

3. EXEMPLE 3

3.1. Objectifs

- Réaliser un test de la linéarité d'une droite de régression (test de lack-of-fit)
- Réaliser un ajustement avec la composante quadratique et en tester la signification

Ces tests sont à réaliser lors de la validation initiale car ils nécessitent des répétitions (plusieurs valeurs de Y pour un X donné).

3.2. Généralités

En général la courbe de calibration "du jour" (c'est-à-dire celle qui est construite quotidiennement) sera une simple droite. Lors de la validation initiale il importera de valider ou d'invalider cette hypothèse par un test de linéarité. Si au cours de la validation initiale, on démontre que dans la gamme des concentrations sélectionnées on a bien une droite, alors cette hypothèse sera systématiquement acceptée pour les droites du jour (en d'autres termes on ne vérifiera pas l'hypothèse de linéarité avec les droites du jour).

Il se peut que dans la gamme des concentrations considérées, l'hypothèse de linéarité soit rejetée mais que la prise en compte d'une **composante quadratique (c'est-à-dire d'un terme significatif en X^2) améliore la calibration**. Cela veut dire que la calibration se fera avec une courbe et non une droite, ce qui peut être un avantage dans la mesure où on pourra avoir une large gamme de concentrations dans la même courbe. L'alternative serait de "rétrécir" la courbe du jour pour en faire une droite mais cela obligerait à faire des dilutions des échantillons ayant les concentrations les plus élevées.

Nous allons examiner dans cet exercice ces différentes questions.

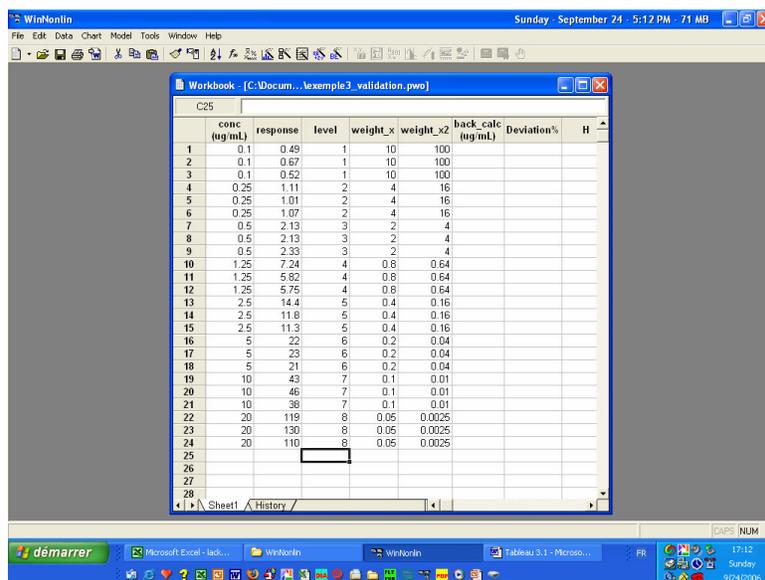
Le tableau 3.1 donne un exemple de résultats analytiques pour construire la courbe de calibration avec 8 niveaux de concentration ayant chacun 3 répétitions.

La figure 3.1. montre le tableau des données à monter dans WinNonlin.

Tableau 3.1 : Gamme des concentrations et réponse chromatographique obtenue – unités arbitraires.

Concentrations théoriques (nominales)	Réponses chromatographiques
0.10	0.49
0.10	0.67
0.10	0.52
0.25	1.11
0.25	1.01
0.25	1.07
0.5	2.13
0.5	2.13
0.5	2.33
1.25	7.24
1.25	5.82
1.25	5.75
2.5	14.4
2.5	11.8
2.5	11.3
5	22
5	23
5	21
10	43
10	46
10	38
20	119
20	130
20	110

Figure 3.1. Les données du tableau 3.1 doivent être saisies dans WinNolin pour faire l'exercice. La colonne "level" contient un code de 1 à 8 correspondant aux 8 niveaux de concentrations; "weight_x" est le poids 1/X et "weight_x2" est le poids 1/X²



3.3. Ajustement des données à une droite non pondérée et pondérée

Commençons par ajuster les données du tableau 3.1 à une droite sans pondération. On obtient l'équation :

$$Y = 5.7682 X - 2.7289 \text{ (avec } r = 0.9857\text{)}$$

La figure 3.2 donne une représentation graphique de la droite ajustée. Son inspection suggère que :

- (i) la réponse chromatographique n'évolue pas de façon linéaire avec les concentrations mais plutôt selon une courbe concave vers le haut
- (ii) la dispersion des réponses chromatographiques augmente avec les concentrations, c'est-à-dire que les variances augmentent avec les concentrations.

Cette double impression est confirmée par l'inspection du graphique des résidus (fig. 3.3). Il montre que les résidus ne sont pas répartis au hasard de part et d'autre de l'horizontal (résidu d'ordonnée 0) mais qu'ils suivent une tendance décroissante de la concentration 0.1 à la concentration 10, pour remonter avec la concentration 20. De plus, la dispersion des résidus augmente avec les concentrations.

Dans WinNonlin il n'est pas possible de faire simplement un test d'homogénéité des variances sur les résidus (test de Bartlett). Néanmoins, l'inspection des variances de résidus montre immédiatement que les variances ne sont pas homogènes pour les différents niveaux de concentration (c'est-à-dire qu'elles ne sont pas du même ordre de grandeur). On peut vérifier cela en faisant appel à l'outil "descriptive statistics" avec "Level" comme "**sort variable**" et "**Response**" comme "**summary variable**" (Sort en anglais signifie trier) (fig. 3.4).

Figure 3.2 : Graphique de la droite de régression obtenue en ajustant les données du tableau 3.1 à une droite sans pondération. On notera que les concentrations mesurées pour les concentrations 5 et 10 sont nettement au-dessous de la droite et que la dispersion des concentrations mesurées augmente avec les concentrations. De plus les réponses prédites pour 0.1 et 0.25 sont négatives !

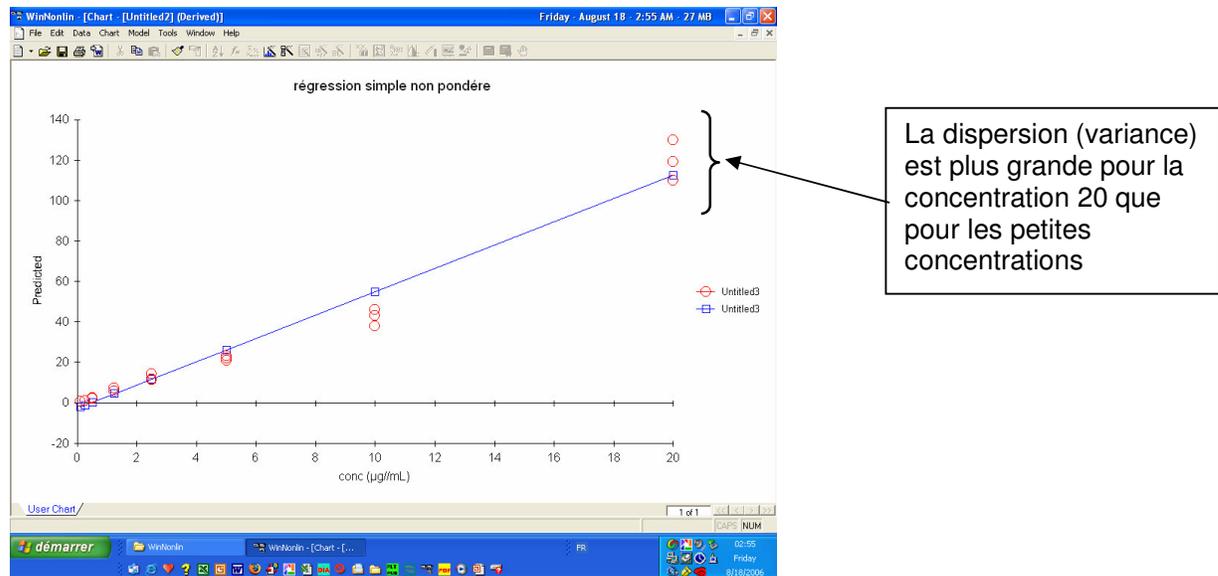


Figure 3.3 : Répartition des résidus obtenus en ajustant les données du tableau 3.1 avec une droite de régression non pondérée. On notera la forme de banane dessinée par les résidus et la dispersion qui augmente de façon croissante avec celle des concentrations.

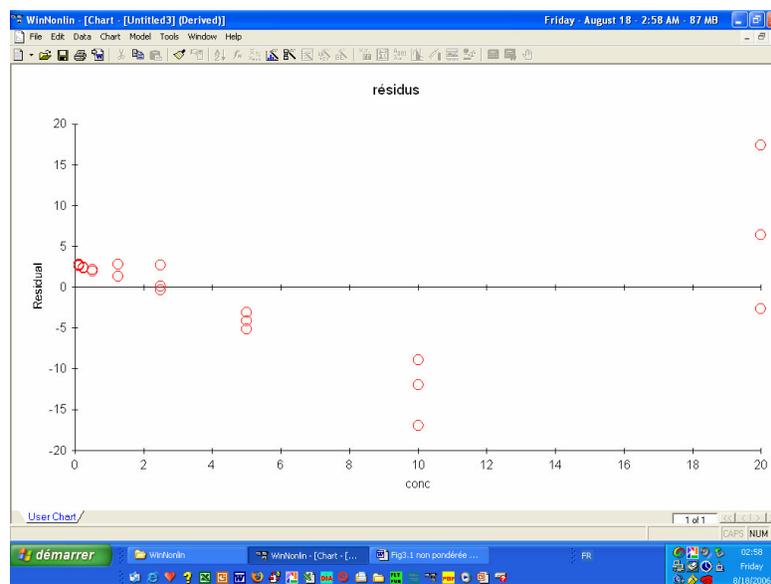
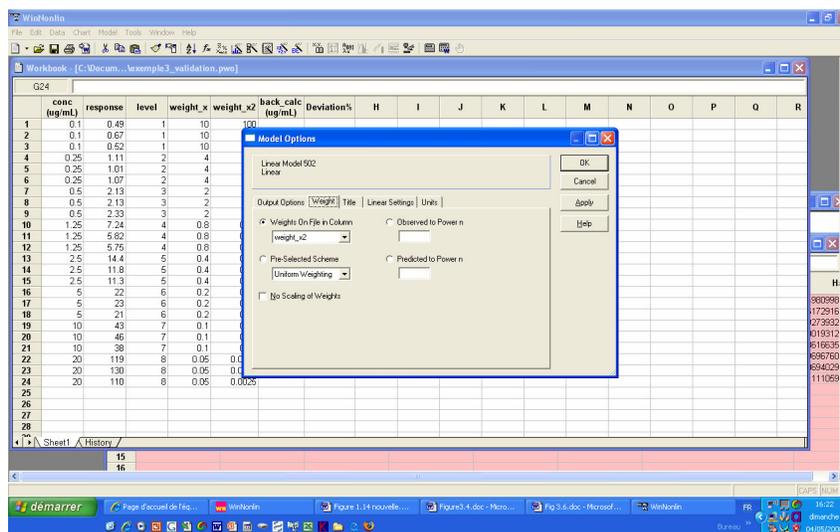


Figure 3.4 : Fenêtre pour effectuer des statistiques descriptives sur les résidus et inspecter les variances et coefficients de variation.



Le tableau des résultats statistiques peut être édité pour faciliter l'inspection des données. Pour cela, avec le click droit de la souris faire "Detach" ce qui permet d'éditer directement la feuille des résultats (fig. 3.5). Ensuite on élimine avec "Delete" les colonnes inutiles (fig 3.6). **L'inspection de la figure 3.6 montre que les variances augmentent avec les concentrations** (de 0.0093 à 100.33) ce qui indique une non-homogénéité (non-égalité) des variances. En revanche, les coefficients de variation sont similaires (de 4.5 à 17%) ce qui nous indique qu'une pondération va s'imposer (en $1/X^2$).

Figure 3.5 : pour pouvoir éditer une feuille de résultats dans WinNonlin il convient de la déverrouiller avec la fonction "Detach" par un click droit de la souris.

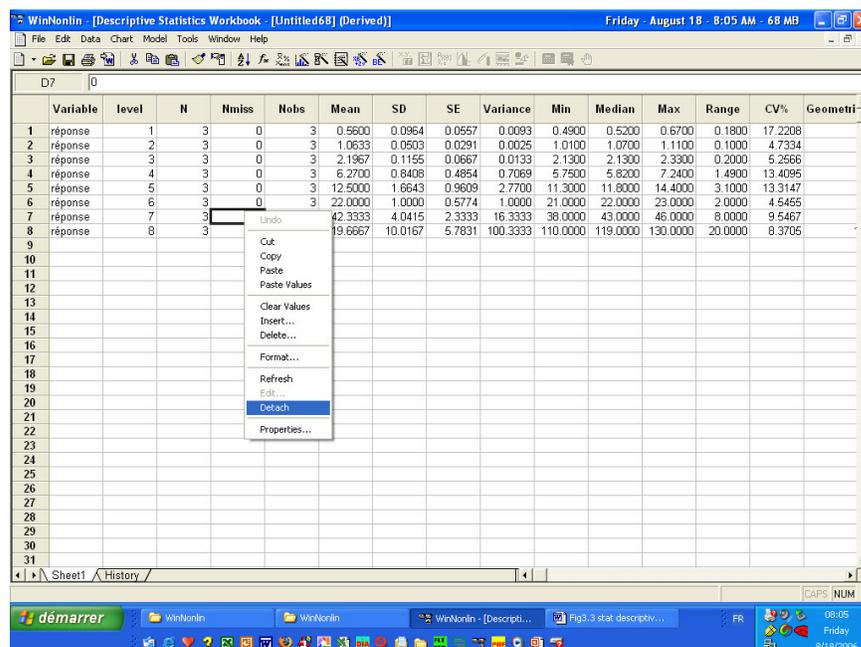
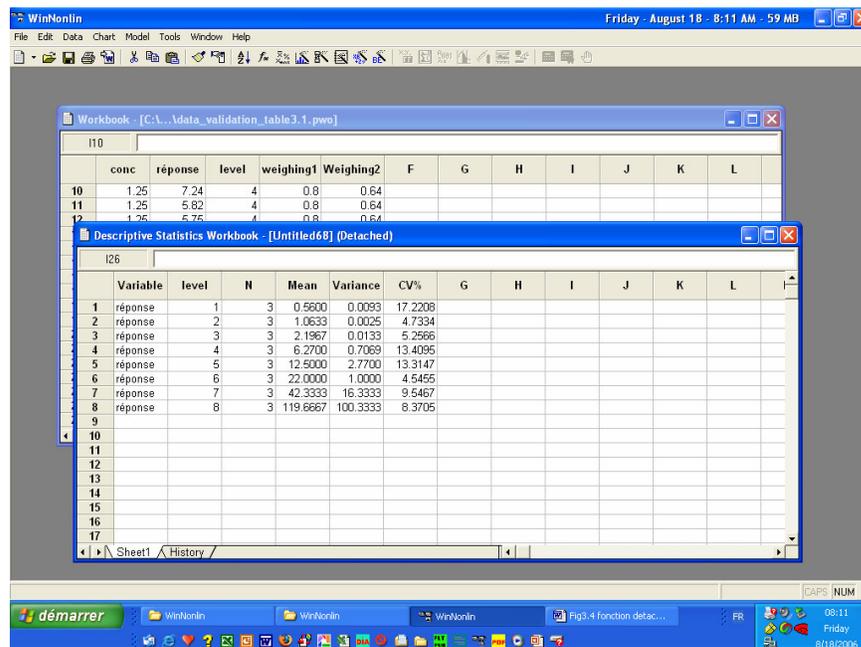


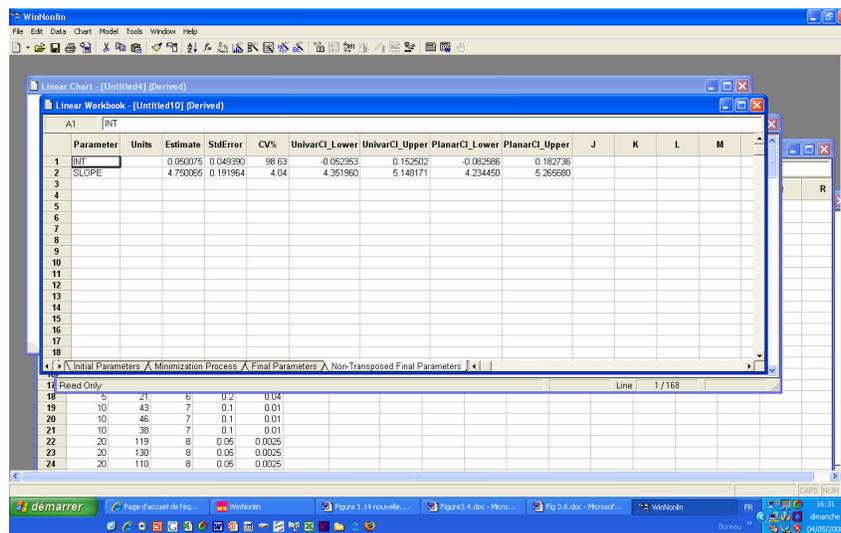
Figure 3.6 : Variances et coefficients de variation des réponses chromatographiques du tableau 3.1 pour les différents niveaux de concentration. On note une grande hétérogénéité des variances ; en revanche les coefficients de variation (CV%) sont homogènes ce qui plaide en faveur d'une pondération par $1/X^2$.



Cela plaide pour un facteur de pondération de $1/X^2$. Nous allons donc modéliser ces données avec un facteur de pondération $1/X^2$.

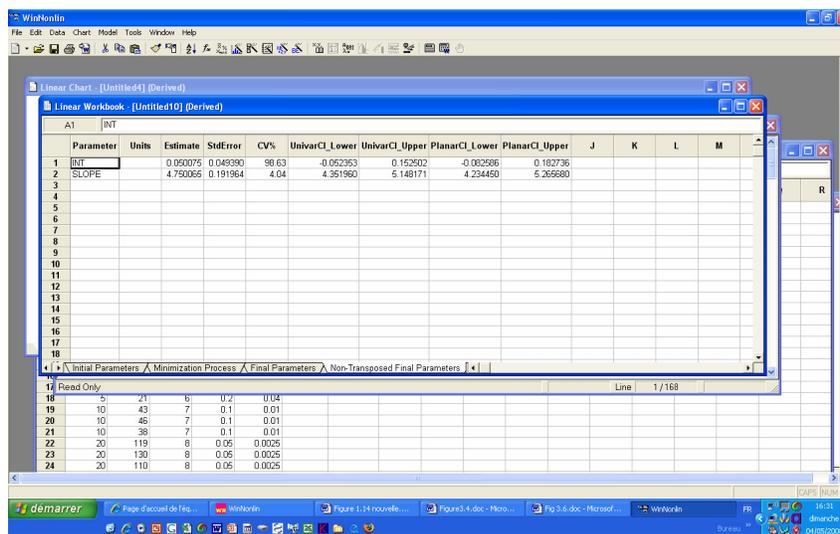
On va réaliser cette régression pondérée en plaçant le vecteur "*Weight_X2*" dans la boîte "*Weights on file columns*" après avoir ouvert les onglets "*Model option > Weight*" (fig.3.7).

Figure 3.7 : Les paramètres de la droite de régression



Les paramètres estimés sont donnés dans l'onglet "*Non transposed final Parameters*" (fig. 3.8) avec "*a*" = **4.7501 (pente)** et "*b*" = **0.0501 (intercept)**.

Figure 3.8 : Paramètres de la droite de régression estimés pour le modèle de régression pondérée par $1/X^2$.



La courbe ajustée et les valeurs observées sont données sur la figure 3.9. Les résidus pondérés sont présentés sur la figure 3.10.

Figure 3.9 : Courbe de calibration obtenue avec un modèle linéaire et une pondération de $1/X^2$. L'inspection de la figure montre que la droite, même pondérée donne un mauvais ajustement.

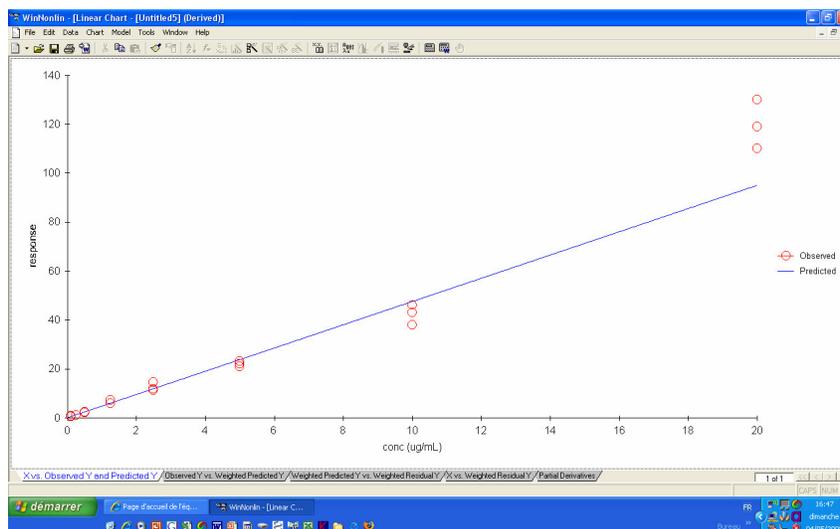
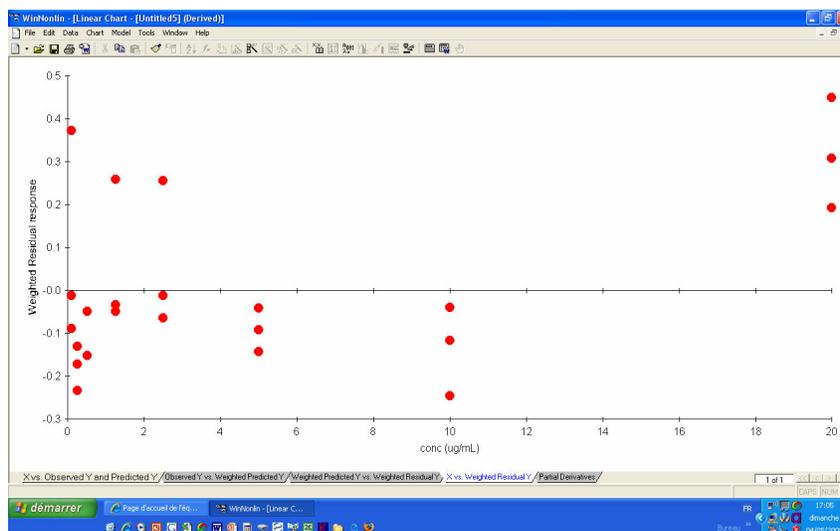


Figure 3.10 : Répartition des résidus obtenue en ajustant les données du tableau 3.1 avec un modèle linéaire simple et une pondération de $1/X^2$. L'inspection des résidus suggère que la dispersion des résidus, pour chaque niveau de concentration, est similaire ce qui plaide en faveur de l'adéquation de la pondération en $1/X^2$. En revanche, la répartition des résidus forme une banane ce qui suggère que le modèle linéaire simple n'est pas adéquat



L'inspection des résidus suggère que le schéma de pondération est adéquat (même dispersion des résidus par niveau de concentration) mais qu'ils forment une sorte de banane (on dit qu'il y a de la structure) ce qui suggère que le modèle de la droite n'est pas satisfaisant. Nous allons vérifier cela avec un test de "lack of fit" qui va tester l'hypothèse que le modèle est bien (ou non) une droite de type $Y = aX + b$.

3.4. Test de linéarité d'une droite

Avant d'accepter la simple droite comme modèle pour la droite de calibration du jour (droite du jour), on doit au moins vérifier lors de la validation initiale qu'il s'agit bien d'une droite de type $Y = aX + b$.

On dispose de 3 approches différentes pour accepter ou refuser la droite comme modèle par défaut.

- 1) l'inspection visuelle de la figure et de celle des résidus. Si manifestement, on observe une répartition en banane des résidus, on peut éliminer la droite comme modèle par défaut sans faire de test.
- 2) Faire un test statistique de linéarité (test of lack of fit) c'est ce que nous allons faire dans cette section

- 3) Faire les calculs de calibration inverse et voir si les valeurs prédites de X par calcul inverse sont acceptables. En effet, il se peut que le test de linéarité rejette l'hypothèse de linéarité mais que les calculs inverses soient acceptables. Ce cas de figure survient avec des techniques analytiques très reproductibles (automates) ; Compte tenu de leur excellente précision, la moindre déviation à la linéarité est détectée par les tests statistiques sans que cela ait une incidence pratique.

On notera que le calcul du coefficient de corrélation (r) n'est pas une bonne approche pour tester la linéarité et un $r = 0.999$ peut correspondre à une courbe plutôt qu'à une droite.

Dans cette section nous allons expliquer comment procéder avec WinNonlin pour tester la linéarité en sachant qu'il n'y a pas de test par défaut offert dans WinNonlin. Il faudra donc le faire à la main !

3.4.1. Test de lack of fit (calcul manuel pour une droite non pondérée)

Pour comprendre le principe nous allons faire intégralement le test de linéarité à la main c'est-à-dire avec la simple aide d'une feuille Excel.

Commençons par en expliquer le principe.

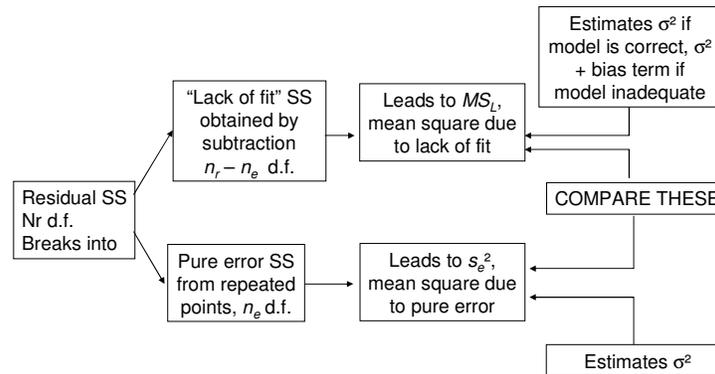
La figure 3.11 montre que la somme des carrés de la résiduelle (SS_{totale}) obtenue à l'issue d'une régression par un modèle simple (ou plus complexe) peut être répartie en ses **2 composantes** : **une partie liée à l'erreur pure (pure error SS) et une partie liée au défaut d'ajustement (SS lack-of-fit)**. On peut donc écrire :

$$SS_{\text{totale}} = SS_{\text{lack of fit}} + SS_{\text{pure}}$$

Et le $SS_{\text{lack of fit}}$ sera obtenu par différence

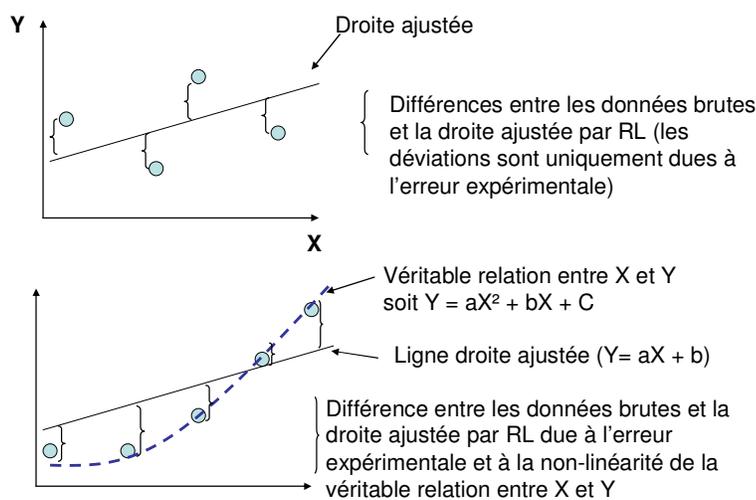
$$SS_{\text{lack of fit}} = SS_{\text{total}} - SS_{\text{pure}}$$

Figure 3.11 Lack of fit et erreur pure (d'après Draper & Smith) (d.f. : degré de liberté)



La figure 3.12 montre les 2 sources d'un défaut d'ajustement : une source d'erreur expérimentale (qui peut être réduite en augmentant la précision de la technique analytique) et un défaut d'ajustement (qui peut être réduit en sélectionnant un meilleur modèle de régression).

Figure 3.12 : représentation graphique des 2 sources d'un défaut d'ajustement des données brutes à un modèle de droite



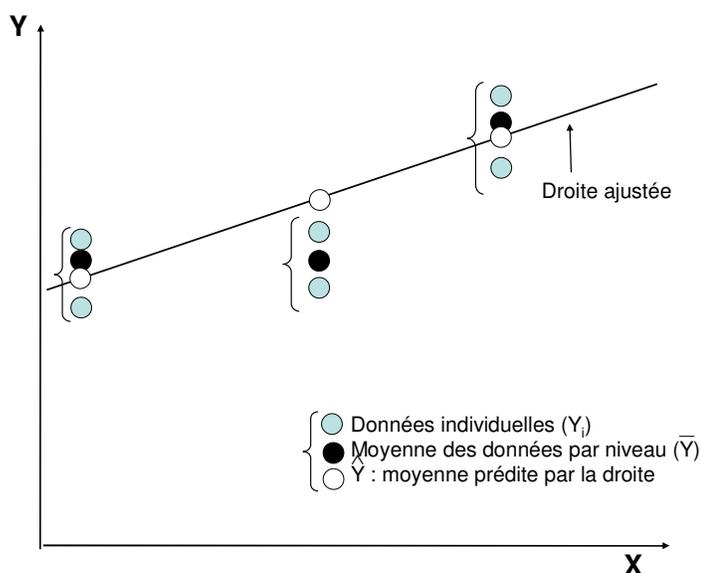
La variance de l'erreur pure notée S_e^2 (encore nommé carré moyen et qui est la "pure SS error" divisée par son nombre de degré de liberté) est une variance qui est indépendante du modèle. Pour être estimée elle nécessite des répliques (plusieurs valeurs observées de Y pour chaque X). Cette variance (S_e^2) est un estimateur de la variance (σ^2) des données brutes et elle calculée en prenant en compte, pour chaque niveau de concentration, la moyenne (\bar{Y}) des données brutes (Y_i).

Cette même variance peut également être estimée en prenant en compte non pas les \bar{Y} (moyennes observées à chaque niveau de concentration), mais les moyennes estimées pour les différents niveaux de concentration par la droite de régression ajustée et qui seront notées \hat{Y} (ou Y chapeau). Si la relation entre X et Y est bien une ligne droite, les déviations des valeurs observées de Y_i à la droite ajustée ne doivent être dues qu'à la variabilité intrinsèque des Y_i (ou encore les \bar{Y} et les \hat{Y} doivent être très similaires). Si la relation entre X et Y n'est pas une droite, la variance, telle que mesurée à partir des déviations (résidus) des Y_i à la droite ajustée, seront augmentées à cause de la non-linéarité.

Le principe des tests de non-linéarité va donc consister à comparer deux variances : la variance de l'erreur pure (the pooled error for the Y_i replicates, or the within mean square) avec la variance des déviations des Y_i à la droite ajustée (deviation from regression).

La figure 3.13 montre la façon de calculer ces deux variances.

Figure 3.13 Estimation des variances pour réaliser un test de non-linéarité (ici 2 points par niveau). Une première variance (celle de l'erreur pure) peut être calculée en ne tenant pas compte du modèle de régression. La variance de l'erreur pure est calculée en prenant en compte la moyenne des données observées à chaque niveau de concentration (\bar{Y} , ●). Une seconde approche pour calculer une autre variance consiste à considérer la moyenne prédite par la droite de calibration (notée \hat{Y} , ○ et non \bar{Y}). Si ces deux variances sont similaires (c'est-à-dire si les \hat{Y} et les \bar{Y} sont similaires), alors cela veut dire que le modèle de régression est adéquat et n'entraîne pas de distorsion (lack of fit).



Nous allons maintenant procéder à ces calculs pour notre exemple du tableau 3.1. en commençant par l'erreur pure (S_e^2).

La SS recherchée est obtenue en calculant pour chaque niveau les $(Y_i - \bar{Y})^2$ avec Y_i les données brutes et \bar{Y} , la moyenne de chaque niveau de concentration. De façon plus formelle :

$$SS_{total} = \sum_{j=1}^{n=8} \sum_{v=1}^{n=3} (Y_{jv} - \bar{Y}_j)^2$$

avec j le nombre de niveau de concentration (ici 8 niveaux de concentration) et v le nombre de réponses par niveau de concentration (ici le même nombre de réponses de 3 par niveau).

Le calcul de la variance de l'erreur pure nécessite de connaître le nombre de degré de liberté (ddl) appelé en anglais degree of freedom (df) avec :

$$ddl(ne) = \sum_{j=1}^8 (nj - 1) = \sum_{j=1}^m nj - m$$

soit pour notre exemple $ne = 3 \times 8 - 8 = 16$

La S_e^2 pure est donc

$$S_e^2 = \frac{\sum_{j=1}^{n=8} \sum_{v=2}^{n=3} (Y_{jv} - \bar{Y}_i)^2}{\sum_{j=1}^m nj - m}$$

Cette S_e^2 est une estimation de σ^2 la variance interne aux données brutes. Pour notre exemple :

$$S_e^2 = 242.33/16 = 15.146$$

Le tableau 3.2 donne le détail des calculs effectués avec Excel.

Tableau 3.2 Calculs à réaliser avec Excel pour estimer l'erreur pure des données du tableau 3.1 (within S²)

Concentrations théoriques (nominales)	Réponses chromatographiques	\bar{Y}	$(Y_i - \bar{Y})^2$	$SS_{(Y_i - \bar{Y})^2}$	Var
0.10	0.49		0.0049		
0.10	0.67	0.560	0.0121	0.0186	0.0093
0.10	0.52		0.0016		
0.25	1.11		0.002177		
0.25	1.01	1.0633	0.002844	0.005066	0.002533
0.25	1.07		0.0000444		
0.5	2.13		0.00444		
0.5	2.13	2.1966	0.00444	0.0266	0.01333
0.5	2.33		0.0177		
1.25	7.24		0.9409		
1.25	5.82	6.270	0.2025	1.4138	0.7069
1.25	5.75		0.2704		
2.5	14.4		3.61		
2.5	11.8	12.50	0.49	5.54	2.77
2.5	11.3		1.44		
5	22		0.00		
5	23	22.00	1.00	2.00	1.00
5	21		1.00		
10	43		0.444		
10	46	42.333	13.444	32.666	16.33
10	38		18.777		
20	119		0.444		
20	130	119.666	106.777	200.666	100.33
20	110		93.444		
SS grand total :				242.33	
S_e^2 :				15.146	

\bar{Y} : moyenne par niveau de concentration

$SS_{(Y_i - \bar{Y})^2}$: résidu élevé au carré. Ex.: $(0.49 - 0.56)^2 = 0.0049$

Moyenne des variances des données par niveau

La droite de régression sans pondération est donnée avec l'équation

$$Y = 5.7682 X - 2.7289 \text{ (voir section 3.3)}$$

Avec cette droite nous pouvons estimer, pour chaque niveau de concentration, la valeur de la réponse prédite (\hat{Y}); par exemple pour la concentration 0.5 : $\hat{Y} = 0.155$.

Avec ces valeurs prédites, nous allons recalculer l'erreur en remplaçant les \bar{Y} du tableau 3.2 par les \hat{Y} ; les calculs intermédiaires sont donnés dans le tableau 3.3.

Tableau 3.3 Calcul de l'erreur résiduelle lorsque les données sont ajustées avec une droite non pondérée.

Concentrations théoriques (nominales)	Réponse chromatographique	\hat{Y}	$(Y_i - \hat{Y})^2$	$SS_{(Y_i - \hat{Y})^2}$
0.10	0.49	-2.152	6.980	22.085
0.10	0.67		7.964	
0.10	0.52		7.1402	
0.25	1.11		5.74507	16.57
0.25	1.01	-1.2868	5.27569	
0.25	1.07		5.55492	
0.5	2.13		3.89997	12.5298
0.5	2.13	0.1552	3.89997	
0.5	2.33		4.729903	
1.25	7.24		7.610255	11.0118
1.25	5.82	4.48133	1.792035	
1.25	5.75		1.609521	
2.5	14.4		7.335398	7.500503
2.5	11.8	11.6916	0.011749	
2.5	11.3		0.153355	
5	22		16.90983	52.72948
5	23	26.1121	9.685515	
5	21		26.134139	
10	43		142.88033	510.45401
10	46	54.9531	80.160793	
10	38		287.41289	
20	119		112.63	348.98
20	130	112.6351	40.504	
20	110		301.527	
SS grand total				981.8672

\hat{Y} : valeur prédite par la droite de régression $Y = 5.76822 X - 2.7889441$



Total qui sera donné par WN avec un ajustement sans pondération

La somme totale des carrés des résidus (SS_{total}) estimée est de 981.867 contre 242.33 pour les calculs faits avec \bar{Y} .

La différence est due à la SS du lack-of-fit qui est de

$$SS_{\text{lack-of-fit}} = 981.867 - 242.33 = 739.54$$

Cette $SS_{\text{lack-of-fit}}$ est à 6 ddl et la variance correspondante du lack-of-fit est de $739.54/6=123.26$.

C'est cette variance qui doit être comparée à la variance résiduelle soit :

$$F_{16}^6 = \frac{123.26}{15.146} = 8.137$$

A comparer au F_{16}^6 critique qui est de 2.74 pour $P=0.05$ et 4.20 pour $P=0.01$.

Il apparaît donc que ce test de linéarité conduit à rejeter la droite comme modèle car 8.137 est supérieur à 2.74 et à 4.20. La droite sera rejetée pour $P<0.01$.

Le tableau 3.4 donne le tableau de l'ANOVA pour faire le test de non-linéarité.

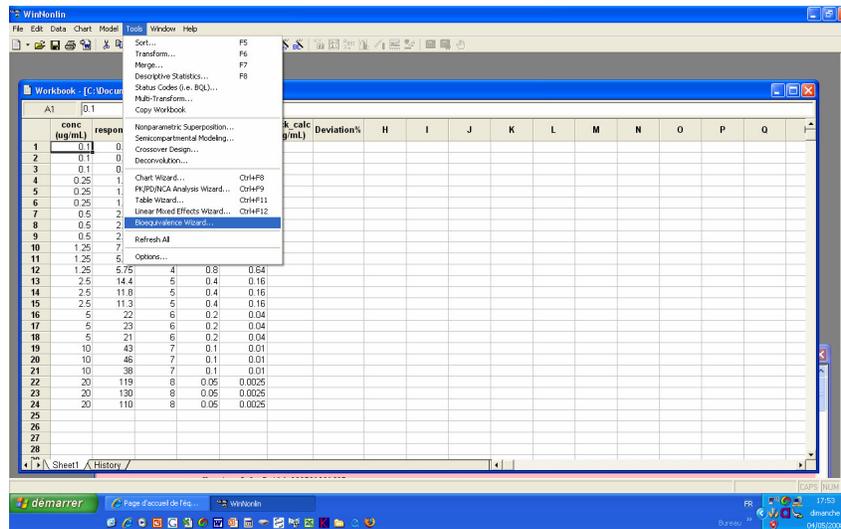
Tableau 3.4 : Tableau de l'ANOVA pour effectuer le test de non-linéarité

Analysis of variance					
Source	Sum of squares	Df	Mean square	F-ratio	Prob.level
Model	33649.849	1	33649.849	753.97	000
Residual	981.86720	22	44.63033		
Lack-of-fit	739.52974	6	123.25496	8.1377	0.0038
Pure error	242.33747	16	15.14609		

3.4.2. Test du lack of fit pour une droite non pondérée avec WinNonlin

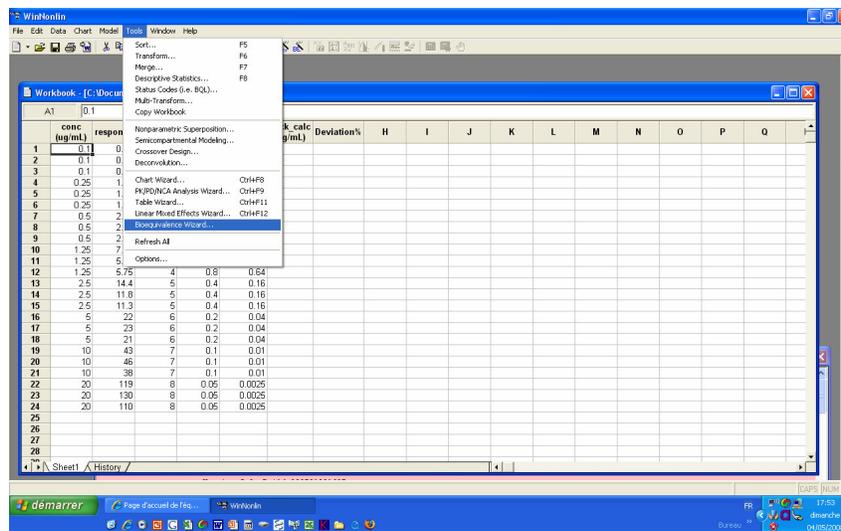
Après avoir ajusté les données du tableau 3.1. à une droite non pondérée, WN donne directement dans la feuille "Diagnostic" la Residual SS" de 981.867 qui est la SS_{totale} (erreur pure et celle du lack-of-fit) avec sa variance de 44.63 (fig 3.14).

Figure 3.14 :



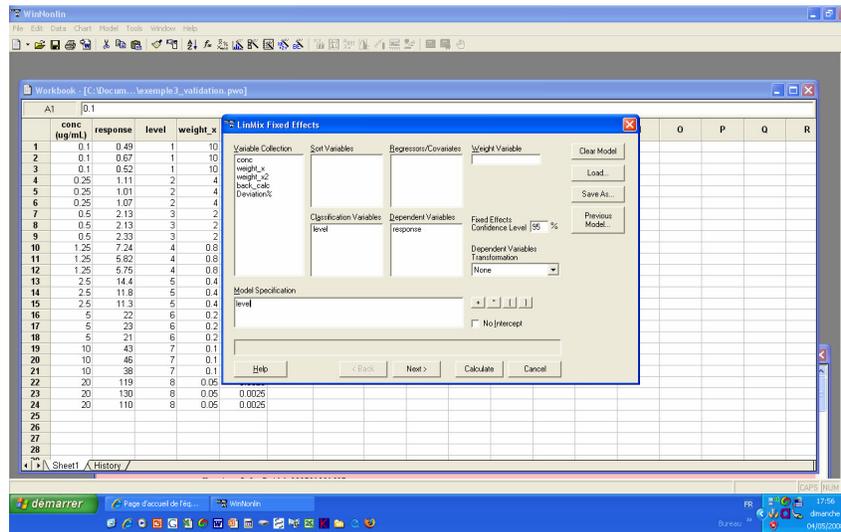
Il ne nous reste qu'à calculer l'erreur pure. Pour cela, on utilisera l'outil WinNonlin "*Linear Mixed Effect Wizard*" ou CTRL+F12. Après avoir fermé, le cas échéant, le modèle en cours et ouvrir l'outil avec "*Tool > Linear Mixed Effect Wizard*" (fig 3.15).

Figure 3.15 : Ouverture du module d'ANOVA de WinNonlin.



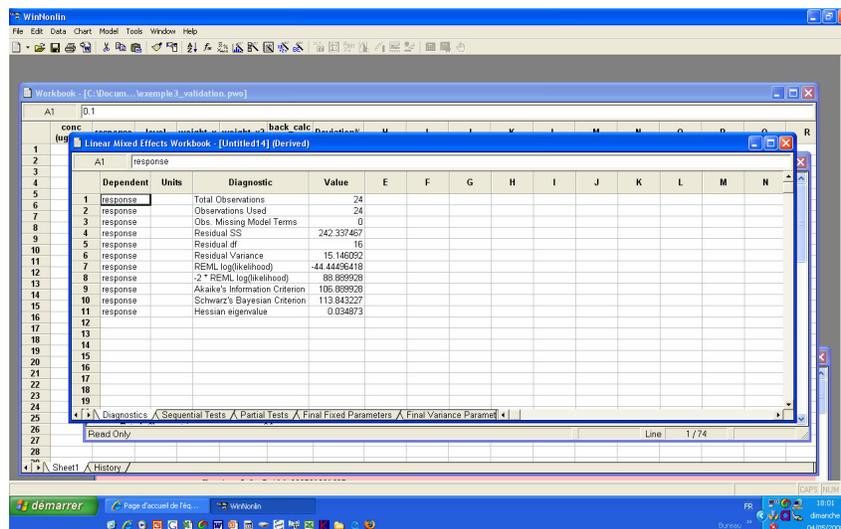
Après avoir ouvert ce module, faire glisser dans la fenêtre "*Classification*" "*level*" et dans "*Dependent variable*" le vecteur "*Response*". Ensuite et ensuite seulement faire glisser "*Level*" dans "*Model specification*" pour préciser qu'il s'agit d'une simple ANOVA à un facteur (level) (Fig. 3.16).

Figure 3.16 : Construction du modèle d'ANOVA a un facteur dans WinNonlin pour estimer la variance de l'erreur pure.



Puis lancer les calculs avec "Calculate". On obtient directement la "SS residual" = 242.33 (fig. 3.17).

Figure 3.17 : fenêtre de WinNonlin donnant la SS des résiduelles et la variance des résiduelles (15.14)



Après avoir fait "Calculate" la première feuille de résultats (c'est-à-dire celle dont l'onglet est intitulé diagnostic) donne la résiduelle pure qui est 242.233.

Ensuite pour faire le test du lack of fit, on procèdera comme cela a été indiqué précédemment avec les calculs manuels.

3.4.3. Test du lack of fit pour une droite pondérée par $1/X^2$

Après avoir ajusté les données avec une droite pondérée, le test du lack of fit se fera selon la même procédure mais il faudra tenir compte de la pondération pour calculer les différents termes d'erreur.

Le terme d'erreur pure est obtenu en faisant l'ANOVA (avec la pondération par Weight_X2 (fig 3.18).

Dans WN on peut normaliser (scaling) les poids de telle façon que le total soit égal au nombre des données. Cela ne modifie pas l'ajustement et augmente la stabilité des calculs.

Figure 3.18 : fenêtre permettant de déclarer la pondération, on notera que la case "No scaling of weights" doit être cochée pour obtenir une résiduelle appropriée (non standardisée par WinNonlin).

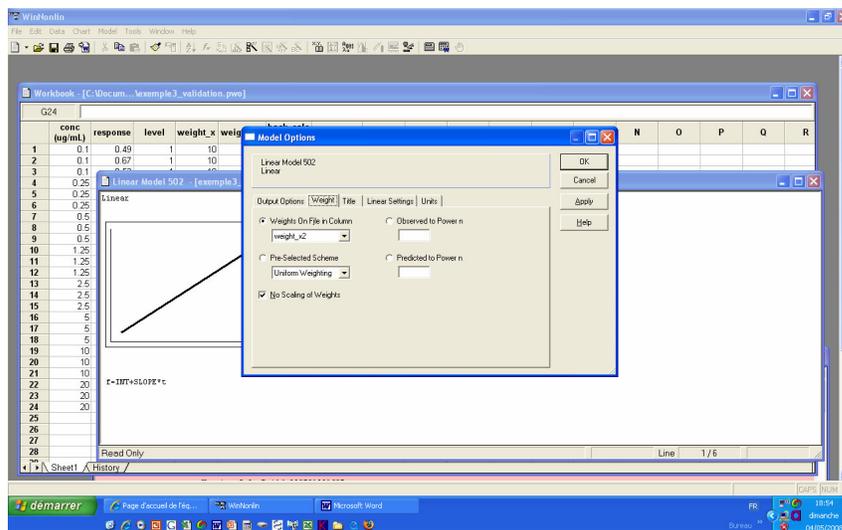


Figure 3.19 : fenêtre pour le calcul de l'erreur pure pour une régression avec une pondération de $1/X^2$. On doit d'abord déclarer les variables (Weight_2 comme regressors/covariates) ensuite et ensuite seulement, glisser le vecteur des "Weight_variable". On doit déclarer "Level" comme classification variables et "Response" comme "Dependent variable". Le modèle est à un facteur (level) à glisser dans "Model specification" puis faire "Calculate".

Déclaration de "weight" "weight" comme une variable

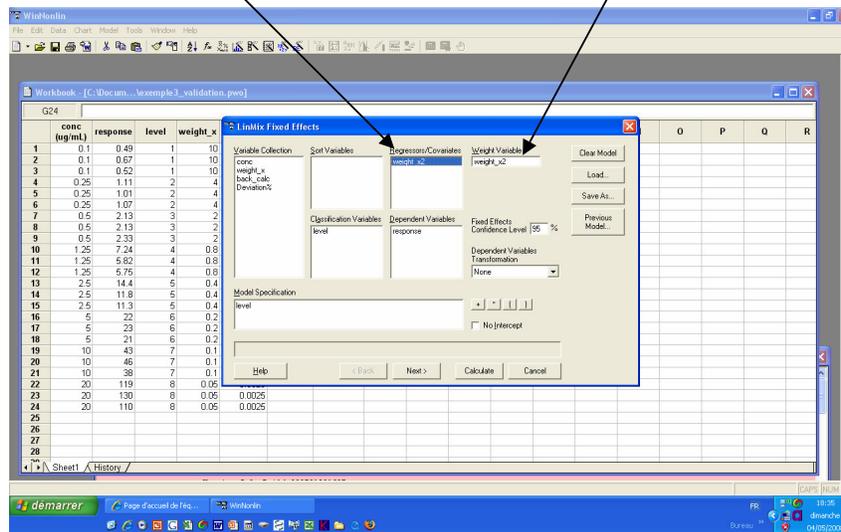
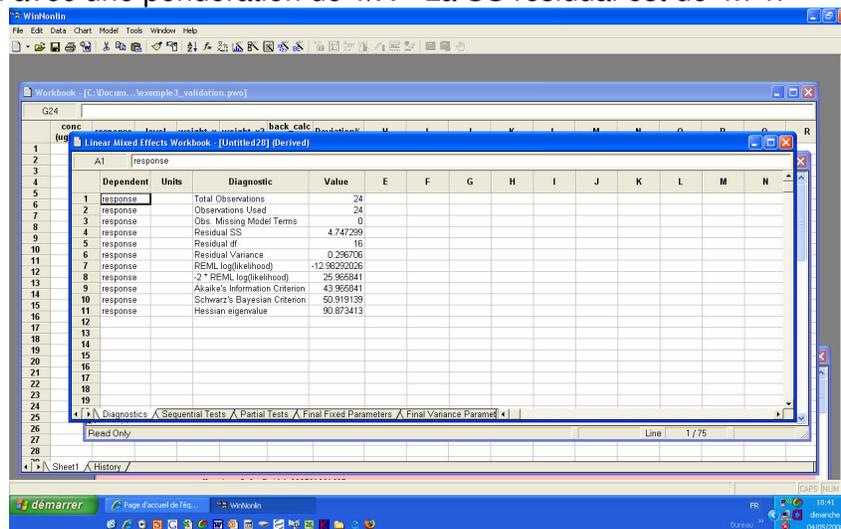


Figure 3.20 : Fenêtre donnant le résultat pour le calcul de l'erreur pure pour une régression avec une pondération de $1/X^2$. La SS residual est de 4.747



Nous allons maintenant estimer la SS totale en faisant une régression pondérée par $1/X^2$. Pour cela nous allons glisser le vecteur "Weight_X2" dans la boîte "Weight on file in colum" et bien vérifier que la case "No Scaling of Weights" est bien cochée (Attention ! jusqu'alors, tous nos calculs ont été réalisés avec des résidus normalisés).

Le calcul de la droite de régression avec une pondération par $1/X^2$ donne une résiduelle de 13.25 pour 22 ddl (fig. 3.21).

Les paramètres de la droite de régression sont obtenus dans la feuille "Final fixed parameters" avec une pente de 4.75 et une ordonnée à l'origine de 0.0501 (fig 3.22).

Nous pouvons faire un test de non-linéarité avec cette droite de régression pondérée, la SS lack-of-fit est de :

$$SS_{lack-of-fit} = SS_{totale} - SS_{pure_error}$$

$$SS_{lack-of-fit} = 13.2584 - 4.7473 = 8.5111$$

et les variances correspondantes sont :

$$\text{Variance résiduelle} = \frac{4.7473}{16} = 0.2967$$

$$\text{Variance lack - of - fit} = \frac{8.5111}{6} = 1.4185$$

$$\text{et le test } F = \frac{1.4185}{0.2967} = 4.781$$

avec $F_{16}^6 = 2.74$ pour $P=0.05$ et 4.20 pour $P=0.01$. L'hypothèse de linéarité est donc rejetée pour $P<0.01$.

Le test de non-linéarité est donc significatif et la droite doit être rejetée.

Face à ce rejet, on peut soit tenter d'ajuster les points avec une courbe c'est-à-dire introduire une composante quadratique soit réduire l'étendue de la gamme.

Nous allons dans la section suivante ajuster les données du tableau 3.1 avec une courbe c'est-à-dire un polynôme du second degré.

Figure 3.21 : Fenêtre donnant le résultat pour le calcul de l'erreur résiduelle pour une régression avec une pondération de $1/X^2$. La SS residual est de 13.258 et elle correspond à la somme de l'erreur pure et du lack of fit

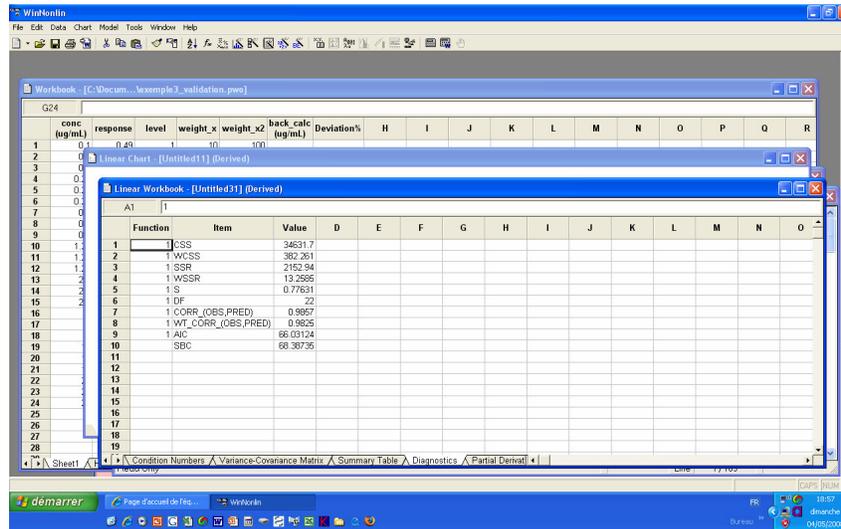
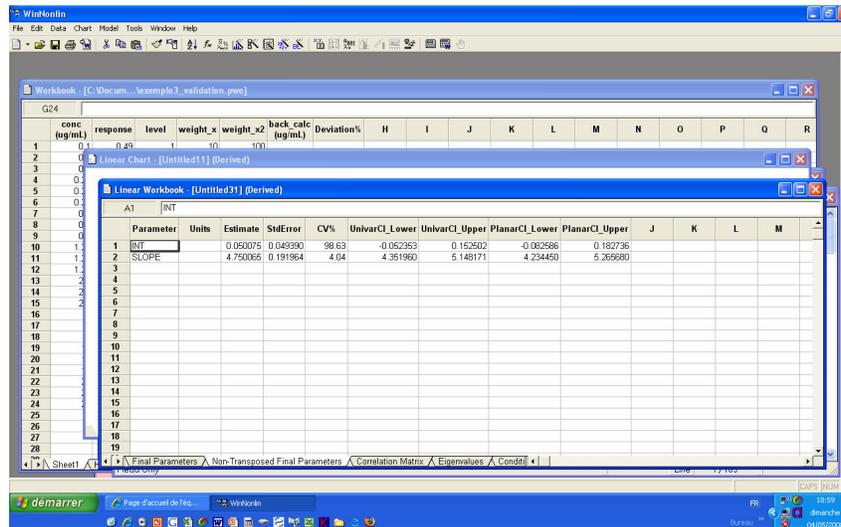
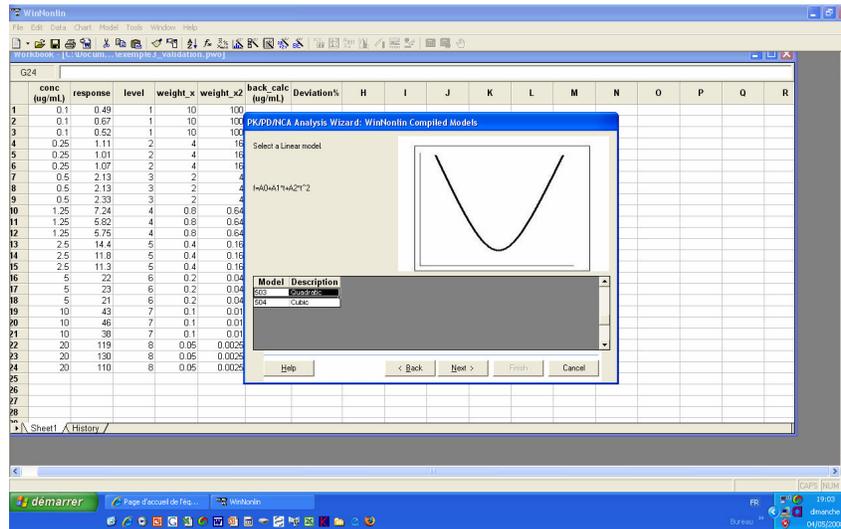


Figure 3.22 : fenêtre donnant les paramètres de la droite de régression avec une pondération de $1/X^2$. la pente est de 4.7501 et l'intercept de 0.051



Pour réaliser un ajustement avec une composante quadratique, on reprend le même module et on sélectionne le modèle quadratique comme indiqué sur la figure 3.23. Le poids sélectionné sera $1/X$; la case "No scaling of Weights" sera cochée.

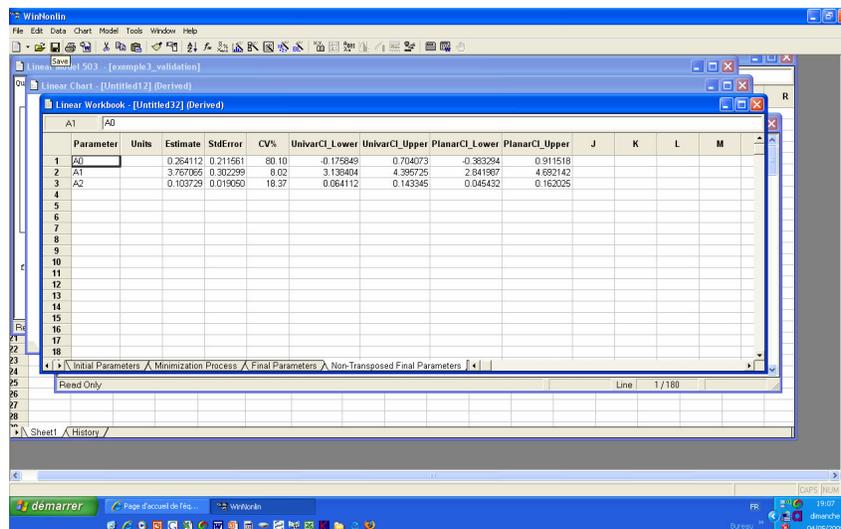
Figure 3.23 : Fenêtre montrant comment monter le modèle avec une composante quadratique et une pondération de 1/X



Les paramètres de la courbe sont :

$$Y = 3.761 X + 0.1037 X^2 + 0.2641 \text{ (fig 3.24).}$$

Figure 3.24 : Fenêtre donnant les paramètres d'un ajustement impliquant une composante quadratique et une pondération de 1/X (A0=intercept, A1 = pente de la composante linéaire ; A2 = pente de la composante quadratique). La résiduelle est de 38.2742 (voir l'onglet "Diagnostic") et la variance résiduelle de 1.822. Il s'agit de la "Residual SS" correspondant à la somme totale des carrés résiduels (lack of fit et pure error).



Nous pourrions à nouveau faire un test de "Linearité" par rapport à cette équation. Une alternative est d'inspecter l'intervalle de confiance de la pente de la composante quadratique (entre 0.045 et 0.162 voir la figure 3.24). Ce qui veut dire que la composante quadratique est significative car l'IC exclut la valeur zéro. Une approche plus simple consiste à faire une évaluation visuelle à la fois de l'ajustement (fig. 3.25) et des résidus pondérés (3.26).

Figure 3.25 : Données du tableau 3.1 ajustées avec une courbe ayant une composante linéaire et quadratique. L'inspection de la figure suggère un bon ajustement par rapport à celui de la figure 3.9.

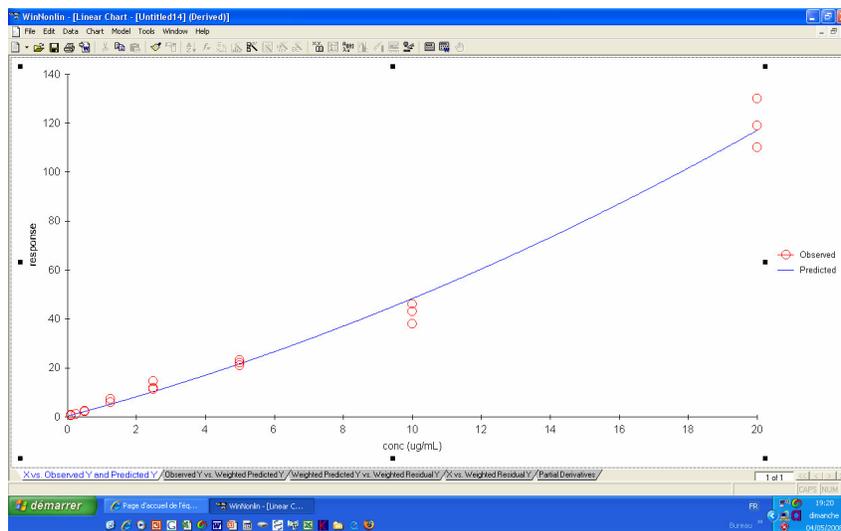
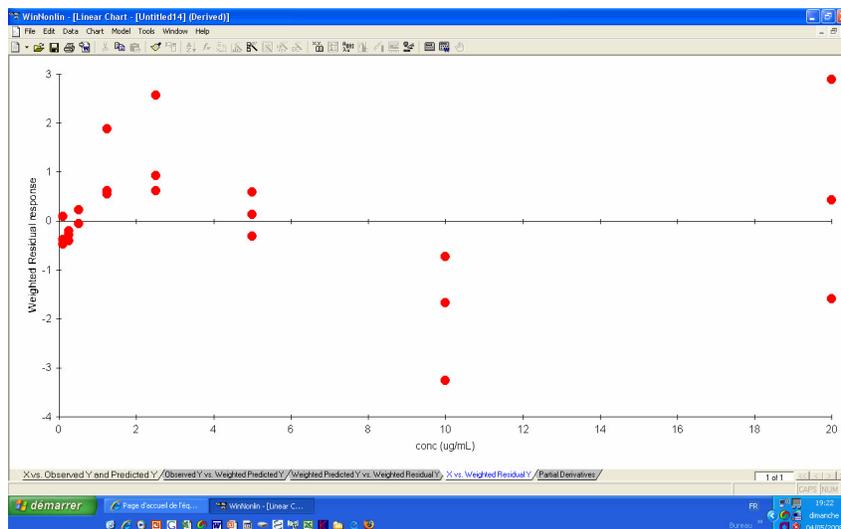


Figure 3.26 : Répartition des résidus pondérés par un ajustement impliquant une composante quadratique. Il apparaît que la répartition des résidus est plutôt satisfaisante (surtout par rapport à ce qui a été vu précédemment sur la fig. 3.10).



L'inspection des résidus montre une amélioration apportée avec la composante quadratique (par rapport à la figure 3.10) même s'il reste encore une forme de banane. **C'est donc la back-calculation qui dira si on peut ou non garder cette courbe. Pour cela on doit trouver les racines d'une équation du second degré c'est-à-dire résoudre l'équation :**

$$Y = aX^2 + bX + c \text{ ou encore } aX^2 + bX + c - Y = 0$$

Soit ici $Y = 0.1037X^2 + 3.7671 X + 0.2641$

Rappelons que le déterminant est :
$$\frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

La figure 3.27 donne les valeurs calculées par étalonnage inverse. On constate que certaines d'entre elles dépassent largement les 20% en terme de déviation.

$$X = \text{solution} = \frac{-b \pm \sqrt{b^2 - 4a \times (c - y)}}{2a}$$

Par exemple pour la réponse 0.49, la concentration prédite est :

$$X = \frac{-3.761 \pm \sqrt{3.761^2 + 4 \times 0.1037 \times 0.22}}{2 \times 0.10} = 0.0599$$

On pourra utiliser pour ces calculs soit une feuille EXCEL soit une feuille de WinNonlin après avoir effectué la commande "Detach".

Figure 3.27 : concentrations calculées par étalonnage inverse par le modèle avec une composante quadratique et une pondération de 1/X

	Dependent	conc	réponse	weighting	Observed	Predicted	c_moins_y	discriminant	back_calc	deviation	K	L	M
1	réponse	0.1	0.49	10	0.49	0.6418556671	-0.2259	3.7795166	0.0596679	-40.13211			
2	réponse	0.1	0.67	10	0.67	0.6418556671	-0.4059	3.78938118	0.107431	7.4309654			
3	réponse	0.1	0.52	10	0.52	0.6418556671	-0.2569	3.781162484	0.0678037	-32.19632			
4	réponse	0.25	1.11	4	1.11	1.2123611	-0.8459	3.813387173	0.2231783	-10.72869			
5	réponse	0.25	1.01	4	1.01	1.2123611	-0.7459	3.807944554	0.1969361	-21.22555			
6	réponse	0.25	1.07	4	1.07	1.2123611	-0.8059	3.811211058	0.2126859	-14.92564			
7	réponse	0.5	2.13	2	2.13	2.173576349	-1.8659	3.868464518	0.4887392	-2.252152			
8	réponse	0.5	2.13	2	2.13	2.173576349	-1.8659	3.868464518	0.4887392	-2.252152			
9	réponse	0.5	2.33	2	2.33	2.173576349	-2.0659	3.8791723	0.5403679	8.0735772			
10	réponse	1.25	7.24	0.8	7.24	5.135018548	-6.9759	4.133257882	1.7659493	41.275966			
11	réponse	1.25	5.82	0.8	5.82	5.135018548	-5.5559	4.061481224	1.4193887	13.551099			
12	réponse	1.25	5.75	0.8	5.75	5.135018548	-5.4859	4.057905081	1.4021461	12.171684			
13	réponse	2.5	14.4	0.4	14.4	10.33007705	-14.1359	4.478237793	3.4288225	37.152901			
14	réponse	2.5	11.8	0.4	11.8	10.33007705	-11.5359	4.396160434	2.8402142	13.60857			
15	réponse	2.5	11.3	0.4	11.3	10.33007705	-11.0359	4.332289664	2.7251189	9.0047567			
16	réponse	5	22	0.2	22	21.69264971	-21.7359	4.817374153	5.0640027	1.2800533			
17	réponse	5	23	0.2	23	21.69264971	-22.7359	4.860236975	5.2706653	5.4133052			
18	réponse	5	21	0.2	21	21.69264971	-20.7359	4.774127536	4.8554847	-2.880305			
19	réponse	10	43	0.1	43	48.30761764	-42.7359	5.649592351	9.0766266	-9.233734			
20	réponse	10	46	0.1	46	48.30761764	-45.7359	5.759571177	9.6029511	-3.974389			
21	réponse	10	38	0.1	38	48.30761764	-37.7359	5.462956501	8.175743	-18.23257			
22	réponse	20	119	0.05	119	117.0968439	-118.7359	7.965092198	20.241042	1.2052121			
23	réponse	20	130	0.05	130	117.0968439	-129.7359	8.246544351	21.598092	7.9904617			
24	réponse	20	110	0.05	110	117.0968439	-109.7359	7.72719184	19.093982	-4.530091			
25													
26													
27													
28													
29													
30													
31													

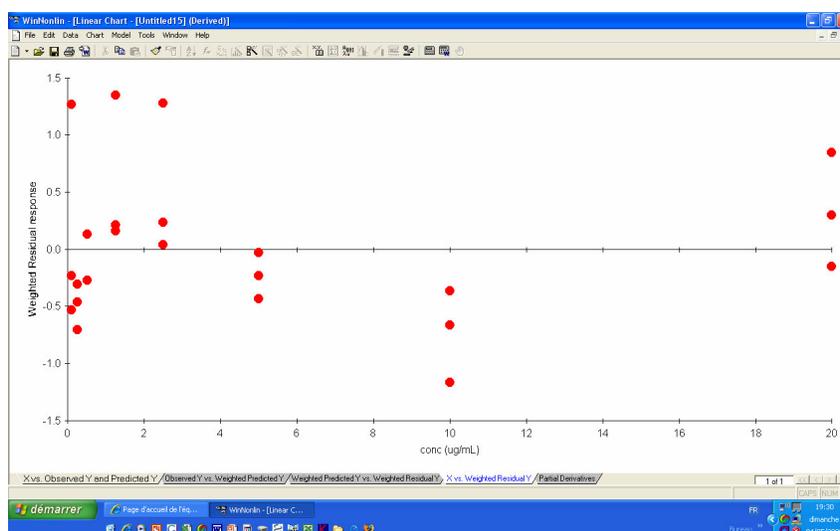
Cela nous incite à explorer la pondération de $1/X^2$. La variance résiduelle est de 0.45 avec une Residual SS de 9.5056.

Les paramètres estimés sont : $Y = 0.1163 X^2 + 4.2625 X + 0.0691$

Nous allons inspecter les résidus pondérés (fig. 3.28). Ils sont ici présentés en %.

L'inspection de la figure 3.28 donne une répartition des résidus similaire à celle qui avait été obtenue avec une pondération de $1/X$.

Figure 3.28 : résidus obtenus par le modèle avec une composante quadratique et une pondération de $1/X^2$



Pour juger de l'intérêt de cette pondération nous allons procéder au calcul par étalonnage inverse en repartant de la feuille obtenue avec "Summary table". Après avoir effectué un "Detach", la racine pour la réponse observée de 0.49 est de 0.0984. Pour la valeur observée de 110, la racine est de 17.466, pour une concentration nominale de 20 soit une déviation de 12.67%

La figure 3.29 donne les résultats. Il apparaît que certaines déviations dépassent les 20% et cette courbe quadratique n'est pas acceptable. La solution va consister à en réduire l'étendue.

Figure 3.29 : Calcul par calibration inverse (back calculation) des concentrations nominales pour une courbe avec une composante quadratique et un facteur de pondération de $1/X^2$

	level	weight_x2	conc_obs (ug/ml)	response_obs	conc (ug/ml)	response	Predicted	C_MOINS_Y	Back_calc	deviation	K	L	M	N	O	P	Q
1	1	100	0.1	0.49	0.1000	0.4900	0.6419	-0.4209	0.0984025353	1.519746465							
2	1	100	0.1	0.67	0.1000	0.6700	0.6419	-0.6009	0.1404354994	-40.43549943							
3	1	100	0.1	0.62	0.1000	0.5200	0.6419	-0.4509	0.1054794268	-5.479426798							
4	2	16	0.25	1.11	0.2500	1.1100	1.2124	-1.0409	0.242593677	2.962529201							
5	2	16	0.25	1.01	0.2500	1.0100	1.2124	-0.9409	0.219425252	12.2298699							
6	2	16	0.25	1.07	0.2500	1.0700	1.2124	-1.0009	0.233329607	6.688077185							
7	3	4	0.5	2.13	0.5000	2.1300	2.1736	-2.0609	0.4772802977	4.543940454							
8	3	4	0.5	2.13	0.5000	2.1300	2.1736	-2.0609	0.4772802977	4.543940454							
9	3	4	0.5	2.33	0.5000	2.3300	2.1736	-2.2609	0.5229546268	-4.590925358							
10	4	0.64	1.25	7.24	1.2500	7.2400	5.1350	-7.1709	1.611463895	-28.91755982							
11	4	0.64	1.25	5.82	1.2500	5.8200	5.1350	-5.7509	1.302970183	-4.229814657							
12	4	0.64	1.25	5.75	1.2500	5.7500	5.1350	-5.6809	1.287531966	-3.00255720							
13	5	0.16	2.5	14.4	2.5000	14.4000	10.3301	-14.2309	3.099901075	-23.99604298							
14	5	0.16	2.5	11.8	2.5000	11.8000	10.3301	-11.2309	2.571671674	-2.868899443							
15	5	0.16	2.5	11.3	2.5000	11.3000	10.3301	-11.2309	2.468550779	1.257968843							
16	6	0.04	5	22	5.0000	22.0000	21.6926	-21.9309	4.574198908	8.516029836							
17	6	0.04	5	23	5.0000	23.0000	21.6926	-22.9309	4.761177168	4.776496833							
18	6	0.04	5	21	5.0000	21.0000	21.6926	-20.9309	4.385680532	12.286389395							
19	7	0.01	10	43	10.0000	43.0000	43.3076	-42.9309	8.225690003	17.74348897							
20	7	0.01	10	46	10.0000	46.0000	43.3076	-45.9309	8.707627395	12.9237285							
21	7	0.01	10	38	10.0000	38.0000	43.3076	-37.9309	7.402510295	-25.9688824							
22	8	0.0025	20	119	20.0000	119.0000	117.0968	-118.9309	18.53162169	7.341891573							
23	8	0.0025	20	130	20.0000	130.0000	117.0968	-128.9309	19.79313739	1.034313044							
24	8	0.0025	20	110	20.0000	110.0000	117.0968	-108.9309	17.46641592	12.66732042							

Nous allons exclure les données obtenues avec la concentration 20 pour ne faire les calculs que sur une étendue de 0.1 à 10 (fig. 3.30).

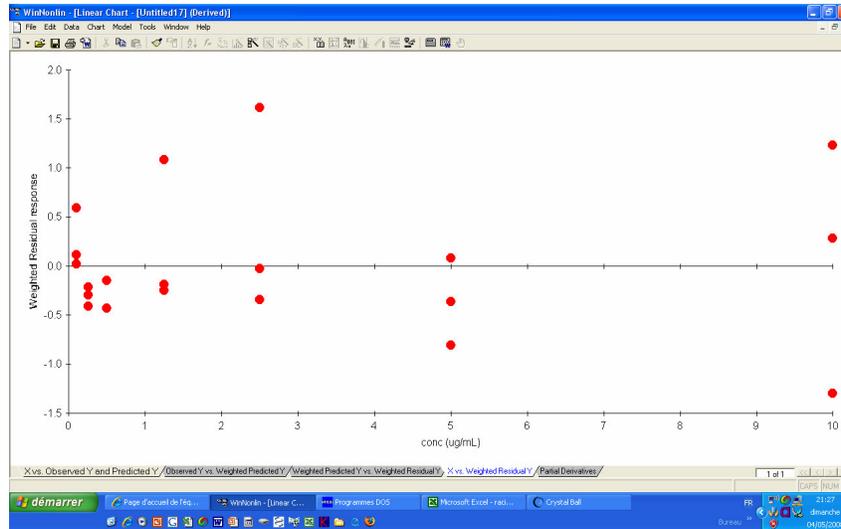
Figure 3.30 : sélection des données pour calculer une courbe de calibration "abrégée". Pour limiter les valeurs à 10, aller dans "Data" et faire "Exclude > selection"

	conc (ug/ml)	response	level	weight_x	weight_x2	back_calc (ug/ml)	Deviation%	H	I	J	K	L	M	N	O	P	Q	R	S
1	0.1	0.49	1	10	100														
2	0.1	0.67	1	10	100														
3	0.1	0.62	1	10	100														
4	0.25	1.11	2	4	16														
5	0.25	1.01	2	4	16														
6	0.25	1.07	2	4	16														
7	0.5	2.13	3	2	4														
8	0.5	2.13	3	2	4														
9	0.5	2.33	3	2	4														
10	1.25	7.24	4	0.8	0.64														
11	1.25	5.82	4	0.8	0.64														
12	1.25	5.75	4	0.8	0.64														
13	2.5	14.4	5	0.4	0.16														
14	2.5	11.8	5	0.4	0.16														
15	2.5	11.3	5	0.4	0.16														
16	5	22	6	0.2	0.04														
17	5	23	6	0.2	0.04														
18	5	21	6	0.2	0.04														
19	10	43	7	0.1	0.01														
20	10	46	7	0.1	0.01														
21	10	38	7	0.1	0.01														
22	20	119	8	0.05	0.0025														
23	20	130	8	0.05	0.0025														
24	20	110	8	0.05	0.0025														

Recalculer avec un facteur de pondération $1/X$

L'inspection des résidus (fig 3.31) indique une bonne répartition des résidus.

Figure 3.31 : Répartition des résidus pour une courbe de calibration avec une composante quadratique obtenue pour les concentrations nominales allant de 0.1 à 10, avec un facteur de pondération de 1/X et une composante quadratique.

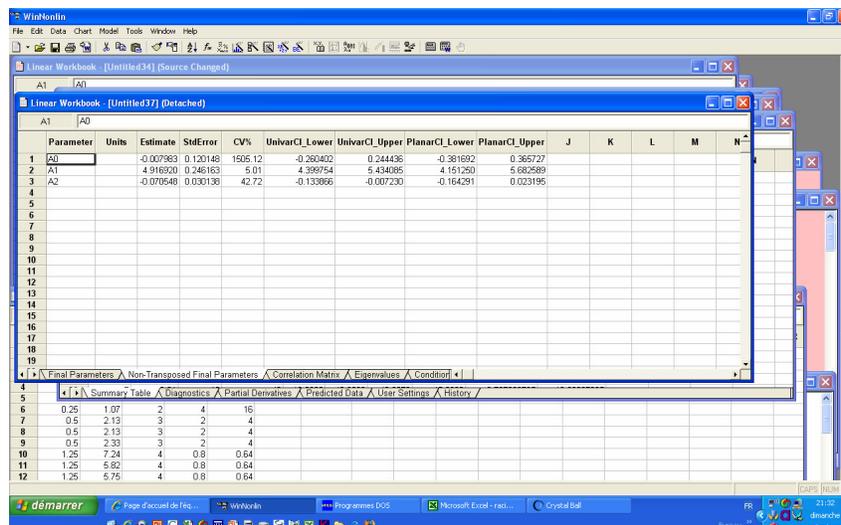


L'équation prédite est :

$$Y = -0.007983X^2 + 4.91629 - 0.070548$$

L'inspection des intervalles de confiance de la pente pour la composante quadratique (0.007983) n'est pas significative (fig. 3.32), ce qui suggère qu'une simple droite peut être adéquate.

Figure 3.32 : Fenêtre donnant les résultats de l'ajustement des données allant de 0.1 à 10 avec une composante quadratique et pondérées par 1/Y²



Avec le modèle linéaire simple et une pondération de $1/X^2$, la pente est de 4.476 et l'ordonnée à l'origine de 0.089185.

La figure 3.33 montre la répartition des résidus qui semble adéquate.

Les back-calculations donnent dans l'ensemble des résultats acceptables seules 2 valeurs sur les 21 dépassant la marge des 20%. Il est probable que l'analyste gardera cette courbe car les déviations sont vues pour deux niveaux différents de concentrations (0.8 et 10).

Figure 3.33 : Répartition des résidus pour un ajustement des données du tableau 3.1 (après avoir supprimé les données de la concentration 20). Les données ont été ajustées avec une simple droite et un facteur de pondération de $1/X^2$.

