Chapitre 1 : Paramètres statistiques de base.

I) Introduction

Il est impossible d'effectuer une analyse chimique qui soit totalement exempte d'erreurs ou d'incertitudes. Tout ce qu'on peut espérer est de minimiser ces erreurs et d'estimer leur grandeur d'une manière acceptable.

La nature des erreurs expérimentales ainsi que leurs conséquences sur les résultats des analyses font l'objet de ce chapitre.

II-Paramètres statistiques de base

II.1) Valeur centrale, justesse et fidélité d'un ensemble de mesures

Quand on répète une mesure faite sur un même échantillon, on obtient des valeurs légèrement différentes. Dans ce cas on estime que pour calculer le résultat final il est préférable de se baser sur:

- La moyenne arithmétique de ces n mesures, désignée par la valeur centrale $\bar{X}$ ou moyenne.

- $$\bar{X} = \frac{X_1 + X_2 \ldots\ldots X_n}{n} = \sum_1^n \frac{X_i}{n} \qquad (1.1)$$

- la médiane (rarement utilisée): après avoir classé les valeurs expérimentales par ordre croissant, on extrait celle qui est située au milieu de la série, sauf si le nombre de valeurs est pair auquel cas on prend la moyenne des deux valeurs situées au centre. La médiane a l'avantage de ne donner aucun poids particulier à une valeur aberrante.

Exemple: Si les valeurs de mesure d'une grandeur sont:
12,01- 12,03- 12,05- 12,68, la moyenne arithmétique est 12,19 et la médiane 12,04.
La médiane est une valeur meilleure que la moyenne arithmétique, car elle ne tient pas compte de la valeur 12,68 qui apparaît anormale (dispersion de 0,67).

II.2) Valeur vraie, erreur absolue et relative

On considère que chaque mesure X_i représente la somme de la valeur vraie X_0 et d'une erreur expérimentale absolue E_a (ΔX). L'erreur absolue E_a sur la mesure i est donc exprimée par :

$$E_a = X_i - X_0 \qquad (1.2)$$

L'exactitude (justesse, erreur absolue) d'un résultat est définit par E_a, plus E_a (ΔX) est faible et plus il y a de l'exactitude sur le résultat. L'***Exactitude*** reflète, à quel point la valeur mesurée est en accord avec la vraie valeur ou la valeur acceptée. En d'autres termes, à quel point la valeur est correcte? Très souvent, la valeur vraie ou acceptée d'une grandeur physique n'est pas connue, ce qui fait qu'il est parfois impossible de déterminer l'exactitude d'une mesure.

L'erreur relative (E_R) sur une mesure (ou sur la valeur centrale) correspond au quotient de la valeur absolue de l'écart correspondant $|E_a|$, par la valeur vraie, E_r peut s'exprimer en % ou en unité de la mesure.

$$E_r = \frac{X - X_0}{X_0} \qquad (1.3)$$

Si on ne connaît pas X_0, ce qui est en général le cas en analyse chimique, on calcule **l'erreur expérimentale** de la mesure i, soit Ei (ou ΔX_i), en remplaçant dans l'équation (1.2), X_0 par la moyenne $\bar{X}$:

$$E_i = X_i - \bar{X} \qquad (1.4)$$

E_i représente l'écart algébrique entre la moyenne et la i^{ème} mesure. L'erreur expérimentale moyenne $\bar{d}$, ou moyenne des écarts, calculée sur les n mesures, permet d'apprécier **la fidélité.**

$$\bar{d} = \frac{\sum |X_i - \bar{X}|}{n} \qquad (1.5)$$

III) Types d'erreurs dans les données expérimentales

On peut distinguer deux types principaux d'erreurs:

- Les erreurs aléatoires (ou indéterminées)
- Les erreurs systématiques (ou déterminées)

III.1) Erreurs systématiques

Les erreurs systématiques ont une valeur définie, on peut leur attribuer une origine et elles sont du même ordre de grandeur et de même signe pour une série de mesures effectuées de la même manière.

III.1.1) : les sources d'erreurs systématiques

Il ya trois types d'erreurs systématiques

✓ **Erreurs Instrumentales :** tous les dispositifs de mesures sont à l'origine d'erreurs systématiques. Ainsi les pipettes, les burettes et les fioles Jaugées peuvent contenir ou délivrer des volumes légèrement différents de ce qu'indiques leurs graduations.

✓ **Erreurs dues à la méthode :** le comportement chimique ou physique non idéal des réactifs et des réactions sur lequel repose une analyse peut être à l'origine d'erreurs systématiques dues à la méthode. Ces causes de non idéalité comprennent la lenteur de certaines réactions, le fait que d'autres ne soient pas complètes, l'instabilité de certaines espèces et l'existence éventuelle de réaction secondaires qui interférent avec la mesure.

Exemple : une erreur courante dans les méthodes volumétriques résulte de petit excès de réactif nécessaire pour qu'un indicateur subisse le changement de couleur qui signale que la réaction est complète. L'exactitude de ce type d'analyse est donc limitée par le phénomène même qui rend le titrage possible.

✓ **Erreurs personnelles :** beaucoup de mesures nécessitent des jugements personnels, comme l'estimation de la position d'une aiguille entres deux graduations, de la couleur d'une solution au point de fin de titrage……..Des jugements de ce types font souvent l'objet d'erreurs systématiques individuelles.

Remarque : Parmi les trois types d'erreurs systématiques rencontrées dans une analyse chimique, les erreurs dues à la méthode sont habituellement les plus difficiles à détecter et à corriger.

III.1.2) L'effet des erreurs systématiques sur les résultats analytiques :

Les erreurs systématiques peuvent être constantes ou proportionnelles.

- **Erreurs constantes :** sont d'autant plus gênantes que la grandeur mesurée est petite.

Exemple : Supposez qu'on perdre 0,5mg de précipité lorsqu'on le traite avec 200ml de liquide de lavage.

Si le précipité pèse 500mg, l'erreur relative due à la solubilité est égale à : $E_r = -(0,5/500) \times 100\% = -0,1\%$

Si l'on perd la même quantité à partir de 50mg de précipité don : $E_r = -1\%$.

- **Erreurs proportionnelles :** la cause fréquente d'erreurs proportionnelle est la présence de contaminants interférant dans l'échantillon.

Exemple : une méthode largement utilisée pour le dosage de cuivre est basée sur la réaction de l'ion cuivre(II) avec l'iodure de potassium pour donner de l'iode. On mesure ensuite la quantité d'iode qui est proportionnelle à la quantité du cuivre dans l'échantillon. Si l'ion fer (III) est présent, il peut aussi réagir avec l'iodure de potassium pour libérer de l'iode

Remarque : les erreurs constantes sont indépendantes de la taille de l'échantillon, les erreurs proportionnelles diminuent ou augmentent selon la taille de l'échantillon.

III.2) Les erreurs aléatoires dans les analyses :

Les erreurs aléatoires, ou indéterminées, apparaissent lorsqu'un système de mesure est poussé à son maximum de sensibilité, ce type d'erreur est causé par les nombreux paramètres incontrôlables qui font inévitablement partie de toute mesure physique ou chimique. Seul ce type d'erreurs qui est pris en considération par l'étude statistique.

III.2.1 Traitement statistique de l'erreur aléatoire

Les erreurs aléatoires, qui affectent les résultats d'une analyse peuvent être évaluées par des méthodes statistiques.

III.2.1.1. L'échantillon et la population

En statistique, on appelle échantillon de données un nombre fini d'observations expérimentales. Les statisticiens appellent le nombre de données infini théorique une population ou un univers de données. Les lois statistiques ont été établies pour une population de données ; elles doivent souvent être sérieusement modifiées quand on les appliqué à un petit échantillon parce qu'un petit nombre de données ne peut plus être représentatif de la population.

III.2.1.2. Définition des termes utilisés en statistique

✓ ***Moyenne de la population μ et Moyenne de l'échantillon $\bar{X}$:***

(Lorsque N→∞) moyenne de la population

$$\mu = \frac{\sum_{i=1}^{N} X_i}{N} \qquad (1.6)$$

(Lorsque N est fini) la moyenne de l'échantillon

$$\bar{X} = \frac{\sum_{1}^{N} x_i}{N} \qquad (1.7)$$

✓ ***Ecart-type*** **:** l'écart type est une mesure de dispersion la plus couramment utilisée en statistique lorsqu'on emploi la moyenne pour calculer une tendance centrale. Il mesure donc la dispersion autour de la moyenne.

✓ ***Ecart type de la population :***

(Théorique) N→∞

$$\sigma = \sqrt{\frac{\sum_{i}^{N} (X_i - \mu)^2}{N}} \qquad (1.8)$$

✓ *Ecart type de l'échantillon :*

$$S = \sqrt{\frac{\sum_i^N (X_i - \bar{X})^2}{N-1}} \qquad (1.9)$$

Propriétés de l'écart –type :

- On utilise l'écart-type que pour mesurer la dispersion autour de la moyenne d'un ensemble de données, l'écart-type est parfois appelée ***flctuation***. Notez que, plus la valeur de l'écart type est grande plus les données sont dispersées autour de la moyenne.
- L'écart-type n'est jamais négatif.
- L'écart-type est sensible aux valeurs aberrantes.
- L'écart-type est nul si toutes les valeurs d'un ensemble de données sont les mêmes (parce que chaque valeur est égale à la moyenne).
- Dans le cas des données ayant approximativement la même moyenne, plus la dispersion est grande, plus l'écart-type est grand.

On peut utiliser l'écart-type parallèlement à la moyenne pour calculer des intervalles de données.

Environ 68% des données se situent à l'intérieur de l'intervalle :

$\bar{X} - S < X < \bar{X} + S$

$$95\% \text{ des données dans l'intervalle: } \bar{X} - 2S < X < \bar{X} + 2S$$

$$99\% \text{ des données dans l'intervalle: } \bar{X} - 3S < X < \bar{X} + 3S$$

La question la plus simple et la plus fréquemment posée est: "Quelle est la valeur typique représentant le mieux les mesures expérimentales, et à quel point est-elle fiable ?

Considérez un ensemble de *N (*=9) mesures d'une certaine propriété (ex., une masse) placées en ordre croissant (c.-à-d., x_1, x_2, x_3, x_4, x_5, x_6. x_7.x_8 et x_9). Plusieurs méthodes utiles et simples permettant de déterminer la valeur la plus probable et son intervalle de confiance et de comparer des résultats de ce type sont disponibles.

Cependant, lorsque le nombre de mesures ***N*** disponibles est restreint, l'utilisation de la ***médiane*** plutôt que de la ***moyenne*** est souvent plus appropriée. En plus de l'***écart-type***, l'***étendue*** est aussi utilisée pour décrire la dispersion dans un ensemble de mesures ou d'observations. L'***étendue*** est tout simplement la différence entre la plus grande et la plus petite des valeurs ou observations d'un ensemble de données.

L'étendue = X_{max}-X_{min}, ou X_{max} et X_{min} sont, respectivement, la plus grande et la plus petite observation d'un ensemble de données.

La *médiane* est définie comme la valeur partageant en deux l'ensemble ordonné de **N** observations, c.à.d c'est le point central dans un jeu ordonné de données. Si ***N*** est **impair**, alors (***N*-1**)/2 mesures sont plus petites que la ***médiane***, et la valeur immédiatement supérieure est appelée la médiane (c.-à-d., la ***médiane*** est le point central de cet ensemble de données). Dans l'illustration ci-dessus, la 5ème mesure (c.-à-d., x_5) serait la ***médiane***.

Si l'ensemble de données contient un nombre **pair** de points, la ***médiane*** sera la moyenne des deux points centraux.

Exemple: Pour N = 6 et x = 2, 3, 3, 5, 6, 7;

La ***médiane*** = (3+5)/2 = 4;

La ***moyenne*** = (2 + 3 + 3 + 5 + 6 + 7)/6 = 4.33; et l'***étendue*** = (7 – 2) = 5.

Remarque : La *médiane* peut ainsi servir à vérifier la moyenne calculée. Dans les échantillons où les erreurs sont distribuées uniformément autour de la moyenne, la ***moyenne*** et la ***médiane*** auront la même valeur. Souvent l'***écart-type relatif*** est plus utile, dans un sens pratique, que l'***écart-type*** car il donne immédiatement une idée du niveau de précision de l'ensemble de données relativement aux valeurs individuelles.

✓ **La variance (S^2):**

$$S^2 = \frac{\sum_1^N (X_i - \overline{x})^2}{N-1} \qquad (1.10)$$

Notez que l'écart-type a les mêmes unités que les données. Alors que la variance a les unités des données élevées au carré.

Ecart-type relatif (S_r) et coefficient de variation(CV)

Les chimistes utilisent souvent l'écart-type relatif plutôt que l'écart-type absolu. Il est souvent exprimé en parties pour mille (‰) ou en pour-cent (%).

$$S_r = \left({}^{S}/_{\overline{X}}\right) x\ 1000 \qquad (1.11)$$

Lorsque l'écart-type relatif est multiplié par 100 (%) on l'appelle le coefficient de variation

$$CV\ (\%) = \left({}^{S}/_{\overline{X}}\right) x\ 100 \qquad (1.12)$$

✓ Erreur-type de la moyenne

$$S_m = {}^{S}/_{\sqrt{N}} \qquad (1.13)$$

La moyenne et l'écart-type d'un ensemble de résultats sont d'une importance primordiale dans tous les domaines de la science et des techniques. La moyenne est importante car elle constitue habituellement la meilleure estimation du paramètre recherché. L'erreur-type de la moyenne est tout aussi importante car elle fournit les informations relatives à la précision et à l'erreur-indéterminée associée aux résultats.

IV) Intervalle de confiance de la moyenne

Quand le nombre N de mesures est petit, (N compris entre 4 et 15 par exemple), et qu'il n'y a pas d'erreur systématique, la moyenne vraie **m** peut être assez différente de la valeur de la moyenne $\overline{X}$. On est donc réduit à faire son estimation en calculant un intervalle de confiance à l'intérieur duquel on se donne une probabilité que l'on s'impose (par exemple 95 %), que la moyenne vraie s'y trouve. Cette opération entraîne un risque d'erreur.

Les limites de confiance pour la moyenne $\overline{X}$ de N mesures peuvent être tirées de **t** par l'équation suivante :

$$LC(\%)\ pour\ m = \overline{X} \pm \frac{t.S}{\sqrt{N}} \qquad (1.14)$$

t : Paramètre de Student est un facteur statistique qui dépend de N et du niveau de confiance choisi (tableau 1).

Tableau 1 : Valeurs du coefficient de Student pour divers degrés de probabilité

N	**niveau de confiance**				
	80%	90%	95%	99%	99,9%
2	3.08	6.31	12.7	63.7	637
3	1.89	2.92	4.30	9.92	31.6
4	1.64	2.35	3.18	5.84	12.9
5	1.53	2.13	2.78	4.60	8.6
6	1.48	2.02	2.57	4.03	6.86
7	1.44	1.94	2.45	3.71	5.96
8	1.42	1.90	2.36	3.50	5.4
9	1.40	1.86	2.31	3.36	5.04
10	1.38	1.83	2.26	3 .25	4.78
11	1.37	1.81	2.23	3.17	4.59
12	1.36	1.80	2.20	3.11	4.44
13	1.36	1.78	2.18	3.06	4.32
14	1.35	1.77	2.16	3.01	4.22
15	1.34	1.76	2.14	2.98	4.14
∞	1.29	1.64	1.96	2.58	3.29

Remarque : Rappelez- vous que le nombre de degrés de liberté n'est pas égal à N, mais à N-1