

ESTIMATION DE LA FIABILITÉ DES RÉSULTATS

Aucune mesure n'est exacte. Toute valeur mesurée comporte une certaine erreur. Le mieux que l'on puisse faire, c'est de déterminer la marge d'erreur. Plus la mesure est faite avec précision, plus la marge d'erreur est restreinte. Les erreurs de mesure sont de deux ordres : *a*) l'erreur de lecture par l'observateur, *b*) la marge d'erreur inhérente à l'instrument lui-même.

Lorsque l'on donne les résultats d'une mesure, il faut indiquer au moins la marge d'erreur associée au résultat final. Ce n'est pas nécessairement une bonne chose de donner un résultat comportant une faible marge d'erreur. Certains instruments ne permettent pas d'obtenir des mesures précises. Souvenez-vous que l'erreur indique la confiance que l'on peut avoir en un résultat, et qu'il vaut mieux surestimer l'erreur que la sous-estimer. Sous-estimer une erreur, cela revient à faire de la fausse représentation.

1 Erreurs statistiques et systématiques

Certaines erreurs surviennent chaque fois que l'on fait une mesure. Par exemple, une graduation inexacte sur un instrument a toujours le même effet sur toutes les mesures effectuées. Ces **erreurs sont dites systématiques** sont très difficiles à détecter et à traiter de manière générale. Il faut les considérer à nouveau pour chaque expérience. Les **erreurs statistiques** ont une valeur différente à chaque mesure. Comme leur nom le suggère, on peut les traiter par des méthodes statistiques. Par exemple, des lectures répétées d'une même valeur donnent un échantillon permettant d'obtenir un meilleur résultat et de faire une estimation de l'erreur statistique.

2 Règles relatives aux combinaisons d'erreurs statistiques

Un résultat final résulte en général d'un calcul portant sur des mesures différentes et indépendantes. Par exemple, on peut calculer la valeur d'une résistance en divisant la mesure provenant d'un voltmètre par la mesure provenant d'un ampèremètre. À chacune de ces deux mesures est associée

une erreur statistique à partir de laquelle il faut calculer l'erreur statistique sur la résistance. Les relations ci-dessous permettent d'obtenir la marge d'erreur associée à toute combinaison algébrique de quantités mesurées. Voici les règles de combinaison des erreurs statistiques associées à deux quantités indépendantes :

Somme et différence

$$(A \pm \Delta A) + (B \pm \Delta B) = (A + B) \pm \sqrt{(\Delta A)^2 + (\Delta B)^2}$$

$$(A \pm \Delta A) - (B \pm \Delta B) = (A - B) \pm \sqrt{(\Delta A)^2 + (\Delta B)^2}$$

Produit et quotient

$$(A \pm \Delta A)(B \pm \Delta B) = (AB) \left[1 \pm \sqrt{\left(\frac{\Delta A}{A}\right)^2 + \left(\frac{\Delta B}{B}\right)^2} \right]$$

$$\frac{A \pm \Delta A}{B \pm \Delta B} = \frac{A}{B} \left[1 \pm \sqrt{\left(\frac{\Delta A}{A}\right)^2 + \left(\frac{\Delta B}{B}\right)^2} \right]$$

Notez que pour les produits et les quotients, l'erreur sur le résultat se calcule à partir des erreurs relatives donc on peut utiliser des pourcentage, alors que pour les sommes et différences, il faut utiliser les erreurs absolues et non les pourcentage.

Les règles ci-dessus représentent des cas particuliers d'une formule d'erreur plus générale. Si le résultat R est fonction des variables mesurées A , B and C etc. ($R = f(A, B, C, \dots)$), alors

$$\Delta R = \sqrt{\left(\frac{\partial f}{\partial A}\right)^2 (\Delta A)^2 + \left(\frac{\partial f}{\partial B}\right)^2 (\Delta B)^2 + \dots}$$

Dans les formules comportant certaines fonctions comme \sin , \log , \exp , etc., il peut être plus commode d'utiliser directement cette formule pour calculer l'erreur sur le résultat.

Exemple:

Lecture de la tension = 19.5 ± 0.5 volts.

Lecture du courant = 4.3 ± 0.2 amp.

Si la résistance est vraiment égale au quotient de ces deux quantités (ce qui peut ne pas être le cas, s'il y a des effets de résistance interne), alors

$$\begin{aligned} \text{Résistance} &= \frac{19.5 \pm 0.5}{4.3 \pm 0.2} \Omega \\ &= \frac{19.5}{4.3} \left(1 \pm \sqrt{\left(\frac{0.5}{19.5}\right)^2 + \left(\frac{0.2}{4.3}\right)^2} \right) \Omega \\ &= 4.53 (1 \pm 0.053) \Omega \end{aligned}$$

Le résultat aurait pu être écrit sous la forme $4.5(1 \pm 0.05) \Omega$ ou bien $4.5 \pm 0.2 \Omega$.

On peut voir que l'erreur due à l'ampèremètre est nettement plus importante que l'erreur due au voltmètre. Re commençons le calcul précédent afin de mieux illustrer cette idée. L'erreur relative sur la tension $\Delta V/V = 0.5/19.5 = 0.026$ est d'environ 3% alors que l'erreur relative sur le courant $\Delta I/I = 0.2/4.3 = 0.047$ est d'environ 5%. Finalement, l'erreur relative calculée sur la résistance $\Delta R/R = \sqrt{(\Delta V/V)^2 + (\Delta I/I)^2} = 0.053$ ce qui est aussi d'environ 5%. On peut voir que dans ce cas l'erreur sur la tension a une contribution négligeable comparée à l'erreur sur l'intensité du courant. Cet exemple illustre que l'erreur calculée avec la formule précédente dépend surtout de la plus grande source d'erreur relative. Si une erreur est disons deux fois plus petite qu'une autre, alors on peut ne pas en tenir compte dans le calcul.

N'oubliez pas que souvent un seul chiffre significatif est nécessaire pour le résultat final et qu'en justifiant votre démarche, vous pouvez simplifier énormément vos calculs d'erreur.

N.B.: Dans le cas des barres d'erreur sur un graphique, la simplification pourrait ne pas s'appliquer sur l'intervalle entier des valeurs utilisées. Ce qui peut être aisément vérifié.

3 Utilisez votre intuition en appliquant les formules d'erreur

1. Ne donnez pas plus de chiffres significatifs dans le résultat qu'il n'y a dans l'erreur.
2. D'habitude, il ne faut qu'un chiffre significatif dans l'estimé de l'erreur
3. N'oubliez pas les erreurs systématiques: si vous trouvez qu'une erreur statistique obtenue en utilisant les formules est trop petite à votre goût, essayez de voir s'il n'y a pas d'erreurs systématiques et faites-en mention dans votre discussion.

4 Mesures répétées d'une même quantité

Il s'agit d'un cas très différent du cas précédent, où les mesures étaient indépendantes. Supposons que l'on fait la même mesure N fois. Chaque fois, on obtient une valeur légèrement différente, à cause de l'erreur statistique. Soit $A_1 A_2 A_3 \dots A_i \dots A_N$ les valeurs obtenues au cours de mesures successives, et ΔA l'erreur statistique sur chacune des mesures. La meilleure estimation de A s'obtient en faisant la moyenne de toutes les valeurs mesurées. Comme cette moyenne constitue une estimation meilleure que chacune des mesures prises séparément, on s'attend à ce que l'erreur soit inférieure à ΔA . Mais de combien ?

La valeur de l'erreur est donnée par :

$$\bar{A} \pm \overline{\Delta A} = \frac{A_1 + A_2 + A_3 + A_4 + \dots + A_N}{N} \pm \frac{\Delta A}{\sqrt{N}} = \left(\frac{1}{N} \sum_{i=1}^{i=N} A_i \right) \pm \frac{\Delta A}{\sqrt{N}}$$

Notez que la formule précédente s'obtient directement de la formule d'erreur pour la somme.

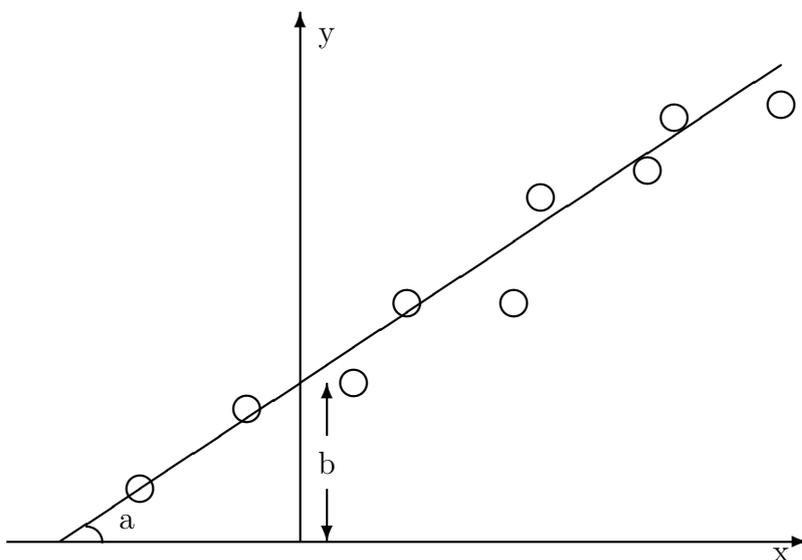
Ainsi, l'erreur statistique sur la valeur moyenne \bar{A} est inversement proportionnelle à la racine carrée du nombre de mesures effectuées. Il est donc utile de faire plusieurs mesures des valeurs critiques, mais il ne vaut pas la peine de le faire un trop grand nombre de fois. Par exemple, 4 mesures d'une même quantité permettent de doubler la précision du résultat final, mais il faut 16 mesures pour doubler à nouveau la précision.

5 Utilisation des calculatrices programmables et ordinateurs

Grâce aux calculatrices et ordinateurs, les calculs se font plus rapidement aujourd'hui qu'à l'époque des tables logarithmiques et de la règle à calcul. Les principes mathématiques n'ont pas changé, mais il est parfois utile de recourir à des méthodes plus élaborées, même si elles exigent des calculs beaucoup plus complexes, afin d'obtenir une meilleure estimation du résultat. Certains procédés mathématiques s'appliquent à plusieurs types d'expériences et des algorithmes standard ont été mis au point pour ces procédés. Certains logiciels et certaines calculatrices de poche possèdent de telles fonctions préprogrammées d'analyse de données, accessibles au moyen d'une simple touche. Il vous est conseillé d'utiliser l'ordinateur. **Mais il n'y a aucun avantage à utiliser ce dernier si vous prenez moins de temps à effectuer la tâche par vous-mêmes.** De plus, la machine ne doit pas vous faire perdre de vue la signification des nombres que vous manipulez. Elle ne sert qu'à effectuer les opérations arithmétiques. Les logiciels d'analyse de données ainsi que les calculatrices ont généralement un programme permettant de faire une régression linéaire par les moindres carrés. Il faut toutefois savoir que les statisticiens utilisent d'autres méthodes pour évaluer la marge de confiance dans leur résultats et n'estiment pas les erreurs de la même façon.

6 Approximation par une droite

La méthode qui consiste à trouver à l'oeil la droite passant le plus près possible des points expérimentaux demeure toujours valable. Si vous utilisez une calculatrice, faites toujours ce genre d'approximation pour confirmer la vraisemblance du résultat.



Soit une série de n mesures y_i , dont chacune correspond à un réglage x_i du montage. La calculatrice utilise comme données en entrée les couples (x_i, y_i) , où $i = 1, 2, \dots, n$, qui représentent les n points d'un graphe.

L'équation de la droite de régression,

$$y = ax + b$$

comporte la pente a et l'ordonnée à l'origine b , dont il faut faire l'estimation à partir des données expérimentales.

Soit les hypothèses suivantes :

1. On néglige les erreurs sur la quantité x .
2. Les erreurs sur la quantité y sont toutes les mêmes, $\sigma_1 = \sigma_2 = \sigma_3$, etc.
3. La distribution des valeurs y_i par rapport à la moyenne \bar{y}_i est de type gaussien.

Si toutes ces hypothèses sont vérifiées, la droite de régression est telle que la somme des carrés des différences d'ordonnée entre le point expérimental et la droite est minimale.

$$\sum_{i=1}^n (y_i - y)^2 = \text{minimum}$$

Pour que l'on ait un minimum, il faut que la dérivée du membre de gauche soit nulle.

En dérivant le membre de gauche par rapport à a , on obtient :

$$\begin{aligned} 0 &= \frac{\partial}{\partial a} \sum (y - y_i)^2 \\ &= \sum \frac{\partial}{\partial a} (ax_i + b - y_i)^2 \\ &= 2 \sum (ax_i + b - y_i) \frac{\partial}{\partial a} (ax_i + b - y_i) \\ &= 2 \sum (ax_i + b - y_i) x_i \end{aligned}$$

Donc,

$$a \sum x_i^2 + b \sum x_i = \sum x_i y_i \quad (1)$$

En dérivant de la même manière par rapport à b , on obtient :

$$\begin{aligned} 0 &= \frac{\partial}{\partial b} \sum (y - y_i)^2 \\ &= \sum \frac{\partial}{\partial b} (ax_i + b - y_i)^2 \\ &= 2 \sum (ax_i + b - y_i) \frac{\partial}{\partial b} (ax_i + b - y_i) \\ &= 2 \sum (ax_i + b - y_i) \end{aligned}$$

Donc,

$$a \sum x_i + bn = \sum y_i \quad (2)$$

La résolution de ces systèmes d'équations donne les valeurs recherchées de a et b . La méthode des déterminants de Kramer donne immédiatement

$$a = \frac{1}{\Delta} \begin{vmatrix} 1 & \bar{y} \\ \bar{x} & \overline{xy} \end{vmatrix} \text{ et } b = \frac{1}{\Delta} \begin{vmatrix} \bar{y} & \bar{x} \\ \overline{xy} & \overline{x^2} \end{vmatrix} \text{ où } \Delta = \begin{vmatrix} 1 & \bar{x} \\ \bar{x} & \overline{x^2} \end{vmatrix}.$$

Dans ce qui précède, $\bar{y} = \frac{1}{n} \sum y_i$, $\overline{x^2} = \frac{1}{n} \sum (x_i^2)$, et $\overline{xy} = \frac{1}{n} \sum (x_i y_i)$. Il est à noter que $\overline{x^2}$ est différent de \bar{x}^2 , et que \overline{xy} est différent de $(\bar{x})(\bar{y})$. On dit que les valeurs a et b sont obtenues par régression par la méthode

des moindres carrés. La plupart des calculatrices programmables actuelles permettent de les calculer. On peut aussi les obtenir avec une calculatrice non programmable, mais cela prend plus de temps. Le danger de tels programmes vient de ce qu'il est facile d'aboutir à des valeurs de a et b ayant 8 chiffres significatifs, sans pour autant savoir jusqu'à quel point elles sont valables.

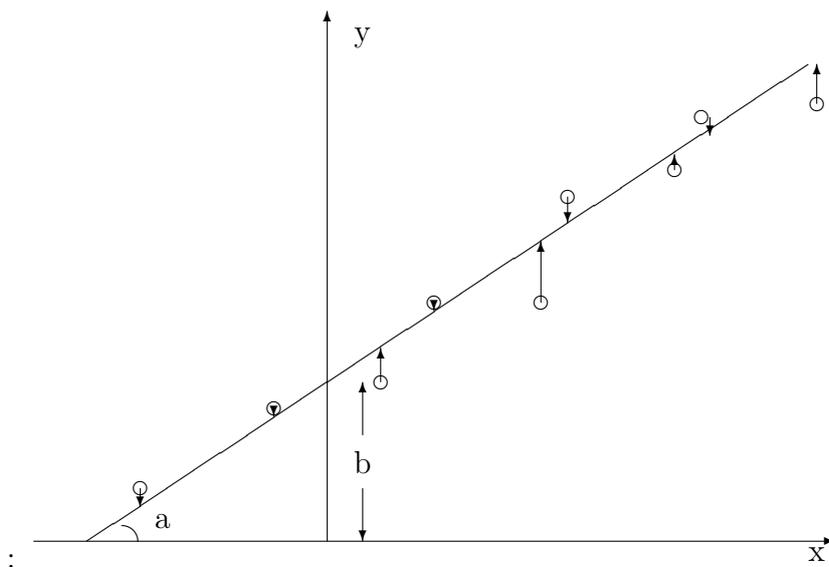
7 Estimation de l'erreur due à la régression par la méthode des moindres carrés

Il y a deux méthodes d'estimation de l'erreur due à la régression par la méthode des moindres carrés. Il faut choisir celle qui convient le mieux à l'expérience en cours.

7.1 Première méthode

Estimation de l'erreur à partir de la distance entre la droite et les points expérimentaux

Dans ce cas, aucune hypothèse n'est émise à propos de la précision des différentes mesures. Les marges d'erreur dues aux instruments n'entrent pas en ligne de compte.



La quantité minimisée par la méthode des moindres carrés est l'erreur σ , souvent appelée **écart-type**, donnée par la relation :

$$\sigma^2 = \frac{1}{n-2} \sum (y - y_i)^2 = \frac{1}{n-2} \sum (ax_i + b - y_i)^2$$

Le $n - 2$ du dénominateur vient du fait que deux degrés de liberté ont servi à trouver a et b , et qu'il ne reste plus donc que $n - 2$ degrés de liberté. Il est évident que s'il n'y a que deux points, la droite passe exactement par ces deux points, et que $\sum (y - y_i)^2 = 0$

L'effet de l'erreur en un point (x_i, y_i) sur les valeurs a et b obtenues par la méthode des moindres carrés peut s'écrire :

$$\sigma_{ai} = \sigma_i \frac{\partial a}{\partial y_i} \quad \text{et} \quad \sigma_{bi} = \sigma_i \frac{\partial b}{\partial y_i}$$

où σ_i est la différence $(y - y_i)$ entre l'ordonnée du point sur la droite de régression et l'ordonnée du point expérimental ayant la même abscisse. L'écart-type calculé sur tous les points précédents constitue une bonne approximation de l'erreur associée à un point donné. On pose donc $\sigma_i = \sigma$. Ainsi, l'incertitude sur la pente σ_a de la droite est donnée par :

$$\begin{aligned} \sigma_a^2 &= \sum_{i=1}^n \left(\sigma \frac{\partial a}{\partial y_i} \right)^2 = \sigma^2 \sum \left(\frac{\partial a}{\partial y_i} \right)^2 = \sigma^2 \sum \frac{1}{\Delta^2} \left(\frac{\partial}{\partial y_i} \left| \begin{array}{cc} 1 & \bar{y} \\ \bar{x} & \bar{xy} \end{array} \right| \right)^2 \\ &= \frac{\sigma^2}{n\Delta^2} \sum \left| \begin{array}{cc} 1 & 1 \\ \bar{x} & x_i \end{array} \right|^2 = \frac{\sigma^2}{n\Delta^2} \sum (x_i - \bar{x})^2 = \frac{\sigma^2}{n\Delta^2} \sum (x_i^2 - 2x_i\bar{x} + \bar{x}^2) \\ &= \frac{\sigma^2}{\Delta^2} (\bar{x}^2 - 2\bar{x}^2 + \bar{x}^2) = \frac{\sigma^2}{\Delta^2} (\bar{x}^2 - \bar{x}^2) \\ &\sigma_a^2 = \frac{\sigma^2}{\Delta} \end{aligned} \tag{3}$$

De la même manière, l'erreur σ_b sur l'ordonnée à l'origine est donnée par :

$$\begin{aligned} \sigma_b^2 &= \sum \left(\sigma \frac{\partial b}{\partial y_i} \right)^2 = \frac{\sigma^2}{\Delta^2} \sum \left(\frac{\partial}{\partial y_i} \left| \begin{array}{cc} \bar{y} & \bar{x} \\ \bar{xy} & \bar{x}^2 \end{array} \right| \right)^2 \\ &= \frac{\sigma^2}{n\Delta^2} \sum \left| \begin{array}{cc} 1 & \bar{x} \\ x_i & \bar{x}^2 \end{array} \right|^2 = \frac{\sigma^2}{n\Delta^2} \sum (\bar{x}^2 - \bar{x}x_i)^2 \end{aligned}$$

$$\begin{aligned}
&= \frac{\sigma^2}{n\Delta^2} \sum (\overline{x^2} - 2\overline{x}x_i + \overline{x^2}x_i^2) = \frac{\sigma^2}{\Delta^2} (\overline{x^2} - 2\overline{x}x^2 + \overline{x^2}x^2) \\
\sigma_b^2 &= \frac{\sigma^2}{\Delta^2} x^2 (\overline{x^2} - \overline{x}^2) = \frac{\sigma^2}{\Delta} x^2
\end{aligned} \tag{4}$$

Le calcul le plus facile de σ^2 est donné par la relation :

$$\sigma^2 = \frac{1}{n-2} \sum (y - y_i)^2 = \frac{n}{n-2} (\overline{y^2} - a\overline{xy} - b\overline{y})$$

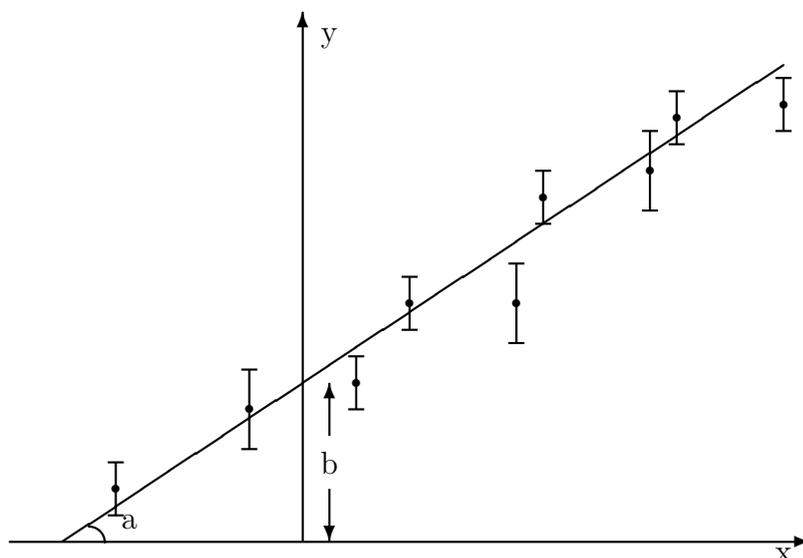
Les valeurs \overline{x} , \overline{y} , $\overline{x^2}$, \overline{xy} , et $\overline{x^2}$ sont déjà connues car elles ont servi au calcul de la pente a . Il reste toutefois à calculer $\overline{y^2}$ pour obtenir l'erreur associée à la pente a et à l'ordonnée à l'origine b .

Il est à noter que ces relations supposent que tous les points ont la même précision. Si certains points sont mesurés avec moins de précision que d'autres (comme dans le cas d'une désintégration radioactive, où les points obtenus à la fin sont moins précis que ceux du début à cause de la diminution du nombre d'événements enregistrés), cette méthode donne non seulement des marges d'erreur fausses, mais aussi des valeurs erronées pour a et b .

7.2 Seconde méthode

Estimation de l'erreur à partir de l'erreur associée à chaque mesure

Dans ce cas, on suppose connue l'erreur de mesure associée à la valeur y_i , de sorte que σ_i est défini pour chaque lecture y_i . Si les σ_i sont tous égaux, cela revient au cas précédent, avec les mêmes valeurs de a et b , à ceci près que σ_a et σ_b se calculent à partir de la valeur σ connue plutôt qu'à partir des approximations des équations (3) et (4).



Si certains points sont connus avec plus de précision que d'autres, il faut leur donner plus d'importance dans la détermination de la droite de régression. On utilise donc une méthode pondérée de régression par la méthode des moindres carrés. Le facteur de pondération associé à chaque point est $\omega_i = \frac{1}{\sigma_i^2}$, où σ_i est l'erreur supposée connue au départ.

Les formules donnant les valeurs de a et b sont semblables aux précédentes, à ceci près que l'on remplace les valeurs moyennes par des moyennes pondérées :

$$\bar{y} = \frac{\sum (y_i/\sigma_i^2)}{\sum (1/\sigma_i^2)} \quad \bar{x} = \frac{\sum (x_i/\sigma_i^2)}{\sum (1/\sigma_i^2)} \quad \overline{xy} = \frac{\sum (x_i y_i/\sigma_i^2)}{\sum (1/\sigma_i^2)} \quad \overline{x^2} = \frac{\sum (x_i^2/\sigma_i^2)}{\sum (1/\sigma_i^2)}$$

Les expressions donnant les erreurs sur a et b sont également les mêmes, sauf que le calcul de σ se fait à l'aide de la formule $\sigma^2 = 1 / \sum(\frac{1}{\sigma_i^2})$, et non à partir de l'écart-type par rapport à la droite.

Il est à noter que si les erreurs dues aux instruments sont sous-estimées, la seconde méthode risque de donner une estimation trop petite des erreurs sur la pente et l'ordonnée à l'origine, auquel cas la première méthode est préférable. Par contre, dans le cas extrême d'un graphe ne comportant que

deux points, la droite passe exactement par ces deux points et la première méthode donne une erreur nulle. Même avec un petit nombre de points, ils peuvent être alignés tout en donnant lieu chacun à une marge d'erreur importante, et la première méthode donne une très mauvaise estimation de l'erreur.

7.3 Cas où la droite passe par l'origine

L'équation d'une droite passant par l'origine est de la forme

$$y = ax$$

et la fonction à minimiser devient $\sum(y_i - y)^2 = \sum(y_i - ax_i)^2$.

Si l'on dérive par rapport à a , la condition d'obtention d'un minimum devient :

$$-2 \sum x_i(y_i - ax_i) = 0$$

Donc, la pente de la droite de régression est donnée par :

$$a = \frac{\sum x_i y_i}{\sum (x_i)^2} = \frac{(\overline{xy})}{(\overline{x^2})}$$

L'erreur σ_a sur la pente se calcule comme précédemment, et l'on obtient $\sigma_a^2 = \sigma^2 / (\overline{x^2})$, comme avec la première méthode. L'écart-type correspondant à l'ensemble des points est donné par :

$$\sigma^2 = \frac{n}{n-2} (\overline{y^2} - a^2 \overline{x^2})$$

Dans le cas de la deuxième méthode, les marges d'erreur sont déterminées par les caractéristiques des instruments utilisés.

Référence P.R.Bevington, *Data Reduction and Error Analysis for the Physical Sciences*, McGraw-Hill, 1969, 1992.