	٢.٥٧١	٤.٠٣٢
٦	١.٩٥٣	٢.٤٤٧	٣.٧٠٧
٧	١.٨٩٥	٢.٣٦٥	٣.٤٩٩
٨	١.٨٦٠	٢.٣٠٦	٣.٣٥٥
٩	١.٨٣٣	٢.٢٦٢	٣.٢٥٠
١٠	١.٨١٢	٢.٢٢٨	٣.١٦٩

أسئلة:

قام محللان بتقدير عنصر الكوبلت في سبيكة جديدة و حصلا على النتائج التالية:

المحلل ١	المحلل ٢
٤٥.٢١	٤٦.٠٦
٤٥.٩٢	٤٥.٩٥
٤٥.٧٥	٤٥.٨٧
٤٥.٠٢	٤٥.٢١
٤٤.٩٩	٤٦.١٤

هل هناك فرق بين دقة و مصداقية المحللين عند مستوى الثقة ٩٥٪؟

٣ ـ ٤ استبعاد النتيجة الشاذة (اختبار Q) Q test:

عند تكرار التحليل عدة مرات (من ٣ الى ١٠) و لما تظهر لنا نتيجة مختلفة بشكل كبير من النتائج الأخرى، يستعمل اختبار Q لكي نقرر هل يمكن استبعادها أو الاحتفاظ بها و نحسب Q كما يلي:

أ ـ اذا كانت النتيجة الأصغر هي الشاذة ترتب النتائج تصاعديا و يحسب Q:

$$Q = \frac{X_2 - X_1}{X_n - X_1}$$

ب ـ اذا كانت النتيجة الأكبر هي الشاذة (أكبر نتيجة هي X_n):

$$Q = \frac{X_n - X_{n-1}}{X_n - X_1}$$

مثال: حصلنا على النتائج المكررة التالية: ١٠٣، ١٠,٠ ، ٩,٨٢ ، ١٠,٩. هل يمكن الأحتفاظ بالنتيجة ١٠,٩ عند ثقة ٩٥٪؟

الحل: أولا نرتب النتائج تصاعديا: ٩,٨٢ ، ١٠,٠ ، ١٠,٣ ، ١٠,٩.

$$Q = \frac{10.9 - 10.3}{10.9 - 9.82} = \frac{0.60}{1.08} = 0.56$$

بما أن القيمة الجدولية (ثقة = ٩٥٪ و $n = 4$) تساوي ٠,٨٢٩ (أكبر من القيمة المحسوبة ٠,٥٦) نستنتج بأنه يمكن الاحتفاظ بنتيجة ١٠,٩ عند مستوى الثقة ٩٥٪ و هناك ٩٥٪ احتمال أن الخطأ هو خطأ عشوائي.

ملاحظة:

١. لا يمكن تطبيق اختبار Q في حالة ثلاثة قياسات عندما يكون اثنان منها متساويين لأن النتيجة الثالثة ستستبعد مهما كانت قيمتها

٢. إذا أكد اختبار Q على الاحتفاظ على النتيجة الشاذة فيكون استخدام المتوسط بدلا من المتوسط لأنه لا يتضمن النتيجة الشاذة.

الجدول (٣ ـ ٣): قيم Q عند حدود ثقة مختلفة

n	%٩٠	%٩٥	%٩٩
٣	٠.٩٤١	٠.٩٧٠	٠.٩٩٤
٤	٠.٧٦٥	٠.٨٢٩	٠.٩٢٦
٥	٠.٦٤٢	٠.٧١٠	٠.٨٢١
٦	٠.٥٦٠	٠.٦٢٥	٠.٧٤٠
٧	٠.٥٠٧	٠.٥٦٨	٠.٦٨٠
٨	٠.٤٦٨	٠.٥٢٦	٠.٦٣٤
٩	٠.٤٣٧	٠.٤٩٣	٠.٥٩٨
١٠	٠.٤١٢	٠.٤٦٦	٠.٥٦٨

٥.٣ مراقبة الجودة Quality control:

مراقبة الجودة هي عبارة عن تخطيط منهجي يحتوي على عمليات و برامج هدفها مراقبة مدى دقة و مصداقية النتائج للحصول في النهاية على نتائج ذات جودة عالية. و يؤخذ في الاعتبار في عملية مراقبة الجودة الآتي:

١. تقدير حد الاكتشاف Detection limit أو الحساسية Sensitivity.
٢. تقدير الدقة في فئة batch من العينات و بين فئات من العينات Between batches و لهذا الغرض تستخدم عينات مزدوجة Replicates samples تدخل عشوائيا في فئة من العينات.
٣. تقدير المصداقية بتحليل مادة قياسية معترف بها Certified reference material.
٤. سجل للنتائج.