

Commentaires et compléments de la fiche de synthèse et des exemples à l'usage du professeur

La fiche de synthèse regroupe les 4 situations que l'élève sera susceptible de rencontrer au cours des activités expérimentales au lycée.

- S'il constate une variabilité du résultat d'un mesurage : Effectuer une estimation en utilisant une série de mesures.
- S'il ne constate pas de variabilité du résultat d'un mesurage¹ : Effectuer une estimation en utilisant une unique mesure.
- Si la grandeur recherchée n'est pas directement mesurée : Utiliser une relation de propagation.
- Exploiter quantitativement le résultat d'une modélisation graphique.

Exprimer le résultat d'un mesurage dans le cas où l'on possède un échantillon de N mesures (x_1, \dots, x_n) du mesurande X (Estimations de type A)

1. Comment effectuer rapidement un étude statistique avec les élèves ?

Pour effectuer les calculs avec les élèves, on peut utiliser un tableur tel que Excel, Libre Office ou bien le logiciel libre GUM_MC de Jean-Marie BIAN SAN (Voir Fiche Methode).

2. Comment utiliser le coefficient de student ?

L'utilisation de l'estimation de l'écart-type $s = \sqrt{\frac{(x_i - \bar{x})^2}{n-1}}$ entraîne une modification significative sur l'évaluation des probabilités d'une variable normale lorsque l'on travaille avec des petits échantillons². La loi t de Student permet d'évaluer les modifications à la loi normale provenant de cette approximation. Il existe des tables de coefficients de student (on peut en fournir une version simplifiée aux élèves)

N	3	4	5	6	7	8	9	10	12	14	16	20
P=95%	4,3	3,82	2,78	2,57	2,45	2,37	2,31	2,26	2,20	2,16	2,13	2,09
P=99%	9,93	5,84	4,60	4,03	3,71	3,50	3,36	3,25	3,11	3,01	2,95	2,86

Figure 1 – Coefficients de student en fonction du nombre N de mesures

ou bien utiliser un logiciel pour les fournir (Exemple : Fonction `LOI.INVERSE.STUDENT` de Libre Office.

3. Comment choisir le nombre de chiffres significatifs et l'arrondis pour exprimer un incertitude ?

Le *Guide pour l'expression de l'incertitude de mesure* publié par le Bureau International des Poids et Mesures émet les recommandations suivantes :

- Les incertitudes seront données avec au plus 2 chiffres significatifs

1. car l'expérience est trop longue pour être répétée ou par manque de précision de l'appareil de mesure
2. On considère un petit échantillon lorsque $n \leq 30$

- « En énonçant les résultats finals, il peut parfois être approprié d'arrondir les incertitudes au chiffre supérieur plutôt qu'au chiffre le plus proche. Par exemple, $u_y = 10,47 \text{ m}\Omega$ pourrait être arrondi à $11 \text{ m}\Omega$. Cependant, le bon sens doit prévaloir et une valeur comme $u_x = 28,05 \text{ kHz}$ doit être arrondie à la valeur inférieure, 28 kHz . Les estimations d'entrée et de sortie doivent être arrondies en accord avec leurs incertitudes ; par exemple, si $y = 10,05762 \Omega$ avec $u_y = 27 \text{ m}\Omega$, y doit être arrondi à $10,058 \Omega$. »

Le professeur peut cependant être amené à imposer une règle plus systématique à ses élèves.

Dans les exemples de ces fiches les incertitudes ont toujours été exprimées avec 2 chiffres significatifs et tronquées, par précaution, à l'excès .

Exprimer le résultat d'un mesurage dans le cas où l'on ne possède qu'une seule mesure du mesurande X (Estimations de type B)

1. Pourquoi utiliser un distribution rectangulaire ?

Il existe d'autres distributions que la distribution rectangulaire (distribution triangulaire, par exemple). La distribution rectangulaire permet cependant de traiter simplement l'ensemble des cas présentés au lycée. Cette distribution majore légèrement l'incertitude.

2. Existe-il un logiciel adapté à ce genre d'estimation ?

Pour effectuer les calculs, on peut utiliser un tableur tel que Excel ou LibreOffice ou bien le logiciel libre GUM_MC de Jean-Marie BIAN SAN (Voir Fiche Methode).

Exprimer le résultat d'un mesurage dans le cas où la grandeur recherchée n'est pas directement mesurée : Calcul d'une incertitude composée

1. Comment propager des incertitudes avec les élèves sans que cela devienne fastidieux ?

Il est possible d'utiliser le logiciel GUM_MC pour calculer le résultat d'une opération de propagation d'incertitudes. (Voire fiche Méthode)

2. Comment propager des incertitudes dans un cas plus complexe ?

Dans le cas où les différentes grandeurs liées par une relation $X = f(G_1 \dots G_N)$ ont été mesurées de manière indépendante (grandeurs décorrélées), la relation de générale de propagation des incertitudes est

$$u_X = \sqrt{\sum_{i=1}^N \left(\frac{\partial f}{\partial G_i} \right)_{\hat{G}_i}^2 u_{G_i}^2}$$

Si les différentes grandeurs n'ont pas été mesurées de manière indépendante (grandeurs corrélées), la relation précédente n'est plus valable.

3. Qu'est que le nombre de degrés de liberté effectifs ? Pourquoi choisir par défaut un coefficient d'élargissement de 2 lors d'une estimation de type B ?

Dans le cas où la relation entre les grandeurs est du type, $Y = f(X_1, \dots, X_n)$, pour déterminer le nombre de degrés de liberté effectif global ν_{eff} , on utilise la relation approchée de WELCH-SATTERTHWAITE

$$\nu_{eff} = \frac{u_Y^4}{\sum_{i=1}^n \frac{(c_i u_{X_i})^4}{\nu_i}}$$

avec ν_i le degré de liberté effectif de la distribution de la grandeur X_i .

Si l'évaluation de l'incertitude-type de X_i est de type A, ν_i est le degré de liberté de la distribution de Student correspondante (donc nombre de mesures-1 si c'est la moyenne d'une série de mesures répétées indépendantes).

Si elle est de type B et qu'on ne possède pas de renseignements spécifiques sur le nombre de degrés de libertés, on peut approximer ν_i par

$$\nu_i \approx \frac{1}{2} \left[\frac{\Delta u_{X_i}}{u_{X_i}} \right]^{-2}$$

où $\frac{\Delta u_{X_i}}{u_{X_i}}$ est l'incertitude relative sur la valeur de l'incertitude-type de X_i ³.

Dans notre exemple,

– $u_{C_b} = 0,0143 \text{ mol/L}$

$$- \sum_{i=1}^n \frac{(c_i u_{X_i})^4}{\nu_i} = \frac{\left(\frac{C_a}{V_b} \cdot u_{V_a}\right)^4}{N-1} = \frac{\left(\frac{0,1}{10,0} \times 0,4\right)^4}{7} = 3,6571428 \times 10^{-11} \text{ mol}^4 \cdot \text{L}^{-4}$$

– On a alors :

$$\nu_{eff} = \frac{u_Y^4}{\sum_{i=1}^n \frac{(c_i u_{X_i})^4}{\nu_i}} = \frac{0,0143^4}{3,6571428 \times 10^{-11}} = 1174,86$$

On a alors un nombre de degré de liberté de $\nu_{eff} = 1174$. On peut alors considérer que l'on est dans le cas d'une distribution de student à 1174 degré de liberté (on peut même considéré que l'on est dans le cadre de l'approximation gaussienne et donc avoir un coefficient d'élargissement $k = 2$ pour un niveau de confiance de 95%.

Exprimer le résultat d'un mesurage dans le cas où l'on possède N couples de valeurs (x_i, y_i) (Méthode des moindres carrés)

1. L'élève doit-il connaître toutes les "formule" présente dans la fiche de synthèse ? En particulier celles relatives à la méthode des moindres carrés ?

3. La plupart du temps, on considèrera que cette erreur relative est totalement négligeable et égale à zéro. Dans ce cas, on aura $\nu_i = +\infty$

4. Les degrés effectifs de liberté ν_{C_a} et ν_{V_b} pour C_a et V_b sont estimés à $+\infty$ donc le poids de ces deux grandeurs pour l'estimation du nombre de degré de liberté effectif global ν_{eff} est nulle

NON seules les compétences précisées par le BO sont exigibles. Une fiche de synthèse peut être proposée à l'élève lors des activités liées à la mesure et à son traitement. Les relations proposées dans la partie modélisation graphique sont données à titre d'illustration.

2. Quelle est l'origine des coefficients présentés dans la fiche et dans l'exemple ?

Dans le cas d'une modélisation linéaire, la méthode des moindres carrés consiste à chercher l'équation de la droite $\hat{y} = \hat{a}x$ par mis tous les ax possibles, qui minimise la grandeur $C = \sum_{i=1}^N (y_i - \hat{y}_i)^2$. Cela revient alors à chercher la valeur de \hat{a} telle que :

$$\left(\frac{dC}{da}\right)_{\hat{a}} = 0$$

En explicitant la condition précédente, on obtient alors l'équation suivant :

$$\left(\frac{dC}{da}\right)_{\hat{a}} = -2 \sum_{i=1}^N (y_i - \hat{a}x_i)x_i = 0$$

d'où

$$\hat{a} = \frac{\sum_{i=1}^N x_i y_i}{\sum_{i=1}^N x_i^2}$$

Chaque expérience réalisée, permet de placer sur le graphe un point de coordonnées (x_i, y_i) . L'ensemble de ces points est supposé normalement distribué par rapport à la droite de régression linéaire. Pour chaque point, la distance à la droite $r_i = y_i - \hat{y}_i$ (appelée « le résidu ») traduit l'erreur que commise lors de chaque mesure.

On quantifie alors la dispersion des points expérimentaux par rapport au modèle selon la relation :

$$s_{stat} = \sqrt{\frac{1}{d} \sum_{i=1}^N (y_i - \hat{y}_i)^2}$$

Remarques :

- Le coefficient d correspond au nombre de degrés de liberté de l'échantillon. Dans le cas d'une modélisation linéaire, de la forme $Y = aX$, $d = N - 1$
- Dans le cas où l'on suppose que les x_i sont parfaitement déterminés, c'est à dire $u_{x_i} = 0$ alors on identifie $s_{stat} = \hat{u}_y$

L'incertitude type associée à la pente a découlent de \hat{u}_y par simple propagation d'incertitudes.

3. Comment faire dans le cas affine ?

Dans le cas d'une modélisation affine (de la forme $Y = aX + b$), la méthode des moindres carrés consiste à chercher l'équation de la droite $\hat{y} = \hat{a}x + \hat{b}$ par mis tous les $ax + b$ possibles, qui minimise la grandeur $C = \sum_{i=1}^N (y_i - \hat{y}_i)^2$. Cela revient alors à chercher les valeurs de a et b telles que :

$$\left(\frac{\partial C}{\partial a}\right)_{\hat{a}} = 0 \quad \text{et} \quad \left(\frac{\partial C}{\partial b}\right)_{\hat{b}} = 0$$

En explicitant la condition précédente, on obtient alors une expression des coefficients de régression linéaire :

$$\hat{a} = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^N (x_i - \bar{x})^2} \quad \text{et} \quad \hat{b} = \bar{y} - \hat{a}\bar{x}$$

On quantifie alors la dispersion des points expérimentaux par rapport au modèle selon la relation :

$$s_{stat} = \sqrt{\frac{1}{N-2} \sum_{i=1}^N (y_i - \hat{a}x_i - \hat{b})^2}$$

Il en découle par propagation des incertitudes les estimateurs des incertitudes associés à la pente et) l'ordonnée à l'origine :

$$\hat{u}_a = \frac{s_{stat}}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2}} \quad \text{et} \quad \hat{u}_b = s_{stat} \sqrt{\frac{1}{N} + \frac{\bar{x}^2}{\sum_{i=1}^N (x_i - \bar{x})^2}}$$

4. Quels outils sont à notre disposition pour travailler efficacement avec des modélisations graphiques ?

Etant donnée la complexité des paramètres mis en jeu, il est intéressant de procéder à un traitement informatisé avec un logiciel tel qu'Excel, Libre Office ou bien un logiciel dédié aux lycéens comme Regressi (Voir Fiche Méthode). La calculatrice de l'élève peut aussi être utilisée avec profit en lien avec le cours de Mathématiques.

Pour un traitement manuel, on peut aussi demander à l'élève de tracer 2 droites extrêmes. Le pente (et éventuellement l'ordonnée à l'origine) sera encadrée et traitée avec une estimation de type B.

5. Comment valider la pertinence du modèle ?

La validation du modèle est une étape importante. On remarque la sensibilité et la simplicité de l'étude visuelle des résidus. A l'inverse, l'étude du coefficient de corrélation r (ou de détermination r^2) ne constitue pas, à elle seule, une méthode pertinente de validation du modèle.

6. Pourquoi ne met on pas de "barres d'incertitudes" comme on le trouve dans beaucoup de documents scientifiques ?

On trouve parfois des graphiques affichant pour chaque point une barre ou une ellipse représentant l'incertitude-type associée à chaque point du graphe. L'exploitation de ces données permet d'affiner l'estimation des paramètres de modélisation à condition que chaque incertitude-type soit issue d'une étude de type A. Cette méthode (appelée méthode des ellipses ou méthode du chi2) semble peu adaptée aux activités expérimentales réalisées au lycée où l'on dispose, le plus souvent, d'un petit nombre de mesures.