Détermination des limites de détection & de quantification des méthodes de dosage en laboratoire hospitalier

PROJET STATISTIQUE 2A

Document de travail

2023 - 2024

Etudiants :
David HERRENSCHMIDT
Camille NAVEL
Raymond WARNOD

Tuteur :
Emmanuel CURIS
Coaching :
Emilie DESOUCHE

1 cas d'une droite d'étalonnage

Soit x la concentration et y le résultat de la mesure physique. Dans le cas d'une courbe d'étalonnage linéaire la relation est donnée par :

$$y = \beta_0 + \beta_1 x + \epsilon$$

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

Une fois la machine étalonnée on ne cherche pas à évaluer la concentration en fonction de la mesure physique, mais l'inverse :

$$\hat{x} = \frac{y - \hat{\beta_0}}{\hat{\beta_1}}$$

Cette équation pose quelques problèmes. Elle défini un estimateur de x qui est biaisé, dont la loi n'est pas une loi bien connue et dont la variance est par conséquent difficile à estimer. Pour pallier à ces difficultés nous allons faire les l'hypothèses que l'estimation de β_0 et β_1 sont très proche de leurs valeurs réelles. Cette hypothèse pourra être contrôler à posteriori. Ainsi, on peut approcher $\hat{x}$ par un développement limité d'ordre 1 :

$$\hat{x} \approx \frac{y - \beta_0}{\beta_1} - \frac{1}{\beta_1}(\hat{\beta_0} - \beta_0) - \frac{y - \beta_0}{\beta_1^2}(\hat{\beta_1} - \beta_1)$$

En remplaçant par y par $\beta_0 + \beta_1 x + \epsilon$ on obtient :

$$\hat{x} \approx x - \frac{\epsilon}{\beta_1} - \frac{1}{\beta_1}(\hat{\beta_0} - \beta_0) - \frac{x}{\beta_1}(\hat{\beta_1} - \beta_1) - \frac{\epsilon}{\beta_1}(\hat{\beta_1} - \beta_1)$$

Enfin en négligeant le terme d'erreur d'ordre 2 $\frac{\epsilon}{\beta_1}(\hat{\beta_1} - \beta_1)$ on obtient :

$$\hat{x} \approx x - \frac{1}{\beta_1}(\hat{\beta_0} - \beta_0) - \frac{x}{\beta_1}(\hat{\beta_1} - \beta_1) - \frac{\epsilon}{\beta_1}$$

Ainsi, $\hat{x}$ est un estimateur approximativement sans biais, approximativement gaussien dont la variance est approximativement :

$$var(\hat{x}) \approx \frac{1}{\beta_1^2}(var(\hat{\beta_0}) + x^2 * var(\hat{\beta_1}) + 2x * cov(\hat{\beta_0}, \hat{\beta_1}) + var(\epsilon))$$

En notant $D = n\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2$ et σ_y^2 la variance de y on obtient :

$$var(\hat{x}) \approx \frac{\sigma_y^2}{\beta_1^2}\left(\frac{1}{D}\left(\sum_{i=1}^n x_i^2 + x^2 n - 2x\sum_{i=1}^n x_i\right) + 1\right)$$

$$var(\hat{x}) \approx \frac{\sigma_y^2}{\beta_1^2}\left(1 + \frac{1}{nD}\left(n\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2 + \left(\sum_{i=1}^n x_i\right)^2 + x^2 n^2 - 2xn\sum_{i=1}^n x_i\right)\right)$$

$$var(\hat{x}) \approx \frac{\sigma_y^2}{\beta_1^2}\left(1 + \frac{1}{n} + \frac{(nx - \sum_{i=1}^n x_i)^2}{nD}\right)$$

Dans le cas d'une mesure répété k fois, si l'on utilise la moyenne des k résultats obtenus comme estimation de x seule par de variance due à l'incertitude sur y est modifié. En notant η_i i variant de 1 à k les termes d'erreur, il suffit de remplacer dans les équation précédentes $var(\epsilon)$ par $\frac{1}{k}var(\sum_{i=1}^k \eta_i) = \frac{\sigma_y^2}{k}$ et l'on obtient :

$$var(\hat{x}) \approx \frac{\sigma_y^2}{\beta_1^2}\left(\frac{1}{k} + \frac{1}{n} + \frac{(nx - \sum_{i=1}^n x_i)^2}{nD}\right)$$

1.1 limite de blanc

Deux cas peuvent être distingués pour calculer la limite de blanc. On peut soit utiliser les données qui on permis de construire la courbe d'étalonnage, soit utiliser des mesures de blancs dédiées à l'estimation de la limite de blanc.

1.1.1 Utilisation de mesures dédiées

Sous l'hypothèse que les mesures de blanc sont des réalisation d'une variable aléatoire gaussienne de moyenne y_0 et d'écart type σ_y, soient $\hat{y_0} = \frac{1}{n}\sum_{i=1}^{n} y_i$ et $\hat{\sigma_y} = \frac{\sum_{i=1}^{n}(yi-\hat{y_0})}{n-1}$ leurs estimateurs respectifs.

Considérons une nouvelle mesure de blanc y_{n+1} :

$$y_{n+1} = y_0 + \epsilon_{n+1}$$

$$var(y_{n+1} - \hat{y_0}) = var(y_0 - \hat{y_0} + \epsilon_{n+1}) = var(y_0 - \hat{y_0}) + var(\epsilon_{n+1}) = \sigma_y^2(\frac{1}{n} + 1)$$

et $\frac{\hat{y_0}-y_0}{\hat{\sigma_y}}$ suit une loi de Student à n-1 degrés de liberté. On peut donc estimer que α% des mesures de blancs seront inférieur à $t_{n-1}(\alpha)\hat{\sigma_y\sqrt{\frac{1}{n}+1}}$. Ainsi la limite de blanc au niveau de confiance $1-\alpha$ sera donnée par :

$$LOB = \hat{y_0} + \frac{t_{n-1}(1-\alpha)\hat{\sigma_y}\sqrt{\frac{1}{n}+1} - \hat{\beta_0}}{\hat{\beta_1}}$$

1.1.2 Utilisation des données d'étalonnage

Lorsqu'on de dispose pas de mesures dédiées à l'estimation de la limite de blanc, elle peut être estimée à partir des données utilisées pour construire la courbe d'étalonnage. Nous avons donné un estimateur d'une concentration x approximativement sans biais et gaussien de variance :

$$\sigma_x^2 \approx \frac{\sigma_y^2}{\beta_1^2}\left(1 + \frac{1}{n} + \frac{(nx - \sum_{i=1}^{n} x_i)^2}{nD}\right)$$

On peut estimer cette variance en remplaçant dans cette équation σ_y ET β_1 par leurs estimations. En négligeant la variabilité de $\hat{\beta_1}$, $\frac{x-\hat{x}}{\hat{\sigma_x}}$ suit une loi de Student à n-2 de degrés de liberté. (on fait cette hypothèse pour retrouver le résultat de l'article mais à vérifier qu'on ne fait pas n'importe quoi). Ainsi la limite de blanc pour un niveau de confiance $1-\alpha$ peut être estimée par :

$$LOB = t_{n-2}(1-\alpha)\frac{\hat{\sigma_y}}{|\hat{\beta_1}|}\sqrt{1 + \frac{1}{n} + \frac{(\sum_{i=1}^{n} x_i)^2}{nD}}$$

1.2 limite de détection

En utilisant à nouveau le fait que $\frac{x-\hat{x}}{\hat{\sigma_x}}$ suit une loi de Student à n-2 de degrés de liberté la limite de détection au niveau de confiance $1-\alpha$ la limite de détection sera solution de l'équation en x :

$$LOB = x - t_{n-2}(\alpha)\frac{\hat{\sigma_y}}{|\hat{\beta_1}|}\sqrt{1 + \frac{1}{n} + \frac{(x\sum_{i=1}^{n} x_i)^2}{nD}}$$

et donc solution de l'équation polynomiale de degré 2 en x :

$$(LOB - x)^2 - t_{n-2}^2(\alpha)\frac{\hat{\sigma_y}^2}{\hat{\beta_1}^2}\left(1 + \frac{1}{n} + \frac{(nx - \sum_{i=1}^{n} x_i)^2}{nD}\right) = 0$$

Il s'agit de intersection de 2 paraboles dont l'une à pour minimum 0 en LOB et l'autre $t^2_{n-2}(\alpha)\frac{\hat{\sigma_y}^2}{\hat{\beta_1}^2}\left(1+\frac{1}{n}\right)$ en $\frac{\sum_{i=1}^n x_i}{n} = \bar{x}$. Si $t^2_{n-2}(\alpha)\frac{\hat{\sigma_y}^2 n}{\hat{\beta_1}^2 D} < 1$, ce qui devrait être le cas si l'hypothèse d'un faible coefficient de variation de $\hat{\beta_1}$ est respectée, nous aurons deux solutions, une de chaque côté de LOB.

Dans l'article une formule est proposée en supposant σ_x^2 constant. Voir si cette hypothèse est réaliste pour les données qui nous sont fournies.

1.3 Limites de quantification

1.3.1 Pour une précision absolue

Soit p la précision souhaitée et $1-\alpha$ le niveau de confiance auquel au souhaite que cette précision soit atteinte. Cette précision sera atteinte si la demi longueur de l'intervalle de confiance associé à x est inférieure ou égale à p :

$$t_{n-2}(1-\frac{\alpha}{2})\frac{\hat{\sigma_y}}{|\hat{\beta_1}|}\sqrt{1+\frac{1}{n}+\frac{(x\sum_{i=1}^n x_i)^2}{nD}} \leq p$$

Les limites de quantification, si elles existent seront les solutions de l'équation polynomiale du second degré en x :

$$t^2_{n-2}(1-\frac{\alpha}{2})\frac{\hat{\sigma_y}^2}{\hat{\beta_1}^2}\left(1+\frac{1}{n}+\frac{(nx-\sum_{i=1}^n x_i)^2}{nD}\right) = p^2$$

1.3.2 Pour une précision relative

Pour estimer la précision relative associé à une mesure, l'usage est d'utiliser le coefficient de variation. Ainsi un précision relative r sera obtenue lorsque :

$$cv(x) = \frac{\sigma_x}{x} \leq r$$

Les limites de quantification seront les solutions de l'équation polynomiale du second degré :

$$\frac{\hat{\sigma_y}^2}{\hat{\beta_1}^2}\left(1+\frac{1}{n}+\frac{(nx-\sum_{i=1}^n x_i)^2}{nD}\right) = r^2x^2$$

Si l'on souhaite associé un niveau de confiance $1-\alpha$ aux limites de quantification relative on cherchera les solutions de l'équation :

$$t^2_{n-2}(1-\frac{\alpha}{2})\frac{\hat{\sigma_y}^2}{\hat{\beta_1}^2}\left(1+\frac{1}{n}+\frac{(nx-\sum_{i=1}^n x_i)^2}{nD}\right) = r^2x^2$$

2 Droite d'étalonnage avec hetéroscédasticité

Nous allons nous intéresser au un cas particulier de courbe d'étalonnage linéaire non homoscédastique, le cas ou la variance du terme d'erreur est proportionnel à la concentration.

Pour mieux visualiser la manière dont cette hypothèse se propage dans nos formules, nous allons commencer par considérer le cas plus général où la variance du terme d'erreur est proportionnelle à une fonction connue de la concentration x notée h :

$$y = \beta_0 + \beta_1 x + \sqrt{h(x)}\epsilon$$

$$y_i = \beta_0 + \beta_1 x_i + \sqrt{h(x_i)}\epsilon_i$$

où il existe σ tel que pour tout i, $var(\epsilon_i] = \sigma^2$

Pour estimé les coefficients β_0, β_1 et σ, nous allons utiliser la méthode des moindres carrés généralisés qui revient à appliquer la méhtode des moindres carrés ordinaires à :

$$\frac{y_i}{\sqrt{h(x_i)}} = \beta_0 \frac{1}{\sqrt{h(x_i)}} + \beta_1 \frac{x_i}{\sqrt{h(x_i)}} + \epsilon_i$$

En notant $D = \sum_{i=1}^n \frac{1}{h(x_i)} \sum_{i=1}^n \frac{x_i^2}{h(x_i)} - \left(\sum_{i=1}^n \frac{x_i}{h(x_i)}\right)^2$ On a :

$$var(\hat{\beta_0}) = \frac{1}{D}\sum_{i=1}^n \frac{x_i^2}{h(x_i)}\sigma^2$$

$$var(\hat{\beta_1}) = \frac{1}{D}\sum_{i=1}^n \frac{1}{h(x_i)}\sigma^2$$

$$cov(\hat{\beta_0}, \hat{\beta_1}) = -\frac{1}{D}\sum_{i=1}^n \frac{x_i}{h(x_i)}\sigma^2$$

En reprenant le développement limité d'ordre 1 vu dans le cas homoscédastique, on obtient :

$$\hat{x} \approx x - \frac{1}{\beta_1}(\hat{\beta_0} - \beta_0) - \frac{x}{\beta_1}(\hat{\beta_1} - \beta_1) - \sqrt{h(x)}\frac{\epsilon}{\beta_1}$$

et :

$$var(\hat{x}) \approx \frac{1}{\beta_1^2}(var(\hat{\beta_0}) + x^2 * var(\hat{\beta_1}) + 2x * cov(\hat{\beta_0}, \hat{\beta_1}) + h(x)var(\epsilon))$$

$$var(\hat{x}) \approx \frac{\sigma^2}{\beta_1^2}\left(\frac{1}{D}\left(\sum_{i=1}^n \frac{x_i^2}{h(x_i)} + x^2\sum_{i=1}^n \frac{1}{h(x_i} - 2x\sum_{i=1}^n \frac{x_i}{h(x_i)}\right) + h(x)\right)$$

$$var(\hat{x}) \approx \frac{\sigma^2}{\beta_1^2}\left(h(x) + \frac{\sum_{i=1}^n \frac{1}{h(x_i)}\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n \frac{x_i}{h(x_i)}\right)^2 + \left(\sum_{i=1}^n \frac{x_i}{h(x_i)}\right)^2 + x^2\left(\sum_{i=1}^n \frac{1}{h(x_i)}\right)^2 - 2x\sum_{i=1}^n \frac{1}{h(x_i)}\sum_{i=1}^n x_i}{D\sum_{i=1}^n \frac{1}{h(x_i)}}\right)$$

$$var(\hat{x}) \approx \frac{\sigma^2}{\beta_1^2}\left(h(x) + \frac{1}{\sum_{i=1}^n \frac{1}{h(x_i)}} + \frac{(x\sum_{i=1}^n \frac{1}{h(x_i)} - \sum_{i=1}^n \frac{x_i}{h(x_i)})^2}{D\sum_{i=1}^n \frac{1}{h(x_i)}}\right)$$

Lorsqu'on h(x) = x la formule se simplifie en :

$$var(\hat{x}) \approx \frac{\sigma^2}{\beta_1^2}\left(x + \frac{1}{\sum_{i=1}^n \frac{1}{x_i}} + \frac{(x\sum_{i=1}^n \frac{1}{x_i} - n)^2}{D\sum_{i=1}^n \frac{1}{x_i}}\right)$$

A noter que ce qui précède n'est applicable que pour h(x) > 0. En particulier pour le cas où h(x) = x, on ne pourra pas utiliser de mesures de blanc pour la construction de la droite de régression.

2.1 limite de blanc

On peut noter que si l'hypothèse de variance proportionnelle à la concentration était vérifier cela signifierait qu'on a aucune incertitude sur résultat du mesure de blanc ce qui est peut réaliste. Toutefois, rien n'empêche de considérer que cette hypothèse est vraie uniquement pour des concentrations supérieures à la limite de détection.

On peut à nouveau distinguer le cas ou l'on dispose de mesures de blancs dédiées du cas où on n'en dispose pas.

2.1.1 Utilisation de mesures dédiées

Ce cas se traite exactement comme dans le cas homoscédastique. La limite de blanc au niveau de confiance $1-\alpha$ sera donnée par :

$$LOB = \hat{y_0} + \frac{t_{n-1}(1-\alpha)\hat{\sigma_y}\sqrt{\frac{1}{n}+1} - \hat{\beta_0}}{\hat{\beta_1}}$$

où $\hat{y_0} = \frac{1}{n}\sum_{i=1}^{n} y_i$, $\hat{\sigma_y} = \frac{\sum_{i=1}^{n}(yi-\hat{y_0}}{n-1}$.

Les estimations de β_o et et β_1 étant ici réalisées avec la méthode des moindres carrés généralisés.

2.1.2 Utilisation des données d'étalonnage

Dans ce cas on peut réutiliser la démarche utilisée dans le cas homoscédastique en mettant à jour la valeur de la variance de l'estimateur de x :

$$\sigma_x^2 \approx \frac{\sigma^2}{\beta_1^2}\left(x + \frac{1}{\sum_{i=1}^{n}\frac{1}{x_i}} + \frac{(x\sum_{i=1}^{n}\frac{1}{x_i} - n)^2}{D\sum_{i=1}^{n}\frac{1}{x_i}}\right)$$

On obtient ainsi :

$$LOB = t_{n-2}(1-\alpha)\frac{\hat{\sigma_y}}{|\hat{\beta_1}|}\sqrt{\frac{1}{\sum_{i=1}^{n}\frac{1}{x_i}} + \frac{n^2}{D\sum_{i=1}^{n}\frac{1}{x_i}}}$$

Ici l'incertitude sur le blanc provient uniquement de l'incertitude due à la construction de la courbe d'étalonnage. L'incertitude due au mesure de blanc est considérée comme nul ce qui n'est pas réaliste.

2.2 limite de détection

En utilisant à nouveau le fait que $\frac{x-\hat{x}}{\hat{\sigma_x}}$ suit une loi de Student à n-2 de degrés de liberté la limite de détection au niveau de confiance α la limite de détection sera solution de l'équation en x :

$$LOB = x - t_{n-2}(1-\alpha)\frac{\hat{\sigma_y}}{|\hat{\beta_1}|}\sqrt{x + \frac{1}{\sum_{i=1}^{n}\frac{1}{x_i}} + \frac{(x\sum_{i=1}^{n}\frac{1}{x_i} - n)^2}{D\sum_{i=1}^{n}\frac{1}{x_i}}}$$

et donc solution de l'équation polynomiale de degré 2 en x :

$$(LOB-x)^2 - t_{n-2}^2(1-\alpha)\frac{\hat{\sigma_y}^2}{\hat{\beta_1}^2}\left(x + \frac{1}{\sum_{i=1}^{n}\frac{1}{x_i}} + \frac{(x\sum_{i=1}^{n}\frac{1}{x_i} - n)^2}{D\sum_{i=1}^{n}\frac{1}{x_i}}\right) = 0$$

On conservera la plus petite solution supérieure à LOB. A voir si l'on peut montrer qu'on à une solution supérieure à LOB et une inférieure.

2.3 Limites de quantification

Là encore on peut utiliser les mêmes raisonnements que dans le cas homoscédastique en adaptant la valeur de la variance de l'estimateur de x

2.3.1 Pour une précision absolue

Soit p la précision souhaitée et $1-\alpha$ le niveau de confiance auquel au souhaite que cette précision soit atteinte. Cette précision sera atteinte si la demi longueur de l'intervalle de confiance associé à x est inférieure ou égale à p :

$$t_{n-2}(1-\frac{\alpha}{2})\frac{\hat{\sigma_y}}{|\hat{\beta_1}|}\sqrt{x+\frac{1}{\sum_{i=1}^{n}\frac{1}{x_i}}+\frac{(x\sum_{i=1}^{n}\frac{1}{x_i}-n)^2}{D\sum_{i=1}^{n}\frac{1}{x_i}}} \leq p$$

Les limites de quantification seront les solutions de l'équation polynomiale du second degré en x :

$$t_{n-2}^2(1-\frac{\alpha}{2})\frac{\hat{\sigma_y}^2}{\hat{\beta_1}^2}\left(x+\frac{1}{\sum_{i=1}^{n}\frac{1}{x_i}}+\frac{(x\sum_{i=1}^{n}\frac{1}{x_i}-n)^2}{D\sum_{i=1}^{n}\frac{1}{x_i}}\right)=p^2$$

2.3.2 Pour une précision relative

Pour estimer la précision relative associé à une mesure, l'usage est d'utiliser le coefficient de variation. Ainsi un précision relative r sera obtenue lorsque :

$$cv(x)=\frac{\sigma_x}{x}\leq r$$

Les limites de quantification seront les solutions de l'équation polynomiale du second degré :

$$\frac{\hat{\sigma_y}^2}{\hat{\beta_1}^2}\left(x+\frac{1}{\sum_{i=1}^{n}\frac{1}{x_i}}+\frac{(x\sum_{i=1}^{n}\frac{1}{x_i}-n)^2}{D\sum_{i=1}^{n}\frac{1}{x_i}}\right)=r^2x^2$$

Si l'on souhaite associé un niveau de confiance $1-\alpha$ aux limites de quantification relative on chercher les solution de l'équation :

$$t_{n-2}^2(1-\frac{\alpha}{2})\frac{\hat{\sigma_y}^2}{\hat{\beta_1}^2}\left(x+\frac{1}{\sum_{i=1}^{n}\frac{1}{x_i}}+\frac{(x\sum_{i=1}^{n}\frac{1}{x_i}-n)^2}{D\sum_{i=1}^{n}\frac{1}{x_i}}\right)=r^2x^2$$